Generación de números aleatorios

Dr. Misael Erikson Maguiña Palma

Distribuciones Discretas

Distribución	Nombre en R
Binomial	binom
Poisson	pois
Geométrica	geom
Hipergeométrica	hyper
Binomial Negativa	nbinom

d: función de densidad o de probabilidad.
p: función de distribución
q: función para el cálculo de cuantiles.
r: función para simular datos con dicha distribución

Distribuciones Continuas

Distribución	Nombre en R
Uniforme	unif
Normal	norm
t Student	t
F Fisher	F
Chi-Cuadrado	chisq
Exponencial	exp
Gamma	gamma
Weibull	weibull
W de Wilcoxon	wilcox

GENERACIÓN DE VARIABLES ALEATORIAS

Los principales métodos para generar las variables aleatorias son:

- Método de la transformada inversa
- Método de convolución
- Métodos de composición

Método de la transformada inversa

Se trataría del método preferible para la simulación de una variable continua (siempre que se disponga de la función cuantil). Está basado en los siguientes resultados:

Si X es una variable aleatoria con función de distribución F continua y estrictamente monótona (invertible), entonces:

$$U = F(X) \sim \mathcal{U}(0,1),$$

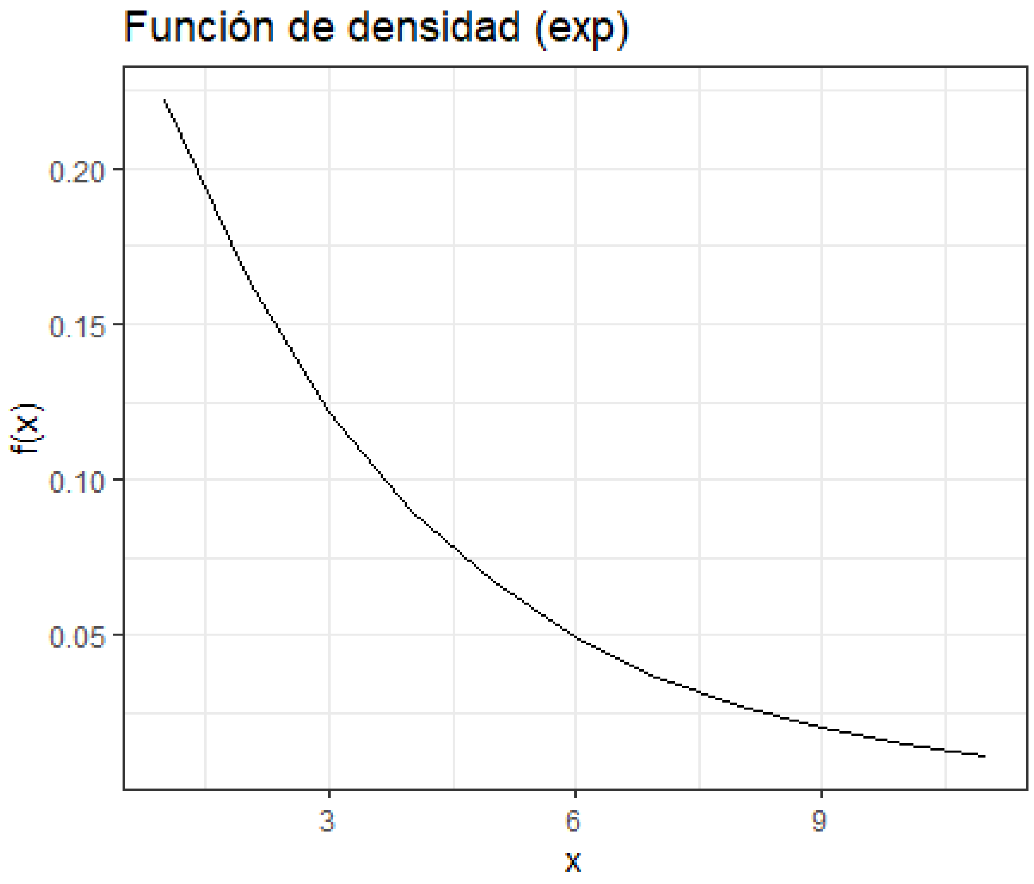
Ya que

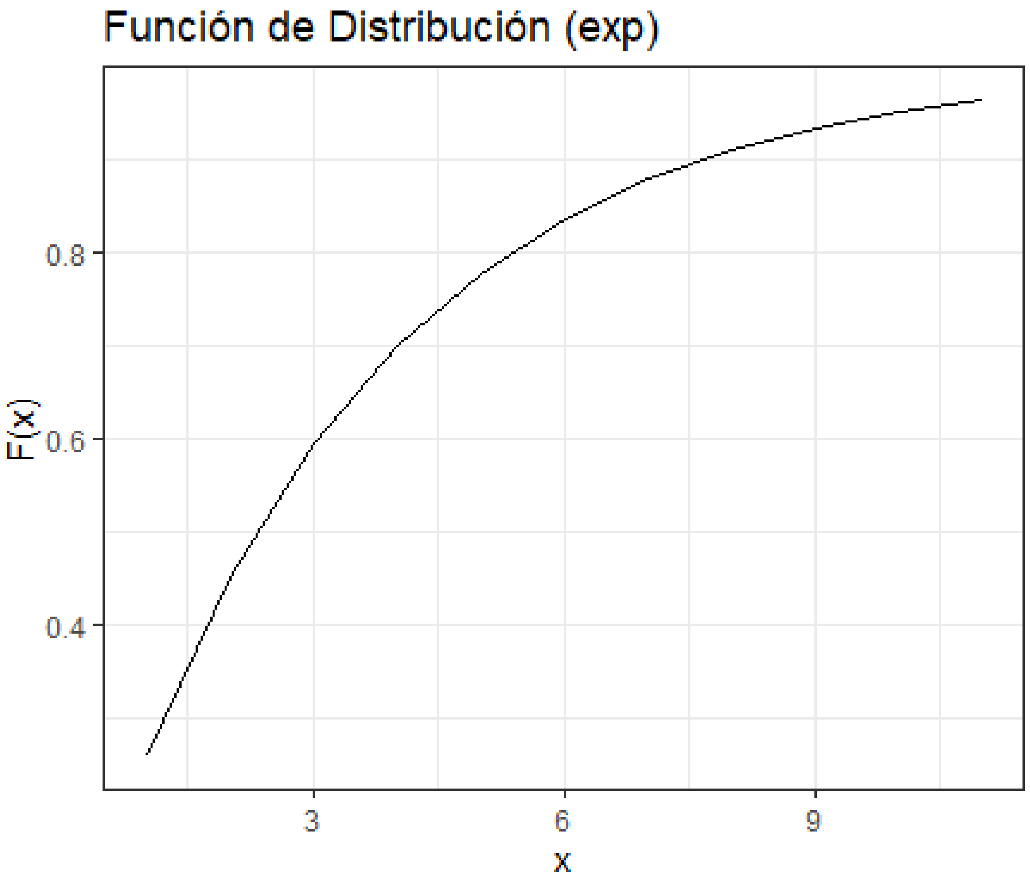
$$G(u) = P(Y \leq u) = P(F(X) \leq u) = P\left(X \leq F^{-1}(u)\right) = F\left(F^{-1}(u)\right) = u.$$

El reciprocó también es cierto. si $U \sim \mathcal{U}(0,1)$ entonces

$$F^{-1}(U) \sim X$$

Método de la transformada inversa





Algunas distribuciones que pueden simularse por el método de inversión

Nombre	Densidad	$F(x)$	$F^{-1}(U)$	Forma simplificada
$\exp(\lambda)$ $(\lambda > 0)$	$\lambda e^{-\lambda x}$, si $x \geq 0$	$1 - e^{-\lambda x}$	$-\dfrac{\ln(1-U)}{\lambda}$	$-\dfrac{\ln U}{\lambda}$
Cauchy	$\dfrac{1}{\pi(1+x^2)}$	$\dfrac{1}{2} + \dfrac{\arctan x}{\pi}$	$\tan\left(\pi\left(U - \dfrac{1}{2}\right)\right)$	$\tan \pi U$
Triangular en $(0, a)$	$\dfrac{2}{a}\left(1 - \dfrac{x}{a}\right)$, si $0 \leq x \leq a$	$\dfrac{2}{a}\left(x - \dfrac{x^2}{2a}\right)$	$a\left(1 - \sqrt{1-U}\right)$	$a\left(1 - \sqrt{U}\right)$
Pareto $(a, b > 0)$	$\dfrac{ab^a}{x^{a+1}}$, si $x \geq b$	$1 - \left(\dfrac{b}{x}\right)^a$	$\dfrac{b}{(1-U)^{1/a}}$	$\dfrac{b}{U^{1/a}}$
Weibull $(\lambda, \alpha > 0)$	$\alpha\lambda^\alpha x^{\alpha-1} e^{-(\lambda x)^\alpha}$, si $x \geq 0$	$1 - e^{-(\lambda x)^\alpha}$	$\dfrac{(-\ln(1-U))^{1/\alpha}}{\lambda}$	$\dfrac{(-\ln U)^{1/\alpha}}{\lambda}$ strut

Distribución Uniforme

A partir de la función de densidad de las variables aleatorias uniformes entre a y b,

$$f(x) = \frac{1}{b-a} \quad a \leq x \leq b$$

se obtiene la función acumulada

$$F(x) = \int_a^x \frac{1}{b-a} dx = \frac{x-a}{b-a} \quad a \leq x \leq b$$

Igualando la función acumulada $F(x)$ con el número pseudoaleatorio $r_i \sim U(0, 1)$, y despejando x se obtiene:

$$x_i = a + (b-a)F(x)_i$$
$$x_i = a + (b-a)r_i$$

Ejemplo 3.6

La temperatura de una estufa se comporta uniformemente dentro del rango de 95 a 100°C. Una lista de números pseudoaleatorios y la ecuación $x_i = 95 + 5r_i$ nos permiten modelar el comportamiento de la variable aleatoria que simula la temperatura de la estufa (vea la tabla 3.6).

Tabla 3.6 Simulación de las temperaturas de una estufa

Medición	r_i	Temperatura °C
1	0.48	97.40
2	0.82	99.10
3	0.69	98.45
4	0.67	98.35
5	0.00	95.00

Distribución exponencial

A partir de la función de densidad de las variables aleatorias exponenciales con media $1/\lambda$,

$$f(x) = \lambda e^{-\lambda x} \qquad \text{para} \qquad x \geq 0$$

se obtiene la función acumulada

$$F(x) = \int_0^x \lambda e^{-\lambda x} dx \qquad = \qquad 1 - e^{-\lambda x} \qquad \text{para} \qquad x \geq 0$$

Igualando la función acumulada $F(x)$ con el número pseudoaleatorio $r_i \sim U(0, 1)$, y despejando x se obtiene:

$$x_i = -\frac{1}{\lambda} \ln(1 - F(x)_i)$$

$$x_i = -\frac{1}{\lambda} \ln(1 - r_i)$$

Los datos históricos del tiempo de servicio en la caja de un banco se comportan de forma exponencial con media de 3 minutos/cliente. Una lista de números pseudoaleatorios $r_i \sim U(0, 1)$ y la ecuación generadora exponencial $x_i = -3\ln(1 - r_i)$ nos permiten simular el comportamiento de la variable aleatoria (vea la tabla 3.7).

Tabla 3.7 Simulación del tiempo de servicio en la caja de un banco

Cliente	r_i	Tiempo de servicio (min)
1	0.64	3.06
2	0.83	5.31
3	0.03	0.09
4	0.50	2.07
5	0.21	0.70

Distribución de Bernoulli

A partir de la distribución de probabilidad de las variables aleatorias de Bernoulli con media

$$p(x) = p^x(1 - p)^{1-x} \qquad \text{para} \qquad x = 0, 1$$

se calculan las probabilidades para $x = 0$ y $x = 1$, para obtener

x	0	1
$p(x)$	$1 - p$	p

Acumulando los valores de $p(x)$ se obtiene:

x	0	1
$P(x)$	$1 - p$	1

Generando números pseudoaleatorios $r_i \sim U(0, 1)$ se aplica la regla:

$$x_i = \begin{cases} \text{si} & r_i \in (0, 1-p) \quad x = 0 \\ \text{si} & r_i \in (1-p, 1) \quad x = 1 \end{cases}$$

Ejemplo 3.8

Los datos históricos sobre la frecuencia de paros de cierta máquina muestran que existe una probabilidad de 0.2 de que ésta falle ($x = 1$), y de 0.8 de que no falle ($x = 0$) en un día determinado. Generar una secuencia aleatoria que simule este comportamiento.

A partir de la distribución de probabilidad de la variable aleatoria de Bernoulli con media 0.8,

$$P(x) = (0.2)^x(0.8)^{1-x} \qquad \text{para} \qquad x = 0, 1$$

se calculan las probabilidades puntuales y las acumuladas para $x = 0$ y $x = 1$, y se obtienen los datos ilustrados en la tabla 3.8:

Tabla 3.8 Cálculo de las probabilidades acumuladas de las fallas de la máquina del ejemplo 3.8

x	0	1
$P(x)$	0.8	0.2
$P(x)$	0.8	1

La regla para generar esta variable aleatoria estaría dada por:

$$x_i = \begin{cases} \text{si} & r_i \in (0 - 0,8) \quad x = 0 \\ \text{si} & r_i \in (0.8 - 1) \quad x = 1 \end{cases}$$

Con una lista de números pseudoaleatorios $r_i \sim U(0, 1)$ y la regla anterior es posible simular el comportamiento de las fallas de la máquina a lo largo del tiempo, considerando que:

- si el número pseudoaleatorio es menor que 0.8, la máquina no fallará, y
- si el número pseudoaleatorio es mayor que 0.8, ocurrirá la falla (vea la tabla 3.9).

Tabla 3.9 Simulación de las fallas de la máquina

Día	r_i	x_i	Evento: la máquina
1	0.453	0	no falla
2	0.823	1	falla
3	0.034	0	no falla
4	0.503	0	no falla
5	0.891	1	falla

Método de Convolución

En algunas distribuciones de probabilidad la variable aleatoria a simular, Y, puede generarse mediante la suma de otras variables aleatorias X de manera más rápida que a través de otros métodos. Entonces, el método de convolución se puede expresar como:

$$Y = X_1 + X_2 + \dots + X_k$$

Las variables aleatorias de cuatro de las distribuciones más conocidas (de Erlang, normal, binomial y de Poisson) pueden ser generadas a través de este método, como se verá a continuación.

Distribución de Erlang

La variable aleatoria k-Erlang con media $1/\lambda$ puede producirse a partir de la generación de k variables exponenciales con media $1/k\lambda$:

$$Y = X_1 + X_2 + \dots + X_4$$

$$Y = -\frac{1}{k\lambda}\ln(1 - r_1) - \frac{1}{k\lambda}\ln(1 - r_2) - \dots - \frac{1}{k\lambda}\ln(1 - r_k)$$

$$Y = -\frac{1}{k\lambda}\left[\ln(1 - r_1) + \ln(1 - r_2) + \dots + \ln(1 - r_k)\right]$$

$$Y = -\frac{1}{k\lambda}\left[\ln\left((1 - r_1)(1 - r_2)\dots(1 - r_k)\right)\right]$$

$$Y = -\frac{1}{k\lambda}\left[\ln\left((1 - r_1)(1 - r_2)\dots(1 - r_k)\right)\right]$$

$$Y = ER_i = -\frac{1}{k\lambda}\left[\ln\prod_{i=1}^{k}(1 - r_i)\right]$$

Ejemplo

- El tiempo de proceso de cierta pieza sigue una distribución 3-Erlang con media $1/\lambda$ de 8 minutos/pieza. Una lista de números pseudoaleatorios r_i (0,1) y la ecuación de generación de números Erlang permite obtener la tabla 3.14, que indica el comportamiento de la variable aleatoria.

$$Y = ER_i = -\frac{8}{3}\left[\ln \prod_{i=1}^{k}(1 - r_i)\right]$$

$$Y = -\frac{8}{3}\ln\left[(1 - r_1)(1 - r_2)(1 - r_k)\right]$$

Tabla 3.14 Simulación del tiempo de proceso para el ejemplo 3.11

Pieza	$1 - r_i$	$1 - r_i$	$1 - r_i$	Tiempo de proceso (min/pza)
1	0.28	0.52	0.64	6.328
2	0.96	0.37	0.83	3.257
3	0.04	0.12	0.03	23.588
4	0.35	0.44	0.50	6.837
5	0.77	0.09	0.21	11.279

Ejemplos de aplicación de esta técnica:

- Una variable Erlang-k es la suma de k exponenciales.
- Una variable Binomial de parámetros n y p es la suma de n variable Bernoulli con probabilidad de éxito p.
- La chi-cuadrado con v grados de libertad es la suma de cuadrados de v normales N (0,1).
- La suma de un gran número de variables de determinada distribución tiene una distribución normal. Este hecho es usado para generar variables normales a partir de la suma de números U (0,1) adecuados.
- Una variable Pascal es la suma de m geométricas.
- La suma de dos uniformes tiene una densidad triangular.

Distribución Normal

- La variable aleatoria normal con media μ y desviación estándar σ puede generarse usando el teorema del limite central

$$x = N_i = \left[\sum_{i=1}^{12} (r_i) - 6 \right] \sigma + \mu$$

Ejemplo

- El volumen de liquido de un refresco sigue una distribución normal con media de 12 onzas y desviación estándar de 0.4 onzas. Generar 5 variables aleatorias con esta distribución para simular el proceso de llenado

Distribución binomial

La variable aleatoria binomial con parámetros N y p puede ser generada a través de la suma de N variables aleatorias con distribución de Bernoulli con parámetro p.

$$Y = B_i = BE_1 + BE_2 + \ldots + BE_N \sim BI(N, p)$$

Ejemplo 3.13

Al inspeccionar lotes de tamaño $N = 5$, la probabilidad de que una pieza sea defectuosa es 0.03. Simular el proceso de inspección para determinar el número de piezas defectuosas por lote.

Este proceso sigue una distribución binomial con $N = 5$ y $p = 0.03$, y será simulado mediante la generación de variables aleatorias de Bernoulli con $p = 0.03$, de acuerdo con el procedimiento señalado en la sección anterior, donde $BE_i = 0$ representa una pieza en buen estado y $BE_i = 1$ una pieza defectuosa. (Observe los resultados en la tabla 3.16.)

$$BE_i = \begin{cases} 0 & \text{si} \quad r_i \in (0 - 0.97) \\ 1 & \text{si} \quad r_i \in (0.97 - 1) \end{cases}$$

$$B_i = BE_1 + BE_2 + \ldots + BE_5$$

Tabla 3.16 Simulación del sistema de inspección del ejemplo 3.13

Lote	r_1	BE_1	r_2	BE_2	r_3	BE_3	r_4	BE_4	r_5	BE_5	Piezas defectuosas
1	0.49	0	0.32	0	0.15	0	0.01	0	0.45	0	0
2	0.11	0	0.85	0	0.93	0	0.99	1	0.61	0	1
3	0.57	0	0.92	0	0.84	0	0.74	0	0.82	0	0
4	0.62	0	0.01	0	0.68	0	0.98	1	0.99	1	2
5	0.34	0	0.98	1	0.99	1	0.02	0	0.98	1	3

Método de composición

Sea f(x) la función de densidad de una variable aleatoria X. El método de composición consiste en lo siguiente:

1. Se descompone f como $f(x) = \sum_{i=1}^{m} p_i f_i(x)$ con fi función de densidad, $p_i \geq 0, i = 1, \ldots, m, \sum_{i=1}^{m} p_i = 1$.

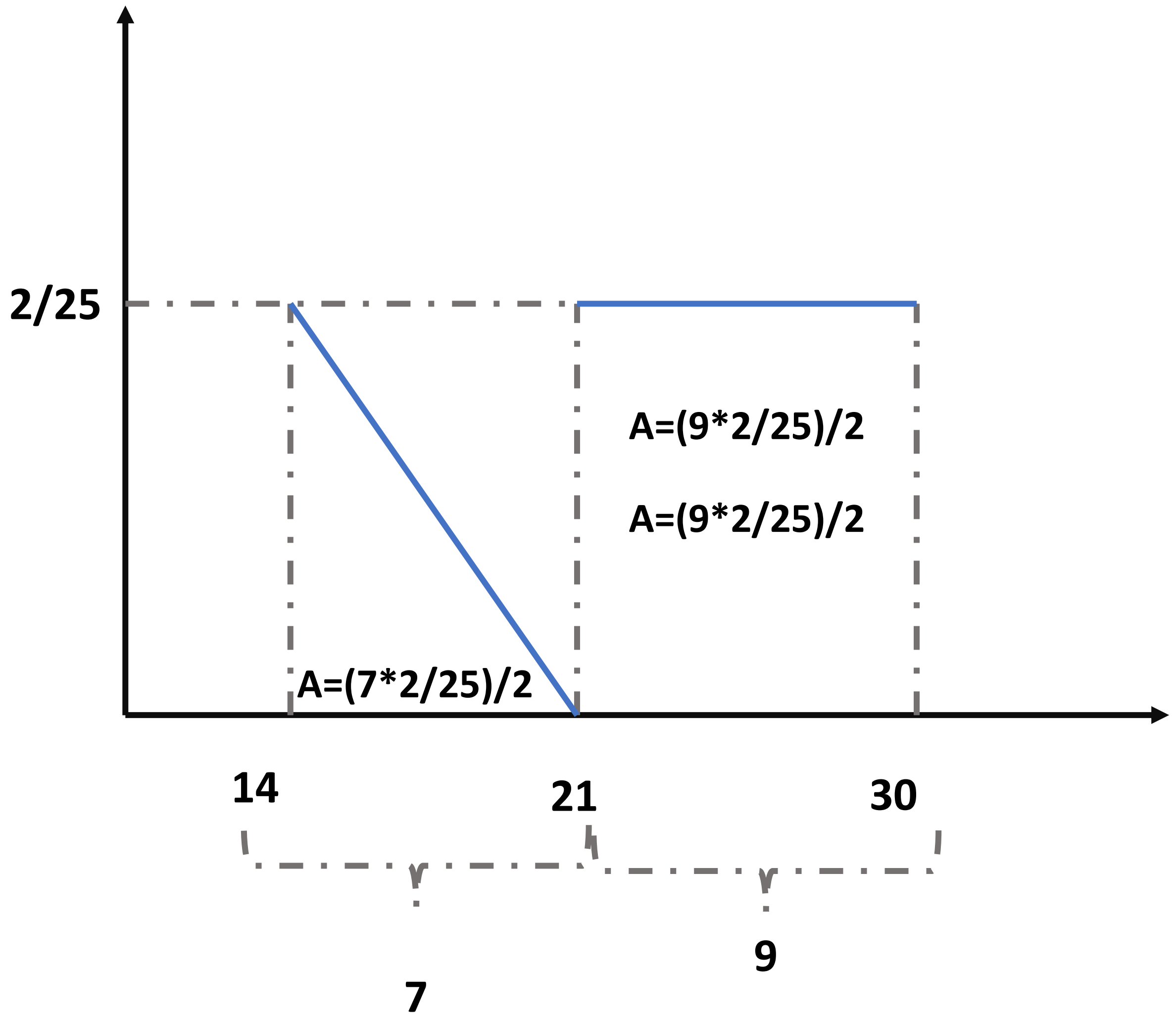
2.Se elige la i-ésima función con probabilidad pi y se genera un valor para la función de densidad fi.

Algoritmo: 1

Se genera u_1, u_2 con distribución U(0,1).

u_1 <prob entonces f1(u_2) y en caso contrario f2(u_2)

Ejemplo



Con probabilidades y simuladores

Región	Prob.	Distribución	Simulador	Valor simulado
Reg1	7/25	Triang (derecha)	21-7*raíz(ri)	18.92
Reg2	18/25	Uniforme	21+9*ri	22.91

r1=0.577
26.169

r2=0.129
18.48

Distribución doble exponencial

A partir de la densidad de la distribución doble exponencial:

$$f\left(x\right)=\frac{\lambda}{2}e^{-\lambda|x|},\ \forall x\in\mathbb{R},$$

se deduce que:

$$f\left(x\right)=\frac{1}{2}f_1\left(x\right)+\frac{1}{2}f_2\left(x\right)$$

siendo:

$$f_1\left(x\right)=\left\{\begin{array}{ll}\lambda e^{-\lambda x} & \text{si } x\geq 0\\ 0 & \text{si} x<0\end{array}\right. \quad f_2\left(x\right)=\left\{\begin{array}{ll}\lambda e^{\lambda x} & \text{si } x<0\\ 0 & \text{si} x\geq 0\end{array}\right.$$

Aceptación y rechazo – ejercicio

$$f_x(x) = \frac{3}{2}x^2, \qquad -1 \le x \le 1$$

$$t(x) = 1.5, \qquad -1 \le x \le 1 \qquad c = \int_{-1}^{1} t(x)dx = 3$$

$$r(x) = \frac{t(x)}{c}$$

$$r(x) = 0.5, \qquad -1 \le x \le 1$$

$$F_r(x) = \int_{-1}^{x} r(y)dy = \int_{-1}^{x} 0.5dy = 0.5y\Big|_{-1}^{x} = 0.5x + 0.5$$

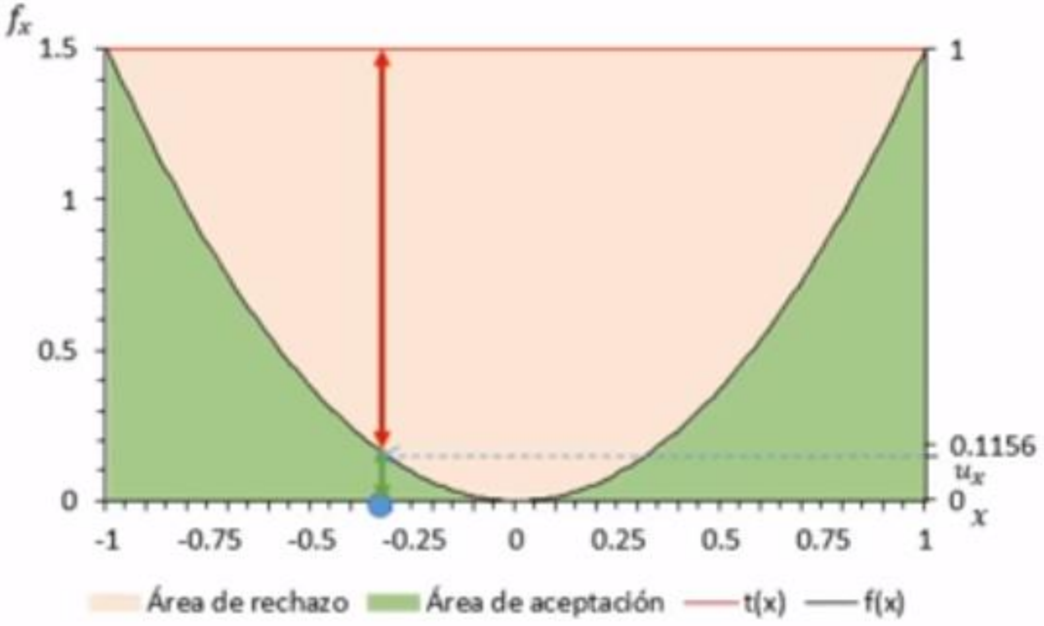
$$x = F_r^{-1}(u_x) = 2u_x - 1$$

$u_r = 0.33$ $\qquad$ $u_x = 0.104$

$$x = 2 \cdot 0.33 - 1 = -0.34$$

$$P\text{ (aceptar } x) = \frac{\frac{3}{2}(-0.34)^2}{\frac{3}{2}} = 0.1156$$

$x = -0.34$ ✓



Pasos:

1. a. Generar $u_r \sim U[0,1]$
   b. Generar $x \sim r(x)$, usando u_r
2. $P\text{ (aceptar } x) = \dfrac{f_x(x)}{t(x)}$
3. Generar $u_x \sim U[0,1]$
4. Si $u_x \le P$ (aceptar x) → devolver x
   Si $u_x > P$ (aceptar x) → ir a 1.