CAPÍTULO 3: DISTRIBUCIONES CONJUGADAS

Para leer

Lee, Capítulo 3, Secciones 3.1,3.2,3.4 y 3.5.

Gelman et al, Capítulo 2, Secciones 2.4 – 2.7.

Ejemplo 18 *Supongamos que en la situación del Ejemplo 13, se usa una distribución a priori de clase Beta, por ejemplo* $\mathcal{B}(\alpha, \beta)$*. Entonces la distribución a posteriori será*

$$f(\theta|\mathbf{x}) \propto \theta^9(1-\theta)^3\theta^{\alpha-1}(1-\theta)^{\beta-1}$$

haciendo caso omiso de las constantes.

Entonces, $f(\theta|\mathbf{x}) \propto \theta^{\alpha+9-1}(1-\theta)^{\beta+3-1}$ *y es fácil ver que la distribución a posteriori es* $\mathcal{B}(\alpha+9, \beta+3)$*. Se dice que la distribución beta es* **conjugada** *con la distribución muestral binomial.*

Definición 5 *Si $\mathcal{F}$ es una clase de distribuciones muestrales $f(x|\boldsymbol{\theta})$ y $\mathcal{P}$ es una clase de distribuciones a priori $p(\boldsymbol{\theta})$ para $\boldsymbol{\theta}$, luego $\mathcal{P}$ es* **conjugada con** *$\mathcal{F}$ si*

$$p(\boldsymbol{\theta}|x) \in \mathcal{P} \; \forall \; f(\cdot|\boldsymbol{\theta}) \in \mathcal{F} \; y \; p(\cdot) \in \mathcal{P}.$$

Ejemplo 19 *Sea $X|\theta \sim \mathcal{P}(\theta)$, con distribución a priori gamma $\theta \sim \mathcal{G}(\alpha, \beta)$.*

Dados los datos $\mathbf{x}$*, la verosimilitud será*

$$\begin{aligned} l(\theta|\mathbf{x}) &= \prod_{i=1}^{n} \frac{\theta^{x_i} e^{-\theta}}{x_i!} \\ &\propto \theta^{\sum_i x_i} e^{-n\theta} \end{aligned}$$

Luego la distribución a posteriori es

$$\begin{aligned} f(\theta|\mathbf{x}) &\propto \frac{\beta}{\Gamma(\alpha)} (\beta\theta)^{\alpha-1} e^{-\beta\theta} \theta^{\sum_i x_i} e^{-n\theta} \\ &\propto \theta^{\alpha+\sum_i x_i - 1} e^{-(\beta+n)\theta} \end{aligned}$$

Entonces $\theta|\mathbf{x} \sim \mathcal{G}(\alpha + n\bar{x}, \beta + n)$. La distribución gamma es conjugada con la distribución muestral Poisson.

Ejemplo 20 *Sea $X|\theta \sim \mathcal{E}(\theta)$. Supongamos una densidad a priori gamma: $\theta \sim \mathcal{G}(\alpha, \beta)$. Entonces, dados los datos $\mathbf{x} = (x_1, \ldots, x_n)$, tenemos*

$$\begin{aligned} f(\theta|bx) &\propto \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta} \times \\ &\quad \times \prod_{i=1}^{n} \theta e^{-\theta x_i} \\ &\propto \theta^{\alpha+n-1}e^{-\theta(\beta+\sum_{i=1}^{n} x_i)} \\ &\propto \theta^{\alpha+n-1}e^{-\theta(\beta+n\bar{x})} \end{aligned}$$

que es el núcleo de una distribución gamma:

$$\theta|\mathbf{x} \sim \mathcal{G}(\alpha + n, \beta + n\bar{x})$$

La distribución gamma tambien es conjugada con la distribución muestral exponencial.

Ejemplo 21 *Supongamos que tenemos datos multinomiales*

$$P(\mathbf{X} = \mathbf{x}|\boldsymbol{\theta}) = \frac{n!}{x_1! \ldots x_k!} \theta_1^{x_1} \cdots \theta_k^{x_k}$$

donde $\mathbf{X} = (X_1, \ldots, X_k)^T$ *y* $\sum_{i=1}^k X_i = n$, *y* $0 < \theta_i < 1$, $\sum_{i=1}^k \theta_i = 1$.

Un ejemplo típico es el resultado de n *tiradas de un dado, cuando* $k = 6$.

Elegimos una distribución a priori Dirichlet

$$f(\boldsymbol{\theta}) \propto \frac{\Gamma(\alpha_1 + \ldots + \alpha_k)}{\Gamma(\alpha_1) \cdots \Gamma(\alpha_k)} \theta_1^{\alpha_1 - 1} \cdots \theta_k^{\alpha_k - 1}$$

Observación 12 *La distribución Dirichlet es la versión multivariante de la distribución beta. Si* $k = 2$, *se tiene la distribución beta introducida anteriormente. Recordamos también que la distribución multinomial es la versión multivariante de la binomial.*

Entonces, la distribución a posteriori dados unos datos $\mathbf{x}$ *es*

$$\begin{aligned} f(\boldsymbol{\theta}|\mathbf{x}) &\propto f(\boldsymbol{\theta}) l(\boldsymbol{\theta}|\mathbf{x}) \\ &\propto \left(\prod_{i=1}^{k} \theta_i^{x_i} \right) \left(\prod_{i=1}^{k} \theta_i^{\alpha_i - 1} \right) \\ &\propto \prod_{i=1}^{k} \theta_i^{x_i + \alpha_i - 1} \end{aligned}$$

es decir otra distribución Dirichlet con parámetro $(\alpha_1 + x_1, \ldots, \alpha_k + x_k)$.

Entonces, la distribución Dirichlet es conjugada con la distribución muestral multinomial.

Ventajas de usar distribuciones conjugadas.

- El proceso de aprendizaje es sencillamente el proceso de cambiar los valores de los parámetros de la distribución a priori.

- Se pueden interpretar los valores de los parámetros de la distribución inicial.

  Ejemplo 22 *Volviendo al Ejemplo 18 veamos que dada una distribución a priori* $\mathcal{B}(\alpha, \beta)$*, la distribución a posteriori es*

  $$\mathcal{B}(\alpha + \#\ \textit{cruces visto}, \beta + \#\ \textit{caras visto})$$

  Podemos interpretar los parámetros a priori $\alpha + \beta$ *y* α *como el número equivalente de tiradas de la moneda y el número de cruces en estas tiradas que habríamos tenido que ver para darnos el nivel de conocimiento representado por la distribución a priori.*

Ejemplo 23 *En el Ejemplo 19, dada una distribución a priori $\mathcal{G}(\alpha, \beta)$, la distribución a posteriori es $\mathcal{G}(\alpha + n\bar{x}, \beta + n)$. La información contenida en la distribución a priori equivale a la información en una muestra de tamaño β con media muestral α/β.*

- En muchos casos, se puede relacionar la media a posteriori con la media a priori y la EMV.

  Ejemplo 24 *Volvemos al Ejemplo 18. La media a posteriori es*

$$\begin{aligned} \frac{\alpha + 9}{\alpha + \beta + 12} &= \frac{(\alpha + \beta)\frac{\alpha}{\alpha+\beta} + 12\frac{9}{12}}{\alpha + \beta + 12} \\ &= wE[\theta] + (1 - w)\widehat{\theta} \end{aligned}$$

  donde $0 < w = \frac{\alpha+\beta}{\alpha+\beta+n} < 1$ *y* $\widehat{\theta} = 9/12$ *es el EMV de* θ*.*

Ejemplo 25 *En el Ejemplo 19, se tiene*

$$\begin{aligned} E[\theta|\mathbf{x}] &= \frac{\alpha + n\bar{x}}{\beta + n} \\ &= w\frac{\alpha}{\beta} + (1-w)\bar{x} \end{aligned}$$

donde $0 \leq w = \frac{\beta}{\beta+n} \leq 1$.

Volviendo al Ejemplo 23, vemos que la media a posteriori es una media ponderada con pesos proporcionales al número de observaciones equivalentes en la distribución a priori y al tamaño de la muestra.

Ejemplo 26 *Retomando al Ejemplo 20, se tiene:*

$$\begin{aligned} E[\theta|\mathbf{x}] &= \frac{\alpha + n}{\beta + n\bar{x}} \\ &= w\frac{\alpha}{\beta} + (1-w)\frac{1}{\bar{x}} \end{aligned}$$

donde

$$0 \leq w = \frac{\beta}{\beta + n\bar{x}} \leq 1$$

y $\frac{\alpha}{\beta}$ *es la media a priori y* $\frac{1}{\bar{x}}$ *es el EMV de* θ.

- La familia de mixturas de distribuciones conjugadas es también conjugada. Si se define la distribución inicial como

$$p(\boldsymbol{\theta}) = \sum_{i=1}^{k} w_i p_i(\boldsymbol{\theta})$$

donde $p_i(\cdot) \in \mathcal{P}$ son conjugadas con una distribución muestral $f(x|\boldsymbol{\theta})$ en el sentido de Definición 5, luego $p(\cdot) \in \mathcal{P}$.

Es posible aproximar cualquier distribución $f(\cdot)$ con una mixtura suficientemente grande de densidades conjugadas.

Ejemplo 27 *Volvemos a la situación del Ejemplo 13. Supongamos ahora una distribución a priori que es una mixtura de tres distribuciones beta:*

$$f(\theta) = 0{,}25\mathcal{B}(2,1) + 0{,}5\mathcal{B}(5,5) + 0{,}25\mathcal{B}(1,2)$$

Ahora, la distribución a posteriori es

$$\begin{aligned}
f(\theta|\mathbf{x}) &\propto l(\theta|\mathbf{x})f(\theta) \\
&\propto \theta^9(1-\theta)^3\left\{0{,}25\frac{1}{B(2,1)}\theta^{2-1}\theta^{1-1}+\right. \\
&\quad \left. 0{,}5\frac{1}{B(5,5)}\theta^{5-1}\theta^{5-1} + 0{,}25\frac{1}{B(1,2)}\theta^{1-1}\theta^{2-1}\right\} \\
&\propto \left\{0{,}25\frac{1}{B(2,1)}\theta^{11-1}\theta^{4-1} + 0{,}5\frac{1}{B(5,5)}\theta^{14-1}\theta^{8-1}+\right. \\
&\quad \left. 0{,}25\frac{1}{B(1,2)}\theta^{10-1}\theta^{5-1}\right\} \\
&\propto \left\{0{,}25\frac{B(11,4)}{B(2,1)}\frac{1}{B(11,4)}\theta^{11-1}\theta^{4-1}+\right. \\
&\quad 0{,}5\frac{B(14,8)}{B(5,5)}\frac{1}{B(14,8)}\theta^{14-1}\theta^{8-1} + \\
&\quad \left. 0{,}25\frac{B(10,5)}{B(1,2)}\frac{1}{B(10,5)}\theta^{10-1}\theta^{5-1}\right\} \\
&= w_1\mathcal{B}(11,4) + w_2\mathcal{B}(14,8) + (1-w_1-w_2)\mathcal{B}(10,5)
\end{aligned}$$

es decir otra mixtura de tres distribuciones beta donde

$$w_1 = \frac{0{,}25\frac{B(11,4)}{B(2,1)}}{0{,}25\frac{B(11,4)}{B(2,1)} + 0{,}5\frac{B(14,8)}{B(5,5)} + 0{,}25\frac{B(10,5)}{B(1,2)}}$$

$$w_2 = \frac{0{,}5\frac{B(14,8)}{B(5,5)}}{0{,}25\frac{B(11,4)}{B(2,1)} + 0{,}5\frac{B(14,8)}{B(5,5)} + 0{,}25\frac{B(10,5)}{B(1,2)}}.$$

La distribución conjugada no siempre es fácil de usar

Ejemplo 28 *Supongamos que* $X|\theta \sim \mathcal{B}(\alpha, \theta)$ *donde* α *es conocido. Entonces*

$$f(x|\theta) \propto \frac{\Gamma(\alpha+\theta)}{\Gamma(\theta)} x^{\theta}$$

y una distribución a priori conjugada será

$$f(\theta) \propto \left(\frac{\Gamma(\alpha+\theta)}{\Gamma(\theta)}\right)^{a} b^{\theta}$$

cuando

$$f(\theta|\mathbf{x}) \propto \left(\frac{\Gamma(\alpha+\theta)}{\Gamma(\theta)}\right)^{a+n} \left(b \prod_{i=1}^{n} x_i\right)^{\theta}$$

y no se puede hallar ni la constante de integración ni la media sin emplear la integración numérica.

Familias Exponenciales

Hay una gran relación entre familias conjugadas y el concepto clásico de la familia exponencial.

Definición 6 *La familia de distribuciones* $f(x|\cdot)$ *con densidades de forma*

$$f(x|\boldsymbol{\theta}) = C(\boldsymbol{\theta})h(x)\exp\left(\mathbf{R}(\boldsymbol{\theta})^T\mathbf{T}(x)\right)$$

donde $C(\cdot)$ *y* $h(\cdot)$ *son funciones y* $\mathbf{R}(\cdot)$ *y* $\mathbf{T}(\cdot)$ *son funciones vectoriales de dimensión* $k = dim(\boldsymbol{\theta})$ *se llama* **una familia exponencial**.

$\phi = \mathbf{R}(\theta)$ *se llama* **el parámetro natural de la familia.**

Si el soporte de X *es independiente de* $\boldsymbol{\theta}$ *se dice que la familia es una familia exponencial regular. Si el soporte depende de* $\boldsymbol{\theta}$*, la familia es irregular.*

Ejemplo 29 *Distribución Poisson*

$$\begin{aligned} P(X = x|\theta) &= \frac{\theta^x e^{-\theta}}{x!} \\ &= e^{-\theta}\frac{1}{x!}\exp\left(x\log(\theta)\right) \end{aligned}$$

El parámetro natural es $\phi = \log(\theta)$.

Ejemplo 30 *Sea* $X|\boldsymbol{\theta} \sim \mathcal{BI}(n, \theta)$ *con* θ *conocido. Luego:*

$$\begin{aligned} P(X = x|\theta) &= \left(\begin{array}{c} n \\ x \end{array}\right)\theta^x(1-\theta)^{n-x} \\ &= \left(\begin{array}{c} n \\ x \end{array}\right)(1-\theta)^n\left(\frac{\theta}{1-\theta}\right)^x \\ &= (1-\theta)^n\left(\begin{array}{c} n \\ x \end{array}\right)\exp\left\{x\log\frac{\theta}{1-\theta}\right\} \end{aligned}$$

y la distribución binomial es una familia exponencial con parámetro natural $\phi = \frac{\theta}{1-\theta}$.

Ejemplo 31 $X|\mu,\sigma^2 \sim \mathcal{N}(\mu,\sigma^2)$. *Supongamos que ambos parámetros son desconocidos.*

$$\begin{aligned} f(x|\mu,\sigma^2) &= \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) \\ &= \left(\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{\mu^2}{2\sigma^2}}\right)\exp\left(\mathbf{R}(\mu,\sigma^2)^T\mathbf{T}(x)\right) \end{aligned}$$

donde

$$\mathbf{R}(\mu,\sigma^2) = \left(-\frac{1}{2\sigma^2},\frac{\mu}{\sigma^2}\right)^T$$

es el parámetro natural y

$$\mathbf{T}(x) = (x^2,x)^T$$

Familias no exponenciales

Aunque la mayoría de las distribuciones comunes son familias exponenciales, existen algunas excepciones.

Ejemplo 32 *Supongamos que* $X|\theta \sim \mathcal{U}(0,\theta)$. *Luego la distribución de* X *es una familia exponencial irregular porque el soporte depende de* θ.

Ejemplo 33 *La distribución de Cauchy*

$$f(x|\theta) \propto \frac{1}{1+(x-\theta)^2}$$

no es una la familia exponencial.

Ejemplo 34 *La distribución F de Fisher*

$$f(x|\alpha,\beta) = \frac{\Gamma((\alpha+\beta)/2)}{\Gamma(\alpha/2)\Gamma(\beta/2)}\alpha^{\alpha/2}\beta^{\beta/2}\frac{x^{\alpha/2-1}}{(\beta+\alpha x)^{(\alpha+\beta)/2}}$$

no es una familia exponencial.

Estadísticos suficientes

Pongamos que $X \sim f(\cdot|\boldsymbol{\theta})$ pertenece a una familia exponencial. Entonces, dados los datos $\mathbf{x}$, la verosimilitud será

$$\begin{aligned} l(\boldsymbol{\theta}|\mathbf{x}) &\propto \prod_{i=1}^{n} C(\boldsymbol{\theta}) h(x_i) \exp\left(\mathbf{R}(\boldsymbol{\theta})^T \mathbf{T}(x_i)\right) \\ &\propto C(\boldsymbol{\theta})^n \exp\left(\mathbf{R}(\boldsymbol{\theta})^T \sum_{i=1}^{n} \mathbf{T}(x_i)\right) (*) \end{aligned}$$

Luego, la verosimilitud depende de (n, el tamaño de la muestra y) $\sum_{i=1}^{n} \mathbf{T}(x_i)$. que es un estadístico suficiente para $\boldsymbol{\theta}$.

Definición 7 *Dados los datos* $\mathbf{x}$, *un estadístico* $\mathbf{S}(\mathbf{x})$ *se llama* **suficiente para** $\boldsymbol{\theta}$ *si*

$$l(\boldsymbol{\theta}|\mathbf{x}) = l(\boldsymbol{\theta}|\mathbf{S}(\mathbf{x}))$$

y (si depende de los datos) se puede expresar el soporte de $\boldsymbol{\theta}$ *como una función de* $\mathbf{S}(\mathbf{x})$.

Ejemplo 35 $X|\theta \sim \mathcal{P}(\theta)$.

Volviendo al Ejemplo 29, $T(x) = x$ *y dados los datos* $\mathbf{x}$*, un estadístico suficiente para* $\boldsymbol{\theta}$ *será* $\sum_{i=1}^{n} x_i$.

La distribución no tiene que ser una familia exponencial regular para que exista un estadístico suficiente. El contraejemplo es la distribución uniforme.

Ejemplo 36 $X|\theta \sim \mathcal{U}(0, \theta)$.

Entonces $l(\mathbf{x}|\theta) \propto \frac{1}{\theta^n}$ *por* $\theta \geq \text{máx}\, x_i$.

máx x_i *es un estadístico suficiente para* θ.

Una familia conjugada con una familia exponencial

Si $f(x|\boldsymbol{\theta})$ es de una familia exponencial, con ecuación $(*)$, está claro que existe una familia conjugada. La distribución

$$f(\boldsymbol{\theta}) \propto C(\boldsymbol{\theta})^{\alpha} \exp\left(\mathbf{R}(\boldsymbol{\theta})^T \boldsymbol{\beta}\right)$$

por constantes α y $\boldsymbol{\beta}$, es conjugada, porque dados los datos $\mathbf{x}$, la distribución a posteriori es

$$f(\boldsymbol{\theta}|\mathbf{x}) \propto C(\boldsymbol{\theta})^{\alpha^*} \exp\left(\mathbf{R}(\boldsymbol{\theta})^T \boldsymbol{\beta}^*\right)$$

donde $\alpha^* = \alpha + n$ y $\boldsymbol{\beta}^* = \boldsymbol{\beta} + \sum_{i=1}^{n} \mathbf{T}(x_i)$.

Ejemplo 37 *Retomemos el Ejemplo 30. Se tiene*

$$P(x|\theta) = (1-\theta)^n \begin{pmatrix} n \\ x \end{pmatrix} \exp\left\{x \log \frac{\theta}{1-\theta}\right\}$$

y entonces, una distribución a priori conjugada sería

$$\begin{aligned} f(\theta) &\propto ((1-\theta)^n)^\alpha \exp\left\{\beta \log \frac{\theta}{1-\theta}\right\} \\ &\propto \theta^\beta (1-\theta)^{n\alpha-\beta} \\ &\propto \theta^{\beta+1-1}(1-\theta)^{n\alpha-\beta+1-1} \\ &\sim \mathcal{B}(\beta+1, n\alpha-\beta+1) \end{aligned}$$

y como la elección de α y β es arbitraria, se sabe que cualquier distribución beta es conjugada a la distribución muestral binomial, como se ha visto anteriormente en el Ejemplo 18

Ejemplo 38 $X|\mu \sim \mathcal{N}(\mu, \sigma^2)$ *(σ^2 conocido).*

$$\begin{aligned} f(x|\mu) &\propto \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) \\ &\propto \exp\left(-\frac{1}{2\sigma^2}\mu^2\right)\exp\left(\frac{\mu}{\sigma^2}x\right) \end{aligned}$$

Entonces, una distribución a priori conjugada será de forma

$$\begin{aligned} f(\mu) &\propto \left[\exp\left(-\frac{1}{2\sigma^2}\mu^2\right)\right]^{\alpha}\exp\left(\frac{\mu}{\sigma^2}\beta\right) \\ &\propto \exp\left(-\frac{\alpha}{2\sigma^2}\left(\mu^2 - 2\frac{\beta}{\alpha}\right)\right) \end{aligned}$$

y completando el cuadrado, se puede ver que la distribución a priori conjugada será normal

$$\mu \sim \mathcal{N}(m, \sigma^2/\alpha)$$

donde $m = \beta/\alpha$.

La distribución a posteriori

Dados los datos $\mathbf{x}$*, la verosimilitud es*

$$l(\mu|\mathbf{x}) \propto (\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right)$$

y entonces, la distribución final será

$$\begin{aligned} f(\mu|\mathbf{x}) &\propto \exp\left(-\frac{1}{2\sigma^2}\left[\alpha(\mu-m)^2+\right.\right. \\ &\qquad \left.\left.+\sum_{i=1}^{n}(\mu-x_i)^2\right]\right) \\ &\propto \exp\left(-\frac{1}{2\sigma^2}\left[(\alpha+n)\mu^2-\right.\right. \\ &\qquad \left.\left.-2\mu(\alpha m+\sum_{i=1}^{n}x_i)\right]\right) \\ &\propto \exp\left(-\frac{\alpha+n}{2\sigma^2}\left(\mu-\frac{\alpha m+n\bar{x}}{\alpha+n}\right)^2\right) \\ \mu|\mathbf{x} &\sim \mathcal{N}\left(\frac{\alpha m+n\bar{x}}{\alpha+n}, \frac{\sigma^2}{\alpha+n}\right) \end{aligned}$$

Observamos que

- *La media a posteriori es*

$$E[\mu|\mathbf{x}] = wE[\mu] + (1-w)\hat{\mu}$$

donde $\hat{\mu} = \bar{x}$ *es el EMV de* μ *y* $w = \frac{\alpha}{\alpha+n}$.

- *Un* **intervalo de credibilidad** *de 95 % para* μ *es*

$$\frac{\alpha m + n\bar{x}}{\alpha + n} \pm 1{,}96\frac{\sigma}{\sqrt{\alpha+n}}$$

- *Si y sólo si* $\alpha = 0$, *el intervalo será igual al intervalo clasíco de confianza.*

$$\bar{x} \pm 1{,}96\frac{\sigma}{\sqrt{n}}$$

En esta situación la distribución a priori sería **impropia**

$$\mu \sim \mathcal{N}(m, \infty) \Rightarrow f(\mu) \propto c$$

- *De vez en cuando, se escribe la distribución a posteriori en la forma* $\theta \sim \mathcal{N}(m, \tau^2)$ *con varianza* $\tau^2 = \sigma^2/\alpha$. *Entonces, sustituyendo se tiene*

$$\mu|\mathbf{x} \sim \mathcal{N}\left(\frac{\frac{1}{\tau^2}m + \frac{n}{\sigma^2}\bar{x}}{\frac{1}{\tau^2} + \frac{n}{\sigma^2}}, \left(\frac{1}{\tau^2} + \frac{n}{\sigma^2}\right)^{-1}\right)$$

En esta expresión, se ve que la media a posteriori es una media ponderada de la media a priori y la media muestral con pesos proporcionales a las precisiones.

La precisión a posteriori es la suma de la precisión a priori y la precisión del EMV.

Distribuciones conjugadas y familias exponenciales irregulares

La distribución muestral no tiene que pertenecer a una familia exponencial regular para que exista una distribución conjugada a priori.

Ejemplo 39 *Consideramos la distribución uniforme ($X \sim \mathcal{U}(0, \theta)$).*

Dados los datos $\mathbf{x}$*, la verosimilitud es*

$$l(\theta|\mathbf{x}) = \theta^{-n} \quad \text{para } \theta > \text{máx}\{x_1, \ldots, x_n\}.$$

Supongamos una distribución a priori Pareto; $\theta \sim \mathcal{PA}(\alpha, \beta)$*. Luego*

$$f(\theta) = \alpha\beta^{\alpha}\theta^{-\alpha-1}$$

para $\theta > \beta$*.*

Entonces la distribución a posteriori es

$$f(\theta|\mathbf{x}) \propto \theta^{-\alpha-n-1}$$

por $\theta > \beta^* = \text{máx}\{\beta, x_1, \ldots, x_n\}$*.*

Luego $\theta|\mathbf{x} \sim \mathcal{PA}(\alpha + n, \beta^*)$*.*