Fundamentos de la inferencia bayesiana

(ver 2019-2020)

Emilio Letón y Elisa M. Molanes-López

04-Ene-2020

Página legal

FUNDAMENTOS DE LA INFERENCIA BAYESIANA

Madrid, versión 2019-2020

ISBN electrónico: xxx.
Edición digital (epub): xxx.

Prefacio

Este material está diseñado con el paradigma del grupo de innovación docente miniXmodular de generar material en formato mini y modular, según se puede ver en la página web www.minixmodular.ia.uned.es, donde se introduce, entre otros, el concepto de mini-libro electrónico modular.

En este mini-libro se dan los conceptos básicos de inferencia bayesiana. Hay que señalar que en la estadística clásica o frecuentista, los parámetros son constantes desconocidas que hay que estimar. Sin embargo, en la llamada estadística bayesiana, los parámetros son variables aleatorias.

Se recomienda refrescar los conceptos frecuentistas de intervalos de confianza (IC) y contrastes de hipótesis (CH) en, por ejemplo, [4] para comprender mejor las diferencias (y semejanzas) entre sus homólogos bayesianos que se verán en este mini-libro.

Mini-capítulo 1

Conceptos básicos

En estadística frecuentista se estudian parámetros θ que se suponen constantes desconocidas. En estadística bayesiana los parámetros θ se suponen que son variables aleatorias (v.a.) con una "distribución" a priori antes de realizar el experimento y con una "distribución" a posteriori después de su realización. Hay que destacar que el término "distribución" es sinónimo de modelo, de modo que si una v.a. sigue una distribución Bernoulli, significa que su modelo es Bernoulli. El motivo de usar la palabra "distribución" es para recordar el concepto de función de distribución, común tanto para las variables aleatorias discretas como continuas. En este mini-capítulo también se verán los conceptos de familias conjugadas, familia exponencial k-paramétrica, distribuciones a priori no informativas y la regla de Jeffrey.

1.1. Distribución a priori y a posteriori

Se supone que se tiene $Y_1, \ldots, Y_n$ una muestra aleatoria simple (m.a.s.) de una v.a. Y con una distribución $p(y|\theta)$ siendo p una función de masa de probabilidad (o simplemente función de masa), en el caso de que Y sea discreta, o una función de densidad de probabilidad (o simplemente función de densidad), en el caso de que Y sea continua. Por otra parte $|\theta$ indica que dicha p depende de θ. En el contexto bayesiano se supone que θ es una v.a. con una distribución o modelo de probabilidad dado por $p(\theta)$, donde de nuevo p puede ser función de masa o de densidad según se suponga θ v.a. discreta o continua. A esta $p(\theta)$ se la conoce como distribución a priori para θ, en contraposición a $p(\theta|y_1, \ldots, y_n)$, que se la conoce como distribución a posteriori y que, de forma análoga al teorema de Bayes, se calcula como

$$p(\theta|y_1, \ldots, y_n) = \frac{p(\theta, y_1, \ldots, y_n)}{p(y_1, \ldots, y_n)} = \frac{p(\theta, y_1, \ldots, y_n)}{\int_\Theta p(\theta, y_1, \ldots, y_n) d\theta} = \frac{p(\theta) p(y_1, \ldots, y_n|\theta)}{\int_\Theta p(\theta) p(y_1, \ldots, y_n|\theta) d\theta}.$$

A efectos prácticos se trabaja con

$$p(\theta|y_1, \ldots, y_n) \propto p(\theta) p(y_1, \ldots, y_n|\theta),$$

donde el símbolo $\propto$ indica que es "proporcional a". Esto quiere decir que a la hora de calcular la distribución a posteriori, se parte de que tiene que ser proporcional al producto de $p(\theta)$ y de $p(y_1, \ldots, y_n|\theta)$. Una vez se haya calculado dicho producto existen dos posibilidades de finalizar la determinación por completo de $p(\theta|y_1, \ldots, y_n)$. La primera posibilidad es calcular dicha constante de proporcionalidad para que $p(\theta|y_1, \ldots, y_n)$ sea realmente una función de masa o una función de densidad (que sume o integre 1, respectivamente) y la segunda posibilidad es tratar de identificar el modelo de probabilidad teórico en un catálogo de los modelos de probabilidad más usuales (ver, por ejemplo, [1], [2] y los ejemplos que se mostrarán más adelante en esta misma mini-sección). Conviene notar que la expresión anterior también se suele escribir como

$$p(\theta|y_1, \ldots, y_n) \propto p(\theta) \ell(\theta|y_1, \ldots, y_n),$$

siendo $\ell(\theta|y_1, \ldots, y_n)$ la función de verosimilitud que es $p(y_1, \ldots, y_n|\theta)$ pero haciendo hincapié en que la muestra está fijada (ver, por ejemplo, [1], [2] y [3]).

Conviene señalar que si t es un estadístico suficiente, se tiene que

$$p(\theta|y_1, \ldots, y_n) = p(\theta|t) \propto p(\theta) \ell(\theta|t),$$

según se puede ver en, por ejemplo, [2].

Ejemplo

1. Se supone que se tiene $Y_1,\ldots,Y_n$ una m.a.s. de $Y \sim Ber(\theta)$ con $p(\theta) \sim Bet(2,3)$, entonces se tiene que $p(\theta|y_1,\ldots,y_n) \sim Bet\big(\sum_{i=1}^n y_i + 2, n+3-\sum_{i=1}^n y_i\big)$.

   SOLUCIÓN:

   Se tiene que

$$\begin{aligned}\ell(\theta|y_1,\ldots,y_n) &= \prod_{i=1}^n p(y_i|\theta) = \theta^{y_1}(1-\theta)^{1-y_1}\cdot\ldots\cdot\theta^{y_n}(1-\theta)^{1-y_n}\\ &\propto \theta^{\sum_{i=1}^n y_i}(1-\theta)^{n-\sum_{i=1}^n y_i},\end{aligned}$$

   por lo que

$$\begin{aligned}p(\theta|y_1,\ldots,y_n) &\propto p(\theta)\ell(\theta|y_1,\ldots,y_n)\\ &= \frac{\Gamma(2+3)}{\Gamma(2)\Gamma(3)}\theta^{2-1}(1-\theta)^{3-1}\theta^{\sum_{i=1}^n y_i}(1-\theta)^{n-\sum_{i=1}^n y_i}\\ &\propto \theta^{1+\sum_{i=1}^n y_i}(1-\theta)^{n+2-\sum_{i=1}^n y_i}\\ &= \theta^{(1+\sum_{i=1}^n y_i+1)-1}(1-\theta)^{(n+2-\sum_{i=1}^n y_i+1)-1}\\ &\propto Bet\big(\sum_{i=1}^n y_i+2, n+3-\sum_{i=1}^n y_i\big)\end{aligned}$$

2. Se supone que se tiene una $Y_1,\ldots,Y_n$ m.a.s. de $Y \sim N(\theta,\sigma^2)$ con $p(\theta) \sim N(\mu,\tau^2)$, con σ^2,μ y τ^2 conocidos, entonces se tiene que $p(\theta|y_1,\ldots,y_n) \sim N\big(\mu(y),\tau^2(y)\big)$, siendo

$$\begin{aligned}\mu(y) &= \frac{\sigma^2}{\sigma^2+n\tau^2}\mu + \frac{n\tau^2}{\sigma^2+n\tau^2}\overline{y} = \frac{\frac{1}{\tau^2}}{\frac{1}{\tau^2}+\frac{n}{\sigma^2}}\mu + \frac{\frac{n}{\sigma^2}}{\frac{1}{\tau^2}+\frac{n}{\sigma^2}}\overline{y}\\ \tau^2(y) &= \frac{\sigma^2\tau^2}{\sigma^2+n\tau^2} = \frac{1}{\frac{1}{\tau^2}+\frac{n}{\sigma^2}}.\end{aligned}$$

   SOLUCIÓN:

   Se tiene que

$$\begin{aligned}p(\theta|y_1,\ldots,y_n) &= p(\theta|\overline{y}) \propto p(\theta)p(\overline{y}|\theta)\\ &\propto exp\Big[-\frac{1}{2}\Big[\frac{(\theta-\mu)^2}{\tau^2}+\frac{n(\overline{y}-\theta)^2}{\sigma^2}\Big]\Big]\\ &= exp\Big[-\frac{1}{2}\Big[\frac{\theta^2-2\mu\theta+\mu^2}{\tau^2}+\frac{n\overline{y}^2-2n\overline{y}\theta+n\theta^2}{\sigma^2}\Big]\Big]\\ &= exp\Big[-\frac{1}{2}\Big[\Big(\frac{1}{\tau^2}+\frac{n}{\sigma^2}\Big)\theta^2-2\Big(\frac{\mu}{\tau^2}+\frac{n\overline{y}}{\sigma^2}\Big)\theta+\Big(\frac{\mu^2}{\tau^2}+\frac{n\overline{y}^2}{\sigma^2}\Big)\Big]\Big]\\ &\propto exp\Big[-\frac{1}{2}\Big(\frac{1}{\tau^2}+\frac{n}{\sigma^2}\Big)\Big(\theta^2-2\frac{\frac{\mu}{\tau^2}+\frac{n\overline{y}}{\sigma^2}}{\frac{1}{\tau^2}+\frac{n}{\sigma^2}}\theta\Big)\Big]\\ &= exp\Big[-\frac{1}{2}\Big(\frac{1}{\tau^2}+\frac{n}{\sigma^2}\Big)\Big(\theta^2-2\frac{\sigma^2\mu+n\tau^2\overline{y}}{\sigma^2+n\tau^2}\theta\Big)\Big]\\ &= exp\Big[-\frac{1}{2}\frac{1}{\tau^2(y)}(\theta^2-2\mu(y)\theta)\Big]\\ &= exp\Big[-\frac{1}{2}\frac{1}{\tau^2(y)}((\theta-\mu(y))^2-\mu^2(y))\Big]\\ &\propto N\big(\mu(y),\tau^2(y)\big)\end{aligned}$$

1.2. Familias conjugadas

Se dice que la familia $\mathcal{P}=\{p(\theta)\}$ es una familia conjugada de la verosimilitud $p(y_1,\ldots,y_n|\theta)$ sii al considerar $p(\theta)$ como distribución a priori se obtiene que $p(\theta|y_1,\ldots,y_n)\in\mathcal{P}$.

Ejemplo

1. La familia de la distribución Beta es conjugada respecto de la verosimilitud Bernoulli. En concreto se tiene que si $Y_1, \ldots, Y_n$ es una m.a.s. de $Y \sim Ber(\theta)$ con $p(\theta) \sim Bet(p_0, q_0)$, entonces se tiene que $p(\theta|y_1, \ldots, y_n) \sim Bet\big(\sum_{i=1}^n y_i + p_0, n + q_0 - \sum_{i=1}^n y_i\big)$. Este resultado configura el modelo Beta-Bernoulli.

   SOLUCIÓN:

   Se tiene que

$$
\begin{aligned}
\ell(\theta|y_1, \ldots, y_n) &= \prod_{i=1}^n p(y_i|\theta) = \theta^{y_1}(1-\theta)^{1-y_1} \cdot \ldots \cdot \theta^{y_n}(1-\theta)^{1-y_n} \\
&\propto \theta^{\sum_{i=1}^n y_i}(1-\theta)^{n-\sum_{i=1}^n y_i}
\end{aligned}
$$

   por lo que

$$
\begin{aligned}
p(\theta|y_1, \ldots, y_n) &\propto p(\theta)\ell(\theta|y_1, \ldots, y_n) \\
&\propto \frac{\Gamma(p_0+q_0)}{\Gamma(p_0)\Gamma(q_0)}\theta^{p_0-1}(1-\theta)^{q_0-1}\theta^{\sum_{i=1}^n y_i}(1-\theta)^{n-\sum_{i=1}^n y_i} \\
&\propto \theta^{\sum_{i=1}^n y_i + p_0 - 1}(1-\theta)^{n-\sum_{i=1}^n y_i + q_0 - 1} \\
&\propto Bet\big(\sum_{i=1}^n y_i + p_0, n + q_0 - \sum_{i=1}^n y_i\big)
\end{aligned}
$$

2. La familia de la distribución Gamma es conjugada respecto de la verosimilitud Poisson. En concreto se tiene que si $Y_1, \ldots, Y_n$ es una m.a.s. de $Y \sim Poi(\theta)$ con $p(\theta) \sim Gam(a_0, p_0)$, entonces se tiene que $p(\theta|y_1, \ldots, y_n) \sim Gam\big(n + a_0, \sum_{i=1}^n y_i + p_0\big)$. Este resultado configura el modelo Gamma-Poisson.

   SOLUCIÓN:

   Se tiene que

$$
\begin{aligned}
\ell(\theta|y_1, \ldots, y_n) &= \prod_{i=1}^n p(y_i|\theta) = \frac{e^{-\theta}\theta^{y_1}}{y_1!} \cdot \ldots \cdot \frac{e^{-\theta}\theta^{y_n}}{y_n!} \\
&\propto e^{-n\theta}\theta^{\sum_{i=1}^n y_i}
\end{aligned}
$$

   por lo que

$$
\begin{aligned}
p(\theta|y_1, \ldots, y_n) &\propto p(\theta)\ell(\theta|y_1, \ldots, y_n) \\
&\propto \frac{a_0^{p_0}}{\Gamma(p_0)}e^{-a_0\theta}\theta^{p_0-1}e^{-n\theta}\theta^{\sum_{i=1}^n y_i} \\
&\propto e^{-(n+a_0)\theta}\theta^{\sum_{i=1}^n y_i + p_0 - 1} \\
&\propto Gam\big(n + a_0, \sum_{i=1}^n y_i + p_0\big)
\end{aligned}
$$

1.3. Familia exponencial k-paramétrica

Se dice que una v.a. Y (discreta o continua) pertenece a la familia exponencial k-paramétrica si $p(y|\theta)$ (que podrá ser la función de masa o de densidad) se puede poner de la forma

$$
p(y|\theta) = h(y)g(\theta)exp\Big\{\sum_{j=1}^k \phi_j(\theta)u_j(y)\Big\} = h(y)g(\theta)exp\Big\{\phi(\theta)^T u(y)\Big\},
$$

donde $\phi(\theta)$ recibe el nombre de parámetro natural. En la definición anterior θ puede ser un vector, al igual que $\phi(\theta)$ y $u(y)$.

Si se toma una m.a.s. $Y_1, \ldots, Y_n$ de una v.a. Y de la familia exponencial, se tiene que

$$
\begin{aligned}
p(y_1, \ldots, y_n|\theta) &= h(y_1)g(\theta)exp\Big\{\phi(\theta)^T u(y_1)\Big\} \cdot \ldots \cdot h(y_n)g(\theta)exp\Big\{\phi(\theta)^T u(y_n)\Big\} \\
&= \Big(\prod_{i=1}^n h(y_i)\Big)\Big(g(\theta)\Big)^n exp\Big\{\phi(\theta)^T \sum_{i=1}^n u(y_i)\Big\},
\end{aligned}
$$

por lo que $T(Y_1, \ldots, Y_n) = \sum_{i=1}^{n} u(Y_i)$ es un estadístico suficiente para θ (ver, por ejemplo, [2]) y además si se toma

$$p(\theta) \propto \Big(g(\theta)\Big)^{\eta} exp\Big\{\phi(\theta)^T \nu\Big\},$$

se tiene que $\mathcal{P} = \{p(\theta)\}$ es la familia natural conjugada respecto de la verosimilitud $p(y_1, \ldots, y_n|\theta)$ ya que

$$\begin{aligned} p(\theta|y_1, \ldots, y_n) &\propto \Big(g(\theta)\Big)^{\eta} exp\Big\{\phi(\theta)^T \nu\Big\} \Big(g(\theta)\Big)^{n} exp\Big\{\phi(\theta)^T t(y_1, \ldots, y_n)\Big\} \\ &= \Big(g(\theta)\Big)^{\eta+n} exp\Big\{\phi(\theta)^T (\nu + t(y_1, \ldots, y_n))\Big\} \in \mathcal{P} \end{aligned}$$

Ejemplo

1. La familia de la distribución Gamma es la familia natural conjugada respecto de la verosimilitud Poisson.

   SOLUCIÓN:

   Se tiene que

   $$p(y|\theta) = \frac{e^{-\theta}\theta^y}{y!} = \frac{1}{y!} e^{-\theta} exp\{Ln(\theta^y)\} = \frac{1}{y!} e^{-\theta} exp\{Ln(\theta) y\}$$

   pertenece a la familia exponencial 1-paramétrica con parámetro natural $\phi(\theta) = Ln(\theta)$, $h(y) = \frac{1}{y!}$, $g(\theta) = e^{-\theta}$ y $u(y) = y$. Además $T(Y_1, \ldots, Y_n) = \sum_{i=1}^{n} Y_i$, es estadístico suficiente para θ y la familia natural conjugada respecto de la verosmilitud Poisson es

   $$p(\theta) \propto e^{-\eta\theta}\theta^{\nu},$$

   que correponde a una $Gam(\eta, \nu + 1)$.

1.4. Distribuciones a priori no informativas

El concepto de distribuciones a priori no informativas intenta reflejar el hecho de que o bien no se sabe a priori nada de θ o bien se desea que la elección de la distribución a priori tenga el menor efecto posible en la distribución de los datos (dejar a los datos hablar por sí mismos, como se dice en [1]).

En el caso de que θ sea un parámetro de localización, es decir que $p(y|\theta) = p(y - \theta)$, la distribución a priori no informativa viene dada por $p(\theta) \propto 1$ (ver, por ejemplo, [2] y [1]). Conviene señalar que aunque la distribución anterior es impropia (en el sentido de que su suma/integral supera 1), sin embargo la distribución a posteriori a partir de ella, en la mayoría de las situaciones, es propia (su suma/integral es 1). En el caso de que θ sea un parámetro de escala, es decir que $p(y|\theta) = \frac{1}{\theta} p(\frac{y}{\theta})$, la distribución a priori $p(\theta) \propto \frac{1}{\theta}$ es no informativa en la escala logaritmo (ver, por ejemplo, [2] y [1]).

Ejemplo

1. Para $\theta \in (0, 1)$ la distribución a priori $p(\theta) \propto \theta^{-1}(1-\theta)^{-1} \sim Bet(0, 0)$ es no informativa en la escala logit.

   SOLUCIÓN:

   Se tiene que si $\Delta = h(\theta)$, se verifica (ver, por ejemplo, [4] y [1]) que

   $$p(\Delta) = p(\theta)\Big|\frac{d\theta}{d\Delta}\Big|,$$

   que aplicado a $\Delta = h(\theta) = logit(\theta)$ resulta que $\theta = \frac{e^{\Delta}}{1+e^{\Delta}}$ y $\frac{d\theta}{d\Delta} = \frac{e^{\Delta}}{(1+e^{\Delta})^2}$ por lo que

   $$p(\Delta) \propto \theta^{-1}(1-\theta)^{-1}\Big|\frac{e^{\Delta}}{(1+e^{\Delta})^2}\Big| = \frac{1+e^{\Delta}}{e^{\Delta}}(1+e^{\Delta})\frac{e^{\Delta}}{(1+e^{\Delta})^2} = 1.$$

2. Para $\theta \in (0, 1)$ la distribución a priori $p(\theta) \propto \theta^{-1/2}(1-\theta)^{-1/2} \sim Bet(\frac{1}{2}, \frac{1}{2})$ es no informativa en la escala arco-seno-raíz.

   SOLUCIÓN:

   En este caso $\Delta = h(\theta) = arcsen\sqrt{\theta}$ con lo que $d\Delta = \frac{1}{\sqrt{1-\theta}} \frac{1}{2} \frac{1}{\sqrt{\theta}} d\theta$ y $\frac{d\theta}{d\Delta} = 2(1-\theta)^{1/2}\theta^{1/2}$ por lo que

   $$p(\Delta) \propto \theta^{-1/2}(1-\theta)^{-1/2}|(1-\theta)^{1/2}\theta^{1/2}| = 1.$$

1.5. Regla de Jeffrey

En su versión unidimensional, la regla de Jeffrey para θ en la verosimilitud $p(y|\theta)$ afirma (ver, por ejemplo, [2] y [1]) que la distribución a priori $p(\theta) \propto \sqrt{J(\theta)}$ es no informativa en la escala $\Delta \propto \int^{\theta} \sqrt{J(t)}dt$ siendo

$$J(\theta) = E_\theta\Big[\frac{d}{d\theta}Ln(p(y|\theta)\Big]^2 = -E_\theta\Big[\frac{d^2}{d\theta^2}Ln(p(y|\theta)\Big].$$

Ejemplo

1. La regla de Jeffrey para $\theta \in (0,1)$ en una verosimilitud binomial, proporciona la distribución a priori $p(\theta) \propto \theta^{-1/2}(1-\theta)^{-1/2} \sim Bet(\frac{1}{2},\frac{1}{2})$ que es no informativa en la escala arco-seno-raíz.

   SOLUCIÓN:

   Se tiene que

$$\begin{aligned}
p(y|\theta) &= \binom{n}{y}\theta^y(1-\theta)^{n-y} \\
Ln(p(y|\theta)) &= cte + yLn(\theta) + (n-y)Ln(1-\theta) \\
\frac{d}{d\theta}Ln(p(y|\theta)) &= y\frac{1}{\theta} - (n-y)\frac{1}{1-\theta} \\
\frac{d^2}{d\theta^2}Ln(p(y|\theta)) &= \frac{-y}{\theta^2} - (n-y)\frac{1}{(1-\theta)^2} \\
J(\theta) &= -E_\theta\Big[\frac{d^2}{d\theta^2}Ln(p(y|\theta))\Big] = -E\Big[\frac{-y}{\theta^2} - (n-y)\frac{1}{(1-\theta)^2}|\theta\Big] \\
&= \frac{1}{\theta^2}E[y|\theta] + \frac{1}{(1-\theta)^2}E[n-y|\theta] = \frac{1}{\theta^2}n\theta + \frac{1}{(1-\theta)^2}(n-n\theta) \\
&= \frac{n}{\theta} + \frac{n}{1-\theta} = \frac{n}{\theta(1-\theta)} \\
p(\theta) &\propto \sqrt{J(\theta)} \propto \theta^{-1/2}(1-\theta)^{-1/2} \sim Bet(\frac{1}{2},\frac{1}{2}),
\end{aligned}$$

   que es no informativa en la escala

$$\Delta \propto \int^{\theta}\sqrt{J(t)}dt \propto \int^{\theta} t^{-1/2}(1-t)^{-1/2}dt = [2arcsen\sqrt{t}]^{\theta} \propto arcsen\sqrt{\theta}.$$

2. La regla de Jeffrey para $\theta \in \mathbb{R}$ en una verosimilitud normal $N(\theta,\sigma^2)$, con σ^2 conocida, proporciona la distribución a priori $p(\theta) \propto 1$ que es no informativa en la escala de θ sin transformar.

   SOLUCIÓN:

   Se tiene que

$$\begin{aligned}
p(y|\theta) &= \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2\sigma^2}(y-\theta)^2} \\
Ln(p(y|\theta)) &= cte - \frac{1}{2\sigma^2}(y-\theta)^2 \\
\frac{d}{d\theta}Ln(p(y|\theta)) &= -\frac{1}{2\sigma^2}2(y-\theta)(-1) = \frac{1}{\sigma^2}(y-\theta) \\
\frac{d^2}{d\theta^2}Ln(p(y|\theta)) &= \frac{-1}{\sigma^2} \\
J(\theta) &= -E_\theta\Big[\frac{d^2}{d\theta^2}Ln(p(y|\theta))\Big] = -E\Big[\frac{-1}{\sigma^2}|\theta\Big] = \frac{1}{\sigma^2} \\
p(\theta) &\propto \sqrt{J(\theta)} \propto \frac{1}{\sigma} \propto 1,
\end{aligned}$$

   que es no informativa en la escala

$$\Delta \propto \int^{\theta}\sqrt{J(t)}dt = \int^{\theta}\frac{1}{\sigma}dt = \frac{\theta}{\sigma} \propto \theta.$$

Mini-capítulo 2

Inferencia bayesiana básica

En este mini-capítulo se tratará de forma introductoria los conceptos de intervalos de credibilidad bayesianos y contrastes de hipótesis bayesianos.

2.1. Intervalos de credibilidad bayesianos

Un intervalo de credibilidad bayesiano de probabilidad $1-\alpha$ es una región $C(y_1,\ldots,y_n)$ de forma que $P(\theta \in C(y_1,\ldots,y_n)) = 1-\alpha$, donde dicha probabilidad se calcula con la distribución a posteriori $p(\theta|y_1,\ldots,y_n)$.

Ejemplo

1. Sea $Y_1,\ldots,Y_n$ una m.a.s. de la v.a. $Y \sim Ber(\theta)$ con $p(\theta) \sim Bet(2,2)$. Se pide construir un intervalo de credibilidad bayesiano al 95 % para θ, sabiendo que después de realizar el experimento se han obtenido los datos experimentales
$$1,0,0,0,0,1,0,0,1,1,1,1,1,1,1,1.$$
En la resolución de este ejercicio habrá que utilizar que $F_{Bet(12,8)}(0.4587) = 0.025$ y $F_{Bet(12,8)}(0.6034) = 0.975$, siendo $F_{Bet(12,8)}$ la función de distribución de una $Bet(12,8)$.
SOLUCIÓN:
Se sabe que $p(\theta|y_1,\ldots,y_n) \sim Bet(\sum_{i=1}^n y_i + 2, n+2-\sum_{i=1}^n y_i) = Bet(12,8)$. Se trata de encontrar a y b de forma que
$$0.95 = P(a < \theta < b).$$
El valor de a tendrá que verificar que $P(Y \leq a) = F_{Bet(12,8)}(a) = 0.05/2$ con lo que $a = 0.4587$ y el valor de b que $P(Y \geq b) = 1 - P(Y < b) = 1 - F_{Bet(12,8)}(b) = 0.05/2$ con lo que $F_{Bet(12,8)}(b) = 0.975$ y por tanto $b = 0.6034$. Todos los cálculos anteriores permiten afirmar que el intervalo de credibilidad bayesiano al 95 % para θ es $(0.4587, 0.6034)$.

2. Sea $Y_1,\ldots,Y_n$ una m.a.s. de la v.a. $Y \sim N(\theta,\sigma^2)$ con σ^2 conocido. Se pide demostrar que si $p(\theta) \propto 1$, distribución a priori no informativa, entonces el intervalo de confianza frecuentista clásico y el intervalo de credibilidad bayesiano coinciden numéricamente.
SOLUCIÓN:
Se sabe que $\overline{Y} \sim N(\theta, \frac{\sigma^2}{n})$ y que es suficiente para θ. Por otra parte
$$\begin{aligned} p(\theta|y_1,\ldots,y_n) = p(\theta|\overline{y}) &\propto p(\theta)p(\overline{y}|\theta) \\ &\propto 1 \cdot e^{-\frac{1}{2\sigma^2/n}(\theta-\overline{y})^2} \end{aligned}$$
Para que la última expresión integre 1 hay que calcular la constante de proporcionalidad. Dado que
$$\int_{-\infty}^{+\infty} e^{-\frac{1}{2\sigma^2/n}(\theta-\overline{y})^2} d\theta = \sqrt{\frac{2\pi\sigma^2}{n}},$$
se tiene que
$$p(\theta|y_1,\ldots,y_n) = \frac{1}{\sqrt{\frac{2\pi\sigma^2}{n}}} e^{-\frac{1}{2\sigma^2/n}(\theta-\overline{y})^2},$$
por lo que $p(\theta|y_1,\ldots,y_n) \sim N(\overline{y}, \frac{\sigma^2}{n})$. Por lo tanto, al tomar $p(\theta) \propto 1$, distribución a priori no informativa, el intervalo de confianza frecuentista clásico y el intervalo de credibilidad bayesiano coinciden numéricamente.

2.2. Contrastes de hipótesis bayesianos

En el contexto de contrastes de hipótesis bayesianos se plantean las siguientes hipótesis:

$$\begin{aligned} H_0 &: \quad \theta \in \Theta_0 \\ H_1 &: \quad \theta \in \Theta_1. \end{aligned}$$

A partir de la distribución a priori de θ se puede calcular la probabilidad a priori de cada hipótesis:

$$\pi_0 = P(\theta \in \Theta_0) = \int_{\Theta_0} p(\theta) d\theta$$

$$\pi_1 = P(\theta \in \Theta_1) = \int_{\Theta_1} p(\theta) d\theta.$$

Análogamente, a partir de la distribución a posteriori de θ también se pueden calcular las probabilidades a posteriori correspondientes:

$$\alpha_0 = P(\theta \in \Theta_0 | y_1, \ldots, y_n) = \int_{\Theta_0} p(\theta | y_1, \ldots, y_n) d\theta$$

$$\alpha_1 = P(\theta \in \Theta_1 | y_1, \ldots, y_n) = \int_{\Theta_1} p(\theta | y_1, \ldots, y_n) d\theta.$$

Utilizando dichas probabilidades se define el factor Bayes B (ver, por ejemplo, [2] y [1]) como el cociente entre la razón de odds a posteriori y la razón de odds a priori, es decir:

$$B = \frac{\alpha_0/\alpha_1}{\pi_0/\pi_1},$$

con lo que

$$\frac{\alpha_0}{\alpha_1} = \frac{\pi_0}{\pi_1} B.$$

A la hora de elegir entre H_0 y H_1, la decisión se realiza en base a α_0 y α_1, de forma que si $\alpha_0 > \alpha_1$ se acepta H_0 y en caso contrario se acepta H_1.

Ejemplo

1. Sea $Y_1, \ldots, Y_n$ una m.a.s. de la v.a. $Y \sim Ber(\theta)$ con $p(\theta) \sim Bet(2,2)$. Se pide realizar un contraste de hipótesis bayesiano entre

$$\begin{aligned} H_0 &: \quad \theta \leq 0.5 \\ H_1 &: \quad \theta > 0.5, \end{aligned}$$

sabiendo que después de realizar el experimento se han obtenido los datos experimentales

$$1, 0, 0, 0, 0, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1.$$

En la resolución de este ejercicio habrá que utilizar que $F_{Bet(12,8)}(0.5) = 0.18$.

SOLUCIÓN:

Se sabe que $p(\theta|y_1, \ldots, y_n) \sim Bet(\sum_{i=1}^n y_i + 2, n + 2 - \sum_{i=1}^n y_i) = Bet(12, 8)$. Se trata de calcular α_0 y α_1 :

$$\begin{aligned} P(\theta \in \Theta_0 | y_1, \ldots, y_n) &= P(\theta \leq 0.5) = F_{Bet(12,8)}(0.5) = 0.18 \\ P(\theta \in \Theta_1 | y_1, \ldots, y_n) &= P(\theta > 0.5) = 1 - F_{Bet(12,8)}(0.5) = 1 - 0.18 = 0.82 \end{aligned}$$

con lo que, al ser $\alpha_0 < \alpha_1$ se acepta H_1.

2. En hipótesis simples, el factor de Bayes es el cociente de verosimilitudes y no depende de la distribución a priori.

SOLUCIÓN:

En este ejemplo $\Theta = \{\theta_0, \theta_1\}$, con lo que trata de realizar un contraste de hipótesis bayesiano entre

$$\begin{aligned} H_0 :& \quad \theta = \theta_0 \\ H_1 :& \quad \theta = \theta_1. \end{aligned}$$

En este contexto

$$\begin{aligned}
\alpha_0 &= P(\theta = \theta_0 | y_1, \ldots, y_n) \propto P(\theta = \theta_0) p(y_1, \ldots, y_n | \theta_0) = \frac{\pi_0 p(y_1, \ldots, y_n | \theta_0)}{\pi_0 p(y_1, \ldots, y_n | \theta_0) + \pi_1 p(y_1, \ldots, y_n | \theta_1)}, \\
\alpha_1 &= P(\theta = \theta_1 | y_1, \ldots, y_n) \propto P(\theta = \theta_1) p(y_1, \ldots, y_n | \theta_1) = \frac{\pi_1 p(y_1, \ldots, y_n | \theta_1)}{\pi_0 p(y_1, \ldots, y_n | \theta_0) + \pi_1 p(y_1, \ldots, y_n | \theta_1)}, \\
\frac{\alpha_0}{\alpha_1} &= \frac{\pi_0}{\pi_1} \frac{p(y_1, \ldots, y_n | \theta_0)}{p(y_1, \ldots, y_n | \theta_1)},
\end{aligned}$$

con lo que $B = \frac{p(y_1, \ldots, y_n | \theta_0)}{p(y_1, \ldots, y_n | \theta_1)}$ y, por tanto, es el cociente de verosimilitudes y no depende en este caso de la distribución a priori.

Referencias

[1] Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., & Rubin, D., *Bayesian data analysis*, Chapman and Hall/CRC, 2014.

[2] Lee, P.M., *Bayesian statistics: an introduction*, Arnold Publication, 1997.

[3] Peña, D., *Fundamentos de Estadística*, Alianza editorial (Madrid), 2014.

[4] Ramos, E., Vélez, R., & Hernández, V., *Modelos probabilísticos y optimización*, Sanz y Torres (Madrid), 2019.

Acerca de

Emilio Letón

Figura 2.1: Emilio Letón

Emilio Letón nace en Madrid en 1966. Es Licenciado en Matemáticas por la UCM en 1989 y doctor en Matemáticas en la misma universidad en 2002. En la actualidad es profesor contratado doctor de la UNED en el departamento de Inteligencia Artificial, al que se incorporó en 2009. Anteriormente fue profesor del departamento de Estadística de la UC3M durante 5 años. Asimismo, ha trabajado durante 15 años en departamentos de Planificación y Estadística dentro del sector bancario y de la industria farmacéutica. Sus líneas de investigación incluyen el Análisis de Supervivencia, tests no paramétricos, PLS, Meta-Análisis, Bioestadística y B-Learning. Ha participado en más de 30 proyectos de innovación docente (siendo coordinador en más de 10 de ellos) colaborando con distintas universidades: UNED, UC3M, UCM y UPM. Ha recibido 1 premio en excelencia en publicaciones científicas (UC3M) y 5 premios en excelencia docente (1 en UC3M, 3 en UNED y 1 en OCW Consortium). En @emilioleton se pueden encontrar sus tweets y su página web personal con información ampliada de su curriculum. https://twitter.com/emilioleton

Elisa M. Molanes-López

Figura 2.2: Elisa M. Molanes-López

Elisa M. Molanes López nace en Vigo en 1976. Se licencia en Matemáticas por la Universidad de Santiago de Compostela en 2000 y consigue el grado de doctora, con acreditación europea, por la Universidad de A Coruña en 2007. Entre 2003 y 2007, al amparo de una beca FPI, realiza varias estancias de investigación en las siguientes universidades extranjeras: Universiteit Hasselt (Bélgica), Université Catholique de Louvain (Bélgica) y The University of Texas (EE.UU.). Durante 8 años, entre 2007 y 2015, es profesora visitante en el departamento de Estadística de la Universidad Carlos III de Madrid. Actualmente, desde octubre de 2015, es profesora ayudante doctora en la Unidad Departamental de Bioestadística del Departamento de Estadística e Investigación Operativa de la Universidad Complutense de Madrid. Durante su etapa docente ha participado en 10 proyectos de innovación docente. En diciembre de 2013, recibe un accésit a la mejor práctica docente en los Premios del Consejo Social de la UNED por su participación en el MOOC "Mini-vídeos docentes modulares: un elemento crítico en el diseño de un MOOC", y en abril de 2014, el OpenCourseWare Consortium le otorga un premio de excelencia por su participación en el curso OCW "Mini-vídeos docentes modulares para diseñar un MOOC". Sus líneas de investigación incluyen la estadística no paramétrica, el análisis de supervivencia, las curvas ROC y las funciones cópula. En la página web personal de Elisa M. Molanes-López, se puede encontrar información más detallada. https://elisamariamolanes.wixsite.com/emolanes

Índice general

Índice de figuras