Análisis Multivariado: Tarea 2

Distribuciones Multivariadas

Fecha de entrega: 14 de marzo

Distribución Normal Multivariada

1. (1 punto) Mostrar que si $\mathbf{x} \sim N_p(\mu, \Sigma)$, entonces $\mathbf{Ax}$ y $\mathbf{Bx}$ son independientes, si y solo si, $\mathbf{A}\Sigma\mathbf{B}^T = 0$.

2. (1 punto) Mostrar que si $\mathbf{y} \sim N_p(\mathbf{0}, \sigma^2\mathbf{I})$ y $\mathbf{A}_{p\times p}$ es una matriz simétrica, entonces $\mathbf{y^T A y} \sim \sigma^2\chi^2_r$, si y solo si, $\mathbf{A}$ es idempotente y tal que $ran(A) = r$. (*Hint: Para la ida obtener la función característica de $y^T\mathbf{A}y$ y compararla con la de la χ^2_r. Deducir que $\mathbf{A}$ tiene r eigenvalores iguales a 1 y así $\mathbf{A}$ es idempotente de rango r*).

3. (1 punto) Mostrar que si $\mathbf{x}_n = (x_{n1}, \ldots, x_{np})$ es una colección de vectores aleatorios independientes e idénticamente distribuidos, con vector de medias μ y matriz (finita) de covarianza Σ. Entonces,

$$\sqrt{n}\left(\bar{\mathbf{x}} - \mu\right) \quad \to \quad N_p\left(\mathbf{0}_p, \Sigma\right).$$

Distribución Wishart

4. (1 punto) Sea $\mathbf{x}_1, \ldots, \mathbf{x}_n$ una colección iid de vectores aleatorios $N_p(\mu, \Sigma)$, $\mathbf{A}_{n\times n}$ una matriz simétrica de rango r y $\mathbf{b}_{n\times 1}$ un vector de constantes. Mostrar que $\mathbf{X}^T b \sim N_p$, $\mathbf{X}^T\mathbf{AX} \sim W_p(r, \Sigma)$ y son independientes, si y solo si, $\mathbf{y}^T b \sim N_1$, $\mathbf{y}^T\mathbf{Ay} \sim \sigma_a^2 \cdot \chi^2_r$ y son independientes.

5. (1 punto) Sea $\mathbf{M} \sim W_p(n, \Sigma)$, entonces si consideramos la siguiente partición de $\mathbf{M}$ y Σ,

$$\mathbf{M} = \begin{pmatrix} \mathbf{M}_{11} & \mathbf{M}_{12} \\ \mathbf{M}_{21} & \mathbf{M}_{22} \end{pmatrix} \qquad\qquad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}.$$

Demostrar que $\mathbf{M}_{11}$ y $\mathbf{M}_{22}$ son independientes si $\Sigma_{12} = \mathbf{0}$.

Distribuciones Elípticas y Esféricas

6. (1 punto) Si $\mathbf{y} \sim S(g)$ y si se considera la transformación a coordenadas polares dada por,

$$\begin{aligned}
y_1 &= r\sin(\theta_1)\\
y_2 &= r\cos(\theta_1)\sin(\theta_2)\\
y_3 &= r\cos(\theta_1)\cos(\theta_2)\sin(\theta_3)\\
&\vdots\\
y_{p-1} &= r\cos(\theta_1)\cos(\theta_2)\cdots\cos(\theta_{p-2})\sin(\theta_{p-1})\\
y_p &= r\cos(\theta_1)\cos(\theta_2)\cdots\cos(\theta_{p-2})\cos(\theta_{p-1}),
\end{aligned}$$

mostrar lo siguiente:

i. La densidad del vector aleatorio $(R, \Theta_1, \ldots, \Theta_{p-1})$ está dada por

$$r^{p-1}\cos(\theta_1)^{p-2}\cos(\theta_2)^{p-3}\cdots\cos(\theta_{p-2})g(r^2).$$

ii. La densidad marginal de R está dada por

$$\frac{2\pi^{\frac{p}{2}} r^{p-1} g(r^2)}{\Gamma\left(\frac{p}{2}\right)}.$$

iii. Las marginales de $\Theta_1, \ldots, \Theta_{p-2}$ están dadas por

$$\frac{\Gamma\left(\frac{p-i}{2}\right)\cos(\theta_i)^{p-i-1}}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{p-i-1}{2}\right)},$$

y la densidad de Θ_{p-1} está dada por

$$\frac{1}{2\pi}.$$

Distribución Dirichlet

7. (1 punto) Sean $y_1, \ldots, y_p \overset{\text{ind}}{\sim} Ga(\alpha_i, \theta)$ mostrar que

$$(x_1, \ldots, x_p) = \left(\frac{y_1}{v}, \ldots, \frac{y_p}{v}\right) \sim Dir(\alpha_1, \ldots, \alpha_p).$$

Además mostrar que,

i. $\mathbb{E}(x_i) = \frac{\alpha_i}{\alpha_0} = \tilde{\alpha}_i$

ii. $\mathsf{Var}(x_i) = \frac{\tilde{\alpha}_i(1-\tilde{\alpha}_i)}{(1+\alpha_0)}$

iii. $\mathsf{Cov}(x_i, x_j) = \frac{-\alpha_i\alpha_j}{\alpha_0^2(\alpha_0+1)}$,

donde $\alpha_0 = \sum_i \alpha_i$.

Ejercicios Prácticos

8. (1 punto) Considera una urna con bolas de K diferentes colores, donde al tiempo 0, hay α_i bolas del i-ésimo color. Al tiempo n, se saca una bola y se regresa a la urna junto con una nueva bola de ese color y se repite el proceso N veces. Si $N \to \infty$, se puede demostrar que las proporciones de bolas en la urna seguirán una distribución Dirichlet.

   i. Considerando $K = 3$, $\alpha = c(2, 5, 1)$ y N "suficientemente grande", implementa este algoritmo y obtén una muestra de tamaño n.

   ii. Diseña e implementa un algoritmo para generar vectores aleatorios Dirichlet a partir de la transformación de variables aleatorias gammas. Obtén una muestra de tamaño n.

   iii. Para diferentes valores de N y n compara la media, el segundo momento y la covarianza muestral contra los resultados teóricos obtenidos en el ejercicio 7. ¿Cuál algoritmo es más eficiente?

9. (1 punto) Considera la base de datos *cork.txt* que contiene los pesos de corcho tomado de las 4 direcciones cardinales de 28 árboles. ¿Se puede asumir que los datos siguen una distribución normal multivariada?

10. (1 punto) Para la base de datos *wine.txt* asumir que las variables *alcohol* y *malic acid* siguen una distribución normal multivariada.

    i. Probar la hipótesis nula de que el vino promedio difiera de 13.15 grados de alcohol y 2.5 unidades de ácido málico.

    ii. Realizar los contrastes de hipótesis necesarios para verificar si existe o no una diferencia para los niveles de alcohol y ácido málico para las clases 1 y 2 de vinos.