Problemas probabilidad y variables aleatorias. Laboratoriao software y problemas 2

Contents

1 Problemas de probabilidad y variables aleatorias lab2

1.1 Problema

Un estuche contiene 2 lápices azules y 3 rojos. Se extraen dos lápices del estuche.

a. Escribe los resultados elementales que definen los sucesos. $M =$ "Solo ha salido un lápiz rojo" y $N =$ "El segundo lápiz extraído es azul". (**0.5 puntos**)
b. Halla las probabilidades de M, N y $M \cap N$. (**1 punto.**)
c. ¿Son los sucesos M y N independientes? (**1 punto.**)

1.1.1 Solución

El espacio muestral que resuelve el problema puede tener varias formas.

Forma 1 Tenemos 5 lápices en total de los que 2 son de color azul y 3 son rojos. Extraemos dos lápices (sí sin reponerlos hemos sacado dos)

Hemos elegido (hay otros modelos para definir el experimento que dan el mismo resultado) definir pares ordenados para el color del primer lápiz extraído y para el del segundo: por ejemplo $(Azul, Rojo)$ será que el primer lápiz es azul y el segundo (notemos que es sin reponer el lápiz en el estuche).

Así es espacio muestral es $\Omega = \{(Azul, Azul), (Azul, Rojo), (Rojo, Azul), (Rojo, Rojo)\}$.

Veamos las probabilidades de cada suceso elemental.

$P(\{(Azul, Azul)\}) = P(Primero \quad Azul) \cdot P(Segundo \quad Azul/PrimeroAzul) = \frac{2}{5} \cdot \frac{1}{4} = \frac{2}{20}$.

$P(\{(Azul, Rojo)\}) = P(Primero \quad Azul) \cdot P(Segundo \quad Rojo/Primero \quad Azul) = \frac{2}{5} \cdot \frac{3}{4} = \frac{6}{20}$.

$P(\{(Rojo, Azul)\}) = P(Primero \quad Rojo) \cdot P(Segundo \quad Azul/Primero \quad Rojo) = \frac{3}{5} \cdot \frac{2}{4} = \frac{6}{20}$.

$P(\{(Rojo, Rojo)\}) = P(Primero \quad Rojo) \cdot P(Segundo \quad Rojo/Primero \quad Rojo) = \frac{3}{5} \cdot \frac{2}{4} = \frac{6}{20}$.

El suceso "solo un lápiz rojo es" $M = \{(Azul, Rojo), (Rojo, Azul)\}$.

El suceso "el segundo lápiz es azul es" $N = \{(Rojo, Azul), (Azul, Azul)\}$

Por lo tanto

$$P(M) = P(\{(Azul, Rojo), (Rojo, Azul)\}) = P((Azul, Rojo)) + P((Rojo, Azul)) = \frac{12}{20} = \frac{3}{5} = 0.6.$$

$$P(N) = P(\{(Rojo, Azul), (Azul, Azul)\}) = P((Rojo, Azul)) + P((Azul, Azul)) = \frac{8}{20} = \frac{2}{5} = 0.4.$$

El suceso $M \cap N = \{(Azul, Rojo), (Rojo, Azul)\} \cup \{(Rojo, Azul), (Azul, Azul)\} = \{(Rojo, Azul)\}$

por lo tanto

$$P(M \cap N) = P(\{(Rojo), Azul)\}) = \frac{6}{20} = \frac{3}{10} = 0.3.$$

Para que My N sean independientes se tiene que cumplir que $P(M \cap N) = P(M) \cdot P(N)$, pero esto no sucede

$$\frac{6}{20} = 0.3 = P(M \cap N) \neq P(M) \cdot P(N) = \frac{12}{20} \cdot \frac{8}{20} = 0.12.$$

Al no cumplirse la anterior igualdad resulta los sucesos M y N no son independientes.

1.2 Problema

En promedio, 3 servidores de cada 20 se bloquea durante una tormenta eléctrica. La compañía Amazonas tiene numerosos servidores repartidos en varios Data Centers. Responder, modelando con una distribución notable las siguientes cuestiones:

a. ¿Calcula la probabilidad de que menos de 5 servidores se bloqueen en un Data Center con 20 servidores? (**0.5 puntos.**)
b. ¿Calcula la probabilidad de que exactamente 5 servidores se hayan bloqueado en un Data Center de 20 servidores? (**1 punto.**)
c. En un Data Center de 60 servidores ¿Cuál es la probabilidad de más de 10 (> 10) servidores se bloqueen? Hacerlo utilizando aproximación por una distribución de Poisson. (**1 punto.**)

1.2.1 Solución

La probabilidad de fallo en un servidor es $P(fallo) = \frac{3}{20} = 0.15$.

Sea X la va.a que nos da el número de servidores que han fallado entre 20. Bajo estas condiciones X sigue una ley $B(n = 20, p = \frac{3}{20} = 0.15)$.

En la pregunta a) nos piden $P(X < 5)$ así que $P(X < 5) = P(X \leq 4) = 0.8298$.

Donde la última igualdad se ha deducido utilizando las tablas de la distribución acumulada de una variable aleatoria $B(n = 20, p = 0.25)$. Con R es `round(pbinom(4,20,0.15))` que da el mismo resultado .

Para el apartado b)

$$\begin{aligned} P(X = 5) &= \binom{20}{5} \cdot 0.15^5 \cdot (1 - 0.15)^{20-5} = \frac{20!}{(20-5)! \cdot 5!} \cdot 0.15^5 \cdot 0.85^{15} \\ &= \frac{16 \cdot 17 \cdot 18 \cdot 19 \cdot 20}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} \cdot 0.15^5 \cdot 0.85^{15} = 0.1028452. \end{aligned}$$

1.3 Problema

El profesor de estadística repite la palabra *muestra* a un ritmo de 20 veces por cada 60 minutos. Sea X_t la variable que cuenta el número de veces que el profesor ha dicho *muestra* en t minutos.

a. Modelizad X_t mediante una distribución Poisson. De X_t dad su parámetro, su valor esperado y varianza. (**0.5 puntos.**)
b. ¿Cuál es la probabilidad de que diga muestra más de 10 veces en media hora? (**1 punto.**)
c. Sea $T =$ el tiempo transcurrido entre la última vez que el profesor dice *muestra* hasta la siguiente vez ¿Cuál es la probabilidad de que $T > 15$? (**1 punto**)

1.3.1 Solución

Sea la λ el promedio de veces que dice "muestra" por minuto como lo dice 20 veces en 60 minutos tenemos que el número promedio de veces es $\lambda = \frac{1}{3}$. por minuto.

Sea $X_t =$ número de veces que dice muestra en t minutos modelizado como un proceso Poisson será que $X_t \equiv Po(\lambda \cdot t = \frac{t}{3})$. Al ser una Poisson tenemos que $E(X_t) = Var(X_t) = \lambda \cdot t = \frac{t}{3}$.

$P($"diga muestra más de 10 veces en media hora"$) = P(X_{30} > 10) = 1 - P(X_{30} \leq 10)$.

Ahora como $X_{30} \equiv Po(\frac{30}{3} = 10)$ consultando la distribución acumulada de una $Po(10)$ tenemos que $P(X_{30} \leq 10) = 0.583$, por lo tanto

$P($"diga muestra más de 10 veces en media hora"$) = 1 - P(X_{30}) = 1 - 0.583 = 0.417$.

Sea $T =$ "tiempo en minutos que tarda en decir muestra desde la última vez que lo dijo''. Como X_t es una $Po(\frac{t}{3})$ sabemos que T sigue una ley $Exp(\frac{1}{3})$ y por lo tanto

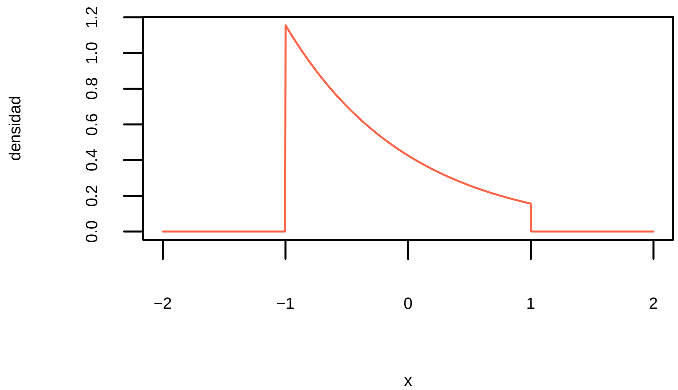
$$P(T > 15) = 1 - P(T \leq 15) = 1 - (1 - e^{-\frac{1}{3} \cdot 15}) = e^{-\frac{1}{3} \cdot 15} = 0.0067.$$

1.4 Problema

Consideremos la va. X con densidad, donde α es un parámetro real.

$$f(x) = \begin{cases} \alpha \cdot e^{1-x} & \text{si } -1 < x < 1 \\ 0 & \text{en otro caso} \end{cases}$$

1. Calculad α para que f sea densidad. (**1 punto.**)
2. Calculad la función de distribución de X. (**1 punto.**)
3. Calculad $P(|X| > \frac{1}{2})$. (**0.5 puntos.**)
4. Calculad $E(X)$. (**Punto extra de esta entrega**. Indicación: hay que integrar por partes)



1.4.1 Solución

Calculemos el área bajo la densidad que debe ser 1.

$$\int_{-\infty}^{\infty} f(x)dx = \int_{-1}^{1} \alpha \cdot e^{1-x} dx = \left[-\alpha \cdot e^{1-x}\right]_{-1}^{1} = -\alpha \cdot e^{0} + \alpha \cdot e^{2} = \alpha \cdot (e^2 - 1).$$

Luego el valor de α para que sea densidad es la solución de $1 = \alpha \cdot (e^2 - 1)$ por tanto $\alpha = \frac{1}{e^2-1} \approx 0.1565$.

Por comodidad de la escritura seguiremos llamando α a esta constante.

Nos piden la función de distribución. Dado x tal que $-1 < x < 1$ tenemos que la función de distribución de X es

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt = \int_{-1}^{x} \alpha \cdot e^{1-t} dt = \left[-\alpha \cdot e^{1-t}\right]_{-1}^{x} = -\alpha \cdot e^{1-x} + \alpha \cdot e^{2} = \alpha \cdot (e^2 - e^{1-x}).$$

Luego la función de distribución es

$$F(x) = P(X \leq x) = \begin{cases} 0 & \text{si } x \leq -1 \\ \alpha \cdot (e^2 - e^{1-x}) & \text{si } -1 < x < 1 \\ 1 & \text{si } 1 < x \end{cases}$$

Calculemos $P\left(|X| > \frac{1}{2}\right)$

$$\begin{aligned} P\left(|X| > \frac{1}{2}\right) &= 1 - P\left(|X| \leq \frac{1}{2}\right) = 1 - P\left(-\frac{1}{2} \leq X \leq \frac{1}{2}\right) = 1 - \left(F\left(\frac{1}{2}\right) - F\left(-\frac{1}{2}\right)\right) \\ &= 1 - \left(\alpha \cdot \left(e^2 - e^{1-\frac{1}{2}}\right) - \alpha \cdot \left(e^2 - e^{1+\frac{1}{2}}\right)\right) = 1 - \alpha \cdot \left(e^{\frac{1}{2}} - e^{\frac{3}{2}}\right) \\ &\approx 1 + 0.1565 \cdot \left(e^{\frac{1}{2}} - e^{\frac{3}{2}}\right) = 0.5566. \end{aligned}$$

Solo nos queda el cálculo del valor esperado (punto extra)

$$
\begin{aligned}
E(X) &= \int_{-\infty}^{+\infty} x\cdot f(x)dx = \int_{-1}^{1} 0.1565\cdot x\cdot e^{1-x}dx = \left| \begin{array}{cc} u=x & du=dx \\ dv=e^{1-x}dx & v=-e^{1-x} \end{array} \right| \\
&= 0.1565\cdot\left[\left[x\left(-e^{1-x}\right)\right]_{-1}^{1} - \int_{-1}^{1}\left(-e^{1-x}\right)dx\right] \\
&= 0.1565\cdot\left[1\cdot\left(-e^{1-1}\right)-(-1)\cdot\left(e^{1-(-1)}\right)-\left[e^{1-x}\right]_{-1}^{1}\right] \\
&= 0.1565\cdot\left[-1-e^{2}-\left(e^{1-1}-e^{1-(-1)}\right)\right] \\
&= 0.1565\cdot\left[-1-e^{2}-1+e^{2}\right] \\
&= 0.1565\cdot(-2) = -0.313.
\end{aligned}
$$

1.5 Problema

(**1 punto.**) Lanzamos un dado de 12 caras numeradas con enteros del 1 al 12 sobre una mesa plana. Observamos el número superior del dado. Calcular la probabilidad de que salga mayor que 8 si el resultado es par.

1.6 Problema

Lanzamos una moneda con probabilidad de cara $p=\frac{1}{2}$ hasta que sale cara dos veces o bien la hemos lanzamos 5 veces, lo primero que ocurra.

Denotemos por X la variable aleatoria que determina el número de tiradas de la moneda.

Se pide:

1. Describir adecuadamente el espacio muestral de la variable X (**0.5 punto.**)
2. Calcular su función de densidad.(**1 punto.**)
3. Calcular $E(X)$.(**1 punto.**)

1.7 Problema

Sea X una variable con distribución uniforme en el intervalo $(1,10)$ con $a>1$. Consideremos la variable $Y=\log_{10}(X)$. Se pide

1. Calcular la función de distribución de Y (**1 punto.**)
2. Calcular la función de densidad de X. (**0.5 puntos.**)
3. Calcular el cuantil 0.95 de X. (**0.5 puntos.**)

1.8 Problema

Consideremos los siguientes sucesos A y B tales que $P(A\cup B)=0.8$, $P(A-B)=0.4$ y $P(B-A)=0.3$. Calcular $P(A\cap B)$, si es posible.(**0.5 puntos.**).

1.9 Problema Ley de Bendford

La ley de Benford es una curiosa distribución de probabilidad que suele aparecer en la distribución del primer dígito de las cantidades registradas en contabilidades y en observaciones científicas o datos numéricos. La variable X sigue una distribución discreta Benford con dominio $D_X = \{1, 2, 3, 4, 5, 7, 8, 9\}$ son 9 dígitos (se elimina el cero) y su función de probabilidad es (log es el logaritmo en base 10)

$$P_X(x) = P(X = x) = \log(x + 1) - \log(x).$$

- a) Calcular la media y la varianza de X.
- b) Calcular la función de distribución de X.
- c) ¿Cuál es el dígito más frecuente (moda)?
- d) Construid con R las funciones de probabilidad y de distribución de X.
- e) Dibujar con R las funciones del apartado anterior.

1.9.1 Solución

1) Recordad que en R `log10` es el logartimo en base 10

Como $\log(d + 1) - \log(d) = \log(\frac{d+1}{d})$. Podemos implementar la función de probabilidad de Bendford con el siguiente código R

```
dBendford = function(x){
  sapply(x, FUN=function(x1)
  {
    if (x1 %in% 1:9)
    {log10(x1+1)-log10(x1)}
    else{0}
  })
}

dBendford(1:9)
```

```
## [1] 0.30103000 0.17609126 0.12493874 0.09691001 0.07918125 0.06694679 0.05799195
## [8] 0.05115252 0.04575749
```

```
sum(dBendford(1:9))
```

```
## [1] 1
```

Así la media μ será

```
mu=sum(c(1:9)*dBendford(1:9))
mu
```

```
## [1] 3.440237
```

```
sumx2=sum(c(1:9)^2*dBendford(1:9))
sumx2
```

```
## [1] 17.89174
```

```
sigma2=sumx2-mu^2
sigma2
```

```
## [1] 6.056513
```

```
sigma=sqrt(sigma2)
sigma
```

```
## [1] 2.460998
```

En resumen La variable de Bendford tiene dominio $D_X = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$ y función de probabilidad

$$P_X(x) = P(X = x) = \begin{cases} 0.30103 & \text{si } x = 1 \\ 0.1760913 & \text{si } x = 2 \\ 0.1249387 & \text{si } x = 3 \\ 0.09691 & \text{si } x = 4 \\ 0.0791812 & \text{si } x = 5 \\ 0.0669468 & \text{si } x = 6 \\ 0.0579919 & \text{si } x = 7 \\ 0.0511525 & \text{si } x = 8 \\ 0.0457575 & \text{si } x = 9 \\ 0 & \text{ en otro caso} \end{cases}$$

$$\begin{aligned} E(X) &= \sum_{k=1}^{9} x \cdot P_X(x) = 1 \cdot 0.30103 + 2 \cdot 0.1760913 + 3 \cdot 0.1249387 + 4 \cdot 0.09691 + 5 \cdot 0.0791812 \\ &+ 6 \cdot 0.0669468 + 7 \cdot 0.0579919 + 8 \cdot 0.0511525 + 9 \cdot 0.0457575 \\ &= 3.440237 \end{aligned}$$

$$\begin{aligned} E(X^2) &= \sum_{k=1}^{9} x^2 \cdot P_X(x) = 1 \cdot 0.30103 + 4 \cdot 0.1760913 + 9 \cdot 0.1249387 + 16 \cdot 0.09691 + 25 \cdot 0.0791812 \\ &+ 36 \cdot 0.0669468 + 49 \cdot 0.0579919 + 64 \cdot 0.0511525 + 81 \cdot 0.0457575 \\ &= 3.440237 \end{aligned}$$

Y por último $Var(X) = E(X^2) - (E(X))^2 = 17.891743 - (3.440237)^2 = 6.0565126$ y la desviación típica es $\sqrt{Var(X)} = 2.4609983$.

2) Ahora nos piden $F_X(x) = P(X \leq x)$. Con R es

```r
pBendford=function(x){
  sapply(x,FUN=function(x){
  probs=cumsum(dBendford(1:9))
  xfloor=floor(x)
  if(xfloor<1){0} else {if(xfloor>8) {1} else {probs[xfloor]}}
})
}

pBendford(0:9)
```

```
##  [1] 0.0000000 0.3010300 0.4771213 0.6020600 0.6989700 0.7781513 0.8450980
##  [8] 0.9030900 0.9542425 1.0000000
```

```r
pBendford(0)
```

```
## [1] 0
```

```r
pBendford(10)
```

```
## [1] 1
```

Así tenemos que

$$F_X(x) = P(X \leq x) = \begin{cases} 0 & \text{si } x < 1 \\ 0.30103 & \text{si } 1 \leq x < 2 \\ 0.4771213 & \text{si } 2 \leq x < 3 \\ 0.60206 & \text{si } 3 \leq x < 4 \\ 0.69897 & \text{si } 4 \leq x < 5 \\ 0.7781513 & \text{si } 5 \leq x < 6 \\ 0.845098 & \text{si } 6 \leq x < 7 \\ 0.90309 & \text{si } 7 \leq x < 8 \\ 0.9542425 & \text{si } 8 \leq x < 9 \\ 1 & \text{si } 9 \leq x \end{cases}$$

c) EL dígito más frecuente es el 1

```r
dBendford(1:9)
```

```
## [1] 0.30103000 0.17609126 0.12493874 0.09691001 0.07918125 0.06694679 0.05799195
## [8] 0.05115252 0.04575749
```

```r
max(dBendford(1:9))
```

```
## [1] 0.30103
```

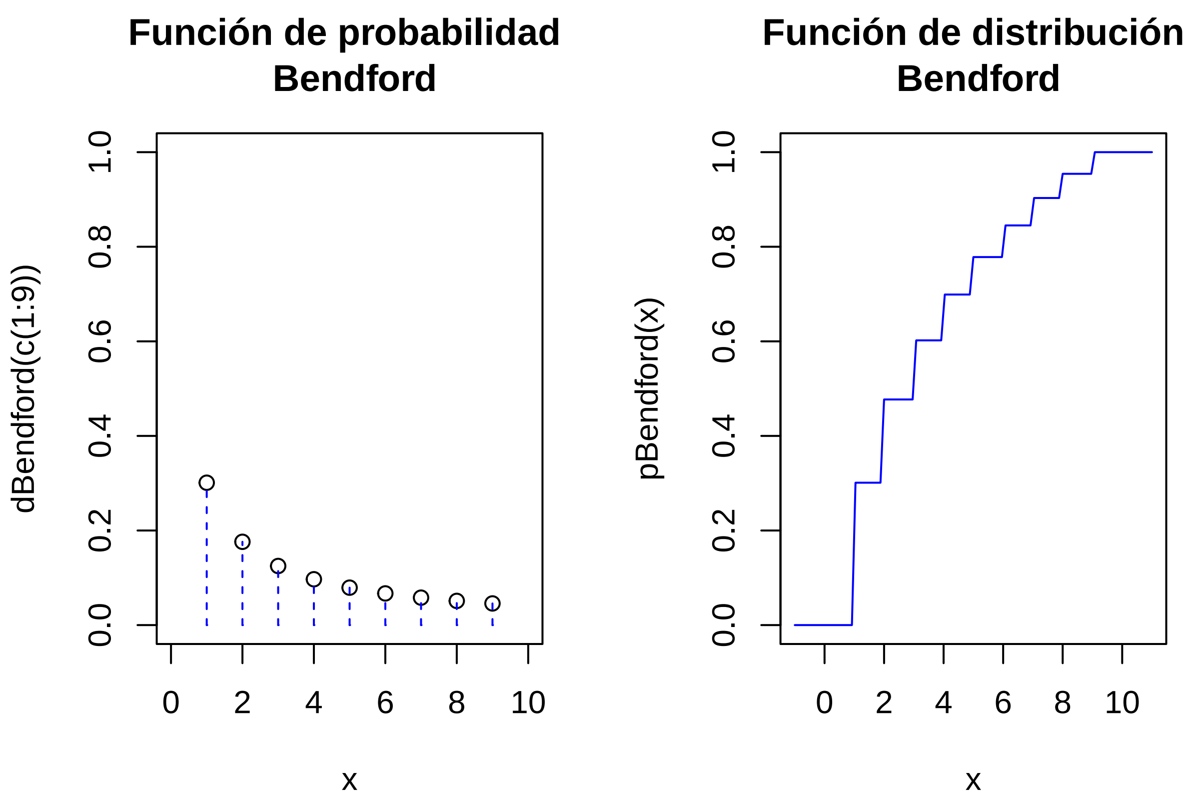
```r
which.max(dBendford(1:9))
```

```
## [1] 1
```

d) Ya lo hemos hecho. . . .

e) Dibujemos

```
par(mfrow=c(1,2))
aux=rep(0,18)
aux[seq(2,18,2)]=dBendford(c(1:9))
x=c(1:9)
plot(x,y=dBendford(c(1:9)),
  ylim=c(0,1),xlim=c(0,10),xlab="x",
  main="Función de probabilidad \n Bendford")
lines(x=rep(1:9,each=2),y=aux, type = "h", lty = 2,col="blue")
curve(pBendford(x), xlim=c(-1,11),col="blue", main="Función de distribución\n Bendford")
```




```
par(mfrow=c(1,1))
```

1.10 Problema Distribución de Pareto (Power law)

- a) Calcular en función de k y γ la densidad de la variable X.
- b) Para $\gamma > 1$ calcular $E(X)$ y $Var(X)$ y su desviación típica.
- c) ¿Qué sucede con $E(X)$ si $0 < \gamma < 1$.
- d) ¿Cómo se calcula está distribución con R y con python?
- e) Dibujar las gráficas de su densidad y distribución para $\gamma = 3$ y $\gamma = 5$.
- f) Explorar por internet (wikipedia) cómo es la distribución **power law** y qué relación tiene el concepto de *scale free* con los resultados del apartado c).

1.10.1 Solución

a) La densidad será la derivada de la distribución F_X respecto de x , si $x \geq x_m > 0$

$$
\begin{aligned}
f_X(x) = (F_X(x))' &= \left(1 - k \cdot \left(\frac{x_m}{x}\right)^{\gamma}\right)' = \left(1 - k \cdot x_m^{\gamma} \cdot x^{-\gamma}\right)' = \left(1 - k \cdot x_m^{\gamma} \cdot x^{-\gamma}\right)' \\
&= -\gamma \cdot (-k \cdot x_m^{\gamma}) \cdot x^{-\gamma-1} = \gamma \cdot k \cdot x_m^{\gamma} \cdot x^{-\gamma-1}
\end{aligned}
$$

Si tenemos $x < x_m$ entonces $f_X(x) = 0$ en resumen

$$
f_X(x) = \begin{cases} \gamma \cdot x_m^{\gamma} \cdot x^{-\gamma-1} & \text{si } x \geq x_m \\ 0 & \text{si } x < x_m \end{cases}
$$

Notemos que γ es un parámetro pero k es na constante a determinar pues la densidad debe integrar 1 en el dominio $D_X = [x_m, +\infty)$

$$
\begin{aligned}
\int_{x_m}^{+\infty} f_X(x)dx &= \int_{x_m}^{+\infty} \gamma \cdot k \cdot x_m^{\gamma} \cdot x^{-\gamma-1} \cdot dx = \left[-k \cdot x_m^{\gamma} \cdot x^{-\gamma}\right]_{x=x_m}^{+\infty} = \lim_{x \to \infty} \left[-k \cdot x_m^{\gamma} \cdot x^{-\gamma}\right] - \left(-k \cdot x_m^{\gamma} \cdot x_m^{-\gamma}\right) \\
&= 0 + k \cdot x_m^{\gamma} \cdot {x_m}^{-\gamma} = k.
\end{aligned}
$$

Notemos que el límite es 0 pues $\gamma > 0$ y $x_m > 0$. Luego $k = 1$ y la función de densidad y la de distribución se puede escribir como damos dos versiones

$$
f_X(x) = \begin{cases} \gamma \cdot x_m^{\gamma} \cdot x^{-(\gamma+1)} & \text{si } x \geq x_m \\ 0 & \text{si } x < x_m \end{cases} = \begin{cases} \frac{\gamma \cdot x_m^{\gamma}}{x^{(\gamma+1)}} & \text{si } x \geq x_m \\ 0 & \text{si } x < x_m \end{cases}
$$

$$
f_X(x) = \begin{cases} \gamma \cdot x_m^{\gamma} \cdot x^{-(\gamma+1)} & \text{si } x \geq x_m \\ 0 & \text{si } x < x_m \end{cases} = \begin{cases} \frac{\gamma \cdot x_m^{\gamma}}{x^{(\gamma+1)}} & \text{si } x \geq x_m \\ 0 & \text{si } x < x_m \end{cases}
$$

y la distribución

$$
F_X(X) = \begin{cases} 1 - x_m^{\gamma} \cdot x^{-\gamma} & \text{si } x > x_m \\ 0 & \text{si } x \leq x_m \end{cases} = \begin{cases} 1 - \left(\frac{x_m}{x}\right)^{\gamma} & \text{si } x > x_m \\ 0 & \text{si } x \leq x_m \end{cases}
$$

c)

Calculemos su esperanza

$$
\begin{aligned}
E(X) &= \int_{x_m}^{+\infty} x \cdot f_X(x) \cdot dx = \int_{x_m}^{+\infty} x \cdot \gamma \cdot x_m^{\gamma} \cdot x^{-\gamma-1} \cdot dx = \int_{x_m}^{+\infty} \gamma \cdot x_m^{\gamma} \cdot x^{-\gamma} \cdot dx = \left[\frac{\gamma}{-\gamma+1} \cdot x_m^{\gamma} \cdot x^{-\gamma+1}\right]_{x=x_m}^{+\infty} \\
&= \lim_{x \to \infty} \left[\frac{\gamma}{-\gamma+1} \cdot x_m^{\gamma} \cdot x^{-\gamma+1}\right] - \left(\frac{\gamma}{-\gamma+1} \cdot x_m^{\gamma} \cdot x_m^{-\gamma+1}\right) = \lim_{x \to \infty} \left[\frac{\gamma}{-\gamma+1} \cdot x_m^{\gamma} \cdot x^{-\gamma+1}\right] + \frac{\gamma \cdot x_m}{\gamma - 1}
\end{aligned}
$$

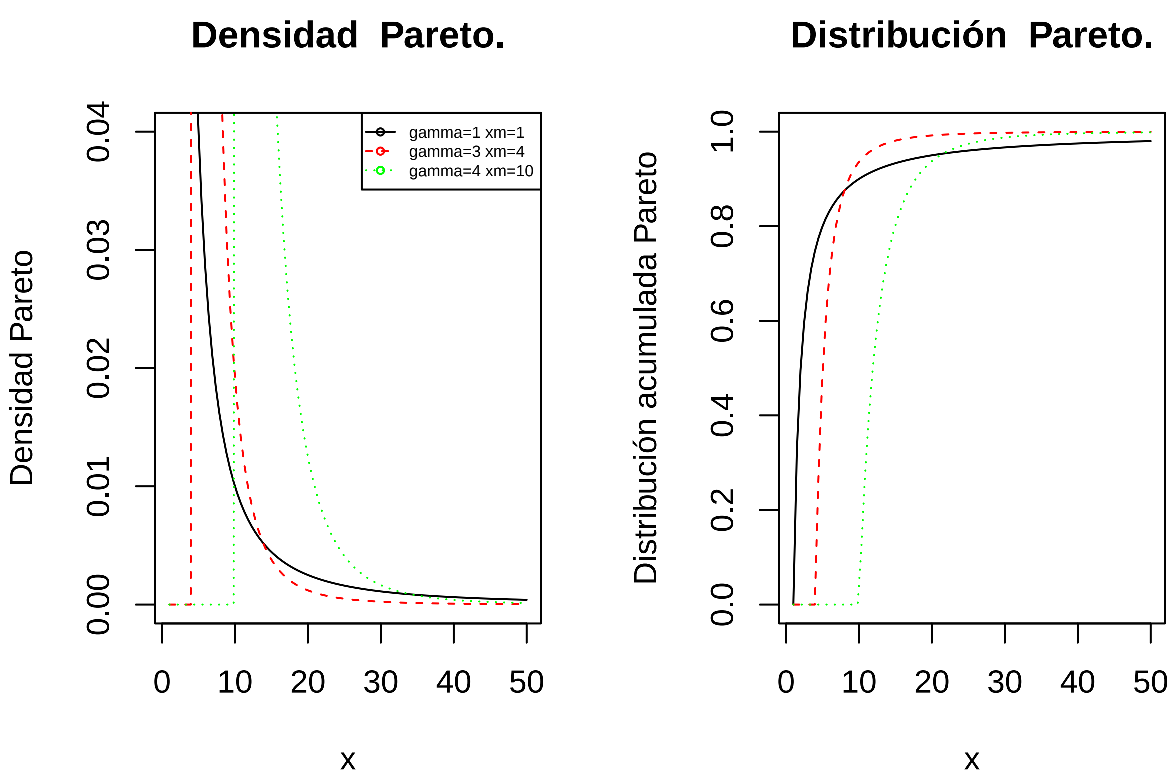
Ahora tenemos dos casos para el límite que $0 < \gamma \leq 1$ o que $\gamma > 1$, es decir que $-\gamma + 1$ sea negativo o positivo, entonces

$$
\lim_{x \to \infty} \left[\frac{\gamma}{-\gamma+1} \cdot x_m^{\gamma} \cdot x^{-\gamma+1}\right] = \begin{cases} +\infty & \text{diverge si } 0 < \gamma \leq 1 \\ \frac{\gamma \cdot x_m}{\gamma - 1} & \text{converge si } \gamma > 1 \end{cases}
$$

Así que **no siempre existe** $E(X)$, si en una distribución Pareto $\gamma \leq 1$ su media diverge se dice entonces que es una distribución **de escala libre**, en inglés **scale free** en el sentido de que carece de media.

e) Podemos programar pero ya lo han hecho en el paquete *Environmental Statistics* (`EnvStats`) y el *Extra Distributions* (`extraDistr`) utilizaremos el segundo paquete en el que las funciones están implementadas en C++) instalarlo si no lo tenéis.

```
par(mfrow=c(1,2))# el parámetro gamma es a y el parámetro m es b
curve(extraDistr::dpareto(x,a=1,b=1),xlim=c(1,50),
      ylim=c(0,0.04),lty=1,main="Densidad  Pareto."
      ,ylab="Densidad Pareto")
curve(extraDistr::dpareto(x,a=3,b=4),
      add=TRUE,col="red",lty=2)
curve(extraDistr::dpareto(x,a=4,b=10),
      add=TRUE,col="green",lty=3)
legend("topright",pch=21,
       legend=c("gamma=1 xm=1","gamma=3 xm=4","gamma=4 xm=10"),
       col=c("black","red","green"),lty=c(1,2,3),cex=0.5)
curve(extraDistr::ppareto(x,a=1,b=1),
      xlim=c(1,50),ylim=c(0,1),lty=1,main="Distribución  Pareto.",
      ylab="Distribución acumulada Pareto")
curve(extraDistr::ppareto(x,a=3,b=4),
      add=TRUE,col="red",lty=2)
curve(extraDistr::ppareto(x,a=4,b=10),
      add=TRUE,col="green",lty=3)
```




```
par(mfrow=c(1,1))
```

f) Buscad los enlaces del la wikipedia. Tenéis que buscar la *Power law* y la *Zipf's law*. Ambas distribuciones son famosas aparecen en la distribución de contactos en una ley social, en la longitud de un mensaje en un foro y en otros aspectos empíricos muy interesantes. Si hay ocasión y el curso es un éxito ampliaremos estas distribuciones.

1.11 Problema Distribución de Pareto (Power law)

La distribución de Gumbel aparece en variables que miden lo que se llama un valor extremo: precipitación máxima de lluvia, tiempo máximo transcurrido entre dos terremotos, o en métodos de *machine learning* el máximo de las puntuaciones de una algoritmo; por ejemplo comparar pares de objetos (fotos, proteínas, etc.).

Una variable aleatoria sigue una ley de distribución Gumbel (de TIPO I) si su distribución es:

$$F_X(x) = \begin{cases} e^{-e^{-\frac{x-\mu}{\beta}}} & \text{si} x \geq 0 \\ 0 & \text{si} x < 0 \end{cases}$$

Para μ y $\beta > 0$ parámetros reales. Llamaremos distribución Gumbel estándar a la que tiene por parámetros $\mu = 0$ y $\beta = 1$.

- a) Si X es una Gumbel estándar calcular su función de densidad y dibujar su gráfica.
- b) Consideremos la función $F(x) = e^{-e^{-x}}$ para $x \geq 0$ y que vale cero en el resto de casos. Comprobar que es la función de distribución $P(X \leq x)$ de una v.a. Gumbel estándar.
- c) Buscad un paquete de R que implemente la distribución Gumbel. Aseguraros de que es la (Gumbel Tipo I). Dejando fijo el parámetro $\beta = 1$ dibujar la densidad Gumbel para varios valores de μ y explicad en que afecta a la gráfica el cambio de μ.
- d) Dejando fijo el parámetro μ dibujad la densidad Gumbel para varios valores de $\beta > 0$ y explicar en qué afecta a la gráfica el cambio de este parámetro.
- e) Buscad cuales son las fórmulas de la esperanza y varianza de una distribución Gumbel en función de α y β.
- f) Repetid los apartados c) y d) con python. Con python se puede pedir con la correspondiente función la esperanza y varianza de esta distribución, comprobar con esta función para algunos valores las fórmulas de la esperanza y la varianza del apartado e).

1.11.1 Solución

a) La Gumbel estándar tiene por distribución

$$F_X(x) = \begin{cases} e^{-e^{-x}} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$$

Entonces si $x > 0$

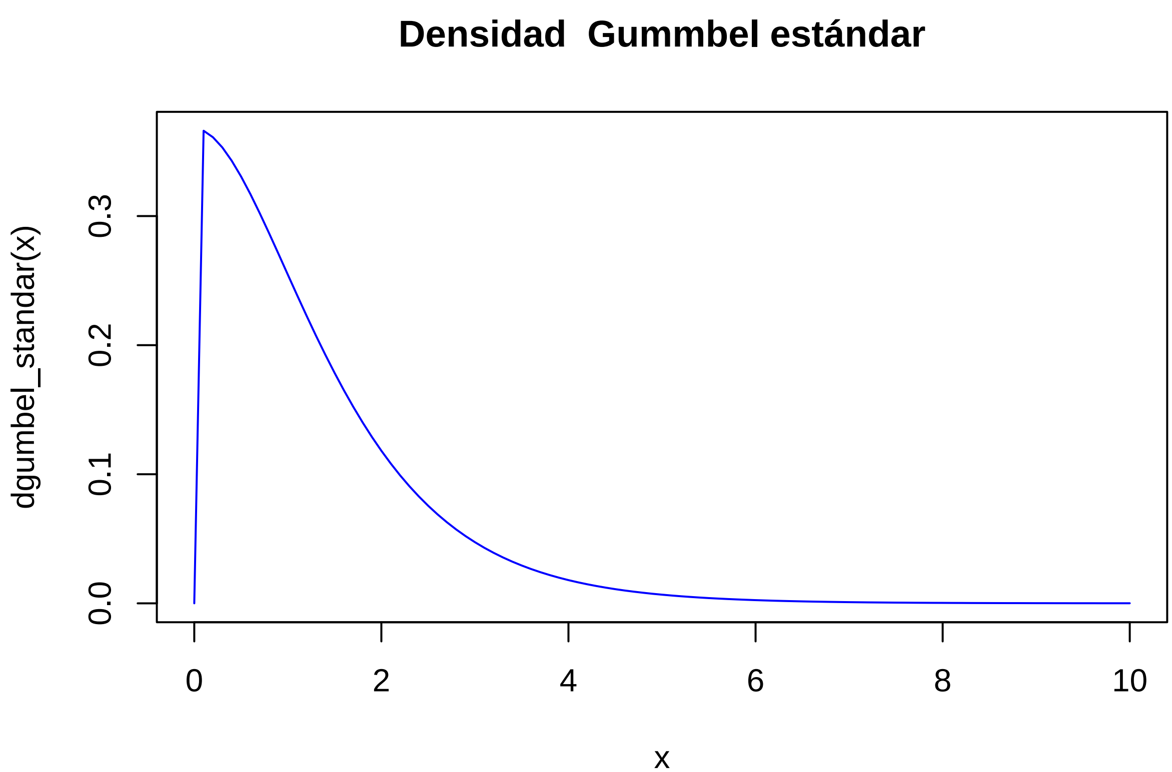
$$f_X(X) = (F_X(x))' = \left(e^{-e^{-x}}\right)' = e^{-e^{-x}} \cdot e^{-x}$$

Luego

$$f_X(x) = \begin{cases} e^{-e^{-x}} \cdot e^{-x} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$$

```r
dgumbel_standar=function(x) {
  sapply(x,
         FUN=function(x) {
           if(x>0) {return(exp(-exp(-x))*exp(-x))} else {return(0)}
         }
  )
  }

curve(dgumbel_standar(x),col="blue",main="Densidad  Gummbel estándar",xlim=c(0,10))
```



b) Consideremos la función $F(x) = e^{-e^{-x}}$ para $x \geq 0$ y que vale cero en el resto de casos. Comprobar que es la función de distribución $P(X \leq x)$ de una v.a. Gumbel estándar.

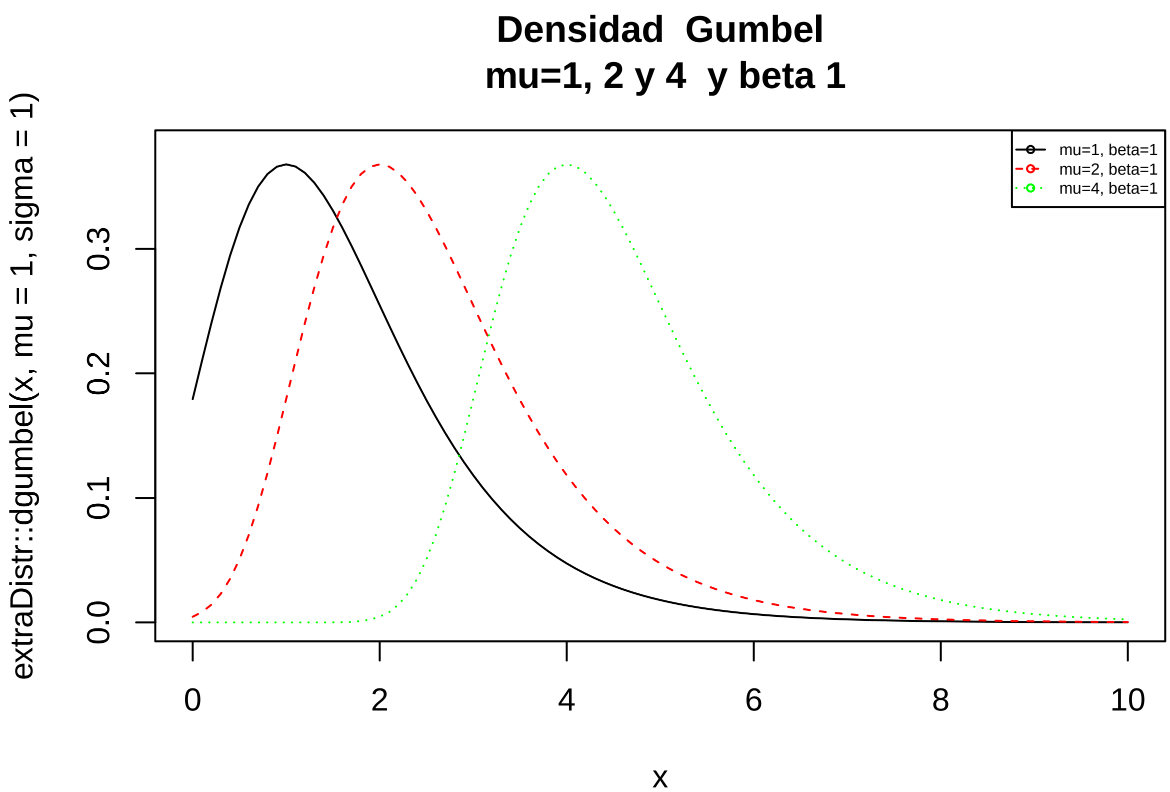
Efectivamente Es suficiente sustituir en la fórmula original.

c) Buscad un paquete de R que implemente la distribución Gumbel. Aseguraros de que es la (Gumbel Tipo I). Dejando fijo el parámetro $\beta = 1$ dibujar la densidad Gumbel para varios valores de μ y explicad en que afecta a la gráfica el cambio de μ.

Un paquete que implementa la Gumbel es `extraDistr` el parámetro `mu` es μ mientras que β es el parámetro `sigma`.

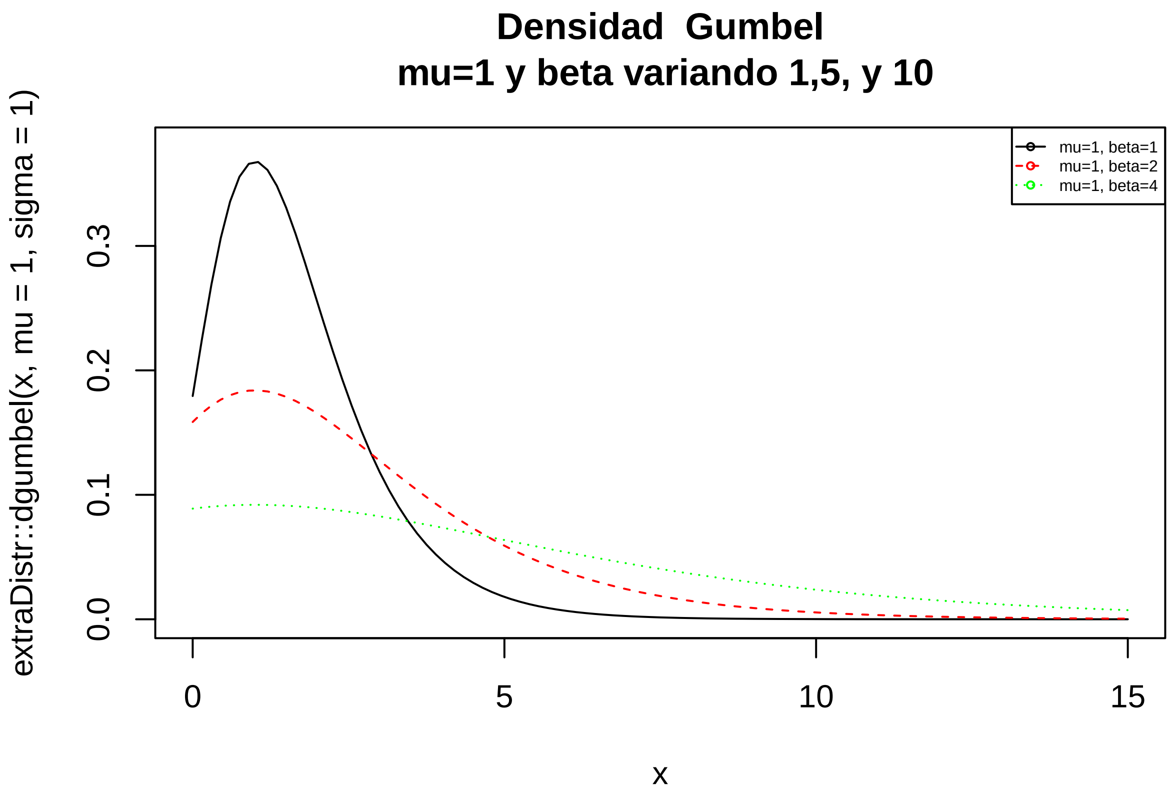
```r
# el parámetro mu es mu y el parámetro beta es sigma
curve(extraDistr::dgumbel(x,mu=1,sigma=1),xlim=c(0,10),
      ylim=c(0,0.38),lty=1,main="Densidad  Gumbel\n mu=1, 2 y 4  y beta 1")
curve(extraDistr::dgumbel(x,mu=2,sigma=1),
      add=TRUE,col="red",lty=2)
```

```
curve(extraDistr::dgumbel(x,mu=4,sigma=1),
      add=TRUE,col="green",lty=3)
legend("topright",pch=21,
       legend=c("mu=1, beta=1","mu=2, beta=1","mu=4, beta=1"),
       col=c("black","red","green"),lty=c(1,2,3),cex=0.5)
```



d) Dejando fijo el parámetro μ dibujad la densidad Gumbel para varios valores de $\beta > 0$ y explicar en que afecta a la gráfica el cambio de este parámetro.

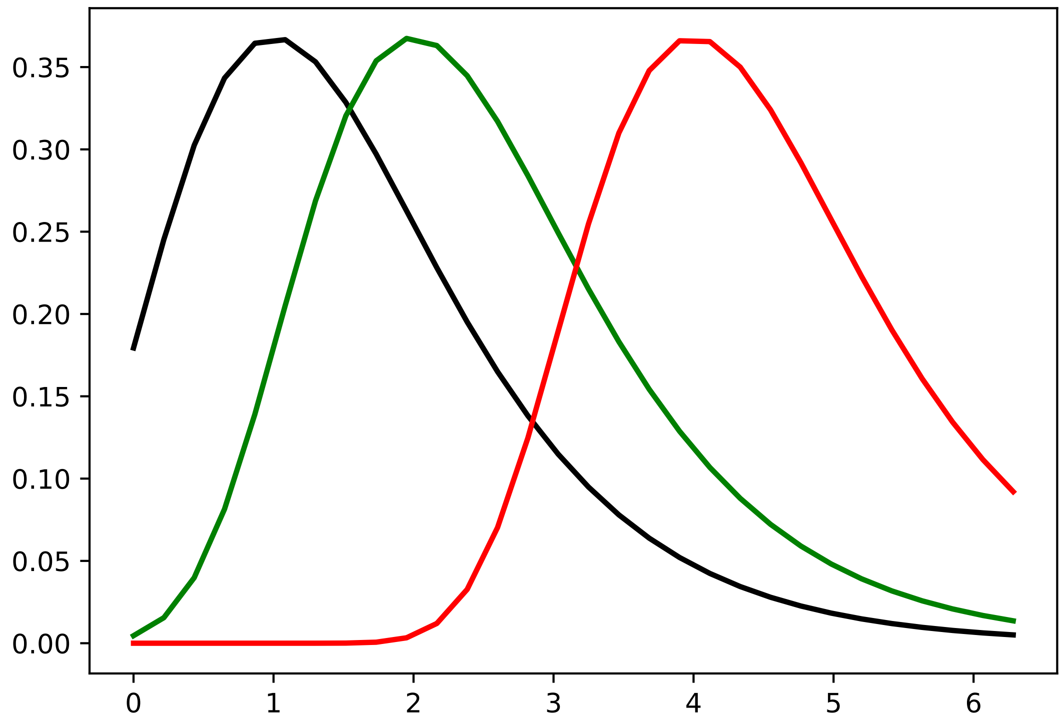
```
# el paramento mu es mu y el parámetro beta es sigma
curve(extraDistr::dgumbel(x,mu=1,sigma=1),xlim=c(0,15),
      ylim=c(0,0.38),lty=1,main="Densidad  Gumbel\n mu=1 y beta variando 1,5, y 10")
curve(extraDistr::dgumbel(x,mu=1,sigma=2),
      add=TRUE,col="red",lty=2)
curve(extraDistr::dgumbel(x,mu=1,sigma=4),
      add=TRUE,col="green",lty=3)
legend("topright",pch=21,
       legend=c("mu=1, beta=1","mu=1, beta=2","mu=1, beta=4"),
       col=c("black","red","green"),lty=c(1,2,3),cex=0.5)
```



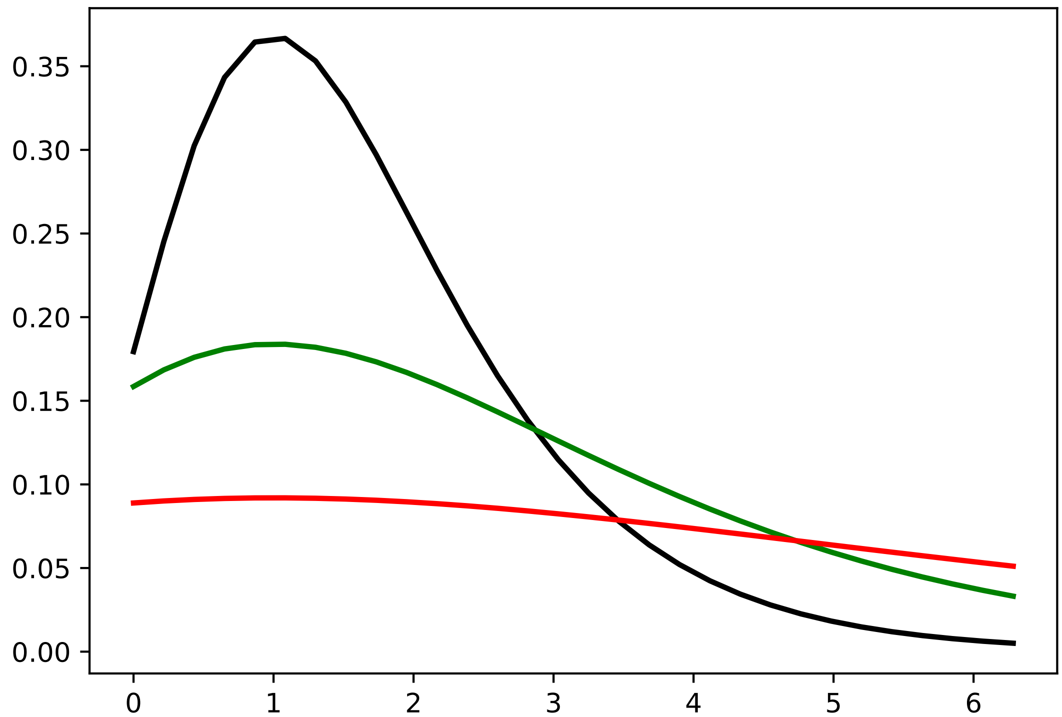
e) Si X sigue un ley Gumbel de parámetros μ y β entonces $E(X) = \mu + \beta \cdot \gamma$ donde gamma es el número de euler $\gamma = 0.577215664\ldots$, y $Var(X) = \frac{\pi^2}{6} \cdot \beta^2$

f) Repetid los apartados c) y d) con python. Con python se puede pedir con la correspondiente función la esperanza y varianza de esta distribución, comprobar con esta función para algunos valores las fórmulas de la esperanza y la varianza del apartado e).

```
import numpy as np
from scipy.stats import gumbel_r
mu, beta = 0, 0.1 # location and scale
x = np.linspace(0, 2 * np.pi, 30)
import matplotlib.pyplot as plt
#count, bins, ignored = plt.hist(s, 30, normed=True)
plt.plot(x,gumbel_r.pdf(x, loc=1, scale=1),linewidth=2, color='black')
plt.plot(x,gumbel_r.pdf(x, loc=2, scale=1),linewidth=2, color='green')
plt.plot(x,gumbel_r.pdf(x, loc=4, scale=1),linewidth=2, color='red')
plt.show()
```




```
import numpy as np
from scipy.stats import gumbel_r
mu, beta = 0, 0.1 # location and scale
x = np.linspace(0, 2 * np.pi, 30)
import matplotlib.pyplot as plt
#count, bins, ignored = plt.hist(s, 30, normed=True)
plt.plot(x,gumbel_r.pdf(x, loc=1, scale=1),linewidth=2, color='black')
plt.plot(x,gumbel_r.pdf(x, loc=1, scale=2),linewidth=2, color='green')
plt.plot(x,gumbel_r.pdf(x, loc=1, scale=4),linewidth=2, color='red')
plt.show()
```



Y los estadísticos

```
from scipy.stats import gumbel_r
gumbel_r.stats(loc=0, scale=1, moments='mv')
```

```
## (array(0.57721566), array(1.64493407))
```

```
print("E(X) = {m}".format(m=gumbel_r.stats(loc=0, scale=1, moments='m')))
```

```
## E(X) = 0.5772156649015329
```

```
print("Var(X) = {v}".format(v=gumbel_r.stats(loc=0, scale=1, moments='v')))
```

```
## Var(X) = 1.6449340668482264
```

Se observa que en este caso la esperanza es la constante de euler $\gamma = 0.577215664\ldots$.