

# Estadística bayesiana

## Tarea 1

Fecha de entrega: 19 de septiembre

1. Suponer que se tiene una urna con 100 bolas, de las cuales  $100 - n$  son blancas y  $n$  son negras y donde  $n$  se distribuye uniformemente en  $\{0, \dots, 100\}$ . Si en el primer intento se saca una bola blanca y no se regresa a la urna, obtener la probabilidad (utilizando el aprendizaje bayesiano) de que la siguiente bola sea blanca.
2. Utilizando el clasificador naive Bayes, diseñar un filtro de spam para la base de datos *SMSSpams.txt*, para esto se recomienda utilizar el 80% de la muestra para entrenar el clasificador y probarlo en el 20% restante. Es importante notar que los signos, símbolos y caracteres numéricos pueden no incluirse en el análisis, así como el hecho de que la capitalización es irrelevante para el proceso de aprendizaje.
3. Considerar  $X_1, \dots, X_n$  una subcolección finita de una sucesión de variables aleatorias intercambi-ables tales que

$$f(x_1, \dots, x_n) = n! \left( 1 + \sum_{i=1}^n x_i \right)^{-(n+1)}.$$

Utilizando el teorema de representación de Bruno de Finetti mostrar que se pueden representar como una colección de variables condicionalmente independientes e idénticamente distribuidas con distribución común exponencial.

4. Demuestre que para toda sucesión intercambiable,  $\{X_i\}_{i=1}^\infty$ , se tiene que  $\text{Cov}(X_i, X_j) \geq 0$ .  
(*Hint: Considerar primero el caso finito*).
5. Sea  $X_1, \dots, X_n$  una colección de variables intercambiables, tales que dado  $\theta$  se considera que  $X_i \mid \theta \stackrel{\text{iid}}{\sim} \text{Po}(\theta)$ . Suponer además que  $\theta \sim \text{Ga}(\alpha, \beta)$ .
  - (a) Encuentre la distribución posterior de  $\theta \mid \mathbf{x}^{(n)}$ .
  - (b) Muestre que la media posterior se puede escribir como un promedio ponderado de la media a priori y el estimador máximo verosímil de  $\theta$ .

- (c) Finalmente, sea  $X_{n+1}$  una observación futura. Encuentre la distribución predictiva, así como la media y la varianza de  $X_{n+1} \mid \mathbf{x}^{(n)}$ .
6. Suponer que  $X \mid \mu \sim N(\mu, \sigma^2)$  y  $Y \mid \mu, \delta \sim N(\mu + \delta, \sigma^2)$  donde  $\sigma^2$  es conocida y  $X$  y  $Y$  son condicionalmente independientes dado  $\mu$  y  $\delta$ .
- (a) Encuentre la distribución conjunta de  $X$  y  $Y$  dado  $\mu$  y  $\delta$ .
- (b) Considerando la distribución impropia  $f(\mu, \delta) \propto 1$ , encuentre la distribución posterior de  $\mu$  y  $\delta$ . ¿Son independientes  $\mu \mid X, Y$  y  $\delta \mid X, Y$ ?
- (c) Encuentre la distribución marginal de  $\delta \mid X, Y$ .
- (d) Encuentre la distribución marginal de  $\mu \mid X, Y$ .
- (e) Considerando una nueva observación  $Z \mid \mu, \delta \sim N(\mu - \delta, \sigma^2)$ , con  $Z$  condicionalmente independiente de  $X$  y  $Y$  dado  $\mu$  y  $\delta$ . Encuentre la distribución predictiva de  $Z \mid X, Y$ .
7. Sea  $X_1, \dots, X_n$  una colección de variables aleatorias condicionalmente independientes dado  $\theta$ , tal que se asume que  $X_i \mid \theta \stackrel{\text{iid}}{\sim} U(0, \theta)$ . Donde  $\theta \sim \text{Pareto}(\alpha, \beta)$ , esto es, para  $\theta > \beta$  se tiene que

$$f(\theta) = \frac{\alpha \beta^\alpha}{\theta^{\alpha+1}}.$$

Obtén la distribución posterior. ¿Es una familia conjugada?

8. Sea  $X_1, \dots, X_n$  una secuencia intercambiable, tales que  $X_i \mid \theta \stackrel{\text{iid}}{\sim} N(\mu, \theta)$  con  $\mu$  conocida. Encuentre la familia conjugada para  $\theta$  y especifique la distribución posterior.
9. Sea  $X_1, \dots, X_n$  una secuencia intercambiable, tales que las  $X_i$  son condicionalmente independientes e idénticamente distribuidas dado un parámetro  $\theta$ . Para las siguientes familias de distribuciones encontrar la distribución de Jeffreys y la distribución posterior de  $\theta$ .
- (a)  $X_i \mid \theta \sim \text{Ber}(\theta)$
- (b)  $X_i \mid \theta \sim \text{Po}(\theta)$
- (c)  $X_i \mid \theta \sim \text{Exp}(\theta)$
- (d)  $X_i \mid \theta \sim \text{Maxwell}(\theta)$ , cuya densidad está dada por

$$f(x \mid \theta) = \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \theta^{\frac{3}{2}} x_i^2 \exp\left(-\frac{\theta x_i^2}{2}\right), \quad x > 0$$

10. Sea  $X_1, \dots, X_n$  una secuencia intercambiable, tales que  $X_i \mid \theta \stackrel{\text{iid}}{\sim} \text{Exp}(\theta)$ , donde  $\mathbb{E}(X_i \mid \theta) = \theta^{-1}$ .
- (a) Encuentre la distribución inicial de Jeffreys y comenta si es impropia.
  - (b) Deriva la distribución posterior.
  - (c) Considerando  $\phi = \log(\theta)$ , obtén la distribución de Jeffreys para  $\phi$  mediante los siguientes métodos:
    - i. Expresando  $\mathcal{L}(\theta)$  como  $\mathcal{L}(\phi)$
    - ii. Transformando la distribución de Jeffreys de  $\theta$  directamente a la de  $\phi$ .

## Actividades de DataCamp

Realizar los siguientes cursos

1. Fundamentals of Bayesian Data Analysis in R
2. Introduction to Data Visualization with ggplot2
3. Intermediate Data Visualization with ggplot2