Gaussian random vectors

October 11, 2011

The weak law of large numbers

Theorem

Let $X_1, X_2, \dots, X_k, \dots$ be a sequence of independent and identically distributed r.v., each having $\mathrm{E}(X_k) = m$ and finite variance.

Let

$$\overline{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n}, \quad n \geq 1.$$

Then, for all $\epsilon > 0$,

$$\lim_{n\to\infty} \mathrm{P}\left(\left|\overline{X}_n - m\right| \geq \epsilon\right) = 0$$

- The r.v. $\overline{X}_n$ has mean m:

$$\mathrm{E}\left(\overline{X}_n\right) = \mathrm{E}\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right)$$

$$= \frac{\mathrm{E}(X_1) + \mathrm{E}(X_2) + \cdots + \mathrm{E}(X_n)}{n}$$

$$= \frac{nm}{n} = m$$

- If $\mathrm{Var}(X_k) = \sigma^2$, then the variance of $\overline{X}_n$ is σ^2/n.

$$\mathrm{Var}\left(\overline{X}_n\right) = \mathrm{E}\left(\left(\overline{X}_n - m\right)^2\right) = \frac{1}{n^2}\,\mathrm{E}\left(\left(\sum_{k=1}^{n}(X_k - m)\right)^2\right)$$

$$= \frac{1}{n^2}\left(\sum_{k=1}^{n}\mathrm{E}\left((X_k - m)^2\right) + 2\sum_{1\le i<j\le n}\mathrm{E}\left((X_i - m)(X_j - m)\right)\right)$$

$$= \frac{\sigma^2}{n}$$

  - Notice that $\mathrm{E}\left((X_i - m)(X_j - m)\right) = 0$ because the r.v. X_i are independent, and, therefore, they are pairwise uncorrelated.

By the Chebyshev's inequality:

$$\mathrm{P}\left(\left|\overline{X}_n - m\right| \ge \epsilon\right) \le \frac{\sigma^2/n}{\epsilon^2}.$$

Hence, as $n \to \infty$,

$$\mathrm{P}\left(\left|\overline{X}_n - m\right| \ge \epsilon\right) \to 0$$

Example: Probability as the limit of the relative frequency

- Let A be a given event with $\mathrm{P}(A) = p$.
- Le us repeat n times the random experiment (independent repetitions).
- Let X_k be the indicator of the event "A happens in the k-th repetition". Hence,

$$\mathrm{E}(X_k) = \mathrm{P}(A) = p.$$

- Then,

$$\overline{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n} = f_n(A)$$

is the relative frequency of the event A.

- Therefore, for all $\epsilon > 0$,

$$\mathrm{P}\left(|f_n(A) - p| \ge \epsilon\right) \to 0 \quad \text{as } n \to \infty$$

Equivalently,

$$\mathrm{P}\left(|f_n(A) - p| < \epsilon\right) \to 1 \quad \text{quan } n \to \infty$$

- Hence, in a certain sense, the relative frequency of the event A converges to its probability.

The central limit theorem

Theorem

Let $X_1, X_2, \dots, X_k, \dots$ be a sequence of independent and identically distributed r.v., each with $E(X_k) = m$ and $\mathrm{Var}(X_k) = \sigma^2$.

Let

$$S_n^\star = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \frac{X_k - m}{\sigma}.$$

Then,

$$\lim_{n\to\infty} F_{S_n^\star}(x) = F_{N(0,1)}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt$$

- We have

$$E(S_n^\star) = 0, \quad \mathrm{Var}(S_n^\star) = 1$$

- Notice that $S_n^\star$ is the normalized arithmetic mean $\overline{X}_n$:

$$S_n^\star = \frac{\overline{X}_n - m}{\sigma/\sqrt{n}}$$

$$\overline{X}_n = \frac{\sigma}{\sqrt{n}} S_n^\star + m$$

- We write

$$S_n^\star \xrightarrow{d} N(0,1),$$

Covariance matrices

The covariance matrix K_X of an n-dimensional r.v.

$$X = (X_1, X_2, \dots, X_n)^t$$

is the square $n \times n$ matrix defined by

$$K_X = E\left((X - m_X)(X - m_X)^t\right)$$

$$= \begin{pmatrix} k_{11} & k_{12} & \cdots & k_{1n} \\ k_{21} & k_{22} & \cdots & k_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ k_{n1} & k_{n2} & \cdots & k_{nn} \end{pmatrix}$$

where

- m_X is the expectation vector

$$m_X = E(X) = (m_{X_1}, m_{X_2}, \dots, m_{X_n})^t$$

- For $i \neq j$,

$$k_{ij} = E\left((X_i - m_{X_i})(X_j - m_{X_j})\right) = \mathrm{Cov}(X_i, X_j)$$

- The diagonal entries of K_X are

$$k_{ii} = E\left((X_i - m_{X_i})^2\right) = \mathrm{Var}(X_i)$$

The matrix K_X is

- symmetric,

$$k_{ij} = \mathrm{Cov}(X_i, X_j) = \mathrm{Cov}(X_j, X_i) = k_{ji}$$

- positive-semidefinite.

  That is, for all $z = (z_1, z_2, \dots, z_n)^t \in \mathbb{R}^n$,

$$z^t K_X z \geq 0$$

- Moreover, the r.v.

$$X_1 - m_{X_1}, X_2 - m_{X_2}, \dots, X_n - m_{X_n}$$

  are linearly independent if and only if K_X is positive-definite. That is, if and only if, for all $z \neq 0$,

$$z^t K_X z > 0$$

Indeed, if

$$Y = z_1 X_1 + \cdots z_n X_n = z^t X,$$

we have

$$z^t K_X z = z^t \,\mathrm{E}\left((X - m_X)(X - m_X)^t\right) z$$

$$= \mathrm{E}\left(z^t (X - m_X)(X - m_X)^t z\right)$$

$$= \mathrm{E}\left((Y - m_Y)(Y - m_Y)^t\right)$$

$$= \mathrm{E}\left((Y - m_Y)^2\right) = \sigma_Y^2 \geq 0.$$

(Notice that $z^t m_X = \sum_{i=1}^n z_i m_{X_i} = m_Y$.)

Moreover,

$z^t K_X z = 0$ for some $z \neq 0$

$\iff \sigma_Y^2 = 0$ for some $z \neq 0$

$\iff Y - m_Y = 0,$ with probability 1, for some $z \neq 0$

$\iff \sum_{i=1}^n z_i (X_i - m_{X_i}) = 0,$ with probability 1,
for some $z = (z_1, z_2, \dots, z_n)^t \neq 0$

$\iff X_1 - m_{X_1}, X_2 - m_{X_2}, \dots, X_n - m_{X_n}$ are linealy dependent,
with probability 1.

Linear transformations

Theorem

Let $X = (X_1, X_2, \dots, X_n)^t$ be an n-dimensional r.v., let A be an $m \times n$ real matrix, and let $Y = (Y_1, Y_2, \dots, Y_m)^t$ be the m-dimensional r.v. defined by

$$Y = AX.$$

Then,

$$m_Y = A\, m_X,$$

$$K_Y = A\, K_X\, A^t$$

We have

$$m_Y = \mathrm{E}(Y) = \mathrm{E}(AX) = A\,\mathrm{E}(X) = Am_X,$$

and

$$\begin{aligned} K_Y &= \mathrm{E}\left((Y - m_Y)(Y - m_Y)^t\right) \\ &= \mathrm{E}\left(A(X - m_X)(X - m_X)^t A^t\right) \\ &= A\,\mathrm{E}\left((X - m_X)(X - m_X)^t\right) A^t = AK_XA^t \end{aligned}$$

Multidimensional gaussian density

Let X_1, X_2, …, X_n be n independent gaussian r.v.

We have,

$$f_X(x_1, x_2, \dots, x_n) = f_{X_1}(x_1) f_{X_2}(x_2) \cdots f_{X_n}(x_n)$$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\, \sigma_{X_i}}\, e^{-\frac{1}{2}\left((x_i - m_{X_i})/\sigma_{X_i}\right)^2}$$

$$= \frac{1}{(2\pi)^{n/2} \sigma_{X_1} \sigma_{X_2} \cdots \sigma_{X_n}}\, e^{-\frac{1}{2}\sum_{i=1}^{n}\left((x_i - m_{X_i})/\sigma_{X_i}\right)^2}$$

$$= \frac{1}{(2\pi)^{n/2}\sqrt{\det(K_X)}}\, e^{-\frac{1}{2}(x - m_X)^t\, K_X^{-1}\, (x - m_X)}$$

Hence,

$$f_X(x_1, x_2, \dots, x_n) =$$

$$\frac{1}{(2\pi)^{n/2}\sqrt{\det(K_X)}}\, \exp\left(-\frac{1}{2}(x - m_X)^t\, K_X^{-1}\, (x - m_X)\right)$$

where

- $x = (x_1, x_2, \ldots, x_n)^t$
- m_X is the expectation vector

$$m_X = (m_{X_1}, \ldots, m_{X_n})$$

- $$K_X = \begin{pmatrix} \sigma^2_{X_1} & . & . & . & 0 \\ & \sigma^2_{X_2} & & & \\ & & . & & \\ & & & . & \\ 0 & . & . & . & \sigma^2_{X_n} \end{pmatrix}$$

  is the covariance matrix.
  It is a diagonal matrix because the r.v. are independent and, hence, pairwise uncorrelated.

Now, let us consider a linear transformation

$$Y = AX,$$

being A a non-singular $n \times n$ matrix.

- The linear system $y = Ax$ has a unique solution $x = A^{-1}y$.
- The jacobian determinant is

$$J(x_1, x_2, \ldots, x_n) = \det(A)$$

Moreover,

- $$m_Y = Am_X \Longrightarrow m_X = A^{-1}m_Y$$
- $$K_Y = AK_XA^t \Longrightarrow$$
  $$K_Y^{-1} = (A^t)^{-1}K_X^{-1}A^{-1} = (A^{-1})^tK_X^{-1}A^{-1}$$
- $$\det(K_Y) = \det(K_X)\det(A)^2$$

In this way, we have

$$f_Y(y_1, y_2, \ldots, y_n) = \left.\frac{f_X(x_1, x_2, \ldots, x_n)}{|J(x_1, x_2, \ldots, x_n)|}\right|_{x=A^{-1}y}$$
$$= \frac{1}{(2\pi)^{n/2}\sqrt{\det(K_X)}\,|\det(A)|} \cdot$$
$$\cdot \exp\left(-\frac{1}{2}(A^{-1}y - A^{-1}m_Y)^t\; K_X^{-1}\,(A^{-1}y - A^{-1}m_Y)\right)$$
$$= \frac{1}{(2\pi)^{n/2}\sqrt{\det(K_Y)}} \cdot$$
$$\cdot \exp\left(-\frac{1}{2}\left(A^{-1}(y - m_Y)\right)^t\;K_X^{-1}\,A^{-1}(y - m_Y)\right)$$

$$= \frac{1}{(2\pi)^{n/2}\sqrt{\det(K_Y)}} \cdot$$
$$\cdot \exp\left(-\frac{1}{2}(y-m_Y)^t\,(A^{-1})^t K_X^{-1} A^{-1}\,(y-m_Y)\right)$$
$$= \frac{1}{(2\pi)^{n/2}\sqrt{\det(K_Y)}} \exp\left(-\frac{1}{2}(y-m_Y)^t\, K_Y^{-1}\,(y-m_Y)\right)$$

- Notice that the above expression is analogous to the one we have in the case of independent r.v.
- But now, the covariance matrix K_Y will not be, in general, a diagonal matrix.

Multidimensional gaussian density

Definition

If the random vector X has probability density

$$f_X(x_1, x_2, \ldots, x_n) = \frac{1}{(2\pi)^{n/2}\sqrt{det(K)}} \exp\left(-\frac{1}{2}(x-m)^t\, K^{-1}\,(x-m)\right),$$

where $x = (x_1, x_2, \ldots, x_n)^t$, m is a column $n \times 1$ vector, and K is a square positive-semidefinite $n \times n$ matrix, we say that X is a n-dimensional gaussian r.v.

- We also say that the r.v. X_1, X_2, ..., X_n are jointly gaussian.

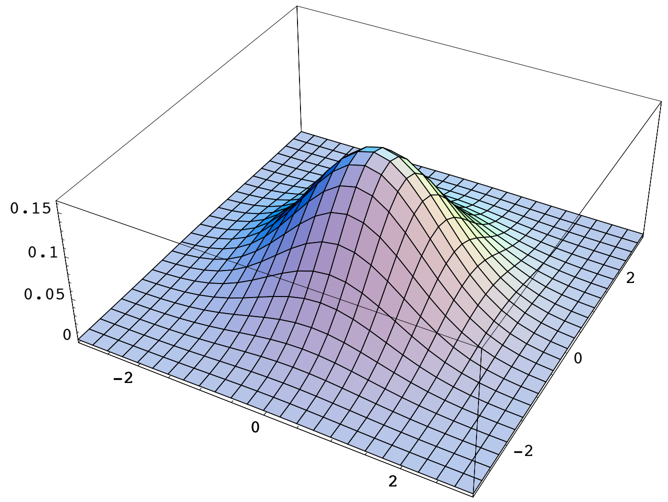
For instance, for $n = 2$ we obtain:

$$f_{XY}(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}\,\sigma_X\sigma_Y} \exp\left(-\frac{1}{2}\cdot\frac{1}{1-\rho^2}\cdot a(x,y)\right),$$

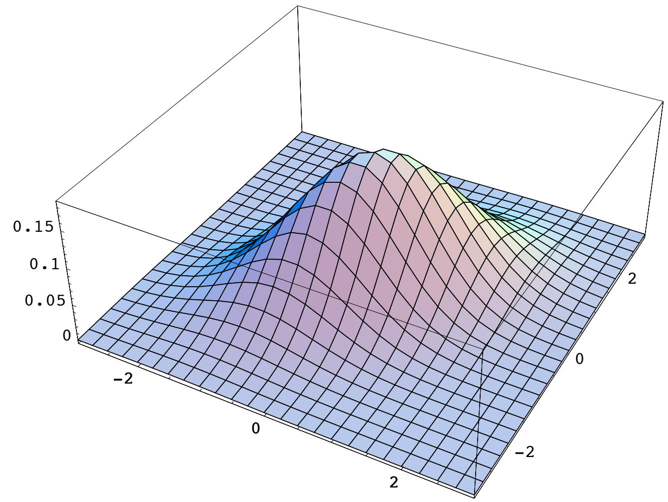
where

$$a(x,y) = \left(\frac{x-m_X}{\sigma_X}\right)^2 - 2\rho\,\frac{x-m_X}{\sigma_X}\cdot\frac{y-m_Y}{\sigma_Y} + \left(\frac{y-m_Y}{\sigma_Y}\right)^2$$

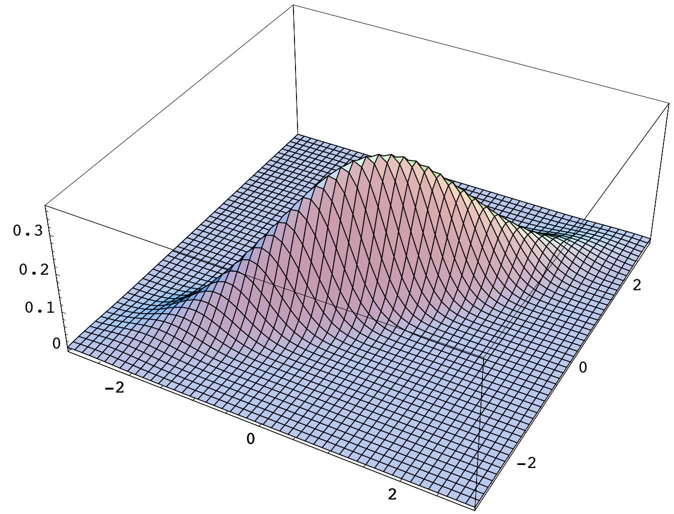
$\sigma_X = \sigma_Y \qquad \rho = 0$



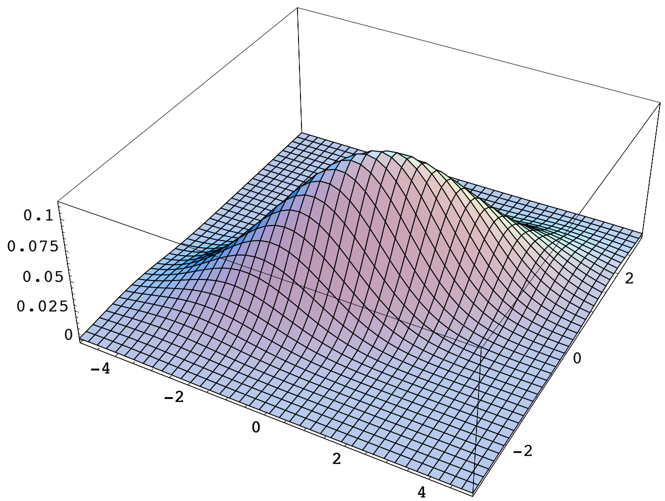
$\sigma_X = \sigma_Y \qquad \rho = 0.5$



$\sigma_X = \sigma_Y \qquad \rho = 0.9$



$\sigma_X = 2\sigma_Y \qquad \rho = 0.7$



Marginal distributions

If $m = (m_i)$ and $K = (k_{ij})$,

- Each component X_i is a (1-dimensional) gaussian r.v. with parameters

$$m_{X_i} = m_i,$$
$$\sigma_i^2 = k_{ii}$$

- K is the covariance matrix of X.

Eigenvalues of the covariance matrix

The matrix K_X is symmetric. Hence, it can be transformed into a diagonal matrix by means of an ortogonal transformation.

That is,

- There exists an ortogonal matrix C such that

$$CK_X C^t = D = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$$

- The numbers $\lambda_i \in \mathbb{R}$ are the eigenvalues of K_X.

Equivalently,

$$K_X = C^t DC$$

Hence,

$$0 \leq z^t K_X z = z^t C^t DCz = (Cz)^t \; D \; (Cz)$$
$$= y^t Dy = \sum_{i=1}^{n} y_i^2 \lambda_i,$$

where $y = Cz$.

- Therefore,

$$\lambda_i \geq 0, \quad i = 1, 2, \ldots, n$$

- K_X is a positive-definite matrix if and only if

$$\lambda_i > 0, \quad i = 1, 2, \ldots, n$$

($X_i - m_{X_i}$, $i = 1, \ldots, n$, are linearly independent r.v.)

- This condition is equivalent to

$$\det(K_X) \neq 0$$

- Gaussian random vectors can be defined in a more general framework, in such a way that K_X is not necessarily invertible.

  That is, $X_i - m_{X_i}$, $i = 1, \ldots, n$, could be not all linearly independent.

  But only in the case that $\det(K_X) \neq 0$ there exists a density

$$f_X(x_1, x_2, \ldots, x_n)$$

Uncorrelation and independence

Theorem

If the r.v. X_1, X_2, …, X_n are jointly gaussian and pairwise uncorrelated, then they are jointly independent.

Indeed,

$$\mathrm{Cov}(X_i, X_j) = 0 \Longrightarrow K_X = \mathrm{diag}(\sigma^2_{X_1}, \sigma^2_{X_2}, \ldots, \sigma^2_{X_n})$$

and, hence,

$$f_X(x_1, x_2, \ldots, x_n) =$$

$$\frac{1}{(2\pi)^{n/2}\sqrt{\det(K_X)}} \exp\left(-\frac{1}{2}(x - m_X)^t\, K_X^{-1}\,(x - m_X)\right)$$

$$= \frac{1}{(2\pi)^{n/2}\sigma_{X_1}\sigma_{X_2}\cdots\sigma_{X_n}}\, e^{-\frac{1}{2}\sum_{i=1}^n\left((x_i - m_{X_i})/\sigma_{X_i}\right)^2}$$

$$= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\,\sigma_{X_i}}\, e^{-\frac{1}{2}\left((x_i - m_{X_i})/\sigma_{X_i}\right)^2} = f_{X_1}(x_1) f_{X_2}(x_2)\cdots f_{X_n}(x_n)$$

Linear combinations

Theorem

Let X be an n-dimensional gaussian r.v., let A be an $m \times n$ real matrix, and let

$$Y = AX.$$

Then, Y is an m-dimensional gaussian r.v. with $m_Y = A\, m_X$ and $K_Y = AK_XA^t$.

- If $m \leq n$ and A has full rang m, the random vector Y has a probaility density $f_Y(y_1, \ldots, y_n)$.

Theorem

The n-dimensional r.v. $X = (X_1, \ldots, X_n)^t$ is gaussian if and only if the 1-dimensional r.v.

$$Y = a_1X_1 + \cdots + a_nX_n = a^tX$$

is gaussian for all $a = (a_1, a_2, \ldots a_n)^t \in \mathbb{R}^n$

Conditional densities

Let X, Y be jointly gaussian. Then,

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

$$= \frac{1}{\sqrt{2\pi}\sqrt{1 - \rho^2}\,\sigma_Y} \exp\left(-\frac{1}{2}\left(\frac{y - m_{Y|X}}{\sigma_{Y|X}}\right)^2\right)$$

- $m_{Y|X}$ is the expected value of Y given X:

$$m_{Y|X} = \mathrm{E}(Y|X = x) = \rho\,\frac{\sigma_Y}{\sigma_X}(x - m_X) + m_Y$$

- $\sigma^2_{Y|X} = (1 - \rho^2)\,\sigma^2_Y$.