Mario Luzardo
Diego Forteza

Modelo no paramétrico multidimensional

para la estimación de los rasgos y de las curvas características del ítem mediante regresión no paramétrica con núcleos

UNIVERSIDAD DE LA REPUBLICA
URUGUAY
CSIC
bibliotecaplural

Mario Luzardo

●

Diego Forteza

Modelo no paramétrico multidimensional para la estimación de los rasgos y de las curvas características del ítem mediante regresión no paramétrica con núcleos

Tabla de contenido

1. Estimación no paramétrica de densidades 1
   1.1. Introducción . . . 1
   1.2. Distribución asintótica del estimador . . . 9
   1.3. Convergencia completa de los estimadores basados en núcleos . . . 14
   1.4. Regresión no paramétrica . . . 17
   1.5. Distribución asintótica . . . 30

2. Teoría de respuesta al ítem 35
   2.1. Introducción . . . 35
   2.2. Dimensionalidad del espacio latente . . . 35
   2.3. Independencia local . . . 36
   2.4. Ausencia de factores de velocidad . . . 37
   2.5. Curva característica del ítem . . . 37
   2.6. Clasificación de los modelos . . . 38
   2.7. Modelos para ítems dicotómicos . . . 39
      2.7.1. Modelo de ojiva normal . . . 39
   2.8. Modelos logísticos . . . 41
      2.8.1. Logits . . . 42
   2.9. Escala de θ . . . 42
      2.9.1. Modelo logístico de un parámetro . . . 43
      2.9.2. Modelo logístico de dos parámetros . . . 43
      2.9.3. Modelo logístico de tres parámetros . . . 44
   2.10. Modelo de Rasch . . . 44
   2.11. Función de información . . . 49
   2.12. Estimación . . . 53
      2.12.1. Método de Newton Raphson . . . 53
      2.12.2. Estimación por máxima verosimilitud . . . 55
      2.12.3. Algoritmo EM . . . 56
      2.12.4. Revisión de los procedimientos . . . 58
   2.13. Estimación de los parámetros de la curva característica del ítem conocida la habilidad . . . 60
      2.13.1. Estimación del modelo de ojiva normal por máxima verosimilitud 60
      2.13.2. Estimación del modelo logístico de dos parámetros por máxima verosimilitud . . . 68
      2.13.3. Estimación del modelo logístico de tres parámetros por máxima verosimilitud . . . 72
      2.13.4. Estimación por mínimo χ^2 . . . 75

2.14. Estimación de la habilidad cuando
se conocen los parámetros de los ítems 76
2.14.1. Estimación de la habilidad por máxima verosimilitud 76
2.14.2. Modelo de ojiva normal 76
2.14.3. Modelo logístico de dos parámetros 78
2.14.4. Modelo logístico de tres parámetros 79
2.15. Estimación conjunta de parámetros del ítem
y habilidad mediante máxima verosimilitud 81
2.16. Estimación del modelo de Rasch por máxima verosimilitud condicional . 83
2.17. Estimación del modelo de Rasch por máxima verosimilitud conjunta . . . 88
2.18. Estimación vía máxima verosimilitud marginal 93
2.18.1. La solución de Bock y Lieberman 94
2.18.2. La solución de Bock y Aitkin 98
2.19. Estimación bayesiana 103
2.19.1. Estimación marginal bayesiana
de los parámetros del ítem en Pc-Bilog 104
2.19.2. Estimación bayesiana del rasgo 108
2.20. Comprobación de los modelos 110
2.20.1. Pruebas basadas en chi-cuadrado 111
2.20.2. Pruebas basadas en la verosimilitud 115
2.20.3. Análisis de residuos 116
2.20.4. Bondad de ajuste del modelo 121
2.21. Modelos no paramétricos clásicos 125
2.21.1. Modelo monótono homogéneo 125
2.21.2. Modelo de doble monoticidad 127
2.21.3. Estimación de la habilidad 128
2.22. Modelo no paramétrico unidimensional
usando regresión no paramétrica 129
2.22.1. Hipótesis 135
2.22.2. Consistencia 136
2.23. Modelos paramétricos multidimensionales 136

3. Modelo no paramétrico multidimensional 143
3.1. Introducción 143
3.2. Estimación monótona 172
3.3. Aplicación al funcionamiento diferencial del ítem 174
3.3.1. Definiciones básicas 174
3.3.2. Técnicas estadísticas para la detección 174
3.3.3. Procesos iterativos 182
3.3.4. Implementación de decisiones 183

3.3.5. Método propuesto 184

4. Simulaciones 187
4.1. Distancia entre curvas características del ítem 187
4.2. Generación de datos 187
4.3. Ajuste de la regresión no paramétrica unidimensional a modelos conocidos y rasgos conocidos 189
4.3.1. Método 189
4.3.2. Resultados 190
4.4. Estudio del modelo unidimensional no paramétrico para rasgo conocido . 193
4.4.1. Método 194
4.4.2. Código de las funciones utilizadas 194
4.4.3. Resultados 197
4.5. Modelo no paramétrico unidimensional con CCI y rasgo desconocido . . . 205
4.5.1. Tamaño muestral 205
4.5.2. Efecto de distintos valores de ancho para la ventana 206
4.5.3. Distintas distribuciones para el parámetro habilidad 207
4.5.4. Efecto de distintos núcleos en la regresión no paramétrica 209
4.5.5. Comparación entre distintos modelos y métodos de estimación . . 210
4.5.6. Otros resultados 212
4.6. Simulaciones en dos dimensiones, ítems dependientes de una sola habilidad 216
4.6.1. Efecto del tamaño muestral 216
4.6.2. Efecto de distintos valores de ancho de ventana 219
4.6.3. Efecto de distintas cópulas para el parámetro habilidad 221
4.6.4. Distintos niveles de correlación entre las habilidades 223
4.6.5. Efecto de distintos núcleos en la regresión no paramétrica 225
4.6.6. Comparación entre valores estimados y reales del parámetro habilidad226
4.7. Simulaciones en dos dimensiones con habilidades dependientes 228
4.7.1. Efecto del tamaño muestral 229
4.7.2. Efecto de distintos anchos de ventana 231
4.7.3. Efecto de distintas distribuciones de θ 233
4.7.4. Efecto de distintos grados de correlación entre las habilidades . . . 235
4.7.5. Efecto de distintos núcleos 237
4.7.6. Comparación entre valores estimados y valores reales del parámetro habilidad 238
4.8. Estimación monótona 240
4.8.1. Tamaño muestral 241
4.8.2. Efecto de distintos valores de ancho para la ventana 243
4.9. Aplicación a datos reales: test CDI 244
4.9.1. Estimación de las CCI diferenciando por rasgo 245

4.9.2. Estudio diferencial del ítem 248
4.9.3. Análisis factorial 249
4.10. Funciones 251
4.10.1. Funciones para calcular las distancias 251
4.10.2. Funciones para la generación de datos 252
4.10.3. Funciones para la estimación de la CCI 255
4.10.4. Distintos núcleos para la regresión no paramétrica 258
4.10.5. Función para transformar el rango 260
4.10.6. Funciones utilizadas para el ánalisis del test CDI 260
4.10.7. Función para transformación monótona creciente 262

Apéndice 262

A. 263
A.1. Modelo de un parámetro 263
A.1.1. Distribución normal 263
A.1.2. Distribución uniforme 266
A.1.3. Distribución t 269
A.1.4. Distribución Beta 270
A.1.5. Distribución Gamma 274
A.2. Modelo de dos parámetros 275
A.3. Modelo de tres parámetros 278
A.4. Modelo logit cúbico 281

Referencias 285

Prefacio

Presentamos en este libro los resultados obtenidos en la investigación "Estimación del rasgo y las curvas características del ítem en el caso multivariado mediante regresión no paramétrica con núcleos" financiada por el Fondo Profesor Clemente Estable en la convocatoria 63 del Plan de Desarrollo Tecnológico Subprograma II, componente B, que se desarrolló entre los años 2007 y 2009.

Exponemos un procedimiento no paramétrico de estimación del rasgo y de las curvas características del ítem (CCI) dependientes de un rasgo latente multidimensional, utilizando para ello el método de regresión no paramétrica con núcleos. Se consideran N sujetos que responden n ítems (dicotómicos, politómicos o de respuesta continua), donde la respuesta a cada uno depende de un rasgo latente multidimensional. Daremos estimadores del rasgo y de las CCI. Además, se encontrarán las condiciones que deben cumplir las CCI, los núcleos, el ancho de la ventana y las funciones que se utilizarán para ordenar a los sujetos para obtener estimadores consistentes cuando N y n tienden a infinito conjuntamente. También obtendremos la distribución asintótica de los estimadores de las CCI. Obtuvimos algoritmos sencillos de programar, no iterativos y más eficientes que los que se utilizan en los modelos paramétricos. Daremos también una estimación monótona de las CCI, para cuando se debe asumir este supuesto.

Mostraremos que este método funciona mejor bajo condiciones generales que los modelos de TRI clásicos, tanto en las simulaciones como en un caso real y para las posibles aplicaciones a tests adaptativos informatizados (TAIs) y funcionamiento diferencial del ítem (DIF).

El capítulo I trata sobre los métodos no paramétricos para la estimación de densidades y regresión. En el capítulo II se analizan los procedimientos paramétricos y no paramétricos desarrollados hasta el momento poniendo especial énfasis en la estimación para mostrar la complejidad de los algoritmos usuales. En el capítulo III se presenta el modelo multidimensional, se demuestra la consistencia conjunta de los estimadores, se expone el estimador monótono y un método para la determinación del funcionamiento diferencial del ítem. Finalmente el capítulo IV está dedicado a las simulaciones, al análisis de datos reales y al código R de las funciones utilizadas.

Los autores queremos agradecer especialmente a Julio Olea y a Ricardo Fraiman, docentes de la Facultad de Psicología de la Universidad Autónoma de Madrid y de la Facultad de Ciencias de la Universidad de la República respectivamente, por su continuo apoyo y asesoria en esta investigación.

Capítulo 1

Estimación no paramétrica de densidades

1.1. Introducción

Supongamos que tenemos una muestra $X_1, ..., X_n$ de variables aleatorias independientes idénticamente distribuidas y que estas variables tienen una densidad f.

Nuestro objetivo es, a partir de esta muestra, estimar f. Para ello consideremos la familia $\Im$ de las funciones de distribución absolutamente continuas con densidad f y exploremos diferentes métodos que nos permitan encontrar f.

Un primer intento es utilizar el método de máxima verosimilitud ya que este es un método universal y asintóticamente óptimo. Lamentablemente es fácil ver que este método no sirve a menos que las funciones de la familia $\Im$ sean monótonas lipchizianas de constante fija.

Otro método sería utilizar como estimador el histograma. Para ello consideremos una serie de intervalos $I_1, ..., I_p$ disjuntos.

Definimos el histograma en el punto x como

$$f_n(x) = \frac{1}{n} \sum_{i=1}^{p} \sum_{j=1}^{n} \chi_{I_i}(X_j)\chi_{I_i}(x)$$

Si observamos este estimador nos damos cuenta que lo que estamos haciendo es contar cuántas observaciones caen en el intervalo que contiene al punto x, es decir,

$$f_n(x) = \frac{\sharp\{j : X_j \in I_k\}}{n}$$

para $x \in I_k$.

Este estimador lo podemos escribir como

$$f_n(x) = \sum_{i=1}^{p} \hat{a}_i \chi_{I_i}(x)$$

donde

$$\hat{a}_i = \frac{1}{n} \sum_{j=1}^{n} \chi_{I_i}(X_j).$$

Por la ley de los grandes números $\hat{a}_i$ converge casi seguramente a la $P(X \in I_i)$, entonces $f_n(x)$ converge casi seguramente a $\sum_{i=1}^{p} P(X \in I_i)\chi_{I_i}(x)$, luego los estimadores basados en histogramas convergen casi seguramente a histogramas lo cual no es satisfactorio ya que este hecho restringe a la familia $\Im$.

Se puede hacer una modificación de este estimador y es considerar la cantidad de intervalos como una función de la cantidad de observaciones, es decir, hacer p(n) y que esta tienda a $+\infty$ con n.

Rosenblatt (1956), Akaike (1956) y Parzen (1962) introducen un nuevo método de estimación basado en núcleos. Lo analizaremos en detalle.

El estimador basado en núcleos piensa la densidad como la derivada de una función de distribución, es decir, $f(x) = F'(x)$ donde F es la función de distribución de X.

Tenemos entonces que

$$\hat{f}_{n,h}(x) = \lim_{h \to 0} \frac{F_n(x+h) - F_n(x-h)}{2h}$$

Analizando este estimador observamos dos hechos; primero por el teorema fundamental de la estadística $F_n \xrightarrow[n]{cs} F$, segundo, el límite en el cociente incremental depende de h, es decir, se nos presenta un problema de límite doble $h \to 0$ y $n \to \infty$. Si $h \to 0$ muy rápidamente se complica la convergencia de F_n a F por lo que es conveniente hacer depender h del tamaño muestral, es decir, considerar h(n). Por lo tanto, se hace depender h del tamaño muestral o sea se considera h(n).

Con esto en mente vemos que

$$\hat{f}_{n,h}(x) = \frac{F_n(x+h(n)) - F_n(x-h(n))}{2h(n)} = \frac{F_n(x-h(n), x+h(n))}{2h(n)} =$$

$$= \frac{1}{nh(n)} \frac{\sharp\{i : X_i \in (x-h(n), x+h(n]\}}{2} = \frac{1}{nh(n)} \sum_{i=1}^{n} \frac{1}{2} \chi_{(x-h(n), x+h(n)]}(X_i)$$

$$= \frac{1}{nh(n)} \sum_{i=1}^{n} \frac{1}{2} \chi_{[-1,1]}(\frac{x - X_i}{h(n)}) = \int \frac{1}{2h(n)} \chi_{[-1,1]}(\frac{x-t}{h(n)}) dF_n(t) = \bar{K}_h * F_n^x$$

donde

$$K_h(z) = \frac{1}{2h(n)} \chi_{[-1,1]}(\frac{z}{h(n)})$$

Analizando este estimador vemos que todo se reduce a una convolución.

Teorema 1.1.1. El estimador $\hat{f}_{n,h}$ definido antes es un estimador consistente de la densidad de f en todo punto de continuidad x de f.

Demostración:

$$E(\hat{f}_{n,h}(x)) = \frac{1}{2h} E(\chi_{[-1,1]}(\frac{x - X_1}{h})) = \frac{1}{2h} E(\chi_{[x-h,x+h]}(X_1)) = \frac{1}{2h} \int_{x-h}^{x+h} f(t)dt \to f(x)$$

si $h \to 0$ por el teorema fundamental del cálculo ya que f es continua en x.

Hemos demostrado que $\hat{f}_{n,h}(x)$ es un estimador asintóticamente insesgado para x fijo y de continuidad de f.

Calculemos ahora la varianza.

$$Var(\hat{f}_{n,h}(x)) = \frac{1}{4nh^2} Var(\chi_{(x-h,x+h)}(X_1)) =$$

$$\frac{1}{4nh^2} P(X_i \in (x-h, x+h))(1 - P(X_i \in (x-h, x+h)) =$$

$$\frac{1}{2nh}(1 - P(X_i \in (x-h, x+h)) \frac{1}{2h} \int_{x-h}^{x+h} f(t)dt \longrightarrow \frac{f(x)}{2nh}$$

cuando $h \to 0$

Si además $nh \to \infty$ $Var(\hat{f}_{n,h}) \longrightarrow 0$ con lo cual $\hat{f}_{n,h}(x) \longrightarrow f(x)$ en probabilidad, lo que prueba la consistencia.

Proposición 1.1.2. Supongamos que la densidad f tiene 3 derivadas acotadas. Entonces $ECM(\hat{f}_{n,h}(x)) = \frac{|f''(x)|^2 h^4}{36} + \frac{f(x)}{2nh} + o(h^4 + \frac{1}{nh})$

Demostración: Usando el desarrollo de Taylor tenemos que

$$f(t) = f(x) + f'(x)(t-x) + f''(x)\frac{(t-x)^2}{2} + f''(\xi)\frac{(t-x)^3}{6}$$

$$E(\hat{f}_{n,h}(x)) = \frac{1}{2h} \int_{x-h}^{x+h} f(t)dt = f(x) + \frac{f'(x)}{2h} \int_{x-h}^{x+h} (t-x)dt + \frac{f''(x)}{2h} \int_{x-h}^{x+h} \frac{(t-x)^2}{2} dt +$$

$$+ \frac{1}{2h} \int_{x-h}^{x+h} f'''(\xi)\frac{(t-x)^3}{6} dt$$

Como $\int_{x-h}^{x+h} (t-x)dt = 0$

$$\int_{x-h}^{x+h} (t-x)^2 dt = \int_{-h}^{h} u^2 du = \frac{2h^3}{3}$$

y

$$\frac{1}{2h} \int_{x-h}^{x+h} f'''(\xi)\frac{(t-x)^3}{6} dt = \frac{1}{2h} \int_{-h}^{h} f'''(\xi)\frac{u^3}{6} du = \frac{1}{2h} \int_{-h}^{h} o(u^2)du = o(h^2)$$

entonces $E(\hat{f}_{n,h}(x)) = f(x) + \frac{f''(x)h^2}{6} + o(h^2)$

Por lo tanto el sesgo es:

$$(E(\hat{f}_{n,h}(x)) - f(x))^2 = (\frac{f''(x)h^2}{6} + o(h^2))^2 = \frac{|\ f''(x)\ |^2\ h^4}{36} + o(h^4)$$

Recordando que el ECM= sesgo+ varianza tenemos

$$ECM(\hat{f}_{n,h}(x)) = \frac{|\ f''(x)\ |^2\ h^4}{36} + \frac{f(x)}{2nh} + o(h^4 + \frac{1}{nh})$$

Observación: Para obtener la velocidad óptima de convergencia hacemos que $h^4 \simeq \frac{1}{nh}$ o sea $h = Cn^{-1/5}$

Con esta elección de h el ECM es del orden de $n^{-4/5}$ y la constante C adecuada la obtenemos minimizando el $ECM(\hat{f}_{n,h}(x))$

$$ECM(\hat{f}_{n,h}(x)) = \frac{|\ f''(x)\ |^2\ C^4 n^{-4/5}}{36} + \frac{f(x)}{2nn^{-1/5}C} = n^{-4/5}[\frac{|\ f''(x)\ |^2\ C^4}{36} + \frac{f(x)}{2C}]$$

$$\frac{\partial ECM(\hat{f}_{n,h}(x))}{\partial C} = \frac{C^3\ |\ f''(x)\ |^2}{9} - \frac{f(x)}{2C^2} = 0$$

de donde $C = (\frac{9f(x)}{2|f''(x)|^2})^{1/5}$

Observemos que al sesgo le conviene que la ventana sea muy chica en tanto a la varianza que sea grande , es decir, existe un compromiso permanente entre sesgo y varianza.

Recordemos que $\hat{f}_{n,h}(x)$ se reduce a una convolución $\hat{f}_{n,h} = \bar{K}_h * F_n$ donde $\bar{K}_h(z) = \frac{1}{2h}\chi_{[-1,1]}(z/h)$.

Cabe preguntarse bajo qué condiciones puede reemplazarse la indicatriz por otra función K.

Bajo esas condiciones entonces el estimador tomaría la forma

$$\hat{f}_{n,h}(x) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{x - x_i}{h})$$

Lema 1.1.3. Sea K una función real que cumple

i) K es acotada y simétrica

ii) $\int |\ K(t)\ |\ dt = C_1 < \infty$

iii) $\lim_{x \to \pm\infty} |\ xK(x)\ | = 0$

Sea g una función real tal que $\int |\ g(t)\ |\ dt = C_2 < \infty$

Consideremos la convolución

$$g_h^x = K_h * g(x) = \frac{1}{h}\int K(\frac{x - t}{h})g(t)dt$$

donde $K_h(z) = \frac{1}{h}K(z/h)$ entonces

$$g_h(x) \longrightarrow g(x) \int K(t)dt$$

si $h \to 0$ para todo punto x de continuidad de g.

Demostración: Sea x un punto de continuidad de g.

Dado $\epsilon > 0$ existe $\delta > 0$ tal que $\mid g(u) - g(x) \mid < \frac{\epsilon}{3C_1}$ si $\mid x - u \mid < \delta$

Consideremos entonces

$$\mid g_h(x) - g(x) \int K(t)dt \mid = \mid \frac{1}{h} \int K(\frac{x-t}{h}) g(t)dt - g(x) \int K(t)dt \mid =$$

$$= \mid \int K(-u)g(x+uh)du - g(x) \int K(u)du \mid \leq \int \mid (K(u) \mid\mid g(x+uh) - g(x) \mid du$$

dado que K es simétrica.

Ahora

$$\int \mid K(u) \mid\mid g(x+uh) - g(x) \mid du < \frac{\epsilon}{3C_1} \int\limits_{|u|<\delta/h} \mid K(u) \mid du +$$

$$+ \int\limits_{|u|\geq\delta/h} \mid K(u) \mid\mid g(x) \mid du + \int\limits_{|u|\geq\delta/h} \mid K(u) \mid\mid g(x+uh) \mid du$$

Como x es un punto de continuidad de g, $\mid g(x) \mid < C_3$ y

$\int\limits_{|u|\geq\delta/h} \mid K(u) \mid du < \frac{\epsilon}{3C_3}$ pues es la cola de una integral convergente. Por último,

$$\int\limits_{|u|\geq\delta/h} \mid K(u) \mid\mid g(x+uh) \mid du \leq \frac{h}{\delta} \int\limits_{|u|\geq\delta/h} \mid uK(u) \mid\mid g(x+uh) \mid du \leq$$

$$< \frac{h}{\delta} \frac{\epsilon\delta}{3C_2} \int\limits_{|u|\geq\delta/h} \mid g(x+uh) \mid du \leq \frac{\epsilon}{3C_2} \int \mid g(t) \mid dt = \frac{\epsilon}{3}$$

pues $\mid uK(u) \mid < \frac{\epsilon\delta}{3C_2}$ para h suficientemente chico.

A partir de estas tres desigualdades obtenemos que $\mid g_h(x) - g(x) \int K(t)dt \mid < \epsilon$ para $h \longrightarrow 0$

Veamos ahora una versión multidimensional del lema anterior.

Lema 1.1.4. Sea $K_h = \frac{1}{h^d}K(\frac{\bullet}{h})$.

Consideremos f en $L^1(R^d)$ y K en $L^1(R^d) \cap L^\infty(R^d)$

Además K cumple

i) $\int\limits_{R^d} K(t)dt = 1$

ii) $K(x) = o(\|x\|^{-d})$ si $\|x\| \longrightarrow +\infty$
Sea $f_h = f * K_h$
entonces $f_h \longrightarrow f$ si $h \longrightarrow 0$ en cada punto de continuidad x de f.

Demostración: Al ser x un punto de continuidad de f, dado $\epsilon > 0$, existe $\delta > 0$ tal que $\mid f(x-t) - f(x) \mid < \frac{\epsilon}{3\|K\|_1}$ si $\|t\| < \delta$

En primer lugar vemos que $\int\limits_{R^d} K_h(t)dt = 1$ ya que

$$\int\limits_{R^d} K_h(t)dt = \int\limits_{R^d} \frac{1}{h^d} K(\frac{t}{h})dt = \int\limits_{R^d} K(u)du = 1$$

Tomando la diferencia

$$\mid f_h(x) - f(x) \mid = \mid \int\limits_{R^d} f(x-t)K_h(t)dt - f(x) \int\limits_{R^d} K_h(t)dt \mid =$$

$$= \mid \int\limits_{R^d} (f(x-t) - f(x))K_h(t)dt \mid \leq \frac{\epsilon}{3 \parallel K \parallel_1} \int\limits_{\|t\|<\delta} \mid K_h(t) \mid dt +$$

$$+ \int\limits_{\|t\|\geqslant\delta} \mid f(x-t) \mid\mid K_h(t) \mid dt + \int\limits_{\|t\|\geqslant\delta} \mid f(x) \mid\mid K_h(t) \mid dt$$

Además $\int\limits_{|t|\geqslant\delta} \mid K_h(t) \mid dt < \frac{\epsilon}{3|f(x)|}$ si h es suficientemente chico por ser la cola de una integral convergente.

Definamos la función $\ell : R^d \longrightarrow R$ mediante $\ell(x) = \mid K(x) \mid \|x\|^d$

Por ii) vemos que $\ell(x) \longrightarrow 0$ si $\|x\| \longrightarrow +\infty$ luego

$$\int\limits_{\|t\|\geq\delta} \mid f(x-t) \mid\mid K_h(t) \mid dt = \int\limits_{\|t\|\geq\delta} \mid f(x-t) \mid \frac{\ell(t/h)}{\|t\|^d} dt \leq$$

$$\leq \frac{1}{\delta^d} \sup_{\|t\| \geq \delta} \ell(t/h) \int\limits_{\|t\|\geq\delta} \mid f(x-t) \mid dt$$

Como $\sup\limits_{\|t\| \geq \delta} \ell(t/h) \longrightarrow 0$ si $h \longrightarrow 0$ pues $\|t/h\| \longrightarrow +\infty$ y

$\int\limits_{\|t\|\geq\delta} \mid f(x-t) \mid dt \leq \parallel f \parallel_1$ entonces $\int\limits_{\|t\|\geq\delta} \mid f(x-t) \mid\mid K_h(t) \mid dt < \frac{\epsilon}{3}$

luego obtenemos que $\mid f_h(x) - f(x) \mid < \epsilon$.

Veamos ahora las propiedades del estimador general basado en núcleos.

Teorema 1.1.5. Sea x un punto de continuidad de f, K un núcleo simétrico y acotado y tal que $\int K(t)dt = 1$. Además $\mid xK(x) \mid \longrightarrow 0$ si $\mid x \mid \longrightarrow +\infty$.

Entonces

$$f_n(x) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{x - X_i}{h})$$

es un estimador asintóticamente insesgado de f(x) si $h \longrightarrow 0$.

Si además $nh \longrightarrow \infty$ es estimador consistente de f(x).

Demostración:

Por el lema 1.1.3

$$E(f_n(x)) = \frac{1}{h}\int K(\frac{x-t}{h})f(t)dt \underset{h}{\to} f(x)$$

$$Var(f_n(x)) = \frac{1}{nh^2}Var(K(\frac{x - X_1}{h})) \leq \frac{1}{nh^2}E(K^2(\frac{x - X_1}{h})) =$$

$$\frac{1}{nh}\frac{1}{h}\int K^2(\frac{x-t}{h})f(t)dt \longrightarrow \frac{f(x)}{nh}\int K^2(t)dt$$

Esta convergencia es válida ya que como $\mid xK(x) \mid \longrightarrow 0$ si $\mid x \mid \longrightarrow +\infty$ entonces $\mid x^2K^2(x) \mid \longrightarrow 0$ si $\mid x \mid \longrightarrow +\infty$ lo que implica que $\mid xK^2(x) \mid \longrightarrow 0$ si $\mid x \mid \longrightarrow +\infty$ pues $xK^2(x) \leq x^2K^2(x)$ para $\mid x \mid > 1$.

Además $\int \mid K^2(t) \mid dt \leq \parallel K \parallel_\infty \int K(t)dt < \infty$

Entonces si $nh \to \infty$ $Var(f_n(x)) \to 0$

Hemos demostrado que si $nh \longrightarrow \infty$ y $h \longrightarrow 0$ entonces $f_n(x) \underset{\longrightarrow}{P} f(x)$

Observación: En realidad hemos probado la convergencia en L^2, es decir que

$$E(f_n(x) - f(x))^2 \longrightarrow 0$$

Observación: el orden de la varianza es $\frac{1}{nh}$ pues

$$Var(\sqrt{nh}f_n(x)) = nhVar(f_n(x)) = \frac{1}{h}Var(K(\frac{x - X_1}{h})) =$$

$$\frac{1}{h}E(K^2(\frac{x - X_1}{h})) - \frac{1}{h}E^2(K(\frac{x - X_1}{h}))$$

luego

$$Var(\sqrt{nh}f_n(x)) \to f(x)\int K^2(u)du$$

si

$$h \to 0$$

ya que

$$\frac{1}{h^2}E^2(K(\frac{x-X_1}{h}) \to (f(x)\int K(u)du)^2$$

y

$$\frac{1}{h}E(K^2(\frac{x-X_1}{h}) \to f(x)\int K^2(u)du$$

de donde

$$\frac{Var(f_n(x))}{1/nh} \to C$$

Proposición 1.1.6. Sea x un punto de continuidad de f y que K cumple las hipótesis del lema 1.1.3.

Además suponemos que:
i) existe f'' y $sup|f''| < M$ y que el núcleo cumple:
ii) $\int uK(u)du = 0$ y
iii) $\int u^2K(u)du = a < \infty$

Entonces la ventana óptima cumple: $h \cong C_x n^{-1/5}$ y $ECM \backsimeq Cn^{4/5}$
Demostración: Sabemos que $ECM = B^2 + Var$ donde B indica el sesgo.

Comencemos calculando el sesgo

$$B(f_n(x)) = E(f_n(x)) - f(x) = \frac{1}{h}\int K(\frac{x-t}{h})f(t)dt - f(x) =$$

$$\int K(u)(f(x+hu) - f(x))du = \int K(u)\left[f'(x)hu + f''(\xi)\frac{h^2u^2}{2}\right]du =$$

$$\frac{h^2}{2}\int u^2K(u)f''(\xi)du \leq \frac{Mah^2}{2}$$

Si tomo $f(x) = x^2$ entonces la desigualdad se transforma en igualdad, luego

$$B_n(f_n(x)) \cong Ch^2$$

Tenemos pues que $ECM = \frac{C_1}{nh} + C_2h^4 + o(\frac{1}{nh} + h^4)$

entonces el h óptimo sería $h \cong Cn^{-1/5}$ y

$$ECM \cong Cn^{-4/5}$$

.

Proposición 1.1.7. Consideremos que x es un punto de continuidad de f y K cumple las hipótesis de los lemas anteriores. Además

i) f es Lipchitz de constante L, es decir pido menos regularidad que en los lemas anteriores.

ii) $\int |K(u)u|du = a_3 < \infty$

Entonces la ventana óptima será $h = C_n^{-1/2}$ y el $ECM(f_n(x)) \cong C_n^{-2/3}$

Demostración: A partir de las hipótesis ya sabemos que:

$$Var(\sqrt{nh} f_n(x)) \to f(x) \int K^2(t)dt$$

El sesgo, usando que $\int K = 1$ queda:

$$| B(f_n(x))| = |E(f_n(x)) - f(x)| = |\frac{1}{h} \int K(\frac{x-\mu}{h} f(u)du - f(x)| \leq$$

$$\leq \int |K(t)||f(x+th) - f(x)|dt \leq L \int |th||K(t)|dt = Lh \int |tK(t)|dt = La_3 h$$

así pues

$$|B(f_n(x))| \leq La_3 h$$

Si utilizo una función lineal $f(t) = LtI_{[x-c,x+c]}(t)$ obtengo la igualdad, entonces:

$$B(f_n(x) \cong hC$$

luego

$$ECM = \frac{C_1}{nh} + C_2 h^2 + o(\frac{1}{nh} + h^2)$$

entonces orden óptimo es: $\frac{1}{nh} \cong h^2$ de donde $h = Cn^{-1/3}$ así pues

$$ECM(f_n(x)) \cong Cn^{-2/3}$$

Observación: A menor regularidad se obtiene menor velocidad de convergencia. Esto no tiene nada que ver con la parte aleatoria del problema.

1.2. Distribución asintótica del estimador

Veremos que adecuadamente normalizados estos estimadores convergen a la distribución normal.

Comencemos dando sin demostración la conocida desigualdad de Jensen.

Teorema 1.2.1. Desigualdad de Jensen

Sea μ una medida positiva sobre una σ-álgebra M en un conjunto Ω tal que $\mu(\Omega) = 1$.

Sea f es una función real en $L^1(\mu)$ con $a < f(x) < b$ para todo $x \in \Omega$.

Si φ es convexa en (a,b) entonces $\varphi(\int_\Omega f du) \leqslant \int_\Omega (\varphi \circ f) du$

Lema 1.2.2. Bajo las condiciones del lema 1.1.3

$$\frac{f_n(x) - Ef_n(x)}{\sigma(f_n(x))} \xrightarrow{w} N(0,1)$$

El estimador de densidad lo podemos escribir como

$$f_n(x) = \frac{1}{nh}\sum_{k=1}^{n}(K\frac{X - X_k}{h}) = \frac{1}{n}\sum_{k=1}^{n} W_{k_n}$$

con

$$W_{k_n} = \frac{1}{h}K(\frac{X - X_h}{h})$$

O sea, se tiene un sistema triangular de variables idénticamente distribuidas por filas.

Utilizaremos la condición de Liapunov para la convergencia a la Normal, de más está decir que esta condición implica las de Lindeberg.

Una condición suficiente para que

$$\frac{\hat{f}_n(x) - E\hat{f}_n(x)}{\sigma(\hat{f}_n(x))} \xrightarrow{w} Z \backsim N(0,1)$$

es que

$$\frac{E(|W_{n_1} - E(W_{n_1})|^{2+\delta})}{n^{\delta/2}\sigma(W_{n_1})^{2+\delta}} \longrightarrow 0$$

con $n \to \infty$ para algún $\delta > 0$

$$E(|W_{n_1}|^{2+\delta}) = \frac{1}{h^{2+\delta}}E(K^{2+\delta}(\frac{x - X_1}{h})) = \frac{1}{h^{1+\delta}}\frac{1}{h}\int K^{2+\delta}(\frac{x-t}{h})(f(t)dt$$

Si se cumplen las hipótesis del lema 1.1.3

$$\frac{1}{h}\int K^{2+\delta}\left(\frac{x-t}{h}\right) f(t)dt \to f(x)\int K^{2+\delta}(t)dt$$

así pues,

$$h^{1+\delta}E\left(|W_{n_1}|^{2+\delta}\right) \to f(x)\int K^{2+\delta}(t)dt$$

De igual modo:

$$\sigma^2(W_{n_1}) = \frac{1}{h^2}Var\left(K(\frac{x-X_1}{h})\right) = \frac{1}{h^2}\int K^2\left(\frac{x-t}{h}\right)f(t)dt$$

Como

$$\frac{1}{h}\int K^2\left(\frac{x-t}{h}\right)f(t)dt$$

se tiene que:

$$h\sigma^2\left(W_{n_1}\right) \to f(x)\int K^2(t)dt$$

También

$$E\left(W_{n_1}\right) \to f(x)$$

si

$$\int K(u)du = 1$$

$$E\left(|W_{n_1} - E\left(W_{n_1}\right)|^{2+\delta}\right) \leqslant E((|W_{n_1}| + |E\left(W_{n_1}\right)|)^{2+\delta})$$

usando la desigualdad:

$$\left(\frac{a+b}{2}\right)^{2+\delta} \leqslant \frac{a^{2+\delta}+b^{2+\delta}}{2}$$

obtenemos:

$$(|W_{n_1}| + |E(W_{n1})|)^{2+\delta} \leqslant \frac{2^{2+\delta}}{2}\left(|W_{n1}|^{2+\delta} + E\left(W_{n1}\right)|^{2+\delta}\right) =$$
$$2^{1+\delta}\left(|W_{n_1}|^{2+\delta} + |E\left(W_{n_1}\right)|^{2+\delta}\right)$$

Usando la desigualdad de Jensen a la función

$$\varphi = x^{2+\delta}$$

tenemos que:

$$(E|W_{n1}|)^{2+\delta} \leqslant E\left(|W_{n1}|^{2+\delta}\right)$$

así pues

$$E\left(|W_{n_1}| + |E\left(W_{n_1}\right)|\right)^{2+\delta} \leqslant E\left(2^{1+\delta}\left(|W_{n1}|^{2+\delta} + |E\left(W_{n1}\right)|^{2+\delta}\right)\right) \leqslant$$

$$\leqslant E\left(2^{1+\delta}\left(|W_{n1}|^{2+\delta} + E\left(|W_{n1}|^{2+\delta}\right)\right)\right) = 2^{1+\delta}\left(E|W_{n1}|^{2+\delta} + E|W_{n1}|^{2+\delta}\right) =$$

$$= 2^{2+\delta}E\left(|W_{n1}|^{2+\delta}\right)$$

luego

$$\frac{E\left(|W_{n1}-E\left(W_{n1}\right)^{2+\delta}|\right)}{n^{\frac{\delta}{2}}\sigma\left(W_{n1}\right)^{2+\delta}} \leqslant \frac{2^{2+\delta}E\left(|W_{n1}|^{2+\delta}\right)}{n^{\frac{\delta}{2}}\sigma\left(W_{n1}\right)^{2+\delta}} =$$

$$= \frac{2^{2+\delta}}{h^{1+\delta}} h^{1+\delta} E\left(|W_{n1}|^{2+\delta}\right) / n^{\frac{\delta}{2}} \frac{h^{1+\frac{\delta}{2}}\sigma\left(W_{n1}\right)^{2+\delta}}{h^{1+\frac{\delta}{2}}}$$

$$\leqslant \frac{C}{h^{1+\delta}} / \frac{n^{\delta/2}}{h^{1+\delta/2}} = \frac{C}{(nh)^{\delta/2}} \to 0$$

si $nh \to \infty$ pues

$$h^{1+\delta}E(|W_{n1}|^{2+\delta}) \to f(x)\int K^2(t)dt$$

y

$$h^{1+d/2}\sigma(w_{n1})^{2+\delta} \to cte$$

así pues se cumple la condición de Liapunov.

Hemos probado que

$$\frac{f_n(x)-Ef_n(x)}{\sigma(f_n(x))} \xrightarrow{w} N(0,1)$$

Teorema 1.2.3. Supongamos que K es acotada y que $\int K(u)du = 1$,
además se cumple que $\mu^2 K(u)$ es acotada, $nh \to \infty$ y $nh^3 \to 0$.
También la densidad f es Lipchitz en un entorno de x,
entonces
a)

$$Var(\sqrt{nh}\widehat{f_n}(x)) \longrightarrow f(x)\int K^2(u)du$$

b)

$$\sqrt{nh}\left(\frac{1}{nh}\sum_{i=1}^{n} E\left(K\left(\frac{x-X_i}{h}\right)\right) - f(x)\right) \to 0$$

c)

$$\sqrt{nh}\left(\widehat{f_n}(x) - f(x)\right) \longrightarrow N(0, f(x)\int K^2(u)du)$$

Demostración:
a) Como K es acotado, $\int K(u)du = 1$ entonces $\int K^2(u)du < \infty$.
Sabemos que $|uK(u)| \to 0$ si $|\ u\ | \to \infty$ pues $u^2K(u)$ acotado luego se cumplen las hipótesis del lema 1.1.3, así que a) se cumple,

b) existe $\delta > 0$ tal que $|f(x+s) - f(x)| \leq c|s|\ ;\ |s| < \delta$

$$|\sqrt{nh}\left(\frac{1}{nh}\sum_{i=1}^{n}E\left(K\left(\frac{x-X_i}{h}\right)\right)-f(x)\right)| = |\sqrt{nh}\left(\int\frac{1}{h}K\left(\frac{x-t}{h}\right)f(t)dt-f(x)\right)| =$$

$$|\sqrt{nh}\left(\int K(u)f(x+uh)du - f(x)\int K(u)du\right)| = |\sqrt{nh}\int K(u)|f(x+uh)-f(x)|du| \leqslant$$

$$\leqslant \sqrt{nh}\left[hC\int_{|hu|<\delta}|uK(u)|du + \int_{|u|>\delta/h}|K(u)||f(x+uh)|du + |f(x)|\int_{|u|>\delta/h}|K(u)|du\right] =$$

Sea $C_1 : \|u^2K(u)\|_\infty$

$$= \sqrt{nh}\left[Ch\int_{|hu|<\delta}|uK(u)|du + \int_{|u|>\delta/h}\frac{|u^2K(u)|}{|u^2|}f(x+uh|du + f(x)|\int_{|u|>\delta/h}\frac{|u^2K(u)|}{u^2}\right] \leq$$

$$\leqslant \sqrt{nh}\left[Ch\int_{|hu|<\delta}|uK(u)|du + \frac{h^2}{\delta^2}C_1\int_{|u|>\delta/h}f(x+uh)du + C_1f(x)\int_{|u|>\delta/h}\frac{du}{u^2}\right] \leqslant$$

$$\leqslant \sqrt{nh}\left[Ch\int_{|hu|<\delta}|uK(u)|du + h\frac{C_1}{\delta^2}\int_R f(t) + C_1f(x)h\int_{|t|>\delta}\frac{dt}{t^2}\right] =$$

$$\sqrt{nh^3}\left[C_2 + C_3 + C_4\right] \to 0$$

si $nh^3 \to 0$ pues

$$\int_{|u|>\delta/h}\frac{du}{u^2} = h\int_{|t|>\delta}\frac{dt}{t^2}$$

$$\int_{|u|>\delta/h}f(x+uh)du = \int_R f(t)\frac{dt}{h}$$

y

$$\int_{|hu|<\delta}|uK(u)|du$$

está acotado al igual que:

$$f(x)\int_{|t|>\delta}\frac{dt}{t^2}$$

c)

$$\sqrt{nh}\,(f_n(x) - f(x)) = \sqrt{nh}\,(f_n(x) - Ef_n(x)) + \sqrt{nh}\,[Ef_n(x) - f(x)]$$

sabemos que:

$$\sqrt{nh}\,[Ef_n(x) - f(x)] \to 0$$

consideremos,

$$\sigma\,(f_n(x))\,\frac{(f_n(x) - Ef_n(x))}{\sigma\,(f_n(x))}$$

se tiene que:

$$\frac{(f_n(x) - Ef_n(x))}{\sigma\,(f_n(x))} \sim N(0,1)$$

además,

$$nh\sigma^2\,(f_n(x)) \longrightarrow f(x)\int K^2(u)du$$

entonces:

$$\sqrt{nh}\sigma\,(f_n(x)) \longrightarrow \left(f(x)\int K^2(u)du\right)^{1/2}$$

así pues,

$$\sqrt{nh}\,(f_n(x) - f(x)) \longrightarrow N\left(0, f(x)\int K^2(u)du\right)$$

1.3. Convergencia completa de los estimadores basados en núcleos

Teorema 1.3.1. Sea

$$f_n(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right), X_1, ..., X_n$$

con densidad f e independientes. El núcleo K cumple las hipótesis del lema 1.1.3, e

$$\int K = 1$$

En estas condiciones el estimador basado en núcleos converge completamente a la densidad f en cada punto de continuidad de f.

Demostración: Queremos probar que

$$\sum_{i=1}^{\infty} P\,(|f_n(x) - f(x)| > \epsilon) < \infty$$

.

$$\{|f_n(x) - f(x)| > \epsilon\} = \left\{ |\frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) - f(x)| > \epsilon \right\} =$$

$$\left\{ |\frac{1}{n} \left[\sum_{i=1}^{n} \left(\frac{1}{h} K\left(\frac{X - x_i}{h}\right) - f(x) \right) \right] | > \epsilon \right\}$$

entonces

$$P\left(|f_n(x) - f(x)| > \epsilon\right) = P\left(|\frac{1}{n} \sum_{i=1}^{n} \left(\frac{1}{h} K\left(\frac{x - X_i}{h}\right) - f(x) \right) | > \epsilon \right) \leq$$

$$\leq P\left(\frac{1}{n} \sum_{i=1}^{n} |\frac{1}{h} K\left(\frac{x - X_i}{h}\right) - f(x)| > \epsilon \right)$$

Consideremos la variable

$$Y_i = \frac{1}{h} K\left(\frac{x - X_i}{h}\right) - f(x)$$

Claramente las Y_i son independientes si las X_i lo son:

$$E(Y_i) = E\left(\frac{1}{h} K\left(\frac{x - X_i}{h}\right) \right) - f(x) = \frac{1}{h} \int K\left(\frac{x - t}{h}\right) f(t)dt - f(x) \to 0$$

pues

$$\frac{1}{h} \int K\left(\frac{x - t}{h}\right) f(t)dt \to f(x)$$

Además,

$$|Y_i| = |\frac{1}{h} K\left(\frac{x - X_i}{h}\right) - f(x)| < C$$

pues K es acotada.

$$E(Y_i^2) = E\left(\frac{1}{h^2} K^2\left(\frac{x - X_i}{h}\right) \right) - 2f(x) E\left(\frac{1}{h} K\left(\frac{x - X_i}{h}\right) \right) + f^2(x) =$$

$$\frac{1}{h}\frac{1}{h} \int K^2\left(\frac{x - t}{h}\right) f(t)dt - 2f(x)\frac{1}{h} \int K\left(\frac{X - t}{h}\right) f(t)dt + f^2(x) \to$$

$$\to \frac{1}{h} f(x) \int K^2(t)dt - f^2(x)$$

ya que

$$\frac{1}{h}\int K^2\left(\frac{x-t}{h}\right)f(t)dt \to f(x)\int K^2(t)dt$$

y

$$(1/h)\int K\left(\frac{x-t}{h}\right)f(x)dt \to f(x)$$

Utilizando la desigualdad de Bernstein tenemos:

$$P\left(\frac{1}{n}\sum_{i=1}^{n}|Y_i| > \epsilon\right) \leq 2exp\left(\frac{-n^2\epsilon^2}{2\left(n(\frac{1}{h}f(x)\int K^2(t)dt - f^2(x))) + cn\epsilon\right)}\right)$$

Operando en el exponente

$$\frac{-n^2\epsilon^2}{2\left(n\left(\frac{f(x)}{h}\int K^2(t)dt - f^2(x) + c\epsilon\right)\right)} = \frac{-n\epsilon^2}{2\left(\frac{C_2}{h} - C_1 + c\epsilon\right)} = \frac{-n\epsilon^2}{2\left(\frac{C_2}{h} + C_3\right)} \leq$$

$$\leq \frac{-n\epsilon^2}{2\frac{C_2}{h}} = -nhC'$$

Así pues

$$P\left(\frac{1}{n}\sum_{i=1}^{n}|Y_i| > \epsilon\right) \leqslant 2exp(-nhC')$$

entonces

$$\sum P\left(\frac{1}{n}\sum_{i=1}^{n}|Y_i| > \epsilon\right) \leqslant \sum 2e^{-nhC'}$$

Si

$$Q(n) = \frac{nh}{\log n}$$

tenemos

$$nh = Q(n) \log n$$

entonces

$$2e^{-C' \log nQ(n)} = 2e^{log n^{-C'Q(n)}} = 2n^{-C'Q(n)}$$

Si $Q(n) \to \infty$ resulta que $\sum 2n^{-C'Q(n)}$ converge, así pues hay convergencia completa.

1.4. Regresión no paramétrica

La función de regresión describe la relación entre una variable explicativa $X \in R^d$ y una variable explicada $Y \in R$. Se puede modelizar su relación mediante la esperanza condicional:

$$E(Y \mid X = x) \tag{1.1}$$

El objetivo del análisis de regresión es encontrar una aproximación de la función $E(Y \mid X = x)$ a partir de n datos observados $\{(X_i, Y_i)\}$.

Para aproximar la función de esperanza condicional se puede utilizar un enfoque paramétrico, donde se asume que la esperanza condicional tiene una forma dada y está totalmente especificada por un número finito de parámetros, o desde un punto de vista no paramétrico, donde no se hace referencia a ninguna forma específica. Este último método se caracteriza por una gran flexibilidad.

Hardle (1990) menciona que la estimación no paramétrica de las curvas de regresión tiene cuatro objetivos fundamentales: "Primero provee un método versátil para explorar la relación entre dos variables. Segundo, da predicciones de las observaciones sin hacer referencia a un modelo paramétrico fijo. Tercero, provee una herramienta para encontrar observaciones espúreas estudiando la influencia de puntos aislados. Cuarto, constituye un método flexible para sustituir valores perdidos o interpolar entre valores de X".

La idea general de la regresión no paramétrica es la del promedio local, es decir, estimar la esperanza condicional mediante un promedio ponderado, donde los pesos son una función que depende del punto donde se va a estimar y de los valores observados. Formalmente:

$$\widehat{E(Y \mid X = x)} = \sum_{i=1}^{n} w_i(x) Y_i \tag{1.2}$$

La regresión no paramétrica por núcleos utiliza para la sucesión de pesos una función real K continua, acotada y simétrica con integral uno, fue introducida por Nadaraya (1964) y Watson (1964).

Supongamos que tenemos n observaciones $(X_1, Y_1).....(X_n, Y_n)$, que son independientes, idénticamente distribuidas con densidad conjunta $f_{X,Y}(x, y)$.

La densidad condicional de Y dada X es:

$$f_{X|Y} = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{f_{X,Y}(x,y)}{\int f_{X,Y}(x,y)dy}$$

Así tenemos que:

$$E(Y \mid X = x) = \frac{\int y f_{X,Y}(x,y)dy}{\int f_{X,Y}(x,y)dy} \tag{1.3}$$

Por lo tanto, para estimar la esperanza condicional utilizamos:

$$\widehat{E(Y \mid X = x)} = \frac{\int y\widehat{f_{X,Y}}(x,y)dy}{\int \widehat{f_{X,Y}}(x,y)dy} \tag{1.4}$$

Consideremos ahora un núcleo $G(x, y)$ en R^2 cualesquiera, o sea, G es una función simétrica, continua y acotada con integral uno.

El estimador por núcleos de la densidad conjunta es:

$$\widehat{f_{X,Y}}(x,y) = \frac{\sum_{i=1}^{n} \frac{1}{h^{d+1}} G(\frac{x-X_i}{h}, \frac{y-Y_i}{h})}{n}$$

donde h es el parámetro de ancho de la ventana, es decir, un número real que controla cuán rápidamente los pesos decrecen a cero.

Si el núcleo es simétrico como función de y tenemos que:

$$\int \frac{1}{h^{d+1}} G(\frac{x-X_i}{h}, \frac{y-Y_i}{h})dy = 0$$

de donde, realizando el cambio de variable, $u = \frac{y-Y_i}{h}$

$$n^{-1}\sum_{i=1}^{n} \int y \frac{1}{h^{d+1}} G(\frac{x-X_i}{h}, \frac{y-Y_i}{h})dy = n^{-1}\sum_{i=1}^{n} Y_i \int \frac{1}{h^d} G(\frac{x-X_i}{h}, u)du$$

Además,

$$\int \widehat{f_{X,Y}}(x,y)dy = n^{-1}\sum_{i=1}^{n} \int \frac{1}{h^d} G(\frac{x-X_i}{h}, u)du$$

Llamemos K al núcleo marginal, o sea, $K(x) = \int G(x,u)du$.

Obtenemos finalmente como estimador:

$$\widehat{E(Y \mid X = x)} = \frac{n^{-1}\sum_{i=1}^{n} Y_i \frac{1}{h^d} K(\frac{x-X_i}{h})}{n^{-1}\sum_{i=1}^{n} \frac{1}{h^d} K(\frac{x-X_i}{h})} = \frac{\sum_{i=1}^{n} Y_i K(\frac{x-X_i}{h})}{\sum_{i=1}^{n} K(\frac{x-X_i}{h})} \tag{1.5}$$

Notemos que el numerador es la convolución de K con los datos y el denominador es la convolución de K con la unidad.

Claramente este es un estimador del tipo de promedio local ponderado con pesos,

$$w_i(x) = \frac{K(\frac{x-X_i}{h})}{\sum_{i=1}^{n} K(\frac{x-X_i}{h})}$$

Algunos núcleos comúnmente usados son:

1. Uniforme $K(u) = 0{,}5$ si $|u| \leq 1$ y 0 en otro caso.
2. Cuadrático o de Epanechnikov $K(u) = 0{,}75(1-u^2)$ si $|u| \leq 1$ y 0 en otro caso.
3. Gaussiano $K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$
4. Biweight $K(u) = \frac{15}{16}\left(1-u^2)^2\right)$ si $|u| \leq 1$ y 0 en otro caso.
5. Triweight $K(u) = \frac{35}{32}\left(1-u^2)^3\right)$ si $|u| \leq 1$ y 0 en otro caso.
6. Epanechnikov de cuarto orden $K(u) = \frac{15}{8}(1-\frac{7}{3}u^2)0{,}75(1-u^2)$ si $|u| \leq 1$ y 0 en otro caso.
7. Gaussiano de cuarto orden $K(u) = \frac{1}{2}(3-u^2)\frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$
8. Biweight de cuarto orden $K(u) = \frac{7}{4}(1-3u^2)\frac{15}{16}\left(1-u^2)^2\right)$ si $|u| \leq 1$ y 0 en otro caso.

Veamos que la estimación es consistente. Consideremos la demostración de Naradaya que supone que $X \sim f_X$

No hace falta pedir densidad. Esto se probó por Greblecki, Kizizak, Parvlack (1984).

Teorema 1.4.1. Supongamos $X \sim f_x$, $f_x(x) > 0$ y continua en x entonces si $h_n \to 0$, $nh_n \to \infty$

$\int K(u)du = 1$ y $|xK(x)| \to 0$ cuando $|x| \to \infty$,$E(|Y|) < \infty$ y si además suponemos alguna de las condiciones a) o b)

a) $E(Y^2) < \infty$

$nh^2 \to \infty$ y K acotada

b) $m_2(x) = E(Y^2/X = x)$ es continua en x.

entonces el estimador de regresión es consistente.

Demostración:

$$\widehat{g}(x) = \frac{\frac{1}{nh}\sum_{i=1}^{n} Y_i K\left(\frac{x-X_i}{h}\right)}{\frac{1}{nh}\sum_{j=1}^{n} K\left(\frac{x-X_j}{h}\right)} = \frac{R_{1n}(x)}{R_{0n(x)}}$$

Bajo estas condiciones se tiene que,

$$R_{0n}(x) \to f(x)$$

Trabajemos ahora con el término $R_{1n}(x)$,

$$E\left(R_{1n}(x)\right) = E\left(\frac{1}{h}K\left(\frac{x-X_i}{h}\right)Y_i\right) =$$

$$= \frac{1}{h}E\left(K\left(\frac{x-X_i}{h}\right)E\left(Y_i/X\right)\right) = \frac{1}{h}\int K\left(\frac{x-t}{h}\right)g(t)f_X(t)dt \to g(x)f(x)$$

en cada punto de continuidad de gf por el lema 1.1.3

Claramente $g.f \in L^1$, pues

$$\int |g(t)|f(t)dt = E(|g(x)|) = E(|E(Y/X)|) \leq E(E(|Y|/X)) = E(|Y|)$$

o sea basta pedir $E(|Y|) < \infty$ y las hipótesis sobre K para aplicar el lema 1.1.3 .

Probemos ahora que la varianza tiende a cero:

$$Var(R_{1n}(x)) = \frac{1}{nh^2}Var\left(K\left(\frac{x-X_1}{h}\right)Y_1\right) \leq \frac{1}{nh^2}E\left(Y_1^2K^2\left(\frac{x-X_1}{h}\right)\right) =$$

$$\frac{1}{nh^2}E\left(K^2\left(\frac{x-X_1}{h}\right)E\left(Y_1^2/X_1\right)\right) = \frac{1}{nh^2}\int K^2\left(\frac{x-t}{h}\right)m_2(t)f(t)dt$$

donde

$$m_2(X) = E(Y^2/X)$$

$$E(m_2(X)) = E(E(Y^2/X)) = E(Y^2)$$

Analicemos ahora el término

$$\frac{1}{nh^2}\int K^2\left(\frac{x-t}{h}\right)m_2(t)f(t)dt$$

Acá caben dos posibilidades:

a) Si $E(Y^2) < \infty$ y $nh^2 \to \infty$ y si K está acotado tenemos,

$$\frac{1}{nh^2} \int K^2 \left(\frac{x-t}{h} \right) m_2(t) f(t) dt \to 0$$

b) si no exigimos que $E(Y^2) < \infty$ entonces,

$$\frac{1}{nh^2} \int K^2 \left(\frac{x-t}{h} \right) m_2(t) f(t) dt = \frac{1}{nh} \frac{1}{h} \int K^2 \left(\frac{x-t}{h} \right) m_2(t) f(t) dt \longrightarrow$$
$$m_2(x) f(x) \int K^2$$

aplicando el lema 1.1.3 siempre y cuando $m_2(x)$ sea continua en x, o sea se necesita que la esperanza condicional sea continua.

Veamos ahora dos resultados, uno de convergencia en probabilidad y otro de convergencia completa.

Sean $(X_1, Y_1) \ldots\ldots (X_n, Y_n)$ vectores aleatorios independientes, idénticamente distribuidas, con $X_i \in R^d, Y_i \in R$.

Consideremos las hipótesis:

H1)

$$h_n \to 0$$

y

$$nh_n^d \to \infty$$

Para el núcleo pidamos:

H2)

$$K : R^d \to R^+$$

existen $C_1, C_2 > 0$ y H tales que

$$C_1 H(\|X\|) \leq K(x) \leq C_2 H(\|X\|)$$

con H acotada, decreciente y tal que $t^d H(t) \to 0$ si $|t| \to \infty$

H3) Existen $c, r > 0$ tal que

$$K(x) \geq c \uparrow | \{\|X\| \leq r\}(x) \chi_{\|x\| \leqslant r}(x)$$

Si f es integrable y X tiene distribución μ entonces definimos:

$$\mu_h = \int K\left(\frac{X-t}{h}\right) f(t)\mu(dt) / \int K\left(\frac{x-t}{h}\right) f(t)\mu(dt) =$$

$$E\left(K\left(\frac{X-\overline{X}}{h}\right) f(\overline{X})\right) / E\left(K\left(\frac{X-\overline{X}}{h}\right)\right)$$

Observaciones:

Veamos que lo que le pedimos al núcleo es menos que suponer continuidad y soporte compacto, o radial, comúnmente se toma $K(\|t\|)$.

La segunda hipótesis sobre el núcleo implica menos que pedir que $K(0) \neq 0$ y K continua en 0, podemos ver que esta condición la cumplen los núcleos usualmete utilizados.

También debemos observar que no hay hipótesis sobre la distribución de las X.

Veamos dos lemas.

Lema 1.4.2. Sea S_r la bola de radio r centrada en x, y sea $a_h(x) = \frac{h^d}{\mu(S_h)}$ entonces $a_h(x)$ tiene un límite finito ctp $x(\mu)$.

Demostración:

Sea λ la medida de Lebesgue, descompongámosla en la parte absolutamente continua respecto a μ en la parte singular.

Notemos por λ_1 a la parte absolutamente continua de λ respecto a μ y λ_2 la parte singular, luego

$$\frac{\lambda(S_h)}{\mu(S_h)} \longrightarrow \frac{d\lambda_1}{d\mu}(x)$$

ctp

$$x(\mu)$$

$$\lambda(S_h) = h^d \lambda(B_1)$$

entonces

$$\frac{h^d}{\mu(S_n)} = \frac{\lambda(S_h)}{\mu(S_n)} \frac{1}{\lambda(B_1)} \longrightarrow \ell(x)$$

ctp

$$x(\mu)$$

donde ℓ es no negativa y finita.

Lema 1.4.3. Sea K no negativo verificando H2) y H3) y f integrable, entonces

$\mu_h(x) \to f(x)$ c.t.p. x (μ) si $h \to 0$

Si K es el indicador de la bola unitaria, tenemos el teorema de diferenciación respecto a la medida μ.

Lo que presentamos es cómo mezclar un teorema de convolución con un teorema de diferenciación.

Demostración:

$$|\mu_h(x) - f(x)| = |\frac{\int K\left(\frac{X-y}{h}\right) f(y)\mu(dy)}{\int K\left(\frac{X-y}{h}\right)\mu(dy)} - f(x)| = |\frac{\int K\left(\frac{X-y}{h}\right)(f(y)-f(x))\mu(dy)}{\int K\left(\frac{X-y}{h}\right)\mu(dy)}| \leq$$

$$\leq \frac{C_2}{C_1}\int H\left(\frac{\|x-y\|}{h}\right)|f(y)-f(x)|\mu(dy)/\int H\left(\frac{\|x-y\|}{h}\right)\mu(dy)$$

por H2)

$$H(t) = \int_0^{H(t)} ds = \int_0^{+\infty} I_{\{H(t)>s\}} ds$$

Si

$$A_{t,h} = \left\{y : H\left(\frac{\|x-y\|}{h}\right) > t\right\}$$

entonces

$$H\left(\frac{\|x-y\|}{h}\right) = \int_0^{+\infty} I_{A_{t,h}}(t)dt$$

Sea

$$\int H\left(\frac{\|x-y\|}{h}\right)\mu(dy) = \int_{-\infty}^{+\infty}\int_0^{+\infty} I_{A_{t,h}}(t)dt\mu(dy) = \int_0^{+\infty}\mu\left(A_{t,h}\right)dt$$

ya que

$$I_{A_{t,h}}$$

es integrable, aplico Fubini.

Ahora

$$\int H\left(\frac{\|x-y\|}{h}\right)|f(y)-f(x)|\mu(dy) = \int_{-\infty}^{+\infty}\int_0^{+\infty} I_{A_{t,h}}(t)|f(y)-f(x)|dt\mu(dy) = \int_0^{+\infty}\int_{A_{t,h}} |f(y)-f(x)|\mu(dy)d$$

aplicando nuevamente Fubini.

Dividiremos el cociente en dos partes, I y II.

Dado $\epsilon > 0$, sea $\delta = \epsilon h^d$

I-

$$\frac{\int_{\delta}^{\infty} \int_{A_{t,h}} |f(y) - f(x)| \mu(dy) dt}{\int_{0}^{\infty} \mu\left(A_{t,h}\right) dt} \leq \frac{\int_{\delta}^{\infty} \int_{A_{t,h}} |f(y) - f(x)| \mu(dy) dt}{\int_{\delta}^{\infty} \mu\left(A_{t,h}\right) dt}$$

$$\leq \sup_{t \geq \delta} \frac{\int_{A_{t,h}} |f(y) - f(x)| \mu(dy)}{\mu\left(A_{t,h}\right) dt}$$

La última desigualdad es cierta ya que si consideramos dos funciones $g \geq 0$ y $h \geq 0$ con sus integrales finitas entonces,

$$\frac{\int_{\delta}^{\infty} g dt}{\int_{\delta}^{\infty} h dt} \leq \sup_{t \geq \delta} \{g/h\}$$

Supongamos que $\frac{g(t)}{h(t)} < A$ para todo $t \geq \delta$

Como g y h son no negativas

$$\int_{\delta}^{+\infty} g dt \leq A \int_{\delta}^{+\infty} h(t)$$

luego

$$\frac{\int_{\delta}^{+\infty} g(t) dt}{\int_{\delta}^{+\infty} h(t) dt} \leq A$$

cualquier cota de g/h cumple la desigualdad, así pues lo cumple el supremo de g/h.

Si $t \geq \delta$, consideremos los conjuntos $A_{t,h}$ y $A_{\delta,h}$, se tiene $radio\left(A_{t,h}\right) < radio\left(A_{\delta,h}\right)$

Tenemos que $radio\left(A_{\delta,h}\right) \to 0$ pues

$$\left\{ y : H\left(\frac{\|x - y\|}{h}\right) > \delta \right\} = \left\{ y : \|x - y\| < H^{-1}(\delta) h \right\}$$

ahora

$$h H^{-1}(\delta) = h H^{-1}(\epsilon h^d) \to 0$$

pues $t^d H(t) \longrightarrow 0_{|t|\to\infty}$

Veamos lo anterior.

Supongamos que $H^{-1}(\epsilon h^d) = t$, entonces se tiene que $h = \left(\frac{H(t)}{\epsilon}\right)^{1/d}$

luego $|hH^{-1}(\epsilon h^d)| = |t\left(\frac{H(t)}{\epsilon}\right)^{1/d}| \longrightarrow 0$ si $h \to 0$ pues $t \to \infty$.

Lo último se deduce ya que $t^d H(t) \longrightarrow 0$ $|t| \to \infty$ y es claro que $t(H(t))^{1/d} \longrightarrow 0$ si $|t| \to \infty$.

Utilizamos un resultado de desigualdades para funciones maximales debido a Hardy y Littlewood.

Concretamente, el resultado nos dice que si $radio(A_{t,h}) \to 0$ entonces

$$\sup_{t \geq \delta} \frac{\int\limits_{A_{t,h}} |f(y) - f(x)|\mu(dy)}{\mu(A_{t,h})} \to 0$$
$$ctpx(\mu)$$

Veamos ahora la segunda parte

II-

$$\frac{\int\limits_0^\delta \int\limits_{A_{t,h}} |f(y) - f(x)|\mu(dy)dt}{\int\limits_0^\infty \mu(A_{t,h})dt}$$

Llamemos $C = \int |f(y)|\mu(dy)$, entonces

$$\int\limits_0^\delta \int\limits_{\Delta_{t,h}} f(y)\mu(dy)\mu \leqslant C\delta$$

y

$$\int\limits_0^\delta \int\limits_{A_{t,h}} f(x)\mu(dy)dt \leq \delta f(x)$$

así pues

$$\int\limits_0^\delta \int\limits_{A_{t,h}} |f(y) - f(x)|\mu(dy)dt \leqslant \delta(C + f(x))$$

También

$$\int\limits_0^\infty \mu(A_{t,h})dt = \int H\left(\frac{\| x - y \|}{h}\right)\mu(dy) \geq C_1\mu\left(S_{rh}\right) = \frac{C_1 h^d r^d}{a_{rh}(x)}$$

luego

$$\frac{\int_0^{\delta}\int_{\Delta_{t,h}}|f(y)-f(x)|\mu(dy)}{\int_0^{\infty}\mu(\Delta_{t,h})} \leq \frac{\delta(C+f(x))a_{rh}(x)}{C_1 r^d h^d} = \frac{(f(x)+C)\epsilon a_{rh}(x)}{C_1 r^d} < \epsilon'$$

$$ctpx(\mu)$$

Aplicamos en la última desigualdad el lema 1.4.2

Teorema 1.4.4. Si $E(|Y|) < \infty$ (esta condición es para que exista la esperanza condicional), entonces bajo H1), H2) y H3)

$$g_n(x) \underset{\rightarrow}{P} g(x)$$

en ctp $x(\mu)$ donde

$$g_n(x) = \frac{\sum_{i=1}^{n} Y_i K\left(\frac{x-X_i}{h}\right)}{\sum_{j=1}^{n} K\left(\frac{x-X_j}{h}\right)}$$

Demostración: dividamos por $E\left(K\left(\frac{x-X_i}{h}\right)\right)$ obteniendo,

$$g_n(x) = \frac{\sum_{i=1}^{n} Y_i K\left(\frac{x-X_i}{h}\right)/E\left(K\left(\frac{x-X_i}{h}\right)\right)}{\sum_{j=1}^{n} K\left(\frac{x-X_i}{h}\right)/E\left(K\left(\frac{x-X_j}{h}\right)\right)}$$

Notemos cómo:

$$V_{ni} = Y_i K\left(\frac{x-X_i}{h}\right)/E\left(K\left(\frac{x-X_i}{h}\right)\right)$$

$$Z_{ni} = K\left(\frac{x-X_i}{h}\right)/E\left(K\left(\frac{x-X_i}{h}\right)\right)$$

$$B_{1n} = \frac{1}{n}\sum_{i=1}^{n}\left(V_{in} - E\left(V_{in}\right)\right)$$

$$B_{2n} = \frac{1}{n}\sum_{i=1}^{n}\left(Z_{in} - E\left(Z_{in}\right)\right)$$

$$A_n = E\left(V_{ni}\right) = E\left(Y_i K\left(\frac{x - x_i}{h}\right)\right) / E\left(K\left(\frac{x - X_i}{h}\right)\right)$$

Obervamos que $E(Z_{ni}) = 1$ y

$$g_n(x) = \frac{(B_{1n} + A_n)}{(B_{2n} + 1)}$$

Ahora,

$$A_n = \frac{E\left(K\left(\frac{x - X_i}{h}\right) E\left(Y/X\right)\right)}{E\left(K\left(\frac{x - X_i}{h}\right)\right)}$$

así pues aplico el lema 1.4.3, tomando $f = g(x)$;

$g(x) \in L^1$, por ser $E(|Y|) < \infty$;

$K \geq 0$ por H3 y $A_n = \mu_h(x)$ entonces

$$A_n \to g(x)$$

ctp

$$x(\mu)$$

Ahora el problema se reduce a probar que

$$B_{1n} \underset{\to}{P} 0$$

y

$$B_{2n} \underset{\to}{P} 0$$

ctp

$$x(\mu)$$

Probemos primero que

$$B_{1n} \underset{\to}{P} 0$$

Sean $N > 0$ $Y' = Y\chi_{\{|Y| \leq N\}}$ e $Y'' = Y - Y'$

$$B'_{1n} = \sum_{i=1}^{n} \left[Y'_i K\left(\frac{x - X_i}{h}\right) - E\left(Y' K\left(\frac{x - X_i}{h}\right)\right)\right] / nE\left(K\left(\frac{x - X}{h}\right)\right)$$

$$B''_{1n} = \sum_{i=1}^{n} \left[Y_i'' K\left(\frac{x-X_i}{h}\right) - E\left(Y'' K\left(\frac{x-X_i}{h}\right)\right)\right] / nE\left(K\left(\frac{x-X}{h}\right)\right)$$

Se tiene por la linealidad de E que

$$B'_{1n} + B''_{1n} = B_{1n}$$

Aplicando Chevicheb

$$P\left(|B'_{1n}| > t\right) \leq \frac{1}{n} Var\left[Y_i' K\left(\frac{x-X_i}{h}\right)\right) / t^2 E^2\left(K\left(\frac{x-X_i}{h}\right)\right) \leq \frac{1}{nt^2} \frac{E\left(Y'^2 K^2\left(\frac{x-X_i}{h}\right)\right)}{E^2\left(K\left(\frac{x-X_i}{h}\right)\right)} \leq$$

$$\leq \frac{N^2 \|K\|_\infty}{nt^2} \frac{1}{E\left(K\left(\frac{x-X_i}{h}\right)\right)}$$

Por H3:

$$E\left(K\left(\frac{x-X_i}{h}\right)\right) \geq cE\left(\chi_{\{\|u\|\leq r\}}\left(\frac{x-X}{h}\right)\right) = c \int\limits_{\|x-t\|\leq rh} \mu(dt) = c\mu(S_{rh})$$

luego

$$\frac{N^2 \|K\|_\infty}{nt^2} \frac{1}{E\left(K\left(\frac{x-X_i}{h}\right)\right)} \leq \frac{N^2 \|K\|_\infty}{nt^2} \frac{r^d h^d}{c\mu(S_{rh}) r^d h^d} = \frac{N^2 \|K\|_\infty}{nt^2 c r^d h^d} a_{rh} = \frac{cte}{nh^d} \to 0$$

ctp $x(\mu)$ para cada N, pues a_{rh} tiene límite finito.

Así pues para cada N fijo, $B'_{1n} \underset{\to}{P} 0$ ctp $x(\mu)$

Aplicando Markov a B''_{1n}

$$P\left(|B''_{1n}| > t\right) \leq \frac{E\left(|B''_{1n}|\right)}{t} \leq \frac{2}{t+} \frac{E\left(|Y''| K\left(\frac{x-X}{h}\right)\right)}{E\left(K\left(\frac{x-X}{h}\right)\right)} = \frac{2}{t} \frac{E\left(K\left(\frac{x-X}{h}\right) E\left(|Y''|/x\right)\right)}{E\left(K\left(\frac{x-X}{h}\right)\right)} =$$

$$= \frac{2}{t} g_{nN}(x)$$

donde

$$g_N(x) = E\left(|Y''|/X\right)$$

Por el lema 1.4.3 con $f = g_N$ se tiene que $\frac{2}{t} g_{nN}(x) \to g_N(x) \frac{2}{t}$ ctp $x(\mu)$

$$g_N(x) = E\left(|Y''|/X\right) = E\left(|Y - Y'/X\right)$$

$$|Y - Y'| < 2|Y|$$

y

$$|Y - Y'| \to 0$$

c.s.

$$N \to \infty$$

Así pues

$$E\left(g_N(x)\right) = E\left(E\left(|Y - Y'|/X\right)\right) = E\left(|Y - Y'|\right) \to 0$$

si

$$N \to \infty$$

además $g_N \geq 0$ y g_N decreciente, luego $g_N(x) \to 0$ ctp $x(\mu)$ con $N \to \infty$,

elijo N tal que $g_N(x)$ sea pequeño.

Si $n \to \infty$ $B_{1n} \underset{\to}{P} 0$

Vale en $A_0 = \{x : g_N(x) \to 0\}$ y $A_N = \{x : g_{Nh} \to g_N\}$ y vale en $\bigcap_{N=0}^{\infty} A_N$

Para B_{2N} se hace lo mismo pero como no están las Y no hay que truncar.

Teorema 1.4.5. Convergencia completa

Si $|Y| < \gamma < \infty$, y se verifican H1), H2), H3), $\frac{nh^d}{logn} \to \infty$

entonces $g_n(x) \underset{\to}{c} g(x)$ ctp $x(\mu)$

Demostración:

Escribimos $g_n(x) = \frac{(B_{1n}+A_n)}{(B_{2n}+1)}$

Sabemos que $A_n \to g(x)$ ctp $x(\mu)$ por el teorema anterior, luego basta ver que $B_{1n} \underset{\to}{c} 0$ y $B_{2n} \underset{\to}{c} 0$ ctp $x(\mu)$

Aplicaremos la desigualdad de Bernstein a V_{ni}

$$|V_{ni}| \leq \frac{\gamma \|K\|_\infty}{E\left(K(\frac{x-X}{h})\right)} \leq \frac{\gamma \|K\|_\infty a_{rh}(x)}{cr^d h^d}$$

que como ya vimos tiene límite finito.

$$Var(V_{ni}) \leq E(V_{ni}^2) = \frac{E\left(Y_i^2 K^2(\frac{x-X}{h})\right)}{E^2\left(Y_i^2 K^2(\frac{x-X}{h})\right)} \leq \frac{\gamma^2 \|K\|_\infty}{E(K(\frac{x-X}{h}))} \leq \frac{\gamma^2 \|K\|_\infty a_{rh}(x)}{cr^d h^d}$$

entonces

$$P(|B_{1n}| > t) \leq 2exp\left[\frac{-nt^2}{2\left(\frac{\gamma^2\|K\|_\infty a_{rh}(x)}{cr^d h^d} + \frac{\gamma\|K\|_\infty a_{rh}(x)t}{cr^d h^d}\right)}\right] =$$

$$= 2exp\left[\frac{-nt^2 cr^d h^d}{2\|K\|_\infty a_{rh}(x)\gamma(\gamma+t)}\right] \leq 2exp\left(\frac{-C_{1n}h^d}{C_2\left(\gamma+t\right)}\right)$$

ctp

$$x(\mu)$$

llamando a $\frac{c_1}{C_2(\gamma+t)} = C$ obtenemos

$$P(|B_{1n}| > t) \leq 2exp(-C(x)nh^d)$$

como

$$\frac{nh^d}{logn} \to \infty$$

entonces

$$\sum 2exp(-C(x)nh^d)$$

converge, luego $\sum P(|B_{1n}| > t)$ converge, por lo que B_{1n} converge completamente.

Análogamente se procede con B_{2n}

1.5. Distribución asintótica

Veamos un resultado sobre la distribución asintótica del estimador de regresión el cual no es con hipótesis minimales.

Teorema 1.5.1. Sean $\{(X_i, Y_i)i \geq 1\}$ variables aleatorias independientes idénticamente distribuidas con $X_i \in R$ y $Y_i \in R$.

Notemos por $F(Y_i/X_i = u)$ a la distribución condicional de Y_i, dado $X_i = u$

y por $g(u) = E(Y_1/X_1 = u)$ y $\sigma^2(u) = E((Y_1 - g(u))^2/X_1 = \mu)$

Supongamos que:

i) g es Lipchitz, y además $\lim_{\varepsilon \to 0} \frac{g(x+\epsilon u) - g(x)}{\epsilon} = g'(x,u)$

es decir, g tiene derivada en la dirección de u (si g es diferenciable se cumplen las 2 condiciones)

ii) $\sigma^2(.)$ es continua en x

iii) $X_i \sim f$ continua en x y acotada

iv) $K : R^d \to R$ acotado no negativo,

$\int K = 1$

$|t|^d K(t) \to 0$ si $|t| \to \infty$ y $t^2 K^2(t)$ integrable

v) existe $0 \leq \beta < \infty$ tal que $nh^{d+2} \to \beta$

Sea

$$g_n(x) = \sum \omega_{ni}(x) Y_i$$

con

$$\omega_{ni}(x) = \frac{K\frac{x-X_i}{h}}{\sum_{i=1}^{n} K\left(\frac{x-X_j}{h}\right)}$$

entonces

$$\left(nh^d\right)^{1/2} (g_n(x) - g(x)) \underset{\to}{W} N\left(b, \frac{\sigma^2(x)}{f(x)} \int K^2(u)du\right)$$

con $b = \beta^{1/2} \int K(t) g'(x,t) dt$

Observación: notando que $nh^{d+2} \to \beta$ es $nh^d h^2 \to \beta$ podemos observar que como $h^2 \to 0$ debe cumplirse que $nh^d \to \infty$

Demostración:

Nuestro objetivo es calcular la distribución asintótica de $g_n(x) - g(x)$

$$\left(nh^d\right)^{1/2} \left[\frac{\sum_i K\left(\frac{x-X_i}{h}\right) Y_i}{\sum_j K\left(\frac{x-X_j}{h}\right)} - g(x)\right] = (nh^d)^{1/2} \left[\frac{\sum_i K\left(\frac{x-X_i}{h}\right) Y_i - g(x) \sum_j K\left(\frac{x-X_i}{h}\right)}{\sum_j K\left(\frac{x-X_j}{h}\right)}\right] =$$

$$\left(nh^d\right)^{1/2} \frac{\left[\frac{1}{nh^d}\sum_i K\left(\frac{x-X_i}{h}\right)Y_i - g(x)\frac{1}{nh^d}\sum_j K\left(\frac{x-X_i}{h}\right)\right]}{\frac{1}{nh^d}\sum_j K\left(\frac{x-X_j}{h}\right)}$$

Como $nh^d \to \infty$ por la observación previa, $h \to 0$ y por iii), iv) y v) obtenemos

$$\frac{1}{nh^d}\sum_j K\left(\frac{x-X_j}{h}\right) \underset{\to}{P} f(x)$$

Analicemos ahora el término

$$\left(nh^d\right)^{1/2}\left[\frac{1}{nh^d}\sum_i K\left(\frac{x-X_i}{h}\right)Y_i - g(x)\frac{1}{nh^d}\sum_j K\left(\frac{x-X_i}{h}\right)\right] =$$

$$\frac{1}{(nh^d)^{1/2}}\left[\sum_i K\left(\frac{x-X_i}{h}\right)(Y_i - g(x))\right] = \frac{1}{(nh^d)^{1/2}}\left[\sum_i K\left(\frac{x-X_i}{h}\right)(Y_i - g(X_i) + g(X_i) - g(x))\right]$$

$$\frac{1}{(nh^d)^{1/2}}\sum_i K\left(\frac{x-X_i}{h}\right)(Y_i - g(X_i)) + \frac{1}{(nh^d)^{1/2}}\sum_i K\left(\frac{x-X_i}{h}\right)(g(X_i) - g(x))$$

Notemos

$$V_{ni} = K\left(\frac{x-X_i}{h}\right)(Y_i - g(X_i))$$

Las variables V_{ni} tienen media 0, y son independientes.

Tenemos un sistema triangular y se verifican las condiciones de Lindeberg.

Calculemos la varianza límite:

Condicionando $(Y_i - g(x))^2$ a X

$$\frac{1}{h^d}E(V_{in}^2) = \frac{1}{h^d}\int K^2\left(\frac{x-t}{h}\right)\sigma^2(t)f(t)dt \to \sigma^2(x)f(x)\int K^2$$

donde hemos aplicado el lema 1.1.3 ya que $\sigma^2 f$ es continua en x por las hipótesis ii) y iii)

$$Var\left(\frac{1}{(nh^d)^{1/2}}\sum_i V_{ni}\right) = \frac{1}{h^d}E(V_{ni}^2) \to \sigma^2 f(x)\int K^2(t)dt$$

Así pues aplicando el teorema de Lindeberg tenemos que:

$$\frac{1}{(nh^d)}^{1/2} \sum_i V_{ni} \to N(0, \sigma^2 f \int K^2)$$

luego,

$$\frac{1}{(nh^d)^{1/2}} \sum_i V_{ni} / \frac{1}{nh^d} \sum_j K\left(\frac{x - X_j}{h}\right) \to N\left(0, \frac{\sigma^2 f(x) \int K^2}{f(x)}\right)$$

Veamos que:

$$\frac{1}{(nh^d)^{1/2}} \sum K\left(\frac{x - X_i}{h}\right)(g(X_i) - g(x)) \overset{P}{\to} \beta^{1/2} f(x) \int K(t) g'(x,t) dt$$

$$E\left(\frac{1}{(nh^d)^{1/2}} \sum_i K\left(\frac{x - X_i}{h}\right)(g(X_i) - g(x))\right) = \frac{1}{(nh^d)^{1/2}} n \frac{h^d}{h^d} \int K\left(\frac{x-t}{h}\right)(g(t) - g(x)) f(t) dt =$$

$$= \left(nh^d\right)^{1/2} \int K(u)(g(x - uh) - g(x))\Big) f(x - uh) du =$$

$$\left(nh^d\right)^{1/2} \int K(u) \frac{(g(x - uh) - g(x))}{h} f(x - uh) h du$$

Ahora si $h \to 0$

$$K(u) \frac{g(x - uh) - g(x)}{h} f(x - uh) \to K(u) g'(x, u) f(x)$$

Además f es acotada,

$$K(u) \frac{g(x - uh) - g(x)}{h} f(x - uh) \leq C |K(u)||u|$$

y esta es una función integrable, pues $u^2 K^2$ es integrable, luego puedo entonces aplicar convergencia dominada.

Nota: En vez de pedir f acotada puedo pedir K de soporte compacto, entonces fuera de un compacto $K(u)$ vale cero, por lo tanto allí puedo hacer uh suficientemente pequeño, de donde $f(x - uh) \leq |f(x) + \epsilon|$ y la acotación sería $C|k(u)||u||f(x) + \epsilon|$.

Aplicando convergencia dominada:

$$\int K(u) \frac{(g(x - uh) - g(x))}{h} f(x - uh) du \to f(x) \int K(t) g'(x, t) dt$$

y

$$\left(nh^d\right)^{1/2} h = \left(nh^{d+2}\right)^{1/2} \to \beta^{1/2}$$

así pues,

$$(nh^d)^{1/2}h\int K(u)\frac{(g(x-uh)-g(x))}{h}f(x-uh)du \to \beta^{1/2}f(x)\int K(t)g'(x,t)dt$$

luego

$$(nh^d)^{1/2}\sum K\left(\frac{x-X_i}{h}\right)(g(x-uh)-g(x))/\frac{1}{nh^d}\sum_j K\left(\frac{x-X_j}{h}\right) \to \beta^{1/2}\int K(t)g'(x,t)dt$$

Así pues tengo $\frac{a_n+b_n}{c_n}$

$$c_n \underset{\to}{P} c$$

$$a_n \underset{\to}{D} N$$

y aplicando Slutsky

$$\frac{b_n}{c_n} \underset{\to}{P} \frac{b}{c}$$

Veamos ahora que la varianza tiende a 0:

$$Var\left(\frac{1}{(nh^d)^{1/2}}\sum_j K\left(\frac{x-X_i}{h}\right)(g(X_i)-g(x))\right) = \frac{1}{(nh^d)}nVar\left(K\left(\frac{x-X_i}{h}\right)(g(X_i)-g(x))\right) =$$

$$\frac{1}{h^d}E\left(K^2\left(\frac{x-X_i}{h}\right)(g(x_i)-g(x))^2\right) = \frac{1}{h^d}\int K^2\left(\frac{x-t}{h}\right)(g(t)-g(x))^2 f(t)dt =$$

$$\int K^2(u)(g(x-uh)-g(x))^2 f(x-uh)dt = h^2\int K^2(u)\frac{(g(x-uh)-g(x))^2}{h}f(x-uh)dt$$

ahora

$$\int K^2(u)\frac{(g(x-uh)-g(x))^2}{h}f(x-uh) \to f(x)\int K^2(u)(g'(x,u))^2du$$

la integral del lado derecho existe pues g es Lipchitz y además K^2u^2 es integrable, así pues aplico convergencia dominada y se tiene que si $h^2 \to 0$ la varianza $\to 0$.

Capítulo 2

Teoría de respuesta al ítem

2.1. Introducción

La teoría de respuesta al ítem se diferencia de la teoría clásica de los test por utilizar modelos basados en las características de los ítems en vez de las del test, donde las características de los ítems son independientes del grupo en que el ítem se ha calibrado y las puntuaciones del rasgo no dependen de las puntuaciones obtenidas en cada test particular. En la teoría de respuesta al ítem se puede obtener una medida de la precisión para cada puntuación del rasgo, lo que la distingue claramente de la teoría clásica (esto lo veremos en la sección destinada a funciones de información) y para evaluar la fiabilidad no se requieren de tests estrictamente paralelos.

La teoría de respuesta al ítem (de ahora en adelante TRI) establece una relación funcional entre la respuesta del examinado a cada ítem y el rasgo latente responsable de tal realización y al que notaremos θ. En la mayoría de los modelos que veremos se asume que esta función depende solo de un rasgo, es decir, son unidimensionales. La función que da la probabilidad de obtener determinada puntuación en el ítem condicionado al rasgo se denomina curva característica del ítem. Comenzaremos viendo las hipótesis que sustentan la TRI.

2.2. Dimensionalidad del espacio latente

Se supone que hay un conjunto de rasgos o aptitudes que subyacen en la respuesta de los individuos a un conjunto de ítems. Cada rasgo es un número real, por lo tanto si existen k rasgos estos determinan puntos del espacio R^k. A los modelos que suponen que un único rasgo es el que determina las respuestas se les llama unidimensionales. Es claro que este supuesto nunca se cumplirá estrictamente, por lo tanto, lo que se pide es que para un conjunto de ítems del test exista un factor dominante que explique la realización del test, a este factor se le llama aptitud medida por el test. Los modelos que consideran más de un factor son llamados multidimensionales.

Si se cumple el supuesto de unidimensionalidad, al aplicar a r subpoblaciones de examinados un test, las distribuciones condicionadas de las puntuaciones del test para cada nivel de rasgo son iguales para todas las subpoblaciones. En caso de no cumplirse lo anterior estamos en presencia de funcionamiento diferencial del ítem. A los efectos de la construcción de tests unidimensionales, Lumdsen recomienda el uso del análisis factorial.

A un conjunto inicial de ítems (construidos a partir de conocimientos previos) se les realiza un análisis factorial y se eliminan aquellos que no carguen en el primer factor. Una vez eliminados esos ítems se repite el procedimiento tantas veces como sea necesario, hasta dar con un solución satisfactoria. Lumdsen (1961) propone que la razón de varianza explicada por el primer factor al segundo se utilice como índice de unidimensionalidad.

2.3. Independencia local

La independencia local refiere a que los n ítems que constituyen el test, condicionados a un valor constante de θ son independientes. Formalmente: sea un test formado por n ítems y sean $U_1, U_2, \ldots, U_n$ las variables que representan las respuestas de los examinados a los ítems. Por ejemplo, los ítems dicotómicos son variables Bernoulli (acierto, error) que indican el resultado en cada ítem. Entonces, si hay independencia local se cumple:

$$P(U_1 \ldots U_n/\theta) = \prod_{j=1}^{n} P(U_j/\theta) \tag{2.1}$$

Es importante notar que la independencia local no implica que los ítems son no correlacionados sino que son independientes condicionalmente (al nivel de rasgo). Hambleton y Swaminathan (1991) muestran que el principio de unidimensionalidad es equivalente a la independencia local (condicionado a solo un rasgo) pues si suponemos que el test es undimensional y mide una variable latente θ, si no se da la independencia local resultaría que algunos sujetos tendrían mayor probabilidad de responder correctamente para el mismo nivel de rasgo, lo cual niega la unidimensionalidad. Recíprocamente, si se da la independencia local tenemos que las probabilidades de respuesta solo dependen de un único rasgo. La independencia local también puede expresarse con datos multidimensionales (condicionando al espacio completo).

Dada la equivalencia entre independencia local y unidimensionalidad, cuando se consideran datos unifactoriales, McDonald (1981) propone usar técnicas de análisis factorial para evaluar el supuesto de independencia local. Define un conjunto de ítems del test como unidimensionales, cuando para sujetos con el mismo nivel de rasgo, la covariación entre los ítems es cero. La condición de la ecuación 2.1 es necesaria pero no suficiente para la independencia dos a dos de las variables. Una alternativa posible para una definición formal de la independencia local con una condición necesaria y suficiente podría ser:

Definición 2.3.1. Los n ítems de un test son localmente independientes condicionados al rasgo θ, si para cualquier subconjunto de índices $i(1), \ldots, i(k)$

$1 \leq k \leq n \quad 1 \leq i(j) \leq n$ entonces:

$$P(U_{i(1)} \ldots U_{i(k)}/\theta) = \prod_{j=i(1)}^{i(k)} P(U_j/\theta) \tag{2.2}$$

2.4. Ausencia de factores de velocidad

En los modelos de TRI se supone que los tests se aplican en ausencia de factores de velocidad, es decir, que la puntuación que obtiene un sujeto en un ítem, se debe solo a su nivel de rasgo, pero no a falta de tiempo. Esta suposición está implícita en el supuesto de unidimensonalidad.

2.5. Curva característica del ítem

Si consideramos un ítem dicotómico i su distribución de frecuencias viene dado por

$$f_i(u/\theta) = P_i(\theta)^u Q_i(\theta)^{1-u} \tag{2.3}$$

donde $u = 1$ o 0 y $Q_i(\theta) = 1 - P_i(\theta)$. La función $P_i(\theta)$ es la esperanza condicional de la puntuación del ítem dado el nivel de rasgo, o sea, es la regresión de la puntuación del ítem sobre la aptitud. A esta función se le llama función o curva característica del ítem (CCI), si estamos en el caso de modelos unidimensionales.

Definición 2.5.1. Llamaremos curva característica del ítem a la función $P(\theta) : [-\infty, \infty] \mapsto [0, 1]$ que da la probabilidad de acertar el ítem condicionada a la variable latente θ, es decir,

$$P(\theta) = P(U = 1/\theta) \tag{2.4}$$

Si el espacio latente está completamente especificado, las curvas características del ítem se mantendrán invariantes para todo subgrupo de la población; luego la probabilidad que un individuo dé una respuesta correcta a determinado ítem solo depende de la forma de la curva característica del ítem y del nivel de rasgo del sujeto, y entonces, es independiente de la distribución del rasgo en la población. Esta propiedad de invarianza de las CCI (no importa en qué región fue calibrado el ítem) constituye una de las mejores características de los modelos de TRI y es de capital importancia en las aplicaciones. La formulación matemática de las CCI es lo que distingue unos modelos de otros.

De manera análoga se puede hablar de curva característica del test (CCT), que se define como la suma de las curvas características de los ítems que componen el test, o sea,

$$CCT = \sum_{i=1}^{n} P_i(\theta) \tag{2.5}$$

También se puede estudiar la curva característica del sujeto que relaciona la dificultad del ítem (luego veremos su significado) y la proporción de ítems respondidos correctamente para cada nivel de dificultad y que constituye una buena herramienta para comparar el proceder de las diferentes personas a través de los ítems, así como para estudiar las curvas empíricas de las personas y las que cabría esperar si se ajustasen a las predicciones del modelo.

2.6. Clasificación de los modelos

En su libro *Handbook of Modern Item Response Theory*, Van der Linden y Hambleton (1997) clasifican a los modelos en varias categorías, las cuales detallamos continuación.

Modelos para ítems dicotómicos:

- ojiva normal 1,2, 3 y 4 parámetros;
- logísticos de 1,2,3 y 4P.

Modelos para ítems politómicos:

- de respuesta nominal;
- para ítems de múltiple opción;
- de escala de clasificación;
- de respuesta graduada;
- de crédito parcial;
- secuencial para respuestas ordenadas;
- parcial generalizado.

Modelos para tiempo de respuesta o ítems de múltiples intentos:

- logístico para test con tiempo límite;
- de test de velocidad con límite de tiempo;
- modelos de respuesta a un único ítem con múltiples intentos.

Modelos para múltiples habilidades o componentes cognitivos:

- de Rasch unidimensionales;
- multidimensionales de ojiva normal;
- de respuesta multicomponente;
- modelos logísticos multidimensionales;
- loglineales para ítems politómicos.

Modelos no paramétricos:

- de ítems dicotómicos;
- de ítems politómicos;

- a partir de análisis de datos funcionales.

Modelos para ítems no monótonos:

- modelo a partir del coseno hiperbólico;
- modelo Parella.

Modelos con supuestos especiales acerca del proceso de respuesta:

- para múltiples grupos;
- logísticos mixtos;
- para respuestas localmente dependientes;
- para tests con formato que permiten información parcial.

Santiesteban y Alvarado (2001) consideran una clasificación más compacta:

- modelos para ítems dicotómicos;
- modelos para ítems politómicos;
- modelos para tiempo de respuesta;
- modelos componenciales y multidimensionales;
- modelos basados en suposiciones más flexibles.

2.7. Modelos para ítems dicotómicos

Los ítems dicotómicos son aquellos en los que una respuesta de un sujeto será clasificada en alguna de dos categorías, según responda correctamente o incorrectamente. Se puntuará el ítem 1 si se responde correctamente y 0 si se hace incorrectamente. Si un individuo no responde a un ítem se considerará su respuesta como incorrecta.

Veremos una serie de modelos para este tipo de ítems.

2.7.1. Modelo de ojiva normal

Fue primeramente propuesto por Lord (1952); y para introducir este modelo seguiremos a Baker (1992).

Consideremos una variable auxiliar $\Gamma_i(\theta)$ que indica la propensión a responder correctamente el ítem i en función del rasgo, con $-\infty \leq \Gamma_i \leq \infty$ y sea γ_i un valor tal que si $\Gamma_i \geq \gamma_i$ el ítem se responde correctamente, o sea $u_{ij} = 1$ y si $\Gamma_i < \gamma_i$ el ítem se falla, o sea $u_{ij} = 0$. La variable Γ_i no es observable. Lo que se tiene son las respuestas de los sujetos a cada ítem, o sea u_{ij} con $j = 1, 2, ..., N$.

Para modelizar a los ítems asumimos:

a)Γ_i es una función lineal de θ, es decir,

$$\Gamma_i = \alpha_i + \beta_i\theta + \epsilon_i \tag{2.6}$$

con $\epsilon_i = N(0, \sigma^2)$ y donde el rasgo se considera sin errores de medida.
Así pues

$$\Gamma_i(\theta_j) \sim N(\alpha_i + \beta_i\theta_j, \sigma_i^2) \tag{2.7}$$

Notemos por μ_{ij} la media de $\Gamma_i(\theta_j)$, o sea $\mu_{ij} = \alpha_i + \beta_i\theta_j$ luego

$$P(u_{ij} = 1|\Theta = \theta_j) = \int_{\gamma_i}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(x-\mu_{ij})^2}{2\sigma_i^2}} dx \tag{2.8}$$

Se tiene además que $\gamma_i = \alpha_i + \beta_i\theta^*$ para algún θ^*, entonces haciendo el cambio de variable $z = \frac{(x-\mu_{ij})}{\sigma_i}$ obtenemos:

$$P(u_{ij} = 1|\Theta = \theta_j) = \int_{\frac{\gamma_i-\mu_{ij}}{\sigma_i}}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \tag{2.9}$$

y además

$$\frac{\gamma_i - \mu_{ij}}{\sigma_i} = \frac{\beta_i(\theta^* - \theta_j)}{\sigma_i} \tag{2.10}$$

Podemos considerar $\beta_i = 1$ sin pérdida de generalidad obteniéndose finalmente que:

$$P(u_{ij} = 1|\Theta = \theta_j) = \int_{-\frac{(\theta_j-\theta^*)}{\sigma_i}}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \tag{2.11}$$

o equivalentemente

$$P(u_{ij} = 1|\Theta = \theta_j) = \int_{-\infty}^{\frac{(\theta_j-\theta^*)}{\sigma_i}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \tag{2.12}$$

Si llamamos $b_i = \theta^*$ y $a_i = \frac{1}{\sigma_i}$ entonces el extremo superior del intervalo de integración queda $a_i(\theta_j - b_i) = a_i\theta_j - a_ib_i$, y este se puede expresar como un modelo lineal poniendo $\lambda_i = a_i$ y $\xi_i = a_ib_i$.

Luego, la probabilidad de responder correctamente el ítem i dado que el sujeto tiene un nivel de rasgo θ_j es

$$P_i(\theta_j) = \int_{-\infty}^{a_i(\theta_j-b_i)} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = \int_{-\infty}^{\xi_i+\lambda_i\theta_j} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \tag{2.13}$$

Al parámetro b_i se le llama dificultad del ítem y representa el punto en la escala de rasgo en el cual la probabilidad de responder correctamente el ítem es 0.5.

Al parámetro a_i se le llama índice de discriminación del ítem i y es proporcional a la pendiente de la CCI en el punto $\theta = b_i$, como se puede comprobar directamente

derivando y observando que $P_i'(b_i) = \frac{a_i}{\sqrt{2\pi}}$. Hasta ahora no se han hecho supuestos sobre la distribución del rasgo en la población. Si la distribución es normal y no hay aciertos al azar en el ítem, se puede establecer una conexión con la teoría clásica de test. Richardson (1936) y Tucker (1946) mostraron que bajo esta condición el coeficiente de correlación biserial $\rho_{\theta U_i}$ cumple $\rho_{\theta U_i} = \frac{a_i}{\sqrt{1+a_i^2}}$.

Además, si notamos la dificultad del ítem (según la teoría clásica)por p_i; Tucker mostró que:

$$p_i = \frac{1}{\sqrt{2\pi}} \int_{\delta_i}^{+\infty} e^{\frac{-1}{2}z^2} dz$$

donde $\delta_i = b_i \rho_{\theta U_i}$.

El modelo que acabamos de presentar es conocido como modelo de ojiva normal de dos parámetros. Existen variaciones de éste . Si suponemos que la discriminación es uno, obtenemos el modelo de un parámetro que toma la forma

$$P_i(\theta_j) = \int_{-\infty}^{(\theta_j - b_i)} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \tag{2.14}$$

Birnbaum (1968) incorpora la posibilidad de aciertos al azar agregando un parámetro y deduce el modelo de tres parámetros.

$$P_i(\theta_j) = c_i + (1 - c_i) \int_{-\infty}^{a_i(\theta_j - b_i)} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \tag{2.15}$$

Barton y Lord (1981) han propuesto un modelo de cuatro parámetros, que no discutiremos aquí, ya que no aporta ventajas significativas respecto a los anteriores.

2.8. Modelos logísticos

Consideremos la función $\Psi(x) = \frac{e^x}{1+e^x}$ y sea $\Phi(x)$ la distribución normal acumulada, Haley (1952) mostró que

$$|\Phi(a_i(\theta_j - b_i)) - \Psi((1{,}702)a_i(\theta_j - b_i))| < {,}01 \tag{2.16}$$

A partir de esta aproximación, utilizándola en los modelos de ojiva normal, Birbaum (1968) obtiene los modelos logísticos correspondientes:

$$1 - P_i(\theta_j) = \frac{e^{D(\theta_j - b_i)}}{1 + e^D(\theta_j - b_i)} \tag{2.17}$$

$$2 - P_i(\theta_j) = \frac{e^{Da_i(\theta_j - b_i)}}{1 + e^{Da_i(\theta_j - b_i)}} \tag{2.18}$$

$$3 - P_i(\theta_j) = c_i + (1 - c_i)\frac{e^{Da_i(\theta_j - b_i)}}{1 + e^{Da_i(\theta_j - b_i)}} \tag{2.19}$$

donde D=1.702

En general el parámetro D se incorpora a los otros, y por lo tanto, de ahora en adelante supondremos que D=1. Como ya hemos visto para los modelos de ojiva normal tenemos que a_i es el parámetro de discriminación, b_i el parámetro de dificultad y c_i el parámetro que indica la posibilidad de aciertos al azar. Estos modelos tienen la ventaja de ser mucho más manejables matemáticamente y son los que comúnmente se usan en TRI.

2.8.1. Logits

Para lograr un mayor significado del modelo en su interpretación y para relacionarlo con las regresiones logísticas, a veces se utilizan los logits.

Llamemos logit a la función de θ :

$$\ln \frac{P(u_{ij}=1|\theta)}{P(u_{ij}=0|\theta)} \tag{2.20}$$

Esto nos indica la relación entre acertar y fallar el ítem, por ejemplo un logit de 0 indica que $\frac{P(u_{ij}=1|\theta)}{P(u_{ij}=0|\theta)} = 1$ o sea $P(u_{ij}=1|\theta) = P(u_{ij}=0|\theta)$. Se verifica inmediatamente que los logits para los diferentes modelos son:

- modelo de un parámetro $\theta - b_i$
- modelo de dos parámetros $a_i(\theta - b_i)$
- modelo de tres parámetros $a_i(\theta - b_i)$ (tomando $\ln \frac{P(u_{ij}=1|\theta)-c_i}{P(u_{ij}=0|\theta)}$)

El logit también es útil para comparar sujetos pues:

$$\ln[\frac{\frac{P(u_{ij}=1|\theta_1)}{P(u_{ij}=0|\theta_1)}}{\ln \frac{P(u_{ij}=1|\theta_2)}{P(u_{ij}=0|\theta_2)}}] = \ln \frac{P(u_{ij}=1|\theta_1)}{P(u_{ij}=0|\theta_1)} - \ln \frac{P(u_{ij}=1|\theta_2)}{P(u_{ij}=0|\theta_2)}$$

Esta diferencia según el caso es igual a:

a) $\theta_1 - \theta_2$ para el modelo de dos parámetros

b) $a_i(\theta_1 - \theta_2)$ para los modelos de dos y tres parámetros.

Es claro que los modelos logísticos se pueden también obtener del hecho de suponer que los logits son una función lineal de la habilidad. Por ejemplo, para el modelo de dos parámetros

$$\ln \frac{P(u_{ij}=1|\Theta=\theta_j)}{P(u_{ij}=0|\Theta=\theta_j)} = a_i(\theta_j - b_i). \tag{2.21}$$

2.9. Escala de θ

Un problema importante es la identificabilidad dentro de la teoría de respuesta al ítem.

Un test no puede dar más que información sobre la ordenación de los examinados.

Para ver esto consideremos una transformacion monótona $\tau = g(\theta)$.

Entonces,

$$P(\theta) = P(g^{-1}(g(\theta))) = P(g^{-1}(\tau)) = P^*(\tau)$$

donde ahora

$$P^* = P \circ g^{-1}$$

es ahora la curva característica del ítem relativa a τ.

Los modelos establecidos no determinan el origen ni las unidades de θ, es decir si se utiliza otro origen y se cambian las unidades de θ y si se corrigen los parámetros adecuadamente, se obtendrán las mismas probabilidades $P(\theta)$. Es claro que surge como pregunta natural cuáles son esas transformaciones. En todos los casos consideramos $D = 1$.

2.9.1. Modelo logístico de un parámetro

Teorema 2.9.1. Dado el ítem i que se describe mediante el modelo de un parámetro

$$P(\theta) = \frac{e^{\theta - b_i}}{1 + e^{\theta - b_i}}.$$

Si se transforman θ y b_i sumando una constante a cada uno, la probabilidad permanece incambiada y el modelo se mantiene.

Demostración:

Sea k un número real y transformemos θ y b_i mediante $\theta' = \theta + k$ y $b'_i = b_i + k$ entonces:

$$P(\theta') = \frac{e^{\theta' - b'_i}}{1 + e^{\theta' - b'_i}} = \frac{e^{\theta + k - b_i - k)}}{1 + e^{\theta + k - b_i - k}} = \frac{e^{\theta - b_i}}{1 + e^{\theta - b_i}} = P(\theta)$$

2.9.2. Modelo logístico de dos parámetros

Teorema 2.9.2. Dado el ítem i que se describe mediante el modelo de dos parámetros:

$$P(\theta) = \frac{e^{a_i(\theta - b_i)}}{1 + e^{a_i(\theta - b_i)}}.$$

Si se transforman θ, b_i y a_i mediante: $\theta' = m\theta + k$ $b'_i = mb_i + k$ $a'_i = \frac{a_i}{m}$ con m y k constantes arbitrarias, entonces la probabilidad permanece incambiada y el modelo se mantiene.

Demostración:

$$P(\theta') = \frac{e^{a'_i(\theta' - b'_i)}}{1 + e^{a'_i(\theta' - b'_i)}} = \frac{e^{\frac{a_i}{m}(mb_i + k - mb_i - k)}}{1 + e^{\frac{a_i}{m}(mb_i + k - mb_i - k)}} == \frac{e^{a_i(\theta - b_i)}}{1 + e^{a_i(\theta - b_i)}} = P(\theta).$$

2.9.3. Modelo logístico de tres parámetros

Teorema 2.9.3. Dado el ítem i que se describe mediante el modelo de tres parámetros

$$P(\theta) = c_i + \frac{e^{a_i(\theta - b_i)}}{1 + e^{a_i(\theta - b_i)}}.$$

Si se transforman θ, b_i y a_i mediante: $\theta' = m\theta + k$ $b'_i = mb_i + k$ $a'_i = \frac{a_i}{m}$ y c_i no se transforma, o sea $c'_i - c_i$ con m y k constantes arbitrarias, entonces la probabilidad permanece incambiada y el modelo se mantiene.

Demostración:

$$P(\theta') = c'_i + (1 - c'_i)\frac{e^{a'_i(\theta' - b'_i)}}{1 + e^{a'_i(\theta' - b'_i)}} = c_i + (1 - c_i)\frac{e^{\frac{a_i}{m}(mb_i + k - mb_i - k)}}{1 + e^{\frac{a_i}{m}(mb_i + k - mb_i - k)}} =$$

$$c_i + (1 - c_i)\frac{e^{a_i(\theta - b_i)}}{1 + e^{a_i(\theta - b_i)}} = P(\theta)$$

2.10. Modelo de Rasch

Al modelo logístico de un parámetro se le llama modelo de Rasch justamente por ser propuesto por Rasch en 1960. El modelo de Rasch es un caso particular del modelo de Birnbaum (1968) de tres parámetros, si consideramos que el parámetro de seudoazar es 0 y la discriminación es igual para todos los ítems (en general se toma igual a 1). Este modelo tiene ventajas importantes pues la función de verosimilitud condicional toma una forma muy sencilla. Una característica importante de este tipo de modelo es que la puntuación total del sujeto es un estadístico suficiente para θ. Esto significa que la probabilidad de un vector de respuestas dado depende únicamente de la puntuación total.

Rasch (1960) en su libro *Probabilistic models for some intelligence and attainment tests* se acerca al análisis de los tests desde un marco probabilístico. Presenta allí tres modelos diferentes: uno para el número de errores en la lectura de un texto, otro para la velocidad de lectura y otro para los ítems de un test. Este último modelo será luego conocido como modelo de Rasch.

La presentación matemática del modelo se basa en una matriz de datos bidimensional obtenida al administrar n ítems a N examinados. Las respuestas son puntuadas en forma dicotómica $u_{ij} = 0, 1$ donde i indica el ítem y j al sujeto. A cada individuo se asocia un vector columna y notaremos por $U = [u_{ij}]$ la matriz nxN, la matriz de datos antes mencionada. El vector que tiene los marginales de las columnas $u_{\cdot j}$ es el vector de las puntuaciones totales r_j de los examinados y el que tiene los marginales de las filas $u_{i\cdot}$ es el vector de las puntuaciones s_i de los ítems.

En este contexto Rasch emplea dos símbolos, $\eta_j \in [0, \infty]$, llamado la habilidad del examinado y $\delta_i \in [0, \infty]$, llamado la dificultad del ítem.

Define la razón $\xi_{ij} = \frac{\eta_j}{\delta_i}$ y la llama parámetro situacional; se nota que la anterior razón también se puede definir como $\xi_{ij} = \eta_j \epsilon_i$ donde ϵ_i es el parámetro de facilidad. Es decir $\delta_i = \frac{1}{\epsilon_i}$ que sirve como parámetro de localización. Es bueno observar que en este caso la dificultad del ítem no es la misma dificultad que hemos visto en los modelos de dos y tres parámetros.

Rasch utilizó la función $f(x) = \frac{x}{1+x}$, la cual es el clásico ejemplo de biyección entre R y [0,1], para poder conectar a la variable ξ_{ij} con una probabilidad. Se tiene entonces que

$$P(u_{ij} = 1|\xi_{ij}) = \frac{\xi_{ij}}{1 + \xi_{ij}} \tag{2.22}$$

y

$$P(u_{ij} = 0|\xi_{ij}) = 1 - \frac{\xi_{ij}}{1 + \xi_{ij}} = \frac{1}{1 + \xi_{ij}} \tag{2.23}$$

luego utilizando que $\xi_{ij} = \frac{\eta_j}{\delta_i}$ y sustituyendo

$$P(u_{ij} = 1|\eta_j, \delta_i) = \frac{\frac{\eta_j}{\delta_i}}{1 + \frac{\eta_j}{\delta_i}} = \frac{\eta_j}{\delta_i + \eta_j} \tag{2.24}$$

y

$$P(u_{ij} = 0|\eta_j, \delta_i) = \frac{1}{1 + \frac{\eta_j}{\delta_i}} = \frac{\delta_i}{\delta_i + \eta_j} \tag{2.25}$$

a partir de esto la forma general de la probabilidad de respuesta queda:

$$P(u_{ij}|\eta_j, \delta_i) = \frac{(\frac{\eta_j}{\delta_i})^{u_{ij}}}{1 + \frac{\eta_j}{\delta_i}} \tag{2.26}$$

El parámetro situacional también puede definirse en términos de odds ratio, o sea:

$$\xi_{ij} = \frac{P(u_{ij} = 1|\eta_j, \delta_i)}{P(u_{ij} = 0|\eta_j, \delta_i)} = \frac{\frac{\eta_j}{\delta_i + \eta_j}}{\frac{\delta_i}{\delta_i + \eta_j}} = \frac{\eta_j}{\delta_i} \tag{2.27}$$

El parámetro situacional permanece incambiado ante cambios de escala, pues si multiplicamos la habilidad y la dificultad por una constante K se tienen que $\xi_{ij} = \frac{K\eta_j}{K\delta_i} = \frac{\eta_j}{\delta_i}$

En el principio dijimos que el modelo de Rasch es un caso particular del modelo de Birnbaum, esto es así pues si definimos $\theta_j = \ln \eta_j$ y $\beta_i = \ln \delta_i$ entonces,

$$P(u_{ij}|\theta_j, \beta_i) = \frac{(e^{\theta_j} e^{-\beta_i})^{u_{ij}}}{1 + e^{\theta_j} e^{-\beta_i}} = \frac{e^{(\theta_j - \beta_i)}}{1 + e^{(\theta_j - \beta_i)}} \tag{2.28}$$

Es decir que podremos trabajar con cualquiera de las dos expresiones.

Probemos ahora lo afirmado anteriormente.

Teorema 2.10.1. La puntuación total del sujeto es un estadístico suficiente para θ.

Demostración:

Recordemos: si dada una muestra aleatoria simple $X_1, X_2, \dots X_n$ de una densidad $f(x, \theta)$, el estadístico $\widehat{\theta}$ es un estadístico suficiente para θ; si dado otro estadístico θ^*, la distribución condicional $p(\theta^*|\widehat{\theta})$ no contiene a θ, o sea θ^* no proporciona ninguna información sobre θ que no haya sido dada por $\widehat{\theta}$.

Además el teorema de factorización nos dice que si la verosimilitud L se descompone como $L = h(\theta, \widehat{\theta})k(X_1, X_2, \dots X_n)$ donde $k(X_1, X_2, \dots X_n)$ no depende de θ, entonces $\widehat{\theta}$ es un estadístico suficiente para θ.

Si calculamos la verosimilitud de un determinado patrón de respuestas tenemos:

$$L = P(u_1, u_2, \dots, u_n|\theta) = \prod_{i=1}^{n} P(u_i|\theta) \tag{2.29}$$

donde u_i puede tomar los valores 0 y 1, luego:

$$L = \prod_{i=1}^{n} P_i(\theta)^{u_i} Q_i(\theta)^{1-u_i} = \prod_{i=1}^{n} (\frac{P_i(\theta)}{1 - P_i(\theta)})^{u_i} Q_i(\theta) \tag{2.30}$$

de donde

$$L = \prod_{i=1}^{n} (\frac{(1 + e^{-D(\theta - b_i)})^{-1}}{1 - (1 + e^{-D(\theta - b_i)})^{-1}})^{u_i} Q_i(\theta) = \prod_{i=1}^{n} (e^{-D(\theta - b_i)})^{u_i} Q_i(\theta) \tag{2.31}$$

luego

$$L = e^{D\theta \sum_1^n u_i} \prod_{i=1}^{n} Q_i(\theta) e^{-D \sum_1^n b_i u_i} \tag{2.32}$$

Si consideramos $\widehat{\theta} = \sum_1^n u_i$, la igualdad anterior nos muestra que es un estadístico suficiente para θ ya que:

$$L = e^{D\theta\widehat{\theta}} \prod_{i=1}^{n} Q_i(\theta) e^{-D \sum_1^n b_i u_i} \tag{2.33}$$

con $h(\theta, \widehat{\theta}) = e^{D\theta\widehat{\theta}} \prod_{i=1}^{n} Q_i(\theta)$ y $k(u_1, u_2, \dots u_n) = e^{-D \sum_1^n b_i u_i}$

Además de ser la puntuación total un estadístico suficiente, el modelo de Rasch posee la propiedad llamada de objetividad específica. Esta propiedad nos dice que si aplicamos n ítems que se ajustan al modelo de Rasch sobre subpoblaciones que tienen rasgos θ_1 y θ_2, entonces para todo ítem i, si comparamos las razones de probabilidades en las subpoblaciones, ésta será igual a la diferencia de los valores de los rasgos, es decir:

$$[\ln(\frac{P_i(\theta_1)}{Q_i(\theta_1)}) - \ln(\frac{P_i(\theta_2)}{Q_i(\theta_2)})]/D = \theta_1 - \theta_2 \tag{2.34}$$

Esto debido a que

$$\ln(\frac{P_i(\theta_1)}{Q_i(\theta_1)}) = \ln(\frac{[1+exp(-Da(\theta_1-b_i))]^{-1}}{1-[1+exp(-Da(\theta_1-b_i))]^{-1}}) = D(\theta_1-b_i) \tag{2.35}$$

e igualmente para θ_2. La propiedad es objetiva pues la relación no depende del ítem i utilizado y es específica pues depende de los n ítems que estiman el rasgo. Cualquier modelo de un parámetro para datos dicotómicos tiene esta propiedad. De los modelos logísticos sólo el modelo de Rasch tiene las propiedades de objetividad específica y de poseer un estadístico suficiente para la aptitud; esto permite que se mantengan las propiedades de la escala de intervalo para examinados e ítems.

Los modelos de Rasch en ítems de respuesta múltiple, en general, no ajustan correctamente, aunque hay desarrollos que tratan de sobreponerse a este problema.

Sigamos con el enfoque de Rasch de su modelo y cómo dedujo las probabilidades utilizando matemática combinatoria.

Tenemos que

$$P(u_{ij}|\eta_j,\delta_i) = \frac{(\frac{\eta_j}{\delta_i})^{u_{ij}}}{1+\frac{\eta_j}{\delta_i}} = \frac{(\eta_j\epsilon_i)^{u_{ij}}}{1+\eta_j\epsilon_i}$$

luego dada la independencia local, la probabilidad de un patrón dado será:

$$P(u_{1j},u_{2j},..,u_{nj}|\eta_j) = \frac{\prod_{i=1}^n(\eta_j\epsilon_i)^{u_{ij}}}{\prod_{i=1}^n(1+\eta_j\epsilon_i)} = \frac{\eta_j^{u_{.j}}\prod_{i=1}^n\epsilon_i^{u_{ij}}}{\prod_{i=1}^n(1+\eta_j\epsilon_i)} \tag{2.36}$$

Si queremos calcular la probabilidad de obtener una puntuación de r se debe observar que esta puntuación se puede obtener de $\binom{n}{r}$ maneras, luego

$$P(u_{.j}=r|\eta_j) = \frac{\eta_j^r\sum_{\{(u_{ij}):u_{.j}=r\}}\prod_{i=1}^n\epsilon_i^{u_{ij}}}{\prod_{i=1}^n(1+\eta_j\epsilon_i)}$$

Tenemos que:

$$\sum_{\{(u_{ij}):u_{.j}=r\}}\prod_{i=1}^n\epsilon_i^{u_{ij}} = (\epsilon_1\epsilon_2...\epsilon_r)+(\epsilon_2\epsilon_3...\epsilon_{r+1})+...+(\epsilon_{n-r}\epsilon_2...\epsilon_n)$$

Como

$$\sum_{r=0}^n P(u_{.j}=r|\eta_j) = \sum_{r=0}^n\frac{\eta_j^r\sum_{\{(u_{ij}):u_{.j}=r\}}\prod_{i=1}^n\epsilon_i^{u_{ij}}}{\prod_{i=1}^n(1+\eta_j\epsilon_i)} = 1$$

entonces

$$\sum_{r=0}^n\eta_j^r\sum_{\{(u_{ij}):u_{.j}=r\}}\prod_{i=1}^n\epsilon_i^{u_{ij}} = \prod_{i=1}^n(1+\eta_j\epsilon_i)$$

Si calculamos la probabilidad de un determinado patrón de respuestas dado que la puntuación total es r se tiene que:

$$P((u_{ij})|u_{.j}=r)=\frac{P((u_{ij})|\eta_j)}{P(u_{.j}=r|\eta_j)}=\frac{\frac{\eta_j^r\prod_{i=1}^n\epsilon_i^{u_{ij}}}{\prod_{i=1}^n(1+\eta_j\epsilon_i)}}{\frac{\eta_j^r\sum_{\{(u_{ij}):u_{.j}=r\}}\prod_{i=1}^n\epsilon_i^{u_{ij}}}{\prod_{i=1}^n(1+\eta_j\epsilon_i)}} \tag{2.37}$$

$$P((u_{ij})|u_{.j}=r)=\frac{\prod_{i=1}^n\epsilon_i^{u_{ij}}}{\sum_{\{(u_{ij}):u_{.j}=r\}}\prod_{i=1}^n\epsilon_i^{u_{ij}}} \tag{2.38}$$

Lo primero que se observa es que en esta probabilidad condicional no aparece el parámetro de habilidad, o sea, dado un puntaje la probabilidad de un determinado patrón, solo depende de los parámetros de los ítems. Esto también nos conduce al hecho de que una vez obtenido $u_{.j}$ toda otra información de cómo se obtuvo el puntaje es innecesaria para estimar la habilidad, o sea $u_{.j}$ es un estadístico suficiente. Para los n sujetos en los n ítems tenemos que:

$$P((u_{ij})|(\eta_j)(\epsilon_i))=\prod_{j=1}^N\prod_{i=1}^n P(u_{ij}|\eta_j,\epsilon_i)=\frac{\prod_{j=1}^N\prod_{i=1}^n(\eta_j\epsilon_i)^{u_{ij}}}{\prod_{j=1}^N\prod_{i=1}^n(1+\eta_j\epsilon_i)} \tag{2.39}$$

queda entonces

$$P((u_{ij})|(\eta_j)(\epsilon_i))=\frac{\prod_{j=1}^N\eta_j^{u_{.j}}\prod_{i=1}^n\epsilon_i^{u_{i.}}}{\prod_{j=1}^N\prod_{i=1}^n(1+\eta_j\epsilon_i)}=\frac{\prod_{j=1}^N\eta_j^{r_j}\prod_{i=1}^n\epsilon_i^{s_i}}{\prod_{j=1}^N\prod_{i=1}^n(1+\eta_j\epsilon_i)} \tag{2.40}$$

Sea $r=r_1+...+r_N$ y $s=s_1+...+s_n$ y sean los conjuntos

$A=\{matrices[u_{ij}]:(u_{i.})=(s_i);(u_{.j})=(r_j)\}$

$A_j=\{matrices[u_{ij}]:(u_{.j})=(r_j)\}$

$A_i=\{matrices[u_{ij}]:(u_{i.})=(s_i)\}$

Tenemos que $\#A$ es el número de formas de obtener este tipo de puntuaciones. A partir de esto:

$$P((u_{.j}=r_j)(u_{i.}=s_i)|(eta_j)(\epsilon_i))=\frac{\prod_{j=1}^N\eta_j^{r_j}\prod_{i=1}^n\epsilon_i^{s_i}}{\prod_{j=1}^N\prod_{i=1}^n(1+\eta_j\epsilon_i)}\#A \tag{2.41}$$

luego

$$P([u_{ij}]|(u_{.j} = r_j)(u_{i.} = s_i)) = \frac{1}{\#A} \tag{2.42}$$

Esto muestra que r_j es un estadístico suficiente para η_j y s_i para ϵ_i ya que la probabilidad condicional es independiente de todos los parámetros. Operando Rasch se llega a que:

$$P((u_{.j} = r_j)|(u_{i.} = s_i)(\eta_j)) = \frac{\#A \prod_{j=1}^{N} \eta_j^{u_{.j}}}{\sum_{A_j} \#A \prod_{j=1}^{N} \epsilon_i^{s_i}} = \frac{\prod_{j=1}^{N} \eta_j^{u_{.j}}}{\sum_{A_j} \prod_{j=1}^{N} \epsilon_i^{s_i}} \tag{2.43}$$

$$P((u_{i.} = s_i)|(u_{.j} = r_j)(\epsilon_i)) = \frac{\#A \prod_{i=1}^{n} \epsilon_i^{u_{i.}}}{\sum_{s_i}^{s} \#A \prod_{i=1}^{n} eta_j^{r_j}} = \frac{\prod_{i=1}^{n} \epsilon_i^{u_{i.}}}{\sum_{s_i}^{s} \prod_{i=1}^{n} eta_j^{r_j}} \tag{2.44}$$

Esto muestra la separación de parámetros dado que la primera ecuación no involucra los parámetros de los ítems y la segunda los parámetros de habilidad. La importancia de este hallazgo radica en que se puede utilizar este resultado para la estimación de los parámetros.

2.11. Función de información

Birnbaum (1968) define la función de información para una puntuación X como la razón entre el cuadrado de la pendiente de la regresión de $X|\theta$ y el error estándar de medida de X para un θ dado:

$$I(\theta, X) = \frac{(\frac{\partial \mu_{X|\theta}}{\partial \theta})^2}{\sigma^2_{X|\theta}} \tag{2.45}$$

y cuando esa puntuación corresponde a la de un ítem i esta función se puede expresar como

$$`I_i(\theta) = \frac{[P_i'(\theta)]^2}{P_i(\theta)Q_i(\theta)} \tag{2.46}$$

Esto tiene conexión con la estimación por máxima verosimilitud de la habilidad y con la distribución muestral. Cuando Birnbaum (1968) desarrolló su trabajo original, el medio de obtener el estimador por máxima verosimilitud de la habilidad, no había sido desarrollado totalmente. Cramer demostró bajo condiciones generales que los estimadores máximo verosímiles, por ejemplo $\widehat{\theta}$, alcanzan la cota de Cramer-Rao $-E(\frac{\partial^2 \ln L}{\partial \theta^2})$ donde

L es la verosimilitud, son insesgados $E(\widehat{\theta}) = \theta$ y tienen distribución asintótica normal con varianza el inverso de dicha cota, o sea, el inverso de la información de Fisher. Concretamente: asintóticamente,

$$\widehat{\theta} \sim N(\theta, \frac{1}{I(\theta)}) = N(\theta, -\frac{1}{E(\frac{\partial^2 \ln L}{\partial \theta^2})}) \tag{2.47}$$

Analicemos la información de Fisher en nuestro caso. Supongamos que un individuo responde a un test de n ítems y sus respuestas son puntuadas 0 si falla o 1 si acierta. Además, supongamos conocidos los parámetros del ítem. Notemos por u_{ij} a la variable dicotómica que representa la respuesta del sujeto j al ítem i.

Sea para el individuo j el vector de respuestas condicionado a su nivel de rasgo $U_j = (u_{1j}, u_{2j}, ..., u_{nj}|\theta_j)$. Bajo el principio de independencia local las u_{ij} son independientes, a partir de esto la verosimilitud queda:

$$P(U_j|\theta_j) = \prod_{i=1}^{n} P_{ij}^{u_{ij}} Q_{ij}^{1-u_{ij}} \tag{2.48}$$

donde $P_{ij} = P_i(\theta_j)$.

Tomando logaritmo:

$$\ln L = \ln P(U_j|\theta_j) = \sum_{i=1}^{n} [u_{ij} \ln P_{ij} + (1 - u_{ij}) \ln Q_{ij}] \tag{2.49}$$

en la última expresión omitimos θ_j en P y Q para simplificar la notación. Ahora,

$$\frac{\partial \ln L}{\partial \theta_j} = \sum_{i=1}^{n} [u_{ij} \frac{\partial \ln P_{ij}}{\partial \theta_j} + (1 - u_{ij}) \frac{\partial \ln Q_{ij}}{\partial \theta_j}] = \sum_{i=1}^{n} [u_{ij} \frac{P'_{ij}}{P_{ij}} + (1 - u_{ij}) \frac{Q'_{ij}}{Q_{ij}}] \tag{2.50}$$

$$\frac{\partial^2 L}{\partial \theta_j^2} = \sum_{i=1}^{n} [u_{ij} \frac{P''_{ij} P_{ij} - P'^2_{ij}}{P_{ij}^2} + (1 - u_{ij}) \frac{Q''_{ij} Q_{ij} - Q'^2_{ij}}{Q_{ij}^2}] \tag{2.51}$$

tomando esperanza:

$$E(\frac{\partial^2 \ln L}{\partial \theta_j^2}) = \sum_{i=1}^{n} [P_{ij} \frac{P''_{ij} P_{ij} - P'^2_{ij}}{P_{ij}^2} + (1 - P_{ij}) \frac{Q''_{ij} Q_{ij} - Q'^2_{ij}}{Q_{ij}^2}] =$$

$$\sum_{i=1}^{n} [\frac{P''_{ij} P_{ij} - P'^2_{ij}}{P_{ij}} + \frac{Q''_{ij} Q_{ij} - Q'^2_{ij}}{Q_{ij}}] = \sum_{i=1}^{n} \frac{P''_{ij} P_{ij} Q_{ij} - P'^2_{ij} Q_{ij} + Q''_{ij} Q_{ij} P_{ij} - Q'^2_{ij} P_{ij}}{P_{ij} Q_{ij}} =$$

$$\sum_{i=1}^{n} \frac{P''_{ij}P_{ij}Q_{ij} - P'^2_{ij}Q_{ij} + (-P''_{ij})Q_{ij}P_{ij} - (-P'_{ij})^2 P_{ij}}{P_{ij}Q_{ij}} = \sum_{i=1}^{n} \frac{-P'^2_{ij}Q_{ij} - P'^2_{ij}P_{ij}}{P_{ij}Q_{ij}} =$$

$$\sum_{i=1}^{n} \frac{-P'^2_{ij}(1 - P_{ij}) - P'^2_{ij}P_{ij}}{P_{ij}Q_{ij}} = -\sum_{i=1}^{n} \frac{P'^2_{ij}}{P_{ij}Q_{ij}}$$

luego

$$I(\theta_j) = -E(\frac{\partial^2 L}{\partial \theta_j^2}) = \sum_{i=1}^{n} \frac{P'^2_i(\theta_j)}{P_i(\theta_j)Q_i(\theta_j)} \tag{2.52}$$

que es la información del test y se descompone en la suma de las contribuciones de cada ítem, por lo cual se define la información del ítem como:

$$I_i(\theta_j) = \frac{[P'_i(\theta_j)]^2}{P_i(\theta_j)Q_i(\theta_j)} \tag{2.53}$$

Luego por lo anterior sabemos que asintóticamente

$$\widehat{\theta} \sim N(\theta, \frac{1}{I(\theta)}) = N(\theta, \frac{1}{\sum_{i=1}^{n} \frac{P'^2_i(\theta_j)}{P_i(\theta_j)Q_i(\theta_j)}}) \tag{2.54}$$

esto nos permite a un nivel dado θ_j establecer intervalos de confianza para θ_j para un nivel α mediante $\widehat{\theta} \pm z_\alpha \frac{1}{\sqrt{I(\theta)}}$ donde z_α se determina a partir de la distribución normal.

Nos interesa ver cómo queda la función de información para cada uno de los modelos. En todos los casos consideramos D=1

Modelo normal:

$$P'_i(\theta) = a_i \frac{e^{\frac{a_i^2(\theta - b_i)^2}{2}}}{\sqrt{2\pi}}$$

y

$$I(\theta) = \sum_{i=1}^{n} a_i^2 \frac{e^{a_i^2(\theta - b_i)^2}}{2\pi P_i(\theta)Q_i(\theta)}$$

Modelo de un parámetro:

$$P'_i(\theta) = P_i(\theta)Q_i(\theta)$$

y

$$I(\theta) = \sum_{i=1}^{n} P_i(\theta)Q_i(\theta)$$

Modelo de dos parámetros:

$$P_i'(\theta) = a_i P_i(\theta)Q_i(\theta)$$

y

$$I(\theta) = \sum_{i=1}^{n} a_i^2 P_i(\theta)Q_i(\theta)$$

Modelo de tres parámetros:

$$P_i'(\theta) = \frac{a_i[P_i(\theta) - c_i]Q_i(\theta)}{1 - c_i}$$

y

$$I(\theta) = \sum_{i=1}^{n} \frac{a_i^2[P_i(\theta) - c_i]^2 Q_i(\theta)}{P_i(\theta)(1 - c_i)^2}$$

Pasemos ahora a comentar algunos hechos interesantes de la función de información de los ítems. Ya vimos que la suma de las funciones de información de los ítems da la función de información del test. Esta propiedad nos permite, por lo tanto, si se tuvieran suficientes ítems, llegar a la función de información del test que deseamos, adicionando ítems convenientes al test. Actualmente, la función de información de los ítems es de las herramientas más utilizadas por quienes construyen tests, permitiéndoles mediante la combinación de los ítems obtener tests ajustados a sus necesidades. La función de información del ítem es un poderoso instrumento para el análisis de los ítems, no solamente por la cantidad de información que el ítem aporta a la medida de θ, sino también a qué nivel de θ aporta dicha información. También la función de información de los ítems se utiliza mucho en los tests adaptativos computarizados.

El valor de rasgo que aporta máxima información para un ítem dado es para el modelo de uno y dos parámetros $\theta = b_i$, con información en ese punto igual a 0,25 en el modelo de un parámetro, y $0{,}25a_i^2$ en el de dos.

Para el modelo de tres parámetros $\theta = b_i + \frac{1}{a_i}\ln[\frac{1+\sqrt{1+8c_i}}{2}]$ y la información máxima es $\frac{a_i^2}{8(1-c_i)^2}[1 - 20c_i - 8c_i^2 + (1 + 8c_i)^{3/2}]$.

Una manera de medir la eficacia de dos tests para un valor dado de θ es mediante su eficiencia relativa definida como el cociente de sus funciones de información:

$$ER = \frac{I_1(\theta)}{I_2(\theta)}$$

2.12. Estimación

Una vez definido el modelo a utilizar deben estimarse sus parámetros lo que usualmente se denomina calibración del ítem. Otro problema es estimar los rasgos o habilidades de los sujetos. Hay varios procedimientos para llevar a cabo la estimación de los ítems, como ser los procedimentos basados en máxima verosimilitud y métodos bayesianos. También se pueden estimar los parámetros de los ítems suponiendo conocida la habilidad, o estimar la habilidad suponiendo conocidos los parámetros de los ítems. Asimismo se puede estimar al mismo tiempo los parámetros de los ítems y el rasgo. La importancia de encontrar procedimientos que provean de estimadores que tengan buenas propiedades estadísticas es clara; pero además que sean eficientes computacionalmente ya que si estimamos los parámetros de n ítems y el rasgo de N sujetos, si utilizamos un modelo de tres parámetros, habría 3n+N parámetros a estimar (aunque en realidad luego de fijar la escala se tendrán que estimar 3n+N-2 parámetros) lo cual para valores grandes de N y n puede ser muy costoso desde el punto de vista computacional. A su vez, para tener una buena estimación se necesita que el tamaño de la muestra sea grande, lo cual todavía acentúa el problema anterior.

Primeramente veamos un proceso para hallar raíces de funciones que nos será de utilidad durante el proceso de estimación.

2.12.1. Método de Newton Raphson

Este es un algoritmo genérico para hallar raíces de funciones. Nosotros lo veremos en el contexto de nuestro problema, por esto nos remitimos a hallar raíces de gradientes.

En lo que viene tendremos que obtener en varias oportunidades el máximo de una función escalar de varias variables $H(x_1, .., x_*)$. Para ello, primeramente, deberemos obtener los puntos estacionarios, o sea los puntos donde se anula el gradiente.

Nuestro problema se reduce entonces a encontrar los puntos donde

$$(\frac{\partial H(x_1, ..., x_k)}{\partial x_1}, ..., \frac{\partial H(x_1, ..., x_k)}{\partial x_k}) = (0, ..., 0) \tag{2.55}$$

En reiteradas ocasiones el sistema anterior no se puede resolver explícitamente. Veremos un método que nos permite aproximarnos a la solución.

Recordemos primero el desarrollo de Taylor de primer orden para campos vectoriales. Sea $f : S \longrightarrow R^m$ con $S \subseteq R^n$ entonces

$$f(a + v) = f(a) + Df(a)v + \|v\|E(a, v) \tag{2.56}$$

donde $E(a, v) \to 0$ si $v \to 0$ con

$$\mathbf{Df(a)} = \begin{pmatrix} D_1 f_1(a) & D_2 f_1(a) & \dots & D_n f_1(a) \\ \dots & \dots & \dots & \dots \\ D_1 f_m(a) & D_2 f_m(a) & \dots & D_n f_m(a) \end{pmatrix} \tag{2.57}$$

Como en nuestro caso particular trabajaremos con 3, 2 o 1 variable, o podremos reducirlo a este caso mediante algunos supuestos, desarrollaremos la solución para el caso de 2 variables, pues ya es un caso multivariado y es lo suficientemente simple para desarrollarlo sin complicaciones de notación. Tomar este caso no tiene pérdida de generalidad y puede extenderse el resultado de manera natural a cualquier dimensión.

Consideremos entonces ahora la función definida por: $f : R^2 \longrightarrow R^2$ que lleva:

$$(x, y) \longrightarrow (\frac{\partial H}{\partial x}, \frac{\partial H}{\partial y})$$

.

Consideremos dos puntos del plano (x_t, y_t) y (x_{t+1}, y_{t+1}) tales que:

$$(x_{t+1}, y_{t+1}) = (x_t, y_t) + (\Delta x_t, \Delta y_t) \tag{2.58}$$

Aplicando el desarrollo anterior a esta función obtenemos:

$$\begin{pmatrix} \frac{\partial H}{\partial x} \\ \frac{\partial H}{\partial y} \end{pmatrix}_{t+1} = \begin{pmatrix} \frac{\partial H}{\partial x} \\ \frac{\partial L}{\partial y} \end{pmatrix}_t + \begin{pmatrix} \frac{\partial^2 H}{\partial x^2} & \frac{\partial^2 H}{\partial x \partial y} \\ \frac{\partial^2 H}{\partial y \partial x} & \frac{\partial^2 H}{\partial y^2} \end{pmatrix}_t \begin{pmatrix} \Delta x_t \\ \Delta y_t \end{pmatrix} +$$

$$+\|(\Delta x_t, \Delta y_t)\| E((x_t, y_t), (\Delta x_t, \Delta y_t))$$

Despreciando los términos de mayor orden; utilizando que

$$\nabla H = (\frac{\partial H}{\partial x}, \frac{\partial L}{\partial y}) = (0, 0)$$

obtenemos el sistema de ecuaciones:

$$\begin{aligned} -[\tfrac{\partial L}{\partial x}]_{(x_t, y_t)} &= [\tfrac{\partial^2 L}{\partial x^2}]_{(x_t, y_t)} \Delta x_t + [\tfrac{\partial^2 H}{\partial x \partial y}]_{(x_t, y_t)} \Delta y_t \\ -[\tfrac{\partial L}{\partial y}]_{(x_t, y_t)} &= [\tfrac{\partial^2 L}{\partial y^2}]_{(x_t, y_t)} \Delta y_t + [\tfrac{\partial^2 H}{\partial y \partial x}]_{(x_t, y_t)} \Delta x_t \end{aligned} \tag{2.59}$$

o sea:

$$\begin{pmatrix} \frac{\partial L}{\partial x} \\ \frac{\partial L}{\partial y} \end{pmatrix}_{(x_t, y_t)} = - \begin{pmatrix} \frac{\partial^2 L}{\partial x^2} & \frac{\partial^2 H}{\partial x \partial y} \\ \frac{\partial^2 L}{\partial y^2} & \frac{\partial^2 H}{\partial y \partial x} \end{pmatrix}_{(x_t, y_t)} \begin{pmatrix} \Delta x_t \\ \Delta y_t \end{pmatrix} \tag{2.60}$$

de donde

$$\begin{pmatrix} \Delta x_t \\ \Delta y_t \end{pmatrix} = - \begin{pmatrix} \frac{\partial^2 L}{\partial x^2} & \frac{\partial^2 H}{\partial x \partial y} \\ \frac{\partial^2 L}{\partial y^2} & \frac{\partial^2 H}{\partial y \partial x} \end{pmatrix}^{-1}_{(x_t, y_t)} \begin{pmatrix} \frac{\partial L}{\partial x} \\ \frac{\partial L}{\partial y} \end{pmatrix}_{(x_t, y_t)} \tag{2.61}$$

Así, esto nos permite definir un proceso iterativo para hallar los puntos estacionarios de la función H. Dado un punto de inicio $(\widehat{x_1}, \widehat{y_1})$ definimos en forma recurrente los sucesivos puntos que aproximan la solución por

$$(x_{t+1}, y_{t+1}) = (x_t, y_t) + (\Delta x_t, \Delta y_t) \tag{2.62}$$

o

$$\begin{pmatrix} x \\ y \end{pmatrix}_{t+1} = \begin{pmatrix} x \\ y \end{pmatrix}_t - \begin{pmatrix} \frac{\partial^2 L}{\partial x^2} & \frac{\partial^2 H}{\partial x \partial y} \\ \frac{\partial^2 L}{\partial y^2} & \frac{\partial^2 H}{\partial y \partial x} \end{pmatrix}_t^{-1} \begin{pmatrix} \frac{\partial L}{\partial x} \\ \frac{\partial L}{\partial y} \end{pmatrix}_t \tag{2.63}$$

el algoritmo continúa hasta que $|(x_{t+1}, y_{t+1}) - (x_t, y_t)| < \epsilon$ con ϵ un valor pequeño prefijado.

A modo de generalización es trivial darse cuenta que si la función H depende de n variables x_i, $i = 1, ...n$ para hallar los ceros del gradiente el algortitmo queda

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}_{t+1} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}_t - \begin{pmatrix} H_{11} & \dots & H_{1n} \\ \vdots & \vdots & \vdots \\ H_{n1} & \dots & H_{nn} \end{pmatrix}_t^{-1} \begin{pmatrix} H_1 \\ \vdots \\ H_n \end{pmatrix}_t \tag{2.64}$$

donde $H_{ij} = \frac{\partial^2 H}{\partial x_i \partial x_j}$.

2.12.2. Estimación por máxima verosimilitud

Supongamos que tenemos una función de densidad $p(x|\Theta)$ determinada por un conjunto de parámetros Θ. Tenemos una muestra aleatoria simple de tamaño n de observaciones con esta distribución, sea $X = \{x_1, ..., x_n\}$. Como son independientes la densidad conjunta de esta muestra es:

$$p(X|\Theta) = \prod_{i=1}^{n} p(x_i|\Theta) = L(\Theta|X) \tag{2.65}$$

La función L se llama verosimilitud de los parámetros dado los datos. El método de máxima verosimilitud propone estimar los parámetros maximizando la función de verosimilitud, o sea los estimadores son:

$$\Theta^* = argmax_{\Theta} L(\Theta|X) \tag{2.66}$$

En general, en vez de maximizar L se maximiza $\ln L$ pues es más sencillo. Los estimadores máximo verosímiles son de los más usados en estadística debido a sus buenas propiedades formales. Estos estimadores:

a) son insesgados o asintóticamente insesgados y consistentes;

b) son estadísticos suficientes o función de ellos;

c) son eficientes;

d) son asintóticamente normales y alcanzan la cota de Cramer-Rao.

Concretamente su distribución en muestras grandes se puede enunciar como: Los estimadores máximo-verosímiles $\widehat{\theta}_1$,$\widehat{\theta}_2$...,$\widehat{\theta}_k$ de los parámetros de una distribución dada por $f(x, \theta_1, \theta_2..., \theta_k)$ para muestras de tamaño n, tienen, para muestras grandes, una distribución que se aproxima a la normal multivariante de medias $\theta_1, \theta_2, ..., \theta_k$ y matriz nR en la forma cuadrática en donde: $r_{ij} = -E[\frac{\partial^2}{\partial\theta_i \partial\theta_j} \ln f(x, \theta_1, \theta_2..., \theta_k)]$. Las varianzas y covarianzas de los estimadores son $(1/n)R^{-1}$.

Esta propiedad se relaciona con la información de Fisher. Asintóticamente,

$$\widehat{\theta} \sim N(\theta, \frac{1}{I(\theta)}) = N(\theta, -\frac{1}{E(\frac{\partial^2 \ln L}{\partial\theta^2})})$$

2.12.3. Algoritmo EM

Este algoritmo fue propuesto por primera vez por Dempster, Laird y Rubin (1977) y es un algoritmo iterativo para obtener estimadores de máxima verosimilitud cuando las observaciones pueden ser vistas como datos incompletos o hay valores perdidos. Hay dos aplicaciones principales de este algoritmo: una es cuando hay valores *missing* debido a

problemas o limitaciones en el proceso de observación. El segundo caso es cuando optimizar la verosimilitud es analíticamente intratable pero puede ser simplificada asumiendo la existencia y valores de parámetros no observados u ocultos. El algoritmo se llama EM porque se compone de un paso donde se toman esperanzas seguido por un paso de maximización. Supongamos que los datos observados X son generados por alguna distribución, llamaremos X a los datos incompletos. Supongamos que los datos completos vienen dados por Z=(X,Y) y tienen densidad conjunta $p(z|\Theta) = p(x, y|\Theta) = p(y|x, \Theta)p(x|\Theta)$

Podemos definir una nueva función de verosimilitud llamada la verosimilitud para los datos completos, definida por:

$$L(\Theta|Z) = L(\Theta|X, Y) = p(X, Y|\Theta)$$

Esta función es una variable aleatoria pues Y es desconocido y aleatorio. Podemos pensar entonces $L(\Theta|X, Y) = h_{X,\Theta}(Y)$ donde X y Θ en esta función son constantes e Y una variable aleatoria.

El algoritmo EM primero encuentra el valor esperado del logaritmo de la verosimilitud para los datos completos con respecto a los datos desconocidos Y, dado los valores observados X y los estimadores de los parámetros en ese momento. Esto es, se define: $Q(\Theta, \Theta^{(i-1)}) = E(\ln p(X, Y|\Theta)|X, \Theta^{(i-1)})$ donde $\Theta^{(i-1)}$ son los parámetros actuales usados para evaluar las esperanzas y Θ son los nuevos parámetros. En la expresión previa X y $\Theta^{(i-1)}$ son constantes, Y es una variable aleatoria que tiene distribución $f(y|X, \Theta^{(i-1)})$ y Θ los parámetros a estimar. Calculando la esperanza condicional tenemos:

$$E(\ln p(X, Y|\Theta)|X, \Theta^{(i-1)}) == \int_{y\in\Omega} \ln p(X, y|\Theta) f(y|X, \Theta^{(i-1)}) dy \tag{2.67}$$

donde $f(y|X, \Theta^{(i-1)})$ es la distribución marginal de los datos no observados y es dependiente de X y de los parámetros observados, Ω es el espacio de valores de y. El cálculo de esta esperanza se denomina paso E. Luego, el segundo paso consiste en maximizar $Q(\Theta, \Theta^{(i-1)})$, o sea:

$$\Theta^i = argmax_\Theta Q(\Theta, \Theta^{(i-1)}) \tag{2.68}$$

Una modificación del algoritmo EM consiste, en vez de maximizar Q, en encontrar algún Θ^i tal que $Q(\Theta^i, \Theta^{(i-1)}) > Q(\Theta, \Theta^{(i-1)})$.
Esta modificación da origen al algoritmo EM generalizado (GEM). Dumpster et al. (1977) prueban que en cada iteración crece el logaritmo de la verosimilitud y que el algoritmo

converge a un máximo local de la verosimilitud. También dan tasas y velocidades de convergencia.

2.12.4. Revisión de los procedimientos

En esta sección se pretende hacer un recuento de los métodos y circunstancias en que se aplican. En las secciones siguientes veremos los algoritmos. En general, se aplican métodos basados en máxima verosimilitud o bayesianos.

Una primera división, no de métodos pero sí de nuestros objetos a estimar, es si:

1- Se estiman los parámetros de los ítems suponiendo conocida la habilidad. Se emplean procedimientos por:

a) máxima verosimilitud,

b) bayesianos y

c) mínimo χ^2.

2- Se estima la habilidad suponiendo conocidos los parámetros de los ítems. Estimación por:

a) máxima verosimilitud,

b) bayesiana.

3- Se estiman los parámetros de los ítems y la habilidad. Estimación:

a) conjunta por máxima verosimilitud,

b) por máxima verosimilitud condicional,

c) por máxima verosimilitud marginal,

d) bayesiana.

4- Métodos alternativos.

Detallamos aquí métodos alternativos que pretenden dar procedimientos de bajo coste computacional. Dentro de éstos está el que utiliza la correlación biserial puntual y la dificutad del ítem (este método se verá luego como parte de los algoritmos para encontrar estimaciones iniciales). También está el método de Urry para estimar los parámetros del modelo 3P. Verhelst y Molenaar (1988) usan regresión logística con métodos iterativos.

En la estimación aparecen ciertos detalles a tener en cuenta. El primero tiene que ver con el problema de la identificación y es producto de lo que vimos en el trabajo de TRI de cómo si sumamos una constante a la dificultad de los ítems y al rasgo, la CCI permanece invariante. Para resolver este problema veremos que se fijará la media de θ a cero y su varianza a 1 (por consiguiente también se transformará la dificultad del ítem).

Se han probado resultados sobre la convergencia de los estimadores:

a) Si los parámetros de los ítems son conocidos el estimador máximo verosímil $\widehat{\theta}$ de la habilidad converge a θ si el número de ítems es suficientemente grande.

b) Si la aptitud es conocida los estimadores máximo verosímiles de los parámetros de los ítems convergen al valor de esos parámetros si el número de examinados es suficientemente grande.

c) Si se estiman conjuntamente los parámetros de los ítems y la habilidad, los parámetros de los ítems son estructurales y los de la habilidad incidentales (pues crece el número de parámetros al aumentar la muestra). En este caso los estimadores de los parámetros de los ítems no convergen a sus parámetros cuando la muestra de examinados se incrementa. Sin embargo, se logra consistencia si además se hace que el número de ítems tienda a infinito. Es decir, para la estimación simultánea se necesitan grandes muestras de examinados y de ítems.

Describiremos ahora diferentes procedimientos y deduciremos los algoritmos, así como sus propiedades.

2.13. Estimación de los parámetros de la curva característica del ítem conocida la habilidad

2.13.1. Estimación del modelo de ojiva normal por máxima verosimilitud

Supongamos que tenemos k grupos, constituido cada uno por f_j sujetos que poseen la habilidad θ_j, donde $j = 1 \dots k$. De cada grupo tenemos que r_j responden correctamente y por lo tanto $f_j - r_j$ responden incorrectamente. Sea $R = (r_1, r_2, \dots, r_k)$ el vector de R^k de respuestas correctas y sea $p(\theta_j) = p_j = \frac{r_j}{f_j}$ la proporción de aciertos y por lo tanto $q(\theta_j) = q_j = \frac{f_j - r_j}{f_j}$ la proporción de errores. Asumimos que r_j se distribuye binomialmente para cada θ_j con parámetros f_j y P_j, donde P_j es la verdadera probabilidad de responder correctamente. Así pues $E(r_j) = f_j P_j$ y $V(r_j) = f_j P_j Q_j$. Podemos entonces calcular la esperanza de la proporción de aciertos obteniendo: $E(p_j) = E(\frac{r_j}{f_j}) = \frac{1}{f_j} E(r_j) = P_j$.

En el modelo de ojiva normal tenemos que:

$$P_j = P_j(\zeta, \lambda, \theta_j) = \int_{-(\zeta+\lambda\theta_j)}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \int_{-\infty}^{(\zeta+\lambda\theta_j)} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \tag{2.69}$$

La función de verosimilitud del vector R es:

$$L = P(R) = \prod_{j=1}^{k} \frac{f_j!}{r_j!(f_j - r_j)!} P_j^{r_j} Q_j^{f_j - r_j} \tag{2.70}$$

entonces el logaritmo de la verosimilitud es

$$\ln L = \sum_{j=1}^{k} \ln(\frac{f_j!}{r_j!(f_j - r_j)!}) + \sum_{j=1}^{k} r_j \ln(P_j) + \sum_{j=1}^{k} (f_j - r_j) \ln(Q_j) \tag{2.71}$$

A los efectos de maximizar el logaritmo de la verosimilitud, encontramos sus puntos estacionarios; resolvemos:

$$\nabla \ln L = (\frac{\partial \ln L}{\partial \zeta}, \frac{\partial \ln L}{\partial \lambda}) = (0, 0) \tag{2.72}$$

Sea $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ la función de densidad normal. Su derivada es:

$$\varphi'(x) = \frac{1}{\sqrt{2\pi}} (-2\frac{x}{2}) e^{-\frac{x^2}{2}} = -x\varphi(x)$$

Procedimientos de estimación
Estimación de la habilidad
Método de máxima verosimilitud
Estimación bayesiana
Estimación de los parámetros de los ítems
Método de mínimo chi-cuadrado
Método de máxima verosimilitud
Estimación bayesiana
Estimación simultánea de los parámetros del ítem y la habilidad
Método de máxima verosimilitud
Máxima verosimilitud conjunta
Máxima verosimilitud marginal
Modelo de Rasch máxima verosimilitud condicional
Estimación bayesiana
Otros

Aplicando derivada de la función compuesta se tiene que:

$$\frac{\partial P_j}{\partial \zeta} = \frac{\partial P_j}{\partial(\zeta + \lambda\theta_j)} \frac{\partial(\zeta + \lambda\theta_j)}{\partial \zeta} = \frac{1}{\sqrt{2\pi}} e^{-\frac{(\zeta+\lambda\theta_j)^2}{2}} = \varphi(\zeta + \lambda\theta_j) \tag{2.73}$$

$$\frac{\partial P_j}{\partial \lambda} = \frac{\partial P_j}{\partial(\zeta + \lambda\theta_j)} \frac{\partial(\zeta + \lambda\theta_j)}{\partial \lambda} = \varphi(\zeta + \lambda\theta_j)\theta_j \tag{2.74}$$

Con lo anterior ahora es fácil calcular el gradiente.

$$\frac{\partial \ln L}{\partial \zeta} = \sum_{j=1}^{k} r_j \frac{1}{P_j} \frac{\partial P_j}{\partial \zeta} + \sum_{j=1}^{k} (f_j - r_j) \frac{1}{Q_j} \frac{\partial Q_j}{\partial \zeta}$$

$$\frac{\partial \ln L}{\partial \zeta} = \sum_{j=1}^{k} r_j \frac{\varphi(\zeta + \lambda\theta_j)}{P_j} - \sum_{j=1}^{k} (f_j - r_j) \frac{\varphi(\zeta + \lambda\theta_j)}{Q_j} = \sum_{j=1}^{k} [r_j - f_j P_j] \frac{\varphi(\zeta + \lambda\theta_j)}{P_j Q_j}$$

Como $r_j = f_j p_j$ entonces:

$$\frac{\partial \ln L}{\partial \zeta} = \sum_{j=1}^{k} [f_j p_j - f_j P_j] \frac{\varphi(\zeta + \lambda\theta_j)}{P_j Q_j} = \sum_{j=1}^{k} \frac{f_j \varphi(\zeta + \lambda\theta_j)}{P_j Q_j} (p_j - P_j) \tag{2.75}$$

Por otro lado,

$$\frac{\partial \ln L}{\partial \lambda} = \sum_{j=1}^{k} r_j \frac{1}{P_j} \frac{\partial P_j}{\partial \lambda} + \sum_{j=1}^{k} (f_j - r_j) \frac{1}{Q_j} \frac{\partial Q_j}{\partial \lambda}$$

$$= \sum_{j=1}^{k} \frac{r_j \varphi(\zeta + \lambda\theta_j)\theta_j}{P_j} - \sum_{j=1}^{k} (f_j - r_j) \frac{\varphi(\zeta + \lambda\theta_j)\theta_j}{Q_j}$$

$$= \sum_{j=1}^{k} [r_j - f_j P_j]\theta_j \frac{\varphi(\zeta + \lambda\theta_j)}{P_j Q_j}$$

entonces,

$$\frac{\partial \ln L}{\partial \lambda} = \sum_{j=1}^{k} f_j [p_j - P_j]\theta_j \frac{\varphi(\zeta + \lambda\theta_j)}{P_j Q_j}. \tag{2.76}$$

Se obtuvo el sistema

$$\sum_{j=1}^{k} f_j \frac{\varphi(\zeta + \lambda\theta_j)}{P_j Q_j}(p_j - P_j) = 0$$

$$\sum_{j=1}^{k} f_j \frac{\varphi(\zeta + \lambda\theta_j)}{P_j Q_j}(p_j - P_j)\theta_j = 0$$

Lema 2.13.1. Sea z_j el desvío normal correspondiente a p_j y Z_j el correspondiente al modelo. Entonces, $z_j - Z_j \cong \frac{p_j - P_j}{\varphi(\zeta + \lambda\theta_j)}$

Demostración:
Sean $P_j = \int_{-Z_j}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$ y $p_j = \int_{-z_j}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$
luego al aplicar el teorema del valor medio obtenemos:

$$p_j - P_j = \int_{-z_j}^{Z_j} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \cong (z_j - Z_j)\varphi(\zeta + \lambda\theta_j)$$

O sea,

$$\frac{p_j - P_j}{\varphi(\zeta + \lambda\theta_j)} \cong (z_j - Z_j)$$

Sustituyendo $\frac{p_j - P_j}{\varphi(\zeta + \lambda\theta_j)}$ por $z_j - Z_j$ en las ecuaciones anteriores obtenemos:

$$\sum_{j=1}^{k} f_j w_j (z_j - Z_j) = 0$$

y

$$\sum_{j=1}^{k} f_j w_j (z_j - Z_j)\theta_j = 0$$

donde $w_j = \dfrac{\varphi(\xi + \lambda\theta_j)^2}{P_j Q_j}$

Es claro ahora que el proceso es ajustar una recta de regresión lineal con pesos. Estas ecuaciones no pueden ser resueltas directamente para obtener las estimaciones de los parámetros. El problema radica en que Z_j depende de los valores de ζ y λ.

Para resolver las ecuaciones de verosimilitud comenzaremos utilizando el procedimiento de Newton Raphson. Para esto debemos determinar la matriz hessiana. Tenemos:

$$\frac{\partial^2 \ln L}{\partial \zeta^2} = \frac{\partial}{\partial \zeta^2}\Big\{ \sum_{j=1}^{k} \frac{f_j \varphi(\zeta + \lambda\theta_j)}{P_j Q_j}[p_j - P_j]\Big\}$$

$$= \sum_{j=1}^{k} f_j \frac{\Big\{ P_j Q_j \frac{\partial[\varphi(\zeta+\lambda\theta_j)(p_j - P_j)]}{\partial \zeta} - [\varphi(\zeta + \lambda\theta_j)(p_j - P_j)\frac{\partial(P_j Q_j)}{\partial \zeta}]\Big\}}{(P_j Q_j)^2}$$

$$= \sum_{j=1}^{k} f_j \frac{\Big\{ P_j Q_j [\varphi'(\zeta+\lambda\theta_j)p_j - (\varphi(\zeta+\lambda\theta_j)^2 + P_j \varphi'(\zeta+\lambda\theta_j))] - [\varphi(\zeta+\lambda\theta_j)(p_j - P_j)(-\varphi(\zeta+\lambda\theta_j)P_j + \varphi(\zeta+\lambda\theta_j)Q_j)]\Big\}}{(P_j Q_j)^2}$$

$$= \sum_{j=1}^{k} f_j \varphi(\zeta+\lambda\theta_j) \frac{\Big\{ P_j Q_j [-(\zeta + \lambda\theta_j)(p_j - P_j) - \varphi(\zeta + \lambda\theta_j)] - [\varphi(\zeta + \lambda\theta_j)(p_j - P_j)(Q_j - P_j)]\Big\}}{(P_j Q_j)^2}$$

$$= \sum_{j=1}^{k} \frac{f_j \varphi(\zeta + \lambda\theta_j)(p_j - P_j)[-(\zeta + \lambda\theta_j)P_j Q_j - \varphi(\zeta + \lambda\theta_j)(Q_j - P_j)]}{(P_j Q_j)^2} - \sum_{j=1}^{k} \frac{f_j \varphi(\zeta + \lambda\theta_j)^2 P_j Q_j}{(P_j Q_j)^2}$$

luego,

$$\frac{\partial^2 \ln L}{\partial \zeta^2} = \sum_{j=1}^{k} \frac{f_j \varphi(\zeta + \lambda\theta_j)(p_j - P_j)}{P_j Q_j}[-(\zeta+\lambda\theta_j) - \frac{\varphi(\zeta + \lambda\theta_j)}{P_j} + \frac{\varphi(\zeta + \lambda\theta_j)}{Q_j}] - \sum_{j=1}^{k} \frac{f_j \varphi(\zeta + \lambda\theta_j)^2}{P_j Q_j}$$

Dado que $\frac{\partial P_j}{\partial \lambda} = \frac{\partial P_j}{\partial \zeta}\theta_j$, podemos obtener fácilmente $\frac{\partial^2 \ln L}{\partial \lambda^2}$ a partir de $\frac{\partial^2 \ln L}{\partial \zeta^2}$.

$$\frac{\partial^2 \ln L}{\partial \lambda^2} = \sum_{j=1}^{k} \frac{f_j \varphi(\zeta + \lambda\theta_j)(p_j - P_j)}{P_j Q_j}[-(\zeta+\lambda\theta_j) - \frac{\varphi(\zeta + \lambda\theta_j)}{P_j} + \frac{\varphi(\zeta + \lambda\theta_j)}{Q_j}]\theta_j^2 - \sum_{j=1}^{k} \frac{f_j \varphi(\zeta + \lambda\theta_j)^2 \theta_j^2}{P_j Q_j}$$

Análogamente obtenemos,

$$\frac{\partial^2 \ln L}{\partial \lambda \partial \zeta} = \frac{\partial^2 \ln L}{\partial \zeta \partial \lambda} =$$

$$= \sum_{j=1}^{k} \frac{f_j \varphi(\zeta + \lambda\theta_j)(p_j - P_j)}{P_j Q_j}[-(\zeta+\lambda\theta_j) - \frac{\varphi(\zeta + \lambda\theta_j)}{P_j} + \frac{\varphi(\zeta + \lambda\theta_j)}{Q_j}]\theta_j - \sum_{j=1}^{k} \frac{f_j \varphi(\zeta + \lambda\theta_j)^2 \theta_j}{P_j Q_j}$$

Para simplificar nuestro problema utilizaremos el método de Fisher que consiste en reemplazar la matriz hessiana por su esperanza. Así tenemos que, si tomamos esperanza como $E(p_j) = P_j$, el primer término de cada ecuación se anula y nos queda:

$$E(\frac{\partial^2 \ln L}{\partial \zeta^2}) = -\sum_{j=1}^{k} \frac{f_j \varphi(\zeta + \lambda\theta_j)^2}{P_j Q_j} = -\sum_{j=1}^{k} f_j w_j$$

$$E(\frac{\partial^2 \ln L}{\partial \lambda^2}) = -\sum_{j=1}^{k} \frac{f_j \varphi(\zeta + \lambda\theta_j)^2 \theta_j^2}{P_j Q_j} = -\sum_{j=1}^{k} f_j w_j \theta_j^2$$

$$E(\frac{\partial^2 \ln L}{\partial \lambda \partial \zeta}) = E(\frac{\partial^2 \ln L}{\partial \zeta \partial \lambda}) = -\sum_{j=1}^{k} \frac{f_j \varphi(\zeta + \lambda\theta_j)^2 \theta_j}{P_j Q_j} = -\sum_{j=1}^{k} f_j w_j \theta_j$$

Así tenemos que resolver el sistema,

$$\sum_{j=1}^{k} f_j w_j \frac{(p_j - P_j)}{\varphi(\zeta + \lambda\theta_j)} = \Delta\widehat{\zeta}_t \sum_{j=1}^{k} f_j w_j + \Delta\widehat{\lambda}_t \sum_{j=1}^{k} f_j w_j \theta_j$$

$$\sum_{j=1}^{k} f_j w_j \frac{(p_j - P_j)}{\varphi(\zeta + \lambda\theta_j)} \theta_j = \Delta\widehat{\zeta}_t \sum_{j=1}^{k} f_j w_j \theta_j + \Delta\widehat{\lambda}_t \sum_{j=1}^{k} f_j w_j \theta_j^2$$

o sea,

$$\begin{pmatrix} \widehat{\zeta} \\ \widehat{\lambda} \end{pmatrix}_{t+1} = \begin{pmatrix} \widehat{\zeta} \\ \widehat{\lambda} \end{pmatrix}_{t} + \begin{pmatrix} \sum_{j=1}^{k} f_j w_j & \sum_{j=1}^{k} f_j w_j \theta_j \\ \sum_{j=1}^{k} f_j w_j \theta_j & \sum_{j=1}^{k} f_j w_j \theta_j^2 \end{pmatrix}_{t}^{-1} \begin{pmatrix} \sum_{j=1}^{k} f_j w_j \frac{(p_j - P_j)}{\varphi(\zeta + \lambda\theta_j)} \\ \sum_{j=1}^{k} f_j w_j \frac{(p_j - P_j)}{\varphi(\zeta + \lambda\theta_j)} \theta_j \end{pmatrix}_{t} \tag{2.77}$$

El método de Fisher, en general, converge más rápidamente que Newton Raphson.

Kale (1962) mostró que los valores iniciales deben estar en un entorno de los parámetros y deben estar basados en estimadores consistentes para que el proceso iterativo converja.

Hay varios esquemas para elegir los valores iniciales, uno es el que utiliza las ecuaciones $\rho_{\theta U_i} = \frac{\alpha_i}{\sqrt{1+\alpha_i^2}}$ donde $\rho_{\theta U_i}$ es el coeficiente de correlación biserial entre θ y la variable de respuesta U y la dificultad del ítem:

$$P_i = \frac{1}{\sqrt{2\pi}} \int_{\delta_i}^{\infty} e^{-\frac{u^2}{2}} du$$

donde $\delta_i = \beta_i \rho_{\theta U_i} = \beta_i \frac{\alpha_i}{\sqrt{1+\alpha_i^2}}$ y de éstas se pueden obtener α_i y β_i

Bajo el modelo de dos parámetros los valores de tendencia e interceptor 1 y 0 muchas veces funcionan adecuadamente.

El procedimiento puede resumirse así:

1- Se comienza con estimaciones primarias de $(\widehat{\zeta}_1, \widehat{\lambda}_1)$.

2- Si en el tiempo t tenemos $(\widehat{\zeta}_t, \widehat{\lambda}_t)$ a partir de estos calculamos $\widehat{\zeta}_t + \widehat{\lambda}_t \theta_j$ para todos los niveles de habilidad.

3- Se calcula $\widehat{P_j} = \int_{-(\widehat{\zeta+\lambda\theta_j})}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$ con $\widehat{Q_j} = 1 - \widehat{P_j}$, $\varphi(\widehat{\zeta + \lambda\theta_j}) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(\widehat{\zeta+\lambda\theta_j})^2}{2}}$ y $w_j = \frac{\varphi(\widehat{\zeta+\lambda\theta_j})^2}{\widehat{P_j}\widehat{Q_j}}$.

4- Se calcula $(\widehat{\zeta}_{t+1}, \widehat{\lambda}_{t+1})$ mediante

$$\begin{pmatrix} \widehat{\zeta} \\ \widehat{\lambda} \end{pmatrix}_{t+1} = \begin{pmatrix} \widehat{\zeta} \\ \widehat{\lambda} \end{pmatrix}_t + \begin{pmatrix} \sum_{j=1}^k f_j w_j & \sum_{j=1}^k f_j w_j \theta_j \\ \sum_{j=1}^k f_j w_j \theta_j & \sum_{j=1}^k f_j w_j \theta_j^2 \end{pmatrix}_t^{-1} \begin{pmatrix} \sum_{j=1}^k f_j w_j \frac{(p_j - P_j)}{\varphi(\zeta+\lambda\theta_j)} \\ \sum_{j=1}^k f_j w_j \frac{(p_j - P_j)}{\varphi(\zeta+\lambda\theta_j)} \theta_j \end{pmatrix}_t \tag{2.78}$$

5- Este proceso se continúa hasta que $|\Delta\widehat{\zeta}_t|$ y $|\Delta\widehat{\lambda}_t|$ son arbitrariamente chicos.

6- A partir de las estimaciones de $\widehat{\zeta}$ y $\widehat{\lambda}$ obtenemos $\widehat{\alpha} = \widehat{\lambda}$ y $\widehat{\beta} = -\frac{\widehat{\zeta}}{\widehat{\lambda}}$

Berkson (1955) halló las varianzas asintóticas de los estimadores:

$$s^2_{\widehat{\lambda}} = \frac{1}{\sum_{j=1}^k f_j w_j (\overline{\theta} - \theta_j)^2} = s^2_{\widehat{\alpha}}$$

donde

$$\overline{\theta} = \frac{\sum_{j=1}^k f_j w_j \theta_j}{\sum_{j=1}^k f_j w_j}$$

$$s^2_{\widehat{\zeta}} = \frac{1}{\sum_{j=1}^k f_j w_j} + \overline{\theta}^2 s^2_{\widehat{\lambda}}$$

y

$$s^2_{\widehat{\beta}} = \frac{1}{\widehat{\alpha}^2}\Big[\frac{1}{\sum_{j=1}^k f_j w_j} + s^2_{\widehat{\alpha}}(\widehat{\beta} - \overline{\theta})^2\Big]$$

Un test de bondad de ajuste obtenido por Garwood (1941) es

$$\chi^2 = \sum_{j=1}^k f_j w_j \frac{(p_j - P_j)^2}{\varphi(\zeta + \lambda\theta_j)^2}$$

que tiene distribución asintótica chi cuadrado con k-2 grados de libertad.

2.13.2. Estimación del modelo logístico de dos parámetros por máxima verosimilitud

En este modelo $P_j = P(a, b, \theta_j) = \frac{1}{1+e^{-a(\theta_j - b)}}$

Si ponemos $\zeta = -ab$ y $\lambda = a$ entonces

$$P_j = P(\zeta, \lambda, \theta_j) = \frac{1}{1 + e^{-(\zeta + \lambda\theta_j)}}$$

y

$$Q_j = 1 - P_j$$

Se prueba fácilmente que:

$$\frac{\partial P_j}{\partial \zeta} = P_j Q_j,$$

$$\frac{\partial P_j}{\partial \lambda} = P_j Q_j \theta_j$$

$$\frac{\partial Q_j}{\partial \zeta} = -P_j Q_j$$

$$\frac{\partial Q_j}{\partial \lambda} = -P_j Q_j \theta_j$$

En el mismo contexto que el modelo de ojiva normal la verosimilitud es

$$L = \prod_{j=1}^{k} \frac{f_j!}{r_j!(f_j - r_j)!} P_j^{r_j} Q_j^{f_j - r_j}$$

Tomando logaritmo:

$$\ln L = \sum_{j=1}^{k} \ln\left(\frac{f_j!}{r_j!(f_j - r_j)!}\right) + \sum_{j=1}^{k} r_j \ln(P_j) + \sum_{j=1}^{k} (f_j - r_j) \ln(Q_j)$$

A los efectos de maximizar el logaritmo de la verosimilitud, encontramos sus puntos estacionarios, resolvemos:

$$\nabla L = \left(\frac{\partial L}{\partial \zeta}, \frac{\partial L}{\partial \lambda}\right) = (0, 0)$$

Aplicaremos, como en el modelo de ojiva normal, el método de Newton Raphson. Haciendo algunos cálculos obtendremos:

$$\frac{\partial L}{\partial \zeta} = \Sigma_{j=1}^{k} f_j (p_j - P_j) \tag{2.79}$$

$$\frac{\partial L}{\partial \lambda} = \Sigma_{j=1}^{k} f_j (p_j - P_j)\theta_j \tag{2.80}$$

$$\frac{\partial^2 L}{\partial \zeta^2} = -\Sigma_{j=1}^{k} f_j P_j Q_j \tag{2.81}$$

$$\frac{\partial^2 L}{\partial \lambda^2} = -\Sigma_{j=1}^{k} f_j P_j Q_j \theta_j^2 \tag{2.82}$$

y

$$\frac{\partial^2 L}{\partial \lambda \partial \zeta} = \frac{\partial^2 L}{\partial \lambda \partial \zeta} = -\Sigma_{j=1}^{k} f_j P_j Q_j \theta_j \tag{2.83}$$

Finalmente, obtenemos:

$$\begin{pmatrix} \widehat{\zeta} \\ \widehat{\lambda} \end{pmatrix}_{t+1} = \begin{pmatrix} \widehat{\zeta} \\ \widehat{\lambda} \end{pmatrix}_{t} + \begin{pmatrix} \sum_{j=1}^{k} f_j w_j & \sum_{j=1}^{k} f_j w_j \theta_j \\ \sum_{j=1}^{k} f_j w_j \theta_j & \sum_{j=1}^{k} f_j w_j \theta_j^2 \end{pmatrix}_{t}^{-1} \begin{pmatrix} \sum_{j=1}^{k} f_j w_j \frac{(p_j - P_j}{P_j Q_j} \\ \sum_{j=1}^{k} f_j w_j \frac{(p_j - P_j)}{P_j Q_j} \theta_j \end{pmatrix}_{t} \tag{2.84}$$

Maxwell (1959) obtuvo las varianzas asintóticas de los estimadores que coinciden en su forma con las del modelo de ojiva normal.

$$s_{\widehat{\lambda}}^2 = \frac{1}{\sum_{j=1}^{k} f_j w_j (\overline{\theta} - \theta_j)^2} = s_{\widehat{\alpha}}^2$$

donde

$$\overline{\theta} = \frac{\sum_{j=1}^{k} f_j w_j \theta_j}{\sum_{j=1}^{k} f_j w_j}$$

$$s_{\widehat{\zeta}}^2 = \frac{1}{\sum_{j=1}^{k} f_j w_j} + \overline{\theta}^2 s_{\widehat{\lambda}}^2$$

$$s^2_{\widehat{\beta}} = \frac{1}{\widehat{\alpha}^2}\Big[\frac{1}{\sum_{j=1}^{k} f_j w_j} + s^2_{\widehat{\alpha}}(\widehat{\beta} - \overline{\theta})^2\Big]$$

Un test de bondad de ajuste también coincide con el del modelo de ojiva normal, obtenido por Garwood:

$$\chi^2 = \sum_{j=1}^{k} f_j w_j \frac{(p_j - P_j)^2}{P_j Q_j}$$

que tiene distribución asintótica chi cuadrado con k-2 grados de libertad.

A continuación se presenta una función para la estimación del modelo logístico de dos parámetros.

```
itempar.mv.2P< −function(th,r,f,maxit,imptodo=1){

# programa para estimar por máxima verosimilitud
# los parámetros del modelo logístico de dos parámetros
# conocida la habilidad
# th vector de habilidades theta
# r vector de repuestas correctas
# f vector de frecuencias
# la función retorna el valor de Zeta, lambda, dificultad, chi2 y los grados de libertad

esti< −c(0,0,0,0,0)

# valores iniciales de los parámetros Z− >zeta
# LAM− > lambda, GL− > grados de libertad

Z< −0 LAM< −1 GL< −length(th)-2
for(n in 1:maxit){
if(imptodo == 0){ ti< −paste(«iteracion»,n)
print(ti)}
AUX1< −0
AUX2< −0
CHI2< −0
AUX3< −0
AUX4< −0
AUX5< −0
```

```
    #loop para theta
    for(k in 1:length(th)){
    if(f[k]==0){next}
    PROP< −r[k]/f[k]
    P< −1/(1+exp(-Z-LAM*th[k]))
    if(P*(1-P)¡.0000009)next
    if(imptodo == 0)
    print(paste("Theta",th[k],"PROPORCION=",PROP,"P  gorro=",P,"PESO=",P*(1-
P))) }
    AUX1< −AUX1+f[k]*P*(1-P)
    AUX2< −AUX2+f[k]*(PROP-P)
    AUX3< −AUX3+f[k]*P*(1-P)*th[k]
    AUX4< −AUX4+f[k]*th[k]*(PROP-P)
    AUX5< −AUX5+f[k]*P*(1-P)*th[k]2̂
    CHI2< −CHI2+f[k]*(PROP-P)2̂/(P*(1-P))
    if(imptodo == 0){
    print(paste(.ªUX1=",AUX1,.ªUX2=",AUX2,.ªUX3=",AUX3,.ªUX4=",
    AUX4,.ªUX5=",AUX5,ÇHI2=",CHI2))}
    }
    if(AUX1<=0){print(.ERROR fuera de limites iteracion")
    return(0)}
    DET< −AUX1*AUX5-AUX3*AUX3
    if(imptodo == 0){
    print(DM) }
    if(DET<=.000099){return(esti)}
    DELTAZ< −(AUX2*AUX5-AUX4*AUX3)/DET
    DELTALAM< −(AUX1*AUX4-AUX3*AUX2)/DET
    Z< −Z+DELTAZ
    LAM< −LAM+DELTALAM
    B< −-Z/LAM
    esti< −c(Z,LAM,B,CHI2,GL)
    if(imptodo == 0){
    print(paste(ïteracion",n, "zeta=",Z,çambio z=",DELTAZ,"lamda=",LAM,çambio l=",DELTALAM
    print(paste("dificultad=",B,ÇHI2=",CHI2,"GL=",GL)) }
    if(abs(Z)>30 — abs(LAM)>20){print(.ERROR fuera de limites iteracion")
    return(0)}
    if(abs(DELTAZ)<=.05 & abs(DELTALAM)<=.05){return(esti)}
    }
    }
```

2.13.3. Estimación del modelo logístico de tres parámetros por máxima verosimilitud

El modelo logístico de tres parámetros toma la forma:

$$P_j = c + (1-c)\frac{1}{1+e^{-a(\theta_j - b)}}$$

Igual que en los casos anteriores la verosimilitud es:

$$L = \prod_{j=1}^{k} \frac{f_j!}{r_j!(f_j - r_j)!} P_j^{r_j} Q_j^{f_j - r_j} \tag{2.85}$$

Tomando logaritmo:

$$\ln L = \sum_{j=1}^{k} \ln(\frac{f_j!}{r_j!(f_j - r_j)!}) + \sum_{j=1}^{k} r_j \ln(P_j) + \sum_{j=1}^{k} (f_j - r_j)\ln(Q_j) \tag{2.86}$$

Nuevamente hallamos los puntos estacionarios.

Resolvemos $\nabla L = (\frac{\partial L}{\partial a}, \frac{\partial L}{\partial b}, \frac{\partial L}{\partial c}) = (0,0,0)$.
Aplicaremos el método de Newton Raphson - Fisher.

Se tiene: $\frac{\partial P_j}{\partial a} = \frac{(\theta_j - b)Q_j(P_j - c)}{1-c}$ $\qquad$ $\frac{\partial P_j}{\partial b} = \frac{-aQ_j(P_j - c)}{1-c}$

Operando se obtiene:

$$\frac{\partial \ln L}{\partial a} = \sum_{j=1}^{k} f_j(p_j - P_j)(\theta_j - b)\frac{1}{(1+e^{-a(\theta_j - b)})P_j} \tag{2.87}$$

$$\frac{\partial \ln L}{\partial b} = -a\sum_{j=1}^{k} f_j(p_j - P_j)\frac{1}{(1+e^{-a(\theta_j - b)})P_j} \tag{2.88}$$

$$\frac{\partial \ln L}{\partial c} = \sum_{j=1}^{k} f_j(p_j - P_j)\frac{1}{P_j - c}\frac{1}{(1+e^{-a(\theta_j - b)})P_j} \tag{2.89}$$

$$E(\frac{\partial^2 \ln L}{\partial a^2}) = -\sum_{j=1}^{k} f_j(\theta_j - b)^2 P_j Q_j [\frac{1}{(1+e^{-a(\theta_j - b)})P_j}]^2 \tag{2.90}$$

$$E(\frac{\partial^2 \ln L}{\partial b^2}) = -a^2\sum_{j=1}^{k} f_j P_j Q_j [\frac{1}{(1+e^{-a(\theta_j - b)})P_j}]^2 \tag{2.91}$$

$$E(\frac{\partial^2 \ln L}{\partial c^2}) = \sum_{j=1}^{k} f_j \frac{Q_j}{1-c} \frac{1}{P_j - c} \frac{1}{(1+e^{-a(\theta_j - b)})P_j} \tag{2.92}$$

$$E(\frac{\partial^2 \ln L}{\partial a \partial b}) \cong \sum_{j=1}^{k} a f_j (\theta_j - b) P_j Q_j [\frac{1}{(1+e^{-a(\theta_j - b)})P_j}]^2 \tag{2.93}$$

$$E(\frac{\partial^2 \ln L}{\partial a \partial c}) \cong -\sum_{j=1}^{k} f_j (\theta_j - b) \frac{Q_j}{1-c} [\frac{1}{(1+e^{-a(\theta_j - b)})P_j}] \tag{2.94}$$

$$E(\frac{\partial^2 \ln L}{\partial b \partial c}) = -\sum_{j=1}^{k} a f_j \frac{Q_j}{1-c} [\frac{1}{(1+e^{-a(\theta_j - b)})P_j}] \tag{2.95}$$

El proceso iterativo se regula por las ecuaciones:

$$\begin{pmatrix} \widehat{a} \\ \widehat{b} \\ \widehat{c} \end{pmatrix}_{t+1} = \begin{pmatrix} \widehat{a} \\ \widehat{b} \\ \widehat{c} \end{pmatrix}_{t} - \begin{pmatrix} E(\frac{\partial^2 \ln L}{\partial a^2}) & E(\frac{\partial^2 \ln L}{\partial a \partial b}) & E(\frac{\partial^2 \ln L}{\partial a \partial c}) \\ E(\frac{\partial^2 \ln L}{\partial a \partial b}) & E(\frac{\partial^2 \ln L}{\partial b^2}) & E(\frac{\partial^2 \ln L}{\partial b \partial c}) \\ E(\frac{\partial^2 \ln L}{\partial a \partial c}) & E(\frac{\partial^2 \ln L}{\partial b \partial c}) & E(\frac{\partial^2 \ln L}{\partial c^2}) \end{pmatrix}_{t}^{-1} \begin{pmatrix} \frac{\partial \ln L}{\partial a} \\ \frac{\partial \ln L}{\partial b} \\ \frac{\partial \ln L}{\partial c} \end{pmatrix}_{t}$$

Damos a continuación un programa de R para la estimación de este modelo.

```
itempar.mv.3P< −function(th,rk,fk,maxit,imptodo=0)

# programa para estimar por máxima verosimilitud
# los parámetros del modelo logistico de tres parámetros
A< −1
B< −0
C< −0.5
PAR< −c(A,B,C)
for(n in 1:maxit){
L1< −0
L2< −0
L3< −0
L11< −0
L12< −0
L13< −0
L23< −0
L22< −0
L33< −0
```

```
for(k in 1:length(th))
PA< −1/(1+exp(-A*(th[k]-B)))
PK< −C+(1-C)*PA
PI< −rk[k]/fk[k]
WK< −PA*(1-PA)/(PK*(1-PK))
if(PK==1){next}
if(WK<.0000009){next}
L1< −L1+fk[k]*(PI-PK)*(th[k]-B)*PA/PK
L2< −L2-A*fk[k]*(PI-PK)*(1-C)*WK
L3< −L3+fk[k]*(PI-PK)*(1/(PK-C))*(PA/PK)
L11< −L11-fk[k]*((th[k]-B)2̂)*PK*(1-PK)*((PA/PK)2̂)
L22< −L22-A2̂*fk[k]*PK*(1-PK)*((PA/PK)2̂)
L33< −L33+fk[k]*((1-PK)/((1-C)*(PK-C)))*(PA/PK)
L12< −L12+A*fk[k]*(th[k]-B)*PK*(1-PK)*((PA/PK)2̂)
L13< −L13-fk[k]*(th[k]-B)*(1-PK)*(1/(1-C))*(PA/PK)
L23< −L23+A*fk[k]*(1-PK)*(1/(1-C))*(PA/PK)
}
MT< −rbind(c(L11,L12,L13),c(L12,L22,L23),c(L13,L23,L33))
VL< −c(L1,L2,L3)
IMT< −solve(MT)
CAMBIO< −IMT %* %VL
PAR< −PAR+CAMBIO
A< −PAR[1]
B< −PAR[2]
C< −PAR[3]
print(A)
print(B)
print(C)
if(imptodo == 0){
print(IMT)
print(VL)
print(paste(ÇAMBIO=",CAMBIO,"PAR=",PAR)) }
if(abs(PAR[1])>20 — abs(PAR[2])>30 —abs(PAR[3])>1)
{print(.ERROR fuera de limites iteracion")
return(0)}
if(abs(CAMBIO[1])<=.05 & abs(CAMBIO[2])<=.05 & abs(CAMBIO[3])<=.05)
{return(PAR)}
}
}
```

2.13.4. Estimación por mínimo χ^2

Este procedimiento fue propuesto por Berkson en 1955. Recordemos el estadístico χ^2 de Pearson para tablas de contingencia.

$$\chi^2 = \sum_{j=1}^{k} \frac{(OBS_j - ESP_j)^2}{ESP_j} \tag{2.96}$$

donde OBS indica los valores observados y ESP los valores esperados. Observamos la tabla de contingencia kx2 considerando los k niveles de habilidad y las dos posibilidades de respuestas (correcto o incorrecto). Para cada nivel θ_j tenemos que la celda correspondiente a correcto contendrá el valor r_j y la celda correspondiente a incorrecto el valor $f_j - r_j$.

El valor esperado de respuestas correctas e incorrectas para el nivel j es $f_j P_j$ y $f_j Q_j$ respectivamente. Luego

$$\chi^2 = \sum_{j=1}^{k} \frac{(r_j - f_j P_j)^2}{f_j P_j} + \frac{((f_j - r_j) - f_j Q_j)^2}{f_j Q_j} \tag{2.97}$$

Como $r_j = f_j p_j$ y si definimos $q_j = \frac{f_j - r_j}{f_j}$ entonces nos queda:

$$\chi^2 = \sum_{j=1}^{k} \frac{(f_j p_j - f_j P_j)}{f_j P_j} + \frac{(f_j q_j - f_j Q_j)}{f_j Q_j} = \sum_{j=1}^{k} \frac{f_j}{P_j Q_j}(p_j - P_j)^2 \tag{2.98}$$

Si queremos optimizar el ajuste lo que hacemos es minimizar el χ^2 considerado como función de los parámetros.

Derivando respecto a ζ y λ e igualando a 0 obtenemos:

$$\sum_{j=1}^{k} f_j \frac{P_j q_j + Q_j p_j}{P_j Q_j}(p_j - P_j) = 0$$

$$\sum_{j=1}^{k} f_j \frac{P_j q_j + Q_j p_j}{P_j Q_j}(p_j - P_j)\theta_j = 0$$

P y Q son funciones de los parámetros y no son lineales en ellos, así que para resolver estas ecuaciones se utiliza nuevamente, como en los casos anteriores, un procedimiento iterativo. No abundamos más en este procedimiento.

2.14. Estimación de la habilidad cuando se conocen los parámetros de los ítems

2.14.1. Estimación de la habilidad por máxima verosimilitud

Supongamos que un individuo responde a un test de n ítems y sus respuestas son puntuadas 0 si falla o 1 si acierta. Además, supongamos conocidos los parámetros del ítem. Notemos por u_{ij} a la variable dicotómica que representa la respuesta del sujeto j al ítem i. Sea para el individuo j el vector de respuestas condicionado a su nivel de rasgo $U_j = (u_{1j}, u_{2j}, ..., u_{nj}|\theta_j)$. Bajo el principio de independencia local las u_{ij} son independientes. A partir de esto la verosimilitud queda:

$$L = \prod_{i=1}^{n} P_{ij}^{u_{ij}} Q_{ij}^{1-u_{ij}} \tag{2.99}$$

donde $P_{ij} = P_i(\theta_j)$.

Tomando logaritmo:

$$\ln L = \ln P(U_j|\theta_j) = \sum_{i=1}^{n} [u_{ij} \ln P_{ij} + (1 - u_{ij}) \ln Q_{ij}] \tag{2.100}$$

Para estimar θ_j maximizamos L. Derivamos respecto a θ_j e igualamos a 0 obteniendo:

$$\frac{\partial L}{\partial \theta_j} = \sum_{i=1}^{n} [u_{ij} \frac{1}{P_{ij}} \frac{\partial P_{ij}}{\partial \theta_j} + (1 - u_{ij}) \frac{1}{Q_{ij}} \frac{\partial Q_{ij}}{\partial \theta_j}] = 0 \tag{2.101}$$

Aplicamos nuevamente el procedimiento de Newton Raphson, que en este caso es unidimensional y como vimos en la primera parte:

$$[\widehat{\theta}_j]_{t+1} = [\widehat{\theta}_j]_t - [\frac{\partial^2 L}{\partial \theta_j^2}]_t^{-1} [\frac{\partial L}{\partial \theta_j}]_t \tag{2.102}$$

Veremos los modelos de ojiva normal y de dos y tres parámetros

2.14.2. Modelo de ojiva normal

Es inmediato ver que:

$$\frac{\partial P_{ij}}{\partial \theta_j} = \lambda_i \varphi(\zeta_i + \lambda_i \theta_j)$$

y

$$\frac{\partial Q_{ij}}{\partial \theta_j} = -\lambda_j \varphi(\zeta_i + \lambda_i \theta_j)$$

Así queda entonces:

$$\frac{\partial \ln L}{\partial \theta_j} = \sum_{i=1}^{n} [u_{ij} \frac{1}{P_{ij}} \lambda_j \varphi(\zeta_i + \lambda_i \theta_j) + (1 - u_{ij}) \frac{1}{Q_{ij}} (-\lambda_j \varphi(\zeta_i + \lambda_i \theta_j))]$$

$$\frac{\partial \ln L}{\partial \theta_j} = \sum_{i=1}^{n} \lambda_j \varphi(\zeta_i + \lambda_i \theta_j) \frac{u_{ij} - P_{ij}}{P_{ij} Q_{ij}} \tag{2.103}$$

Ahora:

$$\frac{\partial^2 \ln L}{\partial \theta_j^2} = \frac{\partial \sum_{i=1}^{n} \lambda_j \varphi(\zeta_i + \lambda_i \theta_j) \frac{u_{ij} - P_{ij}}{P_{ij} Q_{ij}}}{\partial \theta_j} =$$

$$\sum_{i=1}^{n} \frac{\lambda_j^2 \varphi(\zeta_i + \lambda_i \theta_j)}{P_{ij} Q_{ij}} [(u_{ij} - P_{ij})(-(\zeta_i + \lambda_i \theta_j) -$$

$$\frac{\varphi(\zeta_i + \lambda_i \theta_j)}{P_{ij}} + \frac{\varphi(\zeta_i + \lambda_i \theta_j)}{Q_{ij}}) - \varphi(\zeta_i + \lambda_i \theta_j)] =$$

$$\sum_{i=1}^{n} \frac{\lambda_j^2 \varphi(\zeta_i + \lambda_i \theta_j)}{P_{ij} Q_{ij}} [(u_{ij} - P_{ij})(-(\zeta_i + \lambda_i \theta_j) - \frac{\varphi(\zeta_i + \lambda_i \theta_j)}{P_{ij}} + \frac{\varphi(\zeta_i + \lambda_i \theta_j)}{Q_{ij}})] - \sum_{i=1}^{n} \frac{\lambda_j^2 \varphi(\zeta_i + \lambda_i \theta_j)^2}{P_{ij} Q_{ij}}$$

Tomando esperanzas:

$$E(\frac{\partial^2 \ln L}{\partial \theta_j^2}) = -\sum_{i=1}^{n} \frac{\lambda_j^2 \varphi(\zeta_i + \lambda_i \theta_j)^2}{P_{ij} Q_{ij}} \tag{2.104}$$

Finalmente,

$$[\widehat{\theta}_j]_{t+1} = [\widehat{\theta}_j]_t + [\sum_{i=1}^{n} \frac{\lambda_j^2 \varphi(\zeta_i + \lambda_i \theta_j)^2}{P_{ij} Q_{ij}}]_t^{-1} [\sum_{i=1}^{n} \lambda_j \varphi(\zeta_i + \lambda_i \theta_j) \frac{u_{ij} - P_{ij}}{P_{ij} Q_{ij}}]_t \tag{2.105}$$

Como ya sabemos, los estimadores máximo verosímiles bajo hipótesis generales alcanzan la cota de Cramer-Rao y la varianza asintótica es:

$$s^2_{\widehat{\theta}_j} = \frac{1}{-E(\frac{\partial^2 L}{\partial \theta_j^2})} = \frac{1}{\sum_{i=1}^n \lambda_j^2 w_{ij}} = \frac{1}{\sum_{i=1}^n \alpha_j^2 w_{ij}}$$

2.14.3. Modelo logístico de dos parámetros

Tenemos:

$$\frac{\partial P_{ij}}{\partial \theta_j} = \lambda_j P_{ij} Q_{ij}$$

y

$$\frac{\partial Q_{ij}}{\partial \theta_j} = -\lambda_j P_{ij} Q_{ij}$$

luego

$$\frac{\partial \ln L}{\partial \theta_j} = \sum_{i=1}^n [u_{ij} \frac{1}{P_{ij}} \lambda_j P_{ij} Q_{ij} + (1 - u_{ij}) \frac{1}{Q_{ij}} (-\lambda_j P_{ij} Q_{ij})] = \sum_{i=1}^n \lambda_j (u_{ij} - P_{ij}) \quad (2.106)$$

Ahora:

$$\frac{\partial^2 L}{\partial \theta_j^2} = \frac{\partial \sum_{i=1}^n \lambda_j (u_{ij} - P_{ij})}{\partial \theta_j} = -\sum_{i=1}^n \lambda_j^2 P_{ij} Q_{ij} \quad (2.107)$$

de donde

$$[\widehat{\theta}_j]_{t+1} = [\widehat{\theta}_j]_t + [\sum_{i=1}^n \lambda_j^2 P_{ij} Q_{ij}]_t^{-1} [\sum_{i=1}^n \lambda_j (u_{ij} - P_{ij})]_t$$

La varianza asintótica es:

$$s^2_{\widehat{\theta}_j} = \frac{1}{-E(\frac{\partial^2 L}{\partial \theta_j^2})} = \frac{1}{\sum_{i=1}^n \lambda_j^2 P_{ij} Q_{ij}} = \frac{1}{\sum_{i=1}^n \alpha_j^2 P_{ij} Q_{ij}}$$

2.14.4. Modelo logístico de tres parámetros

En este caso:

$$\frac{\partial P_{ij}}{\partial \theta_j} = \lambda_j \frac{P_{ij} - c_i}{1 - c_i} Q_{ij}$$

Operando:

$$\frac{\partial \ln L}{\partial \theta_j} = \sum_{i=1}^{n} [u_{ij} \frac{1}{P_{ij}} \lambda_i Q_{ij} \frac{P_{ij} - c_i}{1 - c_i} + (1 - u_{ij}) \frac{1}{Q_{ij}} (-\lambda_i Q_{ij} \frac{P_{ij} - c_i}{1 - c_i})] =$$

$$\sum_{i=1}^{n} \lambda_i \frac{(u_{ij} - P_{ij})(P_{ij} - c_i)}{(1 - c_i) P_{ij}}$$

y

$$\frac{\partial^2 \ln L}{\partial \theta_j^2} = \sum_{i=1}^{n} \lambda_i^2 [\frac{Q_{ij}(P_{ij} - c_i)}{(1 - c_i)^2} \frac{u_{ij} c_i - P_{ij}^2}{P_{ij}^2}]$$

luego,

$$E(\frac{\partial^2 \ln L}{\partial \theta_j^2}) = \sum_{i=1}^{n} -\lambda_i^2 Q_{ij} [\frac{(P_{ij} - c_i)^2}{(1 - c_i)^2} \frac{1}{P_{ij}}]$$

el proceso iterativo queda:

$$[\widehat{\theta}_j]_{t+1} = [\widehat{\theta}_j]_t + [\sum_{i=1}^{n} -\lambda_i^2 Q_{ij} [\frac{(P_{ij} - c_i)^2}{(1 - c_i)^2} \frac{1}{P_{ij}}]]_t^{-1} [\sum_{i=1}^{n} \lambda_i \frac{(u_{ij} - P_{ij})(P_{ij} - c_i)}{(1 - c_i) P_{ij}}]_t$$

y la varianza asintótica:

$$s^2_{\widehat{\theta}_j} = \frac{1}{-E(\frac{\partial^2 \ln L}{\partial \theta_j^2})} = \frac{1}{\sum_{i=1}^{n} \lambda_i^2 Q_{ij} [\frac{(P_{ij} - c_i)^2}{(1 - c_i)^2} \frac{1}{P_{ij}}]} = \frac{1}{\sum_{i=1}^{n} \alpha_i^2 Q_{ij} [\frac{(P_{ij} - c_i)^2}{(1 - c_i)^2} \frac{1}{P_{ij}}]}$$

Veamos un programa para la estimación de la habilidad.

esttheta.mv< −function(UIJ,CPT,A,C,nitem,bigt,maxit,npara,imptodo){

programa para estimar por máxima verosimilitud

```
if(npara==1){
for (i in 1:length(UIJ)){A[i]< −1.0
}
theta< −0
for(nit in 1: maxit){
sumnum< −0
sumdem< −0
for(i in 1:nitem){
DEV< −CPT[i]+A[i]*theta
PHAT< −1/(1+exp(-DEV))
if(npara¡3){
WIJ< −PHAT*(1-PHAT)
VIJ< −UIJ[i]-PHAT
SUMNUM< −SUMNUM+A[i]*VIJ
SUMDEM< −SUMDEM+A[i]2̂*WIJ
}
if(npara==3){
PT< −C[i]+(1-C[i])*PHAT
if(PT<.00001){PT< −.00001}
if(PT>.99999){PT< −.99999}
WIJ< −PT*(1-PT)
VIJ< −UIJ[i]-PT
PSP< −PHAT/PT
SUMNUM< −SUMNUM+A[i]*VIJ*PSP
SUMDEM< −SUMDEM+A[i]2̂*WIJ*PSP*PSP
}
}
DELTA< −SUMNUM/SUMDEM
if(DELTA < (-BIGT)){
DELTA< −-BIGT }
if(DELTA> BIGT){
DELTA< − BIGT
}
theta< −theta+DELTA
if(abs(DELTA)<.05){
return(theta)
}
} }
```

2.15. Estimación conjunta de parámetros del ítem y habilidad mediante máxima verosimilitud

Ahora se presenta un procedimiento mediante el cual se estiman conjuntamente la habilidad y los parámetros del ítem. El argumento central estará dado por el paradigma de Birnbaum que simplificará los algoritmos a utilizar. Este tipo de estimación exige la eliminación de las puntuaciones extremas tanto para ítems como para sujetos. Por lo tanto, se deben eliminar los sujetos que acierten o fallen todos los ítems. Tampoco puede usarse cuando los ítems pueden dividirse en dos clases, de tal forma que los sujetos acierten todos los ítems de la primera clase y fallen los de la segunda. Su principal limitación es que si tengo N sujetos y n ítems las estimaciones son inconsistentes para n fijo cuando $N \to \infty$. Son consistentes si $N \to \infty$, $n \to \infty$ y $\frac{N}{n} \to \infty$. Como en la mayoría de los casos se tiene un número limitado de ítems, las estimaciones conjuntas de los ítems no son consistentes y tampoco asintóticamente insesgadas.

Supongamos que N sujetos toman un test de n ítems. Tenemos como objetivo estimar conjuntamente los parámetros de los n ítems y los niveles de habilidad de los N individuos. No asumiremos ninguna suposición acerca de la distribución del rasgo. Las respuestas se puntúan en forma dicotómica, igual que antes la respuesta del sujeto j en el ítem i se notará por u_{ij}. Sea para el individuo j el vector de respuestas condicionado a su nivel de rasgo $U_j = (u_{1j}, u_{2j}, ..., u_{nj}|\theta_j)$ y sea U la matriz Nxn cuyo término general es u_{ij}.

Bajo el principio de independencia local las u_{ij} son independientes. A partir de esto la verosimilitud queda:

$$L = \prod_{j=1}^{N} \prod_{i=1}^{n} P_{ij}^{u_{ij}} Q_{ij}^{1-u_{ij}} \tag{2.108}$$

donde $P_{ij} = P_i(\theta_j)$ y entonces

$$\ln L = \sum_{j=1}^{N} \sum_{i=1}^{n} [u_{ij} \ln P_{ij} + (1 - u_{ij}) \ln Q_{ij}] \tag{2.109}$$

Las derivadas parciales son:

$$\frac{\partial \ln L}{\partial \zeta_i} = \sum_{j=1}^{N} [u_{ij} \frac{1}{P_{ij}} \frac{\partial P_{ij}}{\partial \zeta_i} + (1 - u_{ij}) \frac{1}{Q_{ij}} \frac{\partial Q_{ij}}{\partial \zeta_i}]$$

$$\frac{\partial \ln L}{\partial \lambda_i} = \sum_{j=1}^{N} [u_{ij} \frac{1}{P_{ij}} \frac{\partial P_{ij}}{\partial \lambda_i} + (1 - u_{ij}) \frac{1}{Q_{ij}} \frac{\partial Q_{ij}}{\partial \lambda_i}]$$

$$\frac{\partial \ln L}{\partial \theta_j} = \sum_{i=1}^{n} [u_{ij} \frac{1}{P_{ij}} \frac{\partial P_{ij}}{\partial \theta_j} + (1 - u_{ij}) \frac{1}{Q_{ij}} \frac{\partial Q_{ij}}{\partial \theta_j}]$$

Tenemos un sistema de 2n+N derivadas igualadas a 0. Usaremos el algoritmo de Newton Raphson, para ello necesitaremos las derivadas:

$$\frac{\partial^2 \ln L}{\partial \zeta_i^2},$$

$\frac{\partial^2 \ln L}{\partial \lambda_i^2}$, $\frac{\partial^2 \ln L}{\partial \theta_j^2}$, $\frac{\partial^2 \ln L}{\partial \theta_j \partial \zeta_i}$, $\frac{\partial^2 \ln L}{\partial \theta_j \partial \lambda_i}$, $\frac{\partial^2 \ln L}{\partial \lambda_i \partial \zeta_i}$

El algoritmo toma la forma $\psi_{t+1} = \psi_t - H_t^{-1}(\nabla L)_t$ donde ψ es el vector de dimensión 2n+N que contiene los parámetros a estimar; H la matriz hessiana (dimensión 2n+Nx2n+N) y ∇L el vector gradiente del logaritmo de la verosimilitud. La dimensionalidad del problema es grande así que se han tratado de encontrar formas de reducirla y con este objeto se hacen tres supuestos:

1- dado que los examinados se toman aleatoriamente de la población, cada examinado es un objeto independiente, luego las derivadas cruzadas entre pares de examinados tiene esperanza cero.

2- no hay motivo para que exista covariación entre un sujeto y los parámetros de los ítems, así que se asumen independientes y por lo tanto se eliminan las derivadas cruzadas.

3- se asume que los ítems son independientes y por lo tanto se eliminan las derivadas cruzadas entre ítems diferentes.

Después de estas simplificaciones la matriz hessiana sólo está compuesta de submatrices 2x2 para cada ítem y elementos en la diagonal para los sujetos, para ver la representación gráfica recomendamos consultar a Baker (1992: 92).

Aquí entra en juego el paradigma de Birnbaum (1968) que propone un proceso de estimación en dos etapas. En la primera etapa se estiman los parámetros de los ítems asumiendo que las habilidades son conocidas. Para iniciar el proceso iterativo se toman estimaciones crudas de las habilidades, en general, es el puntaje estandarizado del sujeto. Como por los supuestos se pueden considerar los ítems separadamente, se resuelve los estimados iterativamente por:

$$\begin{pmatrix} \widehat{\zeta} \\ \widehat{\lambda} \end{pmatrix}_{t+1} = \begin{pmatrix} \widehat{\zeta} \\ \widehat{\lambda} \end{pmatrix}_t - \begin{pmatrix} E(\frac{\partial^2 L}{\partial \zeta_i^2}) & E(\frac{\partial^2 L}{\partial \zeta_i \partial \lambda_i}) \\ E(\frac{\partial^2 L}{\partial \lambda_i \partial \zeta_i}) & E(\frac{\partial^2 L}{\partial \lambda_i^2}) \end{pmatrix}_t^{-1} \begin{pmatrix} \frac{\partial L}{\partial \zeta_i} \\ \frac{\partial^2 L}{\partial \lambda_i} \end{pmatrix}_t \tag{2.110}$$

A partir de resolver estas ecuaciones se tienen las estimaciones de los parámetros de los ítems y suponiéndolos conocidos podemos estimar las habilidades mediante

$$[\widehat{\theta}_j]_{t+1} = [\widehat{\theta}_j]_t - [E(\frac{\partial^2 L}{\partial \theta_j^2})]_t^{-1} [\frac{\partial L}{\partial \theta_j}]_t \tag{2.111}$$

En este punto un primer ciclo del paradigma de Birnbaum se ha hecho y tenemos estimados de los parámetros de los ítems ζ_i y λ_i para todo i y de las habilidades de los sujetos θ_j para todo j.

Como la métrica de θ es única a menos de transformaciones lineales, aunque no sería necesario fijar la escala, algunos programas como el LOGIST lo hacen mediante la media y la desviación típica de las habilidades. Se tiene entonces, el score estándar:

$$\theta_j^* = \frac{(\widehat{\theta}_j - \overline{\widehat{\theta}})}{s_{\widehat{\theta}}}$$

También se ajustan los parámetros de los ítems mediante

$$b_i^* = \frac{(\widehat{b}_i - \overline{\widehat{\theta}})}{s_{\widehat{\theta}}}$$

y

$$a_i^* = s_{\widehat{\theta}} \widehat{a}_i$$

Por último, se necesita un criterio de parada a los efectos de finalizar los ciclos. Este criterio es arbitrario. Por ejemplo, Wood, Wingersky y Lord (1976) proponen que la diferencia entre dos ciclos sucesivos debe ser menor a un valor específico.

2.16. Estimación del modelo de Rasch por máxima verosimilitud condicional

Rasch indica en 1960 que las expresiones de probabilidad condicional:

$$P((u_{.j} = r_j)|(u_{i.} = s_i)(\eta_j)) = \frac{\#A \prod_{j=1}^{N} \eta_j^{u_{.j}}}{\sum_A j \#A \prod_{j=1}^{N} \epsilon_i^{s_i}} = s_i)(\eta_j)) = \frac{\prod_{j=1}^{N} \eta_j^{u_{.j}}}{\sum_{Aj} \prod_{j=1}^{N} \epsilon_i^{s_i}} \quad (2.112)$$

$$P((u_{i.} = s_i)|(u_{.j} = r_j)(\epsilon_i)) = \frac{\#A \prod_{i=1}^{n} \epsilon_i^{u_{i.}}}{\sum_{Ai} \#A \prod_{i=1}^{n} \eta_j^{r_j}} = \frac{\prod_{i=1}^{n} \epsilon_i^{u_{i.}}}{\sum_{Ai} \prod_{i=1}^{n} \eta_j^{r_j}} \quad (2.113)$$

podrían ser usadas con propósitos de estimación. Rasch provee un algoritmo y su implementación, pero no funcionó bien en la práctica.

Andersen dio un método basado en estimación por máxima verosimilitud basada en probabilidades condicionales. Este procedimiento solo se aplica para estimar los parámetros de los ítems. Una vez que los ítems son calibrados se estima la habilidad. Se basa en la disponibilidad de estadísticos suficientes para la habilidad. En este procedimiento

se calcula la función de verosimilitud condicionada a la puntuación total. Además, se soluciona el problema que se daba en la estimación conjunta respecto a la cantidad de parámetros incidentales, pues la cantidad de posibles puntuaciones es igual a la cantidad de ítems más 1.

El modelo de Rasch establece que:

$$P(u_{ij}|\theta_j\beta_i) = \frac{e^{u_{ij}(\theta_j-\beta_i)}}{1+e^{(\theta_j-\beta_i)}}$$

Luego la probabilidad de que una persona j tenga un vector de respuestas (u_{ij}) para los n ítems viene dada por:

$$L = P((u_{ij})|\theta_j\beta_i) = \prod_{i=1}^{n} P(u_{ij}|\theta_j\beta_i) = \prod_{i=1}^{n} \frac{e^{u_{ij}(\theta_j-\beta_i)}}{1+e^{(\theta_j-\beta_i)}} = \frac{e^{(\sum_{i=1}^{n} u_{ij}\theta_j - \sum_{i=1}^{n} u_{ij}\beta_i)}}{\prod_{i=1}^{n}[1+e^{(\theta_j-\beta_i)}]}$$

luego

$$L = P((u_{ij})|\theta_j\beta_i) = \frac{e^{(r\theta_j - \sum_{i=1}^{n} u_{ij}\beta_i)}}{\prod_{i=1}^{n}[1+e^{(\theta_j-\beta_i)}]}$$

Una persona puede obtener una puntuación r de $\binom{n}{r}$ maneras y la suma de estas probabilidades de la probabilidad de $u_{.j} = r$ entonces:

$$P(u_{.j} = r_j|\theta_j,(\beta_i)) = \sum_{\{(u_{ij}):u_{.j}=r_j\}} P((u_{ij})|\theta_j,\beta_i) = \frac{e^{r\theta_j}\sum_{\{(u_{ij}):u_{.j}=r_j\}} e^{-\sum_{i=1}^{n} u_{ij}\beta_i)}}{\prod_{i=1}^{n}[1+e^{(\theta_j-\beta_i)}]} \tag{2.114}$$

luego la probabilidad condicional de un vector de respuestas (u_{ij}), dado que el puntaje es r es:

$$P((u_{ij})|u_{.j} = r) = \frac{P(u_{ij}|\theta_j,(\beta_i))}{P(u_{.j} = r|\theta_j,(\beta_i))} = \frac{\frac{e^{(r\theta_j - \sum_{i=1}^{n} u_{ij}\beta_i)}}{\prod_{i=1}^{n}[1+e^{(\theta_j-\beta_i)}]}}{\frac{e^{r\theta_j}\sum_{\{(u_{ij}):u_{.j}=r_j\}} e^{-\sum_{i=1}^{n} u_{ij}\beta_i}}{\prod_{i=1}^{n}[1+e^{(\theta_j-\beta_i)}]}} = \frac{e^{-\sum_{i=1}^{n} u_{ij}\beta_i}}{\sum_{\{(u_{ij}):u_{.j}=r_j\}} e^{-\sum_{i=1}^{n} u_{ij}\beta_i}}$$

Si N personas responden al conjunto de ítems –y se asumen independientes las respuestas– entonces la probabilidad condicional de la matriz de respuestas dados los puntajes fila $r_1, r_2, ...r_N$ es:

$$P([u_{ij}]|r_1, r_2, ...r_N\beta_1...\beta_n) = \frac{e^{-\sum_{j=1}^{N}\sum_{i=1}^{n} u_{ij}\beta_i}}{\prod_{j=1}^{N}\sum_{\{(u_{ij}):u_{.j}=r_j\}} e^{-\sum_{i=1}^{n} u_{ij}\beta_i}}$$

Ahora,

$$\prod_{j=1}^{N} \sum_{\{(u_{ij}):u_{.j}=r_j\}} e^{-\sum_{i=1}^{n} u_{ij}\beta_i} = \prod_{r=0}^{n} (\sum_{\{(u_{ij}):u_{.j}=r_j\}} e^{-\sum_{i=1}^{n} u_{ij}\beta_i})^{f_r}$$

donde f_r indica la frecuencia del puntaje r.

Sea $s_i = \sum_{j=1}^{N} u_{ij}$; el puntaje del ítem entonces es:

$$P([u_{ij}]|r_1, r_2, ...r_N\beta_1...\beta_n) = \frac{e^{-\sum_{i=1}^{n} s_i\beta_i}}{\prod_{r=0}^{n}(\sum_{\{(u_{ij}):u_{.j}=r_j\}} e^{-\sum_{i=1}^{n} u_{ij}\beta_i})^{f_r}}$$

Para obtener los estimadores de máxima verosimilitud debe tomarse las derivadas respecto de los parámetros.

Andersen (1973) prueba que:

$$\frac{\partial(\sum_{\{(u_{ij}):u_{.j}=r\}} e^{-\sum_{i=1}^{n} u_{ij}\beta_i})}{\partial\beta_h} = -e^{-\beta_h} \sum_{\{(u_{ij}):u_{.j}=r-1\}} e^{-\sum_{i=1}^{n} u_{ij}\beta_i}$$

y

$$\frac{\partial^2 \sum_{\{(u_{ij}):u_{.j}=r\}} e^{-\sum_{i=1}^{n} u_{ij}\beta_i}}{\partial\beta_h^2} = -e^{-\beta_h} \sum_{\{(u_{ij}):u_{.j}=r-1 \ i\neq h\}} e^{-\sum_{i=1}^{n} u_{ij}\beta_i}$$

donde el ítem h se removió del cálculo de la suma:

$$\frac{\partial^2 \sum_{\{(u_{ij}):u_{.j}=r\}} e^{-\sum_{i=1}^{n} u_{ij}\beta_i}}{\partial\beta_h\partial\beta_k} = -e^{-\beta_h-\beta_k} \sum_{\{(u_{ij}):u_{.j}=r-2 \ i\neq h,k\}} e^{-\sum_{i=1}^{n} u_{ij}\beta_i}$$

donde los ítems h y k se removieron del cálculo de la suma.

El logaritmo de la verosimilitud es:

$$\ln L = -\sum_{i=1}^{n} s_i\beta_i - \sum_{r=1}^{n-1} f_r \ln(\sum_{\{(u_{ij}):u_{.j}=r\}} e^{-\sum_{i=1}^{n} u_{ij}\beta_i})$$

$$L_h = \frac{\partial \ln L}{\partial\beta_h} = -s_h - \sum_{r=1}^{n-1} f_r \frac{\partial \ln(\sum_{\{(u_{ij}):u_{.j}=r\}} e^{-\sum_{i=1}^{n} u_{ij}\beta_i})}{\partial\beta_h} =$$

$$= -s_h + \frac{\sum_{r=1}^{n-1} f_r e^{-\beta_h}\gamma_{r-1}}{\sum_{\{(u_{ij}):u_{.j}=r\}} e^{-\sum_{i=1}^{n} u_{ij}\beta_i}} = 0$$

Además,

$$L_{hh} = \frac{e^{-\beta_h} \sum_{\{(u_{ij}):u_{.j}=r-1 \ i\neq h\}} e^{-\sum_{i=1}^{n} u_{ij}\beta_i}}{\sum_{\{(u_{ij}):u_{.j}=r\}} e^{-\sum_{i=1}^{n} u_{ij}\beta_i}} - [\frac{e^{-\beta_h} \sum_{\{(u_{ij}):u_{.j}=r-1 \ i\neq h\}} e^{-\sum_{i=1}^{n} u_{ij}\beta_i}}{\sum_{\{(u_{ij}):u_{.j}=r\}} e^{-\sum_{i=1}^{n} u_{ij}\beta_i}}]^2$$

y

$$L_{hk} = \frac{e^{-\beta_h-\beta_k} \sum_{\{(u_{ij}):u_{.j}=r-2 \ i\neq h,k\}} e^{-\sum_{i=1}^{n} u_{ij}\beta_i}}{\sum_{\{(u_{ij}):u_{.j}=r\}} e^{-\sum_{i=1}^{n} u_{ij}\beta_i}} -$$

$$- \frac{e^{-\beta_h} \sum_{\{(u_{ij}):u_{.j}=r-1 \ i\neq h\}} e^{-\sum_{i=1}^{n} u_{ij}\beta_i}}{(\sum_{\{(u_{ij}):u_{.j}=r\}} e^{-\sum_{i=1}^{n} u_{ij}\beta_i})(\beta_h)} \frac{e^{-\beta_k} \sum_{\{(u_{ij}):u_{.j}=r-1 \ i\neq h\}} e^{-\sum_{i=1}^{n} u_{ij}\beta_i}}{(\sum_{\{(u_{ij}):u_{.j}=r\}} e^{-\sum_{i=1}^{n} u_{ij}\beta_i})(\beta_k)}$$

El algoritmo iterativo queda:

$$\begin{pmatrix} \widehat{\beta_1} \\ \widehat{\beta_2} \\ \vdots \\ \widehat{\beta_n} \end{pmatrix}_{t+1} = \begin{pmatrix} \widehat{\beta_1} \\ \widehat{\beta_2} \\ \vdots \\ \widehat{\beta_n} \end{pmatrix}_{t} - \begin{pmatrix} L_{11} & L_{12} & ... & L_{1n} \\ L_{21} & L_{22} & ... & L_{2n} \\ \vdots & \vdots & & \vdots \\ L_{n1} & L_{n2} & \vdots & L_{nn} \end{pmatrix}_{t}^{-1} \begin{pmatrix} L_1 \\ L_2 \\ \vdots \\ L_n \end{pmatrix}_{t} \tag{2.115}$$

Hay una indeterminación en la definición de la métrica. Andersen (1973 y 1977) impone la restricción $\sum_{i=1}^{n} \beta_i = 0$ a los efectos de determinar la escala.

Wright y Douglas (1977) implementaron esta restricción sustrayendo la derivada de un ítem de las restantes, por ejemplo el ítem n.

Las derivadas resultantes son:

$$L'_h = L_h - L_n$$

y $L'_{hk} = L_{hk} - L_{hn} - L_{nk} + L_{nn}$

El algoritmo de Newton Raphson queda:

$$\begin{pmatrix} \widehat{\beta_1} \\ \widehat{\beta_2} \\ \vdots \\ \widehat{\beta_{n-1}} \end{pmatrix}_{t+1} = \begin{pmatrix} \widehat{\beta_1} \\ \widehat{\beta_2} \\ \vdots \\ \widehat{\beta_{n-1}} \end{pmatrix}_{t} - \begin{pmatrix} L'_{11} & L'_{12} & ... & L'_{1(n-1)} \\ L'_{21} & L'_{22} & ... & L'_{2(n-1)} \\ \vdots & \vdots & & \vdots \\ L'_{(n-1)1} & L'_{(n-1)2} & \vdots & L'_{(n-1)(n-1)} \end{pmatrix}_{t}^{-1} \begin{pmatrix} L'_1 \\ L'_2 \\ \vdots \\ L'_{(n-1)} \end{pmatrix}_{t} \tag{2.116}$$

Wright y Douglas (1977) implementan el algoritmo vía los siguientes pasos:

1- Obtener estimadores iniciales mediante

$$\widehat{\beta}_i^0 = \ln[\frac{N - s_i}{s_i}] - \frac{\sum_{i=1}^{n} \ln[\frac{N-s_i}{s_i}]}{n}$$

2- Usando el conjunto de $\widehat{\beta}_i$ (al momento) se calculan las funciones

$$\sum_{\{(u_{ij}):u_{.j}=r\}} e^{-\sum_{i=1}^{n} u_{ij}\beta_i}$$

$$\sum_{\{(u_{ij}):u_{.j}=r-1 \ i\neq h\}} e^{-\sum_{i=1}^{n} u_{ij}\beta_i}$$

$$\sum_{\{(u_{ij}):u_{.j}=r-2 \ i\neq h,k\}} e^{-\sum_{i=1}^{n} u_{ij}\beta_i}$$

3- Evaluar $\widehat{L_h}$, y $\widehat{L_{hk}}$

4- Reducir el número de ítems a ser estimados, calculando $\widehat{L'_h}$ y $\widehat{L'_{hk}}$

5- Resolver el algoritmo de N-R anterior para obtener los n-1 estimados, luego $\beta_h = -\sum_{i=1}^{n-1} \widehat{\beta}_i$

6- Repetir los pasos 2-5 hasta el criterio de convergencia:

$$\sum_{i=1}^{n-1} \frac{(\widehat{\beta}_i^{t+1} - \widehat{\beta}_i^{t})^2}{n} < 0{,}001$$

2.17. Estimación del modelo de Rasch por máxima verosimilitud conjunta

La verosimilitud conjunta es:

$$L = \prod_{j=1}^{N} \prod_{i=1}^{n} P(u_{ij}|\theta_j \beta_i) = \frac{e^{\sum_{j=1}^{N} \sum_{i=1}^{n} u_{ij}(\theta_j - \beta_i)}}{\prod_{j=1}^{N} \prod_{i=1}^{n} [1 + e^{(\theta_j - \beta_i)}]}$$

Luego,

$$\ln L = \sum_{j=1}^{N} r_j \theta_j - \sum_{i=1}^{n} s_i \beta_i - \sum_{j=1}^{N} \sum_{i=1}^{n} \ln(1 + e^{(\theta_j - \beta_i)})$$

Derivando y agrupando se tiene:

$$\frac{\partial \ln L}{\partial \theta_g} = r_g - \sum_{i=1}^{n} P_{ig}$$

$$\frac{\partial \ln L}{\partial \beta_i} = L_i = -s_i + \sum_{g=1}^{n-1} f_{.g} P_{ig}$$

donde $f_{.g}$ es la frecuencia del puntaje g (se descartan los puntajes 0 y n)

$$\frac{\partial^2 \ln L}{\partial \theta_g^2} = -\sum_{i=1}^{n} P_{ig} Q_{ig}$$

y

$$\frac{\partial^2 \ln L}{\partial \beta_i^2} = -\sum_{g=1}^{n-1} f_{.g} P_{ig} Q_{ig}$$

Con estas condiciones se aplica el método de Newton Raphson. Se pueden implementar estas soluciones en el siguiente algoritmo:

1- Se obtienen estimados preliminares de $\widehat{\theta}_g$ y $\widehat{\beta}_i$ usando una técnica llamada PROX (debido a Cohen) que toma:

a)

$$\widehat{\beta}_i^0 = \ln[\frac{N - s_i}{s_i}] - \frac{\sum_{i=1}^{n} \ln[\frac{N-s_i}{s_i}]}{n}$$

b)

$$\widehat{\theta}_g^0 = \ln[\frac{r_g}{n - r_g}]$$

c) Estos estimadores basales se reescalan tomando en cuenta sus varianzas y que tengan sus medias iguales a 0.

2- A partir del procedimiento de N-R anterior se estima β como

$$\widehat{\beta}_i^{t+1} = \widehat{\beta}_i^t - [\frac{s_i - \sum_{g=1}^{n-1} f_{.g}\widehat{P_{ig}}}{\sum_{g=1}^{n-1} f_{.g}\widehat{P_{ig}}\widehat{Q_{ig}}}]_t$$

Se itera hasta que:

$$|\widehat{\beta}_i^{t+1} - \widehat{\beta}_i^t| < 0{,}05$$

3- A cada $\widehat{\beta}_i$ se le resta $\overline{\widehat{\beta}}$.

4- Considerando los reescalados $\widehat{\beta}_i$ como conocidos se utiliza N-R nuevamente y se obtiene la iteración:

$$\widehat{\theta}_r^{t+1} = \widehat{\theta}_r^t + [\frac{r_g - \sum_{i=1}^{n} \widehat{P_{ig}}}{\sum_{i=1}^{n} \widehat{P_{ig}}\widehat{Q_{ig}}}]$$

hasta que:

$$|\widehat{\theta}_g^{t+1} - \widehat{\theta}_g^t| < 0{,}05$$

5- Los pasos 2-4 se repiten hasta que se satisface el criterio:

$$\sum_{i=1}^{n-1} \frac{(\widehat{\beta}_i^{y+1} - \widehat{\beta}_i^{y})}{n} < 0{,}025$$

6- Se corrigen los $\widehat{\beta}_i$ del sesgo mediante:

$$\widehat{\beta}_i = \widehat{\beta}_i \frac{n-1}{n}$$

7- Un paso adicional se hace con los nuevos parámetros de los ítems para estimar la habilidad y luego se corrige el sesgo.

8- Se estiman los errores estándares de α y θ mediante

$$SE(\widehat{\beta}_i) = \frac{1}{\sqrt{\sum_{g=1}^{n-1} f_{.g} \widehat{P_{ig}} \widehat{Q_{ig}}}}$$

y

$$SE(\widehat{\theta}_g) = \frac{1}{\sqrt{\sum_{i=1}^{n} \widehat{P_{ig}} \widehat{Q_{ig}}}}$$

La función de información del test es

$$I(\widehat{\theta}_g) = \sum_{i=1}^{n} \widehat{P_{ig}} \widehat{Q_{ig}}$$

A continuación presentamos el programa que implementa lo visto.

```
jmlerasch< −function(FDG,S,nitem,maxscore,maxcycle,maxit,imptodo){
#programa para implementar maxima verosimilitud conjunta de Birnbaum
#modelo de Rasch
B< −0
theta< −0
```

```
#estimados iniciales
N< −0
for(j in 1:maxscore){
N< −N+FDG[j]
}
sumi< −0
for(i in 1:nitem){
B[i]< −log((N-S[i])/S[i])
sumi< −sumi+B[i]
}
meanl< −sumi/nitem
for(i in 1:nitem){
B[i]< −B[i]-meanl
}
#Estimados de habilidad
for(j in 1:maxscore){
theta[j]< −log(j/(nitem-j))
}
first< −"Y"
last< −"N"
for(k in 1:maxcycle){
sumb< −0
#estimacion de los parametros de dificultad
for(i in 1:nitem){
for(kk in 1:maxit){
sumfdgp< −0
sumfdgq< −0
for(j in 1:maxscore){
DEV< −(theta[j]-B[i])
PIJ< −1/(1+exp(-DEV))
sumfdgp< −sumfdgp+FDG[j]*PIJ
sumfdgq< −sumfdgq+FDG[j]*PIJ*(1-PIJ)
}
DELTAB< −(S[i]-sumfdgp)/sumfdgq
B[i]< −B[i]-DELTAB
if(abs(DELTAB)<.05){exit}
}
sumb< −sumb+B[i]
}
bbar< −sumb/nitem
```

```
if(first==”N”){
if(abs(bbar-bbold)>.05){bbold< −bbar
}
else{
#correcion de la dificultad del item por sesgo
for(i in 1:nitem){
B[i]< −B[i]*((nitem-1)/nitem)
}
last< −”Y”
}
}
else {
first< −”N”
bbold< −bbar
}
#estimacion de habilidades
for(j in 1:maxscore){
for(kk in 1:maxit){
sump< −0
sumq< −0
for(i in 1:nitem){
DEV< −(theta[j]-B[i])
PIJ< −1/(1+exp(-DEV))
sump< −sump+PIJ
sumq< −sumq+PIJ*(1-PIJ)
}
DELTA< −(j-sump)/sumq
theta[j]< −theta[j]+DELTA
if(abs(DELTA)¡.05){exit}
}
}
if(last==”N”){next}
for(j in 1:maxscore){
theta[j]< −theta[j]*(nitem-2)/nitem
}
if(last==”Y”){return(theta,B)}
}
#corrección de la dificultad del ítem por sesgo
for(i in 1:nitem){
B[i]< −B[i]*((nitem-1)/nitem)
```

```
}
#estimación de habilidades for(j in 1:maxscore){
for(kk in 1:maxit){
sump< −0
sumq< −0
for(i in 1:nitem){
DEV< −(theta[j]-B[i])
PIJ< −1/(1+exp(-DEV))
sump< −sump+PIJ
sumq< −sumq+PIJ*(1-PIJ)
}
DELTA< −(j-sump)/sumq
theta[j]< −theta[j]+DELTA
if(abs(DELTA)<.05){exit}
}
}
for(j in 1:maxscore){
theta[j]< −theta[j]*(nitem-2)/nitem
} return(theta,B) }
```

2.18. Estimación vía máxima verosimilitud marginal

Aunque este proceso no se considera una técnica bayesiana el proceso de marginalización incorpora el teorema de Bayes en su matemática.

Sea $f(\theta|\tau)$ la densidad de probabilidad de la habilidad en la población (distribución a priori) y τ el vector de parámetros de la densidad, se supondrá que la distribución a priori es la misma para todos los examinados. Sea U_j el vector de respuestas del sujeto j y ξ el vector de parámetros de los ítems. Se aplica el teorema de Bayes para calcular la distribución a posteriori de θ_j. Se tiene que:

$$P(\theta_j|U_j,\tau,\xi) = \frac{P(U_j|\theta_j,\xi)f(\theta|\tau)}{\int P(U_j|\theta_j,\xi)f(\theta_j|\tau)d\theta} \tag{2.117}$$

donde $P(U_j|\theta_j,\xi)$ es la verosimilitud.

Dado que hay independencia local,

$$P(U_j|\theta_j,\xi) = \prod_{i=1}^{n} P_{ij}^{u_{ij}} Q_{ij}^{1-u_{ij}}$$

La integral que aparece en el denominador es la probabilidad marginal (no condicionada a θ_j) del vector de respuestas con respecto a los parámetros de los ítems.

2.18.1. La solución de Bock y Lieberman

Para simplificar la notación notaremos por

$$P(U_j) = \int P(U|\theta,\xi) f(\theta|\tau) d\theta$$

Así la función de verosimilitud marginal es

$$L = \prod_{j=1}^{N} P(U_j)$$

y

$$\ln L = \sum_{j=1}^{N} \ln P(U_j).$$

El sistema a resolver es:

$$\frac{\partial \ln L}{\partial a_i} = 0$$

$$\frac{\partial \ln L}{\partial b_i} = 0$$

$$\frac{\partial \ln L}{\partial c_i} = 0$$

Derivamos obteniendo:

$$\frac{\partial \ln L}{\partial a_i} = \sum_{j=1}^{N} \frac{\partial \ln P(U_j)}{\partial a_i} =$$

$$= \sum_{j=1}^{N} [P(U_j)]^{-1} \int \frac{\partial}{\partial a_i} [P(U_j|\theta, \xi)] f(\theta|\tau) d\theta.$$

$$\frac{\partial}{\partial a_i} [P(U_j|\theta, \xi)] = \frac{\partial}{\partial a_i} [\ln P(U_j|\theta, \xi)] P(U_j|\theta, \xi)$$

Reemplazando:

$$\frac{\partial \ln L}{\partial a_i} = \sum_{j=1}^{N} \int \frac{\partial}{\partial a_i} [\ln P(U_j|\theta, \xi)] [\frac{P(U_j|\theta, \xi) f(\theta|\tau)}{P(U_j)}] d\theta$$

y a partir de la regla de Bayes:

$$\frac{\partial \ln L}{\partial a_i} = \sum_{j=1}^{N} \int \frac{\partial}{\partial a_i} [\ln P(U_j|\theta, \xi)] P(\theta|U_j, \xi, \tau) d\theta$$

luego

$$\frac{\partial \ln L}{\partial a_i} = \sum_{j=1}^{N} \int \frac{\partial}{\partial a_i} [\ln \prod_{i=1}^{n} P_i^{u_{ij}} Q_i^{1-u_{ij}}] P(\theta|U_j, \xi, \tau) d\theta =$$

$$\sum_{j=1}^{N} \int (\prod_{i=1}^{n} P_i^{u_{ij}} Q_i^{1-u_{ij}})^{-1} \frac{\partial}{\partial a_i} [\prod_{i=1}^{n} P_i^{u_{ij}} Q_i^{1-u_{ij}}] P(\theta|U_j, \xi, \tau) d\theta$$

Ahora:

$$\frac{\partial}{\partial a_i} [\prod_{i=1}^{n} P_i^{u_{ij}} Q_i^{1-u_{ij}}] = [\prod_{h \neq i}^{n} P_i^{u_{ij}} Q_i^{1-u_{ij}}] \frac{\partial}{\partial a_i} [P_i^{u_{ij}} Q_i^{1-u_{ij}}]$$

.

Como

$$\frac{\partial P_i(\theta)}{\partial a_i} = (1 - c_i)(\theta - b_i) P_i^*(\theta) Q_i^*(\theta) = K$$

donde

$$P_i^*(\theta) = \frac{e^{a_i(\theta - b_i)}}{1 + e^{a_i(\theta - b_i)}}$$

continuando

$$\frac{\partial}{\partial a_i}[\prod_{i=1}^{n} P_i^{u_{ij}} Q_i^{1-u_{ij}}] = [\prod_{h \neq i}^{n} P_h^{u_{ij}} Q_h^{1-u_{ij}}][P_i(\theta)^{u_{ij}}(1 - u_{ij})]Q_i(\theta)^{1-u_{ij}-1}(-K)+$$

$$+Q_i(\theta)^{1-u_{ij}} u_{ij} P_i(\theta)^{u_{ij}-1}(K)$$

Finalmente,

$$\frac{\partial}{\partial a_i}[\prod_{i=1}^{n} P_i^{u_{ij}} Q_i^{1-u_{ij}}] =$$

$$(-1)^{u_{ij}+1}(1 - c_i)[\prod_{h \neq i}^{n} P_h^{u_{ij}} Q_h^{1-u_{ij}}](\theta - b_i)P_i^*(\theta)Q_i^*(\theta)$$

y sustituyendo

$$\frac{\partial \ln L}{\partial a_i} = \sum_{j=1}^{N}(-1)^{u_{ij}+1}(1 - c_i)\int(\theta_j - b_i)$$

$$[\prod_{i=1}^{n} P_i(\theta_j)^{u_{ij}} Q_i(\theta_j)^{1-u_{ij}}]^{-1}$$

$$[\prod_{h \neq i}^{n} P_h(\theta_j)^{u_{ij}} Q_h(\theta_j)^{1-u_{ij}}]$$

$$P_i^*(\theta_j)Q_i^*(\theta_j)P(\theta_j|U_j, \xi, \tau)d\theta$$

Multiplicando numerador y denominador por:

$$P_i(\theta_j)^{u_{ij}} Q_i(\theta_j)^{1-u_{ij}}$$

el producto sobre $h \neq i$ se convierte en un producto sobre i.

Además,

$$\sum_{j=1}^{N} (-1)^{u_{ij}+1} \frac{1}{P_i(\theta_j)^{u_{ij}} Q_i(\theta_j)^{1-u_{ij}}} = \sum_{j=1}^{N} [u_{ij} - P_i(\theta_j)] \frac{1}{P_i(\theta_j) Q_i(\theta_j)}$$

Finalmente se tiene:

$$\frac{\partial \ln L}{\partial a_i} = (1 - c_i) \sum_{j=1}^{N} \int [u_{ij} - P_i(\theta_j)] w_{ij} (\theta_j - b_i) P(\theta_j | U_j, \xi, \tau) d\theta$$

donde

$$w_{ij} = \frac{P_i^*(\theta_j) Q_i^*(\theta_j)}{P_i(\theta_j) Q_i(\theta_j)}$$

Análogamente, se obtiene:

$$\frac{\partial \ln L}{\partial b_i} = -a_i (1 - c_i) \sum_{j=1}^{N} \int [u_{ij} - P_i(\theta_j)] w_{ij} P(\theta_j | U_j, \xi, \tau) d\theta$$

$$\frac{\partial \ln L}{\partial c_i} = (1 - c_i)^{-1} \sum_{j=1}^{N} \int \frac{[u_{ij} - P_i(\theta_j)]}{P_i(\theta_j)} w_{ij} P(\theta_j | U_j, \xi, \tau) d\theta$$

Las ecuaciones anteriores contienen en su formulación integrales. Para aproximar dichas integrales se aplica la cuadratura de Hermite-Gauss. Si se tiene una densidad continua con momento finito $f(\theta|\tau)$ ésta puede ser aproximada mediante un histograma. Luego la integral puede ser reemplazada por la suma de las áreas de un número finito de rectángulos. Supongamos que tenemos q intervalos y sea X_k el punto medio del intervalo k, al que llamaremos nodo. Cada nodo tiene un peso $A(X_k)$ que toma en cuenta la altura de la función de densidad en un entorno de X_k. Bock y Lieberman (1970) modifican los

nodos y pesos de Stroud y Sechrest para el caso de la densidad normal típica. Sustituyendo la aproximación de la integral por su fórmula de cuadratura resulta:

$$\frac{\partial \ln L}{\partial a_i} = (1 - c_i) \sum_{k=1}^{q} \sum_{j=1}^{N} [u_{ij} - P_i(X_k)] w_{ik} (X_k - b_i) P(X_k | U_j, \xi, \tau)$$

$$\frac{\partial \ln L}{\partial b_i} = -a_i (1 - c_i) \sum_{k=1}^{q} \sum_{j=1}^{N} [u_{ij} - P_i(X_k)] w_{ik} P(X_k | U_j, \xi, \tau)$$

$$\frac{\partial \ln L}{\partial c_i} = (1 - c_i)^{-1} \sum_{k=1}^{q} \sum_{j=1}^{N} \frac{[u_{ij} - P_i(X_k)]}{P_i(X_k)} P(X_k | U_j, \xi, \tau)$$

Ahora se puede aplicar el método de N-R para estimar los 3n parámetros del test simultáneamente. Como en el caso de la estimación conjunta, los u_{ij} son remplazados por su esperanza, o sea se aplica el método de Fisher. El método es no atractivo desde el punto de vista computacional ya que requiere invertir una matriz 3n x 3n y es por esto que este procedimiento se limita a test con muy pocos ítems.

2.18.2. La solución de Bock y Aitkin

Los autores asumen que los ítems son independientes, los sujetos también y que son indpendientes los sujetos de los ítems, es decir, utilizan los mismos supuestos que Birnbaum en su paradigma para la estimación conjunta. De la sección anterior tenemos:

$$\frac{\partial \ln L}{\partial a_i} = (1 - c_i) \sum_{k=1}^{q} (X_k - b_i) [\sum_{j=1}^{N} u_{ij} P(X_k | U_j, \xi, \tau) - P_i(X_k) \sum_{j=1}^{N} P(X_k | U_j, \xi, \tau)] w_{ij}$$

y la distribución a posteriori de la habilidad como

$$P(\theta_j | U_j, \tau, \xi) = \frac{P(U_j | \theta_j, \xi) f(\theta | \tau)}{\int P(U_j | \theta_j, \xi) f(\theta_j | \tau) d\theta}$$

Si la ponemos en su forma de cuadratura:

$$P(X_k | U_j, \tau, \xi) = \frac{\prod_{i=1}^{n} P_i(X_k)^{u_{ij}} Q_i(X_k)^{1-u_{ij}} A(X_k)}{\sum_{k=1}^{q} \prod_{i=1}^{n} P_i(X_k)^{u_{ij}} Q_i(X_k)^{1-u_{ij}} A(X_k)}$$

con

$$P_i(X_k) = c_i + (1 - c_i)\frac{e^{a_i(X_k - b_i)}}{1 + e^{a_i(X_k - b_i)}}$$

definamos

$$\overline{f_{ik}} = \sum_{j=1}^{N} P(X_k|U_j, \xi, \tau) = \sum_{j=1}^{N}[\frac{\prod_{i=1}^{n} P_i(X_k)^{u_{ij}} Q_i(X_k)^{1-u_{ij}} A(X_k)}{\sum_{k=1}^{q}\prod_{i=1}^{n} P_i(X_k)^{u_{ij}} Q_i(X_k)^{1-u_{ij}} A(X_k)}]$$

$$\overline{r_{ik}} = \sum_{j=1}^{N} u_{ij} P(X_k|U_j, \xi, \tau) = \sum_{j=1}^{N}[\frac{\prod_{i=1}^{n} u_{ij} P_i(X_k)^{u_{ij}} Q_i(X_k)^{1-u_{ij}} A(X_k)}{\sum_{k=1}^{q}\prod_{i=1}^{n} P_i(X_k)^{u_{ij}} Q_i(X_k)^{1-u_{ij}} A(X_k)}]$$

Sea ahora $L(X_k)$ la probabilidad condicional del vector de respuestas U_j dado $\theta = X_k$ y los parámetros de los ítems, o sea:

$$L(X_k) = \prod_{i=1}^{n} P_i(X_k)^{u_{ij}} Q_i(X_k)^{1-u_{ij}}$$

Sustituyendo:

$$P(X_k|U_j, \tau, \xi) = \frac{L(X_k)A(X_k)}{\sum_{k=1}^{q} L(X_k)A(X_k)}$$

luego

$$\overline{f_{ik}} = \sum_{j=1}^{N}[\frac{L(X_k)A(X_k)}{\sum_{k=1}^{q} L(X_k)A(X_k)}]$$

y

$$\overline{r_{ik}} = \sum_{j=1}^{N}[\frac{u_{ij}L(X_k)A(X_k)}{\sum_{k=1}^{q} L(X_k)A(X_k)}]$$

Las ecuaciones a resolver son:

$$\frac{\partial \ln L}{\partial a_i} = (1 - c_i)\sum_{k=1}^{q}(X_k - b_i)[\overline{r_{ik}} - \overline{f_{ik}}P_i(X_k)]w_{ij} = 0$$

$$\frac{\partial \ln L}{\partial b_i} = (-a_i)(1-c_i)\sum_{k=1}^{q}[\overline{r_{ik}} - \overline{f_{ik}}P_i(X_k)]w_{ij} = 0$$

$$\frac{\partial \ln L}{\partial c_i} = (1-c_i)^{-1}\sum_{k=1}^{q}\frac{[\overline{r_{ik}} - \overline{f_{ik}}P_i(X_k)]}{P_i(X_k)} = 0$$

A los valores $\overline{r_{ik}}$ $\overline{f_{ik}}$ se les llama datos artificiales. Los parámetros de los ítems se estiman usando el método de Newton Raphson o Fisher. Los estimadores iniciales de los parámetros que suelen emplearse se basan en $\rho_{\theta U_i} = \frac{\alpha_i}{\sqrt{1+\alpha_i^2}}$ donde $\rho_{\theta U_i}$ es el coeficiente de correlación biserial entre θ y la variable de respuesta U y la dificultad de ítem $P_i = \frac{1}{\sqrt{2\pi}}\int_{\delta_i}^{\infty} e^{-\frac{u^2}{2}}du$ donde $\delta_i = \beta_i\rho_{\theta U_i} = \beta_i\frac{\alpha_i}{\sqrt{1+\alpha_i^2}}$ y de estas se pueden obtener α_i y β_i.

En el método MMLE los valores de $\overline{r_{ik}}$ $\overline{f_{ik}}$ dependen de los estimadores de los parámetros de los ítems, por lo tanto, debe establecerse una metodología donde los datos artificiales y los parámetros de los ítems se pueden obtener iterativamente. Para esto Bock y Aitkin emplearon el algoritmo EM de Dempster, Laird y Rubin (1977).

El algoritmo EM como ya dijimos es un procedimiento iterativo que sirve para encontrar los estimadores máximo verosimiles en presencia de variables no observables. Consta de dos pasos: E (esperanza) y M (maximización). En nuestro caso queremos estimar los parámetros de los ítems en presencia de una variable aleatoria no observable (θ). Para realizar inferencia acerca de θ se usa alguna representación observable basada en la respuesta de los sujetos. Sea el par (U, θ) los datos no observados (datos completos) y el vector U los datos observados (datos incompletos).

Sea $f(U, \theta|\xi)$ la densidad de probabilidad conjunta de (U, θ) y ξ^p los parámetros estimados en el ciclo p, entonces ξ^{p+1} se calculan maximizando $E(\ln f(U, \theta|\xi)|U, \xi^p))$ con respecto a ξ. Este proceso se repite hasta satisfacer el criterio de convergencia. Los dos pasos serían entonces:

Paso E: Calcular $E(\ln f(U, \theta|\xi)|U, \xi^p))$

Paso M: Calcular ξ^{p+1} como aquel que maximiza la esperanza posterior.

Comencemos suponiendo que la distribución del rasgo es finita discreta, es decir θ toma finitos valores $\theta_1, \theta_2, ..., \theta_q$ con probabilidades $\pi_1, \pi_2, ..., \pi_q$. y sea f_{ik} la frecuencia

de sujetos en cada valor de rasgo para el ítem i y r_{ik} la frecuencia de respuestas correctas del ítem i para el valor de rasgo k.

Si los N sujetos son muestreados aleatoriamente de una población con la distribución dada, entonces la distribición conjunta de los f_{ik} es una multinomial dada por $[\frac{f!}{f_{i1}!\ldots f_{iq}!}] \prod_{k=1}^{q} \pi_k^{f_{ik}}$.

Dada f_{ik} y θ_k la probabilidad conjunta de los r_{ik} es

$$\prod_{i=1}^{q} \binom{f_{ik}}{r_{ik}} P_i(\theta_k)^{r_{ik}} Q_i(\theta_k)^{f_{ik}-r_{ik}}$$

y la probabilidad conjunta de f y r es

$$\prod_{k=1}^{q} \prod_{i=1}^{q} \binom{f_{ik}}{r_{ik}} P_i(\theta_k)^{r_{ik}} Q_i(\theta_k)^{f_{ik}-r_{ik}} [\frac{f!}{f_{i1}!\ldots f_{iq}!}] \prod_{k=1}^{q} \pi_k^{f_{ik}}$$

luego por el criterio de factorización (f,r), son estadísticos suficientes para (U, θ). Si se ignoran los términos constantes, el logaritmo de la verosimilitud para los datos completos es

$$\ln(L) = \sum_{k=1}^{q} \sum_{i=1}^{q} r_{ik} \ln P_i(\theta_k) + (f_{ik} - r_{ik}) \ln Q_i(\theta_k) + \sum_{k=1}^{q} f_{ik} \ln \pi_k$$

Aquí (f,r) son no observados, pero tomando la esperanza posterior del logaritmo de la verosimilitud dado ξ obtenemos:

$$E(\ln L) = \sum_{k=1}^{q} \sum_{i=1}^{q} E(r_{ik}|U, \xi) \ln P_i(\theta_k) + E[(f_{ik} - r_{ik})|U, \xi] \ln Q_i(\theta_k) +$$

$$+ \sum_{k=1}^{q} E(f_{ik}|U) \ln \pi_k$$

La última suma no depende de ξ. Maximizar los términos remanentes es equivalente a maximizar el paso E. Dado que los ítems se consideran independientes, las derivadas cruzadas segundas de ítems diferentes son 0 en el paso M y la maximización de la esperanza se hace para cada ítem separadamente.

Podemos resumir los pasos como:

PASO E

1- Se usa la forma de cuadratura en:

$$L(X_k) = \prod_{i=1}^{n} P_i(X_k)^{u_{ij}} Q_i(X_k)^{1-u_{ij}}$$

y estimaciones provisionales de los parámetros para calcular la verosimilitud del puntaje de cada individuo en cada uno de los nodos de la cuadratura.

2- Se usa la ecuación:

$$P(X_k|U_j, \tau, \xi) = \frac{L(X_k)A(X_k)}{\sum_{k=1}^{q} L(X_k)A(X_k)}$$

y los pesos de cuadratura $A(X_k)$ en cada uno de los nodos para calcular la probabilidad posterior que la habilidad del j-ésimo sujeto es X_k.

3- Se usan las ecuaciones:

$$\overline{f_{ik}} = \sum_{j=1}^{N} [\frac{L(X_k)A(X_k)}{\sum_{k=1}^{q} L(X_k)A(X_k)}]$$

$$\overline{r_{ik}} = \sum_{j=1}^{N} [\frac{u_{ij}L(X_k)A(X_k)}{\sum_{k=1}^{q} L(X_k)A(X_k)}]$$

para calcular $\overline{f_{ik}}$ y $\overline{r_{ik}}$ para cada ítem en cada nodo.

PASO M

Se resuelven las ecuaciones:

$$\frac{\partial \ln L}{\partial a_i} = (1-c_i)\sum_{k=1}^{q}(X_k - b_i)[\overline{r_{ik}} - \overline{f_{ik}}P_i(X_k)]w_{ij} = 0$$

$$\frac{\partial \ln L}{\partial b_i} = (-a_i)(1-c_i)\sum_{k=1}^{q}[\overline{r_{ik}} - \overline{f_{ik}}P_i(X_k)]w_{ij} = 0$$

$$\frac{\partial \ln L}{\partial c_i} = (1-c_i)^{-1}\sum_{k=1}^{q}\frac{[\overline{r_{ik}} - \overline{f_{ik}}P_i(X_k)]}{P_i(X_k)} = 0$$

para hallar las estimaciones de los parámetros de los ítems usando los datos artificiales $\overline{f_{ik}}$ y $\overline{r_{ik}}$. Estas ecuaciones se resuelven vía el método de Newton Raphson-Fisher.

Como criterio de parada se analiza si la verosimilitud permanece incambiada respecto al ciclo anterior, entonces el proceso finaliza, en caso contrario se repiten los pasos E y M.

2.19. Estimación bayesiana

Supongamos que se desea estimar los parámetros de los ítems cuando los parámetros de habilidad son conocidos. Consideremos la distribución a priori de los parámetros del i-ésimo ítem, sea esta $g(\xi, \eta)$. Asumamos además que todos los ítems tienen la misma distribución a priori. Sea U la matriz Nxn de respuestas a los ítems y sea $L(U|\xi, \theta)$ su función de verosimilitud condicional. La distribución a posteriori entonces es:

$$g(\xi|U, \theta, \eta) \propto L(U|\xi, \theta)g(\xi, \eta)$$

donde el signo de proporcionalidad toma el sentido usual de la estadística bayesiana y

$$L(U|\xi, \theta) = \prod_{j=1}^{N}\prod_{i=1}^{n} P_i(\theta_j)^{u_{ij}}Q_i(\theta_j)^{1-u_{ij}} = \prod_{j=1}^{N} P(U_j|\theta_j, \xi)$$

Mislevy en 1986 presentó un procedimiento de estimación bayesiano que es una extensión de la solución marginal de Bock y Aitkin (1981). Emplea un procedimiento en dos etapas en el cual la información a priori es especificada en un modo jerárquico. Comienza con la distribución conjunta a priori de todos los parámetros:

$$g(\theta, \tau, \xi, \eta) = \prod_{j=1}^{N} g(\theta_j|\tau) \prod_{i=1}^{n} g(\xi_i|\eta) g(\tau) g(\eta)$$

Si se desea hacer inferencia acerca de los parámetros θ, τ, ξ y η. La distribución a posteriori aplicando el teorema de Bayes es:

$$g(\theta, \tau, \xi, \eta|U) \propto L(U|\theta, \xi) g(\theta|\tau) g(\tau) g(\xi|\eta) g(\eta)$$

Esta ecuación se marginaliza respecto a θ y Mislevy sugiere que también se haga respecto a η, pues los considera fuera de interés y los trata como parámetros de ruido. A partir de esto:

$$g(\xi, \tau|U) \propto \int\int L(U|\theta, \xi) g(\theta|\tau) g(\tau) g(\xi|\eta) g(\eta) d\theta d\eta \propto L(U|\xi, \tau) g(\xi) g(\tau)$$

.

donde $L(U|\xi, \tau)$ es la verosimilitud marginal.

2.19.1. Estimación marginal bayesiana de los parámetros del ítem en Pc-Bilog

Se asume que los ítems son independientes, lo cual permite en el proceso de estimación proceder ítem por ítem.

Se deriva el logaritmo de la ecuación anterior con respecto a los parámetros de los ítems.

$$\frac{\partial \ln L(U|\xi, \tau)}{\partial v_i} + \frac{\partial \ln g(\xi)}{\partial v_i} + \frac{\partial g(\tau)}{\partial v_i} = 0$$

donde v_i representa cualquier parámetro.

Estas últimas toman el nombre de ecuaciones modales marginales de Bayes.

Es claro que $\frac{\partial g(\tau)}{\partial v_i} = 0$, luego las ecuaciones quedan:

$$\frac{\partial \ln L(U|\xi, \tau)}{\partial v_i} + \frac{\partial \ln g(\xi)}{\partial v_i} = 0$$

Veremos ahora la solución de esto siguiendo a Baker (1992) del modo más general posible, o sea con el modelo de tres parámetros. Tenemos:

$$L = L(U|\xi,\tau) = \prod_{j=1}^{N} P(U_j|\xi,\tau)$$

y

$$P(U_j|\xi) = \int P(U|\theta,\xi)g(\theta|\tau)d\theta$$

Tomando logaritmos y derivando

$$\frac{\partial \ln L(U|\xi,\tau)}{\partial v_i} = \frac{\partial \sum_{j=1}^{N} \ln P(U_j|\xi,\tau)}{\partial v_i}$$

El Pc-Bilog emplea la transformación $\alpha_i = \ln a_i$ de donde $a_i = e^{\alpha_i}$

Luego el modelo de tres parámetros queda:

$$P_i(\theta_j) = c_i + (1-c_i)[\frac{exp\{(\theta_j - b_i)e^{\alpha_i}\}}{1+exp\{(\theta_j - b_i)e^{\alpha_i}\}}] = c_i + (1-c_i)P_i^*(\theta_j)$$

Así pues derivando con respecto a α_i

$$\frac{\partial \ln L}{\partial \alpha_i} = e^{\alpha_i}(1-c_i)\sum_{j=1}^{N}\int \frac{u_{ij} - P_i(\theta_j)}{P_i(\theta_j)Q_i(\theta_j)}(\theta_j - b_i)P_i^*(\theta_j)Q_i^*(\theta_j)[P(\theta_j|U_j,\xi,\tau)]d\theta$$

Sea $w_{ij} = \frac{P_i^*(\theta_j)Q_i^*(\theta_j)}{P_i(\theta_j)Q_i(\theta_j)}$

pues,

$$\frac{\partial \ln L}{\partial \alpha_i} = e^{\alpha_i}(1-c_i)\sum_{j=1}^{N}\int (u_{ij} - P_i(\theta_j))w_{ij}(\theta_j - b_i)[P(\theta_j|U_j,\xi,\tau)]d\theta$$

Igualmente,

$$\frac{\partial \ln L}{\partial b_i} = -e^{\alpha_i}(1-c_i)\sum_{j=1}^{N}\int (u_{ij} - P_i(\theta_j))w_{ij}[P(\theta_j|U_j,\xi,\tau)]d\theta$$

y

$$\frac{\partial \ln L}{\partial c_i} = (1-c_i)^{-1}\sum_{j=1}^{N}\int \frac{(u_{ij} - P_i(\theta_j))}{P_i(\theta_j)}[P(\theta_j|U_j,\xi,\tau)]d\theta$$

Como nuevamente se tiene una integral se empleará el método de cuadratura, los detalles son los mismos que los ya vistos. Igual que antes, se tiene el número esperado de sujetos y el número esperado de respuestas correctas:

$$\overline{f_{ik}} = \sum_{j=1}^{N} [\frac{L(X_k)A(X_k)}{\sum_{k=1}^{q} L(X_k)A(X_k)}] = \sum_{j=1}^{N} P(X_k|U_j, \xi, \tau)$$

donde

$$L(X_k) = \pi_i^n P_i(\theta_j)^{u_{ij}} Q_i(\theta_j)^{1-u_{ij}}$$

y

$$\overline{r_{ik}} = \sum_{j=1}^{N} [\frac{u_{ij} L(X_k)A(X_k)}{\sum_{k=1}^{q} L(X_k)A(X_k)}] = \sum_{j=1}^{N} u_{ij} P(X_k|U_j, \xi, \tau)$$

con $k = 1, 2, ..., q$

Reescribiendo las ecuaciones usando los datos artificiales tenemos:

$$\frac{\partial \ln L}{\partial \alpha_i} = e^{\alpha_i}(1 - c_i) \sum_{k=1}^{q} (\overline{r_{ik}} - \overline{f_{ik}} P_i(X_k)) w_{ik}(X_k - b_i)$$

$$\frac{\partial \ln L}{\partial b_i} = -e^{\alpha_i}(1 - c_i) \sum_{k=1}^{q} (\overline{r_{ik}} - \overline{f_{ik}} P_i(X_k)) w_{ik}$$

$$\frac{\partial \ln L}{\partial c_i} = (1 - c_i)^{-1} \sum_{k=1}^{q} \frac{(\overline{r_{ik}} - \overline{f_{ik}} P_i(X_k))}{P_i(X_k)}$$

Respecto de las distribuciones a priori de los parámetros, Pc-Bilog asume que cada a_i tiene distribución lognormal sobre $[0, +\infty]$ de donde α_i tiene distribución normal con densidad proporcional a $\exp[-0{,}5((\alpha_i - \mu_\alpha)/\sigma_\alpha)^2]$.

En Pc-Bilog los valores por defecto de los hiperparámetros son $\mu_\alpha = 0$ y $\sigma_\alpha = {,}5$ Como α_i es normal entonces $\frac{\partial \ln g(\alpha)}{\partial \alpha_i} = -\frac{(\alpha_i - \mu_\alpha)}{\sigma_\alpha^2}$

Finalmente con los resultados previos tenemos:

$$L_1 = \frac{\partial \ln[L(U|\xi, \tau) g(\xi)]}{\partial \alpha_i} = e^{\alpha_i}(1 - c_i) \sum_{k=1}^{q} (\overline{r_{ik}} - \overline{f_{ik}} P_i(X_k)) w_{ik}(X_k - b_i) - \frac{(\alpha_i - \mu_\alpha)}{\sigma_\alpha^2} = 0$$

Para b_i se utiliza una distribución normal, luego tenemos

$$L_2 = -e^{\alpha_i}(1-c_i)\sum_{k=1}^{q}(\overline{r_{ik}} - \overline{f_{ik}}P_i(X_k))w_{ik} - \frac{(b_i-\mu_b)}{\sigma_b^2} = 0$$

El parámetro c_i representa una probabilidad. Swaminathan y Gifford (1986) proponen se utilice una distribución beta:

$$g(c) = \frac{[(A-2)+(B-2)]!}{(A-2)!(B-2)!}c_i^{A-2}(1-c_i)^{B-2}$$

obteniéndose:

$$\frac{\partial \ln g(c)}{\partial c_i} = (A-2)/c_i - (B-2)/(1-c_i)$$

así pues,

$$L_3 = (1-c_i)^{-1}\sum_{k=1}^{q}\frac{(\overline{r_{ik}} - \overline{f_{ik}}P_i(X_k))}{P_i(X_k)} + (A-2)/c_i - (B-2)/(1-c_i) = 0$$

2.19.1.1. Estimación vía EM

En el paso E se calculan los datos artificiales, en el paso M los estimadores bayesianos se calculan sobre cada ítem a la vez. Como las ecuaciones son no lineales en los parámetros, se utiliza un procedimiento de Newton Raphson-Fisher. Esto se repite hasta que se satisface el criterio de convergencia.

Operando se llega a

$$\Lambda_{11} = -(e^{\alpha_i})^2\sum_{k=1}^{q}\overline{f_{ik}}(X_k-b_i)^2[\frac{P_i(X_k))-c_i}{1-c_i}]^2\frac{Q_i(X_k)}{P_i(X_k)} - \frac{1}{\sigma_\alpha^2}$$

$$\Lambda_{22} = -(e^{\alpha_i})^2\sum_{k=1}^{q}\overline{f_{ik}}[\frac{P_i(X_k))-c_i}{1-c_i}]^2\frac{Q_i(X_k)}{P_i(X_k)} - \frac{1}{\sigma_b^2}$$

$$\Lambda_{33} = -\sum_{k=1}^{q} \frac{\overline{f_{ik}}}{(1-c_i)^2} \frac{Q_i(X_k)}{P_i(X_k)} - \frac{A-2}{c_i^2} - \frac{B-2}{(1-c_i)^2}$$

$$\Lambda_{12} = \Lambda_{21} = (e^{\alpha_i})^2 \sum_{k=1}^{q} \overline{f_{ik}}(X_k - b_i)[\frac{P_i(X_k)) - c_i}{1-c_i}]^2 \frac{Q_i(X_k)}{P_i(X_k)}$$

$$\Lambda_{13} = \Lambda_{31} = -e^{\alpha_i} \sum_{k=1}^{q} \overline{f_{ik}}(X_k - b_i) \frac{P_i(X_k)) - c_i}{(1-c_i)^2} \frac{Q_i(X_k)}{P_i(X_k)}$$

$$\Lambda_{23} = \Lambda_{32} = e^{\alpha_i} \sum_{k=1}^{q} \overline{f_{ik}} \frac{P_i(X_k)) - c_i}{(1-c_i)^2} \frac{Q_i(X_k)}{P_i(X_k)}$$

.

Finalmente el proceso iterativo queda:

$$\begin{pmatrix} \widehat{\alpha} \\ \widehat{b} \\ \widehat{c} \end{pmatrix}_{t+1} = \begin{pmatrix} \widehat{\alpha} \\ \widehat{b} \\ \widehat{c} \end{pmatrix}_{t} - \begin{pmatrix} \Lambda_{11} & \Lambda_{12} & \Lambda_{13} \\ \Lambda_{21} & \Lambda_{22} & \Lambda_{23} \\ \Lambda_{31} & \Lambda_{32} & \Lambda_{33} \end{pmatrix}_{t}^{-1} \begin{pmatrix} L_1 \\ L_2 \\ L_3 \end{pmatrix}_{t}$$

2.19.2. Estimación bayesiana del rasgo

2.19.2.1. Estimación modal bayesiana

La estimación modal bayesiana o MAP (máximo a posteriori) del rasgo, se basa en la relación:

$$g(\theta_j|U_j, \xi) \propto L(U_j|\theta_j, \xi)g(\theta)$$

Se asume que $g(\theta)$ es $N(\mu_\theta, \sigma_\theta^2)$

Tomando logaritmos

$$\ln g(\theta_j|U_j,\xi) \propto \ln L(U_j|\theta_j,\xi) + \ln g(\theta)$$

La verosimilitud es

$$L(U_j|\theta_j,\xi) = \prod_{i=1}^{n} P_i(\theta_j)^{u_{ij}} Q_i(\theta_j)^{1-u_{ij}}$$

Para obtener el estimador MAP se utiliza el método de Fisher.

Operando,

$$L_\theta = \sum_{i=1}^{n} a_i [\frac{P_i(\theta_j) - c_i}{P_i(\theta_j)(1-c_i)}](u_{ij} - P_i(\theta_j)) - \frac{\theta_j - \mu_\theta}{\sigma_\theta^2}$$

La derivada segunda es:

$$L_{\theta\theta} = \sum_{i=1}^{n} a_i^2 [\frac{P_i(\theta_j) - c_i}{(1-c_i)^2}] \frac{Q_i(\theta_j)}{P_i(\theta_j)} \frac{(u_{ij}c_i - P_i(\theta_j))^2}{P_i(\theta_j)} - \frac{1}{\sigma_\theta^2}$$

Tomando esperanza:

$$\Lambda_{\theta\theta} = -\sum_{i=1}^{n} a_i^2 [\frac{P_i(\theta_j) - c_i}{(1-c_i)}]^2 \frac{Q_i(\theta_j)}{P_i(\theta_j)} - \frac{1}{\sigma_\theta^2}$$

Así el proceso iterativo queda:

$$[\widehat{\theta}_j]_{t+1} = [\widehat{\theta}_j]_t - [\Lambda_{\theta\theta}]_t^{-1} [\Lambda_\theta]_t$$

2.19.2.2. Estimación bayesiana EAP

Se basa en

$$g(\theta_j|U_j,\xi) = \frac{P(U=U_j|\theta_j,\xi)g(\theta)}{P(U=U_j)}$$

Debido a la independencia local:

$$P(U=U_j|\theta_j,\xi) = \prod_{i=1}^{n} P_i(\theta_j)^{u_{ij}} Q_i(\theta_j)^{1-u_{ij}}$$

y

$$P(U=U_j) = \int p(U=U_j|\theta)g(\theta)d\theta$$

luego la esperanza:

$$E(\theta_j|U_j,\xi) = \frac{\int \theta_j \prod_{i=1}^{n} P_i(\theta_j)^{u_{ij}} Q_i(\theta_j)^{1-u_{ij}} g(\theta)d\theta}{\int \prod_{i=1}^{n} P_i(\theta_j)^{u_{ij}} Q_i(\theta_j)^{1-u_{ij}} g(\theta)d\theta}$$

Como la ecuación tiene integrales se utiliza la cuadratura de Hermite-Gauss:

$$E(\theta_j|U_j,\xi) = \overline{\theta_j} = \frac{\sum_{k=1}^{q} X_k L(X_k) A(X_k)}{\sum_{k=1}^{q} L(X_k) A(X_k)}$$

Esta ecuación tiene la interesante característica de que no es iterativa.

2.20. Comprobación de los modelos

Muñiz (1997) nos da una serie de pasos para la selección y confirmación de un modelo dicotómico. Estos son:

1. Definición rigurosa de la variable que se pretende evaluar;

2. Elaboración de los ítems destinados a medir la variable;

3. Aplicación de los ítems a una muestra amplia de personas pertenecientes a la población en la que se utilizará el futuro test y cálculo de los índices clásicos de los ítems;

4. Comprobación de la unidimensionalidad de los ítems;

5. Elección de uno de los modelos de TRI para ítems dicotómicos;

6. Estimación de los parámetros de los ítems y del rasgo de cada sujeto;

7. Comprobación de que el modelo se ajusta a los datos.

Dedicaremos esta sección a analizar el ajuste del modelo a los datos. Una vez estimados los parámetros de los ítems para comprobar el ajuste, se procede como comúnmente se hace, analizando la discrepancia entre los valores pronosticados y los observados. Según Hambleton, Swaminathan y Rogers en (1990) el ajuste debe estudiarse al menos desde tres perspectivas:

a) Evaluando el ajuste entre los valores observados y los esperados bajo el modelo estimado;

b) Evaluando si los datos observados satisfacen los supuestos del modelo que interesa ajustar;

c) Evaluando si se cumplen las ventajas derivadas del modelo, por ejemplo la invarianza de los parámetros de los ítems o de los parámetros de habilidad.

Rost y von Davier (1994) clasifican los índices de bondad de ajuste en

1-Pruebas basadas en chi-cuadrado;

2-Pruebas basadas en verosimilitud;

3-Análisis de residuos.

2.20.1. Pruebas basadas en chi-cuadrado

Para aplicar estos estadísticos deben agruparse las puntuaciones de rasgo estimadas en K intervalos. Se deben seleccionar los intervalos de modo que su amplitud no sea excesiva, ya que podría suceder que los sujetos incluidos en cada intervalo no sean homogéneos en

la habilidad medida. Al mismo tiempo cada intervalo debe incluir un número suficiente de sujetos, ya que los estadísticos basados en χ^2 no funcionan bien si alguno de los intervalos tiene frecuencia nula. Se compara entonces la proporción observada de éxitos con la proporción esperada en cada intervalo. Si las proporciones observadas difieren poco de las esperadas se concluirá que hay un buen ajuste del modelo a los datos en ese ítem. Luego nos interesa obtener índices de esta discrepancia y su distribución a los efectos de poder probar la hipótesis nula que la discrepancia es 0, o sea que el modelo ajusta bien.

Estadístico de Bock (1972)

Sea o_{ik} la proporción observada de respuestas correctas en el intervalo k para el ítem i y e_{ik} la probabilidad teórica de una respuesta correcta en el intervalo k para el ítem i. Para obtener e_{ik} calculamos mediante el modelo la probabilidad en el valor de rasgo obtenido como la mediana de los valores de habilidad que caen en el intervalo k. Notemos además por N_k el número de sujetos en el intervalo k. El estadístico propuesto por Bock es:

$$Q_{Bi} = \sum_{k=1}^{K} \frac{N_k(o_{ik} - e_{ik})^2}{e_{ik}(1 - e_{ik})} \tag{2.118}$$

Este estadístico sigue una distribución asintótica χ^2_{K-m} donde m es el número de parámetros del modelo.

Estadístico de Yen (1981)

Yen propuso la siguiente modificación al estadístico Q_B:

$$Q_{1i} = \sum_{k=1}^{10} \frac{N_k(o_{ik} - e_{ik})^2}{e_{ik}(1 - e_{ik})} \tag{2.119}$$

donde ahora hay 10 intervalos y

$$e_{ik} = \frac{1}{N_k} \sum_{j \in k} P_i(\widehat{\theta}_j)$$

Este estadístico sigue una distribución χ^2_{10-m} donde m es el número de parámetros del modelo.

Estadístico de Wright y Panchapakesan (1969)

Wright y Panchapakesan propusieron un estadístico que se restringe a modelos de un parámetro.

$$Q_{2i} = \sum_{k=1}^{K} \frac{N_k(o_{ik} - e_{ik})^2}{e_{ik}(1 - e_{ik})} \tag{2.120}$$

donde los intervalos se construyen de modo que no existan intervalos con frecuencias nulas, cada intervalo contenga sujetos con el mismo número de respuestas correctas y se eliminan los sujetos que hayan acertado o fallado todos los ítems. Este estadístico sigue una distribución χ^2_{K-1}. También proponen un estadístico que es una generalización del anterior para el caso de n ítems:

$$Q_{2T} = \sum_{k=1}^{K} \sum_{i=1}^{n} \frac{N_k(o_{ik} - e_{ik})^2}{e_{ik}(1 - e_{ik})} \tag{2.121}$$

el cual tiene una distribución $\chi^2_{(K-1)(n-1)}$

Estadístico de Wright y Mead (1978)

Estos autores proponen para el siguiente estadístico apropiado para el modelo de Rasch, el cual está implementado en el programa BICAL:

$$Q_{3i} = \frac{1}{K} \sum_{k=1}^{K} \frac{N_k(o_{ik} - e_{ik})^2}{e_{ik}(1 - e_{ik}) - \sigma^2_{Pk}} \tag{2.122}$$

donde

$$\sigma^2_{Pk} = \frac{1}{N_k} \sum_{j \in k} (P_i(\widehat{\theta_j}) - e_{ik})^2$$

La aplicación de este estadístico supone: dividir a los sujetos de tal forma que se aproximen a una distribución uniforme, que el número de intervalos sea menor o igual a 6 y que el número de sujetos en cada intervalo sea mayor o igual a 15.

Estadístico de Elliot, Murray y Saunders (1977)

También para utilizar con el modelo de Rasch se puede aplicar:

$$Q_{4i} = \sum_{k=1}^{K} \frac{N_k(o_{ik} - e_{ik})^2}{e_{ik}} \tag{2.123}$$

que sigue una distribución χ^2_{K-1}.

Acá cada intervalo está formado por sujetos que obtuvieron la misma puntuación en el test. Además se debe cumplir para cada intervalo $N_k e_{ik} \geq 5$ (esto para cumplir con las condiciones del test chi-cuadrado) y si alguno no lo cumple se fusionará con sus adyacentes hasta lograrlo. Se eliminan también los sujetos que acertaron o fallaron todos los ítems. Los estadísticos Q_3 y Q_4 fueron extendidos por Yen a modelos de dos y tres parámetros, de forma natural.

Limitaciones

Las principales limitaciones y problemas de estos estadísticos derivan de:

1- Dependen fuertemente del tamaño muestral. Al aumentar la cantidad de sujetos aumenta la tasa de error tipo I y viceversa. Esto resulta más grave en los modelos de dos y tres parámetros pues para su estimación se requiere un tamaño muestral grande.

2- La cantidad de intervalos seleccionados, ya que es necesario que estos representen adecuadamente el comportamiento de los datos.

3- Son sensibles a la presencia de *outliers*.

4- No son fácilmente generalizables a los modelos de respuesta politómica.

2.20.2. Pruebas basadas en la verosimilitud

Estadístico de Mckinley y Mills (1985)

Proponen un estadístico de razón de verosimilitud. Consideran, igual que el estadístico de Yen, 10 intervalos que contengan el mismo número de sujetos. Construyen una tabla de contingencia 10x2, donde las dos columnas representan las posibles respuestas del ítem.

Proponen:

$$RL_i = 2\sum_{k=1}^{20} o_{ik} \ln \frac{o_{ik}}{e_{ik}} \tag{2.124}$$

que sigue una distribución χ^2_{10-m}. Este estadístico en el caso de un parámetro es más potente que el de Bock independientemente del tamaño muestral.

Estadístico de Mislevy y Bock (1990)

Proponen el estadístico RL(MB) que se distribuye χ^2_K y está dado por:

$$RL(MB_i) = 2\sum_{k=1}^{K}(o_{ik} \ln \frac{o_{ik}}{N_k e_{ik}} - (N_k - o_{ik}) \ln \frac{N_k o_{ik}}{N_k(1 - e_{ik})}) \tag{2.125}$$

Estadístico de Reise (1990)

Utiliza una técnica denominada bondad de ajuste de personas, Reise adapta el índice de verosimilitud L propuesto por Levine y Rubin. En el estudio de ajuste de personas la función L viene dada por:

$$L_j = \sum_{i=1}^{n}(u_i \ln P_i(\theta_j) + (1 - u_i) \ln Q_i(\theta_j)) \tag{2.126}$$

donde u_i es la respuesta (0-1) del sujeto j y $Q(\theta) = 1 - P()\theta$.
Un valor alto de L indica un buen ajuste del sujeto. Se tiene que:

$$E(L_j) = \sum_{i=1}^{n} (P_i(\theta_j) \ln P_i(\theta_j) + Q_i \ln Q_i(\theta_j))$$

y

$$\sigma_{Lj}^2 = \sum_{i=1}^{n} P_i(\theta_j) Q_i(\theta_j) [\ln \frac{P_i(\theta_j)}{Q_i(\theta_j)}]^2$$

de donde

$$L_{Zj} = \frac{L_j - E(L_j)}{\sigma_{Lj}} \sim N(0,1)$$

Un valor próximo a 0 es adecuado y si es positivo indica una mayor adecuación a lo esperado.

En el estudio de ajuste de los ítems el estadístico viene dado por:

$$L_i = \sum_{j=1}^{N} (u_i \ln P_i(\theta_j) + (1 - u_i) \ln Q_i(\theta_j)) \tag{2.127}$$

Si se utilizan las puntuaciones divididas en K intervalos, toma la forma:

$$L_i = \sum_{k=1}^{K} (A_{ik} \ln P_i(\theta_j) + (N_k - A_{ik}) \ln Q_i(\theta_j))$$

donde A_{ik} es la cantidad de individuos que aciertan el ítem en el intervalo k.

Igualmente que para los sujetos ambos índices se pueden estandarizar y su ecuación es igual a la ya vista, con la única salvedad que ahora la suma es sobre los individuos y no sobre los ítems.

2.20.3. Análisis de residuos

Residual de Wright y Stone (1979)

Ellos proponen el siguiente residual estandarizado:

$$z_{ij} = \frac{[u_i j - P_i(\theta_j)]}{[P_i(\theta_j)(1 - P_i(\theta_j))]}$$

donde u_{ij} es la respuesta del sujeto j al ítem i y $P_i(\theta_j)$ la probabilidad que el sujeto j acierte el ítem i calculada a partir del modelo ajustado.

Se define el residual cuadrático medio para el ítem i como:

$$z_i^2 = \frac{\sum_{j=1}^{N} z_{ij}^2}{N-1}$$

y el residual cuadrático medio para el sujeto j:

$$z_j^2 = \frac{\sum_{i=1}^{n} z_{ij}^2}{n-1}$$

z_i tiene distribución F con $(N-1, \infty)$ y z_j tiene distribución F con $(n-1, \infty)$.

Los autores utilizan la transformación logarítmica para obtener estadísticos con distribución normal.

Los estadísticos quedan entonces:

$$t_i = [\log(z_i^2) + z_i^2 - 1]\frac{\sqrt{N-1}}{\sqrt{8}} \tag{2.128}$$

y

$$t_j = [\log(z_j^2) + z_j^2 - 1]\frac{\sqrt{n-1}}{\sqrt{8}} \tag{2.129}$$

respectivamente.

La dependencia del tamaño muestral de los residuos cuadráticos llevan a los autores a proponer otras medidas como el índice t-total y el t-interítems.

El índice t-total se expresa como:

$$t(T) = \frac{\sum_{i=1}^{n}[u_{ij} - P_i(\theta_j)]^2}{\sum_{i=1}^{n} P_i(\theta_j)Q_i(\theta_j)}$$

De este estadístico se desconoce la distribución, así pues los autores recomiendan se revisen aquellos ítems o sujetos con $t(T) > 2$.

El estadístico t-interítems se expresa como:

$$t(I) = \sum_{k=1}^{K} \frac{[u_{ik} - \sum_{j \in k} N_j P_i(\theta_j)]^2}{\sum_{j \in k} N_j P_i(\theta_j) Q_i(\theta_j)} [\frac{m}{(K-1)(m-1)}]$$

donde m es el número de parámetros del modelo.

Este estadístico sigue una distribución asintótica χ^2 con (K-1)(m-1) grados de libertad. Rogers y Hattie estudiaron estos índices para distintos modelos y encontraron que en general resultan inapropiados para los modelos de Rasch.

Residual de Hambleton, Swaminathan y Rogers (1991)

Definen el residual como $r_{ik} = o_{ik} - e_{ik}$ y definen el residual estandarizado como:

$$z_{rik} = \frac{o_{ik} - e_{ik}}{\sqrt{e_{ik}(1 - e_{ik})/N_k}}$$

La proporción esperada de aciertos se debe calcular para el nivel de habilidad medio o mediano de cada intervalo.

Residual de Ludlow (1986)

Para modelos de respuesta politómica Ludlow (1986) propuso el residual de una observación como:

$$Z_{ij} = \frac{X_{ij} - E_{ij}}{(V_{ij})^{\frac{1}{2}}} \tag{2.130}$$

donde

$$E_{ij} = \sum_{c=0}^{C} cP_{ic}(\theta_j)$$

siendo C el número de categorías del ítem y $P_{ic}(\theta_j)$ la probabilidad de éxito correspondiente a la categoría c. V_{ij} es la varianza asociada a la respuesta esperada que viene dada por:

$$V_{ij} = \sum_{c=0}^{C} (c - E_{ij})^2 P_{ic}(\theta_j)$$

El residual Z_{ij} tiene distribución normal típica.

Ludlow utiliza este residual para proponer una medida global de ajuste del ítem:

$$s_i = W_i^{\frac{1}{3}}(3p_i) + (p_i/3)$$

donde W_i es:

$$W_i = \frac{\sum_{j=1}^{N} V_{ij} Z_{ij}^2}{\sum_{j=1}^{N} V_{ij}}$$

y

$$p_i^2 = \frac{\sum_{j=1}^{N} (C_i j - V_{ij}^2)}{(\sum_{j=1}^{N} V_{ij})^2}$$

donde C_{ij} es el valor de kurtosis de la puntuación X_{ij}.

Índice de Rost y von Davier (1994)

Desarrollaron a partir del modelo de crédito parcial el índice Item-Q (I_{Qi}).

Esta medida utiliza funciones simétricas y se basa en la comparación de los valores de la verosimilitud de los patrones de respuesta observados.

El índice se define como un cociente entre dos verosimilitudes condicionadas. El objetivo para un patrón de respuestas dado es situar su verosimilitud entre el valor máximo y mínimo. Para un conjunto de ítems el valor máximo de la verosimilitud será el correspondiente al patrón de respuestas perfecto al que se llamará patrón Guttman (u_G) y el valor mínimo se dará en patrón aberrante, el que se conoce como patrón anti-Guttman (u_A).

El índice I_{Qi} viene dado por:

$$I_{Qi} = \frac{\ln(RL_{i,G})}{\ln(RL_{A,G})} \tag{2.131}$$

donde:

$$\ln RL_{i,G} = \ln[P(u_i|f_{ic})] - \ln[P(u_G|f_{ic})];$$

$$\ln RL_{A,G} = \ln[P(u_A|f_{ic})] - \ln[P(u_G|f_{ic})];$$

$P(u_i|f_{ic})$ es la probabilidad de obtener el patrón u_i dado el número de respuestas en esa categoría (f_{ic}), y análogamente $P(u_G|f_{ic})$ y $P(u_A|f_{ic})$.

Este índice toma la expresión:

$$I_{Qi} = \frac{\sum_{j=1}^{N} u_{ij}\theta_j - \sum_{j=1}^{N} u_{Gj}\theta_j}{\sum_{j=1}^{N} u_{Aj}\theta_j - \sum_{j=1}^{N} u_{Gj}\theta_j} \tag{2.132}$$

El índice varía entre 0 y 1, donde 0 indica ajuste perfecto y 1 desajuste total.

El I_{Qi} también puede usarse como índice del poder de discriminación del ítem. Así pues un valor de 0 indica discriminación perfecta, el 1 también se da con discriminación perfecta pero negativa y un valor de 0.5 indica que no hay relación entre el rasgo y el ítem.

Rost y von Davier (1994) para obtener una prueba estadística del I_{Qi} trabajan con el numerador del I_{Qi} pues es el único que varía. Definen entonces:

$$I^*_{Qi} = \sum_{j=1}^{N} (u_{ij} - u_{Gj})\theta_j.$$

Se tiene que:

$$E(I^*_{Qi}) = \sum_{j=1}^{N} (\sum_{c=0}^{C} P_{ij}(c)c\theta_j - u_{Gj}\theta_j$$

y

$$V(I^*_{Qi}) = \sum_{j=1}^{N} V(u_{Gj}\theta_j))$$

Donde la varianza de un patrón Guttman, anti-Guttman u observado viene dada por:

$$V(u_{ij}\widehat{\theta_j}) = V(\widehat{\theta_j})V(u_{ij}) + E(\widehat{\theta_j})^2V(u_{ij}) + V(\widehat{\theta_j})E(u_{ij})^2$$

con

$$E(u_{ij}) = \sum_{c=0}^{C} P_{ij}(c)c$$

y

$$V(u_{ij}) = \sum_{c=0}^{C} P_{ij}(c)[c - E(c)]^2$$

A partir de esto se obtiene que:

$$Z(I_{Qi}) = \frac{(I^*_{Qi} - E((I^*_{Qi})}{\sqrt{V((I^*_{Qi})}}$$

que tiene distribución asintótica normal típica.

2.20.4. Bondad de ajuste del modelo

Queremos ahora comprobar el ajuste del modelo al test completo. Veremos algunos procedimientos para comprobar esto.

Procedimiento de Andersen (1973)

Propone un estadístico de razón de verosimilitud para el modelo de un parámetro.

$$RL = \frac{L(\widehat{b})}{\sum_{k=1}^{K} L_k(\widehat{b_b})} \tag{2.133}$$

donde $\widehat{b}$ es el vector de parámetros de dificultad estimados, utilizando el grupo total de sujetos, $\widehat{b_k}$ el vector estimado a partir de los sujetos en el intervalo k, y L indica las funciones de verosimilitud.

Cuando el modelo de un parámetro se ajusta, los parámetros de dificultad son invariantes a través de todos los subgrupos, por lo que RL estará cercano a 1.

Para probar el ajuste utilizamos $G^2 = 2\ln(RL)$ que tiene distribución χ^2 con (K-1)(K-2) grados de libertad. Un valor de G^2 significativamente distinto de cero indica que el modelo de un parámetro no se ajusta bien a los datos.

Procedimiento de Mislevy y Bock (1990)

Proponen el estadístico de razón de verosimilitud:

$$G^2 = 2\Sigma_{k=1}^{2^n} u_k \ln \frac{u_k}{N\overline{P(x_k)}} \tag{2.134}$$

donde u_k son los patrones de respuestas correctas y $\overline{P(x_k)}$ es la probabilidad marginal del patrón de respuestas.

Este test permite probar el ajuste total en test de hasta 10 ítems. Sigue una distribución χ^2 con $2^n - mn - 1$ grados de libertad siendo m el número de parámetros.

Procedimiento de Martin-Lof (1973)

Martin-Lof (1973, citado en Wollenberg, 1988 y en Rost y von Davier, 1994) propone una prueba basada en χ^2.

$$T = \sum_{k=1}^{K} (d_k)'(V_k)^{-1}(d_k)$$

donde V_k es la matriz de varianza covarianza condicionada sobre los valores esperados y d_k el vector de diferencias entre los valores observados y esperados.

$d_k = O_k - E_k$ y $E_{ki} = \frac{N_k \widehat{\epsilon_i} \widehat{Y}_{k-1}^i}{\widehat{Y_k}}$

donde $\widehat{\epsilon_i} = \exp(-b)$ en tanto $\widehat{Y_k}$ y $\widehat{Y}_{k-1}^i$ son funciones simétricas derivadas en la estimación de máxima verosimilitud condicional.

Procedimiento de Wollenberg (1982)

Wollenberg simplifica el estadístico T obteniendo:

$$Q_{W1} = \frac{K-1}{K} \sum_{i=1}^{n} q_i$$

el cual sigue una distribución χ^2 con (K-1)(K-2) grados de libertad y donde

$$q_i = \sum_{i=1}^{K-1} \left(\frac{[O_{ik} - E_{ik}]^2}{E_{ik}} + \frac{[O_{ik} - E_{ik}]^2}{n_k - E_{ik}} \right)$$

.

Procedimiento de Fischer y Scheilblechner (1970)

En este procedimiento se divide el grupo en dos subgrupos. En cada uno de ellos se estiman los parámetros de los ítems y se comparan mediante:

$$S_i = \frac{\widehat{b}_i^{(1)} - \widehat{b}_i^{(2)}}{[S^2_{\widehat{b}_i^{(1)}} - S^2_{\widehat{b}_i^{(2)}}]^{1/2}} \tag{2.135}$$

Luego se utiliza el estadístico

$$S = \sum_{i=1}^{n} S_i^2 \tag{2.136}$$

que sigue una distribución χ^2 con n-1 grados de libertad.

Procedimiento de Reise y Waller (1990)

Proponen utilizar:

$$Q(B) = \sum_{i=1}^{n} Q_{Bi}$$

donde Q_{Bi} es el estadístico de Bock para el ítem i.

Para el modelo de un parámetro se puede medir el ajuste global mediante:

$$Q(2) = \sum_{i=1}^{n} Q_{2i}^2$$

donde Q_{2i}^2 es el estadístico propuesto por Wright y Panchapakesan (1969).
Este sigue una distribución χ^2 con (n-1)(K-1) grados de libertad.

Procedimiento de Yen (1981)

Yen propone seguir los siguientes pasos:

1) Calibrar el test con los modelos de un, dos y tres parámetros.

2) Obtener para cada ítem bajo cada modelo el estadístico Q_1.

3) Calcular los valores medios de Q_1 en los tres modelos.

Si el valor medio en los tres modelos es aproximadamente igual, entonces el modelo de un parámetro es el adecuado. Si es sustancialmente mayor se descarta este modelo y se procede al paso 4:

4) Se eligirá el modelo 2-p si:

las estimaciones de rasgo en los modelos de dos y tres parámetros son linealmente dependientes;

los parámetros de discriminación de ambos modelos correlacionan fuertemente;

las correlaciones entre a y b son similares en ambos modelos.
5) Se eligirá el modelo 3-p si:

las estimaciones de rasgo en los modelos de dos y tres parámetros son linealmente independientes; los parámetros de discriminación de ambos modelos tienen baja correlación; las correlaciones entre a y b son menores en el modelo 2-p que en el 3-p.

2.21. Modelos no paramétricos clásicos

Las primeras referencias a los modelos no paramétricos fueron de Mokken (1971), Henning (1976), Niemöller y Van Schuur (1983), Mokken y Lewis (1982), Sijtsma (1988) y Giampaglia (1990). Más recientemente hay trabajos de Mokken (1997), Sijtsma(1998,2001), Molenaar y Sijtsma (2000,2002), Junker (2000) y Junker y Sijtsma (2001).

El tratamiento matemático elaborado del modelo de monotonía homogéneo lo podemos encontrar en Holland y Rosenbaum (1986), Holland (1990), Stout (1987,1990), Junker (1993) y Ellis y Junker (1997).

Una comparación entre los modelos paramétricos y no paramétricos son comparados en Meijer, Sijtsma y Smid (1990), y en De Koning, Sijstma y Hamers (2001).

Douglas, Kim, Habing y Gao (1998) dan métodos para investigar la independencia local, en tanto métodos para investigar CCI no decrecientes desde un punto de vista práctico son dados en Ramsay (1991), Molenaar y Sijtsma (2000) y Douglas y Cohen (2001).

Consideremos una muestra aleatoria de N sujetos tomados de una población que responden a n ítems dicotómicos.

Se supondrá que se cumplen los supuestos de unidimensionalidad, independencia local y que las CCI cumplen la propiedad de monotonía homogénea respecto al rasgo es decir que para todo i y rasgos θ_l y θ_m entonces $\theta_l \leq \theta_m$ si y solo si $P_i(\theta_l) \leq P_i(\theta_m)$.

Esta propiedad es equivalente a la llamada propiedad de orden similar que establece que un conjunto de CCI ordena similarmente a un conjunto de N sujetos con rasgos $\theta_1, \theta_2, \ldots, \theta_N$ si para algún i $P_i(\theta_1) \leq P_i(\theta_2) \leq \ldots \leq P_i(\theta_N)$ entonces para todo ítem k $P_k(\theta_1) \leq P_k(\theta_2) \leq \ldots \leq P_k(\theta_N)$.

Estas propiedades son suficientes para tener un orden conjunto de los sujetos respecto a un rasgo latente, sin embargo no son suficientes para un orden uniforme de las dificultades de lo ítems. Para esto se requerirá una nueva condición llamada de doble monotonía o monotonía homogénea fuerte.

Formalmente un conjunto de ítems homogéneamente monótonos satisfacen la propiedad de doble monotonía respecto a θ, si para todo par de ítems i y j que para algún θ_0 se cumple que $P_i(\theta_0) \leq P_j(\theta_0)$ entonces para todo θ $P_i(\theta) \leq P_j(\theta)$.

Es decir, las n curvas no se intersectan y las CCI se pueden ordenar de tal manera que $P_1(\theta) \leq P_2(\theta) \leq \ldots \leq P_n(\theta)$ para todo θ. Esto nos permite ordenar los ítems según dificultad.

2.21.1. Modelo monótono homogéneo

Este modelo se soporta en las hipótesis de independencia local y monotonía homogénea.

Sea $X_+ = \sum_{i=1}^{n} X_i$ el puntaje total obtenido en el test; los sujetos son ordenados mediante este estadístico.

Este modelo mide los sujetos en una escala ordinal formalmente, sea ζ un valor arbitrario del rasgo y sean $0 \leq s < t \leq n$ dos posibles valores del puntaje total entonces $P(\theta > \zeta | X_+ = s) \leq P(\theta > \zeta | X_+ = t)$.

De la anterior se deduce que para $0 \leq s < t \leq n$ $E(\theta | X_+ = s) \leq E(\theta | X_+ = t)$ la cual implica que la media del rasgo en el subgrupo de individuos con puntaje t es mayor a la media del rasgo en el subgrupo que tiene puntaje s si $0 \leq s < t \leq n$. Podemos a partir de esto justificar el uso de $X_+ = \sum_{i=1}^{n} X_i$ para ordenar a los sujetos.

También se cumple para $\theta_1 < \theta_2$ que:

$$P(X_+ > x | \theta = \theta_1) \leq P(X_+ > x | \theta = \theta_2)$$

y

$$E(X_+ | \theta = \theta_1) \leq E(X_+ | \theta = \theta_2)$$

Teorema 2.21.1. (Mokken 1971) Si se cumple la propiedad de monotonía homogénea y la independencia local entonces la covarianza entre todos los pares de ítems es no negativa.

Demostración

Sea $F(\theta)$ la función de distribución del rasgo latente y consideremos dos ítems cualesquiera i, j.

Se tiene que:

$$P(X_i = 1) = \int P_i(\theta)\, dF(\theta)$$

y $P(X_i = 1, X_j = 1) = \int P_i(\theta)\, P_j(\theta)\, dF(\theta)$.

Luego, debido a que se trata de ítems dicotómicos, se cumple que:

$$\begin{aligned} &Cov(X_i, X_j) = P(X_i = 1, X_j = 1) - P(X_i = 1)P(X_j = 1) = \\ &\iint P_i(\theta)P_j(\theta)\, dF(\theta)\, dF(\xi) - \int P_i(\theta)\, dF(\theta) \int P_j(\xi)\, dF(\xi) = \\ &\tfrac{1}{2} \iint [P_i(\theta) - P_i(\xi)]\, [P_j(\theta) - P_j(\xi)]\; dF(\theta)\, dF(\xi) \end{aligned}$$

En la región donde $\theta > \xi$ ambos términos son no negativos debido a que CCI son monótonas y en donde $\theta < \xi$ ambos son no positivos, luego la covarianza es nonegativa.

Rosenbaum (1984) y luego Holland y Rosenbaum (1986) probaron el siguiente resultado (asociación condicional) bajo las mismas hipótesis. Si partimos el vector de respuestas u en dos subconjuntos u_1 y u_2 y consideremos f, g dos funciones monótonas no decrecientes, entonces $Cov(f(U_1), g(U_1)|h(U_2)) \geq 0$.

También se puede probar que:

a) Todo par de ítems i y j están condicionalmente asociados en todo subconjunto de sujetos con un puntaje total t en los restantes ítems.

b) Para todo ítem i, P_i es no decreciente si se toma sobre los subgrupos de sujetos ordenados en forma creciente en función del puntaje total sobre los n-1 ítems restantes.

c) Lord y Novick, (1968) prueban que la esperanza condicional (respecto de θ) del puntaje total es una función monótona no decreciente.

d) Grayson (1988) mostró que el puntaje total (condicionado a θ) tiene verosimilitud monótona o sea $\frac{P(X_+=t|\theta)}{P(X_+=s|\theta)} \quad 0 \leq s < t \leq n$ es no decreciente como función de θ, es decir, $X_+|\theta$ es un estadístico óptimo para la clasificación binaria.

2.21.2. Modelo de doble monoticidad

Este modelo incorpora adicionalmente la hipótesis de no intersección de las CCI.

En el caso de ítems dicotómicos tenemos que $E(X_i|\theta) = P_i(\theta)$.

A partir de esto se obtiene, ya que las curvas no se cortan que:

$$E(X_1|\theta) \leq E(X_2|\theta) \leq \ldots \leq E(X_n|\theta)$$

para todo θ. Se dice cuando esto ocurre que los ítems tienen un orden invariante.

Para poder establecer cuál es el orden de precedencia entre los ítems, ya que no es posible observar directamente las esperanzas condicionales, se recurre a la proporción de respuestas correctas. Para justificar esto veamos que como en nuestras suposiciones hemos considerado que el rasgo latente tiene densidad $f(\theta)$, entonces la proporción de respuestas correctas en el ítem i puede determinarse como:

$$P_i = \int_{-\infty}^{+\infty} P_i(\theta)\ f(\theta)\ d\theta = \int_{-\infty}^{+\infty} E(X_i|\theta)\ f(\theta)\ d\theta$$

Si el ítem i tiene menor o igual proporción de respuestas correctas que el ítem j entonces

$P_i - P_j = \int_{-\infty}^{+\infty} [P_i(\theta) - P_j(\theta)]\ \ f(\theta)\ \ d\theta = \int_{-\infty}^{+\infty} [E(X_i|\theta) - E(X_j|\theta)]\ \ f(\theta)\ \ d\theta \leq 0$ de donde se deduce que si $P_i \leq P_j$ y sabemos que los ítems tienen la propiedad de orden invariante entonces $E(X_i|\theta) \leq E(X_j|\theta)$ para todo θ.

Vemos así que hay dos propiedades de monotonía: una sobre las CCI y otra dada por el orden invariante.

Veamos que no asumimos nada sobre la densidad $f(\theta)$, esto implica que si tenemos dos subpoblaciones con densidades $f_1(\theta)$ y $f_2(\theta)$ el orden de la proporción de respuestas correctas en cada subpoblación es el mismo, lo cual es un resultado relevante.

Mokken (1971) prueba que para todo ítem i y todo par de ítems j y k con j<k se cumple:

$$P(X_i = 1, X_j = 1) \leq P(X_i = 1, X_k = 1)$$
$$P(X_i = 0, X_k = 0) \leq P(X_i = 0, X_j = 0)$$

Rosenbaum (1987) probó que si i es más difícil que k entonces $P(U_i = 1|X_+(u_2)) \leq P(U_k = 1|X_+(u_2))$ donde u_2 es el vector de respuestas en los n-2 ítems restantes.

2.21.3. Estimación de la habilidad

Para la estimación de las habilidades hay que recurrir a métodos indirectos. Una primera alternativa es utilizar el puntaje total que con base en las observaciones anteriores está correlacionado positivamente con el rasgo y es un estadístico óptimo para la clasificación binaria.

Schriever (1985) propone utilizar análisis de correspondencia múltiple. La primera componente principal de la matriz de correlaciones de los ítems determina la habilidad

La primera componente puede escribirse como

$$Y = \sum_{i=1}^{n} \alpha_i \ (-1)^{1-u_i} \sqrt{\frac{1-P_i}{P_i}}$$

donde α es el vector propio del valor propio mayor y $\alpha^t\alpha = 1$

Lewis (1983) propone un método bayesiano que permite obtener intervalos de confianza para el rasgo y puede ser usado en tests adaptativos informatizados.

Coeficiente H de escalabilidad

Loevinger (1947,1948) propuso el coeficiente H pero no tuvo mayor aplicabilidad hasta que lo adaptó Mokken (1971) para definir una familia de coeficientes que indican la escalabilidad monótona homogénea para:

a) un par de ítems;

b) un ítem respecto a los restantes;

c) el conjunto total de los ítems.

Consideremos dos ítems i, j dicotómicos, al aplicarse los dos ítems sobre N sujetos se obtiene una tabla 2x2 con las posibles salidas. Si mantenemos fijas las marginales podemos obtener la tabla que produce la máxima covarianza entre los ítems, luego el coeficiente H se define como:

$$H_{ij} = \frac{Cov(X_i, X_j)}{Cov_{\text{máx}}(X_i, X_j)} = \frac{P_{ij} - P_iP_j}{P_i - P_iP_j} = 1 - \frac{P_i - P_{ij}}{P_i(1-P_j)}$$

Donde se utiliza el hecho que el máximo de la covarianza condicionada a las marginales se da cuando $P_{ij} = P_i$.

Se puede definir H a partir de las probabilidades de error de Guttman.

Consideremos dos ítems: i, j tal que i<j o sea $P_i < P_j$ la probabilidad de error de Guttman se define como $e_{ij} = P(X_i = 1, X_j = 0) = P_i - P_{ij}$, y el valor esperado del error bajo la hipótesis de independencia es $e^o_{ij} = P_i(1 - P_j)$. Luego $H_{ij} = 1 - \frac{e_{ij}}{e^o_{ij}}$.

Valores altos de H_{ij} se corresponden con CCI discriminantes ($e_{ij} \approx 0$) y valores bajos con CCI no discriminan (son independientes) en tanto valores negativos contradicen la monotonía homogénea.

El coeficiente de escalabilidad del ítem i con el test se define como:

$$H_i = \frac{Cov(X_i, X_+^{-i})}{Cov_{\text{máx}}(X_i, X_+^{-i})} = \frac{\sum\limits_{j \neq i}(P_{ij} - P_i P_j)}{\sum\limits_{j>i}(P_i - P_i P_j) + \sum\limits_{j<i}(P_i - P_i P_j)} =$$
$$1 - \frac{\sum\limits_{j \neq i}(P_i - P_{ij})}{\sum\limits_{j>i}(P_i - P_i P_j) + \sum\limits_{j<i}(P_i - P_i P_j)}$$

Finalmente el coeficiente de escalabilidad del test es:

$$H = \frac{\sum\limits_i Cov(X_i, X_+^{-i})}{\sum\limits_i Cov_{\text{máx}}(X_i, X_+^{-i})} = 1 - \frac{\sum\limits_i \sum\limits_{j \neq i}(P_i - P_{ij})}{\sum\limits_i \sum\limits_{j>i}(P_i - P_i P_j) + \sum\limits_{j<i}(P_i - P_i P_j)}$$

Se puede ver fácilmente que:

$$\begin{aligned} &\min_j \ H_{ij} \leq H_i \leq \max_j \ H_{ij} \\ &\min_i \ H_i \leq H \leq \max_i \ H_i \\ &\min_{i,j} \ H_{ij} \leq H \leq \max_{i,j} \ H_{ij} \end{aligned}$$

y

$$0 \leq H_{ij} \leq 1 \quad 0 \leq H_i \leq 1 \quad 0 \leq H \leq 1$$

Mokken (1971) propone que una escala es útil si H es mayor o igual a una constante, que debe ser al menos mayor o igual a 0.3 y clasifica las escalas como débiles si $0{,}3 \leq H < 0{,}4$ moderadas si $0{,}4 \leq H < 0{,}5$ y fuertes si $H \geq 0{,}5$.

2.22. Modelo no paramétrico unidimensional usando regresión no paramétrica

Este enfoque del problema en Psicometría fue introducido por Ramsay (1991) y puede ser considerado como una técnica del análisis de datos funcionales.

Consideraremos N sujetos que responden a n ítems dicotómicos o politómicos y queremos estimar el vector de rasgos θ y las curvas características de los ítems (CCI) mediante estimadores no paramétricos.

Uno de los principales cuestionamientos que se pueden hacer al enfoque paramétrico, es que la dependencia entre la probabilidad y el rasgo es forzada a ser de determinada forma funcional y no se toma en cuenta hechos tales como la no monoticidad de los ítems o desviaciones del formato impuesto. También está de por medio la complejidad de los cálculos, tanto si se emplea el paradigma de Birnbaum (máxima verosimilitud marginal con el algoritmo EM) como procedimientos bayesianos.

En los algoritmos usuales como el Logist (Wingersky, Patrick y Lord, 1988) y Bilog (Mislevy y Bock 1982) la velocidad de convergencia es lineal y usualmente lenta.

Existen también modelos cuasiparamétricos, los cuales son más flexibles. Ejemplos de éstos son aquellos que usan la regresión polinómica, o desarrollos en series de Fourier.

Wahba (1990) propone un modelo parcialmente no paramétrico. Otros trabajos que apuntan en la misma dirección son los de Levine (1984, 1985), Drasgow, Levine, Williams, Mc Laughlin y Candell (1989), Ramsay y Abrahamowicz (1989), Ramsay y Winsberg (1991).

Desde el punto de vista no paramétrico tenemos el modelo homogéneo monótono y el modelo homogéneo de doble monoticidad. Para ver un desarrollo detallado de estos modelos para ítems dicotómicos y su extensión a politómicos se puede consultar el libro de Sijtsma y Molenaar (2002) o los capítulos de Mokken y Molenaar en Van der Linden y Hambleton (1997). La fortaleza de este acercamiento se ve disminuida al no dar los autores un estimador de las Curvas Características del Ítem (CCI).

Así pues, Ramsay (1991) propone no sólo estimar el rasgo no paramétricamente, sino utilizar los métodos de regresión no paramétrica mediante núcleos para obtener una estimación de las CCI.

Este método tiene varias ventajas: la primera de ellas es la no imposición de ninguna restricción sobre la forma funcional de las CCI, y además, el método de estimación es no iterativo y muy fácil de programar.

Es de destacar que, para obtener la estimación del rasgo, debe conocerse su distribución, pero si se ignora, lo que se estima es una transformación del rasgo dada por $F^{-1}(\Theta)$ que tiene distribución uniforme en [0,1]. Esto no constituye una limitación, pues como lo expresa Ramsay (1991): «en el contexto del análisis de ítems, un test no puede proporcionarnos más que información sobre el orden de los examinados». Esto es conocido como la pérdida de la identificabilidad.

Sea $P_{jm}(\theta)$ la probabilidad que un examinado con nivel de rasgo θ seleccione la opción m del ítem j con $j = 1, ..., n$ $m = 1, ..., M_j$.

Consideremos N sujetos examinados, entonces tenemos determinadas variables indicadoras Y^i_{jm}, que valen 1 si el sujeto i eligió la opción m en el ítem j, y 0 en caso contrario. O sea, $P_{jm}(\theta) = P(Y_{jm} = 1|\Theta = \theta)$.

Como dijimos, veremos un enfoque del problema simple de entender, fácil de programar, no iterativo, rápido, y eficiente.

Además, estas técnicas permiten tener herramientas de la Teoría de Respuesta al Ítem (TRI) para situaciones de muestras de tamaño moderado.

Ramsay (1991) plantea, a los efectos de la estimación de las curvas características, la utilización de la regresión no paramétrica.

Debido a que la regresión no paramétrica entre dos variables X e Y supone el conocimiento de ambas, y en nuestro caso se desconoce la variable θ, el primer paso que Ramsay (1991) se plantea es el de estimar los rasgos θ. Para ello, supone que éstas tienen alguna distribución conocida $F(\theta)$.

Consideremos un estadístico T y sean T_i los valores de los sujetos. Sea r_i el rango del i-ésimo individuo. Para cada valor r_i consideremos la quartila q_i respecto a alguna distribución. En general, se usa la normal típica, y si consideramos el rasgo como proporciones, se puede usar la distribución beta, por ejemplo, la distribución B(2.5,2.5).

Con base en el valor de T_i se ordena a los individuos, obteniéndose ahora los patrones de respuestas ordenados. Se estiman los P_{jm} mediante regresión no paramétrica entre las variables binarias $Y_{jm}^{(i)}$ y el vector de rasgos $\widehat{\theta_1} = q_1, ..., \widehat{\theta_N} = q_N$.

Se suele utilizar como estadístico el puntaje total, pero Ramsay (1991) propone un estadístico alternativo.

Sea

$$W_{jm} = logit(P_{jm}^{(75)}) - logit(P_{jm}^{(25)})$$

donde $P_{jm}^{(75)}$) y $P_{jm}^{(25)}$) son las proporciones de examinados que caen en el 25 % superior e inferior de la distribución de los puntajes totales de quienes eligen la opción m en el ítem j.

Entonces utiliza el estadístico:

$$T_i = \sum_j \sum_m Y_{jm}^{(i)} W_{jm}$$

o sea, T_i es la suma de pesos para las opciones actualmente elegidas. Este estadístico tiene la ventaja de que produce menos empates y usa información sobre la elección de alternativas incorrectas.

El estimador por núcleos de $P(\theta)$ es el promedio ponderado.

$$\widehat{P}_{jm}(\theta) = \sum_{i=1}^{I} w_i(\theta) Y_{jm}^{(i)} \tag{2.137}$$

donde las ponderaciones $w_i(\theta)$ son definidas para que sean máximas en $\theta = q_i$, y que decrezcan a cero cuando crece $|q_i - \theta|$.

Los pesos $w_i(\theta)$ se definen a partir de los núcleos $K(u)$ como:

$$w_i(\theta) = \frac{K(\frac{\widehat{\theta_i} - \theta}{h})}{\sum_k K(\frac{\widehat{\theta_k} - \theta}{h})} \tag{2.138}$$

Luego,

$$\widehat{P}_{jm}(\theta) = \frac{\sum_{i=1}^{N} K(\frac{\widehat{\theta_i}-\theta}{h}) Y_{jm}^{(i)}}{\sum_{k=1}^{N} K(\frac{\widehat{\theta_k}-\theta}{h})} \tag{2.139}$$

donde el denominador es la estimación de la densidad de las q_i. Esto implica que, como la distribución es conocida, se podría reemplazar el denominador por esta densidad conocida. Esto podría ocasionar valores de P levemente fuera del intervalo [0,1], pero esto puede ser corregido.

Como la diferenciabilidad de $\widehat{P}$ depende de la diferenciabilidad de K, es ventajoso usar los núcleos cuadráticos o gaussianos, por ejemplo, si queremos estimar la función de información del test.

El parámetro h de ancho de banda debe ser elegido adecuadamente. Si h disminuye, también disminuye el sesgo, dado que sólo pocas observaciones muy cercanas a θ son efectivamente ponderadas, pero crece la varianza. Si h crece, disminuye la varianza, pues se ponderan más observaciones, pero aumenta el sesgo.

En Hardle (1990) se dan varios procedimientos para elegir la ventana óptima a los efectos de minimizar el error cuadrático medio o funciones de éste.

Es interesante independizarse de la posible distribución del rasgo, ya que ésta, usualmente, es desconocida.

Notemos por un momento al rasgo como τ y supongamos que tiene distribución F. Sabemos que $F(\tau)$ se distribuye uniforme [0,1]. Por lo tanto, prodríamos estimar en vez de τ una transformación monótona de ella $\theta = F(\tau)$.

Así pues, se puede suponer que el rasgo tiene distribución uniforme, y entonces, este rasgo se estima mediante la proporción empírica.

Si conocemos la distribución F del rasgo, obtenemos la escala original, mediante $\tau = F^{-1}(\theta)$, que tiene distribución F.

Teniendo en cuenta esto, sea para un test de tamaño n $\overline{U}_{n,1}, ..., \overline{U}_{n,N}$, las proporciones de respuestas correctas para cada sujeto, y sea $\overline{U}_{n,-i,1}, ..., \overline{U}_{n,-i,N}$, la proporción de respuestas correctas sin considerar el ítem i.

Tenemos:

$$\overline{U}_{n,k} = \frac{1}{n} \sum_{j=1}^{n} Y_{n,j,k} \tag{2.140}$$

$$\overline{U}_{n,-i,k} = \frac{1}{n-1} \sum_{j=1 j \neq i}^{n} Y_{n,j,k} \tag{2.141}$$

Notaremos por F_n y $F_{n,-i}$ las distribuciones de $\overline{U}_{n,k}$ y $\overline{U}_{n,-i,k}$ respectivamente.

A los efectos de romper los empates, a cada $\overline{U}_{n,-i,k}$ se le agrega una variable aleatoria $\overline{W}_{n,-i,k}$, con distribución uniforme en [0,1/n-1], para obtener una sucesión sin empates $\overline{U}^*_{n,-i,1}, ..., \overline{U}^*_{n,-i,N}$, como:

$$\overline{U}^*_{n,-i,j} = \overline{U}_{n,-i,j} + \overline{W}_{n,-i,j} \tag{2.142}$$

Podemos definir $\widehat{F}_{N,-i}$ y $\widehat{G}_{N,-i}$ como las distribuciones empíricas de $\overline{U}_{n,-i,j}$ y $\overline{U}^*_{n,-i,j}$ respectivamente, o sea:

$$\widehat{F}_{N,-i}(x) = \frac{\#\{\overline{U}_{n,-i,j} \leq x : j = 1, ..., N\}}{N} \tag{2.143}$$

$$\widehat{G}_{N,-i}(x) = \frac{\#\{\overline{U}^*_{n,-i,j} \leq x : j = 1, ..., N\}}{N} \tag{2.144}$$

A partir de éstas se puede calcular el percentil empírico para cada sujeto.

$$\widehat{G}_{N,-i}(\overline{U}^*_{n,-i,1}), ..., \widehat{G}_{N,-i}(\overline{U}^*_{n,-i,N}) \tag{2.145}$$

Si suponemos que la variable latente tiene distribución F, como ya dijimos, se convierten estos percentiles a la escala del rasgo mediante:

$$\widehat{\theta}_{n,-i,k} = F^{-1}\widehat{G}_{N,-i}(\overline{U}^*_{n,-i,k}) \tag{2.146}$$

Nosotros, por esto, consideraremos que F es la uniforme en [0,1] o sea, finalmente,

$$\widehat{\theta}_{n,-i,k} = \widehat{G}_{N,-i}(\overline{U}^{*}_{n,-i,k}) \tag{2.147}$$

Se estima la curva característica del ítem como:

$$\widehat{P}_{n,i}(\theta) = \frac{\sum_{k=1}^{N} K(\frac{\theta-\widehat{\theta}_{n,-i,k}}{h_n}) Y_{n,i,k}}{\sum_{k=1}^{N} K(\frac{\theta-\widehat{\theta}_{n,-i,k}}{h_n})} \tag{2.148}$$

A los efectos de obtener estimadores ordinales del rasgo, se utilizan todos los puntajes de los ítems y no se rompen los empates. Se usa la proporción de respuestas correctas $\overline{U}_{n,k}$ y entonces,

$$\widehat{\theta}_{n,k} = F^{-1}(\widehat{F}_N(\overline{U}_{n,k})) \tag{2.149}$$

Como queremos obtener la consistencia conjunta de las CCI y de los parámetros del rasgo, se requiere que el tamaño muestral y la longitud del test crezcan juntos. Douglas (1997) hace una demostración de esto en un excelente artículo.

Consideremos primero un test de n ítems, administrado a N_n examinados, elegidos al azar. Se estiman las CCI de los n ítems y los N_n rasgos. Entonces ahora, se considera un nuevo test de $n+1$ ítems, que no necesariamente contengan los n anteriores, y se administra a N_{n+1} examinados.

Pretendemos hacer la estimación para cada fila en el arreglo triangular así obtenido.

La teoría asintótica estudia la adecuación de las curvas características del ítem y de los estimados del rasgo a medida que n tiende a infinito.

Tenemos entonces la sucesión de test:

$P_{n,1}, P_{n,2}, \ldots, P_{n,n}$
$P_{n+1,1}, P_{n+1,2}, \ldots, P_{n+1,n}, P_{n+1,n+1}$
$P_{n+2,1}, P_{n+2,2}, \ldots, P_{n+2,n+1}, P_{n+2,n+2}$
. . . . . .
. . . . . .

Sucesión de rasgos:

$\theta_{n,1}, \theta_{n,2}, \ldots, \theta_{n,N_n}$
$\theta_{n+1,1}, \theta_{n+1,2}, \ldots, \theta_{n+1,n}, \theta_{n+1,N_{n+1}}$

$\theta_{n+2,1}, \theta_{n+2,2}, \ldots, \theta_{n+2,n+1}, \theta_{n+2,N_{n+2}}$

. . . . . .

. . . . . .

Cada test viene identificado por sus curvas características de los ítems $P_{n,i}$. Para cada curva $P_{n,i}$ se necesita construir un estimador $\widehat{P}_{n,i}$ y similarmente construiremos estimadores $\widehat{\theta}_{n,k}$ de los rasgos $\theta_{n,k}$.

Notemos por $\overline{P}_n$ a la curva característica del test, o sea,

$$\overline{P}_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} P_{n,j}(\theta).$$

2.22.1. Hipótesis

A los efectos de obtener la consistencia de estas estimaciones, Douglas (1997) asume ciertas hipótesis sobre los ítems, las CCI, los tamaños muestrales N_n, los núcleos y los anchos de banda h_n.

1. Independencia local y unidimensionalidad.
2. Θ tiene distribución uniforme (0,1).
3. $\overline{P}'_n$ existe en (0,1) y existe un número m tal que $0 < m < \overline{P'}_n(\theta)$ para todo $\theta \in (0,1)$.
4. Existe $C > 0$ tal que $np_{nk} < C$ para todo k donde $p_{nk} = P(\overline{U_n} = k)$.
5. Existen constantes M_L y M_U tal que $M_L n^{3/2} < N_n < M_U n^r$.
6. $P'_{n,i}$ existe y es continua en (0,1) para todo n,i.
7. Para cada intervalo compacto $[a,b] \subset (0,1)$ existe una constante $M_{a,b}$ tal que para todo $\theta \in [a,b]$ n y todo n, i $|P'_{n,i}(\theta)| < M_{a,b}$.
8. K tiene soporte acotado, en particular existe una constante C_K, tal que $K(x) > 0$, cuando $|x| < C_K$ y $K(x) = 0$, cuando $|x| > C_K$.
9. K es Lipschitziana de constante L_K.
10. Existe un número $\alpha \in (0, 1/2)$ y números positivos L_α y U_α tal que para todo n,

$$L_\alpha n^{-\alpha} < h_n < U_\alpha n^{-\alpha}$$

.

2.22.2. Consistencia

A partir de estas hipótesis se obtiene primeramente:

Teorema 2.22.1. Dadas las hipótesis A1 a A7 $max_{j=1,...Jn} \mid \widehat{\Theta}_{n,j} - \Theta_{n,j} \mid \rightarrow 0$ con probabilidad 1.

También se obtiene un resultado sobre el error cuadrático.

Teorema 2.22.2. Dadas las hipótesis 1 a 10 para $\theta_0 \in (0,1)$ fijo

$$max\{ECM(\widehat{P}_{n,i}(\theta_0)) : 1 \leq i \leq n\} \leq O(n^{-2\alpha}) \tag{2.150}$$

donde α es la constante de la hipótesis 10.

Definiendo

$$\delta_{n,a,b} = sup_{\theta \in (a,b)} max_{1 \leq i \leq n} \mid \widehat{P}_{n,i}(\theta) - P_{n,i}(\theta) \mid$$

se obtiene:

Teorema 2.22.3. Dadas las hipótesis 1 a 10, $\delta_{n,a,b} \rightarrow 0$ si $n \rightarrow +\infty$ con probabilidad 1.

Finalmente prueba la conistencia conjunta pues a partir de los teoremas 2.22.1 y 3.1.18, y notando que la probabilidad del conjunto donde falla esta convergencia es la unión de dos conjuntos cada uno con probabilidad 0.

Teorema 2.22.4. $max\{\Delta_n, \delta_{n,a,b}\} \rightarrow 0$ si $n \rightarrow +\infty$ con probabilidad 1.

2.23. Modelos paramétricos multidimensionales

Considérense dos vectores de variables aleatorias ν y η de dimensiones n x 1 y p x 1, respectivamente. Sabemos que:

$$\Pr(\nu = u) = \int_{R} \cdots \int g(\nu = u \mid \eta = h) f(h) dh$$

donde u y h son dos realizaciones de las variables ν y η , respectivamente, f(h) es la función de densidad de las variables aleatorias η y la integración se realiza con respecto a η . Se denominan modelos de rasgos latentes multidimensionales a aquellos modelos en los que:

a) Las únicas variables que pueden ser observadas son las ν , mientras que las variables aleatorias η son inobservables y se asume que son continuas.

b) Las variables ν son independientes entre sí para un valor fijo de la variable η , digamos $\eta = h$. Es decir, las variables ν son condicionalmente independientes o localmente independientes.

La primera parte de la definición de los modelos latentes implica que el área de integración p-dimensional R viene dado por el producto de p intervalos $R_i = (-\infty, \infty)$. La segunda parte de la definición implica que podemos escribir:

$$g(\nu = u \,|\eta = h) = \prod_{i=1}^{n} g(\nu_i = u_i \,|\eta = h)$$

y por lo tanto, podemos reescribir como:

$$\Pr(\nu = u) = \int_{R} \ldots \int g(\nu = u \,|\eta = h) f(h) dh = \int_{R} \ldots \int \prod_{i=1}^{n} g(\nu_i = u_i \,|\eta = h)\ f(h) dh$$

En la ecuación anterior es necesario especificar lo siguiente:

a) la dimensionalidad del vector de variables latentes, p;

b) la expresión de la función $g(\nu = u \,|\eta = h)$;

c) la expresión de la función de densidad f(h).

La finalidad última de los modelos de rasgos latentes es la de ofrecer una representación parsimoniosa de datos multivariados. Es decir, estamos interesados en hallar un modelo del tipo anterior que nos proporcione una adecuada representación de los datos en tan pocas dimensiones como sea posible. Además deseamos que p sea menor que n, a ser posible mucho menor que n.

Los modelos de rasgos latentes son un conjunto amplio de modelos entre los que se encuentran:

* el modelo del factor, que es un modelo de rasgos latentes caracterizado por asumir que las variables observadas son continuas, que las funciones $g(\nu = u \,|\eta = h)$ son lineales y que la función de densidad de las variables latentes f(h) es multivariada normal;

* y los modelos de respuesta a los ítems que son aquellos modelos en los que las demás funciones $g(\nu = u \,|\eta = h)$ son no lineales.

Como vimos en los capítulos anteriores existe una amplia literatura sobre modelos unidimensionales de respuesta a los ítems, es decir, con una única variable latente, p=1.

Se han propuesto modelos multidimensionales para datos dicotómicos y para datos policotómicos ordenados y se dispone de métodos de estimación implementados en programas comerciales para este tipo de datos. Pese a que se han propuesto modelos multidimensionales para datos policotómicos no ordenados (Takane y de Leeuw, 1987) no se han implementado aún procedimientos de estimación de dichos modelos.

En la literatura de TRI se denominan funciones de respuesta a la opción i de un ítem a la probabilidad condicional: $g(\nu_i = u_i \,|\eta = h) \equiv \Pr(\nu_i = u_i \,|\eta = h)$. En los modelos paramétricos, estas funciones se expresan como una función de una o más constantes que deben ser estimadas a las que se puede asignar una interpretación psicológica. A estas constantes se las denomina parámetros. En los modelos no paramétricos los parámetros que aparecen en las funciones de respuesta a las opciones de un ítem carecen de interpretación. No se ha propuesto hasta el momento ningún modelo multidimensional de respuesta a los ítems no paramétricos.

Durante años, los efectos de los ítems y los efectos de las variables latentes han sido tratados como efectos fijos en la literatura de teoría de respuesta a los ítems. En este enfoque se especifica una función paramétrica para cada una de las funciones de respuesta

a las opciones a un ítem y se asigna un parámetro a cada sujeto. El objetivo es entonces estimar simultáneamente los parámetros asociados con los ítems y los parámetros asociados con los sujetos. Sin embargo, tal como notaron Neyman y Scott (1948), dado que los parámetros asociados con los sujetos aumentan en número conforme se incremente el tamaño de la muestra, las deseables propiedades asintóticas de los estimadores de máxima verosimilitud o mínimos cuadrados generalizados no se cumplen. Para resolver esto Bock y Lieberman (1970) propusieron considerar a los ítems como un efecto fijo, pero a las variables latentes como un efecto aleatorio. En tal caso, se estimarían los parámetros asociados con cada uno de los ítems, pero unicamente se estimaría la distribución (o densidad) de los parámetros asociados con los sujetos. La conceptualización de los modelos de rasgos latentes como modelos mixtos es la utilizada por Bartholomew (1987), McDonald o Holland (1990).

2.23.0.1. El modelo de respuesta a los ítems multivariado normal para variables dicotómicas

Cuando las variables observadas son dicotómicas, $\nu_i = \{0, 1\}$, y la función que modela la probabilidad de respuesta es la función de distribución de la normal estándar, $\Phi(\bullet)$, las funciones de respuesta a las opciones a un ítem vienen dadas por:

$$\Pr(\nu_i = 1 \,|\eta = h) = \Phi(\alpha_1 + \beta_i^{'} h) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a_i + \beta_i^{'} h} \exp\left(-\frac{y^2}{2}\right) dy$$

$$\Pr(\nu_i = 0 \,|\eta = h) = 1 - \Pr(\nu_i = 1 \,|\eta = h),$$

donde α_i y $\beta_i^{'}$ son el intercepto y vector de pendientes del ítem i y se asume que los rasgos latentes siguen una distribución multivariada normal. Es decir , $f(h) = \varphi(h) \sim N(0, \Phi)$ donde $\varphi(\bullet)$ denota una densidad multivariada normal, y Φ es la matriz de correlaciones entre los factores.

Dado que podemos escribir:

$$\Pr(\nu_i = u_i \,|\eta = h) = [\Pr(\nu_i = 1 \,|\eta = h)]^{\nu_i} \left[1 - \Pr(\nu_i = 1 \,|\eta = h)\right]^{1-\nu_i}$$

$$\nu_i = \{0, 1\}$$

Sustituyendo obtenemos:

$$\Pr(\nu = u) = \int \underset{R}{...} \int \prod_{i=1}^{n} \left[\Pr(\nu_i = 1 \,|\eta = h)\right]^{\nu_i} \left[1 - \Pr(\nu_i = 1 \,|\eta = h)\right]^{1-\nu_i} \varphi(h) dh$$

Las ecuaciones anteriores definen el modelo normal para datos dicotómicos.

Existe una forma alternativa de derivar el modelo normal. Considérese un vector de variables y de dimensiones n x 1 que sigue el modelo del factor común, y asúmase una densidad multivariada normal para y. Por tanto, la densidad marginal de y viene dada por:

$$f^*(y) \sim N(\mu_y, \sum\nolimits_y) = N(0, \Lambda\Phi\Lambda' + \Psi)$$

y la densidad de y condicionada a un valor fijo de η es $g^*(y = y^* \,|\eta = h) \sim N(\Lambda h, \Psi)$, donde Λ es una matriz de cargas factoriales, Ψ es una matriz diagonal de varianzas residuales y Φ es la matriz de correlaciones entre las variables latentes o factores. Las variables aleatorias y y η no son directamente observables, únicamente observamos

$$\nu_i = \begin{cases} 1 & si \quad y_i \geq \tau_i \\ 0 & si \quad y_y < \tau_i \end{cases} \quad i = 1, ..., n$$

donde el vector de parámetros τ se denomina umbrales.

Dado que las variables observables son dicotómicas, las varianzas de las variables y no son identificables, por lo que es habitual fijar su valor igual a la unidad, es decir, $diag(\Lambda\Phi\Lambda' + \Psi) = 1$ mediante $\Psi = I - diag(\Lambda\Phi\Lambda')$

Las ecuaciones anteriores implican que:

$$\Pr(\nu = u) = \int \underset{R^*}{...} \int f^*(y) dy$$

donde $f^*(y)$ es la densidad multivariada y R^* es un área n-dimensional de integración dada por el producto de intervalos:

$$R_i^* = \begin{cases} (\tau_{i'} \infty) & si \quad \nu_i = 1 \\ (-\infty, \tau_i) & si \quad \nu_i = 0 \end{cases}$$

El modelo especificado utilizando es equivalente al primero, pues podemos escribir:

$$\Pr(\nu = u) = \int \underset{R^*}{\cdots} \int f^*(y) dy =$$

$$\int \underset{R^*}{\cdots} \int \left[\int \underset{R}{\cdots} \int g^*(y = y^* \,|\eta = h) f(h) dh \right] dy =$$

$$\int \underset{R^*}{\cdots} \int f(h) \left[\int \underset{R}{\cdots} \int g^*(y = y^* \,|\eta = h) dy \right] dh$$

Operando,

$$\int \underset{R^*}{\cdots} \int g^*(y = y^* \,|\eta = h) dy = \prod_{i=1}^{n} \int_{R_i^*} g^*(y_i = y_i^* \,|\eta = h) dy_i =$$

$$\prod_{i=1}^{n} \left[\int_{\tau_i}^{\infty} g^*(y_i = y_i^* \,|\eta = h) \right]^{v_i} \left[1 - \int_{\tau_i}^{\infty} g^*(y_i = y_i^* \,|\eta = h) dy_i \right]^{1-\nu_i}$$

Dado que la función de densidad de y condicionada a η es normal, hallamos que las funciones de respuesta a las opciones son:

$$\Pr(\nu_i = 1 \,|\eta = h) = \Pr(y_i \,\Big|\, \eta > \tau_i) = \int_{\tau_i}^{\infty} g^*(y_i = y_i^* \,|\eta = h) dy_i =$$

$$= \frac{1}{\sqrt{2\pi}\sigma_{y_i|\eta=h}} \int_{\tau_i}^{\infty} \exp\left[-\frac{1}{2} \left(\frac{y_1 - \mu_{y_i|\eta=h}}{\sigma_{y_i|\eta=h}} \right)^2 \right] dy_i$$

donde

$$\mu_{y_i|\eta=h} = \lambda_i^{'} h$$

$$\sigma_{y_i|\eta=h} = \sqrt{\Psi_i} = \sqrt{1 - \lambda_i^{'} \Phi \lambda_i}$$

Usando el cambio de variable $z_i = \frac{y_i - \mu_{y_i|\eta=h}}{\sigma_{y_i|\eta=h}}$ tenemos que $dy_i = \sigma_{y_i|\eta=h} dz_i$, y cuando $y_i = \tau_i, z_i = \frac{\tau_i - \lambda_i^{'} h}{\sqrt{1 - \lambda_i^{'} \Phi \lambda_i}}$, por lo que

$$\Pr(\nu_i = 1 \,|\eta\, h) = \int_{\frac{\tau_i - \lambda_i^{'} h}{\sqrt{1 - \lambda_i^{'} \Phi \lambda_i}}}^{\infty} \varphi(z_l) dz_i \equiv 1 - \Phi\left(\frac{\tau_i - \lambda_i^{'} h}{\sqrt{1 - \lambda_i^{'} \Phi \lambda_i}} \right) = \Phi\left(\frac{-\tau_i + \lambda_i^{'} h}{\sqrt{1 - \lambda_i^{'} \Phi \lambda_i}} \right)$$

Las ecuaciones se relacionan por:

$$\alpha_i = \frac{-\tau_i}{\sqrt{1 - \lambda_i^{'} \Phi \lambda_i}}$$

y

$$\beta_{ij} = \frac{\lambda_{ij}}{\sqrt{1 - \lambda_i^{'} \Phi \lambda_i}}$$

o sea, ambas ecuaciones son diferentes parametrizaciones del mismo modelo.

De igual forma que en el caso unidimensional del anterior modelo podemos deducir el modelo logístico multidimensional quedando:

$$P(U_{ij} = 1 \mid a_i, d_i, c_i, \theta_j) = c_i + (1 - c_i) \frac{e^{(a_i^t \theta_j + d_i)}}{1 + e^{(a_i^t \theta_j + d_i)}}$$

donde t indica el vector traspuesto.

Capítulo 3

Modelo no paramétrico multidimensional

3.1. Introducción

Presentaremos un modelo no paramétrico multidimensional basado en núcleos. Aquí se utilizará el estimador de regresión no paramétrico con variables independientes multidimensionales. Si bien es natural pensar en aplicar un estimador semejante al unidimensional para ítems multidimensionales, la mayor dificultad se encuentra en estimar el rasgo, ya que éste ahora es un vector. Se necesitan tantos estadísticos para ordenar a los sujetos como rasgos y esto presenta complicaciones técnicas. En el caso unidimensional se utilizaba el promedio o el puntaje total, pero en el caso multidimensional no aparece claro cómo deben ser los estadísticos.

Nos proponemos ahora encontrar las condiciones que deben cumplir dichos estadísticos para que la estimación sea consistente.

A los efectos de estimar cada componente de este vector que representa un rasgo específico, utilizaremos estadísticos para ordenar a los sujetos y así tomar como estimación de ese rasgo en cada sujeto el cuartil empírico correspondiente.

Consideramos ahora que el rasgo es un vector aleatorio en R^d

$$\Theta = (\Theta_1, ..., \Theta_d).$$

Propondremos un método general para estimar el rasgo y un estimador para la curvas características del ítem.

Veremos las condiciones que deben cumplir los estadísticos para obtener estimaciones consistentes del rasgo y la relación entre el tamaño muestral N y la cantidad de ítems n para obtener estimaciones consistentes de las CCI.

Sea (Ω, A, P) un espacio de probabilidad y consideremos sin pérdida de generalidad que el rasgo latente Θ tiene distribuciones marginales uniformes $U[0, 1]$.

Es claro que todo lo que deduzcamos para este supuesto funcionará para cualquier tipo de distribución marginal.

Consideremos una sucesión de funciones vectoriales, Borel medibles, $\mathbf{g_n}$ en R^n a valores en $[0, 1]^d$.

Sea $g_{n,l}$ la componente l-ésima de $\mathbf{g_n}$, para $x \in R^n$ $\mathbf{g_n}(x) = (g_{n,1}(x), ..., g_{n,d}(x))$.

Consideremos un test de n ítems contestados por N sujetos y las variables aleatorias dicotómicas $X_{i,k}$, $i = 1, ..., n$, $k = 1, ...N$ que indican la respuesta del k-ésimo sujeto al i-ésimo ítem. Estas variables dependen de d rasgos latentes (Θ_l) y sea $\mathbf{X} = (X_1, .., X_n)$ el

vector aleatorio que indica las respuestas a los n ítems. Específicamente, $\mathbf{X_k}$ indicará el vector de respuestas del sujeto k.

Supongamos que las funciones $g_{n,l}(X)$ son independientes de $\Theta_1, ..., \Theta_{l-1}, \Theta_{l+1}, .., \Theta_n$ para todo $1 \leq l \leq n$.

Definamos la sucesión de funciones $\{G_{g_n}\} : R^d \longrightarrow R^d$ tal que para cada $x = (x_1, ..., x_d)$ se tiene:

$$G_{g_n}(x) = (P(g_{n,1}(X) \leq x_1), ..., P(g_{n,d}(X) \leq x_d)) \tag{3.1}$$

También tenemos las empíricas:

$$\widehat{F}_{N,l}(x) = \frac{\#\{g_{n,l}(X) \leq x\}}{N} = \frac{\sum_{i=1}^{N} \chi_{\{g_{n,l}(X_i) \leq x\}}}{N}$$

con $l = 1, .., d$ y definamos la función:

$$\widehat{G}_N(x_1, ..., x_d) = (\widehat{F}_{N,1}(x_1), ..., \widehat{F}_{N,d}(x_d)) \tag{3.2}$$

A partir de lo anterior, para estimar la componente l-ésima del rasgo se utiliza la función $g_{n,l}$ para ordenar los sujetos. Luego tomamos la función de distribución empírica, es decir,

$$\widehat{\Theta}_n = \widehat{G}_N(g_n(X)) \tag{3.3}$$

Notaremos además por:

$$F_{n,l}(x) = P(g_{n,l}(X) \leq x).$$

Antes de continuar, veamos algunos resultados sobre desigualdades de concentración, que daremos sin demostración y nos ayudarán en lo que sigue. Para una descripción detallada de estos resultados recomendamos ver el excelente trabajo de Boucheron, Lugosi y Massart (2003).

Consideremos un conjunto $\mathcal{X}$ y $g : \mathcal{X}^n \longrightarrow R$ una función medible.

Sea $Z = g(X_1, ...X_n)$ donde $X_1, ...X_n$ son variables aleatorias independientes (no tienen por qué ser idénticamente distribuidas) que toman valores en $\mathcal{X}$. Consideremos $E(Z)$ y $E_i(Z) = E(Z \mid X_1, .., X_{i-1}, X_{i+1}...X_n)$.

Sea $X'_1, ...X'_n$ una copia independiente de $X_1, ...X_n$ y $Z'_i = g(X_1, ., X'_i, ..X_n)$.

El primer resultado que se obtiene es:

Teorema 3.1.1.

$$V(Z) \leq \sum_{i=1}^{n} E[(Z - E_i(Z))^2]$$

y de éste resulta:

Teorema 3.1.2. Desigualdad de Efron-Stein.

$$V(Z) \leq \frac{1}{2}\sum_{i=1}^{n} E[(Z - Z_i')^2]$$

Definición 3.1.3. Diremos que una función $g : \mathcal{X}^n \longrightarrow R$ tiene la propiedad de diferencias acotadas si existen constantes no negativas $c_1, ..., c_n$, tales que:

$$sup_{x_1,...,x_n,x_i' \in \mathcal{X}} \mid g(x_1, .., x_i, .., x_n) - g(x_1, .., x_i', .., x_n) \mid \leq c_i$$

para $1 \leq c_i \leq n$.

Para funciones de diferencias acotadas se tiene el siguiente corolario:

Corolario 3.1.4. Si g tiene la propiedad de diferencias acotadas con constantes $c_1, ..., c_n$

$$V(Z) \leq \frac{1}{2}\sum_{i=1}^{n} c_i^2$$

A partir de la desigualdad logarítmica simétrica de Sobolev, debida a Massart (2000), se prueba que:

Teorema 3.1.5. Si existe una constante C positiva, tal que, casi seguramente

$$\sum_{i=1}^{n}(Z - Z_i')^2 \leq C$$

entonces para todo $t > 0$

$$P[\mid Z - E(Z) \mid > t] \leq 2e^{-\frac{t^2}{4C}}$$

Para las funciones de diferencias acotadas se tiene el corolario:

Corolario 3.1.6. Si g tiene la propiedad de diferencias acotadas con constantes $c_1, ..., c_n$ entonces,

$$P[\mid Z - E(Z) \mid > t] \leq 2e^{-\frac{t^2}{4\sum_{i=1}^{n} c_i^2}}$$

Esta cota puede ser mejorada y McDiarmid (1989) prueba que, bajo las condiciones del corolario 3.1.6,

$$P[\mid Z - E(Z) \mid > t] \leq 2e^{-\frac{2t^2}{\sum_{i=1}^{n} c_i^2}}$$

En realidad, tampoco es necesario que g sea de diferencias acotadas, pues se puede aplicar el teorema y obtenemos:

Corolario 3.1.7. Si para g existen constantes $c_1, ..., c_n$ tales que:

$$sup_{x_1,..,x_n,x'_1,..x'_n} \sum_{i=1}^{n} \mid g(x_1, ..x_i, ..x_n) - g(x_1, ..x'_i, ..x_n) \mid^2 \leq \sum_{i=1}^{n} c_i^2$$

$$P[\mid Z - E(Z) \mid > t] \leq 2e^{-\frac{t^2}{4\sum_{i=1}^{n} c_i^2}}$$

Esta última es una condición mucho más débil.

Definición 3.1.8. Diremos que una función $f : V \longrightarrow R^n$ con $V \subset R^n$ acotada es de Lipschitz si existe una constante $K > 0$ tal que:

$$\parallel f(x) - f(y) \parallel \leq K \parallel x - y \parallel \forall x, y \in V$$

.

Si f es una función de Lipschitz podemos definir:

$$L = sup_{x \neq y, x, y \in V} \frac{\parallel f(x) - f(y) \parallel}{\parallel x - y \parallel} = inf\{K > 0 : \parallel f(x) - f(y) \parallel \leq K \parallel x - y \parallel \forall x, y \in V\} \tag{3.4}$$

Si tenemos una sucesión $\{f_n\}$ de funciones de Lipschitz tendremos definida también la sucesión $\{L_n\}$ de constantes definidas por 3.4.

En nuestro caso, trabajaremos con funciones de distribución empíricas las cuales no son de Lipschitz, ya que son funciones escalonadas.

Esto motiva la siguiente definición:

Definición 3.1.9. Una sucesión de funciones $f_n : V \longrightarrow R^d$ con $V \subset R^k$ acotado cumple la propiedad (P) si existen sucesiones L_n y $\gamma_n > 0$ tales que: existen $\delta > 0$ y $\lambda > 0$ tal que $\delta < L_n < \lambda$ para todo n y

$$\gamma_n \longrightarrow 0$$

y

$$\parallel f_n(x) - f_n(y) \parallel \leq L_n \parallel x - y \parallel + \gamma_n \forall x, y \in V.$$

Por ejemplo, si tenemos n variables aleatorias Bernoulli X_i y consideramos el promedio $\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$.

Si se cumple que para algún $K > 0$, $nP(\sum_{i=1}^{n} X_i = k) < K$, (esta condición implica que el histograma está uniformemente acotado) entonces para $0 \leq x \leq x' \leq 1$, la sucesión

de funciones de distribución de $\overline{X}$ cumple la propiedad (P).

$$F_n(x') - F_n(x) = \sum_{k=[nx]+1}^{[nx']} P(\sum_{i=1}^{n} X_i = k) \leq \sum_{k=[nx]+1}^{[nx']} \frac{K}{n} =$$

$$= \frac{K}{n}([nx'] - [nx]) \leq \frac{K}{n}(n(x' - x) + 1) = K[(x' - x) + \frac{1}{n}] = K(x' - x) + \frac{K}{n}$$

Consideremos ahora como estimador la media ponderada $\overline{X} = \sum_{i=1}^{n} \omega_i X_i$.

Representemos por $(X_1^j, \ldots, X_i^j, \ldots, X_n^j)$ $j = 1, \ldots, 2^n$ las posibles enuplas de valores que pueden tomarse y sea $I_n = \text{mín}_{j \neq k} \mid \sum_{j,k} \omega_i (X_i^j - X_i^k) \mid$.

Si $\frac{P(\sum \omega_i X_i = k)}{I_n} \leq K$ entonces,

$$F_n(x') - F_n(x) = \sum_{k \in (x, x']} P(\sum \omega_i X_i = k) \leq K I_n(\frac{x' - x}{I_n} + 1) = K(x' - x) + I_n.$$

Luego si

$$I_n \longrightarrow 0$$

se cumple la propiedad (P).

Observemos también que si X_n es una sucesión de variables discretas con función de distribución F_n y si existen sucesiones $f(n) > 0$, $g(n) > 0$ y $h(n) > 0$ y constantes K_1 y K_2 tales que:

$$sup_k\{P(X_n = k)\} \leq \frac{K_1}{f(n)}$$

con $k \in Rec\{X_n\}$, para todo $0 \leq x \leq x' \leq 1$

$$\sum_{k \in Rec\{X_n\}} \chi_{(x,x']}(k) \leq K_2 g(n)(x' - x) + h(n)$$

con $\frac{h(n)}{f(n)} \to 0$ y para algún $\delta > 0$ y $\lambda > 0$; $\delta < \frac{g(n)}{f(n)} < \lambda$ para todo n entonces:

$$F_n(x') - F_n(x) \leq sup_k\{P(X_n = k)\} \sum_{k \in Rec\{X_n\}} \chi_{(x,x']}(k) \leq$$

$$\leq \frac{K_1 K_2 g(n)}{f(n)}(x' - x) + \frac{h(n)}{f(n)}$$

lo que implica que F_n cumple la propiedad (P).

También tenemos que:

Lema 3.1.10. Consideremos una sucesión de estadísticos X_n con función de distribución F_n, que converge completamente a F, la cual es lipschitziana de constante L. Entonces F_n tiene la propiedad (P).

Demostración:
Consideremos $0 \leq x \leq x' \leq 1$.

$$\mid F_n(x') - F_n(x) \mid \leq \mid F_n(x') - F(x') \mid + \mid F(x') - F(x) \mid + \mid F(x) - F_n(x) \mid \leq$$

$$\leq L \parallel x - x' \parallel + 2 \sup_x |F_n(x) - F(x)|$$

Como $|F_n(x) - F(x)|$ está acotado para todo x, notemos por
$\gamma_n = \sup_x |F_n(x) - F(x)|$ entonces,

$$\mid F_n(x') - F_n(x) \mid \leq L \parallel x - x' \parallel + 2\gamma_n$$

donde $\gamma_n \to 0$ pues $F_n \to F$ completamente.

Si tomamos como estimador la media ponderada $\overline{X} = \sum_{i=1}^{n} \omega_i X_i$. donde las X_i son variables aleatorias iid.

Sea

$$a_n = \sum_{i=1}^{n} \omega_i^2$$

Utilizando el resultado de Fisher (1992) tenemos que si:

$$\lim_{t \to \infty} \sup \frac{1}{t} \sharp\{n : \frac{a_n}{\omega_n^2} < t\} < \infty$$

entonces,

$$\frac{S_n}{\sqrt{n}}$$

converge en distribución a la normal típica de donde las distribuciones de:

$$\sum_{i=1}^{n} \frac{\omega_i}{\sqrt{n}} X_i$$

tienen la propiedad (P).

Weber (2006) obtiene resultados combinando condiciones de integrabilidad de las variables aleatorias y condiciones sobre los pesos más débiles.

Consideremos de ahora en adelante en R^d la norma:

$$\|x\| = max_{i=1\ldots d}|x_i|$$

Lema 3.1.11. Sea $G_{g_n}(x)$ definida por 3.1. Supongamos que para las funciones $g_{n,l}$ existen constantes $c_{n,l,1}, ..., c_{n,l,n}$ tales que:

$$sup_{x_1,..,x_n,x'_1,..x'_n} \sum_{i=1}^{n} \mid g_{n,l}(x_1, ..x_i, ..x_n) - g_{n,l}(x_1, ..x'_i, ..x_n) \mid^2 \leq \sum_{i=1}^{n} c^2_{n,l,i} \tag{3.5}$$

para $1 \leq l \leq d$

Supongamos, además, que las $F_{n,l}$ correspondientes cumplen la propiedad (P), (L_n, γ_n) para toda l con $1 \leq l \leq n$.

Entonces para $\epsilon > 0$:

$$P(\| G_{g_n}(g_n(X)) - G_{g_n}(E(g_n(X))) \| > \epsilon) \leq \sum_{l=1}^{d} 2e^{-\frac{(\epsilon-\gamma_n)^2}{4L_n^2 \sum_{i=1}^{n} c^2_{n,l,i}}}$$

En particular si $c_n = \text{máx}_{1\leq l\leq d, 1\leq i\leq n} c_{n,l,i}$

$$P(\| G_{g_n}(g_n(X)) - G_{g_n}(E(g_n(X))) \| > \epsilon) \leq 2de^{-\frac{(\epsilon-\gamma_n)^2}{4L_n^2 n c_n^2}}$$

Observación: En el caso unidimensional

$$g_n(x_1, \ldots, x_n) = \overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}.$$

Ahora $\mid g_{n,l}(x_1, ..x_i, ..x_n) - g_{n,l}(x_1, ..x'_i, ..x_n) \mid = \frac{1}{n}$ entonces

$$sup_{x_1,..,x_n,x'_1,..x'_n} \sum_{i=1}^{n} \mid g_{n,l}(x_1, ..x_i, ..x_n) - g_{n,l}(x_1, ..x'_i, ..x_n) \mid^2 \leq \sum_{i=1}^{n} \frac{1}{n^2} = \frac{1}{n}$$

luego,

$$P(\| G_{g_n}(g_n(X)) - G_{g_n}(E(g_n(X))) \| > \epsilon) = P(\| G_{g_n}(\overline{X}) - G_{g_n}(E(\overline{X})) \| > \epsilon) \leq 2e^{-\frac{(\epsilon - \frac{K}{n})^2}{4K^2 n \frac{1}{n^2}}} = 2e^{-\frac{n(\epsilon - \frac{K}{n})}{4K^2}}$$

Demostración:

$$P(\| G_{g_n}(g_n(X)) - G_{g_n}(E(g_n(X))) \| > \epsilon) \leq$$

$$\leq P(\| g_n(X) - E(g_n(X)) \| > \frac{\epsilon - \gamma_n}{L_n}) =$$

$$= P(max_{1 \leq l \leq d} \mid g_{n,l}(X) - E(g_{n,l}(X)) \mid > \frac{\epsilon - \gamma_n}{L_n}) =$$

$$= P(\bigcup_{1 \leq l \leq d} \mid g_{n,l}(X) - E(g_{n,l}(X)) \mid > \frac{\epsilon - \gamma_n}{L_n}) \leq$$

$$\leq \sum_{l=1}^{d} P(\mid g_{n,l}(X) - E(g_{n,l}(X)) \mid > \frac{\epsilon - \gamma_n}{L_n}) =$$

$$= \sum_{l=1}^{d} \int_{[0,1]^d} P(\mid g_{n,l}(X) - E(g_{n,l}(X)) \mid > \frac{\epsilon - \gamma_n}{L_n} \mid \Theta = \theta) f_\theta(\theta) d\theta \leq$$

$$\leq \sum_{l=1}^{d} 2e^{-\frac{(\epsilon - \gamma_n)^2}{4L_n^2 \sum_{i=1}^{n} c_{n,l,i}^2}}$$

Si $c_n = \text{máx}_{1\leq l\leq d, 1\leq i\leq n} c_{n,l,i}$, entonces,

$$P(\parallel G_n(g_n(X)) - G_n(E(g_n(X))) \parallel > \epsilon) \leq \sum_{l=1}^{d} 2e^{-\frac{(\epsilon-\gamma_n)^2}{2L_n^2 nc_n^2}} = 2de^{-\frac{(\epsilon-\gamma_n)^2}{2L_n^2 nc_n^2}}$$

Lema 3.1.12. Si para toda función $g_{n,l}$ $1 \leq l \leq d$ se cumple la condición dada por la ecuación 3.21, y además,

$$\frac{\partial E(g_{n,l}(X)|\Theta_l)}{\partial \Theta_l} > M_l > 0. \tag{3.6}$$

Sea λ_n una sucesión positiva que tiende a 0, tal que para $0 < \alpha < 1/2$ $\text{máx}_l(\sum_{i=1}^{n} c_{n,l,i}^2)^\alpha = O(\lambda_n)$ entonces para todo $\theta \in [0,1]^d$:

$$\parallel G_{g_n}(E(g_n(X)|\Theta = \theta)) - \theta \parallel < \lambda_n \tag{3.7}$$

Demostración:

Sea $\Theta = (\Theta_1, ..., \Theta_d)$ y $\theta = (\theta_1, ..., \theta_d)$.

Tenemos que:

$$G_{g_n}(E(g_n(X)|\Theta = \theta)) =$$

$$= (P(g_{n,1} \leq E(g_{n,1}(X)|\Theta = \theta)), ..., P(g_{n,d} \leq E(g_{n,d}(X)|\Theta = \theta))) =$$

$$= (P(g_{n,1} \leq E(g_{n,1}(X)|\Theta_1 = \theta_1)), ..., P(g_{n,d} \leq E(g_{n,d}(X)|\Theta_d = \theta_d)))$$

Luego,

$$\parallel G_{g_n}(E(g_n(X)|\Theta = \theta)) - \theta \parallel =$$

$$= max_{1\leq l\leq d} |P(g_{n,l} \leq E(g_{n,l}(X)|\Theta_l = \theta_l)) - \theta_l|$$

Trabajando con la componente l-ésima y considerando que las marginales son uniformes:

$$P(g_{n,l} \leq E(g_{n,l}(X)|\Theta_l = \theta_l) =$$

$$= \int_0^1 P(g_{n,l} \leq E(g_{n,l}(X)|\Theta_l = \theta_l) \mid \Theta_l = \theta_l^*))d\theta_l^* =$$

$$= \int_0^{\theta_l + \frac{\lambda_n}{2}} P(g_{n,l} \leq E(g_{n,l}(X)|\Theta_l = \theta_l) \mid \Theta_l = \theta_l^*))d\theta_l^* +$$

$$+ \int_{\theta_l + \frac{\lambda_n}{2}}^1 P(g_{n,l} \leq E(g_{n,l}(X)|\Theta_l = \theta_l) \mid \Theta_l = \theta_l^*))d\theta_l^* \leq$$

$$\leq \theta_l + \frac{\lambda_n}{2} + \int_{\theta_l + \frac{\lambda_n}{2}}^1 P(g_{n,l} \leq E(g_{n,l}(X)|\Theta_l = \theta_l) \mid \Theta_l = \theta_l^*))d\theta_l^*$$

Tenemos que:

$$\theta_l^* - \theta_l > \frac{\lambda_n}{2}$$

$$E(g_{n,l}(X)|\Theta_l = \theta_l) - E(g_{n,l}(X)|\Theta_l = \theta_l^*) < 0$$

pues $E(g_{n,l}(X)|\Theta_l)$ es creciente.

Además,

$$E(g_{n,l}(X)|\Theta_l = \theta_l^*) - E(g_{n,l}(X)|\Theta_l = \theta_l) =$$

$$= [\frac{\partial E(g_{n,l}(X)|\Theta_l)}{\partial \theta_l}](c)(\theta_l^* - \theta_l) > M_l \frac{\lambda_n}{2}$$

con $c \in (\theta_l, \theta_l^*)$, de donde

$$g_{n,l}(X) - E(g_{n,l}(X)|\Theta_l = \theta_l^*) \leq E(g_{n,l}(X)|\Theta_l = \theta_l) - E(g_{n,l}(X)|\Theta_l = \theta_l^*)$$

si y solo si

$$\mid g_{n,l}(X) - E(g_{n,l}(X)|\Theta_l = \theta_l^*) \mid \geq E(g_{n,l}(X)|\Theta_l = \theta_l^*) - E(g_{n,l}(X)|\Theta_l = \theta_l) > M_l \frac{\lambda_n}{2}$$

Luego,

$$\int_{\theta_l + \frac{\lambda_n}{2}}^{1} P(g_{n,l}(X) \leq E(g_{n,l}(X)|\Theta_l = \theta_l) \mid \Theta_l = \theta_l^*))d\theta_l^* =$$

$$= \int_{\theta_l + \frac{\lambda_n}{2}}^{1} P(g_{n,l}(X) - E(g_{n,l}(X)|\Theta_l = \theta_l^*) \leq E(g_{n,l}(X)|\Theta_l = \theta_l) - E(g_{n,l}(X)|\Theta_l = \theta_l^*) \mid \Theta_l = \theta_l^*))d\theta_l^* <$$

$$< \int_{\theta_l + \frac{\lambda_n}{2}}^{1} P(\mid g_{n,l}(X) - E(g_{n,l}(X)|\Theta_l = \theta_l^*) \mid > M_l \frac{\lambda_n}{2} \mid \Theta_l = \theta_l^*)d\theta_l^* \leq$$

$$\leq 2e^{-\frac{M_l^2 \lambda_n^2}{16 \sum_{i=1}^{n} c_{n,l,i}^2}} (1 - \theta_l - \frac{\lambda_n}{2})$$

De esto obtenemos que:

$$P(g_{n,l} \leq E(g_{n,l}(X)|\Theta_l = \theta_l)) < \theta_l + \frac{\lambda_n}{2} + 2e^{-\frac{M_l^2 \lambda_n^2}{16 \sum_{i=1}^{n} c_{n,l,i}^2}} (1 - \theta_l - \frac{\lambda_n}{2})$$

$$P(g_{n,l} \leq E(g_{n,l}(X)|\Theta_l = \theta_l)) - \theta_l < \frac{\lambda_n}{2} + 2e^{-\frac{M_l^2 \lambda_n^2}{16 \sum_{i=1}^{n} c_{n,l,i}^2}}$$

para todo l con $1 \leq l \leq d$.

Así pues, como

$$\max_l (\sum_{i=1}^{n} c_{n,l,i}^2)^{\alpha} = O(\lambda_n)$$

con $0 < \alpha < 1/2$,

existe K, tal que

$$\frac{\text{máx}_l \left(\sum_{i=1}^n c_{n,l,i}^2\right)^\alpha}{\lambda_n} < K$$

entonces para todo l, tenemos

$$\lambda_n > \frac{1}{K}(\sum_{i=1}^n c_{n,l,i}^2)^\alpha$$

de donde

$$\frac{M_l^2 \lambda_n^2}{16 \sum_{i=1}^n c_{n,l,i}^2} > \frac{M_l^2}{16K^2}(\sum_{i=1}^n c_{n,l,i}^2)^{2\alpha-1}$$

A partir de esto

$$e^{-\frac{M_l^2 \lambda_n^2}{16 \sum_{i=1}^n c_{n,l,i}^2}} < e^{-\frac{M_l^2}{16K^2}(\sum_{i=1}^n c_{n,l,i}^2)^{2\alpha-1}}$$

y luego

$$\frac{e^{-\frac{M_l^2 \lambda_n^2}{16 \sum_{i=1}^n c_{n,l,i}^2}}}{\lambda_n} < K \frac{e^{-\frac{M_l^2}{16K^2}(\sum_{i=1}^n c_{n,l,i}^2)^{2\alpha-1}}}{(\sum_{i=1}^n c_{n,l,i}^2)^\alpha} \longrightarrow 0$$

ya que $2\alpha - 1 < 0$ y $\sum_{i=1}^n c_{n,l,i}^2 \longrightarrow 0$.

Así pues para n suficientemente grande:

$$2e^{-\frac{M_l^2 \lambda_n^2}{16 \sum_{i=1}^n c_{n,l,i}^2}} < \frac{\lambda_n}{2}$$

Entonces para $n \geq n_0$:

$$P(g_{n,l} \leq E(g_{n,l}(X)|\Theta_l = \theta_l)) - \theta_l < \lambda_n$$

Teorema 3.1.13. Supongamos que se cumplen las hipótesis del lema 3.1.11 y del lema 3.1.12 para $0 < \alpha < 1/2$

$$\frac{\gamma_n}{\text{máx}_l(\sum_{i=1}^n c_{n,l,i}^2)^\alpha} \to 0$$

1) Si

$$\max_l(\sum_{i=1}^n c_{n,l,i}^2)^{-1} < N < e^{\frac{32L_n^2+1}{144K^2L_n^2} \text{máx}_l(\sum_{i=1}^n c_{n,l,i}^2)^{\alpha-\frac{1}{2}}}$$

$$max_{k=1,\ldots,N} \parallel \widehat{\Theta}_{n,k} - \Theta_{n,k} \parallel \longrightarrow 0$$

en probabilidad.

2) Si además

$$\lim \log n \max_{l}(\sum_{i=1}^{n} c_{n,l,i}^2)^{1-2\alpha} = 0,$$

entonces

$$max_{k=1,\ldots,N} \parallel \widehat{\Theta}_{n,k} - \Theta_{n,k} \parallel \longrightarrow 0$$

completamente (en particular con probabilidad 1).

Demostración:

Consideremos una sucesión $\epsilon_n \to 0$ positiva, tal que $\max_l(\sum_{i=1}^{n} c_{n,l,i}^2)^{\alpha} = O(\epsilon_n)$. (siendo K una de las cotas).

$$P(max_{k=1,\ldots,N} \parallel \widehat{\Theta}_{n,k} - \Theta_{n,k} \parallel > \epsilon_n) = P(\bigcup_{k=1}^{N} \parallel \widehat{\Theta}_{n,k} - \Theta_{n,k} \parallel > \epsilon_n) \leq$$

$$\leq \sum_{k=1}^{N} P(\parallel \widehat{\Theta}_{n,k} - \Theta_{n,k} \parallel > \epsilon_n) = \sum_{k=1}^{N} P(\parallel \widehat{G}_N(g_n(X_k)) - \Theta_{n,k} \parallel > \epsilon_n) \leq$$

$$\leq \sum_{k=1}^{N} P(\parallel \widehat{G}_N(g_n(X_k)) - G_{g_n}(g_n(X_k)) \parallel > \frac{\epsilon_n}{3}) +$$

$$+ \sum_{k=1}^{N} P(\parallel G_{g_n}(g_n(X_k)) - G_{g_n}(E(g_n(X_k))) \parallel > \frac{\epsilon_n}{3}) +$$

$$+ \sum_{k=1}^{N} P(\parallel G_{g_n}(E(g_n(X_k))) - \Theta_{n,k} \parallel > \frac{\epsilon_n}{3})$$

Ahora:

$$P(\| \widehat{G}_N(g_n(X_k)) - G_{g_n}(g_n(X_k)) \| > \frac{\epsilon_n}{3}) =$$

$$= P(max_{l=1,\dots,d} \mid \widehat{G}_{N,l}(g_n(X_k)) - G_{g_{n,l}}(g_{n,l}(X_k)) \mid > \frac{\epsilon_n}{3}) \leq$$

$$\leq \sum_{l=1}^{d} P(\mid \widehat{G}_{N,l}(g_n(X)) - G_{g_{n,l}}(g_n(X_k)) \mid > \frac{\epsilon_n}{3}) =$$

$$= \sum_{l=1}^{d} P(\mid \widehat{F}_{N,l}(g_n(X_k)) - P(g_{n,l}(X) \leq g_{n,l}(X_k)) \mid > \frac{\epsilon_n}{3}) \leq dCe^{-2\frac{\epsilon_n^2 N}{9}}$$

Para el segundo término, aplicamos el lema 3.1.11.

$$P(\| G_{n,k}(g_n(X)) - G_{n,k}(E(g_n(X))) \| > \frac{\epsilon_n}{3}) \leq 2\sum_{l=1}^{d} e^{-\frac{(\frac{\epsilon_n}{3}-\gamma_n)^2}{4L_n^2 \sum_{i=1}^{n} c_{n,l,i}^2}}$$

El último término, se anula para $n \geq n_0$.
Así hemos obtenido que:

$$P(max_{k=1,\dots,N} \| \widehat{\Theta}_{n,k} - \Theta_{n,k} \| > \epsilon_n) \leq NdCe^{-2\frac{\epsilon_n^2 N}{9}} + 2N\sum_{l=1}^{d} e^{-\frac{(\frac{\epsilon_n}{3}-\gamma_n)^2}{4L_n^2 \sum_{i=1}^{n} c_{n,l,i}^2}}. \quad (3.8)$$

o sea,

$$P(max_{k=1,\dots,N} \| \widehat{\Theta}_{n,k} - \Theta_{n,k} \| > \epsilon_n) \leq Nd\left(Ce^{-2\frac{\epsilon_n^2 N}{9}} + 2e^{-\frac{\epsilon_n^2(1-3\frac{\gamma_n}{\epsilon_n})^2}{36L_n^2 \max_l \sum_{i=1}^{n} c_{n,l,i}^2}}\right). \quad (3.9)$$

Si se cumple cualquiera de las dos condiciones, entonces:

$$e^{-2\frac{\epsilon_n^2 N}{9}} < e^{-\frac{2\,\text{máx}_l(\sum_{i=1}^n c_{n,l,i}^2)^{2\alpha-1}}{9K^2}}$$

Ya que para n suficientemente grande $3\frac{\gamma_n}{\epsilon_n} < 1/2$ para el segundo término tenemos:

$$e^{-\frac{\epsilon_n^2(1-3\frac{\gamma_n}{\epsilon_n})^2}{36L_n^2\,\text{máx}_l\sum_{i=1}^n c_{n,l,i}^2}} < e^{-\frac{\text{máx}_l(\sum_{i=1}^n c_{n,l,i}^2)^{2\alpha-1}}{144K^2L_n^2}}$$

Así pues,

$$P(max_{k=1,\ldots,N} \parallel \widehat{\Theta}_{n,k} - \Theta_{n,k} \parallel > \epsilon_n) < N\left(C_1 e^{-\frac{32L_n^2+1}{144K^2L_n^2}\,\text{máx}_l(\sum_{i=1}^n c_{n,l,i}^2)^{2\alpha-1}}\right)$$

En el caso 1:

$$P(max_{k=1,\ldots,N} \parallel \widehat{\Theta}_{n,k} - \Theta_{n,k} \parallel > \epsilon_n) < C_1 e^{-\frac{32L_n^2+1}{144K^2L_n^2}(\text{máx}_l(\sum_{i=1}^n c_{n,l,i}^2)^{2\alpha-1} - \text{máx}_l(\sum_{i=1}^n c_{n,l,i}^2)^{\alpha-\frac{1}{2}})}$$

o sea,

$$P(max_{k=1,\ldots,N} \parallel \widehat{\Theta}_{n,k} - \Theta_{n,k} \parallel > \epsilon_n) \to 0$$

lo cual implica la convergencia en probabilidad

Si se cumple 2, obtenemos que:

$$\lim \frac{\frac{32L_n^2+1}{144K^2L_n^2}(\text{máx}_l(\sum_{i=1}^n c_{n,l,i}^2)^{2\alpha-1} - \text{máx}_l(\sum_{i=1}^n c_{n,l,i}^2)^{\alpha-\frac{1}{2}})}{\log n} = +\infty$$

lo que implica la convergencia completa pues si tenemos una serie:

$$\sum_{n=1}^{\infty} e^{-f(n)}$$

y

$$Q(n) = \frac{f(n)}{\log n} \to +\infty$$

entonces la serie converge, pues,

$$e^{-f(n)} = e^{\log n^{-Q(n)}} = n^{-Q(n)}$$

y si

$$Q(n) \to +\infty$$

la

$$\sum_{n=1}^{\infty} n^{-Q(n)}$$

es convergente.

Las series exponenciales convergen y la convergencia con probabilidad 1 se deduce de aplicar el lema de Borel Cantelli.

Observación: Si consideramos el caso unidimensional con $g_n(X) = \overline{X}$ y

$$nP\left(\sum_{i=1}^{n} X_i = k\right) < L$$

vemos que: $\gamma_n = \frac{1}{n}$ y $c_n = \frac{L}{n}$, por lo que la condición pedida se cumple ya que

$$\frac{1/n}{1/n^{\alpha}} \to 0$$

.

Si consideramos ϵ_n tal que $\frac{1}{n} = O(\epsilon_n)$ entonces si $n < N < e^{\frac{32L^2+1}{144K^2L^2} n^{\frac{1}{2}-\alpha}}$ y dado que

$$\frac{logn}{n^{1-2\alpha}} \to 0$$

entonces,

$$max_{k=1,\ldots,N} \parallel \widehat{\Theta}_{n,k} - \Theta_{n,k} \parallel \to 0$$

completamente.

Por la forma de los estimadores se pueden producir empates; romperemos los empates sumando a $g_{n,l}$ una variable aleatoria uniforme.

Consideremos una sucesión a_n por ejemplo $a_n = \text{mín}_l \, \text{mín}_{X,X'} \, |g_{n,l}(X) - g_{n,l}(X')|$ con X y X' posibles vectores de respuestas. La sucesión debe cumplir $a_n < \text{máx}_l(\sum_{i=1}^{n} c_{n,l,i}^2)$ Sea:

$$Z_n = g_n(X) + W$$

con W una variable aleatoria con distribución uniforme $U[0, a(n)]^d$ y consideremos las empíricas:

$$\widehat{J}_{N,l}(x) = \frac{\#\{Z_{n,l} \leq x\}}{N} = \frac{\sum_{i=1}^{N} \chi_{\{Z_{n,l} \leq x\}}}{N}$$

con $l = 1, .., d$. Definamos la función:

$$\widehat{T}_N(x_1, ..., x_d) = (\widehat{J}_{N,1}(x_1), ..., \widehat{J}_{N,d}(x_d)) \tag{3.10}$$

Entonces:

$$\widehat{\Theta}^*_n = \widehat{T}_N(Z_n) \tag{3.11}$$

Lema 3.1.14. Supongamos que para $0 < \alpha < 1/2$

$$\frac{\gamma_n}{\text{máx}_l(\sum_{i=1}^n c^2_{n,l,i})^\alpha} \to 0$$

Si $\epsilon_n \to 0$ es una sucesión positiva, tal que:

$$\max_l(\sum_{i=1}^n c^2_{n,l,i})^\alpha = O(\epsilon_n)$$

(siendo K una de las cotas).

Entonces existe una constante C_1 independiente de n , tal que:

$$P(\| \widehat{\Theta}_n - \widehat{\Theta}^*_n \| > \epsilon_n) \leq dC_1 e^{\frac{-2\epsilon_n^2 N}{9}}$$

Demostración:

$$P(\| \widehat{\Theta}_n - \widehat{\Theta}^*_n \| > \epsilon_n) = P(\| \widehat{G}_N(g_n(X)) - \widehat{T}_N(Z_n) \| > \epsilon_n) =$$

$$= P(max_{1 \leq l \leq d} \mid \widehat{F}_{N,l}(g_n(X)) - \widehat{J}_{N,l}(Z_n) \mid > \epsilon_n).$$

Trabajemos con cada componente.

$$\widehat{F}_{N,l}(x - a_n) = \frac{\#\{g_{n,l}(X) \leq x - a_n\}}{N} \leq \frac{\#\{Z_{n,l} \leq x\}}{N} = \widehat{J}_{N,l}(x) \leq \widehat{J}_{N,l}(x + a_n) =$$

$$= \frac{\#\{Z_{n,l} \leq x + a_n\}}{N} \leq \frac{\#\{g_{n,l}(X) \leq x + a_n\}}{N} = \widehat{F}_{N,l}(x + a_n)$$

Luego

$$\widehat{F}_{N,l}(g_{n,l}(X) - a_n) \leq \widehat{J}_{N,l}(Z_n) \leq \widehat{F}_{N,l}(g_{n,l}(X) + a_n)$$

De aquí,

$$\mid \widehat{J}_{N,l}(Z_n) - \widehat{F}_{N,l}(g_{n,l}(X)) \mid \leq \widehat{F}_{N,l}(g_{n,l}(X) + a_n) - \widehat{F}_{N,l}(g_{n,l}(X) - a_n)$$

A partir de esto,

$$P(\mid \widehat{J}_{N,l}(Z_n) - \widehat{F}_{N,l}(g_{n,l}(X)) \mid > \epsilon_n) \leq$$

$$P(\widehat{F}_{N,l}(g_{n,l}(X) + a_n) - \widehat{F}_{N,l}(g_{n,l}(X) - a_n) > \epsilon_n) \leq$$

$$\leq P(\mid \widehat{F}_{N,l}(g_{n,l}(X) + a_n) - F_{N,l}(g_{n,l}(X) + a_n) \mid > \frac{\epsilon_n}{3}) +$$

$$P(\mid F_{N,l}(g_{n,l}(X) + a_n) - F_{N,l}(g_{n,l}(X) - a_n) \mid > \frac{\epsilon_n}{3}) +$$

$$P(\mid \widehat{F}_{N,l}(g_{n,l}(X) - a_n) - F_{N,l}(g_{n,l}(X) - a_n) \mid > \frac{\epsilon_n}{3})$$

Sabemos que:

$$P(\mid \widehat{F}_{N,l}(g_{n,l}(X) + a_n) - F_{N,l}(g_{n,l}(X) + a_n) \mid > \frac{\epsilon}{3}) \leq Ce^{\frac{-2\epsilon_n^2 N}{9}}$$

y

$$P(\mid \widehat{F}_{N,l}(g_{n,l}(X) - a_n) - F_{N,l}(g_{n,l}(X) - a_n) \mid > \frac{\epsilon}{3}) \leq Ce^{\frac{-2\epsilon_n^2 N}{9}}$$

$$F_{N,l}(g_{n,l}(X) + a_n) - F_{N,l}(g_{n,l}(X) - a_n) \leq L_n(g_{n,l}(X) + a_n - g_{n,l}(X) + a_n) + \gamma_n =$$

$$2a_n L_n + \gamma_n$$

Pero para $n > n_0$ tenemos que:

$$\gamma_n < \frac{\text{máx}_l(\sum_{i=1}^{n} c_{n,l,i}^2)^\alpha}{6K} < \frac{\varepsilon_n}{6}$$

y

$$a(n) < \frac{\varepsilon_n}{12L_n}$$

luego para $n > n_0$, el último término se anula, así pues tenemos que existe C_1 tal que para todo n:

$P(\parallel \widehat{\Theta}_n - \widehat{\Theta}_n^* \parallel > \epsilon_n) \leq dC_1 e^{\frac{-2\epsilon_n^2 N}{9}}$

Obtenemos como corolario que:

Corolario 3.1.15. Supongamos que para $0 < \alpha < 1/2$

$$\frac{\gamma_n}{\text{máx}_l(\sum_{i=1}^n c_{n,l,i}^2)^\alpha} \to 0.$$

1) Si $\text{máx}_l(\sum_{i=1}^n c_{n,l,i}^2)^{-1} < N < e^{\frac{32L_n^2+1}{144K^2L_n^2}\text{máx}_l(\sum_{i=1}^n c_{n,l,i}^2)^{\alpha-\frac{1}{2}}}$

$$max_{k=1,\ldots,N} \parallel \widehat{\Theta}_{n,k}^* - \Theta_{n,k} \parallel \longrightarrow 0$$

en probabilidad.

2) Si además $\lim \log n \, \text{máx}_l(\sum_{i=1}^n c_{n,l,i}^2)^{1-2\alpha} = 0$, entonces

$$max_{k=1,\ldots,N} \parallel \widehat{\Theta}_{n,k}^* - \Theta_{n,k} \parallel \longrightarrow 0$$

completamente (en particular con probabilidad 1).

Demostración:

Consideremos una sucesión $\epsilon_n \to 0$ positiva, tal que:

$$\text{máx}_l(\sum_{i=1}^n c_{n,l,i}^2)^\alpha = O(\epsilon_n)$$

(siendo K una de las cotas).

$$\parallel \widehat{\Theta}_{n,k}^* - \Theta_{n,k} \parallel \leq \parallel \widehat{\Theta}_{n,k}^* - \widehat{\Theta}_{n,k} \parallel + \parallel \widehat{\Theta}_{n,k} - \Theta_{n,k} \parallel$$

luego

$$P(\parallel \widehat{\Theta}_{n,k}^* - \Theta_{n,k} \parallel > \varepsilon_n) \leq P(\parallel \widehat{\Theta}_{n,k}^* - \widehat{\Theta}_{n,k} \parallel > \varepsilon_n) + P(\parallel \widehat{\Theta}_{n,k} - \Theta_{n,k} \parallel > \varepsilon_n) \leq$$

$$\leq 2dCe^{\frac{-2\epsilon_n^2 N}{9}} + P(\parallel \widehat{\Theta}_{n,k} - \Theta_{n,k} \parallel > \varepsilon_n)$$

Aplicamos el teorema 3.1.13 a los términos

$$dC_1 e^{\frac{-2\epsilon_n^2 N}{9}}$$

y

$$P(\parallel \widehat{\Theta}_{n,k} - \Theta_{n,k} \parallel > \varepsilon_n)$$

de donde obtenemos el resultado.

Lema 3.1.16. Consideremos que se cumplen las hipótesis del teorema 3.1.13 y que el núcleo K es simétrico acotado y con soporte compacto C_K.

Supongamos que las CCI son Lipchitz de constante M_V en un conjunto V compacto con $V \subset [0,1]^d$.

Además para $0 < \alpha < 1/2$

$$N > \max_l (\sum_{i=1}^{n} c_{n,l,i}^2)^{-\alpha(d+2)}$$

$$\max_l (\sum_{i=1}^{n} c_{n,l,i}^2)^{\alpha} = O(h_n)$$

Existe un entero n_0 tal que para todo $n > n_0$ la siguiente desigualdad ocurre para todo ítem i y $\theta_0 \in U \subset V$, U abierto.

$$P(\mid \widehat{P}_{n,i}(\theta_0) - P_{n,i}(\theta_0) \mid > 3M_V h_n r) \leq e^{-8\frac{m_l^2 M_V^2 N_n T h_n^{d+2}}{9\|K\|_\infty^2 L^2 r^{d-2}}} + N_n (C_1 e^{-\frac{32L_n^2+1}{144K^2 L_n^2} \max_l (\sum_{i=1}^{n} c_{n,l,i}^2)^{2\alpha-1}}) \tag{3.12}$$

donde $r = max\{\| x \| \; x \in C_K\}$

Demostración:
Como

$$\frac{\sum_{k=1}^{N_n} K(\frac{\theta - \widehat{\theta}_{n,i,k}}{h_n})}{N_n h_n^d} \tag{3.13}$$

tiene límite finito con probabilidad 1. Existen m_l y m_u tal que para n suficientemente grande y $\theta \in (0,1)^d$ que:

$$m_l N_n h_n^d < \sum_{k=1}^{N_n} K(\frac{\theta - \widehat{\theta}_{n,i,k}}{h_n}) < m_u N_n h_n^d \tag{3.14}$$

en un conjunto de probabilidad 1.
Sabemos que $K(\frac{\theta - \widehat{\theta}}{h_n}) > 0$ si y solo si $\frac{\theta - \widehat{\theta}}{h_n} \in C_K$.

Como C_K es compacto $C_K \subseteq B_K = \overline{B(0,r)}$ con $r = max\{\| x \| \; x \in C_K\}$

Sea:

$$A_{n,i,\theta} = \{k : \frac{\theta - \widehat{\theta}_{n,i,k}}{h_n} \in B_K\} = \{k : \widehat{\theta}_{n,i,k} \in \theta - h_n B_K\} = \{k : \widehat{\theta}_{n,i,k} \in \overline{B(\theta, h_n r)}\}$$

entonces,

$$\#A_{n,i,\theta} \to N_n \mu_{\widehat{\theta}}(\overline{B(\theta, h_n r)}) \tag{3.15}$$

donde $\mu_{\widehat{\theta}}$ es la distribución de $\widehat{\theta}$
así pues en todo abierto U de $(0,1)^d$ existe n_0 tal que para todo

$\theta \in U$ y $n > n_0$, $|\#A_{n,i,\theta} - N_n \mu_{\widehat{\theta}}(B(\theta, h_n r))| < \epsilon$ de donde

$$\frac{N_n}{2}\mu_{\widehat{\theta}}(B(\theta, h_n r)) < \#A_{n,i,\theta} < \frac{3N_n}{2}\mu_{\widehat{\theta}}(B(\theta, h_n r)) \tag{3.16}$$

Sea $\theta_0 \in U$ y el suceso

$$B_{n,i} = \{\widehat{\theta_j} - \theta_j \in h_n B_K para j = 1, ..., N_n\}$$

$$P(\mid \widehat{P}_{n,i}(\theta_0) - P_{n,i}(\theta_0) \mid > 3M_V h_n r) =$$

$$= P(\mid \widehat{P}_{n,i}(\theta_0) - P_{n,i}(\theta_0) \mid > 3M_V h_n r | B_{n,i}) P(B_{n,i}) +$$

$$+ P(\mid \widehat{P}_{n,i}(\theta_0) - P_{n,i}(\theta_0) \mid > 3M_V h_n r | B^c_{n,i}) P(B^c_{n,i})$$

luego,

$$P(\mid \widehat{P}_{n,i}(\theta_0) - P_{n,i}(\theta_0) \mid > 3M_V h_n r) \leq$$

$$\leq P(\mid \widehat{P}_{n,i}(\theta_0) - P_{n,i}(\theta_0) \mid > 3M_V h_n r | B_{n,i}) + P(B^c_{n,i})$$

Sabemos que:

$$P(B^c_{n,i}) = P(max_{k=1,...,N} \parallel \widehat{\Theta}_{n,k} - \Theta_{n,k} \parallel > h_n r) < N\left(C_1 e^{-\frac{32L_n^2+1}{144K^2L_n^2}\,\text{máx}_l(\sum_{i=1}^n c_{n,l,i}^2)^{2\alpha-1}}\right)$$

ya que:

$$\max_l(\sum_{i=1}^n c_{n,l,i}^2)^\alpha = O(h_n)$$

Tenemos, entonces:

$$P(\mid \widehat{P}_{n,i}(\theta_0) - P_{n,i}(\theta_0) \mid > 3M_V h_n r) \leq$$

$$\leq P(\mid \widehat{P}_{n,i}(\theta_0) - P_{n,i}(\theta_0) \mid > 3M_V h_n r | B_{n,i}) + N\left(C_1 e^{-\frac{32L_n^2+1}{144K^2L_n^2}\,\text{máx}_l(\sum_{i=1}^{n} c_{n,l,i}^2)^{2\alpha-1}}\right)$$

Sea $\Theta_n = (\Theta_{n,1}, ..., \Theta_{n,N_n})$ y análogamente $\widehat{\Theta}_{n,i}$ y sean θ_n y $\widehat{\theta}_{n,i}$ realizaciones de estos vectores.

Sea W la distribución conjunta de estos vectores condicionado a $B_{n,i}$

Tenemos que:

$$P(\mid \widehat{P}_{n,i}(\theta_0) - P_{n,i}(\theta_0) \mid > 3M_V h_n r | B_{n,i}) =$$

$$= \int P(\mid \widehat{P}_{n,i}(\theta_0) - P_{n,i}(\theta_0) \mid > 3M_V h_n r | B_{n,i}, \widehat{\theta}_{n,i}, \theta_n) dW(\widehat{\theta}_{n,i}, \theta_n)$$

Ahora:

$$P(\mid \widehat{P}_{n,i}(\theta_0) - P_{n,i}(\theta_0) \mid > 3M_V h_n r | B_{n,i}, \widehat{\theta}_{n,i}, \theta_n) \leq$$

$$\leq P(\mid \widehat{P}_{n,i}(\theta_0) - E(\widehat{P}_{n,i}(\theta_0)) \mid + \mid E(\widehat{P}_{n,i}(\theta_0)) - P_{n,i}(\theta_0) \mid > 3M_V h_n r | B_{n,i}, \widehat{\theta}_{n,i}, \theta_n)$$

Esta esperanza condicional la podemos calcular como:

$$E(\widehat{P}_{n,i}(\theta_0) | B_{n,i}, \widehat{\theta}_{n,i}, \theta_n) = \frac{1}{\sum_{j \in A_{n,i,\theta_0}} K(\frac{\theta_0 - \widehat{\theta}_{n,i,j}}{h_n})} \sum_{j \in A_{n,i,\theta_0}} K(\frac{\theta_0 - \widehat{\theta}_{n,i,j}}{h_n}) E(X_{n,i,k} | \theta_{n,j}) =$$

$$= \frac{1}{\sum_{j \in A_{n,i,\theta_0}} K(\frac{\theta_0 - \widehat{\theta}_{n,i,j}}{h_n})} \sum_{k \in A_{n,i,\theta_0}} K(\frac{\theta_0 - \widehat{\theta}_{n,i,k}}{h_n}) P_{n,j}(\theta_{n,k})$$

.

Para todo $k \in A_{n,i,\theta_0}$

$$\| \theta_0 - \theta_{n,k} \| \leq \| \theta_0 - \widehat{\theta}_{n,-i,k} \| + \| \widehat{\theta}_{n,-i,k} - \theta_{n,k} \| \leq h_n r + h_n r = 2h_n r$$

Eligiendo n suficientemente grande para que

$$2h_n r < \text{ínf}\{\| x - y \| \; x \in \partial U y \in \partial V\}$$

entonces que

$$\| \theta_0 - \theta_{n,k} \| \leq 2h_n r$$

implica que para todo $k \in A_{n,i,\theta_0}$, $\theta_{n,k} \in V$.

Luego, por ser las CCI Lipchitz en V:

$$|P_{n,i}(\theta_0) - P_{n,j}(\theta_{n,k})| \leq M_V \| \theta_0 - \theta_{n,k} \| \leq 2M_V h_n r$$

Esto implica que:

$$| \; E(\widehat{P}_{n,i}(\theta_0) | B_{n,i}, \widehat{\theta}_{n,i}, \theta_n) - P_{n,i}(\theta_0) \; | =$$

$$= | \; \frac{1}{\sum_{k \in A_{n,i,\theta_0}} K(\frac{\theta_0 - \widehat{\theta}_{n,i,k}}{h_n})} \sum_{k \in A_{n,i,\theta_0}} K(\frac{\theta_0 - \widehat{\theta}_{n,i,k}}{h_n})[P_{n,j}(\theta_{n,k}) - P_{n,i}(\theta_0)] \; | < 2M_V h_n r$$

Entonces para n suficiente grande:

$$P(| \; \widehat{P}_{n,i}(\theta_0) - P_{n,i}(\theta_0) \; | > 3M_V h_n r | B_{n,i}, \widehat{\theta}_{n,i}, \theta_n) \leq$$

$$\leq P(| \; \widehat{P}_{n,i}(\theta_0) - E(\widehat{P}_{n,i}(\theta_0)) \; | > M_V h_n r | B_{n,i}, \widehat{\theta}_{n,i}, \theta_n)$$

Una expresión alternativa para $\widehat{P}_{n,i}(\theta_0)$ viene dada por:

$$\widehat{P}_{n,i}(\theta_0) = \frac{1}{\# A_{n,i,\theta_0}} \sum_{j \in A_{n,i,\theta_0}} w_j X_{n,i,k}$$

con $w_j = \frac{\#A_{n,i,\theta_0} K(\frac{\theta_0 - \widehat{\theta}_{n,i,j}}{h_n})}{\sum_{j \in A_{n,i,\theta_0}} K(\frac{\theta_0 - \widehat{\theta}_{n,i,j}}{h_n})}$

A partir de 3.14 y 3.16 y como K es acotado para n suficientemente grande se tiene:

$$w_j \leq \frac{\frac{3N_n \| K \|_\infty}{2} \mu_{\widehat{\theta}}(B(\theta, h_n r))}{m_l N_n h_n^d} \leq \frac{3 r^d h_n^d \parallel K \parallel_\infty L}{2 m_l h_n^d} = \frac{3 r^d \parallel K \parallel_\infty L}{2 m_l}$$

donde aplicamos el resultado debido a Devroye que:

$$\frac{\mu_{\widehat{\theta}}(B(\theta, h_n r))}{h_n^d r^d}$$

tiene límite finito para casi todo θ con respecto a μ cuando h tiende a 0.

Por lo tanto,

$$T h_n^d r^d < \mu_{\widehat{\theta}}(B(\theta, h_n r)) < L h_n^d r^d$$

Aplicando Hoeffding se obtiene:

$$P(\mid \widehat{P}_{n,i}(\theta_0) - E(\widehat{P}_{n,i}(\theta_0)) \mid > M_V h_n r | B_{n,i}, \widehat{\theta}_{n,i}, \theta_n) =$$

$$P(\mid \frac{1}{\#A_{n,i,\theta_0}} \sum_{k \in A_{n,i,\theta_0}} w_k Y_{n,i,k} - \frac{1}{\#A_{n,i,\theta_0}} \sum_{k \in A_{n,i,\theta_0}} w_k E(X_{n,i,k} | \theta_{n,k}) \mid > M_V h_n r | B_{n,i}, \widehat{\theta}_{n,i}, \theta_n) =$$

$$\leq P(\mid \sum_{k \in A_{n,i,\theta_0}} [X_{n,i,k} - E(X_{n,i,k} | \theta_{n,k})] \mid > \#A_{n,i,\theta_0} \frac{2 m_l M_V h_n}{3 \parallel K \parallel_\infty L r^{d-1}} | B_{n,i}, \widehat{\theta}_{n,i}, \theta_n)$$

$$\leq 2e^{\frac{-8 \#A_{n,i,\theta_0} m_l^2 M_V^2 h_n^2}{9 \| K \|_\infty^2 L^2 r^{2(d-1)}}} \leq 2e^{-8 \frac{m_l^2 M_V^2 N_n T h_n^{d+2}}{9 \| K \|_\infty^2 L^2 r^{d-2}}}$$

Así pues, para n suficientemente grande y para todo $\theta_0 \in U$:

$$P(\mid \widehat{P}_{n,i}(\theta_0) - P_{n,i}(\theta_0) \mid > 3 M_V h_n r) \leq e^{-8 \frac{m_l^2 M_V^2 N_n T h_n^{d+2}}{9 \| K \|_\infty^2 L^2 r^{d-2}}} + N_n \left(C_1 e^{-\frac{32 L_n^2 + 1}{144 K^2 L_n^2} \max_l (\sum_{i=1}^n c_{n,l,i}^2)^{2\alpha - 1}} \right)$$

Teorema 3.1.17. Dadas las hipótesis anteriores y para $\theta_0 \in (0,1)^d$ fijo y

$$N > \eta \frac{9 \parallel K \parallel_\infty^2 L^2 r^{d-2}}{m_l^2 M_V^2 T} \max_l (\sum_{i=1}^n c_{n,l,i}^2)^{-\alpha(d+2)}$$

$$max\{ECM(\widehat{P}_{n,i}(\theta_0)) : 1 \leq i \leq n\} \leq O(\max_l(\sum_{i=1}^{n} c_{n,l,i}^2)^{2\alpha}) + e^{-\eta} \tag{3.17}$$

Observación:
Si suponemos

$$N > \max_l(\sum_{i=1}^{n} c_{n,l,i}^2)^{-\frac{d+2}{2}}$$

entonces

$$max\{ECM(\widehat{P}_{n,i}(\theta_0)) : 1 \leq i \leq n\} \leq O(\max_l(\sum_{i=1}^{n} c_{n,l,i}^2)^{2\alpha}) \tag{3.18}$$

Demostración:

$$ECM(\widehat{P}_{n,i}(\theta_0)) = E((\widehat{P}_{n,i}(\theta_0) - P_{n,i}(\theta_0))^2) =$$

$$= \int_0^1 P[(\widehat{P}_{n,i}(\theta_0) - P_{n,i}(\theta_0))^2 > t]dt \leq$$

$$\leq (3M_V h_n r)^2 + P[(\widehat{P}_{n,i}(\theta_0) - P_{n,i}(\theta_0))^2 > (3M_V h_n r)^2]$$

Esto es

$$ECM(\widehat{P}_{n,i}(\theta_0)) \leq (3M_{c,d}C_k h_n)^2 + P[|\widehat{P}_{n,i}(\theta_0) - P_{n,i}(\theta_0))| > 3M_V h_n r)]$$

Aplicando el lema anterior tenemos para n suficientemente grande y todo ítem i:

$$ECM(\widehat{P}_{n,i}(\theta_0)) \leq (3M_V h_n r)^2 + e^{-8\frac{m_l^2 M_V^2 N_n T h_n^{d+2}}{9\|K\|_\infty^2 L^2 r^{d-2}}} + N_n \left(C_1 e^{-\frac{32L_n^2+1}{144K^2L_n^2} \max_l(\sum_{i=1}^{n} c_{n,l,i}^2)^{2\alpha-1}} \right)$$

$$\leq O(\max_l(\sum_{i=1}^{n} c_{n,l,i}^2)^{2\alpha}) + e^{-\eta}$$

En el caso unidimensional como $0 < \alpha < 1/2$ si se elige α próximo a $1/2$ entonces el ECM se hace cercano a $O(\frac{1}{n})$.

Tenemos entonces el siguiente resultado:

Teorema 3.1.18. Dadas las hipótesis anteriores, suponiendo además que el núcleo es Lipchitz de constante J y además

$N < Mn^r$ con $r > d + 1$

$$sup_{\theta \in (a,b)^d} max_{1 \leq i \leq n} \mid \widehat{P}_{n,i}(\theta) - P_{n,i}(\theta) \mid \to 0$$

si $n \to +\infty$ completamente

con $(a, b)^d \subset (0, 1)^d$

Demostración:

Consideremos una grilla de $2^d N^{2d}$ puntos tal que la distancia (con la norma del máximo) entre dos puntos sea menor o igual a $1/N^2$. Sea θ un punto arbitrario de $[0, 1]^d$ entonces existe k tal que la distancia de θ_k^n a θ no es mayor a $1/N^2$.

Para todo θ tenemos

$$|P_{n,i}(\theta) - \widehat{P}_{n,i}(\theta)| \leq |P_{n,i}(\theta) - P_{n,i}(\theta_k^n)| + |P_{n,i}(\theta_k^n) - \widehat{P}_{n,i}(\theta_k^n)| +$$

$$+ |\widehat{P}_{n,i}(\theta_k^n) - \widehat{P}_{n,i}(\theta)|$$

Consideremos

$$|\widehat{P}_{n,i}(\theta_k^n) - \widehat{P}_{n,i}(\theta)| =$$

$$= |\frac{\sum_{i=1}^N K(\frac{\theta_k^n - \widehat{\theta}_i}{h}) X_{ij}}{\sum_{i=1}^N K(\frac{\theta_k^n - \widehat{\theta}_i}{h})} - \frac{\sum_{i=1}^N K(\frac{\theta - \widehat{\theta}_i}{h}) X_{ij}}{\sum_{i=1}^N K(\frac{\theta - \widehat{\theta}_i}{h})}| \leq |\frac{\sum_{i=1}^N K(\frac{\theta - \widehat{\theta}_i}{h}) X_{ij}}{\sum_{i=1}^N K(\frac{\theta - \widehat{\theta}_i}{h})} - \frac{\sum_{i=1}^N K(\frac{\theta - \widehat{\theta}_i}{h}) X_{ij}}{\sum_{i=1}^N K(\frac{\theta_k^n - \widehat{\theta}_i}{h})}| +$$

$$+ |\frac{\sum_{i=1}^N K(\frac{\theta - \widehat{\theta}_i}{h}) X_{ij}}{\sum_{i=1}^N K(\frac{\theta_k^n - \widehat{\theta}_i}{h})} - \frac{\sum_{i=1}^N K(\frac{\theta_k^n - \widehat{\theta}_i}{h}) X_{ij}}{\sum_{i=1}^N K(\frac{\theta_k^n - \widehat{\theta}_i}{h})}| =$$

$$\frac{\sum_{i=1}^{N} K(\frac{\theta-\widehat{\theta}_i}{h})X_{ij}}{\sum_{i=1}^{N} K(\frac{\theta-\widehat{\theta}_i}{h})\sum_{i=1}^{N} K(\frac{\theta_k^n-\widehat{\theta}_i}{h})}|\sum_{i=1}^{N} K(\frac{\theta_k^n-\widehat{\theta}_i}{h})-\sum_{i=1}^{N} K(\frac{\theta-\widehat{\theta}_i}{h})|+$$

$$+\frac{1}{|\sum_{i=1}^{N} K(\frac{\theta_k^n-\widehat{\theta}_i}{h})|}|\sum_{i=1}^{N} K(\frac{\theta-\widehat{\theta}_i}{h})X_{ij}-\sum_{i=1}^{N} K(\frac{\theta_k^n-\widehat{\theta}_i}{h})X_{ij}|$$

$$\leq \frac{2\sum_{i=1}^{N}|K(\frac{\theta_k^n-\widehat{\theta}_i}{h})-K(\frac{\theta-\widehat{\theta}_i}{h})|}{\sum_{i=1}^{N} K(\frac{\theta_k^n-\widehat{\theta}_i}{h})} \leq \frac{2\sum_{i=1}^{N} J||\theta_k^n-\theta||}{h\sum_{i=1}^{N} K(\frac{\theta_k^n-\widehat{\theta}_i}{h})}$$

Como

$$\sum_{i=1}^{N} K(\frac{\theta_k^n-\widehat{\theta}_i}{h}) > Nh^d m_l$$

obtenemos que:

$$|\widehat{P}_{n,i}(\theta_k^n)-\widehat{P}_{n,i}(\theta)| < \frac{2JN\|\theta_k^n-\theta\|}{Nh^{d+1}m_l} < \frac{2J}{N^2h^{d+1}m_l}$$

Como existe C_1 tal que:

$$h > C_1 \max_l(\sum_{i=1}^{n} c_{n,l,i}^2)^{\alpha}$$

y

$$N > \max_l(\sum_{i=1}^{n} c_{n,l,i}^2)^{-\alpha(d+2)}$$

$$|\widehat{P}_{n,i}(\theta_k^n)-\widehat{P}_{n,i}(\theta)| = o(\frac{1}{n})$$

A partir de

$$sup_{\theta\in[0,1]^d} \mid \frac{\partial P_i}{\partial \theta_j} \mid < c_j$$

para todo j entonces si $c = max_j c_j$ con j=1,...,d

$$|P_{n,i}(\theta)-P_{n,i}(\theta_k^n)| \leq c\|\theta-\theta_k^n\| \leq c\frac{1}{N^2}$$

Sea $\epsilon > 0$ mostraremos que:

$$\sum_{n=1}^{\infty} P(sup_{\theta\in(a,b)^d} max_{1\leq i\leq n} \mid \widehat{P}_{n,i}(\theta)-P_{n,i}(\theta) \mid > \epsilon) < \infty$$

Para probar esto es suficiente probar que:

$$\sum_{n=1}^{\infty}\sum_{i=1}^{n}\sum_{k=1}^{2^d N^{2d}} P[|P_{n,i}(\theta_k^n) - \widehat{P}_{n,i}(\theta_k^n)| > \epsilon/3] < \infty \tag{3.19}$$

ya que para $n > n_0$ suficientemente grande,

$$|P_{n,i}(\theta) - P_{n,i}(\theta_k^n)| \leq \frac{\epsilon}{3}$$

y

$$|\widehat{P}_{n,i}(\theta_k^n) - \widehat{P}_{n,i}(\theta)| \leq \frac{\epsilon}{3}$$

entonces

$$P(\sum_{n=1}^{\infty} P(sup_{\theta\in(a,b)^d} max_{1\leq i\leq n} \mid \widehat{P}_{n,i}(\theta) - P_{n,i}(\theta) \mid > \epsilon) <$$

$$\sum_{n=1}^{\infty}\sum_{i=1}^{n}\sum_{k=1}^{2^d N^{2d}} P[|P_{n,i}(\theta_k^n) - \widehat{P}_{n,i}(\theta_k^n)| > \epsilon/3]$$

Para este n_0 la triple suma es menor que:

$$2^d n_0^2 N_{n_0}^{2d} + \sum_{n>n_0}^{\infty}\sum_{i=1}^{n}\sum_{k=1}^{2N_n^{2d}} P[|P_{n,i}(\theta_k^n) - \widehat{P}_{n,i}(\theta_k^n)| > 3M_V h_n r] \tag{3.20}$$

Se puede aplicar el lema 3.1.16 y entonces 3.20 es menor que:

$$2^d n_0^2 N_{n_0}^{2d} + \sum_{n>n_0}^{\infty} 2^d n N_n^{2d}(e^{-8\frac{m_l^2 M_V^2 N_n T h_n^{d+2}}{9\|K\|_\infty^2 L^2 r^{d-2}}} + N_n(C_1 e^{-\frac{32L_n^2+1}{144K^2L_n^2}\,\text{máx}_l(\sum_{i=1}^n c_{n,l,i}^2)^{2\alpha-1}}))$$

Reemplazando N_n por su cota superior Mn^r entonces lo anterior queda acotado por

$$2^d n_0^2 N_{n0}^{2d} + \sum_{n>n_0}^{\infty} 2^d M^{2d} n^{2dr+1}(e^{-8\frac{m_l^2 M_V^2 T\,\text{máx}_l(\sum_{i=1}^n c_{n,l,i}^2)^{-[d(1/2-\alpha)+1-2\alpha]}}{9\|K\|_\infty^2 L^2 r^{d-2}}} +$$

$$+Mn^r(C_1 e^{-\frac{32L_n^2+1}{144K^2L_n^2}\,\text{máx}_l(\sum_{i=1}^n c_{n,l,i}^2)^{2\alpha-1}}))$$

Como $0 < \alpha < 1/2$ y

$$\lim \log n \max_l (\sum_{i=1}^{n} c_{n,l,i}^2)^{1-2\alpha} = 0$$

la suma anterior es finita.

Ahora considerando en conjunto los resultados anteriores

Teorema 3.1.19. Si se cumplen las hipótesis

1) Independencia Local

2) El rasgo Θ tiene distribuciones marginales uniformes $U[0,1]$

3) para las funciones $g_{n,l}$ existen constantes $c_{n,l,1}, ..., c_{n,l,n}$ tales que:

$$sup_{x_1,..,x_n,x_1',..x_n'} \sum_{i=1}^{n} \mid g_{n,l}(x_1,..x_i,..x_n) - g_{n,l}(x_1,..x_i',..x_n) \mid^2 \leq \sum_{i=1}^{n} c_{n,l,i}^2 \qquad (3.21)$$

para $1 \leq l \leq d$

4) Las $F_{n,l}$ correspondientes cumplen la propiedad (P), (L_n, γ_n) para toda l con $1 \leq l \leq n$

5)

$$\frac{\partial E(g_{n,l}(X)|\Theta_l)}{\partial \Theta_l} > M_l > 0. \qquad (3.22)$$

6) Para $0 < \alpha < 1/2$

$$\frac{\gamma_n}{\max_l (\sum_{i=1}^{n} c_{n,l,i}^2)^{\alpha}} \to 0$$

7) $H \max_l (\sum_{i=1}^{n} c_{n,l,i}^2)^{-\alpha(d+2)} < N < Mn^r$ con $r > d+1$

8)

$$\lim \log n \max_l (\sum_{i=1}^{n} c_{n,l,i}^2)^{1-2\alpha} = 0$$

9) El núcleo K es simétrico, acotado y con soporte compacto

10) Las CCI tienen derivadas parciales continuas en $(0,1)^d$ y en cada campacto $[a,b]^d \subset (0,1)^d$ son acotadas

11) Para $0 < \alpha < 1/2$

$$\max_l (\sum_{i=1}^{n} c_{n,l,i}^2)^{\alpha} = O(h_n)$$

12) El núcleo K es Lipchitz de constante J

entonces

$$max\{max_{k=1,...,N} \parallel \widehat{\Theta}_{n,k} - \Theta_{n,k} \parallel, sup_{\theta \in (a,b)^d} max_{1 \leq i \leq n} \mid \widehat{P}_{n,i}(\theta) - P_{n,i}(\theta) \mid\} \to 0$$

si $n \to +\infty$ con probabilidad 1

Demostración:

Sale directamente de los teoremas 2.22.1 y 3.1.18 notando que la probabilidad del conjunto donde falla esta convergencia es la unión de dos conjuntos cada uno con probabilidad 0.

3.2. Estimación monótona

Veremos ahora un estimador no paramátrico para las CCI monótonas crecientes. Como ya dijimos esta suposición es usual en TRI. Nos basaremos en el método de Dette, Neumeyer y Pilz (2006) que proponen estimar funciones monótonas a partir de un estimador no monótono. Los autores consideran funciones m derivables, estrictamente crecientes en $[0, 1]$.

Si consideramos una muestra $U_1, ..., U_T$ de variables aleatorias independientes con distribución uniforme en $[0, 1]$, sea K_d un núcleo y h_d el ancho de ventana, entonces

$$\frac{1}{TH_d} \sum_{i=1}^{T} K_d(\frac{m(U_i) - u}{h_d})$$

es el estimador de la densidad de $m(U)$.

La densidad de $m(U)$ es $(m^{-1})'(u)\chi_{[m(0),m(1)]}(u)$ luego,

$$\frac{1}{Th_d} \int_{-\infty}^{t} \sum_{i=1}^{T} K_d(\frac{m(U_i) - u}{h_d})du$$

es un estimador consistente de m^{-1} en t.

Entonces en el contexto de nuesto enfoque consideremos una grilla $0, \frac{1}{T}, .., \frac{i}{T}, .., 1$ y consideremos el estimador no paramétrico de la CCI en cada punto:

$$\widehat{P}(\frac{i}{T}) = \frac{\sum_{j=1}^{N} K_r(\frac{\frac{i}{T} - \widehat{\theta}_j}{h_r})Y_j}{\sum_{j=1}^{N} K_r(\frac{\frac{i}{T} - \widehat{\theta}_j}{h_r})}$$

donde K_r y h_r indican el núcleo y la ventana utilizados para la regresión.

Entonces, el estimador monótono de la inversa de la CCI en el punto θ será:

$$\widehat{P_m^{-1}}(\theta) = \frac{1}{Th_d} \int_{-\infty}^{\theta} \sum_{i=1}^{T} K_d(\frac{\widehat{P}(\frac{i}{T}) - u}{h_d})du$$

El estimador de $\widehat{P_m}$ se obtiene mediante la reflexión de $\widehat{P_m^{-1}}$ respecto a la recta $y = x$

Para ilustrar el caso multidimensional veremos como proceder en dos dimensiones Supongamos una CCI creciente en cada rasgo y que el estimador de la CCI se obtuvo como se vio previamente.

$$\widehat{P}_i(\theta_1, \theta_2) = \frac{\sum_{k=1}^{N} K_r(\frac{\widehat{\theta_{1i}} - \theta_1}{h_1}, \frac{\widehat{\theta_{2i}} - \theta_2}{h_2}) X_{ik}}{\sum_{k=1}^{N} K_r(\frac{\widehat{\theta_{1k}} - \theta}{h_1}, \frac{\widehat{\theta_{2k}} - \theta}{h_2})} \tag{3.23}$$

donde e K_r es un núcleo de dimensión dos, con soporte compacto $C \subset [0,1]^2$, y cumple las hipótesis requeridas

Tomemos una grilla $0, \frac{1}{T}, .., \frac{i}{T}, .., 1$ para θ_1 y una grilla $0, \frac{1}{T}, .., \frac{j}{T}, .., 1$ para θ_2

Para cada $\theta_2 \in (0,1)$ fijo, consideremos el estimador

$$\widehat{H^{-1}}(\theta_1|\theta_2) = \frac{1}{Th_d} \int_{-\infty}^{\theta_1} \sum_{i=1}^{T} K_d(\frac{\widehat{P}(\frac{i}{T}, \theta_2) - u}{h_d}) du$$

La función $\widehat{H^{-1}}(\theta_1|\theta_2)$ es estrictamente creciente en θ_1 para cada θ_2 fijo.

Podemos calcular el inverso (como función de θ_1) para obtener $\widehat{H}(\theta_1|\theta_2)$.

Para cada $\theta_1 \in (0,1)$ fijo, calculamos

$$\widehat{P^{*-1}}(\theta_1, \theta_2) = \frac{1}{Th_d} \sum_{j=1}^{T} \int_{-\infty}^{\theta_2} K_d(\frac{H(\theta_1|\frac{j}{T}) - u}{h_d}) du$$

Finalmente, $\widehat{P^*}(\theta_1, \theta_2)$ se calcula mediante la inversa (como función de θ_2).

El algoritmo sería:

Paso 1 Usando el procedimeinto estandar para cada $1 \leq i \leq T$, y $1 \leq j \leq T$ calcular:

$$\widehat{P}(\frac{i}{T}, \frac{j}{T})$$

Paso 2 Para cada $\frac{j}{T}$ con $1 \leq j \leq T$, calcular $H^{-1}(\theta_1|\frac{j}{T})$ mediante

$$\widehat{H^{-1}}(\theta_1|\frac{j}{T}) = \frac{1}{Th_d} \int_{-\infty}^{\theta_1} \sum_{i=1}^{T} K_d(\frac{\widehat{P}(\frac{i}{T}, \frac{j}{T}) - u}{h_d}) du$$

Paso 3 Invirtiendo con respecto a θ_1, obtenemos $\widehat{H}(\theta_1|\frac{j}{T})$ with $1 \leq j \leq T$

Paso 4 Calcular

$$\widehat{P^{*-1}}(\theta_1, \theta_2) = \frac{1}{Th_d} \sum_{j=1}^{T} \int_{-\infty}^{\theta_2} K_d(\frac{H(\theta_1|\frac{j}{T}) - u}{h_d}) du$$

Paso 5 Invirtiendo respecto a θ_2, obtenemos $\widehat{P^*}(\theta_1, \theta_2)$

3.3. Aplicación al funcionamiento diferencial del ítem

3.3.1. Definiciones básicas

Supongamos que tenemos un test que mide un rasgo al que denotaremos por Θ y llamaremos habilidad principal. Diremos que existe DIF si la puntuación obtenida es función no solo del nivel que se tiene en el rasgo medido sino también de otras características secundarias (como puede ser la pertenencia a determinado grupo étnico, cultural, etcétera). Concretamente, tendrán un funcionamiento diferencial aquellos ítems cuya probabilidad de ser acertados, condicionados a un igual valor del rasgo que se pretende medir, difiere entre distintos grupos de la población. El número de grupos a comparar es variable aunque es común la comparación entre dos grupos a los que se denominan grupo focal (grupo de interés) que coincide generalmente con el grupo minoritario y el grupo de referencia, que es el que sirve como base de comparación. Para definir el DIF recurriremos a la teoría multidimensional, que dice que el DIF se da cuando se incumple el supuesto de unidimensionalidad. Si tenemos dos rasgos latentes θ (habilidad principal)y η (habilidades espúreas) habrá DIF si

$$E_\eta(P_1(U = 1/\Theta, \eta)/\Theta) \neq E_\eta(P_2(U = 1/\Theta, \eta)/\Theta) \tag{3.24}$$

Esto implica que las distribuciones de η condicionada a θ en las poblaciones son diferentes. Es importante notar que la multidimensionalidad por sí misma no es la causa de DIF sino las diferencias en las distribuciones condicionadas de las variables espúreas, es decir, la multidimensionalidad es condición necesaria pero no suficiente para la presencia de DIF. Debemos también definir el término impacto y diferenciarlo de funcionamiento diferencial de los ítems. Se llama impacto cuando hay diferencia en el desempeño de un ítem entre grupos y ésta es causada por una diferencia real en la variable medida.

Cabe distinguir entre dos tipos de DIF: uniforme y no uniforme. Se produce DIF uniforme cuando la probabilidad de responder correctamente un ítem es mayor para un grupo respecto de otro para todo valor de θ. En cambio, si existe una región donde la desigualdad se da en un sentido y en su complemento, se da la desigualdad contraria, entonces decimos que tenemos DIF no uniforme.

Akerman (1992) explica que se puede producir DIF uniforme si:

a) los grupos tienen diferentes medias en la habilidad principal y hay una correlación significativa entre la habilidad principal y la espúrea;

b) si existen diferencias en las medias en la habilidad espúrea entre grupos.

Puede ocurrir DIF no uniforme si:

a) la varianza de la habilidad espúrea no es la misma entre los grupos;

b) la correlación entre la habilidad principal y la espúrea difiere entre los grupos.

3.3.2. Técnicas estadísticas para la detección

Sunpogamos que agrupamos a los individuos según una variable V, notemos por Θ el rasgo latente que pretendemos medir mediante nuestra prueba y sea U la variable

aleatoria que representan las puntuaciones en el ítem.

Definición: diremos que el ítem no presenta DIF si

$$P(U/\Theta = \theta, V = v) = P(U/\Theta = \theta)\forall\theta, v$$

es decir, la distribución de U a un nivel de rasgo dado es independiente de la variable V.

Millsap y Everson (1993) clasifican los métodos de detección en torno a dos criterios:

1) el método de medida; y

2) el tipo de respuesta al ítem (dicotómica, politómica, continua, etcétera).

Fidalgo (1996) clasifica los métodos en:

a) procedimientos que no especifican ningún modelo de medida, donde coloca el procedimiento de Mantel-Haenszel, Estandarización, Sibtest, modelos loglineales, modelos logit y regresión logística; y

b) procedimientos basados en la TRI a ser medidas del área, ji cuadrado de Lord y comparación de modelos.

3.3.2.1. Mantel-Haenszel

Dividamos la puntuación en el test en k intervalos y notemos por G_R al grupo de referencia y por G_F al grupo focal. Considerando las respuestas obtenidas por cada grupo se obtienen así k tablas de contingencia 2 x 2. Sean A_k, B_k, C_k y D_k las frecuencias observadas y notemos por N_{Rk}, N_{Fk}, N_{1k} y N_{0k} las marginales. Consideramos la igualdad $\frac{A_k}{B_k} = \alpha_k \frac{C_k}{D_k}$. Si el ítem no presenta DIF es de esperar que $\alpha = 1$ para todo k luego formulamos la prueba:

$H_0 : \alpha_k = 1$ para todo k

$H_1 : \alpha_k \neq 1$ para algún k.

Para contrastar Mantel y Haenszel proponen el estadístico:

$$\chi^2_{MH} = \frac{(\mid \sum_{k=1}^{m} A_k - \sum_{k=1}^{m} \frac{N_{Rk}N_{1k}}{N_k} \mid -0,5)^2}{\sum_{k=1}^{m} \frac{N_{Rk}N_{Fk}N_{1k}N_{0k}}{N_k^2(N_k-1)}} \tag{3.25}$$

el cual sigue una distribución asintótica χ^2 con un grado de libertad. Como estimador de α se utiliza

$$\widehat{\alpha}_{MH} = \frac{\sum_{k=1}^{m} \frac{A_k D_k}{N_k}}{\sum_{k=1}^{m} \frac{B_k C_k}{N_k}} \tag{3.26}$$

Como vemos α varía entre 0 y ∞, es claro que el valor de $\alpha = 1$ implica que no hay DIF. Se puede transformar la escala a otra llamada delta mediante: $\Delta = -2,35 \ln[\widehat{\alpha}_{MH}]$. Si $\Delta \cong 0$ se tiene que no hay DIF en cambio si es negativo tenemos que el ítem favorece al grupo de referencia y si es positivo el favorecido es el grupo focal.

Como defecto principal se le critica a este procedimiento el hecho que puede no detectar el DIF no uniforme. Mazor, Clauser y Hambleton (1994) proponen un procedimiento para utilizar MH con el fin de detectar DIF no uniforme.

Algunas consideraciones acerca de este procedimiento que se hacen tanto en el artículo de Fidalgo (1994) como en el de Mazor, Clauser y Hambleton(1994) son:
1) el tamaño mínimo de la muestra es de 200 sujetos por grupo;
2) los ítems muy fáciles o muy difíciles así como los poco discriminativos tienen alta probabilidad de no ser detectados aunque presenten un DIF importante;
3) si se tienen k ítems las puntuaciones deben dividirse en k+1 intervalos. A medida que disminuye el número de categorías aumenta la cantidad de ítems falsamente identificados;
4) cuanto más se aparte el criterio utilizado para parear los sujetos de la unidimensionalidad mayor será el error tipo I.
En resumen el método MH posee un estimador del DIF, así como un test estadístico de significatividad. Con la modificación propuesta puede detectar el DIF no uniforme y se pueden utilizar los programas informáticos de Hambleton y Rogers; Fidalgo y Nandukamar para calcularlo.

3.3.2.2. Estandarización

Es útil para la descripción del DIF y se basa en la diferencia en la proporción de sujetos que aciertan el ítem en el grupo focal y en el de referencia.
Gráficas que presentan estas diferencias condicionadas indican el monto de DIF que tienen un ítem.
Un índice cuantitativo utilizado es la diferencia de proporciones estandarizadas (DPE) dada por:

$$DPE = \frac{um_{k=1}^{m} W_k (P_{Fk} - P_{Rk})}{\sum_{k=1}^{m} W_k} \tag{3.27}$$

donde W_k es un ponderador y $P_{Fk} = \frac{C_k}{N_{Fk}}$ y $P_{Rk} = \frac{A_k}{N_{Rk}}$ son las proporciones de respuestas correctas en el nivel de puntuación k.
Si se utiliza $W_k = N_{Fk}$ entonces

$$DPE = \frac{\sum_{k=1}^{m} N_{Fk} (P_{Fk} - P_{Rk})}{\sum_{k=1}^{m} N_{Fk}} \tag{3.28}$$

Dorans y Holland (1993) dan una fórmula para el error típico del DPE como en el caso anterior (ver artículo de Fidalgo ecuación 9.12) lo que permite realizar n test de hipótesis de ausencia de DIF mediante el estadístico $E = \frac{DPE}{ET(DPE)}$ en cual tiene una distribución $N(0, 1)$.
Este método no detecta el DIF no uniforme, existe un programa de Hambleton y Rogers que lo calcula.

3.3.2.3. SIBTEST

Este es un procedimiento para estudiar el DIF en uno o más ítems del test simultáneamente y utiliza un modelo de TRI multidimensional no paramétrico propuesto por Shealy

y Stout (1993a, 1993b).

Se dividen los ítems del test en dos grupos: el primero, denominado subtest válido (que incluye a los ítems insesgados) y el segundo llamado subtest estudiado (que incluye a aquellos ítems que pueden presentar DIF). Para determinar qué ítems constituyen cada test se recurre a procedimientos estadísticos, de esta manera se determina qué ítems no cumplen el supuesto de unidimensionalidad. Se agrupan los sujetos dentro del grupo focal y de referencia de acuerdo a su puntaje en el subtest válido para luego comparar su desempeño en el subtest estudiado.

Consideremos que tenemos un test de N ítems y que el subtest válido lo constituyen n de ellos. Sean X y Y las puntuaciones en el test válido y estudiado respectivamente. Sea además β_u el parámetro que mide el sesgo, el cual se puede calcular como:

$$\widehat{\beta_u} = \sum_{k=0}^{n} [\widehat{P_{Fk}}(\overline{Y_{Rk}} - \overline{Y_{Fk}})] \tag{3.29}$$

Luego el estadístico utilizado para el test de hipótesis de ausencia de sesgo (es decir $\beta_u = 0$) está dado por:

$$B = \frac{\widehat{\beta_u}}{\widehat{\sigma}(\widehat{\beta_u})}$$

donde $\widehat{\sigma}(\widehat{\beta_u})$ es el error estándar de β_u, el cual viene dado por la expresión:

$$\widehat{\sigma}(\widehat{\beta_u}) = \sqrt{\sum_{k=0}^{n} P_{Fk}^2 (\frac{1}{J_{Rk}} \sigma^2(Y/k, R) + \frac{1}{J_{Fk}} \sigma^2(Y/k, F))}$$

donde J_{Rk} y J_{Fk} son la cantidad de sujetos que tienen una puntuación igual a k en el subtest válido en cada grupo, y las varianzas que aparecen se calculan con los sujetos que tienen una puntuación k.

Este estadístico tiene una distribución asintótica N(0,1).

Los valores positivos de β_u indican función diferencial del test contra el grupo focal en tanto que negativos en contra del grupo de referencia.

Este procedimiento fue diseñado para detectar DIF uniforme, por lo tanto, no sirve para detectar DIF no uniforme y además debe aplicarse a tests de más de 25 ítems y que no presenten impacto. Un programa debido a Shealy, Stout y Roussos (1995) se puede utilizar para calcularlo.

3.3.2.4. Modelos loglineales

Para este tipo de análisis se utiliza una tabla de contingencia multidimensional del tipo HxGxR donde H indica nivel de habilidad, G los grupos comparados y R la respuesta al ítem. Notaremos a las frecuencias observadas mediante f_{ijk}. Se ajustan varios tipos de modelos jerárquicos para determinar la presencia o no de DIF y si este es uniforme o no uniforme.

Se ajustan tres modelos (F_{ijk} indican las frecuencias esperadas).

1) Modelo saturado que indicará DIF no uniforme:

$$\ln(F_{ijk}) = \lambda + \lambda_{H(i)} + \lambda_{G(j)} + \lambda_{R(k)} + \lambda_{HG(ij)} + \lambda_{HR(jk)} + \lambda_{GR(jk)} + \lambda_{HGR(ijk)}$$

2) Modelo para DIF uniforme:

$$\ln(F_{ijk}) = \lambda + \lambda_{H(i)} + \lambda_{G(j)} + \lambda_{R(k)} + \lambda_{HG(ij)} + \lambda_{HR(jk)} + \lambda_{GR(jk)}$$

3) Modelo que indica ausencia de DIF:

$$\ln(F_{ijk}) = \lambda + \lambda_{H(i)} + \lambda_{G(j)} + \lambda_{R(k)} + \lambda_{HG(ij)} + \lambda_{HR(jk)}$$

En el primero aparecen los efectos principales y todas las posibles interacciones, en el segundo modelo se eliminó la interacción entre el nivel de habilidad, los grupos y la respuesta. En el tercer modelo se elimina la interacción entre grupo y respuesta del modelo 2. La idea es estimar el modelo saturado y a partir de él eliminar términos en forma sucesiva a partir de los de mayor orden hasta quedarnos con el modelo que tenga todos los términos de orden superior significativos. Para decidir con qué modelo nos quedaremos, se utiliza el test de razón de verosimilitudes. La razón de verosimilitud está dada por:

$$LR^2 = 2\Sigma_{i=1}^{p}\Sigma_{j=1}^{q}\Sigma_{k=1}^{r} f_{ijk} \ln(\frac{f_{ijk}}{F_{ijk}}) \tag{3.30}$$

Para comparar el ajuste entre dos modelos usamos la diferencia de razones de verosimilitud que se distribuye χ^2 con tantos grados de libertad como la diferencia entre los grados de libertad de los modelos comparados.
Concretamente si el modelo que mejor ajusta es el tercero concluiremos que no hay DIF; si está presente la interacción entre la respuesta al ítem y el grupo habrá DIF y este será no uniforme o uniforme según esté o no presente la interacción entre habilidad, respuesta y grupo.
Este método es fácilmente implementable en el SPSS o BMDP.

3.3.2.5. Modelos logit

En los modelos logit se considera la variable respuesta al ítem como dependiente de los efectos inducidos por otras variables. Los modelos correspondientes son:

1- $\ln(\frac{F_{ij1}}{F_{ij2}}) = \zeta + \zeta_{H(i)} + \zeta_{G(j)} + \zeta_{HG(ij)}$
2- $\ln(\frac{F_{ij1}}{F_{ij2}}) = \zeta + \zeta_{H(i)} + \zeta_{G(j)}$
3- $\ln(\frac{F_{ij1}}{F_{ij2}}) = \zeta + \zeta_{H(i)}$
Al igual que el anterior los modelos logit se pueden implementar en SPSS o BMDP.

3.3.2.6. Regresión logística

Quienes propusieron esta técnica para detectar DIF fueron Spray y Carlson (1986). En el modelo de regresión logística se establece que:

$$P(Y = 1) = \frac{e^{\sum b_k X_{ik}}}{1 + e^{\sum b_k X_{ik}}}$$

Si llamamos Z_i al exponente entonces este tomará las formas:
1- $Z_i = \tau_0 + \tau_1 H_i + \tau_2 G_j + \tau_3 HG_{ij}$
2- $Z_i = \tau_0 + \tau_1 H_i + \tau_2 G_j$
3- $Z_i = \tau_0 + \tau_1 H_i$
Para determinar el modelo de mejor ajuste se usa el logaritmo de razón de verosimilitudes.

3.3.2.7. Basados en la TRI

Se han desarrollado varios índices basados en la teoría de respuesta al ítem para estudiar el posible DIF. En TRI la curva de respuesta al ítem se determina por curvas logísticas (de uno, dos o tres parámetros). Si para fijar la notación escribimos el modelo de dos parámetros tenemos:

$$P_i(\theta) = \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}} \tag{3.31}$$

donde b_i es el índice de dificultad del ítem y a_i es el índice de discriminación. Se da DIF uniforme cuando $a_R = a_F$ y $b_R \neq b_F$ y DIF no uniforme si $a_R \neq a_F$ y $b_R = b_F$ o $b_R \neq b_F$. Esto propone como estrategias para detectar el DIF la comparación de los parámetros de los ítems entre el grupo focal y el grupo de referencia o comparar el área entre las dos curvas características.

3.3.2.8. Medida del área

Notemos por $P_R(\theta)$ y $P_F(\theta)$ las curvas carácterísticas del ítem para el grupo de referencia y focal. Existirá DIF si la curvas son diferentes más allá de diferencias aleatorias.Una manera de medir esta diferencia es mediante el área entre las dos curvas (o sea la integral de la diferencia).

$$\int_{\theta_1}^{\theta_2} P_R(\theta) - P_F(\theta) d\theta \tag{3.32}$$

Al considerar la diferencia con signo se tiene un indicador de la dirección del DIF; pero en caso de DIF no uniforme debe utilizarse el valor absoluto de la diferencia para evitar cancelaciones espúreas. Así pues un indicador del DIF en este caso será

$$\int_{\theta_1}^{\theta_2} |P_R(\theta) - P_F(\theta)| d\theta \tag{3.33}$$

con θ_1 y θ_2 elegidos para que cubran un amplio rango del rasgo, por ejemplo -3 y 3 respectivamente. Las primeras medidas del área tomaron la forma de una suma discreta, por ejemplo Rudner propone:

$$R = \sum_{\theta=-3}^{\theta=3} |P_R(\theta) - P_F(\theta)| \Delta\theta$$

Kim y Cohen (1991) han encontrado fórmulas para las medidas del área con signo (ACCS) y sin signo (ACSS) para los modelos de uno, dos y tres parámetros. A modo de ejemplo reproduciremos la fórmula de la medida del área para el modelo de un parámetro:

$$ACSS = |\ln(\frac{[1+\exp(\theta_2 - b_R)]^{1/D}[1+\exp(\theta_1 - b_F)]^{1/D}}{[1+\exp(\theta_1 - b_R)]^{1/D}[1+\exp(\theta_2 - b_F)]^{1/D}})| \tag{3.34}$$

En Fidalgo (1996) se encuentran además las expresiones para los modelos 2P y 3P y su discusión según los valores de los parámetros.

Raju (1990) integró entre $-\infty$ y $+\infty$ obteniendo índices exactos para el área sin y con signo. Posteriormente, obtuvo las distribuciones asintóticas de los índices, lo cual nos permite hacer inferencia. Para el modelo de 3P obtiene para el área con signo (ACS):

$$ACS = (1-c)(b_F - b_R)$$

$$E(ACS) = (1-c)(b_F - b_R)$$

$$VAR(ACS) = (1-c)^2(Var(b_F) - Var(b_R))$$

$$Z(ACS) = \frac{(b_F - b_R)}{VAR(ACS)^{1/2}}$$

Los valores positivos de los índices del área con signo indican que el ítem es más fácil para el grupo de referencia y los valores negativos indican lo contrario. Los índices anteriores se pueden normalizar y así realizar test de significación para la detección del DIF.

Se pueden resumir los pasos necesarios para la aplicación de los criterios basados en la TRI en:
1) encontrar modelo de TRI que mejor ajuste a los datos;
2) estimar los parámetros de los ítems en el grupo focal y de referencia. Si además queremos encontrar la significación estadística de los índices se deberá calcular la matriz de varianzas y covarianzas de los ítems;
3) establecer una métrica común para los dos grupos (*equating*) calculando la pendiente y la ordenada en el origen que se necesitan para la equiparación;
4) obtener los índices y los tests de hipótesis.

3.3.2.9. Chi cuadrado de Lord

Este método compara los parámetros del ítem que define la CCI. Sea X_F y X_R el vector que contiene los parámetros de los ítems. Pueden tener dimensión 3x1, 2x1 o 1x1, según el modelo elegido. La hipótesis nula es la igualdad de ambos vectores, es decir,

$$H_0) x_F^{'} - x_R^{'} = 0$$

y se utilizará como estadístico de contraste a la forma cuadrática:

$$\chi^2 = (X_F - X_R)' \Sigma^{-1} (X_F - X_R)$$

donde Σ es la matriz de varianzas-covarianzas de $(X_F - X_R)$ con $\Sigma^{-1} = (S_R + S_F)^{-1}$ siendo las anteriores las matrices de varianzas y covarianzas de los parámetros del ítem. Este estadístico sigue una distribución χ^2 con p grados de libertad siendo p el número de parámetros del modelo.
En el caso de modelo de 1P se obtiene la ecuación:

$$\chi^2 = \frac{(b_F - b_R)}{Var(b_F) + Var(b_R)} \tag{3.35}$$

Es importante puntualizar que cuando se utiliza un modelo 3P, debido a los problemas que lleva la estimación del parámetro c, Lord (1980) propone el siguiente método:
1) ajustar el modelo de 3P tomando los dos grupos conjuntamente;
2) ajustar nuevamente los grupos por separado utilizando como parámetro c el hallado en el paso 1;
3) aplicar el estadístico χ^2.
Se le critica a este método el hecho que puede llevar a que se rechase la hipótesis nula, aun cuando exista poca diferencia entre las curvas en la región donde se encuentra la mayoría de los sujetos.

3.3.2.10. Comparación de modelos

El método se basa en la razón de verosimilitudes para modelos anidados. Específicamente si el modelo que incluye parámetros diferentes para el grupo focal y el de

referencia ajusta mejor que el modelo en el cual los parámetros de los ítems son iguales para ambos grupos. Esto indicaría que el ítem presenta DIF. La razón de verosimilitudes se distribuye como una χ^2 con tantos grados de libertad como la difrerencia entre el número de parámetros entre el modelo aumentado y el compacto bajo la hipótesis nula que los parámetros del modelo aumentados son iguales a 0. Supngamos que tenemos N ítems y que queremos determinar si hay DIF en el ítem k. Para esto determinamos dos modelos: uno que representa la ausencia de DIF y otro que especifica DIF.
Modelo 1 $\{a, b_1, b_2, .., b_k, .., b_n\}$
Modelo 2 $\{a, b_1, b_2, .., b_{kR}, b_{kF}, .., b_n\}$
En el modelo 1 se está imponiendo la restricción $b_{kR} = b_{kF} = b_k$.
Para probar la hipótesis nula de ausencia de DIF se siguen los siguientes pasos:
1) utilizando la totalidad de la muestra se estiman los parámetros del modelo 1 y se calcula el logaritmo de la verosimilitud $(-2 \ln L(1))$;
2) se estiman los parámetros del modelo 2 y el logaritmo de su verosimilitud $(-2 \ln L(2))$;
3) Se calcula ahora el estadístico:

$$LR = -2 \ln \frac{L(1)}{L(2)} = -2 \ln L(1) + 2 \ln L(2)$$

que tiene una distribución χ^2 con grados de libertad igual al número de restricciones impuestas.
Este procedimiento puede utilizarse para comprobar la presencia de DIF en varios ítems simultáneamente y luego si se rechaza la hipótesis nula debe hacerse un posterior análisis para ver cuáles ítems presentan DIF.

3.3.3. Procesos iterativos

Varios autores han propuesto procesos iterativos para evaluar el DIF dentro de la TRI.
Lord (1980) propone el siguiente:

1) Evaluar el DIF en forma habitual;
2) Eliminar los ítems que hayan presentado DIF en el paso 1 y reestimar θ;
3) Volver a evaluar el DIf con los θ calculados en 2.

Segall (1983) propone:
1) estimar los parámetros de los ítems para cada grupo;
2) hallar las constantes A y K que permiten equiparar los ítems;
3) luego de equiparar evaluar DIF en todos los ítems y eliminar aquellos que presenten DIF;
4) recalcular A y K solo con los ítems que no presentan DIF;
5) repetir los pasos 3 y 4 hasta que los ítems identificados con DIF no varíen de una iteración a otra.

Park (1988) propone un procedimiento combinado:

1) Se estiman los parámetos de los ítems para cada grupo. Se ponen en la misma escala y se detecta el DIF;

2) Se colocan nuevamente los ítems en una escala común pero ahora se utilizan para calcular las constantes A y K solo los ítems que no presentaron DIF en el paso 1. Se analiza el DIF nuevamente en todos los ítems. Se continúa este proceso hasta que no varíen los ítems con DIF de una iteración a otra;

3) Se estima θ en cada grupo usando solo los ítems que no presentaron DIF;

4) Usando las estimaciones del paso 3 se estiman los parámetros de todos los ítems dentro de cada grupo (purificación);

5) Se calculan nuevamente las constantes de equiparación utilizando solo los ítems que no presentaron DIF. Se equipara y se evalúa nuevamente el DIF. Se continúa este proceso hasta que no varíen los ítems con DIF entre iteraciones;

6) Si la clasificación de los ítems ha cambiado luego de comenzar el proceso de purificación se regresa al paso 4. Si no ha habido cambios se termina el proceso.

3.3.4. Implementación de decisiones

Mazor, Clauser y Hambleton (1994) describen varias consideraciones esenciales para una implementación apropiada de los procedimientos expuestos:

1) Para los métodos de pareo sugieren que sería mejor obtener un criterio externo ante uno interno, aunque es poco probable poder hacerlo. Cuando se utiliza un criterio interno debe purificarse luego de una primera identificación de los ítems con posible DIF recalculándolo sin utilizarlos. Sin embargo, Holland y Thayer (1988) recomiendan que en el recálculo se utilice el ítem estudiado aun cuando se hubiera identificado en una primera etapa con DIF. Cuando el test no es unidimensional y se utiliza el puntaje total como criterio, debe tenerse en cuenta la mutidimensionalidad y, por ejemplo, identificar subtests unidimensionales. Se debe tener en cuenta la fiabilidad de la variable que se utiliza para parear.

2) Tamaño muestral. Los métodos de detección son sensibles a los tamaños muestrales y se ha visto que pequeñas muestras no detectan DIF sustanciales. Muestras de entre 200 a 250 por grupo son adecuadas para MH, SIBTEST y regresión logística. Para IRT se necesitan muestras de por lo menos 500 sujetos por grupo.

3) DIF no uniforme. Si bien todos los procedimeintos detectan DIF uniforme, si se desea estudiar el DIF no uniforme se deberá utilizar un índice especificamente sensible a este tipo de DIF;

4) Características del ítem. Los ítems con baja discriminación, así como aquellos que son muy fáciles o difíciles, son los que con mayor facilidad no son detectados.

Fidalgo (1996) también da varios consejos para una correcto estudio del DIF:

a) El análisis del DIF debe realizarse dentro de un marco global cuyo objetivo es obtener tests eficientes.

b) El DIF es una condición necesaria pero no suficiente para que un ítem presente sesgo.
c) No deben utilizarse los métodos delta o el anova pues no diferencian entre DIF e impacto. Igualmente desaconseja los métodos ji-cuadrado.
d) Los métodos para detectar DIF no detectan el sesgo generalizado, o sea el caso en que la mayoría de los ítems están sesgados.
e) Los tests estadísticos deben ser complementados con índices del tamaño del DIF.
f) Los métodos basados en la TRI se deben aplicar utilizando métodos de equiparación iterativos y purificando la habilidad estimada de los sujetos.
g) En los métodos que no especifican ningún modelo de medida se debe utilizar el mayor número de categorías de puntuación posible.
h) Con tests cortos es preferible utilizar métodos basados en la TRI.
i) Si se utiliza el χ^2 de Lord se debe utilizar además alguna medida del área para ver el tamaño del DIF.
j) Si tenemos muestras grandes es conveniente utilizar niveles de significación bajos para reducir el error tipo I.

3.3.5. Método propuesto

A partir de la estimación no paramétrica de las CCI proponemos utilizar la medida del área entre las dos curvas estimadas. Trabajaremos en una dimensión aunque lo obtenido es fácilmente generalizabe a más dimensiones. Las dos medidas del área son:

$$\int_{-\infty}^{+\infty} \widehat{P_R(\theta)} - \widehat{P_R(\theta)} d\theta$$

y

$$\int_{-\infty}^{+\infty} \mid \widehat{P_R(\theta)} - \widehat{P_R(\theta)} \mid d\theta$$

Como ejemplo veamos la medida con signo

$$\widehat{P_R(\theta)} = \frac{\sum_{j \in R} K(\frac{\theta - \theta_j^R}{h}) Y_j}{\sum_{j \in R} K(\frac{\theta - \theta_j^R}{h})}$$

y

$$\widehat{P_F(\theta)} = \frac{\sum_{j \in F} K(\frac{\theta - \theta_j \omega F \omega \omega}{h}) Y_j}{\sum_{j \in F} K(\frac{\theta - \theta_j^F}{h})}$$

luego si consideramos a

$$\omega_j^R = \int_{-\infty}^{+\infty} \frac{K(\frac{\theta - \theta_j^R}{h})}{\sum_{j \in R} K(\frac{\theta - \theta_j^R}{h})} d\theta$$

y análogamente ω_j^F entonces

$$\int_{-\infty}^{+\infty} \widehat{P_R(\theta)} - \widehat{P_R(\theta)} d\theta = \sum_{j \in R} \omega_j^R Y_j - \sum_{j \in F} \omega_j^F Y_j$$

Si suponemos que medimos θ_j sin error la varianza de la estimación del área es:

$$\sum_{j \in R} (\omega_j^R)^2 P(\theta_j)(1 - P(\theta_j)) + \sum_{j \in F} (\omega_j^F)^2 P(\theta_j)(1 - P(\theta_j))$$

Luego podemos realizar un test para el DIF utilizando como estadístico a

$$Z = \frac{AREA}{\sqrt{\sum_{j \in R} (\omega_j^R)^2 P(\theta_j)(1 - P(\theta_j)) + \sum_{j \in F} (\omega_j^F)^2 P(\theta_j)(1 - P(\theta_j))}}$$

Capítulo 4

Simulaciones

4.1. Distancia entre curvas características del ítem

Trabajaremos con dos distancias entre las curvas características de los ítems. Sea $\mathcal{A}$ la familia de las CCI.

Definición 4.1.1. Definimos las distancias $d_i : \mathcal{A}x\mathcal{A} \to R$ con i=1,2,3 como:

$$d_1(P_1, P_2) = \{\int \mid P_1(\theta) - P_2(\theta) \mid^2 d\theta\}^{1/2}$$

Esta distancia es la raíz del error cuadrático integrado y lo denominaremos RISE.

Si tomamos la esperanza del error cuadrático medio obtenemos d_2 o RMSE.

$$d_2(P_1, P_2) = \{\int \mid P_1(\theta) - P_2(\theta) \mid^2 f(\theta)d\theta\}^{1/2}$$

Definimos la tercera distancia como la distancia del supremo:

$$d_3(P_1, P_2) = sup_\theta \mid P_1(\theta) - P_2(\theta) \mid$$

4.2. Generación de datos

Existen diversos modelos que permiten relacionar la habilidad del examinado (θ) con las distintas características de los ítems de un test. Llamaremos $\hat{\theta}$ a la estimación de la variable latente habilidad que es medida por los ítems de un test. Hay ciertos supuestos sobre estos modelos, por ejemplo, a medida que aumenta el valor de θ también aumenta la probabilidad de contestar correctamente a un ítem. Otro supuesto es el de unidimensionalidad el cual establece que solo una habilidad es medida por un test.

La probabilidad que el individuo j conteste correctamente el ítem i, bajo la condición de unidimensionalidad, en un modelo logístico de tres parámetros es:

$$P_i(\theta_j) = c_i + (1 - c_i)\frac{e^{Da_i(\theta_j - b_i)}}{1 + e^{Da_i(\theta_j - b_i)}} \qquad i = 1, 2, 3, ..., n \;\; j = 1, 2, 3, ..., N$$

- i hace referencia al ítem en cuestión, hay n ítems distintos.
- j hace referencia al individuo en cuestión, hay N individuos evaluados.
- a_i es el parámetro de discriminación del ítem i, es proporcional a la pendiente de la función de respuesta en el punto $\theta_j = b_i$
- b_i es el parámetro de dificultad del ítem i.
- c_i es el parámetro de azar del ítem i, representa la probabilidad de contestar correctamente para un nivel de habilidad bajo, es la asíntota de la función de respuesta cuando θ tiende a menos infinito.
- θ_j representa la habilidad del individuo j.
- $P_i(\theta_j)$ es la probabilidad del individuo j de contestar correctamente el ítem i.
- D es una constante igual a 1,7.

En muchos casos el supuesto de unidimensionalidad es violado, por ejemplo, un problema matemático puede requerir además de conocimientos matemáticos comprensión lectora. La probabilidad del individuo j de contestar correctamente el ítem i en un modelo logístico de tres parámetros k-dimensional es:

$$P(u_i = 1|\theta_j) = c_i + (1 - c_i)\frac{e^{Da_i'\theta_j - b_i)}}{1 + e^{Da_i'\theta_j - b_i)}} \qquad i = 1, 2, 3, ..., n \;\; j = 1, 2, 3, ..., N$$

En este modelo hay un parámetro de discriminación para cada dimensión del problema, pero un solo parámetro de dificultad.

Para generar tests simulamos las habilidades de los examinados, así como los valores de otros parámetros relacionados con los distintos ítems. Las habilidades (θ) las simulamos de distribuciones N(0;1) iid. En otras secciones evaluamos los efectos de que estas provengan de otras distribuciones. Los parámetros a, b y c los simulamos de variables aleatorias uniformes, y a menos que se especifique otra cosa serán: $a \sim U(0,5;2)$, $b \sim U(-3;3)$ y $c \sim U(0;0,25)$.

Con estos valores calculamos $P_i(\theta_j)$, luego para cada $P_i(\theta_j)$ sorteamos $u \sim Uniforme(0;1)$ si $P_i(\theta_j) > u$ la respuesta del individuo j al ítem i es correcta, en caso contrario incorrecta.

Las funciones utilizadas para la generación de datos se encuentran en la sección 4.10.2.

4.3. Ajuste de la regresión no paramétrica unidimensional a modelos conocidos y rasgos conocidos

En esta sección consideraremos distintas formas funcionales para las CCI y veremos el ajuste de la regresión no paramétrica a estas curvas. Como en este caso tenemos el rasgo conocido y no generamos un test sino los datos de la función, estamos trabajando con el modelo clásico de regresión no paramétrica. Si bien este no es el caso analizado, es interesante hacer esta simulación en primera instancia, para ver cómo se comporta nuestro método. El interés es entonces relativo, pero quisimos empezar con una descripción de lo que sucede cuando lo aplicamos a partir de datos de curvas y rasgos conocidos.

4.3.1. Método

Consideraremos cuatro tipos de curvas. Simularemos modelos logísticos de 1, 2 y 3 parámetros y modelos logit cúbicos, ya que estos modelos tienen puntos de inflexión y si se toman b y c positivos son crecientes.

Los modelos son:

1. $P(\theta) = \frac{1}{1+e^{(\theta-b)}}$
2. $P(\theta) = \frac{1}{1+e^{a(\theta-b)}}$
3. $P(\theta) = c + (1-c)\frac{1}{1+e^{a(\theta-b)}}$
4. $\log \frac{P(\theta)}{1-P(\theta)} = a + b\theta + c\theta^3$

Para la estimación de las regresiones no paramétricas usaremos la función de S-Plus ksmooth y utilizaremos un núcleo normal típico.

También, utilizaremos un ancho de banda del orden de $N^{-1/5}$, donde N es la cantidad de sujetos. Para medir el ajuste utilizaremos las distancias ya definidas.

4.3.1.1. Modelo 1P

Para analizar la bondad de ajuste en función de la cantidad de observaciones utilizamos tamaños muestrales de 100; 250; 500; 1000; 2500; 5000 y 10000 datos de la distribución normal y uniforme [0,1].

También se utilizaron las distribuciones t con 10 gl, (n=10000), beta con distintos parámetros (n=2500) y gama (n=2500) para comprobar que el método funciona bien con distintas distribuciones para θ.

Utilizamos 10 modelos de un parámetro con parámetros de dificultad entre 0.5 y 2 con distribución uniforme. En este caso, consideramos solamente dificultades positivas para no tener tanta variedad de valores. Cuando consideremos los modelos 2P y 3P tomaremos también dificultades negativas.

4.3.1.2. Modelo 2P

Para el modelo logístico de dos parámetros consideraremos 10 díadas de parámetros, con valores de a entre 0.5 y 2 y valores de b entre -2 y 2, ambos con distribución uniforme.

Para el rasgo, consideraremos 5 juegos de sujetos simulados, 5000 con distribución normal 0, 1; 5000 con distribución uniforme [0,1], 2500 con distribución beta de parámetros (0.27,0.44); (2.3,.57) y (5,5).

4.3.1.3. Modelo 3P

Para la curva logística de tres parámetros, tomamos 10 ternas, con valores de a entre 0.5 y 2, valores de b entre -2 y 2 y valores de c entre 0 y 0.5, todos con distribución uniforme.

Para el rasgo consideraremos 5 juegos de sujetos simulados, 5000 con distribución normal 0, 1; 5000 con distribución uniforme [0,1], 2500 con distribución beta de parámetros (0.27,0.44); (2.3,.57) y (5,5).

4.3.1.4. Modelo logit cúbico

Se consideraron 10 modelos con logit cúbico, con valores de a entre -1 y 1, y valores de b entre 1 y 2.5, con distribución uniforme y c 0.75.

Para el rasgo consideraremos 5 juegos de sujetos simulados, 5000 con distribución normal 0, 1; 5000 con distribución uniforme [0,1], 2500 con distribución beta de parámetros (0.27,0.44); (2.3,.57) y (5,5).

4.3.2. Resultados

En el apéndice A se pueden encontrar las tablas completas con todos los resultados de las simulaciones. Mostraremos un resumen de ellas.

4.3.2.1. Modelo 1P

Como era esperado, en el modelo de un parámetro, las distancias disminuyen a medida que crece el tama no muestral. La tabla 4.1 nos lo muestra. Sin embargo, debemos hacer ciertas consideraciones en lo que respecta a la distribución normal. En primer lugar, vemos como la distancia d_1 es mayor que la d_2. Esto es debido a que no se está ponderando. Luego, todo error «pesa» lo mismo, aunque el rasgo esté tomando valores poco probables. Esto también lleva a que se tengan ciertas alteraciones en el máximo error en esa distancia. Esto no se da en las distribuciones con soporte acotado.

Cuando se utiliza la distribución t con 10 gl, d_1 varía entre 0.0034 y 0.0058, teniendo una media de 0.0043 y una desviación típica (D.T.) de 0.0009. En tanto, d_2 varía entre 0.0011 y 0.0014, teniendo una media de 0.0013 y una D.T. de 0.0001.

Para la distribución gamma, d_1 tiene un rango de 0.0027-0.034, una media de 0.0030 y una D.T. de 0.0025 y, d_2 tiene un rango de 0.0008-0.0019, una media de 0.0014 y una D.T. de 0.0004. Ambos resultados son muy satisfactorios.

Tabla 4.1: Estadísticos descriptivos de las distancias en función del tamaño muestral.

Distribución normal				
N	$\overline{d_1}$	D.T. d_1	$\overline{d_2}$	D.T d_2
100	0.021	0.0023	0.0076	0.0005
250	0.0144	0.0007	0.0056	0.0004
500	0.0157	0.0044	0.0046	0.0001
1000	0.0152	0.0065	0.0034	0.00005
2500	0.0067	0.0020	0.0018	0.0001
5000	0.0044	0.0007	0.0014	0.00005
10000	0.0035	0.0011	0.0009	0.00003
Distribución uniforme				
N	$\overline{d_1}$	D.T. d_1	$\overline{d_2}$	D.T. d_2
100	0.0148	0.0032	0.0080	0.0018
250	0.0103	0.0022	0.0056	0.0013
500	0.0077	0.0018	0.0043	0.0011
1000	0.0067	0.0015	0.0037	0.0009
2500	0.0057	0.0011	0.0030	0.0006
5000	0.0045	0.0009	0.0024	0.0005
10000	0.0036	0.0007	0.0019	0.0004
Distribución normal				
N	min d_1	max d_1	min d_2	max d_2
100	0.018	0.0263	0.0067	0.0082
250	0.0134	0.0154	0.0048	0.0061
500	0.0111	0.0231	0.0045	0.0047
1000	0.0091	0.0268	0.0034	0.0036
2500	0.0047	0.010	0.0017	0.0019
5000	0.0035	0.0056	0.0013	0.0014
10000	0.0024	0.0055	0.0008	0.0009
Distribución uniforme				
N	min d_1	max d_1	min d_2	max d_2
100	0.0094	0.0177	0.0049	0.0098
250	0.0066	0.0124	0.0034	0.0069
500	0.0048	0.0095	0.0025	0.0055
1000	0.0042	0.0081	0.0022	0.0046
2500	0.0038	0.0066	0.0020	0.0036
5000	0.030	0.0053	0.0015	0.0029
10000	0.0024	0.0042	0.0012	0.0023

También para la distribución beta se obtiene, para distintos valores de sus parámetros, errores muy pequeños, como lo muestra la tabla 4.2.

Tabla 4.2: Estadísticos descriptivos de las distancias en función del tamaño muestral.

shape 1	shape 2	$\overline{d_1}$	D.T. d_1	$\overline{d_2}$	D.T. d_2
0.27	0.44	0.0052	0.0013	0.0030	0.0008
0.86	0.57	0.0042	0.0008	0.0023	0.0005
1	1	0.0054	0.0011	0.0029	0.0007
1.45	1.81	0.0099	0.0020	0.0053	0.0012
2.3	2.65	0.0118	0.0027	0.0065	0.0016
2.5	2.5	0.0121	0.0027	0.0067	0.0016
5	5	0.0164	0.0041	0.0093	0.0025
shape 1	shape 2	min d_1	max d_1	min d_2	max d_2
0.27	0.44	0.0032	0.0066	0.0017	0.0039
0.86	0.57	0.0029	0.0048	0.0015	0.0028
1	1	0.0034	0.0064	0.0018	0.0036
1.45	1.81	0.0065	0.0116	0.0033	0.0064
2.3	2.65	0.0074	0.0144	0.0039	0.0081
2.5	2.5	0.0076	0.0147	0.0040	0.0083
5	5	0.0097	0.0205	0.0123	0.0420

4.3.2.2. Modelo 2P

En el modelo logístico de dos parámetros obtuvimos con la distribución normal que d_1 varía entre 0.0029-0.0075 y tiene una media de 0.0048 y una D.T. de 0.0018. En tanto, d_2 varía entre 0.0009-0.0021 y tiene una media de 0.0014 y una D.T. de 0.0003.

Si utilizamos la distribución uniforme obtenemos, para d_1, el rango 0.0004-0.0068, una media de 0.0036 y una D.T. de 0.0024 y, para d_2, el rango 0.0008-0.0033, una media de 0.0020 y una D.T. de 0.0012.

Con la distribución beta (0.27, 0.44) tenemos, para d_1, un rango de 0.0007-0.00075, una media de 0.0046 y una D.T. de 0.0026 y, para d_2, un rango de 0.0004-0.0045, una media de 0.0027 y una D.T. de 0.0016.

Con beta (2.3, 0.57) obtuvimos, para d_1, un rango de 0.0008-0.0113, una media de 0.0077 y una D.T. de 0.0047 y, para d_2, un rango de 0.0005-0.0087, una media de 0.0046 y una D.T. de 0.0029.

Con beta (5,5) obtuvimos, para d_1, un rango de 0.0011-0.0232, una media de 0.0126 y una D.T. de 0.0076 y, para d_2, un rango de 0.0007-0.0135, una media de 0.0073 y una D.T. de 0.0044.

4.3.2.3. Modelo 3P

En el modelo de tres parámetros se obtienen los siguientes estadísticos para ambas distancias en las distribuciones anteriores.

En la distribución normal, para d_1, el rango es 0.0018-0.0058, la media 0.0039 y la D.T. 0.0013 y, para d_2, el rango es 0.0006-0.0020, la media 0.0012 y la D.T. 0.0005.

En la distribución uniforme, para d_1, el rango es 0.0017-0.0064, la media 0.0031 y la D.T. 0.0018 y, para d_2, el rango es 0.0009-0.0037, la media 0.0017 y la D.T. 0.0010.

En la distribución Beta (0.27,0.44), para d_1, el rango es 0.0001-0.0088, la media 0.0035 y la D.T. 0.0025 y, para d_2, el rango es 0.0001-0.0054, la media 0.0021 y la D.T. 0.0015.

En la distribución Beta (2.3,0.57), para d_1, el rango es 0.0002-0.0165, la media 0.0061 y la D.T. 0.0049 y, para d_2, el rango, 0.0001-0.0101, la media 0.0036 y la D.T. 0.0030.

En la distribución Beta (5,5), para d_1, el rango es 0.0002-0.0257, la media 0.0098 y la D.T. 0.0078 y, para d_2, el rango es 0.0001-0.0152, la media 0.0057 y la D.T. 0.0046.

4.3.2.4. Modelo logit cúbico

En el modelo logit cúbico se obtiene:

En la distribución normal, para d_1, el rango es 0.0032-0.0041, la media 0.0035 y la D.T. 0.0003 y, para d_2, el rango es 0.0016-0.0019, la media 0.0017 y la D.T. 0.0001.

En la distribución uniforme, para para d_1, el rango es 0.0034-0.0069, la media 0.0053 y la D.T. 0.0012 y, para d_2, el rango es 0.0019-0.0040, la media 0.0031 y la D.T. 0.0007.

En la distribución Beta (0.27,0.44), para d_1, el rango es 0.0043-0.010, la media 0.0075 y la D.T. 0.0018 y, para d_2, el rango es 0.0025-0.0062, la media 0.0046 y la D.T. 0.0012.

En la distribución Beta (2.3,0.57), para d_1, el rango es 0.0077-0.0171, la media 0.0132 y la D.T. 0.0035 y, para d_2 el rango es 0.0047-0.0106, la media 0.0081 y la D.T. 0.0021.

En la distribución Beta (5,5), para d_1, el rango es 0.0131-0.0257, la media 0.0206 y la D.T. 0.0053 y, para d_2, el rango es 0.0075-0.0168, la media 0.0122 y la D.T. 0.0031.

En todos los casos se ha verificado que el método de estimación funciona muy bien cuando se utiliza el modelo determinístico. Los errores disminuyen a medida que aumenta el tama no muestral y no son afectados cuando se utilizan diferentes distribuciones para el rasgo.

4.4. Estudio del modelo unidimensional no paramétrico para rasgo conocido

En este segundo estudio consideraremos el rasgo conocido, y a partir de él, generaremos respuestas a ítems según modelos predeterminados.

Con las respuestas a estos ítems estimamos la CCI y calcularemos las distancias ya definidas para estudiar el ajuste.

4.4.1. Método

Analizaremos dos efectos: el tamaño muestral y el tipo de distribución.

Para la investigación de la incidencia del tamaño muestral tomaremos tamaños muestrales de 100; 250; 500; 1000; 2500; 5000 y 10000 sujetos, con rasgo distribuido normal 0,1. Estimaremos 10 modelos distintos de tres parámetros con cada uno de los tamaños muestrales.

Los valores de los parámetros son aleatorios. Se utilizó para generar la distribución uniforme [0.5,2], para b uniforme [-2,2] y para c uniforme [0,0.5].

Para analizar cómo funciona el método cuando generamos respuestas donde el rasgo sigue distintas distribuciones, simulamos sujetos con rasgo distribuido en forma normal (0,1), uniforme [0,1], beta (0.27,0.44), beta (2.3,0.57) y beta (5,5). En este caso, se utilizaron modelos 3P con parámetros aleatorios y con iguales distribuciones que los anteriores y modelos logit cúbico, con distribución para los parámetros a, uniforme [-1,1], b, uniforme [1,2.5] y c=0.75.

4.4.2. Código de las funciones utilizadas

4.4.2.1. Simulación de respuestas

A los efectos de simular respuestas a ítems cuyas curvas características siguen un modelo determinado utilizamos la función genresp que puede verse en el apartado de funciones.

Esta función retorna una matriz de valores 0,1, y si tot=1, se agrega, además, una columna con los totales obtenidos por cada sujeto y otra columna donde se suma a cada total una variable aleatoria uniforme a los efectos de romper los empates.

Presentamos a continuación una función iccsmoothing que calcula las curvas características de los ítems y los errores estándar para θ conocido. En el apartado de funciones veremos la misma función modificada para el caso general.

```
iccsmoothing< −function(items,h,testsort,ability,points) {
# items un arreglo de índices de las curvas a ser estimadas
# h ancho de banda
# testsort un test dicotómico con I filas (una para cada sujeto) y
# J columnas (una para cada ítem)
# Las filas están preordenadas en orden ascendente según algún estadístico
# como, por ejemplo, la puntuación total
# ability un vector con las I habilidades
# points un vector de thetas donde las curva será evaluada
```

```
# La función retorna dos matrices NxJ conteniendo los estimados de ICC y los errores estándar
#primera fase: computa valores de las curvas en los puntos de evaluación
npoints< −length(points)
nexaminees< −length(ability)
nitems< −length(items)
#inicializa arrays
icc< −matrix(0,npoints,nitems)
weigth< −matrix(0,npoints,nexaminees)
icchat< −matrix(0,nexaminees,nitems)
#cálculo del operador de smoothing en cada punto de evaluación
for(i in 1:npoints){
    theta< −points[i]
    residual< −(ability-theta)/h
# conjunto de índices para pesos significativamente positivos
    windex< −abs(residual)¡=3000
# valores de smoothing para este conjunto de índices
    kernel< −exp(-(residual[windex]2̂)/2)
    weigth[i,windex]< −kernel/sum(kernel)
}
# cálculo de valores de ICC
icc< −weigth
# Segunda fase cálculo de estimados y error estándar
# interpolar valores de curvas de valores de habilidad
for(j in 1:nitems){
    icchat[,j]< −approx(points,icc[,j],ability)$y
    icchat[is.na(icchat),j]< −0
}
# computar los errores estándar con pesos estándar
errest< − sqrt(weigth 2̂ %*% (icchat*(1-icchat)))
# retorno
return(icc,errest)
}
```

Simularemos distintas curvas y veremos su ajuste, para ello utilizaremos la función simul4.

```
simul4< −function(a,b,c,theta,h,gra,nom=.ªuxi.txt")
#a vector de valores de a
#b vector de valores de b
#c vector de valores de c
#theta vector de habilidades
#h ventana a utilizar
# vector de 0 y 1 que indica si se ace el gráfico o no
simm< −genresp(a,b,c,theta)
ancho< −(max(theta)-min(theta))/1000
sali< −matrix(nrow=length(a),ncol=12)
leta< −seq(min(theta),max(theta),ancho)
D = 1.7
Prob< −matrix(0,length(leta),length(a))
ll< −seq(1,length(a),1)
rams< −iccsmoothing(ll,h,simm,theta,leta)
for(j in 1:length(a)) {
Prob[,j] < − c[j] + ((1 - c[j]) * exp( D *a[j] * (leta - b[j])))/(1 + exp(D * a[j] * (leta
- b[j])))
di< −distancias(rams$icc[,j],Prob[,j],leta,ancho)
sali[j,1]< −a[j]
sali[j,2]< −"& "
sali[j,3]< −b[j]
sali[j,4]< −"& "
sali[j,5]< −c[j]
sali[j,6]< −"& "
sali[j,7]< −di$d1
sali[j,8]< −"& "
sali[j,9]< −di$d2
sali[j,10]< −"& "
sali[j,11]< −di$d3
sali[j,12]< −"

"
if(gra[j]==1){
guiPlot(DataSetValues=data.frame(leta,Prob[,j]),
GraphSheet=guiGetGSName(),Page=j,Graph=j)
guiPlot(DataSetValues=data.frame(leta,rams$icc[,j]),
GraphSheet=guiGetGSName(),Graph=j,Page=j)
}
}
```

exportData(sali, nom, type = "FASCII", colNames = F,
format=" %6.4, %3, %6.4, %3, %6.4, %3, %6.4, %3, %6.4, %3, %6.4, %5")
return(rams,Prob,sali) }

4.4.3. Resultados

4.4.3.1. Tamaño muestral

En las tablas 4.4 a 4.13 claramente se observa cómo en todos los modelos simulados las distancias disminuyen al incrementarse el tamaño muestral. En la tabla 4.3 donde mostramos las medias y desviaciones típicas de estas diez simulaciones también se observa esto. Es interesante observar de que forma, a partir de N=250, la distancia d_2 tiene una media de 0.0553, que es una buena aproximación. Por lo tanto, no necesitamos un gran tamaño muestral, y para N=1000 la media es 0.0309, lo cual es excelente.
En las simulaciones 1 y 2 incluimos los gráficos para los distintos tamaños muestrales. Allí podemos visualizar cómo varía la proximidad de las curvas en función de N. Además, se observa de qué manera, para los tamaños mayores a 1000, donde el ajuste es excelente, el mayor error se da en los extremos de la distribución del rasgo (próximo a -3). Esto sucede, en general, cuando utilizamos regresión no paramétrica, ya que se trata de un promedio local.

Tabla 4.3: Media y desviación típica de d_2 en las 10 simulaciones.

N	$\overline{d_2}$	D.T.d_2
100	0.11956	0.0407
250	0.0538	0.0203
500	0.0553	0.0168
1000	0.0309	0.00837
2500	0.0239	0.0038
5000	0.0144	0.0041

Tabla 4.4: Simulación 1,$a = 1{,}3043$ $b = 1{,}7300$ $c = 0{,}2868$

N	d_1	d_2	d_3
100	0.4768	0.1716	0.3195
250	0.2517	0.0664	0.1530
500	0.2116	0.0749	0.1442
1000	0.1831	0.0369	0.2667
2500	0.1158	0.0302	0.1444
5000	0.1153	0.0195	0.0813

Figura 4.1

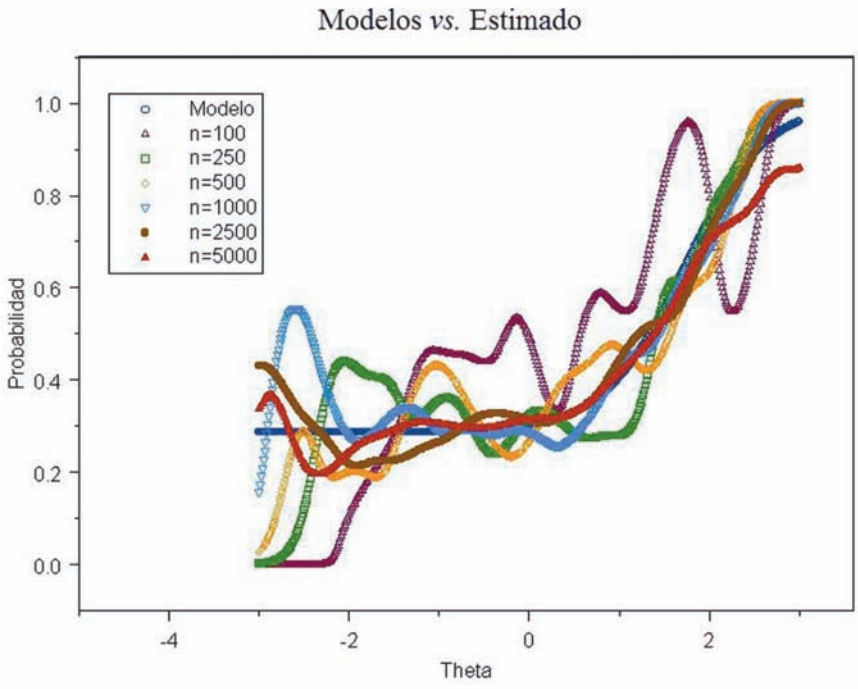


Tabla 4.5: Simulación 2 $a = 1{,}9494$ $b = -1{,}987$ $c = 0{,}1677$

N	d_1	d_2	d_3
100	0.6973	0.1057	0.8042
250	0.0650	0.0165	0.0572
500	0.5965	0.0656	0.8041
1000	0.1269	0.0174	0.0636
2500	0.1620	0.0178	0.3134
5000	0.0491	0.0081	0.0756

Figura 4.2

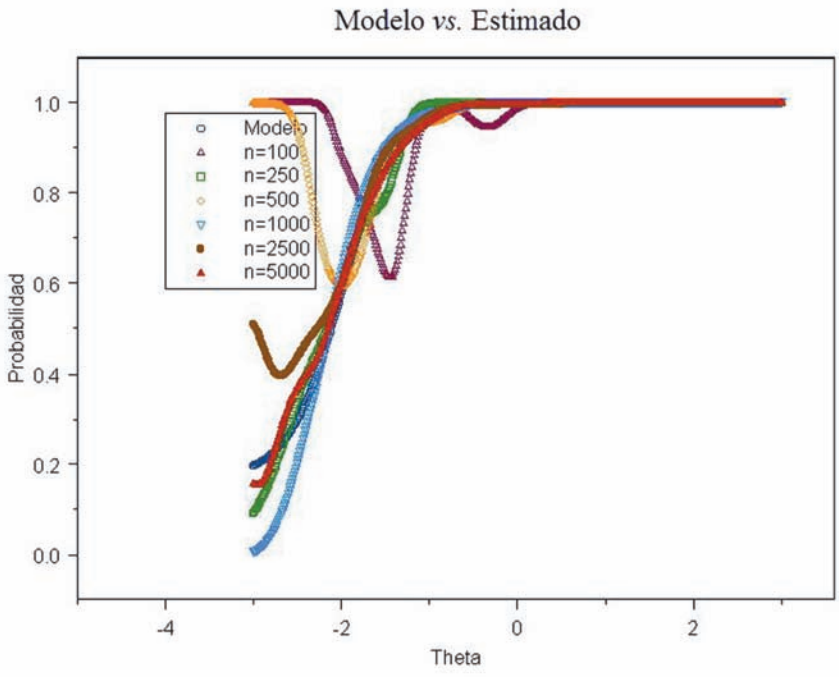


Tabla 4.6: Simulación 3, $a = 0{,}6431419$ $b = -1{,}923517$ $c = 0{,}1906125$

N	d_1	d_2	d_3
100	0.4516	0.0847	0.0649
250	0.3635	0.0637	0.5281
500	0.4973	0.0724	0.6186
1000	0.1274	0.0239	0.3477
2500	0.1197	0.0230	0.1380
5000	0.0592	0.0148	0.0658

Tabla 4.7: Simulación 4, $a = 1{,}2513$ $b = 1{,}0231$ $c = 0{,}0500$

N	d_1	d_2	d_3
100	0.1491	0.0679	0.1226
100	0.0799	0.0338	0.0499
250	0.1262	0.0481	0.0941
500	0.0995	0.0287	0.1126
2500	0.0590	0.0222	0.0367
5000	0.0533	0.0179	0.0229

Tabla 4.8: Simulación 5, $a = 1{,}6675$ $b = -0{,}351$ $c = 0{,}2165$

N	d_1	d_2	d_3
100	0.2639	0.0911	0.2588
250	0.1006	0.0424	0.0893
500	0.2765	0.0425	0.6427
1000	0.1127	0.0312	0.0469
2500	0.0571	0.0188	0.0572
5000	0.0434	0.0090	0.0640

Tabla 4.9: Simulación 6, $a = 0{,}5212$ $b = -0{,}642$ $c = 0{,}1411$

N	d_1	d_2	d_3
100	0.5430	0.1648	0.6221
250	0.4907	0.0878	0.6736
500	0.1776	0.0410	0.0732
1000	0.2535	0.0420	0.3413
2500	0.1004	0.0244	0.1299
5000	0.0428	0.0114	0.0542

Tabla 4.10: Simulación 7, $a = 1{,}8491$ $b = -0{,}807$ $c = 0{,}4264$

N	d_1	d_2	d_3
100	0.4870	0.1104	0.3958
250	0.2079	0.0500	0.1635
500	0.1739	0.0217	0.4342
1000	0.1545	0.0253	0.0827
2500	0.1086	0.0245	0.1641
5000	0.0569	0.0126	0.0889

Tabla 4.11: Simulación 8, $a = 1{,}7506$ $b = 1{,}6515$ $c = 0{,}4589$

N	d_1	d_2	d_3
100	0.5772	0.1734	0.0833
250	0.3665	0.0724	0.4486
500	0.2187	0.0603	0.1077
1000	0.1482	0.0350	0.3104
2500	0.0958	0.0289	0.1263
5000	0.0468	0.0141	0.0294

Tabla 4.12: Simulación 9, $a = 1{,}6405$ $b = -0{,}020$ $c = 0{,}1172$

N	d_1	d_2	d_3
100	0.2011	0.0798	0.1636
250	0.1559	0.0505	0.1053
500	0.1289	0.0582	0.0938
1000	0.0522	0.0256	0.0460
2500	0.0535	0.0252	0.0464
5000	0.0779	0.0168	0.1379

Tabla 4.13: Simulación 10, $a = 0{,}5144$ $b = 1{,}8393$ $c = 0{,}2498$

N	d_1	d_2	d_3
100	0.4572	0.1455	0.2181
250	0.1857	0.0540	0.1896
500	0.2098	0.0689	0.1875
1000	0.2305	0.0430	0.0950
2500	0.1868	0.0237	0.2167
5000	0.1017	0.0196	0.0673

4.4.3.2. Ajuste ante distintas distribuciones del rasgo

En este análisis también se constata un buen ajuste, para todas las distribuciones utilizadas para simular el rasgo. Este buen funcionamiento para diferentes distribuciones nos permite trabajar aunque no sepamos exactamente cómo se distribuye el rasgo. Notemos que para la distribución normal, como ya se dijo, la distancia adecuada para el análisis es d_2.

Veamos a continuación los resultados en las siguientes tablas donde claramente, la distancia d_2 nos muestra un buen ajuste, cualquiera sea la distribución.

Tabla 4.14: Simulación 1, $a = 0{,}5821$ $b = 0{,}5133$ $c = 0{,}1566$

	d_1	d_2	d_3
normal	0.7102	0.0226	0.8378
uniforme	0.0147	0.0083	0.0040
beta(0.27,0.44)	0.0125	0.0075	0.0121
beta(2.3,0.57)	0.0373	0.0227	0.0707
beta(5,5)	0.0164	0.0100	0.0468

Tabla 4.15: Simulación 2, $a = 0{,}5301$ $b = 1{,}8373$ $c = 0{,}1184$

	d_1	d_2	d_3
normal	0.2000	0.0202	0.1476
uniforme	0.0189	0.0113	0.0243
beta(0.27,0.44)	0.0173	0.0100	-0.007
beta(2.3,0.57)	0.0236	0.0133	0.0334
beta(5,5)	0.0098	0.0054	0.0062

Tabla 4.16: Simulación 3,$a = 0{,}8091$ $b = -1{,}065$ $c = 0{,}3418$

	d_1	d_2	d_3
normal	0.4049	0.0293	0.4666
uniforme	0.0105	0.0060	0.0191
beta(0.27,0.44)	0.0024	0.0014	0.0058
beta(2.3,0.57)	0.0154	0.0095	0.0389
beta(5,5)	0.0159	0.0099	0.0406

Tabla 4.17: Simulación 4,$a = 0{,}9086$ $b = 0{,}8393$ $c = 0{,}3772$

	d_1	d_2	d_3
normal	0.4238	0.0222	0.4523
uniforme	0.0081	0.0047	0.0153
beta(0.27,0.44)	0.0221	0.0135	0.0095
beta(2.3,0.57)	0.0271	0.0154	0.0413
beta(5,5)	0.0078	0.0043	0.0161

Tabla 4.18: Simulación 5,$a = 1{,}9820$ $b = 0{,}5811$ $c = 0{,}3091$

	d_1	d_2	d_3
normal	0.3972	0.0306	0.0308
uniforme	0.0393	0.0231	0.0704
beta(0.27,0.44)	0.0165	0.0098	0.0263
beta(2.3,0.57)	0.0781	0.0445	0.1142
beta(5,5)	0.0500	0.0293	0.0857

También cuando el modelo utilizado es el logit cúbico, el ajuste es bueno para las distintas distribuciones del rasgo.
Es más, la media de d_2 para la distribución normal en el modelo 3P fue de 0.0249 en las 5 simulaciones, en tanto, en el modelo logit cúbico fue de 0.0226.
Cuando consideramos la distribución uniforme, la media de d_2 fue de 0.0106 en el modelo 3P y de 0.0153 en el logit cúbico. Es claro, por lo tanto, que el error es aproximadamente igual en ambos modelos utilizados.

Tabla 4.19: Simulación 1 ,$a = 0{,}1305$ $b = 1{,}0722$ $c = 0{,}75$

	d_1	d_2	d_3
normal	0.0382	0.0192	0.0458
uniforme	0.0227	0.0125	0.0387
beta(0.27,0.44)	0.0189	0.0108	0.0326
beta(2.3,0.57)	0.0527	0.0321	0.0946
beta(5,5)	0.0506	0.0282	0.0620

Tabla 4.20: Simulación 2 ,$a = 0{,}5928$ $b = 2{,}3697$ $c = 0{,}75$

	d_1	d_2	d_3
normal	0.0357	0.0184	0.0412
uniforme	0.0232	0.0137	0.0650
beta(0.27,0.44)	0.0301	0.0184	0.0376
beta(2.3,0.57)	0.0542	0.0338	0.1505
beta(5,5)	0.0460	0.0278	0.1150

Tabla 4.21: Simulación 3 ,$a = -0{,}530$ $b = 1{,}4939$ $c = 0{,}75$

	d_1	d_2	d_3
normal	0.0478	0.0224	0.0221
uniforme	0.0301	0.0164	0.0380
beta(0.27,0.44)	0.0278	0.0154	0.0282
beta(2.3,0.57)	0.0702	0.0428	0.1259
beta(5,5)	0.0743	0.0426	0.1118

Tabla 4.22: Simulación 4 ,$a = -0{,}581$ $b = 1{,}7452$ $c = 0{,}75$

	d_1	d_2	d_3
normal	0.0363	0.0195	0.0257
uniforme	0.0362	0.0209	0.0809
beta(0.27,0.44)	0.0298	0.0180	0.0175
beta(2.3,0.57)	0.0925	0.0565	0.1806
beta(5,5)	0.0731	0.0424	0.1279

Tabla 4.23: Simulación 5 ,$a = 0{,}4732$ $b = 2{,}2021$ $c = 0{,}75$

	d_1	d_2	d_3
normal	0.0586	0.0336	0.0412
uniforme	0.0225	0.0130	0.0587
beta(0.27,0.44)	0.0346	0.0212	0.0217
beta(2.3,0.57)	0.0446	0.0276	0.1045
beta(5,5)	0.0554	0.0338	0.1329

4.5. Modelo no paramétrico unidimensional con CCI y rasgo desconocido

4.5.1. Tamaño muestral

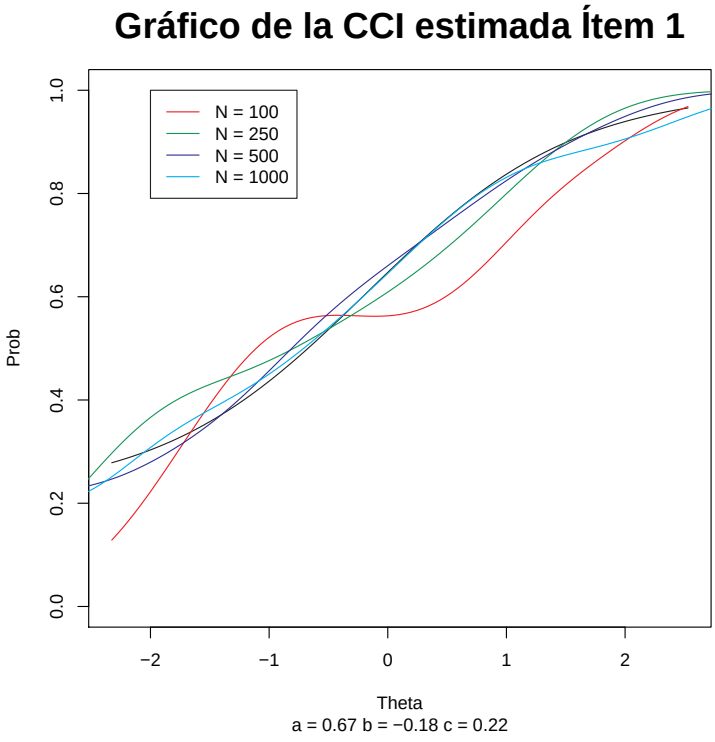
Para evaluar el efecto del tamaño muestral calculamos el promedio de las distancias mencionadas previamente en un test de 30 ítems aplicado a 100, 250, 500 y 1000 individuos cuyas habilidades provienen de una distribución normal (0,1).

En la regresión no paramétrica utilizamos un ancho de ventana (h) igual a 0.5 y un núcleo normal.

En las siguientes secciones evaluaremos los efectos de considerar distintos valores para h y distintos núcleos.

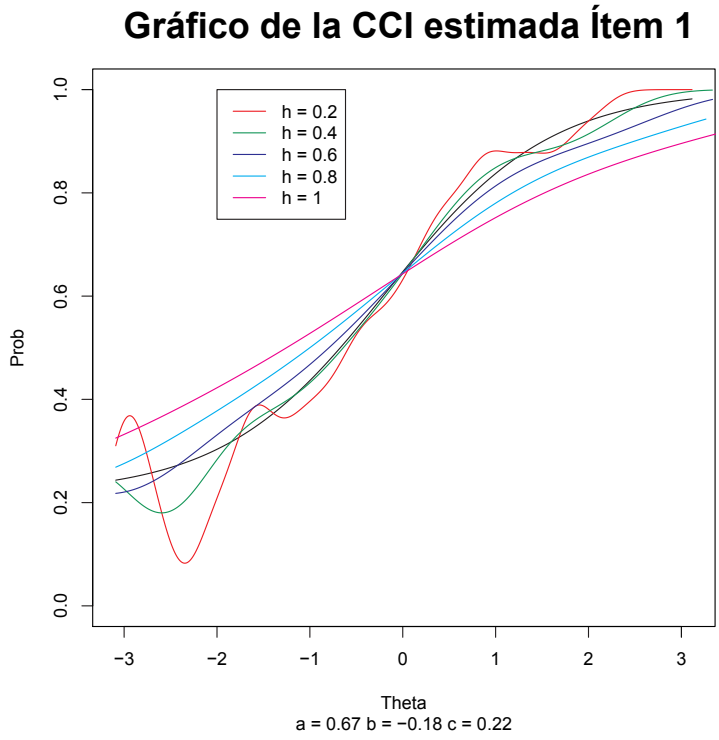
Los resultados fueron:

	media de d1	media de d2	media de d3
N = 100	0.1736	0.0684	0.1197
N = 250	0.1353	0.0493	0.0753
N = 500	0.1688	0.0552	0.0941
N = 1000	0.1357	0.0480	0.0916



4.5.2. Efecto de distintos valores de ancho para la ventana

Utilizando el mismo test de la sección 4.5.1 anterior aplicado a 1000 individuos, estudiamos el efecto de considerar distintos anchos de ventana en la regresión no paramétrica. En el método de estimación utilizamos un núcleo normal.



	media de d1	media de d2	media de d3
h = 0.2	0.1302	0.0339	0.1005
h = 0.4	0.1119	0.0368	0.0784
h = 0.6	0.1773	0.0622	0.1115
h = 0.8	0.2854	0.0916	0.1611
h = 1	0.3935	0.1174	0.2092

4.5.3. Distintas distribuciones para el parámetro habilidad

Para estudiar los efectos de distintas distribuciones para el parámetro habilidad consideramos 5 juegos de datos, aplicados a 500 individuos. En cada uno de estos juegos de datos la habilidad de los individuos proviene de una distribución de probabilidad distinta. Las distribuciones consideradas son:

- Normal (0,1)
- Uniforme (0,1)
- Beta (4,2)
- Beta (2,4)

- Beta (5,5)

Para realizar la comparación entre las distintas distribuciones consideradas, fijamos los valores de a, b y c, con valores de a entre 0.5 y 2, de b entre -3 y 3 y c entre 0 y 0.25.

En los casos de las variables aleatorias uniformes y beta realizamos una transformación debido a que el parámetro habilidad tiene como soporte todos los reales y en las variables aleatorias consideradas es el intervalo $[0, 1]$. La transformación realizada es la siguiente:

Sea x una realización de una variable aleatoria con soporte $[0, 1]$. Si x es menor que 0.5: $y = \frac{2x-1}{2x}$ Si x es mayor igual que 0.5: $y = \frac{2x-1}{-2x+2}$

En todos estos casos utilizamos en la regresión no paramétrica un ancho de ventana igual a 0.5 y un núcleo normal.

Caso a=0.5, b=-3 y c=0.

	d1	d2	d3
Normal(0,1)	0.2521	0.0385	0.0264
Uniforme(0,1)	0.2693	0.0465	0.0336
Beta(4,2)	0.1725	0.0663	0.1249
Beta(2,4)	0.6611	0.1786	0.0019
Beta(5,5)	0.0737	0.0239	0.0729

Caso a=0.875, b=-1.5 y c=0.0625.

	d1	d2	d3
Normal(0,1)	0.0851	0.0289	0.0482
Uniforme(0,1)	0.1711	0.0379	0.0142
Beta(4,2)	0.4436	0.1313	0.3518
Beta(2,4)	0.5032	0.2151	0.0111
Beta(5,5)	0.2312	0.0871	0.2113

Caso a=1.25, b=0 y c=0.125.

	d1	d2	d3
Normal(0,1)	0.1632	0.0656	0.1053
Uniforme(0,1)	0.1386	0.0520	0.0864
Beta(4,2)	0.4256	0.2373	0.3708
Beta(2,4)	0.4720	0.2342	0.0341
Beta(5,5)	0.1121	0.0555	0.0793

Caso a=1.625, b=1.5 y c=0.1875.

	d1	d2	d3
Normal(0,1)	0.2633	0.0916	0.1461
Uniforme(0,1)	0.2313	0.0785	0.1106
Beta(4,2)	0.5329	0.2542	0.4462
Beta(2,4)	0.2831	0.0964	0.0201
Beta(5,5)	0.2960	0.0846	0.0006

Caso a=2, b=3 y c=0.25.

	d1	d2	d3
Normal(0,1)	0.6364	0.1584	0.5160
Uniforme(0,1)	0.7540	0.1910	0.6064
Beta(4,2)	0.9668	0.2889	0.7076
Beta(2,4)	0.5349	0.1335	0.4041
Beta(5,5)	0.6117	0.1478	0.4989

4.5.4. Efecto de distintos núcleos en la regresión no paramétrica

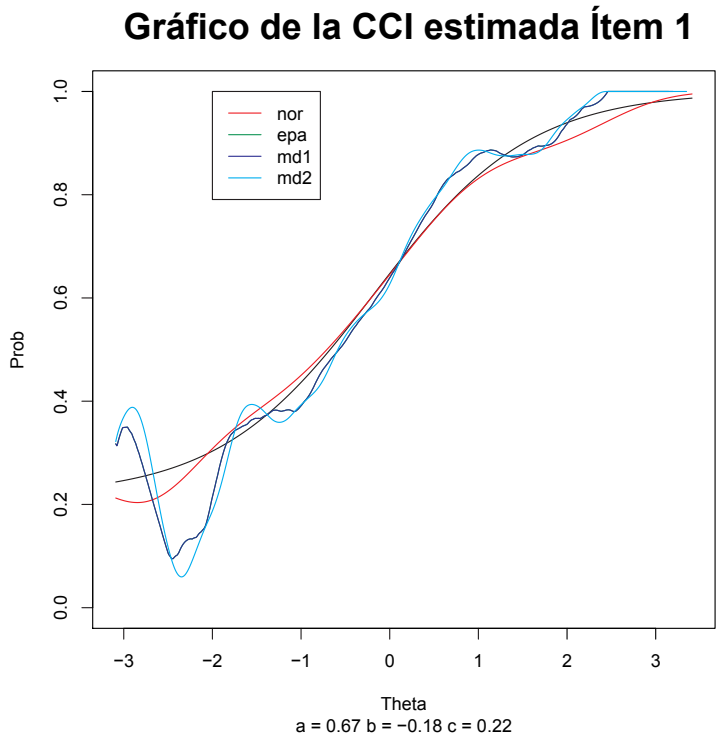
Para ver como funciona el método de estimación al considerar distintos núcleos, utilizamos el test de la sección 4.5.1 aplicado a 1000 individuos con distribución de habilidades normal (0,1) y fijamos un valor de ancho de ventana igual a 0.5. Consideramos los siguientes cuatro núcleos:

1. $\frac{1}{\sqrt{2\pi}}\, e^{\frac{-(x)^2}{2}}$ ('nor')
2. $(1-x^2) \quad |x| \leq 1$ ('epa')

3. $3\pi^{-1}(1-x^2) \quad |x| \leq 1$ ('md1')

4. $4\pi^{-1}(1-x^2)^3 \quad |x| \leq 1$ ('md2')

	media de d1	media de d2	media de d3
nor	0.1356	0.0480	0.0916
epa	0.1244	0.0322	0.0903
md1	0.1246	0.0322	0.0904
md2	0.1391	0.0359	0.1122



4.5.5. Comparación entre distintos modelos y métodos de estimación

Compararemos cuatro modelos: el logístico de tres parámetros, el logístico de dos parámetros, el logístico de un parámetro y el modelo logit cúbico.

Modelo logístico de dos parámetros:

$$P(x_i = 1|\theta) = \frac{e^{Da(\theta-b)}}{1+e^{Da(\theta-b)}}$$

Modelo logístico de un parámetro:

$$P(x_i = 1|\theta) = \frac{e^{D(\theta-b)}}{1 + e^{D(\theta-b)}}$$

En el logístico de tres parámetros consideramos al parámetro $c \sim U(0,0{,}25)$, $b \sim U(-2,2)$ y $a \sim U(0{,}5,2)$. En el caso de logístico de dos parámetros c es igual a cero y en el caso del modelo logístico de un parámetro además de c igual a cero, a es igual a uno.

Modelo logit cúbico:

$$P(x_i = 1|\theta) = \frac{e^{a+b\theta+c\theta^3}}{1 + e^{a+b\theta+c\theta^3}}$$

Para este modelo el parámetro $a \sim U(-1,1)$, $b \sim U(1,2{,}5)$ y c es fijo igual a 0,75.

Para evaluar los distintos modelos consideramos el test de la sección 4.5.1 aplicado a 1000 individuos con distribución de habilidad normal (0,1). En la regresión no paramétrica utilizamos un ancho de ventana igual a 0.5 y un núcleo normal.

Para realizar la comparación entre los distintos métodos de estimación utilizamos el programa Multilog (MLG). Este estima las CCI mediante TRI clásico.

	media de d1	media de d2	media de d3
3 parámetros reg. np	0.1357	0.0480	0.0916
3 parámetros MLG	0.1353	0.0480	0.0933
2 parámetros reg. np	0.1409	0.0525	0.0882
2 parámetros MLG	0.1482	0.0554	0.0912
1 parámetro reg. np	0.1140	0.0427	0.0715
1 parámetros MLG	0.1195	0.0451	0.0744
logit cúbico reg. np	0.1410	0.0667	0.1011
logit cúbico MLG	1.1476	0.3590	0.3700

Para comparar las estimaciones de la habilidad calculamos:

$$Dif = \frac{\sum_{i=1}^{N} \left| \theta_i - \hat{\theta}_i \right|}{N}$$

	Dif Reg. np	Dif Multilog
logístico 1p	0.2355	0.2382
logístico 2p	0.2321	0.2347
logístico 3p	0.2766	0.3801
logit cubico	0.2186	0.4533

4.5.6. Otros resultados

4.5.6.1. Modelo logístico

En la tabla 4.24 vemos cómo disminuye el ECM y la diferencia absoluta a medida que la muestra crece, y que cuando N=500, el error cometido es muy pequeño.

Cuando analizamos el ajuste en los 30 ítems con una muestra de N=500 el máximo error cometido en d_2 es de 0.0856.

Tabla 4.24: Estudio del efecto del tamaño muestral

N	ECM	Dif Abs
100	0.01722711	0.1039181
250	0.01215409	0.08091041
500	0.003302364	0.03618572
1000	0.001594676	0.03143627
2500	0.001311621	0.02055305
5000	0.000552657	0.0154313

Tabla 4.25: Ajuste de los diferentes ítems N=100

ítem	d_1	d_2	d_3
1	0.2972	0.1026	0.1631
2	0.1870	0.0756	0.2172
3	0.1768	0.0505	0.0923
4	0.3434	0.1538	0.2511
5	0.2684	0.0711	0.6053
6	0.1961	0.1066	0.1877
7	0.1531	0.0808	0.1882
8	0.2242	0.0507	0.1664
9	0.3128	0.0937	0.1417
10	0.2574	0.1185	0.3637
11	0.2100	0.0839	0.4033
12	0.2683	0.0964	0.3512
13	0.3439	0.0991	0.6996
14	0.2103	0.0780	0.2630
15	0.2942	0.0866	0.4130
16	0.2290	0.1206	0.2294
17	0.3292	0.1544	0.3427
18	0.3973	0.1377	0.5476
19	0.3205	0.1381	0.3550
20	0.2466	0.1095	0.1653
21	0.1248	0.0402	0.1866
22	0.3880	0.1186	0.5168
23	0.1849	0.0762	0.1363
24	0.2431	0.1030	0.1864
25	0.3117	0.1279	0.2317
26	0.2858	0.0827	0.6013
27	0.1879	0.0626	0.0751
28	0.3398	0.0829	0.5051
29	0.3619	0.1393	0.2406
30	0.1014	0.0333	0.2511

Tabla 4.26: Ajuste de los diferentes ítems N=500

ítem	d_1	d_2	d_3
1	0.1512	0.0668	0.1133
2	0.2517	0.0445	0.3039
3	0.2103	0.0372	0.4830
4	0.3508	0.0747	0.1519
5	0.1679	0.0235	0.3831
6	0.1483	0.0449	0.1026
7	0.1156	0.0436	0.0770
8	0.2006	0.0459	0.2693
9	0.1788	0.0364	0.1969
10	0.1109	0.0428	0.0608
11	0.0983	0.0205	0.0550
12	0.1980	0.0338	0.0848
13	0.1435	0.0430	0.1897
14	0.3022	0.0582	0.0911
15	0.1675	0.0285	0.1110
16	0.1033	0.0367	0.0418
17	0.1826	0.0789	0.3277
18	0.1731	0.0428	0.1192
19	0.1277	0.0562	0.0789
20	0.1712	0.0550	0.0802
21	0.1769	0.0295	0.0003
22	0.3965	0.0835	0.0913
23	0.1786	0.0533	0.2031
24	0.1847	0.0691	0.1601
25	0.1417	0.0357	0.0454
26	0.1240	0.0478	0.0940
27	0.1787	0.0430	0.1864
28	0.3249	0.0431	0.5754
29	0.3177	0.0756	0.3979
30	0.2598	0.0856	0.1208

Tabla 4.27: Ajuste de los diferentes ítems N=2500

ítem	d_1	d_2	d_3
1	0.1682	0.0373	0.2008
2	0.3424	0.0385	0.0583
3	0.1620	0.0227	0.2131
4	0.3825	0.0449	0.1176
5	0.1037	0.0162	0.0718
6	0.1398	0.0205	0.0469
7	0.1081	0.0345	0.0701
8	0.1045	0.0196	0.1580
9	0.0516	0.0145	0.0197
10	0.1515	0.0392	0.0511
11	0.0817	0.0114	0.0222
12	0.2174	0.0440	0.0686
13	0.1660	0.0220	0.0422
14	0.1985	0.0285	0.0431
15	0.0847	0.0170	0.2427
16	0.0992	0.0308	0.0643
17	0.1202	0.0378	0.0717
18	0.3114	0.0371	0.0613
19	0.1176	0.0298	0.0498
20	0.0964	0.0203	0.0862
21	0.1545	0.0228	0.1686
22	0.2501	0.0298	0.0414
23	0.2504	0.0329	0.0142
24	0.1253	0.0234	0.1117
25	0.1048	0.0219	0.1634
26	0.0846	0.0149	0.0289
27	0.2000	0.0181	0.0314
28	0.1716	0.0160	0.1565
29	0.2407	0.0449	0.1394
30	0.2340	0.0528	0.0992

Tabla 4.28

Distribución	ECM	Dif Abs
normal	0.001311621	0.02055305
uniforme	0.00003541157	0.004724083
beta(0.27,0.44)	0.00002913581	0.003250605
beta(2.3,0.57)	0.00004478771	0.004950457
beta(5,5)	0.00004357959	0.005663073

4.5.6.2. Logit cúbico

Se obtuvo para el rasgo un error cuadrático medio de 0.0033 y un error absoluto medio de 0.0362 y el ajuste a las curva fue bueno. Esto muestra la potencia del método. Funciona bien para modelos diferentes a los logísticos, lo que hace que pueda ser ampliamente aplicado. Vease tabla 4.29

4.6. Simulaciones en dos dimensiones, ítems dependientes de una sola habilidad

Para introducirnos en la simulación en dos dimensiones consideramos tests de n ítems, en los que los primeros 30 ítems dependen de una habilidad y los siguientes 30 de otra. Las habilidades las simulamos de distribuciones bivariadas con marginales conocidas y distintos niveles de correlación (cópulas).

4.6.1. Efecto del tamaño muestral

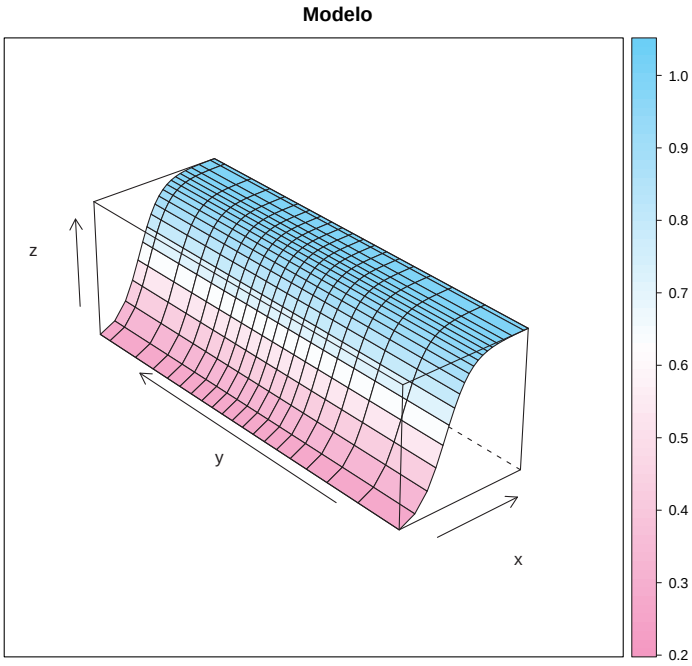
Para evaluar el efecto de distintos tamaños de muestra calculamos el promedio en todos los ítems de las distancias en un test con las caracterśticas previamente descritas aplicado a 100, 500 y 1000 individuos, cuyas habilidades provienen de una cópula con marginales normales(0,1) y correlación de 0,75.

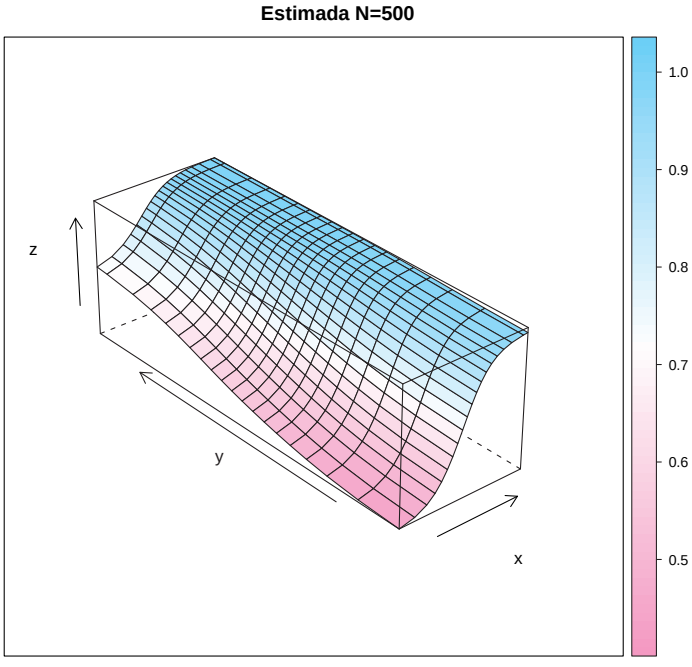
En este caso utilizamos en la regresión no paramétrica un núcleo normal y un ancho de ventana igual a 0,2, más adelante veremos el efecto de considerar otras alternativas.

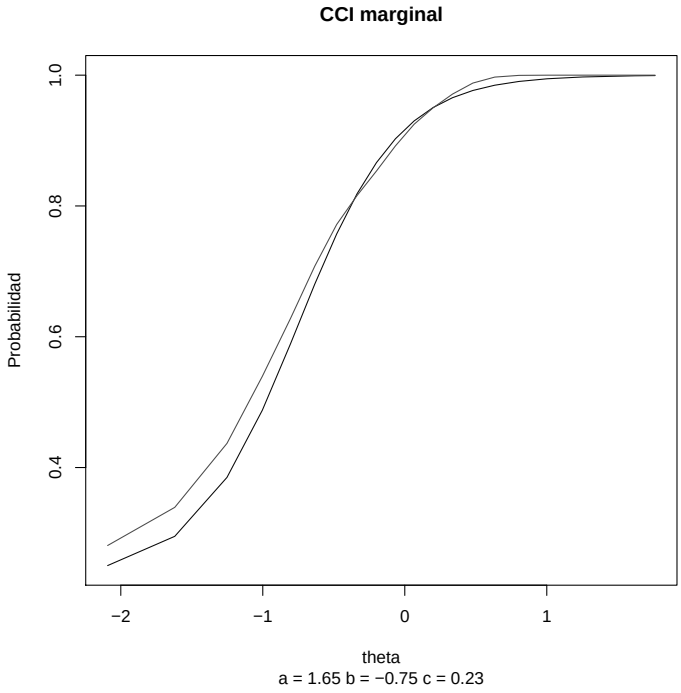
Tabla 4.29

ítem	d_1	d_2	d_3
1	0.1047	0.0580	0.0428
2	0.1065	0.0613	0.0941
3	0.0683	0.0360	0.0827
4	0.0338	0.0192	0.0301
5	0.0846	0.0499	0.1121
6	0.0251	0.0149	0.0319
7	0.0422	0.0236	0.0490
8	0.0286	0.0161	0.0211
9	0.0479	0.0277	0.0586
10	0.0519	0.0294	0.0647
11	0.0755	0.0397	0.0839
12	0.0643	0.0379	0.0483
13	0.0518	0.0278	0.0639
14	0.0554	0.0321	0.0606
15	0.0599	0.0310	0.0504
16	0.0275	0.0146	0.0241
17	0.0397	0.0213	0.0238
18	0.1133	0.0704	0.1592
19	0.0423	0.0252	0.0599
20	0.0630	0.0366	0.0663
21	0.0270	0.0156	0.0348
22	0.0851	0.0488	0.0688
23	0.0485	0.0265	0.0551
24	0.0268	0.0138	0.0302
25	0.0243	0.0124	0.0069
26	0.0609	0.0352	0.0725
27	0.0431	0.0242	0.0533
28	0.0405	0.0223	0.0462
29	0.0468	0.0272	0.0581
30	0.0458	0.0256	0.0486

	d1	d2	d3
N = 100	0.1112	0.0232	0.2565
N = 500	0.0224	0.0042	0.2224
N = 1000	0.0104	0.0021	0.1760


Modelo
z
y
x
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
a = 1.65 b = −0.75 c = 0.23

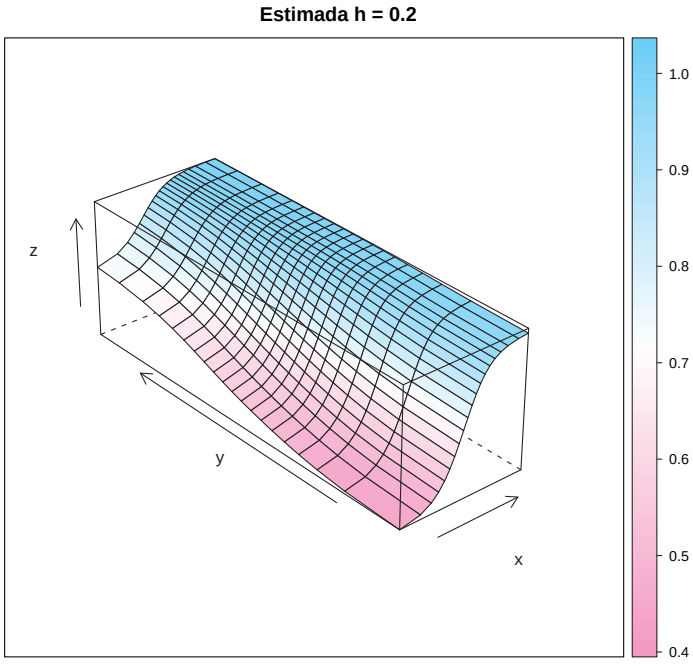

Estimada N=500
z
y
x
1.0
0.9
0.8
0.7
0.6
0.5
a = 1.65 b = −0.75 c = 0.23



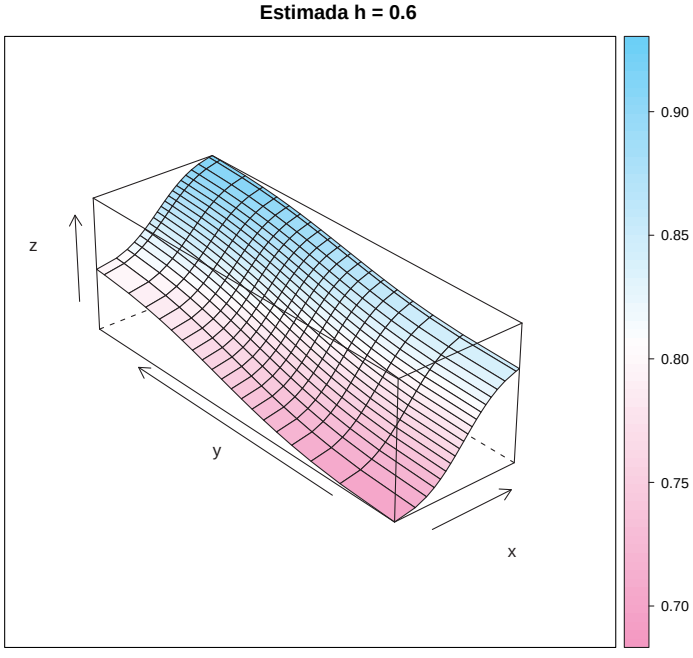
4.6.2. Efecto de distintos valores de ancho de ventana

Para comparar las estimaciones mediante regresión no paramétrica al variar los anchos de ventana consideramos el test de la sección 4.6.1 anterior aplicado a 500 individuos. Luego realizamos las estimaciones utilizando distintos anchos de ventana y calculamos las tres distancias, en todos los casos utilizamos un núcleo normal en la regresión no paramétrica.

	d1	d2	d3
h = 0.2	0.0244	0.0044	0.2490
h = 0.4	0.0303	0.0059	0.2415
h = 0.6	0.0351	0.0069	0.2832
h = 0.8	0.0353	0.0073	0.2278
h = 1	0.0428	0.0083	0.2627



a = 1.65 b = −0.75 c = 0.23 h = 0.2



a = 1.65 b = −0.75 c = 0.23 h = 0.6

4.6.3. Efecto de distintas cópulas para el parámetro habilidad

Evaluamos el efecto de considerar distintas distribuciones bivariadas (cópulas) para las habilidades de los individuos. Para ello utilizamos distintas distribuciones marginales y fijamos el nivel de correlación entre ambas igual a $0,75$. Luego evaluamos a estos individuos con el test de la sección 4.6.1 y calculamos el promedio en todos los ítems de cada una de las tres habilidades. Las cópulas consideradas tienen las siguientes marginales:

- N(0,1) y N(0,1)
- Beta(5,5) y Beta(5,5)
- U(0,1) y U(0,1)

En este caso le aplicamos a las marginales de las cópulas la transformación mencionada en la sección anterior:

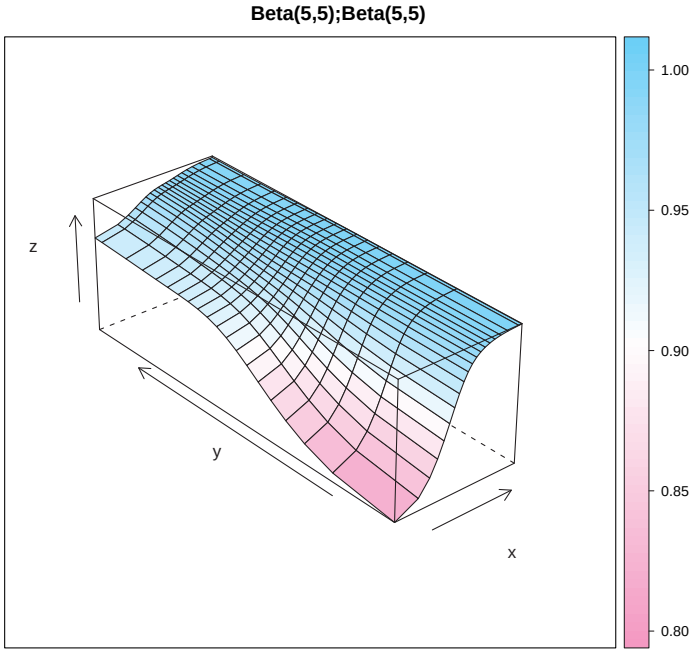
Sea x una realización de una variable aleatoria con soporte $[0, 1]$. Si x es menor que 0,5:

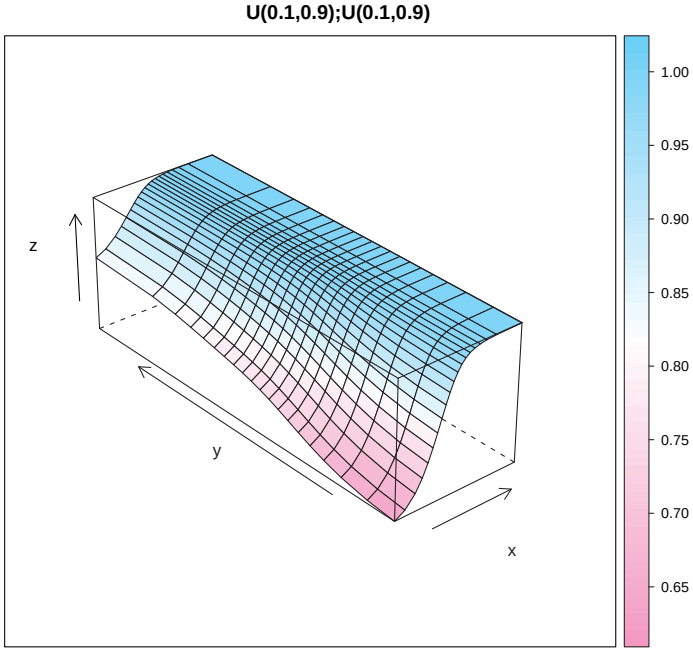
$$y = \frac{2x - 1}{2x}$$

Si x es mayor igual que 0,5:

$$y = \frac{2x - 1}{-2x + 2}$$

	d1	d2	d3
N(0,1);N(0,1)	0.0244	0.0040	0.1874
Beta(5,5);Beta(5,5)	0.0264	0.0047	0.2526
U(0.1,0.9);U(0.1,0.9)	0.0155	0.0028	0.2373


Beta(5,5);Beta(5,5)
z
y
x
1.00
0.95
0.90
0.85
0.80
a = 1.19 b = −1.93 c = 0.19

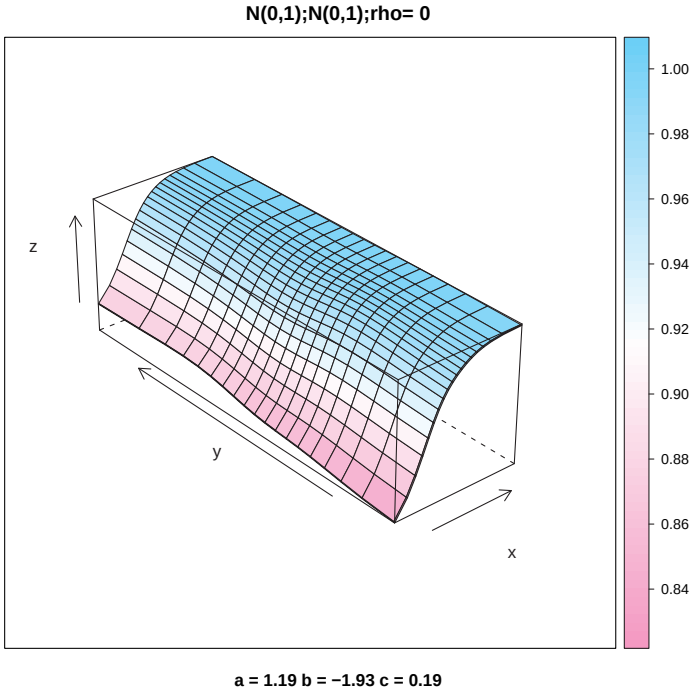

U(0.1,0.9);U(0.1,0.9)
z
y
x
1.00
0.95
0.90
0.85
0.80
0.75
0.70
0.65
a = 1.19 b = −1.93 c = 0.19

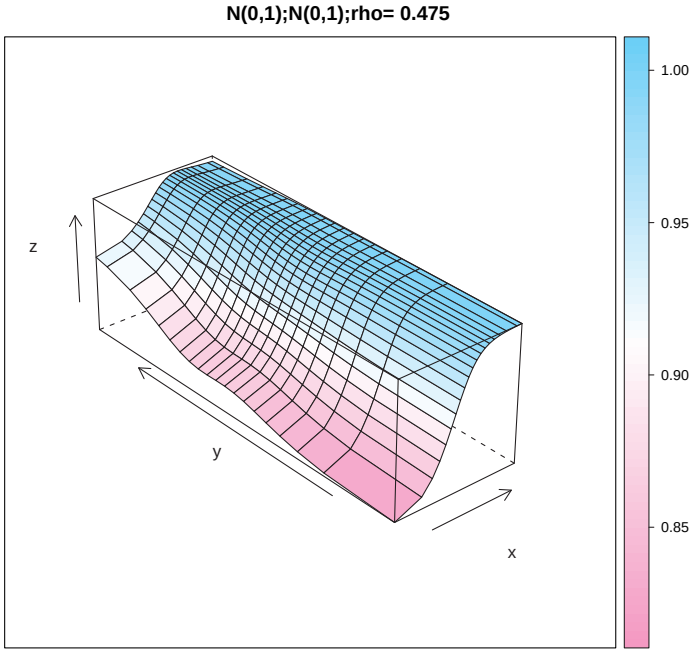
4.6.4. Distintos niveles de correlación entre las habilidades

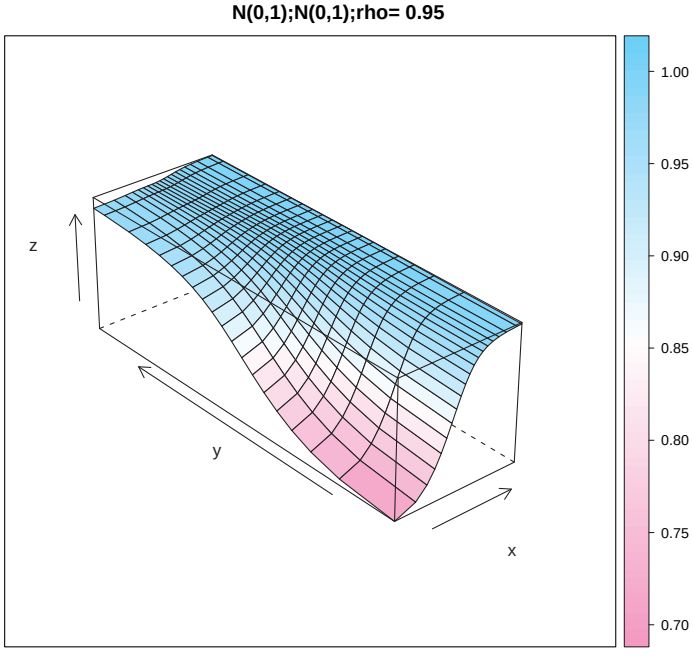
En esta sección consideramos distintos niveles de correlación entre las distribuciones marginales (N(0,1)) que componen la cópula, es decir, consideraremos distintos niveles de correlación entre las dos habilidades de cada individuo. En cada caso sometemos a los individuos al test de la sección 4.6.1.

El caso en el cual el coeficiente de correlación es igual a cero implica que se está suponiendo independencia entre las dos habilidades en cuestión.

	d1	d2	d3
N(0,1);N(0,1);rho= 0	0.0092	0.0017	0.1150
N(0,1);N(0,1);rho= 0.2375	0.0091	0.0017	0.1291
N(0,1);N(0,1);rho= 0.475	0.0114	0.0018	0.2259
N(0,1);N(0,1);rho= 0.7125	0.0132	0.0021	0.2262
N(0,1);N(0,1);rho= 0.95	0.0142	0.0026	0.2751




N(0,1);N(0,1);rho= 0.475
z
y
x
1.00
0.95
0.90
0.85
a = 1.19 b = −1.93 c = 0.19


N(0,1);N(0,1);rho= 0.95
z
y
x
1.00
0.95
0.90
0.85
0.80
0.75
0.70
a = 1.19 b = −1.93 c = 0.19

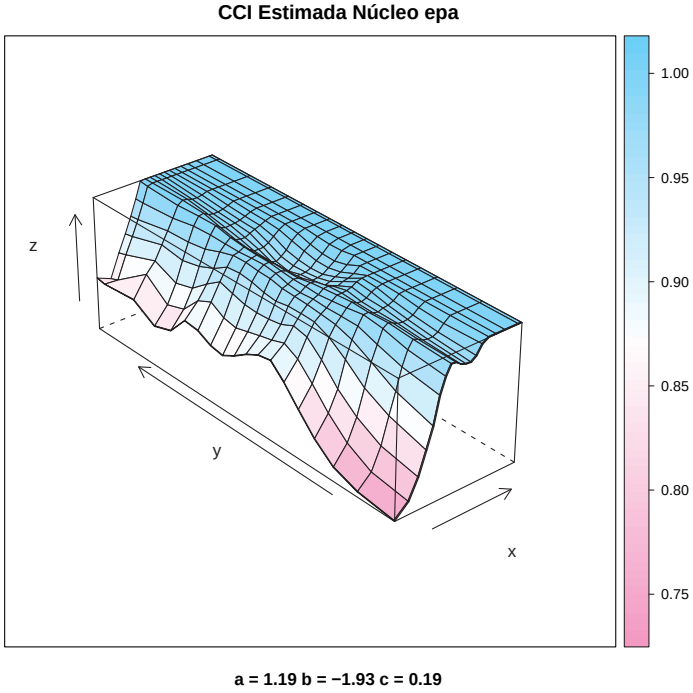
4.6.5. Efecto de distintos núcleos en la regresión no paramétrica

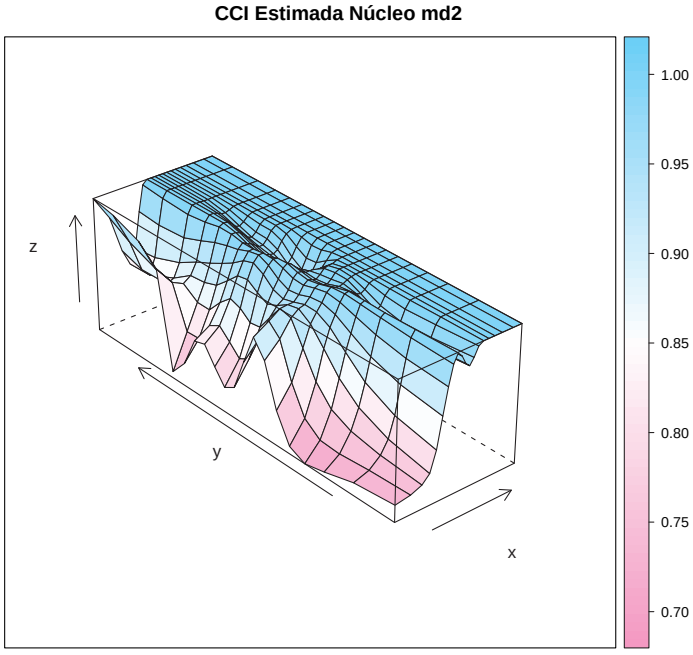
Con el fin de evaluar los efectos de distintos núcleos en la regresión no paramétrica consideramos el mismo test de la sección 4.6.1 aplicado a 500 individuos. Consideramos los siguientes cuatro núcleos:

1. $\frac{1}{\sqrt{2\pi}}\, e^{\frac{-(x)^2}{2}}$ ('nor')
2. $(1-x^2) \quad |x^t x| \leq 1$ ('epa')
3. $3\pi^{-1}(1-x^2) \quad |x^t x| \leq 1$ ('md1')
4. $4\pi^{-1}(1-x^2)^3 \quad |x^t x| \leq 1$ ('md2')

En todos los casos fijamos el ancho de ventana igual a 0,2.

	d1	d2	d3
nor	0.0104	0.0021	0.1760
epa	0.0092	0.0019	0.2184
md1	0.0095	0.0018	0.2572
md2	0.0109	0.0021	0.3218





4.6.6. Comparación entre valores estimados y reales del parámetro habilidad

Para comparar la estimación del parámetro habilidad con el verdadero valor de éste, en una dimensión calculamos:

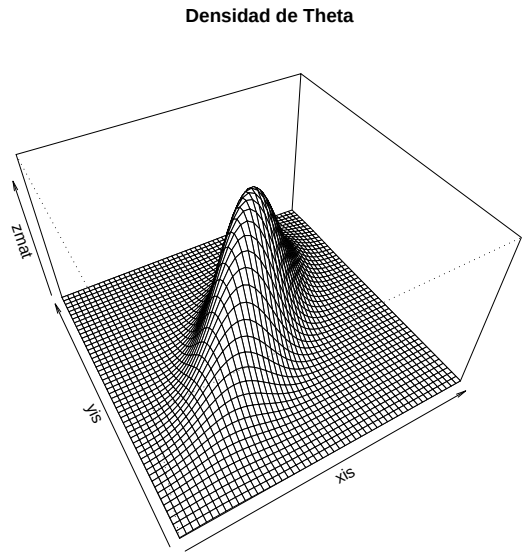
$$Dif = \frac{\sum_{i=1}^{N} \left| \theta_i - \hat{\theta}_i \right|}{N}$$

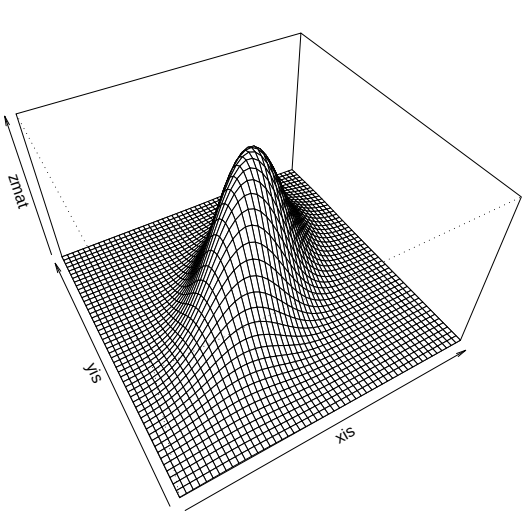
De modo similar en dos dimensiones calculamos:

$$Diff = \frac{\sum_{i=1}^{N} |\hat{\theta_1} - \theta_1| + |\hat{\theta_2} - \theta_2|}{N}$$

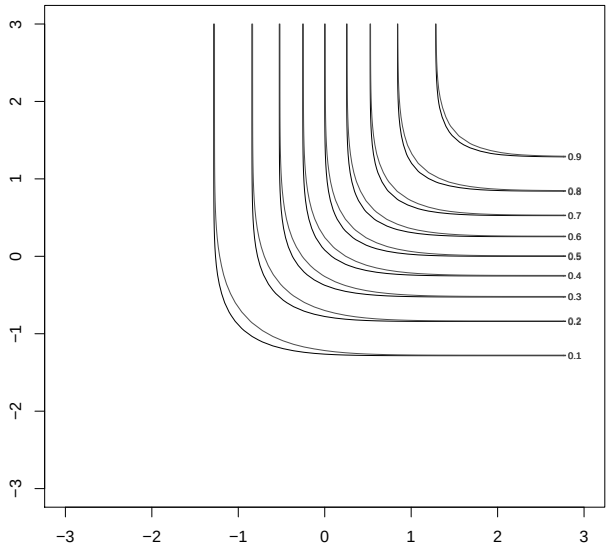
	Diff
N=100	0.71
N=500	0.68
N=1000	0.68

Las siguientes dos gráficas nos muestran la función de densidad del paramétro habilidad y la función de densidad estimada de este parámetro.





En el siguiente gráfico tenemos las curvas de nivel de probabilidad acumulada de la verdadera distribución del parámetro habilidad, en negro, y de la estimación de ésta, en rojo.



4.7. Simulaciones en dos dimensiones con habilidades dependientes

A continución consideraremos ítems dependientes de dos habilidades, como pueden ser idioma español y matemática. Parece lógico suponer que existe correlación entre los niveles de conocimiento de distintas habilidades, por lo que, para simular estas habilidades, utilizaremos distribuciones bivariadas con marginales conocidas y distintos grados de correlación. El modelo que utilizaremos será el logístico de tres parámetros, que en el caso de θ bivariada es:

$P(u_{ij} = 1|a_{i1}, a_{i2}, b_i, c_i, \theta) = c_i + (1 - c_i)\frac{e^{Da_{i1}\theta_{j1}+Da_{i2}\theta_{j2}-b_i}}{1+e^{Da_{i1}\theta_{j1}+Da_{i2}\theta_{j2}-b_i}} \quad i = 1, ..., n \; j = 1, ..., N$

- a_{i1} es el nivel de discriminación en la habilidad 1 del ítem i.
- a_{i2} es el nivel de discriminación en la habilidad 2 del ítem i.
- b_i es el nivel de dificultad del ítem i.
- c_i es el parámetro de azar del ítem i.
- θ_{j1} es la habilidad 1 del individuo j.
- θ_{j2} es la habilidad 2 del individuo j.

Para estimar θ_1 y θ_2 utilizamos respectivamente:

$$T_{\theta_1} = \sum_{i=1}^{n} \frac{a_{i1}}{a_{i2}} X_i \qquad T_{\theta_2} = \sum_{i=1}^{n} \frac{a_{i2}}{a_{i1}} X_i$$

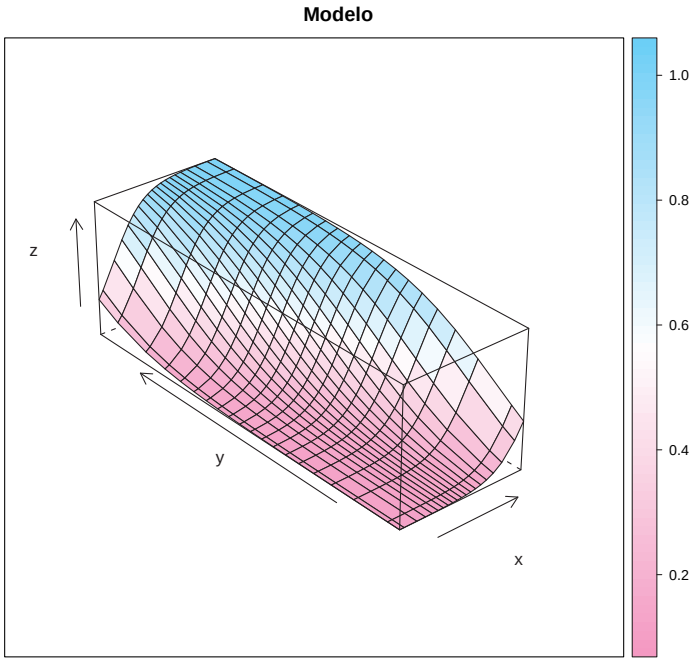
En una primera instancia analizaremos los efectos de distintas distribuciones marginales, manteniendo fijo el nivel de correlación entre las habilidades. Luego analizaremos el efecto de distintos grados de correlación entre las hailidades.

En todos los casos trabajaremos con tests aplicados a 500 individuos y trabajamos con núcleos normales y anchos de ventana igual a 0,2 en las regresiones no paramétricas.

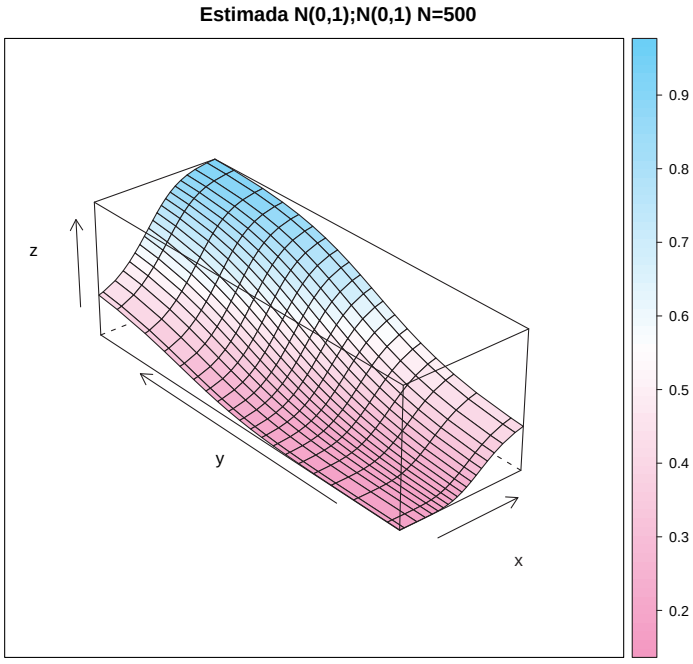
4.7.1. Efecto del tamaño muestral

Evaluaremos los efectos de distintos tamaños de muestra, para ello aplicamos un test de 60 ítems a 100, 500 y 1000 individuos cuyas habilidades fueron simuladas de una distribución bivariada con marginales N(0,1). En la regresión no paramétrica utilizamos un ancho de ventana igual a 0,2 y un núcleo normal.

	media de d1	media de d2	media de d3
N(0,1);N(0,1) N=100	0.1219	0.0238	0.3000
N(0,1);N(0,1) N=500	0.0287	0.0053	0.3062
N(0,1);N(0,1) N=1000	0.0160	0.0029	0.3456



a1 = 1.28 a2 = 1.21 b = 1.06 c = 0.13

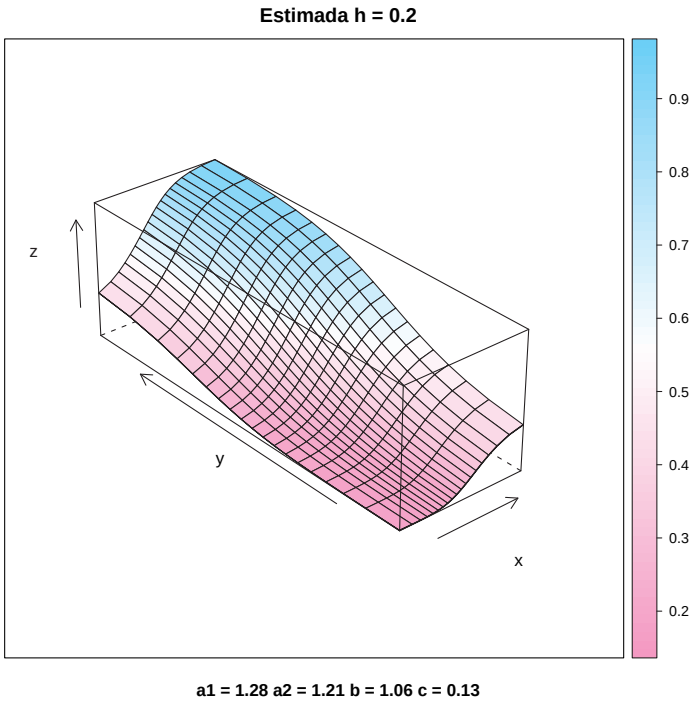


a1 = 1.28 a2 = 1.21 b = 1.06 c = 0.13

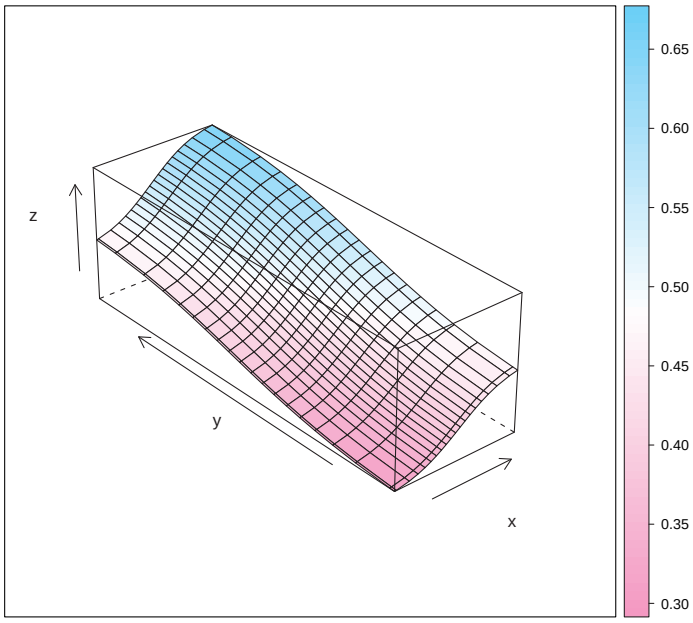
4.7.2. Efecto de distintos anchos de ventana

Utilizando el mismo test que en la sección anterior aplicado a 500 individuos, calcularemos las tres distancias considerando distintos anchos de ventana. Continuamos utilizando en la regresión no paramétrica un núcleo normal.

	media de d1	media de d2	media de d3
h = 0.2	0.0288	0.0054	0.2918
h = 0.4	0.0444	0.0100	0.3876
h = 0.6	0.0552	0.0135	0.3870
h = 0.8	0.0636	0.0150	0.4507
h = 1	0.0663	0.0160	0.4454

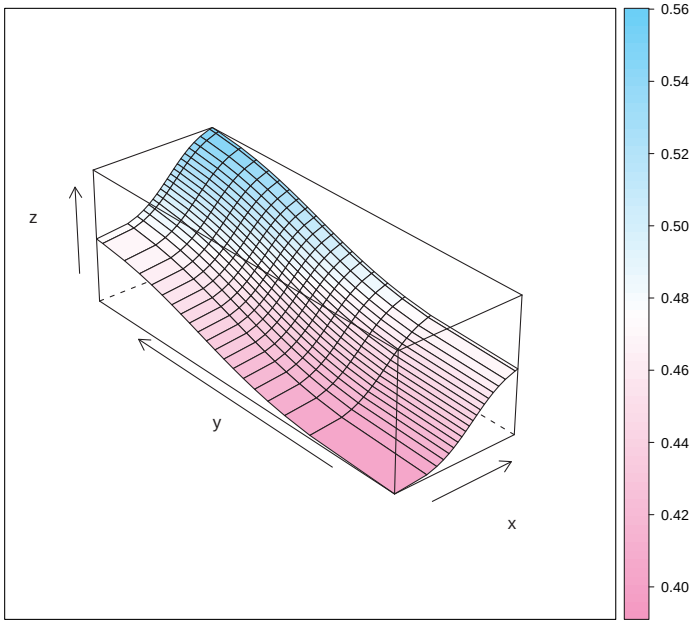


Estimada h = 0.6



a1 = 1.28 a2 = 1.21 b = 1.06 c = 0.13

Estimada h = 1



a1 = 1.28 a2 = 1.21 b = 1.06 c = 0.13

4.7.3. Efecto de distintas distribuciones de θ

Para simular las matrices de habilidades de los individuos se generaron tres distribuciones bivariadas cuyas marginales son conocidas con una correlación igual a 0,75. Estas distribuciones tienen las siguientes marginales:

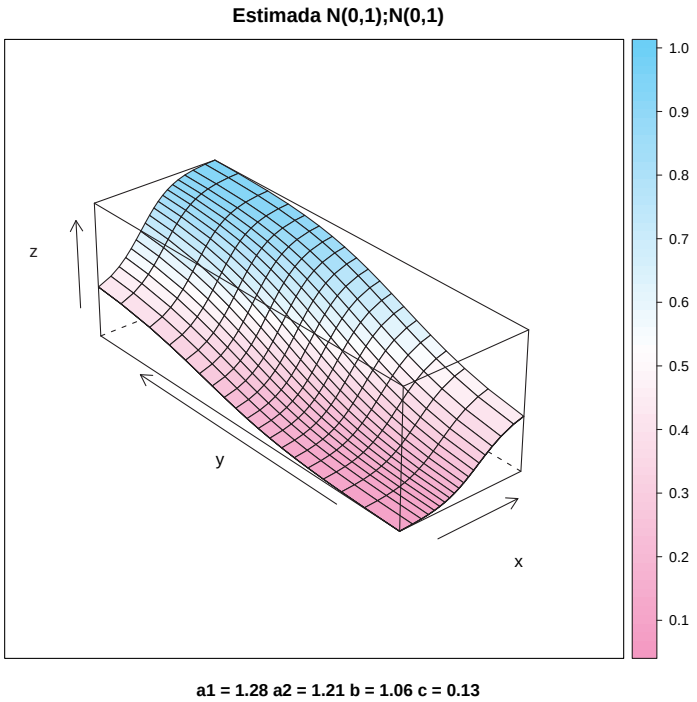
1. N(0,1) y N(0,1)

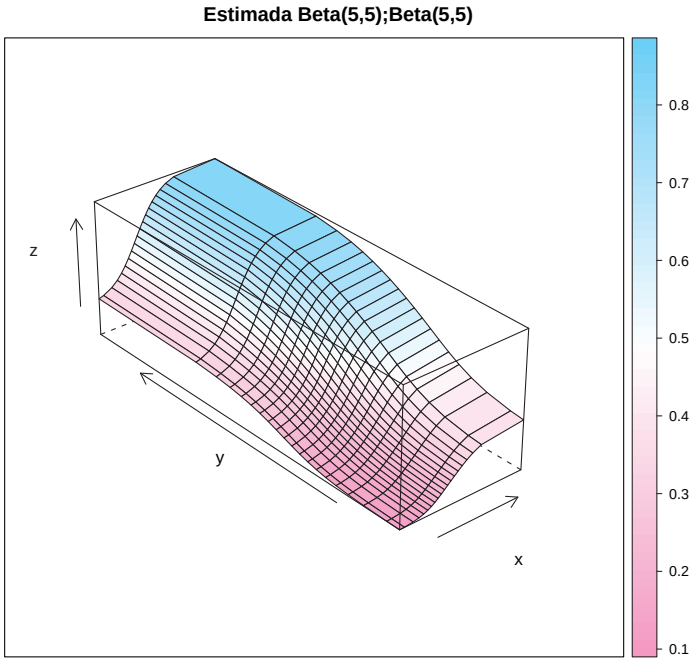
2. B(5,5) y B(5,5)

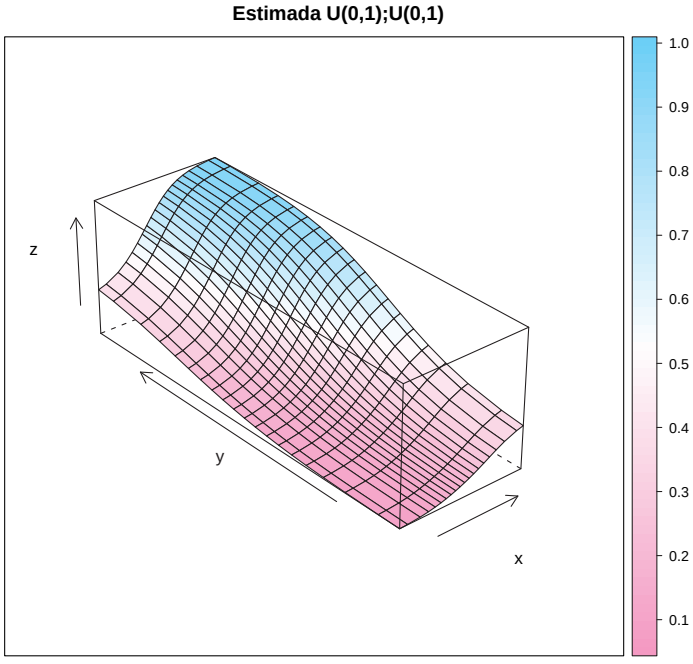
3. U(0,1) y U(0,1)

Para evaluar el efecto de considerar distintas distribuciones para el parámetro habilidad, calculamos las tres distancias y las promediamos.

	media de d1	media de d2	media de d3
N(0,1);N(0,1)	0.0154	0.0029	0.2457
Beta(5,5);Beta(5,5)	0.0308	0.0060	0.3128
U(0,1);U(0,1)	0.0179	0.0025	0.2858



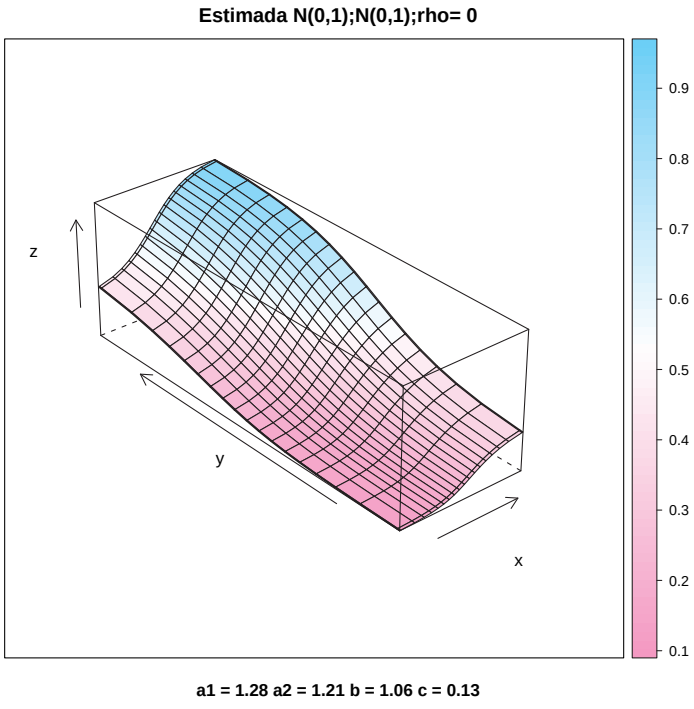

Estimada Beta(5,5);Beta(5,5)
z
y
x
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
a1 = 1.28 a2 = 1.21 b = 1.06 c = 0.13

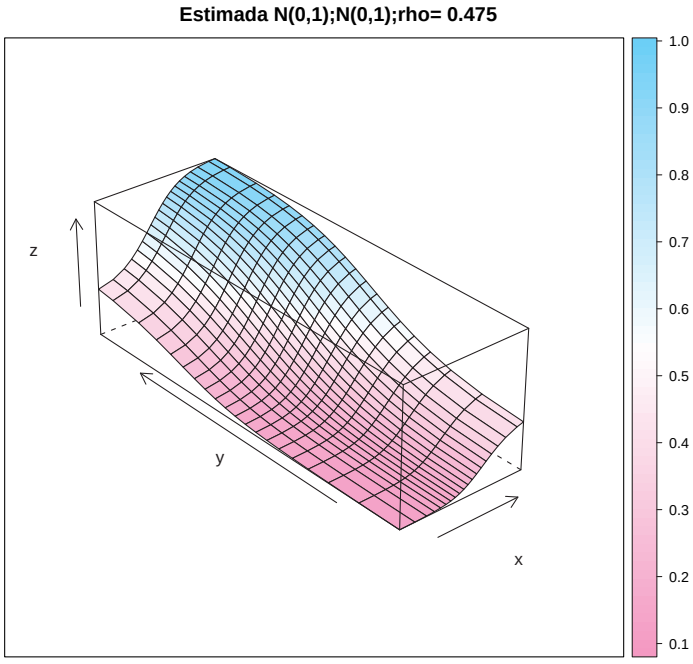

Estimada U(0,1);U(0,1)
z
y
x
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
a1 = 1.28 a2 = 1.21 b = 1.06 c = 0.13

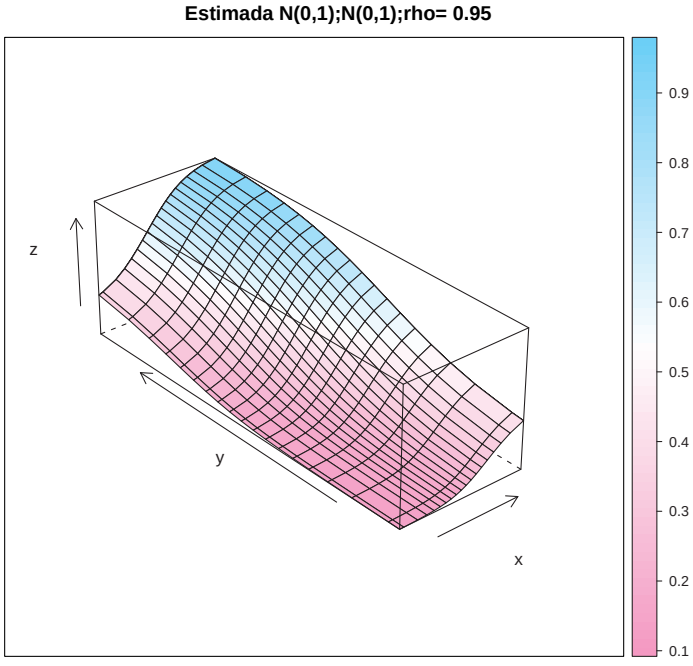
4.7.4. Efecto de distintos grados de correlación entre las habilidades

Para evaluar el efecto de distintos grados de correlación entre las distintas habilidades consideramos una distribución bivariada con marginales con distribución N(0,1) y calculamos las tres distancias para distintos niveles de correlación.

	media de d1	media de d2	media de d3
N(0,1);N(0,1);rho= 0	0.0163	0.0038	0.2330
N(0,1);N(0,1);rho= 0.2375	0.0153	0.0033	0.2423
N(0,1);N(0,1);rho= 0.475	0.0156	0.0032	0.2859
N(0,1);N(0,1);rho= 0.7125	0.0159	0.0029	0.3288
N(0,1);N(0,1);rho= 0.95	0.0138	0.0022	0.2603




Estimada N(0,1);N(0,1);rho= 0.475
z
y
x
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
a1 = 1.28 a2 = 1.21 b = 1.06 c = 0.13


Estimada N(0,1);N(0,1);rho= 0.95
z
y
x
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
a1 = 1.28 a2 = 1.21 b = 1.06 c = 0.13

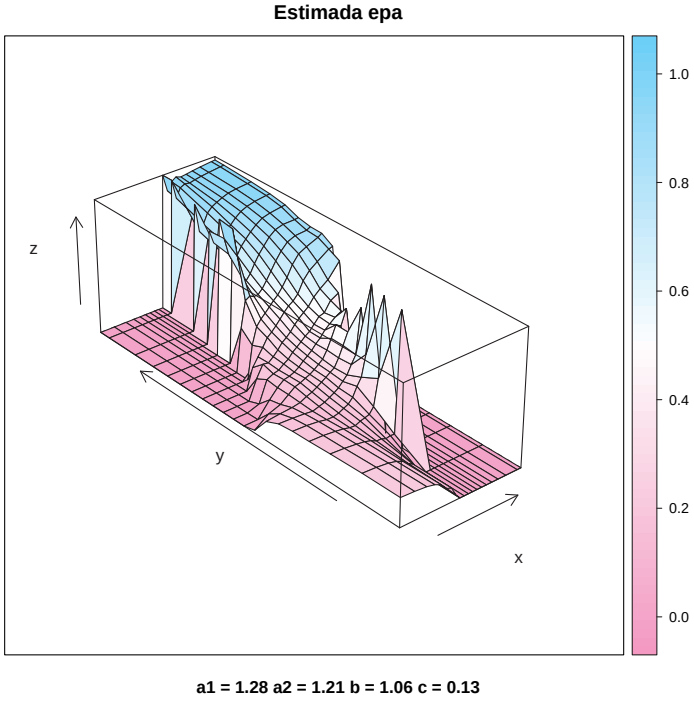
4.7.5. Efecto de distintos núcleos

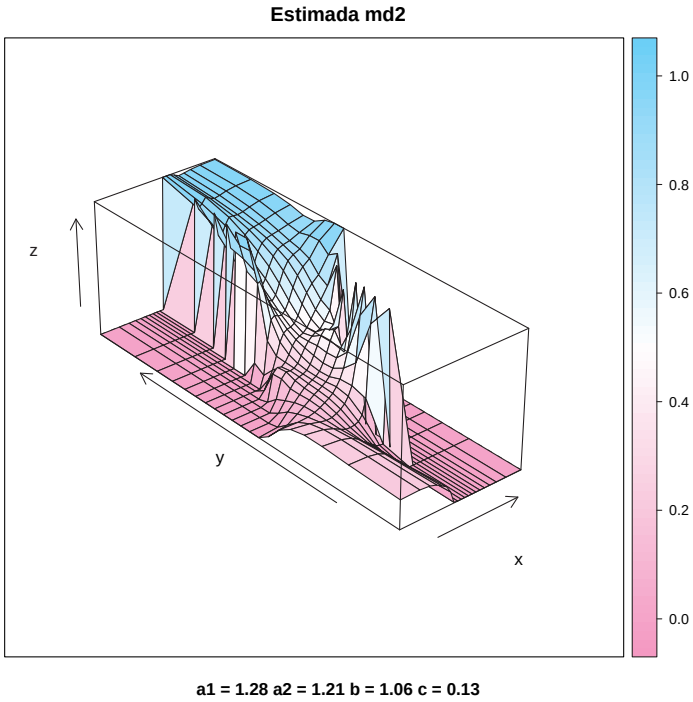
Con el fin de evaluar los efectos de distintos núcleos en la regresión no paramétrica consideramos el mismo test de la sección 4.7.1 aplicado a 500 individuos. Consideramos los siguientes cuatro núcleos:

1. $\frac{1}{\sqrt{2\pi}}\, e^{\frac{-(x)^2}{2}}$ ('nor')
2. $(1-x^2) \quad |x^t x| \leq 1$ ('epa')
3. $3\pi^{-1}(1-x^2) \quad |x^t x| \leq 1$ ('md1')
4. $4\pi^{-1}(1-x^2)^3 \quad |x^t x| \leq 1$ ('md2')

En todos los casos fijamos el ancho de ventana igual a 0,2.

	media de d1	media de d2	media de d3
nor	0.0287	0.0053	0.3062
epa	0.0895	0.0103	0.6855
md1	0.0895	0.0101	0.6867
md2	0.0919	0.0110	0.6985





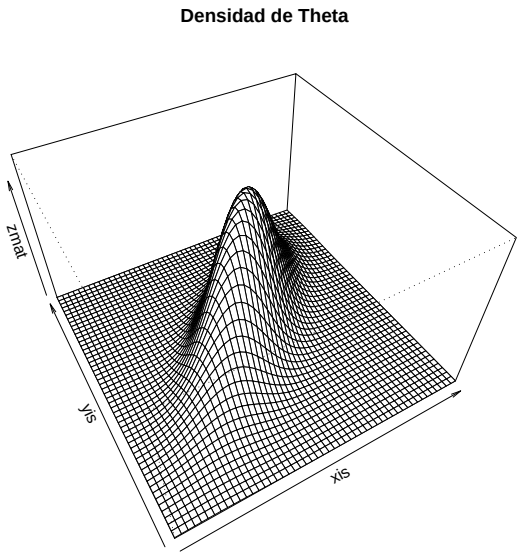
4.7.6. Comparación entre valores estimados y valores reales del parámetro habilidad

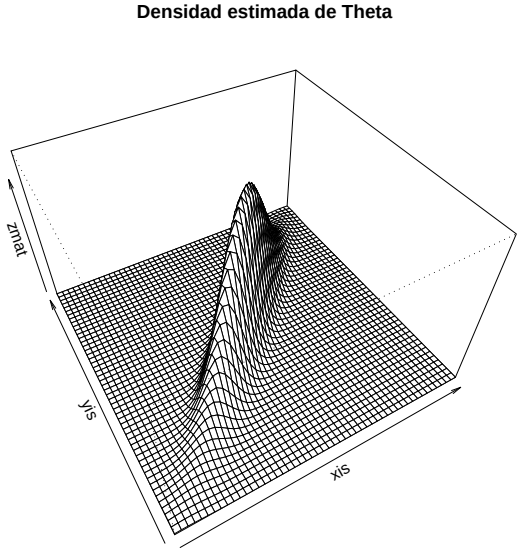
Como en la sección 4.6.6 comparamos el parámetro habilidad y la estimación de este mediante:

$$Diff = \frac{\sum_{i=1}^{N} |\hat{\theta_1} - \theta_1| + |\hat{\theta_2} - \theta_2|}{N}$$

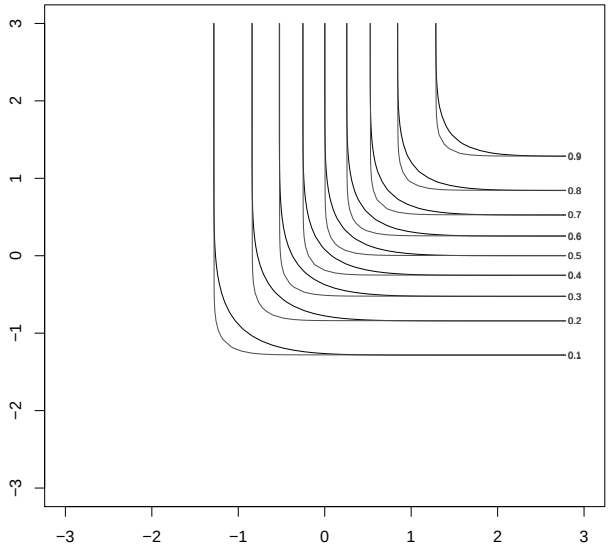
	Diff
N=100	0.72
N=500	0.63
N=1000	0.64

Las siguientes dos gráficas nos muestran la función de densidad del paramétro habilidad y la función de densidad estimada de este parámetro.





En el siguiente gráfico tenemos las curvas de nivel de probabilidad acumulada de la verdadera distribución del parámetro habilidad, en negro, y de la estimación de ésta, en rojo.



4.8. Estimación monótona

Uno de los postulados básicos de la TRI es que la relación entre el desempeño de un examinado en un ítem y los rasgos implícitos del ítem pueden describirse con una función que aumenta monótonamente, llamada curva característica del ítem. Esta función especifica que a medida que aumenta la habilidad de un individuo también aumenta la probabilidad de responder correctamente.

Nuestras estimaciones por regresión no paramétrica no cumplen con este postulado, por lo cual, basados en el artículo de Dette, Neumeyer y Pilz (2006): «A simple nonparametric estimator of a strictly monotone regression function», realizamos una transformación sobre nuestras estimaciones con el objetivo de que éstas sean monótonas crecientes. Este artículo propone:

$$\hat{m}^{-1}(t) := \frac{1}{Nh_d} \sum_{i=1}^{N} \int_{-\infty}^{t} K_d \left(\frac{\hat{m}(i/N) - u}{h_d} \right)$$

como una estimación de $m^{-1}(t)$, donde

$$\hat{m}(x) = \frac{\sum_{i=1}^{N} K_r \left(\frac{X_i - x}{h_r} \right) Y_i}{\sum_{i=1}^{N} K_r \left(\frac{X_i - x}{h_r} \right)}$$

4.8.1. Tamaño muestral

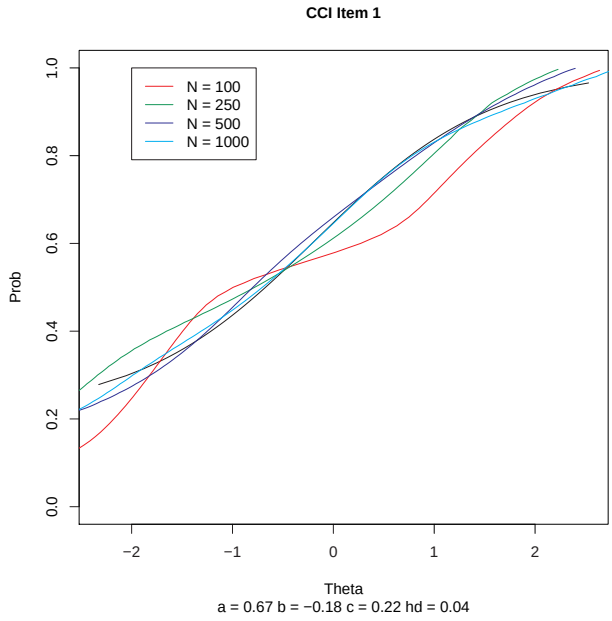
Para evaluar el efecto del tamaño muestral utilizamos un test de 30 ítems aplicado a 100, 250, 500 y 1000 individuos cuyas habilidades provienen de una distribución normal(0,1). Estimamos las CCI mediante regresión no paramétrica utilizando un ancho de ventana de igual a 0.5 y un núcleo normal. A estas estimaciones les aplicamos la transformación descrita anteriormente.

Los siguientes cuadros contienen la media de cada una de las tres distancias consideradas, calculadas en todos los ítems del test. En la primera tabla las distancias fueron calculadas entre las CCI estimadas por la regresión no paramétrica y las CCI reales. En la segunda tabla las distancias fueron calculadas entre las transformaciones de las CCI estimadas y las CCI reales.

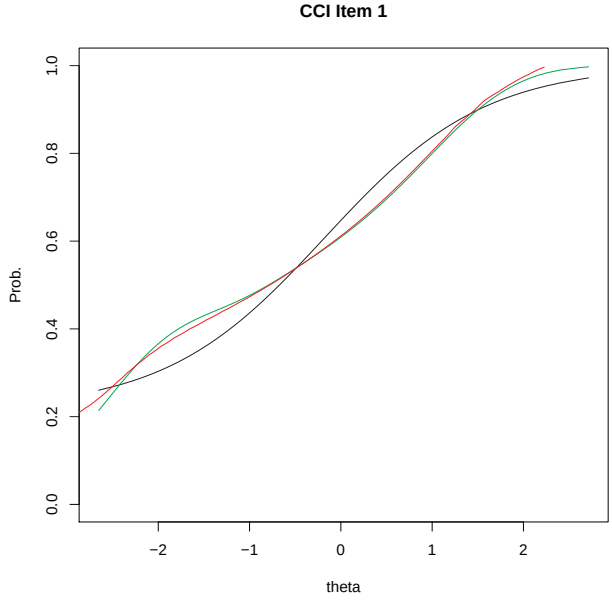
	media de d1	media de d2	media de d3
N = 100	0.1547	0.0748	0.1197
N = 250	0.1187	0.0557	0.0746
N = 500	0.1352	0.0608	0.0941
N = 1000	0.1221	0.0563	0.0916

	media de d1	media de d2	media de d3
N = 100	0.2350	0.0802	0.1190
N = 250	0.1715	0.0591	0.0800
N = 500	0.1922	0.0651	0.0863
N = 1000	0.1489	0.0562	0.0931

El siguiente gráfico contiene las transformaciones de las estimaciones de las CCI, en colores, y la CCI real, en negro.



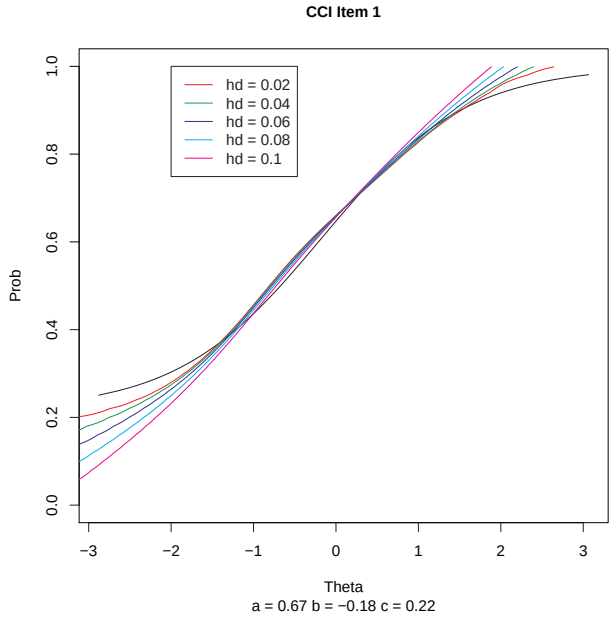
En el siguiente gráfico hay una CCI, una estimación por regresión no paramétrica de dicha CCI y la transformación de esta estimación.



4.8.2. Efecto de distintos valores de ancho para la ventana

Utilizando el mismo test de la sección anterior aplicado a 500 individuos, estudiamos el efecto de considerar distintos anchos de ventana en la transformación.

	media de d1	media de d2	media de d3
reg. n-p	0.1352	0.0608	0.0941
hd = 0.02	0.2336	0.0773	0.0893
hd = 0.04	0.1925	0.0650	0.0863
hd = 0.06	0.1769	0.0599	0.0824
hd = 0.08	0.1629	0.0566	0.0828
hd = 0.1	0.1525	0.0542	0.0837



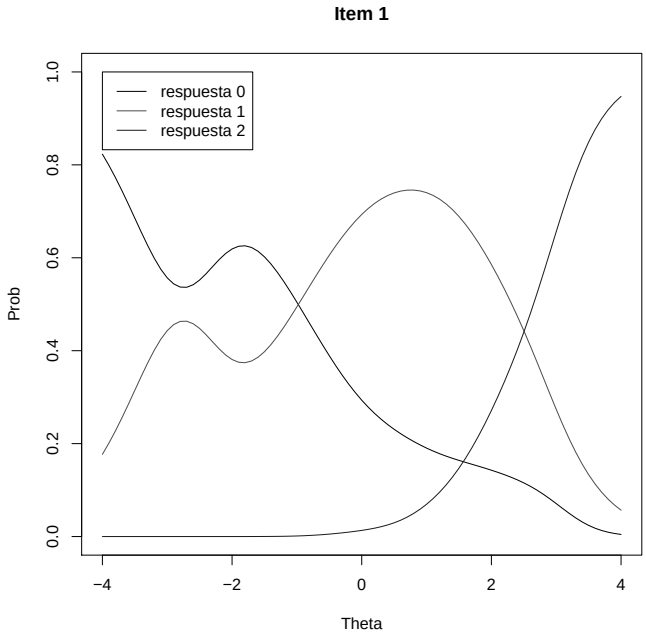
4.9. Aplicación a datos reales: test CDI

El Children Depression Inventory (CDI) es un test psicológico de respuestas politómicas (0, 1 y 2) en el cual se miden cuatro rasgos en distintos ítems tal como lo explica la siguiente tabla.

Rasgo medido	Ítems
1 Autoestima Negativa Retraimiento	1, 4, 7, 10, 14, 20, 21, 22, 25
2 Autoestima Negativa Oposicionamiento	5, 9, 17, 26, 27
3 Ineficacia	2, 3, 13, 15, 23, 24
4 Ansiedad	6, 11, 19

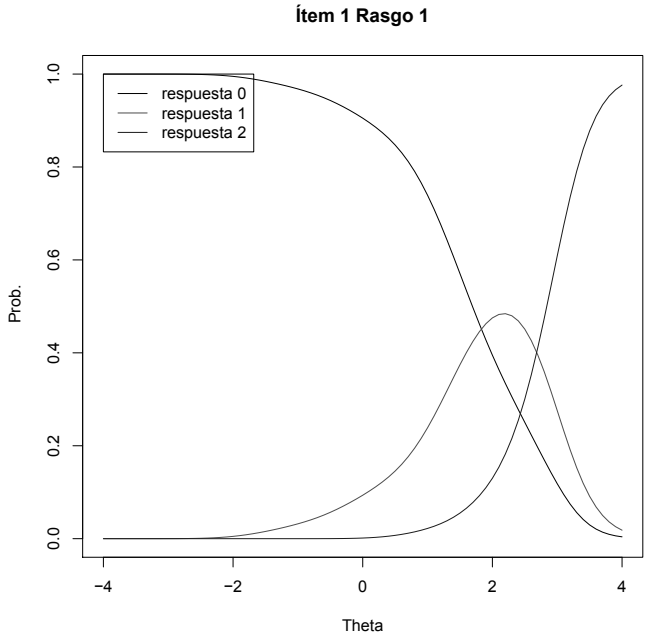
Este test fue aplicado a 605 personas de las cuales 324 son de sexo femenino, 280 de sexo masculino y para 1 se desconoce el género del individuo.

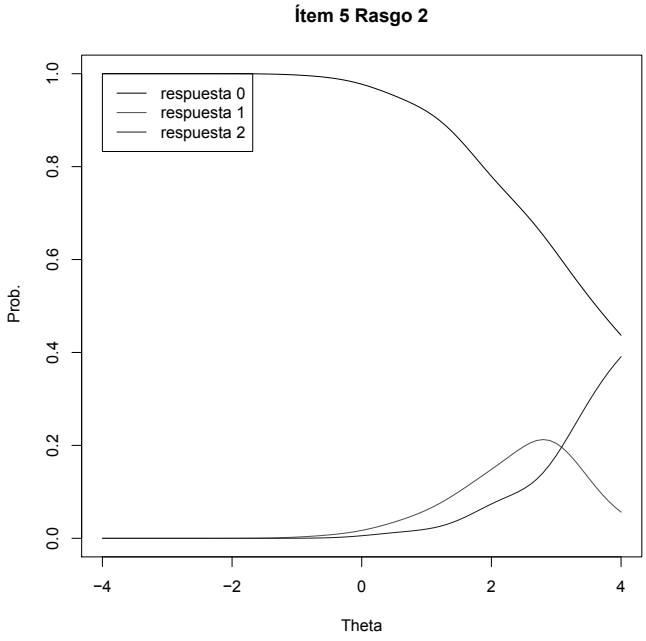
En una primera instancia consideramos el test como unidimensional, por lo tanto utilizaremos como medida de resumen para estimar el rasgo la media de los puntajes obtenidos por cada individuo. A continuación presentamos el gráfico del ítem 1. Las curvas fueron estimadas por regresión no paramétrica utilizando un núcleo normal y un ancho de ventana igual a 0,5.

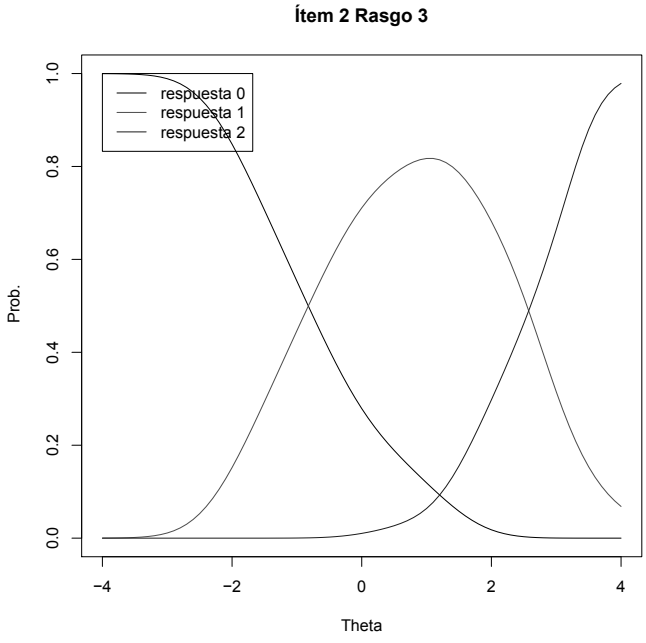


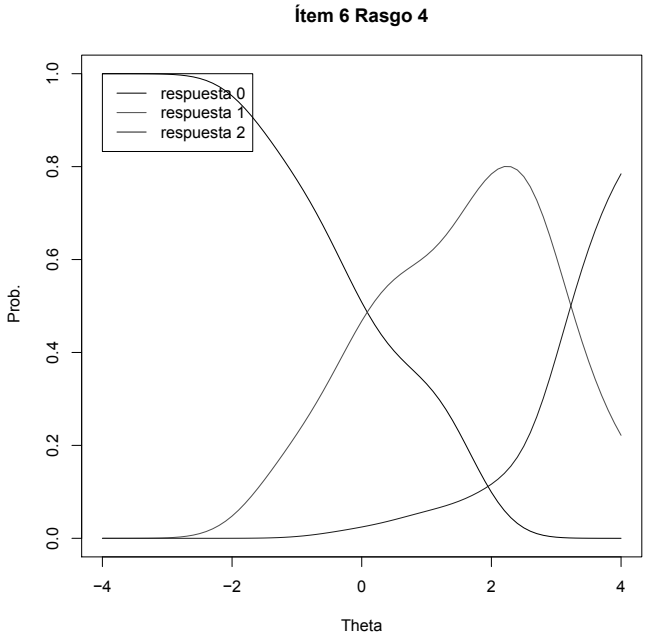
4.9.1. Estimación de las CCI diferenciando por rasgo

En esta sección estimamos nuevamente las CCI para cada ítem pero en este caso diferenciamos por rasgo, es decir, estimamos cada rasgo utilizando la media de los puntajes en los ítems en los cuales el rasgo en cuestión es medido y con base en estas estimaciones de los rasgos estimamos las CCI. Para visualizar esto gráficamente mostraremos las CCI de cuatro ítems distintos, cada uno relacionado a un rasgo distinto.



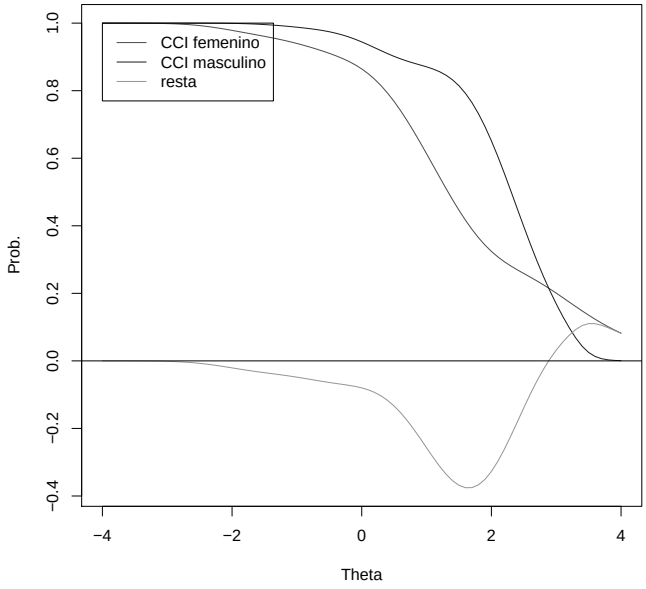

Ítem 5 Rasgo 2
respuesta 0
respuesta 1
respuesta 2
Prob.
0.0
0.2
0.4
0.6
0.8
1.0
−4
−2
0
2
4
Theta


Ítem 2 Rasgo 3
respuesta 0
respuesta 1
respuesta 2
Prob.
0.0
0.2
0.4
0.6
0.8
1.0
−4
−2
0
2
4
Theta


Ítem 6 Rasgo 4
respuesta 0
respuesta 1
respuesta 2
Prob.
0.0
0.2
0.4
0.6
0.8
1.0
−4
−2
0
2
4
Theta

4.9.2. Estudio diferencial del ítem

Para realizar este estudio estimamos las CCI para hombres y mujeres por separado y luego calculamos el área entre ellas. A modo de ejemplo presentamos la siguiente gráfica que contiene la CCI estimada para cada uno de los géneros para el ítem 1 y el rasgo 1.



La siguiente tabla contiene el área promedio para todos los ítems del rasgo en cuestión de la resta de las CCI diferenciadas por género.

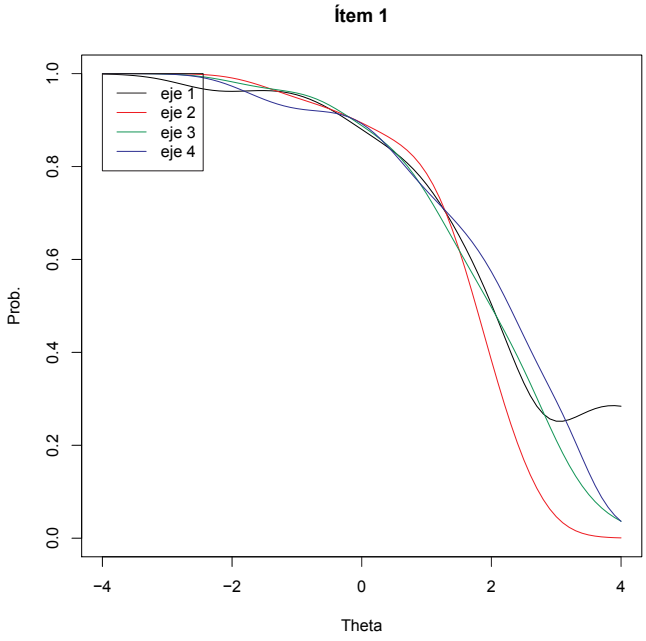
	rasgo 1	rasgo 2	rasgo 3	rasgo 4
área DIF respuesta 0	1.154	1.133	1.027	1.233
área DIF respuesta 1	1.076	1.143	1.009	1.114
área DIF respuesta 2	0.746	0.730	0.635	0.513

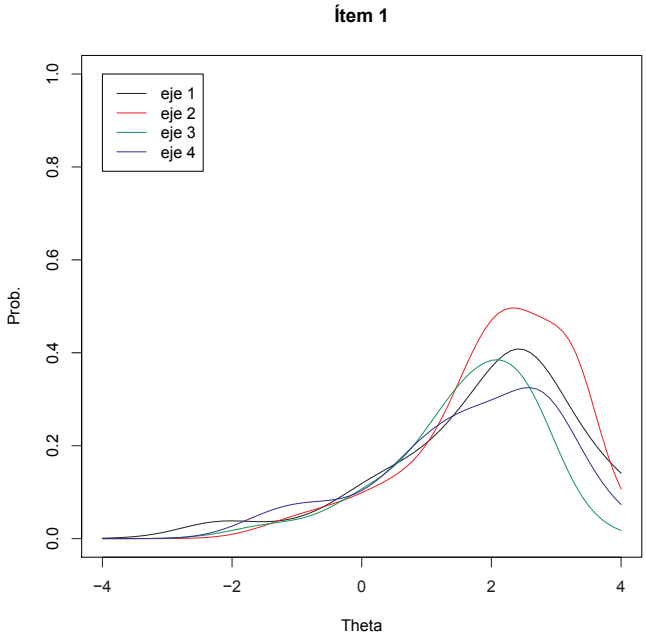
4.9.3. Análisis factorial

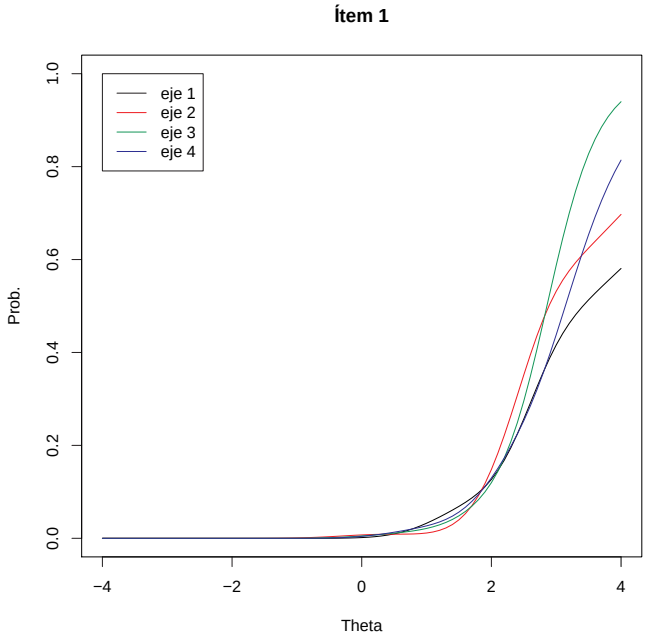
Para realizar el análisis factorial calculamos las correlaciones policóricas entre los ítems, luego encontramos los ejes factoriales y rotamos los mismos según el criterio de varimax. Una vez que contamos con los ejes factoriales rotados proyectamos los resultados de los distintos individuos en ellos y utilizamos estas proyecciones como función para ordenar a los individuos y finalmente estimar las CCI. En este ánalisis consideramos adecuado retener los primeros cuatro que acumulan un $35,6\,\%$ de la varianza total. La siguiente tabla muestra el peso relativo de cada uno de los ítems en cada eje.

	PC1	PC2	PC3	PC4
c1	0.26	0.44	0.29	0.19
c2	0.61	0.12	0.14	-0.07
c3	0.58	0.18	0.11	0.21
c4	0.11	-0.02	0.60	-0.04
c5	-0.04	-0.02	-0.07	0.64
c6	0.11	0.41	0.05	0.26
c7	0.49	0.15	0.30	0.18
c8	0.59	0.26	0.02	-0.01
c9	0.33	0.14	0.23	0.41
c10	0.27	0.48	0.29	0.06
c11	-0.06	0.60	-0.02	0.02
c12	0.16	0.07	0.54	0.37
c13	0.38	0.14	0.18	0.00
c14	0.35	0.07	0.48	0.01
c15	0.43	-0.16	-0.16	0.13
c16	0.36	0.10	0.10	0.12
c17	0.11	0.15	0.16	0.54
c18	0.30	0.01	0.11	0.09
c19	0.01	0.59	-0.15	-0.23
c20	0.36	0.31	0.41	0.31
c21	0.06	-0.13	0.65	0.01
c22	0.04	0.09	0.58	0.07
c23	0.56	-0.28	0.03	0.15
c24	0.49	-0.14	0.13	0.14
c25	0.38	0.24	0.37	0.30
c26	0.15	-0.11	-0.04	0.43
c27	0.16	0.04	0.11	0.67

Las siguientes gráficas muestran las CCI obtenidas a partir de cada uno de los cuatro ejes factoriales.







4.10. Funciones

4.10.1. Funciones para calcular las distancias

Para el cálculo de las distancias antes definidas en la sección 4.1 generamos las siguientes funciones donde:

- y1 es el vector con los valores estimados;
- y2 es el vector con los valores reales;
- y3 es el vector con los valores de theta;
- delta1 es delta theta1;
- delta2 es delta theta2 (solo para los casos de dos habilidades);
- cop es la cópula (solo para los casos de dos habilidades correlacionadas).

```
> distancias

function (y1, y2, y3, delta)
{
    d1 <- sum((y1 - y2)^2 * delta)^(0.5)
    d3 <- abs(max(y1 - y2))
    d2 <- sum((y1 - y2)^2 * dnorm(y3) * delta)^(0.5)
    return(list(d1 = d1, d2 = d2, d3 = d3))
}

> distancias2ind
```

```
function (y1, y2, y3, delta1, delta2)
{
    d1 <- sum((y1 - y2)^2 * delta1 * delta2)^(0.5)
    d2 <- sum((y1 - y2)^2 * dnorm(y3[, 1], mu[1], sigma[1, 1]) *
        dnorm(y3[, 2], mu[2], sigma[2, 2]) * delta1 * delta2)^(0.5)
    d3 <- max(y1 - y2)
    return(list(d1 = d1, d2 = d2, d3 = d3))
}
```

```
> distancias2cop
```

```
function (y1, y2, y3, delta1, delta2, cop)
{
    d1 <- sum((y1 - y2)^2 * delta1 * delta2)^(0.5)
    d2 <- sum((y1 - y2)^2 * dmvdc(cop, y3) * delta1 * delta2)^(0.5)
    d3 <- max(y1 - y2)
    return(list(d1 = d1, d2 = d2, d3 = d3))
}
```

4.10.2. Funciones para la generación de datos

Para generar los tests como fue descrito en la sección 4.2 utilizamos las siguientes funciones para la generación de datos provenientes de modelos logísticos de uno, dos y tres parámetros, provenientes del modelo logit cúbico y tests multidimensionales.

- a valores del parámetro de discriminación;
- b valores del parámetro de dificultad;
- c valores del parámetro de azar;
- theta vector o matriz de habilidades de los examinados;
- nitems (solo para ...).

```
> genresp
```

```
function (a, b, c, theta, tot = 0)
{
    D = 1.7
    if (tot == 0) {
        resp <- matrix(nrow = length(theta), ncol = length(a))
    }
    if (tot == 1) {
        resp <- matrix(nrow = length(theta), ncol = length(a) +
            2)
    }
    for (j in 1:length(a)) {
        print(j)
        for (i in 1:length(theta)) {
            Prob <- c[j] + ((1 - c[j]) * exp(D * a[j] * (theta[i] -
                b[j])))/(1 + exp(D * a[j] * (theta[i] - b[j])))
            aux <- runif(1, 0, 1)
            if (Prob > aux) {
                resp[i, j] = 1
            }
            else {
                resp[i, j] = 0
```

```
            }
        }
    }
    if (tot == 1) {
        su = 1/length(theta)
        resp[, (ncol(resp) - 1)] = rowSums(resp, na.rm = T)
        resp[, ncol(resp)] = resp[, (ncol(resp) - 1)] + runif(nrow(resp),
            0, su)
    }
    resp
}
```

> genresplogit3

```
function (a, b, c, theta, tot = 0)
{
    D = 1.7
    if (tot == 0) {
        resp <- matrix(nrow = length(theta), ncol = length(a))
    }
    if (tot == 1) {
        resp <- matrix(nrow = length(theta), ncol = length(a) +
            2)
    }
    for (j in 1:length(a)) {
        print(j)
        for (i in 1:length(theta)) {
            Prob <- exp(a[j] + b[j] * theta[i] + c[j] * theta[i]^3)/(1 +
                exp(a[j] + b[j] * theta[i] + c[j] * theta[i]^3))
            aux <- runif(1, 0, 1)
            if (Prob > aux) {
                resp[i, j] = 1
            }
            else {
                resp[i, j] = 0
            }
        }
    }
    if (tot == 1) {
        su = 1/length(theta)
        resp[, (ncol(resp) - 1)] = rowSums(resp, na.rm = T)
        resp[, ncol(resp)] = resp[, (ncol(resp) - 1)] + runif(nrow(resp),
            0, su)
    }
    resp
}
```

> genresp2

```
function (nitems, a, b, c, thetas)
{
    D = 1.7
    resp <- matrix(nrow = nrow(thetas), ncol = (sum(nitems) +
        length(nitems)))
    Probs = Probas = list()
    for (j in 1:length(a)) {
        print(j)
        if (j <= nitems[1]) {
            for (i in 1:nrow(thetas)) {
                Prob <- c[j] + (1 - c[j]) * exp(D * a[j] * (thetas[i,
```

```
                1] - b[j]))/(1 + exp(D * a[j] * (thetas[i,
                1] - b[j])))
            aux <- runif(1, 0, 1)
            if (Prob > aux) {
              resp[i, j] = 1
            }
            else {
              resp[i, j] = 0
            }
            resp[i, (ncol(resp) - 1)] = sum(resp[i, 1:nitems[1]]) +
              runif(1, 0, 1/nrow(resp))
            Probs[i] = Prob
        }
    }
    if (j >= nitems[1]) {
        for (i in 1:nrow(thetas)) {
            Prob <- c[j] + (1 - c[j]) * exp(D * a[j] * (thetas[i,
              2] - b[j]))/(1 + exp(D * a[j] * (thetas[i,
              2] - b[j])))
            aux <- runif(1, 0, 1)
            if (Prob > aux) {
              resp[i, j] = 1
            }
            else {
              resp[i, j] = 0
            }
            resp[i, ncol(resp)] = sum(resp[i, (nitems[1] +
              1):sum(nitems)]) + runif(1, 0, 1/nrow(resp))
            Probs[i] = Prob
        }
    }
    Probas[[j]] = do.call(rbind, Probs)
}
return(list(resp = resp, Probas = Probas))
}
```

> genresp2cop

```
function (a, b, c, thetas)
{
    D = 1.7
    resp <- matrix(nrow = nrow(thetas), ncol = (length(b) + 1))
    Probs = Probas = list()
    for (j in 1:length(b)) {
        print(j)
        for (i in 1:nrow(thetas)) {
            Prob <- c[j] + (1 - c[j]) * exp(D * a[j, 1] * thetas[i,
                1] + D * a[j, 2] * thetas[i, 2] - b[j])/(1 +
                exp(D * a[j, 1] * thetas[i, 1] + D * a[j, 2] *
                  thetas[i, 2] - b[j]))
            aux <- runif(1, 0, 1)
            if (Prob > aux) {
                resp[i, j] = 1
            }
            else {
                resp[i, j] = 0
            }
            resp[i, (ncol(resp))] = sum(resp[i, 1:length(b)])
            Probs[i] = Prob
        }
```

```
        Probas[[j]] = do.call(rbind, Probs)
    }
    return(list(resp = resp, Probas = Probas))
}
```

4.10.3. Funciones para la estimación de la CCI

Para las estimaciones de las CCI en pruebas unidimensionales utilizamos la función iccsmoothing, mientras que para las estimaciones en pruebas multidimensionales utilizamos las funciones pesos y thetahat. Con la función thetahat estimamos las habilidades de los individuos y con la función pesos calculamos los pesos de la regresión no paramétrica:

- ítems;
- h valor del ancho de ventana;
- testsort test ordenado;
- ability vector o matriz de habilidades de los examinados estimadas o reales;
- points vector de puntos donde la curva será evaluada;
- nucleo=ñorél núcleo a utilizar en la regresión no paramétrica.

```
> iccsmoothing

function (items, h, testsort, ability, points, nucleo = "nor")
{
    npoints <- length(points)
    nexaminees <- length(ability)
    nitems <- length(items)
    icc <- matrix(0, npoints, nitems)
    weigth <- matrix(0, npoints, nexaminees)
    icchat <- matrix(0, nexaminees, nitems)
    source("nucleos/normal.R")
    source("nucleos/epa.R")
    source("nucleos/md1.R")
    source("nucleos/md2.R")
    for (i in 1:npoints) {
        theta <- points[i]
        residual <- (theta - ability)/h
        if (nucleo == "nor") {
            kernel <- normal(residual)
        }
        if (nucleo == "epa") {
            kernel <- epa(residual)
        }
        if (nucleo == "md1") {
            kernel <- md1(residual)
        }
        if (nucleo == "md2") {
            kernel <- md2(residual)
        }
        weigth[i, kernel$windex] <- kernel$res/sum(kernel$res)
    }
    icc <- weigth %*% testsort[, items]
    for (j in 1:nitems) {
```

```
        icchat[, j] <- approx(points, icc[, j], ability)$y
        icchat[is.na(icchat[, j])] <- 0
    }
    return(list(icc = icc))
}
```

- nitems número de ítems del test;
- test;
- ability vector o matriz de habilidades de los examinados estimadas o reales:

```
> thetahat

function (nitems, test, ability)
{
    test = cbind(test, c(1:dim(test)[1]))
    res = list()
    for (j in 1:length(nitems)) {
        if (j == 1) {
            testsort = test[order(test[, (sum(nitems) + 1)]),
                c(1:nitems[1], (sum(nitems) + 1), ncol(test))]
        }
        if (j == 2) {
            testsort = test[order(test[, (sum(nitems) + 2)]),
                c((nitems[1] + 1):sum(nitems), (sum(nitems) +
                  2), ncol(test))]
        }
        habemp = seq(1, length(test[, 1]), 1)/length(test[, 1])
        habemp[length(test[, 1])] = (length(test[, 1]) - 1 +
            runif(1, 0, 1))/length(test[, 1])
        habilidades = qnorm(habemp)
        res[[j]] = cbind(testsort[, ncol(testsort)], habemp,
            habilidades)
    }
    aux1 = res[[1]]
    aux2 = res[[2]]
    colnames(aux1) = c("indice", "habemp", "f-1")
    colnames(aux2) = c("indice", "habemp", "f-1")
    result = merge(aux1, aux2, by.x = "indice", by.y = "indice")
    return(result)
}
```

- a valores de los parámetros de discriminación;
- b valores del parámetro de dificultad;
- c valores del parámetro de azar;
- test:

```
> thetahatcop

function (a, b, c, test)
{
    t1 = t2 = list()
    for (i in 1:(ncol(test) - 2)) {
        t1[[i]] = (a[i, 1]/a[i, 2]) * test[, i]
```

```
        t2[[i]] = (a[i, 2]/a[i, 1]) * test[, i]
    }
    t1 = do.call(cbind, t1)
    t2 = do.call(cbind, t2)
    test1 = cbind(rowSums(t1), 1:nrow(test))
    test2 = cbind(rowSums(t2), 1:nrow(test))
    habemp = seq(1, length(test[, 1]), 1)/length(test[, 1])
    habemp[length(test[, 1])] = (length(test[, 1]) - 1 + runif(1,
        0, 1))/length(test[, 1])
    test1 = cbind(test1[order(test1[, 1]), 2], habemp, qnorm(habemp))
    test2 = cbind(test2[order(test2[, 1]), 2], habemp, qnorm(habemp))
    colnames(test1) = c("indice", "habemp", "f-1")
    colnames(test2) = c("indice", "habemp", "f-1")
    result = merge(test1, test2, by.x = "indice", by.y = "indice")
    return(result)
}
```

npoints,that,h,nucleo=ñor'

- npoints raíz del número de puntos en que será evaluada la CCI;
- that la salida de las funciones thetahat o thetahatcop;
- h valor del ancho de ventana;
- núcleo=ñorél núcleo a utilizar en la regresión no paramétrica:

```
> pesos

function (npoints, that, h, nucleo = "nor")
{
    source("nucleos 2D/multinorm.R")
    source("nucleos 2D/epa.R")
    source("nucleos 2D/md1.R")
    source("nucleos 2D/md2.R")
    that = that[, seq(2, dim(that)[2], 2)]
    eval = seq(0, 1, length = npoints)
    eval[1] = runif(1, 0, eval[2])
    eval[npoints] = eval[npoints] - runif(1, 0, eval[2])
    nind = c(as.numeric(dim(that)[1]))
    pesos = list()
    puntos = list()
    a = 1
    for (j in 1:length(eval)) {
        print(Sys.time())
        print(j)
        for (i in 1:length(eval)) {
            res = (cbind(rep(eval[i], dim(that)[1]), rep(eval[j],
                dim(that)[1])) - that)/h
            kernel = vector(length = dim(res)[1])
            puntos[[a]] = c(eval[i], eval[j])
            if (nucleo == "nor") {
                kernel = multinorm(res, sigma)
            }
            if (nucleo == "epa") {
                kernel = epa(res)
            }
            if (nucleo == "md1") {
```

```
            kernel = md1(res)
        }
        if (nucleo == "md2") {
            kernel = md2(res)
        }
        weigth = numeric(length(kernel$res))
        weigth[kernel$windex] <- kernel$res[kernel$windex]/sum(kernel$res[kernel$windex])
        pesos[[a]] = weigth
        a = a + 1
    }
  }
  pesos = do.call(rbind, pesos)
  puntos = do.call(rbind, puntos)
  return(list(pesos = pesos, puntos = puntos))
}
```

4.10.4. Distintos núcleos para la regresión no paramétrica

A continuación presentamos las funciones para los distintos núcleos utilizados en la regresión no paramétrica, en una y dos dimensiones. En ambos casos el único argumento es el vector de datos a ser evaluado.

Núcleos en una dimensión

```
> normal

function (x)
{
    res = numeric(length(x))
    windex <- abs(x) <= 3000
    res = exp(-(x[windex]^2)/2)
    return(list(res = res, windex = windex))
}

> epa

function (x)
{
    res = numeric(length(x))
    windex = abs(x * x) <= 1
    res = 1 - x[windex] * x[windex]
    return(list(res = res, windex = windex))
}

> md1

function (x)
{
    res = numeric(length(x))
    windex = abs(x * x) <= 1
    res = 3 * pi^(-1) * (1 - x[windex] * x[windex])
    return(list(res = res, windex = windex))
}

> md2

function (x)
{
    res = numeric(length(x))
    windex = abs(x * x) <= 1
    res = 4 * pi^(-1) * (1 - x[windex] * x[windex])^3
    return(list(res = res, windex = windex))
}
```

```
> multinorm

function (x, sigma)
{
    x = as.matrix(x)
    library(MASS)
    res = numeric(length = (dim(x)[1]))
    n = ncol(sigma)
    for (i in 1:length(res)) {
        res[i] = exp((t(x[i, ]) %*% ginv(sigma) %*% x[i, ])/-2)/(2 *
            pi)^(n/2) * det(sigma)^(1/2)
    }
    windex = abs(res) <= 3000
    return(list(res = res, windex = windex))
}

> epa

function (x)
{
    x = as.matrix(x)
    res = numeric(length = (dim(x)[1]))
    windex = vector(length = (dim(x)[1]))
    for (i in 1:length(windex)) {
        if (t(x[i, ]) %*% x[i, ] <= 1) {
            windex[i] = "TRUE"
        }
        res[i] = 1 - t(x[i, ]) %*% x[i, ]
    }
    windex = as.logical(windex)
    return(list(res = res, windex = windex))
}

> md1

function (x)
{
    x = as.matrix(x)
    res = numeric(length = (dim(x)[1]))
    windex = vector(length = (dim(x)[1]))
    for (i in 1:length(windex)) {
        if (t(x[i, ]) %*% x[i, ] <= 1) {
            windex[i] = "TRUE"
        }
        res[i] = 3 * pi^(-1) * (1 - t(x[i, ]) %*% x[i, ])
    }
    windex = as.logical(windex)
    return(list(res = res, windex = windex))
}

> md2

function (x)
{
    x = as.matrix(x)
    res = numeric(length = (dim(x)[1]))
    windex = vector(length = (dim(x)[1]))
    for (i in 1:length(windex)) {
        if (t(x[i, ]) %*% x[i, ] <= 1) {
```

```
            windex[i] = "TRUE"
        }
        res[i] = 4 * pi^(-1) * (1 - t(x[i, ]) %*% x[i, ])^3
    }
    windex = as.logical(windex)
    return(list(res = res, windex = windex))
}
```

4.10.5. Función para transformar el rango

En algunas secciones se consideró el parámetro habilidad proveniente de variables aleatorias con rango $[0, 1]$, en estos casos se utilizó la siguiente función con el objetivo de transformar el rango a $(-\infty, \infty)$.

```
> cambiosoporte

function (x)
{
    y = x
    y[which(x < 0.5)] = (2 * x[which(x < 0.5)] - 1)/(2 * x[which(x <
        0.5)])
    y[which(x > 0.5)] = (2 * x[which(x > 0.5)] - 1)/(-2 * x[which(x >
        0.5)] + 2)
    return(y)
}
```

4.10.6. Funciones utilizadas para el ánalisis del test CDI

La siguiente función calcula el puntaje total permitiendo diferenciar según el género del individuo y el rasgo de interés, además devuelve el test en forma dicotómica donde es considerada como correcta la respuesta a evaluar.

```
> test.cdi

function (test, respuesta, genero = NULL, rasgo = NULL)
{
    if (!is.null(genero)) {
        test = test[which(test\$sexo == genero), 2:ncol(test)]
    }
    if (is.null(genero)) {
        test = test[, 2:ncol(test)]
    }
    if (is.null(rasgo)) {
        test[, (ncol(test) + 1)] = rowSums(test[, 2:ncol(test)],
            na.rm = T)
    }
    anretraimiento = c(2, 5, 8, 11, 15, 21, 22, 23, 26) - 1
    anoposicionamiento = c(6, 10, 18, 27, 28) - 1
    ineficacia = c(3, 4, 14, 15, 24, 25) - 1
    ansiedad = c(7, 12, 20) - 1
    if (!is.null(rasgo)) {
        if (rasgo == 1) {
            test[, (ncol(test) + 1)] = rowSums(test[, anretraimiento],
                na.rm = T)
```

```
        }
        if (rasgo == 2) {
            test[, (ncol(test) + 1)] = rowSums(test[, anoposicionamiento],
                na.rm = T)
        }
        if (rasgo == 3) {
            test[, (ncol(test) + 1)] = rowSums(test[, ineficacia],
                na.rm = T)
        }
        if (rasgo == 4) {
            test[, (ncol(test) + 1)] = rowSums(test[, ansiedad],
                na.rm = T)
        }
    }
    aux = which(levels(as.factor(test[, 2])) == respuesta)
    for (i in 1:(ncol(test) - 1)) {
        test[, i] = as.factor(test[, i])
        levels(test[, i])[aux] = "9"
        levels(test[, i])[-aux] = "0"
        levels(test[, i])[which(levels(test[, i]) == "9")] = "1"
        test[, i] = as.numeric(as.character(test[, i]))
    }
    test = matrix(unlist(test), ncol = ncol(test), nrow = nrow(test))
    storage.mode(test) = "numeric"
    return(test)
}
```

Para el estudio diferencial del ítem se creó la siguiente función que calcula el área de dos funciones cualesquiera, así como también el área positiva y el área negativa de la resta de éstas. En este caso las funciones utilizadas son las CCI estimadas para hombres y mujeres. También permite graficar ambas funciones y la resta de éstas.

```
> comparacion

function (base1, base2, points, pl = FALSE)
{
    base1 <- as.numeric(base1)
    base2 <- as.numeric(base2)
    sumainf1 = sumainf2 = list()
    for (i in 1:(length(base1) - 1)) {
        sumainf1[i] <- min(base1[i], base1[(i + 1)]) * (points[i +
            1] - points[i])
        sumainf2[i] <- min(base2[i], base2[(i + 1)]) * (points[i +
            1] - points[i])
    }
    suminf1 = sum(unlist(sumainf1))
    suminf2 = sum(unlist(sumainf2))
    sumasup1 = sumasup2 = list()
    for (i in 1:(length(base2) - 1)) {
        sumasup1[i] <- max(base1[i], base1[(i + 1)]) * (points[i +
            1] - points[i])
        sumasup2[i] <- max(base2[i], base2[(i + 1)]) * (points[i +
            1] - points[i])
    }
    sumsup1 = sum(unlist(sumasup1))
    sumsup2 = sum(unlist(sumasup2))
    area1 = mean(c(sumsup1, suminf1))
    area2 = mean(c(sumsup2, suminf2))
```

```
    resta = base1 - base2
    suma.pos = suma.neg = list()
    for (i in 1:(length(resta) - 1)) {
        if (mean(c(resta[i], resta[i + 1])) < 0) {
            suma.neg[i] = mean(c(resta[i], resta[i + 1])) * (points[i +
                1] - points[i])
        }
        if (mean(c(resta[i], resta[i + 1])) > 0) {
            suma.pos[i] = mean(c(resta[i], resta[i + 1])) * (points[i +
                1] - points[i])
        }
    }
    suma.pos = sum(unlist(suma.pos))
    suma.neg = -sum(unlist(suma.neg))
    mini = min(resta)
    if (pl == TRUE) {
        plot(points, base1, type = "l", col = "red", main = "",
            ylim = c(mini, 1), ylab = "Prob.", xlab = "Theta")
        lines(points, base2, type = "l", col = "blue")
        lines(points, resta, type = "l", col = "green")
        abline(h = 0)
        legend(-4, 1, legend = c("CCI femenino", "CCI masculino",
            "resta"), col = c("red", "blue", "green"), lwd = 1)
    }
    return(list(area1 = area1, area2 = area2, area.pos.resta = suma.pos,
        area.neg.resta = suma.neg))
}
```

4.10.7. Función para transformación monótona creciente

La estimación de la CCI por regresión no paramétrica no es necesariamente una curva que crezca monótonamente a medida que crece el parámetro habilidad. Para solucionar este inconveniente utilizamos esta función:

```
> mber

function (icc, hd, nt, puntos)
{
    t = c(1:nt/nt)
    t[length(t)] = 0.9999
    res = numeric(length(t))
    resfin = numeric(length(puntos))
    for (i in 1:length(t)) {
        int = list()
        for (j in 1:length(icc)) {
            integrand = function(x) {
                exp(-((icc[j] - x)/hd)^2/2)/((2 * pi)^0.5)
            }
            int[j] = integrate(integrand, -Inf, t[i])[[1]]
        }
        res[i] = (1/(length(icc) * hd)) * sum(unlist(int))
    }
    t = c(max(t[1] - ((t[2] - t[1])/(res[2] - res[1])) * res[1],
        0), t)
    res = c(5e-04, res)
    resfin <- approx(t, res, puntos) \$ y
    return(list(resfin = resfin, puntos = puntos))
}
```

Apéndice A

A.1. Modelo de un parámetro

Los valores de a utilizados fueron: 0.64; 0.79; 0.81; 0.87; 1.05; 1.20; 1.55; 1.58; 1.70;

A.1.1. Distribución normal

Trabajaremos con 100,250,500,1000,2500,5000,10000 datos con distribución N(0,1)

n=100

a	h	d_1	d_2	d_3
0.6352	0.3981	0.0188	0.0082	0.0110
0.7910	0.3981	0.0196	0.0080	0.0098
0.8082	0.3981	0.0197	0.0080	0.0096
0.8119	0.3981	0.0197	0.0080	0.0096
0.8683	0.3981	0.0201	0.0080	0.0091
1.0525	0.3981	0.0215	0.0078	0.0075
1.1999	0.3981	0.0227	0.0077	0.0077
1.5513	0.3981	0.0253	0.0071	0.0130
1.5768	0.3981	0.0255	0.0070	0.0135
1.6995	0.3981	0.0263	0.0067	0.0159

n=250

a	h	d_1	d_2	d_3
0.6352	0.3314	0.0134	0.0061	0.0079
0.7910	0.3314	0.0139	0.0060	0.0080
0.8082	0.3314	0.0139	0.0060	0.0082
0.8119	0.3314	0.0139	0.0060	0.0083
0.8683	0.3314	0.0141	0.0060	0.0088
1.0525	0.3314	0.0147	0.0059	0.0107
1.1999	0.3314	0.0150	0.0057	0.0122
1.5513	0.3314	0.0154	0.0051	0.0146
1.5768	0.3314	0.0153	0.0051	0.0147
1.6995	0.3314	0.0152	0.0048	0.0149

n=500

a	h	d_1	d_2	d_3
0.6352	0.2885	0.0111	0.0045	0.0072
0.7910	0.2885	0.0122	0.0045	0.0059
0.8082	0.2885	0.0124	0.0045	0.0058
0.8119	0.2885	0.0124	0.0045	0.0058
0.8683	0.2885	0.0129	0.0045	0.0063
1.0525	0.2885	0.0146	0.0046	0.0085
1.1999	0.2885	0.0162	0.0047	0.0108
1.5513	0.2885	0.0207	0.0046	0.0184
1.5768	0.2885	0.0211	0.0046	0.0191
1.6995	0.2885	0.0231	0.0045	0.0227

n=1000

a	h	d_1	d_2	d_3
0.6352	0.2511	0.0091	0.0034	0.0043
0.7910	0.2511	0.0103	0.0034	0.0052
0.8082	0.2511	0.0104	0.0034	0.0053
0.8119	0.2511	0.0105	0.0034	0.0054
0.8683	0.2511	0.0110	0.0034	0.0059
1.0525	0.2511	0.0131	0.0034	0.0081
1.1999	0.2511	0.0152	0.0034	0.0103
1.5513	0.2511	0.0225	0.0033	0.0183
1.5768	0.2511	0.0232	0.0033	0.0190
1.6995	0.2511	0.0268	0.0033	0.0231

n=2500

a	h	d_1	d_2	d_3
0.6352	0.2091	0.0047	0.0019	0.0026
0.7910	0.2091	0.0051	0.0018	0.0025
0.8082	0.2091	0.0052	0.0018	0.0025
0.8119	0.2091	0.0052	0.0018	0.0026
0.8683	0.2091	0.0054	0.0018	0.0028
1.0525	0.2091	0.0061	0.0018	0.0038
1.1999	0.2091	0.0068	0.0018	0.0047
1.5513	0.2091	0.0091	0.0017	0.0079
1.5768	0.2091	0.0093	0.0017	0.0082
1.6995	0.2091	0.0103	0.0017	0.0097

n=5000

a	h	d_1	d_2	d_3
0.6352	0.1820	0.0035	0.0014	0.0021
0.7910	0.1820	0.0037	0.0014	0.0018
0.8082	0.1820	0.0038	0.0014	0.0018
0.8119	0.1820	0.0038	0.0014	0.0018
0.8683	0.1820	0.0039	0.0014	0.0017
1.0525	0.1820	0.0042	0.0014	0.0015
1.1999	0.1820	0.0045	0.0014	0.0013
1.5513	0.1820	0.0053	0.0013	0.0012
1.5768	0.1820	0.0053	0.0013	0.0012
1.6995	0.1820	0.0056	0.0013	0.0015

n=10000

a	h	d_1	d_2	d_3
0.6352	0.1584	0.0024	0.0009	0.0013
0.7910	0.1584	0.0027	0.0009	0.0015
0.8082	0.1584	0.0027	0.0009	0.0016
0.8119	0.1584	0.0027	0.0009	0.0016
0.8683	0.1584	0.0028	0.0009	0.0018
1.0525	0.1584	0.0032	0.0009	0.0024
1.1999	0.1584	0.0036	0.0009	0.0031
1.5513	0.1584	0.0048	0.0009	0.0054
1.5768	0.1584	0.0050	0.0009	0.0057
1.6995	0.1584	0.0055	0.0008	0.0069

A.1.2. Distribución uniforme

Trabajaremos con 100,250,500,1000,2500,5000,10000 observaciones

n=100

a	h	d_1	d_2	d_3
0.6352	0.3981	0.0177	0.0098	0.0380
0.7910	0.3981	0.0174	0.0095	0.0335
0.8082	0.3981	0.0173	0.0094	0.0330
0.8119	0.3981	0.0173	0.0094	0.0329
0.8683	0.3981	0.0170	0.0092	0.0312
1.0525	0.3981	0.0159	0.0085	0.0256
1.1999	0.3981	0.0146	0.0078	0.0215
1.5513	0.3981	0.0109	0.0058	0.0134
1.5768	0.3981	0.0107	0.0056	0.0129
1.6995	0.3981	0.0094	0.0049	0.0108

n=250

a	h	d_1	d_2	d_3
0.6352	0.3314	0.0124	0.0069	0.0309
0.7910	0.3314	0.0121	0.0066	0.0271
0.8082	0.3314	0.0120	0.0066	0.0267
0.8119	0.3314	0.0120	0.0066	0.0266
0.8683	0.3314	0.0119	0.0065	0.0252
1.0525	0.3314	0.0110	0.0059	0.0206
1.1999	0.3314	0.0102	0.0054	0.0173
1.5513	0.3314	0.0077	0.0040	0.0107
1.5768	0.3314	0.0075	0.0039	0.0103
1.6995	0.3314	0.0066	0.0034	0.0086

n=500

a	h	d_1	d_2	d_3
0.6352	0.2885	0.0095	0.0055	0.0288
0.7910	0.2885	0.0091	0.0051	0.0253
0.8082	0.2885	0.0091	0.0051	0.0249
0.8119	0.2885	0.0091	0.0051	0.0248
0.8683	0.2885	0.0089	0.0050	0.0235
1.0525	0.2885	0.0082	0.0045	0.0192
1.1999	0.2885	0.0075	0.0040	0.0161
1.5513	0.2885	0.0056	0.0030	0.0099
1.5768	0.2885	0.0054	0.0029	0.0096
1.6995	0.2885	0.0048	0.0025	0.0080

n=1000

a	h	d_1	d_2	d_3
0.6352	0.2511	0.0081	0.0046	0.0246
0.7910	0.2511	0.0078	0.0044	0.0215
0.8082	0.2511	0.0078	0.0043	0.0211
0.8119	0.2511	0.0078	0.0043	0.0211
0.8683	0.2511	0.0076	0.0042	0.0199
1.0525	0.2511	0.0071	0.0038	0.0163
1.1999	0.2511	0.0065	0.0035	0.0136
1.5513	0.2511	0.0049	0.0026	0.0084
1.5768	0.2511	0.0048	0.0025	0.0081
1.6995	0.2511	0.0042	0.0022	0.0067

n=2500

a	h	d_1	d_2	d_3
0.6352	0.2091	0.0066	0.0036	0.0198
0.7910	0.2091	0.0065	0.0035	0.0173
0.8082	0.2091	0.0065	0.0035	0.0170
0.8119	0.2091	0.0065	0.0035	0.0170
0.8683	0.2091	0.0064	0.0034	0.0160
1.0525	0.2091	0.0061	0.0032	0.0131
1.1999	0.2091	0.0057	0.0030	0.0109
1.5513	0.2091	0.0044	0.0023	0.0067
1.5768	0.2091	0.0043	0.0022	0.0065
1.6995	0.2091	0.0038	0.0020	0.0054

n=5000

a	h	d_1	d_2	d_3
0.6352	0.1820	0.0053	0.0029	0.0178
0.7910	0.1820	0.0052	0.0028	0.0155
0.8082	0.1820	0.0052	0.0028	0.0153
0.8119	0.1820	0.0052	0.0028	0.0152
0.8683	0.1820	0.0051	0.0028	0.0144
1.0525	0.1820	0.0048	0.0025	0.0117
1.1999	0.1820	0.0045	0.0023	0.0098
1.5513	0.1820	0.0034	0.0018	0.0060
1.5768	0.1820	0.0034	0.0017	0.0058
1.6995	0.1820	0.0030	0.0015	0.0048

n=10000

a	h	d_1	d_2	d_3
0.6352	0.1584	0.0042	0.0023	0.0154
0.7910	0.1584	0.0041	0.0022	0.0134
0.8082	0.1584	0.0041	0.0022	0.0132
0.8119	0.1584	0.0041	0.0022	0.0132
0.8683	0.1584	0.0041	0.0022	0.0124
1.0525	0.1584	0.0039	0.0020	0.0101
1.1999	0.1584	0.0036	0.0019	0.0084
1.5513	0.1584	0.0028	0.0014	0.0052
1.5768	0.1584	0.0027	0.0014	0.0050
1.6995	0.1584	0.0024	0.0012	0.0041

A.1.3. Distribución t

Trabajaremos con 10000 observaciones.

gl=10

a	h	d_1	d_2	d_3
0.6352	0.2	0.0034	0.0014	0.0019
0.7910	0.2	0.0036	0.0014	0.0017
0.8082	0.2	0.0037	0.0014	0.0017
0.8119	0.2	0.0037	0.0014	0.0017
0.8683	0.2	0.0038	0.0014	0.0016
1.0525	0.2	0.0041	0.0013	0.0013
1.1999	0.2	0.0044	0.0013	0.0011
1.5513	0.2	0.0053	0.0012	0.0015
1.5768	0.2	0.0054	0.0012	0.0016
1.6995	0.2	0.0058	0.0011	0.0020

A.1.4. Distribución Beta

Trabajaremos con 2500 observaciones

B(0.27,0.44)

a	h	d_1	d_2	d_3
0.6352	0.2091	0.0066	0.0039	0.0094
0.7910	0.2091	0.0062	0.0036	0.0100
0.8082	0.2091	0.0062	0.0036	0.0100
0.8119	0.2091	0.0062	0.0036	0.0100
0.8683	0.2091	0.0060	0.0035	0.0102
1.0525	0.2091	0.0054	0.0031	0.0102
1.1999	0.2091	0.0050	0.0028	0.0099
1.5513	0.2091	0.0037	0.0020	0.0081
1.5768	0.2091	0.0037	0.0020	0.0080
1.6995	0.2091	0.0032	0.0017	0.0072

B(0.86,0.57)

a	h	d_1	d_2	d_3
0.6352	0.2091	0.0048	0.0028	0.0175
0.7910	0.2091	0.0048	0.0027	0.0153
0.8082	0.2091	0.0048	0.0027	0.0151
0.8119	0.2091	0.0048	0.0027	0.0150
0.8683	0.2091	0.0047	0.0026	0.0142
1.0525	0.2091	0.0045	0.0025	0.0116
1.1999	0.2091	0.0042	0.0023	0.0096
1.5513	0.2091	0.0033	0.0018	0.0061
1.5768	0.2091	0.0032	0.0017	0.0060
1.6995	0.2091	0.0029	0.0015	0.0054

B(1,1)

a	h	d_1	d_2	d_3
0.6352	0.2091	0.0064	0.0036	0.0203
0.7910	0.2091	0.0062	0.0034	0.0178
0.8082	0.2091	0.0062	0.0034	0.0175
0.8119	0.2091	0.0062	0.0034	0.0174
0.8683	0.2091	0.0061	0.0033	0.0165
1.0525	0.2091	0.0057	0.0030	0.0134
1.1999	0.2091	0.0052	0.0028	0.0112
1.5513	0.2091	0.0040	0.0021	0.0069
1.5768	0.2091	0.0039	0.0020	0.0066
1.6995	0.2091	0.0034	0.0018	0.0055

B(1.45,1.81)

a	h	d_1	d_2	d_3
0.6352	0.2091	0.0116	0.0064	0.0257
0.7910	0.2091	0.0115	0.0062	0.0225
0.8082	0.2091	0.0115	0.0062	0.0221
0.8119	0.2091	0.0115	0.0062	0.0220
0.8683	0.2091	0.0113	0.0061	0.0208
1.0525	0.2091	0.0107	0.0056	0.0170
1.1999	0.2091	0.0099	0.0052	0.0142
1.5513	0.2091	0.0075	0.0039	0.0088
1.5768	0.2091	0.0074	0.0038	0.0084
1.6995	0.2091	0.0065	0.0033	0.0070

B(2.3,2.65)

a	h	d_1	d_2	d_3
0.6352	0.2091	0.0144	0.0081	0.0316
0.7910	0.2091	0.0140	0.0078	0.0277
0.8082	0.2091	0.0139	0.0077	0.0273
0.8119	0.2091	0.0139	0.0077	0.0272
0.8683	0.2091	0.0136	0.0075	0.0257
1.0525	0.2091	0.0126	0.0069	0.0211
1.1999	0.2091	0.0116	0.0063	0.0177
1.5513	0.2091	0.0086	0.0046	0.0110
1.5768	0.2091	0.0084	0.0045	0.0106
1.6995	0.2091	0.0074	0.0039	0.0088

B(2.5,2.5)

a	h	d_1	d_2	d_3
0.6352	0.2091	0.0147	0.0083	0.0317
0.7910	0.2091	0.0143	0.0080	0.0278
0.8082	0.2091	0.0142	0.0079	0.0274
0.8119	0.2091	0.0142	0.0079	0.0273
0.8683	0.2091	0.0140	0.0077	0.0258
1.0525	0.2091	0.0129	0.0071	0.0212
1.1999	0.2091	0.0119	0.0064	0.0177
1.5513	0.2091	0.0089	0.0047	0.0110
1.5768	0.2091	0.0086	0.0046	0.0106
1.6995	0.2091	0.0076	0.0040	0.0088

B(5,5)

a	h	d_1	d_2	d_3
0.6352	0.2091	0.0205	0.0118	0.0420
0.7910	0.2091	0.0197	0.0113	0.0373
0.8082	0.2091	0.0196	0.0112	0.0368
0.8119	0.2091	0.0196	0.0112	0.0367
0.8683	0.2091	0.0192	0.0109	0.0348
1.0525	0.2091	0.0175	0.0098	0.0289
1.1999	0.2091	0.0158	0.0088	0.0243
1.5513	0.2091	0.0115	0.0063	0.0153
1.5768	0.2091	0.0111	0.0061	0.0147
1.6995	0.2091	0.0097	0.0053	0.0123

A.1.5. Distribución Gamma

Trabajaremos con 2500 observaciones

G(.8,1,25)

a	h	d_1	d_2	d_3
0.6352	0.1584	0.0034	0.0019	0.0131
0.7910	0.1584	0.0032	0.0017	0.0114
0.8082	0.1584	0.0032	0.0017	0.0112
0.8119	0.1584	0.0032	0.0017	0.0112
0.8683	0.1584	0.0031	0.0016	0.0105
1.0525	0.1584	0.0029	0.0014	0.0086
1.1999	0.1584	0.0028	0.0012	0.0071
1.5513	0.1584	0.0027	0.0009	0.0044
1.5768	0.1584	0.0027	0.0009	0.0042
1.6995	0.1584	0.0027	0.0008	0.0035

A.2. Modelo de dos parámetros

N(0,1)

a	b	h	d_1	d_2	d_3
0.8556	-0.632	0.1820	0.0036	0.0012	0.0041
1.1195	-0.298	0.1820	0.0030	0.0015	0.0034
1.6967	1.1229	0.1820	0.0057	0.0021	0.0019
1.4050	1.9946	0.1820	0.0075	0.0014	0.0016
1.1392	0.5502	0.1820	0.0035	0.0016	0.0024
0.9408	0.2258	0.1820	0.0029	0.0013	0.0027
1.3492	1.9498	0.1820	0.0072	0.0014	0.0016
1.4055	-1.599	0.1820	0.0066	0.0016	0.0097
0.5429	1.5994	0.1820	0.0039	0.0009	0.0017
0.8358	0.9438	0.1820	0.0038	0.0013	0.0015

Unif[0,1]

a	b	h	d_1	d_2	d_3
0.8556	-0.632	0.1820	0.0031	0.0019	0.0154
1.1195	-0.298	0.1820	0.0045	0.0027	0.0229
1.6967	1.1229	0.1820	0.0068	0.0034	0.0063
1.4050	1.9946	0.1820	0.0016	0.0008	0.0011
1.1392	0.5502	0.1820	0.0059	0.0033	0.0205
0.9408	0.2258	0.1820	0.0050	0.0029	0.0209
1.3492	1.9498	0.1820	0.0018	0.0009	0.0015
1.4055	-1.599	0.1820	0.0004	0.0002	0.0024
0.5429	1.5994	0.1820	0.0025	0.0014	0.0076
0.8358	0.9438	0.1820	0.0043	0.0024	0.0129

Beta(0.27,0.44)

a	b	h	d_1	d_2	d_3
0.8556	-0.632	0.2091	0.0048	0.0030	0.0067
1.1195	-0.298	0.2091	0.0072	0.0045	0.0100
1.6967	1.1229	0.2091	0.0071	0.0038	0.0169
1.4050	1.9946	0.2091	0.0017	0.0009	0.0042
1.1392	0.5502	0.2091	0.0075	0.0044	0.0098
0.9408	0.2258	0.2091	0.0070	0.0043	0.0091
1.3492	1.9498	0.2091	0.0020	0.0010	0.0048
1.4055	-1.599	0.2091	0.0007	0.0004	0.0010
0.5429	1.5994	0.2091	0.0031	0.0018	0.0051
0.8358	0.9438	0.2091	0.0052	0.0030	0.0085

Beta(2.3,.57)

a	b	h	d_1	d_2	d_3
0.8556	-0.632	0.2091	0.0074	0.0046	0.0264
1.1195	-0.298	0.2091	0.0109	0.0068	0.0394
1.6967	1.1229	0.2091	0.0113	0.0064	0.0176
1.4050	1.9946	0.2091	0.0024	0.0013	0.0043
1.1392	0.5502	0.2091	0.0142	0.0087	0.0406
0.9408	0.2258	0.2091	0.0123	0.0075	0.0390
1.3492	1.9498	0.2091	0.0028	0.0015	0.0049
1.4055	-1.599	0.2091	0.0008	0.0005	0.0035
0.5429	1.5994	0.2091	0.0055	0.0033	0.0148
0.8358	0.9438	0.2091	0.0096	0.0058	0.0255

Beta(5,5)

a	b	h	d_1	d_2	d_3
0.8556	-0.632	0.2091	0.0108	0.0065	0.0316
1.1195	-0.298	0.2091	0.0157	0.0095	0.0471
1.6967	1.1229	0.2091	0.0213	0.0115	0.0179
1.4050	1.9946	0.2091	0.0044	0.0023	0.0033
1.1392	0.5502	0.2091	0.0232	0.0135	0.0488
0.9408	0.2258	0.2091	0.0190	0.0112	0.0467
1.3492	1.9498	0.2091	0.0051	0.0027	0.0041
1.4055	-1.599	0.2091	0.0011	0.0007	0.0042
0.5429	1.5994	0.2091	0.0092	0.0053	0.0177
0.8358	0.9438	0.2091	0.0162	0.0092	0.0306

A.3. Modelo de tres parámetros

N(0,1)

a	b	c	d_1	d_2	d_3
0.6929	0.4904	0.4332	0.0018	0.0006	0.0010
1.9088	-0.840	0.0372	0.0046	0.0020	0.0071
1.4152	1.2086	0.1515	0.0045	0.0015	0.0012
1.6195	1.8686	0.2261	0.0059	0.0012	0.0011
0.9361	0.6804	0.2072	0.0027	0.0011	0.0016
1.3201	0.3602	0.0146	0.0033	0.0017	0.0032
1.8442	-1.975	0.4468	0.0049	0.0008	0.0069
1.7722	-1.615	0.4256	0.0041	0.0011	0.0070
0.6235	0.3824	0.2433	0.0024	0.0008	0.0018
0.5401	-1.035	0.1712	0.0050	0.0007	0.0083

Unif[0,1]

a	b	c	d_1	d_2	d_3
0.6929	0.4904	0.4332	0.0022	0.0012	0.0083
1.9088	-0.840	0.0372	0.0014	0.0009	0.0085
1.4152	1.2086	0.1515	0.0047	0.0024	0.0058
1.6195	1.8686	0.2261	0.0013	0.0007	0.0007
0.9361	0.6804	0.2072	0.0040	0.0022	0.0131
1.3201	0.3602	0.0146	0.0064	0.0037	0.0261
0.6235	0.3824	0.2433	0.0027	0.0015	0.0103
0.5401	-1.035	0.1712	0.0017	0.0010	0.0080

Beta(0.27,0.44)

a	b	c	d_1	d_2	d_3
0.6929	0.4904	0.4332	0.0029	0.0017	0.0037
1.9088	-0.840	0.0372	0.0025	0.0016	0.0037
1.4152	1.2086	0.1515	0.0050	0.0027	0.0115
1.6195	1.8686	0.2261	0.0015	0.0008	0.0037
0.9361	0.6804	0.2072	0.0049	0.0029	0.0072
1.3201	0.3602	0.0146	0.0088	0.0054	0.0113
1.7722	-1.615	0.4256	0.0001	0.0001	0.0002
0.6235	0.3824	0.2433	0.0036	0.0022	0.0045
0.5401	-1.035	0.1712	0.0026	0.0016	0.0035

Beta(2.3,.57)

a	b	c	d_1	d_2	d_3
0.6929	0.4904	0.4332	0.0052	0.0032	0.0156
1.9088	-0.840	0.0372	0.0027	0.0017	0.0118
1.4152	1.2086	0.1515	0.0082	0.0047	0.0132
1.6195	1.8686	0.2261	0.0020	0.0011	0.0037
0.9361	0.6804	0.2072	0.0092	0.0056	0.0258
1.3201	0.3602	0.0146	0.0165	0.0101	0.0509
1.7722	-1.615	0.4256	0.0002	0.0001	0.0008
0.6235	0.3824	0.2433	0.0063	0.0038	0.0193
0.5401	-1.035	0.1712	0.0042	0.0025	0.0141

Beta(5,5)

a	b	c	d_1	d_2	d_3
0.6929	0.4904	0.4332	0.0084	0.0049	0.0188
1.9088	-0.840	0.0372	0.0036	0.0023	0.0140
1.4152	1.2086	0.1515	0.0152	0.0083	0.0159
1.6195	1.8686	0.2261	0.0035	0.0019	0.0021
0.9361	0.6804	0.2072	0.0152	0.0088	0.0310
1.3201	0.3602	0.0146	0.0257	0.0152	0.0610
1.7722	-1.615	0.4256	0.0002	0.0001	0.0010
0.6235	0.3824	0.2433	0.0101	0.0059	0.0231
0.5401	-1.035	0.1712	0.0063	0.0037	0.0169

A.4. Modelo logit cúbico

Consideraremos 10 modelos con valores de a entre -1 y 1 y valores de b entre 1 y 2.5 con distribución uniforme y c 0.75.

Para theta consideraremos 5 juegos de datos, 5000 datos con distribución normal 0, 1; 5000 con distribución uniforme 0,1, 2500 con distribución beta (0.27,0.44); (2.3,.57); (5,5).

N(0,1)

a	b	d_1	d_2	d_3
0.5196	1.2753	0.0035	0.0016	0.0051
0.7026	1.0045	0.0041	0.0017	0.0068
-0.048	2.3643	0.0034	0.0019	0.0044
0.5589	2.3623	0.0035	0.0019	0.0049
0.0431	1.7043	0.0032	0.0016	0.0036
-0.244	2.0731	0.0034	0.0018	0.0036
-0.164	1.5051	0.0033	0.0016	0.0031
-0.525	2.2252	0.0036	0.0019	0.0036
0.0750	1.7002	0.0032	0.0016	0.0036
0.9629	1.9307	0.0037	0.0018	0.0053

Unif[0,1]

a	b	d_1	d_2	d_3
0.5196	1.2753	0.0040	0.0023	0.0158
0.7026	1.0045	0.0034	0.0019	0.0118
-0.048	2.3643	0.0065	0.0040	0.0316
0.5589	2.3623	0.0054	0.0034	0.0284
0.0431	1.7043	0.0054	0.0032	0.0228
-0.244	2.0731	0.0064	0.0037	0.0276
-0.164	1.5051	0.0057	0.0032	0.0201
-0.525	2.2252	0.0069	0.0040	0.0285
0.0750	1.7002	0.0054	0.0031	0.0227
0.9629	1.9307	0.0039	0.0024	0.0199

Beta(0.27,0.44)

a	b	d_1	d_2	d_3
0.5196	1.2753	0.0053	0.0032	0.0068
0.7026	1.0045	0.0043	0.0025	0.0059
-0.048	2.3643	0.0100	0.0062	0.0137
0.5589	2.3623	0.0088	0.0055	0.0124
0.0431	1.7043	0.0075	0.0046	0.0099
-0.244	2.0731	0.0090	0.0055	0.0120
-0.164	1.5051	0.0072	0.0043	0.0093
-0.525	2.2252	0.0094	0.0058	0.0123
0.0750	1.7002	0.0074	0.0046	0.0098
0.9629	1.9307	0.0062	0.0038	0.0087

Beta(2.3,.57)

a	b	d_1	d_2	d_3
0.5196	1.2753	0.0096	0.0058	0.0287
0.7026	1.0045	0.0077	0.0047	0.0217
-0.048	2.3643	0.0171	0.0106	0.0578
0.5589	2.3623	0.0134	0.0083	0.0487
0.0431	1.7043	0.0136	0.0083	0.0421
-0.244	2.0731	0.0165	0.0102	0.0519
-0.164	1.5051	0.0134	0.0081	0.0380
-0.525	2.2252	0.0181	0.0111	0.0550
0.0750	1.7002	0.0134	0.0082	0.0418
0.9629	1.9307	0.0094	0.0059	0.0339

Beta(5,5)

a	b	d_1	d_2	d_3
0.5196	1.2753	0.0155	0.0090	0.0344
0.7026	1.0045	0.0131	0.0075	0.0260
-0.048	2.3643	0.0252	0.0152	0.0692
0.5589	2.3623	0.0190	0.0116	0.0582
0.0431	1.7043	0.0215	0.0126	0.0504
-0.244	2.0731	0.0257	0.0152	0.0622
-0.164	1.5051	0.0224	0.0129	0.0456
-0.525	2.2252	0.0286	0.0168	0.0660
0.0750	1.7002	0.0212	0.0125	0.0500
0.9629	1.9307	0.0137	0.0083	0.0405

Bibliografía

[1] Ackerman, T. A. (1992): A didactic explanation of item bias, item impact, and item validity from multidimensional perspective, Journal of Educational Measurement, 29(1), 67-91.

[2] Akaike, H. (1956): On a 0-1 process and some of his applications, Annals of the Institute of Statistical Mathematics, 8, 87-94.

[3] Andersen, E. B. (1973): Conditional inference for multiple-choice questionnaires, British Journal of Mathematical and Statistical Psychology, 26, 31-44.

[4] Andersen, E. B. (1977): Sufficient statistics and latent trait models, Psychometrika, 42, 69-81.

[5] Baker, F. (1992): Item Response Theory Parameter Estimation Techniques, New York, Marcel Dekker, Inc.

[6] Barton, M.A. y Lord, F. M. (1981): An upper asymptote for the three parameter logistic item response model, Research Bulletin, Princeton, Nueva Jersey, Educational Testing Service.

[7] Bartholomew, D.J. (1987): Latent Variable Models and Factor Analysis, London, Charles Griffin and Company.

[8] Berkson, J. (1955): Maximum likelihood and minimum chi square estimates of the logistic function, Journal of the American Statistical Association, 50, 120-162.

[9] Billingsley, P. (1968): Convergence of Probability Measures, New York, Wiley.

[10] Birnbaum, A. (1968): Some latent trait models and their use in inferring an examinee's ability. En F. M. Lord y M.R. Novick (eds.), Statistical Theories of Mental Test Scores, pp.397-472, Reading, MA, Addinson-Wesley.

[11] Bock, R. D. y Aitkin, M. (1981): Marginal maximum likelihood estimation of item parameters. An application of an EM algorithm. Psychometrika, 46, 443-459.

[12] Bock, R. D. y Lieberman, M. (1970): Fitting a response model for n dichotomously scored items. Psychometrika, 35, 179-197.

[13] Bock, R. D. (1972): Estimating multinomial response relations. En R. C. Bose, I. M. Chakeavarti, P. C. Makalanobis, C. R. Rao, y K.J.C. Smith (eds.), Essays in Probability and Statistics, Chapel Hill, NC, The University of North Carolina Press.

[14] Boucheron, S., Lugosi, G., Massart, P.(2003): Concentration inequalities using the entropy method. Annals of Probability 31, 1583-1614.

[15] Dempster, A. P., Laird, N. M. y Rubin, D. B. (1977): Maximum likelihood from incomplete data via the EM algorithm (with discussion), Journal of the royal Statistical Society, Series B, 39, 1-38.

[16] Dette, H, Neumeyer, N. y Pilz, K. (2006): A simple nonparametric estimator of a strictly monotone regression function. Bernoulli, 12(3), 469-490.

[17] Devorye, L. y Gyorfi, L. (1985): Nonparametric Density Estimation, the L1 view, New York, Wiley.

[18] Dorans, N. J. y Holland, W. P. (1993): DIF detection and description: Mantel-Haenszel and standardization. En W. P. Holland y H. Wainer (eds.), Differential Item Functioning (pp. 35-66), Hillsdale, NJ, LEA.

[19] Douglas, J. (1997): Joint Consistency of Nonparametric Item Characteristic Curve and Ability Estimation. Psychometrika, 62, 7-28.

[20] Douglas, J. y Cohen, A (2001): Nonparametric item response function estimation for assessing parametric model fit. Applied Psychological Measurement, 25, 234-243.

[21] Douglas, J., Kim, H.R., Habing, B. y Gao, F. (1998): Investigating local dependence with conditional covariance functions. Journal of Educational and Behavioral Statistics, 23, 129-151.

[22] Drasgow, F., Levine, M.V., Williams, B., McLaughlin, M.E., y Candell, G.L. (1989): Modeling incorrect responses to multiple-choice item with multilinear formula score theory. Applied Psychological Measurement, 13, 285-299.

[23] Ellis, J. L. y Junker, B. w. (1997): Tail-measurability in monotone latent variable models, Psychometrika, 62, 495-523.

[24] Elliot, C.D., Murray, D.J. y Saunders, R. (1977): Goodness of fit to the Rasch as a criterion of test unidimensionality. Manchester: University of Manchester.

[25] Feller, W.(1978): An introduction to probability theory and its applications, New York, Wiley.

[26] Fisher, E.(1992): A Skorohod representation and an invariance principle for sums of weighted iid random variables, Journal of Mathematics, 22, 169-179.

[27] Fidalgo, A. M. (1996): Funcionamiento diferencial de los ítems. Procedimiento Mantel-Haenzel y modelos loglineales. Tesis doctoral no publicada. Universidad de Oviedo.

[28] _____ (1994): MHDIF: A computer program for detecting uniform and nonuniform differential item functioning with Mantel- Haenszel procedure, Applied Pyschological Measurement, 18(3), 300.

[29] Fischer, G. H. y Scheiblechner, H. H. (1970): Algorithmen und programmen für das probabilitische testmodel von Rasch, Psychologische Beitrage 12, 23-51.

[30] Garwood, F. (1941): The application of maximum likelihood to dosage-mortality curves, Biometrika, 32, 46-58.

[31] Grayson, D.A. (1988): Two-group classification in latent trait theory: Scores with monotone likelihood ratio. Psychometrika 53, 383-392.

[32] Grebliki, W., Krzyzak, A., Pawlak, M. (1984): Distribution-free pointwise consisitency of kernel regression estiamte.The Annals of Statistics, 12, 1570-1575.

[33] Haley, D.C. (1952):Estimation of the dosage mortality relationshop when the dose is subject to error, (Technical Report N° 15), Stanford, CA, Stanford University, Applied Mathematics and Statistics Laboratory.

[34] Hardle,W. (1990): Applied nonparametric regression, Cambridge, Cambridge University Press.

[35] Hambleton, R.; Swaminathan, H. y Rogers, H. (1991): Fundamentals of Item Response Theory,Thousand Oaks, Sage Publications.

[36] Holland, P.W. (1990): On the sampling theory foundations of item response theory models, Psychometrika, 55, 577-601.

[37] Holland, P. W. y Rosenbaum, P.R. (1986): Conditional association and unidimensionality in monotone latent variable models, The Annals of Statistics, 14, 1523-1543.

[38] Holland, P.W. y Thayer, D.T.(1988): Diferential item performance and the Mantel-Haenszel procedure. En H. Wainer y H.I. Braun (Eds). Test validity,(pp.129-145). Hillsdale, NJ: LEA.

[39] Kim, S. H. y Cohen, A. S. (1992): A comparison of two areas measures for detecting differential item functioning across manifest and latent examinee groups, Journal of Educational Measurement, 15(3), 269-278.

[40] Junker, B. W. (1993): Conditional association, essential independence and monotone unidimensional item response models, The Annals of Statistics, 21, 1359-1378.

[41] _____ (2001): On the interplay between nonparametric and parametric IRT, with some thoughts about the future. En A. Boomsma, M.A., J. van Duijn

y T. A. B. Snijders (Eds.), Essays on item response theory (pp.247-276), Nueva York, Springer.

[42] Junker, B. W. y Sijtsma, K. (2001): Nonparametric item response theory in action: An overview of the special issue, Applied Psychological Measurement, 25, 211-220.

[43] Kale, B. K. (1962): On the solution of likelihood equations by iteration processes. The multiparameter case, Biometrika, 49, 479-486.

[44] Levine, M.V. (1984): An introduction to multilinear formula score theory, (Measurement Series 84-5) Champaign, Il. University of Illinois. Departament of Educational Psychology. Model-Based Measurement Laboratory.

[45] ---------- (1985): The trait in latent trait theory. In D. J. Weiss (Ed) Proceedings of the 1982 item response theory/computarized adaptative testing conference. Minneapolis, MN: University of Minnesota. Departament of Psychology. Computarized adaptative testing Laboratory.

[46] Lewis, C. (1983): Bayesian inference for latent abilities. En S. B. Anderson y J. S. Helmick (Eds.), On Educational Testing (pp.224-251), San Francisco, Jossey-Bass.

[47] Loevinger, J. (1947): A systematic approach to the construction and evaluation of tests of ability, Psychological Monographs, 61, N° 4.

[48] ---------- (1948): The technique of homogeneous tests compared with some aspects of 'scale analysisánd factor analysis, Psychological Bulletin, 45, 507-530.

[49] Lord, F. M. (1952): A theory of test scores, Psychometric Monographs, 7.

[50] ---------- (1980): Applications of item response theory to practical testing problems, Hillsdale, LEA.

[51] Lord, F. M. y Novick, M.R. (1968): Statistical theories of mental tests scores, Reading, MA, Addison-Wesley.

[52] Ludlow, L. H. (1986): Graphical analysis of item response theory residuals, Applied Psychological Measurement, 10, 217-229.

[53] Lumsden, J. (1961): The construction of unidimensional tests, Psychological Bulletin, 58, 122-131.

[54] Massart, P. (2000): About the constants in Talagrand's concentration inequalities for empirical processes, Annals of Probability, 28, 863-884.

[55] Maxwell, A. E. (1959): Maximum likelihood estimates of item parameters using the logistic function, Psychometrika, 24, 221-227.

[56] Mazor, K. M., Clauser, B. E. y Hambleton, R. K. (1994): Identification for nonuniform differential item functioning using a variation of the Mantel- Haenszel procedure, Educational and Psychological Measurement, 54(2), 284-291.

[57] McDiarmid, C. (1989): On the method of bounded differences. En: Surveys in Combinatorics 1989. Cambridge University Press,Cambridge, 148-188.

[58] McKinley, R.L y Mills, C.N. (1985): A comparison of several goddness-of-fit statistics. Applied Psychological Measurement 9 (1), 49-57.

[59] McDonald, R. P. (1981): The dimensionality of tests and items, British Journal of Mathematical and Statistical Psychology, 34, 100-117.

[60] Millsap, R. E. y Everson, H. T. (1993): Methodology review: statistical approaches for assessing measurement bias, Applied Psychological Measurement, 16, 389-402.

[61] Mislevy, R.J. y Bock, R.D. (1982): Bilog Item analysis and test scoring with binary logistic models. Mooresville, ID. Scientific Software.

[62] Mislevy, R.J. (1986a): Recent developments in the factor analysis of categorical variables. Journal of Educational Statistics, 11, 3-31.

[63] Mislevy, R.J. (1986b): Bayes modal estimation in item response models Psychometrika, 51, 177-195.

[64] Mokken, R. J. (1971): A theory and procedure of scale analysis, The Hague, Mouton/Berlin, De Gruyter.

[65] _____ (1997): Nonparametric models for dichotomous responses. En W. J. van der Linden y R. K. Hambleton (eds.) Handbook of modern item response theory (pp. 351-367), Nueva York, Springer.

[66] Mokken, R. J. y Lewis, C. (1982): A nonparametric approach to the analysis of dichotomous item responses. Applied Psychological Measurement, 6, 417-430.

[67] Molenaar, I.W.(1997): Nonparametric models for polytomous responses. En W. J. van der Linden y R. K. Hambleton (Eds.) Handbook of modern item response theory (pp. 369-380), Nueva York, Springer.

[68] Molenaar, I.W. y Sijtsma, K.(2000): User's manual MSP5 for Windows, Groningen, iecProGAMMA.

[69] Muñiz, J. (1997): Introducción a la teoría de respuesta a los ítems, Madrid, Pirámide.

[70] Nadaraya, E. A. (1964): On estimating regression. Theory of Probability and its Applications 10, 186-190.

[71] Neyman, J y Scott, E.L. (1948): Consistent estimates based on partially consistent observations.Econometrika 16 (1), 1-32.

[72] Park, D. G. (1988): Investigations of item response theory item bias detection, Universidad de Georgia. Tesis doctoral no publicada.

[73] Parzen, E. (1962): On estimation of a probability density and mode, Annals of Mathematical Statistics, 35, 1065-1076.

[74] Raju, N. S. (1990): Determining the significance of estimated signed and unsigned areas between two item response functions, Applied Psychological Measurement, 14(2), 197-207.

[75] Ramsay, J. O.(1991): Kernel smoothing approaches to nonparametric item characteristic curve estimation. Psychometrika, 56, 611-630.

[76] _____ (1994): TESTGRAF: A program for the graphical analysis of multiple choice test and quetionnaire data. User´s guide.

[77] Ramsay, J. O., y Abrahamowicz, M. (1989): Binomial regression with monotone splines. A psychometric application. Journal of the American Statistical Association, 84, 906-915.

[78] Ramsay, J. O. y Winsberg, S. (1991): Maximum marginal likelihood estimation for semiparametric item analysis, Psychometrika, 56, 365-379.

[79] Rasch, G. (1960): Probabilistic models for some intelligence and attainment tests, Copenhague, The Danish Institute for Educational Research.

[80] Reise, S. P. y Waller, N.G.(1990): Fitting the two-parameter model to personality data, Applied Psychological Measurement, 14, 45-58.

[81] Richardson, M. W. (1936): The relationship between difficulty and the differential validity of a test, Psychometrika 1, 33-49.

[82] Rosenbaum, P. (1984): Testing the conditional independence and monotonicity assumptions of item response theory, Psychometrika, 49, 425-435.

[83] Rosenbaum, P. (1987): Comparing item characteristic curves, Psychometrika, 52(2), 217-233.

[84] Rosenblatt, M. (1956): Remarks on some nonparametric estimates of a density function, Annals of Mathematical Statistics, 27, 642-669.

[85] Rost, J. y von Davier, M. (1994): A conditional item-fit index for Rasch models, Applied Psychological Measurement, 18, 171-182.

[86] Santisteban,C. y Alvarado, J. (2001): Modelos Psicométricos. Madrid, UNED.

[87] Schriever, B. F. (1985): Order Dependence. Tesis doctoral no publicada. Free University, Amsterdam.

[88] Segall, D.O. (1983): Test characteristic curves item bias, and transformation to a common metric in ítem response theory: A metodological artifact with serious consequences and a simple solution. University of Illinois, Department of Psychology, Urbana-Champaign. Manuscrito no publicado.

[89] Serfling, R. J. (1980): Approximations Theorems of matematical statistics, New York, Wiley.

[90] Shealy, R. y Stout, W. (1993a): A model-based standarization approach that separates true bias/DIF from group ability differences and detects test bias/DTF as well as item bias/DIF Psychometrika , 58, 159-194.

[91] _____ (1993b): A model-based standarization approach that separates true bias/DIF from group ability differences and detects test bias/DTF as well as item bias/DIF, Psychometrika , 58, 159-194.

[92] Shealy, R., Stout, W. y Roussos, L.(1995): SIBTEST (versión 1.3), Urbana-Champaign, University of Illinois, Department of Applied Statistics.

[93] Sijtsma, K. (1988): Constributions to Mokken's nonparametric item response theory. Amsterdam, Free University Press.

[94] _____ (1998): Methodology review: Nonparametric IRT approaches to the analysis of dichotomous item scores, Applied Psychological Measurement, 22, 3-32.

[95] _____ (2001): Developments in measurement of persons and items by means of item response models, Behaviormetrika, 28, 65-94.

[96] Sijtsma, K. y Molenaar, I. W. (2002): Introduction to nonparametric item response theory. Thousand Oaks, Sage Publications.

[97] Spray, J. A. y Carlson, J. E. (1986): Comparison of loglinear and logistic regression models for detecting changes in proportions. Comunicación presentada a la Annual Meeting of the American Educational Research Association, San Francisco.

[98] Stout, W. F. (1987): A nonparametric approach for assessing latent trait unidimensionality, Psychometrika, 52, 589-617.

[99] _____ (1990): A new item response theory modleing approach with applications to unidimensionality assessment and ability estimation, Psychometrika, 55, 293-325.

[100] Swaminathan, H. y Gifford, J. A. (1986): Bayesian estimation in the three-parameter logistic model, Psychometrika, 51, 589-601.

[101] Takane, Y. y Leeuw, J. (1987): On the relationship between item response theory and factor analysis of discretized variables, Psychometrika 52, 393-408.

[102] Tucker, L. R. (1946): Maximum validity of a test with equivalent items, Psychometrika 11, 1-13.

[103] Van der Linden, W. y Hambleton, R. (1997): Handbook of Modern Item Response Theory, New York, Springer-Verlag.

[104] Verhelst, N.D., Glas, C.A.W. y van der Sluis, A. (1984): Estimation problems in the Rasch model¿the basic symmetric functions, Computational Statistics Quarterly, 1, 245-262.

[105] Verhelst, N.D. y Molenaar, W. (1988): Logit-based paremeter estimation in Rasch model. Statistica Neerlandica, 42, 273-295.

[106] Wahaba, G (1990): Spline Models for observational data. Philadelphia. Society for Industrial and Applied Mathematics.

[107] Watson, G. S. (1964): Smooth regression analysis Sankyha, Series A,26,101-116.

[108] Wingersky, M.S. Patrick,R. y Lord, F.M. (1988): Logist users guide. Princeton, NJ. Educational Testing Service.

[109] Wollenberg, A. L. van den (1982): A simple and effective method to test the dimensionality axiom of the Rasch model, Applied Psychological Measurement, 6, 83-91.

[110] _____ (1988): Testing a latent trait model. En R. Langeheine y J. Rost (Ed.), Latent trait and latent class models, pp.31-50. Plenum Press: New York.

[111] Wood, R. L., Wingersky, M.S. y Lord, F. M. (1976): LOGIST: A computer program for estimating examinee ability and item characteristic curve parameters (RM-76-6), Princeton, NJ, Educational Testing Service.

[112] Wright, B. D. y Mead, R. J. (1978): BICAL: Calibrating items with the Rasch model (Research Memorandum No. 23A), Chicago, IL, University of Chicago, Statistical Laboratory.

[113] Wright, B.D. y Douglas, G. A.(1977a): Best procedures for sample-free item analysis. Applied Psychological Measurement, 1, 281-295.

[114] _____ (1977b): Conditional versus unconditional procedures for sample-free analysis.Educational and Psychological Measurement, 37, 573-586.

[115] Wright, B.D. y Panchapakesan, N. (1969): A procedure for sample-free item analysis, Educational and Psychological Measurement, 29, 23-48.

[116] Wright, B. D. y Stone, M. H. (1979): Best test design, Chicago, MESA Press University of Chicago.

[117] Yen, M. W. (1981): Using simulation results to choose a latent trait model, Applied Psychological Measurement, 5, 245-262.

Mario Luzardo es licenciado y magíster en Matemáticas por la Universidad de la República, y magíster en Metodologías de las Ciencias del Comportamiento y de la Salud por la Universidad Autónoma de Madrid. Es profesor adjunto del Instituto de Fundamentos de la Facultad de Psicología de la Universidad de la República, donde dicta cursos de grado y posgrado en el área de la estadística y la psicometría. Sus temas de interés son la estadística de datos funcionales, los modelos de rasgo latente y la teoría de respuesta al ítem.

Diego Forteza es Licenciado en Estadística y maestrando en Ingeniería Matemática en la Universidad de la República. Sus áreas de interés son la psicometría y la bioestadística.

ISBN: 978-9974-0-1065-9