Probability (Level M) [STATS 5024]

Lecture 17

The Multinomial and Multivariate Normal distributions

Multinomial distribution I

- The **multinomial distribution** is the generalisation to an arbitrary number of dimensions of the Binomial distribution
- Suppose that n objects are each independently to be placed in one of $p+1$ different categories, each object having probability θ_i of being placed in the ith category $(i = 1, \ldots, p+1)$
- This means that $0 \leqslant \theta_i \leqslant 1$ and that $\theta_1 + \cdots + \theta_{p+1} = 1$

Multinomial distribution II

- Let the random variable X_i denote the total number of objects placed in the ith category $(i = 1, \ldots, p+1)$
- Notice that $X_1 + X_2 + \cdots + X_{p+1} = n$
- This means that only p of the random variables need to be considered explicitly, say $X_1, \ldots, X_p$, since the value of X_{p+1} may be deduced exactly from the values of the other random variables
- The random vector $\boldsymbol{X} = (X_1, \ldots, X_p)$ is said to follow a Multinomial distribution, often written

$$\boldsymbol{X} \sim \mathrm{Mu}(n, \theta_1, \ldots, \theta_p)$$

- Note the restriction $\theta_1 + \cdots + \theta_p \leqslant 1$

Mass function of Multinomial distribution

$\boldsymbol{X}$ has joint range space

$$R_{\boldsymbol{X}} = \{(x_1, \ldots, x_p) : x_1, \ldots, x_p = 0, 1, \ldots, n; x_1 + \cdots + x_p \leqslant n\}.$$

$\boldsymbol{X}$ has joint probability mass function

$$p_{\boldsymbol{X}}(x_1, \ldots, x_p) = \frac{n!}{x_1! \cdots x_p!(n - x_1 - \cdots - x_p)!} \times \theta_1^{x_1} \cdots \theta_p^{x_p} (1 - \theta_1 - \cdots - \theta_p)^{n - x_1 - \cdots - x_p}$$

$$(x_1, \ldots, x_p) \in R_{\boldsymbol{X}}$$

The binomial distribution is the special case of the multinomial when $p = 1$, so $\text{Bi}(n, \theta)$ is the same as $\text{Mu}(n, \theta)$

Marginal mass function of Multinomial

- Marginal probability mass functions can be obtained recursively
- We first find the marginal distribution of $X_1, \ldots, X_{p-1}$, by summing out X_p
- Notice that $(X_1, \ldots, X_{p-1})$ has the joint range space

$$\{(x_1, \ldots, x_{p-1}) : x_1, \ldots, x_{p-1} = 0, 1, \ldots, n; x_1 + \cdots + x_{p-1} \leqslant n\}$$

- (Algebraic details of summing out X_p in notes)
- The marginal distribution of $(X_1, \ldots, X_{p-1})$ is $\mathrm{Mu}(n, \theta_1, \ldots, \theta_{p-1})$

Intuition

- We began with $p + 1$ categories of objects and explicitly modelled the counts in categories $1, 2, \ldots, p$
- Category $p + 1$ may be considered as a 'miscellaneous' category which is only considered implicitly
- When we restrict our attention to the marginal distribution of $X_1, \ldots, X_{p-1}$, we are implicitly combining categories p and $p + 1$ into a new 'miscellaneous' category.

Recursion

- We have established that $(X_1, \ldots, X_{p-1})$ follows a $\text{Mu}(n, \theta_1, \ldots, \theta_{p-1})$ distribution

Recursion

- We have established that $(X_1, \ldots, X_{p-1})$ follows a $\mathrm{Mu}(n, \theta_1, \ldots, \theta_{p-1})$ distribution
- Hence $(X_1, \ldots, X_{p-2})$ follows a $\mathrm{Mu}(n, \theta_1, \ldots, \theta_{p-2})$

Recursion

- We have established that $(X_1, \ldots, X_{p-1})$ follows a $\text{Mu}(n, \theta_1, \ldots, \theta_{p-1})$ distribution
- Hence $(X_1, \ldots, X_{p-2})$ follows a $\text{Mu}(n, \theta_1, \ldots, \theta_{p-2})$
- $\vdots$
- (X_1, X_2) marginally follows a $\text{Mu}(n, \theta_1, \theta_2)$ distribution

Recursion

- We have established that $(X_1, \ldots, X_{p-1})$ follows a $\mathrm{Mu}(n, \theta_1, \ldots, \theta_{p-1})$ distribution
- Hence $(X_1, \ldots, X_{p-2})$ follows a $\mathrm{Mu}(n, \theta_1, \ldots, \theta_{p-2})$
- $\vdots$
- (X_1, X_2) marginally follows a $\mathrm{Mu}(n, \theta_1, \theta_2)$ distribution
- X_1 marginally follows a $\mathrm{Mu}(n, \theta_1)[= \mathrm{Bi}(n, \theta_1)]$ distribution

Recursion

- We have established that $(X_1, \ldots, X_{p-1})$ follows a $\mathrm{Mu}(n, \theta_1, \ldots, \theta_{p-1})$ distribution
- Hence $(X_1, \ldots, X_{p-2})$ follows a $\mathrm{Mu}(n, \theta_1, \ldots, \theta_{p-2})$
- $\vdots$
- (X_1, X_2) marginally follows a $\mathrm{Mu}(n, \theta_1, \theta_2)$ distribution
- X_1 marginally follows a $\mathrm{Mu}(n, \theta_1)[= \mathrm{Bi}(n, \theta_1)]$ distribution
- Symmetry of the joint distribution implies that equivalent results hold for all combinations of the X_js
- For example, $(X_i, X_j) \sim \mathrm{Mu}(n, \theta_i, \theta_j) \quad (X_i \neq X_j)$
- and $X_i \sim \mathrm{Bi}(n, \theta_i)$

Expectation vector and covariance matrix

Solution to Tutorial Problem shows that:

$$\mathrm{E}(\boldsymbol{X}) = \begin{bmatrix} n\theta_1 \\ n\theta_2 \\ \vdots \\ n\theta_p \end{bmatrix}$$

$$\mathrm{Cov}(\boldsymbol{X}) = \begin{bmatrix} n\theta_1(1-\theta_1) & -n\theta_1\theta_2 & \cdots & -n\theta_1\theta_p \\ -n\theta_1\theta_2 & n\theta_2(1-\theta_2) & \cdots & -n\theta_2\theta_p \\ \vdots & \vdots & \ddots & \vdots \\ -n\theta_1\theta_p & -n\theta_2\theta_p & \cdots & n\theta_p(1-\theta_p) \end{bmatrix}$$

The Multivariate Normal distribution

Example 7

Consider the random vector $\boldsymbol{X} = (X_1, \ldots, X_p)$ where $X_1, \ldots, X_p$ are independent random variables and each $X_i \sim \mathrm{N}(\mu_i, \sigma_i^2)$

Because of the independence of the random variables, the joint probability density function of $\boldsymbol{X}$ is:

$$
\begin{aligned}
f_{\boldsymbol{X}}(\boldsymbol{x}) &= \prod_{i=1}^{p} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left[-\frac{1}{2}\left(\frac{x_i - \mu_i}{\sigma_i}\right)^2\right] \\
&= (2\pi)^{-p/2} \frac{1}{\sqrt{\prod_{i=1}^{p} \sigma_i^2}} \exp\left[-\frac{1}{2}\sum_{i=1}^{p}\left(\frac{x_i - \mu_i}{\sigma_i}\right)^2\right]
\end{aligned}
$$

Example 7

- Mean vector of $\boldsymbol{X}$ is $\mathrm{E}(\boldsymbol{X}) = (\mu_1, \ldots, \mu_p) \equiv \boldsymbol{\mu}$
- Covariance matrix of $\boldsymbol{X}$ is $\mathrm{Cov}(\boldsymbol{X}) = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_p^2) \equiv \Sigma$
- This means that

$$\sum_{i=1}^{p} \left(\frac{x_i - \mu_i}{\sigma_i} \right)^2 = (\boldsymbol{x} - \boldsymbol{\mu})^{\mathrm{T}} \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\mu})$$

$$\prod_{i=1}^{p} \sigma_i^2 = \det(\Sigma)$$

- So the (joint) probability density function of $\boldsymbol{X}$ can be written in the form:

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = (2\pi)^{-p/2} [\det(\Sigma)]^{-1/2} \exp\left[-\frac{1}{2} (\boldsymbol{x} - \boldsymbol{\mu})^{\mathrm{T}} \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) \right]$$

Multivariate Normal distribution

Definition

Suppose that a p-dimensional random vector, $\boldsymbol{X}$, has joint probability density function

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = (2\pi)^{-p/2}[\det(\Sigma)]^{-1/2} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right],$$

where Σ is a positive definite $p \times p$ matrix. Then $\boldsymbol{X}$ is said to follow a (non-singular) **Multivariate Normal (MVN) distribution**, sometimes written

$$\boldsymbol{X} \sim \mathrm{N}_p(\boldsymbol{\mu}, \Sigma)$$

Multivariate Normal distribution

- In fact $\boldsymbol{\mu} = \mathrm{E}(\boldsymbol{X})$ and $\Sigma = \mathrm{Cov}(\boldsymbol{X})$
- This explains the restriction of Σ to positive definite matrices in the above definition
- But there is no general requirement for Σ to be diagonal, as it was in Example 7

Proposition

Proposition 3.7

p-dimensional random vector $\boldsymbol{X}$ has a MVN distribution

$$\Leftrightarrow$$

$\boldsymbol{a}^T\boldsymbol{X}$ has a (univariate) Normal distribution, for *every* *p*-dimensional vector $\boldsymbol{a}$

The proof is omitted.

- So, when $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$ it follows that each X_i is marginally distributed as a $\mathrm{N}(\mu_i, \sigma_{ii})$ random variable (where σ_{ii} is the ith diagonal element of Σ)
- But, even when every X_i is marginally normally-distributed, it does not necessarily follow that $\boldsymbol{X}$ is Multivariate Normal

Proposition

Proposition 3.8

Suppose that the p-dimensional random vector $\boldsymbol{X}$ follows the $\mathrm{N}_p(\boldsymbol{\mu}, \Sigma)$ distribution.
If A is a $q \times p$ matrix of constants, and $\boldsymbol{b}$ is a q-dimensional vector of constants, then

$$A\boldsymbol{X} + \boldsymbol{b} \sim \mathrm{N}_q(A\boldsymbol{\mu} + \boldsymbol{b}, A\Sigma A^{\mathrm{T}}).$$

If $\boldsymbol{X}$ follows a non-singular MVN distribution, then the distribution of $A\boldsymbol{X} + \boldsymbol{b}$ is also non-singular if and only if A has rank q. (This requires in particular that $q \leqslant p$.)

Sub-vectors

Corollary

If $\boldsymbol{X} \sim \mathrm{N}_p(\boldsymbol{\mu}, \Sigma)$, then any sub-vector of $\boldsymbol{X}$ has a (joint) marginal distribution that is also MVN.

If $\boldsymbol{X}$ has a non-singular distribution, then so too has any sub-vector of $\boldsymbol{X}$

Proof I

We can permute the elements of $\boldsymbol{X}$ as we please and still obtain a MVN distribution, so we shall assume that the sub-vector whose distribution we wish to find is $\boldsymbol{X}^{(1)}$, consisting of the first r $(1 \leqslant r \leqslant p-1)$ elements of $\boldsymbol{X}$
We partition $\boldsymbol{X}$, μ and Σ conformably as follows:

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}^{(1)} \\ \boldsymbol{X}^{(2)} \end{bmatrix} \qquad \mu = \begin{bmatrix} \mu^{(1)} \\ \mu^{(2)} \end{bmatrix} \qquad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

Proof II

Let $B = [I_r, 0_{r,p-r}]$, where I_r is the $r \times r$ identity matrix and $0_{r,p-r}$ is the $r \times (p-r)$ matrix of zeros. Then B is an $r \times p$ matrix of rank r. Also,

$$B\boldsymbol{X} = \boldsymbol{X}^{(1)} \qquad \text{and} \qquad B\boldsymbol{\mu} = \boldsymbol{\mu}^{(1)} \qquad \text{and} \qquad B\Sigma B^{\mathrm{T}} = \Sigma_{11}.$$

Then, by Proposition 3.8 (with $A = B$ and $\boldsymbol{b} = \mathbf{0}$),

$$\boldsymbol{X}^{(1)} \sim \mathrm{N}_r(\boldsymbol{\mu}^{(1)}, \Sigma_{11})$$

If Σ is non-singular (i.e., positive definite), so too is $B\Sigma B^{\mathrm{T}} = \Sigma_{11}$.

Conditional distribution

Proposition 3.9

Suppose that the random vector $\boldsymbol{X}$ follows the $\mathrm{N}_p(\boldsymbol{\mu}, \Sigma)$ distribution, and partition $\boldsymbol{X}$ as $\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}^{(1)} \\ \boldsymbol{X}^{(2)} \end{bmatrix}$, where as before $\boldsymbol{X}^{(1)}$ consists of the first r $(1 \leqslant r \leqslant p-1)$ elements of $\boldsymbol{X}$. Then,

- the conditional distribution of $\boldsymbol{X}^{(2)}$ given $\boldsymbol{X}^{(1)} = \boldsymbol{x}^{(1)}$ is

$$\mathrm{N}_{p-r}\Big(\boldsymbol{\mu}^{(2)} + \Sigma_{12}^{\mathrm{T}}\Sigma_{11}^{-1}(\boldsymbol{x}^{(1)} - \boldsymbol{\mu}^{(1)}), \Sigma_{22} - \Sigma_{12}^{\mathrm{T}}\Sigma_{11}^{-1}\Sigma_{12}\Big)$$

- the conditional distribution of $\boldsymbol{X}^{(1)}$ given $\boldsymbol{X}^{(2)} = \boldsymbol{x}^{(2)}$ is

$$\mathrm{N}_{r}\Big(\boldsymbol{\mu}^{(1)} + \Sigma_{12}\Sigma_{22}^{-1}(\boldsymbol{x}^{(2)} - \boldsymbol{\mu}^{(2)}), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^{\mathrm{T}}\Big)$$

Correlation and independence

- If $\Sigma_{12} = 0_{r,p-r}$, then the conditional distribution of $\boldsymbol{X}^{(2)}$ given $\boldsymbol{X}^{(1)} = \boldsymbol{x}^{(1)}$ is the same as its marginal distribution for every choice of $\boldsymbol{x}^{(1)}$
- So $\boldsymbol{X}^{(1)}$ and $\boldsymbol{X}^{(2)}$ are independent random vectors
- It follows that, when $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are jointly normally distributed random variables, then they are independent if and only if they are uncorrelated
- (Note: this is *not* true of random variables in general)

Example 8

Suppose that the continuous random vector

$$\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathrm{N}_2\left(\begin{bmatrix} 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 4 & -1 \\ -1 & 9 \end{bmatrix}\right)$$

(a) Identify the marginal distributions of X_1 and X_2 Write down $\mathrm{E}(X_1)$, $\mathrm{Var}(X_1)$, $\mathrm{E}(X_2)$, $\mathrm{Var}(X_2)$, $\mathrm{Cov}(X_1, X_2)$ and $\rho(X_1, X_2)$.

(b) Let $Y_1 = \frac{1}{2}X_1$ and $Y_2 = X_1 + 4X_2$

  (i) Find $\mathrm{Cov}(\boldsymbol{Y})$

  (ii) What is the distribution of Y_2?

  (iii) Are Y_1 and Y_2 independent? Justify your answer

Solution (a)

$$\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathrm{N}_2\left(\begin{bmatrix} 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 4 & -1 \\ -1 & 9 \end{bmatrix} \right)$$

Solution (a)

$$\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathrm{N}_2\left(\begin{bmatrix} 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 4 & -1 \\ -1 & 9 \end{bmatrix}\right)$$

$$\mathrm{E}(X_1) = 2 \qquad \mathrm{Var}(X_1) = 4 \quad X_1 \sim \mathrm{N}(2, 4) \qquad \text{(Prop. 3.7)}$$

Solution (a)

$$\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathrm{N}_2\left(\begin{bmatrix} 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 4 & -1 \\ -1 & 9 \end{bmatrix} \right)$$

$$\begin{aligned} &\mathrm{E}(X_1) = 2 \qquad \mathrm{Var}(X_1) = 4 \quad X_1 \sim \mathrm{N}(2,4) \qquad \text{(Prop. 3.7)} \\ &\mathrm{E}(X_2) = 2 \qquad \mathrm{Var}(X_2) = 9 \quad X_2 \sim \mathrm{N}(2,9) \qquad \text{(Prop. 3.7)} \end{aligned}$$

Solution (a)

$$\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathrm{N}_2\left(\begin{bmatrix} 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 4 & -1 \\ -1 & 9 \end{bmatrix} \right)$$

$$\begin{aligned} \mathrm{E}(X_1) &= 2 \qquad \mathrm{Var}(X_1) = 4 \quad X_1 \sim \mathrm{N}(2,4) \qquad \text{(Prop. 3.7)} \\ \mathrm{E}(X_2) &= 2 \qquad \mathrm{Var}(X_2) = 9 \quad X_2 \sim \mathrm{N}(2,9) \qquad \text{(Prop. 3.7)} \\ \mathrm{Cov}(X_1, X_2) &= -1 \end{aligned}$$

Solution (a)

$$\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathrm{N}_2\left(\begin{bmatrix} 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 4 & -1 \\ -1 & 9 \end{bmatrix} \right)$$

$$\begin{aligned}
\mathrm{E}(X_1) &= 2 \qquad \mathrm{Var}(X_1) = 4 \quad X_1 \sim \mathrm{N}(2,4) \qquad \text{(Prop. 3.7)} \\
\mathrm{E}(X_2) &= 2 \qquad \mathrm{Var}(X_2) = 9 \quad X_2 \sim \mathrm{N}(2,9) \qquad \text{(Prop. 3.7)} \\
\mathrm{Cov}(X_1, X_2) &= -1
\end{aligned}$$

$$\rho_{12} = \frac{\mathrm{Cov}(X_1, X_2)}{\sqrt{\mathrm{Var}(X_1)\mathrm{Var}(X_2)}} = \frac{-1}{\sqrt{4 \times 9}} = -\frac{1}{6}$$

Solution (b) (i)

$$\boldsymbol{Y} = A\boldsymbol{X} + \boldsymbol{b} \qquad A = \begin{bmatrix} \frac{1}{2} & 0 \\ 1 & 4 \end{bmatrix} \qquad b = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Solution (b) (i)

$$\boldsymbol{Y} = A\boldsymbol{X} + \boldsymbol{b} \qquad A = \begin{bmatrix} \frac{1}{2} & 0 \\ 1 & 4 \end{bmatrix} \qquad b = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Use Prop. 3.8: $\boldsymbol{Y} = A\boldsymbol{X} + \boldsymbol{b} \sim \mathrm{N}_q(A\boldsymbol{\mu} + \boldsymbol{b}, A\Sigma A^{\mathrm{T}})$

$$\mathrm{E}(\boldsymbol{Y}) = A\boldsymbol{\mu} + \boldsymbol{b} = \begin{bmatrix} \frac{1}{2} & 0 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 10 \end{bmatrix}$$

Solution (b) (i)

$$\boldsymbol{Y} = A\boldsymbol{X} + \boldsymbol{b} \qquad A = \begin{bmatrix} \frac{1}{2} & 0 \\ 1 & 4 \end{bmatrix} \qquad b = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Use Prop. 3.8: $\boldsymbol{Y} = A\boldsymbol{X} + \boldsymbol{b} \sim \mathrm{N}_q(A\boldsymbol{\mu} + \boldsymbol{b}, A\Sigma A^{\mathrm{T}})$

$$\mathrm{E}(\boldsymbol{Y}) = A\boldsymbol{\mu} + \boldsymbol{b} = \begin{bmatrix} \frac{1}{2} & 0 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 10 \end{bmatrix}$$

$$\begin{aligned} \mathrm{Cov}(\boldsymbol{Y}) = A\Sigma A^{\mathrm{T}} &= \begin{bmatrix} \frac{1}{2} & 0 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} 4 & -1 \\ -1 & 9 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & 1 \\ 0 & 4 \end{bmatrix} \\ &= \begin{bmatrix} 2 & -\frac{1}{2} \\ 0 & 35 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & 1 \\ 0 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 140 \end{bmatrix} \end{aligned}$$

Solution (b) (ii) and (iii)

(ii)

$$Y_2 \sim \mathrm{N}(10, 140) \qquad \text{(Prop. 3.7)}$$

(iii) $\mathrm{Cov}(Y_1, Y_2) = 0$, so $\rho_{12} = 0$ and Y_1 and Y_2 are independent (since uncorrelated $\implies$ independent for MVN random variables)