

TRABAJO FIN DE GRADO

Inferencia Bayesiana

Realizado por:
Íñigo Ortiz Padilla

Tutores:
DR. RAFAEL BLANQUERO BRAVO
DR. EMILIO CARRIZOSA PRIEGO

FACULTAD DE MATEMÁTICAS
Departamento de Estadística e Investigación Operativa
Sevilla, 2018

Índice general

Abstract	5
1. Conceptos previos	7
1.1. Propiedades de la inferencia bayesiana	7
1.2. Fórmula de Bayes	7
1.2.1. Formulación del teorema	7
1.2.2. Fórmula de Bayes y función de verosimilitud	8
1.2.3. Naturaleza secuencial de la fórmula de Bayes	9
1.3. Probabilidad subjetiva	9
1.4. Procedimiento a seguir	10
2. Herramientas de la Inferencia Bayesiana	13
2.1. Estimación bayesiana	13
2.1.1. Estimación puntual	13
2.1.2. Intervalos	15
2.2. Elección de la distribución a priori	18
2.2.1. Distribuciones a priori conjugadas	18
2.2.2. Distribuciones a priori impropias	19
2.2.3. Distribuciones a priori no informativas: Regla de Jeffreys	19
2.3. Distribución a posteriori predictiva	21
3. Inferencia bayesiana sobre una proporción	23
3.1. Cálculo de la distribución a posteriori	23
3.1.1. Distribución a priori discreta para \tilde{p}	23
3.1.2. Distribución a priori continua para \tilde{p}	24
3.1.2.1. Distribución continua a priori para \tilde{p}	24
3.1.2.2. Distribución continua uniforme para \tilde{p}	25
3.1.3. Distribución Beta a priori	27
3.2. Regla de Jeffreys	28
3.3. Distribución predictiva	28
3.4. Contraste de hipótesis	29

4. Distribución de Poisson	31
4.1. Función de verosimilitud	31
4.2. Cálculo de la distribución a posteriori	32
4.2.1. Distribución a priori no informativa	33
4.2.2. Distribución Gamma a priori para Poisson $\tilde{\lambda}$	33
4.3. Distribución predictiva	35
4.4. Contraste de hipótesis	35
5. Distribución normal	37
5.1. Función de verosimilitud para (μ, σ^2)	37
5.2. Regla de Jeffreys	39
5.2.1. Caso $\tilde{\mu}$ desconocido y σ conocido	40
5.2.2. Caso μ conocido, $\tilde{\sigma}$ desconocido	40
5.3. Cálculo de la distribución a posteriori	41
5.3.1. Caso $\tilde{\mu}$ desconocido y σ^2 conocido	41
5.3.1.1. Distribución a priori no informativa	41
5.3.1.2. Distribución a priori conjugada para $\tilde{\mu}$	42
5.3.2. Caso μ conocido y $\tilde{\sigma}^2$ desconocido	42
5.3.2.1. Distribución a priori no informativa	42
5.3.2.2. Distribución a priori conjugada	47
5.3.3. Caso $\tilde{\mu}, \tilde{\sigma}$ desconocidos	47
5.3.3.1. Distribución a priori no informativa	47
5.3.3.2. Distribución conjugada	51
5.4. Distribución predictiva	52
5.5. Test de hipótesis	53
5.5.1. Media de una normal con varianza desconocida	53
5.5.2. Varianza	54
6. Inferencia Bayesiana en R	55
6.1. Inferencia sobre una proporción	55
6.2. Distribución de Poisson	57
6.3. Distribución Normal	58
6.4. Ejemplo práctico	60

Abstract

The subject of this document is Bayesian Inference, an inference system based on Bayes' Formula.

In the first chapter we will state this formula and will discuss how to use it. It will be shown that, according to the formula, posterior distributions are proportional to the likelihood function times the prior distribution. This is the essential notion of Bayesian inference. In the second chapter, definitions and results related to Bayesian Analysis will be given in order to accomplish our inference. These mathematical tools will be of great use in the following chapters, where inference on proportions, Poisson distribution and Normal distribution will be studied. Both conjugate and noninformative prior distributions are considered.

This documents ends with the transcription of an R code, allowing us to compute posterior distributions.

Capítulo 1

Conceptos previos

1.1. Propiedades de la inferencia bayesiana

La inferencia bayesiana es una estrategia de inferencia estadística que tiene su base en el teorema de Bayes. Algunas de sus propiedades, enumeradas a continuación, la hacen particularmente útil para ser usada en la investigación científica.

- Dado un modelo, la inferencia bayesiana hace uso de los datos obtenidos empíricamente.
- Inferencias que resulten, según nuestros propósitos, inaceptables han de ser consecuencia de afirmaciones inapropiadas, no de un inadecuado sistema de inferencia. Por tanto, todos los factores del modelo, incluidas las distribuciones a priori, son susceptibles de ser modificadas.
- Ya que este sistema de inferencia puede ser aplicado a casi cualquier modelo de probabilidad, no se requiere prestar tanta atención a la conveniencia del modelo considerado como al valor científico de éste.

1.2. Fórmula de Bayes

Nos detendremos ahora en analizar un importante teorema que da lugar a la materia de estudio de este trabajo. Aparecido en un artículo del 1764 firmado por Thomas Bayes, este resultado es designado de manera usual como fórmula de Bayes.

1.2.1. Formulación del teorema

Sea $\underline{x} = (x_1, \dots, x_n)$ un vector de n observaciones cuya distribución de probabilidad $P(\underline{x} | \underline{\theta})$ depende de los valores de k parámetros $\underline{\theta}' = (\theta_1, \dots, \theta_k)$. Supongamos además que $\underline{\theta}$ tiene distribución de probabilidad $P(\underline{\theta})$. Entonces se tiene:

$$P(\underline{x}|\underline{\theta})P(\underline{\theta}) = P(\underline{x}, \underline{\theta}) = P(\underline{\theta}|\underline{x})P(\underline{x})$$

Dado el vector de datos observados \underline{x} , la distribución condicionada de $\underline{\theta}$ a \underline{x} es la siguiente:

$$P(\underline{\theta}|\underline{x}) = \frac{P(\underline{x}|\underline{\theta})P(\underline{\theta})}{P(\underline{x})}.$$

Además podemos escribir:

$$P(\underline{x}) = E[P(\underline{x}|\underline{\theta})] = c^{-1}.$$

Para los casos continuo y discreto de θ tenemos entonces, respectivamente:

$$P(\underline{x}) = \int P(\underline{x}|\underline{\theta})P(\underline{\theta})d\underline{\theta}$$

$$P(\underline{x}) = \sum P(\underline{x}|\underline{\theta})P(\underline{\theta})$$

donde la suma y la integral están definidas en el dominio de $\underline{\theta}$, y la esperanza está tomada con respecto a la distribución $P(\underline{\theta})$. Podemos escribir en consecuencia

$$P(\underline{\theta}|\underline{x}) = cP(\underline{x}|\underline{\theta})P(\underline{\theta}).$$

En la expresión precedente, $P(\underline{\theta})$, conocida como distribución **a priori** de $\underline{\theta}$, nos dice qué sabemos de $\underline{\theta}$ sin conocimiento de los datos muestrales, mientras que $P(\underline{\theta}|\underline{x})$ nos dice qué sabemos de $\underline{\theta}$ una vez conocidos los datos muestrales, y es por ello llamada distribución **a posteriori** de $\underline{\theta}$ dado \underline{x} . Por otra parte, la constante c tiene como utilidad normalizar la suma o la integral para que sea igual a uno.

1.2.2. Fórmula de Bayes y función de verosimilitud

Supongamos ahora que conocemos los datos muestrales \underline{x} ; entonces $P(\underline{x}|\underline{\theta})$ puede ser considerada una función de $\underline{\theta}$. Si es éste el caso, entonces dicha función de probabilidad es la función de máxima verosimilitud de $\underline{\theta}$ dado \underline{x} y se escribe $L(\underline{\theta}|\underline{x})$. Podemos por tanto escribir la fórmula de Bayes como

$$P(\underline{\theta}|\underline{x}) = cL(\underline{\theta}|\underline{x})P(\underline{\theta}).$$

Esto quiere decir que la fórmula de Bayes nos asegura que la distribución de probabilidad de $\underline{\theta}$ conocidos los datos \underline{x} es proporcional al producto de la distribución de $\underline{\theta}$ a priori y la función de verosimilitud de $\underline{\theta}$ dado \underline{x} :

$$\text{distribución a posteriori} \propto \text{verosimilitud} \times \text{distribución a priori}.$$

La función de verosimilitud $L(\underline{\theta}|\underline{x})$ juega un importante papel en la fórmula de Bayes, pues es la función que determina cómo los datos \underline{x} modifican a $\underline{\theta}$. Puede ser, por tanto, considerada como una representación de la información sobre $\underline{\theta}$ obtenida a partir de los datos observados.

La función de verosimilitud está definida salvo constante multiplicativa. Esto es coherente con el lugar que ocupa en la fórmula de Bayes, ya que multiplicar la función de verosimilitud por una constante arbitraria no afecta a la distribución a posteriori de $\underline{\theta}$.

1.2.3. Naturaleza secuencial de la fórmula de Bayes

La fórmula de Bayes resulta especialmente útil porque traduce a formulación matemática cómo los conocimientos previos de un suceso o experimento pueden combinarse con nuevos datos. La fórmula permite, en efecto, actualizar de forma continuada la información disponible acerca de los parámetros conforme se van tomando observaciones muestrales.

Supongamos que tenemos una muestra inicial de observaciones \underline{x}_1 ; entonces la fórmula de Bayes se escribe

$$P(\underline{\theta} | \underline{x}_1) \propto P(\underline{\theta})L(\underline{\theta} | \underline{x}_1).$$

Supongamos ahora que disponemos de una segunda muestra de observaciones \underline{x}_2 , independientes de la primera muestra. Entonces

$$\begin{aligned} P(\underline{\theta} | \underline{x}_2, \underline{x}_1) &\propto P(\underline{\theta})L(\underline{\theta} | \underline{x}_1)L(\underline{\theta} | \underline{x}_2) \\ &\propto P(\underline{\theta} | \underline{x}_1)L(\underline{\theta} | \underline{x}_2). \end{aligned}$$

En la expresión precedente, la distribución a posteriori de $\underline{\theta}$ dado \underline{x}_1 , $P(\underline{\theta} | \underline{x}_1)$, hace de distribución a priori de la segunda muestra. Es obvio que este proceso puede repetirse tantas veces como se quiera. En particular, si tenemos n observaciones independientes, la distribución a posteriori puede ser recalculada con cada nueva observación, por lo que en el paso m la verosimilitud asociada con las m observaciones aparece combinada con la distribución a posteriori de $\underline{\theta}$ después de $m - 1$ observaciones, dando como resultado la distribución a posteriori siguiente:

$$P(\underline{\theta} | \underline{x}_1, \dots, \underline{x}_m) \propto P(\underline{\theta} | \underline{x}_1, \dots, \underline{x}_{m-1})L(\underline{\theta} | \underline{x}_m), \quad m = 2, \dots, n$$

donde

$$P(\underline{\theta} | \underline{x}_1) \propto P(\underline{\theta})L(\underline{\theta} | \underline{x}_1).$$

En consecuencia, la fórmula de Bayes describe el proceso de aprendizaje a partir de la experiencia, y muestra cómo el conocimiento del estado del mundo físico representado por los parámetros está continuamente en cambio según aparecen nuevos datos.

1.3. Probabilidad subjetiva

Definición Probabilidad subjetiva: Una probabilidad de que un evento ocurra es un número entre cero y uno que cuantifica la opinión particular (subjetiva) de una persona acerca de cuán probable es que este evento ocurra (o haya ocurrido).

Esta interpretación de probabilidad no está limitada a eventos que se repiten. Observemos que la interpretación subjetiva era de “una”, no “la” probabilidad de un evento. Por tanto, no sólo diferentes personas tendrán diferentes probabilidades subjetivas de un mismo suceso, sino que incluso la misma persona puede cambiar su probabilidad subjetiva al tiempo que se obtiene más información (es entonces cuando la fórmula de Bayes entra en juego).

Existe controversia acerca de qué papel puede tener un concepto como la probabilidad subjetiva en la ciencia. Sin embargo, no deja de ser cierto que dos científicos pueden examinar los mismos datos y llegar a conclusiones diferentes de acuerdo a su conocimiento previo de la materia en cuestión.

1.4. Procedimiento a seguir

Discutiremos brevemente el procedimiento a seguir a la hora de estudiar un problema mediante la Inferencia Bayesiana.

Hipótesis y probabilidad a priori

Consideremos un problema sencillo:

- El suceso A ocurre
- El suceso A no ocurre

Enunciados de esta forma se conocen como hipótesis: afirmaciones acerca de un aspecto concreto del mundo real.

Antes de obtener datos concernientes a este caso particular, sería lógico pensar que la probabilidad del suceso A debería ser similar a otros casos de la misma situación ocurridos con anterioridad a nuestro objeto de estudio. Para ello podemos tomar aleatoriamente un número grande de dichas situaciones y comprobar en cuántas se produjo el suceso A. Haciendo esto estaremos asignando a nuestras dos hipótesis las probabilidades a priori de que ocurran.

Datos

Recogemos los datos relacionados con el problema (ejemplos: temperatura, presión, edad, peso,...). ¿Cómo afectan los datos al suceso A? Si disponemos de resultados y experiencias anteriores, podemos estudiar la probabilidad de que ocurra A dependiendo de los valores de los datos (probabilidad condicionada).

Verosimilitud y probabilidad a posteriori

La fórmula de Bayes es una herramienta matemática que nos permite utilizar los datos para actualizar las probabilidades de nuestro modelo de forma rigurosa y sistemática. La verosimilitud asociada a un modelo es la probabilidad de que sucedan los datos observados bajo las hipótesis del modelo. La fórmula de Bayes combina, como ya hemos visto en un apartado anterior, la probabilidad a priori con la verosimilitud para establecer la probabilidad a posteriori. Esta probabilidad sólo está disponible después de que los datos hayan sido observados.

Capítulo 2

Herramientas de la Inferencia Bayesiana

En este capítulo estudiaremos conceptos propios de la Inferencia Bayesiana que en los capítulos posteriores nos servirán para adaptar distribuciones conocidas al caso bayesiano. Primero veremos cómo estimar de forma puntual y cómo obtener intervalos, para después pasar a estudiar las distintas opciones disponibles al escoger la distribución a priori, cada una de las cuales dará como resultado una distribución a posteriori diferente.

2.1. Estimación bayesiana

2.1.1. Estimación puntual

Desde el punto de vista Bayesiano, el cálculo de un estimador $\hat{\theta}$ de un parámetro θ está basado en la distribución a posteriori. Lo más conveniente es estimar parámetros de localización, como la media o la moda.

- Esperanza: $E(\theta | \underline{x}) = \int_{-\infty}^{\infty} \theta f(\theta | \underline{x}) d\theta$
- Moda: $Mod(\theta | \underline{x}) = arg \max_{\theta} f(\theta | \underline{x})$
- Mediana: el valor a que satisface:

$$\int_{-\infty}^a f(\theta | \underline{x}) d\theta = \frac{1}{2} \text{ y } \int_a^{\infty} f(\theta | \underline{x}) d\theta = \frac{1}{2}$$

Ahora bien, debemos preguntarnos cuál estimación tomar en cada caso específico. Para responder a esto, definiremos la función de pérdida.

Función de pérdida: La función de pérdida $\mathcal{L}(\hat{\theta}, \theta) \in \mathbb{R}$ cuantifica el error obtenido a la hora de estimar el parámetro θ con el estimador $\hat{\theta}$.

Si el estimador es igual al parámetro, entonces a la función de pérdida se le asocia el valor cero: $\mathcal{L}(\hat{\theta}, \theta) = 0$. Definamos ahora algunas funciones de pérdida usadas comúnmente:

- Función de pérdida cuadrática: $\mathcal{L}(\hat{\theta}, \theta) = (\hat{\theta} - \theta)^2$
- Función de pérdida lineal: $\mathcal{L}(\hat{\theta}, \theta) = |\hat{\theta} - \theta|$
- Función de pérdida cero-uno: $\mathcal{L}_\varepsilon(\hat{\theta}, \theta) = \begin{cases} 0, & \text{si } |\hat{\theta} - \theta| \leq \varepsilon \\ 1, & \text{si } |\hat{\theta} - \theta| > \varepsilon \end{cases}$ donde ε es un parámetro fijado.

Ahora elegiremos un estimador puntual $\hat{\theta}$ tal que minimice el error a posteriori esperado con respecto a $f(\theta | \underline{x})$. Este estimador se llama estimador de Bayes.

Definición estimador de Bayes: Un estimador de Bayes de θ con respecto a la función de pérdida $\mathcal{L}(\hat{\theta}, \theta)$ minimiza el error esperado con respecto a la distribución a posteriori $f(\theta | \underline{x})$, es decir, minimiza

$$E[\mathcal{L}(\hat{\theta}, \theta) | \underline{x}] = \int_{-\infty}^{\infty} \mathcal{L}(\hat{\theta}, \theta) f(\theta | \underline{x}) d\theta.$$

Se tienen los siguientes resultados:

Rest. 1: la media a posteriori es el estimador de Bayes con respecto a la función de pérdida cuadrática.

Prueba: Hagamos la prueba para el caso absolutamente continuo. En este caso la esperanza del error cuadrático es

$$E[\mathcal{L}(\hat{\theta}, \theta) | \underline{x}] = \int_{-\infty}^{+\infty} \mathcal{L}(\hat{\theta}, \theta) f(\theta | \underline{x}) d\theta = \int_{-\infty}^{+\infty} (\hat{\theta} - \theta)^2 f(\theta | \underline{x}) d\theta.$$

Derivando con respecto al estimador e igualando la derivada a 0 obtenemos:

$$2 \int_{-\infty}^{+\infty} (\hat{\theta} - \theta) f(\theta | \underline{x}) d\theta = 0 \iff \hat{\theta} - \int_{-\infty}^{+\infty} \theta f(\theta | \underline{x}) d\theta = 0.$$

Y esto conduce a $\hat{\theta} = \int_{-\infty}^{\infty} \theta f(\theta | \underline{x}) d\theta = E(\theta | \underline{x})$.

Rest. 2: la mediana a posteriori es el estimador de Bayes con respecto a la función de pérdida lineal.

Prueba: De nuevo haremos la prueba para el caso particular en el que la variable aleatoria es absolutamente continua. Escribimos el error lineal esperado:

$$\begin{aligned} E[\mathcal{L}(\hat{\theta}, \theta) | \underline{x}] &= \int_{-\infty}^{+\infty} \mathcal{L}(\hat{\theta}, \theta) f(\theta | \underline{x}) d\theta = \int_{-\infty}^{+\infty} |\hat{\theta} - \theta| f(\theta | \underline{x}) d\theta \\ &= \int_{\theta \leq \hat{\theta}} (\hat{\theta} - \theta) f(\theta | \underline{x}) d\theta + \int_{\theta > \hat{\theta}} (\theta - \hat{\theta}) f(\theta | \underline{x}) d\theta. \end{aligned}$$

Calculamos la derivada con respecto al estimador usando la regla de Leibniz para integrales:

$$\begin{aligned} \frac{\partial}{\partial \hat{\theta}} E[\mathcal{L}(\hat{\theta}, \theta) | \underline{x}] &= \frac{\partial}{\partial \hat{\theta}} \int_{-\infty}^{\hat{\theta}} (\hat{\theta} - \theta) f(\theta | \underline{x}) d\theta + \frac{\partial}{\partial \hat{\theta}} \int_{\hat{\theta}}^{+\infty} (\theta - \hat{\theta}) f(\theta | \underline{x}) d\theta \\ &= \int_{-\infty}^{\hat{\theta}} f(\theta | \underline{x}) d\theta - (\hat{\theta} - (-\infty)) f(-\infty | \underline{x}) \cdot 0 + (\hat{\theta} - \hat{\theta}) f(\hat{\theta} | \underline{x}) \cdot 1 \\ &\quad - \int_{\hat{\theta}}^{+\infty} f(\theta | \underline{x}) d\theta - (\hat{\theta} - \hat{\theta}) f(\hat{\theta} | \underline{x}) \cdot 1 + (\infty - \hat{\theta}) f(\infty | \underline{x}) \cdot 0 \\ &= \int_{-\infty}^{\hat{\theta}} f(\theta | \underline{x}) d\theta - \int_{\hat{\theta}}^{+\infty} f(\theta | \underline{x}) d\theta. \end{aligned}$$

Igualando a cero tenemos que la solución debe ser $\hat{\theta} = \text{Med}(\theta | \underline{x})$.

Rest. 3: la moda a posteriori es el estimador de Bayes con respecto a la función de pérdida cero-uno, cuando $\varepsilon \rightarrow 0$.

Prueba: Estudiemos el caso absolutamente continuo. El error esperado de acuerdo a la función de pérdida cero-uno es

$$\begin{aligned} E[\mathcal{L}(\hat{\theta}, \theta) | \underline{x}] &= \int_{-\infty}^{+\infty} \mathcal{L}_\varepsilon(\hat{\theta}, \theta) f(\theta | \underline{x}) d\theta \\ &= \int_{-\infty}^{\hat{\theta}-\varepsilon} f(\theta | \underline{x}) d\theta + \int_{\hat{\theta}+\varepsilon}^{+\infty} f(\theta | \underline{x}) d\theta \\ &= 1 - \int_{\hat{\theta}-\varepsilon}^{\hat{\theta}+\varepsilon} f(\theta | \underline{x}) d\theta. \end{aligned}$$

Se alcanzará el mínimo cuando la integral alcanza el máximo. Para valores pequeños de ε la integral es aproximadamente $2\varepsilon f(\hat{\theta} | \underline{x})$, que alcanza el máximo en la moda a posteriori $\hat{\theta} = \text{Mod}(\theta | \underline{x})$.

2.1.2. Intervalos

Al igual que la estimación puntual, la estimación de intervalos en la inferencia Bayesiana se obtiene a partir de la distribución a posteriori. Para distinguirlos de los intervalos de confianza, que tienen una interpretación diferente, los llamaremos intervalos creíbles.

Definición intervalo creíble: para un $\gamma \in (0, 1)$ fijo, un $\gamma \cdot 100\%$ intervalo creíble para ϑ está definido como intervalo de la forma (t_l, t_u) donde t_l, t_u verifican

$$\int_{t_l}^{t_u} f(\theta | \underline{x}) d\theta = \gamma.$$

La cantidad γ es el **nivel creíble** del intervalo creíble (t_l, t_u) .

Esta definición implica que la variable aleatoria $\theta|\underline{x}$ está contenida en un $\gamma \cdot 100\%$ intervalo creíble con probabilidad γ . La manera más sencilla de construir intervalos creíbles es la de elegir t_l como el cuantil $(1 - \gamma)/2$ y t_u como el cuantil $(1 + \gamma)/2$ de la distribución a posteriori.

Definición intervalos de Máxima Densidad de Probabilidad: Sea $\gamma \in (0, 1)$ un nivel creíble fijo. Un intervalo creíble $I = (t_l, t_u)$ se llama intervalo de Máxima Densidad de Probabilidad, intervalo M.D.P. (*Highest Probability Density* en inglés), si verifica

$$f(\theta|\underline{x}) \geq f(\tilde{\theta}|\underline{x})$$

para todo $\theta \in I$ y todo $\tilde{\theta} \notin I$.

Un intervalo M.D.P. contiene todos los valores del parámetro que tienen mayor densidad de probabilidad a posteriori que aquellos valores del parámetro no contenidos en el intervalo. Bajo ciertas condiciones, la función de densidad a posteriori es igual para los límites del intervalo M.D.P, es decir, $f(t_l|\underline{x}) = f(t_u|\underline{x})$.

Existen contraejemplos en que esto último no se cumple; por ejemplo, cuando la distribución a posteriori es exponencial. El límite menor t_l de un intervalo M.D.P. a cualquier nivel será cero puesto que la densidad de una exponencial es monótona decreciente.

Para espacios de parámetros discretos Θ , la definición de intervalo creíble debe ser modificada ya que no siempre será posible encontrar un intervalo con el exacto nivel creíble γ buscado. En consecuencia, un $\gamma \cdot 100\%$ intervalo creíble $I = (t_l, t_u)$ para el parámetro θ estará definido como

$$\sum_{\theta \in I \cap \Theta} f(\theta|\underline{x}) \geq \gamma.$$

Surge ahora la cuestión de si, al igual que en la estimación puntual, los intervalos creíbles son óptimos con respecto a una cierta función de pérdida. Por simplicidad, asumiremos que el parámetro $\theta \in \Theta$ es un escalar con densidad a posteriori $f(\theta|\underline{x})$. Primero introducimos el concepto de región creíble, una generalización del intervalo creíble. De forma similar puede ser definida una región M.D.P.

Definición Región creíble: Un subconjunto $C \subseteq \Theta$ con

$$\int_C f(\theta|\underline{x})d\theta = \gamma.$$

se llama $\gamma \cdot 100\%$ región creíble para θ con respecto a $f(\theta | \underline{x})$. Si C es un intervalo real, entonces C también se llama intervalo creíble.

No existe una única $\gamma \cdot 100\%$ región creíble para cada γ fijo. Es lógico pensar que la unicidad se obtiene al especificar una cierta función de pérdida $\mathcal{L}(C, \theta)$. La siguiente función de pérdida tiene en cuenta (para cada γ) la medida de Lebesgue $|C|$ de la región creíble, de forma que regiones pequeñas serán preferibles. Además, la función de pérdida también da resultado desfavorable en el caso de que el parámetro real θ no esté contenido en la región creíble. Restringiremos nuestra atención a distribuciones a posteriori continuas.

Proposición: Sea $f(\theta | \underline{x})$ la función de densidad a posteriori de θ y sea, para cada $\gamma \in (0, 1)$ fijo,

$$\mathcal{A} = \{C | P(\theta \in C | \underline{x}) = \gamma\}$$

el conjunto de todas las $\gamma \cdot 100\%$ regiones creíbles para θ . Consideremos ahora la función de pérdida

$$\mathcal{L}(C, \theta) = |C| - I_C(\theta) \text{ para } C \in \mathcal{A}, \theta \in \Theta.$$

Entonces, C es óptimo con respecto a $\mathcal{L}(C, \theta)$ si y sólo si, para todo $\theta_1 \in C$ y $\theta_2 \notin C$,

$$f(\theta_1 | \underline{x}) \geq f(\theta_2 | \underline{x}),$$

es decir, si y sólo si C es una región M.D.P.

Prueba: consideremos un conjunto $C \in \mathcal{A}$ con pérdida esperada

$$\int_{\Theta} \mathcal{L}(C, \theta) f(\theta | \underline{x}) d\theta = |C| - \int_{\Theta} I_C(\theta) f(\theta | \underline{x}) d\theta = |C| - P(\theta \in C | \underline{x}) = |C| - \gamma.$$

Para un γ fijo, la $\gamma \cdot 100\%$ región creíble C con medida más pequeña $|C|$ minimizará por tanto la pérdida esperada. Sea ahora $C \in \mathcal{A}$ verificando $f(\theta_1 | \underline{x}) \geq f(\theta_2 | \underline{x})$ para todo $\theta_1 \in C$ y $\theta_2 \notin C$, y sea $D \in \mathcal{A}$ otro elemento de \mathcal{A} . Tenemos que probar que $|C| \leq |D|$. Sea $A \dot{\cup} B$ la unión disjunta de A y B ($A \dot{\cup} B = A \cup B$ y $A \cap B = \emptyset$). Entonces:

$$C = C \cap \Theta = C \cap (D \dot{\cup} D^c) = (C \cap D) \dot{\cup} (C \cap D^c) \\ \text{y análogamente } D = (C \cap D) \dot{\cup} (C^c \cap D).$$

Queda por tanto probar que $|C \cap D^c| \leq |C^c \cap D|$. Como C es una región M.D.P., se tiene que

$$\sup_{C \cap D} f(\theta | \underline{x}) \leq \inf_{C \cap D^c} f(\theta | \underline{x}),$$

y ya que $C, D \in \mathcal{A}$ se obtiene

$$P(\theta \in C | \underline{x}) = P(\theta \in D | \underline{x}) = \gamma.$$

Usando la precedente descomposición de C y D , obtenemos también lo siguiente:

$$\begin{aligned} P(\theta \in C \cap D | \underline{x}) + P(\theta \in C \cap D^c | \underline{x}) &= P(\theta \in C | \underline{x}) \\ &= P(\theta \in D | \underline{x}) = P(\theta \in C \cap D | \underline{x}) + P(\theta \in C^c \cap D | \underline{x}) \end{aligned}$$

y en consecuencia

$$P(\theta \in C \cap D^c | \underline{x}) = P(\theta \in C^c \cap D | \underline{x}).$$

En definitiva, obtenemos

$$\begin{aligned} \inf_{C \cap D^c} f(\theta | \underline{x}) \cdot |C \cap D^c| &\leq \int_{C \cap D^c} f(\theta | \underline{x}) d\theta \\ &= \int_{C^c \cap D} f(\theta | \underline{x}) d\theta \leq \sup_{C^c \cap D} f(\theta | \underline{x}) \cdot |C^c \cap D| \leq \inf_{C \cap D^c} f(\theta | \underline{x}) \cdot |C^c \cap D|. \end{aligned}$$

Y por tanto, $|C \cap D^c| \leq |C^c \cap D|$ y se tiene $|C| \leq |D|$. La prueba en la otra dirección es similar.

2.2. Elección de la distribución a priori

En la inferencia Bayesiana es fundamental elegir correctamente la distribución a priori, puesto que a través de ella expresamos matemáticamente las ideas previas respecto al modelo. Por tanto, si se dispone de información sobre los datos o el modelo, ésta debe usarse a la hora de construir la distribución a priori. En ciertas situaciones, sin embargo, la elección es algo delicada debido a la ausencia de información previa al modelo. Por otra parte, a menudo es útil restringir el rango de distribuciones a priori a una familia de uno o dos parámetros.

2.2.1. Distribuciones a priori conjugadas

Una solución pragmática al problema de elección de la distribución a priori consiste en seleccionar un miembro de una familia de distribuciones tal que la distribución a posteriori pertenezca a la misma familia.

Definición distribución a priori conjugada: Sea $L(\theta) = f(\underline{x} | \theta)$ una función de verosimilitud basada en la observación $\underline{X} = \underline{x}$. Una clase \mathcal{G} de distribuciones se llama conjugada con respecto a $L(\theta)$ si la distribución a posteriori $f(\theta | \underline{x})$ está en \mathcal{G} para todo \underline{x} siempre que la distribución a priori $f(\theta)$ esté en \mathcal{G} .

La familia $\mathcal{G} = \{\text{todas las distribuciones}\}$ es trivialmente conjugada con respecto a cualquier función de verosimilitud. En la práctica trataremos de encontrar conjuntos \mathcal{G} más pequeños específicos para la verosimilitud $L_{\underline{x}}(\theta)$.

Notemos que es suficiente estudiar si es conjugado un miembro X_i de la muestra aleatoria $X_{1:n}$. En efecto, si la distribución a priori es conjugada, la distribución a posteriori después de haber observado la primera observación es, por definición, del mismo tipo y sirve como la nueva distribución a priori para la próxima observación. La nueva distribución a posteriori, incorporando ahora la segunda observación, pertenece de nuevo a la clase de conjugación. Este proceso de asimilación de datos continúa secuencialmente y sólo cambia el valor de los parámetros de la distribución.

2.2.2. Distribuciones a priori impropias

La distribución a priori tiene influencia sobre la distribución a posteriori. Si se desea minimizar dicha influencia, entonces se puede especificar una distribución a priori imprecisa, por ejemplo una con varianza muy grande. En un caso límite, esto puede conducir a una distribución a priori impropia, con función de densidad cuya integral no es igual a la unidad. Usando estas distribuciones siempre es necesario comprobar que al menos la distribución a posteriori es propia.

Definición distribución a priori impropia: Una distribución a priori con función de densidad $f(\theta) \geq 0$ se llama impropia si

$$\int_{\Theta} f(\theta) d\theta = \infty \quad \sum_{\theta \in \Theta} f(\theta) = \infty$$

para parámetros continuos y discretos, respectivamente.

2.2.3. Distribuciones a priori no informativas: Regla de Jeffreys

Las distribuciones a priori no informativas son útiles cuando deseamos que la inferencia no se vea afectada por información que no provenga de los datos presentes. Estas distribuciones son también apropiadas cuando tenemos muy poco conocimiento previo en comparación con la información contenida en los nuevos datos (en estos casos, es de esperar que la verosimilitud domine a la distribución a priori).

Para establecer una distribución a priori no informativa, la opción más obvia es escoger una distribución a priori localmente uniforme tal que $f_{\theta}(\theta) \propto 1$, en cuyo caso la distribución a posteriori es proporcional a la función de verosimilitud. Si el espacio de parámetros no está acotado, corremos el riesgo de que la distribución localmente uniforme sea impropia.

Ahora bien, surgen problemas en este planteamiento. Supongamos que $h(\theta)$ es una transformación inyectiva y diferenciable del parámetro θ , que tiene distribución a priori localmente uniforme con densidad $f_{\theta}(\theta) \propto 1$. Usando un cambio de variables, obtenemos la correspondiente distribución a priori para ϕ

$$f_\phi(\phi) = f_\theta\{h^{-1}(\phi)\} \cdot \left| \frac{dh^{-1}(\phi)}{d\phi} \right| \propto \left| \frac{dh^{-1}(\phi)}{d\phi} \right|.$$

Este término no es necesariamente constante. En efecto, $f_\phi(\phi)$ será independiente de ϕ únicamente si h es lineal. Si h no lo es, la función de densidad a priori $f_\phi(\phi)$ dependerá de θ y no será, por tanto, localmente uniforme. Aun así, si hemos elegido una parametrización en función de ϕ desde el comienzo, podemos también haber elegido una distribución a priori localmente uniforme $f_\phi(\phi) \propto 1$. Que no exista invarianza bajo reparametrización es un inconveniente de este tipo de distribuciones a priori. Hagamos notar que asumimos implícitamente que podemos aplicar el cambio de variables a funciones de densidad impropias de la misma manera que a las propias.

A pesar de todo lo dicho, existe un caso particular de elección de la distribución a priori que es invariante bajo reparametrización: se trata de la distribución a priori de Jeffreys.

Definición distribución a priori de Jeffreys: Sea X una variable aleatoria con función de verosimilitud $f(x|\theta)$ donde θ es un parámetro escalar desconocido. Entonces la distribución a priori de Jeffreys está definida como

$$f(\theta) \propto \sqrt{J(\theta)}$$

donde $J(\theta)$ es la información de Fisher esperada de θ . La ecuación enunciada también se conoce como Regla de Jeffreys.

La distribución de Jeffreys es proporcional a la raíz cuadrada de la información de Fisher esperada, lo cual puede proporcionar una distribución a priori impropia. A primera vista resulta sorprendente que esta elección sea invariante bajo reparametrización, pero el siguiente resultado nos confirma que, en efecto, así es.

Proposición: Invarianza de la distribución a priori de Jeffreys

La distribución a priori de Jeffreys es invariante bajo parametrizaciones inyectivas de θ : Si

$$f_\theta(\theta) \propto \sqrt{J_\theta(\theta)}$$

entonces la función de densidad $\phi = h(\theta)$ es

$$f_\phi(\phi) \propto \sqrt{J_\phi(\phi)}$$

Para demostrarlo, usaremos el cambio de variables y también el siguiente resultado:

Información de Fisher esperada de una transformación

Sea $J_\theta(\theta)$ la información de Fisher esperada de un parámetro escalar θ , y $\phi = h(\theta)$ una transformación inyectiva de θ . La información de Fisher esperada $J_\phi(\phi)$ de ϕ puede entonces calcularse como sigue:

$$J_\phi(\phi) = J_\theta(\theta) \left\{ \frac{dh^{-1}(\phi)}{d\phi} \right\}^2 = J_\theta(\theta) \left\{ \frac{dh(\theta)}{d\theta} \right\}^{-2}.$$

Con esto ya podemos probar el resultado de invarianza de la distribución a priori de Jeffreys.

Prueba: De $f(\theta) \propto \sqrt{J(\theta)}$ se deduce, usando el cambio de variables y el resultado precedente, que

$$\begin{aligned} f_\phi(\phi) &= f_\theta(\theta) \cdot \left| \frac{dh^{-1}(\phi)}{d\phi} \right| \\ &\propto \sqrt{J_\theta(\theta)} \cdot \left| \frac{dh^{-1}(\phi)}{d\phi} \right| = \sqrt{J_\theta(\theta) \cdot \left\{ \frac{dh^{-1}(\phi)}{d\phi} \right\}^2} = \sqrt{J_\phi(\phi)} \end{aligned}$$

y con esto queda probada la proposición.

Si hubiésemos elegido la parametrización $\phi = h(\theta)$ desde el principio, al aplicar la regla de Jeffreys habríamos obtenido la misma distribución a priori: $f_\phi(\phi) \propto \sqrt{J_\phi(\phi)}$.

2.3. Distribución a posteriori predictiva

Consideremos una muestra aleatoria $X_{1:n}$ de una distribución con función de densidad $f(\underline{x} | \theta)$. Queremos predecir una observación (independiente) que se produzca en el futuro, $Y = X_{n+1}$ de $f(\underline{x} | \theta)$. La distribución a posteriori predictiva de $Y | x_{1:n}$, que puede calcularse fácilmente usando leyes de probabilidad, es la siguiente:

$$\begin{aligned} f(y | x_{1:n}) &= \int f(y, \theta | x_{1:n}) d\theta = \int f(y | \theta, x_{1:n}) f(\theta | x_{1:n}) d\theta \\ &= \int f(y | \theta) f(\theta | x_{1:n}) d\theta. \end{aligned}$$

Observemos que la última línea se deduce de la hipótesis de independencia condicional de Y y $X_{1:n}$, dado θ . Por tanto, para calcular $f(y | x_{1:n})$, basta con integrar el producto de la verosimilitud $f(y | \theta)$ y la función de densidad a posteriori $f(\theta | x_{1:n})$ con respecto de θ . El resultado $f(y | x_{1:n})$ se llama distribución a posteriori predictiva, en contraposición a la distribución a priori predictiva:

$$f(y) = \int f(y | \theta) f(\theta) d\theta.$$

donde $f(y | x_{1:n})$ se ha sustituido por la densidad a priori $f(\theta)$.

Conviene señalar que la distribución a posteriori predictiva $f(y | x_{1:n})$ no es igual a la distribución $f(y | \hat{\theta})$. De hecho, $f(y | \hat{\theta})$ se obtiene al sustituir la distribución a posteriori $f(y | x_{1:n})$ por una medida puntual en $\hat{\theta}$, normalmente el estimador de máxima

verosimilitud, en la expresión para la distribución predictiva. Por su parte, la predicción Bayesiana automáticamente incorpora la falta de certeza en la estimación del parámetro.

Cálculo de la distribución a posteriori predictiva

Ya que la distribución a priori predictiva es sencillamente el denominador en el teorema de Bayes, puede calcularse si se conocen la verosimilitud y las distribuciones a priori y a posteriori:

$$f(y) = \frac{f(y|\theta)f(\theta)}{f(\theta|y)}$$

lo cual está bien definido para todo θ . Una fórmula similar también estará entonces bien definida para la distribución a posteriori predictiva:

$$f(y|\underline{x}) = \frac{f(y|\underline{x},\theta)f(\theta|\underline{x})}{f(\theta|\underline{x},y)} = \frac{f(y|\theta)f(\theta|\underline{x})}{f(\theta|\underline{x},y)}$$

donde la última igualdad es consecuencia de la independencia condicionada de X y Y.

Si $f(\theta)$ es conjugada con respecto a $f(\underline{x}|\theta)$, entonces $f(\theta|\underline{x})$ y $f(\theta|\underline{x},y)$ pertenecen a la misma familia de distribuciones, y la distribución a posteriori predictiva puede obtenerse sin integrar explícitamente.

Capítulo 3

Inferencia bayesiana sobre una proporción

En este capítulo se estudiarán métodos Bayesianos para inferir sobre el valor desconocido p de una distribución Bernoulli asumiendo conocida la distribución a priori de p y habiendo observado datos de un proceso que siga esta distribución.

3.1. Cálculo de la distribución a posteriori

Veremos ahora distintos modos de establecer la distribución a priori, tanto para caso continuo como para discreto, y obtendremos como consecuencia distribuciones a posteriori diferentes en cada caso.

3.1.1. Distribución a priori discreta para \tilde{p}

Enunciamos de nuevo la fórmula de Bayes para m valores de una distribución Bernoulli de parámetro p . Si la probabilidad a priori de que la distribución tome el valor p_i es $P(\tilde{p} = p_i)$, donde $i = 1, 2, \dots, m$ y se han observado unos datos D , la fórmula de Bayes queda como sigue:

$$P(\tilde{p} = p_i | D) = \frac{P(\tilde{p}=p_i)L(p_i|D)}{P(D)} = \left(\frac{1}{P(D)}\right)P(\tilde{p} = p_i)L(p_i | D)$$

donde $\frac{1}{P(D)}$ no depende de p ; podemos por tanto llamarlo k y tenemos:

$$P(\tilde{p} = p_i | D) = kP(\tilde{p} = p_i)L(p_i | D)$$

y por proporcionalidad, deducimos, como hicimos en el capítulo anterior, que

a posteriori \propto verosimilitud \times distribución a priori.

Función verosimilitud de la distribución Bernoulli

Si se observan n experimentos Bernoulli (distribución binomial), los datos observados pueden verse como una secuencia de éxitos y fracasos en un orden específico. Si hay r éxitos y $n - r$ fracasos, entonces

$$L(p_i | D) = p_i^r q_i^{n-r}.$$

Si el orden exacto de éxitos y fracasos no queda registrado en los datos entonces

$$L(p_i | D) = \binom{n}{r} p_i^r q_i^{n-r}$$

aunque esta verosimilitud es equivalente a la anterior puesto que $\binom{n}{r}$ no depende de p .

Caso distribución uniforme

Un caso importante es aquél en que la distribución a priori de \tilde{p} es discreta y uniforme. Usando el resultado de proporcionalidad expuesto más arriba, tenemos lo siguiente:

$$\text{a posteriori} \propto \text{verosimilitud}$$

donde las probabilidades uniformes a priori han sido englobadas por la constante de proporcionalidad. Para obtener la igualdad basta encontrar una constante normalizadora para la verosimilitud.

3.1.2. Distribución a priori continua para \tilde{p}

En la sección anterior, los valores del parámetro \tilde{p} de la distribución Bernoulli estaban restringidos a un conjunto finito de valores; de esta forma se dedujo la relación de proporcionalidad entre distribución a priori y a posteriori.

Esta idea se puede extender para el caso más general en el que p puede tomar cualquier valor entre 0 y 1. Dado que el intervalo entre 0 y 1 es continuo, es necesario considerar funciones de densidad para el valor desconocido de p , tanto para la distribución a priori como para la distribución a posteriori.

3.1.2.1. Distribución continua a priori para \tilde{p}

Sea p_* un valor fijo de p y sea dp la longitud de un intervalo pequeño contenido en los valores de p . Consideremos también un valor r que denote el número de éxitos en n experimentos Bernoulli. Para la función de densidad a priori $f(p)$ se tiene:

$$f(p_*)dp \approx P(p_* - \frac{dp}{2} < \tilde{p} < p_* + \frac{dp}{2}).$$

Es decir, $f(p_*)dp$ es la probabilidad a priori aproximada de que \tilde{p} tome un valor contenido en el intervalo de centro p_* y longitud $2 * dp$.

Por la fórmula de Bayes tenemos:

$$P(p_* - \frac{dp}{2} < \tilde{p} < p_* + \frac{dp}{2} | r) \approx \frac{P(p_* - \frac{dp}{2} < \tilde{p} < p_* + \frac{dp}{2})L(p_* | r, n)}{P(r, n)}.$$

Si dejamos que $P(r|n)$ sea parte de la constante de proporcionalidad, tenemos, de manera aproximada, lo siguiente:

$$P(p_* - \frac{dp}{2} < \tilde{p} < p_* + \frac{dp}{2} | r) \propto P(p_* - \frac{dp}{2} < \tilde{p} < p_* + \frac{dp}{2})L(p_* | r, n).$$

El miembro izquierdo de la ecuación es aproximadamente la probabilidad a posteriori

$$f(p_* | r, n)dp$$

por tanto se obtiene:

$$f(p_* | r, n)dp \approx \frac{f(p_*)dpL(p_* | r, n)}{P(r|n)}.$$

Cuando dp tiende a cero, esta aproximación tiende a la igualdad y se pueden cancelar los dp a ambos lados para obtener la fórmula de Bayes para el caso continuo:

$$f(p_* | r, n) = \frac{f(p_*)L(p_* | r, n)}{P(r|n)}.$$

y usando proporcionalidad

$$f(p_* | r, n) \propto f(p_*)L(p_* | r, n).$$

3.1.2.2. Distribución continua uniforme para \tilde{p}

Al igual que en el caso discreto, cobra importancia el caso uniforme para la distribución continua a priori de \tilde{p} .

Distribución a priori uniforme para \tilde{p} quiere decir que la función de densidad de probabilidad $f(p)$ es constante para todo p , por tanto debe ser un rectángulo con base en el intervalo del cero al uno; además, como debe tener área igual a uno, el rectángulo debe ser un cuadrado, y $f(p)$ debe ser igual a uno para todos los valores de p entre cero y uno.

Usando la ecuación de proporcionalidad del apartado anterior tenemos, dado r ,:

$$f(p | r, n) \propto 1 \times L(p | r, n).$$

Cuando estudiamos antes el caso uniforme para distribución discreta a priori, vimos que la función de verosimilitud era una función discreta en un conjunto discreto de valores de p , y en un p particular tomaba un valor proporcional a $p^r q^{n-r}$. La verosimilitud es ahora la misma salvo por ser esta vez una función continua definida en todo el intervalo de cero a uno:

$$L(p| r, n) \propto p^r(1-p)^{n-r}, \quad 0 < p < 1.$$

En consecuencia la función de densidad a posteriori de $(\tilde{p}| r, n)$ es:

$$f(p| r, n) = kp^r(1-p)^{n-r}$$

donde k es una constante tal que el área bajo $f(p| r, n)$ es uno, es decir, buscamos un k tal que

$$\int_0^1 kp^r(1-p)^{n-r} dp = 1$$

o expresado de otro modo

$$\int_0^1 p^r(1-p)^{n-r} dp = 1/k.$$

Función beta

La integral anterior es un caso concreto de una integral de gran importancia en las matemáticas: la función beta. Esta función, que dará lugar más adelante a la distribución beta, está definida de la siguiente manera:

$$B(a, b) = \int_0^1 p^{a-1}(1-p)^{b-1} dp, \quad a > 0, \quad b > 0.$$

Supongamos ahora que en esta expresión sustituimos a por $r+1$ y b por $n-r+1$ (ambos son mayores que cero) dando como resultado

$$\int_0^1 p^r(1-p)^{n-r} dp = B(r+1, n-r+1)$$

o de manera equivalente

$$\int_0^1 \frac{1}{B(r+1, n-r+1)} p^r(1-p)^{n-r} dp = 1.$$

Observando las ecuaciones, se concluye que el valor que buscábamos para la k no es otro que $\frac{1}{B(r+1, n-r+1)}$.

Esto nos lleva a que la función de densidad de probabilidad a posteriori para p dados r y n , siendo uniforme la distribución a priori, es la siguiente:

$$f(p| r, n) = \frac{1}{B(r+1, n-r+1)} p^r(1-p)^{n-r}.$$

Esta función es efectivamente función de densidad ya que su integral es uno.

La función que se ha definido suele llamarse función de densidad de probabilidad **beta** por su relación con la función beta.

Dos importantes parámetros de la distribución $B(a, b)$:

- Media: $E(\tilde{p}| a, b) = \frac{a}{a+b}$.
- Varianza: $V(\tilde{p}| a, b) = \frac{\frac{a}{a+b} \frac{b}{a+b}}{a+b+1}$.

3.1.3. Distribución Beta a priori

Supongamos que resulta inconveniente considerar una distribución uniforme a priori para nuestra Bernoulli de parámetro \tilde{p} (el tamaño de la muestra es relativamente pequeño, sabemos algo sobre \tilde{p} debido a experimentos anteriores, etc).

Podemos seguir utilizando la fórmula de Bayes si encontramos una apropiada función de distribución a priori para \tilde{p} que pertenezca a la familia beta, es decir, una función de distribución a priori para \tilde{p} de la forma

$$f_{\beta}(p|a, b) \propto p^{a-1}(1-p)^{b-1}$$

para ciertos a y b no negativos. De la expresión

$$\text{a posteriori} \propto \text{verosimilitud} \times \text{distribución a priori}$$

tenemos

$$\text{a posteriori} \propto \text{verosimilitud} \times p^{a-1}(1-p)^{b-1}.$$

Para una muestra de n experimentos Bernoulli con r éxitos, la verosimilitud es, como antes,

$$L(p|r, n) \propto p^r(1-p)^{n-r}.$$

Agrupando con la expresión anterior obtenemos

$$\text{a posteriori} \propto p^r(1-p)^{n-r} \times p^{a-1}(1-p)^{b-1}.$$

Y utilizando que $p^{a-1}p^r = p^{a+r-1}$ y $(1-p)^{b-1}(1-p)^{n-r} = (1-p)^{b+n-r-1}$, llegamos a

$$\text{a posteriori} \propto p^{a+r-1}(1-p)^{b+n-r-1}$$

que es proporcional a una distribución beta $B(a+r, b+n-r)$. Para obtener la igualdad en esta expresión, basta multiplicar el miembro a la derecha por la constante normalizadora apropiada, que en este caso será $\frac{\Gamma(a+b+n)}{\Gamma(a+r)\Gamma(b+n-r)}$ al tratarse de una distribución beta.

Observemos que la media a posteriori es $\frac{a+r}{a+r+b+n-r} = \frac{a+r}{a+b+n}$.

Ya que la distribución a priori para \tilde{p} y la distribución a posteriori pertenecen a la misma familia paramétrica, se deduce que la distribución beta es conjugada para la verosimilitud de una distribución binomial.

3.2. Regla de Jeffreys

Obtendremos la distribución a priori de Jeffreys para la verosimilitud de la distribución binomial. Empezaremos escribiendo la log verosimilitud:

$$\log(L(p| r, n)) = r\log(p) + (n - r)\log(1 - p) + \text{constante}.$$

La segunda derivada de la log verosimilitud es

$$\frac{\partial^2 \log(L(p| r, n))}{\partial p^2} = -\frac{r}{p^2} - \frac{n-r}{(1-p)^2}.$$

Tomar esperanza con respecto a r es sencillo porque aparece linealmente. Podemos por tanto sustituir $E(r| p)$ por r . Si $r \sim \text{Binomial}(n, p)$, entonces $E(r| p) = np$, y tenemos la siguiente expresión para la información de Fisher esperada:

$$J(p| r) = E\left[\frac{\partial^2 \log(L(p| r, n))}{\partial p^2}\right] = \frac{n}{p(1-p)}.$$

Tomando raíz cuadrada y quitando la constante n , obtenemos:

$$f(p) \propto p^{-\frac{1}{2}}(1-p)^{-\frac{1}{2}}, \quad 0 < p < 1.$$

Esta función de densidad se corresponde con la de una distribución beta $B(\frac{1}{2}, \frac{1}{2})$. Por tanto, en el caso de la verosimilitud de una binomial, la distribución a priori de Jeffreys es miembro de la familia conjugada.

3.3. Distribución predictiva

Utilizaremos las fórmulas dadas en el capítulo anterior para obtener la distribución predictiva (a posteriori) para una proporción. Dado n y la distribución a priori para \tilde{p} , tenemos las siguientes distribuciones predictivas para $(\tilde{r}| n)$:

Distribución a priori discreta para \tilde{p}

$$P(\tilde{r} = r) = \sum_i P_b(r| n, p_i)P(\tilde{p} = p_i), \quad r = 0, 1, \dots, n$$

Distribución a priori continua para \tilde{p}

$$P(\tilde{r} = r) = \int_0^1 P_b(r| n, p_i)f(p)dp, \quad r = 0, 1, \dots, n$$

Distribución a priori beta $B(a, b)$ para \tilde{p}

$$P(\tilde{r} = r) = \int_0^1 P_b(r| n, p_i), f_\beta(p| a, b)dp, \quad r = 0, 1, \dots, n$$

Esta distribución discreta para \tilde{r} , que depende de n, a, b , se llama beta-binomial. Escribamos su media y varianza:

- $E(\tilde{r}| n, a, b) = n\frac{a}{a+b}$
- $V(\tilde{r}| n, a, b) = n(n + a + b)\left(\frac{ab}{(a+b)^2(a+b+1)}\right)$

3.4. Contraste de hipótesis

En este apartado veremos cómo podemos estudiar, para el caso de la distribución Bernoulli, un contraste de hipótesis desde el punto de vista de la inferencia bayesiana.

Contraste para una Bernoulli de parámetro p

Consideremos el siguiente contraste:

$$\begin{cases} H_0 : p \leq p_0 \\ H_1 : p > p_0 \end{cases}$$

Si suponemos que se han realizado n experimentos y se han observado r éxitos, el p -valor, en un sentido clásico, es:

$$P(\tilde{r} \geq r | n, p_0).$$

Podemos emplear, suponiendo una distribución a priori uniforme (beta con parámetros $a = b = 1$), una distribución beta como distribución a posteriori para \tilde{p} , cuyos parámetros son $a' = r + 1$ y $b' = n - r + 1$. Podemos entonces calcular

$$P(H_0 \text{ cierta} | \text{datos muestrales}) = P_\beta(\tilde{p} \leq p_0 | a' = r + 1, b' = n - r + 1).$$

Cuando los parámetros a' y b' de la beta son enteros, la función de probabilidad de la binomial y la beta se relacionan de la siguiente manera:

$$P_\beta(\tilde{p} \leq p_0 | r + 1, n - r - 1) \equiv P_b(\tilde{r} \geq r + 1 | n + 1, p_0).$$

Para n suficientemente grande y p_0 no demasiado próximo a cero o uno, se tiene que $P_b(\tilde{r} \geq r | n, p_0)$ no difiere apenas de $P_b(\tilde{r} \geq r + 1 | n + 1, p_0)$, y por tanto el p -valor es similar a la distribución a posteriori para $\tilde{p} \leq p_0$.

Capítulo 4

Distribución de Poisson

En este capítulo estudiaremos métodos Bayesianos para inferir sobre el valor desconocido del parámetro $\tilde{\lambda}$ de una distribución de Poisson. Supondremos conocimiento a priori sobre el parámetro y datos obtenidos mediante un proceso que siga una distribución de Poisson. También veremos cómo hacer predicciones acerca de futuras observaciones.

4.1. Función de verosimilitud

Sean t_i una sucesión de variables independientes idénticamente distribuidas según una exponencial de parámetro λ .

Podemos observar un proceso Poisson por T veces y considerar los t_i en que se obtiene el suceso considerado de éxito; pero igualmente válido es observar el proceso hasta que ocurran r éxitos, considerando los t_i en que sucede y parando cuando se obtengan r éxitos. Fijando T y observando t_1, \dots, t_r se tiene una muestra de Poisson, mientras que fijando r y considerando los t_1, \dots, t_r se tiene lo que se conoce como muestra de distribución de Poisson inversa. En ambos casos, los datos a tener en cuenta son el conjunto de t_i en los que ocurre el éxito.

Calcularemos primero la verosimilitud de la distribución Poisson inversa.

Verosimilitud de la distribución Poisson inversa

Buscamos la probabilidad conjunta de observar sucesos en t_1, \dots, t_r cuando observamos un proceso Poisson de parámetro λ desde el instante 0 hasta el instante t_r , en el que se produce el r -ésimo suceso:

$$L(\lambda | t_1, \dots, t_r) = f(t_1, \dots, t_r | \lambda, r).$$

Por otra parte, si denotamos

$$t'_1 = t_1, t'_2 = t_2 - t_1, t'_3 = t_3 - t_2, \dots, t'_r = t_r - t_{r-1},$$

entonces las variables aleatorias t'_1, \dots, t'_r son independientes y están distribuidas según una exponencial de parámetro λ . Por tanto, ya que

$$f(t_1, \dots, t_r | \lambda, r) = f(t'_1, \dots, t'_r | \lambda, r)$$

y los t'_i son independientes, tenemos:

$$f(t'_1, \dots, t'_r | \lambda, r) = f(t'_1 | \lambda) \times \dots \times f(t'_r | \lambda).$$

Puesto que cada una de estas densidades es de la forma $\lambda e^{-\lambda t'_i}$, su producto es como sigue:

$$\lambda e^{-\lambda t'_1} \times \dots \times \lambda e^{-\lambda t'_r} = \lambda^r e^{-\lambda \sum t'_i}.$$

Y es trivial que $\sum_1^r t'_i = t_r$.

En conclusión, tenemos:

$$L(\lambda | t_1, \dots, t_r) = f(t_1, \dots, t_r | \lambda) = \lambda^r e^{-\lambda t_r}$$

La verosimilitud depende sólo de t_r ; deducimos por tanto que el estadístico t_r es suficiente.

Verosimilitud de la distribución de Poisson

Para una muestra de Poisson, con T fijo, la verosimilitud de λ dados los datos t_1, \dots, t_r y T es una modificación de la verosimilitud obtenida para el caso de la distribución Poisson inversa. Ahora debemos establecer la verosimilitud de forma que no ocurran más sucesos desde t_r hasta T . La probabilidad de cero sucesos entre t_r y T es la probabilidad de observar cero sucesos en un intervalo de tiempo de longitud $(T - t_r)$:

$$P_p(\tilde{r} = 0 | \lambda \times (T - t_r)) = \frac{e^{-\lambda(T-t_r)} (\lambda(T-t_r))^0}{0!} = e^{-\lambda(T-t_r)}.$$

Multiplicando la verosimilitud de la distribución inversa por esta probabilidad obtenemos la verosimilitud para el caso de la distribución de Poisson:

$$\begin{aligned} L(\lambda | t_1, \dots, t_r, T) &= \lambda^r e^{-\lambda t_r} \times e^{-\lambda(T-t_r)} \\ &= \lambda^r e^{-\lambda(t_r + T - t_r)} = \lambda^r e^{-\lambda T}. \end{aligned}$$

El estadístico suficiente es ahora r : basta conocer r para deducir la verosimilitud.

4.2. Cálculo de la distribución a posteriori

En este apartado estudiaremos cómo obtener la distribución a posteriori de una distribución de Poisson, partiendo de una distribución a priori no informativa y también estudiando el caso conjugado.

4.2.1. Distribución a priori no informativa

La verosimilitud de la distribución de Poisson es de la forma

$$L(\lambda | t_1, \dots, t_r) \propto \lambda^r e^{-\lambda T}.$$

Por tanto la log-verosimilitud es:

$$\log(L(\lambda | t_1, \dots, t_r)) = \text{constante} + r \log \lambda - \lambda T.$$

Ahora deduciremos la expresión de una distribución a priori no informativa mediante la regla de Jeffreys. Primero calculamos derivadas sucesivas respecto al parámetro λ :

$$\frac{\partial \log(L(\lambda | t_1, \dots, t_r))}{\partial \lambda} = \frac{r}{\lambda} - T, \quad \frac{\partial^2 \log(L(\lambda | t_1, \dots, t_r))}{\partial \lambda^2} = -\frac{r}{\lambda^2}.$$

Notemos que $r = \sum t_i$, luego $r/T = \bar{t}$. Para \bar{t} distinto de cero, el estimador de máxima verosimilitud de λ es $\hat{\lambda} = \bar{t}$, y tenemos

$$J(\hat{\lambda}) = \left(-\frac{1}{n} \frac{\partial^2 \log(L(\lambda | t_1, \dots, t_r))}{\partial \lambda^2} \right)_{\hat{\lambda}} = \frac{1}{\hat{\lambda}}.$$

La regla de Jeffreys nos dice entonces que la distribución a priori no informativa para $\tilde{\lambda}$ es

$$f(\tilde{\lambda}) \propto \sqrt{J(\tilde{\lambda})} \propto \lambda^{-1/2}.$$

Como consecuencia, la función de densidad a posteriori cumple la siguiente relación de proporcionalidad:

$$f(\tilde{\lambda} | t_1, \dots, t_r) \propto \tilde{\lambda}^{r-\frac{1}{2}} e^{-\tilde{\lambda} T}$$

e integrando se obtiene la constante normalizadora, que es la siguiente:

$$\text{constante} = T^{-(r+\frac{1}{2})} [\Gamma(r + \frac{1}{2})]^{-1}.$$

4.2.2. Distribución Gamma a priori para Poisson $\tilde{\lambda}$

De manera similar a la distribución Beta a priori para una Bernoulli, estudiamos ahora el caso conjugado para la distribución de Poisson.

La familia paramétrica que estudiaremos es la conocida como familia Gamma de distribuciones, llamada así por su relación con la función Gamma, definida como sigue:

$$\Gamma(a) = \int_0^{+\infty} x^{a-1} e^{-x} dx.$$

Por su parte, la función de densidad de probabilidad que usaremos es de la siguiente forma:

$$f(x) = \frac{x^{a-1}e^{-x/b}}{\Gamma(a)b^a}.$$

La familia Gamma depende de los parámetros a y b y, al ser función de densidad, su integral en todo el intervalo real es uno, o expresado de otro modo:

$$\int_0^{+\infty} x^{a-1}e^{-x/b} dx = \Gamma(a)b^a.$$

Vemos ahora la relación de esta función de densidad con la función Gamma (para $b = 1$ es, de hecho, la misma expresión integral). Para adaptar estas consideraciones a la distribución de Poisson, reemplazamos la x por $\tilde{\lambda}$, y los parámetros a y b por r' y $1/T'$ respectivamente. Obtenemos, como consecuencia, la densidad a priori para el parámetro λ de una distribución de Poisson:

$$f_\gamma(\tilde{\lambda} | r', T') = \frac{\tilde{\lambda}^{r'-1}e^{-\tilde{\lambda}T'}}{\Gamma(r')(1/T')^{r'}}$$

para $\tilde{\lambda} > 0$ y $r', T' > 0$. Como eventualmente estudiaremos relaciones de proporcionalidad, las constante pueden obviarse hasta calcular las constante normalizadora de la distribución a posteriori.

Los valores r' y T' de la distribución a priori no tienen por qué ser iguales a los de la función de verosimilitud, de aquí que estén distinguidos con la prima. Escribamos la media y la varianza de la distribución Gamma con respecto a los parámetros de una Poisson.

- Media: $E(\tilde{\lambda} | r', T') = \frac{r'}{T'}$.
- Varianza $V(\tilde{\lambda} | r', T') = \frac{r'}{T'^2}$.

Para finalizar, apliquemos el teorema de Bayes y obtengamos la distribución a posteriori para el parámetro de la distribución de Poisson. Por una parte, la distribución a priori es proporcional a

$$\tilde{\lambda}^{r'-1}e^{-\tilde{\lambda}T'}$$

y por otra, la verosimilitud de $\tilde{\lambda}$ dados los datos (resumidos en los valores r y T) es proporcional a

$$\tilde{\lambda}^r e^{-\tilde{\lambda}T}.$$

En consecuencia, la distribución a posteriori es proporcional a

$$\tilde{\lambda}^{r'-1}e^{-\tilde{\lambda}T'} \times \tilde{\lambda}^r e^{-\tilde{\lambda}T}.$$

Definimos ahora $r'' = r' + r$ y $T'' = T' + T$ y obtenemos la siguiente expresión para la distribución a posteriori:

$$\tilde{\lambda}^{r''-1}e^{-\tilde{\lambda}T''}.$$

Esta expresión se corresponde con una función de densidad Gamma de parámetros r'' y T'' . La constante normalizadora se deduce de la expresión para la función de densidad Gamma que se dio anteriormente, y por tanto la densidad a posteriori (conjugada) que buscamos es

$$f(\tilde{\lambda} | r'', T'') = \frac{\tilde{\lambda}^{r''-1} e^{-\tilde{\lambda} T''}}{\Gamma r'' (1/T'')^{r''}}.$$

4.3. Distribución predictiva

Para $\tilde{\lambda}$ desconocido, las predicciones acerca de $(\tilde{r} | T, \text{distribución a priori})$ vienen expresadas por la distribución predictiva.

Distribución a priori discreta para $\tilde{\lambda}$

$$P(r | T, \text{dist. a priori}) = \sum_i P_{P_0}(r | \lambda_i T) P(\lambda_i), \quad r = 0, 1, \dots$$

Distribución a priori continua para $\tilde{\lambda}$

$$P(r | T, \text{dist. a priori}) = \int_0^{+\infty} P_{P_0}(r | \lambda T) f(\lambda) d\lambda, \quad r = 0, 1, \dots$$

Distribución a priori Gamma (r', T') para $\tilde{\lambda}$

$$P(r | T, r', T') = \int_0^{+\infty} P_{P_0}(r | \lambda T) f_\gamma(\lambda | r', T') d\lambda, \quad r = 0, 1, \dots$$

que corresponde a la función de distribución de probabilidad de una Gamma-Poisson. Escribamos su media y varianza:

- $E(\tilde{r} | T, r', T') = (\frac{r'}{T'}) T$
- $V(\tilde{r} | T, r', T') = (\frac{r'}{T'}) T (\frac{T'+T}{T'})$

4.4. Contraste de hipótesis

Consideremos un test de hipótesis del tipo

$$\begin{cases} H_0 : \lambda \leq \lambda_0 \\ H_1 : \lambda > \lambda_0 \end{cases}$$

donde se observan r sucesos en T experimentos. El p-valor en sentido clásico es

$$P_{P_0}(\tilde{r} \geq r | \lambda_0 \times T).$$

En el análisis Bayesiano, con una distribución a priori no informativa, se tiene a posteriori para $\tilde{\lambda}$ una distribución Gamma con parámetros $r'' = r + 1$ y $T'' = T$. Se calcula lo siguiente:

$$P(H_0 \text{ cierta} | \text{datos}) = P_\gamma(\tilde{\lambda} \leq \lambda_0 | r'' = r + 1, T'' = T).$$

Cuando los parámetros de la gamma r'' y T'' son enteros, la distribución Gamma y la distribución de Poisson son equivalentes, en el sentido siguiente:

$$P_\gamma(\tilde{\lambda} \leq \lambda_0 | r'' = r + 1, T'' = T) \equiv P_{P_0}(\tilde{r} \geq r'' | \lambda_0 \times T'').$$

Y para $\lambda_0 \times T$ grande, se tiene que

$$P_{P_0}(\tilde{r} \geq r | \lambda_0 \times T)$$

no es muy diferente a

$$P_{P_0}(\tilde{r} \geq r'' = r + 1 | \lambda_0 \times T)$$

y el p-valor clásico es en gran medida similar a la probabilidad a posteriori de que $\tilde{\lambda} \leq \lambda$.

Capítulo 5

Distribución normal

5.1. Función de verosimilitud para (μ, σ^2)

Sean x_1, x_2, \dots, x_n una muestra de n observaciones (independientes) de un proceso que sigue una distribución normal, es decir, cada x_i es un valor observado de \tilde{x}_i que se distribuye según una normal $N(\mu, \sigma^2)$ con función de densidad como sigue:

$$f_N(x_i | \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2}$$

Ya hemos visto que la función de la verosimilitud es de la forma

$$L(\text{par'ámetros} | \text{datos}) = f(\text{datos} | \text{par'ámetros}).$$

Ahora tenemos dos parámetros, por lo que la verosimilitud queda

$$L(\mu, \sigma^2 | x_1, x_2, \dots, x_n) = f(x_1, x_2, \dots, x_n | n, \mu, \sigma^2)$$

donde f denota la función de densidad conjunta. Como los \tilde{x}_i son independientes,

$$f(x_1, x_2, \dots, x_n | \mu, \sigma^2) = f_N(x_1 | \mu, \sigma^2) \times f_N(x_2 | \mu, \sigma^2) \times \dots \times f_N(x_n | \mu, \sigma^2).$$

Reemplazando cada función de densidad por su expresión llegamos a

$$L(\mu, \sigma^2 | x_1, x_2, \dots, x_n) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_1-\mu}{\sigma}\right)^2} \times \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_2-\mu}{\sigma}\right)^2} \times \dots \times \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_n-\mu}{\sigma}\right)^2}.$$

Agrupando términos tenemos

$$L(\mu, \sigma^2 | x_1, x_2, \dots, x_n) = (\sigma\sqrt{2\pi})^{-n} e^{-(1/2\sigma^2)[(x_1-\mu)^2+(x_2-\mu)^2+\dots+(x_n-\mu)^2]}.$$

O de manera equivalente:

$$L(\mu, \sigma^2 | x_1, x_2, \dots, x_n) = (\sigma\sqrt{2\pi})^{-n} e^{-(1/2\sigma^2)\sum_1^n (x_i-\mu)^2}.$$

Podemos simplificar esta expresión sustrayendo $(\sqrt{2\pi})^{-n}$ y cambiando el igual por una relación de proporcionalidad:

$$L(\mu, \sigma^2 | x_1, x_2, \dots, x_n) \propto (\sigma)^{-n} \times e^{-(1/2\sigma^2)\sum_1^n (x_i - \mu)^2}.$$

Los dos términos que quedan no pueden quitarse porque dependen de los parámetros.

Ahora haremos algunas transformaciones en lo obtenido para llegar a una expresión que será de utilidad más adelante. Sea \bar{x} la media muestral de las observaciones x_1, x_2, \dots, x_n . Consideremos el término

$$\sum_1^n (x_i - \mu)^2$$

que podemos escribir como

$$\sum_1^n (x_i - \bar{x} + \bar{x} - \mu)^2.$$

Desarrollando obtenemos

$$\sum_1^n (x_i - \bar{x} + \bar{x} - \mu)^2 = \sum_1^n [(x_i - \bar{x}) + (\bar{x} - \mu)]^2 = \sum_1^n (x_i - \bar{x})^2 + 2 \sum_1^n (x_i - \bar{x})(\bar{x} - \mu) + \sum_1^n (\bar{x} - \mu)^2.$$

Estudiemos los tres términos separadamente:

- El primer término está relacionado con la varianza muestral, denotada s^2 y definida como

$$s^2 = \frac{\sum_1^n (x_i - \bar{x})^2}{n}.$$

Por tanto

$$\sum_1^n (x_i - \bar{x})^2 = ns^2.$$

- Para el segundo término tenemos

$$2 \sum_1^n (x_i - \bar{x})(\bar{x} - \mu) = 2(\bar{x} - \mu) \sum_1^n (x_i - \bar{x}).$$

Pero $\sum_1^n (x_i - \bar{x}) = 0$, luego el segundo término se anula.

- En cuanto al tercero, se observa que el sumando es igual para todo n , luego

$$\sum_1^n (\bar{x} - \mu)^2 = n(\bar{x} - \mu)^2.$$

Uniando estos resultados e implementándolos en la función de verosimilitud se obtiene:

$$L(\mu, \sigma^2 | x_1, x_2, \dots, x_n) \propto (\sigma)^{-n} e^{-(1/2\sigma^2)[ns^2 + n(\bar{x} - \mu)^2]}.$$

Lo que conduce a la expresión para la verosimilitud:

$$L(\mu, \sigma^2 | x_1, x_2, \dots, x_n) \propto (\sigma)^{-n} e^{-ns^2/2\sigma^2} \times e^{-n(\mu-\bar{x})^2/2\sigma^2}.$$

Obsevación: antes de proseguir con nuestro estudio de la distribución Normal, conviene aclarar cierto detalle que estará presente a lo largo del capítulo.

Dependiendo de lo que resulte más conveniente para los cálculos, a veces manejaremos expresiones en función de la varianza, σ^2 , y en otras ocasiones en función de la desviación estándar, σ . Ahora bien, es muy sencillo pasar de tener una expresión en función de la primera a tenerla en función de la segunda, y viceversa.

Pongamos como ejemplo que tenemos la función de densidad a priori para σ , pero nos interesa tener dicha densidad en función de σ^2 . Consideremos el siguiente cambio de variable:

$$h : \sigma \mapsto \sigma^2$$

cuya función inversa es

$$h^{-1} : \sigma^2 \mapsto \sqrt{\sigma^2}.$$

Si ahora calculamos el Jacobiano de h^{-1}

$$J(h^{-1}) = \left| \frac{dh^{-1}}{d\sigma^2} \right| = \frac{1}{2\sigma^2}$$

tenemos que la distribución a priori de σ^2 no es más que

$$f(\sigma^2) = f(\sigma)J(h^{-1}) = f(\sigma)\frac{1}{2\sigma^2}.$$

5.2. Regla de Jeffreys

Aplicaremos ahora la regla de Jeffreys para obtener distribuciones a priori no informativas para una normal de parámetros μ y σ^2 . Primero, escribamos lo siguiente:

$$f(x | \mu, \sigma^2) \propto \sigma^{-1} h\left(\frac{x-\mu}{\sigma}\right)$$

donde x no depende de los parámetros y h tiene ciertas condiciones de regularidad. Supongamos también que tenemos una muestra de n observaciones x_1, \dots, x_n de la distribución.

5.2.1. Caso $\tilde{\mu}$ desconocido y σ conocido

La verosimilitud es

$$L(\tilde{\mu} | \sigma, \underline{x}) \propto \prod_{i=1}^n h\left(\frac{x_i - \tilde{\mu}}{\sigma}\right).$$

Ahora tenemos

$$\frac{\partial \log f(x | \tilde{\mu}, \sigma)}{\partial \tilde{\mu}} = -\frac{1}{\sigma} \frac{h'_x}{h(x)}, \text{ donde } y = \left(\frac{x - \tilde{\mu}}{\sigma}\right).$$

Y ahora calculamos la información de Fisher esperada:

$$J(\tilde{\mu}) = E_{x | \tilde{\mu}} \left[\frac{\partial \log f(x | \tilde{\mu}, \sigma)}{\partial \tilde{\mu}} \right]^2 = \frac{1}{\sigma^2} E_y \left[\frac{h'_x}{h(x)} \right]^2$$

donde la esperanza en el último término se toma sobre la distribución $f(y) = h(y)$.

Ya que la expresión para la esperanza no incluye a $\tilde{\mu}$ y σ^2 es conocido, concluimos que $J(\tilde{\mu}) = \text{constante}$. Esto conduce a tomar a $\tilde{\mu}$ localmente uniforme a priori:

$$f(\tilde{\mu} | \sigma) \propto J(\tilde{\mu})^{1/2} = \text{constante}$$

y en consecuencia la correspondiente distribución a posteriori para $\tilde{\mu}$ es

$$f(\tilde{\mu} | \sigma, \underline{x}) \propto \prod_{i=1}^n h\left(\frac{x_i - \tilde{\mu}}{\sigma}\right), \quad -\infty < \tilde{\mu} < \infty.$$

Nota: conviene señalar que la distribución a priori obtenida para la media es impropia, pues la integral de su función de densidad, $\int_{-\infty}^{\infty} \text{constante} d\tilde{\mu}$, no es finita.

5.2.2. Caso μ conocido, $\tilde{\sigma}$ desconocido

La verosimilitud es ahora:

$$L(\tilde{\sigma} | \mu, \underline{x}) \propto \sigma^{-n} \prod_{i=1}^n h\left(\frac{x_i - \mu}{\tilde{\sigma}}\right).$$

Ya que

$$\frac{\partial \log f(x | \mu, \tilde{\sigma})}{\partial \tilde{\sigma}} = -\frac{1}{\tilde{\sigma}} \left[1 + \frac{y h'_x}{h(x)} \right], \text{ donde } y = \left(\frac{x - \mu}{\tilde{\sigma}}\right).$$

se deduce que

$$J(\tilde{\sigma}) = \frac{1}{\tilde{\sigma}^2} E_x \left[1 + \frac{y h'_x}{h(x)} \right]^2 \propto \frac{1}{\tilde{\sigma}^2}.$$

En consecuencia, la regla de Jeffreys nos lleva a la distribución a priori siguiente:

$$f(\tilde{\sigma} | \mu) \propto \frac{1}{\tilde{\sigma}} \text{ ó } f(\log \tilde{\sigma}) \propto \text{constante}.$$

y por tanto la correspondiente distribución a posteriori para $\tilde{\sigma}$ es

$$f(\tilde{\sigma} | \mu, \underline{x}) \propto \sigma^{-(n+1)} \prod_{i=1}^n h\left(\frac{x_i - \mu}{\tilde{\sigma}}\right), \quad \sigma > 0.$$

De nuevo se ha obtenido una distribución a priori impropia (pues la integral de $1/\tilde{\sigma}$ no es finita).

5.3. Cálculo de la distribución a posteriori

5.3.1. Caso $\tilde{\mu}$ desconocido y σ^2 conocido

Sean x_1, x_2, \dots, x_n una muestra de n observaciones de una normal $N(\tilde{\mu}, \sigma^2)$. Asumimos ahora que la varianza σ^2 es conocida. Haremos inferencia sobre la media desconocida $\tilde{\mu}$.

5.3.1.1. Distribución a priori no informativa

En cuanto a la función de verosimilitud, todos los factores, salvo aquellos que dependen de $\tilde{\mu}$, pueden ser absorbidos por la constante de proporcionalidad; por tanto, la función de verosimilitud se reduce a

$$L(\tilde{\mu} | n, \bar{x}, \sigma^2) \propto e^{-n(\tilde{\mu}-\bar{x})^2/2\sigma^2}.$$

Los x_1, x_2, \dots, x_n han sido sustituidos por n y \bar{x} ya que estos dos últimos términos bastan para determinar la verosimilitud.

Como distribución a priori no informativa para $\tilde{\mu}$ usaremos la obtenida mediante la regla de Jeffreys, es decir:

$$f(\tilde{\mu} | \sigma) \propto c.$$

La distribución a posteriori para $\tilde{\mu}$ verifica:

$$f(\tilde{\mu} | \bar{x}) \propto f(\tilde{\mu})L(\tilde{\mu} | \bar{x}) \propto f(\tilde{\mu})f(\bar{x} | \tilde{\mu}, \sigma^2),$$

donde $f(\tilde{\mu})$ es la distribución a priori.

Por tanto tenemos:

$$f(\tilde{\mu} | n, \bar{x}, \sigma^2) \propto e^{-n(\tilde{\mu}-\bar{x})^2/2\sigma^2}.$$

La constante normalizadora que hace que la integral del miembro a la derecha sea uno es

$$c^{-1} = \int_{-\infty}^{\infty} e^{-n(\tilde{\mu}-\bar{x})^2/2\sigma^2} d\tilde{\mu} = \left(\frac{2\pi\sigma^2}{n}\right)^{1/2}$$

y, en consecuencia, la función de densidad a posteriori es

$$f(\tilde{\mu} | n, \bar{x}, \sigma^2) = \left(\frac{2\pi\sigma^2}{n}\right)^{-1/2} e^{-n(\tilde{\mu}-\bar{x})^2/2\sigma^2}, \quad -\infty < \tilde{\mu} < \infty.$$

Concluimos que la distribución que sigue $\tilde{\mu}$ (a posteriori y de manera aproximada) es una normal $N(\bar{x}, \sigma^2/n)$.

5.3.1.2. Distribución a priori conjugada para $\tilde{\mu}$

Buscamos una función de densidad en la que la variable aleatoria aparezca de la misma forma que en la función de verosimilitud para $\tilde{\mu}$. Se observa trivialmente que $(\bar{x} - \tilde{\mu})^2 = (\tilde{\mu} - \bar{x})^2$; combinando esto con la expresión para la verosimilitud, deducimos que una distribución a priori conjugada para $\tilde{\mu}$ viene dada por la función de densidad de la normal (cuando σ^2 es conocido).

Supondremos lo siguiente:

- $\bar{x} \sim N(\tilde{\mu}, \frac{\sigma^2}{n})$
- $\tilde{\mu} \sim N(\mu_0, \sigma_0^2)$

Será a partir de ahora más conveniente expresar la función de densidad en términos de la media y la precisión en lugar de la media y la varianza. La precisión es la inversa de la varianza; a mayor varianza, menor precisión tiene la distribución a tratar. Usando esta indicación tenemos:

- $\bar{x} \sim N(\tilde{\mu}, \frac{n}{\tau^2})$
- $\tilde{\mu} \sim N(\mu_0, \frac{1}{\tau_0^2})$

donde $\tau^2 = 1/\sigma^2$ y $\tau_0^2 = 1/\sigma_0^2$.

Aplicamos ahora el teorema de Bayes para obtener la distribución a posteriori:

$$\begin{aligned} f(\tilde{\mu} | n, \bar{x}) &= \frac{\sqrt{n\tau}}{\sqrt{2\pi}} \exp\left(-\frac{n\tau^2(\tilde{\mu}-\bar{x})^2}{2}\right) \frac{\tau_0}{\sqrt{2\pi}} \exp\left(-\frac{\tau_0^2(\tilde{\mu}-\mu_0)^2}{2}\right) \\ &\propto \exp\left(-\frac{n\tau^2(\tilde{\mu}-\bar{x})^2}{2} - \frac{\tau_0^2(\tilde{\mu}-\mu_0)^2}{2}\right) \\ &\propto \exp\left(-\frac{(n\tau^2+\tau_0^2)\tilde{\mu}^2 + 2\tilde{\mu}(n\tau^2\bar{x} + \tau_0^2\mu_0)}{2}\right) \\ &= \exp\left(-\frac{(n\tau^2+\tau_0^2)(\tilde{\mu}^2 - 2\tilde{\mu}\frac{n\tau^2\bar{x} + \tau_0^2\mu_0}{(n\tau^2+\tau_0^2)})}{2}\right) \\ &\propto \exp\left(-\frac{(n\tau^2+\tau_0^2)(\tilde{\mu} - \frac{n\tau^2\bar{x} + \tau_0^2\mu_0}{(n\tau^2+\tau_0^2)})^2}{2}\right) \end{aligned}$$

Y esto corresponde a una función de densidad de una normal $N\left(\frac{n\tau^2\bar{x} + \tau_0^2\mu_0}{(n\tau^2 + \tau_0^2)}, \frac{1}{n\tau^2 + \tau_0^2}\right)$.

5.3.2. Caso μ conocido y $\tilde{\sigma}^2$ desconocido

5.3.2.1. Distribución a priori no informativa

Como μ es conocido, los factores que dependan de este parámetro en la función de verosimilitud pueden ser absorbidos por la constante de proporcionalidad, y obtenemos:

$$L(\tilde{\sigma}|n, \bar{x}, \tilde{\mu}) \propto \tilde{\sigma}^{-n} e^{(-\frac{ns^2}{2\tilde{\sigma}^2})}.$$

Usamos la distribución a priori de Jeffreys como distribución a priori no informativa para $\tilde{\sigma}$:

$$f(\tilde{\sigma} | \tilde{\mu}) \propto \sigma^{-1}.$$

Usando esta distribución a priori llegamos a la siguiente distribución a posteriori para $\tilde{\sigma}$:

$$f(\tilde{\sigma} | n, \tilde{\mu}, \bar{x}) \propto \tilde{\sigma}^{-(n+1)} e^{(-\frac{ns^2}{2\tilde{\sigma}^2})}.$$

Por otra parte, la distribución a posteriori de $\tilde{\sigma}^2$ debe ser

$$f(\tilde{\mu} | \bar{x}) \propto f(\tilde{\sigma}^2) L(\tilde{\sigma}^2 | \bar{x}) \propto f(\tilde{\sigma}^2) \tilde{\sigma}^{-n} \exp(-\frac{ns^2}{2\tilde{\sigma}^2}).$$

donde $f(\tilde{\sigma}^2)$ es la distribución a priori de $\tilde{\sigma}^2$.

Considereremos el siguiente resultado:

Lema: Sea t suficiente para $\tilde{\mu}$, con distribución conjunta $f(t | \tilde{\mu})$. Entonces:

$$L(\tilde{\mu} | \bar{x}) \propto L_1(\tilde{\mu} | t) \text{ donde } L_1(\tilde{\mu} | t) \propto f(t | \tilde{\mu}).$$

Por otra parte, ya que $L(\tilde{\sigma}^2 | \bar{x})$ involucra sólo al estadístico s^2 , que es suficiente para $\tilde{\sigma}^2$, podemos emplear el lema con el objetivo de deducir la expresión de la distribución a posteriori para $\tilde{\sigma}$ a través de la distribución de s^2 . Observemos que si los sumandos de \bar{x} siguen una distribución Normal con media cero y varianza $\tilde{\sigma}^2$, entonces la distribución de s^2 es $(\tilde{\sigma}^2/n)\chi_n^2$, donde χ^2 es una distribución chi-cuadrado con n grados de libertad. Por tanto:

$$f(s^2 | \tilde{\sigma}^2) = [\Gamma(\frac{n}{2})]^{-1} (\frac{n}{2})^{n/2} \tilde{\sigma}^{-n} (s^2)^{(n/2)-1} \exp(-\frac{ns^2}{2\tilde{\sigma}^2}), \quad s^2 > 0.$$

Como dado s^2 la cantidad $[\Gamma(n/2)]^{-1} (n/2)^{n/2} (s^2)^{(n/2)-1}$ es una constante fija, la distribución a posteriori de $\tilde{\sigma}^2$ dado s^2 es

$$f(\tilde{\sigma}^2 | s^2) \propto f(\tilde{\sigma}^2) P(s^2 | \tilde{\sigma}^2) \propto f(\tilde{\sigma}^2) \tilde{\sigma}^{-n} \exp(-\frac{ns^2}{2\tilde{\sigma}^2})$$

que es lo mismo que anteriormente obtuvimos.

Tomaremos como distribución a priori no informativa una distribución localmente uniforme en $\log \sigma$. Esto implica que

$$f(\tilde{\sigma}^2) \propto \tilde{\sigma}^{-2}.$$

Usando esto junto con la siguiente fórmula integral

$$\int_0^{\infty} x^{-(p+1)} e^{-ax^{-1}} dx = a^{-p} \Gamma(p).$$

obtenemos la constante normalizadora y llegamos a

$$f(\tilde{\sigma}^2 | \bar{x}) = k(\tilde{\sigma}^2)^{-(n/2)+1} \exp(-\frac{ns^2}{2\tilde{\sigma}^2}), \tilde{\sigma}^2 > 0.$$

donde

$$k = [\Gamma(\frac{n}{2})]^{-1} (\frac{ns^2}{2})^{n/2}.$$

Podemos ahora deducir la distribución a posteriori de la desviación estándar $\tilde{\sigma}$:

$$f(\tilde{\sigma} | \bar{x}) = k' \tilde{\sigma}^{-(n+1)} \exp(-\frac{ns^2}{2\tilde{\sigma}^2}), \tilde{\sigma} > 0.$$

donde

$$k' = [\frac{1}{2}\Gamma(\frac{n}{2})]^{-1} (\frac{ns^2}{2})^{n/2}.$$

Podemos además deducir de nuestra expresión para $f(\tilde{\sigma}^2 | \bar{x})$ la distribución a posteriori para $\log \tilde{\sigma}^2 = 2 \log \tilde{\sigma}$:

$$\begin{aligned} f(\log \tilde{\sigma}^2 | \bar{x}) &= [\Gamma(\frac{n}{2} 2^{n/2})]^{-1} (\frac{ns^2}{\tilde{\sigma}^2})^{n/2} \exp[-\frac{1}{2}(\frac{ns^2}{\tilde{\sigma}^2})] \\ &= [\Gamma(\frac{n}{2} 2^{n/2})] \exp[\frac{n}{2}(\log n + \log s^2 - \log \tilde{\sigma}^2) - \frac{1}{2} \exp(\log n + \log s^2 - \log \tilde{\sigma}^2)], \\ &\quad -\infty < \log \tilde{\sigma}^2 < \infty. \end{aligned}$$

Estudiaremos ahora algunas propiedades de las distribuciones de $\tilde{\sigma}^2$, $\tilde{\sigma}$ y $\log \tilde{\sigma}$. Usando notación y resultados precedentes, tenemos que la distribución muestral de

$$\chi_n^2 = \sum_{i=1}^n \frac{(x_i - \bar{\mu})^2}{\tilde{\sigma}^2} = \frac{ns^2}{\tilde{\sigma}^2}$$

es la siguiente:

$$f(\chi_n^2) = [\Gamma(\frac{n}{2}) 2^{n/2}]^{-1} (\chi^2)^{(n/2)-1} \exp(-\frac{1}{2}\chi^2), \chi^2 > 0.$$

que es la distribución χ^2 con n grados de libertad.

Más aún, la distribución muestral de $\chi_n = \sqrt{ns}/\sigma$ (la raíz cuadrada positiva de χ^2) es

$$f(\chi_n) = [\Gamma(\frac{n}{2}) 2^{(n/2)-1}]^{-1} (\chi)^{n-1} \exp(-\frac{1}{2}\chi^2), \chi > 0.$$

A esta distribución se la denomina distribución χ con n grados de libertad. Por último, la distribución muestral de

$$\log \chi_n^2 = (\log n + \log s^2 - \log \tilde{\sigma}^2)$$

es:

$$\begin{aligned} f(\log \chi_n^2) &= [\Gamma(\frac{n}{2})2^{n/2}]^{-1}(\chi^2)^{n/2} \exp(-\frac{1}{2}\chi^2) \\ &= [\Gamma(\frac{n}{2})2^{n/2}]^{-1} \exp[\frac{n}{2} \log \chi^2 - \frac{1}{2} \exp(\log \chi^2)], \quad -\infty < \log \chi^2 < \infty. \end{aligned}$$

En el análisis bayesiano, son los recíprocos χ_n^{-2} y χ_n^{-1} los que aparecen de manera natural. La distribución χ^2 inversa y la distribución χ inversa con n grados de libertad se obtienen haciendo las siguientes transformaciones:

$$\chi_n^{-2} = \frac{1}{\chi_n^2}, \quad \chi_n^{-1} = \frac{1}{\chi_n}$$

llegando a

$$f(\chi_n^{-2}) = [\Gamma(\frac{n}{2})2^{n/2}]^{-1}(\chi^{-2})^{-[(n/2)+1]} \exp(-\frac{1}{2\chi^{-2}}), \quad \chi^{-2} > 0$$

y de manera similar

$$f(\chi_n^{-1}) = [\Gamma(\frac{n}{2})2^{(n/2)-1}]^{-1}(\chi^{-1})^{-(n+1)} \exp(-\frac{1}{2(\chi^{-1})^2}), \quad \chi^{-1} > 0$$

Si comparamos ahora las distribuciones a posteriori para $\tilde{\sigma}$ y $\tilde{\sigma}^2$ con las precedentes observamos que, a posteriori, las cantidades $\tilde{\sigma}^2/ns^2$ y $\tilde{\sigma}/\sqrt{ns}$ se distribuyen de acuerdo a las distribuciones χ_n^{-2} y χ_n^{-1} . Es más, de las distribuciones que obtuvimos para $\log \tilde{\sigma}^2$ y para $\log \chi_n^2$ llegamos a que la distribución a posteriori para $(\log n + \log s^2 - \log \tilde{\sigma}^2)$ es la misma que $\log \chi_n^2$. Al tratar con la distribución a posteriori de la cantidad $\tilde{\sigma}^2/ns^2$ es conveniente recordar que $\tilde{\sigma}^2$ es la variable aleatoria mientras que s^2 es una cantidad fija obtenida a partir de los datos observados.

En resumen, tenemos:

Consideremos s^2 distribuída como $(\tilde{\sigma}^2/\nu)\chi_\nu^2$. Si la distribución a priori de $\log \tilde{\sigma}$ es localmente uniforme, entonces, dado s^2 , la varianza $\tilde{\sigma}^2$ tiene a $(\nu s^2)\chi_\nu^2$ como distribución a posteriori, $\tilde{\sigma}$ sigue la distribución $(\sqrt{\nu}s)\chi_\nu^{-1}$, y finalmente $\log \tilde{\sigma}^2$ hace lo propio con $\log \nu s^2 - \log \chi_\nu^2$ (o $\log \nu s^2 + \log \chi_\nu^{-2}$), donde ν es el número de grados de libertad de cada distribución.

Inferencias sobre la desviación de una Normal

Las distribuciones χ^{-2} , χ^{-1} , y $\log \chi^2$ proporcionan distribuciones a posteriori para $\tilde{\sigma}^2$, $\tilde{\sigma}$, y $\log \tilde{\sigma}^2$. Para agrupar e interpretar la información contenida en estas distribuciones se necesitará usar intervalos de **alta densidad de probabilidad**. Por ejemplo, si queremos hacer inferencia sobre la varianza $\tilde{\sigma}^2$, entonces los extremos de un $(1 - \alpha)$ intervalo M.D.P en $\tilde{\sigma}^2$ vienen dados por dos valores (σ_0^2, σ_1^2) tales que

- $f(\sigma_0^2 | \bar{x}) = f(\sigma_1^2 | \bar{x})$.
- $P(\sigma_0^2 < \sigma^2 < \sigma_1^2 | \bar{x}) = 1 - \alpha$.

Mientras que la probabilidad contenida en un intervalo dado de $\tilde{\sigma}^2$ será la misma que la probabilidad contenida en los correspondientes intervalos de $\tilde{\sigma}$ y $\log \tilde{\sigma}^2$, no es cierto que los límites de un intervalo M.D.P. en $\tilde{\sigma}^2$ corresponderán con los límites de los intervalos M.D.P. en $\tilde{\sigma}$ y $\log \tilde{\sigma}^2$. Esto se comprueba fácilmente observando que las distribuciones a posteriori $f(\tilde{\sigma}^2 | \bar{x})$, $f(\tilde{\sigma} | \bar{x})$, y $f(\log \tilde{\sigma}^2 | \bar{x})$ no son proporcionales entre sí. Por tanto, el par (σ_0^2, σ_1^2) , que satisface

$$f(\sigma_0^2 | \bar{x}) = f(\sigma_1^2 | \bar{x})$$

no verificará, necesariamente,

$$f(\sigma_0 | \bar{x}) = f(\sigma_1 | \bar{x}) \text{ ni } f(\log \sigma_0 | \bar{x}) = f(\log \sigma_1 | \bar{x}).$$

Es decir, los intervalos M.D.P. no son invariantes bajo transformaciones del parámetro a menos que dicha transformación sea lineal. Esto conduce a preguntarse cómo parametrizar de cara a establecer intervalos M.D.P.

Intervalos M.D.P. estandarizados

La inferencia sobre la dispersión de una Normal puede hacerse en términos de la varianza $\tilde{\sigma}^2$, la desviación estándar $\tilde{\sigma}$, o la constante de precisión $\tilde{\sigma}^{-2}$. El intervalo M.D.P. asociado con una probabilidad dada será ligeramente diferente dependiendo de cuál de las tres métricas se use, aunque cada intervalo incluirá la probabilidad establecida. Si usamos distribuciones a priori no informativas para cualquier parámetro, entonces buscamos un intervalo M.D.P. **estandarizado**. Estos intervalos se calculan con la métrica en la que la distribución a priori no informativa es localmente uniforme. Es decir, para la dispersión de una Normal, debemos calcular intervalos que sean M.D.P. para $\log \tilde{\sigma}$. Ya que los intervalos M.D.P. son equivalentes bajo transformaciones lineales de $\log \tilde{\sigma}$, podemos emplear intervalos para cada miembro de la clase de transformaciones

$$c + q * \log \tilde{\sigma}$$

donde c y q son dos constantes arbitrarias. En particular, trabajaremos con

$$\log \chi_n^2 = \log \frac{ns^2}{\tilde{\sigma}^2} = \log ns^2 - 2 * \log \tilde{\sigma}.$$

Los valores de los límites para estos intervalos M.D.P. para $\log \chi_n^2$ pueden ser obtenidos a partir de una tabla de valores para la distribución χ^2 .

5.3.2.2. Distribución a priori conjugada

Buscamos una distribución a priori que sea conjugada con respecto a la verosimilitud para $\tilde{\sigma}^2$ desconocido. Una tal distribución es la distribución Gamma inversa:

$$f(\tilde{\sigma}^2) = \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{1}{(\tilde{\sigma}^2)^{\alpha+1}} \exp\left(-\frac{\beta}{\tilde{\sigma}^2}\right), \quad 0 < \tilde{\sigma}^2 < \infty.$$

La media y la varianza de esta distribución vienen expresadas como:

- Media: $\frac{\beta}{\alpha-1}$, $\alpha > 1$
- Varianza: $\frac{\beta^2}{(\alpha-1)^2(\alpha-2)}$

Calculamos ahora la función de densidad a posteriori, que será función de densidad de una distribución Gamma inversa:

$$\begin{aligned} f(\tilde{\sigma}^2 | \underline{x}) &\propto \frac{1}{(\tilde{\sigma}^2)^{\alpha+1}} \exp\left(-\frac{\beta}{\tilde{\sigma}^2}\right) \times \frac{1}{(\tilde{\sigma}^2)^{n/2}} \exp\left[-\frac{ns^2}{2\tilde{\sigma}^2}\right], \quad 0 < \tilde{\sigma}^2 < \infty \\ &\propto \frac{1}{(\tilde{\sigma}^2)^{\frac{n}{2}+\alpha+1}} \exp\left[-\frac{1}{\tilde{\sigma}^2}\left(\frac{ns^2}{2} + \beta\right)\right], \quad 0 < \tilde{\sigma}^2 < \infty. \end{aligned}$$

Y esto corresponde, efectivamente, a la función de densidad de una Gamma inversa. Por tanto, a posteriori, se tiene:

$$\tilde{\sigma}^2 \sim IG\left(\alpha + \frac{n}{2}, \beta + \frac{ns^2}{2}\right).$$

Hagamos una pequeña observación. Hemos definido la precisión como $\tau^2 = 1/\tilde{\sigma}^2$. Ya que la varianza sigue una distribución Gamma inversa, deducimos que la distribución a posteriori de la precisión es una Gamma con los mismos parámetros, es decir:

$$\tau^2 \sim G\left(\alpha + \frac{n}{2}, \beta + \frac{ns^2}{2}\right).$$

5.3.3. Caso $\tilde{\mu}$, $\tilde{\sigma}$ desconocidos

5.3.3.1. Distribución a priori no informativa

Consideremos una muestra aleatoria de n observaciones independientes obtenida de una población Normal $N(\tilde{\mu}, \tilde{\sigma}^2)$ donde tanto $\tilde{\mu}$ como $\tilde{\sigma}$ son desconocidos. La media muestral \bar{x} y la varianza muestral $s^2 = \nu^{-1} \sum_{i=1}^n (x_i - \bar{x})^2$, $\nu = n - 1$, son suficientes para $(\tilde{\mu}, \tilde{\sigma})$ y están independientemente distribuidos como $N(\tilde{\mu}, \tilde{\sigma}/n)$ y $(\tilde{\sigma}_\nu^2/\nu)\chi_\nu^2$ respectivamente. En este caso, la función de verosimilitud es:

$$\begin{aligned} L(\tilde{\mu}, \tilde{\sigma}^2 | \underline{x}) &\propto f(\bar{x} | \tilde{\mu}, \tilde{\sigma}^2) f(s^2 | \tilde{\sigma}^2) \\ &\propto \tilde{\sigma}^{-n} \exp\left\{-\frac{1}{2\tilde{\sigma}^2}[\nu s^2 + n(\tilde{\mu} - \bar{x})^2]\right\} \end{aligned}$$

donde $f(\bar{x} | \tilde{\mu}, \tilde{\sigma}^2)$ y $f(s^2 | \tilde{\sigma}^2)$ están distribuidas, respectivamente, de la siguiente manera:

$$f(\bar{x} | \tilde{\mu}, \tilde{\sigma}^2) = \frac{\sqrt{n}}{\sqrt{2\pi\tilde{\sigma}}} \exp\left[-\frac{n}{2\tilde{\sigma}^2}(\bar{x} - \tilde{\mu})^2\right]$$

$$f(s^2 | \tilde{\sigma}^2) = \frac{1}{\Gamma[\frac{1}{2}(n-1)]} \left(\frac{n-1}{2\tilde{\sigma}^2}\right)^{\frac{1}{2}(n-1)} (s^2)^{\frac{1}{2}(n-1)-1} \times \exp\left[-\frac{(n-1)s^2}{2\tilde{\sigma}^2}\right], \quad s > 0.$$

Por tanto, dados los datos \underline{x} , la distribución conjunta a posteriori de $(\tilde{\mu}, \tilde{\sigma})$ es:

$$f(\tilde{\mu}, \tilde{\sigma} | \underline{x}) \propto f(\tilde{\mu}, \tilde{\sigma}) f(\bar{x} | \tilde{\mu}, \tilde{\sigma}^2) f(s^2 | \tilde{\sigma})$$

donde $f(\tilde{\mu}, \tilde{\sigma})$ es la distribución a priori.

Es conveniente reparar ahora en el hecho de que no es inadecuado considerar un parámetro de situación independientemente distribuido de un parámetro de escala. En efecto, cualquier idea preconcebida que se tenga sobre el parámetro de situación de una distribución no se verá influenciada, de forma general, por la idea que se tenga acerca del valor del parámetro de escala. Tomando nuestro caso de distribución normal, donde μ y σ son parámetros de situación y escala respectivamente, tenemos que $f(\tilde{\mu} | \tilde{\sigma}) \doteq f(\tilde{\mu})$ (donde hemos asumido que a priori $\tilde{\mu}$ y $\tilde{\sigma}$ son aproximadamente independientes) y, por tanto, concluimos que

$$f(\tilde{\mu}, \tilde{\sigma}) = f(\tilde{\mu} | \tilde{\sigma}) f(\tilde{\sigma}) \doteq f(\tilde{\mu}) f(\tilde{\sigma}).$$

Adoptamos como distribuciones a priori no informativas para los parámetros las siguientes:

$$f(\tilde{\mu}) \propto c \quad f(\log \tilde{\sigma}) \propto c, \quad f(\tilde{\sigma}) \propto \tilde{\sigma}^{-1}.$$

Usando lo anterior llegamos a que la distribución a posteriori es:

$$f(\tilde{\mu}, \tilde{\sigma} | \underline{x}) = k \tilde{\sigma}^{-(n+1)} \exp\left[-\frac{1}{2\tilde{\sigma}^2}[\nu s^2 + n(\tilde{\mu} - \bar{x})^2]\right], \quad -\infty < \tilde{\mu} < \infty, \quad \tilde{\sigma} > 0$$

donde $\nu = n - 1$ y k es la constante de normalización. Usaremos las siguientes fórmulas de integración:

$$\int_0^\infty x^{-(p+1)} e^{-ax^{-2}} dx = \frac{1}{2} a^{-\frac{p}{2}} \Gamma(p/2), \quad a > 0, \quad p > 0$$

$$\int_{-\infty}^\infty \exp\left[-\frac{1}{2}\left(\frac{x-\eta}{c}\right)^2\right] dx = \sqrt{2\pi}c, \quad -\infty < \eta < \infty, \quad c > 0$$

para llegar a

$$k = \sqrt{\frac{n}{2\pi}} \left[\frac{1}{2}\Gamma\left(\frac{\nu}{2}\right)\right]^{-1} \left(\frac{\nu s^2}{2}\right)^{\nu/2}.$$

En conclusión:

$$f(\tilde{\mu}, \tilde{\sigma} | \underline{x}) = \sqrt{\frac{n}{2\pi}} \left[\frac{1}{2}\Gamma\left(\frac{\nu}{2}\right)\right]^{-1} \left(\frac{\nu s^2}{2}\right)^{\nu/2} \tilde{\sigma}^{-(n+1)} \exp\left[-\frac{1}{2\tilde{\sigma}^2}[\nu s^2 + n(\tilde{\mu} - \bar{x})^2]\right], \\ -\infty < \tilde{\mu} < \infty, \quad \tilde{\sigma} > 0.$$

La distribución conjunta a posteriori de $(\tilde{\mu}, \tilde{\sigma})$ obtenida puede escribirse como

$$f(\tilde{\mu}, \tilde{\sigma} | \underline{x}) = f(\tilde{\mu} | \tilde{\sigma}, \underline{x})f(\tilde{\sigma} | \underline{x})$$

donde el primer factor es la distribución condicionada a posteriori de $\tilde{\mu}$ dado $\tilde{\sigma}$, y el segundo factor es la distribución marginal a posteriori de $\tilde{\sigma}$. Estudiemos ahora las distribuciones componentes de $f(\tilde{\mu}, \tilde{\sigma} | \underline{x})$.

Distribución de $\tilde{\mu}$ condicionada a $\tilde{\sigma}$

Si tratamos a $\tilde{\sigma}$ como una constante conocida, la distribución condicional a posteriori de $\tilde{\mu}$ dado $\tilde{\sigma}$ es

$$f(\tilde{\mu} | \tilde{\sigma}, \underline{x}) = \frac{\sqrt{n}}{\sqrt{2\pi\tilde{\sigma}}} \exp\left[-\frac{n}{2\tilde{\sigma}^2}(\tilde{\mu} - \bar{x})^2\right], \quad -\infty < \tilde{\mu} < \infty.$$

Es decir, dado $\tilde{\sigma}$, $\tilde{\mu}$ está distribuida a posteriori según una normal $N(\bar{x}, \tilde{\sigma}^2/n)$.

Distribución marginal de $\tilde{\sigma}$

La distribución marginal a posteriori de $\tilde{\sigma}$ es

$$f(\tilde{\sigma} | \bar{x}) = \frac{f(\tilde{\mu}, \tilde{\sigma} | \bar{x})}{f(\tilde{\mu} | \tilde{\sigma}, \bar{x})}.$$

De la distribución a posteriori y la distribución de $\tilde{\mu}$ condicionada a $\tilde{\sigma}$ llegamos a

$$f(\tilde{\sigma} | \underline{x}) = k' \tilde{\sigma}^{-(\nu+1)} \exp\left(-\frac{\nu s^2}{2\tilde{\sigma}^2}\right), \quad \tilde{\sigma} > 0.$$

donde

$$k' = \left[\frac{1}{2}\Gamma\left(\frac{\nu}{2}\right)\right]^{-1} \left(\frac{\nu s^2}{2}\right)^{\nu/2}.$$

Dicho de otro modo, $\tilde{\sigma}$ está distribuida a posteriori como $\sqrt{\nu s} \chi_{\nu}^{-1}$. Esta distribución es la misma que obtuvimos cuando la media era conocida y la varianza desconocida, pero con la diferencia de que cuando la media es conocida la distribución de $\tilde{\sigma}/(\sqrt{n}s)$, con $ns^2 = \sum_{i=1}^n (x_i - \mu)^2$, tiene n grados de libertad, mientras que cuando la media es desconocida, la distribución de $\tilde{\sigma}/(\sqrt{\nu s})$, con $\nu s^2 = \sum_{i=1}^n (x_i - \bar{x})^2$, tiene $\nu = n - 1$ grados de libertad. Las inferencias sobre $\tilde{\sigma}$, con $\tilde{\mu}$ desconocido, pueden realizarse tal como hicimos en el caso μ conocido y $\tilde{\sigma}$ desconocido.

Distribución marginal a posteriori de $\tilde{\mu}$

La distribución marginal a posteriori de $\tilde{\mu}$ puede ser obtenida integrando con respecto a $\tilde{\sigma}$ en la distribución conjunta a posteriori de $\tilde{\mu}$ y $\tilde{\sigma}$:

$$f(\tilde{\mu} | \underline{x}) = \int_0^{\infty} f(\tilde{\mu}, \sigma | \underline{x}) d\tilde{\sigma} = \int_0^{\infty} f(\tilde{\mu} | \tilde{\sigma}, \underline{x}) f(\tilde{\sigma} | \underline{x}) d\tilde{\sigma}.$$

Usaremos la siguiente fórmula integral

$$\int_0^\infty x^{-(p+1)} e^{-ax^{-2}} dx = \frac{1}{2} a^{-\frac{p}{2}} \Gamma(p/2), \quad a > 0, \quad p > 0$$

para llegar a

$$f(\tilde{\mu} | \underline{x}) = \frac{(s/\sqrt{n})^{-1}}{B(\frac{1}{2}\nu, \frac{1}{2})\sqrt{\nu}} \left[1 + \frac{n(\tilde{\mu} - \bar{x})^2}{\nu s^2}\right]^{-\frac{1}{2}(\nu+1)}, \quad -\infty < \tilde{\mu} < \infty$$

donde $B(p, q)$ es la función beta $B(p, q) = \Gamma(p)\Gamma(q)/\Gamma(p+q)$.

Si escribimos

$$t = \frac{\tilde{\mu} - \bar{x}}{s/\sqrt{n}}$$

entonces tenemos

$$f(t = \frac{\tilde{\mu} - \bar{x}}{s/\sqrt{n}} | \underline{x}) = \frac{1}{B(\frac{1}{2}\nu, \frac{1}{2})\sqrt{\nu}} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{1}{2}(\nu+1)}, \quad -\infty < t < \infty$$

que será una distribución t de Student con $\nu = n - 1$ grados de libertad. Conviene recordar que $\tilde{\mu}$ es la variable aleatoria en la expresión y (\bar{x}, s^2) son valores muestrales conocidos. Notaremos a la distribución obtenida como $t(\bar{x}, s^2/n, \nu)$.

Agrupamos ahora los resultados y resumimos lo que hemos obtenido:

Consideremos los valores muestrales \bar{x} y s^2 distribuidos independientemente según distribuciones $N(\tilde{\mu}, \tilde{\sigma}^2/n)$ y $(\tilde{\sigma}^2/\nu)\chi_\nu^2$ respectivamente. Supongamos que a priori $\tilde{\mu}$ y $\log \tilde{\sigma}$ están distribuidas (aproximadamente) de manera independiente y localmente uniforme. Entonces, dados (\bar{x}, s^2) , se tiene:

- $\tilde{\sigma}$ sigue la distribución $\sqrt{\nu} s \chi_\nu^{-1}$.
- $\tilde{\mu}$ condicionada a $\tilde{\sigma}$ está distribuida según una normal $N(\bar{x}, \sigma^2/n)$.
- $\tilde{\mu}$ tiene distribución marginal $t(\bar{x}, s^2/n, \nu)$.

Intervalos a posteriori para $\tilde{\mu}$

La distribución a posteriori obtenida para $\tilde{\mu}$ es una distribución simétrica centrada en \bar{x} con factor de escala s/\sqrt{n} . Para obtener intervalos M.D.P. para la media debemos usar una tabla de valores de la distribución t de Student.

Si denotamos $t_{(\alpha/2)}(\tau)$ como el valor para el cual

$$P\{t > t_{(\alpha/2)}(\tau)\} = \alpha/2$$

entonces, por la simetría de la distribución t de Student,

$$P\{t < -t_{(\alpha/2)(\tau)}\} = \alpha/2$$

y por tanto

$$P\{|t| < t_{(\alpha/2)(\tau)}\} = 1 - \alpha/2.$$

Se deduce que los límites de un $(1 - \alpha)$ intervalo M.D.P. de $t(\bar{x}, s^2/n, \tau)$ vienen dados por

$$\bar{x} \pm \left(\frac{s}{\sqrt{n}}\right)t_{(\alpha/2)(\tau)}.$$

5.3.3.2. Distribución conjugada

Cuando existe conocimiento a priori sobre los parámetros desconocidos de una Normal, seguramente queramos incorporarlos a nuestro análisis Bayesiano. Usualmente se asume independencia a priori entre los parámetros, y se toma como distribución a priori el producto de una Normal sobre la media y una Gamma inversa para la varianza.

En este caso, ninguna de las distribuciones marginales a posteriori pertenece a la familia paramétrica, y la densidad conjunta a posteriori claramente no será producto de una Normal y una Gamma inversa, por tanto, no estamos afrontando un análisis conjugado propiamente dicho. Para designar esta situación se habla de análisis semi-conjugado, ya que la distribución a priori sería conjugada si uno de los parámetros fuera conocido.

Sin embargo, sí es posible establecer una distribución a priori conjugada cuando los parámetros son desconocidos. La densidad conjunta a priori, que llamaremos densidad Normal-Gamma inversa, es el producto de la marginal Gamma inversa para $\tilde{\sigma}^2$ y la densidad (Normal) de $\tilde{\mu}$ dada $\tilde{\sigma}^2$:

$$f(\tilde{\mu}, \tilde{\sigma}^2) = f(\tilde{\sigma}^2)f(\tilde{\mu} | \tilde{\sigma}^2) = IG(\alpha, \beta) \times N(\mu_0, \frac{1}{k}\tilde{\sigma}^2).$$

Una vez obtenidos los datos y la media muestral \bar{x} , podemos aplicar el teorema de Bayes

$$\begin{aligned} f(\tilde{\mu}, \tilde{\sigma}^2 | \bar{x}) &\propto \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{1}{(\tilde{\sigma}^2)^{\alpha+1}} \exp\left(-\frac{\beta}{\tilde{\sigma}^2}\right) \times \frac{\sqrt{k}}{\sqrt{2\pi\tilde{\sigma}^2}} \exp\left(-\frac{k(\tilde{\mu}-\mu_0)^2}{2\tilde{\sigma}^2}\right) \\ &\quad \times (\tilde{\sigma})^{-n} e^{-ns^2/2\tilde{\sigma}^2} \times e^{-n(\tilde{\mu}-\bar{x})^2/2\tilde{\sigma}^2} \\ &\propto \frac{1}{(\tilde{\sigma}^2)^{\alpha+\frac{n}{2}+1}} \exp\left(-\frac{1}{\tilde{\sigma}^2}\left(\beta + \frac{ns^2}{2}\right)\right) \times \frac{\sqrt{k}}{\tilde{\sigma}} \exp\left(-\frac{k(\tilde{\mu}-\mu_0)^2+n(\tilde{\mu}-\bar{x})^2}{2\tilde{\sigma}^2}\right). \end{aligned}$$

Haremos ahora algunas manipulaciones en el numerador dentro de la última exponencial:

$$\begin{aligned}
k(\tilde{\mu} - \mu_0)^2 + n(\tilde{\mu} - \bar{x})^2 &= k(\tilde{\mu}^2 - 2\tilde{\mu}\mu_0 + \mu_0^2) + n(\tilde{\mu}^2 - 2\tilde{\mu}\bar{x} + \bar{x}^2) \\
&= (k+n)\tilde{\mu}^2 - 2\tilde{\mu}(k\mu_0 + n\bar{x}) + k\mu_0^2 + n\bar{x}^2 \\
&= (k+n)\left[\tilde{\mu}^2 - \frac{2\tilde{\mu}(k\mu_0 + n\bar{x})}{k+n} + \left(\frac{k\mu_0 + n\bar{x}}{k+n}\right)^2\right] - \frac{(k\mu_0 + n\bar{x})^2}{k+n} + k\mu_0^2 + n\bar{x}^2 \\
&= (k+n)\left(\tilde{\mu} - \frac{k\mu_0 + n\bar{x}}{k+n}\right)^2 + \frac{kn(\bar{x} - \mu_0)^2}{k+n}.
\end{aligned}$$

Usando esto en la expresión para la distribución conjunta a posteriori, finalmente tenemos:

$$\begin{aligned}
f(\tilde{\mu}, \tilde{\sigma}^2 | \bar{x}) &\propto \frac{1}{(\tilde{\sigma}^2)^{\alpha + \frac{n}{2} + 1}} \exp\left[-\frac{1}{\tilde{\sigma}^2}\left(\beta + \frac{ns^2}{2} + \frac{kn(\bar{x} - \mu_0)^2}{2(k+n)}\right)\right] \\
&\quad \times \frac{\sqrt{k}}{\tilde{\sigma}} \exp\left[-(k+n)\left(\frac{\tilde{\mu} - \frac{k\mu_0 + n\bar{x}}{k+n}}{2\tilde{\sigma}^2}\right)^2\right].
\end{aligned}$$

Lo obtenido en nuestros cálculos corresponde a una distribución Normal-Gamma inversa:

$$\begin{aligned}
\tilde{\mu}, \tilde{\sigma}^2 | \bar{x} &\sim IG\left(\alpha + \frac{n}{2}, \beta + \frac{ns^2}{2} + \frac{kn(\bar{x} - \mu_0)^2}{2(k+n)}\right) \\
&\quad \times N\left(\frac{k\mu_0 + n\bar{x}}{k+n}, \frac{\tilde{\sigma}^2}{k+n}\right).
\end{aligned}$$

5.4. Distribución predictiva

Estudiaremos la distribución predictiva a posteriori para una distribución normal con media desconocida y varianza conocida.

Sean x_1, \dots, x_n una muestra aleatoria obtenida de una población normal $N(\tilde{\mu}, \sigma^2)$. Nuestro propósito es predecir una observación futura independiente $Y = X_{n+1}$ de la misma distribución. Usando la distribución a priori de Jeffreys para la media, es decir, $f(\tilde{\mu}) \sim 1$, teníamos la distribución a posteriori $\tilde{\mu} \sim N(\bar{x}, \sigma^2/n)$.

La distribución a posteriori predictiva de Y es, usando la fórmula dada en el capítulo segundo, la siguiente:

$$\begin{aligned}
f(y | \underline{x}) &= \int f(y | \tilde{\mu}) f(\tilde{\mu} | \underline{x}) d\tilde{\mu} \\
&\propto \int \exp\left\{-\frac{1}{2}\left[\frac{(\tilde{\mu} - y)^2}{\sigma^2} + \frac{n(\tilde{\mu} - \bar{x})^2}{\sigma^2}\right]\right\} d\tilde{\mu}.
\end{aligned}$$

Consideremos el siguiente resultado:

$$A(x - a)^2 + B(x - b)^2 = C(x - c)^2 + \frac{AB}{C}(a - b)^2$$

donde $C=A+B$ y $c=(Aa+Bb)/C$.

Podemos usar este resultado para obtener

$$\frac{(\tilde{\mu} - y)^2}{\sigma^2} + \frac{n(\tilde{\mu} - \bar{x})^2}{\sigma^2} = C(\tilde{\mu} - c)^2 + \frac{(y - \bar{x})^2}{(1 + \frac{1}{n})\sigma^2}$$

donde $C = (n + 1)/\sigma^2$ y $c = (y + n\bar{x})/(n + 1)$. Observemos que, mientras el primer término de la suma es una forma cuadrática de $\tilde{\mu}$, el segundo término no depende del parámetro desconocido.

Aplicando lo anterior a nuestra expresión para la distribución a posteriori obtenemos

$$\begin{aligned} f(y|\underline{x}) &\propto \int \exp\left\{-\frac{C}{2}(\tilde{\mu} - c)^2 - \frac{1}{2}\frac{(y-\bar{x})^2}{(1+\frac{1}{n})\sigma^2}\right\}d\tilde{\mu} \\ &= \exp\left\{-\frac{1}{2}\frac{(y-\bar{x})^2}{(1+\frac{1}{n})\sigma^2}\right\} \int \exp\left\{-\frac{C}{2}(\tilde{\mu} - c)^2\right\}d\tilde{\mu} \\ &\propto \exp\left\{-\frac{1}{2}\frac{(y-\bar{x})^2}{(1+\frac{1}{n})\sigma^2}\right\}. \end{aligned}$$

donde se ha usado que

$$\int \exp\left\{-\frac{C}{2}(\tilde{\mu} - c)^2\right\}d\tilde{\mu} = \frac{\sqrt{2\pi\sigma}}{\sqrt{n+1}}.$$

Hemos obtenido la función de densidad de una normal de media \bar{x} y varianza $\sigma^2(1 + \frac{1}{n})$, y por tanto la distribución predictiva a posteriori es

$$Y \sim N(\bar{x}, \sigma^2(1 + \frac{1}{n})).$$

5.5. Test de hipótesis

Estudiaremos algunos contrastes de hipótesis relativos a la distribución normal, y veremos la relación existente entre el enfoque Bayesiano y el p-valor clásico.

5.5.1. Media de una normal con varianza desconocida

Consideremos el contraste siguiente:

$$\begin{cases} H_0 : \mu \leq \mu_0 \\ H_1 : \mu > \mu_0 \end{cases}$$

y los estadísticos muestrales \bar{x}_0 y s^2 . El p-valor clásico es

$$P\left(\frac{\bar{x}-\mu_0}{\bar{s}/\sqrt{n}} \geq \frac{\bar{x}_0-\mu_0}{s/\sqrt{n}}\right) = P(\tilde{t}(\tau) \geq \frac{\bar{x}_0-\mu_0}{s/\sqrt{n}}).$$

Un análisis bayesiano con distribución a priori no informativa da lugar a la distribución a posteriori $t(\bar{x}_0, s^2/n, \tau)$ para $\tilde{\mu}$.

$$\begin{aligned} P_t(\tilde{\mu} \leq \mu_0 | \bar{x}_0, s^2/n, \tau) &= P\left(\frac{\tilde{\mu}-\bar{x}_0}{s/\sqrt{n}} \geq \frac{\mu_0-\bar{x}_0}{s/\sqrt{n}}\right) \\ &= P(\tilde{t}(\tau) \geq \frac{\mu_0-\bar{x}_0}{s/\sqrt{n}}). \end{aligned}$$

Ya que la distribución t de Student es simétrica con respecto al cero, el área a la derecha, en sentido clásico

$$\frac{\bar{x}_0-\mu_0}{s/\sqrt{n}}$$

es la misma que el área a la izquierda bajo

$$\frac{\mu_0-\bar{x}_0}{s/\sqrt{n}}.$$

5.5.2. Varianza

Para realizar el contraste siguiente

$$\begin{cases} H_0 : \sigma^2 \leq \sigma_0^2 \\ H_1 : \sigma^2 > \sigma_0^2 \end{cases}$$

dados datos recogidos en los valores n y s_0^2 , el p-valor clásico es

$$P(\tilde{s}^2 \geq s_0^2 | \nu, \sigma_0^2) = P\left(\frac{\nu \tilde{s}^2}{\sigma_0^2} \geq \frac{\nu s_0^2}{\sigma_0^2} | \nu, \sigma_0^2\right) = P_{\chi^2}(\tilde{\chi}^2(\nu) \geq \frac{\nu s_0^2}{\sigma_0^2}).$$

Una aproximación Bayesiana al problema considerando una distribución a priori no informativa resulta en que la distribución a posteriori para

$$\left(\frac{\nu s_0^2}{\tilde{\sigma}^2} | \nu, s_0^2\right)$$

es una distribución χ^2 de parámetro ν , y el interés estaría en calcular lo que sigue:

$$P(\tilde{\sigma}^2 \leq \sigma_0^2 | \nu, s_0^2) = P\left(\frac{\nu s_0^2}{\tilde{\sigma}^2} > \frac{\nu s_0^2}{\sigma_0^2}\right) = P_{\chi^2}(\tilde{\chi}^2(\nu) > \frac{\nu s_0^2}{\sigma_0^2}).$$

Coincide con el p-valor en sentido clásico.

Capítulo 6

Inferencia Bayesiana en R

En este capítulo se definirán instrucciones para el programa R que permitirán realizar inferencia Bayesiana usando los procedimientos descritos en el presente trabajo. Se empleará el paquete “LearnBayes”, disponible en el CRAN de R.

6.1. Inferencia sobre una proporción

Selección de una Beta a priori dados dos cuantiles

Descripción: encuentra los parámetros de una densidad Beta que se corresponda con los valores de dos cuantiles de la distribución.

Función: `beta.select(quantile1, quantile2)`.

Argumentos:

`quantile1`: lista con las componentes p , el valor de la primera probabilidad, y x , el valor del primer cuantil.

`quantile2`: lista con las componentes p , el valor de la segunda probabilidad, y x , el valor del segundo cuantil.

Resultado: vector con los parámetros de la distribución Beta.

Distribución a posteriori para muestra Binomial y Beta a priori

Descripción: Calcula los parámetros y las probabilidades a posteriori para un problema de muestreo binomial donde la distribución a priori es una mezcla discreta de densidades Beta.

Función: `binomial.beta.mix(probs,betapar,data)`

Argumentos:

probs: vector de probabilidades de las componentes Beta de la distribución a priori.
betapar: matriz en la que cada fila contiene los parámetros de una componente Beta de la distribución a priori.

data: vector con el número de éxitos y el número de fracasos.

Resultado:

probs: vector de probabilidades de las componentes Beta a posteriori.
betapar: matriz en la que cada fila contiene los parámetros de una componente Beta de la distribución a posteriori.

Distribución a posteriori de una proporción con distribución a priori discreta

Descripción: calcula la distribución a posteriori de una proporción a partir de una distribución a priori discreta.

Función: pdisc(p, prior, data)

Argumentos:

p: vector con los valores de la proporción.

prior: vector con probabilidades a priori.

data :vector con el número de éxitos y el número de fracasos.

Resultado: vector de probabilidades a posteriori.

Distribución predictiva para una muestra binomial con distribución Beta a priori

Descripción: calcula la distribución predictiva del número de éxitos para un experimento Binomial con distribución a priori Beta para la proporción.

Función: pbetap(ab, n, s)

Argumentos:

ab: vector de parámetros de la Beta a priori.

n: tamaño de la muestra binomial futura.

s: vector con el número de éxitos para el futuro experimento Binomial.

Resultado: vector de probabilidades predecidas para los valores del vector s.

Distribución predictiva para una muestra binomial con distribución discreta a priori

Descripción: calcula la distribución predictiva del número de éxitos para un experimento Binomial con distribución a priori discreta para la proporción.

Función: `pdiscp(p, probs, n, s)`

Argumentos:

`p`: vector de valores de la proporción.

`probs`: vector de probabilidades.

`n`: tamaño de la muestra binomial futura.

`s`: vector con el número de éxitos para el futuro experimento Binomial.

Resultado: vector de probabilidades predecidas para los valores del vector `s`.

Test Bayesiano sobre una proporción

Descripción: contraste de hipótesis bayesiano de que una proporción sea igual a un valor especificado usando una beta a priori.

Función: `pbetat(p0,prob,ab,data)`

Argumentos:

`p0`: valor de la proporción que se quiere contrastar.

`prob`: probabilidad a priori de la hipótesis.

`ab`: vector de parámetros de la Beta a priori bajo la hipótesis alternativa.

`data`: vector con el número de éxitos y el número de fracasos.

Resultado:

`bf`: factor Bayes en favor de la hipótesis nula.

`post`: probabilidad a posteriori de la hipótesis nula.

6.2. Distribución de Poisson

Distribución a posteriori para una muestra Poisson con distribución Gamma a priori

Descripción: calcula los parámetros y las probabilidades para un problema de muestreo de Poisson donde la distribución a priori es una mezcla discreta de densidades gamma.

Función: poisson.gamma.mix(probs,gammapar,data)

Argumentos:

probs: vector de probabilidades de las componentes Gamma a priori.

gammapar: matriz en la que cada fila contiene los parámetros de una Gamma componente de la distribución a priori.

data: lista con componentes y, vector de conteo, y t, vector de intervalos de tiempo.

Resultado:

probs: vector de probabilidades de las componentes Gamma a posteriori.

gammapar: matriz en la que cada fila contiene los parámetros de una Gamma componente de la distribución a posteriori.

6.3. Distribución Normal

Distribución a priori Normal dados dos cuantiles

Descripción: ofrece la media y la desviación estándar de una Normal que concuerde con los dos cuantiles dados.

Función: normal.select(quantile1, quantile2)

Argumentos:

quantile1: lista con las componentes p, el valor de la primera probabilidad, y x, el valor del primer cuantil.

quantile2: lista con las componentes p, el valor de la segunda probabilidad, y x, el valor del segundo cuantil..

Resultado:

mean: media de la distribución Normal correspondiente.

sigma: desviación estándar de la distribución Normal correspondiente.

Distribución a posteriori para una muestra Normal con distribución Normal a priori

Descripción: calcula los parámetros y las probabilidades para un problema de muestreo normal, varianza conocida, donde la distribución a priori es una mezcla discreta de densidades normales.

Función: `normal.normal.mix(probs,normalpar,data)`

Argumentos:

`probs`: vector con las probabilidades de las distribuciones Normales componentes a priori.

`normalpar`: matriz en la que cada fila contiene la media y la varianza de una componente Normal a priori.

`data`: vector de observaciones y varianza muestral.

Resultado:

`probs`: vector con las probabilidades de las distribuciones Normales componentes a posteriori.

`normalpar`: matriz en la que cada fila contiene la media y la varianza de una componente Normal a posteriori.

Distribución predictiva para una Normal

Descripción: dados experimentos simulados de una muestra Normal a posteriori, esta función da experimentos simulados de la distribución predictiva a posteriori de un estadístico de interés.

Función: `normpostpred(parameters,sample.size,f=min)`

Argumentos:

`parameters`: lista de experimentos simulados de la distribución a posteriori donde μ contiene la media Normal y σ^2 la varianza.

`sample.size`: tamaño muestral de la futura muestra.

`f`: función que define el estadístico.

Resultado: muestra simulada de la distribución predictiva a posteriori del estadístico.

Test de hipótesis unilateral sobre la media de una Normal

Descripción: realiza un contraste bayesiano sobre la hipótesis de que la media de una Normal sea menor o igual que un valor específico.

Función: `mnormt.onesided(m0,normpar,data)`

Argumentos:

`m0`: valor de la media que se quiere contrastar.

normpar: vector de media y desviación estándar de la distribución Normal a priori.
data: vector de media muestral, tamaño muestral, y valor conocido de la desviación estándar poblacional.

Resultado:

Bf: factor Bayes en favor de la hipótesis nula.

prior.odds: probabilidades a priori condicionadas a la hipótesis nula.

post.odds: probabilidades a posteriori condicionadas a la hipótesis nula.

postH: probabilidad a posteriori de la hipótesis nula.

Test de hipótesis bilateral sobre la media

Descripción: test Bayesiano de que la media de una Normal sea igual a un valor específico usando una Normal a priori.

Función: `mnormt.twosided(m0, prob, t, data)`

Argumentos:

m0: valor de la media a contrastar.

prob: probabilidad a priori de la hipótesis.

t: vector de los valores de la desviación estándar a priori bajo la hipótesis alternativa.

data: vector de media muestral, tamaño muestral, y valor conocido de la desviación estándar poblacional.

Resultado:

bf: vector de valores del factor Bayes en favor de la hipótesis nula.

post: vector de probabilidades a posteriori de la hipótesis nula.

6.4. Ejemplo práctico

Veremos ahora una situación del mundo real que nos permitirá aplicar nuestros conocimientos sobre la inferencia Bayesiana con ayuda de R.

Consideremos el número de trasplantes, n llevados a cabo en un hospital durante un mes (30 días), donde se registrarán el número de muertes, y . Podemos además predecir la probabilidad de que un determinado paciente muera a partir de su historial clínico, gravedad de la enfermedad, etc. Basándonos en estas probabilidades, podemos obtener el número esperado de muertes, denotado e . Podemos suponer que el número

de muertes sigue una distribución de Poisson con media $e\lambda$.

Nuestra información a priori viene de los datos de trasplantes de hospitales con similares características que el hospital que estudiamos. Supongamos que disponemos de datos extraídos de diez hospitales, y en cada uno de ellos observamos m_i muertes, y l_i pacientes sometidos a la operación, donde $i = 1, \dots, 10$. Los m_i siguen una distribución de Poisson de media $l_i\lambda$. Tomaremos la distribución a priori no informativa $f(\lambda) \propto \lambda^{-1}$. Sea α la suma de todas las m_i , y β la suma de las l_i , entonces nuestra distribución a priori para λ (en el hospital que estudiamos) es una distribución Gamma $G(\alpha, \beta)$. Supongamos que hemos obtenido $\alpha = 13$ y $\beta = 12782$.

En cuanto a nuestro hospital, se han producido 2 muertes (yobs en el programa) en un total de 82 pacientes (pa en el programa).

```
> alpha=13; beta=12782
> yobs=2; pa=87
> y=0:10
> lam=alpha / beta
> py=dpois(y, lam*pa)*dgamma(lam, shape = alpha,
+   rate = beta) / dgamma(lam, shape= alpha + y,
+   rate = beta + pa)
> cbind(y, round(py, 3))
```

```
      y
[1,]  1 0.080
[2,]  2 0.004
[3,]  3 0.000
[4,]  4 0.000
[5,]  5 0.000
[6,]  6 0.000
[7,]  7 0.000
[8,]  8 0.000
[9,]  9 0.000
[10,] 10 0.000
```

Una aproximación de la distribución a posteriori de λ puede ser obtenida al simular 1000 valores de la densidad Gamma:

```
> lambdaA = rgamma(1000, shape = alpha + yobs, rate = beta + pa)
```


Bibliografía

- [1] HANNS L. HARNEY, *Bayesian inference : parameter estimation and decisions*, New York : Springer, 2003.
- [2] GEORGE E.P. BOX Y GEORGE C. TIAO, *Bayesian inference in statistical analysis*, New York [etc.] John Wiley and Sons, 1992.
- [3] MARY KATHRYN COWLES, *Applied Bayesian Statistics: With R and OpenBUGS Examples*, New York, NY : Springer New York : Imprint: Springer, 2013.
- [4] JEAN-MICHEL MARIN, CHRISTIAN P. ROBERT, *Bayesian essentials with R*, segunda edición New York [etc.] : Springer, cop. 2014 .
- [5] LEONHARD HELD, DANIEL SABANÉS BOVÉ, *Applied Statistical Inference: Likelihood and Bayes* , Berlin, Heidelberg : Springer Berlin Heidelberg : Imprint: Springer, 2014 .
- [6] GORDON ANTELMAN, *Elementary bayesian statistics* , Cheltenham, UK [etc.] Edward Elgar, 1997 .
- [7] JIM ALBERT, MARIA RIZZO, *R by Example: Concepts to Code* , New York, NY : Springer New York, 2012 .