Teoría de probabilidad

Martín Andrés Macías Quintero

Ejercicio 12.7

Sean (X, Y) *uniformes sobre la bola unitaria*, es decir,

$$f_{(X,Y)}(x,y) = \begin{cases} \frac{1}{\pi} \text{ si } x^2 + y^2 \leq 1 \\ \\ 0 \text{ si } x^2 + y^2 > 1 \end{cases}$$

Encuentre la distribución de $R = \sqrt{X^2 + Y^2}$. (*Pista*: Introduzca una variable aleatoria auxiliar definida como $S = Arctan\left(\frac{Y}{X}\right)$.) [Respuesta: $f_R(r) = 2rI_{0,1}(r)$.]

Demostración:

Se introduce la variable aleatoria auxiliar $S = tan^{-1}\left(\frac{Y}{X}\right)$. Sea $g(x,y) = \left(\sqrt{x^2+y^2}, \tan^{-1}\left(\frac{y}{x}\right)\right)$. Si se escribe $r = \sqrt{x^2+y^2}$; $0 \leq r \leq 1$ y $\tan^{-1}\left(\frac{y}{x}\right)$, las inversas serían $g_1^{-1}(r,\theta) = (r\cos\theta, r\sin\theta)$ y $g_2^{-1} = (-r\cos\theta, -r\sin\theta)$ puesto que g no es inyectiva. Sus correspondientes jacobianos serían:

$$\left|J_{g_1^{-1}}\right| = \left|\det\begin{bmatrix} \cos\theta & \sin\theta \\ -r\sin\theta & r\cos\theta \end{bmatrix}\right| = |r| = r \tag{1}$$

$$\left|J_{g_2^{-1}}\right| = \left|\det\begin{bmatrix} -\cos\theta & -\sin\theta \\ r\sin\theta & -r\cos\theta \end{bmatrix}\right| = |r| = r \tag{2}$$

De esta forma, por el Corolario 12.1 (*Probability Essentials* de Jacod & Protter) se tiene:

$$\begin{aligned} f_{(R,\Theta)}(r,\theta) &= r\{f_{X,Y}(r\cos\theta, r\sin\theta) + f_{X,Y}(-r\cos\theta, -r\sin\theta)\}I_{(0,1)}(r)I_{\left(\frac{-\pi}{2},\frac{\pi}{2}\right)}(\theta) \\ &= rI_{(0,1)}(r)\left(\frac{1}{\pi}+\frac{1}{\pi}\right)I_{\left(\frac{-\pi}{2},\frac{\pi}{2}\right)}(\theta) \\ &= \frac{2rI_{(0,1)}(r)I_{\left(\frac{-\pi}{2},\frac{\pi}{2}\right)}(\theta)}{\pi} \end{aligned} \tag{3}$$

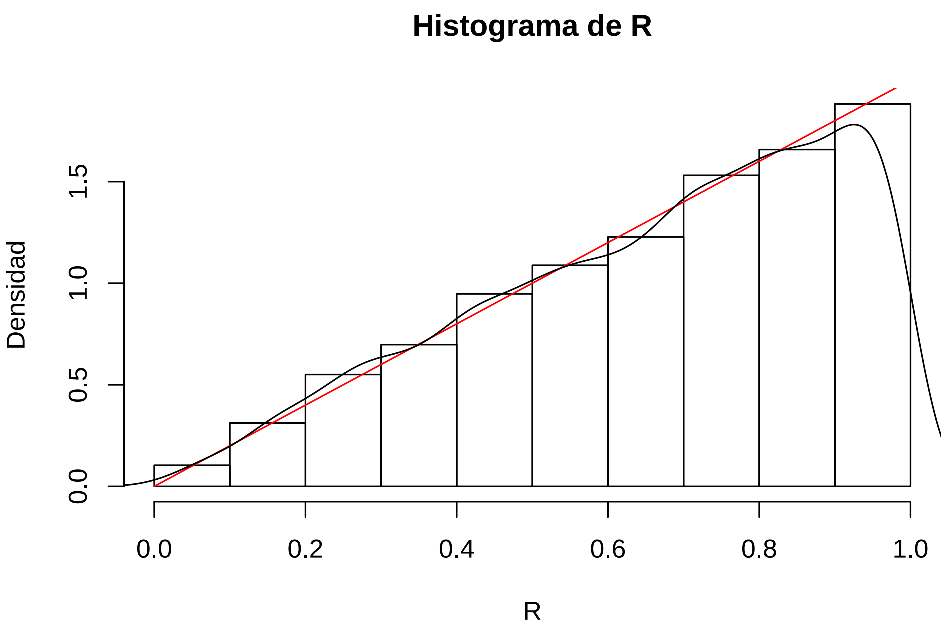
La densidad marginal sería:

$$f_R(r) = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{2rI_{(0,1)}(r)}{\pi} d\theta = 2rI_{(0,1)}(r) \tag{4}$$

Simulación:

Sean X y Y dos variables aleatorias con distribución $U(0,1)$

```
set.seed(163053)
X<-runif(10000)
Y<-runif(10000)

R<-sqrt(X^2+Y^2)
R<-R[R<=1]

hist(R,freq = F, main = "Histograma de R", xlab="R", ylab = "Densidad")
curve(2*x,from = 0,to = 1,col=2,add=T)
lines(density(R))
```



Gracias a la simulación, es posible apreciar que los datos de R simulados a partir de dos variables uniformes, se ajustan a la distribución teórica. La línea de color rojo es la teórica y la de color negro es la de la simulación.

Ejercicio 12.11

Sean (X,Y) normales independientes, ambas con media $\mu = 0$ y varianza σ^2. Sea

$$Z = \sqrt{X^2+Y^2} \quad \text{y} \quad W = Arctan\left(\frac{X}{Y}\right), \qquad -\frac{\pi}{2} < W \leq \frac{\pi}{2}.$$

Demuestre que Z tiene distribución Rayleigh, que W es uniforme sobre $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ y que Z y W son independientes.

Demostración:

Sea $g(x,y) = \left(\sqrt{x^2+y^2}\right), \tan^{-1}\left(\frac{x}{y}\right)$

En este caso las inversas serían $g_1^{-1}(z,w) = (z\sin w, z\cos w)$ y $g_2^{-1}(z,w) = (z\sin w, -z\cos w)$ puesto que g no es inyectiva. Sus correspondientes jacobianos serían:

$$\left|J_{g_1^{-1}}\right| = \left|\det\begin{bmatrix}\sin w & z\cos w\\ \cos w & -z\sin w\end{bmatrix}\right| = |-z| = z \tag{5}$$

$$\left|J_{g_2^{-1}}\right| = \left|\det\begin{bmatrix}\sin w & z\cos w\\ -\cos w & z\sin w\end{bmatrix}\right| = |z| = z \tag{6}$$

De esta forma, se tiene que:

$$\begin{aligned} f_{(Z,W)}(z,w) &= \left(\frac{1}{2\pi\sigma^2}e^{\frac{-z^2}{2\sigma^2}}z + \frac{1}{2\pi\sigma^2}e^{\frac{-z^2}{2\sigma^2}}z\right) I_{\left(-\frac{\pi}{2},\frac{\pi}{2}\right)}(w) I_{(0,\infty)}(z) \\ &= \frac{1}{\pi} I_{\left(-\frac{\pi}{2},\frac{\pi}{2}\right)}(w) \cdot \frac{z}{\sigma^2}e^{\frac{-z^2}{2\sigma^2}} I_{(0,\infty)}(z) \end{aligned} \tag{7}$$

Lo que demuestra que Z tiene distribución Rayleigh(σ^2), W es uniforme sobre el intervalo $\left(-\frac{\pi}{2},\frac{\pi}{2}\right)$ y que Z y W son independientes.

Simulación:

Sean X y Y dos variables aleatorias independientes con distribución normal con media $\mu = 0$ y varianza $\sigma^2 = 4$

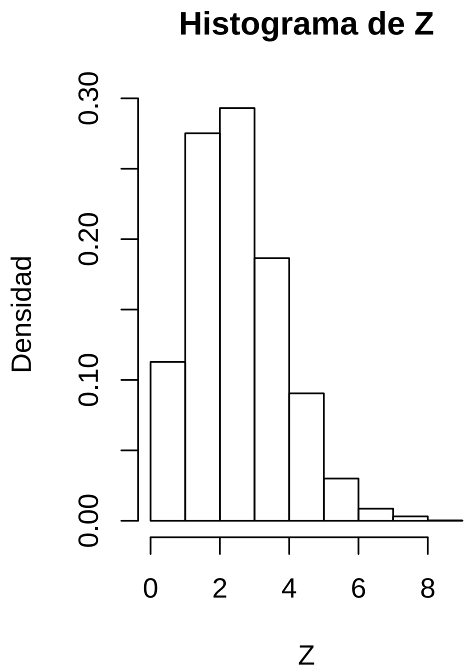
```r
library(VGAM)

set.seed(163053)

X<-rnorm(10000,0,2)
Y<-rnorm(10000,0,2)
Z<-sqrt(X^2+Y^2)
```

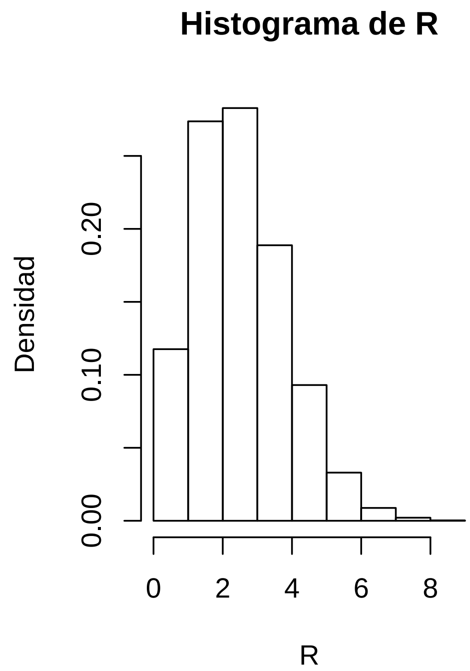
A continuación se simulan datos de una distribución Rayleigh con $\sigma = 2$:

```r
R<-rrayleigh(10000,2)
```

Al comparar los dos histogramas de los datos simulados a partir de X y Y con los de Z:

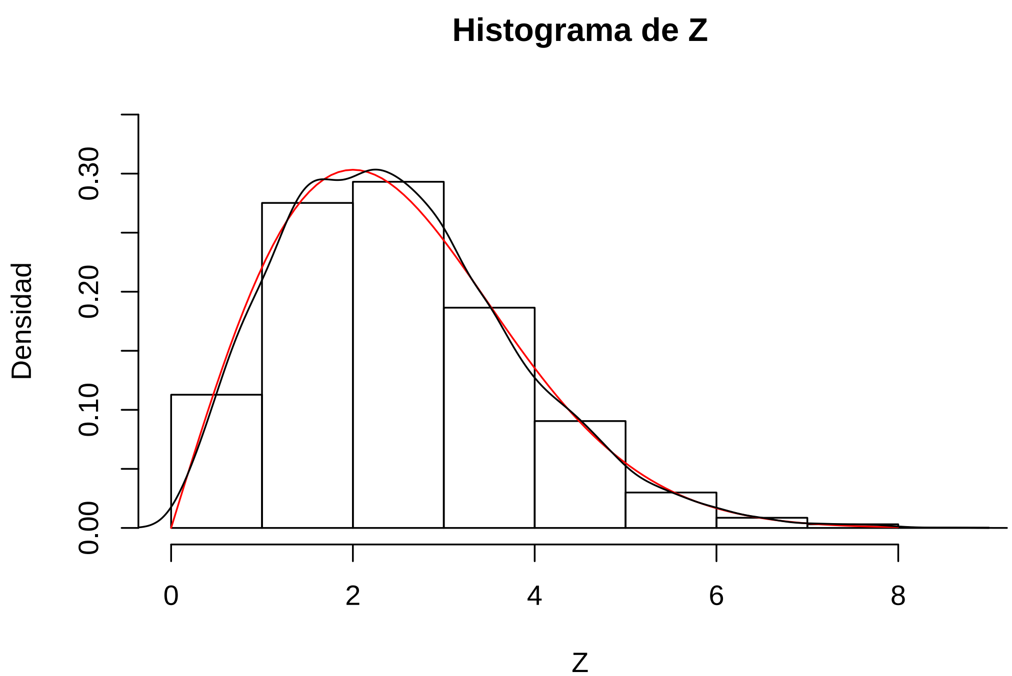
```r
par(mfrow=c(1,2))
hist(Z,freq = F,breaks = 10, main = "Histograma de Z", xlab="Z", ylab = "Densidad")


hist(R,freq = F,breaks = 10, main = "Histograma de R", xlab="R", ylab = "Densidad")
```





Puede verse que los dos histogramas son similares, es decir, que tanto los datos simulados (Z) a partir de dos normales como los obtenidos (R) son equivalentes. Si se comparan con la función de densidad teórica:

```
rayleigh<-function(x,y){
  (x/y^2)*exp((-x^2)/(2*y^2))
}

par(mfrow=c(1,1))
hist(Z,freq = F,breaks = 10,ylim = c(0,0.35),
     main = "Histograma de Z", xlab="Z", ylab = "Densidad")
curve(rayleigh(x,2),from = 0,to = 8,add = T,col=2)
lines(density(Z))
```



```r
ks.test(Z,prayleigh,2)
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  Z
## D = 0.0091383, p-value = 0.3739
## alternative hypothesis: two-sided
```

```r
test <- ks.test(Z,prayleigh,2)
```

De esta forma, es posible confirmar que los datos de la distribución Rayleigh simulados a partir de dos normales coninciden con la función de densidad de una distribución Rayleigh. La línea de color rojo es la teórica y la de color negro es la de la simulación. Igualmente, se utiliza el test de Kolmogorov-Smirnov, con un $\alpha = 0.05$ cuya hipótesis nula afirma que los datos siguen la distribución deseada, en este caso Rayleigh. Como puede verse, el p-valor 0.3739197 permite no rechazar la hipótesis nula indicada, de tal suerte que la prueba confirma que los datos siguen una distribución Rayleigh.

Ahora, sean X y Y dos variables aleatorias normales independientes con media $\mu = 0$ y varianza $\sigma^2 = 4$

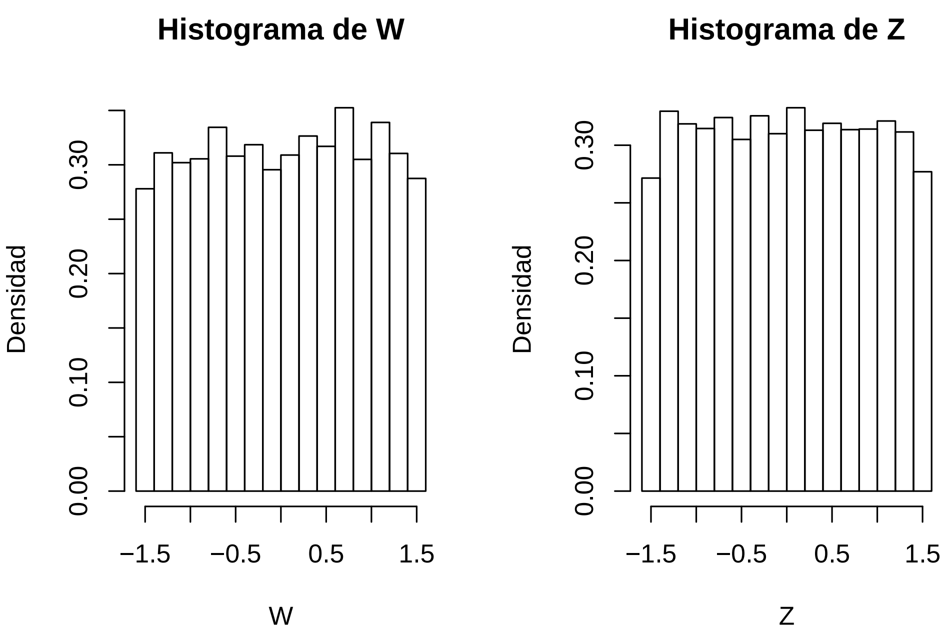
```r
set.seed(163053)

X<-rnorm(10000,0,2)
Y<-rnorm(10000,0,2)

U<-X/Y
W<-atan(U)
```

Al simular datos Z de una distribución uniforme $U(-\pi/2, \pi/2)$, se tiene:.

```r
Z<-runif(10000,-pi/2,pi/2)
```

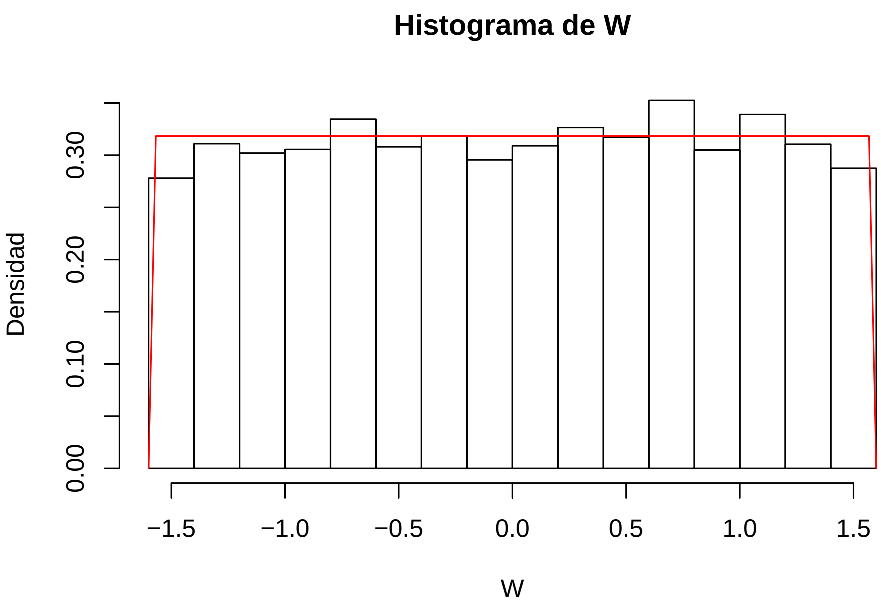
Al comparar los dos histogramas de los datos simulados a partir X y Y y los de Z:

```r
par(mfrow=c(1,2))
hist(W,freq = FALSE, main = "Histograma de W", xlab="W", ylab = "Densidad")
hist(Z,freq = FALSE, main = "Histograma de Z", xlab="Z", ylab = "Densidad")
```



Se observa que los dos histogramas son similares, es decir, que tanto los simulados W a partir de dos normales y los obtenidos (Z) son equivalentes. Al comparar con la función de densidad teórica:

```
unifor<-function(x,y,z){
  if(x>=y&x<=z){1/(z-y)}
  else{0}
}

unif<-Vectorize(unifor)

par(mfrow=c(1,1))
hist(W,freq = FALSE, main = "Histograma de W", xlab="W", ylab = "Densidad")
curve(unif(x,-pi/2,pi/2),add = T,col=2)
```




```
ks.test(W,punif,-pi/2,pi/2)
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  W
## D = 0.012268, p-value = 0.09855
## alternative hypothesis: two-sided
```

```
test <- ks.test(W,punif,-pi/2,pi/2)
```

Se confirma que los datos de la distribución uniforme entre $-\pi/2$ y $\pi/2$ simulados a partir de dos normales coninciden con la función de densidad de la una distribución uniforme. La línea de color rojo es la teórica. Igualmente, se utiliza el test de Kolmogorov-Smirnov, con un $\alpha = 0.05$ cuya hipótesis nula afirma que los datos siguen la distribución deseada. En este caso, el p-valor 0.0985464 permite no rechazar la hipótesis nula indicada y gracias a esta prueba, ratificar que los datos siguen distribución uniforme.

Ejercicio 12.15

(*Simulación de variables aleatorias normales*) Sean U, V dos variables aleatorias uniformes independientes sobre $(0,1)$. Sean $\theta = 2\pi U$ y $S = -\ln(V)$.

(*Pista:* Para la parte *(a)* recuerde que una exponencial es un caso especial de una distribución Gamma: de hecho, esto es χ_2^2. Para la parte (b) invierta el procedimiento del esjericio 12.11).

Observación: El ejercicio 12.15 es conocido como el método Box-Muller para simular variables aleatorias normales.

a. Demuestre que S tiene distribución exponencial, y que $R = \sqrt{2S}$ tiene distribución Rayleigh.

Demostración:

Sea $V \sim U(0,1)$ y $S = -\ln(V)$, luego

$$s = g(v) = -\ln(v) \Rightarrow v = h(s) = e^{-s} \tag{8}$$

Así,

$$|h'(s)| = \left|\frac{dh(s)}{ds}\right| = \left|-e^{-s}\right| = e^{-s} \tag{9}$$

Como

$$f_V(v) = 1I_{(0,1)} \quad f_S(s) = f_V(e^{-s}) \cdot \left|\frac{dv}{ds}\right| = 1 \cdot e^{-s} = e^{-s} \tag{10}$$

De tal suerte que puede concluirse que $S \sim Exp(1)$

Ahora,

$$\begin{aligned} R &= \sqrt{2S} = \sqrt{-2\ln(V)} \\ R &= h(v) = \sqrt{-2\ln(V)} \end{aligned} \tag{11}$$

Luego,

$$\begin{aligned} r = g(v) &= \sqrt{-2\ln(v)} \\ r^2 &= -2\ln(v) \\ -\frac{r^2}{2} &= \ln(v) \\ v = h(r) &= e^{-\frac{r^2}{2}} \end{aligned} \tag{12}$$

De esta forma,

$$|h'(s)| = \left|\frac{dh(s)}{ds}\right| = \left|= -r \cdot e^{-\frac{r^2}{2}}\right| = r \cdot e^{-\frac{r^2}{2}}$$

Entonces,

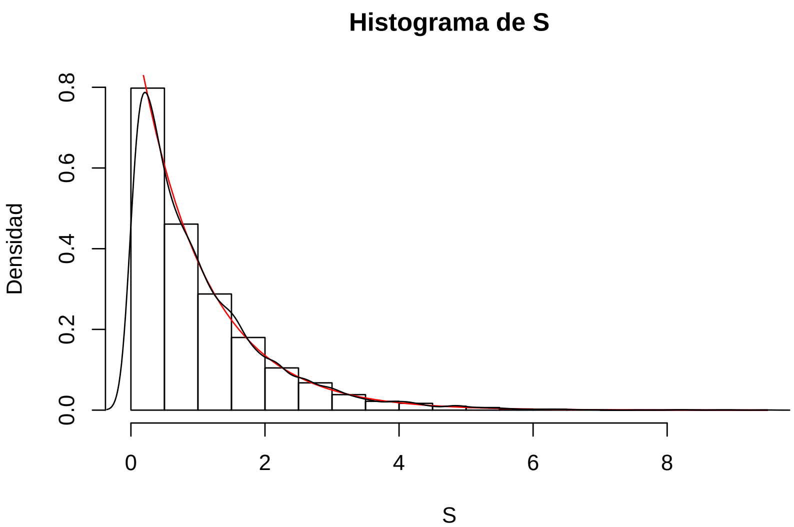
$$f_V(v) = 1I_{(0,1)} \quad f_R(r) = f_V(e^{-\frac{r^2}{2}}) \cdot \left|\frac{dv}{ds}\right| = 1 \cdot re^{-\frac{r^2}{2}} = r \cdot e^{-\frac{r^2}{2}}$$

Que claramente se refiere a una variable aleatoria $R \sim Rayleigh(1)$

Simulación:

Sea V una variable aleatoria con distribución uniforme en $(0, 1)$

```
set.seed(163053)
V<-runif(10000)
S<--log(V)
R<-sqrt(2*S)

hist(S,freq = FALSE, main = "Histograma de S", xlab="S", ylab = "Densidad")
curve(dexp(x,1),add=T,col=2)
lines(density(S))
```




```
ks.test(S,pexp,1)
```

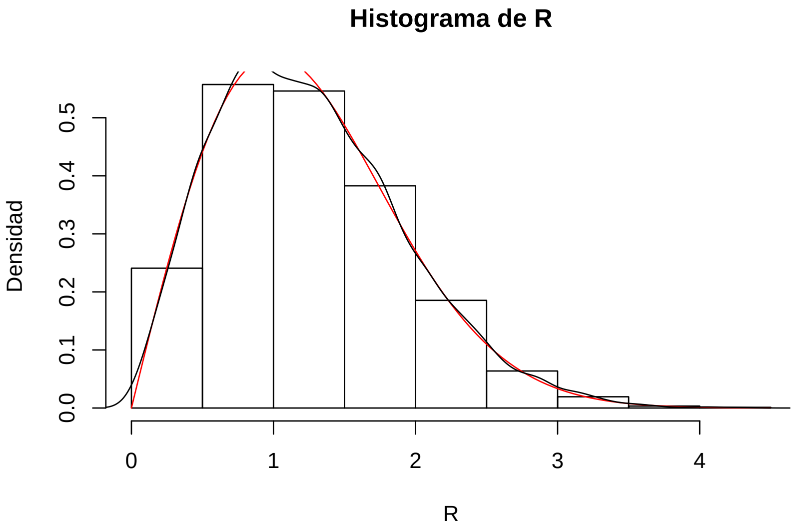
```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  S
## D = 0.0081251, p-value = 0.5239
## alternative hypothesis: two-sided
```

```
test <- ks.test(S,pexp,1)
```

Se observa que los datos de S simulados a partir de una distribución uniforme, se ajusta a una distribución exponencial con parámetro 1. Una vez más, se utiliza el test de Kolmogorov-Smirnov, con un $\alpha = 0.05$ cuya hipótesis nula afirma que los datos siguen la distribución deseada. En este caso, el p-valor 0.5239289 permite no rechazar la hipótesis nula indicada y gracias a esta prueba, ratificar que los datos siguen distribución exponencial.

Por otra parte, los datos simulados de R

```
hist(R,freq = FALSE, main = "Histograma de R", xlab="R", ylab = "Densidad")
curve(drayleigh(x,scale = 1),add=T,col=2)
lines(density(R))
```




```
ks.test(R,prayleigh,2)
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  R
## D = 0.47318, p-value < 2.2e-16
## alternative hypothesis: two-sided
```

```
test <- ks.test(R,prayleigh,2)
```

Se observa que los datos de R simulados a partir de una distribución exponencial de parámetro 1, se ajusta a una distribución Rayleigh con parámetro 1. Una vez más, se utiliza el test de Kolmogorov-Smirnov, con un $\alpha = 0.05$ cuya hipótesis nula afirma que los datos siguen la distribución deseada. En este caso, el p-valor 0 permite no rechazar la hipótesis nula indicada y gracias a esta prueba, ratificar que los datos siguen distribución Rayleigh.

b. Sean $X = R\cos\theta$, $Y = R\sin\theta$. Demuestre que X y Y son normales independientes.

Demostración:

Se sabe que $X = \sqrt{-2\ln(V)}\cos(2\pi U)$ y $Y = \sqrt{-2\ln(V)}\sin(2\pi U)$.

Así, $g(u,v) = (\sqrt{-2\ln(v)}\cos(2\pi u), \sqrt{-2\ln(v)}\sin(2\pi u))$

A continuación se calculan la inversa de g:

$$\frac{Y}{X} = \frac{\sqrt{-2\ln(v)}\cos(2\pi u)}{\sqrt{-2\ln(v)}\sin(2\pi u)} = \tan(2\pi u) \Rightarrow u = \frac{1}{2\pi}\tan^{-1}\left(\frac{y}{x}\right) \tag{13}$$

Por otra parte,

$$
\begin{aligned}
X^2 + Y^2 &= (\sqrt{-2\ln(v)}\cos(2\pi u)^2 + (\sqrt{-2\ln(v)}\sin(2\pi u)^2 \\
&= (-2\ln(v)\cos^2(2\pi u)) + (-2\ln(v)\sin^2(2\pi u) \\
&= -2\ln(v)(\cos^2(2\pi u) + \sin^2(2\pi u) \\
&= -2\ln(v)
\end{aligned} \tag{14}
$$

Así,

$$
v = e^{-\frac{x^2+y^2}{2}} \tag{15}
$$

De esta forma, $g^{-1}(x,y) = \left(\frac{1}{2\pi}\tan^{-1}\left(\frac{y}{x}\right), e^{-\frac{x^2+y^2}{2}}\right)$

El cálculo del jacobiano se presenta a continuación:

$$
\begin{aligned}
\left|J_{g^{-1}}\right| &= \left|\det\begin{bmatrix} \frac{y}{2\pi(x^2+y^2)} & -\frac{x}{2\pi(x^2+y^2)} \\ -x\cdot e^{-\frac{x^2+y^2}{2}} & -y\cdot e^{-\frac{x^2+y^2}{2}} \end{bmatrix}\right| \\
&= \left|\frac{-y^2\cdot e^{-\frac{x^2+y^2}{2}}}{2\pi(x^2+y^2)} - \frac{-x^2\cdot e^{-\frac{x^2+y^2}{2}}}{2\pi(x^2+y^2)}\right| \\
&= \left|\frac{-y^2\cdot e^{-\frac{x^2+y^2}{2}} - x^2\cdot e^{-\frac{x^2+y^2}{2}}}{2\pi(x^2+y^2)}\right| \\
&= \left|\frac{-e^{-\frac{x^2+y^2}{2}}(x^2+y^2)}{2\pi(x^2+y^2)}\right| \\
&= \frac{1}{2\pi}e^{-\frac{x^2+y^2}{2}}
\end{aligned} \tag{16}
$$

De esta forma, la densidad conjunta de X y Y es:

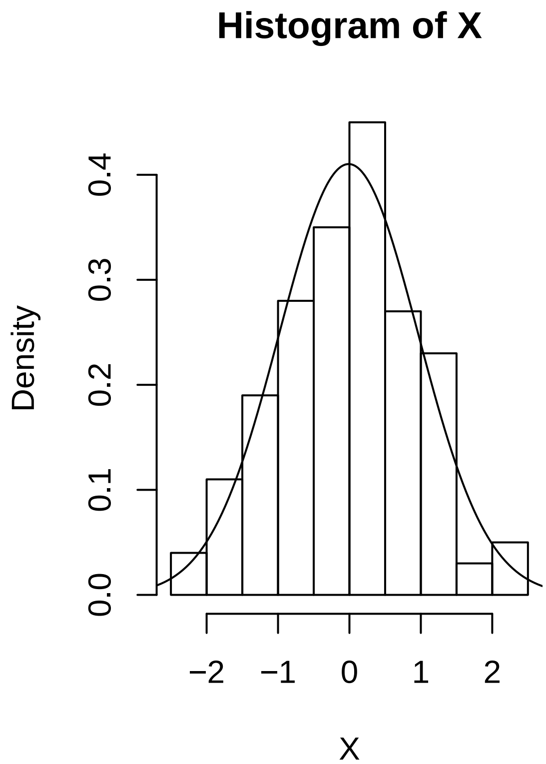
$$
f_{X,Y} = \frac{1}{2\pi}e^{-\frac{x^2+y^2}{2}} = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}} = f(x)f(y) \tag{17}
$$

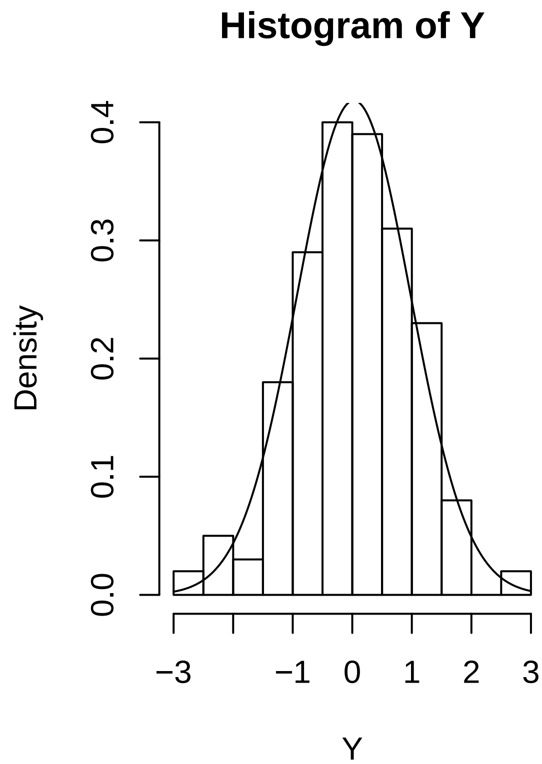
Lo que demuestra que X y Y se distribuyen normales estándar y además son independientes.

Simulación:

Sean U y V variables aleatorias uniformes en $(0,1)$

```
n = 200
U=runif(n)
V=runif(n)
X=sqrt(-2*log(U))*cos(2*pi*V)
Y=sqrt(-2*log(V))*cos(2*pi*U)
x=seq(-3,3,by=.01)
par(mfrow=c(1,2))
hist(X,freq = F);lines(x,dnorm(x,mean(X),sd(X)))
hist(Y,freq = F);lines(x,dnorm(x,mean(Y),sd(Y)))
```





Puede observarse que las variables simuladas de X y Y, corresponden a variables normales con media 0 y varianza 4.

```
ks.test(X,pnorm)
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  X
## D = 0.034158, p-value = 0.9738
## alternative hypothesis: two-sided
```

```
test1 <- ks.test(X,pnorm)
ks.test(Y,pnorm)
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  Y
## D = 0.049467, p-value = 0.712
## alternative hypothesis: two-sided
```

```
test2 <- ks.test(Y,pnorm)
```

Se utiliza el test de Kolmogorov-Smirnov, con un $\alpha = 0.05$ cuya hipótesis nula afirma que los datos siguen la distribución deseada. En este caso, los p-valores 0.9737566 y 0.7119575 permiten no rechazar la hipótesis nula indicada y gracias a esta prueba, ratificar que los datos siguen distribución Normal para cada una de las variables X y Y.