Tema 7

DISTRIBUCIONES CONTINUAS DE PROBABILIDAD

La distribución normal:

La distribución normal, campana de Gauss o, curva normal, también definida por De Moivre.

Características y propiedades:

La siguiente fórmula recoge la función:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad \textit{para } -\infty < x < \infty$$

donde μ y σ, media y desviación típica, son sus parámetros, $\pi = 3{,}1416$ y $e = 2{,}718$ y base de los logaritmos neperianos.
Si una variable X tiene una distribución que se ajusta a la fórmula anterior, es una distribución normal y se expresa $X \rightarrow N$; indicando que tiene una distribución normal con parámetros μ y σ.
Si una variable X le aplicamos una transformación lineal $Y = bX+a$, la nueva variable se distribuirá normalmente pero con media $b\mu_x + a$ y la desviación típica $|b|\sigma_x$. Si restamos la media y dividimos por la desviación típica obtenemos una nueva variable "z". Por tanto:

$$z \rightarrow N(0,1)$$

Y su función de probabilidad será:

$$f(z) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{Z^2}{2}} \qquad \textbf{para } -\infty < z < \infty$$

Una distribución normal es simétrica a su media, μ, coincide con su mediana y su moda.
La curva normal tiene dos puntos de inflexión; dos puntos donde la curva pasa de ser cóncava a convexa. Estos puntos están a la distancia de una desviación típica de la media.
Es asintótica en el eje de abscisas, se extiende desde $-\infty$ hasta $+\infty$ sin tocar nunca el eje.

Casos de utilización de las tablas:

En el supuesto que la tabla no recoja el valor, podemos utilizar el más próximo.

1. Cálculo de la probabilidad para valores menores o iguales que una determinada puntuación típica:

En este caso se mira directamente en la tabla.

2. Cálculo de la probabilidad para valores mayores que una determinada puntuación:

En este supuesto se mira en la tabla la probabilidad que esa puntuación deja por debajo y se resta a 1.

3. Cálculo de la probabilidad entre dos puntuaciones determinadas:

Aquí se restan las probabilidades que dejan por debajo de sí las dos puntuaciones típicas.

Histograma y distribución normal:

Si disponemos de los datos originales de una variable X, y su distribución es normal,utilizaremos las tablas III y IV, pero anteriormente transformaremos las puntuaciones directas en puntuaciones típicas:

$$z_i = \frac{X_i - \bar{X}}{S_x}$$

Aproximación de la binomial a la normal:

Cuando las distribuciones binomiales superan sus valores de 20, se puede aproximar a la binomial normal. Teniendo una variable X, con distribución binomial, su media es $\mu = np$ y su desviación típica $\sigma = \sqrt{npq}$. Podemos realizar:

$$P(X = x) = P\left(\frac{(x-0,5) - \mu}{\sigma} \leq \frac{x-\mu}{\sigma} \leq \frac{(x+0,5) - \mu}{\sigma}\right)$$

$$\Downarrow$$

$$P(X = x) = P\left(\frac{(x-0,5) - np}{\sqrt{npq}} \leq z \leq \frac{(x+0,5) - np}{\sqrt{npq}}\right)$$

Ejemplo:

Tiramos 20 veces una moneda al aire; ¿cuál es la probabilidad de que salgan 12 caras?
Tenemos que número de caras x = 12, con números de intentos n = 20 y p = 0,5. Mirando en la tabla I el valor de la probabilidad de éxito es 0,1201.
Ahora calculamos la media y la desviación típica:

$$\mu = 20 \,.\, 0,5 = 10 \;//\; y;\ \sigma = \sqrt{npq} = \sqrt{20.0,5.0,5} = \sqrt{5} = 2,24$$

Para aproximar la distribución binomial a la normal establecemos un intervalo entre 0,5 a la izquierda y a la derecha:

$$P\,[(12-0,5) \leq x \leq (12+0,5)]$$

Convertimos las puntuaciones en típicas:

$$P\left(\frac{(12-0,5)-\mu}{\sigma} \leq \frac{x-\mu}{\sigma} \leq \frac{(12+0,5)-\mu}{\sigma}\right)$$

y quedaría:

$$P\left(\frac{(12-0,5)-\mu}{\sigma} \leq z \leq \frac{(12+0,5)-\mu}{\sigma}\right)$$

dándole valor a μ y σ:

$$P\left(\frac{(12-0,5)-10}{2,24} \leq z \leq \frac{(12+0,5)-10}{2,24}\right) = P(0,67 \leq z \leq 1,12)$$

finalmente utilizamos las tablas de distribución normal:

$$P(0,67 \leq z \leq 1,12) = 0,8686 – 0,7486 = 0,12$$

la aproximación es muy buena; hay una diferencia de una diezmilésima para 20 intentos. A medida que aumenta n (intentos) mejora la aproximación.

Sumar y restar el valor 0,5 se llama ***corrección por continuidad***, permitiendo utilizar las puntuaciones discretas como continuas.

La distribución “CHI CUADRADO” de Pearson:

En la distribución de Chi cuadrado de Pearson una variable X con distribución $X^2_1, X^2_2, ..., X^2_n$ pasa a ser $X = X^2_n$. Su media y varianza valdrán $\mu = n$ y, $\sigma^2 = 2n$.
Esta distribución se usa para contrastar si la distribución de una variable se ajusta a una distribución determinada.
Entre sus propiedades señalamos:

1. Nunca adopta valores menores de 0.
2. Es asimétrica positiva pero a medida que aumentan sus grados de libertad se va aproximando a la distribución normal.
3. Para n > 30 la podemos aproximar a una distribución N(n, 2n).

En la tabla V se hallan algunos valores de las distribuciones X^2.

Ejemplo:

En una variable con 5 grados de libertad, $X \rightarrow X^2_5$, el valor 11,07 deja por debajo de sí una proporción de 0,95, representándose de la siguiente manera: ${}_{0,95}X^2_5 = 11,07$.
Ahora si quisiéramos calcular P (X > 11,07):

$$P(X > 11,07) = 1 – P(X \leq 11,07) = 1 – 0,95 = 0,05$$

La distribución “t” de Student:

Para definir estas distribuciones,al igual que hemos hecho con el Chi Cuadrado, emplearemos otras distribuciones.
Teniendo dos variables X e Y con una distribución (0,1) y X^2_n. La variable aleatoria $T = \frac{X}{\sqrt{Y/n}}$ sigue una distribución “t”con n grados de libertad y se expresa: $T \rightarrow t_n$.

Su media siempre vale 0 y su varianza $\sigma^2 = \frac{n}{n-2}$.

Una distribución “t” es el cociente entre una variable N(0,1) y la raíz cuadrada de X^2_n dividida por sus grados de libertad.

Sus características son:

1. **1.** Es simétrica, con $\mu = 0$. Su forma es muy parecida a la N(0,1), aunque menos apuntada.
2. **2.** Puede tomar cualquier valor ($-\infty \longleftrightarrow +\infty$).
3. **3.** A medida que aumentan los grados de libertad, la distribución se aproxima más a una distribución normal.
4. **4.** La curva es asintótica al eje de abscisas.

Se emplea en estadística inferencial en contrastes. En la tabla VI se muestran los valores de esta distribución.

La distribución de "F" de Snedecor:

Se emplea principalmente en el contraste de hipótesis.

$$F = \frac{X_1 / n_1}{X_2 / n_2}$$

Sigue una distribución F con n_1 y n_2 grados de libertad ($F_{n1,n2}$). Siendo "n_1"los grados del numerador y "n_2" los del denominador; su media y varianza se definen:

$$\mu = \frac{n_2}{n_2 - 2} \text{ para } n_2 > 2;$$

$$\sigma^2 = \frac{2n^2_2\,(n_1+n_2-2)}{n_1\,(n_2-4)(n_2-2)^2} \text{ para } n_2 > 4.$$

Se caracteriza por:

1. **1.** Es asimétrica positiva por lo que nunca toma valores menores que 0.
2. **2.** Si X es variable con distribución F con n_1 y n_2 grados de libertad, la variable Y = 1/X es también una distribución F (propiedad recíproca):

$${}_{1-p}F_{n1,n2} = \frac{1}{{}_{p}F_{n2,n1}};$$

donde p es la propiedad asociada al valor de la variable. Se emplea para calcular algunos percentiles o probabilidades que no aparecen la tabla. En la tabla VII sólo aparece la probabilidad de que X = 0,900; 0,950; 0,975 y 0,990.

Ejemplo:

Una variable X con distribución $F_{5,10}$:

A) Calcular $P(X \leq 3{,}33)$;
buscamos el valor de 5 grados de libertad en el numerador y 10 grados en el denominador en la tabla VII, donde se encuentra el valor3,33. Observamos que se corresponde con una probabilidad de 0,95; por tanto 3,33 se corresponde con el percentil 95.

B) Determinar el valor del percentil 5 de X, es decir: ${}_{0,05}F_{5,10}$;

en este caso haremos uso de la propiedad recíproca:

$$_{0,05}F_{5,10} = \frac{1}{_{1-0,05}F_{10,5}} = \frac{1}{_{0,95}F_{10,5}}$$

En la tabla VII vemos que: $_{0,95}F_{10,5} = 4{,}74$.
Por tanto:

$$_{0,05}F_{5,10} = \frac{1}{_{0,95}F_{10,5}} = \frac{1}{4{,}74} = 0{,}211$$