Summary

▼ Notations

- $\bar{\mathbb{R}} = [0, \infty] = \mathbb{R}_+ \cup \{\infty\}$
- $X_1 \overset{d}{=} X_2$ means they have the same distribution
- If $P(X_1 \neq X_2) = 1$ then we call them unequal **almost surely**

▼ Probability Theory

▼ Measuring spaces

Important terminology:

- Sample space Ω
- Event $A \subset \Omega$
- Elementary event / outcome $\omega \in \Omega$

Measuring space in the countable case:

1. $P(\Omega) = 1$
2. Sigma additivity for $A_1, ... \subset \Omega$ and $A_i \cup A_j = \emptyset \forall i \neq j$

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

Lemma 1.2

If a sample space Ω is counable, you can specify a probability measure just by (while I is an index-set):

$$P(\{\omega_i\}) = p_i \quad \forall i \in I$$

For every set A it holds:

$$P(A) = \sum_{\omega \in A} P(\{\omega\})$$

σ-algebra:

1. $\emptyset \in \mathcal{A}$
2. if $A \in \mathcal{A}$ then $A^c \in \mathcal{A}$
3. if $A_1, A_2, ... \in \mathcal{A} \Rightarrow \cup_{i=1}^{\infty} A_i \in \mathcal{A}$

Definition of smallest σ-Algebra:

- If the smallest sigma algebra containing set A is called $\mathcal{A}$. Then for every sigma Alegbra $\mathcal{B}$on Ω it holds that:

$$A \subset \mathcal{B} \Rightarrow \mathcal{A} \subset \mathcal{B}$$

Theres also the smallest-σ-Algebra, that is denoted with the notation $\sigma(A)$

Lemma 1.5 → For set $A \subset \mathcal{P}(\Omega)$ $\sigma(A)$ has a solution.

▼ Measure

Defintion Measure:

1. $\mu : \mathcal{A} \to [0, \infty]$
2. $\mu(\emptyset) = 0$
3. $A_1, A_2, ... \in \mathcal{A}$ pairwise disjoint σ-additivity:

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

Definition Probability measure:

1. $\mu : \mathcal{A} \to [0, \infty]$
2. $A_1, A_2, ... \in \mathcal{A}$ pairwise disjoint σ-additivity:

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

3. $P(\Omega) = 1$

Important measures:

- Counting measure $\nu(A) = |A|$
- dirac measure:

$$\delta_\omega(A) = \begin{cases} 1 \text{ if } \omega \in A \\ 0 \text{ else} \end{cases}$$

- Lebesgue measure:

$$\lambda([a, b]) = b - a$$

Furthermore:

$$\lambda(\{a\}) = 0$$

And (only maps from borel-algebras):

$$\lambda : \mathcal{B}([a, b]) \to \mathbb{R}_+$$

Properties of prob. measure (T. 1.9)

1. $P(\emptyset) = 0$
2. Finite additivity
3. $P(A^C) = 1 - P(A)$
4. $A \subset B$ implies $P(A) \leq P(B)$
5. $A \subset B \implies P(B \setminus A) = P(B) - P(A)$
6. Poincare Sylvester: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
7. Cont. from below: If $A_n \subset A_{n+1}$ then

$$P(A_n) \to P\left(\bigcup_{k=1}^{\infty} A_k\right)$$

8. Cont. from above: If $A_{n+1} \subset A_n$ then:

$$P(A_n) \to P\left(\bigcap_{k=1}^{\infty} A_k\right)$$

9. Sub σ additivity

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} P(A_n)$$

▼ Borel sets

Let $A := \{(a,b)|a,b \in \mathbb{R}\}$ then the **Borel sigma field** is defined by:

$$\sigma(A) = \mathcal{B}$$

Each set $C \subset \mathbb{R}$ is called a **borel set** iff $C \in \mathcal{B}$

We will further define a **field** as a family of subsets $\mathcal{A}^* \subset \mathcal{P}(\Omega)$ if:

1. $\emptyset \in \mathcal{A}^*$
2. $A \in \mathcal{A}^* \implies A^\complement \in \mathcal{A}^*$
3. $A_1, A_2, ... \in \mathcal{A}^* \implies A_1 \cup A_2 \in \mathcal{A}^*$

▼ Pre- Measures

Definition: let $\mathcal{A}^*$ be a **field.** Then a function $P^* : \mathcal{A}^* \to [0,\infty]$ is called a **pre-measure** iff for every sequence $A_1, A_2, ... \in \mathcal{A}^*$ with $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}^*$ it holds that:

$$P^*\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P^*(A_i)$$

Theorem of Carathéodory: let $\mathcal{A}^*$ be a field and $P^* : \mathcal{A}^* \to [0,\infty)$ be a **prob-pre-measure.** Then there is one and only one **measure** $P : \sigma(\mathcal{A}^*) \to [0,\infty)$ such that:

$$P(A) = P^*(A) \quad \forall A \in \mathcal{A}^*$$

▼ cdf and Lebesgue Stieltjes measure

Definition of cdf: Let $P : \mathcal{B} \to [0,\infty)$ be a probability measure on $(\mathbb{R}, \mathcal{B})$. Then the **cummulative distribution function** $F : \mathbb{R} \to [0,1]$ is defined by:

$$F(a) = P((-\infty, a]) \quad \forall a \in \mathbb{R}$$

Properties of a destribution function:

1. $P((a,b]) = F(b) - F(a)$
2. $F(a) \leq F(b) \Leftrightarrow a \leq b$
3. For all sequences $(b_n \in \mathbb{R})_{n\in\mathbb{N}}$ monotonously decreasing with $b_n \to b$ it holds that: $F(b_n) \to F(b)$
4. $\lim\limits_{x\to\infty} F(x) = 1$ and $\lim\limits_{x\to-\infty} F(x) = 0$

We now have derived a **distribution function** from a probability measure. **Theorem 1.16** now states, that for every real function $F : \mathbb{R} \to [0,1]$, that satisfies properties 2 -4 from above, there exists one and only one **probability measure** $P : \mathcal{B} \to [0,\infty)$ with: $F(b) = P((-\infty, b])$

Every probability measure, that is characterized by such a function is now called **Lebesgue-stieltjes-measure**

The L**ebesque measure** $\lambda : \mathcal{B} \to [0,\infty)$ is defined by:

$$\lambda((a,b]) = b - a$$

▼ probability mass function and pdf

Definition of pmf: Let $f : \mathbb{R} \to \mathbb{R}_+$. Then f is called a pmf iff:

$$\sum_{x\in\mathcal{S}_f} f(x) = 1 \quad \text{with} \quad \mathcal{S}_f = \{x \in \mathbb{R} : f(x) > 0\}$$

$\mathcal{S}_f$ is called the **support** and must be **countable** in this definition**.** And we can define a corresponding **probability-measure** P **as:**

$$P(A) = \sum_{x \in (A \cap \mathcal{S}_f)} f(x)$$

The cdf is then defined as:

$$F(x) = P((-\infty, x]) = \sum_{a \in \mathcal{S}_f \cap (-\infty, x]} f(a)$$

Definition of pdf:

A function $f : \mathbb{R} \to \mathbb{R}$ is called probability density function, iff:

1. $f(x) \geq 0 \forall x \in \mathbb{R}$
2. f is riemann- (or Lebegue-) integrateble and

$$\int_{\mathbb{R}} f(x)dx = 1$$

For piecewise continuous pdfs we can define:

$$P((a, b]) = \int_a^b f(x)dx$$

▼ Discrete probability measures and pdfs

A probability measure on the measure space $(\mathbb{R}, \mathcal{B})$ is called **discrete iff:**

$$\exists A \subset \mathbb{R} | A \text{ countable} : P(A) = 1$$

Definition pdf: let $f : \mathbb{R} \to \mathbb{R}_+$ be a real and positve mapping. Then f is a pdf iff:

$$\int_{-\infty}^{\infty} f(x) = 1$$

▼ Discrete distributions

Binomial distribution:

$$P(\{i\}) = \binom{n}{i} \pi^i (1 - \pi)^i$$

It holds that:

$$X \sim B(n, p) \implies \mathbb{E}(X) = np, Var(X) = np(1 - p)$$

Geometric distribution:

$$P(\{i\}) = (1 - \pi)^{i-1} \pi$$

Poisson distribution:

$$P(\{i\}) = \frac{\lambda^i}{i!} \exp(-\lambda) = \lim_{n \to \infty} \binom{n}{i} \pi_n^i (1 - \pi_n)^{n-i} | \pi_n \to \lambda$$

It holds that:

$$X \sim Poi(\lambda) \implies \mathbb{E}(X) = Var(X) = \lambda$$

▼ Continuous distributions

Normal:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2} \frac{(x - \mu)^2}{\sigma^2}\right)$$

Uniform dist.:

$$f(x) = \frac{1}{b-a} 1_{[a,b]}(x)$$

It holds that:

$$X \sim U(a,b) \implies \mathbb{E}(X) = \frac{b-a}{2}, Var(X) = \frac{(b-a)^2}{12}$$

Exponential

$$f(x) = \lambda \exp(-\lambda x) 1_{\mathbb{R}_+}(x)$$

It holds that:

$$X \sim Exp(\lambda) \implies \mathbb{E}(X) = \frac{1}{\lambda}, Var(X) = \frac{1}{\lambda^2}$$

- Integration Theory
  - simple functions

$$s(\omega) = \sum_{i=1}^{n} \alpha_i 1_{A_i}(\omega)$$

Also:

$$\int_E s d\mu := \sum_{i=1}^{n} \alpha_i \mu(A_i \cap E)$$

  - Measure Integral

Let $(\Omega, \mathcal{A}, \mu)$ be a measuring space and $f : \Omega \to [0, \infty]$ a non-negative mapping

$$\int_E f d\mu = \sup \left\{ \int_E s d\mu : 0 \leq s \leq f | s \text{ simple} \right\}$$

  - measurable functions

Let $(\Omega, \mathcal{A}, \mu)$ a measuring space $f : \Omega \to [0, \infty]$ a non negative mapping. Then we call f measurable if:

$$f^{-1}(A) \in \mathcal{A} \quad \forall A \in \mathcal{B}(\mathbb{R})$$

Note, that measurability is dependent on $\mathcal{A}$. A function f is measurable for certain σ-algebras but not for all.

We can then say:

Any function $f : \Omega \to [0, \infty]$ is $(\mathcal{P}(\Omega) - \mathcal{B})$ measurable.

Classes of functions where we don’t need to prove measurability:

1. Indicator functions
2. monotone functions
3. cont. functions
4. functions with finitely many discontinuities

  - Lemma 1.27

Lemma 1.27 Let $(\Omega, \mathcal{A}, \mu)$ be a measure space, $f, g : \Omega \to [0, \infty]$ be $(\mathcal{A} - \overline{\mathcal{B}})$-measurable and $E \in \mathcal{A}$. Then, we have

(i) $f \leq g \implies \int_E f d\mu \leq \int_E g d\mu$

(ii) A, B $\in \mathcal{A}$ and $A \subseteq B \implies \int_A f d\mu \leq \int_B f d\mu$

(iii) $c \in [0, \infty] \implies \int_E cf d\mu = c \cdot \int_E f d\mu$

(iv) $f(\omega) = 0 \, \forall \omega \in \Omega \implies \int_E f d\mu = 0$

(v) $\int_E f d\mu = \int_\Omega f \mathbb{1}_E d\mu$

We approach proving properties of the measure-integral by firstly showing it for simple functions, then non-negative functions and lastly integrateable functions.

▼ extension of Integrals to non negative functions

Let setup as in Integral but this time $f : \Omega \to \mathbb{R}$ then we denote $f^- := \max(-f, 0)$ and $f^+ := \max(f, 0)$. Then we can define iff:

$$\int_E f^+ d\mu < \infty \wedge \int_E f^- d\mu < \infty$$

The **measure integral** for f as:

$$\int_E f d\mu = \int_E f^+ d\mu - \int_E f^- d\mu$$

▼ Integrateable functions

Let $(\Omega, \mathcal{A}, \mu)$ a measuring space and $f : \Omega \to \mathbb{R}$ a real mapping. Then we call f integrateable iff:

$$\int_\Omega |f| d\mu < \infty$$

▼ Theorem 1.29

Let f, g be integrateable functions. And $\alpha, \beta \in \mathbb{R}$, then

1. $\alpha f + \beta g$ is integratebale
2. $\int_E (\alpha f + \beta g) d\mu = \alpha \int_E f d\mu + \beta \int_E g d\mu$

▼ Random Variable

Let $(\Omega, \mathcal{A}, P)$ be a probability space, then we call $X : \Omega \to \mathbb{R}^k$ a random variable, iff the mapping is measurable.

▼ Probability distribution

Let $(\Omega, \mathcal{A}, P)$ be a probability space, and $X : \Omega \to \mathbb{R}^k$ a random variable. Then we define the distribution of X as:

$$P^X(B) := P(X^{-1}(B)) \qquad B \in \mathcal{B}(\mathbb{R}^k)$$

▼ Examples for measurable functions in $(\mathbb{R}, \mathcal{B})$

1. (real-valued) indicator functions $1_A(\omega)$

$$
\begin{array}{ll}
\underline{\text{Case 1}} & B \in \mathcal{B} \text{ and } B \cap \{0,1\} = \{0,1\} \\
& X^{-1}(B) = \mathbb{1}_A^{-1}(B) = \Omega = \mathbb{R} \in \mathcal{A} \\
\underline{\text{Case 2}} & B \in \mathcal{B} \text{ and } B \cap \{0,1\} = \emptyset \\
& X^{-1}(B) = \mathbb{1}_A^{-1}(B) = \emptyset \in \mathcal{A} \\
\underline{\text{Case 3}} & B \in \mathcal{B} \text{ and } B \cap \{0,1\} = \{0\} \\
& X^{-1}(B) = \mathbb{1}_A^{-1}(B) = A^{\complement} \in \mathcal{A} \\
\underline{\text{Case 4}} & B \in \mathcal{B} \text{ and } B \cap \{0,1\} = \{1\} \\
& X^{-1}(B) = \mathbb{1}_A^{-1}(B) = A \in \mathcal{A}\,.
\end{array}
$$

2. monotone functions
3. continuous functions
4. functions with only finitely many discontinuities:

$$\left|\left\{f : x \in \mathbb{R} | \lim_{y \downarrow x} f(y) \neq \lim_{y \uparrow x} f(y)\right\}\right| < \infty$$

- Creating measurable functions from sequences (**Theorem 1.33)**

  Let $f, f_n : \Omega \to [0, \infty]$ be measurable on $\mathcal{A}$, then:

  i) $\sup f_n$, $\inf f_n$, $\underline{\lim} f_n$, $\overline{\lim} f_n$ and (if it exists) $\lim_{n\to\infty} f_n$
  ii) $\max\{f_1, \dots, f_n\}$, $\min\{f_1, \dots, f_n\}$
  iii) $\alpha \cdot f_1 + \beta \cdot f_2$, for $\alpha, \beta \in \mathbb{R}$, $f_1 \cdot f_2$
  iv) $f^+ := \max\{f, 0\}$, $f^- := \max\{-f, 0\}$, $|f| = f^+ + f^-$

  are also measurable. This directly translates to random variables.

- stochastic independence for probability function

  Let $(\Omega, \mathcal{A}, P)$ be a probability space. Then $A, B \subset \Omega$ are called stochstically independent if:

  $$P(A \cap B) = P(A)P(B)$$

- stochastic independence for random variables

  Let $(\Omega, \mathcal{A}, P)$ be a probability space and $X_1, ..., X_l : \Omega \to [0, \infty]$ be random variables. Furthermore let $A_1, ..., A_l \subset \Omega$ then $X_1, ..., X_l$ are called **stochastically independent iff**:

  $$P(X_1 \in A_1, ..., X_l \in A_l) = P(X_1^{-1}(A_1)) * \cdots * P(X_l^{-1}(A_l))$$

  You can also generalize this to systems of sets:

  And, more generally, let $(\mathcal{A}_i)_{i \in I}$ be a family of systems of sets. Then, the systems of sets are called **stochastically independent**, if for *any* finite, non-empty index sets $I_0 \subseteq I$ and *any* $A_i \in \mathcal{A}_i$, $i \in I_0$, we have

  $$P\left(\bigcap_{i \in I_0} A_i\right) = \prod_{i \in I_0} P(A_i).$$

  Note that we generate a sequence of sets by taking a set from every $\mathcal{A}_i$.

- Transformation under independence

  Any transformation $G : \mathbb{R}^n \to \mathbb{R}^m$ preserves independence

- cdfs for random variables

  Now we can define:

$$
\begin{gathered}
F : \mathbb{R}^k \to [0,1] \\
F(x) = P(X \leq x) \forall x \in \mathbb{R}^k
\end{gathered}
$$

- marginal densities

$$f_x(x) = \int_{-\infty}^{\infty} f_{x,y}(x,y)dy$$

- density transformation

  Any invertible real function $\phi : \mathbb{R} \to \mathbb{R}$ can be used to find out the density of $Y = \phi(X)$ if X is a r.v.:

$$f_y(y) = f_x(\phi^{-1}(y))|\phi^{-1}{'}(y)|1_{\phi(\mathbb{R})}(y)$$

- expected value for discrete random variables

  Let $X : \Omega \to \mathbb{R}$ with $X(\mathbb{R})$ countable. Then:

$$\mathbb{E}(X) = \sum_{i \in X(\mathbb{R})} iP(X^{-1}(\{i\}))$$

- expected value for continuous random variables

  The expectation of X is just the measure integral over the sample space:

$$\mathbb{E}(X) = \int_{\Omega} XdP$$

  Note, that we use a special sequence of simple functions X_n to estimate this measure integral:

$$X_n(\omega) := \frac{k}{n} \Leftrightarrow \frac{k}{n} \leq X(\omega) \leq \frac{k+1}{n}$$

  The we can rewrite X_n as:

$$X_n = \sum_{k=1}^{\infty} \frac{k}{n} 1_{X \in [k/n,(k+1)/n)}$$

  And then show that:

$$\mathbb{E}(X_n) \to \mathbb{E}(X)$$

  Using pdfs we can get $\mathbb{E}(X)$ by:

$$
\begin{gathered}
\mathbb{E}(X) = \int_{\mathbb{R}} xf(x)dx \\
\mathbb{E}(g(X)) = \int_{\mathbb{R}} g(x)f(x)dx
\end{gathered}
$$

- Theorem 1.41 expectation after transformation

  Let X be a real r.v. with piecewise continuous density f_x. Then any real function g yields the property:

$$\mathbb{E}(g(X)) = \int g(x)f_x(x)dx$$

- Theorem 1.42 (properties of expectation)

  1. $|\mathbb{E}[X_1]| \leq \sup_{\omega} |X_1(\omega)|$

- absolute expectation never exceeds the supremum

2. $\mathbb{E}[\alpha_1 X_1 + \alpha_2 X_2] = \alpha_1 \mathbb{E}(X_1) + \alpha_2 \mathbb{E}(X_2)$
   - linearity
3. $X_1 \leq X_2 \implies \mathbb{E}[X_1] \leq \mathbb{E}[X_2]$
4. Independence implies $\mathbb{E}[X_1 X_2] = \mathbb{E}[X_1]\mathbb{E}[X_2]$
   - Proof in lecture with Fubini and productspaces

▼ s-moments

Let X be a real r.v. then the **sth-moment** is defined by $\mathbb{E}(X^s)$

Furthermore the **sth-central-moment** is defined by $\mathbb{E}((X - \mathbb{E}(X))^s)$

Furthermore the **sth-absolute-moment** is defined by $\mathbb{E}(|X|^s)$

And the

▼ Definition of covariance

Let X, Y be r.v. with finite second moments. We call Cov of X, Y:

$$Cov(X,Y) = \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))] = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

▼ Theorem 1.46 (Additivity of Variance)

Let $X = (X_1, ..., X_n)^\intercal$ where X_i are real r.v. and $\beta = (\beta_1, ..., \beta_n)$. Then we can denote:

$$Var(\beta X) = \beta^\intercal \Sigma \beta$$

While Σ is the covariance matrix with $\Sigma := (\sigma_{i,j})_{\{1,..,n\}\times\{1,..,n\}} = Cov(X_i, X_j)$

Furthermore we call $\mathbb{E}(X)$ the **expectation vector**

We can also compute Σ by:

$$\Sigma = \mathbb{E}[(X - \mathbb{E}(X))(X - \mathbb{E}(X))^\intercal]$$

Some handy properties, that arise from this:

Theorem 1.46 If $X_1, X_2, \ldots, X_n$ are independent with finite variance, then

$$Var(X_1 + \cdots + X_n) = Var(X_1) + \cdots + Var(X_n). \tag{10}$$

Proposition 1.48 For X, Y, Z random variables with finite second moments and $a, b \in \mathbb{R}$, the following properties hold true:

- $Var(X) = 0 \;\Leftrightarrow\; P(X = E[X]) = 1$,
- $Var(aX) = a^2\, Var(X)$,
- $Cov(aX + b, Y) = a\, Cov(X, Y)$,
- $Cov(X + Y, Z) = Cov(X, Z) + Cov(Y, Z)$,
- $|Cov(X, Y)| \leq \sqrt{Var(X)\, Var(Y)}$ (Cauchy-Schwarz inequality). This implies in (9) that $-1 \leq \rho(X, Y) \leq 1$.

▼ Quantiles

Let X be a real r.v. then we call a number $q_\alpha \in \mathbb{R}$ the α-quantile iff:

$$P(X \leq q_\alpha) \geq \alpha \wedge P(X \geq q_\alpha) \geq 1-\alpha$$

We can call the **kth-α-quantile q_α^k if:**

$$P(X \leq q_\alpha^k) \geq k\alpha \wedge P(X \geq q_\alpha^k) \geq 1-k\alpha$$

▼ Proposition 1.51 (optimization problem)

Let X be a real r.v. with $\mathbb{E}(X^2) < \infty$ then:

$$\arg\min_x \mathbb{E}[|X-x|] = q_{0.5} \tag{1}$$
$$\arg\min_x \mathbb{E}[|X-x|^2] = \mathbb{E}(X) \tag{2}$$

▼ Asymptotic Theory

▼ convergence in probability

$$X_n \xrightarrow{P} X \iff \forall\varepsilon > 0 : P(|X_n - X| > \varepsilon) \underset{n\to\infty}{\longrightarrow} 0$$

▼ Pointwise convergence

Let $X, X_1, ..., X_n$ be real valued r.v. then we say the sequence X_i converges pointwise towards X if:

$$\lim_{n\to\infty} X_n(\omega) = X(\omega) \quad \forall\omega \in \Omega$$

Additionally when $0 \leq X_n \leq X_{n+1}$ we can say:

$$\lim_{n\to\infty} \mathbb{E}(X_n) = \mathbb{E}(X)$$

Note that we have to have monotonicity to say this. **Pointwise conversion is not sufficient for convergence of expectations.**

▼ almost sure convergence

Is pointwise convergence for at least a subset $A \subset \Omega$ for which $P(A) = 1$.

▼ convergence in the p-th mean

Let X_n be a sequence of real r.v., then $X_n \xrightarrow{L_p} X$ iff:

$$\mathbb{E}[||X_n - X||^p] \to 0$$

▼ convergence in distrubtion

Let X_n be a sequence of r.v. and X be a r.v. on $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$. Then X_n converges in distribution iff:

$$E(f(X_n)) \to E(f(X)) \quad \forall f \in C_b(\mathbb{R}^k)$$

Whereas $C_b(\mathbb{R}^k)$ is the set of all continuous and bounded functions.

You can also write:

$$X_n \xrightarrow{\mathcal{L}} X \quad X_n \xrightarrow{d} X$$

You can also prove convergence in distribution by showing that:

$$F_n(x) \to F(x)$$

Or that the characteristic funcitons are equal:

$$\varphi_{X_n}(x) \to \varphi_X(x) \quad \forall x \in \mathbb{R}$$

▼ Lipschitz functions

Let $f : \mathbb{R} \to \mathbb{R}$. Then f is called a Lipschitz function iff:

$$|f(x) - f(y)| \leq L||x - y|| \quad \forall x, y \in \mathbb{R}, L \in \mathbb{R}$$

▼ Markov's inequality

This is a more general inequality, than Chebychef:

$$P(||X|| \geq \varepsilon) \leq \frac{\mathbb{E}(g(X))}{g(\varepsilon)}$$

While $g : \mathbb{R}_+ \to \mathbb{R}_+$ is a monotone increasing function. With $g(0) \neq 0$

▼ Convergence in expectation

A popular use case of asymptotic theory is analysing the distribution of an estimator. A part of this analysis is checking the expected value of the distribution. Often times we need the following statement to make life easier in this usecase:

$$\mathbb{E}(X_n) \underset{n\to\infty}{\longrightarrow} \mathbb{E}(X) \Leftrightarrow \int_\Omega \lim_{n\to\infty} X_n dP = \lim_{n\to\infty} \int_\Omega X_n dP$$

We learnt, pointwise conversion ($X_n(\omega) \to X(\omega) \forall \omega \in \Omega$) is not sufficient for this statement.

For demonstration, Prof. Jensch provided a simple example:

Let $(\mathbb{N}, \mathcal{P}(\mathbb{N}), \nu)$ be a measuring space and $\nu(A) = |A|$ the counting measure. Furhtermore let:

$$X_n(\omega) = \sum_{i=1}^{\infty} \frac{1}{n} 1_{\{1,\ldots,n\}}(\omega)$$

Then we can say $X_n \to X$ with $X(\omega) = 0 \forall \omega$ so:

$$\lim_{n\to\infty} \int_\Omega X_n d\nu = \lim_{n\to\infty} \sum_{i=1}^{n} \frac{1}{n} = \lim_{n\to\infty} \frac{n}{n} = 1 \neq 0 = \int_\Omega X d\nu = \int_\Omega \lim_{n\to\infty} X_n d\nu$$

So:

$$\lim_{n\to\infty} \int_\Omega X_n d\nu \neq \int_\Omega \lim_{n\to\infty} X_n d\nu$$

Of course this begs the question: what additional assumptions do we need?

In the lecture, we have gotten to know two additional assumptions, that each yield conversion in expectation:

1. monotone convergence

   $X_n(\omega) \leq X_{n+1}(\omega) \forall \omega$ i.e. the sequence grows monotonically

2. dominated convergence

   There is a function $Y : \Omega \to \mathbb{R}$ such that $\forall \omega \in \Omega : Y(\omega) \geq X_n(\omega) \forall n$ and $\mathbb{E}(Y) < \infty$

We later found out that we can losen the statements up a little bit, by not putting these constraints on all $\omega \in \Omega$ but on a subset $A \subset \Omega$ such that $P(A) = 1$ (see slide 80 / 82 for more detail). In that case we can speak about the assumptions holding **almost surely**.

Own example: We can look a at a lebesgue measureable functions $X_n : [0, 1] \to \mathbb{R}$ if there is a countable infinite (at max) set $A \subset \mathbb{R}$ for which it holds that:$X_n(x) \to X(x) \forall x \in \mathbb{R} \setminus A$ and there is another countable infinite set (at max) $B \subset \mathbb{R}$ for which it holds that $\exists Y : Y(x) \geq X_n(x) \forall x \in \mathbb{R} \setminus B$ and $\mathbb{E}(Y) < \infty$ then we can say:

$$\mathbb{E}(X_n) \to \mathbb{E}(X)$$

▼ Theorem 2.11

It holds that:

$$X_n \xrightarrow{a.s.} X \implies X_n \xrightarrow{P} X$$

▼ Theorem 2.14

It holds that:

$$X_n \xrightarrow{P} a \Longleftrightarrow X_n \xrightarrow{d} a$$

▼ Continuous mapping theorem

Let $X_n \to X$ and $g : \mathbb{R}^k \to \mathbb{R}^l$ cont. then we get:

$$X_n \longrightarrow X \implies g(X_n) \longrightarrow g(X)$$

This holds for all modes of convergence except L_i.

▼ Weak law of large numbers 1&2

Let $(X_n)_n$ be a sequence of r.v. uncorrelated with $\mathbb{E}[X_1] = \mathbb{E}[X_2] = ... = \mu \in \mathbb{R}$ and finte second central moments. Then:

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \mu$$

We can derive the same statement by choosing $X_n \overset{iid}{\sim} F$ and just assuming $\mathbb{E}(X_i) < \infty \forall i$ (which then is the law of large numbers 2).

▼ Fatou's Lemma

Let X_n be a real sequence of r.v. on the probability space $(\Omega, \mathcal{A}, P)$. Then it holds that:

$$\int \underline{\lim}_{n\to\infty} X_n dP \leq \underline{\lim}_{n\to\infty} \int X_n dP$$

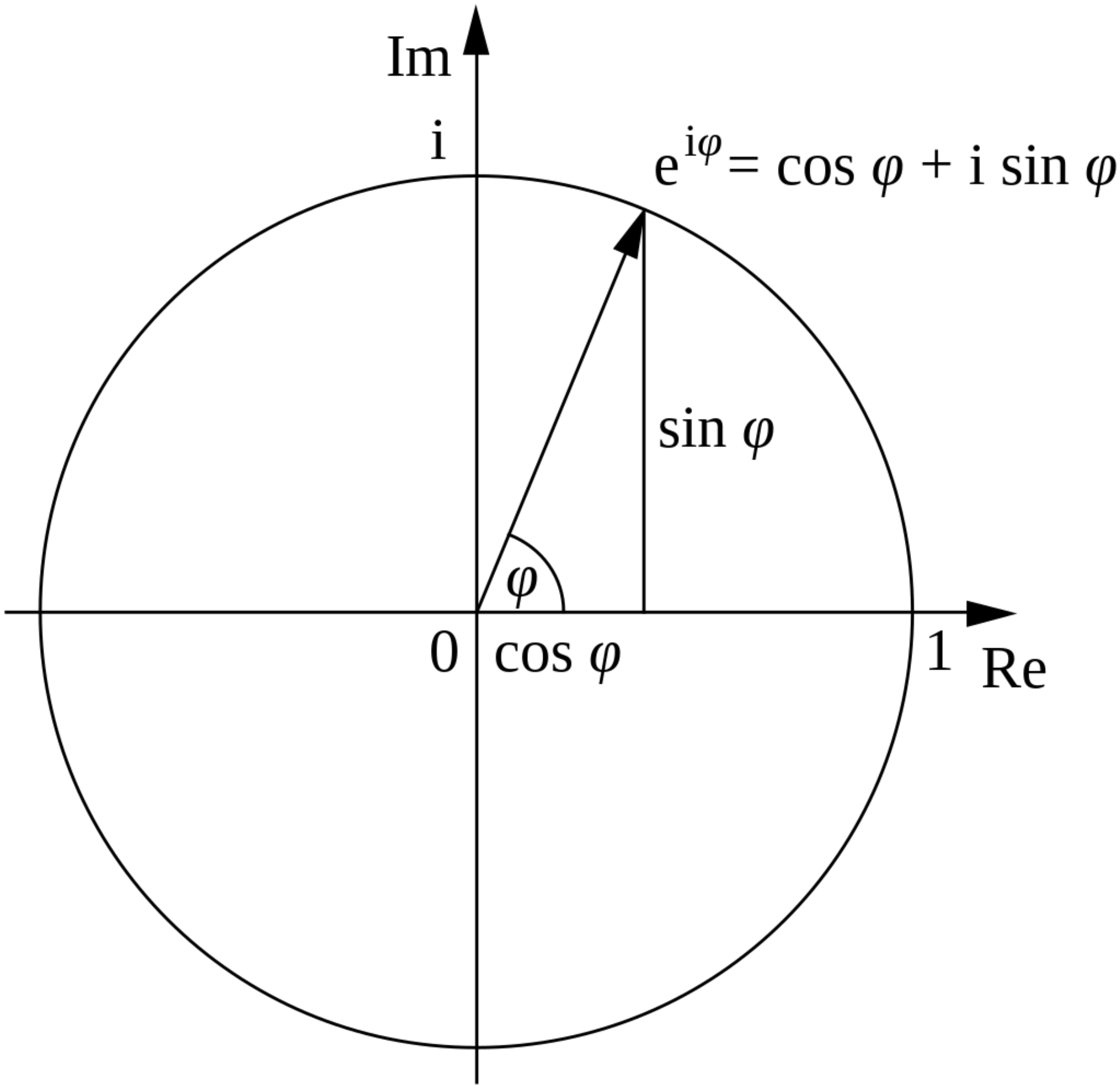
This implies:

$$\underline{\lim}_{n\to\infty} X_n = X \implies \mathbb{E}(X) \leq \underline{\lim}_{n\to\infty} \mathbb{E}(X)$$

▼ characteristic functions

Let a random vector $X : \Omega \to \mathbb{R}^k$. Then the characteristic function $\varphi_X : \mathbb{R}^k \to \mathbb{C}$ is defined as:

$$\varphi_X(t) = \mathbb{E}[e^{it'X}] = \mathbb{E}(cos(t'X)) + i\mathbb{E}(sin(it'X))$$



For X discrete:

$$\varphi_X(t) = \sum_{i=1}^{\infty} e^{it'x_i} P(X = X_i)$$

For X continuous:

$$\varphi_X(t) = \int e^{it'x} f_X(x) dx$$

▼ example

$X \sim Poi(\lambda), \lambda > 0$ that means:

$$p_X(k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

Therefore X is discrete and:

$$\varphi_X(t) = \sum_{i=1}^{\infty} e^{it'k} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{i=1}^{\infty} \frac{(\lambda e^{it'})^k}{k!} = \exp(\lambda(\exp(it) - 1))$$

▼ Why do all this?

Even tho convergence in distribution is the weakest form of convergence, we need this form for the **central limit theorem!**

We can say:

$$X_n \xrightarrow{d} X \iff \varphi_{X_n} \to \varphi_X$$

▼ Levy's cont. theorem

Let X_n, X be r.v. on prob. space $\{\Omega, \mathcal{A}, P\}$. Then it holds that:

$$X_n \xrightarrow{d} X \iff \varphi_{X_n} \underset{n\to\infty}{\longrightarrow} \varphi_X \forall t \in \mathbb{R}^k$$

▼ Slutsky's Lemma

We can avoid showing convergene in distribution for X_n iff:

$$||X_n - Z_n|| \xrightarrow{P} 0, Z_n \xrightarrow{d} X \implies X_n \xrightarrow{d} X$$

▼ Stochastic boundedness

Let $(X_n)_n$ be a sequence of real r.v. then $(X_n)_n$ is stochastically bounded iff:

$$\forall \varepsilon > 0 \exists! n_0 \in \mathbb{N}, C \in \mathbb{R} : \quad P(||X_n|| \leq C) \;\; \forall n \geq n_0$$

▼ stochastic Landau symbols

We say $X = O_p(\tilde{C})$ iff $X/\tilde{C}$ is stochastically bounded

We say $X_n = o_p(C_n)$ iff $X_n/C_n \xrightarrow{p} 0$

Special cases

$\tilde{C} = 1$ then we say: $X = O_p(1) \iff X$ is stochastically bounded

$C_n = 1$ then it holds that: $X_n = o_p(1) \iff X_n \xrightarrow{p} 0$

▼ Properties of stochastic Landau symbols

1. $X_n \xrightarrow{d} X \implies X_n = O_p(1)$
2. $X_n = C + o_p(1) \implies X_n = O_p(1)$
    - If a sequence is equal to a constant and a random part, that converges in probability to 0, then every element of the sequence is stochastically bounded
3. The following operations are possible:
    a. $o_p(1) + o_p(1) = o_p(1)$
    b. $O_p(1) + O_p(1) = O_p(1)$
    c. $O_p(1)O_p(1) = O_p(1)$
    d. $O_p(1)o_p(1) = o_p(1)$
4. $g : \mathbb{R}^k \to \mathbb{R}^l$ continuous at x_0. Then:

$$X_n = x_0 + o_p(1) \implies g(X_n) = g(x_0) + o_p(1)$$

▼ Strong LLN

For $X_1, X_2, ...$ iid with $\mathbb{E}(\mu_i) = \mu < \infty \forall i$, it holds that:

$$\bar{X}_n \xrightarrow{a.s.} \mu$$

- CLT of Lindeberg-Levy

  Let $(X_n)_{n\in\mathbb{N}}$ iid real r.v. with finite mean and variance and $Var(X_i) = \sigma^2 \in (0, \infty)$. Then:

$$\frac{\sum_{i=1}^n X_i - \mu}{\sqrt{n}\sigma} \xrightarrow{d} \mathcal{N}(0,1)$$

- Lyapunov CLT

  Let $X_1, X_2, ...$ be independent real valued r.v. with $\mathbb{E}(X_t) = \mu_t < \infty, Var(X_t) = \sigma_t^2 < \infty, m_{3,t} = \mathbb{E}(|X_t - \mu_t|^3) < \infty$. Then:

$$\frac{\sum_{i=1}^n (X_i - \mu_i)}{\sqrt{\sum_{i=1}^n \sigma_i^2}} \xrightarrow{d} \mathcal{N}(0,1)$$

- Multivariate CLT

  Let $X_1, X_2, ...$ iid $\mathbb{R}^k$ valued r.v. with $\mathbb{E}(X_i) = \mu \in \mathbb{R}^k \forall i$. And finite covariance matrix Σ. Then:

$$\frac{1}{\sqrt{\sigma}}\left(\sum_{i=1}^n X_i - n\mu\right) \xrightarrow{d} \mathcal{N}(0_k, \Sigma)$$

- Delta method

  Let $X_1, X_2, ...$ a sequence of $\mathbb{R}^k$ valued r.v. and $c \in \mathbb{R}^k$ with:

$$\sqrt{n}(X_n - c) \xrightarrow{d} \mathcal{N}(0, \Sigma)$$

  Furthermore let $\phi : U \subset \mathbb{R}^k \to \mathbb{R}^m$ be differentiable at c then:

$$\sqrt{n}(\phi(X_n) - \phi(c)) \xrightarrow{d} \mathcal{N}(0, \phi'(c)\Sigma(\phi'(c))')$$

- Berry Esseen

  Speed of convergence for CLT. So let $X_1, X_2, ...$ be an iid sequence of r.v. and:

$$Z_n = \frac{1}{\sqrt{n}\sigma}\sum_{i=1}^n (X_i - \mu)$$

  Then we want to know something about:

$$\sup_x |F_{Z_n}(x) - \phi(x)|$$

  Berry Esseen says, that if a finite third central moment exists we can say:

$$\sup_x |F_{Z_n}(x) - \phi(x)| \leq c\frac{\mathbb{E}(|X_1 - \mu|^3)}{\sigma^3\sqrt{n}}$$

  With some constant $0.4097 \leq c \leq 0.7655$.

- Conditional Expectation
  - Definition of conditional prob. measure

    Let $(\Omega, \mathcal{A}, P)$ and $A \in \mathcal{A}$ with $P(A) > 0$. Then $P(\cdot|A)$ exists and is defined by:

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \quad \forall B \in \mathcal{A}$$

  - Regression and definition of conditional expectation

    Consider real r.v. X and $\mathbb{R}^k$-valued r.v. with $\mathbb{E}(Y^2) < \infty$. We want to find a measurable function $g : \mathbb{R}^k \to \mathbb{R}$ with the property, that it minimizes:

$$\mathbb{E}((Y - g(X))^2)$$

It turns out, that **the random variable** $g(X)$ is actually the conditional expectation:

$$g(X) = \mathbb{E}(Y|X)$$

In notation we call:

$$g(x) = \mathbb{E}(Y|X = x)$$

▼ Conditional probability and distribution

Let $X : \Omega \to \mathbb{R}^k$ and $Y : \Omega \to \mathbb{R}^l$ be to random variables. Then for any $B \in \mathcal{B}(\mathbb{R}^l)$ we call the conditional Probability of B given X:

$$P^{Y|X}(B) = P(Y \in B|X) = \mathbb{E}(1_B(Y)|X)$$

Also if we assume a concrete outcome of X then we call the resulting measure a conditional distribution:

$$P^{Y|X=x}(B) = P(Y \in B|X = x) = \mathbb{E}(1_B(Y)|X = x)$$

▼ Conditional expectation discrete case

Let X be r.v. on prob space $(\Omega, \mathcal{A}, P)$ and $A \in \mathcal{A}$ with $P(A) > 0$. Then we get:

$$\mathbb{E}(X|A) = \int_\Omega X dP(\cdot|A)$$

▼ Theorem of Fubini

Let X_1 and X_2 be r.v. with different prob. spaces. Then we get that, the product space of the two prob. spaces is defined by:

$$\{\Omega_1 \times \Omega_2, \mathcal{A}_1 \times \mathcal{A}_2, P_1 \cdot P_2\}$$

Whereas $P_1 \cdot P_2 : \Omega_1 \times \Omega_2 \to \mathbb{R} \times \mathbb{R}$ is defined by:$(\omega_1, \omega_2) \to P_1(\omega_1) \cdot P_2(\omega_2)$

Then it holds that:

$$\begin{aligned} \int_{\Omega_1 \times \Omega_2} X_1 X_2 dP_1 \cdot P_2 &= \int_{\Omega_1} \int_{\Omega_2} X_1 X_2 dP_1 dP_2 \\ &= \int_{\Omega_1} \int_{\Omega_2} X_1 dP_1 X_2 dP_2 \\ &= \int_{\Omega_2} \int_{\Omega_1} X_2 dP_2 X_1 dP_2 \end{aligned}$$

We obtain the conditional expectation by an optimization problem. If we look at:

$$\mathbb{E}((Y - g(X))^2) = \sum_y (y - g(x))^2 P(X = x, Y = y)$$

And then multiply with $1/P(X = x)$ (because the optimization problem does not change with scaling), we get:

$$\mathbb{E}((Y - g(X))^2) = \sum_y (y - g(x))^2 \frac{P(X = x, Y = y)}{P(X = x)}$$

This yields the result that:

$$g(x) = \sum_y y \frac{P(X=x, Y=y)}{P(X=x)}$$

▼ conditional expectation for continuous case

We now consider X, Y r.v. with $f_{X,Y}(x,y)$ as a pdf. Then we get the conditional expectation by minimizing:

$$\min_g \mathbb{E}((Y-g(X))^2) \Leftrightarrow \min_g \int_{\mathbb{R}} (y-g(x))^2 f_{X,Y}(x,y)dy$$

FOC yields:

$$\begin{aligned} \int_{\mathbb{R}} (y-g(x)) f_{X,Y}(x,y)dy &\overset{!}{=} 0 \\ \int_{\mathbb{R}} y f_{X,Y}(x,y)dy - g(x)\int_{\mathbb{R}} f_{X,Y}(x,y)dy &= 0 \\ \int_{\mathbb{R}} y f_{X,Y}(x,y)dy - g(x) f_X(x) &= 0 \\ \int_{\mathbb{R}} y \frac{f_{X,Y}(x,y)}{f_X(x)}dy &= g(x) \end{aligned}$$

So this means that the conditional expectation is:

$$\mathbb{E}(Y|X) = g(x) = \int_{\mathbb{R}} y \frac{f_{X,Y}(x,y)}{f_X(x)}dy \quad \text{a.s.}$$

▼ conditional pdfs under continuous case

Let X, Y be random variables continuous with $X, Y \sim F_{X,Y}$. Then we say:

$$f_{X|Y}(y|x) = \begin{cases} \frac{f_{X,Y}(x,y)}{f_X(x)} \text{ if } f_X(x) > 0 \\ \text{any density else} \end{cases}$$

Also we then know that:

$$\mathbb{E}(Y|X=x) = \int_{\mathbb{R}} y f_{Y|X}(y|x)dy$$

▼ Alternative definitions of conditional expectation

Let $h : \mathbb{R}^k \to [0,1]$ be $\mathcal{B}(\mathbb{R}^k) - \mathcal{B}([0,1])$-measureable. $g(X)$ is the conditional expectation iff:

$$\mathbb{E}((Y-g(X))h(X)) = 0$$

This condition can also be written as:

$$\mathbb{E}(Y1(X \in B)] = \mathbb{E}(g(X)(X \in B)] \quad \forall B \in \mathcal{B}(\mathbb{R}^k)$$

▼ Uniqueness of conditional expectation

It holds that two minimizers g_1 and g_2 are equal almost surely:

$$g_1(X) = g_2(X) \text{ a.s.} \quad \Leftrightarrow \quad P(g_1(X) = g_2(X)) = 1$$

▼ Theorem 3.9 (iterated expectation)

Let $Y : \Omega \to \mathbb{R}^k$ and $Z : \Omega \to \mathbb{R}^l$ each r.v. with prob. space $\{\Omega, \mathcal{A}, P\}$. Then the following holds:

$$\begin{align}
\mathbb{E}(\mathbb{E}(Y|X)) &= \mathbb{E}(Y) \tag{3}\\
\mathbb{E}(\mathbb{E}(Y|X,Z)|Z) &= \mathbb{E}(Y|Z) \tag{4}\\
\mathbb{E}(\mathbb{E}(Y|X)|X,Z) &= \mathbb{E}(Y|X) \tag{5}\\
\mathbb{E}(f(X)Y|X) &= f(X)\mathbb{E}(Y|X) \tag{6}\\
&\quad \text{with } \mathbb{E}(f^2(X)) + \mathbb{E}(Y|f(X)) < \infty \tag{7}\\
\mathbb{E}(\mathbb{E}(Y|X)|f(X)) &= \mathbb{E}(Y|f(X)) \tag{8}
\end{align}$$

2- 6 hold only almost surely. Meaning the set for which these equalities hold, is measured at 1.

▼ Theorem 3.10 (properties of conditional expectation)

1. Linearity

$$\mathbb{E}(a_1Y_1 + a_2Y_2|X) = a_1\mathbb{E}(Y_1|X) + a_2\mathbb{E}(Y_2|X)$$

2. monotonicity

$$Y_1 \leq Y_2 \implies \mathbb{E}(Y_1|X) \leq \mathbb{E}(Y_2|X)$$

3. Cauchy schwarz

$$(\mathbb{E}(XY|Z))^2 \leq \mathbb{E}(X^2|Z)\mathbb{E}(Y^2|Z)$$

4. Jensen meckling

   Let $\rho : \mathbb{R} \to \mathbb{R}$ convex then:

$$\rho(\mathbb{E}(Y|X)) \leq \mathbb{E}(\rho(Y)|X)$$

5. Approximation of integrals

$$0 \leq Y_n \uparrow Y \implies \mathbb{E}(Y_n|X) \uparrow \mathbb{E}(Y|X)$$

6. Relation of probability and expectation

$$P(|Y| \geq \epsilon|X) \leq \frac{\mathbb{E}(Y^2|X)}{\epsilon^2}$$

▼ independence and image measures / expectations

Let $Y : \mathcal{A} \to \mathbb{R}$ with Y^2 integrateable and $X : \mathcal{A} \to \mathbb{R}^k$. If Y and X are independent, then:

$$\begin{align}
P^{Y|X}(B) &= P^Y(B) \quad \forall B \in \mathcal{B}(\mathbb{R}) \tag{9}\\
\mathbb{E}(Y|X) &= \mathbb{E}(Y) \text{ a.s.} \tag{10}
\end{align}$$

▼ Law of total probability

Given a partition of the sample space B_i with $\biguplus_{i=1}^{\infty} B_i = \Omega$. It holds for all events in $\mathcal{A}$ that:

$$P(A) = \sum_{i=1}^{\infty} P(A|B_i)P(B_i)$$

▼ Conditional variance

Let $Y : \mathcal{A} \to \mathbb{R}$ borel-measurable (real r.v.) with finite fourth moment. Let $X : \mathcal{A} \to \mathbb{R}^k$ be borel-measurable. Then we define the conditional variance of $Y|X$ as:

$$Var(Y|X) = \mathbb{E}((Y - \mathbb{E}(Y|X))^2|X)$$

Note that this is still a random variable!!!!

▼ Basic properties of the conditional variance

Let $a, b : \mathbb{R}^k \to \mathbb{R}$ borel-measurable, then we have that:

$$Var(a(X)Y + b(X)|X) = a^2(X)Var(Y|X)$$

And given $c \in \mathbb{R}$ a constant:

$$Var(c|X) = 0$$

If X and Y are independent we get:

$$Var(Y|X) = Var(Y)$$

- Advanced statements about the conditional variance

  We can decompose the variance into its mean value and the conditional variance of the variance of the conditional expected value:

$$Var(Y) = \mathbb{E}(Var(Y|X)) + Var(\mathbb{E}(Y|X))$$

  Furthermore we get the result, that the expected value decreases, the more r.v. are conditioned on it:

$$\mathbb{E}(Var(Y|X)) \geq \mathbb{E}(Var(Y|X, Z))$$

- Extending to multivariate case

  If $Y : \mathcal{A} \to \mathbb{R}^k$

- Proofs in lecture
  - Theorem 1.42 iv

    This can be done directly but we use a more general construction of so-called productspaces and an application of Fubinis-theorem

    1. **Productspaces (only for prob. measures and spaces)**

       Let $(\Omega_1, \mathcal{A}_1, P_1)$ and $(\Omega_2, \mathcal{A}_2, P_2)$ be two prob.-spaces. Then their product space is defined by: $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2, P_1 \otimes P_2)$, where:

       $\Omega_1 \times \Omega_2 := \{(\omega_1, \omega_2 | \omega_1 \in \Omega_1, \omega_2 \in \Omega_2\}$

       $\mathcal{A}_1 \otimes \mathcal{A}_2 := \sigma(\{A_1 \times A_2 | A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2\})$

       $P_1 \otimes P_2(A_1 \times A_2) := P_1(A_1)P_2(A_2)$ product probability measure of P_1 and P_2

       Suppose we have two $\mathbb{R}$ valued random variables X_1 on $(\Omega_1, \mathcal{A}_1, P_1)$ and X_2 on $(\Omega_2, \mathcal{A}_2, P_2)$. Then there exists a **joint** probspace $(\Omega, \mathcal{A}, P) := (\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2, P_1 \otimes P_2)$

       where X_1 and X_2 are random variables and stochstically independent.

    2. **Theorem Fubini**

       let $X : \Omega_1 \times \Omega_2 \to \bar{\mathbb{R}}$ be a r.v. on the product prob space $(\Omega, \mathcal{A}, P) := (\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2, P_1 \otimes P_2)$. Then the function:

$$\omega_1 \to \int_{\Omega_2} X_{\omega_1} dP_2$$

$$\omega_2 \to \int_{\Omega_1} X_{\omega_2} dP_1$$

       Where $X_{\omega_1}(\omega_2) := X(\omega_1, \omega_2)$ with ω_1 fixed and vice versa. P_1 and P_2 are almost everywhere defined.

       Then

$$\int_{\Omega} X dP = \int_{\Omega_1 \times \Omega_2} X dP_1 \otimes P_2 = \int_{\Omega_1} \left[\int_{\Omega_2} X_{\omega_1} dP_2 \right] dP_1 = \int_{\Omega_2} \left[\int_{\Omega_1} X_{\omega_2} dP_1 \right] dP_2$$

Order of integration does not matter

3. **Proof of iv**

   Let X_1 and X_2 be stochastically independent on $(\Omega, \mathcal{A}, P)$. That is we have X_1 on $(\Omega_1, \mathcal{A}_1, P_1)$ and X_2 on $(\Omega_2, \mathcal{A}_2, P_2)$ s.t. (X_1, X_2) is a r.v. on $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2, P_1 \otimes P_2)$.

   Then we have $(X_1, X_2) : \Omega_1 \times \Omega_2 \to \mathbb{R}^2$ and $X_1 \cdot X_2 : \Omega_1 \times \Omega_2 \to \mathbb{R}$. Then we have

   $$\mathbb{E}(X_1 \cdot X_2) = \int_\Omega X_1 X_2 dP = \int_{\Omega_1 \times \Omega_2} X_1 X_2 dP_1 \otimes P_2 = \int_\Omega X_1(\omega_1) X_2(\omega_2) dP_1 \otimes P_2(\omega_1, \omega_2) \underset{s.F.}{=} \int_{\Omega_2} \Big[\int_\Omega$$

▼ Theorem 1.46

s.t. $X_1, ..., X_n$ independent $\implies Var(X_1 + ... + X_n) = Var(X_1) + ... + Var(X_n)$

$$Var(\sum_{i=1}^n X_i) = \mathbb{E}((\sum_{i=1}^n (X_i - \mathbb{E}(\sum_{i=1}^n X_i))^2) = \mathbb{E}((\sum_{i=1}^n (X_i - \mathbb{E}(X_i))^2) = ... = \sum_{i=1}^n \sum_{j=1}^n Cov(X_i, X_j) = \sum_{i=1}^n Co$$

▼ complex numbers and their absolute value

Shot that:

$$\lim_{u \to \infty} |\exp(u(it - \lambda))| \to 0$$

The proof relies on $|\exp(ix)| = 1 \forall x \in \mathbb{R}$. The expression above can be written as:

$$\lim_{u \to \infty} \left| \frac{\exp(uit)}{\exp(u\lambda)} \right| = 0$$