VARIABLE ALEATORIA

Jesús García de Jalón de la Fuente

1. Distribuciones de probabilidad

Consideremos un espacio muestral E correspondiente a un experimento aleatorio, por ejemplo, lanzar dos dados. A cada elemento del espacio muestral, es decir, a cada resultado del experimento, podemos asociarle un número, por ejemplo, la diferencia de las puntuaciones obtenidas. Estos valores numéricos asociados a cada resultado forman la variable aleatoria. A cada valor de la variable aleatoria le corresponde una probabilidad. Por ejemplo, en el caso anterior, estas probabilidades serían:

VARIABLE	PROBABILIDAD
0	$\frac{6}{36}$
1	$\frac{10}{36}$
2	$\frac{8}{36}$
3	$\frac{6}{36}$
4	$\frac{4}{36}$
5	$\frac{2}{36}$

Se llama distribución de probabilidad al conjunto de una variable aleatoria y de las probabilidades asociadas a cada valor.

La variable aleatoria puede ser discreta o continua. En este último caso solo tiene sentido asignar probabilidades a intervalos de valores. La probabilidad asociada a un valor particular de la variable aleatoria es cero.

En general, una distribución de probabilidad de variable discreta se puede expresar mediante una tabla:

VARIABLE	PROBABILIDAD
x_1	p_1
x_2	p_2
$\cdots$	$\cdots$
x_n	p_n

Como en el caso de las frecuencias relativas, la suma de todas las probabilidades debe ser igual a 1.

Para una distribución de probabilidad de variable discreta, la media (o esperanza matemática) y la desviación típica se definen de la misma forma que para la variable estadística sustituyendo la frecuencia relativa por la probabilidad:

$$\mu = \sum p_i x_i \qquad \sigma^2 = \sum p_i (x_i - \mu)^2 = \sum p_i x_i^2 - \mu^2$$

2. La distribución binomial

La distribución binomial es un caso particualar de distribución de probabilidad de variable discreta. Para definirla debemos establecer en primer lugar un experimento aleatorio, a partir de él definir la variable aleatoria y finalmente calcular las probabilidades correspondientes.

En el caso de la distribución binomial, el experimento aleatorio consiste en una prueba con dos posibles resultados que se repite n veces. Las pruebas que se repiten son independientes (el resultado de una prueba no influye en la siguiente) y a los dos resultados los llamaremos éxito y fracaso. La variable aleatoria será el número de éxitos obtenidos en las n pruebas.

Sii en cualquiera de las pruebas que se repiten, la probabilidad de éxito es p, la probabilidad de fracaso es $q = 1 - p$ y el número de veces que se repite la prueba es n, hablaremos de la distribución binomial $B(n, p)$.

Puede demostrarse, que la probabilidad de obtener k éxitos en las n pruebas está dadas por la siguiente fórmula:

$$p(X = k) = \binom{n}{k} p^k q^{n-k}$$

y que en base a esta distribución de probabilidad, la media y la desviación típica valen:

$$\mu = np \qquad \sigma^2 = npq \qquad \sigma = \sqrt{npq}$$

3. La distribución normal

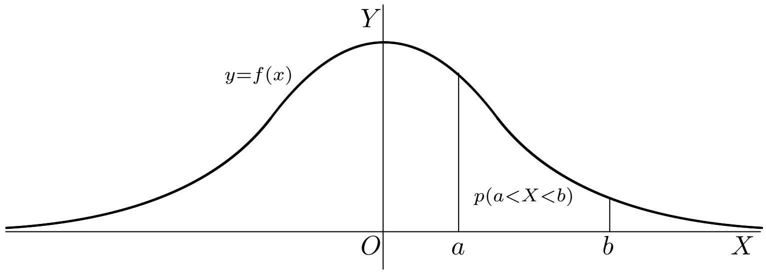
La variable aleatoria es continua si puede tomar los infinitos valores de un cierto intervalo $[a, b]$ (a y b pueden ser infinitos). Una distribución de probabilidad de variable continua debe asignar una probabilidad a cualquier intervalo de valores de la variable aleatoria. Esta asignación se hace con ayuda de dos funciones, una función de densidad $f(x)$ y una función de distribución $F(x)$. Veamos el significado de estas funciones.

La función de densidad es una función positiva que cumple que la probabilidad de que la variable aleatoria se encuentre en el intervalo $[a, b]$ es igual al área bajo la curva en ese intervalo; o expresado en términos de integrales:

$$p(a < X < b) = \int_a^b f(x)\, dx$$

Como la suma de todas las probabilidades debe ser igual a 1, la función de densidad debe cumplir que el área total bajo la curva debe ser 1, o lo que es lo mismo:

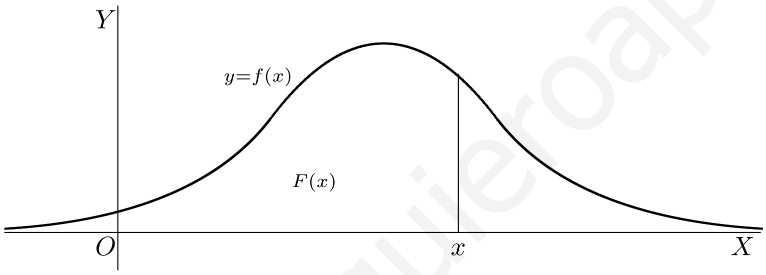
$$\int_{-\infty}^{+\infty} f(x)\,dx = 1$$



La media y la desviación típica de una distribución de probabilidad de variable continua se obtiene mediante fórmulas similares de las expuestas para variable discreta sustituyendo las sumas por integrales:

$$\mu = \int_{-\infty}^{\infty} x f(x)\,dx \qquad \sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx = \int_{-\infty}^{\infty} x^2 f(x)\,dx - \mu^2$$

La función de distribución $F(x)$ representa probabilidades acumuladas. Así, $F(x)$ es la probabilidad de obtener un resultado menor (o menor o igual cuya probabilidad es la misma) que x.



Conocida la función de distribución $F(x)$ puede calcularse fácilmente la probabilidad asociada a cualquier intervalo mediante diferencias:

$$p(a < X < b) = F(b) - F(a)$$

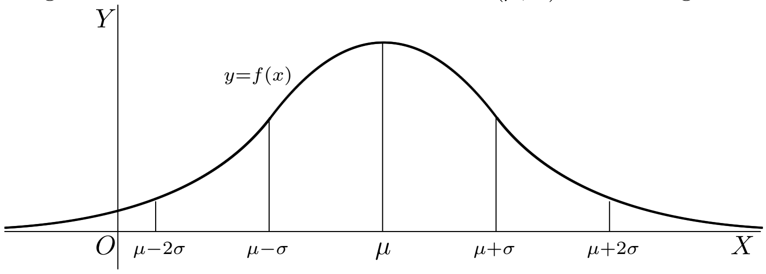
Una distribución de probabilidad de variable continua cuya importancia se comprenderá en el tema de inferencia estadística es la distribución normal. La distribución normal de media μ y desviación típica σ tiene la siguiente función de densidad:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

La distribución normal de media 0 y desviación típica 1 se indica mediante $N(0,1)$. Los valores de la función de distribución de $N(0,1)$ se encuentran en las tablas de la distribución normal. A partir de estos valores puede obtenerse la función de distribución de $N(\mu,\sigma)$ tipificando la variable, esto es, si z es la variable de $N(0,1)$ y x la variable de $N(\mu,\sigma)$, se pasa de una a otra mediante el cambio:

$$z = \frac{x-\mu}{\sigma} \qquad \text{o bien} \qquad x = \mu + z\sigma$$

La gráfica de la función de densidad de $N(\mu, \sigma)$ tiene la siguiente forma:



La función es simétrica respecto a $x = \mu$ y es tanto más aplanada cuanto mayor sea σ. La probabilidad correspondiente al intervalo $(\mu - \sigma, \mu + \sigma)$ (es decir al intervalo $(-1, 1)$ para la variable tipificada z) es aproximadamente de $0,68$. Para el intervalo $(\mu - 2\sigma, \mu + 2\sigma)$ es aproximadamente de $0,95$ y para $(\mu - 3\sigma, \mu + 3\sigma)$ de $0,99$.

4. Distribuciones binomial y normal

Consideremos la distribución binomial $B(n, p)$. Para valores grandes de n, o más exactamente, cuando los valores de np y nq son suficientemente grandes (en la práctica, cuando son ambos mayores que 5), las probabilidades de la distribución binomial peden calcularse aproximadamente a partir de la correspondiente distribución normal $N(np, \sqrt{npq})$.

Sea x' la variable aleatoria de $B(n, p)$ y x la variabla aleatoria de $N(np, \sqrt{npq})$. Puesto que x' es una variable discreta y x una variable continua, para relacionar las probabiidades entre ambas distribuciones, será preciso asociar a cada valor de x' un intervalo de valores de x. De forma natural, se la asocia a un valor $x' = k$ el intervalo de valores de x, $k - 0{,}5; k + 0{,}5$.

De esta forma, tenemos la siguiente relación entre las probabilidades de ambas distribuciones:

- $p(x' = k) \simeq p(k - 0{,}5 < x < k + 0{,}5)$
- $p(x' < k) \simeq p(x < k - 0{,}5)$
- $p(x' \leq k) \simeq p(x < k + 0{,}5)$
- $p(x' > k) \simeq p(x > k + 0{,}5)$
- $p(x' \geq k) \simeq p(x > k - 0{,}5)$