

EYP1113 - Probabilidad y Estadística
Tarea I

Objetivo

El siguiente proyecto tiene como objetivo que el grupo de trabajo sea capaz de ajustar una distribución de probabilidad a un conjunto de observaciones y predecir a partir de representaciones de las esperanzas condicionales el valor de una variable.

Introducción

Durante el transcurso del semestre se han visto distintos modelos de probabilidad y algunas técnicas estadísticas para la estimación de ellos. Una herramienta muy utilizada en ingeniería para el ajuste de modelos de probabilidad a un conjunto de observaciones son los llamados “gráficos de probabilidad”, los cuales permiten a partir de una representación lineal de los percentiles teóricos de un modelo ajustar una recta cuyo intercepto y pendiente son útiles para estimar sus parámetros. Por otra parte, al finalizar el capítulo 3 se vio que el mejor predictor de una variable es siempre la esperanza condicional, y para ver ilustrar esto, representaremos el comportamiento de dicha esperanza mediante funciones “simples” en base a la información que se disponga para predecir.

Parte I [40 %]

Importe los datos que se encuentran en el archivo `Datos01.txt` y grafique el histograma de densidad.

Como podrá ver, el comportamiento de estos datos es asimétrico, por esta razón la distribución normal no deja contento a unos especialistas que proponen las distribuciones SEV y Weibull como mejor opción.

La función de distribución de probabilidad acumulada $F_X(x)$ y función de densidad $f_X(x)$ de una SEV está dada por:

$$F_X(x) = 1 - \exp \left[-\exp \left(\frac{x - \mu}{\sigma} \right) \right] \quad \text{y} \quad f_X(x) = \frac{1}{\sigma} \cdot \exp \left[\left(\frac{x - \mu}{\sigma} \right) - \exp \left(\frac{x - \mu}{\sigma} \right) \right]$$

con $E(X) = \mu - \sigma \gamma$ y $\text{Var}(X) = \frac{\sigma^2 \pi^2}{6}$, donde $\gamma = 0,5772$ y $\pi = 3,1416$.

Para **R**, copiar y pegar en la consola las funciones que se encuentran en `SEV.txt`.

- `f_X(x) = dsev(x, location = μ, scale = σ)`
- `F_X(q) = psev(q, location = μ, scale = σ)`
- `F_X-1(p) = qsev(p, location = μ, scale = σ)`

Mientras que la distribución Weibull está dada por:

$$F_X(x) = 1 - \exp \left[-\left(\frac{x}{b} \right)^a \right] \quad \text{y} \quad f_X(x) = \left(\frac{a}{b} \right) \left(\frac{x}{b} \right)^{a-1} \exp \left[-\left(\frac{x}{b} \right)^a \right]$$

con $E(X) = \frac{b}{a} \Gamma \left(\frac{1}{a} \right)$ y $\text{Var}(X) = b^2 \left\{ \Gamma \left(1 + \frac{2}{a} \right) - \left[\Gamma \left(1 + \frac{1}{a} \right) \right]^2 \right\}$.

En **R**.

- $f_X(x) = \text{dweibull}(x, \text{shape} = a, \text{scale} = b)$
- $F_X(q) = \text{pweibull}(q, \text{shape} = a, \text{scale} = b)$
- $F_X^{-1}(p) = \text{qweibull}(p, \text{shape} = a, \text{scale} = b)$

Construya un gráfico de probabilidad para la distribución Normal, SEV y Weibull. Entregue una tabla resumen con las estimaciones obtenidas y pegue al histograma los tres ajustes realizados.

Parte II [60 %]

Desde SINCA (<http://sinca.mma.gob.cl/>) y CSN (<http://www.sismologia.cl>) se extrajo información meteorológica desde el 1ro de abril de este año hasta el día viernes 5 de mayo y los movimientos telúricos registrados en este período de tiempo. La base `Datos02.txt` empalma a cada sismo, las variables meteorológicas de la hora en que ocurrió y `Datos03.txt` empalma a las variables meteorológicas, conteo de sismos y promedios las variables de latitudes, longitudes, magnitudes y profundidades.

- (a) [40 %] Describa a partir de medidas empíricas y gráficos adecuados las variables: Latitud, Longitud, Magnitud, Profundidad, Radiacion, Presion, Humedad, Temperatura y VelocidadViento, que se encuentran en la `Datos02.txt`.
- (b) [20 %] Represente el comportamiento del número esperado de sismos por hora mediante una función simple con respecto al tiempo u otra variables, desde la ocurrencia del terremoto 6.9° Richter e ilustre gráficamente. La función `lm()` y `nls()` son útiles para estimar coeficientes de funciones lineales y no lineales. El conteo de sismos se encuentra en `Datos03.txt`.

Importante

Debe presentar un informe con lo solicitado en que se explique en detalle lo realizado. Puede agregar los script de **R** utilizados en una Anexo. La presentación y escritura del informe será considerada en su nota. El plazo de entrega finaliza el día viernes 19 de mayo de 2017 a las 23:59 horas.