Estadística Bayesiana

Inferencia Bayesiana

Ms Carlos López de Castilla Vásquez

Universidad Nacional Agraria La Molina

2017-1

Introducción

- La *teoría de decisión* es un área de suma importancia en estadística ya que muchos problemas del mundo real pueden tomar la forma de un *problema de decisión*.
- El proceso de inferencia estadística puede ser entendido como un problema de decisión.
- Luego de observar un conjunto de datos: ¿qué valor debo utilizar para estimar el parámetro?
- Para algunos autores los temas como intervalos de confianza y pruebas de hipótesis carecen de importancia ya que en la estadística bayesiana la *distribución posterior es la inferencia*.

Elementos en la teoría de decisión

- Existen muchos enfoques en la teoría de la decisión pero quizás el más coherente sea el enfoque bayesiano ya que considera *toda la información disponible* para la toma de decisiones.
- Se tiene un espacio paramétrico Θ que describe todos los posibles *estados de la naturaleza*.
- Se considera un *conjunto de acciones* A sobre las cuales tomar la decisión.
- Se define una *función pérdida* $L(\theta, a)$ que determina la *pérdida* en la que se incurre al tomar la decisión a cuando el verdadero estado de la naturaleza es θ.

Pérdida esperada posterior

- Se requiere una distribución posterior $f(\theta|y)$ que contenga toda la información disponible sobre el parámetro.
- La idea es optar por la acción que minimice la *pérdida esperada posterior*.
- Se busca minimizar la siguiente expresión:

$$\rho(a, y) = \int_{\theta} L(\theta, a) f(\theta|y)\, d\theta$$

- El resultado anterior representa la pérdida esperada al tomar la acción a habiendo observado el valor y.

Regla de decisión y riesgo de Bayes

- Los diferentes valores de y determinarán distintos valores de a que minimicen la pérdida esperada posterior.
- El conjunto de estos valores, denotados por $d(y)$, constituyen la *regla de decisión de Bayes*.
- Para una regla de Bayes se puede calcular el riesgo asociado a este conjunto de acciones usando el *riesgo de Bayes*:

$$RB\left(d\right) = \int_{y} \rho\left(d\left(y\right), y\right) f\left(y\right) dy$$

- El riesgo de Bayes será menor al costo esperado de cualquier otra estrategia o decisión.

Ejemplo: Política de vacunación

- Una autoridad en salud pública desea desarrollar una política de vacunación contra una enfermedad la cual es causa de ausentismo en el trabajo.
- Estudios realizados sugieren que el 60 % de la población ya está inmunizada.
- Se ha desarrollado una prueba simple en la piel pero no es completamente confiable.
- La pérdida por no vacunar a un individuo vulnerable es 20 mientras que la pérdida por vacunar a una persona inmune es 8. No hay pérdida cuando se vacuna a una persona vulnerable o cuando no se vacuna a una persona inmune.

Ejemplo: Política de vacunación

- Las probabilidades de reacción a la prueba mencionada se presentan a continuación:

	Insignificante	Suave	Moderada	Fuerte
Inmune	0.35	0.30	0.21	0.14
Vulnerable	0.09	0.17	0.25	0.49

- Calcular las probabilidades posteriores.
- Hallar las pérdidas esperadas posteriores y la correspondiente regla de Bayes.
- Calcular el riesgo de Bayes.

Ejemplo: Densidad de un material

- Un material es clasificado por la compañía que lo produce según su densidad. Sea θ la verdadera densidad del material.
- Si $0 < \theta < 1$ el material es de tipo A, si $1 < \theta < 2$ el material es de tipo B y si $\theta > 2$ el material es de tipo C.
- El material es sometido a una prueba cuyo resultado, y, definirá como será clasificado.
- Suponga que:

$$f(y|\theta) = \exp\{-y + \theta\} \qquad y > \theta$$

y que $\theta \sim \mathcal{E}(1)$.

Ejemplo: Densidad de un material

- La pérdida obtenida por clasificar incorrectamente el material se da en la siguiente tabla:

Verdadero tipo	Clasificado como A	B	C
A	0	2	4
B	1	0	1
C	2	1	0

- Hallar la regla de Bayes si se observa que $y = 4$.
- Calcular el riesgo de Bayes asociado a la regla anterior.

Estimación puntual

- La distribución posterior ofrece un resumen completo del proceso de inferencia sobre el parámetro θ.
- Sin embargo, en algunos casos es necesario obtener un valor estimado del parámetro desconocido: ¿qué valor elegir como estimador de θ?
- El problema de estimación se convierte en un problema de decisión y para resolverlo se requiere especificar una función pérdida.
- La elección de una función pérdida en un problema específico dependerá del contexto.

Función pérdida cuadrática

Teorema

La función pérdida cuadrática se define por:

$$L(\theta, a) = (\theta - a)^2$$

El estimador de Bayes para θ bajo la función pérdida cuadrática es la media de la distribución posterior.

Función pérdida valor absoluto

Teorema

La función pérdida valor absoluto se define por:

$$L(\theta, a) = |\theta - a|$$

El estimador de Bayes para θ bajo la función pérdida valor absoluto es la mediana de la distribución posterior.

Función pérdida 0-1

Teorema

La función pérdida 0-1 se define por:

$$L(\theta, a) = \begin{cases} 0 & \text{si } |a - \theta| \leq \epsilon \\ 1 & \text{si } |a - \theta| > \epsilon \end{cases}$$

El estimador de Bayes para θ bajo la función pérdida 0-1 es la moda de la distribución posterior.

Ejemplo: Tiempo de vida de un componente

- Suponga que el tiempo de vida de un componente eléctrico (en miles de horas) tiene distribución exponencial con parámetro θ y que su distribución a priori $\theta \sim \mathcal{G}(45, 5)$.
- En una muestra de 100 componentes se encontró una media de 0.1125 miles de horas.
- Estime el tiempo medio de vida de los componentes usando las funciones pérdida cuadrática, valor absoluto y 0-1.

Mediana de la distribución gamma inversa

```
> library(pscl)
> qigamma(p=0.5, alpha=145, beta=16.25)
[1] 0.1123271
```

Intervalos de credibilidad

- Los intervalos de credibilidad pueden ser interpretados como los intervalos dentro de los cuales se encuentra el parámetro con cierta probabilidad.
- La interpretación anterior no es posible y representa un problema en los intervalos clásicos en los que el parámetro es considerado un valor constante (en este caso, la interpretación es frecuencial).
- Una dificultad con los intervalos de credibilidad bayesianos es que estos pueden ser definidos de varias maneras: los intervalos centrales y los intervalos de máxima densidad posterior.

Intervalos de credibilidad

- El intervalo de credibilidad central posterior del $100\,(1-\alpha)\,\%$ consiste de dos cuantiles debajo y por encima de los cuales se tiene exactamente $100\,(\alpha/2)\,\%$ de la probabilidad posterior.
- El intervalo de credibilidad de máxima densidad posterior (highest posterior density) es la región que contiene el $100\,(1-\alpha)\,\%$ de la probabilidad posterior pero que además tiene la característica de que la densidad dentro de la región nunca es menor a la de cualquier punto fuera de la misma.
- Los intervalos anteriores son idénticos si la distribución posterior es unimodal y simétrica.

Ejemplo: Número de incendios semanales

- Suponga que el número de incendios semanales ocurridos en una ciudad tiene distribución de Poisson con parámetro θ.
- Se decide utilizar una distribución a priori conjugada:

$$\theta \sim \mathcal{G}\left(\alpha = 2, \beta = 1\right)$$

- Suponga que el número de incendios observados en una muestra de 10 semanas fue 4.
- Hallar el intervalo de credibilidad central y de máxima densidad posterior del 95 % para θ.

Ejemplo: Número de incendios semanales

Intervalo central posterior

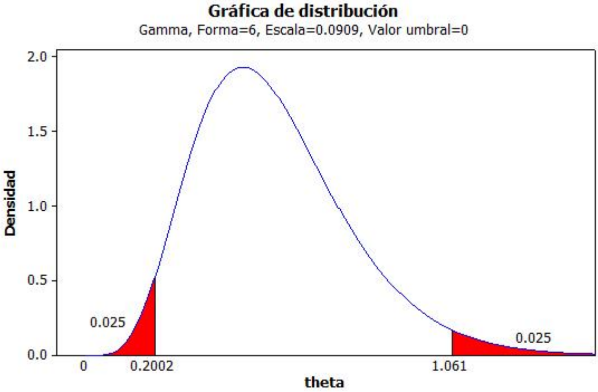
```
> qgamma(p=0.025, shape=6, rate=11)
[1] 0.2001722
> qgamma(p=0.975, shape=6, rate=11)
[1] 1.060757
```

Intervalo de máxima densidad posterior

```
> library(TeachingDemos)
> hpd(qgamma, shape=6, rate=11, conf=0.95)
[1] 0.1598254 0.9876772
```

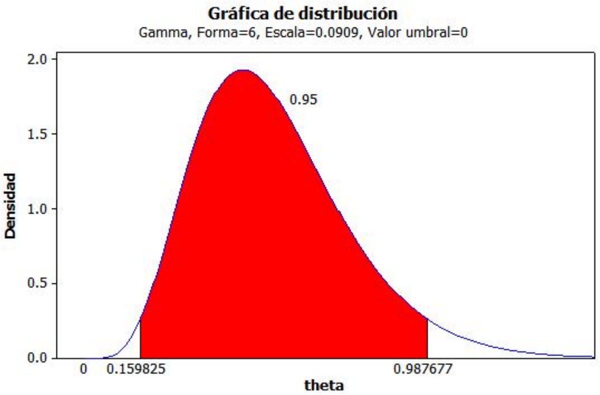
Ejemplo: Número de incendios semanales

Figura 1: Intervalo de credibilidad central posterior



Ejemplo: Número de incendios semanales

Figura 2: Intervalo de credibilidad de máxima densidad posterior



Pruebas de hipótesis

- Una prueba de hipótesis consiste en tomar una decisión entre dos alternativas: $H_0 : \theta \in \Omega_0$ y $H_1 : \theta \in \Omega_1$.
- Las pruebas de hipótesis clásicas son evaluadas en términos de las probabilidades de cometer error tipo I y tipo II.
- En el análisis bayesiano se calculan las probabilidades a priori y a posteriori para H_0 y H_1:

	A favor de H_0	A favor de H_1
A priori	$\pi_0 = \Pr(\theta \in \Omega_0 \mid f(\theta))$	$\pi_1 = \Pr(\theta \in \Omega_1 \mid f(\theta))$
Posterior	$\alpha_0 = \Pr(\theta \in \Omega_0 \mid f(\theta \mid y))$	$\alpha_1 = \Pr(\theta \in \Omega_1 \mid f(\theta \mid y))$

Pruebas de hipótesis

- La razón π_1/π_0 es llamada *odds a priori* y la razón α_1/α_0 es llamada *odds posterior*.
- El *factor de Bayes* a favor de H_1 es:

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

  y nos informa de los cambios en nuestras creencias luego de observar los datos.
- En el caso más simple las hipótesis son $H_0 : \theta = \theta_0$ y $H_1 : \theta = \theta_1$.

Ejemplo: Tiempo de espera por un autobús

- Se sabe que el tiempo de espera (en minutos) por un autobús en un paradero a cierta hora del día tiene una distribución uniforme en el intervalo $(0, \theta)$.
- Por información de otras rutas similares se considera que a priori:
$$\theta \sim \mathcal{PA}(\alpha = 5, \beta = 8)$$
- Suponga que en este paradero son observados al azar los siguientes tiempos de espera: 9, 3, 2, 5 y 4 minutos.
- ¿Se puede afirmar que el tiempo máximo de espera por un autobús en el paradero es de como mínimo 10 minutos?

Ejemplo: Tiempo de espera por un autobús

Odds a priori

```
> library(VGAM)
> pi1 <- ppareto(q=10, scale=5, shape=8)
> odds.priori <- pi1/(1 - pi1) [1] 255
```

Odds a posteriori

```
> alpha1 <- ppareto(q=10, scale=9, shape=13)
> odds.posteriori <- alpha1/(1 - alpha1) [1] 2.934118
```

Factor de Bayes

```
> B <- odds.posteriori/odds.priori [1] 0.01150634
```

Introducción

- Con una muestra grande los valores de los parámetros en la distribución a priori suelen ser poco importantes.
- Si $y|\lambda \sim \mathcal{P}(\lambda)$ y a priori $\lambda \sim \mathcal{G}(\alpha, \beta)$ entonces:

$$\mathsf{E}[\lambda|\mathbf{y}] = \frac{\alpha + n\bar{y}}{\beta + n} \longrightarrow \bar{y}$$

  cuando $n \to \infty$.
- Es común que los estimadores bayesianos se aproximen a los estimadores de *máxima verosimilitud* en muestras grandes.
- Si la distribución posterior es *unimodal* y *aproximadamente simétrica* puede ser conveniente aproximarla por una distribución normal.

Serie de Taylor

- Usando expansión en serie de Taylor para el logaritmo de la distribución posterior centrada en $\tilde{\theta}$:

$$\begin{aligned}\log f(\theta|\mathbf{y}) &= \log f(\tilde{\theta}|\mathbf{y}) + \frac{1}{2}(\theta-\tilde{\theta})^2\left[\frac{d^2}{d\theta^2}\log f(\theta|\mathbf{y})\right]\Bigg|_{\theta=\tilde{\theta}} + \\ &\quad \frac{1}{6}(\theta-\tilde{\theta})^3\left[\frac{d^3}{d\theta^3}\log f(\theta|\mathbf{y})\right]\Bigg|_{\theta=\tilde{\theta}} + \cdots\end{aligned}$$

- El término lineal en la expansión es cero y los términos de orden superior pierden importancia con relación al término cuadrático cuando el tamaño de muestra es grande.

Primera aproximación

Teorema

La primera aproximación a la distribución normal es:

$$f(\theta|\mathbf{y}) \approx \mathcal{N}\left(\mathrm{E}(\theta|\mathbf{y}), \mathrm{Var}(\theta|\mathbf{y})\right)$$

suponiendo que la media y variancia de la distribución posterior existen.

Segunda aproximación

Teorema

La segunda aproximación a la distribución normal es:

$$f(\theta|\mathbf{y}) \approx \mathcal{N}\left(\tilde{\theta}, I_1^{-1}(\tilde{\theta})\right)$$

donde $\tilde{\theta}$ es la moda de la distribución posterior y:

$$I_1(\theta) = -\frac{d^2}{d\theta^2}\log f(\theta|\mathbf{y})$$

es la información observada.

Tercera aproximación

Teorema

La tercera aproximación a la distribución normal es:

$$f\left(\theta|\mathbf{y}\right) \approx \mathcal{N}\left(\hat{\theta}, I_2^{-1}(\hat{\theta})\right)$$

donde $\hat{\theta}$ es el estimador de máxima verosimilitud de θ y:

$$I_2\left(\theta\right) = -\frac{d^2}{d\theta^2}\log f\left(\mathbf{y}|\theta\right)$$

Cuarta aproximación

Teorema

La cuarta aproximación a la distribución normal es:

$$f\left(\theta|\mathbf{y}\right) \approx \mathcal{N}\left(\hat{\theta}, I_3^{-1}(\hat{\theta})\right)$$

donde $\hat{\theta}$ es el estimador de máxima verosimilitud de θ y:

$$I_3\left(\theta\right) = -n\mathrm{E}\left(\frac{d^2}{d\theta^2}\log f\left(Y|\theta\right)\Big|\theta\right)$$

Ejemplos

- Se toma una muestra de tamaño n desde la distribución $\mathcal{B}(\theta)$. Suponga que a priori $\theta \sim \mathcal{BE}(\alpha, \beta)$. Hallar todas las aproximaciones de la distribución posterior a la distribución normal asumiendo que la muestra es grande.
- Se toma una muestra de tamaño n desde la distribución $\mathcal{P}(\lambda)$. Suponga que a priori $\lambda \sim \mathcal{G}(\alpha, \beta)$. Hallar todas las aproximaciones de la distribución posterior a la distribución normal asumiendo que la muestra es grande.
- Se toma una muestra de tamaño n desde la distribución $\mathcal{E}\left(\lambda^{-1}\right)$. Suponga que a priori $\lambda \sim \mathcal{GI}(\alpha, \beta)$. Hallar todas las aproximaciones de la distribución posterior a la distribución normal asumiendo que la muestra es grande.