

Inferencia

Examen final – Problema 2

27 de mayo de 2024

La siguiente muestra aleatoria simple, $c(1.25, 1.47, 0.98, 1.09, 0.39, 1.54, 0.52, 0.62, 1.19, 0.17, 0.56, 1.07, 0.66, 0.96, 0.54, 0.59, 1.44, 0.26, 2.73, 1.48, 2.6, 0.84, 0.73, 0.51, 0.53, 1.73, 0.09, 0.34, 0.43, 1.08, 0.8, 0.4, 0.69, 1.13, 0.3, 1.11, 0.38, 0.75)$, proviene de una población cuya densidad es

$$f(x) = e^{-2x} \cdot x^{p-1} \cdot 2^p / \Gamma(p) \cdot I(x > 0)$$

con $p > 0$.

1. Obtén un estadístico mínimamente suficiente para p .

Verosimilitud:

$$L(p) = e^{-2 \sum x_i} \cdot \prod x_i^{p-1} \cdot 2^{np} / \Gamma(p)^n \cdot I(x_{(1)} > 0) = h(\vec{x}) \cdot g(\prod x_i, p)$$

con $h(\vec{x}) = e^{-2 \sum x_i} \cdot I(x_{(1)} > 0)$, $g(t, p) = t^{p-1} \cdot 2^{np} / \Gamma(p)^n \implies \prod X_i$ es suficiente

$$\frac{L(\vec{x})}{L(\vec{y})} = \frac{e^{-2 \sum x_i} \cdot \prod x_i^{p-1} \cdot 2^{np} / \Gamma(p)^n \cdot I(x_{(1)} > 0)}{e^{-2 \sum y_i} \cdot \prod y_i^{p-1} \cdot 2^{np} / \Gamma(p)^n \cdot I(y_{(1)} > 0)} = e^{-2(\sum x_i - \sum y_i)} \cdot \left(\frac{\prod x_i}{\prod y_i} \right)^{p-1} \cdot \frac{I(x_{(1)} > 0)}{I(y_{(1)} > 0)}$$

\implies independiente de p si y sólo si $\frac{\prod x_i}{\prod y_i} = 1$, o sea, $\prod x_i = \prod y_i$

2. Obtén una estimación puntual por el método de los momentos para p .

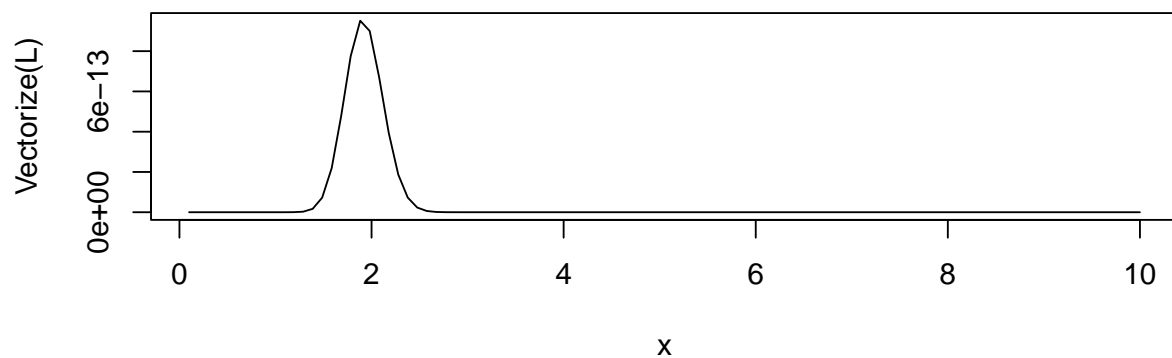
$$X \hookrightarrow \gamma(p, 2) \implies \mathcal{E}(X) = p/2 \implies \hat{p}_{MM} = 2\bar{X}$$

```
x <- c(1.25, 1.47, 0.98, 1.09, 0.39, 1.54, 0.52, 0.62, 1.19, 0.17, 0.56,
      1.07, 0.66, 0.96, 0.54, 0.59, 1.44, 0.26, 2.73, 1.48, 2.6, 0.84,
      0.73, 0.51, 0.53, 1.73, 0.09, 0.34, 0.43, 1.08, 0.8, 0.4, 0.69,
      1.13, 0.3, 1.11, 0.38, 0.75)
(pMM <- 2*mean(x))

## [1] 1.786842
```

3. Obtén una estimación puntual por el método de máxima verosimilitud para p .

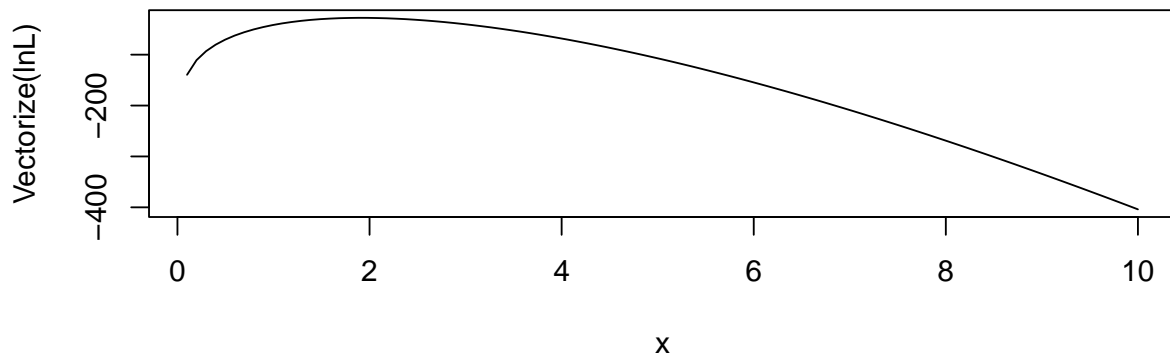
```
n <- length(x)
L <- function (p) exp(-2*sum(x)) * prod(x^(p-1)) * 2^(n*p) / gamma(p)^n *
  (min(x)>0)
L <- function (p) prod (dgamma (x, p, 2)) # equivalentemente
plot(Vectorize(L), 0.1, 10)
```



```
(pMV <- optimize(L, c(0,10), maximum=TRUE) $ maximum)

## [1] 1.909611

lnL <- function (p) sum (dgamma (x, p, 2, log=TRUE)) # logverosimilitud
plot(Vectorize(lnL), 0.1, 10)
```



```
optimize(lnL, c(0,10), maximum=TRUE) $ maximum      # pMV de nuevo
## [1] 1.909601
```

4. Obtén un intervalo de confianza para p a nivel $1 - \alpha = 0,95$.

$$X_i \hookrightarrow \gamma(p, 2)$$

No se puede obtener un pivote exacto, porque el parámetro p cambia la forma de la distribución. Podemos usar *bootstrap* o, como $n > 30$, intervalos asintóticos:

$$\sum X_i \hookrightarrow \gamma(np, 2) \Rightarrow \mathcal{E}(2\bar{X}) = p$$

bootstrap percentil paramétrico:

```
quantile (replicate(1e5, 2*mean(rgamma(n, pMV, 2))), c(.025, .975))
##      2.5%      97.5%
## 1.495590 2.374693
```

bootstrap percentil no paramétrico:

```
quantile (replicate(1e5, 2*mean(sample(x, , TRUE))), c(.025, .975))
##      2.5%      97.5%
## 1.437368 2.176316
```

bootstrap básico no paramétrico:

```
pMM - quantile (replicate(1e5, 2*mean(sample(x, TRUE))-pMM), c(.975, .025))
##      97.5%      2.5%
## 1.394211 2.138947
```

t-bootstrap paramétrico:

```
pMM - quantile (replicate(1e5,
                        ((pMMb <- 2*mean(rgamma(n, pMV, 2))) - pMM)
                        /sqrt(pMMb/n)),
                c(.975, .025)) *
sqrt(pMM/n)
##      97.5%      2.5%
## 1.277247 2.106379
```

TCL pues $n = 38 > 30$:

$$\begin{aligned}
 X &\hookrightarrow \gamma(p, 2) \Rightarrow \bar{X} \overset{\sim}{\hookrightarrow} \mathcal{N}\left(\frac{p}{2}, \sqrt{\frac{p}{4n}}\right) \\
 1 - \alpha &\approx \Pr\left[-z_{1-\alpha/2} \leq \frac{\bar{X} - p/2}{\sqrt{p}/2\sqrt{n}} = \frac{2\bar{X} - p}{\sqrt{p}/\sqrt{n}} \leq z_{1-\alpha/2}\right] \\
 &= \Pr\left[-z \frac{\sqrt{p}}{\sqrt{n}} \leq 2\bar{X} - p \leq z \frac{\sqrt{p}}{\sqrt{n}}\right] = \Pr\left[(2\bar{X} - p)^2 \leq z^2 \frac{p}{n}\right] \\
 &= \Pr\left[4\bar{X}^2 - 4\bar{X}p + p^2 - \frac{z^2}{n}p \leq 0\right] = \Pr\left[p^2 \cdot (1) + p \cdot \left(-4\bar{X} - \frac{z^2}{n}\right) + (4\bar{X}^2) \leq 0\right] \\
 &= \Pr\left[p \in \frac{4\bar{X} + z^2/n \pm \sqrt{(4\bar{X} + z^2/n)^2 - 4 \cdot 4\bar{X}^2}}{2}\right]
 \end{aligned}$$

```
z <- qnorm(.975)
m <- mean(x)
(4*m+z^2/n + c(-1,1)*sqrt((4*m+z^2/n)^2-4*4*m^2))/ 2
## [1] 1.409382 2.265393
```

Método delta:

$$\bar{X} \overset{\sim}{\hookrightarrow} \mathcal{N}\left(\frac{p}{2}, \sqrt{\frac{p}{4n}}\right) \implies \sqrt{\bar{X}} \overset{\sim}{\hookrightarrow} \mathcal{N}\left(\sqrt{\frac{p}{2}}, \sqrt{\frac{p}{4n}} \cdot \left|\frac{-1}{2\sqrt{\frac{p}{2}}}\right|\right) = \mathcal{N}\left(\sqrt{\frac{p}{2}}, \frac{1}{\sqrt{8n}}\right)$$

$$\begin{aligned}
1 - \alpha &\approx \Pr \left[-z \leq \frac{\sqrt{\bar{X}} - \sqrt{p/2}}{1/\sqrt{8n}} \leq z \right] = \Pr \left[\frac{-z}{\sqrt{8n}} \leq \sqrt{\bar{X}} - \sqrt{\frac{p}{2}} \leq \frac{z}{\sqrt{8n}} \right] \\
&= \Pr \left[-\sqrt{\bar{X}} + \frac{-z}{\sqrt{8n}} \leq -\sqrt{\frac{p}{2}} \leq -\sqrt{\bar{X}} + \frac{z}{\sqrt{8n}} \right] \\
&= \Pr \left[\sqrt{\bar{X}} - \frac{-z}{\sqrt{8n}} \geq \sqrt{\frac{p}{2}} \geq \sqrt{\bar{X}} - \frac{z}{\sqrt{8n}} \right] \\
&= \Pr \left[\sqrt{\bar{X}} + \frac{z}{\sqrt{8n}} \geq \sqrt{\frac{p}{2}} \geq \sqrt{\bar{X}} - \frac{z}{\sqrt{8n}} \right] \\
&= \Pr \left[\sqrt{\bar{X}} - \frac{z}{\sqrt{8n}} \leq \sqrt{\frac{p}{2}} \leq \sqrt{\bar{X}} + \frac{z}{\sqrt{8n}} \right] \\
&= \Pr \left[2 \left(\sqrt{\bar{X}} - \frac{z}{\sqrt{8n}} \right)^2 \leq p \leq 2 \left(\sqrt{\bar{X}} + \frac{z}{\sqrt{8n}} \right)^2 \right]
\end{aligned}$$

```
2 * (sqrt(m) + c(-1,1)*z/sqrt(8*n)) ^ 2
```

```
## [1] 1.387105 2.237125
```

5. Comprueba si la muestra sigue la distribución del enunciado.

El paquete `goftest` de R permite¹ aplicar CvM y AD para contrastarlo. Por ejemplo:

```
goftest::ad.test (x, "pgamma", pMV, 2)

##
## Anderson-Darling test of goodness-of-fit
## Null hypothesis: Gamma distribution
## Parameters assumed to be fixed
##
## data: x
## An = 0.41803, p-value = 0.8298

goftest::ad.test (x, "pgamma", pMV, 2) $ p.value
```

¹Poco rigurosamente, porque tales procedimientos deben aplicarse sólo cuando hay cambios de localización y escala, no de forma.

```
## [1] 0.8297679

gofest::ad.test (x, "pgamma", pMV, 2) $ p.value # constante

## [1] 0.8297679

gofest::ad.test (x, "pgamma", pMV, 2, estimated=TRUE)

##
## Anderson-Darling test of goodness-of-fit
## Braun's adjustment using 6 groups
## Null hypothesis: Gamma distribution
## Parameters assumed to have been estimated from data
##
## data: x
## Anmax = 1.6259, p-value = 0.6235

gofest::ad.test (x, "pgamma", pMV, 2, estimated=TRUE) $ p.value

## [1] 0.9113858

gofest::ad.test (x, "pgamma", pMV, 2, estimated=TRUE) $ p.value # varía

## [1] 0.5227473
```

En este caso habría que considerar p como *estimated* y $a = 2$ como conocido, pero `ad.test` no permite eso. En todo caso, los p -valores tan altos evitan rechazar H_0 .
 χ^2 de bondad de ajuste:

```
k <- floor(n/5) # para que no haya problemas con la distribución asintótica
cortes <- qgamma ((0:k)/k, pMV, 2)
obs <- table (cut (x, cortes))
1 - pchisq (chisq.test(obs, p=rep(1/k,k)) $ statistic, k-2)

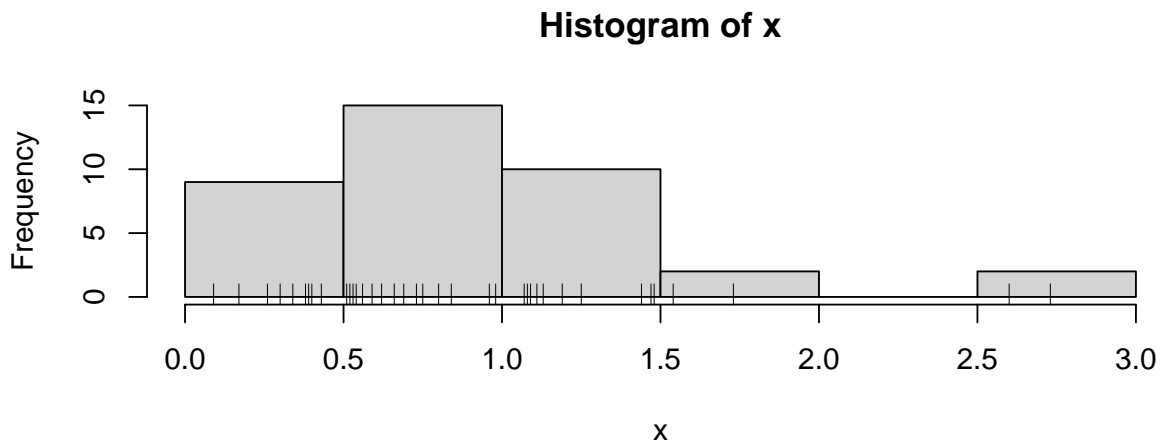
## X-squared
## 0.549416
```

6. Describe gráfica y numéricamente la muestra.

```
round (c (summary(x), n=n, dt=sd(x), cv=sd(x)/mean(x)), 2)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.     n      dt      cv
##  0.09   0.51   0.74   0.89   1.12   2.73  38.00  0.59  0.66
```

```
hist(x) ; rug(x, .1)
```



Es una distribución de valores positivos, con bastante dispersión (coeficiente de variación = 66 %) y asimétrica hacia la derecha.

7. Contrasta a nivel $\alpha = 0,05$ las hipótesis $H_0 : p = 1, H_1 : p = 2$.

Neyman-Pearson:

$$\frac{L(2)}{L(1)} = \frac{e^{-2 \sum x_i} \cdot \prod x_i \cdot 2^{2n} / \Gamma(2)^n \cdot I(x_{(1)} > 0)}{e^{-2 \sum x_i} \cdot 2^n / \Gamma(1)^n \cdot I(x_{(1)} > 0)} = \prod x_i \cdot 2^n > k \Leftrightarrow \prod x_i > k'$$

$H_0 \Rightarrow X_i \leftrightarrow \gamma(1, 2)$:

```
distri <- replicate (1e5, prod(rgamma(n,1,2)))
mean (distri > prod(x)) # p-valor aprox
```

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

Menos potente:

```

ks.test (x, pgamma, 1, 2)

##
## Exact one-sample Kolmogorov-Smirnov test
##
## data: x
## D = 0.40256, p-value = 4.497e-06
## alternative hypothesis: two-sided

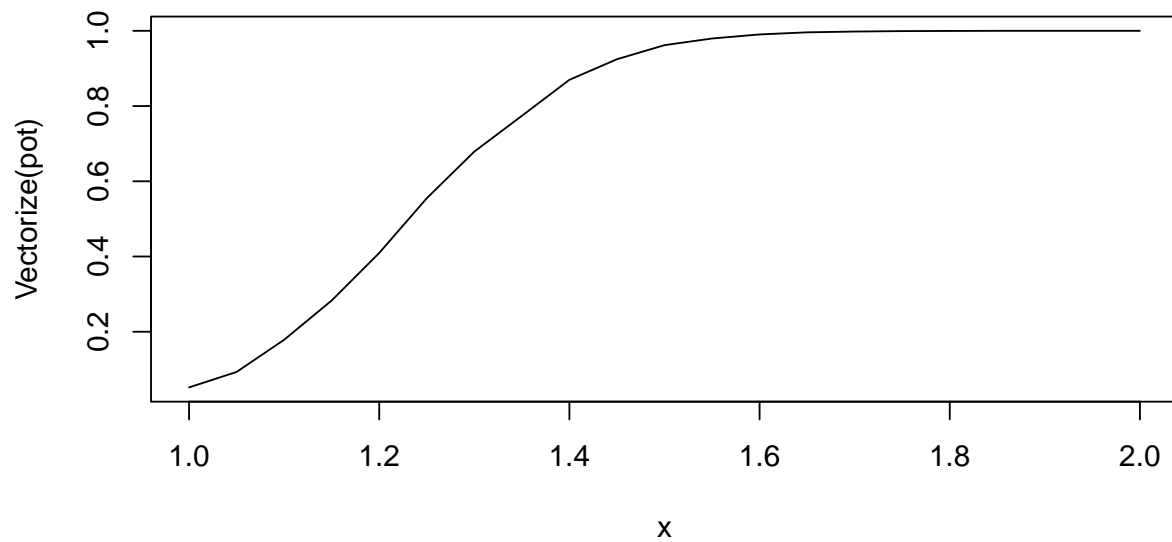
```

8. Representa gráficamente la función potencia para el contraste $H_0 : p = 1, H_1 : p > 1$ con $\alpha = 0,05$.

```

RC <- quantile (distrib, .95)
pot <- function (p) mean (replicate (1e4, prod(rgamma(n,p,2)) > RC))
plot (Vectorize(pot), 1, 2, n=21)

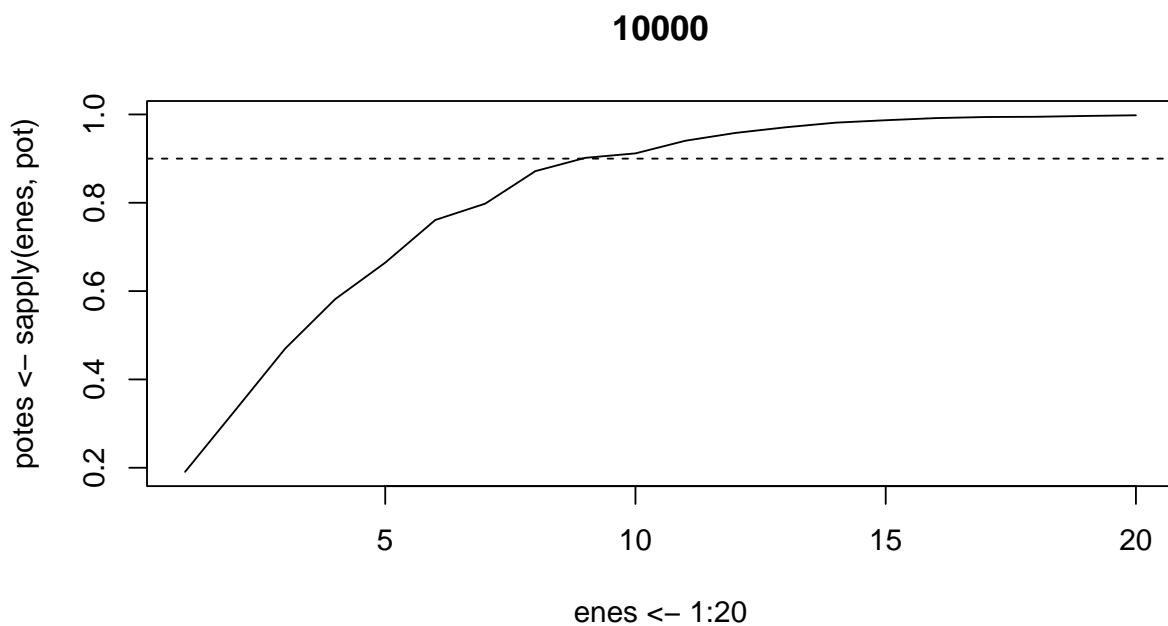
```



9. **Calcula el tamaño de muestra necesario para obtener una potencia mayor que 0,9 en el contraste $H_0 : p = 1, H_1 : p = 2$ con $\alpha = 0,05$.**

Usando Neyman-Pearson (el más potente):

```
B <- 1e4
pot <- function (n)
{
  RC <- quantile (replicate (B, prod(rgamma(n,1,2))), .95)
  mean (replicate (B, prod(rgamma(n,2,2)) > RC))
}
plot (enes <- 1:20, potes <- sapply(enes, pot), type="l", main=B)
abline (.9, 0, lty=2)
```



```
enes [potes>.9] [1]
```

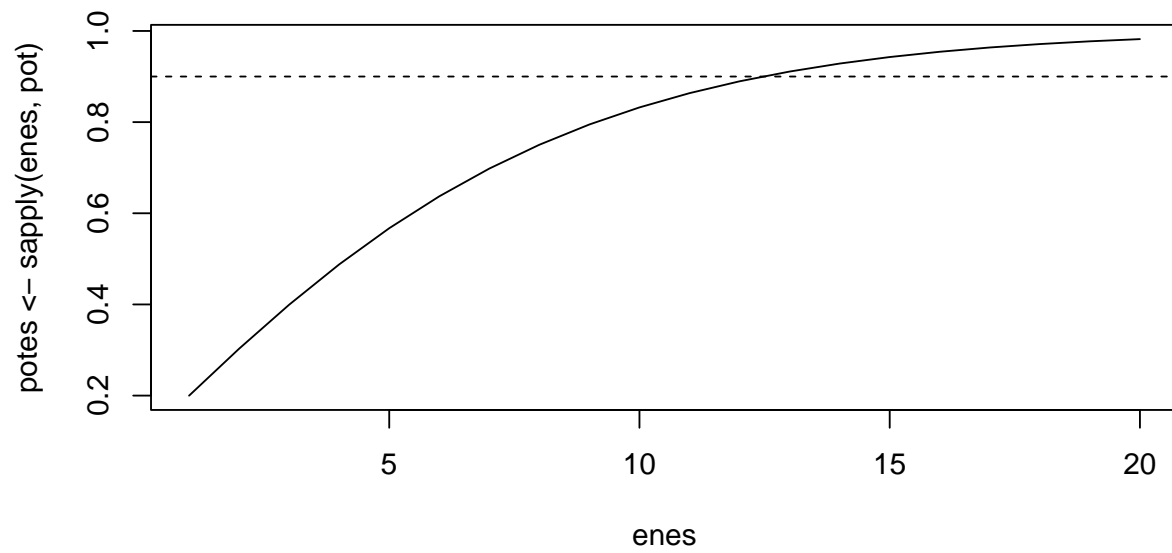
```
## [1] 9
```

Usando el estadístico $\sum X_i \leftrightarrow \gamma(np, 2)$, basado en el estimador del método de los momentos:

```

pot <- function (n)
{
  RC <- qgamma (.95, n * 1, 2)
  1 - pgamma (RC, n * 2, 2)
}
plot (enes, potes <- sapply(enes,pot), type="l")
abline (.9, 0, lty=2)

```



```

enes [potes>.9] [1]
## [1] 13

```