MATH 556: MATHEMATICAL STATISTICS I

MULTIVARIATE PROBABILITY DISTRIBUTIONS

1. **The Multinomial Distribution**

The *multinomial distribution* is a multivariate generalization of the binomial distribution. The binomial distribution can be derived from an "infinite urn" model with two types of objects being sampled without replacement. Suppose that the proportion of "Type 1" objects in the urn is θ (so $0 \leq \theta \leq 1$) and hence the proportion of "Type 2" objects in the urn is $1-\theta$. If n objects are sampled, and X is the random variable corresponding to the number of "Type 1" objects in the sample. Then $X \sim Bin(n, \theta)$.

Now consider a generalization; suppose that the urn contains $d+1$ types of objects ($d = 1, 2, \ldots$), with θ_i being the proportion of Type i objects, for $i = 1, \ldots, d+1$. Let X_i be the random variable corresponding to the number of type i objects in a sample of size n, for $i = 1, \ldots, d$. Then the joint pmf of vector $\mathbf{X} = (X_1, \ldots, X_d)^\top$ is given by

$$f_{X_1,\ldots,X_d}(x_1, \ldots, x_d) = \frac{n!}{x_1! \ldots, x_d! x_{d+1}!} \theta_1^{x_1} \ldots . \theta_d^{x_d} \theta_{d+1}^{x_{d+1}} = \frac{n!}{x_1! \ldots, x_d! x_{d+1}!} \prod_{i=1}^{d+1} \theta_i^{x_i}$$

where $0 \leq \theta_i \leq 1$ for all i, and $\theta_1 + \cdots + \theta_d + \theta_{d+1} = 1$, and where x_{d+1} is defined by

$$x_{d+1} = n - (x_1 + \ldots + x_d).$$

This is the joint pmf for the **multinomial distribution**. We write

$$\mathbf{X} \sim Multinomial(n, \theta_1, \ldots, \theta_d).$$

2. **The Dirichlet Distribution**

The *Dirichlet distribution* is a multivariate generalization of the Beta distribution that gives a distribution for random variables on a simplex. The joint pdf of vector $\mathbf{X} = (X_1, \ldots, X_d)^\top$ where $0 \leq X_i \leq 1$ for $i = 1, \ldots, d$, and

$$0 \leq \sum_{i=1}^{d} X_i \leq 1.$$

is given by

$$f_{X_1,\ldots,X_d}(x_1, \ldots, x_d) = \frac{\Gamma(\alpha)}{\Gamma(\alpha_1) \ldots \Gamma(\alpha_d) \Gamma(\alpha_{d+1})} x_1^{\alpha_1 - 1} \ldots, x_d^{\alpha_d - 1} {x_{d+1}}^{\alpha_{d+1} - 1}$$

for $0 \leq x_i \leq 1$ for all i such that $x_1 + \cdots + x_d + x_{d+1} = 1$, where $\alpha = \alpha_1 + \cdots + \alpha_{d+1}$ and where x_{d+1} is defined by

$$x_{d+1} = 1 - (x_1 + \cdots + x_d).$$

This is the density function which reduces to the Beta distribution if $d = 1$. It can also be shown that the marginal distribution of X_i is $Beta(\alpha_i, \alpha - \alpha_i)$. We write

$$\mathbf{X} \sim Dirichlet(\alpha_1, \ldots, \alpha_{d+1}).$$

The Dirichlet distribution can be generated by considering independent random variables $Z_1, \ldots, Z_{d+1}$, with $Z_j \sim Gamma(\alpha_j, 1)$, and then defining

$$X_j = \frac{Z_j}{\sum\limits_{k=1}^{d+1} Z_k} \qquad j = 1, \ldots, d+1.$$

3. **The Multivariate Normal Distribution**

The **multivariate normal distribution** is a multivariate generalization of the normal distribution. The joint pdf of $\mathbf{X} = (X_1, \dots, X_d)^\top$ takes the form

$$f_{X_1,\dots,X_d}(x_1, \dots, x_d) = \left(\frac{1}{2\pi}\right)^{d/2} \frac{1}{|\Sigma|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \mu)^\top \Sigma^{-1}(\mathbf{x} - \mu)\right\}$$

where $\mathbf{x} = (x_1, \dots, x_d)^\top$, μ is a $d \times 1$ vector, and Σ is a symmetric, positive-definite $d \times d$ matrix. The distribution is obtained by taking a vector $\mathbf{Z} = (Z_1, \dots, Z_d)^\top$ of independent standard Normal random variables with joint pdf

$$f_{Z_1,\dots,Z_d}(z_1, \dots, z_d) = \left(\frac{1}{2\pi}\right)^{d/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{d} z_i^2\right\} = \left(\frac{1}{2\pi}\right)^{d/2} \exp\left\{-\frac{1}{2}\mathbf{z}^\top\mathbf{z}\right\}$$

and taking the linear transformation

$$\mathbf{X} = \mathbf{L}\mathbf{Z} + \mu$$

where $\mathbf{L}$ is the Cholesky factor of Σ, that is,

$$\Sigma = \mathbf{L}\mathbf{L}^\top.$$

Using the multivariate transformation result, we can deduce the multivariate Normal joint pdf. It can be shown that for any linear combination

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$$

for constant matrix $\mathbf{A}$ and vector $\mathbf{b}$ (compatible in dimension) also has a multivariate normal distribution; this result can be derived using moment generating functions; we have for $\mathbf{t} = (t_1, \dots, t_d)^\top \in \mathbb{R}^d$, by independence

$$M_{\mathbf{Z}}(\mathbf{t}) = \exp\left\{\frac{1}{2}\sum_{i=1}^{d} t_i^2\right\} = \exp\left\{\frac{1}{2}\mathbf{t}^\top\mathbf{t}\right\}$$

so therefore

$$\begin{aligned}
M_{\mathbf{X}}(\mathbf{t}) = \mathbb{E}_{\mathbf{X}}[\exp\{\mathbf{t}^\top\mathbf{X}\}] &= \mathbb{E}_{\mathbf{Z}}[\exp\{\mathbf{t}^\top(\mathbf{L}\mathbf{Z} + \mu)\}] \\
&= \exp\{\mathbf{t}^\top\mu\}\mathbb{E}_{\mathbf{Z}}[\exp\{(\mathbf{t}^\top\mathbf{L})\mathbf{Z})\}] \\
&= \exp\{\mathbf{t}^\top\mu\}M_{\mathbf{Z}}(\mathbf{L}^\top\mathbf{t}) \\
&= \exp\{\mathbf{t}^\top\mu\}\exp\left\{\frac{1}{2}(\mathbf{L}^\top\mathbf{t})^\top(\mathbf{L}^\top\mathbf{t})\right\} \\
&= \exp\left\{\mathbf{t}^\top\mu + \frac{1}{2}\mathbf{t}^\top(\mathbf{L}\mathbf{L}^\top)\mathbf{t}\right\} \\
&= \exp\left\{\mathbf{t}^\top\mu + \frac{1}{2}\mathbf{t}^\top\Sigma\mathbf{t}\right\}.
\end{aligned}$$

The distribution of $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$ can be deduced using similar methods as

$$\mathbf{Y} \sim Normal_d(\mathbf{A}\mu + \mathbf{b}, \mathbf{A}\Sigma\mathbf{A}^\top).$$

Marginal And Conditional Distributions

All marginal and all conditional distributions derived from the multivariate normal are also multivariate normal; for the marginal distributions, the result follows immediately from the derivation above

Suppose that vector random variable $\mathbf{X} = (X_1, X_2, \ldots, X_d)^\top$ has a multivariate normal distribution with pdf given by

$$f_{\mathbf{X}}(\mathbf{x}) = \left(\frac{1}{2\pi}\right)^{d/2} \frac{1}{|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}\mathbf{x}^\top \Sigma^{-1} \mathbf{x}\right\} \tag{1}$$

where Σ is the $d \times d$ variance-covariance matrix (we can consider here the case where the expected value μ is the $d \times 1$ zero vector; results for the general case are easily available by transformation).

Consider partitioning $\mathbf{X}$ into two components $\mathbf{X}_1$ and $\mathbf{X}_2$ of dimensions d_1 and $d_2 = d - d_1$ respectively, that is,

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix}.$$

We attempt to deduce

(a) the **marginal** distribution of $\mathbf{X}_1$, and
(b) the **conditional** distribution of $\mathbf{X}_2$ **given** that $\mathbf{X}_1 = \mathbf{x}_1$.

First, write

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

where Σ_{11} is $d_1 \times d_1$, Σ_{22} is $d_2 \times d_2$, $\Sigma_{21} = \Sigma_{12}^\top$, and

$$\Sigma^{-1} = \mathbf{V} = \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{bmatrix}$$

so that $\Sigma\mathbf{V} = \mathbf{I}_d$ ($\mathbf{I}_r$ is the $r \times r$ identity matrix) gives

$$\begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_{d_1} & 0 \\ 0 & \mathbf{I}_{d_2} \end{bmatrix}$$

where 0 represents the zero matrix of appropriate dimension. More specifically,

$$\Sigma_{11}\mathbf{V}_{11} + \Sigma_{12}\mathbf{V}_{21} = \mathbf{I}_{d_1} \tag{2}$$
$$\Sigma_{11}\mathbf{V}_{12} + \Sigma_{12}\mathbf{V}_{22} = 0 \tag{3}$$
$$\Sigma_{21}\mathbf{V}_{11} + \Sigma_{22}\mathbf{V}_{21} = 0 \tag{4}$$
$$\Sigma_{21}\mathbf{V}_{12} + \Sigma_{22}\mathbf{V}_{22} = \mathbf{I}_{d_2}. \tag{5}$$

From the multivariate normal pdf in equation (1), we can re-express the term in the exponent as

$$\mathbf{x}^\top \Sigma^{-1} \mathbf{x} = \mathbf{x}_1^\top \mathbf{V}_{11}\mathbf{x}_1 + \mathbf{x}_1^\top \mathbf{V}_{12}\mathbf{x}_2 + \mathbf{x}_2^\top \mathbf{V}_{21}\mathbf{x}_1 + \mathbf{x}_2^\top \mathbf{V}_{22}\mathbf{x}_2. \tag{6}$$

In order to compute the marginal and conditional distributions, we must complete the square in $\mathbf{x}_2$ in this expression. We can write

$$\mathbf{x}^\top \Sigma^{-1} \mathbf{x} = (\mathbf{x}_2 - \mathbf{m})^\top \mathbf{M}(\mathbf{x}_2 - \mathbf{m}) + \mathbf{c} \tag{7}$$

and by comparing with equation (6) we can deduce that, for quadratic terms in $\mathbf{x}_2$,

$$\mathbf{x}_2^\top \mathbf{V}_{22}\mathbf{x}_2 = \mathbf{x}_2^\top \mathbf{M}\mathbf{x}_2 \qquad \therefore \qquad \mathbf{M} = \mathbf{V}_{22}$$

for linear terms

$$\mathbf{x}_2^\top \mathbf{V}_{21}\mathbf{x}_1 = -\mathbf{x}_2^\top \mathbf{M}\mathbf{m} \qquad \therefore \qquad \mathbf{m} = -\mathbf{V}_{22}^{-1}\mathbf{V}_{21}\mathbf{x}_1$$

and for constant terms

$$\mathbf{x}_1^\top \mathbf{V}_{11}\mathbf{x}_1 = \mathbf{c} + \mathbf{m}^\top \mathbf{M}\mathbf{m} \qquad \therefore \qquad \mathbf{c} = \mathbf{x}_1^\top (\mathbf{V}_{11} - \mathbf{V}_{21}^\top \mathbf{V}_{22}^{-1}\mathbf{V}_{21})\mathbf{x}_1$$

thus yielding all the terms required for equation (7), that is

$$\mathbf{x}^\top \Sigma^{-1}\mathbf{x} = (\mathbf{x}_2 + \mathbf{V}_{22}^{-1}\mathbf{V}_{21}\mathbf{x}_1)^\top \mathbf{V}_{22}(\mathbf{x}_2 + \mathbf{V}_{22}^{-1}\mathbf{V}_{21}\mathbf{x}_1) + \mathbf{x}_1^\top (\mathbf{V}_{11} - \mathbf{V}_{21}^\top \mathbf{V}_{22}^{-1}\mathbf{V}_{21})\mathbf{x}_1,$$

which, crucially, is a sum of two terms, where the first can be interpreted as a function of $\mathbf{x}_2$, given $\mathbf{x}_1$, and the second is a function of $\mathbf{x}_1$ only.

Hence we have a factorization of the joint pdf using the chain rule for random variables;

$$f_{\mathbf{X}}(\mathbf{x}) = f_{\mathbf{X}_2|\mathbf{X}_1}(\mathbf{x}_2|\mathbf{x}_1) f_{\mathbf{X}_1}(\mathbf{x}_1) \tag{8}$$

where

$$f_{\mathbf{X}_2|\mathbf{X}_1}(\mathbf{x}_2|\mathbf{x}_1) \propto \exp\left\{-\frac{1}{2}(\mathbf{x}_2 + \mathbf{V}_{22}^{-1}\mathbf{V}_{21}\mathbf{x}_1)^\top \mathbf{V}_{22}(\mathbf{x}_2 + \mathbf{V}_{22}^{-1}\mathbf{V}_{21}\mathbf{x}_1)\right\}$$

giving that

$$\mathbf{X}_2|\mathbf{X}_1 = \mathbf{x}_1 \ \sim Normal_{d_2}\left(-\mathbf{V}_{22}^{-1}\mathbf{V}_{21}\mathbf{x}_1, \mathbf{V}_{22}^{-1}\right) \tag{9}$$

and

$$f_{\mathbf{X}_1}(\mathbf{x}_1) \propto \exp\left\{-\frac{1}{2}\mathbf{x}_1^\top (\mathbf{V}_{11} - \mathbf{V}_{21}^\top \mathbf{V}_{22}^{-1}\mathbf{V}_{21})\mathbf{x}_1\right\}$$

giving that

$$\mathbf{X}_1 \sim Normal_{d_1}\left(0, (\mathbf{V}_{11} - \mathbf{V}_{21}^\top \mathbf{V}_{22}^{-1}\mathbf{V}_{21})^{-1}\right). \tag{10}$$

But, from equation (3), $\Sigma_{12} = -\Sigma_{11}\mathbf{V}_{12}\mathbf{V}_{22}^{-1}$, and then from equation (2), substituting in Σ_{12},

$$\Sigma_{11}\mathbf{V}_{11} - \Sigma_{11}\mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21} = \mathbf{I}_d \qquad \therefore \qquad \Sigma_{11} = (\mathbf{V}_{11} - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21})^{-1} = (\mathbf{V}_{11} - \mathbf{V}_{21}^\top \mathbf{V}_{22}^{-1}\mathbf{V}_{21})^{-1}.$$

Hence, by inspection of equation (10), we conclude that

$$\boxed{\mathbf{X}_1 \sim Normal_{d_1}\left(0, \Sigma_{11}\right),}$$

that is, we can extract the Σ_{11} block of Σ to define the marginal sigma matrix of $\mathbf{X}_1$.

Using similar arguments, we can define the conditional distribution from equation (9) more precisely. First, from equation (3), $\mathbf{V}_{12} = -\Sigma_{11}^{-1}\Sigma_{12}\mathbf{V}_{22}$, and then from equation (5), substituting in $\mathbf{V}_{12}$

$$-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\mathbf{V}_{22} + \Sigma_{22}\mathbf{V}_{22} = \mathbf{I}_{d-d} \qquad \therefore \qquad \mathbf{V}_{22}^{-1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} = \Sigma_{22} - \Sigma_{12}^\top \Sigma_{11}^{-1}\Sigma_{12}.$$

Finally, from equation (3), taking transposes on both sides, we have that $\mathbf{V}_{21}\Sigma_{11} + \mathbf{V}_{22}\Sigma_{21} = 0$. Then pre-multiplying by $\mathbf{V}_{22}^{-1}$, and post-multiplying by Σ_{11}^{-1}, we have

$$\mathbf{V}_{22}^{-1}\mathbf{V}_{21} + \Sigma_{21}\Sigma_{11}^{-1} = 0 \qquad \therefore \qquad \mathbf{V}_{22}^{-1}\mathbf{V}_{21} = -\Sigma_{21}\Sigma_{11}^{-1},$$

so we have, substituting into equation (9), that

$$\boxed{\mathbf{X}_2|\mathbf{X}_1 = \mathbf{x}_1 \ \sim Normal_{d_2}\left(\Sigma_{21}\Sigma_{11}^{-1}\mathbf{x}_1, \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\right).}$$

Thus any marginal, and any conditional distribution of a multivariate normal joint distribution is also multivariate normal, as the choices of $\mathbf{X}_1$ and $\mathbf{X}_2$ are arbitrary.