The Multivariate Normal Distribution

Why should we consider the multivariate normal distribution? It would seem that applied problems are so complex that it would only be interesting from a mathematical perspective.

1. It is mathematically tractable for a large number of problems, and, therefore, progress towards answers to statistical questions can be provided, even if only approximately so.
2. Because it is tractable for so many problems, it provides insight into techniques based upon other distributions or even non-parametric techniques. For this, it is often a benchmark against which other methods are judged.
3. For some problems it serves as a reasonable model of the data. In other instances, transformations can be applied to the set of responses to have the set conform well to multivariate normality.
4. The sampling distribution of many (multivariate) statistics are normal, regardless of the parent distribution (Multivariate Central Limit Theorems). Thus, for large sample sizes, we may be able to make use of results from the multivariate normal distribution to answer our statistical questions, even when the parent distribution is not multivariate normal.

Consider first the univariate normal distribution with parameters μ (the mean) and σ (the variance) for the random variable x,

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} \tag{1}$$

for $-\infty < x < \infty$, $-\infty < \mu < \infty$, and $\sigma^2 > 0$.

Now rewrite the exponent $(x-\mu)^2/\sigma^2$ using the linear algebra formulation of

$$(x-\mu)'(\sigma^2)^{-1}(x-\mu).$$

This formulation matches that for the generalized or Mahalanobis squared distance

$$(\mathbf{x}-\mu)'\mathbf{\Sigma^{-1}}(\mathbf{x}-\mu),$$

where both $\mathbf{x}$ and μ are vectors. The multivariate normal distribution can be derived by substituting the Mahalanobis squared distance formula into the univariate formula and normalizing the distribution such that the total probability of the distribution is 1. This yields,

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{p/2}\left|\Sigma\right|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mu)'\mathbf{\Sigma^{-1}}(\mathbf{x}-\mu)} \tag{2}$$

for $-\infty < \mathbf{x} < \infty$, $-\infty < \mu < \infty$, and for Σ positive definite.

In the bivariate normal case the squared distance formula, in terms of the individual means μ_1 and μ_2, variances σ_{11} and σ_{22}, and correlation ρ_{12}, is

$$(\mathbf{x}-\mu)'\mathbf{\Sigma^{-1}}(\mathbf{x}-\mu) = \frac{1}{1-\rho_{12}^2}\left[\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)^2 + \left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)^2 - 2\rho_{12}\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)\right]. \tag{3}$$

Properties of the Multivariate Normal Distribution

Let $\mathbf{X} \sim \mathbf{N_p}(\mu, \mathbf{\Sigma})$ be p-variate multivariate normal with mean μ and variance-covariance matrix Σ, where

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix} \qquad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} \qquad \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}.$$

1. The solid ellipsoid of all $\mathbf{x}$, such that $(\mathbf{x}-\mu)'\mathbf{\Sigma^{-1}}(\mathbf{x}-\mu) \leq \chi^2_p(\alpha)$, contains $(1-\alpha)100\%$ of the probability in the distribution. This also implies that contours delineating regions of constant probability about μ are given by $(\mathbf{x}-\mu)'\mathbf{\Sigma^{-1}}(\mathbf{x}-\mu) = \chi^2_p(\alpha)$.

2. The semiaxes of the ellipsoid containing $(1-\alpha)100\%$ probability are given by the eigenvalues (λ_i) and eigenvectors $(\mathbf{e_i})$ of Σ such that the semiaxes are

   $$\pm c\sqrt{\lambda_i}\mathbf{e_i}$$

   where $c^2 = \chi^2_p(\alpha)$.

3. All subsets of $\mathbf{X}$ are themselves multivariate normal.

4. Any linear combination of the X_i, say $\mathbf{c}'\mathbf{X} = c_1X_1 + c_2X_2 + \cdots + c_pX_p$, is normally distributed as

   $$\mathbf{c}'\mathbf{X} \sim N(\mathbf{c}'\mu, \mathbf{c}'\mathbf{\Sigma c}).$$

   Further, q linear combinations of the X_i, say $C'\mathbf{X}$, is (q-variate) multivariate normal. Let

   $$C'\mathbf{X} = \begin{bmatrix} c_{11}X_1 & + & c_{12}X_2 & + & \cdots & + & c_{1p}X_p \\ c_{21}X_1 & + & c_{22}X_2 & + & \cdots & + & c_{2p}X_p \\ \vdots & + & \vdots & + & \vdots & + & \vdots \\ c_{q1}X_1 & + & c_{q2}X_2 & + & \cdots & + & c_{qp}X_p \end{bmatrix},$$

   then

   $$C'\mathbf{X} \sim N_q(C'\mu, C'\Sigma C).$$

5. Subdivide the vector $\mathbf{X}$ into two subsets $\mathbf{X_1}$ and $\mathbf{X_2}$,

   $$\mathbf{X} = \begin{bmatrix} \mathbf{X_1} \\ \mathbf{X_2} \end{bmatrix}, \ \mathbf{X_1} = \begin{bmatrix} X_{11} \\ X_{12} \\ \vdots \\ X_{1p} \end{bmatrix}, \text{ and } \mathbf{X_2} = \begin{bmatrix} X_{21} \\ X_{22} \\ \vdots \\ X_{2q} \end{bmatrix},$$

   and so that

   $$\mathbf{X} = \begin{bmatrix} \mathbf{X_1} \\ \mathbf{X_2} \end{bmatrix} \sim N_{p+q}\left(\begin{bmatrix} \mu_\mathbf{1} \\ \mu_\mathbf{2} \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right).$$

   The conditional distribution of $\mathbf{X_1}$ given $\mathbf{X_2} = \mathbf{x_2}$, $f(\mathbf{X_1}|\mathbf{X_2} = \mathbf{x_2})$ is (p-variate) multivariate normal,

   $$N_p(\mu_\mathbf{2} + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x_2} - \mu_\mathbf{2}), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}).$$

6. If two variates, say X_1 and X_2, of the multivariate normal are uncorrelated, $\rho_{12} = 0$ and implies $\sigma_{12} = 0$, then X_1 and X_2 are independent. This property is *not* in general true for other distributions. However, it is always true that if two variates are independent, then they are uncorrelated, no matter what their joint distribution is.

Sampling Distributions of the Multivariate Normal

1. Let
$$\bar{\mathbf{X}} = \begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \vdots \\ \bar{X}_p \end{bmatrix}$$
be the vector of sample means from a sample of size n from the multivariate normal distribution for $\mathbf{X}$, then
$$\bar{\mathbf{X}} \sim N_p(\mu, \frac{1}{n}\Sigma).$$

2. Let S be the sample variance-covariance matrix computed from a sample of size n from the multivariate normal distribution for $\mathbf{X}$, then
$$(n-1)S \sim W_{(n-1)}(\Sigma),$$
the Wishart distribution with $(n-1)$ degrees of freedom.

3. The density function W for S does not exist when $n \leq p$. Further, S must be positive definite $(\lambda_i > 0 \; \forall \; i = 1, 2, \cdots, p)$ for the density to exist.

4. $\bar{\mathbf{X}}$ and S are stochastically independent.

5. Let $(n-1)S \sim W_{(n-1)}(\Sigma)$, then
$$(n-1)C'SC \sim W_{(n-1)}(C'\Sigma C).$$

6. Let $A_1 = (n_1 - 1)S_1 \sim W_{(n_1-1)}(\Sigma)$ and $A_2 = (n_2 - 1)S_2 \sim W_{(n_2-1)}(\Sigma)$, where S_1 and S_2 are independent estimates of Σ, then
$$A_1 + A_2 \sim W_{(n_1+n_2-2)}(\Sigma)$$
and
$$(\frac{1}{n_1 + n_2 - 2})(A_1 + A_2)$$
is a "pooled" estimate of Σ.

7. Let $\mathbf{X_1}, \mathbf{X_2}, \ldots, \mathbf{X_n}$ be a simple random sample of size n, where $\mathbf{X_i} \sim N_p(\mu, \Sigma)$, then approximately for $(n-p)$ large,
$$n(\bar{\mathbf{X}} - \mu)'\Sigma^{-1}(\bar{\mathbf{X}} - \mu) \sim \chi^2_p.$$
A central limit theorem says that for very large $n - p$ we can relax the requirement that the $\mathbf{X_i}$ be multivariate normal. Further, for $n - p$ large, an approximate $(1-\alpha)100\%$ confidence region for μ is given by the set of all μ such that
$$n(\bar{\mathbf{X}} - \mu)'S^{-1}(\bar{\mathbf{X}} - \mu) \leq \chi^2_p(\alpha).$$

Assessing Multivariate Normality

The methods for assessing multivariate normality of a set of data make use of the properties of the multivariate normal distribution discussed earlier. You should also note that the tools assume a common multivariate normal distribution for the data, i.e., the same mean μ and covariance matrix Σ. This means that for many sets of data, checks on multivariate normality will need to be performed on the residuals rather than the raw data. Some ideas to consider are:

1. All marginal distributions must be normal. Check the normality of each variable. If a variable does not conform to the normal distribution, then the set of variables can not be multivariate normal.

   Steps for the q-q normal distribution plot:

   (a) Order the observations from smallest to largest ($X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$). These are the order statistics for this random variable and they estimate the quantiles of the distribution from which they were sampled. The quantile is the value at which a certain proportion of the distribution is less than or equal to that value.

   (b) Estimate the proportion of the distribution that should be less than or equal to the value of each order statistic. One such estimate is

   $$\frac{i - 1/2}{n}$$

   where i is the rank of each observation.

   (c) Compute the expected quantiles from the normal distribution as

   $$q_i = \Phi^{-1}\left(\frac{i - 1/2}{n}\right),$$

   where Φ^{-1} is the inverse of the standard normal cumulative distribution function.

   (d) Plot the observed quantiles, $X_{(i)}$, versus the expected quantiles, q_i, and check for linearity of the plot. If the observed quantiles correspond with a normal distribution, then the points will plot on a straight line. If not, reject (multivariate) normality. Note, you should have a minimum sample size of 20 to begin to have confidence in the plot.

2. All pairs of variables must be bivariate normal. Produce scatter plots of all pairs of variables. Density regions should correspond roughly to elliptical patterns with linear relationships among pairs of variables.

3. Linear combinations of the variables are normal. Check any meaningful linear combinations for normality (sums, differences). Further, check the principal components (linear combinations corresponding to the eigenvectors of Σ) for normality. Check pairs of linear combinations and principal components for bivariate normality. Rejection of normality or bivariate normality for linear combinations also rejects multivariate normality.

4. Squared distances about the population mean vector are distributed as chi-square with p degrees of freedom. Estimate the population mean vector with the sample mean vector, and estimate the population covariance matrix with the sample covariance matrix. Compute the squared distances of each observation to the sample mean vector and check to see that they are chi-square distributed.

   Steps for the q-q chi-square distribution plot:

(a) Compute $d_i^2 = (\mathbf{X_i} - \bar{\mathbf{X}})' S^{-1} (\mathbf{X_i} - \bar{\mathbf{X}})$.

(b) Order the d_i^2 from smallest to largest to get observed quantiles of the distribution as $d_{(1)}^2 \leq d_{(2)}^2 \leq \cdots \leq d_{(n)}^2$.

(c) Compute expected quantiles from the χ_p^2 distribution where

$$q_i = \chi_p^2 \left(\frac{i - 1/2}{n} \right)$$

corresponding with each d_i^2, $i = 1, 2, \ldots, n$.

(d) Plot d_i^2 versus q_i for $i = 1, 2, \ldots, n$ and check for linearity in the plot. If the points do not form a straight line, then the observed quantiles do not follow from the chi-square distribution, so reject multivariate normality.

These steps can be repeated for subsets of the variables and linear combinations. This may be useful in identifying whether "problems" of multivariate normality are associated with a set of variables or a single variable.

If (multivariate) normality is rejected, then

1. check for outliers or errors in the data. If outliers are identified, then use methods for dealing with outliers to determine their impact on the analysis. Alternative robust methods may also be available. Remember that an outlier may be the most informative observation in the data set as it is "different" from all the others.

2. consider transformations of one or more of the variables. A variable may, for example, follow the lognormal distribution, so a logarithm transformation would be in order. Note that transformations also affect the variable's associations with the other variables.

3. consider robust or alternative multivariate methods if available. Some techniques are much less sensitive to outliers or the distribution of the data than others. See, for example, Silverman, B.W., 1986, *Density Estimation for Statistics and Data Analysis*, Chapman & Hall, New York, 175pp.

4. consider basing inference on results using resampling methods (Monte Carlo, bootstrap, or permutation methods). See, for example, Manly, B.F.J., 1997, *Randomization, Bootstrap and Monte Carlo Methods in Biology*, second edition, Chapman & Hall, New York, 399pp.