Probability and Statistics

Nicola Studer
nicstuder@student.ethz.ch

■ General ■ Discrete ■ Continuous

1 Mathematical framework

Probability Space $(\Omega, \mathcal{F}, \mathbb{P})$

- $\boldsymbol{\Omega \neq \emptyset}$: **Sample space** with $\omega \in \Omega$ as outcome.
- $\boldsymbol{\mathcal{F}} \subseteq \mathcal{P}(\Omega)$: $\boldsymbol{\sigma}$**-algebra**
  1. $\Omega \in \mathcal{F}$
  2. $A \in \mathcal{F} \implies A^{\complement} \in \mathcal{F}$ where A is an event.
  3. $A_1, A_2, \ldots \in \mathcal{F} \implies \bigcup_{i=1}^{\infty} \in \mathcal{F}$
- $\mathbb{P}$: **Probability measure** on $(\Omega, \mathcal{F})$

  $\mathbb{P} : \mathcal{F} \to [0,1], A \mapsto \mathbb{P}(A)$
  1. $\mathbb{P}(\Omega) = \sum_{\omega \in \Omega} p(\omega) = 1$ if $p(\omega) := \mathbb{P}(\{\omega\})$.
  2. $P(A) = \sum_{i=1}^{\infty} P(A_i)$ if $A = \bigcup_{i=1}^{\infty} A_i$ disjoint.

We have some further consequences of this definition:

1. $\emptyset \in \mathcal{F}$
2. $A_1, \ldots \in \mathcal{F} \Rightarrow \bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$
3. $A, B \in \mathcal{F} \implies A \cup B \in \mathcal{F}$
4. $A, B \in \mathcal{F} \implies A \cap B \in \mathcal{F}$

and

1. $\mathbb{P}(\emptyset) = 0$
2. $\mathbb{P}(A^{\complement}) = 1 - \mathbb{P}(A)$
3. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$
4. $A_1, \ldots A_k$ disjoint $\implies \mathbb{P}(\bigcup_{i=1}^{k} A_i) = \sum_{i=1}^{k} \mathbb{P}(A_i)$
5. $\mathbb{P}(\bigcup_{i=1}^{k} A_i) = \sum_{i=1}^{k} \sum_{1 \leq j_1 < \ldots < j_i \leq k} \mathbb{P}(A_{j_1} \cap \ldots \cap A_{j_i})$
6. (Union Bound) $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} \mathbb{P}(A_i)$
7. (Monotonicity) $A \subseteq B \implies \mathbb{P}(A) \leq \mathbb{P}(B)$
8. (De Morgan's Law) $(\bigcup_{i=1}^{\infty} A_i)^{\complement} = \bigcap_{i=1}^{\infty} (A_i)^{\complement}$
9. $A \setminus (A \cap B^{\complement}) = A \cap B$

2 Conditional probabilities

Conditional probability

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be some probability space and $A, B \in \mathcal{F}$ with $\mathbb{P}(B) > 0$. Then the **conditional probability of** A **given** B is defined by:

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

Proposition. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be some probability space and $B \in \mathcal{F}$. then $\mathbb{P}(\cdot \mid B)$ is a probability measure on $(\Omega, \mathcal{F})$.*

Law of total probability

Let $\mathcal{B} = (B_i)_{i \in I}$ be a countable partition of Ω. Then

$$\forall A \in \mathcal{F} \quad \mathbb{P}(A) = \sum_{i \in I : \mathbb{P}(B_i) > 0} \mathbb{P}(A \mid B_i) \cdot \mathbb{P}(B_i)$$

Chain Rule of Probability

If $\mathbb{P}(\bigcap_{i=1}^{n} A_i) > 0$, then we can write it as:

$\mathbb{P}(\cap_{i=1}^{n} A_i) = \mathbb{P}(A_1)\mathbb{P}(A_2 \mid A_1) \cdots \mathbb{P}(A_n \mid A_1 \cap \ldots \cap A_{n-1})$

Bayes' theorem

Let $\mathcal{B} = (B_i)_{i \in I}$ be a countable partition of Ω where $\mathbb{P}(B_i) > 0$ for all i. Then for all $A \in \mathcal{F}$ with $\mathbb{P}(A) > 0$:

$$\mathbb{P}(B_i \mid A) = \frac{\mathbb{P}(A \mid B_i) \cdot \mathbb{P}(B_i)}{\sum_{j \in I} \mathbb{P}(A \mid B_j) \cdot \mathbb{P}(B_j)}$$

3 Independence

Independence of two events

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A collection of events $A_1, \ldots, A_n$ is called independent, if

$$\forall I \subseteq \{1, \ldots, n\} \quad \mathbb{P}\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} \mathbb{P}(A_i)$$

A, B ind. $\iff \mathbb{P}(A \mid B) = \mathbb{P}(A) \iff \mathbb{P}(B \mid A) = \mathbb{P}(B)$

- If A, B are independent, then likewise $A, B^{\complement}$.
- $\mathbb{P}(A) \in \{0, 1\} \implies A$ is independent for all events.
- If A is independent of itself, then $\mathbb{P}(A) \in \{0, 1\}$

4 Random variables

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A random variable is **measurable** map $X : \Omega \to \mathbb{R}$ such that for all $a \in \mathbb{R}$:

$$\{\omega \in \Omega \mid X(\omega) \leq a\} \in \mathcal{F}$$

Furthermore we can define **events** in terms of a **random variable** where we use an abuse of notation:

$$\{X \leq a\} = \{\omega \in \Omega \mid X(\omega) \leq a\}$$

Which has a probability of $\mathbb{P}(X \leq a) = \mathbb{P}(\{X \leq a\})$.

Preimage of a random variable X

$$X^{-1}(A) = \{\omega \in \Omega \mid X(\omega) \in A\} = \text{"}\boldsymbol{X \in A}\text{"} \quad A \subset \mathbb{R}$$

X is $\boldsymbol{\mathcal{F}}$**-measurable** if $\forall B \subsetneq \mathbb{R}$, B closed: $X^{-1}(B) \in \mathcal{F}$.

Distribution of a Random Variable

For $A \in \mathcal{B}(\mathbb{R})$, we define the distribution (**law**) of X as

$\mu(A) = \mathbb{P}_X(A) := \mathbb{P}(X^{-1}(A)) = \mathbb{P}(\{\omega \in \Omega \mid X(\omega) \in A\})$

Cumulative Distribution Function (CDF)

Let X be a random variable on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, the CDF $F_X : \mathbb{R} \to [0,1]$ is defined by

$$\forall x \in \mathbb{R} \quad F_X(x) = \mathbb{P}(X^{-1}((-\infty, x))) =: \mathbb{P}(X \leq x)$$

Properties of the CDF (characterized by RCLL, Càdlàg):

- $a < b \implies \mathbb{P}(a < X \leq b) = F_X(b) - F_x(a)$
- F_X is monotonically increasing
- F_X is right-continuous, i.e. $\lim_{y \to x^+} F_X(y) = F_X(x)$
- $\lim_{x \to -\infty} F_X(x) = 0$ and $\lim_{x \to \infty} F_X(x) = 1$

Independence of random variables

The RVs $X_1, \ldots, X_n$ are called independent, if

$$\forall x_1, \ldots, x_n \in \mathbb{R} : \quad \mathbb{P}(X_1 \leq x_1, \ldots, X_n \leq x_n) = \mathbb{P}(X_1 \leq x_1) \cdot \ldots \cdot \mathbb{P}(X_n \leq x_n)$$

4.1 Discrete random variables

A random variable X is called **discrete** if the set of values it can output $\mathcal{X}$, is a discrete (a finite or countable set).

Probability Mass Function (PMF)

Let X be a discrete random variable and $\mathcal{X}$ be the set of all its values. The probability mass function $p_x(x)$ for each $x \in \mathcal{X}$ is defined by:

$$p_X(x) = \mathbb{P}(X = x) = \mathbb{P}_X(\{x\}) = \mathbb{P}(X^{-1}(\{x\}))$$

CDF of a discrete random variable

$$F_X(x) = \mathbb{P}(X \le x) = \sum_{x' \in \mathcal{X} \le x} p_X(x')$$

4.2 Continuous Random Variable

A random variable X is called **continuous** if the set of values it can produce is uncountably infinite and the probability of attaining a single value is zero.

Probability Density Function (PDF)

Let X be a continuous random variable. If there exists a (measureable) function $f_x : \mathbb{R} \to [0, \infty)$, such that

$$\mathbb{P}_X(I) = \mathbb{P}(X \in I) = \int_I f_x(x)\, dx$$

for all intervals I in $\mathbb{R}$, we call it the **PDF**.

Notice that $\int_{\mathbb{R}} f_x(x)\, dx = 1$ since $\mathbb{P}_X(\mathbb{R}) = 1$.

CDF of a continuous random variable

$$F_x(x) = \mathbb{P}(X \le x) = \int_{-\infty}^{x} f_x(t)\, dt = \mathbb{P}_X((-\infty, x])$$

The definition leads to the following consequences:

- $\mathbb{P}(a \le x \le b) = \mathbb{P}(a < x < b)$
- $\mathbb{P}(X = x) = 0$
- $\mathbb{P}(X \in [a, b]) = \mathbb{P}(X \in (a, b))$
- F_x differentiable $\implies \frac{dF_x}{dx}(x_0) = f_x(x_0)$

5 Expectation

For any random variable $X : \Omega \to \mathbb{R}_+$ with non-negative values, the expected value is defined with the CDF as:

$$\mathbb{E}(X) = \int_{\mathbb{R}} x\, dF_X(x) = \int_{\mathbb{R}} x \frac{F_X(x)}{dx}\, dx$$

Expected Value, Discrete

Let X be a discrete random variable with value in $\mathcal{X}$. The expected value is defined as (if the sum is well defined):

$$\mathbb{E}(X) := \sum_{x \in \mathcal{X}} x \cdot p_X(x) = \sum_{\omega \in \Omega} X(\omega) \cdot p(\omega)$$

Proposition. $\mathbb{E}(1_A) = \mathbb{P}(A)$.

Proposition. *If* $\mathcal{X} \subseteq \mathbb{N}_0$, *then* $\mathbb{E}(X) = \sum_{j=0}^{\infty} \mathbb{P}(X > j)$.

Law of the unconscious statistician (LOTUS)

$$\mathbb{E}(\phi(X)) = \sum_{x \in \mathcal{X}} \phi(x) \cdot p_X(x)$$

Expected Value, Continuous

Let X be a continuous random variable with density f_x. The expected value is defined as:

$$\mathbb{E}(X) := \int_{\mathbb{R}} x \cdot f_X(x)\, dx$$

LOTUS

$$\mathbb{E}(\phi(X)) = \int_{\mathbb{R}} \phi(x) \cdot f_X(x)\, dx$$

Proposition. *Falls* $\mathcal{X} = \mathbb{R}_{\ge 0}$, *dann gilt*

$$\mathbb{E}(X) = \int_0^{\infty} (1 - F_X(x))\, dx$$

Linearity of expected value

$$\mathbb{E}(\alpha X + \beta Y) = \alpha \mathbb{E}(X) + \beta \mathbb{E}(Y) \quad (a, b \in \mathbb{R})$$

Expected value of XY, independence

$$X, Y \text{ ind.} \Leftrightarrow \mathbb{E}(\phi(X)\psi(Y)) = \mathbb{E}(\phi(X))\mathbb{E}(\psi(Y))$$

This theorem also holds for $X_1, \dots X_n$ and $\phi_1, \dots, \phi_n$.

5.1 Variance and Covariance

Variance

Let X be a random variable such that $\mathbb{E}(X^2) < \infty$.

$$\mathrm{Var}(X) = \sigma_X^2 = \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$$

σ_X is called the **standard deviation** of X.

- If $\mathbb{E}(X^2) < \infty, \lambda, \alpha \in \mathbb{R}$, then $\mathrm{Var}(\lambda X + \alpha) = \lambda^2 \mathrm{Var}(X)$.
- If $S = X_1 + \ldots + X_n$, with $X_1, \ldots, X_n$ pairwise independent (or uncorrelated), then $\sigma_S^2 = \sum_{i=1}^n \sigma_{X_i}^2$

Covariance / Correlation

Let X, Y be two random variables with $\mathbb{E}(X^2) < \infty$ and $\mathbb{E}(Y^2) < \infty$, then their covariance is defined as

$$\begin{aligned}\mathrm{Cov}(X, Y) &= \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))) \\ &= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)\end{aligned}$$

If $\mathrm{Cov}(X, Y) = 0$, then they are uncorrelated.

$$\varrho(X, Y) = \mathrm{corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}}$$

- $\mathrm{Cov}(X, X) = \sigma_X^2 = \mathrm{Var}(X)$
- X, Y independent $\implies \mathrm{Cov}(X, Y) = 0$

5.2 Inequalities

Lemma (Monotonicity). $X \le Y \implies \mathbb{E}(X) \le \mathbb{E}(Y)$

Markov's Inequality

Let X be a R.V, $g : \mathcal{X} \to [0, \infty)$ monotonically increasing, then for $a \in \mathbb{R}$:

$$\mathbb{P}(X \ge a) \le \frac{\mathbb{E}(g(X))}{g(a)}$$

Jensen's Inequality

Let X be a R.V, $\phi : \mathbb{R} \to \mathbb{R}$ a convex function, then

$$\phi(\mathbb{E}(X)) \le \mathbb{E}(\phi(X))$$

Chebychev's Inequality

Let X be a R.V. s.t. $\mathbb{E}(X^2) < \infty$, then for $t > 0$

$$\mathbb{P}(|X - \mathbb{E}(X)| \ge t) \le \frac{\mathrm{Var}(X)}{t^2}$$

6 Joint Distributions

Joint Distribution

Let $X_1, \dots, X_n$ be R.V. defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Consider $X = (X_1, \dots, X_n)$. The joint distribution $\mathbb{P}_X$ on the hyper set $A \in \mathcal{B}^n$ where $\mathcal{B}^n = \sigma(\{A_1 \times \cdots \times A_n \mid A_i \in \mathcal{B}\})$ is defined as:

$$\mathbb{P}_X(A) = \mathbb{P}(X^{-1}(A)) = \mathbb{P}(\{\omega \in \Omega \mid X(\omega) \in A\})$$

Joint PMF

Let $X = X_1, \dots, X_n$ be discrete R.V., defined on the same probability space.

$$p_X(x_1, \dots, x_n) = \mathbb{P}(X_1 = x_1, \dots, X_n = x_n)$$

Joint CDF, discrete

Let $X = X_1, \dots, X_n$ be discrete R.V., defined on the same probability space.

$$F_X(x_1,\dots,x_n) = \mathbb{P}(X_1 \le x_1, \dots, X_n \le x_n) = \sum_{y_1 \le x_1, \dots, y_n \le x_n} p(y_1, \dots, y_n)$$

Proposition. *Let $X_1, \dots X_n$ be discrete R.V. with $X_i \in \mathcal{X}_i$. Then $Z = \phi(X_1, \dots, X_n)$ is a discrete random variable with values in $\mathcal{X} = (\mathcal{X}_1 \times \dots \times \mathcal{X}_n)$ and with distribution given by $\forall z \in \mathcal{X}$:*

$$\mathbb{P}(Z = z) = \sum_{\substack{x_1 \in \mathcal{X}_1, \dots, x_n \in \mathcal{X}_n \\ \phi(x_1, \dots, x_n) = z}} \mathbb{P}(X_1 = x_1, \dots, X_n = x_n)$$

Marginal distribution, discrete

Let $X_1, \dots, X_n$ be discrete R.V. with joint PMF. For every i we have $\forall z \in \mathcal{X}_i$:

$$\mathbb{P}(X_i = z) = \sum_{x_1, \dots x_{i-1}, x_{i+1}, \dots, x_n} p(x_1, \dots, x_{i-1}, z, x_{x+1}, \dots, x_n)$$

Proposition. *The expected values of joint discrete R.V. can be written as:*

$$\mathbb{E}(\phi(X_1, \dots, X_n)) = \sum_{x_1 \in \mathcal{X}_1, \dots, x_n \in \mathcal{X}_n} \phi(x_1, \dots, x_n) \cdot p(x_1, \dots, x_n)$$

Joint PDF

Let the R.V. $X = X_1, \dots, X_n$ be defined on the same probability space. We say these R.V. have a joint PDF $f_{X_1, \dots, X_n} = f_X$ if for all subsets $A \in \mathcal{B}^n$ it holds

$$\mathbb{P}_X(A) = \mathbb{P}((X_1, \dots, X_n) \in A)$$
$$= \int_A f_X(x_1, \dots, x_n)\, dx_1 \dots dx_n$$

Joint CDF, continuous

Let $X = X_1, \dots, X_n$ be continuous R.V., defined on the same probability space.

$$F_X(x_1, \dots, x_n) = \mathbb{P}(X_1 \le x_1, \dots, X_n \le x_n)$$
$$= \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f_X(x_1, \dots, x_n)\, dx_n \dots dx_1$$

Proposition. *Let $X_1, \dots, X_n$ be continuous R.V. with joint PDF and $\phi : \mathbb{R}^n \to \mathbb{R}$. The expected value is defined as*

$$\mathbb{E}(\phi(X_1, \dots, X_n)) =$$
$$\int_{\mathbb{R}} \dots \int_{\mathbb{R}} \phi(x_1, \dots, x_n) \cdot f(x_1, \dots, x_n)\, dx_n \dots dx_1$$

Marginal density, continuous

Let $X_1, \dots, X_n$ be continuous R.V. with a joint PDF. For every i we have $\forall z \in \mathcal{X}_i$:

$$f_{X_i}(z) = \int_{x_1, \dots, x_{i-1}, x_{x+1}, \dots, x_n} f(x_1, \dots, x_{i-1}, z, x_{i+1}, \dots, x_n)\, dx_n \dots dx_1$$

Independence of joint Random Variables

Let $X_1, \dots, X_n$ be R.V. with $p_{X_1}, \dots, p_{X_n}$.

$X_1, \dots X_n$ are independent

$\Leftrightarrow\ p(x_1, \dots, x_n) = p_{X_1}(x_1) \cdot \dots \cdot p_{X_n}(x_n)$

$\Leftrightarrow\ \mathbb{E}(\phi_1(X_1) \cdot \dots \cdot \phi_n(X_n)) = \mathbb{E}(\phi_1(X_1)) \cdot \dots \cdot \mathbb{E}(\phi_n(X_n))$

6.1 Useful lemmas and theorems

Distribution of transformed Random Variables

Let $X = (X_1, \dots, X_n)$ and $Y(\omega) = g(X(\omega))$ be two random variable with $g : \mathbb{R}^n \to \mathbb{R}^m$. Then for any $A \in \mathcal{B}(\mathbb{R})$

$$\mu_Y(A) = \mu_X(g^{-1}(A)).$$

Transformation Theorem

Let $g(x) = m + Bx$ with $\det(B) \neq 0$. If μ_X is abs. continuous, then μ_Y is abs. continuous and furthermore

$$f_Y(x) = \frac{1}{|\det(B)|} f_X(B^{-1}(x - m)).$$

Exp. $X_1, X_2 \sim U(0,1)$, CDF of $Y = X_1 + X_2, Z = X_1 - X_2$? Then $g(x_1, x_2) = (x_1 + x_2, x_1 - x_2)$ and $B = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ with $\det(B) = 2$, $g(x) = Bx$. Thus $(Y, Z) = g(X_1, X_2)$ and with the transformation theorem we get

$$f_{Y,Z}(y, z) = f_{X_1,X_2}(g^{-1}(y, z)) = \frac{1}{2} f_{X_1,X_2}\left(\frac{y+z}{2}, \frac{y-z}{2}\right)$$

Proposition. *If the R.V. (X, Y) has a density function f, then $Z = X + Y$ has a density $f_Z(z) = \int_{\mathbb{R}} f(x, z - x)\, dx$.*

Convolution

If X, Y indep., then $Z = X + Y$ has a density function:

$$(f_X * f_Y) := f_Z(z) = \int_{\mathbb{R}} f_X(x) f_Y(z - x)\, dx$$

$$\forall k \in \mathbb{N}_0 : \mathbb{P}[X + Y = k] = \sum_{j=0}^{k} \mathbb{P}[X = j]\mathbb{P}[Y = k - j]$$

Proposition. *Suppose $X_1, \dots, X_n$ are i.i.d with CDF F. Then $V = \max\{X_1, \dots, X_n\}$ has CDF $F_V(x) = F^n(x)$ as $\mathbb{P}(\max\{X_1, \dots, X_n\} < x) = \mathbb{P}(X_1 < x, \dots, X_n < x)$ and $U = \min\{X_1, \dots, X_n\}$ has CDF $F_U(x) = 1 - (1 - F(x))^n$.*

Transformation of Bivariate Random Variables

Let (X, Y) be two R.V. with $f_{X,Y}$. If we have a bijective transformation $U = g_1(X, Y), V = g_2(X, Y)$ with inverses $X = h_1(U, V)$ and $Y = h_2(U, V)$, then:

$$f_{U,V}(u, v) = f_{X,Y}(h_1(u, v), h_2(u, v))|J| \quad J = \begin{bmatrix} \frac{\partial h_1}{\partial u} & \frac{\partial h_1}{\partial v} \\ \frac{\partial h_2}{\partial u} & \frac{\partial h_2}{\partial v} \end{bmatrix}$$

CDF/PDF of $Z = g(X, Y)$

1. For each z, find the set $A_z = \{(x, y) \mid g(x, y) \le z\}$
2. Find the CDF $F_Z(z) = \mathbb{P}(\{(x, y) \mid g(x, y) \le z\}) = \int_{\mathbb{R}} \int_{\mathbb{R}} p_{X,Y}(x, y) 1_{\{(x,y) \in A_z\}}\, dx\, dy$
3. The pdf is $p_Z(z) = F'_Z(z)$

7 Conditional Distribution

Conditional PMF

Let X, Y be two discrete R.V. on the same probability space taking values in $\mathcal{X}$ and $\mathcal{Y}$. For $y \in \mathcal{Y}$ such that $\mathbb{P}(Y = y) \neq 0$, we define the *conditional distribution of X given $Y = y$* as

$$p_{X|Y}(x \mid y) := \frac{p_{X,Y}(x,y)}{p_Y(y)} = \frac{p_{X,Y}(x,y)}{\int_{\mathbb{R}} p_{X,Y}(x,y)\,dx}$$

Proposition. *X, Y independent $\implies p(x \mid y) = p(x)$.*

Proposition. *$p_{X|Y}(\cdot \mid y)$ is a PMF if $p_Y(y) > 0, y \in \mathcal{Y}$. Thus, we can consider a R.V. "$X \mid Y = y$".*

Multiple variables

$$p_{X_{k+1},\dots,X_n|X_1,\dots,X_k}(x_{k+1},\dots,x_n \mid x_1,\dots x_k) := \frac{p_{X_1,\dots,X_n}(x_1,\dots,x_n)}{p_{X_1,\dots,X_k}(x_1,\dots,x_k)}$$

Conditional Density

Let X, Y be two continuous random variables with joint density $f_{X,Y}$. The *conditional density of X given $Y = y$* is defined as

$$f_{X|Y}(x \mid y) = \begin{cases} \frac{f_{X,Y}(x,y)}{f_Y(y)} & \text{if } f_Y(y) > 0 \\ 0 & \text{otherwise} \end{cases}$$

7.1 Mixed case

Conditional probability density

Let X be a continuous and Y a discrete R.V. For each $y \in \mathcal{Y}$, we have the conditional probability distribution $\mathbb{P}_{X|y}$ on $\mathbb{R}$ for each subset $A \in \mathcal{B}(\mathbb{R})$ defined as

$$\mathbb{P}_{X|y}(A) := \frac{\mathbb{P}(X \in A, Y = y)}{p_Y(y)}$$

If this conditional distribution has a density, we call it the *conditional probability density of X given $Y = y$*

Law of Total Probability

$$f_X(x) = \int f_{X|Y}(x \mid y) f_Y(y)\,dy$$

If Y discrete, the integral is replaced by sum over $y \in \mathcal{Y}$.

Bayes' Rule, X continuous

$$f_{X|Y}(x \mid y) = \frac{f_{Y|X}(y \mid x) f_X(x)}{\int f_{Y|X}(y \mid x') f_X(x')\,dx'}$$

If X discrete, the integral is replaced by sum over $x' \in \mathcal{X}$.

Chain Rule

$$p(x_1,\dots,x_n) = p(x_1)p(x_2 \mid x_1)\cdots p(x_n \mid x_1,\dots,x_n)$$

Conditional Independence

X and Y are conditionally independent given Z if $\forall x, y, z$:

$$p_{X,Y|Z}(x,y \mid z) = p_{X|Z}(x \mid z) p_{Y|Z}(y \mid z)$$

8 Convergence to Expectation

Independent with identical distribution (iid.)

$X_1, X_2, \dots$ on the same probability space are iid. if they are independent and each X_i has the same distribution.

Convergence of Random Variables

Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of random variables. Then:

1. X_n converges to X **almost surely** if
$$\mathbb{P}(\{w \in \Omega \mid \lim_{n\to\infty} X_n(\omega) = X(\omega)\}) = 1$$
and we write $X_n \overset{a.s.}{\to} X$ as $n \to \infty$.
2. X_n converges to X **in probability**, if for any $\epsilon > 0$
$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \epsilon) = 0$$
and we write $X_n \overset{\mathbb{P}}{\to} X$ as $n \to \infty$.
3. X_n converges to X **in distribution**, if for all continuity points x of F_X we have
$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$
and we write $X_n \overset{\mathcal{D}}{\to} X$ as $n \to \infty$.

Lemma. *$X_n \overset{a.s.}{\to} X \implies X_n \overset{\mathbb{P}}{\to} X \implies X_n \overset{\mathcal{D}}{\to} X$.*

Sample Mean / Sample Variance

$X_1, \dots, X_n$ iid. with $\mathbb{E}(X_i) < \infty$.
$\bar{X}_n := \frac{1}{n}\sum_{i=1}^n X_i$. $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$.

Laws of Large Numbers (LLN)

Let $X_1, X_2, \dots$ be iid. R.V. with $\mu := \mathbb{E}(X_i) < \infty$ and $\sigma^2 := \mathrm{Var}(X_i) < \infty$. Then the weak LLN (**WLLN**) states that as $n \to \infty$

$$\bar{X}_n \overset{\mathbb{P}}{\to} \mu.$$

The strong LLN (**SLLN**) states as $n \to \infty$

$$\bar{X}_n \overset{a.s.}{\to} \mu.$$

Central Limit Theorem (CLT)

Let $X_1, X_2, \dots$ be iid. R.V. with $\mu := \mathbb{E}(X_i) < \infty$ and $\sigma^2 := \mathrm{Var}(X_i) < \infty$. Then for $S_n := \sum_{i=1}^n X_i$ we have

$$\forall x \in \mathbb{R}: \quad \mathbb{P}\left(\frac{S_n - n\mu}{\sqrt{\sigma^2 n}} \leq x\right) \underset{n\to\infty}{\to} \Phi(x)$$

where Φ is the CDF of the $\mathcal{N}(0,1)$-distribution.

9 Conditional Expectation

Conditional expectation on an event

Let $B \in \mathcal{F}$ denote an event with $\mathbb{P}(B) > 0$, X a R.V.

$$\mathbb{E}(X \mid B) = \frac{\mathbb{E}(1_B X)}{\mathbb{P}(B)} = \sum_{x\in\mathcal{X}} x \cdot \mathbb{P}(X = x \mid B) = \sum_{\omega\in\Omega} X(\omega)\mathbb{P}(\{\omega\} \mid B)$$

Conditional expectation on R.V.

$$\mathbb{E}(X \mid Y = y) := \int x\, f_{X|Y}(x \mid y)\,dx$$

If X is discrete, replace integral by sum over $x \in \mathcal{X}$.

This introduces a new random variable:

$$\omega \in \Omega \mapsto Y(\omega) \mapsto \mathbb{E}(X \mid Y = Y(\omega))$$

We define $\mathbb{E}(X \mid Y)$ to denote this new random variable and call it the conditional expectation of X given Y.

Tower Property

$$\mathbb{E}(X) = \mathbb{E}(\mathbb{E}(X \mid Y))$$

Law of Total Variance

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

10 Estimators

We make the following assumptions:

- Parameter space $\Theta \subseteq \mathbb{R}^m$
- Family of probability distributions $\{\mathbb{P}_\theta \mid \theta \in \Theta\}$ on the sample space $(\Omega, \mathcal{F})$; for each element $\theta = (\theta_1, \ldots, \theta_m)$ in the parameter space, there exists a probability model $\mathbb{P}_\theta$, which describes the relationship between the data and the unknown parameters.
- Random variables $X_1, \ldots, X_n$ on $(\Omega, \mathcal{F})$
- We refer to the observed data $x_1, \ldots, x_n$ or the random variables $X_1, \ldots, X_n$ as a sample.
- For each $\theta_1, \ldots, \theta_m$ we search for an estimator T_j.

Estimator

An estimator is a random variable $T_j : \Omega \to \mathbb{R}$ of form

$$T_j = t_j(X_1, \ldots X_n) \qquad t_j : \mathbb{R}^n \to \mathbb{R}$$

Notation: $T = (T_1, \ldots, T_m)$ and $\theta = (\theta_1, \ldots, \theta_m)$.

Unbiased

An estimator T is unbiased, if $\forall \theta \in \Theta : \mathbb{E}_\theta(T) = \theta$.

Bias

The bias of T in the model $\mathbb{P}_\theta$ is defined as $\mathbb{E}_\theta(T) - \theta$.

Consistent

The sequence $T^{(n)}$ is **consistent** if $T^{(n)} \xrightarrow{\mathbb{P}_\theta} \theta$ as $n \to \infty$.

Mean Squared Error (MSE)

$$\begin{aligned}\mathrm{MSE}_\theta(T) &= \mathbb{E}_\theta((T - \theta)^2) \\ &= \mathrm{Var}_\theta(T) + (\mathbb{E}_\theta(T) - \theta)^2\end{aligned}$$

An estimator T gives us a random single value for θ, drawn from the sampling distribution of the estimator, which represents the distribution of possible values that we would get if we were to calculate the estimator for many different samples of data.

10.1 Maximum-Likelihood-Method

Let $X_1, \ldots, X_n$ be random variables on $\mathbb{P}_\theta$. The **Likelihood-function** $L : \Theta \to \mathbb{R}$ is defined as:

$$L(x_1, \ldots, x_n; \theta) = \begin{cases} p(x_1, \ldots, x_n; \theta) & \text{discrete case} \\ f(x_1, \ldots, x_n; \theta) & \text{continuous case} \end{cases}$$

with p (or f) being the joint PMF/PDF. Then for each $x_1, \ldots, x_n \in \mathcal{X}$ let $t_{\mathrm{ML}}(x_1, \ldots, x_n)$ be the value which maximizes the function $\theta \mapsto L(X_1, \ldots, X_n; \theta)$. The **Maximum-Likelihood-Estimator** is then defined as

$$T_{\mathrm{ML}} = t_{\mathrm{ML}}(X_1, \ldots, X_n) \in \arg\max_\theta L(X_1, \ldots, X_n; \theta).$$

Alternative notation of MLE (from IML)

$$\hat{\theta}_{\mathrm{MLE}} = \arg\max_\theta L(\theta; X, y) = p(y_1, \ldots, y_n \mid x_1, \ldots, x_n, \theta)$$

10.2 Method of Moments

k-th sample moment

$$\hat{\mu_k} = \frac{1}{n} \sum_{i=1}^{n} x_i^k$$

Suppose we know the formula of the first k moments $\mu_1 = f_1(\theta_1, \ldots, \theta_n)$, …, $\mu_n = f_n(\theta_1, \ldots, \theta_n)$, then the method of moments estimator for $\theta_1, \ldots, \theta_n$ is denoted by $\hat{\theta}_1, \ldots, \hat{\theta}_n$ and can be solved with the following system of equations:

$$\hat{\mu_1} = f_1(\hat{\theta}_1, \ldots, \hat{\theta}_n) \quad \ldots \quad \hat{\mu}_n = f_n(\hat{\theta}_1, \ldots, \hat{\theta}_n)$$

10.3 Bayesian Inference

Instead of making a point estimator, Bayes' inference provides a full posterior distribution that captures uncertainty in parameter estimation.

We assume $\theta \sim \pi_0$ ($\mathbb{P}(\theta = t) = \pi_0(t)$) as **prior**. Then we can calculate the **posterior distribution**:

$$\begin{aligned}\mathbb{P}(\theta \mid x_1, \ldots, x_n) &= \frac{\mathbb{P}(x_1, \ldots, x_n, \theta)}{\mathbb{P}(x_1, \ldots, x_n)} \\ &= \frac{\mathbb{P}(x_1, \ldots, x_n \mid \theta)\pi_0(\theta)}{\int f(x_1, \ldots, x_n \mid \theta = \theta')\pi_0(\theta')\, d\theta'}\end{aligned}$$

11 Tests

Null hypothesis H_0, alternative Hypothesis H_A

Two subsets $\Theta_0 \subseteq \Theta, \Theta_A \subseteq \Theta$, s.t. $\Theta_0 \cap \Theta_A = \emptyset$. A hypothesis is **simple** if $\Theta_0 = \{\alpha \in \Theta\}$ or else **composite**.

Test

A **test** is a tuple (T, K), where T is a random variable of form $T = t(X_1, \ldots, X_n)$ and $K \subseteq \mathbb{R}$ a deterministic subset of $\mathbb{R}$. T is called the **test statistic** and K is the **rejection region** or critical region.

We reject a hypothesis $H_0 \iff 1_{t(x_1,\ldots,x_n) \in K} = 1$.

Error types

- **Type I Error**: Rejecting a true H_0.
  Probability: $\mathbb{P}_\theta(T \in K),\ \theta \in \Theta_0$.
- **Type II Error**: Failing to reject a false H_0.
  Probability: $\mathbb{P}_\theta(T \notin K) = 1 - \mathbb{P}_\theta(T \in K),\ \theta \in \Theta_A$.

1. Minimize type I errors:

Significance level of a test

A test has a significance level $\alpha \in [0, 1]$ if

$$\sup_{\theta \in \Theta_0} \mathbb{P}_\theta(T \in K) \le \alpha$$

2. Maximize the power of the test (reducing type II errors)

Power of a test

The power of a test is defined as:

$$\beta : \Theta_A \to [0, 1], \quad \theta \mapsto \beta(\theta) := \mathbb{P}_\theta(T \in K)$$

As this is asymmetric and thus makes it harder to reject H_0 than accepting it, we take the **negation** of the wanted statement as H_0.

11.1 How to find the hypothesis

1. List the things to avoid.
2. Take that as Type I Error.
3. Conclude the H_0 with reverse engineering.

11.2 z-test

Prerequisites: $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ iid; $\boldsymbol{\sigma^2}$ **known**.
Hypothesis: $H_0 : \theta = \theta_0$, $H_A : \theta > \theta_0$ or $\theta < \theta_0$ or $\theta \neq \theta_0$.
Teststatistic: $T := \frac{\bar{X_n} - \theta_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$ under $\mathbb{P}_{\theta_0}$.
Rejection Region:

$$H_A : \theta > \theta_0 \implies K_> = (c_>, \infty)$$
$$H_A : \theta < \theta_0 \implies K_< = (-\infty, c_<)$$
$$H_A : \theta \neq \theta_0 \implies K_{\neq} = (-\infty, -c_{\neq}) \cup (c_{\neq}, \infty)$$

How to find $c_>, c_<, c_{\neq}$:

$$\alpha = \mathbb{P}_{\theta_0}(T > c_>) \implies c_> = \Phi^{-1}(1-\alpha)$$
$$\alpha = 1 - \mathbb{P}_{\theta_0}(T \leq -c_<) \implies c_< = -\Phi^{-1}(1-\alpha)$$
$$\alpha = 2(1 - \mathbb{P}(T \leq c_{\neq})) \implies c_{\neq} = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$$

11.3 t-test

Prerequisites: $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ iid.
Hypothesis: $H_0 : \mu = \mu_0$, H_A : same as z-Test.
Teststatistic: $T := \frac{\bar{X_n} - \mu_0}{S/\sqrt{n}} \sim t_{n-1}$ $\quad (S = \sqrt{S^2})$.
Rejection Region: Same as z-Test.
How to find $c_>, c_<, c_{\neq}$:

$c_> = t_{n-1,1-\alpha} \quad c_< = t_{n-1,\alpha} = -t_{n-1,1-\alpha} \quad c_{\neq} = t_{n-1,1-\frac{\alpha}{2}}$

11.4 Unpaired two-sample z-, t-test

Pre.: $X_1, \dots, X_n \sim \mathcal{N}(\mu_1, \sigma^2)$, $Y_1, \dots, Y_m \sim \mathcal{N}(\mu_2, \sigma^2)$ iid.
Hypothesis: $H_0 : \mu_1 - \mu_2 = \omega_0$, H_A : same as z-Test.
Rejection Region: Same as z-Test.
Teststatistic (known $\boldsymbol{\sigma^2}$):

$$T := \frac{(\bar{X_n} - \bar{Y_m}) - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim \mathcal{N}(0,1).$$

Teststatistic (unknown $\boldsymbol{\sigma^2}$):
$T := \frac{(\bar{X_n} - \bar{Y_m}) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{m+n-2}$ with sample variances S_1, S_2 and $S_p = \sqrt{\frac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2}}$.
How to find $c_>, c_<, c_{\neq}$: Same as z-, t-Test.

11.5 Paired two-sample z-, t-test

$Z_i := X_i - Y_i \sim \mathcal{N}(\mu_1 - \mu_2, 2\sigma^2)$. Then do z-, t-Test.

11.6 Construct Tests

Let $X_1, \dots, X_n$ be discrete or continuous under $\mathbb{P}_{\theta_0}$ and $\mathbb{P}_{\theta_A}$, where $\theta_0 \neq \theta_A$ both simple. Then the likelihood-quotient is:

$$R(x_1, \dots, x_n) = \frac{L(x_1, \dots, x_n; \theta_A)}{L(x_1, \dots, x_n; \theta_0)}.$$

$L(x_1, \dots, x_n; \theta_0) = 0 \implies R(x_1, \dots, x_n) = \infty$. Furthermore $R \gg 1 \implies H_A > H_0$ and $R \ll 1 \implies H_A < H_0$.

LQ-Test $(c \geq 0)$

$$T = R(X_1, \dots, X_n) \quad \text{and} \quad K = (c, \infty)$$

Neyman-Pearson-Lemma
Let (T, K) be a LQ-Test with $\alpha^* = \mathbb{P}_{\theta_0}(T \in K)$ and (T', K') be a any test with significance level $\alpha \leq \alpha^*$, then:

$$\mathbb{P}_{\theta_A}(T' \in K') \leq \mathbb{P}_{\theta_A}(T \in K)$$

11.7 Binomial Test

Assume $X_1, \dots, X_n \sim \text{Ber}(\theta)$ iid. under $\mathbb{P}_\theta$, then the teststatistic $T = \sum_{i=1}^n X_i$ with derivation:

$$R(x_1, \dots, x_n; \theta_0, \theta_A) = \left(\frac{\theta_A(1-\theta_0)}{\theta_0(1-\theta_A)}\right)^{\sum_{i=1}^n x_i} \left(\frac{1-\theta_A}{1-\theta_0}\right)^n$$

If $H_A : \theta > \theta_0$ then $\frac{\theta_A(1-\theta_0)}{\theta_0(1-\theta_A)} > 1$ and thus we can deduce:

$$R(x_1, \dots, x_n; \theta_0, \theta_A) \gg 1 \iff \sum_{i=1}^n x_i \gg 1$$

11.8 p-Value

Ordered Teststatistic

A family of tests $(T, K_t)_{t \geq 0}$ is ordered according to T if $K_t \subset \mathbb{R}$ and $s \leq t \implies K_t \subseteq K_s$.

- $K_t = (t, \infty)$: Right tailed $(H_A : \theta > \theta_0)$
- $K_t = (-\infty, -t)$: Left tailed $(H_A : \theta < \theta_0)$
- $K_t = (-\infty, -t) \cup (t, \infty)$: Two-tailed $(H_A : \theta \neq \theta_0)$

***p*-Value**

Let $H_0 : \theta = \theta_0$ and $(T, K_t)_{t \geq 0}$ an ordered family of tests.

$$p\text{-Value} : \Omega \to [0,1], \quad \omega \mapsto \mathbb{P}_{\theta_0}(T \in K_{T(\omega)})$$

1. If T continuous and $K_t = (t, \infty)$, then the p-Value under $\mathbb{P}_{\theta_0}$ is $\mathcal{U}([0,1])$.
2. For p-Value γ holds, that all tests with significance level $\alpha > \gamma$ H_0 will be rejected.

Thus a small p-Value $\implies$ H_0 is likely to be rejected.

12 Confidence Interval

Confidence Interval

A confidence interval with **confidence level** $1 - \alpha$ is a subset $C(x_1, \dots, x_n) \subseteq \Theta$ such that

$$\forall \theta \in \Theta \quad \mathbb{P}_\theta(C(X_1, \dots, X_n) \ni \theta) \geq 1 - \alpha.$$

Mostly this is a random interval $[u(X), v(X)]$.

Confidence Interval → Test

$C(x_1, \dots, x_n)$ with confidence level $1-\alpha$ known. We define a test statistic $1_{\{\theta_0 \notin C(X_1, \dots, X_n)\}}$. Thus we get:

$\mathbb{P}_{\theta_0}(\theta_0 \notin C(X_1, \dots, X_n)) = 1 - \mathbb{P}_{\theta_0}(C(X_1, \dots, X_n) \ni \theta_0) \leq \alpha$.

Confidence Interval ← Test

Given $\forall \theta_0 : T_{\theta_0}(x_1, \dots, x_n) \leq \alpha$. We define the confidence interval $C(X_1, \dots, X_n) = \{\theta \in \Theta \mid T_\theta(X_1, \dots, X_n) \notin K_\theta\}$. Thus we get:

$$\begin{aligned}\mathbb{P}_\theta(\theta \in C(X_1, \dots, X_n)) &= \mathbb{P}_\theta(T_\theta(X_1, \dots, X_n) \notin K_\theta) \\ &= 1 - \mathbb{P}_\theta(T_\theta(X_1, \dots, X_n) \in K) \\ &\geq 1 - \alpha.\end{aligned}$$

Approximative Confidence Interval

If $X_i \sim \mathbb{P}(\cdot \mid \theta)$ and we want to find a confidence interval with some confidence level $1 - \alpha$ for θ, we can use the CLT:

1. Calculate $c_{\neq} = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$
2. Transform $\mathbb{P}(-c_{\neq} \leq \frac{S_n - n\mathbb{E}(X_1)}{\sqrt{\text{Var}(X_1)n}} \leq c_{\neq}) = \mathbb{P}(\alpha \leq \theta \leq \beta)$
3. Then $C(X_1, \dots, X_n) = [\alpha, \beta]$

Useful (introduced) inequalities

- $\theta \in [\frac{1}{2}, 1] \implies \frac{\sqrt{1-\theta}}{\theta} \leq \sqrt{2}$
- $\theta \in [0,1] \implies \theta(1-\theta) \leq \frac{1}{4}$

Can be used for approximative confidence intervals.

13 Additional Stuff

Random Number Generation

Let F be a continuous, strictly monotone increasing CDF with inverse F^{-1}. Then

$$X \sim \mathcal{U}([0,1]), Y = F^{-1}(X) \implies F_Y = F.$$

Empirical distribution function

Just like the sample mean $\bar{X}_n$ and sample variance S^2 there is a empirical CDF defined as:

$$\hat{F}_n(t) := \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i<t\}} \overset{a.s.}{\to} \mathbb{E}(1_{(-\infty,t]}(X_1)) = F_X(t)$$

Monte-Carlo Integration

Used to approximate the value of an integral.

$$\int_{[0,1]^m} h(x_1,\ldots,x_m)\,dx_1 \ldots dx_m$$
$$= \mathbb{E}(h(U_1,\ldots,U_m)) \approx \frac{1}{N}\sum_{i=1}^{N} h(u_1^i,\ldots,u_m^i)$$

Where $U_1,\ldots,U_m \sim \mathcal{U}([0,1])$.

Moment generating function

Let X be a R.V and $t \in \mathbb{R}$, then the MGF is defined as:

$$M_X(t) := \mathbb{E}(e^{tX})$$

This is always well defined for $[0,\infty]$. Furthermore:

$$\frac{d^k}{dt^k}M_X(t)|_{t=0} = \mathbb{E}(X^k) = m_X^k \qquad (k\text{-th Moment}).$$

Chernoff Bound

$X_1,\ldots,X_n$ iid. s.t. $\forall t \in \mathbb{R} : \mathbb{E}(e^{tX}) < \infty$, $S_n := \sum_{i=1}^{n} X_i$.

$$\forall b \in \mathbb{R} \quad \mathbb{P}(S_n \geq b) \leq \exp\left(\inf_{t\in\mathbb{R}}(n \log M_X(t) - tb)\right)$$

For $X \sim \text{Bin}(n,p)$

- $\Pr[X \geq (1+\delta)\mathbb{E}[X]] \leq e^{-\frac{1}{3}\delta^2\mathbb{E}[X]} \quad \forall 0 < \delta \leq 1$
- $\Pr[X \leq (1-\delta)\mathbb{E}[X]] \leq e^{-\frac{1}{2}\delta^2\mathbb{E}[X]} \quad \forall 0 < \delta \leq 1$
- $\Pr[X \geq t] \leq 2^{-t}$ für $t \geq 2e\mathbb{E}[X]$

14 Analysis Stuff

Derivative and integration rules

Linearity: $(\alpha \cdot f(x) + g(x))' = \alpha \cdot f'(x) + g'(x)$
Product rule: $(f \cdot g)'(x) = f'(x) \cdot g(x) + f(x) \cdot g'(x)$
Quotient rule: $\left(\frac{f}{g}\right)'(x) = \frac{f'(x)\cdot g(x) - f(x)\cdot g'(x)}{g(x)^2}$
Chain rule: $(f \circ g)'(x) = f'(g(x)) \cdot g'(x)$
Inverse: $(f^{-1})'(y_0) = \frac{1}{f'(x_0)} = \frac{1}{f'(f^{-1}(y_0))}$, $y_0 = f(x_0)$
Part. Int.: $\int f(x)g'(x)\,dx = f(x)g(x) - \int f'(x)g(x)\,dx$

- Choose g': exp $\to$ trig $\to$ poly $\to$ inverse trig. $\to$ logs
- Choose f: logs $\to$ inverse trig. $\to$ poly $\to$ trig $\to$ exp
- Sometimes it is necessary to multiply by 1.
  E.g.: $\int \ln x \;dx = \int \ln x \cdot 1 \;dx \Rightarrow f(x) = \ln x,\; g'(x) = 1.$

Substitution

$$\int_{\phi(a)}^{\phi(b)} f(x)\,dx = \int_a^b f(\phi(t))\phi'(t)dt = (F\circ\phi)(b) - (F\circ\phi)(a)$$

since $F' = f$ then $f(\phi(t))\phi'(t) = (F \circ \phi)'(t)$.

14.1 Series

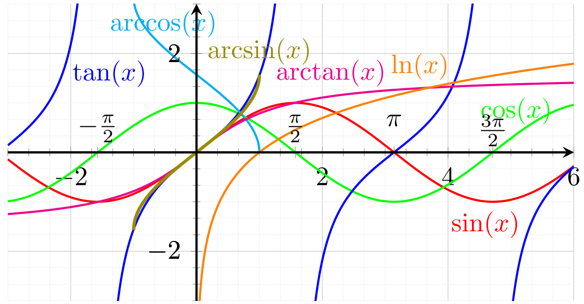
- Geometric: $\sum_{n=0}^{\infty} q^n = \frac{1}{1-q}$ if $|q| < 1$
- Harmonic: $\sum_{n=1}^{\infty} \frac{1}{n}$ diverges
- Telescope: $\sum_{n=0}^{\infty} \frac{1}{n(n+1)} = 1$
- $\exp(z) := \sum_{n=0}^{\infty} \frac{z^n}{n!} = \lim_{n\to\infty}(1 + \frac{z}{n})^n = e^z$
- $\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$ converges $s > 1$ $\left(\frac{1}{1-\frac{1}{2^{s-1}}}\right)$
- $\text{p}(z) = \sum_{k=0}^{\infty} c_k z^k$ conv. abs. $|z| < \rho = \frac{1}{\limsup |c_k|^{1/k}}$

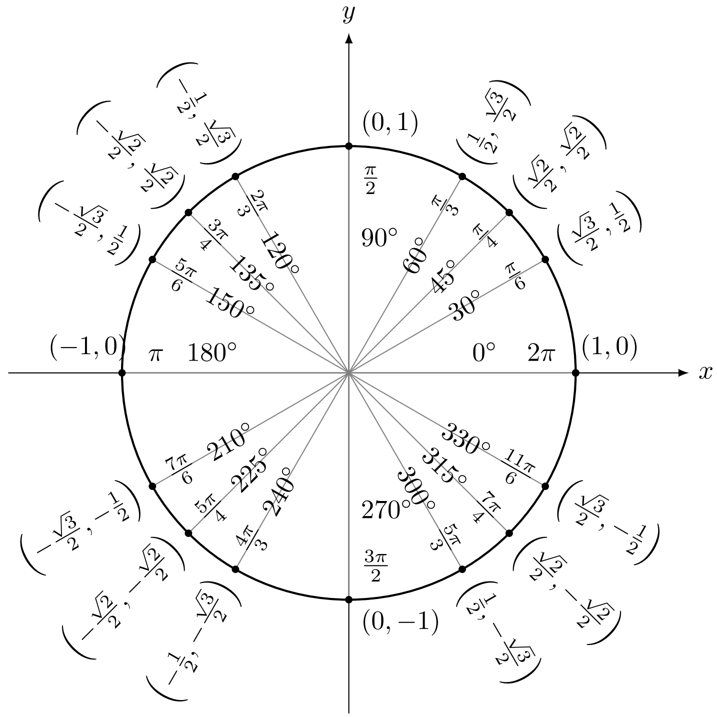
$\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$ $\qquad$ $\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}$
$\sum_{i=1}^{n} i^3 = \frac{n^2(n+1)^2}{4}$ $\qquad$ $\sum_{i=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$

14.2 Logarithm Rules

$\ln(1) = 0$	$\ln(e) = 1$
$\ln(xy) = \ln(x) + \ln(y)$	$\ln(x/y) = \ln(x) - \ln(y)$
$\ln(x^y) = y \cdot \ln(x)$	$x^\alpha \cdot x^\beta = x^{\alpha+\beta}$
$(x^\alpha)^\beta = x^{\alpha\cdot\beta}$	$\frac{x-1}{x} \leq \ln(x) \leq x - 1$
$\ln(1 + x^\alpha) \leq \alpha x$	$\log_\alpha(x) = \frac{\ln(x)}{\ln(\alpha)}$

14.3 Important Functions





$\binom{n}{k} = \frac{n!}{(n-k)!k!}$ $\qquad$ $(x+y)^n = \sum_{k=0}^{n}\binom{n}{k}x^{n-k}y^k$

$x = \frac{-b\pm\sqrt{b^2-4ac}}{2a}$

Derivatives and Integrals (src: dcamenisch)

$\mathbf{F(x)}$	$\mathbf{f(x)}$
c	0
x^a	$a \cdot x^{a-1}$
$\frac{1}{a+1}x^{a+1}$	x^a
$\frac{1}{a\cdot(n+1)}(ax+b)^{n+1}$	$(ax+b)^n$
$\frac{x^{a+1}}{a+1}$	$x^a,\ a \neq -1$
$\frac{1}{x}$	$-\frac{1}{x^2}$
$\sqrt{x}$	$\frac{1}{2\sqrt{x}}$
$\sqrt[n]{x}$	$\frac{1}{n}x^{\frac{1}{n}-1}$
$\frac{2}{3}x^{\frac{3}{2}}$	$\sqrt{x}$
$\frac{n}{n+1}x^{\frac{1}{n}+1}$	$\sqrt[n]{x}$
e^x	e^x
$\ln(\lvert x\rvert)$	$\frac{1}{x}$
$\log_a(\lvert x\rvert)$	$\frac{1}{x\ln(a)} = \log_a(e^{\frac{1}{x}})$
$\sin(x)$	$\cos(x)$
$\cos(x)$	$-\sin(x)$
$\tan(x) = \frac{\sin(x)}{\cos(x)}$	$\frac{1}{\cos^2(x)} = 1 + \tan^2(x)$
$\cot(x) = \frac{\cos(x)}{\sin(x)}$	$\frac{1}{-\sin^2(x)}$
$\arcsin(x)$	$\frac{1}{\sqrt{1-x^2}}$
$\arccos(x)$	$\frac{-1}{\sqrt{1-x^2}}$
$\arctan(x)$	$\frac{1}{1+x^2}$
$\sinh(x) = \frac{e^x+e^{-x}}{2}$	$\cosh(x)$
$\cosh(x) = \frac{e^x-e^{-x}}{2}$	$\sinh(x)$
$\tanh(x) = \frac{\sinh(x)}{\cosh(x)}$	$\frac{1}{\cosh^2(x)} = 1 - \tanh^2(x)$
$\frac{1}{f(x)}$	$\frac{-f'(x)}{(f(x))^2}$
a^{cx}	$a^{cx} \cdot c\ln(a)$
x^x	$x^x \cdot (1+\ln(x)),\ x > 0$
$(x^x)^x$	$(x^x)^x(x + 2x\ln(x)),\ x > 0$
x^{x^x}	$x^{x^x}(x^{x-1} + \ln(x) \cdot x^x(1+\ln(x)))$

$\mathbf{F(x)}$	$\mathbf{f(x)}$
$\frac{1}{a}\ln(\lvert ax+b\rvert)$	$\frac{1}{ax+b}$
$\frac{ax}{c} - \frac{ad-bc}{c^2}\ln(\lvert cx+d\rvert)$	$\frac{ax+b}{cx+d}$
$\frac{1}{2a}\ln\left(\left\lvert\frac{x-a}{x+a}\right\rvert\right)$	$\frac{1}{x^2-a^2}$
$\frac{x}{2}\sqrt{a^2+x^2} + \frac{a^2}{2}\ln\left(x+\sqrt{a^2+x^2}\right)$	$\sqrt{a^2+x^2}$
$\frac{x}{2}\sqrt{a^2-x^2} - \frac{a^2}{2}\arcsin\left(\frac{x}{\lvert a\rvert}\right)$	$\sqrt{a^2-x^2}$
$\frac{x}{2}\sqrt{x^2-a^2} - \frac{a^2}{2}\ln\left(x+\sqrt{x^2-a^2}\right)$	$\sqrt{x^2-a^2}$
$\ln\left(x+\sqrt{x^2 \pm a^2}\right)$	$\frac{1}{\sqrt{x^2\pm a^2}}$
$\arcsin\left(\frac{x}{\lvert a\rvert}\right)$	$\frac{1}{\sqrt{a^2-x^2}}$
$\frac{1}{a}\arctan\left(\frac{x}{a}\right)$	$\frac{1}{x^2+a^2}$
$-\frac{1}{a}\cos(ax+b)$	$\sin(ax+b)$
$\cos(ax+b)$	$-a\sin(ax+b)$
$\frac{1}{a}\sin(ax+b)$	$\cos(ax+b)$
$\sin(ax+b)$	$a\cos(ax+b)$
$-\ln(\lvert\cos(x)\rvert)$	$\tan(x)$
$\ln(\lvert\sin(x)\rvert)$	$\cot(x)$
$\ln\left(\left\lvert\tan\left(\frac{x}{2}\right)\right\rvert\right)$	$\frac{1}{\sin(x)}$
$\ln\left(\left\lvert\tan\left(\frac{x}{2}+\frac{\pi}{4}\right)\right\rvert\right)$	$\frac{1}{\cos(x)}$
$\frac{1}{2}(x - \sin(x)\cos(x))$	$\sin^2(x)$
$\frac{1}{2}(x + \sin(x)\cos(x))$	$\cos^2(x)$
$\frac{1}{4}(\frac{1}{3}\cos(3x) - 3\cos(x))$	$\sin^3(x)$
$\frac{1}{4}(\frac{1}{3}\sin(3x) + 3\sin(x))$	$\cos^3(x)$
$\tan(x) - x$	$\tan^2(x)$
$-\cot(x) - x$	$\cot^2(x)$
$x\arcsin(x) + \sqrt{1-x^2}$	$\arcsin(x)$
$x\arccos(x) - \sqrt{1-x^2}$	$\arccos(x)$
$x\arctan(x) - \frac{1}{2}\ln\left(1+x^2\right)$	$\arctan(x)$
$\ln(\cosh(x))$	$\tanh(x)$
$\ln(\lvert f(x)\rvert)$	$\frac{f'(x)}{f(x)}$

$\mathbf{F(x)}$	$\mathbf{f(x)}$
$x(\ln(\lvert x\rvert) - 1)$	$\ln(\lvert x\rvert)$
$\frac{1}{n+1}(\ln x)^{n+1} \quad n \neq -1$	$\frac{1}{x}(\ln x)^n$
$\frac{1}{2n}(\ln x^n)^2 \quad n \neq 0$	$\frac{1}{x}\ln x^n$
$\ln(\lvert\ln(x)\rvert) \quad x > 0, x \neq 1$	$\frac{1}{x\ln(x)}$
$\frac{1}{b\ln(a)}a^{bx}$	a^{bx}
$\frac{cx-1}{c^2} \cdot e^{cx}$	$x \cdot e^{cx}$
$\frac{1}{c}e^{cx}$	e^{cx}
$\frac{x^{n+1}}{n+1}\left(\ln(x) - \frac{1}{n+1}\right) \quad n \neq -1$	$x^n\ln(x)$
$\frac{e^{cx}(c\sin(ax+b) - a\cos(ax+b))}{a^2+c^2}$	$e^{cx}\sin(ax+b)$
$\frac{e^{cx}(c\cos(ax+b) + a\sin(ax+b))}{a^2+c^2}$	$e^{cx}\cos(ax+b)$
$\sin(x)\cos(x)$	$\frac{\sin^2(x)}{2}$
$\frac{1}{2}(f(x))^2$	$f'(x)f(x)$
$\sqrt{\pi}$	$\int_{-\infty}^{\infty} e^{-x^2}\,dx$
$\frac{1}{a(n+1)}(ax+b)^{n+1}$	$(ax+b)^n$
$\frac{(ax+b)^{n+2}}{(n+2)a^2} - \frac{b(ax+b)^{n+1}}{(n+1)a^2}$	$x(ax)^n$
$\frac{(ax^p+b)^{n+1}}{ap(n+1)}$	$(ax^p+b)^n x^{p-1}$
$\frac{1}{ap}\ln\lvert ax^p+b\rvert$	$(ax^p+b)^{-1}x^{p-1}$
$\frac{ax}{c} - \frac{ad-bc}{c^2}\ln\lvert cx+d\rvert$	$\frac{ax+b}{cx+d}$
$-x\cos(x) + \sin(x)$	$x\sin(x)$
$x\sin(x) + \cos(x)$	$x\cos(x)$
$\operatorname{arccot}(x)$	$-\frac{1}{1+x^2}$
$\coth(x)$	$1 - \coth^2 x = -\frac{1}{\sinh^2(x)}$
$\operatorname{arcoth}(x)$	$\frac{1}{1-x^2}$

15 List of distributions

Bernoulli ($X \sim \mathrm{Ber}(p)$)

Prerequisites: $\mathcal{X} = \{0,1\}$, $p \in [0,1]$

$p_X(x) = p^x(1-p)^{1-x}$ $\quad \mathbb{E}(X) = p$

$F_X(x) = 1 - p \ (x \in [0,1))$ $\quad \mathrm{Var}(X) = p(1-p)$

Binomial ($X \sim \mathrm{Bin}(n,p)$)

Prerequisites: $\mathcal{X} = \{0,1,\dots,n\}$, $p \in [0,1]$, $n \in \mathbb{N}$

$p_X(x) = \binom{n}{x} p^x (1-p)^{n-x}$ $\quad \mathbb{E}(X) = np$

$F_X(x) = \sum_{k=0}^{x} \binom{n}{k} p^k (1-p)^{n-k}$ $\quad \mathrm{Var}(X) = np(1-p)$

Negative Binomial ($X \sim \mathrm{NB}(r,p)$)

Prerequisites: $\mathcal{X} = \{r, r+1, \dots\}$, $p \in [0,1]$, $r > 0$

$p_X(x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}$ $\quad \mathbb{E}(X) = \frac{r(1-p)}{p}$

$\mathrm{Var}(X) = \frac{r(1-p)}{p^2}$

Geometric ($X \sim \mathrm{Geom}(p)$) **(memoryless)**

Prerequisites: $\mathcal{X} = \mathbb{N}_1$, $p \in (0,1]$

$p_X(x) = p(1-p)^{x-1}$ $\quad \mathbb{E}(X) = \frac{1}{p}$

$F_X(x) = 1 - (1-p)^x$ $\quad \mathrm{Var}(X) = \frac{1-p}{p^2}$

Poisson ($X \sim \mathrm{Poisson}(\lambda)$)

Prerequisites: $\mathcal{X} = \mathbb{N}_0$, $\lambda > 0$

$p_X(x) = e^{-\lambda} \frac{\lambda^x}{x!}$ $\quad \mathbb{E}(X) = \lambda$

$F_X(x) = e^{-\lambda} \sum_{k=0}^{x} \frac{\lambda^k}{k!}$ $\quad \mathrm{Var}(X) = \lambda$

Hypergeometric ($X \sim \mathbf{HG}(N,K,n)$)

Urn with n objects: r type 1, $n-r$ type 2. Draw m w/o replacement. X is type 1 count in sample.

Prereq.: $\mathcal{X} = \{0,1,\dots,\min(m,r)\}$; $n \in \mathbb{N}$; $m, r \in [n]$

$k \in \mathcal{X} : p_X(k) = \binom{r}{k}\binom{n-r}{m-k} / \binom{n}{m}$ $\quad \mathbb{E}(X) = mr/n$

Gamma ($X \sim \mathrm{Ga}(\alpha,\lambda)$)

Prerequisites: $\mathcal{X} = \mathbb{R}_{>0}$, $\alpha > 0$, $\lambda > 0$

$f_X(x) = \frac{1}{\Gamma(\alpha)} \lambda^\alpha x^{\alpha-1} e^{-\lambda x}$ $\quad \mathbb{E}(X) = \frac{\alpha}{\lambda}$ $\quad \mathrm{Var}(X) = \frac{\alpha}{\lambda^2}$

Uniform ($X \sim \mathcal{U}([a,b])$)

Prerequisites: $a < b$, $x \notin [a,b] \implies f_X(x) = 0$

$f_X(x) = \frac{1}{b-a}$ $\quad \mathbb{E}(X) = \frac{a+b}{2}$

$F_X(x) = \frac{x-a}{b-a}$ if $x \in [a,b]$ $\quad \mathrm{Var}(X) = \frac{1}{12}(b-a)^2$

Lemma. $X \sim \mathcal{U}([0,1]) \implies \mathbb{P}(X \le F(t)) = F(t)$

Exponential ($X \sim \mathrm{Exp}(\lambda)$) **(memoryless)**

Prerequisites: $\mathcal{X} = \mathbb{R}_+$, $\lambda > 0$

$f_X(x) = \lambda e^{-\lambda x}$ $\quad \mathbb{E}(X) = \frac{1}{\lambda}$

$F_X(x) = 1 - e^{-\lambda x}$ $\quad \mathrm{Var}(X) = \frac{1}{\lambda^2}$

Normal/Gaussian ($X \sim \mathcal{N}(\mu,\sigma^2)$)

Prerequisites: $\mathcal{X} = \mathbb{R}$, $\sigma^2 \in \mathbb{R}_{>0}$

$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ $\quad \mathbb{E}(X) = \mu$

$F_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} dy$ $\quad \mathrm{Var}(X) = \sigma^2$

Lemma. $X_1 \sim \mathcal{N}(\mu_1,\sigma_1^2), \dots, X_n \sim \mathcal{N}(\mu_n,\sigma_n^2)$ *ind., then* $\alpha + \sum_i \lambda_i X_i \sim \mathcal{N}(\alpha + \lambda_1\mu_1 + \dots + \lambda_n\mu_n, \lambda_1^2\sigma_1^2 + \dots + \lambda_n^2\sigma_n^2)$

Chi-Squared ($X \sim \chi^2(k)$)

Prerequisites: $k \in \mathbb{N}$, $\mathcal{X} = \mathbb{R}_{>0}$ if $k = 1$, otw. $\mathcal{X} = \mathbb{R}_+$

$f_X(x) = \frac{1}{2^{\frac{k}{2}}\Gamma(\frac{k}{2})} x^{\frac{k}{2}-1} e^{-\frac{x}{2}}$ $\quad \mathbb{E}(X) = k$ $\quad \mathrm{Var}(X) = 2k$

Student's t with parameter ν

Prerequisites: $\mathcal{X} = \mathbb{R}$, $\nu > 0$

$f_X(x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\cdot\Gamma(\frac{n}{2})} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}$ $\quad \mathbb{E}(X) = 0$ if $\nu > 1$

$\mathrm{Var}(X)$: $\frac{\nu}{\nu-2}$ for $\nu > 0$, ∞ for $2 < \nu \le 4$, 0 otw.

Lemma. $X \sim \mathcal{N}(0,1)$, $Y \sim \chi^2(k)$ *ind.* $\implies \frac{X}{\sqrt{\frac{1}{k}Y}} \sim t_k$

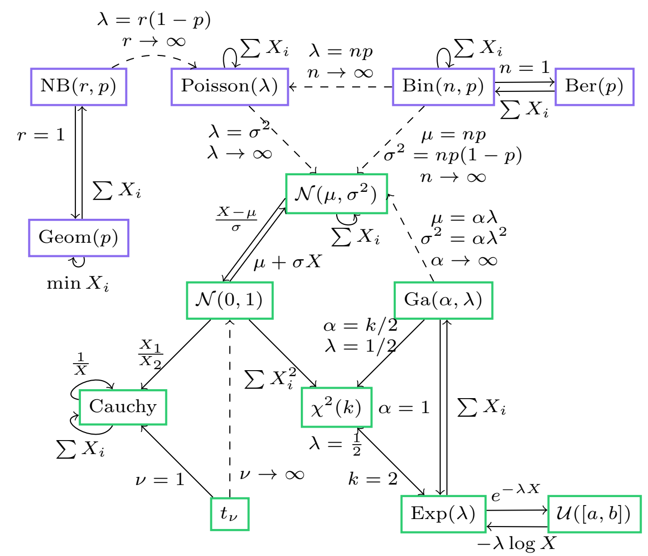
Cauchy

Prerequisites: $\mathcal{X} = \mathbb{R}$

$f_X(x) = \frac{1}{\pi} \cdot \frac{1}{1+x^2}$ $\quad \mathbb{E}(X)$ undef.

$F_X(x) = \frac{1}{2} + \frac{1}{\pi}\arctan(x)$ $\quad \mathrm{Var}(X)$ undef.

Relationship between distributions



Distribution statements ($X_1, \dots, X_n$ **iid.** $\sim \mathcal{N}(\mu,\sigma^2)$)

- $\bar{X}_n \sim \mathcal{N}(\mu, \frac{1}{n}\sigma^2)$ and thus $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$.
- $\frac{n-1}{\sigma^2} S^2 = \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2 \sim \chi^2(n-1)$.
- $\bar{X}_n$ and S^2 are independent.
- $\frac{\bar{X}_n - \mu}{S/\sqrt{n}} = \frac{\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{1}{n-1}\frac{n-1}{\sigma^2}S^2}} \sim t_{n-1}$

Gamma Function

$$\Gamma(\alpha) := \int_0^\infty u^{\alpha-1} e^{-u} \, du \quad (\alpha > 0)$$

Memorylessness

X memoryless $\overset{\text{def}}{\iff} \mathbb{P}(X \ge s + t \mid X \ge s) = \mathbb{P}(X \ge t)$

15.1 Potenzreihen

$\sin(z) = \sum_{n=0}^{\infty} \frac{(-1)^n z^{2n+1}}{(2n+1)!}$ $\quad$ $\sinh(z) = \sum_{n=0}^{\infty} \frac{x^{2n+1}}{(1n+1)!}$

$\cos(z) = \sum_{n=0}^{\infty} \frac{(-1)^n z^{2n}}{(2n)!}$ $\quad$ $\cosh(z) = \sum_{n=0}^{\infty} \frac{x^{2n}}{(2n)!}$

$\exp(z) = \sum_{n=0}^{\infty} \frac{z^n}{n!}$ $\quad$ $\ln(1+x) = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} x^n$

Matrices

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = aei - afh - bdi + bfg + cdh - ceg$$

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \implies A^{-1} = \frac{1}{|A|} \cdot \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

15.1.1 Charakterisierung der Dichte durch $\mathbb{E}$

Sei $\phi : \mathbb{R}^n \to \mathbb{R}$ eine Abbildung und $X_1, \dots, X_n$ ZV mit gemeinsamer Dichte f. Dann lässt sich $\mathbb{E}(Z)$ für die Zufallsvariable $Z = \phi(X_1, \dots, X_n)$ mit

$$\mathbb{E}(Z) = \int \cdots \int_{\mathbb{R}^n} \phi(x_1, \dots, x_n) \cdot f(x_1, \dots, x_n)\, dx_n \dots\, dx_1$$

berechnen.

Dies reicht aber nicht, um die Dichte einer transformierten ZV zu berechnen. Mehrere Zufallsvariablen mit unterschiedlichen Dichten können den gleichen Erwartungswert haben.

Sei $f : \mathbb{R} \to \mathbb{R}_+$ eine Abbildung, sodass $\int_{-\infty}^{\infty} f(z)\, dz = 1$.
Dann sind folgende Aussagen äquivalent

- Z ist stetig mit Dichte f
- Für jede stückweise stetige, beschränkte Abbildung $\psi : \mathbb{R} \to \mathbb{R}$ gilt

$$\mathbb{E}(\psi(Z)) = \int_{-\infty}^{\infty} \psi(z) f(z)\, dz$$

Beispielrechnung

Wir können diese Erkenntnis nutzen, um die Dichte einer transformierten Zufallsvariable zu berechnen.

Seien X und Y zwei Zufallsvariablen mit gemeinsamer Dichtefunktion

$$f(x,y) = \begin{cases} \frac{1}{x^2 y^2} & \text{für } x \geq 1, y \geq 1 \\ 0 & \text{sonst.} \end{cases}$$

Bestimme die Dichtefunktion f_V der Zufallsvariable $V = XY$.

Sei $\psi : \mathbb{R} \to \mathbb{R}$ stückweise stetig und beschränkt. Wir definieren $\phi(x,y) = \psi(xy) = \psi(v)$ und berechnen

$$\begin{aligned} \mathbb{E}(\psi(V)) = \mathbb{E}(\phi(X,Y)) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi(x,y) f(x,y) dx\, dy \\ &= \int_{1}^{\infty} \int_{1}^{\infty} \psi(xy) \frac{1}{x^2 y^2} dx\, dy \\ &\text{Substition } v = xy,\ dv = y\, dx \\ &= \int_{1}^{\infty} \int_{y}^{\infty} \psi(v) \frac{1}{v^2} \frac{dv}{y}\, dy \end{aligned}$$

$$\begin{aligned} A &= \{(v,y) \in \mathbb{R}^2 \mid 1 \leq y < \infty, y \leq v < \infty\} \\ &= \{(v,y) \in \mathbb{R}^2 \mid 1 \leq y \leq v, 1 \leq v < \infty\} \end{aligned}$$

Zeichnung hilft

$$\begin{aligned} &= \int_{1}^{\infty} \int_{1}^{v} \psi(v) \frac{1}{v^2 y}\, dy\, dv \\ &= \int_{1}^{\infty} \psi(v) \frac{\ln(v)}{v^2}\, dv \\ &= \int_{-\infty}^{\infty} \psi(v) \cdot \frac{\ln(v)}{v^2} 1_{v \in [1,\infty)}\, dv \end{aligned}$$

$$\implies f_V(t) = \frac{\ln(v)}{v^2} 1_{v \in [1,\infty)}$$

genereller Transformationssatz
Sei Z ein n-dimensionaler Zufallsvektor mit Dichtefunktion $f_Z : \mathbb{R}^n \to \mathbb{R}_+$ und $\phi : \mathbb{R}^n \to \mathbb{R}^n$ stetig differenzierbar mit stetig differenzierbarer Umkehrabbildung ϕ^{-1}. Dann gilt für die Dichte f_U von $U = \phi(Z)$:

$$f_U(\vec{u}) = f_Z\left(\phi^{-1}(\vec{u})\right) \cdot |\det(J_{\phi^{-1}}(\vec{u}))|$$

Beispielrechnung

Wir haben $Z = (X,Y)$, wobei X, Y unabhängig und exponentialverteilt mit $\lambda > 0$. Berechne die Dichtefunktion f_U von

$$U := \frac{X}{X+Y}$$

Wir definieren ϕ, so dass $(U,Y) = \phi(X,Y)$.

$$\phi(x,y) = \begin{pmatrix} \frac{x}{x+y} \\ y \end{pmatrix} \text{ und } \phi^{-1}(u,y) = \begin{pmatrix} \frac{uy}{1-u} \\ y \end{pmatrix}$$

Check: $\phi^{-1}\left(\frac{x}{x+y}, y\right) = \left(\frac{\frac{x}{x+y} y}{1 - \frac{x}{x+y}}, y\right) = \left(\frac{xy}{x+y-x}, y\right) = (x,y)$.

We then have

$$\begin{aligned} \left|\det\left(J_{\phi^{-1}}(u,y)\right)\right| &= \left|\det \begin{pmatrix} \frac{y}{1-u} + \frac{uy}{(1-u)^2} & 0 \\ \frac{u}{1-u} & 1 \end{pmatrix}\right| \\ &= \left|\frac{y(1-u) + uy}{(1-u)^2}\right| = \left|\frac{y}{(1-u)^2}\right| \end{aligned}$$

Per genereller Transformationssatz gilt

$$\begin{aligned} f_{U,Y}(u,y) &= f_{X,Y}\left(\frac{uy}{1-u}, y\right) \left|\frac{y}{(1-u)^2}\right| \\ &= \begin{cases} \lambda^2 e^{-\lambda\left(\frac{uy}{1-u} + y\right)} \left|\frac{y}{(1-u)^2}\right| & \text{if } \frac{uy}{1-u} \geq 0 \wedge y \geq 0 \\ 0 \cdot \left|\frac{y}{(1-u)^2}\right| & \text{sonst.} \end{cases} \end{aligned}$$

$$\begin{aligned} f_U(u) &= \int_{-\infty}^{\infty} f_{U,Y}(u,y)\, dy \\ &= \int_{0}^{\infty} \frac{\lambda^2}{(1-u)^2} e^{-\frac{\lambda}{1-u} y} y 1_{u \in [0,1]}\, dy \end{aligned}$$

per partielle Integration

$$= 1_{u \in [0,1]}$$