SF2940: Probability theory
Lecture 6: Multivariate Normal Distribution

Timo Koski

17.9.2008

KTH Matematik

Lecture: Contents

- Random vectors, mean vector, covariance matrix, rules of transformation
- Multivariate normal RV, moment generating functions, characteristic function, rules of transformation
- Density of a multivariate normal RV
- Joint PDF of bivariate normal RVs
- Conditional distributions

PART 1: Mean vector, Covariance matrix, MGF, Characteristic function

KTH Matematik

Vector Notation: Random Vector

A random vector $\mathbf{X}$ is a column vector

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} = (X_1, X_2, \ldots, X_n)^T$$

Each X_i is a random variable.

Sample Value Random Vector

A column vector

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = (x_1, x_2, \ldots, x_n)^T$$

We can think of x_i is an outcome of X_i.

KTH Matematik

Joint CDF, Joint PDF

The joint CDF (=cumulative distribution function) of a continuous random vector $\mathbf{X}$ is

$$F_{\mathbf{X}}(\mathbf{x}) = F_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = P(\mathbf{X} \leq \mathbf{x}) =$$

$$= P(X_1 \leq x_1, \ldots, X_n \leq x_n)$$

Joint probability density function (PDF)

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{\partial^n}{\partial x_1 \ldots \partial x_n} F_{X_1,\ldots,X_n}(x_1,\ldots,x_n)$$

KTH Matematik

Mean Vector

$$\mu_{\mathbf{X}} = E[\mathbf{X}] = \begin{pmatrix} E[X_1] \\ E[X_2] \\ \vdots \\ E[X_n] \end{pmatrix},$$

a column vector of means (=expectations) of **X**.

KTH Matematik

If $\mathbf{X}^T$ is the transposed column vector (=a row vector), then

$$\mathbf{X}\mathbf{X}^T$$

is a $n \times n$ matrix, and

$$\mathbf{X}^T\mathbf{X}$$

is a scalar product, a real number.

Covariance Matrix of A Random Vector

Covariance matrix

$$C_{\mathbf{X}} := E\left[\left(\mathbf{X} - \mu_{\mathbf{X}}\right)\left(\mathbf{X} - \mu_{\mathbf{X}}\right)^{T}\right]$$

where the element (i, j)

$$C_{\mathbf{X}}(i,j) = E\left[\left(X_i - \mu_i\right)\left(X_j - \mu_j\right)\right]$$

is the covariance of X_i and X_j.

KTH Matematik

A Quadratic Form

$$\mathbf{x}^T C_{\mathbf{X}} \mathbf{x} = \sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j C_{\mathbf{X}}(i,j).$$

We see that

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j E\left[(X_i - \mu_i)(X_j - \mu_j)\right]$$

$$= E\left[\sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j (X_i - \mu_i)(X_j - \mu_j)\right] \quad (*)$$

KTH Matematik

Properties of a Covariance Matrix

- Covariance matrix is nonnegative definite, i.e., for all $\mathbf{x}$ we have

$$\mathbf{x}^T C_{\mathbf{X}} \mathbf{x} \geq 0$$

  Hence

$$\det C_{\mathbf{X}} \geq 0.$$

- The covariance matrix is symmetric

$$C_{\mathbf{X}} = C_{\mathbf{X}}^T$$

Properties of a Covariance Matrix

The covariance matrix is symmetric

$$C_{\mathbf{X}} = C_{\mathbf{X}}^T$$

since

$$C_{\mathbf{X}}(i,j) = E\left[(X_i - \mu_i)(X_j - \mu_j)\right]$$
$$= E\left[(X_j - \mu_j)(X_i - \mu_i)\right] = C_{\mathbf{X}}(j,i)$$

KTH Matematik

Properties of a Covariance Matrix

A covariance matrix is positive definite,

$$\mathbf{x}^T C_{\mathbf{X}} \mathbf{x} > 0$$

for all $\mathbf{x} \neq \mathbf{0}$ iff

$$\det C_{\mathbf{X}} > 0$$

(i.e. $C_{\mathbf{X}}$ is invertible).

Proposition

$$\mathbf{x}^T C_{\mathbf{X}} \mathbf{x} \geq 0$$

Pf: By $(*)$ above

$$\mathbf{x}^T C_{\mathbf{X}} \mathbf{x} = \mathbf{x}^T E\left[\left(\mathbf{X} - \mu_{\mathbf{X}}\right)\left(\mathbf{X} - \mu_{\mathbf{X}}\right)^T \right] \mathbf{x}$$

$$= E\left[\mathbf{x}^T \left(\mathbf{X} - \mu_{\mathbf{X}}\right)\left(\mathbf{X} - \mu_{\mathbf{X}}\right)^T \mathbf{x} \right] = E\left[\mathbf{x}^T \mathbf{w} \cdot \mathbf{w}^T \mathbf{x} \right]$$

where we have set $\mathbf{w} = \left(\mathbf{X} - \mu_{\mathbf{X}}\right)$. Then by linear algebra $\mathbf{x}^T \mathbf{w} = \mathbf{w}^T \mathbf{x}$ $= \sum_{i=1}^{n} w_i x_i$. Hence

$$E\left[\mathbf{x}^T \mathbf{w} \mathbf{w}^T \mathbf{x} \right] = E\left[\left(\sum_{i=1}^{n} w_i x_i \right)^2 \right] \geq 0.$$

Coefficient of Correlation

The *Coefficient of Correlation* ρ of X and Y is defined as

$$\rho := \rho_{X,Y} := \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X) \cdot \mathrm{Var}(Y)}},$$

where $\mathrm{Cov}(X,Y) = E\left[(X - \mu_X)(Y - \mu_Y)\right]$. This is normalized

$$-1 \leq \rho_{X,Y} \leq 1$$

For random variables X and Y,

- $\mathrm{Cov}(X,Y) = \rho_{X,Y} = 0$ does not always mean that X, Y are independent.

Special case: Covariance Matrix of A Bivariate Vector

$\mathbf{X} = (X_1, X_2)^T$.

$$C_{\mathbf{X}} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix},$$

where ρ is the coefficient of correlation of X_1 and X_2, and $\sigma_1^2 = \mathrm{Var}(X_1)$, $\sigma_2^2 = \mathrm{Var}(X_2)$. $C_{\mathbf{X}}$ is invertible iff $\rho^2 \neq 1$, for proof we note that

$$\det C_{\mathbf{X}} = \sigma_1^2\sigma_2^2\left(1 - \rho^2\right)$$

Special case: Covariance Matrix of A Bivariate Vector

$$\mathbf{\Lambda} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix},$$

if $\rho^2 \neq 1$, the inverse exists and

$$\mathbf{\Lambda}^{-1} = \frac{1}{\sigma_1^2\sigma_1^2(1-\rho^2)} \begin{pmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix},$$

KTH Matematik

$\mathbf{Y} = B\mathbf{X} + \mathbf{b}$

Proposition

$\mathbf{X}$ is a random vector with mean vector $\mu_{\mathbf{X}}$ and covariance matrix $C_{\mathbf{X}}$. B is a $m \times n$ matrix. If $\mathbf{Y} = B\mathbf{X} + \mathbf{b}$, then

$$E\mathbf{Y} = B\mu_{\mathbf{X}} + \mathbf{b}$$

$$C_{\mathbf{Y}} = BC_{\mathbf{X}}B^T$$

Pf: For simplicity of writing, take $\mathbf{b} = \mu = \mathbf{0}$. Then

$$C_{\mathbf{Y}} = E\mathbf{Y}\mathbf{Y}^T = EB\mathbf{X}\left(B\mathbf{X}\right)^T =$$

$$= EB\mathbf{X}\mathbf{X}^T B^T = BE\left[\mathbf{X}\mathbf{X}^T\right]B^T = BC_{\mathbf{X}}B^T$$

Moment Generating and Characteristic Functions

Definition

Moment generating function of $\mathbf{X}$ *is defined as*

$$\psi_{\mathbf{X}}(\mathbf{t}) \stackrel{\text{def}}{=} Ee^{\mathbf{t}^T\mathbf{X}} = Ee^{t_1X_1+t_2X_2+\cdots+t_nX_n}$$

Definition

Characteristic function of $\mathbf{X}$ *is defined as*

$$\varphi_{\mathbf{X}}(\mathbf{t}) \stackrel{\text{def}}{=} Ee^{i\mathbf{t}^T\mathbf{X}} = Ee^{i(t_1X_1+t_2X_2+\cdots+t_nX_n)}$$

Special cases: take $t_1 = 1, t_2 = t_3 = \ldots = t_n = 0$, then $\varphi_{\mathbf{X}}(\mathbf{t}) = \varphi_{X_1}(t_1)$.

PART 2: Def I of a multivariate normal distribution

We recall first some of the properties of univariate normal distribution

Normal (Gaussian) One-dimensional RVs

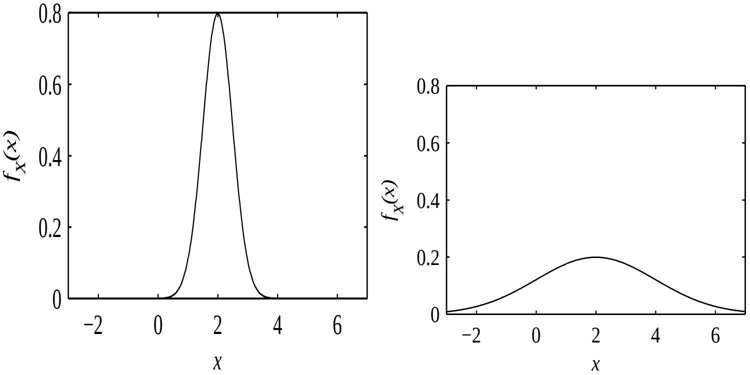
- X is a normal random variable if

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

  where μ is real and $\sigma > 0$.
- Notation: $X \in N(\mu, \sigma^2)$
- Properties: $E(X) = \mu$, $\mathrm{Var} = \sigma^2$

KTH Matematik

Normal (Gaussian) One-dimensional RVs



(a) $\mu = 2,\ \sigma = 1/2$, **(b)** $\mu = 2,\ \sigma = 2$

Linear Transformation

- $X \in N(\mu_X, \sigma^2) \Rightarrow Y = aX + b$ is $N(a\mu_X + b, a^2\sigma^2)$
- Thus $Z = \frac{X-\mu_X}{\sigma_X} \in N(0,1)$ and

$$\mathrm{P}(X \leq x) = \mathrm{P}\left(\frac{X - \mu_X}{\sigma_X} \leq \frac{x - \mu_X}{\sigma_X}\right)$$

or

$$F_X(x) = \mathrm{P}\left(Z \leq \frac{x - \mu_X}{\sigma_X}\right) = \Phi\left(\frac{x - \mu_X}{\sigma_X}\right)$$

KTH Matematik

Normal (Gaussian) One-dimensional RVs

$X \in N(\mu, \sigma^2)$ then the moment generating function is

$$\psi_X(t) = E\left[e^{tX}\right] = e^{t\mu + \frac{1}{2}t^2\sigma^2},$$

and the characteristic function is

$$\varphi_X(t) = E\left[e^{itX}\right] = e^{it\mu - \frac{1}{2}t^2\sigma^2}$$

as found in Lectures 4 and 5.

Multivariate Normal Def. I

Definition

An $n \times 1$ random vector $\mathbf{X}$ has a normal distribution iff for **every** *$n \times 1$-vector* **a** *the one-dimensional random vector $a^T\mathbf{X}$ has a normal distribution.*

We write $\mathbf{X} \in N(\mu, \mathbf{\Lambda})$, when μ is the mean vector and $\mathbf{\Lambda}$ is the covariance matrix.

Consequences of Def. I (1)

An $n \times 1$ vector $\mathbf{X} \in N(\mu, \mathbf{\Lambda})$ iff the one-dimensional random vector $\mathbf{a}^T\mathbf{X}$ has a normal distribution for **every** n-vector $\mathbf{a}$.
Now we know that (take $\mathbf{B} = \mathbf{a}^T$ in the preceding)

$$Ea^T\mathbf{X} = \mathbf{a}^T\mu, \mathrm{Var}\left[\mathbf{a}^T\mathbf{X}\right] = \mathbf{a}^T\mathbf{\Lambda}\mathbf{a}$$

KTH Matematik

Consequences of Def. I (2)

Hence, if $Y = \mathbf{a}^T\mathbf{X}$, then $Y \in N\left(\mathbf{a}^T\mu, \mathbf{a}^T\mathbf{\Lambda}\mathbf{a}\right)$ and the moment generating function of Y is

$$\psi_Y(t) = E\left[e^{tY}\right] = e^{t\mathbf{a}^T\mu + \frac{1}{2}t^2\mathbf{a}^T\mathbf{\Lambda}\mathbf{a}}.$$

Therefore

$$\psi_{\mathbf{X}}(\mathbf{a}) = Ee^{\mathbf{a}^T\mathbf{X}} = \psi_Y(1) = e^{\mathbf{a}^T\mu + \frac{1}{2}\mathbf{a}^T\mathbf{\Lambda}\mathbf{a}}.$$

Hence we have shown that if $\mathbf{X} \in N(\mu, \mathbf{\Lambda})$, then

$$\psi_{\mathbf{X}}(\mathbf{t}) = Ee^{\mathbf{t}^T\mathbf{X}} = e^{\mathbf{t}^T\mu + \frac{1}{2}\mathbf{t}^T\mathbf{\Lambda}\mathbf{t}}.$$

is the moment generating function of $\mathbf{X}$.

Consequences of Def. I (4)

In the same way we can find that

$$\varphi_{\mathbf{X}}(\mathbf{t}) = Ee^{i\mathbf{t}^T\mathbf{X}} = e^{i\mathbf{t}^T\mu - \frac{1}{2}\mathbf{t}^T\mathbf{\Lambda}\mathbf{t}}.$$

is the characteristic function of $\mathbf{X} \in N(\mu, \mathbf{\Lambda})$.

Consequences of Def. I (5)

Let $\mathbf{\Lambda}$ be a diagonal covariance matrix with λ_i^2s on the main diagonal, i.e.,

$$\mathbf{\Lambda} = \begin{pmatrix} \lambda_1^2 & 0 & 0 & \dots & 0 \\ 0 & \lambda_2^2 & 0 & \dots & 0 \\ 0 & 0 & \lambda_3^2 & \dots & 0 \\ 0 & \ddots & \vdots & \dots & 0 \\ 0 & 0 & 0 & \dots & \lambda_n^2 \end{pmatrix},$$

Proposition

If $\mathbf{X} \in N(\mu, \mathbf{\Lambda})$, then $X_1, X_2, \dots, X_n$ are independent normal variables.

KTH Matematik

Consequences of Def. I (6)

Pf: $\mathbf{\Lambda}$ is diagonal, the quadratic form becomes a single sum of squares.

$$\varphi_{\mathbf{X}}(\mathbf{t}) = e^{i\mathbf{t}^T\mu - \frac{1}{2}\mathbf{t}^T\mathbf{\Lambda}\mathbf{t}} =$$

$$= e^{i\sum_{i=1}^{n}\mu_i t_i - \frac{1}{2}\sum_{i=1}^{n}\lambda_i^2 t_i^2}$$

$$= e^{i\mu_1 t_1 - \frac{1}{2}\lambda_1^2 t_1^2} e^{i\mu_2 t_2 - \frac{1}{2}\lambda_2^2 t_2^2} \cdots e^{i\mu_n t_n - \frac{1}{2}\lambda_n^2 t_n^2}$$

is the product of the characteristic functions of $X_i \in N\left(\mu_i, \lambda_i^2\right)$, which are thus seen to be independent $N\left(\mu_i, \lambda_i^2\right)$. $\square$

KTH VETENSKAP OCH KONST

KTH Matematik

Further properties of the multivariate normal

$\mathbf{X} \in N(\mu, \mathbf{\Lambda})$

- Every component X_k is one-dimensional normal. To prove this we take

$$\mathbf{a} = (0, 0, \ldots, \underbrace{1}_{\text{position } k}, 0, \ldots, 0)^T$$

  and the conclusion follows by Def. I.

- $X_1 + X_2 + \cdots X_n$ is one-dimensional normal. Note: The terms in the sum need not be independent.

$\mathbf{X} \in N(\mu, \mathbf{\Lambda})$

- Every marginal distribution is normal. To prove this we consider any k variables $X_{i_1}, X_{i_2} \ldots X_{i_k}$ and then take $\mathbf{a}$ such that $a_j = 0$ for $j \neq i_1, \ldots i_k$ and then apply Def. I.

Proposition

$\mathbf{X} \in N(\mu, \mathbf{\Lambda})$ *and* $\mathbf{Y} = B\mathbf{X} + \mathbf{b}$. *Then*

$$\mathbf{Y} \in N\left(B\mu + \mathbf{b}, B\mathbf{\Lambda}B^T\right).$$

Pf:

$$\psi_{\mathbf{Y}}(\mathbf{s}) = E\left[e^{\mathbf{s}^T\mathbf{Y}}\right] = E\left[e^{\mathbf{s}^T(\mathbf{b}+B\mathbf{X})}\right] =$$

$$= e^{\mathbf{s}^T\mathbf{b}} E\left[e^{\mathbf{s}^T B\mathbf{X}}\right] = e^{\mathbf{s}^T\mathbf{b}} E\left[e^{\left(B^T\mathbf{s}\right)^T\mathbf{X}}\right]$$

$$E\left[e^{\left(B^T\mathbf{s}\right)^T\mathbf{X}}\right] = \psi_{\mathbf{X}}\left(B^T\mathbf{s}\right).$$

KTH Matematik

Properties of multivariate normal

$\mathbf{X} \in N(\mu, \mathbf{\Lambda})$

$$\psi_{\mathbf{X}}\left(B^T\mathbf{s}\right) = e^{\left(B^T\mathbf{s}\right)^T\mu + \frac{1}{2}\left(B^T\mathbf{s}\right)^T\mathbf{\Lambda}\left(B^T\mathbf{s}\right)}.$$

$$\left(B^T\mathbf{s}\right)^T\mu = \mathbf{s}^T B\mu,$$

$$\left(B^T\mathbf{s}\right)^T\mathbf{\Lambda}\left(B^T\mathbf{s}\right) = \mathbf{s}^T B\mathbf{\Lambda}B^T\mathbf{s},$$

$$e^{\left(B^T\mathbf{s}\right)^T\mu + \frac{1}{2}\left(B^T\mathbf{s}\right)^T\mathbf{\Lambda}\left(B^T\mathbf{s}\right)} = e^{\mathbf{s}^T B\mu + \frac{1}{2}\mathbf{s}^T B\mathbf{\Lambda}B^T\mathbf{s}}$$

KTH Matematik

Properties of multivariate normal

$$\psi_{\mathbf{X}}\left(B^T\mathbf{s}\right) = e^{\mathbf{s}^T B\mu + \frac{1}{2}\mathbf{s}^T B\mathbf{\Lambda} B^T\mathbf{s}}.$$

$$\psi_{\mathbf{Y}}(\mathbf{s}) = e^{\mathbf{s}^T\mathbf{b}}\psi_{\mathbf{X}}\left(B^T\mathbf{s}\right) = e^{\mathbf{s}^T\mathbf{b}} e^{\mathbf{s}^T B\mu + \frac{1}{2}\mathbf{s}^T B\mathbf{\Lambda} B^T\mathbf{s}}$$

$$\psi_{\mathbf{Y}}(\mathbf{s}) = e^{\mathbf{s}^T(\mathbf{b}+B\mu) + \frac{1}{2}\mathbf{s}^T B\mathbf{\Lambda} B^T\mathbf{s}},$$

which proves the claim as asserted. $\square$

KTH Matematik

PART 3: Multivariate normal, Def. II: characteristic function, DEF III: density

Multivariate normal, Def. II: char. fnctn

Definition

A random vector $\mathbf{X}$ *with mean vector* μ *and a covariance matrix* $\mathbf{\Lambda}$ *is* $N(\mu, \mathbf{\Lambda})$ *if its characteristic function is*

$$\varphi_{\mathbf{X}}(\mathbf{t}) = Ee^{i\mathbf{t}^T\mathbf{X}} = e^{i\mathbf{t}^T\mu - \frac{1}{2}\mathbf{t}^T\mathbf{\Lambda}\mathbf{t}}.$$

Multivariate normal, Def. II implies Def. I

We need to show that the one-dimensional random vector $Y = \mathbf{a}^T\mathbf{X}$ has a normal distribution.

$$\varphi_Y(t) = E\left[e^{itY}\right] = E\left[e^{it\sum_{i=1}^{n} a_i \cdot X_i}\right] =$$

$$= E\left[e^{it\mathbf{a}^T\mathbf{X}}\right] = \varphi_{\mathbf{X}}\left(t\mathbf{a}\right) =$$

$$= e^{it\mathbf{a}^T\mu - \frac{1}{2}t^2\mathbf{a}^T\mathbf{\Lambda}\mathbf{a}}$$

and this is the characteristic function of $N\left(\mathbf{a}^T\mu, \mathbf{a}^T\mathbf{\Lambda}\mathbf{a}\right)$.

KTH Matematik

Multivariate normal, Def. III: joint PDF

Definition

A random vector $\mathbf{X}$ *with mean vector* μ *and an invertible covariance matrix* $\mathbf{\Lambda}$ *is* $N(\mu, \mathbf{\Lambda})$, *if the density is*

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det(\mathbf{\Lambda})}} e^{-\frac{1}{2}(\mathbf{x}-\mu)^T \mathbf{\Lambda}^{-1}(\mathbf{x}-\mu)}$$

KTH Matematik

Multivariate normal

It can be checked by a computation taht

$$e^{i\mathbf{t}^T\mu-\frac{1}{2}\mathbf{t}^T\mathbf{\Lambda}\mathbf{t}} = \int_{R^n} e^{i\mathbf{t}^T\mathbf{x}} \frac{1}{(2\pi)^{n/2}\sqrt{\det(\mathbf{\Lambda})}} e^{-\frac{1}{2}(\mathbf{x}-\mu)^T\mathbf{\Lambda}^{-1}(\mathbf{x}-\mu)} d\mathbf{x}$$

(complete the square) Hence Def. III implies the property in Def. II. The three definitions are equivalent, in the case inverse of the covariance matrix exists.

KTH Matematik

PART 4: Bivariate normal with density

Multivariate Normal: the bivariate case

As soon as $\rho^2 \neq 1$, the matrix

$$\mathbf{\Lambda} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix},$$

is invertible, and the inverse is

$$\mathbf{\Lambda}^{-1} = \frac{1}{\sigma_1^2\sigma_1^2(1-\rho^2)} \begin{pmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix},$$

Multivariate Normal: the bivariate case

$\rho^2 \neq 1$, and $\mathbf{X} = (X_1, X_2)^T$, then

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{2\pi)\sqrt{\det \mathbf{\Lambda}}} e^{-\frac{1}{2}(\mathbf{x}-\mu_{\mathbf{X}})^T \mathbf{\Lambda}^{-1}(\mathbf{x}-\mu_{\mathbf{X}})}$$

$$= \frac{1}{(2\pi)\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2}Q(x_1,x_2)}$$

Multivariate Normal: the bivariate case

where

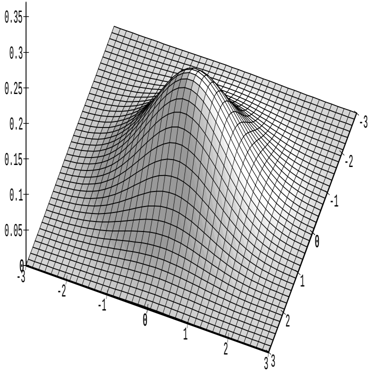
$$Q(x_1, x_2) =$$

$$\frac{1}{(1-\rho^2)} \cdot \left[\left(\frac{x_1 - \mu_1}{\sigma_1} \right)^2 - \frac{2\rho(x_1 - \mu_1)(x_2 - \mu_2)}{\sigma_1 \sigma_2} + \left(\frac{x_2 - \mu_2}{\sigma_2} \right)^2 \right]$$

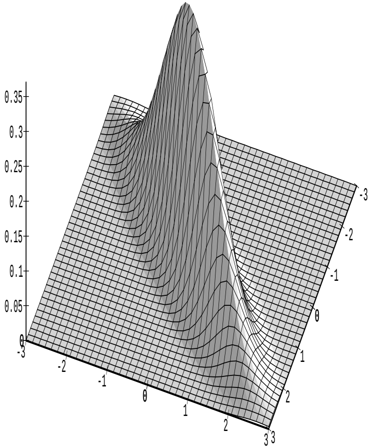
For this, invert the matrix Λ and expand the quadratic form !

KTH Matematik

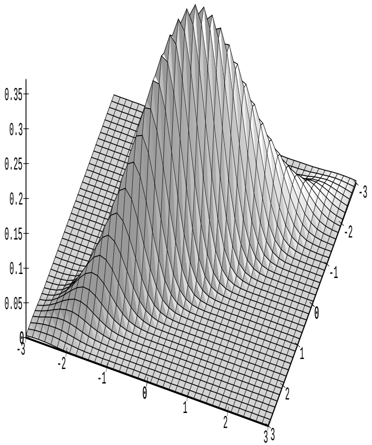
$\rho = 0$



$\rho = 0.9$



$\rho = -0.9$



Conditional densities for the bivariate normal

Complete the square of the exponent to write

$$f_{X,Y}(x,y) = f_X(x) f_{Y|X}(y)$$

where

$$f_X(x) = \frac{1}{\sigma_1\sqrt{2\pi}} e^{-\frac{1}{2\sigma_1^2}(x-\mu_1)^2}$$

$$f_{Y|X}(y) = \frac{1}{\tilde{\sigma}_2\sqrt{2\pi}} e^{-\frac{1}{2\tilde{\sigma}_2^2}(y-\tilde{\mu}_2(x))^2}$$

$$\tilde{\mu}_2(x) = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x-\mu_1), \tilde{\sigma}_2 = \sigma_2\sqrt{1-\rho^2}$$

Bivariate normal properties

- $E(X) = \mu_1$
- Given $X = x$, Y is Gaussian
- Conditional mean of Y given $X = x$:

$$\tilde{\mu}_2(x) = \mu_2 + \rho \frac{\sigma_2}{\sigma_1}(x - \mu_1) = E(Y|X = x)$$

- Conditional variance of Y given $X = x$:

$$\mathrm{Var}(Y|X = x) = \sigma_2^2 \left(1 - \rho^2\right)$$

KTH VETENSKAP OCH KONST
KTH Matematik

Bivariate normal properties

- Conditional mean of Y given $X = x$:

$$\tilde{\mu}_2(x) = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x - \mu_1) = E(Y|X = x)$$

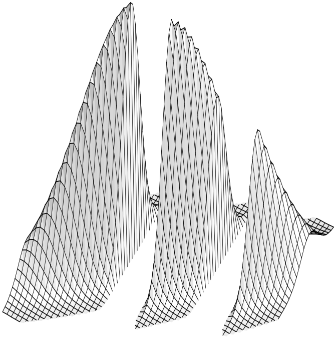
- Conditional variance of Y given $X = x$:

$$\mathrm{Var}(Y|X = x) = \sigma_2^2\left(1 - \rho^2\right)$$

Recall Theorem. 5.2. and Theorem. 5.3 in A. Gut: An Intermediate Course in Probability. Hence the conditional mean of Y given X variable in a bivariate normal distribution is also the best LINEAR predictor of Y based on X, and the conditional variance is the residual variance.

KTH Matematik

Marginal PDFs



Proof of conditional pdf

Consider

$$\frac{f_{X,Y}(x,y)}{f_X(x)} =$$

$$\frac{\sigma_1\sqrt{2\pi}}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2}Q(x,y)+\frac{1}{2\sigma_1^2}(x-\mu_1)^2}$$

KTH Matematik

Proof of conditional pdf

$$-\frac{1}{2}Q(x,y)+\frac{1}{2\sigma_1^2}(x-\mu_1)^2$$
$$=-\frac{1}{2}H(x,y),$$

Proof of conditional pdfs

$$H(x,y) =$$

$$\frac{1}{(1-\rho^2)} \cdot \left[\left(\frac{x-\mu_1}{\sigma_1} \right)^2 - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \left(\frac{y-\mu_2}{\sigma_2} \right)^2 \right]$$

$$- \left(\frac{x-\mu_1}{\sigma_1} \right)^2$$

KTH Matematik

Proof of conditional pdf

$$H(x,y) =$$

$$\frac{\rho^2}{(1-\rho^2)}\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2(1-\rho^2)} + \left(\frac{y-\mu_2}{\sigma_2^2(1-\rho^2)}\right)^2$$

KTH Matematik

Proof of conditional pdf

$$H(x,y) = \frac{\left(y - \mu_2 - \rho\frac{\sigma_2}{\sigma_1}(x-\mu_1)\right)^2}{\sigma_2^2(1-\rho^2)}$$

Conditional pdf

$$\frac{f_{X,Y}(x,y)}{f_X(x)} =$$

$$\frac{1}{\sqrt{1-\rho^2}\sigma_2\sqrt{2\pi}} e^{\left[-\frac{1}{2}\frac{\left(y-\mu_2-\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1)\right)^2}{\sigma_2^2(1-\rho^2)}\right]}$$

This establishes the bivariate normal properties claimed above.

KTH Matematik

Proposition

(X, Y) *bivariate normal* $\Rightarrow \rho = \rho_{X,Y}$

Proof:

$$
\begin{aligned}
& E\left[(X-\mu_1)(Y-\mu_2)\right] \\
&= E(E([(X-\mu_1)(Y-\mu_2)]\,|X)) \\
&= E((X-\mu_1)E\left[Y-\mu_2\right]|X))
\end{aligned}
$$

Bivariate normal properties : ρ

$$
\begin{aligned}
&= E((X-\mu_1)E\left[(Y-\mu_2)\right]|X)) \\
&= E(X-\mu_1)\left[E(Y|X)-\mu_2\right] \\
&= E((X-\mu_1)\left[\mu_2+\rho\frac{\sigma_2}{\sigma_1}(X-\mu_1)-\mu_2\right] \\
&= \rho\frac{\sigma_2}{\sigma_1}E(X-\mu_1)((X-\mu_1))
\end{aligned}
$$

KTH Matematik

Bivariate normal properties : ρ

$$
\begin{aligned}
&= \rho\frac{\sigma_2}{\sigma_1}E(X-\mu_1)(X-\mu_1)\\
&= \rho\frac{\sigma_2}{\sigma_1}E(X-\mu_1)^2\\
&= \rho\frac{\sigma_2}{\sigma_1}\sigma_1^2\\
&= \rho\sigma_2\sigma_1
\end{aligned}
$$

Bivariate normal properties : ρ

In other words we have checked that

$$\rho = \frac{E\left[(X-\mu_1)(Y-\mu_2)\right]}{\sigma_2\sigma_1}$$

$\square$

$\rho = 0 \Leftrightarrow$ bivariate normal X, Y are independent.

PART 5: Generating a multivariate normal variable

Standard Normal Vector: definition

$\mathbf{Z} \in N(\mathbf{0}, \mathbf{I})$ is a standard normal vector.
$\mathbf{I}$ is the $n \times n$ identity matrix.

$$f_{\mathbf{Z}}(\mathbf{z}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det(\mathbf{I})}} e^{-\frac{1}{2}(\mathbf{z}-\mathbf{0})^T \mathbf{I}^{-1}(\mathbf{z}-\mathbf{0})}$$

$$= \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}\mathbf{z}^T\mathbf{z}}$$

KTH Matematik

Distribution of $\mathbf{X} = A\mathbf{Z} + \mathbf{b}$

$\mathbf{X} = A\mathbf{Z} + \mathbf{b}$, $\mathbf{Z}$ is standard Gaussian, then

$$\mathbf{X} = N\left(\mathbf{b}, AA^T\right)$$

(follows by a rule in the preceding)

KTH Matematik

Multivariate Normal: the bivariate case

If

$$\mathbf{\Lambda} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix},$$

then $\mathbf{\Lambda} = AA^T$, where

$$A = \begin{pmatrix} \sigma_1 & 0 \\ \rho\sigma_2 & \sigma_2\sqrt{1-\rho^2} \end{pmatrix},$$

Standard Normal Vector

$\mathbf{X} \in N\left(\mu_{\mathbf{X}}, \mathbf{\Lambda}\right)$, and A is such that

$$\mathbf{\Lambda} = AA^{T}$$

(An invertible matrix A with this property exists always, if Λ is positive definite (we need the symmetry of $\mathbf{\Lambda}$, too.) Then

$$\mathbf{Z} = A^{-1}\left(\mathbf{X} - \mu_{\mathbf{X}}\right)$$

is a standard Gaussian vector.
Proof: We give the first idea of his proof, a rule of transformation.

KTH Matematik

Rule of transformation

If $\mathbf{X}$ has density $f_{\mathbf{X}}(\mathbf{x})$, $Y = A\mathbf{X} + \mathbf{b}$,
A is invertible, then

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{|\det A|} f_{\mathbf{X}}\left(A^{-1}(\mathbf{y} - \mathbf{b})\right)$$

Note that if $\mathbf{\Lambda} = AA^T$, then

$$\det \mathbf{\Lambda} = \det A \cdot \det A^T = \det A \cdot \det A = \det A^2,$$

so that $|\det A| = \sqrt{\det \mathbf{\Lambda}}$.

KTH Matematik