Generación de observaciones aleatorias para una distribución de probabilidad.

Dada una sucesión de números aleatorios, ¿cómo se puede generar una sucesión de observaciones aleatorias que sigan una distribución dada?.

El ingrediente básico para cada método de generación de variables aleatorias de cualquier distribución o proceso aleatorio es una fuente de variables U(0,1) *iid*. Es necesario un generador de números pseudoaleatorios U(0,1).

Existen muchas técnicas para generar variables aleatorias y los algoritmos usados dependen, claro está, de la distribución que se desea generar. Sin embargo, tales técnicas pueden ser clasificadas de acuerdo a su base teórica (enfoques generales)[1]:

1. Transformada inversa.
2. Composición
3. Convolución.
4. Aceptación-Rechazo.

Transformada inversa.

Sea X la variable aleatoria (v.a.) bajo estudio y sea $F(x) = P\{X \le x\}$ la función de distribución acumulada. Si X es una v.a. continua, entonces $F(x)$ es continua y estrictamente creciente con $0 < F(x) < 1$, es decir, si $x_1 < x_2 \Rightarrow 0 < F(x_1) \le F(x_2) < 1$.

Si X es continua, $\exists f(x)$ no negativa tal que $P(x \in B) = \int_B f(x)dx$ y $\int_{-\infty}^{\infty} f(x)dx = 1$.

$$F(x) = P(X \in (-\infty, x]) = \int_{-\infty}^{x} f(y)dy; \quad \forall -\infty < x < \infty$$

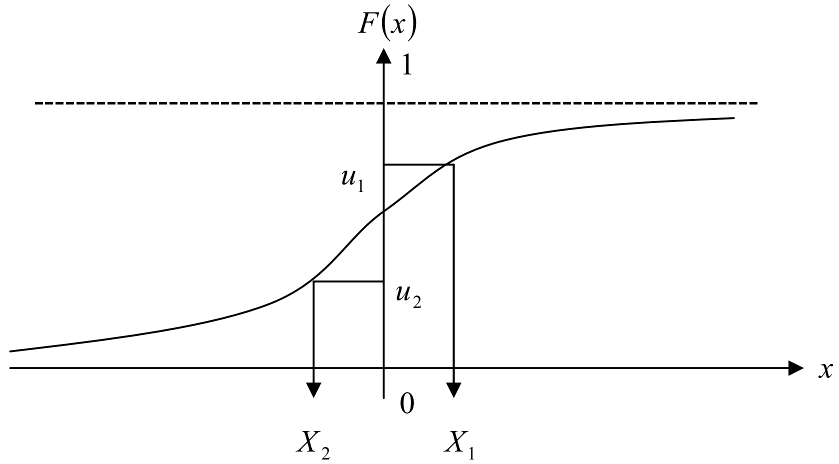
$$\boxed{f(x) = F'(x); \quad P(X \in I = [a,b]) = \int_a^b f(y)dy = F(b) - F(a)}$$ [2]

El *algoritmo* para generar una variable $\boldsymbol{x}$ (continua) que tiene a $\boldsymbol{F}$ como función de distribución es el siguiente:

1. Generar un número aleatorio $u \sim$ U(0,1).
2. Retornar $X = F^{-1}(u)$ (se establece $F(X) = u$ y se despeja X, que es entonces la observación aleatoria a partir de la distribución de probabilidad deseada).

($F^{-1}(u)$ siempre estará definida debido a que $0 \le u \le 1$ y el rango de F es [0,1])

[1] *A. M. Law, W. Kelton,* "Simulation Modeling and Analysis", NewYork: McGraw-Hill, 1991.
[2] Teorema fundamental del cálculo.



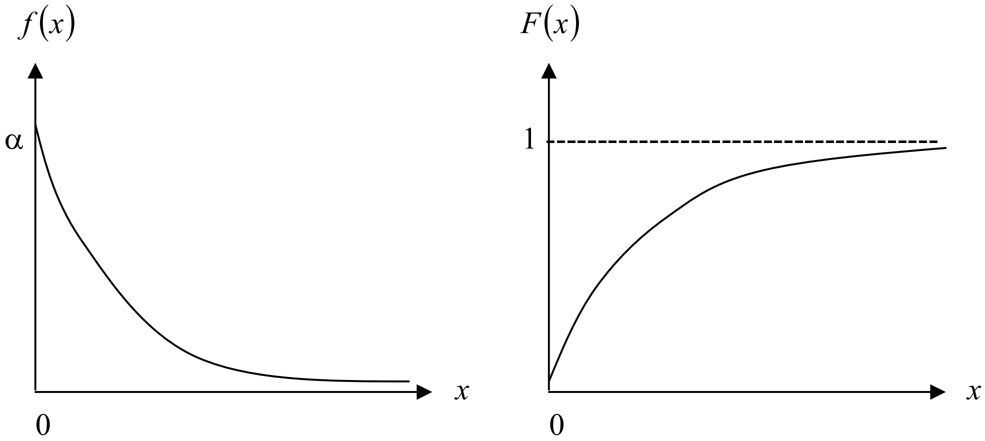
Ejemplo:

Sea T exponencial con parámetro α. $E(T) = \dfrac{1}{\alpha}$ $\quad Var(T) = \dfrac{1}{\alpha^2}$

$$f_T(t) = \begin{cases} \alpha e^{-\alpha t}, & t \geq 0 \\ 0, & t < 0 \end{cases}$$

$$f_T(t) = \begin{cases} \dfrac{1}{\beta} e^{-t/\beta}, & t \geq 0 \\ 0, & t < 0 \end{cases}$$ (otra forma de la misma distribución, $E(T) = \beta \quad Var(T) = \beta^2$)

$$F(x) = \int_0^x \alpha e^{-\alpha t} dt = -e^{-\alpha t}\Big|_0^x = -e^{-\alpha x} + 1 \Rightarrow F(x) = \begin{cases} 1 - e^{-\alpha x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$



$$1-e^{-\alpha x}=u \Leftrightarrow e^{-\alpha x}=1-u \Leftrightarrow \ln\left(e^{-\alpha x}\right)=\ln(1-u) \Leftrightarrow -\alpha x=\ln(1-u) \Leftrightarrow x=-\frac{1}{\alpha}\ln(1-u)$$

$$\therefore F^{-1}(u)=-\frac{1}{\alpha}\ln(1-u) \qquad \text{ó} \qquad F^{-1}(u)=-\beta\ln(1-u)$$

Algoritmo:

1. Se genera $u \sim$ U(0,1).
2. Se obtiene $x=-\frac{1}{\alpha}\ln(u)$.

Nota: $(1-u)$ y u tienen la misma distribución U(0,1).

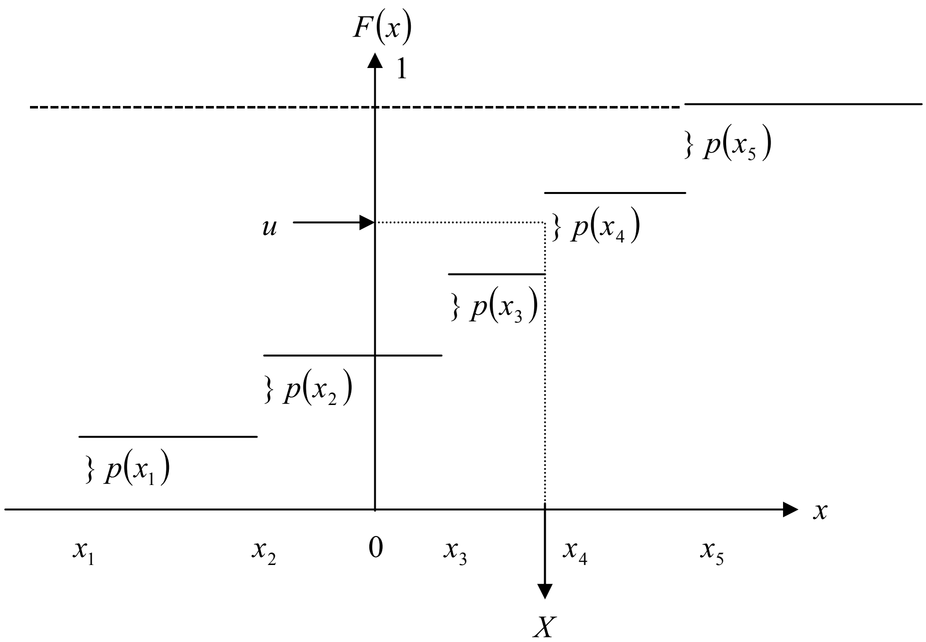
Si X es discreta, $p(x_i)=P(X=x_i)$; $i=1,2,\ldots$ con $\sum_{i=1}^{\infty} p(x_i)=1$.

$p(x)=$ función de masa de probabilidad para X.

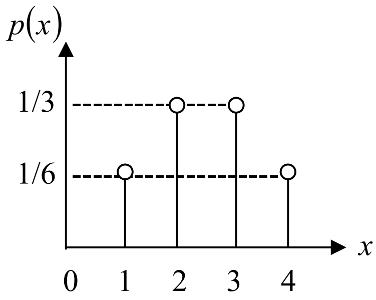
$$\therefore F(x)=\sum_{x_i \le x} p(x_i) \quad \forall -\infty < x < \infty$$

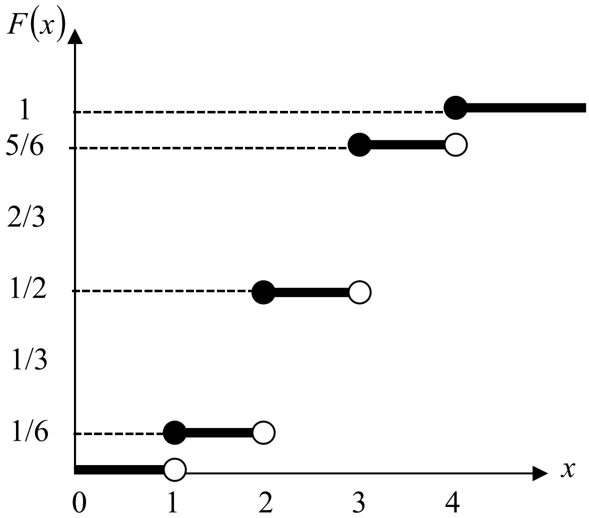
El *algoritmo* para generar una variable $\boldsymbol{x}$ (discreta) que tiene a $\boldsymbol{F}$ como función de distribución es el siguiente:

1. Generar un número aleatorio $u \sim$ U(0,1).
2. Determinar el entero positivo más pequeño I tal que $u \le F(x_I)$ y retornar $X=x_I$.



Ejemplo: (inventario)





$$P(2 \le x \le 3) = p(2) + p(3) = \frac{1}{3} + \frac{1}{3} = \frac{2}{3}$$

Composición.

Tal técnica aplica cuando la función de distribución *F* (de la variable que deseamos generar) puede ser expresada como una combinación convexa de otras funciones de distribución $F_1, F_2, F_3, \ldots$ y cuando se espera obtener muestras más fácilmente de F_j que de la original *F*.

$$F(x) = \sum_{j=1}^{\infty} \lambda_j F_j(x) \text{ donde } \lambda_j \ge 0, \sum_{j=1}^{\infty} \lambda_j = 1$$

(para algún *k*, $\lambda_k > 0$ pero $\lambda_j = 0$ para $j > k \Rightarrow$ suma finita)

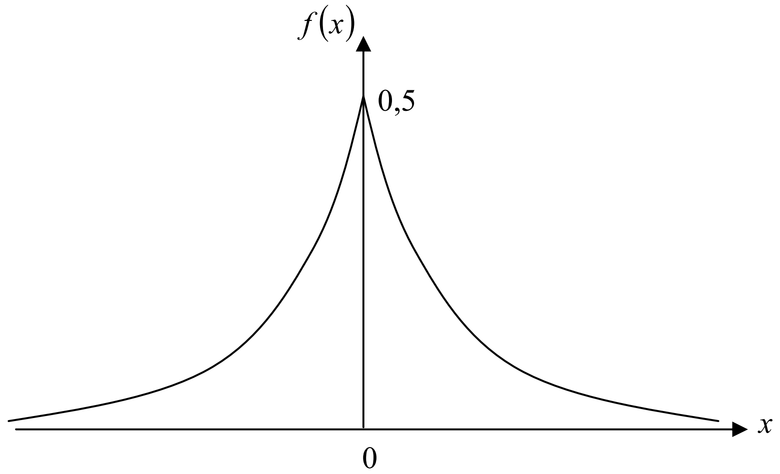
Equivalentemente, si *X* tiene una función de densidad *f* y $f(x) = \sum_{j=1}^{\infty} \lambda_j f_j(x)$ donde las f_j's son otras densidades, el método aplica.

Algoritmo:

1. Generar un entero aleatorio positivo *J* tal que $P(J = j) = \lambda_j ; j = 1,2,\ldots$
2. Retornar *X* con función de distribución F_J.

Ejemplo: Distribución doble exponencial o distribución de Laplace.

$$f(x) = 0{,}5e^{-|x|} \text{ (densidad)}$$



$$f(x) = 0{,}5e^{x} I_{(-\infty,0)}(x) + 0{,}5e^{-x} I_{[0,\infty)}(x) \text{ (densidad reescrita)}$$

$$I_A(x) = \begin{cases} 1, & x \in A \\ 0, & x \notin A \end{cases} \text{ (función indicador)}$$

f es una combinación convexa de $f_1(x) = e^{x} I_{(-\infty,0)}(x)$ y $f_2(x) = e^{-x} I_{[0,\infty)}(x)$ (densidades), con $p_1 = p_2 = 0{,}5(= \lambda_i 's)$

Algoritmo:

1. Generar $u_1, u_2 \sim$ U(0,1) *iid*.
2. Si $u_1 \leq 0{,}5$, retornar $X = \ln(u_2)$.
3. Si $u_1 > 0{,}5$, retornar $X = -\ln(u_2)$.

Convolución.

Para algunas distribuciones, la variable aleatoria deseada *X* puede ser escrita como suma de otras variables aleatorias que son *iid* y que pueden ser generadas más fácilmente que la generación directa de *X*. Asumimos que $Y_1, Y_2, \ldots, Y_m$ (*m* fijo) son *iid* tal que $Y_1 + Y_2 + \ldots + Y_m$ tiene la misma distribución de *X*. En procesos estocásticos se dice que la distribución de *X* es la "*m-fold convolution*" de la distribución de Y_j.

Algoritmo: Sean *F* y *G* las funciones de distribución de *X* e Y_j.

1. Generar $Y_1, Y_2, \ldots, Y_m$ *iid* cada una con una función de distribución *G*.
2. Retornar $X = Y_1 + Y_2 + \ldots + Y_m$.

Ejemplo: Binomial.

$X \sim bin$(t,p), $X = Y_1 + Y_2 + \ldots + Y_t$ con $Y_j \sim Bernoulli$(p)

Algoritmo:

1. Generar $u \sim$ U(0,1).
2. Si $u \le p$ retornar $X = 1$, sino retornar $X = 0$.

Ejemplo: m-Erlang.

$$X = Y_1 + Y_2 + \ldots + Y_m,\ Y_j \sim Exponencial(\beta/m).$$

Nota: Si se desea generar una uniforme $u \sim$ U(a,b), se procede de la siguiente manera:

1. Generar $u \sim$ U(0,1).
2. Retornar $X = a + (b - a)u$.

Aceptación-Rechazo.[3]

Procedimiento general para la generación de variables aleatorias correspondientes a cualquier función de densidad $f(x)$, continua y acotada, es decir, $0 \le f(x) \le f_{\max}$ en $a \le x \le b$.

Algoritmo:

1. Generar un par de números aleatorios uniformemente distribuidos en el intervalo (0,1), es decir, $u_1, u_2 \sim$ U(0,1).

2. Obtener una variable uniforme z en $a \le x \le b$ usando, $z = a + (b - a)u_1$.

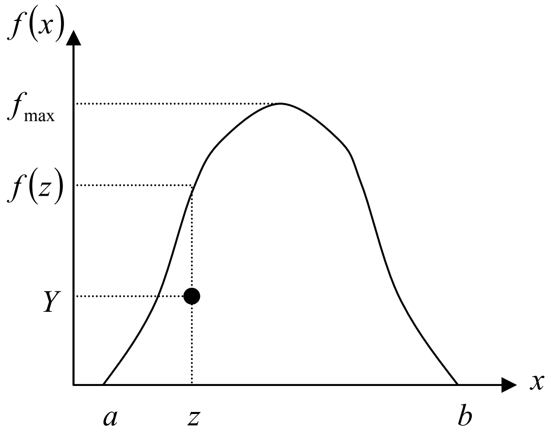
3. Determinar (evaluar) $f(z)$.

4. Obtener una variable uniforme Y en el intervalo $0 \le Y \le f_{\max}$ usando la relación $Y = f_{\max} u_2$.

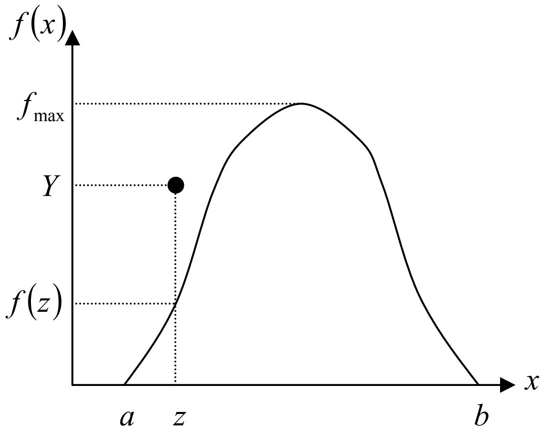
5. Comparar Y con $f(z)$.
   - Si $Y \le f(z) \Rightarrow (Y, z)$ cae dentro del área bajo f o sobre f. Luego, se acepta z como la variable deseada, $X = z$.
   - Si $Y > f(z) \Rightarrow (Y, z)$ está por encima de f. Luego, se rechaza z y se regresa al paso 1.

6. Se repiten los pasos 1 al 5 sucesivamente hasta encontrar la aceptación en el paso 5.

[3] *Byron S. Gottfried,* "Elements of Stochastic Process Simulation", Prentice Hall Inc., Englewood Cliffs, New Jersey, 1984.



$Y \leq f(z)$
(aceptación de z, $X = z$)



$Y > f(z)$
(rechazo de z)