

Departamento de Ingeniería Matemática  
 MA3402-1 Estadística  
 3 de septiembre de 2019



## Tarea 1: Estimador de máxima verosimilitud

**Profesor:** Felipe Tobar

**Auxiliares:** Diego Marchant, Francisco Vásquez

**Fecha de entrega:** 11 de septiembre 2019

**Contexto:** Esta tarea contiene partes teóricas y prácticas, donde el objetivo es evaluar su conocimiento sobre estimadores de máxima verosimilitud y cómo construirlos en situaciones reales. Por esta razón, es fundamental que sus respuestas estén presentadas de forma clara y evidencien su entendimiento de los conceptos evaluados y su análisis científico.

**Indicaciones:** Esta tarea debe ser realizada en grupos de 2 alumnos y si bien puede ser producida en cualquier formato (L<sup>A</sup>T<sub>E</sub>X recomendado) debe ser entregada en formato .pdf con una extensión máxima de 3 páginas incluyendo figuras, título, discusión y desarrollo.

**P1.** Considere las siguientes distribuciones uniformes:

(a)  $p_1(x|\theta) = 1/\theta, 0 \leq x \leq \theta$

(b)  $p_2(x|\theta) = 1/\theta, 0 < x < \theta$

(c)  $p_3(x|\theta) = 1, \theta \leq x \leq \theta + 1$

y asuma que para cada una de ellas tiene un conjunto de observaciones (o muestras IID)  $x = (x_1, \dots, x_n)$ . Calcule la verosimilitud para el parámetro  $\theta$  en cada caso y analice la existencia y unicidad de estimador de máxima verosimilitud.

**P2.** Usted ha sido contratado como *chief data science officer* en una importante aerolínea internacional. Su primer proyecto, de alta prioridad para la aerolínea, es predecir el número de pasajeros que vuelan en la aerolínea cada mes. Para esto, la unidad de estadística de la compañía indica que el número de pasajeros en el tiempo crece de forma cuadrática y exhibe una estacionalidad anual. Además, estudios preliminares de dicha división entregan evidencia para asumir que la dispersión sobre las tendencias recién descritas puede ser modelada como una VA gaussiana. Con esto, su equipo propone modelar el número de pasajeros  $X$  en el tiempo  $t$  de acuerdo a

$$X|t \sim \mathcal{N}\left(\theta_0 + \theta_1 t + \theta_2 t^2 + \theta_3 e^{\theta_4 t} \cos(\theta_5 t + \theta_6), \theta_7^2\right), \quad (1)$$

donde  $\{\theta_0, \theta_1, \theta_2\}$  modelan la tendencia cuadrática,  $\{\theta_3, \theta_4, \theta_5, \theta_6\}$  modelan la amplitud y frecuencia de la parte oscilatoria, y  $\theta_7$  es la desviación estándar del modelo. Nos referiremos a  $\theta$  como en el vector que contiene todos los parámetros antes descrito. El objetivo de este ejercicio es usar datos reales para identificar todos estos parámetros y hacer predicciones. Para este fin, siga las siguientes instrucciones.

(a) Considere un conjunto de observaciones de instantes de tiempo y de cantidad de pasajeros de la forma  $\{(t_i, x_i)\}_{i=1}^n$ . Calcule la log-verosimilitud de  $\theta$ .

- (b) Asuma que  $\{\theta_4, \theta_5, \theta_6\}$  son conocidos y encuentre estimador de máxima verosimilitud para el resto de los parámetros de forma analítica. **Hint:** escriba la media de la normal como  $\hat{\theta}^\top \phi(t)$ , donde  $\hat{\theta}^\top = [\theta_0, \theta_1, \theta_2, \theta_3]$  y  $\phi(t)$  no depende de ningún parámetro desconocido, y luego resuelva para  $\hat{\theta}$  y para  $\theta_7$ . **¡Observe que la primera componente de  $\phi$  es 1!**
- (c) Obtenga los datos de <https://www.kaggle.com/abhishekmamidi/air-passengers> y gráfíquelos.
- (d) Utilice estos datos para calcular los parámetros de acuerdo a su estimador de máxima verosimilitud y grafique la función de media junto a los datos. Elija valores apropiados para los parámetros que se asumieron conocidos. **Hint:** para elegir estos valores, interprete los parámetros (e.g., tasa de crecimiento y frecuencia) y proceda mediante “prueba y error” usando su gráfico.
- (e) Encuentre el estimador de máxima verosimilitud para todos los parámetros del modelo. Para esto, programe su log-verosimilitud calculada en el punto (a) y optimícela usando la función *minimize* del paquete *SciPy*. Grafique nuevamente, pero esta vez indique la región confianza del 95%. **Hint:** El intervalo de confianza del 95% de la distribución Gaussiana  $\mathcal{N}(\mu, \sigma^2)$  es  $[\mu - 2\sigma, \mu + 2\sigma]$ .
- (f) ¿Cómo puede verificar y convencer al resto de la compañía que su estimador predice efectivamente la cantidad de pasajeros? ¿Puede su modelo *generalizar a datos no vistos anteriormente*?