Instituto Tecnológico de Costa Rica **Escuela de Computación** Programa de Especialidad en Ciencias de los Datos **Curso: Estadistica** Profesor: Ph. D. Saúl Calderón Ramírez	QUIZ 0 Entrega: Domingo 17 de setiembre del 2023 Debe subir un *pdf* con la respuesta. A través del TEC digital Valor: 100 pts. Puntos Obtenidos: ____________ Nota: ________________

Nombre del (la) estudiante: Gabriel Valentine Fonseca

Carné: ___________________________

1. Suponga que, usted desea ajustar una función de densidad de distribución a un conjunto de datos que se refiere a la probabilidad de que un paciente sea atendido o no en menos de 5 minutos. Su objetivo es predicir la probabilidad con la que el paciente será atendido o no. Para ello, su equipo de cientificos de datos observa que el histograma de frecuencias puede ajustarse muy bien a un modelo paramétrico binomial.

$$p\,(t|\rho) = \rho^{t}\,(1-\rho)^{(t-1)}$$

Este modelo también se conoce como un modelo de Bernoulli. Como se observa, el modelo cuenta con un solo parámetro a estimar, en este caso ρ.

(a) **(50 pts) Suponga que se utilizará el algoritmo del descenso del gradiente, para obtener el ρ óptimo**, maximizando la verosimlitud de tal modelo, para la serie de valores (conjunto de datos)

$$\ln\left(\prod_{i=1}^{n} p^{t_i}(1-p)^{t_i-1}\right)$$

En este caso, el algoritmo del descenso del gradiente implementa la actualización del parámetro $\lambda\,(\tau+1)$ para una iteración $\tau+1$, como sigue:

$$\rho\,(\tau+1) = \rho\,(\tau) + \Delta\rho\,(\tau)\,.$$

Calcule el cambio en el parámetro ρ, $\Delta\rho\,(\tau)$. Muestre todos los pasos intermedios.

Dada la función:

$$\ln\left(\prod_{i=1}^{n} \rho^{t_i}(1-\rho)^{1-t_i}\right)$$

Aplicando propiedades:

$$\sum_{i=1}^{n} \ln\left(\rho^{t_i}(1-\rho)^{1-t_i}\right)$$

$$\sum_{i=1}^{n} \left(t_i\ln(\rho) + (1-t_i)\ln(1-\rho)\right)$$

$$\sum_{i=1}^{n} \left(t_i\ln(\rho) + (1-t_i)\ln(1-\rho)\right)$$

Se deriva respecto a p:

$$\frac{d}{d\rho}(t_i \ln(\rho)) = \frac{t_i}{\rho}$$

$$\frac{d}{d\rho}((1-t_i)\ln(1-\rho)) = \frac{1-t_i}{1-\rho}$$

Por lo tanto,

$$\Delta p(t) = \sum_{i=1}^{n}\left(\frac{t_i}{\rho} + \frac{1-t_i}{1-\rho}\right)$$

(b) **(50 pts)** Un miembro de su equipo sospecha que puede usarse una forma más eficiente para optimizar el parámetro ρ: derivando respecto a tal parámetro, igualando la ecuación a 0 y despejandolo. Demuestre si es posible hacerlo, y en caso de ser posible, muestre todos los pasos para llegar a la expresión del ρ óptimo. Finalmente responda, porqué el esquema sugerido por su colega es más efectivo que el implementado en el apartado anterior?

Derivando e igualando a 0:

$$\sum_{i=1}^{n}\left(\frac{t_i}{\rho} + \frac{1-t_i}{1-\rho}\right) = 0$$

$$\sum_{i=1}^{n}(t_i(1-\rho) - \rho(1-t_i)) = 0$$

$$\sum_{i=1}^{n} t_i - \sum_{i=1}^{n} t_i\rho - \sum_{i=1}^{n}\rho + \sum_{i=1}^{n}\rho t_i = 0$$

$$\sum_{i=1}^{n} t_i - N\rho = 0$$

$$\rho = \frac{1}{N}\sum_{i=1}^{n} t_i$$

El método propuesto por el miembro de mi equipo es más eficiente en este caso debido a que ofrece una solución analítica directa sin requerir iteraciones. Esto puede resultar en una mayor eficiencia computacional y una convergencia más rápida, especialmente en problemas que cuentan con una solución cerrada, como el que estamos abordando.