COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

STAT 4224/5224

Bayesian Statistics

Dobrin Marchev

Introduction

- Up until now all of our models have been *univariate* models, that is, models for a single measurement on each member of a sample of individuals or each run of a repeated experiment.
- However, datasets are frequently *multivariate*, having multiple measurements for each individual or experiment.
- We now cover what is perhaps the most useful model for multivariate data, the *multivariate normal model*, which allows us to jointly estimate population means, variances and correlations of a collection of variables.
- The model can also be used to impute missing data.

Univariate Normal (Gaussian) Distribution

- Bell-shaped distribution with tendency for individuals to clump around the group median/mean
- Used to model many biological phenomena
- Many *estimators* have approximately normal sampling distributions (Central Limit Theorem)
- Notation: $X \sim \mathrm{N}(\mu, \sigma^2)$ where μ is mean and σ^2 is the variance

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}, -\infty < x < \infty, -\infty < \mu < \infty, \sigma > 0$$

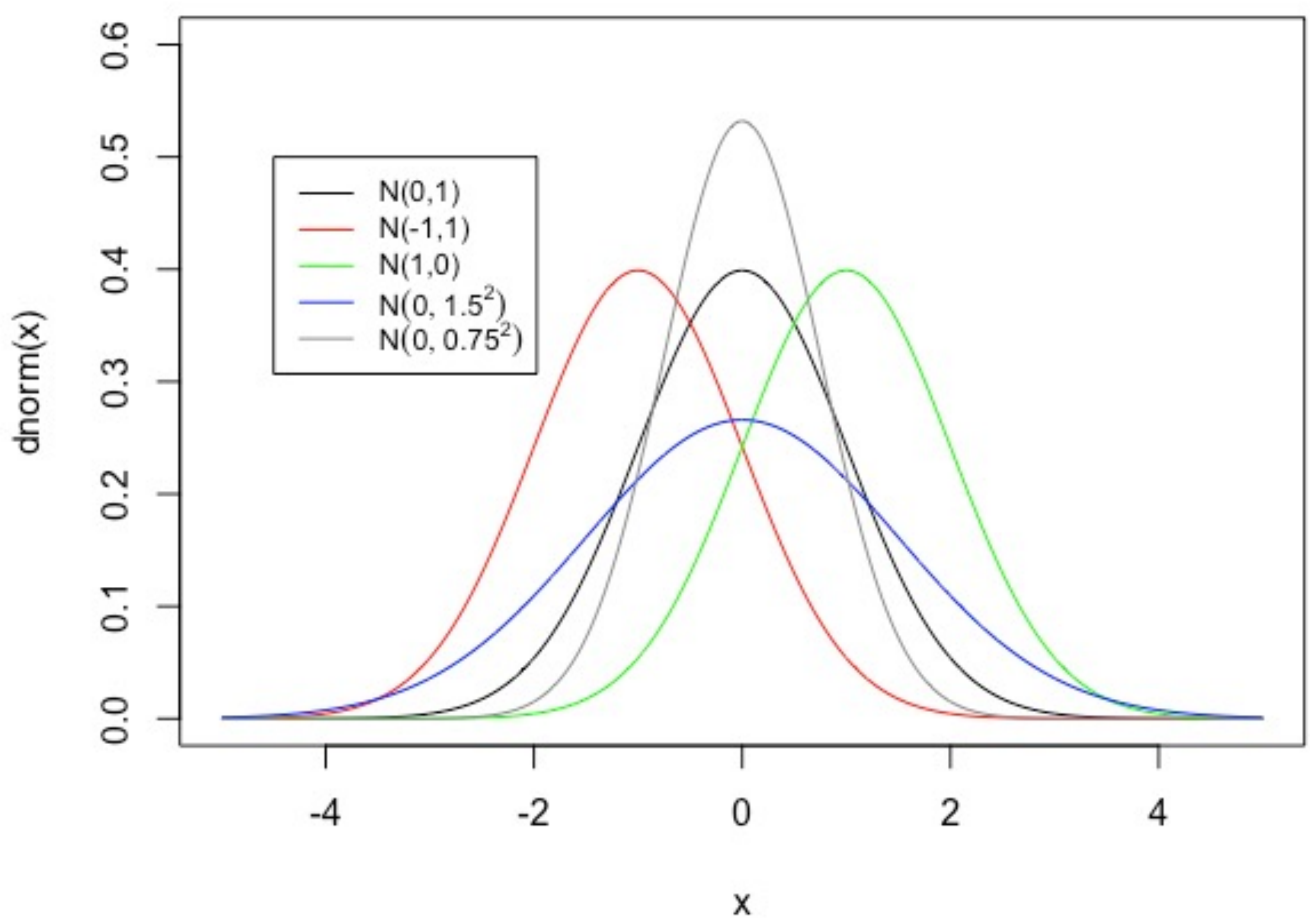
Obtaining Probabilities and Quantiles in `R`:

To obtain: $F(x) = P(X \leq x)$ → Use Function: `pnorm`(x, μ, σ)

To obtain the p^{th} quantile: $P(X \leq x_p) = p$ → Use Function: `qnorm`(p, μ, σ)

Virtually all statistics textbooks give the cdf for standardized normal random variables:
$z = (x - \mu)/\sigma \sim \mathrm{N}(0,1)$

Normal Distribution – Density Functions (pdf)



Chi-Square Distribution

- Indexed by "degrees of freedom (ν)" $X \sim \chi_\nu^2$
- $Z \sim N(0,1) \Rightarrow Z^2 \sim \chi_1^2$
- Assuming Independence:

$$X_1,\dots,X_n \sim \chi^2_{\nu_i} \quad i=1,\dots,n \quad \Rightarrow \quad \sum_{i=1}^{n} X_i \sim \chi^2_{\sum \nu_i}$$

Density Function:

$$f(x) = \frac{1}{\Gamma\left(\frac{\nu}{2}\right)2^{\nu/2}} x^{(\nu/2)-1} e^{-x/2} \quad x>0, \nu>0 \qquad E\{X\}=\nu \quad V\{X\}=2\nu$$

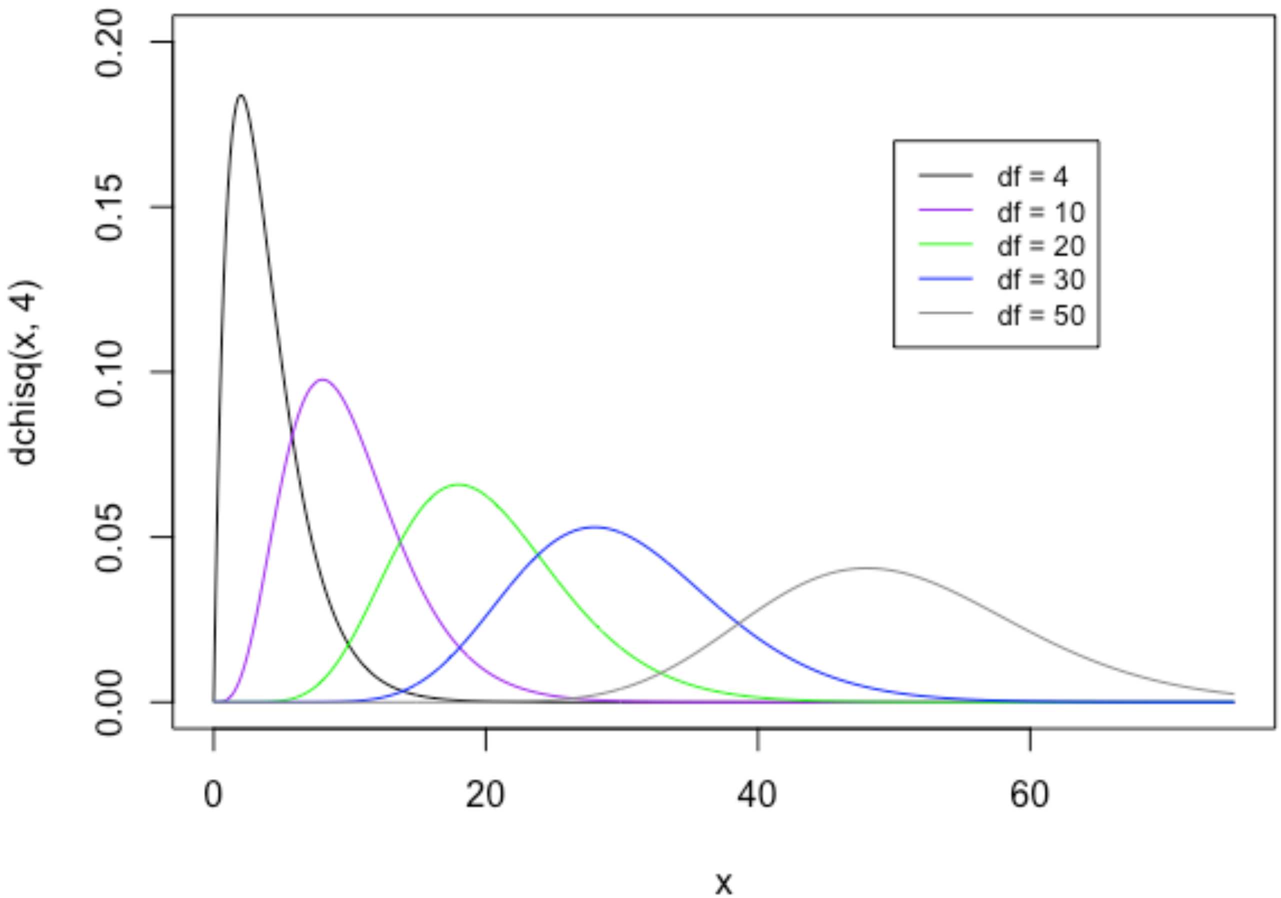
Obtaining Probabilities in R:

To obtain: 1-F(x) = P(X≥x) Use Function: pchisq(x, ν)

To obtain quantiles: P(X≤x_p) = p Use Function: qchisq(x, ν)

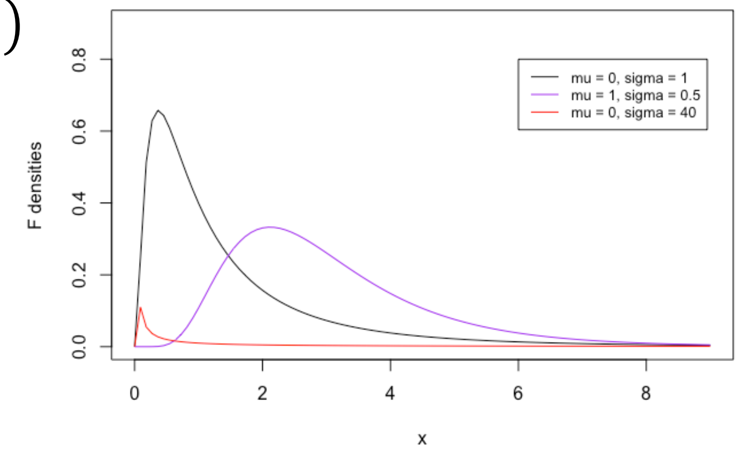
Many statistics textbooks give upper tail cut-off values for commonly used upper (and sometimes lower) tail probabilities

Chi-Square Distributions



Log-normal Distribution

- If $Z \sim N(0,1), \mu \in \mathbb{R}, \sigma > 0$, then
- $X = e^{\mu + \sigma Z}$ is called log-normal distribution
- The pdf is $f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{\sigma\sqrt{2}}}, x > 0$
- One of the most common applications where log-normal distributions are used in finance is in the analysis of stock prices and income distributions.
- The mean is $e^{\mu + \frac{1}{2}\sigma^2}$
- The second moment is $\mu_2 = e^{2(\mu + \sigma^2)}$
- The median is e^{μ}



Obtaining Probabilities/Quantiles in R:

Density *f*(*x*):	dlnorm(x, μ, σ)
To obtain: *F*(*x*) = P(*X* ≤ *x*):	plnorm(x, μ, σ)
p^{th} quantile:	qlnorm(x, μ, σ)

Bivariate Normal Distribution

Let the joint pdf of two variables be:

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\, e^{-½\, Q(x_1,x_2)}$$

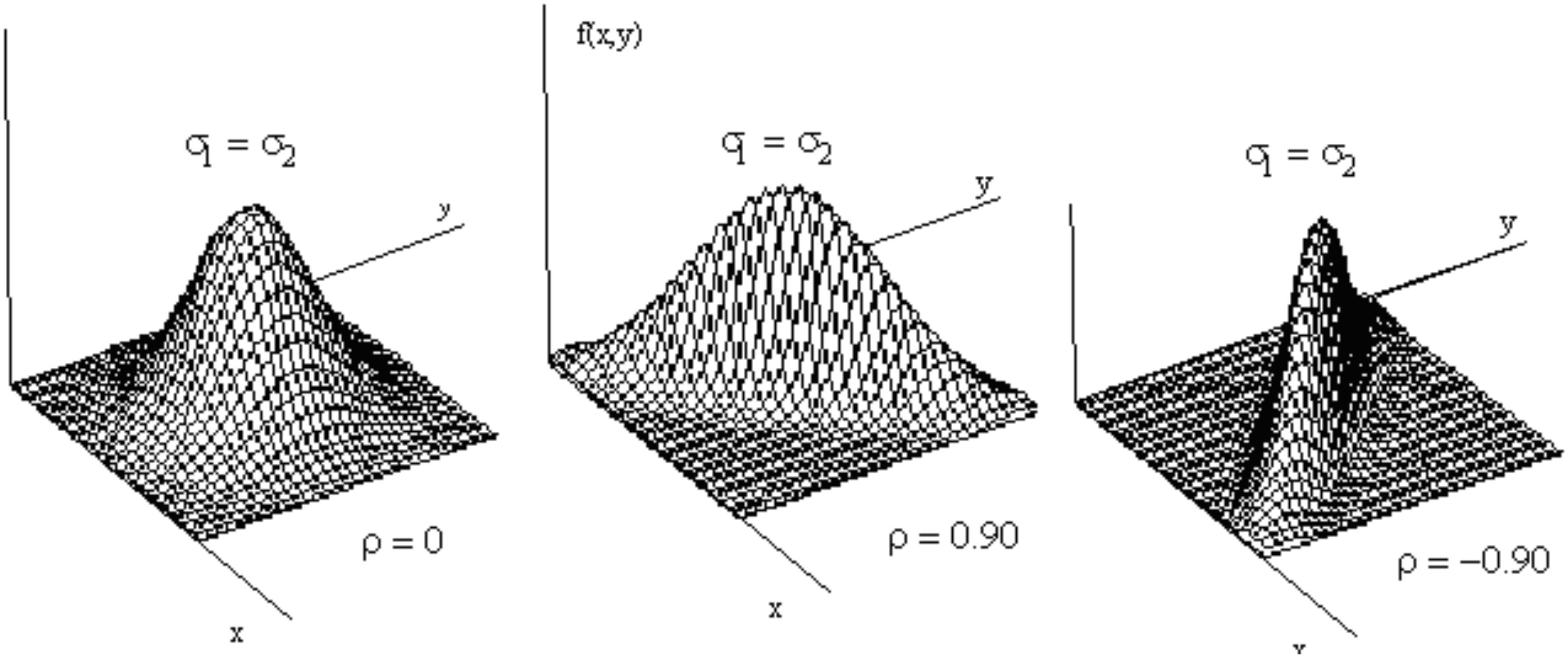
where

$$Q(x_1, x_2) = \frac{\left\{\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right\}}{1-\rho^2}$$

This distribution is called the **bivariate Normal distribution.**

The parameters are μ_1, μ_2 , σ_1, σ_2 and ρ.

Surface Plots of the bivariate Normal distribution



Note:

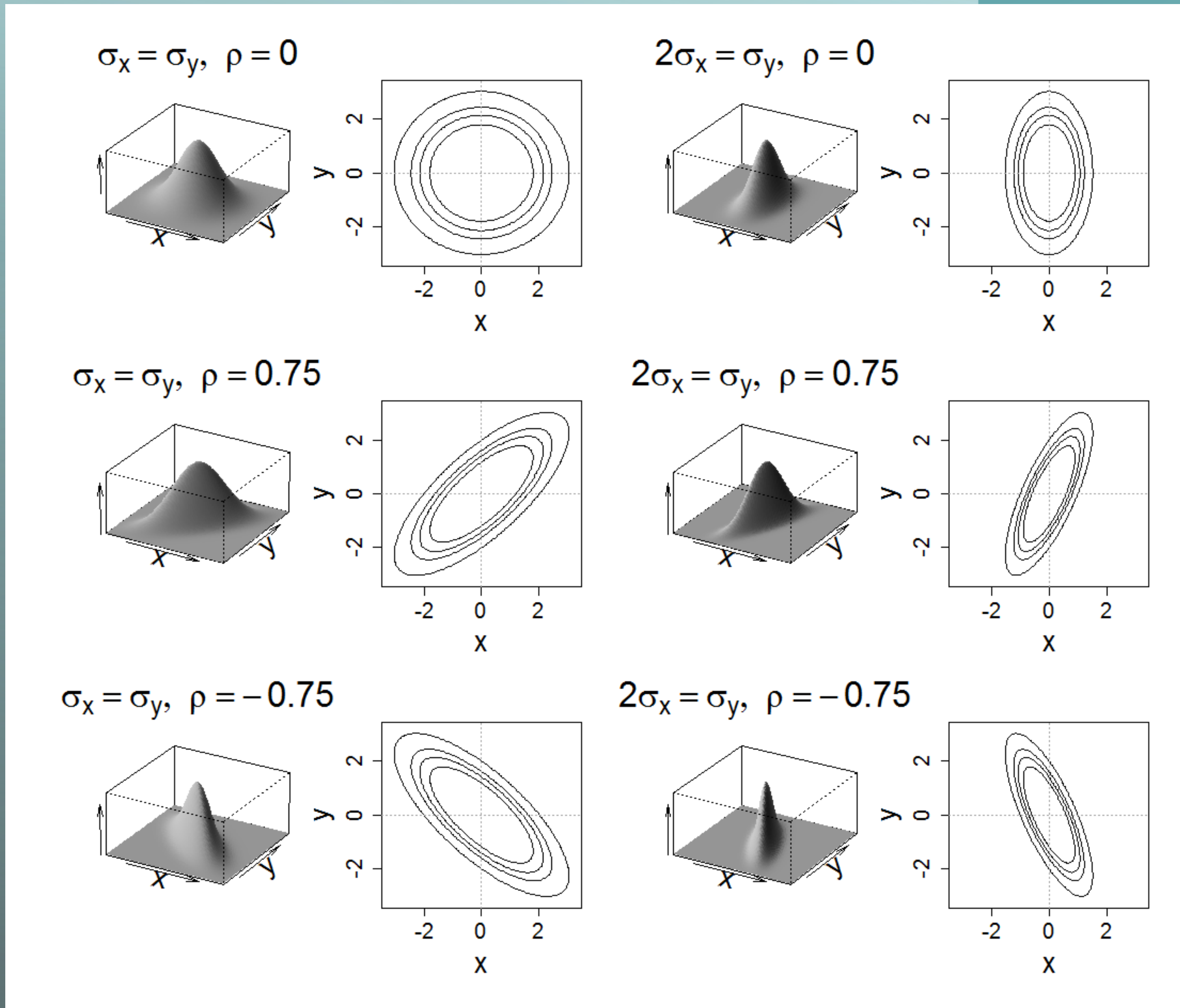
$$f(x_1, x_2) = \frac{1}{(2\pi)\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2}Q(x_1, x_2)}$$

is constant when

$$Q(x_1, x_2) = \frac{\left\{\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right\}}{1-\rho^2}$$

is constant.

This is true when x_1, x_2 lie on an ellipse centered at μ_1, μ_2 .



$\sigma_x = \sigma_y,\ \rho = 0$
$2\sigma_x = \sigma_y,\ \rho = 0$
$\sigma_x = \sigma_y,\ \rho = 0.75$
$2\sigma_x = \sigma_y,\ \rho = 0.75$
$\sigma_x = \sigma_y,\ \rho = -0.75$
$2\sigma_x = \sigma_y,\ \rho = -0.75$
x
y
-2
0
2

Marginal Distributions

Recall the definition of marginal distributions for continuous random variables:

$$f_1(x_1) = \int_{-\infty}^{\infty} f(x_1, x_2)\, dx_2 \quad \text{and} \quad f_2(x_2) = \int_{-\infty}^{\infty} f(x_1, x_2)\, dx_1$$

It can be shown that in the case of the bivariate normal distribution the marginal distribution of x_i is Normal with mean μ_i and standard deviation σ_i, $i = 1, 2$

Thus the marginal distribution of x_2 is Normal with mean μ_2 and standard deviation σ_2.

Similarly, the marginal distribution of x_1 is Normal with mean μ_1 and standard deviation σ_1.

Conditional Distributions

Theorem: Show that in the case of the bivariate normal distribution the conditional distribution of x_i given x_j is normal with:

- Mean = $\mu_{i|j} = \mu_i + \rho \frac{\sigma_i}{\sigma_j}(x_j - \mu_j)$
- Standard deviation = $\sigma_{i|j} = \sigma_i \sqrt{1 - \rho^2}$

Proof

$$f_{2|1}\left(x_2 \,|\, x_1\right) = \frac{f\left(x_1, x_2\right)}{f_1\left(x_1\right)}$$

$$= \frac{\mathrm{e}^{-\frac{1}{2}Q(x_1,x_2)}}{(2\pi)\sigma_1\sigma_2\sqrt{1-\rho^2}} \Bigg/ \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{1}{2}\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2}$$

$$= \frac{\mathrm{e}^{-\frac{1}{2}Q(x_1,x_2)-\frac{1}{2}\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2}}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}} = \frac{\mathrm{e}^{-\frac{1}{2}\left[\left(\frac{x_1-a}{b}\right)^2+c\right]-\frac{1}{2}\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2}}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}}$$

where $$b = \sigma_1 \sqrt{1 - \rho^2}$$

$$a = \mu_1 + \rho \frac{\sigma_1}{\sigma_2}\left(x_2 - \mu\right)$$

and $$c = \left(\frac{x_2 - \mu_2}{\sigma_2}\right)^2$$

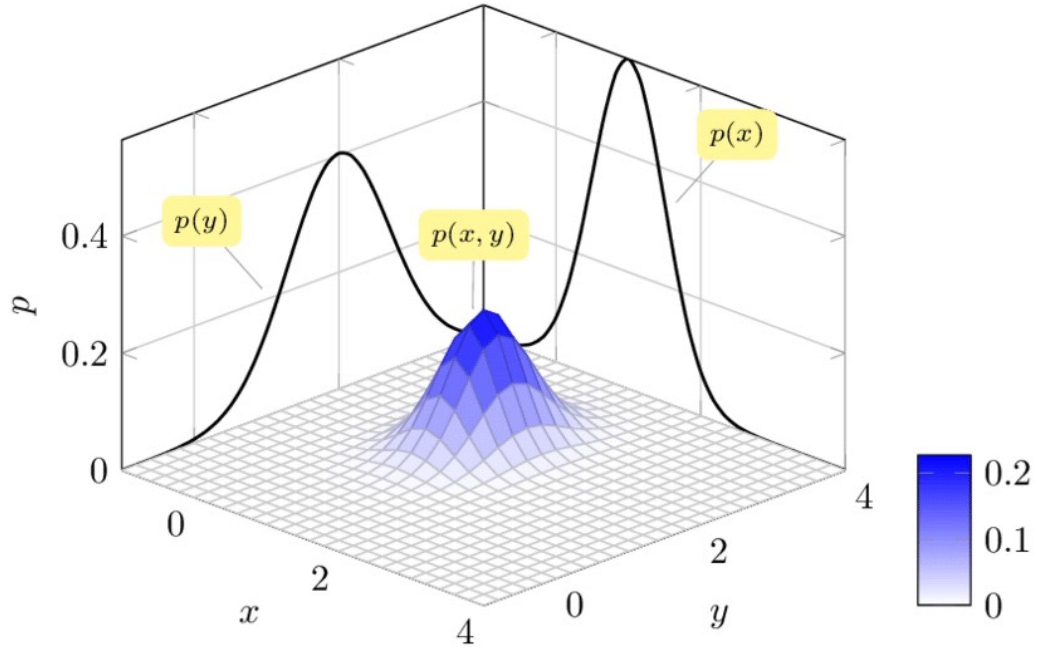
Hence $$f_{1|2}\left(x_1 \middle| x_2\right) = \frac{1}{\sqrt{2\pi b}}\; e^{-\frac{1}{2}\left(\frac{x_1 - a}{b}\right)^2}$$

Thus the conditional distribution of x_2 given x_1 is Normal with:

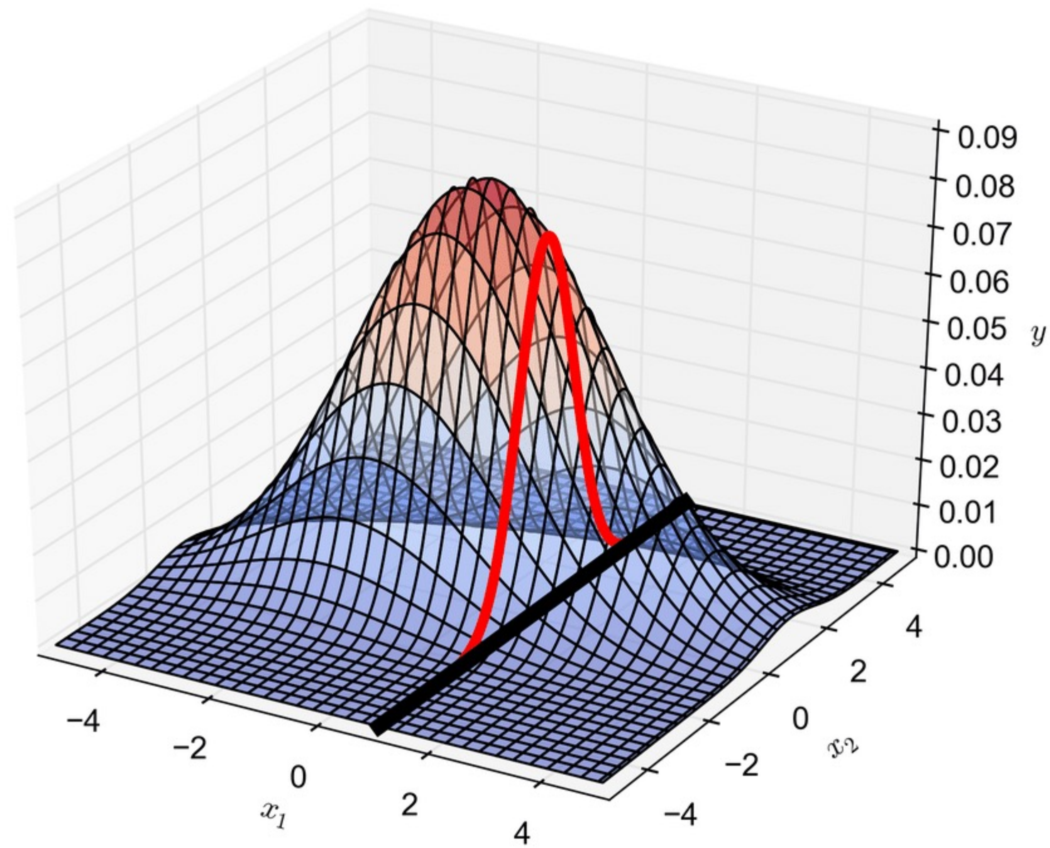
mean $a = \mu_{1|2} = \mu_1 + \rho \frac{\sigma_1}{\sigma_2}\left(x_2 - \mu_2\right)$ and

standard deviation $b = \sigma_{1|2} = \sigma_1 \sqrt{1 - \rho^2}$

Bivariate normal distribution with marginal distributions



Bivariate normal distribution with conditional distribution



Multivariate Normal Distribution

Notation: $\boldsymbol{X} \sim \mathrm{N}_p(\boldsymbol{\mu_X}, \boldsymbol{\Sigma_X})$, where:

$$\boldsymbol{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}, \boldsymbol{\mu_X} = E(\boldsymbol{X}) = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix}, \boldsymbol{\Sigma_X} = \sigma^2(\boldsymbol{X}) = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1p} & \sigma_{2p} & \cdots & \sigma_{pp} \end{bmatrix}$$

Multivariate normal density function:

$$f(\boldsymbol{x}) = (2\pi)^{-p/2} \left| \boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1/2} \right| e^{-\frac{1}{2}(\boldsymbol{X}-\boldsymbol{\mu_X})' \boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1} (\boldsymbol{X}-\boldsymbol{\mu_X})}$$

Results:

$$X_i \sim N(\mu_i, \sigma_{ii}), \qquad i = 1, \ldots, p$$
$$\mathrm{cov}(X_i, X_j) = \sigma_{ij}, \qquad i \neq j$$

Multivariate Normal – Conditional Distributions

Let $\boldsymbol{X}_1$ be $q \times 1$ and $\boldsymbol{X}_2$ be $(p$ -q$) \times 1$ and

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{bmatrix} \sim \mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \qquad \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \qquad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$$

Then $\boldsymbol{X}_1 \mid \boldsymbol{X}_2 = \boldsymbol{x}_2 \sim \mathrm{N}_q$ with
mean $\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2)$ and covariance matrix $\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$

Note: the conditional mean depends on the specific value $\boldsymbol{x}_2$ but the covariance matrix does not.

Special case: $p = 2$, $q = 1$.
Then $X_1 \mid X_2 = x_2 \sim \mathrm{N}(\mu_{1|x2}, \sigma_{1|x2})$
where

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

This result is the basis of linear regression models!

Example with $p = 2$

Joint Distribution:

$$f(x_1, x_2) = \frac{1}{2\pi\sqrt{\sigma_1^2\sigma_2^2\left(1-\rho^2\right)}} \exp\left\{-\left(\frac{1}{2\left(1-\rho^2\right)}\right)\left[\frac{\left(x_1-\mu_1\right)^2}{\sigma_1^2} - \frac{2\rho\left(x_1-\mu_1\right)\left(x_2-\mu_2\right)}{\sigma_1\sigma_2} + \frac{\left(x_2-\mu_2\right)^2}{\sigma_2^2}\right]\right\} \quad -\infty < x_1, x_2 < \infty$$

Marginal (aka Unconditional) Distributions:

$$f_1\left(x_1\right) = \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left\{-\frac{\left(x_1-\mu_1\right)^2}{2\sigma_1^2}\right\} \quad -\infty < x_1 < \infty$$

$$f_2\left(x_2\right) = \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left\{-\frac{\left(x_2-\mu_2\right)^2}{2\sigma_2^2}\right\} \quad -\infty < x_2 < \infty$$

$$X_1 \sim N\left(\mu_1, \sigma_1^2\right) \qquad X_2 \sim N\left(\mu_2, \sigma_2^2\right)$$

Conditional Distributions:

$$f\left(x_2 \mid x_1\right) = \frac{1}{\sqrt{2\pi\sigma_2^2\left(1-\rho^2\right)}} \exp\left\{-\left(\frac{1}{2\left(1-\rho^2\right)\sigma_2^2}\right)\left[x_2 - \left(\mu_2 + \frac{\left(x_1-\mu_1\right)\rho\sigma_2}{\sigma_1}\right)\right]^2\right\} \quad -\infty < x_2 < \infty$$

$$f\left(x_1 \mid x_2\right) = \frac{1}{\sqrt{2\pi\sigma_1^2\left(1-\rho^2\right)}} \exp\left\{-\left(\frac{1}{2\left(1-\rho^2\right)\sigma_1^2}\right)\left[x_1 - \left(\mu_1 + \frac{\left(x_2-\mu_2\right)\rho\sigma_1}{\sigma_2}\right)\right]^2\right\} \quad -\infty < x_1 < \infty$$

$$X_2 \mid X_1 = x_1 \sim N\left[\mu_2 + \frac{\left(x_1-\mu_1\right)\rho\sigma_2}{\sigma_1}, \sigma_2^2\left(1-\rho^2\right)\right] \qquad X_1 \mid X_2 = x_2 \sim N\left[\mu_1 + \frac{\left(x_2-\mu_2\right)\rho\sigma_1}{\sigma_2}, \sigma_1^2\left(1-\rho^2\right)\right]$$

Multivariate Normal Properties

- If $\boldsymbol{X} \sim \mathrm{N}_p(\boldsymbol{\mu_x}, \ \boldsymbol{\Sigma})$, then for $1 \leq k \neq l \leq p$,

$$\mathrm{cov}(X_k, X_l) = 0 \text{ iff } X_k \text{ and } X_l \text{ are independent}$$

- If $\boldsymbol{A}$ is a full rank matrix of constants, then:

$$\boldsymbol{W} = \boldsymbol{AX} \sim \mathrm{N}\left(\boldsymbol{A\mu_x}, \boldsymbol{A\Sigma_x A'}\right)$$

- Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be independent with a common covariance matrix:

$$\boldsymbol{X}_i \sim \mathrm{N}_p(\boldsymbol{\mu}_i, \ \boldsymbol{\Sigma}),\ i = 1, \ldots, n$$

Define

$$\boldsymbol{W} = a_1\boldsymbol{X}_1 + \ldots + a_n\boldsymbol{X}_n$$

Then

$$\boldsymbol{W} \sim N_p\left(\sum_{i=1}^{n} a_i \boldsymbol{\mu}_i \, , \left(\sum_{i=1}^{n} a_i^2\right)\boldsymbol{\Sigma}\right)$$

Results Involving Multivariate Normal

Theorem: Let $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Then:

(a) $(\boldsymbol{X} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1}(\boldsymbol{X} - \boldsymbol{\mu}) \sim \chi^2_p$

(b) The probability that $\boldsymbol{X}$ is inside the solid ellipsoid

$$\{\boldsymbol{X}: (\boldsymbol{X} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1}(\boldsymbol{X} - \boldsymbol{\mu}) \leq \chi^2_p(\alpha)\}$$

is $1 - \alpha$, where $\chi^2_p(\alpha)$ denotes the upper α percentile of the χ^2_p distribution.

Univariate case: $X \sim N_1(\mu, \sigma^2)$

$$Z = \frac{X - \mu}{\sigma} \sim N(0,1)$$

$$(x - \mu)\sigma^{-2}(x - \mu) = Z^2 \sim \chi^2_1$$

Example

Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{60}$ be a random sample of size 60 from a four-variate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Specify the distribution of:

(a) $\overline{\boldsymbol{X}}$

(b) $(\boldsymbol{X}_1 - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_1 - \boldsymbol{\mu})$

(c) $n(\overline{\boldsymbol{X}} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1}(\overline{\boldsymbol{X}} - \boldsymbol{\mu})$

Example - Solution

Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{60}$ be a random sample of size 60 from a four-variate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Specify the distribution of:

(a) $\bar{\boldsymbol{X}}$

(b) $(\boldsymbol{X}_1 - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_1 - \boldsymbol{\mu})$

(c) $n(\bar{\boldsymbol{X}} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{X}} - \boldsymbol{\mu})$

Solution:

Given: $n = 60$, $p = 4$ and $X_i \sim \mathrm{N}_4(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $i = 1, \ldots, 60$

a) $\bar{X}$ is a 4×1 vector:

$$\bar{\boldsymbol{X}} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{1}{n} x_1 + \frac{1}{n} x_2 + \cdots + \frac{1}{n} x_n$$

$$\Rightarrow \bar{\boldsymbol{X}} \sim N_4\left(\frac{1}{n}\boldsymbol{\mu} + \frac{1}{n}\boldsymbol{\mu} + \ \ldots + \frac{1}{n}\boldsymbol{\mu}, \frac{1}{n^2}\boldsymbol{\Sigma} + \ \ldots + \frac{1}{n^2}\boldsymbol{\Sigma}\right) = \mathrm{N}_4\left(\boldsymbol{\mu}, \frac{1}{n}\boldsymbol{\Sigma}\right)$$

b) $(\boldsymbol{X}_1 - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_1 - \boldsymbol{\mu}) \sim \chi^2{}_4$

Multivariate Normal Likelihood Function

Let $\boldsymbol{X}_1, \dots, \boldsymbol{X}_{\mathrm{n}} \sim \mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Then the joint pdf is:

$$\prod_{i=1}^{n}\left[(2\pi)^{-\frac{p}{2}}|\Sigma|^{-\frac{1}{2}}e^{-\frac{1}{2}(x_i-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu})}\right] = (2\pi)^{-\frac{np}{2}}|\Sigma|^{-\frac{n}{2}}e^{-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{x}_i-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu})}$$

Now, working with the exponential term multiplied by -1/2, we have:

$$\sum_{i=1}^{n}(\boldsymbol{x}_i-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu}) = \mathrm{tr}\left[\sum_{i=1}^{n}(\boldsymbol{x}_i-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu})\right]$$

$$= \mathrm{tr}\left[\sum_{i=1}^{n}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu})(\boldsymbol{x}_i-\boldsymbol{\mu})'\right] = \mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}\sum_{i=1}^{n}(\boldsymbol{x}_i-\boldsymbol{\mu})(\boldsymbol{x}_i-\boldsymbol{\mu})'\right]$$

and

$$\sum_{i=1}^{n}(\boldsymbol{x}_i-\boldsymbol{\mu})(\boldsymbol{x}_i-\boldsymbol{\mu})' = \sum_{i=1}^{n}(\boldsymbol{x}_i-\overline{\boldsymbol{x}}+\overline{\boldsymbol{x}}-\boldsymbol{\mu})(\boldsymbol{x}_i-\overline{\boldsymbol{x}}+\overline{\boldsymbol{x}}-\boldsymbol{\mu})'$$

$$= \sum_{i=1}^{n}(\boldsymbol{x}_i-\overline{\boldsymbol{x}})(\boldsymbol{x}_i-\overline{\boldsymbol{x}})' + \sum_{i=1}^{n}(\overline{\boldsymbol{x}}-\boldsymbol{\mu})(\overline{\boldsymbol{x}}-\boldsymbol{\mu})' + 2\sum_{i=1}^{n}(\boldsymbol{x}_i-\overline{\boldsymbol{x}}))(\overline{\boldsymbol{x}}-\boldsymbol{\mu})'$$

$$= \sum_{i=1}^{n}(\boldsymbol{x}_i-\overline{\boldsymbol{x}})(\boldsymbol{x}_i-\overline{\boldsymbol{x}})' + n(\overline{\boldsymbol{x}}-\boldsymbol{\mu})(\overline{\boldsymbol{x}}-\boldsymbol{\mu})'$$

Therefore, the joint likelihood function is:

$$(2\pi)^{-\frac{np}{2}}|\Sigma|^{-\frac{n}{2}}e^{-\frac{1}{2}\boldsymbol{tr}\left[\boldsymbol{\Sigma}^{-1}\left(\sum_{i=1}^{n}(\boldsymbol{x}_i-\overline{\boldsymbol{x}})(\boldsymbol{x}_i-\overline{\boldsymbol{x}})'+n(\overline{\boldsymbol{x}}-\boldsymbol{\mu})(\overline{\boldsymbol{x}}-\boldsymbol{\mu})'\right)\right]}$$

Maximum Likelihood Estimator of $\boldsymbol{\mu}$

Likelihood Function:

$$(2\pi)^{-np/2}|\boldsymbol{\Sigma}|^{-n/2}\exp\left\{-\frac{1}{2}\mathrm{tr}\left\{\boldsymbol{\Sigma}^{-1}\left[\sum_{j=1}^{n}(\mathbf{x}_j-\bar{\mathbf{x}})(\mathbf{x}_j-\bar{\mathbf{x}})'+n(\bar{\mathbf{x}}-\boldsymbol{\mu})(\bar{\mathbf{x}}-\boldsymbol{\mu})'\right]\right\}\right\}=$$

$$(2\pi)^{-np/2}|\boldsymbol{\Sigma}|^{-n/2}\exp\left\{-\frac{1}{2}\left[\mathrm{tr}\left\{\boldsymbol{\Sigma}^{-1}\left[\sum_{j=1}^{n}(\mathbf{x}_j-\bar{\mathbf{x}})(\mathbf{x}_j-\bar{\mathbf{x}})'\right]\right\}+\mathrm{tr}\left\{\boldsymbol{\Sigma}^{-1}n(\bar{\mathbf{x}}-\boldsymbol{\mu})(\bar{\mathbf{x}}-\boldsymbol{\mu})'\right\}\right]\right\}$$

$$(2\pi)^{-np/2}|\boldsymbol{\Sigma}|^{-n/2}\exp\left\{-\frac{1}{2}\left[\mathrm{tr}\left\{\boldsymbol{\Sigma}^{-1}\left[\sum_{j=1}^{n}(\mathbf{x}_j-\bar{\mathbf{x}})(\mathbf{x}_j-\bar{\mathbf{x}})'\right]\right\}+n(\bar{\mathbf{x}}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\bar{\mathbf{x}}-\boldsymbol{\mu})\right]\right\}$$

Maximum Likelihood Estimator for $\boldsymbol{\mu}$:

$$L(\boldsymbol{\mu},\boldsymbol{\Sigma})=(2\pi)^{-np/2}|\boldsymbol{\Sigma}|^{-n/2}\exp\left\{-\frac{1}{2}\left[\mathrm{tr}\left\{\boldsymbol{\Sigma}^{-1}\left[\sum_{j=1}^{n}(\mathbf{x}_j-\bar{\mathbf{x}})(\mathbf{x}_j-\bar{\mathbf{x}})'\right]\right\}\right]\right\}\exp\left\{-\frac{1}{2}\left[n(\bar{\mathbf{x}}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\bar{\mathbf{x}}-\boldsymbol{\mu})\right]\right\}$$

maximized when $\hat{\boldsymbol{\mu}}=\bar{\mathbf{X}} \quad \Rightarrow$

$\exp\left\{-\frac{1}{2}\left[n(\bar{\mathbf{x}}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\bar{\mathbf{x}}-\boldsymbol{\mu})\right]\right\}=1$ is at its maximum since $\boldsymbol{\Sigma}^{-1}$ is positive definite

Maximum Likelihood Estimator of $\boldsymbol{\Sigma}$

Result: $\mathbf{B} \equiv p \times p$ positive definite, scalar $b > 0$, $\boldsymbol{\Sigma} \equiv$ positive definite:

$$\frac{1}{|\boldsymbol{\Sigma}|^{b}} \exp\left\{-\frac{1}{2}\operatorname{tr}\left\{\boldsymbol{\Sigma}^{-1}\mathbf{B}\right\}\right\} \le \frac{1}{|\mathbf{B}|^{b}}(2b)^{bp} e^{-bp} \quad \text{with equality holding at } \boldsymbol{\Sigma} = \left(\frac{1}{2b}\right)\mathbf{B}$$

Maximum Likelihood Estimator for $\boldsymbol{\Sigma}$ evaluated at $\hat{\boldsymbol{\mu}}$:

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-np/2} |\boldsymbol{\Sigma}|^{-n/2} \exp\left\{-\frac{1}{2}\left[\operatorname{tr}\left\{\boldsymbol{\Sigma}^{-1}\left[\sum_{j=1}^{n}(\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})'\right]\right\}\right]\right\} \quad \text{setting} \quad b = \frac{n}{2} \quad \mathbf{B} = \sum_{j=1}^{n}(\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})'$$

$$\Rightarrow \quad \hat{\boldsymbol{\Sigma}} = \left(\frac{1}{2(n/2)}\right)\sum_{j=1}^{n}(\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})' = \frac{1}{n}\sum_{j=1}^{n}(\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})' = \frac{n-1}{n}\mathbf{S}$$

Likelihood Function evaluated at the observed ML estimates:

$$L(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}) = (2\pi)^{-np/2} \left|\hat{\boldsymbol{\Sigma}}\right|^{-n/2} \exp\left\{-\frac{1}{2}\left[\operatorname{tr}\left\{\left(\hat{\boldsymbol{\Sigma}}\right)^{-1}\left[\sum_{j=1}^{n}(\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})'\right]\right\}\right]\right\} = (2\pi)^{-np/2} \left|\hat{\boldsymbol{\Sigma}}\right|^{-n/2} e^{-np/2}$$

Note: $\left|\hat{\boldsymbol{\Sigma}}\right| = \left|\frac{n-1}{n}\mathbf{S}\right| = \left(\frac{n-1}{n}\right)^{p} |\mathbf{S}| \quad \Rightarrow$

$$L(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}) = (2\pi)^{-np/2} e^{-np/2} \left(\frac{n-1}{n}\right)^{p} |\mathbf{S}| \quad = \text{ constant} \times \text{generalized inverse}$$

Results for ML Estimators and Large-Sample Properties

The likelihood function $L(\boldsymbol{\mu}, \boldsymbol{\Sigma};\ \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ depends on the observed data only through $\overline{\boldsymbol{x}}$ and $\boldsymbol{S} = \frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{x}_i - \overline{\boldsymbol{x}})'(\boldsymbol{x}_i - \overline{\boldsymbol{x}})$, that is, $\overline{\boldsymbol{x}}$ and $\boldsymbol{S}$ are *sufficient* statistics.

Sampling distributions:

$$\overline{\boldsymbol{X}} \sim N_p\left(\boldsymbol{\mu}, \frac{1}{n}\boldsymbol{\Sigma}\right), (n-1)\boldsymbol{S} \sim \text{Wishart with df} = n-1$$

Also, $\overline{\boldsymbol{X}}$ and $\boldsymbol{S}$ are independent r.v.s

Multivariate Law of Large Numbers:

$$\overline{\boldsymbol{X}} \xrightarrow{P} \boldsymbol{\mu}$$

Multivariate Central Limit Theorem:

Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be iid random vectors with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Then:

$$\sqrt{n}(\overline{\boldsymbol{X}} - \boldsymbol{\mu}) \xrightarrow{d} N_p(0, \boldsymbol{\Sigma})$$

and

$$\sqrt{n}(\overline{\boldsymbol{X}} - \boldsymbol{\mu})'\boldsymbol{S}^{-1}(\overline{\boldsymbol{X}} - \boldsymbol{\mu}) \xrightarrow{d} \chi_p^2$$

Wishart Distribution

- It is a generalization to multidimensions of the Chi-Square distribution.
- The Wishart distribution is a sum of outer products of random vectors.
- It is a *random matrix* which is symmetric and positive definite.
- Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \sim \mathrm{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$ be independent. Then the distribution of the $p \times p$ random matrix $\boldsymbol{M} = \sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i'$ is said to have the Wishart distribution with df = n.
- It defines the distribution of the sample covariance matrix.
- Notation: $\boldsymbol{M} \sim W_p(\boldsymbol{\Sigma}, n)$

Obtaining samples in `R`:
To generate *n* random matrices, distributed according to the Wishart distribution with parameters Σ and df, W_p(Σ,m), where m = df.
Use Function: rWishart(n, `df`, Σ)

Wishart Distribution Properties

- Let $\boldsymbol{M} \sim W_p(\boldsymbol{\Sigma}, n)$
- $E(\boldsymbol{M}) = n\boldsymbol{\Sigma}$
- $\boldsymbol{M} \sim \boldsymbol{A}W_p(\boldsymbol{I}_p, n)\boldsymbol{A}'$, where $\boldsymbol{\Sigma} = \boldsymbol{A}\boldsymbol{A}'$ is the LU-decomposition
- Assume $n > p$ and $\boldsymbol{\Sigma}$ is invertible. Then the pdf of $\boldsymbol{M}$ is

$$f(\boldsymbol{m}, n, \boldsymbol{\Sigma}) = \frac{|\boldsymbol{m}|^{\frac{n-p-1}{2}} e^{-½\,\mathrm{tr}(\boldsymbol{m}\boldsymbol{\Sigma}^{-1})}}{2^{\frac{pn}{2}} \pi^{\frac{p(p-1)}{4}} |\boldsymbol{\Sigma}|^{\frac{n}{2}} \prod_{i=1}^{p} \Gamma\left(\frac{n+1-1}{2}\right)}$$

where the support is all symmetric positive definite matrices $\boldsymbol{m}$.