

## CAPÍTULO 2: DISTRIBUCIONES CONJUGADAS

### Para leer

Gelman et al (1995), Capítulo 2.

Lee (1997), Capítulo 3, Secciones 3.1–3.5.

Bernardo y Smith (1994), Sección 5.2.

**Ejemplo 11** *Retomando el Ejemplo 6, sea la distribución a priori una Beta,  $\mathcal{B}(\alpha, \beta)$ . Entonces la distribución a posteriori será*

$$\begin{aligned} f(\theta|x) &\propto f(\theta)l(\theta|\mathbf{x}) \\ &\propto \frac{1}{B(\alpha, \beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1} \times \\ &\quad \binom{12}{9} \theta^9 (1-\theta)^3 \\ &\propto \theta^{\alpha+9-1} (1-\theta)^{\beta+3-1} \\ \theta|x &\sim \mathcal{B}(\alpha+9, \beta+3). \end{aligned}$$

*La distribución a posteriori es de la misma familia como la distribución a priori.*

*Generalizando, dada la distribución a priori  $\mathcal{B}(\alpha, \beta)$ , si se tira la moneda  $n$  veces observando  $x$  cruces y  $n - x$  caras, entonces la distribución a posteriori es  $\mathcal{B}(\alpha + x, \beta + n - x)$ .*

*Entonces para cualquier distribución de muestreo de tipo Bernoulli, binomial, geométrica o binomial negativa, si se utiliza una distribución a priori beta, luego la distribución a posteriori es de la misma familia.*

*Se dice que la distribución beta es **conjugada** (Raiffa y Schlaifer 1961) con las distribuciones de muestreo Bernoulli, binomial, etc..*

La idea de la distribución conjugada es que la distribución a posteriori tiene la misma forma algebraica como la distribución a priori.

Para formalizarla, recordamos la Definición 1 de una estadística suficiente  $t$ . Si  $t$  es suficiente para  $\theta$ , entonces

$$\begin{aligned} f(\mathbf{x}|\theta) &= f(t|\theta)f(\mathbf{x}|t) \\ l(\theta|\mathbf{x}) &\propto f(t|\theta). \end{aligned}$$

Luego, esta claro que si se elige una distribución a priori  $f(\theta)$  con la misma estructura de  $f(t|\theta)$ , pensando en esto como función de  $\theta$ , entonces, la distribución a posteriori tendrá la misma forma.

Sigue una definición formal.

**Definición 3** (Bernardo y Smith 1994) La familia conjugada de densidades a priori para  $\theta$  con respecto a la verosimilitud  $l(\theta|\mathbf{x})$  con estadístico suficiente  $\mathbf{t} = \mathbf{t}(\mathbf{x}) = \{n, \mathbf{s}(\mathbf{x})\}$  (de dimensión fija  $k$ , independiente de  $\mathbf{x}$ ) es

$$f(\theta|\tau = (\tau_0, \tau_1, \dots, \tau_k) \in \mathcal{T})$$

donde

$$\mathcal{T} = \left\{ \tau : \int_{\Theta} f(\mathbf{s} = (\tau_1, \dots, \tau_k) | \theta, n = \tau_0) d\theta < \infty \right\}$$

y

$$f(\theta|\tau) = \frac{f(\mathbf{s} = (\tau_1, \dots, \tau_k) | \theta, n = \tau_0)}{\int_{\Theta} f(\mathbf{s} = (\tau_1, \dots, \tau_k) | \theta, n = \tau_0) d\theta}.$$

**Ejemplo 12** Sea  $X|\theta \sim \mathcal{P}(\theta)$ , una densidad Poisson.

Dada una muestra de datos  $\mathbf{x}$ , la verosimilitud es

$$l(\theta|\mathbf{x}) = \prod_{i=1}^n \frac{\theta^{x_i} e^{-\theta}}{x_i!} \propto \theta^{n\bar{x}} e^{-n\theta}$$

donde  $\bar{x}$  es suficiente.

Entonces, una distribución conjugada sería de forma

$$f(\theta) \propto \theta^{\tau_1} e^{-\tau_0 \theta}$$

es decir una distribución gamma, por ejemplo  $\mathcal{G}(\alpha, \beta)$  (donde  $\alpha = \tau_1 + 1$  y  $\beta = \tau_0$ ).

Luego si  $\theta \sim \mathcal{G}(\alpha, \beta)$ , dada una muestra de tamaño  $n$ , se tiene  $\theta|\mathbf{x} \sim \mathcal{G}(\alpha + n\bar{x}, \beta + n)$ .

La distribución gamma es conjugada con la distribución muestral Poisson.

## Familias Exponenciales y Distribuciones Conjugadas

Dada la definición de una distribución conjugada, es obvio que está relacionado con el concepto de familias exponenciales de distribuciones.

**Definición 4** *La familia de distribuciones  $f(x|\cdot)$  con densidades de forma*

$$f(x|\boldsymbol{\theta}) = C(\boldsymbol{\theta})h(x) \exp \left( \sum_{i=1}^k R_i(\boldsymbol{\theta})s_i(x) \right)$$

*se llama **una familia exponencial** (con  $k$  parámetros).*

$\boldsymbol{\phi} = \mathbf{R}(\boldsymbol{\theta}) = (R_1(\boldsymbol{\theta}), \dots, R_k(\boldsymbol{\theta}))$  *se llama **el parámetro natural de la familia.***

*Si el soporte de  $X$  no depende de  $\theta$ , la familia se llama regular. Si el soporte de  $X$  depende de  $\theta$ , la familia es irregular.*

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

**Ejemplo 13** *La distribución uniforme  $X|\theta \sim \mathcal{U}(0, \theta)$  es una familia exponencial irregular.*

**Ejemplo 14** *La distribución de Cauchy*

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

*no pertenece a ninguna familia exponencial.*

## Una familia conjugada con una familia exponencial regular

Si  $f(x|\boldsymbol{\theta})$  es de una familia exponencial, con densidad de la forma de la Definición 4, está claro que existe una familia conjugada.

**Teorema 2** *Sea  $f(x|\boldsymbol{\theta})$  una familia exponencial como en Definición 4. Entonces la distribución a priori  $f(\boldsymbol{\theta}) = g(\boldsymbol{\theta}|\boldsymbol{\tau})$  con*

$$g(\boldsymbol{\theta}|\boldsymbol{\tau}) = K(\boldsymbol{\tau})C(\boldsymbol{\theta})^{\boldsymbol{\tau}_0} \exp\left(\sum_{i=1}^k R_i(\boldsymbol{\theta})\tau_i\right)$$

*es conjugada y dada una muestra de tamaño  $n$ , la distribución a posteriori es*

$$f(\boldsymbol{\theta}|\mathbf{x}) = g(\boldsymbol{\theta}|\boldsymbol{\tau} + \mathbf{t})$$

*donde  $\mathbf{t}(\mathbf{x}) = \left\{n, \sum_{i=1}^n s_1(x_i), \dots, \sum_{i=1}^n s_k(x_i)\right\}$  es el estadístico suficiente.*



Igualmente, si la distribución predictiva a priori de unos futuros datos  $\mathbf{y}$  es

$$\begin{aligned} f(\mathbf{y}) &= \int f(\mathbf{y}|\boldsymbol{\theta})f(\boldsymbol{\theta}) d\boldsymbol{\theta} \\ &= \int f(\mathbf{y}|\boldsymbol{\theta})g(\boldsymbol{\theta}|\boldsymbol{\tau}) d\boldsymbol{\theta} \\ &= h(\mathbf{y}|\boldsymbol{\tau}) \end{aligned}$$

entonces la distribución predictiva a posteriori es

$$f(\mathbf{y}|\mathbf{x}) = h(\mathbf{y}|\boldsymbol{\tau} + \mathbf{t}).$$

## **Demostración**

Dada una muestra  $\mathbf{x}$ , la verosimilitud es

$$l(\boldsymbol{\theta}|\mathbf{x}) \propto C(\boldsymbol{\theta})^n \exp \left( \sum_{j=1}^n \sum_{i=1}^k R_i(\boldsymbol{\theta}) s_i(x_j) \right)$$

$\mathbf{t}(\mathbf{x}) = \left\{ n, \sum_{i=1}^n s_1(x_i), \dots, \sum_{i=1}^n s_k(x_i) \right\}$  es suficiente.

Entonces, dada la distribución a priori

$$\begin{aligned} f(\boldsymbol{\theta}) &= g(\boldsymbol{\theta}|\boldsymbol{\tau}) \\ &\propto C(\boldsymbol{\theta})^{\boldsymbol{\tau}_0} \exp\left(\sum_{i=1}^k R_i(\boldsymbol{\theta})\tau_i\right) \end{aligned}$$

la distribución a posteriori es

$$\begin{aligned} f(\boldsymbol{\theta}|\mathbf{x}) &\propto C(\boldsymbol{\theta})^{\boldsymbol{\tau}_0 + \mathbf{n}} \exp\left(\sum_{i=1}^k R_i(\boldsymbol{\theta})(\tau_i + \sum_{j=1}^n s_i(x_j))\right) \\ &= g(\boldsymbol{\theta}|\boldsymbol{\tau} + \mathbf{t}). \end{aligned}$$

Igualmente

$$\begin{aligned} f(\mathbf{y}|\mathbf{x}) &= \int f(\mathbf{y}|\boldsymbol{\theta})f(\boldsymbol{\theta}|\mathbf{x}) d\boldsymbol{\theta} \\ &= \int f(\mathbf{y}|\boldsymbol{\theta})g(\boldsymbol{\theta}|\boldsymbol{\tau} + \mathbf{t}) d\boldsymbol{\theta} \\ &= h(\boldsymbol{\tau} + \mathbf{t}). \end{aligned}$$

◇

El teorema demuestra que la familia conjugada de distribuciones está cerrada bajo muestreo (Weatherill 1961).

La información en la distribución a priori es equivalente a la información contenida en una muestra con estadístico suficiente  $\tau$ .

Si se quiere ser objetivo, una posibilidad es considerar la distribución a priori derivada cuando  $\tau \rightarrow 0$ ). En este caso, la distribución a priori puede ser *impropia*.

**Ejemplo 15** Sea  $X|\theta \sim \mathcal{E}(\theta)$ . Entonces

$$f(x|\theta) = \theta e^{-x\theta}$$

es una familia exponencial y una distribución a priori conjugada para  $\theta$  es

$$f(\theta) \propto \theta^{\tau_0} \exp(-\beta\tau_1)$$

es decir una distribución gamma  $\mathcal{G}(\alpha, \beta)$  donde  $\alpha = \tau_0 + 1$  y  $\beta = \tau_1$ .

Entonces, la distribución a priori gamma es conjugada con la distribución exponencial y dada una muestra de tamaño  $n$ , la distribución a posteriori es  $\mathcal{G}(\alpha + n, \beta + \sum_{i=1}^n x_i)$ .

Dejando  $\tau_0, \tau_1 \rightarrow 0$ , este implica que la distribución a priori es  $(\mathcal{G}(-1, 0))$   $f(\theta) \propto 1$ , es decir una distribución impropia. En este caso,  $\theta|\mathbf{x} \sim \mathcal{G}(n, n\bar{x})$  y la media a posteriori coincide con el EMV.

**Ejemplo 16** Sea  $X$  un ensayo de Bernoulli con parámetro  $\theta$ . Luego

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

Una distribución conjugada es

$$\begin{aligned} f(\theta) &\propto (1 - \theta)^{\tau_0} \exp\left(\tau_1 \log \frac{\theta}{1 - \theta}\right) \\ &\propto \theta^{\tau_1} (1 - \theta)^{\tau_0 - \tau_1} \end{aligned}$$

es decir una distribución beta,  $\mathcal{B}(\alpha, \beta)$  donde  $\alpha = \tau_1 + 1$  y  $\beta = \tau_0 - \tau_1 + 1$ .

Dada una muestra de tamaño  $n$ , el estadístico suficiente es  $t(\mathbf{x}) = \{n, \sum_{i=1}^n x_i\}$  y luego la distribución a posteriori es  $\mathcal{B}(\alpha^*, \beta^*)$  donde

$$\begin{aligned}\alpha^* &= \tau_1 + \sum_{i=1}^n x_i + 1 \\ &= \alpha + \# \text{ cruces} \\ \beta^* &= \tau_0 + n - \tau_1 - \sum_{i=1}^n x_i + 1 \\ &= \beta + \# \text{ caras}\end{aligned}$$

Se ha visto en el Ejemplo 11 que la distribución beta es conjugada con la binomial.

## Interpretación de la información proporcionada por una distribución a priori conjugada

Se ha visto que el parámetro  $\tau_0$  juega el mismo papel en la distribución a priori como el tamaño de la muestra, y los demás parámetros  $(\tau_1, \dots, \tau_k)$  equivalen a los estadísticos suficientes. Entonces, se puede interpretar la información proporcionada por la distribución a priori como la información de un experimento de tamaño  $n$  donde se observan estadísticos suficientes  $(\tau_1, \dots, \tau_k)$ .

**Ejemplo 17** *Volviendo al Ejemplo 11 se ve que dada una distribución a priori  $B(\alpha, \beta)$ , la distribución a posteriori es*

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

*La información contenida en la distribución a priori es equivalente a una muestra de  $\alpha + \beta$  tiradas de la moneda,  $\alpha$  de ellas siendo cruces.*

## La media a posteriori como media ponderada

Se puede expresar la media a posteriori como una media ponderada de la media a priori y el EMV.

**Ejemplo 18** *En el Ejemplo 12, tenemos*

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

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

*La media a posteriori es una media ponderada con pesos proporcionales al número de observaciones equivalentes en la distribución inicial ( $\beta$ ) y al número de datos ( $n$ ).*



## Formalización

Supongamos una familia exponencial con parámetro natural  $\boldsymbol{\theta}$  (de dimensión  $k$ ),

$$f(x|\boldsymbol{\theta}) \propto D(\boldsymbol{\theta}) \exp \left( \sum_{i=1}^k \theta_i s_i(x) \right).$$

La distribución a priori conjugada es

$$\begin{aligned} f(\boldsymbol{\theta}) &\propto D(\boldsymbol{\theta})^{\tau_0} \exp \left( \sum_{i=1}^k \theta_i \tau_i \right) \\ &= \exp \left( \sum_{i=1}^k \theta_i \tau_i - \tau_0 G(\boldsymbol{\theta}) \right) \quad \text{donde} \\ G(\boldsymbol{\theta}) &= -\log D(\boldsymbol{\theta}) \end{aligned}$$

y la distribución a posteriori es

$$f(\boldsymbol{\theta}|\mathbf{x}) \propto \exp \left( \sum_{i=1}^k \theta_i (\tau_i + \sum_{j=1}^n s_i(x_j)) - (\tau_0 + n) G(\boldsymbol{\theta}) \right).$$

Luego se tiene el siguiente teorema.

### Teorema 3

$$E[\nabla G(\boldsymbol{\theta})] = w\bar{s} + (1 - w)\bar{\tau} \quad \text{donde}$$

$$[\nabla G(\boldsymbol{\theta})]_i = \frac{\partial}{\partial \theta_i} G(\boldsymbol{\theta})$$

$$w = \frac{n}{\tau_0 + n} \quad \text{y}$$

$$\bar{\tau} = \frac{1}{\tau_0} (\tau_1, \dots, \tau_k)$$

$$\bar{s} = \frac{1}{n} \left( \sum_{j=1}^n s_1(x_j), \dots, \sum_{j=1}^n s_k(x_j) \right)$$

El teorema demuestra que la media a posteriori de  $\nabla G(\boldsymbol{\theta})$  es una media ponderada de la media a priori y la media muestral de los estadísticos suficientes  $\bar{s}$ , es decir  $\nabla G(\hat{\boldsymbol{\theta}})$  donde  $\hat{\boldsymbol{\theta}}$  es el EMV de  $\boldsymbol{\theta}$  con ponderaciones proporcionales a  $\tau_0$  y el tamaño de la muestra.

## Demostración

Abusando de la notación, sea  $\boldsymbol{\tau} = (\tau_1, \dots, \tau_k)$ .  
Luego

$$\begin{aligned}\tau_0 (\boldsymbol{\tau} - E[\nabla G(\boldsymbol{\theta})]) &= \int \tau_0 (\boldsymbol{\tau} - \nabla G(\boldsymbol{\theta})) f(\boldsymbol{\theta}) d\boldsymbol{\theta} \\ &= \int \nabla f(\boldsymbol{\theta}) d\boldsymbol{\theta} \quad \text{donde} \\ \nabla [f(\boldsymbol{\theta})]_i &= \frac{\partial}{\partial \theta_i} f(\boldsymbol{\theta}) \quad \text{que implica} \\ E[\nabla G(\boldsymbol{\theta})] &= \boldsymbol{\tau}\end{aligned}$$

Entonces, a posteriori, mediante el mismo argumento, se tiene:

$$E[\nabla G(\boldsymbol{\theta})|\mathbf{x}] = \boldsymbol{\tau} + \sum_{j=1}^n \mathbf{s}(\mathbf{x}_j)$$

donde  $\mathbf{s}(x) = (s_1(x), \dots, s_k(x))$ . Luego

$$\begin{aligned}E[\nabla G(\boldsymbol{\theta})|\mathbf{x}] &= \frac{\tau_0 \bar{\boldsymbol{\tau}} + n \bar{\mathbf{s}}}{\tau_0 + n} \\ &= w \bar{\mathbf{s}} + (1 - w) \bar{\boldsymbol{\tau}}\end{aligned}$$

◇

**Ejemplo 19** Retomamos el Ejemplo 18. Se tiene  $X|\theta \sim \mathcal{P}(\theta)$ . Entonces

$$f(x|\theta) \propto \exp(x \log \theta - \theta)$$

y reparameterizando,  $\phi = \log \theta$ ,

$$f(x|\phi) = \exp(\phi x - e^\phi)$$

Una distribución a priori conjugada para  $\phi$  es

$$f(\phi) \propto \exp(\phi \tau_1 - \tau_0 e^\phi)$$

y entonces, la media a posteriori de  $\frac{\partial}{\partial \phi} e^\phi = e^\phi = \theta$  es

$$E[\theta|\mathbf{x}] = \frac{\tau_0 \bar{\tau}_1 + n \bar{x}}{\tau_0 + n} \quad \text{donde } \bar{\tau}_1 = \frac{\tau_1}{\tau_0}.$$

Haciendo la transformación al revés,  $\theta = \log \phi$ , observamos que la distribución a priori implícita para  $\theta$  es

$$f(\theta) \propto \theta^{\tau_1 - 1} \exp(-\tau_0 \theta) \quad \text{es decir}$$
$$\theta \sim \mathcal{G}(\tau_1, \tau_0)$$

y sustituyendo  $\alpha = \tau_1$  y  $\beta = \tau_0$ , se ha demostrado el resultado del Ejemplo 18.

En el caso más general, las ponderaciones no necesariamente son proporcionales a  $n$  y  $\tau_0$ .

**Ejemplo 20** *Volvemos al Ejemplo 15 con datos exponenciales y distribución a priori gamma. En este caso, la distribución a posteriori es  $\mathcal{G}(\alpha + n, \beta + n\bar{x})$  y la media a posteriori es*

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

donde  $w = \frac{\bar{x}}{\bar{x} + \beta}$ .

No obstante, se puede utilizar el teorema para demostrar que la media a posteriori de  $\frac{1}{\theta}$  es una combinación lineal de la media a priori  $\frac{\beta}{\alpha - 1}$  y  $\bar{x}$ :

$$E\left[\frac{1}{\theta} \mid \mathbf{x}\right] = w\bar{x} + (1 - w)\frac{\beta}{\alpha - 1}$$

donde  $w = \frac{n}{n + \alpha - 1}$ .

## Inferencia para la media de una distribución normal

**Ejemplo 21**  $X|\mu \sim \mathcal{N}(\mu, \sigma^2)$  ( $\sigma^2$  conocido).

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

es decir, una familia exponencial con  $C(\mu) = \exp\left(-\frac{1}{2\sigma^2}\mu^2\right)$ ,  $s_1(x) = x$  y  $R_1(\mu) = \frac{\mu}{\sigma^2}$ .

Entonces, una distribución a priori conjugada será de forma

$$\begin{aligned} f(\mu) &\propto \left[\exp\left(-\frac{1}{2\sigma^2}\mu^2\right)\right]^{\tau_0} \exp\left(\frac{\mu}{\sigma^2}\tau_1\right) \\ &\propto \exp\left(-\frac{\tau_0}{2\sigma^2}\left(\mu^2 - 2\frac{\tau_1}{\tau_0}\mu\right)\right) \end{aligned}$$

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

$$\mu \sim \mathcal{N}(m, \delta^2)$$

donde  $m = \tau_1/\tau_0$  y  $\delta^2 = \sigma^2/\tau_0$ .

## La distribución a posteriori. Cálculo directo mediante el teorema de Bayes.

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

$$\begin{aligned}l(\mu|\mathbf{x}) &\propto (\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right) \\ &\propto (\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \left[(n-1)s^2 + n(\mu - \bar{x})^2\right]\right)\end{aligned}$$

*donde  $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$  es el estimador clásico de la varianza.*

*Entonces, dada una distribución a priori normal, la distribución a posteriori es*

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

## Cálculo usando las propiedades de la familia exponencial.

Se ha visto que la distribución a priori conjugada es  $\mathcal{N}(m, \delta^2)$  donde  $m = \tau_1/\tau_0$  y  $\delta^2 = \sigma^2/\tau_0$  y que

$$\sum_{j=1}^n s_1(x_j) = \sum_{j=1}^n x_j = n\bar{x}.$$

Entonces, la distribución a posteriori es  $\mathcal{N}(m^*, \delta^{2*})$  donde

$$m^* = \frac{\tau_1 + \sum_{i=1}^n x_i}{\tau_0 + n}$$
$$\delta^{2*} = \frac{\sigma^2}{\tau_0 + n}$$

y entonces, sustituyendo por  $\tau_0$  y  $\tau_1$ ,

$$\tau_0 = \sigma^2/\delta^2$$
$$\tau_1 = m\sigma^2/\delta^2.$$



Luego, se tiene

$$\begin{aligned}\delta^{2*} &= \frac{\sigma^2}{\sigma^2/\delta^2 + n} \\ &= \left( \frac{1}{\delta^2} + \frac{n}{\sigma^2} \right)^{-1} \\ m^* &= \frac{m\sigma^2/\delta^2 + \sum_{i=1}^n x_i}{\sigma^2/\delta^2 + n} \quad \text{y dividiendo por } \sigma^2 \\ m^* &= \frac{\frac{1}{\delta^2}m + \frac{n}{\sigma^2}\bar{x}}{\frac{1}{\delta^2} + \frac{n}{\sigma^2}}\end{aligned}$$

es decir el mismo resultado como anteriormente.

## Propiedades de la distribución a posteriori

### Observación 17

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

donde  $\hat{\mu} = \bar{x}$  es el EMV de  $\mu$  y  $w = \frac{\frac{1}{\delta^2}}{\frac{1}{\delta^2} + \frac{n}{\sigma^2}}$ . Las ponderaciones son proporcionales a las precisiones del estimador a priori ( $m$ ) y el EMV.

**Observación 18** *Un intervalo de credibilidad a posteriori de 95 % para  $\mu$  es*

$$\frac{\frac{m}{\delta^2} + \frac{n}{\sigma^2}\bar{x}}{\frac{1}{\delta^2} + \frac{n}{\sigma^2}} \pm 1,96\sqrt{\left(\frac{1}{\delta^2} + \frac{n}{\sigma^2}\right)^{-1}}$$

**Observación 19** Si y sólo si  $\frac{1}{\delta^2} = 0$ , el intervalo será igual al intervalo clásico de confianza.

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

En esta situación la distribución a priori es impropia

$$\mu \sim \lim_{\delta^2 \rightarrow \infty} \mathcal{N}(m, \delta^2) \Rightarrow f(\mu) \propto c$$

**Observación 20** A menudo se expresa la distribución a priori de otra forma, escribiendo  $\delta^2 = \sigma^2/\alpha$ . Entonces, la distribución a priori es

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

y la distribución a posteriori reduce a

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

Es posible derivar este resultado directamente a través del Teorema 3.

## Mixturas de distribuciones conjugadas

La familia de mixturas de distribuciones conjugadas es también conjugada. Sea la distribución a priori

$$p(\boldsymbol{\theta}) = \sum_{i=1}^r w_i f_i(\boldsymbol{\theta})$$

donde  $f_i(\cdot)$  es conjugada con una distribución muestral  $f(x|\boldsymbol{\theta})$  en el sentido de Definición 3. Entonces

$$f(\boldsymbol{\theta}|x) = \sum_{i=1}^r w_i^* f_i^*(\boldsymbol{\theta})$$

donde  $f_i^*(\cdot)$  es también conjugada.

**Observación 21** *Es posible aproximar cualquier distribución  $f(\cdot)$  con una mixtura suficientemente grande de densidades conjugadas.*

**Ejemplo 22** Volviendo al Ejemplo 11, supongamos que la distribución a priori es una mixtura de una distribución uniforme y una distribución más informativa:

$$\theta \sim 0,25\mathcal{B}(1, 1) + 0,75\mathcal{B}(5, 5).$$

$$\begin{aligned} f(\theta|x) &\propto \left( 0,25 + 0,75 \frac{1}{B(5, 5)} \theta^{5-1} (1 - \theta)^{5-1} \right) \\ &\quad \theta^9 (1 - \theta)^3 \\ &\propto 0,25 \theta^{10-1} (1 - \theta)^{4-1} + \\ &\quad \frac{0,75}{B(5, 5)} \theta^{14-1} (1 - \theta)^{8-1} \\ &\propto 0,25 B(10, 4) \frac{1}{B(10, 4)} \theta^{10-1} (1 - \theta)^{4-1} + \\ &\quad \frac{0,75 B(14, 8)}{B(5, 5)} \frac{1}{B(14, 8)} \theta^{14-1} (1 - \theta)^{8-1} \\ &= w \mathcal{B}(10, 4) + (1 - w) \mathcal{B}(14, 8) \end{aligned}$$

$$\text{donde } w = \frac{0,25 B(10, 4)}{0,25 B(10, 4) + \frac{0,75 B(14, 8)}{B(5, 5)}}.$$

## Inferencia para la distribución uniforme

**Ejemplo 23** Aunque la distribución uniforme ( $X \sim \mathcal{U}(0, \theta)$ ) es una familia exponencial irregular, si existe una distribución conjugada.

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

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

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

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

donde  $\theta > \alpha$ .

La distribución a posteriori es

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

para  $\theta > \alpha^* = \text{máx}\{\alpha, x_1, \dots, x_n\}$  es decir que  $\theta|\mathbf{x} \sim \mathcal{PA}(\alpha^*, \beta + n)$ .

## Distribuciones conjugadas no siempre son fáciles de utilizar

Si se conocen las propiedades de la familia conjugada de distribuciones, entonces la inferencia es sencilla. No obstante, existen algunas excepciones.

**Ejemplo 24** *Supongamos que  $X|\theta \sim \mathcal{B}(\alpha, \theta)$  donde  $\alpha$  es conocido. Entonces*

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

*y una distribución a priori conjugada será*

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

*cuando*

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

*y es imposible calcular la constante de integración, la media ... sin el uso de integración numérica.*

## Aplicación 1: Inferencia para colas Markovianas

Consideramos inferencia para el sistema de colas  $M/M/1$  siguiendo Armero y Bayarri (1994a).

El sistema de colas  $M(\lambda)/M(\mu)/1$  es un sistema con tiempos entre llegadas  $X \sim \mathcal{E}(\lambda)$ , tiempos de servicio  $Y \sim \mathcal{E}(\mu)$  y un servidor. Si llega un cliente y el servidor no está vacío, entonces espera su turno en la cola.

Típicamente se interesa por variables como el tamaño de la cola o el tiempo de espera de un cliente. Si, en principio, el sistema está vacío la distribución de  $N(t)$ , el número de clientes en el sistema después de un tiempo  $t$  tiene una fórmula complicada (ver Gross y Harris 1985):



$$\begin{aligned}
P(N(t) = n) = e^{-(\lambda+\mu)t} & \left[ \left(\frac{\lambda}{\mu}\right)^{(n-n_0)/2} I_{n-n_0} \left(2\sqrt{\lambda\mu t}\right) \right. \\
& + \left(\frac{\lambda}{\mu}\right)^{(n-n_0-1)/2} I_{n+n_0+1} \left(2\sqrt{\lambda\mu t}\right) \\
& \left. + \left(1 - \frac{\lambda}{\mu}\right) \left(\frac{\lambda}{\mu}\right)^{n/2} \sum_{j=n+n_0+1}^{\infty} I_j \left(2\sqrt{\lambda\mu t}\right) \right]
\end{aligned}$$

donde  $I_j(c)$  es una función Bessel modificada, es decir

$$I_j(c) = \sum_{k=0}^{\infty} \frac{(c/2)^{j+2k}}{k!(j+k)!}$$

Es interesante considerar bajo cuáles condiciones converge esta distribución cuando  $t \rightarrow \infty$ . Claramente, si la tasa de llegadas es más alta que la tasa de servicios, se esperaría que el número de personas esperando crecería con el tiempo.

## Condición del equilibrio del sistema

El sistema es estable si y sólo si la *intensidad de tráfico*  $\rho = \lambda/\mu < 1$ .

En este caso la distribución de  $N(t)$  se acerca a un límite cuando  $t \rightarrow \infty$ :

$$\begin{aligned} P(N = n|\rho) &= \lim_{t \rightarrow \infty} P(N(t) = n|\rho) \\ &= (1 - \rho)\rho^n. \end{aligned}$$

Ver por ejemplo Gross y Harris (1985).

Igualmente, el tiempo,  $W$ , pasado en el sistema por un cliente llegando en equilibrio es exponencial

$$W|\lambda, \mu \sim \mathcal{E}(\mu - \lambda).$$

**Observación 22** *También existen expresiones explícitas para el tiempo de espera, periodos de ocupación etc. Ver Gross y Harris (1985).*

Para tomar decisiones sobre la introducción de nuevos servidores etc. se necesita estudiar la estabilidad del sistema y las distribuciones en el equilibrio.

## Experimento y Inferencia

Se supone que  $\lambda$  y  $\mu$  son desconocidos.

Un experimento sencillo es el de observar  $n_l$  tiempos entre llegadas y  $n_s$  tiempos de servicio. En este caso, la verosimilitud es

$$l(\lambda, \mu | \mathbf{x}, \mathbf{y}) \propto \lambda^{n_l} e^{-\lambda t_l} \mu^{n_s} e^{-\mu t_s}$$

donde  $t_l$  y  $t_s$  son las sumas de los tiempos de llegadas y servicios respectivamente.

Suponiendo unas distribuciones a priori (independientes) gamma;

$$\lambda \sim \mathcal{G}(\alpha_l, \beta_l) \quad \mu \sim \mathcal{G}(\alpha_s, \beta_s)$$

entonces las distribuciones a posteriori son también (independientes) gamma:

$$\begin{aligned} \lambda | \mathbf{x} &\sim \mathcal{G}(\alpha_l + n_l, \beta_l + t_l) \\ \mu | \mathbf{y} &\sim \mathcal{G}(\alpha_s + n_s, \beta_s + t_s) \end{aligned}$$

## Estimación de la intensidad de tráfico

Es fácil estimar la intensidad de tráfico  $\rho$ . Su esperanza es:

$$\begin{aligned} E[\rho|\mathbf{x}, \mathbf{y}] &= E[\lambda/\mu|\mathbf{x}, \mathbf{y}] \\ &= E[\lambda|\mathbf{x}]E[1/\mu|\mathbf{y}] \\ &= \frac{\alpha_l + n_l}{\beta_l + t_l} \frac{\beta_s + t_s}{\alpha_s + n_s - 1} \end{aligned}$$

Además es fácil evaluar la distribución de  $\rho$ .

A posteriori,  $\lambda$  y  $\mu$  son independientes y

$$2(\beta_l + t_l)\lambda|\mathbf{x} \sim \chi_{2(\alpha_l + n_l)}^2 \quad 2(\beta_s + t_s)\mu|\mathbf{y} \sim \chi_{2(\alpha_s + n_s)}^2$$

y entonces

$$\frac{(\beta_l + t_l)(\alpha_s + n_s)}{(\alpha_l + n_l)(\beta_s + t_s)}\rho|\mathbf{x}, \mathbf{y} \sim \mathcal{F}_{\frac{2(\alpha_l + n_l)}{2(\alpha_s + n_s)}}.$$

La probabilidad a posteriori de que el sistema sea estable es  $p = P(\rho < 1|\mathbf{x}, \mathbf{y})$ . Se la puede evaluar mediante integración numérica.

## Estimación del tamaño de la cola en equilibrio

Si  $p$  es grande, es natural suponer que la cola sea estable. En este caso, se puede hacer inferencia sobre el tamaño de la cola en equilibrio.

$$\begin{aligned} P(N = n | \mathbf{x}, \mathbf{y}, \text{equilibrio}) &= P(N = n | \mathbf{x}, \mathbf{y}, \rho < 1) \\ &= \int_0^1 (1 - \rho) \rho^n f(\rho | \mathbf{x}, \mathbf{y}, \rho < 1) d\rho \\ &= \frac{1}{p} \int_0^1 (1 - \rho) \rho^n f(\rho | \mathbf{x}, \mathbf{y}) d\rho \end{aligned}$$

Se puede expresar este integral en términos de la función Gauss hipergeométrica:

$$P(N = n|\cdot) = \frac{\alpha_l^* \Gamma(\alpha_l^* + n)}{\Gamma(\alpha_l^* + n + 2)} \times \frac{{}_2F_1\left(\alpha_l^* + \alpha_s^*, \alpha_l^* + n; \alpha_l^* + n + 2; -\frac{\beta_l^*}{\beta_s^*}\right)}{{}_2F_1\left(\alpha_l^* + \alpha_s^*, \alpha_l^*; \alpha_l^* + 1; -\frac{\beta_l^*}{\beta_s^*}\right)}$$

donde  $\alpha_l^* = \alpha_l + n_l$ ,  $\beta_l^* = \beta_l + t_l$ ,  $\alpha_s^* = \alpha_s + n_s$ ,  $\beta_s^* = \beta_s + t_s$  y la función Gauss hipergeométrica es

$${}_2F_1(a, b; c; d) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)} \int_0^1 z^{b-1} (1-z)^{c-b-1} (1-dz)^{-a} dz.$$

No obstante, se necesitan métodos numéricos para evaluar la integral. Existen varias posibilidades (integración numérica, Monte Carlo ...)

**Observación 23** Si  $\alpha_s$  es un número entero, se puede evaluar el integral explícitamente.

## Resultados explícitos cuando $\alpha_s$ es un número entero

$$\begin{aligned}
 P(N = n | \mathbf{x}, \mathbf{y}) &= E [(1 - \rho)\rho^n | \mathbf{x}, \mathbf{y}, \rho < 1] \\
 &= E [\rho^n | \mathbf{x}, \mathbf{y}, \rho < 1] - \\
 &\quad E [\rho^{n+1} | \mathbf{x}, \mathbf{y}, \rho < 1]
 \end{aligned}$$

Hallamos los momentos de  $\rho$  condicionado en el equilibrio.

$$\begin{aligned}
 E [\rho^r | \mathbf{x}, \mathbf{y}, \rho < 1] &= E \left[ \left( \frac{\lambda}{\mu} \right)^r | \mathbf{x}, \mathbf{y}, \lambda < \mu \right] \\
 &= \frac{1}{p} \int_0^\infty \int_\lambda^\infty \left( \frac{\lambda}{\mu} \right)^r f(\lambda | \mathbf{x}) f(\mu | \mathbf{y}) d\mu d\lambda \\
 &= \frac{1}{p} \int_0^\infty \lambda^r f(\lambda | \mathbf{x}) \int_\lambda^\infty \mu^{-r} f(\mu | \mathbf{y}) d\mu d\lambda \\
 &= \frac{1}{p} \int_0^\infty \lambda^r f(\lambda | \mathbf{x}) \frac{\Gamma(\alpha_s^* - r)}{\Gamma(\alpha_s^*)} (\beta_s^*)^r \\
 &\quad \int_\lambda^\infty g(\mu | \alpha_s^* - r, \beta_s^*) d\mu d\lambda
 \end{aligned}$$



donde  $g(\mu|\alpha_s^* - r, \beta_s^*)$  representa una función de densidad gamma (Erlang) con parámetros  $\alpha_s^* - r$  y  $\beta_s^*$ .

Recordamos un resultado para un proceso Poisson. Si el número de sucesos en una unidad de tiempo es Poisson con tasa  $\theta$ , entonces el tiempo hasta el primer suceso es exponencial con la misma tasa y el tiempo hasta el suceso  $r$  es Erlang o gamma con parámetros  $r$  y  $\theta$ .

Luego  $\int_{\lambda}^{\infty} g(\mu|\alpha_s^* - r, \beta_s^*) d\mu$  es la probabilidad de que el suceso número  $\alpha_s^* - r$  ocurra después de un tiempo  $\lambda$ . Esto es igual a decir que el número de sucesos en  $[0, \lambda)$  es menor de  $\alpha_s^* - r$  y luego

$$\int_{\lambda}^{\infty} g(\mu|\alpha_s^* - r, \beta_s^*) d\mu = \sum_{j=0}^{\alpha_s^* - r - 1} \frac{(\beta_s^* \lambda)^j}{j!} e^{-\beta_s^* \lambda}.$$

Entonces,

$$\begin{aligned}
E[\rho^r | \mathbf{x}, \mathbf{y}, \rho < 1] &= \frac{1}{p} \int_0^\infty \lambda^r f(\lambda | \mathbf{x}) \frac{\Gamma(\alpha_s^* - r)}{\Gamma(\alpha_s^*)} (\beta_s^*)^r \\
&\quad \sum_{j=0}^{\alpha_s^* - r - 1} \frac{(\beta_s^* \lambda)^j}{j!} e^{-\beta_s^* \lambda} d\lambda \\
&= \frac{1}{p} \frac{\Gamma(\alpha_s^* - r)}{\Gamma(\alpha_s^*)} (\beta_s^*)^r \\
&\quad \sum_{j=0}^{\alpha_s^* - r - 1} \frac{\beta_s^{*j}}{j!} \int_0^\infty \lambda^r e^{-\beta_s^* \lambda} f(\lambda | \mathbf{x}) d\lambda \\
&= \frac{1}{p} \frac{\Gamma(\alpha_s^* - r)}{\Gamma(\alpha_s^*)} (\beta_s^*)^r \sum_{j=0}^{\alpha_s^* - r - 1} \frac{\beta_s^{*j}}{j!} \\
&\quad \frac{\Gamma(\alpha_l^* + j)}{\Gamma(\alpha_l^*)} \frac{\beta_l^{*\alpha_l^*}}{(\beta_l^* + \beta_s^*)^{\alpha_l^* + j}} \\
&= \frac{1}{p} \frac{\Gamma(\alpha_s^* - r)}{\Gamma(\alpha_s^*)} (\beta_s^*)^r P(T \leq \alpha_s^* - r - 1)
\end{aligned}$$

donde  $T \sim \mathcal{BN}\left(\alpha_l^*, \frac{\beta_l^*}{\beta_l^* + \beta_s^*}\right)$ .

Luego se tiene una expresión explícita para  $P(N = n | \mathbf{x}, \mathbf{y})$  y de la misma manera es posible hallar la distribución predictiva de  $W$ .

## Estimación de la distribución de $N(t)$ por simulación Monte Carlo

La idea es simular una muestra de las distribuciones a posteriori de  $\lambda$  y  $\mu$  suponiendo la condición del equilibrio y usar los datos muestrales para estimar las probabilidades de interés.

Se usa el siguiente algoritmo para estimar las distribuciones predictivas de  $N$  y  $W$  a través de una muestra Monte Carlo de tamaño  $S$ . Para estimar la distribución de  $N(t)$ , no es necesario imponer la condición del equilibrio.

## Algoritmo

1. Para  $i = 1, \dots, S$

a) Generar  $\lambda_i \sim \mathcal{G}(\alpha_l + n_l, \beta_l + t_l)$

b) Generar  $\mu_i \sim \mathcal{G}(\alpha_s + n_s, \beta_s + t_s)$

c) Definir  $\rho_i = \frac{\lambda_i}{\mu_i}$ .

2.

$$P(N(t) = n | \text{data}) \approx \frac{1}{S} \sum_{i=1}^S P(N(t) = n | \lambda_i, \mu_i)$$

Se puede usar el mismo tipo de algoritmo para estimar la densidad de un periodo de ocupación.

## Estimación de los momentos

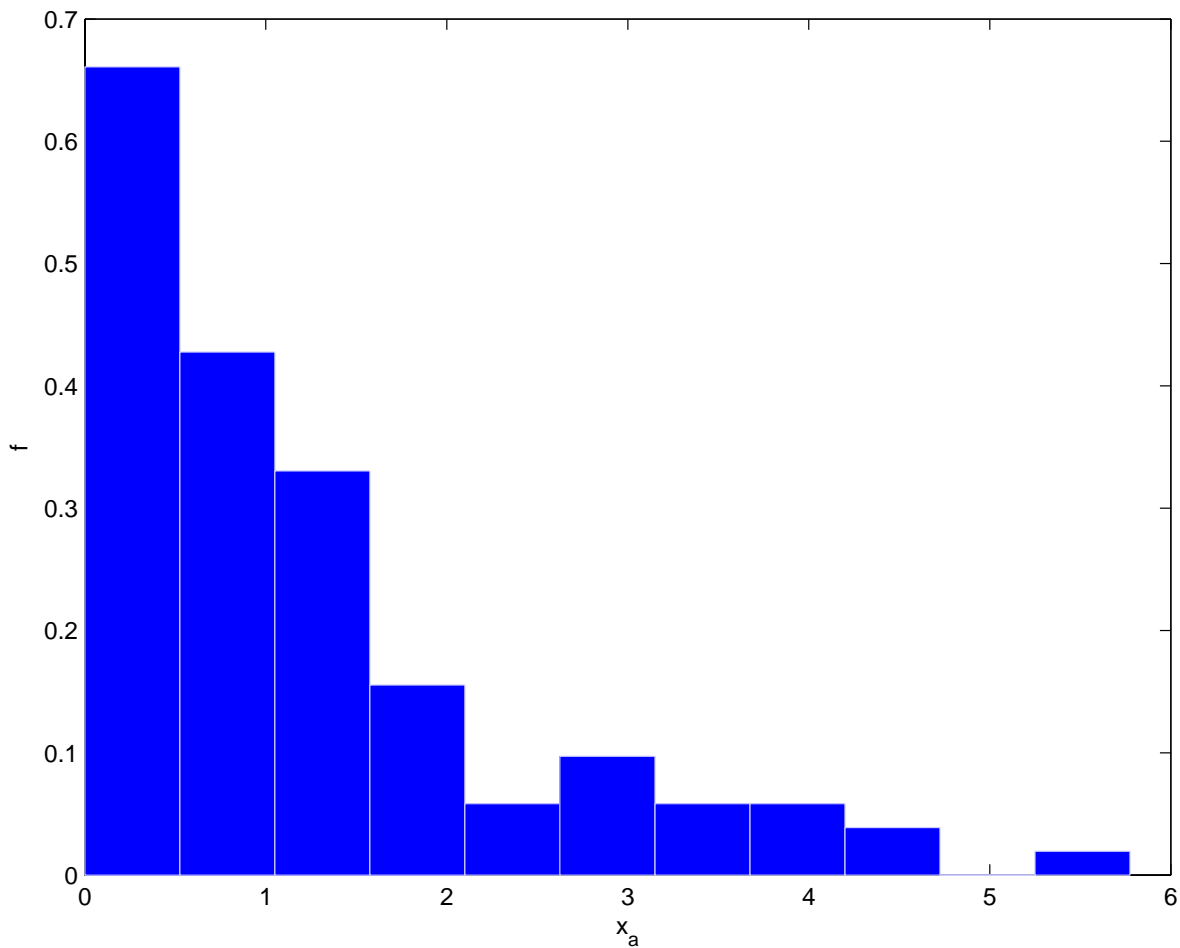
A pesar de que los momentos de  $N(t)$  existen para cualquier  $t$ , las variables  $N$  y  $W$  no tienen media.

$$\begin{aligned} E[N|\mathbf{x}, \mathbf{y}, \rho < 1] &= E[E[N|\rho]|\mathbf{x}, \mathbf{y}, \rho < 1] \\ &= E\left[\frac{\rho}{1-\rho}|\mathbf{x}, \mathbf{y}, \rho < 1\right] \\ &= \frac{1}{p} \int_0^1 \frac{\rho}{1-\rho} f(\rho|\mathbf{x}, \mathbf{y}) d\rho \\ &= \infty \end{aligned}$$

**Observación 24** *Es una característica general en inferencia para colas. El mismo resultado ocurre con el tiempo de espera o el periodo de ocupación. Ver Armero y Bayarri (1994a), Wiper (1998).*

*Se resuelve el problema parcialmente suponiendo a priori que  $P(\rho \geq 1) = 0$ . Ver Armero y Bayarri (1994b) o Ruggeri et al (1996).*

**Ejemplo 25** *Hall (1991) presenta datos de un cajero automático en Berkeley. El siguiente histograma muestra los tiempos entre llegadas. Parece que se ajusten a una distribución exponencial.*



*(Los servicios no se ajustan tan bien pero ...)*

*Los estadísticos suficientes son*

$$n_a = n_s = 98, \quad t_a = 119,71 \quad t_s = 81,35$$

*Suponiendo distribuciones a priori no-informativas*

$$f(\lambda) \propto \frac{1}{\lambda} \quad f(\mu) \propto \frac{1}{\mu}$$

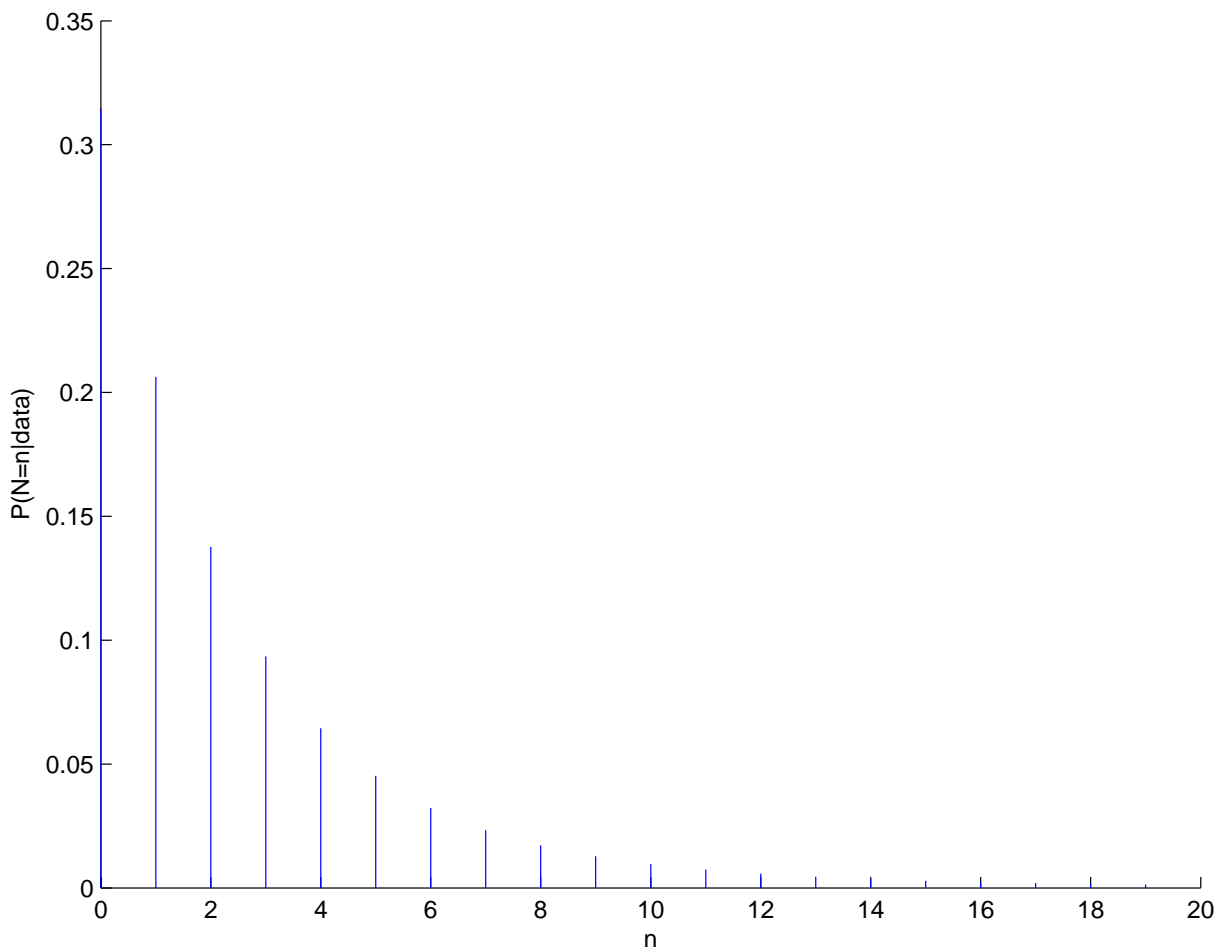
*las distribuciones a posteriori son*

$$\lambda|\mathbf{x} \sim \mathcal{G}(98, 119,71) \quad \mu|\mathbf{y} \sim \mathcal{G}(98, 81,35)$$

*y entonces  $E[\rho|\mathbf{x}, \mathbf{y}] \approx 0,668$  y se puede demostrar que  $P(\rho < 1|\mathbf{x}, \mathbf{y}) \approx 0,997$ .*

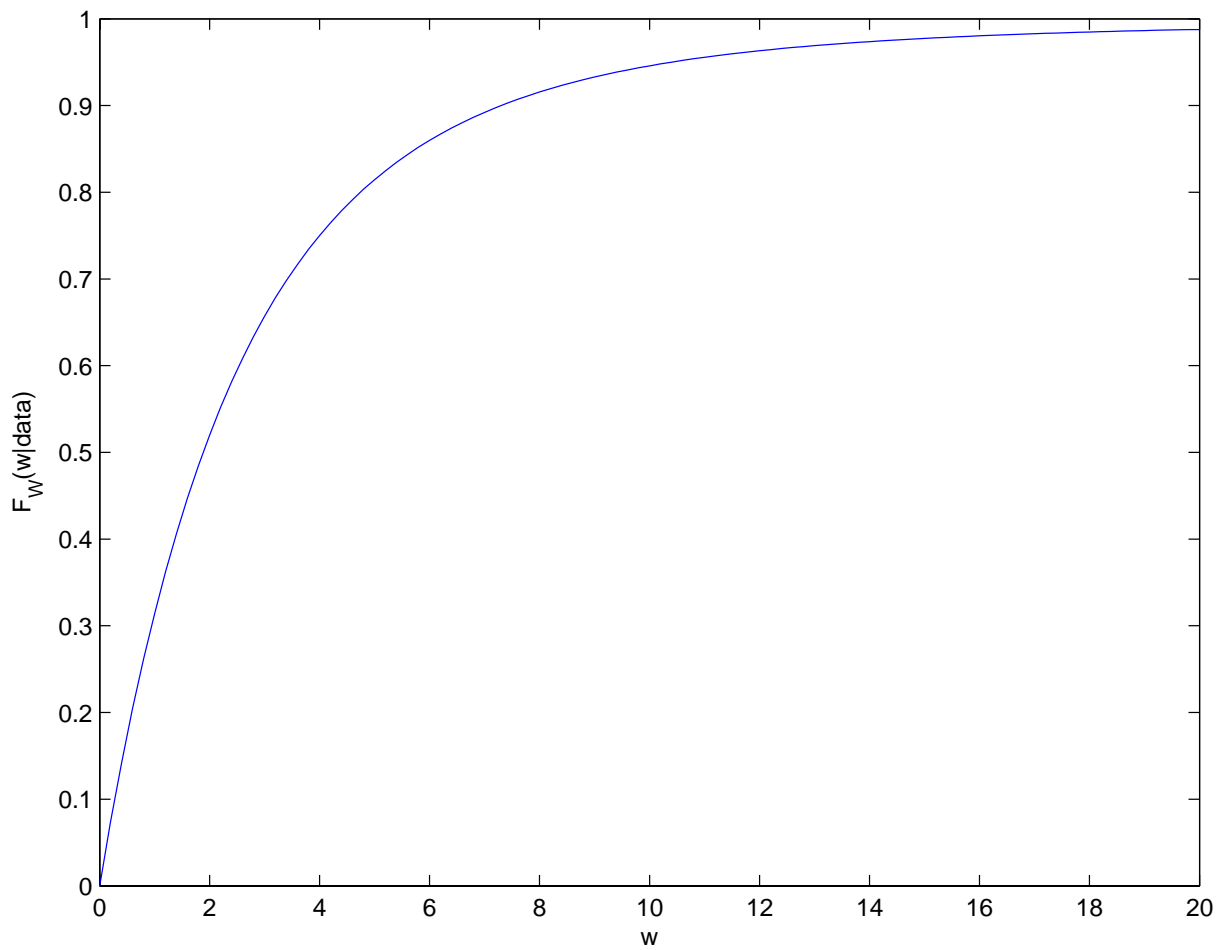
*Parece razonable suponer equilibrio.*

*El primer dibujo muestra las probabilidades para diferentes valores del tamaño de la cola en equilibrio.*

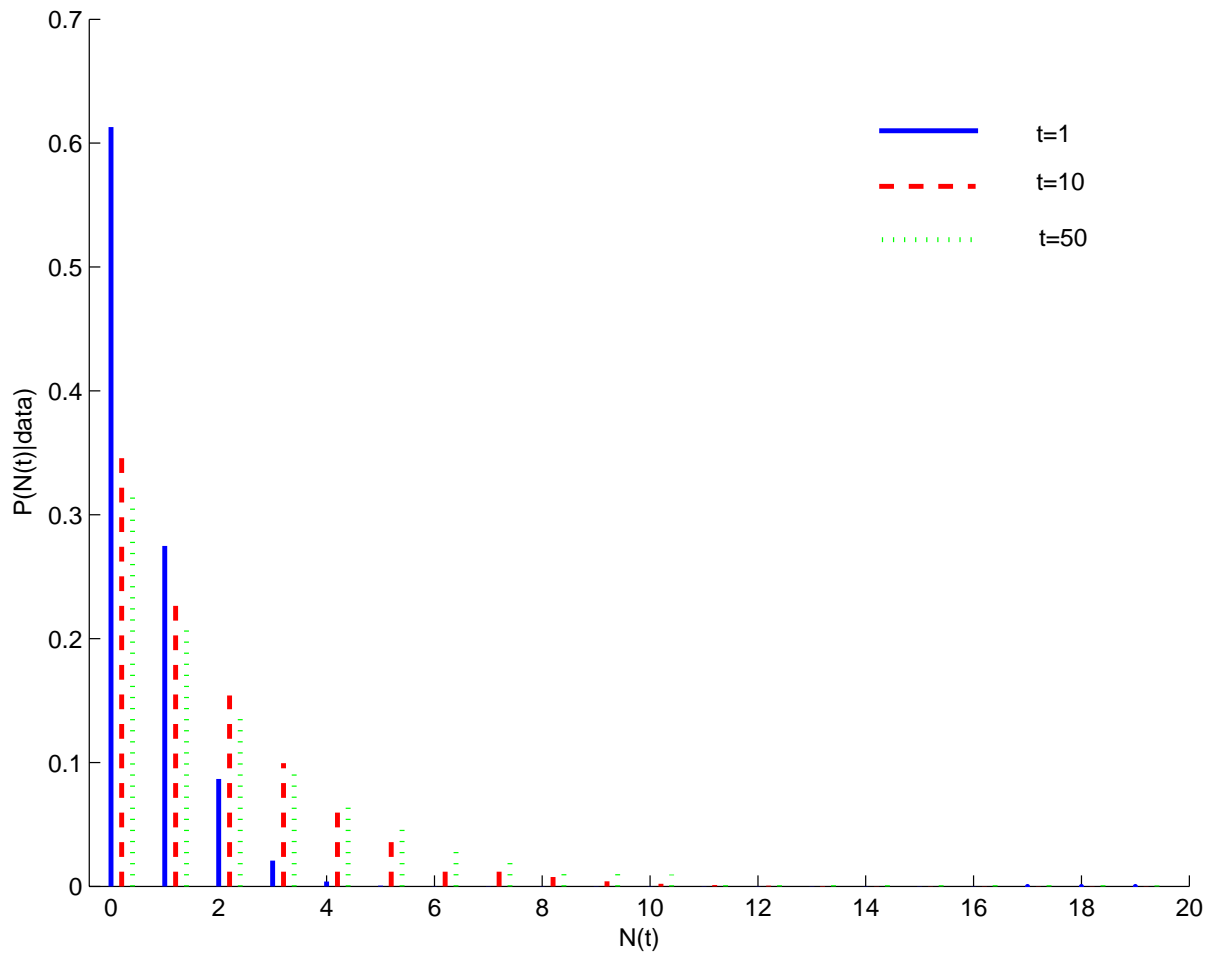


El siguiente diagrama muestra la distribución predictiva de  $W$





El último dibujo muestra la distribución de  $N(t)$  para varios tiempos  $t$  sin suponer el equilibrio.



Se ve la convergencia a la distribución de  $N$ .

## Aplicación 2: Comprobación de software con múltiples usuarios

Estudiamos el problema de la detección de errores en software siguiendo Wiper y Wilson (2007).

¿Porqué comprobar el software?

Se quiere detectar y depurar los errores (*bugs* o *faults*) en el programa y mejorar la calidad del producto final.

El modelo tipo académico es lo siguiente.

Se generan entradas del perfil de un usuario típico y se graban los tiempos entre fallos (*failures*) sucesivos. Después de cada fallo, se depura el programa.

A lo largo del tiempo, la fiabilidad del software se mejora y se lanza el producto al mercado cuando sea suficientemente fiable. Ver, por ejemplo Singpurwalla y Wilson (1999).

En comprobación beta (*beta testing*), varios usuarios comprueban el software durante un periodo del tiempo comentando los varios fallos del programa que encuentran.

- Cada usuario puede observar varios fallos debidos al mismo error.
- Puede que un usuario olvide a mandar el informe sobre un error encontrado.
- Al final del periodo, se depura el programa y se decide entre seguir comprobando o lanzar el producto.

## Un modelo para comprobación beta

- El programa contiene  $N$  errores. Cada fallo es debido a exactamente un error.
- Para un usuario dado, el tiempo  $S_j$  hasta observar el primer fallo debido al error  $j$  se distribuye como

$$S_j | \lambda_j \sim \mathcal{E}(\lambda_j)$$

independientemente de los otros errores.

- $\Rightarrow$  el tiempo,  $T$  para observar el primer fallo  $T = \min\{S_1, \dots, S_N\}$ .

$$T | N, \boldsymbol{\lambda} \sim \mathcal{E}(\lambda_0) \quad \text{donde } \lambda_0 = \sum_{j=1}^N \lambda_j.$$

- Si  $Z$  es un indicador de la causa de un fallo, entonces

$$P(Z = j|N, \boldsymbol{\lambda}) = \frac{\lambda_j}{\lambda_0} \quad \text{para } j = 1, \dots, N$$

- $Z$  y  $T$  son independientes.

## Los usuarios

- $M_0$  usuarios intercambiables comprueban el software durante un tiempo  $T_0$ .
- A menudo, si se observa un fallo, se puede identificar la causa, pero no siempre.
- Supongamos:  $p = P(\text{se identifica la causa del fallo})$  para todos los fallos y todos los usuarios.

## Función de verosimilitud e inferencia

La verosimilitud es:

$$l(N, \lambda, p | \text{data}) = p^{r-r_0} (1-p)^{r_0} \prod_{k=1}^K \lambda_k^{r_k} \exp(-M_0 T_0 \lambda_0) \quad \text{para } N \geq K,$$

donde se han identificado  $K$  errores y  $r_k$  fallos debidos a estos errores, para  $k = 1, \dots, K$ .

$r_0$  es el número de fallos no identificados y  $r = \sum_{k=1}^K r_k$ .

El EMV para  $N$  es  $N = K$ , lo que no sirve de mucho. Mejor emplear inferencia bayesiana.

## Distribuciones a priori

A menudo, se tiene información a priori sobre  $N$  y  $\lambda_0$ , pero no sobre los errores individuales.

Se reparameteriza en términos de  $\lambda_0$  y  $\boldsymbol{\rho}$  donde  $\rho_j = \lambda_j/\lambda_0$ . Entonces:

$$l(p, N, \lambda_0, \boldsymbol{\rho} | \text{datos}) = p^{r-r_0} (1-p)^{r_0} \prod_{k=1}^N \rho_k^{r_k} \lambda_0^r \exp(-M_0 T_0 \lambda_0)$$

Distribuciones a priori  $p \sim \mathcal{B}(v, w)$  y  $\lambda_0 \sim \mathcal{G}(a, b)$  son conjugadas.



## Distribuciones a priori de $N$ y $\rho$

Supongamos que  $N \sim \mathcal{P}(\theta)$ . Consideramos dos posibilidades para  $\rho$ .

### 1. Modelo fijo ( $\mathcal{F}$ )

$$\rho_j = 1/N \quad \text{para } j = 1, \dots, N.$$

Se supone que todos los errores tienen la misma importancia, parecido al modelo de Jelinski y Moranda (1972).

### 2. Distribución Dirichlet. ( $\mathcal{D}$ )

$$\rho|N \sim \mathcal{D}(\phi, \dots, \phi) \quad \text{para algún } \phi > 0$$

Algunos errores pueden ser más grandes que otros.

Se generaliza este modelo poniendo una distribución jerárquica sobre  $\phi$ .

## Distribuciones a posteriori

$$p|\text{datos} \sim \mathcal{B}(v + r - r_0, w + r_0)$$
$$\lambda_0 \sim \mathcal{G}(a + r, b + M_0 T_0)$$

1. Modelo fijo.

$$P(N|\mathcal{F}, \text{datos}) \propto \frac{\theta^N}{N!} \frac{1}{N^{r-r_0}}$$

para  $N = K, K + 1, \dots$ . Es fácil estimar la media de errores en el programa etc. truncando los sumatorios.

2. Modelo Dirichlet.

$$\rho|\mathcal{D}, N, \text{datos} \sim \mathcal{D}(\phi + r_1, \dots, \phi + r_K, \phi, \dots, \phi)$$
$$P(N|\mathcal{D}, \text{datos}) \propto \frac{\theta^N}{N!} \frac{\Gamma(N\phi)}{\Gamma(N\phi + r - r_0)}$$

## Un segundo periodo de comprobación

Después del primer periodo de ocupación, se depura el software y se debe decidir seguir comprobando o no.

Consideramos sólo la posibilidad de comprobar con  $M_1$  usuarios durante un periodo  $T_1$ . ¿Cómo elegimos  $M_1$  y  $T_1$ ?

Definimos una función de costes y minimizamos el coste esperado.

**Observación 25** *Elegimos la decisión Bayes, es decir la decisión que maximiza la utilidad esperada. Ver el capítulo 5.*

## Una función de costes

- $c_1$  = coste por cada usuario por unidad de tiempo. (Sueldo)
- Coste  $c_2$  por unidad de tiempo. (Coste de oportunidad)
- Coste  $c_3$  por cada error encontrado durante el periodo. (Coste de depurar)
- Coste  $c_4$  por cada fallo por unidad de tiempo después del lanzamiento del producto.

La función de costes es

$$C(M_1, T_1) = c_1 M_1 T_1 + c_2 T_1 + c_3 B + c_4 \lambda_0 \left( \sum_{i=K+1}^N \rho_i (1 - I(i)) \right)$$

donde  $B$  es el número de errores encontrados en el periodo de comprobación y  $I(i)$  es un indicador de si se ha descubierto el error  $i$ .

## Evaluación del coste esperado

Observamos que  $B = \sum_{i=K+1}^N I(i)$  y entonces

$$C(M_1, T_1) = c_1 M_1 T_1 + c_2 T_1 + c_4 \lambda_0 \sum_{i=K+1}^N \rho_i + \sum_{i=K+1}^N (c_3 - c_4 \lambda_0 \rho_i) I(i)$$

Se necesita hallar el valor de  $E[I(i)|N, p, \lambda_0, \rho]$ .  
¿Cómo hacerlo?

$$E[I(i)|N, p, \lambda_0, \boldsymbol{\rho}] = P(\text{se descubre el error } i)$$

y condicionamos sobre el número de fallos  $F$  observados en el periodo de comprobación.

Se tiene

$$P(\text{descubierto}|F = f) = 1 - \left( 1 - p + p \frac{\sum_{j=K+1, j \neq i}^N \rho_j}{\sum_{j=K+1}^N \rho_j} \right)^f$$

Además

$$F|N, \lambda_0, \boldsymbol{\rho}, \text{datos} \sim \mathcal{P} \left( \lambda_0 M_1 T_1 \sum_{i=K+1}^N \rho_i \right)$$

y, combinando estos dos resultados, se puede hallar el coste esperado dados los parámetros del modelo.

Ahora es posible sacar el coste predictivo bajo cada modelo integrando sobre las distribuciones a posteriori de  $p$ ,  $\lambda_0$ ,  $\boldsymbol{\rho}$  y  $N$ .

## Coste esperado para el modelo fijo

$$\begin{aligned}
 &= c_1 M_1 T_1 + c_2 T_1 + \\
 & c_4 \frac{a+r}{b+M_0 T_0} \sum_{N=K+1}^{\infty} \frac{N-K}{N} P(N|\mathcal{F}, \text{datos}) \\
 & + \sum_{N=K+1}^{\infty} (N-K) P(N|\mathcal{F}, \text{datos}) \sum_{f=1}^{\infty} \left[ 1 - \sum_{s=0}^f \binom{f}{s} \right. \\
 & \left. \frac{B(v+r-r_0+s, w+r_0+f-s)}{B(v+r-r_0, w+r_0)} \left( \frac{N-K-1}{N-K} \right)^s \right] \\
 & \frac{\Gamma(a+r+f)}{f! \Gamma(a+r)} p_N^f (1-p_N)^{a+r} \\
 & \left( c_3 - c_4 \frac{a+r+f}{N(b+M_0 T_0 + M_1 T_1 \frac{N-K}{N})} \right)
 \end{aligned}$$

$$\text{donde } p_N = \frac{M_1 T_1 \frac{N-K}{N}}{b+M_0 T_0 + M_1 T_1 \frac{N-K}{N}}.$$

Existe una expresión parecida para el modelo Dirichlet. (Se necesita integración numérica unidimensional).

## Ejemplo

- Un programa contiene  $N = 20$  errores con tasa de fallos total  $\lambda_0 = 10$ .
- $\rho$  generado de Dirichlet(1, ..., 1).
- Un usuario con periodo de comprobación de 20 días.
- $p = 0,9$ .
- Fallos debidos a 14 errores distintos.
- $r_0 = 15$  fallos no identificados en  $r = 199$ .
- Media a priori de  $N = 20$ . A prioris no informativos para los demás parámetros.



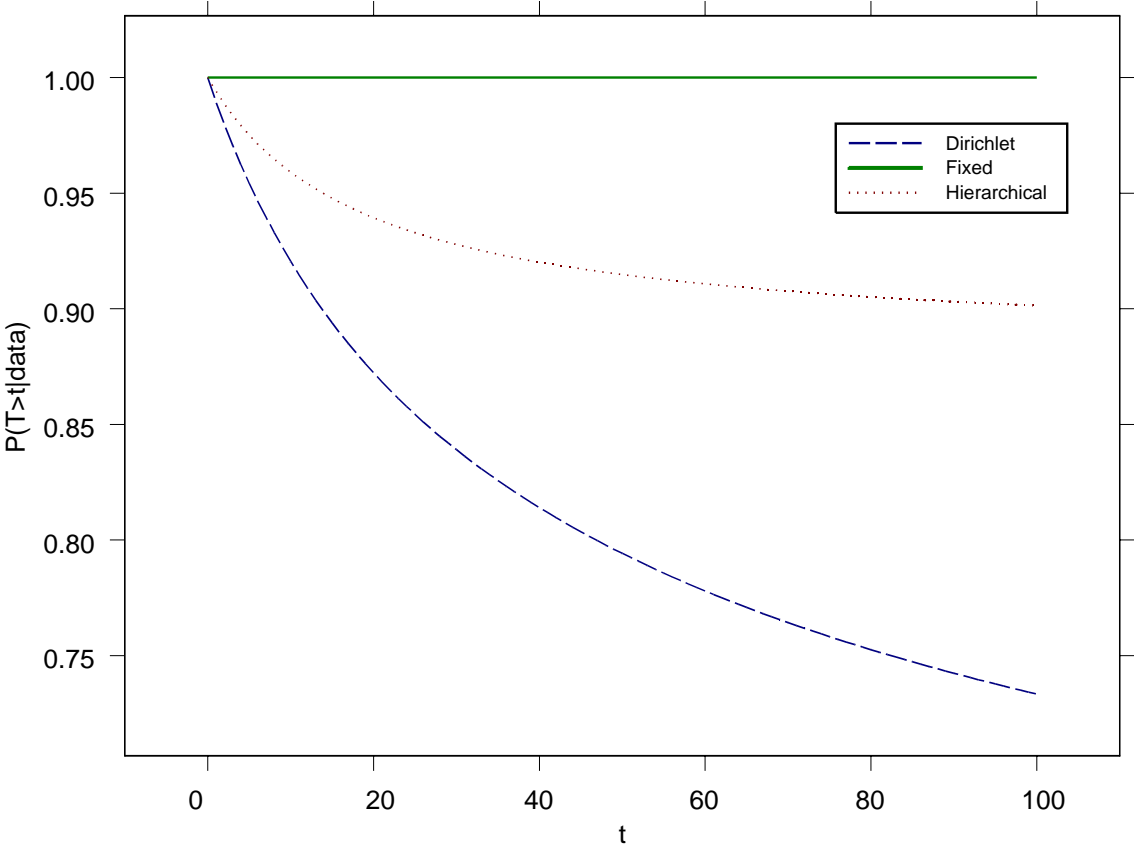
## Distribución a posteriori de $N$

$N$	Modelo		
	$\mathcal{F}$	$\mathcal{D}$	$\mathcal{H}$
14	1,0000	,2836	,8794
15	0	,2241	,1096
16	0	,1676	,0101
17	0	,1189	,0008
18	0	,0804	,0001
19	0	,0518	,0000
20	0	,0320	,0000
> 20	0	,0416	0
$E[N \text{datos}]$	14,0000	16,0256	14,1784

$\mathcal{H}$  es un modelo jerárquico, más avanzado.

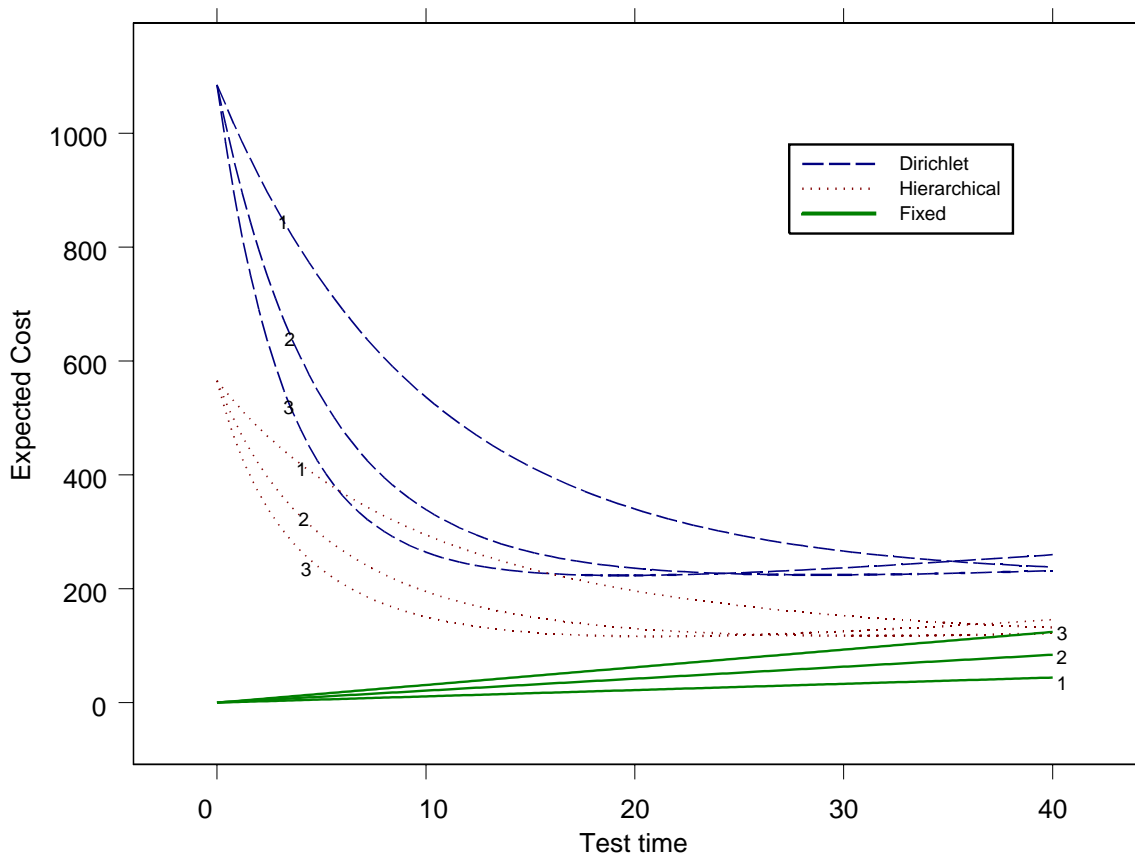
# Funciones de fiabilidad esperadas

Se supone la depuración anterior del software.  
La fiabilidad es  $P(T > t)$  donde  $T =$  tiempo al siguiente fallo.



# Coste esperado de más comprobación

Consideramos hasta 3 usuarios durante un tiempo máximo de 20 días. Costes  $c_1 = 1$ ,  $c_2 = c_3 = 0,1$ , y  $c_4 = 100000$



## Extensiones

- Aplicaciones a datos del internet. Ver Wiper y Wislon (2007).
- ¿Cómo elegimos un modelo? Utilizamos factores Bayes. Ver el capítulo 5. Otra posibilidad es hacer un promedio de las predicciones mediante los distintos modelos (*model averaging*).
- Usuarios no homogéneos.
- ¿Que hacer si no podemos fijar el valor de  $M_1$ ?