EYP1113 - Probabilidad y Estadística

Laboratorio 7

Profesoras: Natalia Venegas y Pilar Tello

Facultad de Matemáticas
Departamento de Estadística
Pontificia Universidad Católica de Chile

Segundo Semestre 2019

Determinación de Modelos Probabilísticos

El modelo de distribución de probablidad apropiado para describir un fenómeno es generalmente, desconocido.

Bajo ciertas circunstancias, las propiedades básicas del proceso físico subyacente del fenómeno aleatorio sugiere la forma de la distribución de probabilidades.

Ejemplos

- Cumple vs. No cumple $\rightarrow$ Bernoulli.
- Número de ”eventos" en periodos $\rightarrow$ Poisson.
- Tiempos de duración o espera $\rightarrow$ Exponencial.
- Suma de eventos individuales $\rightarrow$ Normal.
- Condiciones extremas de un proceso $\rightarrow$ Valor extremo.

En muchas situaciones, la distribución de probabilidad debe ser determinada empíricamente a partir de los datos.

Inicialmente, aproximaciones gráficas (Histograma v/s Densidad) nos pueden ayudar a inferir ”visualmente" sobre la distribución.

También, con datos disponibles, pueden obtenerse los gráficos de probabilidad (Probability Papers) para distribuciones dadas (si los puntos están en línea recta, la distribución es apropiada).

Por último, dada una distribución a priori puede evaluarse la ”bondad de ajuste" (Test χ^2, Test de Kolmogorov-Smirnov o el Test de Anderson-Darling, entre otros).

En este laboratorio nos enfocaremos en la construcción de un gráfico de probabilidad.

Es la representación gráfica de los datos observados y sus correspondientes frecuencias acumuladas.

Para un conjunto de N observaciones, $x_1, \ldots, x_N$ ordenados de menor a mayor, el m-ésimo valor es graficado contra la probabilidad acumulada de $\frac{m}{N+1}$.

La utilidad del ”papel" de probabilidad es reflejar ”el ajuste" que presentan los datos con respecto a la distribución subyacente.

La linealidad o falta de esta nos indica lo adecuado o inadecuado de la distribución.

Sean $x_{(1)}, \ldots, x_{(N)}$ observaciones ordenadas de menor a mayor y $p_1 = \frac{1}{N+1}, \ldots, p_N = \frac{N}{N+1}$ sus respectivas probabilidades empíricas. Calculemos los percentiles teóricos $\Phi^{-1}(p_i)$ de una distribución Normal Estándar para cada p_i, con $i = 1, \ldots, N$.

Si los x's distribuyen *Normal*(μ, σ) entonces la siguiente relación lineal se debe cumplir:

$$x_{(q)} = \mu + \sigma \cdot \Phi^{-1}(p_q)$$

Gráficos de probabilidad: Distribución Log-Normal

Sean $x_{(1)}, \ldots, x_{(N)}$ observaciones ordenadas de menor a mayor y $p_1 = \frac{1}{N+1}, \ldots, p_N = \frac{N}{N+1}$ sus respectivas probabilidades empíricas. Calculemos los percentiles teóricos $\Phi^{-1}(p_i)$ de una distribución Normal Estándar para cada p_i, con $i = 1, \ldots, N$.

Si los x's distribuyen $log - Normal(\lambda, \zeta)$ entonces la siguiente relación lineal se debe cumplir:

$$\ln x_{(q)} = \lambda + \zeta \cdot \Phi^{-1}(p_q)$$

1. Construya un gráfico de probabilidad para una distribución Normal. Simule datos provenientes de una distribución Normal para probar el gráfico construido.
2. Construya un gráfico de probabilidad para una distribución Log-Normal. Simule datos provenientes de una distribución Log-Normal para probar el gráfico construido.
3. Simule datos provenientes de una distribución $Gamma(k = 10, \nu = 0.1)$ y ajuste los modelos anteriores en un histograma según la estimación obtenida del gráfico de probabilidad.