Probability Theory

November 6, 2023

Contents

1 Random variables

1.1 Independence

A family $(A_i)_{i\in I} \subset \mathcal{A}$ of measurable sets on a **probability space** $(\Omega;\mathcal{A};P)$ is **independent**, if $P\left(\bigcap_{i\in F} A_i\right) = \prod_{i\in F} P(A_i)$ for every finite subset $F \subset I$. A family $(\mathcal{E}_i)_{i\in I}$ of set systems $\mathcal{E}_i \subset \mathcal{A}$ with $i \in I$ is independent if the families $\left(A_{i_f}\right)_{i_f\in F}$ are independent with $A_{i_f} \in \mathcal{E}_{i_f}$ for $i_f \in F$ and every nonempty and finite subset $F \subset I$. For two independent systems $\mathcal{E},\mathcal{D} \subset \mathcal{A}$ on a probability space$(\Omega;\mathcal{A};P)$ the corresponding **Dynkin-systems** $\delta(\mathcal{E})$ and $\delta(\mathcal{D})$ are independent too since the family $\mathcal{I}(\mathcal{D}) := \{A \in \mathcal{A} : P(A\cap D) = P(A)\cdot P(D)\,\forall D \in \mathcal{D}\}$ already is a Dynkin-system: Obviously we have $\Omega \in \mathcal{I}(\mathcal{D})$ and for $A \in \mathcal{I}(\mathcal{D})$ and $D \in \mathcal{D}$ we have $P((\Omega\setminus A)\cap D) = P(D\setminus(A\cap D)) = P(D) - P(A\cap D) = P(D) - P(A)\cdot P(D) = P(D)\cdot(1-P(A)) = P(X\setminus A)\cdot P(D)$ such that $X\setminus A \in \mathcal{I}(\mathcal{D})$. For pairwise disjoint $(A_n)_{n\in\mathbb{N}} \subset \mathcal{I}(\mathcal{D})$ we have $P\left(\left(\mathring{\bigcup}_{n\in\mathbb{N}} A_n\right)\cap D\right) = P\left(\mathring{\bigcup}_{n\in\mathbb{N}}(A_n\cap D)\right) = \sum_{n\in\mathbb{N}} P(A_n\cap D) = \sum_{n\in\mathbb{N}} P(A_n)\cdot P(D) = P(D)\cdot\sum_{n\in\mathbb{N}} P(A_n) = P(D)\cdot P\left(\mathring{\bigcup}_{n\in\mathbb{N}} A_n\right)$ and hence $\mathring{\bigcup}_{n\in\mathbb{N}} A_n \subset \mathcal{I}(\mathcal{D})$. On account of $\mathcal{E} \subset \mathcal{I}(\mathcal{D})$ follows $\delta(\mathcal{E}) \subset \mathcal{I}(\mathcal{D})$ and hence the assertion. Since independence refers to finite subfamilies this property extends to arbitrary independent families $(\mathcal{E}_i)_{i\in I}$ and their **Dynkin-systems** $(\delta(\mathcal{E}_i))_{i\in I}$ and with [4, p. 1.6] even to their **σ-algebrae** $(\sigma(\mathcal{E}_i))_{i\in I} = (\delta(\mathcal{E}_i))_{i\in I}$ if the $(\mathcal{E}_i)_{i\in I}$ are **closed** with respect to **intersections**. Applying this property to the σ-algebrae $\sigma(\{A\}) = \{\emptyset; A; \Omega\setminus A; \Omega\}$ resp. $\sigma(\{B\})$ generated by two independents sets A and B shows the **independence of the complements**.

1.2 Borel's zero-one-law

For an **independent** sequence $(A_n)_{n\geq 1}$ of measurable sets $A_n \in \mathcal{A}$ on a probability space $(\Omega;\mathcal{A};P)$ we have $P\left(\bigcap_{n\geq 1}\bigcup_{k\geq n} A_k\right) \in \{0;1\}$.

Proof: Due to 1.1 for every $n \geq 1$ the σ-algebrae $\mathcal{T}_{n+1} = \sigma\left(\left\{\bigcap_{m=0}^{j} A_{k_m} : k_m \geq n+1; 0 \leq m \leq j \in \mathbb{N}\right\}\right)$ and $\mathcal{A}_n = \sigma\left(\left\{\bigcap_{m=0}^{j} A_{k_m} : k_m \leq n; 0 \leq m \leq j \in \mathbb{N}\right\}\right)$ are **independent**. Also for every $n \geq 1$ we have $T = \bigcap_{n\geq 1}\bigcup_{k\geq n} A_k \in \mathcal{T}_n$ and hence $\mathcal{A}_n \in \mathcal{I}(T) := \{A \in \mathcal{A} : P(A\cap T) = P(A)\cdot P(T)\}$ as well as $\mathcal{T}_n \in \sigma(\mathcal{A})$ with $\mathcal{A} = \bigcup_{n\geq 1}\mathcal{A}_n$. Since $\mathcal{I}(T)$ is a **Dynkin-system** including the **π-system** $\mathcal{A}$ and consequently $\sigma(\mathcal{A}) = \delta(\mathcal{A}) \subset \mathcal{I}(T)$ follows $T \in \mathcal{I}(T)$, i.e. T is **independent of itself** and hence $P(T) = P(T\cap T) = P(T)\cdot P(T) \in \{0;1\}$.

1.3 Random variables

Measurable mappings $X : \Omega \to Y$ on probability spaces $(\Omega;\mathcal{A};P)$ are called random variables with their **expectation** $E(X) := \int X dP$ and **probability distribution** $P_X := X(P)$. The random variables $(X_i)_{i\in I}$ with $X_i : (\Omega;\mathcal{A};P) \to (Y_i;\mathcal{A}_i)$ are **independent** if the σ-algebrae $\left(X_i^{-1}(\mathcal{A}_i)\right)_{i\in I}$ with $X_i^{-1}(\mathcal{A}_i) \subset \mathcal{A}$ are **independent**, i.e. for $i,j \in I$ and $A_i \in \mathcal{A}_i, A_j \in \mathcal{A}_j$ holds $P\left(X_i^{-1}[A_i]\cap X_j^{-1}[A_j]\right) = P_{X_i}(A_i)\cdot P_{X_j}(A_j)$. In the case of a finite $J = \{1;...;n\}$ we have $P\left(\bigcap_{i=1}^{n}\{X_i \in A_i\}\right) = \prod_{i=1}^{n} P(X_i \in A_i)$ such that according to [4, th. 8.15] the **common distribution** is given by the **product measure** $P_{(X_1;...;X_n)} = \bigotimes_{i=1}^{n} P_{X_i}$ on the **product σ-algebra** $\bigotimes_{i\in I}\mathcal{A}_i$ whence from [2, th. 7.3] follows that the **distribution of the sum** $S_n = s_n(X_1;...;X_n) = X_1 + ... + X_n$ coincides with the **convolution** $P_{S_n} = s_n \circ P_{(X_1;...;X_n)} = P_{X_1} * ... * P_{X_n}$(cf. 3.14.4). For **real-valued** random variables $X : \Omega \to \mathbb{R}$ we have $0 \leq E\left((X - E(X))^2\right) = E(X^2) - (E(X))^2$ and hence $E(X^2) \geq (E(X))^2$. The **variance** $VAR(X) = E\left((X - E(X))^2\right) = E(X^2) - E^2(X)$ resp. the **standard deviation**

$\sigma(X) := \|X - E(X)\|_2 = \sqrt{VAR(X)} = \sigma(X - E(X))$ are independent of the expected value and hence are preserved if we examine the **centered random variable** $X - E(X)$.

1.4 Chebyshev's inequality

For every random variable $X : \Omega \to \mathbb{R}^+$ on a probability space $(\Omega; \mathcal{A}; P)$ and every $t > 0$ we have $t \cdot P(X \geq t) \leq \int X dP$.

Proof: $\alpha \cdot P(\{X \geq \alpha\}) \leq \int_{\{X \geq \alpha\}} X dP \leq \int X dP$.

.

1.5 Expectations of products of independent random variables

For **independent** and **real** random variables $X, Y \in \mathcal{B}(\Omega; \mathbb{R})$ we have $E(X \cdot Y) = E(X) \cdot E(Y)$.

Proof: On account of $E(\chi_A \cdot \chi_B) = E(\chi_{A \cap B}) = P(A \cap B) = P(A) \cdot P(B) = E(\chi_A) \cdot E(\chi_B)$ the proposition holds for **characteristic** functions and due to the linearity of the integral also for **step** functions $\varphi, \psi \in \mathcal{S}(\Omega; \mathbb{R})$. For **integrable** functions $X, Y \in \mathcal{B}(\Omega; \mathbb{R})$ with P-a.e. $X = \lim_{n\to\infty} X_n$ resp. $Y = \lim_{n\to\infty} Y_n$ for sequences $(X_n)_{n\in\mathbb{N}}, (Y_n)_{n\in\mathbb{N}} \subset \mathcal{S}(\Omega; \mathbb{R})$ according to [4, p. 5.5] we have P-a.e. $X \cdot Y = \lim_{n\to\infty} (X_n \cdot Y_n)$. According to the hypothesis $E(X_n \cdot Y_n) = E(X_n) \cdot E(Y_n) \leq 2E(X) \cdot E(Y) < \infty$ holds for $n \geq N$ and some $N \in \mathbb{N}$ so that we can apply **monotone convergence** [4, p. 5.12] to obtain $E(X \cdot Y) = \lim_{n\to\infty} E(X_n \cdot Y_n) = \lim_{n\to\infty} (E(X_n) \cdot E(Y_n)) = \lim_{n\to\infty} E(X_n) \cdot \lim_{n\to\infty} E(Y_n) = E(X) \cdot E(Y)$.

1.6 The median

The real number $m(X)$ is a **median** of the random variable $X : \Omega \to \mathbb{R}$ iff $P(X \leq m(X)) \geq \frac{1}{2} \leq P(X \geq m(X))$. Obviously for two medians $m_1(X) < m_2(X)$ every intermediate value $m_1(X) < \alpha < m_2(X)$ is a median too. The **minimal median** is $m_{min}(X) = \inf\left\{\lambda \in \mathbb{R} : P(X \leq \lambda) \geq \frac{1}{2}\right\} = \inf\left\{\lambda \in \mathbb{R} : P(X > \lambda) \leq \frac{1}{2}\right\}$ since due to the **continuity from above**[4, p. 2.2.3] on the one hand we have $P(X \leq m_{min}(X)) = P\left(\bigcap_{n\geq 1}\left\{X \leq m_{min}(X) + \frac{1}{n}\right\}\right) = \inf_{n\geq 1} P\left(X \leq m_{min}(X) + \frac{1}{n}\right) \geq \frac{1}{2}$ and on the other hand $P(X \geq m_{min}(X)) = P\left(\bigcap_{n\geq 1}\left\{X \geq m_{min}(X) - \frac{1}{n}\right\}\right) = \inf_{n\geq 1} P\left(X \geq m_{min}(X) - \frac{1}{n}\right)$ $= 1 - \sup_{n\geq 1} P\left(X < m_{min}(X) - \frac{1}{n}\right) \geq \frac{1}{2}$, i.e. $m_{min}(X)$ is itself a **median** and since for every $\epsilon > 0$ holds $P(X \leq m_{min}(X) - \epsilon) < \frac{1}{2}$ it is the **minimal median**. Correspondingly the **maximal median** is $m_{max}(X) = \sup\left\{\lambda \in \mathbb{R} : P(X \geq \lambda) \geq \frac{1}{2}\right\} = \sup\left\{\lambda \in \mathbb{R} : P(X < \lambda) \leq \frac{1}{2}\right\}$. The relation $m_{min}(X) \leq m_{max}(X)$ holds since otherwise we had $\sup_{n\geq 1} P\left(X \geq m_{max}(X) + \frac{1}{n}\right) = P\left(\bigcup_{n\geq 1}\left\{X \geq m_{max}(X) + \frac{1}{n}\right\}\right)$ $= P(X > m_{max}(X)) > \frac{1}{2}$, i.e. there existed a $\lambda = m_{max}(X) + \frac{1}{n}$ with $P(X \geq \lambda) \geq \frac{1}{2}$ contrary to the definition of $m_{max}(X)$. Obviously we have **linearity** in the form $c \cdot m(X) = m(c \cdot X)$ and $m(X) + c = m(X + c)$ for every $c \in \mathbb{R}$.

1.7 Lévy's inequality

For **independent** and **real** random variables $X_i : (\Omega, A, P) \to \mathbb{R}$, $1 \leq i \leq n$ with sums $S_m := \sum_{i=1}^{m} X_i$ and every $\epsilon > 0$ we have $\mu\left(\max_{1\leq i\leq n} |S_i + m(S_n - S_i)| \geq \epsilon\right) \leq 2P(|S_n| \geq \epsilon)$.

Note: This inequality allows us to obtain an estimate for the maximal deviation $|S_i + m(S_n - S_i)|$ of **all** partial sums S_i given the measure of the deviation $|S_n|$ of the **single** sum S_n.

Proof: For $S_0 := 0$ and $T = \min_{1\leq i\leq m} \{|S_i + m(S_n - S_i)| \geq \epsilon\}$ if such an i exists and $T := n+1$ otherwise the **pairwise disjoint** sets $A_i := \{T = i\} \in \sigma(X_1, ..., X_i)$ are **independent** of $B_i = \{S_n - S_i \geq m(S_n - S_i)\} \in \sigma(X_i, ..., X_n)$. Hence from $P(B_i) \geq \frac{1}{2}$ follows $P(S_n \geq \epsilon) \geq P\left(\bigcup_{i=1}^{n} A_i \cap B_i\right) = \sum_{i=1}^{n} P(A_i \cap B_i) = \sum_{i=1}^{n} P(A_i) \cdot P(B_i) \geq \frac{1}{2} P(1 \leq T \leq n) = \frac{1}{2}\mu\left(\max_{1\leq i\leq n} S_i + m(S_n - S_i) \geq \epsilon\right)$. Since the same inequality holds for $-X_i$ resp. $-S_i$ with $m(-S_n + S_i) = -m(S_n - S_i)$ and all corresponding sets are **disjoint** we can use the **additivity** of P and simply **add** the two inequalities to obtain the assertion.

1.8 Lévy's convergence theorem

For the sequence $(S_n)_{n\geq 1}$ of the sums $S_n := \sum_{i=1}^{n} X_i$ of **real** and **independent** random variables $(X_i)_{i\geq 1}$ the P**-a-e- convergence** is **equivalent** to the **convergence in measure.**

Proof:

$\Rightarrow$: **Lebesgue's convergence theorem**[4, p. 4.11].

$\Leftarrow$: **Riesz' convergence theorem**[4, p. 4.13.3] provides for every $\frac{1}{4} > \epsilon > 0$ an $n_\epsilon \geq 1$ with $P(|S_n - S_m| \geq \epsilon) < \epsilon$ for all $n > m \geq n_\epsilon$. In particular we have $P(|S_n - S_m| \geq \epsilon) < \frac{1}{2}$ and hence $|m(S_n - S_m)| \leq \epsilon$ for $n > m \geq n_\epsilon$. The preceding inequality yields $P\left(\max_{m<i\leq n} |S_i - S_m| \geq 2\epsilon\right) \leq 2P(|S_n - S_m| \geq \epsilon) < 2\epsilon$. For $n \to \infty$ follows $P\left(\sup_{m<i} |S_i - S_m| \geq 2\epsilon\right) \leq 2\epsilon$ and due to the **completeness** [4, p. 4.14] of the P-a-e- convergence we obtain the assertion.

1.9 Abel's partial summation

1. For two **real** sequences $(a_i)_{i\geq 0}, (b_i)_{i\geq 0} \subset \mathbb{R}$ and $A_n = \sum_{i=0}^{n} a_i$ we have
   $\sum_{i=1}^{n} a_i b_i = A_n b_n - A_0 b_1 - \sum_{i=1}^{n-1} A_i (b_{i+1} - b_i)$ for $n \geq 1$.
2. If also $\lim_{n\to\infty} A_n = A_0^* < \infty$ with $A_n^* = \sum_{i>n} a_i$ holds we have
   $\sum_{i=1}^{n} a_i b_i = A_0^* b_1 - A_n^* b_n + \sum_{i=1}^{n-1} A_i^* (b_{i+1} - b_i)$ für $n \geq 1$.
3. If additionally $a_i \geq 0$ and $b_{i+1} \geq b_i \geq 0$ for all $i \geq 0$ is satisfied we have
   $\sum_{i=1}^{n} a_i b_i = A_0^* b_1 + \sum_{i=1}^{n-1} A_i^* (b_{i+1} - b_i)$ for $n \geq 1$.

Proof:

1. $\sum_{i=1}^{n} a_i b_i = \sum_{i=0}^{n-1} (A_{i+1} - A_i) b_{i+1} = A_n b_n - \sum_{i=1}^{n-1} A_i (b_{i+1} - b_i) - A_0 b_1$.
2. Follows from 1. with $a_0 = -\sum_{i=1}^{\infty} a_i = -A_0^*$.
3. In the case of $\lim_{n\to\infty} A_n^* b_n > 0$ with $\sum_{i>n} a_i b_i \geq A_n^* b_n$ and 2. we have $A_0^* b_1 + \sum_{i\geq 1} A_i^* (b_{i+1} - b_i) \geq \sum_{i>1} a_i b_i = \infty$ and hence the assertion. For $\lim_{n\to\infty} A_n^* b_n = 0$ it directly follows from 2. with $n \to \infty$.

1.10 Kronecker's lemma

For a **positive real** and **increasing** sequence $(b_i)_{i\geq 1}$ with $\lim_{i\to\infty} \frac{1}{b_i} = 0$ and a further **real** sequence $(a_i)_{i\geq 1}$ with $\sum_{i\geq 1} \frac{a_i}{b_i} < \infty$ we have $\lim_{n\to\infty} \frac{1}{b_n} \sum_{i=1}^{n} a_i = 0$.

Proof: From 1.9.2 with $c_i = \frac{a_i}{b_i}$ and $\lim\limits_{n\to\infty} C_n = C_0^* = \sum_{i\geq 1} \frac{a_i}{b_i} < \infty$ resp. $\lim\limits_{n\to\infty} C_n^* = 0$ we have the decomposition $\frac{1}{b_n} \sum\limits_{i=1}^{n} a_i = \frac{1}{b_n} \sum\limits_{i=1}^{n} c_i b_i = \frac{1}{b_n} C_0^* b_1 + C_n^* + \frac{1}{b_n} \sum\limits_{i=1}^{n-1} C_i^* (b_{i+1} - b_i)$. For $n \to \infty$ the first two summands converge to zero. This also holds for the third summand since for every $\epsilon > 0$ there is an $m \geq 1$ with $|C_i^*| < \epsilon$ for all $i \geq m$ such that on the one hand $\left| \frac{1}{b_n} \sum\limits_{i=m}^{n-1} C_i^* (b_{i+1} - b_i) \right| < \epsilon \frac{1}{b_n} \sum\limits_{i=m}^{n-1} (b_{i+1} - b_i) = \epsilon \left(1 - \frac{b_m}{b_n}\right) < \epsilon$ and on the other hand $\left| \frac{1}{b_n} \sum\limits_{i=1}^{m-1} C_i^* (b_{i+1} - b_i) \right| < \epsilon$ for a sufficiently large $n \geq 1$.

1.11 The Khintchin-Kolmogorov convergence theorem

For every sequence $(X_n)_{n\geq 1}$ **of independent and centered random variables** $X_n \in L^2(P)$ with $\sum_{n\geq 1} E(X_n^2) < \infty$ the **sums** $S_m := \sum\limits_{n=1}^{m} X_n$ **converge** P**-a.e.** and in **quadratic mean** to a $S = \lim\limits_{m\to\infty} S_m \in L^2(P)$ with $E(S)^2 = \sum_{n\geq 1} E(X_n^2)$.

Proof : Owing to 1.5, $E(X_n) = 0$ for all n≥ 1 and by the hypothesis we have $\lim\limits_{k\to\infty} \sup\limits_{m\geq k} E(S_m - S_k)^2 = \lim\limits_{k\to\infty} \sup\limits_{m\geq k} \sum\limits_{i=k}^{m} E(X_i^2) = 0$ such that due to[4, p. 6.7] there is an $S = \lim\limits_{k\to\infty} S_{m(k)} \in L^2(P)$ with a μ-**a.e. convergent partial sequence** $\left(S_{m(k)}\right)_{k\geq 1}$ as well as convergence of the complete sequence in the **quadratic mean**: $\lim\limits_{m\to\infty} E(S - S_m)^2 = 0$. Owing to[4, p. 6.9] we can infer the convergence **in measure** and due to **Lévy's theorem** 1.8 μ**-a.e.** convergence of the **complete series.** Due to 1.5 and $E(X_n) = 0$ we also obtain $E(S)^2 = \lim\limits_{m\to\infty} E(S_m)^2 = \sum_{n\geq 1} E(X_n^2)$.

1.12 Kolmogorov's strong law of large numbers

The **mean values** $\frac{1}{n} S_n = \frac{1}{n} \sum\limits_{k=1}^{n} X_k$ of every sequence $(X_k)_{k\geq 1}$ of **independent, identically distributed** and **integrable random variables** P**-almost sure converge** to the **common expectation :** $\lim\limits_{n\to\infty} \frac{1}{n} S_n = E(X_1)$.

Note: The strong law of large numbers provides a mathematical basis for the principle of learning from experience and every statistical method in science. From the mean results $\frac{1}{n} S_n$ of independent trials executed under similar conditions in the **past** we infer the expected outcome $E(X_1)$ in the **future**.

Proof: At first we prove the proposition for **truncated random variables** $Y_k = \frac{1}{k} \cdot X_k \cdot \chi_{\{|X_k| \leq k\}}$: With the sets $A_n = \{n - 1 < |X_1| \leq n\}$ we obtain $\sum_{k\geq 1} E\left(|Y_k|^2\right) = \sum_{k\geq 1} \sum_{k\geq n\geq 1} n^{-2} \int_{A_n} |X_1|^2 dP$ $= \sum_{n\geq 1} \sum_{k\geq n} n^{-2} \int_{A_m} |X_1|^2 dP \leq \sum_{n\geq 1} \frac{2}{n} \int_{A_m} |X_1|^2 dP \leq 2 \sum_{n\geq 1} \int_{A_n} |X_1| dP \leq 2E(|X_k|) < \infty$ so that due to **Khintchin - Kolmogorov** 1.11 we have P-a.s. $\sum_{k\geq 1} (Y_k - E(Y_k)) < \infty$.

The deviations have the measure $\sum_{k\geq 1} P\left(\frac{1}{k} X_k \neq Y_k\right) = \sum_{k\geq 1} P(|X_1| > k) \leq \sum_{k\geq 1} \sum_{n\geq k} P(n + 1 \geq |X_1| > n)$ $\leq \sum_{n\geq 1} \sum_{n\geq k\geq 1} P(n + 1 \geq |X_1| > n) = \sum_{k\geq 1} (k + 1) \cdot P(n + 1 \geq |X_1| > n) \leq E(|X_1|) < \infty$ such that according to **Borel-Cantelli** [4, th. 4.12] follows $P\left(\bigcap_{n\geq 1} \bigcup_{k\geq n} \left\{\frac{1}{k} X_k \neq Y_k\right\}\right) = 0$ and with the first estimate above we obtain P-a.e. $\sum_{k\geq 1} \frac{1}{k} (X_k - E(k \cdot Y_k)) = \sum_{k\geq 1} \left(\frac{1}{k} \cdot X_k - E(Y_k)\right) < \infty$. On account of $\lim\limits_{n\to\infty} \frac{1}{n} \sum\limits_{k=1}^{n} E(k \cdot Y_k) = \lim\limits_{n\to\infty} \frac{1}{n} \sum\limits_{k=1}^{n} E\left(X_1 \cdot \chi_{\{|X_1| \leq k\}}\right) = \lim\limits_{n\to\infty} E\left(X_1 \cdot \chi_{\{|X_1| \leq k\}}\right) = E(X_1)$ and **Kronecker** 1.10 follows $\lim\limits_{n\to\infty} \frac{1}{n} S_n - E(X_1) = \lim\limits_{n\to\infty} \frac{1}{n} \sum\limits_{k=1}^{n} (X_k - E(k \cdot Y_k)) = 0$.

2 Stochastic processes

2.1 Definition of the bold game strategy

A gambler enters the casino with **capital** $C_0 > 0$ and takes **independently** and **identically distributed** bets with $P(Y_k = 1) = p$ resp. $P(Y_k = -1) = 1 - p = q$ for $k \geq 1$ until his fortune $C_0 + S_n$ with $S_n = \sum_{k=1}^{n} Y_k$ reaches either in the case $S_{C_0,n} = \{C_0 + S_n = c\} \cap \bigcap_{k=1}^{n-1} \{0 < C_0 + S_n < c\}$ of **success** the goal c or in the case $R_{C_0,n} = \{C_0 + S_n = 0\} \cap \bigcap_{k=1}^{n-1} \{0 < C_0 + S_n < c\}$ of **ruin** the value 0. The **probability of ultimate success** is $s_c(C_0) = P\left(\bigcup_{n\geq 1} S_{C_0,n}\right) = \sum_{n\geq 1} P(S_{C_0,n})$ and correspondingly the **probability of ultimate ruin** is $r_c(C_0) = P\left(\bigcup_{n\geq 1} R_{C_0,n}\right) = \sum_{n\geq 1} P(R_{C_0,n})$. The cases $S_{C_0,0} = S_{0,n} = S_{c,n} = \emptyset$ for $C_0 < c$ resp. $S_{c,0} = \Omega$ yield the **boundary conditions** $s_c(0) = 0$ and $s_c(c) = 1$. Similarly $R_{C_0,0} = R_{0,n} = R_{c,n} = \emptyset$ for $C_0 < c$ resp. $R_{0,0} = \Omega$ give $r_c(0) = 1$ and $r_c(c) = 0$. Since the bets are **independently** and **identically distributed** we have the **recursive formulae**

$$s_c(C_0) = p \cdot s_c(C_0 + 1) + q \cdot s_c(C_0 - 1) \text{ resp. } r_c(C_0) = q \cdot r_c(C_0 + 1) + p \cdot r_c(C_0 - 1).$$

In general these recursions have the **explicit solutions**

$$s_c(C_0) = \begin{cases} A + B \cdot \rho^{C_0} & \text{if } p \neq q \\ A + B \cdot a & \text{if } p = q \end{cases} \text{ for } \rho = \frac{q}{p}.$$

The boundary conditions result in

$$s_c(C_0) = \begin{cases} \frac{\rho^{C_0}-1}{\rho^c - 1} & \text{if } p \neq q \\ \frac{C_0}{c} & \text{if } p = q \end{cases} \text{ resp. } r_c(C_0) = \begin{cases} \frac{\rho^{C_0-c}-1}{\rho^{-c}-1} & \text{if } p \neq q \\ \frac{c-C_0}{c} & \text{if } p = q \end{cases}$$

Hence $s_c(C_0) + r_c(C_0) = 1$, i.e. the game will P**-almost sure** not continue forever.

In the n-th game the **wager** $W_n(C_0; Y_1; ...; Y_{n-1}) \geq 0$ results in the **win** $W_n Y_n$ and the **capital** $C_n = C_{n-1} + W_n Y_n$. The random variables $(Y_k)_{k\geq 1}$ generate an **increasing filtration** $(\mathcal{F}_n)_{n\geq 1}$ with $\mathcal{F}_n = \sigma(Y_1; ...; Y_n)$ representing the **knowledge** up to the n-th game. Since the σ-algebrae $\sigma(Y_n)$ are independent of the $\mathcal{F}_{n-1} = \sigma(Y_1; ...; Y_{n-1})$ due to 1.5 we have $E(Y_n \cdot W_n) = E(Y_n) \cdot E(W_n) = (p - q) \cdot E(W_n)$. Consequently in the **subfair** case with $p < q$ the sequence $(E(C_n))_{n\geq 1}$ of expected capital is **decreasing**.

The **stopping time** $\tau : \mathbb{R} \times \Omega \to \mathbb{N}$ denotes the number $\tau(C_0; \omega)$ of trials the gambler plays before he decides to stop. This decision depends only in the knowledge gathered up to τ, i.e. $\{\tau = n\} \in \mathcal{F}_n$. Also we assume that $P(\tau < \infty) = 1$. The **capital** then is

$$C_n^* = \begin{cases} C_n & \text{if } \tau \geq n \\ C_\tau & \text{if } \tau \leq n \end{cases} \text{ with the wager } W_n^* = \begin{cases} W_n & \text{if } \tau \geq n \\ 0 & \text{if } \tau \leq n \end{cases} = W_n \chi_{\{\tau \geq n\}}$$

so that we arrive at the recursive formula $C_n^* = C_{n-1}^* + W_n^* \cdot Y_n$. Since $\{\tau \geq n\} = \Omega \setminus \{\tau < n\} \in \mathcal{F}_{n-1}$ the random variables C_n^* resp. W_n^* are $\mathcal{F}_{n-1}$- measurable whence the argument from above applies whence the sequence $(E(C_n^*))_{n\geq 1}$ of expected capital still is **decreasing**. If we assume a finite line of credit of the gambler as well as a finite capital of the bank, i.e. $-M \leq C_n^* \leq M$ for an $M > 0$ and every $n \geq 1$ and consider that P-a.s. $\lim_{n\to\infty} C_n^* = C_\tau$ the **dominated convergence theorem**[4, p. 5.14] yields $\lim_{n\to\infty} E(C_n^*) = E(C_\tau)$ and in particular $E(C_\tau) \leq E(C_n) \leq E(C_1) \leq C_0$: No gambling system may reverse the odds of a subfair game.

Nonetheless it is possible to optimize the (still unfavourable) success probability in a **subfair** game in a striking way leading to a P**-a.e. differentiable** function with **fractal character** and **outside**

the domain of the fundamental theorem of calculus. To this end we scale the **initial fortune** to $0 \leq C_0 \leq 1$and the **goal** to $c = 1$. The **bold game** strategy is defined by

$$W_n = \begin{cases} C_{n-1} & \text{if } 0 \leq C_{n-1} \leq \frac{1}{2} \\ 1 - C_{n-1} & \text{if } \frac{1}{2} \leq C_{n-1} \leq 1 \end{cases} \text{ and } \tau \left(C_0; \omega\right) = n \text{ iff } C_n \in \{0; 1\}.$$

Under the condition that the play has not terminated at time $k-1$ it will continue beyond k iff either $Y_k = 1$ in the case of $C_{k-1} \leq \frac{1}{2}$ or$Y_k = -1$ in the case of $C_{k-1} \geq \frac{1}{2}$. Hence we have $P(\tau \geq k+1 | \tau \geq k) \leq m = \max\{p; q\}$ whence $P(\tau \geq k+1) \leq m^n$ and consequently $P(\tau = \infty) = 0$. Thus the game will terminate P-a.s. The mapping $C_\tau : \Omega \to \{0; 1\}$ is a $\bigcup_{n \in \mathbb{N}} \mathcal{F}_n$-measurable random variable since $\{C_\tau = y\} = \bigcup_{n \in \mathbb{N}} (\{\tau = n\} \cap \{C_n = y\})$ for $y \in \{0; 1\}$. We now examine the success probability of the initial capital $0 \leq x \leq 1$ expressed by the function $F : [0; 1] \to [0; 1]$ with $F(x) = P(C_\tau = 1)$ for $C_0 = x$.

2.2 Properties of the bold game strategy

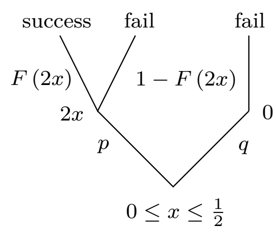
1. In the subfair case $p \leq q$ of a sequence of trials with **independently** and **identically distributed** outcomes $Y_i : \Omega \to \{-1; 1\}$ and $P(Y_i = 1) = p$ resp. $P(Y_i = -1) = q$ for every $0 \leq x \leq 1$ the success probability F of the **bold game** strategy as described above satisfies the **functional equation** $F(x) = \begin{cases} p \cdot F(2x) & \text{if } 0 \leq x \leq \frac{1}{2} \\ p + q \cdot F(2x-1) & \text{if } \frac{1}{2} \leq x \leq 1 \end{cases}$.

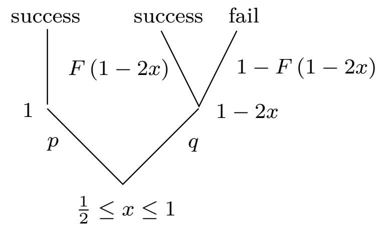
2. It is also the **distribution function** $F(x) = P(X \geq x)$ of the **random variable** $X = \sum_{i \geq 1} \frac{X_i}{2^i} : \Omega \to [0; 1]$ with **independently** and **identically distributed coefficients** $X_i : \Omega \to \{0; 1\}$ and $P(X_i = 1) = p$ resp. $P(X_i = 0) = q$.

3. The function $F : [0; 1] \to [0; 1]$ is **continuous**, **increasing** and P-a.e. **differentiable** with P-a.e. $\frac{dF}{dx}(x) = 0$ for $p < q$.

Note: The functional equation expresses the **fractal character** of the distribution function F in terms of self-similarity: The values $F(y) \in [0; 1]$ **on** the whole domain $y \in [0; 1]$ are replicated in the lower part $F\left(\frac{1}{2}y\right) = \frac{1}{p}F(y) \in [0; p]$ and the values $F(y) \in [p; 1]$ **in** the upper part are also repeated in the interval $F\left(\frac{1}{2}(y+1)\right) = \frac{1}{q}(F(y) - p) \in \left[0; \frac{1}{q}\right]$.

Proof: The **functional equation** follows from the **event tree** at the right hand side based on the independently and identically distributed probabilities of the separate trials. Applying it we obtain





$F(1) = P(1.0)_2 = 1;$

$F\left(\frac{1}{2}\right) = P(0.1)_2 = p;$

$F\left(\frac{1}{4}\right) = F(0.01)_2 = P(1; 1) = p^2;$

$F\left(\frac{3}{4}\right) = F(0.11)_2 = P(1 \vee 0; 1) = p + qp;$

$F\left(\frac{1}{8}\right) = F(0.001)_2 = P(1; 1; 1) = p^3;$

$F\left(\frac{3}{8}\right) = F(0.011)_2 = P(1; 1 \vee 1; 0; 1) = p^2 + pqp;$

$F\left(\frac{5}{8}\right) = F(0.101)_2 = P(1 \vee 0; 1; 1) = p + qpp;$

$F\left(\frac{7}{8}\right) = F(0.111)_2 = P(1 \vee 0; 1 \vee 0; 0; 1) = p + qp + qqp$

In general for a **dyadic number** $x = \sum_{i=1}^{n} \frac{x_i}{2^i} = (0.x_1...x_n)_2$ of **rank** $n \geq 1$ we have

- either $(0.x_1...x_n)_2 + \frac{1}{2^n} \leq \frac{1}{2}$ hence $x_1 = 0$ and $F(x) = p \cdot F(2x)$ so that $F\left((0.x_1...x_n)_2 + \frac{1}{2^n}\right) - F(0.x_1...x_n)_2 = p\left(F\left((0.x_2...x_n)_2 + \frac{1}{2^{n-1}}\right) - F(0.x_2...x_n)_2\right)$
- or $(0.x_1...x_n)_2 + \frac{1}{2^n} \geq \frac{1}{2}$ hence $x_1 = 1$ and $F(x) = p + q \cdot F(2x-1)$ so that due to $2 \cdot \left((0.x_1...x_n)_2 + \frac{1}{2^n}\right) - 1 = (1.x_2...x_n)_2 + \frac{1}{2^{n-1}} - 1 = (0.x_2...x_n)_2 + \frac{1}{2^{n-1}}$ we have $F\left((0.x_1...x_n)_2 + \frac{1}{2^n}\right) - F(0.x_1...x_n)_2 = q\left(F\left((0.x_2...x_n)_2 + \frac{1}{2^{n-1}}\right) - F(0.x_2...x_n)_2\right)$.

Subsuming both cases and skewing the **outcomes** Y_i slightily so that
they fit as **coefficients** $X_i = 1 - \frac{1}{2}(1 - Y_i)$ such that $P(X_i = 1) = P(Y_i = 1) = p$ resp. $P(X_i = 0) = P(Y_i = -1) = q$ we obtain

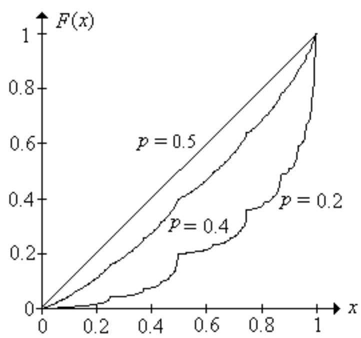
$$
\begin{aligned}
F\left((0.x_1...x_n)_2 + \frac{1}{2^n}\right) - F(0.x_1...x_n)_2 &= p(x_1)\left(F\left((0.x_2...x_n)_2 + \frac{1}{2^{n-1}}\right) - F(0.x_2...x_n)_2\right) \\
&\vdots \\
&= P(X_1 = x_1) \cdot ... \cdot P(X_n = x_n) \cdot (F(1) - F(0)) \\
&= P(X_1 = x_1) \cdot ... \cdot P(X_n = x_n) \\
&= P((X_1, ..., X_k) = (x_1, ..., x_k)) \\
&\leq m^n \text{ but also} \\
&> 0
\end{aligned}
$$

whence immediately follow the **increasing** character as well as the **continuity** of F. Also we can compute the **explicit formula** using the **Kronecker symbol** $\delta_{x1} = \begin{cases} 1 & \text{if } x = 1 \\ 0 & \text{if } x \neq 1 \end{cases}$ to exclude the cases $x_k = 0$ resp. $F(0.x_1...x_k)_2 = F(0.x_1...x_{k-1})_2$ so that

$$
\begin{aligned}
&F(0.x_1...x_n)_2 \\
&= F\left(0.x_1...x_{n-1} + \frac{1}{2^n}\right)_2 \\
&= P(X_1 = x_1) \cdot ... \cdot P(X_n = x_n) \cdot \delta_{x_n 1} + F(0.x_1...x_{n-1})_2 \\
&= P(X_1 = x_1) \cdot ... \cdot P(X_n = x_n) \cdot \delta_{x_n 1} + P(X_1 = x_1) \cdot ... \cdot P(X_{n-1} = x_{n-1}) \cdot \delta_{x_{n-1} 1} + ... + P(X_1 = x_1) \cdot \delta_{x_1 1} \\
&= \sum_{k=1}^{n} P(X_1 = x_1) \cdot ... \cdot P(X_k = x_k) \cdot \delta_{x_k 1} \\
&= \sum_{k=1}^{n} P((X_1, ..., X_k) = (x_1, ..., x_k)) \\
&= P\left(: \sum_{i=1}^{k} \frac{X_i}{2^i} \leq \sum_{i=1}^{n} \frac{x_i}{2^i}; k \leq n\right)
\end{aligned}
$$

Hence for every **dyadic number** $x = \sum_{i=1}^{n} \frac{x_i}{2^i}$ of **rank** $n \geq 1$ we have $F(x) = P(X \leq x)$ for the **random variable** $X = \sum_{i=1}^{n} \frac{X_i}{2^i}$. Since the dyadic numbers of finite rank are **dense** in $[0;1]$ and F is **continuous** this formula extends to every real $x \in [0;1]$, i.e. F is the **distribution function** for the random variable $X = \sum_{i \geq 1} \frac{X_i}{2^i}$.

In order to compute the **derivative** for a given $x \in]0;1[$ (conveniently excluding the λ-null set $\{0;1\}$) and every $n \geq 1$ we choose $0 \leq k_n \leq 2^n - 1$ such that $x \in I_n = \left]\frac{k_n}{2^n}; \frac{k_n+1}{2^n}\right[$. According to **Lebesgue's differentiation theorem** [4, p. 12.4] the derivative $\frac{dF}{d\lambda}(x) = \lim_{n\to\infty} \frac{F\left(\frac{k_n}{2^n}+\frac{1}{2^n}\right)-F\left(\frac{k_n}{2^n}\right)}{2^n} = \lim_{n\to\infty} \frac{P(X\in I_n)}{2^n}$ exists λ-a.e. on $[0;1]$. If we assume $\frac{dF}{d\lambda}(x) > 0$ it follows that $P(X \in I_n) > 0$ and from $\lim_{n\to\infty} \frac{P(X\in I_{n+1})}{2^{n+1}} = \lim_{n\to\infty} \frac{P(X\in I_n)}{2^n}$ we infer $\lim_{n\to\infty} \frac{P(X\in I_{n+1})}{P(X\in I_n)} = \frac{1}{2}$. From $\frac{k_n}{2^n} = (0.x_1...x_n)_2$ follows $\frac{k_{n+1}}{2^n} = (0.x_1...x_n x_{n+1})_2$ with $x_{n+1} = 0$ iff $I_{n+1} \subset I_n$ lies in the left half of I_n and $x_{n+1} = 0$ iff it is the right half of I_n. Due to the explicit formula shown above we infer $\frac{P(X\in I_{n+1})}{P(X\in I_n)} = \frac{P(X_1=x_1)\cdot\ldots\cdot P(X_n=x_{n+1})}{P(X_1=x_1)\cdot\ldots\cdot P(X_n=x_n)} = P(X_n = x_{n+1}) \in \{p;q\}$ contrary to the assumption $p < q$. Hence the proof is complete.



2.3 Convexity of the success probability

In the subfair case $p \leq q$ for every $0 \leq x-t \leq x \leq x+t \leq 1$ we have $F(x) \geq p \cdot F(x+t) + q \cdot F(x-t)$.

Proof: We prove the inequality $\Delta(r,s) = F(a) - pF(s) - qF(r) \geq 0$ by **induction** over n for **dyadic numbers** $0 \leq r \leq s \leq 1$ of rank n and **mean** $a = \frac{1}{2}(r+s)$ of rank $n+1$. By the continuity of F this result then extends to real arguments. We assume that the inequality holds for r, s of rank $n \geq 1$. There are four cases to consider:

Case I: $s \geq \frac{1}{2}$. The first part of the functional equation gives $\Delta(r,s) = p\Delta(2r,2s)$. Since $2r, 2s$ are of rank n the induction hypothesis implies that $\Delta(2r,2s) \geq 0$.

Case II: $\frac{1}{2} \leq r$. By the second part of the functional equation we have $\Delta(r,s) = q\Delta(2r-1, 2s-1) \geq 0$.

Case III: $r \leq a \leq \frac{1}{2} \leq 2$. The functional equation delivers $\Delta(r,s) = pF(2a) - p(p + qF(2s-1)) - q(pF(2r))$. From $\frac{1}{2} \leq s \leq r+s = 2a \leq 1$ follows $F(2a) = p + qF(4a-1)$ and from $0 \leq 2a - \frac{1}{2} \leq \frac{1}{2}$follows $F\left(2a - \frac{1}{2}\right) = pF(4a-1)$. Therefore $pF(2a) = p^2 + qF\left(2a - \frac{1}{2}\right)$ whence $\Delta(r,s) = q\left(F\left(2a - \frac{1}{2}\right) - pF(2s-1) - pF(2r)\right)$. Since $p \leq q$the right side does not increase if either of the two p is changed to q. Hence $\Delta(r,s) \geq q\max\{\Delta(2r, 2s-1), \Delta(2s-1, 2r)\}$. Since we may apply the induction hypothesis either to $2r \leq 2s-1$ or to $2s-1 \leq 2r$ at least one of the two Δ on the right is nonnegative.

Case IV: $r \leq \frac{1}{2} \leq a \leq s$. The functional equation gives $\Delta(r,s) = pq + qF(2a-1) - pqF(2s-1) - pqF(2r)$. From $0 \leq 2a-1 = r+2-1 \leq \frac{1}{2}$ follows $F(2a-1) = pF(4a-2)$ and from $\frac{1}{2} \leq 2a - \frac{1}{2} = r+s-\frac{1}{2} \leq 1$ follows $F\left(2a - \frac{1}{2} = p + qF(4a-2)\right)$. Therefore $qF(2a-1) = pF\left(2a - \frac{1}{2}\right)$and it follows that $\Delta(r,s) = p\left(q - p + F\left(2a - \frac{1}{2}\right) - qF(2s-1) - qF(2r)\right)$. On the one hand if $2s-1 \leq 2r$ the right side becomes $p((q-p)(1-F(2r)) + \Delta(2s-1, 2r)) \geq 0$. On the other hand if $2r \leq 2s-1$ it is $p((q-p)(1-F(2s-1)) + \Delta(2r, 2s-1)) \geq 0$. This completes the proof.

2.4 The Dubins-Savage Theorem

The bold play strategy is the optimal strategy in the subfair case $p \leq q$, i.e. for every other strategy π and every initial capital $0 \leq x \leq 1$ we have $F_\pi(x) \leq F(x)$.

Proof: We consider the conditional chance $F\left(C^*_{\pi,n}\right)$ of success if the strategy π is replaced by bold game after the n-th trial and the capital $C^*_{\pi,n}(C_0, Y_1, ..., Y_n)$ depending on the initial capital $0 \leq C_0 \leq 1$ and the independently as well as identically distributed outcomes $Y_i \in \{-1, 1\}$ in the trials $1 \leq i \leq n$. We abbreviate $C^*_{\pi,n-1} = x$ and $W^*_{\pi,n} = t$ so that we can write $C^*_{\pi,n} = x + tY_n$ and $F\left(C^*_{\pi,n}\right) = \sum_{x,t} \chi_{\left\{C^*_{\pi,n-1}=x, W^*_{\pi,n}=t\right\}} F(x + tY_n)$ where x resp. t vary over the **finite** ranges of

$C^*_{\pi,n-1}$ resp. $W^*_{\pi,n}$. Since $C^*_{\pi,n-1}$ and $W^*_{\pi,n}$ are $\sigma(Y_1,...,Y_{n-1})$-measurable and $F(x+tY_n)$ is $\sigma(Y_n)$-measurable for the now **fixed** (!) s and t in the sum by **independence** we obtain $E\left(F\left(C^*_{\pi,n}\right)\right) = \sum_{x,t} P\left(C^*_{\pi,n-1}=x, W^*_{\pi,n}=t\right)\cdot E\left(F\left(x+tY_n\right)\right)$. According to the preceding lemma 2.3 we have $E\left(F\left(x+tY_n\right)\right)\le F(x)$if $0\le x-t\le x\le x+t\le 1$. We assume that the alternative strategy π keeps to the same capital limits as the bold game, i.e. $W^*_{\pi,n}\le \min\left\{C^*_{\pi,n-1}, 1-C^*_{\pi,n-1}\right\}$ and consequently $C^*_{\pi,n}\in[0;1]$ whence $E\left(F\left(C^*_{\pi,n}\right)\right)\le\sum_{x,t}P\left(C^*_{\pi,n-1}=x,W^*_{\pi,n}=t\right)\cdot F(x) = \sum_x P\left(C^*_{\pi,n-1}=x\right)\cdot F(x) = E\left(F\left(C^*_{\pi,n-1}\right)\right)$. This inequality already implies that every trial the gambler waits before changing to bold play **diminishes his expected chance of success in the overall game** played in the first n trials in some arbitrary alternative strategy and from the $n+1$ th game on with bold play. But we can sharpen this statement considerably: Since the estimate is true for each $n\ge 1$ and $F\left(C^*_{\pi,\tau_\pi}\right) = F\left(C^*_{\pi,n}\right) = \begin{cases}1 & \text{if } C^*_{\pi,\tau_\pi}=1\\ 0 & \text{if } C^*_{\pi,\tau_\pi}\neq 1\end{cases}$ for $n\ge\tau_\pi$ with $P(\tau_p<\infty)$ guaranteed by the alternative strategy π (cf. 2.1) we obtain $E\left(F\left(C^*_{\pi,\tau_\pi}\right)\right) = E\left(F\left(C^*_{\pi,n}\right)\right)\le E\left(F\left(C^*_{\pi,0}\right)\right) = E\left(F\left(C_0\right)\right) = F\left(C_0\right)$. Since $F_\pi\left(C_0\right) = 1\cdot P\left(C^*_{\pi,\tau_\pi}=1\right) = F\left(C^*_{\pi,\tau_\pi}=1\right)\cdot P\left(C^*_{\pi,\tau_\pi}=1\right)\le E\left(F\left(C_\tau\right)\right)\le F\left(C_0\right)$ we have proven the assertion.

3 Weak convergence

3.1 Simple discontinuities of monotone functions

Every **monotone** function $f:\,]a;b[\,\to\mathbb{R}$ is **continuous** except at a **countable** set of points and the **discontinuity** at each of such point $c\in\,]a;b[$ is **simple**, i.e.

$$-\infty < \sup_{a<x<c} f(x) = \lim_{n\to\infty} f\left(c-\frac{1}{n}\right) < \lim_{n\to\infty} f\left(c+\frac{1}{n}\right) = \inf_{c<x<b} f(x) < \infty$$

.

Notes:

1. In [4, th. 11.1] it is shown that for every (not necessarily measurable) $f:(X;d)\to(Y;D)$ between **metric spaces** the set of **discontinuities**

   $$D_f = \{x\in X:\ \exists\epsilon>0:\ \forall\delta>0\,\exists y;z\in B_\delta(x):\ D\left(f(y);f(z)\right)\ge\epsilon\}$$

   is **$\mathcal{B}(X)$-measurable**.

2. In [2, th. 1.2] it is proved that for every **real** $f:\mathbb{R}\to\mathbb{R}$ the set of **jump** and **vertex** points with **existing** but **differing Dini derivatives**

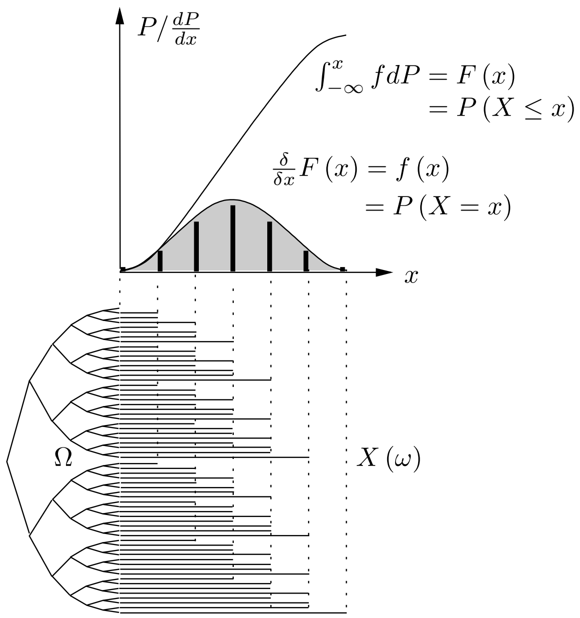
   $$\left\{D_+f = D^+f = D_+^+f \neq D_-^-f = D_-f = D^-f\right\}$$

   is **countable.**

Proof: W.l.o.g. we assume f to be **nondecreasing** whence $a<x<c<y<b$ implies $-\infty < f(x) < f(c) < f(y) < \infty$ and consequently $-\infty<\alpha=\sup_{a<x<c} f(x)\le f(c)\le \inf_{c<x<b} f(x)=\beta<\infty$. In order to prove that $\alpha = f(c-) = \lim_{n\to\infty} f\left(c-\frac{1}{n}\right)$ resp. $\beta = f(c+) = \lim_{n\to\infty} f\left(c+\frac{1}{n}\right)$ we observe that the nondecreasing character of f implies that for every $\epsilon>0$ there is an $m\ge 1$ such that for every $n\ge m$ holds $\alpha-\epsilon < f\left(c-\frac{1}{n}\right)\le\alpha$ whence follows $f(c-)=\alpha$ and analogously $f(c+)=\beta$. Also we remark that $a<c<x<d<b$ implies $f(c+)\le f(x)\le f(d-)$. Hence for every $c;d\in D_f = \{x\in\,]a;b[\,: f(x-)<f(x+)\}$ there are **rational** $f(c-)<r_c<f(c+)<f(d-)<r_d<f(d+)$, i.e. the map $r:D\to\mathbb{Q}$ defined by $r(c)=r_c$is **injective**.

3.2 Distribution functions

Every **random variable** $X:\Omega\to\mathbb{R}$ on a **probability space** $(\Omega;\mathcal{A};P)$ determines a probability measure $P_X = X\circ P$ on $(\mathbb{R};\mathcal{B}(\mathbb{R}))$ and according to [4, th. 3.7] a **nondecreasing** and **right continuous distribution function** $F_X:\mathbb{R}\to[0;1]$ with **existing left limits** such that $F_X(x) = (X\circ P)(]-\infty;x]) = P(X\leq x)$. According to the preceding theorem 3.1 every distribution function has at most a **countable number of simple discontinuities**. Conversely every such distribution function $F:\mathbb{R}\to[0;1]$ determines a **unique probability measure** P_F on $(\mathbb{R};\mathcal{B}(\mathbb{R}))$ and **many possible probability spaces** $(\Omega;\mathcal{A};P)$ with corresponding **random variables** $X:\Omega\to\mathbb{R}$ such that $F(x) = P(X\leq x) = P(]-\infty;x])$, among them the **trivial random variable** $X=\mathrm{id}:\mathbb{R}\to\mathbb{R}$. E.g. the **binomial distribution** $b_{3;0,5}$ with $b_{3;0,5}(0) = b_{3;0,5}(3) = \frac{1}{8}$ resp. $b_{3;0,5}(1) = b_{3;0,5}(2) = \frac{3}{8}$ may be realized by three tosses of a coin as well as by the single throw of an octagonal die with corresponding labels.



3.3 Expectations and distribution functions

For every random variable $X:\Omega\to\mathbb{R}$ on a probability space $(\Omega;\mathcal{A};P)$ we have

1. $E(X) = \int_0^\infty P(X\geq x)\,dx - \int_{-\infty}^0 P(X\leq x)\,dx$.

In the case of a **continuous distribution function** $F:\mathbb{R}\to[0;1]$ with $F(x) = P(X\leq x)$ holds

2. $E(X) = \int_0^\infty (1-F(x))\,dx - \int_{-\infty}^0 F(x)\,dx$.

Proof: By Fubini [4, th. 8.9] we have the expectation of the **positive** part $E(X^+) = \int X^+dP$ $= \int t dP_{X^+}(t) = \int\int \chi_{[0\leq x\leq t]}(x)\,dxdP_X(t) = \int\int \chi_{[0\leq x\leq t]}(t)\,dP_X(t)\,dx = \int_0^\infty P(X\geq x)\,dx$ and in the case of a **continuous distribution function** $F:\mathbb{R}\to[0;1]$ with $P(X=x) = P_X(\{x\}) = P_X\left(\bigcap_{n\geq 1}\left]x-\frac{1}{n};x+\frac{1}{n}\right]\right) = \lim_{n\to\infty}\left(F\left(x+\frac{1}{n}\right)-F\left(x+\frac{1}{n}\right)\right) = \lim_{n\to\infty}F\left(x+\frac{1}{n}\right) - \lim_{n\to\infty}F\left(x+\frac{1}{n}\right) = F(x)-F(x) = 0$ follows $E(X^+) = \int_0^\infty P(X>x)\,dx = \int_0^\infty(1-F(x))\,dx$. The **negative** part is computed by $E(X^-) = \int X^-dP = \int t dP_{X^-}(t) = \int\int\chi_{[t\leq x\leq 0]}(x)\,dxdP_X(t) = \int\int\chi_{[t\leq x\leq 0]}(t)\,dP_X(t)\,dx$ $= \int_{-\infty}^0 P(X\leq x)\,dx = \int_{-\infty}^0 F(x)\,dx$ whence by $E(X) = E(X^+-X^-) = E(X^+)-E(X^-)$ follows the assertion.

3.4 The primitive of a distribution function

The **distribution** $P_X:\mathcal{B}(\mathbb{R})\to[0;1]$ of a random variable $X:\Omega\to\mathbb{R}$ on a probability space $(\Omega;\mathcal{A};P)$ is **λ-absolutely continuous** iff its **distribution function** $F:\mathbb{R}\to[0;1]$ is **absolutely continuous** and in this case there is a **probability density function** $f = \frac{dF}{d\lambda}:\mathbb{R}\to\mathbb{R}_0^+$ which is the **primitive** of F with $P(X\leq x) = F(x) = \int_{-\infty}^x f d\lambda$.

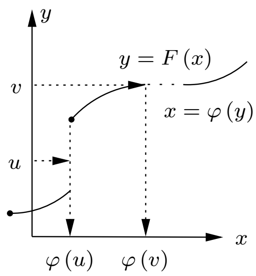
Proof:

$\Rightarrow$: According to [4, def. 9.6] for every $\epsilon>0$ there is a $\delta>0$ such that for any disjoint collection $(]\alpha_i;\beta_i[)_{1\leq i\leq n}$ of segments with overall length $\sum_{i=1}^n(\beta_i-\alpha_i) = \sum_{i=1}^n\lambda(]\beta_i-\alpha_i]) = \lambda\left(\bigcup_{i=1}^n]\beta_i-\alpha_i]\right)<\delta$ holds $\sum_{i=1}^n|F(\beta_i)-F(\alpha_i)| = \sum_{i=1}^n P_X(]\beta_i-\alpha_i]) = P_X\left(\bigcup_{i=1}^n]\beta_i-\alpha_i]\right)<\epsilon$ whence from [2, def. 2.7] follows the **absolute continuity** of F. The existence of the **probability density function** then is a consequence of the **fundamental theorem of calculus** [2, th. 12.10].

$\Leftarrow$: Follows at once from the **fundamental theorem of calculus** [2, th. 12.10] and the definition [4, def. 9.5] of absolute continuity with regard to λ.

3.5 Skorohod's representation theorem

For every sequence $(X_n)_{n\geq 1}$ of **random variables** $X_n : \Omega \to \mathbb{R}$ on a probability space $(\Omega; \mathcal{A}; P)$ converging **in measure** to a random variable $X : \Omega \to \mathbb{R}$ there is a sequence $(\varphi_n)_{n\geq 1}$ of **random variables** $\varphi_n :]0;1[\to \mathbb{R}$ on the probability space $(]0;1[\,; \mathcal{B}\,(]0;1[)\,; \lambda)$ with **identical distributions** $\varphi_n \circ \lambda = X_n \circ P$ with regard to the **Lebesgue measure** λ converging **pointwise everywhere** to a random variable $\varphi :]0;1[\to \mathbb{R}$ with $\varphi \circ \lambda = X \circ P$.



Proof: With the **distribution functions** $F_n; F : \mathbb{R} \to [0;1]$ defined by $F_n(x) = P_n(X_n \leq x)$ resp. $F(x) = P(X \leq x)$ we define the **quantile function** $\varphi_n(y) = \inf\{x \in \mathbb{R} : y \leq F_n(x)\}$ resp. $\varphi(y) = \inf\{x \in \mathbb{R} : y \leq F(x)\}$ such that due to the **nondecreasing** character and the **right continuity** of F we have $\varphi(y) \leq x \Leftrightarrow \forall \epsilon > 0 : y \leq F(x + \epsilon) \Leftrightarrow y \leq F(x) = P(X \leq x)$ whence $\lambda(\varphi \leq x) = F(x)$, i.e. $X \circ P = \varphi \circ \lambda$ and likewise $\lambda(\varphi_n \leq x) = F_n(x)$, i.e. $X_n \circ P = X_n \circ \lambda$. In particular $\varphi(y)$ is the smallest x such that $y \leq F(x)$ whence $(\varphi \circ F)(x) \leq x$ with **equality** in the case of F **strictly increasing** in x. Conversely $y \leq (F \circ \varphi)(y)$ with **equality** in the case of F being **left continuous** in $\varphi(y)$, i.e. $P(\{\varphi(y)\}) = 0$: The quantile function φ is again **nondecreasing** and **right continuous**; in the strictly increasing and continuous case it is the **inverse** of the distribution function F. It remains to show that $\lim\limits_{n\to\infty} \varphi_n(y) = \varphi(y)$ for every $y \in]0;1[$:

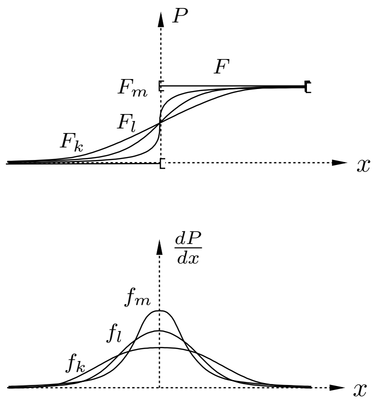
According to the note in[4, th. 2.2] there are at most countably many $x \in \mathbb{R}$ with $P(X = x) > 0$ such that every interval $]a - \epsilon; a[$ contains an x with $P(X = x) = 0$.

Consequently for every $\epsilon > 0$ there is an x with $P(X = x) = 0$ and $\varphi(y) - \epsilon < x < \varphi(y)$ such that $F(x) < y$. Since F is **continuous** in x we have $\lim\limits_{n\to\infty} F_n(x) = F(x)$ so that for n large enough $F_n(x) < y$ holds whence $\varphi(y) - \epsilon < x < \varphi_n(y)$ and consequently $\liminf\limits_{n\to\infty} \varphi_n(y) \geq \varphi(y)$.

Analogously for every $y' > y$ exists an x with $P(X = x) = 0$ and $\varphi(y') < x < \varphi(y') + \epsilon$ so that $y < y' \leq (F \circ \varphi)(y') \leq F(x)$. Since F is **continuous** in x we have $\lim\limits_{n\to\infty} F_n(x) = F(x)$ so that for n large enough $y \leq F_n(x)$ holds whence $\varphi_n(y) \leq x < \varphi(y') + \epsilon$ and consequently $\limsup\limits_{n\to\infty} \varphi_n(y) \leq \varphi(y')$ for $y < y'$. Hence $\lim\limits_{n\to\infty} \varphi_n(y) = \varphi(y)$ if φ is **continuous** at y. Since φ is nondecreasing on $]0;1[$ it has at most countably many points y_k with $\lim\limits_{n\to\infty} \varphi\left(y_k - \frac{1}{n}\right) < \varphi(y_k)$ and we may simply define $\varphi(y_k) = \varphi(y_k) = 0$ to obtain $\lim\limits_{n\to\infty} \varphi_n(y) = \varphi(y)$ for every $y \in]0;1[$ without changing their distribution.

3.6 Convergence in measure and in distribution

A sequence $(X_n)_{n\geq 1}$ of **random variables** $X_n : \Omega \to \mathbb{R}$ on a measure space $(\Omega; \mathcal{A}; P)$ **converging in measure** to a random variable $X : \Omega \to \mathbb{R}$ also **converges in distribution** to X, i.e. at every **point of continuity** t the **distribution functions** F_n defined by $F_n(t) = P(X_n \leq x)$ converge to F defined by $F(x) = P(X \leq x)$: $\lim\limits_{n\to\infty} F_n(x) = F(x)$.



Proof: For every $\epsilon > 0$ we have $\lim\limits_{n\to\infty} P(|X_n - X| > \epsilon) = 0$ and also for every $n \geq 1$ the inequality $P(X \leq x - \epsilon) - P(|X - X_n| \geq \epsilon) \leq P(X_n \leq x) \leq P(X \leq x + \epsilon) + P(|X_n - X| \geq \epsilon)$. For $n \to \infty$ and then $\epsilon \to 0$ we obtain $P(X < x) \leq \liminf\limits_{n\to\infty} P(X_n \leq x) \leq \limsup\limits_{n\to\infty} P(X_n \leq x) \leq P(X \leq x)$. Hence for every point $x \in \mathbb{R}$ of **(left) continuity** with $P_X(\{x\}) = P_X\left(\bigcap_{n\geq 1} \left]x - \frac{1}{n}; x\right]\right) \overset{2.2.3}{=} \lim\limits_{n\to\infty} P_X\left(\left]x - \frac{1}{n}; x\right]\right) = \lim\limits_{n\to\infty}\left(F(x) - F\left(x - \frac{1}{n}\right)\right) = 0$ we have $\lim\limits_{n\to\infty} F_n(x) = \lim\limits_{n\to\infty} P(X_n \leq x) = P(X \leq x) = F(x)$.

3.7 The weak law of large numbers

For the **mean values** $\frac{1}{n}S_n = \frac{1}{n}\sum\limits_{k=1}^{n} X_k$ of every sequence $(X_k)_{k\geq 1}$ of **independent, identically distributed** and **integrable random variables** with expectations $\mu = E(X_1)$ the following statements concerning their asymptotic behaviour hold:

1. **P-almost sure convergence**: $\lim\limits_{n\to\infty} \frac{1}{n}S_n = \mu$
   due to the **strong law of large numbers** 1.12.
2. **Convergence in measure**: $\lim\limits_{n\to\infty} P\left(\left|\frac{1}{n}S_n - \mu\right| \leq \epsilon\right) = 0$ for every $\epsilon > 0$
   due to **Lebesgue's convergence theorem** [4, th. 4.11]
3. **Weak convergence**: $\lim\limits_{n\to\infty} F_n(x) = \lim\limits_{n\to\infty} P\left(\frac{1}{n}S_n \leq t\right) = \begin{cases} 1 & \text{for } x > \mu \\ 0 & \text{for } x < \mu \end{cases} = P\left(\lim\limits_{n\to\infty}\frac{1}{n}S_n \leq t\right) = F(x)$ for every **point of continuity** $x \neq \mu$
   due to the preceding theorem 3.6.

Note: Concerning the asymptotic behavoiur at the **point of discontinuity** $x = \mu$ the **strong law of large numbers** asserts that P-a.e. $\lim\limits_{n\to\infty}\frac{1}{n}S_n = \mu$ whence $P\left(\lim\limits_{n\to\infty}\frac{1}{n}S_n \leq \mu\right) = 1$. Choosing a **symmetric** distribution e.g. $P(X_k = 0) = P(X_k = 2\mu) = \frac{1}{2}$ we obtain $P(X_k \leq \mu) = P(X_k \geq \mu) = \frac{1}{2}$ whence $P\left(\lim\limits_{n\to\infty}\frac{1}{n}S_n \leq \mu\right) = \frac{1}{2}$ for every $n \geq 1$ such that $F_n(\mu)$ **does not converge** to $F(\mu)$.

3.8 The Helly-Bray theorem

For **probability measures** $P_n; P : \mathcal{B}(\mathbb{R}) \to [0;1]$ with **distribution functions** $F_n; F : \mathbb{R} \to [0;1]$ defined by $F_n(x) = P_n(]-\infty; x])$ resp. $F(x) = P(]-\infty; x])$ and the **linear functionals** $\Lambda_n; \Lambda \in \mathcal{C}_b^*(\mathbb{R};\mathbb{R})$ defined in [3, def. 5.8] by $\Lambda_n f = \int f dP_n$ resp. $\Lambda f = \int f dP$ for $f \in \mathcal{C}_b(\mathbb{R};\mathbb{R})$ the following three conditions are equivalent:

1. $(P_n)_{n\geq 1}$ converges **in distribution** to P, i.e.
   $\lim\limits_{n\to\infty} F_n(x) = F(x)$ at every **continuity point** $x \in \mathbb{R}$ of F.
2. $(P_n)_{n\geq 1}$ **weakly** converges to P, i.e.
   $\lim\limits_{n\to\infty} \int f dP_n = \int f dP$ for every **bounded** and **continuous** $f \in \mathcal{C}_b(\mathbb{R};\mathbb{R})$.
3. $(\Lambda_n)_{n\geq 1}$ **weakly*** converges to Λ.
4. $\lim\limits_{n\to\infty} P_n(A) = P(A)$ for every **λ-continuity set** $A \in \mathcal{B}(\mathbb{R})$ with $\lambda(\delta A) = 0$.

Note: The parts 2. - 4. are a corollary to the **Portmanteau theorem** [4, th. 11.5].

Proof:

1. $\Rightarrow$ 2. : According to **Skorohod's theorem** 3.5 for the **quantile functions** $\varphi_n; \varphi :]0;1[\to \mathbb{R}$ with $P_n = \varphi_n \circ \lambda$ resp. $P = \varphi \circ \lambda$ and every $0 < y < 1$ holds $\lim_{n\to\infty} (f \circ \varphi_n)(y) = (f \circ \varphi)(y)$ whence by the **mapping theorem** [4, th. 11.7] follows $\lim_{n\to\infty} (f \circ \varphi_n)(y) = (f \circ \varphi)(y)$ at every **point of continuity** $y \in]0;1[$, hence λ-a.e. Excluding the **countable λ-null set of discontinuities** according to 3.1 we infer λ-a.e. $\lim_{n\to\infty} f \circ \varphi_n = f \circ \varphi$ whence by the **dominated convergence** theorem [4, th. 5.14] follows $\lim_{n\to\infty} \int f dP_n = \lim_{n\to\infty} \int f d\lambda_{\varphi_n} = \lim_{n\to\infty} \int (f \circ \varphi_n) d\lambda = \lim_{n\to\infty} \int (f \circ \varphi) d\lambda = \lim_{n\to\infty} \int f d\lambda_\varphi = \int f dP$.

2. $\Rightarrow$ 1. : For $x < y$ consider the function $f : \mathbb{R} \to [0;1]$ defined by

$$f(t) = \begin{cases} 1 & \text{for } t \le x \\ \frac{y-t}{y-x} & \text{for } x \le t \le y \\ 0 & \text{for } y \le t \end{cases}$$

Since $\chi_{]-\infty;x]} \le f \le \chi_{]-\infty;y]}$ we have $\limsup_{n\to\infty} F_n(x) = \limsup_{n\to\infty} \int \chi_{]-\infty;x]} dP_n \le \lim_{n\to\infty} \int f dP_n = \int f dP = \int \chi_{]-\infty;y]} dP = F(y)$ and since this is true for every $y > x$ we obtain $\limsup_{n\to\infty} F_n(x) \le F(x)$. Similarly for $y < x$ holds $F(y) \le \liminf_{n\to\infty} F_n(x)$ and hence $\lim_{n\to\infty} F\left(x - \frac{1}{n}\right) \le \liminf_{n\to\infty} F_n(x)$, i.e. convergence at every point of continuity.

1. $\Leftrightarrow$ 3. : Follows directly from the definition of **weak* convergence** in [3, def. 5.8].

1. $\Rightarrow$ 4. : Follows directly from [4, th. 11.7] since according to the hypothesis $f = \chi_A$ is λ-a.e. **continuous**.

4. $\Rightarrow$ 1. : Obvious since $\delta(]-\infty;t]) = \{t\}$.

3.9 Helly's selection theorem

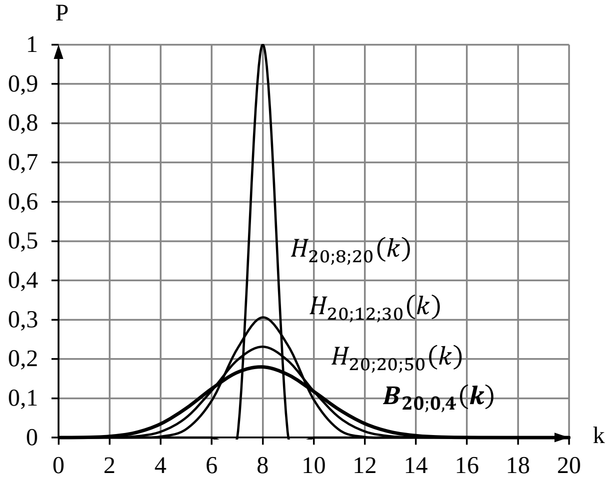
Every **tight** sequence $(P_n)_{n\in\mathbb{N}}$ of **probability measures** on the real numbers $P_n : \mathcal{B}(\mathbb{R}) \to [0;1]$ includes a **subsequence weakly converging to a probability measure** $P : \mathcal{B}(\mathbb{R}) \to [0;1]$ iff it is **tight**, i.e. for every $\epsilon > 0$ exists numbers $a_\epsilon < b_\epsilon \in \mathbb{R}$ such that $P_n(]a_\epsilon; b_\epsilon]) = F_n(b_\epsilon) - F_n(a_\epsilon) > 1 - \epsilon$ for every $n \in \mathbb{N}$.

Note: This is a corollary to **Prohorov's theorem** [4, th. 11.10].

Proof: With the **distribution functions** $F_n : \mathbb{R} \to [0;1]$ defined as usual by $F_n(x) = P_n(]-\infty;x])$ according to the **diagonal principle** [4, th. 11.8] there is a sequence $(n_k)_{k\in\mathbb{N}}$ such that the limit $G(r) = \lim_{k\to\infty} F_{n_k}(r)$ exists for every rational $r \in \mathbb{Q}$. Then $F : \mathbb{R} \to [0;1]$ with $F(x) = \inf\{G(r) : r > x\}$ is **nondecreasing** and obviously **right continuous**. If F is continuous at $x \in \mathbb{R}$ for every $\epsilon > 0$ there is an $y < x$ such that $F(y) > F(x) - \epsilon$. Furthermore there are rational $r, s \in \mathbb{Q}$ with $y < r < x < s$ such that $F(x) - \epsilon < F(r) \le F(x) \le F(s) < F(x) + \epsilon$ whence $F(x) - \frac{\epsilon}{2} < F_{n_k}(r) \le F_{n_k}(x) \le F_{n_k}(s) \le F(x) + \frac{\epsilon}{2}$ for every $k \ge K$ and some $K \in \mathbb{N}$. Thus $F(x) = \lim_{k\to\infty} F_{n_k}(x)$ at every point x of continuity of F. Due to the **tightness hypothesis** for every $\epsilon > 0$ we can find **continuity points** $a < b$ such that $F(b) - F(a) = \lim_{n\to\infty} (F_n(b) - F_n(a)) \ge 1 - \epsilon$. On account of the nondecreasing character of F follows $\lim_{m\to\infty} (F(m)) - F(-m) \ge 1$ and since $0 \le F(x) \le 1$ for all $x \in \mathbb{R}$ we arrive at $\lim_{m\to\infty} F(m) = 1$ resp. $\lim_{M\to\infty} F(-m) = 0$.

3.10 Approximation of the Binomial distribution

The random variable $S_n : \Omega \to \mathbb{N}$ denotes the number of red balls when $n \geq 1$ balls are drawn **without replacement** from an urn containing M red balls and $N-M$ black ones. Its distribution is **hypergeometric** with



$$P\left(S_n = k\right)$$

$$= H_{n,M,N}\left(k\right)$$

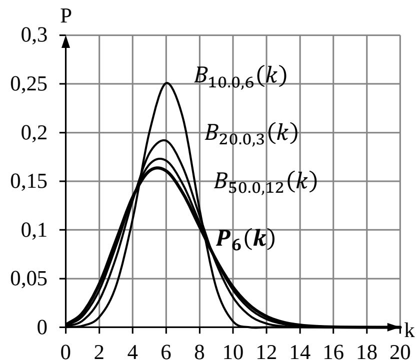
$$= \frac{\begin{pmatrix} M \\ k \end{pmatrix} \cdot \begin{pmatrix} N-M \\ n-k \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}}$$

$$= \begin{pmatrix} n \\ k \end{pmatrix} \cdot \underbrace{\frac{M \cdot \ldots \cdot (M-K)}{N \cdot \ldots \cdot (N-K)}}_{\to p^k} \cdot \underbrace{\frac{(N-M) \cdot \ldots \cdot (N-M-n-k+1)}{(N-K-1) \cdot \ldots \cdot (N-n+1)}}_{\to (1-p)^{n-k}}$$

and for large N the replacement becomes irrelevant such that the hypergeoemtric distribution converges in to the **Binomial distribution:** $\lim\limits_{N\to\infty} H_{n,M,N}\left(k\right) = B_{n,p}\left(k\right)$ with $p = \frac{M}{N}$.

3.11 Approximation of the Poisson distribution

For identically and independently distributed random variables $X_i : \Omega \to \{0;1\}$ with $P\left(X_i = 1\right)$ and $P\left(X_i = 0\right) = 1-p$ and $S_n = \sum\limits_{i=1}^{n} X_i$ we have the **Binomial distribution**



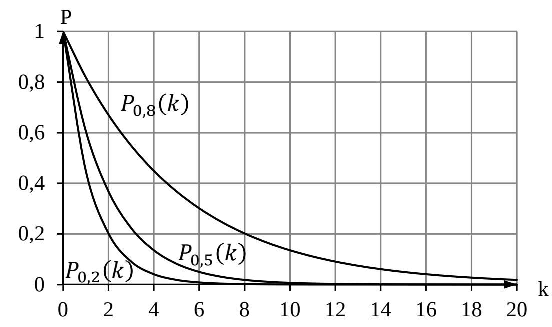
$$P\left(S_n = k\right) = B_{n,k}\left(k\right) = \begin{pmatrix} n \\ k \end{pmatrix} p^k \left(1-p\right)^{n-k}$$

. For $\lambda = n \cdot p$ and fixed k the limit

$$\begin{aligned}
\lim_{n\to\infty} B_{n,k}\left(k\right) &= \lim_{n\to\infty} \begin{pmatrix} n \\ k \end{pmatrix} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\
&= \lim_{n\to\infty} \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \cdot \frac{1}{\left(1 - \frac{\lambda}{n}\right)^k} \cdot \prod_{i=0}^{k-1} \left(1 - \frac{i}{n}\right) \\
&= \frac{\lambda^k \cdot e^\lambda}{k!} \cdot \lim_{n\to\infty} \prod_{i=0}^{k-1} \frac{1 - \frac{i}{n}}{1 - \frac{\lambda}{n}} \\
&= \frac{\lambda^k \cdot e^\lambda}{k!} \\
&= P_\lambda\left(k\right)
\end{aligned}$$

is the **Poisson distribution**. By **Scheffés theorem** [4, p. 9.15] the limit extends to every measurable set whence we obtain **convergence in distribution**.

3.12 The exponential distribution



We define the **independently and identically distributed** random variables $X_t : \Omega \to \{0;1\}$ by $X_t(\omega) = 1$ if an event (e.g. the arrival of a customer at a queue or the call at a telephone exchange) occurs at time $t > 0$ with $P(X_t = 1) = \alpha$. For reasons of compatibility we assume that P-a.s. no event occurs at the start: $P(X_0 = 0) = 1$. Also the number of events in a finite time interval shall always be finite: $\forall \omega \in \Omega\, \forall t > 0 :$ $\text{card}\{0 \le \tau \le t : X_t(\omega) = 1\} < \infty$. Hence we can define the **sample path** $N_t : \Omega \to \mathbb{N}$ by the finite sum $N_t(\omega) = \sum_{\tau=0}^{t} X_\tau$. The **waiting times** $\Delta T_n : \Omega \to \mathbb{R}_0^+$ between the $n-1$ th and the n th event are $\Delta T_n = T_n - T_{n-1}$ for $n \ge 1$ with the **arrival times** $T_n = \inf\{\tau \ge 0 : N_\tau = n\}$ resp. $T_0 = 0$ are also **identically dristributed** since $P(\Delta T_n > s) = P\left(\bigcap_{T_{n-1} \le \tau \le T_{n-1}+s} \{X_\tau = 0\}\right) = P\left(\bigcap_{0 \le \tau \le s} \{X_\tau = 0\}\right) = P(\Delta T_1 > s)$. Since the events $\{\Delta T_n > s\} = \bigcap_{T_{n-1} \le \tau \le T_{n-1}+s} \{X_\tau = 0\}$ and $\bigcap_{T_{n-1}+s < \tau \le T_{n-1}+s+t} \{X_\tau = 0\}$ are **independent** and $P(X_0 = 0) = 1$ there is **no memory effect**, i.e.

$$
\begin{aligned}
P(\Delta T_n > s+t) &= P\left(\bigcap_{0 \le \tau \le s+t} \{X_\tau = 0\}\right) \\
&= P\left(\left(\bigcap_{\tau \le s} \{X_\tau = 0\}\right) \cap \left(\bigcap_{s < \tau \le s+t} \{X_\tau = 0\}\right)\right) \\
&= P\left(\bigcap_{\tau \le s} \{X_\tau = 0\}\right) \cdot P\left(\bigcap_{s < \tau \le s+t} \{X_\tau = 0\}\right) \\
&= P\left(\bigcap_{\tau \le s} \{X_\tau = 0\}\right) \cdot P\left(\bigcap_{0 < \tau \le t} \{X_\tau = 0\}\right) \\
&= P\left(\bigcap_{\tau \le s} \{X_\tau = 0\}\right) \cdot P\left(\bigcap_{0 \le \tau \le t} \{X_\tau = 0\}\right) \\
&= P(\Delta T_n > s) \cdot P(\Delta T_n > t)
\end{aligned}
$$

This is the **functional equation** of the **exponential function** and since $0 \le P \le 1$ with scaling factor $P(\Delta T_n > 0) = P(\Delta T_1 > 0) = P(X_0 = 0) = 1$ the exponent must be negative whence $P(\Delta T_n > t) = e^{-\alpha t}$. Also we have $\lim_{t \to 0} \frac{1}{t} P\left(\bigcup_{\tau \le t} \{X_t = 1\}\right) = \lim_{t \to 0} \frac{1}{t} P(\Delta T_n \le t) = \lim_{t \to 0} \frac{1}{t}\left(1 - e^{-\alpha t}\right) = \frac{dF}{dt}(0) = \alpha$. Note that the probability of an event occurring at a **fixed time** is $P(X_t = 1) = 0$. The **distribution function** $F : \mathbb{R}_0^+ \to [0;1]$ for the waiting times satisfies the functional equation $1 - F(s+t) = (1 - F(s)) \cdot (1 - F(t))$ with the explicit formula

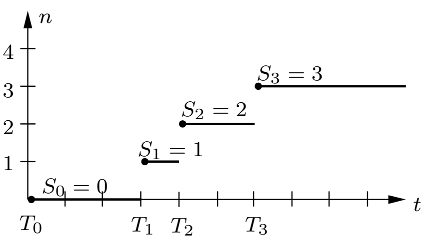
$$
F(t) = P(\Delta T_n \le t) = \begin{cases} 0 & \text{if } t \le 0 \\ 1 - e^{-\alpha t} & \text{if } t > 0 \end{cases}
$$

and **mean waiting time**

$$
E(\Delta T) = \int_\Omega \Delta T dP = \int_{[0;\infty[} t dF(t) = \int_{[0;\infty[} t \cdot \frac{dF}{d\lambda} d\lambda(t) = \int_{[0;\infty[} \alpha t \cdot e^{-\alpha t} dt = \frac{1}{\alpha}.
$$

3.13 The Poisson process

Every **stochastic process** $N : \mathbb{R}_0^+ \times \Omega \to \mathbb{N}$ defined by the **measurable** number $N_I(\omega) \in \mathbb{N}$ of events or **increments** occurring in the time intervall $I \subset \mathbb{R}_0^+$ and in particular $N_t = N_{[0;t]}$ with **arrival times** $T_n(\omega) = \inf\{\tau \geq 0 : N_\tau(\omega) = n\}$ and **waiting times** $\Delta T_n = T_n - T_{n-1}$ satisfying the following conditions:



1. **start** $N_0(\omega) = 0$ for **every** $\omega \in \Omega$
2. **nondecreasing càdlàg sample paths** $t \mapsto N_t(\omega)$ for **every** $\omega \in \Omega$
3. P**-a.s. single events:** $P\left(N_t - \sup\limits_{s<t} N_s \leq 1\right) = 1$
4. P**-a.s. no accumulations**: $P(N_t - N_s < \infty) = 1$ for every $s < t$
5. **independent** occurrence $P(N_{I\cup J} = 0) = P(N_I = 0) \cdot P(N_J = 0)$ for any **disjoint** intervals $I, J \subset \mathbb{R}_0^+$
6. **identically distributed** occurrence $P(N_I = 0) = P(N_J = 0)$ for any **disjoint** intervals $I, J \subset \mathbb{R}_0^+$ of **equal length**

has

1. **Poisson distributed increments** with $P(N_t = n) = e^{-\alpha t} \cdot \frac{(\alpha t)^n}{n!}$ and
2. **Exponentially distributed waiting times** with $P(\Delta T_n > t) = e^{-\alpha t}$.

Proof: Note that we **assume** the **existence** and **measurability** of the random variables N_t on a suitable **measure space** $\left(\mathbb{R}_0^+ \times \Omega; \mathcal{F}; P\right)$. The construction of the corresponding σ-algebra $\mathcal{F}$ requires **Kolmogorov's existence theorem** and is not the subject of this proof.

Proof of 1.: Let $p(t) = P(N_t \geq 1)$, $q(t) = 1 - p(t)$ and $q = q(1)$. From 4. and 5. follows $q\left(\frac{k}{n}\right) = q^{k/n}$ for every rational $\frac{k}{n} > 0$. Due to 2. this relation extends to real $t > 0$ since $q(t) = P(N_t = 0) = P(N_t < 1) = P\left(\inf\limits_{k/n>t} N_{k/n} < 1\right) = P\left(\bigcup_{k/n>t}\left\{N_{k/n} < 1\right\}\right) = \sup\limits_{k/n>t} P\left(N_{k/n} < 1\right) = \sup\limits_{k/n>t} q^{k/n} = q^t$. On account of 5. we have $P\left(N_{]s;s+t[} = 0\right) = q^t$ with $q > 0$ since otherwise we had $P\left(N_{]s;s+t[} \geq 1\right) = 1$ for every $t > 0$ whence $P\left(N_{]s;s+t[} = \infty\right) = 1$ contrary to 4. Hence we obtain $P(\Delta T_n > t) = P\left(N_{]s;s+t[} = 0\right) = e^{-\alpha t}$ with $\alpha = -\ln(q)$.

Proof of 2.: Let $N_{n,t}(\omega) = \sum\limits_{i=1}^{n} \chi_{N_{I_i} \geq 1}$ the number of intervals $I_i = \left]\frac{t(i-1)}{n}; \frac{ti}{n}\right]$ with length $\frac{t}{n}$ and $P(N_{I_i} \geq 1) = 1 - e^{-\frac{\alpha t}{n}}$ in a disjoint partition of $]0;1]$ with at least an occurrence. Owing to 5. and 6. we have $P(N_{n,t} = k) = \begin{pmatrix} n \\ k \end{pmatrix} \left(1 - e^{-\frac{\alpha t}{n}}\right)^k \cdot \left(e^{-\frac{\alpha t}{n}}\right)^{n-k}$ whence the **Poisson approximation** 3.11 yields $\lim\limits_{n\to\infty} P(N_{n,t} = k) = \lim\limits_{n\to\infty} \begin{pmatrix} n \\ k \end{pmatrix} \left(1 - e^{-\alpha\frac{t}{n}}\right)^k \cdot \left(e^{-\alpha\frac{t}{n}}\right)^{n-k} = \lim\limits_{n\to\infty} \begin{pmatrix} n \\ k \end{pmatrix} \left(\frac{\alpha t}{n}\right)^k \left(1 - \frac{\alpha t}{n}\right)^{n-k} = P_{\alpha t}(k)$. According to 4. for every $t > 0$ there is an $n \geq 1$ such that the probability for the event $D_{t,n} = \bigcap_{\Delta T_k \leq t}\left\{\Delta T_k > \frac{t}{n}\right\}$ of **every** waiting time between two events in the interval $]0;t]$ exceeding $\frac{t}{n}$ is $P(D_{t,n}) = 1$. Hence the sequence $(D_{t,n})_{n\geq 1}$ of cases $D_{t,n} = \bigcap_{\Delta T_k \leq t}\left\{\Delta T_k > \frac{t}{n}\right\} \subset \Omega$ is increasing with $\lim\limits_{n\to\infty} P(D_{t,n}) = P\left(\bigcup_{n\geq 1} D_{t,n}\right) = 1$. Since $N_{n,t}(\omega) = N_t(\omega)$ for every $\omega \in D_{t,n}$ we have $P(N_{n,t} \neq N_t) = 1 - P(D_{n,t})$ whence P-a.e. $\lim\limits_{n\to\infty} N_{n,t} = N_t$. By **Lebesgue's convergence theorem** [4, p. 4.11], theorem 3.6 and **Helly-Bray** 3.8.2 we infer $P(N_t = k) = \lim\limits_{n\to\infty} P(N_{n,t} = k) = P_{\alpha t}(k)$.

3.14 Characteristic functions of independent random variables

The **characteristic function** of a **real random variable** $X : \Omega \to \mathbb{R}$ is the **Fourier transform** $\hat{\varphi}_X(\xi) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \varphi_X(x) \cdot e^{-ix\xi} dx = \frac{1}{\sqrt{2\pi}} \int e^{-ix\xi} dP_X = \frac{1}{\sqrt{2\pi}} E\left(e^{-iX\xi}\right)$ of its **probability density function** $\varphi_X = \frac{dP_X}{d\lambda}$ according to [2, p. 3.6]. The properties of the Fourier transform translate into the following equations for characteristic functions:

1. $\hat{\varphi}_{X/\sigma}(\xi) = \sigma \cdot \hat{\varphi}_X(\sigma\xi)$ from [2, p. 3.8.2]
2. $\hat{\varphi}_{X+\mu}(\xi) = e^{-i\mu t} \cdot \hat{\varphi}_X(\xi)$ from [2, p. 3.8.3]
3. $(x^n \cdot \varphi)^\wedge(\xi) = i^n \frac{\delta^n \hat{\varphi}}{\delta \xi^n}(\xi)$ for $x^n \varphi \in L^1$ from [2, p. 3.8.6]

For **independent** random variables X, Y the **exponentials** $e^{-iX\xi}, e^{-iY\xi}$ are again **independent** for every $\xi \in \mathbb{R}$ such that from 1.5 and [2, p. 3.8.4] we obtain

4. $\hat{\varphi}_{X+Y}(\xi) = \frac{1}{\sqrt{2\pi}} E\left(e^{-iX\xi}\right) \cdot E\left(e^{-iY\xi}\right) = \sqrt{2\pi} \cdot \hat{\varphi}_X(\xi) \cdot \hat{\varphi}_Y(\xi) = (\varphi_X * \varphi_Y)^\wedge(\xi)$.

Due to 3.14.3 the n-th **moments** $E(X^n) = \int x^n \cdot \varphi(x)\, dx$ are of prominent interest in the **Taylor expansion** of probability density functions used in the proof of the **central limit theorem** 3.20. Due to [4, p. 6.6.1] the k-th moments exist for all $k \leq n$ if the n-th **absolute moment** $E(|X|^n) < \infty$ is finite. Obviously for the **normal density** all moments are finite so that integrating by parts we obtain $E(X) = 0$, $E\left(X^k\right) = \frac{1}{\sqrt{2\pi}} \int x^k \cdot e^{-x^2/2} dx = \frac{k-1}{\sqrt{2\pi}} \int x^{k-2} \cdot e^{-x^2/2} dx = (k-1) \cdot E\left(X^{k-2}\right)$ for all $k \geq 2$ whence $E\left(X^{2k}\right) = 1 \cdot 3 \cdot \ldots \cdot (2k-1)$ and $E\left(X^{2k+1}\right) = 0$.

3.15 Laplace transforms and moment generating functions

The **moment generating function** of a **real random variable** $X : \Omega \to \mathbb{R}$ is the **Laplace transform** $L\varphi_X : U \to \mathbb{C}$ of its **probability density function** $\varphi_X = \frac{dP_X}{d\lambda}$ defined by $L\varphi_X(\xi) = \int_{\mathbb{R}} \varphi_X(x) \cdot e^{x\xi} dx = \int e^{X\xi} dP = E\left(e^{X\xi}\right)$ for every $\xi \in U \subset \mathbb{C}$ provided that the integral is finite. The **Fourier transform** is a special case of the Laplace transform with imaginary $\xi \in i\mathbb{R}$. For $\mathrm{Re}\xi \geq 0$ we have $\int_{\mathbb{R}^-} \varphi_X(x) \cdot e^{x\xi} dx < \infty$ and for $0 < \mathrm{Re}\xi_1 < \mathrm{Re}\xi_2$ holds $\left|\int_{\mathbb{R}^+} \varphi_X(x) \cdot e^{x\xi_1} dx\right| < \left|\int_{\mathbb{R}^+} \varphi_X(x) \cdot e^{x\xi_2}\right| dx$ whence $L\varphi_X(\xi) < \infty$ for $0 \leq \mathrm{Re}\xi \leq \mathrm{Re}\xi_0$ and since the analogous estimates hold for $\mathrm{Re}\xi \leq 0$ the integral converges for $|\mathrm{Re}\xi| \leq \xi_0$ for some $\xi_0 \geq 0$. As the discrete probability measure $P : \mathcal{P}(\mathbb{Z}) \to [0;1]$ with $P(z) = \frac{\pi^2}{12z^2}$ for $z \neq 0$ and $P(0) = 0$ shows the area of convergence may actually be restricted to the **imaginary axis**, i.e. the **Fourier transform**.

In the case of convergence in the strip $\{|\mathrm{Re}\xi| \leq \xi_0\}$ for some $\xi_0 > 0$ we have an P_X-integrable majorant $e^{x\xi_0} + e^{-x\xi_0}$ for every $\left|X^k\right| \leq \sum_{k\geq 0} \frac{|\xi|^k}{k!} \left|X^k\right| = e^{|x\xi|}$such that all moments $E\left(X^k\right) < \infty$ exist and the **dominated convergence theorem** [4, p. 5.14] gives

$$L\varphi_X(\xi) = \int e^{X\xi} dP = \int \sum_{k\geq 0} \frac{(X\xi)^k}{k!} dP = \sum_{k\geq 0} \frac{E\left(X^k\right)}{k!} \xi^k.$$

By the **Taylor expansion** [2, p. 1.13] we conclude that $\frac{d^k L\varphi_X}{d\xi^k}(0) = E\left(X^k\right)$. Furthermore the measure P_{X_ξ} defined by $P(X_\xi < y) = \int_{\{X<y\}} \frac{e^{\xi X}}{L\varphi_X(\xi)} dP = \int_{-\infty}^{y} \frac{e^{\xi x} \cdot \varphi_X(x)}{L\varphi_X(\xi)} dx$ has the Laplace transform $L\varphi_{X_\xi}(\eta) = \int e^{\eta x} \cdot \frac{e^{\xi x} \cdot \varphi_X(x)}{L\varphi_X(\xi)} dx = \frac{L\varphi_X(\xi+\eta)}{L\varphi_X(\xi)}$ whence $\frac{1}{L\varphi_X(\xi)} \cdot \frac{d^k L\varphi_X}{dx^k}(\xi) = \frac{d^k L\varphi_{X_\xi}}{dx^k}(0) = E\left(X_\xi^k\right) = \int \frac{x^k \cdot e^{\xi x} \cdot \varphi_X(x)}{L\varphi_X(\xi)} dx$. Thus we obtain

$$\frac{d^k L\varphi_X}{dx^k}(\xi) = \int X^k \cdot e^{\xi X} dP = \int x^k \cdot e^{\xi x} \cdot \varphi(x)\, dx \text{ for } |\mathrm{Re}\xi| \leq \xi_0.$$

3.16 Moments and the characteristic function

For a **real random variable** $X : \Omega \to \mathbb{R}$ with **k-th absolute moments** $E\left(|X|^k\right) < \infty$ and $\xi \in \mathbb{R}$ we have

1. $\left|\hat{\varphi}_X(\xi) - \sum_{k=0}^{n} \frac{E(X^k)}{k!}(i\xi)^k\right| \leq E\left(\min\left\{\frac{|\xi X|^{n+1}}{(n+1)!}; \frac{2|\xi X|^n}{n!}\right\}\right)$.
2. $\frac{d^k}{dx^k}\hat{\varphi}_X(\xi) = E\left((iX)^k \cdot e^{i\xi X}\right)$, in particular $\frac{d^k}{dx^k}\hat{\varphi}_X(0) = i^k E\left(X^k\right)$
3. In the case of a **finite Laplace transform** $L\varphi_X(r) = \sum_{k\geq 0}\frac{E(X^k)}{k!}r^k < \infty$ for **some** $r > 0$ the random variable X is **uniquely determined** by its moments $E\left(X^k\right)$ for $k \geq 1$.

Proof:

1.: An **integration by parts** [2, p. 1.5] of the remainder of the **Taylor expansion** [2, p. 2.2] for the complex valued exponential function yields

$$\begin{aligned} e^{ix} =& \sum_{k=0}^{n} \frac{(ix)^k}{k!} + \frac{i^{n+1}}{n!}\int_0^x (x-t)^n e^{it} dt \\ =& \sum_{k=0}^{n} \frac{(ix)^k}{k!} + \frac{i^n}{(n-1)!}\int_0^x (x-t)^{n-1}\left(e^{it}-1\right) e^{it} dt \end{aligned}$$

whence

$$\left|e^{ix} - \sum_{k=0}^{n}\frac{(ix)^k}{k!}\right| \leq E\left(\min\left\{\frac{|x|^{n+1}}{(n+1)!}; \frac{2|x|^n}{n!}\right\}\right)$$

so that the assertion follows from the definition of the **Fourier transform** [2, p. 4.6].

2.: According to 1. and considering $E(E(X)) = E(X)$ we have

$$\begin{aligned} \left|\frac{\hat{\varphi}_X(\xi+h) - \hat{\varphi}_X(\xi)}{h} - E\left(iXe^{i\xi X}\right)\right| &= \left|\frac{1}{h}E\left(e^{i\xi X}\cdot\left(e^{ihX} - 1 - ihX\right)\right)\right| \\ &\leq \frac{1}{|h|}\cdot E\left(\left|e^{ihX} - 1 - ihX\right|\right) \\ &\leq \frac{1}{|h|}E\left(\min\left\{\frac{|hX|^2}{2}; \frac{2|hX|}{1}\right\}\right) \\ &= E\left(\min\left\{\frac{1}{2}|h|\cdot|X|^2; 2|X|\right\}\right) \end{aligned}$$

With the majorant $2|X|$for $h \to 0$ by **dominated convergence** [4, p. 5.14] we obtain $\frac{d}{d\xi}\hat{\varphi}_X(\xi) = E\left(iXe^{i\xi X}\right)$. Repeating this argument inductively proves the assertion for $k \leq n$ with $E(X^n) < \infty$.

3.: For any $\xi < r$ there is a $k_0 \geq 1$ such that $2k\xi^{2k-1} < r^{2k}$ for $k \geq k_0$. Since $|x|^{2k-1} \leq 1 + |x|^{2k}$ for every $x \in \mathbb{R}$ we have

$$\frac{E\left(|X|^{2k-1}\right)\cdot\xi^{2k-1}}{(2k-1)!} \leq \frac{\xi^{2k-1}}{(2k-1)!} + \frac{E\left(|X|^{2k}\right)\cdot\xi^{2k}}{(2k-1)!} \leq \frac{\xi^{2k-1}}{(2k-1)!} + \frac{E\left(|X|^{2k}\right)\cdot r^{2k}}{(2k)!}$$

such that because of $\sum_{k\geq 0}\frac{E(X^k)}{k!}r^k < \infty$ follows $\lim_{k\to\infty}\frac{E(|X|^k)\cdot\xi^k}{k!} \leq \lim_{k\to\infty}\frac{E(X^k)\cdot r^k}{k!} = 0$. As in the proof of 1. from

$$\left|e^{i\eta x}\left(e^{i\xi x} - \sum_{k=0}^{n}\frac{(i\xi x)^k}{k!}\right)\right| \leq \frac{|\xi x|^{n+1}}{(n+1)!} \text{ and } \frac{d^k}{dx^k}\hat{\varphi}_X(\xi) = E\left((iX)^k\cdot e^{i\xi X}\right) \text{ for } \xi, \eta \in \mathbb{R}$$

we infer

$$\left|\hat{\varphi}_X(\eta+\xi) - \sum_{k=0}^{n} \frac{1}{k!} \cdot \frac{d^k}{dx^k}\hat{\varphi}_X(\eta) \cdot \xi^k\right| \leq \frac{|\xi|^{n+1} \cdot E\left(|X|^{n+1}\right)}{(n+1)!} \text{ for } \xi, \eta \in \mathbb{R}$$

whence

$$\hat{\varphi}_X(\eta+\xi) = \sum_{k=0}^{n} \frac{1}{k!} \cdot \frac{d^k}{dx^k}\hat{\varphi}_X(\eta) \cdot \xi^k \text{ for } |\xi| \leq r \text{ and } \eta \in \mathbb{R}$$

Assuming a second random variable Y with equal moments by analogous arguments we obtain the characteristic function

$$\hat{\varphi}_Y(\eta+\xi) = \sum_{k=0}^{\infty} \frac{1}{k!} \cdot \frac{d^k}{dx^k}\hat{\varphi}_Y(\eta) \cdot \xi^k \text{ for } |\xi| \leq r \text{ and } \eta \in \mathbb{R}$$

In order to show equality we apply a process of **analytic continuation**: According to 2. for $\eta = 0$ we have $\frac{d^k}{dx^k}\hat{\varphi}_X(0) = i^k E\left(X^k\right) = \frac{d^k}{dx^k}\hat{\varphi}_Y(0)$ whence $\hat{\varphi}_X(\xi) = \hat{\varphi}_Y(\xi)$ for $|\xi| \leq r - \epsilon$ and any $\epsilon > 0$. But then $\frac{d^k}{dx^k}\hat{\varphi}_X(\pm(r-\epsilon)) = \frac{d^k}{dx^k}\hat{\varphi}_Y(\pm(r-\epsilon))$ and by expansion around $\pm(r-\epsilon)$ we obtain equality for $|\xi| \leq 2r - \epsilon$ etc.., so that the assertion follows by the uniqueness of the Fourier transform [2, p. 4.13].

Note: The moments $E\left(X^k\right)$of a random variable X give an estimate for the probability of large values resp. deviations from the mean $E(X)$ resp. the weight of their **tails**. They also coincide with the **derivatives of the characteristic function**, i.e. its **smoothness** and hence determine its **asymptotic behavior** and thus its suitability for convergence. The more moments X has, the more derivatives φ_X has. in particular 3 tranlsates into a completion of the **Helly-Bray theorem** 3.8:

3.17 The moment criterion for weak convergence

A sequence $(X_n)_{n\geq 1}$ of **real random variable** $X : \Omega \to \mathbb{R}$ with **k-th absolute moments** $E\left(|X|^k\right) < \infty$ and **finite Laplace transform** $L\varphi_X(r) = \sum_{k\geq 0} \frac{E\left(X^k\right)}{k!} r^k < \infty$ for **some** $r > 0$ **converges in distribution** to a random variable X if **all moments converge** $\lim_{n\to\infty} E\left(X_n^k\right) = E\left(X^k\right)$ for every $k \geq 1$.

3.18 Examples

1. A random variable $X : \Omega \to \mathbb{R}$ with distribution $P(X \leq x) = \Phi_{\mu,\sigma}(x) = \int_{-\infty}^{x} \phi_{\mu,\sigma}(t)\,dt$ for the **normal density function** $\phi_{\mu;\sigma}(t) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right)$ is **normally distributed** and a short calculation involving **change of variables** as well as **integration by parts** according to the definitions in 1.3 yield the **expectation** $E(X) = \mu$ and the **variance** $VAR(X) = \sigma^2$. The **moment generating function** is $M_\varphi(s) = \frac{1}{\sqrt{2\pi}} \int e^{s\xi} \cdot e^{-\xi^2/2} d\xi = \frac{1}{\sqrt{2\pi}} e^{-s^2/2} \int e^{-(\xi-s)^2/2} d\xi = e^{-s^2/2} = \sum_{k\geq 0} \frac{1}{k!}\left(\frac{s^2}{2}\right)^k = \sum_{k\geq 0} \frac{1\cdot 3\cdot \ldots \cdot (2k-1)}{(2k)!} \cdot s^{2k}$ whence we obtain the moments $E\left(X^{2k}\right) = 1 \cdot 3 \cdot \ldots \cdot (2k-1)$ resp. $E\left(X^{2k+1}\right) = 0$ for $k \geq 0$.
2. For the **exponential distribution** from 3.12 for Re$s < \alpha$ we have the **moment generating function** $M_p(s) = \int_0^\infty e^{s\xi} \cdot \alpha e^{-\alpha\xi} d\xi = \frac{\alpha}{\alpha - s} = \sum_{k\geq 0} \frac{s^k}{\alpha^k}$ whence $E\left(X^k\right) = \frac{k!}{\alpha^k}$, in particular **expectation** $E(X) = \frac{1}{\alpha}$ and **variance** $VAR(X) = \frac{1}{\alpha^2}$.
3. For the **Poisson distribution** from 3.13 the **moment generating function** is $M_\lambda(s) = \sum_{k\geq 0} e^{ks} \cdot e^{-\lambda} \frac{\lambda^k}{k!} = e^{\lambda(e^s - 1)}$ whence $\frac{dM}{d\xi}(s) = \lambda e^s \cdot M(s)$ and $\frac{d^2M}{d\xi^2}(s) = \left(\lambda^2 e^{2s} + \lambda \cdot e^s\right) \cdot M(s)$ whence $\frac{dM}{d\xi}(0) = \lambda$ and $\frac{d^2M}{d\xi^2}(0) = \left(\lambda^2 + \lambda\right)$; in particular the **expectation** resp. the **variance** are both $E(X) = VAR(X) = \lambda$.

4. For **independent random variables** X, Y with moment generating functions $M(X)$, $M(Y)$ in $\{|\mathrm{Re}s| \leq s_0\}$ the exponents e^{sX}, e^{sY} are stilll independent such that theorem 1.5 gives

$$M(X+Y) = E\left(e^{s(X+Y)}\right) = E\left(e^{sX} \cdot e^{sY}\right) = E\left(e^{sX}\right) \cdot E\left(e^{sX}\right) = M(X) \cdot M(Y)$$

3.19 Approximation of complex products

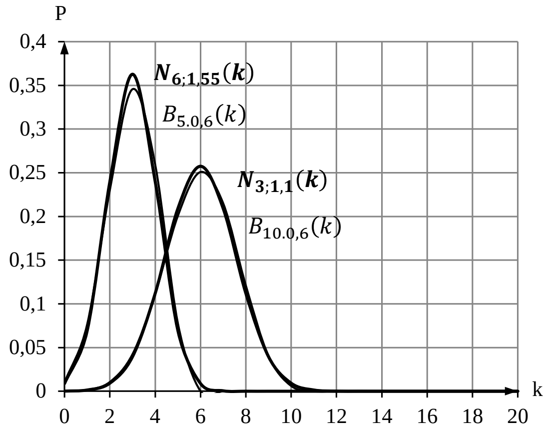
For complex numbers $z_1; ...; z_n; w_1; ...; w_n \in \overline{B_1(0)}$ we have $\left|\prod\limits_{k=1}^{n} z_k - \prod\limits_{k=1}^{n} w_k\right| \leq \sum\limits_{k=1}^{n} |z_k - w_k|$.

Proof: By induction with $\prod\limits_{j=1}^{0} w_j = \prod\limits_{i=n+1}^{n} z_i = 1$ from

$$\begin{aligned}
\prod_{k=1}^{n} z_k - \prod_{k=1}^{n} w_k &= (z_1 - w_1) \cdot \prod_{i=2}^{n} z_i + w_1 \cdot \left(\prod_{i=2}^{n} z_i - \prod_{i=2}^{n} w_i\right) \\
&= (z_1 - w_1) \cdot \prod_{i=2}^{n} z_i + w_1 \cdot (z_2 - w_2) \cdot \prod_{i=3}^{n} z_i + w_1 \cdot w_2 \cdot \left(\prod_{i=3}^{n} z_i - \prod_{i=3}^{n} w_i\right) \\
&\vdots \\
&= \sum_{k=1}^{n} \prod_{j=1}^{k-1} w_j \cdot (z_k - w_k) \cdot \prod_{i=k+1}^{n} z_i
\end{aligned}$$

3.20 The central limit theorem

For a **triangular array** $\left((X_{n;k})_{k \leq k_n}\right)_{n \geq 1}$ of **independent** families $X_{n;1}; ...; X_{n;k_n} : \Omega_n \to \mathbb{R}$ of random variables with



- **sums** $S_n = \sum\limits_{k=1}^{k_n} X_{n;k}$
- **expectations** $\eta_{n;k} = E(X_{n;k}) < \infty$
- **variances** $\sigma_{n;k}^2 = E\left(X_{n;k}^2\right) - E^2(X_{n;k}) < \infty$
- **sum variances** $s_n = E(S_n^2) - E^2(S_n) = \sum\limits_{k=1}^{k_n} \sigma_{n;k}^2$

the **normalized** sums $\overline{S}_n = \sum\limits_{k=1}^{k_n} \overline{X}_{n;k}$ of $\overline{X}_{n;k} = \frac{X_{n;k} - \eta_{n;k}}{s_n}$ with

- **expectations** $E\left(\overline{S}_n\right) = E\left(\overline{X}_{n;k}\right) = 0$
- **variances** $\overline{\sigma}_{n;k}^2 = E\left(\overline{X}_{n;k}^2\right) = \frac{\sigma_{n;k}^2}{s_n^2}$
- **sum variances** $E\left(\overline{S}_n^2\right) = \sum\limits_{k=1}^{k_n} \overline{\sigma}_{n;k}^2 = 1$

satisfying the **Lindeberg condition**

$$\lim_{n \to \infty} \sum_{k=1}^{k_n} \int_{|\overline{X}_{n;k}| \geq \epsilon} \overline{X}_{n;k}^2 dP = 0 \text{ for every } \epsilon > 0$$

converge **weakly** to the **Normal distribution**: $\lim\limits_{n \to \infty} P\left(\overline{S}_n \leq x\right) = \mathcal{N}_{0;1}(]-\infty; x]) = \int_{-\infty}^{x} \phi_{0;1}(\xi)\, d\xi$.

Proof: The **Lindeberg condition** yields $\lim_{n\to\infty}\max_{1\le k\le k_n}\overline{\sigma}^2_{n;k}\le\lim_{n\to\infty}\max_{1\le k\le k_n}\left(\epsilon^2+\int_{\left|\overline{X}_{n;k}\right|\ge\epsilon}\overline{X}^2_{n;k}dP\right)=0$ such that for every $\xi\in\mathbb{R}$ there is an $n_\xi\ge 1$ with $\frac{1}{2}\overline{\sigma}^2_{n;k}\xi^2\le 1$. Hence we can apply

- the lemma 3.19 to both of the **product differences** in the **second** line
- the estimate 3.16.1 to both of the **quadratic order Taylor approxiations** of $\hat{\varphi}_{\overline{X}_{n;k}}(\xi)$ and $e^{-\xi^2\overline{\sigma}^2_{n;k}/2}$ in the **third** line

of the following estimate for fixed $\xi\in\mathbb{R}$ and every $\epsilon>0$:

$$\begin{aligned}\left|\hat{\varphi}_{\overline{S}_n}(\xi)-\hat{\phi}(\xi)\right| &= \left|\prod_{k=1}^{k_n}\hat{\varphi}_{\overline{X}_{n;k}}(\xi)-\prod_{k=1}^{k_n}e^{-\xi^2\overline{\sigma}^2_{n;k}/2}\right| \\ &\le \left|\prod_{k=1}^{k_n}\hat{\varphi}_{\overline{X}_{n;k}}(\xi)-\prod_{k=1}^{k_n}\left(1-\frac{1}{2}\overline{\sigma}^2_{n;k}\xi^2\right)\right|+\left|\prod_{k=1}^{k_n}e^{-\xi^2\overline{\sigma}^2_{n;k}/2}-\prod_{k=1}^{k_n}\left(1-\frac{1}{2}\overline{\sigma}^2_{n;k}\xi^2\right)\right| \\ &\le\sum_{k=1}^{k_n}\left|\hat{\varphi}_{\overline{X}_{n;k}}(\xi)-1+\frac{1}{2}\overline{\sigma}^2_{n;k}\xi^2\right|+\sum_{k=1}^{k_n}\left|e^{-\xi^2\overline{\sigma}^2_{n;k}/2}-1+\frac{1}{2}\overline{\sigma}^2_{n;k}\xi^2\right| \\ &\le\sum_{k=1}^{k_n}E\left(\min\left\{\left|\xi\overline{X}_{n;k}\right|^2;\frac{1}{6}\left|\xi\overline{X}_{n;k}\right|^3\right\}\right)+\sum_{k=1}^{k_n}\xi^4e^{\xi^2}\overline{\sigma}^4_{n;k} \\ &\le\sum_{k=1}^{k_n}\xi^2\int_{\left|\overline{X}_{n;k}\right|<\epsilon}\overline{X}^2_{n;k}dP+\sum_{k=1}^{k_n}\xi^2\int_{\left|\overline{X}_{n;k}\right|\ge\epsilon}\overline{X}^2_{n;k}dP+\xi^4e^{\xi^2}\sum_{k=1}^{k_n}\overline{\sigma}^4_{n;k} \\ &\le\epsilon\xi^2\sum_{k=1}^{k_n}\overline{\sigma}^2_{n;k}+\xi^2\int_{\left|\overline{X}_{n;k}\right|\ge\epsilon}\overline{X}^2_{n;k}dP+\xi^4e^{\xi^2}\sum_{k=1}^{k_n}\overline{\sigma}^4_{n;k}\end{aligned}$$

Due to the Lindeberg condition resp. its consequence $\lim_{n\to\infty}\max_{1\le k\le k_n}\overline{\sigma}^2_{n;k}=0$ all three summand vanish for $n\to\infty$ and by **Lévy's continuity theorem** [2, th. 7.18] the assertion is proved.

3.21 Lyapunov's condition

The **Lyapunov condition** for some $\delta>0$ on the right hand side of the following estimate is **stronger** than the Lindeberg condition but sometimes easier to prove:

$\sum_{k=1}^{k_n}\int_{\left|\overline{X}_{n;k}\right|\ge\epsilon}\overline{X}^2_{n;k}dP\le\frac{1}{\epsilon^\delta}\sum_{k=1}^{k_n}\int_{\left|\overline{X}_{n;k}\right|\ge\epsilon}\left|\overline{X}\right|^{2+\delta}_{n;k}dP\le\sum_{k=1}^{k_n}\int_{\left|\overline{X}_{n;k}\right|\ge\epsilon}\left|\overline{X}\right|^{2+\delta}_{n;k}dP<\infty.$

Note: In [1, th 27.4] a variant of the central limit theorem being very useful for **Markov processes** is proved for sequences in which random variables **far apart from each other** are **nearly independent**.

Index

References

[1] Patrick Billingsley. *Probability and Measure*. 3rd ed. Wiley, 1994.

[2] Arne Vorwerg. "Analysis". In: (2023). URL: http://www.vorwerg-net.de/Mathematik/Analysis.pdf.

[3] Arne Vorwerg. "Functional Analysis". In: (2023). URL: http://www.vorwerg-net.de/Mathematik/Functional%20Analysis.pdf.

[4] Arne Vorwerg. "Measure Theory". In: (2023). URL: http://www.vorwerg-net.de/Mathematik/Measure%20Theory.pdf.