Ejercicios Tema 3 - Distribuciones Notables: más distribuciones notables

Ricardo Alberich, Juan Gabriel Gomila y Arnau Mir

Curso de Probabilidad y Variables Aleatorias con R y Python

Distribuciones Notables: más distribuciones notables.

Ley de Bendford

La ley de Benford es una curiosa distribución de probabilidad que suele aparecer en la distribución del primer dígito de las cantidades registradas en contabilidades y en observaciones científicas o datos numéricos. La variable X sigue una distribución discreta Benford con dominio $D_X = \{1, 2, 3, 4, 5, 7, 8, 9\}$ son 9 dígitos (se elimina el cero) y sin función de probabilidad es

$$P_X(x) = P(X = x) = \log(d + 1) - \log(d).$$

- a) Calcular la media y la varianza de X.
- b) Calcular la función de distribución de X.
- c) ¿Cuál es el dígito más frecuente (moda)?
- d) Construid con R las funciones de probabilidad y de distribución de X.
- e) Dibujar con R las funciones del apartado anterior.

Distribución de Pareto (Power law)

Es una distribución que aparece en muchos ámbitos. Consideremos el económico. Supongamos que en un gran país consideramos la población activa económicamente; desde el más humilde becario al directivo más adinerado.

Escogemos un individuo al azar de esta población y observamos la variable $X =$ sus ingresos en euros (digamos que anuales).

Un modelo razonable es el que supone que:

- Hay un ingreso mínimo $x_m > 0$.
- La probabilidad de un ingreso mayor que x decrece de forma inversamente proporcional al ingreso x, es decir proporcional a $\left(\frac{x_m}{x}\right)^{\gamma \cdot x}$ para algún número real $\gamma > 1$.

Maś formalmente. dado $x > x_m$

$$P(X > x) = k \cdot \left(\frac{x_m}{x}\right)^{\gamma}.$$

Luego su función de distribución es

$$F_X(X) = P(X \leq x) = \begin{cases} 1 - P(X > x) = 1 - k \cdot \left(\frac{x_m}{x}\right)^{\gamma} & \text{si } x > x_m \\ 0 & \text{si } x \leq x_m \end{cases}$$

Se pide

- a) Calcular en función de k y γ la densidad de la variable X.
- b) Para $\gamma > 1$ calcular $E(X)$ y $Var(X)$ y su desviación típica.

- c) ¿Qué sucede con $E(X)$ si $0 < \gamma < 1$.
- d) ¿Cómo se calcula está distribución con R y con python?
- e) Dibujar las gráficas de su densidad y distribución para $\gamma = 3$ y $\gamma = 5$.
- f) Explorar por internet (wikipedia) cómo es la distribución **power law** y qué relación tiene el concepto de *scale free* con los resultados del apartado c).

Distribución de Gumbel (teoría del valor extremo).

La distribución de Gumbel aparece en variables que miden lo que se llama un valor extremo: precipitación máxima de lluvia, tiempo máximo transcurrido entre dos terremotos, o en métodos de machine learning el máximo de las puntuaciones de una algoritmo; por ejemplo comparar pares de objetos (fotos, proteínas, etc.).

Una variable aleatoria sigue una ley de distribución Gumbel (de TIPO I) si su distribución es:

$$F_X(x) = \begin{cases} e^{-e^{-\frac{x-\mu}{\beta}}} & \text{si} x \geq 0 \\ 0 & \text{si} x < 0 \end{cases}$$

Para μ y $\beta > 0$ parámetros reales. Llamaremos distribución Gumbel estándar a la que tiene por parámetros $\mu = 0$ y $\beta = 1$.

- a) Si X es una Gumbel estándar calcular su función de densidad y dibujar su gráfica.
- b) Consideremos la función $F(x) = e^{-e^{-x}}$ para $x \geq 0$ y que vale cero en el resto de casos. Comprobar que es la función de distribución $P(X \leq x)$ de una v.a. Gumbel estándar.
- c) Buscad un paquete de R que implemente la distribución Gumbel. Aseguraros de que es la (Gumbel Tipo I). Dejando fijo el parámetro $\beta = 1$ dibujar la densidad Gumbel para varios valores de μ y explicad en que afecta a la gráfica el cambio de μ.
- d) Dejando fijo el parámetro μ dibujad la densidad Gumbel para varios valores de $\beta > 0$ y explicar en que afecta a la gráfica el cambio de este parámetro.
- e) Buscad cuales son las fórmulas de la esperanza y varianza de una distribución Gumbel en función de α y β.
- f) Repetid los apartados c) y d) con python. Con python se puede pedir con la correspondiente función la esperanza y varianza de esta distribución, comprobar con esta función para algunos valores las fórmulas de la esperanza y la varianza del apartado e).