Proyecto 2

Sara Leiva, Pedro Pablo Sanín

Introducción

La distribución Gamma con parámetros α y β se usa en las ciencias para modelar tiempos de vida de organismos o artefactos. Su densidad de probabilidad está dada por

$$\frac{x^{\alpha-1}e^{-x/\beta}}{\Gamma(\alpha)\beta^{\alpha}}$$

y está soportada para $x > 0$. Se tiene además que

$$\mu = \mathbb{E}(X) = \alpha\beta, \quad \sigma^2 = \text{Var}(X) = \alpha\beta^2$$

.

El objetivo de este proyecto es calcular el Estimador de Máxima Verosimilitud del parámetro $\theta = (\alpha, \beta)$ de este modelo, y verificar la validez del conjunto elíptico de confianza obtenido al estimar el parámetro con el Teorema del Límite Central para Estimadores de Máxima Verosimilitud biparamétrico.

A lo largo del proyecto, supondremos que contamos con n muestras iid. $\mathcal{X} = (X_1, \ldots, X_n)$ con distribución Gamma(α, β) para $\alpha = \alpha_0 = 2.5$ y $\beta = \beta_0 = 4.6$.

Cálculo numérico del Estimador de Máxima Verosimilitud para el parámetro θ

Método de Newton-Raphson para calcular el EMV La verosimilitud del parámetro θ para la muestra $\mathcal{X}$ está dada por la función bivariada

$$L(\theta) = \prod_{i=1}^{n} f(X_i; \alpha, \beta) = \prod_{i=1}^{n} \frac{{X_i}^{\alpha-1}e^{-X_i/\beta}}{\Gamma(\alpha)\beta^{\alpha}}$$

. Así, la log-verosimilitud es:

$$l(\theta) = \sum_{i=1}^{n}\left((\alpha-1)\ln(X_i) - \frac{X_i}{\beta} - \ln(\Gamma(\alpha)) - \alpha\ln(\beta)\right) = -n\ln(\Gamma(\alpha)) - n\alpha\ln(\beta) + \sum_{i=1}^{n}\left((\alpha-1)\ln(X_i) - \frac{X_i}{\beta}\right).$$

Para encontrar el valor de θ que minimiza $l(\theta)$, es decir, el estimador de máxima verosimilitud para los valores observados, debemos solucionar las ecuaciones

$$\frac{\partial l(\theta)}{\partial \alpha} = 0, \qquad \frac{\partial l(\theta)}{\partial \beta} = 0.$$

Vea entonces que

$$\frac{\partial l(\theta)}{\partial \alpha} = -n\frac{\Gamma'(\alpha)}{\Gamma(\alpha)} - n\ln(\beta) + \sum_{i=1}^{n}\ln(X_i).$$

Asimismo

$$\frac{\partial l(\theta)}{\partial \beta} = -\frac{n\alpha}{\beta} + \frac{1}{\beta^2}\sum_{i=1}^{n} X_i.$$

Para solucionar el sistema de ecuaciones, podemos usar métodos numéricos, pues la solución simbólica es difícil. En este caso, usaremos el método de Newton-Raphson bivariado. Para esto, debemos identificar un estimador inicial $\tilde{\theta}_0$ y a partir de esto construimos las aproximaciones a la raíz por la fórmula:

$$\tilde{\theta}_{k+1} = \tilde{\theta}_k - [Df(\tilde{\theta}_k)]^{-1} f(\tilde{\theta}_k),$$

siendo $f : X \subseteq \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ con

$$f(\alpha, \beta) = (\frac{\partial l(\theta)}{\partial \alpha}(\alpha, \beta), \frac{\partial l(\theta)}{\partial \beta}(\alpha, \beta))$$

Así, si $\tilde{\theta}_k = (\tilde{\alpha}_k, \tilde{\beta}_k)$, tenemos que

$$Df(\tilde{\theta}_k) = \begin{pmatrix} -n\dfrac{\Gamma''(\tilde{\alpha}_k)\Gamma(\tilde{\alpha}_k) - \Gamma'(\tilde{\alpha}_k)^2}{\Gamma(\tilde{\alpha}_k)^2} & -\dfrac{n}{\tilde{\beta}_k} \\ -\dfrac{n}{\tilde{\beta}_k} & \dfrac{n\tilde{\alpha}_k}{\tilde{\beta}_k^2} - 2\dfrac{\sum_{i=1}^n X_i}{\tilde{\beta}_k^3} \end{pmatrix}.$$

Invirtiendo la matriz, obtenemos:

$$[Df(\tilde{\theta}_k)]^{-1} = \frac{1}{-n\dfrac{\Gamma''(\tilde{\alpha}_k)\Gamma(\tilde{\alpha}_k) - \Gamma'(\tilde{\alpha}_k)^2}{\Gamma(\tilde{\alpha}_k)^2}\left(\dfrac{n\tilde{\alpha}_k}{\tilde{\beta}_k^2} - 2\dfrac{\sum_{i=1}^n X_i}{\tilde{\beta}_k^3}\right) - \dfrac{n^2}{\tilde{\beta}_k^2}} \begin{pmatrix} \dfrac{n\tilde{\alpha}_k}{\tilde{\beta}_k^2} - 2\dfrac{\sum_{i=1}^n X_i}{\tilde{\beta}_k^3} & \dfrac{n}{\tilde{\beta}_k} \\ \dfrac{n}{\tilde{\beta}_k} & -n\dfrac{\Gamma''(\tilde{\alpha}_k)\Gamma(\tilde{\alpha}_k) - \Gamma'(\tilde{\alpha}_k)^2}{\Gamma(\tilde{\alpha}_k)^2} \end{pmatrix}.$$

Sabemos que, para asegurarnos de que el método de Newton-Raphson converge, debemos elegir un estimador inicial apropiado. Sabemos que podemos encontrar estimadores suficientemente buenos para α y β usando los momentos de la distribución. Sabemos de las ecuaciones anteriorese que

$$\alpha = \frac{\mu^2}{\sigma^2}, \qquad \beta = \frac{\sigma^2}{\mu}$$

y así, podemos definir

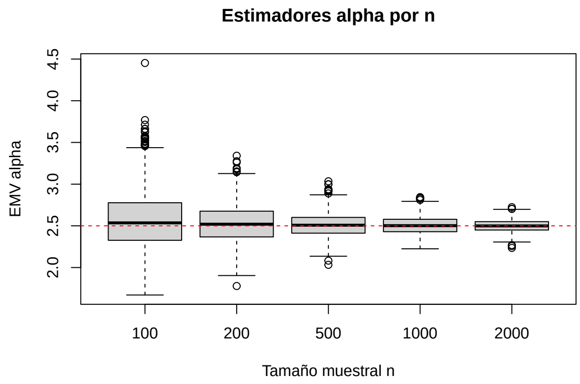
$$\tilde{\alpha}_0 = \frac{\overline{X}^2}{s^2}, \qquad \tilde{\beta}_= \frac{s^2}{\overline{\overline{X}}}$$

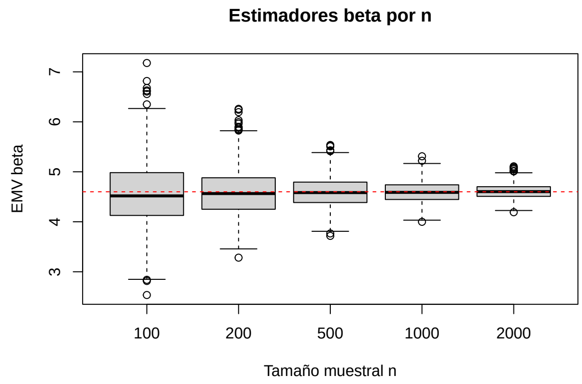
Calculando el EMV A continuación mostramos el código R usado para calcular el estimador

```
set.seed(1)

mle_gamma_nr <- function(sample, precision = 1e-6, max_iteraciones = 50){
  # Estadísticos básicos
  sample.size     <- length(sample)
  sample.mean     <- mean(sample)
  sample.variance <- var(sample)
  sample.sum      <- sum(sample)
  sample.sumlog   <- sum(log(sample))

  iteraciones <- 0L
  diferencia <- Inf

  # Gradiente de la log-verosimilitud
  likelihood_gradient <- function(theta_k){
    alpha_k <- theta_k[1]
```

```r
    beta_k  <- theta_k[2]
    dalpha  <- -sample.size * digamma(alpha_k) - sample.size * log(beta_k) + sample.sumlog
    dbeta   <- (-sample.size * alpha_k / beta_k) + (sample.sum / beta_k^2)
    return(c(dalpha, dbeta))
  }

  # Inversa del Hessiano
  inverse_hessian <- function(theta_k){
    alpha_k <- theta_k[1]
    beta_k  <- theta_k[2]
    a <- -sample.size * trigamma(alpha_k)
    b <- -sample.size / beta_k
    d <- (sample.size * alpha_k / beta_k^2) - 2 * (sample.sum / beta_k^3)
    detA <- a * d - b^2
    return((1 / detA) * matrix(c(d, -b, -b, a), nrow = 2, byrow = TRUE))
  }

  # Inicialización por momentos
  theta_ant <- c((sample.mean)^2 / sample.variance, sample.variance / sample.mean)

  #Verificamos que los estimadores estén bien definidos y sean positivos.
  if (any(!is.finite(theta_ant)) || any(theta_ant <= 0)) {
    #theta_ant <- pmax(theta_ant, 1e-8)
  }

  repeat {
    # Paso de Newton: theta_nuevo = theta_ant - inv(H) %*% grad
    paso <- drop(inverse_hessian(theta_ant) %*% likelihood_gradient(theta_ant))
    theta_nuevo <- theta_ant - paso

    theta_nuevo[!is.finite(theta_nuevo) | theta_nuevo <= 0] <- 1e-8

    iteraciones <- iteraciones + 1L
    diferencia <- max(abs(theta_nuevo - theta_ant))

    if (diferencia < precision || iteraciones >= max_iteraciones) break
    theta_ant <- theta_nuevo
  }

  return(list(
    theta = theta_nuevo,
    iteraciones = iteraciones,
    convergio = (diferencia < precision)
  ))
}
```

Ahora, generaremos 1000 muestras de tamaños $n = 100, 200, 500, 1000 y 2000$, y dibujaremos para comparar los distintos estimadores obtenidos para cada n.





Podemos ver que a medida que aumenta el tamaño muestral, la varianza de los estimadores se reduce y la media tiende al parámetro real según el cual se generaron los datos.

Conjuntos de confianza

Sabemos del TLC para el EMV que

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \to \mathcal{N}(0, \frac{1}{I(\theta_0)})$$

en distribución. Por el teorema de Slutski tenemos que

$$\sqrt{n} I_F(\hat{\theta}_n)^{1/2}(\hat{\theta}_n - \theta_0) \to I_F(\theta_0)^{1/2}\mathcal{N}(0, \frac{1}{I(\theta_0)})$$

en distribución y así

$$\sqrt{n} I_F(\hat{\theta}_n)^{1/2}(\hat{\theta}_n - \theta_0) \to \mathcal{N}(0, I)$$

(definimos $A^{1/2}$ tal que $A = A^{1/2}A^{1/2}$ asumimos la continuidad de la función $A \mapsto A^{1/2}$).

Vea que

Ahora, de la convergencia deducimos que

$$n(\hat{\theta}_n - \theta_0)^T I_F(\hat{\theta}_n)(\hat{\theta}_n - \theta_0) \to \chi^2_2$$

. Así, si ρ es el cuantil 95% de la distribución χ^2_2, el conjunto de confianza

$$\mathcal{C} = \{\theta : n(\hat{\theta}_n - \theta)^T I_F(\hat{\theta}_n)(\hat{\theta}_n - \theta) \leq \rho\}$$

contiene al θ_0 con probabilidad aproximada de 95%.

A continuación, basado en los datos calculados podemos encontrar la proporción de datos que cumplen la condición anterior. Para eso considere el siguiente código:

```
nivel <- 0.95
rho <- qchisq(nivel, 2)

I_F <- function(theta) {
  alpha <- theta[1]
  beta <- theta[2]

  a <- trigamma(alpha)
```

```
  b <- 1/beta
  c <- 1/beta
  d <- alpha/(beta^2)

  return(matrix(c(a, b, c, d), nrow=2, ncol=2, byrow=TRUE))
}

esta_en_C <- function(theta_n, theta, n, rho) {
  diff <- theta_n - theta
  IF <- I_F(theta_n)
  value <- n * t(diff) %*% IF %*% diff
  return(value <= rho)
}
```

Al calcular el porcentaje de parámetros en los resultados que cumplen con la condición respecto a θ_0 obtenemos:

```
##      n porcentaje
## 1  100       91.5
## 2  200       92.6
## 3  500       94.2
## 4 1000       94.3
## 5 2000       94.8
```

Podemos graficar los resultados así: