

N.º 8
2 EJ.



UNIVERSIDAD NACIONAL
AUTONOMA DE MEXICO

FACULTAD DE CIENCIAS

GIBBS SAMPLER

T E S I S

Que para obtener el Titulo de

ACTUARIA

P r e s e n t a

MARIA ESTHER CAAMAÑO SIERRA

México, D.F.

Agosto 1992

FALLA DE ORIGEN



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 está 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.

Contenido

Introducción	1
--------------	---

CAPITULO I. El Algoritmo de Gibbs

I.1. Muestreo de Gibbs	5
I.2. Fundamentos Matemáticos del Algoritmo	9
I.3. Convergencia en el caso bivariado continuo	15
I.4. El caso con más de dos variables	21
I.5. Selección del tamaño de k	23

CAPITULO II. Otros Algoritmos

II.1. El Algoritmo de Rubin "Importance Sampling"	25
II.2. El Algoritmo de Data Augmentation	27

CAPITULO III. Aplicación

III.1. Aplicación del Algoritmo de Gibbs	28
Conclusiones	42
Apéndice: Programas para obtener las muestras	43
Referencias	44

Introducción

La estadística puede ser considerada como un conjunto de métodos que son útiles para analizar adecuadamente conjuntos de datos y extraer conclusiones para, típicamente, alimentar un proceso de toma de decisiones. Actualmente los métodos estadísticos constituyen una herramienta necesaria en los más diversos campos de actividad profesional. Dentro de la estadística, existen diferentes enfoques con los que se puede abordar un problema de inferencia, de entre ellos sobresalen los llamados frecuentista y bayesiano.

Históricamente, en el panorama estadístico ha predominado el muy conocido enfoque frecuentista (Kendall Maurice Sir, *The advanced Theory of Statistics, vol.II*). En este trabajo sin embargo, el contenido estadístico de la discusión se referirá al enfoque bayesiano que si bien es más reciente, presenta ventajas metodológicas.

El enfoque bayesiano plantea cualquier problema de inferencia como un problema de decisión: primeramente se establece el conjunto de decisiones u opciones alternativas $D = \{d_1, d_2, \dots, d_n\}$, que debe ser exhaustivo y excluyente; para cada decisión d_i se establece el conjunto de sucesos inciertos $\{\theta_{i1}, \theta_{i2}, \dots\}$ mutuamente excluyentes que condicionan las consecuencias $\{c_{i1}, c_{i2}, \dots\}$ de cada una de las decisiones. El problema de decisión consiste entonces en elegir una opción del conjunto D sin saber cuál de los sucesos inciertos tendrá lugar. Este enfoque establece una base axiomática para la solución de estos problemas y como consecuencia de los axiomas el decisor debe especificar una medida de probabilidad que describa la información que posee sobre la verosimilitud de los distintos sucesos inciertos, y una función de utilidad, para medir sus preferencias entre las posibles consecuencias de cada opción. El criterio de decisión (el criterio de Bayes) consiste en elegir aquella opción que maximice la utilidad esperada, dada por:

$$u^*(d_i) = \sum_{j=1}^{m_i} u(c_{ij})p(\theta_{ij}) \quad (1)$$

donde:

$u^*(d_i)$ denota la utilidad esperada de d_i .

$u(c_{ij})$ denota la utilidad que tiene para el decisor la consecuencia c_{ij}

$p(\theta_{ij})$ es la probabilidad del suceso incierto θ_{ij} que origina el decisor al momento de tomar la decisión.

m_i indica el número de sucesos inciertos para la decisión d_i .

En lo que se refiere a la medida de probabilidad es razonable suponer que esta será asignada de acuerdo a la información disponible. De esta manera, si la información se basa esencialmente en resultados muestrales previos, las probabilidades asignadas podrían basarse en las frecuencias relativas, si por otra parte, la información procede de la percepción de determinadas simetrías se puede esperar una asignación que utilice la definición clásica de probabilidad y si no se presenta ninguno de los casos anteriores, mediante una apreciación subjetiva del decisor.

Ahora bien, cuando se cuenta con información adicional sobre los sucesos inciertos esa información se incorpora para modificar o actualizar las probabilidades iniciales $p(\theta_{ij}) \quad j = 1, \dots, m_i, \quad i = 1, \dots, n$.

Así, si se llama X a la información adicional, los nuevos valores de las probabilidades asignadas a los sucesos inciertos, después de haber obtenido la información X , serán $p(\theta_{ij}|X) \quad j = 1, \dots, m_i, \quad i = 1, \dots, n$. Utilizando el teorema de Bayes se pueden obtener estas probabilidades, llamadas finales, a partir de las probabilidades iniciales $p(\theta_{ij}) \quad j = 1, \dots, m_i, \quad i = 1, \dots, n$. y de la relación que existe entre X y los sucesos inciertos θ_{ij} ; esto es; $P(X|\theta_{ij})$ que se conoce como la función de verosimilitud de θ_{ij} . En resumen,

$$P(\theta_{ij}|X) = \frac{P(\theta_{ij})P(X|\theta_{ij})}{P(X)} \quad (2)$$

donde $P(X)$ es la función de probabilidad marginal de X . Equivalentemente y puesto que $P(X)$ no depende de θ , se tiene que

$$P(\theta_{ij}|X) \propto P(\theta_{ij})P(X|\theta_{ij}) \quad (3)$$

como función de θ_{ij} , en donde la constante de proporcionalidad se puede calcular utilizando la expresión

$$\int_{-\infty}^{\infty} P(\theta_{ij}|X) d\theta_{ij} = 1 \quad (4)$$

$$\left(\sum_i P(\theta_{ij}|X) = 1 \right) \quad (5)$$

según se trate de un caso continuo o discreto. En particular, entonces, si θ_{ij} es continuo y c denota la constante referida, se tiene que

$$c = \int P(\theta_{ij})P(X|\theta_{ij})d\theta_{ij} \quad (6)$$

Esta última expresión puede ser resuelta en muchos casos con procedimientos analíticos. Sin embargo, no es raro que se requieran métodos numéricos. Teniendo las distribuciones finales puede calcularse para cada decisión la utilidad esperada y finalmente es posible tomar la decisión que la maximice.

Hay que hacer notar que las probabilidades finales no sólo son útiles para obtener utilidades esperadas sino también para obtener características concretas de la distribución correspondiente, como la esperanza, la varianza y en general expresiones del tipo:

$$\int_{\Theta} P(\theta|X)g(\theta)d\theta \quad (7)$$

donde $g(\theta)$ es una función totalmente conocida de θ . Estos resúmenes son de ayuda para describir algunos aspectos particulares del conocimiento que se posee sobre θ .

En ocasiones la distribución de X no sólo depende de un suceso incierto de interés (parámetro de interés), sino también de otros, $\omega_1, \omega_2, \dots, \omega_n$, de manera que la información relevante tiene una distribución de probabilidad $P(X|\theta, \omega_1, \dots, \omega_n)$. En estos

casos, para obtener la correspondiente distribución final marginal de θ es necesario tener una distribución inicial conjunta para $\theta, \omega_1, \dots, \omega_n$, aplicar el teorema de Bayes, encontrar la constante de proporcionalidad y posteriormente marginalizar, esto es:

$$P(\theta, \omega_1, \dots, \omega_n | X) \propto P(X | \theta, \omega_1, \dots, \omega_n) P(\theta, \omega_1, \dots, \omega_n) \quad (8)$$

$$P(\theta, \omega_1, \dots, \omega_n | X) = \frac{1}{c} P(X | \theta, \omega_1, \dots, \omega_n) P(\theta, \omega_1, \dots, \omega_n) \quad (9)$$

donde

$$c = \int P(\theta, \omega_1, \dots, \omega_n) P(X | \theta, \omega_1, \dots, \omega_n) d\theta d\omega_1 \dots d\omega_n \quad (10)$$

y posteriormente obtener:

$$P(\theta | X) = \int_{\omega_n} \int_{\omega_{n-1}} \dots \int_{\omega_1} P(\theta, \omega_1, \dots, \omega_n | X) d\omega_1 d\omega_2 \dots d\omega_n \quad (11)$$

Al igual que (6), la expresión (11) puede ser resuelta mediante procedimientos analíticos, o bien por métodos numéricos. Es de notar que en cualquier caso la dificultad típicamente aumenta con el número de parámetros secundarios $(\omega_1, \dots, \omega_n)$.

En los últimos años ha aparecido en la literatura otro tipo de métodos para resolver este problema, que utiliza algoritmos de muestreo y simulación. Algunos de ellos son el conocido método de Montecarlo, el algoritmo "the Data Augmentation" descrito por Tanner y Wong (1987), el "Gibbs Sampler" introducido por Geman y Geman (1984) y el "Importance Sampling" propuesto por Rubin (1987).

La finalidad de este trabajo es analizar el algoritmo "Gibbs sampler" además de comentar brevemente los otros dos algoritmos.

CAPITULO I

Muestreo de Gibbs

Como se mencionó anteriormente, en ocasiones la función de distribución de la información adicional depende no sólo del suceso o parámetro de interés. En tal caso y como resultado de la aplicación del teorema de Bayes, se obtiene la probabilidad final conjunta $P(\theta, \omega_1, \dots, \omega_n | X)$ a partir de la cual, vía marginalización, se debe obtener $P(\theta | X)$. En lugar de calcular o aproximar numéricamente la integral que da origen a $P(\theta | X)$, el Gibbs sampler genera una muestra aleatoria $\theta^{(1)}, \dots, \theta^{(m)} \sim P(\theta | X)$ sin requerir para ello la función $P(\theta | X)$; con un argumento que se basa en la ley de los grandes números, esta muestra se utiliza para calcular características tales como la media, varianza y en general expresiones del tipo $\int P(\theta | X) g(\theta) d\theta$ donde $g(\theta)$ es una función totalmente conocida de θ . La forma para aproximar las expresiones de este tipo es como sigue:

$$\bar{E}(g(\theta)) = \frac{1}{m} \sum_{i=1}^m g(\theta^{(i)}) \quad (12)$$

En lo que sigue, primero se describe, en términos generales, el algoritmo para el caso de 2 variables, θ y ω . Se aplica a un ejemplo particular y posteriormente se desarrolla el caso para N variables. Por facilidad, $P(\theta | X)$ se sustituye por $f_{\Theta}(\theta)$ en lo que sigue.

Para dos variables (θ, ω) , el Gibbs sampler genera una muestra de $f_{\Theta}(\theta)$ a partir de las distribuciones condicionales $f_{\Theta|W}(\theta|\omega)$ y $f_{W|\Theta}(\omega|\theta)$. Es decir, para aplicar este algoritmo es necesario tener disponibles las distribuciones condicionales en el sentido de que dado un valor de la variable que condiciona debe ser posible generar correcta y eficientemente observaciones de la variable que está condicionada.

El primer paso es dar un valor inicial específico $\omega = \omega_0$; con éste y mediante un proceso iterativo se obtiene una secuencia de Gibbs de valores

$$\omega_0, \theta_0, \dots, \omega_k, \theta_k \quad (13)$$

El proceso iterativo es el siguiente: teniendo $\omega = \omega_0$, este se sustituye en $f_{\Theta|W}(\theta|\omega)$. A partir de $f_{\Theta|W}(\theta|\omega = \omega_0)$ se genera un valor $\theta = \theta_0$ que se utiliza en $f_{W|\Theta}(\omega|\theta)$ para obtener $f_{W|\Theta}(\omega|\theta = \theta_0)$. De nuevo, se genera un valor $\omega = \omega_1$ que se emplea para obtener un valor $\theta = \theta_1$ y así sucesivamente hasta obtener a ω_k y a θ_k .

Al proceso que permite generar (13) se le conoce como "Gibbs sampling". Posteriormente se verá que bajo ciertas condiciones, la distribución de θ_k converge a $f_{\Theta}(\theta)$ conforme $k \rightarrow \infty$. Así, para k suficientemente grande la observación final de (13) se puede considerar una muestra de tamaño uno para $f_{\Theta}(\theta)$. Si se repite m veces este procedimiento se obtiene la muestra $\theta_k^{(1)}, \theta_k^{(2)}, \dots, \theta_k^{(m)}$ de $f_{\Theta}(\theta)$ con la que se pueden calcular aproximaciones del tipo (12).

Por otra parte, con los valores $\omega_k^{(1)}, \theta_k^{(1)}, \dots, \omega_k^{(m)}, \theta_k^{(m)}$ es posible estimar $f_{\Theta}(\theta)$ por medio de la expresión:

$$\widehat{f_{\Theta}(\theta)} = \frac{1}{m} \sum_{i=1}^m f(\theta|\omega = \omega_k^{(i)}) \quad (14)$$

donde $\omega_k^{(1)}, \dots, \omega_k^{(m)}$ es la muestra de $f_W(\omega)$ obtenida mediante el mismo algoritmo.

Ejemplo:

Sean Θ y W variables aleatorias tales que su f.d.p. conjunta es la siguiente:

$$f_{\Theta W}(\theta, \omega) \propto \binom{n}{\theta} \omega^{\theta+a-1} (1-\omega)^{n-\theta+b-1} \quad (15)$$

con $\theta = 0, 1, 2, \dots, n$, $0 \leq \omega \leq 1$

Se desea obtener la f.d.p. de Θ , donde n, a, b son enteros positivos conocidos.

Se puede observar en (15) que para todo valor fijo $\omega = \omega_0$ se tiene que:

$$f_{\Theta|W}(\theta|\omega = \omega_0) \propto \binom{n}{\theta} \omega_0^{\theta+a-1} (1-\omega_0)^{n-\theta+b-1}$$

como función de θ . Más aun,

$$f_{\Theta|W}(\theta|\omega = \omega_0) \propto \binom{n}{\theta} \omega_0^\theta (1 - \omega_0)^{n-\theta} [\omega_0^{a-1} (1 - \omega_0)^{b-1}]$$

de manera que necesariamente,

$$f_{\Theta|W}(\theta|\omega = \omega_0) = \binom{n}{\theta} \omega_0^\theta (1 - \omega_0)^{n-\theta}$$

En otras palabras, dado $\omega = \omega_0$, $\Theta \sim \text{Bin}(\omega_0, n)$. Por otra parte, para todo valor fijo $\theta = \theta_0$ se tiene

$$f_{W|\Theta}(\omega|\theta = \theta_0) \propto \binom{n}{\theta_0} \omega^{\theta_0+a-1} (1 - \omega)^{n-\theta_0+b-1}$$

como función de ω . De hecho se tiene que

$$f_{W|\Theta}(\omega|\theta = \theta_0) \propto \omega^{\theta_0+a-1} (1 - \omega)^{n-\theta_0+b-1}$$

de modo que la distribución condicional de ω , dado $\theta = \theta_0$, resulta una $\text{Beta}(\theta_0 + a, n - \theta_0 + b)$. Es decir,

$$f_{W|\Theta}(\omega|\theta = \theta_0) = \frac{\Gamma(a + b + n)}{\Gamma(\theta_0 + a)\Gamma(n - \theta_0 + b)} \omega^{\theta_0+a-1} (1 - \omega)^{n-\theta_0+b-1}$$

En este ejemplo, la f.d.p de Θ puede ser obtenida analíticamente:

$$f_{\Theta}(\theta) \propto \int \binom{n}{\theta} \omega^{\theta+a-1} (1 - \omega)^{n-\theta+b-1} d\omega$$

$$f_{\Theta}(\theta) \propto \binom{n}{\theta} \frac{\Gamma(\theta + a)\Gamma(n - \theta + b)}{\Gamma(n + a + b)}$$

Pero

$$\begin{aligned} \frac{\binom{n}{\theta} \Gamma(\theta + a)\Gamma(n - \theta + b)}{\Gamma(n + a + b)} &= \frac{n!(\theta + a - 1)!(n - \theta + b - 1)!}{\theta!(n - \theta)!(n + a + b - 1)!} \\ &= \left[\frac{n!(b - 1)!(a - 1)!}{(n + a + b - 1)!} \right] \left[\frac{(\theta + a - 1)!}{\theta!(a - 1)!} \right] \left[\frac{(n - \theta + b - 1)!}{(n - \theta)!(b - 1)!} \right] \\ &= \frac{\Gamma(b)\Gamma(a)n!}{\Gamma(n + a + b)} \binom{\theta + a - 1}{\theta} \binom{n - \theta + b - 1}{n - \theta} \end{aligned}$$

Sea c la constante de proporcionalidad, como $\theta = 0, 1, \dots, n$ tiene que suceder que:

$$c \frac{\Gamma(b)\Gamma(a)n!}{\Gamma(n+a+b)} \sum_{\theta=0}^n \binom{\theta+a-1}{\theta} \binom{n-\theta+b-1}{n-\theta} = 1$$

pero $\sum_{\theta=0}^n \binom{\theta+a-1}{\theta} \binom{n-\theta+b-1}{n-\theta} = \binom{a+b+n-1}{n}$ (Feller, 1972 pág 122), quedando la expresión anterior como

$$c \frac{\Gamma(b)\Gamma(a)n!}{\Gamma(n+a+b)} \frac{\Gamma(a+b+n)}{\Gamma(n+1)\Gamma(a+b)} = 1$$

teniendo que

$$c = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$$

por lo que

$$f_{\Theta}(\theta) = \binom{n}{\theta} \frac{\Gamma(a+b)\Gamma(\theta+a)\Gamma(n-\theta+b)}{\Gamma(a)\Gamma(b)\Gamma(n+a+b)}$$

En cualquier caso, para utilizar el Gibbs Sampler es necesario fijar n, a, b . Como ilustración, sean $n = 16$, $a = 2$ y $b = 4$; sustituyendo estos valores en $f_{\Theta|W}(\theta|\omega)$ y $f_{W|\Theta}(\omega|\theta)$:

$$f_{\Theta|W}(\theta|\omega) = \binom{16}{\theta} \omega^{\theta} (1-\omega)^{16-\theta}$$

$$f_{W|\Theta}(\omega|\theta) = \frac{\Gamma(22)\omega^{\theta+1}(1-\omega)^{19-\theta}}{\Gamma(\theta+2)\Gamma(20-\theta)}$$

Dando un valor inicial arbitrario a ω (por ejemplo 0.5), éste se sustituye en $f_{\Theta|W}(\theta|\omega)$ para que de este modo se genere un valor $\theta = \theta_0$ y así sucesivamente hasta obtener a ω_k y θ_k . Este proceso iterativo se repitió $m = 500$ veces con $k = 10$. Los resultados se pueden ver en la siguiente página.

VALORES DE Θ Y W OBTENIDOS MEDIANTE EL PROCESAMIENTO DE GIBBS
CON $K=10$ Y $M=500$

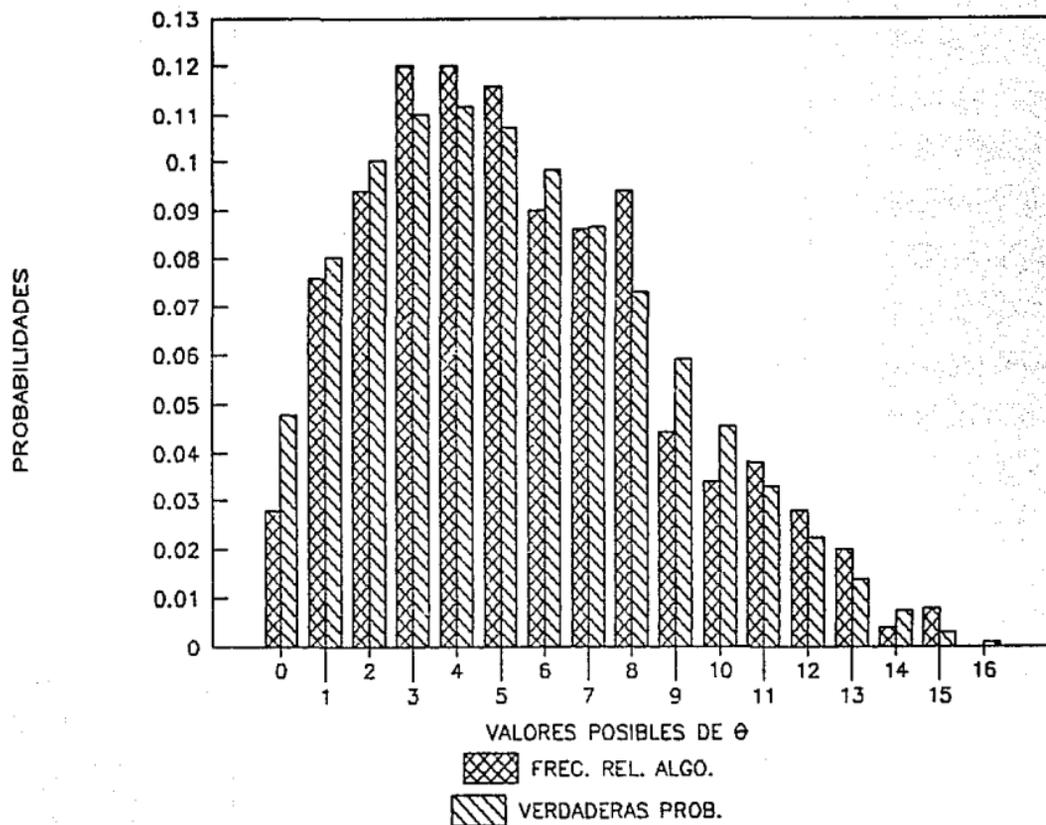
	Θ	W								
1	3	0.115113	5	0.286217	5	0.410210	5	0.552941	3	0.163279
2	4	0.259412	8	0.430294	1	0.254096	4	0.258230	4	0.327598
3	2	0.242184	6	0.315273	8	0.605315	3	0.080354	11	0.574936
4	8	0.511483	10	0.478660	3	0.143516	6	0.219545	13	0.346986
5	2	0.208672	7	0.378425	6	0.291175	2	0.257391	8	0.279196
6	0	0.103247	6	0.346220	4	0.330110	5	0.493491	12	0.733847
7	9	0.534682	8	0.590837	4	0.355206	5	0.511738	9	0.383000
8	2	0.145538	5	0.428457	0	0.037002	9	0.661328	4	0.364015
9	7	0.204888	2	0.259510	4	0.222238	5	0.541212	5	0.314541
10	2	0.077084	9	0.603253	3	0.164046	1	0.121212	12	0.733460
11	1	0.088552	1	0.121437	5	0.372725	7	0.573619	4	0.315117
12	1	0.363297	11	0.482283	4	0.159034	3	0.262558	4	0.423855
13	15	0.818245	3	0.192874	2	0.221371	2	0.166818	0	0.237046
14	1	0.237135	4	0.282121	4	0.180655	4	0.320999	2	0.182895
15	1	0.172725	3	0.274159	7	0.412353	6	0.441452	8	0.432168
16	4	0.328918	0	0.619407	13	0.694879	5	0.179575	10	0.602010
17	8	0.406681	9	0.528758	11	0.544427	0	0.174253	5	0.347867
18	4	0.132048	6	0.339415	9	0.377428	3	0.351192	7	0.570484
19	5	0.210935	4	0.182717	7	0.526323	7	0.231048	3	0.284263
20	7	0.282706	8	0.596973	4	0.244902	4	0.337114	11	0.403481
21	12	0.587001	5	0.355178	11	0.702968	11	0.544392	5	0.375328
22	3	0.194497	6	0.306477	3	0.206499	2	0.038959	8	0.395721
23	7	0.522415	2	0.468320	4	0.282675	8	0.462820	5	0.222758
24	8	0.640611	5	0.493555	5	0.458595	3	0.201328	1	0.290835
25	0	0.328571	7	0.414344	10	0.573405	1	0.110289	7	0.444818
26	8	0.341998	7	0.462274	4	0.359151	10	0.501539	14	0.609092
27	0	0.054330	4	0.117820	2	0.038251	7	0.218882	2	0.062375
28	3	0.272211	4	0.424043	7	0.333579	7	0.330879	4	0.427533
29	8	0.545706	5	0.456287	9	0.814630	9	0.393814	7	0.697307
30	0	0.024475	2	0.219723	0	0.091917	9	0.530302	12	0.626822
31	7	0.402882	3	0.253782	3	0.085696	7	0.422965	7	0.482114
32	8	0.458313	2	0.241879	6	0.322910	13	0.520106	5	0.436997
33	8	0.481373	8	0.382035	3	0.240872	3	0.184681	1	0.184788
34	4	0.331116	3	0.286005	10	0.599747	2	0.343376	5	0.233395
35	8	0.390867	8	0.486791	3	0.111428	10	0.638174	11	0.449590
36	1	0.185140	9	0.449767	5	0.236060	2	0.196303	6	0.259378
37	12	0.583776	5	0.548725	5	0.258232	6	0.321735	4	0.222777
38	10	0.562128	10	0.541952	13	0.718174	2	0.087715	5	0.348830
39	4	0.183585	4	0.148377	5	0.293860	8	0.566323	9	0.484028
40	7	0.403121	6	0.282918	3	0.140818	0	0.062316	6	0.290698
41	3	0.071772	1	0.112068	3	0.120121	5	0.242205	1	0.186264
42	2	0.272259	3	0.208897	5	0.380491	0	0.039854	4	0.239997
43	3	0.270274	8	0.509396	13	0.529291	4	0.216012	11	0.607602
44	5	0.437152	8	0.555055	7	0.368839	7	0.339479	6	0.174053
45	8	0.384832	8	0.430651	3	0.054213	1	0.129597	4	0.229122
46	2	0.149494	6	0.384747	6	0.425328	5	0.250818	6	0.481923
47	5	0.401020	4	0.193226	1	0.100095	6	0.375903	2	0.080846
48	7	0.529321	1	0.065497	6	0.445153	5	0.219276	5	0.241515
49	7	0.276248	12	0.612040	4	0.234385	8	0.330491	8	0.299138
50	3	0.251721	1	0.140840	2	0.109176	3	0.071414	4	0.398860

VALORES DE Θ Y W OBTENIDOS MEDIANTE EL PROCEDIEMTO DE GIBBS
CON $K=10$ Y $M=500$

	Θ	W								
1	5	0.131308	4	0.165747	6	0.383367	8	0.579698	2	0.153593
2	8	0.412919	3	0.204399	8	0.662791	8	0.312992	8	0.528504
3	1	0.146024	2	0.082811	7	0.531447	8	0.388966	11	0.585490
4	5	0.316293	9	0.585530	0	0.138540	5	0.520513	1	0.251464
5	3	0.179749	5	0.301613	7	0.265019	3	0.083078	4	0.096789
6	13	0.689995	2	0.134614	3	0.298306	3	0.214801	1	0.230113
7	11	0.513962	9	0.482912	7	0.266613	8	0.588300	13	0.638158
8	8	0.538285	11	0.469766	4	0.228211	5	0.270289	3	0.254832
9	9	0.448015	7	0.470818	7	0.492821	4	0.179352	5	0.299457
10	8	0.539371	4	0.202629	6	0.347106	2	0.148278	7	0.349364
11	11	0.518095	8	0.388581	3	0.143207	4	0.271200	8	0.475943
12	8	0.361936	4	0.293007	11	0.584912	8	0.347719	4	0.226036
13	3	0.257012	5	0.350795	4	0.149063	9	0.574189	5	0.309515
14	6	0.370708	8	0.423510	12	0.475351	0	0.501098	5	0.615815
15	10	0.564257	5	0.195070	3	0.155148	1	0.262379	1	0.126135
16	0	0.022123	13	0.678154	5	0.385646	1	0.043599	12	0.710535
17	2	0.149422	4	0.294811	5	0.419542	1	0.196168	10	0.637592
18	8	0.395990	1	0.385134	10	0.543153	0	0.103082	6	0.441703
19	3	0.113421	1	0.071417	7	0.384671	8	0.546660	3	0.207196
20	8	0.385812	8	0.579528	15	0.642734	6	0.266358	5	0.291127
21	7	0.415372	4	0.343004	9	0.480074	2	0.142447	6	0.805325
22	6	0.231933	11	0.593085	11	0.679148	1	0.081801	7	0.343238
23	6	0.209399	5	0.346981	1	0.048335	2	0.121160	2	0.227941
24	7	0.337955	0	0.110383	3	0.202979	12	0.486139	4	0.327307
25	15	0.799352	12	0.789983	1	0.258193	11	0.749891	14	0.788874
26	4	0.226142	10	0.584192	1	0.155696	0	0.131220	11	0.673701
27	3	0.325538	9	0.656959	13	0.485946	6	0.429627	10	0.543691
28	3	0.273675	1	0.060153	12	0.483067	5	0.242885	2	0.226085
29	9	0.690620	3	0.296162	3	0.145186	4	0.296492	6	0.410399
30	2	0.158045	2	0.141778	8	0.758861	1	0.183968	4	0.251190
31	7	0.395330	2	0.166615	2	0.254818	3	0.377367	7	0.377731
32	5	0.341751	5	0.344540	2	0.259283	5	0.386181	11	0.475553
33	7	0.400459	5	0.273592	1	0.150634	6	0.415464	3	0.157509
34	3	0.173862	4	0.285629	7	0.394428	4	0.134750	12	0.578805
35	15	0.871126	13	0.460372	4	0.197256	5	0.348102	10	0.481143
36	8	0.289317	5	0.336870	4	0.343114	7	0.231854	1	0.077237
37	3	0.302462	2	0.091030	8	0.542038	8	0.385839	12	0.850047
38	4	0.295275	6	0.370413	9	0.725342	1	0.122545	4	0.160595
39	3	0.429753	5	0.361145	2	0.110788	3	0.299385	8	0.476134
40	1	0.072505	10	0.371709	1	0.200220	2	0.150268	7	0.461778
41	3	0.252927	4	0.413587	2	0.204582	3	0.268075	6	0.532851
42	9	0.493422	8	0.617502	2	0.126884	1	0.125890	5	0.399225
43	2	0.163905	10	0.612303	2	0.298939	4	0.298398	8	0.458766
44	2	0.213962	4	0.434208	11	0.657413	3	0.290601	8	0.378384
45	2	0.318313	3	0.202336	9	0.717218	6	0.353150	6	0.384389
46	8	0.555673	3	0.170563	1	0.105434	10	0.577002	2	0.145072
47	5	0.224624	8	0.629427	5	0.199020	12	0.721547	3	0.354508
48	8	0.377656	4	0.148643	7	0.297179	7	0.535274	8	0.469484
49	3	0.341513	3	0.454962	3	0.266390	4	0.250054	4	0.293182
50	6	0.575891	5	0.288780	7	0.423858	2	0.161692	8	0.277937

GRAFICA COMPARATIVA

DE LA DISTRIBUCION DE θ



Fundamentos Matemáticos del Algoritmo.

Todos los autores de diversos artículos (por ejemplo Gelfand (1990 a), Gelfand (1990 b), Zeger (1990)) que hacen uso del Gibbs Sampler dentro de su investigación dan como referencia principal el artículo de Geman y Geman (1981), en donde se demuestra la convergencia del algoritmo; esta demostración, es obscura y compleja para desarrollar aquí, pues, además de no utilizar un lenguaje estadístico, hace uso de conceptos de reconstrucción de imágenes digitales, redes neuronales y sistemas expertos.

Posteriormente, aparece en la literatura estadística, el artículo de Tanner y Wong (1987), quienes desarrollan una demostración alternativa de un procedimiento que tiene relación con el algoritmo de Gibbs. Las ideas de esa demostración se utilizan en esta sección.

Dada una colección de variables aleatorias W_1, W_2, \dots, W_n el Gibbs Sampler produce aproximaciones para las respectivas distribuciones marginales.

Por hipótesis se cuenta con todas las distribuciones condicionales $f_{W_i|W_{(i)}}(\omega_i|\omega_{(i)})$ en donde $\omega_{(i)} = \{\omega_j; j = 1, 2, \dots, n; j \neq i\}$. Esto es, dado un vector fijo de $W_{(i)}$ ($W_{(i)} = \omega_{(i)}$), se puede generar, correcta y eficientemente, observaciones o muestras de la distribución de W_i condicionada a $\{W_{(i)} = \omega_{(i)}\}$.

Por facilidad, considere el caso de sólo 2 variables, W, Θ . Por definición, se tiene que:

$$f_W(\omega) = \int f_{W\Theta}(\omega, \theta) d\theta \quad (16)$$

Usando el hecho de que $f_{W\Theta}(\omega, \theta) = f_{W|\Theta}(\omega|\theta)f_{\Theta}(\theta)$ podemos escribir (16) de la siguiente manera:

$$f_W(\omega) = \int f_{W|\Theta}(\omega|\theta)f_{\Theta}(\theta)d\theta \quad (17)$$

De la misma manera, se tiene que

$$f_{\Theta}(\theta) = \int f_{\Theta|W}(\theta|\omega) f_W(\omega) d\omega \quad (18)$$

Para facilitar la notación en lo que sigue se puede escribir

$$f_{\Theta}(\theta) = \int f_{\Theta|W}(\theta|t) f_W(t) dt \quad (18')$$

Ahora, sustituyendo (18') en (17), e intercambiando el orden de integración, se observa que $f_W(\omega)$ debe satisfacer la ecuación funcional

$$s = M(s) \quad (19)$$

con

$$M(s) = \int \left[\int f_{W|\Theta}(\omega|\theta) f_{\Theta|W}(\theta|t) d\theta \right] s(t) dt \quad (20)$$

o equivalentemente,

$$s(\omega) = \int H(\omega, t) s(t) dt \quad (21)$$

con

$$H(\omega, t) = \int f_{W|\Theta}(\omega|\theta) f_{\Theta|W}(\theta|t) d\theta \quad (22)$$

y

$$M(s) = \int H(\omega, t) s(t) dt \quad (23)$$

Lo que hace el "Gibbs Sampler", con una aproximación inicial para $f(\omega)$ que se denota $f_0(\omega)$, es calcular sucesivamente

$$f_{i+1} = M(f_i) \text{ con } i = 0, 2, \dots, k-1. \quad (24)$$

Para clarificar cómo el procedimiento de Gibbs coincide con esta solución iterativa de una ecuación de punto fijo es conveniente considerar un caso particular. Casella y George (1992), ilustran el comportamiento del Gibbs Sampler con un ejemplo muy interesante. Considere el caso de dos variables con función de densidad conjunta:

$$f_{W\Theta}(\omega, \theta) = \begin{bmatrix} f_{W\Theta}(0,0) & f_{W\Theta}(1,0) \\ f_{W\Theta}(0,1) & f_{W\Theta}(1,1) \end{bmatrix} = \begin{bmatrix} p_1 & p_2 \\ p_3 & p_4 \end{bmatrix}$$

con $p_i \geq 0$ y $\sum_i p_i = 1$.

Para esta distribución se tiene que $f_{\Theta}(\theta)$ y $f_W(\omega)$ están dadas por:

$$f_W(\omega) = [f_W(0) \quad f_W(1)] = [p_1 + p_3 \quad p_2 + p_4]$$

$$f_{\Theta}(\theta) = [f_{\Theta}(0) \quad f_{\Theta}(1)] = [p_1 + p_2 \quad p_3 + p_4]$$

de manera que, estrictamente, no es necesario aplicar Gibbs sampling. De cualquier forma, se tiene que las distribuciones condicionales para $W|(\Theta = \theta)$ y para $\Theta|(W = \omega)$ pueden ser calculadas sin problema alguno, y expresadas en forma matricial, como sigue:

$$\begin{aligned} A_{\theta|w} &= \begin{pmatrix} f_{\Theta|W}(0|0) & f_{\Theta|W}(1|0) \\ f_{\Theta|W}(0|1) & f_{\Theta|W}(1|1) \end{pmatrix} = \begin{pmatrix} \frac{p_1}{p_1+p_3} & \frac{p_2}{p_1+p_3} \\ \frac{p_3}{p_2+p_4} & \frac{p_4}{p_2+p_4} \end{pmatrix} \\ A_{w|\theta} &= \begin{pmatrix} f_{W|\Theta}(0|0) & f_{W|\Theta}(1|0) \\ f_{W|\Theta}(0|1) & f_{W|\Theta}(1|1) \end{pmatrix} = \begin{pmatrix} \frac{p_1}{p_1+p_2} & \frac{p_2}{p_1+p_2} \\ \frac{p_3}{p_3+p_4} & \frac{p_4}{p_3+p_4} \end{pmatrix} \end{aligned} \quad (25)$$

El proceso iterativo de Gibbs aplicado en este caso da como resultado una secuencia de 0's y 1's y las matrices $A_{\theta|w}$ y $A_{w|\theta}$ se pueden interpretar como las matrices de transición, que se forman con las probabilidades de paso de un estado ω a un estado θ y viceversa.

Si se está interesado en generar, la distribución marginal de W sólo se requiere el último de los valores de cada iteración $\omega_0, \omega_1, \dots, \omega_k$, de la secuencia de Gibbs. Para

ilustrar el proceso de convergencia de las distribución marginal de ω_k , es interesante observar que ω_1 se obtiene a partir de ω_0 , con un paso intermedio que involucra a θ_1 , formando así una cadena de Markov con probabilidad de transición:

$$P(W_1 = \omega_1 | W_0 = \omega_0) = \sum_{\theta} P(W_1 = \omega_1 | \Theta_1 = \theta) P(\Theta_1 = \theta | W_0 = \omega_0) \quad (26)$$

De esta forma, la evolución del sistema del estado ω_0 al ω_1 queda descrito por la matriz de transición:

$$A_{\omega|\omega} = A_{\theta|\omega} A_{\omega|\theta} \quad (27)$$

Así, de manera similar, para ir de ω_0 a ω_k se forma una cadena de Markov con probabilidad de transición $P(W_k = \omega_k | W_0 = \omega_0)$, la cual está descrita por la matriz de transición $(A_{\omega|\omega})^k$.

Si f_k denota la distribución marginal de W_k , se sigue de las propiedades de las cadenas de Markov (Hoel, Port and Stone; 1972), que:

$$f_k = f_0 (A_{\omega|\omega})^k = (f_0 (A_{\omega|\omega})^{k-1}) A_{\omega|\omega} \quad (28)$$

o equivalentemente

$$f_k = f_{k-1} A_{\omega|\omega} \quad (28')$$

Como todas las entradas de la matriz $A_{\omega|\omega}$ son positivas, la ecuación (28) implican que para cualquier probabilidad inicial f_0 conforme $k \rightarrow \infty$, f_k converge a la única función de distribución f que es punto estacionario de (28') (Hoel, Port and Stone; (1972)). Es decir, a la función f que satisface la ecuación

$$f A_{\omega|\omega} = f \quad (29)$$

En el ejemplo se tiene que

$$A_{\omega|\omega} = \begin{bmatrix} \frac{p_1}{p_1+p_3} & \frac{p_3}{p_1+p_3} \\ \frac{p_2}{p_2+p_4} & \frac{p_4}{p_2+p_4} \end{bmatrix} \begin{bmatrix} \frac{p_1}{p_1+p_2} & \frac{p_2}{p_1+p_2} \\ \frac{p_3}{p_3+p_4} & \frac{p_4}{p_3+p_4} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{p_1^2}{(p_1+p_3)(p_1+p_2)} + \frac{p_3^2}{(p_1+p_3)(p_3+p_4)} & \frac{p_1 p_2}{(p_1+p_3)(p_1+p_2)} + \frac{p_3 p_4}{(p_1+p_3)(p_3+p_4)} \\ \frac{p_2 p_1}{(p_1+p_2)(p_2+p_4)} + \frac{p_4 p_3}{(p_2+p_4)(p_3+p_4)} & \frac{p_2^2}{(p_1+p_2)(p_2+p_4)} + \frac{p_4^2}{(p_2+p_4)(p_3+p_4)} \end{bmatrix}$$

Por otra parte

$$[p_1 + p_3 \quad p_2 + p_4] \begin{bmatrix} \frac{p_1^2}{(p_1+p_3)(p_1+p_2)} + \frac{p_3^2}{(p_1+p_3)(p_3+p_4)} & \frac{p_1 p_2}{(p_1+p_3)(p_1+p_2)} + \frac{p_3 p_4}{(p_1+p_3)(p_3+p_4)} \\ \frac{p_2 p_1}{(p_1+p_2)(p_2+p_4)} + \frac{p_4 p_3}{(p_2+p_4)(p_3+p_4)} & \frac{p_2^2}{(p_1+p_2)(p_2+p_4)} + \frac{p_4^2}{(p_2+p_4)(p_3+p_4)} \end{bmatrix}$$

$$= \left[\frac{p_1^2 + p_2 p_1}{p_1 + p_2} + \frac{p_3^2 + p_3 p_4}{p_3 + p_4} \quad \frac{p_1^2 + p_2 p_1}{p_1 + p_2} + \frac{p_3^2 + p_3 p_4}{p_3 + p_4} \right]$$

$$= \left[\frac{p_1(p_1+p_2)}{p_1+p_2} + \frac{p_3(p_3+p_4)}{p_3+p_4} \quad \frac{p_2(p_1+p_2)}{p_1+p_2} + \frac{p_4(p_3+p_4)}{p_3+p_4} \right]$$

$$= [p_1 + p_3, p_2 + p_4]$$

Por lo tanto $f = f_W(\omega)$. Así, Casella y George (1992) comprueban que la secuencia de Gibbs para este problema converge a la marginal que interesa.

Si el proceso iterativo se detiene en una k suficientemente grande puede considerarse que la distribución marginal de W_k es aproximadamente igual a $f_W(\omega)$.

El desarrollo algebraico realizado para matrices de transición 2×2 es el mismo cuando las matrices de transición son $n \times n$; teniendo así que la distribución estacionaria será la distribución marginal de ω .

Si una de las dos o ambas variables, θ y ω ; son continuas, el argumento es parecido. La ecuación análoga a (26) resulta

$$f_{W_1|W_0}(\omega_1|\omega_0) = \int f_{W_1|\Theta_1}(\omega_1|\theta) f_{\Theta_1|W_0}(\theta|\omega_0) d\theta \quad (30)$$

en donde el papel de la matriz de transición en k pasos ($A_{\omega|\omega}$)^k del caso discreto, aquí lo juega una especie de "matriz infinita de transición" cuyas entradas satisfacen la siguiente relación:

$$f_{W_k|W_0}(\omega|\omega_0) = \int f_{W_k|W_{k-1}}(\omega|t) f_{W_{k-1}|W_0}(t|\omega_0) dt \quad (31)$$

de manera, que la siguiente ecuación corresponde a las expresiones (28) y (28')

$$\begin{aligned} f_{W_k}(\omega) &= \int f_{W_0}(\omega_0) f_{W_k|W_0}(\omega|\omega_0) d\omega_0 \\ &= \int f_{W_0}(\omega_0) \left[\int f_{W_k|W_{k-1}}(\omega|t) f_{W_{k-1}|W_0}(t|\omega_0) dt \right] d\omega_0 \\ &= \int \left[\int f_{W_{k-1}|W_0}(t|\omega_0) f_{W_0}(\omega_0) d\omega_0 \right] f_{W_k|W_{k-1}}(\omega|t) dt \\ &= \int f_{W_{k-1}}(t) f_{W_k|W_{k-1}}(\omega|t) dt \end{aligned} \quad (32)$$

en donde el operador integral juega el papel que; en el caso discreto, correspondía a la suma que se obtiene al efectuar el producto matricial.

Examinando con más detalle la ecuación (32)

$$f_{W_k}(\omega) = \int f_{W_{k-1}}(t) f_{W_k|W_{k-1}}(\omega|t) dt \quad (33)$$

se puede observar que $f_{W_{k-1}}(t)$ juega el papel de $H(\omega, t)$ en la expresión (21) y que por tanto, (33) es equivalente a (24).

En otras palabras, el procedimiento de Gibbs para el caso continuo equivale a la solución iterada de la ecuación de punto fijo $s = M(s)$. La validez del Gibbs Sampler en este caso ha sido estudiado en detalle por Tanner y Wong (1987).

Convergencia en el Caso Bivariado Continuo

En esta sección se presenta y analiza una serie de resultados que utilizan Tanner y Wong (1987) para establecer la convergencia de $f_{W_k}(\omega)$ a $f_{W}(\omega)$.

Suponga que W es un subconjunto compacto y convexo de R ; L_1 es el espacio de funciones integrables (Lebesgue) de $\omega \in W$, con $\|g\| = \int |g(\omega)|d\omega$ para $g \in L_1$.

Sean $f_i(\omega)$, $H(\omega, t)$ y M definidas como se hizo anteriormente, donde M es un operador lineal. Sea f_* como la densidad verdadera de W .

Recordando el desarrollo para obtener la ecuación (22); f_* es un punto fijo de M , es decir, $M(f_*) = f_*$.

Teorema 1. Las distancias en L_1 a la distribución verdadera son no crecientes en las iteraciones, es decir

$$\|f_{i+1} - f_*\| \leq \|f_i - f_*\|$$

Demostración. La demostración de Tanner y Wong (1987) hace uso de los siguientes resultados

a)

$$\begin{aligned} \int H(\omega, t)d\omega &= \int \left[\int f_{W|\Theta}(\omega|\theta) f_{\Theta|W}(\theta|t) d\theta \right] d\omega \\ &= \int \left[\int f_{W|\Theta}(\omega|\theta) f_{\Theta|W}(\theta|t) d\omega \right] d\theta \\ &= \int f_{\Theta|W}(\theta|t) \left[\int f_{W|\Theta}(\omega|\theta) d\omega \right] d\theta \\ &= \int f_{\Theta|W}(\theta|t) * 1 d\theta \\ &= 1 \end{aligned}$$

así si $g(\omega) \geq 0 \forall \omega$ entonces

$$\begin{aligned}
\|M(g)\| &= \int |M(g(\omega))| d\omega \\
&= \int \left| \int H(\omega, t) g(t) dt \right| d\omega \\
&= \int \int H(\omega, t) g(t) dt d\omega \\
&= \int \int H(\omega, t) g(t) d\omega dt \\
&= \int |g(t)| \int H(\omega, t) d\omega dt \\
&= \int g(t) * 1 dt \\
&= \int |g(t)| dt \\
&= \|g\|
\end{aligned}$$

b) Si $g(\omega) \geq f(\omega)$ para todo ω entonces $Mg(\omega) \geq Mf(\omega)$ para todo ω (esto se debe a que M es un operador integral).

Sea

$$g = f_i - f_*$$

entonces

$$Mg = f_{i+1} - f_*$$

$$\|Mg\| = \int |Mg(\omega)| d\omega \leq \int |(M|g|)(\omega)| d\omega = \|M|g|\| = \| |g| \| = \|g\|$$

es decir

$$\|f_{i+1} - f_*\| \leq \|f_i - f_*\|$$

Resta ahora asegurar que las distancias a la distribución verdadera son estrictamente decrecientes en las iteraciones y comprobar que f_* la única densidad que satisface la ecuación de punto fijo.

Para poder tener respuestas afirmativas a estas cuestiones, Tanner y Wong (1987) requieren las siguientes condiciones:

a) $H(\omega, t)$ es uniformemente acotada y equicontinua en ω . Para cualquier $\omega_0 \in W$ existe una vecindad abierta U de ω de tal manera que $H(\omega, t) > 0 \forall \omega, t \in U$; esto es, si ω, t están cercanas, entonces es posible generar algún valor θ de $\int_{\Theta|W}(\theta|t)$ tal que $f_{W|\Theta}(\omega|\theta)$ es diferente de cero.

Incorporando estas hipótesis, los autores proceden a la prueba del siguiente

Lema 1. Bajo las condiciones mencionadas, cualquier densidad f que es punto fijo de M debe ser continua y estrictamente positiva.

Demostración.

Continuidad: Por hipótesis $f(\omega) \geq 0$, $f(\omega) = \int H(\omega, t)f(t)dt$.

Como $H(\omega, t)$ es equicontinua se tiene que si $\omega_1, \omega \in W$ $|H(\omega_1, t) - H(\omega, t)| \rightarrow 0$ conforme $\omega_1 \rightarrow \omega$, por lo tanto

$$\int |H(\omega_1, t) - H(\omega, t)|f(t)dt \rightarrow 0$$

y

$$\begin{aligned} |f(\omega_1) - f(\omega)| &= \left| \int H(\omega_1, t)f(t)dt - \int H(\omega, t)f(t)dt \right| \\ &= \left| \int [H(\omega_1, t) - H(\omega, t)]f(t)dt \right| \\ &\leq \int |[H(\omega_1, t) - H(\omega, t)]f(t)|dt \\ &= \int |H(\omega_1, t) - H(\omega, t)|f(t)dt \end{aligned}$$

de donde se deduce, utilizando convergencia dominada, que f es continua.

Estrictamente Positiva: Considerese el conjunto $A = \{\omega \in W; f(\omega) > 0\}$. Suponga que $A \neq W$.

Entonces debe haber al menos $\omega_0 \in W$ que es punto frontera A , ya que W es conexo.

Por la condición a), existe una vecindad U de ω_0 tal que $H(\omega, t) > 0 \forall \omega, t \in U$.

Como ω_0 es punto frontera de A , existe $\omega_1, t \in U(\omega_0)$, tal que $f(\omega_1) = 0$, y $f(t) > 0$

para algún subconjunto abierto de U ; teniendo así:

$$0 = f(\omega_1) = \int_W H(\omega_1, t)f(t)dt \geq \int_U H(\omega_1, t)f(t)dt \geq 0$$

Por lo tanto debe suceder que $A = W$.

El siguiente paso de Tanner y Wong consiste en probar un resultado intermedio más:

Lema 2. Bajo las condiciones, si $g \in L_1$ es una función tal que ni su parte positiva, g^+ ; y ni su parte negativa, g^- ; es igual a cero entonces

$$\|M(g)\| < \|g\|$$

La prueba de Tanner y Wong(1987) se basa en demostrar que el soporte de Mg^+ intersección el soporte de Mg^- es no vacía. Sin embargo, es posible considerar una función g que satisface la hipótesis del lema y tal que $sopMg^+ \cap sopMg^- = \emptyset$.

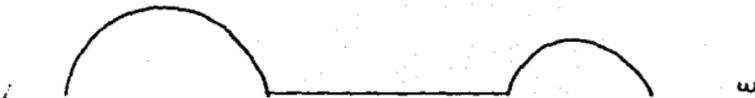
g

Función g



M(g)

Función Mg-



Por otra parte, la evidencia empírica provista por el empleo masivo del Gibbs Sampler sugiere que el resultado es correcto aun cuando no sea la de Tanner y Wong (1987) la prueba apropiada. La autora de este trabajo intentó formular una demostración alternativa pero la búsqueda resultó infructuosa al menos en el sentido de que la hipótesis adicionales se consideran excesivamente restrictivas.

En cualquier caso, Tanner y Wong (1987) presentan el siguiente

Corolario. Bajo las condiciones, la distancia de f_i a f_* en cada iteración es estrictamente decreciente.

y finalmente, el Teorema principal cuya demostración queda en duda puesto que emplea el resultado del lema 2.

Teorema 2. Bajo las condiciones, la densidad f_* es la única que satisface la ecuación de punto fijo.

Demostración. Ya se vió que f satisface la ecuación de punto fijo, sólo se requiere demostrar que es la única que los satisface.

Suponga que existe otra función f_{**} diferente a f_* tal que $Mf_{**} = f_{**}$. Defina

$$g = f_* - f_{**}$$

Por el lema 1, g debe ser continua; además $\int g(\omega)d\omega = 0$ ya que g está definida como una resta de funciones de densidad.

Por hipótesis, $g \neq 0$; pero como g puede escribirse en términos de g^+, g^- se tiene que:

$$0 \neq g = g^+ - g^- \Leftrightarrow g^+ \neq g^- \Leftrightarrow g^+ \neq 0 \vee g^- \neq 0$$

es decir, ni g^+ ni g^- pueden ser idénticamente cero. Haciendo uso del lema 2,

$$\|Mg\| < \|g\|$$

pero

$$Mg = Mf_* - Mf_{**} = f_* - f_{**} = g$$

y, consecuentemente,

$$\|Mg\| = \|g\|$$

lo cual es una contradicción. Por lo tanto f_* es única.

El caso con más de dos variables

Conforme el número de variables en un problema se incrementa, aumenta también el número de distribuciones marginales y condicionales que se pueden obtener de la distribución conjunta, de tal manera que es posible usar diferentes conjuntos de distribuciones condicionales para calcular la distribución marginal de interés, y se tienen diferentes caminos para establecer la ecuación de punto fijo correspondiente a (23).

Así, si se tienen tres variables Θ, W y Y variables, el Gibbs sampler requiere de las condicionales $f_{\Theta|W,Y}(\theta|w,y), f_{W|\Theta,Y}(w|\theta,y), f_{Y|W,\Theta}(y|w,\theta)$ y no de todas las condicionales que pueden ser generadas con tres variables.

Lo que hace el Gibbs Sampler es generar una muestra en forma iterativa a partir de $f_{\Theta|W,Y}(\theta|w,y), f_{W|\Theta,Y}(w|\theta,y), f_{Y|W,\Theta}(y|w,\theta)$ como sigue: dados los valores iniciales $\omega = \omega_0, \theta = \theta_0$ estos se sustituyen en $f_{Y|W,\Theta}(y|w,\theta)$. A partir de $f_{Y|W,\Theta}(y|w = \omega_0, \theta = \theta_0)$ se genera un valor $y = y_0$ que se utiliza en $f_{W|\Theta,Y}(w|\theta, y)$ para obtener $f_{W|\Theta,Y}(w|\theta = \theta_0, y = y_0)$ y de esta generar $\omega = \omega_1$, el cual es utilizado junto con $y = y_0$ para obtener $f_{\Theta|W,Y}(\theta|w = \omega_1, y = y_0)$. Procediendo, sucesivamente, se tiene en la j -ésima iteración que y_j es generada de $f_{Y|\Theta,W}(y|\theta = \theta_j, w = \omega_j), W = \omega_{j+1}$, se obtiene a partir de $f_{W|Y,\Theta}(w|y = y_j, \theta = \theta_j)$ y $\Theta = \theta_{j+1}$ es generada de $f_{\Theta|W,Y}(\theta|y = y_j, w = \omega_{j+1})$. De esta manera se la obtiene la sucesión

$$\omega_0, \theta_0, y_0, \dots, \omega_k, \theta_k, y_k \quad (34)$$

con la propiedad de que cuando $k \rightarrow \infty, \omega_k$ resulta una muestra de tamaño 1 de $f_W(\omega)$, θ_k lo es de $f_{\Theta}(\theta)$ y y_k de $f_Y(y)$.

El desarrollo teórico que da sustento al Gibbs Sampler cuando se tienen más de dos variables es más complicado pero esencialmente se reduce a plantear nuevamente algunas ecuaciones de punto fijo.

Así, por ejemplo.

$$f_W(\omega) = \int \int f_{W,\Theta,Y}(\omega, \theta, y) dy d\theta \quad (35)$$

Como $f_{W,\Theta,Y}(\omega, \theta, y) = f_{W|\Theta,Y}(w|\theta, y) f_{\Theta,Y}(\theta, y)$ resulta que

$$f_W(\omega) = \int \int f_{W|\Theta, Y}(\omega|\theta, y) f_{\Theta, Y}(\theta, y) dy d\theta \quad (36)$$

Pero $f_{\Theta, Y}(\theta, y) = \int f_{\Theta, Y|W}(\theta, y|t) f_W(t) dt$, entonces la ecuación anterior resulta

$$f_W(\omega) = \int \int f_{W|\Theta, Y}(\omega|\theta, y) dy d\theta \left[\int f_{\Theta, Y|W}(\theta, y|t) f_W(t) dt \right] \quad (37)$$

$$f_W(\omega) = \int \left[\int \int f_{W|\Theta, Y}(\omega|\theta, y) f_{\Theta, Y|W}(\theta, y|t) dy d\theta \right] f_W(t) dt \quad (38)$$

o, equivalentemente,

$$\begin{aligned} f_W(\omega) &= \int H(\omega, t) f_W(t) dt \\ &= M(f_W) \end{aligned} \quad (39)$$

con

$$H(\omega, t) = \int \int f_{W|\Theta, Y}(\omega|\theta, y) f_{\Theta, Y|W}(\theta, y|t) dy d\theta$$

De manera análoga se puede proceder con $f_{\Theta}(\theta)$ y $f_Y(y)$.

Es importante observar que la funcional $M(f)$ en este caso, no es única.

Es posible desarrollar las distribuciones conjuntas con otras descomposiciones dando lugar a núcleos alternativos a $H(\omega, t)$. En cualquier caso no es evidente que la iteración para tres variables; y en general para N variables, resuelva la ecuación de punto fijo en (39). Los lectores interesados pueden recurrir a Geman y Geman (1984).

Selección del tamaño de k

Como se ilustró anteriormente, el Gibbs Sampler genera una cadena de Markov de variables aleatorias, que converge a la distribución de interés. Muchos investigadores que trabajan con este algoritmo explotan esta propiedad para hacer la selección adecuada del valor de k .

Una estrategia general para seleccionar k , es determinar la convergencia de algún aspecto de la secuencias de Gibbs. Distintos autores que han trabajado con este algoritmo han producido varias sugerencias como, por ejemplo:

Tanner (1991) sugiere determinar una secuencia de pesos que midan la discrepancia entre lo muestreado y la distribución deseada.

Gerfald, Hills, Racine-Ponn y Smith (1990), después de haber aplicado el algoritmo a un extenso número de problemas (modelos jerárquicos, problemas de componentes de varianza, inferencia predictiva, etc.), dicen que para la implementación completa del algoritmo hay que determinar el tamaño de k y que, al incrementar las iteraciones, la selección de m debe ser especificada. Estos autores observan que los valores de k y m dependen de la aplicación particular. No obstante, para todas las aplicaciones que han desarrollado sólo han manejado valores $k \leq 50$ y $m \leq 1000$.

Primeramente utilizan el algoritmo con problemas donde tienen alternativas para comparar los resultados; trabajan con los datos generados permitiendo correr el programa hasta obtener la distribución marginal posterior para cada parámetro de interés haciéndolo de la siguiente manera: para una m fija incrementan k , graficando la densidad estimada (14) y observando si las estimaciones son visualmente distinguibles. Posteriormente incrementan el valor de m para apreciar la "estabilidad" de la densidad.

Teniendo los datos generados, se seleccionan entre 40 y 100 puntos espaciados igualmente en el dominio de la distribución, evaluando estos puntos en la densidad estimada (14). Después se obtiene una curva spline (cúbica) a través de esos valores y se com-

para con la verdadera. En ocasiones, es recomendable realizar una integración "simple trapezoidal" sobre la colección de los puntos de la densidad estimada para verificar que el resultado es muy cercano a 1.

Desafortunadamente estas sugerencias no cuentan con una teoría para el caso general.

De cualquier forma, parece ser más sencillo utilizar el algoritmo de Gibbs; aunque esto requiera de hacer diferentes variaciones en el valor de k y m , que hacer el cálculo de integrales similares a (11), tarea que se dificulta a medida que aumenta el número de variables.

CAPITULO II

Algoritmo de Rubin " Importance Sampling "

En la literatura estadística, con otros fines e incluso siguiendo otros paradigmas, existen algunos términos técnicos para el cálculo de distribuciones que en alguna forma están relacionados con el Gibbs Sampler. Con el propósito de completar la información que se proporciona en este trabajo, se incluye una breve descripción de dos de estos algoritmos.

Rubin(1987) sugirió un método de Monte Carlo no iterativo para generar distribuciones marginales.

Primeramente se presenta la idea básica para el caso de dos variables y posteriormente se generaliza.

Suponga que se trata de obtener la distribución marginal de Θ , teniendo sólo la forma funcional (módulo la constante de normalización) de la densidad conjunta $f_{\Theta|W}(\theta, \omega)$ y además se tiene disponible la densidad condicional $f_{\Theta|W}(\theta|\omega)$.

Se elige primeramente una distribución aproximada para la variable W , sea $f_{1W}(\omega)_s$.

La función $f_{\Theta|1W}(\theta|\omega)_s f_{1W}(\omega)_s$ provee una aproximación a la distribución de (Θ, W) .

Se obtienen los valores $(\Theta = \theta_i, W = \omega_i)$ con $i = 1, 2, \dots, M$ para esta distribución, de la siguiente manera: se obtiene un valor ω_0 de $f_{1W}(\omega)_s$, éste se sustituye en $f_{\Theta|1W}(\theta|\omega = \omega_0)$ para obtener el valor $\theta = \theta_0$ y así sucesivamente hasta obtener los M valores (θ, ω) .

Se calcula el valor $r_i = f_{\Theta|W}(\theta = \theta_i, \omega = \omega_i) / [f_{\Theta|1W}(\theta = \theta_i|\omega = \omega_i) * f_{1W}(\omega = \omega_i)_s]$ con $i = 1, 2, \dots, M$.

La densidad marginal de Θ se puede aproximar por:

$$\widehat{f_{\Theta}}(\theta) = \frac{\sum_{i=1}^M [f_{\Theta|1W}(\theta|\omega = \omega_i) * r_i]}{\sum_{i=1}^M r_i} \quad (40)$$

Dividiendo el numerador y el denominador de la ecuación anterior por M y utilizando

la ley de los grandes números se llega al siguiente resultado:

$$\widehat{f_{\Theta}}(\theta) \rightarrow f_{\Theta}(\theta) \text{ con probabilidad 1 conforme } M \rightarrow \infty \text{ para casi toda } \Theta.$$

El funcionamiento de (40) depende fuertemente, como es de esperar, de la selección de $f_W(\omega)_s$.

Además, si se tiene $f_{W|\Theta}(\omega|\theta)$ disponible, el estimador de la distribución marginal de W es:

$$\widehat{f_W}(\omega) = \sum_{i=1}^M [f_{W|\Theta}(\omega|\theta = \theta_i) \cdot r_i] / \sum_{i=1}^M r_i$$

Para 3 variables, se tiene la siguiente situación:

Suponga que se quiere obtener la distribución marginal de Θ , dada la forma funcional de $f_{\Theta W Y}(\theta, \omega, y)$ y se tiene disponible la función condicional $f_{\Theta|W Y}(\theta|\omega, y)$. En este caso la tupla (ω, y) juega el papel de ω en el caso de dos variables y, en general, se necesita especificar una distribución aproximada para W, Y , sea $f_{W Y}(\omega, y)_s$.

Sin embargo, si por ejemplo, $f_{W Y}(\omega, y)$ está disponible, sólo se requiere especificar a $f_Y(y)_s$. En cualquier caso; se obtienen las tuplas $(\theta_i, \omega_i, y_i)$ con $i = 1, 2, \dots, M$, como se hizo en el caso de dos variables, y se calcula

$$r_i = f(\theta_i, \omega_i, y_i) / [f(\theta_i|\omega_i, y_i) f(\omega_i, y_i)_s]$$

Obteniendo la densidad marginal estimada:

$$\widehat{f_{\Theta}}(\theta) = \sum_{i=1}^M [f_{\Theta|W Y}(\theta|\omega = \omega_i, y = y_i) \cdot r_i] / \sum_{i=1}^M r_i$$

Algoritmo de "Data Augmentation"

Tanner y Wong (1987) adoptan un procedimiento alternativo al Gibbs Sampler para resolver la integración en (23). En particular, la identidad (17) lleva al siguiente esquema iterativo. Si se conoce $f_{Y|\Theta}(\omega|\theta)$,

Dada una aproximación inicial $g_i(\omega)$ a $f_{Y|\Theta}(\omega)$:

a) Generan una muestra $\theta_1, \theta_2, \dots, \theta_n$ para la aproximación inicial para $f_{\Theta}(\theta)$.

b) Se renueva la aproximación inicial de $f_{Y|\Theta}(\omega)$ de la siguiente manera:

$$g_{i+1}(\omega) = \sum_{j=1}^n f(\omega|\theta_j)/n$$

y el proceso se repite hasta satisfacer algún criterio de convergencia.

Ahora bien, la primera etapa está compuesta por:

a1) la generación ω a partir de $g_i(\omega)$

a2) generar θ a partir de $f_{\Theta|Y}(\theta|t)$, con $t = \omega$ obtenido en el paso anterior.

A la etapa (a) se le llama "etapa de imputación" y a la (b), "etapa posterior". El algoritmo de Data Augmentation consiste de iteraciones sucesivas de la etapa de imputación y la posterior.

Dos consideraciones son interesantes para implementar este algoritmo: el criterio de la convergencia y la selección de imputaciones n por iteración. Tanner y Wong (1987) afirman que cuando el valor n es suficientemente grande, las dos etapas proveerán una aproximación muy cercana a una iteración de (24).

CAPITULO III

Aplicación del Algoritmo de Gibbs

Mendoza (1988) aborda el problema estadístico de producir inferencias sobre un cociente de combinaciones lineales de coeficientes de un modelo de regresión múltiple. El caso más simple, de medias Normales ha sido investigado por diversos autores y desde enfoques diferentes. En su versión más común, este problema se plantea de la siguiente manera:

Sean X_1, X_2, \dots, X_n y Y_1, Y_2, \dots, Y_m dos muestras aleatorias independientes entre sí, tales que X_i con $i = 1 \dots n$ tiene una distribución normal con media μ_1 y varianza σ^2 y cada Y_i con $i = 1 \dots m$ tiene una distribución Normal con media μ_2 y varianza σ^2 ; con μ_1, μ_2, σ^2 parámetros desconocidos. Suponiendo que $\mu_2 \neq 0$, el problema del cociente de medias consiste en producir inferencias acerca del parámetro:

$$\rho = \mu_1/\mu_2$$

que describe la magnitud relativa de las medias.

La situación en la que se requiere un análisis estadístico de este tipo aparece en muy diversas situaciones pero en particular en el contexto de Bioensayo.

Desde un punto de vista estructural, un bioensayo o ensayo biológico puede describirse como un conjunto de estímulos (ocasionalmente con un sólo elemento) aplicado bajo circunstancias específicas y utilizando varios niveles a determinadas unidades biológicas con la finalidad de producir un conjunto de respuestas observables. El objetivo general de un bioensayo es producir una descripción apropiada de la relación dosis-respuesta involucrada y uno de los problemas particulares más usuales es el de establecer la potencia relativa de dos diferentes estímulos. Si se considera un ensayo directo, el parámetro que describe la potencia relativa coincide con un cociente de medias.

El problema de cociente de medias ha sido considerado desde la perspectiva bayesiana, entre otros, por Kapperman, Geisser y Antle (1970) y Bernardo (1977).

Ante este panorama, considerese una situación mucho más general, el problema de obtener inferencias sobre el cociente de combinaciones lineales de coeficientes de un modelo de regresión múltiple:

Sean $Y \in R^N$ un vector aleatorio con distribución Normal Multivariada, con vector de medias $X\theta$ y matriz de covarianzas $\sigma^2 I$; donde X es una matriz ($N \times k$) de rango completo, $\theta \in R^k$ ($k < N$) es un vector de coeficientes desconocidos y σ también es un parámetro desconocido.

Bajo estas condiciones el parámetro de interés, sobre el que se requiere obtener inferencias, está definido por

$$\rho = \lambda_1^t \theta / \lambda_2^t \theta$$

en donde λ_1 y λ_2 son dos vectores fijos y linealmente independientes en R^k tales que $\lambda_2^t \theta \neq 0$.

Claramente, este problema incluye el del cociente de medias que se comentó anteriormente. Sea $L \in M_{k \times k}$ de rango completo tal que su transpuesta está dada por $L^t = (l_1, l_2, \dots, l_k)$ con $l_1 = \lambda_1$ y $l_2 = \lambda_2$, entonces, $\beta = L\theta$ satisface las relaciones

$$\beta_1 = \lambda_1^t \theta$$

y

$$\beta_2 = \lambda_2^t \theta$$

En otras palabras, es posible reparametrizar el modelo de manera que:

$$\begin{aligned} X\theta &= X(L^{-1}L)\theta \\ &= (XL^{-1})(L\theta) \\ &= Z\beta \end{aligned}$$

con $Z = XL^{-1}$ y $\beta = L\theta$, teniendo ahora

$$Y \sim N(Z\beta, \sigma^2 I)$$

y el parámetro de interés está dado entonces, por:

$$\rho = \beta_1/\beta_2$$

es decir

$$f(Y|Z, \beta, \sigma) = (2\pi\sigma^2)^{-N/2} \exp[-(Y - Z\beta)^t(Y - Z\beta)/(2\sigma^2)]$$

Haciendo una transformación adicional del espacio de parámetros para expresar esta densidad en términos de ρ se tiene que:

Sea $\gamma \in R^k$ un vector tal que

$$\gamma_1 = \rho$$

$$\gamma_i = \beta_i, i = 2 \dots k$$

Esta transformación es inyectiva cuya inversa está dada por:

$$\beta_1 = \gamma_1 \gamma_2$$

$$\beta_i = \gamma_i, i = 2 \dots k$$

de manera que β puede expresarse en términos de γ ($\beta = \beta(\gamma)$). Así, la función de verosimilitud de (γ, σ) puede expresarse como:

$$L(\gamma, \sigma|Y) = (2\pi)^{-N/2} \exp[-(Y - Z\beta(\gamma))^t(Y - Z\beta(\gamma))/(2\sigma^2)]$$

que involucra explícitamente a $\rho = \gamma_1$.

Aplicando el teorema de Bayes, partiendo de una densidad inicial $p(\gamma, \sigma)$ y dado un conjunto específico de observaciones Y , la densidad final se obtiene de la relación:

$$p(\gamma, \sigma|Y) \propto p(\gamma, \sigma)L(\gamma, \sigma|Y)$$

Si la densidad inicial conjunta se escribe como:

$$p(\gamma, \sigma) = p(\rho, \beta_2, \dots, \beta_k, \sigma) = p(\beta_2, \dots, \sigma|\rho)p(\rho)$$

se observa que la información inicial se puede descomponer en 2 partes; la que se posee sobre el parámetro de interés descrita por $p(\rho)$ y la que se conoce sobre los demás parámetros cuando se supone conocido el parámetro de interés y que se describe mediante $p(\beta_2, \dots, \beta_k, \sigma | \rho)$. Mendoza (1988) propone, tomando en cuenta que el interés se centra en ρ , que se utilice para $\beta_2, \beta_3, \dots, \beta_k | \rho$ una distribución de referencia. Específicamente propone:

$$p(\beta_2, \dots, \beta_k, \sigma | \rho) \propto \sigma^{-r}$$

para algún $r > 0$. Teniendo así, que:

$$p(\gamma, \sigma) \propto p(\rho) \sigma^{-r}$$

por lo tanto

$$\begin{aligned} p(\gamma, \sigma | Y) &\propto p(\gamma, \sigma) L(\gamma, \sigma | Y) \\ &\propto [p(\rho) \sigma^{-r}] (2\pi\sigma^2)^{-N/2} \exp[-(Y - Z\beta(\gamma))^2 / (2\sigma^2)] \end{aligned}$$

de donde la densidad marginal final de γ se puede obtener como:

$$\begin{aligned} p(\gamma | Y) &= \int p(\gamma, \sigma | Y) d\sigma \\ &\propto \int p(\gamma, \sigma) L(\gamma, \sigma | Y) d\sigma \\ &\propto p(\rho) \int \sigma^{-(N+r)} \exp[-B/\sigma^2] d\sigma \end{aligned}$$

con $B = \{(Y - Z\beta(\gamma))^2 / (2\sigma^2)\}$.

Aplicando el cambio de variable $u = \sigma^{-2}$ se tiene:

$$\begin{aligned} \int \sigma^{-(N+r)} \exp[-B/\sigma^2] d\sigma &= - \int_0^\infty u^{(N+r)/2} \exp[-Bu] u^{-3/2} (1/2) du \\ &= (1/2) \int_0^\infty u^{(N+r-3)/2} \exp[-Bu] du \\ &= (1/2) \int_0^\infty u^{A-1} \exp[-Bu] du \end{aligned}$$

con $A = (N + r - 1)/2$. El integrando de la última expresión corresponde al núcleo de una densidad gamma con parámetros A y B, de manera que se puede concluir que:

$$p(\gamma|Y) \propto p(\rho)B^{-A} \\ = p(\rho)[(Y - Z\beta(\gamma))^t(Y - Z\beta(\gamma))]^{-(N+r-1)/2} \quad (41)$$

Recordando que $\gamma^t = (\gamma_1, \dots, \gamma_k) = (\rho, \gamma_2, \dots, \gamma_k) = (\rho, \gamma_{2-})$ y haciendo uso de un poco de álgebra lineal se tiene que:

$$(Y - Z\beta(\gamma)) = Y - (\rho\gamma_2 z_1 + \gamma_2 z_2 + \dots + \gamma_k z_k) \\ = Y - (\gamma_2(z_2 + \rho z_1) + \gamma_3 z_3 + \dots + \gamma_k z_k) \quad (42)$$

Definiendo la matriz $W = (w_1, \dots, w_{k-1})$ ($N \times (k-1)$) como

$$(w_1, \dots, w_{k-1}) = (\rho z_1 + z_2, z_3, \dots, z_k)$$

se tiene que la expresión (42) queda como:

$$Y - (\gamma_2 w_1 + \gamma_3 w_2 + \dots + \gamma_k w_{k-1}) = Y - W\gamma_{2-}$$

Recurriendo a esta última expresión, se puede escribir lo siguiente:

$$p(\gamma|Y) \propto p(\rho)[(Y - W\gamma_{2-})^t(Y - W\gamma_{2-})]^{-(N+r-1)/2}$$

donde (ver Mendoza (1988), pág 36 y 37):

$$(Y - W\gamma_{2-})^t(Y - W\gamma_{2-}) = [1 + (\gamma_{2-} - \mu)^t(W^t W)(\gamma_{2-} - \mu)/H(\rho)]H(\rho)$$

con

$$\mu = (W^t W)^{-1} W^t Y$$

$$H(\rho) = (Y - W\mu)^t(Y - W\mu)$$

utilizando estas últimas expresiones, se tiene:

$$p(\gamma|Y) \propto p(\rho)H(\rho)^{-(N+r-1)/2} [1 + (\gamma_{2-} - \mu)^t \frac{W^t W}{H(\rho)} (\gamma_{2-} - \mu)]^{-(N+r-1)/2}$$

con $[1 + (\gamma_{2-} - \mu)^t \frac{W^t W}{H(\rho)} (\gamma_{2-} - \mu)]^{-(N+r-1)/2}$ el núcleo de una distribución t multivariada con $N + r - k$ grados de libertad, matriz de precisión $T = (N + r - k)(W^t W)/H(\rho)$ y vector de localización μ . Posteriormente Mendoza (1988) obtiene $P(\rho|Y)$ integrando $P(\gamma|Y)$ respecto a $\gamma_2, \dots, \gamma_k$.

En contraste con ese desarrollo teórico, aquí se obtendrá la $P(\rho|Y)$ utilizando el algoritmo de Gibbs; para ello, es conveniente observar que;

$$p(\gamma|Y) = p(\gamma_1 = \rho, \gamma_{2-}|Y) \propto p(\rho)[(Y - W\gamma_{2-})^t(Y - W\gamma_{2-})]^{-(N+r-1)/2}]$$

así que las condicionales tienen las formas siguientes

$$\begin{aligned} p(\gamma_{2-}|\rho, Y) &= p(\rho, \gamma_{2-}|Y)/P(\rho|Y) \\ &= c p(\rho) H(\rho)^{-(N+r-1)/2} \left[1 + \frac{1}{N+r-k} (\gamma_{2-} - \mu)^t T(\gamma_{2-} - \mu) \right]^{-(N+r-1)/2} / p(\rho|Y) \\ &= c H(\rho)^{-(N+r-1)/2} \frac{p(\rho)}{p(\rho|Y)} \left[1 + \frac{1}{N+r-k} (\gamma_{2-} - \mu)^t T(\gamma_{2-} - \mu) \right]^{-(N+r-1)/2} \\ &= c H(\rho)^{-(N+r-1)/2} c(\rho) \left[1 + \frac{1}{N+r-k} (\gamma_{2-} - \mu)^t T(\gamma_{2-} - \mu) \right]^{-(N+r-1)/2} \end{aligned}$$

en donde el factor $c(\rho)$ mide el cambio en credibilidad del valor ρ .

Por lo tanto la distribución de $(\gamma_{2-}|\rho, Y)$ es una t multivariada con $N + r - k$ grados de libertad, matriz de precisión $T = (N + r - k)(W^t W)/(Y - W\mu)^t(Y - W\mu)$ y vector de localización $\mu = (W^t W)^{-1} W^t Y$.

$$\begin{aligned} p(\rho|\gamma_{2-}, Y) &= p(\rho, \gamma_{2-}|Y)/p(\gamma_{2-}|Y) \\ &= c_2 p(\rho) [(Y - W\gamma_{2-})^t(Y - W\gamma_{2-})]^{-(N+r-1)/2} / p(\gamma_{2-}|Y) \\ &= c_3 p(\rho) [(Y - W\gamma_{2-})^t(Y - W\gamma_{2-})]^{-(N+r-1)/2} \end{aligned}$$

pero

$$[(Y - W\gamma_2)^t(Y - W\gamma_2)] = \sum_{i=1}^N (y_i - (z_{i2} + \rho z_{i1})\gamma_2 - z_{i3}\gamma_3 - \dots - z_{ik}\gamma_k)^2$$

Sea $a_i = -z_{i1}\gamma_2$ y $b_i = y_i - z_{i2}\gamma_2 - \dots - z_{ik}\gamma_k$ con $i = 1, \dots, N$, y $M = N + r - 1$.

Se tiene que:

$$[(Y - W\gamma_2)^t(Y - W\gamma_2)]^{-M/2} = \left[\sum_{i=1}^N (a_i\rho + b_i)^2 \right]^{-M/2}$$

pero

$$(a_i\rho + b_i)^2 = a_i^2\rho^2 + 2a_ib_i\rho + b_i^2$$

$$\begin{aligned} \sum_{i=1}^N (a_i\rho + b_i)^2 &= \rho^2 \sum a_i^2 + 2\rho \sum a_ib_i + \sum b_i^2 \\ &= (\sum a_i^2)[\rho^2 + 2\rho \sum a_ib_i / \sum a_i^2] + \sum b_i^2 \\ &= (\sum a_i^2)[\rho^2 + 2\rho \sum a_ib_i / \sum a_i^2 + (\sum a_ib_i / \sum a_i^2)^2 \\ &\quad - (\sum a_ib_i / \sum a_i^2)^2] + \sum b_i^2 \\ &= (\sum a_i^2)[\rho - (-\sum a_ib_i / \sum a_i^2)]^2 - (\sum a_i^2)(\sum a_ib_i / \sum a_i^2)^2 \\ &\quad + \sum b_i^2 \\ &= S \left[\frac{(\sum a_i^2)[\rho - (-\sum a_ib_i / \sum a_i^2)]^2}{S} + 1 \right] \end{aligned}$$

con $S = -(\sum a_i^2)(\sum a_ib_i / \sum a_i^2)^2 + \sum b_i^2$

por lo tanto

$$[(Y - W\gamma_2)^t(Y - W\gamma_2)]^{-M/2} = S^{-M/2} \left[\frac{\sum a_i^2 [\rho - (-\sum a_ib_i / \sum a_i^2)]^2}{S} + 1 \right]^{-M/2}$$

donde $B = \left[\frac{\sum a_i^2 (\rho - (-\sum a_i b_i / \sum a_i^2))^2}{S} + 1 \right]$ corresponde al kernel de una distribución t de student univariada con $\alpha = N + r - 2$ grados de libertad, $\mu = -\sum a_i b_i / \sum a_i^2$ parámetro de localización y $\tau = (\sum a_i^2) \alpha / S$ parámetro de precisión.

Si en particular se toma $P(\rho)$ constante, se tiene que la distribución de $(\rho | \gamma_2, Y)$ es entonces una t de student con los parámetros mencionados anteriormente.

Así, teniendo estas dos condicionales, lo que resta es hacer la simulación correspondiente.

A continuación se describe con más detalle el modelo de cociente de pendientes para, posteriormente, utilizar el algoritmo de Gibbs en un ejemplo numérico de este modelo y poder hacer comparaciones de los resultados obtenidos a través de este camino con los resultados obtenidos por Mendoza (1988); quién estudia el mismo ejemplo.

Modelo: Suponga que se realiza un experimento en el que p dosis X_{11}, \dots, X_{1p} de un primer estímulo y q dosis X_{21}, \dots, X_{2q} de un segundo estímulo se ensayan n veces de manera que se obtiene un conjunto $Y_{1jk}; j = 1, \dots, p, k = 1, \dots, n; Y_{2jk}; j = 1, \dots, q, k = 1, \dots, n$ de observaciones condicionalmente independientes, con distribución Normal y varianza común σ^2 tales que

$$E(Y_{1jk}) = \alpha + \pi X_{1j}$$

$$E(Y_{2jk}) = \alpha + \rho \pi X_{2j}$$

Así, Y_{ijk} representa la k -ésima medición de la variable Y asociada a la aplicación de X_{ij} ; la j -ésima dosis del i -ésimo estímulo.

Si las dosis X_1 del primer estímulo y X_2 del segundo estímulo son equivalentes; esto es si las correspondientes respuestas esperadas son iguales [$E(Y_1) = E(Y_2)$], entonces su potencia relativa se define simplemente como el cociente X_1/X_2 . Utilizando la estructura anterior se llega a que:

$$X_1/X_2 = \rho(\pi / \pi) = \rho$$

Este valor existe si $\pi \neq 0$ y es independiente de la particular pareja de dosis equivalentes, como consecuencia de la ordenada común en la estructura. Esta estructura se conoce en la literatura con el nombre de cociente de pendientes (Bliss (1970), Armitage (1977)).

El modelo que describe el ensayo de cociente de pendientes es un caso particular del modelo de regresión lineal múltiple, sólo se requiere establecer las siguientes relaciones:

$$Y = \begin{bmatrix} Y_{111} \\ Y_{112} \\ \vdots \\ Y_{1pn} \\ Y_{211} \\ Y_{212} \\ \vdots \\ Y_{2qn} \end{bmatrix}$$

$$X = \begin{bmatrix} 1 & X_{11} & 0 \\ 1 & X_{11} & 0 \\ \vdots & \vdots & \vdots \\ 1 & X_{1p} & 0 \\ 1 & 0 & X_{21} \\ 1 & 0 & X_{21} \\ \vdots & \vdots & \vdots \\ 1 & 0 & X_{2q} \end{bmatrix}$$

el vector $Y (\in R^N, n = n(p+q))$ se obtiene de ordenar lexicográficamente las respuestas observadas, la matriz $X (N \times 3)$ tiene la primera columna con todas sus entradas iguales a uno, la segunda columna con las últimas nq entradas iguales a cero y las primeras np entradas en p bloques de tamaño n de manera que todas las entradas del j -ésimo bloque sean iguales a X_{1j} ; la tercera columna de X tiene las np primeras entradas iguales a cero y las restantes nq agrupadas en q bloques de tamaño n de forma que todas las entradas del j -ésimo bloque son iguales a X_{2j} .

Si se define a:

$$\theta = \begin{bmatrix} \alpha \\ \pi \\ \rho\pi \end{bmatrix}$$

$$\lambda_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

$$\lambda_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

se tiene que

$$Y \sim N(X\theta, \sigma^2 I) \quad (43)$$

Para pasar al modelo reparametrizado (43), es necesario construir la matriz L tal que

$$L^t = (\lambda_1, \lambda_2, \lambda_3)$$

sea de rango completo; puesto que $\lambda_1 = e_3, \lambda_2 = e_1$, con e_i el i -ésimo vector canónico en R^3 , se tiene que $\lambda_3 = e_1$. De esta forma,

$$L = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

y entonces

$$L^{-1} = L$$

de donde se sigue que si

$$Z = XL^{-1}$$

entonces

$$Z = \begin{bmatrix} 0 & X_{11} & 1 \\ 0 & X_{11} & 1 \\ \vdots & \vdots & \vdots \\ 0 & X_{1p} & 1 \\ X_{21} & 0 & 1 \\ X_{21} & 0 & 1 \\ \vdots & \vdots & \vdots \\ X_{2q} & 0 & 1 \end{bmatrix}$$

En forma análoga, como $\beta = L\theta$, entonces

$$\begin{aligned}\beta &= \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} \\ &= \begin{bmatrix} \rho\pi \\ \pi \\ \alpha \end{bmatrix}\end{aligned}$$

y se tiene el modelo reparametrizado

$$Y \sim N(Z\beta, \sigma^2 I)$$

y la reexpresión del parámetro de interés

$$\rho = \beta_1 / \beta_2$$

Haciendo la reparametrización adicional del espacio de parámetros; se define al vector $\gamma \in R^3$, como:

$$\begin{aligned}\gamma &= \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{bmatrix} \\ &= \begin{bmatrix} \rho \\ \beta_2 \\ \beta_3 \end{bmatrix} \\ &= \begin{bmatrix} \rho \\ \pi \\ \alpha \end{bmatrix}\end{aligned}$$

Ejemplo Numérico: La siguiente tabla, representa los datos obtenidos de un ensayo biológico reportado por Wood (1964); apareciendo en el libro de Finney (1978, pág 161), el cual es analizado por Mendoza (1988).

Datos de Wood ($n = 4, p = 3, q = 2$)

Estímulo

1

2

ESTA TESIS NO DEBE
SALIR DE LA BIBLIOTECA

Dosis	0.0	0.5	1.0	0.5	1.0
	38	97	167	80	121
Respuesta	45	100	164	88	124
	40	105	159	90	122
	44	98	156	82	122

Utilizando las relaciones especificadas anteriormente, con este ejemplo se tiene que:

$$Y = \begin{bmatrix} 38 \\ 45 \\ 40 \\ 44 \\ 97 \\ 100 \\ 105 \\ 98 \\ 167 \\ 164 \\ 159 \\ 156 \\ 80 \\ 88 \\ 90 \\ 82 \\ 121 \\ 124 \\ 122 \\ 122 \end{bmatrix}$$

$$X^t = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & .5 & .5 & .5 & .5 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .5 & .5 & .5 & .5 & 1 & 1 & 1 & 1 \end{bmatrix}$$

$$Z^t = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .5 & .5 & .5 & .5 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & .5 & .5 & .5 & .5 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

$$W^t = \begin{bmatrix} 0 & 0 & 0 & 0 & .5 & .5 & .5 & .5 & 1 & 1 & 1 & 1 & .5\rho & .5\rho & .5\rho & .5\rho & \rho & \rho & \rho & \rho \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

$$W^t W = \begin{bmatrix} 5 + 5\rho^2 & 6 + 6\rho \\ 6 + 6\rho & 20 \end{bmatrix}$$

$$(W^t W)^{-1} = \frac{1}{(64 + \rho(64\rho - 72))} \begin{bmatrix} 20 & -6(1 + \rho) \\ -6(1 + \rho) & 5(1 + \rho^2) \end{bmatrix}$$

$$W^t Y = \begin{bmatrix} 846 + 659\rho \\ 2042 \end{bmatrix}$$

$$\mu = \frac{1}{(64 + \rho(64\rho - 72))} \begin{bmatrix} 4668 + 928\rho \\ 5134 - 9030\rho + 6256\rho^2 \end{bmatrix}$$

$$T = \frac{20}{c} \begin{bmatrix} 5(1 + \rho^2)(64 + 64\rho^2 - 72\rho)^2 & 6(1 + \rho)(64 + 64\rho^2 - 72\rho)^2 \\ 6(1 + \rho)(64 + 64\rho^2 - 72\rho)^2 & 20(64 + 64\rho^2 - 72\rho)^2 \end{bmatrix}$$

$$\text{con } c = 60154624 + 127012928\rho^4 - 316680192\rho^3 + 382667824\rho^2 - 24143456\rho$$

Considerando una $r = k = 3$ (Mendoza (1988)) se tiene que la distribución de $(\alpha, \beta | \rho, Y)$ es una t-multivariada con $N + r - k = 20$ grados de libertad, con matriz de precisión T , y vector de localización μ .

La distribución de $(\rho | \alpha, \beta, Y)$ es una t-student con $N + r - 2 = 21$ grados de libertad, con media $\mu_1 = (-659 + 6\beta + 3500\alpha)/(5000\alpha)$ y precisión

$$\tau = (105000000\alpha^2)/(766709\alpha^2 + 25\alpha^4 + 64\alpha^2\beta^2 + 7968\beta\alpha^2 - 8460\alpha^3 - 20420\alpha\beta).$$

Los resultados vía Gibbs Sampler se pueden observar en la siguiente página.

VALORES DE RHO OBTENIDOS MEDIANTE EL PROC.DE GIBBS
 CON K = 30 Y M = 500 (VALORES YA ORDENADOS)

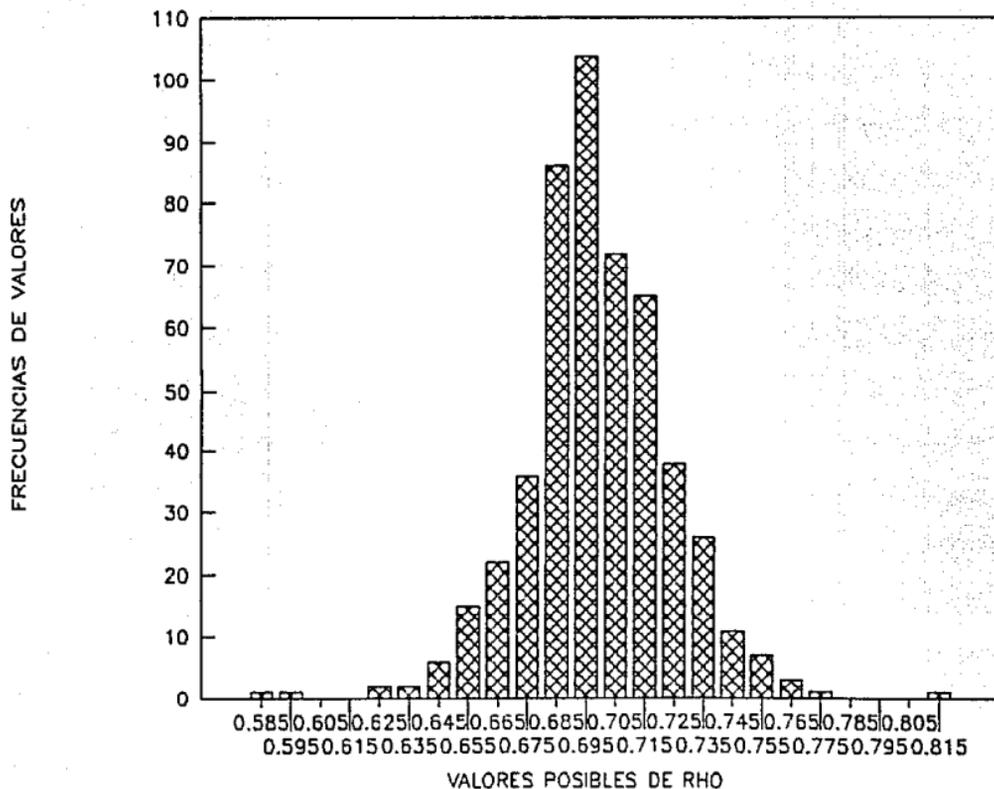
1	0.5842903	0.6707418	0.6820506	0.6879170	0.6932453
2	0.5921388	0.6711855	0.6820899	0.6879945	0.6932823
3	0.6212789	0.6712037	0.6821288	0.6880286	0.6934532
4	0.6245181	0.6713173	0.6822132	0.6880659	0.6934630
5	0.6392385	0.6714526	0.6823514	0.6882339	0.6935771
6	0.6399429	0.6723709	0.6825255	0.6882439	0.6935828
7	0.6420659	0.6727214	0.6826054	0.6884750	0.6937883
8	0.6431045	0.6729662	0.6826307	0.6886485	0.6938176
9	0.6447887	0.6736825	0.6826738	0.6886768	0.6941069
10	0.6463383	0.6737787	0.6829101	0.6887188	0.6941254
11	0.6463444	0.6739896	0.6829758	0.6887228	0.6941500
12	0.6471183	0.6741272	0.6829864	0.6887479	0.6941861
13	0.6490130	0.6743698	0.6830999	0.6888419	0.6942969
14	0.6507910	0.6744284	0.6836879	0.6890711	0.6943471
15	0.6515603	0.6745985	0.6844464	0.6891323	0.6946389
16	0.6530784	0.6746310	0.6844890	0.6892016	0.6946551
17	0.6535182	0.6747461	0.6845683	0.6893047	0.6948335
18	0.6551132	0.6749987	0.6847752	0.6893073	0.6950791
19	0.6555227	0.6750429	0.6848105	0.6894275	0.6952457
20	0.6558382	0.6753168	0.6848588	0.6894402	0.6952947
21	0.6561035	0.6755472	0.6848992	0.6894841	0.6954828
22	0.6564348	0.6758175	0.6849597	0.6898132	0.6955105
23	0.6569791	0.6759584	0.6852501	0.6901188	0.6955329
24	0.6570305	0.6768956	0.6853843	0.6903967	0.6955412
25	0.6572817	0.6769646	0.6853858	0.6904436	0.6956695
26	0.6574594	0.6770522	0.6854220	0.6904685	0.6958718
27	0.6590239	0.6772996	0.6854702	0.6905464	0.69595754
28	0.6598714	0.6776091	0.6855429	0.6908007	0.6960051
29	0.6608925	0.6776502	0.6855851	0.6907238	0.6960921
30	0.6613277	0.6779056	0.6856886	0.6910579	0.6961231
31	0.6627594	0.6779718	0.6857783	0.6910654	0.6961873
32	0.6632434	0.6782658	0.6860014	0.6912919	0.6963158
33	0.6634041	0.6789758	0.6860124	0.6913197	0.6964187
34	0.6649035	0.6791109	0.6861734	0.6913212	0.6965402
35	0.6652782	0.6791861	0.6862454	0.6913831	0.6965488
36	0.6652964	0.6799263	0.6862600	0.6915242	0.6967447
37	0.6653718	0.6800281	0.6862943	0.6917497	0.6968198
38	0.6657240	0.6802670	0.6862980	0.6918980	0.6970142
39	0.6661888	0.6802922	0.6864995	0.6919093	0.6970830
40	0.6664850	0.6804310	0.6865559	0.6921687	0.6970838
41	0.6665875	0.6804904	0.6866320	0.6921697	0.6971364
42	0.6666321	0.6805142	0.6867488	0.6922132	0.6973049
43	0.6668075	0.6805384	0.6868751	0.6922760	0.6974069
44	0.6672800	0.6806207	0.6868883	0.6924948	0.6974192
45	0.6675181	0.6807598	0.6871648	0.6924994	0.6977338
46	0.6677180	0.6814123	0.6872689	0.6925659	0.6977413
47	0.6682212	0.6815772	0.6873344	0.6925948	0.6977630
48	0.6682904	0.6817350	0.6874993	0.6927843	0.6979357
49	0.6692991	0.6817714	0.6875580	0.6928049	0.6979624
50	0.6693287	0.6819871	0.6875771	0.6932408	0.6980427

VALORES DE RHO OBTENIDOS MEDIANTE EL PROC.DE GIBBS
 CON K=30 Y M=500 (VALORES YA ORDENADOS)

51	0.6981580	0.7027310	0.7101480	0.7188253	0.7298804
52	0.6982668	0.7028035	0.7102381	0.7188683	0.7301597
53	0.6982819	0.7028137	0.7107390	0.7189023	0.7308644
54	0.6983052	0.7028834	0.7108153	0.7189918	0.7309827
55	0.6983626	0.7030656	0.7110202	0.7191429	0.7314386
56	0.6984715	0.7031040	0.7110446	0.7192878	0.7317483
57	0.6985214	0.7031752	0.7111224	0.7194454	0.7322120
58	0.6985545	0.7031989	0.7113411	0.7194565	0.7324449
59	0.6986739	0.7034902	0.7113543	0.7194841	0.7326804
60	0.6987231	0.7035641	0.7113719	0.7196742	0.7326812
61	0.6988423	0.7037548	0.7117291	0.7197565	0.7327694
62	0.6988956	0.7037659	0.7118886	0.7199368	0.7335442
63	0.6989263	0.7040118	0.7119784	0.7199713	0.7337863
64	0.6990619	0.7040245	0.7120072	0.7201631	0.7339552
65	0.6990852	0.7046520	0.7120858	0.7207442	0.7342634
66	0.6991042	0.7046970	0.7121259	0.7211230	0.7344089
67	0.6993151	0.7049231	0.7121397	0.7212654	0.7354425
68	0.6993517	0.7050826	0.7125621	0.7222483	0.7354986
69	0.6993980	0.7053048	0.7126511	0.7228401	0.7355083
70	0.6994572	0.7059995	0.7130631	0.7226868	0.7357689
71	0.6996214	0.7061400	0.7131252	0.7227336	0.7362628
72	0.6997697	0.7062606	0.7131432	0.7227801	0.7370071
73	0.6998200	0.7065173	0.7132695	0.7228425	0.7374092
74	0.6999190	0.7065683	0.7135360	0.7228584	0.7376146
75	0.6999578	0.7068608	0.7136042	0.7230865	0.7385014
76	0.6999811	0.7068835	0.7136047	0.7233793	0.7392598
77	0.7001387	0.7068072	0.7136845	0.7234789	0.7398889
78	0.7001392	0.7068282	0.7139643	0.7240695	0.7418090
79	0.7001499	0.7069650	0.7139914	0.7244634	0.7421836
80	0.7001839	0.7070170	0.7139993	0.7247875	0.7432981
81	0.7002639	0.7070601	0.7141121	0.7256981	0.7436486
82	0.7004063	0.7074808	0.7142808	0.7260578	0.7438209
83	0.7004115	0.7077351	0.7148726	0.7262215	0.7447879
84	0.7007162	0.7078981	0.7150076	0.7263955	0.7448508
85	0.7007562	0.7079101	0.7150109	0.7268570	0.7483489
86	0.7008491	0.7079195	0.7150652	0.7267946	0.7492191
87	0.7010754	0.7079769	0.7153247	0.7268915	0.7493598
88	0.7011404	0.7080594	0.7156385	0.7269135	0.7496116
89	0.7013967	0.7083161	0.7160006	0.7271125	0.7506490
90	0.7015141	0.7084273	0.7164688	0.7271859	0.7507541
91	0.7015394	0.7088408	0.7166695	0.7274276	0.7511692
92	0.7016442	0.7089748	0.7171435	0.7276644	0.7523615
93	0.7018720	0.7089908	0.7171741	0.7277331	0.7530287
94	0.7016821	0.7090457	0.7172667	0.7278111	0.7537182
95	0.7017064	0.7091607	0.7174631	0.7279959	0.7572370
96	0.7019678	0.7092476	0.7175457	0.7284580	0.7843908
97	0.7020708	0.7096602	0.7178120	0.7286113	0.7662623
98	0.7021819	0.7098006	0.7183902	0.7289859	0.7669267
99	0.7022835	0.7100030	0.7184720	0.7292447	0.7754517
100	0.7027270	0.7100831	0.7185757	0.7296911	0.8133573

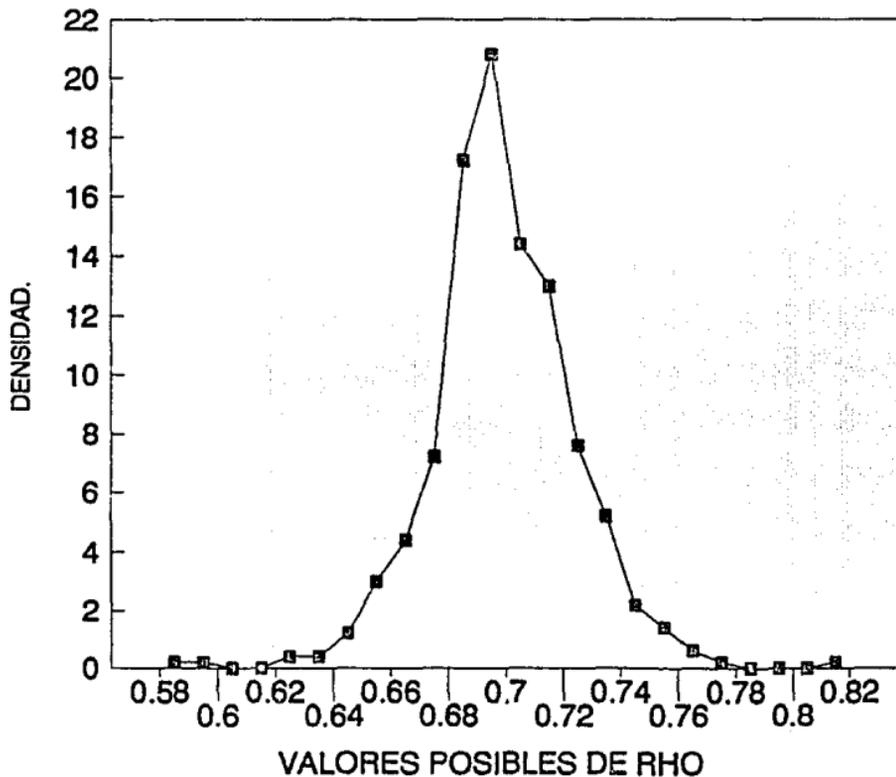
GRAFICA DE FRECUENCIAS DE RHO

CON K=30 Y M=500

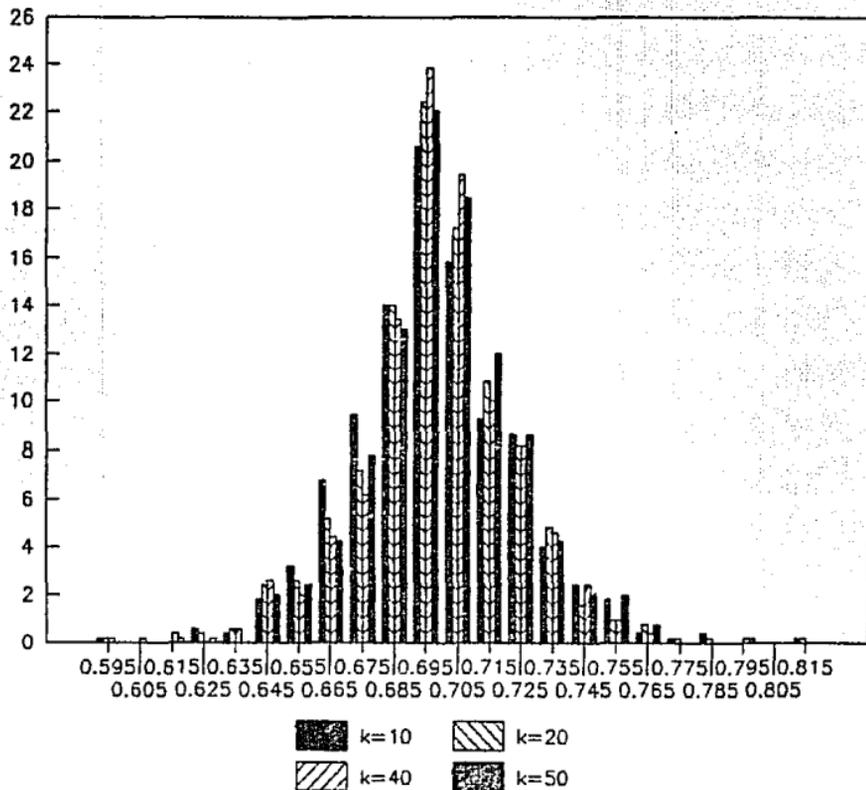


GRAFICA DE LA FUN. DE DENSIDAD DE RHO

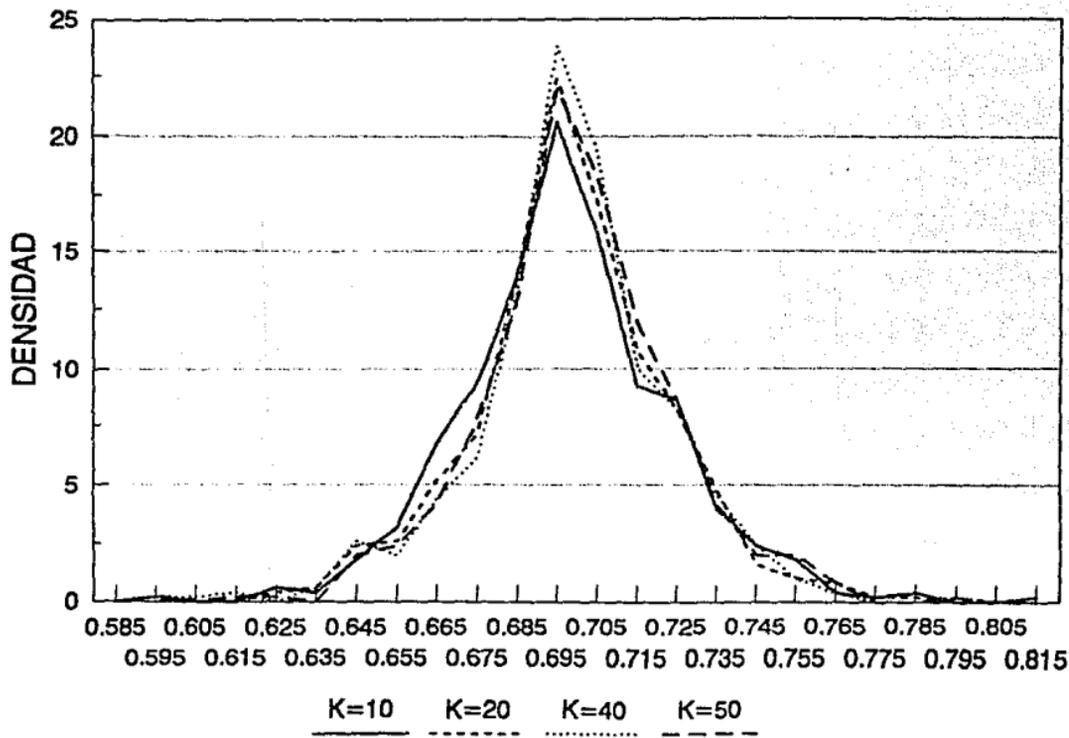
UTILIZANDO LA MUESTRA OBTENIDA CON $K=30$, $M=500$

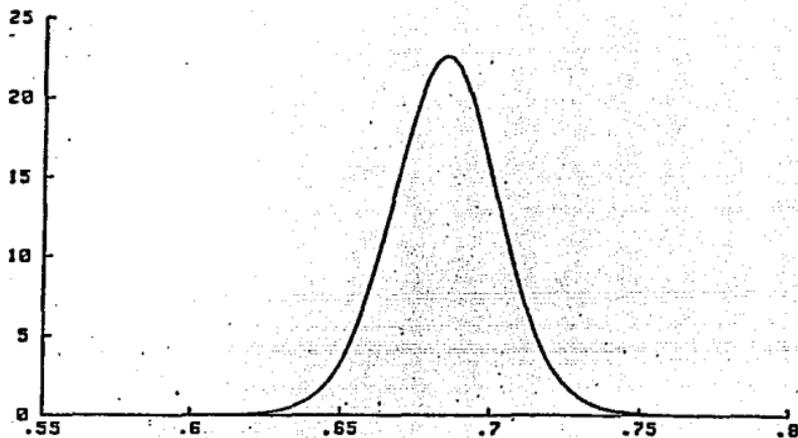


DISTINTAS K



GRAFICA COMPARATIVA UTILIZANDO DISTINTOS VALORES DE K





Gráfica obtenida por Mendoza (1988)



Gráfica obtenida mediante el algoritmo (gráf. de frec.)

Conclusiones.

Las dificultades técnicas que aparecen en el cálculo de densidades marginales posteriores, necesarias para la inferencia bayesiana, han sido un impedimento para la aplicación completa de la teoría de Bayes a datos Reales.

A pesar de que actualmente existen un gran número de avances técnicos de aproximaciones analíticas y numéricas, la implementación de éstas técnicas frecuentemente requieren de experiencia en el manejo de este tipo de aproximaciones, así como del conocimiento de Software especializado.

Una alternativa en estas condiciones son los métodos de cálculo de de distribuciones vía simulación. En particular, en los últimos años ha logrado popularidad el llamado Gibbs Sampler. Este procedimiento resuelve el problema simulando muestras aleatorias de las distribuciones de interés que pueden ser utilizadas para obtener aproximaciones de cualquier característica de la distribución bajo estudio. El Gibbs Sampler requiere que el usuario sea capaz de generar observaciones de algunas distribuciones condicionales y en la práctica se ha observado que si bien es un método alternativo, su tasa de convergencia es muy alta de manera que además de ser sencillo, típicamente es rápido y por tanto, barato.

El Gibbs Sampler, lo mismo que algunos otros algoritmos similares, ha tenido un gran impacto sobre el desarrollo de la inferencia bayesiana. Una gran cantidad de problemas de aplicación que se consideraban relativamente inaccesibles son ahora fácilmente abordables.

Aunque la mayoría de las aplicaciones del Gibbs Sampler se refieren a problemas bayesianos, también es útil en cálculos de inferencia clásica.

De esta manera, el Gibbs Sampler constituye una herramienta de una indudable utilidad en la Estadística Aplicada.

APENDICE

```
PROGRAM SAMPLER_T(LUMI,MULT);
```

```
uses crt;
```

```
{ ESTE PROGRAMA GENERA DOS MUESTRAS ALEATORIAS DE TAMAÑO  
300, QUE PROVIENEN DE DOS DIFERENTES VECTORES ALEATORIOS,  
UNA, PROVIENE DE UNA VARIABLE ALEATORIA QUE TIENE DIST. _  
CONDICIONAL; A UN VECTOR ALEATORIO, T DE STUDENT. LA OTRA _  
PROVIENE DE UN VECTOR ALEATORIO QUE TIENE DIST. CONDICIO-  
NAL; A LA VARIABLE ALEATORIA ANTERIOR, T MULTIVARIADA EN _  
R^2. ESTE PROGRAMA SE BASA EN EL ALGORITMO DE GIBBS SAM-  
PLER.}
```

```
TYPE
```

```
VECTOR=ARRAY[1..2] OF REAL;
```

```
VAR
```

```
A,X,MU,B;VECTOR;  
RHOO,RHO1,Z;REAL;  
K,M,I;INTEGER;  
ARCHI;TEXT;
```

```
PROCEDURE NORMAL(VAR U1,U2:REAL; VAR X,Y:REAL);
```

```
(ESTE PROC. GENERA DOS MUESTRAS DE TAM.1 DE VAR. ALEATO.)  
(NORMAL(0,1) INDEPENDIENTES)
```

```
VAR
```

```
T,E,R,C,AUX;REAL;  
BEGIN  
  REPEAT  
    T:=(U1-0.5)*2.0;  
    E:=(U2-0.5)*2.0;  
    R:=SQRT(T)+SQRT(E);  
    U1:=random;  
    U2:=random;  
  UNTIL R<=1;  
  AUX:=ABS((2+1)*R)/R;  
  C:=SQRT(AUX);  
  X:=C*T;  
  Y:=C*E  
END;
```

```
FUNCTION JI_CUAD(GRADLIB:INTEGER):REAL;
```

```
(ESTA FUNCION GENERA UNA MUESTRA DE TAM.1 DE UNA JI_CUAD CON)  
(*GRADLIB* GRADOS DE LIBERTAD)
```

```
VAR
```

```
J1,U1,U2,X1,X2:REAL;  
A;INTEGER;  
BEGIN  
  J1:=0.00;  
  X1:=0.00;  
  X2:=0.00;
```

```

IF GRADLIB MOD 2=0 THEN
  A:=GRADLIB DIV 2
ELSE
  A:=(GRADLIB+1) DIV 2;
FOR I:=1 TO A DO
  BEGIN
    U1:=RANDOM;
    U2:=RANDOM;
    NORMAL(U1,U2,X1,X2);
    J1:=J1+SQR(X1)+SQR(X2);
    IF ((I=1) AND (GRADLIB MOD 2 > 0)) THEN
      J1:=J1-SQR(X1);
    END;
  J1_CUAD:=J1;
END;

```

```

FUNCTION T_UNIV(U1,U2:REAL):REAL;
(FUNCION QUE GENERA UNA T DE ESTUDENT CENTRAL)
(COM 21 GRADOS DE LIBERTAD)
VAR
  X1,X2,CHI,NUM,DEN:REAL;
BEGIN
  X1:=0.00;
  X2:=0.00;
  CHI:=J1_CUAD(21);
  NORMAL(U1,U2,X1,X2);
  NUM:=X1*SQR(21);
  DEN:=SQR(CHI);
  T_UNIV:=NUM/DEN;
END;

```

```

FUNCTION TRANSFORMA(B2,B3:REAL):REAL;
(FUNCION QUE GENERA UNA T DE ESTUDENT CON MEDIA PAR2)
(Y PRECISION PARI COM 21 GRADOS DE LIBERTAD)
VAR
  U1,U2,PARI,PAR2,TST,A,B,AUI:REAL;
BEGIN
  U1:=RANDOM;
  U2:=RANDOM;
  AUI:=ABS(766709+SQR(B2)+25*SQR(B2)+SQR(B2)+64*SQR(B2*B3)+7968*B3+SQR(B2)-8460*B2+SQR(B2)-20420*B2*B3);
  PARI:=((105+SQR(1000*B2))/AUI);
  PAR2:= (61B3-659+3500*B2)/(15000*B2);
  TST:=T_UNIV(U1,U2);
  A:=SQR(PARI);
  B:=TST/A;
  TRANSFORMA:=B*PAR2;
END;

```

```

FUNCTION NORMAL_MULT(I(T:REAL):REAL;

```

(FUNCION QUE OBTIENE LA PRIMERA ENTRADA DE UN VECTOR ALEATORIO)
(EN R^2 CON DISTRIB. NORMAL MULT CON VECTOR DE MEDIAS O Y CON)
(MATRIZ DE COVARIANZAS, LA MATRIZ DE PREC. DE LA DIST. T MULTI)
{VAR(ADA)

```
VAR
  A,X1,M,AUX1,AUX2,AUX3,AUX4,Y1,Y2,U1,U2:REAL;
BEGIN
  U1:=RANDOM;
  U2:=RANDOM;
  NORMAL(U1,U2,Y1,Y2);
  AUX1:=SQRT(100*(1+SQRT(T)))*(64+T*(64*T-72));
  A:=ABS(60154624+127012928*SQRT(T)*SQRT(T)-316680192*T*SQRT(T)+382667824*SQRT(T)-24143456*T);
  AUX2:=SQRT(A);
  AUX3:=AUX1/AUX2;
  X1:=AUX3*Y2;
  NORMAL_MULT1:=X1;
END;
```

FUNCTION NORMAL_MULT2(T:REAL):REAL;
(FUNCION SIMILAR A LA ANTERIOR; OBTENIENDO LA SEGUNDA ENTRADA)
{DEL VECTOR ALEAT. EN R^2)

```
VAR
  A,B,C,D,Y1,Y2,X2,E,U3,U4,AUX1,AUX2,AUX3:REAL;
BEGIN
  U3:=RANDOM;
  U4:=RANDOM;
  NORMAL(U3,U4,Y1,Y2);
  AUX1:=120*(1+T)*(64+T*(64*T-72));
  AUX2:=ABS(60154624+127012928*SQRT(T)*SQRT(T)-316680192*T*SQRT(T)+382667824*SQRT(T)-24143456*T)+ABS(100*(1+SQRT(T)));
  A:=SQRT(AUX2);
  AUX3:=AUX1/A;
  B:=400*SQRT(64+T*(64*T-72));
  C:=B/AUX2;
  D:=ABS(C-AUX3);
  E:=SQRT(D);
  X2:=Y1*E+AUX3*Y2;
  NORMAL_MULT2:=X2
END;
```

FUNCTION MEDIA_VECT1(X:REAL):REAL;
(ESTA FUNCION OBTIENE LA PRIMERA ENTRADA DEL VECTOR)
(DE LOCALIZACION O MEDIA DE LA T-MULTIVARIADA)

```
VAR
  X1:REAL;
BEGIN
  X1:=(4668+928*(X1))/(64+X1*(64*X1-72));
  MEDIA_VECT1:=X1;
END;
```

```

FUNCION MEDIA_VECT2(X:REAL):REAL;
(ESTA FUNCION OBTIENE LA SEGUNDA ENTRADA DEL VECTOR)
(DE LOCALIZACION DE LA T-MULTIVARIADA)
VAR
  X2:REAL;
BEGIN
  X2:=(-903081X+62561SDR(X)+5134)/(64+X*(648X-721));
  MEDIA_VECT2:=X2;
END;

```

```

FUNCION T_MULT(X,M,I:REAL):REAL;
(ESTA FUNCION OBTIENE UNA MUESTRA DE TAMAÑO I DE)
(UNA ENTRADA DE UN VECTOR ALEATORIO CON DISTRIB.)
(T MULTIVARIADA CON 20 GRADOS DE LIBERTAD, MATRIZ )
(DE PREC. Y VECTOR DE LOCALIZACION OBTENIDOS EN EL)
(DESARROLLO TEORICO DEL PROBLEMA)
VAR
  A,B:REAL;
BEGIN
  A:=X*SDRT(20);
  B:=SDR(2);
  T_MULT:=A/B *M;
END;

```

```

BEGIN (DEL PROGRAMA PRINCIPAL)
  ASSIGN(ARCHI,'A:EIA.PAS');
  REWRITE(ARCHI);
  B[1]:=-2;
  B[2]:=RANDOM;
  FOR M:=1 TO 500 DO
  BEGIN
    FOR K:=1 TO 30 DO
    BEGIN
      RHO1:=TRANSFORMA(B[1],B[2]);
      X[1]:=NORMAL_MULT1(RHO1);
      X[2]:=NORMAL_MULT2(RHO1);
      Z:=J1_CUAD(20);
      MU[1]:=MEDIA_VECT1(RHO1);
      MU[2]:=MEDIA_VECT2(RHO1);
      FOR I:=1 TO 2 DO
        B[I]:=T_MULT(X[I],MU[I],Z);
      IF K=30 THEN
        BEGIN
          WRITELN(ARCHI,RHO1);
          CLRSCR;
          GOTDY(10,10);
          WRITE('LLEVO ', M, ' DATOS');
        END;
      END;
    END;
  END;

```

```
CLOSE(ARCHI);
CLRSCL;
END.
```

```

PROGRAM GIBBS_SAMPLER;
uses crt;
(ESTE PROGRAMA SE BASA EN EL ALGORITMO DE GIBBS)
(PARA OBTENER DOS MUESTRAS PROVENIENTES DE UTILIZAR)
(UN ALGORITMO PARA OBTENER MUESTRAS DE UNA DISTRIBUCION)
(BETA Y DE UNA DISTRIBUCION BINOMIAL)
TYPE
STANDARREGLO=ARRAY[0..21] OF REAL;
VAR
N:INTEGER;
I:INTEGER;
M:INTEGER;
K:INTEGER;
B:INTEGER;
X:INTEGER;
Y:REAL;
U1:REAL;
U2:REAL;
V1:REAL;
V2:REAL;
W:REAL;
A:REAL;
ARCHI:TEXT;

PROCEDURE AUX(VAR O,P:REAL);
(PROCEDIMIENTO QUE ES AUXILIAR PARA EL PROCEDIMIENTO QUE ORDENA UN VECTOR DE)
(NUMEROS REALES)
VAR
Q:REAL;
BEGIN
Q:=O;
O:=P;
P:=Q;
END;

PROCEDURE ORDE(INICIO,ALTO:INTEGER; VAR BET:STANDARREGLO);
(PROCEDIMIENTO BASADO EN EL ALGORITMO DE LA BURBUJA PARA ORDENAR UN ARREGLO)
VAR
INDICE:INTEGER;
BANDERA:BOOLEAN;
BEGIN
REPEAT
BANDERA:=FALSE;
FOR INDICE:=INICIO TO ALTO-1 DO
BEGIN
IF BET[INDICE]>BET[INDICE+1] THEN
BEGIN
AUX(BET[INDICE],BET[INDICE+1]);
BANDERA:=TRUE;
END
END
END

```

```
END;  
UNTIL BANDERA=FALSE  
END;
```

```
FUNCTION BETA(X:INTEGER):REAL;  
{FUNCION QUE OBTIENE UNA MUESTRA DE TAMAÑO 1 DE UNA DISTRIBUCION BETA CON }  
{PARAMETROS X+2 Y 20-X}
```

```
VAR  
L:INTEGER;  
BET:STANDARDREGLO;  
BEGIN  
FOR L:=1 TO 21 DO  
  BET(L):=RANDOM;  
  ORDE(1,21,BET);  
  BETA:=BET[2*X]  
END;
```

```
FUNCTION BINOMIAL(Y:REAL):INTEGER;  
{FUNCION QUE OBTIENE UNA MUESTRA DE TAMAÑO 1 DE UNA DISTRIBUCION BINOMIAL }  
{CON PARAMETROS Y Y M=16}
```

```
BEGIN  
N:=0;  
B:=0;  
WHILE N<16 DO  
  BEGIN  
    A:=RANDOM;  
    IF A<=Y THEN  
      B:=B+1;  
    N:=N+1  
  END;  
BINOMIAL:=B  
END;
```

```
BEGIN  
ASSIGN(ARCHI,'A:GUARDA.Prn');  
REWRITE(ARCHI);  
X:=0;  
Y:=0;  
FOR M:=1 TO 500 DO  
  BEGIN  
    FOR X:=1 TO 10 DO  
      BEGIN  
        IF 1=1 THEN  
          Y:=RANDOM  
        ELSE  
          Y:=BETA(X);  
        Z:=BINOMIAL(Y);  
        IF X=10 THEN  
          BEGIN  
            WRITELN(ARCHI,X,Y);
```

```
CLRSR;  
GOTOXY(10,10);  
WRITE('LLEVO ', N, ' DATOS');  
END;  
END;  
CLOSE (ARCH1);  
READLN;  
CLRSR;  
END.
```

Referencias.

Ahrens, J.H. and Pete, U.(1974). Computer methods for sampling from Gamma, Beta, Poisson and Binomial distributions. *Computing*, 12, 223-246.

Armitage, P.(1977). *Statistical methods in medical research*. Oxford: Blackwell Scientific Publications.

Bartle, R.G.(1980). *Introducción al Análisis Matemático*. 1era. ed. México: Limusa.

Bernardo, J.M(1977). *Inferences about the ratio of normal means: a bayesian approach to the Fieller-Creasy problem. en Recent Developments Statistics (J.R Barra et al. eds.)* New York: North Holland.

Bernardo, J.M(1981). *Bioestadística, una perspectiva bayesiana*. 1era. ed. Barcelona: Vicens - Vives.

Besag, J.(1974). Spatial interaction and statistical analysis of line systems. *Journal of the Royal Statistical Society, series B* 36, 192-236.

Bliss, C.I.(1970). *Statistics in biology*. vol.2. New York: McGraw-Hill.

Casella, G. and George E.I.(1992). Explaining the Gibbs Sampler. Technical Report 96. Statistics Research Center, University of Chicago.

DeGroot, M.H.(1970). *Optimal Statistical Decisions*. New York: McGraw-Hill.

Feller, W.(1975). *Introducción a la teoría de probabilidades y sus aplicaciones*. México: Limusa.

Finney, D.J. (1952/78). *Statistical method in biological assay*. 3rd. ed. London: Griffin.

Gelfand, A.E. and Smith, A.F.M.(1990). Sampling based approaches to calculations marginal desities. *Journal of the American Statistical Association*, 85, 395-409.

Gelfand, A.E; Hills, E.S; Racine-Poon, A. and Smith, A.F.M(1990). Illustration of Bayesian Inference in Normal Data Models using Gibbs Sampling. *Journal of the American Statistical Association*, 85, 972-985.

Geman, S. and Geman, D.(1984). Stochastic relaxation, Gibbs distributions and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721-741.

Hoel, P.G; Port S.C; and Stone C.J (1972). *Introduction to Stochastic Processes*. New York: Houghton Mifflin.

Kapperman, R.F; Geisser, S. and Antle, C.E. (1979). Bayesian and fiducial solutions for de Fieller-Creasy problem. *Sankhya*, 32, 331-340.

Kendall, M. and Stuart, A.(1979). *The advanced Theory of Statistics*. vol 2. 4th ed. London: Griffin.

Kinderman, A.J; Monahan, J.F. and Ramage, J.G.(1977). Computer methods for sampling from student's t distributions. *Mathematics and computation*, 31, 1009-1018.

Kenneth, H. and Ray, K.(1982). *Algebra Lineal*. 2da. ed. México: Prentice Hall.

Kotz J.(1982). *Encyclopedia of Statistical Sciences*. vol.1 y 2. New York: John Wiley and Sons.

Lindley, D.V.(1977). *Teoría de la Decisión*. 1era. ed. Barcelona: Vicens-Vives.

Maindonald, J.H.(1984). *Statistical Computation*. USA: Wiley.

Mendoza, M(1988). *Inferencia bayesiana sobre cocientes de combinaciones lineales en un modelo de regresión múltiple*. Tesis Doctoral. UNAM.

Ripley, B.D.(1987). *Stochastic Simulation*. New York: Wiley.

Tanner; M.A. and Wong, W.(1987). The calculation of posterior distributions by data augmentation (with discution). *Journal of American Statistical Association*. 82, 528-550.

Tanner, M.A.(1991). *Tools for Statistical Inference*. New York: Verlag.

Rubin, D.B.(1987). Comment on The calculation of posterior distributions by data augmentation. by Tanner, M.A. and Wong, W.H. *Journal of the American Statistical Association*, 82, 543-546.