Auxiliar #11

Estadística Bayesiana I

Parte 1: Análisis de casos

Comente si las siguientes aseveraciones son verdaderas o falsas y justifique en cada caso.

1. El parámetro θ_{MAP} obtenido de maximizar la posterior es equivalente a al parámetro θ_{MLE} que resulta de maximizar la verosimilitud siempre y cuando el prior represente una distribución uniforme.

Solución

Recordemos que en la inferencia Bayesiana, la relación entre la posterior $p(\theta|y)$ y la verosimilitud $p(y|\theta)$:

$$p(\theta|y) \propto p(y|\theta)p(\theta)$$

Donde $p(\theta)$ es el prior. Y si este sigue una distribución $Uniforme([a, b])$ con a y b números reales, sabemos que:

$$p(\theta) = \begin{cases} \frac{1}{b-a} & \text{para } a \leq \theta \leq b \\ 0 & \text{para } \theta < a \text{ o } \theta > b \end{cases}$$

Entonces, asumiendo que $\theta \in [a, b]$, la aproximación del posterior queda como:

$$p(\theta|y) \propto p(y|\theta)\frac{1}{b-a}$$

Y como el término $\frac{1}{b-a}$ es una constante, es decir, no depende de θ, la posterior queda proporcional a la verosimilitud:

$$p(\theta|y) \propto p(y|\theta)$$

Por lo tanto, la maximización de la posterior en función de θ es equivalente a la maximización de la verosimilitud en función del mismo parámetro. De esta manera, $\theta_{MAP} = \theta_{MAP}$ cuando el prior es $Uniforme([a, b])$.

2. La distribución Bernoulli no pertenece a la familia exponencial puesto que su función de densidad de probabilidad $P(X = x) = p^x(1-p)^{1-x}$ con $x \in \{0, 1\}$ no tiene un término exponencial.

Solución

Esta afirmación es falsa. Porque la distribución Bernoulli pertenece a la familia exponencial pese a no contar con una exponencial en su fórmula de densidad de probabilidad más habitual. Ya que, de hecho, esta expresión puede reescribirse de manera que puedan ser identificables los elementos y funciones características de la familia exponencial. Haremos eso a continuación:

$$\begin{aligned} P(x) &= p^x(1-p)^{1-x} \\ &= \exp\left[\log\left(p^x(1-p)^{1-x}\right)\right] \\ &= \exp[x \log p + (1-x)\log(1-p)] \\ &= \exp\left[x \log \frac{p}{1-p} + \log(1-p)\right] \end{aligned}$$

Definimos $\eta = log(\frac{p}{1-p})$ y queda:

$$P(x) = \exp\left[x\eta - \log\left(1 + e^{\eta}\right)\right]$$

Donde:

- $T(x) = x$
- $\eta = log(\frac{p}{1-p})$
- $A(p) = \log\left(1 + e^{\eta(p)}\right)$

3. Para realizar predicciones sobre los datos, es decir, calcular la probabilidad de que x ocurra teniendo en cuenta la función de densidad $P(X = x|\theta)$, basta con obtener θ_{MAP}, reemplazarlo en la función y así, finalmente, tener una probabilidad predicha para x. Por lo tanto, no tiene sentido el uso de la predictive posterior distribution.

Solución

Ocupar el θ_{MAP} de la manera propuesta en el enunciado es un método válido. Sin embargo, con esto se está perdiendo la incertidumbre que existe sobre θ al tratarlo como un valor fijo, ya que según la estadística Bayesiana, el parámetro es una variable aleatoria y el uso de la predictive posterior distribution permite integrar la dicha variabilidad.

Además, bajo la lógica de que θ es una variable aleatoria, θ_{MAP} podría no ser un buen indicador en casos donde la distribución es relativamente uniforme.

Entonces, el comentario es falso. Ya que el uso de la predictive posterior distribution está justificado por un principio de la estadística de Bayesiana de que θ es una variable aleatoria.

Parte 2: Desarrollo

1. Considere datos que siguen una distribución Exponencial de parámetro θ, o sea, la función de densidad de probabilidad de estos datos es $p(y_i|\theta) = \theta e^{-y_i\theta}$, para todo $y_i > 0$. Además, considere que θ sigue una distribución $Gamma(\theta|\alpha, \beta)$, es decir, la función de densidad de probabilidad de θ es $p(\theta|\alpha, \beta) = \frac{\theta^{\alpha-1}e^{-\beta\theta}\beta^{\alpha}}{\Gamma(\alpha)}$, para todo $\theta > 0, \alpha > 0, \beta > 0$.

   (a) Determine la distribución del la posterior $p(\theta|y)$ y responda si el prior es conjugado de la distribución Exponencial. Justifique.

   Solución

   En la inferencia Bayesiana, la relación entre la posterior $p(\theta|y)$, la verosimilitud $p(y|\theta)$ y el prior $p(\theta)$ es:

$$p(\theta|y) \propto p(y|\theta)p(\theta)$$

   Donde la verosimilitud, por definición es:

$$\begin{aligned} p(y|\theta) &= p(y_1, y_2, .., y_n|\theta) \\ &= \prod_{i=1}^{n} p(y_i|\theta) \end{aligned}$$

   Y sabemos que la verosimilitud representa una distribución Exponencial.

$$\begin{aligned} p(y|\theta) &= p(y_1, y_2, .., y_n|\theta) \\ &= \prod_{i=1}^{n} \theta e^{-y_i\theta} \\ &= \theta^n e^{-\theta\sum_{i=1}^{n} y_i} \\ &= \theta^n e^{-\theta n\hat{y}} \end{aligned}$$

   Por otro lado, se dice que el prior es $Gamma(\theta|\alpha, \beta)$. Por lo tanto:

$$\begin{aligned} p(\theta|\alpha, \beta) &= \frac{\theta^{\alpha-1}e^{-\beta\theta}\beta^{\alpha}}{\Gamma(\alpha)} \\ &\propto \theta^{\alpha-1}e^{-\beta\theta} \end{aligned}$$

   Entonces, la aproximación de la posterior queda como:

$$
\begin{aligned}
p(\theta|y) &\propto \theta^n e^{-\theta n\hat{y}} \theta^{\alpha-1} e^{-\beta\theta} \\
&= \theta^{\alpha+n-1} e^{-\theta(\beta+n\hat{y})}
\end{aligned}
$$

De esta manera la posterior representa una distribución $Gamma(\alpha+n-1, \beta+n\hat{y})$ Luego, notar que el prior tiene la misma clase de distribución que el posterior. Por lo tanto, se puede decir que el prior *Gamma* es conjugado de la distribución *Exponencial*, dada por la verosimilitud.

(b) La distribución Exponencial es comúnmente utilizada para modelar tiempos de espera. Por lo tanto, en lo que sigue, se buscará concluir sobre el parámetro θ de esta distribución para representar la probabilidad de tiempos de espera en un paradero de un bus interurbano. Por lo tanto, se tiene una muestra de $n = 100$ tiempos de espera en minutos registrados por pasajeros del bus, quienes en promedio esperaron $\hat{y} = 5$ minutos para abordar.

Teniendo esto en cuenta y asumiendo que $\alpha = 3$ y $\beta = 1$, calcule el MAP y el HDCI al 80% para θ según la posterior.

Solución

Se calcula computacionalmente.

(c) Exprese la probabilidad de que una persona tenga que esperar 10 minutos el bus en términos de la predictive posterior distribution.

Solución

$$
\begin{aligned}
P(y_i = 10|y) &= \int_0^\infty p(y_i = 10|\theta) p(\theta|y) d\theta \\
&= \int_0^\infty \theta e^{-\theta 10} \frac{\theta^{\alpha+n-1} e^{-\theta(\beta+n\bar{y})} (\beta+n\bar{y})^{(\alpha+n-1)}}{\Gamma(\alpha)} d\theta \\
&= \int_0^\infty \frac{\theta^{\alpha+n} e^{-\theta(\beta+n\bar{y}+10)} (\beta+n\bar{y})^{(\alpha+n-1)}}{\Gamma(\alpha)} d\theta
\end{aligned}
$$

2. (Propuesto) Se tiene que la distribución $Normal(\mu, \sigma^2)$ pertenece a la familia exponencial. Por lo tanto, describa la distribución $Normal(\mu, \sigma^2)$ como familia exponencial, es decir, identifique la función log-normalizadora, estadísticos suficientes y parámetros naturales.

Solución

La función de densidad de probabilidad de la normal es:

$$
p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/\left(2\sigma^2\right)}
$$

La cual se puede re-escribir de la siguiente forma:

$$
\begin{aligned}
p(x) &= \frac{1}{\sqrt{2\pi}} \exp\left(-\log\sigma - \frac{x^2}{2\sigma^2} + \frac{\mu x}{\sigma^2} - \frac{\mu^2}{2\sigma^2}\right) \\
&= \underbrace{\frac{1}{\sqrt{2\pi}}}_{h(x)} \exp\left(\eta^\top T(x) - \underbrace{\log\sigma - \mu^2/\left(2\sigma^2\right)}_{A(\theta)}\right)
\end{aligned}
$$

Donde:

- $T(x) = \begin{pmatrix} x \\ x^2 \end{pmatrix}$
- $\eta = \begin{pmatrix} \mu/\sigma^2 \\ -1/\left(2\sigma^2\right) \end{pmatrix}$
- $$\begin{aligned} A(\theta) &= \frac{\mu^2}{2\sigma^2} + \log\sigma \\ &= -\frac{[\theta]_1^2}{4[\theta]_2} - \frac{1}{2}\log\left(-2[\theta]_2\right) \end{aligned}$$