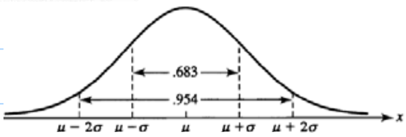
Multivariate Normal Distribution

• Density of Normal Distribution

➢ univariate case: normal density with mean μ and variance σ^2

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-[(x-\mu)/\sigma]^2/2} \qquad -\infty < x < \infty$$



➢ multivariate case: a p-dimensional normal density for the random vector $\mathbf{X}' = [X_1, X_2, \ldots, X_p]$ has the form

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

where $-\infty < x_i < \infty$, $i = 1, 2, \ldots, p$, $\boldsymbol{\mu}$ is the mean of $\mathbf{X}$

$\mathbf{\Sigma}$ is the variance–covariance matrix of $\mathbf{X}$. (Note. $\mathbf{\Sigma}^{-1}$ must exist, i.e., $|\Sigma| \neq 0$)

We shall denote this p-dimensional normal density by $N_p(\boldsymbol{\mu}, \mathbf{\Sigma})$

- the term $\left(\frac{x-\mu}{\sigma}\right)^2 = (x-\mu)(\sigma^2)^{-1}(x-\mu)$ | $(\mathbf{x}-\boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})$

  is a distance measure

- Contours of constant density for the p-dimensional normal distribution are ellipsoids defined by $\mathbf{x}$ such the that

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

  These ellipsoids are centered at $\boldsymbol{\mu}$ and have axes $\pm c\sqrt{\lambda_i}\,\mathbf{e}_i$, where $\mathbf{\Sigma}\mathbf{e}_i = \lambda_i\mathbf{e}_i$ for $i = 1, 2, \ldots, p$. $(\lambda, \mathbf{e})$ is an eigenvalue–eigenvector pair for $\mathbf{\Sigma}$

➢ Example: bivariate normal density

$$\mu_1 = E(X_1), \quad \mu_2 = E(X_2),$$

$\sigma_{11} = \text{Var}(X_1)$, $\sigma_{22} = \text{Var}(X_2)$, and $\rho_{12} = \sigma_{12}/(\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}) = \text{Corr}(X_1, X_2)$.

- $$\mathbf{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} \qquad \mathbf{\Sigma}^{-1} = \frac{1}{\sigma_{11}\sigma_{22} - \sigma_{12}^2}\begin{bmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{12} & \sigma_{11} \end{bmatrix}$$

- by writing $\sigma_{12} = \rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}$, we obtain $\sigma_{11}\sigma_{22} - \sigma_{12}^2 = \sigma_{11}\sigma_{22}(1-\rho_{12}^2)$

- $(\mathbf{x}-\boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})$

$$= [x_1 - \mu_1, x_2 - \mu_2]\frac{1}{\sigma_{11}\sigma_{22}(1-\rho_{12}^2)} \begin{bmatrix} \sigma_{22} & -\rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}} \\ -\rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}} & \sigma_{11} \end{bmatrix}\begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix}$$

$$= \frac{\sigma_{22}(x_1-\mu_1)^2 + \sigma_{11}(x_2-\mu_2)^2 - 2\rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}(x_1-\mu_1)(x_2-\mu_2)}{\sigma_{11}\sigma_{22}(1-\rho_{12}^2)}$$

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

- written in terms of the standardized values $(x_1 - \mu_1)/\sqrt{\sigma_{11}}$ and $(x_2 - \mu_2)/\sqrt{\sigma_{22}}$.
- if X_1 and X_2 are uncorrelated, they are independent

- contour for the case $\sigma_{11} = \sigma_{22},\ \sigma_{12} \neq 0$

- $0 = |\mathbf{\Sigma} - \lambda\mathbf{I}| = \begin{vmatrix} \sigma_{11} - \lambda & \sigma_{12} \\ \sigma_{12} & \sigma_{11} - \lambda \end{vmatrix} = (\sigma_{11} - \lambda)^2 - \sigma_{12}^2$
  $= (\lambda - \sigma_{11} - \sigma_{12})(\lambda - \sigma_{11} + \sigma_{12})$
- the eigenvalues are $\lambda_1 = \sigma_{11} + \sigma_{12}$ and $\lambda_2 = \sigma_{11} - \sigma_{12}$
- $\begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{11} \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} = (\sigma_{11} + \sigma_{12}) \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} \Rightarrow \begin{array}{l} \sigma_{11}e_1 + \sigma_{12}e_2 = (\sigma_{11} + \sigma_{12})e_1 \\ \sigma_{12}e_1 + \sigma_{11}e_2 = (\sigma_{11} + \sigma_{12})e_2 \end{array}$

$$\lambda_1 = \sigma_{11} + \sigma_{12}, \qquad \mathbf{e}_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

- $\lambda_2 = \sigma_{11} - \sigma_{12}$ yields the eigenvector $\mathbf{e}_2' = [1/\sqrt{2}, -1/\sqrt{2}]$.
- When the covariance σ_{12} (or correlation ρ_{12}) is positive,

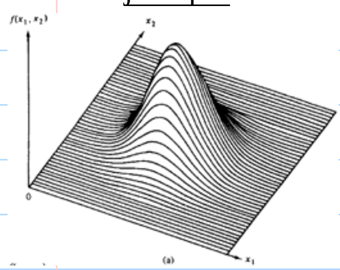
  When the covariance (correlation) is negative,

What if $|\rho_{12}|$ increase?

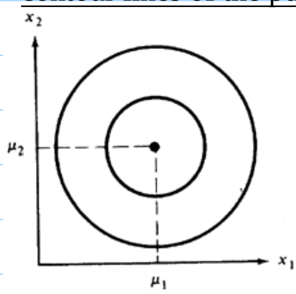
- when $\rho_{12} = 0$

(a) $\sigma_1=\sigma_2,\ \rho=0 \Rightarrow$ independent and equal variance

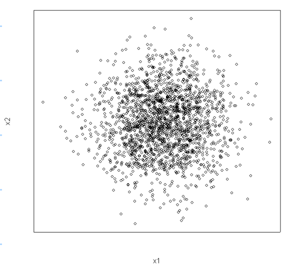
joint pdf



contour lines of the pdf

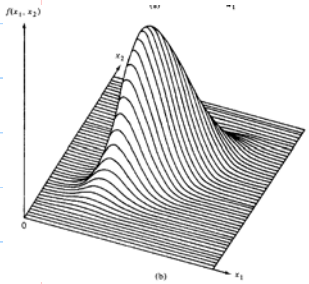


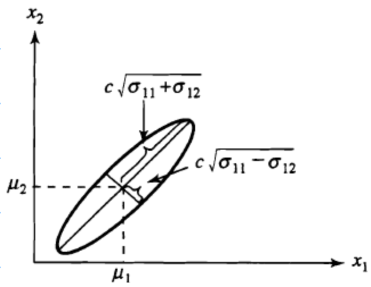
data generated from the pdf

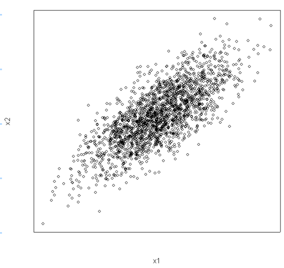


(b) $\sigma_1=\sigma_2,\ \rho=0.75 \Rightarrow$ correlated and equal variance

Q: what should the contour lines look like if $\sigma_1 \neq \sigma_2$?







when $\sigma_1=\sigma_2,\ \rho\neq 0$, the major/minor axis of the ellipse is parallel to $x_1=x_2$ or $x_1=-x_2$

contour of Normal pdf is an ellipse because it can be expressed as $(\boldsymbol{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})=c$

➢ random sample from a multivariate normal distribution

Let us assume that the $p \times 1$ vectors $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ represent a random sample from a multivariate normal population with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Since $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ are mutually independent and each has distribution $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the joint density function of all the observations is the product of the marginal normal densities:

$$\left\{\begin{array}{c}\text{Joint density}\\ \text{of } \mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n\end{array}\right\} = \prod_{j=1}^{n} \left\{ \frac{1}{(2\pi)^{p/2} |\boldsymbol{\Sigma}|^{1/2}} e^{-(\mathbf{x}_j - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x}_j - \boldsymbol{\mu})/2} \right\}$$

$$= \frac{1}{(2\pi)^{np/2}} \frac{1}{|\boldsymbol{\Sigma}|^{n/2}} e^{-\sum_{j=1}^{n} (\mathbf{x}_j - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x}_j - \boldsymbol{\mu})/2}$$

➢ **Q**: why normal?

- While real data are never exactly multivariate normal, the normal density is often a useful approximation to the "true" population distribution
- The multivariate normal density is mathematically tractable and "nice" results can be obtained
- The distribution of many multivariate statistics are approximately normal, regardless of the form of the parent population because of a *central limit* effect

• Some properties of multivariate normal distribution

➢ **Result 4.3.** If $\mathbf{X}$ is distributed as $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the q linear combinations

$$\underset{(q\times p)}{\mathbf{A}}\underset{(p\times 1)}{\mathbf{X}} = \begin{bmatrix} a_{11}X_1 + \cdots + a_{1p}X_p \\ a_{21}X_1 + \cdots + a_{2p}X_p \\ \vdots \\ a_{q1}X_1 + \cdots + a_{qp}X_p \end{bmatrix}$$

are distributed as $N_q(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$. Also, $\underset{(p\times 1)}{\mathbf{X}} + \underset{(p\times 1)}{\mathbf{d}}$, where $\mathbf{d}$ is a vector of constants, is distributed as $N_p(\boldsymbol{\mu} + \mathbf{d}, \boldsymbol{\Sigma})$.

Alternative definition of multivariate normal distribution: Let $z_1, \ldots, z_p$ be i.i.d from $N(0,1)$ and $Z = [z_1, \ldots, z_p]^T$. For a p-dim vector μ and a $p \times p$ symmetric, positive definite matrix Σ, X is said to have a multivariate normal distribution $N_p(\mu, \Sigma)$ if it has the same distribution as $$\Sigma^{1/2} Z + \mu$$

- $\mathbf{Z} = \boldsymbol{\Sigma}^{-\frac{1}{2}}(\mathbf{X} - \boldsymbol{\mu})$ has a $N_p(\mathbf{0}, \mathbf{I}_p)$ distribution

$$(\mathbf{X} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu}) = (\mathbf{X} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{\Sigma}^{-\frac{1}{2}} (\mathbf{X} - \boldsymbol{\mu})$$

$$= \mathbf{Z}'\mathbf{Z} = Z_1^2 + Z_2^2 + \cdots + Z_p^2$$

➢ **Result 4.4.** All subsets of $\mathbf{X}$ are normally distributed. If we respectively partition $\mathbf{X}$, its mean vector $\boldsymbol{\mu}$, and its covariance matrix $\boldsymbol{\Sigma}$ as

$$\underset{(p\times 1)}{\mathbf{X}} = \begin{bmatrix} \underset{(q\times 1)}{\mathbf{X}_1} \\ \hdashline \underset{((p-q)\times 1)}{\mathbf{X}_2} \end{bmatrix} \qquad \underset{(p\times 1)}{\boldsymbol{\mu}} = \begin{bmatrix} \underset{(q\times 1)}{\boldsymbol{\mu}_1} \\ \hdashline \underset{((p-q)\times 1)}{\boldsymbol{\mu}_2} \end{bmatrix}$$

and

$$\underset{(p\times p)}{\boldsymbol{\Sigma}} = \left[\begin{array}{c:c} \underset{(q\times q)}{\boldsymbol{\Sigma}_{11}} & \underset{(q\times(p-q))}{\boldsymbol{\Sigma}_{12}} \\ \hdashline \underset{((p-q)\times q)}{\boldsymbol{\Sigma}_{21}} & \underset{((p-q)\times(p-q))}{\boldsymbol{\Sigma}_{22}} \end{array}\right]$$

then $\mathbf{X}_1$ is distributed as $N_q(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$.

➢ **Result 4.5.**

(a) If $\underset{(q_1\times 1)}{\mathbf{X}_1}$ and $\underset{(q_2\times 1)}{\mathbf{X}_2}$ are independent, then $\text{Cov}(\mathbf{X}_1, \mathbf{X}_2) = \mathbf{0}$, a $q_1 \times q_2$ matrix of zeros.

(b) If $\begin{bmatrix} \mathbf{X}_1 \\ \hdashline \mathbf{X}_2 \end{bmatrix}$ is $N_{q_1+q_2}\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \hdashline \boldsymbol{\mu}_2 \end{bmatrix}, \left[\begin{array}{c:c} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \hdashline \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{array}\right]\right)$, then $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent if and only if $\boldsymbol{\Sigma}_{12} = \mathbf{0}$.

(c) If $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent and are distributed as $N_{q_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$ and $N_{q_2}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$, respectively, then $\begin{bmatrix} \mathbf{X}_1 \\ \hdashline \mathbf{X}_2 \end{bmatrix}$ has the multivariate normal distribution

$$N_{q_1+q_2}\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \hdashline \boldsymbol{\mu}_2 \end{bmatrix}, \left[\begin{array}{c:c} \boldsymbol{\Sigma}_{11} & \mathbf{0} \\ \hdashline \mathbf{0}' & \boldsymbol{\Sigma}_{22} \end{array}\right]\right)$$

Note. Suppose that X_1 and X_2 are multivariate normal. $X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ may not be a multivariate normal.

➢ **Result 4.6.** Let $\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \hdashline \mathbf{X}_2 \end{bmatrix}$ be distributed as $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \hdashline \boldsymbol{\mu}_2 \end{bmatrix}$, $\boldsymbol{\Sigma} = \left[\begin{array}{c:c} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \hdashline \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{array}\right]$, and $|\boldsymbol{\Sigma}_{22}| > 0$. Then the conditional distribution of $\mathbf{X}_1$, given that $\mathbf{X}_2 = \mathbf{x}_2$, is normal and has

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

and

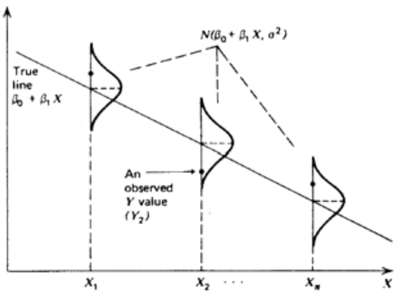
$$\text{Covariance} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$$

Note that the covariance does not depend on the value $\mathbf{x}_2$ of the conditioning variable.

- Example: conditional density of a bivariate normal distribution
  - The conditional density of X_1, given that $X_2 = x_2$, is

$$N\left(\mu_1 + \frac{\sigma_{12}}{\sigma_{22}}(x_2 - \mu_2), \sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}}\right)$$

Here $\sigma_{11} - \sigma_{12}^2/\sigma_{22} = \sigma_{11}(1 - \rho_{12}^2)$.



  - conditional density and regression model

➢ **Result 4.7.** Let $\mathbf{X}$ be distributed as $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $|\boldsymbol{\Sigma}| > 0$. Then

(a) $(\mathbf{X} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\mu})$ is distributed as χ_p^2, where χ_p^2 denotes the chi-square distribution with p degrees of freedom.

(b) The $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution assigns probability $1 - \alpha$ to the solid ellipsoid $\{\mathbf{x}: (\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) \le \chi_p^2(\alpha)\}$, where $\chi_p^2(\alpha)$ denotes the upper (100α)th percentile of the χ_p^2 distribution.