Estadística e inferencia I
Primer Cuatrimestre — 2024
Segundo Parcial

Apellido y nombre: ....

---

1. Se modela el tiempo de espera y hasta un cierto evento como proveniente de una distribución exponencial con parámetro θ. Es decir,

$$p(y|\theta) = \varepsilon(y|\theta) = \theta e^{-\theta y}$$

(a) Demostrar que $\text{Gamma}(\alpha, \beta) = \frac{e^{-\beta\theta}\beta^{\alpha}\theta^{\alpha-1}}{\Gamma(\alpha)}$ es un prior conjugado para θ. ¿Qué distribución sigue la posterior? (Ayuda: obtener la distribución a menos de una constante de normalización)

(b) Demostrar que si se realizan n observaciones $\{y_1, ..., y_n\}$, la verosimilitud tiene la misma forma funcional que una distribución $\text{Gamma}(n+1, n\bar{y})$ donde $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$. A partir de este resultado, ¿cómo se interpreta el prior conjugado en términos de resultados de hipotéticas observaciones exponenciales anteriores?

(c) Se mide el tiempo de espera y de cinco personas en la fila de un banco y se obtienen los siguientes resultados en minutos: $[4, 5, 4, 10, 7]$. Se modela $y \sim \varepsilon(\theta)$ y se propone un prior $\text{Gamma}(2, 4)$. Obtener la distribución posterior de θ, su esperanza y varianza posterior.

2. Sea $y \sim \text{Bi}(n, \theta)$ con un prior conjugado para θ:

(a) Construir una aproximación normal de $p(\theta|y)$ a partir de aproximar su logaritmo por una función cuadrática de θ.

(b) Elegir valores numéricos para y, n y los hiperparámetros del prior. Graficar y comparar la distribución obtenida con la distribución exacta y con la aproximación normal basada en el TCL. Dado un n no muy grande, ¿cuál aproximación conviene cuando se tienen muy pocos (o muchos) éxitos? Demostrar gráficamente.

3. (a) Tres fábricas F_1, F_2 y F_3 producen el 45, 30 y 25%, respectivamente, del total de azulejos disponibles en el mercado. La proporción de azulejos defectuosos de cada fábrica es 3, 4 y 5%, respectivamente. Si agarramos un azulejo defectuoso, ¿cuál es la probabilidad de que este azulejo sea de la segunda fábrica?

(b) Fijemos $N = 100$. Hay fábricas $F_1, \ldots, F_N$ tales que F_n produce $\frac{2n}{N(N+1)}$ del total de azulejos disponibles en el mercado, para cada N entre 1 y N. La proporción de azulejos defectuosos la fábrica n es $\frac{1}{2n}$. Supongamos que agarramos un azulejo defectuoso. Calcular, para cada n entre 1 y N, la probabilidad de que este azulejo sea de la fábrica F_n.

4. Consideremos los datos disponibles en https://tinyurl.com/3t6kvv94, que representan mediciones de horas mensuales de estudio y notas de alumnos. Vamos a pensar que la variable de respuesta es la nota, y, que modelamos como dependiente de las horas de estudio, x.

(a) Ajustar un modelo lineal de la forma $y \sim \mathbf{N}(\beta_0 + \beta_1 x, \varepsilon)$ usando `PyMC`. Graficar y describir las posterior de los parámetros β_0, β_1 y ε de la regresión.

(b) Calcular el 90% HDI de la nota que le corresponde a una persona que estudió 50 horas mensuales.

(c) Calcular el 90% HDI de la distribución de $\beta_0 + 50\beta_1$.