

Reducción de Varianza

Definiciones y Técnicas Básicas

Leslie Murray

Facultad de Ciencias Exactas, Ingeniería y Agrimensura
Universidad Nacional de Rosario
Rosario, Argentina

Mayo, 2024

- Sea \mathbf{X} una v.a. en el espacio Ω con:
 - función de densidad de probabilidad (fdp) $f_{\mathbf{X}}(x)$, si \mathbf{X} es continua,
 - densidad de masa de probabilidad $\mathbb{P}\{\mathbf{X} = x_i\} = p_i, \forall x_i \in \Omega$, si \mathbf{X} es discreta.
- Sean $\mathbb{E}\{\mathbf{X}\} = \boldsymbol{\mu}$ y $\mathbb{V}\{\mathbf{X}\} = \mathbb{E}\{(\mathbf{X} - \boldsymbol{\mu})^2\} = \mathbb{E}\{\mathbf{X}^2\} - \boldsymbol{\mu}^2 = \boldsymbol{\sigma}^2$, media y varianza de \mathbf{X} , respectivamente.
- Interesa conocer el valor de $\boldsymbol{\mu}$.
- Cuando el cálculo de $\boldsymbol{\mu}$ es difícil o directamente imposible, si el problema lo permite, se puede usar en su lugar un estimador.
- Sea $\hat{\boldsymbol{\mu}}$ el estimador Monte Carlo Estándar o crudo de $\boldsymbol{\mu}$.

¿CÓMO SE CALCULA $\hat{\boldsymbol{\mu}}$?

- Siendo $\{\mathbf{X}^{(i)}\}_{i=1}^n$ una sucesión de lecturas u observaciones independientes de \mathbf{X} ,
- cada $\mathbf{X}^{(i)}$ es una v.a. aleatoria en si misma llamada *copia* o *replicación* de \mathbf{X} .
- $\{\mathbf{X}^{(i)}\}_{i=1}^n$ es, por lo tanto, una sucesión de v.a. *i.i.d.*
- Llamando $S_n = \sum_{i=1}^n \mathbf{X}^{(i)}$, el estimador $\hat{\boldsymbol{\mu}}$ se calcula como:

$$\hat{\boldsymbol{\mu}} = \frac{S_n}{n}$$

- $\mathbb{E}\{\hat{\mu}\} = \mathbb{E}\left\{\frac{S_n}{n}\right\} = \frac{1}{n}\mathbb{E}\{S_n\} = \frac{1}{n}n\mu = \mu$
- $\mathbb{V}\{\hat{\mu}\} = \mathbb{V}\left\{\frac{S_n}{n}\right\} = \frac{1}{n^2}\mathbb{V}\{S_n\} = \frac{1}{n^2}n\mathbb{V}\{\mathbf{X}\} = \frac{\sigma^2}{n}$

$\mathbb{V}\{\hat{\mu}\}$ es un indicador de la calidad del estimador $\hat{\mu}$, véase por qué:

Intervalo de Confianza: $[I_1, I_2] = \left[\hat{\mu} - z_{\alpha/2} \sigma/\sqrt{n}, \hat{\mu} + z_{\alpha/2} \sigma/\sqrt{n}\right]$

μ está dentro del intervalo $[I_1, I_2]$ con probabilidad $1 - \alpha$.

Error Relativo: $ER = \frac{\sigma/\sqrt{n}}{\mu}$.

ER es proporcional al tamaño del Intervalo de Confianza y por lo tanto es un indicador de qué tan cerca de μ puede estar $\hat{\mu}$.

Ambos indicadores están determinados por $\mathbb{V}\{\hat{\mu}\} = \sigma^2/n$

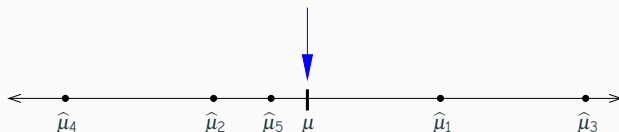
La varianza del estimador estándar sólo puede bajarse a costa del incremento de n .

Objetivo de la Reducción de Varianza

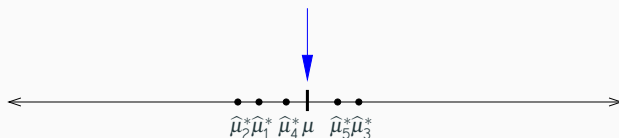
Siendo $\hat{\mu}$ el estimador Monte Carlo Estándar de μ , es de interés encontrar otro estimador, llamémosle $\hat{\mu}^*$, con las siguientes propiedades:

- $\mathbb{E}\{\hat{\mu}^*\} = \mathbb{E}\{\hat{\mu}\} = \mu$
- $\mathbb{V}\{\hat{\mu}^*\} < \mathbb{V}\{\hat{\mu}\}$

Una serie de replicaciones del estimador $\hat{\mu}$, visto como v.a., podría ser la siguiente:



En tanto una serie de replicaciones del estimador $\hat{\mu}^*$ podría ser:



Hay gran cantidad de métodos de Reducción de Varianza para hacer estimaciones eficientes de la media, μ , de una variable aleatoria \mathbf{X} . Por ejemplo:

- **Muestreo de Importancia.** Se modifica la distribución de probabilidad de la variable cuya media se quiere estimar, procurando encontrar una distribución que de lugar a un estimador con menor varianza. Al final se corrige el resultado.
- **Variables Antitéticas.** Se modifica el método de muestreo para conseguir correlación negativa entre muestras. La estimación realizada a partir de muestras negativamente correlacionadas, tiene menor varianza que la estándar.
- **Monte Carlo Condicional.** Se condiciona la v.a. cuya media es de interés a otra v.a. del modelo. Sorteando adecuadamente, el estimador de la variable condicionada tiene igual media y menor varianza que el de la variable original.
- **Muestreo Estratificado.** Se particiona el espacio de estado en subespacios (estratos) y se transforma la simulación global en una simulación por estratos.
- **Método basado en Cotas.** A partir de cotas (inferiores y superiores) para la función de estructura y para el valor a estimar, se crean una v.a. y una distribución de probabilidad nuevas. La v.a. nueva tiene, bajo la distribución nueva, igual media y menor varianza.

Se presentan seguidamente algunos.

Muestreo de Importancia

Muestreo de Importancia (I)

En la estimación Monte Carlo crudo se considera (implícitamente) la probabilidad de cada valor de \mathbf{X} por la proporción de veces que ese valor aparece:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}^{(i)}$$

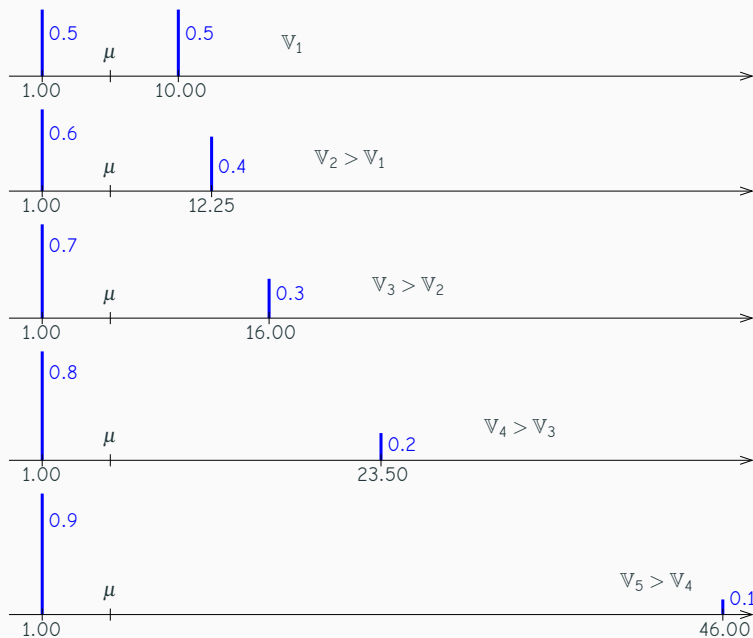
Los valores $\{\mathbf{x}^{(i)}\}_{i=1}^n$ se sortean en forma aleatoria, luego, valores con probabilidad baja o extremadamente baja, podrían no sortearse nunca.

Bajo determinadas circunstancias, la presencia de valores de \mathbf{X} con probabilidad muy baja (*evento raro*) tiende a incrementar la varianza de \mathbf{X} y, consecuentemente, de $\hat{\mu}$.

EJEMPLO 1: Sea \mathbf{X} una v.a. en $\Omega = \{1, k\} \rightarrow \mathbb{E}\{\mathbf{X}\} = 1 \times \mathbb{P}\{\mathbf{X} = 1\} + k \times \mathbb{P}\{\mathbf{X} = k\}$. Se varían las probabilidades y se ajusta k para que, en todos los casos, $\mathbb{E}\{\mathbf{X}\} = 5.50$.

1.00	$\mathbb{P}\{\mathbf{X} = 1\}$	k	$\mathbb{P}\{\mathbf{X} = k\}$	$\mathbb{E}\{\mathbf{X}\}$	$\mathbb{V}\{\mathbf{X}\}$
1.00	0.5000	10.00	0.5000	5.50	20.25
1.00	0.6000	12.25	0.4000	5.50	30.38
1.00	0.7000	16.00	0.3000	5.50	47.25
1.00	0.8000	23.50	0.2000	5.50	81.00
1.00	0.9000	46.00	0.1000	5.50	182.25
1.00	0.9900	451.00	0.0100	5.50	2004.75
1.00	0.9990	4501.00	0.0010	5.50	20229.75
1.00	0.9999	45001.00	0.0001	5.50	202479.75

Muestreo de Importancia (II)



Aquí la misma tabla anterior:

1.00	$\mathbb{P}\{\mathbf{X} = 1\}$	k	$\mathbb{P}\{\mathbf{X} = k\}$	$\mathbb{E}\{\mathbf{X}\}$	$\mathbb{V}\{\mathbf{X}\}$
1.00	0.5000	10.00	0.5000	5.50	20.25
1.00	0.6000	12.25	0.4000	5.50	30.38
1.00	0.7000	16.00	0.3000	5.50	47.25
1.00	0.8000	23.50	0.2000	5.50	81.00
1.00	0.9000	46.00	0.1000	5.50	182.25
1.00	0.9900	451.00	0.0100	5.50	2004.75
1.00	0.9990	4501.00	0.0010	5.50	20229.75
1.00	0.9999	45001.00	0.0001	5.50	202479.75

- Cada una de las 8 líneas está asociada a una v.a. con un espacio Ω y una distribución diferente, pero para todas la media es la misma.
- Si el objetivo es estimar $\mu = \mathbb{E}\{\mathbf{X}\}$, da lo mismo hacerlo mediante la simulación de cualquiera de las ocho v.a.
- Sin embargo, la eficiencia de cada una de las 8 formas de hacer la estimación resultará muy diferente.
- La esencia del Muestreo de Importancia está en cambiar la distribución de probabilidad (*cambio de medida*) de una v.a. \mathbf{X} , por otra que conserve la media y permita estimarla en condiciones más favorables (menor varianza, menor error).

Una forma posible de hacer el *cambio de medida*, es la siguiente:

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

a condición de que $\int_{-\infty}^{+\infty} h(x) dx = 1$ y que $h(x) > 0$, $\forall x$ donde $\phi(x)f(x) > 0$.

Luego, da lo mismo estimar μ generando valores de \mathbf{X} a partir de f y aplicarlos a $\phi(\mathbf{X})$, que generarlos a partir de h y aplicarlos a $\phi(\mathbf{X})R(\mathbf{X})$, donde $R(\mathbf{X}) = f(\mathbf{X})/h(\mathbf{X})$:

$$\text{M.I.} \rightarrow \tilde{\mu} = \frac{1}{n} \sum_{i=1}^n \phi(\mathbf{X}^{(i)}) R(\mathbf{X}^{(i)}), \quad \mathbf{X} \sim h \qquad \text{Estándar} \rightarrow \hat{\mu} = \frac{1}{n} \sum_{i=1}^n \phi(\mathbf{X}^{(i)}), \quad \mathbf{X} \sim f$$

Las varianzas son:

$$\begin{aligned} \mathbb{V}_f\{\phi(\mathbf{X})\} &= \int_{-\infty}^{+\infty} \phi(x)^2 f(x) dx - \mathbb{E}_f^2\{\phi(\mathbf{X})\} \\ \mathbb{V}_h\{\phi(\mathbf{X}) R(\mathbf{X})\} &= \int_{-\infty}^{+\infty} \phi(x)^2 R(x)^2 h(x) dx - \mathbb{E}_h^2\{\phi(\mathbf{X}) R(\mathbf{X})\} \end{aligned}$$

Variables Antitéticas

Sean \mathbf{X}_1 y \mathbf{X}_2 dos v.a. con distribución conocida,

- Si \mathbf{X}_1 e \mathbf{X}_2 son independientes:

$$\mathbb{V}\{\mathbf{X}_1 \pm \mathbf{X}_2\} = \mathbb{V}\{\mathbf{X}_1\} + \mathbb{V}\{\mathbf{X}_2\}$$

- Si \mathbf{X}_1 e \mathbf{X}_2 "no" son independientes:

$$\mathbb{V}\{\mathbf{X}_1 \pm \mathbf{X}_2\} = \mathbb{V}\{\mathbf{X}_1\} + \mathbb{V}\{\mathbf{X}_2\} \pm 2\text{Cov}\{\mathbf{X}_1, \mathbf{X}_2\}$$

Sean, $\hat{\mu}$ el estimador crudo previamente definido y $\hat{\mu}^*$ otro estimador crudo basado en muestras $\{\mathbf{X}^{(i)}\}_{i=1}^n$ que no son independientes:

$$\begin{aligned}\mathbb{V}\{\hat{\mu}^*\} &= \mathbb{V}\left\{\frac{1}{n}\sum_{i=1}^n \mathbf{x}^{(i)}\right\} \\ &= \frac{1}{n^2} \mathbb{V}\left\{\sum_{i=1}^n \mathbf{x}^{(i)}\right\} \\ &= \frac{1}{n^2} \sum_{1 \leq i \leq n} \mathbb{V}\{\mathbf{x}^{(i)}\} + \frac{2}{n^2} \sum_{1 \leq i < j \leq n} \text{Cov}\{\mathbf{x}^{(i)}, \mathbf{x}^{(j)}\} \\ &= \mathbb{V}\{\hat{\mu}\} + \frac{2}{n^2} \sum_{1 \leq i < j \leq n} \text{Cov}\{\mathbf{x}^{(i)}, \mathbf{x}^{(j)}\}\end{aligned}$$

$$\mathbb{V}\{\hat{\mu}^*\} = \mathbb{V}\{\hat{\mu}\} + \frac{2}{n^2} \sum_{1 \leq i < j \leq n} \text{Cov}\{\mathbf{x}^{(i)}, \mathbf{x}^{(j)}\}$$

- Si la suma de covarianzas es negativa, $\mathbb{V}\{\hat{\mu}^*\} < \mathbb{V}\{\hat{\mu}\}$.
- La suma contempla todos los pares $(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})$ considerando cada uno una sola vez (notar que: $\text{Cov}\{\mathbf{X}, \mathbf{Y}\} = \text{Cov}\{\mathbf{Y}, \mathbf{X}\}$).
- No hace falta que “todos” los pares tengan correlación negativa, basta con que “algunos” la tengan y que los demás sean independientes.

Base del Método

El método de Variables Antitéticas consiste sortear las n muestras como $n/2$ pares, a condición de que las variables de cada par tengan, entre sí, correlación negativa.

Una forma de generar correlación negativa entre pares de muestras es la siguiente:

Toda vez que como parte de una replicación se use, directa o indirectamente, un valor aleatorio uniformemente distribuido, U , en un instante futuro de la simulación deberá repetirse esa replicación usando el valor $(1 - U)^{(*)}$.

Si, parte de una simulación se basa en la secuencia:

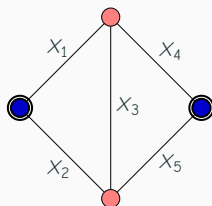
$$U^{(1)}, U^{(2)}, U^{(3)}, \dots, U^{(n)},$$

en otra parte de la simulación deberá usarse la secuencia:

$$1 - U^{(1)}, 1 - U^{(2)}, 1 - U^{(3)}, \dots, 1 - U^{(n)},$$

^(*) Es simple probar que dos v.a. de Bernoulli, X e Y , generadas mediante números aleatorios U y $1 - U$, respectivamente, están correlacionados negativamente.

EJEMPLO 1:



$$1 - \widehat{\zeta} = 1 - \frac{1}{n} \sum_{i=1}^n \phi(\mathbf{X})^{(i)} \rightarrow \phi(\mathbf{X})^{(i)} = \phi(X_1^{(i)}, X_2^{(i)}, X_3^{(i)}, X_4^{(i)}, X_5^{(i)})$$

En la i -ésima replicación se generan los cinco valores necesario de la siguiente forma:

$$X_j^{(i)} = \begin{cases} 1 & \text{si } U_j^{(i)} < r_j \\ 0 & \text{caso contrario} \end{cases} \quad U_j^{(i)} \sim \text{Unif}(0,1) \quad j = 1, 2, \dots, 5.$$

La $i+1$ -ésima replicación de cada enlace tendrá la siguiente forma:

$$X_j^{(i+1)} = \begin{cases} 1 & \text{si } 1 - U_j^{(i)} < r_j \\ 0 & \text{caso contrario} \end{cases}$$

siendo $U_j^{(i)}$, $j = 1, 2, \dots, 5$, los mismos de la i -ésima replicación (hay que guardarlos).

Monte Carlo Condicional

Siendo \mathbf{X} la variable cuya media interesa estimar e \mathbf{Y} alguna otra variable del modelo, es simple probar que:

$$\mu = \mathbb{E}\{\mathbf{X}\} = \mathbb{E}\{\mathbb{E}\{\mathbf{X} \mid \mathbf{Y}\}\}$$

- Luego, tanto \mathbf{X} como $\mathbb{E}\{\mathbf{X} \mid \mathbf{Y}\}$ son variables cuya media es el valor de interés.
- ¿Cómo se podría hacer una simulación de tipo Monte Carlo de $\mathbb{E}\{\mathbf{X} \mid \mathbf{Y}\}$ y qué ventaja o desventaja tendría sobre la simulación directa de \mathbf{X} ?
- Según una igualdad simple de probar:

$$\mathbb{V}\{\mathbf{X}\} = \mathbb{E}\{\mathbb{V}\{\mathbf{X} \mid \mathbf{Y}\}\} + \mathbb{V}\{\mathbb{E}\{\mathbf{X} \mid \mathbf{Y}\}\}$$

luego, dado que los tres términos son positivos:

$$\mathbb{V}\{\mathbf{X}\} \geq \mathbb{V}\{\mathbb{E}\{\mathbf{X} \mid \mathbf{Y}\}\}$$

- Conviene entonces hacer la estimación a partir de n muestras independientes de $\mathbb{E}\{\mathbf{X} \mid \mathbf{Y}\}$ en lugar de hacerlo a partir de n muestras independientes de \mathbf{X} .

¿Cuál es la forma de obtener muestras de $\mathbb{E}\{\mathbf{X} \mid \mathbf{Y}\}$?

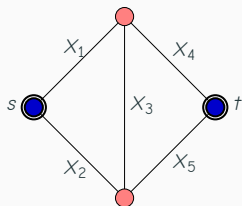
Monte Carlo Condicional (II)

Para obtener muestras de $\mathbb{E}\{\mathbf{X} | \mathbf{Y}\}$:

- Se sortean n valores *i.i.d.* de la variable \mathbf{Y} : $\{\mathbf{Y}^{(i)}\}_{i=1}^n$.
- Fijando el valor de \mathbf{Y} al sorteado $\mathbf{Y}^{(i)}$, se calcula $\mathbb{E}\{\mathbf{X}\}$.
- Este último cálculo resulta ser, en realidad, $\mathbb{E}\{\mathbf{X} | \mathbf{Y}^{(i)}\} = \mathbb{E}\{\mathbf{X} | \mathbf{Y}\}^{(i)}$.
- El estimador Monte Carlo Condicional, en definitiva un estimador de $\mathbb{E}\{\mathbf{X} | \mathbf{Y}\}$, es entonces:

$$\hat{\mu}^* = \frac{1}{n} \sum_{i=1}^n \mathbb{E}\{\mathbf{X} | \mathbf{Y}\}^{(i)} = \frac{1}{n} \sum_{i=1}^n \mathbb{E}\{\mathbf{X} | \mathbf{Y}^{(i)}\}$$

EJEMPLO 2:



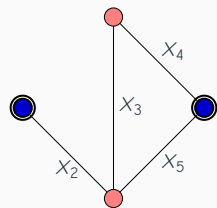
$$\mathbb{P}\{X_i = 1\} = r_i$$

$$\mathbb{P}\{X_i = 0\} = q_i = 1 - r_i$$

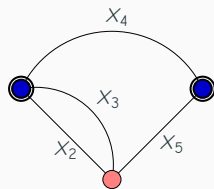
$$i = 1, 2, 3, 4, 5$$

$$\gamma(\mathbf{X}) = \begin{cases} 1 & \text{si } \mathbf{X} \text{ es tal que } s \text{ y } t \text{ "no" están conectados} \\ 0 & \text{en caso contrario} \end{cases} \quad \leftarrow \quad \gamma(\mathbf{X}) = 1 - \phi(\mathbf{X})$$

Elegimos X_1 , para estimar $1 - \zeta$ a través de la v.a. $\mathbb{E}\{\gamma(\mathbf{X}) \mid X_1\}$:



$X_1 = 0$



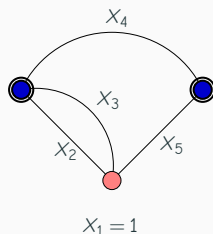
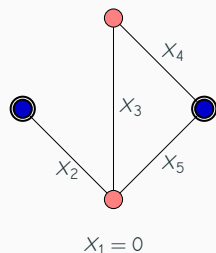
$X_1 = 1$

El mecanismo resultante es entonces:

- Sortear n valores *i.i.d.* de la variable X_1 , esto es $\{X_1^{(i)}\}_{i=1}^n$.
- Fijar X_1 al valor sorteado, $X_1^{(i)}$, y calcular $\mathbb{E}\{\gamma(\mathbf{X}) \mid X_1^{(i)}\}$.
- Calcular el estimador Monte Carlo Condicional como:

$$\hat{\mu}^* = \frac{1}{n} \sum_{i=1}^n \mathbb{E}\{\gamma(\mathbf{X}) \mid X_1\}^{(i)} = \frac{1}{n} \sum_{i=1}^n \mathbb{E}\{\gamma(\mathbf{X}) \mid X_1^{(i)}\}$$

Se deben calcular (antes de la simulación) $\gamma(\mathbf{X})$ si $X_1 = 0$ y $\gamma(\mathbf{X})$ si $X_1 = 1$:



$$\mathbb{E}\{\gamma(\mathbf{X})|X_1 = 0\} = q_2 + q_4q_5 + q_3q_5 - q_2q_4q_5 - q_2q_3q_5 - q_3q_4q_5 + q_2q_3q_4q_5, \quad (1)$$

$$\mathbb{E}\{\gamma(\mathbf{X})|X_1 = 1\} = q_2q_3q_4 + q_4q_5 - q_2q_3q_4q_5. \quad (2)$$

Mecanismo

Se sortean valores para X_1 , si sale 0 se acumula (1), si sale 1, se acumula (2).

$$\hat{\mu}^* = \frac{1}{n} \sum_{i=1}^n \mathbb{E}\{\gamma(\mathbf{X}) | X_1^{(i)}\}$$

q	$E\{E\{\gamma(\mathbf{X}) X_1\}\}$	$V\{E\{\gamma(\mathbf{X}) X_1\}\}$	$V\{\gamma(\mathbf{X})\}/V\{E\{\gamma(\mathbf{X}) X_1\}\}$
1	1.00000E+00	0.00000E+00	---
0.1	2.15200E-02	1.01506E-03	20.74
0.01	2.01950E-04	1.00910E-06	200.09
0.001	2.00200E-06	1.00099E-09	2,000.01
0.0001	2.00020E-08	1.00010E-12	20,000.00
0.00001	2.00002E-10	1.00001E-15	200,000.00
0.000001	2.00000E-12	1.00000E-18	2,000,000.00



- Monte Carlo Condicional es una combinación entre *simulación* y *cálculo exacto*.
- De dónde esté la frontera que separa *simulación* y *cálculo exacto* dependerá la precisión del método.
- A modo de ejemplo, si en lugar de condicionar contra X_1 , se lo hace contra (X_1, X_2) :
 - En cada replicación se sortea uno de los cuatro posibles valores de (X_1, X_2) .
 - Por cada valor sorteado (X_1, X_2) , se calcula $\gamma(\mathbf{X})$ en la red resultante.

Ahora la varianza del estimador será **mayor** que en el caso anterior (más *simulación* y menos *cálculo exacto*), pero el cálculo exacto será más sencillo.



S.M. Ross. *Introduction to Probability Models*. 10th ed. Elsevier Science, 2006. ISBN: 9780123756879.



R.Y. Rubinstein and D.P. Kroese. *Simulation and the Monte Carlo Method*. Wiley, 2008. ISBN: 9780470230374.