Mathematical statistics
Week 1: Essentials in probability theory

Péter Elek and Ádám Reiff

17th September, 2013

1. Random variables and probability distributions

1.1. Univariate random variables and distributions

Key concept: random variables (r.v.).

- (Revision: basic rules for the calculation of probabilities)
- Tentative description: a variable that can take different values in subsequent "experiments"
- Notation: X is the r.v., x-s are the values that it can take
- More formal definition: A function from a sample space S into the real numbers
  - Sample space: all possible outcomes of a particular experiment
  - Experiment: a procedure that can be repeated an infinite number of times, which has a well-defined set of possible outcomes
- Example: tossing a coin 10 times, and calculating the number of heads
- Types: discrete and non-discrete random variables

Discrete random variables.

- Can take finite (or countably infinite) number of values $x_1, x_2, \ldots, x_k, \ldots$
- Examples: tossing a coin, throwing a dice
- Probability density (or mass) function (pdf, pmf): gives the probability of each possible value, $\Pr(X = x_j) = p_j$, with $\sum_{i=1}^{k} p_i = 1$ (or $\sum_{i=1}^{\infty} p_i = 1$)
  - For the dice-throwing example, $\Pr(X = 1) = 1/6, \Pr(X = 2) = 1/6, \ldots, \Pr(X = 6) = 1/6$

- For the coin-throwing example, $\Pr(X = Head) = 1/2, \Pr(X = Tail) = 1/2$

Discrete distributions: examples.

- Bernoulli (p): X can take two values, 0 and 1. $\Pr(X = 1) = p$, $\Pr(X = 0) = 1 - p$.
  - Result of a random coin toss is Bernoulli.
- Binomial (n, p): sum of n (independent) Bernoulli distributed r.v-s: $Y = X_1 + X_2 + \cdot + X_n$, where X_i are all Bernoulli. Y can take $n + 1$ values: $0, 1, \cdot, n$. $\Pr(Y = k) = \binom{n}{k} p^k (1 - p)^{n-k}$.
  - Number of heads after tossing a coin ten times is Binomial.
- Poisson (λ): $\Pr(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$ for all $k \geq 0$ integer.
  - Number of doctor visits during a year for a particular person can be Poisson.
- Geometric (p): $\Pr(X = k) = p(1 - p)^{k-1}$ for all $k \geq 1$ integer.
  - The time until a head occurs in the coin-tossing example

Continuous random variables.

- Non-discrete r.v.: Can take (uncountably) infinitely many values (like $x \in \Re$, $x \in \Re^+$, $x \in [0; 1]$)
- Continuous r.v.: takes on any real value with zero probability
- Example: uniform distribution on $[0; 1]$.
  - It can take values in this interval, each value is equally likely. (Similar to throwing with a fair dice with infinitely many sides.)
- Example for a non-continuous and non-discrete variable: a mixture of a zero (with probability p) and a uniform (with probability $1 - p$)

Cumulative distribution function (cdf).

- Cumulative distribution function (cdf): $F(x) = \Pr(X \leq x)$
  - this is the probability of X taking a value not larger than x.
- Can be defined for any (e.g. discrete or continuous) r.v.
- Properties:

- $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$
- $F(x)$ is non-decreasing
- $F(x)$ is right-continuous
- $\Pr(X > a) = 1 - F(a)$
- $\Pr(a < X \leq b) = \Pr(X \leq b) - \Pr(X \leq a) = F(b) - F(a)$

Probability density function (pdf) for continuous r.v.-s.

- Probability density function (pdf) of a continuous variable: $f(x) = F'(x)$, the derivative of the cdf
- It gives probabilities of X taking value *in a given range*: $\Pr(a < X \leq b) = \int_a^b f(t)dt$
- Note: the probability density function $f(x)$ of a continuous r.v. does *not* represent the probability of any particular value, since $\Pr(X = x) = 0$ for each x.
- But $f(x)$ is still informative about the "typical" values: $\Pr(x \leq X \leq x + dx) \approx f(x)dx$ if dx is small.
- Properties:
  - $f(x) \geq 0$
  - $\int_{-\infty}^{\infty} f(x)dx = 1$

Continuous distributions: examples.

- Uniform (a, b): takes values on $[a; b]$ and each value is equally likely. $f(x) = 1/(b - a)$ if $a \leq x < b$ and zero otherwise.
- Normal (μ, σ): $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, where μ and $\sigma > 0$ are parameters
- Exponential (θ): $f(x) = \frac{1}{\theta} e^{-\frac{x}{\theta}}$, $(x > 0)$, θ is the parameter
- Lognormal and Gamma distributions (see later)

Transformation of random variables.

- Let X be a r.v. and g a real-valued function. Then $Y = g(X)$ is a r.v. as well.
- Let X be a random variable with cdf $F_X(x)$ and pdf $f_X(x) = F'(x)$. Suppose $Y = g(X)$ for some strictly increasing function $g()$. Then the cdf $F_Y(y)$ and pdf $f_Y(y)$ of random variable Y are $F_Y(y) = F_X\left(g^{-1}(y)\right)$ and $f_Y(y) = \frac{1}{g'(y)} f_X\left[g^{-1}(y)\right]$.

– Proof: for each a, $F_X(a) = \Pr(X \leq a)$. Then $F_Y(a) = \Pr(Y \leq a) = \Pr(g(X) \leq a) = \Pr\left(X \leq g^{-1}(a)\right) = F_X\left[g^{-1}(a)\right]$. (One can take the inverse as $g()$ is strictly monotone.) $f_Y(a)$ is obtained simply by taking the derivative of $F_Y(a)$.

Simulation of distributions.

- A useful result: u is a uniformly distributed random variable on $[0, 1]$, and $F()$ is the cdf of an arbitrary continuous distribution. Then the random variable $F^{-1}(u)$ is distributed according to the distribution defined by F. (F^{-1} is the inverse of function F.)
- Proof: in the problem set

1.2. Joint distributions

Joint distribution and independence.

- Let X and Y two discrete random variables with possible values $x_1, x_2, \ldots, x_k$ and $y_1, y_2, \ldots, y_l$. Then the *joint pdf* of X and Y is $f_{X,Y}(x, y) = \Pr(X = x, Y = y)$. Usual notation: $\Pr(X = x_i, Y = y_j) = p_{ij}$ with $\sum_{i=1}^{k} \sum_{j=1}^{l} p_{ij} = 1$.
- A similar definition exists for continuous r.v.-s.
- For more than two r.v.-s: $f(x_1, x_2, \ldots x_n) = \Pr(X_1 = x_1, X_2 = x_2, \ldots X_n = x_n)$
- X and Y are *independent* if and only if $\Pr(X = x, Y = y) = \Pr(X = x) \cdot \Pr(Y = y)$ for each (x, y).
- Similarly for more than two r.v.-s.
- Note: this can also be written as $f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y)$, where $f_{X,Y}$ is the joint pdf, and f_X, f_Y are the *marginal probability density functions.*

Example.

- X can take x_1 and x_2, Y can take y_1 and y_2, with the following pdf:

	y_1	y_2	$p_{i\cdot}$
x_1	0.4	0.3	0.7
x_2	0.2	0.1	0.3
$p_{\cdot j}$	0.6	0.4	1.0

- Then the marginal distributions:

- $\Pr(X = x_1) = 0.7$, $\Pr(X = x_2) = 0.3$
- $\Pr(Y = y_1) = 0.6$, $\Pr(Y = y_2) = 0.4$

- Is it true that $\Pr(X = x, Y = y) = \Pr(X = x) \cdot \Pr(Y = y)$ for each (x, y)? Are X and Y independent?
- Modify the joint pdf of X and Y in such a way that they become independent (with the same marginal pdf-s)!

Conditional distributions.

- X and Y are two random variables. *Conditional distribution* of Y on X: the distribution of Y given that X takes a certain value x.
- Conditional pdf: $f_{Y|X}(y \mid x) = \Pr(Y = y \mid X = x)$, i.e. the probability of $Y = y$ given that $X = x$.
- Note: $f_{Y|X}(y \mid x)$ can also be written as $f_{X,Y}(x, y)/f_X(x)$.
- Independence of X and Y means that the conditional distribution of Y on X does not depend on X: $f_{Y|X}(y \mid x) = f_{X,Y}(x, y)/f_X(x) = f_X(x) \cdot f_Y(y)/f_X(x) = f_Y(y)$
- Example (cont.): in the previous example,
  - The conditional distribution of Y given that $X = x_1$: $\Pr(Y = y_1|X = x_1) = 4/7$, $\Pr(Y = y_2|X = x_1) = 3/7$.
  - The conditional distribution of Y given that $X = x_2$: $\Pr(Y = y_1|X = x_2) = 2/3$, $\Pr(Y = y_2|X = x_2) = 1/3$.

2. Numerical measures of probability distributions

2.1. Measures of central tendency

Numerical measures of central tendency.

- These show the "typical" element of the distribution.
- Expected value: the "average" value, weighted according to the probability distribution
- Median (of a continuous r.v.): the number m such that $F(m) = 1/2$
  - i.e. 50% of all elements are smaller, and 50% of all elements are higher.
  - For discrete r.v.-s it may not be unique. E.g. in the coin-tossing example every $0 \leq m \leq 1$ may be a median.

- Mode: the element of which the probability is highest, or where $f(x)$ takes its maximum value
  - It may not be unique.
- For symmetric distributions, all measures yield the same result.

Expected value.

- Expected value: $E(X)=\sum_{i=1}^{\infty} p_i x_i$ or $E(X)=\int_{-\infty}^{\infty} x f(x) dx$
- Properties:
  - Expected value of constant c: $E(c)=c$
  - If a, b constant, X is a random variable, then $E(aX+b)=aE(X)+b$
  - Expected value of a sum (or a linear combination) equals the sum (or linear combination) of expected values: for all $(a_1, a_2, \ldots, a_n)$ real numbers and $(X_1, X_2, \ldots, X_n)$ r.v.-s $E(a_1X_1+\ldots+a_nX_n)=a_1E(X_1)+\ldots+a_nE(X_n)$.
- Minimizing property of the expected value: the expression $E\left((X-b)^2\right)$ is minimized if $b=E(X)$.

Examples.

- Bernoulli (p): $E(X)=p*1+(1-p)*0=p$
- Binomial (n,p): $E(X)=np$ (why?)
- Normal (μ,σ): $E(X)=\mu$ (why?)
- Exponential (θ): $E(X)=\theta$ (why?)
- What is the median and mode of these distributions?

Expected value of transformations.

- Let X be a r.v. and g a real-valued function. Then $Y=g(X)$ is a r.v. as well.
- To calculate $E(g(X))$, we do not need to determine the distribution of $g(X)$ since it can be calculated as $E(g(X))=\sum_{i=1}^{\infty} p_i g(x_i)$ or $E(g(X))=\int_{-\infty}^{\infty} g(x) f(x) dx$.
- Note: for nonlinear g functions $E(g(X))\neq g(E(X))$.
  - Jensen's inequality: if g is convex, then $E(g(X))\geq g(E(X))$.
- Example: calculate $E(X^2)$ for the dice-throwing example!

2.2. Measures of variation

Numerical measures of variation.

- These measure the variability of the random variable.
- Range: the difference between the largest and smallest (possible) element
- Mean absolute deviation: the expected value of the absolute deviation from the mean: $E\left(|X - E(X)|\right)$
- Variance: the expected value of the squared deviation from the mean: $Var(X) = E\left[(X - E(X))^2\right] = E\left(X^2\right) - (E(X))^2$
  - For discrete random variables, $E(X^2) = \sum_{i=1}^{k} p_i x_i^2$, for continuous random variables $E\left(X^2\right) = \int_{-\infty}^{\infty} x^2 f(x)dx$.
- Standard deviation: the square root of the variance: $sd(X) = \sigma_X = \sqrt{E\left(X^2\right) - (E(X))^2}$

Variance and standard deviation.

- Properties:
  - For a, b constants and an X random variable, $Var(aX+b) = a^2 Var(X)$ and $sd(aX + b) = |a|\, sd(X)$.
  - The variance and standard deviation of a constant c are 0: $Var\left(c\right) = 0$.
- Standardization of a random variable: let $E\left(X\right) = \mu$ and $sd\left(X\right) = \sigma$. Then $Z = \frac{X-\mu}{\sigma}$ is the standardized r.v., for which $E\left(Z\right) = 0$ and $sd\left(Z\right) = 1$.

Examples.

- Variance of a throw with dice: $\frac{35}{12}$
- Variance of a Bernoulli (p) variable is $p(1-p)$
- Variance of a Binomial (n, p) variable is $np(1-p)$
- Variance of a standard normal random variable (with $\mu = 0$, $\sigma = 1$) is $Var(X) = 1$ (proof: with integration by parts)
- Variance of a Normal (μ, σ) variable is $Var(X) = \sigma^2$.

2.3. Other measures

Higher moments of the distributions.

- n-th moment of a distribution: $E\left(X^n\right)$
- n-th central moment of a distribution: $E\left\{\left[X - E\left(X\right)\right]^n\right\}$
- Variance = second central moment
- Skewness (standardized third central moment) = $E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right]$
  - Skewness: It measures the asymmetry of the distribution.
- Kurtosis (standardized fourth central moment) = $E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right]$
  - Kurtosis of the normal distribution is 3.
  - It measures the "peakedness" of the distribution and that how "heavy" its tails are.

Quantiles: other descriptive measures of the distributions.

- Quartiles: the "thresholds" between the quarters of distributions.
  - 25% of the distribution is smaller than the first quartile, and 75% is bigger.
  - 50% of the distribution is smaller than the second quartile, and 50% is bigger. (So second quartile = median.)
- Deciles: the "thresholds" between each 10% of the distribution.
  - Example: 30% of the distribution is smaller than the third decile, and 70% is bigger.
- Percentile: the "thresholds" between each 1% of the distribution.
  - Example: 72% of the distribution is smaller than the seventy-second percentile, and 28% is bigger.

2.4. Measures for joint distributions

Measures of association.

- Covariance: $Cov(X,Y) = \sigma_{XY} = E\left[\left(X - E(X)\right)\left(Y - E(Y)\right)\right]$
- A useful property: $Cov(X,Y) = E(XY) - E(X)E(Y)$.
- Correlation: $Corr(X,Y) = \frac{Cov(X,Y)}{sd(X) \cdot sd(Y)} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$

Properties of covariance and correlation.

- If X and Y are independent, then $Cov(X,Y) = 0$
  - To see why, use the definition of expected value.
  - The inverse is not true!!! Example: $\Pr(X = 1) = 0.25$, $\Pr(X = 0) = 0.5$, $\Pr(X = -1) = 0.25$, and let $Y = X^2$.
- If a, b, c, d constant, X, Y are random variables, then $Cov(aX+b, cY+d) = acCov(X,Y)$.
- Cauchy-Schwartz-inequality: $|Cov(X,Y)| \leq sd(X) \cdot sd(Y)$
- $-1 \leq Corr(X,Y) \leq 1$ (follows from Cauchy-Schwartz)
  - $Corr(X,Y) = 1$ if and only if there is a perfect positive linear relationship.
  - $Corr(X,Y) = -1$ if and only if there is a perfect negative linear relationship.
- Correlation is scale-invariant, i.e. $Corr(aX+b, cY+d) = sign(ac)Corr(X,Y)$.

Further properties of variance.

- $Var(aX + bY) = a^2Var(X) + b^2Var(Y) + 2abCov(X,Y)$
- So $Var(aX + bY) = a^2Var(X) + b^2Var(Y)$ *if only if* X and Y are uncorrelated.
- Similarly, if X and Y are uncorrelated, then $Var(X - Y) = Var(X) + Var(Y)$.
- If $\{X_1, \ldots, X_n\}$ are pairwise uncorrelated r.v.-s and $\{a_1, \ldots, a_n\}$ are real numbers, then $Var\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i^2 Var(X_i)$.

Conditional expectation.

- $E(Y \mid X = x)$ is the expected value of Y given that X takes a certain value of x. It is just a function of x which tells how the expected value of Y varies with the values of X.
- $E(Y \mid X = x) = \sum_{j=1}^k y_j \Pr(Y = y_j \mid X = x) = \sum_{j=1}^k y_j f_{Y|X}(y_j \mid x)$
- Example: X is years of education, Y is yearly wage. Then $E(Y \mid X = 12)$ (the expected wage of those who went to school for 12 years) is probably higher than $E(Y \mid X = 6)$.
- Note: we can use the notation $E(Y \mid X)$ for the *random variable* which takes the value $E(Y \mid X = x)$ for $X = x$. This r.v. is a function of X.

Properties of conditional expectation.

- For any function $c()$: $E[c(X) \mid X] = c(X)$.
- For any functions $a()$ and $b()$: $E[a(X)Y + b(X) \mid X] = a(X)E(Y \mid X) + b(X)$
- If X and Y are independent, then $E(Y \mid X) = E(Y)$.
  - As a consequence, if $E(U) = 0$ and U is independent of X, then $E(U \mid X) = 0$.
- Minimizing property: $E(Y \mid X)$ minimizes the expected squared prediction error for Y among all functions of X.
  - Let $\mu(X) = E(Y \mid X)$. Then, for every $g()$ function, $E\left[(Y - \mu(X))^2 \mid X\right] \leq E\left[(Y - g(X))^2 \mid X\right]$ and $E\left[(Y - \mu(X))^2\right] \leq E\left[(Y - g(X))^2\right]$.

Law of iterated expectations.

- Law of iterated expectations: $E[E(Y \mid X)] = E(Y)$.
- A generalization: $E[E(Y \mid X)] = E[E(Y \mid X, Z) \mid X]$.
- A consequence: if $E(Y \mid X) = E(Y)$ then $Cov(X, Y) = 0$. Moreover, every function of X is uncorrelated with Y.
  - The converse is not true. Example?

Conditional variance.

- Similar to the unconditional case: $Var(Y \mid X = x) = E\left(Y^2 \mid X = x\right) - (E(Y \mid X = x))^2$
- A useful property:
  - If X and Y are independent, then $Var(Y \mid X = x) = Var(Y)$ for each x

Summary: alternative definitions of "independence".

- "Stochastic" independence: if $f(Y \mid X = x)$ is the same for all x. (This is what we defined, and we will understand "independence" as this.)
- Mean independence: if $E(Y \mid X = x)$ is the same for all x.
- Uncorrelatedness: if $Corr(X, Y) = \sigma_{XY} = 0$
- "Stochastic" independence implies mean independence. Mean independence implies uncorrelatedness. But the opposites are not true!!!

Summary of some common distributions.

Discrete	$\Pr(X=k)$		$E(X)$	$sd(X)$
Bernoulli	$p; 1-p$		p	$\sqrt{p(1-p)}$
Binomial	$\binom{n}{k}p^k(1-p)^{n-k}$		np	$\sqrt{np(1-p)}$
Poisson	$\frac{\lambda^k}{k!}e^{-\lambda}$		λ	$\sqrt{\lambda}$
Continuous	$f(x)$	$F(x)$	$E(X)$	$sd(X)$
Uniform	$\frac{1}{b-a}$	$\frac{x-a}{b-a}$	$\frac{a+b}{2}$	$\frac{b-a}{\sqrt{12}}$
Normal	$\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	$\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}}\,dt$	μ	σ
Exp.	$\frac{1}{\theta}e^{-\frac{x}{\theta}}$	$1-e^{-\frac{x}{\theta}}$	θ	θ

3. The normal distribution and related distributions

3.1. The normal distribution and its properties

It is of high importance because.

- Any linear combination of independent normal random variables is also normally distributed.
- It is the basis of many other distributions that are frequently used in statistics and econometrics: lognormal, chi-square, t-distribution, F-distribution.
- Many real variables are normally distributed (like body height of people, IQ-level of people etc). (Why?)
- Asympotics: many distributions have some relationship with the normal distribution in asymptotic terms (see later).

Properties of the normal distribution I.

- If $X \sim N(\mu, \sigma)$, then $aX + b \sim N(a\mu + b, |a|\sigma)$.
  - Proof: use the rule for the pdf of transformed variables.
- Hence, if $X \sim N(\mu, \sigma)$, then $(X-\mu)/\sigma \sim N(0,1)$ (the standard normal distribution).
- If X and Y are jointly normally distributed, then they are independent if and only if $Cov(X,Y) = 0$.
  - This is a special feature of the normal distribution!

Properties of the normal distribution II.

- Let $X \sim N(\mu_X, \sigma_X)$, $Y \sim N(\mu_Y, \sigma_Y)$ and the two variables independent. Then $X + Y \sim N\left(\mu_X + \mu_Y, \sqrt{\sigma_X^2 + \sigma_Y^2}\right)$.
  - Proof: start from $F_{X+Y}(a) = \Pr(X+Y \leq a) = \int_{-\infty}^{\infty} \Pr(Y \leq a-x \mid X) f_X(x)dx = \int_{-\infty}^{\infty} F_Y(a-x) \cdot f_X(x)dx$, and take the derivative with respect to a. Then tedious calculations show that $f_{X+Y}(a) = \frac{1}{\sqrt{\sigma_X^2+\sigma_Y^2}\sqrt{2\pi}} e^{-\frac{(a-\mu_X-\mu_Y)^2}{2\left(\sigma_X^2+\sigma_Y^2\right)}}$.
- Let $X \sim N(\mu_X, \sigma_X)$, $Y \sim N(\mu_Y, \sigma_Y)$ and the two variables independent. Then, for a and b constants, $aX+bY \sim N\left(a\mu_X + b\mu_Y, \sqrt{a^2\sigma_X^2 + b^2\sigma_Y^2}\right)$.
  - Proof: it follows from the previous results.

Calculation of normal probabilities.

- The standard normal cdf, $\Phi(z)$ cannot be determined in a closed form integral but can be calculated numerically.
- All normal probabilities can be expressed in terms of the standard normal cdf.
- Let $X \sim N(\mu, \sigma)$. Then $\Pr(a < X \leq b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$.
- Example: $\Pr\left(|\frac{X-\mu}{\sigma}| \leq 1.96\right) = \Phi(1.96) - \Phi(-1.96) = 2 \cdot \Phi(1.96) - 1 = 0.95$

3.2. Related distributions

Lognormal distribution.

- X is lognormally distributed if its logarithm is normally distributed
- If $Y \sim N(\mu, \sigma)$, then the random variable $X = e^Y$ is lognormally distributed. One can show: $E\left(e^Y\right) = e^{\mu+\frac{\sigma^2}{2}}$, $Var\left(e^Y\right) = \sigma^2 e^{2\mu+\sigma^2}$.

Chi-squared distribution.

- If $Z_1, Z_2, \ldots, Z_n$ are *independent standard normal* random variables, then $X = \sum_{i=1}^{n} Z_i^2$ follows a chi-squared distribution with n degrees of freedom.
- Its expected value and variance depends on the degrees of freedom: $E(X) = n$, $Var(X) = 2n$ (why?).

- It can take only positive values.
- Its distribution is very asymmetric.

t-distribution.

- If $Z \sim N(0,1)$ and $X \sim \chi^2_n$, independent from each other, then $t = \frac{Z}{\sqrt{X/n}}$ follows a t-distribution with n degrees of freedom.
- The shape is similar to the shape of standard normal distribution: it is symmetric, but has "heavier tails" (i.e. more extreme observations occur with higher frequency).
- Expected value is 0 for $n > 1$, variance is $\frac{n}{n-2}$ for $n > 2$ (otherwise the moments do not exist).
- As $n \longrightarrow \infty$, the t-distribution approaches the standard normal distribution. (A proof requires the law of large numbers.)

F-distribution.

- If $X_1 \sim \chi^2_{k_1}$ and $X_2 \sim \chi^2_{k_2}$, independent from each other, then $F = \frac{X_1/k_1}{X_2/k_2}$ follows an F-distribution with k_1 and k_2 degrees of freedom.
- It can take only positive values.
- $t_n^2 \sim F_{1,n}$.

Material.

- W Appendix B
- CB 1, 2.1-2.3 (pages 47-62, until Definiton 2.3.6), 3.1-3.3, 3.5, 4.1-4.3, 4.5-4.6.
  - CB is needed only to the extent covered in the lectures.