Random vectors, multivariate normal (Gaussian) distributions

February 29, 2024

Motto

Everyone believes in it (Gauss distribution): experimentalists believing that it is a mathematical theorem, mathematicians believing that it is an empirical fact.

Quote attributed to Henri Poincaré by de Finetti. However, Cramer attributes the remark to Lippman and quoted by Poincaré.

Gabriel Lippman – a Nobel prize winner in physics,
Henri Poincaré – a mathematician, theoretical physicist, engineer, and a philosopher of science

A vector of random variables

Random vector: A vector of random variables

$$\mathbf{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_p \end{bmatrix}$$

The mean of **X** – the vectors of means of X_j's

$$\mu_j = \mathbb{E}(X_j)$$

define

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix}$$

Covariance and correlation matrices: $\mathbf{\Sigma}$, $\boldsymbol{\rho}$

Given variances/covariances between pairs (X_i, X_j)'s:

$$\sigma_{ij} = \mathbb{E}((X_i - \mu_i)(X_j - \mu_j)) = \mathbb{E}(X_i X_j) - \mu_i \mu_j$$

between variables, define the covariance and correlation matrices

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}, \boldsymbol{\rho} = \begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{21} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & & \vdots \\ \rho_{p1} & \rho_{p2} & \cdots & 1 \end{pmatrix},$$

where

$$\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}}$$

Multivariate normal (Gaussian) distribution – notation

For compactness of notation a column vector $\begin{bmatrix} a \\ b \end{bmatrix}$ will be written as a pair (a, b) (the row notation with commas). Note that (a, b) is not treated as a matrix $\begin{bmatrix} a & b \end{bmatrix}$. The same applies to p-dimensional vectors.

Multivariate normal (Gaussian) distribution – notation

For compactness of notation a column vector $\begin{bmatrix} a \\ b \end{bmatrix}$ will be written as a pair (a, b) (the row notation with commas). Note that (a, b) is not treated as a matrix $\begin{bmatrix} a & b \end{bmatrix}$. The same applies to p-dimensional vectors.

- The multivariate normal or Gaussian random vector $\mathbf{X} = (X_1, \ldots, X_p)$ is given by a continuous pmf

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

that is characterized by two parameters: a vector parameter $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)$ and a matrix parameter $\boldsymbol{\Sigma}$.

Multivariate normal (Gaussian) distribution – notation

For compactness of notation a column vector $\begin{bmatrix} a \\ b \end{bmatrix}$ will be written as a pair (a, b) (the row notation with commas). Note that (a, b) is not treated as a matrix $\begin{bmatrix} a & b \end{bmatrix}$. The same applies to p-dimensional vectors.

- The multivariate normal or Gaussian random vector $\mathbf{X} = (X_1, \dots, X_p)$ is given by a continuous pmf

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

  that is characterized by two parameters: a vector parameter $\boldsymbol{\mu} = (\mu_1, \dots, \mu_p)$ and a matrix parameter $\boldsymbol{\Sigma}$.
- The notation $\mathbf{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ should be read as "the random vector $\mathbf{X}$ has multivariate normal (Gaussian) distribution with the vector parameter $\boldsymbol{\mu}$ and the matrix parameter $\boldsymbol{\Sigma}$."

Multivariate normal (Gaussian) distribution – properties

We often drop the dimension *p* from the notation writing $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Multivariate normal (Gaussian) distribution – properties

We often drop the dimension p from the notation writing $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

- The vector parameter $\boldsymbol{\mu}$ is equal to the mean of $\mathbf{X}$ and the matrix parameter $\boldsymbol{\Sigma}$ is equal to the covariance matrix of $\mathbf{X}$.

Multivariate normal (Gaussian) distribution – properties

We often drop the dimension p from the notation writing $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

- The vector parameter $\boldsymbol{\mu}$ is equal to the mean of $\mathbf{X}$ and the matrix parameter $\boldsymbol{\Sigma}$ is equal to the covariance matrix of $\mathbf{X}$.
- Any coordinate X_i of $\mathbf{X}$ is also normally distributed, i.e. X_i has $\mathcal{N}(\mu_i, \sigma_i^2)$.

Multivariate normal (Gaussian) distribution – properties

We often drop the dimension p from the notation writing $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

- The vector parameter $\boldsymbol{\mu}$ is equal to the mean of $\mathbf{X}$ and the matrix parameter $\boldsymbol{\Sigma}$ is equal to the covariance matrix of $\mathbf{X}$.
- Any coordinate X_i of $\mathbf{X}$ is also normally distributed, i.e. X_i has $\mathcal{N}(\mu_i, \sigma_i^2)$.
- If $\mathbf{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\mathbf{A}$ is a $q \times p$ (non-random) matrix, $q \leq p$, (and the matrix $\mathbf{A}$ is of the rank q), then

$$\mathbf{AX} \sim \mathcal{N}_q(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T)$$

Subsetting from coordinates of MND

Any vector made of a subset of different coordinates of $\mathbf{X}$ is also multivariate normal with the corresponding vector mean and covariance matrix.

Subsetting from coordinates of MND

Any vector made of a subset of different coordinates of $\mathbf{X}$ is also multivariate normal with the corresponding vector mean and covariance matrix.

More precisely, if $\mathbf{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix}$$

are partitioned into sub-vectors $\mathbf{X}_1 : q \times 1$ and $\mathbf{X}_2 : (p-q) \times 1$ then with

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$$

$$\mathbf{X}_1 \sim \mathcal{N}_q(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}) \text{ and } \mathbf{X}_2 \sim \mathcal{N}_{p-q}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$$

Chi-square distribution – review

Chi-square distribution – review

- Chi-square distribution with k degrees of freedom is the distribution of $\chi_k^2 = Z_1^2 + \cdots + Z_k^2$, where Z_i's are independent standard normal variables $\mathcal{N}(0, 1)$.

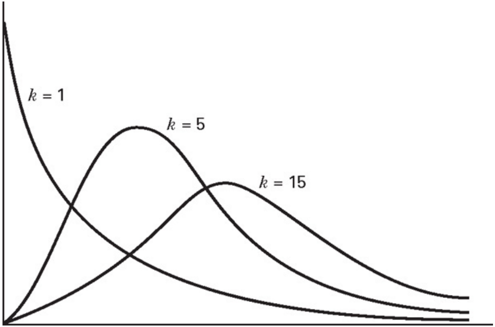
Chi-square distribution – review

- Chi-square distribution with k degrees of freedom is the distribution of $\chi_k^2 = Z_1^2 + \cdots + Z_k^2$, where Z_i's are independent standard normal variables $\mathcal{N}(0, 1)$.
- **Exercise:** Find that $E\left(\chi_k^2\right) = k$.

Chi-square distribution – review

- Chi-square distribution with k degrees of freedom is the distribution of $\chi_k^2 = Z_1^2 + \cdots + Z_k^2$, where Z_i's are independent standard normal variables $\mathcal{N}(0, 1)$.
- **Exercise:** Find that $E\left(\chi_k^2\right) = k$.
- Variance $V\left(\chi_k^2\right) = 2k$.

Chi-square distribution – review

- Chi-square distribution with k degrees of freedom is the distribution of $\chi^2_k = Z_1^2 + \cdots + Z_k^2$, where Z_i's are independent standard normal variables $\mathcal{N}(0, 1)$.
- **Exercise:** Find that $E\left(\chi^2_k\right) = k$.
- Variance $V\left(\chi^2_k\right) = 2k$.
- 

Chi-square distribution, cont

Chi-square distribution, cont

- Let $Y_1, \ldots, Y_n$ be a sample from $\mathcal{N}(\mu, \sigma^2)$ and S^2 be the sample variance. Then

$$(n-1)S^2/\sigma^2$$

is distributed the same as χ^2_{n-1}.

Constant level ellipsoid as a confidence regions

- The random variable

$$(\mathbf{X} - \mu)^T \mathbf{\Sigma}^{-1} (\mathbf{X} - \mu)$$

is distributed as chi-square with p degrees of freedom

Constant level ellipsoid as a confidence regions

- The random variable

$$(\mathbf{X} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu})$$

is distributed as chi-square with p degrees of freedom

- From this: The probability that $\mathbf{X}$ belongs to the region

$$\{\mathbf{x} \in \mathbb{R}^p; (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \leq \chi_p^2(\alpha)\}$$

is α.

Constant level ellipsoid as a confidence regions

- The random variable

$$(\mathbf{X} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu})$$

  is distributed as chi-square with p degrees of freedom
- From this: The probability that $\mathbf{X}$ belongs to the region

$$\{\mathbf{x} \in \mathbb{R}^p; (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \leq \chi_p^2(\alpha)\}$$

  is α.
- The region is called a confidence ellipsoid.

Graphical interpretation of the eigenvalues and eigenvectors

Graphical interpretation of the eigenvalues and eigenvectors

- Let $(\lambda_1, \mathbf{e}_1), \ldots, (\lambda_p, \mathbf{e}_p)$ be eigenvalue-eigenvector pairs corresponding to the covariance matrix $\boldsymbol{\Sigma}$ of a multivariate normal distribution.

Graphical interpretation of the eigenvalues and eigenvectors

- Let $(\lambda_1, \mathbf{e}_1), \ldots, (\lambda_p, \mathbf{e}_p)$ be eigenvalue-eigenvector pairs corresponding to the covariance matrix $\boldsymbol{\Sigma}$ of a multivariate normal distribution.
- The constant density ellipsoids for this distribution have the following properties:

Graphical interpretation of the eigenvalues and eigenvectors

- Let $(\lambda_1, \mathbf{e}_1), \ldots, (\lambda_p, \mathbf{e}_p)$ be eigenvalue-eigenvector pairs corresponding to the covariance matrix $\mathbf{\Sigma}$ of a multivariate normal distribution.
- The constant density ellipsoids for this distribution have the following properties:
    - The eigenvectors $\mathbf{e}_i$'s point in the direction of the axes of the ellipsoids.
    - The lengths of the axes are proportional to the square roots of λ_i's.
    - The distribution of $\mathbf{P}^T(\mathbf{X} - \mu)$ is the same as the $(\sqrt{\lambda_1} Z_1, \ldots, \sqrt{\lambda_p} Z_p)$, where Z_i's are independent standard normal random variables.

Special case – two dimensions

Definition of bivariate normal distribution

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \cdot \exp\left[-\frac{(x-\mu_X)^2}{2\sigma_X^2(1-\rho^2)} + \frac{\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y(1-\rho^2)} - \frac{(y-\mu_Y)^2}{2\sigma_Y^2(1-\rho^2)}\right]$$

Special case – two dimensions

Definition of bivariate normal distribution

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \cdot$$

$$\cdot \exp\left[-\frac{(x-\mu_X)^2}{2\sigma_X^2(1-\rho^2)} + \frac{\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y(1-\rho^2)} - \frac{(y-\mu_Y)^2}{2\sigma_Y^2(1-\rho^2)}\right]$$

- The mean $\boldsymbol{\mu} = (\mu_Y, \mu_Y)$ and $\boldsymbol{\Sigma}$ denote the 2×2 symmetric matrix

$$\begin{bmatrix} \sigma_X^2 & \sigma_X\sigma_Y\rho \\ \sigma_Y\sigma_X\rho & \sigma_Y^2 \end{bmatrix},$$

where $\rho \in [-1, 1]$ and both σ_X and σ_Y are positive.

Special case – two dimensions

Definition of bivariate normal distribution

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \cdot$$
$$\cdot \exp\left[-\frac{(x-\mu_X)^2}{2\sigma_X^2(1-\rho^2)} + \frac{\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y(1-\rho^2)} - \frac{(y-\mu_Y)^2}{2\sigma_Y^2(1-\rho^2)}\right]$$

- The mean $\boldsymbol{\mu} = (\mu_Y, \mu_Y)$ and $\boldsymbol{\Sigma}$ denote the 2×2 symmetric matrix
$$\begin{bmatrix} \sigma_X^2 & \sigma_X\sigma_Y\rho \\ \sigma_Y\sigma_X\rho & \sigma_Y^2 \end{bmatrix},$$
where $\rho \in [-1, 1]$ and both σ_X and σ_Y are positive.
- How many numerical parameters in total?

Special case – two dimensions

Definition of bivariate normal distribution

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \cdot$$
$$\cdot \exp\left[-\frac{(x-\mu_X)^2}{2\sigma_X^2(1-\rho^2)} + \frac{\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y(1-\rho^2)} - \frac{(y-\mu_Y)^2}{2\sigma_Y^2(1-\rho^2)}\right]$$

- The mean $\boldsymbol{\mu} = (\mu_Y, \mu_Y)$ and $\boldsymbol{\Sigma}$ denote the 2×2 symmetric matrix
$$\begin{bmatrix} \sigma_X^2 & \sigma_X\sigma_Y\rho \\ \sigma_Y\sigma_X\rho & \sigma_Y^2 \end{bmatrix},$$
where $\rho \in [-1, 1]$ and both σ_X and σ_Y are positive.
- How many numerical parameters in total? 2+3=5

Bivariate normal distribution – properties

The following properties hold

Bivariate normal distribution – properties

The following properties hold

1. X is normally distributed with mean μ_X and standard deviation σ_X, i.e. $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$.

Bivariate normal distribution – properties

The following properties hold

1. X is normally distributed with mean μ_X and standard deviation σ_X, i.e. $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$.
2. Y is normally distributed with mean μ_Y and standard deviation σ_Y, i.e. $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$.

Bivariate normal distribution – properties

The following properties hold

1. X is normally distributed with mean μ_X and standard deviation σ_X, i.e. $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$.
2. Y is normally distributed with mean μ_Y and standard deviation σ_Y, i.e. $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$.
3. $Corr(X, Y) = \rho$, and $Cov(X, Y) = \sigma_X \sigma_Y \rho$.

Bivariate normal distribution – properties

The following properties hold

1. X is normally distributed with mean μ_X and standard deviation σ_X, i.e. $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$.
2. Y is normally distributed with mean μ_Y and standard deviation σ_Y, i.e. $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$.
3. $Corr(X, Y) = \rho$, and $Cov(X, Y) = \sigma_X \sigma_Y \rho$.
4. X and Y are independent if and only if $\rho = 0$.

Bivariate normal distribution – properties

The following properties hold

1. X is normally distributed with mean μ_X and standard deviation σ_X, i.e. $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$.
2. Y is normally distributed with mean μ_Y and standard deviation σ_Y, i.e. $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$.
3. $Corr(X, Y) = \rho$, and $Cov(X, Y) = \sigma_X \sigma_Y \rho$.
4. X and Y are independent if and only if $\rho = 0$.
5. Any linear combination $aX + bY + c$ is also a normal variable.

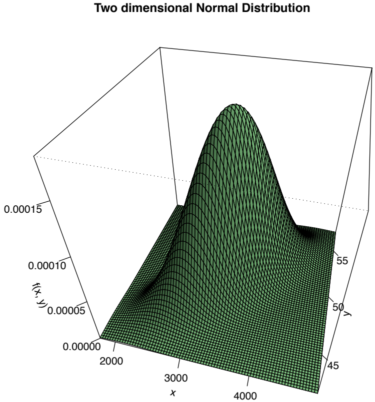
Bivariate normal distribution – properties

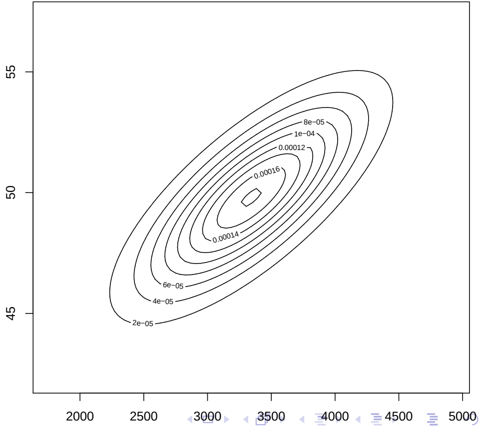
The following properties hold

1. X is normally distributed with mean μ_X and standard deviation σ_X, i.e. $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$.
2. Y is normally distributed with mean μ_Y and standard deviation σ_Y, i.e. $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$.
3. $Corr(X, Y) = \rho$, and $Cov(X, Y) = \sigma_X \sigma_Y \rho$.
4. X and Y are independent if and only if $\rho = 0$.
5. Any linear combination $aX + bY + c$ is also a normal variable.
6. Exercise: What are the parameters of the distribution of $aX + bY + c$?

Plots of bivariate normal density $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

$$\boldsymbol{\mu} = (3343, 49.8) \text{ and } \boldsymbol{\Sigma} = \begin{bmatrix} 278784 & 990 \\ 990 & 6.25 \end{bmatrix}$$





R-code – Computing densities

```
mu=c(3343,49.8)
Sigma=matrix(c(278784,990,990,6.25),byrow=T,
nrow=2,ncol=2)

#The range of x (3 sigma rule is used)
A=3*sqrt(Sigma[1,1])
x=seq(mu[1]-A,mu[1]+A,by=0.025*A)

B=3*sqrt(Sigma[2,2])
y=seq(mu[2]-B,mu[2]+B,by=0.025*B)

library('emdbook')
DenVal=function(x, y, m, s) {
  z = cbind(x,y)
  return(dmvnorm(z, mu=m, Sigma=s))
}
dva=outer(x,y,DenVal ,mu,Sigma)
```

R-code – graphical tools

Simple:

```
contour(x = x, y = y, z = dva)
```

R-code – graphical tools

Simple:

```
contour(x = x, y = y, z = dva)
```

and more fancy:

```
persp(x = x, y = y, z = dva,
main="Two dimensional Normal Distribution",
xlab = "x", ylab = "y", zlab = expression(f(x, y)),
theta = 15, phi=40,
axes = TRUE, col="lightgreen",shade=0.75, box = TRUE,
ticktype="detailed", nticks=5, ltheta=140,lphi=20)
```

Conditioning of random variables given values of other variables – review

Conditioning of random variables given values of other variables – review

- Suppose that we are dealing with a pair of variables (X, Y) and we learn about the value one of them, say $Y = y$.

Conditioning of random variables given values of other variables – review

- Suppose that we are dealing with a pair of variables (X, Y) and we learn about the value one of them, say $Y = y$.
- We would like to adjust our knowledge about X by accounting that $Y = y$.

Conditioning of random variables given values of other variables – review

- Suppose that we are dealing with a pair of variables (X, Y) and we learn about the value one of them, say $Y = y$.
- We would like to adjust our knowledge about X by accounting that $Y = y$.
- In other words we are interested in the conditional distribution of X given that Y took some specific value y

Conditioning of random variables given values of other variables – review

- Suppose that we are dealing with a pair of variables (X, Y) and we learn about the value one of them, say $Y = y$.
- We would like to adjust our knowledge about X by accounting that $Y = y$.
- In other words we are interested in the conditional distribution of X given that Y took some specific value y

Definition of the conditional pdf of X given $Y = y$

In the conctinuous variable case the conditional distribution of X given $Y = y$ is given by the conditional pmf

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}.$$

We observe that the conditional distribution of X is obtained by dividing the joint pdf of (X, Y) by the value of the marginal pdf of Y at y.

Conditional distributions for bivariate normal distributions

- As we have seen, in order to get the conditional distributions we need to know the formulas for the joint and marginal distributions.

Conditional distributions for bivariate normal distributions

- As we have seen, in order to get the conditional distributions we need to know the formulas for the joint and marginal distributions.
- The joint distribution for bivariate normal distribution is known explicitely.

Conditional distributions for bivariate normal distributions

- As we have seen, in order to get the conditional distributions we need to know the formulas for the joint and marginal distributions.
- The joint distribution for bivariate normal distribution is known explicitely.
- The marginal can be obtained by integrating in one variable the joint distribution.

Conditional distributions for bivariate normal distributions

- As we have seen, in order to get the conditional distributions we need to know the formulas for the joint and marginal distributions.
- The joint distribution for bivariate normal distribution is known explicitely.
- The marginal can be obtained by integrating in one variable the joint distribution.
- Fortunately, the marginal distribution for the jointly normal distribution is also normal.

Conditional distributions for bivariate normal distributions

- As we have seen, in order to get the conditional distributions we need to know the formulas for the joint and marginal distributions.
- The joint distribution for bivariate normal distribution is known explicitely.
- The marginal can be obtained by integrating in one variable the joint distribution.
- Fortunately, the marginal distribution for the jointly normal distribution is also normal.
- This allows to compute the conditional distribution of X given that $Y = y$ and the result is given below.

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{1}{\sqrt{2\pi\sigma_X^2(1-\rho^2)}} \exp\left(-\frac{(x-(\mu_X+\rho\sigma_X(y-\mu_Y)/\sigma_Y))^2}{2\sigma_X^2(1-\rho^2)}\right).$$

Conditional distributions for bivariate normal distributions – parameters

Thus the conditional distribution of X given $Y = y$ is also normal (Gaussian), with parameters as follows

$$E(X|Y = y) = \mu_X + \rho\sigma_X \frac{y - \mu_Y}{\sigma_Y}$$

$$Var(X|Y = y) = \sigma_X^2(1 - \rho^2)$$

Conditional distributions for bivariate normal distributions – parameters

Thus the conditional distribution of X given $Y = y$ is also normal (Gaussian), with parameters as follows

$$E(X|Y = y) = \mu_X + \rho\sigma_X \frac{y - \mu_Y}{\sigma_Y}$$
$$Var(X|Y = y) = \sigma_X^2(1 - \rho^2)$$

- From that we see how we need to change the parameters of X to account for additional information given by Y taking value y.

Conditional distributions for bivariate normal distributions – parameters

Thus the conditional distribution of X given $Y = y$ is also normal (Gaussian), with parameters as follows

$$
\begin{aligned}
E(X|Y = y) &= \mu_X + \rho\sigma_X \frac{y - \mu_Y}{\sigma_Y} \\
Var(X|Y = y) &= \sigma_X^2(1 - \rho^2)
\end{aligned}
$$

- From that we see how we need to change the parameters of X to account for additional information given by Y taking value y.
- We note that the variance of X is reduced, so by learning about Y taking a value we reduce uncertainty about X.

Conditional distributions for bivariate normal distributions – parameters

Thus the conditional distribution of X given $Y = y$ is also normal (Gaussian), with parameters as follows

$$E(X|Y = y) = \mu_X + \rho\sigma_X \frac{y - \mu_Y}{\sigma_Y}$$
$$Var(X|Y = y) = \sigma_X^2(1 - \rho^2)$$

- From that we see how we need to change the parameters of X to account for additional information given by Y taking value y.
- We note that the variance of X is reduced, so by learning about Y taking a value we reduce uncertainty about X.
- More variables are correlated ($\rho \approx 1$) less uncertainty about X.

Conditional distributions – general case

If $\mathbf{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix}$$

are partitioned into sub-vectors $\mathbf{X}_1 : q \times 1$ and $\mathbf{X}_2 : (p-q) \times 1$ then with

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$$

the conditional distribution of $\mathbf{X}_1$ given $\mathbf{X}_2$, is

$$\mathbf{X}_1 | \mathbf{X}_2 = \mathbf{x}_2 \sim \mathcal{N}_q(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})$$

Further properties

- Conditional distributions are multivariate normal.
- The expectation of $\mathbf{X}_1$ depends (linearly) on $\mathbf{x}_2$.

$$\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2)$$

- The covariance matrix of $\mathbf{X}_1$ conditioned on $\mathbf{X}_2 = \mathbf{x}_2$ is

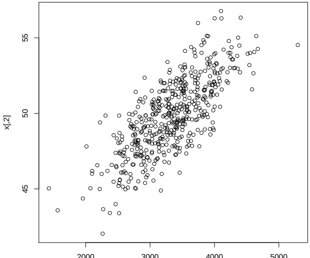
$$\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$$

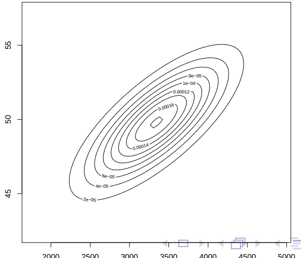
  and is independent of $\mathbf{x}_2$.

Bivariate normal data

Bivariate data are represented as pairs of points that can be represented graphically as a scatter plot.

```
library('SimDesign')
 mu=c(3343,49.8)
 Sigma=matrix(c(278784,990,990,6.25),
 byrow=T,nrow=2,ncol=2)
 x=rmvnorm(500,mu,Sigma)
 plot(x)
```





Organisation of observations

The observations in a sample are arranged in a $n \times p$ matrix $\mathbf{X}$ where n is the number of experimental units (the size of the sample) and p is the number of variables.

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2k} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots & & \vdots \\ x_{j1} & x_{j2} & \cdots & x_{jk} & \cdots & x_{jp} \\ \vdots & \vdots & & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} & \cdots & x_{np} \end{bmatrix}$$

Vector notation

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \mathbf{x}_j^T \\ \vdots \\ \mathbf{x}_n^T \end{bmatrix}$$

Row j in this matrix

$$\mathbf{x}_j^T = [x_{j1}\ x_{j2}\ \ldots\ x_{jk}\ \ldots, x_{jp}]$$

is a p-dimensional observation.

Sample mean vector

Given a sample mean for a variable i

$$\bar{x}_i = \frac{1}{n} \sum_{k=1}^{n} x_{ki}$$

we define the sample mean vector as

$$\bar{\mathbf{X}} = [\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_p]$$

Sample covariance matrix

Given sample covariances

$$s_{ij} = \frac{1}{n-1} \sum_{k=1}^{n} (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)$$

between variables i and j we define the (sample) covariance matrix

$$\mathbf{S} = \begin{bmatrix} s_{11} & s_{12} & \dots & s_{1p} \\ s_{21} & s_{22} & \dots & s_{2p} \\ \vdots & \vdots & & \vdots \\ s_{p1} & s_{p2} & \dots & s_{pp} \end{bmatrix}.$$

Matrix equation for the sample covariance matrix

$$S = \frac{1}{n-1}(X - 1_n\bar{X})^T(X - 1_n\bar{X})$$

where 1_n is the column vector consisting of n ones.

Sample correlation matrix

Given sample correlations

$$r_{ij} = \frac{s_{ij}}{\sqrt{s_{ii}s_{jj}}}$$

between variables i and j, we define the (sample) correlation matrix

$$\mathbf{R} = \begin{bmatrix} 1 & r_{12} & \dots & r_{1p} \\ r_{21} & 1 & \dots & r_{2p} \\ \vdots & \vdots & & \vdots \\ r_{p1} & r_{p2} & \dots & 1 \end{bmatrix}.$$

Estimation of μ and $\mathbf{\Sigma}$

Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be n independent observations of $\mathbf{X}$ and let

$$\bar{\mathbf{X}} = \frac{1}{n}\sum_{k=1}^{n} \mathbf{X}_k^T$$

Then

$$\mathbb{E}(\bar{\mathbf{X}}) = \frac{1}{n}\sum_{k=1}^{n} \mathbb{E}(\mathbf{X}_k^T) = \mathbb{E}(\mathbf{X}_k^T)$$

Estimation of μ and Σ

Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be n independent observations of $\mathbf{X}$ and let

$$\bar{\mathbf{X}} = \frac{1}{n}\sum_{k=1}^{n} \mathbf{X}_k^T$$

Then

$$\mathbb{E}(\bar{\mathbf{X}}) = \frac{1}{n}\sum_{k=1}^{n} \mathbb{E}(\mathbf{X}_k^T) = \mathbb{E}(\mathbf{X}_k^T)$$

Estimation of μ

The mean vector $\bar{\mathbf{X}}$ is an unbiased estimator of μ^T.

Estimation of μ and $\boldsymbol{\Sigma}$

Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be n independent observations of $\mathbf{X}$ and let

$$\bar{\mathbf{X}} = \frac{1}{n}\sum_{k=1}^{n} \mathbf{X}_k^T$$

Then

$$\mathbb{E}(\bar{\mathbf{X}}) = \frac{1}{n}\sum_{k=1}^{n} \mathbb{E}(\mathbf{X}_k^T) = \mathbb{E}(\mathbf{X}_k^T)$$

Estimation of μ

The mean vector $\bar{\mathbf{X}}$ is an unbiased estimator of μ^T.

Estimation of $\boldsymbol{\Sigma}$

It holds that

$$\mathbb{E}(\mathbf{S}) = \boldsymbol{\Sigma}$$

Consistency for $\bar{\mathbf{X}}$ and $\mathbf{S}$

Multivariate versions of LLN (law of large numbers) and CLT (central limit theorem)

Given a random sample $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ from a distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

Consistency for $\bar{\mathbf{X}}$ and $\mathbf{S}$

Multivariate versions of LLN (law of large numbers) and CLT (central limit theorem)

Given a random sample $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ from a distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

LLN:
$\bar{\mathbf{X}}$ converges in probability to $\boldsymbol{\mu}^T$.
$\mathbf{S}$ converges in probability to $\boldsymbol{\Sigma}$.

Maximum likelihood estimates

- Based on observations from the multivariate normal distributions the likelihood for $\boldsymbol{\mu}, \boldsymbol{\Sigma}$ is

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

Maximum likelihood estimates

- Based on observations from the multivariate normal distributions the likelihood for $\boldsymbol{\mu}, \boldsymbol{\Sigma}$ is

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

- Maximum likelihood estimates of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are obtained by maximizing the likelihood function.

Maximum likelihood estimates

- Based on observations from the multivariate normal distributions the likelihood for $\boldsymbol{\mu}, \boldsymbol{\Sigma}$ is

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

- Maximum likelihood estimates of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are obtained by maximizing the likelihood function.
- The maximum likelihood estimates of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are given by

$$\hat{\boldsymbol{\mu}} = \bar{\mathbf{X}}^T$$

$$\hat{\boldsymbol{\Sigma}} = \frac{n-1}{n}\mathbf{S} = \mathbf{S}_n$$

Wishart distribution

- The distribution of
$$\sum_{k=1}^{m} \mathbf{X}_j \mathbf{X}_j^T$$
where $\mathbf{X}_j \sim \mathcal{N}_p(\mathbf{0}, \mathbf{\Sigma})$ is called a *p dimensional Wishart distribution with m degrees of freedom*

Wishart distribution

- The distribution of
$$\sum_{k=1}^{m} \mathbf{X}_j \mathbf{X}_j^T$$
where $\mathbf{X}_j \sim \mathcal{N}_p(\mathbf{0}, \mathbf{\Sigma})$ is called a *p dimensional Wishart distribution with m degrees of freedom*
- If rows of $X_{m \times p}$ are independent random vectors from $\mathcal{N}_p(\mathbf{0}, \mathbf{\Sigma})$ then $X^T X$ has the p dimensional Wishart distribution with m degrees of freedom.
- We use the notation $W_p(m, \mathbf{\Sigma})$ for a Wishart distribution with m degrees of freedom

Wishart distribution

- The distribution of

$$\sum_{k=1}^{m} \mathbf{X}_j \mathbf{X}_j^T$$

  where $\mathbf{X}_j \sim \mathcal{N}_p(\mathbf{0}, \mathbf{\Sigma})$ is called a *p dimensional Wishart distribution with m degrees of freedom*
- If rows of $X_{m\times p}$ are independent random vectors from $\mathcal{N}_p(\mathbf{0}, \mathbf{\Sigma})$ then $X^T X$ has the p dimensional Wishart distribution with m degrees of freedom.
- We use the notation $W_p(m, \mathbf{\Sigma})$ for a Wishart distribution with m degrees of freedom
- It is a multivariate generalization of the chi-square distribution.

Simultaneous distribution for $\bar{\mathbf{X}}$ and $\mathbf{S}$

It holds that if samples are from multivariate normal distributions

- $\bar{\mathbf{X}}^T \sim \mathcal{N}_p(\boldsymbol{\mu}, \frac{1}{n}\boldsymbol{\Sigma})$
- $(n-1)\mathbf{S} \sim W_p(n-1, \boldsymbol{\Sigma})$
- $\bar{\mathbf{X}}$ and $\mathbf{S}$ are independent