Inferencia Estadística - Guia 5

Nicolas Ferrer
e-mail: nicolas.ferrer.747@gmail.com

Agosto 2020

1 Ejercicio 1

Sea $X|\theta \sim \text{Pois}(\theta)$, $P(\theta = 2) = 1/3$ y $P(\theta = 3) = 2/3$. Supón que observamos $x_1 = 2$ y $x_2 = 4$.
Queremos calcular la probabilidad a posteriori de θ:

$$P(\theta|X_1 = 2, X_2 = 4) = \frac{P(X_1 = 2, X_2 = 4|\theta)}{P(X_1 = 2, X_2 = 4)}P(\theta)$$

Recordar que para el modelo Poisson:

$$P(\mathbf{X}|\theta) = \frac{\theta^{\sum x_i} e^{-n\theta}}{\prod x_i!} \Rightarrow P(X_1 = 2, X_2 = 4|\theta) = \frac{\theta^6 e^{-2\theta}}{2 * 4!}$$

Por otro lado, usando la ayuda provista:

$$\begin{aligned} P(X_1 = 2, X_2 = 4) &= P(X_1 = 2, X_2 = 4|\theta = 2)P(\theta = 2) + P(X_1 = 2, X_2 = 4|\theta = 3)P(\theta = 3) \\ &= \frac{2^6 e^{-2*2}}{2 * 4!} * \frac{1}{3} + \frac{3^6 e^{-2*3}}{2 * 4!} * \frac{2}{3} \\ &\approx 0.033 \end{aligned}$$

Por lo tanto, podemos escribir la probabilidad a posteriori de θ como:

$$P(\theta|X_1 = 2, X_2 = 4) = \frac{P(\theta)}{P(X_1 = 2, X_2 = 4)} \frac{\theta^6 e^{-2\theta}}{2 * 4!}$$

Notar que $P(X_1 = 2, X_2 = 4)$ es independiente de θ, por lo tanto, podemos escribir:

$$P(\theta|X_1 = 2, X_2 = 4) \propto \theta^6 e^{-2\theta} P(\theta)$$

Para realizar inferencia sobre θ, podemos reportar θ_{MAP}, el valor que maximiza la probabilidad a posteriori. Calculando los valores de $P(\theta|\mathbf{X})$ para cada caso:

$$\begin{aligned} P(\theta = 2|X_1 = 2, X_2 = 4) &\propto 2^6 e^{-2*2} * \frac{1}{3} \approx 0.39 \\ P(\theta = 3|X_1 = 2, X_2 = 4) &\propto 3^6 e^{-2*3} * \frac{2}{3} \approx 1.2 \end{aligned}$$

Por lo tanto, $\theta_{MAP} = 3$.

2 Ejercicio 2

Definimos la probabilidad a posteriori de θ condicional en haber observado $\{\mathbf{x}_1, \mathbf{x}_2\}$ como:

$$\pi(\theta|\mathbf{x}_1, \mathbf{x}_2) \propto L(\theta|\mathbf{x}_1, \mathbf{x}_2)\pi(\theta)$$

Si las observaciones provienen de una secuencia de variables aleatorias independientemente distribuidas, podemos escribir $L(\theta|\mathbf{x}_1, \mathbf{x}_2) = L(\theta|\mathbf{x}_1)L(\theta|\mathbf{x}_2)$. Por lo tanto:

$$\begin{aligned}\pi(\theta|\mathbf{x}_1, \mathbf{x}_2) &\propto L(\theta|\mathbf{x}_2)\underbrace{L(\theta|\mathbf{x}_1)\pi(\theta)}_{\pi(\theta|\mathbf{x}_1)} \\ &\propto L(\theta|\mathbf{x}_2)\pi(\theta|\mathbf{x}_1)\end{aligned}$$

Que era lo que queríamos mostrar. Intuitivamente, la independencia entre variables aleatorias nos permite calcular la probabilidad a posteriori de toda la muestra "actualizando" secuencialmente probabilidades a posteriori.

3 Ejercicio 3

3.1 Inciso a

Queremos calcular:

$$\pi(\theta|\mathbf{X}) \propto L(\theta|\mathbf{x})\pi(\theta)$$

Para el modelo Poisson(θ):

$$L(\theta|\mathbf{X}) \propto \theta^{\sum x_i} e^{-n\theta}$$

Por lo tanto, utilizando que $\theta \sim \Gamma(\alpha, \beta)$:

$$\begin{aligned}\pi(\theta|\mathbf{X}) &\propto \theta^{\sum x_i} e^{-n\theta} \frac{\theta^{\alpha-1} e^{-\theta/\beta}}{\Gamma(\alpha)\beta^\alpha} \\ &\propto \theta^{(\alpha+\sum x_i)-1} e^{-\theta(n+1/\beta)} \\ &\propto \theta^{\alpha_n-1} e^{-\theta/\beta_n}\end{aligned}$$

Esta expresión corresponde (por proporcionalidad) a la verosimilitud de una variable $\Gamma(\alpha_n, \beta_n)$. En el segundo paso, podemos eliminar el denominador porque α, β están dados. Para el último paso, notar que:

$$n + \frac{1}{\beta} = \frac{\beta n + 1}{\beta} = \frac{1}{\beta_n}$$

3.2 Inciso b

Para $\theta \sim \Gamma(\alpha_n, \beta_n)$ definida como en la consigna, tenemos:

$$\begin{aligned}E(\theta) &= \alpha_n\beta_n \\ V(\theta) &= \alpha_n\beta_n^2\end{aligned}$$

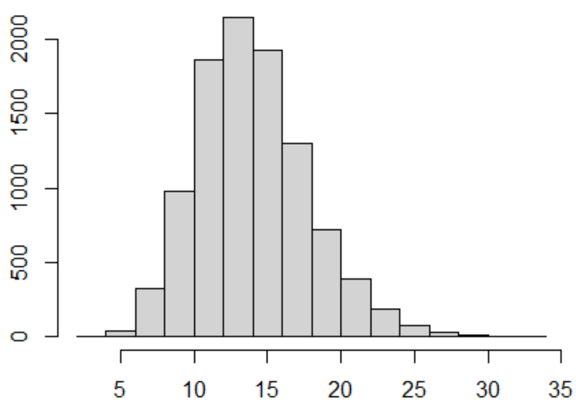
Notar que podemos escribir $\alpha_n = \alpha + n\bar{x}$. Por lo tanto:

$$\begin{aligned}\lim_{n\to\infty} E(\theta) &= (\alpha + n\bar{x})\left(\frac{\beta}{\beta n + 1}\right) = \bar{x} = \hat{\theta}_{MLE} \\ \lim_{n\to\infty} V(\theta) &= (\alpha + n\bar{x})\left(\frac{\beta}{\beta n + 1}\right)^2 = 0\end{aligned}$$

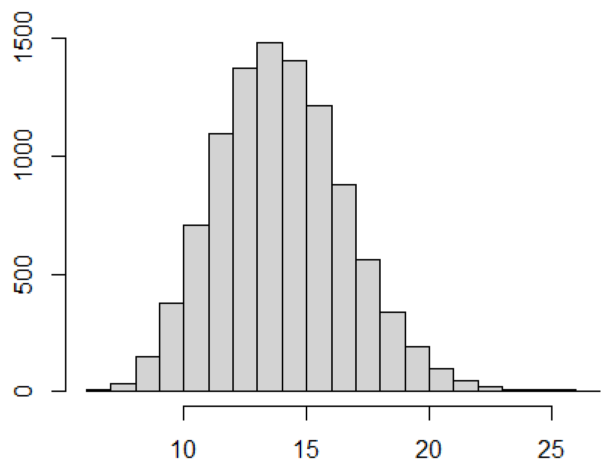
Puede verse entonces que a medida que n crece la distribución a priori "pierde peso" a favor de la información contenida en la muestra.

3.3 Inciso c

A la hora de elegir valores de α, β, querríamos comunicar nuestra creencia previa respecto a la distribución de los posibles valores de la media de crímenes diarios. Dado que observamos $\bar{x} = 14$, resulta razonable alguna combinación de $\{\alpha, \beta\}$ tal que $\alpha * \beta = 14$. No obstante, existen infinitas combinaciones que satisfacen dicha condición, por lo tanto, debemos tomar cierta postura respecto a la varianza del parámetro. Si consideramos que existe un amplio rango de valores posibles para θ, favoreceríamos una mayor valor de β. Presentamos por ejemplo los histogramas asociados a dos muestras generadas con diferentes parametrizaciones:



(a) $\alpha = 14$, $\beta = 1$



(b) $\alpha = 28$, $\beta = 1/2$

La primera parametrización resulta en una media igual pero mayor varianza, por lo tanto, la elegiríamos si considerásemos que existe mayor "incertidumbre" sobre θ. Si quisiesemos una prior cercana a ser "no informativa", elegiríamos un valor de β muy grande en relación a α. Para resolver el resto del ejercicio elegimos arbitrariamente $\alpha = 14$ y $\beta = 1$.

Dado lo visto en el inciso a, sabemos que la distribución a posteriori de θ será $\Gamma(\alpha_n, \beta_n)$. Reemplazando nuestros valores a priori para (α, β) y la información de la muestra:

$$\pi(\theta|\mathbf{x}) \propto \Gamma(\alpha_n = 154, \beta_n = 1/11)$$

3.4 Inciso d

Por lo visto en el inciso b):

$$E(\theta) = \alpha_n \beta_n = \frac{154}{11} = 14$$
$$V(\theta) = \alpha_n \beta_n^2 = \frac{154}{11^2} \approx 1.272$$

3.5 Inciso e

En este inciso, queremos encontrar el intervalo de credibilidad de mayor densidad a posteriori, es decir el *highest posterior density* (HPD) credible set. Este conjunto puede definirse como:

$$HPD = \{\theta \in \Theta : \pi(\theta|\mathbf{x}) \geq \nu_\alpha\}$$

Donde ν_α es el valor más alto tal que:

$$P(\theta \in HPD(\nu_\alpha)|\mathbf{x}) = 1 - \alpha$$

Existen diferentes maneras en las cuales podríamos aproximar el intervalo de credibilidad HPD. Si la distribución es unimodal pero no simétrica, podríamos por ejemplo computar todos los intervalos que acumulan $(1-\alpha)\%$ de las observaciones y elegir el más estrecho[1].

No obstante, si la distribución es unimodal y aproximadamente simétrica (como en este caso) alrededor de la moda M, será suficiente buscar los cuantiles $\alpha/2$ y $(1-\alpha/2)$ de la distribución, ya que estos concentrarán $(1-\alpha)\%$ de los valores posibles con la mayor densidad. En este caso, para una variable $\Gamma(154, 11^{-1})$[2]

$$HPD_{95\%} = [q_{0.025}, q_{0.975}] = [11.876, 16.295]$$
$$HPD_{99\%} = [q_{0.005}, q_{0.995}] = [11.264, 17.076]$$

Bajo el enfoque bayesiano, podemos afirmar que la probabilidad a posteriori condicional en haber observado $\mathbf{x}$ de que θ caiga en el intervalo HPD de cobertura $(1-\alpha)$ es igual a $(1-\alpha)\%$.

4 Ejercicio 4

Una vez más, queremos obtener:

$$\pi(\theta|\mathbf{X}) \propto L(\theta|\mathbf{x})\pi(\theta)$$

El modelo $Ber(\theta)$ se puede representar con la verosimilitud:

$$L(\theta|\mathbf{X}) = \theta^{\sum x_i}(1-\theta)^{n-\sum x_i}$$

Para la distribución uniforme continua, $\pi(\theta) = 1$ para todo θ, por lo tanto:

$$\pi(\theta|\mathbf{X}) \propto \theta^{\sum x_i}(1-\theta)^{n-\sum x_i} = \theta^{n\bar{x}}(1-\theta)^{n(1-\bar{x})}$$

4.1 Inciso a

En este caso, utilizar una prior $U(0,1)$ resultaría *no informativo*, en tanto asignamos probabilidad equivalente a todos los valores posibles de θ. Notar, entonces, que toda la información referente a la distribución de θ contenida en la probabilidad a posteriori es la proveniente de la muestra.

4.2 Inciso b

Recordar que la función de densidad de una variable aleatoria $X \sim \text{Beta}(\alpha, \beta)$ es igual a:

$$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}$$

Donde $B(\alpha,\beta)$ es la función beta incompleta que no depende de x. Por lo tanto, la probabilidad a posteriori calculada anteriormente es equivalente a una distribución $\text{Beta}\,(n\bar{x}+1, n(1-\bar{x})+1)$.

Para una variable aleatoria $X \sim \text{Beta}(\alpha, \beta)$:

$$E(X) = \frac{\alpha}{\alpha+\beta}$$
$$V(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

Por lo tanto, reemplazando α y β por los valores obtenidos anteriormente:

$$E(\theta|\mathbf{X}) = \frac{n\bar{x}+1}{n+2}$$
$$V(\theta|\mathbf{X}) = \frac{n^2\bar{x}(1-\bar{x})+n+1}{(n+2)^2(n+3)}$$

[1] Si quieren ver un ejemplo de esto, adjunto un código de R en el campus.

[2] En `R` puede calcular los cuantiles con la funcion `qgamma`. Recuerde que `R` parametriza el modelo Gamma como $\Gamma(\alpha, 1/\beta)$.

5 Ejercicio 5

$$
\begin{aligned}
\pi(\theta|\mathbf{x}) &\propto \exp\left\{-\frac{1}{2\tau_0^2}(\theta-\mu_0)^2 - \frac{1}{2\sigma_0^2}\sum_{i=1}^{n}(\theta-x_i)^2\right\} \\
&\propto \exp\left\{-\frac{1}{2\tau_0^2}(\theta^2-2\theta\mu_0+\mu_0^2) - \frac{1}{2\sigma_0^2}\sum_{i=1}^{n}(\theta^2-2\theta x_i+x_i^2)\right\} \\
&\propto \exp\left\{-\frac{1}{2\tau_0^2}(\theta^2-2\theta\mu_0+\mu_0^2) - \frac{1}{2\sigma_0^2}\left(n\theta^2-2\theta\sum_{i=1}^{n}x_i+\sum_{i=1}^{n}x_i^2\right)\right\} \\
&\propto \exp\left\{-\frac{1}{2\tau_0^2}(\theta^2-2\theta\mu_0+\mu_0^2) - \frac{1}{2\sigma_0^2}\left(n\theta^2-2\theta n\bar{x}+\sum_{i=1}^{n}x_i^2\right)\right\}
\end{aligned}
$$

Cualquier término dentro del exponencial que no incluya θ puede ser considerado un factor de proporcionalidad, por lo cual:

$$
\begin{aligned}
\pi(\theta|\mathbf{x}) &\propto \exp\left\{-\frac{1}{2\tau_0^2}(\theta^2-2\theta\mu_0) - \frac{n}{2\sigma_0^2}\left(\theta^2-2\theta\bar{x}\right)\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left(\frac{1}{\tau_0^2}(\theta^2-2\theta\mu_0) + \frac{n}{\sigma_0^2}\left(\theta^2-2\theta\bar{x}\right)\right)\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left(\theta^2\frac{\sigma_0^2+n\tau_0^2}{\tau_0^2\sigma_0^2} - 2\theta\frac{\mu_0\sigma_0^2+n\bar{x}\tau_0^2}{\tau_0^2\sigma_0^2}\right)\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[\theta^2\left(\frac{1}{\tau_0^2}+\frac{n}{\sigma_0^2}\right) - 2\theta\left(\frac{\mu_0}{\tau_0^2}+\frac{n\bar{x}}{\sigma_0^2}\right)\right]\right\}
\end{aligned}
$$

Reemplazando las definicionnes de μ_n y σ_n^2:

$$
\pi(\theta|\mathbf{x}) \propto \exp\left\{-\frac{1}{2\sigma_n^2}\left(\theta^2-2\theta\mu_n\right)\right\}
$$

Dado que μ_n y σ_n^2 no dependen de θ, podemos completar el cuadrado con un factor de proporcionalidad:

$$
\begin{aligned}
\pi(\theta|\mathbf{x}) &\propto \exp\left\{-\frac{1}{2\sigma_n^2}\left(\theta^2-2\theta\mu_n+\mu_n^2\right)\right\} \\
&\propto \exp\left\{-\frac{1}{2\sigma_n^2}\left(\theta-\mu_n\right)^2\right\}
\end{aligned}
$$

Que era el resultado que queríamos obtener.

5.1 Inciso a

Notar que:

$$
\begin{aligned}
\lim_{n\to\infty}\mu_n &= \bar{x} \\
\lim_{n\to\infty}\sigma_n^2 &= 0
\end{aligned}
$$

5.2 Inciso b

Este resultado nos indica que a medida que crece n, la distribución a posteriori del parámetro θ se degenera a la información provista en la muestra, en este caso, a través de $\bar{x}$.

5.3 Inciso c

Recordar que la distribución normal resulta ser unimodal, simétrica y que su moda coincide con su esperanza. Entonces, dado que θ sigue una distribución a posteriori $N(\mu_n, \sigma_n)$, podemos afirmar que el intervalo de credibilidad HPD estará dado por:

$$HPD_{95\%} = \mu_n \pm \sigma_n z_{0.975}$$

Donde z_α es el cuantil α de la distribución $N(0, 1)$.

6 Ejercicio 6

6.1 Inciso a

Tal como en escenarios anteriores en los cuales analizamos el desempleo, podemos plantear el parámetro de interés θ como la probabilidad de éxito asociada a una secuencia de variables aleatorias $\{X_i\} \sim \text{Ber}(\theta)$.

Tal como en el ejercicio 3, una opción *no informativa* sería utilizar una distribución $U(0, 1)$ como prior para θ. No obstante, ello resulta poco razonable, dado que sabemos que la tasa de desempleo suele encontrarse entre 5% y el 15%. Por lo tanto, otra alternativa sería utilizar una distribución $U(a, b)$ con límites que consideremos razonables, por ej., en base a datos históricos.

Otra opción sería utilizar una prior $\text{Beta}(\alpha, \beta)$. Esta elección tiene la ventaja de que forma un modelo *conjugado*, lo cual significa que la distribución a posteriori también será un modelo Beta. Además, nos permite ajustar los parámetros α y β para definir cuan informativa queremos que sea nuestra prior. Para el resto de este ejercicio, elegimos esta distribución y trabajamos sin fijar valores de α y β.

6.2 Inciso b

Para $\pi(\theta) = \text{Beta}(\alpha, \beta)$:

$$\pi(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

Por otro lado, la verosimilitud para el modelo Bernouilli es:

$$L(\theta|\mathbf{X}) = \theta^{\sum x_i}(1-\theta)^{n-\sum x_i}$$

Por lo tanto, la distribución a posteriori será:

$$\begin{aligned}\pi(\theta|\mathbf{X}) &\propto \theta^{\alpha+\sum x_i-1}(1-\theta)^{\beta+n-\sum x_i-1} \\ &\propto \theta^{\alpha_n-1}(1-\theta)^{\beta_n-1}\end{aligned}$$

Entonces, $\pi(\theta|\mathbf{X}) = \text{Beta}(\alpha_n, \beta_n)$, con:

$$\begin{aligned}\alpha_n &= \alpha + n\bar{x} \\ \beta_n &= \beta + n(1-\bar{x})\end{aligned}$$

6.3 Inciso c

Dado que sabemos que la distribución a posteriori es Beta(α_n, β_n), podemos expresar sus momentos como:

$$E(\theta|\mathbf{X}) = \frac{\alpha_n}{\alpha_n + \beta_n} = \frac{\alpha + n\bar{x}}{\alpha + \beta + n}$$
$$M(\theta|\mathbf{X}) = \frac{\alpha_n - 1}{\alpha_n + \beta_n - 2} = \frac{\alpha + n\bar{x} - 1}{\alpha + \beta + n - 2}$$
$$V(\theta|\mathbf{X}) = \frac{\alpha_n\beta_n}{(\alpha_n + \beta_n)^2(\alpha_n + \beta_n + 1)} = \frac{(\alpha + n\bar{x})(\beta + n(1 - \bar{x}))}{(\alpha + \beta + n)^2(\alpha + \beta + n + 1)}$$

La expresión utilizada para la moda es válida para valores de α, β mayores a uno, de lo contrario, la distribución sería bimodal.

6.4 Inciso d

Si consideramos que la distribución de la tasa de desempleo es unimodal, elegiríamos valores de α, β mayores a 1 para nuestra prior . No obstante, la distribución β sólo preserva la simetría en tanto $\alpha = \beta$. Por lo tanto, preservar la simetría implica que la distribución se centre alrededor de 1/2, lo cual no resulta razonable para la tasa de desempleo.

No obstante, si graficamos la distribución por ejemplo para $\alpha = 20, \beta = 180$, vemos que la misma es aproximadamente simétrica alrededor del 10%:

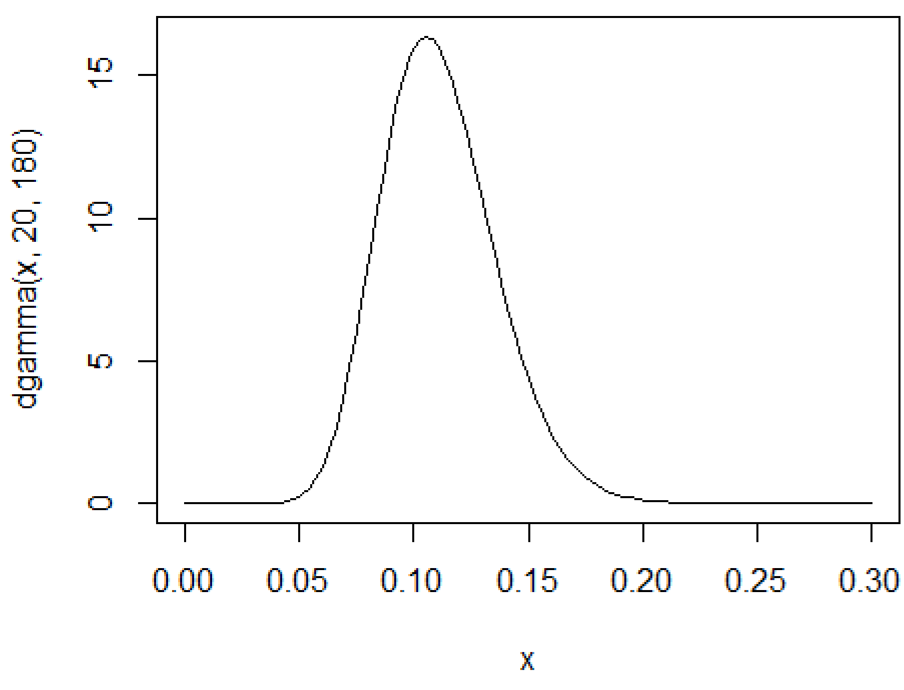


Figure 2: Función de densidad para Beta(20, 180)

Entonces, una posible metodología es elegir valores de α y β tales que la distribución se centre alrededor de algún valor de θ que consideremos razonable. Luego, si escalamos ambos factores en forma equivalente, podemos modificar la dispersión y simetría de la distribución[3].

Una vez elegidos los valores de nuestra prior, podemos calcular los valores a posteriori y analizar si los mismos resultan en una distribución razonablemente simétrica y unimodal. De ser así, podríamos simplemente reportar los cuantiles de la distribución:

$$HPD_{95\%} = [q_{0.025}, q_{0.975}]$$

Alternativamente, si la distribución resulta ser muy asimétrica, podemos aplicar la metodología comentada en el inciso 3.e.

6.5 Inciso e

Si nuestro amigo eligió una prior considerablemente diferente a la nuestra, ya sea por ser más o menos informativa o encontrarse centrada en algún valor diferente de θ, los resultados reportados podrían diferir.

6.6 Inciso f

Al incrementar el tamaño de la muestra, el valor de los parámetros de la distribución a posteriori tenderán a verse más influenciados por lo observado en los datos que por la elección de prior. Por lo tanto, esperaríamos que la "distancia" entra las conclusiones obtenidas sea menor.

[3] Valores más grandes resultarán en una distribución con menor varianza pero más simétrica.