TEMA 5.- Algunos modelos de distribución multivariantes

Asignatura: PROBABILIDAD

Grado en Ingeniería Informática y Matemáticas

(3er Curso - 1er semestre)

©Prof. Dr. José Luis Romero Béjar

Departamento de Estadística e Investigación Operativa
Facultad de Ciencias (Despacho XX)

Periodo de docencia: 13/09/2021 a 22/12/2021

Outline

Outline

1. Distribución Multinomial
2. Distribución Normal Bidimensional

Introducción al modelo multinomial

La **distribución multinomial** es una distribución discreta multivariante, que **extiende a la distribución binomial** cuando el **experimento aleatorio tiene más de dos resultados posibles**, no tan solo éxtio o fracaso.

Deducción del modelo probabilístico multinomial

Sean $A_1, A_2, \ldots, A_{k+1}$ los $k+1$ resultados o sucesos exhaustivos y mutuamente excluyentes de un determinado experimento aleatorio, es decir:

$$\cup_{i=1}^{k+1} A_i = \Omega, \quad A_i \cap A_j = \emptyset, \forall i \neq j.$$

Denotamos $p_i = P(A_i), i = 1, 2, \ldots, k+1$.
Entonces $\sum_{i=1}^{k+1} p_i = \sum_{i=1}^{k+1} P[A_i] = P(\cup_{i=1}^{k+1} A_i) = P(\Omega) = 1$, y por tanto $p_{k+1} = 1 - \sum_{i=1}^{k} p_i$.

Supongamos que **se realizan n repeticiones independientes** del experimento en las **mismas condiciones**, de modo que las **probabilidades p_i se mantienen constantes** en todas las repeticiones.
Si consideramos $x_1, x_2, \ldots, x_k$ enteros no negativos tales que $x_1 + x_2 + \cdots + x_k \leq n$, entonces **la probabilidad de que en las n repeticiones ocurra exactamente x_i veces el suceso A_i**, $\forall i = 1, 2, \ldots, k$, y por tanto $x_{k+1} = n - (x_1 + x_2 + \cdots + x_k)$ veces el suceso A_{k+1} es:

$$\frac{n!}{x_1! x_2! \ldots x_k! (n - \sum_{i=1}^{k} x_i)!} p_1^{x_1} p_2^{x_2} \ldots p_k^{x_k} \left(1 - \sum_{i=1}^{k} p_i\right)^{n - \sum_{i=1}^{k} x_i}.$$

Distribución multinomial

Consideramos el vector aleatorio $(X_1, X_2, \ldots, X_{k+1})$ cuyas componentes X_i **cuentan el número de veces que ocurre el suceso** A_i, $\forall i = 1, 2, \ldots, k+1$, de modo que X_{k+1} queda completamente determinado por $x_{k+1} = n - \sum_{i=1}^{k} x_i$.

Entonces se dice que $(X_1, X_2, \ldots, X_k)$ sigue una **distribución multinomial** k-dimensional con parámetros n y $p_1, \ldots p_k$, que **denotamos**,

$$(X_1, \ldots, X_k) \sim M_k(n; p_1, \ldots, p_k),$$

si y sólo si su **función masa de probabilidad** k-dimensional viene dada por:

$$P(X_1 = x_1, \ldots, X_k = x_k) = \frac{n!}{x_1! \cdots x_k!(n - \sum_{i=1}^{k} x_i)!} p_1^{x_1} \cdots p_k^{x_k} \left(1 - \sum_{i=1}^{k} p_i\right)^{n - \sum_{i=1}^{k} x_i},$$

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

Ejercicio voluntario: probar que esta expresión define una verdadera función masa de probabilidad. (Ver solución propuesta por el alumnado en la sección de este tema en PRADO).

Función generatriz de momentos

$$
\begin{aligned}
M_{X_1,\ldots,X_k}(t_1,\ldots,t_k) &= E\left[\exp\left(\sum_{i=1}^{k} t_i X_i\right)\right] \\
&= \left(p_1 e^{t_1} + \cdots + p_k e^{t_k} + \left(1 - \sum_{i=1}^{k} p_i\right)\right)^n, \quad t_1,\ldots,t_k \in \mathbb{R}.
\end{aligned}
$$

Ejercicio voluntario: justificar la expresión anterior para la función generatriz de momentos. (Ver solución propuesta por el alumnado en la sección de este tema en PRADO).

Distribuciones Marginales

Sea $(X_1, \ldots, X_k) \sim M_k(n; p_1, \ldots, p_k)$, entonces para cualquier subvector $(X_{i_1}, \ldots X_{i_l})$, con $i_1, \ldots, i_l \in \{1, \ldots, k\}$, siendo $i_m \neq i_p$, $m \neq p$, $m, p = 1, \ldots, l$, se cumple:

$$(X_{i_1}, \ldots X_{i_l}) \sim M_l(n; p_1, \ldots, p_l), \quad l \in \{1, \ldots, k-1\}.$$

En particular, para las **marginales unidimensionales** se cumple:

$$X_i \sim B(n, p_i), \quad i = 1, \ldots, k.$$

La prueba es directa, sin más que obtener las funciones generatrices de momentos marginales de X_i o conjuntas del subvector $(X_{i_1}, \ldots X_{i_l})$, a partir de la función de generatriz de momentos conjunta anterior.

Por ejemplo, para el **caso unidimensional**:

$$M_{X_i}(t_i) = M_{X_1, \ldots, X_k}(0, \ldots, 0, t_i, 0, \ldots, 0) = \left(p_i e^{t_i} + (1 - p_i)\right)^n, \; t_i \in \mathbb{R}.$$

(Ver la justificación de estos resultados propuesta por el alumnado en la sección de este tema en PRADO).

Distribuciones Condicionadas

Las **distribuciones condicionadas** de una multinomial **siguen también una distribución multinomial**.

Más concretamente, sea $(X_1,\ldots,X_k) \sim M_k(n;p_1,\ldots,p_k)$, entonces:

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

En particular, las **condicionadas unidimensionales** cumplen:

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

(Ver la justificación de estos resultados propuesta por el alumnado en la sección de este tema en PRADO).

Reproductividad

Sean $X_1, \ldots, X_p$, vectores aleatorios **independientes** k-dimensionales, con distribución multinomial con parámetros n_i, $i = 1, \ldots, p$, y $p_1, \ldots, p_k$. Se tiene entonces

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

Ejercicio voluntario: obtener la función de generatriz de momentos de $\sum_{i=1}^{p} X_i$ para deducir la distribución anterior.

Vector de medias y covarianzas

Sea $X = (X_1, \ldots, X_k)$, vector aleatorio k-dimensional, con distribución multinomial con parámetros n, y $p_1, \ldots, p_k$. Se tiene entonces:

i). $E[X] = (np_1, np_2, \ldots, np_k)$

ii). $Cov(X_i, X_j) = -np_i p_j$, $\forall i \neq j$

Demostración:

i). $E[X] = (np_1, np_2, \ldots, np_k)$ porque las distribuciones marginales son binomiales de parámetros n y p_i, $\forall i = 1, 2, \ldots, k$.

ii). $Cov(X_i, X_j) = E[X_i X_j] - E[X_i]E[X_j]$, $\forall i \neq j$

$$\begin{aligned} E[X_i X_j] &= \frac{\partial^2 M_X(t_1, \ldots, t_k))}{\partial t_i \partial t_j}|_{(t_1,\ldots,t_k)=(0,\ldots,0)} \\ &= n(n-1)\left(p_1 e^{t_1} + \cdots + p_k e^{t_k} + \left(1 - \sum_{i=1}^{k} p_i\right)\right)^{n-2} p_i p_j e^{t_i} e^{t_j} |(t_1, \ldots, t_k) = (0, \ldots, 0) \\ &= n(n-1) p_i p_j \end{aligned}$$

Por otro lado,

$E[X_i] = np_i$ y $E[X_j] = np_j$ porque las marginales tienen distribuciones binomiales.

Por tanto,

$$Cov(X_i, X_j) = E[X_i X_j] - E[X_i]E[X_j] = n(n-1)p_i p_j - n^2 p_i p_j = -np_i p_j$$

Regresión y correlación para el caso bidimensional

- La **curva (recta) de regresión** de X_i/X_j, viene dada por

$$x_i = \frac{np_i}{1-p_j} - \frac{p_i}{1-p_j}x_j.$$

- Las **razones de correlación**, que **coinciden con el coeficiente de determinación**, vienen dadas por:

$$\eta^2_{X_i/X_j} = \eta^2_{X_j/X_i} = \rho^2_{X_iX_j} = \frac{p_ip_j}{(1-p_i)(1-p_j)}, \quad i \neq j,\ i,j \in \{1,\ldots,k\}.$$

- El **coeficiente de correlación lineal** viene dado por

$$\rho_{X_iX_j} = -\sqrt{\frac{p_ip_j}{(1-p_i)(1-p_j)}}, \quad i \neq j,\ i,j \in \{1,\ldots,k\}.$$

Outline

Distribución Normal Bidimensional

Se dice que (X_1, X_2) se distribuye según una **normal bidimensional** con **vector de medias** $\boldsymbol{\mu} = (\mu_1, \mu_2)$ y con **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 de probabilidad** viene dada por:

$$\begin{aligned} f_{(X_1,X_2)}(x_1, x_2) &= \frac{1}{2\pi\left[\det(\Sigma)\right]^{1/2}} e^{\left(-\frac{1}{2}(\mathrm{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathrm{x}-\boldsymbol{\mu})\right)} \\ &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}-2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right)\right]\right)}, \end{aligned}$$

donde $\mathrm{x} = (x_1, x_2) \in \mathbb{R}^2$, σ_i^2 es la varianza de X_i, $i = 1, 2$, y ρ es el coeficiente de correlación lineal de X e Y.

Marginales y condicionadas

Ambas distribuciones, **marginales y condicionadas**, son **normales**.

- En el caso de las **marginales**, se tiene

$$X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2), \quad X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2).$$

- Para las **condicionadas** se obtienen las siguientes distribuciones normales:

$$X_1/X_2 = x_2 \sim \mathcal{N}\left(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2 - \mu_2), \sigma_1^2(1-\rho^2)\right)$$
$$X_2/X_1 = x_1 \sim \mathcal{N}\left(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x_1 - \mu_1), \sigma_2^2(1-\rho^2)\right)$$

(Ver la justificación de estos resultados propuesta por el alumnado en la sección de este tema en PRADO).

Regresión y correlación

- La **curva de regresión**, que **coincide con la recta de regresión** de X_i/X_j, para $i,j = 1,2,\ i \neq j$, viene dada por la **media de las distribuciones normales condicionadas**, anteriormente especificadas.

  Más concretamente, viene dada por

  $$x_i = \mu_i + \rho\frac{\sigma_i}{\sigma_j}(x_j - \mu_j), \quad i,j = 1,2,\ i \neq j.$$

- Equivalentemente, las **razones de correlación coinciden con el coeficiente de determinación** ρ^2, ya que las curvas de regresión coinciden con las rectas de regresión.
- El **E.C.M. asociado a la curva de regresión** de X_i sobre X_j viene dado por $\sigma_i^2(1-\rho^2)$, $i = 1,2$, para cada $j = 1,2$, con $i \neq j$.

Independencia $\Leftrightarrow$ incorrelación $\Leftrightarrow \Sigma$ Diagonal

Observamos que

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{pmatrix},$$

matriz de covarianzas de una normal bidimensional, es no singular $(\rho^2 \neq 1)$ y semidefinida positiva.

Es fácil comprobar que bajo el supuesto de normalidad bivariante:

$$\textbf{Independencia} \;\Leftrightarrow\; \textbf{Incorrelación} \;\Leftrightarrow \Sigma \textbf{ Diagonal}$$

De hecho la segunda equivalencia es evidente, y la primera es debida a que

$$\rho = 0 \Leftrightarrow f(x,y) = f_X(x) f_Y(y).$$

Importante: esta equivalencia no es válida si cada variable tiene distribución normal unidimensional pero la conjunta no.

Teorema para la función generatriz de momentos de la normal bivariante

Para $\mathbf{t} = (t_1, t_2) \in \mathbb{R}^2$,

$$\begin{aligned} M_{X_1,X_2}(t_1,t_2) &= M_{X_1,X_2}(\mathbf{t}) = E\left[\exp\left(\langle \mathbf{t}, \mathbf{X}\rangle\right)\right] \\ &= \exp\left(\langle \boldsymbol{\mu}, \mathbf{t}\rangle + \frac{\mathbf{t}\Sigma\mathbf{t}^T}{2}\right), \\ &= \exp\left(t_1\mu_1 + t_2\mu_2 + \frac{t_1^2\sigma_1^2 + t_2^2\sigma_2^2 + 2t_1t_2\rho\sigma_1\sigma_2}{2}\right). \end{aligned}$$

Teorema para la función generatriz de momentos de la normal bivariante (demostración)

$$M_{X_1,X_2}(t_1,t_2) = E\left[e^{(\langle \mathbf{t},\mathbf{X}\rangle)}\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{(t_1x_1+t_2x_2)} f_{X_1,X_2}(x_1,x_2)dx_1dx_2$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{(t_1x_1+t_2x_2)} f_{X_1/X_2=x_2}(x_1)f_{X_2}(x_2)dx_1dx_2$$

$$= \int_{-\infty}^{\infty} e^{(t_2x_2)} f_{X_2}(x_2)\left[\int_{-\infty}^{\infty} e^{(t_1x_1)} f_{X_1/X_2=x_2}(x_1)dx_1\right]dx_2 = \int_{-\infty}^{\infty} e^{(t_2x_2)} M_{X_1/X_2=x_2}(t_1)f_{X_2}(x_2)dx_2$$

$$= \int_{-\infty}^{\infty} e^{(t_2x_2)} e^{\left(t_1\left(\mu_1+\rho\frac{\sigma_1}{\sigma_2}(x_2-\mu_2)\right)+\frac{t_1^2\sigma_1^2(1-\rho^2)}{2}\right)} f_{X_2}(x_2)dx_2$$

$$= e^{\left(t_1\left(\mu_1-\rho\frac{\sigma_1}{\sigma_2}\mu_2\right)+\frac{t_1^2\sigma_1^2(1-\rho^2)}{2}\right)} \int_{-\infty}^{\infty} e^{\left(t_2x_2+t_1\rho\frac{\sigma_1}{\sigma_2}x_2\right)} f_{X_2}(x_2)dx_2$$

$$= e^{\left(t_1\left(\mu_1-\rho\frac{\sigma_1}{\sigma_2}\mu_2\right)+\frac{t_1^2\sigma_1^2(1-\rho^2)}{2}\right)} M_{X_2}\left(t_2+t_1\rho\frac{\sigma_1}{\sigma_2}\right)$$

$$= e^{\left(t_1\left(\mu_1-\rho\frac{\sigma_1}{\sigma_2}\mu_2\right)+\frac{t_1^2\sigma_1^2(1-\rho^2)}{2}\right)} e^{\left(\left(t_2+t_1\rho\frac{\sigma_1}{\sigma_2}\right)\mu_2+\frac{\left(t_2+t_1\rho\frac{\sigma_1}{\sigma_2}\right)^2\sigma_2^2}{2}\right)}$$

$$= e^{\left(t_1\mu_1+t_2\mu_2+\frac{t_1^2\sigma_1^2+t_2^2\sigma_2^2+2t_1t_2\rho\sigma_1\sigma_2}{2}\right)}, \quad t_1,t_2\in\mathbb{R}.$$

Teorema para la función generatriz de momentos de la normal bivariante (demostración)

En la derivación de la función generatriz de momentos, se ha aplicado la definición de la densidad de probabilidad marginal de $X_1/X_2 = x_2$, en términos de la conjunta y condicionada, es decir,

$$f_{X_1/X_2=x_2}(x_1) = \frac{f_{X_1,X_2}(x_1,x_2)}{f_{X_2}(x_2)} \Rightarrow f_{X_1,X_2}(x_1,x_2) = f_{X_1/X_2=x_2}(x_1)f_{X_2}(x_2). \quad (1)$$

La ecuación (1) se puede verificar directamente a partir de la expresión de la densidad de probabilidad conjunta de la normal bidimensional. Más concretamente se obtiene, operando en el argumento de la exponencial que define $f_{X_1,X_2}(x_1,x_2)$ como sigue:

$$f_{X_1,X_2}(x_1,x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}-2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right)\right]\right)}$$

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

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

Normalidad para combinaciones lineales de las componentes

Sea $\mathrm{X} \sim \mathcal{N}_2(\boldsymbol{\mu}, \Sigma)$. Se considera $\mathrm{A}_{2\times q}$ una matriz de rango máximo q, con $q = 1, 2$. Entonces:

$$\mathrm{Y} = \mathrm{X}\mathrm{A}_{2\times q} \sim \mathcal{N}_q\left(\boldsymbol{\mu}\mathrm{A}, \mathrm{A}^T\Sigma\mathrm{A}\right).$$

Demostración:

La demostración se obtiene de forma directa a partir del teorema anterior.

Sea $t \in \mathbb{R}^q$, entonces:

$$E\left[e^{tY^T}\right] = E\left[e^{(tA^T)X^T}\right] = M_X(tA^T) = e^{(tA^T)\mu^T + \frac{(tA^T)\Sigma(tA^T)^T}{2}} = e^{t(\mu A)^T + \frac{t(A^T\Sigma A)t^T}{2}}$$

Dado que A es de rango máximo, $|A^T\sigma A| \neq 0$, se deduce que $Y = XA \sim \mathcal{N}_q(\mu A, A^T\Sigma A)$.

Observación:

Se tiene que $Y = XA_{2x2} = (a_{11}X_1 + a_{12}X_2, a_{21}X_1 + a_{22}X_2)$. De modo que **cualquier vector bidimensional cuyas componentes sean combinación lineal** de X_1 y X_2, componentes de un vector normal bidimensional, de modo que la **matriz** A que define esta combinaciones sea **no singular**, tiene una distribución **normal bidimiensional**.

Corolario

Cualquier **combinación lineal de las componentes de un vector normal bidimensional** tiene **distribución normal unidimensional**.

Sea X $= (X_1, X_2) \sim \mathcal{N}_2 \left(\mu = (\mu_1, \mu_2), \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{pmatrix} \right)$, entonces:

$$a_1X_1 + a_2X_2 = (X_1, X_2) \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}_{2x1} \sim \mathcal{N}_1 \left((\mu_1, \mu_2) \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}, (a_1, a_2) \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \right).$$

Y por tanto,

$$a_1X_1 + a_2X_2 \sim \mathcal{N}_1 \left(a_1\mu_1 + a_2\mu_2, a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + 2\rho\sigma_1\sigma_2 a_1 a_2 \right)$$

siempre que $rg(A) = 1, (a_1$ ó $a_2 \neq 0)$.