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 $1-\theta$. Suppose that 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 $k+1$ types of objects ($k = 1, 2, ...$), with θ_i being the proportion of Type i objects, for $i = 1, ..., k+1$. Let X_i be the random variable corresponding to the number of type i objects in a sample of size n, for $i = 1, ..., k$. Then the joint pmf of vector $\mathbf{X} = (X_1, ..., X_k)^\top$ is given by

$$f_{X_1,...,X_k}(x_1, ..., x_k) = \frac{n!}{x_1!...x_k!x_{k+1}!}\theta_1^{x_1}....\theta_k^{x_k}\theta_{k+1}^{x_{k+1}} = \frac{n!}{x_1!...x_k!x_{k+1}!}\prod_{i=1}^{k+1}\theta_i^{x_i}$$

where $0 \leq \theta_i \leq 1$ for all i, and $\theta_1 + ... + \theta_k + \theta_{k+1} = 1$, and where x_{k+1} is defined by

$$x_{k+1} = n - (x_1 + \ldots + x_k).$$

This is the mass function for the **multinomial distribution** which reduces to the binomial if $k = 1$.

2 The Dirichlet Distribution

The Dirichlet distribution is a multivariate generalization of the Beta distribution. The joint pdf of vector $\mathbf{X} = (X_1, ..., X_k)^\top$ is given by

$$f_{X_1,...,X_k}(x_1, ..., x_k) = \frac{\Gamma(\alpha)}{\Gamma(\alpha_1)...\Gamma(\alpha_k)\Gamma(\alpha_{k+1})}x_1^{\alpha_1-1}...x_k^{\alpha_k-1}x_{k+1}{}^{\alpha_{k+1}-1}$$

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

$$x_{k+1} = 1 - (x_1 + ... + x_k).$$

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

3 The Multivariate Normal Distribution

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

$$f_{X_1,...,X_k}(x_1, ..., x_k) = \left(\frac{1}{2\pi}\right)^{k/2}\frac{1}{|\Sigma|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)^\top\Sigma^{-1}(\mathbf{x}-\mu)\right\}$$

where $\mathbf{x} = (x_1, ..., x_k)^\top$, μ is a $k \times 1$ vector, and Σ is a symmetric, positive-definite $k \times k$ matrix. It can be shown that all marginal and all conditional distributions derived from the multivariate normal are also multivariate normal, and that any linear combination

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

for matrix A also has a multivariate normal distribution.

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

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

where Σ is the $k \times k$ variance-covariance matrix (we can consider here the case where the expected value $\underset{\sim}{\mu}$ is the $k \times 1$ 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 k_1 and $k_2 = k - k_1$ 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 $k_1 \times k_1$, Σ_{22} is $k_2 \times k_2$, $\Sigma_{21} = \Sigma_{12}^\top$, and

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

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

$$\begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_{k_1} & 0 \\ 0 & \mathbf{I}_{k_2} \end{bmatrix}$$

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

$$\Sigma_{11}\mathbf{V}_{11} + \Sigma_{12}\mathbf{V}_{21} = \mathbf{I}_{k_1} \tag{2}$$
$$\Sigma_{11}\mathbf{V}_{12} + \Sigma_{12}\mathbf{V}_{22} = 0 \tag{3}$$
$$\Sigma_{21}\mathbf{V}_{11} + \Sigma_{22}\mathbf{V}_{21} = 0 \tag{4}$$
$$\Sigma_{21}\mathbf{V}_{12} + \Sigma_{22}\mathbf{V}_{22} = \mathbf{I}_{k_2}. \tag{5}$$

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 $\mathbf{x}_2$,

$$\mathbf{x}_2^\top \mathbf{V}_{22}\mathbf{x}_2 = \mathbf{x}_2^\top \mathbf{M}\mathbf{x}_2 \qquad \therefore \qquad \mathbf{M} = \mathbf{V}_{22} \tag{8}$$

for linear terms

$$\mathbf{x}_2^\top \mathbf{V}_{21}\mathbf{x}_1 = -\mathbf{x}_2^\top \mathbf{M}\mathbf{m} \qquad \therefore \qquad \mathbf{m} = -\mathbf{V}_{22}^{-1}\mathbf{V}_{21}\mathbf{x}_1 \tag{9}$$

and for constant terms

$$\mathbf{x}_1^\top \mathbf{V}_{11}\mathbf{x}_1 = \mathbf{c} + \mathbf{m}^\top \mathbf{M}\mathbf{m} \qquad \therefore \qquad \mathbf{c} = \mathbf{x}_1^\top (\mathbf{V}_{11} - \mathbf{V}_{21}^\top \mathbf{V}_{22}^{-1}\mathbf{V}_{21})\mathbf{x}_1 \tag{10}$$

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

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

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 an immediate factorization of the full joint pdf using the chain rule for random variables;

$$f_{\mathbf{X}}(\mathbf{x}) = f_{\mathbf{X}_2|\mathbf{X}_1}(\mathbf{x}_2|\mathbf{x}_1) f_{\mathbf{X}_1}(\mathbf{x}_1) \tag{12}$$

where

$$f_{\mathbf{X}_2|\mathbf{X}_1}(\mathbf{x}_2|\mathbf{x}_1) \propto \exp\left\{-\frac{1}{2}(\mathbf{x}_2 + \mathbf{V}_{22}^{-1}\mathbf{V}_{21}\mathbf{x}_1)^\top \mathbf{V}_{22}(\mathbf{x}_2 + \mathbf{V}_{22}^{-1}\mathbf{V}_{21}\mathbf{x}_1)\right\} \tag{13}$$

giving that

$$\mathbf{X}_2|\mathbf{X}_1 = \mathbf{x}_1 \ \sim \mathcal{N}_{k_2}\left(-\mathbf{V}_{22}^{-1}\mathbf{V}_{21}\mathbf{x}_1, \mathbf{V}_{22}^{-1}\right) \tag{14}$$

and

$$f_{\mathbf{X}_1}(\mathbf{x}_1) \propto \exp\left\{-\frac{1}{2}\mathbf{x}_1^\top (\mathbf{V}_{11} - \mathbf{V}_{21}^\top \mathbf{V}_{22}^{-1}\mathbf{V}_{21})\mathbf{x}_1\right\} \tag{15}$$

giving that

$$\mathbf{X}_1 \sim \mathcal{N}_{k_1}\left(0, (\mathbf{V}_{11} - \mathbf{V}_{21}^\top \mathbf{V}_{22}^{-1}\mathbf{V}_{21})^{-1}\right). \tag{16}$$

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 (16), we conclude that

$$\boxed{\mathbf{X}_1 \sim \mathcal{N}_{k_1}\left(0, \Sigma_{11}\right),} \tag{17}$$

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 (14) more precisely. First, from equation (3), $\mathbf{V}_{12} = -\Sigma_{11}^{-1}\Sigma_{12}\mathbf{V}_{22}$, and then from equation (5), substituting in $\mathbf{V}_{12}$

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

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 (14), that

$$\boxed{\mathbf{X}_2|\mathbf{X}_1 = \mathbf{x}_1 \ \sim \mathcal{N}_{k_2}\left(\Sigma_{21}\Sigma_{11}^{-1}\mathbf{x}_1, \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\right).} \tag{18}$$

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.