Simulación de la distribución triangular

Silvia Suárez Crespo

17 de mayo de 2009

1. Enunciado

Programar una función en R para generar valores de una variable aleatoria dada por la distribución triangular en (0,c), cuya función de densidad es la siguiente:

$$f(x) = \frac{2}{c}(1 - \frac{x}{c})$$

Deberemos repetir 100 veces los siguiente:

- Generar una secuencia de tamaño 1000 de la distribución propuesta.
- Calcular el estadístico chi-cuadrado o de Kolmogoroff-Smirnoff para la secuencia generada
- Resolver los contrastes con un nivel de significación del 5

Una vez realizado lo anterior deberemos estimar la probabilidad de que los estadísticos superen el valor crítico y el promedio de los valores de los estadísticos.

En los casos en que sea posible,dibujar las gráficas de las funciones de densidad.

Calcular la media de las observaciones $\bar{X} = \sum_{i=1}^{n} \frac{X_i}{n}$ para n=50,100,200,2000 para comprobar empíricamente la ley fuerte de los grandes números.

$$\text{lím}_{n\to\infty}(\sum_{i=1}^{n} \frac{X_i}{n}) = \mu = \mathbb{E}[X_i], \quad \text{si } \mathbb{E}[X_i]\text{existe}$$

Dar los valores convenientes de los parámetros de la distribución, en este caso c.

2. La distribución triangular

Notemos en primer lugar que la función de densidad que aparece en el enunciado pertenece a una distribución triangular cuyos parámetros están escogidos de una forma particular.

En general, en la distrubución triangular constamos de tres parámetros: el límite inferior (a),el modo (b) y el límite superior (c). La función de densidad correspondiente es la que sigue:

$$f(x) = \begin{cases} \frac{2(x-a)}{(c-a)(b-a)} & \text{si } a \leq x \leq b \\ \frac{2(c-x)}{(c-a)(c-b)} & \text{si } b \leq x \leq c \end{cases}$$

y cuya función de distribución es:

$$F(x) = \begin{cases} \frac{(x-2)^2}{(c-a)(b-a)} & \text{si } a \leq x \leq b \\ 1 - \frac{(c-x)^2}{(c-a)(c-b)} & \text{si } b \leq x \leq c \end{cases}$$

En nuestro caso, basta tomar $a = b = 0$ y c como parámetro, que será el límite superior para obtener la función de densidad del enunciado:

$$f(x) = \frac{2}{c}(1 - \frac{x}{c})$$

Para obtener la función de distribución bastaría substituír nuestros parámetros particulares en la expresión de la función de distribución triangular.

Supongamos que no disponemos de tal información. En ese caso, bastaría con calcular la integral de la función de densidad:

$$\begin{aligned} F(x) &= \int_0^x f(x)dx = \int_0^x \frac{2}{c}(1 - \frac{x}{c})dx = \\ &= \frac{2}{c}\int_0^x 1dx - \frac{2}{c^2}\int_0^x xdx = \\ &= \left[\frac{2}{c}x\right]_0^x - \left[\frac{x^2}{c^2}\right]_0^x = \frac{2x}{c} - \frac{x^2}{c^2} = \frac{x}{c}\left(2 - \frac{x}{c}\right) \end{aligned}$$

Notemos que la expresión de la función de distribución que nos aparece aparenta ser facilmente invertible, luego ya podemos suponer que el método a utilizar para simular será con casi toda seguridad el método de inversión.

3. Método a utilizar: MÉTODO DE LA INVERSA

Representemos en primer lugar la gráfica de la función de densidad y de la función de distribución, con objeto de ver si esta última pueda ser invertible o no:

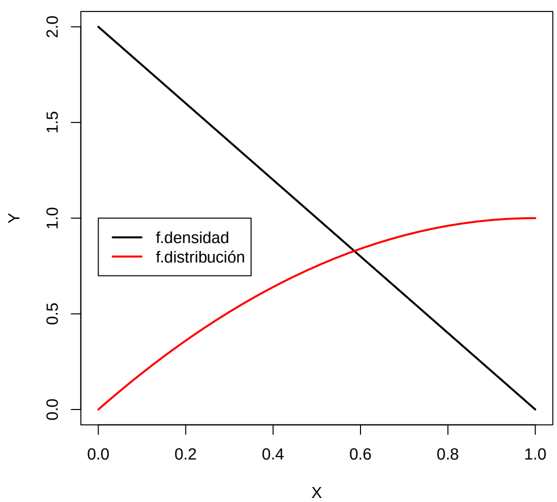

Figura 1: *Función de densidad y función de distribución de una distribución triangular en (0,c),con c=1.*

Como vemos, la función de distribución es claramente invertible. Procedamos entonces a aplicar el método de inversión o de la transformación inversa.

Dicho método está basado en el siguiente resultado teórico:

Teorema 1. (DE INVERSIÓN)

Sea X una variable aleatoria con función de distribución F, continua e invertible. Entonces, la variable aleatoria U=F(X), transformada de la original mediante su propia distribución, tiene distribución U(0,1).
Como consecuencia, si U tiene distuibución uniforme en el intervalo (0,1), entonces la variable aleatoria $F^{-1}(U)$ tiene función de distribución F (la misma distribución que la de X).

Consideremos pues nuestra función de distribución:

$$F(x) = \frac{x}{c}\left(2 - \frac{x}{c}\right)$$

Vemos que la función de distribución es continua e invertible. Consideremos entonces $U = F(X) \in U(0,1)$, es decir:

$$F(x) = \frac{x}{c}\left(2 - \frac{x}{c}\right) = u$$

Despejemos entonces x de nuestra ecuación, obteniendo:

$$-\frac{x^2}{c^2} + \frac{2x}{c} - u = 0 \Leftrightarrow x = \frac{-\frac{2}{c} \pm \sqrt{\frac{4}{c^2} - \frac{4}{c^2}u}}{-\frac{2}{c}} \Leftrightarrow x = \frac{-\frac{2}{c} \pm \frac{2}{c}\sqrt{1-u}}{-\frac{2}{c^2}} \Leftrightarrow x = c \pm c\sqrt{1-u}$$

Dado que x debe estar en el intervalo $(0, a)$,la ecuación tiene solución única (es decir, la inversa está bien definida) y es la siguiente:

$$x = c - c\sqrt{1-u}$$

Dado que $U \sim U(0,1)$, entonces $1-U \sim U(0,1)$, por lo que a la hora de generar la variable aleatoria generaremos U en lugar de $1-U$.Luego la variable a generar será:

$$X = c - c\sqrt{1-U} \quad \text{con} \quad U \sim U(0,1)$$

Por tanto el proceso general será, para cada dato que generemos (n datos),será hacer:

1. Generar $U_i \sim U \sim U(0,1) \quad i \in \{1,...,n\}$
2. Devolver $X_i = c - c\sqrt{U_i} \quad i \in \{1,...,n\}$

Una vez resuelto el tema de la generación de la variable aleatoria, procederemos al contraste de la bondad de ajuste. Dado que se trata de una distribución continua, el contraste más apropiado es el de Kolmogoroff-Smirnoff. Resolveremos los contrastes con un nivel de significación del 5 % y estimaremos la probabilidad de que los estadísticos superen el valor crítico y el promedio de los valores de los estadísticos.

Por último, calcularemos la media de los observaciones para distintos tamaños muestrales ($n = 50, 100, 200, 2000$) para comprobar empíricamente la ley fuerte de los grandes números. Notemos que la media de una distribución triangular general de parámetros a,b,c definida en la sección anterior es:

$$\mathbb{E}(X) = \frac{a+b+c}{3}$$

En nuestro caso,como $a = b = 0$, la expresión se nos reduce a $\mathbb{E}(X) = \frac{c}{3}$

4. Implementación en R y conclusiones

Dibujemos para comenzar nuestra función de distribución.

```
#GRÁFICA PARA PONER EN EL MÉTODO DE INVERSIÓN

curve(f_distr(x),axes=FALSE,0,1,ylab="",xlab="",col="green",lwd=1.9)

axis(1,pos=0)
axis(2,pos=0)

abline(h=1,col="gray40",lwd=2)

title(sub="Distribución triangular(0,1)",col.sub="gray50",cex.sub=0.75,font.sub=4,
line=2)

arrows(0.2,f_distr(0.2),0.2,0,col="dark red",lty="dotted")
segments(0,f_distr(0.2),0.2,f_distr(0.2),col="dark red",lty="dotted")
text(-0.01,f_distr(0.2),"u",col="dark red",cex=0.8)

text(0.2,-0.03,expression({F^-1}(u)==x),col="dark red",cex=0.8,pos=4,offset=-0.6)
```

Así, obtenemos la siguiente gráfica:

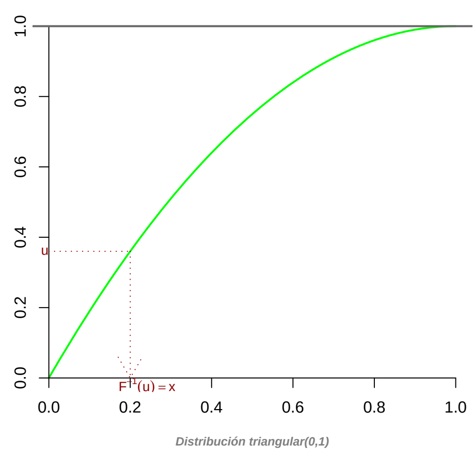

Figura 2: *Función de distribución de una distribución triangular en (0,c),con c=1.*

Pasemos a definir ahora la función de distribución triangular, ya que la necesitaremos a la hora de aplicar el contraste de Kolmogoroff-Smirnoff

```
#FUNCIÓN DE DISTRIBUCIÓN TRIANGULAR
f_distr<-function(x){x/c*(2-x/c)}
```

A continuación implementamos la generación de la variable aleatoria proveniente de la distribución triangular, en función de la variable uniforme $U(0,1)$, proveniente del método de la inversión.Lo haremos en una función, que tendrá como argumentos de entrada el número de variables a generar y el parámetro del que depende, y como salida la variable generada.

```
#GENERACIÓN DE LA VARIABLE ALEATORIA
triangular<-function(n,c){
x<-runif(n,0,1)
x<-c-c*sqrt(x)
return(x)
}
```

Vayamos ahora con el contraste de Kolmogoroff-Smirnoff. Será una función en la cual dejaremos abierta la lista de argumentos entrantes, de manera que no fijamos la función para ninguna distribución en particular,y como argumentos salientes tendremos los p-valores y los distintos valores que obtenemos del estadístico del contraste.

```
#CONTRASTE DE KOLMOGOROFF-SMIRNOFF
TEST_KS<-function(x,...){
result<-ks.test(x,...)
sal<-c(result$p.value,result$statistic)
return(sal)
}
```

Para finalizar completaremos la implementación del método de la inversa, creando un programa desde el que hacer las llamadas necesarias a las funciones anteriormente definidas y completar los puntos requeridos. Empezamos fijando un valor del parámetro, en este caso escogemos $c = 1$. Fijamos el número de réplicas a realizar y llamamos a la función «**triangular**»para generar los valores de la variable aleatoria.Tras esto contrastamos via función «**TEST_KS**»(test de Kolmogoroff-Smirnoff)la bondad del ajuste.

Acto seguido imprimimos por pantalla los p-valores y los valores del estadísto del contraste y calculamos la probabilidad de que los estimadores superen el valor crítico y el promedio de los valores de los estadísticos via el comando «summary».

```
#PROGRAMA
c=1
k=100
x<-triangular(1000*k,c)
dim(x)<-c(1000,k)
res_test<-apply(x,2,TEST_KS,"f_distr")#Obtenemos una matriz de dos filas por cien
columnas.En la primera de ellas tenemos los p-valores y en la segunda fila los
valores del estadístico.
(vector_pvalores<-res_test[1,])
(vector_estadistico<-res_test[2,])

rechazados<-vector_pvalores[vector_pvalores<0.05]#Son los p-valores del contraste
(pvalor_estimador<-length(rechazados)/k)
summary(vector_estadistico)
```

Los resultados que obtenemos son los siguientes:

$$p - valor = 0{,}07$$

lo que nos dice que estamos practicamente en el límite entre aceptar y rechazar la hipótesis de que los datos provengan de una distribución triangular, lo cual no es un buen dato. Aceptamos la hipótesis pero podríamos estar generando mal nuestra variable aleatoria.

Vayamos ahora con la estimación de la media. Notemos que la media teórica es $\mathbb{E}(X) = \frac{c}{3} = \frac{1}{3}$. Los resultados obtenidos para los distintos tamaños muestrales son:

$$\begin{aligned} n = 50 &\quad \longrightarrow \quad mean = 0{,}3350144 \\ n = 100 &\quad \longrightarrow \quad mean = 0{,}3362187 \\ n = 200 &\quad \longrightarrow \quad mean = 0{,}3336682 \\ n = 2000 &\quad \longrightarrow \quad mean = 0{,}3335336 \end{aligned}$$

Por tanto vemos que se verifica la Ley Fuerte de los Grandes Números.