Multivariate Normal Distribution

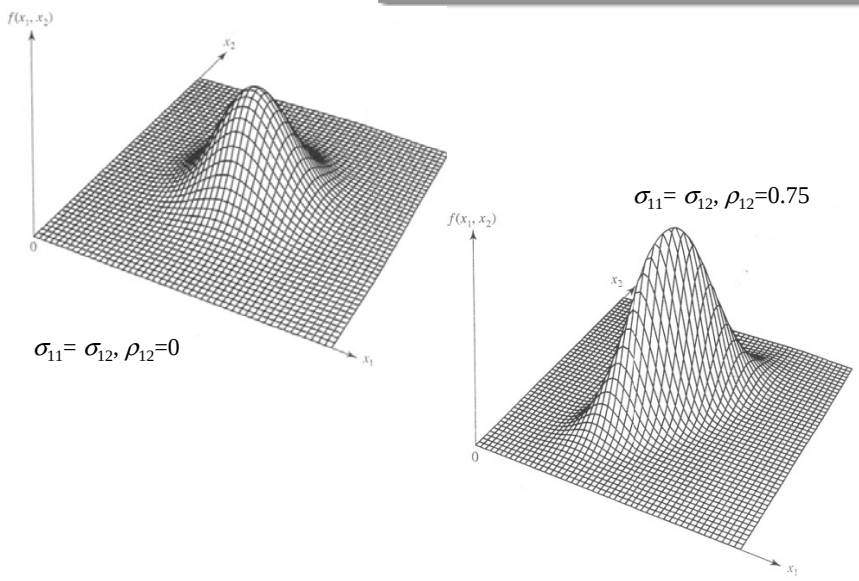
The Multivariate Normal Density

- Normal density is often a useful approximation to the "true" population distribution
- Multivariate normal distribution is mathematically tractable
- The p-dimensional normal density, denoted by $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ for the random vector $\mathbf{X}^T=[\mathbf{X}_1, \ldots, \mathbf{X}_p]$ has the form

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{p/2} |\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}-\boldsymbol{\mu})/2\right\}$$

where $-\infty<x_i<\infty$, $i=1, \ldots, p$, $\boldsymbol{\mu}$ is the expected value (mean) of $\mathbf{X}$, and the $p\times p$ matrix $\boldsymbol{\Sigma}$ is the variance-covariance matrix of $\mathbf{X}$.

Plot of Bivariate Normal Density



Contour of Constant Density

Contours of constant density for the *p*-dimensional normal distribution are ellipsoids defined by **x** such that

$$(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})=c^2$$

These ellipsoids are centered at $\boldsymbol{\mu}$ and have axes

$$\pm c\sqrt{\lambda_i}\cdot\mathbf{e}_i, \text{where } \boldsymbol{\Sigma}\mathbf{e}_i=\lambda_i\mathbf{e}_i, i=1,\ldots,p, \|\mathbf{e}\|=1$$

Properties of (Multivariate) Normal Distribution

If **X** has a multivariate normal distribution,

- Linear combinations of the components of **X** are normally distributed
- All subsets of the components of **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

23

Chi-Square Distribution

Let **X** be distributed as $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $|\boldsymbol{\Sigma}|>0$. Then $(\mathbf{X}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu})$ is distributed as χ^2_p, where χ^2_p denotes the chi-square distribution with p degrees of freedom.

24

Central Limit Theorem

- Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be independent observations from any population with mean $\boldsymbol{\mu}$ and finite covariance $\boldsymbol{\Sigma}$. Then

$$\sqrt{n}(\overline{\mathbf{X}} - \boldsymbol{\mu})$$

has an approximate $N_p(\mathbf{0}, \boldsymbol{\Sigma})$ distribution for large sample sizes. Here n should also be large relative to p.

25

Testing of Normality (Not Required)

- It is difficult to construct a "good" overall test of joint normality in more than two dimensions
- For most practical work, one-dimensional and two-dimensional investigations are fairly sufficient
- Testing univariate marginal distributions
  - Q-Q plot
  - Statistical testing
- Testing bivariate normality: based on chi-square distribution of the squared generalized distance

26

Transformations to Near Normality (Not Required)

- If normality is not a viable assumption, non-normal data can be made more “normal looking” by considering transformations of the data such as taking log, square roots, logit transformation
- Power transformation for univariate data: Box-Cox method
- Transforming Multivariate Observations: In practical applications, it may be good enough to make each (univariate) marginal distribution approximately normal