Una distribución muy usada para el análisis de datos bimodales es la distribución de mixtura de normales cuya función de densidad es dada por

$$f_X(x) = p\phi(x|\mu_1, \sigma_1^2) + (1-p)\phi(x|\mu_2, \sigma_2^2), x \in \mathbb{R}$$

donde $\mu_1, \mu_2 \in \mathbb{R}$, $\sigma_1^2, \sigma_2^2 > 0$, $p \in (0,1)$ y $\phi(.|a,b^2)$ representa la densidad de una distribución normal con media a y varianza b^2 . Utilizaremos la siguiente notación para una variable aleatoria X que siga esta distribución: $X \sim \text{MN}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, p)$.

a) Encuentre una expresión para la media y la varianza de esta distribución.

Solución. Definimos $X_i \sim N(\mu_i, \sigma_i^2)$ y $\phi_i = \phi(x|\mu_i, \sigma_i^2)$, teniendo lo siguiente:

$$E[X_i] = \int_{-\infty}^{\infty} x \phi_i = \mu_i \tag{1}$$

$$Var[X_i] = E[X_i^2] - E[X_i]^2 = \int_{-\infty}^{\infty} x^2 \phi_i - \mu_i^2 = \sigma_i^2$$
 (2)

Procedemos a hallar la media de X:

$$E[X] = \int_{-\infty}^{\infty} x(p\phi_1 + (1-p)\phi_2)dx$$

$$= \int_{-\infty}^{\infty} (xp\phi_1 + x(1-p)\phi_2)dx$$

$$= p\int_{-\infty}^{\infty} x\phi_1 dx + (1-p)\int_{-\infty}^{\infty} x\phi_2 dx$$

$$\stackrel{\text{por } (1)}{=} p\mu_1 + (1-p)\mu_2$$

$$(3)$$

Para hallar la varianza, primero hallamos E $\left[X^{2}\right]$:

$$E[X^{2}] = \int_{-\infty}^{\infty} x^{2} (p\phi_{1} + (1-p)\phi_{2}) dx$$

$$= \int_{-\infty}^{\infty} (x^{2}p\phi_{1} + x^{2}(1-p)\phi_{2}) dx$$

$$= p \int_{-\infty}^{\infty} x^{2}\phi_{1} dx + (1-p) \int_{-\infty}^{\infty} x^{2}\phi_{2} dx$$

$$\stackrel{\text{por}}{=} (2) = p \left(\sigma_{1}^{2} + \mu_{1}^{2}\right) + (1-p) \left(\sigma_{2}^{2} + \mu_{2}^{2}\right)$$
(4)

Por tanto, usando (3) y (4):

$$Var[X] = E[X^{2}] - E[X]^{2}$$

$$= p(\sigma_{1}^{2} + \mu_{1}^{2}) + (1 - p)(\sigma_{2}^{2} + \mu_{2}^{2}) - (p\mu_{1} + (1 - p)\mu_{2})^{2}$$

$$= p(1 - p)(\mu_{1} - \mu_{2})^{2} + p\sigma_{1}^{2} + (1 - p)\sigma_{2}^{2}.$$
(5)

b) Pruebe la siguiente propiedad: Sea $W \sim \text{Bernoulli}(p)$ y X otra v.a. tal que su distribución condicional a W es dada por

$$X|W = 1 \sim N(\mu_1, \sigma_1^2)$$

 $X|W = 0 \sim N(\mu_2, \sigma_2^2)$

entonces $X \sim \text{MN}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, p)$.

Solución. Tenemos:

$$f_W(w) = \begin{cases} p, & \text{, si } w = 1\\ 1 - p, & \text{, si } w = 0 \end{cases}$$

$$(6)$$

$$f_{X|W}(x|w) = \begin{cases} \phi_1 & , \text{ si } w = 1\\ \phi_2 & , \text{ si } w = 0 \end{cases}$$

$$(7)$$

Nos interesa calcular f_X , la densidad marginal de X, por lo que usamos la regla de probabilidad total con la partición dada por los valores de W:

$$f_{X}(x) = f_{XW}(x,1) + f_{XW}(x,0)$$
Regla del producto
$$= f_{W}(1)f_{X|W}(x|1) + f_{W}(0)f_{X|W}(x|0)$$

$$por (6) y (7)$$

$$= p\phi_{1} + (1-p)\phi_{2}$$

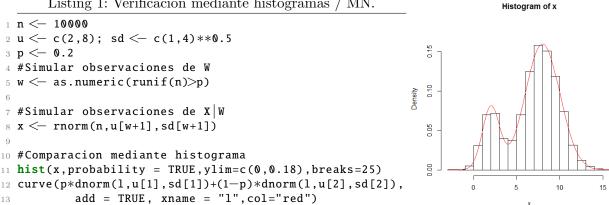
$$= p\phi(x|\mu_{1}, \sigma_{1}^{2}) + (1-p)\phi(x|\mu_{2}, \sigma_{2}^{2})$$
(8)

Concluimos que $X \sim \text{MN}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, p)$.

c) Utilice la propiedad dada en b) para proponer un algoritmo para simular valores de una v.a. $X \sim \text{MN}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, p)$. Implemente el algoritmo propuesto en R. Simule 10 000 valores de una MN(2, 8, 1, 4, 0.2) y verifique mediante gráficos de histograma con función de densidad y cuantiles que el método propuesto funciona correctamente.

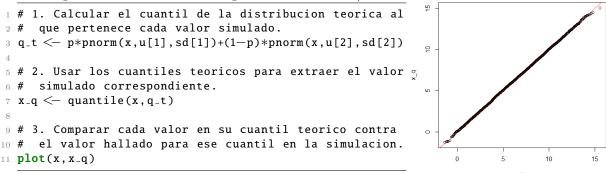
Solución. Una manera de simular X consiste en generar primero un valor de W y luego uno de X|W=w. El segundo valor tendrá la distribución deseada.

Listing 1: Verificación mediante histogramas / MN.



Vemos que el histograma de los valores simulados se ajusta bien a la curva de densidad de probabilidad de la distribución objetivo.

Listing 2: Verificación mediante gráfico de cuantiles / MN.



La gráfica de cuantiles muestra una línea casi perfecta, lo cual significa que el valor simulado y el valor teórico en cada cuantil tienen gran cercanía y, por lo tanto, nuestra simulación reproduce adecuadamente la distribución objetivo.

d) Presente los pasos que debe seguir el algoritmo de Metropolis-Hastings considerando como distribución generadora de candidatos una distribución normal centrada en el punto anterior. Implemente en R el algoritmo propuesto. Simule 10 000 valores de una MN(2, 8, 1, 4, 0.2) y verifique mediante gráficos de cadenas, histograma con función de densidad y cuantiles que el método propuesto funciona correctamente.

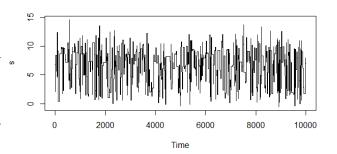
Solución.

Listing 3: Algoritmo de Metropolis-Hastings.

```
3 # Valor inicial basado en la media de la distribucion objetivo
s < c(p*u[1]+(1-p)*u[2],
5 numeric(n))
  # Recomendacion de Gelman para la varianza de distribucion generadora
s \, sdq \leftarrow (p*(sd[1]**2 + u[1]**2) +
9 (1-p)*(sd[2]**2 + u[2]**2)-
10 (p*u[1]+(1-p)*u[2])**2)*(2.4**2)
11
12 # Definimos funciones objetivo y candidato
13 d_obj \leftarrow function(x) \{p*dnorm(x,u[1],sd[1])+(1-p)*dnorm(x,u[2],sd[2])\}
14 d_{can} \leftarrow function(x,uc) \{dnorm(x,uc,sdq)\}
15
16 # Definimos funcion alpha
17 alpha \leftarrow function(x,y,f,q)
  \min(1, f(y)*q(x,y)/(f(x)*q(y,x)))
19
20
21 # Bucle central del algoritmo
22 for(i in 1:(length(s)-1)){
24 # Obtener un valor de la distribucion generadora, condicionado a valor previo.
25 candidato <- rnorm(1,s[i],sdq)</pre>
26
```

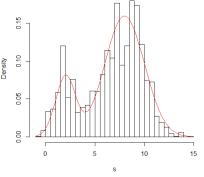
Listing 4: Revisión del gráfico de cadena / MN.

```
1 ts.plot(s)
2 # Distribucion parece estable
3 # desde el inicio, no haria falta
4 # retirar observaciones.
```



Listing 5: Verificación mediante histograma / MN.

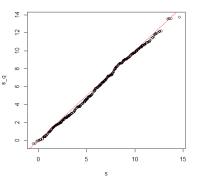
```
hist(s,probability = TRUE,ylim=c(0,0.18),breaks=25)
curve(p*dnorm(1,u[1],sd[1])+(1-p)*dnorm(1,u[2],sd[2]),
dadd = TRUE, xname = "l",col="red")
```



Histogram of s

Listing 6: Verificación mediante gráfica de cuantiles / MN.

```
# 1. Calcular el cuantil de la distribucion teorica al
2 # que pertenece cada valor simulado.
3 q_t <- p*pnorm(s,u[1],sd[1])+(1-p)*pnorm(s,u[2],sd[2])
4
5 # 2. Usar los cuantiles teoricos para extraer el valor
6 # simulado correspondiente.
7 s_q <- quantile(s,q_t)
8
9 # 3. Comparar cada valor en su cuantil teorico contra
10 # el valor hallado para ese cuantil en la simulacion.
11 plot(s,s_q)</pre>
```



Vemos que los valores simulados por Metropolis-Hastings tienen cierta cercanía con la curva teórica en el histograma y se mantienen cercanos a la diagonal en la gráfica de cuantiles, pero en general su ajuste no es tan bueno como el método anterior. \Box

Una variable aleatoria definida en toda la recta, tiene distribución normal asimétrica (Azzalini, 1985) con parámetro de asimetría λ si su función de densidad es dada por la siguiente expresión:

$$f_X(x) = 2\phi(x)\Phi(\lambda x), x \in \mathbb{R}.$$

donde $\phi(\cdot)$ es la función de densidad y $\Phi(\cdot)$ es la función de distribución acumulada de una distribución normal estándar y $\lambda \in \mathbb{R}$. Se utiliza la notación $X \sim SN(\lambda)$.

a) Pruebe que una distribución normal estándar puede ser considerada como distribución generadora de candidatos en el algoritmo de aceptación y rechazo.

Solución. La implementación del método de aceptación y rechazo para la simulación de $f_X(x)$ toma una función $g_X(x)$, llamada generadora, la cual es fácil de simular y con la que debe cumplirse lo siguiente:

$$\frac{f_X(x)}{g_X(x)} \le c, \ \forall x \in \mathbb{R}. \tag{9}$$

para una constante c. Veamos lo siguiente:

$$\frac{f_X(x)}{\phi(x)} = \frac{2\phi(x)\Phi(\lambda x)}{\phi(x)}$$

$$= 2\Phi(\lambda x)$$

$$= 2\int_{-\infty}^{\lambda x} \phi(t)dt$$

$$\leq 2\underbrace{\int_{-\infty}^{\infty} \phi(t)dt}_{=1}$$

$$= 2.$$
(10)

Tenemos que $\frac{f_X(x)}{\phi(x)} \le 2$, por lo que $\phi(x)$, además de ser fácil de simular, puede ser considerada como función generadora dado que cumple (9).

b) Presente los pasos que debe seguir el algoritmo de aceptación y rechazo considerando como distribución generadora de candidatos una distribución normal estándar. Implemente en R el algoritmo propuesto. Simule 10000 valores de una SN(8) y verifique mediante gráficos de histograma con función de densidad y cuantiles que el método propuesto funciona correctamente.

Solution. La función a simular es la siguiente:

$$f_X(x) = 2\phi(x)\Phi(8x), \ x \in \mathbb{R}. \tag{11}$$

Los pasos que se deberían seguir en el algoritmo de aceptación y rechazo serían los siguientes:

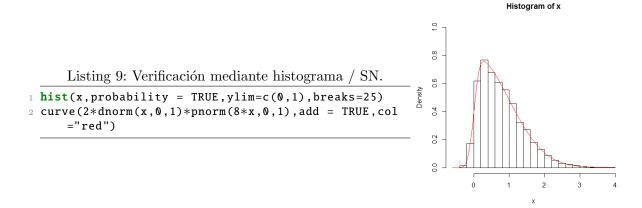
- 1) Generamos $y \sim N(0, 1)$.
- 2) Generamos $u \sim U(0,1)$.
- 3) Si $u \leq \frac{f_X(x)}{2\phi(x)} = \Phi(8x) \Rightarrow x = y$. Caso contrario ir al paso 1.

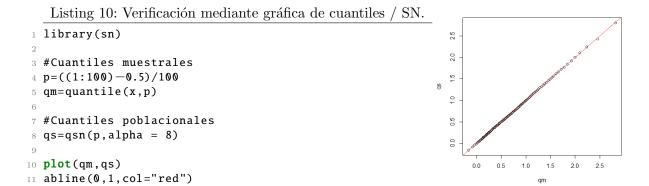
A continuación, se presenta un gráfico de la función a simular y la función generadora

Distribución SN con propuesta normal estandar Listing 7: Distribución de interés y distribución normal estándar / SN. cq(y) curve(2*dnorm(x,0,1)*pnorm(8*x,0,1),0,3,col=1,lwd=2, 0.1 ylim=c(0,1.5),2 xla="x", 3 ylab="Densidad"; 0.5 4 main="Distribucion SN con propuesta normal estandar") 5 curve(2*dnorm(x,0,1),col=2,lwd=2,add=T) 6 legend(1.5,1.5,c("f(y)","cg(y)"),col=c(1,2),lty=1,lwd =2,cex=1.2)0.0 3.0

Listing 8: Algoritmo de aceptación y rechazo.

```
_{1} M<-10000
2 x<-numeric(M)</pre>
4 for(h in 1:M){
     cond < -0
     while(cond==0){
       y<-rnorm(1) # Generar un candidato
       u \leftarrow runif(1) \# Generar u \sim U(0,1)
       if(u \le pnorm(8*v, 0, 1))
9
          # Aceptar el candidato
10
          x[h] < -y
11
          cond < -1
12
14
15
```





Se observa, según el histograma, que efectivamente nuestros datos se ajustan correctamente usando una simulación con el método de aceptación y rechazo implementándose como función generadora a la función de densidad normal estandar, $g_X(x) = \phi(x)$.

Además, una segunda verificación, mediante un gráfico de cuantiles, nos indica la alta correspondencia entre los datos teóricos y los simulados, corroborando de nuevo la correcta simulación.

Una variable aleatoria X definida en toda la recta, tiene distribución logística potencia con parámetro de asimetría α si su función de densidad es dada por la siguiente expresión:

$$f_X(x) = \alpha l(x) L(x)^{\alpha - 1}, x \in \mathbb{R}.$$

donde $l(x) = \frac{e^x}{(1+e^x)^2}$ es la función densidad y $L(x) = \frac{e^x}{1+e^x}$ es la función de distribución acumulada de una distribución logística estándar, y $\alpha > 0$. Se utiliza la notación $X \sim \text{PL}(\alpha)$.

a) Encuentre la distribución acumulada de una $PL(\alpha)$.

Solución. Resolviendo $f_X(x)$:

$$f_X(x) = \alpha l(x)L(x)^{\alpha-1} = \alpha \frac{e^x}{(1+e^x)^2} \left(\frac{e^x}{1+e^x}\right)^{\alpha-1} = \alpha e^{-x} (1+e^{-x})^{-\alpha-1}.$$
 (12)

Derivando L(x) tenemos lo siguiente:

$$\frac{d(L(x))}{dx} = \frac{d\left(\frac{e^x}{1+e^x}\right)}{dx} = \frac{e^x}{(1+e^x)^2} = l(x) \quad \Rightarrow \quad d(L(x)) = l(x)dx. \tag{13}$$

Determinemos ahora $F_X(x)$, la distribución acumulada de $f_X(x)$:

$$F_X(x) = \int_{-\infty}^x f_X(t)dt$$

$$= \int_{-\infty}^x \alpha l(t)L(t)^{\alpha-1}dt$$

$$= \int_{-\infty}^x \alpha L(t)^{\alpha-1}l(t)dt$$

$$= \int_{-\infty}^x \alpha L(t)^{\alpha-1}d(L(t))$$

$$= L(t)^{\alpha} \begin{bmatrix} x \\ -\infty \end{bmatrix}$$

$$= L(x)^{\alpha} - \lim_{t \to -\infty} L(t)^{\alpha}$$

$$= L(x)^{\alpha} - \lim_{t \to -\infty} \left(\frac{e^t}{1+e^t}\right)^{\alpha}$$

$$= L(x)^{\alpha} - \left(\lim_{t \to -\infty} \frac{e^t}{1+e^t}\right)^{\alpha}$$

$$= L(x)^{\alpha}$$

$$= L(x)^{\alpha}$$

$$= \left(\frac{e^x}{1+e^x}\right)^{\alpha}, \forall x \in \mathbb{R}.$$
(14)

b) Presente los pasos que debe seguir el algoritmo de transformada inversa. Implemente en R el algoritmo propuesto. Simule 10000 valores de una PL(0.5) y verifique mediante gráficos de histograma y cuantiles que el método propuesto funciona correctamente.

Solución. Para usar el método de la transformada inversa, se debe hallar $X = F^{-1}(U)$. De esta manera:

$$u = F_X(x) = \left(\frac{e^x}{1 + e^x}\right)^{\alpha} = (1 + e^{-x})^{-\alpha} \quad \Leftrightarrow \quad u^{-1/\alpha} = 1 + e^{-x}$$

$$\Leftrightarrow \quad e^{-x} = u^{-1/\alpha} - 1$$

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

$$\Rightarrow \quad F_X^{-1}(u) = x = \ln\left(\frac{1}{u^{-1/\alpha} - 1}\right) \tag{15}$$

Listing 11: Verificación mediante histograma / PL.

```
u = runif(10000)
a = 0.5
x = log(1/(u^(-1/a)-1))
fx = function(x) {a*exp(-x)*(1+exp(-x))^(-a-1)}

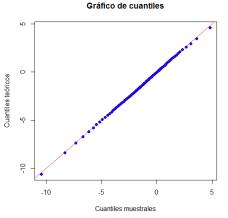
#Grafico de Histograma
par(mfrow=c(1,2))
hist(x,probability = TRUE,xlab="Valores simulados",
    ylab="Densidad",main="Histograma de los valores simulados")
curve(fx,col=2,lwd=2,add=TRUE)
```

Histograma de los valores simulados

Listing 12: Verificación mediante gráfica de cuantiles / PL.

```
p = ((1:100)-0.5)/100

2
3 #Calculando los cuantiles muestrales
4 qm = quantile(x,p)
5
6 #Calculando los cuantiles poblacionales
7 qp = log(1/(p^(-1/a)-1))
8
9 plot(qm,qp,pch=16,lwd=2,col=4,main=paste("Grafico de cuantiles"),xlab="Cuantiles muestrales",ylab="Cuantiles teoricos")
10 abline(0,1,col="red")
```



Para verificar si nuestros datos simulados son correctos, se grafica un histograma con los datos generados por el método de la transformada inversa. Ahora, superponiendo la curva de la distribución logística potencia en el histograma, notamos que los datos se ajustan, por lo que se puede afirmar que la simulación de los datos es apropiada. Asimismo, en el gráfico de cuantiles se aprecia similitud entre los datos simulados y los datos teóricos.

Sea X una variable aleatoria con distribución de Gumbel, entonces su función de densidad de probabilidad (f.d.p.) y su función de distribución acumulada (f.d.a.) es dada por

$$f(x) = e^{-x}e^{-e^{-x}}$$
 y $F(x) = e^{-e^{-x}}$, $x \in \mathbb{R}$.

Una posible generalización es la distribución Gumbel potencia cuya f.d.p. y f.d.a son dadas por

$$f(x) = \alpha e^{-x} \left[e^{-e^{-x}} \right]^{\alpha}$$
 y $F(x) = \left[e^{-e^{-x}} \right]^{\alpha}$, $x \in \mathbb{R}$, $\alpha > 0$.

Otra posibilidad es la distribución Gumbel potencia recíproca cuya f.d.p. y f.d.a son dadas por

$$g(x) = \alpha e^x [e^{-e^x}]^{\alpha}$$
 y $G(x) = 1 - [e^{-e^x}]^{\alpha}$, $x \in \mathbb{R}$, $\alpha > 0$.

a) Verificar si las f.d.p. de las distribuciones: Gumbel potencia y Gumbel potencia reciproca son log-concava.

Solución. Sabemos para que una función h(x) sea log-cóncava, debe cumplirse lo siguiente:

$$\frac{d^2(\log(h(x)))}{dx^2} < 0, \ x \in \mathbb{R}. \tag{16}$$

• Verificación sobre la distribución Gumbel potencia:

$$f(x) = \alpha e^{-x} \left[e^{-e^{-x}} \right]^{\alpha} = \alpha e^{-x - \alpha e^{-x}}$$

$$\Rightarrow \log(f(x)) = \log(\alpha) - x - \alpha e^{-x}$$
(17)

Veamos que la función descrita en (17) cumple ser log-cóncava,

$$\frac{d(\log(f(x)))}{dx} = -1 + \alpha e^{-x}$$

$$\frac{d^2(\log(f(x)))}{dx^2} = -\alpha e^{-x} < 0 \quad \forall x \in \mathbb{R}.$$
(18)

Esto último se debe a que $\alpha > 0$ y $e^{-x} > 0$ para todo $x \in \mathbb{R}$.

• Verificación sobre la distribución Gumbel potencia recíproca:

$$g(x) = \alpha e^x \left[e^{-e^x} \right]^{\alpha} = \alpha e^{x - \alpha e^x}$$

$$\Rightarrow \log(g(x)) = \log(\alpha) + x - \alpha e^x$$
(19)

Veamos que la función descrita en (19) cumple ser log-cóncava,

$$\frac{d(\log(g(x)))}{dx} = 1 - \alpha e^x$$

$$\frac{d^2(\log(g(x)))}{dx^2} = -\alpha e^x < 0 \quad \forall x \in \mathbb{R}.$$
(20)

Esto último se debe a que $\alpha > 0$ y $e^x > 0$ para todo $x \in \mathbb{R}$.

b) Implemente el método ARS en R. Simule 10,000 valores para cada distribución, considere $\alpha = 2$. Realice un gráfico de histograma con la función de densidad y un gráfico de cuantiles.

Solución.

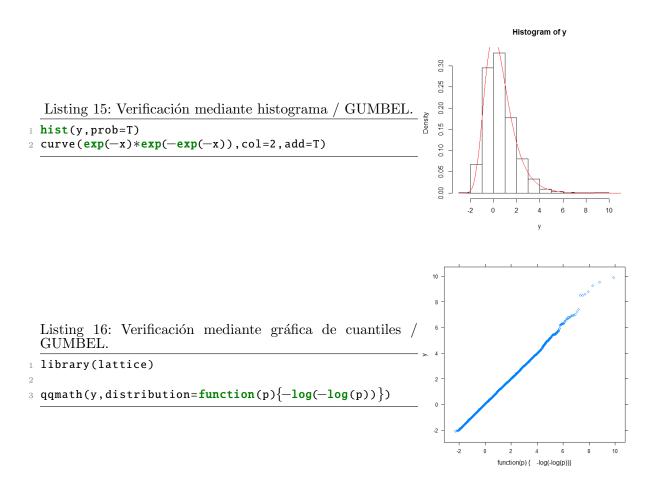
Listing 13: Método ARS / Definiciones Iniciales.

```
1 M<-10000
2 a<-2
3 p<-((1:100)-0.5)/100
```

• Método de ARS en R para la función de distribución de Gumbel:

Listing 14: Algoritmo ARS / GUMBEL.

```
1 library(ars)
2
3 f1<-function(x,mu=0,sigma=1){-x-exp(-x)}
4 f1prima<-function(x,mu=0,sigma=1){-1+exp(-x)}
5 y<-ars(10000,f1,f1prima,mu=0,sigma=1)</pre>
```



Método de ARS en R para la función de distribución de Gumbel potencia:

Listing 17: Algoritmo ARS / GUMBEL POTENCIA.

```
library(ars)

f function(x,a) {log(a)-x-a*exp(-x)}

fprima function(x,a) {-1+a*exp(-x)}

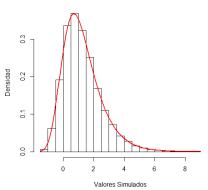
gp ars(M,f,fprima,a=2)
```

Listing 18: Verificación mediante histograma / GUMBEL POTENCIA.

hist(gp,prob=T,xlab="Valores Simulados",ylab="
 Densidad",main="Histograma de los valores
 simulados")

fx<-function(x) {a*exp(-x)*(exp(-exp(-x)))^a}

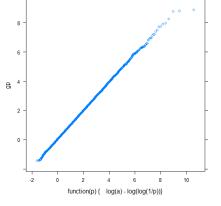
curve(fx,col=2,lwd=2,add=T)</pre>



Histograma de los valores simulados

Listing 19: Verificación mediante gráfica de cuantiles / GUMBEL POTENCIA.

qqmath(gp,distribution=function(p) $\{log(a)-log(log(1/p))\}$)



Método de ARS en R para la función de distribución de Gumbel potencia recíproca:

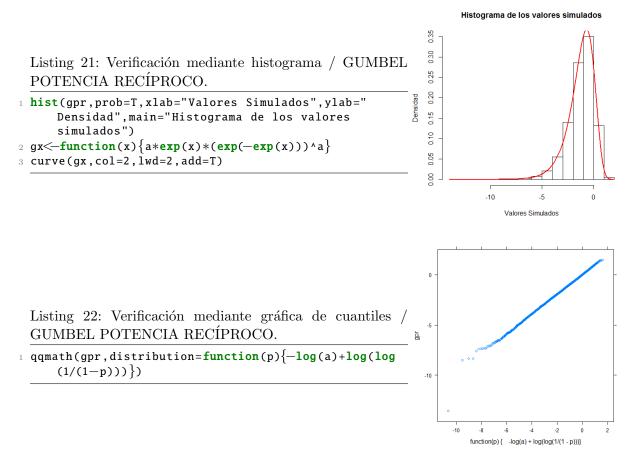
Listing 20: Algoritmo ARS / GUMBEL POTENCIA RECÍPROCO.

```
library(ars)

g<-function(x,a){log(a)+x-a*exp(x)}

gprima<-function(x,a){1-a*exp(x)}

gpr<-ars(M,g,gprima,a=2)</pre>
```



Se ha podido demostrar que las 3 funciones, f.d.p. de Gumbel, f.d.p. de Gumbel potencia y f.d.p. de Gumbel potencia recíproca, se ha podido generar las muestras de variables aleatorias, dado que cada una ha cumplido dos condiciones:

i) Las 3 funciones son continuas y diferenciables en todo su dominio R_X . Es diferenciable porque las funciones admiten derivadas en cualquier dirección y puede aproximarse al menos hasta primer orden. Es continua pues para cualquier punto x=a, existe f(a), asimismo también existe el límite de la función en el punto x=a y finalmente f(a)coincide con el límite de la función en el punto.

$$\exists f(a)$$
 (21)

$$\exists \lim_{x \to a} f(x) \Leftrightarrow \lim_{x \to a^{-}} f(x) = \lim_{x \to a^{+}} f(x)$$

$$f(a) = \lim_{x \to a} f(x)$$
(22)

$$f(a) = \lim_{x \to a} f(x) \tag{23}$$

ii) La función log-cóncava de cada una cumple que su segunda derivada es menor que CERO, para todo $x \in R_X$.

Gracias a que log f(x), log f(x) potencia y log g(x) son funciones cóncavas, estas han podido ser acotadas superiormente por sus tangentes, lo cual ha permitido que se haya podido acercar cada vez más a las funciones.

Gráfico de las 3 funciones:

