Sampling From The Normal Density

* We will now consider the analysis of random samples from the multivariate normal distribution.

* As usual a random sample is a set of vectors $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ which are mutually independent and identically distributed.

* In this case the common distribution will be a p-dimensional normal distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$.

* In general we shall assume that $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are unknown and that interest is in their estimation.

The Multivariate Normal Likelihood

* Maximum likelihood is a very common method to estimate unknown parameters.

* We maximize the likelihood function which is proportional to the joint density of all of the data.

* If we observe the multivariate normal random sample $\boldsymbol{X}_1 = \boldsymbol{x}_1, \boldsymbol{X}_2 = \boldsymbol{x}_2, \ldots, \boldsymbol{X}_n = \boldsymbol{x}_n$ then the likelihood is

$$
\begin{aligned}
L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) &\propto \prod_{i=1}^{n} f_p(\boldsymbol{x}_i; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \\
&\propto \prod_{i=1}^{n} \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{- \frac{1}{2}(\boldsymbol{x}_i - \boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu})\right\} \\
&= \frac{1}{|\boldsymbol{\Sigma}|^{n/2}} \exp\left\{- \frac{1}{2} \sum_{i=1}^{n} (\boldsymbol{x}_i - \boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu})\right\}
\end{aligned}
$$

Alternative Expressions for the Exponent

$$
\begin{aligned}
&\sum_{i=1}^{n} (\boldsymbol{x}_i - \boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu}) \\
&\quad = \sum_{i=1}^{n} (\boldsymbol{x}_i - \overline{\boldsymbol{x}})^t \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_i - \overline{\boldsymbol{x}}) - n(\overline{\boldsymbol{x}} - \boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1} (\overline{\boldsymbol{x}} - \boldsymbol{\mu}) \\
&\quad = \mathrm{tr}\left(\boldsymbol{\Sigma}^{-1} \sum_{i=1}^{n} (\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})^t \right) - n(\overline{\boldsymbol{x}} - \boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1} (\overline{\boldsymbol{x}} - \boldsymbol{\mu})
\end{aligned}
$$

Lemma 3

Suppose that $\boldsymbol{B}$ is a symmetric positive definite matrix and $b > 0$ is a constant. Then

$$
\frac{1}{|\boldsymbol{\Sigma}|} e^{-\mathrm{tr}(\boldsymbol{\Sigma}^{-1} B)/2} \leqslant \frac{1}{|\boldsymbol{B}|}^{b} (2b)^{pb}
$$

Maximum Likelihood Estimators

Theorem 13

Suppose that

$$\boldsymbol{X}_1 = \boldsymbol{x}_1, \boldsymbol{X}_2 = \boldsymbol{x}_2, \ldots, \boldsymbol{X}_n = \boldsymbol{x}_n$$

is a random sample from the p-dimensional normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Then the maximum likelihood estimates are

$$\hat{\boldsymbol{\mu}} = \overline{\boldsymbol{x}} \qquad \hat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})^t = \frac{n-1}{n}\boldsymbol{S}$$

The corresponding maximum likelihood estimators are

$$\hat{\boldsymbol{\mu}} = \overline{\boldsymbol{X}} \qquad \hat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{X}_i - \overline{\boldsymbol{X}})(\boldsymbol{X}_i - \overline{\boldsymbol{X}})^t$$

Sampling Distributions

* The maximum likelihood estimates are important but they are not sufficient for inference.

* To conduct inference we need to consider the maximum likelihood estimators.

* These are random quantities since they are functions of the underlying random vectors $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$.

* As all random quantities their behaviour can be determined from their distribution.

* Specifying the distribution will allow us to find confidence regions and/or test if a sample satisfies some pre-specified hypothesis.

Sampling Distributions of $\overline{X}$ and S

Theorem 14

Suppose that $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ is a random sample from a p-variate normal distribution with mean $\boldsymbol{\mu}$ and covariacne $\boldsymbol{\Sigma}$. Then

1. *The Sample mean has distribution*

$$\overline{\boldsymbol{X}} \sim N_p\left(\boldsymbol{\mu}, \frac{1}{n}\boldsymbol{\Sigma}\right)$$

2. $(n-1)\boldsymbol{S}$ *is a random matrix with a Wishart distribution with $n-1$ degrees of freedom and covariance matrix $\boldsymbol{\Sigma}$.*

3. $\overline{\boldsymbol{X}}$ *and $\boldsymbol{S}$ are independent.*

The Wishart Distribution

* A distribution defined on the space of positive definite symmetric matrices.

* First derived in 1928 by John Wishart at Rothamstead Experimantal Station in England.

* The pdf of the Wishart distribution with k degrees of freedom and $p \times p$ covariance matrix $\boldsymbol{\Sigma}$ is

$$w(\boldsymbol{A}; k, \boldsymbol{\Sigma}) \;=\; \frac{|\boldsymbol{A}|^{(k-p-1)/2} \exp\left\{-\mathrm{tr}\left(\boldsymbol{A}\boldsymbol{\Sigma}^{-1}\right)/2\right\}}{2^{kp/2}\pi^{p(p-1)/4} \prod_{j=1}^{p} \Gamma((n-j)/2)}$$

for $\boldsymbol{A}$ positive definite.

* When $\boldsymbol{\Sigma} = 1$ (and $p = 1$) the Wishart distribution becomes the chi-squared distribution with k degrees of freedom.

Properties of the Wishart Distribution

* If $\boldsymbol{Y} \sim \mathsf{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$ then $\boldsymbol{Y}\boldsymbol{Y}^t \sim \mathsf{W}(1, \boldsymbol{\Sigma})$.

* If $\boldsymbol{A}_1 \sim \mathsf{W}(k_1, \boldsymbol{\Sigma})$ and $\boldsymbol{A}_2 \sim \mathsf{W}(k_2, \boldsymbol{\Sigma})$ and $\boldsymbol{A}_1, \boldsymbol{A}_2$ are independent then

$$\boldsymbol{A}_1 + \boldsymbol{A}_2 \;\sim\; \mathsf{W}(k_1 + k_2, \boldsymbol{\Sigma})$$

* If $\boldsymbol{A} \sim \mathsf{W}(k, \boldsymbol{\Sigma})$ and $\boldsymbol{C}$ is a $q \times p$ matrix of rank $q \leqslant p$ then

$$\boldsymbol{C}\boldsymbol{A}\boldsymbol{C}^t \;\sim\; \mathsf{W}\left(k, \boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}^t\right)$$

* A consequence of this is that if $\boldsymbol{A} \sim \mathsf{W}(k, \boldsymbol{\Sigma})$ then $A_{ii} \sim \sigma_{ii}\chi^2_k$.

Large Sample Approximate Distributions

Theorem 15 (Multivariate Central Limit Theorem)
Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be a random sample from a p-dimensional distribution with mean $\boldsymbol{\mu}$ and non-singular covariance matrix $\boldsymbol{\Sigma}$. Then

$$\sqrt{n}\left(\overline{\boldsymbol{X}} - \boldsymbol{\mu}\right) \quad \xrightarrow{d} \quad N_p(\mathbf{0}, \boldsymbol{\Sigma})$$

Theorem 16
Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be a random sample from a p-dimensional distribution with mean $\boldsymbol{\mu}$ and non-singular covariance matrix $\boldsymbol{\Sigma}$. Then

$$n\left(\overline{\boldsymbol{X}} - \mu\right) \boldsymbol{S}^{-1} \left(\overline{\boldsymbol{X}} - \boldsymbol{\mu}\right)^t \quad \xrightarrow{d} \quad \chi_p^2$$

* Note that for these results to be approximately valid we require $n - p$ to be large.

Assessing Normality

* The derivation of the maximum likelihood estimators and most of our distributional results (such as in Theorem 14) require that the the random sample come from a p-dimensional normal distribution.

* Given a real dataset, we would like to use these (and subsequent) results.

* To do that safely we need to be reasonably sure that the underlying distribution is at least close to multivariate normal.

* How can we assess if that is the case for a given dataset?

The Univariate Case

* Suppose that $p = 1$ and we wish to assess normality.

* We can examine a histogram and see if it looks unimodal and reasonably symmetric.

* This is helpful but is not always sufficient.

* As well as the histogram being the correct shape we would want the spread of the data to be in agreement with what we would expect from a normal distribution (most observations within 2-3 standard deviations of the mean).

* It is hard to assess that from a histogram.

The Normal Q-Q Plot

* One way of assessing if the data may come from a normal distribution is to compare the ordered data with what we would expect if the underlying distribution truly was normal.

* It suffices to look at the standard normal since any normal can be written as a linear transformation of the standard normal.

* A Quantile-Quantile Plot plots quantiles of the standard normal (x-axis) against the observed order statistics (y-axis).

* If the data truly come from a normal distribution then the resulting plot should approximate a straight line.

Extending The Q-Q Plot

* A crucial step in the univarite Q-Q plot is putting the data in increasing order.

* For higher dimensions, such ordering is not so easy.

* One way to order multivariate observations is in terms of their Mahalanobis Distance

$$\delta_i^2 = (\boldsymbol{x}_i - \boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu})$$

* Of course $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are generally not known so instead we use

$$d_i^2 = (\boldsymbol{x}_i - \overline{\boldsymbol{x}})^t \boldsymbol{S}^{-1} (\boldsymbol{x}_i - \overline{\boldsymbol{x}})$$

* From Theorem 16, these should look like a sample from a χ_p^2 distribution as long as $n - p$ is sufficiently large.

The Chi-Squared Q-Q Plot

1. Calculate the squared distances

$$d_i^2 = (\boldsymbol{x}_i - \overline{\boldsymbol{x}})^t \boldsymbol{S}^{-1} (\boldsymbol{x}_i - \overline{\boldsymbol{x}})$$

2. Sort them in order $d_{(1)}^2 \leqslant d_{(2)}^2 \leqslant \cdots \leqslant d_n^2$.

3. Find $q_i(p, n)$ such that

$$\mathsf{P}\left(\chi_p^2 \leqslant q_i(p, n)\right) = \frac{i - 0.5}{n}$$

4. Plot the $\left(q_i(p, n), d_{(i)}^2\right)$ pairs.

5. If the true population distribution is normal then the plot should be close to a 45° line.

Outliers

* Many real datasets contain outliers which must be examined.

* These will generally be visible in the extremes of the Q-Q plot.

* It is always important to examine why a point is an outlier.

* Simply removing the point from the dataset may be the wrong thing to do as the outlier may contain very valuable information.

* Of course, if it can be determined that there is an error in the recorded values for this observation then it should be corrected or removed if correction is not possible.

Approximate Normality and Transformations

* The results presented in this section are actually quite robust to minor violations of normality.

* That is why we do not need to do any formal strict testing, just visual inspection of the Q-Q plot is sufficient.

* There will, of course, be situations in which even approximate normality is not a reasonable assumption.

* In many such cases we can transform one or more components of the random vector to make the overall vector closer to multivariate normal.

Useful Transformations

* If one variable is a count then taking square root is useful.

* For proportions (constrained to be between 0 and 1), taking the logit usually helps

$$\text{logit}(p) \; = \; \log\left(\frac{p}{1-p}\right)$$

* For variables representing correlations, the Fisher transformation

$$z(r) \; = \; \frac{1}{2}\log\left(\frac{1+r}{1-r}\right)$$

will make the correlations approximately normal if the data used to construct the correlations is bivariate normal.

* For positively skewed data, taking logs or square roots will often help.