Simulación Estocástica

Generación de variables aleatorias

Vanda Inácio de Carvalho

Primer Semestre 2015

- Suponga que estamos interesados en generar el valor de una variable aleatoria discreta X con función masa de probabilidad

$$\Pr(X = x_j) = p_j, \quad j = 0, 1, \ldots, \quad \sum_j p_j = 1.$$

- Si $U \sim \text{Unif}(0, 1)$, entonces podemos ocupar el siguiente algoritmo para generar una muestra de X

Algoritmo

$$X = \begin{cases} x_0 & si\ U < p_0 \\ x_1 & si\ p_0 \leq U < p_0 + p_1 \\ x_2 & si\ p_0 + p_1 \leq U < p_0 + p_1 + p_2 \\ \vdots & \\ x_j & si\ \sum_{i=0}^{j-1} p_i \leq U < \sum_{i=0}^{j} p_i \\ \vdots & \end{cases}$$

- Es decir, hacemos $X = x_j$ si $F(x_j - 1) \leq U < F(x_j)$.
- Verifiquemos que usando esta construcción de hecho se tiene $\Pr(X = x_j) = p_j$

$$
\begin{aligned}
\Pr(X = x_j) &= \Pr\left(\sum_{i=0}^{j-1} p_i \leq U < \sum_{i=0}^{j} p_i\right) \\
&= \Pr\left(U < \sum_{i=0}^{j} p_i\right) - \Pr\left(U < \sum_{i=0}^{j-1} p_i\right) \\
&= F_U\left(\sum_{i=0}^{j} p_i\right) - F_U\left(\sum_{i=0}^{j-1} p_i\right) \\
&= \sum_{i=0}^{j} p_i - \sum_{i=0}^{j-1} p_i \quad (F_U(u) = u, \forall u \in [0,1]). \\
&= p_j
\end{aligned}
$$

Ejemplo

- *Suponga que* X *es una función aleatória con masa de probabilidad* $p_j = \Pr(X = j)$, $j = 1, 2, 3, 4$ *dada por*

$$p_1 = 0.2, \quad p_2 = 0.15, \quad p_3 = 0.25, \quad p_4 = 0.4.$$

  Simular una muestra de X.

- ***Algoritmo***

  $U \sim \mathit{Unif}(0, 1)$
  Si $U < 0.2$, $X = 1$
  Si $0.2 \leq U < 0.35$, $X = 2$
  Si $0.35 \leq U < 0.6$, $X = 3$
  Si $U \geq 0.6$, $X = 4$

```
nsim=1000; x=numeric(nsim)
set.seed(123)
for(i in 1:nsim){
u=runif(1)
if(u<0.2) x[i]=1
if(u>=0.2 & u<0.35) x[i]=2
if(u>=0.35 & u<0.6) x[i]=3
if(u>=0.6) x[i]=4
}
table(x)
```

Generación de variables aleatorias discretas

Método de la transformada inversa (caso discreto)

Alternativamente

```
gvad=function(x,p){
n=length(x); cp=c(0,cumsum(p))
u=runif(1)
for(i in 1:n){
if(u>cp[i] & u<=cp[i+1])
return(x[i])
}
}


x=1:4; p=c(0.2,0.15,0.25,0.4)
nsim=1000
x.sim=array(nsim)
set.seed(123)
for(j in 1:nsim) x.sim[j]=gvad(x,p)
table(x.sim)
```

Generación de variables aleatorias discretas

Método de la transformada inversa (caso discreto)

- Vamos a usar em método de la transformada inversa para generar variables aleatorias con distribución binomial.
- $X \sim \text{Binomial}(n, p)$ si

$$\Pr(X = j) = \binom{n}{j} p^j (1-p)^{n-j}, \quad j = 0, \ldots, n.$$

- X puede ser interpretada como el el número de sucesos en n pruebas de Bernoulli independientes y con probabilidad de suceso p.
- Así, si $Y_1, \ldots, Y_n$ son variables aleatorias independientes tales que $Y_i \sim \text{Bernoulli}(p)$, entonces

$$X = \sum_{i=1}^{n} Y_i \sim \text{Binomial}(n, p).$$

- Por lo tanto, para generar una realización de X, necesitamos generar n valores de Y.
- Note que Y sólo puede tomar dos valores: 0 y 1, con probabilidades $1 - p$ y p, respectivamente.

- El algoritmo es entonces el siguiente

Ejemplo

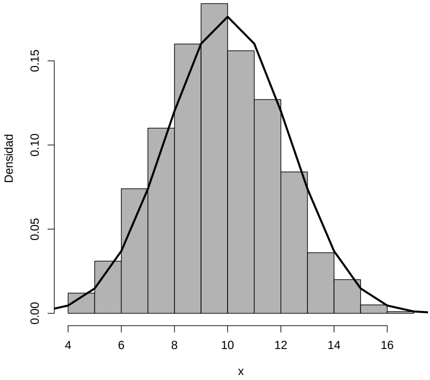
1. *Simular* $U_1, \ldots, U_n \sim Unif(0,1)$.
2. *Para* $i = 1, \ldots, n$, *si* $U_i \in [0, p)$, *entonces* $Y_i = 1$. *En otro caso,* $Y_i = 0$.
3. *Hacer,* $X = \sum_{i=1}^{n} Y_i$

```
rbinomial=function(n,p){
u=runif(n,0,1)
y=u<=p
x=sum(y)
return(x)
}

m=1000; x=numeric(m)
for(j in 1:m) x[j]=rbinomial(20,0.5)
table(x)
```

Generación de variables aleatorias discretas

Método de la transformada inversa (caso discreto)



Generación de variables aleatorias contínuas

Método de la transformada inversa (caso contínuo)

- Sea X una variable aleatoria contínua con función de distribución F.
- Sea $U \sim \text{Unif}(0,1)$. **Hecho**: $F(X) \sim U$. Tenemos entonces que $X \sim F^{-1}(U)$.
- Verifiquemos que se tiene $\Pr(X \leq x) = F(x)$

$$
\begin{aligned}
\Pr(X \leq x) &= \Pr(F^{-1}(U) \leq x) \\
&= \Pr(F(F^{-1}(U)) \leq F(x)) \qquad (F \text{ es no decreciente}) \\
&= \Pr(U \leq F(x)) \\
&= F_U(F(x)) \\
&= F(x)
\end{aligned}
$$

Generación de variables aleatorias contínuas

Método de la transformada inversa (caso contínuo)

- Si F es bijectiva, entonces $F^{-1}(u)$ es definida como siendo el valor de x tal que $F(x) = u$.

- Si no este el caso, se puede siempre usar la definición de inversa generalizada, dada por

$$F^{-1}(u) = \inf\{x \in \mathbb{R} : F(x) \geq u\}.$$

- Así, si $u_1, u_2, \ldots$ son realizaciones de U, con $U \sim \text{Unif}(0, 1)$, entonces $x_1, x_2, \ldots$ son realizaciones de X, donde $x_i = F^{-1}(u_i)$.

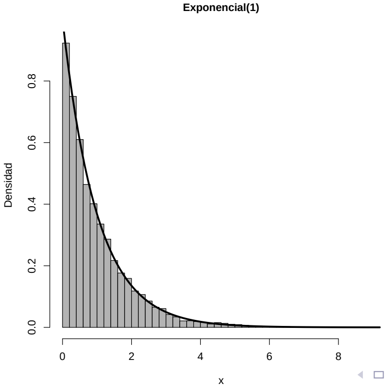
Ejemplo

- *Suponga que* $X \sim Exp(\lambda)$. *Entonces,* $F_X(x) = 1 - e^{-\lambda x}, \quad x \geq 0$.
- *Sea* $u \sim Unif(0, 1)$.

$$
\begin{aligned}
u = F_X(x) &\Rightarrow u = 1 - e^{-\lambda x} \\
&\Rightarrow 1 - u = e^{-\lambda x} \\
&\Rightarrow x = -\frac{1}{\lambda}\log(1 - u)
\end{aligned}
$$

- *Así,* $X = -\frac{1}{\lambda}\log(1 - U)$

```
m=10000; u=runif(m)
lambda=1; x=-log(1-u)
hist(x,freq=F,col="gray70",breaks=50)
curve(dexp(x,1),add=T,lwd=3)
```



Generación de variables aleatorias contínuas

Método de la transformada inversa (caso contínuo)

Ejemplo

- *Suponga que* $X \sim$ *Weibull*(α, β). *Tenemos que* $F_X(x) = 1 - e^{-(x/\alpha)^\beta}, \quad x \geq 0$.
- *Sea* $u \sim$ *Unif*$(0, 1)$.

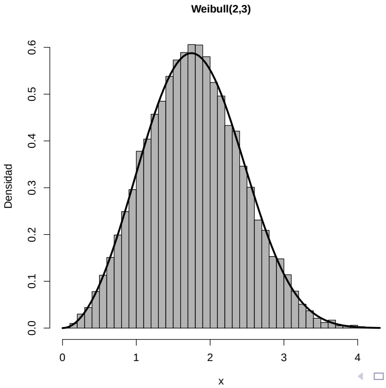
$$
\begin{aligned}
u = F_X(x) &\Rightarrow 1 - e^{-(x/\alpha)^\beta} \\
&\Rightarrow x = \alpha[-\log(1-u)]^{1/\beta}
\end{aligned}
$$

- *Así,* $X = \alpha[-\log(1-U)]^{1/\beta}$.

Generación de variables aleatorias contínuas

Método de la transformada inversa (caso contínuo)

```
m=10000; u=runif(m)
alpha=2; beta=3; x=alpha*((-log(1-u))**(1/beta))
hist(x,freq=F,col="gray70",breaks=50)
curve(dweibull(x,3,2),add=T,lwd=3)
```



Generación de variables aleatorias contínuas

Método de la transformada inversa (caso contínuo)

Ejemplo

- *Sea X una variable aleatoria tal que su función de densidad está dada por*

$$f(x) = \left|\frac{x}{4}\right|, \quad -2 \leq x \leq 2.$$

- *Podemos entonces escribir f de la siguiente manera*

$$f(x) = \begin{cases} -\frac{x}{4} & si\ -2 \leq x \leq 0 \\ \frac{x}{4} & si\ 0 \leq x \leq 2 \end{cases}$$

- *Determinemos ahora la función de distribución, $F(x)$.*
- *Para $x < -2$ tenemos que $F(x) = 0$, mientras que para $x > 2$, tenemos $F(x) = 1$.*
- *Para $-2 \leq x \leq 0$, tenemos que*

$$F(x) = \int_{-2}^{x} -\frac{z}{4} dz = -\left[\frac{z^2}{8}\right]_{-2}^{x} = -\frac{x^2}{8} + \frac{1}{2}.$$

Ejemplo

- *Para* $0 \leq x \leq 2$*, tenemos que*

$$F(x) = \int_{-2}^{0} -\frac{x}{4}dx + \int_{0}^{x} \frac{z}{4}dz = \left[-\frac{x^2}{8}\right]_{-2}^{0} + \left[\frac{z^2}{8}\right]_{0}^{x} = \frac{x^2}{8} + \frac{1}{2}.$$

- $F(x)$ *está dada entonces por*

$$F(x) = \begin{cases} 0 & si\ x < 0 \\ -\frac{x^2}{8} + \frac{1}{2} & si\ -2 \leq x \leq 0 \\ \frac{x^2}{8} + \frac{1}{2} & si\ 0 \leq x \leq 2 \\ 1 & si\ x > 2 \end{cases}$$

- *Determinemos ahora la inversa de* F*. Sea* $u \sim Unif(0,1)$ *y verifiquemos, primeramente, los limites de* u *para los cuales* $F^{-1}(u)$ *está definida*

$$x = -2 \Rightarrow u = 0, \quad x = 0 \Rightarrow u = 1/2, \quad x = 2 \Rightarrow u = 1.$$

Generación de variables aleatorias contínuas

Método de la transformada inversa (caso contínuo)

Ejemplo

- *Para* $-2 \leq x \leq 0$

$$-\frac{x^2}{8} + \frac{1}{2} = u \Rightarrow x^2 = 4 - 8u \underset{-2\leq x\leq 0}{\Rightarrow} x = -2\sqrt{1-2u}.$$

- *Para* $0 \leq x \leq 2$

$$\frac{x^2}{8} + \frac{1}{2} = u \Rightarrow x^2 = 8u - 4 \underset{0\leq x\leq 2}{\Rightarrow} x = 2\sqrt{2u-1}.$$

- *Así*

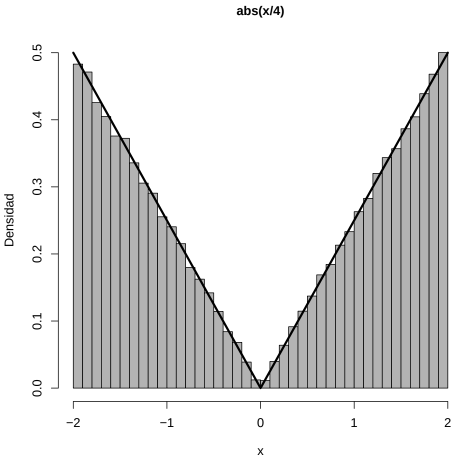
$$F^{-1}(u) = \begin{cases} -2\sqrt{1-2u} & si\ 0 \leq u \leq 0.5 \\ 2\sqrt{2u-1} & si\ 0.5 \leq u \leq 1 \end{cases}$$

```
modulo=function(m){
x=numeric(m); u=runif(m)
for(i in 1:m){
if(u[i]<0.5) x[i]=-2*sqrt(1-2*u[i])
else x[i]=2*sqrt(2*u[i]-1)
}
return(x)
}


x=modulo(100000)
hist(x,freq=F,col="gray70",breaks=50)
z=seq(-2,2,by=0.005); dens=ifelse(z<0,-z/4,z/4)
lines(z,dens,lwd=3)
```

Generación de variables aleatorias contínuas

Método de la transformada inversa (caso contínuo)



Generación de variables aleatorias contínuas

Método basados en exponenciales

- Hemos visto como generar variables aleatorias exponenciales usando variables aleatorias uniformes.
- Algunas otras variables pueden ser generadas a partir de la distribución exponencial.
- Si $Y_i \overset{\text{iid}}{\sim} \text{Exp}(1)$, entonces

$$X = 2\sum_{i=1}^{\nu} Y_i \sim \chi^2_{2\nu}, \quad \nu \in \{1, 2, \ldots\},$$

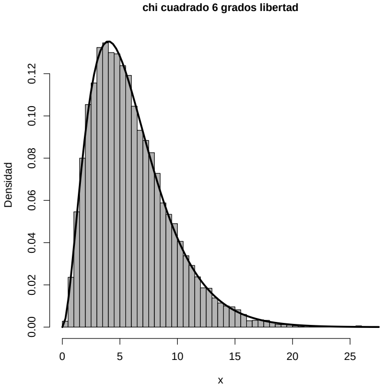
$$X = \beta\sum_{i=1}^{\alpha} Y_i \sim \text{Gamma}(\alpha, \beta), \quad \alpha \in \{1, 2, \ldots\},$$

$$X = \frac{\sum_{i=1}^{a} Y_i}{\sum_{i=1}^{a+b} Y_i} \sim \text{Beta}(a, b), \quad a, b \in \{1, 2, \ldots\}$$

Generación de variables aleatorias contínuas

Método de la transformada inversa (caso contínuo)

```
nsim=10000
u1=runif(nsim); u2=runif(nsim); u3=runif(nsim)
x=-2*log(u1*u2*u3)
```



- Estas transformaciones son simples de usar pero su uso es limitado; por ejemplo

  (i) no podemos generar variables aleatorias gamma con parámertro de forma α no entero;

  (ii) no podemos generar un chi cuadrado con número de grados de libertad impar, y en particular, no podemos generar una χ_1^2, lo que nos permitiria generar $N(0, 1)$s.

Generación de variables aleatorias contínuas

Método de Box-Muller

- El método de Box-Muller produce dos números aleatorios normales independientes a partir de dos números aleatorios uniformes independientes.

- Sean $X_1, X_2 \overset{\text{iid}}{\sim} \text{N}(0,1)$, La función de densidad conjunta está dada por

$$f_{X_1,X_2}(x_1,x_2) \overset{\text{ind}}{=} f_{X_1}(x_1) f_{X_2}(x_2) = \frac{1}{2\pi} \exp\left\{-\frac{1}{2}(x_1^2 + x_2^2)\right\}, \quad x_1, x_2 \in \mathbb{R}.$$

- Es más conveniente escribir la densidad conjunta en coordenadas polares. Las coordenadas polares de (x_1, x_2) son

$$\begin{cases} x_1 = r\cos(\theta) \\ x_2 = r\sin(\theta) \end{cases}$$

con $r = \sqrt{x_1^2 + x_2^2}$ y $\theta = \tan^{-1}(x2/x1)$.

- El jacobiano de la transformación está dado por

$$\left|\frac{\partial(x_1,x_2)}{\partial(r,\theta)}\right| = \begin{vmatrix} \frac{\partial x_1}{\partial r} & \frac{\partial x_1}{\partial \theta} \\ \frac{\partial x_2}{\partial r} & \frac{\partial x_2}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos(\theta) & -r\sin(\theta) \\ \sin(\theta) & r\cos(\theta) \end{vmatrix} = r$$

- Así, la densidad conjunta de $R = \sqrt{x_1^2 + X_2^2}$ y $\Theta = \tan^{-1}(X_2/X_1)$ es

$$f_{R,\Theta}(r,\theta) = f_{X_1,X_2}(r,\theta)\left|\frac{\partial(x_1,x_2)}{\partial(r,\theta)}\right| = \frac{1}{2\pi}\exp\left\{-\frac{1}{2}r^2\right\} r, \quad r > 0, 0 \leq \theta < 2\pi$$

- Nuestro objetivo es generar una muestra desde X_1 y X_2, pero si logramos saber como generar de R y Θ, sabemos como generar desde X_1 y X_2.

- Analisemos entonces las distribuciones marginales.

Generación de variables aleatorias contínuas

Método de Box-Muller

- La distribución marginal de Θ es

$$f_\Theta(\theta) = \int_{r=0}^{2\pi} \frac{1}{2\pi} \exp\left\{-\frac{1}{2}r^2\right\} r dr = \frac{1}{2\pi}\left[-\exp\left\{-\frac{1}{2}r^2\right\}\right]_0^\infty = \frac{1}{2\pi}, \quad 0 \leq \theta < 2\pi.$$

- Concluídos entonces que $\Theta \sim \text{Unif}(0, 2\pi)$.

- La distribución marginal de R es

$$f_R(r) = \int_0^{2\pi} \frac{1}{2\pi} r \exp\left\{-\frac{1}{2}r^2\right\} d\theta = \frac{1}{2\pi}[\theta]_0^{2\pi} r \exp\left\{-\frac{1}{2}r^2\right\} = r \exp\left\{-\frac{1}{2}r^2\right\}, \quad r > 0.$$

- Notemos que $f_{R,\Theta}(r,\theta) = f_R(r) f_\Theta(\theta)$, o sea, R y Θ son independientes.

- No reconocemos la densidad de $f_R(r)$ como una densidad conocida. Usaremos entonces el mı́odo de la transformada inversa para generar una muestra desde R.

- Determinemos entonces $F_R(r)$

$$F_R(r) = \int_0^r z \exp\left\{-\frac{1}{2}z^2\right\} \mathrm{d}z = -\left[\exp\left\{-\frac{1}{2}z^2\right\}\right]_0^r = 1 - \exp\left\{-\frac{1}{2}r^2\right\}, \quad r > 0.$$

- Sea $u \sim \text{Unif}(0, 1)$,

$$F_R(r) = u \underset{r>0}{\Rightarrow} r = \sqrt{-2\log(1-u)}.$$

- Como $U \sim \text{Unif}(0, 1)$, entonces $1 - U \sim \text{Unif}(0, 1)$ y, por lo tanto, podemos considerar $R = \sqrt{-2\log U}$.

- Así, el procedimiento para generar los dos normales independientes a partir de los dos uniformes independientes es el siguiente:

  1. Generar $U_1, U_2 \overset{\text{iid}}{\sim} \text{Unif}(0, 1)$.
  2. Hacer $\Theta = 2\pi U_1$ y $R = \sqrt{-2 \log U_2}$.
  3. Hacer $X_1 = R\cos(\Theta) = \sqrt{-2 \log U_2}\cos(2\pi U_1)$ y
     $X_2 = R\sin(\Theta) = \sqrt{-2 \log U_2}\sin(2\pi U_1)$.

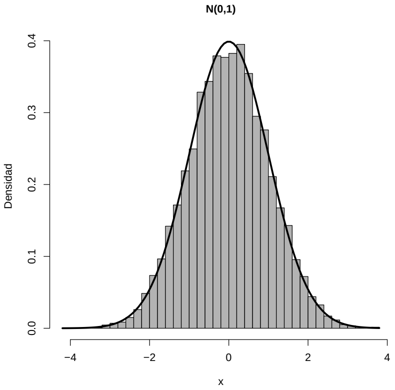
- Si el objetivo es generar $X \sim \text{N}(\mu, \sigma^2)$, entonces basta hacer $X = \mu + \sigma Z$, con $Z \sim \text{N}(0, 1)$.

- Una desventaja del algoritmo es su lentitud, debido a la necesidad de calcular funciones tales como el log, cos y sin.

Generación de variables aleatorias contínuas

Método de Box-Muller

```
m=10000; u1=runif(m); u2=runif(m)
theta=2*pi*u1; r=sqrt(-2*log(u2))
x1=r*cos(theta); x2=r*sin(theta)
```



- Supongamos que estamos interesados en obtener una muestra de una variable aleatoria con función de densidad dada por

$$f(x) = pf_1(x) + (1-p)f_2(x), \tag{1}$$

donde $0 < p < 1$.

- Una forma de simular de esta variable aleatoria es observar que si X_1 y X_2 son variables aleatorias con densidades f_1 y f_2, respectivamente, entonces la variable aleatoria X definida como

$$X = \begin{cases} X_1 & \text{con prob. } p \\ X_2 & \text{con prob. } 1-p \end{cases}$$

tendrá la densidad dada en (1).

- Luego, para generar un valor de tal variable aleatoria, generamos un número $u \sim \text{Unif}(0,1)$ y hacemos

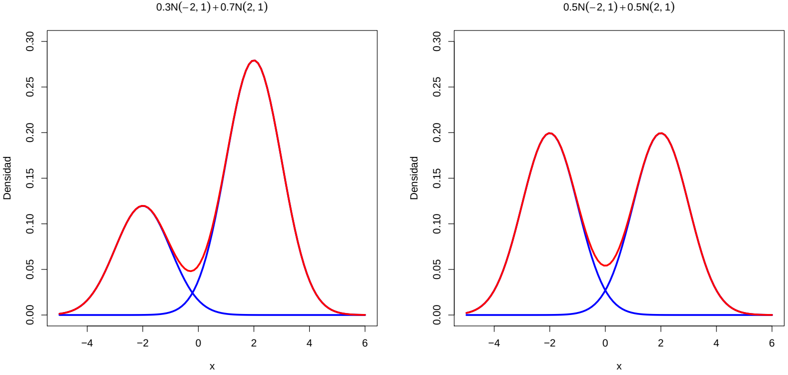
$$X = \begin{cases} X_1 & \text{si } u \leq p \\ X_2 & \text{si } u > p \end{cases}$$

- En general, si f se puede escribir como

$$f(x) = \sum_{k=1}^{K} p_k f_k(x), \quad \sum_{k=1}^{K} p_k = 1, \quad p_k \geq 0, \quad k = 1, \ldots, K.$$

- Una forma de simular a partir de f es la siguiente
  1. Escoger f_k con probabilidad p_k.
  2. Generar una observación desde f_k.
- Note que, equivalentemente, también se puede hacer la decomposición de $F(x)$ en una combinación convexa de otras funciones de distribución.

Mezclas de dos normales



Simular desde mezclas

```
m=10000; x=numeric(m); u=runif(m)
for(i in 1:m){
x[i]=ifelse(u[i]<0.5,rnorm(1,-3,1),rnorm(1,3,1))
}
hist(x,freq=F)

Alternativamente

mu=c(-3,3); sigma=c(1,1)
ind=x=numeric(m)
for(i in 1:m){
ind=sample(1:2,1,replace=T)
x[i]=rnorm(1,mu[ind],sigma[ind])
}

mu=c(-1,2); sigma=c(1,1)
ind=x=numeric(nsim)
for(i in 1:nsim){
ind=sample(1:2,1,replace=T,prob=c(0.3,0.7))
x[i]=rnorm(1,mu[ind],sigma[ind])
}
```

Ejemplo

Se desea generar una muestra aleatoria desde la siguiente densidad (conocida como densidad triangular)

$$f(x) = \begin{cases} 1 + x & si \; -1 \leq x \leq 0 \\ 1 - x & si \; 0 \leq x \leq 1 \end{cases}$$

usando

(a) *método de la transformada inversa,*

(b) *método de composición.*

Generación de variables aleatorias contínuas

Método de composición

Ejemplo

- *Usemos entonces el método de la transformada inversa para generar una muestra desde f. Para eso, necesitamos de determinar la función de distribución, F.*

- *Para $-1 \leq x \leq 0$, tenemos que*

$$F(x) = \int_{-1}^{x} (1+z)dz = \left[z + \frac{z^2}{2}\right]_{-1}^{x} = \frac{x^2}{2} + x + \frac{1}{2}$$

- *Para $0 \leq x \leq 1$, tenemos que*

$$F(x) = \int_{-1}^{0} (1+x)dx + \int_{0}^{x} (1-z)dz = \left[x + \frac{x^2}{2}\right]_{-1}^{0} + \left[z - \frac{z^2}{2}\right]_{0}^{x} = -\frac{x^2}{2} + x + \frac{1}{2}$$

Ejemplo

- *$F(x)$ está dada entonces por*

$$F(x) = \begin{cases} 0 & si\ x < 0 \\ \frac{x^2}{2} + x + \frac{1}{2} & si\ -1 \leq x \leq 0 \\ -\frac{x^2}{2} + x + \frac{1}{2} & si\ 0 \leq x \leq 1 \\ 1 & si\ x > 1 \end{cases}$$

- *Determinar $F^{-1}(u)$, con $u \sim Unif(0,1)$.*
- *Para $-1 \leq x \leq 0$*

$$\frac{x^2}{2} + x + \frac{1}{2} = u \Rightarrow x = -1 + \sqrt{2u},\ \text{dado que } -1 \leq x \leq 0.$$

- *Para $0 \leq x \leq 1$*

$$-\frac{x^2}{2} + x + \frac{1}{2} = u \Rightarrow x = 1 - \sqrt{2(1-u)},\ \text{dado que } 0 \leq x \leq 1.$$

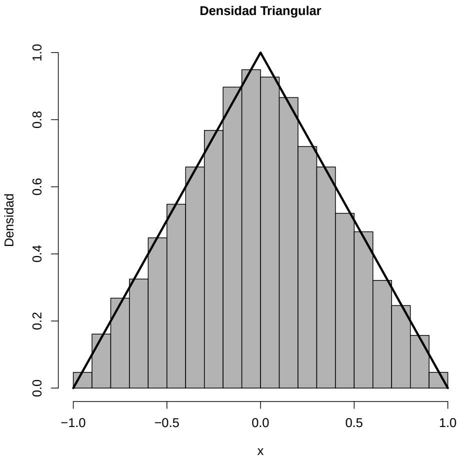
Ejemplo

- *Tenemos así*

$$F^{-1}(u) = \begin{cases} -1+\sqrt{2u} & si\, 0 \leq u \leq 0.5 \\ 1-\sqrt{2(1-u)} & si\, 0.5 \leq u \leq 1 \end{cases}$$

```
m=10000; x=numeric(m); u=runif(m)
for(i in 1:m){
x[i]=ifelse(u[i]<0.5,-1+sqrt(2*u[i]),1-sqrt(2*(1-u[i])))
}

hist(x,freq=F,col="gray70",ylim=c(0,1),xlim=c(-1,1))
z=seq(-1,1,by=0.005); denstriang=ifelse(z<0,1+z,1-z)
lines(z,denstriang,lwd=3)
```


Densidad Triangular
Densidad
0.0
0.2
0.4
0.6
0.8
1.0
−1.0
−0.5
0.0
0.5
1.0
x

Ejemplo

- *Usemos, ahora, el método de composición para generar una muestra desde f.*
- *Notemos que f puede ser equivalentemente escrita como*

$$f(x) = (1 + x)I_{[-1,0]}(x) + (1 - x)I_{[0,1]}(x),$$

y aún como

$$f(x) = 0.5 \times 2 \times (1 + x)I_{[-1,0]}(x) + 0.5 \times 2 \times (1 - x)I_{[0,1]}(x),$$

y, entonces, tenemos que

$$\begin{aligned} p &= 0.5, \\ f_1(x) &= 2(1 + x)I_{[-1,0]}(x), \\ f_2(x) &= 2(1 - x)I_{[0,1]}(x). \end{aligned}$$

Ejemplo

- *Ahora, el objetivo, es generar una muestra desde f_1 y f_2, respectivamente.*
- *Veamos primero como generar desde f_1.*

$$F_1(x) = \int_{-1}^{x} 2(z+1)dz = 2\left[\frac{z^2}{2} + z\right]_{-1}^{x} = x^2 + 2x + 1, \quad -1 \leq x \leq 0$$

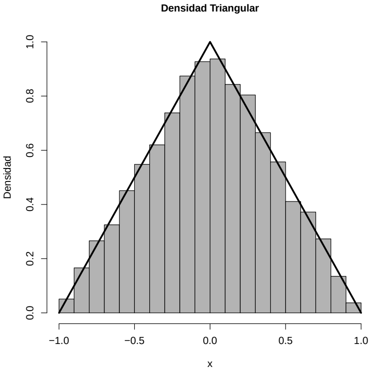
- *Sea $u \sim Unif(0,1)$. $x^2 + 2x + 1 = u \Rightarrow x = -1 + \sqrt{u}$, una vez que $-1 \leq x \leq 0$.*
- *Generemos ahora desde f_2.*

$$F_2(x) = \int_{0}^{x} 2(1-z)dz = -x^2 + 2x, \quad 0 \leq x \leq 1.$$

- *Tenemos, $-x^2 + 2x = u \Rightarrow x = 1 - \sqrt{1-u}$, dado que $0 \leq x \leq 1$.*

Generación de variables aleatorias contínuas
Método de composición

```
m=10000; xc=numeric(m); u1=runif(m); u2=runif(m)
for(i in 1:m){
xc[i]=ifelse(u1[i]<0.5,-1+sqrt(u2[i]),1-sqrt(1-u2[i]))
}
```



Generación de variables aleatorias contínuas

Algoritmo de aceptación rechazo

- Existen muchas distribuciones de las cuales es difícil, o mismo imposible, simular usando directamente la transformada inversa.
- Además, en algunos casos, no podemos representar la distribución en una forma útil, tal como en el caso de transformaciones o mezclas.
- La idea del método de aceptación-rechazo es usar una densidad desde la cual sea facil muestrear (denominada densidad instrumental o candidata) y rechazar los valores muestreados que sean poco probables bajo la densidad objetivo, la densidad de la cual deseamos muestrear.

- Supongamos entonces que estamos interesados en generar una muestra de una variable aleatoria X con densidad $f(x)$, pero que no es fácil hacerlo.
- Además, supongamos que se puede facilmente simular desde una densidad $g(x)$ (la densidad candidata o instrumental).
- Supongamos que existe una constante M tal que

$$f(x) \leq Mg(x), \quad \forall x.$$

- El algoritmo para generar desde $f(x)$ está entonces dado por

  1. Generar un valor candidato , x_c, desde $g(x)$.
  2. Generar $u \sim \text{Unif}(0, 1)$.
  3. Si
  $$u \leq \frac{f(x_c)}{Mg(x_c)},$$
  entonces hacer $x = x_c$. En otro caso, rechazar x_c.
  4. Repetir hasta obtener el tamaño de muestra deseado.

- Observaciones:

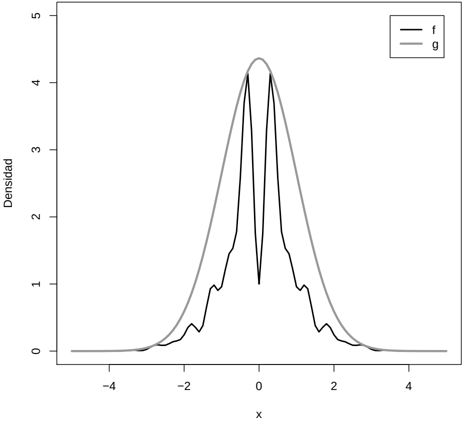
  (i) Es posible muestrear demostrar que X efectivamente tiene densidad f.

  (ii) La probabilidad de aceptación de un valor generado es $1/m$ y, el número de intentos necesários para que x_c sea aceptado tiene distribución geometrica con média M.

  (iii) Por lo tanto, por cuestiones de eficiencia M, debe ser pequeño.

  (iv) No es necesario conocer f, basta que se conozca la forma funcional de f (o sea no es necesario las constantes).

  (v) La condición $f(x) \leq Mg(x)$ es necesaria para generar efectivamente desde f.



f
g
Densidad
x
0
1
2
3
4
5
−4
−2
0
2
4

Generación de variables aleatorias contínuas

Algoritmo de aceptación rechazo

Ejemplo

- *Suponga que queremos simular* $X \sim Beta(3, 2)$*, cuya densidad está dada por*

$$f(x) = \frac{4!}{2!1!} x^2(1 - x) = 12x^2(1 - x), \quad 0 \leq x \leq 1.$$

- *La función de distribución es*

$$F(x) = \int_0^x 12z^2(1 - z)dz = x^3(4 - 3x), \quad 0 \leq x \leq 1.$$

- *Determinar la inversa* F^{-1} *no es fácil por lo que vamos a usar el algoritmo de aceptación-rechazo.*

Ejemplo

- *Como el soporte de f es el intervalo $(0,1)$, consideraremos*

$$g(x) = 1, \quad 0 \leq x \leq 1,$$

  o sea, g es la distribución uniforme en $(0,1)$.

- *Vamos, ahora, determinar el valor de M más pequeño tal que*

$$f(x)/g(x) \leq M.$$

- *Es decir, buscaremos el máximo de $f(x)/g(x)$.*
- *Sea*

$$h(x) = \frac{f(x)}{g(x)} = f(x) = 2x^2(1-x), \quad 0 \leq x \leq 1.$$

- *Tenemos que*

$$h'(x) = 0 \Rightarrow x = 0 \vee x = 2/3.$$

- *Además, $f(0) = 0$ y $f(2/3) = 16/9$. Por lo tanto, $M = 16/9$.*

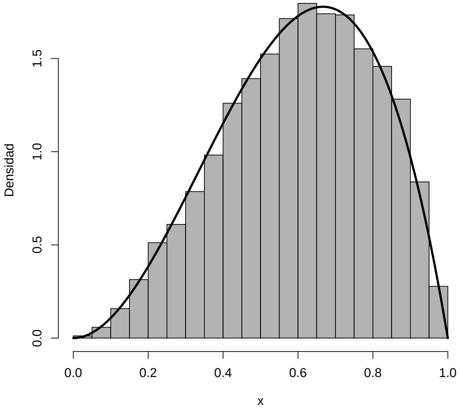
```
n=10000
x=0; t=0; M=16/9; count=0
while(t<n){
count=count+1
x.c=runif(1)
u=runif(1)
if(u<(1/M)*(12*(x.c**2)*(1-x.c))){
t=t+1
x[t]=x.c
}
}

hist(x,freq=F,xlim=c(0,1),col="gray70",ylim=c(0,1.8))
curve(dbeta(x,3,2),from=0,to=1,lwd=3,add=T)

(10000/count)*100 #número de valores candidatos aceptados
```

Generación de variables aleatorias contínuas

Algoritmo de aceptación rechazo



Generación de variables aleatorias contínuas

Algoritmo de aceptación rechazo

Ejemplo

- *Hemos ya visto el siguiente ejemplo, donde $x \sim f$, con*

$$f(x) = \begin{cases} 1 + x, & si \ -1 \leq x \leq 0 \\ 1 - x, & si \, 0 \leq x \leq 1 \end{cases}$$

- *Como el soporte de X es $[-1, 1]$, entonces hace sentido considerar $g \sim Unif(-1, 1)$, o sea,*

$$g(x) = \frac{1}{2}, \quad \forall x \in [-1, 1].$$

- *Si consideramos $M = 2$, tenemos que*

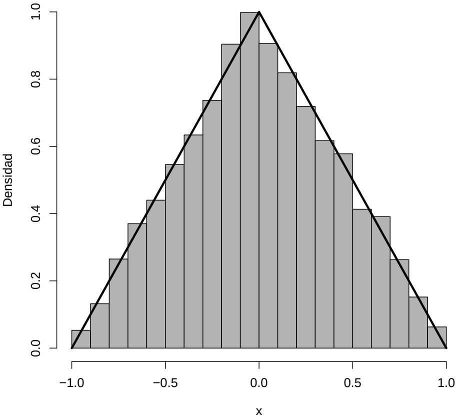
$$f(x) \leq 2 \times \frac{1}{2} \Rightarrow f(x) \leq 1, \quad \forall x \in [-1, 1].$$

```
n=10000
t=0; x=0; M=2; count=0
while(t<n){
count=count+1
x.c=runif(1,-1,1)
u=runif(1,0,1)
if(u<=ifelse(x.c<0,1+x.c,1-x.c)/(M*0.5)){
t=t+1
x[t]=x.c
}
}

hist(x,freq=F,xlim=c(-1,1),col="gray70")
z=seq(-1,1,by=0.005); denstriang=numeric(nz)
denstriang=ifelse(z<0,1+z,1-z)
lines(z,denstriang,lwd=3)
```

Generación de variables aleatorias contínuas

Algoritmo de aceptación rechazo



Ejemplo

- *Suponga que queremos generar* $X \sim Gamma(3/2, 1)$*, cuya densidad está dada por*

$$f(x) = Kx^{1/2}e^{-x}, \quad x \leq 0,$$

con $K = 1/\Gamma(3/2) = 2/\sqrt{\pi}$*.*

- *Como* X *tiene soporte en* $[0, \infty)$ *y media igual a* $3/2$*, entonces es natural considerar para* g *la distribuición exponencial con média* $3/2$*.*
- *Así,*

$$g(x) = \frac{2}{3}e^{-\frac{2}{3}x}, \quad x \leq 0.$$

- *Determinemos el menor valor de* M *tal que*

$$\frac{f(x)}{g(x)} \leq M, \quad \forall x \leq 0,$$

lo que es quivalente a determinar el máximo de $f(x)/g(x)$ *y considerar* M *igual al máximo.*

Ejemplo

- *Sea*

$$h(x) = \frac{f(x)}{g(x)} = \frac{\frac{2}{\sqrt{\pi}} x^{1/2} e^{-x}}{\frac{2}{3} e^{-2x/3}} = \frac{3}{\sqrt{\pi}} x^{1/2} e^{-x/3}, \quad x \leq 0$$

- *Determinaremos el máximo de h en* `R` *usando la funciń* `optimize` *(o* `optimise`*).*

Generación de variables aleatorias contínuas

Algoritmo de aceptación rechazo

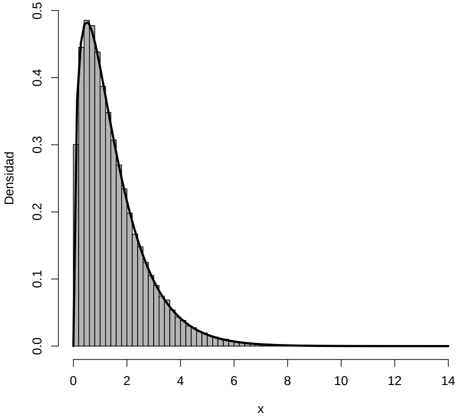
```
curve(dgamma(x,1.5,1),from=0,to=15,ylim=c(0,0.8))
curve(dexp(x,2/3),col="red",add=T)

f=function(x)(3/sqrt(pi))*(x**0.5)*exp(-x/3)
optimize(f,interval=c(0,1000),maximum=TRUE)$objective

n=10000
x=0; t=0; counter=0
while(t<n){
u=runif(1)
counter=counter+1
x.c=-1.5*log(runif(1))
alpha=(3/sqrt(pi))*(x.c**0.5)*exp(-x.c/3)/1.26
if(u<=alpha){
t=t+1
x[t]=x.c
}
}
```

Generación de variables aleatorias contínuas

Algoritmo de aceptación rechazo



Ejemplo

- *Generar una muetsra de* $X \sim f$*, con*

$$f(x) = \begin{cases} \frac{2}{\sqrt{2\pi}} e^{-x^2/2}, & si\ x \leq 0 \\ 0, en\ otro\ caso \end{cases}$$

- f *es la llamada distribución 'half normal'.*
- *Consideremos para* g *la distribución exponencial con média* 1*,*

$$g(x) = e^{-x}, \quad x \leq 0.$$

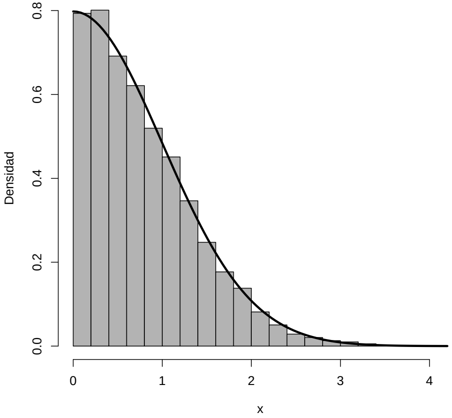
- *Determinemos el máximo de* $h(x)$ *a fin de obtener el valor de* M

$$h(x) = \frac{f(x)}{g(x)} = \frac{2}{\sqrt{2\pi}} e^{x - x^2/2}.$$

- *El máximo de* h *ocurre en el valor de* x *que maximiza* $x - x^2/2$*, que haciendo los cálculos verificamos ser* $x = 1$*. Así,* $M = h(1) = \sqrt{\frac{2e}{\pi}}$*.*

```
curve(2*dnorm(x,0,1),from=0,to=15,ylim=c(0,1.5),lwd=2)
curve(sqrt(2*exp(1)/pi)*dexp(x,1),col="blue",add=T,lwd=3)

n=10000
x=0; t=0; counter=0
while(t<n){
u=runif(1)
counter=counter+1
x.c=-log(runif(1)) x.c=rexp(1,1)
alpha=2*dnorm(x.c,0,1)/(sqrt(2*exp(1)/pi)*dexp(x.c,1))
if(u<=alpha){
t=t+1
x[t]=x.c
}
}
```



Densidad
0.0
0.2
0.4
0.6
0.8
0
1
2
3
4
x

Generación de variables aleatorias contínuas

Correcciones/comentarios/sugerencias:

vanda.kinets@gmail.com o icalhau@mat.puc.cl