

Entrega 2 - Parte 2

Ignacio Spiousas

6 de junio de 2023

Ejercicio 5 de la práctica 2

Enunciado: La cantidad de lluvia (en pulgadas) que cae en una tormenta en cierta región puede modelarse mediante una distribución $\Gamma(\alpha, \lambda)$. En 100 tormentas se observan las cantidades que encontrarás en el campus virtual.

a) Los estimadores

Enunciado: Hallar estimaciones basadas en el método de momentos y de máxima verosimilitud de α y λ . Para el estimador de MV, explorar el comando `fitdistr(x, densfun = "gamma")` del paquete MASS.

En este caso vamos a tener que cada medición de precipitaciones es independiente y puede modelarse mediante una distribución Γ de parámetros α y λ

$$X_i \stackrel{i.i.d}{\sim} \Gamma(\alpha, \lambda)$$

Como ahora tenemos dos parámetros vamos a necesitar 900 los dos primeros momentos para poder plantear un sistema de ecuaciones y despejar los estimadores. En el caso de la distribución gamma, la densidad tiene la siguiente forma:

$$f_X(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}$$

El primer momento

Planteo la ecuación del primer momento de la siguiente forma:

$$\frac{1}{n} \sum_{i=1}^n X_i = E_{\hat{\alpha}\hat{\lambda}}(X_1)$$

Donde la esperanza de la gamma está definida por:

$$E_{\hat{\alpha}\hat{\lambda}}(X_1) = \frac{\hat{\alpha}}{\hat{\lambda}}$$

Entonces podemos decir que:

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n X_i &= \frac{\hat{\alpha}}{\hat{\lambda}} \\ \bar{X}_n &= \frac{\hat{\alpha}}{\hat{\lambda}} \end{aligned}$$

El segundo momento

Planteo la ecuación del segundo momento de la siguiente forma:

$$\frac{1}{n} \sum_{i=1}^n X_i^2 = E_{\hat{\alpha}\hat{\lambda}}(X_1^2)$$

La esperanza de la V.A. al cuadrado se puede reescribir como:

$$\begin{aligned} E_{\hat{\alpha}\hat{\lambda}}(X_1^2) &= Var_{\hat{\alpha}\hat{\lambda}}(X_1) + (E_{\hat{\alpha}\hat{\lambda}}(X_1))^2 \\ E_{\hat{\alpha}\hat{\lambda}}(X_1^2) &= \frac{\hat{\alpha}}{\hat{\lambda}^2} + \frac{\hat{\alpha}^2}{\hat{\lambda}^2} \\ E_{\hat{\alpha}\hat{\lambda}}(X_1^2) &= \frac{\hat{\alpha}}{\hat{\lambda}^2} (1 + \hat{\alpha}) \end{aligned}$$

Entonces, la segunda ecuación del sistema de ecuaciones queda:

$$\frac{1}{n} \sum_{i=1}^n X_i^2 = \frac{\hat{\alpha}}{\hat{\lambda}^2} (1 + \hat{\alpha})$$

El sistema de ecuaciones

Pasando en limpio, el sistema de ecuaciones queda:

$$\begin{cases} \frac{1}{n} \sum_{i=1}^n X_i^2 &= \frac{\hat{\alpha}}{\hat{\lambda}^2} (1 + \hat{\alpha}) \\ \frac{1}{n} \sum_{i=1}^n X_i &= \frac{\hat{\alpha}}{\hat{\lambda}} \end{cases}$$

Operando podemos despejar los valores de los estimadores:

$$\begin{aligned} \hat{\alpha} &= \frac{\bar{X}_n^2}{\frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}_n^2} \\ \hat{\lambda} &= \frac{\hat{\alpha}}{\bar{X}_n} = \frac{\bar{X}_n}{\frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}_n^2} \end{aligned}$$

Ahora para los estimadores de máxima verosimilitud vamos a utilizar el comando `fitdistr(x, densfun = "gamma")`. A modo de ejemplo vamos a simular unos datos y a recuperar el valor de sus parámetros. Vamos a simular 100 realizaciones de una V.A. X (en el vector `x`).

Supongamos que $X_i \stackrel{i.i.d}{\sim} \Gamma(\alpha = 4, \lambda = 2)$, en la función `rgamma`, que genera realizaciones de la gamma, los parámetros equivalente se llaman `shape` (α) y `rate` (λ). Simulamos 100 realizaciones y recuperamos los parámetros.

```
set.seed(15)
N = 100
alpha = 4
lambda = 2
x = rgamma(n = N, shape = alpha, rate = lambda)

fitdistr(x, densfun = "gamma")

##      shape      rate
## 4.3516231 1.9585061
## (0.5933280) (0.2830585)
```

Alternativamente podemos utilizar los estimadores de momentos para los mismos datos simulados.

```
alpha_mom <- mean(x)^2/(1/N*sum(x^2)-mean(x)^2)
lambda_mom <- mean(x)/(1/N*sum(x^2)-mean(x)^2)
cat(paste("El alpha estimado por momentos es:", round(alpha_mom, digits = 4),
        "\nEl lambda estimado por momentos es:", round(lambda_mom, digits = 4)))
```

```
## El alpha estimado por momentos es: 4.7322
## El lambda estimado por momentos es: 2.1298
```

b) La estimación

Enunciado: Dar una estimación de la probabilidad de que la cantidad de lluvia en una tormenta en esta región supere las 97 pulgadas, basado en los estimadores de máxima verosimilitud.

Los parámetros estimados por cada método son:

```
x = read.csv(file = here("metodos_estimacion/TPs/entrega_2/data/lluvia.csv"))[["x"]]
alpha_mom <- mean(x)^2/(1/N*sum(x^2)-mean(x)^2)
lambda_mom <- mean(x)/(1/N*sum(x^2)-mean(x)^2)
alpha_mv <- fitdistr(x, densfun = "gamma")$estimate[1]
lambda_mv <- fitdistr(x, densfun = "gamma")$estimate[2]
cat(paste("El alpha estimado por momentos es:", round(alpha_mom, digits = 4),
        "\nEl lambda estimado por momentos es:", round(lambda_mom, digits = 4)))
```

```
## El alpha estimado por momentos es: 0.5507
## El lambda estimado por momentos es: 0.0037
```

```
cat(paste("El alpha estimado por MV es:", round(alpha_mv, digits = 4),
        "\nEl lambda estimado por MV es:", round(lambda_mv, digits = 4)))
```

```
## El alpha estimado por MV es: 0.5017
## El lambda estimado por MV es: 0.0033
```

Y ahora podemos usar los de MV para estimar la probabilidad de que $X > 126$.

$$P(X > 126) = 1 - P(X < 126) = 1 - F(x = 126)$$

```
cat(paste("La probabilidad estimada es:", round(1-pgamma(q = 126,
        shape = round(alpha_mv, digits = 4),
        rate = round(lambda_mv, digits = 4)),
        digits = 4)))
```

```
## La probabilidad estimada es: 0.363
```

c) Comparemos las estimaciones

Enunciado: Hacer un histograma de las observaciones, superponer en color negro una estimación no paramétrica de la densidad, en color rojo, la densidad de la gamma con los parámetros estimados por el método de los momentos y, en color azul, la densidad de la gamma con los parámetros estimados por el método de máxima verosimilitud. ¿Qué observa?

Primero hagamos un eje x común para todas las estimaciones:

```
n_x <- 1024
x_i <- seq(min(x), max(x), (max(x)-min(x))/(n_x-1))
```

Ahora definamos una función que calcule la densidad de una gamma dados X , α y λ

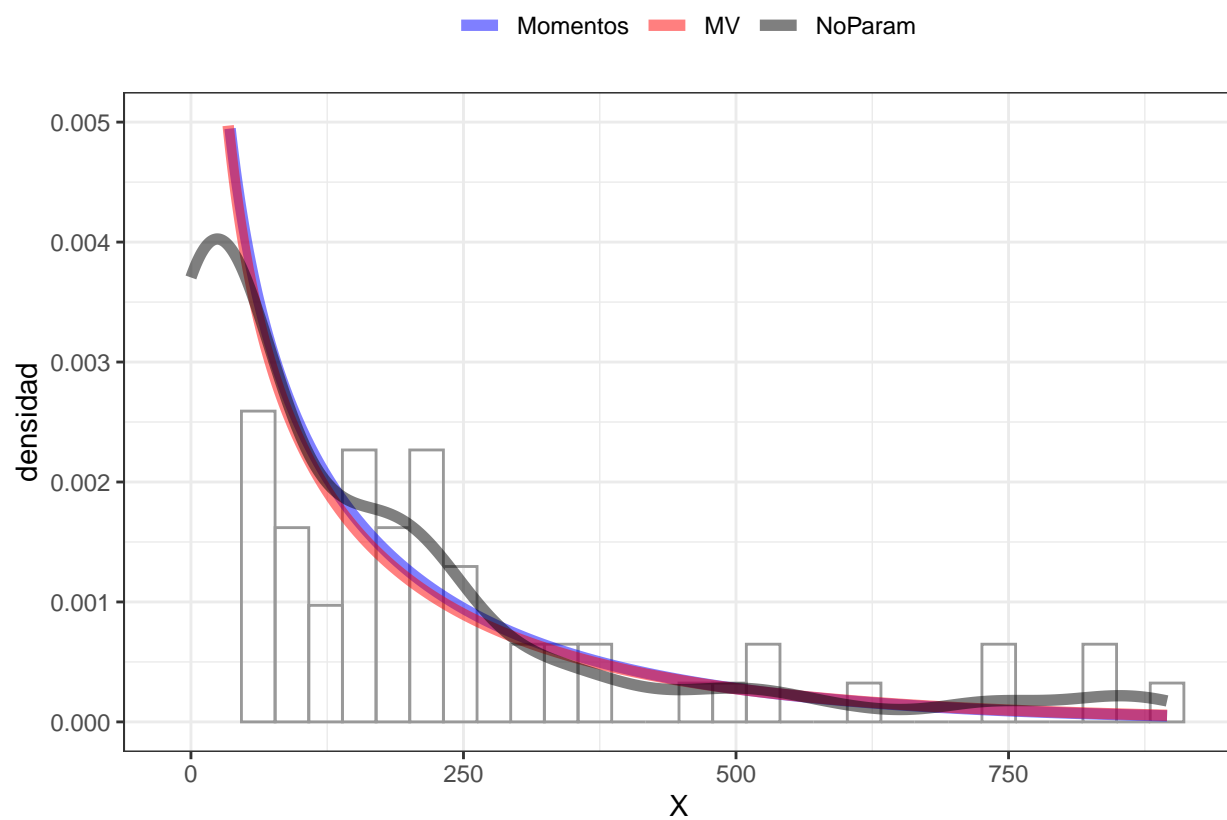
```
gamma_dens <- function(alpha, lambda, x) {
  lambda^alpha/gamma(alpha) * x^(alpha-1) * exp(-lambda*x)
}
```

Ahora armemos un tibble con todas las estimaciones de la densidad:

```
estimaciones <- tibble(NoParam = density(x,
                                         from = min(x),
                                         to = max(x),
                                         n = n_x)$y,
                      Momentos = gamma_dens(alpha = alpha_mom,
                                              lambda = lambda_mom,
                                              x = x_i),
                      MV = gamma_dens(alpha = alpha_mv,
                                       lambda = lambda_mv,
                                       x = x_i),
                      x = x_i) %>%
  pivot_longer(cols = !x, names_to = "metodo", values_to = "estim")
```

Ahora grafiquemos:

```
tibble(x = x) %>%
  ggplot(aes(x = x, y = after_stat(density))) +
  geom_histogram(color = "gray60",
                alpha = 0,
                bins = 30) +
  geom_line(data = estimaciones,
            aes(x = x, y = estim, color = metodo),
            linewidth = 2,
            alpha = .5) +
  scale_color_manual(values = c("blue", "red", "black")) +
  labs(x = "X",
       y = "densidad",
       color = NULL) +
  ylim(c(0, 0.005)) +
  theme_bw() +
  theme(legend.position = "top")
```



La estimación no paramétrica da valores más bajos de densidad para valores chicos de X mientras que las estimadas a partir de momento y MV dar valores más grandes.