

Familias de Distribuciones Conjugadas: Exponencial y Gamma

José Jaén Delgado

Segundo Trabajo

En el presente proyecto estudiamos las distribuciones conjugadas Exponencial y Gamma utilizando los métodos propios de la Estadística Bayesiana. El código de R relevante para las siguientes operaciones se encuentra adjunto en Aula Global.

Suponiendo $Y|\theta \sim \text{Exp}(\theta)$ y que $\theta \sim \text{Gamma}(\alpha, \beta)$, dado que trabajamos con distribuciones de probabilidad continuas, la distribución marginal o predictiva de Y se calcula como:

$$f(y) = \int f(y|\theta)f(\theta) d\theta$$

Por tanto, $Y \sim \text{Pareto}(\alpha, \beta)$, demostrándolo a continuación:

$$\begin{aligned} f(y) &= \int \theta e^{-\theta y} \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta} d\theta \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \int \theta^{(\alpha+1)-1} e^{-\theta(\beta+y)} d\theta \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \int \theta^{\alpha^*-1} e^{-\theta\beta^*} d\theta \quad (\alpha^*, \beta^*) = (\alpha + 1, \beta + y) \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha^*)}{\beta^{*\alpha^*}} \int \frac{\beta^{*\alpha^*}}{\Gamma(\alpha^*)} \theta^{\alpha^*-1} e^{-\theta\beta^*} d\theta \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha + 1)}{(\beta + y)^{\alpha+1}} \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\alpha\Gamma(\alpha)}{(\beta + y)^{\alpha+1}} = \frac{\alpha\beta^\alpha}{(\beta + y)^{\alpha+1}} \end{aligned}$$

Dada una muestra $\{y_i\}_{i=1}^n$ el Estimador Máximo Verosímil (EMV) de θ se obtiene:

$$\hat{\theta} = \arg \max_{\theta} \prod_{i=1}^n \theta \exp \{-\theta y_i\}$$

Las Condiciones de Primer Orden (CPO) dan como resultado $\hat{\theta} = \frac{1}{\bar{y}}$

Para hallar la distribución a posteriori de θ aplicamos el Teorema de Bayes, que guarda la siguiente relación de proporcionalidad entre la verosimilitud y nuestro a priori:

$$\begin{aligned} f(\theta|\text{datos}) &\propto l(\theta|\text{datos}) * f(\theta) \\ &\propto \prod_{i=1}^n \theta e^{-\theta y_i} \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta} \\ &\propto \theta^{\alpha+n-1} e^{-\theta(\beta+n\bar{y})} \\ &= \frac{(\beta + n\bar{y})^{\alpha+n}}{\Gamma(\alpha + n)} \theta^{\alpha+n-1} e^{-\theta(\beta+n\bar{y})} \end{aligned}$$

Es posible pasar de la proporcionalidad a una razón de igualdad ya que introducimos una constante de integración mediante la cual convertimos toda la expresión en una distribución de probabilidad conocida. Además, se tiene que la expresión para la verosimilitud se puede reescribir como $\theta^n e^{-\theta n\bar{y}}$.

Consecuentemente, $\theta|\text{datos} \sim \text{Gamma}(\alpha + n, \beta + n\bar{y})$ y aplicando las propiedades de la Distribución Gamma:

$$E[\theta|\text{datos}] = \frac{\alpha + n}{\beta + n\bar{y}}$$

Es fácil demostrar que conforme $(\alpha, \beta) \rightarrow 0$ la distribución a posteriori se acerca a $\theta|\text{datos} \sim \text{Gamma}(n, n\bar{y})$, y procediendo de manera similar al anterior cálculo:

$$E[\theta|\text{datos}] = \frac{n}{n\bar{y}} = \frac{1}{\bar{y}}$$

Lo cual revela que la media a posteriori tiende al EMV. En la práctica supone adoptar una perspectiva frecuentista, pues eliminamos la subjetividad propia de la Estadística Bayesiana. Los datos dominan por completo al a priori no informativo.

A partir de las conclusiones analíticas que hemos mostrado en las páginas anteriores y sirviéndonos del software estadístico R, es posible graficar las distribuciones a priori, a posteriori y la verosimilitud escalada. Aprovechamos para recordar que el código utilizado para la realización de este proyecto se encuentra adjunto en Aula Global.

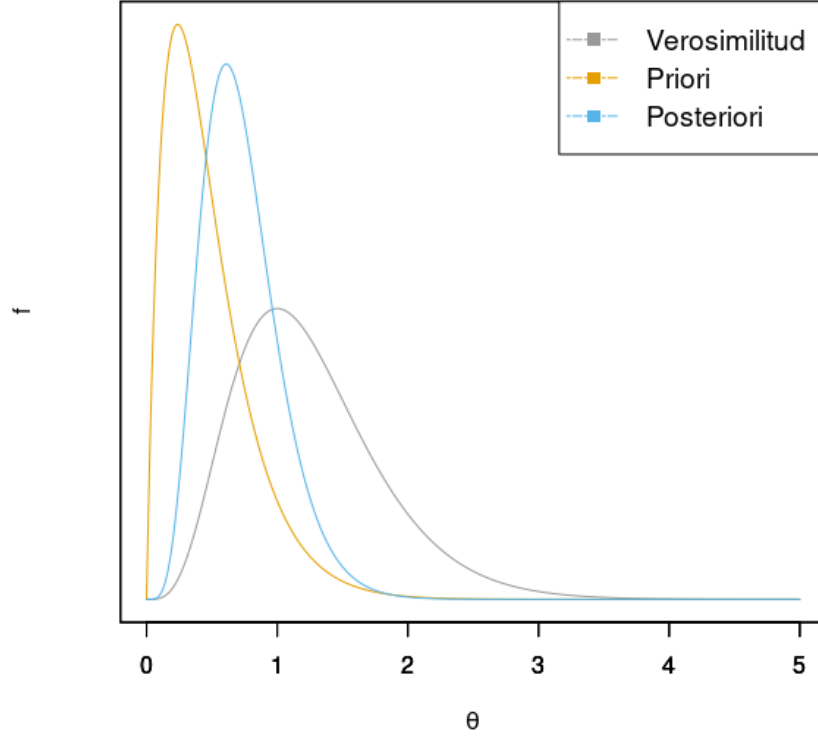


Figure 1: Distribuciones a priori, a posteriori y Verosimilitud Escalada

Asimismo, al objeto de llevar a cabo predicciones para la variable Y , se puede aprovechar que ya conocemos su distribución marginal para incorporar a la misma los parámetros de la función a posteriori obtenida previamente. De esta manera, trabajamos con una distribución predictiva a posteriori (distribución marginal a posteriori) de la variable de interés. Por tanto:

$$f(\hat{y}) = \frac{(\alpha + n)(\beta + n\bar{y})^{\alpha+n}}{(\beta + n\bar{y})^{\alpha+n+1}}$$

A continuación mostramos los datos representados en un histograma, sobreponiendo sobre éste la curva de la distribución predictiva a posteriori presentada en la página anterior. Se puede observar que el ajuste es satisfactorio, por lo que cabe esperar que la predicción sea razonable.

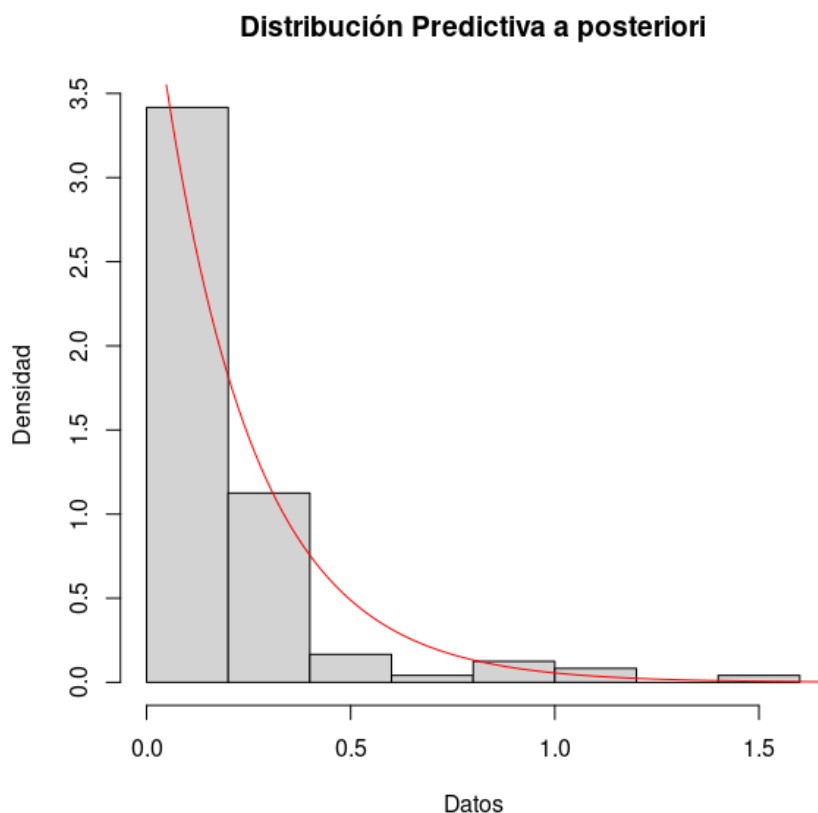


Figure 2: Histograma de datos con predictiva a posteriori

Por último, dado que los Métodos Bayesianos permiten incorporar información subjetiva (normalmente referida a opinión de expertos), puede darse el caso de que el investigador se encuentre inseguro entre dos opiniones o más. Gracias a la mixtura de distribuciones a priori, se permite combinar diferentes puntos de vista asignando a cada uno ponderaciones:

$$f(\theta) = \sum_{i=1}^n w_i f(\theta | \alpha_i, \beta_i)$$

Para el caso de una mixtura tal que $\theta \sim w^* \text{Gamma}(\alpha_1, \beta_1) + (1 - w) \text{Gamma}(\alpha_2, \beta_2)$:

$$f(\theta|\text{datos}) \propto \theta^n e^{-\theta n \bar{y}} [w \text{Gamma}(\alpha_1, \beta_1) + (1 - w) \text{Gamma}(\alpha_2, \beta_2)]$$

$$\propto \theta^n e^{-\theta n \bar{y}} \left[w \frac{\beta_1^{\alpha_1}}{\Gamma(\alpha_1)} \theta^{\alpha_1-1} e^{-\beta_1 \theta} + (1 - w) \frac{\beta_2^{\alpha_2}}{\Gamma(\alpha_2)} \theta^{\alpha_2-1} e^{-\beta_2 \theta} \right]$$

$$\propto w \frac{\beta_1^{\alpha_1}}{\Gamma(\alpha_1)} \theta^{\alpha_1+n-1} e^{-\theta(\beta_1+n\bar{y})} + (1 - w) \frac{\beta_2^{\alpha_2}}{\Gamma(\alpha_2)} \theta^{\alpha_2+n-1} e^{-\theta(\beta_2+n\bar{y})}$$

$$\propto w \frac{\beta_1^{\alpha_1}}{\Gamma(\alpha_1)} \frac{\Gamma(\alpha_1 + n)}{(\beta_1 + n\bar{y})^{\alpha_1+n}} \left[\frac{(\beta_1 + n\bar{y})^{\alpha_1+n}}{\Gamma(\alpha_1 + n)} \theta^{\alpha_1+n-1} e^{-\theta(\beta_1+n\bar{y})} \right]$$

$$+ (1 - w) \frac{\beta_2^{\alpha_2}}{\Gamma(\alpha_2)} \frac{\Gamma(\alpha_2 + n)}{(\beta_2 + n\bar{y})^{\alpha_2+n}} \left[\frac{(\beta_2 + n\bar{y})^{\alpha_2+n}}{\Gamma(\alpha_2 + n)} \theta^{\alpha_2+n-1} e^{-\theta(\beta_2+n\bar{y})} \right]$$

$$= w * \frac{\beta_1^{\alpha_1}}{\Gamma(\alpha_1)} \frac{\Gamma(\alpha_1 + n)}{(\beta_1 + n\bar{y})^{\alpha_1+n}} + (1 - w) * \frac{\beta_2^{\alpha_2}}{\Gamma(\alpha_2)} \frac{\Gamma(\alpha_2 + n)}{(\beta_2 + n\bar{y})^{\alpha_2+n}}$$

$$w^* = \frac{w_i \frac{\beta_i^{\alpha_i}}{\Gamma(\alpha_i)} \frac{\Gamma(\alpha_i + n)}{(\beta_i + n\bar{y})^{\alpha_i+n}}}{\sum_{i=1}^n w_i \frac{\beta_i^{\alpha_i}}{\Gamma(\alpha_i)} \frac{\Gamma(\alpha_i + n)}{(\beta_i + n\bar{y})^{\alpha_i+n}}}$$

Aplicando los algoritmos necesarios en R para programar las soluciones analíticas propuestas, encontramos que para los datos y elección de θ de nuestro caso, la verosimilitud domina a los a priori del investigador. En buena parte esto se debe al tamaño muestral de los 120 datos exponenciales que han servido de base para el proyecto. Asimismo, se aprecia que los segundos componentes del vector de parámetros, a saber, (α_2, β_2) presentan una mayor ponderación final, al acercarse al verdadero valor de θ en mayor medida que el otro a priori.

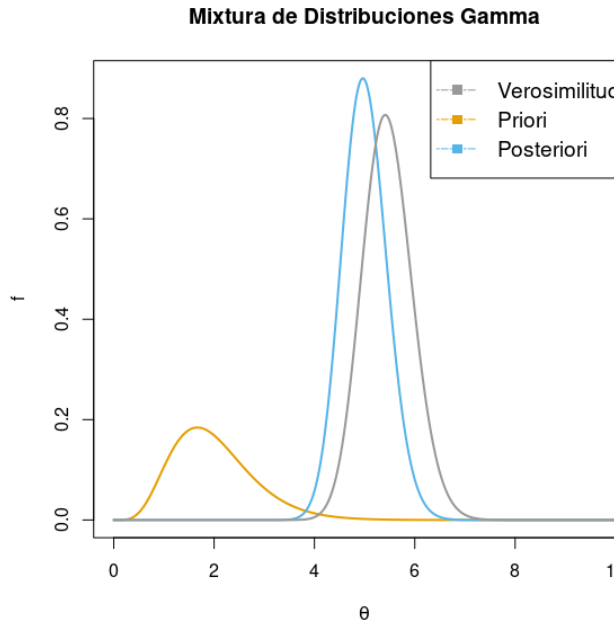


Figure 3: Distribuciones con mixtura

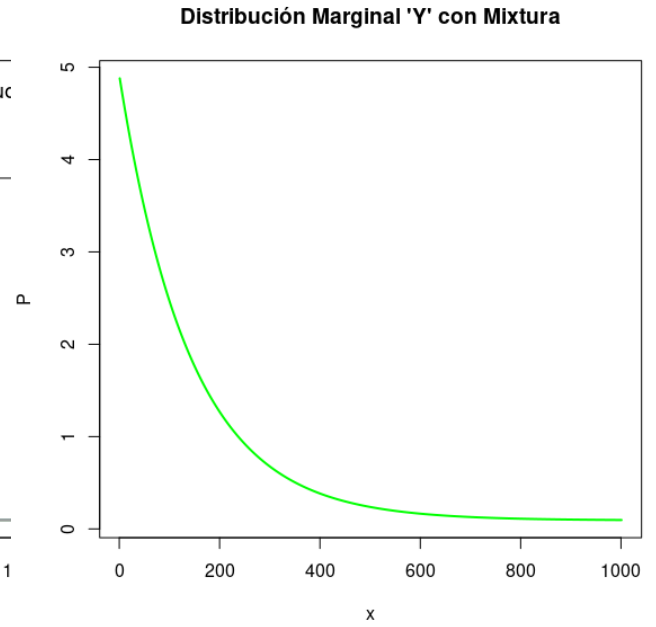


Figure 4: Marginal 'Y' con mixtura

En cuanto a la distribución marginal de Y , es posible llevar a cabo una predicción a posteriori implementando la lógica de la mixtura previamente explicada. Modificamos los parámetros de la distribución de Pareto de acuerdo con las ponderaciones de las distribuciones Gamma. Observamos que el ajuste es razonable, tal y como se muestra en el gráfico de la derecha.