

94
201



UNIVERSIDAD NACIONAL AUTONOMA
DE MEXICO

FACULTAD DE CIENCIAS

“ALGUNAS TECNICAS DE
INTEGRACION MONTE CARLO”

T E S I S
QUE PARA OBTENER EL TITULO DE
A C T U A R I A
P R E S E N T A
LIZBETH ROMAN PADILLA

Director de Tesis:
DR. JOSE RODOLFO MENDOZA BLANCO



MEXICO, D. F.

JUNIO 1997

CIENCIAS
BLANCO



TESIS CON
FALLA DE ORIGEN



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

PAGINACION VARIA

COMPLETA LA INFORMACION



UNIVERSIDAD NACIONAL
AUTÓNOMA DE
MÉXICO

M. en C. Virginia Abrín Batule
Jefe de la División de Estudios Profesionales de la
Facultad de Ciencias
Presente

Comunicamos a usted que hemos revisado el trabajo de Tesis:

"Algunas técnicas de integración Monte Carlo"

realizado por Lizbeth Román Padilla

con número de cuenta 8926228-1 , pasante de la carrera de Actuaría

Dicho trabajo cuenta con nuestro voto aprobatorio.

Atentamente

Director de Tesis	
Propietario	Dr. José Rodolfo Mendoza Blanco
Propietario	Dr. Eduardo Arturo Gutiérrez Peña
Propietario	Mat. Margarita Elvira Chávez Cano
Suplente	M. en C. José Salvador Zamora Muñoz
Suplente	M. en A.B. María del Pilar Alonso Reyes


Consejo Departamental de Matemáticas

José Rodolfo Mendoza Blanco
M. en C. Margarita Elvira Chávez Cano
José Salvador Zamora Muñoz
María del Pilar Alonso Reyes

**A mis tres mejores amigas
A las compañeras de lágrimas y aventuras
A mi mamá y hermanas**

Blanca, Bety y Abby

Con cariño a mi padre

A mis abuelitos, Odilón Padilla y Esperanza Robles.

A mis tíos: Fernando, Alicia, José Luis, Javier, Lilia, Paula y Marco Antonio.

A mis queridos primos: Carlos, Enrique, Lucy, Ricardo, Mario, Fernandito, Nancy, Karina, Luis Antonio, Hermaín, Claudia y Francisco Javier.

A mis amigas Lilian Mendoza Z., Erika Rodríguez, Claudia Villegas A. y Martha A. López C.

A los compañeros tesisistas.

A las profesoras Concepción Salinas, Concepción Ortiz y Carmen Arteaga de quienes tengo un grato recuerdo.

A las instituciones de educación pública y en especial a la U.N.A.M. por acogerme, enseñarme a crecer, y brindarme la oportunidad de conocer a un magnífico profesor, el Dr. José R. Mendoza Blanco a quien agradezco sinceramente por su apoyo y confianza.

Gracias a todas la personas que se encuentran cerca de nosotras y que de alguna forma nos han ayudado a seguir adelante.

Agradezco a los profesores

Mat. Margarita Chávez Cano,

Dr. Eduardo Gutiérrez Peña,

M. en A.P. Pilar Alonso Reyes y

M. en C. Salvador Zamora Muñoz

por el tiempo dedicado a la lectura del presente trabajo y por sus observaciones.

NOTACION

Antes de iniciar la lectura de la presente, se recomienda leer este apartado para familiarizarse con la simbología y notación que se hace uso en casi la totalidad del contenido.

En el extremo izquierdo se encuentra la simbología y a la derecha el significado.

$v.a$	Variable aleatoria;
$i.i.d.(I.I.D.)$	Independientes e idénticamente distribuidos;
$i.i.d.(I.I.D.) (\mu, \sigma^2)$	Independientes e idénticamente distribuidos con media μ y varianza σ^2 ;
$f.d.p.$	Función de densidad de probabilidad;
$f_Y(\bullet)$	Función de densidad de probabilidad de Y ;
$I_A(x)$	Función Indicadora de X sobre el conjunto A ;
X, Y, Z, \dots	Variable aleatoria, denotadas con mayúsculas;
$\hat{f}(\bullet)$	Densidad estimada;

los diferentes modos de convergencia se denotarán como

$\xrightarrow{\text{c.s.}}$	Convergencia casi segura;
\xrightarrow{p}	Convergencia en probabilidad;
\implies	Convergencia en distribución;

Desde el punto de vista Bayesiano la siguiente es la notación:

f_θ	Función de densidad <i>a priori</i> ;
$L(\bullet)$	Función de verosimilitud;
$p(\bullet)$	Función de densidad posterior;

y

A' La matriz A transpuesta.

PREFACIO

El problema de calcular ciertas integrales ha dado origen a numerosos métodos para su evaluación y una de las herramientas más utilizadas en el Análisis Numérico para aproximarlas es la integración Monte Carlo. Su aplicación se realiza principalmente en el cálculo de integrales con k dimensiones cuya solución en forma analítica es difícil de obtener.

En la presente se hace referencia principalmente a el algoritmo de Muestreo Gibbs como una alternativa para la obtención de integrales difíciles junto con la integración Monte Carlo, presentando pruebas de convergencia e ilustrando su implementación mediante ejemplos sencillos.

El algoritmo de Muestreo Gibbs es similar al algoritmo de Sustitución Sucesiva, que se describirá en el primer Capítulo, y ambos se utilizan en la integración Monte Carlo desembocando así en una estimación de la integral buscada.

Dentro del primer Capítulo se contempla el problema a resolver dando herramientas matemáticas necesarias para que el lector no experimentado se familiarice con los conceptos, además se incluye la metodología de la integración Monte Carlo y se detallan los algoritmos de Muestreo de Sustitución Sucesiva y Muestreo Gibbs, finalizando con una breve introducción sobre Inferencia Bayesiana.

El Capítulo 2 retoma la última parte del primer Capítulo; y aquí se discute una forma de aproximar la esperanza de una función objetivo con respecto a la función de densidad posterior, incluyendo algunos supuestos generales para la convergencia casi segura a la verdadera esperanza. Se introduce el concepto de la densidad de Muestreo de Importancia, cuya utilidad se ejemplificará ampliamente en el Capítulo 4. Para finalizar el Capítulo 2 se presenta una sección donde se aplican los resultados obtenidos originando así la estimación

por intervalo de la esperanza posterior de la función objetivo.

El Capítulo 3 se basa principalmente en la prueba de la convergencia del algoritmo de Sustitución Sucesiva (donde el Muestreo Gibbs es equivalente a el muestreo de Sustitución Sucesiva, bajo ciertas condiciones), introduciendo los conceptos y resultados preliminares necesarios a lo largo de la demostración.

El Capítulo 4 presenta un estudio de simulación y una serie de gráficas que ilustran los métodos descritos en los Capítulos 1 a 3. En particular se ejemplifica el concepto de densidad de Muestreo de Importancia, se hace una estimación de la varianza para los valores encontrados y se presentan diferentes formas que visualizan la convergencia así como algunas sugerencias.

En el Capítulo 5 se incluye un problema real procedente de un estudio sobre el crecimiento de un grupo de ratas. La finalidad del análisis es estimar las densidades posteriores marginales de los parámetros poblacionales que describen dicho crecimiento. En este Capítulo se resumen muchos de los conceptos discutidos anteriormente, principalmente de Simulaciones. Además es de gran importancia porque aplica de manera práctica el Muestreo Gibbs detallando la metodología de su aplicación en forma breve.

Dentro de las conclusiones se discuten los resultados obtenidos y se presentan algunas observaciones y sugerencias que aparecieron a lo largo del trabajo.

En la parte final se anexa un Apéndice con los listados codificados de los programas que generaron las gráficas y tablas de valores de los Capítulos 1,2,4 y 5 así como los desarrollos matemáticos de acuerdo a la finalidad del programa. Todos se encuentran codificados en el lenguaje de programación *S-PLUS 3.3*. La razón por la cual se hizo uso de este lenguaje es su flexibilidad en su manejo y la presentación de las gráficas.

El método que se describirá no resulta ser más eficiente que lo desarrollados dentro del Análisis Numérico, pero en el caso de dimensiones grandes presenta simplicidad y facilidad de implementación para los usuarios con conocimientos básicos de computación y experiencia dentro de la Estadística.

Después de este breve bosquejo se hace una invitación al lector a adentrarse en forma detallada a las pruebas de convergencia y a los ejemplos prácticos.

INDICE

NOTACION	ii
PREFACIO	iii
1 INTRODUCCION	1
1.1 MODOS DE CONVERGENCIA	1
1.2 INTEGRACION MONTE CARLO	10
1.3 MUESTREO DE SUSTITUCION SUCESIVA	14
1.4 MUESTREO DE GIBBS	17
1.5 INFERENCIA BAYESIANA	20
2 INTEGRACION VIA MUESTREO DE IMPORTANCIA	24
2.1 ESTIMACION POR INTERVALO DE LA MEDIA POSTERIOR DE UNA FUNCION	32
3 CONVERGENCIA DEL MUESTREO DE SUSTITUCION SUCESIVA	36
3.1 EL ESPACIO DE HILBERT	36
3.1.1 OPERADORES ADJUNTOS.	39
3.2 PRUEBA DE LA CONVERGENCIA DEL M.S.S.	43

4	SIMULACIONES	53
5	UN MODELO JERARQUICO.	73
5.1	INTRODUCCION	73
5.2	IMPLEMENTACION DE UN MODELO JERARQUICO	74
	CONCLUSIONES	87
	APENDICE	91
	BIBLIOGRAFIA	138

Capítulo 1

INTRODUCCION

Las técnicas de integración Monte Carlo tienen su fundamento en teoría asintótica, por tanto, se hace una introducción a los modos de convergencia a fin de justificar los resultados obtenidos posteriormente. El lector que se encuentre familiarizado con los conceptos siguientes puede omitir su lectura y comenzar con la sección 1.2, a partir de la cual se describe el objetivo principal de este trabajo; en caso contrario se recomienda leerla.

1.1 MODOS DE CONVERGENCIA

Con el objetivo de que el lector tenga una noción de las herramientas matemáticas usadas para hacer inferencias basadas en aproximaciones para muestras grandes, se definen conceptos básicos de convergencia, y también se incluyen algunos ejemplos.

Sea $\{a_n, n = 1, 2, \dots\}$ una sucesión de números reales estrictamente positivos y sea $\{X_n, n = 1, 2, \dots\}$ una sucesión de variables aleatorias todas definidas en el mismo espacio de probabilidad

Definición 1.1 (*Convergencia en Probabilidad*). Se dice que X_n converge en probabilidad a cero, escrito como

$$X_n = o_p(1) \quad \text{o} \quad X_n \xrightarrow{p} 0,$$

si para cada $\varepsilon > 0$

$$p(|\mathbf{X}_n| > \varepsilon) \rightarrow 0$$

cuando $n \rightarrow \infty$.

Definición 1.2 (*Acotamiento en probabilidad*). Se dice que la sucesión $\{\mathbf{X}_n\}$ es acotada en probabilidad, escrito como $\mathbf{X}_n = O_p(1)$, si para cada $\varepsilon > 0$ existe $\delta(\varepsilon) \in (0, \infty)$ tal que

$$p(|\mathbf{X}_n| > \delta(\varepsilon)) < \varepsilon$$

para toda n .

Definición 1.3 (*Convergencia en Probabilidad y Orden en Probabilidad*).

(a) \mathbf{X}_n converge en probabilidad a la variable aleatoria X , escrito como

$$\mathbf{X}_n \xrightarrow{p} X,$$

si y sólo si

$$\mathbf{X}_n - X = o_p(1).$$

(b) $\mathbf{X}_n = o_p(a_n)$ si y sólo si $a_n^{-1}\mathbf{X}_n = o_p(1)$.

(c) $\mathbf{X}_n = O_p(a_n)$ si y sólo si $a_n^{-1}\mathbf{X}_n = O_p(1)$.

Proposición 1.1 Si \mathbf{X}_n y $Y_n, n = 1, 2, \dots$, son variables aleatorias definidas sobre el mismo espacio de probabilidad y $a_n > 0, b_n > 0, n = 1, 2, \dots$, entonces

(a) Si $\mathbf{X}_n = o_p(a_n)$ y $Y_n = o_p(b_n)$, se tiene que

$$\mathbf{X}_n Y_n = o_p(a_n b_n),$$

$$\mathbf{X}_n + Y_n = o_p(\max(a_n, b_n)),$$

y

$$|\mathbf{X}_n|^r = o_p(a_n^r), \quad \text{para } r > 0;$$

(b) si $\mathbf{X}_n = o_p(a_n)$ y $Y_n = O_p(b_n)$, se tiene

$$\mathbf{X}_n Y_n = o_p(a_n b_n).$$

Por otra parte

(c) el primer inciso permanece válido si o_p es reemplazado por O_p .

Demostración. Ver Brockwell & Davis (1991, págs. 199,200) \square

Ejemplo 1.1 Considere una serie infinita de lanzamientos independientes de una moneda. Sea p la probabilidad de éxito y sea X_n la proporción de sucesos exitosos en los primeros n volados. Entonces

$$X_n \xrightarrow{p} p.$$

Ejemplo 1.2 Sea $\{X_n\}$ una sucesión de v.a. tales que $E[X_n] = m$ y $\text{Var}[X_n] = \sigma_n^2 > 0$ para toda n , donde $\sigma_n^2 \rightarrow 0$ cuando $n \rightarrow \infty$.

Sea $Z_n = \sigma_n^{-1}(X_n - m)$ entonces $Z_n = O_p(1)$.

Definición 1.4 (Orden en Probabilidad para vectores aleatorios).

(a) $\mathbf{X}_n = o_p(a_n)$ si y sólo si $X_{nj} = o_p(a_n), j = 1, \dots, k$.

(b) $\mathbf{X}_n = O_p(a_n)$ si y sólo si $X_{nj} = O_p(a_n), j = 1, \dots, k$.

(c) \mathbf{X}_n converge en probabilidad a el vector aleatorio \mathbf{X} , escrito como

$$\mathbf{X}_n \xrightarrow{p} \mathbf{X}.$$

si y sólo si $\mathbf{X}_n - \mathbf{X} = o_p(1)$.

La convergencia en probabilidad de \mathbf{X}_n a \mathbf{X} puede también caracterizarse en términos de la distancia euclidiana

$$|\mathbf{X}_n - \mathbf{X}| = \left[\sum_{j=1}^k (X_{nj} - X_j)^2 \right]^{\frac{1}{2}}.$$

Como puede verse en la siguiente Proposición.

Proposición 1.2 $\mathbf{X}_n - \mathbf{X} = o_p(1)$ si y sólo si $|\mathbf{X}_n - \mathbf{X}| = o_p(1)$.

Demostración. Ver Brockwell & Davis (1991, págs. 200) \square

Definición 1.5 (Convergencia en Media r -ésima, $r > 0$). La sucesión de variables aleatorias $\{\mathbf{X}_n\}$ se dice que converge en media r -ésima a X , escrito como

$$\mathbf{X}_n \xrightarrow{r} X,$$

si

$$E|\mathbf{X}_n - X|^r \rightarrow 0$$

cuando $n \rightarrow \infty$.

Definición 1.6 (Convergencia en Distribución). La sucesión $\{X_n\}$ de k -vectores aleatorios con funciones de distribución $\{F_{X_n}(\cdot)\}$ se dice que converge en distribución si existe un k -vector aleatorio X tal que

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x) \quad (1.1)$$

para toda $x \in C$, donde C es el conjunto de puntos de continuidad de la función de distribución $F_X(\cdot)$ de X . Si (1.1) se cumple puede decirse que \mathbf{X}_n converge en distribución a X . Esta convergencia será denotada por

$$\mathbf{X}_n \Longrightarrow X$$

o

$$F_{\mathbf{X}_n} \Longrightarrow F_X.$$

Proposición 1.3 Si $\mathbf{X}_n \xrightarrow{r} X$ entonces

(a) $E[\exp(it'X_n) - \exp(it'X)] \rightarrow 0$ cuando $n \rightarrow \infty$ para cada $t \in \mathbb{R}^k$

y

(b) $X_n \Rightarrow X$.

Demostración. Ver Brockwell & Davis (1991, pág. 205) \square

Teorema 1.1. (*Ley Débil de los Grandes Números*). Si $\{X_n\}$ es una sucesión de variables aleatorias i.i.d. con una media finita μ , entonces

$$\bar{X}_n \xrightarrow{P} \mu$$

donde

$$\bar{X}_n \equiv (X_1 + \dots + X_n) / n. \quad (1.2)$$

Demostración. Ver Brockwell & Davis (1991, págs. 206, 207) \square

Teorema 1.2 (*Ley Fuerte de los Grandes Números*). Si $\{X_n\}$ es una sucesión de variables aleatorias i.i.d. y $E[X_j] = \mu, j = 1, \dots, n$, entonces

$$P \left[\lim_n \bar{X}_n = \mu \right] = 1,$$

denotada por:

$$\bar{X}_n \xrightarrow{c.s.} \mu.$$

Demostración. Ver Billingsley (1979, pág. 70) \square

Ejemplo 1.3 Mediante un ejemplo simple se demuestra que la convergencia en probabilidad no implica la convergencia casi segura.

Sean las variables $Z_{ij} : \Omega \rightarrow R, i = 1, 2, \dots, j = 1, \dots, i$; con función de densidad

$$Z_{ij}(w) = \begin{cases} 1 & \text{si } \frac{i-1}{i} < w \leq \frac{i}{i} \\ 0 & \text{en otro caso} \end{cases}.$$

Se denota una nueva sucesión por

$$X_n(w) = Z_{ij}(w)$$

con

$$n = \frac{i(i-1)}{2} + j.$$

Sea $X(w) = 0$, $w \in [0, 1]$.

Como

$$P[|X_n - X| \geq \varepsilon] = P[X_n \geq \varepsilon]$$

X_n toma sólo valores 0 ó 1.

Si $\varepsilon > 1$

$$P[|X_n - X| \geq \varepsilon] = P[\emptyset] = 0$$

entonces

$$\lim_{n \rightarrow \infty} P[|X_n - X| \geq \varepsilon] = 0,$$

si $0 < \varepsilon \leq 1$

$$\begin{aligned} P[|X_n - X| \geq \varepsilon] &= P[X_n \geq \varepsilon] \\ &= P[X_n = 1] \\ &= P[Z_{ij} = 1] \\ &= \frac{1}{j} \quad j \leq i. \end{aligned}$$

Como

$$n < \frac{i(i-1)}{2} + \frac{3}{2}i + \frac{1}{2} = \frac{(i+1)^2}{2},$$

despejando i

$$i > \sqrt{2n} - 1.$$

entonces

$$P[|X_n - X| \geq \varepsilon] < \frac{1}{\sqrt{2n} - 1} \rightarrow 0$$

cuando $n \rightarrow \infty$; de donde se sigue que $\varepsilon > 0$.

$$\lim_{n \rightarrow \infty} P[|X_n - X| \geq \varepsilon] = 0.$$

Ahora se demuestra que

$$\lim_{n \rightarrow \infty} X_n(w) \neq X(w).$$

Para cada $w \in (0, 1)$ y $M \in \mathbb{N}$ existe $n > M$ tal que $X_n(w) = 1$, n es de la forma

$$n = \frac{i(i-1)}{2} + j$$

donde $0 \leq j < i+1$ y $M < \frac{i(i-1)}{2}$. Se tiene que

$$\frac{j-1}{i+1} < w \leq \frac{j}{i+1}.$$

Entonces

$$X_n(w) = 1$$

y por lo tanto

$$\lim_{n \rightarrow \infty} X_n(w) = 1 \neq X(w).$$

Proposición 1.4 Sea $X_n, n = 1, 2, \dots$ y $Y_{n,j}, j = 1, 2, \dots; n = 1, 2, \dots$, k -variables aleatorias tal que

- (a) $Y_{n,j} \Rightarrow Y_j$ cuando $n \rightarrow \infty$ para cada $j = 1, 2, \dots$.
- (b) $Y_j \Rightarrow Y$ cuando $j \rightarrow \infty$. y
- (c)

$$\lim_{j \rightarrow \infty} \limsup_{n \rightarrow \infty} p(|X_n - Y_{n,j}| > \varepsilon) = 0$$

para cada $\varepsilon > 0$, entonces

$$X_n \Rightarrow Y \text{ cuando } n \rightarrow \infty.$$

Demostración. Ver Brockwell & Davis (1991, págs. 207, 208). \square

Definición 1.7 Una sucesión de variables aleatorias $\{X_n\}$ se dice que es asintóticamente normal con "media" μ_n y "desviación estándar" σ_n si $\sigma_n > 0$ para n suficientemente grande y

$$\sigma_n^{-1}(X_n - \mu_n) \Rightarrow Z.$$

donde $Z \sim N(0, 1)$.

La Definición anterior puede escribirse como

$$X_n \text{ es } AN(\mu_n, \sigma_n^2).$$

Observación 1.1 Si X_n es $AN(\mu_n, \sigma_n^2)$ no es necesariamente el caso que $\mu_n = E[X_n]$ o que $\sigma_n^2 = Var(X_n)$.

Teorema 1.3 (Teorema del Límite Central). Si $\{X_n\} \sim \text{IID}(\mu, \sigma^2)$ y $\bar{X}_n = (X_1 + \dots + X_n)/n$, entonces

$$\bar{X}_n \text{ es } AN(\mu, \sigma^2/n).$$

Demostración. Ver Brockwell & Davis (1991, pág. 210). \square

Proposición 1.5 Si X_n es $AN(\mu, \sigma_n^2)$, donde $\sigma_n \rightarrow 0$ cuando $n \rightarrow \infty$, y si g es una función diferenciable en μ , entonces

$$g(X_n) \text{ es } AN(g(\mu), g'(\mu)^2 \sigma_n^2).$$

Demostración. Ver Brockwell & Davis (1991, pág. 210). \square

Ejemplo 1.4 Suponga que $\{X_n\} \sim \text{I.I.D.}(\mu, \sigma^2)$ donde $\mu \neq 0$ y $0 < \sigma^2 < \infty$. Si

$$\bar{X}_n = (X_1 + \dots + X_n)/n$$

entonces por el Teorema 1.3

$$\bar{X}_n \text{ es } AN(\mu, \sigma^2/n),$$

y por la Proposición anterior,

$$\bar{X}_n^{-1} \text{ es } AN(\mu^{-1}, \mu^{-4} \sigma^2/n).$$

Definición 1.8 La sucesión $\{X_n\}$ de k -vectores aleatorios es asintóticamente normal con "vector de medias" μ_n y "matriz de covarianzas" Σ_n si

(a) Los elementos de la diagonal de Σ_n no son cero para una n suficientemente grande.

y

(b) $\lambda'X_n$ es $AN(\lambda'\mu_n, \lambda'\Sigma_n\lambda)$ para cada $\lambda \in \mathbb{R}^k$ tal que $\lambda'\Sigma_n\lambda > 0$ para toda n suficientemente grande.

La Definición anterior puede escribirse como $X_n \sim AN(\mu_n, \Sigma_n)$.

Proposición 1.6 Suponga que X_n es $AN(\mu, c_n^2 \Sigma)$ donde Σ es una matriz simétrica no negativa y $c_n \rightarrow 0$ cuando $n \rightarrow \infty$. Si $g(\mathbf{X}) = (g(\mathbf{X}_1), \dots, g(\mathbf{X}_m))'$ es un mapeo de \mathbb{R}^k a \mathbb{R}^m tal que $g_i(\bullet)$ es continuamente diferenciable en una vecindad de μ , y si $D\Sigma D'$ tiene en toda su diagonal elementos no cero, donde D es la matriz $[(\partial g_i / \partial x_j)(\mu)]$ de $m \times k$, entonces

$$g(X_n) \text{ es } AN(g(\mu), c_n^2 D \Sigma D').$$

Demostración. Ver Brockwell & Davis (1991, págs. 211,212). \square

Ejemplo 1.5 (El coeficiente de variación muestral). Suponga que $\{X_n\} \sim \text{IID}(\mu, \sigma^2)$, $\sigma > 0$,

$$E[X_n^4] = \mu_4,$$

$$E[X_n^3] = \mu_3,$$

$$E[X_n^2] = \mu^2 + \sigma^2.$$

y

$$E[X_n] = \mu_1 \neq 0$$

El coeficiente de variación se define como $Y_n = S_n / \bar{X}_n$ con \bar{X}_n como se definió en (1.2) y

$$S_n^2 = n^{-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

Es fácil verificar que

$$\begin{bmatrix} \bar{X}_n \\ n^{-1} \sum_{i=1}^n X_i^2 \end{bmatrix} \text{ es } AN \left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, n^{-1} \Sigma \right)$$

donde Σ es la matriz con componentes

$$\Sigma_{ij} = \mu_{i+j} - \mu_i \mu_j, \quad i, j = 1, 2.$$

Ahora se define $Y_n = g(\bar{X}_n, n^{-1} \sum_{i=1}^n X_i^2)$ donde $g(u, b) = a^{-1}(b - a^2)^{1/2}$. Aplicando la Proposición 1.6, con

$$D = \left[\frac{\partial g}{\partial x}(\mu), \frac{\partial g}{\partial y}(\mu) \right] = \left[-\mu_2 / (\sigma \mu_1^2), 1 / (2\sigma \mu_1) \right],$$

se tiene que

$$Y_n \text{ es } AN\left(\sigma / \mu_1, n^{-1} D \Sigma D'\right).$$

Teorema 1.4 Considere el vector aleatorio $X = (X_1, X_2, \dots, X_d)$, donde X_i tiene dimensión $p_i, i = 1, 2, \dots, d$. Suponga también que X tiene densidades marginales $f_{X_i}(x_i), X_i \in \Theta_i, i = 1, 2, \dots, d$, y la densidad conjunta el producto de espacios $\Theta = \Theta_1 \times \Theta_2 \times \dots \times \Theta_d$ como dominio. Entonces las distribuciones condicionales completas determinan únicamente la distribución conjunta $f_X(x)$.

Demostración. Considere un vector arbitrario $X_1^0, X_2^0, \dots, X_d^0$ en Θ . La densidad conjunta puede escribirse en términos de sus condicionales como

$$f(X_1, X_2, \dots, X_d) \propto \frac{f(X_1 | X_2^0, \dots, X_d^0) f(X_2 | X_1, X_3^0, \dots, X_d^0) \dots f(X_d | X_1, \dots, X_{d-1})}{f(X_2^0 | X_1, X_3^0, \dots, X_d^0) f(X_3^0 | X_1, X_2, \dots, X_d^0) \dots f(X_d^0 | X_1, \dots, X_{d-1})}.$$

lo cual prueba el resultado. \square

1.2 INTEGRACION MONTE CARLO

Las herramientas numéricas para evaluar integrales han tenido un gran desarrollo ya que en muchos casos la evaluación en forma directa presenta dificultades. Uno de los ejemplos más simples podría ser el cálculo de la siguiente integral

$$\int_0^{1.96} e^{-\frac{x^2}{2}} dx, \quad (1.3)$$

cuyo valor exacto es $\frac{0.95}{2}\sqrt{2\pi}$.

Con base en este tipo de problemas se han desarrollado diferentes métodos de integración, uno de ellos es el método Monte Carlo.

La integración con el método Monte Carlo se apoya en la Ley Fuerte de los Grandes Números (convergencia en probabilidad y casi segura) y consiste en que la integral de una función puede expresarse como un valor esperado con respecto a una densidad de probabilidad, lo cual sugiere una enfoque estadístico al problema de integración. La esperanza es estimada entonces como el promedio de los valores que toma la función; esto se representa como

$$\begin{aligned} \int_X h(x) dx &= \int_X \frac{h(x)}{f_X} f_X dx \\ &= E \left[\frac{h(X)}{f_X} \right] \\ &\stackrel{e.s.}{\approx} \frac{1}{n} \sum_{i=0}^n \frac{h(X_i)}{f_X} \end{aligned} \quad (1.4)$$

donde X_1, X_2, \dots, X_n son v.a. i.i.d. con densidad f_X .

De acuerdo a este método se puede hacer una estimación del valor real de la integral anterior (1.3) como sigue:

$$\begin{aligned} \int_0^{1.96} e^{-\frac{x^2}{2}} dx &= \int_{-\infty}^{\infty} I_{(0,1.96)}(x) e^{-x^2/2} \frac{\sqrt{2\pi}}{\sqrt{2\pi}} dx \\ &= E \left[I_{(0,1.96)}(x) \sqrt{2\pi} \right] \\ &\stackrel{e.s.}{\approx} \frac{1}{n} \sum_{i=0}^n I_{(0,1.96)}(X_i) \sqrt{2\pi}, \end{aligned} \quad (1.5)$$

donde X_1, X_2, \dots, X_n son v.a. i.i.d. $N(0, 1)$ y así obtener la esperanza de la función original aproximada. En la tabla 1.1 se presentan los valores aproximados de la integral (1.3) para distintos tamaños de muestra.

Puede observarse que los valores estimados se aproximan al verdadero valor que es $\frac{0.95}{2}\sqrt{2\pi}$; los valores fueron generados bajo los comandos de simulación del lenguaje *S-PLUS* y cuyo listado se encuentra en el apéndice bajo el nombre de "Introduc".

Otro ejemplo puede ser el cálculo de la esperanza de una distribución normal bivariada, con parámetros (μ, Σ) ; donde la integral a evaluar tiene más de una variable. Con los

Tamaño	Estimación
10	1.2533141373155
100	1.27838042006181
1000	1.17059540425268
10000	1.18914445348495
15000	1.18630360810703
20000	1.19503502993033
30000	1.19073198472555
50000	1.18959564657438
70000	1.19283277794619
100000	1.19333052270358

TABLA 1.1: Valores estimados de la integral (1.3), con diferentes tamaños de muestra, bajo el método Monte Carlo de integración

parámetros anteriores el valor real de la esperanza puede calcularse como

$$E(XY) = \text{cov}(X, Y) + E(X)E(Y) \quad (1.6)$$

y estimada, (en caso de que sean desconocidos los parámetros) por la integración Monte Carlo a través de

$$E(XY) \approx \frac{1}{n} \sum_{i=1}^n X_i Y_i$$

donde $(X_i, Y_i), i = 1, \dots, n$, son v.a. i.i.d. $N_2(\mu, \Sigma)$.

En la tabla 1.2 se presenta estimaciones de (1.6) para distintos tamaños de muestra de una normal bivariada con matriz de varianzas-covarianzas

$$\Sigma = \begin{pmatrix} 1 & .5 \\ .5 & 1 \end{pmatrix} \quad (1.7)$$

y vector de medias

$$\mu^t = (12, 12) \quad (1.8)$$

El valor real de $E(XY)$ es 144.5 y puede verse que a medida que se incrementa el tamaño de muestra, el valor estimado tiene una mejor aproximación. Estas estimaciones fueron calculadas por generación de variables aleatorias bajo el lenguaje de programación *S-PLUS* cuyo código tiene el nombre de "introduc".

Una de las aplicaciones de este método es la posibilidad de calcular funciones de densidad marginales, cuando la densidad conjunta es difícil de simular en forma directa pero es

Tamaño	Estimación
10	135.514314758913
100	143.358272718973
1000	144.695187794262
10000	144.418256454752
15000	144.559766260599
20000	144.527972930245
30000	144.527972930245
50000	144.530863614139

TABLA 1.2: Valores estimados de $E(XY)$ bajo el método de integración Monte Carlo para diferentes tamaños de muestra.

posible generar de las densidades condicionales. Ahora la densidad conjunta se puede expresar como

$$f_{X,Y}(x,y) = f_{X|Y}(x|y) f_Y(y), \quad (1.9)$$

y obtener los valores $Y \sim f_Y(y)$ y $X \sim f_{X|Y}(x|y)$ de las densidades marginal y condicional. Entonces los valores generados (X, Y) tienen distribución conjunta $f_{X,Y}(x, y)$. Si X y Y son variables aleatorias discretas

$$\begin{aligned} P[X = x, Y = y] &= P[Y = y] P[X = x|Y = y] \\ &= f_Y(y) f_{X|Y}(x|y) \\ &= f_{X,Y}(x, y). \end{aligned}$$

Ahora, aplicando la integración Monte Carlo:

$$\begin{aligned} f_X(x) &= \int f_{X,Y}(x, y) dy \\ &= \int \frac{f_{X,Y}(x, y)}{f_Y(y)} f_Y(y) dy \\ &= \int f_{X|Y}(x|y) f_Y(y) dy \\ &= E[f_{X|Y}(x|y)] \\ &\stackrel{c.a.}{\approx} \frac{1}{n} \sum_{i=1}^n f_{X|Y}(x|y_i). \end{aligned} \quad (1.10)$$

La figura 1.1 presenta las densidades marginales estimadas para distintos tamaños de muestra, basadas en la misma normal bivariada de la página anterior, y se comparan con la

función de densidad verdadera (líneas continuas).

Puede verse que a medida que se incrementa el tamaño de muestra, hay una mejor convergencia de la función de densidad estimada hacia la verdadera, se tiene una mejor aproximación en base al número de elementos muestreados.

Los valores de la densidad estimada se obtuvieron mediante el lenguaje de programación *S-PLUS* cuyas instrucciones se encuentran en el listado "introduc".

¿Pero qué pasa cuando no se cuenta con un valor inicial para generar la muestra, esto es, que sea imposible obtener valores de f_Y para aplicarlos en la distribución condicional?. En este caso se da un valor inicial arbitrario Y_0 para obtener X de f_X , como consecuencia la pareja (Y_0, X) no tiene distribución conjunta $f_{X,Y}$, o en el caso general, cuando se cuenta con distribuciones multivariadas en donde se dificulta la generación de muestras de variables aleatorias directamente de las densidades condicionales.

El método Monte Carlo no especifica la forma de obtener los valores para el estimador de la esperanza. Cuando se cuenta con este inconveniente existen algoritmos que proporcionan la facilidad de generar muestras de distribuciones multivariadas y así aplicar la integración Monte Carlo en la estimación del valor de una integral o de una función de distribución. Estos algoritmos se describen en la siguiente sección.

1.3 MUESTREO DE SUSTITUCION SUCESIVA

Siguiendo con el problema presentado en la sección anterior, se procede a describir brevemente el *Muestreo de Sustitución Sucesiva (M.S.S.)*, basado en el descrito por Gelfand and Smith (1990); comenzando con la notación usada por Schervish and Bradley (1992) donde el proceso de marginalización se denota por

$$f_{X|Y}(x|y) = \int f_{X|Y,Z,W}(x|y,z,w) f_{Z|W,Y}(z|w,y) f_{W|Y}(w|y) dz dw.$$

Se suponen 2 variables aleatorias X, Y con densidades condicionales disponibles (donde disponibles significa que pueden ser directa y eficientemente generados dados valores específicos de las variables condicionantes), denotadas como

$$f_{X|Y}(x|y)$$

y

$$f_{Y|X}(y|x),$$

que pueden generarse por algún método obteniendo las marginales como sigue:

$$f_X(x) = \int_R f_{X|Y}(x|y)f_Y(y)dy \quad (1.11)$$

y

$$f_Y(y) = \int_R f_{Y|X}(y|x)f_X(x)dx. \quad (1.12)$$

Si ahora se sustituye la ecuación (1.12) en (1.11) resulta

$$\begin{aligned} f_X(x) &= \int_R f_{X|Y}(x|y) \int_R f_{Y|X}(y|x)f_X(x)dx dy \\ &= \int_R h(X, Y) f_X(x)dx \end{aligned} \quad (1.13)$$

donde

$$h(X, Y) = \int_R f_{X|Y}(x|y)f_{Y|X}(y|x)dy.$$

El proceso anterior genera muestras de estas densidades marginales como sigue:

Teniendo un valor inicial $X^{(0)}$ con densidad $f_X(x_0)$, dado que $f_{Y|X}$ está disponible se puede generar $Y^{(1)} \sim f_{Y|X}(y|x_0)$: y la densidad marginal es

$$f_{Y_1}(y) = \int f_{Y|X}(y|x)f_X(x_0)dx.$$

Ahora para completar un ciclo generar $X^{(1)} \sim f_{X|Y}(x|y^{(1)})$, de donde

$$f_{X_1}(x) = \int f_{X|Y}(x|y)f_{Y_1}(y)dy = \int h(X, Y) f_X(x_0)dx.$$

Repetiendo este ciclo produce $Y^{(2)}$ y $X^{(2)}$, y eventualmente, después de i -iteraciones pueden generarse m pares idénticamente distribuidos $(X_j^{(i)}, Y_j^{(i)})$, $j = 1, \dots, m$.

A este esquema de generación se le llama Muestreo de Sustitución Sucesiva. Gelfand and Smith (1990).

Si se terminan todas las repeticiones en la i -ésima iteración, la densidad estimada de X (o en forma análoga la de Y) es la aproximación Monte Carlo de (1.11) dada por

$$f_{\hat{X}}(x) = \frac{1}{m} \sum_{j=1}^m f_{X|Y_j}(x|y_j^{(i)}). \quad (1.14)$$

Una generalización para el caso de tres variables X, Y, Z , suponiendo que las siguientes distribuciones condicionales están disponibles, está dada por

$$f_X(x) = \int_{Y,Z} f_{X,Z|Y}(x, z|y) f_Y(y) dy dz$$

$$f_Y(y) = \int_{X,Z} f_{Y,X|Z}(y, x|z) f_Z(z) dz dx$$

$$f_Z(z) = \int_{X,Y} f_{Z,Y|X}(z, y|x) f_X(x) dx dy.$$

Tomando un valor arbitrario inicial para $X^{(0)}$, el algoritmo básico es el siguiente

- (a) generar $(Z^{(0)'}, Y^{(0)'})$ de $f_{Z,Y|X}(z, y|x^{(0)})$
- (b) generar $(Y^{(1)}, X^{(0)'})$ de $f_{Y,X|Z}(y, x|z^{(0)'})$
- (c) y finalmente generar $(X^{(1)}, Z^{(1)})$ de $f_{X,Z|Y}(x, z|y^{(1)})$.

Un ciclo completo del algoritmo (i.e. para generar $X^{(1)}$ comenzando de $X^{(0)}$) requiere de seis variables generadas: más de dos como se vió para el caso de dos variables X, Y . Repitiendo i -veces el procedimiento, en forma similar a los tres incisos anteriores produce el vector $(X^{(i)}, Y^{(i)}, Z^{(i)})$, y el proceso entero después de m -veces se obtienen los vectores $(X_j^{(i)}, Y_j^{(i)}, Z_j^{(i)})$, $j = 1, \dots, m$.

Cabe hacer notar que el *M.S.S.* no requiere de la especificación de la densidad conjunta, sino sólo de la disponibilidad de las densidades condicionales, y por ejemplo, el generar variables aleatorias con densidad $f_{Z,Y|X}$ requiere de $f_{Z|Y,X}$ y $f_{Y|X}$, donde las densidades marginal y condicional reducida sean de fácil evaluación.

En forma paralela, de acuerdo con la ecuación (1.14), se puede estimar la densidad marginal para X , (en forma análoga para Y, Z) como

$$\hat{f}_X(x) = \frac{1}{m} \sum_{j=1}^m f_{X|Y,Z}(x|y_j^{(i)}, z_j^{(i)}).$$

Para el caso de k -variables X_1, \dots, X_k , el algoritmo de *M.S.S.* requiere de $k(k-1)$ generaciones de variables aleatorias para completar un ciclo. Después de i -repeticiones corriendo en paralelo m -veces ($mik(k-1)$ generaciones aleatorias) se obtienen los vectores

aleatorios $(X_{1j}^{(i)}, \dots, X_{kj}^{(i)})$, $j = 1, \dots, m$, con el estimador de densidad para X_s dado por la aproximación Monte Carlo

$$\hat{f}_{X_s}(x_s) = \frac{1}{m} \sum_{j=1}^m f_{X_s | \{X_{ij}, i \neq s\}}(x_s | \{x_{ij}^{(i)}, i \neq s\}). \quad (1.15)$$

El mayor problema para los usuarios del *M.S.S.* es decidir cuándo "converge" el algoritmo, esto es, cuántas iteraciones son necesarias para obtener una buena aproximación al muestreo de la densidad de interés.

1.4 MUESTREO DE GIBBS

El muestreo de Gibbs (*Gibbs Sampler*, Geman and Geman (1984)), es una técnica muy parecida al *M.S.S.* para generación de variables aleatorias de una densidad conjunta simulando de las densidades condicionales completas, donde la densidad completa es la densidad condicional de una variable dada todas las restantes. El Muestreo Gibbs se basa en algunas propiedades elementales de las Cadenas de Markov suponiendo que las densidades condicionales completas para cada variable están disponibles o son de fácil evaluación, y además, que determinen (módulo la constante de normalización) a la densidad conjunta. lo anterior se puede verificar con el Teorema 1.4.

El algoritmo básico del Muestreo Gibbs con tamaño de muestra fijo se da a continuación:

Algoritmo básico de Muestreo Gibbs

(Tamaño de muestra fijo)

1. **Generación de valores iniciales.** Seleccionar un vector inicial $X^{(0)} = (X_1^{(0)}, \dots, X_k^{(0)})$, $X_i \in \Theta_i$, en el espacio $\Theta = \Theta_1 \times \dots \times \Theta_k$. Sea $t=0$ y $j=1$.
2. **Muestreo de las Distribuciones Condicionales Completas.**
 - (a) Muestrear $X_{(1,j)}^{(t+1)}$ de $f(x_{(1,j)} | x_{(2,j)}^{(t)}, \dots, x_{(k,j)}^{(t)})$
 - (b) Muestrear $X_{(2,j)}^{(t+1)}$ de $f(x_{(2,j)} | x_{(1,j)}^{(t+1)}, x_{(3,j)}^{(t)}, \dots, x_{(k,j)}^{(t)})$
 - ⋮

(c) Muestrear $X_{(k,j)}^{(t+1)}$ de $f(x_{(k,j)} | x_{(1,j)}^{(t+1)}, \dots, x_{(k-1,j)}^{(t+1)})$, para tener el vector $X^{(t+1)} = (X_{(1,j)}^{(t+1)}, X_{(2,j)}^{(t+1)}, \dots, X_{(k,j)}^{(t+1)})$.

3. **Iteración para la convergencia del proceso de Muestreo.** Sea $t=t+1$ y repetir el paso 2 hasta que $t=i$, donde i es un número específico de iteraciones.

4. **Repetición del algoritmo m veces.** Sea $j=j+1$ y repetir el paso 3 hasta que $j=m$, donde m es el tamaño de muestra para obtener k -tuplas independientes de

$$X_j^{(i)} = (X_{(1,j)}^{(i)}, \dots, X_{(k,j)}^{(i)}), j = 1, \dots, m \quad (1.16)$$

después de i -iteraciones.

Estas variables aleatorias pueden ahora ser usadas para estimar integrales, como son la densidad marginal de X por el método de integración Monte Carlo similarmente a la ecuación (1.15):

$$\hat{f}_X(x) = \frac{1}{m} \sum_{j=1}^m f_{X_s | \{X_{(r,j)}, r \neq s\}}(x_s | \{x_{(r,j)}^{(i)}, r \neq s\}).$$

Bajo ciertas condiciones, se puede decir que el vector (1.16) converge en distribución a una observación aleatoria de f_{X_1, \dots, X_k} cuando $i \rightarrow \infty$ (ver Capítulo 3 donde se prueba la convergencia).

El algoritmo anterior produce vectores aleatorios como en la ecuación (1.16) con densidad aproximada $f_X(x)$. Se nota que los vectores $X_j^{(1)}, \dots, X_j^{(i)}$, para cualquier X_j , provienen de una cadena de Markov con función de transición de X'_j a X_j dada por

$$K(X_j, X'_j) = f_{X_1 | X'_2, \dots, X'_k_j}(x_1 | x'_2, \dots, x'_k_j) f_{X_2 | X_1, X'_3, \dots, X'_k_j}(x_2 | x_1, x'_3, \dots, x'_k_j) \cdots f_{X_{k_j} | X_1, \dots, X_{k-1,j}}(x_{k_j} | x_1, \dots, x_{k-1,j}). \quad (1.17)$$

Hasta este momento se ha contemplado un tamaño de muestra fijo, pero existe otro algoritmo que maneja el tamaño de muestra variable. Dicho algoritmo proporciona la posibilidad de determinar después de cierto número de iteraciones, si el proceso se vuelve estacionario o la influencia del valor inicial ha sido superada, y con esto decidir si debe aumentarse el tamaño de muestra o mantenerlo constante.

Si se comienza con valores $X^{(0)}$ generados por la densidad $f^{(0)}(X)$, la densidad de $X^{(t)}$ puede obtenerse recursivamente usando la transformación:

$$f^{(t+1)}(X) = T(f^{(t)}(X)) = \int K(x, x') f^{(t)}(x') dx'. \quad (1.18)$$

Que en el caso de tener la verdadera densidad $f_X(x)$, cumple con ser un punto fijo para la transformación T , esto quiere decir que

$$T(f_X) = f_X.$$

Es difícil encontrarse que la integral (1.18) sea de fácil evaluación y entonces se aproxima por medio de la integración Monte Carlo.

Algoritmo de Muestreo Gibbs (Tamaño de Muestra Variable)

- 1. Generación de valores iniciales.** Obtener una muestra aleatoria de tamaño m_0 de una aproximación inicial a $f_X(x)$ denotada por $\hat{f}^{(0)}(x)$. Sea $X_j^{(0)} = (X_{1j}^{(0)}, X_{2j}^{(0)}, \dots, X_{k_j}^{(0)})$, $j = 1, \dots, m_0$, la muestra. Sea $t=0$ y escoger el nuevo tamaño de muestra m_1 para la siguiente iteración.

Muestreo de la aproximación a la verdadera f_X . Usando la muestra aleatoria

$X_j^{(t)}$, $j = 1, \dots, m_t$, la iteración Gibbs ideal sería

$$f^{(t+1)}(X) = \int K(x, x') f^{(t)}(x') dx',$$

la cual es aproximada, usando la función $K(\bullet, \bullet)$ definida en (1.17), por

$$\hat{f}^{(t+1)}(X) = \frac{1}{m_{t+1}} \sum_{j=1}^{m_{t+1}} K(X, X_j^{(t)}).$$

Generar $X_j^{(t+1)} = (X_{1j}^{(t+1)}, X_{2j}^{(t+1)}, \dots, X_{k_j}^{(t+1)})$, $j = 1, \dots, m_{t+1}$ de $\hat{f}^{(t+1)}(X)$. Cabe señalar que un vector aleatorio X con distribución $\hat{f}^{(t+1)}(X)$ puede obtenerse generando primero un número aleatorio R teniendo una distribución discreta uniforme sobre los enteros $1, 2, \dots, m_t$, y entonces (condicional sobre $X_R^{(t)}$) usando las distribuciones condicionales como en (1.17), para generar el vector X . Sea $t=t+1$ y el próximo tamaño de muestra m_{t+1} .

Iteración a la Convergencia del Proceso de Muestreo. Repetir el paso 2, hasta $t=i$ para formar la muestra final $X_j^{(i)} = (X_{1j}^{(i)}, X_{2j}^{(i)}, \dots, X_{kj}^{(i)})$, $j = 1, 2, \dots, m_t$.

Ambos algoritmos (Gibbs y Sustitución Sucesiva) son equivalentes para el caso bivariado ($k = 2$). Para el caso general Gibbs supone la disponibilidad de k distribuciones condicionales completas (el conjunto mínimo para determinar la densidad conjunta), en tanto que el *M.S.S.* requiere de $k(k-1)$ distribuciones condicionales, incluyendo las distribuciones condicionales completas. Por tanto el *M.S.S.* presenta una acelerada convergencia (es más eficiente), ya que muestrea de las distribuciones correctas cada vez, mientras que el Muestreo de Gibbs muestrea únicamente de las distribuciones condicionales completas. Gelfand & Smith (1990). Debe notarse, sin embargo, que en la práctica difícilmente se cuenta con todas las distribuciones condicionales requeridas por el *M.S.S.*

Con base en estos algoritmos y el operador T de transición definido en (1.18), en el Capítulo 3 se deriva una demostración basada en la demostración hecha por Geweke (1989), donde se prueba la convergencia del *M.S.S.* y se discuten condiciones para la convergencia débil y fuerte.

1.5 INFERENCIA BAYESIANA

Retomando y resumiendo la idea básica de la integración Monte Carlo descrito en la sección 1.2, se desea estimar el valor de una integral cuya solución numérica es difícil de obtener

$$\int_{\Theta} f(\theta) d\theta \quad (1.19)$$

Claramente (1.19) puede escribirse como

$$\int_{\Theta} \frac{f(\theta)}{MI(\theta)} \cdot MI(\theta) d\theta \quad (1.20)$$

donde $MI(\theta)$ es una función de densidad de probabilidad sobre Θ . En otras palabras (1.20) es

$$E \left[\frac{f(\theta)}{MI(\theta)} \right]$$

y a la densidad $MI(\theta)$ se el conoce como la densidad de Muestreo de Importancia.

Entonces para una integral dada existe una infinidad de estimadores insesgados con precisiones distintas. Un aspecto que contempla la integración Monte Carlo es el diseño de técnicas de reducción de varianza para dichos estimadores, una de las cuales se refiere a la elección de $MI(\theta)$ y que generalmente requiere satisfacer las condiciones siguientes: fácil de simular, forma similar a la función que se desea integrar, y colas más pesadas que $f(\theta)$ (el soporte de $MI(\theta)$ debe incluir al de $f(\theta)$ pues de otra forma la varianza del estimador sería muy grande o infinita).

En la Estadística, la integración Monte Carlo tiene una importante aplicación dentro de la teoría Bayesiana, en donde es más socorrido para el cálculo de integrales de una función con respecto a una función de densidad de probabilidad posterior, y a continuación se procede a hacer una breve introducción a la Inferencia Bayesiana.

En muchos de los casos existe una función de densidad de probabilidad $f_{X|\Theta}(x|\theta)$ basada en un vector de parámetros θ desconocido el cual especifica la densidad de probabilidad de X . La inferencia Bayesiana procede de la función de verosimilitud asociada a $f_{X|\Theta}(x|\theta)$, ($L(\theta)$) que ahora depende de los parámetros condicionantes, y de la función de densidad a priori $f_{\Theta}(\theta)$. La función de densidad posterior, que es la función de densidad de los parámetros desconocidos θ , es proporcional al producto de la función de verosimilitud y la función de distribución a priori.

Aplicando el Teorema de Bayes, resulta lo siguiente:

$$\begin{aligned} f_{\Theta|X}(\theta|x) &= \frac{f_{X|\Theta}(x|\theta) f_{\Theta}(\theta)}{f_X(x)} \\ &= \frac{f_{X|\Theta}(x|\theta) f_{\Theta}(\theta)}{\int f_{X|\Theta}(x|\theta) f_{\Theta}(\theta) d\theta} \\ &= \frac{f_{X|\Theta}(x|\theta) f_{\Theta}(\theta)}{\int f_{X|\Theta}(x|\theta) f_{\Theta}(\theta) d\theta}. \end{aligned}$$

Entonces puede decirse que

$$f_{\Theta|X}(\theta|x) \propto f_{X|\Theta}(x|\theta) f_{\Theta}(\theta).$$

Sea

$$p(\theta) = f_{X|\Theta}(x|\theta) f_{\Theta}(\theta),$$

y

$$c = \int f_{X|\Theta}(x|\theta) f_{\Theta}(\theta) d\theta.$$

Entonces

$$\frac{\int p(\theta) d\theta}{c} = 1.$$

Cabe señalar que muchos de los problemas de inferencia Bayesiana se reducen a la evaluación del valor esperado de alguna función de interés $g(\theta)$ con respecto a la función de distribución posterior; i.e.

$$E[g(\theta)] = \frac{\int g(\theta) p(\theta) d\theta}{\int p(\theta) d\theta}.$$

La integral anterior puede estimarse por el método Monte Carlo si se escribe en forma equivalente

$$E[g(\theta)] = \frac{\int \frac{g(\theta)p(\theta)}{MI(\theta)} MI(\theta) d\theta}{\int \frac{p(\theta)}{MI(\theta)} MI(\theta) d\theta} \quad (1.21)$$

donde $MI(\theta)$ es la función de densidad de Muestreo de Importancia. En otras palabras (1.21) se reduce a calcular

$$E[g(\theta)] = \frac{E\left[\frac{g(\theta)p(\theta)}{MI(\theta)}\right]}{E\left[\frac{p(\theta)}{MI(\theta)}\right]}.$$

Si se genera una muestra $\theta_1, \dots, \theta_n$ de $MI(\theta)$ pueden estimarse por separado las integrales de (1.21) como

$$E[g(\theta)] \stackrel{c.s.}{\leftarrow} \frac{\sum_{i=1}^n \frac{g(y_i)p(y_i)}{MI(y_i)}}{\sum_{i=1}^n \frac{p(y_i)}{MI(y_i)}} = \bar{g}_n. \quad (1.22)$$

Cada una de las esperanzas anteriores convergen en forma casi segura a la integral (1.21) de acuerdo con los resultados obtenidos en la primera sección de modos de convergencia.

La precisión de \bar{g}_n depende tanto del tamaño de muestra, como de la densidad MI , además de ser invariante a escalamientos de p y MI .

Ya que \bar{g}_n converge en forma casi segura a $E[g(\theta)]$, en el siguiente Capítulo, bajo algunos supuestos más adelante se demuestra que

$$\sqrt{n} \{\bar{g}_n - E[g(\theta)]\} \implies N(0, \sigma^2)$$

donde σ^2 puede ser estimado consistentemente. Además se muestra cómo calcular el tamaño de muestra necesario para una longitud deseada del intervalo de confianza estimado para \bar{g}_n .

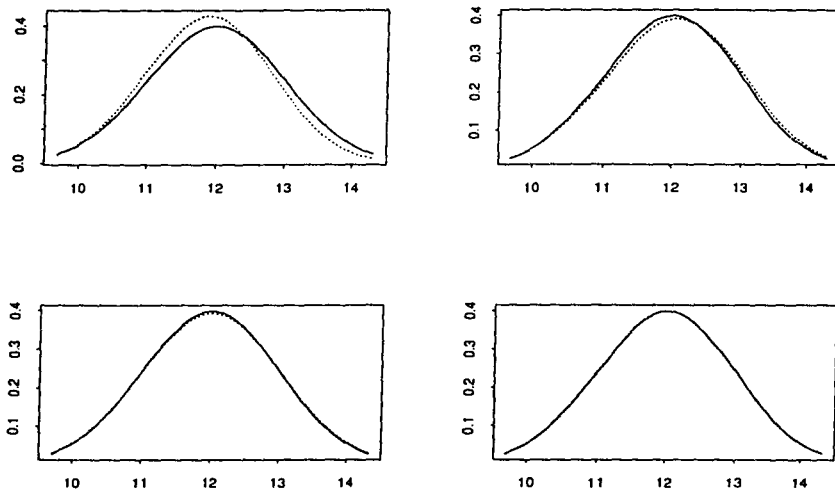


FIGURA 1.1: Densidad verdadera (línea continua) contra la densidad estimada por Monte Carlo (línea punteada) con tamaños de muestra, de izquierda a derecha y de abajo hacia arriba 10, 100, 1000 y 10000 respectivamente.

Capítulo 2

INTEGRACION VIA MUESTREO DE IMPORTANCIA

En este capítulo se discute una forma de aproximar el valor esperado de una función $g(\theta)$ respecto a una función de densidad posterior, como se describió en la sección 1.5. Adicionalmente, se presentan condiciones suficientes para convergencia de la aproximación a $E[g(\theta)]$.

Se comienza con notación y algunos supuestos generales.

El valor esperado de una función $h(\theta)$ cuando la densidad de θ es proporcional a $f(\theta)$ se denota y define por

$$E_I[h(\theta)] = \left[\int f(\theta) d\theta \right]^{-1} \int_{\Theta} h(\theta) f(\theta) d\theta.$$

donde $f(\theta)$ es el núcleo de la densidad posterior y Θ su dominio.

SUPUESTO 1: El producto de la densidad a priori, $f_{\Theta}(\theta)$, y de la función de verosimilitud, $L(\theta)$, es proporcional a una f.d.p. propia definida sobre Θ .

SUPUESTO 2: $\{\theta_i\}_{i=1}^{\infty}$ es una sucesión de vectores aleatorios i.i.d., que tienen como densidad común a la función de densidad de Muestreo de Importancia $MI(\theta)$.

SUPUESTO 3: El soporte de $MI(\theta)$ incluye a Θ .

SUPUESTO 4: $E[h(\theta)]$ es finita.

Los supuestos 2 y 3 son específicos para el método de integración Monte Carlo con Muestreo de Importancia, pero el tercer supuesto no siempre se cumple.

Bajo ciertas circunstancias es recomendable no trabajar con las constantes de normalización para $MI(\theta)$, y reexpresar \bar{g}_n , ecuación (1.22), en función de $MI^*(\theta) = d^{-1}MI(\theta)$. Denotando la función de peso como

$$w(\theta) \equiv \frac{p(\theta)}{MI^*(\theta)}, \quad (2.1)$$

entonces de la ecuación (1.22) se deduce

$$\bar{g}_n = \frac{\sum_{i=1}^n \frac{g(\theta_i)p(\theta_i)}{MI^*(\theta_i)}}{\sum_{i=1}^n \frac{p(\theta_i)}{MI^*(\theta_i)}} = \frac{\sum_{i=1}^n g(\theta_i)w(\theta_i)}{\sum_{i=1}^n w(\theta_i)}. \quad (2.2)$$

Teorema 2.1 *Bajo los 4 supuestos anteriores y denotando a $E[g(\theta)] = \bar{g}$, entonces*

$$\bar{g}_n \xrightarrow{c.s.} E[g(\theta)] = \bar{g}.$$

Demostración. Como $\theta_1, \dots, \theta_n$ son i.i.d. y utilizando la Ley Fuerte de los Grandes Números (Teorema 1.2) se tiene que

$$\begin{aligned} n^{-1} \sum_{i=1}^n \frac{g(\theta_i)p(\theta_i)}{MI^*(\theta_i)} &\xrightarrow{c.s.} E_{MI} \left[\frac{g(\theta)p(\theta)}{MI^*(\theta)} \right] = \int_{\theta} \frac{g(\theta)p(\theta)}{MI^*(\theta)} MI(\theta) d\theta \\ &= d \int_{\theta} \frac{g(\theta)p(\theta)}{MI(\theta)} MI(\theta) d\theta \\ &= d \int_{\theta} g(\theta) p(\theta) d\theta, \end{aligned} \quad (2.3)$$

y adicionalmente

$$\begin{aligned} n^{-1} \sum_{i=1}^n \frac{p(\theta_i)}{MI^*(\theta_i)} &\xrightarrow{c.s.} E_{MI} \left[\frac{p(\theta_i)}{MI^*(\theta_i)} \right] = d \int \frac{p(\theta)}{MI(\theta)} MI(\theta) d\theta \\ &= cd. \end{aligned} \quad (2.4)$$

Las convergencias (2.3 y 2.4) pueden escribirse haciendo uso del Teorema 1.2 como

$$\lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \frac{g(\theta_i)p(\theta_i)}{MI^*(\theta_i)} = d \int_{\theta} g(\theta) p(\theta) d\theta$$

converge con probabilidad 1, y

$$\lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \frac{p(\theta_i)}{MI^*(\theta_i)} = cd$$

converge con probabilidad 1, su cociente cumple que

$$\lim_{n \rightarrow \infty} \frac{n^{-1} \sum_{i=1}^n \frac{g(\theta_i)p(\theta_i)}{MI^*(\theta_i)}}{n^{-1} \sum_{i=1}^n \frac{p(\theta_i)}{MI^*(\theta_i)}} = \frac{d \int_{\Theta} g(\theta) p(\theta) d\theta}{cd}$$

converge con probabilidad 1, como el cociente de las integrales anteriores es igual a \bar{g}_n (ecuación 2.2). aplicando de nuevo el Teorema 1.2; \bar{g}_n converge a \bar{g} con probabilidad 1, escrito como:

$$\bar{g}_n \xrightarrow{c.s.} \frac{d \int_{\Theta} g(\theta) p(\theta) d\theta}{cd} = \bar{g}.$$

□

Estos resultados proporcionan una aproximación a $g(\theta)$ que pudiera considerarse pobre dado que nada puede decirse acerca de la velocidad de convergencia, que depende de la selección de $MI(\theta)$. De hecho, para algunas $MI(\theta)$, \bar{g}_n puede tener un mal comportamiento (ver el Capítulo de Simulaciones gráficas 4.3, 4.4 y 4.5). Una posible solución es encontrar la distribución de \bar{g}_n y tratar de minimizar su varianza.

Teorema 2.2 Si en adición a los supuestos anteriores se tiene que

$$\begin{aligned} E_p[w(\theta)] &= c^{-1} \int_{\Theta} w(\theta) p(\theta) d\theta \\ &= c^{-1} \int_{\Theta} \frac{p(\theta)^2}{MI^*(\theta)} d\theta \end{aligned}$$

y

$$\begin{aligned} E_p[g(\theta)^2 w(\theta)] &= c^{-1} \int_{\Theta} g(\theta)^2 w(\theta) p(\theta) d\theta \\ &= c^{-1} \int_{\Theta} \frac{g(\theta)^2 p(\theta)^2}{MI^*(\theta)} d\theta \end{aligned}$$

son finitas, definiendo

$$\begin{aligned} \sigma^2 &\equiv d^{-1} E_p[(\bar{g}(\theta) - \bar{g}_n)^2 w(\theta)] \\ &= c^{-1} d^{-1} \int_{\Theta} (g(\theta) - \bar{g}_n)^2 w(\theta) p(\theta) d\theta, \end{aligned}$$

y

$$\hat{\sigma}_n^2 \equiv \frac{\sum_{i=1}^n [y(\theta_i) - \bar{y}_n]^2 w(\theta_i)}{[\sum_{i=1}^n w(\theta_i)]^2} \quad (2.5)$$

entonces

$$n^{1/2}(\bar{y}_n - \bar{y}) \implies N(0, \sigma^2),$$

y

$$n\hat{\sigma}_n^2 \xrightarrow{c.s.} \sigma^2.$$

Demostración. Si se define $A(\theta) \equiv g(\theta)w(\theta)$, se observa que

$$\bar{y}_n = \frac{n^{-1} \sum_{i=1}^n A(\theta_i)}{n^{-1} \sum_{i=1}^n w(\theta_i)}.$$

Por el Teorema 1.2 se tiene que:

$$\begin{aligned} n^{-1} \sum_{i=1}^n A(\theta_i) &\xrightarrow{c.s.} E_{MI} [A(\theta_i)] = \int A(\theta) MI(\theta) d\theta \\ &= d \int g(\theta) p(\theta) d\theta \\ &= cdE_p[g(\theta)] \\ &= cd\bar{y}. \end{aligned}$$

y

$$n^{-1} \sum_{i=1}^n w(\theta_i) \xrightarrow{c.s.} E[w(\theta_i)] = cd.$$

A continuación y con base en el Teorema 1.3 (Teorema del Límite Central) se demuestra que

$$n^{1/2}(\bar{y}_n - \bar{y}) \implies N(0, \sigma^2).$$

De acuerdo a la Definición 1.8, sea

$$T_n = n^{1/2} \begin{pmatrix} n^{-1} \sum_{i=1}^n A(\theta_i) - E[n^{-1} \sum_{i=1}^n A(\theta_i)] \\ n^{-1} \sum_{i=1}^n w(\theta_i) - E[n^{-1} \sum_{i=1}^n w(\theta_i)] \end{pmatrix}. \quad (2.6)$$

Ahora se demuestra que T_n se distribuye asintóticamente Normal Bivariada con vector de medias

$$\mu_n = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

y matriz de varianzas-covarianzas.

$$\begin{aligned} \text{Var}(T_n) &= \sum_n \\ &= \begin{pmatrix} \text{var} \left[\frac{\sqrt{n}}{n} \sum_{i=1}^n A(\theta_i) \right] & \text{cov} \left[\left(\frac{\sqrt{n}}{n} \sum_{i=1}^n A(\theta_i) \right), \left(\frac{\sqrt{n}}{n} \sum_{i=1}^n w(\theta_i) \right) \right] \\ \text{cov} \left[\left(\frac{\sqrt{n}}{n} \sum_{i=1}^n A(\theta_i) \right), \left(\frac{\sqrt{n}}{n} \sum_{i=1}^n w(\theta_i) \right) \right] & \text{var} \left[\frac{\sqrt{n}}{n} \sum_{i=1}^n w(\theta_i) \right] \end{pmatrix}. \end{aligned}$$

Comenzando por el cálculo de las varianzas y la covarianza,

$$\begin{aligned} \text{var} \left[\frac{\sqrt{n}}{n} \sum_{i=1}^n A(\theta_i) \right] &= \frac{1}{n} \sum_{i=1}^n \text{var} [A(\theta_i)] \\ &= \text{var} [A(\theta_1)] \\ &= \int [A(\theta) - cd\bar{y}]^2 MI(\theta) d\theta \\ &= \int A(\theta)^2 MI(\theta) d\theta + (cd\bar{y})^2 - 2cd\bar{y} \int A(\theta) MI(\theta) d\theta \\ &= \int g(\theta)^2 w(\theta) \frac{p(\theta)}{d-1} MI(\theta) d\theta + (cd\bar{y})^2 - 2cd\bar{y} \int g(\theta) \frac{p(\theta)}{d-1} MI(\theta) d\theta \\ &= \int g(\theta)^2 w(\theta) p(\theta) d\theta + (cd\bar{y})^2 - 2cd^2\bar{y} \int g(\theta) p(\theta) d\theta \\ &= cd \int g(\theta)^2 w(\theta) \frac{p(\theta)}{c} d\theta + (cd\bar{y})^2 - 2c^2 d^2 \bar{y} \int g(\theta) \frac{p(\theta)}{c} d\theta \\ &= cd E_p [g(\theta)^2 w(\theta)] + (cd\bar{y})^2 - 2c^2 d^2 \bar{y} E [g(\theta)] \\ &= cd E_p [g(\theta)^2 w(\theta)] + (cd\bar{y})^2 - 2(cd\bar{y})^2 \\ &= cd E_p [g(\theta)^2 w(\theta)] - (cd\bar{y})^2 \end{aligned}$$

y

$$\begin{aligned} \text{var} \left[\left(\frac{\sqrt{n}}{n} \sum_{i=1}^n w(\theta_i) \right) \right] &= \frac{1}{n} \sum_{i=1}^n \text{var} [w(\theta_i)] \\ &= \text{var} [w(\theta_1)] \\ &= \int [w(\theta) - cd]^2 MI(\theta) d\theta \\ &= \int w(\theta)^2 MI(\theta) d\theta + (cd)^2 - 2cd \int w(\theta) MI(\theta) d\theta \\ &= \int \frac{w(\theta)p(\theta)MI(\theta)}{d-1} d\theta + (cd)^2 - 2cd \int \frac{p(\theta)MI(\theta)}{d-1} d\theta \\ &= \int w(\theta) p(\theta) d\theta + (cd)^2 - 2cd^2 \bar{y} \int p(\theta) d\theta \\ &= cd \int \frac{w(\theta)}{c} p(\theta) d\theta + (cd)^2 - 2c^2 d^2 \bar{y} \int \frac{p(\theta)}{c} d\theta \\ &= cd E_p [w(\theta)] + (cd)^2 - 2(cd)^2 \\ &= cd E_p [w(\theta)] - (cd)^2, \end{aligned}$$

por último el cálculo de la covarianza:

$$\begin{aligned}
 \text{cov} \left(\frac{\sqrt{n} \sum_{i=1}^n A(\theta_i)}{n}, \frac{\sqrt{n} \sum_{i=1}^n w(\theta_i)}{n} \right) &= \frac{1}{n} \text{cov} \left[\sum_{i=1}^n A(\theta_i), \sum_{i=1}^n w(\theta_i) \right] \\
 &= \frac{1}{n} \sum_{i=1}^n \text{cov} [A(\theta_i), w(\theta_i)] \\
 &= \int [A(\theta) - cd\bar{g}] [w(\theta) - cd] MI(\theta) d\theta \\
 &= \int [A(\theta)w(\theta) - cdA(\theta) - cd\bar{g}w(\theta) + (cd)^2\bar{g}] MI(\theta) d\theta \\
 &= cdE[A(\theta)] - (cd)^2\bar{g} - (cd)^2\bar{g} + (cd)^2\bar{g} \\
 &= \int [dA(\theta) - cd^2g(\theta) - cd^2\bar{g}] p(\theta) d\theta + (cd)^2\bar{g} \\
 &= cdE[g(\theta)w(\theta)] - (cd)^2\bar{g}.
 \end{aligned}$$

Se nota que las varianzas son positivas y no dependen del tamaño de la muestra, verificando así el primer inciso de la Definición 1.8.

Sea λ un vector de escalares $\lambda' = (\lambda_1, \lambda_2)'$ para verificar el segundo inciso y ver la normalidad asintótica bivariada de T_n , entonces λ multiplica a la matriz (2.6) como sigue

$$\begin{aligned}
 \lambda' T_n &= \lambda_1 \left[\sqrt{n} \left(n^{-1} \sum_{i=1}^n A(\theta_i) - E[A(\theta_i)] \right) \right] \\
 &\quad + \lambda_2 \left[\sqrt{n} \left(n^{-1} \sum_{i=1}^n w(\theta_i) - E[w(\theta_i)] \right) \right] \\
 &= \sqrt{n} \left(n^{-1} \sum_{i=1}^n (\lambda_1 A(\theta_i) + \lambda_2 w(\theta_i)) \right) \\
 &\quad - \sqrt{n} (\lambda_1 E[A(\theta_i)] + \lambda_2 E[w(\theta_i)]).
 \end{aligned}$$

Cumpliendo lo siguiente

$$\frac{\lambda' T_n}{\delta_n} \Rightarrow N(0, 1),$$

donde

$$\begin{aligned}
 \delta_n^2 &= \text{Var}(T_n) \\
 &= \lambda \sum_n \lambda'
 \end{aligned}$$

y

$$\frac{\lambda' T_n}{\delta_n} \stackrel{d}{=} \frac{\sqrt{n} \{n^{-1} \sum_{i=1}^n (\lambda_1 A(\theta_i) + \lambda_2 w(\theta_i)) - (\lambda_1 E[A(\theta_i)] + \lambda_2 E[w(\theta_i)])\}}{\delta_n}.$$

Entonces, aplicando el Teorema 1.3

$$\lambda' T_n \sim AN(0, \delta_n^2),$$

y como consecuencia

$$T_n \sim A.N_2 \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma_n \right].$$

Se denota una nueva variable por

$$\frac{1}{\sqrt{n}} T_n = \left(\begin{array}{c} n^{-1} \sum_{i=1}^n A(\theta_i) - E[A(\theta_i)] \\ n^{-1} \sum_{i=1}^n w(\theta_i) - E[w(\theta_i)] \end{array} \right) \sim A.N_2 \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \frac{1}{n} \Sigma_n \right]$$

que cumple claramente con

$$\frac{\frac{1}{\sqrt{n}} \lambda' T_n}{\delta_n / \sqrt{n}} \implies N(0, 1)$$

entonces

$$\frac{1}{\sqrt{n}} \lambda' T_n \sim A.N \left[0, \frac{1}{n} \delta_n^2 \right].$$

Puede decirse que el vector

$$\frac{1}{\sqrt{n}} T_n \sim A.N_2 \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \frac{1}{n} \Sigma_n \right],$$

sobre la Proposición 1.6 (pág. 9) sea $c_n^2 = 1/n$, verificando lo siguiente.

$$c_n^2 = \frac{1}{n} \rightarrow 0 \quad \text{cuando} \quad n \rightarrow \infty.$$

y sea la transformación

$$H \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \frac{X_1 + E[A(\theta_1)]}{X_2 + E[w(\theta_1)]}$$

que es continuamente diferenciable, aplicando la transformación H al vector $(1/\sqrt{n}) T_n$ queda como

$$H \left(\frac{1}{\sqrt{n}} T_n \right) = \frac{n^{-1} \sum_{i=1}^n A(\theta_i)}{n^{-1} \sum_{i=1}^n w(\theta_i)} \quad (2.7)$$

y sea D la matriz de 1×2 como se describe en la Proposición 1.6 igual a

$$\begin{aligned} D &= \left(\frac{\partial H}{\partial X_1}, \frac{\partial H}{\partial X_2} \right)_{X_1=0, X_2=0} \\ &= \left(\frac{1}{X_2 + E[w(\theta_1)]}, -\frac{X_1 - E[A(\theta_1)]}{(X_2 + E[w(\theta_1)])^2} \right) \\ &= \left(\frac{1}{cd}, -\frac{\bar{a}}{cd} \right). \end{aligned}$$

De acuerdo en la forma como se definió \bar{g}_n , ecuación (2.2), se expresa como en la ecuación (2.7)

$$\bar{g}_n = H \left(\frac{1}{\sqrt{n}} T_n \right) = \frac{n^{-1} \sum_{i=1}^n g(\theta_i) w(\theta_i)}{n^{-1} \sum_{i=1}^n w(\theta_i)};$$

y la Proposición 1.6 dice que

$$H \left(\frac{1}{\sqrt{n}} T_n \right) \sim AN \left(\frac{E[A(\theta_1)]}{E[w(\theta_1)]} = \bar{g}, \frac{1}{n} D \sum_n D' \right),$$

donde

$$\begin{aligned} D \sum_n D' &= D \begin{pmatrix} \text{var} [(A(\theta_1))] & \text{cov} [A(\theta_1), w(\theta_1)] \\ \text{cov} [A(\theta_1), w(\theta_1)] & \text{var} [w(\theta_1)] \end{pmatrix} D' \\ &= \left(\frac{1}{cd}, -\frac{\bar{g}}{cd} \right) \begin{pmatrix} \text{var} [(A(\theta_1))] & \text{cov} [A(\theta_1), w(\theta_1)] \\ \text{cov} [A(\theta_1), w(\theta_1)] & \text{var} [w(\theta_1)] \end{pmatrix} \begin{pmatrix} \frac{1}{cd} \\ -\frac{\bar{g}}{cd} \end{pmatrix} \\ &= \left(\frac{1}{cd} \right) (\text{var} [A(\theta_1)] - \bar{g} \text{cov} [A(\theta_1), w(\theta_1)] \cdot \text{cov} [A(\theta_1), w(\theta_1)] - \bar{g} \text{var} [w(\theta_1)]) \begin{pmatrix} \frac{1}{cd} \\ -\frac{\bar{g}}{cd} \end{pmatrix} \\ &= \left(\frac{1}{cd} \right)^2 (\text{var} [(A(\theta_1))] - \bar{g} \text{cov} [A(\theta_1), w(\theta_1)] - \bar{g} \text{cov} [A(\theta_1), w(\theta_1)] + \bar{g}^2 \text{var} [w(\theta_1)]) \\ &= (cd)^{-2} (\text{var} [(A(\theta_1))] - 2\bar{g} \text{cov} [A(\theta_1), w(\theta_1)] + (\bar{g})^2 \text{var} [w(\theta_1)]) \\ &= (cd)^{-2} \left(cd E_p [g(\theta_1)^2 w(\theta_1)] - (cd\bar{g})^2 - 2\bar{g} cd E_p [g(\theta_1) w(\theta_1)] \right) \\ &\quad + 2(cd\bar{g})^2 + \bar{g}^2 cd E_p [w(\theta_1)] - (cd\bar{g})^2 \\ &= (cd)^{-2} \left(cd \int [g(\theta)^2 w(\theta) - \bar{g}g(\theta)w(\theta) + (\bar{g})^2 w(\theta)] \frac{p(\theta)}{c} d\theta \right) \\ &= (cd)^{-2} \left(d \int [g(\theta)^2 - \bar{g}g(\theta) + (\bar{g})^2] w(\theta) p(\theta) d\theta \right) \\ &= c^2 d^{-1} \int [g(\theta) - \bar{g}]^2 w(\theta) p(\theta) d\theta. \end{aligned}$$

Aplicando el Teorema 1.3 y de acuerdo con la Definición 1.7 se tiene

$$n^{1/2} (\bar{g}_n - \bar{g}) \implies N(0, \sigma^2),$$

donde

$$\sigma^2 = c^{-2} d^{-1} \int [g(\theta) - \bar{g}]^2 p(\theta) w(\theta) d\theta.$$

Como σ^2 es desconocida y difícil de calcular en forma analítica se construye una variable (ecuación 2.5) tal que

$$n\hat{\sigma}_n^2 \xrightarrow{c} \sigma^2.$$

Sea

$$n\hat{\sigma}_n^2 = \frac{\frac{1}{n} \sum_{i=1}^n [g(\theta_i) - \bar{g}_n]^2 w(\theta_i)^2}{\left[\frac{1}{n} \sum_{i=1}^n w(\theta_i) \right]^2} \quad (2.8)$$

$$\begin{aligned}
&= \frac{\frac{1}{n} \sum_{i=1}^n [g(\theta_i)^2 - 2\bar{g}_n g(\theta_i) + (\bar{g}_n)^2] w(\theta_i)^2}{\left[\frac{1}{n} \sum_{i=1}^n w(\theta_i) \right]^2} \\
&= \frac{\frac{1}{n} \sum_{i=1}^n g(\theta_i)^2 w(\theta_i)^2 - 2\bar{g}_n \frac{1}{n} \sum_{i=1}^n g(\theta_i) w(\theta_i)^2 + (\bar{g}_n)^2 \frac{1}{n} \sum_{i=1}^n w(\theta_i)^2}{\left[\frac{1}{n} \sum_{i=1}^n w(\theta_i) \right]^2}
\end{aligned}$$

Entonces cada una de las sumas tiene valores θ_i independientes y por el Teorema 1.2 cada elemento de (2.8) converge en forma casi segura a

$$\begin{aligned}
\int [(g(\theta) w(\theta))^2 - 2\bar{g} g(\theta) w(\theta)^2 + (\bar{g} w(\theta))^2] M I(\theta) d\theta &= \frac{\int [g(\theta)^2 - 2\bar{g} g(\theta) + \bar{g}^2] u(\theta)^2 M I(\theta) d(\theta)}{\left[\int u(\theta) M I(\theta) d(\theta) \right]^2} \\
&= \frac{d \int (g(\theta) - \bar{g})^2 u(\theta)^2 p(\theta) d(\theta)}{(cd)^2} \\
&= c^{-2} d^{-1} \int [g(\theta) - \bar{g}]^2 p(\theta) w(\theta) d\theta \\
&= \sigma^2
\end{aligned}$$

lo cual finaliza la demostración.

□

En la siguiente sección se estima el intervalo de confianza para $E[g(\theta)]$ utilizando las herramientas demostradas anteriormente.

2.1 ESTIMACION POR INTERVALO DE LA MEDIA POSTERIOR DE UNA FUNCION

En esta sección, se hace un breve cálculo para la estimación de un intervalo de confianza de la esperanza de la función $g(\theta)$ graficando para diferentes longitudes, los tamaños de muestra respectivos que se necesitarían.

Se sabe que

$$n\hat{\sigma}_n^2 \xrightarrow{c.s.} \sigma^2.$$

Una forma alternativa de expresar a $n\hat{\sigma}_n^2$ sería

$$n\hat{\sigma}_n^2 = n * \frac{\sum_{i=0}^n (g(\theta_i) - \bar{g}_n)^2 w(\theta_i)^2}{\sum_{i=0}^n w(\theta_i)} * \frac{1}{\frac{1}{n} \sum_{i=0}^n w(\theta_i)}. \quad (2.9)$$

De acuerdo al Teorema 1.2 se cumple lo siguiente

$$\frac{1}{n} \sum_{i=0}^n w(\theta_i) \xrightarrow{c.s.} cd$$

y si se multiplica por (2.9) se tiene la siguiente expresión de la varianza

$$\tilde{\sigma}^2 = \frac{\sum_{i=0}^n (g(\theta_i) - \bar{g}_n)^2 w(\theta_i)^2}{\sum_{i=0}^n w(\theta_i)} \xrightarrow{c.s.} cd\sigma^2.$$

Por el Teorema 1.3.

$$\frac{\sqrt{n}(\bar{g}_n - \bar{g})}{\tilde{\sigma}} \xrightarrow{D} Z, \quad (2.10)$$

donde $Z \sim N(0, 1)$. Con las definiciones de la sección sobre modos de Convergencia puede verse

$$\sqrt{\frac{cd\sigma^2}{\tilde{\sigma}^2}} = \frac{\sqrt{cd}\sigma}{\tilde{\sigma}} \xrightarrow{c.s.} 1. \quad (2.11)$$

Si se multiplican las dos ecuaciones anteriores. (2.10 y 2.11), y si se altera la convergencia en distribución por una constante, entonces

$$\frac{\sqrt{n}(\bar{g}_n - \bar{g})}{\tilde{\sigma}} * \frac{\sqrt{cd}\sigma}{\tilde{\sigma}} = \frac{\sqrt{cdn}(\bar{g}_n - \bar{g})}{\tilde{\sigma}} \xrightarrow{D} Z,$$

donde $Z \sim N(0, 1)$. De esta forma se pueden calcular los intervalos de confianza.

Se desea encontrar un valor de a tal que

$$p \left[-a < \frac{\sqrt{cdn}(\bar{g}_n - \bar{g})}{\tilde{\sigma}} < a \right] = 0.95.$$

Trabajando con la parte central se deduce el siguiente intervalo para la media

$$\bar{g} \in \left(\bar{g}_n - a \frac{\tilde{\sigma}}{\sqrt{cdn}}, \bar{g}_n + a \frac{\tilde{\sigma}}{\sqrt{cdn}} \right),$$

donde, de acuerdo a valores en tabla de una distribución normal estándar $a = 1.96$. En cierto modo puede considerarse al segundo factor de la suma como una longitud. Basándose en esto surge una pregunta ¿para que tamaño de muestra se tiene la longitud deseada?. Esta es una de las ventajas que asegura el Teorema 2.2 del Capítulo 2.

Haciendo unos pequeños despejes se encuentra el tamaño. Sea L la longitud propuesta del intervalo, entonces

$$L = 2 * a \frac{\tilde{\sigma}}{\sqrt{cdn}}$$

y por lo tanto

$$n = \left(a \frac{\tilde{\sigma}}{2 * L\sqrt{cd}} \right)^2.$$

Así para un tamaño de muestra inicial se encuentra otro nuevo tamaño de acuerdo a una longitud propuesta y así sucesivamente hasta asegurar tener una buena aproximación para la esperanza.

En la siguiente figura (figura 2.2) se grafican los tamaños de muestra estimados contra la longitud de intervalo con un tamaño de muestra inicial.

Las gráficas 2.2 fueron generadas con el programa llamado "mn" que se encuentra codificado en el Apéndice.

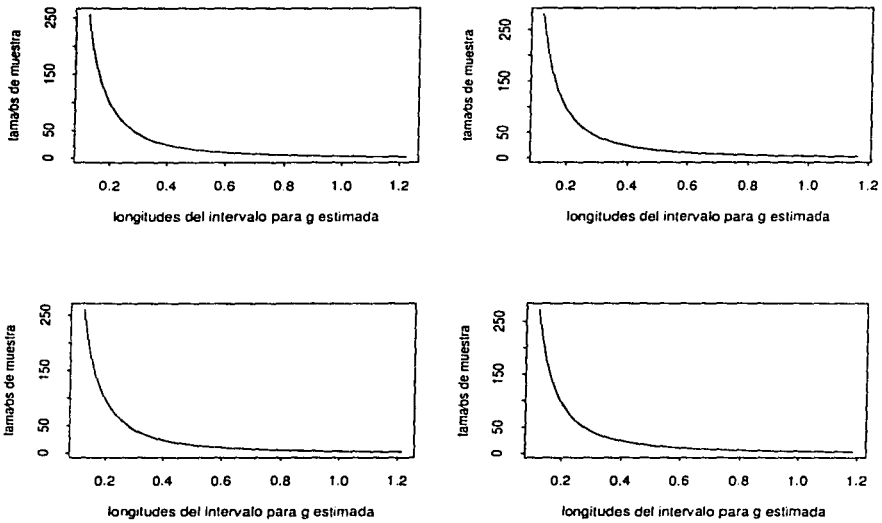


FIGURA 2.2: Longitud del tamaño estimado con diferentes tamaños de muestra.

Capítulo 3

CONVERGENCIA DEL MUESTREO DE SUSTITUCION SUCESIVA

Dentro de este capítulo se detalla una prueba general de convergencia desarrollada por Schervish M. y Bradley (1992) para distribuciones arbitrarias continuas, la cual supone que las distribuciones condicionales usadas en el algoritmos descritos en la introducción (págs. 17 y 19) son estrictamente positivas y que el kernel del operador de transición es doblemente integrable con respecto a una medida en particular.

Antes se introducen algunos conceptos básicos del espacio de Hilbert que serán utilizados en el resto del capítulo.

3.1 EL ESPACIO DE HILBERT

Definición 3.1 *Un espacio vectorial V es un conjunto de objetos X, Y, Z, \dots llamados vectores, donde se define a $\underline{0}$ el vector cero y para cada vector X_i existe un vector negativo $-X_i$, cumpliendo los elementos de V las siguientes axiomas:*

la adición de los vectores se sujeta a las siguientes reglas:

- (a) $X + Y = Y + X$ (ley conmutativa);
- (b) $X + (Y + Z) = (X + Y) + Z$ (ley asociativa);
- (c) $X + \underline{0} = X$;
- (d) $X + (-X) = \underline{0}$;

sean λ, μ escalares, la multiplicación de los vectores se sujeta a las siguientes reglas:

- (a) $\lambda(X + Y) = \lambda X + \lambda Y$;
- (b) $(\lambda + \mu)X = \lambda X + \mu X$;
- (c) $(\lambda\mu)X = \lambda(\mu X)$;
- (d) $1X = X$.

Definición 3.2 Una estructura $V = (V, +, \cdot, \langle \rangle)$ es un espacio de producto interno si

1. $(V, +, \cdot)$ es un espacio vectorial;
2. $\langle \cdot, \cdot \rangle$ es una función que asocia un número con cada par de vectores en V sujetos a las siguientes reglas:

para todo $X, Y, Z \in V$ y $\lambda \in R$

- (a) $\langle X, Y \rangle = \langle Y, X \rangle$
- (b) $\langle X + Y, Z \rangle = \langle X, Z \rangle + \langle Y, Z \rangle$ y $\langle \lambda X, Y \rangle = \lambda \langle X, Y \rangle$
- (c) $\langle \lambda X, Y \rangle = \lambda \langle X, Y \rangle$
- (d) $\langle X, X \rangle$ es un número real no negativo, y $\langle X, X \rangle = 0$ si y sólo si X es el vector cero en V .

Un espacio de producto interno con la propiedad adicional de completitud es llamado espacio de *Hilbert*. Para definir completitud es necesario el concepto de una sucesión de Cauchy.

Definición 3.3 La norma de producto interior sobre V es una función de V a \mathbb{R} definida para cada $X \in V$ por

$$\|X\| = \sqrt{\langle X, X \rangle}.$$

Teorema 3.1 La norma de producto interior satisface lo siguiente:

para todo $X, Y \in V$ y $\lambda \in \mathbb{R}$

- (a) $\|X\| \geq 0$, y la igualdad se cumple si $X = 0$.
- (b) $\|\lambda X\| = |\lambda| \|X\|$;
- (c) $\|X + Y\|^2 + \|X - Y\|^2 = 2\|X\|^2 + 2\|Y\|^2$.

Demostración. Ver Cohen (1989, pág.16). \square

Teorema 3.2 (La desigualdad de Cauchy-Schwarz).

Para todo $X, Y \in V$

$$|\langle X, Y \rangle| \leq \|X\| \|Y\|.$$

Demostración. Ver Cohen (1989, pág. 17). \square

Definición 3.4 Una sucesión $\{X_n\}$ en V converge en norma a un vector $Y \in V$ si y sólo si

$$\lim_{n \rightarrow \infty} \|X_n - Y\| = 0.$$

Definición 3.5 (Sucesión de Cauchy). Una sucesión $\{X_n, n = 1, 2, \dots\}$ de elementos de un espacio de producto interno se dice que es una sucesión de Cauchy si

$$\|X_n - X_m\| \rightarrow 0 \quad \text{cuando } m, n \rightarrow \infty,$$

i.e. si para cada $\epsilon > 0$ existe un entero positivo $N(\epsilon)$ tal que

$$\|X_n - X_m\| < \epsilon \quad \text{para toda } m, n > N(\epsilon).$$

Definición 3.6 (Espacio de Hilbert). Un espacio de Hilbert H es un espacio de producto interior el cual es completo, i.e., un espacio de producto interno en el cual cada sucesión de Cauchy $\{X_n\}$ converge en norma a algún elemento $X \in H$.

Definición 3.7 Un conjunto de vectores X_α en un espacio de Hilbert H , donde α recorre algún conjunto de índices A , se dice ortonormal si satisface las relaciones de ortogonalidad $\langle X_\alpha, X_\beta \rangle = 0$ para todo $\alpha \neq \beta$, $\alpha \in A$, $\beta \in A$ y si está normalizado de modo que $\|X_\alpha\| = 1$ para toda $\alpha \in A$. En otras palabras $\{X_\alpha\}$ es ortonormal si

$$\langle X_\alpha, X_\beta \rangle = \begin{cases} 1 & \text{si } \alpha = \beta \\ 0 & \text{si } \alpha \neq \beta \end{cases}.$$

Definición 3.8 Una sucesión de $\{f_n\}_{n=1}^{\infty}$ en el espacio de Hilbert H , converge (o converge fuertemente) a f si $\langle f_n - f, f_n - f \rangle$ converge a cero. La sucesión converge débilmente a f si para cada $u \in H$, $\langle f_n, u \rangle$ converge a $\langle f, u \rangle$.

3.1.1 OPERADORES ADJUNTOS.

Definición 3.9 Sea una colección τ de subconjuntos de un conjunto X , se dice ser una topología X si τ cumple las tres siguientes propiedades:

1. $\emptyset \in \tau$ y $X \in \tau$.
2. si $V_i \in \tau$, $i = 1, 2, \dots, n$, entonces $V_1 \cap V_2 \cap \dots \cap V_n \in \tau$,
3. si $\{V_\alpha\}$ es una colección arbitraria de miembros de τ (finita, numerable o no numerable), entonces $\bigcup_\alpha V_\alpha \in \tau$.

Si τ es una topología en X entonces es llamado un espacio topológico y los miembros de τ son llamados los conjuntos abiertos de X .

Definición 3.10 Se denota como $L_2(\mu, \Sigma, S)$ a la colección de todas las funciones complejas medibles f sobre un campo Σ de subconjuntos de un espacio topológico S para las que

$$\int |f|^2 d\mu < \infty.$$

Los elementos de $L_2(\mu, \Sigma, S)$ se denominan funciones integrables Lebesgue (con respecto a μ) o funciones sumables.

Definición 3.11 Cualquier mapeo lineal $T : H \rightarrow H$ será llamado un operador.

En particular I es el operador identidad, 0 es el operador cero y λI es el operador escalar.

Definición 3.12 El adjunto de un operador U es el operador U^* tal que para toda $f, g \in H$, $\langle U(f), g \rangle = \langle f, U^*(g) \rangle$. Un operador U es auto-adjunto si $U = U^*$, (operador Hermitiano).

H, K, L denotan espacios de Hilbert.

Para abreviar se utiliza la siguiente notación donde T es un operador y $X \in H$

$$T(X) \equiv TX.$$

Teorema 3.3 . Si $S : H \rightarrow K$ y $T : H \rightarrow K$ son mapeos lineales continuos y λ un escalar;

$$1) (S + T)^* = S^* + T^*$$

$$2) (\lambda T)^* = \lambda^* T^*$$

$$3) (T^*Y|X) = (Y|TX) \quad \text{para toda } X \in H, Y \in K$$

$$4) (T^*)^* = T$$

$$5) \|T^*T\| = \|TT^*\| = \|T\|^2$$

$$6) T^*T = 0 \iff T = 0.$$

Demostración. Ver Berberian (1976, secc. 8, pág.132). \square

Teorema 3.4 . Si $T : H \rightarrow K$ es un mapeo lineal continuo, existe uno y sólo un mapeo lineal continuo

$$T^* : K \rightarrow H$$

tal que

$$(T X|Y) = (X|T^*Y)$$

para toda $X \in H, Y \in K$, se tiene que $\|T^*\| = \|T\|$.

Demostración. Ver Berberian (1976, secc. 8, pág. 131). \square

Teorema 3.5 . Si $T : H \rightarrow K$ y $S : K \rightarrow L$ son mapeos lineales continuos, entonces $(ST)^* = T^*S^*$.

Demostración. Ver Berberian (1976, secc.8, pág. 133). \square

Definición 3.13 . (Operadores Hilbert-Schmidt). Si el espacio de Hilbert es representado como un espacio $L_2(S, \Sigma, \mu)$, con medida positiva μ , entonces los operadores Hilbert-Schmidt son aquellos operadores k teniendo una representación en la forma

$$(kf)(s) = \int_S k(s,t) f(t) \mu(dt). \quad f \in L_2(S, \Sigma, \mu),$$

donde

$$\int_S \int_S |k(s,t)|^2 \mu(ds) \mu(dt) < \infty.$$

En algunas de las interpretaciones de la teoría de los operadores de Hilbert-Schmidt se supone que el kernel k es Hermitiano simétrico de tal manera que el correspondiente operador es auto-adjunto. A veces será mejor trabajar con un espacio definido en el L_2 -espacio y pueden ser definidos los operadores como:

sea $\{X_\alpha, \alpha \in A\}$ es un conjunto ortonormal completo en el espacio de Hilbert. Un operador lineal acotado T se dice ser un "operador Hilbert-Schmidt" si la cantidad $\|T\|$ definida por la ecuación

$$\|T\| = \left\{ \sum_{\alpha \in A} |TX_\alpha|^2 \right\}^{1/2}$$

es finita.

Definición 3.14 Un operador U es completamente continuo si cada conjunto acotado $B \subseteq H$ es mapeado sucesivamente a un conjunto compacto por U . (Esto es, cada sucesión en $U(B)$ tiene una subsucesión convergente).

Teorema 3.6 Operadores completamente continuos.

Si para cada $\varepsilon > 0$ existe un operador completamente continuo A_ε tal que

$$\|(A - A_\varepsilon)\| \leq \varepsilon \|f\|$$

para $f \in H$ entonces el operador A es completamente continuo.

Demostración. Ver Akhiezer and Glazman (1961, sección 28, pág. 27). \square

Teorema 3.7 . (El espectro de un operador completamente continuo auto-adjunto en R).

Cada operador no-cero completamente continuo auto-adjunto A tiene una sucesión de eigenvectores (finita o infinita) e_1, e_2, e_3, \dots pertenecientes a sus eigenvalores no-cero $\lambda_1, \lambda_2, \lambda_3, \dots$

$$(|\lambda_1| \geq |\lambda_2| \geq |\lambda_3| \geq \dots)$$

tal que para cada vector f de la forma $f = Ah$ la relación de Parseval

$$\|f\|^2 = \sum_k |(f, e_k)|^2$$

se cumple.

Demostración. Ver Akhiezer and Glazman (1961, secc. 55, pág. 127).

Teorema 3.8 (Teorema espectral para espacios dimensionalmente finitos).

Si A es un operador Hermitiano en H , entonces existe una base ortonormal $B_A = \{X_1, \dots, X_n\}$ para H con respecto a la matriz diagonal para A (todas las entradas no diagonales iguales a cero)

$$A = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$$

y

(a) λ_k está en la diagonal si y sólo si es una eigenvalor de A ;

(b) para toda k , $x_k \in S(A, \lambda)$;

(c) para toda $\lambda \in A$, $S(A, \lambda) = \text{span} \{x_k | x_k \in B_A \text{ y } \lambda = \lambda_k\}$;

donde $S(A, \lambda) = \{x | x \in H \text{ y } Ax = \lambda x\}$, y si $S \subseteq H$ se define el span de S como la intersección de los siguientes conjuntos $\cap \{k | k \text{ es una subespacio en } H \text{ con } S \subseteq k\}$.

Demostración. Ver Cohen (1989, págs. 75,76). \square

3.2 PRUEBA DE LA CONVERGENCIA DEL M.S.S.

Sea $Y = (Y_1, \dots, Y_p)$ un vector aleatorio. Para cada $i = 1, \dots, p$, sea

$$f^{(i)}(x) = f_{Y, \{Y_r, r \neq i\}}(x),$$

que es, la densidad condicional dado $Y_j = x_j$ para $i \neq j$ de Y_i en x_i , con respecto a una medida λ_i . Para dos vectores Y y Y' , definen para cada $i = 1, \dots, p-1$,

$y^{(i')} = (y_1, \dots, y_i, y'_{i+1}, \dots, y'_p)$. Por conveniencia, sea

$$\begin{aligned} Y^{(0')} &= (Y'_1, \dots, Y'_p) = Y', \\ Y^{(p')} &= (Y_1, \dots, Y_p) = Y, \end{aligned}$$

y

$$f^{(0)}(y) = f_Y(y).$$

Sea μ la medida definida por

$$\mu(A) = \int_A \frac{1}{f_Y(y)} d\lambda^p(y),$$

donde λ^p es el producto de las medidas $\lambda_1 \times \dots \times \lambda_p$. Sea H el espacio de Hilbert de funciones que son doblemente integrables con respecto a una medida μ .

$$H = \left\{ g : \int |g|^2 d\mu < \infty \right\}. \quad (3.12)$$

El producto interno en H es

$$\langle f, g \rangle = \int f(y) g(y) d\mu(y).$$

Nótese que $f_Y \in H$ ya que

$$\begin{aligned} \int [f_Y(y)]^2 \frac{1}{f_Y(y)} d\mu(y) &= \int f_Y(y) d\lambda^p(y) \\ &= 1 < \infty \end{aligned}$$

y por tanto H no es vacío. También se nota que el producto interno depende de la densidad f_Y y no puede calcularse explícitamente si la función no es conocida. La razón por definir así el producto interno es que hace más clara la demostración y garantiza que f_Y está en el espacio de Hilbert.

Se define el operador T sobre H por $T(f) = b$ si

$$b(y) = \int K(y', y) f(y') d\mu(y'),$$

donde

$$\begin{aligned} K(y', y) &= \prod_{i=0}^p f^{(i)}(y^{(i)}) \\ &= f^{(0)}(y') \cdot f^{(1)}(y^{(1)}) \cdot \dots \cdot f^{(p)}(y) \\ &= f_{Y'}(y') f_{Y_1|Y_2, \dots, Y_p}(y_1 | y_2', \dots, y_p') \cdot \dots \cdot f_{Y_p|Y_1, \dots, Y_{p-1}}(y_p | y_1, \dots, y_{p-1}). \end{aligned}$$

Es claro que este operador proporciona la transición de la función de densidad conjunta para Y en una iteración del M.S.S. a la densidad de la siguiente iteración. Esto es, si f es la densidad conjunta (con respecto a λ^p) de las observaciones muestreadas a la iteración i , entonces $T(f)$ es la densidad conjunta a la iteración $i + 1$. De aquí que sea cierto que $T(f_Y) = f_Y$, y puede verse que f_Y es una eigenfunción de T con eigenvalor 1.

Se demuestra que $K(y', y) / f_Y(y')$ es una función de densidad con respecto a λ^p como función de y , de acuerdo al Teorema 1.4

$$\int \frac{K(y', y)}{f_Y(y')} d\lambda^p(y) = \int (f_Y(y'))^{-1} f_Y(y') f_{Y_1|Y_2, \dots, Y_p}(y_1 | y_2', \dots, y_p') \cdot \dots \quad (3.13)$$

$$\cdot \dots \cdot f_{Y_p|Y_1, \dots, Y_{p-1}}(y_p | y_1, \dots, y_{p-1}) d\lambda^p(y) \quad (3.14)$$

$$\begin{aligned} &= \int f_{Y_1|Y_2, \dots, Y_p}(y_1 | y_2', \dots, y_p') \cdot \dots \cdot f_{Y_p|Y_1, \dots, Y_{p-1}}(y_p | y_1, \dots, y_{p-1}) d\lambda^p(y) \\ &= 1. \end{aligned}$$

entonces

$$\int K(y', y) d\lambda^p(y) = f_Y(y'). \quad (3.15)$$

Lema 3.1 *El valor más grande del eigenvalor del operador T es 1.*

Demostración. Se supone que c es un eigenvalor de T con eigenfunción f . Entonces $T(f) = cf$ por la desigualdad de Cauchy-Schwarz se cumple que $\int \|f(y)\| d\lambda^p(y) < \infty$ junto con el desarrollo de la ecuación (3.13) se tiene

$$\begin{aligned} |c| \int \|f(y)\| d\lambda^p(y) &= \int \|T(f)(y)\| d\lambda^p(y) \\ &= \int \left\| \int K(y', y) f(y') d\mu(y') \right\| d\lambda^p(y) \\ &\leq \int \int K(y', y) d\lambda^p(y) \|f(y')\| d\mu(y') \\ &= \int \|f(y')\| f_Y(y') d\mu(y') \\ &= \int \|f(y')\| d\lambda^p(y') \end{aligned}$$

entonces

$$|c| \int \|f(y)\| d\lambda^p(y) \leq \int \|f(y')\| d\lambda^p(y')$$

de donde se sigue que $|c| \leq 1$ \square .

De acuerdo a la Definición 3.13, el operador T es de tipo Hilbert-Schmidt si

$$\iint |K(y', y)|^2 d\mu(y') d\mu(y) < \infty.$$

En este caso el operador T es completamente continuo de acuerdo al Teorema 3.6, según Akhiezer and Glazman (1961).

El adjunto de T es T^* , definido por $T^*(f) = b$ si

$$b(y) = \int K(y, y') f(y') d\mu(y').$$

El siguiente Lema demuestra que ambos operadores son muy parecidos.

Lema 3.2 *El operador T^* corresponde a una versión del M.S.S. en la cual las coordenadas están actualizadas en el orden opuesto.*

Demostración. Sea $K^*(y', y) = K(y, y')$. Para cualquier vector p -dimensional a , sea $a_{\setminus i}$ representa el vector $(p-1)$ -dimensional, el cual es igual al vector a pero no tiene la i -ésima coordenada. Con esta notación es posible escribir

$$\begin{aligned} K(y', y) &= \frac{\prod_{i=0}^p f_{Y'}(y^{(i)})}{\prod_{i=1}^p f_{Y_{\setminus i}}(y^{(i)})} \\ &= f_{Y'}(y') \frac{f_{Y'}(y_1, y'_2, \dots, y'_p)}{f_{Y_{\setminus 1}}(y'_2, \dots, y'_p)} \frac{f_{Y'}(y_1, y_2, y'_3, \dots, y'_p)}{f_{Y_{\setminus 2}}(y_1, y'_3, \dots, y'_p)} \dots \frac{f_{Y'}(y)}{f_{Y_{\setminus p}}(y_1, y_2, \dots, y_{p-1})} \\ &= f_{Y'}(y') f_{Y_{\setminus 1} | Y_2, \dots, Y_p}(y_1 | y'_2, \dots, y'_p) \cdot \dots \cdot f_{Y_p | Y_1, \dots, Y_{p-1}}(y_p | y_1, \dots, y_{p-1}) \\ &= \prod_{i=0}^p f^{(i)}(y^{(i)}) \end{aligned}$$

Para vectores Y, Y' , se define, para cada $i = 2, \dots, p$, $Y_{(i)} = (Y'_1, \dots, Y'_{i-1}, Y_i, \dots, Y_p)$. Por conveniencia

$$\begin{aligned} Y_{(1)} &= (Y_1, \dots, Y_p) = Y, \\ Y_{(p)} &= Y'. \end{aligned}$$

Se sigue de la igualdad anterior que

$$\begin{aligned}
 K^*(y', y) &= \frac{\prod_{i=0}^p f_{Y_i}(y^{(i)})}{\prod_{i=1}^p f_{Y_i}(y^{(i)}, y_i)} \\
 &= f_{Y_1}(y') \frac{f_{Y_2}(y)}{f_{Y_2, \dots, Y_p}(y_2, \dots, y_p)} \cdot \frac{f_{Y_3}(y'_1, y_2, \dots, y_p)}{f_{Y_3, Y_3, \dots, Y_p}(y_1, y_3, \dots, y_p)} \\
 &\quad \dots \frac{f_{Y_p}(y'_1, \dots, y'_{p-1}, y_p)}{f_{Y'_1, \dots, Y'_{p-1}}(y'_1, \dots, y'_{p-1})} \\
 &= f_{Y_1}(y') f_{Y_2|Y_2, \dots, Y_p}(y_1 | y_2, \dots, y_p) \cdot \dots \cdot f_{Y_p|Y'_1, \dots, Y'_{p-1}}(y_p | y'_1, \dots, y'_{p-1}) \\
 &= \prod_{i=0}^p f^{(i)}(y^{(i)}).
 \end{aligned}$$

el cual es precisamente el kernel del operador de transición (con respecto a μ) para el M.S.S. en el orden opuesto \square .

Se demuestra en el siguiente Teorema que las iteraciones sucesivas de M.S.S. convergen fuertemente a f_Y y que la tasa de convergencia es geométrica en el espacio normado de Hilbert.

Teorema 3.9 Con la notación anterior y suponiendo que se cumple para el operador T la propiedad de tipo Hilbert-Schmidt (Definición 3.13) y para alguna $\mu > 0$ con respecto a $\mu \times \mu$. Existe un número $c \in [0, 1)$ tal que, para cada densidad $u_0 \in H$, la sucesión de funciones $u_n = T(u_{n-1}) = T^n(u_0)$ para $n = 1, 2, \dots$ satisface

$$\|u_n - f_Y\| \leq \|u_0\| c^n$$

para toda n .

Demostración. Sea $U = T^*T$, que es, U el operador correspondiente a dos iteraciones del M.S.S., una hecha en orden inverso y la otra en el orden natural. De acuerdo al Teorema 3.5 (Berberian (1961)) U es auto-adjunto; también es completamente continuo porque T es continuo. De acuerdo al Teorema 3.8 H tiene una base ortonormal finita o infinita de eigenfunciones. Es fácil ver entonces que, para toda f

$$\int f(y) d\lambda^p(y) = \int T(f)(y) d\lambda^p(y)$$

$$\int f(y) d\lambda^p(y) = \int T^*(f)(y) d\lambda^p(y)$$

de lo cual se sigue que

$$\int f(y) d\lambda^p(y) = \int U(y) d\lambda^p(y). \quad (3.16)$$

Se define el siguiente operador como

$$V(f) = U(f) - f_Y \langle f_Y, f \rangle,$$

en particular

$$\begin{aligned} V(f_Y) &= U(f_Y) - \langle f_Y, f_Y f_Y \rangle \\ &= f_Y - f_Y \int f_Y d\lambda^p(y) \\ &= 0. \end{aligned}$$

Entonces V es auto-adjunto y completamente continuo, y su norma es el valor absoluto de su eigenvalor más grande. Además, es el segundo valor más grande de eigenvalores de U . Entonces basta probar que $\|V\| < 1$. Sea r el eigenvalor más grande de V y $V(g) = rg$. Si $r = 0$ la demostración es trivial, así que se supone que $r \neq 0$. Pero antes se demuestra que

$$\begin{aligned} \langle g, f_Y \rangle &= \frac{1}{r} \langle V(g), f_Y \rangle \\ &= \frac{1}{r} \langle g, V(f_Y) \rangle \\ &= 0. \end{aligned}$$

Dado que g no es idénticamente 0, puede escribirse como la suma de sus partes negativa y positiva, $g = g^+ - g^-$, definidos por los siguientes conjuntos

$$\begin{aligned} B &= \{y : g(y) > 0\}, \\ C &= \{y : g(y) < 0\}. \end{aligned}$$

Sea $0 = \langle f_Y, g \rangle = \langle f_Y, g^+ - g^- \rangle = \langle f_Y, g^+ \rangle - \langle f_Y, g^- \rangle$, como g no es cero con probabilidad uno y $f_Y > 0$ entonces $\langle f_Y, g^+ \rangle > 0$ y $\langle f_Y, g^- \rangle > 0$ las medidas de B y C son positivas ($\lambda^p(B) > 0$ y $\lambda^p(C) > 0$). Se muestra por contradicción que $|r| < 1$, suponiendo que $|r| = 1$ entonces

$$V(g) = U(g) = U(g^+ - g^-)$$

$$\begin{aligned}
&= U(g^+) - U(g^-) \\
&= \begin{cases} g^+ - g^- & \text{si } r = 1 \\ g^- - g^+ & \text{si } r = -1 \end{cases}
\end{aligned}$$

Dado que $k > 0$ (el operador 3.15), se sigue que $U(g^+)(y) > 0$ y $U(g^-)(y) > 0$ para toda y . De aquí que

$$\begin{aligned}
g^+(y) &< \begin{cases} U(g^+)(y) & \text{si } r = 1 \\ U(g^-)(y) & \text{si } r = -1 \end{cases} \quad \text{para } y \in B, \\
g^-(y) &< \begin{cases} U(g^-)(y) & \text{si } r = 1 \\ U(g^+)(y) & \text{si } r = -1 \end{cases} \quad \text{para } y \in C.
\end{aligned}$$

Como consecuencia $U(g^+ + g^-) > g^+ + g^-$ para toda y . Como $|g| = g^+ + g^-$, esto quiere decir que

$$U(|g|) > |g|,$$

por lo cual

$$\int U(|g|) d\lambda^p > \int |g| d\lambda^p,$$

que contradice la propiedad 3.16. De aquí que $|r| < 1$ y también que el valor más grande del operador V tenga valor absoluto $|r| < 1$ y por tanto se concluye que $\|V\| = |r| < 1$.

Si $W(f) = T(f) - f_Y < f_Y \cdot f >$ el operador W^*W es fácil ver que $V = W^*W$

$$\begin{aligned}
W^*W(f) &= W^*(T(f) - f_Y < f_Y \cdot f >) \\
&= T^*[T(f) - f_Y < f_Y \cdot f >] - f_Y < f_Y, T(f) - f_Y < f_Y \cdot f > > \\
&= T^*T(f) - f_Y < f_Y \cdot f > - f_Y \int [T(f)(y) - f_Y(y) \int f(x) d\lambda^p(x)] d\lambda^p(y) \\
&= T^*T(f) - f_Y < f_Y \cdot f > - f_Y \left[\int T(f)(y) d\lambda^p(y) - \int f_Y(y) \int f(x) d\lambda^p(x) d\lambda^p(y) \right] \\
&= T^*T(f) - f_Y < f_Y \cdot f > - f_Y \left[\int T(f)(y) d\lambda^p - \int f_Y(y) d\lambda^p \right] \\
&= U(f) - f_Y < f_Y \cdot f > - f_Y(0) \\
&= V(f).
\end{aligned}$$

De acuerdo al Teorema 3.3 (inciso 5), se verifica lo siguiente

$$\begin{aligned}
\|V\| &= \|W^*W\| \\
&= \|W\|^2
\end{aligned}$$

de donde

$$\begin{aligned}\|W\| &= |r|^{\frac{1}{2}} \\ &= c < 1.\end{aligned}$$

Si f es una densidad, entonces $\langle f_Y, f \rangle = 1$ y

$$\begin{aligned}W(f) &= T(f) - f_Y \\ &= T(f - f_Y).\end{aligned}$$

Si $\langle f_Y, g \rangle = 0$, entonces $W(g) = T(g)$ y $\langle f_Y, W(g) \rangle = 0$, de lo cual se sigue que para toda n ,

$$W^n(g) = T^n(g).$$

Dado que

$$\langle f_Y, f - f_Y \rangle = \int (f(y) - f_Y(y)) d\lambda^p(y) = 0$$

para cada densidad f , se sigue para toda n

$$\begin{aligned}W^n(f) &= W^n(f - f_Y) \\ &= T^n(f - f_Y) \\ &= T^n(f) - f_Y.\end{aligned}$$

Así que

$$\begin{aligned}\|u_n - f_Y\| &= \|T^n(u_0) - f_Y\| \\ &\leq \|W\|^n \|u_0\| \\ &= c^n \|u_0\|\end{aligned}$$

lo cual finaliza la demostración. \square .

El Teorema 3.9 señala que una sucesión de funciones de densidad generada por las iteraciones converge fuertemente a la distribución de Y , f_Y , y por tanto converge débilmente.

Corolario 3.1 *Bajo las condiciones del Teorema 3.9. Sea F_0 una distribución absolutamente continua con respecto a μ y se supone que $dF_0/d\mu = u_0 \in H$. Sea F la distribución de Y . Sea F_0 la distribución muestreada al comenzar el M.S.S. y F_n la distribución después*

de n iteraciones. Existe un número $c \in [0, 1)$ tal que, para cada conjunto A de Borel en R^p ,

$$|F_n(A) - F(A)| \leq c^n \|u_0\|.$$

Demostración. Sea $u_n = T(u_{n-1})$ como en el Teorema 3.9 y sea c el número que garantiza el acotamiento. Sea A el conjunto de Borel en R^p y sea I_A su función indicadora. Se define $g = I_A f_Y$,

$$\begin{aligned} \|g\|^2 &= \begin{cases} \|I_A f_Y\|^2 = \|f_Y\|^2 & y \in A \\ 0 & y \notin A \end{cases} \\ &= F(A) \leq 1. \end{aligned}$$

Se tiene que

$$\begin{aligned} F(A) &= \int_A f_Y(y) d\lambda^p(y) \\ &= \int f_Y(y) I_A d\lambda^p(y) \\ &= \int f_Y(y)^2 I_A d\mu(y) \\ &= \langle f_Y, g \rangle, \end{aligned}$$

$$\begin{aligned} F_n(A) &= \int_A u_n(y) d\lambda^p(y) \\ &= \int u_n(y) I_A d\lambda^p(y) \\ &= \int f_Y(y) I_A u_n(y) d\mu(y) \\ &= \langle u_n, g \rangle. \end{aligned}$$

La desigualdad de Cauchy-Schwarz implica

$$\begin{aligned} |F_n(A) - F(A)| &= |\langle u_n, g \rangle - \langle f_Y, g \rangle| \\ &= |\langle u_n - f_Y, g \rangle| \\ &\leq \|u_n - f_Y\| \|g\| \\ &\leq c^n \|u_0\| \end{aligned}$$

□.

Existen varios trabajos que desarrollan estudios sobre la convergencia del *M.S.S.* Algunos de los más importantes son el realizado por Tanner y Wong (1987), el trabajo presentado por Chan (1993) sobre la ergodicidad geométrica del Muestreo Gibbs y tres criterios de convergencia que postulan Zellner y Min (1995), los cuales son aplicables de acuerdo a la investigación que se realiza.

Capítulo 4

SIMULACIONES

Dentro de este capítulo se presentan ejemplos de simulaciones para ilustrar en forma gráfica la inferencia descrita en el Capítulo 2 y la convergencia del *M.S.S.*

Todas las gráficas fueron generadas dentro del lenguaje de programación *S-Plus* y los listados de codificación pueden encontrarse en el Apéndice.

Ahora, brevemente se retoman algunos problemas surgidos dentro del capítulo 2. Empezando con ejemplificar, mediante el uso de tres diferentes densidades, el efecto de la densidad de Muestreo de Importancia en la aproximación de la integral

$$\int_0^{1.96} e^{-x^2/2} dx. \quad (4.17)$$

Aquí se retoma el problema de elegir la mejor densidad de Muestreo de Importancia para tener una buena aproximación.

Ejemplo 4.1 *En las tres primeras gráficas (figuras 4.3, 4.4 y 4.5) pueden verse las fluctuaciones del valor estimado al iterarse varias veces con diferentes tamaños de muestra. Las densidades usadas fueron, en orden respectivo: Exponencial con parámetro 2, Ji-cuadrada con 8 grados de libertad y Normal estándar; todas se iteraron 100 veces con los siguientes tamaños de muestra: 1000, 3000, 5000, 7000, 10000, 150000. La forma en que se calculó la estimación es similar a la de la ecuación (1.5) para cada caso.*

Se nota que para igual tamaño de muestra existen diferencias en cada densidad. Puede verse que la convergencia para la densidad Ji cuadrada es lenta comparada con las otras dos; y a su vez, en lo que respecta a la densidad Normal y Exponencial, la segunda presenta una convergencia más rápida. A esto se agrega la importancia del tamaño de muestra, por ejemplo, en el caso de la densidad Normal, se debe considerar un tamaño de muestra mayor comparado con el que se necesita para la densidad Exponencial, con el fin de tener valores similares en precisión.

Para el caso de la solución de la integral (4.17) se distingue en forma gráfica que, comparativamente, la densidad Exponencial es una buena densidad de Muestreo de Importancia para estimarla.

Después del cálculo de las estimaciones se procede a obtener las varianzas de éstas de acuerdo a lo descrito en el Capítulo 2, suponiendo a la densidad Normal como la densidad de Muestreo de Importancia adecuada. Esta suposición se hace para ejemplificar la posibilidad de elegir una variedad de formas para las funciones "g" y "p" dando por resultado diferencias en los valores calculados de las varianzas que se resumen en los siguientes dos ejemplos.

Ejemplo 4.2 *Se quiere calcular el valor esperado de una función de interés con respecto a una densidad posterior. El problema aparece en un caso concreto donde se multiplica y divide por la densidad de Muestreo de Importancia para obtener la esperanza de la función original, modificada por algunos valores más de acuerdo al método de integración Monte Carlo con una suma. Aquí debe identificarse cuál es la función de interés y cuál la densidad posterior porque influye en la estimación de la varianza.*

Si se estima el valor de la integral (4.17)

$$\int_0^{1.96} \exp(-x^2/2) dx$$

con una densidad Normal con media μ y varianza σ^2 como la densidad de Muestreo de Importancia, se calcula el valor estimado de la integral como sigue:

Caso 1. *Sean*

$$g(x) = \sqrt{2\pi} I_{(0,1.96)}(x), \quad p(x) = \exp\left\{\frac{-x^2}{2}\right\}, \quad MI^*(x) = \exp\left\{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\} \quad (4.18)$$

con

$$\begin{aligned} \int_0^{1.96} \exp(-x^2/2) dx &= \frac{\int_{-\infty}^{\infty} \sqrt{2\pi} I_{(0,1.96)}(x) \exp(-x^2/2) dx}{\int_{-\infty}^{\infty} \exp(-x^2/2) dx} \\ &= \frac{\int_{-\infty}^{\infty} \frac{\sqrt{2\pi} I_{(0,1.96)}(x) \exp(-x^2/2)}{\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx}{\int_{-\infty}^{\infty} \frac{\exp\left(-\frac{x^2}{2}\right)}{\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx} \\ &= \frac{E\left[\sqrt{2\pi} I_{(0,1.96)}(x) \exp\left\{-\frac{x^2}{2} + \frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}\right]}{E\left[\exp\left\{-\frac{x^2}{2} + \frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}\right]} \\ &\stackrel{2.5}{=} \frac{n^{-1} \sum_{i=1}^n \left[\sqrt{2\pi} I_{(0,1.96)}(X_i) \exp\left\{-\frac{X_i^2}{2} + \frac{1}{2}\left(\frac{X_i-\mu}{\sigma}\right)^2\right\}\right]}{n^{-1} \sum_{i=1}^n \left[\exp\left\{-\frac{X_i^2}{2} + \frac{1}{2}\left(\frac{X_i-\mu}{\sigma}\right)^2\right\}\right]} \end{aligned}$$

donde \bar{y}_n es

$$\bar{y}_n = \frac{n^{-1} \sum_{i=1}^n \left[\sqrt{2\pi} I_{(0,1.96)}(X_i) \exp\left\{-\frac{X_i^2}{2} + \frac{1}{2}\left(\frac{X_i-\mu}{\sigma}\right)^2\right\}\right]}{n^{-1} \sum_{i=1}^n \left[\exp\left\{-\frac{X_i^2}{2} + \frac{1}{2}\left(\frac{X_i-\mu}{\sigma}\right)^2\right\}\right]}$$

y

$$w(x) = \frac{p(x)}{MI^*(x)} = \exp\left\{-\frac{x^2}{2} + \frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}$$

(ecuación (2.1)), calculando la varianza como en la ecuación (2.5)

$$\hat{\sigma}_n^2 = \frac{\sum_{i=1}^n (g(X_i) - \bar{y}_n)^2 \exp\left\{2\left(-\frac{X_i^2}{2} + \frac{1}{2}\left(\frac{X_i-\mu}{\sigma}\right)^2\right)\right\}}{\left[\sum_{i=1}^n \exp\left\{-\frac{X_i^2}{2} + \frac{1}{2}\left(\frac{X_i-\mu}{\sigma}\right)^2\right\}\right]^2}$$

o también

Caso 2. Sean

$$g(x) = 2I_{(0,1.96)}(x)\epsilon^{x(1-x)/2}, \quad p(x) = \epsilon^{-x/2}, \quad MI^*(x) = \exp\left\{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\} \quad (4.19)$$

y \bar{y}_n de la siguiente forma

$$\bar{y}_n = \frac{n^{-1} \sum_{i=1}^n \left[2I_{(0,1.96)}(X_i) \exp\left\{-\frac{X_i^2}{2} + \frac{1}{2}\left(\frac{X_i-\mu}{\sigma}\right)^2\right\}\right]}{n^{-1} \sum_{i=1}^n \left[\exp\left\{-\frac{X_i^2}{2} + \frac{1}{2}\left(\frac{X_i-\mu}{\sigma}\right)^2\right\}\right]}$$

y

$$w(x) = \frac{p(x)}{MI} = \frac{e^{-X_i/2}}{e^{\frac{1}{2}(\frac{X_i - \bar{g}_n}{\sigma_n})^2}}$$

$$\hat{\sigma}_n^2 = \frac{\sum_{i=1}^n (g(X_i) - \bar{g}_n)^2 w(X_i)^2}{\left[\sum_{i=1}^n w(X_i) \right]^2}$$

Siguiendo con lo anterior, se presentan a continuación dos gráficas (figuras (4.6) y (4.7)) que exhiben las varianzas de las estimaciones para los casos 1 y 2 anteriores, ambas confrontan los valores estimados de la varianza (eje Y) contra el tamaño de muestra (eje X), la primera toma el caso 1 y la segunda el caso 2.

Nótese que los valores de la gráfica 4.7 son más pequeños comparados con los de la gráfica 4.6 aún partiendo de que son estimados con la misma densidad de Muestreo de Importancia, pero con diferencias en la asignación de las funciones "g" y "p".

Otro ejemplo interesante surgió con la suposición de que la densidad Normal es una buena densidad de Muestreo de Importancia para el caso de la aproximación de la integral (4.17). No sólo con considerarla una buena densidad de acuerdo a las estimaciones obtenidas, proporciona información suficiente acerca de los valores que pudiesen tomar sus parámetros, con el fin de minimizar la varianza de las estimaciones y asegure una buena aproximación. Véase el siguiente ejemplo que describe lo anterior.

Ejemplo 4.3 1. *El primer caso es con una densidad Normal estándar siguiendo el mismo desarrollo como en ecuación (1.5) con las funciones g, p, MI y w definidas como:*

$$g(x) = \sqrt{2\pi} J_{(0,1,96)}(x),$$

$$p(x) = e^{-x^2/2},$$

$$MI(x) = e^{-x^2/2},$$

$$w(x) = 1,$$

con varianza estimada

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n \left(\sqrt{2\pi} J_{(0,1,96)}(X_i) - \bar{g}_n \right)^2}{n^2}$$

contra;

2. una densidad Normal no estándar con media diferente de cero y varianza unitaria detallándose los cálculos para este caso:

se definen las funciones g , p y MI como

$$\begin{aligned} g(x) &= I_{(0,1.96)}(x)\sqrt{2\pi}, \\ p(x) &= \exp\{-x^2/2\}, \\ MI(x) &= \exp\left\{-\frac{(x-\mu)^2}{2}\right\}, \end{aligned}$$

si se completa la ecuación (1.3) de la siguiente forma

$$\begin{aligned} \int_0^{1.96} e^{-x^2/2} dx &= \frac{\int_{-\infty}^{\infty} \frac{I_{(0,1.96)}(x) \exp\{-x^2/2\} \exp\left\{-\frac{(x-\mu)^2}{2}\right\}}{\exp\left\{-\frac{(x-\mu)^2}{2}\right\}} dx}{\int_{-\infty}^{\infty} \frac{\exp\{-x^2/2\} \exp\left\{-\frac{(x-\mu)^2}{2}\right\}}{\exp\left\{-\frac{(x-\mu)^2}{2}\right\}} dx} \\ &= \frac{E[\sqrt{2\pi} I_{(0,1.96)}(x) \exp\{-x\mu + \mu^2/2\}]}{E[\exp\{-x\mu + \mu^2/2\}]} \\ &\stackrel{c.s.}{=} \frac{\sum_{i=1}^n \sqrt{2\pi} I_{(0,1.96)}(X_i) \exp\{-X_i\mu + \mu^2/2\}}{\sum_{i=1}^n \exp\{-X_i\mu + \mu^2/2\}} \\ &= \bar{g}_n. \end{aligned}$$

con función de peso y varianza igual a

$$\begin{aligned} w(x) &= \exp\{-x\mu + \mu^2/2\}, \\ \hat{\sigma}^2 &= \frac{\sum_{i=1}^n (g(X_i) - \bar{g}_n)^2 (\exp\{-X_i\mu + \mu^2/2\})^2}{\left(\sum_{i=1}^n \exp\{-X_i\mu + \mu^2/2\}\right)^2}. \end{aligned}$$

Las 2 gráficas siguientes (4.8) muestran las estimaciones de las varianzas (ecuación 1.5) para cada caso. Se puede observar que existen diferencias entre los valores estimados siendo que provienen de la misma densidad de Muestreo de Importancia pero con diferente parámetro, demostrando así que es importante encontrar una densidad de Muestreo de Importancia adecuada y no sólo esto, sino, que además los valores de los parámetros de dicha densidad con los cuales se tenga una mejor aproximación.

Como una consecuencia de no elegir adecuadamente las funciones "g" y "p" con el fin de minimizar la varianza, se observa que el definir las dos funciones anteriores puede no

ser lo más conveniente de acuerdo a lo obtenido en el ejemplo 4.2, con fluctuaciones de los valores estimados. A esto puede anexarse un nuevo tema de investigación; en este caso se podría sugerir tomar sólo una función $H(x)$ en lugar de descomponerla y así se calcula el cociente $H(x)/MI$ bajo el dominio de la función MI , tema que no pudo continuarse por el límite del tiempo. Pero no solo el encontrar gráficamente una buena densidad de Muestreo de Importancia asegura tener buenas aproximaciones, falta la especificación de los valores óptimos que pudiesen tomar los parámetros de dicha densidad, como se explica en el ejemplo 4.3.

En el ejemplo siguiente se describen varias gráficas. En ellas se observa que la velocidad de convergencia de la densidad estimada depende de tres variables: el tamaño de muestra, el número de iteraciones y el valor inicial.

Ejemplo 4.4 *Siguiendo el desarrollo de la ecuación (1.10) se calcula la densidad marginal de una densidad normal bivariada generando valores de las densidades condicionadas y después se aplica la integración Monte Carlo. La densidad normal bivariada tiene como parámetros el vector de medias siguiente $\mu' = (12, 12)$ y matriz de varianzas $\Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 4 \end{pmatrix}$.*

1. *En las gráficas contenidas en la figura 4.9 se tienen las variaciones que presentan las densidades estimadas al fijar el número de iteraciones y variar el tamaño de muestra junto con el valor inicial. El primer renglón corresponde a un tamaño de muestra pequeño con tres valores iniciales que parten del valor real de la media a valores completamente alejados de ésta. En los siguientes renglones se incrementa el tamaño de muestra, también se nota que en el último renglón, a pesar de tener un valor inicial grande y un número de iteraciones fijas, la aproximación es mejor debido al número de elementos muestreados.*
2. *Las siguientes 9 gráficas, figura 4.10, corresponden a un tamaño de muestra fijo contra variaciones en las iteraciones (renglones, de menor a mayor) y valores iniciales (columnas, de menor a mayor). El primer renglón corresponde a iteraciones con valores que parten del exacto a uno muy alejado del valor real. Se nota de que al tener*

sólo una iteración se presenta una aproximación regular, y observando las siguientes gráficas se ve que las aproximaciones son mejores conforme aumenta el número de iteraciones, aún para valores iniciales alejados del verdadero.

3. *En las gráficas de la figura 4.11, se fija un valor inicial grande, variando el número de iteraciones (renglones) y tamaños de muestra (columnas). Puede distinguirse que la densidad estimada con valor inicial grande se aproxima a la verdadera densidad conforme aumenta el número de iteraciones y el tamaño de muestra.*

Con las figuras anteriores se muestra que los valores obtenidos después de un proceso iterativo convergen a la verdadera densidad, comprobando en forma gráfica el resultado del Capítulo 3.

Otra forma de ver la convergencia se basa en la propuesta por Geweke (1989) que consiste en los cuartiles para poder monitorear la convergencia con diferentes tamaños de muestra y que se discute en el siguiente ejemplo.

Ejemplo 4.5 *Siguiendo el algoritmo de Muestreo Gibbs con tamaño de muestra variable, (pág. 19): al incrementar el tamaño de muestra, indica que en forma aleatoria se obtengan los datos que generarán a los nuevos datos, hasta completar el nuevo tamaño de muestra. La gráfica 4.12 presenta las estimaciones de los tres cuartiles de las densidades marginales de una normal bivariable descrita como en el ejemplo 4.4 pero con covarianza igual a 1, con los tamaños de muestra siguientes (100,500,5000,10000,15000), y por cada tamaño 10 iteraciones.*

El eje X corresponde al número de iteraciones y el eje Y corresponde los valores de los cuartiles estimados, los tamaños de muestra varían de acuerdo al comportamiento asintótico de los cuartiles.

En el algoritmo para tamaño de muestra variable, se observa que al generar la muestra anterior que dará origen a la nueva muestra se pierde información que podría considerarse como buena y en forma paralela podría tomar valores generadores "malos", y en el peor de

los casos hasta más de una ocasión (por la forma aleatoria en que son escogidos). Una forma alternativa de obtener el vector generador podría ser "pegar" los datos anteriores hasta tener el nuevo tamaño de muestra. Así, los valores más aproximados podrían considerarse al menos una vez y no habría desperdicio de información. Vea la gráfica 4.13 que describe lo anterior tomando los mismos parámetros de las gráfica (4.12).

El graficar sólo los cuartiles no asegura que el proceso de muestrear haya alcanzado una convergencia satisfactoria, porque los valores de los cuartiles estarán cerca a los reales pero en caso de tener un tamaño de muestra pequeño, los valores estimados de los cuantiles extremos (por decir los deciles) tendrán fluctuaciones más grandes. Surge así la idea de monitorear la convergencia de algunos de los cuantiles extremos observando lo anterior en el siguiente ejemplo haciendo uso del algoritmo de Muestreo Gibbs con tamaño de muestra variable.

Ejemplo 4.6 *En las siguientes gráfica (fig. 4.14) se verán valores estimados de los cuantiles (0.01,0.5,0.99) y su estimación después de cierto número de iteraciones.*

Los siguientes son los parámetros que utilizaron para la generación de las gráficas de dos cuantiles extremos y la mediana de X e Y , vector de medias $(12,12)$; $var=(1,0,1)$; valor inicial 0; cuantiles $(0.01,0.5,0.99)$; tamaño de muestra $(100,500,5000,10000,150000)$, y por cada tamaño 10 iteraciones.

Veáse la gráfica 4.14 que ilustra la convergencia de las densidades marginales estimadas después de 50 iteraciones. Conforme aumenta el tamaño de muestra los cuantiles extremos van convergiendo y el valor de la mediana se encuentra en un proceso estacionario. De acuerdo a esto puede sugerirse la graficación de los valores para cuantiles extremos para monitorear su convergencia, en vez de los cuartiles solamente. Es intuitivamente claro que cuando hayan convergido los extremos, los centrales ya lo habrán hecho. De acuerdo a lo anterior puede suponerse que para el tamaño de muestra en el que los extremos tengan un buen comportamiento servirá para estimar la verdadera densidad.

Y con esto finaliza la sección de simulaciones. Se hace énfasis en que las gráficas anteriores fueron simuladas en el lenguaje de programación *S-Plus* y los listados correspondientes se encuentran en el Apéndice.

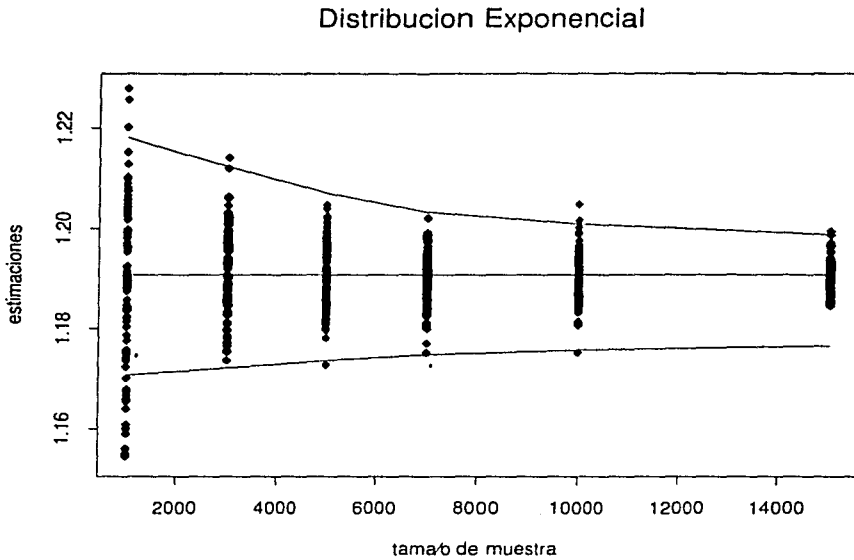


FIGURA 4.3: Estimación de la integral (4.1) con diferentes tamaños de muestra utilizando a la distribución Exponencial como la distribución de Muestreo por Importancia.

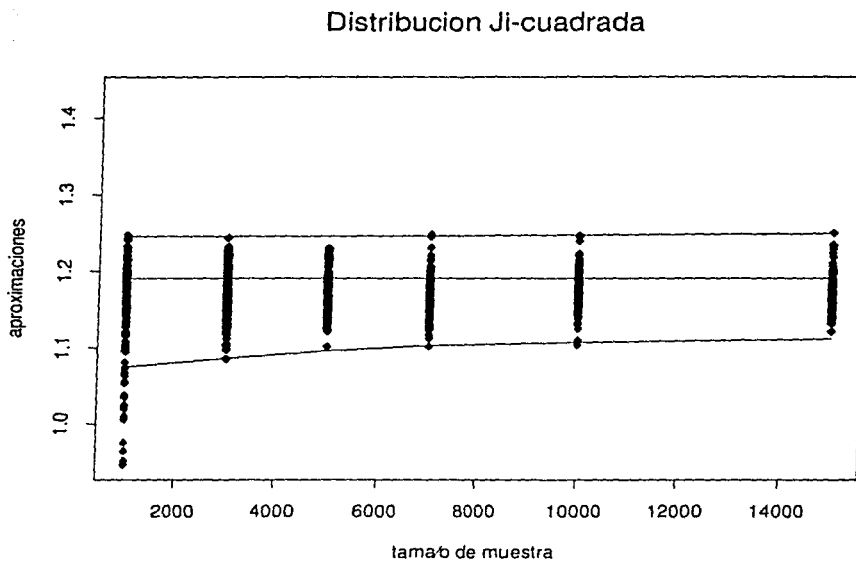


FIGURA 4.4: Estimación de la integral (4.1) con diferentes tamaños de muestra utilizando a la distribución Ji cuadrada (8 g.l.) como distribución de Muestreo por Importancia.

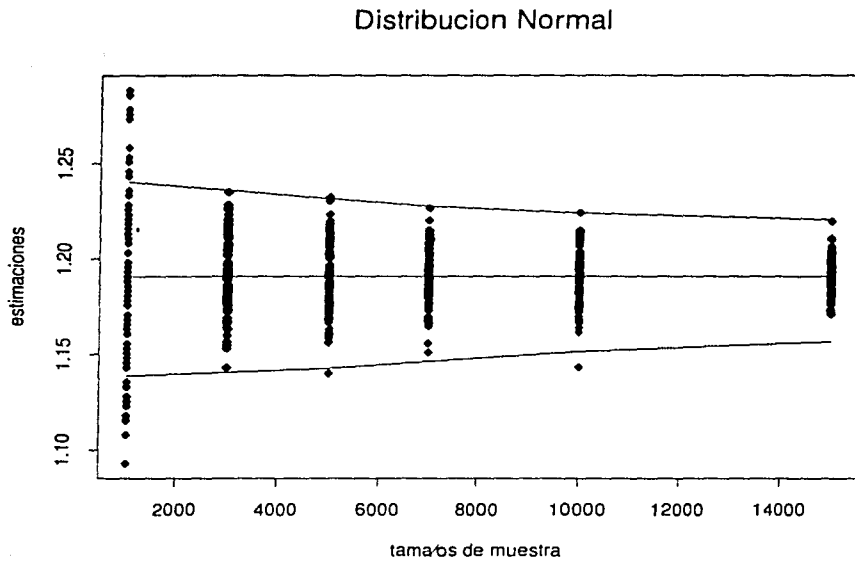


FIGURA 4.5: Estimación de la Integral (4.1) con diferentes tamaños de muestra utilizando a la distribución Normal (estándar) como distribución de Muestreo por Importancia.

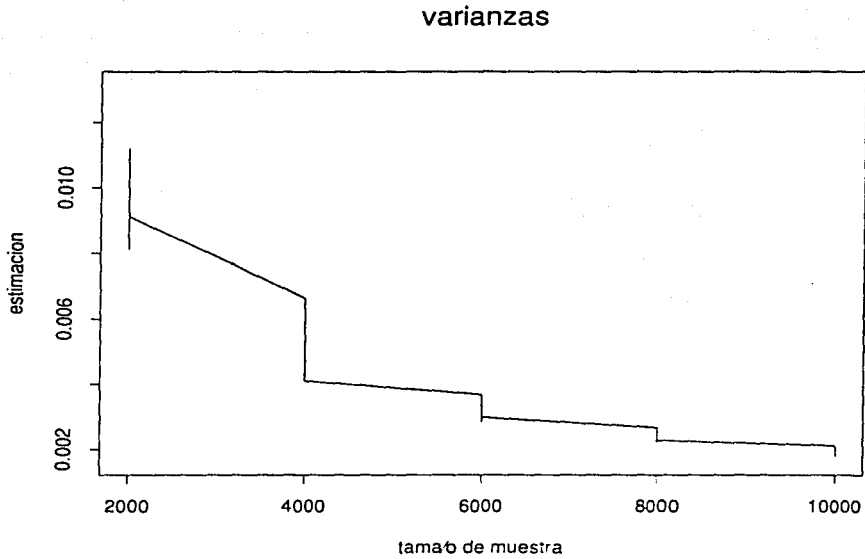


FIGURA 4.6: Estimación de la varianza para diferentes tamaños de muestra, con funciones "g" y "p" asignadas en la forma del caso 1 (ejemplo 4.2).

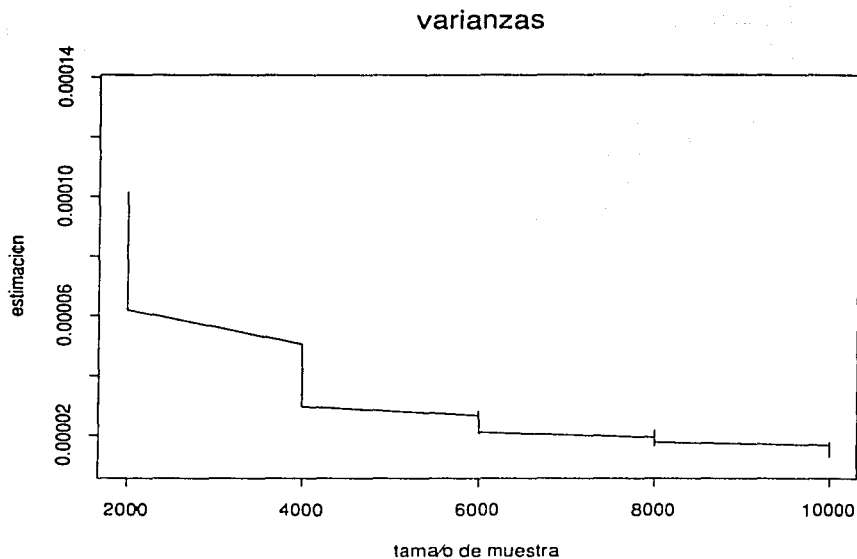


FIGURA 4.7: Estimación de la varianza para diferentes tamaños de muestra con funciones "g" y "p" asignadas como se discute en el caso 2 del ejemplo 4.2.

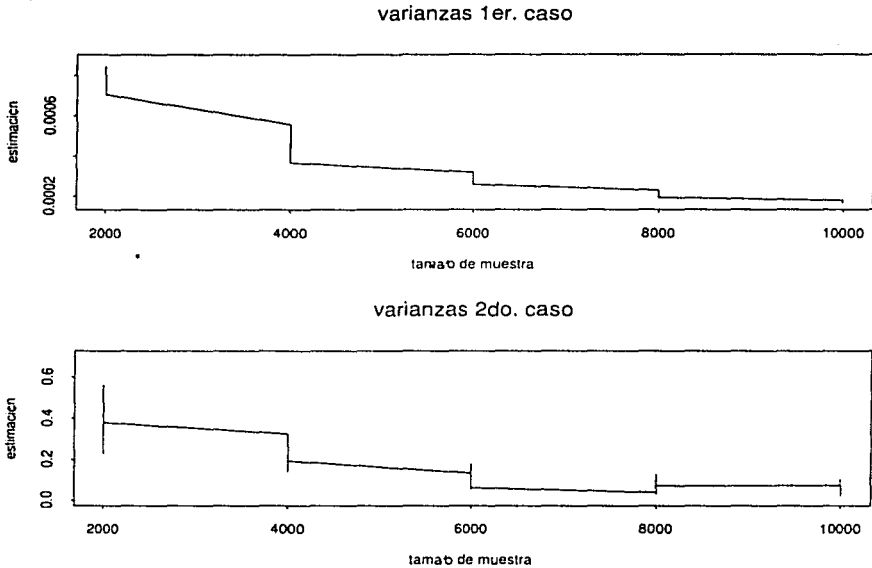


FIGURA 4.8: Gráficas correspondientes a las estimaciones de la varianza cuando se supone a la distribución Normal como la distribución de muestreo por importancia, pero, con diferentes valores paramétrales. La 1a. y 2a. gráfica representan los casos 1 y 2 del ejemplo 4.3, respectivamente.

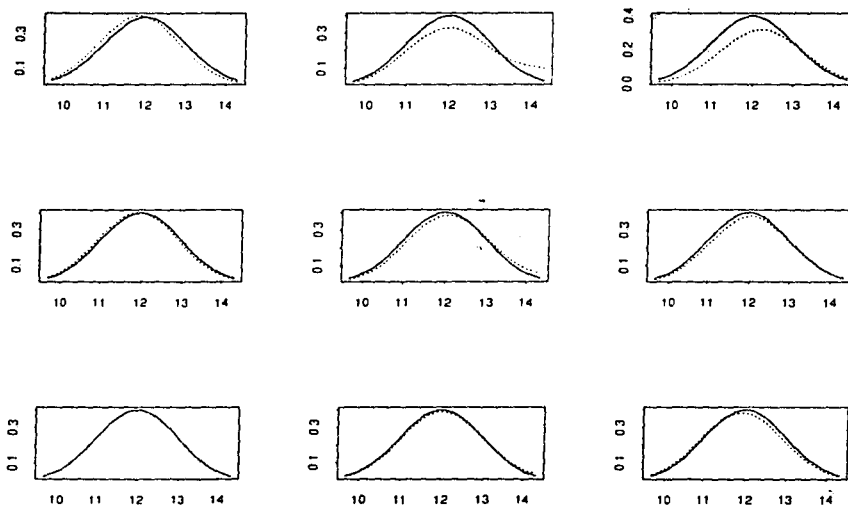


FIGURA 4.9: Densidad estimada (línea punteada) con variación en el tamaño de muestra (renglones, menor a mayor) y con diferentes valores iniciales (columnas, menor a mayor) teniendo un número de iteraciones fijo

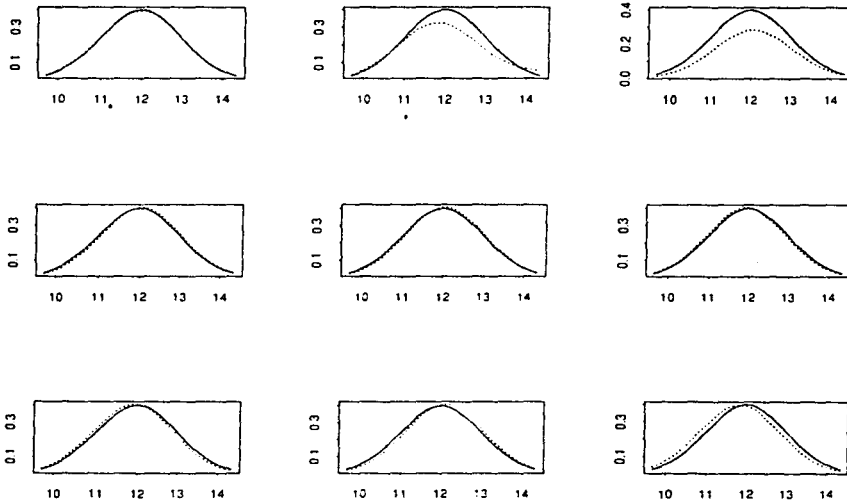


FIGURA 4.10: Estimación de la densidad estimada (línea punteada) con variación en el número de iteraciones (renglones, menor a mayor) y el valor inicial (columnas, menor a mayor) manteniendo fijo el tamaño de muestra.

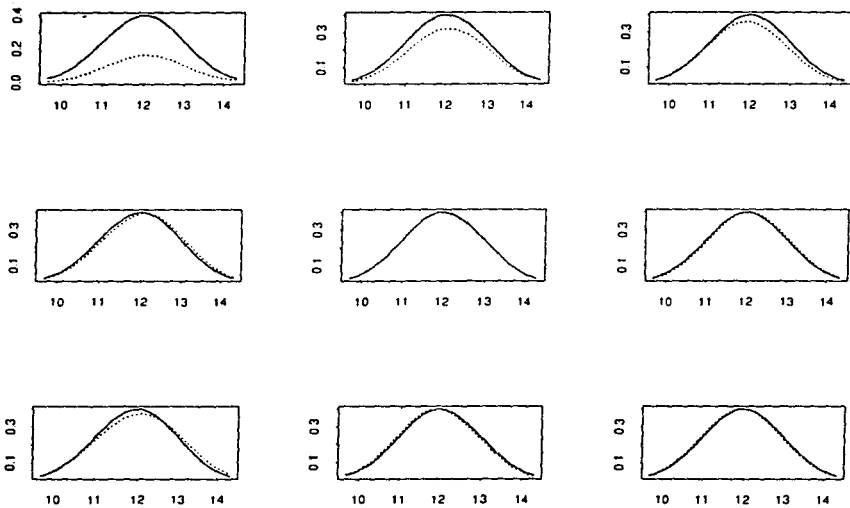


FIGURA 4.11: Densidad estimada (línea punteada) con diferentes tamaños de muestra (renglones) y valores iniciales (columnas) manteniendo fijo el valor inicial.

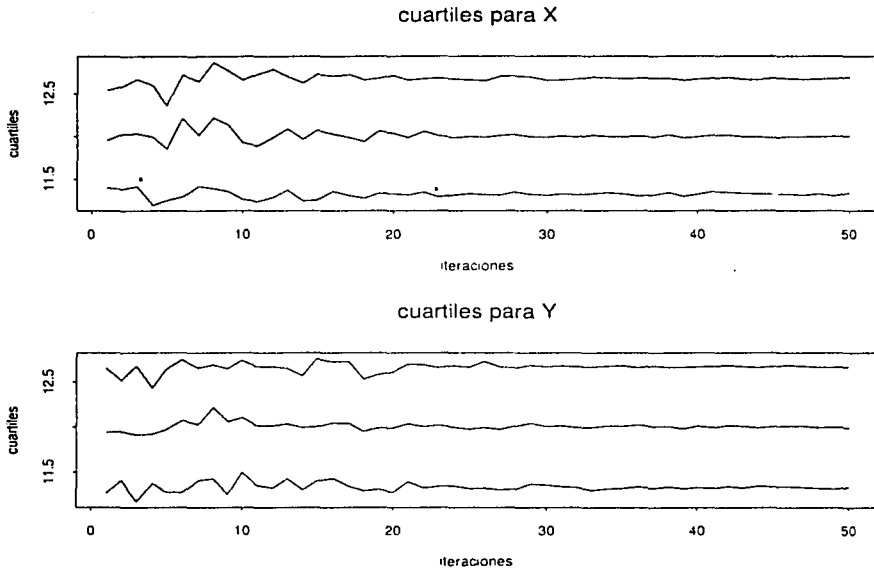


FIGURA 4.12: Comportamiento asintótico de los cuartiles estimados siguiendo el algoritmo de Muestreo Gibbs con tamaño variable.

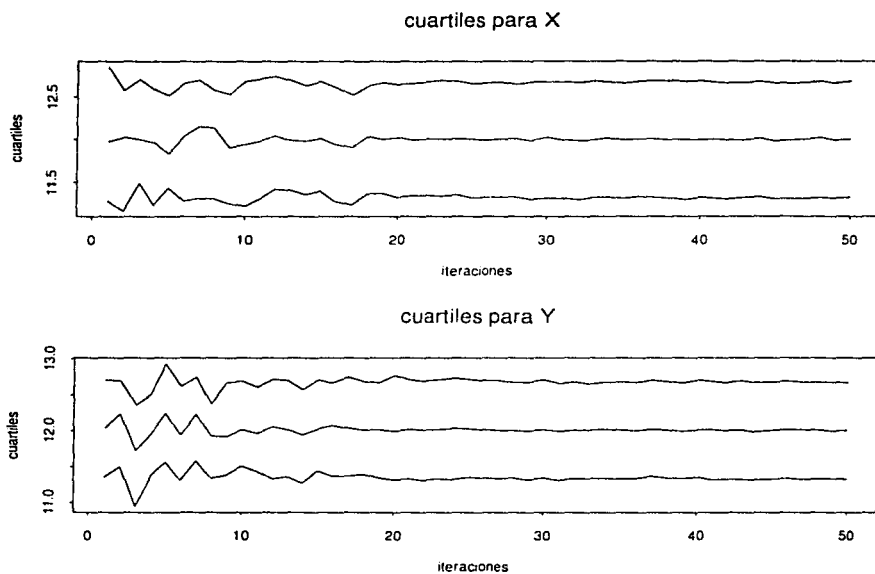


FIGURA 4.13: Comportamiento asintótico de los cuartiles bajo el algoritmo sugerido que "pega" los valores.

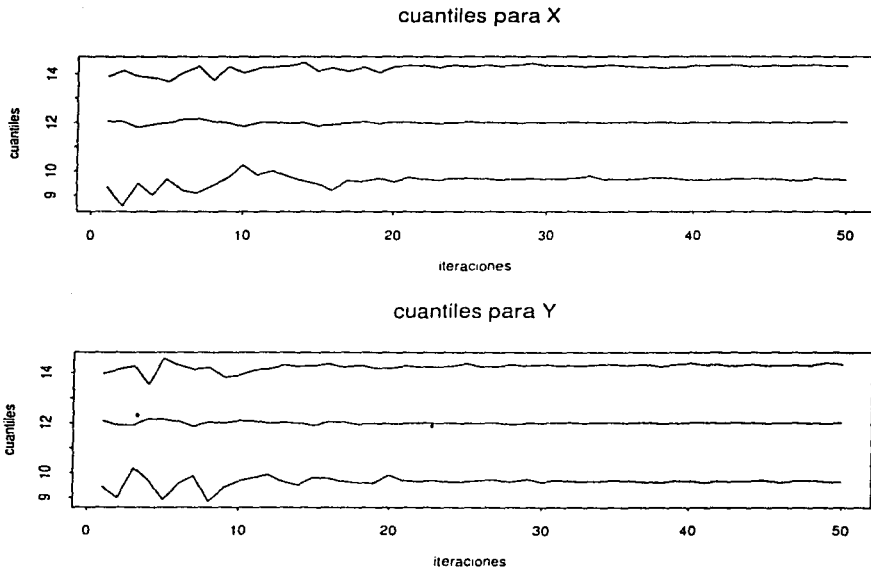


FIGURA 4.14: Comportamiento asintótico de cuantiles extremos después de 50 iteraciones bajo el algoritmo de Muestreo Gibbs con tamaño de muestra variable.

Capítulo 5

UN MODELO JERARQUICO.

5.1 INTRODUCCION

En este Capítulo se discute un problema de estimación, para ejemplificar la técnica de Muestreo Gibbs en un estudio realizado por Gelfand et al. (1990) cuyo objetivo es la estimación de las densidades marginales posteriores de los parámetros poblacionales en un grupo control de ratas.

Por conveniencia se hará uso de una notación diferente a la que se había llevado anteriormente: en los sucesivos las funciones de densidad de probabilidad serán denotadas por corchetes cuadrados, por ejemplo: $[X, Y]$ densidad conjunta, $[X | Y]$ densidad condicional, $[Y]$ densidad marginal y el proceso de marginalización por integración como

$$[X] = \int [X | Y] [Y].$$

El siguiente Teorema será utilizado más adelante.

Teorema 5.1 Sean X_1, X_2, \dots, X_n vectores aleatorios, entonces X_1, X_2, \dots, X_n son mutuamente independientes si y sólo si existen funciones $g_i(X_i), i = 1, \dots, n$, tales que la función de densidad conjunta de (X_1, X_2, \dots, X_n) puede ser escrita como

$$[X_1, X_2, \dots, X_n] = g_1(X_1) g_2(X_2) \dots g_n(X_n).$$

rata	x_{i1}	x_{i2}	x_{i3}	x_{i4}	x_{i5}	rata	x_{i1}	x_{i2}	x_{i3}	x_{i4}	x_{i5}
1	151	199	246	283	320	16	160	207	248	288	324
2	145	199	249	293	351	17	142	187	234	280	316
3	117	214	263	312	328	18	156	203	243	283	317
4	155	200	237	272	297	19	157	212	259	307	336
5	135	188	230	280	323	20	152	203	246	286	321
6	159	210	252	298	331	21	154	205	253	298	334
7	141	189	231	275	305	22	139	190	225	267	302
8	159	201	248	297	338	23	146	191	229	272	302
9	177	236	285	340	376	24	157	211	250	285	323
10	134	182	220	260	296	25	132	185	237	286	331
11	160	208	261	313	352	26	160	207	257	303	345
12	143	188	220	273	314	27	169	216	261	295	333
13	154	200	244	289	325	28	157	205	248	289	316
14	171	221	270	326	358	29	137	180	219	258	191
15	163	216	242	281	312	30	153	200	244	286	324

TABLA 5.3: Pesos de la pob. de ratas: grupo control donde $x_{i1} = 8, x_{i2} = 15, x_{i3} = 22, x_{i4} = 29, x_{i5} = 36$ denotan días: $i = 1, \dots, 30$.

5.2 IMPLEMENTACION DE UN MODELO JERARQUICO

El siguiente problema es un estudio conducido por la compañía CIBA-GEIGY sobre el crecimiento de una población de ratas en donde se registraron los pesos de un grupo control medidos durante cinco semanas. Los datos están disponibles en la tabla 5.3.

Para el periodo considerado se suponen curvas de crecimiento recto y también se supone homocedasticidad normal en los errores, así que

$$Y_{ij} \sim N(\alpha_i + \beta_i x_{ij}, \sigma_c^2), \quad i = 1, \dots, k; j = 1, \dots, n_i;$$

donde

k — número total de ratas;

n_i — número de mediciones de la i -ésima rata;

Y_{ij} — peso de la i -ésima rata cuando la medida j -ésima fué tomada;

x_{ij} — edad en días de la rata i en la medida j .

La estructura de la población es modelada suponiendo independencia condicional

$$\left(\begin{array}{c} \alpha_i \\ \beta_i \end{array} \right) \left| \left(\begin{array}{c} \alpha_c \\ \beta_c \end{array} \right), \Sigma_c \right. \sim N \left[\left(\begin{array}{c} \alpha_c \\ \beta_c \end{array} \right), \Sigma_c \right]; \quad i = 1, \dots, k. \quad (5.20)$$

Un análisis Bayesiano completo requiere de la especificación de una densidad a priori para $\sigma_c^2, \mu_c = (\alpha_c, \beta_c)'$ y Σ_c . Una inferencia típica incluiría el cálculo de las densidades marginales posteriores para los parámetros poblacionales (α_c, β_c) y la predicción del peso de las ratas en forma individual.

Para la especificación a priori, se supone independencia, tomando

$$[\mu_c, \Sigma_c^{-1}, \sigma_c^2] = [\mu_c] [\Sigma_c^{-1}] [\sigma_c^2]$$

para tener una forma Normal-Wishart-Gamma Inversa :

$$\begin{aligned} [\mu_c] &= N(\eta, C), \\ [\Sigma_c^{-1}] &= W((\rho R)^{-1}, \rho), \\ [\sigma_c^2] &= IG\left(\frac{\gamma_0}{2}, \frac{\gamma_0 \tau_0^2}{2}\right). \end{aligned}$$

Reescribiendo el modelo para la i -ésima forma individual como

$$(Y_i | \theta_i, \sigma_c^2) \sim N(X_i \theta_i, \sigma_c^2 I_{n_i})$$

donde X_i es el vector de días de edad y $\theta_i = (\alpha_i, \beta_i)'$ el vector de parámetros para la rata i -ésima y definiendo

$$\theta = (\theta_1, \dots, \theta_k),$$

$$Y = (Y_1, \dots, Y_k),$$

$$\bar{\theta} = k^{-1} \sum_{i=1}^k \theta_i,$$

$$n = \sum_{i=1}^k n_i.$$

$$D_i = (\sigma_c^{-2} X_i' X_i + \Sigma_c^{-1})^{-1}, \quad (5.21)$$

$$V = (k \Sigma_c^{-1} + C^{-1})^{-1}; \quad (5.22)$$

entonces el muestreo Gibbs para $\theta = (\theta_1, \dots, \theta_k)$, Σ_c^{-1} y σ_c^2 (un total de 66 parámetros) es especificado con las densidades condicionales siguientes:

La función de densidad posterior para los valores de θ_i se obtiene de la siguiente forma:

$$[\theta | Y, \sigma_c^2, \mu_c, \Sigma_c^{-1}] \propto [\theta, Y, \sigma_c^2, \mu_c, \Sigma_c^{-1}]$$

donde la densidad conjunta puede obtenerse como

$$\begin{aligned} [\theta, Y, \sigma_c^2, \mu_c, \Sigma_c^{-1}] &= [Y | \theta, \sigma_c^2, \mu_c, \Sigma_c^{-1}] [\theta, \sigma_c^2, \mu_c, \Sigma_c^{-1}] \\ &= [Y | \theta, \sigma_c^2] [\theta | \sigma_c^2, \mu_c, \Sigma_c^{-1}] [\sigma_c^2, \mu_c, \Sigma_c^{-1}], \end{aligned}$$

de donde

$$[\theta, Y, \sigma_c^2, \mu_c, \Sigma_c^{-1}] \propto [Y | \theta, \sigma_c^2] [\theta | \mu_c, \Sigma_c^{-1}],$$

además las observaciones en las ratas (Y_i) son independientes entre sí (dado θ_i y σ_c^2), obteniendo la densidad conjunta como la multiplicación de las densidades marginales. Por otra parte, $\theta_1, \dots, \theta_k$ supone independencia condicional; entonces lo anterior se reescribe como:

$$\begin{aligned} [\theta | Y, \sigma_c^2, \mu_c, \Sigma_c^{-1}] &= \prod_{i=1}^k |2\pi\sigma_c^2 I_n|^{-1/2} \exp \left\{ -\frac{1}{2} (Y_i - X_i\theta_i)' (\sigma_c^2 I_n)^{-1} (Y_i - X_i\theta_i) \right\} * \\ &\quad \prod_{i=1}^k |2\pi \Sigma_c|^{-1/2} \exp \left\{ -\frac{1}{2} (\theta_i - \mu_c)' \Sigma_c^{-1} (\theta_i - \mu_c) \right\} \\ &= \prod_{i=1}^k \exp \left\{ - (2\sigma_c^2)^{-1} (Y_i - X_i\theta_i)' (Y_i - X_i\theta_i) \right\} * \\ &\quad \prod_{i=1}^k \left\{ -\frac{1}{2} (\theta_i - \mu_c)' \Sigma_c^{-1} (\theta_i - \mu_c) \right\}; \end{aligned}$$

como se tiene una multiplicación de funciones $g_i(\theta_i)$ y donde cada θ_i es independiente (por la independencia de las ratas) de acuerdo al Teorema 5.1, el producto define la densidad conjunta para θ y se calcula la f.d.p. para cada θ_i .

$$[\theta_i | Y_i, \sigma_c^2, \mu_c, \Sigma_c^{-1}] \propto \exp \left\{ -\frac{1}{2\sigma_c^2} (Y_i - X_i\theta_i)' (Y_i - X_i\theta_i) - \frac{1}{2} (\theta_i - \mu_c)' \Sigma_c^{-1} (\theta_i - \mu_c) \right\}$$

y tan solo se trabaja con los valores bajo la función exponencial

$$\sigma_c^{-2} (Y_i - X_i\theta_i)' (Y_i - X_i\theta_i) + (\theta_i - \mu_c)' \Sigma_c^{-1} (\theta_i - \mu_c) =$$

$$\begin{aligned}
&= \theta_i' \left(\sigma_c^{-2} X_i' X_i + \sum_c^{-1} \right) \theta_i - 2\theta_i' \left(\sigma_c^{-2} X_i' Y_i + \sum_c^{-1} \mu_c \right) + \sigma_c^{-2} Y_i' Y_i + \mu_c' \sum_c^{-1} \mu_c \\
&= \theta_i' D_i \theta_i - 2\theta_i' \left(\sigma_c^{-2} X_i' Y_i + \sum_c^{-1} \mu_c \right) + \sigma_c^{-2} Y_i' Y_i + \mu_c' \sum_c^{-1} \mu_c \\
&= \left(\theta_i - D_i \left(\sigma_c^{-2} X_i' Y_i + \sum_c^{-1} \mu_c \right) \right)' D_i^{-1} \left(\theta_i - D_i \left(\sigma_c^{-2} X_i' Y_i + \sum_c^{-1} \mu_c \right) \right)
\end{aligned}$$

donde D_i se definió en la ecuación (5.21), entonces

$$\left[\theta_i \mid Y_i, \sigma_c^2, \mu_c, \sum_c^{-1} \right] \sim N \left(D_i \left(\sigma_c^{-2} X_i' Y_i + \sum_c^{-1} \mu_c \right), D_i \right).$$

Se continua con el cálculo de la densidades posteriores para μ_c, \sum_c^{-1} y σ_c^2 haciendo uso de los Teoremas 1.4 y 5.1.

Densidad posterior de μ_c .

$$\begin{aligned}
\left[\mu_c \mid Y, \sigma_c^2, \theta, \sum_c^{-1} \right] &\propto \left[\theta, Y, \sigma_c^2, \mu_c, \sum_c^{-1} \right] \\
&= \left[Y \mid \theta, \sigma_c^2, \mu_c, \sum_c^{-1} \right] \left[\theta, \sigma_c^2, \mu_c, \sum_c^{-1} \right] \\
&= \left[Y \mid \theta, \sigma_c^2 \right] \left[\theta \mid \sigma_c^2, \mu_c, \sum_c^{-1} \right] \left[\sigma_c^2, \mu_c, \sum_c^{-1} \right] \\
&\propto \left[\theta \mid \mu_c, \sum_c^{-1} \right] \left[\mu_c \right],
\end{aligned}$$

que es igual a

$$\begin{aligned}
\left[\mu_c \mid Y, \sigma_c^2, \theta, \sum_c^{-1} \right] &= \prod_{i=1}^k \left| 2\pi \sum_c \right|^{-1/2} \exp \left\{ -\frac{1}{2} (\theta_i - \mu_c)' \sum_c^{-1} (\theta_i - \mu_c) \right\} * \\
&\quad \left| 2\pi C \right|^{-1/2} \exp \left\{ -\frac{1}{2} (\mu_c - \eta)' C^{-1} (\mu_c - \eta) \right\} \\
&\propto \prod_{i=1}^k \exp \left\{ -\frac{1}{2} (\theta_i - \mu_c)' \sum_c^{-1} (\theta_i - \mu_c) \right\} * \\
&\quad \exp \left\{ -\frac{1}{2} (\mu_c - \eta)' C^{-1} (\mu_c - \eta) \right\}
\end{aligned}$$

trabajando con el exponente queda como

$$\begin{aligned}
&\sum_{i=1}^k (\theta_i - \mu_c)' \sum_c^{-1} (\theta_i - \mu_c) + (\mu_c - \eta)' C^{-1} (\mu_c - \eta) = \\
&= \mu_c' \left(k \sum_c^{-1} + C^{-1} \right) \mu_c - 2\mu_c' \left(\sum_c^{-1} k \bar{\theta} + C^{-1} \eta \right) + \sum_c^{-1} \sum_{i=1}^k \theta_i^2 + \eta' C^{-1} \eta \\
&= \mu_c' V \mu_c - 2\mu_c' V^{-1} V \left(\sum_c^{-1} k \bar{\theta} + C^{-1} \eta \right) + \sum_c^{-1} \sum_{i=1}^k \theta_i^2 + \eta' C^{-1} \eta \\
&= \left(\mu_c - V \left(\sum_c^{-1} k \bar{\theta} + C^{-1} \eta \right) \right)' V^{-1} \left(\mu_c - V \left(\sum_c^{-1} k \bar{\theta} + C^{-1} \eta \right) \right),
\end{aligned}$$

donde V es como en la ecuación (5.22), entonces

$$[\mu_c | Y, \sigma_c^2, \theta, \Sigma_c^{-1}] \sim N \left(V \left(\Sigma_c^{-1} k \bar{\theta} + C^{-1} \eta \right), V \right).$$

Densidad posterior para Σ_c^{-1} .

$$\begin{aligned} [\Sigma_c^{-1} | Y, \sigma_c^2, \theta, \mu_c] &\propto [\Sigma_c^{-1}, Y, \sigma_c^2, \theta, \mu_c] \\ &= [Y | \theta, \sigma_c^2] [\theta | \sigma_c^2, \mu_c, \Sigma_c^{-1}] [\sigma_c^2, \mu_c, \Sigma_c^{-1}] \\ &\propto [\theta | \mu_c, \Sigma_c^{-1}] [\Sigma_c^{-1}]. \end{aligned}$$

quedando de la siguiente manera

$$\begin{aligned} [\Sigma_c^{-1} | Y, \sigma_c^2, \theta, \mu_c] &\propto \prod_{i=1}^k |2\pi \Sigma_c|^{-1/2} \exp \left\{ -\frac{1}{2} (\theta_i - \mu_c)' \Sigma_c^{-1} (\theta_i - \mu_c) \right\} * \\ &\quad |\Sigma_c^{-1}|^{-\left(\frac{n-p-1}{2}\right)} \exp \left\{ -\frac{1}{2} \text{traza} (\rho R \Sigma_c^{-1}) \right\} \\ &= |\Sigma_c^{-1}|^{\left(\frac{k+p-n-1}{2}\right)} \exp \left\{ \text{traza} \left(\sum_{i=1}^k (\theta_i - \mu_c) (\theta_i - \mu_c)' + \rho R \right) \Sigma_c^{-1} \right\}. \end{aligned}$$

entonces

$$[\Sigma_c^{-1} | Y, \sigma_c^2, \theta, \mu_c] \sim W \left(\left(\sum_{i=1}^k (\theta_i - \mu_c) (\theta_i - \mu_c)' + \rho R \right)^{-1}, k + \rho \right).$$

Finalmente la densidad posterior de σ_c^2 , que se obtuvo así

$$\begin{aligned} [\sigma_c^2 | Y, \Sigma_c^{-1}, \theta, \mu_c] &\propto [\Sigma_c^{-1}, Y, \sigma_c^2, \theta, \mu_c] \\ &= [Y | \theta, \sigma_c^2] [\theta | \sigma_c^2, \mu_c, \Sigma_c^{-1}] [\sigma_c^2, \mu_c, \Sigma_c^{-1}] \\ &\propto [Y | \theta, \sigma_c^2] [\sigma_c^2]. \end{aligned}$$

trabajando con lo anterior

$$\begin{aligned} [\sigma_c^2 | Y, \Sigma_c^{-1}, \theta, \mu_c] &\propto \prod_{i=1}^k |2\pi \sigma_c^2 I_n|^{-1/2} \exp \left\{ -\frac{1}{2} (Y_i - X_i \theta_i)' (\sigma_c^2 I_n)^{-1} (Y_i - X_i \theta_i) \right\} \\ &\quad \exp \left\{ -\frac{\gamma_0 \tau_0^2}{2\sigma_c^2} \right\} \left(\frac{1}{\sigma_c^2} \right)^{\frac{\gamma}{2} + 1} \\ &= \prod_{i=1}^k |2\pi \sigma_c^2 I_n|^{-1/2} \exp \left\{ -\frac{1}{2\sigma_c^2} \left((Y_i - X_i \theta_i)' (Y_i - X_i \theta_i) + \gamma_0 \tau_0^2 \right) \right\} \end{aligned}$$

$$\begin{aligned}
& * \left(\frac{1}{\sigma_c^2} \right)^{\frac{n}{2}+1} \\
& \propto (\sigma_c^2)^{-\sum_{i=1}^k n_i/2} \exp \left\{ -\frac{1}{2\sigma_c^2} \left(\sum_{i=1}^k (Y_i - X_i \theta_i)' (Y_i - X_i \theta_i) + \gamma_0 \tau_0^2 \right) \right\} \\
& * \left(\frac{1}{\sigma_c^2} \right)^{\frac{n}{2}+1} \\
& = \left(\frac{1}{\sigma_c^2} \right)^{\frac{n+\gamma_0}{2}+1} \exp \left\{ -\frac{1}{2\sigma_c^2} \left(\sum_{i=1}^k (Y_i - X_i \theta_i)' (Y_i - X_i \theta_i) + \gamma_0 \tau_0^2 \right) \right\},
\end{aligned}$$

entonces

$$[\sigma_c^2 | Y, \Sigma_c^{-1}, \theta, \mu_c] \sim GI \left(\frac{n + \gamma_0}{2}, \frac{1}{2} \left(\sum_{i=1}^k (Y_i - X_i \theta_i)' (Y_i - X_i \theta_i) + \gamma_0 \tau_0^2 \right) \right),$$

donde GI significa una función de densidad Gamma Inversa.

Para el análisis de crecimiento anterior, la especificación de los hiperparámetros a priori fue la siguiente

$$\begin{aligned}
C^{-1} &= 0, \\
\gamma_0 &= 0, \\
\rho &= 2, \\
R &= \begin{pmatrix} 100 & 0 \\ 0 & 0.1 \end{pmatrix}.
\end{aligned}$$

Resumiendo lo anterior se tienen las siguientes densidades posteriores para cada uno de los parámetros

$$[\theta_i | Y_i, \sigma_c^2, \mu_c, \Sigma_c^{-1}] \sim N \left(D_i \left(\sigma_c^{-2} X_i' Y_i + \Sigma_c^{-1} \mu_c \right), D_i \right); \quad (5.23)$$

$$[\mu_c | Y, \sigma_c^2, \theta, \Sigma_c^{-1}] \sim N \left(V \left(\Sigma_c^{-1} k \bar{\theta} + C^{-1} \eta \right), V \right); \quad (5.24)$$

$$[\Sigma_c^{-1} | Y, \sigma_c^2, \theta, \mu_c] \sim W \left(\left(\sum_{i=1}^k (\theta_i - \mu_c) (\theta_i - \mu_c)' + \rho R \right)^{-1}, k + \rho \right); \quad (5.25)$$

$$[\sigma_c^2 | Y, \Sigma_c^{-1}, \theta, \mu_c] \sim IG \left(\frac{n + \gamma_0}{2}, \frac{1}{2} \left(\sum_{i=1}^k (Y_i - X_i \theta_i)' (Y_i - X_i \theta_i) + \gamma_0 \tau_0^2 \right) \right), \quad (5.26)$$

y el orden en que se generaron los parámetros se describe a continuación:

**ESTA TESIS NO DEBE
SALIR DE LA BIBLIOTECA**

1. generar σ_c^2 de (5.26),
2. generar Σ_c^{-1} de (5.25),
3. generar μ_c de (5.24),
4. generar θ_i de (5.23);
5. repetir los cuatro pasos anteriores siguiendo la estructura del algoritmo de Muestreo Gibbs con tamaño de muestra variable (1.4).

Para el caso de la primera iteración los valores iniciales para θ_i son los obtenidos después de aplicar regresión lineal a cada vector de pesos de las ratas, paso 1, y el promedio de éstos para el vector μ_c inicial, paso 2.

De acuerdo al trabajo desarrollado por Gelfand et. al (1990) monitorean el proceso anterior observado gráficas Q-Q de muestras para α_c , β_c , σ_c^2 y los eigenvalores de Σ_c . con un tamaño de muestra 50 después de 35 ciclos (iteraciones). En este problema los valores para α_c , β_c son de menor interés.

Basándose en el resultado anterior, se presentan las gráficas 5.15 y 5.16 de los cuartiles del vector μ_c tomadas en cada ciclo por cada tamaño de muestra, monitoreando su convergencia.

Después se calculan las densidades marginales para α_c, β_c mediante las densidad posterior de μ_c , ecuación (5.24),

$$\begin{aligned} [\mu_c] &= \int [\mu_c | Y, \sigma_c^2, \theta, \Sigma_c^{-1}] [Y, \sigma_c^2, \theta, \Sigma_c^{-1}] \\ &\approx n^{-1} \sum_{i=1}^n [\mu_c | Y, \sigma_{c(i)}^2, \theta_{(i)}, \Sigma_{c(i)}^{-1}] \\ &= \widehat{[\mu_c]} \end{aligned}$$

donde $[\mu_c | Y, \sigma_{c(i)}^2, \theta_{(i)}, \Sigma_{c(i)}^{-1}]$, $i = 1, \dots, n$, que se distribuye como se detalla en (5.24). Obteniendo $\widehat{[\mu_c]}$ pueden graficarse las densidades marginales de α_c y β_c (gráficas 5.17 y 5.18).

A continuación se procede a estimar los pesos para la rata i -ésima. Como los pesos son funciones independientes entre ratas se obtendrían de la siguiente forma

$$[Y_i] \propto \int [Y_i | \theta_i, \sigma_c^2] [\theta_i, \sigma_c^2] d(\theta_i, \sigma_c^2)$$

$$\begin{aligned}
 &= \sum_{j=1}^n [Y_i | \theta_{i(j)}, \sigma_{c(j)}^2] \\
 &= \widehat{[Y_i]}
 \end{aligned}$$

donde $[Y_i | \theta_{i(j)}, \sigma_c^2] = N(X_i, \theta_{i(j)}, \sigma_{c(j)}^2)$, $i = 1, \dots, n$, y se obtiene $\widehat{[Y_i]}$ sobre el promedio de toda la muestra generada por el proceso Gibbs. La figura 5.19 muestra una estimación promedio de los pesos esperados de la raza 25 después del proceso de estimación de los parámetros poblacionales.

El programa que generó lo anterior (gráficas 5.15, 5.16, 5.17, 5.18 y 5.19) se encuentra en la sección del Apéndice bajo el nombre de "ratas". Cabe mencionar que debido a la manejo excesivo del lenguaje la simulación resultó ser muy lenta y aunque el lenguaje "S-Plus" es muy eficiente se hicieron algunas interfaces con FORTRAN para generar datos con las densidades Wishart y Normal Multivariada. Después de todo existen ciertas líneas que pudiesen haber agilizado la simulación, pero la razón para la cual fué hecho el programa "ratas" es el proporcionar un ejemplo de aplicación del Muestreo Gibbs y no una alternativa de calculo para el usuario común.

cuartiles estimados para ALFA

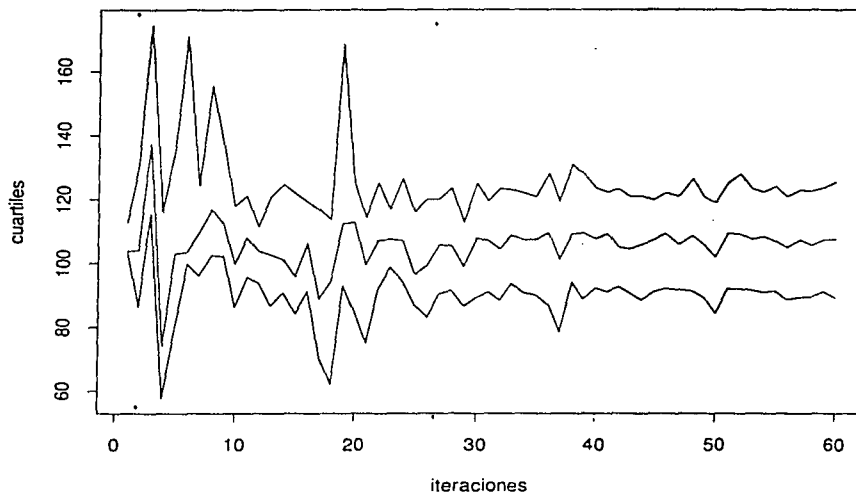


FIGURA 5.15: Cuartiles estimados para Alfa.

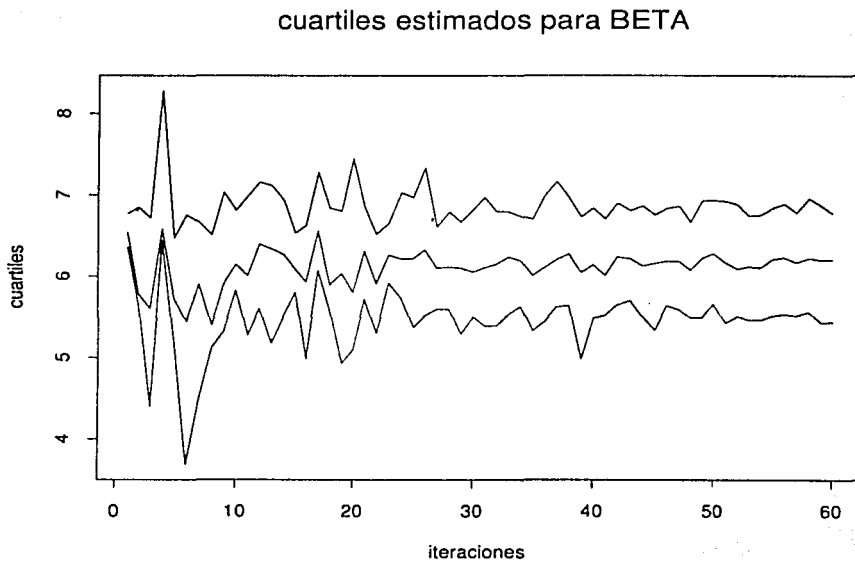


FIGURA 5.16: Cuartiles estimados para Beta

Densidad estimada para ALFA

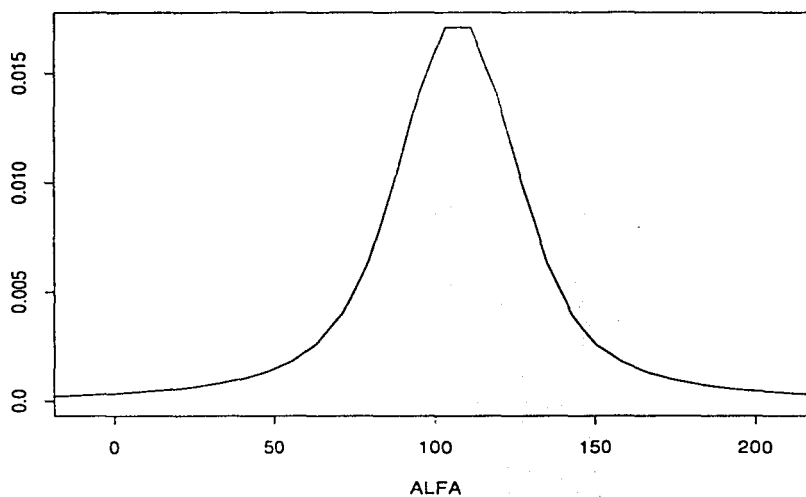


FIGURA 5.17: Densidad marginal posterior estimada para Alfa

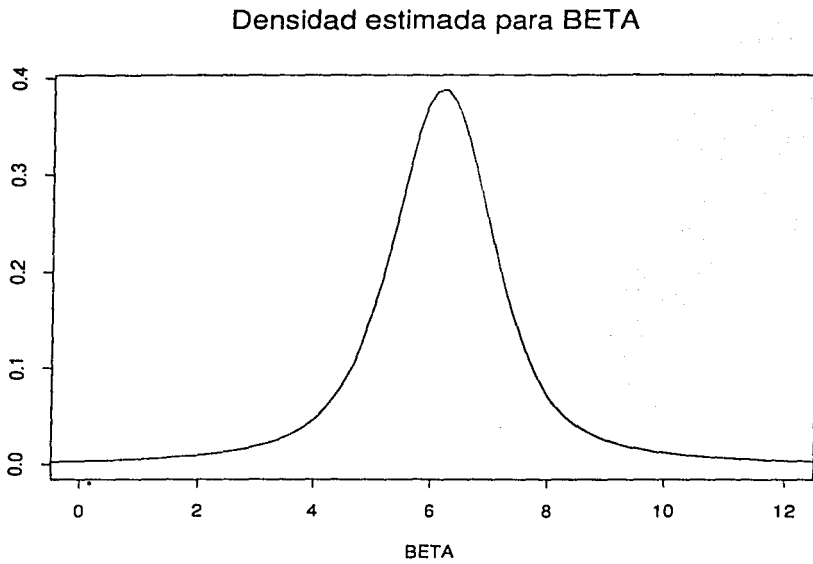


FIGURA 5.18: Densidad marginal posterior estimada para Beta

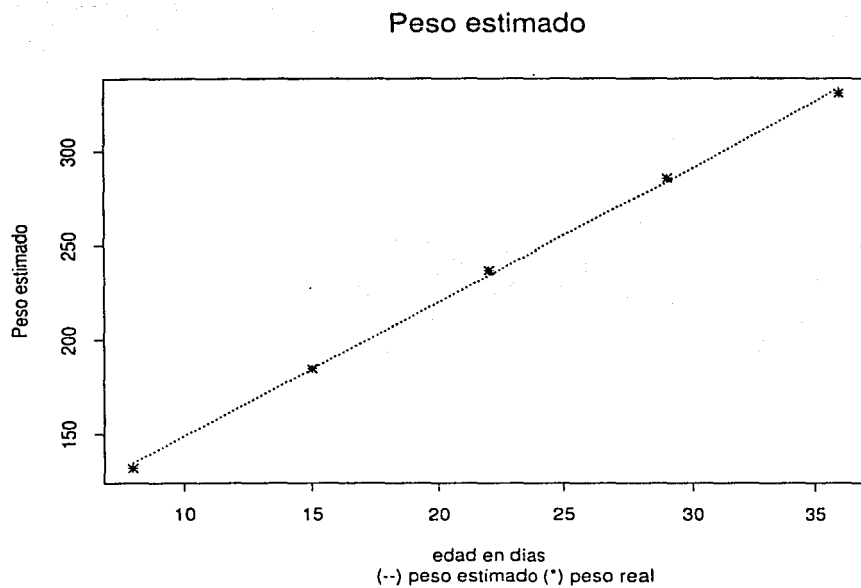


FIGURA 5.19: Peso promedio estimada de la rata 25 después del proceso Gibbs.

CONCLUSIONES

Resultado de encontrarse con integrales de difícil solución, el Análisis Numérico ha desarrollado una teoría extensa basada en métodos numéricos para resolverlas, surgiendo sofisticadas herramientas computacionales. Muchos de estos métodos, sin embargo, resultan ser inadecuados (lentos) al encontrarse con integrales de grandes dimensiones.

A lo largo de la presente se toma como tema principal la aportación de la integración Monte Carlo junto con los algoritmos muestreo por Sustitución Sucesiva y Gibbs, a los métodos numéricos para aproximar integrales. El algoritmo de muestreo Gibbs tiene su mayor aplicación, dentro de la Estadística, en la obtención aproximada de las densidades posteriores necesarias para hacer inferencias Bayesianas. Dicho algoritmo puede ser implementado fácilmente y ha sido extensamente probado en diversas situaciones, presentando gran versatilidad. La teoría que lo sustenta es relativamente simple: cadenas de Markov e integración Monte Carlo.

El algoritmo de Muestreo Gibbs es similar al algoritmo *M.S.S.* descrito en el primer Capítulo, cuyas relaciones se discutieron al final de la sección 1.4, y se basa en la simulación de una cadena de Markov con función de transición K (ecuación (1.17)) existiendo dos modalidades: tamaño de muestra fijo y variable (págs. 17 y 19). En este último caso, dado que la obtención de la solución de la integral de la función de transición es difícil, se hace uso de la integración Monte Carlo para aproximarla por medio de sumas. De esta forma se procede a estimar la integral de K . En relación a la integración por Muestreo de Importancia en el Capítulo 2 se presentan condiciones suficientes para la convergencia de la integral (Geweke, 1989), encontrando cierta ambigüedad en la tercera porque supone que el soporte de la densidad de Muestreo de Importancia se encuentra definida la densidad posterior (afirmación que debe verificarse). En caso contrario se debe especificar para

$$g(\theta)p(\theta) = 0, \quad \theta \notin \Theta;$$

equivalentemente introducir la función indicadora

$$I_{\Theta}(\theta) = \begin{cases} 1 & \text{si } \theta \in \Theta \\ 0 & \text{si } \theta \notin \Theta \end{cases},$$

ya que puede darse el caso en que $p(\theta)$ no esté definida bajo el soporte de la densidad $MI(\theta)$. De ahí la importancia de la selección adecuada de $MI(\theta)$ que aproxime en forma más precisa al valor exacto de la integral.

Geweke (1989) postula condiciones adicionales que debe cumplir $MI(\theta)$ para minimizar la varianza encontrada en la ecuación (2.5), pero en la práctica es difícil encontrar alguna densidad $MI(\theta)$ que las cumpla. (ver Geweke (1989) págs. 1321-1323). Esto se ejemplifica con las gráficas 4.3.4.4 y 4.5 dentro del Capítulo de Simulaciones las cuales tratan de aproximar el valor real de la integral (4.17). Con base en dichas gráficas se intuye que una buena densidad $MI(\theta)$ es la Exponencial. Para ejemplificar las diferencias en la estimación de la varianza, de acuerdo al trabajo de Geweke (1989), se supone a la densidad Normal como la densidad de Muestreo de Importancia (ejemplos 4.2 y 4.3). Como ya se había mencionado, las funciones "g" y "p" pudiesen ser asignadas de diferentes formas encontrando variaciones en los resultados (ver gráficas 4.6, 4.7 y 4.8). Lo anterior sugiere un tema de investigación al suponer el uso de una función $H(\theta)$ en lugar de la descomposición $g(\theta)p(\theta)$ y de este modo calcular la esperanza del cociente $H(\theta)/MI(\theta)$ bajo el dominio de $MI(\theta)$: trabajo en el cual no se hizo una profunda investigación debido al límite de tiempo y tan sólo se señala como una alternativa para nuevos proyectos de investigación.

Después de haber encontrado una buena densidad de Muestreo de Importancia, existe el inconveniente de la asignación de los valores que pudiesen tomar sus parámetros, ya que, de acuerdo a la discusión presentada en el ejercicio 4.3, existen diferencias gráficas en la estimación de la varianza.

Una demostración formal de la convergencia del proceso iterativo se detalla en el Capítulo 3 haciendo uso de conceptos sobre el espacio de Hilbert y sus operadores. Aquí se sugiere comprobar la condición dada por la integral (3.12) para la densidad Normal, ya que esta presenta un comportamiento asintótico bajo el proceso iterativo (ver ejemplo 4.4)

y en caso contrario hacer más flexibles los supuestos generales para la demostración de la convergencia del M.S.S.

Existen varios trabajos sobre la convergencia del algoritmo de Muestreo Gibbs, algunos de los cuales se mencionan al finalizar el Capítulo 3. Cabe hacer notar que existe cierta similitud entre los presentados por Tanner & Wong (1987) y el de Chan (1993). Este último supone algunos conocimientos sobre procesos ergódicos. El trabajo desarrollado por Zellner & Min (1995) resulta atractivo debido a que presentan tres criterios de convergencia que son aplicables de acuerdo al tipo de investigación en que se quiera utilizar el Muestreo Gibbs, y están diseñados para los investigadores cuyo campo de trabajo es más aplicado.

Siguiendo con la convergencia puede verse que esta depende de tres factores: el número de iteraciones, el tamaño de muestra y el valor inicial, (ver ejemplo 4.4). Lo anterior puede resumirse en una sola idea, propuesta por Tanner & Wong (1987), y que consiste en graficar los cuartiles de las muestras generadas vs. el número de iteraciones. A través de éstas gráficas puede verificarse el comportamiento asintótico y como se mencionó, esta forma de monitorear la convergencia revela información sobre la fluctuación debida al valor inicial, el incremento del tamaño de muestra y el incremento del número de iteraciones. Después de haber eliminado el efecto del valor inicial con un determinado tamaño de muestra, producto de varias iteraciones, se cuenta con la facilidad de incrementar en forma abrupta el tamaño de muestra y/o las iteraciones; de esta manera se puede estar "jugando" con la muestra y las iteraciones hasta tener una "convergencia" aceptable para el investigador, (ver gráfica 4.12). Como se discutió anteriormente (ejemplo 4.5, Capítulo 4) se pierde información "buena" al no tomarla en cuenta para generar los nuevos valores. Para este caso se anexa una nueva forma de obtener valores nuevos a partir de pegar los valores viejos hasta tener el nuevo tamaño de muestra, (ver gráfica 4.13), obteniendo así una convergencia más rápida.

Bajo la idea de Tanner & Wong (1987) se puede sugerir no sólo el monitoreo de los cuartiles, sino también de algunos de los cuantiles extremos, por ejemplo los deciles (0.1.0.9), y estudiar su comportamiento. Suponiendo que cuando éstos tengan una convergencia aceptable, los cuartiles también la tendrán, en ese caso los valores obtenidos al final del proceso pueden suponerse cercanos a la verdadera densidad posterior que se desea estimar, (ver gráfica 4.14).

En el último Capítulo se presenta el trabajo realizado sobre la implementación de un

modelo jerárquico. Con esto se pretende ilustrar las ventajas del muestreo Gibbs en un estudio real. Debido al excesivo uso del lenguaje *S-PLUS* en la codificación de dicho programa se hicieron interfaces con FORTRAN para agilizar la simulación, aún así, existen líneas en el código que pudiesen ser reemplazadas para acelerar el proceso: líneas que no se modificaron porque el objetivo principal del programa es el de presentar una aplicación del muestreo de Gibbs. Por tanto las gráficas de los cuartiles para los parámetros presentan oscilaciones, pero en forma intuitiva se encuentran estacionados cerca del valor real. (gráficas 5.15 y 5.16). Las densidades posteriores marginales estimadas presentan una alta concentración en el valor de su media y las estimaciones de los pesos para la rata 25 son aceptables.

Para finalizar se señala la gran utilidad que presenta el Muestreo Gibbs como un método numérico que ofrece grandes ventajas debido a su fácil implementación.

Se hace una recomendación para las personas interesadas en la codificación de un programa que implemente el algoritmo de Muestreo Gibbs, es que trate de codificarlo en algún lenguaje de programación (C o FORTRAN) debido a la velocidad que estos proporcionan. Además que se disponga de suficiente memoria RAM (más de 8 megas preferentemente), que se imprima en pantalla el número de iteración en que se encuentra el proceso para controlar el estado del cálculo y detectar posibles "loops" infinitos.

APENDICE

Los siguientes son los listados de los programas utilizados en el Capítulo 1 A 3, codificados dentro del lenguaje *S-PLUS* para Windows.

Este programa generó los datos de las tablas 1.1 y 1.2, y la gráfica 1.1, que se encuentran en el Capítulo 1.

Programa "introduc".

#Este programa calcula el valor estimado de la esperanza de XY donde (X,Y) tiene densidad normal bivariada, con la integración Monte Carlo además calcula la densidad marginal de X en base a valores obtenidos con la distribución condicionada a Y .

```
suma <- numeric(0)
```

```
aux <- numeric(0)
```

```
espE <- numeric(0) #esperanza estimada
```

```
espR <- numeric(0) #esperanza real
```

```
p1 <- 2
```

```
pren <- numeric(0)
```

la siguiente función genera valores de una distribución normal multivariada con vector de medias "men" y matriz de varianzas "van".

```
gnm <- function( n=1, p=1, men=rep(0,p), van=diag(rep(1,p)) )
```

```
{
```

```

y <- matrix( rnorm(n*p), nrow=n, ncol=p )
ra <- chol( van )
y <- y %*% ra + matrix ( rep(men,n), nrow=n, ncol=p, byrow=T )
y
}
#####MENU PRINCIPAL #####
menu1 <- c(" INTRODUCIR PARAMETROS",
" INTRODUCIR LOS TAMAÑOS DE MUESTRA",
" ESTIMACION Y VALOR REAL DE LA ESPERANZA DE XY",
" GRAFICA DE LA DENSIDAD MARGINAL DE X",
" SALIR",),.
#####
repeat
resp <- menu(menu1)
#####1#####
if (resp == 1)
{
cat ("Vector de Medias ")
men1 <- scan(n=2)
men1 <- matrix(men1,ncol=p1,nrow=1)
cat ("teclea la matriz de var-cov ( (1,1),(1,2),(2,1),(2,2) \n" )
van1 <- scan(n=4)
van1 <- matrix(van1,ncol=p1,byrow=T)
vxdy <- sqrt ( van1[1,1] - van1[2,1]*van1[2,1] / van1[2,2] )

```



```

c1 <- van1[2,1] van1[2,2]
r1 <- men1[1] -men1[2]*c1
vest <- numeric(0)
}
#####1#####
#####2#####
if (resp==2)
{
  cat ("numero de simulaciones deseadas ")
  tam <- as.numeric(readline())
  cat ("tamaños de muestra para las ",tam," simulaciones \n")
  n1 <- scan(n=tam)
}
#####2#####
#####3#####
if (resp==3)
{
  mesp <- men1[1,1]*men1[1,2]
  #Densidad Verdadera para X
  qe <- sqrt(van1[1,1])
  px <- seq (.01,.99,length=200)
  qx <- qnorm (px,men1[1,1],qe)
  dy <- dnorm (qx,men1[1,1],qe)
  for (i in 1:length(n1))

```

```

#generación de los datos con la normal bivariada
aa <- gnm(n1[i],p1,men1,van1)
#estimación de la esperanza del producto de las variables
mpt <- 1
for (j in 1:p1) mpt <- aa[,j]*mpt
espE <- c(espE,sum(mpt)/n1[i])
#densidades estimadas para X dado Y
pren <- rep(0,length=200)
for (k in seq(n1[i]) )
pren <- pren + dnorm(qx,r1+c1*aa[k,2],vxdy)
pren <- pren/n1[i]
vest <- cbind(vest,pren)
espR <- van1[1,2]+mest
cat ("el valor real es ",espR,"\n")
cat ("tamaño estimación \n")
for (k in 1:length(n1)) cat(n1[k]," ".espE[k]," \n" )
}
#####3#####
#####4#####
if (resp==4)
{
#Graficación de la densidad marginal de X dado Y
for (i in 1:length(n1))
plot (qx,dy,lty=1,ylim=range(vest[,i],dy),type="l" , xlab="", ylab="")

```

```
lines(qx.vest[,i].lty=2)
# dens. real (líneas continuas) y densidad estimada (líneas punteadas)
bb <- readline()
}
#####4#####
if (resp==5) break
}
```

Los tres siguientes listados generan muestras con distribuciones exponencial, ji-cuadrada y Normal Estándar; para verificar cuál es una buena Densidad de Muestreo de Importancia, como se señala en el Capítulo 2 .

La manera en que se calculó fue la siguiente:

queremos estimar el valor de la integral

$$\int_0^{1.96} \exp\{-x^2/2\} dx$$

y desarrollando para cada caso se tiene que :

Exponencial

$$\begin{aligned} \int_0^{1.96} \exp\{-x^2/2\} dx &= \frac{\int_{-\infty}^{\infty} \sqrt{2\pi} I_{(0.1, 96)}(x) \exp\{-x^2/2\} dx}{\int_{-\infty}^{\infty} \exp\left\{-\frac{x^2}{2}\right\} dx} \\ &= \frac{\int_0^{\infty} \frac{\sqrt{2\pi} I_{(0.1, 96)}(x) \exp\{-x^2/2\}}{\exp\{-\lambda x\}} \exp(-\lambda x) dx}{\int_0^{\infty} \frac{\exp\{-x^2/2\}}{\exp\{-\lambda x\}} \exp(-\lambda x) dx} \\ &= E\left[\frac{\sqrt{2\pi} I_{(0.1, 96)}(X) \exp\{-x^2/2\}}{\exp(-\lambda X)}\right] \\ &= \frac{E\left[\frac{\exp\{-x^2/2\}}{\exp(-\lambda X)}\right]}{\frac{1}{n} \sum_{i=1}^n \frac{\sqrt{2\pi} I_{(0.1, 96)}(X_i) \exp\{-x^2/2\}}{\exp(-\lambda X_i)}} \\ &\approx \frac{\frac{1}{n} \sum_{i=1}^n \frac{\exp\{-X_i^2/2\}}{\exp(-\lambda X_i)}}{\frac{1}{n} \sum_{i=1}^n \frac{\exp\{-X_i^2/2\}}{\exp(-\lambda X_i)}} \\ &= \bar{g}_n, \end{aligned}$$

donde cada $X_i \sim \exp(\lambda)$, con las funciones g , p y MI^* definidas como

$$\begin{aligned} g(x) &= \sqrt{2\pi} I_{(0.1, 96)}(x), \\ p(x) &= \exp(-x^2/2), \quad y \\ MI^*(x) &= \exp(-\lambda x) \end{aligned}$$

Programa que simuló la gráfica 4.3 dentro del Capítulo 4.

Programa "expon".

```

#Este programa calcula iteraciones para una distribución exponencial
tmues <- c(1000,3000,5000,7000,10000,15000)
iter <- 100
param <- 2
muestra <- numeric(0)
valor <- numeric(0)
cte <- sqrt(2*pi) * (.95/2) #valor real de la integral a estimar
maxmin <- numeric(0)
for (i in 1:length(tmues))
{
  n <- tmues[i]
  m2 <- numeric(0)
  for (j in seq(iter))
  {
    x <- rexp(n,param)
    y <- x[x>0 & x<1.96]
    nume <- sqrt(2*pi) * exp( -.5 (y*y) + (param*y) )
    deno <- exp( -.5 (x*x) + (param*x) )
    resul <- sum(nume) / ( 2*sum(deno) )
  }
  m2 <- c(m2,resul)
}
maxmin <- rbind( maxmin,range(m2) )
valor <- c(valor,m2)
muestra <- c(muestra,rep(n,iter))

```

```
z <- max(abs(valor-cte))
}
plot(muestra,valor, xlim=range(tmues),
      ylim=range(cte+z,z-cte),
      xlab="tamaño de muestra", ylab="estimaciones", main = "Dist. Exponencial")
lines ( c(tmues[1],tmues[length(tmues)]), c(cte,cte))
lines (tmues, smooth (maxmin[,1]) )
lines (tmues, smooth (maxmin[,2]) ).
```

Ji-cuadrada con k grados de libertad .

$$\begin{aligned}
 \int_0^{1.96} \exp\{-x^2/2\} dx &= \frac{\int_{-\infty}^{\infty} \sqrt{2\pi} I_{(0.1, 96)}(x) \exp\{-x^2/2\} dx}{\int_{-\infty}^{\infty} \exp\{-x^2/2\} dx} \\
 &= \frac{\int_0^{\infty} \sqrt{2\pi} I_{(0.1, 96)}(x) \exp\{-x^2/2\} dx}{2 \int_0^{\infty} \frac{\exp\{-x/2\} x^{\frac{k}{2}-1}}{\exp\{-x^2/2\}} \exp\{-x/2\} x^{\frac{k}{2}-1} dx} \\
 &= \frac{E\left[\frac{\sqrt{2\pi} I_{(0.1, 96)}(X) \exp\{-X^2/2\}}{\exp\{-X/2\} X^{\frac{k}{2}-1}}\right]}{2E\left[\frac{\exp\{-X^2/2\}}{\exp\{-X/2\} X^{\frac{k}{2}-1}}\right]} \\
 &\approx \frac{\frac{1}{n} \sum_{i=1}^n \frac{\sqrt{2\pi} I_{(0.1, 96)}(X_i) \exp\{-X_i^2/2\}}{\exp\{-X_i/2\} X_i^{\frac{k}{2}-1}}}{\frac{2}{n} \sum_{i=1}^n \frac{\exp\{-X_i^2/2\}}{\exp\{-X_i/2\} X_i^{\frac{k}{2}-1}}} \\
 &= \bar{y}_n.
 \end{aligned}$$

donde cada $X_i \sim \chi^2(k)$, con funciones g , p y MI^* definidas como:

$$\begin{aligned}
 g(x) &= \sqrt{2\pi} I_{(0.1, 96)}(x), \\
 p(x) &= \exp\left(-\frac{x^2}{2}\right), \quad y \\
 MI^*(x) &= \exp\left(-\frac{x}{2}\right) x^{\frac{k}{2}-1}.
 \end{aligned}$$

Programa que simuló la gráfica 4.4 dentro del Capítulo 4.

Programa "ji-cuad"

#Este programa genera una muestra para con una distribución Ji-cuadrada con 8 grados de libertad

k <- 8 # grados de libertad

repe <- 100 # número de repeticiones para cada número de muestra

tm <- c(1000,3000,5000,7000,10000,15000)

#tm es el vector que contiene los tamaños de muestra a calcular

```

A <- k / 2
valor <- numeric(0)
muestra <- numeric(0)
maxmin <- numeric(0)
cte <- sqrt(2*pi)*( .95/2) #valor real de la integral
for ( i in 1:length(tm) )
{
  n <- tm[i]
m2 <- numeric(0)
  for (j in seq(repe) )
  {
    x <- rchisq(n, k)
    y <- x[x<1.96] #porque solo tiene valores positivos
    if ( length(y) > 0 )
    {
      nume <- sqrt(2*pi) * exp( -.5*y*(y-1) ) / ( y**(A-1) )
deno <- exp( -.5*x*(x-1) ) / ( x**(A-1) )
      resul <- sum(nume) / (2*sum(deno))
m2 <- c(m2,resul)
    }
  }
else
  cat("No hay observaciones disponibles ...\n")
}
muestra <- c(muestra,rep(n,iter))

```



```
valor <- c(valor,m2)
maxmin <- rbind(maxmin,range(m2))
}
z <- max( abs(valor-cte2) )
plot(muestra,valor, xlim=range(tm),
      ylim=range(cte+z, cte-z),
      xlab="tamaño de muestra", ylab="estimaciones",
      main="Dist. Ji-cuadrada")
lines( c(tm[1] , tm[length(tm)] ) , c( cte2,cte2 ) )
lines( tm, smooth(maxmin[,1]) )
lines( tm, smooth(maxmin[,2]) ).
```

Normal estándar

$$\begin{aligned} \int_0^{1.96} \exp\{-x^2/2\} dx &= \frac{\int_{-\infty}^{\infty} \sqrt{2\pi} I_{(0.1, 0.96)}(x) \exp\{-x^2/2\} dx}{\int_{-\infty}^{\infty} \exp\{-x^2/2\} dx} \\ &= \int_{-\infty}^{\infty} \sqrt{2\pi} I_{(0.1, 0.96)}(x) \exp\{-x^2/2\} dx \\ &= E\left[\sqrt{2\pi} I_{(0.1, 0.96)}(X)\right] \\ &\approx \frac{1}{n} \sum_{i=0}^n \sqrt{2\pi} I_{(0.1, 0.96)}(X_i), \end{aligned}$$

donde cada $X_i \sim N(0, 1)$, con funciones g , p y MI^* definidas como

$$\begin{aligned} g(x) &= \sqrt{2\pi} I_{(0.1, 0.96)}(x), \\ p(x) &= \exp\{-x^2/2\}, \quad y \\ MI^*(x) &= \exp\{-x^2/2\}. \end{aligned}$$

Programa que simuló la gráfica 4.5 dentro del Capítulo 4.

Programa "normal".

#Este programa calcula iteraciones para una distribución normal estándar

```
iter <- 100
```

```
tmues <- c (1000,3000,5000,7000,10000,15000)
```

```
muestra <- numeric(0)
```

```
valor <- numeric(0)
```

```
cte <- (sqrt(2*pi)) * (.95/2) #valor real de la integral
```

```
ymax <- numeric(0)
```

```
ymin <- numeric(0)
```

```
for (i in 1:length(tmues))
```

```

{
n <- tmues[i]
tmin <- cte
tmax <- cte
for (j in seq(iter))
{
x <- rnorm(n,0,1)
y <- x[x>0 & x<1.96]
z <- length(y)
resul <- (sqrt(2*pi)) * (1/n) * z
muestra <- c(muestra,n)
valor <- c(valor,resul)
if ( tmax < resul )    tmax <- resul
if ( tmin > resul )    tmin <- resul
}
ymin <- c(ymin,tmin)
ymax <- c(ymax,tmax)
}
z <- max(abs(valor - cte))
plot ( muestra , valor,
      xlab="tamaños de muestra", ylab="aproximaciones" ,
      xlim=range(muestra,0) , ylim=range(valor,z+cte,cte-z),
      main="Normal")
lines ( c(tmues[1] , tmues[length(tmues)]) ,c(cte,cte))

```

```
lines (tmues, smooth (ymin) )
```

```
lines (tmues, smooth (ymax) ).
```

Dentro del Capítulo 2 mediante cálculos algebraicos se encuentran los valores estimados de las varianzas de las aproximaciones. a continuación. A continuación se presentan cuatro ejemplos de su estimación para distintas desiciones de las funciones "g" y "p". Los programas utilizan como densidad de Muestreo de Importancia la densidad Normal ya que esta proporciona una mejor aproximación de la integral (1.3).

Este primer programa utiliza las funciones (4.18).

Programa que simuló la gráfica 4.6 dentro del Capítulo de Simulaciones.

Programa "varianza1"

#Calcula las varianzas para aproximarnos para obtener una media de precisión del error en la integral (1.3) verifica las diferencias de tomar las funciones "g" y "p" en forma diferente

```
tmues <- c (2000,4000,6000,8000,10000)
```

```
iter <- 100
```

```
vartot <- numeric(0)
```

```
media <- 4
```

```
v <- 4
```

```
muestra <- numeric(0)
```

```
valor <- numeric(0)
```

```
for (i in 1:length(tmues))
```

```
{
```

```
  n <- tmues[i]
```

```
  for (j in seq(iter))
```

```
  {
```

```
    x <- rnorm(n,media,sqrt(v))
```

```
    y <- x[x>0 & x<1.96]
```

```

if (length(y) > 0)
{
  nume <- sqrt(2*pi) * exp(.5*(-(y*y) + ((y-media)**2)/v) )
deno <- exp(.5*(-(x*x) + ((x-media)**2)/v) )
  gnbarra <- sum(nume) / sum(deno)
  valor <- c(valor,gnbarra)
  y1 <- x*(x>0 & x<1.96)
  y2 <- sqrt(2*pi*v) * y1 - gnbarra
  w <- exp(0.5*( -(x^2) + ((x-media)^2)/v ) )
  var <- sum( (y2^2)*(w)^2 ) / ( (sum(w))^2 )
  vartot <- c(vartot, var)
}
}
muestra <- c(muestra,rep(n,iter))
}
plot(muestra, vartot, xlab="tamaño de muestra",
ylab="estimación", main="Varianzas" )
lines (muestra,smooth(vartot)).

```

Este programa utiliza las funciones g, p y MP (4.19) para el segundo caso.

Programa que simuló la gráfica 4.7 dentro del Capítulo de Simulaciones.

Programa "varianza2"

#Calcula las varianzas con un dens. normal no estándar tomando las funciones "g" y "p" distintas a el pgm. anterior

```

tmues <- c (2000,4000,6000,8000,10000)
iter <- 100
vartot <- numeric(0)
media <- - 4
v <- 4
muestra <- numeric(0)
for (i in 1:length(tmues))
{
  n <- tmues[i]
  for (j in seq(iter))
  {
    x <- rnorm(n,media,sqrt(v))
    y <- x[x>0 & x<1.96]
    if (length(y) > 0)
    {
      nume <- 2 * exp( .5*( -(y*y) + ((y-media)**2)/v) )
      deno <- exp(-.5*(( x-(x-media)**2)/v))
    }
  }
}

```

```

gnbarra <- sum(ume) / sum(deno)
muestra <- c(muestra,n)
valor <- c(valor,gnbarra)
if (tmax < gnbarra) tmax <- gnbarra
if (tmin > gnbarra) tmin <- gnbarra
y1 <- x*(x>0 & x<1.96)
y2 <- 2*y1*exp(-.5*(y1*(1-y1))) - gnbarra
w <- exp(-0.5 *(-x+((x-media)^2)/v))
var <- sum( (y2^2) * (w)^2 ) / (sum(w))^2
vartot <- c(vartot, var)
}
}
ymax <- c(ymax,tmax)
ymin <- c(ymin,tmin)
}
z <- max (abs(valor-cte))
plot (muestra,vartot, xlab="", ylab="", main="varianzas")
lines (muestra,smooth(vartot) )

```


Este programa utiliza las funciones de los casos 1 y 2 del ejemplo 4.3 y simuló la gráfica 4.8 dentro del Capítulo de Simulaciones.

Programa "conjunta"

#Este pgm. utiliza una normal estándar vs. una normal con media diferente de cero tomando en forma distintas las funciones "g" y "p", presenta las graficas de las estimaciones de las varianzas.

```

tmues <- c (2000,4000,6000,8000,10000)
iter <- 100
media <- 4 #vectores para almacenar la información
vartot1<- numeric(0)
vartot2<- numeric(0)
muestra<- numeric(0)
valor1 <- numeric(0)
valor2 <- numeric(0)
cte <- sqrt(2*pi) * 0.95/2 #valor real de la integral
y1max <- numeric(0)
y1min <- numeric(0)
y2max <- numeric(0)
y2min <- numeric(0)
for (i in 1:length(tmues))
{
  n <- tmues[i]
  for (j in seq(iter))

```

```

{
  t1min <- - cte
  t1max <- - cte
  t2max <- - cte
  t2min <- - cte
  x1 <- rnorm(n,0,1)
  y1 <- x1[x1>0 & x1<1.96]
  x2 <- rnorm(n,media,1)
  y2 <- x2[x2>0 & x2<1.96]
  if (length(y1) & length(y2) >0)
  {
#primer caso (normal estándar)
    z1 <- length (y1)
    g1barra <- sqrt(2*pi) * (1/n) * z1
    valor1 <- c(valor1,g1barra)
    if (t1max < g1barra) t1max <- g1barra
    if (t1min > g1barra) t1min <- g1barra
    l1 <- x1*(x1>0 & x1<1.96)
    e2 <- sqrt(2*pi) * l1 - g1barra
    var1 <- sum( (e2^2)) / (n ^2)
    vartot1 <- c(vartot1, var1)
#segundo caso (normal med=4)
    z2 <- exp (-y2*media + (media* * 2)/2)
    g2barra <- sqrt(2*pi) * (1/n) * sum(z2)
  }
}

```

```

valor2 <- c(valor2,g2barra)
if (t2max > g2barra) t2max <- g2barra
if (t2min > g2barra) t2min <- g2barra
l2 <- x2*(x2>0 & x2<1.96)
n2 <- sqrt(2*pi) * l2 - g2barra
w2 <- exp(-l2*media + (media^2) / 2)
var2 <- sum ( (n2^2) * (w2^2) ) / (sum (w2) )^2
muestra <- c(muestra,n)
}
}
y1max <- c(y1max,t1max)
y1min <- c(y1min,t1min)
y2max <- c(y2max,t2max)
y2min <- c(y2min,t2min)
}
z1 <- max (abs(valor1-cte))
z2 <- max (abs(valor2-cte))
par (mfrow=c(2,2))
plot (muestra , vartot1, xlab="", ylab="", main="varianzas 1er. caso")
plot (muestra ,valor1, xlab="", ylab="", main="aproximaciones 1er. caso",
      xlim=range(muestra), ylim=range(valor1,cte-z1, cte+z1) )
lines ( c (tmues[1], tmues[length(tmues)] ) , c (cte,cte) )
plot (muestra,vartot2, xlab="", ylab="", main="varianzas 2do. caso")
plot (muestra,valor2, xlab="", ylab="", main="aproximaciones 2do. caso",

```

```
xlim=range(muestra), ylim=range(valor2,cte+z2, cte-z2 ) )  
lines ( c (tmues[1], tmues[length(tmues)] ) , c (cte,cte) )
```

Convergencia de Distribuciones Estimadas.

El siguiente listado calcula la distribución estimada de X dado Y . Para diferentes tamaños de muestra, iteraciones y valores iniciales, presentamos tres gráficas dentro del Capítulo 4 en donde se mantiene fijo uno de los valores anteriores y variables los demás.

Programa que simuló las gráficas 4.9, 4.10 y 4.11 dentro del Capítulo de Simulaciones.

Programa "gibbs"

#Este programa calcula la dist. marginal de X a partir de la dist. condicional de X dado Y , graficándolas para cada valor fijo y los demás variables (valor inicial, número de iteraciones y tamaño de muestra).

```
#valores que permanecen constantes
mm <- c(12,12)
vv <- c(1,.5,4)
se1 <- sqrt ( vv[1] - vv[2]*vv[2] / vv[3] )
se2 <- sqrt ( vv[3] - vv[2]*vv[2] / vv[1] )
c1 <- - vv[2]/vv[3]
c2 <- - vv[2]/vv[1]
r1 <- mm[1] -mm[2]*c1
r2 <- mm[2] -mm[1]*c2
# Densidad Verdadera para X
qe <- sqrt(vv[1])
px <- seq (.01,.99,length=200)
qx <- qnorm (px,mm[1],qe)
dy <- dnorm (qx,mm[1],qe)
par (mfrow=c(3,3)) #abro modo gráfico
```

```
totiter <- c(1,5,10) #vector de iteraciones
tammue <- c(5,15,30) #vector de tamaños
valini <- c(150000) #vector de valores iniciales
for (iter in totiter)
{
  for ( tam in tammue )
  {
    for ( yini in valini )
    {
      #####
      y <- yini
      muestra <- numeric(0)
      for (k in 1:tam)
      {
        for (l in 1:iter)
        {
          x <- rnorm (1,r1 + y*c1,se1)
          y <- rnorm (1,r2 + x*c2,se2)
        }
        muestra <- rbind(muestra,c(x,y))
      }
      pren <- rep(0,length=200)
      for (i in seq(tam) )
      {
```

```
    pren <- pren + dnorm(qx,r1+c1*muestra[i,2],se1)
  }
  pren <- pren / tam
  plot (qx,dy,lty=1,ylim=range(pren,dy),type="l",
        xlab="", ylab="") lines(qx,pren,lty=2)
#####
}
}
}.
```

Tamaño de muestra variable generado en forma aleatoria.

El siguiente es el listado para generar iteraciones con tamaño de muestra variables en dos casos (en forma aleatoria y no aleatoria), y para el caso de monitorear cuantiles que no sean los centrales; teniendo la opción de cambiar o añadirse acuerdo al menú de opciones.

En el caso en que la nueva muestra no se haya obtenido de la anterior en forma aleatoria solo cambian algunas líneas, que dentro del programa son señaladas. Si se desea calcular diferentes cuantiles se tecléa la opción 3 modificando los cuantiles por (0.01,0.5,0.99).

Programa que simuló la gráfica 4.12 para el caso de los cuantiles centrales y la gráfica 4.14 dentro del Capítulo 4.

Programa "tamvar"

#Este programa calcula los cuantiles estimados de la dist., grafica los valores para que el usuario puede calcular el tamaño de muestra de acuerdo a la grafica correspondiente. Trabaja bajo el concepto de tamaño de muestra variable, calculando los valores aleatoriamente conforme a los obtenidos anteriormente, teniendo la opción también de graficar cuantiles diferentes a los centrales.

```
me <- numeric(0)
va <- numeric(0)
totiter <- 0
cuant <- numeric(0)
ini <- numeric(0)
muestra <- numeric(0)
cuantil <- c(.25,.5,.75) #cuantiles iniciales
##### MENU PRINCIPAL #####
menu1 <- c("1 INTRODUCIR PARAMETROS",
           "2 INTRODUCIR VALOR INICIAL",
           "3 INTRODUCIR CUANTILES",
```


"4 INTRODUCIR TAMAÑO DE MUESTRA E ITERACIONES",

"5 GRAFICA DE CUARTILES",

"6 MARGINALES",

"7 SALIR")

#####3

repeat

{

resp <- menu(menu1)

1

if (resp == 1)

{

cat ("teclea el vector de medias (m1,m2)\n")

me <- scan(n=2)

dpo <- "n" #bandera para matrices definidas positivas

while (dpo != "s")

{

cat ("teclea la matriz de var-cov ((1,1),(2,1),(2,2))\n")

va <- scan(n=3)

if (va[1]*va[3]-va[2]*va[2] <= 0)

dpo <- "n"

else

dpo <- "s"

}

parámetros constantes para la dist. condicionada

```

se1 <- -sqrt ( va[1] - va[2]*vv[2] / va[3] )
se2 <- -sqrt ( va[3] - va[2]*va[2] / va[1] )
c1 <- -va[2]/va[3]
c2 <- -va[2]/va[1]
r1 <- -me[1] -me[2]*c1
r2 <- -me[2] -me[1]*c2
}

##### 1111111 #####
##### 2222222 #####

if (resp == 2)
{
  cat ("valor inicial para Y :") # pide el valor inicial
  ini <- as.numeric(readline())
}

##### 2222222 #####
##### 3333333 #####

if (resp == 3)
{
  if (totiter ==0)
  {
    cat ("no. de cuantiles \n ")
    cc <- as.numeric(readline())
    cat ("cuantiles \n")
    xy <- scan (n=cc)
  }
}

```

```

cat ("quieres agregar(a), modificar(m) los cuantiles ?")
rp <- readline()
if (rp == "a")
{
  cuantil <- c(cuantil,xy)
  cuantil <- sort(cuantil) #ordena los cuantiles
}
if (rp == "m")
cuantil <- sort(xy)
cat ("los cuantiles son ",cuantil)
}
else
cat ("\nuna vez corrido el programa no puedes alterar los cuantiles\n") }
##### 3333333 #####
##### 4444444 #####
if (resp == 4)
{
if (length(me)==0 || length(ini)==0) #verifica que estén los parámetros adicionales
{
  cat ("no has introducido los parámetros iniciales")
}
else
{
  cat ("tamaño de muestra ")

```

```

s2 <- as.numeric(readline())
cat ("no. iteraciones ")
iter <- as.numeric(readline())
if (totiter == 0)
{
  tam <- s2
  y <- rep(ini,times=tam)
}
## Aquí es donde se agrega un "else" con nuevas instrucciones para la muestra
no aleatoria ##
totiter <- totiter + iter
pp <- ceiling(s2*cuantil) #cuantiles
for (i in 1:iter)
{
  x <- rnorm (s2,r1+y[ceiling(tam*runif(tam,0,1))]) * c1,se1)#genera X de los valores
aleatorios, línea que se modifica
  y <- rnorm (s2,r2+x*c2,se2)
  cuant <- rbind(cuant, c( sort(x)[pp] , sort(y)[pp] )) #pega por renglones los cuantiles
  tam <- s2
}
}
}
#####4444444 #####
##### 555555555555 #####3

```

```

if (resp == 5)
{
if (length(cuant) == 0)
{
cat ("necesitas teclear la opcion 4")
}
else
{
xx <- seq(totiter)
tc <- length(cuantil)
plot(xx,cuant[,1],type="l",ylim=range(cuant[,1:tc]),
      xlab="iteraciones", ylab="cuartiles",
      main="cuartiles para X")
for (j in seq(tc)) lines (xx,cuant[,j])
aa <- readline()
plot(xx,cuant[,tc+1],type="l",ylim=range(cuant[, (tc+1):(2*tc)]),
      xlab="iteraciones", ylab="cuartiles",
      main="cuartiles para Y")
for (j in (tc+1):(2*tc)) lines (xx,cuant[,j])
cat ("\nMUESTRA TOTAL ",tam)
cat ("\nITERACIONES TOTALES ",totiter, "\n")
}
}
##### 55555555555555 #####3

```

```
##### 6666666666666666 #####  
if (resp ==6)  
{  
if (length(x)== 0)  
cat (" no has teclado la opcion 4")  
else  
{  
# Densidades Verdaderas para X y Y  
px <- seq (.01,.99,length=300)  
qx <- qnorm (px,me[1],sqrt(va[1]))  
dx <- dnorm (qx,me[1],sqrt(va[1]))  
qy <- qnorm (px,me[2],sqrt(va[3]))  
dy <- dnorm (qy,me[2],sqrt(va[3]))  
# Densidades estimadas para X y Y  
xdy <- rep(0,length=300)  
ydx <- rep(0,length=300)  
for (i in 1:300)  
{  
xdy[i] <- mean ( dnorm(qx[i],r1+c1*y,se1) )  
ydx[i] <- mean ( dnorm(qy[i],r2+c2*x,se2) ) }  
plot (qx,dx,lty=1,ylim=range(xdy,dx),type="l",  
xlab="", ylab="",  
main="densidad verdadera y estimada para X")  
lines(qx,xdy,lty=2)
```

```

aa <- readline()
plot (qy.dy,lty=1,ylim=range(ydx,dy),type="l",
      xlab="", ylab="",
      main="densidad verdadera y estimada para Y")
lines(qy,ydx,lty=2)
}
}
##### 666666666666 #####
if (resp==7) break
}.
Líneas para la muestra no aleatoria
else
{
mxc <- numeric(0)
aux <- (s2/tam)
for (h in 1:trunc(aux)) mxc <- c(mxc,y)
mxc <- c(mxc, y[1:(aux-trunc(aux))]) # pega los valores hast tener el nvo. tam. de
muestra
y <- mxc
}
y la línea que se modifica en la generación de la muestra aleatoria queda como
x <- rnorm (s2, r1+y*c1, se1)

```

Estimación del tamaño de muestra.

Ahora el listado del Capítulo 2, donde se calcula el tamaño de muestra que se necesitaría para una longitud dada.

Los cálculos ya han sido desarrollados anteriormente y solo anexo la codificación del programa.

Programa que simuló las gráficas de la figura 2.2.

Programa "mn"

#Este programa calcula el tamaño de muestra necesaria para acercarnos al verdadero valor de la integral (1.3), utilizando una dist. normal estándar.

```

cat ("teclea el tamaño de muestra ")
n <- as.numeric(readline())
x <- rnorm(n,0,1)
y <- x[x>0 & x<1.96]
z <- length(y)
gnbarra <- (sqrt(2*pi)) * (z/n)
y1 <- rep(1,length=n)
y1 <- y1*(x>0 & x<1.96)
y2 <- sqrt(2*pi) * y1 - gnbarra
vari <- sum(y2*y2) / n
long <- seq (0.001,1,length=100) * abs(gnbarra)
a <- 2 * 1.96 * sqrt(vari)
b <- long * sqrt(2*pi)
muestra <- (a/b)*(a/b)
plot(long,muestra,
```



```
xlab="longitudes del intervalo para g estimada",  
ylab="tamaños de muestra")  
cat ("valor estimado ", gnbarr,  
      "\n la varianza " ,vari/n, "\n").
```

El siguiente listado está basado en el capítulo que presenta un modelo jerárquico como ejemplo ilustrativo del Muestreo Gibbs para el cálculo de densidades posteriores de los parámetros poblacionales. Dos de las funciones no se encuentran (las que generan valores con f.d.p. Gamma y Wishart) por ser codificadas en FORTRAN. Se realizó una interface para agilizar el proceso Gibbs.

Programa que simuló las gráficas 5.15.5.16.5.17.5.18 y 5.19.

RATAS

#Programa basado en el ejemplo de las Ratas hecho por CYBA-GEIGY verificando el peso de éstas durante 5 semanas datos de las ratas

#matriz de edades en días de las ratas

días < - matrix(c(rep(1,length=5),8,15,22,29,36), ncol=5,byrow=T)

#pesos de las ratas por renglones para cada una de ellas

peso < - matrix(c(151, 199, 246, 283, 320,
145,199,249,293,354, 147,214,263,312,328, 155,200,237,272,297,
135,188,230,280,323, 159,210,252,298,331, 141,189,231,275,305,
159,201,248,297,338, 177,236,285,340,376, 134,182,220,260,296,
160,208,261,313,352, 143,188,220,273,314, 154,200,244,289,325,
171,221,270,326,358, 163,216,242,281,312, 160,207,248,288,324,
142,187,234,280,316, 156,203,243,283,317, 157,212,259,307,336,
152,203,246,286,321, 154,205,253,298,334, 139,190,225,267,302,
146,191,229,272,302, 157,211,250,285,323, 132,185,237,286,331,
160,207,257,303,345, 169,216,261,295,333, 157,205,248,289,316,
137,180,219,258,291, 153,200,244,286,324,),nrow=5,byrow=F)

#parámetros R, rho, C , nu

```

R <- matrix(c(100.0,0.0,1),ncol=2,byrow=T)
rho <- - 2
#matriz de parámetros theta obtenidos de la regresión
theta <- matrix( c(107.1714, 6.0286,
87.08571, 7.31429, 108.2286, 6.5714, 120.3143, 5.0857, 84.11429, 6.68571,
114.2286, 6.1714, 98.08571, 5.91429, 105.9143, 6.4857, 125.0286, 7.1714,
92.05714, 5.74286, 105.1143, 6.9857, 93.4, 6.1, 106.9429, 6.1571, 118.6571,
6.8429, 128.7143, 5.1857, 116.8571, 5.8429, 93.2, 6.3, 114.0571,
5.7429, 111.8286, 6.4714, 109.2857, 6.0143, 106.4286, 6.4714, 97.94286,
5.75714, 104.4857, 5.6143, 117.6, 5.8, 77.37143, 7.12857, 107.9429, 6.6571,
126.8857, 5.8143, 116.6571, 5.7429, 95.68571, 5.51429, 106.8857, 6.1143 ),
ncol=30,byrow=F )
#número de mediciones
ni <- - 5
#número total de ratas
k <- - 30
#número total de observaciones
n <- - ni * k
##Arreglos multidimensionales que guardan los valores generados obtenidos de las itera-
ciones
#theta <- - array (c(0),c(30,2,tam))
thetabarra <- - array ( c(0),c(2,tam))
##FUNCIONES QUE GENERAN LAS DISTRIBUCIONES A UTILIZAR ###
##primero se carga la subrutina que contiene a las f.d.p. siguientes

```

```

load.code <- function(name)
{
if ( !exists("carga",frame=0) )
{
dyn.load(name) #, size=3000000)
assign("carga",1,frame=0) #Session object
cat("Subrutina Cargada en Memoria...\n\n")
load.code("c:\lizbeth\aa\NUE1.obj")
IX <- 100
IY <- 200
IZ <- 300

#####F.D.P. GAMMA INVERSA #####
aa <- .Fortran ("GAM",resp=as.double(numeric(oo)), as.double(oo), as.integer(c(0)) )
DIGamma <- function(alfa=1,beta=1)
{
return (beta/rgamma(1,alfa))
}

#####F.D.P. GAMMA INVERSA #####
#####F.D.P. WISHART #####
Ord <- 2
aa <- .Fortran( "GW", Resp=as.double(numeric(Ord*Ord)),
as.double(numeric(Ord*Ord)), as.integer(Ord),
as.integer(Ord*Ord), as.integer(numeric(1)), as.integer(c(0)) )
#función para generar matrices con dist. WISHART

```

```

GWishart <- function(sigma Ord=1 .Glib=1 )
#Mresp matriz de respuesta
#sigma matriz de entrada
#Ord orden de la matriz
#Tam el cuadrado de Ord
#Glib grados de libertad
{
aa <- .Fortran( "GW", Resp=as.double(numeric(Ord*Ord)),
as.double(numeric(Ord*Ord)), as.integer(Ord),
as.integer(Ord*Ord), as.integer(numeric(1)), as.integer(c(0)) )
return ( matrix (aa$Resp, nrow=Ord,byrow=T ))
}
#####F.D.P. WISHART #####
#####F.D.P. NORMAL MULTIVARIADA #####
Dim <- 2
aa <- .Fortran( "GNM", resp=as.double(numeric(Dim)),
as.double(numeric(Dim)), as.double(Dim*Dim),
as.integer(Dim), as.integer(Dim*Dim), as.integer(c(0)) )
#función para generar matrices con dist. Normal Multivariada
GNM <- function( med , sigma, Dim=1 )
{
#med vector de medias
#sigma matriz de varianzas
#Dim dimensión de la distribución

```

```

aa <- .Fortran( "GNM", resp=as.double(numeric(Dim)),
as.double(med), as.double(sigma),
as.integer(Dim), as.integer(Dim*Dim), as.integer(c(1)) )
return (aa$resp)
}

```

```
#####SIMULACION DE DIST. CONDICIONADAS DE alfa Y beta #####
```

#Este programa calcula los cuantiles estimados de la dist., grafica los valores, el usuario puede monitorear el tamaño de muestra gráficamente y de acuerdo a esto calcular el tamaño de muestra necesaria. Trabaja bajo el concepto de tamaño de muestra variable, calculando los valores aleatoriamente conforme a los obtenidos anteriormente, teniendo la opción también de graficar cuantiles diferentes a los centrales.

```
totiter <- 0
```

```
cuant <- numeric(0)
```

```
muestra <- numeric(0)
```

```
cuantil <- c(.25,.5,.75)
```

```
#####MENU PRINCIPAL #####
```

```
menu1 <- c("INTRODUCIR CUANTILES",
```

```
"INTRODUCIR TAMAÑO DE MUESTRA E ITERACIONES (generación de datos)",
```

```
"GRAFICA DE CUANTILES",
```

```
"GRAFICA DE DENSIDADES MARGINALES PARA ALFA Y BETA",
```

```
"PREDICCION PARA LA RATA i-ESIMA",
```

```
"SALIR")
```

```
#####
```

```
repeat
```

```

resp <- menu(menu1)
##### 1 #####
if (resp == 1)
{
if (totiter == 0)
{
cat ("no. de cuantiles \n")
cc <- as.numeric(readline())
cat ("cuantiles \n")
xy <- scan (n=cc)
cat ("quieres agregar(a), modificar(m) los cuantiles ?")
rp <- readline()
if (rp == "a") { cuantil <- c(cuantil,xy)
cuantil <- sort(cuantil) }
if (rp == "m") cuantil <- sort(xy)
cat ("los cuantiles son ",cuantil)
}
else cat ("\nuna vez corrido el programa no puedes alterar los cuantiles\n")
}
##### 1 #####
##### 2 #####
if (resp == 2)
{
u1 <- numeric(0)

```

```

u2 <- numeric(0)
cat (" tamaño de muestra ")
s2 <- as.numeric(readline())
cat (" no. iteraciones ")
iter <- as.numeric(readline())
aux <- 600
if (totiter == 0) tam <- s2
#inicializo los primeros parámetros
thetabarra <- array ( c( mean(theta[1,]),mean(theta[2,]) ), c(2,aux))
theta <- array ( t(theta),c(30,2,aux) )
mu <- thetabarra
totiter <- totiter + iter
pp <- ceiling(s2*cuantil)
sigma <- numeric(aux)
SIGMA <- array (c(0),c(2,2,aux))
for (i in 1:iter)
{
for (j in 1:s2)
{
j1 <- ceiling( tam*runif(1,0,1) )
suma1 <- 0
#qw <- t(peso) - theta[.,j1]%*%
#suma1 <- sum(diag(qw%*%t(qw)))
suma2 <- matrix(c(0),ncol=2,nrow=2)

```



```

for (l in 1:k)
{
qw <- t(peso[,l]) - theta[l,,j1]^o*o*dias
suma1 <- suma1 + (qw%*%t(qw))
qx <- theta[l,,j1] - mu[,j1]
suma2 <- suma2 + ( qx %*%t(qx))
}
sigma[j] <- DIGamma( n/2, (suma1)/2 )
SIGMA[,j] <- GWishart(suma2 + rho*R , 2 , k+rho)
#variables auxiliares para calcular el vector MU
V <- solve( k* SIGMA[,j] )
V[1,2] <- V[2,1]
V2 <- V%*%(k*SIGMA[,j])
V3 <- V2 %*%thetabarra[,j1]
mu[,j] <- GNM(V3,V,2)
u1 <- rbind(u1,c(V3[1,1],V3[2,1]))
u2 <- rbind(u2,c(V[1,1],V[2,2]))
}
#creo los valores de theta
a2 <- dias%*%t(dias)
for (m in 1:s2)
{
a3 <- 1/sigma[m]
Di <- solve(a3*a2 + SIGMA[,m])

```

```

Di[1,2] <- - Di[2,1]
a4 <- SIGMA[,,m]%%mu[m]
for (n in 1:k)
{
ti <- a3*dias%%peso[n]+a4
theta[n,,m] <- GNM(Di%%ti,Di,2)
}
thetabarra[,m] <- c(mean(theta[,1,m]),mean(theta[,2,m]))
}
cuant <- rbind(cuant, c( sort(mu[1,1:s2])[pp] , sort(mu[2,1:s2])[pp] ))
tam <- s2
}
}
##### 2 #####
##### 3 #####
if (resp == 3)
{
if (length(cuant) == 0)
cat ("necesitas teclear la opcion 2")
else
{
xx <- seq(totiter)
tc <- length(cuantil)
plot(xx,cuant[,1], type="l",ylim=range(cuant[,1:tc]), xlab="iteraciones",

```

```

ylab="cuantiles", main="cuantiles para ALFA")
for (j in seq(tc)) lines(xx,cuant[,j])
dev.print(postscript, "CALFA.eps")
aa <- readline()
plot(xx,cuant[,tc+1], type="l",ylim=range(cuant[,:(tc+1):(2*tc)]), xlab="iteraciones",
ylab="cuantiles", main="cuantiles para BETA")
for (j in (tc-1):(2*tc)) lines(xx,cuant[,j])
dev.print(postscript, "CBETA.eps")
cat ("\nTOTAL ",tam)
cat ("\nTOTALES ",totiter, "\n")
}
}
##### 3 #####
##### 4 #####
if (resp == 4 )

```

#Gráfica de las densidades marginales estimadas de ALFA y BETA obtenidas a partir de la densidad de MU ya que presentó una comparable convergencia, el desarrollo matemático se encuentra descrito en el capítulo del ejemplo de las ratas, se sugiere leerlo antes de comenzar a descifrar el presente pgm.

```

# x representa ALFA y y representa a BETA
{
pres <- 300
px <- seq(0.01,0.99,length=pres)
dx <- rep(0,length=pres)

```

```
dy <- rep(0,length=pres)
min1 <- - 1023
max1 <- - 0
min2 <- - 1023
max2 <- - 0
for (i in 1:s2)
{
a1 <- min(qnorm(px,u1[i,1],sqrt(u2[i,1])))
a2 <- max(qnorm(px,u1[i,1],sqrt(u2[i,1])))
b1 <- min(qnorm(px,u1[i,2],sqrt(u2[i,2])))
b2 <- max(qnorm(px,u1[i,2],sqrt(u2[i,2])))
if ( min1 > a1 ) min1 <- a1
if ( max1 < a2 ) max1 <- a2
if ( min2 > b1 ) min2 <- b1
if ( max2 < b2 ) max2 <- b2
}
qx <- seq(min1,max1,length=pres)
qy <- seq(min2,max2,length=pres)
for (i in 1:s2)
{
dx <- dx + dnorm(qx,u1[i,1],sqrt(u2[i,1]))
dy <- dy + dnorm(qy,u1[i,2],sqrt(u2[i,2]))
}
dx <- dx/s2
```

```

dy <- dy/s2
plot(qx,dx, xlab="ALFA", ylab="", main="Densidad estimada para ALFA", type="l")
dev.print(postscript, "ALFA.eps")
aa <- readline()
plot(qy,dy, xlab="BETA", ylab="", main="Densidad estimada para BETA", type="l")
dev.print(postscript, "BETA.eps")
##### 4 #####
##### 5 #####
if (resp == 5)
#en esta opción se grafica el peso estimado de la rata que el usuario elige
cat ("elige el número de la rata que deseas estime sus pesos (1,...,30)")
cc <- as.numeric(readline())
Pest <- rep(0,length=ni)
for (i in 1:s2)
  Pest <- Pest + rnorm(5,theta[cc,i]%*%dias,sqrt(sigma[i]))
Pest <- Pest/s2
plot(dias[2,],peso[,cc], xlab="edad en dias (-) peso estimado (*) peso real",
  main="Peso estimado", ylab="Peso estimado")
lines(dias[2,],Pest,lty=2)
dev.print(postscript, "Pest.eps")
##### 5 #####
##### 6 #####
if (resp == 6) break
##### 6 #####

```

BIBLIOGRAFIA

1. AKHIEZER N.I., GLAZMAN I.M. (1961). *Theory of Linear Operators in Hilbert Space*. Volume I. Frederick Ungar Publishing C.O. New York.
2. BERBERIAN S.F. (1976). *Introduction to Hilbert Space*. Chelsea Publishing Company. New York.
3. BROCKWELL P.J. & DAVIS R.A. (1991). *Time Series: Theory and Methods*. Second Edition. Springer-Verlag. New York.
4. BILLINGSLEY P. (1979). *Probability and Measure*. John Wiley & Sons. New York.
5. CAAMAÑO M.E. (1992). Gibbs Sampler. *Tesis de Licenciatura (Actuaría)*. Facultad de Ciencias. UNAM. México D.F.
6. CASELLA G. & GEORGE E.I. (1992). Explaining the Gibbs Sampler. *The American Statistician*, 3. vol. 46. 167-174.
7. CHAN K.S. (1993). Asymptotic Behavior of the Gibbs Sampler. *Journal of the American Statistical Association*, vol. 88, no. 421. 320-326.
8. COHEN D.W. (1989). *An Introduction to Hilbert Spaces and Quantum Logic*. Springer-Verlag. New York.
9. DEGROOT M.H. (1986). *Probability & Statistics*. 2nd. Edition. Addison Wesley. México.
10. DUNFORD N. & SCHWARTZ J.T. (1963). *Linear Operators (Part II Spectral Theory)*. Interscience Publishers, John Wiley & Sons. New York.

11. GELFAND A.E. & SMITH A.F.M. (1990). Sampling-Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association*, 410, vol. 85, 398-409.
12. GELFAND A.E., HILLS S.E., RACINE-POON A. & SMITH A.F.M. (1990). Illustration of Bayesian Inference in Normal Data Models Using Gibbs Sampling. *Journal of the American Statistical Association*, 412, vol. 85, 972-985.
13. GEMAN S. & GEMAN D. (1984). Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721-741.
14. GEWEKE J. (1989). Bayesian Inference in Econometrics Models using Monte Carlo Integration. *Econometrica*, 57, 1317-1339.
15. HAMMERSLEY J.M. & HANDSCOMB D.C. (1979). *Monte Carlo Methods*. Chapman & Hall. London.
16. MENDOZA J.R. (1987). La Distribución Normal Multivariada y su relación con otras Distribuciones. *Tesis de Licenciatura (Actuaría)*. Facultad de Ciencias. UNAM. México D.F.
17. RITTER C. & TANNER M.A. (1992). Facilitating the Gibbs Sampler: The Gibbs Stopper and the Griddy-Gibbs Sampler. *Journal of the American Statistical Association*, 419, vol. 87, 861-868.
18. RUDIN W. (1985). *Análisis Real y Complejo*. Reimpresión. Alhambra. Madrid.
RUDIN W. (1988). *Análisis Real y Complejo*. Tercera edición. Mc. Graw Hill. Cd. de México.
19. SAVILLE D.J. & WOOD G.R. (1991). *Statistical Methods : The Geometric Approach*. Springer-Verlag. New York.
20. SHERVISH M.J. & BRADLEY P.C. (1992). On the Convergence of Successive Substitution Sampling. *Journal of Computational and Graphical Statistics*, 2, 111-127.
21. SWOKOWSKI E.W. (1982). *Cálculo con Geometría Analítica*. Segunda edición. versión en Español. Wadsworth, Inc. México.

22. TANNER M.A. (1993). *Tools for Statistical Inference*. Second edition. Springer-Verlag. New York.
23. TANNER M.A. & WONG W.H. (1987). The Calculation of Posterior Data Augmentation. *Journal of the American Statistical Association*, 398, vol. 2. 528-550.
24. WILLIAMSON R.E., CROWLL R.H. & TROTTER R.F. (1972). *Calculus of vector functions*. Third Edition. Prentice Hall Inc. Englewood Clifs.
25. ZELLNER A. & MIN C. (1995). Gibbs Sampler Convergence Criteria. *Journal of the American Statistical Association*, vol. 90, no.431, 921-927.