

Comparación entre Cinco Métodos de Monte Carlo para Estimar la Confiabilidad Diámetro Acotada de Redes de Comunicaciones

Leslie Murray

Facultad de Ciencias Exactas, Ingeniería y Agrimensura

Universidad Nacional de Rosario

Rosario, Argentina

Resumen

El cálculo exacto de la confiabilidad de redes por medio de la mayoría de los criterios de operatividad es un problema NP-difícil. Como alternativa, los métodos de *Monte Carlo* son una práctica simple, comunmente aceptada y a veces muy efectiva para hacer estimaciones de confiabilidad. En este informe se revisan cinco métodos de *Monte Carlo*: *Monte Carlo Crudo*, *Dagger*, *Monte Carlo Permutación*, y *Entropía Cruzada* sobre *Crudo* y *Permutación*, y se los aplica en la determinación de la *Confiabilidad Diámetro Acotada*, que es una generalización del clásico problema de \mathcal{K} -terminales, como función de estructura. Dentro del contexto de la *Confiabilidad Diámetro Acotada*, los cinco métodos se implementan y se estudia su rendimiento para estimar y analizar la *anti-confiabilidad* de tres redes elementales. A partir de los resultados se derivan algunas conclusiones y consideraciones sobre la robustez y eficiencia de los distintos métodos.

Palabras Clave – Monte Carlo, Redes, Confiabilidad, Enlaces, Reducción de Varianza.

1 Introducción

Una Red de Comunicaciones se dice *confiable* si es capaz de cumplir con los propósitos para los cuales ha sido concebida. Para “medir” qué tan bien una red cumple con estos propósitos, es preciso definir criterios de operatividad y encontrar la forma de decir qué tan cerca o lejos está la red de ser *operativa* bajo esos criterios. En un sentido matemático, la confiabilidad se define como la probabilidad de que la red esté *operativa* (en perfecto estado de funcionamiento) bajo los criterios de operatividad propuestos, y la *anti-confiabilidad* como la probabilidad de que la red se encuentre *no-operativa* o *fallada* (totalmente fuera de servicio) bajo los mismos criterios.

Un modelo de red de comunicaciones, apropiado para estimar su confiabilidad mediante simulación es un grafo $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, en el que $\mathcal{V} = \{v_1, v_2, \dots, v_n\}$ es el conjunto de nodos y $\mathcal{E} = \{e_1, e_2, \dots, e_m\}$ el conjunto de enlaces. En este modelo tanto nodos como enlaces pueden encontrarse en uno de dos estados posibles: *operativos* o *fallados*. Ambos estados se identifican por el valor de sendas variables aleatorias binarias: X_{v_i} para el i -ésimo nodo y X_{e_i} para el i -ésimo enlace. Estas variables toman valor 1 cuando el componente asociado está *operativo* y 0 cuando el componente está *fallado*. Una simplificación consiste en aceptar que $X_{v_i} = 1$, $i = 1, 2, \dots, n$, es decir que todos los nodos están siempre *operativos*. Esta simplificación permite modificar la notación, llamando $X_i = X_{e_i}$, $i = 1, 2, \dots, m$, es decir descartando la indicación v o e en las variables.

Es natural entonces definir la confiabilidad r_i de cada enlace como la probabilidad de que el mismo se encuentre *operativo*: $r_i = \mathbb{P}[X_i = 1]$, $i = 1, 2, \dots, m$, y, recíprocamente, la anti-confiabilidad como la probabilidad de que el enlace se encuentre *fallado*: $q_i = 1 - r_i = \mathbb{P}[X_i = 0]$, $i = 1, 2, \dots, m$. El vector $\mathbf{X} = (X_1, X_2, \dots, X_m)$ define el estado global de la red desde el punto de vista de su capacidad operativa por cuanto indica el estado de todos los componentes en condiciones potenciales de fallar. Los valores de \mathbf{X} se ajustan a una función de masa de probabilidad $f_{\mathbf{X}}(\mathbf{x})$.

Para definir la confiabilidad global de la red se establece una cláusula o condición tal que cuando es satisfecha implica para la red un estado de operatividad aceptable, mientras que cuando no lo es, define o representa una capacidad de operación no aceptable para la red. Una posible condición es que haya conectividad entre todos los nodos de un determinado subconjunto $\mathcal{K} \subseteq \mathcal{V}$. Otra condición es que haya conectividad entre dos nodos designados, v_s y v_t , es decir $\mathcal{K} = \{v_s, v_t\}$. La determinación de confiabilidad por medio de estas cláusulas se conoce como problema de \mathcal{K} -terminales. El último caso ($|\mathcal{K}| = 2$) se conoce también como problema fuente-terminal. Ninguno de estos dos enfoques toma en consideración la longitud del camino entre pares de nodos conectados. La *Confiabilidad Diámetro Acotada* [1, 13] es una generalización del concepto de \mathcal{K} -terminales según la cual la condición de operatividad se considera satisfecha si todos los nodos de \mathcal{K} (ya sea $|\mathcal{K}| = 2$ o $|\mathcal{K}| > 2$), están conectados por un camino de longitud D o menor.

En cualquier caso el cumplimiento de la condición depende del valor del vector \mathbf{X} por lo que en términos formales la confiabilidad es una función $\Phi(\mathbf{X})$, llamada función de estructura. Esta función se define de manera que para aquellos valores de \mathbf{X} para los que la condición es satisfecha, $\Phi(\mathbf{X}) = 1$ y para los demás valores, $\Phi(\mathbf{X}) = 0$. Vista como variable aleatoria la función de estructura tendrá un valor medio o esperado $0 \leq \mathbb{E}\{\Phi(\mathbf{X})\} \leq 1$, tanto más cercano a 1 cuanto mayor sea la cardinalidad del conjunto de valores de \mathbf{X} para los que $\Phi(\mathbf{X}) = 1$, es decir cuanto mayor sea el número de estados de la red que satisfacen la condición impuesta para definir su capacidad operativa como aceptable. El valor medio o esperado de $\Phi(\mathbf{X})$ es entonces un indicador de la confiabilidad R de la red mientras que el valor medio de $(1 - \Phi(\mathbf{X}))$ es un indicador de su anti-confiabilidad Q :

$$R = \mathbb{E}\{\Phi(\mathbf{X})\} = \sum_{i=1}^{2^m} \mathbb{P}[\mathbf{X}^i] \Phi(\mathbf{X}^i) \quad Q = \mathbb{E}\{1 - \Phi(\mathbf{X})\} = \sum_{i=1}^{2^m} \mathbb{P}[\mathbf{X}^i] (1 - \Phi(\mathbf{X}^i)) \quad (1)$$

siendo \mathbf{X}^i cada uno de los valores posibles de \mathbf{X} y $\mathbb{P}[\mathbf{X}^i]$ su correspondiente valor de probabilidad.

El cálculo exacto de (1) (en cualquiera de sus dos formas) define un problema NP-difícil por lo que, a medida que m crece, se convierte en intratable desde el punto de vista computacional. Una forma de proceder es descartar el cálculo exacto y realizar, en su lugar, una estimación. Los métodos de *Monte Carlo* [7] son una práctica simple, comúnmente aceptada y a veces muy efectiva para hacer estas estimaciones.

En la Sección 2 se presenta el concepto de *Confiabilidad Diámetro Acotada* [1, 2]. En las secciones siguientes se revisan cinco métodos de *Monte Carlo* propuestos en la literatura para estimar medidas clásicas de confiabilidad, generalmente definidas mediante el problema de \mathcal{K} -terminales: *Monte Carlo Crudo* [7], *Dagger* [5, 11], *Monte Carlo Permutación* [6, 9, 10], y *Entropía Cruzada* sobre *Crudo* y *Permutación* [3, 4, 9, 10]. Estos métodos son adaptados sin dificultad a la estimación de *Confiabilidad Diámetro Acotada*. La Sección 7 presenta resultados experimentales obtenidos de la aplicación de los cinco métodos a la estimación de la anti-confiabilidad de tres redes elementales de comunicaciones (según el concepto de *Confiabilidad Diámetro Acotada*). Al final del informe se presentan las conclusiones y se indican algunas líneas de trabajo futuro.

2 Confiabilidad Diámetro Acotada

De acuerdo a los propósitos para los cuales ha sido concebida, existen diferentes formas de definir la función de estructura $\Phi(\mathbf{X})$ de una red de comunicaciones. El enfoque clásico se apoya en el problema de \mathcal{K} -terminales. En este caso la red se considera *operativa* ($\Phi(\mathbf{X}) = 1$) si todos los pares de nodos de $\mathcal{K} \subseteq \mathcal{V}$ están conectados, mientras que si no todos los pares de nodos de \mathcal{K} están conectados, la red se considera *no operativa* o *fallada* ($\Phi(\mathbf{X}) = 0$).

La *Confiabilidad Diámetro Acotada* [13] es una generalización del problema de \mathcal{K} -terminales, en la cual el concepto de “conectado”, es reemplazado por el de “ D -conectado” con el siguiente significado: dos nodos están “ D -conectados” si existe entre ambos un camino formado por D o menos enlaces *operativos*. Luego, $\Phi(\mathbf{X}) = 1$ si todos los nodos de $\mathcal{K} \subseteq \mathcal{V}$ están D -conectados y, si no todos los nodos de \mathcal{K} están D -conectados, $\Phi(\mathbf{X}) = 0$. Si a cualquier mensaje le toma un tiempo (retardo) T atravesar un enlace de la red, la *Confiabilidad Diámetro Acotada* será la probabilidad de que un mensaje pueda viajar entre cualquier par de nodos de \mathcal{K} en un tiempo no mayor que DT .

El siguiente algoritmo implementa la función de estructura en el problema de *Confiabilidad Diámetro Acotada*, para el caso $\mathcal{K} = \mathcal{V}$, es decir cuando interesa verificar si todos los nodos de \mathcal{V} están D -conectados, lo que equivale a decir si la red completa es o no D -conectada. El algoritmo ejecuta D iteraciones de una *Búsqueda Primero en Anchura* (BFS–Breadth First Search) desde cada nodo de \mathcal{V} ; si al cabo de la D -ésima iteración, los $\mathcal{V} - 1$ nodos restantes son visitados, la red es D -conectada y $\Phi(\mathbf{X}) = 1$, en caso contrario, la red no es D -conectada y $\Phi(\mathbf{X}) = 0$.

```

para(todos los nodos s de la red){
  enqueue_Q(s)
  num=1
  hacer D veces{
    mientras(queue_Q no esté vacía){
      d=dequeue_Q()
      marcar d como visitado
      para(todos los vecinos n de d)
        si(n no ha sido visitado){
          marcar n como visitado
          enqueue_B(n)
          num=num+1
        }
      }
    mientras(queue_B no esté vacía)
      enqueue_Q(dequeue_B())
    }
  si(num < número de nodos de la red){
    "la red no es D-conectada"
    terminar
  }
}
"la red es D-conectada"

```

Hasta donde sabemos, no hay en la literatura estudios de métodos de *Monte Carlo* para estimar la medida de *Confiabilidad Diámetro Acotada*; sólo algoritmos de cálculo exacto (de tiempo exponencial, en el caso general) y algunos límites han sido estudiados [1, 2, 8, 13].

3 Monte Carlo Crudo

El más simple de los algoritmos de *Monte Carlo* (conocido como *Monte Carlo Estandar* o *Monte Carlo Crudo*) consiste en sortear N valores independientes de $\mathbf{X}^{(i)}$ a partir de $f_{\mathbf{X}}(\mathbf{x})$ y determinar estimaciones para R y Q como:

$$\hat{R} = \frac{1}{N} \sum_{i=1}^N \Phi(\mathbf{X}^{(i)}) \quad \hat{Q} = \frac{1}{N} \sum_{i=1}^N (1 - \Phi(\mathbf{X}^{(i)})) \quad (2)$$

donde la distribución de probabilidad de cada $\mathbf{X}^{(i)}$ es la misma que la de \mathbf{X} . \hat{R} (\hat{Q}) es un estimador insesgado de R (Q) porque $\mathbb{E}\{\hat{R}\} = R$ ($\mathbb{E}\{\hat{Q}\} = Q$). La principal debilidad de estos estimadores es su baja precisión relativa, especialmente cuando la medida a estimar es muy pequeña (cerca de 0). Esto puede apreciarse en el caso de Q : su varianza exacta es $\mathbb{V}\{\hat{Q}\} = Q(1 - Q)/N$, y su error relativo, expresado mediante el coeficiente de variación, es $\mathbb{V}\{\hat{Q}\}^{1/2}/\mathbb{E}\{\hat{Q}\} = [(1 - \hat{Q})/((N - 1)\hat{Q})]^{1/2}$ (dado que $\hat{Q}(1 - \hat{Q})/(N - 1)$ es un estimador insesgado de $\mathbb{V}\{\hat{Q}\}$). Luego, para el caso de redes altamente confiables, para las que Q es un valor pequeño, una estimación precisa requiere de un gran número de muestras N .

Los métodos presentados en las secciones siguientes procuran mejorar el rendimiento de *Monte Carlo Crudo*, fundamentalmente a través de una reducción de varianza.

4 Variables Antitéticas, el Método *Dagger*

Si las muestras utilizadas en (2) para realizar la estimación son independientes, serán por lo tanto no correlacionadas. Es posible, no obstante, trabajar con muestras correlacionadas y si el tipo de correlación entre muestras es adecuadamente elegido es posible conseguir, a costa de esa correlación, una disminución en el valor de la varianza del estimador. Una forma de implementar estas ideas es mediante el Método de las Variables Antitéticas y sus variantes, una de las cuales, el método *Dagger* [5, 11], se presenta en esta sección.

Cada $\Phi(\mathbf{X}^{(i)})$ de (2) es una variable aleatoria binaria (de Bernoulli) con probabilidad R de valer 1 y $(1 - R)$ de valer 0, por lo que la varianza de \hat{R} es $R(1 - R)$. Luego:

$$\mathbb{V}\{\hat{R}\} = \mathbb{V}\left\{\frac{1}{N}\sum_{i=1}^N\Phi(\mathbf{X}^{(i)})\right\} \quad (3)$$

$$= \frac{1}{N^2}\sum_{i=1}^N\mathbb{V}\{\Phi(\mathbf{X}^{(i)})\} = \frac{1}{N^2}\sum_{i=1}^N\mathbb{V}\{\Phi(\mathbf{X})\} \quad (4)$$

$$= \frac{1}{N^2}N R(1 - R) \quad (5)$$

$$= \frac{R(1 - R)}{N} \quad (6)$$

Si bien R no se conoce, es simple probar que:

$$E\left\{\frac{\hat{R}(1 - \hat{R})}{N - 1}\right\} = \frac{R(1 - R)}{N} \quad (7)$$

es decir, $\hat{R}(1 - \hat{R})/(N - 1)$ es un estimador insesgado de $\mathbb{V}\{\hat{R}\}$.

Es posible probar que también la siguiente expresión corresponde a un estimador insesgado de $\mathbb{V}\{\hat{R}\}$:

$$\frac{1}{N - 1}\sum_{i=1}^N[\Phi(\mathbf{X}^{(i)}) - \hat{R}]^2 = \frac{1}{N - 1}\left[\left(\sum_{i=1}^N\Phi(\mathbf{X}^{(i)})\right)^2 - N\hat{R}^2\right] \quad (8)$$

La determinación de \hat{R} presentada hasta aquí corresponde al método de *Monte Carlo Crudo*. No se contempla ninguna intervención sobre las muestras o replicaciones $\Phi(\mathbf{X}^{(i)})$, las que simplemente resultan elegidas en forma independiente entre todos los valores posibles. Pero volviendo a la determinación del valor exacto de $\mathbb{V}\{\hat{R}\}$, la ecuación (4) toma en cuenta el carácter independiente de las muestras, según lo anticipado. Si se elimina esta hipótesis de independencia resulta otro estimador de R , \hat{R}' , también insesgado, cuya varianza es:

$$\mathbb{V}\{\hat{R}'\} = \mathbb{V}\left\{\frac{1}{N}\sum_{i=1}^N\Phi(\mathbf{X}^{(i)})\right\} \quad (9)$$

$$= \frac{1}{N^2}\mathbb{V}\left\{\sum_{i=1}^N\Phi(\mathbf{X}^{(i)})\right\} \quad (10)$$

$$= \frac{1}{N^2}\left\{\sum_{1 \leq i \leq N}\mathbb{V}\{\Phi(\mathbf{X}^{(i)})\} + 2\sum_{1 \leq i < j \leq N}\mathbb{Cov}(\Phi(\mathbf{X}^{(i)}), \Phi(\mathbf{X}^{(j)}))\right\} \quad (11)$$

$$= \frac{R(1 - R)}{N} + \frac{2}{N^2}\sum_{1 \leq i < j \leq N}\mathbb{Cov}(\Phi(\mathbf{X}^{(i)}), \Phi(\mathbf{X}^{(j)})) \quad (12)$$

$$= \mathbb{V}\{\hat{R}\} + \frac{2}{N^2}\sum_{1 \leq i < j \leq N}\mathbb{Cov}(\Phi(\mathbf{X}^{(i)}), \Phi(\mathbf{X}^{(j)})) \quad (13)$$

A partir de la expresión (13) resulta evidente que si la dependencia entre los pares de muestras ($\Phi(\mathbf{X}^{(i)}), \Phi(\mathbf{X}^{(j)})$) es tal que la suma de las covarianzas da un valor negativo, el estimador \hat{R}' será más preciso que \hat{R} y, consecuentemente, permitirá satisfacer las mismas especificaciones de error con un menor número de muestras o replicaciones. Alcanzar este objetivo implica un beneficio sobre otro: por una lado, en tanto método de *Monte Carlo*, la estimación se realiza con un número de muestras “reducido”, por otro si las muestras se distribuyen de modo que en conjunto tengan

correlación negativa, la calidad del estimador equivaldrá a la de otro obtenido con muestras independientes pero con mayor número de replicaciones.

No es sencillo establecer criterios de dependencia entre las muestras. En la expresión (13) se evalúan todos los pares, por lo que si fuera posible sortear a partir de una distribución tal que todo par de muestras resulte correlacionado negativamente, la reducción de varianza sería alcanzada sin lugar a dudas. No es necesario de todas maneras apuntar a tan exigente objetivo, basta con que “algunos” pares de muestras estén correlacionados negativamente, mientras los demás sean independientes y consecuentemente su covarianza sea nula. Por otro lado la correlación negativa debe pesar sobre pares $(\Phi(\mathbf{X}^{(i)}), \Phi(\mathbf{X}^{(j)}))$, pero los valores que hay que generar en la simulación, y sobre cuya distribución de probabilidad se tiene acceso o capacidad de influir, son los $\mathbf{X}^{(i)}$ y no los $\Phi(\mathbf{X}^{(i)})$. Se puede probar, no obstante, que si la función de estructura $\Phi(\mathbf{X})$ es monótona, cualquier par $(\mathbf{X}^{(i)}, \mathbf{X}^{(j)})$ negativamente correlacionado hará que el par $(\Phi(\mathbf{X}^{(i)}), \Phi(\mathbf{X}^{(j)}))$ tenga también correlación negativa. Y más aún, la correlación negativa de los pares $(\Phi(\mathbf{X}^{(i)}), \Phi(\mathbf{X}^{(j)}))$ puede garantizarse tan sólo con que la correlación negativa se de sobre alguna de las coordenadas de \mathbf{X} , sea ésta X_1, X_2, \dots , o X_m , es decir sobre alguno de los enlaces del modelo de red. Resumiendo, es posible garantizar la correlación negativa entre pares $(\Phi(\mathbf{X}^{(i)}), \Phi(\mathbf{X}^{(j)}))$ de muestras logrando correlación negativa sobre al menos sobre un enlace, es decir $\text{Cov}(X_k^{(i)}, X_k^{(j)})$ para al menos algún k , siendo independientes los restantes pares.

Existen mecanismos que permiten generar muestras correlacionadas por bloques. Conviene entonces adecuar a este caso las expresiones presentadas. Supóngase que es posible generar conjuntos de L muestras de \mathbf{X} , $\{\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \dots, \mathbf{X}^{(L)}\}$, tales que tengan correlación negativa todos los pares $(\mathbf{X}^{(i)}, \mathbf{X}^{(j)})$ del conjunto. A partir de estas muestras, es decir dentro del conjunto, es posible definir el siguiente estimador parcial con su correspondiente varianza:

$$\hat{R}_L = \frac{1}{L} \sum_{i=1}^L \Phi(\mathbf{X}^{(i)}) \quad (14)$$

$$\mathbb{V}\{\hat{R}_L\} = \frac{1}{L^2} \left\{ \sum_{1 \leq i \leq L} \mathbb{V}\{\Phi(\mathbf{X}^{(i)})\} + 2 \sum_{1 \leq i < j \leq L} \text{Cov}(\Phi(\mathbf{X}^{(i)}), \Phi(\mathbf{X}^{(j)})) \right\} \quad (15)$$

Para el caso de N muestras, con $N = BL$, es decir tomando B conjuntos de L muestras correlacionadas, es posible reescribir la expresión del estimador \hat{R}' por medio del estimador parcial. Dado que el estimador parcial toma valores distintos para cada uno de los B conjuntos de L muestras, conviene redefinir el estimador parcial agregando un índice, b (tanto al estimador como a las muestras), que indique sobre cuál de los B conjuntos están definidos:

$$\hat{R}_L^b = \frac{1}{L} \sum_{i=1}^L \Phi(\mathbf{X}^{(i,b)}) \quad b = 1, 2, \dots, B \quad (16)$$

Finalmente el estimador \hat{R}' de $\mathbb{E}\{\Phi(\mathbf{X})\}$ y su respectiva varianza pueden reescribirse como:

$$\hat{R}' = \frac{1}{B} \sum_{b=1}^B \hat{R}_L^b \quad (17)$$

$$\mathbb{V}\{\hat{R}'\} = \mathbb{V}\{\hat{R}\} + \frac{2}{BL^2} \sum_{1 \leq i < j \leq L} \text{Cov}(\Phi(\mathbf{X}^{(i)}), \Phi(\mathbf{X}^{(j)})) \quad (18)$$

Resta determinar un estimador para la varianza de \hat{R}' . Si se toma en cuenta que la expresión

$$\frac{1}{B-1} \sum_{b=1}^B (\hat{R}_L^b - \hat{R}')^2 \quad (19)$$

es un estimador insesgado de la varianza del estimador parcial, entonces

$$\frac{1}{B(B-1)} \sum_{b=1}^B (\hat{R}_L^b - \hat{R}')^2 = \frac{1}{(B-1)} \left[\left(\frac{1}{B} \sum_{b=1}^B (\hat{R}_L^b)^2 \right) - \hat{R}'^2 \right] \quad (20)$$

es un estimador insesgado de la varianza de la varianza de \hat{R}' .

Similares consideraciones y propiedades valen para la estimación de anti-confiabilidad \hat{Q} para la que, además, los estimadores tienen la misma forma que los de \hat{R} .

4.1 Dagger

Sea un vector aleatorio $\mathbf{Y} = (Y_1, Y_2)$ con Y_1 distribuida uniformemente en $[0, 1]$ e $Y_2 = 1 - Y_1$. Una forma de sortear valores para \mathbf{Y} a partir de un generador de números aleatorios uniformemente distribuidos en $[0, 1]$ es asignar la salida U del generador a la componente Y_1 y su complemento, $1 - U$, a la componente Y_2 . Dado que $\mu_{Y_1} = \mu_{Y_2} = \frac{1}{2}$:

$$\text{Cov}(Y_1, Y_2) = \mathbb{E}\{(Y_1 - \mu_{Y_1})(Y_2 - \mu_{Y_2})\} = -\frac{1}{12} \quad (21)$$

Luego, dos variables aleatorias uniformemente distribuidas en $[0, 1]$, cuyos valores se sortean a partir de U y $1 - U$ respectivamente, están correlacionadas negativamente.

Sea ahora un vector $\mathbf{X} = (X_1, X_2)$ donde X_1 y X_2 son dos variables aleatorias de Bernoulli de parámetro r . Si se sortean valores para cada una de ellas a partir de un generador de números aleatorios U uniformemente distribuidos en el $[0, 1]$, de acuerdo al siguiente algoritmo:

$$\begin{array}{ll} \text{if } U < r & X_1 = 1 \quad \text{else } X_1 = 0 \\ \text{if } (1 - U) < r & X_2 = 1 \quad \text{else } X_2 = 0 \end{array}$$

es posible probar, mediante un análisis caso por caso [5], que

$$\text{Cov}(X_1, X_2) = -q^2 \quad (22)$$

donde:

$$q = \begin{cases} r & \text{si } r < 0.5 \\ 1 - r & \text{si } r \geq 0.5 \end{cases}$$

En el análisis caso por caso, a través del cual se determina $\text{Cov}(X_1, X_2)$, es necesario conocer el valor esperado de cada par (X_1, X_2) , para lo cual es preciso conocer el espacio de valores de U que le corresponde, es decir el conjunto de valores de U que en caso de ser sorteado da lugar al correspondiente par de valores (X_1, X_2) . Todos los casos posibles se muestran en el siguiente detalle:

X_1	X_2	Conjunto de valores de U ($r < 0.5$)	Conjunto de valores de U ($r \geq 0.5$)
1	0	$[0, r[$ 	$[0, 1 - r[$ 
0	1	$[1 - r, 1[$ 	$[r, 1[$ 
0	0	$[r, 1 - r[$ 	0 
1	1	0 	$[1 - r, r[$ 

La interpretación que corresponde es la siguiente: para cada posible valor de U se aplica el algoritmo de asignación de valores a X_1 y X_2 . Según resulten las comparaciones de U y $1 - U$ contra r corresponde un valor para cada variable. En definitiva cada par de valores de X_1 y X_2 puede ponerse en correspondencia con un conjunto de valores de U . La gama de valores de U asociada a cada par X_1 y X_2 es la que se indica con trazo grueso en el esquema de líneas del detalle.

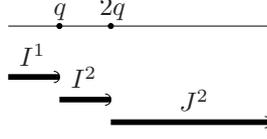
Dado que la variable U está distribuida uniformemente en el intervalo $[0, 1]$, desde el punto de vista de los valores resultantes lo que cuenta es el ancho del sector de trazo grueso y no su ubicación relativa dentro del intervalo $[0, 1]$. Es posible entonces redistribuir los sectores de trazo grueso según se muestra en el siguiente detalle:

X_1	X_2	Conjunto de valores de U ($r < 0.5$)	Conjunto de valores de U ($r \geq 0.5$)
1	0	$[0, r[$ 	$[1 - r, 2(1 - r)[$ 
0	1	$[r, 2r[$ 	$[0, 1 - r[$ 
0	0	$[2r, 1[$ 	0 
1	1	0 	$[2(1 - r), 1[$ 

Este nuevo esquema no genera los mismos valores de X_1 y X_2 que el anterior por cada valor de U , pero si conserva la misma relación probabilística entre las variables. Consecuentemente en grandes cantidades genera valores de X_1 y

X_2 en la misma proporción y, lo que es más importante, da lugar a la misma covarianza entre los valores resultantes que el esquema anterior.

Reemplazando r por q (según se definió antes) y nombrando los sectores resultantes como: $I^1 = [0, q[$, $I^2 = [q, 2q[$ y $J^2 = [2q, 1[$ respectivamente, como se muestra en la siguiente figura:



Este último mecanismo de asignación de valores para X_1 y X_2 a partir de U , puede resumirse en la siguiente expresión:

$$X_i = (1 - X_{default}) 1_{(U \in I^i)} + X_{default} 1_{(U \notin I^i)} \quad i = 1, 2 \quad (23)$$

donde:

$$X_{default} = \begin{cases} 0 & \text{si } r < 0.5 \\ 1 & \text{si } r \geq 0.5 \end{cases}$$

Este esquema puede generalizarse al caso en que el intervalo $[0, 1]$ se particione en L subintervalos, siendo L el máximo número de subintervalos I^i de ancho q que caben en el $[0, 1]$, quedando al final un único intervalo J de dimensión menor o igual a q . Se puede probar que en este caso la expresión (23) toma la siguiente forma general, conservándose las propiedades de correlación entre todo par de muestras:

$$X_i = (1 - X_{default}) 1_{(U \in I^i)} + X_{default} 1_{(U \notin I^i)} \quad i = 1, 2, \dots, L \quad (24)$$

Esta expresión define un mecanismo cuya interpretación es la siguiente. En función del valor de q se elige L , pudiendo éste ser un entero cualquiera tal que $1 < L \leq \lfloor 1/q \rfloor$. Por cada valor de U se genera una L -upla de valores $\{X_1, X_2, \dots, X_L\}$ con $X_i \in \{0, 1\}$, $i = 1, 2, \dots, L$. En grandes cantidades resultará:

$$Cov(X_i, X_j) = -q^2 \quad (25)$$

para todo par de variables (X_i, X_j) de la L -upla.

Los algoritmos deberán explotar el hecho de que a lo sumo uno de los X_i de la L -upla resultará igual a $(1 - X_{default})$ ya que cada U caerá dentro de (a lo sumo) uno de los I^i , mientras que los restantes (eventualmente todos) se mantendrán en el valor $X_{default}$.

Numéricamente la cantidad de valores posibles para la L -upla es de 2^L . Pero el algoritmo de la expresión (24) restringe ese espacio a sólo $L + 1$ valores ya que los posibles valores de U (cada uno de los cuales permite armar una L -upla completa) pueden caer en uno de $L + 1$ subintervalos $(I^1, I^2, \dots, I^L, J^L)$.

Es evidente que la elección de un valor elevado para L (el más elevado posible, llegado el caso) no hará descender la varianza del estimador más allá de $-q^2$. El beneficio de elegir un valor elevado de L está en la reducción de la cantidad de accesos al generador de números aleatorios. Para $L = 2$ con un único valor aleatorio U es posible generar dos instancias de la variable de interés. En definitiva por cada valor de U es posible obtener L instancias de la variable de interés (para cualquier valor válido de L).

Supóngase ahora que la secuencia $\{X_1, X_2, \dots, X_L\}$ constituye una serie de L replicaciones de la variable asociada al estado de un enlace del modelo de red de comunicaciones, por ejemplo el k -ésimo enlace, esto es: $\{X_k^{(1)}, X_k^{(2)}, \dots, X_k^{(L)}\}$. Valen todas las consideraciones previas en cuanto a que basta con que se sorteen valores correlacionados negativamente sobre un único enlace para que en el cálculo de la varianza de un estimador de $\mathbb{E}\{\Phi(\mathbf{X})\}$ se vea reducida producto de las covarianzas negativas de los correspondientes pares $(\Phi(\mathbf{X}^{(i)}), \Phi(\mathbf{X}^{(j)}))$. Es factible no obstante sortear valores para todos los enlaces del modelo con el mismo algoritmo. La dificultad está en el hecho de que la confiabilidad r_i de cada enlace no tiene por qué ser la misma lo que hará que, en el caso más general, no exista un único valor L para el modelo sino un L_i para cada enlace.

EJEMPLO (Aplicación del método *Dagger*):

Sea una red de 3 enlaces, $\mathcal{E} = \{e_1, e_2, e_3\}$, con los siguientes valores característicos (el significado de los n_i se explica inmediatamente),

$$\begin{array}{llll} r_1 = 0.95 & q_1 = 0.05 & L_1 = 20 & n_1 = 1 \\ r_2 = 0.90 & q_2 = 0.10 & L_2 = 10 & n_2 = 2 \\ r_3 = 0.80 & q_3 = 0.20 & L_3 = 5 & n_3 = 4 \end{array}$$

Por cada valor aleatorio U será posible generar las siguientes instancias para cada enlace:

$$\begin{aligned} & \{X_1^{(1)}, X_1^{(2)}, X_1^{(3)}, X_1^{(4)}, X_1^{(5)}, X_1^{(6)}, X_1^{(7)}, X_1^{(8)}, X_1^{(9)}, X_1^{(10)}, X_1^{(11)}, X_1^{(12)}, \dots, X_1^{(20)}\} \\ & \{X_2^{(1)}, X_2^{(2)}, X_2^{(3)}, X_2^{(4)}, X_2^{(5)}, \dots, X_2^{(10)}\} \\ & \{X_3^{(1)}, X_3^{(2)}, \dots, X_3^{(5)}\} \end{aligned}$$

Una forma de proceder es la que propone el algoritmo correspondiente al *Método Dagger*. La idea es determinar un valor de L como el mínimo común múltiplo de todos los L_i (20 en este caso, pero no siempre uno de los L_i) y ver cuántas veces cabe cada L_i dentro de L . Esta relación, $n_i = L/L_i$, cuyos valores se muestran al comienzo del ejemplo, indica la cantidad de valores de U necesarios para generar L instancias en cada enlace.

Existen otras formas de elegir el valor de L y su respectiva relación con los L_i las que, sin apartarse en esencia de las ideas precedentes, dan lugar a variantes del algoritmo que en líneas generales se conoce como *Monte Carlo con Variables Antitéticas*. En su versión más común el algoritmo se presenta para el caso $L = 2$, mientras que si $L = 1$ el método se convierte en el de *Monte Carlo Crudo*.

5 Monte Carlo Permutación

Es posible estimar la confiabilidad de las redes basándose en la observación de la evolución temporal del estado de los enlaces. Una forma de hacerlo es asumir que al comienzo todos están *fallados* (*operativos*) y que se van poniendo *operativos* (*fallados*) a lo largo del tiempo. Eligiendo adecuadamente las distribuciones de probabilidad que regulan las conmutaciones y analizando el orden en el que los distintos enlaces van cambiando de estado (permutación) es posible determinar un estimador de la confiabilidad que presenta ventajas respecto al de *Monte Carlo Crudo* desde el punto de vista de su varianza [6, 10, 9].

El análisis correspondiente al método de *Monte Carlo Crudo* es de tipo *estático*. En un determinado instante se observa la red y se decide si está *operativa* o no. Para ello se hace el recuento de enlaces que se encuentran *operativos*, de los que se hallan *fallados* y se aplica el criterio que determina la operatividad de la red. El experimento se repite gran cantidad de veces, sin consideración alguna respecto a “cuál” es el instante de observación, y la proporción de experimentos que encuentran a la red en estado *operativo* resulta un indicador de su confiabilidad.

En la práctica no se realizan las observaciones de la red tal como se acaba de describir. Lo que se hace es recrear en forma artificial ese escenario, para lo cual se sortea un estado probable para cada enlace, quedando así definido un estado probable para la red y sobre el estado de la red así definido se aplica el criterio que determina sus condiciones de operatividad.

Supóngase ahora que en lugar de este modelo *estático*, en el que se observan los enlaces en forma instantánea y atemporal, se plantea un modelo *dinámico* que permita analizar la red a través del tiempo. Es preciso entonces establecer el comportamiento de los enlaces como función del tiempo y determinar los instantes de inicio y fin de la observación. Supóngase que al comienzo ($t = 0$) todos los enlaces están *fallados* y que van pasando al estado *operativo* en instantes $\tau(i)$, $i = 1, 2, \dots, m$, distribuidos exponencialmente con tasa $\lambda(i)$, esto es: $\mathbb{P}[\tau(i) \leq t] = 1 - e^{-\lambda(i)t}$. Supóngase además que los enlaces evolucionan en forma mutuamente independiente. El vector $\mathbf{X} = (X_1, X_2, \dots, X_m)$ se convierte entonces en $\mathbf{X}(t) = (X_1(t), X_2(t), \dots, X_m(t))$, vector éste que describe por completo el estado de la red. $\mathbf{X}(t)$ es un proceso de Markov con espacio de estados $\{0, 1\}^m$ llamado *Proceso de Construcción* [6].

En un instante t' cualquiera, existe una probabilidad $(1 - e^{-\lambda(i)t'})$ de que el i -ésimo enlace haya pasado del estado *fallado* al estado *operativo*. Nótese que si se elige $\lambda(i) = -\log(q_i)$, con $q_i = 1 - r_i$, en el instante $t = 1$ la probabilidad de que el i -ésimo enlace esté *operativo* es exactamente r_i . Si en $t = 1$ cada enlace tiene una probabilidad r_i de encontrarse *operativo*, entonces en ese instante la red tiene una probabilidad R de encontrarse en un estado de operatividad aceptable² y, consecuentemente, una probabilidad Q de encontrarse en un estado de operatividad no aceptable. Por esta razón el instante $t = 1$ puede perfectamente hacer las veces del instante de observación del modelo *estático*, en el modelo *dinámico*. Basta entonces con recrear un escenario que permita chequear la condición de operatividad, habiendo estado todos los enlaces *fallados* en $t = 0$ y con probabilidad de pasar al estado *operativo* en instantes $\tau(i)$, $i = 1, 2, \dots, m$. Para algunos enlaces la transición entre estados ocurrirá en $t \leq 1$ con lo que, si el experimento se repite gran cantidad de veces, la proporción de los mismos para los que resulte $\Phi(\mathbf{X}(1)) = 1$ será un indicador equivalente (el mismo, concretamente) al de confiabilidad R obtenido mediante el método de *Monte Carlo Crudo*. Del mismo modo la proporción de experimentos para los que resulte $\Phi(\mathbf{X}(1)) = 0$ (o, lo que es lo mismo, $1 - \Phi(\mathbf{X}(1)) = 1$) será un indicador de la anti-confiabilidad Q :

$$R = \mathbb{E}\{\Phi(\mathbf{X}(1))\} \qquad Q = \mathbb{E}\{1 - \Phi(\mathbf{X}(1))\} \qquad (26)$$

¹ r_i es la probabilidad de encontrar al i -ésimo enlace en estado *operativo* en un instante cualquiera, según se definió en la Sección 1

² R es el valor de confiabilidad de la red, según se definió en la Sección 1

Debe notarse que se han definido los mismos indicadores de confiabilidad y anti-confiabilidad cuya estimación fue abordada mediante el método de *Monte Carlo Crudo*, con la sola diferencia de que el estado de los enlaces se sorteó mediante un artificio que es computacionalmente costoso (y claramente innecesario). Consecuentemente la varianza del estimador así obtenido será la misma que la del método de *Monte Carlo Crudo*. La verdadera ventaja del modelo *dinámico* (que quedará demostrada más adelante) no se obtiene entonces sorteando el estado de los enlaces para $t = 1$ y viendo si la red está *operativa* o no, sino sorteando, cada vez, una permutación, viendo que probabilidad le corresponde a la permutación sorteada de poner la red *operativa* en $t = 1$, y luego promediando todas esas probabilidades [9, 10]. Por permutación se entiende lo siguiente: para cada realización del proceso $\mathbf{X}(t)$ existe un orden en el que los enlaces van cambiando su estado, este orden es una variable aleatoria Ω que toma valores en el espacio de permutaciones de \mathcal{E}^3 .

NOTACIÓN: Hasta el momento se ha utilizado una notación según la cual cada enlace e es identificado dentro de la red con un subíndice i , esto es e_i , $i = 1, 2, \dots, m$. Mediante un argumento con el mismo valor del subíndice se identifican también la tasa correspondiente, $\lambda(i)$, y el instante de conmutación, $\tau(i)$.

$$e_i \rightarrow \lambda(i), \tau(i) \quad i = 1, 2, \dots, m$$

Al considerar las permutaciones es preciso identificar, además, la ubicación de cada enlace dentro de la permutación. Esto conduce a la necesidad de una indicación adicional en cada enlace, por ejemplo un supraíndice de la forma e^j , $j = 1, 2, \dots, m$. Así una permutación, en notación completa, toma la forma $(e_1^1, e_2^2, e_3^3, \dots, e_m^m)$, cuando $\tau(i) < \tau(j) < \tau(k) < \dots < \tau(l)$. Puede ser que en algunos casos sólo interese una de las indicaciones, siendo irrelevante la otra. Si se trata de tomar en cuenta únicamente la ubicación dentro de la permutación, basta con el supraíndice. En ese caso podría ser necesaria también una referencia a la tasa y al instante de conmutación del enlace correspondiente, lo que puede notarse mediante el mismo supraíndice:

$$e^j \rightarrow \lambda^j, \tau^j \quad j = 1, 2, \dots, m$$

Algunas propiedades de las permutaciones pueden ponerse en función del conjunto de enlaces que, para cada subintervalo de tiempo, se encuentra todavía *fallado*. Esto es:

$$\begin{cases} \mathcal{E}_0 = \mathcal{E} & 0 \leq t < \tau^1 \\ \mathcal{E}_i = \mathcal{E}_{i-1} - e^i & \tau^i \leq t < \tau^{i+1} \\ \mathcal{E}_m = \emptyset & \tau^m \leq t \end{cases} \quad i = 1, 2, \dots, m-1$$

entonces:

$$\mathbb{P}[\Omega = \omega] = \prod_{i=1}^m \frac{\lambda^i}{\lambda(\mathcal{E}_{i-1})} \quad (27)$$

donde $\lambda(\mathcal{E}_i) = \sum_{e^j \in \mathcal{E}_i} \lambda^j = \sum_{j=i+1}^m \lambda^j$.

Por otro lado los tiempos $A_i = \tau^{i+1} - \tau^i$, $i = 0, 1, \dots, m-1$ (aceptando $\tau^0 = 0$) se distribuyen exponencialmente con parámetro $\lambda(\mathcal{E}_i)$.

La red alcanza el estado de operatividad aceptable (es decir la condición $\Phi(\mathbf{X}(t)) = 1$) en un instante distinto cada vez, según cuál de las posibles permutaciones ω ocurra. En rigor, para conocer el instante en que se alcanza la condición $\Phi(\mathbf{X}(t)) = 1$ no basta con conocer la permutación ω , es decir el orden de conmutación de los enlaces, sino también los tiempos asociados a la conmutación de cada uno. El índice $b(\omega)$, correspondiente al enlace gracias al cual se llega a la condición $\Phi(\mathbf{X}(t)) = 1$ para la permutación ω , se denomina *número crítico* de ω :

$$b(\omega) = \min\{i \mid \Phi(\mathbf{X}(\tau^i)) = 1\} \quad (28)$$

Dado que la forma de recrear el escenario que permite hacer las evaluaciones (tanto de confiabilidad R como de anti-confiabilidad Q) consiste en generar distintas permutaciones ω , conviene condicionar las expresiones (26) al valor de la variable aleatoria Ω . Para el caso de la anti-confiabilidad resulta entonces:

$$Q = \mathbb{E}\{1 - \Phi(\mathbf{X}(1))\} = \sum_{\omega} \mathbb{P}[\Phi(\mathbf{X}(1)) = 0 \mid \Omega = \omega] \mathbb{P}[\Omega = \omega] \quad (29)$$

Donde la sumatoria se hace sobre todas las permutaciones ω . El término que se suma, es decir la probabilidad de que, dada una permutación, la red no haya alcanzado el estado de operatividad aceptable en $t = 1$, puede calcularse reemplazando t por 1 en la siguiente expresión:

$$\mathbb{P}[\Phi(\mathbf{X}(t)) = 0 \mid \Omega = \omega] = \mathbb{P}[A_0 + A_1 + \dots + A_{b(\omega)-1} > t \mid \Omega = \omega] \quad (30)$$

$$= \mathbb{P}\left[\text{Conv}_{0 \leq i < b(\omega)-1} \{e^{\lambda(\mathcal{E}_i)t}\} > t\right] \quad (31)$$

³ \mathcal{E} es el conjunto de todos los enlaces de la red, según se definió en la Sección 1

Cada A_i , se dijo, está distribuido exponencialmente con parámetro $\lambda(\mathcal{E}_i)$, por lo que la distribución de la suma de los A_i , desde 0 hasta $b(\omega) - 1$, es la convolución entre las $b(\omega)$ exponenciales correspondientes. Se puede prescindir del condicional $\Omega = \omega$ en (31) ya que los valores \mathcal{E}_i dependen de la permutación. Fijando $t = 1$ en (30) queda definida una función de la variable aleatoria Ω , $G(\Omega)$, cuyos valores corresponden a la probabilidad de que, dada una determinada permutación la red no satisfaga la condición de operatividad en $t = 1$. Esta función es:

$$G(\omega) = \mathbb{P}[\Phi(\mathbf{X}(1)) = 0 \mid \Omega = \omega] \quad (32)$$

de lo que surge inmediatamente que:

$$Q = \mathbb{E}\{G(\Omega)\} = \sum_{\omega} G(\omega) \mathbb{P}[\Omega = \omega] \quad (33)$$

Descartando el cálculo exacto, es posible realizar una estimación, \hat{Q} , por *Monte Carlo*, promediando N valores de la función G aplicada sobre replicaciones independientes e idénticamente distribuidas de la variable aleatoria Ω :

$$\hat{Q} = \frac{1}{N} \sum_{i=1}^N G(\Omega^{(i)}) \quad (34)$$

Cabe preguntarse entonces dónde está el beneficio del método. Cuando la anti-confiabilidad se define sobre modelos *estáticos* como por ejemplo el método de *Monte Carlo Crudo*, la misma corresponde al valor esperado de una variable de Bernoulli, siendo entonces su varianza:

$$\mathbb{V}_{Crudo}\{Q\} = Q(1 - Q), \quad (35)$$

Para el modelo *dinámico* actual (permutación), la anti-confiabilidad se define en (33) como el valor esperado de la variable aleatoria $G(\Omega)$, que no es de Bernoulli sino que toma valores en el intervalo $[0, 1]$. Su varianza es:

$$\mathbb{V}_{Perm}\{Q\} = \sum_{\omega} G(\omega)^2 \mathbb{P}[\Omega = \omega] - \left(\sum_{\omega} G(\omega) \mathbb{P}[\Omega = \omega] \right)^2 \quad (36)$$

Desde luego Q es la misma en ambos casos ya que la idea es calcular por distintos métodos un valor de anti-confiabilidad que representa el mismo criterio de operatividad para la red. Reemplazando entonces la expresión (33) en (35):

$$\mathbb{V}_{Crudo}\{Q\} = \sum_{\omega} G(\omega) \mathbb{P}[\Omega = \omega] \left(1 - \sum_{\omega} G(\omega) \mathbb{P}[\Omega = \omega] \right) \quad (37)$$

$$= \sum_{\omega} G(\omega) \mathbb{P}[\Omega = \omega] - \left(\sum_{\omega} G(\omega) \mathbb{P}[\Omega = \omega] \right)^2 \quad (38)$$

$$= \sum_{\omega} G(\omega) \mathbb{P}[\Omega = \omega] - \left(\sum_{\omega} G(\omega) \mathbb{P}[\Omega = \omega] \right)^2 \pm \sum_{\omega} G(\omega)^2 \mathbb{P}[\Omega = \omega] \quad (39)$$

$$= \mathbb{V}_{Perm}\{Q\} + \sum_{\omega} G(\omega) \mathbb{P}[\Omega = \omega] - \sum_{\omega} G(\omega)^2 \mathbb{P}[\Omega = \omega] \quad (40)$$

$$= \mathbb{V}_{Perm}\{Q\} + \sum_{\omega} G(\omega) \mathbb{P}[\Omega = \omega] (1 - G(\omega)) \quad (41)$$

de donde resulta $\mathbb{V}_{Perm}\{Q\} \leq \mathbb{V}_{Crudo}\{Q\}$.

Resta plantear la forma de hacer las replicaciones $G(\Omega^{(i)})$ que permitan estimar la anti-confiabilidad mediante la expresión (34). Pero antes vale la pena hacer una interpretación del significado de la variable aleatoria $G(\Omega)$. Los valores que toma esta función corresponden al valor de probabilidad de que, dada una permutación determinada, es decir un cierto orden en el que los enlaces pasan al estado *operativo*, la red no llegue a ponerse *operativa* en $t \leq 1$. Tal como se acaba de decir, las permutaciones ω hacen referencia únicamente al orden y no a los instantes en que los enlaces cambian de estado. La Figura 1 muestra dos realizaciones de una misma permutación: en el caso (a) la red alcanza el estado *operativo* en un instante posterior a $t = 1$, mientras que en el caso (b) ocurre lo contrario. El enlace que define los dos casos posibles es el que corresponde al *número crítico*, $b(\omega)$. Luego, dada una permutación, la condición que debe verificar y cuya probabilidad se debe evaluar, es: $A_0 + A_1 + \dots + A_{b(\omega)-1} > 1$ o, lo que es lo mismo, $\text{Conv} \{e^{\lambda(\mathcal{E}_i)}\}_{0 \leq i < b(\omega)-1} > 1$.

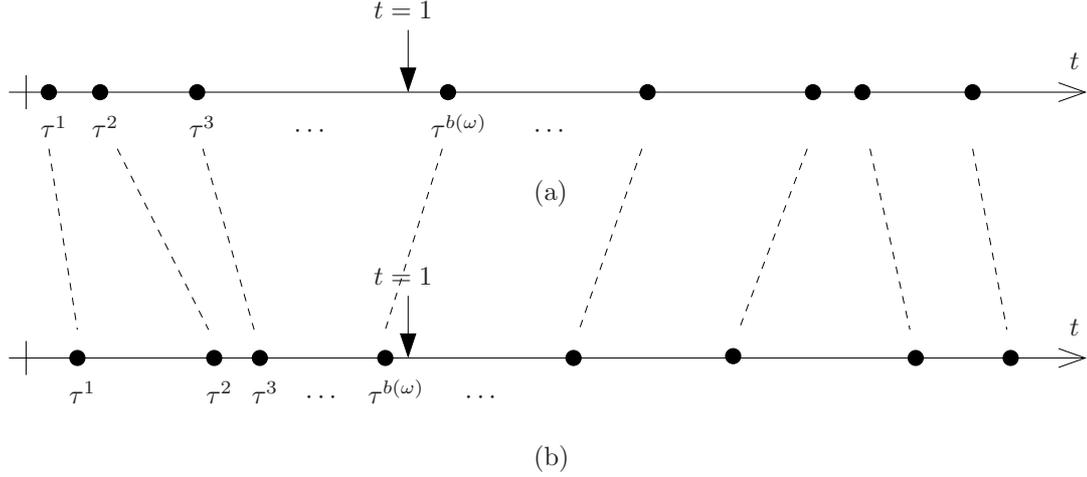


Figura 1: Dos Realizaciones de la Misma Permutación

El mecanismo para implementar este método puede resumirse entonces en los siguientes pasos:

1. Generar una permutación ω para lo cual es necesario sortear, a partir de distribuciones exponenciales⁴ de parámetro $\lambda(i)$, los tiempos $\tau(i)$ de conmutación de todos los enlaces y ordenarlos según los tiempos sorteados, esto es:

$$(e_i^1, e_j^2, e_k^3, \dots, e_l^m) \text{ si } \tau(i) < \tau(j) < \tau(k) < \dots < \tau(l)$$

2. Identificar el *número crítico* para lo cual se debe aplicar la función de estructura sobre los grafos resultantes de ir agregando cada vez un enlace, en el orden indicado por la permutación sorteada. Una vez que la función de estructura deja de valer 0, es decir cuando vale 1 por primera vez, el *número crítico* $b(\omega)$ es el número de orden (en la permutación sorteada) del último enlace agregado⁵.
3. Calcular la siguiente *fdp* (función de densidad de probabilidad): $\text{Conv}_{0 \leq i < b(\omega) - 1} \{e^{\lambda(\mathcal{E}_i)t}\}$ correspondiente a la permutación sorteada. Esta *fdp* es (su expresión matemática, una vez calculada) una suma de exponenciales. Calcular luego la función de probabilidad acumulada de esta *fdp*, para lo cual basta reemplazar cada exponencial de la forma $\lambda(i)e^{\lambda(i)t}$ por su acumulada, $(1 - e^{\lambda(i)t})$. Valorizar luego la función de probabilidad acumulada en $t = 1$ y acumular el valor resultante.
4. Dividir los valores acumulados por su cantidad y el resultado obtenido es un estimador de la anti-confiabilidad de la red.

Para la determinación de la convolución del punto 3 tómesese en cuenta que la suma de dos variables aleatorias independientes, distribuidas exponencialmente: $f_0(t) = \lambda_0 e^{-\lambda_0 t}$ y $f_1(t) = \lambda_1 e^{-\lambda_1 t}$, con $\lambda_0 \neq \lambda_1$, se distribuye según:

$$f_{0,1}(t) = f_0(t) * f_1(t) \quad (42)$$

$$= \lambda_0 e^{-\lambda_0 t} * \lambda_1 e^{-\lambda_1 t} \quad (43)$$

$$= \frac{\lambda_0 \lambda_1}{\lambda_0 - \lambda_1} (e^{-\lambda_1 t} - e^{-\lambda_0 t}) \quad (44)$$

$$= \frac{\lambda_0 f_1(t) - \lambda_1 f_0(t)}{\lambda_0 - \lambda_1} \quad (45)$$

Sumando una tercer variable, $f_2(t) = \lambda_2 e^{-\lambda_2 t}$, la distribución de la suma correspondiente resulta $f_{0,1,2}(t) = f_0(t) * f_1(t) * f_2(t)$ lo que, una vez desarrollado se convierte en:

$$f_{0,1,2}(t) = \frac{\lambda_0 \frac{\lambda_1 f_2(t) - \lambda_2 f_1(t)}{\lambda_1 - \lambda_2} - \lambda_1 \frac{\lambda_0 f_2(t) - \lambda_2 f_0(t)}{\lambda_0 - \lambda_2}}{\lambda_0 - \lambda_1} \quad (46)$$

Agrupando los términos de este modo se puede conservar la forma de la distribución buscada como una suma de distribuciones exponenciales. Resulta entonces inmediata la determinación de la distribución acumulada de esa

⁴Valores ajustados a una distribución exponencial de parámetro $\lambda(i)$ se pueden sortear como $(-\log U)/\lambda(i)$ donde U es una variable aleatoria uniformemente distribuida en $(0, 1]$.

⁵Se acepta que las redes bajo análisis verifican propiedades de monotonicidad tales que el agregado de enlaces nunca empeora sus condiciones de operatividad. Es decir, si $\mathcal{E}' \subseteq \mathcal{E}$ entonces si $\mathcal{G} = (\mathcal{V}, \mathcal{E}')$ es *operativa*, $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ también lo es.

suma. Basta reemplazar cada función de densidad de la forma $f_i(t) = \lambda_i e^{-\lambda_i t}$ por su acumulada correspondiente: $F_i(t) = 1 - e^{-\lambda_i t}$. Luego $F_{0,1,2}(t)$ se convierte en:

$$F_{0,1,2}(t) = \frac{\lambda_0 \frac{\lambda_1 F_2(t) - \lambda_2 F_1(t)}{\lambda_1 - \lambda_2} - \lambda_1 \frac{\lambda_0 F_2(t) - \lambda_2 F_0(t)}{\lambda_0 - \lambda_2}}{\lambda_0 - \lambda_1} \quad (47)$$

Dado que lo que interesa es el comportamiento de la suma en $t = 1$ resulta útil simplificar la notación llamando $F_i = F_i(1)$, y en cuanto a la suma, $F_{S_j} = F_{0,1,\dots,j}(1)$. Luego:

$$F_{S_2} = \frac{\lambda_0 \frac{\lambda_1 F_2 - \lambda_2 F_1}{\lambda_1 - \lambda_2} - \lambda_1 \frac{\lambda_0 F_2 - \lambda_2 F_0}{\lambda_0 - \lambda_2}}{\lambda_0 - \lambda_1} \quad (48)$$

Expresión que puede obtenerse mediante el siguiente algoritmo:

$$\Psi(x, y) = \frac{\lambda_y \Psi(x+1, x) - \lambda_x \Psi(x+1, y)}{\lambda_y - \lambda_x} \quad (49)$$

El algoritmo debe iniciarse siempre como $\Psi(1, 0)$ y (cuando se ha solicitado F_{S_j}) llega a la condición de parada la primera vez que se pide $\Psi(x, y)$ con $x > j$, devolviendo F_y . Una vez que ha sido explorado todo el árbol, $\Psi(1, 0)$ tiene el valor F_{S_j} . Por ejemplo, el cálculo para F_{S_2} es:

$$\Psi(1, 0) = \frac{\lambda_0 \Psi(2, 1) - \lambda_1 \Psi(2, 0)}{\lambda_0 - \lambda_1} \quad (50)$$

$$(51)$$

$$\Psi(2, 1) = \frac{\lambda_1 \Psi(3, 2) - \lambda_2 \Psi(3, 1)}{\lambda_1 - \lambda_2} = \frac{\lambda_1 F_2 - \lambda_2 F_1}{\lambda_1 - \lambda_2} \quad (52)$$

$$\Psi(2, 0) = \frac{\lambda_0 \Psi(3, 2) - \lambda_2 \Psi(3, 0)}{\lambda_0 - \lambda_2} = \frac{\lambda_0 F_2 - \lambda_2 F_0}{\lambda_0 - \lambda_2} \quad (53)$$

$$(54)$$

$$\Psi(1, 0) = \frac{\lambda_0 \frac{\lambda_1 F_2 - \lambda_2 F_1}{\lambda_1 - \lambda_2} - \lambda_1 \frac{\lambda_0 F_2 - \lambda_2 F_0}{\lambda_0 - \lambda_2}}{\lambda_0 - \lambda_1} = F_{S_3} \quad (55)$$

6 Entropía Cruzada

Tanto en *Monte Carlo Crudo* como en *Monte Carlo Permutación* es posible aplicar el *Muestreo de Importancia* [7] para cambiar las distribuciones de las cuales sortear, de tal forma que el muestreo a partir de las nuevas distribuciones resulte más eficiente. Existen diversos métodos para hacer el cambio de las distribuciones, uno de los cuales se conoce con el nombre de *Entropía Cruzada* [3, 4, 9, 10]. Antes de abordar esta metodología se procederá a revisar algunas cuestiones que tienen que ver con los principios básicos del muestreo de *Monte Carlo* y con el *Muestreo de Importancia*.

6.1 Muestreo de Importancia

Una de las áreas en la que los métodos de *Monte Carlo* son de gran utilidad es el cálculo de integrales. Un caso elemental (si no el más elemental), corresponde a la determinación de una integral definida de una función de una variable: $I_1 = \int_a^b g(x) dx$. Esta integral puede llevarse a las siguientes formas equivalentes:

$$I_1 = \int_a^b g(x) 1_{[a,b]} dx = (b-a) \int_a^b g(x) u_{ab}(x) dx = (b-a) \mathbb{E}_u \{g(X)\} \quad (56)$$

donde: $1_{[a,b]}$ es una función que vale 1 entre a y b , y 0 en el resto y $u_{ab}(x)$ es la expresión correspondiente a la distribución uniforme entre a y b o sea una función constante de valor $1/(b-a)$ entre a y b , y 0 en el resto. $\mathbb{E}_u \{g(X)\}$ indica la esperanza de $g(X)$ cuando X es una variable aleatoria distribuida uniformemente (de ahí el subíndice u). Luego, el valor I_1 se puede estimar haciendo una estimación del valor medio o esperado de la función $g(X)$ a partir de $u_{ab}(x)$, para lo cual es preciso sortear valores $x^{(i)}$ uniformemente distribuidos entre a y b y promediar los valores $g(x^{(i)})$:

$$\hat{I}_1 = (b-a) \frac{1}{N} \sum_{i=1}^N g(x^{(i)}) \quad (57)$$

Lo interesante de este análisis está en la resolución de la integral y su transformación en el promedio de valores de $g(x)$ tomados de la distribución uniforme. Pueden distinguirse dos partes en el integrando de la expresión (56), cada una de las cuales cumple una función perfectamente identificable. Una, la distribución (uniforme, en este caso), que genera valores, y la otra ($g(x)$, en este caso) que es valorizada en los valores generados. Ese es el mecanismo

a observar, mecanismo también aplicable en la determinación por *Monte Carlo* del valor medio o esperado de una variable aleatoria X , distribuida según cualquier distribución $f(x)$:

$$I_2 = \int_{-\infty}^{+\infty} x f(x) dx \quad (58)$$

Se sabe que $\mathbb{E}_f\{X\} = I_2$ se puede estimar mediante una expresión similar a (57):

$$\hat{I}_2 = \frac{1}{N} \sum_{i=1}^N x^{(i)} \quad (59)$$

sólo que las muestras $x^{(i)}$ se sortean de la distribución $f(x)$ y la función $g(x)$ ahora es directamente la función x .

Por último, un caso más general: el cálculo del valor esperado de una función de variable aleatoria $g(X)$ cuando la variable X se distribuye según una determinada $f(x)$:

$$I_3 = \int_{-\infty}^{+\infty} g(x) f(x) dx = \mathbb{E}_f\{g(X)\} \quad (60)$$

Con base en las ideas precedentes la integral, o lo que es lo mismo el valor esperado de $g(X)$, se puede estimar como:

$$\hat{I}_3 = \frac{1}{N} \sum_{i=1}^N g(x^{(i)}) \quad (61)$$

donde las muestras $x^{(i)}$ se sortean de la distribución $f(x)$.

Queda entonces aclarado el mecanismo de resolución de estas integrales que caen dentro de la categoría conocida como integral de Lebesgue-Stieltjes [7]. Resumiendo, si es posible detectar que el integrando está formado por el producto de dos funciones, una de ellas ($g(x)$) “cualquiera” y la otra ($f(x)$) una función de distribución de probabilidad ($\int_{-\infty}^{+\infty} f(x) dx = 1$), la estimación por *Monte Carlo* de la integral puede hacerse sorteando valores $x^{(i)}$ a partir de $f(x)$ y promediando los valores $g(x^{(i)})$. Véase que, si bien desde el punto de vista del cálculo ambas funciones cumplen la misma función, es decir, si la determinación de I_3 mediante el cálculo ordinario es factible, tanto $f(x)$ como $g(x)$ tienen el mismo significado (ambas forman parte del integrando y son tratadas por igual). Pero en la estimación por *Monte Carlo* los roles de ambas funciones son esencialmente diferentes: desde la función que es posible hacerlo, se sortean valores que se promedian valorizados en la otra función.

Un caso muy particular es aquel en que tanto $f(x)$ como $g(x)$ son distribuciones de probabilidad. En ese caso ambas funciones pueden cumplir cualquiera de los dos roles. Surge entonces una interesante observación: en cualquiera de ambos casos \hat{I}_3 resultará un estimador insesgado de I_3 ya que $\mathbb{E}\{\hat{I}_3\} = I_3$, donde \mathbb{E} puede ser tanto \mathbb{E}_f como \mathbb{E}_g , pero no puede decirse lo mismo acerca de otros parámetros del estimador como por ejemplo su varianza. A través de un análisis informal supóngase, por ejemplo, que $f(x)$ asume valores “grandes” donde $g(x)$ asume valores “pequeños” y viceversa. Entre los sumandos de la expresión (61) aparecerán entonces “muchos” valores “pequeños” y pocos valores “grandes” de $g(x^{(i)})$. Pero si se invierten los roles y los valores se sortean de $g(x)$, en la suma de (61) habrá “muchos” valores “pequeños” y pocos valores “grandes” de $f(x^{(i)})$. Sería muy casual que en ambas variantes los valores de $g(x^{(i)})$ y de $f(x^{(i)})$ estén dispersos de un modo tal que el cálculo de la varianza de \hat{I}_3 arroje exactamente el mismo valor. Este hecho no puede afirmarse sino que en cambio cabe esperar distintos valores de esa varianza. Consecuentemente una de las dos elecciones resultará más favorable que la otra para hacer la estimación.

Los ejemplos precedentes tienen por finalidad ilustrar el mecanismo de estimación de las integrales a través del papel que desempeñan las funciones del integrando y también presentar la posibilidad de encontrar soluciones eficientes haciendo elecciones y/o modificaciones apropiadas. Tal es el caso del *Muestreo de Importancia*[7] cuya idea central es la siguiente. Supóngase que la estimación de I_3 indicada en (61) no es eficiente, ya sea porque es difícil sortear muestras de $f(x)$ o porque $f(x)$ genera “muchos” valores en zonas del dominio de $g(x)$ donde no son “muy” necesarios y “pocos” valores en zonas donde sí lo son, razón por la cual la varianza de \hat{I}_3 es grande. La solución propuesta por el *Muestreo de Importancia* pasa por elegir una nueva distribución $h(x)$, tal que no valga 0 salvo en los puntos para los cuales $f(x)$ vale 0, y transformar la expresión (60) de la siguiente forma:

$$\mathbb{E}_f\{g(X)\} = \int_{-\infty}^{+\infty} g(x) f(x) dx = \int_{-\infty}^{+\infty} g(x) \frac{f(x)}{h(x)} h(x) dx = \mathbb{E}_h \left\{ \frac{g(x)f(x)}{h(x)} \right\} \quad (62)$$

Con esta modificación la integral puede estimarse como:

$$\hat{I}_3 = \frac{1}{N} \sum_{i=1}^N \frac{g(x^{(i)}) f(x^{(i)})}{h(x^{(i)})} \quad (63)$$

donde las muestras $x^{(i)}$ se sortean de la distribución $h(x)$.

Llamando *Función de Importancia* a $R(x) = f(x)/h(x)$, las expresiones del cálculo de la integral y de su correspondiente estimador se convierten en:

$$I_3 = \int_{-\infty}^{+\infty} g(x) R(x) h(x) dx = \mathbb{E}_h \{g(x) R(x)\} \quad \hat{I}_3 = \frac{1}{N} \sum_{i=1}^N g(x^{(i)}) R(x^{(i)}) \quad (64)$$

Dada la coincidencia (ya comentada) entre el valor de la integral I_3 y el cálculo de una esperanza ($\mathbb{E}_f\{g(X)\}$ o $\mathbb{E}_h\{R(x) g(x)\}$, según el caso), es posible identificar también una varianza en cada uno de los cálculos. Si bien la varianza así determinada no coincidirá numéricamente con la del estimador correspondiente, sí estará estrechamente relacionada y reflejará el comportamiento de esta última.

$$\mathbb{V}_f\{g(X)\} = \int_{-\infty}^{+\infty} g(x)^2 f(x) dx - \mathbb{E}_f^2\{g(X)\} \quad (65)$$

$$\mathbb{V}_h\{g(X) R(X)\} = \int_{-\infty}^{+\infty} g(x)^2 R(x)^2 h(x) dx - \mathbb{E}_h^2\{g(X) R(x)\} \quad (66)$$

Dado que por cualquiera de los dos caminos el valor esperado es el mismo, la diferencia entre varianzas es la diferencia entre los momentos de orden 2:

$$\mathbb{V}_f\{g(X)\} - \mathbb{V}_h\{g(X) R(X)\} = \int_{-\infty}^{+\infty} g(x)^2 f(x) dx - \int_{-\infty}^{+\infty} g(x)^2 R(x)^2 h(x) dx \quad (67)$$

$$= \int_{-\infty}^{+\infty} [1 - R(x)] g(x)^2 f(x) dx \quad (68)$$

Para lograr una reducción efectiva de la varianza del estimador es necesario entonces encontrar una $R(x)$ que haga de (68) un valor positivo.

Observando la expresión (66) resulta evidente que si:

$$h(x) = \frac{g(x) f(x)}{\mathbb{E}_f\{g(X)\}} \quad (69)$$

entonces $\mathbb{V}_h\{g(X) R(X)\} = 0$. Desde luego se trata de un resultado imposible, dado que la $h(x)$ propuesta contiene en su definición a $\mathbb{E}_f\{g(X)\}$, que es justamente el valor que permite estimar. De todas maneras, la expresión (69) es interesante y servirá (más adelante), no para ser aplicada rigurosamente sino para intentar encontrar, dentro de una familia de funciones candidatas, la que más se le parezca.

En definitiva la elección de $R(x)$ (o de $h(x)$) permite transformar el problema de estimar I_3 promediando valores de $g(x)$ sorteados de $f(x)$ en el promedio de valores de $R(x)g(x)$ sorteados de $h(x)$. La elección de $h(x)$ no es un problema trivial, nótese que esta función forma parte tanto de la generación de muestras como de la valoración de las muestras generadas, es decir participa de los dos roles. Se puede, no obstante hacer algunas observaciones acerca de la forma de elegir $h(x)$. Por ejemplo, la expresión (68) puede verse como el cálculo de una esperanza, concretamente la esperanza de la función $[1 - R(x)] g(x)^2$ a partir de la distribución $f(x)$. Si $[1 - R(x)] g(x)^2$ es no negativa para todos los valores que pueda generar $f(x)$ la esperanza resultará no negativa. Dado que $g(x)^2 \geq 0$, una condición suficiente para que la esperanza en cuestión resulte no negativa es: $[1 - R(x)] \geq 0$, o $R(x) \leq 1$ o, en definitiva, $h(x) \geq f(x)$. Ahora bien, tanto $f(x)$ como $h(x)$ son distribuciones de probabilidad (siempre positivas), por lo que si una de ellas resulta mayor o igual que la otra en todo su dominio, su integral en $(-\infty, +\infty)$ dará mayor o igual a 1!. La condición $h(x) \geq f(x)$ es entonces suficiente pero no necesaria y desde luego imposible de cumplir sobre todo el dominio de las funciones. Sólo se podrá aspirar a que se verifique en determinadas zonas, en particular en aquellas en las que $g(x)^2 f(x)$ asuma valores "grandes" (para darle más peso a los valores positivos en la integral (68)) siendo tolerable la inversa, es decir $h(x) < f(x)$, en el resto o sea en las zonas donde $g(x)^2 f(x)$ asume valores "bajos". A través de un análisis más detallado se demuestra en [7] que la reducción de varianza se alcanza cuando se verifica la condición:

$$h(x) \geq \frac{g(x) - G_I}{G_S - G_I} f(x) \quad (70)$$

donde G_I y G_S son cotas inferior y superior de $g(x)$ respectivamente, es decir $G_I \leq g(x) \leq G_S$.

EJEMPLO:

$$\text{Sean } f(x) = e^{-x} \text{ con } x \geq 0 \text{ y } g(x) = \begin{cases} 0 & \text{si } x < T \\ 1 & \text{si } x \geq T \end{cases} \text{ con } T > 0.$$

La integral

$$I = \int_{-\infty}^{+\infty} g(x) f(x) dx = \int_T^{+\infty} 1 e^{-x} dx$$

puede estimarse por *Monte Carlo* como:

$$\hat{I} = \frac{1}{N} \sum_{i=1}^N g(x^{(i)})$$

sorteando los valores $x^{(i)}$ de la distribución e^{-x} . Pero existe un inconveniente, puesto en evidencia en la Figura 2(a). Allí se muestra el caso $T = 5$ y se puede apreciar que para valores de T del orden de 5 o superiores es extremadamente bajo el número de muestras que darán $g(x^{(i)}) = 1$. En la sumatoria la inmensa mayoría de los sumandos será 0. Estimaciones realizadas con “pocas” muestras darán entonces resultados muy alejados de la realidad siendo necesarias “muchas” muestras para obtener resultados de una precisión razonable. Las estimaciones serán, en general, dispersas lo que implica que la varianza del estimador será elevada.

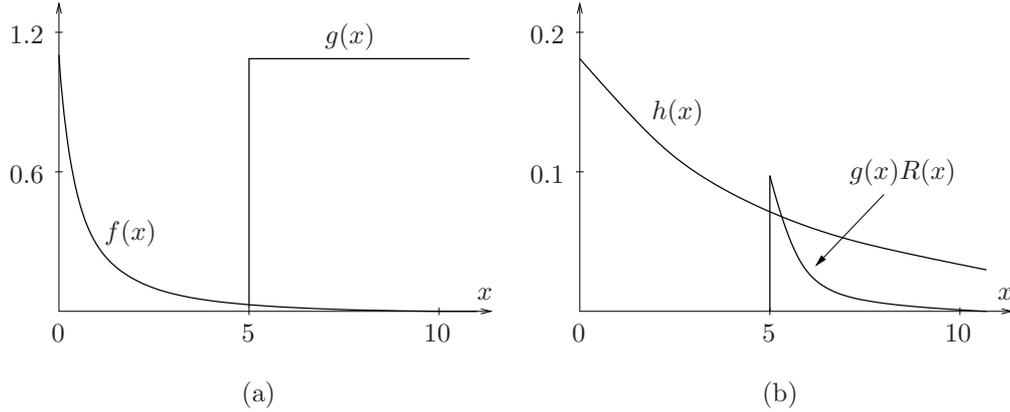


Figura 2: Ejemplo de Aplicación de *Muestreo de Importancia*

Se propone el *Muestreo de Importancia* con: $h(x) = b e^{-bx}$, $x \geq 0$ y $0 < b \leq 1$. El problema se convierte entonces en:

$$I = \int_T^{+\infty} \underbrace{\frac{e^{(b-1)x}}{b}}_{g(x)R(x)} \underbrace{b e^{-bx}}_{h(x)} dx$$

La Figura 2(b) muestra, para el caso $T = 5$ y $b = 0.18$, la distribución $h(x)$ desde la que se sorteán valores, y la función $g(x)R(x)$ a ser evaluada en los valores sorteados. Se aprecia claramente que el dominio de la función a evaluar no presenta sectores perjudicados que reciban muy escasa cantidad de muestras.

El ejemplo es tan simple que además puede resolverse analíticamente en forma exacta. En tal sentido, antes del *Muestreo de Importancia*:

$$\begin{aligned} \mathbb{E}_f\{g(X)\} &= e^{-T} \\ \mathbb{V}_f\{g(X)\} &= e^{-T}(1 - e^{-T}) \end{aligned}$$

y después del *Muestreo de Importancia*:

$$\begin{aligned} \mathbb{E}_h\{R(X) g(X)\} &= e^{-T} \\ \mathbb{V}_h\{g(X) R(X)\} &= \frac{e^{-T(2-b)}}{b(2-b)} - e^{-2T} \end{aligned}$$

La máxima reducción posible de varianza para la $h(x)$ propuesta se alcanza haciendo mínimo el valor de $\mathbb{V}_h\{g(X) R(X)\}$, cosa que sucede para $b = \tilde{b}$:

$$\tilde{b} = 1 + \frac{1}{T} - \sqrt{1 + \frac{1}{T^2}}$$

que para $T = 5$ produce $\tilde{b} = 0.18$, que es el valor utilizado en la Figura 2(b).

Por último cabe acotar que todas las consideraciones presentadas acerca del muestreo, en particular el *Muestreo de Importancia*, valen para el caso de variables de mayor orden, es decir vectores, y distribuciones de probabilidad conjuntas en lugar de funciones de distribución de una única variable.

6.2 Entropía Cruzada en la Estimación de Eventos Raros

Sea $\mathbf{X} = (X_1, X_2, \dots, X_m)$ el vector aleatorio correspondiente al estado de los enlaces de la red y sea $f(\mathbf{x}, \mathbf{v})$ una familia de densidades de probabilidad sobre el dominio de \mathbf{X} , caracterizadas por el valor de un parámetro real \mathbf{v} , también de dimensión m . Supóngase que interesa conocer la probabilidad ℓ de que cierta función real $S(\mathbf{X})$ supere determinado valor γ cuando X sigue la distribución $f(\mathbf{x}, \mathbf{u})$, siendo \mathbf{u} un valor determinado del parámetro \mathbf{v} :

$$\ell = \mathbb{P}_{\mathbf{u}}[S(\mathbf{X}) \geq \gamma] \quad (71)$$

Si el valor de ℓ es muy bajo, se dice que $S(\mathbf{X}) \geq \gamma$ es un evento raro. Para su tratamiento resulta más apropiado presentar ℓ en términos de una esperanza. Se define entonces la función $I_{\{S(\mathbf{x}) \geq \gamma\}}$, que vale 1 si $S(\mathbf{X}) \geq \gamma$ y 0 en caso contrario. Entonces:

$$\ell = \mathbb{E}_{\mathbf{u}}\{I_{\{S(\mathbf{x}) \geq \gamma\}}\} \quad (72)$$

Resulta inmediato obtener un estimador insesgado de ℓ aplicando el método de *Monte Carlo Crudo*:

$$\hat{\ell}_{\mathbf{u}} = \frac{1}{N} \sum_{i=1}^N I_{\{S(\mathbf{x}^{(i)}) \geq \gamma\}} \quad (73)$$

donde las muestras $\mathbf{X}^{(i)}$ son sorteadas de la distribución $f(\mathbf{x}, \mathbf{u})$. El hecho es que la inmensa mayoría de los sumandos serán 0, siendo necesario un valor de N muy grande para obtener una estimación confiable. Una mejora a través del *Muestreo de Importancia* es:

$$\hat{\ell}_{\mathbf{u},g} = \frac{1}{N} \sum_{i=1}^N I_{\{S(\mathbf{x}^{(i)}) \geq \gamma\}} \frac{f(\mathbf{X}^{(i)}, \mathbf{u})}{g(\mathbf{X}^{(i)})} \quad (74)$$

donde las muestras $\mathbf{X}^{(i)}$ ya no se sortean de $f(\mathbf{x}, \mathbf{u})$ sino de $g(\mathbf{x})$. Retomando las ideas comentadas acerca de la expresión (69), la elección ideal de $g(\mathbf{x})$ sería:

$$g^*(\mathbf{x}) = \frac{I_{\{S(\mathbf{x}) \geq \gamma\}} f(\mathbf{x}, \mathbf{u})}{\ell} \quad (75)$$

Dada la imposibilidad de esta elección, se propone como alternativa una de las distribuciones de la familia $f(\mathbf{x}, \mathbf{v})$, con un nuevo valor del parámetro, esto es: $f(\mathbf{x}, \mathbf{v}^*)$.

El problema se convierte entonces en determinar \mathbf{v}^* , o sea decidir cuál de todas las distribuciones de la familia de candidatas $f(\mathbf{x}, \mathbf{v})$ es la que más se parece a $g^*(\mathbf{x})$. Para ello es preciso contar con una medida de distancia entre distribuciones y elegir el valor \mathbf{v}^* que identifique, entre todas las f , aquella que esté a la mínima distancia de $g^*(\mathbf{x})$. Una medida de utilidad es la *Distancia de Kullback-Leibler*, también conocida como *Entropía Cruzada*. Según esta definición, la distancia $\mathcal{D}(d_1, d_2)$ entre dos densidades de probabilidad $d_1(\mathbf{X})$ y $d_2(\mathbf{X})$ está dada por:

$$\mathcal{D}(d_1, d_2) = \mathbb{E}_{d_1} \left\{ \ln \frac{d_1(\mathbf{X})}{d_2(\mathbf{X})} \right\} \quad (76)$$

$$= \mathbb{E}_{d_1} \{ \ln d_1(\mathbf{X}) \} - \mathbb{E}_{d_1} \{ \ln d_2(\mathbf{X}) \} \quad (77)$$

$$= \int \ln d_1(\mathbf{x}) d_1(\mathbf{x}) d\mathbf{x} - \int \ln d_2(\mathbf{x}) d_1(\mathbf{x}) d\mathbf{x} \quad (78)$$

Una forma apropiada de elegir d_1 y d_2 (nótese que \mathcal{D} no es simétrica: $\mathcal{D}(d_1, d_2) \neq \mathcal{D}(d_2, d_1)$) es: $d_1(\mathbf{x}) = g^*(\mathbf{x})$ y $d_2(\mathbf{x}) = f(\mathbf{x}, \mathbf{v})$, con lo que:

$$\mathcal{D}(d_1, d_2) = \int \ln g^*(\mathbf{x}) g^*(\mathbf{x}) d\mathbf{x} - \int \ln f(\mathbf{x}, \mathbf{v}) g^*(\mathbf{x}) d\mathbf{x} \quad (79)$$

El primer término no depende de \mathbf{v} , por lo que la mínima distancia $\mathcal{D}(d_1, d_2)$, según \mathbf{v} , corresponde a:

$$\min_{\mathbf{v}} \mathcal{D}(d_1, d_2) = \max_{\mathbf{v}} \int \ln f(\mathbf{x}, \mathbf{v}) g^*(\mathbf{x}) d\mathbf{x} \quad (80)$$

$$= \max_{\mathbf{v}} \int \ln f(\mathbf{x}, \mathbf{v}) \frac{I_{\{S(\mathbf{x}) \geq \gamma\}} f(\mathbf{x}, \mathbf{u})}{\ell} d\mathbf{x} \quad (81)$$

$$= \max_{\mathbf{v}} \mathbb{E}_{\mathbf{u}} \{ I_{\{S(\mathbf{x}) \geq \gamma\}} \ln f(\mathbf{X}, \mathbf{v}) \} \quad (82)$$

Es posible elegir una distribución $f(\mathbf{x}, \mathbf{w})$ más favorable que $f(\mathbf{x}, \mathbf{u})$ para aplicar *Muestreo de Importancia* al cálculo de la \mathbb{E} . Siendo entonces $W(\mathbf{x}, \mathbf{u}, \mathbf{w}) = f(\mathbf{x}, \mathbf{u})/f(\mathbf{x}, \mathbf{w})$:

$$\min_{\mathbf{v}} \mathcal{D}(d_1, d_2) = \max_{\mathbf{w}} \mathbb{E}_{\mathbf{u}} \{ I_{\{S(\mathbf{x}) \geq \gamma\}} W(\mathbf{X}, \mathbf{u}, \mathbf{w}) \ln f(\mathbf{X}, \mathbf{v}) \} \quad (83)$$

Pero lo que interesa no es el máximo valor de $\mathbb{E}_{\mathbf{w}}$ sino el argumento \mathbf{v}^* que da lugar al mismo:

$$\mathbf{v}^* = \underset{\mathbf{v}}{\operatorname{argmax}} \mathbb{E}_{\mathbf{w}} \{ I_{\{S(\mathbf{x})\} \geq \gamma} W(\mathbf{X}, \mathbf{u}, \mathbf{w}) \ln f(\mathbf{X}, \mathbf{v}) \} \quad (84)$$

La \mathbb{E} de la expresión (83) puede estimarse, con lo que resulta factible también una estimación de \mathbf{v}^* :

$$\hat{\mathbf{v}}^* = \underset{\mathbf{v}}{\operatorname{argmax}} \frac{1}{N} \sum_{i=1}^N I_{\{S(\mathbf{X}^{(i)})\} \geq \gamma} W(\mathbf{X}^{(i)}, \mathbf{u}, \mathbf{w}) \ln f(\mathbf{X}^{(i)}, \mathbf{v}) \quad (85)$$

donde las muestras $\mathbf{X}^{(i)}$, $i = 1, 2, \dots, N$ se sortean de la distribución $f(\mathbf{x}, \mathbf{w})$.

No obstante en muchos casos (por ejemplo cuando las f pertenecen a familias de distribuciones exponenciales) la determinación de \mathbf{v}^* , pueden resolverse analíticamente.

Existe también un proceso iterativo para estimar \mathbf{v}^* :

$$\mathbf{v}_k = \underset{\mathbf{v}}{\operatorname{argmax}} \sum_{i=1}^N I_{\{S(\mathbf{X}^{(i)})\} \geq \gamma} W(\mathbf{X}^{(i)}, \mathbf{u}, \mathbf{v}_{k-1}) \ln f(\mathbf{X}^{(i)}, \mathbf{v}) \quad (86)$$

El proceso se inicia con la elección $\mathbf{v}_0 = \mathbf{u}$, y las muestras $\mathbf{X}^{(i)}$, $i = 1, 2, \dots, N$ se sortean, cada vez, de la distribución $f(\mathbf{x}, \mathbf{v}_k)$. El cálculo de \mathbf{v}_1 en (86) se hace entonces sorteando valores de \mathbf{X} a partir de $f(\mathbf{x}, \mathbf{u})$ por lo que si el evento $S(\mathbf{X}) \geq \gamma$ es “muy” raro, o sea la probabilidad de que ocurra es extremadamente baja, en la primera iteración la inmensa mayoría de los sumandos tendrán valor 0. Una variante para salvar este inconveniente es iterar también sobre el valor de γ . Para ello el evento se convierte en $S(\mathbf{X}) \geq \gamma_k$, y γ_k se va actualizando en cada iteración, en la expectativa de que los primeros valores hagan de $S(\mathbf{X}) \geq \gamma_k$ un evento no tan raro y de que, finalmente, γ_k converja a γ . La idea es, bajo cada \mathbf{v}_{k-1} , elegir como γ_k al $(1 - \rho)$ cuantil⁶ de $S(\mathbf{X})$ (es decir un valor tal que $\mathbb{P}_{\mathbf{v}_{k-1}}[S(\mathbf{X}) \geq \gamma_k] \geq \rho$ y $\mathbb{P}_{\mathbf{v}_{k-1}}[S(\mathbf{X}) \leq \gamma_k] \geq (1 - \rho)$). Este valor puede estimarse de la siguiente forma: en cada iteración se ordenan los N valores de $S(\mathbf{X}^{(i)})$, de menor a mayor, esto es $S_1 \leq S_2 \leq \dots \leq S_N$ y se elige $\gamma_k = S_{\lceil (1-\rho)N \rceil}$ donde $\lceil (1 - \rho)N \rceil$ indica la parte entera de $(1 - \rho)N$, a menos que $\gamma_k > \gamma$, en cuyo caso se elige $\gamma_k = \gamma$.

Queda definido entonces un mecanismo más o menos general para aplicar el criterio de *Entropía Cruzada* a la estimación de eventos raros:

1. Setear $k = 1$ y $\mathbf{v}_{k-1} = \mathbf{u}$.
2. Sortear N valores $\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \dots, \mathbf{X}^{(N)}$ a partir de $f(\mathbf{x}, \mathbf{v}_{k-1})$ y con estas muestras determinar el valor de γ_k .
3. Mediante la expresión (86) reemplazando γ por γ_k y utilizando el mismo juego de muestras para todas las componentes, o a través de una determinación analítica si el problema lo permite, calcular \mathbf{v}_k .
4. Si $\gamma_k < \gamma$ hacer $k = k + 1$ y volver al punto 2. Caso contrario ir al punto 5.
5. Estimar la probabilidad de evento raro ℓ como:

$$\hat{\ell} = \frac{1}{M} \sum_{i=1}^M I_{\{S(\mathbf{X}^{(i)})\} \geq \gamma} \frac{f(\mathbf{X}^{(i)}, \mathbf{u})}{f(\mathbf{X}^{(i)}, \mathbf{v}_K)}$$

donde K es el numero total de iteraciones requeridas, es decir el último valor de k y M el número total de muestras elegido para realizar la estimación. Las muestras $\mathbf{X}^{(i)}$ se sortean de $f(\mathbf{x}, \mathbf{v}_K)$.

6.3 Entropía Cruzada en la Estimación de una Esperanza

En la expresión (72) de la Sección 6.2 se plantea el cálculo de la esperanza de una variable aleatoria cuyo valor se sabe a priori es muy próximo a 0. De ahí en más, todo el desarrollo del método de *Entropía Cruzada* apunta a resolver ese cálculo en forma eficiente. Obviando la parte concerniente a γ y a la forma de proceder mediante la variable auxiliar γ_k , es totalmente lícito reemplazar, en el planteo inicial, la variable $I_{\{S(\mathbf{x})\} \geq \gamma}$ por una función cualquiera de \mathbf{X} , por ejemplo $F(\mathbf{X})$ y avanzar en los cálculos hacia la determinación de su esperanza (sin perder de vista que se trata de un valor muy bajo). El cálculo de ℓ queda convertido entonces en:

$$\ell = \mathbb{E}_{\mathbf{u}} \{ F(\mathbf{X}) \} \quad (87)$$

y los resultados del método, por ejemplo los correspondientes a las expresiones (85) y (86), en:

$$\hat{\mathbf{v}}^* = \underset{\mathbf{v}}{\operatorname{argmax}} \frac{1}{N} \sum_{i=1}^N F(\mathbf{X}^{(i)}) W(\mathbf{X}^{(i)}, \mathbf{u}, \mathbf{w}) \ln f(\mathbf{X}^{(i)}, \mathbf{v}) \quad (88)$$

$$\mathbf{v}_k = \underset{\mathbf{v}}{\operatorname{argmax}} \sum_{i=1}^N F(\mathbf{X}^{(i)}) W(\mathbf{X}^{(i)}, \mathbf{u}, \mathbf{v}_{k-1}) \ln f(\mathbf{X}^{(i)}, \mathbf{v}) \quad (89)$$

⁶Se debe elegir un valor de $\rho < 1$, no muy pequeño, en el orden de 10^{-2} .

La estimación de ℓ se resuelve entonces mediante le siguiente mecanismo:

1. Setear $k = 1$ y $\mathbf{v}_{k-1} = \mathbf{u}$.
2. Mediante la expresión (89), utilizando el mismo juego de muestras para todas las componentes, o a través de una determinación analítica si el problema lo permite, calcular \mathbf{v}_k .
3. Si los valores de \mathbf{v}_k no están estabilizados hacer $k = k + 1$ y volver al punto 2. Caso contrario ir al punto 4.
4. Estimar la esperanza de $F(\mathbf{X})$ como:

$$\hat{\ell} = \frac{1}{M} \sum_{i=1}^M F(\mathbf{X}^{(i)}) \frac{f(\mathbf{X}^{(i)}, \mathbf{u})}{f(\mathbf{X}^{(i)}, \mathbf{v}_K)}$$

donde K es el numero total de iteraciones requeridas, es decir el último valor de k y M el número total de muestras elegido para realizar la estimación. Las muestras $\mathbf{X}^{(i)}$ se sortean de $f(\mathbf{x}, \mathbf{v}_K)$.

6.4 Entropía Cruzada sobre Monte Carlo Crudo

Retomando las ideas de la Sección 5, supóngase que al comienzo ($t = 0$) todos los enlaces están *fallados* y que van pasando al estado *operativo* en instantes $\tau(i)$, $i = 1, 2, \dots, m$, distribuidos exponencialmente con tasa $\lambda(i)$, esto es: $\mathbb{P}[\tau(i) \leq t] = 1 - e^{-\lambda(i)t}$. Supóngase además que los enlaces evolucionan en forma mutuamente independiente y que $\lambda(i) = -\log(q_i)$, con lo que en el instante $t = 1$ la probabilidad de que el i -ésimo enlace esté *operativo* es exactamente r_i , esto es:

$$r_i = \mathbb{P}[\tau(i) \leq 1] \quad (90)$$

Siguiendo la misma línea de razonamiento, la confiabilidad R de la red es (numéricamente) igual a la probabilidad de que la misma esté en un estado de de operatividad aceptable en $t = 1$. Llamando S al instante en que la red se pone *operativa*, $R = \mathbb{P}[S \leq 1]$. Pero el instante S depende del instante de cambio de estado de cada uno de los enlaces, luego $S = S(\mathbf{T})$, donde $\mathbf{T} = (\tau(1), \tau(2), \dots, \tau(m))$. Finalmente:

$$R = \mathbb{P}[S(\mathbf{T}) \leq 1] \quad Q = \mathbb{P}[S(\mathbf{T}) > 1] \quad (91)$$

La variable aleatoria \mathbf{T} , indicadora de los tiempos de conmutación de los m enlaces, sigue una distribución exponencial $f(\mathbf{t}, \mathbf{u})$, donde el parámetro \mathbf{u} es el conjunto de las m medias. Si la red es altamente confiable, Q será un valor extremadamente bajo y, $S(\mathbf{T}) > 1$ un evento raro. El problema se presta entonces para ser resuelto mediante el mecanismo de 5 puntos presentado en la Sección 6.2 donde:

$$\begin{aligned} \ell &= R \\ \mathbf{u} &= (1/\lambda(1), 1/\lambda(2), \dots, 1/\lambda(m)) \\ \mathbf{X} &= \mathbf{T} \\ \gamma &= 1 \end{aligned}$$

El procedimiento se convierte entonces en:

1. Setear $k = 1$ y $\mathbf{v}_{k-1} = \mathbf{u} = (1/\lambda(1), 1/\lambda(2), \dots, 1/\lambda(m))$.
2. Sortear N valores $\mathbf{T}^{(1)}, \mathbf{T}^{(2)}, \dots, \mathbf{T}^{(N)}$ a partir de $f(\mathbf{t}, \mathbf{v}_{k-1})$ y con estas muestras determinar el valor de γ_k .
3. Utilizando el mismo juego de muestras para todas las componentes, calcular \mathbf{v}_k mediante la siguiente expresión exacta:

$$v_{k,j} = \frac{\sum_{i=1}^N I_{\{S(\mathbf{T}^{(i)}) \geq \gamma_k\}} W(\mathbf{T}^{(i)}; \mathbf{u}, \mathbf{v}_{k-1}) \tau(j)^{(i)}}{\sum_{i=1}^N I_{\{S(\mathbf{T}^{(i)}) \geq \gamma_k\}} W(\mathbf{T}^{(i)}; \mathbf{u}, \mathbf{v}_{k-1})}$$

donde $\mathbf{v}_k = (v_{k,1}, v_{k,2}, \dots, v_{k,m})$ y:

$$W(\mathbf{t}; \mathbf{u}, \mathbf{v}) = \frac{f(\mathbf{t}, \mathbf{u})}{f(\mathbf{t}, \mathbf{v})} = \exp \left(- \sum_{j=1}^m \tau(j) \left(\frac{1}{u_j} - \frac{1}{v_j} \right) \right) \prod_{j=1}^m \frac{v_j}{u_j}$$

4. Si $\gamma_k < 1$ hacer $k = k + 1$ y volver al punto 2. Caso contrario ir al punto 5.
5. Estimar la probabilidad de evento raro Q como:

$$Q = \frac{1}{M} \sum_{i=1}^M I_{\{S(\mathbf{T}^{(i)}) > 1\}} W(\mathbf{T}^{(i)}; \mathbf{u}, \mathbf{v}_K)$$

donde K es el numero total de iteraciones requeridas, es decir el último valor de k y M el número total de muestras elegido para realizar la estimación. Las muestras $\mathbf{T}^{(i)}$ se sortean de $f(\mathbf{t}, \mathbf{v}_K)$.

6.5 Entropía Cruzada sobre Monte Carlo Permutación

Aplicando el método de *Monte Carlo Permutación*, la confiabilidad R de la red se determina a partir de la expresión:

$$R = \mathbb{E}\{G(\Omega)\} \quad (92)$$

donde la función $G(\Omega)$ indica la probabilidad de que, dada la permutación $\Omega = \omega$, la red esté *operativa* en $t = 1$, esto es:

$$G(\omega) = \mathbb{P}[\Phi(\mathbf{X}(1)) = 1 \mid \Omega = \omega] \quad (93)$$

Del mismo modo:

$$Q = \mathbb{E}\{\bar{G}(\Omega)\} \quad (94)$$

donde la función $\bar{G}(\Omega)$ indica la probabilidad de que, dada la permutación $\Omega = \omega$, la red no esté *operativa* en $t = 1$, esto es:

$$\bar{G}(\omega) = \mathbb{P}[\Phi(\mathbf{X}(1)) = 0 \mid \Omega = \omega] \quad (95)$$

Las permutaciones están determinadas por el orden en que se ponen operativos los enlaces, lo que en definitiva está relacionado con la familia de distribuciones exponenciales $f(\mathbf{t}, \mathbf{u})$ que regula el el cambio de estado de los mismos. Es posible explicitar esa relación diciendo que $\Omega = \Omega(\mathbf{T})$ con lo que (94) se convierte en:

$$Q = \mathbb{E}_{\mathbf{u}}\{\bar{G}(\Omega(\mathbf{T}))\} = \mathbb{E}_{\mathbf{u}}\{F(\mathbf{T})\} \quad (96)$$

donde $F(\mathbf{T}) = \bar{G}(\Omega(\mathbf{T}))$ y $\mathbf{T} = (\tau(1), \tau(2), \dots, \tau(m))$. El valor de Q puede estimarse entonces mediante el mecanismo presentado en la Sección 6.3 donde:

$$\begin{aligned} \ell &= Q \\ \mathbf{u} &= (1/\lambda(1), 1/\lambda(2), \dots, 1/\lambda(m)) \\ \mathbf{X} &= \mathbf{T} \end{aligned}$$

El procedimiento se convierte entonces en:

1. Setear $k = 1$ y $\mathbf{v}_{k-1} = \mathbf{u} = (1/\lambda(1), 1/\lambda(2), \dots, 1/\lambda(m))$.
2. Sortear N valores $\mathbf{T}^{(1)}, \mathbf{T}^{(2)}, \dots, \mathbf{T}^{(N)}$ a partir de $f(\mathbf{t}, \mathbf{v}_{k-1})$ y con estas muestras determinar \mathbf{v}_k mediante la siguiente expresión exacta:

$$v_{k,j} = \frac{\sum_{i=1}^N F(\mathbf{T}^{(i)}) W(\mathbf{T}^{(i)}; \mathbf{u}, \mathbf{v}_{k-1}) \tau(j)^{(i)}}{\sum_{i=1}^N F(\mathbf{T}^{(i)}) W(\mathbf{T}^{(i)}; \mathbf{u}, \mathbf{v}_{k-1})}$$

donde $\mathbf{v}_k = (v_{k,1}, v_{k,2}, \dots, v_{k,m})$ y:

$$W(\mathbf{t}; \mathbf{u}, \mathbf{v}) = \frac{f(\mathbf{t}, \mathbf{u})}{f(\mathbf{t}, \mathbf{v})} = \exp\left(-\sum_{j=1}^m \tau(j) \left(\frac{1}{u_j} - \frac{1}{v_j}\right)\right) \prod_{j=1}^m \frac{v_j}{u_j}$$

3. Si los valores de \mathbf{v}_k no están estabilizados hacer $k = k + 1$ y volver al punto 2. Caso contrario ir al punto 4.
4. Estimar Q , la esperanza de $F(\mathbf{T})$, como:

$$Q = \frac{1}{M} \sum_{i=1}^M F(\mathbf{T}^{(i)}) W(\mathbf{T}^{(i)}; \mathbf{u}, \mathbf{v}_K)$$

donde K es el numero total de iteraciones requeridas, es decir el último valor de k y M el número total de muestras elegido para realizar la estimación. Las muestras $\mathbf{T}^{(i)}$ se sortean de $f(\mathbf{t}, \mathbf{v}_K)$.

7 Resultados Numéricos

En esta sección se aplican los métodos de *Monte Carlo* presentados en secciones anteriores a la estimación y el análisis de la *Confiabilidad Diámetro Acotada* de las redes mostradas en la Figura 3: una topología triangular (TRI), una reja cuadrada (GRID) y una hexagonal (HEX). En todos los casos $\mathcal{K} = \mathcal{V}$. Las implementaciones se ajustan a las ideas discutidas en cada sección.

El primer conjunto de experimentos se realiza en tres etapas. En cada una se trabaja con un valor de confiabilidad individual por enlace (el mismo para cada enlace en las tres redes), y con el mismo valor de D en cada red. En todos los casos se utiliza un tamaño de muestra de 10^5 para hacer la estimación de anti-confiabilidad Q y de su correspondiente Desviación Estandar. Los resultados se presentan en las Tablas 1, 2 y 3. Tal como es de esperar, la peor estimación, en todos los casos, proviene de *Monte Carlo Crudo*. *Dagger* no reporta mejoras en la precisión sino que se mantiene

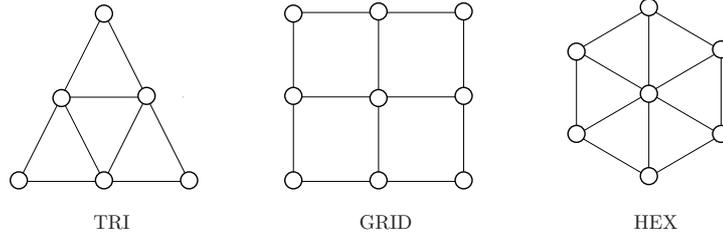


Figura 3: Topología de las Redes Bajo Análisis

en el mismo orden que *Monte Carlo Crudo*. La verdadera mejora aparece con *Monte Carlo Permutación*. La relación entre Desviaciones Estandar entre *Permutación* y *Crudo* está en el orden de 2 para confiabilidad individual por enlace de 0.80 y 20 para 0.99. *Entropía Cruzada* produce en ambos casos, *Crudo* y *Permutación*, una ligera mejora con respecto a las versiones sin *Entropía Cruzada*. Una prueba adicional, que no se muestra en las tablas, indica que, para un valor de confiabilidad individual por enlace de 0.90, la misma precisión se consigue con los siguientes ordenes de tamaño de muestra: *Monte Carlo Crudo*: 10^5 , *Dagger*: 10^5 , *Monte Carlo Permutación*: 2×10^3 , *EC-Monte Carlo Crudo*: 10^4 y *EC-Monte Carlo Permutación*: 2×10^3 .

	TRI $D = 3$	GRID $D = 5$	HEX $D = 3$
<i>MC Crudo</i>	1.91×10^{-1} (3.63×10^{-3})	3.71×10^{-1} (4.99×10^{-3})	1.72×10^{-1} (3.45×10^{-3})
<i>Dagger</i>	1.91×10^{-1} (3.63×10^{-3})	3.72×10^{-1} (5.00×10^{-3})	1.72×10^{-1} (3.45×10^{-3})
<i>MC Permutación</i>	1.92×10^{-1} (1.93×10^{-3})	3.72×10^{-1} (3.39×10^{-3})	1.73×10^{-1} (1.78×10^{-3})
<i>EC-MC Crudo</i>	1.91×10^{-1} (3.18×10^{-3})	3.71×10^{-1} (4.74×10^{-4})	1.72×10^{-1} (3.01×10^{-3})
<i>EC-MC Permut</i>	1.92×10^{-1} (1.92×10^{-3})	3.71×10^{-1} (3.39×10^{-3})	1.73×10^{-1} (1.76×10^{-3})

Tabla 1: Anti-confiabilidad Q (Desviación Estandar) - $r_i = 0.80$, $i = 1, 2, \dots, m$

	TRI $D = 3$	GRID $D = 5$	HEX $D = 3$
<i>MC Crudo</i>	4.31×10^{-2} (1.74×10^{-3})	1.10×10^{-1} (2.78×10^{-3})	2.93×10^{-2} (1.44×10^{-3})
<i>Dagger</i>	4.39×10^{-2} (1.76×10^{-3})	1.11×10^{-1} (2.79×10^{-3})	2.82×10^{-2} (1.41×10^{-3})
<i>MC Permutación</i>	4.42×10^{-2} (6.03×10^{-4})	1.12×10^{-1} (1.30×10^{-3})	2.88×10^{-2} (3.87×10^{-4})
<i>EC-MC Crudo</i>	4.35×10^{-2} (1.28×10^{-3})	1.11×10^{-1} (2.34×10^{-3})	2.84×10^{-2} (9.96×10^{-4})
<i>EC-MC Permut</i>	4.41×10^{-2} (5.72×10^{-4})	1.11×10^{-1} (1.31×10^{-3})	2.88×10^{-2} (3.82×10^{-4})

Tabla 2: Anti-confiabilidad Q (Desviación Estandar) - $r_i = 0.90$, $i = 1, 2, \dots, m$

	TRI $D = 3$	GRID $D = 5$	HEX $D = 3$
<i>MC Crudo</i>	3.00×10^{-4} (1.46×10^{-4})	1.37×10^{-3} (3.12×10^{-4})	4.00×10^{-5} (5.33×10^{-5})
<i>Dagger</i>	2.90×10^{-4} (1.43×10^{-4})	1.34×10^{-3} (3.08×10^{-4})	2.00×10^{-5} (3.77×10^{-5})
<i>MC Permutación</i>	3.19×10^{-4} (8.30×10^{-6})	1.19×10^{-3} (2.18×10^{-5})	3.49×10^{-5} (7.00×10^{-7})
<i>EC-MC Crudo</i>	3.44×10^{-4} (7.35×10^{-5})	1.19×10^{-3} (2.68×10^{-4})	2.40×10^{-5} (8.50×10^{-6})
<i>EC-MC Permut</i>	3.22×10^{-4} (7.30×10^{-6})	1.19×10^{-3} (2.20×10^{-5})	3.48×10^{-5} (7.00×10^{-7})

Tabla 3: Anti-confiabilidad Q (Desviación Estandar) - $r_i = 0.99$, $i = 1, 2, \dots, m$

El segundo conjunto de experimentos tiene que ver con el valor de D . Numéricamente el rango de valores para este parámetro es: $0 < D < \infty$. De todas maneras, desde un punto de vista práctico, deben considerarse límites diferentes: $D_d < D < D_u$. Para cada red, el valor de D_d es la distancia del camino más corto entre los dos nodos más alejados entre sí. Es claro que un camino de distancia 1 no garantiza la conectividad entre varios pares de nodos de TRI y HEX. De hecho un mínimo de 2 es necesario para estas dos redes. En el caso de GRID los nodos de los extremos (vértices del cuadrado) están a distancia 4 del opuesto (diagonal) siendo por lo tanto $D_d = 4$. Por otro lado

existe el camino más largo entre cualquier par de nodos de cada red (que será a los sumo de distancia $\mathcal{V} - 1$). Estos caminos se muestran en la Figura 4, correspondiendo valores de 5, 8 y 6 para TRI, GRID y HEX respectivamente. Si $D < D_d$, el valor de la anti-Confabilidad Diámetro Acotada Q , valdrá siempre 1.00 indicando una condición de falla permanente, cualquiera sea el conjunto de estados sorteado para los enlaces. Por otro lado, para cualquier valor $D > D_u$ el valor de Q será siempre el mismo.

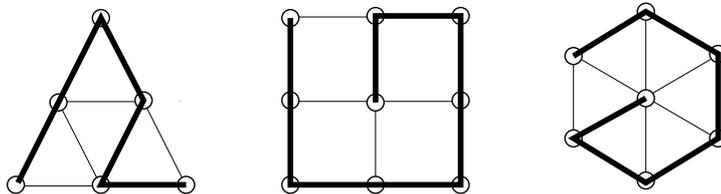


Figura 4: Ejemplos del Camino Más Largo Para Cada Red

Las Tablas 4 y 5 presentan las estimaciones de Q y de la Desviación Estandar para dos métodos de *Monte Carlo*, *Crudo* y *Permutación*. Las tres redes son evaluadas en el rango correspondiente de D de cada una, esto es $D_d < D < D_u$. Los resultados muestran que los valores de Q caen a medida que D crece. Esta propiedad es predecible para cualquier red porque cuando D crece el evento correspondiente a encontrar la red en estado de operatividad no aceptable se hace más raro. A la vez, a incrementos de D le corresponden decrementos de la Desviación Estandar, para ambos métodos. Estos decrementos son ligeramente más notorios para *Permutación* que para *Crudo*.

D	TRI	GRID	HEX
2	4.83×10^{-1} (5.65×10^{-3})		4.70×10^{-1} (5.58×10^{-3})
3	4.31×10^{-2} (1.74×10^{-3})		2.93×10^{-2} (1.44×10^{-3})
4	3.08×10^{-2} (1.47×10^{-3})	1.93×10^{-1} (3.65×10^{-3})	8.36×10^{-3} (7.70×10^{-4})
5	3.02×10^{-2} (1.46×10^{-3})	1.10×10^{-1} (2.78×10^{-3})	6.62×10^{-3} (6.85×10^{-4})
6		6.58×10^{-2} (2.15×10^{-3})	6.56×10^{-3} (6.82×10^{-4})
7		5.38×10^{-2} (1.95×10^{-3})	
8		5.28×10^{-2} (1.93×10^{-3})	

Tabla 4: Anti-confiabilidad Q (Desviación Estandar) – *MC Crudo*

D	TRI	GRID	HEX
2	4.84×10^{-1} (4.08×10^{-3})		4.69×10^{-1} (4.30×10^{-3})
3	4.42×10^{-2} (6.03×10^{-4})		2.88×10^{-2} (3.87×10^{-4})
4	3.13×10^{-2} (5.69×10^{-4})	1.94×10^{-1} (1.87×10^{-3})	8.28×10^{-3} (1.73×10^{-4})
5	3.06×10^{-2} (5.69×10^{-4})	1.12×10^{-1} (1.30×10^{-3})	6.52×10^{-3} (1.67×10^{-4})
6		6.64×10^{-2} (8.62×10^{-4})	6.49×10^{-3} (1.67×10^{-4})
7		5.40×10^{-2} (8.13×10^{-4})	
8		5.31×10^{-2} (8.13×10^{-4})	

Tabla 5: Anti-confiabilidad Q (Desviación Estandar) – *MC Permutación*

Para los experimentos consignados en las Tablas 1, 2 y 3 se registraron tiempos de ejecución (no mostrados en las tablas) que, tomados en promedio con referencia a *Monte Carlo Crudo* son: *Monte Carlo Crudo*: 1.00, *Dagger*: 1.20, *Monte Carlo Permutación*: 3.90, *EC-Monte Carlo Crudo*: 2.70 y *EC-Monte Carlo Permutación*: 4.00.

Todos los experimentos fueron realizados por medio de programas escritos en C (compilador gcc) haciendo uso del generador de números aleatorios *Mersenne Twister* [12]. De acuerdo a lo señalado por [9], para garantizar la convergencia y evitar oscilaciones en el método de *Entropía Cruzada sobre Monte Carlo Crudo* el valor de ρ fue adaptado al de la confiabilidad individual por enlace ($\rho = 0.40, 0.20$ y 0.10 para $r_i = 0.80, 0.90$ y 0.99 respectivamente). El tamaño de muestra para estimar los valores de γ_k fue 2000.

8 Conclusiones

En este trabajo se han estudiado e implementado cinco métodos de *Monte Carlo* para estimar la anti-confiabilidad de redes de comunicaciones: *Monte Carlo Crudo*, *Dagger*, *Monte Carlo Permutación*, y *Entropía Cruzada* sobre *Crudo* y *Permutación*. Las estimaciones fueron hechas dentro del criterio de *Confiabilidad Diámetro Acotada* habiendo resultado sencilla la adaptación de los métodos a dicho criterio.

Para estudiar el rendimiento de los distintos métodos el trabajo presenta los resultados obtenidos sobre tres topologías elementales de redes de comunicaciones, mostrando la precisión de cada método y su relación con la confiabilidad individual por enlace y con el parámetro D .

Con *Monte Carlo Permutación* se obtienen muy buenos resultados mientras que *Crudo* y *Dagger* exhiben un pobre desempeño, manteniéndose ambos en el mismo orden de precisión. Los modelos con *Entropía Cruzada* son modificaciones sobre *Monte Carlo Permutación* y *Monte Carlo Crudo*. En ambos casos la *Entropía Cruzada* produce algunas mejoras, pero de hecho la mejora es mucho más notable sobre *Crudo* que sobre *Permutación*.

Teniendo en cuenta que los modelos con *Entropía Cruzada* requieren un esfuerzo importante, tanto desde el punto de vista de su programación como de su uso (elección y ajuste de parámetros, tamaños de muestra, etc.) el método de *Monte Carlo Permutación* aparece como la mejor opción entre todas las implementadas (mejor relación “simplicidad de implementación–eficiencia”).

La influencia de D es más pronunciada en *Monte Carlo Permutación* que en *Crudo*, lo que equivale a decir que el método más preciso es más influenciado por el valor de D .

Dado que hay una vasta literatura sobre el análisis de confiabilidad de redes a partir de funciones de estructura clásicas (el problema de \mathcal{K} -terminales, por ejemplo) pero (hasta donde sabemos) ninguna sobre el enfoque de *Confiabilidad Diámetro Acotada*, los resultados de este trabajo resultan una útil contribución para la elección del método más apropiado en este último caso.

Cabe aclarar que para dar soporte a la parte experimental del presente trabajo (en particular a las implementaciones de *Monte Carlo Permutación* y *Entropía Cruzada* sobre *Permutación*) fue programado el algoritmo de convolución desarrollado al final de la Sección 5. Pero ya en etapa de revisión del informe fue encontrada en [14] una fórmula que permite determinar la densidad de probabilidad de una suma de distribuciones exponenciales según la cual, siendo $\{X_i\}_{i=1, \dots, n}$, $n \geq 2$ un conjunto de variables aleatorias independientes, distribuidas exponencialmente con todos los λ_i distintos, entonces:

$$f_{X_1+X_2+\dots+X_n}(x) = \left[\prod_{i=1}^n \lambda_i \right] \sum_{j=1}^n \frac{e^{-\lambda_j x}}{\prod_{k=1, k \neq j}^n (\lambda_k - \lambda_j)} \quad x > 0$$

Esta fórmula permite obviar la determinación explícita de la convolución y conseguir una notable reducción en el costo computacional que tiene la determinación de probabilidad de que una dada permutación no llegue a poner la red operativa en $t = 1$.

Las líneas de trabajo futuro que se desprenden del presente trabajo incluyen (i) ensayar estimaciones de *Confiabilidad Diámetro Acotada* sobre redes de mayor tamaño (aprovechando la reducción de costos debida a la fórmula presentada más arriba) y para la conectividad entre un determinado subconjunto de nodos ($\mathcal{K} \subset \mathcal{V}$) en lugar de conectividad entre todos los nodos ($\mathcal{K} = \mathcal{V}$), (ii) aplicar otros métodos de *Monte Carlo* a la resolución del mismo problema y (iii) mejorar la implementación de los mecanismos de convergencia en los modelos de *Entropía Cruzada*.

Referencias

- [1] H. Cancela and L. Petingi. Reliability of communication networks with delay constraints: Computational complexity and complete topologies. *International Journal of Mathematics and Mathematical Sciences*, 2004:29:1551–1562, 2004.
- [2] Héctor Cancela and Louis Petingi. Diameter constrained network reliability: exact evaluation by factorization and bounds. Technical Report 01-03, Facultad de Ingeniería, UDELAR, Montevideo, Uruguay, 2001. ISSN: 0797-6410.
- [3] Pieter-Tjerk de Boer, Dirk Kroese, Shie Mannor, and Reuven Rubinstein. A tutorial on the cross-entropy method. *Annals of Operations Research*, 134:19–67(49), February 2005.
- [4] Tito Homem de Mello. Rare event estimation for static models via cross-entropy and importance sampling.
- [5] Mohamed El-Khadiri and Gerardo Rubino. A Monte Carlo method based on antithetic variates for network reliability computations. Technical Report Rapports de Recherche No.1609, Institut National de Recherche en Informatique et en Automatique, 1992.
- [6] T. Elperin, I. B. Gertsbakh, and M. Lomonosov. Estimation of network reliability using graph evolution models. *IEEE Transactions on Reliability*, 40(5):572–581, Dec 1991.
- [7] G. S. Fishman. *Monte Carlo: Concepts, Algorithms, and Applications*. Springer Series in Operations Research. Springer-Verlag, New York, 1996.
- [8] H. Cancela and L. Petingi. On the characterization of the domination of a diameter constrained network reliability model. *Discrete Applied Mathematics*, 154:13:1885–1896, 2006.
- [9] K-P. Hui, N. Bean, M. Kraetzl, and Dirk Kroese. The cross-entropy method for network reliability estimation. *Annals of Operations Research*, 134:101–118(18), February 2005.
- [10] Kin-Ping Hui. *Network Reliability Estimation*. PhD thesis, The University of Adelaide, Adelaide, Australia, 2005.
- [11] H. Kumamoto, K. Tanaka, K. Inoue, and E. J. Henley. Dagger-Sampling Monte Carlo for system unavailability evaluation. *IEEE Transactions on Reliability*, R-29:122–125, 1980.
- [12] M. Matsumoto and T. Nishimura. Mersenne twister: A 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation*, 8(1):3–30, 1998.
- [13] L. Petingi and J. Rodriguez. Reliability of networks with delay constraints. *Congr. Numer.*, 152:117–123, 2001.
- [14] <http://www.math.bme.hu/~balazs/sumexp.pdf>. [11/07/07].