Tema 5. Algunos Modelos de Distribución de Probabilidad Multidimensionales

Úrsula Torres Parejo

Departamento de Estadística e I.O. Universidad de Granada

Contenido.

- Distribución Multinomial
  - Definición
  - Función generatriz de momentos. Distribuciones marginales y condicionadas
  - Regresión y correlación. Caso bidimensional
  - Reproductividad
- Distribución Normal Bivariante
  - Definición
  - Distribuciones marginales y condicionadas
  - Regresión y correlación. Caso bidimensional
  - Normalidad de las combinaciones lineales de las componentes
  - Función generatriz de momentos

Distribución Multinomial.

Es una generalización de la distribución binomial cuando el experimento aleatorio considerado no tiene sólo dos posibles resultados (éxito o fracaso), sino 3 o más.
Indica el número de veces que aparecen k sucesos excluyentes y no necesariamente exhaustivos con probabilidades $p_1, ..., p_k$ en repeticiones independientes de un experimento.

$$\text{Si } X = (X_1, ..., X_k) \sim M_k(n; p_1, ..., p_k) \Rightarrow$$

$$P[X_1 = x_1, ..., X_k = x_k] =$$

$$= \frac{n!}{x_1! \cdots x_k! (n - \sum_{i=1}^k x_i)!} p_1^{x_1} \cdots p_k^{x_k} \left(1 - \sum_{i=1}^k p_i\right)^{n - \sum_{i=1}^k x_i}$$

$$n \in \mathbb{N}, 0 < p_i < 1, \sum_{i=1}^k p_i \leq 1, \sum_{i=1}^k x_i \leq n,\ x_i \in \{0, ..., n\}$$

$$i = 1, ..., k$$

F.G.M. Distribución marginales y condicionadas.

Función generatriz de momentos

$$M_X(t_1,...,t_k) = \left(p_1 e^{t_1} + ... + p_k e^{t_k} + \left(1 - \sum_{i=1}^{k} p_i\right)\right)^n \forall t_1,...,t_k \in \mathbb{R}$$

Distribuciones marginales

$M_{X_i}(t_i) = M_X(0,...,0,t_i,0,...,0) = (p_i e^{t_i} + (1-p_i))^n \, \forall t_i \in \mathbb{R}$
de donde se deduce que $X_i \sim B(n,p_i),\ i = 1,..,k$.
De forma análoga, la distribución de cualquier subvector $X_{i1},...,X_{il},\ l < k$ es $M_l(n;p_1,...,p_l)$.

Definición.

Sea $(X_1,...,X_k) \sim M_k(n;p_1,...,p_k)$. Para cualquier subvector $X_{i1},...,X_{il}$ con $i_1,...,i_l \in \{1,...,k\}$, $X_{i1},...,X_{il} \sim M_l(n;p_1,...,p_l)$.
En particular, $X_i \sim B(n,p_i), i = 1,...,k$.

F.G.M. Distribución marginales y condicionadas.

Distribuciones condicionadas

Las distribuciones condicionadas de una distribución Multinomial, también se distribuyen según una Multinomial.

$$(X_1, ..., X_l | X_{l+1} = x_{l+1}, ..., X_k = x_k) \sim$$
$$M_l\left(n - \sum_{i=l+1}^{k} x_i; \frac{p_1}{1 - \sum_{i=l+1}^{k} p_i}, ..., \frac{p_l}{1 - \sum_{i=l+1}^{k} p_i}\right)$$

Y en particular:

$$X_i | X_j = x_j \sim B\left(n - x_j; \frac{pi}{1 - p_j}\right), i, j = 1, ..., k \quad i \neq j$$

Regresión y correlación. Caso bidimensional.

La curva de regresión de X_i sobre X_j es:

$$X_i = E[X_i|X_j = x_j] = \frac{np_i}{1-p_j} - \frac{p_i}{1-p_j}X_j$$

que coincide con la recta de regresión.

Las razones de correlación coinciden con el coeficiente de determinación.

$$\eta^2_{X_i|X_j} = \eta^2_{X_j|X_i} = \rho^2_{X_iX_j} = \frac{p_i p_j}{(1-p_i)(1-p_j)}$$

$i, j = 1, ..., k \quad i \neq j$

Reproductividad

La distribución Multinomial es reproductiva respecto al parámetro n:

$X_i \sim M(n_i; p_1, ..., p_k)$ $i = 1, ..., p$ v.a. independientes $\Rightarrow$

$$\sum_{i=1}^{p} X_i \sim M\left(\sum_{i=1}^{p} n_i; p_1, ..., p_k\right)$$

Distribución Normal Bidimensional

Definición

Se dice que (X_1, X_2) se distribuye según una Normal Bidimensional con vector de medias (μ_1, μ_2) y matriz de varianzas-covarianzas $\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{pmatrix}$ si su función de densidad viene dada por:

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \cdot e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right)\right]}$$

con $x_i \in \mathbb{R}$, $\mu_i \in \mathbb{R}, \sigma_i > 0. -1 < \rho < 1, i = 1, 2$

Expresión matricial: $f(x_1, x_2) = \frac{1}{2\pi\sqrt{|\Sigma|}} e^{\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right)}$

Distribuciones marginales y condicionadas

Distribuciones marginales

$$X_1 \sim N(\mu_1, \sigma_1^2), \qquad X_2 \sim N(\mu_2, \sigma_2^2)$$

Distribuciones condicionadas

$$X_1 | X_2 = x_2 \sim N\left(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2 - \mu_2), \sigma_1^2(1-\rho^2)\right)$$

$$X_2 | X_1 = x_1 \sim N\left(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x_1 - \mu_1), \sigma_2^2(1-\rho^2)\right)$$

Regresión y correlación

Las curvas de regresión coinciden con las rectas de regresión y vienen dadas por la media de las distribuciones normales condicionadas:

$$X_1 = E[X_1|X_2 = x_2] = \mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2 - \mu_2)$$

$$X_2 = E[X_2|X_1 = x_1] = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x_1 - \mu_1)$$

Los E.C.M asociados son:

$$ECM[X_1|X_2] = E[Var[X_1|X_2]] = \sigma_1^2(1 - \rho^2)$$

$$ECM[X_2|X_1] = E[Var[X_2|X_1]] = \sigma_2^2(1 - \rho^2)$$

Además:

$$\eta^2_{X_1|X_2} = \eta^2_{X_2|X_1} = \rho^2_{X_1,X_2}$$

Normalidad de las Combinaciones Lineales y Función Generatriz de Momentos

Normalidad de las Combinaciones Lineales

$$X \sim N_2(\mu, \Sigma) \Rightarrow Y = X \cdot A_{2xq} \sim N_q(\mu A, A^T \Sigma A), \quad q = 1, 2$$

Función Generatriz de Momentos

$$M_{X_1,X_2}(t_1, t_2) = e^{t_1\mu_1 + t_2\mu_2 + \frac{t_1^2\sigma_1^2 + t_2^2\sigma_2^2 + 2t_1t_2\rho\sigma_1\sigma_2}{2}} \quad \forall t_1, t_2 \in \mathbb{R}^2$$