

Introducción a distribuciones de probabilidad

Ramon Ceballos

6/2/2021

Distribuciones de probabilidad

Distribución de probabilidad. En teoría de la probabilidad y estadística, **la distribución de probabilidad de una variable aleatoria** es una función que asigna a cada suceso definido sobre la variable la probabilidad de que dicho suceso ocurra.

1. Sesgo de una distribución de probabilidad

El **sesgo (skewness) de una distribución de probabilidad** mide que tan simétrica o asimétrica es una distribución en torno al valor central.

En el caso de las distribuciones unimodales es fácil detectarlo. Pero en distribuciones bimodales o formadas de diferentes subdistribuciones será más complejo determinar el sesgo de forma visual.

El sesgo e una distribución de probabilidad se puede calcular de diversas formas, no existiendo un método formal para dicho cálculo.

El sesgo puede ser negativo o positivo. Se dice que el **sesgo es negativo** (sesgo hacia la dcha) porque la cola de la parte izda de la curva es demasiado alargada en relación a donde se concentran la mayor parte de valores d ela distribución. El caso contrario es el del **sesgo positivo**. Para mayor claridad busca en wikipedia skewness. Cuanto más cercano sea el **sesgo a 0**, mayor será el grado de simetría de la distribución.

En matemáticas muchas veces el sesgo se ha definido como una relación entre la media y la mediana. Por tanto se define como:

$$sesgo = \frac{\mu - mediana}{\sigma}$$

siendo μ la media, σ la desviación estándar.

También se puede definir con la fórmula de los momentos de Pearson (**Pearson's moment coefficient of skewness**). Esta es la definición que emplea **r**.

2. Distribuciones en R y funciones

Dada cualquier variable aleatoria, **va**, **R** nos da cuatro funciones para poder trabajar con ellas:

- **dva(x,...)**: *Función de densidad o de probabilidad $f(x)$* de la variable aleatoria para el valor x del dominio de definición. Para una distribución normal utilizaremos **dnorm()**, para una de Poisson usaremos **dpois()** y así con todas.
- **pva(x,...)**: *Función de distribución* (es la probabilidad) $F(x)$ de la variable aleatoria para el valor x del dominio de definición. Se emplea la misma sintaxis que la anterior cambiando la **d** por la **p**.
- **qva(p,...)**: *Cuantil p -ésimo* de la variable aleatoria (el valor de x más pequeño tal que $F(x) \geq p$).
- **rva(n,...)**: *Generador de n observaciones aleatorias* siguiendo la distribución de la variable aleatoria.

Distribuciones en Python y funciones

Dada cualquier variable aleatoria, en **Python** tenemos las mismas cuatro funciones, sin que su nombre dependa del nombre de la función como pasaba en **R**. En este caso será el nombre de la distribución seguido de un pto y seguido del nombre de una de las cuatro funciones siguientes:

- **pmf(k, ...)** (probability mass function o función de probabilidad) o **pdf(x, ...)** (probability density function o función de densidad): *Función de probabilidad $f(k)$ o de densidad $f(x)$* de la variable aleatoria para los valores k o x del dominio.
- **cdf(x, ...)** (cumulative density function): *Función de distribución $F(x)$* de la variable aleatoria para el valor k del dominio.
- **ppf(p, ...)**: *Cuantil p -ésimo* de la variable aleatoria (el valor de x más pequeño tal que $F(x) \geq p$).
- **rvs(size, ...)** (random value generator) : *Generador de $size$ observaciones* siguiendo la distribución de la variable aleatoria elegida.

También vale la pena conocer la función **stats(moments='mvsk')** que nos devuelve cuatro valores con los estadísticos de la media **m**, la varianza **v**, el sesgo **s** y la curtosis **k** de la distribución.