

# Advanced Mathematical Methods

WS 2019/20

## 4 Mathematical Statistics

PD Dr. Thomas Dimpfl

*Chair of Statistics, Econometrics and Empirical Economics*

EBERHARD KARLS  
UNIVERSITÄT  
TÜBINGEN



WIRTSCHAFTS- UND  
SOZIALWISSENSCHAFTLICHE  
FAKULTÄT

# Outline: Mathematical Statistics

- 4.1 Measure spaces
- 4.2 Random Variables
- 4.3 pdf and cdf
- 4.4 Expectation, Variance and Moments

# Readings

A. Papoulis and A. U. Pillai. *Probability, Random Variables and Stochastic Processes*.

Mc Graw Hill, fourth edition, 2002 Chapters 1-4

# Online References

MIT Course on Probabilistic Systems Analysis and Applied Probability (by John Tsitsiklis)

- ▶ Discrete RVs I: Concept of random variables, probability mass function, expected value, variance  
<https://www.youtube.com/watch?v=3MOahpLxj6A>
- ▶ Continuous RVs: probability density function, cumulative distribution function, expected value, variance  
[https://www.youtube.com/watch?v=mHfn\\_7ym6to](https://www.youtube.com/watch?v=mHfn_7ym6to)

## 4.1 Measure spaces

Notation:  $\Omega$

- ▶ fundamental measure (or probability, or sample) space
- ▶ consists of all points (singletons)  $\omega$  possible as the outcome to an experiment

Definition: Event

An Event  $A$  is a subset of  $\Omega$ . The empty event  $\emptyset$  and the whole space  $\Omega$  are also events.

## 4.1 Measure spaces

### Definition: Topological space

A topological space  $(\Omega, \mathcal{F})$  is a space  $\Omega$  together with a class  $\mathcal{F}$  of subsets of  $\Omega$ . The members of the set  $\mathcal{F}$  are called open sets.  $\mathcal{F}$  has the property that unions of any number of the sets in  $\mathcal{F}$  (finite or infinite, countable or uncountable) remain in  $\mathcal{F}$ , and intersections of finite numbers of sets in  $\mathcal{F}$  also remain in  $\mathcal{F}$ . The closed sets are those whose complements are in  $\mathcal{F}$ .

## 4.1 Measure spaces

### Definition: Sigma-Algebra

$\mathcal{F}$  is a sigma algebra if

- (i)  $A_k \in \mathcal{F}$  for all  $k$  implies  $\cup_{k=1}^{\infty} A_k \in \mathcal{F}$ ,
- (ii)  $A \in \mathcal{F}$  implies  $\bar{A} \in \mathcal{F}$ ,
- (iii)  $\emptyset \in \mathcal{F}$ .

### Theorem: Properties of a Sigma-Algebra

If  $\mathcal{F}$  is a sigma algebra, then

- (iv)  $\Omega \in \mathcal{F}$ ,
- (v)  $A_k \in \mathcal{F}$  for all  $k$  implies  $\cap_{k=1}^{\infty} A_k \in \mathcal{F}$ .

## 4.1 Measure spaces

### Definition: Measurable space

A pair  $(\Omega, \mathcal{F})$  where the former is a set and the latter a sigma-algebra of subsets of  $\Omega$  is called a measurable space.

### Definition: Probability measure

A probability measure is a measure  $P$  in the measurable space  $(\Omega, \mathcal{F})$  which satisfies the following properties:

- (i)  $P(A) \geq 0$  for all  $A$
- (ii)  $P(\Omega) = 1$
- (iii)  $P(\emptyset) = 0$
- (iv)  $P(\bar{A}) = 1 - P(A)$
- (v) monotonicity, subadditivity



## 4.1 Measure spaces

### Definition: Probability space

The triple  $(\Omega, \mathcal{F}, \mathcal{P})$  is called a probability space.

### Theorem: Conditional probability

For  $B \in \mathcal{F}$  with  $P(B) > 0$ ,  $Q(A) = P(A | B) = P(A \cap B) / P(B)$  is a probability measure on the same space  $(\Omega, \mathcal{F})$

## 4.2 Random Variables

### Definition: Measurable function

Let  $f$  be a function from a measurable space  $(\Omega, \mathcal{F})$  into the real numbers. The function  $f$  is measurable if for each Borel set  $B \in \mathcal{B}$ , the set  $\{\omega; f(\omega) \in B\} \in \mathcal{F}$ .

### Definition: Random variable

A random variable  $X$  is a measurable function from a probability space  $(\Omega, \mathcal{F}, \mathcal{P})$  into the real numbers  $\mathbb{R}$ .

## 4.3 Cumulative Distribution Functions

Probability distribution function: discrete case

$$f_X(x) = P(X = x)$$

requirements:

- ▶  $0 \leq P(X = x) \leq 1$
- ▶  $\sum_x f_X(x) = 1$

## 4.3 Cumulative Distribution Functions

(Probability) Density function: continuous case

it holds that  $P(X = x) = 0$

requirements:

$$\blacktriangleright P(a \leq X \leq b) = \int_a^b f_X(x) dx \geq 0$$

$$\blacktriangleright \int_{-\infty}^{\infty} f_X(x) dx = 1$$

## 4.3 Cumulative Distribution Functions

### Definition: Cumulative distribution function

The cumulative distribution function (cdf) of a random variable  $X$  is defined to be the function  $F_X(x) = P(X \leq x)$ , for  $x \in \mathbb{R}$ .

to get the cdf:

**discrete:**

$$F_X(x) = \sum_{X \leq x} f_X(x) = P(X \leq x)$$

**continuous:**

$$F_X(x) = \int_{-\infty}^x f_X(t) dt$$

## 4.3 Cumulative Distribution Functions

### Properties

- (i)  $F_X(+\infty) = 1; F_X(-\infty) = 0$
- (ii)  $F_X(x)$  is a nondecreasing function of  $x$ :  
if  $x_1 < x_2$ ,  $F_X(x_1) \leq F_X(x_2)$   
note: the event  $\{X \leq x_1\}$  is a subset of  $\{X \leq x_2\}$
- (iii) if  $F_X(x_0) = 0$ , then  $F_X(x) = 0 \quad \forall \quad x \leq x_0$
- (iv)  $P(X > x) = 1 - F_X(x)$   
events  $\{X \leq x\}$  and  $\{X > x\}$  are mutually exclusive and  
 $\{X \leq x\} \cup \{X > x\} = \Omega$
- (v)  $F_X(x)$  is continuous from the right:  
 $\lim_{x \rightarrow a^+} F_X(x) = F_X(a)$
- (vi)  $P(x_1 \leq X \leq x_2) = F_X(x_2) - F_X(x_1)$

## 4.4 Expectation, Variance and Moments

### Expectations of a random variable

$$E[X] = \begin{cases} \sum x_i f_X(x_i) & \text{if } x \text{ is discrete} \\ \int_{-\infty}^{\infty} x f_X(x) dx & \text{if } x \text{ is continuous} \end{cases}$$

$g(X)$  a measurable function of  $x$ , then:

$$E[g(X)] = \begin{cases} \sum g(x_i) f_X(x_i) & \text{if } x \text{ is discrete} \\ \int_{-\infty}^{\infty} g(x) f_X(x) dx & \text{if } x \text{ is continuous} \end{cases}$$

## 4.4 Expectation, Variance and Moments

### Calculation rules

- ▶  $E[a] = a$
- ▶  $E[bX] = b \cdot E[X]$
- ▶ linear transformation  $E[a + bX] = a + bE[X]$
- ▶  $E[g_1(X) + g_2(X)] = E[g_1(X)] + E[g_2(X)]$



## 4.4 Expectation, Variance and Moments

### Variance of a random variable

let  $g(X) = (X - E[X])^2$

$$\begin{aligned} \text{Var}[X] &= \sigma^2 = E[(X - E[X])^2] \\ &= \begin{cases} \sum (x_i - E[X])^2 f_X(x_i) & \text{if } x \text{ is discrete} \\ \int_{-\infty}^{\infty} (x - E[X])^2 f_X(x) dx & \text{if } x \text{ is continuous} \end{cases} \end{aligned}$$

## 4.4 Expectation, Variance and Moments

### Calculation rules

- ▶  $Var[a] = 0$
- ▶  $Var[X + a] = Var[X]$
- ▶  $Var[bX] = b^2 Var[X]$
- ▶  $Var[a + bX] = b^2 Var[X]$

**important result:**

$$Var[X] = E[X^2] - E[X]^2$$

## 4.4 Expectation, Variance and Moments

### Standardization

an important transformation: standardization of a random variable  $X$

$$\text{let } g(X) = \frac{X - \mu}{\sigma} = Z$$

$$Z = \frac{X - \mu}{\sigma} = \frac{-\mu}{\sigma} + \frac{1}{\sigma}X$$

$$\Rightarrow E[Z] = 0$$

$$\Rightarrow \text{Var}[Z] = 1$$

## 4.4 Expectation, Variance and Moments

### Chebychev Inequality

for any random variable  $X$  with finite expected value  $\mu$  and finite variance  $\sigma^2 > 0$  and a positive constant  $k$

$$P(\mu - k\sigma \leq X \leq \mu + k\sigma) \geq 1 - \frac{1}{k^2}$$

## 4.4 Expectation, Variance and Moments

### Skewness and Kurtosis

central moments of a random variable:

$$\mu_r = E[(X - \mu)^r]$$

as  $r$  grows,  $\mu_r$  tends to explode

solution: normalization

▶ skewness coefficient:  $\gamma = \frac{E[(X - \mu)^3]}{\sigma^3}$

▶ kurtosis:  $\kappa = \frac{E[(X - \mu)^4]}{\sigma^4}$

often reported as excess kurtosis  $\kappa - 3$