Análisis Bayesiano
(Borradores, Curso 23)

Sebastian Grynberg

17-19 de junio de 2013

Aquí no valen Dotores,
Solo vale la esperiencia,
Aquí verían su inocencia
Esos que todo lo saben;
Por que esto tiene otra llave
Y el gaucho tiene su ciencia.
(Martín Fierro)

Índice

1. Análisis Bayesiano

Si se lo compara con el modelado probabilístico, el propósito del análisis estadístico es fundamentalmente un propósito de *inversión*, ya que se propone inferir las causas (los parámetros del mecanismo aleatorio) a partir de los efectos (las observaciones). En otras palabras, cuando observamos un fenómeno aleatorio regulado por un parámetro θ, los métodos estadísticos nos permiten deducir de las observaciones una *inferencia* (esto es, un resumen, una caracterización) sobre θ, mientras que el modelado probabilístico caracteriza el comportamiento de las observaciones futuras *condicionales* a θ. Este aspecto de la estadística es obvio en la noción de función de verosimilitud, puesto que, formalmente, es la densidad conjunta de la muestra reescrita en el orden propio

$$L(\theta|\mathbf{x}) = f(\mathbf{x}|\theta), \tag{1}$$

i.e., como una función de θ, que es *desconocida*, que depende de los valores observados $\mathbf{x}$.

La *regla de Bayes* es una descripción general de la inversión de probabilidades: si A y E son eventos de probabilidad positiva, $\mathbb{P}(A|E)$ y $\mathbb{P}(E\,|A)$ están relacionados por

$$\mathbb{P}(A|E) = \frac{\mathbb{P}(E\,|A)\mathbb{P}(A)}{\mathbb{P}(E)} = \frac{\mathbb{P}(E\,|A)\mathbb{P}(A)}{\mathbb{P}(E\,|A)\mathbb{P}(A) + \mathbb{P}(E\,|A^c)\mathbb{P}(A^c)}.$$

En su versión continua, la regla de Bayes establece que dadas dos variables aleatorias X e Y, con distribución condicional $f_{X|Y=y}(x)$ y distribución marginal $f_Y(y)$, la distribución condicional de Y dado que $X = x$ es

$$f_{Y|X=x}(y) = \frac{f_{X|Y=y}(x)f_Y(y)}{\int f_{X|Y=y}(x)f_Y(y)dy}.$$

1.1. Distribuciones *a priori* y *a posteriori*

Desde el punto de vista probabilístico el teorema de inversión es bastante natural. Bayes y Laplace fueron más allá y consideraron que la incerteza sobre el parámetro desconocido de un modelo paramétrico puede modelarse mediante una distribución de probabilidad sobre el espacio paramétrico.

La esencia del enfoque Bayesiano consiste en que el parámetro desconocido, θ, se considera como *variable aleatoria* con cierta función densidad de probabilidades

$$\pi_\theta(t), \qquad t \in \Theta.$$

La densidad $\pi_\theta(t)$ se llama densidad *a priori*, o sea, dada *antes* del experimento. El enfoque Bayesiano supone que el parámetro desconocido θ se ha escogido aleatoriamente de la distribución cuya densidad es $\pi_\theta(t)$.

Definición 1.1. Un modelo estadístico Bayesiano está hecho de un modelo paramétrico $\mathcal{F} = \{f(x|t) : t \in \Theta\}$ para las observaciones y una distribución de probabilidad *a priori* $\pi_\theta(t)$ sobre el espacio paramétrico Θ.

Nota Bene. En un modelo Bayesiano, la "densidad" muestral $f(x|t)$, $t \in \Theta$, es la "densidad" condicional de la variable aleatoria X dado que $\theta = t$.

Dado un modelo Bayesiano podemos construir varias distribuciones, a saber:

1. La distribución *conjunta* del parámetro θ y la muestra aleatoria $\mathbf{X} = (X_1, \dots, X_n)$:

$$f_{\theta,\mathbf{X}}(t, \mathbf{x}) = f(\mathbf{x}|t)\pi_\theta(t) = \left(\prod_{i=1}^{n} f(x_i|t)\right)\pi_\theta(t). \tag{2}$$

2. La distribución *marginal* de la muestra aleatoria $\mathbf{X} = (X_1, \dots X_n)$:

$$f_{\mathbf{X}}(\mathbf{x}) = \int_\Theta f_{\theta,\mathbf{X}}(t, \mathbf{x})dt = \int_\Theta f(\mathbf{x}|t)\pi_\theta(t)dt. \tag{3}$$

3. La distribución *a posteriori* (o sea, *después del experimento*) de la variable aleatoria θ, obtenida mediante la fórmula de Bayes:

$$\pi(t|\mathbf{x}) = \frac{f_{\theta,\mathbf{X}}(t, \mathbf{x})}{\int_\Theta f_{\theta,\mathbf{X}}(t, \mathbf{x})dt} = \frac{f(\mathbf{x}|t)\pi_\theta(t)}{\int_\Theta f(\mathbf{x}|t)\pi_\theta(t)dt}. \tag{4}$$

Nota Bene. Si el parámetro θ es una variable aleatoria discreta, la "densidad" a priori $\pi_\theta(t)$ debe interpretarse como la función de probabilidades y las expresiones del tipo $\int dt$ deben reemplazarse por expresiones del tipo $\sum_t$.

Ejemplo 1.2 (Bayes (1764))**.** Se echa a rodar una bola de billar B_1 sobre una línea de longitud 1, con probabilidad uniforme de que se detenga en cualquier lugar. Se detiene en θ. Una segunda bola B_2 se echa a rodar 5 veces bajo las mismas condiciones que la primera y X denota la cantidad de veces que la bola B_2 se detuvo a la izquierda de donde lo hizo B_1. Dado que $X = x$, *¿qué se puede inferir sobre θ?*

El problema consiste en hallar la distribución *a posteriori* de θ dado que $X = x$, cuando la distribución *a priori* de θ es uniforme sobre $(0, 1)$ y $X \sim$ Binomial$(5, \theta)$. Puesto que

$$f(x|t) = \binom{5}{x} t^x (1-t)^{5-x} \qquad \text{y} \qquad \pi_\theta(t) = \mathbf{1}\{t \in (0, 1)\},$$

la distribución conjunta del parámetro θ y la variable aleatoria X es

$$f_{\theta,X}(t,x) = \binom{5}{x} t^x (1-t)^{5-x} \mathbf{1}\{t \in (0,1)\}$$

y la distribución marginal de la variable X es

$$\begin{aligned} f_X(x) &= \int_0^1 \binom{5}{x} t^x (1-t)^{5-x} dt = \binom{5}{x} \int_0^1 t^x (1-t)^{5-x} dt = \binom{5}{x} \frac{\Gamma(x+1)\Gamma(6-x)}{\Gamma(7)} \\ &= \frac{5!}{x!(5-x)!} \frac{x!(5-x!)}{6!} = \frac{1}{6}, \qquad x = 0, 1, \dots, 5 \end{aligned}$$

(En palabras, los 6 posibles valores de X son igualmente probables.)

De lo anterior se deduce que la distribución a posteriori de θ dado que $X = x$

$$\pi(t|x) = 6\binom{5}{x} t^x (1-t)^{5-x} \mathbf{1}\{t \in (0,1)\},$$

i.e., la distribución de θ condicional a que $X = x$ es la distribución $\beta(x+1, 6-x)$. □

Ejemplo 1.3 (Laplace (1773))**.** En una urna hay 12 bolas blancas y negras. Si la primer bola extraída es blanca, ¿cuál es la probabilidad de que la proporción θ de bolas blancas sea 2/3? Asumiendo *a priori* que las cantidades 2 a 11 de bolas blancas son igualmente probables, i.e., que θ es equiprobable sobre $\{2/12, \dots, 11/12\}$. La distribución a posteriori de θ se deduce usando el teorema de Bayes:

$$\pi(2/3|\text{datos}) = \frac{(2/3)(1/10)}{\sum_{p=2/12}^{11/12} p(1/10)} = \frac{(2/3)}{\sum_{n=2}^{11} n/12} = \frac{8}{(11 \times 12)/2 - 1} = \frac{8}{65}.$$

□

Principio de verosimilitud. La fórmula de Bayes (4) puede leerse del siguiente modo: observado que la muestra aleatoria $\mathbf{X}$ arrojó los valores $\mathbf{x}$, la distribución a posteriori de θ es proporcional a la función de verosimilitud $L(t|\mathbf{x}) = f(\mathbf{x}|t)$ multiplicada por la distribución a priori de θ. En símbolos

$$\pi(t|\mathbf{x}) \propto L(t|\mathbf{x})\pi_\theta(t).$$

Esto significa que la información sobre la variable θ que viene en una muestra $\mathbf{x}$ está completamente contenida en la función de verosimilitud $L(t|\mathbf{x})$. Más aún, cuando $\mathbf{x}_1$ y $\mathbf{x}_2$ son dos observaciones que dependen del mismo parámetro θ y existe una constante c que satisface

$$L_1(t|\mathbf{x}_1) = cL_2(t|\mathbf{x}_2)$$

para cada $t \in \Theta$, entonces $\mathbf{x}_1$ y $\mathbf{x}_2$ tienen la misma información sobre θ y deben conducir a inferencias idénticas. Esto es así porque el análisis Bayesiano se basa completamente en la distribución *a posteriori* $\pi(t|\mathbf{x})$ que depende de $\mathbf{x}$ solo a través de $L(t|\mathbf{x})$. □

Ejemplo 1.4. Trabajando sobre el ranking de una serie televisiva un investigador encontró 9 espectadores que la miran y 3 que no la miran. Si no se dispone de más información sobre el experimento, se pueden proponer al menos dos modelos. Si $\theta \in (0,1)$ representa la proporción de los espectadores que mira la serie:

(1) El investigador encuestó a 12 personas y por lo tanto observó $X \sim \text{Binomial}(12, \theta)$ con $X = 9$.

(2) El investigador encuestó Y personas hasta que encontró 3 que no miraban la serie y por lo tanto observó $Y \sim \text{Pascal}(3, 1-\theta)$ con $Y = 12$.

El punto importante es que, en cualquiera de los dos modelos, la verosimilitud es proporcional a

$$\theta^3(1-\theta)^9.$$

Por lo tanto, el principio de verosimilitud implica que la inferencia sobre θ debe ser idéntica para ambos modelos. □

1.2. Distribuciones predictivas

Sea $\mathbf{X} = (X_1, \ldots, X_n)$ una muestra aleatoria de una distribución indexada por θ. Se observa que $\mathbf{X} = \mathbf{x}$ y se quiere *predecir* una el comportamiento de una nueva observación $Y \sim g(y|\theta)$, donde Y es una variable aleatoria que depende del mismo parámetro θ. En el contexto probabilístico *predecir* significa contestar preguntas del tipo: ¿con qué probabilidad se observaran valores en un intervalo dado? En otras palabras ¿cuál será la distribución de la nueva observación Y?

Este problema se puede resolver usando la fórmula de probabilidad total. Dado que se observó $\mathbf{X} = \mathbf{x}$, *la función densidad predictiva (o incondicional)* de la nueva observación Y será

$$g(y|\mathbf{x}) = \int g(y|t)\pi(t|\mathbf{x})dt. \tag{5}$$

El primer factor del integrando que aparece en (5) corresponde a las densidades de la variable aleatoria Y condicionadas al conocimiento de que $\theta = t$. El segundo factor corresponde a la densidad a posteriori del parámetro aleatorio θ.

Si tuviésemos la capacidad de observar qué valor arrojó la variable θ y observáramos que $\theta = t$, la predicción de Y quedaría determinada por la *densidad condicional* $g(y|t)$. Sin embargo, la hipótesis fundamental de este enfoque es que el parámetro θ no puede ser observado y lo único que podemos observar es la muestra aleatoria $\mathbf{X}$. El calificativo de *incondicional* que se le otorga a la densidad $g(y|\mathbf{x})$ obtenida en (5) está puesto para destacar que su construcción no utiliza observaciones del parámetro θ.

Ejemplo 1.5 (Bayes (1764) Continuación.)**.** Supongamos ahora que la bola B_2 se detuvo exactamente 3 veces a la izquierda de donde lo hizo la bola B_1, *¿cuál es la probabilidad p de que al echar a rodar una tercera bola de billar B_3 también se detenga a la izquierda de donde se detuvo B_1?*

Sea $Y \sim \text{Bernoulli}(\theta)$ la variable aleatoria que vale 1 si la bola B_3 se detiene a la izquierda de donde se detuvo B_1 y 0 en caso contrario. Para calcular p usamos la distribución predictiva:

$$p = \mathbb{P}(Y = 1|X = 3) = \int_0^1 \mathbb{P}(Y = 1|t)\pi(t|3)dt = \int_0^1 t\pi(t|3) = \mathbb{E}[\theta|X = 3].$$

Como $\theta|X = 3 \sim \beta(4, 2)$, resulta que $p = 4/6$. □

1.3. Estimadores Bayesianos

1. **Estimación bayesiana por esperanza condicional.** En el contexto Bayesiano θ es una variable aleatoria. Entre todas las funciones (de la muestra aleatoria $\mathbf{X}$) $\hat{\theta} = \varphi(\mathbf{X})$ la mejor estimación para θ (desde el punto de vista de minimizar el error cuadrático medio $\mathbb{E}[(\theta - \varphi(\mathbf{X}))^2]$) es la esperanza condicional $\mathbb{E}[\theta|\mathbf{X}]$:

$$\hat{\theta}(\mathbf{X}) = \mathbb{E}[\theta|\mathbf{X}] = \int t\pi(t|\mathbf{X})dt. \tag{6}$$

2. **Estimación bayesiana por máximo a posteriori.** Otro estimador, de uso frecuente, es el llamado *máximo a posteriori (o moda)* definido por

$$\hat{\theta}_{map}(\mathbf{X}) := \arg\max_{t\in\Theta} \pi(t|\mathbf{X}). \tag{7}$$

Ejemplo 1.6 (Bayes (1764) Continuación.)**.** Supongamos ahora que la bola B_2 se detuvo exactamente 3 veces a la izquierda de donde lo hizo la bola B_1. En tal caso

$$\hat{\theta}(3) = \mathbb{E}[\theta|X = 3] = \frac{4}{6}$$

y

$$\hat{\theta}_{map}(3) = \arg\max_{t\in(0,1)} 6\binom{5}{3}t^3(1-t)^2 = \arg\max_{t\in(0,1)} t^3(1-t)^2.$$

Como el logaritmo es una función creciente, el argumento que maximiza a la función $t^3(1-t)^2$ coincide con el argumento maximizador de la función $\psi(t) = \log(t^3(1-t)^2) = 3\log(t) + 2\log(1-t)$. Observando que

$$0 = \frac{d}{dt}\psi(t) = \frac{3}{t} - \frac{2}{1-t} \iff 3(1-t) - 2t = 0 \iff t = \frac{3}{5},$$

se puede deducir que

$$\hat{\theta}_{map}(3) = \frac{3}{5}.$$

□

1.4. Estimación por intervalo para parámetro continuo

Dada la muestra aleatoria $\mathbf{X}$ se desea construir intervalos (acotados) que capturen casi toda la variabilidad del parámetro aleatorio θ. Si el intervalo $[a, b]$ es tal que

$$\mathbb{P}(\theta \in [a, b]|\mathbf{X}) = 1 - \alpha, \tag{8}$$

será llamado *intervalo estimador de nivel* $1-\alpha$. En la práctica, los valores de α son pequeños: 0.1 o 0.05 o 0.01. En general, los valores de a y b dependerán de los valores de la muestra aleatoria $\mathbf{x}$. Dado que $\mathbf{X} = \mathbf{x}$, los intervalos estimadores de nivel $1-\alpha$ se obtienen resolviendo la siguiente ecuación de las variables a y b:

$$\int_a^b \pi(t|\mathbf{x})dt = 1 - \alpha. \tag{9}$$

De todas las soluciones posibles de la ecuación (9) se prefieren aquellas que producen intervalos de longitud lo más pequeña posible.

Una solución particular de la ecuación (9) puede obtenerse mediante el siguiente razonamiento: como la distribución a posteriori del parámetro θ está centrada alrededor de su esperanza, $\hat{\theta}(\mathbf{x}) := \mathbb{E}[\theta|\mathbf{X} = \mathbf{x}]$, y no puede desviarse demasiado de allí, los intervalos que la contengan deben ser relativamente pequeños. Esto sugiere la siguiente construcción: dividir a la mitad el nivel y tratar de capturar cada una de las mitades a izquierda y a derecha de $\hat{\theta}(\mathbf{x})$. En otras palabras, se trata de resolver las siguientes ecuaciones:

$$\int_a^{\hat{\theta}(\mathbf{x})} \pi(t|\mathbf{x})dt = \frac{1-\alpha}{2}, \qquad \int_{\hat{\theta}(\mathbf{x})}^b \pi(t|\mathbf{x})dt = \frac{1-\alpha}{2}. \tag{10}$$

Ejemplo 1.7. Se considera el siguiente modelo Bayesiano: $X \sim \mathcal{N}(\theta, 1)$ con distribución a priori $\theta \sim \mathcal{N}(0, 10)$. Sobre la base de una muestra de tamaño 1 de X se quiere determinar un intervalo de nivel $1-\alpha$ para la variable θ.

Dado que $X = x$ tenemos que

$$\pi(t|x) \propto L(\theta|x)\pi_\theta(t) \propto \exp\left(-\frac{(x-t)^2}{2} - \frac{t^2}{20}\right) \propto \exp\left(-\frac{11}{20}\left(t - \frac{10x}{11}\right)^2\right)$$

y por lo tanto $\theta|X = x \sim \mathcal{N}\left(\frac{10x}{11}, \frac{10}{11}\right)$. Como la variable

$$Z = \frac{(\theta|X = x) - (10x/11)}{\sqrt{10/11}} \sim \mathcal{N}(0,1)$$

tenemos que $\mathbb{P}\left(|Z| < z_{1-\alpha/2}\right) = 1-\alpha$ y de allí se deduce dado que $X = x$ el intervalo

$$\left[\frac{10x}{11} - z_{1-\alpha/2}\sqrt{\frac{10}{11}}, \frac{10x}{11} + z_{1-\alpha/2}\sqrt{\frac{10}{11}}\right]$$

es un intervalo estimador de nivel $1-\alpha$. □

1.5. Sobre la distribución a priori uniforme.

Cuando el parámetro θ tiene distribución a priori $\mathcal{U}[a,b]$, esto es $\pi_\theta(t) = \frac{1}{b-a}\mathbf{1}\{t \in [a,b]\}$ el enfoque Bayesiano se simplifica abruptamente.

La fórmula de Bayes para la distribución a posteriori (4) adopta la forma

$$\pi(t|\mathbf{x}) = \frac{L(t|\mathbf{x})\frac{1}{b-a}\mathbf{1}\{t \in [a,b]\}}{\int L(t|\mathbf{x})\frac{1}{b-a}\mathbf{1}\{t \in [a,b]\}dt} = \frac{L(t|\mathbf{x})\mathbf{1}\{t \in [a,b]\}}{\int_a^b L(t|\mathbf{x})dt}. \tag{11}$$

En palabras, si la distribución a priori del parámetro es uniforme, la densidad de su distribución a posteriori es proporcional a la función de verosimilitud: $\pi(t|\mathbf{x}) \propto L(t|\mathbf{x})$.

Nota Bene. En cierto sentido, que puede precisarse, la distribución $\mathcal{U}[a,b]$ es la *menos informativa* entre todas las distribuciones continuas a valores en $[a,b]$.

En *teoría de la información* la indeterminación de una variable aleatoria X se mide con la *entropía* definida por $H(X) := \mathbb{E}[-\log f(X)]$, donde $f(x)$ es la densidad de probabilidades de la variable aleatoria X. En otros términos

$$H(X) := -\int f(x)\log f(x)dx. \tag{12}$$

Teorema 1.8. *Entre todas las variables aleatorias continuas a valores en* $[a, b]$ *la que maximiza la entropía es la* $\mathcal{U}[a, b]$.

Demostración. No se pierde generalidad si se supone que $[a, b] = [0, 1]$. Si $X \sim \mathcal{U}[0, 1]$, entonces

$$H(X) = -\int_0^1 1 \log(1) dx = 0.$$

El resultado se obtiene mostrando que si X es una variable aleatoria continua a valores en el $[0, 1]$, entonces $H(X) \leq 0$.

Es fácil ver que para todo $x > 0$ vale la desigualdad

$$\log(x) \leq x - 1 \tag{13}$$

Poniendo $x = \frac{1}{u}$, $u > 0$, en la desigualdad (13) se obtiene

$$-\log u = \log\left(\frac{1}{u}\right) \leq \frac{1}{u} - 1 \tag{14}$$

La desigualdad (14) se usa para obtener

$$H(X) = -\int_0^1 f(x) \log f(x) dx \leq \int_0^1 f(x) \left(\frac{1}{f(x)} - 1\right) dx = \int_0^1 1 dx - \int_0^1 f(x) dx = 0.$$

□

Comentario Bibliográfico. Una exposición elemental de la noción de entropía y de las distribuciones menos informativas puede leerse en Pugachev, V.S., (1973). *Introducción a la Teoría de Probabilidades*, Mir, Moscu.

Enfoque Bayesiano generalizado. Si la función de verosimilitud $L(t|\mathbf{x})$ es integrable, i.e., $0 < \int_{-\infty}^{\infty} L(t|\mathbf{x}) dt < \infty$, la expresión

$$\pi(t|\mathbf{x}) := \frac{L(t|\mathbf{x})}{\int_{-\infty}^{\infty} L(t|\mathbf{x}) dt} \tag{15}$$

define una densidad de probabilidades en $\mathbb{R}$. Por abuso del lenguaje, algunos autores suelen llamarla la densidad a posteriori correspondiente a la distribución a priori "*uniforme sobre la recta*"[1] No hay ningún problema en utilizar este enfoque siempre que no se pierda de vista que no existe ninguna distribución uniforme sobre regiones de longitud infinita. El enfoque que postula una densidad a posteriori de la forma (15) será llamado *Bayesiano generalizado*.

2. Ejemplos

2.1. Las distribuciones β y el problema del "control de calidad"

Control de calidad. La calidad de un proceso de producción puede medirse por el porcentaje, $100\,\theta\,\%$, de artículos defectuosos producidos. Cada artículo producido tiene asociada

[1]**Nota histórica:** la denominación para esta a priori impropia se debe a Laplace.

una variable aleatoria de Bernoulli, $X \sim \text{Bernoulli}(\theta)$, cuyo parámetro θ denota la probabilidad de que el artículo sea defectuoso.

El punto de partida del enfoque Bayesiano es la distribución a priori del parámetro. Supongamos que, a priori, $\theta \sim \mathcal{U}(0,1)$. Se observa una muestra aleatoria $\mathbf{X} = (X_1, \dots, X_n)$ y usando la fórmula de Bayes (4) se obtiene la densidad, $\pi(t|\mathbf{x})$, de la distribución a posteriori de θ dado que $\mathbf{X} = \mathbf{x}$. Cuando la densidad a priori es uniforme la densidad a posteriori es proporcional a la verosimilitud. Por lo tanto,

$$\pi(t|\mathbf{x}) \propto L(t|\mathbf{x}) = t^{k(\mathbf{x})}(1-t)^{n-k(\mathbf{x})}\mathbf{1}\{t \in (0,1)\}, \tag{16}$$

donde $k(\mathbf{x}) = \sum_{i=1}^n x_i$. De la identidad (16) se concluye que $\theta|\mathbf{X} = \mathbf{x}$ tiene una distribución beta de parámetros $k(\mathbf{x})+1$ y $n-k(\mathbf{x})+1$. En consecuencia la constante de proporcionalidad será

$$\frac{\Gamma(n+2)}{\Gamma(k(\mathbf{x})+1)\Gamma(n-k(\mathbf{x})+1)} = \frac{(n+1)!}{k(\mathbf{x})!(n-k(\mathbf{x}))!} = (n+1)\binom{n}{k(\mathbf{x})}. \tag{17}$$

Conclusión. Sea $\mathbf{X} = (X_1, \dots, X_n)$ una muestra aleatoria de volumen n correspondiente a una variable aleatoria $X \sim \text{Bernoulli}(\theta)$. Si la distribución a priori del parámetro θ es uniforme sobre el intervalo $(0,1)$ y se observa que $\mathbf{X} = \mathbf{x}$, entonces la distribución a posteriori (del parámetro θ) es una $\beta(k+1, n-k+1)$, donde k es la cantidad de éxitos observados. En otras palabras, la densidad de $\theta|\mathbf{X} = \mathbf{x}$ es

$$\pi(t|\mathbf{x}) = (n+1)\binom{n}{k} t^k (1-t)^{n-k}\mathbf{1}\{t \in (0,1)\}, \tag{18}$$

donde $k = \sum_{i=1}^n x_i$. □

Función de probabilidad marginal. Cuál es la probabilidad de que en una muestra de volumen n se observen exactamente k artículos defectuosos. La cantidad de artículos defectuosos será $N = \sum_{i=1}^n X_i$. Dado que $\theta = t$, las variables $X_1, \dots, X_n$ serán independientes, cada una con distribución de Bernoulli(t) y en tal caso $N \sim \text{Binomial}(n, t)$

$$\mathbb{P}(N = k|t) = \binom{n}{k} t^k (1-t)^{n-k}, \qquad k = 0, 1, \dots, n \tag{19}$$

Por lo tanto, condicionando sobre $\theta = t$ y usando la fórmula de probabilidad total, obtenemos que

$$\begin{aligned}
\mathbb{P}(N = k) &= \int_0^1 \mathbb{P}(N = k|t)\pi_\theta(t)dt = \int_0^1 \binom{n}{k} t^k (1-t)^{n-k} dt \\
&= \binom{n}{k} \int_0^1 t^k (1-t)^{n-k} dt = \binom{n}{k} \frac{k!(n-k)!}{(n+1)!} \\
&= \frac{1}{n+1} \qquad k = 0, 1, \dots, n
\end{aligned} \tag{20}$$

En otras palabras, los $n+1$ valores posibles de N son igualmente probables.

Función de probabilidad predictiva Supongamos ahora que en una muestra de volumen n se observaron exactamente k artículos defectuosos. Cuál es la probabilidad p de que un nuevo artículo resulte defectuoso?

Para calcular p usamos la función de probabilidad predictiva obtenida en (5):

$$p = f(1|\mathbf{x}) = \int_0^1 f(1|t)\pi(t|\mathbf{x})dt = \int_0^1 t\pi(t|\mathbf{x})dx = \mathbb{E}[\theta|\mathbf{X} = \mathbf{x}] = \frac{k+1}{n+2}. \tag{21}$$

Esto es, si los primeros n artículos resultaron en k defectuosos, entonces el próximo artículo será defectuoso con probabilidad $(k+1)/(n+2)$.

De la ecuación (21) resulta una descripción alternativa del proceso de producción examinado: Hay una urna que inicialmente contiene una bola blanca y una bola negra. En cada paso se extrae al azar una bola de la urna y se la repone junto con otra del mismo color. Después de cada extracción la cantidad de bolas del color extraído aumenta una unidad y la cantidad de bolas del color opuesto se mantiene constante. Si de las primeras n bolas elegidas, k fueron blancas, entonces en la urna al momento de la $n+1$-ésima extracción hay $k+1$ blancas y $n-k+1$ negras, y por lo tanto la siguiente bola será blanca con probabilidad $(k+1)/(n+2)$. Identificando la extracción de una bola blanca con un artículo defectuoso, tenemos una descripción alternativa del modelo original. Esté último se llama *modelo de urna de Polya*.

Estimadores Bayesianos

1. Utilizando la esperanza condicional de $\theta|\mathbf{X} = \mathbf{x}$ obtenemos la siguiente estimación

$$\hat{\theta}(\mathbf{x}) = \mathbb{E}[\theta|\mathbf{X} = \mathbf{x}] = \frac{1}{n+2}\left(1 + \sum_{i=1}^n x_i\right). \tag{22}$$

2. El estimador máximo a posteriori se obtiene observando que

$$\begin{aligned}\hat{\theta}_{map}(\mathbf{x}) &= \arg \max_{t\in(0,1)} (n+1)\binom{n}{k} t^k(1-t)^{n-k} = \arg \max_{t\in(0,1)} t^k(1-t)^{n-k} \\ &= \arg \max_{t\in(0,1)} \log t^k(1-t)^{n-k} = \arg \max_{t\in(0,1)} (k\log t + (n-k)\log(1-t)) \\ &= \frac{k}{n},\end{aligned}$$

donde $k = \sum_{i=1}^n x_i$. Por lo tanto,

$$\hat{\theta}_{map}(\mathbf{x}) = \bar{x}. \tag{23}$$

Nota Bene. Notar que

$$\hat{\theta}(\mathbf{x}) = \frac{n}{n+2}\bar{x} + \frac{1}{n+2} = \frac{n}{n+2}\bar{x} + \frac{2}{n+2}\mathbb{E}[\mathcal{U}(0,1)],$$

donde $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$.

Estimación por intervalo Se quiere construir un intervalo estimador (de nivel $1-\alpha$) para θ sabiendo que en una muestra de volumen n se observaron k artículos defectuosos.

En este caso la ecuación (9) adopta la forma

$$1-\alpha = \int_a^b \frac{(n+1)!}{k!(n-k)!} t^k (1-t)^{n-k} dt. \tag{24}$$

El problema equivale a encontrar las raíces de un polinomio de grado $n+1$ en las variables a y b y no hay métodos generales para encontrarlas. El problema se puede resolver mediante alguna técnica de cálculo numérico para aproximar raíces de polinomios implementada en un computador. Para $3 \leq n+1 \leq 4$ pueden utilizarse las fórmulas de Tartaglia para resolver ecuaciones de tercer y cuarto grado. Estas fórmulas pueden consultarse en el Tomo 1 del *Análisis matemático* de Rey Pastor.

Cuando $k=0$ o $k=n$ la ecuación (24) se puede resolver "a mano": si $k=0$ la ecuación (24) adopta la forma

$$\begin{aligned}
1-\alpha &= \int_a^b (n+1)(1-t)^n dt = (n+1)\left(-\frac{(1-t)^{n+1}}{n+1}\bigg|_a^b\right) \\
&= (n+1)\left(\frac{(1-a)^{n+1}}{n+1} - \frac{(1-b)^{n+1}}{n+1}\right) \\
&= (1-a)^{n+1} - (1-b)^{n+1}.
\end{aligned}$$

Fijado un valor "razonable" de a se puede despejar el valor de b

$$b = 1 - \sqrt[n+1]{(1-a)^{n+1} - (1-\alpha)}, \qquad 0 \leq a \leq 1 - \sqrt[n+1]{1-\alpha} \tag{25}$$

Hemos visto que, para $k=0$ el máximo a posteriori es 0, poniendo $a=0$ se obtiene $b = 1 - \sqrt[n+1]{\alpha}$. Por lo tanto, el intervalo

$$\left[0, 1 - \sqrt[n+1]{\alpha}\right]$$

es un intervalo estimador de nivel $1-\alpha$.

Ejemplo 2.1. Sea X una variable aleatoria Bernoulli de parámetro θ. A priori se supone que la distribución de θ es uniforme sobre el intervalo $[0,1]$. Supongamos que una muestra aleatoria de volumen $n=20$ arroja los siguientes resultados:

$$\mathbf{x} = (0,0,1,0,1,0,1,0,0,1,0,0,1,1,1,1,1,1,0,1)$$

Distribución a posteriori. Como la cantidad de éxitos observados es $k=11$, tenemos que $\theta|\mathbf{X}=\mathbf{x} \sim \beta(12,10)$. En otras palabras, la densidad a posteriori es de la forma

$$\pi(t|\mathbf{x}) = \frac{21!}{11!9!} t^{11}(1-t)^9 \mathbf{1}\{t \in [0,1]\}. \tag{26}$$

En la Figura 1 se muestran los gráficos de la distribución a priori de θ y de la distribución a posteriori de θ vista la muestra.

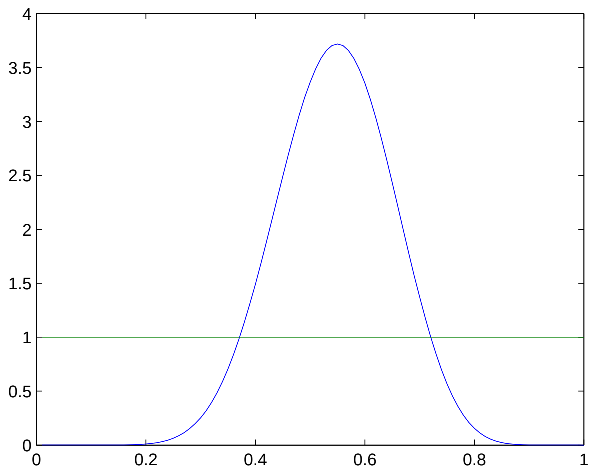

Figura 1: Gráficos de las densidades a priori y a posteriori: en verde el gráfico de la densidad de la distribución $\mathcal{U}[0,1]$ y en azul el de la distribución $\beta(12,10)$.

Predicción. ¿Cuál es la probabilidad de que en una nueva muestra de volumen 5 resulten exactamente 2 éxitos?

En primer lugar hay que observar que dado que $\theta = t$ la cantidad de éxitos N en una muestra de volumen 5 tiene distribución Binomial$(5,t)$. Por lo tanto,

$$\mathbb{P}(N = 2|t) = \binom{5}{2} t^2(1-t)^3 = 10t^2(1-t)^3.$$

Como la densidad a posteriori de θ resultó ser

$$\pi(t|\mathbf{x}) = \frac{21!}{11!9!} t^{11}(1-t)^9 \mathbf{1}\{t \in [0,1]\},$$

de la fórmula de probabilidad total se deduce que

$$\begin{aligned}
\mathbb{P}(N = 2|\mathbf{x}) &= \int_0^1 \mathbb{P}(N = 2|t) f(t|\mathbf{x}) dt = \int_0^1 10t^2(1-t)^3 \frac{21!}{11!9!} t^{11}(1-t)^9 dt \\
&= 10\frac{21!}{11!9!} \int_0^1 t^{13}(1-t)^{12} dt = 10\frac{21!}{11!9!} \frac{13!12!}{26!} = \frac{6}{23} = 0.26\ldots
\end{aligned}$$

Estimadores Bayesianos

1. Esperanza condicional:

$$\hat{\theta} = \mathbb{E}[\theta|\mathbf{X} = \mathbf{x}] = \frac{12}{22} = \frac{6}{11} = 0.5454\ldots.$$

2. Máximo a posteriori:

$$\hat{\theta}_{map} = \bar{x} = \frac{11}{20} = 0.55.$$

Estimación por intervalo Para construir un intervalo $[a, b]$, de nivel 0.95, para θ podemos resolver las siguientes ecuaciones

$$\int_0^a \frac{21!}{11!9!} t^{11}(1-t)^9 dt = 0.025, \qquad \int_0^b \frac{21!}{11!9!} t^{11}(1-t)^9 dt = 0.975.$$

Utilizando una herramienta de cálculo obtenemos que $a = 0.3402$ y $b = 0.7429$. $\square$

2.2. Normales de varianza conocida y media normal

Sea $\mathbf{X} = (X_1, \dots, X_n)$ una muestra aleatoria de una familia normal $\mathcal{N}(\theta, \sigma^2)$, con σ^2 conocido. Supongamos que la distribución a priori del parámetro θ es una normal $\mathcal{N}(\mu, \rho^2)$

Distribución a posteriori. Por definición, ver (4), la densidad a posteriori de θ, dado que $\mathbf{X} = \mathbf{x}$, queda caracterizada por la relación de proporcionalidad $\pi(t|\mathbf{x}) \propto L(t|\mathbf{x})\pi_\theta(t)$, donde $L(t|\mathbf{x})$ es la función de verosimilitud y $\pi_\theta(t)$ la densidad a priori de θ.

Primero calculamos la función de verosimilitud. De las igualdades

$$\begin{aligned} L(\mu, \sigma^2|\mathbf{x}) &= \prod_{i=1}^n f(x_i|\mu, \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right) \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2\right) \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left(-\frac{\sum_{i=1}^n (x_i-\bar{x})^2}{2\sigma^2}\right) \exp\left(-\frac{n(\bar{x}-\mu)^2}{2\sigma^2}\right), \qquad (27) \end{aligned}$$

donde $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$,[2] se deduce que

$$L(t|\mathbf{x}) \propto \exp\left(-\frac{n(\bar{x}-t)^2}{2\sigma^2}\right). \qquad (28)$$

Por hipótesis, $\theta \sim \mathcal{N}(\mu, \rho^2)$. En consecuencia,

$$\pi_\theta(t) \propto \exp\left(-\frac{(t-\mu)^2}{2\rho^2}\right) \qquad (29)$$

De (28) y (29), la densidad a posteriori satisface

$$\pi(t|\mathbf{x}) \propto \exp\left(-\left[\frac{n(\bar{x}-t)^2}{2\sigma^2} + \frac{(t-\mu)^2}{2\rho^2}\right]\right). \qquad (30)$$

Completando cuadrados respecto de t se obtiene

$$\frac{n(\bar{x}-t)^2}{2\sigma^2} + \frac{(t-\mu)^2}{2\rho^2} = \frac{n\rho^2+\sigma^2}{2\sigma^2\rho^2}\left(t - \frac{n\rho^2\bar{x}+\sigma^2\mu}{n\rho^2+\sigma^2}\right)^2 + \text{otras cosas} \qquad (31)$$

[2] La última igualdad de (27) se obtiene observando que

$$\sum_{i=1}^n (x_i-\mu)^2 = \sum_{i=1}^n (x_i-\bar{x})^2 + n(\bar{x}-\mu)^2.$$

donde "otras cosas" son expresiones que no dependen de t. En consecuencia,

$$\pi(t|\mathbf{x}) \propto \exp\left(-\frac{n\rho^2+\sigma^2}{2\sigma^2\rho^2}\left(t-\frac{n\rho^2\bar{x}+\sigma^2\mu}{n\rho^2+\sigma^2}\right)^2\right). \tag{32}$$

Por lo tanto, la distribución a posteriori de θ dado que $\mathbf{X}=\mathbf{x}$ es una normal

$$\mathcal{N}\left(\frac{n\rho^2\bar{x}+\sigma^2\mu}{n\rho^2+\sigma^2}, \frac{\sigma^2\rho^2}{n\rho^2+\sigma^2}\right). \tag{33}$$

Función densidad predictiva. Comenzamos calculando el producto de la densidad condicional de X dado que $\theta=t$ por la densidad a posteriori de θ dado que $\mathbf{X}=\mathbf{x}$:

$$\begin{aligned} f(x|t)\pi(t|\mathbf{x}) &= \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-t)^2}{2\sigma^2}\right)\frac{1}{\sqrt{2\pi}\rho_*}\exp\left(-\frac{(t-\mu_*)^2}{2\rho_*^2}\right) \\ &= \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{2\pi}\rho_*\sigma}\exp\left(-\left[\frac{(x-t)^2}{2\sigma^2}+\frac{(t-\mu_*)^2}{2\rho_*^2}\right]\right), \end{aligned} \tag{34}$$

donde μ_* y ρ_*^2 son la media y la varianza de la distribución a posteriori de θ dado que $\mathbf{X}=\mathbf{x}$

$$\mu_* = \frac{n\rho^2\bar{x}+\sigma^2\mu}{n\rho^2+\sigma^2} \qquad \text{y} \qquad \rho_*^2 = \frac{\sigma^2\rho^2}{n\rho^2+\sigma^2} \tag{35}$$

Con un poco de paciencia, puede verse que

$$\frac{(x-t)^2}{2\sigma^2}+\frac{(t-\mu_*)^2}{2\rho_*^2} = \frac{\rho_*^2+\sigma^2}{2\sigma^2\rho_*^2}\left(t-\frac{\rho_*^2x+\sigma^2\mu_*}{\rho_*^2+\sigma^2}\right)^2+\frac{(x-\mu_*)^2}{2(\rho_*^2+\sigma^2)} \tag{36}$$

En consecuencia,

$$\begin{aligned} &f(x|t)\pi(t|\mathbf{x}) \\ &= \frac{1}{\sqrt{2\pi}\sigma}\frac{1}{\sqrt{2\pi}\rho_*}\exp\left(-\left[\frac{\rho_*^2+\sigma^2}{2\sigma^2\rho_*^2}\left(t-\frac{\rho_*^2x+\sigma^2\mu_*}{\rho_*^2+\sigma^2}\right)^2+\frac{(x-\mu_*)^2}{2(\rho_*^2+\sigma^2)}\right]\right) \\ &= \frac{1}{\sqrt{2\pi(\rho_*^2+\sigma^2)}}\exp\left(-\frac{(x-\mu_*)^2}{2(\rho_*^2+\sigma^2)}\right) \\ &\quad\times\frac{1}{\sqrt{2\pi\frac{\rho_*^2\sigma^2}{\rho_*^2+\sigma^2}}}\exp\left(-\frac{\rho_*^2+\sigma^2}{2\sigma^2\rho_*^2}\left(t-\frac{\rho_*^2x+\sigma^2\mu_*}{\rho_*^2+\sigma^2}\right)^2\right). \end{aligned} \tag{37}$$

Integrando respecto de t, ambos lados de identidad (37), obtenemos la expresión de la densidad predictiva

$$f(x|\mathbf{x}) = \int f(x|t)\pi(t|\mathbf{x})dt = \frac{1}{\sqrt{2\pi(\rho_*^2+\sigma^2)}}\exp\left(-\frac{(x-\mu_*)^2}{2(\rho_*^2+\sigma^2)}\right). \tag{38}$$

En otras palabras, la distribución de la variable aleatoria X dado que $\mathbf{X}=\mathbf{x}$, es una normal de media μ_* y varianza $\sigma^2+\rho_*^2$. El resultado obtenido nos permite calcular todas las probabilidades de la forma $\mathbb{P}(X\in A|\mathbf{X}=\mathbf{x})$.

Estimadores Bayesianos. En este caso, como el máximo de la normal se alcanza en la media ambos estimadores coinciden:

$$\hat{\theta} = \frac{n\rho^2\bar{x} + \sigma^2\mu}{n\rho^2 + \sigma^2}. \tag{39}$$

Nota Bene. Note que

$$\hat{\theta} = \frac{n\rho^2}{n\rho^2 + \sigma^2}\bar{x} + \frac{\sigma^2}{n\rho^2 + \sigma^2}\mu = \frac{n\rho^2}{n\rho^2 + \sigma^2}\bar{x} + \frac{\sigma^2}{n\rho^2 + \sigma^2}\mathbb{E}[\mathcal{N}(\mu, \rho^2)] \tag{40}$$

Estimación por intervalo. En lo que sigue construiremos un intervalo estimador de nivel $1-\alpha$ para θ sabiendo que $\mathbf{X} = \mathbf{x}$. Sabemos que $\theta|\mathbf{X} = \mathbf{x}$ se distribuye como una normal de media μ_* y varianza ρ_*^2. Proponiendo un intervalo centrado en la media μ_* de la forma

$$[\mu_* - \epsilon, \mu_* + \epsilon] \tag{41}$$

y usando la simetría de la normal con respecto a su media, el problema se reduce a encontrar el valor de ϵ que resuelve la ecuación siguiente

$$1 - \frac{\alpha}{2} = \mathbb{P}\left(\theta \leq \mu_* + \epsilon|\mathbf{X} = \mathbf{x}\right) = \mathbb{P}\left(\frac{\theta - \mu_*}{\rho_*} \leq \frac{\epsilon}{\rho_*}\middle| \mathbf{X} = \mathbf{x}\right) = \Phi\left(\frac{\epsilon}{\rho_*}\right). \tag{42}$$

En consecuencia,

$$\epsilon = \rho_*\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) = \sqrt{\frac{\sigma^2\rho^2}{n\rho^2 + \sigma^2}}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) = \frac{\sigma\rho}{\sqrt{n\rho^2 + \sigma^2}}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \tag{43}$$

Por lo tanto, el intervalo

$$\left[\frac{n\rho^2\bar{x} + \sigma^2\mu}{n\rho^2 + \sigma^2} - \frac{\sigma\rho}{\sqrt{n\rho^2 + \sigma^2}}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right), \frac{n\rho^2\bar{x} + \sigma^2\mu}{n\rho^2 + \sigma^2} + \frac{\sigma\rho}{\sqrt{n\rho^2 + \sigma^2}}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right] \tag{44}$$

es un intervalo estimador de nivel $1-\alpha$ para θ sabiendo que $\mathbf{X} = \mathbf{x}$. Note que la longitud del intervalo no depende los valores arrojados por la muestra y es del orden de $\frac{1}{\sqrt{n}}$.

Curva peligrosa. Para una muestra de una $\mathcal{N}(\theta, \sigma^2)$ con distribución a priori para θ de la forma $\mathcal{N}(\mu, \rho^2)$ obtuvimos que la distribución a posteriori satisface

$$f(t|\mathbf{x}) \propto \exp\left(-\frac{n\rho^2 + \sigma^2}{2\sigma^2\rho^2}\left(t - \frac{n\rho^2\bar{x} + \sigma^2\mu}{n\rho^2 + \sigma^2}\right)^2\right). \tag{45}$$

A medida que aumentamos el valor de ρ^2 la información contenida en la distribución a priori se va "destruyendo" y la densidad a posteriori se va aproximando a la densidad de una normal de media $\bar{x}$ y varianza σ^2/n:

$$\lim_{\rho^2 \to \infty} f(t|\mathbf{x}) \propto \exp\left(-\frac{n\,(t - \bar{x})^2}{2\sigma^2}\right) \propto L_t(\mathbf{x}). \tag{46}$$

En palabras informales y poco rigurosas, si se destruye la información contenida en la distribución a priori $\mathcal{N}(\mu, \rho^2)$ mediante el procedimiento de hacer $\rho^2 \to \infty$ se obtiene una densidad de probabilidades proporcional a la verosimilitud. Vale decir, en el caso límite se obtiene el *enfoque Bayesiano generalizado*. Desde esta perspectiva, el enfoque Bayesiano generalizado puede interpretarse como una metodología orientada a destruir toda la información contenida en las distribuciones a priori del parámetro.

Ejemplo 2.2. Se tiene la siguiente muestra aleatoria de volumen $n = 10$ de una población $\mathcal{N}(\theta, 1)$

2.0135	0.9233	0.0935	0.0907	0.3909
0.3781	-1.9313	-0.8401	3.4864	-0.6258

Si, a priori, suponemos que $\theta \sim \mathcal{N}(0, 1)$, entonces la distribución a posteriori de θ es una normal, ver (33), $\mathcal{N}\left(\frac{10\bar{x}}{11}, \frac{1}{11}\right)$. Observando la muestra se obtiene que $\bar{x} = 0.3979$. Por lo tanto, la distribución a posteriori del parámetro es una normal $\mathcal{N}(\frac{3.979}{11}, \frac{1}{11})$.

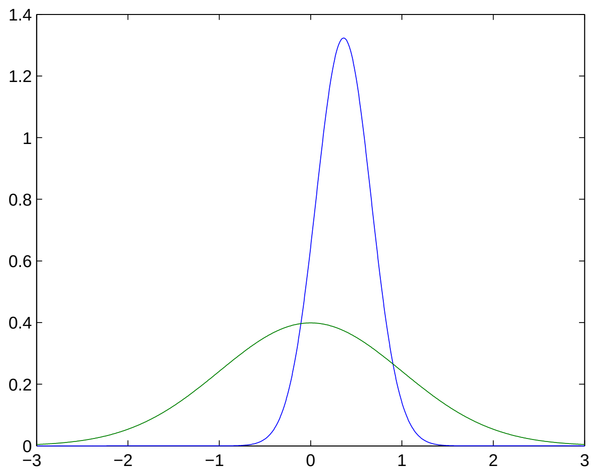

Figura 2: Gráficos de las densidades a priori (en verde) y a posteriori (en azul).

Como la moda y la media de la distribución normal coinciden, el estimador puntual Bayesiano resulta ser $\hat{\theta} = 3.979/11 = 0.3617\ldots$.

Utilizando la tabla de la normal estándar puede verse que $I = [-0.22920.9527]$ es un intervalo de nivel 0.95.

Etcétera... □

2.3. Distribuciones Poisson con a priori Gamma

Sea $\mathbf{X} = (X_1, \ldots, X_n)$ una muestra aleatoria de una distribución Poisson de parámetro θ, $\theta > 0$. Supongamos que la distribución a priori del parámetro θ es una Gamma de parámetros ν y λ. Esto es, la densidad a priori del parámetro es de la forma

$$\pi_\theta(t) \propto t^{\nu-1} e^{-\lambda t} \mathbf{1}\{t > 0\} \tag{47}$$

.

Distribución a posteriori. La densidad a posteriori de θ, dado que $\mathbf{X} = \mathbf{x}$, queda caracterizada por la relación de proporcionalidad $\pi(t|\mathbf{x}) \propto L(t|\mathbf{x})\pi_\theta(t)$, donde $L(t|\mathbf{x})$ es la función de verosimilitud y $\pi_\theta(t)$ es la densidad a priori de θ. En este caso la función de verosimilitud es de la forma

$$L(t|\mathbf{x}) \propto e^{-nt} t^{\sum_{i=1}^n x_i}. \tag{48}$$

De (47) y (48) se deduce que la densidad a posteriori de θ dado que $\mathbf{X} = \mathbf{x}$ satisface

$$\pi(t|\mathbf{x}) \propto e^{-nt} t^{\sum_{i=1}^n x_i} t^{\nu-1} e^{-\lambda t} \mathbf{1}\{t > 0\} = t^{\sum_{i=1}^n x_i + \nu - 1} e^{-(n+\lambda)t} \mathbf{1}\{t > 0\}. \tag{49}$$

Por lo tanto, la distribución a posteriori de θ dado que $\mathbf{X} = \mathbf{x}$ es una Gamma

$$\Gamma\left(\sum_{i=1}^n x_i + \nu, n + \lambda\right).$$

Estimadores Bayesianos.

1. Utilizando la esperanza condicional de $\theta|\mathbf{X} = \mathbf{x}$ obtenemos la siguiente estimación.

$$\hat{\theta} = \mathbb{E}[\theta|\mathbf{X} = \mathbf{x}] = \frac{\sum_{i=1}^n x_i + \nu}{n + \lambda} \tag{50}$$

2. La estimación por máximo a posteriori se obtiene observando que

$$\arg\max_{t>0} t^a e^{-bt} = \arg\max_{t>0} \log t^a e^{-bt} = \arg\max_{t>0}(a \log t - bt) = \frac{b}{a}.$$

Por lo tanto,

$$\hat{\theta}_{map} = \frac{\sum_{i=1}^n x_i + \nu - 1}{n + \lambda}. \tag{51}$$

Nota Bene. Notar que

$$\begin{aligned} \hat{\theta} &= \frac{\sum_{i=1}^n x_i + \nu}{n + \lambda} = \frac{n}{n + \lambda}\left(\frac{\sum_{i=1}^n x_i}{n}\right) + \frac{\lambda}{n + \lambda}\left(\frac{\nu}{\lambda}\right) \\ &= \frac{n}{n + \lambda}\bar{x} + \frac{\lambda}{n + \lambda}\mathbb{E}[\Gamma(\nu, \lambda)]. \end{aligned} \tag{52}$$

Función de probabilidad predictiva. El producto de la probabilidad condicional de X dado que $\theta = t$ por la densidad a posteriori de θ dado que $\mathbf{X} = \mathbf{x}$:

$$\begin{aligned} f(x|t)\pi(t|\mathbf{x}) &= e^{-t}\frac{t^x}{x!}\frac{(n+\lambda)^{\nu(\mathbf{x})}}{\Gamma(\nu(\mathbf{x}))} t^{\nu(\mathbf{x})-1} e^{-(n+\lambda)t} \mathbf{1}\{t > 0\} \\ &= \frac{(n+\lambda)^{\nu(\mathbf{x})}}{x!\Gamma(\nu(\mathbf{x}))} t^{\nu(\mathbf{x})+x-1} e^{-(n+\lambda+1)t} \mathbf{1}\{t > 0\}, \end{aligned} \tag{53}$$

donde $\nu(\mathbf{x}) = \sum_{i=1}^n x_i + \nu$. Integrando respecto de t ambos lados de la identidad (53), obtenemos la expresión de la función de probabilidad incondicional (o predictiva)

$$\begin{aligned} f(x|\mathbf{x}) &= \frac{(n+\lambda)^{\nu(\mathbf{x})}}{x!\Gamma(\nu(\mathbf{x}))} \int_0^\infty t^{\nu(\mathbf{x})+x-1} e^{-(n+\lambda+1)t} dt \\ &= \frac{(n+\lambda)^{\nu(\mathbf{x})}}{x!\Gamma(\nu(\mathbf{x}))} \frac{\Gamma(\nu(\mathbf{x})+x)}{(n+\lambda+1)^{\nu(\mathbf{x})+x}} \\ &= \frac{\Gamma(\nu(\mathbf{x})+x)}{\Gamma(\nu(\mathbf{x}))x!} \frac{(n+\lambda)^{\nu(\mathbf{x})}}{(n+\lambda+1)^{\nu(\mathbf{x})+x}} \\ &= \frac{\Gamma(\nu(\mathbf{x})+x)}{\Gamma(\nu(\mathbf{x}))x!} \left(\frac{1}{n+\lambda+1}\right)^x \left(\frac{n+\lambda}{n+\lambda+1}\right)^{\nu(\mathbf{x})}. \end{aligned} \tag{54}$$

Una expresión que con un poco de paciencia (o una computadora a la mano) se puede calcular para cada valor de x.

Caso $\nu \in \mathbb{N}$. En este caso la expresión para la función de probabilidad incondicional (54) adopta la forma

$$\begin{aligned} f(x|\mathbf{x}) &= \frac{(\nu(\mathbf{x})+x-1)!}{(\nu(\mathbf{x})-1)!x!} \left(\frac{1}{n+\lambda+1}\right)^x \left(\frac{n+\lambda}{n+\lambda+1}\right)^{\nu(\mathbf{x})} \\ &= \binom{\nu(\mathbf{x})+x-1}{\nu(\mathbf{x})-1} \left(\frac{1}{n+\lambda+1}\right)^x \left(\frac{n+\lambda}{n+\lambda+1}\right)^{\nu(\mathbf{x})}. \end{aligned} \tag{55}$$

La expresión (55) para la función de probabilidad condicional $f(x|\mathbf{x})$ admite la siguiente interpretación probabilística: *Dado que $\mathbf{X} = \mathbf{x}$, la probabilidad incondicional de que la variable Poisson asuma el valor x es igual a la probabilidad de que en una sucesión de ensayos Bernoulli independientes de parámetro $\frac{n+\lambda}{n+\lambda+1}$ el $\nu(\mathbf{x})$-ésimo éxito ocurra en el $(\nu(\mathbf{x})+x)$-ésimo ensayo.*

Estimación por intervalo. Dado que $\mathbf{X} = \mathbf{x}$, podemos construir un intervalo estimador de nivel $1-\alpha$ para θ observando que

$$2(n+\lambda)\theta \sim \Gamma\left(\frac{2\nu(\mathbf{x})}{2}, \frac{1}{2}\right).$$

Si además $\nu \in \mathbb{N}$, entonces

$$2(n+\lambda)\theta \sim \chi^2_{2\nu(\mathbf{x})}.$$

En tal caso,

$$\mathbb{P}\left(2(n+\lambda)\theta \in \left[\chi^2_{2\nu(\mathbf{x}),\alpha/2}, \chi^2_{2\nu(\mathbf{x}),1-\alpha/2}\right]\right) = 1-\alpha.$$

Por lo tanto, si $\nu \in \mathbb{N}$ y sabiendo que $\mathbf{X} = \mathbf{x}$ el intervalo

$$\left[\frac{\chi^2_{2\nu(\mathbf{x}),\alpha/2}}{2(n+\lambda)}, \frac{\chi^2_{2\nu(\mathbf{x}),1-\alpha/2}}{2(n+\lambda)}\right],$$

donde $\nu(\mathbf{x}) = \sum_{i=1}^n x_i + \nu$, es un intervalo estimador de nivel $1-\alpha$ para θ.

Ejemplo 2.3. La cantidad de errores de tipeo por hoja que comete una secretaria profesional puede modelarse con una distribución de Poisson de parámetro θ (¿Por qué?). A priori, se supone que el parámetro θ sigue una distribución exponencial de intensidad 1 (Esta hipótesis sobre la distribución de θ es la menos informativa si se supone que la media de la distribución es 1). Se analizan 10 hojas tipeadas por la mencionada secretaria y resulta que la cantidad de errores por página es

1	3	3	3	4	6	3	2	2	2

Si la secretaria tipea una nueva hoja, cuál es la probabilidad de que cometa como máximo un error?

Solución. Para resolver este problema utilizaremos la función de probabilidad predictiva. De acuerdo con (54), como la distribución a priori de θ es una $\text{Exp}(1) = \Gamma(1,1)$, dicha función es de la forma

$$f(x|\mathbf{x}) = \binom{\nu(\mathbf{x}) + x - 1}{\nu(\mathbf{x}) - 1}\left(\frac{1}{n+\lambda+1}\right)^{x}\left(\frac{n+\lambda}{n+\lambda+1}\right)^{\nu(\mathbf{x})} = \binom{29+x}{29}\left(\frac{1}{12}\right)^{x}\left(\frac{11}{12}\right)^{30},$$

debido a que $n = 10$, $\nu(\mathbf{x}) = \sum_{i=1}^{n} x_i + 1 = 30$ y $\lambda = 1$. Por lo tanto, la probabilidad de que la secretaria cometa como máximo un error al tipear una nueva hoja será

$$\begin{aligned} f(0|\mathbf{x}) + f(1|\mathbf{x}) &= \binom{29}{29}\left(\frac{1}{12}\right)^{0}\left(\frac{11}{12}\right)^{30} + \binom{30}{29}\left(\frac{1}{12}\right)^{1}\left(\frac{11}{12}\right)^{30} \\ &= \left(\frac{11}{12}\right)^{30}\left(1 + 30\left(\frac{1}{12}\right)\right) = \left(\frac{11}{12}\right)^{30}\left(\frac{7}{2}\right) = 0.257\ldots \end{aligned}$$

$\square$

3. Bibliografía consultada

Para redactar estas notas se consultaron los siguientes libros:

1. Bolfarine, H., Sandoval, M. C.: Introdução à Inferência Estatística. SBM, Rio de Janeiro. (2001)

2. Borovkov, A. A.: Estadística matemática. Mir, Moscú. (1984)

3. Hoel P. G.: Introducción a la estadística matemática. Ariel, Barcelona. (1980)

4. Pugachev, V. S.: Introducción a la Teoría de Probabilidades. Mir, Moscu. (1973)

5. Robert, C. P.: The Bayesian Choice. Springer, New York. (2007)

6. Ross, S. M.: Introduction to Probability and Statistics for Engieneers and Scientists. Elsevier Academic Press, San Diego. (2004)