



UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO

FACULTAD DE CIENCIAS

Muestreo de Importancia:

T E S I S

QUE PARA OBTENER EL TÍTULO DE :

A C T U A R I O

P R E S E N T A :

CÉSAR ALMENARA MARTÍNEZ



FACULTAD DE CIENCIAS
UNAM

TUTOR:
DRA. ANA MEDA GUARDIOLA

2008

Índice general

| | |
|--|-----------|
| 1. Introducción. | 3 |
| §1.1. El Método. | 6 |
| 2. Muestreo de Importancia. | 13 |
| §2.1. Simulación Monte Carlo. | 13 |
| §2.2. Teorema del Límite Central e Intervalos de Confianza. | 15 |
| §2.3. Reducción de Varianza. | 17 |
| §2.4. El Muestreo de Importancia. | 18 |
| §2.5. Simulación de Eventos Raros. | 19 |
| §2.6. Razón de Verosimilitudes. | 24 |
| §2.7. Cambio Exponencial de Medida. | 25 |
| 3. Grandes Desviaciones. | 27 |
| §3.1. Motivación e Idea Heurística. | 27 |
| §3.2. Bases Teóricas. | 28 |
| 4. Contra-ejemplo. | 41 |
| §4.1. Descripción del Problema. | 42 |
| §4.2. Suma de Variables. | 47 |
| 5. Aplicaciones. | 55 |
| §5.1. VaR. | 55 |
| §5.1.1. Aproximaciones Delta y Delta-Gamma. | 58 |
| §5.1.2. Reducción de Varianza. | 59 |
| §5.1.3. Reducción de Varianza con Base en la Aproximación Delta. | 62 |
| §5.1.4. Reducción de Varianza con Base en la Aproximación Delta-Gamma. | 64 |
| §5.2. Opciones. | 67 |
| §5.2.1. Horizonte Aleatorio. | 69 |

| | |
|--|-----------|
| §5.2.2. Horizonte Largo. | 71 |
| §5.2.3. Opción de Choque: <i>Knock in Option</i> | 72 |
| §5.2.4. Opciones de Trayectoria Dependiente. | 75 |
| §5.2.5. Cambio de Deriva: Linealización. | 76 |
| §5.2.6. Cambio de Deriva: Aproximación Normal y Trayectoria Ópti- ma. | 78 |
| §5.2.7. Opciones Asiáticas. | 80 |
| §5.2.8. Muestreo de Importancia más Estratificación. | 81 |
| §5.3. El Problema de la Ruina. | 81 |
| §5.3.1. El Proceso de Riesgo. | 82 |
| §5.3.2. Muestreo de Importancia Via Conjugación de Lundberg. . . | 86 |
| §5.3.3. Muestreo de Importancia para el caso de Horizonte Finito. . | 91 |
| 6. Apéndice A | 97 |
| §6.1. MCCN.m | 98 |
| §6.2. ISN.m | 98 |
| §6.3. IC95.m | 99 |
| §6.4. call-asiatico-MCC.m | 99 |
| §6.5. call-asiatico-IS.m | 103 |

Notación.

| | |
|----------------------|---|
| \mathbb{P} | Probabilidad. |
| \mathbb{E} | Operador Esperanza. |
| Var | Operador Varianza. |
| $\mathbb{I}_{\{A\}}$ | Función indicadora sobre el conjunto $\{A\}$. |
| \Rightarrow | Implica. |
| \therefore | Por lo tanto. |
| \forall | Para todo. |
| $:=$ | Definición (el lado izquierdo se define como el lado derecho). |
| F, G | Distribuciones de probabilidad. |
| f, g | Funciones de densidad. |
| $N(\mu, \sigma^2)$ | Distribución Normal con media μ y varianza σ^2 . |
| i.e | <i>Es decir.</i> |
| i.i.d | <i>Independientes e idénticamente distribuidas.</i> |
| v.a | <i>Variable aleatoria.</i> |
| c.s | <i>Convergencia casi segura ($\xrightarrow{c.s.}$).</i> |
| S_n | Suma de las variables aleatorias X_1, \dots, X_n . |
| $M(\theta)$ | Función generadora de momentos (f.g.m) o transformada de Laplace. |
| $\psi(\theta)$ | Función generadora acumulativa o logaritmo de la f.g.m. ($\log M(\theta)$). |
| $>>$ | Representa la continuidad absoluta entre dos medidas. |
| \propto | Proporcional a. |
| \approx | Aproximadamente a. |
| \equiv | Equivalente a. |
| \underline{v} | Vector $v = (v_1, \dots, v_n)$ donde n depende del contexto del problema. |
| \underline{v}^T | Vector v transpuesto, cuya dimensión depende del contexto del problema. |
| \mathbb{I} | Matriz identidad, cuya dimensión depende del contexto del problema. |

Capítulo 1

Introducción.

En la actualidad la simulación se ha convertido en una herramienta estadística muy importante ya que la complicada estructura de los fenómenos en estudio dificulta encontrar resultados analíticos que ayuden a su interpretación. Sin embargo la simulación puede ser una herramienta sumamente costosa en términos computacionales, por lo que el verdadero problema es desarrollar métodos que reduzcan el número de simulaciones sin reducir la exactitud de la estimación. Estos métodos son conocidos con el nombre de *técnicas de reducción de varianza*. En este trabajo mostraremos el funcionamiento del *Muestreo de Importancia*, técnica que a pesar de ser utilizada como un método de estimación *Monte Carlo* en la práctica, en gran parte de la literatura referente a estos métodos (métodos Monte Carlo) es considerada únicamente como una técnica de reducción de varianza, lo que le resta relevancia y generalidad al método en si. Nosotros trataremos al muestreo de importancia como una técnica general de simulación y explicaremos el aspecto teórico de la efectividad de la misma, utilizando la amplia relación que mantiene con las *Grandes Desviaciones*¹. Con esto en mente, comencemos por explicar algunos conceptos básicos que nos ayudarán a entender el mundo de la simulación estocástica.

El método **Monte Carlo** consiste en generar valores X_i de una distribución F para $i = 1, \dots, n$ y calcular $\theta_i = \theta(X_i)$ para cada i , donde al generar valores de θ_i para un n suficientemente grande, la distribución empírica con pesos $\frac{1}{n}$ en cada valor de θ_i es una estimación de $\theta(X)$.

$\theta(X)$ representa información relevante como valores de integrales o en un con-

¹Las *Grandes Desviaciones* son básicamente un conjunto de resultados asintóticos a probabilidades de eventos de baja ocurrencia así como un conjunto de métodos para derivar dichos resultados.

texto probabilístico, probabilidades de eventos específicos o cantidades como la esperanza y la varianza de alguna población o fenómeno que se desea describir o entender, y que debido a su estructura, encontrar una forma cerrada para calcularla es imposible o sumamente difícil, por lo que la alternativa es crear métodos que nos proporcionen aproximaciones confiables y de bajo costo en términos del tiempo empleado para dicho proceso. Los métodos Monte Carlo se ajustan particularmente bien para este propósito, basando su estimación de la realidad en la *Ley Fuerte de los Grandes Números* y asegurando con esto que para muestras de gran tamaño la convergencia hacia el valor exacto es casi segura. Sin embargo, en la realidad simular muestras lo suficientemente grandes para garantizar dicha convergencia se refleja en costos elevados para quien realiza el cálculo, donde estos costos se refieren al uso de recursos computacionales, monetarios, de infraestructura y en la mayoría de los casos al tiempo empleado. Es por eso que en el fondo el problema se transforma en la elección de estimadores que proporcionen una varianza pequeña para así asegurar una menor inversión de recursos.

El **Muestreo de Importancia** o **IS** como sus siglas en inglés lo denominan, es un método de simulación vía Monte Carlo descrito como el proceso de estimar algo acerca de una distribución mediante observaciones de una distribución diferente. Bajo el muestreo de importancia (**IS**) cada dato se obtiene de una *distribución de muestreo* diferente de la distribución original, de modo que si nombramos a esta última distribución como G , entonces $X_i \sim G$ y los valores observados $\theta(X_i)$ para pesos tales que se compense el sesgo del método de muestreo propuesto por la distribución estimada, representan nuestra nueva estimación.

La *distribución de muestreo* se elige con la intención de mejorar la eficiencia² de la simulación, y así asignar mayor densidad a la región de **importancia** o área de interés en comparación con la distribución verdadera. Existen diferentes tipos de esquemas en el muestreo de importancia, ya que la elección del cambio de medida está en función del evento o fenómeno en estudio. En el caso de probabilidades de eventos de baja ocurrencia, mejor conocidos como eventos “raros”, el esquema propuesto por las *Grandes Desviaciones* proporciona el cambio de medida óptimo en un sentido asintótico, y muchas veces dicho cambio proporciona estimaciones de varianza mínima, con lo cual se asegura su efectividad en términos de eficiencia.

²La eficiencia en términos estadísticos, se refiere al mínimo uso de recursos para la obtención de una estimación, es decir que una estimación es más eficiente que otra si se obtienen los mismos resultados con un menor empleo de recursos.

El concepto de eficiencia será tratado a detalle en los capítulos subsecuentes, pero en esencia se refiere a obtener una mayor reducción de varianza³ utilizando la menor cantidad de recursos posible. La eficiencia y la precisión son dos conceptos distintos pero ampliamente ligados, ya que para obtener estimaciones eficientes debemos asegurar un cierto grado de precisión. A lo largo de este documento explicaremos debidamente estos conceptos y enfatizaremos su importancia para efectos prácticos.

En resumen, como consecuencia del cambio de distribución impuesto por el muestreo de importancia, tenemos las siguientes ventajas:

- Se obtiene gran ganancia en la eficiencia de la simulación, es decir que el orden de magnitud del número de iteraciones que se requieren para alcanzar cierto grado de efectividad se reduce considerablemente.
- En algunos de los casos más difíciles de estimar, como los eventos de baja probabilidad se obtiene la mayor precisión registrada.
- Facilita la estimación de eventos “raros”, es decir que este método es muy común en la estimación de eventos poco probables ya que de hacerse de forma adecuada (esquema via grandes desviaciones), la eficiencia de la simulación en términos de reducción de varianza, es óptima.
- Es consistente con la *Teoría de Grandes Desviaciones*, que nos habla de aspectos asintóticos y comportamientos límite de eventos o fenómenos ubicados en la cola de la distribución.
- Es un método viable en aplicaciones con más de una distribución verdadera.
- El muestreo de importancia puede usarse para estimar resultados de muchas distribuciones en una sola simulación.
- El **IS** puede utilizarse para calcular esperanzas con respecto a parámetros de entrada en la distribución o distribuciones.

³La reducción de varianza como ya indicamos es el proceso utilizado para incrementar la precisión de estimaciones que pueden ser obtenidas mediante un número fijo de iteraciones.

§1.1. El Método.

Para describir rápidamente el método del *Muestreo de Importancia*, debemos tomar en cuenta que para cualquier aplicación de este, una muestra se obtiene bajo la “distribución de muestreo”, sin embargo se requiere de una estimación basada en la distribución verdadera. Esta estimación se obtiene bajo la distribución empírica pero en lugar de suponer uniformidad sobre los pesos, se asignan diferentes pesos a cada observación.

Dichos pesos se eligen a modo de insesgar los resultados, es decir, para impedir sesgos hechos al utilizar la “distribución de muestreo” en lugar de la verdadera. Para poder realizar esto, se escogen pesos rigurosamente proporcionales a los de la “función de pesos”, definida de la siguiente manera

$$\mathbf{W}(X) := \frac{f(X)}{g(X)}, \quad (1.1)$$

donde f, g representan las densidades de F y G respectivamente. Cabe mencionar que a pesar de la condición rigurosa de proporcionalidad, la elección de diferentes pesos es posible, de tal forma que diferentes filosofías de **IS** conducen a elecciones distintas de pesos y por lo tanto a diferentes estimaciones.

Existen dos amplias interpretaciones del **IS**, la Integración y el Muestreo. Desde el punto de vista tradicional, la simulación Monte Carlo es una forma de integración ya que para estimar la esperanza de $\theta(X)$ bajo la densidad f , debemos realizar lo siguiente

$$\mathbb{E}_f(\theta(X)) = \int \theta(x) f(x) dx = \int \theta(x) \frac{f(x)}{g(x)} g(x) dx.$$

Por lo tanto, si $g(x) > 0$ cuando $f(x) > 0$, definimos a $Y(x) := \theta(x) f(x) / g(x)$, para escribir a la media como

$$\mu := \mathbb{E}_f(\theta(X)) = \mathbb{E}_g(Y(X)),$$

y claramente μ puede estimarse bajo $\bar{Y} = 1/n \sum_{i=1}^n Y(x_i)$, la muestra promedio de los valores obtenidos bajo g para una cierta muestra de tamaño n .

Al hacer **IS** de esta forma, se involucran tanto un esquema modificado de muestreo como una transformación inducida de θ a Y , de tal forma que si Y tiene

menor varianza (bajo g) que la que tiene θ (bajo f), nuestra estimación es más eficiente (en el siguiente capítulo trataremos a fondo con la reducción de varianza). Bajo este esquema de **IS**, la *integración estimada* de una esperanza es

$$\hat{\mu}_{int} := \frac{1}{n} \sum_{i=1}^n Y(x_i) = \bar{Y}. \quad (1.2)$$

Aquí el muestreo de importancia involucra la concentración de esfuerzos, es decir, requiere de especial atención en las regiones relevantes del espacio muestral. De esta forma el muestreo de importancia puede confundirse con un muestreo estratificado⁴, sin embargo en el transcurso de este trabajo nos daremos cuenta de que el **IS** es una herramienta diferente y mucho más flexible para la estimación de eventos poco probables.

La aproximación muestral necesita que la distribución tenga probabilidad total igual a 1. Cabe resaltar que la *integración estimada* puede reescribirse como un promedio pesado de la siguiente forma

$$\hat{\mu}_{int} = \sum_{i=1}^n V_{int}(X_i) \theta(X_i), \quad (1.3)$$

en donde

$$V_{int}(X_i) = \frac{\mathbf{W}(X_i)}{n} \quad (1.4)$$

representa el peso asignado a cada una de las observaciones i . Por lo tanto tenemos como resultado un promedio pesado sobre los valores observados, con pesos que por lo general no suman 1; sin embargo podemos asegurar que

$$\mathbb{E}_g(\mathbf{W}) = \int \mathbf{W}(x) g(x) dx = \int f(x) dx = 1. \quad (1.5)$$

Como la aproximación muestral requiere que los pesos sumen 1, sólo necesitamos hacer la siguiente ponderación

$$V_{razón}(X_i) := \frac{\mathbf{W}(X_i)}{\sum_j \mathbf{W}(X_j)}. \quad (1.6)$$

Esto nos da como resultado lo que se denomina *Razón Estimada*

⁴El muestreo estratificado se define como aquel en el que se divide a una población de N individuos en k sub-poblaciones o *estratos*, atendiendo a criterios que puedan ser importantes en el estudio, y posteriormente realizar un muestreo aleatorio simple en cada uno de estos *estratos*.

$$\hat{\mu}_{\text{razón}} = \frac{1}{n} \sum_{i=1}^n V_{\text{razón}}(X_i) \theta(X_i) = \frac{\bar{Y}}{\bar{W}}. \quad (1.7)$$

La aproximación muestral y la estimación via integración, proporcionan mayor eficiencia en el cálculo de integrales y valores esperados (esperanzas), particularmente para distribuciones con expresiones o parámetros difíciles de calcular, esto incluye la estimación de probabilidades pequeñas o probabilidades que tienden a cero. Por otro lado la estimación via integración no es equivariante, lo que repercute en un cierto número de consecuencias inesperadas, que deberán ser evitadas en muchas de las aplicaciones. Una de estas dificultades ocurre en simulaciones con más de un parámetro o cantidad de salida, esto se debe a que la misma transformación propuesta por el método tiene que aplicarse a cada una de estas cantidades o parámetros.

La aproximación muestral es más general, pues la razón estimada produce una distribución estimada de puntos y pesos, mientras que la estimación via Integración sólo estima esperanzas. Esta aproximación puede utilizarse en algunas aplicaciones que caen fuera del objetivo tradicional de la estimación via integración, el cual es la reducción de varianza. Aún más, esta estimación es responsable de garantizar rigidez en las estrategias de muestreo para un parámetro de salida. De cualquier forma, la *Razón Estimada* es menos útil como técnica pura para la reducción de varianza con una cota inferior distinta del cero en la posible reducción. Afortunadamente, las extensiones de ambos métodos convergen en la producción de los mismos resultados. Estas estimaciones proporcionan lo mejor de ambos mundos: estimaciones de distribuciones con masa unitaria, equivarianza y estimaciones de esperanzas con menor varianza que las proporcionadas por la integración básica y la estimación muestral. La aplicación más amplia en la mejora de estimaciones es la *regresión*. Para cualquiera de estas estimaciones, la efectividad u obtención de buenos resultados depende de la distribución muestral elegida para realizar los cálculos.

Para entender el funcionamiento del método, supongamos que se desea calcular la esperanza de una función conocida y que tenemos X una variable aleatoria con distribución normal de media 0 y varianza 1.

Una forma de calcularla sería hacer una estimación directa, es decir, calcular

$$\mathbb{E}(f(X)), \quad \text{con} \quad X \sim N(0, 1),$$

estimando $\mathbb{E}(f(X))$, mediante la siguiente aproximación

$$\mathbb{E}(f(X)) \approx \frac{1}{n} \sum_{i=1}^n f(X_i), \quad (1.8)$$

con n un entero suficientemente grande y X_i el i -ésimo número generado aleatoriamente bajo la distribución normal estándar de manera independiente. Sin embargo grandes fallas pueden ocurrir en el cálculo de esta estimación, pues aunque se puede evadir el tema de qué tan válidos son los números generados y argumentar una regla de selección de un entero n apropiado (ver sección 2.2) o incluso si los números aleatorios son perfectos y n es muy grande, existen funciones muy sencillas para las cuales esta simulación *ingenua* falla aún para la obtención de un dígito significativo.

Como ejemplo de lo anterior, tomemos $f = \mathbb{I}_{\{x \geq 4\}}$ y supongamos que lo que deseamos calcular es $\mathbb{E}(f(x))$ utilizando (1.8). Si pensamos en el número de términos de (1.8) hasta el primer sumando distinto de cero, fácilmente nos daremos cuenta de que el rápido decaimiento en la cola de la distribución de una variable aleatoria normal implica que este número es muy grande (ver **cuadro 1** y **sección 2.5**). Peor aún para garantizar un cierto grado de confianza para que el primer dígito de la estimación sea correcto, probablemente no deberíamos detener nuestra simulación hasta haber obtenido un centenar o más de sumandos distintos de cero. Por lo tanto estas dos observaciones revelan que no es factible realizar el cálculo bajo simulación directa. Esto puede ser un argumento para creer que no se puede calcular $\mathbb{E}(f(x))$ via una simulación, pero nada puede estar más lejos de la realidad, ya que de hecho un apropiado diseño del tipo de simulación nos permitirá obtener un tamaño razonable de n y generar números aleatorios de forma adecuada, a modo de obtener 3 dígitos significativos en sólo unos segundos. Todo lo que necesitamos hacer es encontrar una forma de concentrar nuestros esfuerzos en lo que realmente es importante o significativo para nuestro contexto.

El problema con la simulación directa es que uno puede obtener muestras $\{X_i\}_{i \geq 1}$ que se encuentren lejos del punto en donde la información importante de F está ubicada. Si nuestro problema hubiera sido estimar $\mathbb{E}(g(x))$, donde $g(x) = \mathbb{I}_{\{x \leq 0\}}$, entonces tanto las muestras de normales estandarizadas y lo más informativo sobre la conducta de g podrían ser centralizadas en el cero. En tal caso, la simulación directa podría realizar un trabajo satisfactorio. Esto sugiere que para realizar una simulación adecuada, deberíamos ser capaces de encontrar alguna forma de transformar nuestro problema original en uno en donde las muestras que obtengamos estén concentradas directamente en donde se encuentra la conducta relevante de F .

La idea es buscar la manera de ganar un poco de flexibilidad sobre el lugar

donde se observan las muestras, y afortunadamente existe una forma natural para realizar los cálculos que hacen factible esta idea.

Si denotamos a \mathbb{E}_μ como la esperanza bajo el modelo $N(\mu, 1)$ entonces, es bastante claro que siempre podemos reescribir $\mathbb{E}_0[f(X)]$, con X una variable aleatoria $N(0, 1)$ como

$$\begin{aligned}\mathbb{E}_0[f(X)] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) e^{-\frac{x^2}{2}} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) e^{-\frac{(x-\mu)^2}{2}} e^{\frac{\mu^2}{2}} e^{-\mu x} dx \\ &= \mathbb{E}_\mu[f(X) e^{-\mu X + \frac{\mu^2}{2}}].\end{aligned}\tag{1.9}$$

El argumento fuerte, es que uno tiene una familia paramétrica completa como alternativa a la simulación *ingenua* dada por (1.8), de tal forma podríamos usar simplemente cualquiera de las estimaciones

$$\mathbb{E}_0[f(X)] \approx \frac{1}{n} \sum_{i=1}^n g(Y_i) \quad \text{en donde } g(y) = f(y) e^{-\mu y + \frac{\mu^2}{2}}, \quad y \quad Y_i \sim N(\mu, 1).$$

Como la elección de μ está a nuestra entera disposición, entonces podemos asegurar una mejoría en la eficiencia de la estimación directa (1.8). Por ejemplo, si podemos evaluar el coeficiente de variación, que no es más que una medida de dispersión definida como $c(\mu) = |\mathbb{E}g[Y_1]| / \text{Var}[g(Y_1)]^{\frac{1}{2}}$ y evaluar esta cantidad como función de μ , entonces la elección apropiada de μ será el valor que minimice $c(\mu)$.

Desafortunadamente tal esquema de optimización no es realista, pues como tratamos de usar la simulación para encontrar $\mathbb{E}[f(X)]$, que es igual a $\mathbb{E}[g(Y_1)]$ para $\mu = 0$, entonces si somos en cierto sentido estrictos debemos admitir que no podemos esperar conocer $c(\mu)$. Sin embargo, no nos encontramos tan lejos de saberlo ya que aunque no podamos encontrar el valor óptimo de μ , para obtener una buena estimación, basta con realizar una buena elección de este parámetro basándonos en una intuición de la región en donde se encuentra la información relevante de F , y con esto podemos asegurar que casi siempre mejoraremos nuestra estimación si podemos elegir μ de tal forma que nuestra muestra se cambie cerca de esta zona de importancia. Por ejemplo, en nuestro problema original sería suficiente tomar $\mu = 4$, para mejorar nuestra estimación de manera sustancial.

En la siguiente tabla se muestra una comparación numérica entre el muestreo de importancia y la simulación ingenua, para la estimación de $\mathbb{E}[\mathbb{I}_{\{x \geq 4\}}]$, con $x \sim N(0, 1)$.

Cuadro 1.1: Simulación ingenua contra Muestreo de importancia. El cuadro indica el número de simulaciones (N), el valor estimado (Est.), su varianza (Var.), el número de valores aceptados (M) y un intervalo del 95 % de confianza para la estimación.

| Estimación de $\mathbb{E}[\mathbb{I}_{\{x \geq 4\}}]$, con $x \sim N(0, 1)$ | | | | | |
|--|---------|----------------------------------|----------------------------------|--------|--|
| Método | N | Est. | Var. | M | IC |
| Simulación Ingenua (MCC) | 1000 | 0 | 0 | 0 | (0, 0) |
| | 10000 | 0 | 0 | 0 | (0, 0) |
| | 100000 | $1 \times 10^{-4} \times 0.3$ | $1 \times 10^{-4} \times 0.3999$ | 3 | $1 \times 10^{-4} \times (-0.03947, 0.639478)$ |
| | 1000000 | $1 \times 10^{-4} \times 0.34$ | $1 \times 10^{-4} \times 0.3998$ | 34 | $1 \times 10^{-4} \times (0.22571, 0.639478)$ |
| Muestreo de Importancia | 1000 | $1 \times 10^{-4} \times 0.302$ | $1 \times 10^{-10} \times 0.425$ | 496 | $1 \times 10^{-4} \times (0.26196, 0.34282)$ |
| | 10000 | $1 \times 10^{-7} \times 0.3277$ | $1 \times 10^{-11} \times 0.471$ | 5046 | $1 \times 10^{-4} \times (0.31428, 0.34117)$ |
| | 100000 | $1 \times 10^{-8} \times 0.3164$ | $1 \times 10^{-12} \times 0.45$ | 49809 | $1 \times 10^{-4} \times (0.31228, 0.32059)$ |
| | 1000000 | $1 \times 10^{-9} \times 0.3174$ | $1 \times 10^{-13} \times 0.5$ | 499809 | $1 \times 10^{-4} \times (0.31619, 0.31875)$ |

En este trabajo explicaremos brevemente la estructura del método conocido como *muestreo de importancia* introduciendo algunos conceptos básicos y definiciones para el desarrollo de este documento [Capítulo 2]. Motivaremos el uso de herramientas básicas de la teoría de *Grandes Desviaciones* en aplicaciones estadísticas [Capítulos 3 y 4] y estableceremos el objetivo principal de este documento: utilizar la teoría de grandes desviaciones para crear un esquema de muestreo de importancia óptimo en el estudio de eventos “raros” o de baja ocurrencia [Capítulo 5].

La principal referencia de este trabajo es el artículo escrito por *Paul Glasserman y Yashan Wang* [11], en el que se basa el Capítulo 4 y que motivó el objetivo principal de este documento. La parte introductoria a la teoría de grandes desviaciones tratada en el Capítulo 3, se encuentra desarrollada en el libro de *Richard Durrett* [7], mientras que las aplicaciones mostradas en el Capítulo 5 fueron una recopilación de [14], [9], [2], [10] y las referencias mencionadas en cada uno de los problemas.

Capítulo 2

Muestreo de Importancia.

En este capítulo hablaremos brevemente de las herramientas básicas en **simulación** y **métodos Monte Carlo**. Únicamente nos preocuparemos por mencionar algunos aspectos fundamentales y explicaremos por qué estas herramientas son útiles para el análisis realizado por el **muestreo de importancia (IS)**, lo que es nuestra principal motivación. Mencionaremos también tres criterios de eficiencia que serán importantes para la evaluación de la efectividad del *IS*. Estos criterios serán ampliamente explicados desde un aspecto teórico en el siguiente capítulo. Sin embargo es conveniente ir familiarizándonos con dichos conceptos ya que serán trascendentes en el análisis de la práctica y nos ayudarán a ver cuándo la aplicación del muestreo de importancia bajo un esquema de grandes desviaciones es óptimo.

§2.1. Simulación Monte Carlo.

Los métodos *Monte Carlo* se basan en la analogía existente entre *Probabilidad* y *Volumen*. Las Matemáticas de medida formalizan la noción intuitiva de *Probabilidad* al asociar a un evento con un conjunto de resultados posibles y posteriormente definir la *Probabilidad* del evento como el volumen o medida relativo al universo de posibles resultados. La simulación *Monte Carlo* usa esta identidad de manera inversa, calcula primero el volumen del conjunto para interpretarlo como una probabilidad. En el más simple de los casos, esto significa muestrear aleatoriamente a partir de un universo de posibles resultados y tomar la fracción aleatoria que cae en un conjunto dado como una estimación del volumen del mismo conjunto. La *Ley Fuerte de los Grandes Números* asegura que esta estimación converge al valor correcto conforme la fracción aleatoria de resultados simulados crece. El *Teorema*

del *Límite Central* provee información acerca de la probable magnitud del error en la estimación después de un número finito de simulaciones.

En nuestro caso particular nos interesa realizar simulaciones sobre esperanzas o probabilidades de eventos “raros”, por lo cual nos restringiremos a realizar simulaciones via *Monte Carlo* de la siguiente forma:

Sea Z una variable aleatoria (v.a.), y supongamos que lo que deseamos simular es $\zeta = \mathbb{E}[Z]$ en una situación tal que ζ no sea fácil de expresar analíticamente pero pueda ser simulada. El método *Monte Carlo Crudo* (MCC) consiste en simular repeticiones independientes e idénticamente distribuidas (i.i.d) de Z, Z_1, \dots, Z_N , y estimar ζ via la media empírica

$$\bar{\zeta} = \frac{Z_1 + \dots + Z_N}{N},$$

y de manera análoga estimar la varianza de Z utilizando la varianza empírica

$$s^2 = \frac{1}{N} \sum_{i=1}^N (Z_i - \bar{\zeta})^2 = \frac{1}{N} \sum_{i=1}^N Z_i^2 - \bar{\zeta}^2.$$

De este modo, si consideramos el problema de estimar

$$a = \mathbb{E}[h(X)] = \int h(x)f(x)dx,$$

donde X es una v.a. en \mathbb{R}^d con densidad f y $h : \mathbb{R}^d \rightarrow \mathbb{R}$, entonces para realizar nuestra estimación via *Monte Carlo*, proponemos un estimador de la forma

$$\hat{a} = \frac{1}{n} \sum_{i=1}^n h(X_i),$$

con X_1, \dots, X_n v.a.i.i.d. con distribución f . Este estimador a pesar de tener una forma sencilla, basa su estructura en la *Ley Fuerte de los Grandes Números*, con lo que aseguramos un comportamiento asintótico de \hat{a} sobre a para muestras de gran tamaño. Sin embargo, como ya se ha mencionado previamente, este tipo de estimadores no presentan un buen desempeño en algunos casos. Existe una gran variedad de alteraciones a este tipo de estimadores para así mejorar los resultados, en la mayor parte de las ocasiones estas mejorías buscan reducir la varianza de las estimaciones y con esto disminuir el costo de la simulación.

§2.2. Teorema del Límite Central e Intervalos de Confianza.

El resultado que a continuación se menciona, es uno de los teoremas límite más importantes de la teoría de probabilidad, y una de las herramientas más utilizadas para realizar estadística. Para fines de este documento, la prueba de este teorema será referida a cualquier libro de teoría de probabilidad, en particular se recomienda al lector consultar [6]. El objetivo de enunciar este importante resultado es dar una expresión cerrada para el tamaño de muestra requerido para alcanzar cierta precisión en las estimaciones realizadas via Monte Carlo.

La versión más sencilla y conocida del *Teorema del Límite Central* establece lo siguiente: Si X_1, X_2, \dots son variables aleatorias independientes e idénticamente distribuidas, con media μ y varianza σ^2 ($0 < \sigma < \infty$). Entonces la media muestral

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i,$$

satisface que

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \rightarrow N(0, 1), \quad (2.1)$$

donde la convergencia es en distribución. Este hecho se prueba al mostrar que la función característica de la expresión de la izquierda converge a la función característica de una variable aleatoria con distribución normal estándar. Si X_1, X_2, \dots son vectores aleatorios independientes e idénticamente distribuidos con media $\underline{\mu}$ y matriz de varianzas y covarianzas Σ , entonces

$$\sqrt{n}(\bar{X}_n - \underline{\mu}) \rightarrow N(\underline{0}, \Sigma),$$

con $N(\underline{0}, \Sigma)$ denota la distribución normal multivariada con vector de medias cero y matriz de varianzas y covarianzas Σ . Aquí, nuevamente la convergencia es en distribución y este resultado puede deducirse fácilmente a partir de (2.1) considerando todas las combinaciones lineales del vector \bar{X}_n .

De la definición de convergencia en distribución, (2.1) significa que para toda $x \in \mathbb{R}$

$$\mathbb{P}\left\{\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq x\right\} \rightarrow \Phi(x),$$

en donde Φ representa la distribución acumulativa de una normal estándar.

De modo que la probabilidad de que un intervalo de la forma

$$\left(\bar{X}_n - c_1 \frac{\sigma}{\sqrt{n}}, \bar{X}_n + c_2 \frac{\sigma}{\sqrt{n}} \right),$$

con $0 < c_1, c_2 < \infty$, contenga a μ se aproxima a $\Phi(c_2) - \Phi(-c_1)$ conforme n tiende a infinito. Se puede elegir a c_1 y c_2 de modo que este límite sea igual a $1 - \delta$, para algún $\delta > 0$.

De entre todas las elecciones posibles de c_1 y c_2 para las cuales $\Phi(c_2) - \Phi(-c_1) = 1 - \delta$, los valores que minimizan la longitud del intervalo $(-c_1, c_2)$ son $c_1 = c_2 = q_{\delta/2}$, donde $q_{\delta/2}$ es el cuantil $\delta/2$ de la distribución normal estándar que resuelve la ecuación $1 - \Phi(q_{\delta/2}) = \delta/2$. Por lo tanto el intervalo

$$\bar{X}_n \pm q_{\delta/2} \frac{\sigma}{\sqrt{n}} \quad (2.2)$$

contiene a μ con probabilidad aproximadamente igual $1 - \delta$ conforme n tiende a infinito y es en este sentido un intervalo de confianza asintóticamente válido para μ . Por la *Ley Fuerte de los Grandes Números* sabemos que

$$s_n = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2} \rightarrow \sigma.$$

Lo que implica que s_n es un estimador consistente de σ , y ya que $\sigma > 0$ podemos modificar a s_n de modo que siempre sea positivo sin afectar la propiedad de consistencia del estimador, y obtener que σ/s_n converge a 1. Entonces, utilizando este hecho y (2.1), tenemos la siguiente convergencia en distribución

$$\frac{\bar{X}_n - \mu}{s_n/\sqrt{n}} \rightarrow N(0, 1).$$

De modo que

$$\bar{X}_n \pm q_{\delta/2} \frac{s_n}{\sqrt{n}}, \quad (2.3)$$

es un $1 - \delta$ intervalo de confianza asintóticamente válido. Como nuestro objetivo es proveer información sobre la precisión del estimador en términos del tamaño de muestra, utilizamos (2.2) para mostrar que el número de repeticiones necesarias para alcanzar un intervalo de confianza de la mitad del ancho de ϵ es

$$n_\epsilon = \frac{q_{\delta/2}^2 \sigma^2}{\epsilon^2}. \quad (2.4)$$

Si σ es desconocida, entonces se utiliza un procedimiento de dos etapas en donde la primera consiste en usar un conjunto inicial de repeticiones para estimar dicho valor, y posteriormente utilizarlo en (2.4) para estimar el tamaño total de la muestra. Cabe mencionar que n_ϵ crece en función de que tan pequeño sea el error ϵ que se desea admitir.

En (2.1) solamente se ha presentado la forma más elemental del *Teorema del Límite Central*. La media muestral y otros estimadores son asintóticamente normales, bajo condiciones más generales. El resultado elemental aquí mostrado, es suficiente cada vez que se simulan repeticiones independientes e idénticamente distribuidas, no obstante es importante mencionar que para tratar con otro tipo de condiciones que nacen en el ambiente de la simulación, se necesita de resultados más generales. Por mencionar algunas de estas condiciones podemos decir que:

- Las técnicas de reducción de varianza muchas veces introducen dependencia a través de las repeticiones. En algunos casos la dependencia es insignificante conforme el número de repeticiones crece, y en los casos en donde esto no ocurre, simular conjuntos y permitir dependencia entre ellos, pero no dentro de ellos, reduce el problema a uno de repeticiones independientes. Aún cuando esta limitante no se presente, una versión más general del teorema del límite central es necesaria. Como ejemplo de lo anterior se recomienda al lector revisar la discusión establecida en el análisis de salida de la sección 4.4 de [9].
- A menudo nuestro interés se enfoca en como el error de un estimador cambia conforme algún parámetro de la simulación cambia en función del crecimiento del número de repeticiones. Por ejemplo, cuando se discretiza un modelo con paso de tiempo h , se puede analizar la convergencia de un estimador conforme h tiende a cero y n tiende a infinito. Esto se debe a que cambiar h , cambia la distribución dependiendo de donde sea muestreada la repetición. Casos como este requieren de un teorema del límite central para arreglos de variables aleatorias en lugar de uno para sucesiones (ver sección 7.2 de [4]).

§2.3. Reducción de Varianza.

El propósito de las técnicas de reducción de varianza es encontrar un estimador MCC Z de ζ y posteriormente transformar Z en un estimador alternativo Z'

que cumpla que $\mathbb{E}[Z'] = \mathbb{E}[Z] = \zeta$ y $\mathbb{V}ar[Z'] < \mathbb{V}ar[Z]$. Este problema es un área clásica en la literatura de simulación, y muchas ideas sofisticadas se han desarrollado a lo largo de los años. La reducción de varianza involucra una idea teórica (en algunos casos, cálculos matemáticos), un esfuerzo de programación y un menor tiempo para producir una simulación. Por lo tanto, uno podría argumentar que a menos de que la $\mathbb{V}ar[Z']$ sea considerablemente menor que la $\mathbb{V}ar[Z]$, la reducción de varianza difícilmente valdría la pena.

Por ejemplo, si tenemos que la $\mathbb{V}ar[Z'] = \frac{1}{2}\mathbb{V}ar[Z]$, entonces al reemplazar el número de repeticiones N por $2N$ tendremos la misma precisión para el método MCC que cuando simulamos $N' = N$ repeticiones de Z' y en la mayoría de los casos, este modesto incremento de N no provoca problema alguno. Existen muchos métodos de reducción de varianza utilizados para la estimación de eventos “raros” o poco probables, algunos de ellos son: **Ajustes de Regresión, Control de Variedades, Números Aleatorios Comunes** y el **Método Monte Carlo Condicional**. Sin embargo, nosotros enfocaremos nuestro estudio en el *Muestreo de Importancia (IS)*, ya que es de gran utilidad para simular probabilidades de baja ocurrencia y está ampliamente ligado con la *Teoría de Grandes Desviaciones*, además de ser una de las técnicas más sofisticadas y eficientes que se han estudiado hasta el momento para hacer estimaciones de eventos tales como la *Ruina*.

§2.4. El Muestreo de Importancia.

El IS consiste en calcular $\zeta = \mathbb{E}[Z]$ simulando bajo una medida de probabilidad diferente de la inicial, es decir que si \mathbb{P} representa nuestra medida inicial y $\tilde{\mathbb{P}}$ es la nueva medida de probabilidad con la propiedad de que exista una v.a. L tal que

$$\zeta = \mathbb{E}_{\mathbb{P}}[Z] = \mathbb{E}_{\tilde{\mathbb{P}}}[LZ], \quad (2.5)$$

donde $\mathbb{E}_{\mathbb{P}}$ es el operador esperanza bajo \mathbb{P} y $\mathbb{E}_{\tilde{\mathbb{P}}}$ es la esperanza bajo $\tilde{\mathbb{P}}$. Al implementar el método MCC para generar parejas $(Z_1, L_1), \dots, (Z_N, L_N)$ a partir de $\tilde{\mathbb{P}}$, utilizamos el estimador

$$\hat{\zeta}_{\tilde{\mathbb{P}}} = \frac{1}{N} \sum_{i=1}^N L_i Z_i.$$

Por lo tanto para cumplir con (2.5), la obvia posibilidad para nuestra nueva medida, es tomar $\tilde{\mathbb{P}}$ equivalente a \mathbb{P} y viceversa, es decir que \mathbb{P} y $\tilde{\mathbb{P}}$ deben ser mutuamente equivalentes. Posteriormente debemos definir a $L = d\mathbb{P}/d\tilde{\mathbb{P}}$ como la

razón de verosimilitudes, herramienta de la que hablaremos un poco más adelante.

Como es de suponerse la reducción de variancia puede o no ocurrir, ya que todo depende de la elección de $\tilde{\mathbb{P}}$, y por ende el problema radica en hacer una elección adecuada de esta nueva medida. Para este fin, una observación crucial es que la elección óptima para el cambio de medida $\tilde{\mathbb{P}}$ es tal que $d\tilde{\mathbb{P}}/d\mathbb{P} = Z/\mathbb{E}[Z] = Z/\zeta$, es decir que $L = \zeta/Z$ (donde el evento $\{Z = 0\}$ no altera el análisis pues $\tilde{\mathbb{P}}\{Z = 0\} = 0$), por lo tanto

$$\text{Var}_{\tilde{\mathbb{P}}}[LZ] = \mathbb{E}_{\tilde{\mathbb{P}}}[(LZ)^2] - [\mathbb{E}_{\tilde{\mathbb{P}}}[LZ]]^2 = \mathbb{E}_{\tilde{\mathbb{P}}}\left[\frac{\zeta^2}{Z^2}Z^2\right] - \mathbb{E}_{\tilde{\mathbb{P}}}[Z]^2 = \zeta^2 - \zeta^2 = 0.$$

Aparentemente bajo esta elección hemos producido un estimador de variancia cero, sin embargo, el argumento es un tanto engañoso ya que hemos simulado bajo el supuesto de que ζ no es factible analíticamente, de tal manera que no podemos calcular $L = Z/\zeta$. No obstante, aún cuando el cambio de medida óptimo no es factible, proporciona una clara guía que nos indica elegir $\tilde{\mathbb{P}}$ tal que $d\tilde{\mathbb{P}}/d\mathbb{P}$ sea proporcional a Z tanto como sea posible. Esto puede ser difícil de valorar, pero tentativamente uno puede tomar \mathbb{P} de tal forma que los valores grandes de Z sean más probables.

§2.5. Simulación de Eventos Raros.

El problema consiste en estimar $\mathbb{P}\{A\}$ cuando esta probabilidad es muy pequeña, es decir que A es un evento “raro” o poco probable. De este modo, se propone estimar $\mathbb{P}\{A\}$ a través de $Z = \mathbb{I}_{\{A\}}$ para efectos de simulación y entonces calcular $\zeta = \mathbb{E}[Z]$.

Si estamos en el contexto que envuelve el problema de la Ruina (por mencionar uno de estos eventos de baja ocurrencia), el evento A puede describirse de la siguiente forma: $A = \{\inf\{n \geq 0 : S_n > u\} \leq T\}$ si es que estamos ubicados en un horizonte de tiempo finito (T) y como $A = \{\inf\{n \geq 0 : S_n > u\} < \infty\}$ para un horizonte de tiempo infinito, en donde en cada una de estas representaciones se inicia con un capital igual a u . Más adelante se describirán a detalle los conceptos aquí nombrados, pero mientras tanto es de gran utilidad conocer a grandes rasgos el trato de este tipo de eventos. El método MCC proporciona una variancia igual a $\sigma_Z^2 = \zeta(1 - \zeta)$, cantidad que tiende a cero conforme ζ tiende a cero. Sin embargo, a pesar de que esta precisión es buena, la precisión relativa es mala, pues

$$\frac{\sigma_Z}{\zeta} = \frac{\sqrt{\zeta(1-\zeta)}}{\zeta} \sim \frac{1}{\zeta} \longrightarrow \infty, \quad \text{conforme } \zeta \rightarrow 0.$$

En otras palabras, si ζ es sumamente pequeña esta estimación es inútil para obtener información sobre nuestro evento en cuestión, o visto desde el aspecto económico sería considerablemente costoso realizar un estudio basado en esta estimación, ya que el número de simulaciones requeridas para alcanzar una cierta efectividad es considerablemente grande. Para ilustrar este problema en términos del tamaño de muestra N necesario para alcanzar una relativa precisión, digamos del 10 %, sabemos que a partir del teorema del límite central, esto se reduce a resolver la siguiente ecuación

$$q_{\delta/2} \frac{\sigma_Z}{\zeta \sqrt{N}} = 0.1,$$

lo que implica que

$$N = \frac{100 \cdot q_{\delta/2}^2 \zeta (1 - \zeta)}{\zeta^2} \sim \frac{100 \cdot q_{\delta/2}^2}{\zeta},$$

y como $q_{\delta/2}$ es una constante igual al cuantil 0.05 de una distribución normal estándar, el valor de N crece conforme ζ tiende a cero. Por lo tanto, si ζ es muy pequeña, el tamaño de muestra necesario para alcanzar esta precisión relativa será muy grande.

A partir de este momento y hasta el final de este trabajo, adoptaremos el **IS** como forma potencial (aunque no la única) para lidiar con este tipo de problemas. Sabemos que el cambio de medida proporcionado por el estimador de varianza cero es

$$\tilde{\mathbb{P}}\{B\} = \mathbb{E}\left[\frac{Z}{\zeta}; B\right] = \frac{1}{\zeta} \mathbb{P}\{AB\} = \mathbb{P}\{B|A\}.$$

Es decir que el cambio óptimo (por ser el de menor varianza) $\tilde{\mathbb{P}}$ es la distribución condicional dado A . No obstante, no contamos con conocimiento alguno sobre ζ lo cual nos impide hacer el cálculo de la razón de verosimilitudes y por lo tanto tampoco podemos obtener el **IS**-estimador, es decir, el estimador propuesto por el muestreo de importancia.

Usualmente no es práctico simular a partir de la probabilidad condicional $\mathbb{P}\{\cdot|A\}$, por lo que trataremos de hacer nuestra elección de $\tilde{\mathbb{P}}$ de forma tan parecida a $\mathbb{P}\{\cdot|A\}$ como sea posible.

En lo referente a la efectividad de la simulación de eventos “raros” se establecen tres criterios de eficiencia como los más comunes y útiles para evaluar el

desempeño de estimaciones de eventos de baja probabilidad. El primero de ellos se denomina *Eficiencia Asintótica*. Este concepto, como observaremos en el siguiente capítulo, esta ampliamente ligado al análisis de *grandes desviaciones* sobre el comportamiento de eventos “raros”, y se define como

$$\limsup_{x \rightarrow \infty} \left\{ -\frac{1}{x} \log \mathbb{E}_{\tilde{\mathbb{P}}}[\hat{\alpha}^2(x)] \right\} = 2\gamma, \quad (2.6)$$

en donde $\hat{\alpha} = L_x \mathbb{I}_{\{A_x\}}$, A_x representa el evento o fenómeno de baja probabilidad, L_x es la razón de verosimilitudes, $\tilde{\mathbb{P}}$ es el cambio de medida seleccionado y

$$\gamma = -\lim_{x \rightarrow \infty} \frac{1}{x} \log \mathbb{P}\{A_x\}.$$

El segundo criterio se denomina *Eficiencia logarítmica*, y basa su fuerza en un sentido asintótico que implica acercarse “tanto como se desee” a la potencia 2, de manera que $\text{Var}_{\tilde{\mathbb{P}}}[\hat{\alpha}(x)]$ debe tender a 0 al menos tan rápido como $\mathbb{P}\{A_x\}^{2-\epsilon}$ para algún $\epsilon > 0$, es decir

$$\limsup_{x \rightarrow \infty} \frac{\text{Var}_{\tilde{\mathbb{P}}}[\hat{\alpha}(x)]}{\mathbb{P}\{A_x\}^{2-\epsilon}} < \infty. \quad (2.7)$$

Esto implica que la varianza del estimador (calculada bajo la medida de cambio $\tilde{\mathbb{P}}$) decrece ligeramente más lento que $\mathbb{P}\{A_x\}^2$. En la práctica este criterio es conocido con el nombre de *Eficiencia Logarítmica* debido a su expresión equivalente

$$\liminf_{x \rightarrow \infty} \frac{-\log(\text{Var}_{\tilde{\mathbb{P}}}[\hat{\alpha}(x)])}{-\log(\mathbb{P}\{A_x\})} \geq 2 \quad (2.8)$$

Proposición 1. *Las expresiones (2.7) y (2.8) son equivalentes.*

Demostración

(2.7) se cumple si y sólo si para toda $\epsilon' > 0$, existe $x_0 > 0$ tal que para todo $x > x_0$

$$\frac{\text{Var}_{\tilde{\mathbb{P}}}[\hat{\alpha}(x)]}{\mathbb{P}\{A_x\}^{2-\epsilon}} \leq k + \epsilon',$$

con k constante. Entonces, utilizando que la función logaritmo es creciente

$$\log \text{Var}_{\tilde{\mathbb{P}}}[\hat{\alpha}(x)] \leq \log(k + \epsilon') + \log(\mathbb{P}\{A_x\}^{2-\epsilon}).$$

Por lo tanto

$$\frac{-\log \text{Var}_{\tilde{\mathbb{P}}}[\hat{\alpha}(x)] + \log(k + \epsilon')}{-\log(\mathbb{P}\{A_x\})} \geq 2 - \epsilon.$$

Como el denominador de la expresión del lado izquierdo es positivo, aseguramos que

$$\liminf_{x \rightarrow \infty} \frac{-\log \text{Var}_{\tilde{\mathbb{P}}}[\hat{\alpha}(x)]}{-\log(\mathbb{P}\{A_x\})} + \liminf_{x \rightarrow \infty} \frac{\log(k + \epsilon')}{-\log(\mathbb{P}\{A_x\})} \geq 2 - \epsilon.$$

Es claro que el segundo término de la suma del lado izquierdo de la ecuación, tiende a cero conforme x crece a infinito, dando como resultado

$$\liminf_{x \rightarrow \infty} \frac{-\log \text{Var}_{\tilde{\mathbb{P}}}[\hat{\alpha}(x)]}{-\log(\mathbb{P}\{A_x\})} \geq 2 - \epsilon.$$

Ya que todos los pasos a quí mostrados son un si sólo si y la elección de ϵ es arbitraria, se concluye el resultado.

◇

El tercer criterio, es probablemente el más utilizado en la práctica y como se puede observar en la siguiente definición, también está relacionado con la conducta del segundo momento del estimador en cuestión. Diremos que un estimador tiene *Error Relativo Acotado* si satisface

$$\limsup_{x \rightarrow \infty} \frac{\sqrt{\mathbb{E}_{\tilde{\mathbb{P}}}[\hat{\alpha}^2(x)]}}{\mathbb{P}\{A_x\}} < \infty \quad (2.9)$$

En la literatura de simulación existen varias definiciones alternativas para decir cuando un estimador tiene error relativo acotado o no, sin embargo todas coinciden en la siguiente condición: se dice que un estimador tiene *error relativo acotado* si

$$\lim_{x \rightarrow \infty} \frac{\text{Var}_{\tilde{\mathbb{P}}}[\hat{\alpha}(x)]}{\mathbb{P}\{A_x\}^2} < \infty. \quad (2.10)$$

Ambas definiciones nos dicen que de cumplirse, la varianza del estimador en cuestión no crece desmesuradamente respecto a su valor esperado (bajo $\tilde{\mathbb{P}}$). Como se puede observar, tanto (2.9) como (2.10) se basan en mantener el crecimiento del segundo momento del estimador controlado.

Proposición 2. (2.10) implica (2.9).

Demostración

Si (2.10) se cumple, implica que

$$\limsup_{x \rightarrow \infty} \frac{\mathbb{E}_{\tilde{\mathbb{P}}}[\hat{\alpha}^2(x)] - \mathbb{E}_{\tilde{\mathbb{P}}}^2[\hat{\alpha}(x)]}{\mathbb{P}\{A_x\}^2} < \infty.$$

Entonces, utilizando que $\mathbb{P}\{A_x\} = \mathbb{E}_{\tilde{\mathbb{P}}}[\hat{\alpha}(x)]$ tenemos

$$\limsup_{x \rightarrow \infty} \frac{\mathbb{E}_{\tilde{\mathbb{P}}}[\hat{\alpha}^2(x)]}{\mathbb{P}\{A_x\}^2} - 1 < \infty,$$

de modo que existe $x_0 > 0$ tal que para todo $x > x_0$ y $\epsilon > 0$

$$\frac{\mathbb{E}_{\tilde{\mathbb{P}}}[\hat{\alpha}^2(x)]}{\mathbb{P}\{A_x\}^2} < k + 1 + \epsilon$$

con k constante. Utilizando que la raíz cuadrada es una función creciente se obtiene

$$\frac{\sqrt{\mathbb{E}_{\tilde{\mathbb{P}}}[\hat{\alpha}^2(x)]}}{\mathbb{P}\{A_x\}} < (k + 1 + \epsilon)^{1/2},$$

con lo que se concluye que

$$\limsup_{x \rightarrow \infty} \frac{\sqrt{\mathbb{E}_{\tilde{\mathbb{P}}}[\hat{\alpha}^2(x)]}}{\mathbb{P}\{A_x\}} < \infty$$

◇

El recíproco **no** es cierto ya que de manera análoga al procedimiento antes expuesto, sólo podemos garantizar que

$$\limsup_{x \rightarrow \infty} \frac{\sqrt{\mathbb{E}_{\tilde{\mathbb{P}}}[\hat{\alpha}^2(x)]}}{\mathbb{P}\{A_x\}} < \infty \quad \Rightarrow \quad \limsup_{x \rightarrow \infty} \frac{\text{Var}_{\tilde{\mathbb{P}}}[\hat{\alpha}(x)]}{\mathbb{P}\{A_x\}^2} < \infty.$$

No obstante, si la condición (2.9) se satisface, se puede garantizar un crecimiento “controlado” de la varianza del estimador $\hat{\alpha}$. Es por esto que para efectos prácticos y del desarrollo de este trabajo, basta con verificar (2.9) para medir la eficiencia de un estimador monte carlo en términos de varianza.

Más adelante trataremos con estos conceptos como herramientas básicas para medir la eficiencia del muestreo de importancia y estableceremos condiciones para que los estimadores propuestos en las secciones subsecuentes cumplan estas propiedades, así como también enfatizaremos su uso en el estudio de casos específicos.

§2.6. Razón de Verosimilitudes.

Hasta este momento hemos tratado con el concepto de Razón de verosimilitudes sin haberlo definido propiamente, sin embargo es claro que el objetivo primordial de esta razón es asignar pesos a nuestras variables simuladas y que la correcta elección de este cociente reeditara en una simulación eficiente. Para aclarar estas ideas trataremos con funciones de variables aleatorias y no sólo con variables aleatorias como en un principio, en el entendido de que una función de una variable aleatoria no es más que otra variable aleatoria. Entonces si recordamos nuestro problema central, lo que deseamos es estimar

$$a = \mathbb{E}[h(X)] = \int h(x)f(x)dx,$$

con h una función medible, f la densidad de la v.a. X . Si g es cualquier otra función de densidad en \mathbb{R}^d que cumpla que $f(x) > 0$ implica que $g(x) > 0$ para toda $x \in \mathbb{R}^d$, entonces nuestra representación alternativa de a proporcionada por el **IS** es:

$$a = \int h(x) \frac{f(x)}{g(x)} g(x) dx.$$

Esta integral puede interpretarse como la esperanza de un producto de funciones con respecto de la densidad g , de tal forma que podemos reescribir esta expresión como

$$a = \mathbb{E}_g \left[h(X) \frac{f(X)}{g(X)} \right],$$

donde \mathbb{E}_g representa la esperanza bajo la densidad g . Si X_1, \dots, X_n son simulaciones independientes bajo g , el **IS**-estimador asociado a g es

$$\hat{a}_g = \hat{a}_g(n) = \frac{1}{n} \sum_{i=1}^n h(X_i) \frac{f(X_i)}{g(X_i)} \quad (2.11)$$

El peso $f(X_i)/g(X_i)$ es lo que hemos definido todo este tiempo como la *Razón de verosimilitudes*, que no es otra cosa más que la *derivada de Radon-Nikodym* de f con respecto a g , evaluada en X_i . Como se puede observar, lo único que hemos hecho aquí es pasar del uso de variables aleatorias a funciones de variables aleatorias, y por lo tanto todo lo propuesto para el caso de variables es análogo para funciones de variables.

Hasta ahora hemos supuesto por simplicidad, que X es \mathbb{R}^d -valuada, pero estas ideas pueden ser extendidas para X que toma valores en conjuntos más generales. También, hemos supuesto que X tiene densidad f , pero las mismas observaciones son aplicables si f es una función de masa de probabilidades (o en un contexto más general, podemos tomar una densidad con respecto a alguna medida de referencia en \mathbb{R}^d , diferente de la medida de *Lebesgue*).

§2.7. Cambio Exponencial de Medida.

A continuación hablaremos del cambio de medida más socorrido en el estudio del *Muestreo de Importancia*, el *Cambio Exponencial de Medida*, también conocido como *Medida de Gibbs*. Este cambio de medida es de mucha utilidad debido a la gran variedad de distribuciones que abarca y a la “facilidad” de su cálculo.

Para una función de distribución F en \mathbb{R} , definamos a

$$\psi(\theta) = \log \int_{-\infty}^{\infty} e^{\theta x} dF(x) = \log \mathbb{E}[e^{\theta x}].$$

A esta función se le da el nombre de función generadora acumulativa de F o simplemente logaritmo de la transformada de *Laplace* de F . Sea $\Theta = \{\theta : \psi(\theta) < \infty\}$ y supongamos además que Θ es no vacío, entonces para cada $\theta \in \Theta$ definamos ahora

$$F_{\theta}(x) = \int_{-\infty}^x e^{\theta u - \psi(\theta)} dF(u).$$

Cada una de estas F_{θ} es una función de distribución de probabilidades, y el conjunto $\{F_{\theta}, \theta \in \Theta\}$ forma lo que se denomina *Familia Exponencial* de distribuciones, mientras que a la transformación de F a F_{θ} se le conoce como cambio exponencial de medida. Si F tiene densidad f , entonces F_{θ} tiene densidad

$$f_{\theta}(x) = e^{\theta x - \psi(\theta)} f(x).$$

Supongamos ahora que X_1, \dots, X_n son v.a.i.i.d con distribución común F y que aplicamos un cambio exponencial de medida bajo el cual estas variables se conservan i.i.d pero ahora con distribución común F_θ , entonces la razón de verosimilitudes para esta transformación es

$$\prod_{i=1}^n \frac{dF_0(x_i)}{dF_\theta(x_i)} = \exp\left\{-\theta \sum_{i=1}^n x_i + n\psi(\theta)\right\}. \quad (2.12)$$

Una propiedad importante del cambio exponencial de medida es que la razón de verosimilitudes, que es en principio función de X_1, \dots, X_n se reduce a una función de la suma de X_i 's con $i = 1, \dots, n$, lo cual en términos estadísticos se denomina como *Estadístico Suficiente*⁵ para θ .

La función ψ almacena gran parte de la información relevante de la distribución F_θ , además de ser una función relativamente fácil de maniobrar. Un ejemplo claro de esto es que al derivar $\psi(\theta)$ con respecto a θ , obtenemos la media de F_θ . Denotemos a \mathbb{E}_θ como la esperanza respecto a la distribución F_θ y notemos que $\psi(\theta) = \log \mathbb{E}[\exp(\theta X)]$, con \mathbb{E} igual a la esperanza bajo F . Por lo tanto al diferenciar obtenemos

$$\psi'(\theta) = \frac{\mathbb{E}[X e^{\theta X}]}{\mathbb{E}[e^{\theta X}]} = \mathbb{E}[X e^{\theta X - \psi(\theta)}] = \mathbb{E}_\theta[X].$$

Cálculos similares se pueden hacer sobre las siguientes derivadas de ψ y así verificar por ejemplo que $\psi''(\theta)$ es la varianza de F_θ . La función ψ pasa por el origen y la desigualdad de Hölder nos muestra que es una función convexa, de tal forma que ψ'' es positiva, (para mayores detalles sobre los aspectos teóricos de la *Familia Exponencial*, ver Barndorff-Nielson [3]).

⁵Un Estadístico T es suficiente para θ si la distribución condicional de la v.a. X dado $\{T = t\}$ no depende de θ .

Capítulo 3

Grandes Desviaciones y Muestreo de Importancia.

En el presente capítulo desarrollaremos la parte fundamental de la *Teoría de Grandes Desviaciones* marcando especial énfasis en la obtención de la cota inferior de las grandes desviaciones, ya que esta cota se encuentra al realizar cierto cambio de medida que como podremos observar en los capítulos subsecuentes, será el cambio de medida utilizado en el **IS** bajo el esquema que propone este documento. Además notaremos que este cambio de medida mantiene una fuerte relación con uno de los criterios de eficiencia mencionados en el capítulo anterior, la *Eficiencia Asintótica*, lo que en la práctica será de gran utilidad para garantizar la obtención de un método que simule eventos poco probables con gran efectividad al menos asintóticamente.

La exposición de este capítulo se basa fundamentalmente en [7] y recomendamos al lector consultar [5] para mayor comprensión sobre la teoría antes mencionada.

§3.1. Motivación e Idea Heurística.

Sean X_1, X_2, \dots , variables aleatorias independientes idénticamente distribuidas con distribución común F y media finita ($\mu < \infty$). Tomemos un número $a > 0$ tal que $a > \mathbb{E}(X_1)$. Entonces por la *Ley Débil de los Grandes Números*, la probabilidad del evento $\{X_1 + \dots + X_n > na\}$ decrece a cero conforme n tiende a infinito.

Nuestro interés es saber qué tan rapido decrece esta probabilidad, por lo que para tratar de responder esto calcularemos lo siguiente:

Sea k un entero tal que n/k es entero, entonces $X_{(jk)+1} + \dots + X_{(j+1)k} \geq ak$ para todo $j = 0, \dots, (\frac{n}{k}) - 1$, lo que implica que

$$\begin{aligned} \mathbb{P}\left\{\frac{X_1 + \dots + X_n}{n} \geq a\right\} &= \mathbb{P}\{X_1 + \dots + X_n \geq an\} \\ &\geq \mathbb{P}\{X_{(jk)+1} + \dots + X_{(j+1)k} \geq ak \quad \forall \quad j = 0, \dots, (\frac{n}{k}) - 1\}. \end{aligned}$$

Es decir que

$$\begin{aligned} \mathbb{P}\left\{\frac{X_1 + \dots + X_n}{n} \geq a\right\} &\geq \mathbb{P}\{X_{(jk)+1} + \dots + X_{(j+1)k} \geq ak \quad \forall \quad j = 0, \dots, (\frac{n}{k}) - 1\} \\ &= \mathbb{P}\{X_1 + \dots + X_k \geq ak\}^{\frac{n}{k}}, \end{aligned}$$

por independencia de las X_i 's. De modo que la tasa de convergencia debe ser a lo más exponencial. Esto nos lleva a pensar que la probabilidad del evento

$$\left\{\left|\frac{X_1 + \dots + X_n}{n} - \mathbb{E}\left[\frac{X_1 + \dots + X_n}{n}\right]\right| > \epsilon\right\},$$

decrece de manera **Exponencial**.

Las *Grandes Desviaciones* analizan el comportamiento asintótico de sistemas estocásticos, razón por la cual se adecuan perfectamente al tratamiento de eventos de baja probabilidad y como hemos visto hasta ahora el muestreo de importancia se basa esencialmente en encontrar un cambio de medida óptimo para realizar estimaciones. A continuación mostraremos cómo la cota inferior de las *grandes desviaciones* nos brinda las bases teóricas para realizar un cambio de medida y obtener aproximaciones óptimas en un sentido asintótico.

§3.2. Grandes Desviaciones.

En esta sección daremos una prueba del *Teorema de Chernoff* en \mathbb{R} , también conocido como el *Teorema de Cramér*, para variables aleatorias independientes e idénticamente distribuidas.

Sean $n, m \in \mathbb{N}$, X_1, X_2, \dots v.a.i.i.d, identifiquemos a

$$\gamma(a) = \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\{S_n \geq na\},$$

para un $a > \mu = \mathbb{E}(X_i)$.

Nuestro primer paso es probar que este límite existe, por lo cual necesitamos hacer la siguiente observación. Si definimos a $\Pi_n = \mathbb{P}\{S_n \geq na\}$ entonces

$$\Pi_{n+m} = \mathbb{P}\{S_{n+m} \geq (n+m)a\} \geq \mathbb{P}\{S_m \geq ma, S_{n+m} - S_m \geq na\}.$$

Al utilizar la propiedad de independencia obtenemos que

$$\mathbb{P}\{S_m \geq ma, S_{n+m} - S_m \geq na\} = \mathbb{P}\{S_m \geq ma\} \mathbb{P}\{S_n \geq na\} = \Pi_m \Pi_n.$$

Definamos ahora a

$$\gamma_n = \frac{1}{n} \log \Pi_n = \frac{1}{n} \log \mathbb{P}\{S_n \geq na\},$$

para transformar así el producto en una adición.

Con el siguiente lema, veremos que el límite $\gamma(a)$ efectivamente existe.

Lema 1. Si $\gamma_{n+m} \geq \gamma_n + \gamma_m$, entonces

$$\lim_{n \rightarrow \infty} \frac{\gamma_n}{n} = \sup_{m \geq 1} \left\{ \frac{\gamma_m}{m} \right\}.$$

Demostración

Claramente el $\limsup_{n \rightarrow \infty} \{\gamma_n/n\} \leq \sup_{m \geq 1} \{\gamma_m/m\}$, por lo que para completar nuestra prueba es suficiente demostrar que para toda $m \in \mathbb{N}$,

$$\liminf_{n \rightarrow \infty} \frac{\gamma_n}{n} \geq \frac{\gamma_m}{m}.$$

Escribamos $n = km + l$ con $0 \leq l < m$. Como $\gamma_{n+m} \geq \gamma_m + \gamma_n$, podemos observar que $\gamma_n = \gamma_{km+l} \geq \gamma_{km} + \gamma_l = k\gamma_m + \gamma_l$, lo que implica que al dividir entre $n = km + l$ tenemos

$$\frac{\gamma_n}{n} \geq \frac{k\gamma_m + \gamma_l}{km + l} = \left(\frac{km}{km + l} \right) \frac{\gamma_m}{m} + \frac{\gamma_l}{n} = \left(\frac{1}{1 + \frac{l}{km}} \right) \frac{\gamma_m}{m} + \frac{\gamma_l}{n}.$$

Por lo tanto si hacemos crecer n a infinito, dado que $n = km + l$ con $0 \leq l < m$, entonces

$$\liminf_{n \rightarrow \infty} \frac{\gamma_n}{n} \geq \frac{\gamma_m}{m}.$$

◇

El *Lema* anterior prueba la existencia de $\gamma(a)$, que además es negativo ya que $\mathbb{P}\{S_n \geq na\} < 1$.

Por lo tanto

$$\mathbb{P}\{S_n \geq na\} \leq e^{n\gamma(a)+\epsilon}, \quad \text{para un } \epsilon > 0 \text{ y } \forall n \in \mathbb{N}.$$

De aquí en adelante supondremos que X_i tiene transformada de *Laplace* finita para algún $\theta > 0$, salvo cuando se especifique lo contrario.

Sea $\theta_+ = \sup\{\theta : M(\theta) < \infty\}$, $\theta_- = \inf\{\theta : M(\theta) < \infty\}$. Es claro que $M(\theta) < \infty$ para $\theta \in (\theta_-, \theta_+)$. Por lo tanto, ya que X_i tiene transformada de *Laplace* finita, tenemos que $\mathbb{E}[X_i^+] < \infty$ y además $\mu = \mathbb{E}[X^+] - \mathbb{E}[X^-] \in [-\infty, \infty)$.

Ahora bien, si $\theta > 0$, la desigualdad de *Chebyshev* implica

$$e^{\theta na} \mathbb{P}\{S_n \geq na\} \leq \mathbb{E}[\exp(\theta S_n)] = M^n(\theta).$$

De donde podemos concluir que

$$\mathbb{P}\{S_n \geq na\} \leq \exp\{-n(a\theta - \psi(\theta))\}. \quad (3.1)$$

Para entender un poco mejor el comportamiento de esta cota superior, establezcamos el siguiente *Lema*.

Lema 2. Si $a > \mu$, entonces existe θ_0 tal que $a\theta - \psi(\theta) > 0$, para todo $\theta \in [0, \theta_0]$.

Demostración

Comenzaremos por notar que $\psi(0) = \log M(0) = 0$, por lo que basta con probar:

1. ψ es continua en 0.
2. ψ es diferenciable en $(0, \theta_+)$, y
3. $\psi'(\theta)$ converge a μ conforme θ tiende a 0.

Por lo tanto, existe θ tal que

$$a\theta - \psi(\theta) = \int_0^\theta (a - \psi'(x))dx < 0.$$

Primero mostremos que efectivamente la derivada de ψ existe, para esto, sea $F(x) = \mathbb{P}\{X_i \leq x\}$, como

$$|e^{hx} - 1| = \left| \int_0^{hx} e^y dy \right| \leq |hx|(e^{hx} + 1). \quad (3.2)$$

Entonces al hacer uso del *Teorema de Convergencia Dominada* con (3.2), logramos mostrar que

$$\begin{aligned} M'(\theta) &= \lim_{h \rightarrow 0} \frac{M(\theta + h) - M(\theta)}{h} = \lim_{h \rightarrow 0} \int \frac{e^{hx} - 1}{h} e^{\theta x} dF(x) \\ &= \int x e^{\theta x} dF(x) \quad , \text{ para } \theta \in (\theta_-, \theta_+). \end{aligned}$$

Por la última ecuación se sigue que $\psi(\theta) = \log M(\theta)$ tiene derivada $\psi'(\theta) = M'(\theta)/M(\theta)$. Ahora si tomamos el límite cuando θ decrece a 0, notamos que por el *Teorema de Convergencia Monótona* para las $x \leq 0$ ⁶ y del *Teorema de Convergencia Dominada* para las $x \geq 0$ tenemos que $M'(\theta)$ tiende a μ conforme θ tiende a 0.

Un razonamiento similar nos muestra que si θ decrece a 0, entonces $M(\theta)$ converge a 1.

Con lo anterior hemos probado de 1. a 3., y dada la condición de suficiencia, nuestro *Lema* queda demostrado.

◇

Hasta este momento hemos encontrado una cota superior para $\mathbb{P}\{S_n \geq na\}$, y de manera natural nos preguntamos por el valor óptimo de la misma. Es por esta razón que el siguiente paso es buscar optimizar esta cota encontrando el valor máximo de $\theta a - \psi(\theta)$. De este modo al derivar con respecto de θ tenemos

$$\frac{d}{d\theta} \{\theta a - \psi(\theta)\} = a - \frac{M'(\theta)}{M(\theta)}.$$

⁶Es claro que tenemos una sucesión decreciente como función de θ que converge a la identidad y por lo tanto la integral converge a la media

Entonces, si suponemos que las cosas se comportan “bien”, el máximo ocurre cuando $a = M'(\theta)/M(\theta)$. Para ver que efectivamente las cosas se comportan “bien” hagamos de esta condición subjetiva una *Hipótesis* matemática. Para esto definamos

$$F_\theta(x) = \frac{1}{M(\theta)} \int_{-\infty}^x e^{\theta y} dF(y).$$

Si utilizamos la prueba del *Lema 2*, tenemos que para $\theta \in (\theta_-, \theta_+)$, F_θ es una distribución con media

$$\mathbb{E}_\theta(X) := \int x dF_\theta(x) = \frac{1}{M(\theta)} \int_{-\infty}^{\infty} x e^{\theta x} dF(x) = \frac{M'(\theta)}{M(\theta)}.$$

Por lo tanto, si repetimos la misma prueba podemos notar que si $\theta \in (\theta_-, \theta_+)$, entonces

$$M''(\theta) = \int_{-\infty}^{\infty} x^2 e^{\theta x} dF(x).$$

De esta forma

$$\frac{d}{d\theta} \frac{M'(\theta)}{M(\theta)} = \frac{M''(\theta)}{M(\theta)} - \left(\frac{M'(\theta)}{M(\theta)} \right)^2 = \int x^2 dF_\theta(x) - \left(\int x dF_\theta(x) \right)^2 \geq 0.$$

Como esta última expresión representa la varianza de F_θ , si pensamos que la distribución de F no tiene masa en el punto μ , entonces $M'(\theta)/M(\theta)$ es estrictamente creciente y por lo tanto la función $a\theta - \log M(\theta)$ es cóncava. Además como $M'(0)/M(0) = \mu$ tenemos que para cada $a > \mu$ existe al menos un $\theta_a \geq 0$ que resuelve la ecuación $a = M'(\theta_a)/M(\theta_a)$ y este valor de θ es el que maximiza la función $a\theta - \log M(\theta)$.

Para poder aclarar la existencia de θ_a analicemos el siguiente escenario

Teorema 1. *Supongamos que $M(\theta)$ existe para algún $\theta > 0$, que la distribución F no tiene masa en μ , y que además existe un $\theta_a \in (0, \theta_+)$ tal que $a = M'(\theta_a)/M(\theta_a)$. Entonces*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\{S_n \geq na\} = -a\theta_a + \log M(\theta_a).$$

Demostración

Iniciemos por decir que

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\{S_n \geq na\} \leq -a\theta_a + \log M(\theta_a),$$

es consecuencia inmediata de

$$\mathbb{P}\{S_n \geq na\} \leq \exp\{-n(a\theta - \log M(\theta))\},$$

ya que si $\theta > 0$ podemos aplicar la desigualdad de *Chebyshev* para obtener

$$\begin{aligned} e^{n\theta a} \mathbb{P}\{S_n \geq na\} &\leq M^n(\theta) \\ \Leftrightarrow \mathbb{P}\{S_n \geq na\} &\leq e^{-n\theta a} M^n(\theta) \\ \Leftrightarrow \mathbb{P}\{S_n \geq na\} &\leq \exp\{-n\theta a + n \log M(\theta)\}. \end{aligned}$$

Entonces si aplicamos la función logaritmo en ambos lados de la desigualdad

$$\log \mathbb{P}\{S_n \geq na\} \leq -n\theta a + n \log M(\theta) = -n\{\theta a - \log M(\theta)\}$$

$$\Rightarrow \frac{1}{n} \log \mathbb{P}\{S_n \geq na\} \leq -\theta a + \log M(\theta).$$

Por lo tanto, al cambiar θ por θ_a tenemos que

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\{S_n \geq na\} \leq -a\theta_a + \log M(\theta_a).$$

Ahora, para probar que

$$\liminf_{n \rightarrow \infty} \log \mathbb{P}\{S_n \geq na\} \geq -\theta_a a + \log M(\theta_a),$$

tomemos un $\lambda \in (\theta_a, \theta_+)$ y $X_1^\lambda, X_2^\lambda, \dots$ v.a.i.i.d con distribución F_λ .

Definamos para esta sucesión de variables aleatorias a $S_n^\lambda = X_1^\lambda + \dots + X_n^\lambda$.

Entonces, si escribimos dF/dF_λ como la derivada de *Radon-Nikodym* asociada a las medidas en cuestión, se sigue inmediatamente de la definición que

$$\frac{dF}{dF_\lambda} = e^{\lambda x} M(\lambda).$$

De modo que si F_λ^n es la distribución de S_n^λ y F^n la distribución de S_n , entonces

Lema 3. La n -ésima derivada de Radon-Nikodym de las distribuciones F_λ y F está dada por

$$\frac{dF^n}{dF_\lambda^n} = e^{\lambda x} M^n(\lambda).$$

Demostración

Para poder probar esto utilizaremos inducción sobre n .

El caso $n = 1$ queda cubierto inmediatamente de la definición de F y F_λ .

Si $n > 1$, debemos notar que

$$F^n(z) = F^{n-1} * F(z) = \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} dF(y) dF^{n-1}(x),$$

es decir que F^n es la n -ésima convolución de la distribución F . Por lo tanto bajo la hipótesis de inducción tenemos

$$dF^{n-1}(x) = e^{-\lambda x} M^{n-1}(\lambda) dF_\lambda^{n-1}(x) \quad y \quad dF(y) = e^{-\lambda y} M(\lambda) dF_\lambda(y),$$

lo que implica que

$$\begin{aligned} F^n(z) &= \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} e^{-\lambda y} M(\lambda) dF_\lambda(y) e^{-\lambda x} M^{n-1}(\lambda) dF_\lambda^{n-1}(x) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\mathbb{I}_{\{x+y \leq z\}}} e^{-\lambda(x+y)} M^n(\lambda) dF_\lambda(y) dF_\lambda^{n-1} \quad \text{variables con dist. } F_\lambda. \\ &= \mathbb{E}[\mathbb{I}_{\{S_{n-1}^\lambda + X_n^\lambda\}} e^{\lambda(S_{n-1}^\lambda + X_n^\lambda)} M^n(\lambda)] \quad \text{haciendo } X_n^\lambda = S_n^\lambda - S_{n-1}^\lambda. \\ &= \mathbb{E}[\mathbb{I}_{\{S_n^\lambda \leq z\}} e^{-\lambda S_n^\lambda} M^n(\lambda)] \\ &= \int_{-\infty}^z e^{-\lambda u} M^n(\lambda) dF_\lambda^n(u). \end{aligned}$$

De modo que

$$dF^n(z) = e^{-\lambda z} M^n(\lambda) dF_\lambda^n(z) \quad \therefore \quad \frac{dF^n}{dF_\lambda^n} = e^{-\lambda z} M^n(\lambda).$$

◇

De esta forma si tenemos que $\nu > a$, entonces al utilizar el *Lema* anterior y la monotonía podemos afirmar que

$$\mathbb{P}\{S_n \geq na\} \geq \int_{na}^{\nu} e^{-\lambda x} M^n(\lambda) dF_{\lambda}^n(x) \geq M^n(\lambda) e^{-\lambda n\nu} (F_{\lambda}^n(n\nu) - F_{\lambda}^n(na)). \quad (3.3)$$

Ahora como ya sabemos que la media de la distribución F_{λ} es $M'(\lambda)/M(\lambda)$, si tenemos que $a < M'(\lambda)/M(\lambda) < \nu$ con probabilidad 1, utilizando la *Ley Débil de los Grandes Números* (LDGN) obtenemos que

$$F_{\lambda}^n(n\nu) - F_{\lambda}^n(na) \longrightarrow 1 \quad \text{conforme} \quad n \rightarrow \infty.$$

Por otro lado, el producto $M^n(\lambda)e^{-\lambda n\nu}$ converge a cero a una tasa menos rápida, por lo que

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\{S_n \geq na\} \geq -\lambda\nu + \log M(\lambda).$$

Y ya que $\lambda > \theta_a$ y $\nu > M'(\lambda)/M(\lambda)$ es arbitraria, nuestra prueba está completa.

◇

Si nos fijamos ahora en los valores de θ para los que no podemos resolver la ecuación $a = M'(\theta)/M(\theta)$, debemos observar que si x_0 representa el supremo de la función de distribución sobre todos los valores de x para los cuales $F(x) < 1$ (extremo del soporte de la distribución), entonces si F no tiene masa en x_0 , podemos afirmar que $M'(\theta)/M(\theta)$ crece a x_0 conforme θ crece a infinito, pero $M'(\theta)/M(\theta) < x_0$ para todo $\theta < \infty$. Sin embargo el resultado para $a = x_0$ es trivial ya que

$$\frac{1}{n} \log \mathbb{P}\{S_n \geq nx_0\} = \log \mathbb{P}\{X_i = x_0\} \quad \forall n.$$

Cuando $x_0 = \infty$ se tiene que $M'(\theta)/M(\theta)$ crece a infinito conforme θ crece a infinito, de tal forma que el único caso por analizar queda cubierto con el siguiente *Teorema*

Teorema 2. *Supongamos que $x_0 = \infty$, $\theta_+ < \infty$ y que además $M'(\theta)/M(\theta)$ crece a un límite finito a_0 conforme θ crece a θ_0 , entonces si $a_0 \leq a < \infty$*

$$\frac{1}{n} \log \mathbb{P}\{S_n \geq na\} \longrightarrow -a\theta_+ + \log M(\theta_+),$$

implicando con esto la linealidad de $\gamma(a)$ para valores $a \geq a_0$.

Demostración

Como $\log(M(\theta))' = M'(\theta)/M(\theta)$, entonces al integrar de 0 a θ_+ tenemos

$$\int_0^{\theta_+} \frac{M'(\theta)}{M(\theta)} d\theta = \log(M(\theta_+)) - \log(M(0)),$$

donde claramente $\log(M(0)) = 0$ y ya que $M(\theta)$ es finito para todo θ , garantizamos que $\log(M(\theta))$ también es finito.

Si hacemos $\theta = \theta_+$ en $\mathbb{P}\{S_n \geq na\} \leq \exp\{-n(a\theta - \psi(\theta))\}$ obtenemos que

$$\frac{1}{n} \log \mathbb{P}\{S_n \geq na\} \leq -a\theta_+ + \log(M(\theta_+)) = -a\theta_+ + \psi(\theta_+).$$

Por lo tanto

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\{S_n \geq na\} \leq -a\theta_+ + \psi(\theta_+).$$

Para verificar la otra desigualdad, usaremos la distribución F_λ con $\lambda = \theta_+$, entonces

$$\mathbb{P}\{S_n \geq na\} \geq \int_{na}^{n\nu} e^{-\theta_+ x} M^n(\theta_+) dF_{\theta_+}^n(x) \geq M^n(\theta_+) e^{-\theta_+ n\nu} [F_{\theta_+}^n(n\nu) - F_{\theta_+}^n(na)].$$

Y como esto es válido para $\nu > a$, tenemos que por la *Ley Débil de los Grandes Números* la parte entre corchetes converge a 1, y por lo tanto

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\{S_n \geq na\} \geq -\nu\theta_+ + \psi(\theta_+).$$

Ya que $\theta_+ > \theta_a$ y $\nu > M'(\theta_+)/M(\theta_+)$ es arbitrario, según nuestras hipótesis, se sigue que

$$\frac{1}{n} \log \mathbb{P}\{S_n \geq na\} \longrightarrow -a\theta_+ + \psi(\theta_+).$$

Al hacer que θ crezca a θ_+ y aplicando el *Teorema de Convergencia Dominada* para $x \leq 0$ y el *Teorema de Convergencia Monótona* para $x \geq 0$, de manera análoga al punto 3 de la demostración del *Lema 2*, podemos observar que F_λ tiene media a_0 . Entonces como

$$\mathbb{P}\{S_n \geq na\} \geq M^n(\lambda) e^{-n\lambda\nu} [F_\lambda^n(n\nu) - F_\lambda^n(na)],$$

si $a_0 \leq a < \nu = a + 3\epsilon$, se obtiene lo siguiente

$$\frac{1}{n} \log \mathbb{P}\{S_n \geq na\} \geq M(\lambda) - \lambda\nu + \frac{1}{n} \log \mathbb{P}\{S_n^\lambda \in (na, n\nu)\}.$$

Al tomar $X_1^\lambda, X_2^\lambda, \dots$ v.a.i.i.d con distribución F_λ y considerar S_n^λ , tenemos que

$$\mathbb{P}\{S_n^\lambda \in (na, n\nu)\} \geq \mathbb{P}\{S_{n-1}^\lambda \in ((a_0 - \epsilon)n, (a_0 + \epsilon)n)\} \cdot \mathbb{P}\{X_n^\lambda \in ((a - a_0 + \epsilon)n, (a - a_0 + 2\epsilon)n)\}.$$

Es decir que primero ubicamos a S_{n-1}^λ y después observamos en donde cae X_n^λ . Ahora, si utilizamos el hecho de que la distribución F_λ tiene media a_0 , por un lado sabemos que por la LDGN

$$\mathbb{P}\{S_{n-1}^\lambda \in ((a_0 - \epsilon)n, (a_0 + \epsilon)n)\} \xrightarrow{n \rightarrow \infty} 1.$$

Por lo que

$$\begin{aligned} \mathbb{P}\{S_n^\lambda \in (na, n\nu)\} &\geq \mathbb{P}\{S_{n-1}^\lambda \in ((a_0 - \epsilon)n, (a_0 + \epsilon)n)\} \\ &\quad \cdot \mathbb{P}\{X_n^\lambda \in ((a - a_0 + \epsilon)n, (a - a_0 + 2\epsilon)n)\} \\ &\geq \frac{1}{2} \mathbb{P}\{X_n^\lambda \in ((a - a_0 + \epsilon)n, (a - a_0 + \epsilon)(n + 1))\}. \end{aligned} \tag{3.4}$$

Cabe resaltar que esta cota inferior es válida, pues si analizamos los intervalos en los que se pide se encuentre la variable X_n^λ por separado, podemos observar que el extremo izquierdo de ambos intervalos es el mismo $((a - a_0 + \epsilon)n)$, por lo que únicamente analizaremos el extremo derecho.

Por un lado,

$$(a - a_0 + 2\epsilon)n = (a - a_0 + \epsilon)n + \epsilon + \epsilon(n - 1),$$

mientras que

$$(a - a_0 + \epsilon)(n + 1) = (a - a_0 + \epsilon)n + \epsilon + (a - a_0).$$

Por lo tanto al eliminar los términos iguales, obtenemos $\epsilon(n - 1)$ en el primer intervalo, en tanto que en el otro nos resta $(a - a_0)$. Entonces resaltando que este último término es constante, tenemos que para un n suficientemente grande $\epsilon(n - 1) > (a - a_0)$, con lo cual confirmamos que la cota inferior expuesta en (3.4) es correcta.

Así que para poder completar la prueba de nuestro *Teorema*, sólo nos resta saber cuál es el comportamiento de

$$\mathbb{P}\{X_n^\lambda \in ((a - a_0 + \epsilon)n, (a - a_0 + \epsilon)(n + 1))\}.$$

De esta forma, al hacer un análisis de primer paso afirmamos que

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\{X_n^\lambda \in ((a - a_0 + \epsilon)n, (a - a_0 + \epsilon)(n + 1))\} = 0, \epsilon > 0.$$

El argumento para esta afirmación es el siguiente : podemos decir que este límite superior es igual a cero, ya que de otra forma seríamos capaces de elegir un $\eta < 0$ tal que

$$\frac{1}{n} \log \mathbb{P}\{X_n^\lambda \in ((a - a_0 + \epsilon)n, (a - a_0 + \epsilon)(n + 1))\} \leq \eta < 0.$$

Entonces

$$\mathbb{P}\{X_\lambda^1 \in (wn, w(n + 1))\} \leq e^{n\eta} < 1 \quad \text{en donde } w = (a - a_0 + \epsilon).$$

De modo que

$$\sum_{n=0}^{\infty} \mathbb{P}\{X_\lambda^1 \in (wn, w(n + 1))\} \leq \sum_{n=0}^{\infty} e^{n\eta} < \infty.$$

Puesto que la función exponencial decrece rápidamente y se encuentra elevada a un exponente negativo, implica que para la parte positiva de las variables la esperanza es finita, es decir $\mathbb{E}[X_1^{\lambda^+}] < \infty$.

Por lo tanto, al hacer un análisis similar tomando $\vartheta > 0$ tal que $\vartheta < -\eta$, entonces como $\vartheta + \eta < 0$

$$\begin{aligned} \mathbb{E}[e^{\vartheta X_1^\lambda}] &= \int_{-\infty}^{\infty} e^{\vartheta x^\lambda} dF_\lambda(x) \\ &= \int_{-\infty}^0 e^{\vartheta x^\lambda} dF_\lambda(x) + \int_0^{Nw} e^{\vartheta x^\lambda} dF_\lambda + \int_{Nw}^{\infty} e^{\vartheta x^\lambda} dF_\lambda(x), \end{aligned} \tag{3.5}$$

para un $N_w \in \mathbb{R}^+$, que depende de w . Ahora, dada la ecuación (3.5) únicamente basta verificar que estas tres últimas integrales son finitas para obtener el resultado que estamos buscando. Es claro que por las propiedades de la función exponencial en cero, en la primera integral tenemos

$$\int_{-\infty}^0 e^{\vartheta x_1^\lambda} dF_\lambda(x) < 1 \cdot \mathbb{P}\{X_1^\lambda \in (-\infty, 0)\} < \infty,$$

mientras que en la segunda

$$\int_0^{Nw} e^{\vartheta x_1^\lambda} dF_\lambda < e^{\vartheta Nw} \cdot \mathbb{P}\{X_1^\lambda \in (0, Nw)\} < \infty.$$

Por último, para garantizar que la tercer integral es finita, podemos realizar lo siguiente

$$\begin{aligned} \int_{Nw}^{\infty} e^{\vartheta x_1^\lambda} dF_\lambda(x) &= \sum_{n=Nw}^{\infty} \int_n^{n+1} e^{\vartheta x_1^\lambda} dF_\lambda(x) \leq \sum_{n=Nw}^{\infty} e^{\vartheta(n+1)} \mathbb{P}\{X_1^\lambda \in (n, n+1]\} \\ &\leq \sum_{n=Nw}^{\infty} e^{\vartheta(n+1)} e^{n\eta} \quad \text{por lo visto arriba} \\ &= \sum_{n=Nw}^{\infty} e^{\vartheta} e^{n\vartheta} e^{n\eta} = e^{\vartheta} \sum_{n=Nw}^{\infty} e^{(\vartheta+\eta)n} < \infty \quad \text{pues } (\vartheta + \eta) < 0. \end{aligned}$$

Ahora bien, como $\mathbb{E}[e^{\vartheta X_1^\lambda}] < \infty$ para algún $\vartheta > 0$ implica que $\mathbb{E}[e^{(\lambda+\vartheta)X_1}] < \infty$. No obstante, esto no es posible ya que contradice el supuesto $\lambda = \theta_+ = \sup\{\theta : M(\theta) < \infty\}$.

De esta forma se concluye que

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\{S_n \geq na\} \geq \liminf_{n \rightarrow \infty} [M(\lambda) - \lambda\nu] + \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\{S_n^\lambda \in (na, n\nu)\},$$

implica

$$\liminf_{n \rightarrow \infty} [M(\lambda) - \lambda\nu] + \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\{S_n^\lambda \in (na, n\nu)\} \geq M(\lambda) - \lambda\nu.$$

Por lo tanto

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\{S_n \geq na\} = \gamma(a).$$

◇

Como acotación importante al desarrollo teórico antes expuesto, observamos que F_θ es justamente el *Cambio Exponencial de Medida* definido en la sección 2.7 y que $\gamma(a)$ puede traducirse como la constante γ definida para la *Eficiencia Asintótica*, concepto del cual hablamos en la sección 2.5, considerando un a fijo y el hecho de que $\{S_n \geq na\}$ es un evento de baja probabilidad. Por lo tanto la esencia del muestreo de importancia bajo un esquema de Grandes Desviaciones, radica en encontrar el θ_a que optimiza la función $a\theta - \psi(\theta)$ y posteriormente realizar el cambio de medida sobre este parámetro.

Capítulo 4

Contra-ejemplo en Muestreo de Importancia para Probabilidades de Grandes Desviaciones.

Como ya hemos mencionado en repetidas ocasiones, el *Muestreo de Importancia* es un estupendo método para obtener estimaciones eficientes de probabilidades de eventos *raros*, via *Monte Carlo*. Sabemos que el muestreo de importancia puede basarse en el cambio de medida que sugiere el análisis de *Grandes Desviaciones* y que bajo éste esquema garantizamos la asintoticidad de nuestros resultados. Sin embargo, aún no podemos asegurar reducir sustancialmente la varianza de estas estimaciones.

En una gran variedad de problemas ésta aproximación puede sugerir el uso de estimadores que resultan ser óptimos en un sentido asintótico. No obstante, en este capítulo brindaremos un ejemplo para el cual, el estimador propuesto bajo el esquema que sugieren las grandes desviaciones, presenta un pobre desempeño en lo que a eficiencia se refiere, ya que este estimador puede tener una varianza que decrezca a una tasa más lenta que la de un estimador “ingenuo” o “burdo”, una varianza que se incremente con la rareza del evento o incluso una varianza infinita. La característica principal de este ejemplo es el hecho de permitir más de una forma para que el evento “raro” ocurra, lo que debe ser analizado minuciosamente o reeditaré en una baja eficiencia. Esta es la idea mostrada en el artículo de *Paul Glasserman y Yashan Wang* [11], en el cual se basa la exposición de este capítulo. En el desarrollo del ejemplo se propone un estimador alternativo y se discute a fondo el por qué de esta idea, así como también se enfatiza el hecho de que esta alterna-

tiva otorga pesos específicos a las trayectorias de baja probabilidad omitidas por los términos asintóticos.

Como ya sabemos las grandes desviaciones buscan identificar la curva o trayectoria más parecida al evento que se desea simular, sin embargo esto puede tener efectos adversos en lo que a la varianza se refiere. Para hacer más evidente esto, simplemente tomemos en cuenta que el análisis de primer momento no nos puede garantizar nada acerca de la conducta del segundo momento, por lo que es natural pensar en la necesidad de ir más allá del límite de las grandes desviaciones para garantizar la efectividad del **IS** en términos de varianza. A pesar de existir casos para los cuales el estimador propuesto via grandes desviaciones, no muestra un gran desempeño, estos casos son contados y es claro que el cambio de medida propuesto bajo este esquema siempre será un buen punto de partida, ya que de ser coherente con estos principios garantizamos propiedades sobre la convergencia de nuestras estimaciones, y en caso de ser necesario siempre podemos refinar esta primer propuesta para obtener los resultados deseados en términos de reducción de varianza.

§4.1. Descripción del Problema.

Comencemos por dar una formulación genérica del problema al que nos enfrentaremos, y para esto consideremos una familia indexada de eventos $\{A_x, x > 0\}$ que satisfaga cualquiera de las siguientes dos condiciones

$$\lim_{x \rightarrow \infty} \frac{1}{x} \log \mathbb{P}\{A_x\} = -\gamma, \quad (4.1)$$

para algún $\gamma > 0$, o la asintoticidad exponencial

$$\mathbb{P}\{A_x\} \sim C e^{-\gamma x}, \quad (4.2)$$

para alguna constante $C > 0$. En la sección 2.4 hablamos de que para estimar $\alpha(x) = \mathbb{P}\{A_x\}$ via una simulación directa, debemos generar repeticiones independientes de la variable aleatoria $\mathbb{I}_{\{A_x\}}$ o algo que se le aproxime.

Bajo este tipo de simulación tenemos que la varianza del estimador es

$$\text{Var}[\mathbb{I}_{\{A_x\}}] = \mathbb{E}[\mathbb{I}_{\{A_x\}}^2] - \mathbb{E}[\mathbb{I}_{\{A_x\}}]^2 = \alpha - \alpha^2,$$

por lo tanto, si $\mathbb{P}\{A_x\}$ tiende a 0, implica que la varianza se aproxima a cero, pero el *Error Relativo* del estimador (el cociente de la desviación estándar y la media)

satisface

$$Error\ Relativo = \frac{\sqrt{\alpha(x) - \alpha^2(x)}}{\alpha(x)} \geq \frac{1}{\sqrt{\alpha(x)}} \longrightarrow \infty, \quad \text{conforme } x \rightarrow \infty.$$

Si (4.1) se cumple, el incremento a infinito es exponencial, lo que indica que el número de repeticiones independientes de $\mathbb{I}_{\{A_x\}}$ requeridas para alcanzar un *Error Relativo* fijo crece exponencialmente con x , ya que de manera análoga al procedimiento mostrado en la sección 2.5 observamos que el tamaño de muestra N para alcanzar una cierta efectividad δ se expresa como

$$N = \frac{q_{\delta/2}^2(\alpha(x) - \alpha^2(x))}{\delta^2\alpha(x)} \sim \frac{q_{\delta/2}^2}{\delta^2 C e^{-\gamma x}}.$$

Hemos visto que bajo el IS se generan muestras independientes sobre una diferente medida de probabilidad digamos $\tilde{\mathbb{P}}$ con operador esperanza $\mathbb{E}_{\tilde{\mathbb{P}}}$ y se usa la representación

$$\mathbb{P}\{A_x\} = \mathbb{E}_{\tilde{\mathbb{P}}}[L_x; A_x] := \mathbb{E}_{\tilde{\mathbb{P}}}[L_x \mathbb{I}_{\{A_x\}}], \quad (4.3)$$

donde L_x es la razón de verosimilitudes entre las medidas $\tilde{\mathbb{P}}$ y \mathbb{P} , o de manera más precisa es la razón de verosimilitudes de una restricción de $\tilde{\mathbb{P}}$ a una restricción de \mathbb{P} , la restricción que se hace de una sub- σ -álgebra contenida en A_x en la cual $\tilde{\mathbb{P}} \gg \mathbb{P}$. Muchas veces la continuidad absoluta de las medidas no restringidas falla. Este es justamente el caso de nuestro ejemplo y de hecho de todos los ejemplos en donde el cambio de \mathbb{P} a $\tilde{\mathbb{P}}$ transforme la distribución en común de una sucesión infinita de variables aleatorias independientes idénticamente distribuidas.

Basado en (4.3) obtenemos un estimador insesgado de $\mathbb{P}\{A_x\} = \alpha(x)$ promediando repeticiones independientes de

$$\hat{\alpha}(x) = L_x \mathbb{I}_{\{A_x\}},$$

con $\hat{\alpha}$ generado bajo $\tilde{\mathbb{P}}$. Ahora bien, si la desviación estándar de este estimador es más pequeña que la de $\mathbb{I}_{\{A_x\}}$, al menos para x grande, implica que $\hat{\alpha}(x)$ es un mejor estimador para nuestros propósitos.

Si (4.1) se satisface, entonces la no negatividad de la varianza implica que

$$\begin{aligned}
& 0 \leq \text{Var}_{\tilde{\mathbb{P}}}\{\hat{\alpha}\} = \mathbb{E}_{\tilde{\mathbb{P}}}\{\hat{\alpha}^2\} - \mathbb{E}_{\tilde{\mathbb{P}}}\{\hat{\alpha}\}^2 \\
\Rightarrow & -\mathbb{E}_{\tilde{\mathbb{P}}}\{\hat{\alpha}^2\} \leq -\mathbb{E}_{\tilde{\mathbb{P}}}\{\hat{\alpha}\}^2 \\
\Rightarrow & -\frac{1}{x} \log \mathbb{E}_{\tilde{\mathbb{P}}}\{\hat{\alpha}^2\} \leq -\frac{1}{x} \log \mathbb{E}_{\tilde{\mathbb{P}}}\{\hat{\alpha}\}^2 \\
\Rightarrow & \limsup_{x \rightarrow \infty} -\frac{1}{x} \log \mathbb{E}_{\tilde{\mathbb{P}}}\{\hat{\alpha}^2\} \leq \lim_{x \rightarrow \infty} -\frac{1}{x} \log \mathbb{E}_{\tilde{\mathbb{P}}}\{\hat{\alpha}\}^2 \\
\Rightarrow & \limsup_{x \rightarrow \infty} -\frac{1}{x} \log \mathbb{E}_{\tilde{\mathbb{P}}}\{\hat{\alpha}^2\} \leq \lim_{x \rightarrow \infty} -\frac{1}{x} \log \mathbb{P}^2\{A_x\} = 2\gamma.
\end{aligned}$$

Por lo tanto

$$\limsup_{x \rightarrow \infty} \left\{ -\frac{1}{x} \log \mathbb{E}_{\tilde{\mathbb{P}}}\{\hat{\alpha}^2(x)\} \right\} \leq 2\gamma.$$

Esto significa que la tasa logarítmica de crecimiento del segundo momento del estimador puede ser a lo más dos veces γ .

Recordaremos ahora las siguientes definiciones

Definición 1. Se dice que un estimador $\hat{\alpha}(x)$ es **Asintóticamente Eficiente** o **Asintóticamente Óptimo**, si la tasa logarítmica de crecimiento máxima es alcanzada, es decir

$$\lim_{x \rightarrow \infty} \left\{ -\frac{1}{x} \log \mathbb{E}_{\tilde{\mathbb{P}}}\{\hat{\alpha}^2(x)\} \right\} = 2\gamma.$$

Definición 2. Decimos que el estimador $\hat{\alpha}(x)$ es **Logarítmicamente Eficiente**, si existe $\epsilon > 0$ tal que

$$\limsup_{x \rightarrow \infty} \frac{\text{Var}_{\tilde{\mathbb{P}}}\{\hat{\alpha}(x)\}}{\mathbb{P}\{A_x\}^{2-\epsilon}} < \infty,$$

o de manera equivalente

$$\liminf_{x \rightarrow \infty} \frac{-\log(\text{Var}_{\tilde{\mathbb{P}}}\{\hat{\alpha}(x)\})}{-\log(\mathbb{P}\{A_x\})} \geq 2$$

Definición 3. Un estimador $\hat{\alpha}(x)$ tiene **Error Relativo Acotado** si satisface

$$\limsup_{x \rightarrow \infty} \frac{\sqrt{\mathbb{E}_{\tilde{\mathbb{P}}}\{\hat{\alpha}^2(x)\}}}{\mathbb{P}\{A_x\}} < \infty.$$

Las siguientes dos proposiciones, enfatizan la importancia de la **Definición 3**.

Proposición 3. Si $\hat{\alpha}(x)$ tiene error relativo acotado, entonces $\hat{\alpha}(x)$ es asintóticamente eficiente.

Demostración

Sabemos que debido a la no negatividad de la varianza, siempre se cumple que

$$\limsup_{x \rightarrow \infty} \left\{ -\frac{1}{x} \log \mathbb{E}_{\tilde{\mathbb{P}}}[\hat{\alpha}^2(x)] \right\} \leq 2\gamma,$$

por lo que para probar que $\hat{\alpha}(x)$ es asintóticamente eficiente, basta con demostrar

$$\liminf_{x \rightarrow \infty} \left\{ -\frac{1}{x} \log \mathbb{E}_{\tilde{\mathbb{P}}}[\hat{\alpha}^2(x)] \right\} \geq 2\gamma.$$

De este modo, utilizando que $\hat{\alpha}(x)$ tiene error relativo acotado, tenemos que para todo $\epsilon > 0$, existe $x_0 > 0$ tal que para todo $x > x_0$

$$\frac{\sqrt{\mathbb{E}_{\tilde{\mathbb{P}}}[\hat{\alpha}^2(x)]}}{\mathbb{P}\{A_x\}} = \frac{\sqrt{\mathbb{E}_{\tilde{\mathbb{P}}}[\hat{\alpha}^2(x)]}}{\mathbb{E}_{\tilde{\mathbb{P}}}[\hat{\alpha}(x)]} \leq k + \epsilon,$$

con k constante. Entonces

$$\mathbb{E}_{\tilde{\mathbb{P}}}[\hat{\alpha}^2(x)] \leq (k + \epsilon)^2 \mathbb{E}_{\tilde{\mathbb{P}}}^2[\hat{\alpha}(x)].$$

Haciendo uso de la monotonía de la función logaritmo observamos que

$$\log \mathbb{E}_{\tilde{\mathbb{P}}}[\hat{\alpha}^2(x)] \leq 2 \log(k + \epsilon) + 2 \log \mathbb{E}_{\tilde{\mathbb{P}}}[\hat{\alpha}(x)],$$

de donde se sigue

$$-\frac{1}{x} \log \mathbb{E}_{\tilde{\mathbb{P}}}[\hat{\alpha}^2(x)] \geq -\frac{2}{x} \log(k + \epsilon) - \frac{2}{x} \log \mathbb{E}_{\tilde{\mathbb{P}}}[\hat{\alpha}(x)].$$

Por lo tanto, al tomar el límite inferior de esta desigualdad

$$\liminf_{x \rightarrow \infty} -\frac{1}{x} \log \mathbb{E}_{\tilde{\mathbb{P}}}[\hat{\alpha}^2(x)] \geq 2 \liminf_{x \rightarrow \infty} -\frac{1}{x} \log \mathbb{E}_{\tilde{\mathbb{P}}}[\hat{\alpha}(x)],$$

ya que el primer sumando de la parte del lado derecho, tiende a cero conforme x tiende a infinito. Por último como $\mathbb{E}_{\tilde{\mathbb{P}}}[\hat{\alpha}(x)] = \mathbb{P}\{A_x\}$, se concluye el resultado.

◇

Proposición 4. Si $\hat{\alpha}(x)$ tiene error relativo acotado, entonces $\hat{\alpha}(x)$ es logarítmicamente eficiente.

Demostración

Como $\hat{\alpha}(x)$ tiene error relativo acotado, existe $x_0 > 0$ tal que para toda $\epsilon > 0$ y $x > x_0$

$$\frac{\sqrt{\mathbb{E}_{\tilde{\mathbb{P}}}\left[\hat{\alpha}^2(x)\right]}}{\mathbb{P}\{A_x\}} \leq k + \epsilon,$$

con k constante. Al elevar al cuadrado y restar de ambos lados de la desigualdad un 1, tenemos que

$$\frac{\mathbb{E}_{\tilde{\mathbb{P}}}\left[\hat{\alpha}^2(x)\right]}{\mathbb{P}^2\{A_x\}} - \frac{\mathbb{E}_{\tilde{\mathbb{P}}}^2\left[\hat{\alpha}(x)\right]}{\mathbb{E}_{\tilde{\mathbb{P}}}^2\left[\hat{\alpha}(x)\right]} \leq (k + \epsilon)^2 - \frac{\mathbb{E}_{\tilde{\mathbb{P}}}^2\left[\hat{\alpha}(x)\right]}{\mathbb{E}_{\tilde{\mathbb{P}}}^2\left[\hat{\alpha}(x)\right]},$$

de donde facilmente se sigue

$$\frac{\text{Var}_{\tilde{\mathbb{P}}}\left[\hat{\alpha}(x)\right]}{\mathbb{P}\{A_x\}^2} < (k + \epsilon)^2 - 1.$$

Por lo tanto

$$\limsup_{x \rightarrow \infty} \frac{\text{Var}_{\tilde{\mathbb{P}}}\left[\hat{\alpha}(x)\right]}{\mathbb{P}\{A_x\}^{2-\epsilon}} < \infty$$

◇

Estas tres propiedades son de gran utilidad para el análisis de la eficiencia de los estimadores, y ya que el objetivo primordial de este capítulo es mostrar como los estimadores sugeridos en base a la asintoticidad de $\mathbb{P}\{A_x\}$, pueden no cumplir dichas propiedades e inclusive pueden tener menor efectividad que estimadores directos como la función indicadora del evento, necesitamos precisar perfectamente el mecanismo mediante el cual un estimador es sugerido. Uno puede imaginar pocos sentidos por los cuales un cambio de medida es consistente con el resultado de grandes desviaciones, lo que sugiere que es poco razonable pensar que alguno de ellos pueda cubrir todos los aspectos a estudiar. Tal vez la noción más atractiva sería la existencia de un *Teorema Límite Condicionado*, que nos muestre estrictamente que la medida de probabilidad \mathbb{P} condicionada en A_x converge a $\tilde{\mathbb{P}}$ conforme x tiende a infinito y así simular sobre la *Ley Condicional* y producir un estimador de varianza cero.

Una noción relacionada es la existencia de un resultado de grandes desviaciones para medidas empíricas en trayectorias de A_x , generalizando así el *Teorema de Sanov* (ver [5]).

Sin embargo, nosotros trabajaremos con dos nociones aparentemente más débiles que intentarán acompañar a (4.1) y (4.2), respectivamente.

Definición 4. Diremos que $\tilde{\mathbb{P}}$ es consistente con el resultado de grandes desviaciones para $\mathbb{P}\{A_x\}$ si el $\liminf_{x \rightarrow \infty} \tilde{\mathbb{P}}\{A_x\} > 0$ y bajo $\tilde{\mathbb{P}}$, cualquiera de las dos siguientes condiciones se cumple:

$$-\frac{1}{x} \log L_x \longrightarrow \gamma \quad \text{en probabilidad,} \quad (4.4)$$

o

$$e^{\gamma x} L_x \quad \text{tiene una distribución límite propia,} \quad (4.5)$$

donde “propio” significa excluir distribuciones con masa en $\pm\infty$.

Esto claramente no intenta cubrir todos los casos, pero sí trata de establecer un eslabón entre el resultado de grandes desviaciones y el cambio de medida propuesto. Tanto (4.4) como (4.5), sugieren que \mathbb{P} condicionada en A_x es aproximada por $\tilde{\mathbb{P}}$ debido a que la razón de verosimilitudes no se esparce demasiado: la *Ley condicional* por sí sola tiene razón de verosimilitudes constante con respecto a \mathbb{P} en A_x . En algunas ocasiones es preferible multiplicar por la función indicadora $\mathbb{I}_{\{A_x\}}$ en la condición (4.5), pero en el ejemplo que mostraremos aquí, tenemos que $\tilde{\mathbb{P}}$ se encuentra en la frontera de una familia paramétrica de medidas bajo la cual la probabilidad de A_x converge a 1. Podemos argumentar nociones más fuertes de “Consistencia con Grandes Desviaciones” que serán preferibles, no obstante, en las aplicaciones este tipo de nociones pueden ser imposibles de verificar, por lo que en la práctica, uno muchas veces encuentra que no hay nexos precisos entre la asintoticidad de los eventos “raros” y la medida sugerida por el muestreo de importancia.

A continuación mostraremos un ejemplo basado en la asintoticidad logarítmica, para el cual el estimador propuesto es consistente con (4.4) y sin embargo al hacer un análisis de varianza, podemos observar que este estimador tiene varianza infinita. Posteriormente se propone un estimador alternativo que cumple con ser asintóticamente eficiente.

§4.2. Probabilidades de Cola para la Suma de Variables Aleatorias.

Sean X_1, X_2, \dots v.a.i.i.d, con función generadora acumulativa $\psi(\theta)$ cuyo dominio tiene interior no vacío, y supongamos también que existe $a > 0$ tal que

$|\mathbb{E}[X_1]| < a$, para el cual además existen soluciones θ_a y θ_{-a} en el interior del dominio de ψ a las ecuaciones $\psi'(\theta_a) = a$ y $\psi'(\theta_{-a}) = -a$. Sea $S_n = \sum_{i=1}^n X_i$, $n = 1, 2, \dots$. Un hecho conocido a partir de la prueba del *Teorema de Cramér* mostrada en el capítulo anterior, es que

$$-\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\{S_n \geq na\} = \theta_a a - \psi(\theta_a) = \gamma_a, \quad (4.6)$$

y

$$-\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\{S_n \leq -na\} = -\theta_{-a} a - \psi(\theta_{-a}) = \gamma_{-a}. \quad (4.7)$$

El único cambio de medida asintóticamente eficiente, debe cambiar la distribución común de las X_i por θ_a , es decir

$$\mathbb{P}_{\theta_a}\{X_1 \leq x\} = \mathbb{E}[\exp\{\theta_a X_1 - \psi(\theta_a)\}; X_1 \leq x].$$

Cabe resaltar que bajo \mathbb{P}_{θ_a} , las X_i 's son aún v.a.i.i.d, ya que las medidas \mathbb{P}_{θ_a} y \mathbb{P} son mutuamente absolutamente continuas cuando nos restringimos a la σ -álgebra generada por $\{X_1, \dots, X_n\}$, sin embargo no lo son en la σ -álgebra generada por la sucesión infinita $\{X_1, X_2, \dots\}$. Entonces podemos generalizar esto para obtener

$$\mathbb{P}_{\theta_a}\{S_n \leq an\} = \mathbb{E}_{\theta_a}[\exp\{-\theta_a S_n + n\psi(\theta_a)\}; S_n \leq an].$$

Más adelante, se verificará que si se cumple la condición $\gamma_a < \gamma_{-a}$ entonces \mathbb{P}_{θ_a} es consistente con la conducta de grandes desviaciones que describe $\mathbb{P}\{|S_n| \geq an\}$. Por lo que la obvia extensión del método ya descrito, sería generar repeticiones independientes de

$$\hat{\alpha}(n) = \exp\{-\theta_a S_n + n\psi(\theta_a)\} \mathbb{I}_{\{|S_n| \geq an\}} = L_n \mathbb{I}_{\{|S_n| \geq an\}},$$

que es un estimador generado bajo \mathbb{P}_{θ_a} .

Al analizar el comportamiento de este estimador nos daremos cuenta de algunos inconvenientes, como el hecho de que $\hat{\alpha}$ tiene varianza infinita (*Teorema 3.1*), por lo que se buscará una alternativa a nuestro método tradicional para generar un estimador y así refinarlo, obteniendo la siguiente propuesta

$$\hat{\beta}_+(n) = \exp\{\theta_a S_n + n\psi(\theta_a)\} \mathbb{I}_{\{S_n \geq an\}},$$

generado bajo \mathbb{P}_{θ_a} y

$$\hat{\beta}_-(n) = \exp\{\theta_{-a}S_n + n\psi(\theta_{-a})\}\mathbb{I}_{\{S_n \leq -an\}},$$

generado independientemente bajo $\mathbb{P}_{\theta_{-a}}$. Al sumar estos dos últimos estimadores, obtendremos una mejor estimación del evento en cuestión.

De tal forma que $\hat{\beta}(n) = \hat{\beta}_+(n) + \hat{\beta}_-(n)$, es el estimador propuesto, y como es natural también se realizará un análisis sobre la eficiencia del mismo (*Proposición 5.3*).

Proposición 5. *Supongamos $\gamma_a < \gamma_{-a}$, entonces:*

1. $\lim_{n \rightarrow \infty} -\frac{1}{n} \log \mathbb{P}\{|S_n| \geq an\} = \gamma_a$.
2. L_n en $\hat{\alpha}(n)$ satisface (4.4) con $\gamma = \gamma_a$.
3. $\hat{\beta}(n)$ es asintóticamente eficiente.

Demostración

Para probar el punto 1, recordamos los resultados del *Teorema de Cramér* previamente expuestos, es decir, nuestras ecuaciones (4.6) y (4.7).

Si utilizamos una propiedad general de la función logaritmo que establece que

$$\log(x + y) = \log(x) + O\left(\frac{y}{x}\right), \quad \text{si } \frac{y}{x} \rightarrow 0,$$

podemos mostrar lo siguiente

$$\begin{aligned} \lim_{n \rightarrow \infty} -\frac{1}{n} \log \mathbb{P}\{|S_n| \geq an\} &= \lim_{n \rightarrow \infty} -\frac{1}{n} \left[\log (\mathbb{P}\{S_n \geq na\} + \mathbb{P}\{S_n \leq -an\}) \right] \\ &= \lim_{n \rightarrow \infty} -\frac{1}{n} \log \mathbb{P}\{S_n \geq na\} + \lim_{n \rightarrow \infty} O\left(\frac{\mathbb{P}\{S_n \geq na\}}{\mathbb{P}\{S_n \leq -an\}}\right) \\ &= \theta_a a - \psi(\theta_a) \\ &= \gamma_a. \end{aligned}$$

En el punto 2 tenemos

$$\hat{\alpha}(n) = \exp\{-\theta_a S_n + n\psi(\theta_a)\}\mathbb{I}_{\{|S_n| \geq an\}} = L_n \mathbb{I}_{\{|S_n| \geq an\}},$$

lo que implica que

$$L_n = \exp\{-\theta_a S_n + n\psi(\theta_a)\},$$

de donde

$$-\frac{1}{n} \log L_n = \theta_a \frac{S_n}{n} - \psi(\theta_a).$$

Al tomar en cuenta que la $\mathbb{E}_\theta[X_1] = \psi'(\theta)$ y basados en el resultado de la *Ley Fuerte de los Grandes Números*, sabemos que S_n/n converge a la media μ con probabilidad igual a 1. Entonces ya que $\psi'(\theta_{\pm a}) = \pm a$ (por hipótesis) obtenemos el resultado buscado, es decir

$$-\frac{1}{n} \log L_n = \theta_a \frac{S_n}{n} - \psi(\theta_a) \longrightarrow \theta_a a - \psi(\theta_a),$$

donde esta convergencia es casi segura con respecto a \mathbb{P}_{θ_a} , conforme n crece.

Para poder probar 3., nos basamos en que el

$$\lim_{n \rightarrow \infty} -\frac{1}{n} \log \mathbb{E}_{\theta_{\pm a}}[\hat{\beta}_{\pm}^2] = 2\gamma_{\pm a},$$

puesto que $\hat{\beta}_{\pm}$ es un estimador de la forma $L_x \mathbb{I}_{\{A_x\}}$. Entonces al utilizar la desigualdad de *Jensen* y aplicar un razonamiento similar al realizado en la descripción del problema general, tenemos que para β_+ se cumple

$$\mathbb{E}_{\theta_a}[\hat{\beta}_+^2] \geq \mathbb{E}_{\theta_a}^2[\hat{\beta}_+] = \mathbb{P}_{\theta_a}^2\{S_n \geq an\}.$$

Por lo que

$$-\frac{1}{n} \log \mathbb{E}_{\theta_a}[\hat{\beta}_+^2] \leq -\frac{1}{n} \log \mathbb{P}_{\theta_a}^2\{S_n \geq an\}.$$

La no negatividad de la varianza nos da

$$2\gamma_{+a} \leq \limsup_{n \rightarrow \infty} -\frac{1}{n} \log \mathbb{E}_{\theta_a}[\hat{\beta}_+^2] \leq \limsup_{n \rightarrow \infty} -\frac{1}{n} \log \mathbb{P}_{\theta_a}^2\{S_n \geq an\} \leq 2\gamma_{+a}. \quad (4.8)$$

Por lo tanto

$$\limsup_{n \rightarrow \infty} -\frac{1}{n} \log \mathbb{E}_{\theta_a}[\hat{\beta}_+^2] = 2\gamma_{+a}.$$

Notando que los límites inferiores también se encuentran acotados como en (4.8), obtenemos

$$\liminf_{n \rightarrow \infty} -\frac{1}{n} \log \mathbb{E}_{\theta_a}[\hat{\beta}_+^2] = 2\gamma_{+a}.$$

De tal forma que

$$\lim_{n \rightarrow \infty} -\frac{1}{n} \log \mathbb{E}_{\theta_a}[\hat{\beta}_+^2] = 2\gamma_{+a}.$$

Un razonamiento análogo sobre el segundo momento de $\hat{\beta}_-$ nos lleva a concluir que

$$\lim_{n \rightarrow \infty} -\frac{1}{n} \log \mathbb{E}_{\theta_{-a}}[\hat{\beta}_-^2] = 2\gamma_{-a}.$$

Por lo tanto, basándonos en que $\hat{\beta}(n)$ es la suma de $\hat{\beta}_+(n)$ y $\hat{\beta}_-(n)$, implica que el logaritmo del segundo momento decrece asintóticamente a una tasa *no menor* que la más pequeña de las tasas de convergencia para los dos estimadores por separado. Esto significa que la tasa de decrecimiento de $\hat{\beta}(n)$ es *mayor o igual* que el $\min\{2\gamma_{+a}, 2\gamma_{-a}\} = 2\gamma_{+a}$, ya que $\gamma_{+a} < \gamma_{-a}$. Y al utilizar nuevamente la no negatividad de la varianza de $\hat{\beta}(n)$ tenemos que la tasa debe ser *menor o igual* que este mismo mínimo, por lo tanto resulta que esta tasa de decrecimiento es exactamente el mínimo de las dos por separado. Es decir que

$$\lim_{n \rightarrow \infty} -\frac{1}{n} \log \mathbb{E}_{\theta}[\hat{\beta}^2] = 2\gamma_a.$$

Con lo que se concluye que nuestro estimador alternativo es asintóticamente eficiente.

◇

El siguiente *Teorema* analiza el comportamiento de $\hat{\alpha}(n)$, el estimador sugerido por el término asintótico de $\mathbb{P}\{|S_n| \geq an\}$. En este caso notaremos que la eficiencia del estimador puede ser muy pobre, ya que a pesar de que \mathbb{P}_{θ_a} asigna poca probabilidad al evento $\{S_n \leq -an\}$, L_n puede ser muy grande en este conjunto, estableciendo lo que en este capítulo hemos denominado como *Contra-ejemplo*.

Teorema 3. *Supongamos que $\gamma_a < \gamma_{-a}$*

1. *Si $\psi(-\theta_a) = \infty$, entonces $\hat{\alpha}(n)$ tiene varianza infinita.*
2. *Si $\theta_a + \theta_{-a} \geq 0$ y $\psi(-\theta_a) < \infty$, entonces $\mathbb{E}_{\theta_a}[\hat{\alpha}^2(n)] \rightarrow \infty$.*
3. *Si $\theta_a + \theta_{-a} < 0$, entonces $\frac{1}{n} \log \mathbb{E}_{\theta_a}[\hat{\alpha}^2(n)] \rightarrow \min\{2\gamma_a, -\psi(\theta_a) - \psi(\theta_{-a}) - a(\theta_a + \theta_{-a})\}$.*

Demostración

Para demostrar el punto 1, notamos que

$$\begin{aligned}
\mathbb{E}_{\theta_a}[\hat{\alpha}^2(n)] &= \mathbb{E}_{\theta_a}[\exp\{-2(\theta_a S_n - n\psi(\theta_a))\}; |S_n| \geq an] \\
&= \int_{\{|S_n| \geq an\}} \exp\{-2(\theta_a S_n - n\psi(\theta_a))\} e^{\theta_a S_n} M(\theta_a)^{-n} dF^n \\
&= \int_{\{|S_n| \geq an\}} \exp\{-(\theta_a S_n - n\psi(\theta_a))\} dF^n \\
&= \mathbb{E}[\exp\{-(\theta_a S_n - n\psi(\theta_a))\}; |S_n| \geq an] \\
&= \mathbb{E}[\exp\{-(\theta_a S_n - n\psi(\theta_a))\}; S_n \geq an] \\
&\quad + \mathbb{E}[\exp\{-(\theta_a S_n - n\psi(\theta_a))\}; S_n \leq -an].
\end{aligned}$$

Lo que implica que

$$\begin{aligned}
\mathbb{E}_{\theta_a}[\hat{\alpha}^2(n)] &\geq \mathbb{E}[\exp\{-(\theta_a S_n - n\psi(\theta_a))\}; S_n \leq -an] \\
&= \mathbb{E}[\exp\{-(\theta_a S_n - n\psi(\theta_a))\}] - \mathbb{E}[\exp\{-(\theta_a S_n - n\psi(\theta_a))\}; S_n > -an] \\
&= \exp\{n\psi(\theta_a)\} \left(\mathbb{E}[\exp\{-\theta S_n\}] - \mathbb{E}[\exp\{-\theta S_n\}; S_n > -an] \right) \\
&\geq \exp\{n\psi(\theta_a)\} \left(\mathbb{E}[\exp\{-\theta S_n\}] - \exp\{-\theta_a a\} \right) \\
&= \exp\{n\psi(\theta_a)\} \left(\exp\{n\psi(-\theta_a)\} - \exp\{n\theta_a a\} \right) \\
&= \infty.
\end{aligned}$$

Ya que $\psi(-\theta_a) = \infty$, tenemos que como el segundo momento del estimador analizado no es acotado, entonces $\hat{\alpha}$ tiene varianza infinita.

Para probar el punto 2, escribamos

$$\begin{aligned}
&\mathbb{E}[\exp\{-(\theta_a S_n - n\psi(\theta_a))\}; S_n \leq -an] \\
&= \mathbb{E}_{-\theta_a}[\exp\{-\theta_a S_n + n\psi(\theta_a) - (-\theta_a)S_n + n\psi(\theta_a)\}; S_n \leq -an] \quad (4.9) \\
&= \exp\{n\psi(\theta_a)\} + n\psi(-\theta_a) \mathbb{P}_{-\theta_a}\{S_n \leq -an\}.
\end{aligned}$$

De tal forma que si $\theta_a + \theta_{-a} \geq 0$ entonces $\mathbb{E}_{-\theta_a}[X_1] = \psi'(-\theta) \leq \psi'(\theta_{-a}) = -a$, pues $\theta_{-a} \geq -\theta_a$ y ψ es convexa.

Si notamos que el $\liminf_{n \rightarrow \infty} \mathbb{P}_{-\theta_a} \{S_n \leq -an\} > 0$ y hacemos uso de la convexidad de ψ , tenemos que $\psi(\theta_a) + \psi(\theta_{-a}) > 2\psi(0) = 0$, lo que implica

$$\mathbb{E}_{\theta_a}[\hat{\alpha}^2(n)] \geq \exp\{n(\psi(\theta_a) + \psi(\theta_{-a}))\} \mathbb{P}_{-\theta_a} \{S_n \leq -an\}.$$

Y como $\psi(\theta_a) + \psi(\theta_{-a}) > 0$ podemos asegurar que existe un $v > 0$ tal que $\psi(\theta_a) + \psi(\theta_{-a}) > v$, de modo que

$$\begin{aligned} \mathbb{E}_{\theta_a}[\hat{\alpha}^2(n)] &\geq \exp\{nv\} \mathbb{P}_{-\theta_a} \{S_n \leq -an\} \\ &\geq \liminf_{n \rightarrow \infty} \{ \exp\{nv\} \mathbb{P}_{-\theta_a} \{S_n \leq -an\} \} \\ &= \infty. \end{aligned}$$

Por lo tanto

$$\mathbb{E}_{\theta_a}[\hat{\alpha}^2(n)] \rightarrow \infty.$$

Con este resultado, lo único que nos resta probar es el punto 3, y para poder demostrarlo, debemos observar que la hipótesis principal, $\theta_a + \theta_{-a} < 0$, implica que $\theta_{-a} < -\theta_a$, entonces

$$-a = \psi'(\theta_{-a}) < \psi'(-\theta_a) = \mathbb{E}_{-\theta_a}[X_1].$$

Definamos ahora a

$$\begin{aligned} \psi_{-\theta_a} &:= \log \mathbb{E}_{-\theta_a}[\exp\{\theta X_1\}] \\ &= \log \mathbb{E}[\exp\{-\theta_a X_1 - \psi(-\theta_a) + \theta X_1\}] \\ &= \log \mathbb{E}[e^{X_1(\theta - \theta_a)}] - \log \mathbb{E}[e^{-\theta_a X_1}] \\ &= \psi(\theta - \theta_a) - \psi(-\theta_a) \end{aligned} \tag{4.10}$$

Al derivar ambos miembros de la igualdad con respecto a θ , obtenemos que

$$\psi'_{-\theta_a} = \psi'(\theta - \theta_a),$$

de modo que $\bar{\theta}_{-a} \equiv \theta_a + \theta_{-a}$ resuelve la ecuación

$$\psi'_{-\theta_a}(\bar{\theta}_{-a}) = \psi'(\theta_{-a} - \theta_a) = \psi'(\theta_{-a}) = -a.$$

Al hacer uso de este resultado junto con en el *Teorema de Cramér* logramos ver lo siguiente

$$\begin{aligned}
\lim_{n \rightarrow \infty} -\frac{1}{n} \log \mathbb{P}_{-\theta_a} \{S_n \leq an\} &= -\bar{\theta}_{-a}a - \psi_{-\theta_a}(\bar{\theta}_{-a}) \\
&= -a(\theta_a + \theta_{-a}) - \psi(\bar{\theta}_{-a} - \theta_a) + \psi(-\theta_a) \\
&= -a(\theta_a + \theta_{-a}) - \psi(\theta_{-a}) + \psi(-\theta_a).
\end{aligned}$$

Entonces bajo la expresión (4.9) tenemos

$$\begin{aligned}
&-\frac{1}{n} \log \mathbb{E}[\exp\{-(\theta_a S_n - n\psi(\theta_a))\}; S_n \leq -an] \\
&= -\frac{1}{n} \log \left\{ \exp\{n\psi(\theta_a) + n\psi(-\theta_a)\} \mathbb{P}_{-\theta_a} \{S_n \leq -an\} \right\} \\
&= -\frac{1}{n} \log \left\{ \exp\{n\psi(\theta_a) + n\psi(-\theta_a)\} \right\} - \frac{1}{n} \log \mathbb{P}_{-\theta_a} \{S_n \leq -an\} \\
&= -\psi(\theta_a) - \psi(-\theta_a) - \frac{1}{n} \log \mathbb{P}_{-\theta_a} \{S_n \leq -an\} \\
&= -\psi(\theta_a) - \psi(-\theta_a) - a(\theta_a + \theta_{-a}) - \psi(\theta_{-a}) + \psi(-\theta_a) \\
&= -\psi(\theta_a) - \psi(\theta_{-a}) - a(\theta_a + \theta_{-a}).
\end{aligned}$$

Ahora, utilizando parte de la prueba del punto 3 de la *Proposición 5*, sabemos que

$$-\frac{1}{n} \log \mathbb{E}[\exp\{-(\theta_a S_n - n\psi(\theta_a))\}; S_n \geq an] \longrightarrow 2\gamma_a, \quad \text{conforme } n \rightarrow \infty.$$

Por lo tanto

$$-\frac{1}{n} \log \mathbb{E}_{\theta_a} [\hat{\alpha}^2(n)] \longrightarrow \min\{2\gamma_a, -\psi(\theta_a) - \psi(\theta_{-a}) - a(\theta_a + \theta_{-a})\}, \quad \text{conforme } n \rightarrow \infty.$$

◇

Capítulo 5

Aplicaciones.

El objetivo principal de este capítulo es mostrar algunas de las aplicaciones del *Muestreo de Importancia* a problemas específicos. Analizaremos el aspecto práctico del método propuesto en este trabajo en escenarios tales como el cálculo de medidas de riesgo como el **VaR**, la optimización en la elección de **Portafolios de Inversión**, la **Valuación de Opciones** y el problema de la **Ruina**.

§5.1. Muestreo de Importancia para el Valor en Riesgo.

El llamado *Valor en Riesgo*, mejor conocido como *VaR* por sus siglas en inglés, se ha convertido en una importante medida para estimar y administrar un portafolio de inversión. Esta medida de riesgo se define como un cierto cuantil del cambio en el valor del portafolio de inversión durante un periodo de tiempo específico. Para ser más claros, supongamos que el valor de un portafolio de inversión al tiempo t , está determinado por la función $V(t)$, que el periodo de tiempo al que está sujeto es Δt , y que el valor del portafolio al tiempo $t + \Delta t$ es $V(t + \Delta t)$.

Denominaremos a la pérdida en el valor del portafolio durante el periodo de tiempo al que está sujeto como $\ell = -\Delta V$, en donde $\Delta V = [V(t + \Delta t) - V(t)]$ y definiremos al VaR_{x_p} asociado con una probabilidad p , por la siguiente relación

$$\mathbb{P}\{\ell > x_p\} = p. \tag{5.1}$$

Es decir que el VaR_{x_p} es el cuantil $(1 - p)$ de la distribución de pérdida.

Comunmente Δt es tomado como un día o dos semanas y $p \leq 0.05$. Muchas veces $p = 0.01$ es de gran interés ya que con este valor de p se asegura una confianza del 99% sobre el nivel de pérdida.

Para valuar (5.1), los métodos Monte Carlo son utilizados en la mayoría de las ocasiones debido al amplio material existente. De modo que aproximar los cambios de los factores de riesgo del portafolio de inversión se hace posible y con esto re-valorar el portafolio y calcular la distribución de pérdida. No obstante, la obtención de estimaciones exactas del VaR puede ser computacionalmente costosa, ya que:

1. En un portafolio de inversión, por lo general se encuentra un gran número de instrumentos, es decir que nuestra variable en cuestión es de considerable complejidad y por lo tanto para poder hacer la valuación del portafolio se requiere de presupuestos altos.
2. Cuando p es pequeña, una gran cantidad de simulaciones son requeridas para lograr estimaciones acertadas sobre la cola de la distribución de probabilidad.

Nuestro propósito radica entonces en describir una técnica de reducción de varianza, que nos ayude a disminuir sustancialmente la varianza para alcanzar una determinada precisión. Bajo este contexto, nuestra propuesta obvia es el muestreo de importancia pues la clave para reducir la varianza de una estimación del VaR_{x_p} es obtener estimaciones exactas o certeras de $\mathbb{P}\{\ell > x\}$, para valores de x cercanos a x_p y de esta forma concentrar nuestros esfuerzos en esta zona. Nuestro fundamento básico es explotar el conocimiento que tenemos sobre la distribución de una aproximación a la función de pérdida y así idear esquemas de muestreo Monte Carlo más efectivos. Las aproximaciones a la función de pérdida empleadas en este problema son expansiones de series de Taylor de primer y segundo orden sobre ℓ . Dichas expansiones, son las aproximaciones Delta y Delta-Gamma respectivamente. Estas aproximaciones son utilizadas ya que si los factores de riesgo tienen una distribución normal multivariada, como se asume en muchos de los casos prácticos, la distribución de la aproximación Delta se conoce en su forma analítica, mientras que la distribución de la aproximación Delta-Gamma puede calcularse numéricamente.

Estas aproximaciones **No** siempre son lo suficientemente exactas para proveer estimaciones precisas del VaR , sin embargo de la correlación entre la aproxima-

ción y la pérdida actual, el conocimiento sobre la aproximación puede ponernos en gran ventaja para cumplir nuestro propósito y reducir la varianza. Si estamos interesados en estimar $\mathbb{P}\{\ell > x\}$, entonces podemos hacer uso de $\mathbb{I}_{\{Q > x\}}$ como variable de control, en donde Q es la aproximación a la pérdida. Específicamente esto significa reemplazar la estimación estandar $\mathbb{I}_{\{\ell > x\}}$ con

$$\mathbb{I}_{\{\ell > x\}} - \beta[\mathbb{I}_{\{Q > x\}} - \mathbb{P}\{Q > x\}],$$

donde ℓ y Q son valuadas en el mismo escenario de precios y $\mathbb{P}\{Q > x\}$ se calcula numericamente. El coeficiente β puede estimarse a partir de la simulación que minimice la varianza.

El principal problema de estimar $\mathbb{P}\{\ell > x\}$ para valores grandes de x , es el hecho de que en muy pocas muestras se encontrará el evento $\{\ell > x\}$, y como consecuencia en muchas de las muestras se tiene que $\mathbb{I}_{\{\ell > x\}} = 0$. Como sabemos, una principal característica del muestreo de importancia, es la buena adaptación en la simulación de eventos “raros” o de baja probabilidad. Entonces si suponemos que la densidad conjunta del cambio de los factores de riesgo es f , al aplicar nuestro método tenemos que en lugar de simular con esta densidad, una diferente densidad g será utilizada, de tal forma que podemos escribir a la probabilidad de pérdida de la siguiente manera

$$\mathbb{P}\{\ell > x\} = \int \mathbb{I}_{\{\ell > x\}} f(z) dz = \int \mathbb{I}_{\{\ell > x\}} \frac{f(z)}{g(z)} g(z) dz = \mathbb{E}_g[\mathbb{I}_{\{\ell > x\}} L(Z)]. \quad (5.2)$$

A continuación mostraremos cómo utilizar las aproximaciones Delta y Delta-Gamma como guía para la elección de una efectiva distribución de muestreo.

Observaremos además en el desarrollo de este ejemplo, que la aplicación de esta técnica de reducción de varianza (el **IS**) puede mejorarse al combinarse con la estratificación y que en particular la combinación de estos dos métodos no involucra alguna clase de costo extra. Por el contrario los gastos generales de una implementación de este tipo son sumamente pequeños. Para ser más claros, si suponemos que la aproximación Delta-Gamma está dada, los únicos gastos son:

1. Costo en términos del tiempo empleado para calcular el cambio de medida del muestreo de importancia.
2. Tiempo empleado para calcular los cuantiles relacionados con el muestreo estratificado.

3. Un costo adicional por cada muestra al generar los factores de riesgo a partir de la distribución condicional de los estratos y calcular su respectiva razón de verosimilitudes, costo que es insignificante comparado con lo que se necesitaría para evaluar incluso un portafolio de inversión de tamaño modesto.

§5.1.1. Aproximaciones Delta y Delta-Gamma.

Comenzaremos por describir la forma de las aproximaciones Delta y Delta-Gamma como sumas de términos que envuelven variables aleatorias independientes normales estándar. Esto facilita el cálculo de las cantidades requeridas para el control del muestreo de importancia y la estratificación.

Supondremos por el momento que hay m factores de riesgo, y que

$$S(t) = (S_1(t), \dots, S_m(t)),$$

denota el valor de estos factores al tiempo t . Definamos a $\Delta S = [S(t + \Delta t) - S(t)]^T$ como el cambio en los factores de riesgo durante el intervalo de tiempo $[t, t + \Delta t]$. Por lo tanto la aproximación Delta-Gamma está dada por la siguiente ecuación

$$\Delta V = -\ell \approx \Theta \Delta t + \delta^T \Delta S + \frac{1}{2} \Delta S^T \Gamma \Delta S, \quad (5.3)$$

donde $\Theta = \frac{\partial V}{\partial t}$, $\delta_i = \frac{\partial V}{\partial S_i(t)}$, y $\Gamma_{ij} = \frac{\partial^2 V}{\partial S_i(t) \partial S_j(t)}$, (todas la derivadas parciales son evaluadas en $S(t)$). Mientras que la aproximación Delta está dada por

$$\Delta V = -\ell \approx \Theta \Delta t + \delta^T \Delta S$$

Si ΔS tiene distribución normal multivariada con vector de medias cero y matriz de varianzas y covarianzas Σ , entonces para obtener muestras de ΔS podemos fijar $\Delta S = CZ$, donde Z es un vector de m normales estándar independientes entre si y C es una matriz tal que $CC^T = \Sigma$, de tal forma que para la aproximación Delta tenemos que

$$\mathbb{P}\{\ell > x\} \approx \mathbb{P}\{b^T Z > x + \Theta \Delta t\} \equiv \mathbb{P}\{Y > y_x\}, \quad (5.4)$$

en donde

$$b^T = -\delta^T C, \quad Y = b^T Z, \quad y_x = x + \Theta \Delta t. \quad (5.5)$$

Para la aproximación Delta-Gamma, buscamos expresar $\ell \approx c + b^T Z + Z^T \Lambda Z \equiv c + Q$ donde cada una de las Z 's sea normal estándar independientes entre si y Λ

sea una matriz diagonal. Entonces lo único que nos falta, es elegir C de tal forma que

$$C^T C = \Sigma \quad y \quad \frac{1}{2} C^T \Gamma C \quad \text{sea diagonal.} \quad (5.6)$$

Para mostrar que tal elección es posible, iniciemos con una matriz C arbitraria para la cual se cumpla que $CC^T = \Sigma$ y escribamos $\frac{1}{2} C^T \Gamma C = -U\Lambda U^T$, donde U es una matriz ortogonal con la propiedad de que sus columnas son los vectores propios de $\frac{1}{2} C^T \Gamma C$ y $-\Lambda$ es la matriz de valores propios de $\frac{1}{2} C^T \Gamma C$ (y en consecuencia también de $\frac{1}{2} \Gamma \Sigma$), lo cual es viable al aplicar el *Teorema de Descomposición Espectral*. Ahora al reemplazar la matriz original C con CU obtenemos

$$\begin{aligned} CU(CU)^T &= CUU^T C^T = C C^T \quad \text{pues } U \text{ es ortogonal} \\ \Rightarrow \frac{1}{2} (CU)^T \Gamma CU &= \frac{1}{2} U^T C^T \Gamma CU = -U^T U \Lambda U^T U = -\mathbb{I} \Lambda \mathbb{I} = -\Lambda \quad \text{que es diagonal.} \end{aligned}$$

De modo que la nueva elección satisface (5.6). Entonces, al seguir la estructura de (5.6) observamos que

$$\ell \approx -\Theta \Delta t - \delta^T CZ = -\Theta \Delta t + b^T Z + Z^T \Lambda Z. \quad (5.7)$$

Para $b^T = -\delta^T C$ y Z es un vector de normales estándar independientes. Entonces

$$\mathbb{P}\{\ell > x\} \approx \mathbb{P}\{Q > x + \Theta \Delta t\} = \mathbb{P}\{Q > y_x\}, \quad (5.8)$$

donde $Q = b^T Z + Z^T \Lambda Z = \sum (b_i Z_i + \lambda_i Z_i^2)$ es la parte estocástica de la aproximación cuadrática de ℓ . Por lo tanto al completar el cuadrado tenemos que la distribución de Q puede ser relacionada con una suma de variables aleatorias ji-cuadradas no centradas.

§5.1.2. Preliminares para la Reducción de Varianza.

En esta sección mostraremos la flexibilidad del IS enfatizando el hecho de que puede utilizarse no sólo como técnica inicial, si no también puede combinarse con otros métodos como el *Muestreo Estratificado* para obtener mejores resultados o reparar problemas que aparezcan en el proceso de aplicación y de este modo nos daremos cuenta de que combinaciones de estos dos métodos muestran un mejor

desempeño bajo circunstancias específicas. Como ya se ha descrito en (5.2), podemos reescribir a $\mathbb{P}\{\ell > x\}$ como $\mathbb{E}_g[\mathbb{I}_{\{\ell > x\}}L]$, en donde este último denota la esperanza bajo la distribución de muestreo para Z y su respectiva razón de verosimilitudes. Sabemos que el producto de estas dos últimas es el estimador propuesto por el muestreo de importancia. En el contexto descrito en la sección anterior Z es un vector de normales estándar, entonces bajo este supuesto, podemos considerar dos tipos de **IS**, uno en el que únicamente cambiemos la media de la distribución de Z y otro donde ambos parámetros de la distribución de nuestro vector aleatorio normal estándar sean cambiados. Entonces si la media de la distribución de Z se cambia de 0 a v , tenemos

$$L(Z) = \frac{f(Z)}{g(Z)} = \frac{\frac{1}{\sqrt{2\pi}}e^{\{-\frac{1}{2}Z^T Z\}}}{\frac{1}{\sqrt{2\pi}}e^{\{-\frac{1}{2}(Z-v)^T \mathbb{I}(Z-v)\}}} = e^{\{\frac{1}{2}v^T v - v^T Z\}}. \quad (5.9)$$

Si el cambio es en ambos parámetros, cambiando la media de 0 a v y la matriz de varianzas y covarianzas de \mathbb{I} a B con $B > 0$ (definida positiva) implica que

$$L(Z) = \frac{f(Z)}{g(Z)} = \frac{e^{\{-\frac{1}{2}Z^T Z\}}}{|B|^{-\frac{1}{2}}e^{\{-\frac{1}{2}(Z-v)^T B^{-1}(Z-v)\}}}, \quad (5.10)$$

en donde $|B|$ es el determinante de B . Por lo tanto si simulamos n muestras (Z^1, \dots, Z^n) y calculamos la pérdida asociada al vector Z y la denotamos como (ℓ_1, \dots, ℓ_n) , tenemos que el estimador propuesto para $\mathbb{P}\{\ell > x\}$ es

$$\hat{\mathbb{P}}_x = \frac{1}{n} \sum_{j=1}^n \mathbb{I}_{\{\ell > x\}} L(Z^j). \quad (5.11)$$

De este modo si X es una variable aleatoria con media conocida, la variable de control propuesta bajo el **IS** deberá ser $X \cdot L(Z)$, que tiene la misma media.

Por otro lado, en el muestreo estratificado se tiene que identificar la variable de estratificación X y posteriormente partir el rango de X en k intervalos o estratos (S_1, \dots, S_k) lo que nos lleva a escribir

$$\mathbb{P}\{\ell > x\} = \sum_{i=1}^k \mathbb{P}\{\ell > x | X \in S_i\} \mathbb{P}\{X \in S_i\}. \quad (5.12)$$

Comunmente se utiliza $k = 25$ hasta $k = 100$ estratos equiprobables (es decir, $\mathbb{P}\{X \in S_i\} = 1/k$), posteriormente se sacan n_i muestras de X de cada uno de los estratos i existentes, y se define a X_{ij} como la j -ésima muestra del estrato i . Tomemos Z^{ij} una muestra de Z con distribución condicional $Z | X = X_{ij}$, y denotemos

a ℓ_{ij} como la correspondiente pérdida del portafolio de la muestra Z^{ij} . Entonces $\mathbb{P}\{\ell > x\}$ puede estimarse como

$$\hat{\mathbb{P}}_x = \sum_{i=1}^k \mathbb{P}\{X \in S_i\} \times \frac{1}{n_i} \sum_{j=1}^{n_i} \mathbb{I}_{\{\ell_{ij} > x\}}, \quad (5.13)$$

en donde claramente tenemos la libertad de asignar los estratos, es decir, elegir n_i de una manera arbitraria. Para obtener mejores resultados debemos aplicar la asignación óptima para un número total de muestras específico con estratos equiprobables (hecho conocido), que es hacer n_i proporcional a la desviación estándar de $\mathbb{I}_{\{\ell_{ij} > x\}}$.

Ahora, cuando combinamos el IS con el muestreo estratificado, se puede pensar en la aplicación indistinta de cada uno de los métodos primero, es decir aplicar IS primero y después el muestreo estratificado o viceversa y así obtener dos distintos estimadores. Al aplicar primero el muestreo estratificado tenemos

$$\begin{aligned} \mathbb{P}\{\ell > x\} &= \sum_{i=1}^k \mathbb{P}\{\ell > x | X \in S_i\} \mathbb{P}\{X \in S_i\} \\ &= \sum_{i=1}^k \mathbb{E}[\mathbb{I}_{\{\ell_{ij} > x\}} L(Z) | X \in S_i] \mathbb{P}\{X \in S_i\}. \end{aligned} \quad (5.14)$$

Es decir que X se toma de la distribución original y después Z se obtiene de la distribución de muestreo dado X , y por lo tanto el estimador asociado con (5.14) es

$$\hat{\mathbb{P}}_x = \sum_{i=1}^k \mathbb{P}\{X \in S_i\} \times \frac{1}{n_i} \sum_{j=1}^{n_i} \mathbb{I}_{\{\ell > x\}} L(Z^{ij}) \quad (5.15)$$

Por otro lado, si aplicamos primero el muestreo de importancia y posteriormente la estratificación obtendremos

$$\mathbb{P}\{\ell > x\} = \mathbb{E}_g[\mathbb{I}_{\{\ell > x\}} L] = \sum_{i=1}^k \mathbb{E}_g[\mathbb{I}_{\{\ell > x\}} L(Z) | X \in S_i] \mathbb{P}_g\{X \in S_i\}, \quad (5.16)$$

en donde \mathbb{P}_g denota la probabilidad bajo la densidad de cambio g . Es claro que en este método, tanto X como Z dado X se obtienen de la distribución de muestreo, por lo tanto el estimador asociado es

$$\hat{\mathbb{P}}_x = \sum_{i=1}^k \mathbb{P}_g \{X \in S_i\} \times \frac{1}{n_i} \sum_{j=1}^{n_i} \mathbb{I}_{\{\ell_{ij} > x\}} L(Z^{ij}). \quad (5.17)$$

Para ver la distinción entre (5.14) y (5.16), supongamos que en ambos métodos empleamos k estratos equiprobables (bajo \mathbb{P} y \mathbb{P}_g respectivamente). Entonces si $X = b^T Z$ y $\ell \approx X$, al utilizar (5.14), la media de X es 0 y sólo unos cuantos estratos tendrán indicadora $\mathbb{I}_{\{\ell > x\}}$ positiva para un x grande. Ahora si usamos (5.16), la media de Z se cambia de 0 a v donde $b^T v = x$, entonces la media de X es x y en aproximadamente la mitad de los estratos se tiene una función indicadora positiva; de tal forma que al combinar estos dos métodos, primero se debe aplicar **IS** y posteriormente la *Estratificación*. Un esquema eficiente de muestreo para una variable de estratificación lineal $X = b^T Z$ se describe en el artículo escrito por *P. Glasserman, P. Heidelberger y P. Shahabuddin* [10], en donde a grandes rasgos, definen una variable $v = b/\sqrt{(b^T b)}$ y denotan a Φ^{-1} como la inversa de la distribución normal estándar de modo que al tomar una variable U uniforme en el $(0, 1)$ y un vector φ de m normales estándar independientes obtienen la siguiente transformación

$$Z = \Phi^{-1}(U)v + (I - vv^T)\varphi,$$

que por supuesto también es un vector de variables aleatorias normales estándar. Al reemplazar las U 's muestreadas independientemente por una muestra estratificada utilizando subintervalos del intervalo unitario como estratos, obtienen el efecto de estratificación sobre $v^T Z$ y por lo tanto también de $b^T Z$ (revisar [10] para mayores detalles).

§5.1.3. Reducción de Varianza con Base en la Aproximación Delta.

La técnica más obvia de reducción de varianza basada en la aproximación Delta es usar a la cola de la distribución de probabilidades de la aproximación como variable de control. Específicamente, (5.4) sugiere usar $\mathbb{I}_{\{Y > y_x\}}$ como variable de control, donde $Y = b^T Z$ y $y_x = x + \Theta \Delta t$. La media de esta variable se puede calcular fácilmente, no obstante como ya se ha discutido, la efectividad de esta aproximación disminuye según crece x . Cabe resaltar que puede existir otra $y' \neq y_x$ para la cual la variable de control $\mathbb{I}_{\{Y > y'\}}$ tenga menor varianza.

El muestreo estratificado en $Y = b^T Z$ usa mayor información sobre la aproximación Delta y evita algunas de las dificultades envueltas en la aplicación de

variables de control (es decir, seleccionamos la mejor y' y estimamos la variable de control óptima, múltiplo que podría introducir un sesgo e incluso generar estimaciones negativas de la probabilidad) sin embargo, a menos de que la asignación del estrato $\{n_i\}$ sea diseñada de tal forma que se coloquen más muestras en el estrato “prometedor”, entonces más muestras resultarán en $\ell < x$.

En la estimación de probabilidades de eventos “raros” como $\mathbb{P}\{\ell > x\}$, un criterio de elección para la distribución de muestreo se basa en el análisis de grandes desviaciones y como ya se ha mencionado repetidamente en este trabajo, dicho análisis nos dice que bajo ciertas condiciones técnicas, la probabilidad de un evento raro es aproximadamente igual a la probabilidad de “la trayectoria más parecida” al evento raro. Con esta idea en mente, la distribución de muestreo es elegida como la distribución que describa la trayectoria más parecida al evento en cuestión. Para distribuciones Gaussianas multivariadas esta aproximación ya se ha estudiado ampliamente cambiando la media de las variables de 0 a $\mu = (\mu_1, \dots, \mu_m)$ donde este punto es el que maximiza la probabilidad del evento. Por lo tanto es claro que al utilizar la aproximación dada por (5.4), encontraremos μ al resolver el siguiente problema de optimización

$$\text{máx} -\frac{1}{2} \sum_{i=1}^m z_i^2 \quad \text{tal que} \quad \sum_{i=1}^m b_i z_i \geq y_x. \quad (5.18)$$

Vía multiplicadores de Lagrange, obtenemos la siguiente solución: $\mu = by_x / (b^T b)$, lo que implica que la razón de verosimilitudes apropiada está dada por (5.9), con $v = \mu$. El segundo momento de este estimador es $\mathbb{E}_g[\mathbb{I}_{\{\ell > x\}} L^2(Z)] = \mathbb{E}[\mathbb{I}_{\{\ell > x\}} L(Z)]$, de modo que una condición suficiente para lograr la reducción de varianza es $L(Z) \leq 1$ para todos los puntos Z tales que $L > x$. La condición $L(Z) \leq 1$ es equivalente a $\frac{1}{2}\mu^T \mu - \mu^T Z \leq 0$, que al completar el cuadrado nos da como resultado que cada punto $Z \in \{\ell > x\}$ debe ser más cercano a μ que a 0. De hecho, en general se obtiene reducción de varianza con sólo cambiar la media de 0 a v , donde v es cualquier vector (distinto de μ), haciendo que cada punto en $\{\ell > x\}$ sea más cercano a v que a 0. Con un procedimiento análogo, si $L \leq f$ para toda $Z \in \{\ell > x\}$, el segundo momento del IS-estimador se reduce por lo menos en un factor de f . La varianza del IS-estimador puede reducirse aún más si utilizamos una variable de control como $\mathbb{I}_{\{Y > y_x\}} L(Z)$, que es la elección más obvia para nuestro caso.

Para combinar el IS con la estratificación, uno puede estratificar virtualmente sobre cualquier variable. En el caso de distribuciones normales es particularmente fácil estratificar sobre una combinación lineal de las Z 's, resaltando que cuando la media de la distribución de muestreo es μ , entonces $L(Z) = c_1 \exp(-\mu^T Z) = c_1 \exp(-c_2 b^T Z)$ para algunas constantes c_1 y c_2 . Estrictamente esto sugiere estratifi-

car sobre $\mu^T Z$ (equivalentemente $b^T Z$), ya que esto simultáneamente remueve toda la variabilidad en la razón de verosimilitudes y la mayor parte de la variabilidad en la función indicadora $\mathbb{I}_{\{\ell > x\}}$, lo que provoca que la aproximación Delta sea más cercana a la pérdida ℓ .

§5.1.4. Reducción de Varianza con Base en la Aproximación Delta-Gamma.

Como ya hemos visto, la aproximación Delta-Gamma puede usarse para derivar técnicas de reducción de varianza, primero (5.8) sugiere usar a $\mathbb{I}_{\{Q > y_x\}}$ como variable de control, o usar a Q como variable de estratificación. Para estratificar en Q , debemos poder simular Q y Z dado Q , puesto que $\mathbb{P}\{Q \in S_i\}$ puede calcularse numéricamente, entonces en principio se debe simular Q bajo el método de la transformada inversa⁷. Sin embargo, no contamos con un método directo para simular Z dada la forma cuadrática de Q , pero un método simple para generar Z a partir de la distribución condicional correcta es el siguiente

Primero, se debe generar un vector Z de variables normales estándar independientes y después calcular Q . Si $Q \in S_i$, entonces este vector Z tiene distribución Z dado $Q \in S_i$. Si hay menos que n_i muestras del estrato i entonces debemos utilizar esta Z para valuar el portafolio, de otra forma debemos descartarla. Es necesario continuar con el muestreo hasta obtener el número adecuado de muestras de cada estrato. Como esta idea está basada en el método de aceptación y rechazo (ver capítulo 2 de [9]), esto desecha algunas muestras, sin embargo, la experiencia de quienes han realizado este procedimiento ha sido que a excepción del $\{n_i\}$ más asimétrico u oblicuo (skewed), los gastos generales son modestos, especialmente en comparación con el costo de valuar el portafolio (ver [14]).

También es posible la estratificación sobre otras variables. Primero, supongamos que λ_1 es mucho mayor que λ_i , donde cada λ_i representa cada uno de los elementos de la matriz Λ con ($i > 1$) y $\lambda_1 > 0$. De esta forma gran parte de la variabilidad en la parte positiva de Z es explicada por Z_1^2 , lo cual sugiere estratificar en Z_1^2 , que claramente tiene distribución ji-cuadrada, al provenir de variables normales estándar. Esta distribución debe ser fácil de simular utilizando el método

⁷Método que afirma que si deseamos generar una v.a. X con distribución F , entonces lo que se debe hacer es resolver $X = F^{-1}(U)$ con U una v.a. uniforme en el $(0, 1)$. (ver capítulo 2 de [9] para mayores detalles)

de la transformada inversa. Si las λ_i 's son todas aproximadamente iguales, entonces gran parte de la variabilidad en Q es explicada por $R^2 = Z_1^2 + \dots + Z_m^2$, lo que nos motiva a estratificar sobre R^2 (estratificación radial). Cabe resaltar que R^2 tiene distribución ji-cuadrada con m grados de libertad. Ahora para completar el muestreo estratificado, primero hay que simular R^2 utilizando el método de la transformada inversa, y posteriormente simular m variables normales estándar independientes X_1, \dots, X_m y calcular $Z_i = X_i \sqrt{R} / \sqrt{X_1^2 + \dots + X_m^2}$. La ventaja primordial de estratificar sobre R^2 en lugar de hacerlo sobre Q es que existe este método directo de simular R^2 , y posteriormente obtener simulaciones de Z dado R^2 , no obstante, estratificar en Q es típicamente más efectivo que sobre R^2 .

La aproximación Delta-Gamma se puede utilizar en una variedad de formas distintas para así seleccionar la distribución de muestreo. Ya que contamos con un análisis previo del comportamiento de grandes desviaciones para este tipo de eventos (de manera similar a lo visto en la sección anterior), podemos aplicar la aproximación de la *trayectoria más parecida* en la que solamente el vector de medias es cambiado. En esta aproximación, deberíamos resolver el mismo problema de optimización que en (5.18), excepto porque la restricción cambia a $\sum (b_i z_i + \lambda_i z_i^2) \geq y_x$ la cual se deriva de la aproximación cuadrática más que de la lineal, como en la aproximación Delta. Al resolver este problema via multiplicadores de Lagrange, obtenemos que el vector de medias óptimo ahora tiene la forma $\mu_i = \beta b_i / (1 - 2\beta \lambda_i)$ para alguna constante de normalización β . No obstante, esta aproximación puede sufrir del mismo problema descrito en la sección anterior, es decir, nuestros resultados están sujetos a que el evento $\{\ell > x\}$ sea dominado por el punto μ , o de otra forma los resultados obtenidos serán poco informativos.

Si examinamos el cambio de medida propuesto por el muestreo de importancia en el cual no solamente el vector de medias es alterado, si no que además la matriz de varianzas y covarianzas de Z es cambiada, nos restringiremos a una forma particular de vector de medias y matriz de varianzas y covarianzas, que proviene de considerar un *cambio exponencial de medida* en la forma cuadrática de Q .

Sea θ el parámetro de cambio y definamos a

$$B_\theta = (\mathbb{I} - 2\theta\Lambda)^{-1} \quad y \quad \mu_\theta = \theta B_\theta b. \quad (5.19)$$

Cuando el muestreo de importancia está hecho a modo de fijar el vector de medias en μ_θ y la matriz de varianzas y covarianzas en B_θ , la razón de verosimilitudes de (5.19) se simplifica de la siguiente manera

$$L(Z) = \exp\{-\theta(b^T Z + Z^T \Lambda Z) + \psi(\theta)\} = \exp\{-\theta Q + \psi(\theta)\}, \quad (5.20)$$

donde

$$\psi(\theta) = \frac{1}{2} \sum_{i=1}^m \left(\frac{(\theta b_i)^2}{1 - 2\theta \lambda_i} - \ln(1 - 2\theta \lambda_i) \right), \quad (5.21)$$

es el logaritmo natural de la función generadora de momentos de la variable aleatoria Q . Debemos notar que cuando el \mathbf{IS} está hecho de esta forma, el único término aleatorio en la razón de verosimilitudes es Q . La gran pregunta es cómo elegir θ de manera adecuada para que se cumplan nuestros propósitos. Para ello, supongamos que la aproximación Delta-Gamma es exacta. Entonces, debido a la estructura de la muestra, el segundo momento del estimador es

$$\mathbb{E}_\theta[L(Z)^2 \mathbb{I}_{\{Q > y_x\}}] = \mathbb{E}_\theta[\exp\{-2\theta Q + 2\psi(\theta)\} \mathbb{I}_{\{Q > y_x\}}] \leq \exp\{-2\theta y_x + 2\psi(\theta)\}, \quad (5.22)$$

donde \mathbb{E}_θ denota la esperanza bajo la nueva medida cuando el parámetro de cambio es θ . Al elegir $\theta = \theta_x$ para minimizar el lado izquierdo de (5.22) obtenemos la mejor cota superior posible para el segundo momento (aunque no necesariamente el valor más pequeño del segundo momento), mientras no exista una forma cerrada para la expresión de θ_x , podemos fácilmente encontrar una al resolver la ecuación no lineal $\psi'(\theta_x) = y_x$. La función ψ es estrictamente convexa y $|\psi'(\theta)|$ tiende a infinito conforme $|\theta|$ tiende a infinito, asegurando la existencia de una única solución que en teoría se puede calcular numéricamente. Con este valor de θ_x la media de Q bajo el \mathbf{IS} es igual a y_x , aún más, como ya se ha demostrado, hacer la selección del parámetro de cambio bajo este método nos proporciona una “técnica asintóticamente óptima” conforme x crece. Bajo el supuesto de que la aproximación Delta-Gamma es exacta, estrictamente hablando, el hecho de ser asintóticamente óptimo significa que el segundo momento tiende a cero a una tasa del doble de la tasa con la que el primer momento tiende a cero, la cual es la mejor tasa posible ya que la varianza es no negativa. Sin embargo, la simulación puede ser no requerida si la aproximación Delta-Gamma es exacta, este pequeño análisis indica que el procedimiento que implica el muestreo de importancia es muy efectivo en la práctica.

§5.2. Valuación de Opciones Exóticas.

La valuación de opciones es una de las áreas de más desarrollo e importancia en las matemáticas financieras de los últimos años, esto se debe en gran parte a que ayudan a disminuir la incertidumbre que existe sobre el mercado de precios al determinar su valor mediante activos fundamentales tratados en cualquier tipo de Mercado Financiero⁸. En otras palabras las opciones son instrumentos financieros de transferencia de riesgo que basan su valor en el valor de un activo fundamental denominado *subyacente*.⁹ Las opciones fueron diseñadas para que el poseedor se beneficie de los movimientos del mercado a su favor pero no sufra pérdidas como consecuencia de movimientos del mercado en su contra. A grandes rasgos, una opción es un contrato que le da al propietario el derecho, más no la obligación de comprar o vender el subyacente negociado. Los dos tipos básicos de opciones son: la opción de compra mejor conocida como *Call* y la opción de venta o *Put*. Un *Call* le da a su dueño el derecho, pero no la obligación de comprar el subyacente en cuestión a un cierto precio pactado en una fecha específica, mientras que un *Put* por su parte le otorga a su poseedor el derecho pero no la obligación de vender el subyacente a un precio y fecha determinados.

Existe una gran variedad de tipos de opciones financieras como reflejo de la complejidad de los mercados financieros y los instrumentos que ahí se manejan. Sin embargo para propósitos de este documento nos basta con saber que dependiendo de la tipología del contrato las opciones se dividen de manera básica en:

1. *Europeas*: Son aquellas opciones que se pueden ejecutar únicamente al final del contrato, es decir que pueden ejercerse solamente en la fecha de vencimiento estipulada por las partes involucradas.
2. *Americanas*: Son aquellas opciones que se pueden ejecutar durante toda la duración del contrato, es decir que pueden ejercerse en cualquier momento dentro de los límites de tiempo convenidos en el contrato.

Dependiendo de la complejidad del contrato, las opciones se clasifican en *convencionales* y *exóticas*, donde las primeras son las más usuales y sencillas, mientras que las segundas adquieren su nombre por ser combinaciones de las primeras con

⁸Un Mercado Financiero se define en términos económicos como un mecanismo que permite a los agentes económicos el intercambio de dinero por valores o materias primas.

⁹Los subyacentes pueden ser bienes como el oro, productos como el petróleo, una acción individual, una canasta de acciones, un instrumento de deuda, indicadores como un índice bursátil o incluso el precio de otro instrumento financiero.

variaciones en los perfiles de rentabilidad o ser productos completamente diferentes con opcionalidad incluida, es decir que en su estructura se encuentra aparejada una opción. Este tipo de instrumento financiero se manejan por lo regular en los mercados de moneda extranjera, o sea que el activo subyacente es la divisa. Para mayores especificaciones de las opciones y los productos derivados en general se recomienda al lector revisar [16].

Con el objetivo de mostrar la utilidad del muestreo de importancia en el mundo de las opciones, trataremos con algunos ejemplos de opciones exóticas enfatizando la flexibilidad del método bajo un esquema de grandes desviaciones. Antes de comenzar a tratar el caso de las opciones exóticas, describamos algunos conceptos básicos de la teoría matemática del muestreo de importancia aplicado a las opciones financieras para así poder entender de manera clara la naturaleza del problema.

Iniciemos por explicar el cálculo de la *Razón de Verosimilitudes* bajo un cambio exponencial de medida inducido por el esquema de muestreo de importancia propuesto por las grandes desviaciones. En la valuación de opciones, es natural pensar que X representa una trayectoria discreta de un activo subyacente, de esta forma la densidad de la trayectoria (si es que existe alguna) no se encuentra especificada de manera directa, pero puede construirse a través de elementos un poco más primitivos. Consideremos por ejemplo una trayectoria discreta $S(t_i)$, $i = 0, 1, \dots, m$, de activos subyacentes o variables de estado evaluada al tiempo t_i , y supongamos que el proceso es Markoviano. Adicionalmente supongamos que la distribución condicional de $S(t_i)$ dado $S(t_{i-1}) = x$ tiene densidad $f_i(x, \cdot)$. Si consideramos un cambio de medida bajo el cual las densidades de transición f_i sean reemplazadas con densidades de transición g_i , entonces la razón de verosimilitudes para este cambio de medida es

$$\prod_{i=1}^m \frac{f_i(S(t_{i-1}), S(t_i))}{g_i(S(t_{i-1}), S(t_i))}.$$

De manera más precisa, si \mathbb{E} denota la esperanza bajo la medida original (f) y \mathbb{E}_g denota el operador esperanza bajo el cambio de medida (g), entonces

$$\mathbb{E}[h(S(t_1), \dots, S(t_m))] = \mathbb{E}_g \left[h(S(t_1), \dots, S(t_m)) \prod_{i=1}^m \frac{f_i(S(t_{i-1}), S(t_i))}{g_i(S(t_{i-1}), S(t_i))} \right], \quad (5.23)$$

para todas las funciones h , tales que la esperanza sobre la medida original exista y sea finita.

De forma implícita en el ejemplo anterior, se supone que $S(t_0)$ es constante. Ahora, de un modo más general podemos admitir tener densidad f_0 bajo la medida original y densidad g_0 bajo la nueva medida. Esto podría resultar en un factor adicional $f_0(S(t_0))/g_0(S(t_0))$ en la razón de verosimilitudes.

En la mayoría de las ocasiones la simulación de una trayectoria $S(t_0), \dots, S(t_m)$ se realiza utilizando un mecanismo recursivo de la forma

$$S(t_{i+1}) = G(S(t_i), X_{i+1}), \quad (5.24)$$

basándonos en vectores aleatorios i.i.d X_1, X_2, \dots, X_m . Si suponemos que las X_i tienen densidad común f para $i = 1, \dots, m$ y aplicamos un cambio de medida que preserve la independencia entre las X_i , pero cambie a una densidad común g , entonces la correspondiente razón de verosimilitudes es

$$\prod_{i=1}^m \frac{f(X_i)}{g(X_i)}.$$

Esto significa que

$$\mathbb{E}[h(S(t_1), \dots, S(t_m))] = \mathbb{E}_g \left[h(S(t_1), \dots, S(t_m)) \prod_{i=1}^m \frac{f(X_i)}{g(X_i)} \right]. \quad (5.25)$$

Claramente esta última ecuación recae en el hecho de que $S(t_1), \dots, S(t_m)$ son funciones de X_1, \dots, X_m .

§5.2.1. Horizonte Aleatorio.

El *Horizonte aleatorio* es un tiempo de paro proveniente de la extensión de un número fijo de pasos m a un número aleatorio de pasos. Si aplicamos esta idea a las ecuaciones (5.23) y (5.25) para demostrar que esto se cumple en el caso de entradas i.i.d., debemos suponer que $n = 1, 2, \dots$ y definir h_n como una función de n argumentos, para poder pensar ahora que lo que deseamos estimar es

$$\mathbb{E}[h_N(S(t_1), \dots, S(t_N))], \quad (5.26)$$

con N una variable aleatoria que toma valores en $\{1, 2, \dots\}$. Por ejemplo si pensamos en el caso de una opción con barrera y denominamos a esta barrera como

b , podemos definir a N como el índice más pequeño del tiempo t_i para el cual $S(t_i) > b$, tomando $N = m$ si todos los $S(t_0), \dots, S(t_m)$ caen debajo de la barrera. De esta forma podemos expresar el pago descontado de un put (opción de venta) “up-and-out” como $h_N(S(t_1), \dots, S(t_N))$ con

$$h_n(S(t_1), \dots, S(t_N)) = \begin{cases} e^{-rt_m}(K - S(t_m))^+ & \text{si } n = m. \\ 0 & \text{si } n = 0, \dots, m - 1. \end{cases}$$

Así el precio de la opción puede ser expresado como en (5.26). Supongamos ahora que la ecuación (5.24) se cumple y que además $S(t_0)$ es fijo bajo ambas medidas (f la original y g la de cambio). Sea N un tiempo de paro para la sucesión X_1, X_2, \dots ; por ejemplo, N podría ser el tiempo de paro para $S(t_1), S(t_2), \dots$ como en el ejemplo de la opción con barrera. Entonces

$$\mathbb{E}[h_N(S(t_1), \dots, S(t_N))\mathbb{I}_{\{N < \infty\}}] = \mathbb{E}_g \left[h_N(S(t_1), \dots, S(t_N)) \prod_{i=1}^N \frac{f(X_i)}{g(X_i)} \mathbb{I}_{\{N < \infty\}} \right],$$

con la condición de que la esperanza de la izquierda sea finita. Esta identidad (algunas veces llamada identidad de *Wald* o identidad fundamental del análisis secuencial) se establece de la siguiente forma

$$\begin{aligned} \mathbb{E}[h_N(S(t_1), \dots, S(t_N))\mathbb{I}_{\{N < \infty\}}] &= \sum_{n=1}^{\infty} \mathbb{E}[h_n(S(t_1), \dots, S(t_N))\mathbb{I}_{\{N=n\}}] \\ &= \sum_{n=1}^{\infty} \mathbb{E}_g \left[h_n(S(t_1), \dots, S(t_N))\mathbb{I}_{\{N=n\}} \prod_{i=1}^n \frac{f(X_i)}{g(X_i)} \right] \\ &= \mathbb{E}_g \left[h_N(S(t_1), \dots, S(t_N)) \prod_{i=1}^N \frac{f(X_i)}{g(X_i)} \mathbb{I}_{\{N < \infty\}} \right], \end{aligned} \tag{5.27}$$

en donde para la segunda igualdad se usa la propiedad de tiempo de paro. Como N es un tiempo de paro, el evento $\{N = n\}$ queda determinado por X_1, \dots, X_n , lo que nos permite aplicar (5.25) a cada término en la suma infinita. Debemos notar que es completamente posible que el evento $\{N < \infty\}$ tenga probabilidad igual a 1 bajo una de las medidas pero no bajo la otra.

§5.2.2. Horizonte Largo.

Continuemos considerando dos medidas de probabilidad bajo las cuales los vectores aleatorios X_1, X_2, \dots son i.i.d. Denotemos por \mathbb{P} a la medida que asigna a las X_i 's la densidad f y \mathbb{P}_g a la que asigna la densidad g a las variables aleatorias.

Debe notarse que aunque f y g sean mutuamente absolutamente continuas, las medidas de probabilidad \mathbb{P} y \mathbb{P}_g no necesariamente lo son. Sin embargo, la continuidad absoluta se mantiene para las restricciones de estas medidas a los eventos definidos por un segmento inicial finito de una secuencia infinita.

Si para $A \subseteq \mathbb{R}^d$, el evento

$$\left\{ \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{i=1}^m \mathbb{I}_{\{X_i \in A\}} = \int_A f(x) dx \right\},$$

tiene probabilidad igual a 1 bajo la medida \mathbb{P} , el evento debe tener probabilidad cero bajo \mathbb{P}_g a menos que f y g sean idénticas casi donde sea. La *Ley Fuerte de los Grandes Números* garantiza que \mathbb{P} y \mathbb{P}_g no concuerdan sobre aquellos eventos con probabilidad 0. Este colapso de continuidad absoluta en el límite, se refleja en una conducta un tanto patológica de la razón de verosimilitudes conforme el número de términos crece, argumento mostrado por *Glynn e Iglehart* en [12]. Supongamos que

$$\mathbb{E}_g \left[\left| \log \left(\frac{f(X_1)}{g(X_2)} \right) \right| \right] < \infty.$$

Entonces bajo la *Ley Fuerte de los Grandes Números* esto implica que

$$\frac{1}{m} \sum_{i=1}^m \log \left(\frac{f(X_i)}{g(X_i)} \right) \longrightarrow \mathbb{E}_g \left[\left| \log \left(\frac{f(X_1)}{g(X_1)} \right) \right| \right] \equiv C \quad \text{con } C \text{ constante,} \quad (5.28)$$

con probabilidad 1 bajo \mathbb{P}_g . Si aplicamos la desigualdad de *Jensen*, por la concavidad de la función logaritmo obtenemos que

$$C \leq \log \mathbb{E}_g \left[\frac{f(X_1)}{g(X_1)} \right] = \log \int \frac{f(x)}{g(x)} g(x) dx = 0,$$

en donde la desigualdad es estricta a menos que $\mathbb{P}_g\{f(X_1) = g(X_1)\} = 1$ (ya que el logaritmo es estrictamente cóncavo). Pero si $C < 0$, (5.28) implica que

$$\sum_{i=1}^m \log \left(\frac{f(X_i)}{g(X_i)} \right) \rightarrow -\infty,$$

exponencialmente. Por lo tanto

$$\sum_{i=1}^m \log \left(\frac{f(X_i)}{g(X_i)} \right) \rightarrow Cm \quad \Rightarrow \quad \prod_{i=1}^m \frac{f(X_i)}{g(X_i)} \rightarrow \exp\{Cm\},$$

que conforme m tiende a infinito, se acerca a cero con probabilidad $\mathbb{P}_g = 1$. De tal manera que la razón de verosimilitudes converge a 0 aunque su esperanza sea igual a 1 para todo m . Esto indica que la razón de verosimilitudes tiende a hacerse un poco “oblicua o sesgada”, tomando valores cada vez más grandes con una pequeña pero no despreciable probabilidad. Esto en cambio resulta en un incremento en la varianza si el cambio de medida no se elige cuidadosamente.

§5.2.3. Opción de Choque: *Knock in Option*.

Como un ejemplo sumamente ilustrativo del muestreo de importancia a través de un cambio exponencial de medida, trataremos con el caso de una “opción de barrera baja y adentro” (*down and in barrier option*¹⁰). El instrumento es una opción digital¹¹ de choque con función de pago

$$\mathbb{I}_{\{S(T) > K\}} \cdot \mathbb{I}_{\{\min_{1 \leq K \leq m} \{S(t_m) < H\}\}},$$

con $t_0 = 0 < t_1 < \dots < t_m = T$, donde T es el tiempo de maduración de la opción, es decir el tiempo que marca el vencimiento del contrato, $S(t_i)$ para $i = 0, \dots, m$ representa el valor del subyacente al tiempo t_i , K es el precio del contrato previamente convenido, mejor conocido como *Strike* y H representa la barrera especificada en el contrato.

De esta forma si H es mucho más pequeña que $S(0)$, es claro que en una simulación ordinaria una gran cantidad de trayectorias del subyacente resultarán en un pago igual a cero. Dadas estas condiciones, nos encontramos con un evento de probabilidad pequeña y como bien sabemos el muestreo de importancia se adapta perfectamente a este tipo de escenarios haciendo que los choques con la barrera

¹⁰Estas opciones son un call o put ordinarios con una barrera (H) que hace que la opción adquiera un valor fijo estipulado en el contrato de acuerdo a si el máximo o el mínimo de los valores del subyacente han topado o chocado con la barrera.

¹¹Son opciones parecidas al put y call tradicionales, pero tienen valor en el vencimiento cuando el subyacente es inferior o superior al precio de ejercicio, respectivamente.

sean menos raros.

Supongamos que el activo subyacente puede modelarse a través de un proceso de la forma

$$S(t_n) = S(0) \exp\{\xi_n\}, \quad \text{donde} \quad \xi_n = \sum_{i=1}^n X_i,$$

para X_1, X_2, \dots v.a.i.i.d y $\xi_0 = 0$. Por lo tanto la función de pago se convierte en

$$\mathbb{I}_{\{\xi_m > c, \tau < m\}},$$

donde $c = \ln(K/S(0))$ y τ es la primera vez que la caminata aleatoria ξ_n cae por debajo de $-b = \ln(H/S(0))$. Entonces si b o c son grandes, tenemos que la probabilidad de que se presente un pago distinto de cero es muy pequeña, de modo que para hacer crecer esta probabilidad de pago, debemos guiar a ξ_n hacia abajo de $-b$ y posteriormente hacerla subir hasta c .

Supongamos que X_i tiene función generadora acumulativa ψ . Al realizar un esquema de muestreo de importancia en donde primero apliquemos un cambio exponencial de medida que altere la distribución de X_i bajo un parámetro θ_- para el cual la deriva $\psi'(\theta_-)$ sea estrictamente negativa hasta el momento en el que la barrera sea cruzada. Posteriormente, cambiamos la distribución de las restantes $X_{\tau+1}, \dots, X_m$ bajo un parámetro θ_+ para el cual la deriva $\psi'(\theta_+)$ sea estrictamente positiva, guiando así al proceso hacia arriba hasta alcanzar el *strike*. Entonces el evento $\{\tau < m\}$ tiene razón de verosimilitudes

$$\begin{aligned} L_x &= \exp\{-\theta_- \xi_\tau + \psi(\theta_-)\tau\} \cdot \exp\{-\theta_+[\xi_m - \xi_\tau] + \psi(\theta_+)[m - \tau]\} \\ &= \exp\{(\theta_+ - \theta_-)\xi_\tau - \theta_+ \xi_m + (\psi(\theta_-) - \psi(\theta_+))\tau + m\psi(\theta_+)\}. \end{aligned}$$

Por lo tanto el estimador que propone el muestreo de importancia bajo este esquema es el producto de esta razón de verosimilitudes con la función de pago descontada. Apliquemos ahora un argumento heurístico para seleccionar los parámetros θ_+ y θ_- . Ya que para b y c grandes suponemos que $\xi_\tau \approx -b$ y $\xi_m \approx c$ en el evento $\{\tau < m, \xi_m > c\}$, es decir que el exceso por encima de c y el déficit por debajo de $-b$ deben ser pequeños, es natural suponer que la mayor parte de la variabilidad del estimador sea resultado del tiempo de cruce τ (tiempo de cruce de la barrera). Por lo tanto seleccionamos los parámetros θ_+ y θ_- de modo que se satisfaga que $\psi(\theta_-) = \psi(\theta_+)$, simplificándose así la razón de verosimilitudes de la siguiente forma

$$\exp\{(\theta_+ - \theta_-)\xi_\tau - \theta_+\xi_m + m\psi(\theta_+)\},$$

donde claramente podemos observar que cualquier tipo de dependencia sobre τ queda eliminada. Por último supondremos que la caminata aleatoria se mueve en línea recta de 0 a $-b$ con una tasa $|\psi'(\theta_-)|$ y que posteriormente sigue una trayectoria en línea recta de $-b$ a c con una tasa $\psi(\theta_+)$ y que además alcanza c en el tiempo m , es decir que:

$$\frac{-b}{\psi'(\theta_-)} + \frac{c+b}{\psi'(\theta_+)} = m.$$

Esta condición garantiza la unicidad en la elección de θ_\pm , al menos si el dominio de ψ es suficientemente grande.

Para el caso del Movimiento Browniano Geométrico $MBG(\mu, \sigma^2)$ con espacio de tiempo $t_n = nh$, tenemos que

$$X_i \sim N\left(\left(\mu - \frac{1}{2}\sigma^2\right)h, \sigma^2h\right),$$

que la función generadora de cumulantes es

$$\psi(\theta) = \left(\mu - \frac{1}{2}\sigma^2\right)h\theta + \frac{1}{2}\sigma^2h\theta^2.$$

Como es una función cuadrática para θ , entonces debe ser simétrica con respecto a su mínimo, por lo que la condición $\psi(\theta_-) = \psi(\theta_+)$ implica que $\psi(\theta_-) = -\psi(\theta_+)$ y dado el cambio de medida propuesto, tenemos que la caminata aleatoria se mueve a una velocidad constante igual a $|\psi'(\theta_\pm)|$. La trayectoria debe pasar por debajo de la barrera y subir hasta alcanzar el *strike* en m pasos, por lo tanto

$$\psi(\theta_-) = -\psi(\theta_+) \Rightarrow \frac{2b+c}{|\psi'(\theta_\pm)|} = m$$

\Rightarrow

$$|\psi'(\theta_\pm)| = \frac{2b+c}{m}.$$

Al resolver esta última ecuación encontramos que los parámetros de cambio son

$$\theta_{\pm} = \left(\frac{1}{2} - \frac{\mu}{\sigma^2}\right) \pm \frac{2b + c}{m\sigma^2 h}.$$

Es fácil ver que el punto en el cual ψ se minimiza es $(\frac{1}{2} - \frac{\mu}{\sigma^2})$, y que los parámetros de cambio son simétricos con respecto a este punto, cumpliéndose así todos nuestros supuestos.

§5.2.4. Opciones de Trayectoria Dependiente.

En esta sección trataremos la valuación opciones de *Trayectoria Dependiente*, considerando modelos para los activos subyacentes conducidos por el Movimiento Browniano (o simplemente por vectores aleatorios normales después de la discretización) y un cambio de deriva del Movimiento Browniano que guíe al subyacente hacia las regiones de “importancia”, donde la “importancia” será determinada por el pago de la opción. Además identificaremos un específico cambio de medida a través del planteamiento de un problema de optimización.

El método descrito a continuación contiene únicamente aspectos fundamentales del tratamiento del problema, pero que ayudarán a comprender la idea básica de esta aplicación. Si el lector de este trabajo desea indagar más sobre el desarrollo teórico del problema, se recomienda revisar [10].

Este método se restringe a cambios determinísticos de deriva en tiempos discretos. Es teóricamente posible eliminar toda la varianza de la estimación a través de un cambio estocástico de deriva en tiempo continuo, sin embargo esto requiere del conocimiento previo del precio de la opción, lo que es literalmente imposible. Esto se debe a que esencialmente la reducción de varianza se logra al tomar la opción, valuarla como un activo numerado y aplicar un cambio de medida asociado al cambio de la numeración, además se pueden producir sesgos para las aproximaciones.

Para simplificar las cosas, nos restringiremos al tratamiento del problema en tiempo discreto, considerando la partición $0 = t_0 < t_1 < \dots < t_m = T$. Supondremos que la única fuente de aleatoriedad involucrada en el modelo simulado es un Movimiento Browniano de dimensión d . Los incrementos del Movimiento Browniano de t_{i-1} a t_i son simulados como $\sqrt{t_i - t_{i-1}}Z_i$, donde Z_1, \dots, Z_m son vectores d -dimensionales con distribución normal estándar. Si denotamos por Z a la concatenación de las Z_i s en un solo vector de longitud $n \equiv md$, entonces cada simulación de Z determinará una trayectoria de los activos subyacentes o variables de estado, mientras que cada trayectoria determinará el pago descontado de

una opción. Sea G la composición de estos mapeos, por lo tanto $G(Z)$ es el pago descontado derivado de Z . Nuestra tarea es estimar $\mathbb{E}[G(Z)]$, es decir la esperanza tomada para Z con distribución Normal estándar de dimensión n .

Para aclarar estas ideas, consideremos un solo activo subyacente modelado como un $MBG(r, \sigma^2)$ y simulémoslo de la siguiente manera

$$S(t_i) = S(t_{i-1}) \exp\left\{\left(r - \frac{1}{2}\sigma^2\right)(t_i - t_{i-1}) + \sigma\sqrt{t_i - t_{i-1}}Z_i\right\}, \quad i = 1, \dots, m. \quad (5.29)$$

De este modo si tomamos un call asiático¹² con tiempo de maduración T , tasa de rendimiento r y strike K , entonces podemos observar que para el promedio aritmético \bar{S} de los valores $S(t_i)$, el pago de la opción como función de Z_i se puede escribir de la siguiente manera

$$G(Z) = G(Z_1, \dots, Z_m) = e^{-rT}[\bar{S} - K]^+.$$

Por lo tanto valuar la opción significa calcular la $\mathbb{E}[G(Z)]$, para $Z \sim N(\underline{0}, \mathbb{I})$.

§5.2.5. Cambio de Deriva: Linealización.

A través del muestreo de importancia podemos cambiar la distribución de Z y obtener un estimador insesgado de $\mathbb{E}[G(Z)]$, donde claramente proporcionamos ciertos pesos a cada evento de acuerdo a su importancia según la razón de verosimilitudes. Si nos restringimos únicamente a cambios de medida que cambien la media de Z de cero a algún otro vector $\underline{\mu}$, e identificamos a la nueva medida de probabilidad como $\mathbb{P}_{\underline{\mu}}$ y a su respectivo operador esperanza como $\mathbb{E}_{\underline{\mu}}$, entonces tenemos que $Z \sim N(\underline{\mu}, \mathbb{I})$. Bajo las condiciones anteriormente expuestas tenemos que

$$\mathbb{E}[G(Z)] = \mathbb{E}_{\underline{\mu}}[G(Z) \exp\{-\underline{\mu}^T Z + \frac{1}{2}\underline{\mu}^T \underline{\mu}\}],$$

para cualquier vector $\underline{\mu} \in \mathbb{R}^n$. Un algoritmo fácil de simular es

¹²Se denominan opciones asiáticas a aquellas cuyo valor depende del promedio de los valores que ha tenido el subyacente durante la vida (o parte de ella) de la opción.

Para repeticiones $i = 1, \dots, N$
 Generar $Z^i \sim N(\underline{\mu}, I)$
 Asignar $Y^i \leftarrow G(Z^i) \exp \left\{ -\underline{\mu}^T Z^i + \frac{1}{2} \underline{\mu}^T \underline{\mu} \right\}$
 Regresar $\frac{Y^1 + \dots + Y^N}{N}$.

Claramente este estimador es insesgado para cualquier valor de $\underline{\mu}$, no obstante, estamos interesados en la elección del parámetro $\underline{\mu}$ que proporcione el estimador de menor varianza. Con este objetivo en mente, tomemos las funciones G con dominio no negativo (como es típico para las funciones de pago de opciones descontadas) y escribamos $G(Z) = \exp \{F(Z)\}$, con la convención de que $F(Z) = -\infty$ si $G(Z) = 0$. Cabe resaltar que tomar la esperanza de Z bajo $\mathbb{P}_{\underline{\mu}}$ es equivalente a reemplazar Z por $\underline{\mu} + Z$ y tomar la esperanza bajo la medida original. En el algoritmo expuesto previamente, esto simplemente significa que podemos obtener muestras de la distribución $N(\underline{\mu}, \mathbb{I})$ muestreando a partir de una $N(\underline{0}, \mathbb{I})$ y sumar posteriormente el parámetro $\underline{\mu}$, de modo que

$$\begin{aligned} \mathbb{E}[G(Z)] &= \mathbb{E}[G(\underline{\mu} + Z)] = \mathbb{E}_{\underline{\mu}}[\exp \{F(\underline{\mu} + Z)\} \exp \left\{ -\underline{\mu}^T Z + \frac{1}{2} \underline{\mu}^T \underline{\mu} \right\}] \\ &= \mathbb{E}[\exp \{F(\underline{\mu} + Z)\} \exp \left\{ -\underline{\mu}^T (\underline{\mu} + Z) + \frac{1}{2} \underline{\mu}^T \underline{\mu} \right\}] \quad (5.30) \\ &= \mathbb{E}[\exp \{F(\underline{\mu} + Z)\} \exp \left\{ -\underline{\mu}^T Z - \frac{1}{2} \underline{\mu}^T \underline{\mu} \right\}]. \end{aligned}$$

Además para todo vector $\underline{\mu}$ la expresión que se encuentra dentro del operador esperanza en (5.30) es un estimador insesgado con $Z \sim N(\underline{0}, \mathbb{I})$.

Para ver una elección particular de $\underline{\mu}$, desarrollemos a F en su expansión de Taylor de grado 1, entonces podemos aproximar al estimador de la siguiente forma

$$\exp \{F(\underline{\mu} + Z)\} \exp \left\{ -\underline{\mu}^T Z - \frac{1}{2} \underline{\mu}^T \underline{\mu} \right\} \approx \exp \{F(\underline{\mu}) + \nabla F(\underline{\mu}) Z\} \exp \left\{ -\underline{\mu}^T Z - \frac{1}{2} \underline{\mu}^T \underline{\mu} \right\}, \quad (5.31)$$

con $\nabla F(\underline{\mu})$ el gradiente de F evaluado en $\underline{\mu}$. Si recordamos un poco lo realizado en el caso del VaR, entonces debemos poder elegir $\underline{\mu}$ de tal forma que se satisfaga la siguiente condición de punto fijo

$$\nabla F(\underline{\mu}) = \underline{\mu}^T. \quad (5.32)$$

De este modo la expresión del lado derecho de (5.31) colapsará hacia una constante que no depende en absoluto de Z . Por lo tanto si aplicamos un esquema de muestreo de importancia para el cual se satisfaga (5.32) debemos lograr producir un estimador de varianza cero, si y sólo si (5.31) se cumple estrictamente, es decir en igualdad, mientras que para el caso en que sólo se cumpla aproximadamente lograremos producir un estimador de varianza muy pequeña.

§5.2.6. Cambio de Deriva: Aproximación Normal y Trayectoria Óptima.

A continuación presentaremos un argumento alternativo al de la sección anterior, que conduce esencialmente a una elección equivalente del parámetro $\underline{\mu}$. Recordemos que para estimar $\mathbb{E}[h(X)]$ con $h : \mathbb{R}^d \rightarrow \mathbb{R}$, el muestreo de importancia propone un estimador de la forma

$$\frac{1}{n} \sum_{i=1}^n h(X_i) \frac{f(X_i)}{g(X_i)},$$

donde f representa la densidad del problema original, g es la densidad de cambio y X_1, \dots, X_n son variables aleatorias simuladas bajo la densidad g .

Entonces si pensamos solamente en las funciones h que son no-negativas, esto implicaría que el producto $h(x)f(x)$ es también no-negativo y además es claro que puede ser normalizado para formar una densidad. Supongamos que esta densidad es g , por lo tanto

$$g(x) \propto h(x)f(x), \quad (5.33)$$

donde si no consideramos el valor de X_i obtendremos que la constante de normalización es igual a

$$h(X_i) \frac{f(X_i)}{g(X_i)}.$$

En este caso el estimador propuesto por el muestreo de importancia es un estimador de varianza cero, sin embargo como ya se ha expuesto con anterioridad en este documento, este estimador es inútil en la práctica pues para lograr normalizar el producto $h \cdot f$ necesitamos dividir el mismo entre el valor de su integral, es

decir la $\mathbb{E}[h]$.

Con esto en mente, sabemos que la densidad óptima del muestreo de importancia es el producto normalizado del integrando y la densidad original, de modo que al adaptar estas ideas a nuestro problema de estimación para la valuación de opciones de trayectoria dependiente podemos decir que la densidad óptima es proporcional a

$$\exp\left\{F(z) - \frac{1}{2}z^T z\right\},$$

pues el integrando es igual a $\exp\{F(z)\}$ y $\exp\{-z^T z/2\}$ es proporcional a la densidad normal estándar. Al normalizar esta función y dividirla entre su integral producimos una nueva densidad que en general no pertenece a la familia de densidades normales.

Ya que hemos restringido nuestros cálculos a cambios de medida que sólo alteren la media de la distribución, podemos tratar de seleccionar $\underline{\mu}$ de tal forma que la densidad $N(\underline{\mu}, \mathbb{I})$ se aproxime a la distribución óptima. Una manera de hacer esto es escoger $\underline{\mu}$ igual a la moda de la distribución, es decir elegir el valor $\underline{\mu}$ que resuelve

$$\underset{z}{\text{máx}}\left\{F(z) - \frac{1}{2}z^T z\right\}. \quad (5.34)$$

Dado este problema de maximización, observamos que la condición de primer orden para el óptimo es $\nabla F(z) = z^T$, que coincide con (5.32). Si por ejemplo tenemos que la función objetivo en (5.34) es estrictamente cóncava, entonces si la condición de primer orden tiene solución, podemos asegurar que esta es la única solución existente y por lo tanto el único punto óptimo.

Podemos interpretar a la solución de (5.34) como una trayectoria óptima, ya que para cada $z \in \mathbb{R}^n$ se determina una trayectoria discreta del Movimiento Browniano y esta a su vez describe una trayectoria del activo subyacente. Si estamos midiendo la importancia via el producto de la función de pago $\exp\{F(Z)\}$ y la densidad $\exp\{-z^T z/2\}/(2\pi)^{n/2}$, se dice que la solución de (5.34) es la trayectoria más “importante” del subyacente. Por último, sólo nos falta resaltar el hecho de que al hacer la elección de este punto $\underline{\mu}$, estamos eligiendo una nueva deriva para el Movimiento Browniano que guía al proceso a lo largo de la trayectoria óptima.

§5.2.7. Opciones Asiáticas.

En la sección anterior hemos mostrado la selección y aplicación del cambio de deriva óptimo para el caso de un call asiático, dándonos cuenta de que resolver (5.34) es equivalente a maximizar $G(z) \exp\{-z^T z/2\}$, con G la función de pago de una opción asiática. En este ejemplo podemos ver claramente que el factor de descuento $\exp\{-rT\}$ es constante, sin embargo para propósitos de optimización podemos ignorarlo y redefinir $G(z)$ como $[\bar{S} - K]^+$. Además es fácil ver que para obtener este máximo, es suficiente con considerar los puntos z para los cuales $\bar{S} > K$ ya que son justamente los puntos donde G es diferenciable.

Entonces para la condición de primer orden debemos diferenciar

$$[\bar{S} - K] \exp\left\{-\frac{z^T z}{2}\right\},$$

para posteriormente resolver

$$\frac{\partial \bar{S}}{\partial z_j} - [\bar{S} - K] z_j = 0.$$

Al utilizar (5.29), encontramos que

$$\frac{\partial \bar{S}}{\partial z_j} = \frac{1}{m} \sum_{i=j}^m \frac{\partial S(t_i)}{\partial z_j} = \frac{1}{m} \sum_{i=j}^m \sigma \sqrt{t_i - t_{i-1}} S(t_i),$$

transformando nuestra condición de primer orden en

$$z_j = \frac{\sum_{i=j}^m \sigma \sqrt{t_i - t_{i-1}} S(t_i)}{mG(z)}.$$

Si consideramos ahora el caso particular donde el espaciado de tiempo es uniforme con $t_i - t_{i-1} \equiv h$, obtenemos la siguiente regla recursiva

$$z_1 = \frac{\sigma \sqrt{h}(G(z) + K)}{G(z)}, \quad z_{j+1} = z_j - \frac{\sigma \sqrt{h} S(t_j)}{mG(z)} \quad j = 1, \dots, m-1. \quad (5.35)$$

De donde dado el valor de $G(z)$, (5.35) y (5.29) determinan el valor de z . De hecho si $y \equiv G(z)$, podemos aplicar (5.35) para calcular z_j para $j = 1, \dots, m-1$.

A través de esta iteración, cada valor de y determina un valor de $z(y)$ y una trayectoria $S(t_i, y)$, con $i = 1, \dots, m$. Al resolver la condición de primer orden, nos reducimos a encontrar la y para la cual la función de pago en $S(t_1, y), \dots, S(t_m, y)$ es de hecho y , es decir que debemos encontrar la raíz de la ecuación

$$\frac{1}{m} \sum_{j=1}^m S(t_j, y) - K - y = 0.$$

§5.2.8. Muestreo de Importancia más Estratificación.

Como ya hemos visto la reducción de varianza puede ser difícil de obtener en algunos casos prácticos dadas las condiciones del problema, sin embargo sabemos también que el muestreo de importancia es una técnica sumamente poderosa y flexible para poder alterarse y combinarse con otros métodos como el muestreo estratificado (como vimos en el caso del VaR), y así lograr mejores resultados. Si realizamos una estratificación sobre una proyección lineal de Z , tendremos que $v^T Z$ está estratificado para algún $v \in \mathbb{R}^n$, y además es fácil de implementar (ver sección 4.3.2 de [9]).

El cambio de medida no afecta en absoluto esta condición sobre la muestra, ya que de hecho podemos obtener muestras de una distribución $N(\underline{\mu}, \mathbb{I})$ a partir de una $N(\underline{0}, \mathbb{I})$ y posteriormente añadir el parámetro $\underline{\mu}$, de modo que podemos aplicar primero un Muestreo Estratificado sobre la distribución $N(\underline{0}, \mathbb{I})$ y posteriormente sumar $\underline{\mu}$.

En [10] se consideran dos tipos de estrategias para seleccionar la dirección v de la estratificación. La primera es fijar $v = \underline{\mu}$ en los grupos donde $\underline{\mu}$ describe una trayectoria importante, y por ende una potencial dirección para la estratificación. La segunda estrategia es expandir (5.31) de la siguiente manera

$$\begin{aligned} \exp\{F(\underline{\mu} + Z)\} \exp\{-\underline{\mu}^T Z - \frac{1}{2}\underline{\mu}^T \underline{\mu}\} &\approx \exp\{F(\underline{\mu}) + \nabla F(\underline{\mu})Z + \frac{1}{2}Z^T H(\underline{\mu})Z\} \\ &\cdot \exp\{-\underline{\mu}^T Z - \frac{1}{2}\underline{\mu}^T \underline{\mu}\}, \end{aligned}$$

con $H(\underline{\mu})$ como el Hessiano de la matriz que proporciona F evaluado en $\underline{\mu}$. Bajo un esquema de muestreo de importancia en donde $\underline{\mu}^T = \nabla F(\underline{\mu})$, se elimina el término lineal en el exponente indicándonos que la estratificación debe aplicarse al término cuadrático, bajo algún criterio de optimización (sección 4.3.2 de [9]).

§5.3. El Problema de la Ruina.

Una aplicación clásica del muestreo de importancia nace a partir de la estimación de probabilidades de ruina en la *Teoría de Riesgo* de los Seguros. Si el lector

de este trabajo desea profundizar más en los aspectos teóricos de la *Teoría de la Ruina*, se recomienda ampliamente el libro de S. Asmussen, *Ruin Probabilities* [2], en el cual se basa la exposición de esta sección.

§5.3.1. El Proceso de Riesgo.

Un *proceso de riesgo de reservas* $\{R_t\}_{t \geq 0}$, es un modelo que describe la evolución de las reservas de una compañía de seguros a través del tiempo. Denotaremos a la reserva inicial como $u = R_0$, de esta manera si la compañía de seguros se aruina, implica que sus reservas han caído por debajo del cero. La probabilidad de ruina en un horizonte de tiempo infinito es

$$\Psi(u) = \mathbb{P}\{\inf_{t \geq 0} R_t < 0\} = \mathbb{P}\{\inf_{t \geq 0} R_t < 0 | R_0 = u\}, \quad (5.36)$$

mientras que la probabilidad de ruina para un horizonte de tiempo finito, es decir, antes de un tiempo T fijo, se define como

$$\Psi(u, T) = \mathbb{P}\{\inf_{0 \leq t \leq T} R_t < 0\}. \quad (5.37)$$

A veces, es más conveniente trabajar con el *proceso del excedente de las reclamaciones* $\{S_t\}_{t \geq 0}$ definido como $S_t = u - R_t$, para el cual

$$\tau(u) = \inf\{t \geq 0 : R_t < 0\} = \inf\{t \geq 0 : S_t > u\} \quad (5.38)$$

$$M = \sup_{0 \leq t < \infty} S_t, \quad \text{y} \quad M_T = \sup_{0 \leq t \leq T} S_t, \quad (5.39)$$

representan el tiempo de ruina y el valor máximo del proceso para un horizonte de tiempo infinito y uno finito, respectivamente. De manera alternativa las probabilidades de ruina pueden escribirse como

$$\Psi(u) = \mathbb{P}\{\tau(u) < \infty\} = \mathbb{P}\{M > u\} \quad (5.40)$$

$$\Psi(u, T) = \mathbb{P}\{M_T > u\} = \mathbb{P}\{\tau(u) \leq T\}. \quad (5.41)$$

Para que el proceso de riesgo de las reservas describa apropiadamente la realidad, necesitamos que se cumplan las siguientes condiciones:

- Sólo puede haber un número finito de reclamaciones en intervalos de tiempo finitos, esto significa que si N_t denota el número de reclamaciones en el intervalo $[0, t]$, entonces $N_t < \infty$. Denotaremos al tiempo entre reclamaciones como T_i para $i = 1, 2, \dots$, donde T_1 representa el tiempo de llegada de la primera reclamación. Por lo tanto, el tiempo de llegada de la n -ésima reclamación es $\sigma_n = T_1 + \dots + T_n$ y $N_t = \min\{n \geq 0 : \sigma_{n+1} > t\} = \max\{n \geq 0 : \sigma_n \leq t\}$.
- El tamaño de la n -ésima reclamación lo denotaremos por U_n .
- El flujo de primas se presenta a una tasa constante p por unidad de tiempo.

De este modo al juntar todos nuestros supuestos tenemos que

$$R_t = u + pt - \sum_{k=1}^{N_t} U_k, \quad S_t = \sum_{k=1}^{N_t} U_k - pt, \quad (5.42)$$

donde las variables U_k son v.a.i.i.d, mayores o iguales que cero, que describen la distribución de las reclamaciones y son independientes del proceso N_t .

Dadas las condiciones del problema, el *proceso Poisson Compuesto* es uno de los modelos más utilizados para describir el fenómeno denominado como ruina, sin embargo, procesos más generales como los procesos de *Lévy* pueden ser utilizados para modelar este problema.

A continuación introduciremos algunos resultados y propiedades de gran relevancia para el cálculo de probabilidades de ruina modeladas a través de un proceso Poisson Compuesto con intensidad β y de igual forma realizaremos el cálculo de las herramientas básicas del muestreo de importancia.

Una de las propiedades más importantes del proceso Poisson es que existe una constante ρ tal que

$$\frac{1}{t} \sum_{k=1}^{N_t} U_k \xrightarrow{c.s.} \rho, \quad t \rightarrow \infty. \quad (5.43)$$

La demostración de este hecho se sigue de inmediato de las propiedades del proceso Poisson de la siguiente manera

$$\lim_{t \rightarrow \infty} \mathbb{E}\left[\frac{1}{t} \sum_{k=1}^{N_t} U_k\right] = \lim_{t \rightarrow \infty} \mathbb{E}\left[\mathbb{E}\left[\frac{1}{t} \sum_{k=1}^{N_t} U_k \mid N_t = n\right]\right] = \lim_{t \rightarrow \infty} \frac{1}{t} \beta t \mathbb{E}[U_1] = \beta \mathbb{E}[U_1] = \rho.$$

A la constante ρ se le interpreta como el monto promedio de las reclamaciones por unidad de tiempo. Una cantidad de mayor relevancia, es la llamada “carga de

seguridad" η (*safety loading condition*), definida como el monto relativo por el cual la tasa del flujo de primas excede el monto promedio de las reclamaciones, donde

$$\eta = \frac{p - \rho}{\rho}. \quad (5.44)$$

Esta cantidad es de gran importancia ya que gracias a ella podemos determinar fácilmente cuándo ocurre la ruina. De hecho las compañías aseguradoras siempre buscarán mantener $\eta > 0$, para hacer nula la posibilidad de una eventual ruina. Para establecer de manera clara lo anterior tenemos las siguientes dos proposiciones

Proposición 6. *Para el proceso del excedente de las reclamaciones $\{S_t\}_{t \geq 0}$, se cumple:*

- a) *Sin importar el valor de η , S_t/t converge casi seguramente a $\rho - p$ si t tiende a infinito.*
- b) *Si $\eta < 0$, entonces $S_t \xrightarrow{c.s.} \infty$.*
- c) *Si $\eta > 0$, entonces $S_t \xrightarrow{c.s.} -\infty$.*
- d) *Si $\eta = 0$, entonces el $\liminf_{t \rightarrow \infty} S_t = -\infty$ y $\limsup_{t \rightarrow \infty} S_t = \infty$.*

Para probar la proposición anterior necesitamos establecer el siguiente *Lema*:

Lema 4. *Si $nh \leq t \leq (n+1)h$, entonces*

$$S_{nh} - h \leq S_t \leq S_{(n+1)h} + h.$$

Demostración

Primero debemos notar que para $u, v \geq 0$, se tiene que $S_{u+v} \geq S_u - v$, de hecho $S_{u+v} - S_u$ alcanza su mínimo cuando no se presentan reclamaciones en el intervalo $(u, u+v)$ y el valor que toma es precisamente v . En particular, si $t = nh + v$ con $0 \leq v \leq h$ entonces

$$S_t \geq S_{nh} - v \geq S_{nh} - h.$$

Análogamente,

$$S_{nh+h} \geq S_{nh} - h.$$

Con lo que probamos ambos lados de la desigualdad mostrada en el *Lema 4*.

◇

A continuación probaremos la *Proposición 2*.

Demostración

Para un h fijo, tenemos que la sucesión $\{S_{nh}\}_{n \geq 0}$ es una caminata aleatoria en tiempo discreto y por lo tanto al aplicar la *Ley Fuerte de los Grandes Números* podemos asegurar que como se cumple (5.43) entonces

$$\frac{S_{nh}}{n} = \frac{\sum_{k=1}^{N_{nh}} U_k - pn h}{n} \xrightarrow{c.s.} \mathbb{E}[S_h] = h(\rho - p), \quad n \rightarrow \infty.$$

De modo que al aplicar el resultado del *Lema* anterior, concluimos que

$$\liminf_{t \rightarrow \infty} \frac{S_t}{t} = \liminf_{n \rightarrow \infty} \inf_{nh \leq t \leq (n+1)h} \frac{S_t}{t} \geq \frac{1}{h} \liminf_{n \rightarrow \infty} \frac{S_{nh} - h}{n} = \frac{1}{h} \mathbb{E}[S_h] = \rho - p.$$

De manera análoga

$$\limsup_{t \rightarrow \infty} \frac{S_t}{t} = \limsup_{n \rightarrow \infty} \sup_{nh \leq t \leq (n+1)h} \frac{S_t}{t} \leq \frac{1}{h} \limsup_{n \rightarrow \infty} \frac{S_{(n+1)h} + h}{n} = \frac{1}{h} \mathbb{E}[S_h] = \rho - p.$$

Con lo que probamos el resultado expuesto en *a)*. Los incisos *b)* y *c)* son consecuencia inmediata de *a)* y la definición de η . Para *d)*, debemos aplicar un resultado mucho más general de las caminatas aleatorias expuesto en [1] (pp. 169), que nos dice que el $\liminf_{n \rightarrow \infty} S_{nh} = -\infty$ y que el $\limsup_{n \rightarrow \infty} S_{nh} = \infty$ concluyendo así la prueba de la proposición.

◇

Para observar la relación que existe entre la probabilidad de ruina y la llamada carga de seguridad η , establezcamos la siguiente proposición:

Proposición 7. *Supongamos que (5.43) se cumple, entonces si $\eta \leq 0$ implica que $M = \infty$ casi seguramente, y por lo tanto $\Psi(u) = 1$ para toda u , mientras que si $\eta > 0$ entonces $M < \infty$ casi seguramente y $\Psi(u) < 1$ para toda u suficientemente grande.*

Demostración

Como se cumple (5.43) tenemos que

$$\frac{S_t}{t} = \frac{\sum_{k=1}^{N_t} U_k - pt}{t} \xrightarrow{c.s.} \rho - p, \quad t \rightarrow \infty,$$

de modo que si $\eta \leq 0$, entonces $\rho - p > 0$ lo que implica que $S_t \xrightarrow{c.s.} \infty$ y $M = \sup_{0 \leq t < \infty} \{S_t\} = \infty$, $\Psi(u) = \mathbb{P}\{\tau(u) < \infty\} = \mathbb{P}\{M > u\} = 1$ para toda u . Por otra parte si $\eta > 0$ implica que $S_t/t < 0$ y entonces S_t converge casi seguramente a $-\infty$ y con esto $M < \infty$ casi seguramente, lo que nos lleva a concluir que $\Psi(u) < 1$.

◇

§5.3.2. Muestreo de Importancia Via Conjugación de Lundberg.

En las secciones anteriores hemos desarrollado las herramientas necesarias para la aplicación del muestreo de importancia al problema de la ruina. Únicamente nos falta enunciar un resultado de la teoría de riesgo, conocido como la *Aproximación de Cramér-Lundberg*, que nos dice que para un proceso Poisson Compuesto con intensidad β , tasa del flujo de primas $p = 1$ y distribución del tamaño de las reclamaciones igual a B con función generadora de momentos o transformada de Laplace $M_B(\theta)$, se tiene que

$$\Psi(u) \sim C e^{-\gamma u}, \quad u \rightarrow \infty, \quad (5.45)$$

donde $C = (1 - \rho)/(\beta M_B(\gamma) - 1)$ representa la constante de *Cramér-Lundberg*, ρ está definida como en (5.43) y $\gamma > 0$ igual a la solución de la ecuación de *Lundberg*

$$\beta(M_B(\gamma) - 1) - \gamma = 0, \quad (5.46)$$

la cual puede escribirse de manera equivalente como

$$M_B(\gamma) = 1 + \frac{\gamma}{\beta}. \quad (5.47)$$

A la constante γ se le conoce como el *coeficiente de ajuste*.

Para aplicar el **IS** debemos considerar que

$$\Psi(u) = \mathbb{P}\{\tau(u) < \infty\} = e^{-\gamma u} \tilde{\mathbb{E}}[e^{-\gamma \xi(u)}],$$

donde ξ representa el excedente de primas y que nuestras simulaciones deberán obtenerse a partir de la nueva medida de probabilidad $\tilde{\mathbb{P}}$, lo que implica utilizar $\tilde{\beta}$ y \tilde{B} como nuevos parámetros en lugar de los originales β y B , con el propósito de almacenar información de $Z(u) = e^{-\gamma S_{\tau(u)}}$, el estimador propuesto por el muestreo de importancia.

Para efectos prácticos, el proceso continuo $\{S_t\}$ se simula considerando tiempos discretos $\{\sigma_k\}$, que correspondan a la llegada de cada una de las reclamaciones. Con lo anterior podemos proponer el siguiente algoritmo para generar simulaciones del estimador $Z = Z(u)$:

1. Calcular el coeficiente de ajuste $\gamma > 0$ como solución de la ecuación de Lundberg

$$0 = \kappa(\gamma) = \beta(M_B(\gamma) - 1) - \gamma,$$

y posteriormente definir a $\tilde{\beta}$ y \tilde{B} como

$$\tilde{\beta} = \beta M_B(\gamma) \quad y \quad \tilde{B}(dx) = e^{\gamma x} \frac{B(dx)}{M_B(\gamma)}.$$

2. Iniciar S en cero.
3. Generar T como una variable aleatoria exponencial de parámetro $\tilde{\beta}$ y U a partir de \tilde{B} . Posteriormente asignar $S \leftarrow S + U - T$.
4. Si $S > u$, debemos asignar $Z \leftarrow e^{\gamma S}$, de otra forma regresamos al paso 3 y repetimos el procedimiento.

Existen varias razones intuitivas para creer que el algoritmo anterior es bueno, en primera instancia, como $\tilde{\mathbb{P}}\{\tau(u) < \infty\} = 1$, entonces se resuelve el problema para un horizonte infinito, además de que dado el conocimiento sobre la forma de $\Psi(u)$, podemos esperar obtener poca varianza en nuestras estimaciones al aislar la información desconocida, $\tilde{\mathbb{E}}[e^{-\gamma \xi(u)}]$ y así evitar simular la parte conocida $e^{-\gamma u}$. Si deseamos encontrar una trayectoria muestral del proceso de riesgo que se aproxime a la ruina, debemos considerar que $\tilde{\mathbb{P}}$ y $\mathbb{P}\{\cdot | \tau(u) < \infty\}$ (ambas medidas restringidas a $\mathcal{F}_{\tau(u)}$) coincidan asintóticamente en $\{\tau(u) < \infty\}$, lo que nos indica que la medida de cambio $\tilde{\mathbb{P}}$ está cerca de ser la medida óptima de cambio.

A continuación analizaremos la eficiencia del estimador propuesto por el muestreo de importancia.

Teorema 4. *El estimador $Z(u) = e^{-\gamma S_{\tau(u)}}$ simulado bajo $\tilde{\mathbb{P}}$ tiene error relativo acotado.*

Demostración

$$\tilde{\mathbb{E}}[Z(u)^2] = \tilde{\mathbb{E}}[e^{-2\gamma S_{\tau(u)}}] \leq e^{-2\gamma u} \sim \frac{\Psi(u)^2}{C^2} \sim e^{-2\gamma u} < \infty, \quad \forall u.$$

◇

En este punto, es natural preguntarnos si existen parámetros distintos a los propuestos por el muestreo de importancia bajo un esquema de grandes desviaciones, que mejoren la varianza de la simulación. La respuesta a este cuestionamiento es **no**. Si tenemos β_1 y B_1 distintos de $\tilde{\beta}$ y \tilde{B} (los parámetros propuestos bajo el **IS**), y pensamos en el problema con horizonte infinito, entonces podemos restringir nuestra atención al caso en que $\beta_1 \mu_{B_1} \geq 1$ (con μ_{B_1} la media de la distribución B_1), ya que de esta forma aseguramos que la carga de seguridad η sea negativa y por consecuencia la probabilidad de una eventual ruina es igual a 1. Entonces el estimador propuesto por el muestreo de importancia tiene la siguiente forma

$$Z(u) = \prod_{i=1}^{m(u)} \frac{\beta e^{-\beta T_i}}{\beta_1 e^{-\beta_1 T_i}} \frac{dB}{dB_1}(U_i), \quad (5.48)$$

donde $m(u)$ representa el número de reclamaciones que llevan a la compañía de seguros a la ruina y recordamos que $T_i = \sigma_i - \sigma_{i-1}$ es el intervalo de tiempo de la reclamación i .

Teorema 5. *El estimador (5.48) simulado con los parámetros β_1 y B_1 no es logarítmicamente eficiente cuando $(\beta_1, B_1) \neq (\tilde{\beta}, \tilde{B})$.*

Demostración

La demostración de este *Teorema* es *Corolario* del *Teorema* 6.

◇

Cabe resaltar que podemos generalizar de manera sencilla este algoritmo a un modelo de renovación, de la siguiente manera: Sean X_1, X_2, \dots v.a.i.i.d, con distribución F y supongamos que esta distribución tiene media $\mu_F < 0$. Tomemos a S_n y definamos a $m(u) = \inf\{n : S_n > u\}$. Supongamos también que la función generadora de momentos de F , $M_F(\gamma) = 1$ y que $M'_F(\gamma) < \infty$ para alguna constante positiva γ . Sabemos que bajo el muestreo de importancia la distribución de cambio \tilde{F} , se obtiene al hacer $\tilde{F}(dx) = e^{\gamma x} F(dx)$, de modo que el estimador propuesto por el **IS** es $Z(u) = e^{-\gamma S_{M(u)}}$, es decir que la distribución de cambio \tilde{F} debe ser equivalente a F y

$$Z(u) = \prod_{i=1}^{m(u)} \frac{dF}{d\tilde{F}}(X_i). \quad (5.49)$$

Teorema 6. *El estimador (5.49), simulado bajo la distribución F_1 de X_i tiene error relativo acotado cuando $F_1 = \tilde{F}$. Cuando $F_1 \neq \tilde{F}$ este estimador **no** es logarítmicamente eficiente.¹⁵*

Demostración

El hecho de que (5.49) tenga *error relativo acotado* se prueba exactamente igual que el *Teorema 5*, notando que $F_1(dx) = \tilde{F}(dx) = e^{\gamma x} F(dx)$.

Para probar la segunda parte del *Teorema*, utilicemos la regla de la cadena para las derivadas de Radon-Nikodym y escribamos

$$W(F|F_1) = \frac{dF}{dF_1}(X_1) \cdots \frac{dF}{dF_1}(X_{m(u)}),$$

de modo que

$$\begin{aligned} \mathbb{E}_{F_1}[Z^2(u)] &= \mathbb{E}_{F_1}[W^2(F|F_1)] = \mathbb{E}_{F_1}[W^2(F|\tilde{F})W^2(\tilde{F}|F_1)] \\ &= \tilde{\mathbb{E}}[W^2(F|\tilde{F})W(\tilde{F}|F_1)] = \tilde{\mathbb{E}}[\exp\{K_1 + \cdots + K_{m(u)}\}], \end{aligned}$$

donde

$$\begin{aligned} K_i &= \log \left(\frac{d\tilde{F}}{dF_1}(X_i) \left(\frac{dF}{d\tilde{F}}(X_i) \right)^2 \right) = \log \left(\frac{d\tilde{F}}{dF_1}(X_i) \right) + 2 \log \left(\frac{dF}{d\tilde{F}}(X_i) \right) \\ &= -\log \left(\frac{dF_1}{d\tilde{F}}(X_i) \right) + 2 \log \left(\frac{dF}{e^{\gamma x} dF}(X_i) \right) = -\log \left(\frac{dF_1}{d\tilde{F}}(X_i) \right) - 2\gamma X_i. \end{aligned}$$

Por lo tanto

$$\tilde{\mathbb{E}}[K_i] = \tilde{\mathbb{E}} \left[-\log \left(\frac{dF_1}{d\tilde{F}}(X_i) \right) \right] - 2\gamma \tilde{\mathbb{E}}[X_i] = \epsilon' - 2\gamma \tilde{\mathbb{E}}[X_i].$$

Y al utilizar la desigualdad de *Jensen* y la concavidad de la función logaritmo aseguramos que $\epsilon' > 0$. Como último paso hay que darse cuenta de que las variables K_1, K_2, \dots son i.i.d, entonces la desigualdad de *Jensen* y la identidad de *Wald* nos dicen que

$$\begin{aligned} \mathbb{E}_{F_1}[Z(u)^2] &= \tilde{\mathbb{E}}[\exp\{K_1 + \cdots + K_{m(u)}\}] \\ &\geq \exp\{\tilde{\mathbb{E}}[K_1 + \cdots + K_{m(u)}]\} \\ &= \exp\{\tilde{\mathbb{E}}[m(u)](\epsilon' - 2\gamma \tilde{\mathbb{E}}[X_i])\}. \end{aligned}$$

¹⁵Este *Teorema* Prueba que los parámetros de cambio óptimos son únicos.

Y ya que $\tilde{\mathbb{E}}[m(u)/u]$ converge a $1/\tilde{\mathbb{E}}[X_i]$, podemos concluir que para $0 < \epsilon < \epsilon'/\tilde{\mathbb{E}}[X_i]$,

$$\begin{aligned}
\limsup_{u \rightarrow \infty} \frac{\mathbb{E}_{F_1}[Z(u)^2]}{Z(u)^2 e^{\epsilon u}} &\geq \limsup_{u \rightarrow \infty} \frac{\exp\{\tilde{\mathbb{E}}[m(u)](\epsilon' - 2\gamma\tilde{\mathbb{E}}[X_i])\}}{C^2 e^{-2\gamma u + \epsilon u}} \\
&= \limsup_{u \rightarrow \infty} \frac{\exp\{(u/\tilde{\mathbb{E}}[X_i])(\epsilon' - 2\gamma\tilde{\mathbb{E}}[X_i])\}}{C^2 e^{-2\gamma u + \epsilon u}} \\
&= \limsup_{u \rightarrow \infty} \frac{\exp\{(u/\tilde{\mathbb{E}}[X_i])\epsilon' - 2\gamma u\}}{C^2 e^{-2\gamma u + \epsilon u}} \\
&\geq \limsup_{u \rightarrow \infty} \frac{e^{u\epsilon - 2\gamma u}}{C^2 e^{-2\gamma u + \epsilon u}} \\
&= \frac{1}{C^2} > 0.
\end{aligned}$$

Lo que completa la demostración.

◇

A continuación mostraremos la prueba del *Teorema 6*:

Consideremos el proceso Poisson Compuesto de riesgo con intensidades β', β'' , intervalos de tiempo genéricos T', T'' , distribución del tamaño de las reclamaciones B', B'' y tamaño genérico de las reclamaciones U', U'' . Bajo estas hipótesis, de acuerdo al *Teorema 7* todo lo que necesitamos probar es que si $U' - T' \stackrel{D}{=} U'' - T''$ (i.e. igualdad en distribución), ya que con esto aseguramos que $\beta' = \beta''$ y que $B' = B''$. Entonces si hacemos uso de la propiedad de *pérdida de memoria* de la distribución exponencial, tenemos que $U' - T'$ tiene cola exponencial izquierda con tasa β' , mientras que $U'' - T''$ tiene cola exponencial izquierda con tasa β'' , lo que implica que $\beta' = \beta''$. De este modo tenemos por un lado que

$$\mathbb{P}\{U' - T' > x\} = \int_0^\infty \beta' e^{-\beta' y} \bar{B}'(x + y) dy = \beta' e^{-\beta' x} \int_x^\infty e^{-\beta' z} \bar{B}'(z) dz,$$

donde hemos realizado el cambio de variable $z = x + y$. Y por otro lado si hacemos lo mismo pero ahora con $U'' - T''$ tenemos

$$\mathbb{P}\{U'' - T'' > x\} = \int_0^\infty \beta'' e^{-\beta'' y} \bar{B}''(x + y) dy = \beta'' e^{-\beta'' x} \int_x^\infty e^{-\beta'' z} \bar{B}''(z) dz.$$

Al diferenciar ambas probabilidades e igualar los resultados, obtenemos que

$$\beta' e^{-\beta' x} (-e^{-\beta' x} \bar{B}'(x)) = \beta'' e^{-\beta'' x} (-e^{-\beta'' x} \bar{B}''(x)).$$

Y ya que $x > 0$, $\beta' = \beta''$ y $U' - T' \stackrel{D}{=} U'' - T''$ podemos concluir que $\hat{B}'(x) = \hat{B}''(x)$ para todo $x > 0$, i.e. $B' = B''$.

◇

§5.3.3. Muestreo de Importancia para el caso de Horizonte Finito.

El problema central es producir un estimador eficiente para $\Psi(u, T)$ con $T < \infty$. Con esto en mente, primero debemos establecer algunas ideas con respecto al momento en el que ocurre la ruina, para después aplicar este conocimiento al muestreo de importancia.

Para el modelo Poisson Compuesto general, los resultados conocidos son poco explícitos y básicamente toman la forma de desigualdades y aproximaciones. El primer resultado importante de esta sección es que el valor $u\tilde{m}$, donde u representa el capital inicial y

$$\tilde{m} = \frac{1}{k'(\gamma)} = \frac{1}{\beta M'_B(\gamma) - 1} = \frac{1}{\tilde{\beta} \tilde{\mathbb{E}}[U] - 1} = \frac{C}{1 - \rho}.$$

Es en un sentido crítico apropiado, el tiempo más parecido al tiempo de ruina. Los resultados posteriores tratan con mayor precisión y versiones refinadas de este resultado.

Teorema 7. *Supongamos que $\eta > 0$. Entonces, $\tau(u)/u$ converge en probabilidad a \tilde{m} conforme u tiende a infinito.*

$$\mathbb{P} \left\{ \left| \frac{\tau(u)}{u} - \tilde{m} \right| > \epsilon \mid \tau(u) < \infty \right\} \xrightarrow{u \rightarrow \infty} 0. \quad (5.50)$$

Esto implica que

$$\frac{\Psi(u, mu)}{\Psi(u)} \rightarrow \begin{cases} 0 & \text{si } m < \tilde{m} \\ 1 & \text{si } m > \tilde{m} \end{cases} \quad (5.51)$$

Para poder probar esto, necesitamos el siguiente resultado auxiliar:

Proposición 8. Si $\eta < 0$, i.e. $\rho = \beta\mu_B > 1$, entonces conforme u tiende a infinito,

$$\frac{\tau(u)}{u} \xrightarrow{c.s} m = \frac{1}{\rho - 1} \quad y \quad \frac{\mathbb{E}[\tau(u)]}{u} \rightarrow \frac{1}{\rho - 1}. \quad (5.52)$$

Además

$$\frac{\tau(u) - \mu}{\sqrt{u}} \xrightarrow{D} N(0, w^2), \quad \text{donde } w^2 = \beta\mu_B^2 m^3. \quad (5.53)$$

Demostración

El supuesto $\eta < 0$ asegura que $\mathbb{P}\{\tau(u) < \infty\} = 1$ y $\tau(u)$ converge casi seguramente a infinito. Además por la *Proposición 2*, sabemos que S_t/t converge casi seguramente a $1/m$ y por lo tanto tenemos la convergencia casi segura de

$$m = \lim_{t \rightarrow \infty} \frac{t}{S_t} = \lim_{u \rightarrow \infty} \frac{\tau(u)}{S_\tau(u)} = \lim_{u \rightarrow \infty} \frac{\tau(u)}{u + \xi_{\tau(u)}} = \lim_{u \rightarrow \infty} \frac{\tau(u)}{u}.$$

Aquí utilizamos que si $\xi(t) = \inf\{(T_k - T_{k-1}) - t \mid t < T_k - T_{k-1}\}$, entonces $\xi_{\tau(u)} = o(u)$ c.s (*Proposición A1.6* de [2]) ya que para el proceso de renovación asociado, si $\mu_B < \infty$ entonces $\xi(t)/t$ converge casi seguramente a 0 y si además $\mathbb{E}[T_0] < \infty$ tenemos que $\mathbb{E}[\xi(t)/t]$ converge a 0. Para probar esto último debemos considerar que N_t , el número de renovaciones antes de t , satisface que N_t/t converge con probabilidad 1 a μ_B , por lo que para un t lo suficientemente grande podemos acotar $\xi(t)$ por $M(t) = \max\{T_k - T_{k-1} \mid k < 2t/\mu_B\}$. Y ya que el máximo M_n de n variables aleatorias i.i.d. con media finita satisface que M_n/n converge casi seguramente a 0 (por *Borel-Cantelli*), implica que $\xi(t)/t$ converge con probabilidad 1 a 0. Si además $\mathbb{E}[T_0] < \infty$, y suponemos que $\mathbb{E}_o[\xi(t)]$ satisface la ecuación

$$z(t) = \mathbb{E}[(T_1 - T_0) - t; (T_1 - T_0) > t],$$

entonces

$$\mathbb{E}_o[\xi(t)] = \int_0^t U(dy)z(t-y) = \int_0^t U(t-dy)z(y) \leq c \sum_{k=0}^{\lfloor t \rfloor + 1} z(k),$$

con $U = \sum_0^\infty F^{*n}$, donde F^{*n} representa la n -ésima convolución de F y $c = \sup_x \{U(x+1) - U(x)\}$ ($c < \infty$ ya que $U(x+1) - U(x) \leq U(1)$). Como $z(k) \leq \mathbb{E}[(T_1 - T_0); (T_1 - T_0) > t]$ converge a 0, la suma es $o(t)$ y por lo tanto $\mathbb{E}_o[\xi(t)/t]$ converge a 0.

De este modo el primer resultado de la *Proposición* queda demostrado, i.e.

$$\frac{\tau(u)}{u} \xrightarrow{c.s.} m.$$

Para probar que

$$\frac{\mathbb{E}[\tau(u)]}{u} \rightarrow \frac{1}{\rho - 1},$$

utilizamos la identidad de *Wald* de la siguiente forma

$$u + \mathbb{E}[\xi(u)] = \mathbb{E}[S_{\tau(u)}] = \mathbb{E}[\tau(u)]\mathbb{E}[S_1] = (\rho - 1)\mathbb{E}[\tau(u)],$$

y ya que $\mathbb{E}[\xi(u)/u]$ converge a 0, afirmamos la segunda parte de (5.52).

Por último, para demostrar (5.53) aplicamos el *Teorema del Límite Central* para concluir que

$$\frac{S_t - t/m}{\sqrt{t}} \xrightarrow{D} N(0, \beta\mu_B^2).$$

Entonces de acuerdo con el *Teorema de Anscombe* (*Teorema* 7.3.2 de [4]) y (5.52), el mismo resultado se mantiene para $t = \tau(u)$. De modo que si Z se distribuye $N(0, 1)$, entonces

$$\begin{aligned} \frac{S_{\tau(u)} - \tau(u)/m}{\sqrt{\tau(u)}} &= \frac{u + \xi(u) - \tau(u)/m}{\sqrt{\tau(u)}} \approx \sqrt{\beta\mu_B^2}Z \Rightarrow \\ \frac{\tau(u) - mu}{\sqrt{\tau(u)}} &\approx -m\sqrt{\beta\mu_B^2}Z \xrightarrow{D} m\sqrt{\beta\mu_B^2}Z \Rightarrow \\ \frac{\tau(u) - mu}{\sqrt{u}} &\approx m^{3/2}\sqrt{\beta\mu_B^2}Z = wZ. \end{aligned}$$

◇

Demostración del Teorema 7

Comencemos por notar que el lado izquierdo de la ecuación (5.50) es igual a

$$\begin{aligned} \frac{\mathbb{P}\left\{\left|\frac{\tau(u)}{u} - \tilde{m}\right| > 0, \tau(u) < \infty\right\}}{\mathbb{P}\{\tau(u) < \infty\}} &= \frac{e^{-\gamma u} \tilde{\mathbb{E}}\left[e^{-\gamma\xi(u)}; \left|\frac{\tau(u)}{u} - \tilde{m}\right| > \epsilon, \tau(u) < \infty\right]}{\Psi(u)} \\ &\leq \frac{e^{-\gamma u} \tilde{\mathbb{P}}\left\{\left|\frac{\tau(u)}{u} - \tilde{m}\right| > \epsilon\right\}}{O(e^{-\gamma u})}. \end{aligned}$$

Como $\tilde{\mathbb{P}}\{\cdot\}$ converge a 0 por la *Proposición 8*, entonces

$$\mathbb{P}\left\{\left|\frac{\tau(u)}{u} - \tilde{m}\right| > \epsilon \mid \tau(u) < \infty\right\} \longrightarrow 0.$$

La ecuación (5.51) queda inmediatamente demostrada con el siguiente argumento

$$\frac{\Psi(u, T)}{\Psi(u)} = \frac{\mathbb{P}\{\tau(u) \leq mu\}}{\mathbb{P}\{\tau(u) < \infty\}} = \frac{\mathbb{P}\left\{\frac{\tau(u)}{u} \leq m\right\}}{\mathbb{P}\{\tau(u) < \infty\}}.$$

Por lo tanto ya que el denominador es menor que 1 pues $\eta > 0$, entonces si $m < \tilde{m}$ tenemos que $\tau(u)/u$ converge en probabilidad a \tilde{m} , ya que (5.50) se cumple. Y como el numerador tiende a cero bajo estas condiciones, concluimos que el cociente mostrado arriba tiende a cero y si $m > \tilde{m}$, entonces tiende a 1 ya que tanto numerador como denominador coinciden. Por lo tanto (5.51) se cumple.

◇

Este último resultado nos indica que podemos esperar diferentes conclusiones dependiendo de si $y < 1/k'(\gamma)$ o $y > 1/k'(\gamma)$. El caso fácil es cuando $y > 1/k'(\gamma)$, donde $\Psi(u, T)$ es cercana a $\Psi(u)$, de modo que es natural esperar que el cambio de medida de \mathbb{P} a $\tilde{\mathbb{P}}$ produzca resultados cercanamente óptimos. De hecho

Proposición 9. *Si $y > 1/k'(\gamma)$, entonces el estimador $Z(u) = e^{-\gamma S_{\tau(u)}} \mathbb{I}_{\{\tau(u) \leq yu\}}$ simulado bajo los parámetros $\tilde{\beta}$ y \tilde{B} tiene error relativo acotado.*

Demostración

Por el *Teorema 7* tenemos que $\Psi(u, T)/\Psi(u)$ tiende a 1, de modo que $z(u) = \Psi(u, yu)$ es del orden de magnitud de $e^{-\gamma u}$ i.e. son asintóticamente parecidos. Entonces al realizar un procedimiento parecido al de la demostración del *Teorema 4* aseguramos que

$$\tilde{\mathbb{E}}[Z(u)^2] \leq e^{-\gamma u}.$$

Por lo tanto $Z(u)$ tiene error relativo acotado.

◇

Nuestro siguiente paso es analizar el caso $y < 1/k'(\gamma)$, para el cual consideraremos a α_y como la solución de la ecuación $k'(\alpha) = 1/y$, y que $\gamma_y = \alpha_y - yk(\alpha_y)$ determina el orden de magnitud de $\Psi(u, yu)$ en el sentido

$$\frac{-\log \Psi(u)}{u} \rightarrow \gamma_y.$$

La demostración de este hecho se muestra en el *Teorema IV.4.8* de [2], utilizando que $\gamma_y > \gamma$. Aún más, basándonos en ese resultado podemos afirmar que

$$\Psi(u, yu) = e^{-\gamma_y u} \mathbb{E}_{\alpha_y}[\exp\{-\alpha_y \xi(u) + \tau(u)k(\alpha_y)\}; \tau(u) \leq yu]. \quad (5.54)$$

Ya que la definición de α_y es equivalente a decir que $\mathbb{E}_{\alpha_y}[\tau(u)]$ es asintóticamente parecido a yu , podemos esperar que el cambio de medida de \mathbb{P} a $\tilde{\mathbb{P}}$ sea óptimo. El estimador correspondiente a este esquema de muestreo de importancia es

$$Z(u) = \exp\{-\alpha_y \xi(u) + \tau(u)k(\alpha_y)\} \mathbb{I}_{\{\tau(u) \leq yu\}}, \quad (5.55)$$

para el cual tenemos el siguiente resultado

Teorema 8. *El estimador (5.55) simulado con los parámetros β_{α_y} y B_{α_y} es logarítmicamente eficiente.*

Demostración

Ya que $\gamma_y > \gamma$, tenemos que $k(\alpha_y) > 0$ entonces

$$\begin{aligned} \mathbb{E}_{\alpha_y}[Z(u)^2] &= \mathbb{E}_{\alpha_y}[\exp\{-2\alpha_y \xi(u) + 2\tau(u)k(\alpha_y)\}; \tau(u) \leq yu] \\ &\leq e^{-2\gamma_y u} \mathbb{E}_{\alpha_y}[e^{-2\alpha_y \xi(u)}; \tau(u) \leq yu] \\ &\leq e^{-2\gamma_y u}. \end{aligned}$$

Por lo tanto, como $-\log(\Psi(u))/u$ converge a γ_y tenemos que

$$\liminf_{u \rightarrow \infty} \frac{-\log(\text{Var}[Z(u)])}{-\log(z(u))} = \liminf_{u \rightarrow \infty} \frac{-\log(\text{Var}[Z(u)])}{-\log(\Psi(u))} = \liminf_{u \rightarrow \infty} \frac{-\log(\text{Var}[Z(u)])}{\gamma_y u} \geq 2.$$

◇

Capítulo 6

Apéndice A

En esta sección se realiza una simulación basada en el desarrollo teórico mostrado en la **Sección 5.2.7**, para calcular el valor de un *call asiático* y así mostrar la efectividad de la aplicación del muestreo de importancia desde un aspecto numérico en uno de los casos estudiados en el capítulo anterior.

Cuadro 6.1: El cuadro muestra la simulación hecha para calcular el valor de un *call asiático* (C^K), su desviación estándar (Desv.), varianza (Var.), error estándar (E) y un intervalo del 95 % de confianza para la estimación (IC). En todos los caso se tomó $S(0) = 100$, $r = 0.2$, $T = 4$ y $m = 365$.

| Monte Carlo Crudo | | | | | | | |
|-------------------------|-----|--------|---------|---------|----------|--------|--------------------|
| σ | K | τ | C^K | Desv. | Var. | E | IC |
| 0.2 | 90 | 0 | 28.7755 | 17.4253 | 303.6418 | 0.5510 | (27.6950, 29.8551) |
| | | 1 | 34.2191 | 23.0101 | 529.4639 | 0.7276 | (32.7930, 35.6453) |
| | | 2 | 42.6921 | 28.2891 | 800.2714 | 0.8946 | (40.9387, 44.4455) |
| | | 3 | 51.4290 | 35.1816 | 1237.7 | 1.1125 | (49.2484, 53.6096) |
| Muestreo de Importancia | | | | | | | |
| σ | K | τ | C^K | Desv. | Var. | E | IC |
| 0.2 | 90 | 0 | 32.2985 | 0.2174 | 0.0473 | 0.0069 | (32.2850, 32.3119) |
| | | 1 | 33.3323 | 0.2544 | 0.0647 | 0.0080 | (33.2850, 33.3481) |
| | | 2 | 32.2604 | 0.2613 | 0.0683 | 0.0083 | (32.2442, 32.2766) |
| | | 3 | 32.5676 | 0.2776 | 0.0771 | 0.0088 | (32.5504, 32.5848) |

El número de simulaciones hechas para cada caso fue $n = 1000$. Se muestran resultados para un valor de volatilidad $\sigma = 0.2$, precio pactado (*strike*) $K = 90$ y periodo de inicio de ponderación $\tau = 0, 1, 2, 3$.

Los resultados numéricos nos muestran una notable y amplia mejoría de la estimación en cuanto a varianza se refiere.

A continuación se muestran los códigos de los programas elaborados en *Matlab* utilizados en el **Capítulo 1** y el **Apendice A**.

§6.1. MCCN.m

%Esta rutina estima $p=P[x \geq \mu]$, cuando X se distribuye $N(0,1)$
%haciendo Monte Carlo Crudo, sobre una muestra de tamaño n .

```
function [E,IC]=MCCN(n,mu)

x=randn(n,1);           %Simula n normales estándar.
I=[x>=mu];             %Función indicadora sobre el
                        %evento de interés.
N=sum(I);               %Número total de valores aceptados.
p_est=mean(I);          %Estimación MCC.
st=std(I);              %Calculo de la desviación estándar.
E=[p_est,st^2,N];      %Regresa el valor estimado de p,
                        %su varianza y N.
IC=IC95(p_est,st,n);   %Calcula el intervalo del 95%
                        %de confianza para p_est.
```

§6.2. ISN.m

%Esta rutina estima $p=P[x \geq \mu]$, cuando X se distribuye $N(0,1)$
%bajo el Muestreo de Importancia con μ arbitraria, sobre
%una muestra de tamaño n .

```
function [I,IC]=ISN(n,mu)

x=randn(1,n)+ mu;      %Cambio de medida.
I=[x>=mu];             %Función indicadora sobre
                        %el evento de interés.
```

```

N=sum(I); %Número total de valores
          %aceptados.
I=[x>=mu].*exp(-mu*x + mu^2/2); %Función indicadora sobre
          %el evento de interés.
p_est=mean(I); %Estimación MCC.
st=std(I); %Calculo de la desviación
          %estándar.
I=[p_est,st^2,N]; %Regresa el valor estimado
          %de p, su varianza y N.
IC=IC95(p_est,st,n); %Calcula el intervalo del
          %95% de confianza para
          %p_est.

```

§6.3. IC95.m

```

%Esta función calcula el intervalo del 95% de confinaza
%para p_est
%con desviación st y tamaño de muestra n.

```

```

function C=IC95(p_est,st,n)

C=[p_est-1.96*st/sqrt(n),p_est+1.96*st/sqrt(n)];

```

§6.4. call-asiatico-MCC.m

```

%Esta función calcula el valor,la desviación estándar,
%la varianza,el error estandar y un intervalo de
%confianza del 95% de un call asiático utilizando MCC.
%Además grafica la trayectoria descrita por el
%instrumento.
%Recibe como parámetros de entrada:
%S0=el valor actual del activo.
%K=strike.
%r=tasa de interés libre de riesgo, por unidad de
%tiempo.
%T=tiempo en que se debe ejercer el contrato.
%m=número de puntos por unidad de tiempo, que se usan

```

```

%para construir el promedio. i.e. 1/m es la longitud
%de cada subintervalo de tiempo.
%sigma=volatilidad del instrumento.
%n=número de simulaciones.
%tau=tiempo en el que comienza el promedio, en la
%mayoría de los casos,por simplicidad se toma tau=0.

function [val_opc,desv_est,varianza,error_est,IC]=...
    call_asiatico_MCC(S0,K,r,T,m,sigma,n,tau)

D=zeros(1,T-1);           %D=vector de dimensión
                           %(T-1) que representa
                           %el pago de los dividen-
                           %dos al final de cada
                           %unidad de tiempo.

a=[1:n];                 %Vector de trayectorias
                           %simuladas.
b=a;                     %Vector que indica las
                           %trayectorias posibles.
c=n;                     %Número de elementos en
                           %b.
final=zeros(1,n);       %Vector de trayectorias
                           %completas.

while c>0                %Mientras tengamos
                           %trayectorias incompletas
                           %(es decir,que existen
                           %trayectorias
                           %imposibles en los
                           %ciclos previos).

    S=S0*ones(1,n);     %Comienza cada trayectoria
                           %con S0.

    t=1;
    while (t<=tau) & (c>0) %Calcula S(t), para t en
                           %{1,2,...,tau} ya para

```

```

%todas las k para las que
%aún no existan trayecto-
%rias completas.

S(b)=S(b).*exp((r-(1/2)*(sigma^2))+...
                sigma*randn(1,c));
b=b(S(b)>=D(t));      %Las trayectorias sólo
                    %son posibles si S(t)>=D(t)
S(b)=S(b)-D(t);      %Entonces el valor del
                    %siguiente periodo es
                    %S(t)-D(t).
c=length(b);         %Actualiza las variables.
t=t+1;
end

prom=S;              %Se inicia el promedio
                    %al tiempo tau

while (t<T) & (c>0) %Calcula S(t), para t en
                    %{tau+1/m,...,T-1} para
                    %todas a aquellas k, para
                    %las cuales no existen
                    %aún trayectorias completas.

    for i=1:m         %Calcula s(t-1+i/m), i en
                    %{1,...,m} y actualiza prom.

        S(b)=S(b).*exp(((r-(1/2)*(sigma^2))/m)...
                        +(sigma/sqrt(m))*randn(1,c));

        prom(b)=prom(b)+S(b);
    end

    b=b(S(b)>=D(t)); %Nuevamente, las trayecto-
                    %rias sólo son posibles si
                    %S(t)>=D(t).

```

```

        S(b)=S(b)-D(t);           %Entonces el valor inicial
                                %para el siguiente periodo
                                %es S(t)-D(t).
        c=length(b);           %Actualiza las variables.
        t=t+1;
    end                           %En este momento tenemos
                                %el valor al principio del
                                %último periodo, y los
                                %valores correctos para
                                %los índices en b.

        final(b)=ones(1,c);     %Repetimos el ciclo para
                                %aquellos índices k para
                                %los el último ciclo resul-
                                %to en una trayectoria
                                %imposible.

        b=a(final==0);
        c=length(b);
        S(b)=S0*ones(1,c);
    end                           %Ahora cada trayectoria
                                %es correcta y S es un
                                %vector de n trayectorias
                                %al inicio del T-ésimo
                                %periodo.

    for i=1:m
        U=randn(1,n);
        S=S.*exp((r-(1/2)*(sigma^2))/m+(sigma/sqrt(m))*U);
        prom=prom+S;
                                %Calcula S(T-1+i/m), i en
                                %{1,...,m} y actualiza
                                %el valor de prom.
                                %S es ahora S(T)
    end

    prom=prom/(m*(T-tau)+1);     %Calcula el promedio.

```

```

V=exp(-r*T)*max([(prom-K) ; zeros(1,n)]);

                                %Calcula el valor de la
                                %opción para cada
                                %trayectoria.
val_opc=mean(V);                %Calcula el valor de
                                %la opción.
desv_est=std(V);                %Calcula la desviación
                                %estándar del valor
                                %de la opción.
varianza=var(V);                %Calcula su varianza.

error_est=sqrt(sum((V-val_opc).^2)/(n-1)/n);
                                %Calcula el error
                                %relativo estimado.
IC=IC95(val_opc,desv_est,n);    %Calcula un intervalo
                                %de confianza del 95%.
plot(V);                          %Grafica la trayectoria
                                %de la opción.

```

§6.5. call-asiatico-IS.m

```

%Esta función calcula el valor, la desviación estándar,
%la varianza,el error estandar y un intervalo de
%confianza del 95% de un call asiático utilizando IS.
%Además muestra la grafica de la trayectoria descrita
%por el instrumento.
%Recibe como parámetros de entrada:
%S0=el valor actual del activo.
%K=strike.
%r=tasa de interés libre de riesgo,por unidad de
%tiempo.
%T=tiempo en que se debe ejercer el contrato.
%m=número de puntos por unidad de tiempo, que se
%usan para construir el promedio. i.e. 1/m es la
%longitud de cada subintervalo de tiempo.

```

```

%sigma=volatilidad del instrumento.
%n=número de simulaciones.
%tau=tiempo en el que comienza el promedio, en la
%mayoría de los casos por simplicidad se toma tau=0.

function [val_opc,desv_est,varianza,error_est,IC]=...
    call_asiatICO_IS(S0,K,r,T,m,sigma,n,tau)

D=zeros(1,T-1);           %D=vector de dimensión
                           %(T-1) que representa
                           %el pago de los
                           %dividendos al final
                           %de cada unidad de
                           %tiempo.

a=[1:n];                 %Vector de trayectorias
                           %simuladas.
b=a;                     %Vector que indica las
                           %trayectorias posibles.
c=n;                     %Número de elementos en b.
final=zeros(1,n);       %Vector de trayectorias
                           %completas.

while c>0                %Mientras tengamos
                           %trayectorias incompletas,
                           %es decir que existen
                           %trayectorias imposibles
                           %en los ciclos previos.

    S=S0*ones(1,n);     %Comienza cada trayecto-
                           %ria con S0.

    t=1;
    while (t<=tau) & (c>0) %Calcula S(t), para t en
                           % {1,2,...,tau} ya para
                           % todas las k para las que
                           % aún no existan trayecto-
                           % rias completas.

```

```

S(b)=S(b).*exp((r-(1/2)*(sigma^2))+...
                sigma*(randn(1,c)));

b=b(S(b)>=D(t));           %Las trayectorias sólo
                           %son posibles si S(t)>=D(t).
S(b)=S(b)-D(t);          %Entonces el valor del
                           %siguiente periodo es
                           %S(t)-D(t).

c=length(b);             %Actualiza las variables.
t=t+1;

end

prom=S;                   %Se inicia el promedio al
                           %tiempo tau.

while (t<T) & (c>0)      %Calcula S(t), para t en
                           %{tau+1/m,...,T-1} para
                           %todas aquellas k, para
                           %las cuales no existen
                           %aún trayectorias completas.

for i=1:m                  %Calcula s(t-1+i/m), i en
                           %{1,...,m} y actualiza prom.

    S(b)=S(b).*exp(((r-(1/2)*(sigma^2))/m)...
                    +(sigma/sqrt(m))*(randn(1,c)));

    prom(b)=prom(b)+S(b);
end

b=b(S(b)>=D(t));           %Nuevamente, las trayecto-
                           %rias sólo son posibles si
                           %S(t)>=D(t).
S(b)=S(b)-D(t);          %Entonces el valor inicial
                           %para el siguiente periodo
                           %es S(t)-D(t).

```

```

        c=length(b);           %Actualiza las variables.
        t=t+1;
    end                         %En este momento tenemos
                                %el valor al principio del
                                %último periodo, y los
                                %valores correctos para
                                %los índices en b.

        final(b)=ones(1,c);    %Repetimos el ciclo para
                                %aquellos índices k para
                                %los el último ciclo resul-
                                %to en una trayectoria
                                %imposible.

        b=a(final==0);
        c=length(b);
        S(b)=S0*ones(1,c);
    end                         %Ahora cada trayectoria es
                                %correcta y S es un vector
                                %de n trayectorias al
                                %inicio del T-ésimo periodo.

    for i=1:m                   %Calcula S(T-1+i/m), i
                                %en {1,...,m} y actualiza
                                %el valor de prom.

        U=randn(1,n);
        S=S.*exp((r-(1/2)*(sigma^2))/m+(sigma/sqrt(m))*U);
        prom=prom+S;           %S es ahora S(T).
    end

    prom=prom/(m*(T-tau)+1);    %Calcula el promedio.
    v=sum(prom);

                                %Encuentra el parámetro
                                %óptimo de manera numérica.

    f=@(x) (1/n)*v-K-x;

```

```

y=fzero(f,0); %Raíz
z=zeros(1,n); %Recursión de las ecuaciones
               %de primer orden
               %(problema de maximización).

z(1)=((sigma*sqrt((1/m))*(y+K))/y);
for j=1:n-1
    z(j+1)=z(j)-((sigma*sqrt(1/m)*S(j))/(n*y));
end

mu=z; %Parámetro de cambio óptimo.

%Se realiza una nueva
%simulación via IS con el
%parámetro de cambio mu.
dt = T/n; %Longitud de cada
           %sub-intervalo de tiempo.
SS = S0 * ones(1,n); %Vector de trayectorias.
AA = zeros(1,n);
for i=1:n %Simulación de trayecto-
           %rias y cálculo del
           %promedio.
U= randn(1,n)+mu;
dWW = U* sqrt(dt);
SS = SS .* exp((r - .5* sigma ^2)* dt + sigma *dWW);
AA = AA + SS;
end
AA = AA/n;

V=exp(-r*T)* mean(max(AA-K,0))*...
exp(-mu'*U+(1/2)*mu'*mu);

%Valor de la opción por
%cada trayectoria.

VV=mean(V);
val_opc=mean(VV); %Valor de la opción.
desv_est=std(VV); %Desviación estándar.

```


Bibliografía

- [1] S. Asmussen. *Applied Probability and Queues*. Wiley, 1987.
- [2] S. Asmussen. *Ruin Probabilities*. World Scientific, 2000.
- [3] O.E. Barndorff-Nielsen. *Information and Exponential Families in Statistical Theory*. Wiley, 1978.
- [4] K.L. Chung. *A Course in Probability Theory*. Academic Press, 1974.
- [5] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*. Springer, 1998.
- [6] R. M. Dudley. *Real Analysis and Probability*. Wadsworth, 1989.
- [7] R. Durrett. *Probability. Theory and Examples*. Wadsworth, 1991.
- [8] G. B. Folland. *Real Analysis. Modern Techniques and Their Applications*. Wiley, 1999.
- [9] P. Glasserman. *Monte Carlo Methods in Financial Engineering*. Springer, 2004.
- [10] P. Glasserman, P. Heidelberger, and P. Shahabuddin. Asymptotically optimal importance sampling and estratification for pricing path-dependent options. *Mathematical Finance*, pages 117 – 152, 1999.
- [11] P. Glasserman and Y. Wang. Counterexamples in importance sampling for large deviations probabilities. *The Annals of Applied Probability*, pages 731 – 806, 1997.
- [12] P. W. Glynn and D. L. Iglehart. Importance sampling for stochastic simulations. *Management Science*, pages 1367 – 1392, 1989.
- [13] J. Grandell. *Aspects of Risk Theory*. Springer-Verlag, 1991.

- [14] P. Heidelberger, P. Shahabuddin, and P. Glasserman. Importance sampling and stratification for value-at-risk.
- [15] T. C. Hesterberg. *Advances in Importance Sampling*. PhD thesis, Stanford University, 1988.
- [16] J. C. Hull. *Options, Futures and Other Derivatives*. Prentice-Hall, 2000.
- [17] C. Klupperberg, T. Mikosch, and P. Embrechts. *Modelling Extremal Events for Insurance and Finance*. Springer, 1997.
- [18] P.Ñey, J. S. Sadowsky, and J. A. Bucklew. Monte carlo simulation and large deviations theory for uniformly recurrent markov chains. *The Annals of Applied Probability*, pages 44 – 59, 1990.