Probability Space

The **sample space** Ω contains all **outcomes** $\omega \in \Omega$ of an experiment.

Sigma-Algebra

$\mathcal{F} \subseteq \mathcal{P}(\Omega)$ defines a sigma-algebra on Ω if and only if

i) $\Omega \in \mathcal{F}$
ii) $A \in \mathcal{F} \implies A^C \in \mathcal{F}$ (complements included)
iii) $A_1, \ldots \in \mathcal{F} \implies \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$ (ORs included)

P1.6 Consequently,

iv) $\emptyset \in \mathcal{F}$
v) $A_1, \ldots \in \mathcal{F} \implies \bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$
vi) $A, B \in \mathcal{F} \implies A \cup B \in \mathcal{F}$
vii) $A, B \in \mathcal{F} \implies A \cap B \in \mathcal{F}$

Probability Measure

For a sigma-algebra $\mathcal{F}$ on Ω, a probability measure on $(\Omega, \mathcal{F})$ is a map of form

$$\begin{aligned} \mathbb{P} : \mathcal{F} &\to [0,1] \\ A &\mapsto \mathbb{P}[A] \end{aligned}$$

that satisfies

i) $\mathbb{P}[\Omega] = 1$
ii) $A = \biguplus_{i=1}^{\infty} A_i \implies \mathbb{P}[A] = \sum_{i=1}^{\infty} \mathbb{P}[A_i]$ (countable additivity)

P1.7 Consequently,

iii) $\mathbb{P}[\emptyset] = 0$
iv) (Additivity) For finitely many pairwise disjoint events $A_1, \ldots, A_k \in \mathcal{F}$, $\mathbb{P}\left[\biguplus_{i \in [k]} A_k\right] = \sum_{i \in [k]} \mathbb{P}[A_i]$
v) For any $A \in \mathcal{F}$, $\mathbb{P}[A^C] = 1 - \mathbb{P}[A]$
vi) For arbitrary $A, B \in \mathcal{F}$, $\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B]$

Probability Space. The tuple $(\Omega, \mathcal{F}, \mathbb{P})$ defines a probability space.

Monotonicity. For $A, B \in \mathcal{F}$, $A \subseteq B \implies \mathbb{P}[A] \leq \mathbb{P}[B]$.

Almost Sure Events. $A \in \mathcal{P}(\Omega)$ is said to occur almost surely (a.s.) if $\exists A' \in \mathcal{F}, A' \subseteq A : \mathbb{P}[A'] = 1$.

Union Bound. Generally,

$$A_1, \ldots \in \mathcal{F} \implies \mathbb{P}\Big[\bigcup_{i=1}^{\infty} A_i\Big] \leq \sum_{i=1}^{\infty} \mathbb{P}[A_i]$$

Continuity Properties. For $A_1, \ldots \in \mathcal{F}$,

i) $A_k \subseteq A_{k+1}$ for all $k \in \mathbb{N}^+ \implies \lim_{n\to\infty} \mathbb{P}[A_n] = \mathbb{P}\Big[\bigcup_{n=1}^{\infty} A_n\Big]$
ii) $A_k \supseteq A_{k+1}$ for all $k \in \mathbb{N}^+ \implies \lim_{n\to\infty} \mathbb{P}[A_n] = \mathbb{P}\Big[\bigcap_{n=1}^{\infty} A_n\Big]$

Conditional Probability

For events $A, B \in \mathcal{F}$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with $\mathbb{P}[B] > 0$,

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}$$

defines the *probability of* A *given* B.

P1.15 If $\mathbb{P}[B] \neq 0$, then the function $A \mapsto \mathbb{P}[A|B]$ is a probability measure on Ω.

Total Probability

For any partition $B_1 \uplus \ldots \uplus B_n = \Omega$ with $\mathbb{P}[B_i] > 0$ for all $i \in [n]$,

$$\forall A \in \mathcal{F} : \mathbb{P}[A] = \sum_{i=1}^{n} \mathbb{P}[A|B_i]\, \mathbb{P}[B_i]$$

Bayes Formula

For any partition $B_1 \uplus \ldots \uplus B_n = \Omega$ with $\mathbb{P}[B_i] > 0$ for all $i \in [n]$ and event A with $\mathbb{P}[A] > 0$,

$$\forall i \in [n] : \mathbb{P}[B_i|A] = \frac{\mathbb{P}[A|B_i]\, \mathbb{P}[B_i]}{\sum_{j=1}^{n} \mathbb{P}[A|B_j]\, \mathbb{P}[B_j]}$$

Applying the total probability formula yields

$$\mathbb{P}[B|A] = \frac{\mathbb{P}[A|B]\, \mathbb{P}[B]}{\mathbb{P}[A]}$$

Event Independence

Two events $A, B \in \mathcal{F}$ are said to be independent if

$$\mathbb{P}[A \cap B] = \mathbb{P}[A]\, \mathbb{P}[B]$$

A collection of events $A_i \in \mathcal{F}$, $i \in I$ is said to be independent if

$$\forall J \subseteq I,\ |J| < \infty : \mathbb{P}\Big[\bigcap_{j \in J} A_j\Big] = \prod_{j \in J} \mathbb{P}[A_j]$$

Remark 1.19 An event $A \in \mathcal{F}$ is independent of

i) every event in $\mathcal{F} \Leftrightarrow \mathbb{P}[A] \in \{0, 1\}$
ii) $B \in \mathcal{F} \Leftrightarrow A$ is independent of B^C

P1.20 Two events $A, B \in \mathcal{F}$ with $\mathbb{P}[A], \mathbb{P}[B] > 0$ are independent $\Leftrightarrow \mathbb{P}[A|B] = \mathbb{P}[A] \Leftrightarrow \mathbb{P}[B|A] = \mathbb{P}[B]$

Random Variables and Distribution Functions

A **random variable** X on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a map of signature $\Omega \to \mathbb{R}$ such that for all $a \in \mathbb{R}$,

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

Its **distribution function** is defined as

$$\begin{aligned} F_X : \mathbb{R} &\to [0,1] \\ a &\mapsto \mathbb{P}[X \leq a] \end{aligned}$$

Theorem 2.4. F_X

i) is nondecreasing
ii) is right continuous, i.e. for all $a \in \mathbb{R}$, $F_X(a) = \lim_{h \downarrow 0} F_X(a+h)$
iii) satisfies $\lim_{a \to -\infty} F(a) = 0$ and $\lim_{a \to \infty} F(a) = 1$

Variable Independence (I)

A set of variables $X_1, \ldots, X_n$ is said to be independent if

$$\forall x_1, \ldots, x_n \in \mathbb{R} : \mathbb{P}[X_1 \leq x_1, \ldots, X_n \leq x_n] = \prod_{i \in [n]} F_{X_i}(x_i)$$

An infinite sequence of variables $X_1, \ldots$ is said to be

i) independent if $X_1, \ldots, X_m$ is independent for all $m \in \mathbb{N}^+$
ii) independent and identically distributed (iid) if it is independent and $\forall i, j \in \mathbb{N}^+ : F_{X_i} = F_{X_j}$

Indicator Function. For an event $A \in \mathcal{F}$, the shorthand $\mathbb{1}_A$ describes its indicator function, i.e.

$$\mathbb{1}_A : \omega \mapsto \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{otherwise} \end{cases}$$

$$F_{\mathbb{1}_A} : a \mapsto \begin{cases} 0 & \text{if } a < 0 \\ 1 - \mathbb{P}[A] & \text{if } 0 \leq a < 1 \\ 1 & \text{otherwise} \end{cases}$$

Discrete Random Variables

X is said to be discrete if a finite or countable set $W \subset \mathbb{R}$ exists such that $X \in W$ a.s.

The **distribution of** X is defined as

$$\begin{aligned} p_X : W &\to [0,1] \\ x &\mapsto \mathbb{P}[X = x] \end{aligned}$$

and satisfies

$$\sum_{x \in W} p_X(x) = 1$$

The distribution function can be derived through

$$F_X(a) = \sum_{y \in W \cap]-\infty, a]} p_X(y)$$

The distribution can be determined by splitting F_X into piecewise constant functions at the locations $x_1, \dots, x_n$ of missing left continuity:

$$\begin{aligned} p_X(x) &= \begin{cases} 0 & \text{for } x < x_1 \\ F_X(x_1) & \text{for } x_1 \le x < x_2 \\ F_X(x_m) - F_X(x_{m-1}) & \text{otherwise} \end{cases} \\ &= F_X(x) - \underbrace{F_X(x-)}_{\lim_{h \downarrow 0} F(x-h)} \end{aligned}$$

Bernoulli

For $p \in [0,1]$, $X \sim \mathrm{Ber}(p)$ defines

$$\begin{aligned} p_X : \{0,1\} &\to [0,1] \\ a &\mapsto \begin{cases} 1-p & \text{for } a = 0 \\ p & \text{for } a = 1 \end{cases} \end{aligned}$$

$$\mathbb{E}[X] = p, \ \mathrm{Var}[X] = p(1-p)$$

Existence Theorem of Kolmogorov. There exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with an infinite sequence of random variables $X_1, \dots$ that are iid as $\mathrm{Ber}(\frac{1}{2})$.

Binomial

For $n \in \mathbb{N}_0$ and $p \in [0,1]$, $X \sim \mathrm{Bin}(n,p)$ defines

$$\begin{aligned} p_X : \{0, \dots, n\} &\to [0,1] \\ a &\mapsto \binom{n}{a} p^a (1-p)^{n-a} \end{aligned}$$

$$\mathbb{E}[X] = np, \ \mathrm{Var}[X] = np(1-p)$$

P3.13 Sum of Bernoulli. For iid $X_1, \dots, X_n \sim \mathrm{Ber}(p)$, $\sum_{i=1}^n X_i \sim \mathrm{Bin}(n,p)$.

Remark 3.14 Binomial Sum. For independent $X \sim \mathrm{Bin}(n,p)$ and $Y \sim \mathrm{Bin}(m,p)$, $X + Y \sim \mathrm{Bin}(n+m, p)$.

Geometric

For $p \in]0,1]$, $X \sim \mathrm{Geom}(p)$ defines

$$\begin{aligned} p_X : \mathbb{N}^+ &\to [0,1] \\ a &\mapsto (1-p)^{a-1} p \end{aligned}$$

$$\mathbb{E}[X] = {}^1\!/\!_p, \ \mathrm{Var}[X] = \tfrac{1-p}{p^2}$$

P3.18 Tries to success. For an infinite sequence of iid $X_1, \dots \sim \mathrm{Ber}(p)$, $\min\{n \in \mathbb{N}^+ \mid X_n = 1\} \sim \mathrm{Geom}(p)$.

P3.20 Absence of memory. For $n \ge 0$, $k \ge 1$, and $X \sim \mathrm{Geom}(p)$,

$$\mathbb{P}[X \ge n + k \mid X > n] = \mathbb{P}[X \ge k]$$

Negative Binomial

For $r \in \mathbb{N}^+$ and $p \in]0,1]$, $X \sim \mathrm{NB}(r,p)$ defines

$$\begin{aligned} p_X : \mathbb{N}^+ \setminus [r-1] &\to [0,1] \\ a &\mapsto \tbinom{a-1}{r-1} (1-p)^{a-r} p^r \end{aligned}$$

$$\mathbb{E}[X] = {}^r\!/\!_p, \ \mathrm{Var}[X] = \tfrac{r(1-p)}{p^2}$$

Sum of Geometric. For iid $X_1, \dots, X_n \sim \mathrm{Geom}(s)$, $\sum_{i=1}^n X_i \sim \mathrm{NB}(n,p)$.

Poisson

For $\lambda \in]0, \infty[$, $X \sim \mathrm{Poisson}(\lambda)$ defines

$$\begin{aligned} p_X : \mathbb{N}_0 &\to [0,1] \\ a &\mapsto \frac{\lambda^a}{a!} e^{-\lambda} \end{aligned}$$

$$\mathbb{E}[X] = \lambda, \ \mathrm{Var}[X] = \lambda$$

P3.23 Approximating Binomial. For $\lambda > 0$ and $X_n \sim \mathrm{Bin}(n, \frac{\lambda}{n})$, $N \sim \mathrm{Poisson}(\lambda)$ satisfies $p_N = \lim_{n \to \infty} p_{X_n}$.

Continuous Random Variables

X is said to be continuous if its distribution function can be written as $F_X(a) = \int_{-\infty}^a f_X(x)\,dx$ for all $a \in \mathbb{R}$ with $f_X \ge 0$. The density of X is the above function of signature

$$f_x : \mathbb{R} \to [0, \infty[$$

and satisfies

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

The distribution function can be constructed from its definition.

The density can often be determined by splitting F_X into piecewise $\mathcal{C}^1$-functions at the locations $x_1, \dots, x_n$ of missing left-continuity:

$$f_X = \begin{cases} \frac{\partial f_X}{\partial x} & \text{for } x \in]a,b[\text{ where} \\ & (a,b) \in \{(-\infty, x_1), (x_n, \infty)\} \\ & \cup \{(x_i, x_{i+1}) \mid i \in [n-1]\} \\ \text{arbitrary} \in \mathbb{R} & \text{otherwise} \end{cases}$$

Uniform

For $a, b \in \mathbb{R}$, $a < b$, $X \sim \mathcal{U}([a,b])$ defines

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{for } x \in [a,b] \\ 0 & \text{otherwise} \end{cases}$$

$$F_X(x) = \begin{cases} 0 & \text{for } x < a \\ \frac{x-a}{b-a} & \text{for } x \in [a,b] \\ 1 & \text{for } x > b \end{cases}$$

$$\mathbb{E}[X] = \tfrac{a+b}{2}, \ \mathrm{Var}[X] = \tfrac{(b-a)^2}{12}$$

Construction from Bernoulli. Using iid $X_1, \dots \sim \mathrm{Ber}(1/2)$, one can construct $\sum_{i=1}^{\infty} 2^{-i} X_i \sim \mathcal{U}([0,1])$.

Constructing Random Variables

For the target distribution function F, construct the **generalized inverse**

$$\begin{aligned} F^{-1} : \ ]0,1[\ &\to \mathbb{R} \\ \alpha &\mapsto \inf\{x \in \mathbb{R} \mid F(x) \ge \alpha\} \end{aligned}$$

Then, Theorem 2.13 guarantees that $X := F^{-1}(\mathcal{U}([0,1]))$ is a random variable with $F_X = F$

Exponential

For $\lambda \in]0, \infty[$, $X \sim \mathrm{Exp}(\lambda)$ defines

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{for } x \ge 0 \\ 0 & \text{otherwise} \end{cases}$$

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x} & \text{for } x \ge 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbb{E}[X] = {}^1\!/\!_\lambda, \ \mathrm{Var}[X] = {}^1\!/\!_{\lambda^2}$$

Absence of memory. For $t, s \ge 0$ and $X \sim \mathrm{Exp}(\lambda)$,

$$\mathbb{P}[X > t + s \mid X > t] = \mathbb{P}[T > s]$$

Normal

For mean $\mu \in \mathbb{R}$ and variance $\sigma^2 \in]0,\infty[$, $X \sim \mathcal{N}(\mu, \sigma^2)$ defines

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$F_X = \Phi\left(\frac{x-m}{\sigma}\right)$$

$$\text{where } \Phi(x) := \tfrac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt$$

$$\mathbb{E}[X] = \mu, \ \operatorname{Var}[X] = \sigma^2$$

Combining Normal. For independent $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$, ..., $X_n \sim \mathcal{N}(\mu_n, \sigma_n^2)$, $\lambda_1, \dots, \lambda_n \in \mathbb{R}$, and $m \in \mathbb{R}$,

$$m + \sum_{i=1}^{n} \lambda_i X_i \sim \mathcal{N}\Big(m + \sum_{i=1}^{n} \lambda_i \mu_i, \sum_{i=1}^{n} \lambda_i^2 \sigma_i^2\Big)$$

Chi^2

For degree of freedom $k \in \mathbb{N}^+$, $X \sim \chi^2(k)$ defines

$$f_X(x) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1} e^{-x/2}$$

$$\mathbb{E}[X] = k, \ \operatorname{Var}[X] = 2k$$

Sum of Standard Normal. For iid $X_1, \dots, X_n \sim \mathcal{N}(0,1)$, $\sum_{i=1}^{n} X_i^2 \sim \chi^2(n)$

Student's t

For degree of freedom $\nu \in]0,\infty[$, $X \sim t(\nu)$ defines

$$f_X(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

$$\mathbb{E}[X] = \begin{cases} 0 & \text{for } \nu > 0 \\ \text{undef.} & \text{otherwise} \end{cases}$$

$$\operatorname{Var}[X] = \begin{cases} \frac{\nu}{\nu-2} & \text{for } \nu > 2 \\ \text{undef. } / \infty & \text{otherwise} \end{cases}$$

P2.6 For independent $X \sim \mathcal{N}(0,1)$ and $Y \sim \chi^2(m)$, $\frac{X}{\sqrt{\frac{1}{m}Y}} \sim t(m)$

Normal is t's limit. For $(Z_n)_{n\geq 1} \sim t(n)$ and $Z \sim \mathcal{N}(0,1)$, $\lim_{n\to\infty} F_{Z_n} = F_Z$.

Expectation

The **expected value** of a nonnegative random variable $Y : \Omega \to]0,\infty[$ is defined as

$$\mathbb{E}[Y] = \int_0^{\infty} (1 - F_Y(x))\, dx$$

More generally, for $X : \Omega \to \mathbb{R}$ with $\mathbb{E}[|X|] < \infty$, one defines

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{1}_{X\geq 0} X] - \mathbb{E}[-\mathbb{1}_{X\leq 0} X]$$

Expectation of a Discrete RV

For an arbitrary function $\phi : \mathbb{R} \to \mathbb{R}$ and discrete random variable $X : \Omega \to \mathbb{R}$ with $X \in W \subset \mathbb{R}$ a.s.,

$$\mathbb{E}[\phi(X)] = \sum_{x\in W} \phi(x)\, p_X(x)$$

if defined.

Expectation of a Continuous RV

For a function $\phi : \mathbb{R} \to \mathbb{R}$ and continuous random variable $X : \Omega \to \mathbb{R}$ that guarantee that $\phi(X)$ is a random variable too,

$$\mathbb{E}[\phi(X)] = \int_{-\infty}^{\infty} \phi(x)\, f_X(x)\, dx$$

if defined.

Theorem 4.10 Linearity For arbitrary random variables $X, Y : \Omega \to \mathbb{R}$ and $\lambda \in \mathbb{R}$, if defined,

i) $\mathbb{E}[\lambda X] = \lambda\, \mathbb{E}[X]$
ii) $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$

Theorem 4.13

$$X, Y \text{ independent} \implies \mathbb{E}[XY] = \mathbb{E}[X]\, \mathbb{E}[Y]$$

Tailsum Formulas. For any random variable X with $X \geq 0$ a.s.,

$$\mathbb{E}[X] = \textstyle\int_0^{\infty} \mathbb{P}[X > x]\, dx$$

If X is discrete with $W = \mathbb{N}_0$, then

$$\mathbb{E}[X] = \textstyle\sum_{i=1}^{\infty} \mathbb{P}[X \geq n]$$

Theorem 4.17 Random variables $X_1, \dots, X_n$ are independent $\Leftrightarrow$ For all bounded piecewise $\mathcal{C}^0$ functions $\phi_1, \dots, \phi_n : \mathbb{R} \to \mathbb{R}$

$$\mathbb{E}\big[\textstyle\prod_{i=1}^{n} \phi_n(X_n)\big] = \prod_{i=1}^{n} \mathbb{E}[\phi_n(X_n)]$$

P4.19 For random variables X, Y,

$$X \leq Y \text{ a.s.} \implies \mathbb{E}[X] \leq \mathbb{E}[Y]$$

Markov's Inequality

For any random variable X with $X \geq 0$ a.s.,

$$\forall a \in]0,\infty[\ B: \ \mathbb{P}[X \geq a] \leq \frac{\mathbb{E}[X]}{a}$$

Jensen's Inequality

For any random variable X and convex function $\phi : \mathbb{R} \to \mathbb{R}$, if defined,

$$\phi(\mathbb{E}[X]) \leq \mathbb{E}[\phi(X)]$$

More Inequalities. Consequently,

i) $|\mathbb{E}[X]| \leq \mathbb{E}[|X|]$
ii) $\mathbb{E}[|X|] \leq \sqrt{\mathbb{E}[X^2]}$

Variance

The **variance** of a random variable X with $\mathbb{E}[X^2] < \infty$ is defined as

$$\operatorname{Var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2] = \sigma_X^2$$

The square root σ_X is its **standard deviation**.

Chebyshev's Inequality

For any random variable X satisfying $\mathbb{E}[X^2] < \infty$,

$$\forall a \in [0,\infty[: \mathbb{P}[|X - \mathbb{E}[X]| \geq a] \leq \frac{\operatorname{Var}[X]}{a^2}$$

P4.25 For pairwise independent random variables $X, X_1, \dots, X_n$ with defined variances,

i) $\operatorname{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$
ii) $\forall \lambda \in \mathbb{R} : \operatorname{Var}[\lambda X] = \lambda^2 \operatorname{Var}[X]$
iii) $\operatorname{Var}\big[\sum_{i=1}^{n} X_i\big] = \sum_{i=1}^{n} \operatorname{Var}[X_i]$

Covariance. For random variables X, Y with with $\mathbb{E}[X^2], \mathbb{E}[Y^2] < \infty$, the covariance is defined as $\operatorname{Cov}[X, Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$. To disprove independence, one might make use of the fact that

$$X, Y \text{ independent} \implies \operatorname{Cov}(X, Y) = 0$$

Joint Distributions

The **joint distribution of discrete random variables** $X_1, \dots, X_n$ with $X_i \in W_i$ a.s. for $i \in [n]$ and $W = \bigtimes_{i=1}^{n} W_i$ is defined as

$$p : W \to [0, 1]$$
$$x_{1:n} \mapsto \mathbb{P}[X_{i:n} = x_{1:n}]$$

and satisfies $\sum_{x_{1:n}\in W} p(x_{1:n}) = 1$.

The **joint density of continuous random variables** $X_1, \dots, X_n$ is (if existent) the function $f : \mathbb{R}^n \to [0,\infty[$ such that

$$\mathbb{P}[X_1 \leq a_1, \dots, X_n \leq a_n]$$
$$= \int_{-\infty}^{a_1} \dots \int_{-\infty}^{a_n} f(x_{1:n})\, dx_n \dots dx_1$$

holds for all $a_{1:n} \in \mathbb{R}^n$. For $a_{1:n} = \infty$, the above expression evaluates to 1.

Variable Independence (II)

A set of discrete / continuous random variables $X_1, \ldots, X_n$ with a joint distribution p / density f are independent $\iff$ $p(x_{1:n}) = \prod_{i=1}^n p_{X_i}(x_i)$ / $f(x_{1:n}) = \prod_{i=1}^n f_{X_i}(x_i)$ for all $x_{i:n} \in W / \mathbb{R}^n$

Asymptotic Behavior

Law of Large Numbers

For iid $X_1, \ldots, X_n$ with $\mathbb{E}[|X_1|] < \infty$,

$$\lim_{n\to\infty} \underbrace{\frac{\sum_{i=1}^n X_i}{n}}_{\text{empirical average}} = \mathbb{E}[X_1]$$

Convergence in Distribution. A sequence of random variables $X_1, \ldots$ is said to satisfy "$X_n \overset{\text{approx}}{\approx} X$ as $n \to \infty$" if

$$\forall x \in \mathbb{R} : \lim_{n\to\infty} \mathbb{P}[X_n \leq x] = \mathbb{P}[X \leq x]$$

Central Limit Theorem

For iid $X_1, \ldots, X_n$ with $\mathbb{E}[X_1^2] < \infty$, the random variable

$$Z_n = \frac{\sum_{i=1}^n X_i - \mathbb{E}[X_1]\, n}{\sqrt{\mathrm{Var}[X_1]\, n}}$$

satisfies

$$Z_n \overset{\text{approx}}{\approx} Z \text{ as } n \to \infty$$

with $Z \sim \mathcal{N}(0, 1)$.

Consequently,

$$\forall a \in \mathbb{R} : \lim_{n\to\infty} \mathbb{P}\big[Z_n \leq a\big] = \Phi(a)$$

Inductive Statistics

→ On a fixed **base space** $(\Omega, \mathcal{F})$, observe the

→ **realization** $x_1, \ldots, x_n$ (i.e. empirical data points) of

→ **samples** $X_1, \ldots, X_n$ to determine an optimal

→ **model parameter** $\theta \in \Theta$ and

→ **confidence interval** $I \subset \Theta$ for chosing from a suitable

→ **model family** $\mathbb{P}_\theta$. Check the validity of one's assumptions using

→ **statistical tests**.

Estimator

A random variable of form

$$T = t(X_{1:n}), \quad t : \mathbb{R}^n \to \mathbb{R}$$

attempting to estimate θ is referred to as an estimator.

Bias

An estimator T under the model $\mathbb{P}_\theta$ shows a

- bias of $\mathbb{E}_\theta[T] - \theta$
- mean squared error of $\mathrm{MSE}_\theta[T] = \mathbb{E}_\theta[(T - \theta)^2]$

It is said to be **unbiased** ("erwartungstreu") if $\forall \theta \in \Theta : \mathbb{E}_\theta[T] = \theta$.

MSE Decomposition. One can show that $\mathrm{MSE}_\theta[T] = \mathrm{Var}_\theta[T] + (\mathbb{E}_\theta[T] - \theta)^2$.

Likelihood Function. For samples $X_{1:n}$, define

$$L(x_{1:n}; \theta) = \begin{cases} p_{X_{1:n}}(x_{1:n}; \theta) \\ \quad \text{if } \forall i \in [n] : X_i \text{ discrete} \\ f_{X_{1:n}}(x_{1:n}; \theta) \quad \text{otherwise} \end{cases}$$

Working with $\log L$ is often more convenient since for iid variables, products turn into sums.

Maximum Likelihood Estimator

A random variable

$$T_{ML} = t_{ML}(X_{1:n})$$

with

$$\begin{aligned} t_{ML} : \mathbb{R}^n &\to \mathbb{R} \\ x_{1:n} &\mapsto \arg\max_{\theta \in \Theta} L(x_{1:n}; \theta) \end{aligned}$$

is a maximum likelihood estimator for θ using samples $X_{1:n}$.

Empirical Mean. An ML-Estimator for θ assuming iid $X_{1:n} \sim \mathrm{Ber}(\theta)$ is

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

Empirical Variance. An ML-Estimator for σ_θ^2 assuming iid $X_{1:n} \sim \mathcal{N}(\mu_\theta, \sigma_\theta^2)$ is

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

P2.7 For iid $X_{1:n} \sim \mathcal{N}(\mu, \sigma^2)$, $\bar{X}_n$ and S^2 are independent.

Confidence Interval

A confidence interval $I = [A, B]$ for θ of level $1 - \alpha$, $\alpha \in \mathbb{R}$ satisfies

$$\forall \theta \in \Theta : \mathbb{P}_\theta[A \leq \theta \leq B] \geq 1 - \alpha$$

with derived random variables $A = a(X_{1:n})$ and $B = b(X_{1:n})$ using $a, b : \mathbb{R}^n \to \mathbb{R}$.

Statistical Tests

A statistical test is a pair (T, K) of a derived **test statistic** $T = t(X_{1:n})$, $t : \mathbb{R}^n \to \mathbb{R}$ and a **critical region** ("Verwerfungsbereich") $K \subseteq \mathbb{R}$.

One rejects the **null hypothesis** H_0 : "$\theta \in \Theta_0 \subseteq \Theta$" if and only if $T(\omega) \in K$.

Since rejection is harder than acceptance, one should choose the negation of what one intends to show for H_0 and an similar **alternative** H_A : "$\theta \in \Theta_A \subseteq \Theta \setminus \Theta_0$".

Significance Level

A test (T, K) possesses a significance level $\alpha \in [0, 1]$ if the probability of **Type 1 errors**, i.e. erroneously rejecting H_0, satisfies

$$\forall \theta \in \Theta_0 : \mathbb{P}_\theta[T \in K] \leq \alpha$$

Power

A test (T, K)'s power is defined as the function

$$\begin{aligned} \beta : \Theta_A &\to [0, 1] \\ \theta &\mapsto \mathbb{P}_\theta[T \in K] \end{aligned}$$

It is *inversely* tied to the **Type 2 error**, i.e. erroneously *not* rejecting H_0.

Neyman-Pearson Construction

A good test can usually be constructed systematically:

For T, choose the generalized **Likelihood-Quotient**

$$R(x_{1:n}) = \frac{\sup_{\theta_A \in \Theta_A} L(x_{1:n}; \theta_A)}{\sup_{\theta_0 \in \Theta_0} L(x_{1:n}; \theta_0)}$$

For K, choose an interval $]c_0, \infty[$ with c_0 set to suit α.

p-value

For a simple $H_0 : \theta = \theta_0$ and ordered family of tests $(T, (K_t)_{t \geq 1})$, i.e. $s \leq t \Rightarrow K_s \supseteq K_t$, the p-value is the random variable $G(T)$ derived using

$$\begin{aligned} G : \ ]0, \infty[&\to [0, 1] \\ t &\mapsto \mathbb{P}_{\theta_0}[T \in K_t] \end{aligned}$$

All tests with $\alpha > G(T(\omega))$ reject H_0.

Quantile

The lower α-quantile of a random variable X is defined as

$$\inf\{x \mid \mathbb{P}[X \leq x] \geq \alpha\}$$

For $T \sim t(n)$, one often denotes the lower α-quantile as $t_{n,\alpha}$.

For $Z \sim \mathcal{N}(0,1)$, one often denotes the lower α-quantile as $z_\alpha := t_{\infty,\alpha} = \Phi^{-1}(\alpha)$.

Symmetry. For $n \in \mathbb{N}_\infty^+$ and $\alpha \in [0,1]$, $t_{n,\alpha} = 1 - t_{n,1-\alpha}$.

Guides

Approximate Confidence Intervals

Bernoulli – Interval for p

Assuming iid $X_1, \dots, X_n \sim \mathrm{Ber}(\theta)$, estimate $\theta \in [0,1]$ using $\bar{X}_n$. For $n \to \infty$, the Central Limit Theorem guarantees that

$$\frac{n(\bar{X}_n - \theta)}{\sqrt{\theta(1-\theta)n}} \overset{\text{approx}}{\approx} Z$$

where $Z \sim \mathcal{N}(0,1)$. Consequently, one can approximate

$$\mathbb{P}[\bar{X}_n - c_\theta \le \theta \le \bar{X}_n + c_\theta] \ge 1 - \alpha$$

for

$$c_\theta = \frac{z_{1-\alpha/2}\sqrt{\theta(1-\theta)n}}{n} \le \frac{z_{1-\alpha/2}}{2\sqrt{n}} = \tilde{c}$$

which leads to

$$I = [\bar{X}_n - \tilde{c}, X_n + \tilde{c}]$$

Geometric – Interval for p

Assuming iid $X_1, \dots, X_n \sim \mathrm{Geom}(\theta)$, estimate $\theta \in [m, 1]$, $m \in [0, 1[$, using

$$T_{ML} = \frac{n}{\sum_{i=1}^{n} X_i}$$

For $n \to \infty$, the Central Limit Theorem guarantees that

$$\frac{n(1/T_{ML} - 1/\theta)}{\sqrt{\frac{1-\theta}{\theta^2}n}} \overset{\text{approx}}{\approx} Z$$

where $Z \sim \mathcal{N}(0,1)$. Consequently, one can approximate

$$\mathbb{P}\left[(T_{ML}^{-1} + c_\theta)^{-1} \le \theta \le (T_{ML}^{-1} - c_\theta)^{-1}\right] \ge 1 - \alpha$$

for

$$c_\theta = \frac{z_{1-\alpha/2}}{\sqrt{n}} \frac{\sqrt{1-\theta}}{\theta} \le \frac{z_{1-\alpha/2}}{\sqrt{n}} \frac{\sqrt{1-m}}{m} = \tilde{c}$$

which leads to

$$I = \left[(T_{ML}^{-1} + \tilde{c})^{-1}, (T_{ML}^{-1} - \tilde{c})^{-1}\right]$$

Tests

Normal – Test for μ

Assuming iid $X_1, \dots, X_n \sim \mathcal{N}(\theta, \sigma^2)$, test $H_0 : \theta = \theta_0$ for some $\theta_0 \in \mathbb{R}$ with $H_A : \theta > \theta_0$ using

$$T = \frac{\bar{X}_n - \theta_0}{\sigma/\sqrt{n}}$$

If H_0 holds, $T \sim \mathcal{N}(0,1)$. Build

$$K =]z_{1-\alpha}, \infty[$$

Choosing $H_A : \theta < \theta_0$ requires

$$K =]-\infty, \underbrace{-z_{1-\alpha}}_{z_\alpha}[$$

Choosing $H_A : \theta \neq \theta_0$ requires

$$K =]-\infty, -z_{1-\alpha/2}[\;\cup\;]z_{1-\alpha/2}, \infty[$$

Normal – Test for μ without σ²

Assuming iid $X_1, \dots, X_n \sim \mathcal{N}(\mu_\theta, \sigma_\theta^2)$, test $H_0 : \mu_\theta = \mu_0$ for some $\theta \in \mathbb{R}^2$ with $H_A : \mu_\theta > \mu_0$ using

$$T = \frac{\bar{X}_n - \mu_0}{S/\sqrt{n}}$$

If H_0 holds, $T \sim t(n-1)$. Build

$$K =]t_{n-1,1-\alpha}, \infty[$$

Choosing $H_A : \mu_\theta < \mu_0$ requires

$$K =]-\infty, \underbrace{-t_{n-1,1-\alpha}}_{t_{n-1,\alpha}}[$$

Choosing $H_A : \theta \neq \theta_0$ requires

$$K =]-\infty, -t_{n-1,1-\alpha/2}[\;\cup\;]t_{n-1,1-\alpha/2}, \infty[$$

Bernoulli – Test for p

Assuming iid $X_1, \dots, X_n \sim \mathrm{Ber}(\theta)$, test $H_0 : \theta = \theta_0$ for some $\theta_0 \in [0,1]$ with $H_A : \theta > \theta_0$ using

$$T = \sum_{i=1}^{n} X_i$$

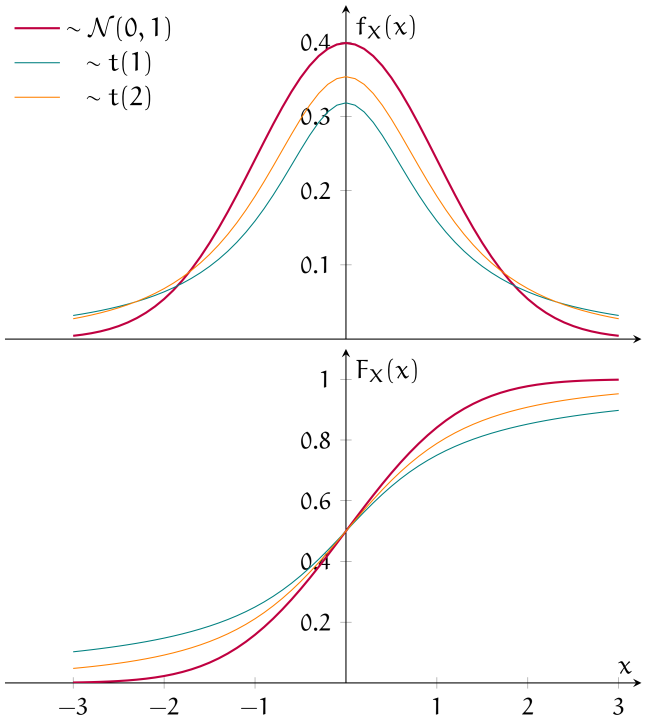
If H_0 holds, $T \sim \mathrm{Bin}(n, \theta_0)$. Build

$$K = \{k_0 + 1, \dots, n\} \quad \text{using}$$

$$k_0 = \min\{k \in \mathbb{N}_0 \mid \mathbb{P}[T \le k] > 1 - \alpha\}$$

Graphs

Normal & Student's t



Diverses

Summensätze.

i) $\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$

ii) $\sum_{i=0}^{n} k^i = \frac{1-k^{n+1}}{1-k}$

iii) $\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}$

Binomialsatz. $\forall x, y \in \mathbb{C}\ \ \forall n \geq 1:$

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

Binomialkoeffizient. Mit "n choose k" bezeichnet man

$$\binom{n}{k} := \frac{n!}{(n-k)!\,k!}$$

Differentiationsregeln. $\forall f, g \in \mathbb{R}^D$, $D \subseteq \mathbb{R}$, in x_0 differenzierbar,

i) $(f+g)'(x_0) = f'(x_0) + g'(x_0)$

ii) $(fg)'(x_0) = f'(x_0)g(x_0) + f(x_0)g'(x_0)$

iii) $\left(f/g\right)'(x_0) = \frac{f'(x_0)g(x_0) - f(x_0)g'(x_0)}{g^2(x_0)}$, falls $g(x_0) \neq 0$.

Partielle Integration. Für stetig differenzierbare $f, g \in \mathbb{R}^{[a,b]}$ gilt

$$\int_a^b (fg')(x)\,dx = \underbrace{\Big[(fg)(x)\Big]_a^b}_{(fg)(b)-(fg)(a)} - \int_a^b (f'g)(x)\,dx$$

Substitution. Für stetige $f \in \mathbb{R}^I$ auf dem Intervall $I \subseteq \mathbb{R}$ und stetig differenzierbarer $\phi \in D^{[a,b]}$ auf $D \subseteq I$ gilt

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

Angewandt heisst dies z.B.

i) $\int_{a+c}^{b+c} f(x)\,dx = \int_a^b f(t+c)\,dt$

ii) $\int_a^b f(ct)\,dt = \frac{1}{c}\int_{ac}^{bc} f(x)\,dx$

Variablensubstitution. Für eine geeignete $\phi \in C^1(U, \mathbb{R}^2), U \subset \mathbb{R}^2$ offen, mit $\phi(B) = A$ und $f \in C^0(A, \mathbb{R})$, wobei A, B kompakt, $\partial A, \partial B$ Nullmengen, $\phi|_{B\setminus N \to A}$ injektiv für Nullmengen $N \subset B$ gilt

$$\int_A f(s)\,ds = \int_B (f \circ \phi)(s)\,|\det J_\phi(s)|\,ds$$

Polarkoordinaten. Verwendet man

$$\phi:\ ]0,\infty[\, \times\,]0, 2\pi[\ \to \mathbb{R}^2$$
$$\begin{pmatrix} r \\ \theta \end{pmatrix} \mapsto \begin{pmatrix} r\cos(\theta) \\ r\sin(\theta) \end{pmatrix}$$

ist wegen $\det(J_\phi) = r$ beispielsweise

$$\int_{x^2+y^2<R^2} f(s)\,ds = \int_0^{2\pi}\int_0^R f\begin{pmatrix} r\cos(\theta) \\ r\sin(\theta) \end{pmatrix} r\,dr\,d\theta$$

Sphärische Koordinaten. Verwendet man

$$\phi: \mathbb{S} \times\,]0, \pi[\ \to \mathbb{R}^3$$
$$\begin{pmatrix} r \\ \theta \\ \varphi \end{pmatrix} \mapsto \begin{pmatrix} r\cos(\theta)\sin(\varphi) \\ r\sin(\theta)\cos(\varphi) \\ r\cos(\varphi) \end{pmatrix}$$

ist $\det(J_\phi) = r^2\sin(\theta)$.

deg	x	$\sin(x)$	$\cos(x)$	$\tan(x)$
$0°$	0	0	1	0
$30°$	$\frac{\pi}{6}$	$\frac{1}{2}$	$\frac{\sqrt{3}}{2}$	$\frac{\sqrt{3}}{3}$
$45°$	$\frac{\pi}{4}$	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{2}}{2}$	1
$60°$	$\frac{\pi}{3}$	$\frac{\sqrt{3}}{2}$	$\frac{1}{2}$	$\sqrt{3}$
$90°$	$\frac{\pi}{2}$	1	0	

Die trigonometrischen Funktionen

S3.41 $\sin: \mathbb{C} \to \mathbb{C}$

$$z \mapsto \sum_{n=0}^{\infty} \frac{(-1)^n z^{2n+1}}{(2n+1)!} \quad \text{und}$$

$\cos: \mathbb{C} \to \mathbb{C}$

$$z \mapsto \sum_{n=0}^{\infty} \frac{(-1)^n z^{2n}}{(2n)!}$$

sind in $\mathbb{R}$ stetige Funktionen. Es gelten

S3.42 $\forall z, w \in \mathbb{C},\ \forall x \in \mathbb{R}$,

i) $\exp(iz) = \cos(z) + i\sin(z)$

ii) $\sin(-z) = -\sin(z)$, $\cos(-z) = \cos(z)$

iii) $\sin(z) = \frac{\exp(iz) - \exp(-iz)}{2i}$, $\cos(z) = \frac{\exp(iz) + \exp(-iz)}{2}$

iv) $\sin(z+w) = \sin(z)\cos(w) + \cos(z)\sin(w)$, $\cos(z+w) = \cos(z)\cos(w) - \sin(z)\sin(w)$

v) $\cos^2(z) + \sin^2(z) = 1$

vi) $\sin(2z) = 2\sin(z)\cos(z)$

vii) $\cos(2z) = \cos^2(z) - \sin^2(z)$

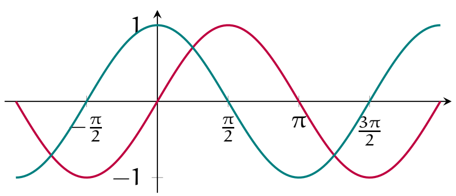
viii) $\sin\left(x + \pi/2\right) = \cos(x)$, $\cos\left(x + \pi/2\right) = -\sin(x)$

ix) $\sin(x+\pi) = -\sin(x)$, $\sin(x + 2\pi) = \sin(x)$

x) $\cos(x+\pi) = -\cos(x)$, $\cos(x+2\pi) = \cos(x)$

xi) $\{k\pi : k \in \mathbb{Z}\}$ Nullstellen von sin

xii) $\{\pi/2 + k\pi : k \in \mathbb{Z}\}$ Nullstellen von cos



Tangens. Für $z \notin \frac{\pi}{2} + \pi \cdot \mathbb{Z}$ ist $\tan(z) := \frac{\sin(z)}{\cos(z)}$ und für $z \notin \pi \cdot \mathbb{Z}$ ist $\cot(z) := \frac{\cos(z)}{\sin(z)}$ definiert.

Die reelle Exponentialfunktion

S3.24 $\exp: \mathbb{R} \to\,]0, \infty[$

$$x \mapsto \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

ist streng monoton wachsend, stetig und surjektiv (also bijektiv). Es gelten

i) $\exp(x+y) = \exp(x)\exp(y)$

ii) $\exp(x) > 0\ \ \forall x \in \mathbb{R}$

iii) $\exp(x) \geq 1 + x\ \ \forall x \in \mathbb{R}$

iv) $\exp(i\pi) = -1$, $\exp(2i\pi) = 1$

Der natürliche Logarithmus

K3.28 Die Umkehrabbildung von exp, $\ln:\ ]0,\infty[\ \to \mathbb{R}$, ist streng monoton wachsend, stetig und bijektiv. Ferner gilt

$$\ln(ab) = \ln(a) + \ln(b)\ \ \forall a, b \in\,]0, \infty[$$