Estadística bayesiana

Tarea 1

Fecha de entrega: 8 de marzo

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 $\{0, \ldots, 100\}$. 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 recordar 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 $\{X_1, \ldots, X_n\}$ una subcolección finita de una sucesión de variables aleatorias intercambiables tales que

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

   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 $\mathsf{Cov}(X_i, X_j) \geq 0$. *(Hint: Considerar primero el caso finito).*

5. Sea $\{X_1, \ldots, X_n\}$ una colección de variables intercambiables, tales que dado θ se considera que $X_i|\theta \overset{\text{iid}}{\sim} Po(\theta)$. Suponer además que $\theta \sim Ga(\alpha, \beta)$.

   (a) Encontrar la distribución posterior de $\theta|\mathbf{x}$.

   (b) Mostrar que la media posterior se puede escribir como un promedio ponderado de la media a priori y el estimador máximo verosímil de θ.

(c) Sea X_{n+1} una observación futura. Encontrar la distribución predictiva, así como la media y la varianza de $X_{n+1}|\mathbf{X}$.

6. Suponer que $X|\mu \sim N(\mu, \sigma^2)$ y $Y|\mu, \delta \sim N(\mu + \delta, \sigma^2)$ donde σ^2 es conocida y X y Y son condicionalmente independientes dado μ y δ.

   (a) Encontrar la distribución conjunta de X y Y dado μ y δ.

   (b) Considerando la distribución impropia $f(\mu, \delta) \propto 1$, encontrar la distribución posterior de μ y δ. ¿Son independientes $\mu|X, Y$ y $\delta|X, Y$?

   (c) Encontrar la distribución marginal de $\delta|X, Y$.

   (d) Encontrar la distribución marginal de $\mu|X, Y$.

   (e) Considerando una nueva observación $Z|\mu, \delta \sim N(\mu - \delta, \sigma^2)$, con Z condicionalmente independiente de X y Y dado μ y δ. Encontrar la distribución predictiva de $f(z|x, y)$.

7. Sea $X_1, \ldots, X_n$ una colección de variables aleatorias condicionalmente independientes dado θ, tal que se asume que $X_i|\theta \sim U(0, \theta)$.

   (a) Mostrar que $M = \max\{X_i\}$ es una estadística suficiente para $\mathbf{X} = (X_1, \ldots, X_n)$.

   (b) Considerar que $\theta \sim \text{Pareto}(\alpha, \beta)$, esto es, para $\theta > \beta$ se tiene que

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

   Obtener la distribución posterior ¿es una familia conjugada?

8. Sea $X_1, \ldots, X_n$ una secuencia intercambiable, tales que $X_i|\theta \overset{\text{iid}}{\sim} N(\mu, \theta)$ con μ conocida.

   (a) Encontrar el estadístico suficiente para θ.

   (b) Encontrar la familia conjugada y especificar la distribución posterior.

9. Sea $X_1, \ldots, X_n$ una secuencia intercambiable, tales que las X_i son condicionalmente independientes e idénticamente distribuidas dado un parámetro θ. Para las siguientes familias de distribuciones encontrar la distribución de Jeffreys y la distribución posterior de θ.

   (a) $X_i|\theta \sim \text{Bernoulli}(\theta)$.

   (b) $X_i|\theta \sim Po(\theta)$.

   (c) $X_i|\theta \sim Exp(\theta)$.

   (d) $X_i|\theta \sim N(\mu, \theta)$ con μ conocida.

10. Sea $X_1, \ldots, X_n$ una secuencia intercambiable, tales que $X_i|\theta \overset{\text{iid}}{\sim} Exp(\theta)$, donde $\mathbb{E}(X_i|\theta) = \theta^{-1}$.

(a) Encontrar la distribución inicial de Jeffreys y comenta si es impropia.

(b) Derivar la distribución posterior.

(c) Considerando $\phi = \log(\theta)$, obtener la distribución de Jeffreys para ϕ mediante los métodos:

   i. Expresando $\mathcal{L}(\theta)$ como $\mathcal{L}(\phi)$

   ii. Transformando la distribución de Jeffreys de θ directamente a la de ϕ. ¿Concuerdan ambos métodos?