# Convergencia del muestrador de Gibbs

Basado en Cassella, George (1992)

*Diego Uriarte*

*24/9/2018*

## ¿Cómo funciona el muestrador de Gibbs?

Ilustraremos con un ejemplo bivariado. Queremos muestras de $f(x)$ (distribuida de manera conjunta con $y$). Conocemos las condicionales $f(x|y)$ y $f(y|x)$. Se especifica un valor inicial para $Y_0' = y_0'$. Se obtiene apartir de las condicionales una sucesión de muestras para x e y:

$$X_j' \sim f(x|Y_j' = y_j')$$
$$Y_{j+1}' \sim f(y|X_j' = x_j')$$

Así, obtenemos una secuencia: $Y_0', X_0', Y_1', X_1', \cdots Y_k', X_k',$. Resulta que bajo algunas condiciones generales, cuando $k \to \infty$, $X_k'$ es efectivamente una muestra de $f(x)$

## Una prueba simple de la convergencia del muestrador de Gibbs (caso discreto)

Mostraremos como converge el muestreo tomando como ejemplo una distribución bivariada, donde X e Y se distribuyen Bernoulli de manera conjunta.

La distribución conjunta de X, Y, además de la marginal de X están dadas por:

$$\begin{bmatrix} f_{x,y}(0,0) & f_{x,y}(1,0) \\ f_{x,y}(0,1) & f_{x,y}(1,1) \end{bmatrix} = \begin{bmatrix} p_1 & p_2 \\ p_3 & p_4 \end{bmatrix} |f_x = \begin{bmatrix} p_1 + p_3 & p_2 + p_4 \end{bmatrix}$$

Determinamos las probabilidades condicionales $X|Y = y$ e $Y|X = x$, formulando las siguientes matrices:

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

Dado que estamos interesados en muestrar de la marginal de X, nos interesan los valores $X_0', X_1', X_2', \cdots$. Sin embargo, para ir de $X_0' \to X_1'$ debemos muestrar $Y_1'$, es decir, $X_0' \to Y_1' \to X_1'$.

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

Para este caso en particular, se tiene que la matrices de transición para x está dada por $[A_{x|x} = A_{y|x} A_{x|y}$:

Ahora, es directo calcular la distribución marginal de realización k, $X_k'$:

$f_k = f_0 A_{x|x}^k = (f_0 A_{x|x}^{k-1}) A_{x|x} = f_{k-1} A_{x|x}$

Bajo ciertas condiciones (matriz con elementos positivos), se tiene un punto fijo tal que: $f = f A_{x|x}$:

En nuestro caso simple:

$$\begin{aligned} f_x A_{x|x} = f_x A_{y|x} A_{x|y} &= \begin{bmatrix} p_1 + p_3 & p_2 + p_4 \end{bmatrix} \begin{bmatrix} \frac{p_1}{p_1+p_3} & \frac{p_3}{p_1+p_3} \\ \frac{p_2}{p_2+p_4} & \frac{p_4}{p_2+p_4} \end{bmatrix} \begin{bmatrix} \frac{p_1}{p_1+p_2} & \frac{p_2}{p_1+p_2} \\ \frac{p_3}{p_3+p_4} & \frac{p_4}{p_3+p_4} \end{bmatrix} \\ &= \begin{bmatrix} p_1 + p_3 & p_2 + p_4 \end{bmatrix} = f_x \end{aligned}$$

### Intuición para el caso bivariado (continuo)

Partamos que conocemos $f_{X|Y}(x|y)$ y $f_{Y|X}(y|x)$. Si queremos la marginal, por definición: $f_X(x) = \int f_{xy}(x,y)dy$

Por Bayes sabemos que: $f_{xy}(x,y) = f_{X|Y}(x|y)f_Y(y)$, por tanto: $f_X(x) = \int f_{X|Y}(x|y)f_Y(y)dy$

De manera similar, tenemos $f_{xy}(x,y) = f_{Y|X}(y|x)f_X(y)$ con $f_Y(y) = \int f_{Y|X}(y|x)f_X(x)dx$.

Reemplazando:

$$f_X(x) = \int f_{X|Y}(x|y) \int f_{Y|X}(y|t)f_X(t)dtdy = \int \left[\int f_{X|Y}(x|y)f_{Y|X}(y|t)dy\right] f_X(t)dt = \int h(x,t)f_X(t)dt$$

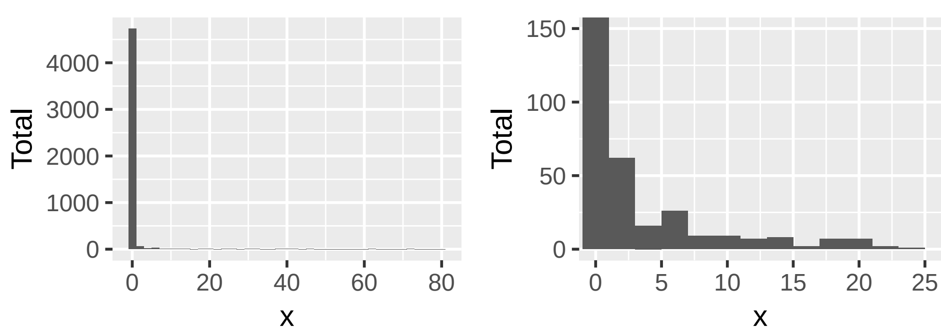
con $h(x,t) = \left[\int f_{X|Y}(x|y)f_{Y|X}(y|t)dy\right]$. La ecuación con $f_x$ se conoce como una ecuación de punto fijo integral y la secuencia de Gibbs es una manera de converger a la solución de dicha ecuación.

Sin embargo, el algoritmo de Gibbs no siempre nos lleva a la marginal que deseamos. Veamos el siguiente ejemplo de distribuciones condicionales exponenciales definidas en $(0, \infty)$.

$$f(x|y) = ye^{-yx}, 0 < x < \infty \quad | \quad f(y|x) = xe^{-xy}, 0 < y < \infty$$

$$f_X(x) = \int \left[\int y \exp^{-yx} t \exp^{-ty} dy\right] f_X(t)dt = \int \left[\frac{t}{(x+t)^2}\right] f_X(t)dt$$

Vemos que $f_X(t) = 1/t$ resuelve la ecuación (reemplazando). Pero cuando intentamos aplicar el muestrador de Gibbs a estas condicionales, no obtenemos convergencia (es decir, una aproximación a $1/x$), la mayor parte de los valores se acumula cerca 0:




Histograma completo

Zoom en el eje y

Podemos evitar esto si truncamos las distribuciones condicionales de manera que se cumpla:

$$\int_0^B f_X(x)dx < \infty$$

Por contraposición, en el caso sin truncar la marginal no integra a un número finito: $\int_0^\infty \frac{1}{x}dx \to \infty$

Una condición para que algoritmo converga es que la marginal sea una función de densidad $\int f_X(x)dx < \infty$. Una manera de segurar esto es limitando las condicionales.