Multivariate Statistics

Sudipta Das

Assistant Professor,
Department of Data Science,
Ramakrishna Mission Vivekananda University, Kolkata
Slides adapted from Jhonson & Winchern

Outline I

1. Multivariate Normal Distribution
   - Multivariate Normal Density
   - MVN Likelihood and Maximum Likelihood Estimation of μ and Σ
   - The Sampling Distribution of $\bar{X}$ and S
   - Large Sample Behavior of $\bar{X}$ and S
   - Assessing the Assumption of Normality
   - Detecting Outliers and Cleaning Data
   - Transformations to Near Normality

Multivariate Normal Distribution I

- The pdf of multivariate normal random vector $\mathbf{X}' = [X_1, X_2, \ldots, X_p]$ with mean μ and covariance matrix Σ is given by

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

  where $-\infty < x_i < \infty$.
- Notation $\mathbf{X} \sim N_p(\mu, \Sigma)$

Multivariate Normal Distribution II

- Bivariate Normal Density

$$f(x_1, x_2) = \frac{e^{-\frac{1}{2(1-\rho_{12}^2)}\left[\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)^2 + \left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)^2 - 2\rho_{12}\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)\right]}}{2\pi\sqrt{\sigma_{11}\sigma_{22}(1-\rho_{12}^2)}}$$

- Figure 4.2 (Page 152)

Multivariate Normal Distribution III

- Observation
  - Contours of constant density for the $p-$dimensional normal distribution are ellipsoids defined by $\mathbf{x}$ such that
$$(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu) = c^2.$$
  - These ellipsoids are centerded at μ and have axes $\pm c\sqrt{\lambda_i}e_i$,
    - where $\Sigma e_i = \lambda_i e_i$, for $i = 1, 2, \ldots, p$.
- Example 4.2 (Page 154)
- Code 01

Properties of Multivariate Normal Distribution I

- Properties of Multivariate Normal Distribution
  - Linear combinations of the components of $\mathbf{X}$ are normally distributed.
  - All subsets of the components of $\mathbf{X}$ have a (multivariate) normal distribution.
  - Zero covariance implies that the corresponding components are independently distributed.
  - The conditional distributions of the components are (multivariate) normal.

Properties of Multivariate Normal Distribution II

- Result 1:
  - If $\mathbf{X}$ is distributed as $N_p(\mu, \Sigma)$, then any linear combination of variables $\mathbf{a}'\mathbf{X} = a_1X_1 + a_2X_2 + \ldots + a_pX_p$ is distributed as $N(a'\mu, a'\Sigma a)$.
  - Also,if $\mathbf{a}'\mathbf{X}$ is distributed as $N(a'\mu, a'\Sigma a)$ for every $\mathbf{a}$, then $\mathbf{X}$ must be $N_p(\mu, \Sigma)$.

Properties of Multivariate Normal Distribution III

- Result 2:
  - If $\mathbf{X}$ is distributed as $N_p(\mu, \Sigma)$, the q linear combinations

$$A\mathbf{X} = \begin{bmatrix} a_{11}X_1 + \ldots + a_{1p}X_p \\ a_{21}X_1 + \ldots + a_{2p}X_p \\ \vdots \\ a_{q1}X_1 + \ldots + a_{qp}X_p \end{bmatrix} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_p \end{bmatrix} = \mathbf{Y}$$

    are distributed as $N_q(A\mu, A\Sigma A')$.
    - Sketch of proof: For any $\mathbf{c}$, $\mathbf{c}'\mathbf{Y} \sim N()$, as it is a linear combination of $X_1, X_2, \ldots, X_p$.
  - Also, $\mathbf{X} + \mathbf{d}$, where d is a vector of constants, is distributed as $N_p(\mu + d, \Sigma)$.

Properties of Multivariate Normal Distribution IV

- Result 3:
  - All subsets of **X** are normally distributed.
  - If we respectively partition **X**, its mean vector μ, and its covariance matrix Σ as

$$\mathbf{X}_{p\times 1} = \begin{bmatrix} \mathbf{X_1}_{q\times 1} \\ \mathbf{X_2}_{(p-q)\times 1} \end{bmatrix}, \mu = \begin{bmatrix} \mu_{\mathbf{1}\,q\times 1} \\ \mu_{\mathbf{2}\,(p-q)\times 1} \end{bmatrix}$$

and

$$\Sigma_{p\times p} = \begin{bmatrix} \Sigma_{11\,q\times q} & \Sigma_{12\,q\times(p-q)} \\ \Sigma_{21\,(p-q)\times q} & \Sigma_{22\,(p-q)\times(p-q)} \end{bmatrix}$$

then $\mathbf{X_1}$ is distributed as $N_q(\mu_1, \Sigma_{11})$.

Properties of Multivariate Normal Distribution V

- Sketch of proof:
  For example if we are interested in the subset $\{X_k, X_l\}$, then choose the matrix A as in Result 2 as

$$A = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 0 & 1_{(k,k)} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1_{(l,l)} & 0 \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix}$$

Properties of Multivariate Normal Distribution VI

- Result 4:
  - If $\begin{bmatrix} \mathbf{X_1} \\ \mathbf{X_2} \end{bmatrix}$ is

$$N_{q_1+q_2}\left(\begin{bmatrix} \mu_{\mathbf{1}} \\ \mu_{\mathbf{2}} \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}\right),$$

then $\mathbf{X_1}$ and $\mathbf{X_2}$ are independent iff

$$Cov(\mathbf{X_1}, \mathbf{X_2}) = \Sigma_{12} = 0.$$

- Sketch of proof:
  Joint density function

$$
\begin{aligned}
f(\mathbf{x}) &= \frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)} \\
&= \frac{1}{(2\pi)^{\frac{q_1+q_2}{2}}\left|\begin{bmatrix}\Sigma_{11} & 0 \\ 0 & \Sigma_{22}\end{bmatrix}\right|^{\frac{1}{2}}} e^{-\frac{1}{2}\begin{bmatrix}\mathbf{x_1}-\mu_1 \\ \mathbf{x_2}-\mu_2\end{bmatrix}'\begin{bmatrix}\Sigma_{11} & 0 \\ 0 & \Sigma_{22}\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{x_1}-\mu_1 \\ \mathbf{x_2}-\mu_2\end{bmatrix}} \\
&= \frac{1}{(2\pi)^{\frac{q_1+q_2}{2}}|\Sigma_{11}|^{\frac{1}{2}}|\Sigma_{22}|^{\frac{1}{2}}} e^{-\frac{1}{2}\begin{bmatrix}\mathbf{x_1}-\mu_1 & \mathbf{x_2}-\mu_2\end{bmatrix}\begin{bmatrix}\Sigma_{11}^{-1} & 0 \\ 0 & \Sigma_{22}^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{x_1}-\mu_1 \\ \mathbf{x_2}-\mu_2\end{bmatrix}} \\
&= \frac{1}{(2\pi)^{\frac{q_1}{2}}|\Sigma_{11}|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{x_1}-\mu_1)'\Sigma_{11}^{-1}(\mathbf{x_1}-\mu_1)} \frac{1}{(2\pi)^{\frac{q_2}{2}}|\Sigma_{22}|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{x_2}-\mu_2)'\Sigma_{22}^{-1}(\mathbf{x_2}-\mu_2)} \\
&= f(\mathbf{x_1})f(\mathbf{x_2})
\end{aligned}
$$

Properties of Multivariate Normal Distribution VIII

- Result 5:
  - Let $\mathbf{X} = \begin{bmatrix} \mathbf{X_1} \\ \mathbf{X_2} \end{bmatrix}$ be distributed as $N_p(\mu, \Sigma)$ with

$$\mu = \begin{bmatrix} \mu_\mathbf{1} \\ \mu_\mathbf{2} \end{bmatrix}, \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \text{ and } |\Sigma_{22}| > 0.$$

    Then the conditional distribution of $\mathbf{X_1}$, given that $\mathbf{X_2} = \mathbf{x_2}$, is normal and has

$$\text{Mean } = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x_2} - \mu_2)$$

    and

$$\text{Covariance } = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}.$$

Properties of Multivariate Normal Distribution IX

- Sketch of the proof:

Choose

$$A = \begin{bmatrix} I_{q_1 \times q_1} & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I_{q_2 \times q_2} \end{bmatrix}.$$

Now,

$$A(\mathbf{X} - \mu) = \begin{bmatrix} \mathbf{X_1} - \mu_1 - \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{X_2} - \mu_2) \\ \mathbf{X_2} - \mu_2 \end{bmatrix} \sim N_p\left(\mathbf{0}, \begin{bmatrix} \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} & 0 \\ 0 & \Sigma_{22} \end{bmatrix}\right)$$

Thus,

$$\mathbf{X_1} - \mu_1 - \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{X_2} - \mu_2) \sim N_{q_1}(\mathbf{0}, \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$$

and it is independent of $\mathbf{X_2}$.
Hence, given $\mathbf{X_2} = \mathbf{x_2}$ for any $\mathbf{x_2}$,

$$\mathbf{X_1}|\mathbf{X_2} = \mathbf{x_2} \sim N\left(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x_2} - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)$$

- Result 6:
  - Let $\mathbf{X}$ be distributed as $N_p(\mu, \Sigma)$ with $|\Sigma| > 0$. Then
    - $(\mathbf{X} - \mu)'\Sigma^{-1}(\mathbf{X} - \mu)$ is distributed as χ^2_p.
    - The $N_p(\mu, \Sigma)$ distribution assigns probability $1 - \alpha$ to the solid ellipsoid $\{\mathbf{x} : (\mathbf{X} - \mu)'\Sigma^{-1}(\mathbf{X} - \mu) \leq \chi^2_p(\alpha)\}$ where denotes the upper (100α)th percentile of the χ^2_p distribution.

Properties of Multivariate Normal Distribution XI

- Sketch of proof:

  Choose $A = \begin{bmatrix} \frac{1}{\sqrt{\lambda_1}} e_1' \\ \vdots \\ \frac{1}{\sqrt{\lambda_p}} e_p' \end{bmatrix}$, where $(\lambda_i, \mathbf{e}_i)s$ are eigen-value and eigen-vector pairs of $\Sigma = \sum_{i=1}^{p} \lambda_i e_i e_i'$.

  Now
  $$\mathbf{Z} = A(\mathbf{X} - \mu) \sim N_p(\mathbf{0}, A\Sigma A' = I).$$

  Thus,
  $$\begin{aligned} \mathbf{Z}'\mathbf{Z} = (\mathbf{X} - \mu)' A' A (\mathbf{X} - \mu) &= (\mathbf{X} - \mu)' \left[\frac{1}{\sqrt{\lambda_1}} e_1 \cdots \frac{1}{\sqrt{\lambda_p}} e_p \right] \begin{bmatrix} \frac{1}{\sqrt{\lambda_1}} e_1' \\ \vdots \\ \frac{1}{\sqrt{\lambda_p}} e_p' \end{bmatrix} (\mathbf{X} - \mu) \\ &= (\mathbf{X} - \mu)' \left[\sum_{i=1}^{p} \frac{1}{\lambda_i} e_i e_i' \right] (\mathbf{X} - \mu) = (\mathbf{X} - \mu)' \Sigma^{-1} (\mathbf{X} - \mu) \sim \chi_p^2 \end{aligned}$$

- Code02

Properties of Multivariate Normal Distribution XII

- Result 7:
  - Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be mutually independent with $\mathbf{X}_j$ distributed as $N_p(\mu_j, \Sigma)$. Then

$$\mathbf{V}_1 = c_1\mathbf{X}_1 + c_2\mathbf{X}_2 + \ldots + c_n\mathbf{X}_n$$

    is distributed as $N_p\left(\sum_{j=1}^{n} c_j\mu_j, \left(\sum_{j=1}^{n} c_j^2\right)\Sigma\right)$.
  - Moreover, $\mathbf{V}_1$ and $\mathbf{V}_2 = b_1\mathbf{X}_1 + b_2\mathbf{X}_2 + \ldots + b_n\mathbf{X}_n$ are jointly multivariate normal with covariance matrix

$$\begin{bmatrix} \mathbf{c}'\mathbf{c}\Sigma & \mathbf{b}'\mathbf{c}\Sigma \\ \mathbf{b}'\mathbf{c}\Sigma & \mathbf{b}'\mathbf{b}\Sigma \end{bmatrix}.$$

    Consequently, $\mathbf{V}_1$ and $\mathbf{V}_2$ are independent if $\mathbf{b}'\mathbf{c} = 0$.

Sampling from Multivariate Normal Distribution I

- Let the sample size be n
- Samples are taken independently from a multivariate normal population with mean vector μ and covariance matrix Σ
- Likelihood function

$$
\begin{aligned}
f_{\mathbf{X_1},\mathbf{X_2},\ldots,\mathbf{X_n}}(\mathbf{X_1},\mathbf{X_2},\ldots,\mathbf{X_n}) &= \prod_{j=1}^{n} \frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(x_j-\mu)'\Sigma^{-1}(x_j-\mu)} \\
&= \frac{e^{-\frac{1}{2}\sum_{j=1}^{n}(x_j-\mu)'\Sigma^{-1}(x_j-\mu)}}{(2\pi)^{\frac{np}{2}}|\Sigma|^{\frac{n}{2}}} \\
&= L(\mu,\Sigma).
\end{aligned}
$$

Sampling from Multivariate Normal Distribution II

- Result 8:
  - Let $\mathbf{X_1}, \mathbf{X_2}, \ldots, \mathbf{X_n}$ be a random sample from a normal population with mean μ and covariance Σ. Then

$$\hat{\mu} = \bar{\mathbf{X}} \text{ and } \hat{\Sigma} = \frac{1}{n}\sum_{j=1}^{n}(\mathbf{X_j} - \bar{\mathbf{X}})(\mathbf{X_j} - \bar{\mathbf{X}})' = \frac{n-1}{n}S$$

are the *maximum likelihood estimator*s of μ and Σ, respectively.

Their observed values, $\bar{\mathbf{x}}$ and $\frac{1}{n}\sum_{j=1}^{n}(\mathbf{x_j} - \bar{\mathbf{x}})(\mathbf{x_j} - \bar{\mathbf{x}})'$, are called the maximum likelihood estimates of μ and Σ.

- Sketch of proof:

$$L(\mu, \Sigma) \propto \frac{1}{|\Sigma|^{n/2}} e^{-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_i-\mu)'\Sigma^{-1}(\mathbf{x}_i-\mu)}$$

  - Note

$$\begin{aligned}\sum_{i=1}^{n}(\mathbf{x}_i-\mu)'\Sigma^{-1}(\mathbf{x}_i-\mu) &= \sum_{i=1}^{n}(\mathbf{x}_i-\bar{\mathbf{x}}+\bar{\mathbf{x}}-\mu)'\Sigma^{-1}(\mathbf{x}_i-\bar{\mathbf{x}}+\bar{\mathbf{x}}-\mu) \\ &= \sum_{i=1}^{n}(\mathbf{x}_i-\bar{\mathbf{x}})'\Sigma^{-1}(\mathbf{x}_i-\bar{\mathbf{x}})+n(\bar{\mathbf{x}}-\mu)'\Sigma^{-1}(\bar{\mathbf{x}}-\mu),\end{aligned}$$

is minimized at $\mu = \bar{\mathbf{x}}$

- Note

$$L(\mu = \bar{\mathbf{x}}, \Sigma) \propto \frac{1}{|\Sigma|^{n/2}} e^{-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}})'\Sigma^{-1}(\mathbf{x}_i - \bar{\mathbf{x}})} = \frac{1}{|\Sigma|^{n/2}} e^{-\frac{1}{2} tr\left[\Sigma^{-1}\sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})'\right]}.$$

Let, $B = \sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})'$.
Thus, $L(\mu = \bar{\mathbf{x}}, \Sigma) \propto \frac{1}{|\Sigma|^{n/2}} e^{-\frac{1}{2} tr\left(\Sigma^{-1}B\right)}$.
Let η_is are eigenvalues of $B^{1/2}\Sigma^{-1}B^{1/2}$. Hence,

$$\begin{aligned}
L(\mu = \bar{\mathbf{x}}, \Sigma) \propto \frac{1}{|\Sigma|^{n/2}} e^{-\frac{1}{2} tr\left(\Sigma^{-1}B\right)} &= \frac{|B|^{n/4}|\Sigma|^{-n/2}|B|^{n/4}}{|B|^{n/2}} e^{-\frac{1}{2} tr\left(B^{1/2}\Sigma^{-1}B^{1/2}\right)} \\
&= \frac{|B^{1/2}\Sigma^{-1}B^{1/2}|^{\frac{n}{2}}}{|B|^{\frac{n}{2}}} e^{-\frac{1}{2} tr\left(B^{1/2}\Sigma^{-1}B^{1/2}\right)} \\
&= \frac{\left(\prod_{i=1}^{p}\eta_i\right)^{\frac{n}{2}}}{|B|^{\frac{n}{2}}} e^{-\frac{1}{2}\sum_{i=1}^{p}\eta_i} = \frac{\prod_{i=1}^{p}\eta_i^{\frac{n}{2}} e^{-\frac{\eta_i}{2}}}{|B|^{\frac{n}{2}}}.
\end{aligned}$$

Since, $x^{n/2}e^{-x/2}$ is maximized at $x = n$.
Thus $\max(L(\bar{\mathbf{x}}, \Sigma)) = \frac{n^{np/2}e^{-np/2}}{|B|^{n/2}}$.
Now at $\Sigma = \frac{1}{n}B$, the $L\left(\bar{\mathbf{x}}, \Sigma = \frac{1}{n}B\right) \propto \frac{1}{|\frac{1}{n}B|^{n/2}} e^{-\frac{1}{2} tr\left(nB^{-1}B\right)} = \frac{n^{\frac{np}{2}}}{|B|^{n/2}} e^{-np/2}$, we achieve the maximum.

Sampling from Multivariate Normal Distribution V

- Obsevation:

$$
\begin{aligned}
L(\hat{\mu}, \hat{\Sigma}) &= \frac{1}{(2\pi)^{\frac{np}{2}}} e^{-\frac{np}{2}} \frac{1}{|\hat{\Sigma}|^{\frac{n}{2}}} \\
&\propto (\text{Generalized Variance})^{-\frac{n}{2}},
\end{aligned}
$$

since

$$
|\hat{\Sigma}| = |S_n| = \left(\frac{n-1}{n}\right)^p |S|.
$$

The Sampling Distribution of $\bar{X}$ and S

- Properties of $\bar{X}$ and S
  - $\bar{X}$ is distributed as
    $$N_p\left(\mu, \frac{1}{n}\Sigma\right)$$
  - $(n-1)S$ is distributed as
    $$W_{p,(n-1)} = \sum_{i=1}^{n-1} ZZ',$$
    where $Z \sim N_p(0, \Sigma)$.
    - Wishart random matrix of order $p \times p$ with $n-1$ d.f.
  - $\bar{X}$ and S are independent.

- Large Sample Behavior of $\bar{X}$ and S
  - (Law of large numbers). Let $X_1, X_2, \ldots, X_n$ be independent observations from a population with mean $E(X_i) = \mu$. Then

$$\bar{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

    converges in probability to μ as n increases without bound.
    - S converges to Σ in probability.
  - (The central limit theorem). Let $X_1, X_2, \ldots, X_n$ be independent observations from any population with mean μ and finite covariance Σ. Then

$$\sqrt{n}(\bar{X} - \mu) \sim N_p(0, \Sigma)$$

    for large sample sizes. Here n should also be large relative to p.
  - $n(\bar{X} - \mu)' S^{-1} (\bar{X} - \mu)$ is approximately χ^2_p.

Assessing the Assumption of Normality I

- Most of the statistical techniques discussed, assume that each vector observation X_i comes from a multivariate normal distribution.
  - Do the marginal distributions of the elements of X appear to be normal?
  - Do the scatter plots of pairs of observations on different characteristics give the elliptical appearance expected from normal populations?

Assessing the Assumption of Normality II

- It has proved difficult to construct a "good" overall test of joint normality in more than two dimension.
    - It is possible, for example, to construct a nonnormal bivariate distribution with normal marginals.
    - For most practical work, one-dimensional and two-dimensional investigations are ordinarily sufficient

- Evaluating the Normality of the Univariate Marginal Distributions.
  1. A univariate normal distribution assigns probability .683(.954) to the interval $(\mu_i - 1(2)\sqrt{\sigma_{ii}}, \mu_i + 1(2)\sqrt{\sigma_{ii}})$.
     - Consequently, with a large sample size n, we expect the observed proportion of the observations lying in the interval $(\bar{x}_i - 1(2)\sqrt{s_{ii}}, \bar{x}_i + 1(2)\sqrt{s_{ii}})$. to be about .683(.954).

2. Q-Q plot
   - Let $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ represent these observations after they are ordered according to magnitude.
   - For a standard normal distribution, the quantiles, $q_{(j)}$ are defined by the relation
   $$\int_{-\infty}^{q(j)} \frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}} dz = \frac{j - .5}{n}.$$
   - If the data arise from a normal distribution the pairs $(q_{(j)}, x_{(j)})$ will be approximately linearly related.

   Example 4.9 (Page 179)

- Evaluating the Normality of the Bivariate Distributions.
  - If the observations were generated from a multivariate normal distribution, each bivariate distribution would be normal.

1. Check the set of bivariate outcomes **x** such that $(x-\mu)'\Sigma^{-1}(x-\mu) \leq \chi_2^2(.5)$ has probability 0.5.
   - μ is replaced by $\bar{x}$ and Σ^{-1} is replaced by S^{-1}.

Assessing the Assumption of Normality VI

2. More formal method for judging the joint normality of a data set is based on squared generalized distance $d_j^2 = (x_j - \bar{x})' S^{-1} (x_j - \bar{x})$, where x_js are sample observations.
   - When the parent population is multivariate normal and both n and $n - p$ are large each of the squared distances $d_1^2, d_2^2, \ldots, d_n^2$ should behave like a chi-square random variable.
   - The resulting plot is called chi-square plot or gamma plot.
   - Note: It can be used for any $p \geq 2$.

Assessing the Assumption of Normality VII

- Constructing chi-square plot
  - Order the squared distance $d^2_{(1)}, d^2_{(2)}, \dots, d^2_{(n)}$
  - Graph the pairs $(q_{c,p}\frac{(j-.5)}{n}, d^2_{(j)})$, where $q_{c,p}(\frac{j-.5}{n})$ is the $100(j-.5)/n$ quantile of the chi-square distribution with p degrees of freedom.
- The plot should resemble a straight line through the origin having slop 1.
- One or two points far above the line indicate large distances, or outlying observations, that merit further attention.
  Example 4.13 (Page 184)

Detecting Outliers and Cleaning Data I

- Most data sets contain one or few unusual observations that do not seem to belong to the pattern of variability produced by other obseravtions.
- Outliers are not wrong numbers, they need further investigations.

Detecting Outliers and Cleaning Data II

- Methods of detecting outliers
  1. Make a dot plot for each variable.
     - Calculate the standardized values $z_{jk} = \frac{x_{jk} - \bar{x}_k}{\sqrt{s_{kk}}}$, for $j = 1, \ldots, n$ and $k = 1, \ldots, p$.
     - Examine these standardized values for large and small values.

     Figure 4.10 (Page 188)
  2. Make a scatter plot for each pair of variables.
     - Calculate the generalized sqared distance $(x_j - \bar{x})' S^{-1} (x_j - \bar{x})$.
     - Examine these distances for unusually large values.
     - In a *chi-square* plot, these would be the points farthest form the origin.

     Figure 4.11 (Page 191)

Transformations to Near Normality I

- If normality assumption is violated
  - Transform the data
  - For example:
    - Count data (y) take the square roots ($\sqrt{y}$)
    - Proportion data (p) take logit transformation $\left(\frac{1}{2}\log\frac{p}{1-p}\right)$
    - Correlation coefficients (r) take Fisher's z-transform $\left(\frac{1}{2}\log\frac{1+r}{1-r}\right)$

Transformations to Near Normality II

- It is convenient to let the data suggest a transformation
  - A useful transformation for this purpose is the family of power transformations
  - For positive r.v.
  - Shrinking $\ldots, x^{-1}, x^{-1/2}, \ln x, x^{1/4}, x^{1/2}$
  - Expanding $x^2, x^3, \ldots$

Transformations to Near Normality III

- Box and Cox family of power transformations

$$x^{(\lambda)} = \begin{cases} \frac{x^\lambda - 1}{\lambda}, & \text{if } \lambda \neq 0 \\ \ln x, & \text{if } \lambda = 0, \end{cases}$$

Transformations to Near Normality IV

- Choice of an appropriate power λ is the solution of that maximizes the expression

$$l(\lambda) = -\frac{n}{2} \ln \left[\frac{1}{n} \sum_{j=1}^{n} \left(x_j^{(\lambda)} - \bar{x_j^{(\lambda)}} \right)^2 \right] + (\lambda - 1) \sum_{j=1}^{n} \ln x_j.$$

- After the transformation, one should also check for adequacy of normality.
  Example 4.16 (Page 194)

Transformations to Near Normality V

- Multivariate Data:- Try to make each marginal distribution approximately normal.
  - For all k in $1, \ldots, p$ maximize

$$l_k(\lambda) = -\frac{n}{2} \ln \left[\frac{1}{n} \sum_{j=1}^{n} \left(x_{jk}^{(\lambda_k)} - \overline{x_{jk}^{(\lambda_k)}} \right)^2 \right] + (\lambda_k - 1) \sum_{j=1}^{n} \ln x_{jk}.$$

  - Hence, $\hat{\lambda} = [\hat{\lambda}_1, \ldots, \hat{\lambda}_p]$.
  - Therefore, the jth transformed multivariate observation is $x_j^{(\hat{\lambda})} = [\frac{x_{j1}^{\hat{\lambda}_1} - 1}{\hat{\lambda}_1}, \ldots, \frac{x_{jp}^{\hat{\lambda}_p} - 1}{\hat{\lambda}_p}]'$.
  - It's equivalent to maximizing the univariate likelihood for kth feature over the parameters μ_k, σ_{kk} and λ_k.

Transformations to Near Normality VI

- If the normal marginals are not sufficient to ensure that the joint distribution is normal, one can start with initial values as $\lambda = [\hat{\lambda}_1, \ldots, \hat{\lambda}_p]$ to maximize the multivariate function

$$l(\lambda_1, \ldots, \lambda_p) = -\frac{n}{2} \ln |S(\lambda)| + \sum_{k=1}^{p} \left((\lambda_k - 1) \sum_{j=1}^{n} \ln x_{jk} \right),$$

where $S(\lambda)$ is the sample covariance matrix computed from $x_j^{(\lambda)} = [\frac{x_{j1}^{\lambda_1} - 1}{\lambda_1}, \ldots, \frac{x_{jp}^{\lambda_p} - 1}{\lambda_p}]'$.

- It's equivalent to maximizing the multivariate likelihood over the parameters μ, Σ and λ.