Descripción e implementación del Muestreador de Gibbs en versión bivariada

MANUEL ALEJANDRO MORENO ARÉVALO

UNIVERSIDAD DISTRITAL FRANCISCO JOSE DE CALDAS
FACULTAD DE CIENCIAS Y EDUCACIÓN
PROYECTO CURRICULAR DE MATEMÁTICAS
BOGOTÁ, D.C.
DICIEMBRE DE 2016

Descripción e implementación del Muestreador de Gibbs en versión bivariada

Manuel Alejandro Moreno Arévalo

Disertación presentada para optar al título de Matemático

Director
M. Sc. Luis Alejandro Másmela Caita.

Universidad Distrital Francisco Jose de Caldas
Facultad de Ciencias y Educación
Proyecto Curricular de Matemáticas
Bogotá, D.C.
Diciembre de 2016

Título en español

Descripción e implementación del Muestreador de Gibbs en versión bivariada

Title in English

Description and implementation of Gibbs Sampler in bivariate version

Resumen: Este trabajo de grado presenta un revisión bibliográfica de los fundamentos matemáticos del muestreador de Gibbs y una descripción específica de la mecánica de este. El muestreador de Gibbs es un algoritmo que permite generar una muestra aleatoria a partir de la distribución de probabilidad conjunta de dos o más variables aleatorias haciendo uso de sus distribuciones condicionales.

El trabajo se estructuró en tres capítulos:

- Capítulo 1: Se muestran los fundamentos de la teoría de probabilidad en el caso multivariado, delimitando este estudio a los conceptos empleados en el muestreador de Gibbs.
- Capítulo 2: Presenta la teoría de las Cadenas de Markov a tiempo discreto con espacio de estados discreto. Por ser las cadenas de Markov un soporte teórico para el muestreador de Gibbs.
- Capítulo 3: Expone una descripción del Muestreador de Gibbs en donde se discute su aplicación y su eficacia.

.

Abstract: This degree work presents a literature review of the mathematical foundations of the Gibbs sampler and a specified description of the mechanics of this. The Gibbs sampler is an algorithm that allows to generate a random sample from the joint probability distribution of two or more random variables making use of their conditional distributions.

The work was divided into three chapters:

- In chapter 1 the basics of probability theory in the multivariate case are shown, this study is limited to the concepts used in the Gibbs sampler.
- In chapter 2 it presents the theory of Markov chains with discrete time discrete state space. The Markov chains being theoretical support for the Gibbs sampler
- In chapter 3 expose a description of the Gibbs sampler where its application and effectiveness is discussed.

.

Palabras clave: Distribución condicional, Cadenas de Markov, Sucesión de Gibbs.

Keywords: Conditional distribution, Markov chains, Gibbs sequence.

Nota de aceptación

Trabajo de tesis

Jurado
Gabriel Córdoba Suárez

Director
Luis Alejandro Másmela Caita

Bogotá, D.C., Diciembre 7 de 2016

Dedicado a

Mi mamá Rosalia Arévalo y mi abuelita Clementina Arévalo.

Agradecimientos

Quiero agradecer a Dios por brindarme las capacidades para desarrollar mis estudios y por todas las otras bendiciones que ha derramado sobre mi. A mi mamá Rosalia Arévalo y mi abuelita Clementina Arévalo quienes han sido mi apoyo y guía en las decisiones que he tomado en mi vida. Quiero agradecer también a los profesores del Proyecto Curricular de Matemáticas de la Universidad Distrital por sus lecciones académicas y consejos de vida que han compartido conmigo a lo largo de estos años. Agradezco especialmente al profesor Luis Alejandro Másmela quien dirigió este trabajo de grado porque sin sus conocimientos y paciencia no se hubiese podido culminar satisfactoriamente. Por último, pero no por ello menos importante, al profesor Gabriel Córdoba, por las correcciones y sugerencias hechas a este documento.

Índice general

Índice de tablas

Índice de figuras

Introducción

Algunos cálculos en la teoría de probabilidad requieren de una cierta habilidad matemática. Por ejemplo, si se tiene la función de densidad de probabilidad conjunta de dos variables aleatorias y se está interesado en obtener la función de densidad de una variable aleatoria en particular. Aquí presentamos un método conocido como muestreador de Gibbs que entre sus bondades se encuentra facilitar este tipo de operaciones. Es precisamente en el Capítulo 3 que el lector podrá observar la mecánica simple y el gran alcance que tiene dicho método. Para tal fin en el citado capítulo se encontrará un algoritmo que describe al muestreador de Gibbs y con el objetivo de que se observe la bondad del método, este se ha implementado para dos distribuciones. Se podrá apreciar las casi indistinguibles diferencias que existen entre las distribuciones teóricas encontradas por métodos analíticos y las simuladas mediante el muestreador de Gibbs. Dichas simulaciones fueron realizadas con el software estadístico R.

El muestreador de Gibbs es una técnica novedosa desarrollada por los hermanos Alan Geman y Donald Jay Geman en 1984. Al igual que los demás algoritmos conocidos como MCMC su origen guarda una relación fuerte con la física y su desarrollo a la computación. Como si fuera poco es aplicado en campos como las finanzas, la economía y la biología entre otros.

Se ha estructurado el presente trabajo en tres capítulos.

En el capítulo 1, se exponen conceptos básicos de la teoría de la probabilidad en el caso multivariado, restringiendo el estudio a conceptos que se emplean para desarrollar el muestreador de Gibbs.

En el capítulo 2, se presenta dichos elementos con espacio de estados discreto. Para tal fin se ha seguido muy de cerca el texto [8]. Este capítulo comprende desde la definición de Cadena de Markov a tiempo discreto con espacio de estado discreto hasta cadenas reversibles, pasando por la distribución límite que es de vital importancia para el muestreador de Gibbs.

El capítulo 3, comienza presentando una muy breve revisión histórica del muestreador de Gibbs, expone el algoritmo del muestreador para el caso bivariado, continua con dos ejemplos para los cuales se realizó en cada caso una simulación empleando el muestreador de Gibbs y se contrastanlos resultados con la distribución teórica. Luego se presentan mediante ejemplos la característica markoviana del muestreador. Culminamos este capítulo con la sección llamada evaluación de la matriz P^n

CAPÍTULO 1

Preliminares

El objetivo de este capítulo es presentar algunos conceptos básicos de la teoría de la probabilidad en el caso multivariado. Al igual que como sucede en el caso univariado donde a partir de la definición de variable aleatoria se cimientan un gran número de definiciones, teoremas y/o proposiciones, entre otros, en el caso multivariado la piedra angular es la definición de vector aleatorio $n-$dimensional.

En el campo de las aplicaciones es común estar interesado en observar fenómenos numéricos de dos, tres o más características simultáneas. Así, por analogía con el caso univariado se presenta la definición de función de distribución conjunta acumulada y algunas de sus propiedades. Se brinda también la definición de densidad de probabilidad condicional, que es pieza clave para el muestreador de Gibbs.

1.1. Algunos conceptos de la teoría de la probabilidad en el caso multivariado

Un concepto básico en probabilidad tiene que ver con la definición de vector aleatorio. Este extiende el concepto de variable aleatoria por una función de valores en $\mathbb{R}^n$.

Definición 1. *Sea $X_1, X_2, \cdots X_n$ n-variables aleatorias reales definidas sobre el mismo espacio de probabilidad $(\Omega, \Im, P)$. La función $\boldsymbol{X} : \Omega \to \mathbb{R}^n$ definida por*

$$\boldsymbol{X}(w) := (X_1(w), \cdots, X_n(w))$$

es llamada un $n-$dimensional vector aleatorio.

Se puede asociar a un vector aleatorio n-dimensional su función de distribución como lo hace la siguiente definición. La distribución permite calcular la probabilidad de que el valor del vector aleatorio, pertenezca a un conjunto.

Definición 2. *Sea $\boldsymbol{X}$ un $n-$dimensional vector aleatorio. La medida de probabilidad definida por:*

$$P_{\boldsymbol{X}}(B) := P(\boldsymbol{X} \in B); \quad B \in \mathscr{B}_{\boldsymbol{n}}$$

Es llamada la distribución del vector aleatorio $\boldsymbol{X}$.

Nota 1. *$\mathscr{B}_{\boldsymbol{n}}$ denota la $\sigma-$álgebra de Borel sobre $\mathbb{R}^n$.*

La naturaleza discreta de las variables aleatorias de un vector aleatorio $\boldsymbol{X}$ permite caracterizar a este como un vector discreto. La siguiente definición evidencia esto e incluye lo que se entenderá en adelante por función de distribución conjunta.

Definición 3. *Sea $\boldsymbol{X} = (X_1, X_2, \cdots, X_n)$ un vector aleatorio $n-$dimensional. Si las variables aleatorias X_i, con $i = 1, \cdots, n$ son todas discretas, se dice que el vector aleatorio $\boldsymbol{X}$ es discreto. En este caso la función de masa de probabilidad de $\boldsymbol{X}$, también llamada función de distribución conjunta de las variables aleatorias $X_1, X_2, \cdots, X_n$, esta definida por:*

$$p_{\boldsymbol{X}}(x) := \begin{cases} P(\boldsymbol{X} = x) & \text{si } x \text{ pertenece a la imagen de } \boldsymbol{X} \\ 0 & \text{en cualquier otro caso.} \end{cases}$$

Es importante añadir que si en la definición anterior las variables aleatorias X_i, con $i = 1, \cdots, n$ son todas continuas, el vector aleatorio $\boldsymbol{X}$ se dice continuo. Puede también presentarse el caso en que una variable aleatoria digamos X_j para $1 \leq j \leq n$ pueda ser continua y las demás discretas o viceversa.

Definición 4. *Sea $\boldsymbol{X} = (X_1, X_2, \cdots, X_n)$ un vector aleatorio $n-$dimensional. La función definida por:*

$$F(x_1, x_2, \cdots, x_n) := P(X_1 \leq x_1, X_2 \leq x_2, \cdots, X_n \leq x_n),$$

para toda $(x_1, x_2, \cdots, x_n) \in \mathbb{R}^n$ es llamada la función de distribución conjunta acumulada de las variables aleatorias $X_1, X_2, \cdots, X_n$, o simplemente la función de distribución del vector aleatorio $n-$dimensional X.

Existe un procedimiento para encontrar la función de distribución marginal acumulada de una variable aleatoria X_j para $1 \leq j \leq n$ de un vector aleatorio $n-$dimensional X, dado este y su función de distribución conjunta acumulada digamos F. Tal procedimiento es el que establece el siguiente teorema.

Teorema 1. *Sea $\boldsymbol{X} = (X_1, X_2, \cdots, X_n)$ un vector aleatorio $n-$dimensional con función de distribución conjunta acumulada F. Para cada $j = 1, \cdots, n$, la función de distribución acumulada de la variable aleatoria X_j está dada por*

$$F_{X_j}(x) = \lim_{x_1 \to \infty} \cdots \lim_{x_{j-1} \to \infty} \lim_{x_{j+1} \to \infty} \cdots \lim_{x_n \to \infty} F(x_1, \cdots x_n).$$

La función de distribución F_{X_j} es llamada la funcion de distribución marginal acumulada de la variable aleatoria X_j.

Dada la importancia de la función de distribución de vectores aleatorios, es de interés presentar características de los mismos. Los Teoremas 2 y 3 ilustran dichas características. Por ejemplo, cómo identificar si una función es en efecto función de distribución conjunta acumulada en el caso bivariado. La demostración de estos resultados puede consultarse en [4].

Teorema 2. *Sea $\boldsymbol{X} = (X, Y)$ un vector aleatorio $2-$dimensional. La función de distribución conjunta acumulada F de las variables aleatorias X, Y tiene las siguientes propiedades:*

1.
$$\Delta_a^b F := F(b_1, b_2) + F(a_1, a_2) - F(a_1, b_2) - F(b_1, a_2) \geq 0$$
donde $a = (a_1, a_2), b = (b_1, b_2) \in \mathbb{R}^2$ con $a_1 \leq b_1$ y $a_2 \leq b_2$.

2.
$$\lim_{x \searrow x_0} F(x, y) = F(x_0, y) \tag{1.1}$$

3.
$$\lim_{y \searrow y_0} F(x, y) = F(x, y_0) \tag{1.2}$$

4.
$$\lim_{x \to -\infty} F(x, y) = 0$$
y
$$\lim_{y \to -\infty} F(x, y) = 0$$

5.
$$\lim_{(x,y) \to (\infty, \infty)} F(x, y) = 1$$

Teorema 3. *Sea $\boldsymbol{X} = (X_1, X_2, \cdots, X_n)$ un vector aleatorio $n-$dimensional. La función de distribución conjunta acumulada F de las variables aleatorias $X_1, X_2, \cdots, X_n$ tiene las siguientes propiedades:*

1. *$\Delta_a^b F \geq 0$ para todo $a = (a_1, \cdots, a_n)$, $b = (b_1, \cdots, b_n) \in \mathbb{R}^n$ con $a \leq b$, donde:*
$$\Delta_a^b F := \sum_{(\epsilon_1, \cdots \epsilon_n) \in \{0,1\}^n} (-1)^{\left(\sum_{i=1}^n \epsilon_i\right)} F(\epsilon_1 a_1 + (1 - \epsilon_1) b_1, \cdots, \epsilon_n a_n + (1 - \epsilon_n) b_n).$$

2. *F es continua a derecha en cada componente.*

3. *Para todo $a_1, \cdots, a_{i-1}, a_{i+1}, \cdots, a_n \in \mathbb{R}$ con $i = 1, \cdots, n$, tenemos:*
$$\lim_{x \searrow -\infty} F\left(a_1, \cdots, a_{i-1}, \underbrace{x}_{\text{i-ésima posición}}, a_{i+1}, \cdots, a_n\right) = 0$$

4.

$$\lim_{(x_1,x_2,\cdots x_n)\to(\infty,\infty,\cdots,\infty)} F\left(x_1,x_2,\cdots x_n\right)=1$$

Definición 5. *Sean $X_1, X_2, \cdots, X_n$ n variables aleatorias de valor real definidas sobre el mismo espacio de probabilidad. Se dice que las variables aleatorias son conjuntamente continuas, si existe una función integrable $f : \mathbb{R}^n \to [0, +\infty)$ tal que para todo conjunto de Borel C de $\mathbb{R}^n$*

$$P\left(\left(X_1,X_2,\cdots,X_n\right)\in C\right)=\int\underset{C}{\cdots}\int f\left(x_1,x_2,\cdots,x_n\right)dx_1dx_2\cdots dx_n.$$

La función f es llamada la función de densidad de probabilidad conjunta de las variables aleatorias $X_1, X_2, \cdots X_n$.

Nota 2. *Es costumbre notar fdp a la función de densidad de probabilidad.*

Los teoremas siguientes exhiben dos relaciones notables el Teorema 4 establece cómo dado un vector aleatorio 2−dimensional y su función de distribución conjunta se puede encontrar la función de densidad de probabilidad conjunta en las regiones en que esta sea continua. La demostración de este hecho se puede encontrar en [4].

Teorema 4. *Sean X y Y variables aleatorias continuas con funcion de distribución conjunta F. Entonces, la función de densidad de probabilidad conjunta f es:*

$$f\left(x,y\right)=\frac{\partial^2F\left(x,y\right)}{\partial x\partial y}=\frac{\partial^2F\left(x,y\right)}{\partial y\partial x}$$

para todos los puntos (x, y) donde $f(x, y)$ es continua.

Algunas veces dado un conjunto de variables aleatorias, se está interesado en conocer la función de densidad de una variable en particular. El teorema siguiente establece los elementos que se deben conocer para llevar a buen fin esta tarea.

Teorema 5. *Si $X_1, X_2, \cdots, X_n$ son n variables aleatorias de valor real, teniendo f (función de densidad de probabilidad conjunta), entonces*

$$f_{X_j}\left(x\right)=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}f\left(x_1,\cdots,x_{j-1},x,x_{j+1},\cdots,x_n\right)dx_1\cdots dx_{j-1}dx_{j+1}\cdots dx_n$$

es la función de densidad de la variable aleatoria X_j para $j = 1, 2, \cdots, n$.

La Definición 6 y 8 son indispensables para formar la sucesión de Gibbs.

Definición 6. *Sean X y Y dos variables aleatorias discretas. La función de probabilidad condicional de X dado $Y = y$ se define como*

$$p_{X|Y}\left(x|y\right):=P\left(X=x|Y=y\right)=\frac{P\left(X=x,Y=y\right)}{P\left(Y=y\right)}.$$

Para todo y para el cual $P(Y=y)>0$.
Definición 7. *La función de distribución condicional de X dado $Y=y$ se define como:*

$$F_{X|Y}(x|y) := P(X\leq x|Y=y) = \sum_{k\leq x} p_{X|Y}(k|y).$$

Para todo y para el cual $P(Y=y)>0$.

La Definición 8 caracteriza la relación entre la función de densidad de una variable aleatoria y la función de densidad de probabilidad conjunta, llamaremos a esta la función de probabilidad condicional. En la Definición 6 se presentó para el caso discreto.

Definición 8. *Sean X y Y variables aleatorias continuas con función de densidad de probabilidad conjunta f. La función de densidad de probabilidad condicional de X dado $Y=y$ se define como*

$$f_{X|Y}(x|y) := \frac{f(x,y)}{f_Y(y)}$$

para todo y con $f_Y(y)>0$.

Definición 9. *La función de distribución condicional de X dado $Y=y$ se define como*

$$F_{X|Y}(x|y) := P(X\leq x|Y=y) = \int_{-\infty}^{x} f_{X|Y}(t|y)\,dt$$

Para todo y con $f_Y(y)>0$.

1.2. La prueba Ji-cuadrado

Uno de los objetivos del presente trabajo es exponer al muestreador de Gibbs como un método optimo para estimar la distribución marginal de una variable aleatoria. Es así que con los valores obtenidos mediante la sucesión de Gibbs, se quiere presentar que su comportamiento se asemeja a la distribución teórica. En [5] utilizan las gráficas para mostrar esto, nosotros además lo presentamos por medio de la prueba ji-cuadrado implementada en R.

El test ji-cuadrado es un estadístico de prueba formulado en 1990 por Karl Pearson. En este supone que desconocemos la distribución de probabilidad y la hipótesis nula es que dicha distribución es alguna conocida. Para determinar si la hipótesis se acepta o rechaza se emplea una muestra. El estadístico de prueba es

$$X^2 = \sum_{i=1}^{k} \frac{[n_i - np_i]^2}{np_i}$$

Donde n es el tamaño de la muestra y np_i son lo valores medios teóricos. Un hecho que es conocido en estadística pero no es presentado aquí, pues su tratamiento escapa de los objetivos del trabajo es que la distribución del estadístico X^2 es un distribución χ^2 con $k-1$ grados de libertad, cuando la hipótesis nula es cierta.

CAPÍTULO 2

Cadenas de Markov

Debido a la naturaleza Markoviana de la iteración del muestreador de Gibbs se hace indispensable estudiar las cadenas de Markov que es el propósito del presente capítulo, para hacerlo se ha decidido seguir muy de cerca el texto [8].

Las cadenas Markov datan su aparición entre 1905 a 1907 y se deben al matemático ruso Andrey Markov considerado por esto como precursor de los procesos estocásticos. El estudio de cadenas de Markov se extendió ampliamente en las primeras décadas del siglo XX debido en gran parte a que muchos fenómenos físicos, financieros y sociales encontraban su representación en este tipo de modelo.

Definición 10. *Una cadena de Markov es un proceso estocástico a tiempo discreto* $\{X_n : n = 0, 1, \ldots\}$, *con espacio de estados discreto, y que satisface la propiedad de Markov, esto es, para cualquier entero* $n \geq 0$, *y para cualesquiera estados* $x_0, \ldots, x_{n+1}$, *se cumple*

$$p\left(x_{n+1} | x_0, \ldots, x_n\right) = p\left(x_{n+1} | x_n\right) \tag{2.1}$$

La propiedad de Markov establece que la probabilidad condicional de un evento futuro $X_{n+1} = k$ dado el estado presente $X_n = i$ es independiente de cualquier evento en el pasado. Por otro lado dentro de la teoría de las cadenas de Markov un concepto indispensable es la llamada probabilidad de transición siendo esta $P\left(X_n = k | X_{n-1} = i\right)$ la probabilidad de que el proceso pase del estado i al estado k en un solo paso o unidad de tiempo, se suele denotar por p_{ik}. Así mismo se dice que una cadena de Markov es homogénea o cadena de Markov con probabilidad estacionaria si la probabilidad de transición es independiente del tiempo en el que se encuentre la cadena.

Definición 11. *Sea*

$$P = (p_{ij}) = \begin{pmatrix} p_{00} & p_{01} & p_{02} & p_{03} \cdots \\ p_{10} & p_{11} & p_{12} & p_{13} \cdots \\ p_{20} & p_{21} & p_{22} & p_{23} \cdots \\ p_{30} & p_{31} & p_{32} & p_{33} \cdots \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix} \tag{2.2}$$

La matriz para la cual sus elementos son las probabilidades de transición entre los estados del proceso.

La matriz que se definió se conoce como matriz estocástica o matriz de probabilidad de transición. La Proposición 1 señala dos propiedades de esta matriz y su demostracion puede ser consultada en [8].

Proposición 1. *La matriz de probabilidad de transición* $P = (p_{ij})$ *satisface*

1. $p_{ij} \geq 0$
2. $$\sum_{j=0}^{\infty} p_{ij} = 1 \quad i = 0, 1, 2, 3, \ldots$$

Definición 12. *Una distribución inicial para una cadena de Markov con espacio de estados* $\{0, 1, 2, 3, \ldots\}$ *es una distribución de probabilidad sobre este conjunto.*

La Proposición 2 recibe su nombre gracias a dos matemáticos de nacionalidades distintas pero apasionados por los procesos estocásticos, esto les permitió de manera independiente llegar a su formulación; los matemáticos aquí relacionados son Sydney Chapman y Andréi Kolmogórov de nacionalidades inglesa y rusa respectivamente.

Proposición 2. ***(Ecuación de Chapman-Kolmogorov)*** *Para cualesquiera enteros* m, n *tales que* $0 \leq m \leq n$*, y cualesquiera estados* i *y* j *se tiene que*

$$p_{ij}(n) = \sum_{k} p_{ik}(m) p_{kj}(n-m). \tag{2.3}$$

La demostración de la Proposición 2 requiere del teorema de probabilidad total y de la propiedad de Markov. La Proposición 3 es un resultado de la Ecuación de Chapman-Kolmogorov y ambas demostraciones pueden ser consultadas en [8].

Proposición 3. *La probabilidad de transición en* n *pasos.* $p_{ij}(n)$*, está dada por la entrada* (i, j) *de la* n*-ésima potencia de la matriz* P*, es decir,*

$$p_{ij}(n) = (P^n)_{ij}.$$

2.1. Comunicación

La mayoría de textos de procesos estocásticos, afirman que la comunicación es una relación de equivalencia, la demostración de este hecho se basa en la definición de $p_{ij}(0)$ como la delta Kronecker, es decir,

$$p_{ij}(0) = \delta_{ij} = \begin{cases} 0 & \text{si } i \neq j \\ 1 & \text{si } i = j \end{cases}$$

y la ecuación de Chapman Kolmogorov. Así pues dada la relación de estados que se comunican es posible descomponer las cadenas de Markov por los subconjuntos del espacio de estados que están comunicados.

Definición 13. *Se dice que el estado j es accesible desde el estado i si existe un entero $n \geq 0$ tal que $p_{ij}(n) > 0$, esto se escribe simplemente como $i \to j$. Se dice además que los estados i y j son comunicantes, y se escribe $i \leftrightarrow j$, si se cumple que $i \to j$ y $j \to i$.*

Se puede decir que una Cadena de Markov se dice irreducible cuando la relación de comunicación genera una única clase de equivalencia o como se presenta en la Definición 14, ambas definiciones son equivalentes.

Definición 14. *Se dice que una cadena de Markov es irreducible si todos los estados se comunican entre sí.*

2.2. Periodo

Definición 15. *El periodo de un estado i es un número entero no negativo denotado por $d(i)$, y definido como sigue:*

$$d(i) = m.c.d. \{n \geq 1 : p_{ii}(n) > 0\},$$

en donde m.c.d. significa máximo común divisor. Cuando $p_{ii}(n) = 0$ para toda $n \geq 1$, se define $d(i) = 0$. En particular, se dice que un estado i es aperiódico si $d(i) = 1$. Cuando $d(i) = k \geq 2$ se dice que i es periódico de periodo k.

Una cadena de Markov en la cual todos sus estados son aperiódicos es llamada una cadena de Markov aperiódica. El siguiente conjunto de Proposiciones 4, 5 y 6 son algunas propiedades del periodo. La Proposición 4 se puede interpretar de diversas maneras, por ejemplo, se puede decir que si dos estados están comunicados entonces tendrán el mismo periodo lo que sugiere que el periodo es una propiedad de clase.

Proposición 4. *Si los estados i y j pertenecen a la misma clase de comunicación, entonces tienen el mismo periodo.*

La Proposición 5 afirma que para cada estado es posible retornar a este con múltiplos lo suficientemente grandes del periodo $d(i)$.

Proposición 5. *Para cada estado i, existe un entero N tal que para toda $n \geq N$, se cumple $p_{ii}(nd(i)) > 0$.*

La Proposición 6 es un resultado de las Proposiciones 5 y 2.

Proposición 6. *Si $p_{ij}(m) > 0$ para algún entero m, entonces existe un entero N tal que para toda $n \geq N$ se cumple $p_{ij}(m + nd(j)) > 0$.*

2.3. Primeras visitas

Esta sección trata del tiempo de primera visita de un subconjunto del espacio de estados de una cadena de Markov. Existen diversas aplicaciones, por ejemplo, el tiempo de mantenimiento de cierta máquina, es decir, dado que la máquina funciona correctamente en qué periodo n de tiempo la máquina presentará un daño y necesitará mantenimiento. Esta aplicación sugiere que se puede relacionar la Definición 16 con la Definición 17. Es decir, dado de que exista una primera visita al subconjunto del espacio de estados $A :=$\{*la máquina presenta avería*\}, con un único estado al que se puede llamar k y suponiendo que la cadena inicia en el estado $j :=$ *la máquina funciona correctamente*. Entonces según la Definición 17 $f_{ij}(n)$ representa la probabilidad de que τ_{ij} tome el valor n. La Proposición 7 es una formula bastante útil y la demostración requiere de la ecuación 2.1.

Definición 16. *Sea A un subconjunto del espacio de estados de una cadena de Markov $\{X_n : n \geq 0\}$. El tiempo de primera visita al conjunto A es la variable aleatoria*

$$\tau_A = \begin{cases} \text{mín}\{n \geq 1 : X_n \in A\} & \textit{si } X_n \in A \textit{ para algún } n \geq 1, \\ \infty & \textit{otro caso.} \end{cases}$$

Definición 17. *Para cada $n \geq 1$, el número $f_{ij}(n)$ denota la probabilidad de que una cadena que inicia en el estado i, llegue al estado j por primera vez en exactamente n pasos, es decir,*

$$f_{ij}(n) = P(X_n = j, X_{n-1} \neq j, \ldots, X_1 \neq j | X_0 = i).$$

Adicionalmente se define $f_{ij}(0) = 0$, incluyendo el caso $i = j$.

Proposición 7. *Para cada $n \geq 1$,*

$$p_{ij}(n) = \sum_{k=1}^{n} f_{ij}(k)\, p_{ij}(n-k).$$

2.4. Recurrencia y transitoriedad

En las secciones anteriores se han tratado las probabilidades de transición de un estado a otro. El objetivo de esta sección es presentar una clasificación de los estados de una cadena de Markov teniendo en cuenta su eventual probabilidad de regresar al estado dado que se partió del mismo estado.

Definición 18. *Se dice que un estado i es recurrente si la probabilidad de eventualmente regresar a i, partiendo de i, es uno, es decir, si*

$$P\left(X_n = i \quad \textit{para alguna} \quad n \geq 1 | X_0 = i\right) = 1$$

Un estado que no es recurrente se llama transitorio, y en tal caso la probabilidad anterior es estrictamente menor a uno.

Definición 19. *Un estado i es recurrente si $f_{ii} = 1$, es decir, si la probabilidad de regresar a él en un tiempo finito es uno. Análogamente, un estado i es transitorio si $f_{ii} < 1$.*

Proposición 8. *El estado i es:*

1. *recurrente si, y sólo si $\sum_{n=1}^{\infty} p_{ii}(n) = \infty$.*
2. *transitorio si, y sólo si $\sum_{n=1}^{\infty} p_{ii}(n) < \infty$.*

La Proposición 8 es empleada en algunos textos de procesos estocásticos como la definición de estado recurrente y estado transitorio.

Proposición 9. *La recurrencia es una propiedad de clase, es decir,*

1. *Si i es recurrente e $i \leftrightarrow j$, entonces j es recurrente.*
2. *Si i es transitorio e $i \leftrightarrow j$, entonces j es transitorio.*

Proposición 10. *Sea j un estado transitorio. Para cualquier estado inicial i, se cumple que $\sum_{n=1}^{\infty} p_{ij}(n) < \infty$. En consecuencia, $\lim_{n\to\infty} p_{ij}(n) = 0$.*

¿Se puede decir que todos los estados de una cadena de Markov finita son recurrentes?, de la Proposición 11 se puede extraer respuesta.

Proposición 11. *Toda cadena de Markov finita tiene por lo menos un estado recurrente.*

Las pruebas de las Proposiciones 8, 9, 10 y 11, pueden ser consultadas en [8].

2.5. Tiempo medio de recurrencia

Definición 20. *El tiempo medio de recurrencia de un estado recurrente j, a partir del estado i, se define como la esperanza de τ_{ij}, y se denota por μ_{ij}, es decir,*

$$\mu_{ij} = E\left(\tau_{ij}\right) = \sum_{n=1}^{\infty} n f_{ij}\left(n\right)$$

Notación 1. *Se denota τ_j al tiempo de primera visita cuando el estado de inicio y de llegada coinciden, para este caso μ_j denota el tiempo medio de recurrencia y representa el promedio del número de pasos que la cadena se toma en regresar al estado j que es recurrente.*

2.6. Clases cerradas

En esta sección se considera brevemente el concepto de clase cerrada como una colección de estados no vacía tales que para cualquier estado i que pertenezca a esta colección y cualquier j que no pertenezca se tiene que $p_{ij}(n) = 0$ para todo $n \in \mathbb{N}$.

Definición 21. *Una colección de estados no vacía $\mathscr{C}$ es cerrada si ningún estado fuera de $\mathscr{C}$ es accesible desde algún estado dentro de $\mathscr{C}$, es decir, si para cualquier $i \in \mathscr{C}$ y $j \notin \mathscr{C}, i \nrightarrow j$.*

La Proposición 12 establece una forma de caracterizar una clase de comunicación vía colección de estados que es cerrado e irreducible.

Proposición 12. *Toda colección de estados que es cerrada e irreducible es una clase de comunicación.*

2.7. Número de visitas

Esta primera parte de la sección concierne a definir una nueva variable aleatoria que representa el número de visitas que una cadena de Markov hace sobre un estado particular para un tiempo finito n. La Proposición 13 presenta algunas propiedades para esta variable aleatoria su demostración puede consultarse en [8].

Definición 22. *La variable aleatoria que registra el número de visitas que una cadena realiza sobre un estado j a partir del estado i, para cualquier tiempo finito n se define por:*

$$N_{ij}\left(n\right) = \sum_{k=1}^{n} 1_{X_k=j}.$$

Cuando $X_0 = i$.

Notación 2. *Se acostumbra a denotar* $N_i(n)$ *cuando los estados* j *e* i *coinciden.*

Proposición 13. *Para cualesquiera estados* i *y* j,

1. $$P\left(N_{ij} \geq k\right) = \begin{cases} 1 & si\ k = 0, \\ f_{ij}\left(f_{jj}\right)^{k-1} & si\ k \geq 1. \end{cases}$$

2. $$P\left(N_{ij} = k\right) = \begin{cases} 1 - f_{ij} & si\ k = 0, \\ f_{ij}\left(f_{jj}\right)^{k-1}\left(1 - f_{jj}\right) & si\ k \geq 1. \end{cases}$$

3. $$E\left(N_{ij}\right) = \sum_{n=1}^{\infty} p_{ij}\left(n\right) = \begin{cases} 0 & si\ f_{ij} = 0, \\ \frac{f_{ij}}{1-f_{jj}} & si\ 0 \leq f_{jj} < 1, \\ \infty & si\ f_{ij} \neq 0 \ \ y \ \ f_{jj} = 1. \end{cases}$$

4. $$P\left(N_{ij} = \infty\right) = \begin{cases} 0 & si\ j\ \ es\ transitorio, \\ f_{ij} & si\ j\ \ es\ recurrente. \end{cases}$$

5. $$P\left(N_{ij} < \infty\right) = \begin{cases} 1 & si\ j\ \ es\ transitorio, \\ 1 - f_{ij} & si\ j\ \ es\ recurrente. \end{cases}$$

La siguiente definición permite extender la propiedad de Markov a la Proposición 14.

Definición 23. *Una variable aleatoria* $\tau : \Omega \to \{0, 1, \ldots \infty\}$ *es un tiempo de paro respecto del proceso estocástico indicado si para cada entero* $n \geq 0$, *el evento* $(\tau = n)$ *depende únicamente de* $X_0, X_1, \ldots, X_n$.

La propiedad fuerte de Markov establece que una cadena de Markov es invariante si es observada a partir de un tiempo aleatorio τ. La demostración de la Proposición 14 se divide en dos partes verificar que las probabilidades de transición siguen siendo estacionarias y que se satisface la ecuación 2.4.

Proposición 14. ***(Propiedad fuerte de Markov)*** *Sea* $\{X_n : n \geq 0\}$ *una cadena de Markov y sea* τ *un tiempo de paro respecto de este proceso. Condicionado al evento* $(\tau < \infty)$, *el proceso* $\{X_{\tau+n} : n \geq 0\}$ *es una cadena de Markov, es decir, la probabilidad*

$$P\left(X_{\tau+n+1} = j | X_0 = x_0, \ldots, X_{\tau+n-1} = x_{n-1}, X_{\tau+n} = i\right) \tag{2.4}$$

Es igual a $P\left(X_{\tau+n+1} = j | X_{\tau+n} = i\right)$

Para finalizar esta sección se presenta el teorema ergódico para cadenas de Markov junto con un corolario. *El teorema ergódico proporciona el comportamiento límite del promedio temporal del número de vistas a un estado de una cadena de Markov*, tanto el teorema ergódico como su corolario hacen parte de los teoremas límites para cadenas de Markov. La demostración de estos resultados puede ser consultada en [8].

Teorema 6. ***(Teorema ergódico para cadenas de Markov)*** *Para cualesquiera estados i y j de una cadena de Markov irreducible se cumple que*

$$\lim_{n\to\infty} \frac{N_{ij}(n)}{n} = \frac{1}{\mu_j} \quad c.s. \tag{2.5}$$

siendo este límite cero cuando $\mu_j = \infty$.

Notación 3. *c.s.:= casi seguramente.*

Corolario 1. *Si una cadena es irreducible y recurrente, entonces*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} P_{ij}(k) = \frac{1}{\mu_j}$$

2.8. Recurrencia positiva y nula

En la Sección 2.4 se presentó una primera división para toda cadena de Markov por su espacio de estados en dos subconjuntos disjuntos a saber recurrentes y transitorios. El objetivo de esta sección es realizar una nueva división del espacio de estados empleando la Definición 20 y responder la pregunta ¿Existen estados recurrentes nulos en Cadenas de Markov finitas? Para lo cual es conveniente ver la Proposición 16.

Definición 24. *Se dice que un estado recurrente i es:*

1. *recurrente positivo si $\mu_i < \infty$.*
2. *recurrente nulo si $\mu_i = \infty$.*

Proposición 15. *Sea i un estado recurrente. Entonces,*

1. *si i es recurrente positivo e $i \leftrightarrow j$, entonces j es recurrente positivo.*
2. *si i es recurrente nulo e $i \leftrightarrow j$, entonces j es recurrente nulo.*

Proposición 16. *No existen estados recurrentes nulos en cadenas de Markov finitas.*

2.9. Evolución de distribuciones

Esta sección es una adaptación de [8] de la sección del mismo nombre. El propósito es describir cómo se puede obtener una sucesión infinita de distribuciones de probabilidad teniendo como partida *un espacio de estados finito* $\{0, 1, \ldots, n\}$ *, una distribución inicial* $\pi^0 = (\pi_0^0, \pi_1^0, \ldots, \pi_n^0)$ *. y una matriz de probabilidad de transición en un paso. La cadena se encuentre en alguno de sus posibles estados de acuerdo a la distribución* $\pi^1 = (\pi_0^1, \pi_1^1, \ldots, \pi_n^1)$ *, una vez pasada la primera unidad de tiempo, para esta distribución la* $j-$*ésima entrada de este vector es*

$$\begin{aligned}\pi_j^1 &= P(X_1 = j) \\ &= \sum_{i=0}^{n} P(X_1 = j | X_0 = i) P(X_0 = i) \\ &= \sum_{i=0}^{n} \pi_i^0 p_{ij}\end{aligned}$$

Esto evidencia que mediante la multiplicación $\pi^1 = \pi^0 P$*, se puede obtener el vector* π^1 *vía el vector* π^0 *y la matriz de probabilidades de transición* P*, es decir*

$$\left(\pi_0^1, \pi_1^1, \ldots, \pi_n^1\right) = \left(\pi_0^0, \pi_1^0, \ldots, \pi_n^0\right) \begin{pmatrix} p_{00} & \cdots & p_{0n} \\ \vdots & & \vdots \\ p_{n0} & \cdots & p_{nn} \end{pmatrix}$$

De manera análoga se puede obtener el vector π^2 *mediante* $\pi^2 = \pi^1 P = \pi^0 P^2$*, si se continuara con este procedimiento, se llegaría que para* $m \geq 1$*,*

$$\pi^m = \pi^{m-1} P = \pi^0 P^m \tag{2.6}$$

Así se obtendría sucesión infinita de distribuciones de probabilidad $\pi^0, \pi^1, \pi^2, \ldots,$ *en donde a excepción de la primera, todas se obtuvieron de su inmediata anterior multiplicando a derecha por la matriz de probabilidades de transición en un paso.*

2.10. Distribuciones estacionarias

Definición 25. *Una distribución de probabilidad* $\pi = (\pi_0, \pi_1, \ldots)$ *es estacionaria o invariante para una cadena de Markov con matriz de probabilidades de transición* $P = (p_{ij})$ *si*

$$\pi_j = \sum_i \pi_i p_{ij}.$$

Al igual que sucede en álgebra lineal donde un sistema de ecuaciones lineales puede tener infinitas soluciones o única solución o no existir solución una distribución estacionaria para una cadena de Markov puede también presentarse de manera única, múltiple o no existir. En libros de procesos estocásticos pueden encontrarse un buen número de ejemplos.

Por lo dicho anteriormente es óptimo enunciar una definición alterna de distribución estacionaria a saber: *una distribución de probabilidad* $\pi = (\pi_0, \pi_1, \ldots)$ *es estacionaria o invariante para una cadena de Markov con matriz de probabilidades de transición* $P = (p_{ij})$ si,

$$\pi = \pi P. \tag{2.7}$$

Producto de esta definición alterna se facilita obtener algunos resultados, por ejemplo, si se tuviera una matriz de transición P para una cadena de Markov y se quisiera conseguir una probable distribución estacionaria se podría resolver el sistema de ecuaciones $\pi = \pi P$ con la condición $\sum_j \pi_j = 1$. una interesante propiedad matemática de las distribuciones estacionaria es que forma un conjunto convexo.

Se conoce como distribución estacionaria debido a la propiedad de ser invariante por el paso del tiempo, esto es una consecuencia de la ecuación 2.7 y se enuncia como: *una distribución* π *estacionaria satisface que para todo* $m \in \mathbb{N}$ *se tiene que* $\pi = \pi P^m$.

Es preciso destacar que aunque el vector nulo o vector de ceros cumple la ecuación 2.7 este no es una distribución de probabilidad.

Proposición 17. ***(Soporte de una distribución estacionaria)*** *Sea* π *una distribución estacionaria para una cadena de Markov. Si* j *es un estado transitorio o recurrente nulo, entonces* $\pi_j = 0$.

La demostración de las Proposiciones 17 y 18 pueden ser consultadas en [8]. La importancia de la Proposición 18 es que concibe una fórmula de la distribución estacionaria para una cadena de Markov irreducible y recurrente positiva, tomando el inverso del tiempo medio de recurrencia para el correspondiente estado. Como también establece que toda cadena finita e irreducible tiene una única distribución estacionaria.

Proposición 18. ***(Existencia y unicidad de la distribución estacionaria)*** *Toda cadena de Markov que es irreducible y recurrente positiva tiene una única distribución estacionaria dada por*

$$\pi_j = \frac{1}{\mu_j} > 0,$$

en donde μ_j *es el tiempo medio de recurrencia del estado* j. *En particular, toda cadena finita e irreducible tiene una única distribución estacionaria.*

2.11. Distribuciones límite

Definición 26. *Considere una cadena de Markov con matriz de probabiliades de transición $P = (p_{ij})$ y distribución inicial π^0. Se le llama distribución límite de esta cadena al vector*

$$\pi = \lim_{n\to\infty} \pi^0 P^n = \lim_{n\to\infty} \sum_i \pi_i^0 p_{ij}(n)$$

La definición de distribución límite para una cadena de Markov establece que esta se determina por el límite de las potencia de P y por tanto no depende de la distribución inicial. Además es una distribución estacionaria.

No siempre una distribución estacionaria es una distribución límite, es natural entonces preguntarse qué condiciones debe satisfacer una distribución límite para ser estacionaria, para responder esta pregunta se presenta la Proposición 19 que es válida tanto para espacios finitos como infinitos.

Obsérvese también que la Proposición 19 no hace la salvedad en que los límites puedan ser todos nulos pero finaliza recalcando que se obtiene una distribución de probabilidad verdadera cuando el espacio de estados es finito. La importancia de este argumento es que existen vectores que satisfacen la Definición 26 pero no son distribuciones de probabilidad.

Proposición 19. *Considere una cadena de Markov con probabilidades de transición p_{ij} tales que los límites $\pi_j = \lim_{n\to\infty} p_{ij}(n)$ existen para cada j, y no dependen del estado i. Entonces*

1. $$\sum_j \pi_j \leq 1.$$

2. $$\pi_j = \sum_i \pi_i p_{ij}.$$

Cuando el espacio de estados es finito se cumple la igualdad en el primer resultado obteniéndose una distribución de probabilidad verdadera.

La Proposición 20 amplia el conocimiento de la distribución límite pues enuncia qué condiciones se deben tener para la existencia de los límites de probabilidad. Suele pensarse debido a la condición 3 que la Proposición 20 aparenta ser un reciproco de la Proposición 19.

Proposición 20. ***(Convergencia a la distribución estacionaria)*** *Considere una cadena de Markov que es:*

1. *irreducible,*
2. *aperiódica, y*

3. con distribución estacionaria π

entonces para cualesquiera estados i *y* j, $\lim_{n\to\infty} p_{ij}(n) = \pi_j$.

La Proposición 21 juega un papel importante dentro de las cadenas de Markov, pues establece las condiciones para la existencia de probabilidades limite y suministra además una forma explícita para hallarlas. Por otro lado es un argumento que relaciona una buena cantidad de elementos matemáticos de las cadenas de Markov y fruto de esto su demostración es bastante corta. La prueba de las proposiciones presentadas en esta sección puede consultarse en [8].

Proposición 21. ***(Convergencia para cadenas de Markov)*** *Considere una cadena de Markov que es:*

1. *irreducible,*
2. *recurrente positiva, y*
3. *aperiódica.*

Entonces las probabilidades límite $\pi_j = \lim_{n\to\infty} p_{ij}(n)$ *existen, están dadas por* $\pi_j = 1/\mu_j$, *y constituyen la única solución al sistema de ecuaciones*

$$\pi_j = \sum_i \pi_i p_{ij} \tag{2.8}$$

sujeto a las condiciones $\pi_j \geq 0$, *y* $\sum_j \pi_j = 1$.

2.12. Cadenas regulares

Existen cadenas de Markov para las cuales alguna potencia n de la correspondiente matriz de transición tiene todos sus elementos no nulos, es decir, P^n es una matriz con todas sus entrada mayores que cero para algún $n \in \mathbb{N}$. A este tipo de cadenas de Markov se le conocen como cadenas regulares; un método no muy eficiente para conocer si una cadena es regular consiste en calcular las potencias de la matriz de transición asociada a la cadena hasta encontrar un n tal que P^n tenga todas sus entradas estrictamente positivas.

La Proposición 22 es equivalente a la Definición 27. La Proposición 23 enuncia el comportamiento límite de algunas cadenas finitas.

Definición 27. *Se dice que una cadena finita o su matriz de probabilidades de transición es regular si existe un natural* n *tal que* $p_{ij}(n) > 0$, *para cualesquiera estados* i *y* j.

Proposición 22. *Una matriz estocástica es regular si, y sólo si, es finita, irreducible y aperiódica.*

Proposición 23. *Toda cadena finita que además es:*

1. *regular, tiene como distribución límite la única solución no negativa del sistema de ecuaciones 2.8.*
2. *regular y doblemente estocástica, tiene como distribución límite la distribución uniforme.*
3. *irreducible, aperiódica y doblemente estocástica, tiene como distribución límite la distribución uniforme.*

2.13. Cadenas reversibles

Se cierra este capítulo con un muy breve estudio de un tipo particular de cadenas irreducibles conocidas como cadenas reversibles. La condición de reversibilidad suele describirse afirmando que para cualesquiera dos estados j e i de una cadena de Markov la tasa de transición de i a j coincide con la tasa de transición de j a i. La Proposición 24 describe como caracterizar una cadena de Markov reversible para la cual exista una distribución que satisfaga la identidad 2.9.

Definición 28. *Se dice que una cadena de Markov irreducible con probabilidades de transición p_{ij} y con distribución estacionaria π es reversible en el tiempo si para cualesquiera estados i y j,*

$$\pi_i p_{ij} = \pi_j p_{ji}. \tag{2.9}$$

Proposición 24. *Considere una cadena irreducible para la cual existe una distribución π que cumple la identidad 2.9. Entonces π es una distribución estacionaria y la cadena es reversible.*

CAPÍTULO 3

Muestreador de Gibbs

El muestreador de Gibbs es uno de los algoritmos markovianos, estos son conocidos como métodos MCMC (proviene del inglés Monte Carlo Markov Chain). Los cuales se basan en emplear el método de Monte Carlo vía cadenas de Markov. A continuación, presentaremos un muy breve desarrollo histórico de los métodos MCMC. Para esto, es conveniente iniciar por el método de Montecarlo, ya que como se ha dicho es la raíz de los algoritmos MCMC. El método de Montecarlo surge entre 1944 y 1946, su descubrimiento lo comparten los matemáticos Stanislaw Ulam y John von Neumann. Es preciso destacar que la idea original del método se debe a Ulam, pero es Neumann quien realiza mejoras y un aporte significativo para que el método de Monte Carlo se hubiese implementado en 1948, para obtener los valores singulares de la ecuación de Schrödinger aunque también se recibió aportes del físico Enrico Fermi.

Más tarde, en 1953 el físico-matemático Nicholas Constantine Metrópolis propone el primer algoritmo que lleva su nombre, para generar muestras de la distribución de Boltzmann. Dicho algoritmo es generalizado en 1970 por el estadístico canadiense W.K. Hastings

Vale la pena señalar, la estrecha relación que ha existido entre la física y los algoritmos MCMC. Del cual el muestreador de Gibbs no es la excepción, pues surge en el campo de la física estadística con el propósito de poder simular campos aleatorios markovianos, en especial los campos aleatorios de Gibbs.

El muestreador de Gibbs fue propuesto en 1984 por los hermanos y matemáticos Stuart Alan Geman y Donald Jay Geman en el paper “Stochastic Relaxation, Gibbs Distribution, and the Bayesian Restoration of Images”. Pero, es hasta en 1990 con la publicación de Alan E. Gelfan y Adrian F.M. Smith de “Sampling-Based Approaches to Calculating Marginal Densities ”y por sus conferencias impartidas que el muestreador de Gibbs se considera como una revelación en estadística. Además debido a los recientes avances de la computación el muestreador de Gibbs se aplica ampliamente en campos como las finanzas, la economía y la biología entre otros.

3.1. Algoritmo del muestreador de Gibbs para el caso bivariado

Esta sección es un primer acercamiento a la mecánica del muestreador de Gibbs. Pues aquí se presenta un algoritmo para implementarlo en el caso bivariado. El objetivo de dicho algoritmo es que el lector comprenda cómo surge la sucesión de Gibbs para el caso que estamos considerando y que observe también, el mecanismo sencillo que describe al muestreador de Gibbs.

Algoritmo 1.

Entrada:
- Distribuciones condicionales, $f_{X|Y}(x \mid y)$ y $f_{Y|X}(y \mid x)$.
- Número máximo de iteraciones k.
- Valor inicial de $Y_0' = y_0'$.

Salida: Sucesión de Gibbs

Paso 1. Sea $i = 0, Y_0' = y_0'$.

Paso 2. Mientras $i + 1 \leq k$ realizar pasos 3 y 4.

Paso 3. Obtener X_i' mediante $f(x \mid Y_i' = y_i')$.

Paso 4. Obtener Y_{i+1}' mediante $f(y \mid X_i' = x_i')$.

- **Salida** $\left(X_i', Y_{i+1}'\right)$ (Algoritmo culminado satisfactoriamente).
- **PARAR.**

Paso 5. **Salida** (Se pidió la implementación del método como máximo para k iteraciones).

- **PARAR.**

Los valores

$$Y_0', X_0', Y_1', X_1', Y_2', X_2', \ldots, Y_k', X_k', \tag{3.1}$$

obtenidos al aplicar el Algoritmo 1 se conocen como "Sucesión de Gibbs".

El éxito del muestreador de Gibss se basa en que haciendo k lo suficientemente grande, la observación de $X_k' = x_k'$ es un punto muestral de $f(x)$, es decir, cuando $k \to \infty$ la distribución de $X_k' \to f(x)$ la verdadera marginal de X.

3.2. Ilustrando el muestreador de Gibss

Ahora ya familiarizados con el muestreador de Gibbs podemos implementarlo. Esta sección lo hace, su objetivo es comparar los resultados obtenidos mediante métodos analíticos y el muestreador de Gibbs. Para esto se consideran dos distribuciones conjuntas, a saber, las que establecen las ecuaciones 3.6 y 3.9. Para dichas distribuciones se encuentra por medios analíticos sus distribuciones marginales y condicionales. Luego se emplea el muestreador de Gibbs para simular la distribución marginal de X, para las distribuciones en 3.6 y 3.9. Por último se comparan los histogramas entre la distribución teórica y la simulada. Esto se desarrolla en la subsecciones 3.2.2 y 3.2.3. Pues en la primera se quiso realizar una introducción a la distribución marginal que se simulará mediante Gibbs en la subsección 3.2.2. Además, porque dicha distribución es frecuente en la estadística Bayesiana y se presenta una forma para obtenerla con métodos de estadística clásica.

3.2.1. Distribución Beta-binomial

En estadística clásica los parámetros de una distribución son siempre constantes. Ahora bien, en estadística Bayesiana se consideran distribuciones cuyos parámetros no son constantes sino aleatorios. Un ejemplo, de tal situación es la familia de distribuciones discretas de probabilidad conocida como Beta-binomial. Esta se obtiene considerando que la probabilidad de éxito p, en una distribución $Bin(n,p)$, por cada ensayo no es fija sino que se distribuye $Beta(\alpha,\beta)$.

Notación 4. *La función Beta es,*

$$B\left(\alpha,\beta\right)=\int_0^1 y^{\alpha-1}\left(1-y\right)^{\beta-1}dy.$$

La notaremos por $B\left(\alpha,\beta\right)$ *.*

3.2.1.1. Distribución Beta-binomial como distribución compuesta

Sea X una variable aleatoria con parámetro P que se distribuye $Beta(\alpha,\beta)$. Sabemos que,

$$p_{X|P}\left(x\mid p\right)=\binom{n}{x}p^{k}\left(1-p\right)^{n-x}, \tag{3.2}$$

entonces

$$f\left(p,x\right)=p_{X|P}\left(x\mid p\right)=\binom{n}{x}p^{x}\left(1-p\right)^{n-x}\frac{p^{\alpha-1}\left(1-p\right)^{\beta-1}}{B\left(\alpha,\beta\right)}. \tag{3.3}$$

Así que la distribución compuesta está dada por,

$$\begin{aligned} p_X(x) &= \int_0^1 \binom{n}{x} p^x (1-p)^{n-x} \frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha,\beta)} dp \\ &= \binom{n}{x} \frac{1}{B(\alpha,\beta)} \int_0^1 p^x (1-p)^{n-x} p^{\alpha-1} (1-p)^{\beta-1} dp \\ &= \binom{n}{x} \frac{1}{B(\alpha,\beta)} \int_0^1 B(x+\alpha, n-x+\beta) \frac{p^{x+\alpha-1}(1-p)^{n-x+\beta-1}}{B(x+\alpha,n-x+\beta)} dp \\ &= \binom{n}{x} \frac{B(x+\alpha,n-x+\beta)}{B(\alpha,\beta)}. \end{aligned} \tag{3.4}$$

Empleando la identidad $B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$, se puede reescribir la ecuación 3.4 como sigue,

$$\begin{aligned} p_X(x) &= \binom{n}{x} \frac{B(x+\alpha,n-x+\beta)}{B(\alpha,\beta)} \\ &= \binom{n}{x} \frac{\frac{\Gamma(x+\alpha)\Gamma(n-x+\beta)}{\Gamma(n+\alpha+\beta)}}{\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}} \\ &= \binom{n}{x} \frac{\Gamma(\alpha+\beta)\Gamma(x+\alpha)\Gamma(n-x+\beta)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\alpha+\beta+n)} \end{aligned} \tag{3.5}$$

la ecuación 3.5, corresponde a la función de probabilidad de la distribución beta-binomial.

3.2.2. Aplicando el Muestreador de Gibbs

Se busca exponer que el muestreador de Gibbs es un método óptimo. Por ejemplo, para generar una muy buena aproximación a la densidad marginal $f(x)$ de una distribución conjunta $f(x,y)$. Dicho lo anterior, dada una función de distribución conjunta se obtendrá de manera analítica y haciendo uso del muesteador de Gibbs la distribución marginal para una variable aleatoria distribuida beta-binomial.

A continuación, se presenta el desarrollo analítico para las variables aleatorias X y Y con función de probabilidad conjunta dada por:

$$f_{X,Y}(x,y) = \binom{n}{x} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{x+\alpha-1}(1-y)^{n-x+\beta-1} \quad x = 0,1,2,3,\ldots,n \quad 0 \leq y \leq 1. \tag{3.6}$$

Con respecto a 3.6, es posible encontrar las distribuciones $f_X(x), f_Y(y), f_{X|Y}(x,y)$ y $f_{Y|X}(y|x)$ como sigue

$$
\begin{aligned}
f_X(x) &= \int_0^1 \binom{n}{x} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{x+\alpha-1}(1-y)^{n-x+\beta-1}\,dy \\
&= \binom{n}{x} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 y^{x+\alpha-1}(1-y)^{n-x+\beta-1}\,dy \\
&= \binom{n}{x} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(x+\alpha)\Gamma(n-x+\beta)}{\Gamma(\alpha+n+\beta)}.
\end{aligned} \tag{3.7}
$$

Así que $f_X(x)$ es beta-binomial de parámetros n, α y β. A continuación,

$$
\begin{aligned}
f_Y(y) &= \sum_{x=0}^{n} \binom{n}{x} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{x+\alpha-1}(1-y)^{n-x+\beta-1} \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1}(1-y)^{\beta-1} \sum_{x=0}^{n} \binom{n}{x} y^{x}(1-y)^{n-x} \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1}(1-y)^{\beta-1}.
\end{aligned}
$$

Es decir, $f_Y(y)$ es $Beta(\alpha,\beta)$. Luego,

$$
\begin{aligned}
f_{Y|X}(y \mid x) &= \frac{\binom{n}{x} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{x+\alpha-1}(1-y)^{n-x+\beta-1}}{\binom{n}{x} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(x+\alpha)\Gamma(n-x+\beta)}{\Gamma(\alpha+n+\beta)}} \\
&= \frac{\Gamma(\alpha+n+\beta)}{\Gamma(x+\alpha)\Gamma(n-x+\beta)} y^{x+\alpha-1}(1-y)^{n-x+\beta-1} \\
&= \left(\frac{\Gamma(x+\alpha)\Gamma(n-x+\beta)}{\Gamma(n+\alpha+\beta)}\right)^{-1} y^{x+\alpha-1}(1-y)^{n-x+\beta-1} \\
&= \frac{1}{B(x+\alpha, n-x+\beta)} y^{x+\alpha-1}(1-y)^{n-x+\beta-1},
\end{aligned}
$$

donde se hizo uso de $B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$, para obtener que $f_{Y|X}(y \mid x)$ es Beta$(x+\alpha, n-x+\beta)$. Por último,

$$
\begin{aligned}
f_{X|Y}(x \mid y) &= \frac{\binom{n}{x} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{x+\alpha-1}(1-y)^{n-x+\beta-1}}{\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1}(1-y)^{\beta-1}} \\
&= \binom{n}{x} y^{x}(1-y)^{n-x},
\end{aligned}
$$

es decir $f_{X|Y}(x \mid y)$ es Binomial(n, y).

Hemos realizado una simulación en R con $k = 100000$ iteraciones, de valor inicial $Y_0' = 0.4$, para la distribución marginal de X de la distribución conjunta 3.6, implementando el muestreador de Gibbs. Esto con el fin de contrastar los resultados analíticos y el muestreador de Gibbs. Ahora, el lector puede apreciar el objetivo de haber encontrado por metodos analíticos la ecuación 3.7. Pues esta indica que la distribución marginal de X es beta-binomial y que la gráfica producida por la marginal simulada se debe parecer al comportamiento de dicha distribución. En efecto, esto lo muestra la figura 3.1, donde el histograma de color azul representa el obtenido empleando el muestreador de Gibbs y el de color rojo se obtuvo directamente de la distribución beta-binomial con parámetros $n = 10$, $\alpha = 2$ y $\beta = 5$. El lector puede apreciar también que ambos histogramas son similares, hecho que evidencia lo eficaz del muestreador de Gibss.

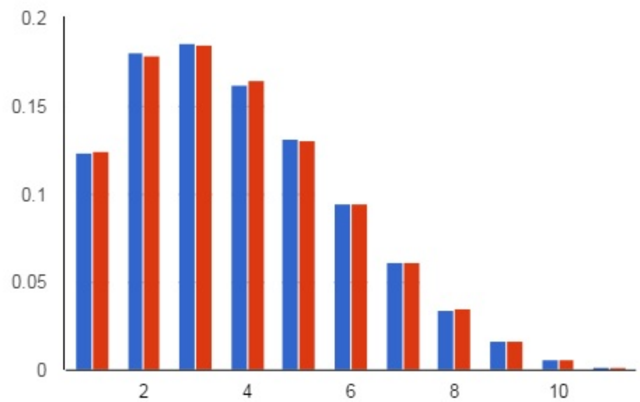

Figura 3.1. Comparación de histogramas

Hemos también querido implementar la prueba ji-cuadrado. Con el fin de comparar los valores obtenidos mediante la sucesión de Gibbs y los teóricos. El tamaño de la muestra es 100000, se construyeron $k = 11$ celdas. La hipótesis nula es H_0 : Los datos provienen de una distribución Beta-binomial con parámetros $n = 10$, $\alpha = 2$ y $\beta = 5$. La hipótesis alterna es H_1: Los datos no provienen de una distribución Beta-binomial con parámetros $n = 10$, $\alpha = 2$ y $\beta = 5$. Con el código 3.4 se obtuvó un p-valor de 1 y un valor χ^2 de 0.0001129436 dado que este es menor que $\chi^2_{0.05} = 18.3070$ para 10 grados de libertad. Se acepta H_0, es decir, no hay evidencia para decir que estos datos no provienen de una distribución Beta-binomial con parámetros $n = 10$, $\alpha = 2$ y $\beta = 5$.

También se ha simulado la distribución beta-binomial con el muestreador de Gibbs, con el ánimo de observar que para valores pequeños la distribución simulada se asemeja rápidamente a la teórica. La figura 3.2 permite apreciar esto. Los histogramas en azul y verde representan la distribución simuladas empleando Gibbs con $k = 1000$ y $k = 10000$ respectivamente y el histograma de color rojo la distribución teórica, es decir, la distribución beta-binomial con parámetros $n = 10$, $\alpha = 2$ y $\beta = 5$.

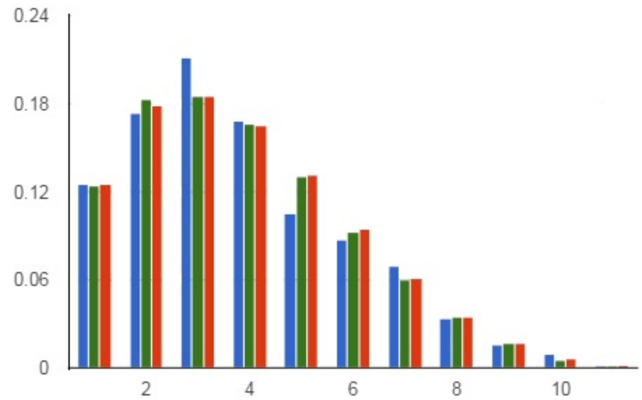

FIGURA 3.2. Comparando el muestreador de Gibbs.

3.2.3. Distribución trinomial

La distribución trinomial es un caso particular de la distribución multinomial, que es la análoga de la distribución binomial en el caso multivariado. La distribución trinomial cuantifica la probabilidad de que en n ensayos independientes se obtenga precisamente x_i veces un evento E_i, donde $i = 1, 2, 3$. Con la condición de que en un único ensayo se obtenga uno de los E_i, Además, p_i denota la correspondiente probabilidad. La distribución trinomial tiene función de probabilidad dada por

$$f(x_1, x_2, x_3; n, p_1, p_2, p_3) = \begin{cases} \frac{n!}{x_1!x_2!x_3!} p_1^{x_1} p_2^{x_2} p_3^{x_3} & \text{si } \sum_{i=1}^{3} x_i = n, \\ 0 & \text{en otro caso.} \end{cases} \tag{3.8}$$

También 3.8 satisface $\sum_{i=1}^{3} p_i = 1$, de donde se sigue que $p_3 = 1 - p_1 - p_2$. Por otro lado $x_3 = n - x_1 - x_2$. Así que reemplazando estos valores en 3.8, se logra suprimir la dependencia de x_3 y p_3, es decir, podemos reescribir 3.8 como sige,

$$f(x_1, x_2; n, p_1, p_2) = \frac{n!}{x_1!x_2!\,(n - x_1 - x_2)!} p_1^{x_1} p_2^{x_2} (1 - p_1 - p_2)^{n - x_1 - x_2}\,. \tag{3.9}$$

Encontremos ahora la distribución marginal de X_1 y X_2.

$$\begin{aligned} f_{X_1}(x_1) &= \sum_{x_2=0}^{n-x_1} \frac{n!}{x_1!x_2!\,(n - x_1 - x_2)!} p_1^{x_1} p_2^{x_2} (1 - p_1 - p_2)^{n - x_1 - x_2} \\ &= \frac{p_1^{x_1}}{x_1!} \sum_{x_2=0}^{n-x_1} \frac{n!}{x_2!\,(n - x_1 - x_2)!} p_2^{x_2} (1 - p_1 - p_2)^{n - x_1 - x_2} \end{aligned} \tag{3.10}$$

antes de realizar el siguiente paso observe que $x_1 + x_2 + x_3 = n$, de donde $n \geq x_1$, así que

$$n! = \underbrace{(1)(2)(3)\cdots(n - x_1)}_{(n-x_1)!}(n - x_1 + 1)(n - x_1 + 2)\cdots(n)$$

regresando a 3.10

$$
\begin{aligned}
&= \frac{p_1^{x_1}}{x_1!}(n-x_1+1)(n-x_1+1)\cdots(n)\sum_{x_2=0}^{n-x_1}\frac{(n-x_1)!}{x_2!\,(n-x_1-x_2)!}p_2^{x_2}\left(1-p_1-p_2\right)^{n-x_1-x_2} \\
&= \frac{p_1^{x_1}}{x_1!}(n-x_1+1)\cdots(n)\sum_{x_2=0}^{n-x_1}\binom{n-x_1}{x_2}p_2^{x_2}\left(1-p_1-p_2\right)^{n-x_1-x_2} \\
&= \frac{p_1^{x_1}}{x_1!}\frac{(n-x_1)!(n-x_1+1)\cdots(n)}{(n-x_1)!}\left(p_2+(1-p_1-p_2)\right)^{n-x_1} \\
&= \frac{n!}{x_1!\,(n-x_1)!}p_1^{x_1}\left(1-p_1\right)^{n-x_1} \\
&= \binom{n}{x_1}p_1^{x_1}\left(1-p_1\right)^{n-x_1}.
\end{aligned} \tag{3.11}
$$

Es así que $f_{X_1}(x_1)$ es Binomial(n,p_1). En seguida para $f_{X_2}(x_2)$, se tiene que

$$
\begin{aligned}
f_{X_2}(x_2) &= \sum_{x_1=0}^{n-x_2}\frac{n!}{x_1!x_2!\,(n-x_1-x_2)!}p_1^{x_1}p_2^{x_2}\left(1-p_1-p_2\right)^{n-x_1-x_2} \\
&= \frac{p_2^{x_2}n!}{x_2!}\sum_{x_1=0}^{n-x_2}\frac{(n-x_2)!}{(n-x_2)!x_1!(n-x_2-x_1)!}p_1^{x_1}\left(1-p_1-p_2\right)^{n-x_1-x_2} \\
&= \frac{p_2^{x_2}n!}{(n-x_2)!x_2!}\sum_{x_1=0}^{n-x_2}\frac{(n-x_2)!}{x_1!(n-x_2-x_1)!}p_1^{x_1}\left(1-p_1-p_2\right)^{n-x_2-x_1} \\
&= \binom{n}{x_2}p_2^{x_2}\sum_{x_1=0}^{n-x_2}\binom{n-x_2}{x_1}p_1^{x_1}\left(1-p_1-p_2\right)^{n-x_2-x_1} \\
&= \binom{n}{x_2}p_2^{x_2}\left(p_1+(1-p_1-p_2)\right)^{n-x_2} \\
&= \binom{n}{x_2}p_2^{x_2}\left(1-p_2\right)^{n-x_2}.
\end{aligned}
$$

De donde $f_{X_2}(x_2)$ es Binomial(n,p_2). Encontremos a continuación la distribución condicional, $f_{X_1|X_2}(x_1 \mid x_2)$,

$$
\begin{aligned}
f_{X_1|X_2}(x_1 \mid x_2) &= \frac{\frac{n!}{x_1!x_2!(n-x_1-x_2)!}p_1^{x_1}p_2^{x_2}\left(1-p_1-p_2\right)^{n-x_1-x_2}}{\frac{n!}{x_2!(n-x_2)!}p_2^{x_2}\left(1-p_2\right)^{n-x_2}} \\
&= \frac{n!x_2!\left(n-x_2\right)!}{n!x_1!x_2!\left(n-x_1-x_2\right)!}\frac{p_2^{x_2}}{p_2^{x_2}}\frac{p_1^{x_1}\left(1-p_1-p_2\right)^{n-x_1-x_2}}{\left(1-p_2\right)^{n-x_2}}\frac{\left(1-p_2\right)^{x_1}}{\left(1-p_2\right)^{x_1}} \\
&= \frac{\left(n-x_2\right)!}{x_1!\left(n-x_2-x_1!\right)}\frac{p_1^{x_1}}{\left(1-p_2\right)^{x_1}}\frac{\left(1-p_1-p_2\right)^{n-x_1-x_2}}{\left(1-p_2\right)^{-x_1}\left(1-p_2\right)^{n-x_2}} \\
&= \binom{n-x_2}{x_1}\left(\frac{p_1}{1-p_2}\right)^{x_1}\left(\frac{1-p_1-p_2}{1-p_2}\right)^{n-x_2-x_1} \\
&= \binom{n-x_2}{x_1}\left(\frac{p_1}{1-p_2}\right)^{x_1}\left(\frac{1-p_2}{1-p_2}-\frac{p_1}{1-p_2}\right)^{n-x_2-x_1} \\
&= \binom{n-x_2}{x_1}\left(\frac{p_1}{1-p_2}\right)^{x_1}\left(1-\frac{p_1}{1-p_2}\right)^{n-x_2-x_1}.
\end{aligned}
$$

Por lo tanto $f_{X_1|X_2}(x_1 \mid x_2)$ se distribuye binomial de parametros $n-x_2$ y $\frac{p_1}{1-p_2}$. Para la distribución condicional, $f_{X_2|X_1}\left(x_2 \mid x_1\right)$, se tiene que,

$$
\begin{aligned}
f_{X_2|X_1}\left(x_2 \mid x_1\right) &= \frac{\frac{n!}{x_1!x_2!(n-x_1-x_2)!}p_1^{x_1}p_2^{x_2}\left(1-p_1-p_2\right)^{n-x_1-x_2}}{\frac{n!}{x_1!(n-x_1)!}p_1^{x_1}\left(1-p_1\right)^{n-x_1}} \\
&= \frac{n!x_1!\left(n-x_1\right)!}{n!x_1!x_2!\left(n-x_1-x_2\right)!}\frac{p_1^{x_1}}{p_1^{x_1}}\frac{p_2^{x_2}\left(1-p_1-p_2\right)^{n-x_1-x_2}}{\left(1-p_1\right)^{n-x_1}}\frac{\left(1-p_1\right)^{x_2}}{\left(1-p_1\right)^{x_2}} \\
&= \frac{\left(n-x_1\right)!}{x_2!\left(n-x_1-x_2\right)!}\frac{p_2^{x_2}}{\left(1-p_1\right)^{x_2}}\frac{\left(1-p_1-p_2\right)^{n-x_1-x_2}}{\left(1-p_1\right)^{n-x_1}\left(1-p_1\right)^{-x_2}} \\
&= \binom{n-x_1}{x_2}\left(\frac{p_2}{1-p_1}\right)^{x_2}\left(\frac{1-p_1-p_2}{1-p_1}\right)^{n-x_1-x_2} \\
&= \binom{n-x_1}{x_2}\left(\frac{p_2}{1-p_1}\right)^{x_2}\left(\frac{1-p_1}{1-p_1}-\frac{p_2}{1-p_1}\right)^{n-x_1-x_2} \\
&= \binom{n-x_1}{x_2}\left(\frac{p_2}{1-p_1}\right)^{x_2}\left(1-\frac{p_2}{1-p_1}\right)^{n-x_1-x_2}.
\end{aligned}
$$

Es decir, $f_{X_2|X_1}\left(x_2 \mid x_1\right)$ se distribuye binomial de parametros $n-x_1$ y $\frac{p_2}{1-p_1}$.

La figura 3.3 contrasta la distribución teórica en 3.11 y la simulada empleando el muestreador de Gibbs, que ha sido creada en R con k=200000 iteraciones. En dicha figura puede apreciarse lo casi indistinguible que es el comportamiento de ambas distribuciones, en donde el histograma en color azul representa la distribución simulada mediante Gibbs y el histograma en color rojo la distribución teórica en 3.11.

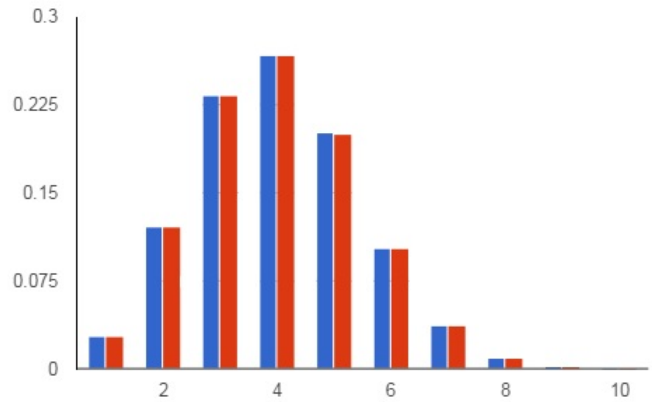

FIGURA 3.3. Comparación de dos Histogramas

A continuación se presenta la prueba ji-cuadrado con el fin de comparar los valores obtenidos mediante la sucesión de Gibbs y los teóricos. El tamaño de la muestra es 200000, se construyeron $k = 11$ celdas. La hipótesis nula es H_0 : Los datos provienen de una distribución trinomial con parámetros $n = 10$ $p_1 = 0.3$, $p_2 = 0.2$ y $p_3 = 0.5$. La hipótesis alterna es H_1: Los datos no provienen de una distribución trinomial con parámetros $n = 10$ $p_1 = 0.3$, $p_2 = 0.2$ y $p_3 = 0.5$. En R se obtuvó un p-valor de 0.9999997 y un valor χ^2 de 0.2569682 como este valor es menor que $\chi^2_{0.05} = 18.3070$ con 10 grados de libertad. Se acepta H_0, es decir, no hay evidencia para decir que estos datos no provienen de una distribución distribución trinomial con parámetros $n = 10$ $p_1 = 0.3$, $p_2 = 0.2$ y $p_3 = 0.5$

3.3. Una prueba de Convergencia simple

En la sección anterior se obtuvo de manera analítica y por medio del muestreador de Gibbs la distribución marginal para la variable aleatoria X de dos distribuciones conjuntas $f(X, Y)$. Ahora, esta nueva sección presenta una matriz que notaremos por $A_{X|X}$ y posee la propiedad de que su distribución estacionaria es f_X.Si bien su desarrollo se hace en el caso discreto también presentamos un paralelo entre la versión discreta y la continua.

El objetivo de esta sección es que el lector se percate de la naturaleza markoviana del muestreador de Gibbs, para ello se han hecho comentarios y consideraciones que permitan apreciarlo.

Para comenzar consideremos primero un ejemplo, Sean X y Y dos variables aleatorias cada una Bernoulli, con función de distribución de probabilidad conjunta como lo indica la tabla siguiente:

Y \ X	0	1	Suma
0	0.2	0.1	0.3
1	0.4	0.3	0.7
Suma	0.6	0.4	1

TABLA 3.1. Distribución de probabilidades de (X, Y).

podemos expresar la tabla 3.1 en términos de la función de probabilidad conjunta, es decir,

$$\begin{bmatrix} f_{X,Y}(0,0) & f_{X,Y}(1,0) \\ f_{X,Y}(0,1) & f_{X,Y}(1,1) \end{bmatrix} = \begin{bmatrix} 0.2 & 0.1 \\ 0.4 & 0.3 \end{bmatrix}$$

así que las distribuciones marginales de X y Y están dadas por

$$f_X = [f_X(0), f_X(1)] = [0.6, 0.4]$$

$$f_Y = [f_Y(0), f_Y(1)] = [0.3, 0.7]\,,$$

que se distribuyen Bernoulli con probabilidad de éxito 0.4 y 0.7. Luego podemos calcular la distribución marginal de $X \mid Y = y$,

$$f_{X|Y}\left(0 \mid 0\right) = \frac{P\left(0,0\right)}{P\left(0\right)} = \frac{0.2}{0.3}$$

$$f_{X|Y}\left(1 \mid 0\right) = \frac{P\left(1,0\right)}{P\left(0\right)} = \frac{0.1}{0.3}$$

$$f_{X|Y}\left(0 \mid 1\right) = \frac{P\left(0,1\right)}{P\left(1\right)} = \frac{0.4}{0.7}$$

$$f_{X|Y}\left(1 \mid 1\right) = \frac{P\left(1,1\right)}{P\left(1\right)} = \frac{0.3}{0.7}.$$

Para construir una matriz que contenga las probabilidades condicionales de X dado $Y = y$

$$A_{X|Y} = \begin{bmatrix} \frac{0.2}{0.3} & \frac{0.1}{0.3} \\ \frac{0.4}{0.7} & \frac{0.3}{0.7} \end{bmatrix}$$

y de manera análoga podemos calcular la distribución marginal de $Y \mid X = x$, y obtener la siguiente matriz

$$A_{Y|X} = \begin{bmatrix} \frac{0.2}{0.6} & \frac{0.4}{0.6} \\ \frac{0.1}{0.4} & \frac{0.3}{0.4} \end{bmatrix}.$$

Si consideramos el espacio de estados $\{0, 1\}$, es natural pensar en las matrices $A_{X|Y}$ y $A_{Y|X}$ como matrices de transición. Donde por ejemplo $A_{X|Y}$ contiene las probabilidades de ir del estado X al estado Y y $A_{Y|X}$ posee las probabilidades de acceder del estado Y al estado X.

Con esta información uno estaría interesado en obtener la matriz de probabilidad de transición de la distribución marginal de X (Denotada por $A_{X|X}$). Para tal fin el interés recae sobre la subsucesión X' de la sucesión 3.1. Dado que 3.1 produce $X_0' \rightarrow Y_1' \rightarrow X_1' \rightarrow Y_2'$ y ya que $X_0' \rightarrow X_1'$ forma una cadena de Markov e implica tener que pasar por Y_1'. Podemos escribir esto en términos de la probabilidad condicional gracias a la ecuacion 2 para obtener

$$P\left(X_1' = x_1 | X_0' = x_0\right) = \sum_y P\left(X_1' = x_1 | Y_1' = y\right) \times P\left(Y_1' = y | X_0' = x_0\right),$$

con esto $A_{X|X}$ viene dado por

$$A_{X|X} = A_{Y|X} A_{X|Y} = \begin{bmatrix} \frac{0.2}{0.6} & \frac{0.4}{0.6} \\ \frac{0.1}{0.4} & \frac{0.3}{0.4} \end{bmatrix} \begin{bmatrix} \frac{0.2}{0.3} & \frac{0.1}{0.3} \\ \frac{0.4}{0.7} & \frac{0.3}{0.7} \end{bmatrix} = \begin{bmatrix} \frac{38}{63} & \frac{25}{63} \\ \frac{25}{42} & \frac{17}{42} \end{bmatrix}. \tag{3.12}$$

En resumen hemos construidos la matriz de transición para ir de X_0' a X_1'. Esto es fundamental para el presente trabajo pues se puso en evidencia la naturaleza Markoviana de la iteración del Muestreador de Gibbs al considerar que ir de X_0' a X_1' forma una Cadena de Markov y poder escribir $A_{X|X}$ como un producto de matrices de transición. En general la idea fundamental de los algoritmos MCMC, es el diseño de una cadena de Markov cuya distribución estacionaria es precisamente la distribución que se quiere muestrear. Para el caso del muestreador de Gibbs presentamos la generalización del ejemplo anterior y uno empleando la distribución trinomial.

En este orden de ideas consideremos consideremos la tabla 3.2, con muestreo multinomial dado que X y Y son variables aleatorias con distribución conjunta Bernoulli.

Y \ X	0	1
0	p_1	p_2
1	p_3	p_4

TABLA 3.2. Distribución de probabilidades conjunta de la variable aleatoria discreta (X, Y).

Teniendo en cuenta que $p_i \geq 0$, para $i \in \{1, 2, 3, 4\}$ y es tal que $p_1 + p_2 + p_3 + p_4 = 1$. Luego expresando la tabla 3.2 en términos de la función de probabilidad conjunta, tenemos que

$$\begin{bmatrix} f_{X,Y}(0,0) & f_{X,Y}(1,0) \\ f_{X,Y}(0,1) & f_{X,Y}(1,1) \end{bmatrix} = \begin{bmatrix} p_1 & p_2 \\ p_3 & p_4 \end{bmatrix}$$

se sigue también que la distribución marginal de X es una distribución Bernoulli con probabilidad de éxito $p_2 + p_4$.

$$f_X = [f_X(0), f_X(1)] = [p_1 + p_3 \quad p_2 + p_4] \tag{3.13}$$

de manera análoga a como se hizo en el ejemplo anterior se puede construir dos matrices con todas las distribuciones condicionales de $X|Y$ y $Y|X$ a saber:

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

y

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

Si consideramos el espacio de estados $\{0,1\}$, es natural pensar en las matrices $A_{X|Y}$ y $A_{Y|X}$ como matrices de transición. Donde por ejemplo $A_{X|Y}$ contiene las probabilidades de ir del estado X al estado Y y $A_{Y|X}$ posee las probabilidades de acceder del estado Y al estado X. Además sí aplicamos el esquema iterativo (3.1) se obtendría una sucesión de ceros y unos.

Nuevamente nos preguntamos ¿cómo obtener la distribución marginal de X?, para hacerlo centraremos nuestra atención en dos hechos, primero en la subsucesión de X' de la sucesión de Gibbs y segundo en que los valores en 3.1 se obtuvieron fijando un valor inicial y evaluando en las condicionales respectivas su inmediato anterior. Así que si consideramos la sucesión de Gibbs como $X_0' \to Y_1' \to X_1'$ esta nos permite acceder de X_0' a X_1' pasando por Y_1'. Además ya que $X_0' \to X_1'$ forma una cadena de Markov podemos aplicar la ecuación 2 para expresar esto en términos de la probabilidad de transición, es decir,

$$P\left(X_1' = x_1 | X_0' = x_0\right) = \sum_y P\left(X_1' = x_1 | Y_1' = y\right) \times P\left(Y_1' = y | X_0' = x_0\right) \quad (3.14)$$

de donde resulta que la matriz de probabilidad de transición $A_{X|X}$ de la sucesión X', viene dada por

$$A_{X|X} = A_{Y|X}A_{X|Y} = \begin{bmatrix} \frac{1}{p_1+p_3}\left(\frac{p_1^2}{p_1+p_2} + \frac{p_3^2}{p_3+p_4}\right) & \frac{1}{p_1+p_3}\left(\frac{p_1p_2}{p_1+p_2} + \frac{p_3p_4}{p_3+p_4}\right) \\ \frac{1}{p_2+p_4}\left(\frac{p_2p_1}{p_1+p_2} + \frac{p_4p_3}{p_3+p_4}\right) & \frac{1}{p_2+p_4}\left(\frac{p_2^2}{p_1+p_2} + \frac{p_4^2}{p_3+p_4}\right) \end{bmatrix}. \quad (3.15)$$

De manera que la matriz de transición para $P\left(X_k' = x_k | X_0' = x_0\right)$ es $\left(A_{X|X}\right)^k$. Esto implica que podemos calcular fácilmente la distribución de probabilidad de cualquier X_k' en la sucesión 3.1. En efecto, si denotamos la distribución marginal de X_k' por

$$f_k = \left[f_k(0), f_k(1)\right]$$

entonces para cualquier k, tendremos que,

$$f_k = f_0 A_{X|X}^k = \left(f_0 A_{X|X}^{k-1}\right) A_{X|X} = f_{k-1} A_{X|X}. \quad (3.16)$$

El lector en este punto puede llegarse a preguntar ¿que relación tiene esto con el muestreador de Gibbs? la respuesta la daremos de inmediato. En efecto, ya que como se ha indicado desde el inició de este capítulo el muestreador de Gibbs es uno de los algoritmos MCMC, los cuales tienen como idea fundamental el diseño de una cadena de Markov cuya distribución estacionaria es precisamente la distribución que se quiere muestrear. Esto y el hecho de que cuando la matriz $A_{X|X}$ tiene todas sus entradas positivas, la ecuación 3.16 garantiza que cuando $k \to \infty$, f_k converge a la distribución f que es en un punto estacionario de 3.16. Sin importar la probabilidad inicial f_0. Mas aún el hecho de que se satisface,

$$fA_{X|X} = f. \tag{3.17}$$

Responde al lector la pregunta planteada, es decir, la relación de este desarrollo y el muestreador de Gibbs es que hemos visto la naturaleza markoviana del muestreador de Gibbs. Así mismo de la ecuación 3.17 para el caso descrito, podemos entender porque el muestreador de Gibbs es uno de los algoritmos MCMC.

Veamos ahora que para f_X en 3.13 se verifica 3.17,

$$\begin{aligned}
f_X A_{X|X} &= [p_1 + p_3 \quad p_2 + p_4] \begin{bmatrix} \frac{1}{p_1+p_3}\left(\frac{p_1^2}{p_1+p_2} + \frac{p_3^2}{p_3+p_4}\right) & \frac{1}{p_1+p_3}\left(\frac{p_1p_2}{p_1+p_2} + \frac{p_3p_4}{p_3+p_4}\right) \\ \frac{1}{p_2+p_4}\left(\frac{p_2p_1}{p_1+p_2} + \frac{p_4p_3}{p_3+p_4}\right) & \frac{1}{p_2+p_4}\left(\frac{p_2^2}{p_1+p_2} + \frac{p_4^2}{p_3+p_4}\right) \end{bmatrix} \\
&= [p_1 + p_3 \quad p_2 + p_4] \\
&= f_X
\end{aligned}$$

Además, si la sucesión 3.1 converge, la f que satisface 3.17 es la distribución marginal de X. Para indicar este hecho se ha implementado el siguiente ejemplo con el software estadístico R.

Ejemplo 1.

Consideremos la distribución dada por la tabla 3.3

Y \ X	0	1	2	3	suma
0	0.008	0.036	0.054	0.027	0.125
1	0.06	0.18	0.135	0	0.375
2	0.150	0.225	0	0	0.375
3	0.125	0	0	0	0.125
suma	0.343	0.441	0.189	0.027	1

Tabla 3.3. Distribución de probabilidades conjunta para el ejemplo 1.

Sea P la matriz obtenida de expresar la tabla 3.3 en términos de la función de probabilidad conjunta,es decir,

$$\begin{bmatrix} f_{X,Y}(0,0) & f_{X,Y}(1,0) & f_{X,Y}(2,0) & f_{X,Y}(3,0) \\ f_{X,Y}(0,1) & f_{X,Y}(1,1) & f_{X,Y}(2,1) & f_{X,Y}(3,1) \\ f_{X,Y}(0,2) & f_{X,Y}(1,2) & f_{X,Y}(2,2) & f_{X,Y}(3,2) \\ f_{X,Y}(0,3) & f_{X,Y}(1,3) & f_{X,Y}(2,3) & f_{X,Y}(3,3) \end{bmatrix} = \begin{bmatrix} 0.008 & 0.036 & 0.054 & 0.027 \\ 0.06 & 0.18 & 0.135 & 0 \\ 0.150 & 0.225 & 0 & 0 \\ 0.125 & 0 & 0 & 0 \end{bmatrix} = P.$$

Los renglones en la matriz $A_{Y|X}$ se obtienen dividiendo los respectivos renglones de P^T por la suma de la correspondiente columna en P. Por ejemplo calculemos la entrada $(2,1)$ de $A_{Y|X}$ esta es $0.036/0.441 = 0.08163265$. Así obtenemos

$$A_{Y|X} = \begin{bmatrix} 0.02332362 & 0.1749271 & 0.4373178 & 0.3644315 \\ 0.08163265 & 0.4081633 & 0.5102041 & 0 \\ 0.28571429 & 0.7142857 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

para obtener $A_{X|Y}$ dividimos cada renglon de P por la suma de todos los elementos en este renglón, por ejemplo calculemos la entrada $(1,1)$ en la matriz $A_{X|Y}$, esto es, $0.008/0.125 = 0.064$.

$$A_{X|Y} = \begin{bmatrix} 0.064 & 0.288 & 0.432 & 0.216 \\ 0.160 & 0.480 & 0.360 & 0 \\ 0.40 & 0.6 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

de manera que la matriz de transición de la subsucesión X' en la sucesión de Gibbs viene dada por

$$A_{X|X} = A_{Y|X}A_{X|Y} = \begin{bmatrix} 0.5688397 & 0.3530729 & 0.07304956 & 0.005037901 \\ 0.2746122 & 0.5255510 & 0.18220408 & 0.017632653 \\ 0.1325714 & 0.4251429 & 0.38057143 & 0.061714286 \\ 0.064 & 0.2880 & 0.4320 & 0.2160 \end{bmatrix} \quad (3.18)$$

la cadena de Markov de las X' es irreducible pues todos sus estados se comunican, es aperiodica ya que todos sus estados tienen periodo 1, como lo muestra la matriz 3.18. Además también posee distribución estacionaria $f = [0.343, 0.441, 0.189, 0.027]$ (que se corresponde con la marginal de X en la tabla 3.3). Así que por el teorema 20, se tiene que para cualesquiera estados i y j, $\lim_{n\to\infty} p_{ij}(n) = f_j$, en efecto, para $n = 21$ se tiene que:

$$A^{21}_{X|X} = \begin{bmatrix} 0.343 & 0.441 & 0.189 & 0.027 \\ 0.343 & 0.441 & 0.189 & 0.027 \\ 0.343 & 0.441 & 0.189 & 0.027 \\ 0.343 & 0.441 & 0.189 & 0.027 \end{bmatrix}.$$

Como un primer resultado se puede afirmar que para valores suficientemente grandes de k la distribución marginal de $X_k^{'}$ tiende a f_X. Así si detenemos la iteración (3.1) a un valor suficientemente grande de k, podemos suponer que la distribución de $X_k^{'}$ es aproximadamente f_X.

Casella y George (1992) argumentan que es posible considerar X, Y alguna o ambas continuas, si bien no presentan qué condiciones se deben imponer para extender los argumentos de dimensión finita, si afirman que la teoría continúa sirviendo y que por tanto el muestreador de Gibbs determina una muy buena aproximación a la distribución marginal de X. Es en este contexto que presentaremos un paralelo entre la versión discreta que se ha desarrollado y el caso continuo. Para ello es pertinente presentar pimero una versión para este último caso de la ecuación 3.14, es decir, una formula que represente la densidad condicional de X_1' dado X_0', esta es,

$$f_{X_1'|X_0'}(x_1 \mid x_0) = \int f_{X_1'|Y_1'}(x_1 \mid y)\, f_{Y_1'|X_0'}(y \mid x_0)\, dy.$$

Una vez hecha esta precisión, nuestro interés recae sobre tratar de describir una matriz análoga a $\left(A_{x|x}\right)^k$, es decir, la matriz de transición de $k-$pasos. Teniendo presente que X o Y pueden ser continuas, entonces el elemento buscado es “una matriz de transición infinita ”. Cuyas entradas satisfacen la siguiente ecuación:

$$f_{X_k'|X_0'}(x \mid x_0) = \int f_{X_k'|X_{k-1}'}(x \mid t)\, f_{X_{k-1}'|X_0}(t \mid x_0)\, dt, \tag{3.19}$$

Por ultimo vale la pena mencionar que la ecuación 3.19 representa la versión continua 3.16, así que la densidad $f_{X_k'|X_{k-1}'}$ representa una transición de un solo paso, y las otras dos densidades hacen referencia a f_k y f_{k-1}. También en [5] se describe que cuando $k \to \infty$ se deduce que el punto estacionario de (3.19) es la densidad marginal de X, la densidad para la cual $f_{X_k'|X_{k-1}'}$ converge.

El siguientes aspecto a tratar es revelar el argumento por el que funciona el muestreador de Gibbs además establecemos la forma limite de la ecuación 3.19 todo esto lo presentamos aquí en versión bivariada. Para comenzar sabemos que conocidas las distribuciones condicionales podemos determinar la distribución conjunta si esta existe, Lo que parece poner al muestreador de Gibbs como una aplicación de este hecho. Pero es conveniente señalar que no siempre conocidas las distribuciones condicionales estas determinen una distribución marginal adecuada. Para tal situación el lector interesado puede consultar [5] donde se presenta un ejemplo de tal situación.

Hay que recordarle al lector que el muestreador de Gibbs es capaz de simular la distribución marginal mediante las distribuciones condicionales correspondientes. Esta idea es clave para el siguiente desarrollo. Para comenzar sean X y Y dos variables aleatorias para las cuales conocemos las densidades condicionales $f_{X|Y}(x \mid y)$, $f_{Y|X}(y \mid x)$ y desconocemos $f_{XY}(X, Y)$. Encontraremos ahora la densidad marginal de X, $f_X(x)$, para hacerlo por definición sabemos que

$$f_X(x) = \int f_{XY}\left(x, y\right) dy,$$

donde $f_{XY}(x, y)$ es la densidad conjunta desconocida, pero sabemos que $f_{XY}\left(x, y\right) = f_{X|Y}\left(x \mid y\right) f_Y(y)$, se deduce que

$$f_X(x) = \int f_{X|Y}\left(x|y\right) f_Y\left(y\right) dy, \tag{3.20}$$

ahora reemplazando $f_Y(y) = \int f_{Y|X}\left(y|t\right) f_X\left(t\right) dt$, en 3.20 tenemos que

$$\begin{aligned} f_X(x) &= \int f_{X|Y}\left(x|y\right) \int f_{Y|X}\left(y|t\right) f_X\left(t\right) dtdy \\ &= \int \left[\underbrace{\int f_{X|Y}\left(x|y\right) f_{Y|X}\left(y|t\right) dy}_{h(x,t)} \right] f_X\left(t\right) dt \qquad (3.21) \\ &= \int h\left(x, t\right) f_X\left(t\right) dt. \end{aligned}$$

Es decir, escribimos la distribución marginal de X en términos de las condicionales. Razón por la cual el lector debe apreciar el porque del exito del muestreador de Gibbbs. Además la ecuación (3.21) define una ecuación integral de punto fijo para la cual $f_X(x)$ es solución. La unicidad de la solución se explica en [**?**].

Además de que la ecuación 3.21 coloca a la distribución marginal de X en términos de las condicionales. Un hecho que no se mostrara aquí pues su desarrollo excede el objetivo del presente trabajo, es que cuando $k \to \infty$ en (3.19), se tiene que

$$f_{X'_k|X'_0}\left(x|x_0\right) \to f_X\left(x\right) \quad \text{y} \quad f_{X'_k|X'_{k-1}}\left(x|t\right) \to h\left(x, t\right) \tag{3.22}$$

muestran que 3.21 es la forma limite de 3.19.

3.4. Evaluación de P^n

La ecuación 3.16 en la sección anterior motiva trabajar con la k-ésima potencia de la matriz de transición. Por otro lado, evaluar una matriz cuadrada digamos $n \times n$ con un exponente $k \geq 5$, es una trabajo bastante tedioso en especial si $n \geq 4$. Así pues, el objetivo de esta sección es presentar la matriz de transición para una cadena de Markov con espacio de estados finito, cuando es evaluada para alguna potencia en los enteros mayores que cero. Se ha decidido seguir el texto [2], para el presente desarrollo.

Sea P la matriz de transición para una cadena de Markov con k estados. Supongamos que los autovalores de P son distintos entre sí. Además la ecuación

$$P\boldsymbol{x} = \lambda \boldsymbol{x} \qquad \text{y} \qquad \boldsymbol{y}^{'} P = \lambda \boldsymbol{y}^{'}$$

tiene solución $\boldsymbol{x}$ y $\boldsymbol{y}$ distinta de cero si y sólo si λ es una autovalor de P. Se sigue que para cada λ_j, podemos encontrar los correspondientes autovectores a derecha y a izquierda $\boldsymbol{x}_{\boldsymbol{j}}^{'}$ y $\boldsymbol{y}_{\boldsymbol{j}}^{'}$.

Ahora si definimos $\boldsymbol{H} = [\boldsymbol{x_1}, \boldsymbol{x_2}, \cdots, \boldsymbol{x_k}]$ y $\boldsymbol{K} = [\boldsymbol{y_1}, \boldsymbol{y_2}, \cdots, \boldsymbol{y_k}]$, es posible mediante algunos cálculos con estos elementos, llegar a caracterizar la matriz P^n como

$$P^n = \sum_{j=1}^{k} \lambda_j^n A_j,$$

dado que $\boldsymbol{y}_{\boldsymbol{j}}^{'} \boldsymbol{x_j} = 1$ y A_j definido como $\boldsymbol{x_j} \boldsymbol{y}_{\boldsymbol{j}}^{'}$. Sin embargo en [2], también se considera el caso donde no se ajusta la condición $\boldsymbol{y_j} \boldsymbol{x_j} = 1$ sino $\boldsymbol{y}_{\boldsymbol{j}}^{'} \boldsymbol{x_j} = c_j$, con el fin de obtener

$$P^n = \sum_{j=1}^{k} c_j^{-1} \lambda_j^n A_j.$$

También es posible presentar la matriz P^n cuando P tiene multiples autovalores, dada por,

$$\boldsymbol{P^n} = \sum_j \frac{1}{(r_j - 1)!} \left[\frac{d^{r_j - 1}}{d\lambda^{r_j - 1}} \frac{\lambda^n adj\left(\lambda \boldsymbol{I} - \boldsymbol{P}\right)}{\psi_j\left(\lambda\right)} \right]_{\lambda = \lambda_j}$$

donde $\psi_j\left(\lambda\right) = \prod_{i \neq j} \left(\lambda - \lambda_i\right)^{r_i}$ y r_j es la multiplicidad de λ_j (Frazer, R.A., Duncan, W.J., AND Collar, A.R., 1946).

Conclusiones

- El muestreador de Gibbs es un algoritmo versátil para simular las distribuciones marginales de una distribución conjunta. Además dado que la mecánica que lo describe es sencilla su aplicación en software como R no presenta mayor dificultad y permite apreciar el buen comportamiento que tiene la distribución simulada mediante el muestreador de Gibbs a la distribución teórica. Por otro lado la simulación desarrollada también permite apreciar hechos como lo rápido que converge la matriz de transición a $A_{X|X}$ a la distribución marginal de la variable aleatoria X.

- Las cadenas de Markov hacen parte del soporte teórico del muestreador de Gibbs. Pues permiten construir la matriz de transición en k pasos para la variable aleatoria que se quiere simular, considerando los valores de la variable en la sucesión de Gibbs como una cadena de Markov. La importancia de construir dicha matriz es que su distribución estacionaria coincide con la que se quiere simular, como se mencionó y se ilustró por medio de ejemplos en el trabajo.

- El trabajo desarrollado también permite apreciar que el éxito del muestreador de Gibbs se basa en que la función marginal de X se puede escribir en términos de las distribuciones condicionales $X \mid Y$ y $Y \mid X$.

APÉNDICE

Códigos en R

Empleando el Muestreador de Gibbs para la distribución Beta-binomial

```
x<-c()
y<-c(0.4) ##Valor inicial para la sucesión de Gibbs##
##Algoritmo del Muestreador de Gibbs##
for(i in 1:100000){
       x[i] <- rbinom(1,10,y[i])
       y[i+1] <-rbeta(1,x[i]+2,15-x[i])
   }
###Sucesión de Gibbs####
datos<-table(x)/100000

####Distribución Teórica Beta-binomial
con parámetros n=10, alfa=2 y beta=5#######
dbb <- function(x){
       (choose(10,x)*30/factorial(16))*gamma(x+2)*gamma(15-x)
}
#### Evaluando la distribución Beta-binomial
con parámetros n=10, alfa=2 y beta=5#######
a <- 0:10
t <- dbb(a)
```

Prueba ji-cuadrado para la distribución Beta-binomial

```
x<-c()
y<-c(0.4) ##Valor inicial para la sucesión de Gibbs##
##Algoritmo del Muestreador de Gibbs##
for(i in 1:100000){
       x[i] <- rbinom(1,10,y[i])
       y[i+1] <-rbeta(1,x[i]+2,15-x[i])
   }

####Distribución Teórica Beta-binomial
con parámetros n=10, alfa=2 y beta=5#######
dbb <- function(x){
       (choose(10,x)*30/factorial(16))*gamma(x+2)*gamma(15-x)
}
#### Evaluando la distribución Beta-binomial
con parámetros n=10, alfa=2 y beta=5#######
a <- 0:10
t <- dbb(a)


###Cuadrado de bondad de ajuste####
frecbb <- as.vector(table(x))/100000
estad_chi <- sum((frecbb-t)^2/t)
pvalor <- p_valor<- pchisq(estad_chi,10,lower.tail=FALSE)
```

Empleando el Muestreador de Gibbs para la distribución Binomial

```
x<-c()
y<-c(2) ##Valor inicial para la sucesión de Gibbs##
##Algoritmo del Muestreador de Gibbs##
for(i in 1:200000){
x[i] <- rbinom(1,10-y[i],0.3/(1-0.2))
y[i+1] <-rbinom(1,10-x[i],0.2/(1-0.3))
}
###Sucesión de Gibbs####
datos2<-table(x)/200000
#### Evaluando la distribución Binomial #######
db <- function(x){
    choose(10,x)*((0.3)^(x))*((0.7)^(10-x))
```

```
}
l<-db(0:10)
```

Prueba ji-cuadrado para la distribución Binomial

```
x<-c()
y<-c(2) ##Valor inicial para la sucesión de Gibbs##
##Algoritmo del Muestreador de Gibbs##
for(i in 1:200000){
x[i] <- rbinom(1,10-y[i],0.3/(1-0.2))
y[i+1] <-rbinom(1,10-x[i],0.2/(1-0.3))
}
#### Evaluando la distribución Binomial #######
db <- function(x){
    choose(10,x)*((0.3)^(x))*((0.7)^(10-x))
}
l<-db(0:10)
###Cuadrado de bondad de ajuste####
frecb <- as.vector(table(x))/200000
estad_chi <- sum((frecb-l)^2/l)
pvalor <- p_valor<- pchisq(estad_chi,10,lower.tail=FALSE)
```

Código Ejemplo 1

```
install.packages("Matrix",dependencies=TRUE)
install.packages("expm",dependencies=TRUE)
#################################
#Distribución conjunta trinomial#
#################################
trinomial<- function(x,y,n,p,q){
if(x+y<=n)
(factorial(n)/(factorial(x)*factorial(y)*factorial(n-x-y)))
*p^x*q^y*(1-p-q)^(n-x-y)
else 0 }
#################################
#Matriz probabilidades trinomial#
#################################
conjunta <- function(n,p,q){
T <- matrix(nrow=n+1,ncol=n+1)
for(i in 0:n){
for(j in 0:n){
T[i+1,j+1] <- trinomial(i,j,n,p,q)
```

```
}
}
print(T)}
########################################################
#Obtención de la matriz de probabilidades de transición#
########################################################
##########
#Conjunta#
##########
fxy <- conjunta(3,0.5,0.3)
###############
#Marginal de x#
###############
fx <- apply(fxy,2,sum)
###############
#Marginal de y#
###############
fy <- apply(fxy,1,sum)
###################################################
#Marginal de X a partir de la matriz de transición#
###################################################
Ayx <- t(fxy)/fx
Axy <- fxy/fy
Ax <- Ayx%*%Axy
PAx<- Ax %^% 21
###################################################
#Marginal de Y a partir de la matriz de transición#
###################################################
Ay <- Axy%*%Ayx
PAy <- Ay %^% 21
```

Bibliografía

[1] Alan, G AND Smith, A. *Sampling-Based Approaches to Calculating Marginal Densities*. Journal of the American Statistical Association, 85(40), 398–409, 1990.

[2] Bailey, N. *The elements of Stochastic Processes with applications to the natural sciences*. United States of America: John Willey & Sons, Inc., 1964.

[3] Blanco, L. *Probabilidad*. Bogotá, Colombia: Unibiblos, 2004.

[4] Blanco, L., Arunachalam W., AND Dharmaraja, D. *Introduction to Probability and Stochastic Processes with Applications*. New Jersey, USA: John Wiley & Sons, Inc., 2012.

[5] Casella, G AND Edward, G. *Explaining the Gibbs Sampler*. The American Statistical Association, 46(3), 167–174, 1992.

[6] Frazer, R.A., Duncan, W.J., AND Collar, A.R. *Elementary Matrices*. Cambridge University Press, 1946.

[7] Rincón, L. *Curso intermedio de Probabilidad*. Mexicó DF, Mexicó: Departamento de Matemáticas Facultad de Ciencias UNAM, 2007.

[8] Rincón, L. *Introducción a los procesos estocásticos*. Mexicó DF, Mexicó: Departamento de Matemáticas Facultad de Ciencias UNAM, 2012.

[9] Ross, S. *Introduction to probability Models*. San Diego, United States of America: John Willey & Sons, Inc., 2010.

[10] Prieto, V. *VII Coloqio colombiano de matemáticas, Cadenas de Markov*. Bogotá, Colombia: Universidad Nacional de Colombia, 1977