Estadística bayesiana

Maestría en Modelado y Simulación/ Maestría en Ingeniería y analítica de datos

Carenne Ludeña

Departamento de Ciencias Básicas y Modelado
Facultad de Ciencias Naturales e Ingeniería
Universidad Jorge Tadeo Lozano

Probabilidad condicional

- Si $P(A) \neq 0$: $P(B|A) = P(A \cap B)/P(A)$
- $$P(A \cap B) = P(A) \times P(B|A)$$
- Variables discretas: $P(X|Y) = P(X, Y)/P(Y)$
- Variables continuas: $f(x|y) = \frac{f(x,y)}{f(y)}$
- Mezcla: Y discreta y X continua $X|Y = y \sim f_y(x)$

Probabilidad total

- Si $\{A_i\}_i$ es una partición de Ω
- $P(A) = \sum_i P(A \cap A_i) = \sum_i P(A|A_i)P(A_i)$
- Bayes: $P(A_i|A) = P(A|A_i)P(A_i)/P(A)$

Esperanza condicional

- Caso discreto: $E_{X|Y=y} = \sum_i x_i p(X = x_i|Y = y)$
- Caso continuo o de mezcla: $E_{X|Y=y} = \int xf(x|Y = y)$
- $E_{X|Y=y}$ es una variable aleatoria: depende del valor de Y
- $E(X) = E(E(X|Y))$
- $Var(X) = E(Var(X|Y)) + Var(E(X|Y))$

Estadística bayesiana

- Principios: Thomas Bayes (1701–1761)
- Una probabilidad es un grado de creencia. No depende de llevar a cabo un experimento
- Los parámetros que definen una distribución son a su vez variables aleatorias
- la distribución de los parámetros se actualiza a medida que se observa el fenómeno
- Modelo $X|\theta \sim f(x|\theta)$
- Densidad a priori $\theta \sim f(\theta)$
- Marginal $f(x) = \int f(x|\theta)f(\theta)d\theta$
- Densidad a posteriori $\theta|X \sim \frac{f(x|\theta)f(\theta)}{f(x)}$
- Densidad de predicción $y|x \sim f(y|x) = \int f(y|x,\theta)f(\theta|x)$

Verosimilitudes

- Si $X = (X_1, \ldots, X_n)$ muestra independiente
- $L_n(x) = \prod_i f(x_i|\theta)$
- $f(\theta|x) = \frac{f(\theta)\prod_i f(x_i|\theta)}{f(x)}$
- $f(\theta|x) \propto f(\theta)L_n(x)$
- La constante de normalización en general no la usaremos: se buscan maneras de no necesitar la constante de normalización.
- Contar con una distribución a posteriori de θ permite cambiar el paradigma de estimación puntual por el conocimiento de la distribución
- Valor medio a posteriori: $\overline{\theta} = \int \theta f(\theta|x)d\theta$
- MAP: θ que maximiza $f(\theta|x)f(\theta)$
- Intervalo de confianza: a, b tales que $\int_a^b f(\theta|x) = 1 - \alpha$
- Se puede estimar el intervalo de confianza usando los cuantiles estimados usando la distribución

Ejemplo Bernoulli

- $p \sim Unif(0,1)$, $X|p \sim Ber(p)$
- $X_1, \ldots, X_n$, $S = \sum_i X_i$
- $f(p|X) \sim Beta(S+1, n-S+1)$
- $Z \sim Beta(\alpha, \beta)$: $f(z) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} z^{\alpha-1} z^{\beta-1}$ y $E(Z) = \frac{\alpha}{\alpha+\beta}$. La uniforme es un caso especial de una Beta.
- $\overline{p} = \frac{S+1}{n+2}$
- $\overline{p} = \lambda_n \hat{p} + (1-\lambda)\tilde{p}$
- Con $\hat{p} = S/n$, $\tilde{p} = 1/2 = E(p)$, $\lambda_n = n/(n+2)$
- Si las distribuciones previas y posteriores pertenecen a la misma familia se dice que la previa es conjugada con respecto al modelo.

Ejemplo Bernoulli

Graficar betas

```
 library(stats)
 par(mfrow=c(3,2))
 x=seq(0,1,by=o.1)
 alpha=c(0,2,10,20,50,500)
 beta=c(0,2,8,11,27,232)
 for(i in 1:length(alpha)){
y<-dbeta(x,shape1=alpha[i],shape2=beta[i])
plot(x,y,type="l")
}
```

Ejemplo Gaussiano

- $\theta \sim N(a, b^2)$, $X|p \sim N(\theta, \sigma^2)$
- $X_1, \ldots, X_n$, $\overline{X} = \sum_i X_i / n$
- $f(\theta|X) \sim N(\overline{\theta}, \tau^2)$
- $\overline{\theta} = w\overline{X} + (1 - w)a$
- $w = \frac{\frac{n}{\sigma^2}}{\frac{n}{\sigma^2} + \frac{1}{b^2}}$
- $\frac{1}{\tau^2} = \frac{n}{\sigma^2} + \frac{1}{b^2}$
- $w \to 1$ cuando $n \to \infty$, pero si $b \to \infty$ produce el mismo efecto (o muchos datos o una previa muy poco informativa)
- El intervalo de confianza se calcula con los parámetros de la normal: $[\overline{\theta} - \tau * q_{1-\alpha/2}, \overline{\theta} + \tau * q_{1-\alpha/2}]$

propiedades de muestras grandes

- Sea $\hat{\theta}$ el estimador máximo verosímil (maximiza la verosimilitud)
- Sea $s = \sqrt{(Var(\hat{\theta}))}$
- Entonces $C_n = [\hat{\theta} - q_{1-\alpha/2}s, \hat{\theta} - q_{1-\alpha/2}s]$ es el intervalo de confianza clásico
- Se tiene que $P(\theta \in C_n|X) \to 1 - \alpha$

Previas no informativas e impropias

- Previas no informativas: que no dicen mucho
- Ejemplo Bernoulli $p \sim Unif(0, 1)$
- Ejemplo Gaussiano $b \to \infty$
- Se usan cuando no sabemos mucho
- Previas impropias: no integran $\int f(\theta) d\theta = \infty$

Pruebas de hipótesis Bayesianas

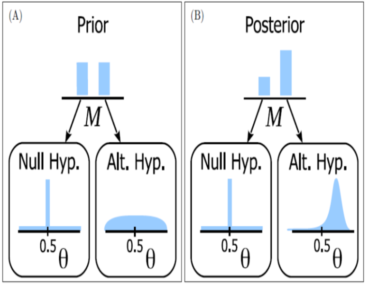
- Se necesita una previa para θ bajo cada hipótesis (nula y alternativa)
- Se incluye también una previa sobre las hipótesis:por ejemplo $P(H_0) = P(H_1) = 1/2$
- Ejemplo: $H_0 : \theta = \theta_0$, $H_1 : \theta \neq \theta_0$
- Bajo H_1: $f(\theta)$ es la previa para θ
- $P(H_0|X) = \frac{f(x|H_0)P(H_0)}{f(x|H_0)P(H_0)+f(x|H_1)P(H_1)} = \frac{f(x|\theta_0)}{f(x|\theta_0)+\int f(x|\theta)f(\theta)d\theta}$
- Esto produce resultados diferentes a los usuales!

Pruebas de hipótesis Bayesianas

Factor de bayes (BF)

$$BF = \frac{P(H_0|X)}{P(H_1|X)} / \frac{P(H_0)}{P(H_1)}$$

Para rechazar H_0, BF debe ser pequeña ($< 0.05, 0.1$)



Estimadores de Bayes

- Funciones de riesgo $R(\theta,\hat{\theta}) = E(L(\theta,\hat{\theta})) = \int L(\theta,\hat{\theta(x)})f(x|\theta)dx$, con L una función de pérdida
- Visión clásica: minimizar una cota superior de $R(\theta,\hat{\theta})$. $\tilde{\theta} = \min \max_{\theta} R(\theta,\hat{\theta})$
- Enfoque bayesiano: minimizar la esperanza con respecto a la distribución de θ
- $R_B(f,\hat{\theta}) = \int R(\theta,\hat{\theta})f(\theta)d\theta$
- Este estimador se llama el estimador de Bayes