Clase práctica 8

Importance sampling

Problema 1

1. Utilice el algoritmo de Metropolis para muestrear la distribución normal en una dimensión. Estudie el porcentaje de aceptación de movimientos y la función correlación en función del tamaño del paso de exploración.

2. Use las variables aleatorias generadas para calcular,

$$I = \int_{-\infty}^{\infty} dx\, x^2\, e^{-x^2/2} \tag{8.1}$$

Estudie la dependencia del resultado con el tiempo de termalización y con la posición de inicio del sampleo.

Solución

En las primeras clases se dio un algoritmo para obtener números pseudo-aleatorios distribuidos uniformemente en (0,1). Ahora queremos obtener una secuencia, $x_0, x_1, ..., x_i, ...$ de valores distribuidos normalmente con media $\mu = 0$ y desviación $\sigma = 1$, es decir, con distribución,

$$f(x_i) = \frac{1}{\sqrt{2\pi}}\, e^{-x_i^2/2} \tag{8.2}$$

Lo primero que se nos ocurre es transformar la variable $u \sim \mathrm{Unif}(0,1)$ por medio de alguna función matemática $g(u)$, de manera que los valores $y = g(u)$ se distribuyan normalemente. Tenemos, entonces, que encontrar $g(u)$.

$y = g(u)$ establece una relación bi-unívoca entre u e y. Por lo tanto, la probbilidad de ocurrencia de ambas debe ser la misma,

$$f_y(y)\, dy = f_u(u)\, du \qquad , \qquad f_y(y) = \frac{f_u(u)}{dy/du} \tag{8.3}$$

donde $f_u(u) = 1$, por ser $\mathrm{Unif}(0,1)$. Resolviendo la Ec. (8.3) podemos identificar,

$$\frac{dF_y(y)}{dy} = f_y(y) = \frac{du}{dy} = \Big(g^{-1}(y)\Big)' \qquad , \qquad F_y^{-1}(u) = g(u) \tag{8.4}$$

Este es el método de *inversión* y es válido para cualquier distribución $f_y(y)$ en la que existe una relación bi-unívoca $y = g(u)$. Tiene el inconveniente de que requiere de una expresión explícita de F_y^{-1}. En el caso de la distribución normal, no la tenemos.

Debemos buscar un método que sólo requiera de la distribución de densidad $f_y(y)$. Por ejemplo, si tomamos valores $x_0, x_1, ..., x_i, ...$ y calculamos para cada uno las probabilidades,

$$p_i = f_x(x_i)\, dx \cong \frac{e^{-x_i^2/2}\, dx}{\sum_{j=1}^{N} e^{-x_j^2/2}\, dx} = \frac{1}{\sum_{j=1}^{N} e^{-(x_j^2 - x_i^2)/2}} \tag{8.5}$$

tendremos un conjunto $\{p_i\}, i = 1...N$ de probabilidades aproximadamente normales. Luego, podemos elegir aleatoriamente cualquier valor de i y aceptar x_i con probabilidad p_i. La probabilidad final de aceptación será $p^*.p_i$, donde $p^* = 1/N$ es la probabilidad del *trial*. Ahora, la secuencia formada por los valores aceptados $y_0, y_1, ..., y_i, ...$ se distribuye normalmente.

Este método es aproximado cuando $N \to \infty'$. Esto exige mucho esfuerzo computacional en el denominador de la Ec. (8.5). Por otro lado, la elección de los valores i se realiza aleatoriamente. También podría hacerse un barrido cíclico, pero en cualquiera de los dos casos, no es una elección eficiente, porque pasaríamos mucho tiempo muestreando regiones que no nos interesan.

Podemos mejorar el método si observamos el cociente de la Ec. (8.5). El denominador puede interpretarse como la suma de todas las probabilidades condicionadas f_j,

$$p_i = \frac{1}{N}\frac{e^{-x_i^2/2}}{\sum_{j=1}^{N} e^{-x_j^2/2}\,\frac{1}{N}} = \frac{1}{N}\frac{f_i}{\sum_{j=1}^{N} f_j\left(\frac{1}{N}\right)} \tag{8.6}$$

donde son todas equiprobables y deben ir multiplicadas por $1/N$. Es decir, al muestrear aleatoriamente, cualquier f_j puede ocurrir antes que f_i, y ocurre con probabilidad $1/N$. Pero si muestreamos de manera distinta, alterando la probabilidad $1/N$ por,

$$\begin{cases} 1 & \text{si } f_j \rightarrow f_i \\ 0 & \text{caso contrario} \end{cases} \tag{8.7}$$

no tenemos que calcular la sumatoria. Ahora tenemos, $p_i = e^{-x_i^2/2}/e^{-x_j^2/2}$.

Este tipo de muestreo se implementa con un caminador aleatorio. Es decir, si parto del valor x_j, elijo x_i dentro de la distancia $|x_i - x_j| \leq \delta$ (donde δ es el tamaño del paso de exploración). Este es en síntesis el algoritmo de Metropolis. Su expresión completa es,

$$P_{i,i+1} = \begin{cases} p^* \dfrac{\omega(x_{i+1})}{\omega(x_i)} & \text{si } \omega(x_{i+1}) \leq \omega(x_i) \\ p^* & \text{si } \omega(x_{i+1}) > \omega(x_i) \end{cases} \tag{8.8}$$

$$p^* = \begin{cases} 1/2\delta & \text{si } |x_{i+1} - x_i| \leq \delta \\ 0 & \text{si } |x_{i+1} - x_i| > \delta \end{cases} \tag{8.9}$$

donde $\omega(x) = e^{-x^2/2}$.

¿Qué valor de δ usamos para optimizar el muestreo?

Por un lado, queremos muestrear la mayor región posible en el menor tiempo, así que pensamos en un δ grande. Pero por otro lado, no queremos perder detalle de las zonas que más nos interesan. Pensamos en un δ chico.

En realidad, si el paso es muy grande, $P_{i,i+1}$ será chico y la mayoría de los pasos serán rechazados. Por el contrario, si δ es chico, $\omega(x_{i+1}) \approx \omega(x_i)$ y las muestras estarán muy correlacionadas. La exploración se vuelve muy lenta. En general, se considera que la probabilidad de aceptación de pasos debe ubicarse entre 0,3 y 0,6. Preferentemente, pedimos que la mitad de los pasos sean aceptados.

Investiguemos qué valor de δ corresponde a una probabilidad de aceptación de 1/2. Si $\omega(x_{i+1}) \leq \omega(x_i)$, la probabilidad de aceptación será $p^* \times \omega(x_{i+1})/\omega(x_i)$, mientras que para $\omega(x_{i+1}) > \omega(x_i)$ será sólo de p^*. Por lo tanto, debemos hallar con qué probabilidad ocurren ambas situaciones y la expresión de $f(x_{i+1}, x_i)$.

Ejercicio: determinación de $f(\omega(x_{i+1})/\omega(x_i))$

Encuentre la expresión para función densidad de distribución $f(\omega(x_{i+1})/\omega(x_i))$. Obtenga la probabilidad de que $f(\omega(x_{i+1})/\omega(x_i)) \leq 1$.

Queremos hallar la distribución de la expresión $e^{-(x_{i+1}^2 - x_i^2)/2}$, suponiendo que x_{i+1} y x_i tienen una distribución normal y está correlacionadas con correlación ρ.

Primero observamos que $x_{i+1}^2 - x_i^2 = (x_{i+1} - x_i)(x_{i+1} + x_i) = y_1.y_2$. Las variables $y_1 = x_{i+1} - x_i$ e $y_2 = x_{i+1} + x_i$ representan una transformación lineal respecto de x_{i+1} y x_i. Si (x_{i+1}, x_i) tienen una distribución normal conjunta (recordemos que *no* son independientes entre sí),

$$f(x_{i+1}, x_i) = \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} \exp\Big(- \frac{1}{2\sigma^2(1-\rho^2)}(x_{i+1}^2 + x_i^2 - 2\rho x_{i+1} x_i)\Big) \tag{8.10}$$

y se relacionan con (y_1, y_2) por medio de la función $g(x_{i+1}, x_i)$, entonces la distribución de (y_1, y_2) viene dada por,

$$f_y(y_1, y_2) = f_x(g^{-1}(y_1, y_2)) \times |J(g^{-1}(y_1, y_2))| \tag{8.11}$$

donde $|J(g^{-1}(y_1, y_2))|$ es el módulo del determinante jacobiano (en este caso, vale 1/2). Haciendo cuentas, se obtiene,

$$f_y(y_1, y_2) = \frac{1}{2\pi(2\sigma^2)\sqrt{1-\rho^2}} \exp\Big[- \frac{1}{2(2\sigma^2)(1-\rho^2)}(y_1^2(1-\rho) + y_2^2(1+\rho))\Big] \tag{8.12}$$

Esta expresión muestra que y_1 e y_2 son independientes entre sí. Cada una se distribuye según $y_1 \sim N(0, \sigma\sqrt{2(1-\rho)})$ e $y_2 \sim N(0, \sigma\sqrt{2(1+\rho)})$. La distribución del producto de dos variables aleatorias independientes, de media nula y desviaciones σ_1 y σ_2 respectivamente es [1] ,

$$f_y(y_1.y_2) = \frac{1}{\pi\sigma_1\sigma_2} K_0\Big(\frac{|y_1|}{\sigma_1}.\frac{|y_2|}{\sigma_2}\Big) \tag{8.13}$$

donde $K_0(z)$ es la función de Bessel modificada de segunda especie y grado cero. Esta distribución está muy concentrada entorno de $y_1.y_2 = 0$. Además es simétrica respecto del cero por lo que la probabilidad de que $y_1.y_2 < 0$ es $1/2$.

La distribución $f(\omega(x_{i+1})/\omega(x_i))$ se obtiene fácilmente, transformando la Ec. (8.13) por mendio de la función $g(y_1.y_2) = \exp(-y_1.y_2/2)$ y calculando su jacobiano. Por ejemplo, si $g(y_1.y_2) = \exp(-y_1.y_2/2)$, obtenemos,

$$f\left(e^{-(x_{i+1}^2 - x_i^2)/2}\right) = \frac{2}{\pi\sigma_1\sigma_2|z|} K_0\Big(\frac{2|\ln z|}{\sigma_1\sigma_2}\Big) \tag{8.14}$$

donde $z = \exp(-y_1.y_2/2)$.

La probabilidad de aceptación es,

$$\langle P_a\rangle = \frac{1}{2}p^*.\Big\langle\frac{\omega(x_{i+1})}{\omega(x_i)}\Big\rangle + \frac{1}{2}p^* \tag{8.15}$$

donde el valor $1/2$ se agregó debido a que los casos $y_1.y_2 < 0$ e $y_1.y_2 > 0$ son equiprobables, como se mostró más arriba. Si asuminos que $\langle\omega(x_{i+1})/\omega(x_i)\rangle \approx \exp(\langle -y_1.y_2/2\rangle)$, se obtiene que,

$$\langle -y_1.y_2/2\rangle = -\frac{1}{2}\,\langle y_1.y_2\rangle = -\frac{1}{2}\frac{\int_0^\infty \frac{u}{\pi\sigma_1\sigma_2}K_0\Big(\frac{|u|}{\sigma_1\sigma_2}\Big)\,du}{1/2} = -\sigma_1\sigma_2\int_0^\infty \frac{t}{\pi}K_0(t)\,dt \tag{8.16}$$

donde la última integral da $1/\pi$ y $\sigma_1\sigma_2 = 2\sigma^2\sqrt{1-\rho^2}$ (ver Abramowitz). En el caso particular en que $\sigma = 1$ y $\rho = 0$ resulta,

[1] ver Weisstein, Eric W. *Normal Product Distribution.*. From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/NormalProductDistribution.html

$$\langle P_a \rangle \approx \frac{1}{4\delta} \, e^{-2/\pi} + \frac{1}{4\delta} \tag{8.17}$$

Si pedimos que $\langle P_a \rangle = 1/2$ resulta $\delta = 0,76$. Para el caso $\sigma = 1$ y $\rho = 1$ resulta $\delta = 1$. Observamos que el tamaño del paso depende de la correlación, si queremos mantener el nivel de aceptación constante. Cuanto más correlacionadas estén las muestras, mayor debe ser el paso porque sino el algoritmo comienza a aceptar todos los movimientos.