Estadística Bayesiana

Modelos uniparamétricos

Ms Carlos López de Castilla Vásquez

Universidad Nacional Agraria La Molina

2017-1

Distribución binomial

- Se tiene una secuencia de ensayos independientes de Bernoulli $y_1, \cdots, y_n$. Sea y el número total de éxitos en los n ensayos.
- El modelo binomial queda definido por:

$$f\left(y|\theta\right) = \binom{n}{y}\theta^y\left(1-\theta\right)^{n-y}$$

- Si la distribución a priori es $\theta \sim \mathcal{U}\left(0,1\right)$ entonces:

$$\theta|y \sim \mathcal{BE}\left(y+1, n-y+1\right)$$

- La distribución predictiva posterior para un nuevo ensayo es:

$$\tilde{y}|y \sim \mathcal{BB}\left(1, y+1, n-y+1\right)$$

Distribuciones a priori no informativas

- ¿Cómo se puede justificar la elección de una distribución a priori uniforme para θ?
- La distribución a priori debe incluir todos los valores posibles de θ, pero no tiene que estar necesariamente concentrada en torno al verdadero valor ya que la información en los datos modificará y dominará cualquier especificación probabilística inicial.
- El razonamiento de Laplace para la densidad a priori uniforme es llamado *principio de la razón insuficiente*, el cual indica que si nada es conocido acerca de θ, entonces no hay ninguna razón para asignar probabilidades diferentes a algunos de sus valores.

Relaciones entre la media a priori y posterior

- Es natural esperar que existan algunas relaciones entre la distribución a priori y la distribución posterior.
- Se podría esperar que la distribución posterior, que incorpora la información de los datos, sea menos variable que la distribución a priori.
- La media posterior se encuentra entre la media a priori y el EMV de θ si:

$$\mathrm{E}\left(\theta|y\right) = w\mathrm{E}\left(\theta\right) + \left(1 - w\right)\hat{\theta} \qquad 0 < w < 1$$

- En el ejemplo binomial anterior la media posterior se encuentra entre la media a priori y la proporción muestral de éxitos.

Familias conjugadas

- La función de verosimilitud binomial es de la forma:

$$f\left(y|\theta\right) \quad \propto \quad \theta^{y}\left(1-\theta\right)^{n-y}$$

- Si la distribución a priori es $\theta \sim \mathcal{BE}\left(\alpha,\beta\right)$ entonces la distribución posterior es:

$$\theta|y \sim \mathcal{BE}\left(y+\alpha, n-y+\beta\right)$$

- Se dice que la distribución beta es una *familia conjugada* con la verosimilitud binomial.
- La distribución predictiva posterior para m nuevos ensayos es:

$$\tilde{y}|y \sim \mathcal{BB}\left(m, y+\alpha, n-y+\beta\right)$$

Ejemplo: Tratamiento para cáncer

- Un médico sugiere un nuevo tratamiento para una forma de cáncer. Sea θ la probabilidad de que un paciente con el nuevo tratamiento sobreviva más de seis meses. Se sabe que $\mathrm{E}(\theta) = 0.30$ y $\mathrm{Var}(\theta) = 0.019$.
- Suponga que el médico decide representar sus creencias usando la distribución beta ¿cuáles serían los parámetros?
- Una muestra de 15 pacientes recibio el tratamiento y 9 de ellos sobrevivieron más de seis meses. Hallar la distribución posterior para θ.
- ¿Cuál es la probabilidad que en una nueva muestra de 5 pacientes solo 2 de ellos sobrevivan más de seis meses?

Distribución normal con varianza conocida

- Se tiene una muestra $\mathbf{y} = (y_1, \cdots, y_n)$ obtenida a partir de la distribución normal:

$$f(y|\theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(y-\theta)^2\right\}$$

- La función de verosimilitud es:

$$\begin{aligned} L(\theta|\mathbf{y}) &\propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\theta)^2\right\} \\ &\propto \exp\left\{-\frac{n}{2\sigma^2}\left(\theta^2 - 2\theta\overline{y}\right)\right\} \end{aligned}$$

Distribución normal con varianza conocida

- La distribución a priori conjugada es:

$$\theta \sim \mathcal{N}\left(\mu_0, \tau_0^2\right)$$

- La distribución posterior es:

$$\theta|\mathbf{y} \sim \mathcal{N}\left(\mu_n = \frac{\frac{1}{\tau_0^2}\mu_0 + \frac{n}{\sigma^2}\overline{y}}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}}, \tau_n^2 = \frac{1}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}}\right)$$

- La distribución predictiva posterior para $\tilde{y}$ es:

$$\tilde{y}|\mathbf{y} \sim \mathcal{N}\left(\mu_n, \tau_n^2 + \sigma^2\right)$$

Ejemplo: Presión sanguínea

- Un pequeño estudio piloto evaluó la efectividad de una droga experimental en el control de la presión sanguínea de pacientes con moderada hipertensión.
- El cambio en la presión sanguínea de los pacientes (en mmHg) después de tomar la droga tiene una distribución normal con media desconocida y variancia 25.
- El experimento fue llevado a cabo con una muestra de 8 pacientes obteniéndose los siguientes resultados (valores negativos significan que la droga disminuyó la presión sanguínea, valores positivos que la incrementó): 3.7, −6.7, 10.5, −6.1, −17.6, 2.3, −7.9 y −8.9.

Ejemplo: Presión sanguínea

- Calcule la distribución posterior para la media a partir de una distribución a priori:

$$\theta \sim \mathcal{N}\left(0, 100\right)$$

- La droga será aceptable si la probabilidad de que le disminuya la presión sanguínea a un paciente es mayor al 90 %. ¿Cuál es el valor de esta probabilidad usando la distribucion posterior?
- Hallar la distribución predictiva posterior para el cambio en la presión sanguínea de un nuevo paciente.

Distribución normal con media conocida

- La función de verosimilitud para $\mathbf{y} = (y_1, \cdots, y_n)$ es:

$$L\left(\sigma^2|\mathbf{y}\right) \quad \propto \quad \left(\sigma^2\right)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\theta)^2\right\}$$

- La distribución a priori conjugada es:

$$\sigma^2 \sim \chi^2 \text{ inversa de escala } \left(\upsilon_0, \sigma_0^2\right)$$

- La distribución posterior es:

$$\sigma^2|\mathbf{y} \sim \chi^2 \text{ inversa de escala } \left(\upsilon_0 + n, \frac{\upsilon_0\sigma_0^2 + \sum_{i=1}^{n}(y_i-\theta)^2}{\upsilon_0 + n}\right)$$

Ejemplo: Máquina empaquetadora

- Una máquina empaqueta un determinado producto de tal manera que el peso de los paquetes se distribuye normalmente con una media igual a 500 gramos.
- Ultimamente se ha recibido un número importante de quejas relacionadas al peso de los paquetes por lo que se considera la posibilidad que el proceso no se encuentre funcionando de forma correcta.
- El proceso de control de calidad toma en cuenta la variabilidad que es un indicador de confiabilidad del producto.
- Se considera que el proceso se encuentra bajo control cuando la varianza no supera los 0.05 gramos^2.

Ejemplo: Máquina empaquetadora

- Se elige al azar una muestra de 14 paquetes cuyos pesos (en gramos) se muestran a continuación:

499.75	500.42	500.20	500.16	500.26	500.18	500.06
500.10	500.28	499.85	500.27	499.75	499.81	499.95

- Suponga que la distribución a priori es:

$$\sigma^2 \sim \chi^2 \text{ inversa de escala } \left(\upsilon_0 = 4, \sigma_0^2 = 0{,}025\right)$$

- Hallar la distribución posterior y luego calcular la probabilidad que el proceso se encuentre fuera de control.

Ejemplo: Máquina empaquetadora

Usando números aleatorios

```
> vn <- 18 ; sigma2n <- 0.04405556; library(geoR); set.seed(200)
> sigma2 <- rinvchisq(n=10000,df=vn,scale=sigma2n)
> prob <- length(sigma2[sigma2>0.05])/10000
> prob
[1] 0.3987
```

Usando la función de densidad

```
> prob <- 1-integrate(f=dinvchisq,lower=0,upper=0.05,df=vn,
scale=sigma2n)$value
> prob
[1] 0.3976827
```

Distribución de Poisson

- Se tiene una muestra $\mathbf{y} = (y_1, \cdots, y_n)$ obtenida de la distribución de Poisson con parámetro θ:

$$f(y|\theta) = \frac{\theta^y \exp\{-\theta\}}{y!} \qquad y = 0, 1, \cdots$$

- La distribución a priori conjugada es :

$$\theta \sim \mathcal{G}(\alpha, \beta)$$

- La distribución posterior es:

$$\theta|\mathbf{y} \sim \mathcal{G}(\alpha + n\bar{y}, \beta + n)$$

Ejemplo: Incendios forestales

- 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:

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

- Hallar la distribución posterior si el número total de incendios observados en una muestra de 10 semanas fue 4.
- Calcular la probabilidad que para una semana futura el número de incendios ocurridos en la ciudad sea menor que dos.

Distribución exponencial

- La distribución exponencial es usada para modelar tiempos de espera. Su función de densidad es:

$$f\left(y|\theta\right) = \theta \exp\left\{-y\theta\right\} \qquad y > 0$$

  y el parámetro θ es llamado la *tasa*.
- La distribución a priori conjugada es:

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

- La correspondiente distribución posterior es:

$$\theta|\mathbf{y} \sim \mathcal{G}\left(\alpha + n, \beta + n\overline{y}\right)$$

Ejemplo: Tiempo de vida

- Suponga que el tiempo de vida de un componente eléctrico tiene distribución exponencial con parámetro θ y que su distribución a priori es $\theta \sim \mathcal{G}(\alpha = 45, \beta = 5)$.
- En una muestra de 10 componentes se encontraron las siguientes medidas de resumen (en miles de horas):

0.052	0.105	0.234	0.031	0.061
0.243	0.118	0.173	0.042	0.007

- Hallar la distribución posterior para θ y luego la probabilidad que el tiempo medio de vida sea mayor de 100 horas.

Distribuciones no informativas

- Muchas veces se desea contar con distribuciones a priori que puedan garantizar una mínima influencia en la distribución posterior.
- Tales distribuciones son algunas veces llamadas *distribuciones a priori de referencia* y la densidad a priori es descrita como *vaga*, *plana*, *difusa* o *no informativa*.
- La razón para utilizar distribuciones a priori no informativas es dejar que *los datos hablen por sí mismos* de modo que el proceso de inferencia no se encuentre afectado por información externa a ellos.

Distribuciones a priori propias e impropias

- Sea $y|\theta \sim \mathcal{N}\left(\theta, \sigma^2\right)$ con σ^2 conocida. Si la distribución a priori es $f\left(\theta\right) \propto 1$ entonces:

$$\theta|\mathbf{y} \sim \mathcal{N}\left(\overline{y}, \frac{\sigma^2}{n}\right)$$

- La distribución a priori no es técnicamente posible ya que su integral es diferente de uno.
- Se llamará a una densidad a priori *propia* si no depende de los datos y su integral es igual a uno.
- A pesar de que la distribución a priori del ejemplo es impropia la distribución posterior resultante es propia.

Principio de invariancia de Jeffreys

- Una aproximación usada para definir distribuciones a priori no informativas fue desarrollada por Jeffreys, quien se basó en transformaciones uno a uno del parámetro $\phi = h(\theta)$.
- Por transformación de variables la densidad a priori es equivalente a la siguiente densidad a priori para ϕ:

$$f(\phi) = f(\theta)\left|\frac{d\theta}{d\phi}\right|$$

- El principio de Jeffreys considera que cualquier regla para determinar $f(\theta)$ debe conducir a un resultado equivalente si se aplica al parámetro transformado.

Principio de invariancia de Jeffreys

- La distribución a priori no informativa de Jeffreys es:

$$f(\theta) \propto [J(\theta)]^{1/2}$$

donde $J(\theta)$ es la *información de Fisher*:

$$J(\theta) = \mathrm{E}\left[\left(\frac{d \log f(y|\theta)}{d\theta}\right)^2 \middle| \theta\right] = -\mathrm{E}\left[\left(\frac{d^2 \log f(y|\theta)}{d\theta^2}\right) \middle| \theta\right]$$

- Los resultados anteriores pueden ser extendidos a modelos multiparamétricos pero los resultados son controvertidos.

Ejemplos

- Sea la distribución fundamental:

$$y|\theta \sim \mathcal{BI}\left(n, \theta\right)$$

Hallar la distribución a priori no informativa de Jeffreys para θ.

- Sea la distribución fundamental:

$$y|\theta, \sigma^2 \sim \mathcal{N}\left(\theta, \sigma^2\right)$$

con σ^2 conocida. Hallar la distribución a priori no informativa de Jeffreys para θ.

- Hallar la distribución a priori no informativa de Jeffreys para σ^2 considerando que θ es conocida.

Mixtura de distribuciones a priori

- Suponga que $f_1(\theta), \cdots, f_k(\theta)$ son distribuciones a priori conjugadas para θ que conducen a las distribuciones posteriores:
$$f_1(\theta|y), \cdots, f_k(\theta|y)$$
respectivamente.
- Considere la siguiente mixtura de distribuciones:
$$f(\theta) = \sum_{i=1}^{k} w_i f_i(\theta)$$
donde w_i es el peso o ponderación de la distribución a priori $f_i(\theta)$ en $f(\theta)$.

Mixtura de distribuciones a priori

- La distribución posterior es:

$$\begin{aligned} f(\theta|y) &\propto f(y|\theta) \sum_{i=1}^{k} w_i f_i(\theta) \\ &\propto \sum_{i=1}^{k} w_i f(y|\theta) f_i(\theta) = \sum_{i=1}^{k} w_i^* f_i(\theta|y) \end{aligned}$$

- First Bayes permite usar mixturas a priori para los cuatro modelos conjugados mencionados en esta presentación.

Ejemplo: Moneda

- Suponga que una moneda se hace girar sobre una mesa y sea θ la probabilidad de que aparezca cara.
- Dadas las posibles imperfecciones en el borde de la moneda, se podría pensar que θ es aproximadamente 0.3 o 0.7.
- Como la distribución beta es una familia conjugada para la verosimilitud binomial podría usarse la distribución a priori:

$$f(\theta) = \frac{1}{2}\mathcal{BE}(6, 14) + \frac{1}{2}\mathcal{BE}(14, 6)$$

- Hallar la distribución posterior asumiendo que en diez lanzamientos de la moneda se observaron 4 caras y 6 sellos.