Estadística e inferencia I

Primer cuatrimestre — 2024

Segundo recuperatorio

Apellido y nombre: ....

---

1. Sea y una variable aleatoria con distribución normal con media θ y varianza conocida σ^2

(*a*) Demostrar que una distribución $\mathcal{N}(\mu_0, \tau_0^2)$ es un prior conjugado para θ y resulta en una distribución posterior $\mathcal{N}(\theta|\mu_1, \tau_1^2)$ * con $\mu_1 = (\frac{\mu_0}{\tau_0^2} + \frac{y}{\sigma^2})/(\frac{1}{\tau_0^2} + \frac{1}{\sigma^2})$ y $\frac{1}{\tau_1^2} = \frac{1}{\tau_0^2} + \frac{1}{\sigma^2}$.

*Ayuda: obtener la distribución a menos de una constante de normalización.

(*b*) Simular un conjunto de datos de tamaño $n = 10$ a partir de una distribución normal con media $\theta_{true} = 5$ y $\sigma = 1$. A partir de estos datos, calcular y graficar la likelihood en función de θ.

(*c*) Elegir un prior $\theta \sim \mathcal{N}(0, 1)$ y calcular la distribución posterior de θ. Graficar las distribuciones prior y posterior. ¿Cómo se interpretan las diferencias entre prior y posterior en cuanto a su posición y ancho? ¿Cómo se podría modificar el prior para que la posterior se le parezca más?

2. Sean $\{y_1, ..., y_n\}$ provenientes de una distribución exponencial con parámetro θ. Usando un prior conjugado $p(\theta) = \text{Gamma}(\alpha, \beta) = \frac{e^{-\beta\theta}\beta^\alpha\theta^{\alpha-1}}{\Gamma(\alpha)}$, se obtiene que la posterior de θ es una distribución $\text{Gamma}(\alpha + n, \beta + n\bar{y})$.

(*a*) Derivar el logaritmo de la posterior y obtener su moda.

(*b*) Construir la aproximación normal de la posterior basada en la derivada segunda su logaritmo evaluado en la moda.

3. Generar datos sintéticos de una regresión lineal simple con ruido gaussiano, como los que generábamos cuando veíamos regresión lineal frecuentista. Por ejemplo, considerar la función $f : \mathbb{R} \to \mathbb{R}$ dada por $f(x) = 2x - 1$, $X \sim \mathcal{U}(0, 1)$, $Y := f(X)$ y $n = 100$; tomar muestras $x_1, \ldots, x_n$ de X y aplicar la función f a cada muestra para calcular $y_i = f(x_i)$, con $1 \leq i \leq n$.

(*a*) Ajustar un modelo lineal de la forma $y \sim \text{N}(\beta_0 + \beta_1 x, \epsilon)$ usando `PyMC`. Graficar y describir las posterior de los parámetros de la regresión.

(*b*) ¿Cuál es el valor que estimamos que tenga $f(\frac{1}{2})$? ¿Con qué error?

(*c*) Graficar y dar percentiles de la distribución de $\beta_0 + \frac{1}{2}\beta_1$.

4. Estamos interesados en la proporción de éxito θ de una distribución Bernoulli. Tenemos una muestra observada que consiste en el número de éxitos en 30 ensayos independientes y con idéntica distribución de una binomial con parámetro θ; en esta muestra hay precisamente 19 éxitos. Asumimos que la distribución prior viene dada por una uniforme en $[0, 0.5]$ con peso 0.3 y otra uniforme en $[0.5, 1]$ con peso 0.7.

(*a*) Calcular la distribución posterior de θ y graficarla. ¿Cuál es el 90% HPDI?

(*b*) Calcular la probabilidad de obtener al menos 9 éxitos en 10 experimentos.