

**Statistics**  
**PhD course Tinbergen Institute**  
2002

Peter Spreij (UvA)

`www.science.uva.nl/~spreij`  
`spreij@science.uva.nl`

## Lecture 1a (10 september 2002)

- sample spaces
- probability measures
- computing probabilities: counting methods and the multiplication principle
- conditional probability
- independent events

## Lecture 1b (11 september 2002)

- random variables (discrete or continuous)
- distribution function
- frequency (or probability mass) function
- density function
- specific discrete distributions
- specific continuous distributions
- transformations of a random variable

## Lecture 2a (17 september 2002)

- random vectors (discrete or continuous)
- joint distribution function
- frequency (or probability mass) function
- density function
- independent random variables
- sums of independent random variables
- transformations of a random variable

**the density of a bivariate random vector**

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

## transformation rule

If  $X$  is a random variable and  $Y = g(X)$ , with  $g$  monotone and differentiable with inverse  $h$ , then

$$f_Y(y) = \frac{f_X(h(y))}{|g'(h(y))|}.$$

If  $X$  is a random variable and  $Y = g(X)$ , where  $g$  is invertible (with inverse  $h$ ) and differentiable, then

$$f_Y(y) = \frac{f_X(h(y))}{|J(h(y))|},$$

where

$$J(x) = \det \begin{pmatrix} \frac{\partial}{\partial x_1} g_1(x) & \cdots & \frac{\partial}{\partial x_n} g_1(x) \\ \vdots & & \vdots \\ \frac{\partial}{\partial x_1} g_n(x) & \cdots & \frac{\partial}{\partial x_n} g_n(x) \end{pmatrix}$$

## Lecture 2b (18 september 2002)

- expected value of a random variable (discrete and continuous)
- expectations of functions of a random variable
- expectation of linear combinations
- variance and standard deviation
- covariance and correlation
- conditional probability and conditional expectation
- prediction

## Lecture 3a (24 september 2002)

- Some extras
- Limit theorems
  - Laws of large numbers
  - Convergence in distribution
  - the Central Limit Theorem
- Distributions derived from the normal distribution
  - $\chi^2$  distribution
  - (student)  $t$  distribution
  - $F$  distribution



# Dependence and Correlation

**Important implication:** if  $X$  and  $Y$  are independent, then  $\text{Cov}(X, Y) = 0$ , so they are uncorrelated.

**BUT**, if  $X$  and  $Y$  are uncorrelated, they are not necessarily independent.

Example:

$Y \setminus X$	-1	0	+1	
0	1/4	0	1/4	1/2
1	0	1/2	0	1/2
	1/4	1/2	1/4	1

We see that  $\mathbb{E}X = 0$ ,  $\mathbb{E}(XY) = 0$ ,  
 so  $\text{Cov}(X, Y) = 0$ ,  
 but  $\mathbb{P}(X = 0, Y = 0) \neq \mathbb{P}(X = 0)\mathbb{P}(Y = 0)$ .

However.....

Remember that in general

$$\rho = \rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}.$$

Let  $X$  and  $Y$  have a bivariate normal distribution with parameters  $\mu_X$  (the expected value of  $X$ ),  $\mu_Y$ ,  $\sigma_X$  (the standard deviation of  $X$ ),  $\sigma_Y$  and (correlation coefficient)  $\rho$ .

We know that IN THIS CASE  $X$  and  $Y$  are independent iff  $\rho = 0$ .

**Hence for bivariate normal  $(X, Y)$  independence is equivalent to being uncorrelated!**

**Warning:** If  $X$  is normal and  $Y$  is normal, then it does NOT necessarily follow that  $(X, Y)$  is bivariate normal. But this certainly happens if one also knows that  $X$  and  $Y$  are independent.

## Covariance matrix

If  $X = (X_1, \dots, X_m)^\top$  and  $Y = (Y_1, \dots, Y_n)^\top$  are random vectors, then  $\text{Cov}(X, Y)$  is the  $m \times n$  matrix with elements

$$\text{Cov}(X, Y)_{ij} = \text{Cov}(X_i, Y_j).$$

For  $X = Y$  we write  $\text{Cov}(X)$  instead of  $\text{Cov}(X, Y)$ .

### Properties:

1.  $\text{Cov}(X)$  is a symmetric nonnegative definite matrix.
2. If a sub-vector of  $X$  is independent of a sub-vector of  $Y$ , then their corresponding covariance matrix is the zero matrix.
3. If  $X$  has expectation vector  $\mu$  and covariance matrix  $\Sigma$ , then  $Y = AX + b$  has expectation vector  $A\mu + b$  and covariance matrix  $A\Sigma A^\top$ .

## the multivariate normal distribution

Let a random  $n$ -vector  $X$  have expectation vector  $\mu$  and covariance matrix  $\Sigma$ . Assume that  $\Sigma$  is invertible. Then  $X$  is said to have multivariate normal distribution if the density of  $X$  is

$$\frac{1}{(2\pi)^{n/2} \det(\Sigma)^{1/2}} \exp\left(-\frac{1}{2}(x - \mu)^\top \Sigma^{-1}(x - \mu)\right).$$

### Properties:

1. Two non-overlapping sub-vectors of  $X$  are independent *iff* their covariance matrix is zero.
2. If  $X$  has a multivariate normal distribution with expectation  $\mu$  and covariance matrix  $\Sigma$ , then  $Y = AX + b$  ( $A$  a square invertible matrix,  $b$  a vector) also has a multivariate normal distribution, with expectation vector  $A\mu + b$  and covariance matrix  $A\Sigma A^\top$ .

## Relations between different types of convergence

Let  $X, X_1, X_2, \dots$  and  $Y_1, Y_2, \dots$  be random variables,  $c$  a real constant.

1. If  $X_n \xrightarrow{P} X$ , then also  $X_n \xrightarrow{d} X$ .
2. If  $X_n \xrightarrow{d} c$ , then also  $X_n \xrightarrow{P} c$ .
3. If  $X_n \xrightarrow{P} c$ , then also  $g(X_n) \xrightarrow{P} g(c)$ , if  $g$  is a continuous function.
4. If  $X_n \xrightarrow{d} X$  and  $Y_n \xrightarrow{d} c$ , then  $g(X_n, Y_n) \xrightarrow{d} g(X, c)$ , if  $g$  is continuous function (on  $\mathbb{R}^2$ ).

## Lecture 3b (25 september 2002)

- parameter estimation, consistency
- method of moments
- maximum likelihood, asymptotic distribution
- Cramer-Rao bound, optimality

## Lecture 4a (2 october 2002)

- hypothesis testing
- Neyman-Pearson, optimal tests
- properties of normal distribution
- (student)  $t$  distribution
- confidence intervals, relation with tests

## $\chi^2$ distribution

A random variable  $X$  is said to have a  $\chi^2$  distribution with  $n$  degrees of freedom ( $\chi_n^2$  distribution) if it has the same distribution as  $\sum_{i=1}^n Z_i^2$ , where the  $Z_i$  are *iid* standard normal random variables:

$$X \stackrel{d}{=} \sum_{i=1}^n Z_i^2.$$

## (student) $t$ distribution

A random variable  $X$  is said to have a  $t$  distribution with  $n$  degrees of freedom ( $t_n$  distribution) if

$$X \stackrel{d}{=} \frac{Z}{W/n},$$

where  $Z$  and  $W$  are independent random variables,  $Z$  having a standard normal distribution and  $W$  having a  $\chi_n^2$  distribution.

For large  $n$ , the  $t_n$  distribution is approximately normal.



**Theorem** Let  $X_1, \dots, X_n$  be a sample from a  $N(\mu, \sigma^2)$  distribution. Then

- (1)  $\bar{X}$  and  $\sum_{i=1}^n (X_i - \bar{X})^2$  are independent.
- (2)  $\frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2$  has a  $\chi^2$  distribution with  $n - 1$  degrees of freedom.
- (3) The statistic

$$\frac{\bar{X} - \mu}{s/\sqrt{n}}$$

has a  $t$ -distribution with  $n - 1$  degrees of freedom.

## Application - $t$ -statistic

The Gauss test statistic for  $\mu$  when we deal with a sample from the  $N(\mu, \sigma^2)$  distribution is

$$\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma},$$

which we can only use when  $\sigma$  is known. If this is not the case, we replace it in the above statistic with  $S = \left(\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2\right)^{1/2}$ . The resulting statistic is

$$\frac{\sqrt{n}(\bar{X} - \mu)}{S},$$

which has a  $t_{n-1}$  distribution.

## Confidence intervals based on MLE

Recall that (under some assumptions)

$$\sqrt{nI(\theta_0)}(\hat{\theta} - \theta_0) \stackrel{d}{\approx} N(0, 1).$$

Hence  $(1 - \alpha)$ -confidence interval for  $\theta_0$  would have limits

$$\hat{\theta} \pm \frac{z(\alpha/2)}{\sqrt{nI(\theta_0)}}.$$

But, since  $\theta_0$  is unknown this does not work. Instead we take the calculable confidence interval

$$\hat{\theta} \pm \frac{z(\alpha/2)}{\sqrt{nI(\hat{\theta})}}.$$

Justification: if  $I$  is continuous, also

$$\sqrt{nI(\hat{\theta})}(\hat{\theta} - \theta_0) \stackrel{d}{\approx} N(0, 1).$$

## Lecture 4b (3 october 2002)

- (generalized) likelihood ratio test
- regression
- least squares estimators
- matrix approach
- statistical properties of the estimators (mean, variance, confidence intervals)

## (generalized) likelihood ratio test

Neyman-Pearson test to testing  $H_0 : \theta = \theta_0$  against  $H_A : \theta = \theta_A$  rejects  $H_0$  for small values of

$$\frac{f_{\theta_0}(X)}{f_{\theta_A}(X)},$$

when  $X$  is observed and where the  $f_\theta$  are 'densities'.

For composite hypotheses testing this approach is generalized as follows. We consider  $H_0 : \theta \in \Theta_0$  and  $H_1 : \theta \in \Theta_A$ , where  $\Theta_0 \cap \Theta_A = \emptyset$ . Let  $\Theta = \Theta_0 \cup \Theta_A$ . The GLR test rejects the null-hypothesis for small values of

$$\Lambda = \Lambda(X) = \frac{\sup_{\theta \in \Theta_0} f_\theta(X)}{\sup_{\theta \in \Theta} f_\theta(X)}.$$

Remark: notice that the denominator is maximized by the Maximum likelihood estimator (if it exists).

To find the rejection region, one needs the distribution of  $\Lambda$  (under the null-hypothesis). Usually  $\Lambda$  and its distribution are difficult to handle. Therefore one uses an asymptotic result for the case when we observe a large sample  $X = (X_1, \dots, X_n)$ .

Under certain conditions one has the following result:

The distribution of  $-2 \log \Lambda(X)$  is approximately  $\chi_{d-d_0}^2$ , where  $d = \dim \Theta$  and  $d_0 = \dim \Theta_0$ . Hence the rejection set  $R$  is approximated by the set  $\{x : -2 \log \Lambda(x) \leq \chi_{d-d_0}^2(1 - \alpha)\}$ .