Métodos Estadísticos
Tarea 1

1. Considere una muestra aleatoria $X_1, ..., X_n$ con función de densidad
$$f(x|\theta) = \mathbb{1}_{\{\theta < x < \theta+1\}}, \quad \theta \in \mathbb{R}$$
   - Calcule un estimador para θ por el método de momentos y por el método de máxima verosimilitud, diga si estos son insesgados y suficientes para theta.
   - Con ayuda de la muestra aleatoria observada, **dataEj1.txt** , calcule la estimaciones en ambos casos y quédese con alguna de las dos, justificando su respuesta.
   - Finalmente haga un histograma (recuerdo normalizarlo) con los datos y encime la función de densidad $f(x|\hat{\theta})$, con $\hat{\theta}$ la estimación elegida del inciso anterior.
2. Considere una muestra aleatoria $X_1, ..., X_n$ con función de densidad
$$f(x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}, x = 0, 1, ... \quad \lambda > 0$$
   - Encuentre el MLE para $g(\lambda) = \lambda e^{-\lambda}$ y diga si este es insesgado.
   - Proponga un estimador insesgado para $g(\lambda) = \lambda e^{-\lambda}$, £Es el UMVUE para $g(\lambda)$?
   - Con ayuda del teorema de Rao-Blackwell encuentre el UMVUE para $g(\lambda)$
   - Con ayuda de la muestra aleatoria observada, **dataEj2.txt** , encuentre las estimaciones para cada uno de los estimadores anteriores, en este caso ya sabes cual es el mejor.
3. Un profesor debe entregar el promedio de las calificaciones de sus alumnos, sin embargo, sólo ha calificado a 20 de ellos, sabe que las calificaciones de sus alumnos se pueden modelar con siguiente función de densidad
$$f(x|\mu) = \frac{1}{\sqrt{3\pi}} e^{-\frac{(x-\mu)^2}{3}} \quad x \in \mathbb{R}$$
   - Calcule el MLE para μ de manera general, ¿es el UMVUE para μ?
   - Calcule el MLE para la probabilidad de que un alumno repruebe
   - Proponga un estimador insesgado para la probabilidad de que un alumno repruebe.
   - Con ayuda del archivo **dataEj3.txt** diga las estimaciones anteriores.
4. Considere que una aseguradora de automóviles tiene un registro por día de la cantidad de siniestros y su monto en el último año, **dataEj4.txt**, se sabe, que el número de siniestros

por día, tiene una distribución poisson con parámetro λ desconocido, además el monto de cada siniestro es independiente uno de otro y tienen la siguiente distribución

$$f(y|\alpha) = \frac{\alpha(90000)^{\alpha}}{(90000+x)^{\alpha+1}}\mathbb{1}_{\{x>0\}}, \quad \alpha > 0$$

Se sabe que, por día, el modelo que describe el monto total de siniestros es el siguiente

$$S = Y_1 + Y_2 + ... + Y_N$$

donde N indica la variable aleatoria poisson de los siniestros.

- Calcule el MLE del parámetro λ y de α, considere que se pueden trabajar cada estimación por separado.
- ¿Son los estimadores UMVUE?
- Calcule las estimaciones con ayuda de los datos.
- Haga un histograma de las observaciones del monto siniestros y dibuje encima la función de densidad $f(x|\hat{\alpha})$, con $\hat{\alpha}$ la estimación elegida del inciso anterior.
- Ahora que ya tiene las estimaciones de λ y de α haga una simulación del número de siniestros y el monto de cada uno de los siguientes 30 días.

5. Supongamos que tenemos observaciones $(x_1, y_1), ..., (x_n, y_n)$ y nosotros queremos utilizar un modelo que nos permita ayudar a encontrar el valor de y_{n+1} si yo tengo conocimiento de un valor x_{n+1}, es decir la idea es intentar predecir el valor de y a partir de un valor conocido x, el modelo que podemos utilizar, es el modelo de regresión lineal simple. (RLSI)

$$Y_i = \alpha + \beta X_i + U_i$$

donde $U_1, ...U_n$ tienen distribución $Normal(0, \sigma^2)$ y por lo tanto $Y_i \sim Normal(\alpha + \beta X_i, \sigma^2)$

Notemos que $Y_1, ..., Y_n$ son variables independientes, sin embargo no son idénticamente distribuidas, ademas α, β, σ^2 son parámetros desconocidos.

- Calcule los MLE's de α, β
- Calcule la media, varianza y covarianza muestral para cada uno de los 3 archivos, debe calcular la varianza tanto de la variable x como de la y, ¿qué observa?
- Diga cuales son las estimaciones de α y β para cada archivo, y ahora tiene una recta de regresión para cada archivo $\hat{y} = \hat{\alpha} + \hat{\beta}x$, dibuje este junto con el diagrama de dispersión de cada archivo, el diagrama de dispersión de cada archivo es la gráfica entre la coordenadas de la variable x y la variable y.