MA 2C, Third Term, 1995

Oliver Knill

ORGANISATION

Course information:

Ma 2c: Spring 95

Instructor: Oliver Knill, 172 Sloan, X 4325 ,
e-mail: knill@cco.caltech.edu

Lectures: MWF at 10:00 AM, in 22 Gates

Text: T. Apostol, Calculus II, (2nd edition) Some notes from the instructor.

Topics: We finish at the beginning with the vector analysis (see chapter 12 of Apostol), and continue with the theory of probability (chapter 13,14) of Apostol. Especially in the probability part, we will also treat some topics **not** treated in Apostol's book.

Recitation: Every Thursday, as scheduled for each section.

Homework: The solutions need to be turned in every Monday on the teaching assistant. Some problems will be from the book of Apostol, some problems will be given by the instructor. The problems will be available at least one week in advance. Only the problems with a (*) need to be done, though it is advisable to do (or at least try) all of them. The problems with a double (**), should be done without consulting others.

Examinations: There will be one Mid-term examination and a final exam, both on a timed, three hours, take-home bases.

Grading policy: The midterm will count toward 30% of the grade, homework 30% and the final exam 40%.

MA 2c:
CALCULUS/PROBABILITY

Oliver Knill, Room 172, Tel.: 4325, e-mail: knill@cco.caltech.edu

Course given in the third term 1995 Caltech: content according to the weeks:

CHAPTER 0: (CALCULUS END OF MA 2b)
1. Week : **1. Surface integrals, 2. Theorem of Stokes, 3. Theorem of Gauss.**

CHAPTER 1: (FINITE PROBABILITY SPACES)
2. Week : **1. Boolean algebras, 2. Probability spaces, 3. Three constructions.**
3. Week : **4. Independent events, 5. Random variables, 6. Expectation.**
4. Week : **7. Variance, Covariance, Correlation. 8. Independent random variables. 9. Random walk and Game systems.**

CHAPTER 2: (GENERAL PROBABILITY SPACES)
5. Week : **1. Cardinality of sets. 2. General probability spaces.**
6. Week : **3. Discrete Random variables, Expectation and variance of discrete random variables. 4. The Borel Cantelli lemma.**
7. Week : **5. Integration and expectation, 6. Distribution of random variables. 7. Examples of continuous distributions 8. Expectation and variance of continuous distributions.**
8. Week : **9. Multidimensional distributions. 10. Transformations of distributions. 11. Characteristic functions. 12 Distributions of sums of independent random variables.**

CHAPTER 3: (LIMIT THEOREMS)
9. Week : **1. Chebychev inequality, 2. Weak law of large numbers.**
10. Week : **3. Strong law of large numbers. 4. Central limit theorem.**

THEORY

1. Week, (Summary of the theory)

NOTATION:

$r = (x, y, z)$ **point in** $\mathbf{R}^3$

$C = \alpha(I)$ **curve,** $\alpha : t \in I = [a, b] \to \alpha(t) \in \mathbf{R}^3$

$S = r(B)$ **surface,** $(u, v) \in B \mapsto r(u, v) \in \mathbf{R}^3$

U **region in** $\mathbf{R}^3$.

δC, δS, δU **boundaries of curve, surface, region**

$F = (P, Q, R)$ **vector field ,** f **scalar field**

DIFFERENTIAL OPERATORS:

$\partial_x f = \frac{\partial}{\partial x} f = f_x$ **(Partial derivative)**

$\nabla = (\partial_x, \partial_y, \partial_z)$ **(Nabla)**

$\text{grad}F = \nabla F = (\frac{\partial F}{\partial x}, \frac{\partial F}{\partial y}, \frac{\partial F}{\partial z})$

$$\text{curl}F = \nabla \wedge F = \begin{pmatrix} \partial_x \\ \partial_y \\ \partial_z \end{pmatrix} \wedge \begin{pmatrix} P \\ Q \\ R \end{pmatrix} = \begin{pmatrix} R_y - Q_z \\ P_z - R_x \\ Q_x - P_y \end{pmatrix}.$$

$\text{div}F = \nabla \cdot F = (\partial_x, \partial_y, \partial_z) \cdot (P, Q, R) = P_x + Q_y + R_z$.

IDENTITIES:

$\text{div curl}F = \nabla \cdot (\nabla \wedge F) = 0$

$\text{curl grad}f = \nabla \wedge \nabla f = 0$.

GEOMETRY OF SURFACE:

$S = r(B)$ **surface.**

r_u, r_v **tangent vectors**

$N = r_u \wedge r_v$ **normal vector**

$n = N/||N||$ **unit normal vector.**

DEFINITIONS:

line integral
$\int_C F \cdot ds = \int_a^b F(\alpha(t)) \cdot \alpha'(t)\, dt$

surface integral
$\int\int_S F \cdot dS = \int\int_B F(r(u, v)) \cdot N(r(u, v))\, dudv$

volume integral
$\int\int\int_U f\, dV = \int\int\int_U f(x, y, z)\, dxdydz$

FUNDAMENTAL THEOREM OF CALCULUS:

$$\int_C \text{grad}f \cdot ds = f(\alpha(a)) - f(\alpha(b)) = \int_{\delta C} f$$

THEOREM OF STOKES:

$$\int\int_S \text{curl}F \cdot dS = \int_{\delta S} F \cdot ds$$

THEOREM OF GAUSS:

$$\int\int\int_U \text{div}F\, dV = \int\int_{\delta U} F\, dS$$

3. Week, (Summary of the theory)

DEFINITION:
$(\Omega, \mathcal{A}, P)$ finite probability space.
$A, B \in \mathcal{A}$ are **independent** if and only if

$$P[A \cap B] = P[A] \cdot P[B] .$$

A finite set $\{A_i\}_{i \in I}$ of events is called **independent** if and only if for all $J \subset I$

$$P[\bigcap_{i \in J} A_i] = \prod_{i \in J} P[A_i] .$$

PROPERTIES:
$A, B \in \mathcal{A}$ are independent, if and only if either $P[B] = 0$ or $P[A|B] = P[A]$.

$(\Omega, \mathcal{A}, P) = (\Delta, \mathcal{B}, Q)^n$ (product space). Given $B_i \in \mathcal{B}$ then

$$A_i = \{\omega = (\omega_1, \dots, \omega_n) \mid \omega_i \in B_i\}$$

are all independent.

DEFINITION: A **random variable** on a finite probability space $(\Omega, \mathcal{A}, P)$ is a map $X : \Omega \to \mathbf{R}$ such that for all $a \in \mathbf{R}$, we have $\{X = a\} \in \mathcal{A}$.

DEFINITION: The **expectation** of a random variable X is defined as

$$E[X] = \sum_{a \in X(\Omega)} a \cdot P[X = a] = \sum_{A \in \mathcal{A}, A \text{ atom}} X(A) \cdot P[A] ,$$

where an **atom** is a set in $\mathcal{A}$ so that $B \subset A, B \in \mathcal{A} \Rightarrow B = A$ or $B = \emptyset$. If $\mathcal{A} = \{A \subset \Omega\}$, then the atoms are all of the form $\{\omega\}$ and

$$E[X] = \sum_{\omega \in \Omega} X(\omega) P[\{\omega\}] .$$

By the definition of a random variable, X must be constant on each atom A and $X(A)$ is defined as the common value, X takes on A. The two expressions for $E[X]$ in the box to the left are seen to be the same using $a = X(A)$ and $P[X = a] = \sum_{A \text{ atom } X(A)=a} P[A]$.

PROPERTIES OF EXPECTATION: For random variables X, Y and $\lambda \in \mathbf{R}$

$E[X + Y] = E[X] + E[Y]$	$E[\lambda X] = \lambda E[X]$
$X \leq Y \Rightarrow E[X] \leq E[Y]$	$E[X^2] = 0 \Leftrightarrow X = 0$
$E[X] = c$ if $X(\omega) = c$ is constant	$E[X - E[X]] = 0 .$

PROOF OF THE ABOVE PROPERTIES:

$E[X + Y] = \sum_{A \text{ atom}} (X + Y)(A) \cdot P[A] = \sum_{A \text{ atom}} (X(A) + Y(A)) \cdot P[A] = E[X] + E[Y]$
$E[\lambda X] = \sum_{A \text{ atom}} (\lambda X)(A) P[A] = \lambda \sum_{A \text{ atom}} X(A) P[A] = \lambda E[X]$
$X \leq Y \Rightarrow X(A) \leq Y(A)$, for all atoms A and $E[X] \leq E[Y]$
$E[X^2] = 0 \Leftrightarrow X^2(A) = 0$ for all atoms A $\Leftrightarrow X = 0$
$X(\omega) = c$ is constant $\Rightarrow E[X] = c \cdot P[X = c] = c \cdot 1 = c$
$E[X - E[X]] = E[X] - E[E[X]] = E[X] - E[X] = 0$

4. Week, (Summary of the theory)

DEFINITION: $(\Omega, \mathcal{A}, P)$ probability space, X, Y random variables.

Variance

$$\mathrm{Var}[X] = \mathrm{E}[\,(X - \mathrm{E}[X])^2\,] .$$

Standard deviation

$$\sigma[X] = \sqrt{\mathrm{Var}[X]} .$$

Covariance

$$\mathrm{Cov}[X,Y] = \mathrm{E}[(X - \mathrm{E}[X]) \cdot (Y - \mathrm{E}[Y])] .$$

Correlation of $\mathrm{Var}[X] \neq 0, \mathrm{Var}[Y] \neq 0$

$$\mathrm{Corr}[X,Y] = \frac{\mathrm{Cov}[X,Y]}{\sigma[X]\sigma[Y]} .$$

$\mathrm{Corr}[X,Y] = 0$: **uncorrelated** X and Y.

PROPERTIES of VAR, COV, and CORR:

$\mathrm{Var}[X] \geq 0$.
$\mathrm{Var}[X] = \mathrm{E}[X^2] - \mathrm{E}[X]^2$.
$\mathrm{Var}[\lambda X] = \lambda^2 \mathrm{Var}[X]$.

$\mathrm{Cov}[X,Y] = \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y]$.
$\mathrm{Cov}[X,Y] \leq \sigma[X]\sigma[Y]$ (Schwarz inequality).

$-1 \leq \mathrm{Corr}[X,Y] \leq 1$.
$\mathrm{Corr}[X,Y] = 1$ if $X - \mathrm{E}[X] = Y - \mathrm{E}[Y]$
$\mathrm{Corr}[X,Y] = -1$ if $X - \mathrm{E}[X] = -(Y - \mathrm{E}[Y])$.

BERNOULLI DISTRIBUTED RANDOM VARIABLES: $(\Omega = \{0,1\}^n, \mathcal{A}, P = Q^n)$, where $Q[\{1\}] = p, Q[\{0\}] = q = 1 - p$.

$$X(\omega) = \sum_{i=1}^{n} \omega_i$$

$$P[X = k] = \binom{n}{k} p^k q^{n-k}$$

$$E[X] = \sum_{k=1}^{n} k \binom{n}{k} p^k q^{n-k} = pn$$

$$\mathrm{Var}[X] = \sum_{k=1}^{n} k^2 \binom{n}{k} p^k q^{n-k} - E[X]^2 = npq$$

DEFINITION: X, Y are **independent** if for all $a, b \in \mathbf{R}$

$$P[X = a; Y = b] = P[X = a] \cdot P[Y = b] .$$

A finite collection $\{X_i\}_{i \in I}$ of random variables are **independent**, if for all $J \subset I$ and $a_i \in \mathbf{R}$

$$P[X_i = a_i, i \in J] = \prod_{i \in J} P[X_i = a_i] .$$

PROPERTIES:

- If X and Y are independent, then $\mathrm{E}[X \cdot Y] = \mathrm{E}[X] \cdot \mathrm{E}[Y]$.
- If X_i is a set of independent random variables, then $\mathrm{E}[\prod_{i=1}^n X_i] = \prod_{i=1}^n \mathrm{E}[X_i]$.
- If X, Y are independent then $\mathrm{Cov}[X,Y] = 0$.
- A constant random variable is independent to any other random variable.

DEFINITION: The **regression line** of two random variables X, Y is defined as $y = ax + b$, where

$$a = \frac{\mathrm{Cov}[X,Y]}{\mathrm{Var}[X]}, b = \mathrm{E}[Y] - a\mathrm{E}[X] .$$

PROPERTY: Given $X, \mathrm{Cov}[X,Y], \mathrm{E}[Y]$, and the regression line $y = ax + b$ of X, Y. The random variable $\tilde{Y} = aX + b$ minimizes $\mathrm{Var}[Y - \tilde{Y}]$ under the constraint $\mathrm{E}[Y] = \mathrm{E}[\tilde{Y}]$ and is the best guess for Y, when knowing only $\mathrm{E}[Y]$ and $\mathrm{Cov}[X,Y]$. We check $\mathrm{Cov}[X,Y] = \mathrm{Cov}[X,\tilde{Y}]$.

5. Week, (Summary of the theory)

DEFINITION: Important example: **One dimensional random walk**
($\Omega = \{-1,1\}^N = \{(\omega = (\omega_1 < \ldots, \omega_N) \mid \omega_i \in \{-1,1\}\}, \mathcal{A} = \{A \subset \Omega\}, \mathrm{P}[A] = |A|/|\Omega|)$.
The random variables $X_k(\omega) = \omega_k$ define the k'th step. The random variables $S_n = \sum_{k=1}^n X_k(\omega)$ describe the location of the random walker (drunken sailor) at time n. If X_k is the win or loss in a game at time k, then S_n is the total win or loss up to time n. Ω is the set of all possible trajectories up to time N.

PROPERTIES Random walk:

a) $I \subset \{1, \ldots, N\}$, $x_i \in \{-1,1\}$

$$\mathrm{P}[X_i = x_i, i \in I] = 2^{-|I|}.$$

b) $\mathrm{E}[X_k] = 0$,
c) $\mathrm{E}[S_k] = 0$.
d) $n + x$ even, $\mathrm{P}[X_n = x] = 2^{-n} \binom{n}{\frac{n+x}{2}}$.
$n + x$ odd, $\mathrm{P}[X_n = x] = 0$.

DEFINITION: A **gambling system** attached to the random walk is sequence of random variables V_k such that every event $\{V_n = c\}$ is a union of sets of the form $\{\omega_1 = x_1, \ldots, \omega_{n-1} = x_{n-1}\}$.
Let V_k be a gambling system, then

$$S_n^V = \sum_{i=1}^n V_k X_k$$

is the **total winnings** with this system.

PROPERTY of gambling systems:
You can't beat the system: $\mathrm{E}[S_N^V] = 0$.

DEFINITION: Cardinality.
$f : A \to B$ is $1:1$ or injective: $x \neq y \Rightarrow f(x) \neq f(y) \forall x, y \in A$.
$f : A \to B$ is **onto** or surjective): $f(A) = B$.
f is **bijectiv** $\Leftrightarrow$ f is 1:1 and onto.
A, B are called **equivalent**, if there exists a bijection $f : A \to B$.
A equivalent to $\mathbb{N}$: **countable infinite.**
A equivalent to finite set: **finite.**
A neither finite nor countable infinite: **uncountable.**

DEFINITION:

A **σ-algebra** on Ω is a set $\mathcal{A}$ of subsets of Ω satisfying
- $\Omega \in \mathcal{A}$,
- $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$,
- $\{A_1, A_2, \ldots\} \subset \mathcal{A}$ countable $\Rightarrow \bigcup_{i=1}^\infty A_i \in \mathcal{A}$.

$P : \mathcal{A} \to \mathbb{R}$ is a **probability measure** if
- $\mathrm{P}[A] \geq 0$, (nonnegativity)
- $\mathrm{P}[\Omega] = 1$, (normalisation)
- $\{A_1, A_2, \ldots\}$ countable set of disjoint sets $\Rightarrow \mathrm{P}[\bigcup_{i=1}^\infty A_i] = \sum_{i=1}^\infty \mathrm{P}[A_i]$ (**σ-additivity**)

A **probability space** $(\Omega, \mathcal{A}, P)$ sonsists of a set Ω, a σ-algebra $\mathcal{A}$ on Ω and a probability measure P on $\mathcal{A}$.

If Ω is **finite**, the probability space is called a finite probability space. If Ω is countable, it is called **discrete**.

6. Week, (Summary of the theory)

DEFINITION: A function $d : \Omega \times \Omega \to \mathbf{R}$ is called a **metric** if

(i) $d(x,y) \geq 0$,
(ii) $d(x,z) \leq d(x,y) + d(y,z)$
(iii) $d(x,y) = 0 \Leftrightarrow x = y$

The pair (Ω, d) where Ω is a set and d is a metric is called a **metric space**. Examples. $(\mathbf{R}^n, d(x-y) = ||x-y||$, $(\{0,1\}^{\mathbf{N}}, d(x,y) = \sum_{i=1}^{\infty} |x_i - y_i| 2^{-i})$.

PROPOSITION:
Let $\{\mathcal{A}_i\}_{i \in I}$ be a collection of σ-algebras in Ω. Then $\bigcap_{i \in I} \mathcal{A}_i$ is a σ-algebra.
It follows that if $\mathcal{S}$ is a set of subsets of Ω, then there exists a **smallest σ-algebra**, which contains $\mathcal{S}$.

DEFINITION: Let (Ω, d) be a metric space and let $\mathcal{S}$ be the set of open balls $B_r(x) = \{y \in \Omega \mid d(x,y) < r\}$. The smallest σ-algebra which contains $\mathcal{S}$ is called the **Borel $\sigma-$ algebra** on Ω.

DEFINITION: Given a sequence of independent events in a probability space. Define $A_\infty := \limsup_{n\to\infty} A_n := \bigcap_{m=1}^{\infty} \bigcup_{n \geq m} A_n$. We have $A_\infty = \{\omega \mid \omega \text{ is in infinitely many } A_i\}$.

BOREL CANTELLI LEMMA: (Monkey typing Shakespeare)

a) If $\sum_n \mathrm{P}[A_n] < \infty$, then $\mathrm{P}[A_\infty] = 0$.
b) If $\sum_n \mathrm{P}[A_n] = \infty$, then $\mathrm{P}[A_\infty] = 1$.

DEFINITION: A random variable X on a probability space $(\Omega, \mathcal{A}, P)$ is called **discrete**, if $\Omega(X)$ is countable or finite. In this case, the expectation of X is defined as

$$\mathrm{E}[X] = \sum_{a \in X(\Omega)} a \cdot \mathrm{P}[X = a]$$

if the sum converges. We denote with $\mathcal{L}^1$ the set of random variables, for which $\mathrm{E}[|X|] < \infty$. Variance, Covariance etc. are defined as in the finite case (keep always an eye on convergence). Note that if $f(X) \in \mathcal{L}^1$, then

$$\mathrm{E}[f(X)] = \sum_{a \in X(\Omega)} f(a) \cdot \mathrm{P}[X = a] \,.$$

For example if $X^2 \in \mathcal{L}^1$, then

$$\mathrm{Var}[X] = \mathrm{E}[(X - \mathrm{E}[X])^2] = \sum_{a \in X(\Omega)} (a - m)^2 \mathrm{P}[X = a], \; m = \mathrm{E}[X] \,.$$

EXAMPLES:

$\mathrm{P}[X = k] = \frac{\lambda^k}{k!} e^{-\lambda}$	Poisson	$\mathrm{E}[X] = \lambda$	Electrons from cathode
$\mathrm{P}[X = k] = (1-p)^{k-1} p$	Geometric	$\mathrm{E}[X] = 1/p$	Waiting time for success
$\mathrm{P}[X = k] = \zeta(s)^{-1} k^{-s}$	Zeta	$\mathrm{E}[X] = \zeta(s+1)/\zeta(s)$	$\zeta(s) = \sum_{n=1}^{\infty} n^{-s}$

7. Week, (Summary of the theory)

DEFINITION: Integration, Expectation: Denote with $\mathcal{S}$ the set of random variables taking finitely many values: Define for $X \in \mathcal{S}$

$$\mathrm{E}[X] := \sum_{a \in X(\Omega)} a \cdot \mathrm{P}[X = a] .$$

Let $\mathcal{L}^1$ be the set of random variables X for which $\sup_{Y \in \mathcal{S}, Y \leq |X|} \mathrm{E}[Y] < \infty$. For $X \in \mathcal{L}^1$ and $X \geq 0$, the integral or expectation is defined as

$$\mathrm{E}[X] := \sup_{Y \in \mathcal{S}, Y \leq X} \mathrm{E}[Y] .$$

In general, we decompose X into $X = X^+ - X^-$ with $X^\pm \geq 0$ and put $\mathrm{E}[X] = \mathrm{E}[X^+] - \mathrm{E}[X^-]$. We write also $\int_\Omega X \, dP$ for $\mathrm{E}[X]$ since expectation is integration. Variance, Covariance etc. are defined as in the finite case: $\mathrm{Var}[X] = \mathrm{E}[(X - E[X])^2], \mathrm{Cov}[X,Y] = \mathrm{E}[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])]$.

DEFINITION: The Distribution function of a random variable X is $F(t) = \mathrm{P}[X \leq t]$. Absolutely continuous random variable: the probability density function $F' = f$ exists. Discrete random variable: F is piecewise constant with countably many jump discontinuities. The expectation, variance and $\mathrm{E}[g(X)]$ for $g(X) \in \mathcal{L}^1$ is in the continuous case

$$m = \mathrm{E}[X] = \int_{-\infty}^{\infty} x f(x) \, dx, \ \mathrm{Var}[X] = \int_{-\infty}^{\infty} (x - m)^2 f(x) \, dx \ , \mathrm{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) \, dx$$

For discrete random variables this is (repetition)

$$m = \mathrm{E}[X] = \sum_{a \in X(\Omega)} a\mathrm{P}[X = a], \ \mathrm{Var}[X] = \sum_{a \in X(\Omega)} (a - m)^2 \mathrm{P}[X = a], \mathrm{E}[g(X)] = \sum_{a \in X(\Omega)} g(a)\mathrm{P}[X = a]$$

Sometimes, one does not know the distribution of the random variable, then $\mathrm{E}[X], \mathrm{Var}[X]$ and $\mathrm{E}[g(X)]$ have to be computed by integrating (rsp. summing) over Ω.

EXAMPLES OF DISCRETE DISTRIBUTIONS:

Distribution	$\mathrm{P}[x = k] =$	Parameters	Domain	Mean	Variance
Binomial	$\binom{n}{k} p^k (1-p)^{n-k}$	$n \in \mathbb{N}, p \in [0,1]$	$\{0, \dots, n\}$	np	$np(1-p)$
Poisson	$\frac{\lambda^k}{k!} e^{-\lambda}$	$\lambda > 0$	$\{0, 1, \dots\}$	λ	λ
Geometric	$(1-p)^{k-1} p$	$p \in (0,1)$	$\{1, 2, \dots\}$	$1/p$	$1/p^2$

EXAMPLES OF ABSOLUTELY CONTINUOUS DISTRIBUTIONS:

Distribution	Density $f(x) =$	Parameters	Domain	Mean	Variance
Uniform	$1_{[a,b]} \cdot (b-a)^{-1}$	$a < b$	$[a,b]$	$(a+b)/2$	$(b-a)^2/12$
Exponential	$\lambda e^{-\lambda x}$	$\lambda > 0$	$\mathbb{R}^+$	$1/\lambda$	$1/\lambda^2$
Normal	$(2\pi\sigma^2)^{-1/2} \, e^{-\frac{(x-m)^2}{2\sigma^2}}$	$m \in \mathbb{R}, \sigma^2 > 0$	$\mathbb{R}$	m	σ^2
Erlang	$\frac{\lambda^k x^{k-1}}{(k-1)!} e^{-\lambda x}$	$\mathbb{R}^+$	$\lambda > 0, k \in \mathbb{N}$	k/λ	k/λ^2

8. Week, (Summary of the theory)

PROPERTIES OF DISTRIBUTION FUNCTIONS:

$$\begin{array}{ll} F(t) \in [0,1] & \mathrm{P}[a < X \leq b] = F(b) - F(a) \\ a \leq b \Rightarrow F(a) \leq F(b) & \lim_{t\to-\infty} F(t) = 0, \lim_{t\to\infty} F(t) = 1 \\ \lim_{\epsilon\searrow 0} F(a+\epsilon) = F(a) & \lim_{\epsilon\searrow 0} F(a-\epsilon) = F(a) - \mathrm{P}[X=a] \end{array}$$

Every function $F : \mathbf{R} \to \mathbf{R}$ satisfying the above properties belongs to a random variable X: define the probability space $(\Omega, \mathcal{A}, P)$, where $\mathcal{A}$ is the Borel σ-algebra on $\Omega = \mathbf{R}$ and P is defined by $\mathrm{P}[[a,b)] = F(b) - F(a) = \mathrm{P}[X \in [a,b)]$.

DEFINITION: $X = (X_1, X_2, \dots, X_d)$ is called a **random vector** if X_i are random variables. The **distribution function** of X (also called **joint distribution** of $X_1, X_2, \dots, X_d$) is defined as

$$F(t_1, \dots, t_d) = \mathrm{P}[X_1 \leq t_1, X_2 \leq t_2, \dots, X_d \leq t_d] .$$

The distribution is **continuous** if there exists a function $f : \mathbf{R}^d \to \mathbf{R}^+$ such that

$$F(t) = \int_{-\infty}^{t_1} \int_{-\infty}^{t_2} \cdots \int_{-\infty}^{t_d} f(y_1, y_2, \dots, y_d)\, dy_d \dots dy_2 dy_1 .$$

TRANSFORMATIION OF RANDOM VARIABLES:

- Let F be a continuous invertible distribution function. Let X be a random variable which is uniformly distributed in $[0,1]$. Then $Y = F^{-1}(X)$ gives random numbers with distribution F.
- Given a continuous random variable X with density f and a differentiable invertible function $\phi : \mathbf{R} \to \mathbf{R}$ with inverse ψ. The random variable $Y = \phi(X)$ has the density

$$g(t) = f(\psi(t))|\psi'(t)| .$$

- Given a continuous random vector X with density f and a differentiable invertible function $\phi : \mathbf{R}^d \to \mathbf{R}^d$ with inverse ψ. The random variable $Y = \phi(X)$ has the density

$$g(u) = f(\psi(u))|\mathrm{Det}(D\psi(u))| .$$

DEFINITION: The **characteristic function** of X is defined as $\phi_X(t) = \mathrm{E}[e^{itX}]$.
Discrete case: $\phi_X(t) = \sum_{a\in X(\Omega)} e^{ita}\mathrm{P}[X = a]$. Continuous case: $\phi_X(t) = \int_{-\infty}^{\infty} e^{itx} f(x)\, dx$.
CALCULATION OF MOMENTS: $\mathrm{E}[X^k] = (-i)^k \phi_X^{(k)}(0)$. Especially, $\mathrm{E}[X] = -i\phi'_X(0)$.

SUMS OF INDEPNDENT RANDOM VARIABLES: X_i independent with distribution $\phi_i, S = \sum_{i=1}^n X_i$, then $\phi_S(t) = \phi_1(t) \cdot \phi_2(t) \cdots \phi_n(t)$.

THE GAMMA FUNCTION. Some distributions use the Gamma function:

$$\Gamma(n) = \int_0^\infty z^{n-1} e^{-z}\, dz .$$

For $n \in \mathbf{N}$, we have $(n-1)!$. Proof. $\Gamma(1) = 1$, $\Gamma(n) = (n-1)\Gamma(n-1)$ by partial integration. Computations like $\Gamma(1/2) = \sqrt{\pi}, \Gamma(3/2) = \sqrt{\pi}/2$ use $\int_{\mathbf{R}} e^{-x^2/2} = \sqrt{2\pi}$.

Distribution	Parameter	Charact. function
Normal	$m \in \mathbf{R}, \sigma^2 > 0$	$e^{mit - \sigma^2 t^2/2}$
Standard normal		$e^{-t^2/2}$
Uniform	$[-a, a]$	$\sin(at)/(at)$
Exponential	$\lambda > 0$	$\lambda/(\lambda - it)$
Binomial	$n \in \mathbf{N}, p \in [0,1]$	$(p + (1-p)e^{it})^n$
Poisson	$\lambda > 0$	$e^{\lambda(e^{it}-1)}$
Geometric	$p \in (0,1)$	$\frac{pe^{it}}{(1-(1-p)e^{it})}$

9. Week, (Summary of the theory)

CHEBYCHEV-MARKOV INEQUALITY.
Let $h : \mathbf{R}^+ \to \mathbf{R}^+$ be a monoton function and $X \geq 0$ a random variable with $h(X) \in \mathcal{L}^1$. Then for all $c > 0$

$$h(c) \cdot \mathrm{P}[X \geq c] \leq \mathrm{E}[h(X)] .$$

Proof. Take the expectation of $h(c)1_{X \geq c}(\omega) \leq h(X)(\omega)$. Use the monotonicity and linearity of the expectation.

CHEBYCHEV INEQUALITY.
If $X \in \mathcal{L}^2$, then for all $c > 0$

$$\mathrm{P}[|X - \mathrm{E}[X]| \geq c] \leq \frac{\mathrm{Var}[X]}{c^2} .$$

Proof. Apply Chebychev-Markov's inequality to $Y = |X - \mathrm{E}[X]|$ and $h(x) = x^2$.

DEFINITION.
A sequence of random variables X_n **converges in probability** to a random variable X, if for all $\epsilon > 0$,

$$\lim_{n \to \infty} \mathrm{P}[|X_n - X| \geq \epsilon] = 0 .$$

WEAK LAW OF LARGE NUMBERS.
Assume X_i have common expectation $E[X_i] = m$ and satisfy $\sup_n \frac{1}{n} \sum_{i=1}^n \mathrm{Var}[X_i] < \infty$. If X_n are pairwise uncorrelated, then $\sum_{i=1}^n X_i/n \to m$ in probability.
Proof. Since in generl $\mathrm{Var}[X + Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2 \cdot \mathrm{Cov}[X, Y]$ and X_n are pairwise uncorrelated, we have $\mathrm{Var}[X_n + X_m] = \mathrm{Var}[X_n] + \mathrm{Var}[X_m]$ for $n \neq m$ and by induction $\mathrm{Var}[S_n] = \sum_{i=1}^n \mathrm{Var}[X_i]$. Using linearity, we obtain $E[S_n/n] = m$ and

$$\mathrm{Var}[S_n/n] = E[(S_n)^2/n^2] - E[S_n]^2/n^2 = \mathrm{Var}[S_n]/n^2 = \frac{1}{n^2} \sum_{i=1}^n \mathrm{Var}[X_n] \to 0 .$$

With Chebychev's inequality, we obtain

$$P[|S_n/n - m| \geq \epsilon] \leq \frac{\mathrm{Var}[S_n/n]}{\epsilon^2} \to 0 \ (n \to \infty) .$$

IMPORTANT SPECIAL CASE.
If X_i are independent random variables with the same distribution for which the mean m and variance exist both, then $\sum_{i=1}^n X_i/n \to m$ in probability.

EXISTENCE OF INDEPENDENT RANDOM VARIABLES: (don't read this!)
Given a distribution function F, there exists a probability space $(\Omega, \mathcal{A}, P)$ and independent random variables $X_1, X_2, \ldots$ which have all the distribution F.
Proof. We know how to construct a single random variable X with distribution F on a probability space $(\mathbf{R}, \mathcal{B}, Q)$. Form the product space

$$(\Omega, \mathcal{A}, P) = (\mathbf{R}^{\mathbf{N}}, \mathcal{B}^{\mathbf{N}}, Q^{\mathbf{N}}) .$$

Ω contains sequences $\omega = (\omega_1, \omega_2, \ldots)$ and the probability measure P is defined by

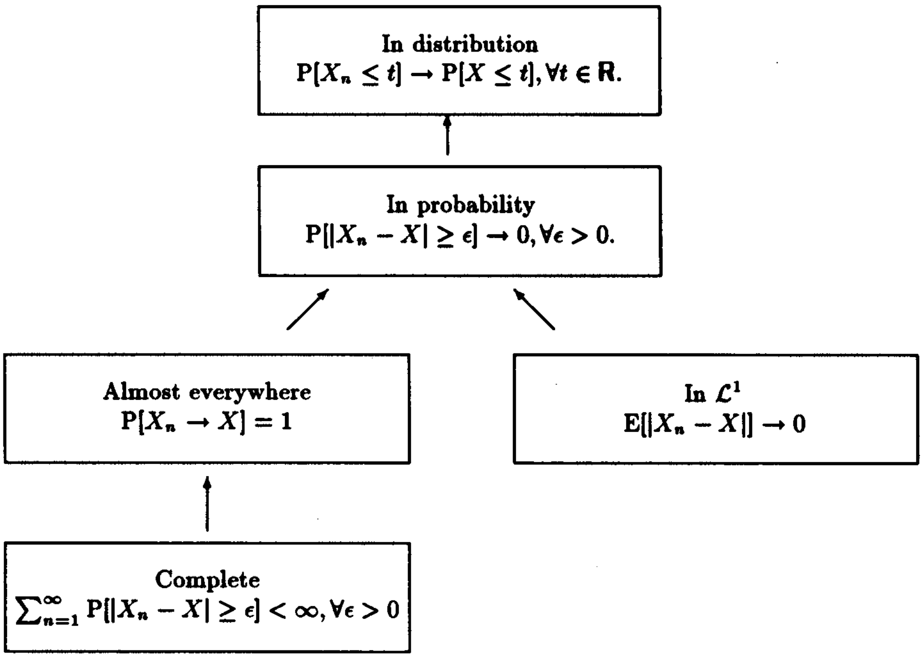
$$P[A_1 \times A_2 \times A_3 \cdots \times A_n \times \mathbf{R} \times \mathbf{R} \times \ldots] = Q[A_1] \cdot Q[A_2] \ldots Q[A_n] .$$

The σ-algebra $\mathcal{A}$ is the smallest σ-algebra containing all sets of the form $A_1 \times A_2 \times A_3 \cdots \times A_n \times \mathbf{R} \times \mathbf{R} \times \ldots$ with $A_i \in \mathcal{B}$. The random variables $X_i(\omega) = \omega_i$ are independent and have all the distribution F.

10. Week, (Summary of the theory)

DEFINITION. A sequence of random variables X_n converges **almost everywhere** to a random variable X, if $\mathrm{P}[X_n \to X, n \to \infty] = 1$.

RELATION BETWEEN CONVERGENCE OF RANDOM VARIABLES: an arrow stands for "implies"



STRONGER WEAK LAW OF LARGE NUMBERS:
Assume X_i have common expectation $E[X_i] = m$ and satisfy $M = \sup_n \mathrm{E}[X^4] < \infty, \sup_n \mathrm{E}[X^2]^2 < \infty$. If X_i are independent, then $\sum_n \mathrm{P}[|S_n/n - m| \geq \epsilon]$ converges for all $\epsilon > 0$.
Proof. Estimation of $\mathrm{E}[X_n^4]$ with Chebychev-Markov's inequality gives $\mathrm{P}[|S_n/n - m| \geq \epsilon] \leq C/n^2$ for some constant C.

STRONG LAW OF LARGE NUMBERS:
Assume X_n are independent random variables with $M = \sup_n \mathrm{E}[X_n^4] < \infty$, $\sup_n \mathrm{E}[X_n^2] < \infty$ with common expectation $\mathrm{E}[X_n] = m$. Then $S_n/n \to m$ almost everywhere.
Proof. Direct consequence of the stronger weak law above since complete convergence implies convergence almost everywhere.

DEFINITION. A sequence of random variables X_n converges **in distribution** to a random variable X, if for all $t \in \mathbf{R}$, $\mathrm{P}[X_n \leq t] \to \mathrm{P}[X \leq t]$ for $n \to \infty$.

CENTRAL LIMIT THEOREM:
Given X_n which are independent with mean m and variance σ^2. Let X be a random variable with standard normal distribution. Then

$$\frac{S_n - nm}{\sigma\sqrt{n}} \to X$$

in distribution, where $S_n = X_1 + X_2 + \ldots + X_n$.
Proof. A calculation shows that the characteristic functions of $S_n^* = (S_n - E[S_n])/\sigma[S_n]$ converge to the characteristic function of X.

HOMEWORK

Ma2c, O. Knill, **April 1995**

Week 1

Due: Monday, April 10, 1995

Topics: Theorem of Stokes, Theorem of Gauss

Reading: Apostol: 12.1, 12.2, 12.3, 12.5, 12.7, 12.9, 12.11, 12.12, 12.18, 12.19, 12.20

1) (*) Topic: **Surfaces.** Given the surface $S = r(B)$, where $B = [0, 2\pi] \times [a, b]$ and

$$r : (u, v) \mapsto (v\cos(u), v\sin(u), f(v)) ,$$

where $f : [a, b] \to \mathbf{R}$ is a differentiable function.

a) Compute the two tangent vectors r_u and r_v at each point of S.
b) Determine the normal vector $N = r_u \wedge r_v$ at each point of S.

2) (*) Topic: **Surface integrals.** Let Q be the surface

$$Q = \{(x, y, z) \in \mathbf{R}^3 \mid -1 \leq x \leq 1, -1 \leq y \leq 1, z = 0\}$$

given by $r(u, v) = (u, v, 0)$ and let H be the semi-sphere

$$H = \{(x, y, z) \in \mathbf{R}^3 \mid x^2 + y^2 + z^2 = 1, z \geq 0\}$$

given by $r(u, v) = (u, v, +\sqrt{1 - (u^2 + v^2)})$. Compute the flux $\int_S F \, dS$ of the vector field F through the surface S in the following cases:

a) $F(x, y, z) = (0, 0, 1), S = Q$, b) $F(x, y, z) = (1, 1, 0), S = Q$,
c) $F(x, y, z) = (1, 1, 1), S = Q$, d) $F(x, y, z) = (x, y, z), S = H$.

3) (*) Topic: **Theorem of Stokes.** Let H be the surface

$$H = \{(x, y, z) \in \mathbf{R}^3 \mid x^2 + y^2 + z^2 = 1 \ z \geq 0\}$$

and D the disc

$$D = \{x^2 + y^2 \leq 1, z = 0\} .$$

(In both cases the orientation is such that the normal vector n has a positive z component).

Consider the vector field $F(x, y, z) = (-y, x, 0)$.

a) Compute the line integral of F around the circle δD (counterclockwise oriented in the xy plane).
b) Compute curl$F$.
c) Compute the flux $\int_H \text{curl}(F)\, dS$ through the surface H.
d) Compute the flux $\int_D \text{curl}(F)\, dS$ through the surface D.
e) Why are the two integrals the same?

4) Topic: **Theorem of Stokes.** Let γ be the closed path in $\mathbf{R}^3$ consisting of piecewise straight lines connecting the points $P_1 = (0, 0, \pi/2) \to P_2 = (0, \pi/2, \pi/2) \to P_3 = (0, \pi/2, 0) \to P_4 = (\pi/2, \pi/2, 0) \to P_5 = (\pi/2, 0, 0) \to P_6 = (\pi/2, 0, \pi/2) \to P_1$. Let F be the vector field

$$F(x, y, z) = (\sin y, \sin z, \sin x) .$$

Compute the integral

$$\int_\gamma F \, ds$$

a) directly.
b) with the Theorem of Stokes.

5) Topic: **Theorem of Stokes.** Let F a vector field in $\mathbf{R}^3$ and S a surface such that F is orthogonal to S at every point of S. Prove that F is orthogonal to curl(F) at every point of S. Hint: Assume the claim is wrong at one point in S and apply Stokes for a small disc around this point.

6) (**) Topic: **Theorem of Gauss.** Let $a, b, c \in \mathbf{R}$ be positive numbers. Consider the surface $r(B)$, where

$$r: \ (u,v) \mapsto r(u,v) = (a \cos u \cos v, b \sin u \cos v, c \sin v)$$

and $B = \{(u,v) \mid -\pi < u < \pi, -\frac{\pi}{2} < v < \frac{\pi}{2}\}$.

a) Describe S geometrically. (Hint: Compute $x^2/a^2 + y^2/b^2 + z^2/c^2$).
b) Compute r_u and r_v.
c) Compute $dS = N\ dudv$.
d) Let $F(x,y,z) = (0,0,z)$. Compute the flux of F through S and give an interpretation of the result with the theorem of Gauss.

7) Topic: **Theorem of Gauss.** Calculate the flux of the vector field

$$F(x,y,z) = (x + x^2yz, y + xy^2z, z - 2xyz^2)$$

through the semi-sphere

$$S = \{(x,y,z) \mid x^2 + y^2 + z^2 = 4, \ z \geq 0\}.$$

(The orientation of S is such that the normal vectors in S have the third component ≥ 0).

8) (*) Topic: **Theorem of Gauss.** Let S be a closed surface in $\mathbf{R}^3$ enclosing a domain G. Assume G contains the origin $O = (0,0,0)$ in its interior. Consider the Coulomb field

$$F(x,y,z) = \frac{1}{4\pi r^2}\frac{(x,y,z)}{r}, r = \sqrt{x^2 + y^2 + z^2}$$

which is defined for $\mathbf{R}^3 \setminus \{O\}$. What is the flux of F through the boundary $S = \delta G$ of G?

9) Topic: **Theorem of Gauss.** Given the function

$$f(x,y,z) = a(x^2 + y^2 + z^2) + b(xy + yz + zx) + c(x + y + z)\ ,$$

where $a, b, c \in \mathbf{R}$ are constants. Compute the flux of $\operatorname{grad} f$ through the sphere $S = \{x^2 + y^2 + z^2 = 1\}$.

Ma2c, O. Knill, **April 1995**

Week 2

Due: Monday, April 17, 1995

Topics: Boolean algebras, Finite probability spaces

Reading: Apostol: 13.1, 13.2, 13.3, 13.5, 13.6

1) Topic: **Boolean algebras.** Let $\mathcal{A}$ be a Boolean algebra over a finite set Ω. Prove that for all $A, B \in \mathcal{A}$
a) $A \cap B^c$ and B are disjoint.
b) $A \cup B = (A \cap B^c) \cup B$.
c) $A \cap B$ and $A \cap B^c$ are disjoint.

2) (*) Topic: **Boolean algebras.** Let $\mathcal{A}$ be a Boolean algebra over a finite set Ω. Define $A \Delta B = (A \cup B) \setminus (A \cap B)$. Show that $(\mathcal{A}, \Delta)$ is a commutative group (this means the following properties hold:)
a) For all A, B, C the **associativity law** $(A\Delta B)\Delta C = A\Delta(B\Delta C)$ holds.
b) There exists a **zero element** $A_{zero} \in \mathcal{A}$ such that $A \Delta A_{zero} = A$ for all $A \in \mathcal{A}$.
c) For all A, there exists an **inverse** B such that $A\Delta B = A_{zero}$.
d) For all A, B the **commutativity law** holds $A\Delta B = B\Delta A$.

3) Topic: **Boolean algebras.** Let $\mathcal{A}$ be a Boolean algebra over a finite nonempty set Ω. Show that $(\mathcal{A}, \cap)$ is a commutative semi-group: (this means the following properties hold:)
a) For all A, B, C the **associativity law** holds $A \cap (B \cap C) = (A \cap B) \cap C$.
b) There exists a **one element** A_{one} such that $A \cap A_{one} = A$ for all $A \in \mathcal{A}$.
c) Show that $(\mathcal{A}, \cap)$ is not a group (except if Ω has only 1 element).
d) The **commutativity law holds:** $A \cap B = B \cap A$.

4) Topic: **Boolean algebras.**
a) Show that $(\mathcal{A}, \Delta, \cap)$ is a ring: (this means that $(\mathcal{A}, \Delta)$ is a group and $(\mathcal{A}, \cap)$ is a semi-group and that the **distributivity law**

$$A \cap (B\Delta C) = (A \cap B)\Delta(A \cap C)$$

holds.
b) (for fun only) Show that $(\mathbb{Z},+,\cdot)$ is a ring too. Explore the similarities and differences between this ring and the ring $(\mathcal{A},\Delta,\cap)$.

5) (*) Topic: **Probability spaces.** $\Omega = \{1,2,3\}$.
a) Find all Boolean algebras on Ω. b) Find all probability spaces on Ω.

6) (*) Topic: **Probability spaces.** What is the probability that a 2×2 matrix having random entries 0 or 1 (both with equal probability 1/2) is invertible? (Construct the probability space $(\Omega, \mathcal{A}, P)$ and the event that a matrix is invertible).

7) Topic: **Probability spaces.** Consider a part of the internet computer network consisting of 3 nodes (computers). Assume that two of the three computers are conneced with probability p and that with probability $q = 1-p$, a connection is broken.
a) Construct the probability space $(\Omega, \mathcal{A}, P)$.
b) What is the event that all nodes have connection to all other nodes.
c) What is the probability that one can reach from any of the computers any other computer.

8) (*) Topic: **Probability spaces. The birthday paradox**: assume 20 people are in a room. What is the probability that at least two people have the same birthday? (Hint: Compute the event that no people have the same birthday. The year is assumed to have 365 days.)

9) (**) Topic: **Probability spaces. The wardrobe problem**: we distribute randomly $n \in \mathbb{N}$ coats of n people.
a) Set up the probability space $(\Omega, \mathcal{A}, P)$ for the situation.
b) Define for $1 \leq i \leq n$ the event A_i that at least person number i gets his (or her) coat. What is the probability of the event $A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}$, where $1 \leq i_1 < i_2 < \ldots < i_k \leq n$.
b) What is the probability $p(n)$ that no person gets his (or her) own coat? Hint: Consider the event that at least one person gets his (or her) coat and use the formula from class

$$P[\bigcup_{i=1}^{n} A_i] = \sum_{k=1}^{n}(-1)^{k-1} \sum_{1 \leq i_1 < i_2 < \ldots < i_k \leq n} P[A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}] .$$

c) What happens with the probability $p(n)$ for $n \to \infty$?

Ma2c, O. Knill, **April 1995**

Week 3

Due: Monday, April 24, 1995

Topics: constructions of new probability spaces: (i) change of algebra, (ii) product space (iii) conditional probability space, independence of events

Reading: Apostol: 13.10, 13.12, 13.13, 13.15

1) Topic: **Boolean algebras.** Consider the probability space $(\Omega, \mathcal{A}, P)$, where Ω consists of all N-tupels $\omega = (\omega_1, \dots, \omega_N)$ with $\omega_i \in \{0,1\}$, where $\mathcal{A}$ is the set of subsets of Ω and P is the Laplace probability measure $P[A] = |A|/|\Omega|$, where $|A|$ denotes the number of elements in A. Let $\mathcal{A}_n$ be the set of subsets A of Ω which can be written as a union of sets $\{\omega \mid \omega_1 = x_i, \dots, \omega_n = x_n\}$ with $x_i \in \{0,1\}$. Show that for every n, the set $\mathcal{A}_n$ is a Boolean algebra. How many elements does $\mathcal{A}_n$ have?

2) (*) Topic: **Conditional probability.** Given a finite probability space $(\Omega, \mathcal{A}, P)$ and events $A_i \in \mathcal{A}$. Prove the formula:

$$P[A_1 \cap \dots \cap A_n] = P[A_1] \cdot P[A_2 \mid A_1] \cdot P[A_3 \mid A_1 \cap A_2] \cdots P[A_n \mid A_1 \cap \dots \cap A_{n-1}] .$$

3) Topic: **Conditional probability.** A random number generator of a computer gives randomly an integer in the set $\{1, 2, 3, \dots, 49, 50\}$.
a) What is the probability that the random number generator gives a prime number?
b) The random number generator is modified in the following way. Instead of giving a number out, it checks, whether the obtained number is divisable by 2 or 3. If this is the case, it produces an other random number, etc, until it finds a number which is neither divisible by 2 nor by 3. The first so obtained number is then given out. What is the probability that this random number generator with filter gives a prime number?

4) Topic: **Conditional probability**. Let $(\Omega, \mathcal{A}, P)$ be a finite probability space. Given $B \subset \Omega$ with $P[B] > 0$, we have proven in class that $(\Omega, \mathcal{B}, Q)$ with $\mathcal{B} = \mathcal{A} \cap B$ and $Q[A] = P[A \mid B]$ is a probability space. Give an example which shows that it is in general not possible to reconstruct the probability space $(\Omega, \mathcal{A}, P)$ from $(\Omega, \mathcal{B}, Q)$.

5) (*) Topic: **Conditional probability**. You hear somebody say: "I have three children, and one of them is a girl". What is the probability that the other two children are both boys?

6) (*) Topic: **Independence**. Take a standard stack of 52 cards. (It contains as usual 4 queens and 4 kings).
a) We take randomly one card. Is the event that this card is a queen independent from the event that this card is a king?
b) We take randomly a card, put it back and take randomly a second card. Is the event that the first card is a queen independent from the event that the second card is a king?
c) We take randomly two cards, one after the other, this time without returning the first card. Is the event that the first card is a queen independent from the event that the second card is a king?
Justify in a),b) and c) your answers by computing the relevant probabilities. (Hint: Experiments with a small stack of cards can be helpful).

7) Topic: **Change of the Boolean algebra, independence**. Given n independent events $A_1, A_2, \dots A_n$ in a finite probability space $(\Omega, \mathcal{A}, P)$. Let $\mathcal{B} \subset \mathcal{A}$ be a Boolean algebra. If $A_i \in \mathcal{B}$, then A_i are also independent in the probability space $(\Omega, \mathcal{B}, P)$.

8) (*) Topic: **Product probability space, indepencence**. Given two finite probability spaces $(\Omega_i, \mathcal{A}_i, P_i)$ for $i = 1, 2$. Assume that A_1, B_1 are independent in $(\Omega_1, \mathcal{A}_1, P_1)$ and A_2, B_2 are independent in $(\Omega_2, \mathcal{A}_2, P_2)$. Show that the four sets $A_1 \times \Omega_2, B_1 \times \Omega_2, \Omega_1 \times A_2, \Omega_1 \times B_2$ are independent in the product space $(\Omega = \Omega_1 \times \Omega_2, \mathcal{A}, P) = (\Omega_1, \mathcal{A}_1, P_1) \times (\Omega_2, \mathcal{A}_2, P_2)$.

9) (**) Topic: **Conditional probability, independence**. Given a finite probability space $(\Omega, \mathcal{A}, P)$. Assume that A, B, C are independent events. Show that the events $A \cap C, B \cap C$ are independent in the conditional probability space $(C, \mathcal{A} \cap C, P[\cdot \mid C\,])$.

Ma2c, O. Knill, **April 1995**

Week 4

Due: Monday, May 1, 1995

Topics: Random variables, Expectation, Variance and Covariance, Correlation, Independent random variables.

Reminder of some definitions:

A **random variable** on a finite probability space $(\Omega, \mathcal{A}, P)$ is a map $X : \Omega \to \mathbf{R}$ such that for all $a \in \mathbf{R}$, we have $\{X = a\} \in \mathcal{A}$.

The **expectation** of a random variable X is defined as

$$\mathrm{E}[X] = \sum_{a \in X(\Omega)} a \cdot P[X = a] .$$

The **variance** of a random variable X is defined as

$$\mathrm{Var}[X] = \mathrm{E}[\,(X - \mathrm{E}[X])^2\,] .$$

We know from class that $\mathrm{Var}[X] = \mathrm{E}[X^2] - \mathrm{E}[X]^2$.
The **standard deviation** of X is $\sigma[X] = \sqrt{\mathrm{Var}[X]}$.

The **covariance** of two random variables X, Y is defined as

$$\mathrm{Cov}[X, Y] = \mathrm{E}[(X - \mathrm{E}[X]) \cdot (Y - \mathrm{E}[Y])] .$$

We know from class that $\mathrm{Cov}[X, Y] = \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y]$.

Given two random variables X, Y with $\mathrm{Var}[X] > 0, \mathrm{Var}[Y] > 0$, the **correlation** between X and Y is defined as

$$\mathrm{Corr}[X, Y] = \frac{\mathrm{Cov}[X, Y]}{\sigma[X]\sigma[Y]} .$$

If $\mathrm{Corr}[X, Y] = 0$, then X, Y are called **uncorrelated**.

Two random variables X, Y are **independent** if for all $a, b \in \mathbf{R}$

$$P[X = a; Y = b] = P[X = a] \cdot P[Y = b] \,.$$

1) (*) Topic: **Expectation.** Two computers **piccolo**[1] and **violin**[2] of a local well known computer net are connected with exactly n connections all going from piccolo to violin. Each of the connections goes broken with probability p, $0 \leq p \leq 1$, independent of the other connections. What is the expected number of broken connections?

2) (*) Topic: **Probability, Expectation.** Consider the random walker (drunken sailor) in the plane, who makes randomly n steps (of length 1) in the four different directions u, d, l, r, with probability $p_r = p_l = p_u = p_d = 1/4$. He starts at $(0,0)$
a) What is the probability that he is after 6 steps again in $(0,0)$?
b) What is the probability that he is after 23 steps again at the same place. (Hint: do not immediately start computing)?
c) Consider the random variable $X(\omega)$ which gives the euclidean distance of the walker from the origin after n steps. Compute $\mathrm{E}[X]$ in the cases, $n = 0, 1, 2$.
d) Compute $\mathrm{Var}[X]$ in the cases $n = 0, 1, 2$.

3) Topic: **Random variables.** Give an example of a probability space and a function $X : \Omega \to \mathbf{R}$ which is not a random variable.

4) (**) Topic: **Expectation, Variance, Covariance, Correlation.** Given the probability space $(\Omega = \{1,2,3,4,5,6\}, \mathcal{A} = \{A \subset \mathcal{A}\}, P[A] = |A|/|\Omega|)$ ("throwing a dice once"). We use the notation $1_A(\omega) = \begin{cases} 1, \omega \in A \\ 0, \omega \notin A \end{cases}$ for the **characteristic function** of an event A. Define the random variables

$$X(\omega) = \omega^3, Y(\omega) = 1_{\{\omega \leq 3\}}, Z(\omega) = 1_{\{1\}} \,.$$

Compute

a) $\mathrm{E}[X], \mathrm{Var}[X], \sigma[X]$, *b)* $\mathrm{E}[Y], \mathrm{Var}[Y], \sigma[Y]$.
c) $\mathrm{Cov}[X,Y], \mathrm{Cov}[X,Z], \mathrm{Cov}[Y,Z]$ *d)* $\mathrm{Corr}[X,Y], \mathrm{Corr}[X,Z], \mathrm{Corr}[Y,Z]$.

---

[1] **Any possible coincidence with living computers would be a pure accident.**
[2] **Name changed.**

5) Topic: **Covariance, Independent random variables.** Given the probability space ($\Omega = \{1,2,3,4,5,6\}^2, \mathcal{A} = \{A \subset \Omega, P[A] = |A|/|\Omega|$) ("throwing a dice twice"). Define the random variables $X(\omega) = \omega_1$, $Y(\omega) = \omega_2$. Compute $\mathrm{Cov}[X,Y]$. Hint. Use a result in theory instead of computing the value directly.

6) (*) Topic: **Expectation, Variance.** Let $(\Omega, \mathcal{A}, P)$ be the probability space, where Ω is set of 2×2 matrices with entries $0, 1$ taken randomly with probability $P[\omega_{11} = 1] = P[\omega_{12} = 1] = P[\omega_{21} = 1] = 1/3, P[\omega_{22} = 1] = 0$, (compaire alsoProblem 6) in Week 2 where we had an other measure). Consider the random variables $X(\omega) = \det(\omega), Y(\omega) = \mathrm{tr}(\omega)$, where det is the determinant and tr is the trace (the sum of the diagonal elements). Compute $\mathrm{E}[X], \mathrm{E}[Y], \mathrm{Var}[X], \mathrm{Var}[Y], \mathrm{Cov}[X,Y], \mathrm{Corr}[X,Y]$.

7) (*) Topic: **Correlation, Independent random variables.**
Given two random variables X, Y on a finite probability space $(\Omega, \mathcal{A}, P)$ which have both positive variance.
a) For each $\theta \in [0, 2\pi]$, define

$$X_\theta = X\cos(\theta) - Y\sin(\theta) ,$$
$$Y_\theta = X\sin(\theta) + Y\cos(\theta) .$$

Show that there exists a value of θ for which X_θ, Y_θ are uncorrelated.
b) If X, Y are uncorrelated then

$$\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] .$$

c) Why is the formula in b) also called Pythagoras law?
d) Compute $\mathrm{Var}[X+Y]$ if X, Y are independent both with standard deviation 1.

8) Topic: **Expectation, Variance.** Given two random variables X, Y on the probability space $(\Omega = \{0,1\}, \mathcal{A} = \{A \subset \Omega\}, P[\{0\}] = P[\{1\}] = 1/2)$. Determine all random variables X and Y which have zero expectation $\mathrm{E}[X] = \mathrm{E}[Y] = 0$ and the same variance $\mathrm{Var}[X] = \mathrm{Var}[Y] = 1$.

9) Topic: **Expectation.** For $n \in \mathbb{N}$, let Ω_n be the number of alcohols $C_nH_{2n+1}OH$. Two molecules which can be deformed into each other are considered as identical. For $n = 3$ for example, there are two different such molecules:

```
    H   H   H               H   H    H
    |   |   |               |   |    |
H - C - C - C - OH      H - C - C  - C - H
    |   |   |               |   |    |
    H   H   H               H   OH   H
```

Every OH group is connected to one carbon atom C which has either one or two carbon neighbors (and this property has some influence on chemical properties of the molecule). Assume, that each of these molecules is produced with the same probability. We consider now the case $n = 4$ and so the alcohol C_4H_9OH. How many carbon neighbors do we expect for the carbon atom attached to OH?

This is the last homework assignment before the midterm examination. The topics of the midterm assignment are identical to the material of the first four weeks, which is:

1. Week **: 1. Surface integrals, 2. Theorem of Stokes, 3. Theorem of Gauss.**

2. Week **: 1. Boolean algebras, 2. Probability spaces, 3. Three constructions.**

3. Week **: 4. Independent events, 5. Random variables, 6. Expectation.**

4. Week **: 7. Variance, Covariance, Correlation. 8. Independent random variables.**

The problems will be in a similar style as the Week1-Week4 homework problems.

Ma2c, O. Knill, **May 1995**

Week 6

Due: Monday, May 15, 1995

Topics: Cardinality of sets, general probability spaces, random variables on discrete probability spaces

Reading: Apostol: 13.19, 13.21, 14.1, 14.2

DEFINITION: A map $f : A \to B$ is called **1:1** (or injective), if $x \neq y \Rightarrow f(x) \neq f(y)$ for all $x, y \in A$. The map f is called **onto** (or surjective), if $f(A) = B$. The map f is called **bijectiv**, if it is 1:1 and onto.
Two sets A, B are called **equivalent**, if there exists a bijection $f : A \to B$. A set which is equivalent to $\mathbb{N}$, the set of natural numbers, is called **countable infinite**. A set which is not finite and not countable is called **uncountable**. Every set is either **finite, countable infinite or uncountable** and these attributes determines the **cardinality** of a set. A set which is either finite or countable infinite is called **countable**. To avoid confusion, we do not use the word "countable" in the restricted sense "countable infinite".

DEFINITION:

A **σ-algebra** on Ω is a set $\mathcal{A}$ of subsets of Ω satisfying

$\Omega \in \mathcal{A}$,
$A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$,
$\{A_1, A_2, \dots\} \subset \mathcal{A}$ countable $\Rightarrow$ $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$.

DEFINITION:

$P : \mathcal{A} \to \mathbb{R}$ is a **probability measure** if

$\mathrm{P}[A] \geq 0$, (nonnegativity)
$\mathrm{P}[\Omega] = 1$, (normalisation)
$\{A_1, A_2, \dots\}$ countable set of disjoint sets $\Rightarrow$ $\mathrm{P}[\bigcup_{i=1}^{\infty} A_i] = \sum_{i=1}^{\infty} \mathrm{P}[A_i]$ (**σ-additivity**)

DEFINITION: Probability space: $(\Omega, \mathcal{A}, P)$ is called a **probability space** if $\mathcal{A}$ is a σ-algebra of sets in Ω and $P : \mathcal{A} \to \mathbb{R}$ is a probability measure. If Ω is countable infinite or finite, the probability space is called **discrete**.

1) (*) Topic: **Cardinality of sets.**
$\mathbb{N}$ denotes the set of natural numbers $\{1,2,3,\ldots\}$. Show that the following sets are countable infinite:
a) $\mathbb{N}^2 = \{1,4,9,16,\ldots\}$, the set of squares.
b) $P = \{2,3,5,7,\ldots,\}$, the set of primes.
c) $\mathbb{Z} = \{\ldots,-4,-3,-2,-1,0,1,2,3,4,\ldots\}$, the set of integers.
d) $\mathbb{N} \times \mathbb{N}$.
e) $A = \{2^n 3^m,\ n,m \in \mathbb{N}\}$.
f) Assume, the set of **prime twins** $T = \{3,5,11,17,29,41,\ldots\} = \{p \in \mathbb{N} \mid p, p+2 \text{ are both prime }\}$ is infinite. Show that it is then countable infinite. For brave people: decide, whether the assumption T is true. [1] (The following mathematica 'one line program' shows a graph giving for each n the n'th prime-twin $p_n \leq 10000$.

```
ListPlot[Union[Table[If[PrimeQ[p] && PrimeQ[p+2],p,3],{p,3,10000}]]]
```

The output of this line is in the appendix of this exercice. Experiment yourself!)

2) (*) Topic: **Cardinality of sets.**
In class, we have seen, how Cantor's diagonal argument showed that the set $[0,1]$ is uncountable. This was done by writing every point of $[0,1]$ using a decimal expansion and showing that the set of these sequences is not countable. We will repeat this here in more detail but using the binary expansion. Since some binary sequences like $0.01111111\ldots$ and $0.100000\ldots$ represent the same number, we define

$$S = \{(a_1, a_2, \ldots,),\ a_i \in \{0,1\},\ \begin{array}{l}\text{either } a_i = 0, \text{ for infinitely many } i, \\ \text{or } a_i = 1,\ \forall\, i\}\end{array} .$$

a) Verify that $f: S \to [0,1]$ given by $f(a) = \sum_{i=1}^{\infty} a_i 2^{-i}$ is a bijection.
b) Show that S and $S \times S$ are equivalent by constructing a bijection between these two sets.

[1] This is a famous unsolved problem in mathematics and many prizes are waiting for the proof (:-)] or disproof (:-(] of this conjecture.

c) Show that $[0,1]$ and $[0,1] \times [0,1]$ are equivalent. Hint: use a) and b).
d) Show with Cantor's diagonal argument that S is uncountable.

3) Topic: **Cardinality of sets**
a) Prove that every infinite subset A of a countable infinite set S is countable infinite. Hint: Prove the fact first for $S = \mathbb{N}$. Construct then a bijective function $f : \mathbb{N} \to A$.
b) Prove that the Cartesian product $A \times B$ of two countable infinite sets A, B is countable infinite. Hint: Prove the fact first for $A = B = \mathbb{N}$. Use then the bijective function $f : \mathbb{N} \to \mathbb{N} \times \mathbb{N}$ which has been done in question 1.d).

4) (*) Topic: **General probability measures.**
Let $(\Omega, \mathcal{A}, P)$ be a probability space. A point $\omega \in \Omega$ is called an **atom**, if $\{\omega\} \in \mathcal{A}$ and $P[\{\omega\}] > 0$. Show that the set of atoms in Ω is either finite or countable infinite.

5) Topic: **General probability spaces.**
Given the probability space $(\Omega = [0,1], \mathcal{A}, P)$, where $\mathcal{A}$ is the Borel σ-algebra on $[0,1]$ (the smallest σ-algebra which contains all intervals (a,b) with $a < b$) and P is the Lebesgue measure defined by $P[(a,b)] = b - a$. Show that the set of points with $P[\{a\}] > 0$ is finite or countable infinite.

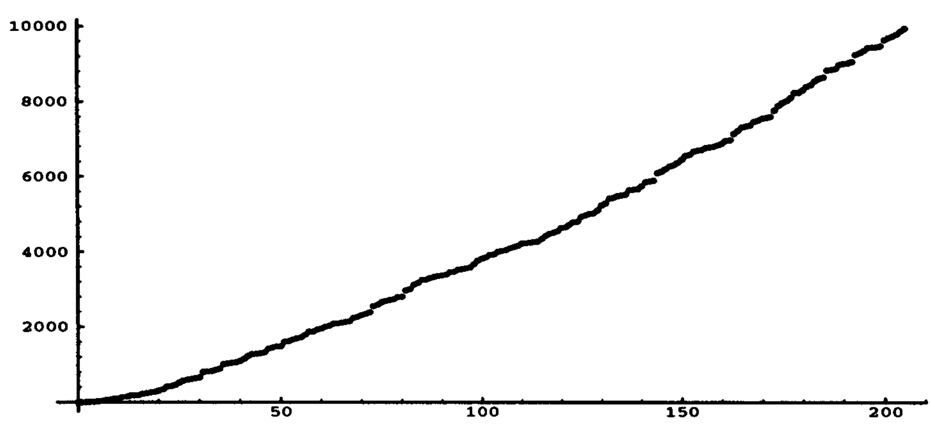
6) (**) Topic: **General probability spaces.**
Given the probability space $(\Omega = [0,1]^2, \mathcal{A}, P)$, where $\mathcal{A}$ is the Borel σ-algebra on $[0,1]^2$ (the smallest σ-algebra which contains all sets $(a,b) \times (d,c)$) and where P is defined by $P[(a,b) \times (c,d)] = (b-a) \cdot (d-c)$. Let $A = \{(x,y) \in [0,1]^2 \mid x = y\}$ (the diagonal in the square Ω). Show that $P[A] = 0$.

7) Topic: **General probability spaces.** (A simplified version of Bertrand's paradox)
Let Ω be the disc $\{x^2 + y^2 \leq 1\} \subset \mathbb{R}^2$. Problem: "What is the probability that a point has no more than distance $1/2$ from the center."
First answer: the probability is $1/2$, since the problem is radially symmetric, it is the probability that the radius in $[0,1]$ is smaller than $1/2$.
Second answer: the probability is $1/2$ since the area of the disc with radius $1/2$ is $1/4$.

Determine the probability space $(\Omega, \mathcal{A}, P)$ for which the first answer is right and the probability space $(\Omega, \mathcal{A}, P)$ for which the second answer is right.

8) Topic: **General probability spaces.**
Model each of the three solutions of (the in class discussed) paradox of Bertrand with a probability space.

9) (*) Topic: **Random variables on discrete probability spaces.**
Let $\Omega = \{1, 2, 3, \ldots\} = \mathbb{N}, \mathcal{A} = \{A \subset \Omega\}$. Define the function P on $\mathcal{A}$ through $P[A] = \sum_{i \in A} 2^{-(i+1)}$.
a) Verify that $(\Omega, \mathcal{A}, P)$ is a probability space.
b) Compute the value of the infinite sum $\sum_{n=1}^{\infty} np^{-n}$ for any $p > 0$. Hint: call the sum $f(p)$ and use a differentiation trick.
c) Compute the expectation of the random variable $X(\omega) = \omega$. Hint: use the answer in b).



Graph showing for each n the n'th prime twin.

Ma2c, O. Knill, **May 1995**

Week 7

Due: Monday, May 22, 1995

Topics: Borel Cantelli, Random variables, Integration, Expectation, Variance, Distribution function

DEFINITION: Given a sequence of independent events in a probability space. Define $A_\infty := \limsup_{n\to\infty} A_n := \bigcap_{m=1}^\infty \bigcup_{n\geq m} A_n$. We have $A_\infty = \{\omega \mid \omega \text{ is in infinitely many } A_i\}$.
BOREL CANTELLI LEMMA: (Monkey typing Shakespeare)

a) If $\sum_n \mathrm{P}[A_n] < \infty$, then $\mathrm{P}[A_\infty] = 0$.
b) If $\sum_n \mathrm{P}[A_n] = \infty$, then $\mathrm{P}[A_\infty] = 1$.

DEFINITION: Integration, Expectation: Denote with $\mathcal{S}$ the set of random variables taking finitely many values: Define for $X \in \mathcal{S}$

$$\mathrm{E}[X] := \sum_{a\in X(\Omega)} a \cdot \mathrm{P}[X = a] \; .$$

Let $\mathcal{L}^1$ be the set of random variables X for which $\sup_{Y\in\mathcal{S}, Y\leq|X|} \mathrm{E}[Y] < \infty$. For $X \in \mathcal{L}^1$ and $X \geq 0$, the **integral** or **expectation** is defined as

$$\mathrm{E}[X] := \sup_{Y\in\mathcal{S}, Y\leq X} \mathrm{E}[Y] \; .$$

In general, we decompose X into $X = X^+ - X^-$ with $X^\pm \geq 0$ and put $\mathrm{E}[X] = \mathrm{E}[X^+] - \mathrm{E}[X^-]$. We write also $\int_\Omega X \, dP$ for $\mathrm{E}[X]$ since **expectation is integration.** Variance, Covariance etc. are defined as in the finite case: $\mathrm{Var}[X] = \mathrm{E}[(X - E[X])^2], \mathrm{Cov}[X, Y] = \mathrm{E}[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])]$.
The **distribution function** of a random variable is defined as $F(t) = \mathrm{P}[x \leq t]$. If $F(t) = \int_{-\infty}^t f(s) \, ds$, then f is the **probability density function** of X.

1) Topic: **General probability spaces.**
Consider the space $[0,1]$ and the algebra $\mathcal{A}$ which is the smallest σ-algebra containing all open intervals (a,b), $a < b$.
a) Verify: every set $\{\omega\}$, $\omega \in [0,1]$ is in $\mathcal{A}$. Hint: Show first $\{0,1\} \in \mathcal{A}$.
b) Show that $\mathrm{P}[\{\omega\}] = 0$.
c) Use b) to show that in the definition of the probability measure, the σ-additivity can not be replaced by the general additivity $\mathrm{P}[A] = \sum_{s \in S} \mathrm{P}[A_s]$ for any set S. Hint: take $S = \Omega$ and $A_\omega = \{\omega\}$.

2) Topic: **General probability spaces, Banach-Tarsky paradox**
We mentioned in class the Banach Tarsky paradox which is the mathematical theorem[1] that one can write the unitball S in $\mathbb{R}^3$ as a disjoint union of 5 sets $A = A_1 \cup A_2 \cup A_3 \cup A_4 \cup A_5$ such that after some rotation and translation of each set A_i, the sets $A_1 \cup A_2 \cup A_3$ and $A_4 \cup A_5$ are both again unitballs.
Show that at least one of the functions $X_i = 1_{A_i} : \mathbb{R}^3 \to \mathbb{R}$ is not a random variable. Read *critically* the definition of a random variable in Apostol's book Vol 2, p. 512.

3) (*) Topic: **Borel Cantelli Lemma.**
We are throwing dices infinitely often. The probability space $(\Omega, \mathcal{A}, P)$ is given by $\Omega = \{1,2,3,4,5,6\}^{\mathbb{N}} = (\omega_1, \omega_2, \dots)$, $\mathcal{A}$ is the Borel σ-algebra, which is the smallest σ-algebra containing all the sets of the form $C_{n,j} = \{\omega \mid \omega_n = j\}$ with $n \in \mathbb{N}, j \in \{1,2,3,4,5,6\}$ and where P is determined by $\mathrm{P}[C_{n,j}] = 1/6$. Prove with Borel-Cantelli, that for almost all experiments ω, there are infinitely many n, for which $\omega_n = \omega_{n+1} \dots = \omega_{n+1000} = 6$. (Almost all means that the set of exceptions has measure zero).

4) (*) Topic: **Expectation, Variance.**
A random number generator produces random numbers in the interval $[0,1]$ such that the probability that the number is in an interval (a,b) is $b^2 - a^2$.
a) Model this situation with a random variable X on the Lebesgue probability space $([0,1], \mathcal{A}, P)$.

---

[1] **As you know, every mathematical theorem is either 1. obvious or 2. not so obvious or 3. obviously false. The Banach-Tarsky theorem belongs clearly to the third category. You can find a proof in Amer. Mathem. Monthly, 86, 151-161, 1979.**

b) Compute E[X] which is the expected average number produced by X.
c) Compute Var[X], which measures the fluctuations.

5) Topic: **Expectation.**
For which $\alpha \in \mathbb{R}$ does the random variables $X(x) = x^\alpha$ on the Lebesgue probability space $([0,1], \mathcal{A}, P)$ have a finite mean E[X] and for which α does the variance Var[X] exist?

6) (**) Topic: **Expectation.**
Let $\Omega = [-1,1] \times [-1,1]$, $\mathcal{A}$ the Borel σ-algebra on Ω and P the measure defined by $\mathrm{P}[(a,b) \times (c,d)] = (b-a)(d-c)/4$. Define the random variable $X(\omega) = |\omega| = \sqrt{x^2+y^2}$ which measures the euclidean distance of $\omega = (x,y)$ to the origin $(0,0)$. Compute E[X], the expected distance from 0 and $\sigma[X] = \sqrt{\mathrm{Var}[X]}$, the expected deviation from E[X]. Hint. We recommend that you consult a symbolic friend[2] like Mathematica for the computation of the integral(s). In Mathematica, one computes for example with *Integrate*[*Sqrt*[x], x] symbolically the indefinite integral $\int \sqrt{x}\, dx$.

7) (*) Topic: **Expectation, Petersburg paradox.**
The Petersburg casino (discussed in class) is modeled by the probability space $(\{0,1\}^{\mathbb{N}}, \mathcal{A}, P)$, where $\mathcal{A}$ is the Borel σ-algebra (the smallest σ-algebra containing all the sets $A_{n,j} = \{\omega_n = j\}$ which is the event that at the n'th time either $j = 0$ or $j = 1$ appears) and P is the probability measure defined by $\mathrm{P}[A_{n,j}] = 1/2$. The game is defined by the winning function $X(\omega) = 2^{T(\omega)}$, where $T(\omega) = \{\text{smallest } n \mid \omega_n = 0\}$. We have seen that E[$X$] $= \infty$. So, any finite fee is actually very favourable for us. The paradox is that nobody would go to the Petersburg casino and pay even 10 dollars and you have seen in class how fast I lost 30 dollars. One problem is that for example the case $X(\omega) = 2^{300}$ is very, very unprobable but you would win 2^{300} dollars (!), which is more than the number of protons in the universe[3]
a) We know from class that $\mathrm{P}[T = k] = 2^{-k}$. Compute E[$T$], the expected number of times we have to wait, until 1 occurs. Hint: use HW Week 6).
b) D.Bernoulli proposed to replace the useless fee E[X] by an expected

---

[2]Reference books for integrals have statisticallly an error of 15 %, which means that every 7'th integral, you look up, is wrong.

[3]The intuition sais that there is a problem to make so many bank notes in this case since every bank note needs at least one proton. (Pigeon-hole-principle)

profit $\mathrm{E}[\sqrt{X}]^2$. Compute that number which is Bernoulli's suggestion for an entrance fee for the Petersburg casino.

8) Topic: **Expectation.**
A and B are throwing darts onto a wall modeled by the complex plane $\mathbf{C}$. Their target is the disc $\Omega = \{|z| < 1\}$ which is hit randomly. More precicely, every sector $\{\arg(z) \in (\alpha, \beta) \subset [0, 2\pi)\}$ is hit with probability $(2\pi)^{-1}(\beta - \alpha)$ and every ring $\{a < |z| < b\}$ is hit with probability $\int_a^b f(r)\, dr$, where $f(r)$ is a function depending on the player, A throws the darts with a precision so that $f(r) = 2r$, B throws the darts so that $f(r) = 3r^2$. The disc is partitioned into 3 rings of equal radius and giving points as follows. For $\omega \in \Omega$, the win is given by

$$X(\omega) = 5 \cdot 1_{\{|\omega| \leq 1/3\}} + 3 \cdot 1_{\{1/3 \leq |\omega| \leq 2/3\}} + 1 \cdot 1_{\{2/3 \leq |z| \leq 1\}} \,.$$

a) Determine probability spaces $(\Omega, \mathcal{A}, \mathrm{P}_A)$ and $(\Omega, \mathcal{A}, \mathrm{P}_B)$ for the two players.
b) What is the expected win $\mathrm{E}_A[X]$ for A and the expected win $\mathrm{E}_B[X]$ for B, where E_A is the expectation value with respect to P_A and E_B is the expectation value with respect to P_B.

9) (*) Topic: **Expectation, Distribution function, distribution of prime numbers.**
The prime number theorem says that the n'th prime number p_n is roughly $p_n \sim n \log(n)$ in the sense that this becomes better and better for $n \to \infty$. So, $q_n = p_{n+1}/\log(p_n)$ is roughly equally spaced. What is the fine structure of the distribution of prime numbers? Experimentally, it seems that they behave quite randomly, namely that $n \mapsto (q_{n+1} - q_n)/\log(n)$ behaves like an exponential distributed random variable X satisfying[4]

$$\mathrm{P}[X \in (a, b)] = \int_a^b \lambda e^{-\lambda t}\, dt \,.$$

a) Show that $([0, \infty), \mathcal{A}, \mathcal{P})$ is a probability space, where $\mathcal{A}$ is the Borel σ-algebra and where P is the above measure with some parameter $\lambda > 0$. Hint: you need only to show that $\mathrm{P}[\Omega] = 1$.
b) Define the random variable $X(\omega) = \omega$. What is $\mathrm{E}[X^3]$?
c) Compute the distribution function of X.

[4] It is an unsolved mathematical problem if such a relation can be proven asymptotically. P. Sarnak was talking about that in the Mathematics Colloquium last week.

Ma2c, O. Knill, **May 1995**

Week 8

Due: Tuesday, May 30, 1995 (day after Memorial day)

Topics: (Multidimensional) distribution functions, Expectation and variance of absolutely continuous random variables, Transformations of distributions, Characteristic functions

DEFINITION: The **Distribution function** of a random variable X is $F(t) = \mathrm{P}[X \leq t]$. **Absolutely continuous random variable**: the **probability density function** $F' = f$ exists. **Discrete random variable**: F is piecewise constant with countably many jump discontinuities. The **expectation** and **variance** of a continuous distribution is

$$m = \mathrm{E}[X] = \int_{-\infty}^{\infty} x f(x)\, dx, \ \mathrm{Var}[X] = \int_{-\infty}^{\infty} (x-m)^2 f(x)\, dx \ .$$

The **expectation** and **variance** of a discrete distribution is

$$m = \mathrm{E}[X] = \sum_{a \in X(\Omega)} a \mathrm{P}[X = a], \ \mathrm{Var}[X] = \sum_{a \in X(\Omega)} (a-m)^2 \mathrm{P}[X = a] \ .$$

EXAMPLES OF ABSOLUTELY CONTINUOUS DISTRIBUTIONS:

Distribution	Density $f(x) = F'(x) =$	Parameters	Mean	Variance
Uniform	$1_{[a,b]} \cdot (b-a)^{-1}$	$a < b$	$(a+b)/2$	$(b-a)^2/12$
Exponential	$\lambda e^{-\lambda x}$	$\lambda > 0$	$1/\lambda$	$1/\lambda^2$
Normal	$(2\pi\sigma^2)^{-1/2}\, e^{-\frac{(x-m)^2}{\sigma^2}}$	$m \in \mathbf{R}, \sigma^2 > 0$	m	σ^2

EXAMPLES OF DISCRETE DISTRIBUTIONS:

Distribution	$\mathrm{P}[x = k] =$	Parameters	Mean	Variance
Binomial	$\binom{n}{k} p^k (1-p)^{n-k}$	$n \in \mathbf{N}, p \in [0,1]$	np	$np(1-p)$
Poisson	$\frac{\lambda^k}{k!} e^{-\lambda}$	$\lambda > 0$	λ	λ
Geometric	$(1-p)^{k-1} p$	$p \in (0,1)$	$1/p$	$1/p^2$

The **characteristic function** of X is defined as $\phi_X(t) = \mathrm{E}[e^{itX}]$. Discrete case: $\phi_X(t) = \sum_{a \in X(\Omega)} e^{ita} \mathrm{P}[X = a]$. Continuous case: $\phi_X(t) = \int_{-\infty}^{\infty} e^{itx} f(x)\, dx$.

1) (*) Topic: **Distribution functions, Normal distribution.** We consider the normal distribution with probability density $f = f_{m,\sigma} = (2\pi\sigma^2)^{-1/2} \, e^{-\frac{(x-m)^2}{2\sigma^2}}$.
a) Show that $f(x)$ is maximal for $x = m$.
b) For $x = m \pm \sigma$, the second derivative of f is vanishing.
c) Show that $\int_{-\infty}^{\infty} f(x) \, dx = 1$, $\mathrm{E}[X] = m, \mathrm{Var}[X] = \sigma^2$ (this has been done in class and you can of course use your notes).

2) (*) Topic: **Distribution functions, Erlang distribution.** Consider a random variable X with distribution function $F(t) = \mathrm{P}[X \leq t]$ defined by the probability density function $f(x) = F'(x)$ which is zero on $\{x < 0\}$ and

$$f(x) = \frac{\lambda^k x^{k-1}}{(k-1)!} e^{-\lambda x}, \text{ for } x \geq 0 \, ,$$

where λ, k are parameters.
a) Show that the exponential distribution is a special case.
b) Verify that f is indeed a density functions, that means, show that $\int_0^\infty f(t) \, dt = 1$.
c) Compute $\mathrm{E}[X]$.
d) Compute $\mathrm{Var}[X]$.

3) (*) Topic: **Distribution functions, Exponential distribution.** A radioactive sample containing **Lutetium** (Lu) [1] emits α rays (He^4 nuclei). Assume the waiting time X (measured in seconds) for a decay is exponentially distributed with parameter $\lambda = 3$.
a) What is the probability to get a decay in 1 second?
b) How long does one have to wait in average to measure a decay?

4) (*) Topic: **Multi-dimensional distributions.** Given $\sigma_i > 0$, and $m_i \in \mathbb{R}$. Show that

$$f(t_1, t_2) = \frac{1}{\sqrt{(2\pi)^2 \sigma_1^2 \sigma_2^2}} \, e^{-\frac{(t_1-m_1)^2}{2\sigma_1^2} - \frac{(t_2-m_2)^2}{2\sigma_2^2}}$$

is the multi-dimensional density function for some random vector (X_1, X_2). Hint: Show that $f = f_1 \cdot f_2$, where f_1, f_2 are density functions which belong

---

[1] A Mathematica standard package gives more information: <<Miscellaneous ‘ ChemicalElements ‘ AtomicNumber[Lutetium] AtomicWeight[Lutetium]

to normal distributed random variables. Use also that $f_{(X,Y)} = f_X f_Y$ means that X, Y are independent.

5) **Transformation of multi-dimensional distributions.**
Let X have a uniform distribution on $[0, \pi]$.
a) Compute the distribution function of the random variable $Z = \cos(X)$.
b) Let Φ, R be two independent random variable, where Φ has the uniform distribution on $[0, 2\pi]$ and R has an exponential distribution on $[0, \infty)$ with parameter $\lambda = 1$. Compute the multidimensional density function of (Φ, R). Hint: Use a fact from theory about density functions of random vectors with independent components.
c) Compute the multi-dimensional density function of the random vector $(R\cos(\Phi), R\sin(\Phi))$ using the transformation rule.

6) (**) Topic: **Characteristic functions.**
a) Compute the characteristic function of a discrete random variable X which is Poisson distributed:

$$\mathrm{P}[X = k] = \frac{\lambda^k}{k!} e^{-\lambda} .$$

b) Use *a*) and the theory, to compute $\mathrm{E}[X^4]$ for a Poisson distributed random variable with parameter λ.

7) Topic: **Characteristic functions.** Compute the characteristic function of a random variable with the Erlang distribution with parameter k and λ (see problem above).

8) Topic: **Erlang distribution.** Show that the sum of k independent λ-exponential distributed random variables has the Erlang distribution with parameter (k, n). Hint: Use the solution of the last problem and a fact about characteristic functions of sums of independent random variables.

9) Topic: **Binomial distribution.**
a) Show that every Binomial distribution is the distribution of the sum of n random variables.
b) Use a) to compute the characteristic function of a random variable which is (n, p)-Bernoulli distributed.

Preliminary information about the final examination. The topics of the final examination are the material of the second until 10'th week (no vector analysis). The content of the course:

CHAPTER 1: (FINITE PROBABILITY SPACES)

2. Week : 1. Boolean algebras, 2. Probability spaces, 3. Three constructions.
3. Week : 4. Independent events, 5. Random variables, 6. Expectation.
4. Week : 7. Variance, Covariance, Correlation. 8. Independent random variables. 9. Random walk and Game systems.

CHAPTER 2: (GENERAL PROBABILITY SPACES)

5. Week : 1. Cardinality of sets. 2. General probability spaces.
6. Week : 3. Discrete Random variables, Expectation and variance of discrete random variables. 4. The Borel Cantelli lemma.
7. Week : 5. Integration and expectation, 6. Distribution of random variables. 7. Examples of continuous distributions 8. Expectation and variance of continuous distributions.
8. Week : 9. Multidimensional distributions. 10. Transformations of distributions. 11. Characteristic functions. 12 Distributions of sums of independent random variables.

CHAPTER 3: (LIMIT THEOREMS)

9. Week : 1. Chebychev inequality, 2. Weak law of large numbers.
10. Week : 3. Strong law of large numbers. 4. Central limit theorem.

The problems will be in a similar style as the homework problems.

Ma2c, O. Knill, **June 1995**

Week 9

Topics: Chebychev inequality, Weak law of large numbers

CHEBYCHEV-MARKOV INEQUALITY.
Let $h : \mathbf{R}^+ \to \mathbf{R}^+$ be a monoton function and $X \geq 0$ a random variable with $h(X) \in \mathcal{L}^1$. Then for all $c > 0$

$$h(c) \cdot \mathrm{P}[X \geq c] \leq \mathrm{E}[h(X)] \; .$$

Proof. Take the expectation of $h(c) 1_{X \geq c}(\omega) \leq h(X)(\omega)$ using the monotonicity and linearity of the expectation.

CHEBYCHEV INEQUALITY.
If $X \in \mathcal{L}^2$, then for all $c > 0$

$$\mathrm{P}[|X - \mathrm{E}[X]| \geq c] \leq \frac{\mathrm{Var}[X]}{c^2} \; .$$

Proof. Apply Chebychev-Markov to $Y = |X - \mathrm{E}[X]|$ and $h(x) = x^2$.

DEFINITION.
A sequence of random variables X_n **converges in probability** to a random variable X, if for all $\epsilon > 0$,

$$\lim_{n \to \infty} \mathrm{P}[|X_n - X| \geq \epsilon] = 0 \; .$$

WEAK LAW OF LARGE NUMBERS.
Assume X_i have common expectation $E[X_i] = m$ and satisfy $\sup_n \frac{1}{n} \sum_{i=1}^n \mathrm{Var}[X_i] < \infty$. If X_n are pairwise uncorrelated, then $\sum_{i=1}^n X_i / n \to m$ in probability.

IMPORTANT SPECIAL CASE.
If X_i are independent random variables with the same distribution for which the mean m and variance exists, then $\sum_{i=1}^n X_i / n \to m$ in probability.

1) Topic: **Chebychev inequality**
Given a sequence of independent random variables X_n which have the standard normal distribution with density

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} .$$

Call $S_n = X_1 + X_2 + \ldots + X_n$. Estimate with Chebychev's inequality

$$\mathrm{P}[|S_n/n| \geq \epsilon]$$

from above in terms of ϵ and n.

2) Topic: **Chebychev-Markov inequality.**
Show that every random variable satisfies for $\epsilon > 0$

$$\mathrm{P}[|X| \geq \epsilon] \leq \frac{\mathrm{E}[e^{\sqrt{|X|}}]}{e^{\sqrt{\epsilon}}} .$$

3) Topic: **Convergence in probability**
Consider the probability space $(\Omega, \mathcal{A}, P)$, where $\Omega = [0, 1]$, $\mathcal{A}$ is the Borel σ-algebra of $[0, 1]$ and P is the Lebesgue probability measure defined by $\mathrm{P}[[a, b)] = b - a$.
a) Show that the sequence of random variables $X_n(x) = n^2 \cdot x^n$ converges in probability to a limit X and compute this limit.
b) Show that $\mathrm{E}[|X_n - X|]$ does not converge to zero.

4) Topic: **Weak law of large numbers**
A coin is tossed repeatedly so that heads turns up with probability p in each toss and tail turns up with probability $1 - p$. Let H_n be the number of heads and T_n be the number of tails in n tosses.
a) Show that $(H_n - T_n)/n$ converges in probability to a limit c.
b) Compute that limit.
Hint: Introduce a random variable X_n, which is 1, if head turns up and which is -1, if tail turns up.

This is the last homework assignment and has not to be turned in. In an emergency case, if you feel in trouble with not having enough homeworks, there is the following possibility: you turn in your solution of this homework to the TA or to me until Monday morning 11^{00}. This can give you credit **for maximal 50% of one homework assignment** and can replace a homework, which you have missed or in which you got less then 50% of the points. In any case, it is advisable to look over this problem set. On Monday, June 5, the solutions will be distributed.
The final will be given out Monday, June 5 and due Monday, June 12, $11^{00}, AM$ (For the senior students, the final is given out Friday, June 2 and due: Wednesday, June 7, $11^{00}AM$). The topics of the final examination are the material of the second until beginning 10'th week (no vector analysis).

CHAPTER 1: (FINITE PROBABILITY SPACES)
2. Week: 1. Boolean algebras, 2. Probability spaces, 3. Three constructions.
3. Week: 4. Independent events, 5. Random variables, 6. Expectation.
4. Week: 7. Variance, Covariance, Correlation. 8. Independent random variables. 9. Random walk and Game systems.
CHAPTER 2: (GENERAL PROBABILITY SPACES)
5. Week: 1. Cardinality of sets. 2. General probability spaces.
6. Week: 3. Discrete Random variables, Expectation and variance of discrete random variables. 4. The Borel Cantelli lemma.
7. Week: 5. Integration and expectation, 6. Distribution of random variables. 7. Examples of continuous distributions 8. Expectation and variance of continuous distributions.
8. Week: 9. Multidimensional distributions. 10. Transformations of distributions. 11. Characteristic functions. 12 Distributions of sums of independent random variables.
CHAPTER 3: (LIMIT THEOREMS)
9. Week: 1. Chebychev inequality, 2. Weak law of large numbers.
10. Week: 3. Strong law of large numbers. 4. Central limit theorem.

The problems will be in a similar style as the homework problems.

EXAMS

MA 2c: PROBABILITY

Oliver Knill, Room 172, Tel.: 4325, e-mail: knill@cco.caltech.edu

Course given in the third term 1995 Caltech: Content according to the weeks:

CHAPTER 1: (FINITE PROBABILITY SPACES)
2. Week : 1. Boolean algebras, 2. Probability spaces, 3. Three constructions.
3. Week : 4. Independent events, 5. Random variables, 6. Expectation.
4. Week : 7. Variance, Covariance, Correlation. 8. Independent random variables. 9. Random walk and Game systems.

CHAPTER 2: (GENERAL PROBABILITY SPACES)
5. Week : 1. Cardinality of sets. 2. General probability spaces.
6. Week : 3. Discrete Random variables, Expectation and variance of discrete random variables. 4. The Borel Cantelli lemma.
7. Week : 5. Integration and expectation, 6. Distribution of random variables. 7. Examples of continuous distributions 8. Expectation and variance of continuous distributions.
8. Week : 9. Multidimensional distributions. 10. Transformations of distributions. 11. Characteristic functions. 12 Distributions of sums of independent random variables.

CHAPTER 3: (LIMIT THEOREMS)
9. Week : 1. Chebychev inequality, 2. Weak law of large numbers.
10. Week : 3. Strong law of large numbers. 4. Central limit theorem.

Contents

1 Finite probability spaces **4**

1.1 Boolean algebra of a finite set 4
1.2 Finite probability spaces 6
1.3 Constructions of new probability spaces 9
1.4 Independent events 13
1.5 Random variables 14
1.6 Some more examples 15
1.7 Expectation 18
1.8 Variance, Covariance, Correlation 18
1.9 More examples about variance 20
1.10 Correlation line 21
1.11 Independent random variables 22

2 General probability spaces **24**

2.1 Cardinality of sets 24
2.2 General probability spaces 25
2.3 Appendix: Banach-Tarski's paradox 28
2.4 Discrete Random variables, expectation 28
2.5 Appendix: the Petersburg paradox 31
2.6 Independence and the Borel Cantelli lemma 32
2.7 Integration and expectation 33
2.8 Distribution of a random variable 35
2.9 Examples of probability density functions 38
2.10 Joint distribution of random variables 42
2.11 Transformation of random variables 43
2.12 Characteristic functions 44
2.13 Distributions of independent random variables 47
2.14 Inequalities 48
2.15 The weak law of large numbers 49
2.16 Experiment 51

2.17 The strong law of large numbers 52
2.18 Appendix. Normality of numbers 54
2.19 The central limit theorem 54

3 Summaries **56**

Chapter 1

Finite probability spaces

1.1 Boolean algebra of a finite set

Definition. Let Ω be a finite set. A set $\mathcal{A}$ of subsets of Ω is called a **Boolean algebra** if

(i) $\Omega \in \mathcal{A}$,
(ii) $A \in \mathcal{A} \to A^c \in \mathcal{A}$,
(iii) $A, B \in \mathcal{A} \Rightarrow A \cup B \in \mathcal{A}$

The pair $(\Omega, \mathcal{A})$ is also called a **measurable space.**

A Boolean algebra is closed under all set-theoretical operations:
Notation:
$\bigcup_{k=1}^n A_k$ is a short hand notation for $A_1 \cup A_2 \dots, A_n$.
$\bigcap_{k=1}^n A_k$ is a short hand notation for $A_1 \cap A_2 \cap \dots \Delta A_n$.
$\Delta_{k=1}^n A_k$ is a short hand notation for $A_1 \Delta A_2 \dots \Delta A_n$.
$A \setminus B = \{\omega \in A \mid \omega \notin B\}$.

$$\omega \in A \cap B \Leftrightarrow \omega \in A \text{ and } \omega \in B$$

$$\omega \in A \cup B \Leftrightarrow \omega \in A \text{ or } \omega \in B$$

$$\omega \in A \Delta B \Leftrightarrow \omega \in A \text{ xor } \omega \in B$$

$$\omega \in A \setminus B \Leftrightarrow \omega \in A \text{ andnot } \omega \in B$$

Lemma 1.1.1 *Let $\mathcal{A}$ be a Boolean algebra. Then:*
a) $\emptyset \in \mathcal{A}$.
b) $A, B \in \mathcal{A} \Rightarrow A \cap B \in \mathcal{A}$.
c) $A, B \in \mathcal{A} \Rightarrow A \setminus B \in \mathcal{A}$.
d) $A, B \in \mathcal{A} \Rightarrow A \Delta B \in \mathcal{A}$.

Proof. a) $\emptyset = \Omega^c$.
b) $A \cap B = (A^c \cup B^c)^c$.
c) $A \setminus B = A \cap B^c$
d) $A \Delta B = (A \cup B) \setminus (A \cap B)$ □

Examples.
1) Let $\mathcal{A}$ be the set of all subsets of Ω. This is a Boolean algebra.
2) Let $\mathcal{A}$ be the set $\{\emptyset, \Omega\}$. It is a Boolean algebra.
3) Let $\mathcal{A}$ be the set $\{\emptyset, A, A^c, \Omega\}$. It is a Boolean algebra. It is the smallest algebra containing A.
4) Let S be a finite set of measureable sets. Then $\mathcal{A}(S)$, the smallest algebra containing S is a finite algebra.

Lemma 1.1.2 *If $\mathcal{A}$ and $\mathcal{B}$ are Boolean algebras over Ω, then $\mathcal{A} \cap \mathcal{B}$ is a Boolean algebra over Ω.*

Proof. $\Omega \in \mathcal{A}, \Omega \in \mathcal{B} \Rightarrow \Omega \in \mathcal{A} \cap \mathcal{B}$.
$A \in \mathcal{A} \cap \mathcal{B}$, $B \in \mathcal{A} \cap \mathcal{B}$, then $A \cap B \in \mathcal{A} \cap \mathcal{B}$.
$A \in \mathcal{A} \cap \mathcal{B}$, then $A^c = \in \mathcal{A} \cap \mathcal{B}$. □

Remark. It is not true that if $\mathcal{A}$ and $\mathcal{B}$ are Boolean algebras, then $\mathcal{A} \cup \mathcal{B}$ are Boolean algebras.

Probabilistic interpretation: think of Ω as the set of possible **experiments** and of $\mathcal{A}$ as a set of **events**. A translation into common language is:

$A \cup B$: Events A or B occurs.
$A \cap B$: Both events A and B occur.
$A^c \cap B^c$: Neither A nor B occur.
$A \Delta B$: Exactly one of A or B occurs.
$A \cap B = \emptyset$: The events A and B exclude each other.
$A \subset B$: If A occurs, then B occurs.

Example 1). We are throwing a dice two times. Each experiments is a point in a finite set

$$\Omega = \{(a, b) \mid a, b \in \{1, 2, 3, 4, 5, 6\}\}$$

containing 36 elements. An example of an event is

$$A = \{(a, b) \in \Omega \mid a + b = 7\} .$$

It contains 6 points in Ω. (Make the experiment).

Example 2). Consider a chessboard with 8×8 plaquettes. Occupy with probability $1/2$ every plaquette of the chessboard. The probability space has 2^{64} elements. $\mathcal{A}$, the set of subsets of Ω has $2^{2^{64}}$ elements.

We met the last time a problem:

Given two events A, B. What is the interpretation of $A \subset B$?

Clearly $\omega \in A \Rightarrow \omega \in B$. So, if A is given by a statement S, and B by a statment T, then $S \Rightarrow T$.

On the other hand, if A has the property S and B has the property T, then the property T is a part of property S and I ment property T implies property S.

The problematic is the following: our language is not precice. Sometimes, adjectives are used for narrowing properties, sometimes used to enrich properties. This is reflected best in the following joke:

All thiefs are human beeings. Therefore, all good thiefs are good human beeings.

The problem is that the first time, good is used in a enriching way, the second time in a narrowing way.

Example. Throwing dices:

THE PROBABILITY SPACE:

$$\Omega = \{(a,b) \mid a \in \{1,2,\dots,6\}, b \in \{1,2,\dots,6\} .$$

$$\mathcal{A} = \{A \subset \Omega\} .$$

$$P[A] = \frac{|A|}{|\Omega|} .$$

EXAMPLES OF EVENTS:

$$A = \{a+b=7\}$$

$$B = \{a+b \leq 7\}$$

We have $A \subset B$ and the statement "$a+b=7$" implies the statement "$a+b \geq 7$". Properties are used always in a narrowing sense.

1.2 Finite probability spaces

Definition. Let $\mathcal{A}$ be a Boolean algebra over a finite set Ω. A function $P : \mathcal{A} \to \mathbf{R}$ is called **finitely additive** if

$$P[A \cup B] = P[A] + P[B]$$

for all disjoint sets $A, B \in \mathcal{A}$.

Definition. A finitely additive function P on $\mathcal{A}$ is called a **measure** if $P[A] \geq 0$ for all $A \in \mathcal{A}$. A measure is called a **probability measure** if it is a measure and $P[\Omega] = 1$.

Definition. If P is a probability measure on $(\Omega, \mathcal{A})$, where Ω is a finite set and $\mathcal{A}$ is a Boolean algebra on Ω, we call $(\Omega, \mathcal{A}, P)$ a **finite probability space.**

Proposition 1.2.1 *Let $(\Omega, \mathcal{A}, P)$ be a finite probability space. Then:*
a) $P[\emptyset] = 0$.
b) $A \subset B \Rightarrow P[A] \leq P[B]$.
c) $P[A^c] = 1 - P[A]$.
d) $P[A \cup B] = P[A] + P[B] - P[A \cap B] \leq P[A] + P[B]$.
e) $P[\bigcup_{i=1}^n A_i] = \sum_{k=1}^n (-1)^{k-1} \sum_{1 \leq i_1 < i_2 < \ldots < i_k \leq n} P[A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}]$.

Proof. a) Since $\emptyset$ is disjoint from $\emptyset$, we have by the finite additivity

$$P[\emptyset] = P[\emptyset \cup \emptyset] = P[\emptyset] + P[\emptyset]$$

and so $P[\emptyset] = 0$.
b) If $A \subset B$, then A is the disjoint union $A = B \cup (A \setminus B)$ so that by the finite additivity and nonnegativity of P

$$P[A] = P[B] + P[A \setminus B] \geq P[B] .$$

c) A and A^c are disjoint. By the finite additivity

$$P[A] + P[A^c] = P[A \cup A^c] = P[\Omega] = 1 .$$

d) is clear and a special case of $e)$
e) If we take $\sum_i P[A_i]$, then we have counted twice the sets $A_{i_1} \cap A_{i_2}$ with $i_i < i_2$. We take them away. But then, we have taken away twice the sets $A_{i_1} \cap A_{i_2} \cap A_{i_3}$ we again add them. And so on.
□

Notation. Given $A \in \mathcal{A}$, we denote with $|A|$, the number of elements in A.

Examples.
Let $\mathcal{A}$ be the Boolean algebra of all sets of subset of Ω.
1)

$$P[A] = \frac{|A|}{|\Omega|}$$

is a probability measure on $(\Omega, \mathcal{A})$. It is called the **normalized counting measure.**
2) For every $\omega \in \Omega$,

$$P[A] = 1_A(\omega) = \begin{cases} 1, & \omega \in A \\ 0, & \omega \in A^c \end{cases}$$

is a measure on $(\Omega, \mathcal{A})$.
3) Given $\Omega = \{\omega_1, \omega_2 \ldots, \omega_n\}$ and n nonnegative numbers $p_\omega \in \mathbf{R}$ with $\sum_{\omega \in \Omega} p_\omega = 1$. Then

$$P[A] = \sum_{\omega \in A} p_\omega$$

is a probability measure on Ω. Both examples 1) and 3) are special cases. Every probability measure P is of this form with $p_\omega = P[\{\omega\}]$.

Probabilistic interpretation. A probability space $(\Omega, \mathcal{A}, P)$ determines the experiments as well as its possible events as well as the probabilities that each event occurs. One reads:

$P[A] = p$: The probability that the event A occurs is p.
$P[A] = 0$: The event A doesn't occur almost surely.
$P[A] = 1$: The event A occurs almost surely.

It is important to distinguish between: the event A does not occur $A = \emptyset$ and the event A does not occur almost surely, $P[A] = 0$.

Cultural remark. Finite probability spaces seem to be very special and not so useful. Only in the 80'th of this century, Nelson has shown that all the calculus of probabilities can be done on finite probability spaces without loss of generality. This is possible because of Nonstandard analysis with which much of the current mathematics gets much simpler. However, it will probabily some decades until this knowledge will enter in courses of calculus or probability.

Problem. Dave has two children. One of them is a girl. What is the probability that the other child is a boy.

Model the probability space: Without knowledge, we had the probability space $\Omega = \{BG, GB, BB, GG\}$ with the algebra of all subsets and with the Laplace probability measure. Let A be the event that one of them is a girl. We have

$$A = \{BG, GB, GG\} .$$

We have to model this new probability space so that every of these three possibilities have the same probability. Clearly, the probability that the other child is a boy is 2/3.

We will come back to this example.

(In average, 83 give the right answer.) Computing probabilities of events in a probability space $(\Omega, \mathcal{A}, P)$ is a combinatorial problem.

Let Ω be the set of subsets of a finite set X with n elements.

1) Then Ω contains

$$n! = n \cdot (n-1) \cdot (n-2) \cdots 2 \cdot 1$$

elements.

2) Let A be the event in Ω that ω contains k elements.

$$|A| = \left(\begin{array}{c} k \\ n \end{array} \right) = \frac{n!}{k! \cdot (n-k)!}$$

elements.

3) Let A be the event that ω contains k different elements.

$$|A| = \frac{n!}{n-k!} = n \cdot (n-1) \cdots (n-k+1)$$

elements.

Let Ω be the set of vectors $(a_1, \dots, a_d)$ with $a_i \in X$, where X is again a finite set with n elements.

1) Ω contains $(n!)^d$ elements.

2) The set $A \subset \Omega$

$$A = \{a_i \in A_i\}, A_i \subset X$$

then A contains $\prod_{i=1}^{d} |A_i|$ elements.

Many combinatorial problems are not so easy:

1) Let Ω be the set of permutations of of $\{1, 2, \ldots, n\}$. Let A_n be the set of permutations in Ω_n which have no fixed point. One can show that

$$|A_n| = n! \sum_{i=0}^{n} \frac{(-1)^i}{i!} .$$

We see that $|A_n|/|\Omega_n| \to e$.

2) Let Ω_n be the set of all subsets R of $\mathbf{N} = \{1, 2, 3 \ldots\}$ such that $\sum_{r \in R} r = n$. The number $|\Omega_n|$ is finite and denoted with $p(n)$ and called the number of partitions of n into positive summands.

3) Let Ω_n be the set of $n \times n$ matrices with entries 0 or 1. How many of them are invertible? For $n = 2$ for example, there are exactly two invertible matrices. We do not intend to solve the general problem.

4) Let Ω_n be the number of alcohols $C_n H_{2n+1} OH$. Two molecules which can go into each other by deformation are considered as identical. For $n = 3$ for example, there are two different molecules.

```
    H   H   H                H   H    H
    |   |   |                |   |    |
H - C - C - C - OH       H - C - C  - C - H
    |   |   |                |   |    |
    H   H   H                H   OH   H
```

The problem is a problem in denumerating trees and we do not intend to solve it here.

1.3 Constructions of new probability spaces

1. Construction: CHANGING OF THE ALGEBRA
Given a probability space $(\Omega, \mathcal{A}, P)$. If $\mathcal{B} \subset \mathcal{A}$ is an other algebra, we call it a **subalgebra.** We get a new probability space

$$(\Omega, \mathcal{B}, P)$$

by restricting P to $\mathcal{B}$.

Interpretation. The probability space $(\Omega, \mathcal{B}, P)$ models a situation, where we know less events than before. This can be interpreted that we can no more observe the experiment with the same accurancy. Example. We are throwing dices and suddenly, we can no more read all of the numbers of the dice but only see whether they are even or odd. $\Omega = \{1, 2, \ldots, 6\}$ with $\mathcal{A} = \{A \subset \Omega\}$ goes over into $\mathcal{A} = \{\emptyset, \Omega, \{1, 3, 5\}, \{2, 4, 6\}\}$.

The size of the algebra measures the accuracy of the measurements.

In applications, one often has not only one subalgebra but many subalgebras and so a family of probability spaces. This is especially important in gambling situations, where with time, the knowledge about the system is changing.

Proposition 1.3.1 *The space $(\Omega, \mathcal{B}, P = P|\mathcal{B})$ is again a probability space.*

2. Construction: PRODUCT PROBABILITY SPACE
Given two probability spaces $(\Omega_i, \mathcal{A}_i, P_i)$ for $i = 1, 2$.
Define the new probability space

$$(\Omega, \mathcal{A}, P)$$

where $\Omega = \Omega_1 \times \Omega_2 = \{(\omega_1, \omega_2) \mid \omega_1 \in \Omega_1, \omega_2 \in \Omega_2\}$ and $\mathcal{A}$ is the smallest Boolean algebra, which contains the sets (A_1, A_2), $A_i \in \mathcal{A}_i$.

(Since there are only finitely many Boolean algebras on $\mathcal{B}$ we can take the intersection of all Boolean algebras, which contain these sets). By requiring

$$P[(A_1, A_2)] = P_1[A_1]P_2[A_2]$$

we determine P on $\mathcal{A}$ since every $A \in \mathcal{A}$ is a finite disjoint union of elements of the form (A_1, A_2).

Proposition 1.3.2 *The product space is again a probability space.*

Proof. Enumerate the atoms of Ω_1 and Ω_2 and attach to them the probabilities a_i, b_i. $P[\{a\}] = p_a$, $P[\{(a_1, a_2)\}] = p_{a_1}p_{a_2}$ by definition. Additivity positivity is clear, normalisation $\sum_i p_i = 1$, $\sum_j q_j = 1$. Therefore, $\sum_{ij} p_i q_j = 1$. □

Example. We are throwing dices twice. The probability space $(\Omega_1, \mathcal{A}_1, P_1)$ describes the first experiment, the probability space $(\Omega_2, \mathcal{A}_2, P_2)$ the second one. The product space models the two experiments together.

The product probability space models different experiments which have do not influence each other.

We can iterate this construction and get the product space $(\Omega, \mathcal{A}, P)^n$.

Example. If we are throwing dices n times, then this experiment is described by the probability space $(\Omega, \mathcal{A}, P)^n$, if $(\Omega, \mathcal{A}, P)$ describes a single case.

Example. We are throwing a dime n times. The probability of the event that we have k times head is

$$|A| = \left(\begin{array}{c} k \\ n \end{array} \right) = \frac{n!}{k! \cdot (n-k)!} \, .$$

3. Construction: CONDITIONAL PROBABILITY SPACE

Definition. Let $(\Omega, \mathcal{A}, P)$ be a finite probability space and assume $B \in \mathcal{A}$ satisfies $P[B] > 0$. Define the **conditional probability of A given B** by

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

We get a new probability space

$$(B, \mathcal{A} \cap B, P[\cdot|B]) \, ,$$

where $\mathcal{A} \cap A = \{A \cap B \mid B \in \mathcal{A}\}$ and $P[B \mid A] = P[A \cap B]/P[A]$ is the **conditional probabiliy measure** with respect to A.

Interpretation. The probability space $(A, \mathcal{A} \cap A, P[\cdot|A])$ models a new situation, where we know that A occurs.
$P[A|B]$ is the probability that A occurs under the condition that B occurs.

The conditional probability space is the situation, where the experimentalist works with more information and this information excludes certain experiments.

Example. Look again at Dave who has two children and we know that one child is a girl.
We model first the situation without the additional knowledge. Then, $(\Omega, \mathcal{A}, P) = (\{GB, BG, GG, BB\}, \mathcal{A}, P)$ and each of the events $\{GB\}, \{BG\}, \{GG\}, \{BB\}$ has equal probability 1/2. Knowing that one child is a girl restricts the probability space to $A = \{GB, BG, GG\}$ and we compute now the conditional probability that we have one boy $P[\{BG, GB, BB\}|A] = P[\{BG, GB\}]/P[A] = 1/24/3 = 2/3$.

Proposition 1.3.3 *Given a finite probability space $(\Omega, \mathcal{A}, P)$ and $B \in \mathcal{A}$ with $P[B] > 0$, then $(B, \mathcal{A} \cap B, P[\cdot \mid B])$ is a probability space.*

Proof. $\mathcal{A} \cap B$ is a Boolean algebra and $Q[A] := P[A|B]$ is a measure. □

Example. A dice is thrown and the result is known to be an even number. What is the probability that this number is divisible by 3?
$\Omega = \{1, 2, 3, 4, 5, 6\}$, $B = \{2, 4, 6\}$, $A = \{3, 6\}$. Since only one element in B is divisible by 3 and $P[B] = 1/2$, we have

$$P[A|B] = P[A \cap B]/P[B] = 1/6 : 1/2 = 1/3 \, .$$

Example. An urn contains 10 white and 6 red balls. Let A_1 be the event that one obtains a white ball the first time one takes a ball and A_2 the event that one gets a white ball, the second time one takes a ball. We have

$$P[A_2|A_1] = P[A_2 \cap A_1]/P[A_1] = 9/15 \, .$$

In applications, it is often dangerous that when measuring a probability, not hidden conditional probabilities are measured. See in Appostol the example of the Biology department which wants to find out, if there are more boys or girls in the U.S. population.

Even simple problems can be surprisingly puzzling:

Problem. **The three door problem , Monty Hall problem (1991)**

Suppose you're on a game show and you are given a choice of three doors. Behind one door is a car and behind the others are goats. You pick a door-say No. 1 - and the host, who knows what's behind the doors, opens another door-say, No. 3-which has a goat. (In all games, he opens a door to reveal a goat). He then says to you, "Do you want to pick door No. 2?" (In all games he always offers an option to switch). Is it to your advantage to switch your choice?

Answer: Switching the door doubles the chances to win:

The problem was discussed by Marilyn vos Savant in a "Parade" column in 1991 and provoked a big controversy in the next months. Thousends of letters were written. The problem is that intuitive argumentation can easily lead to the conclusion that it does not matter whether to change the door or not. Modeling the problem with a probability space and taking conditional probabilities is a bit confusing. It is simpler to model seperately the two strategies "switching" and "no switching" in two probability spaces and to compute the winning event in both cases. In both cases, the probability space has three elements!

No switching: The probability space consists of three elements

$$\Omega = \{goat, goat, car\} .$$

The event that you win is

$$A = \{car\}$$

You choose a door and win with probability 1/3. The opening of the host does not affect any more your choice.

Switching: you choose the door with the car. You loose since you switch. You choose a door with a goat. The host opens the other door with the goat. You win and there are two cases, where you win. There is one case, where you loose. The probability to win is 2/3. Again, we model the thing with a probability space: $\Omega = \{car, goat, goat\}$. This time, the event that you win is $A = \{goat, goat\}$ and it has probability 2/3.

Proposition 1.3.4 *Given disjoint events A_i with $\bigcup_{i=1}^n A_i = \Omega$. Then*

$$P[A] = \sum_{i, P[A_i]>0} P[A|A_i] \cdot P[A_i] .$$

Proof. Since $A = \bigcup_{i=1}^n A \cap A_i$ and the sets $A \cap A_i$ are disjoint we have

$$P[A] = \sum_{i=1}^n P[A \cap A_i] = \sum_{i=1}^n P[A|A_i]P[A_i] .$$

□

Interpretation. The probabiliy of A can be computed in a case to case study: Compute $P[A|A_i]$ in different cases and add the probabiliies up.

1.4 Independent events

Definition. Two events A and B are called **independent** if $P[A \cap B] = P[A] \cdot P[B]$.

Proposition 1.4.1 *Two events A and B are independent, if and only if either $P[B] = 0$ or $P[A|B] = P[A]$.*

Proof. If $P[B] = 0$, then B is independent of any other event. If $P[B] > 0$ and A and B are independent then $P[A|B] = P[A \cap B]/P[B] = P[A]$. If $P[B] > 0$ and $P[A \cap B]/P[B] = P[A]$, then $P[A \cap B] = P[A] \cdot P[B]$ and A and B are independent. □

Definition. A finite set $\{A_i\}_{i \in I}$ of events is called independent if

$$P[\bigcap_{i \in J} A_i] = \prod_{i \in J} P[A_i]$$

for all $J \subset I$.

Example. Let $\Omega = \{a, b, c, d\}$ and P the uniform probability giving to each of the elements probability $1/4$. The events $A = \{a, d\}, B = \{b, d\}, C = \{c, d\}$ are pairwise independent since they have all probability $1/2$ and their intersection has probability $1/4$. The intersection $A \cap B \cap C$ is $\{d\}$ and has probabiliy $1/4$ while $P[A] \cdot P[B] \cdot P[C] = 1/8$, so the events A, B, C are not independent.

Independent events are very natural when doing an experiment with randomness several times. This situation occurs in gambling.

Remark. Independence can sometimes be overcome in games. For example, in Black Jack, there are people called "counters" who keep track about the partial sum over all played cards. This leads to an advantage of the player since the next experiments are no more independent. In the extreme case, when all cards except one card has been played, the "counter" knows for sure the value of the next card. (See Los Angeles magazene). This advantage is so big that one can live with that 30000$ per year. But one has to do it without been caught!

Definition. Let $(\Delta, \mathcal{F}, Q)$ be a probability space. For every n, we can form the **product space** $\Omega, \mathcal{A}, P)$ defined as follows: Ω is the set of n tupels with entries in Δ, $\mathcal{A}$ is the smallest Boolean algebra which contains all the sets $(A_1, A_2, \dots, A_n)$ with $A_i \in \mathcal{F}$. The probability measure P is determined if we define $P[(A_1, A_2, \dots, A_n)] = \prod_{i=1}^n Q[A_i]$.

An important example is the case when $\Delta = \{0, 1\}$ and $Q[\{0\}] = q, Q[\{1\}] = p$. The event $A = \{\omega \in \Omega \mid \sum_i \omega_i = k\}$ has probability

$$P[A] = \binom{n}{k} p^k q^{n-k} .$$

This is called **Bernoulli's formula.**

Probabilistic interpretation. If $(\Delta, \mathcal{F}, Q)$ describes a single experiment (like throwing a dice) then $(\Omega, \mathcal{A}, P)$ describes n independent experiments.

Proposition 1.4.2 *Let $(\Omega, \mathcal{A}, P)$ is the n-fold product space of a probability space $(\Delta, \mathcal{B}, Q)$. Let B_i be events, which depend only on the i-th outcome:*

$$A_i = \{\omega = (\omega_1, \dots, \omega_n) \mid \omega_i \in B_i \subset \Delta\}$$

Then B_i are all independent.

Proof. We have by definition of the probability measure P that $P[B_i] = Q[A_i]$ and

$$P[\bigcap_{i=1}^{k} B_{j_i}] = \prod_{i=1}^{k} Q[A_{j_i}] = \prod_{i=1}^{k} P[B_{j_i}] .$$

□

Together with a problem in the homework, we see that chaning $\mathcal{A}$, taking the product space or taking the conditional probability space are all leaving invariant independent events.

Appendix: Lotto. Geometric filling out of the lotto matrix. Does this decrease the probability of winning? What happens, if you use the date of the day as your guess?

1.5 Random variables

Definition. Let $(\Omega, \mathcal{A}, P)$ be a probability space. A function $X : \Omega \to \mathbf{R}$ is called a **random variable**, if for all values $a \in \mathbf{R}$, we have $X^{-1}(\{a\}) \in \mathcal{A}$. One says, that X is measureable.

Example. If $\mathcal{A}$ is the set of subsets of Ω, then every function $X : \Omega \to \mathbf{R}$ is a random variable.

Example. Given a set of functions $X_1, X_2, \dots, X_n$ on a set Ω. It is natural to consider the measurable space $(X, \mathcal{A})$, where $\mathcal{A}$ is the smallest algebra such that X_i are measureable.

Example. We are throwing dices 2 times. We have already seen that this probability space has 36 elements. Every experiment is given by a pair $\omega = (\omega_1, \omega_2)$. Look at the random variables $X(\omega) = \omega_1$ or $Y(\omega) = \omega_2$.

Proposition 1.5.1 *The set $\mathcal{L}$ of random variables on the finite probability space $(\Omega, \mathcal{A}, P)$. form an algebra: This means a) X, Y are random variables, then $X + Y$ is a random variables.*
b) X is a random variable and $\lambda \in \mathbf{R}$, then λX is a random variable.
c) X, Y are random variables, then XY is a random variable.
d) Given a function $f : \mathbf{R} \to \mathbf{R}$, then $f \circ X$ is a random variable.

Interpretation. Random variables are outcomes of measurements. For every experiment ω, we have the outcome $X(\omega)$. If $(\Omega, \mathcal{A}, P)$ models an experiment a random variable X is a measurement.

1.6 Some more examples

We do not address the question, how big is the probability that there is life on an other planet, or if the time for evolution of human beeings was enough or if it needed a kick from outside. Such questions are subtle but involve a lot of probability. The main problem with such questions is that the set-up is not easy at all.

1) **The random number generator with filter (see Week 3 homework.**

```
Generator:=Random[Integer,{1,50}]
Filtered:=Module[{n=Generator},While[Mod[n,3]==0 !! Mod[n,2]==0, n=Generator];n];
Experiment[n_] :=Table[Filtered,{n}];
CountPrimes[s_]:=Module[{k=0},Do[If[PrimeQ[s[[i]]],k++],{i,Length[s]}];k];
CountDivisible[s_]:=Module[{k=0},
  Do[If[Mod[n,3]==0 || Mod[n,2]==0,k++],{n,Length[s]}];k];
f[n_]:=CountPrimes[Experiment[n]]/n;
NumberOfPrimes=CountPrimes[Table[n,{n,50}]];
NumberOfDivisible=CountDivisible[Table[n,{n,50}]]
```

What is the average outcome of the random generator and the random generator with filter? The original generator has in average a value $51/2$ (Gauss). There are 15 prime numbers and the generator gives a prime number with probability $3/10$. The tuned generator gives the numbers $1,5,7,11,13,17,19,23,25,29,31,35,37,41,43,47,49$ randomly and there are 13 prime numbers. The tuned operator gives a prime number with probability $13/17$. In average, it gives a value $253/17$ which is around 15.

2) **Balls in a box.** Assume we have a box with two rooms of size p and q with $p+q=1$. Assume we have n particles are in this box modelling a gas. We assume the particles move around in a random independent way so that a particle is with probability p in the left box and with probability q in the right box. What is the probability that we have k particles in one box and n particles in the other box. We have computed the probability already. $Omega = \{left, right\}^n$. Bernoulli's formula gave the right probability:

$$B_k(p,q) = \binom{n}{k} p^k q^{n-k} .$$

How many particles do we expect to be in the left room?
This is a nice calculation: Since the function $B_k(p,q)$ gives summed up $(p+q)^n = 1$. we differentiate the sum with respect to p, which gives n and with the other formula $\sum_k k \binom{n}{k} p^{k-1} q^{n-k}$ so that we compute the average np.

3) **The random walker in the plane.** We look at the one-dimensional model in detail later on.

The problem. Assume a walker walks on the chessboard starting at the middle and moves at each step with probability p_r, p_l, p_u, p_d into one of the four directions.

a) What is the probabilty that it is after 6 steps again at the same place?

b) How far is he expected to be in 6 steps in the x direction, how far does he expect to be in 6 steps?

Let us do it first with the computer:

The probability space: $\Omega = \{(l,l,l,l),(l,l,l,u),(l,l,l,r),(l,l,l,d),...\}$ has $4*4*4*4*4 = 4096$ elements.
The random variabes $X(\omega) = x$ coordinate after 6 steps". $Y(\omega) = y$ coordinate after 6 steps.

```
Omega=Partition[Flatten[Table[{a1,a2,a3,a4,a5,a6},
                    {a1,4},{a2,4},{a3,4},{a4,4},{a5,4},{a6,4}]],6];
Travel[omega_]:=Module[{s={0,0}},
  Do[If[omega[[i]]==1,s=s+{ 1,0},
     If[omega[[i]]==2,s=s+{ 0,1},
     If[omega[[i]]==3,s=s+{-1,0},
     If[omega[[i]]==4,s=s+{0,-1}]]]],{i,Length[omega]}];s];
Distance[{a_,b_}]:=N[Sqrt[a^2+b^2]];
Distances=Map[Distance,Travels];
Travels=Table[Travel[Omega[[i]]],{i,Length[Omega]}];
Count[Travels,{0,0}];
ProbabilityOfReturn=Count[Travels,{0,0}]/Length[Omega]
AverageTravel=Sum[Travels[[i]],{i,Length[Travels]}]/Length[Omega]
```

Let us see if we can do it with the brain better.

How many possibilities do we have to make 6 steps and to return. We must make an even number of steps in each direction.
1) In x direction only: we must move 3 times to the right and 3 times to the left. There are

$$\binom{6}{3} = 20$$

possibilities.
2) In x direction 2 steps and in y direction 4 steps.

$$lrudud, lruudd, lrduud, lrdduu, lrdudu, lruddu$$

So, 6 times

$$2 * \binom{6}{2} = 180$$

possibilities.
3) In y direction 2 steps and in x direction 4 steps. again 180.
4) In y direction only 20.
Total: 400 possibilities.

The average walk is easy to discribe. Since the event that the walker ends at the point (k, l) is the same then the probability that the walker ends at the point $(-k, -l)$. We get

$$E[X] = 0, E[Y] = 0 .$$

Let us now illustrate the result in a fancy way:

```
<<Graphics'Graphics'
Stat={Count[Travels,{0,0}],Count[Travels,{0,2}],
        Count[Travels,{0,4}],Count[Travels,{0,6}]};
Stat={400, 225, 36, 1};
PieChart[Stat];
BarChart[Stat];
```

```
Stat={400., 225., 36., 1.};
MusicChart[data_]:=
  Play[Sum_{i=1}^{Length[Stat]} Stat[[i]] Sin[Stat[[i]]*10*t], {t,0,10}];
```

Here is the random walk in two and three dimensions:

```
Ways={{1.,0.},{0.,1.},{-1.,0.},{0.,-1.}};
T[{x_,y_}]:={x,y}+Ways[[Random[Integer,{1,4}]]];
Show[Graphics[Line[NestList[T,{0.,0.},1000]]],
  Frame->True,FrameTicks->None,AspectRatio->1]

Ways={{1,0,0},{0,1,0},{0,0,1},{-1,0,0},{0,-1,0},{0,0,-1}};
T[{x_,y_,z_}]:={x,y,z}+Ways[[Random[Integer,{1,6}]]];
Show[Graphics3D[Line[NestList[T,{0.,0.,0.},1000]]],
  Boxed->True, AspectRatio->1]
```

Here is a film, how the random walker walks:

```
Ways={{1.,0.},{0.,1.},{-1.,0.},{0.,-1.}};
T[{x_,y_}]:={x,y}+Ways[[Random[Integer,{1,4}]]];
Module[{s={},v={0.,0.},w}, Do[w=T[v]; s=Append[s,Line[{v,w}]]; v=w;
  Show[Graphics[s],Frame->False,
      PlotRange->{{-20,20},{-20,20}}],{100}]];
```

1.7 Expectation

Definition. Let X be a random variable on the probability space $(\Omega, \mathcal{A}, P)$. If $\mathcal{A}$ is the set of subsets of Ω, then the **expectation** of X is defined as

$$E[X] = \sum_{\omega \in \Omega} P[\{\omega\}] X(\omega) .$$

In general, we define

$$E[X] = \sum_{a, X(\omega)=a} a \; P[\{\omega | X(\omega) = a\}] .$$

Notice that this is defined since $X^{-1}(a) = \{\omega | X(\omega) = a\} \in \mathcal{A}$.

Probabilistic interpretation. As the name says, the expectation of X is the expected outcome of the measurement. In a game, the expectation value of the winning gives your average win.

Proposition 1.7.1 (Properties expectation) *Let X, Y be random variables on $(\Omega, \mathcal{A}, P)$ and let λ, μ be real numbers.*
a) $E[X + Y] = E[X] + E[Y]$
b) $E[\lambda X] = \lambda E[X]$.
c) $X \leq Y \Rightarrow E[X] \leq E[Y]$.
d) $E[X^2] = 0 \Leftrightarrow X = 0$.
e) $E[X - E[X]] = 0$.

Proof. Obvious. □

1.8 Variance, Covariance, Correlation

Definition. Let X be a random variable. The **variance** of X is defined as

$$\mathrm{Var}[X] = E[(X - E[X])^2] .$$

The square $\sigma[X] = \sqrt{\mathrm{Var}[X]}$ is called the **standard deviation** of X. If X and Y are random variables, their **covariance** is defined as

$$\mathrm{Cov}[X, Y] = \mathrm{E}[(X - E[X])(Y - \mathrm{E}[Y])] .$$

Example. Let $\Omega = \{1, 2, 3, 4, 5, 6\}$ be the possible outcomes of throwing a dice, where $P[\{i\}] = 1/6$ have all the same probability. The random variable $X[\omega] = \omega$ has the expectation

$$\mathrm{E}[X] = \sum_{i=1}^{6} i/6 = 21/6 = 7/2$$

and the variance

$$\mathrm{Var}[X] = \sum_{i=1}^{6} \frac{1}{6}(i - 7/2)^2 = 35/12 .$$

The standard deviation is 1.70783. This is the expected error we make, when guessing the value 7/2.

Interpretation. The variance is the quadratic mean error and the standard deviation is should be thought of as the typical deviation from the mean.

The **covariance** of two random variables X, Y is defined as

$$\mathrm{Cov}[X, Y] = \mathrm{E}[(X - \mathrm{E}[X]) \cdot (Y - \mathrm{E}[Y])] .$$

Interpretation. The covariance is the scalar product between the centered random variables of E and Y.

Proposition 1.8.1 $\mathrm{Var}[X] \geq 0$.
$\mathrm{Var}[X] = \mathrm{E}[X^2] - \mathrm{E}[X]^2$.
$\mathrm{Cov}[X, Y] = \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y]$.

Proof. a) obvious.
b) Use the properties of expectation: lel $m = E[X]$.

$$\mathrm{Var}[X] = \mathrm{E}[(X - m)^2] = \mathrm{E}[X^2 - 2mX + m^2] = \mathrm{E}[X^2] - 2m\mathrm{E}[X] + m^2 = \mathrm{E}[X^2] - m^2 .$$

c) Let $m = E[X]$ and $n = E[Y]$. We have

$$\mathrm{Cov}[X, Y] = \mathrm{E}[(X - m)(Y - n)] = \mathrm{E}[XY] - m\mathrm{E}[Y] - n\mathrm{E}[X] + mn = \mathrm{E}[XY] - mn$$

□

Example. Consider the probability space $(\Omega = \{0, 1\}^2, \mathcal{A}, P)$ where $P[\{1\} \times \{0, 1\}] = p = P[\{0, 1\} \times \{1\}]$ ("throwing two dimes"). Given the random variables $X(\omega) = \omega_1$ and $Y(\omega) = \omega_2$. We compute

$$\mathrm{Cov}[X, Y] = E[XY] - E[X] \cdot E[Y] = p^2 - p^2 = 0 .$$

Definition: Given two random variables X, Y with $\mathrm{Var}[X] > 0, \mathrm{Var}[Y] > 0$, the **correlation** between X and Y is defined as

$$\mathrm{Corr}[X, Y] = \frac{\mathrm{Cov}[X, Y]}{\sigma[X], \sigma[Y]} .$$

If $\mathrm{Corr}[X, Y] = 0$, then X, Y are called **uncorrelated.**

Interpretation: This means that the centered random variables are orthogonal to each other if looked at as vectors. If the correlation is 1 or -1, there exists a linear relation between X and Y.

Proposition 1.8.2 *a)* $-1 \leq \text{Corr}[X,Y] \leq 1$.
b) $\text{Corr}[X,Y] = 1$ *if* $X - \text{E}[X] = Y - \text{E}[Y]$ *and* $\text{Corr}[X,Y] = -1$ *if* $X - \text{E}[X] = -(Y - \text{E}[Y])$.

Proof. Recall the Schwarz inequality for vectors in $\mathbf{R}^n$:

$$|(x,y)|^2 \leq (x,x) \cdot (y,y)$$

where $x, y \in \mathbf{R}^n$. Proof: (recently appeared in Mathematics Magazine)

$$\frac{\sum_k x_k y_k}{(\sum_k x_k^2)^{1/2}(\sum_k y_k^2)^{1/2}} = 1 - \frac{1}{2}\sum_k \left(\frac{x_k}{(\sum_k x_k^2)^{1/2}} - \frac{y_k}{(\sum_k y_k^2)^{1/2}} \right)^2 .$$

Apply this inequality to $x = X - E[X]$ and $y = Y - E[Y]$.
If $\text{Corr}[X,Y] = 1$, then the two vectors $X - E[X]$ and $Y - E[Y]$ are parallel. □

1.9 More examples about variance

Example 1. Given n **particles in a box.** We have seen that in average

$$\text{E}[X] = \sum_k k \binom{n}{k} p^{k-1} q^{n-k} = pn$$

particles are in the left box. X was the random variable counting the number of particles in the left box.
With a similar differentiation trick like before we get

$$p^2 n(n-1) = \sum_k k(k-1) \binom{n}{k} p^{k-1} q^{n-k}$$

so that $E[X^2] - E[X] = p^2 n(n-1)$. We get therefore

$$\text{E}[X^2] = p^2 n(n-1) - np$$

and

$$\text{Var}[X] = E[X^2] - E[X]^2 = np(1-p) = npq .$$

Example 2. **roulette.**
Let $\Omega = \{0, 1, 2, \ldots, 36\}$ each with the same probability. Consider the three random variables

$$X_1(\omega) = \begin{cases} 1 & \omega \text{ even not } 0 \\ -1 & \omega \text{ odd} \end{cases}$$

$$X_2(\omega) = \begin{cases} 35 & \omega = 3 \\ -1 & \text{else} \end{cases}$$

$$X_3(\omega) = \begin{cases} 2 & 1 < \omega \leq 12 \\ -1 & \text{else} \end{cases}$$

We have

$$\begin{aligned}
\mathrm{P}[X_1 = 1] = 18/37 \quad &, \quad \mathrm{P}[X_1 = -1] = 19/37 \\
\mathrm{P}[X_2 = 35] = 1/37 \quad &, \quad \mathrm{P}[X_2 = -1] = 36/37 \\
\mathrm{P}[X_3 = 3] = 12/37 \quad &, \quad \mathrm{P}[X_3 = -1] = 25/37 .
\end{aligned}$$

We compute $\mathrm{E}[X_i] = -1/37$ and

$$\mathrm{Var}[X_1] = 1^2\mathrm{P}[X_1 = 1] + (-1)^2\mathrm{P}[X_1 = -1] - 1/37^2 ,$$

$$\mathrm{Var}[X_2] = (35^2\mathrm{P}[X_2 = 1] + (-1)^2\mathrm{P}[X_2 = -1] - 1/37^2 .$$

1.10 Correlation line

Example. (Linear regression).
Given two random variables X, Y for which we know X, the expectation of Y and the covariance of X, Y. In order to find the best prediction for Y, we want to find the random variable $\tilde{Y} = aX + b$, which minimizes $\mathrm{Var}[Y - \tilde{Y}]$ and satisfies $E[Y] = E[\tilde{Y}]$.

Proposition 1.10.1 *Given random variables X, Y, for which we know the expectation, the covariance and the variance of X.*
Among all the random variables $Y_{a,b} = aX + b$ the random variable $\tilde{Y}$ with

$$a = \frac{\mathrm{Cov}[X,Y]}{\mathrm{Var}[X]}, b = E[Y] - aE[X]$$

if $\mathrm{Var}[X] > 0$ and $\tilde{Y} = E[Y]$ if $\mathrm{Var}[X] = 0$. minimizes $\mathrm{Var}[Y - \tilde{Y}]$ under the constraint $E[Y] = E[\tilde{Y}]$.

Proof. From $E[\tilde{Y}] = aE[X] + b = E[Y]$, we get $b = E[Y] - aE[X]$. Minimize $\mathrm{Var}[Y - aX] = \mathrm{Var}[Y] + 2a\mathrm{Cov}[Y, X] - a^2\mathrm{Var}[X]$ gives the solution of a. □

Definition. The line $y = ax + b$ is called the **regression line.**

Interpretation. Assume, we can measure one of the random variables X, Y and the covariance of them, the random variable $aX + b$ is the best guess for Y. If the two random variables are uncorrelated, then $a = 0$ and $\tilde{Y} = E[Y]$ is the best guess for the random variable Y.

Empirical variance and standard deviation.
There are statistical reasons, why one also considers the empirical variance

$$\tilde{E}[(X - E[X])^2]$$

where $\tilde{E}[X] = n/(n-1)E[X]$ for $n = |\Omega|$.

Standardisation.
Given a random variable X, we denote with $X^* = \frac{X - E[X]}{\sigma[X]}$ the standardisation of X. This random variable has mean zero and variance 1.

Example. For a Bernoulli distributed random variable, we get

$$X_n^* = (X - pn)/\sqrt{np(1-p)} .$$

As we will see later, the distribution of X_n^* converges to the Normal distribution. This is a special case of the central limit theorem and calld the DeMoivre-Laplace global limit theorem.

1.11 Independent random variables

Definition. Two random variables X, Y are called **independent**, if for all $a, b \in \mathbf{R}$, the events $A = \{X = a\}$ and $B = \{Y = b\}$ are independent.
More generally, a set of random variables X_i are **independent** if for all $a_i \subset \mathbf{R}$, the sets $\{X_i = a_i\}$ are independent events.

Remark. It follows that if X, Y are independent, then $P[X \in a, Y \in b] = P[X \in a] \cdot P[Y \in b]$.

Proposition 1.11.1 *If X and Y are independent random variables, then $E[X \cdot Y] = E[X] \cdot E[Y]$.*
More generally, if X_i is a set of independent random variables, then

$$E[\prod_{i=1}^{n} X_i] = \prod_{i=1}^{n} E[X_i] \,.$$

Proof.

$$\begin{aligned} E[XY] &= \sum_{a \in X(\Omega), b \in Y(\omega)} abP[X(\omega) = a, Y(\omega) = b] \\ &= \sum_{a \in X(\Omega), b \in Y(\omega)} abP[X(\omega) = a]P[Y(\omega) = b] \\ &= \sum_{a \in X(\Omega)} aP[X(\omega) = a] \sum_{b \in Y(\omega)} bP[Y(\omega) = b] \\ &= E[X] \cdot E[Y] \end{aligned}$$

The general claim goes by induction. □

Example. Let $(\Omega, \mathcal{A}, P)$ be the product space of two probability spaces. Define

$$X((\omega_1, \omega_2)) = \omega_1, Y((\omega_1, \omega_2)) = \omega_2 \,.$$

Then X and Y are independent.
Remark. This example is a prototype and one can show that if we have given two random variables which are independent, one can always weeken the algebra $\mathcal{A}$ (and rename the atoms to elements in Ω) so that the probability space is a product space.

Proposition 1.11.2 *If X, Y are independent and f, g are measureable functions, then $f(X)$ and $g(Y)$ are independent.*

Proof. $\mathrm{P}[X = a] = \mathrm{P}[f(X) = f(a)]$, $\mathrm{P}[Y = b] = \mathrm{P}[g(Y) = g(b)]$. Since $\mathrm{P}[X = a, Y = b] = \mathrm{P}[X = a]\mathrm{P}[Y = b]$, we have also $\mathrm{P}[f(X) = f(a), g(Y) = g(b)] = \mathrm{P}[f(X) = f(a)]\mathrm{P}[g(Y) = g(b)]$. □

Proposition 1.11.3 *Two independent random variables X, Y are uncorrelated.*

Proof. Since X, Y are independent, also $X - E[X]$ and $X - E[Y]$ are independent and so $\mathrm{Cov}[X,Y] = \mathrm{E}[(X - E[X])]\mathrm{E}[((Y - E[Y])] = 0$. □

It follows that if $\mathrm{Cov}[X,Y] \neq 0$, then X, Y are dependent. The converse is however not true:

Example: Given a nonconstant random variable U such that $\mathrm{E}[U^3] = \mathrm{E}[U^5] = 0$. Define $X = U^3$ and $Y = U^2$. They are uncorrelated since

$$E[U^3(U^2 - E[U^2])] = E[U^5] - E[U^2]E[U^3] = 0\,.$$

or better $E[XY] - E[X]E[Y] = 0$. The variables X, Y are however not independent. Let a be a value attained by U. Then

$$\mathrm{P}[X = a^2, Y = a^3] = \mathrm{P}[U^2 = a^2, U^3 = a^3] = \mathrm{P}[U = a] \notin \{0,1\}$$

and

$$\mathrm{P}[U^2 = a^2]\mathrm{P}[U^3 = a^3] = \mathrm{P}[U = a]^2 \neq \mathrm{P}[U = a]$$

so that X, Y are not independent.

Chapter 2

General probability spaces

2.1 Cardinality of sets

Definition. A set is called **countable** if we can enumerate its elements. In other words, a set S is countable if there exists a bijection between N and S or if S is finite. A set is called **uncountable** if it is not countable.

Lemma 2.1.1 (Cantor) *If S is countable, then the set $\mathcal{S}$ of its subsets is uncountable.*

Proof. Assume $\mathcal{P}$ is countable. By definition there exists a bijection f between $S \to \mathcal{S}$ so that $f(s) \in \mathcal{S}$ for all $s \in S$. Define

$$D = \{s \in S \mid s \notin f(s)\} .$$

Clearly $D \in \mathcal{S}$ and there exists $t \in S$ with $f(t) = D$. There are two cases: $t \in D$ and $b \notin D$. In the first case $t \notin D$ and in the second case $t \in D$. Both situations are not possible. The contradiction shows that the existence of f is not possible. □

Examples.

- The set of finite subset of a countable set is countable.
- The set of subsets of a countable set is uncountable as we have just seen.
- A subset of a countable set is countable.
- The union or intersection of countable sets is countable.
- The cartesian product of countable sets is countable.
- The set of rational numbers is countable.
- The set of real numbers in $[0, 1]$ is uncountable.

There exists a direct proof of this fact:
Assume, we can number the real numbers. Put them one after each other

1	.123342344343444323
2	.123428342834832348
3	.234823484328384832
4	.029394823482934842

5	.1727347374737730340
6	.1233423443434444993
7	.9823443344234434444
8	.4434455454333345255

Now, we build a new number which differs from the n-th number at the n-th place.

Appendix: Bertrand's paradox

Colloquial language is not precice enough to treat probabilistic problems. Paradoxons appear, when the definition of objects allows different interpretations which leads then to funny or wrong conclusions. We give three famous problems. These three examples should serve as a motivation to introduce probability theory on a precice footing.

Bertrand's paradox (Bertrand 1889)
Throw at random straight lines onto the unit disc. What is the probability that the straight line intersects the disc with a length $\geq \sqrt{3}$, the length of the equilateral triangle inscribed in the circle?

First answer: take an arbitrary point P in the disc and look at the set of all lines passing through that point parametrized by an angle ϕ. In order that the chord be longer than $\sqrt{3}$, the line has to lie within an angle of 60° out of 180°. The probability is thus 1/3.

Second answer: take all lines perpendicular to a fixed diameter. The chord is longer than $\sqrt{3}$, when the point of intersection lies on the middle half of the diameter. The probability is thus 1/2.

Third answer: look at the midpoints of the chords. If they lie in the disc of radius 1/2 which has area 1/4 of the whole disc, the chord is longer than $\sqrt{3}$. The probability is thus 1/4.

2.2 General probability spaces

Definition. Let Ω be any set. A set $\mathcal{A} \subset \mathcal{P}(S)$ is called a σ-algebra, if it is a Boolean algebra and if A_i is a sequence in $\mathcal{A}$, then $\bigcup_i A_i \in \mathcal{A}$. in other words, we require

(i) $\Omega \in \mathcal{A}$,
(ii) $A \in \mathcal{A} \rightarrow A^c \in \mathcal{A}$,
(iii) $A_i \in \mathcal{A} \Rightarrow \bigcup_i A_i \in \mathcal{A}$

The pair $(\Omega, \mathcal{A})$ is called a measurable space.

Example. Let Ω be any set and let $\mathcal{A}$ be the set of subsets of Ω. Then $\mathcal{A}$ is a σ-algebra.

As in the case of finite probability spaces, we conclude that a σ-algebra is closed under all countable operations.

Definition. Let $\mathcal{A}$ be a σ-algebra over a set Ω. A function $P : \mathcal{A} \to \mathbf{R}$ is called **countably additive** if

$$P[\bigcup_i A_i] = \sum_i P[A_i]$$

for all disjoint sets $A_i \in \mathcal{A}$.

Definition. A countably additive function P on a σ-algebra $\mathcal{A}$ is called a **measure**, if $P[A] \geq 0$ for all $A \in \mathcal{A}$. A measure is called a **probability measure** if it is a measure and $P[\Omega] = 1$.

Definition. $(\Omega, \mathcal{A}, P)$ is called a **probability space** if P is a probability measure on the measurable space $(\Omega, \mathcal{A})$.

Examples.
1) Let $\Omega = \mathbf{N}$ and $\mathcal{A}$ be the set of subsets of Ω. For every $\lambda > 0$

$$P[A] = \sum_{n \in \Omega} e^{-\lambda} \frac{\lambda^n}{n!}$$

defines a probability measure on the measurable space $(\Omega, \mathcal{A})$.
2) For every $s > 1$

$$P[A] = \sum_{n \in \Omega} \zeta(s)^{-1} n^{-s}$$

is a probability measure on the measurable space $(\Omega, \mathcal{A})$. The function $s \mapsto \zeta(s) = \sum_{n \in \Omega} n^{-s}$ is the Riemann zeta function).
3) The measure

$$P[A] = \sum_{n \in \Omega} 2^{-n+1}$$

is a probabiliy measure. (see exercice).

Definition. A function $d : \Omega \times \Omega \to \mathbf{R}$ is called a **metric** if

(i) $d(x,y) \geq 0$, (ii) $d(x,y) + d(y,z) \leq d(x,z)$ (iii) $d(x,y) = 0 \Leftrightarrow x = y$

The pair (Ω, d) where Ω is a set and d is a metric is called a **metric space.**

Examples.
1) $\Omega = \mathbf{R}, d(x,y) = |x - y|$.
2) $\Omega = \mathbf{R}^2, d(x,y) = |x_1 - y_1| + |x_2 - y_2|$.
3) Ω arbitrary $d(x,y) = 1, x \neq y$ and $d(x,x) = 0$.
4) $\Omega = \{z \in \mathbf{C} \mid |z| = 1\}$, $d(x,y) = |\sin(x - y)|$.

Definition. Let (Ω, d) be a metric space. A subset $A \subset \Omega$ is **open**, if for every point $x \in A$ there exists $r > 0$ such that the ball $\{y \in \Omega \mid |d(x,y) < r\}$ is contained in A.

Lemma 2.2.1 *Let $\{\mathcal{A}_i\}_{i \in I}$ be a collection of σ-algebras in Ω. Then $\bigcap_{i \in I} \mathcal{A}_i$ is a σ-algebra.*

Proof. Check properties (i), (ii), (iii):
If Ω is in all algebras $\mathcal{A}_i$, then it is in the intersection. If A is in all algebras $\mathcal{A}_i$, then A^c is in all algebras and so in the intersection. If $A_1, A_2 \dots,$ is a sequence of sets which are all in $\mathcal{A}_i$ then their intersection is in all $\mathcal{A}_i$ and so in the set. □

Corollary 2.2.2 *Let $\mathcal{S}$ be a set of subsets of Ω. Then there exists a smallest σ-algebra $\mathcal{A}$, which contains $\mathcal{S}$. One says $\mathcal{S}$* **generates** *$\mathcal{A}$.*

Proof. Take the collection of all σ-algebras, which contain $\mathcal{S}$. This set is not empty and its intersection $\mathcal{A}$ has the required properties. □

Definition. Let (Ω, d) be a metric space and $\mathcal{S}$ be the set of open sets in Ω. The σ-algebra generated by $\mathcal{S}$ is called the **Borel σ-algebra.** Together with a probability measure P on $(\Omega, \mathcal{A})$, one gets a Borel probability measure.

Definition. Let $(\Omega, \mathcal{A}, P)$ be a probability space. A point ω satisfying $P[\{\omega\}] > 0$ is called an **atom.**

Lemma 2.2.3 *A probability space $(\Omega, \mathcal{A}, P)$ has only countably many atoms.*

Proof. The set $T_n = \{(n+1)^{-1} < P^{-1}(\omega) \leq n^{-1}\}$ contains maximal n elements. The union $T = \bigcup_n T_n$ is countable and every atom has to be in T. □

Corollary 2.2.4 *A probability space, for which every element is an atom is countable.*

Proof. We have seen that an uncountable sum of positive numbers is infinite. □

We come now to the construction of new probability spaces. There is not much to say here since the definition is the same as in the finite case.

1) Restriction of $\mathcal{B}$.
The same definition. 2) Product probability space.
Note here, that the product σ algebra is the smallest σ algebra which contains all sets $A \times B$, where $A \in \mathcal{A}_1$ and $B \in \mathcal{A}_2$.
3) Conditional probability space.

The same definition.

What is new now, that we can also define an **infinite product space**: (we restrict ourselves to countable products). Let $(\Omega_i, \mathcal{A}_i, P_i)$ be probability spaces. Then the product space Ω consists of all sequences $i \mapsto \omega_i$ and $\mathcal{A}$ is the smallest σ algebra, which contains all the sets of the form $(\dots, \Omega, \Omega, A_n, A_{n+1}, \dots, A_m, \Omega, \Omega \dots)$, with $A_i \in \mathcal{A}_i$.

2.3 Appendix: Banach-Tarski's paradox

Denote with $\mathbf{R}^d$ the d dimensional space and with $\mathcal{A}$ the set of subsets of $\mathbf{R}^d$. Banach has proven in 1923 that in dimensions $d \leq 2$, there exists a function $P : \mathcal{A} \to [0, \infty]$ such that $P[\emptyset] = 0$ and $P[\bigcup_{i=1}^n A_i] = \sum_{i=1}^n P[A_i]$ if A_i are disjoint sets in $\mathbf{R}^d$ and such that $P[R(A)] = P[A]$ for any rotation or translation R and $P[\prod_{i=1}^n [a_i, b_i]] = \prod_i (b_i - a_i)$.

In dimensions $d \geq 3$ however, there exists the following situation: the unit ball B can be written as a union of 5 disjoint sets A_i, such that after suitable rotations and translations, the sets A_1, A_2, A_3 can be put together to a unitball and the sets A_4 and A_5 can be put together to a unitball.

It is also known that there exists no countable additive translational invariant measure P on the real line equiped with the σ-algebra of all subsets of $\mathbf{R}$:

An example in one dimensions is due to Vitali.

If one skips one of the basic axioms in mathematics, the axiom of choice, then one could add the axiom that all subsets of $\mathbf{R}$ are measureable.

2.4 Discrete Random variables, expectation

Definition: A probability space is called **discrete** if Ω is countable. In this case, many things are very similar to the case of finite probability spaces.

Doing a similar simplification as in the finite case, one can assume without loss of generality that $\Omega = \mathbf{N}$ and that $\mathcal{A}$ is the set of subsets of $\mathbf{N}$.

Remark. The σ-algebra is uncountable or finite. In the later case, it is a Boolean algebra.
Proof. Assume the $\sigma-$ algebra is infinite. Then there exist infinitely many disjoint sets in $\mathcal{A}$. Take a sequence $A_1, A_2, \dots$ of such sets. Then for any subset S of $\mathbf{N}$, we have $A_S = \bigcup_{i \in S} A_i \in \mathcal{A}$ by definition. and for different sets $S, T \subset \mathbf{N}$, we get different sets A_S, A_T.
So, when dealing with infinite probability spaces, we have to deal with uncountable σ-algebras.

Discrete probability spaces are used to model situations, where the set of outcomes is countable.

Definition. A **random variable** is a function $X : \Omega \to \mathbf{R}$ such that $X^{-1}(a) \in \mathcal{A}$ for all $a \in \mathbf{R}$.

Remark. A random variable on a discrete probability space takes only countably many (or finitely many) values.

Definition. The **expectation** of a random variable X is defined as

$$E[X] = \sum_{a \in X(\Omega)} aP[X = a] .$$

This is in general an infinite sum and is therefore not always defined. See the Petersburg paradox!

Definition. We call $\mathcal{L}^1$ the set of random variables, for which the sum $E[X]$ converges. These random variables are called **integrable.** Nonintegrable random variables do seldom occur.

Example. (Poisson measure) Let $\Omega = \{0, 1, 2, 3, 4, \dots\}$ and $\mathcal{A} = \{A \subset \Omega\}$ and

$$P[\{\omega\}] = e^{-\lambda} \frac{\lambda^\omega}{\omega!} ,$$

where λ is a parameter. Consider the random variable

$$X(\omega) = \omega .$$

This random variable discribes the number of calles in a telephone center or the number of electrons emmitted from a cathode. In some sense, it is the most natural distribution on an infinite set and random variables, which have values in **N** and which models a random process has this distribution. We compute the expectation of this random variable:

$$\mathrm{E}[X] = \sum_{k=0}^{\infty} kP[X = k] = \sum_{k=0}^{\infty} ke^{-\lambda} \frac{\lambda^k}{k!} = \lambda .$$

The Poisson measure is a good approximation for large Bernoulli n. Assume $np_n = \lambda$. Then

$$\begin{aligned} \mathrm{P}[S_N = k] &= B(N,k) \frac{\lambda^k}{N^k} (1 - \frac{\lambda}{N})^{N-k} \\ &= \frac{\lambda^k}{k!} \frac{N(N-1) \cdot (N-k+1)}{N^k} \cdot (1 - \frac{\lambda}{N})^{N-k} \\ &\to \frac{\lambda^k}{k!} e^{-\lambda} . \end{aligned}$$

Example. (Geometric measure) Let $\Omega = \{1, 2, 3, \dots\}$ and $P[\{k\}] = (1-p)^{k-1}p$. This is the expected waiting time for success of independent p distributed random variables $P[X_1 = 0, \dots X_{k-1} = 0, X_k = 1]$. How long do we expect to have to throw a coin in order to get head? The expected value of $X(\omega) = \omega$ is

$$\mathrm{E}[X] = \sum_{k=1}^{\infty} k(1 - p^{k-1})p = 1/p .$$

Example. (ζ function measure)
Let $\Omega = \{1, 2, 3, 4, \dots\}$ and

$$P[A] = \sum_{n \in A} \zeta(s)^{-1} n^{-s} .$$

Let us compute the value of $E[X]$ for $X(\omega) = \omega$.

$$E[X] = \sum_{n} \zeta(s)^{-1} n^{-s+1} = \zeta(s+1)/\zeta(s) .$$

Plot3D[Zeta[x+I y+1.00000001]/Zeta[x+I y],x,-5,5,y,-5,5] This probability space is actually useful for an other reason. Take the events

$$A_p = \{n \in \mathbf{N} \mid p \text{divides } n\}$$

Take 2 different primes p, q. Then $A_p \cap A_q = \{n\ pq\text{divides } n\}$. Let us compute the probabilities of A_p and A_q and A_{pq}. We have

$$P[A_p] = \zeta(s)^{-1} \sum_n (np)^{-s} = p^{-s}$$

$$P[A_q] = \zeta(s)^{-1} \sum_n (np)^{-s} = q^{-s}$$

$$P[A_p \cap A_q] = \zeta(s)^{-1} \sum_n (npq)^{-s} = (qp)^{-s}$$

More generally, we see that all the events A_p with different primes are independent. Now let us compute the probability of the set of natural numbers which are not divisible by any prime number. This is $\prod_p (1 - p^{-s})$. On the other hand, this is the set $\{1\}$ which has the probability $\zeta(s)^{-1}$. We have shown Euler's formula about prime numbers.

We were discussing already examples of probability spaces which were infinite. We had computed the expectation value of some random variables taking countably many values.

1. Poisson $P[X = k] = \frac{\lambda^k}{k!} e^{-\lambda}$.
We have seen that this is a limit of Bernoulli probability spaces.
2. Petersburg $P[X = 2^k] = 2^{-k}$.
We have seen that this is an example, where $E[X] = \infty$.
3. Zeta $P[X = k] = \zeta^{-s} \lambda^k / k!$.
We have seen a proof of Euler's formula $\zeta^{-s} = \prod_{pprime} (1 - p^{-s})$.
4. Homework: $P[X = k] = 2^{-k}$. This is a special case of the geometric distribution.

We look first at more examples.

Example. (Geometric measure) Let $\Omega = \{1, 2, 3, \dots\}$ and $P[\{k\}] = (1-p)^{k-1} p$. This is the expected waiting time for success of independent p distributed random variables $P[X_1 = 0, \dots X_{k-1} = 0, X_k = 1]$. How long do we expect to have to throw a coin in order to get head? The expected value of $X(\omega) = \omega$ is

$$E[X] = \sum_{k=1}^{\infty} k(1 - p)^{k-1})p = 1/p \,.$$

How did we compute that value? Again with a differentiation trick. Since

$$\sum_k k(1 - p)^{k-1} p = \frac{d}{dp} \sum_{k=1}^{\infty} (1 - p)^k p + \sum_{k=1}^{\infty} (1 - p)^{(k-1)} = 1/p \,.$$

For random variables taking countably many values, we can also easily compute the value

$$E[f(X)] = \sum_{a \in X(\Omega)} f(a) P[X = a] \,.$$

One has to be careful however with the convergence of the integral. For example, let Y be the number of ones before a zero in the Petersburg case. Then, $X = e^Y$ was the win in the Petersburg game. The expectation of Y exists, it is

$$E[Y] = \sum_k kP[Y = k] = \sum_{k=1}^{\infty} k(1-p)^k$$

which is $(1-p)/p * E[X_{geometric}] = (1-p)/p^2$. The expectation value of X is however ∞.

In all these examples, it is not essential that Ω is discrete. We can take $\Omega = [0, 1]$ and define X as piecewise constant function. The expectation value is then computed by integration and we will come to that in the next week.

Variance and correlation.
Given a random variable taking only countably many values. (Discrete random variable). Then

$$\text{Var}[X] = \sum_{a \in X(\Omega)} (a - E[X])^2 P[X = a] \, .$$

This is as before. For example $\text{Var}[X] = E[X^2] - E[X]^2$. Now, we need however that $E[X^2] < \infty$. This is not authomatic. Take $P[X = k] = k^{-3}$. Then $\text{E}[X] = \sum_k kk^{-3} < \infty$ but $\text{E}[X^2] = \sum_k k^{-1} = \infty$.
All the other definitions are analogue.

2.5 Appendix: the Petersburg paradox

Petersburger paradox (D.Bernoulli, 1738)
In the casino, you pay an entrence fee c and you get the prize 2^T, where T is the number of times, one has to throw a coin until "head" appears. Fair would be an entrence fee of

$$c = \sum_{k=1}^{\infty} 2^k P[T = k] = \sum_{k=1}^{\infty} 1 = \infty \, .$$

The paradox is that nobody would agree to pay $c = 10$.

The following Mathematica procedures allow to simulate the game of Petersburg.

```
Game[n_]:=Table[Module[{s=1},While[Random[Integer]==1,s=2*s];s],{k,n}];
AverageProfit[n_]:=N[Apply[Plus,Game[n]]/n];
ListPlot[Table[AverageProfit[k],{k,100}]]
```

This program produced the following graph:

Figure 1. The average profit per game of the player in dependece on the number of games which were played.

What is the problem?:
The problem with the situation is that it is not quite clear, what is fair. For example, the situation $T = 20$ is so unprobably that it never occurs in the life time of a person so that one actually has not to worry about so big values of T. But how to solve the paradox? Bernoulli proposed to take not the the expectation value $E[G]$ of the profit $G = 2^T$ but an expectation value say $(E[\sqrt{G}])^2$ which would lead to a fair entrence

$$(E[\sqrt{G}])^2 = (\sum_{n=1}^{\infty} 2^{k/2} 2^{-k})^2 = (1/(\sqrt{2}-1))^2 \sim 6.25 .$$

It is not so clear if that is the way out of the paradox. Such reasoning play however a role in economical and social sciences.

2.6 Independence and the Borel Cantelli lemma

Definition. Let $\{A_n\}_{n \in \mathbf{N}}$ be a sequence of of subsets of Ω. Define the set

$$A_\infty := \limsup_{n \to \infty} A_n = \bigcap_{m=1}^{\infty} \bigcup_{n \geq m} A_n .$$

Probabilistic interpretation. The set A_∞ consists of the set $\{\omega \in \Omega\}$ such that $\omega \in A_n$ for infinitely many $n \in \mathbf{N}$. It is the event that infinitely many of the events A_n happen.

Theorem 2.6.1 (Borel-Cantelli-lemma) *1) Given a sequence $A_n \in \mathcal{A}$.*

$$\sum_{n \in \mathbf{N}} P[A_n] < \infty \Rightarrow P[A_\infty] = 0 .$$

2) Given a sequence $A_n \in \mathcal{A}$ of independent sets, then

$$\sum_{n \in \mathbf{N}} P[A_n] = \infty \Rightarrow P[A_\infty] = 1 .$$

Proof.
(1) $\mathrm{P}[A_\infty] = \lim_{n\to\infty} \mathrm{P}[\bigcup_{k \geq n} A_k] \leq \lim_{n \to \infty} \sum_{k \geq n} \mathrm{P}[A_k] = 0$.

(2) For every $n \in \mathbf{N}$, we have

$$\mathrm{P}[\bigcap_{k \geq n} A_k^c] = \prod_{k \geq n} \mathrm{P}[A_k^c] = \prod_{k \geq n} (1 - \mathrm{P}[A_k]) \leq \prod_{k \geq n} \exp(-\mathrm{P}[A_k]) = \exp(-\sum_{k \geq n} \mathrm{P}[A_k]) = 0 .$$

(We have used in the first equality the independence of the sets A_i and in the inequality step the elementary estimate $1 - x \leq e^{-x}$ for $x \geq 0$.)

From

$$\mathrm{P}[A_\infty^c] = \mathrm{P}[\bigcup_{n \in \mathbf{N}} \bigcap_{k \geq n} A_k^c] \leq \sum_{n \in \mathbf{N}} \mathrm{P}[\bigcap_{k \geq n} A_k^c] = 0$$

follows $\mathrm{P}[A_\infty^c] = 0$ and so $\mathrm{P}[A_\infty] = 1$. □

Example. (The independence is necessary in 2): Take the probability space $([0,1],\mathcal{B},P)$, where P is the Lebesgue measure on the Borel σ-algebra of $[0,1]$. For $A_n = [0,1/n]$ we get $A_\infty = \emptyset$ and $P(A_n) = 1/n$.

Appendix: Monkey typing Shakespeare

Silly example. (Monkey typing Shakespeare) Writing the collected works of Shakespeare amounts to type a sequence of N symbols on a typewriter. A monkey types symbols at random, one per unit time, producing a random sequence X_n of identically distributed sequence of random variables in the set of all possible symbols. If each letter occurs with probability at least ϵ, then the probability that Shakespeare's work appears when typing the first N letters is ϵ^N. Call A_1 this event. Call A_k the event that this happens when typing the $(k-1)N+1$ until the kN'th letter. These sets A_k are all independent and have all equal probability. The Borel-Cantelli lemma, the events occur infinitely often. This means that Shakespeare is not only written once but infinitely many times. To model this example precicely, we have to construct a probability space $(\Omega,\mathcal{A},P)$. Take a finite alphabet Δ which is a compact topological space with the discrete topology. Form the product space $\Omega = \Delta^{\mathbf{Z}}$ which is by Tychonov compact and let $\mathcal{A}$ the Borel-σ-algebra on Ω. If we put on Δ a probability measure Q, the probability measure $Q^{\mathbf{Z}}$ exists on $(\Omega,\mathcal{A})$. It has the property that any zylinder set $Z(w) = \{\omega \in \Omega \mid \omega_k = r_k, \dots \omega_n = r_n\}$ defined by a word $w = [r_k, , \dots r_n]$ has the measure $P(Z(w)) = \prod_{i=k}^n P(\omega_i = r_i) = \prod_{i=k}^n Q(r_i)$. Finite unions of zylinder sets form an algebra $\mathcal{R}$ satisfying $\sigma(\mathcal{R}) = \mathcal{A}$. One can show that P is σ additive on this algebra. By Carathéodori's continuation theorem of measure theory, there exists a measure P on $(\Omega,\mathcal{A})$. The process of typing is given by the sequence of independent random variables $X_n(\omega) = \omega_n$. The event that Sheakespeare is written in the time interval $[Nk+1, N(k+1)]$ is given by a zylinder set A_k. They have all the same probability. Use Borel-Cantelli 2).

The following Mathematica procedures allow to simulate the typing of the Monkey:

```
Alphabet={a,b,c,d,e,f,g,h,i,j,k,l,m,n,o,p,q,r,s,t,u,v,w,x,y,z};
Monkey[n_]:=Table[Alphabet[[Random[Integer,{1,26}]]],{n}]
```

We let him type 700 letters on the computer and save it in a file Sheakspeare.

```
Shakespeare=Monkey[700];
Save["Shakespeare",Shakespeare];
```

2.7 Integration and expectation

Definition. Let $(\Omega,\mathcal{A},P)$ be a proability space. Let $\mathcal{B}$ be the Borel $\sigma-$ algebra on $\mathbf{R}, d(x,y) = |x-y|$. A function $X : \Omega \to \mathbf{R}$ is called **a random variable** if for all $B \in \mathcal{B}$, $X^{-1}(B) \in \mathcal{A}$.

Example.
1) If $\mathcal{A}$ is the set of subsets of $\mathcal{A}$, then every function X is measurable.
2) Let $\Omega = \mathbf{R}$ and $\mathcal{A} = \mathcal{B}$ the Borel σ-algebra. A continuous function $X : \Omega \to \mathbf{R}$ is measurable.

Definiton. A **step function** is a random variable which is of the form $X = \sum_{i=1}^{n} \alpha_i \cdot 1_{A_i}$ with $\alpha_i \in \mathbf{R}$ and where $A_i \in \mathcal{A}$ are disjoint. Call $\mathcal{S}$ the set of such random variables. These functions can alternatively be defined as random variables which take only finitely many values. For $X \in \mathcal{S}$ we can define the *integral*

$$\mathrm{E}[X] := \int_\Omega X \, dP = \sum_{i=1}^{n} \alpha_i P(A_i) = \sum_{a \in X(\Omega)} a \cdot \mathrm{P}[X = a] \, .$$

Definition. Define $\mathcal{L}^1$ as the set of random variables X, for which

$$\sup_{Y \in \mathcal{S}, Y \leq |X|} \int Y \, dP$$

is finite. For $X \in \mathcal{L}^1$, we can define the **integral or expectation**

$$\mathrm{E}[X] := \int X \, dP = \sup_{Y \in \mathcal{S}, Y \leq X^+} \int Y \, dP - \sup_{Y \in \mathcal{S}, Y \leq X^-} \int Y \, dP$$

where $X^+ = X \vee 0 = \max(X, 0)$ and $X^- = -X \vee 0 = \max(-X, 0)$. The vector space $\mathcal{L}^1$ is called the space of *integrable random variables.*

Examples.
1) If Ω is a finite, then the expectation is the expectation we had considered in the first Chapter.
2) If Ω is a countable set, then

$$\mathrm{E}[X] = \sum_{\omega \in \Omega} P[\omega] X(\omega) \, .$$

We have seen that in a previous section also.
3) Let $\Omega = [a, b]$ with $a < b$, $\mathcal{A}$ the Borel σ algebra and $P[(c, d)] = (d - c)/(b - a)$.

Proposition 2.7.1 *Every differentiable function $f = X$ on $[a, b]$ is a random variable.*

Proof. The idea is that $\int_a^b f \, dx$ can be written as a limit of Riemann sums. More precicely, we can approximate f from below and above by functions $f_n^- \leq f \leq f_n^+$, where

$$\begin{aligned} f_n^- &= \frac{1}{n} \sum_{k=1}^{n} \inf_{x \in I_k} f(x) 1_{I_k} \\ f_n^+ &= \frac{1}{n} \sum_{k=1}^{n} \inf_{x \in I_k} f(x) 1_{I_k} \end{aligned}$$

where $I_k = [(k-1)/n, k/n)$. These functions are both in $\mathcal{S}$. We know that $\int_a^b f_n^+ - f_n^- \, dx \leq C/n$, where $C = \sup_x f'(x)$. □

Actually, much more functions are random variables. All continuous functions for example are random variables and if X_n is a sequence of random variables converging pointwise to a bounded function X, then X is a random variable. In order that a function is not a random variable, it must be pathological and Apostol's simplified definition in VOlume 2 is wrong but harmless.

2.8 Distribution of a random variable

Definition. Let X be a random variable. The **distribution function** of X is the function

$$F(t) = P[X \leq t] .$$

One reason to introduce distribution functions is that one can replace integrals on the probability space Ω by integrals on $\mathbf{R}$ which is more convenient.

See Apostol 14.5, 14.6, 14.7, 14.9, 14.10, 14.11, 14.13, 14.14, 14.15

We show now that the distribution function is a monotone function on $\mathbf{R}$ which takes values in $[0, 1]$.

Proposition 2.8.1 *A distribution function has the following properties.*
(i) $F(t) \in [0, 1]$.
(ii) $P[a < X \leq b] = F(b) - F(a)$.
(iii) $a \leq b \Rightarrow F(a) \leq F(b)$.
(iv) $\lim_{t \to -\infty} F(t) = 0, \lim_{t \to \infty} F(t) = 1$.

Definition. A random variable has a **continuous distribution**, if its distribution function is continuous. If $X(\Omega)$ is countable, then the distribution function is called **discrete**. In this case, the function is piecewise constant and has jump discontinuities. In the case of finite probabiliy space, there are finitely many discontinuities and they are exacly at the places a, where X takes values and where the jumpheight $P[X = a]$ is positive.

Distribution functions are continuous to the right:

Proposition 2.8.2 *Let F be a distribution function of a random variable X, then $F(a + 0) = F(a)$ and $F(t - 0) = F(a) - P[X = a]$.*

Proof. The limits exist because of the monotonicity and boundedness of F. For the proof, note that for any probability space and any nested sequence of sets $A_n \subset A_{n-1}$ with $\bigcap_N A_n = A$, we have $P[\bigcap_n A_n] = P[A]$.

$$F(a + 1/n) - F(a) = P[a < X \leq a + 1/n] \to 0 .$$

On the other hand

$$F(a) - F(a - 1/n) = P[a - 1/n < X \leq a] = P[a - 1/n < X < a] + P[X = a] \to P[X = a] .$$

(the proof in Apostol p 518 is too complicated). □

Definition. If $F(t) = \int_{-\infty} f(s)\, ds$ with some continuous function f, then f is called the **probability density** of the random variable X.

Remark. If X has a probabiliy density, then F is is continuous and the random variable therefore a continuous distribution.

Definition. The **expectation** of a random variable X with density $f(x)$ is given by

$$E[X] = \int_{\mathbf{R}} x\, f(x)\, dx\,.$$

Proposition 2.8.3 *Assume, we have given a function F which satisfies the above three properties. Then $P[[a,b)] = F(b) - F(a) = P[X \in [a,b)]$ defines a measure on the Borel σ-algebra of $\Omega = \mathbf{R}$ and the random variable $X(x) = x$ has the distribution F.*

Proof. We check that P is a measure when restricted to the set of halfopen intervals $\{[a,b)$. General theory assures that P can be extended to a measure P on the σ algebra generated by these sets. □

This proposition tells that we can speak of distribution functions without having to refer to explicit probability spaces.

One can see it also differently without using the distribution function. Definition. Given a random variable X. The measure μ on $\mathbf{R}$ defined by

$$\mu(B) = P[X \in B]$$

is called the **law** of X.

The random variable $Y(x) = x$ on $(\mathbf{R}, \mathcal{B}, \mu)$ and the random variable X have the same distribution. The above proposition says that we actually do not have to know the random variable itself but only the distribution function. And from this distribution, we can construct the probabiliy space and the random variable.

REPETITION:

Definition. Let X be a random variable. The **distribution function** of X is the function

$$F(t) = P[X \leq t]\,.$$

Proposition 2.8.4 *A distribution function has the following properties.*
(i) $F(t) \in [0,1]$.
(ii) $P[a < X \leq b] = F(b) - F(a)$.
(iii) $a \leq b \Rightarrow F(a) \leq F(b)$.
(iv) $\lim_{t \to -\infty} F(t) = 0, \lim_{t \to \infty} F(t) = 1$.

Definition. A random variable has a **continuous distribution**, if its distribution function is continuous. If $X(\Omega)$ is countable, then the distribution function is called **discrete.**

Proposition 2.8.5 *Let F be a distribution function of a random variable X, then $F(a+0) = F(a)$ and $F(t-0) = F(a) - P[X = a]$.*

Definition. If $F' = f$, then f is called the **probability density** of the random variable X.

Remark. If X has a probabiliy density, then F is is continuous and the random variable therefore a continuous distribution.

Proposition 2.8.6 *Assume, we have given a function F which satisfies the above three properties. Then $P[[a,b)] = F(b) - F(a) = P[X \in [a,b)]$ defines a measure on the Borel σ-algebra of $\Omega = \mathbf{R}$ and the random variable $X(x) = x$ has the distribution F.*

Definition. The **expectation** of a random variable X with density $f(x)$ is defined as

$$\mathrm{E}[X] = \int_{\mathbf{R}} x\, f(x)\, dx .$$

It follows that if g is a continuous function on $\mathbf{R}$, then

$$\mathrm{E}[g(X)] = \int_{\mathbf{R}} g(x)\, f(x)\, dx .$$

Especially,

$$\mathrm{Var}[X] = E[X^2] - E[X]^2 = \int_{\mathbf{R}} x^2\, f(x)\, dx - (\int_{\mathbf{R}} x^2\, f(x)\, dx)^2 .$$

Definition. Given a random variable X. The measure μ on $\mathbf{R}$ defined by

$$\mu(B) = P[X \in B]$$

is called the **law** of X.

DISTRIBUTION FUNCTIONS OF $\phi(X)$.

Proposition 2.8.7 *Given a continuous random variable X with density f and a differentiable invertible function $\phi : \mathbf{R} \to \mathbf{R}$ with inverse ψ. The random variable $Y = \phi(X)$ has the density*

$$g(t) = f(\psi(t))\psi'(t) .$$

Proof. Let G be the distribution function of Y:

$$G(t) = P[Y \leq t] = P[X \leq \psi(t)] = F(\psi(t)) .$$

Therefore

$$g(t) = G'(t) = F'(\psi(t))\psi'(t) = f(\psi(t)) .$$

□

2.9 Examples of probability density functions

We consider now some examples with continuous densities. A random variable X on a probability space $(\Omega, \mathcal{A}, P)$ with a given distribution f can always be written as a random variable $Y(x) = x$ on the probability space $(\mathbf{R}, \mathcal{A}, f(x)\, dx)$ because this random variable has also the distribution f.

1. $(\Omega = [a, b], \mathcal{A} = \text{Borel}, P = dx)$.
If X has the constant density $f(x) = (b-a)^{-1}$, then the distribution of X is the **uniform distribution** on the interval $[a, b]$. We compute

$$\mathrm{E}[X] = \int_a^b x\, f(x)\, dx = \frac{1}{2}(a+b)\,.$$

More generally, we have

$$\mathrm{E}[X^p] = \int_a^b x^p\, f(x)\, dx = \frac{1}{p+1}\frac{b^{p+1} - a^{p+1}}{b-a}\,.$$

and

$$\mathrm{Var}[X] = \frac{b^3 - a^3}{3(b-a)} - \frac{(a-b)^2}{4} = (b-a)^2/12\,.$$

The uniform distribution is used for errors, which are known to be in some interval like for example a random error of an angle.

2. $(\Omega = [0, \infty), \mathcal{A} = \text{Borel}, P = \alpha e^{-\alpha x}\, dx)$.
If X has the density $f(x) = \alpha e^{-\alpha x}$, then the distribution of X is the **exponential distribution** on $\mathbf{R}$. We compute recursively

$$\mathrm{E}[X^p] = \int_0^\infty x^p \alpha e^{-\alpha x}\, dx = \frac{p}{\alpha}\mathrm{E}[X^{p-1}]\,.$$

Especially, $\mathrm{E}[X] = \alpha^{-1}, \mathrm{E}[X^2] = 2/\alpha^2$ and $\mathrm{Var}[X] = \alpha^{-2}$.

The exponential distribution is used to describe the waiting time for an event like the time for getting a phonecall or the time for writing a computer program.

3. If X has the density

$$f_{m,\sigma}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}} \, dx ,$$

then the distribution of X is called the **Normal or Gauss distribution.** Important is the special case $m = 0, \sigma = 1$

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \, dx$$

which is the **standard normal distribution.** We have

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} x f(x) \, dx = m + \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} (x-m) f(x) \, dx$$

The second summand is a constant times the integral of the $d/dx f(x)$ and so vanishing. By a linear change of variables, we can assume $m = 0$ now. For odd p, we have

$$\mathrm{E}[X^p] = \int_{\mathbf{R}} x^p f(x) \, dx = 0$$

since f is an even function and so $x^p f(x) = -(-x)^p f(-x)$ is an odd function. Also

$$\mathrm{E}[|X|^p] = 2\frac{1}{\sqrt{2\pi\sigma^2}} \int_0^{\infty} x^p e^{-\frac{x^2}{2\sigma^2}} \, dx$$

A change of variables $z = x^2/(2\sigma^2)$ gives

$$\mathrm{E}[|X|^p] = \pi^{-1/2} 2^{p/2} \sigma^p \int_0^{\infty} z^{\frac{1}{2}(p+1)-1} e^{-z} \, dz = \pi^{-1/2} 2^{p/2} \sigma^p \Gamma(\frac{1}{2}(p+1)) .$$

Especially, with $1 = \pi^{-1/2}\Gamma(1/2)$,

$$\mathrm{E}[X^2] = \sigma^2$$

so that σ is the standard deviation.

$\Gamma(n) = \int_0^{\infty} z^{n-1} e^{-z} \, dz = (n-1)!$ for $n \in \mathbf{N}$. Proof. $\gamma(1) = 1$, $\Gamma(n) = (n-1)\Gamma(n-1)$ by partial integration.

The normal distribution is used to measure errors of measurements.

4. The distribution of X with density

$$f(x) = \frac{1}{\pi(1+x^2)} ,$$

is called **Cauchy distribution.** It has no expectiation since

$$\mathrm{E}[X] = \int_{\mathbf{R}} \frac{x}{\pi(1+x^2)}$$

this integral does not converge since it is majorized by $g(x) = Cx^{-1}$ for some $C > 0$. Also the Variance does not exist.

5. The distribution X with density

$$f(x) = \frac{\lambda^k}{(k-1)!} x^{k-1} e^{-\lambda x}$$

on $\mathbf{R}^+$ is called the **Erlang distribution.** The parameters are k and λ. The sum of k exponential distributed independent random variables are **Erlang distributed.** A generalisation is the **Gamma** distribution, where n is replaced by a general real number α and $(\alpha-1)!$ by $\Gamma(\alpha)$.
It follows from the fact that we have a sum of random variables that $E[X] = k/\lambda$ and $\text{Var}[X] = k/\lambda^2$.

Distribution	Parameters	Mean	Variance
ac1) Normal	$m \in \mathbf{R}$, $\sigma^2 > 0$	m	σ^2
ac2) Cauchy	-	"0"	∞
ac3) Uniform	$a < b$	$(a+b)/2$	$(b-a)^2/12$
ac4) Exponential	$\lambda > 0$	$1/\lambda$	$1/\lambda^2$
pp1) Binomial	$n \in \mathbf{N}$, $p \in [0,1]$	np	$np(1-p)$
pp2) Poisson	$\lambda > 0$	λ	λ
pp3) Uniform	$n \in \mathbf{N}$	$(1+n)/2$	$(n^2-1)/12$
pp4) Geometric	$p \in (0,1)$	$1/p$	$1/p^2$

```
Needs["Statistics`NormalDistribution`"]
Variance[NormalDistribution[0,1]]
Random[NormalDistribution[0,1]]
Plot[PDF[NormalDistribution[0,1],x],{x,-3,3}]  (*Prob. Density. Funct*)
CharacteristicFunction[NormalDistribution[0,1],t]
```

Exercice type. The random time which is used to repair a TV set is exponentially distributed with $\lambda = 0.5$. What is the probability thgat the repairtime is $\geq 3h$? How long does it take in average to repair the TV?

$X = randomvariable.$

$$P[X \geq 3] = 1 - P[X < 3] = 1 - F(3) = 1 - (1 - e^{-\lambda 3}) = e^{-\lambda 3} = 0.223 \,.$$

$E[X] = \lambda^{-1} = 2.$

Some addendum.

1) We have not yet shown that the density of the normal distribution is really a density function, i.e. that

$$\int_{-\infty}^{\infty} f_{m,\sigma}(x)\, dx = 1 \,.$$

We use a trick to compute first $I = (2\pi)^{-1/2} \int_{\mathbf{R}} e^{-x^2/2}\, dx$:
We have

$$I^2 = (2\pi)^{-1} \int_{\mathbf{R}} \int_{\mathbf{R}} e^{-x^2-y^2}\, dx\, dy$$

$$= (2\pi)^{-1} \int_0^{2\pi} d\phi \int_0^{\infty} dr r e^{-r^2/2} = 1 \,.$$

so that $I = 1$. The verification of the density is now obtained by a change of variables $x \mapsto (x-m)/\sigma$.

2) We also want to verify that $\gamma(1/2) = \sqrt{\pi}$. We do that directly with a change of variables $z = u^2, dz = 2u$.

$$\begin{aligned}
\Gamma(1/2) &= \int_0^{\infty} \frac{1}{\sqrt{z}} e^{-z} \, dz \\
&= 2 \int_0^{\infty} e^{-u^2} \, du \\
&= \sqrt{2} \int_0^{\infty} e^{-u^2/2} \, du = \sqrt{2\pi} 2
\end{aligned}$$

3) The distribution with $m = 0, \sigma = 1$ is called the **standard normal distribution.**

4) If F is the distribution function with density f, then $1 - F(t)$ is the area of the region below f over the interval $(-\infty, t]$.

Reminder.

If X has the density

$$f_{m,\sigma}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}} \, dx ,$$

then the distribution of X is called the **Normal or Gauss distribution.** Important is the special case $m = 0, \sigma = 1$

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \, dx$$

which is the **standard normal distribution.** We have

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} x f(x) \, dx = m + \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} (x-m) f(x) \, dx$$

The second summand is a constant times the integral of the $d/dx f(x)$ and so vanishing. By a linear change of variables, we can assume $m = 0$ now. For odd p, we have

$$\mathrm{E}[X^p] = \int_{\mathbf{R}} x^p f(x) \, dx = 0$$

since f is an even function and so $x^p f(x) = -(-x)^p f(-x)$ is an odd function.
Also

$$\mathrm{E}[|X|^p] = 2 \frac{1}{\sqrt{2\pi\sigma^2}} \int_0^{\infty} x^p e^{-\frac{x^2}{2\sigma^2}} \, dx$$

A change of variables $z = x^2/(2\sigma^2)$ gives

$$\mathrm{E}[|X|^p] = \pi^{-1/2} 2^{p/2} \sigma^p \int_0^{\infty} z^{\frac{1}{2}(p+1)-1} e^{-z} \, dz = \pi^{-1/2} 2^{p/2} \sigma^p \Gamma(\frac{1}{2}(p+1)) .$$

Especially, with $1 = \pi^{-1/2}\Gamma(1/2)$,

$$\mathrm{E}[X^2] = \sigma^2$$

so that σ is the standard deviation.

$\Gamma(n) = \int_0^\infty z^{n-1} e^{-z} \, dz = (n-1)!$ for $n \in \mathbf{N}$. Proof. $\gamma(1) = 1$, $\Gamma(n) = (n-1)\Gamma(n-1)$ by partial integration.

The normal distribution is used to measure errors of measurements.

Remark about Mathematica.

2.10 Joint distribution of random variables

Definition. Given d random variables $X_1, X_2, \dots, X_d$. We call $X = (X_1, X_2, \dots, X_d)$ a **random vector**.

Definition. The **distribution function** of X (also called **joint distribution** of $X_1, X_2, \dots, X_d$ is defined by the function F on $\mathbf{R}^d$:

$$F(x_1, \dots, x_d) = P[X_1 \le x_1, X_2 \le x_2, \dots, X_d \le x_d] .$$

Definition. Like in the one dimensional case, a distribution function is called **discrete** if there exists a countable set C in $\mathbf{R}^d$ such that $P[X \in C] = 1$. It is called **continuous** if there exists a function $f : \mathbf{R}^d \to \mathbf{R}^+$ such that

$$F(t) = \int_{-\infty}^{t_1} \int_{-\infty}^{t_1} \cdots \int_{-\infty}^{t_d} f(y_1, y_2, \dots, y_d) \, dy \, .$$

Lemma 2.10.1 *If F is a continuous distribution function with continuous density f then $f = F_{12,\dots,d}$.*

Proof. There exists at most one continuous density function. The formula is checked by integration. □

2.11 Transformation of random variables

Repetition: The following proposition is of practical value for the generation of absolutely continuous random variables with distribution F.

Proposition 2.11.1 *(Quantile transformation) Let F be a distribution function which is continuous and invertible. Let X be a random variable which is uniformly distributed in $[0,1]$. Then $Y = F^{-1}(X)$ gives random numbers with distribution F.*

Proof. Since $F_x(t) = t$, we have

$$F_Y(t) = P[Y \le t] = P[F^{-1}(X) \le t] = P[X \le F(t)] = F_X F(t) = F(t)$$

□

Example. $F(t) = 1 - e^{-\lambda t} = u$, then

$$F^{-1}(u) = -\lambda^{-1} \log(1-u)$$

So $Y = -\lambda^{-1} \log(1 - X)$ generates exponential distributed random variables.

More generally

Proposition 2.11.2 *Given a continuous random variable X with density f and a differentiable invertible function $\phi : \mathbf{R} \to \mathbf{R}$ with inverse ψ. The random variable $Y = \phi(X)$ has the density*

$$g(t) = f(\psi(t)) |\psi'(t)| \, .$$

Proof. Let G be the distribution function of Y:

$$G(t) = P[Y \le t] = P[X \le \psi(t)] = F(\psi(t)) \, .$$

Therefore

$$g(u) = G'(u) = F'(\psi(u))\psi'(u) = f(\psi(u))\psi'(u) \, .$$

□

Example.
Assume, X has a normal distribution with parameter m, σ. We want to compute the probability density function of $Y = e^X$.
Answer, we have $g(u) = f(\log(u))u^{-1}$ and so

$$g(u) = \frac{1}{\sqrt{2\pi\sigma^2}u} e^{-\frac{(\log(t)-m)^2}{2\sigma^2}}$$

for $u \geq u$ and $Wg(u) = 0$ else.
We compute the mean $E[Y] = e^{m+\sigma^2/2}$ and $\text{Var}[X] = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1)$.

This distribution is called the **Logarithmic Normal** distribution.

For multidimensional distributions, we have:

Proposition 2.11.3 *Given a continuous random vector X with density f and a differentiable invertible function $\phi : \mathbf{R}^d \to \mathbf{R}^d$ with inverse ψ. The random variable $Y = \phi(X)$ has the density*

$$g(u) = f(\psi(u))|Det(D\psi(u))| .$$

Proof. This is the usual formula for transformations of volume integrals. □

Example. Given a random vector (X, Y) with joint density f. Determine the density of $(U, V) = \phi(X, Y) = (X + Y, X - Y)$.
The inverse is $(x, y) = \psi(u, v) = ((u + v)/2, (u - v)/2)$ which has the determinant $-1/2$ since $Det(D\psi) = Det(D\phi^{-1}) = 1/DetD\phi$. Therefore $f_{(U,V)}(u, v) = \frac{1}{2}f_{(X,Y)}((u + v)/2, (u - v)/2)$.

2.12 Characteristic functions

Definition. The **characteristic function** of X is defined as $\phi_X(t) = \mathrm{E}[e^{itX}]$.
Discrete case:

$$\phi_X(t) = \sum_{a \in X(\Omega)} e^{ita} \mathrm{P}[X = a] .$$

Continuous case:

$$\phi_X(t) = \int_{-\infty}^{\infty} e^{itx} f(x) \, dx .$$

Remark. Characteristic functions are a nice technical tool which simplify many calculations. Especially, when we deal with sums of independent random variables, characteristic functions behave nicely. Mathematically, the characteristic function is the Fourier transform of the law of X.

Examples of characteristic functions.

Distribution	Parameter	Characteristic function
Normal	$m \in \mathbf{R}$, $\sigma^2 > 0$	$e^{mit-\sigma^2 t^2/2}$
St. normal		$e^{-t^2/2}$
Uniform	$[-a, a]$	$\sin(at)/(at)$
Exponential	$\lambda > 0$	$\lambda/(\lambda - it)$
Binomial	$n \in \mathbf{N}$, $p \in [0, 1]$	$(p + (1-p)e^{it})^n$
Poisson	$\lambda > 0$, λ	$e^{\lambda(e^{it}-1)}$
Geometric	$p \in (0, 1)$	$\frac{pe^{it}}{(1-(1-p)e^{it}}$

Proof. 1) Normal: Let $c = \sqrt{2\pi\sigma^2}^{-1}$.

$$\mathrm{E}[e^{iX}] = \int e^{itx - \frac{(x-m)^2}{2\sigma^2}} \, dx$$

The simplest is to verify this fomula instead of deriving it. We prove

$$\sqrt{2\pi\sigma^2}^{-1} \int_{-\infty}^{\infty} e^{it(x-m)+\sigma^2 t^2/2 - \frac{(x-m)^2}{2\sigma^2}} \, dx = 1$$

By a change of variables $z = (x - m)/\sigma$ this is

$$\sqrt{2\pi}^{-1} \int_{-\infty}^{\infty} e^{itz\sigma + \sigma^2 t^2/2 - z^2/2} \, dx = 1$$

which is true since the left hand side is

$$\sqrt{2\pi}^{-1} \int_{-\infty}^{\infty} e^{-(z+i\sigma t)^2/2} \, dz = 1 \, .$$

2) Uniform: $\int_{-a}^{a} e^{it}/(2a) \, dt = \sin(at)/(at)$.

3) Exponential: $\int_0^\infty \lambda e^{(it-\lambda)x} dx = \lambda/(\lambda - it)$.

4) Poisson is an exercice.

5) Binomial: $\sum_{k=0}^{n} B(n,k) e^{itk} p^k (1-p)^{(n-k)} =?$. To compute that, we need to know more about the characteristic function of sums of independent random variables. Let us compute the case $n = 1$:

$$p + e^{it}(1-p) \, .$$

□

Remark. For finite random variables, the characteristic function is a periodic function since it is a sum of periodic functions. This is not true in general any more as the above examples show.

Characteristic functions are an important tool to compute moments $\mathrm{E}[X^k]$ of a random variable X:

Proposition 2.12.1 $\mathrm{E}[X^k] = (-i)^k \phi_X^{(k)}$.

Proof. We have

$$\phi_X(t) = \mathrm{E}[e^{itX}] = \mathrm{E}[\sum_{k=0}^{\infty} \frac{i^k t^k X^k}{k!}] = \sum_{k=0}^{\infty} i^k \mathrm{E}[X^k] \frac{t^k}{k!} .$$

If we make the Taylor expansion of ϕ:

$$\phi_X(t) = \sum \frac{d}{dt^k} \phi_X(t) \frac{t^k}{k!}$$

and compaire coefficients, we get

$$\mathrm{E}[X^k] = i^k \phi_X^{(k)} .$$

□

Proposition 2.12.2 *Given a sequence of independent random variables X_j with characteristic functions ϕ_j. The characteristic function of $\sum_{j=1}^n X_j$ is $\phi(t) = \prod_{j=1}^n \phi_j(t)$.*

Proof. Since X_j are independent, we get for any complex-valued Borel-measureable functions g_j for which $\mathrm{E}[g_j(X_j)]$ exists that

$$\mathrm{E}[\prod_{j=1}^{n} g_j(X_j)] = \prod_{j=1}^{n} \mathrm{E}[g_j(X_j)]$$

since $g_j(X_j)$ are independent random variables. One can prove this first for random variables $X_j = 1_{A_j}$ and then for sums and so on but also by realizing that the σ-algebra generated by $g_j(X_j)$ is a subalgebra of the algebra generated by X_j. The claim is the special case $g_j(x) = e^{itx}$. □

Example. Take two random variables X, Y which have normal distribution $N(\sigma_1, m_1)$ and $N(\sigma_2, m_2)$. Then $\phi_{X+Y} = e^{m_1 it - \sigma_1^2 t^2/2 + m_2 it - \sigma_2^2 t^2/2}$ is $N(m_1 + m_2, \sqrt{\sigma_1^2 + \sigma_2^2})$ distributed.
This is one reason, why normal distributed random variables are so important. The sum of independent such random variables is again in the same class.

Definition. The **convolution** of f and g is given by

$$f \star g(x) = \int_{\mathbf{R}} f(x-y) g(y) \, dy .$$

Corollary 2.12.3 *Given two random variables X, Y with densities f, g. Then $X + Y$ has the density $f \star g$. More generally, the distribution of the random variable $\sum_{j=1}^n X_j$ is given by $f_1 \star f_2 \star \ldots \star f_n$, if X_j are independent with distributions f_j.*

Proof. This follows immediately from the last proposition and the algebraic isomorphisms between the set of characteristic functions with convolution product and the set of distribution functions with pointwise multiplication

$$\hat{f} \star \hat{g} = \widehat{f \cdot g} .$$

Proof:

$$\chi_f = \chi_X = \int_{-\infty}^{\infty} e^{itx} f(x)\, dx$$

$$\chi_g = \chi_Y = \int_{-\infty}^{\infty} e^{ity} g(y)\, dy$$

$$\chi_{X+Y} = \chi_X \chi_Y = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{it(x+y)} f(x)g(y)\, dy\, dx$$

We make now a change of variables $x \to z = (x+y)$ so that $x = z - y$, $dx = dz$ and

$$\chi_{X+Y} = \int_{-\infty}^{\infty} e^{itz} \int_{-\infty}^{\infty} f(z-y)g(y)\, dy = \chi_{f*g}\,.$$

□

Lemma 2.12.4 (Characteristic functions determine the distribution)
The characteristic function ϕ_X determines the distribution of X.

Proof. (Proof by literature):
Since the Fourier transform of a measure determines the measure uniquely, the characteristic function of a distribution determines the distribution.
(Proof by hand):
Since a distribution function F has only countably many points of discontinuities, it is enough to determine $F(b) - F(a)$ in terms of ϕ if a and b are continuity points of F. A computation shows that for such points a, b, we have

$$F(b) - F(a) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{e^{-ita} - e^{-itb}}{it}\phi(t)\, dt\,.$$

We verify that in the case when X has a continuous distribution. Then

$$f(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-itx}\phi(t)\, dt$$

which gives after integration from a to b the above result.
In the discrete case, we have

$$P[X = k] = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-itk}\phi_X(t)\, dt$$

which is the k'th Fourier coefficient of the periodic function ϕ_X. □

2.13 Distributions of independent random variables

Definition. Two random variables X, Y are independent, if for all intervals $[a, b), [c, d)$, we have

$$P[X \in [a,b); Y \in [c,d)] = P[X \in [a,b)]P[Y \in [c,d)]\,.$$

This generalises the definition for discrete random variables, where we required

$$P[X = a; Y = b] = P[X = a] \cdot P[Y = b]\,.$$

We get this statement by choosing intervals which contain only one point in $X(\Omega)$.

For finitely many random variables, the definition is analogue and for an arbitrary set of random variables, we require that every finite subset is independent.

Proposition 2.13.1 *Two random variable X, Y are independent if and only if their distribution functions F_X, F_Y satisfy*

$$F_{(X,Y)}(s,t) = F_X(s)F_Y(t) .$$

Proof. If X, Y are independent, then

$$F_{(X,Y)}(s,t) = P[X \leq s, Y \leq t] = P[X \leq s] \cdot P[Y \leq t] = F_X(s) \cdot F_Y(t) .$$

Assume now that $F_{(X,Y)}(s,t) = F_X(s)F_Y(t)$. Take two intervals $[a,b), [c,d)$. We have

$$\begin{aligned} P[\{X \in [a,b)\} \cap \{Y \in [c,d)\}] &= P[\{X \leq b\} \cap \{Y \leq d\}] - P[\{X \leq b\} \cap \{Y \leq c\}] - P[\{X \leq a\} \cap \{Y \leq d\}] + P[\{X \\ &= F_{X,Y}(b,d) - F_{X,Y}(b,c) - F_{(X,Y)}(a,d) + F_{X,Y}(a,b) \\ &= F_X(b)F_Y(d) - F_X(b)F_Y(c) - F_X(a)F_Y(d) + F_X(a)F_Y(b) \\ &= (F_X(b) - F_X(a))(F_Y(d) - F_X(c)) = P[X \in [a,b)]P[Y \in [c,d)] \end{aligned}$$

□

Corollary 2.13.2 *Two continuous random variables are independent if and only if their density functions f_X, f_Y satisfy*

$$f_{(X,Y)}(x,y) = f_X(x)f_Y(y) .$$

Proof. If $f_{(X,Y)}(x,y) = f_X(x)f_Y(y)$, then $F_{(X,Y)}(s,t) = F_X(s)F_Y(t)$ which implies the claim. The other direction follows by differentiation. □

2.14 Inequalities

The Chebychev inequality

Theorem 2.14.1 (Chebychev-Markov inequality) *Let h be a monoton function on $\mathbf{R}$ with $h \geq 0$. For every $c > 0$, and $h(X) \in \mathcal{L}^1$ we have*

$$h(c) \cdot \mathrm{P}[X \geq c] \leq \mathrm{E}[h(X)] .$$

Proof. Integrate the inequality $h(c)1_{X \geq c} \leq h(X)$ gives directly the above inequality. We used the monotonicity and linearity of the expectation. □

Example. $h(x) = |x|$ leads to $P(|X| \geq c) \leq ||X||_1/c$ which implies for example that $\mathrm{E}[|X|] = 0 \Rightarrow \mathrm{P}[X = 0] = 1$.

Important is the following special case:

Theorem 2.14.2 (Chebychev inequality) *If $X \in \mathcal{L}^2$, then*

$$\mathrm{P}[|X - \mathrm{E}[X]| \geq c] \leq \frac{\mathrm{Var}[X]}{c^2} .$$

Proof. Take $h(x) = x^2$ and apply the Chebychev-Markov inequality to the random variable $Y = X - \mathrm{E}[X] \in \mathcal{L}^2$ satisfying $h(Y) \in \mathcal{L}^1$. □

The Chebychev inequality is used to estimate the probabiliy that an experiment is more than c away from its mean. This inequality shows that the variation really measures the fluctuation from the mean.

Example. Given a standard normal distributed random variable X. What is the probability that $|X| > \epsilon$? We can say that the probability is less or equal than $1/c^2$. The exact probabiliy is $2F(c)$ which could be computed by an integral. The nice thing about the Chebychev inequality is that it gives a handy estimate about the probabiliy of an error.

Example. Uniform distribution on $[-1, 1]$. Compare the error given by Chebychev with the real error.

Example. Gambling in the casino with roulette. We have seen that in all the games, the mean loosing is the same and only the variance changes. The probability to make an amoutn c more or less can be estimated by the variance.

2.15 The weak law of large numbers

Let $X_1, X_2, \dots$ a sequence of random variables on a probability space $(\Omega, \mathcal{A}, P)$. We are interested in the asymptotic behavior of the sums $S_n = X_1 + X_2 + \dots X_n$ for $n \to \infty$ and especially in the convergence of the averages S_n/n. This behavior is described by "laws of large numbers". Depending on the definition of convergence, one speaks of "weak" and "strong" laws of large numbers.

How can we generate a sequence of independent random variables?

Proposition 2.15.1 (Construction of independent random variables) *Given a distribution function F, there exists a probability space $(\Omega, \mathcal{A}, P)$ and independent random variables $X_1, X_2, \dots$ which have all the distribution F.*

Proof. We have already seen, how we can construct a single random variable X with distribution F. Call this probability space $(\mathbf{R}, \mathcal{B}, Q)$. Now, we form the product space

$$(\Omega, \mathcal{A}, P) = (\mathbf{R}^{\mathbf{N}}, \mathcal{B}^{\mathbf{N}}, Q^{\mathbf{N}}) .$$

It consists of sequences $\omega = (\omega_1, \omega_2, \dots)$ and the probability measure P is defined by

$$P[A_1 \times A_2 \times A_3 \times A_n \cdots \times A_n \times \mathbf{R} \times \mathbf{R} \times \dots] = Q[A_1] \cdot Q[A_2] \dots Q[A_n] .$$

The random variables $X_i(\omega) = \omega_i$ are now independent and have all the distribution F. □

We first prove the weak law of large numbers. There exist different versions of this theorem since more assumptions on X_n can allow stronger statements.

Definition. A sequence of random variables Y_n converges **in probability** to a random variable Y, if for all $\epsilon > 0$,

$$\lim_{n \to \infty} P[|Y_n - Y| \geq \epsilon] = 0 .$$

Comparison of convergences: if for some $p \in [1, \infty)$, $||X_n - X||_p \to 0$, then $X_n \to X$ in probability since by the Chebychev-Markov inequality, $P[|X_n - X| \geq \epsilon] \leq ||X - X_n||^p/\epsilon^p$.

The convergence defined by $X_n \to_p X$ if $E[|X_n - X|] \to 0$ is stronger than the convergence in probability.

Note that we do not need to need the random variables to be integrable in order to define the convergence in probability.

Theorem 2.15.2 (Weak law of large numbers for uncorrelated $X_i \in \mathcal{L}^2$) *Assume $X_i \in \mathcal{L}^2$ have common expectation $E[X_i] = m$ and satisfy* $\sup_n \frac{1}{n} \sum_{i=1}^n \mathrm{Var}[X_i] < \infty$. *If X_n are pairwise uncorrelated* ($\mathrm{Cov}[X_i, X_j] = 0$ *for $i \neq j$) then $S_n/n \to m$ in probability.*

Proof. Since $\mathrm{Var}[X + Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2 \cdot \mathrm{Cov}[X, Y]$ and X_n are pairwise uncorrelated, we get $\mathrm{Var}[X_n + X_m] = \mathrm{Var}[X_n] + \mathrm{Var}[X_m]$ and by induction $\mathrm{Var}[S_n] = \sum_{i=1}^n \mathrm{Var}[X_n]$. Using linearity, we obtain $E[S_n/n] = m$ and

$$\mathrm{Var}[S_n/n] = E[(S_n)^2/n^2] - E[S_n]^2/n^2 = \mathrm{Var}[S_n]/n^2 = \frac{1}{n^2} \sum_{i=1}^n \mathrm{Var}[X_n] \to 0 .$$

With Chebychev's inequality, we obtain

$$P[|S_n/n - m| \geq \epsilon] \leq \frac{\mathrm{Var}[S_n/n]}{\epsilon^2} .$$

□

In the first version of the weak law of large numbers Theorem 2.15.2, we only assumed the random variables to be uncorrelated. Under the stronger condition of independece and a stronger conditions on the moments, the convergence can be accelerated:

Jensen Inequality Definition. A function h is *convex*, if there exists for all $x_0 \in \mathbf{R}$ a linear map $l(x) = ax + b$ such that $l(x_0) = h(x_0)$ and for all $x \in \mathbf{R}$ the inequality $l(x) \leq h(x)$ holds.

Theorem 2.15.3 (Jensen inequality) *Given $X \in \mathcal{L}^1$. For any convex function $h : \mathbf{R} \to \mathbf{R}$, we have*

$$\mathrm{E}[h(X)] \geq h(\mathrm{E}[X])$$

where the left hand side can also be infinite.

Proof. Let l be the linear map at $x_0 = \mathrm{E}[X]$. By the linearity and monoticity of the expectation, we get

$$h(\mathrm{E}[X]) = l(\mathrm{E}[X]) = \mathrm{E}[l(X)] \leq \mathrm{E}[h(X)] .$$

□

Example. Given $p \leq q$. Define $h(x) = |x|^{q/p}$. Jensen's inequality gives $\mathrm{E}[|X|^q] = \mathrm{E}[h(|X|^p)] \leq h(\mathrm{E}[|X|^p] = \mathrm{E}[|X|^p]^{q/p}$. This implies that $||X||_q := \mathrm{E}[|X|^q]^{1/q} \leq \mathrm{E}[|X|^p]^{1/p} = ||X||_p$ for $p \leq q$ and so $\mathcal{L}^p \subset \mathcal{L}^q$ for $p \geq q$.

Theorem 2.15.4 (Weak law of large numbers for independent $X_i \in \mathcal{L}^4$) *Assume $X_i \in \mathcal{L}^4$ have common expectation $E[X_i] = m$ and satisfy $M = \sup_n ||X||_4 < \infty$. If X_i are independent, then $S_n/n \to m$ in probability. More precicely, $\sum_n P[|S_n/n| \geq \epsilon]$ converges for all $\epsilon > 0$.*

Proof. We can assume without loss of generality that $m = 0$. Because the X_i are independent, we get

$$E[S_n^4] = \sum_{i_1,i_2,i_3,i_4=1}^{n} E[X_{i_1} X_{i_2} X_{i_3} X_{i_4}] .$$

Again by independence, a summand $E[X_{i_1} X_{i_2} X_{i_3} X_{i_4}]$ is vanishing if an index $i = i_k$ occurs allone, is $E[X_i^4]$ if all indices are the same and $E[X_i^2]E[X_j^2]$, if there are two pairwise equal indices. Since by Jensen's inequality $E[X_i^2]^2 \leq E[X_i^4] \leq M$, we get

$$E[S_n^4] \leq nM + n(n+1)M .$$

Use now the generalized Chebychev inequality with $h(x) = x^4$ to get

$$\begin{aligned} P[|S_n/n| \geq \epsilon] &\leq \frac{E[(S_n/n)^4]}{\epsilon^4} \\ &\leq M\frac{n+n^2}{\epsilon^4 n^4} \leq 2M\frac{1}{\epsilon^4 n^2} . \end{aligned}$$

□

2.16 Experiment

```
<<Statistics'Master'

X1:=Random[Integer,{-1,1}];
X2:=Random[NormalDistribution[0,1]];
X3:=Random[CauchyDistribution[1,1]];

X:=X1;T[x_]:=x+X;S=NestList[T,0.,1000]; ListPlot[Table[S[[n]]/n,{n,Length[S]}]];
```

2.17 The strong law of large numbers

The weak laws of large numbers made statements about stochastic convergence of sums $S_n/n = (X_1 + \ldots + X_n)/n$ of random variables X_n. The strong laws of large numbers make analogue statements about almost everywhere convergence.

In order to formulate the strong law of large numbers, we need some other notions of convergence. The first three of the following definitions we have met already.

Definition. A sequence of random variables X_n converges **in probability** to a random variable X, if $P[|X_n - X| \geq \epsilon] \to 0$ for all $\epsilon > 0$.

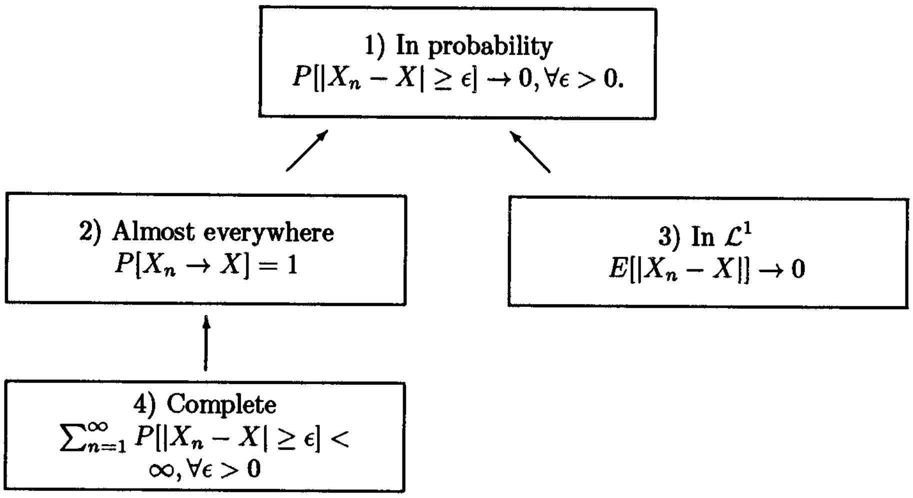
Definition. A sequence of random variables X_n converges **almost everywhere** to a random variable X, if $P[X_n \to X] = 1$.

Definition. A sequence of $\mathcal{L}^p$ random variables X_n **converges in $\mathcal{L}^p$** to a random variable X, if $||X_n - X||_p \to 0$ for $n \to \infty$.

Definition. A sequence of random variables X_n converges **fast in probability**, or **completely** $\sum_n P[|X_n - X| \geq \epsilon] < \infty$ for all $\epsilon > 0$.

We have now four notions of convergence of random variables $X_n \to X$:

Proposition 2.17.1 (Relations between convergences) :



Proof. 2) ⇒ 1): Since

$$\{X_n \to X\} = \bigcap_k \bigcup_m \bigcap_{n \geq m} \{|X_n - X| \leq 1/k\}$$

"almost everywhere convergence" is equivalent with

$$1 = P[\bigcup_m \bigcap_{n \geq m} \{|X_n - X| \leq 1/k\}] = \lim_{n \to \infty} P[\bigcap_{n \geq m} \{|X_n - X| \leq 1/k\}]$$

for all k. Therefore

$$P[|X_m - X| \geq \epsilon] \leq P[\bigcap_{n \geq m} \{|X_n - X| \geq \epsilon\}] \to 0$$

for all $\epsilon > 0$.

4) $\Rightarrow$ 2): From Borel Cantelli, we get for all $\epsilon > 0$

$$P[|X_n - X| \geq \epsilon, \text{ infinitely often}] = 0$$

We get so for $\epsilon_n \to 0$

$$P[\bigcup_n |X_n - X| \geq \epsilon_k, \text{ infinitely often}] \leq \sum_n P[|X_n - X| \geq \epsilon_k, \text{ infinitely often}] = 0$$

from which we obtain $P[X_n \to X] = 1$.

3) $\Rightarrow$ 1): Use Chebychev-Markov inequality to get $P[|X_n - X| \geq \epsilon] \leq E[|X_n - X|^p]/\epsilon^p$. □

Theorem 2.17.2 (Weak law of large numbers for independent $X_i \in \mathcal{L}^4$) *Assume $X_i \in \mathcal{L}^4$ have common expectation $E[X_i] = m$ and satisfy $M = \sup_n ||X||_4, ||X||_2 < \infty$. If X_i are independent, then $S_n/n \to m$ in probability. More precicely, $\sum_n P[|S_n/n - m| \geq \epsilon]$ converges for all $\epsilon > 0$.*

Proof. We can assume without loss of generality that $m = 0$. Because the X_i are independent, we get

$$E[S_n^4] = \sum_{i_1,i_2,i_3,i_4=1}^{n} E[X_{i_1} X_{i_2} X_{i_3} X_{i_4}] .$$

Again by independence, a summand $E[X_{i_1} X_{i_2} X_{i_3} X_{i_4}]$ is vanishing if an index $i = i_k$ occurs allone, is $E[X_i^4]$ if all indices are the same and $E[X_i^2]E[X_j^2]$, if there are two pairwise equal indices. Since by Jensen's inequality $E[X_i^2]^2 \leq E[X_i^4] \leq M$, we get

$$E[S_n^4] \leq nM + n(n+1)M .$$

Use now the generalized Chebychev inequality with $h(x) = x^4$ to get

$$\begin{aligned} P[|S_n/n| \geq \epsilon] &\leq \frac{E[(S_n/n)^4]}{\epsilon^4} \\ &\leq M\frac{n+n^2}{\epsilon^4 n^4} \leq 2M\frac{1}{\epsilon^4 n^2} . \end{aligned}$$

□

Theorem 2.17.3 (Strong law for independent X_n, bounded in $\mathcal{L}^4$) *Assume X_n are independent random variables in $\mathcal{L}^4$ with common expectation $E[X_n] = m$ and satisfying $M = \sup_n ||X_n||_4^4 < \infty$. Then $S_n/n \to m$ almost everywhere.*

Proof. In the proof of the weak law of large numbers dealing with $\mathcal{L}^4$ random variables, we got

$$P[|S_n/n - m| \geq \epsilon] \leq 2M\frac{1}{\epsilon^4 n^2} .$$

This means that $S_n/n \to m$ fast in probability which implies convergence almost everywhere. □

Theorem 2.17.4 (Strong law for pairwise independent $X_n \in \mathcal{L}^1$) *Assume $X_n \in \mathcal{L}^1$ are pairwise independent and identically distributed. Then $S_n/n \to E[X_1]$ almost everywhere.*

2.18 Appendix. Normality of numbers

```
(* UNSOLVED PROBLEM; IS PI NORMAL ? IT WOULD IMPLY THAT THE MEAN  *)
(* OF THE DIGITS EXISTS AND IS 4.5                                *)

Mean[s_]:=N[Sum[s[[i]],{i,Length[s]}]/Length[s]];
PiDigits[n_]:=RealDigits[N[Pi,n]][[1]];
ListPlot[Table[Mean[PiDigits[j]],{j,300}]]

Mean[PiDigits[50000]]

(* THE SAME QUESTION FOR E                                        *)

EDigits[n_]:=RealDigits[N[E,n]][[1]];
ListPlot[Table[Mean[EDigits[j]],{j,300}]];
Mean[EDigits[50000]]

(* A TWO DIMENSIONAL TEST                                         *)

PiDigits[n_]:=Partition[RealDigits[N[Pi,n]][[1]],2];
ListPlot[Partition[PiDigits[1000],2]];
```

2.19 The central limit theorem

Definition. A sequence of random variables X_n converges **in distribution** to a random variable X, if for all $t \in \mathbf{R}$, $\mathrm{P}_n[X_n \leq t] \to \mathrm{P}[X \leq t]$ for $n \to \infty$.

Relations between convergences. See summary. The convergence in distribution is clearly the weakest version.

Lemma 2.19.1 *X_n converges in distribution to X if its characteristic functions $\phi_{X_n}(t)$ converge pointwise to the characteristic function $\phi_X(t)$.*

Proof. We can recover the distribution function F_X from X:

$$F(b) - F(a) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-ita} - e^{-itb}}{it} \phi(t) \, dt \, .$$

In the continuous case, it is even easier to recover the density

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-itx} \phi(t) \, dt$$

or in the discrete case

$$\mathrm{P}[X = k] = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-itk} \phi_X(t) \, dt$$

□

Given a random variable X for which its mean $\mathrm{E}[X]$ and standard deviation $\sigma[X]$ exists, we define $X^* = (X - \mathrm{E}[X])/\sigma[X]$. Note that if we take sums of independent random variables, then the mean $\mathrm{E}[S_n] = n\mathrm{E}[X]$ and standard deviation satisfy $\sigma[S_n] = \sigma[X_n]\sqrt{n}$.

Theorem 2.19.2 (Central limit theorem for IID random variables) *Given $X_n \in \mathcal{L}^2$ which are IID with mean m and variance σ^2. Then $S_n^* \to N(0,1)$ in distribution.*

Proof. Since the characteristic function of $N(0,1)$ is $e^{-t^2/2}$ we have to show that for all $t \in \mathbf{R}$

$$\mathrm{E}[e^{it\frac{S_n}{\sigma\sqrt{n}}}] \to e^{-t^2/2} \, .$$

Denote with ϕ the characteristic function of X_n. Since by assumption $\mathrm{E}[X_n] = 0, \mathrm{E}[X_n^2] = \sigma^2$, we have

$$\phi(t) = 1 - \frac{\sigma^2}{2} t^2 + o(t^2) \, .$$

Therefore

$$\begin{aligned} \mathrm{E}[e^{it\frac{S_n}{\sigma\sqrt{n}}}] &= \phi(\frac{t}{\sigma\sqrt{n}})^n \\ &= (1 - \frac{1}{2}\frac{t^2}{n} + o(1/n))^n \\ &= e^{-t^2/2} + o(1) \, . \end{aligned}$$

□

Chapter 3

Summaries

2. Week, (Summary of the theory)

DEFINITION: Boolean algebra

Ω be a finite set. A set $\mathcal{A}$ of subsets of Ω is a **Boolean algebra**, if

$\Omega \in \mathcal{A}$,
$A \in \mathcal{A} \to A^c \in \mathcal{A}$,
$A, B \in \mathcal{A} \Rightarrow A \cup B \in \mathcal{A}$.

PROPERTIES: A Boolean algebra $(\Omega, \mathcal{A})$ is closed under all set theoretical operations: $A, B \in \mathcal{A}$, then

$\emptyset \in \mathcal{A}$	$A \cap B \in \mathcal{A}$.
$A \setminus B \in \mathcal{A}$	$A \Delta B \in \mathcal{A}$.

DEFINITION: Probability measure

A function $P : \mathcal{A} \to \mathbf{R}$ is a **probability measure** if

$\mathrm{P}[A] \geq 0$, (nonnegativity)
$\mathrm{P}[\Omega] = 1$, (normalisation)
$\mathrm{P}[\bigcup_{i=1}^{n} A_i = \sum_{i=1}^{n} \mathrm{P}[A_i], if A_i \cap A_j = \emptyset$, all i, j, (additivity)

DEFINITION: Finite probability space

We say $(\Omega, \mathcal{A}, P)$ is a finite probability space if $\mathcal{A}$ is a Boolean algebra on the finite set Ω and P is a probability measure on $(\Omega, \mathcal{A})$.

PROPERTIES:
$A \subset B \Rightarrow \mathrm{P}[A] \leq \mathrm{P}[B]$.
$\mathrm{P}[A^c] = 1 - \mathrm{P}[A]$
$\mathrm{P}[\emptyset] = 0$

SWITCH ON, SWITCH OFF formula:

$$\mathrm{P}[\bigcup_{i=1}^{n} A_i] = \sum_{k=1}^{n}(-1)^{k-1} \sum_{1 \leq i_1 < i_2 < \ldots < i_k \leq n} \mathrm{P}[A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}]$$

CONSTRUCTION OF NEW PROBABILITY SPACES:

Change of algebra: $(\Omega, \mathcal{B}, Q)$ is a probability space, where $\mathcal{B} \subset \mathcal{A}$ is a Boolean algebra, Q restriction of P to $\mathcal{B}$.
Product space:
$(\Omega, \mathcal{A}, P) = (\Omega_1, \mathcal{A}_1, P_1) \times (\Omega_2, \mathcal{A}_2, P_2)$, where $\Omega = \Omega_1 \times \Omega_2$, $\mathcal{A}$ is the smallest algebra containing $\mathcal{A}_1 \times \mathcal{A}_2 = \{A_1 \times A_2 \mid A_i \in \mathcal{A}_i\}$ and $\mathrm{P}[(A_1 \times A_2)] = P_1[A_1] \cdot P_2[A_2]$.
Conditional probability space:
$(B, \mathcal{A} \cap B, \mathrm{P}[\cdot|B])$, if $\mathrm{Pr}[B] > 0$ and $\mathcal{A} \cap A = \{A \cap B \mid A \in \mathcal{A}\}$

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

3. Week, (Summary of the theory)

DEFINITION:
$(\Omega, \mathcal{A}, P)$ finite probability space.
$A, B \in \mathcal{A}$ are **independent** if and only if

$$P[A \cap B] = P[A] \cdot P[B] .$$

A finite set $\{A_i\}_{i \in I}$ of events is called **independent** if and only if for all $J \subset I$

$$P[\bigcap_{i \in J} A_i] = \prod_{i \in J} P[A_i] .$$

PROPERTIES:
$A, B \in \mathcal{A}$ are independent, if and only if either $P[B] = 0$ or $P[A|B] = P[A]$.

$(\Omega, \mathcal{A}, P) = (\Delta, \mathcal{B}, Q)^n$ (product space). Given $B_i \in \mathcal{B}$ then

$$A_i = \{\omega = (\omega_1, \ldots, \omega_n) \mid \omega_i \in B_i\}$$

are all independent.

DEFINITION: A **random variable** on a finite probability space $(\Omega, \mathcal{A}, P)$ is a map $X : \Omega \to \mathbf{R}$ such that for all $a \in \mathbf{R}$, we have $\{X = a\} \in \mathcal{A}$.

DEFINITION: The **expectation** of a random variable X is defined as

$$E[X] = \sum_{a \in X(\Omega)} a \cdot P[X = a] = \sum_{A \in \mathcal{A}, A \text{ atom}} X(A) \cdot P[A] ,$$

where an **atom** is a set in $\mathcal{A}$ so that $B \subset A, B \in \mathcal{A} \Rightarrow B = A$ or $B = \emptyset$. If $\mathcal{A} = \{A \subset \Omega\}$, then the atoms are all of the form $\{\omega\}$ and

$$E[X] = \sum_{\omega \in \Omega} X(\omega) P[\{\omega\}] .$$

By the definition of a random variable, X must be constant on each atom A and $X(A)$ is defined as the common value, X takes on A. The two expressions for $E[X]$ in the box to the left are seen to be the same using $a = X(A)$ and $P[X = a] = \sum_{A \text{ atom } X(A)=a} P[A]$.

PROPERTIES OF EXPECTATION: For random variables X, Y and $\lambda \in \mathbf{R}$

$E[X + Y] = E[X] + E[Y]$	$E[\lambda X] = \lambda E[X]$
$X \leq Y \Rightarrow E[X] \leq E[Y]$	$E[X^2] = 0 \Leftrightarrow X = 0$
$E[X] = c$ if $X(\omega) = c$ is constant	$E[X - E[X]] = 0$.

PROOF OF THE ABOVE PROPERTIES:

$E[X + Y] = \sum_{A \text{ atom}} (X + Y)(A) \cdot P[A] = \sum_{A \text{ atom}} (X(A) + Y(A)) \cdot P[A] = E[X] + E[Y]$
$E[\lambda X] = \sum_{A \text{ atom}} (\lambda X)(A) P[A] = \lambda \sum_{A \text{ atom}} X(A) P[A] = \lambda E[X]$
$X \leq Y \Rightarrow X(A) \leq Y(A)$, for all atoms A and $E[X] \leq E[Y]$
$E[X^2] = 0 \Leftrightarrow X^2(A) = 0$ for all atoms A $\Leftrightarrow X = 0$
$X(\omega) = c$ is constant $\Rightarrow E[X] = c \cdot P[X = c] = c \cdot 1 = c$
$E[X - E[X]] = E[X] - E[E[X]] = E[X] - E[X] = 0$

4. Week, (Summary of the theory)

DEFINITION: $(\Omega, \mathcal{A}, P)$ probability space, X, Y random variables.

Variance

$$\mathrm{Var}[X] = \mathrm{E}[\,(X - \mathrm{E}[X])^2\,]\,.$$

Standard deviation

$$\sigma[X] = \sqrt{\mathrm{Var}[X]}\,.$$

Covariance

$$\mathrm{Cov}[X,Y] = \mathrm{E}[(X - \mathrm{E}[X]) \cdot (Y - \mathrm{E}[Y])]\,.$$

Correlation of $\mathrm{Var}[X] \neq 0, \mathrm{Var}[Y] \neq 0$

$$\mathrm{Corr}[X,Y] = \frac{\mathrm{Cov}[X,Y]}{\sigma[X]\sigma[Y]}\,.$$

$\mathrm{Corr}[X,Y] = 0$: **uncorrelated** X and Y.

PROPERTIES of VAR, COV, and CORR:

$\mathrm{Var}[X] \geq 0$.
$\mathrm{Var}[X] = \mathrm{E}[X^2] - \mathrm{E}[X]^2$.
$\mathrm{Var}[\lambda X] = \lambda^2 \mathrm{Var}[X]$.

$\mathrm{Cov}[X,Y] = \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y]$.
$\mathrm{Cov}[X,Y] \leq \sigma[X]\sigma[Y]$ (Schwarz inequality).

$-1 \leq \mathrm{Corr}[X,Y] \leq 1$.
$\mathrm{Corr}[X,Y] = 1$ if $X - \mathrm{E}[X] = Y - \mathrm{E}[Y]$
$\mathrm{Corr}[X,Y] = -1$ if $X - \mathrm{E}[X] = -(Y - \mathrm{E}[Y])$.

BERNOULLI DISTRIBUTED RANDOM VARIABLES: $(\Omega = \{0,1\}^n, \mathcal{A}, P = Q^n)$, where $Q[\{1\}] = p, Q[\{0\}] = q = 1 - p$.

$$X(\omega) = \sum_{i=1}^{n} \omega_i$$

$$P[X = k] = \binom{n}{k} p^k q^{n-k}$$

$$E[X] = \sum_{k=1}^{n} k \binom{n}{k} p^k q^{n-k} = pn$$

$$\mathrm{Var}[X] = \sum_{k=1}^{n} k^2 \binom{n}{k} p^k q^{n-k} - E[X]^2 = npq$$

DEFINITION: X, Y are **independent** if for all $a, b \in \mathbf{R}$

$$P[X = a; Y = b] = P[X = a] \cdot P[Y = b]\,.$$

A finite collection $\{X_i\}_{i \in I}$ of random variables are **independent**, if for all $J \subset I$ and $a_i \in \mathbf{R}$

$$P[X_i = a_i, i \in J] = \prod_{i \in J} P[X_i = a_i]\,.$$

PROPERTIES:

- If X and Y are independent, then $\mathrm{E}[X \cdot Y] = \mathrm{E}[X] \cdot \mathrm{E}[Y]$.
- If X_i is a set of independent random variables, then $\mathrm{E}[\prod_{i=1}^{n} X_i] = \prod_{i=1}^{n} \mathrm{E}[X_i]$.
- If X, Y are independent then $\mathrm{Cov}[X,Y] = 0$.
- A constant random variable is independent to any other random variable.

DEFINITION: The **regression line** of two random variables X, Y is defined as $y = ax+b$, where

$$a = \frac{\mathrm{Cov}[X,Y]}{\mathrm{Var}[X]}, b = \mathrm{E}[Y] - a\mathrm{E}[X] .$$

PROPERTY: Given $X, \mathrm{Cov}[X,Y], \mathrm{E}[Y]$, and the regression line $y = ax+b$ of X, Y. The random variable $\tilde{Y} = aX+b$ minimizes $\mathrm{Var}[Y-\tilde{Y}]$ under the constraint $\mathrm{E}[Y] = \mathrm{E}[\tilde{Y}]$ and is the best guess for Y, when knowing only $\mathrm{E}[Y]$ and $\mathrm{Cov}[X,Y]$. We check $\mathrm{Cov}[X,Y] = \mathrm{Cov}[X,\tilde{Y}]$.

5. Week, (Summary of the theory)

DEFINITION: Important example: **One dimensional random walk** $(\Omega = \{-1,1\}^N = \{(\omega = (\omega_1 < \ldots, \omega_N) \mid \omega_i \in \{-1,1\}\}, \mathcal{A} = \{A \subset \Omega\}, \mathrm{P}[A] = |A|/|\Omega|)$. The random variables $X_k(\omega) = \omega_k$ define the k'th step. The random variables $S_n = \sum_{k=1}^n X_k(\omega)$ describe the location of the random walker (drunken sailor) at time n. If X_k is the win or loss in a game at time k, then S_n is the total win or loss up to time n. Ω is the set of all possible trajectories up to time N.

PROPERTIES Random walk:

a) $I \subset \{1,\ldots,N\}$, $x_i \in \{-1,1\}$

$$\mathrm{P}[X_i = x_i, i \in I] = 2^{-|I|}.$$

b) $\mathrm{E}[X_k] = 0$,
c) $\mathrm{E}[S_k] = 0$.
d) $n+x$ even, $\mathrm{P}[X_n = x] = 2^{-n}\binom{n}{\frac{n+x}{2}}$.
$n+x$ odd, $\mathrm{P}[X_n = x] = 0$.

DEFINITION: A **gambling system** attached to the random walk is sequence of random variables V_k such that every event $\{V_n = c\}$ is a union of sets of the form $\{\omega_1 = x_1, \ldots, \omega_{n-1} = x_{n-1}\}$.
Let V_k be a gambling system, then

$$S_n^V = \sum_{i=1}^n V_k X_k$$

is the **total winnings** with this system.

PROPERTY of gambling systems:
You can't beat the system: $\mathrm{E}[S_N^V] = 0$.

DEFINITION: Cardinality.
$f : A \to B$ is $1:1$ or injective: $x \neq y \Rightarrow f(x) \neq f(y) \forall x, y \in A$.
$f : A \to B$ is **onto** or surjective): $f(A) = B$.
f is **bijectiv** $\Leftrightarrow$ f is 1:1 and onto.
A, B are called **equivalent**, if there exists a bijection $f : A \to B$.
A equivalent to **N**: **countable infinite.**
A equivalent to finite set: **finite.**
A neither finite nor countable infinite: **uncountable.**

DEFINITION:

A **σ-algebra** on Ω is a set $\mathcal{A}$ of subsets of Ω satisfying

$\Omega \in \mathcal{A}$,
$A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$,
$\{A_1, A_2, \ldots\} \subset \mathcal{A}$ countable $\Rightarrow$ $\bigcup_{i=1}^\infty A_i \in \mathcal{A}$.

$P : \mathcal{A} \to \mathbf{R}$ is a **probability measure** if

$\mathrm{P}[A] \geq 0$, (nonnegativity)
$\mathrm{P}[\Omega] = 1$, (normalisation)
$\{A_1, A_2, \ldots\}$ countable set of disjoint sets $\Rightarrow$
$\mathrm{P}[\bigcup_{i=1}^\infty A_i] = \sum_{i=1}^\infty \mathrm{P}[A_i]$ (**σ-additivity**)

A **probability space** $(\Omega, \mathcal{A}, P)$ sonsists of a set Ω, a σ-algebra $\mathcal{A}$ on Ω and a probability measure P on $\mathcal{A}$.

If Ω is **finite**, the probability space is called a finite probability space. If Ω is countable, it is called **discrete**.

6. Week, (Summary of the theory)

DEFINITION: A function $d : \Omega \times \Omega \to \mathbf{R}$ is called a **metric** if

(i) $d(x,y) \geq 0$,
(ii) $d(x,z) \leq d(x,y) + d(y,z)$
(iii) $d(x,y) = 0 \Leftrightarrow x = y$

The pair (Ω, d) where Ω is a set and d is a metric is called a **metric space**. Examples. $(\mathbf{R}^n, d(x-y) = ||x-y||$, $(\{0,1\}^{\mathbf{N}}, d(x,y) = \sum_{i=1}^{\infty} |x_i - y_i| 2^{-i})$.

PROPOSITION:
Let $\{\mathcal{A}_i\}_{i \in I}$ be a collection of σ-algebras in Ω. Then $\bigcap_{i \in I} \mathcal{A}_i$ is a σ-algebra.
It follows that if $\mathcal{S}$ is a set of subsets of Ω, then there exists a **smallest σ-algebra**, which contains $\mathcal{S}$.

DEFINITION: Let (Ω, d) be a metric space and let $\mathcal{S}$ be the set of open balls $B_r(x) = \{y \in \Omega \mid d(x,y) < r\}$. The smallest σ-algebra which contains $\mathcal{S}$ is called the **Borel $\sigma-$ algebra** on Ω.

DEFINITION: Given a sequence of independent events in a probability space. Define $A_\infty := \limsup_{n \to \infty} A_n := \bigcap_{m=1}^{\infty} \bigcup_{n \geq m} A_n$. We have $A_\infty = \{\omega \mid \omega \text{ is in infinitely many } A_i\}$.

BOREL CANTELLI LEMMA: (Monkey typing Shakespeare)

a) If $\sum_n \mathrm{P}[A_n] < \infty$, then $\mathrm{P}[A_\infty] = 0$.
b) If $\sum_n \mathrm{P}[A_n] = \infty$, then $\mathrm{P}[A_\infty] = 1$.

DEFINITION: A random variable X on a probability space $(\Omega, \mathcal{A}, P)$ is called **discrete**, if $\Omega(X)$ is countable or finite. In this case, the expectation of X is defined as

$$\mathrm{E}[X] = \sum_{a \in X(\Omega)} a \cdot \mathrm{P}[X = a]$$

if the sum converges. We denote with $\mathcal{L}^1$ the set of random variables, for which $\mathrm{E}[|X|] < \infty$. Variance, Covariance etc. are defined as in the finite case (keep always an eye on convergence). Note that if $f(X) \in \mathcal{L}^1$, then

$$\mathrm{E}[f(X)] = \sum_{a \in X(\Omega)} f(a) \cdot \mathrm{P}[X = a] .$$

For example if $X^2 \in \mathcal{L}^1$, then

$$\mathrm{Var}[X] = \mathrm{E}[(X - \mathrm{E}[X])^2] = \sum_{a \in X(\Omega)} (a - m)^2 \mathrm{P}[X = a],\ m = \mathrm{E}[X] .$$

EXAMPLES:

$P[X = k] = \frac{\lambda^k}{k!} e^{-\lambda}$	Poisson	$E[X] = \lambda$	Electrons from cathode
$P[X = k] = (1 - p)^{k-1} p$	Geometric	$E[X] = 1/p$	Waiting time for success
$P[X = k] = \zeta(s)^{-1} k^{-s}$	Zeta	$E[X] = \zeta(s+1)/\zeta(s)$	$\zeta(s) = \sum_{n=1}^{\infty} n^{-s}$

7. Week, (Summary of the theory)

DEFINITION: Integration, Expectation: Denote with $\mathcal{S}$ the set of random variables taking finitely many values: Define for $X \in \mathcal{S}$

$$\mathrm{E}[X] := \sum_{a \in X(\Omega)} a \cdot \mathrm{P}[X = a] .$$

Let $\mathcal{L}^1$ be the set of random variables X for which $\sup_{Y \in \mathcal{S}, Y \leq |X|} \mathrm{E}[Y] < \infty$. For $X \in \mathcal{L}^1$ and $X \geq 0$, the **integral** or **expectation** is defined as

$$\mathrm{E}[X] := \sup_{Y \in \mathcal{S}, Y \leq X} \mathrm{E}[Y] .$$

In general, we decompose X into $X = X^+ - X^-$ with $X^\pm \geq 0$ and put $\mathrm{E}[X] = \mathrm{E}[X^+] - \mathrm{E}[X^-]$. We write also $\int_\Omega X \, dP$ for $\mathrm{E}[X]$ since **expectation is integration.** Variance, Covariance etc. are defined as in the finite case: $\mathrm{Var}[X] = \mathrm{E}[(X - E[X])^2], \mathrm{Cov}[X, Y] = \mathrm{E}[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])]$.

DEFINITION: The **Distribution function** of a random variable X is $F(t) = \mathrm{P}[X \leq t]$. **Absolutely continuous random variable**: the **probability density function** $F' = f$ exists. **Discrete random variable:** F is piecewise constant with countably many jump discontinuities. The **expectation, variance** and $\mathrm{E}[g(X)]$ for $g(X) \in \mathcal{L}^1$ is in the continuous case

$$m = \mathrm{E}[X] = \int_{-\infty}^{\infty} x f(x) \, dx, \ \mathrm{Var}[X] = \int_{-\infty}^{\infty} (x - m)^2 f(x) \, dx \ , \mathrm{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) \, dx$$

For discrete random variables this is (repetition)

$$m = \mathrm{E}[X] = \sum_{a \in X(\Omega)} a \mathrm{P}[X = a], \ \mathrm{Var}[X] = \sum_{a \in X(\Omega)} (a - m)^2 \mathrm{P}[X = a], \mathrm{E}[g(X)] = \sum_{a \in X(\Omega)} g(a) \mathrm{P}[X = a]$$

Sometimes, one does not know the distribution of the random variable, then $\mathrm{E}[X], \mathrm{Var}[X]$ and $\mathrm{E}[g(X)]$ have to be computed by integrating (rsp. summing) over Ω.

EXAMPLES OF DISCRETE DISTRIBUTIONS:

Distribution	$\mathrm{P}[x = k] =$	Parameters	Domain	Mean	Variance
Binomial	$\binom{n}{k} p^k (1-p)^{n-k}$	$n \in \mathbf{N}, p \in [0,1]$	$\{0, \ldots, n\}$	np	$np(1-p)$
Poisson	$\frac{\lambda^k}{k!} e^{-\lambda}$	$\lambda > 0$	$\{0, 1, \ldots\}$	λ	λ
Geometric	$(1-p)^{k-1} p$	$p \in (0,1)$	$\{1, 2, \ldots\}$	$1/p$	$1/p^2$

EXAMPLES OF ABSOLUTELY CONTINUOUS DISTRIBUTIONS:

Distribution	Density $f(x) =$	Parameters	Domain	Mean	Variance
Uniform	$1_{[a,b]} \cdot (b-a)^{-1}$	$a < b$	$[a, b]$	$(a+b)/2$	$(b-a)^2/12$
Exponential	$\lambda e^{-\lambda x}$	$\lambda > 0$	$\mathbf{R}^+$	$1/\lambda$	$1/\lambda^2$
Normal	$(2\pi\sigma^2)^{-1/2} \, e^{-\frac{(x-m)^2}{2\sigma^2}}$	$m \in \mathbf{R}, \sigma^2 > 0$	$\mathbf{R}$	m	σ^2
Erlang	$\frac{\lambda^k x^{k-1}}{(k-1)!} e^{-\lambda x}$	$\mathbf{R}^+$	$\lambda > 0, k \in \mathbf{N}$	k/λ	k/λ^2

8. Week, (Summary of the theory)

PROPERTIES OF DISTRIBUTION FUNCTIONS:

$F(t) \in [0,1]$	$\mathrm{P}[a < X \le b] = F(b) - F(a)$
$a \le b \Rightarrow F(a) \le F(b)$	$\lim_{t\to-\infty} F(t) = 0, \lim_{t\to\infty} F(t) = 1$
$\lim_{\epsilon \searrow 0} F(a+\epsilon) = F(a)$	$\lim_{\epsilon \searrow 0} F(a-\epsilon) = F(a) - \mathrm{P}[X = a]$

Every function $F : \mathbf{R} \to \mathbf{R}$ satisfying the above properties belongs to a random variable X: define the probability space $(\Omega, \mathcal{A}, P)$, where $\mathcal{A}$ is the Borel σ-algebra on $\Omega = \mathbf{R}$ and P is defined by $\mathrm{P}[[a,b)] = F(b) - F(a) = \mathrm{P}[X \in [a,b)]$.

DEFINITION: $X = (X_1, X_2, \ldots, X_d)$ is called a **random vector** if X_i are random variables. The **distribution function** of X (also called **joint distribution** of $X_1, X_2, \ldots, X_d$) is defined as

$$F(t_1, \ldots, t_d) = \mathrm{P}[X_1 \le t_1, X_2 \le t_2, \ldots, X_d \le t_d] .$$

The distribution is **continuous** if there exists a function $f : \mathbf{R}^d \to \mathbf{R}^+$ such that

$$F(t) = \int_{-\infty}^{t_1} \int_{-\infty}^{t_2} \cdots \int_{-\infty}^{t_d} f(y_1, y_2, \ldots, y_d) \, dy_d \ldots dy_2 dy_1 .$$

TRANSFORMATIION OF RANDOM VARIABLES:

- Let F be a continuous invertible distribution function. Let X be a random variable which is uniformly distributed in $[0,1]$. Then $Y = F^{-1}(X)$ gives random numbers with distribution F.
- Given a continuous random variable X with density f and a differentiable invertible function $\phi : \mathbf{R} \to \mathbf{R}$ with inverse ψ. The random variable $Y = \phi(X)$ has the density

$$g(t) = f(\psi(t))|\psi'(t)| .$$

- Given a continuous random vector X with density f and a differentiable invertible function $\phi : \mathbf{R}^d \to \mathbf{R}^d$ with inverse ψ. The random variable $Y = \phi(X)$ has the density

$$g(u) = f(\psi(u))|\mathrm{Det}(D\psi(u))| .$$

DEFINITION: The **characteristic function** of X is defined as $\phi_X(t) = \mathrm{E}[e^{itX}]$.
Discrete case: $\phi_X(t) = \sum_{a \in X(\Omega)} e^{ita} \mathrm{P}[X = a]$. Continuous case: $\phi_X(t) = \int_{-\infty}^{\infty} e^{itx} f(x) \, dx$.
CALCULATION OF MOMENTS: $\mathrm{E}[X^k] = (-i)^k \phi_X^{(k)}(0)$. Especially, $\mathrm{E}[X] = -i\phi'_X(0)$.

SUMS OF INDEPNDENT RANDOM VARIABLES: X_i independent with distribution $\phi_i, S = \sum_{i=1}^n X_i$, then $\phi_S(t) = \phi_1(t) \cdot \phi_2(t) \cdots \phi_n(t)$.

THE GAMMA FUNCTION. Some distributions use the Gamma function:

$$\Gamma(n) = \int_0^\infty z^{n-1} e^{-z} \, dz \, .$$

For $n \in \mathbf{N}$, we have $(n-1)!$. Proof. $\Gamma(1) = 1$, $\Gamma(n) = (n-1)\Gamma(n-1)$ by partial integration. Computations like $\Gamma(1/2) = \sqrt{\pi}, \Gamma(3/2) = \sqrt{\pi}/2$ use $\int_{\mathbf{R}} e^{-x^2/2} = \sqrt{2\pi}$.

Distribution	Parameter	Charact. function
Normal	$m \in \mathbf{R}$, $\sigma^2 > 0$	$e^{mit-\sigma^2 t^2/2}$
Standard normal		$e^{-t^2/2}$
Uniform	$[-a, a]$	$\sin(at)/(at)$
Exponential	$\lambda > 0$	$\lambda/(\lambda - it)$
Binomial	$n \in \mathbf{N}, p \in [0, 1]$	$(p + (1-p)e^{it})^n$
Poisson	$\lambda > 0$	$e^{\lambda(e^{it}-1)}$
Geometric	$p \in (0, 1)$	$\frac{pe^{it}}{(1-(1-p)e^{it})}$

9. Week, (Summary of the theory)

CHEBYCHEV-MARKOV INEQUALITY.
Let $h : \mathbf{R}^+ \to \mathbf{R}^+$ be a monoton function and $X \geq 0$ a random variable with $h(X) \in \mathcal{L}^1$. Then for all $c > 0$

$$h(c) \cdot \mathrm{P}[X \geq c] \leq \mathrm{E}[h(X)] .$$

Proof. Take the expectation of $h(c)1_{X \geq c}(\omega) \leq h(X)(\omega)$. Use the monotonicity and linearity of the expectation.
CHEBYCHEV INEQUALITY.
If $X \in \mathcal{L}^2$, then for all $c > 0$

$$\mathrm{P}[|X - \mathrm{E}[X]| \geq c] \leq \frac{\mathrm{Var}[X]}{c^2} .$$

Proof. Apply Chebychev-Markov's inequality to $Y = |X - \mathrm{E}[X]|$ and $h(x) = x^2$.

DEFINITION.
A sequence of random variables X_n **converges in probability** to a random variable X, if for all $\epsilon > 0$,

$$\lim_{n \to \infty} \mathrm{P}[|X_n - X| \geq \epsilon] = 0 .$$

WEAK LAW OF LARGE NUMBERS.
Assume X_i have common expectation $E[X_i] = m$ and satisfy $\sup_n \frac{1}{n} \sum_{i=1}^n \mathrm{Var}[X_i] < \infty$. If X_n are pairwise uncorrelated, then $\sum_{i=1}^n X_i/n \to m$ in probability.
Proof. Since in generl $\mathrm{Var}[X + Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2 \cdot \mathrm{Cov}[X, Y]$ and X_n are pairwise uncorrelated, we have $\mathrm{Var}[X_n + X_m] = \mathrm{Var}[X_n] + \mathrm{Var}[X_m]$ for $n \neq m$ and by induction $\mathrm{Var}[S_n] = \sum_{i=1}^n \mathrm{Var}[X_i]$. Using linearity, we obtain $E[S_n/n] = m$ and

$$\mathrm{Var}[S_n/n] = E[(S_n)^2/n^2] - E[S_n]^2/n^2 = \mathrm{Var}[S_n]/n^2 = \frac{1}{n^2} \sum_{i=1}^n \mathrm{Var}[X_n] \to 0 .$$

With Chebychev's inequality, we obtain

$$P[|S_n/n - m| \geq \epsilon] \leq \frac{\mathrm{Var}[S_n/n]}{\epsilon^2} \to 0 \ (n \to \infty) .$$

IMPORTANT SPECIAL CASE.
If X_i are independent random variables with the same distribution for which the mean m and variance exist both, then $\sum_{i=1}^n X_i/n \to m$ in probability.

EXISTENCE OF INDEPENDENT RANDOM VARIABLES: (don't read this!)
Given a distribution function F, there exists a probability space $(\Omega, \mathcal{A}, P)$ and independent random variables $X_1, X_2, \dots$ which have all the distribution F.
Proof. We know how to construct a single random variable X with distribution F on a probability space $(\mathbf{R}, \mathcal{B}, Q)$. Form the product space

$$(\Omega, \mathcal{A}, P) = (\mathbf{R}^{\mathbf{N}}, \mathcal{B}^{\mathbf{N}}, Q^{\mathbf{N}}) .$$

Ω contains sequences $\omega = (\omega_1, \omega_2, \dots)$ and the probability measure P is defined by

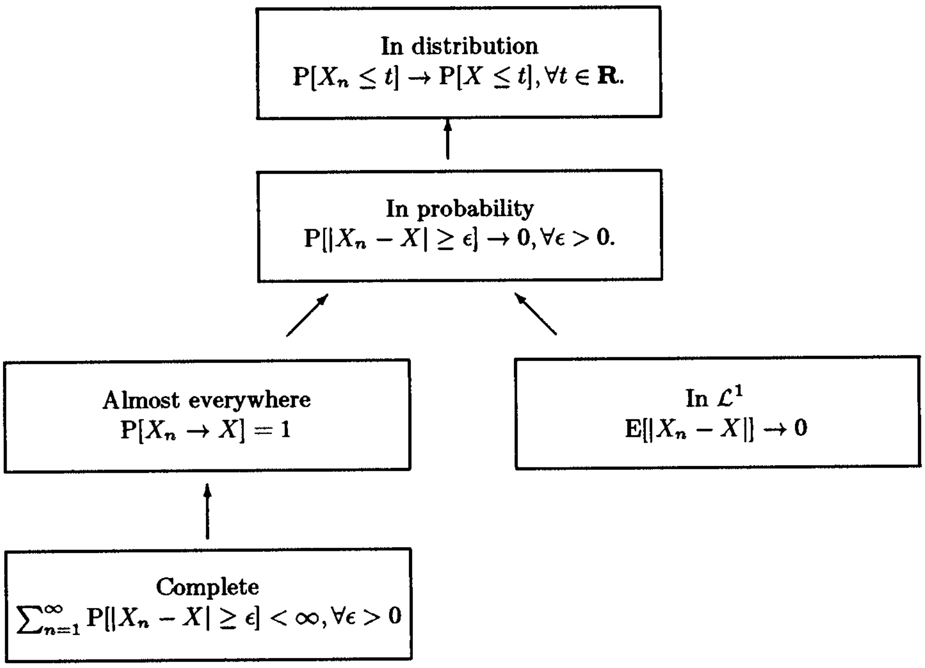
$$P[A_1 \times A_2 \times A_3 \cdots \times A_n \times \mathbf{R} \times \mathbf{R} \times \dots] = Q[A_1] \cdot Q[A_2] \dots Q[A_n] .$$

The σ-algebra $\mathcal{A}$ is the smallest σ-algebra containing all sets of the form $A_1 \times A_2 \times A_3 \cdots \times A_n \times \mathbf{R} \times \mathbf{R} \times \dots$ with $A_i \in \mathcal{B}$. The random variables $X_i(\omega) = \omega_i$ are independent and have all the distribution F.

10. Week, (Summary of the theory)

DEFINITION. A sequence of random variables X_n converges **almost everywhere** to a random variable X, if $\mathrm{P}[X_n \to X, n \to \infty] = 1$.

RELATION BETWEEN CONVERGENCE OF RANDOM VARIABLES: an arrow stands for "implies"



STRONGER WEAK LAW OF LARGE NUMBERS:
Assume X_i have common expectation $E[X_i] = m$ and satisfy $M = \sup_n \mathrm{E}[X^4] < \infty, \sup_n \mathrm{E}[X^2]^2 < \infty$. If X_i are independent, then $\sum_n \mathrm{P}[|S_n/n - m| \geq \epsilon]$ converges for all $\epsilon > 0$.
Proof. Estimation of $\mathrm{E}[X_n^4]$ with Chebychev-Markov's inequality gives $\mathrm{P}[|S_n/n - m| \geq \epsilon] \leq C/n^2$ for some constant C.

STRONG LAW OF LARGE NUMBERS:
Assume X_n are independent random variables with $M = \sup_n \mathrm{E}[X_n^4] < \infty$, $\sup_n \mathrm{E}[X_n^2] < \infty$ with common expectation $\mathrm{E}[X_n] = m$. Then $S_n/n \to m$ almost everywhere.
Proof. Direct consequence of the stronger weak law above since complete convergence implies convergence almost everywhere.

DEFINITION. A sequence of random variables X_n converges **in distribution** to a random variable X, if for all $t \in \mathbf{R}$, $\mathrm{P}[X_n \leq t] \to \mathrm{P}[X \leq t]$ for $n \to \infty$.

CENTRAL LIMIT THEOREM:
Given X_n which are independent with mean m and variance σ^2. Let X be a random variable with standard normal distribution. Then

$$\frac{S_n - nm}{\sigma\sqrt{n}} \to X$$

in distribution, where $S_n = X_1 + X_2 + \ldots + X_n$.
Proof. A calculation shows that the characteristic functions of $S_n^* = (S_n - E[S_n])/\sigma[S_n]$ converge to the characteristic function of X.

Ma2c, O. Knill, **June 1995**

Ma2c Final Exam

Material: Open book, open notes, open homework, no computer.

Time: **4 hours in one sitting.** No credit is given for work done in **overtime.**

Form: **A blue book is required.** Write Ma2c, name, **section number and TA's name** on the blue book. Use whenever possible a **new page** when starting a new problem and indicate on each page, at which problem you are working on there.

Due: Return the blue book to the slot marked Ma2C outside **Sloan 255 by 11:00 AM, Monday, June 12, 1995.** Late papers are not accepted.

Points: There are 8 problems giving in total maximally 100 points. The number of points for each question and subquestion is indicated in each question and subquestions. Write down, how you obtained the solution and which definitions/facts you need from the theory. The problems have motivating stories which help to understand the problem better. It is **recommended but not necessary** to read the information in boxes in order to solve the problem.

1) Topic: **Theorem of Stokes**. Consider the surface $S = r(B)$ defined by $B = [0, 2\pi] \times [0, 2]$ and

$$r : (u, v) \mapsto (v \sin(\alpha)\ \cos(u), v \sin(\alpha)\ \sin(u), v \cos(\alpha)) \ ,$$

where $\alpha \in \mathbf{R}$ is a parameter. Define the vector field $F(x, y, z) = (y, -x, z^3)$.
a) (5 points) Compute the flux of the vector field $\operatorname{curl}(F)$ through the surface S using a surface integral.
b) (5 points) Using the theorem of Stokes, determine the value of the line integral $\int_\gamma F\ ds$, where

$$\gamma : t \in [0, 8\pi] \mapsto (2 \sin(\alpha) \cos(t), 2 \sin(\alpha) \sin(t), 2 \cos(\alpha)) \ .$$

2) Topic: **Theorem of Gauss**. Let U be the region in $\mathbf{R}^3$ which is bounded by the two tori

$$S_1 : (u, v) \mapsto ((8 + 3 \cos(v)) \cos(u), (8 + 3 \cos(v)) \sin(u), 3 \sin(v))$$

$$S_2 : (u, v) \mapsto ((8 + 2 \cos(v)) \cos(u), (8 + 2 \cos(v)) \sin(u), 2 \sin(v))$$

which are both parametrized by $B = [0, 2\pi] \times [0, 2\pi]$.
a) (5 points) Compute with the theorem of Gauss the flux of $F(x, y, z) = (2x + y, 3y + x^3, 5z + y^2)$ through S_1 and do the same also for the surface S_2.
b) (5 points) Compute the flux of F through the boundary δU of U. (As usual, the boundary is oriented in such a way that the normal vector N points outside U at all points of δU.)

3) Topic: **Boolean algebras, Probability measure**. Let $(\Omega, \mathcal{A}, P)$ be a finite probability space, where $\mathcal{A}$ is the set of subsets of Ω and $\mathrm{P}[A] = |A|/|\Omega|$. The following five problems are independent of each other and can therefore be solved in any order.
a) (3 points) Compute $\sum_{A \in \mathcal{A}} \mathrm{P}[A]$.
b) (2 points) Show: $\mathrm{P}[A \cap B] = \mathrm{P}[A \cup B] - \mathrm{P}[A \Delta B]$ for all $A, B \in \mathcal{A}$.
c) (2 points) Compute $\sum_{\omega \in \Omega} \mathrm{P}[\{\omega\}] \log(\mathrm{P}[\{\omega\}])$.
d) (3 points) Fix $B \in \mathcal{A}$. Let $\mathcal{B} = \{A \in \mathcal{A} \mid B \cap A \neq \emptyset \Rightarrow B \subset A\}$. Show that $\mathcal{B}$ is a Boolean algebra.

4) Topic: **Finite probability spaces.** (Generalisation of the in class discussed and illustrated three door problem). You are a guest in a show. You see $n \geq 3$ closed doors. Behind one of the doors is hidden a car and behind the other $n-1$ doors are goats. You can guess one door. The host, who knows, what is behind each door, opens $n-2$ doors with goats so that all theses doors are different from your chosen door. You have now the two possibilities: *no switch* which is to keep the door you had chosen at the beginning (and which is still closed), or *switch* which is to an other of the two still closed doors. In any of the two cases, if you have now the door with a car, you win this car, else you win nothing.
a) (3 points) Model the situation *switch* with a probability space $(\Omega, \mathcal{A}, P)$ and determine the winning event.
b) (3 points) Model the situation *noswitch* with a probability space and determine the winning event.
c) (4 points) Compute the probability of the winning event with switching and the probability of the winning event without switching.

5) Topic: **Conditional probability space.** Given a finite probabiliy space $(\Omega, \mathcal{A}, P)$ and disjoint sets $B_i \in \mathcal{A}$ such that $\bigcup_{i=1}^{n} B_i = \Omega$.
a) (5 points) Show that $\mathrm{P}[A] = \sum_{i=1}^{n} \mathrm{P}[A|B_i] \cdot \mathrm{P}[B_i]$.
b) (5 points) Using a) show that

$$\mathrm{P}[B_j|A] = \frac{\mathrm{P}[A|B_j]\mathrm{P}[B_j]}{\sum_{i=1}^{n} \mathrm{P}[A|B_i]\mathrm{P}[B_i]} .$$

6) Topic: **Random variables.** Given the probability space $(\Omega = \{1,2,3,4\}, \mathcal{A} = \{\emptyset, \{1,2\}, \{3,4\}, P)$, where P is arbitrary. Determine for the following statements, whether they are true or not true and justify your answer:
a) (2 points) $X = 1_{\{\omega \leq 3\}}$ is a random variable.
b) (2 points) $X = 1_{\{\omega \leq 2\}}$ is a random variable.
c) (2 points) $X = 17$ is a random variable.
d) (2 points) For any three random variables X, Y, Z on $(\Omega, \mathcal{A}, P)$, there exist $a, b \in \mathbf{R}$ such that $Z = aX + bY$.
e) (2 points) For any two random variables X, Y on $(\Omega, \mathcal{A}, P)$, there exist $a \in \mathbf{R}$, such that $X = aY$.

7) Topic: **Expectation.** The probability space in roulette is defined by $(\Omega = \{0,1,2,\ldots,36\}, \mathcal{A} = \{A \subset \Omega\}, \mathrm{P}[\{i\}] = 1/37)$. A new house in Las

Vegas offers to play the following option: as usually you bet 1 dollar. If the ball hits a number different from zero which is divisible by 5 then you get back 5 dollars, (you win then $4 = 5 - 1$ dollars). In any other case, you loose your dollar.
a) (2 points) Determine the random variable $X : \Omega \to \mathbf{R}$ which measures your win in one game.
b) (4 points) Determine $\mathrm{E}[X]$, which is the expected win or loss.
c) (4 points) Compute $\mathrm{Var}[X]$ which is the risk of this option.

8) Topic: **Variance, Covariance.** Given random variables X, Y and constants $c, d \in \mathbf{R}$.
a) (2 points) Show that X and $X + c$ have the same variance.
b) (2 points) Show that $\mathrm{Cov}[X, Y] = \mathrm{Cov}[X + c, Y + d]$.
c) (3 points) Assume $\mathrm{Cov}[X, Y] = 1, \mathrm{Var}[X] = \mathrm{Var}[Y] = 1$. Compute $\mathrm{Var}[3X + 5Y]$.
d) (3 points) Assume $\mathrm{Cov}[X, Y] = 0, \mathrm{Var}[X] = \mathrm{Var}[Y] = 1$. Compute $\mathrm{Cov}[5X + 7Y, Y + X]$.

9) Topic: **Independent random variables.** Given the probability space $(\Omega = \{0,1\}^n, \mathcal{A} = \{A \subset \Omega\}, P = Q^n)$, $Q[\{1\}] = p, Q[\{0\}] = 1 - p$. We know that the random variable $X = X_1 + X_2 \dots + X_n$ is Bernoulli distributed: $P[X = k] = \binom{n}{k} p^k q^{n-k}$.
a) (3 points) Compute $\mathrm{E}[X^3]$.
b) (3 points) What is the correlation of X and X^2? (Use the result in a))
c) (2 points) Assume $n \geq 4$. Compute the covariance of the two random variables $X_1 + X_2$ and $X_3 + X_4$.
d) (2 points) Assume $n \geq 2$. Compute the covariance of the two random variables $X_1 + X_2$ and X.

10) Topic: **Independent events and random variables.**
a) (3 points) Prove: A, B are independent events if and only if $X = 1_A$ and $Y = 1_B$ are independent random variables.
b) (3 points) If $X - Y$ and $X + Y$ are independent and $\mathrm{E}[X] = \mathrm{E}[Y] = 0$, then $\mathrm{Var}[X] = \mathrm{Var}[Y]$.
c) (4 points) If $X = 1_A, Y = 1_B$ are uncorrelated, then X, Y are independent.

Ma2c, O. Knill, **May 1995**

Ma2c Midterm Exam (Solutions)

1) Topic: **Theorem of Stokes.** From

$$r : (u,v) \mapsto (v\sin(\alpha)\ \cos(u), v\sin(\alpha)\ \sin(u), v\cos(\alpha)) ,$$

we compute

$$N = r_u \wedge r_v = \begin{pmatrix} -v\sin(\alpha)\sin(u) \\ v\sin(\alpha)\cos(u) \\ 0 \end{pmatrix} \wedge \begin{pmatrix} \sin(\alpha)\cos(u) \\ \sin(\alpha)\cos(u) \\ \cos(\alpha) \end{pmatrix} = \begin{pmatrix} v\sin(\alpha)\cos(u) \\ v\sin(\alpha)\cos(\alpha)\sin(u) \\ -v\sin^2(\alpha) \end{pmatrix} .$$

The vector field $F(x,y,z) = (y,-x,z^3)$ has the curl

$$\text{curl}(F) = (0,0,-2) .$$

The flux of the vector field curl(F) through the surface S using a surface integral is

$$\int\int_B \text{curl}(F)\ dS = \int_0^{2\pi} du \int_0^{2\pi} 2v\sin^2(\alpha)\ dv = 8\pi\sin^2(\alpha) .$$

b) The line integral $\int_\gamma F\ ds$ is -4 times the integral $\int_{\delta S} F\ ds$ which is by Stokes given by $\int\int_B \text{curl}(F)\ dS = 8\pi\sin^2(\alpha)$. The line integral has therefore the value $\boxed{-32\pi\sin^2(\alpha)}$.

There was no point reduction given for direct computations of the lineintegral. It is just more work.

2) Topic: **Theorem of Gauss.** a) Let U_1 be the region bounded by

$$S_1 : (u,v) \mapsto ((8+3\cos(v))\cos(u), (8+3\cos(v))\sin(u), 3\sin(v))$$

and let U_2 be the region bounded by

$$S_2 : (u,v) \mapsto ((8+2\cos(v))\cos(u), (8+2\cos(v))\sin(u), 2\sin(v)) .$$

The divergence of the vector field $F(x,y,z) = (2x+y, 3y+x^3, 5z+y^2)$ is 10. By Gauss, the flux through S_1 is

$$\int\int\int_{U_1} \text{div}(F)\ dV = 10\text{Vol}(U_1) = 10(8\cdot 2\pi)(3^2\pi) = 1440\pi^2 .$$

Analoguous, the flux through S_2 is

$$\int\int\int_{U_2} \text{div}(F)\ dV = 10\text{Vol}(U_2) = 10(8\cdot 2\pi)(2^2\pi) = 640\pi^2 .$$

b) The flux of F through the boundary δU of U is the flux through S_1 minus the flux through S_2 and is so $\boxed{800\pi^2}$.

3) Topic: **Boolean algebras, Probability measure.** Let n denote the number of elements in Ω. There are $\binom{n}{k}$ sets with k elements. The measure of such a set is k/n.

a) $\sum_{A\in\mathcal{A}} \mathrm{P}[A] = \sum_{k=1}^{n} (k/n) \binom{n}{k} = 2^{n-1}$.

We computed the value of the sum as follows:

Define the function

$$f(x) = \sum_{k=0}^{n} \binom{n}{k} x^k = (1+x)^n .$$

Differentiation with respect to x and multipliing with x gives

$$xf'(x) = \sum_{k=1}^{n} k \binom{n}{k} x^k = nx(1+x)^{n-1} .$$

This gives for $x = 1$ the identity

$$\sum_{k=1}^{n} k \binom{n}{k} = n2^{n-1} .$$

b) The sets $A \cap B$ and $A\Delta B$ are disjoint and their union is $A \cup B$. The additivity of the measure P gives

$$\mathrm{P}[A \cap B] + \mathrm{P}[A\Delta B] = \mathrm{P}[A \cup B] .$$

c) We have $\mathrm{P}[\{\omega\}] = 1/n$ for all ω. We get therefore $\sum_{\omega\in\Omega} \mathrm{P}[\{\omega\}] \log(\mathrm{P}[\{\omega\}]) = -\log(n)$. Remark beside: The negative value of this sum is called the **entropy** of the measure.

d) Fix $B \in \mathcal{A}$. Let $\mathcal{B} = \{A \in \mathcal{A} \mid B \cap A \neq \emptyset \Rightarrow B \subset A\}$. To show that $\mathcal{B}$ is a Boolean algebra we have to check three properties:

(i) $\Omega \in \mathcal{B}$ because $B \subset \Omega$ is always true.

(ii) If $A \in \mathcal{B}$ we prove that the complement A^c is also in $\mathcal{B}$: If $A \cap B = \emptyset$, then $B \subset A^c$ which implies $A^c \in \mathcal{B}$. If $A \cap B \neq \emptyset$, then $A^c \cap B \neq \emptyset$ and $A^c \in \mathcal{B}$.

(iii) If $A, C \in \mathcal{B}$, then $A \cup C \in \mathcal{B}$. If $A \cap B \neq \emptyset$ (which implies $B \subset A$) or $C \cap B \neq \emptyset$ (which implies $B \subset C$), then $B \subset A \cup B$ and therefore $A \cup B \in \mathcal{B}$. Assume therefore $A \cap B = \emptyset$ and $C \cap B = \emptyset$. Then $(A \cup C) \cap B = \emptyset$ so that also in this case $A \cup B \in \mathcal{B}$.

4) Topic: **Finite probability spaces.**

a) In the situation *switch*, there are n possibilities, namely the door, you choose at the beginning. In $n-1$ cases, you choose first a goat, call these experiments $goat_1, \dots goat_{n-1}$. Or you choose a car and this experiment is denoted by car. The probability space has therefore n elements $\Omega = \{goat_1, \dots, goat_{n-1}, car\}$. The Boolean algebra is the set of subsets of Ω and each of the experiments has the same probability $P[\{\omega\}] = 1/n$. The winning event is $A = \{goat_1, goat_2, \dots, goat_{n-1}\}$ because in each of these cases, the host will open the $n-2$ doors without car (he has no other choice

since he is not allowed to open the door with the car). If you switch to the other, still closed door, you win the car.
b) In the situation *noswitch*, there are also n possibilities and again $\Omega = \{goat_1, \dots, goat_{n-1}, car\}$, which are the possible choices of your door. Of course, the winning event is $\boxed{A = \{car\}}$. c) In the case of switching, you win with probability $\boxed{(n-1)/n}$ and in the case of noswitching, you win with probability $\boxed{1/n}$.

5) Topic: **Conditional probability space.** a) With disjoint sets $B_i \in \mathcal{A}$ such that $\bigcup_{i=1}^n B_i = \Omega$, every event A can be written as a union of disjoint events $A \cap B_i$. By the additivity property of the probability measure, we have $P[A] = \sum_{i=1}^n P[A \cap B_i]$. By the definition of conditional probability, we have $\mathrm{P}[A|B_i] \cdot \mathrm{P}[B_i] = P[A \cap B_i]$ and this is true also for $P[B_i] = 0$.

$$\sum_{i=1}^n \mathrm{P}[A|B_i] \cdot \mathrm{P}[B_i] = \sum_{i=1}^n P[A \cap B_i] = P[A] .$$

b) We assume $P[A] > 0$ because only then $P[B_j|A]$ is defined. By multipling both sides of the equation with the denominator, we see that we have to show

$$\sum_{i=1}^n \mathrm{P}[A|B_i]\mathrm{P}[B_i]\mathrm{P}[B_j|A] = \mathrm{P}[A|B_j]P[B_j] .$$

By a), the left hand side is $\mathrm{P}[A]\mathrm{P}[B_j|A] = \mathrm{P}[A \cap B_j]$. which is also the value of the right hand side by the definition of the conditional probability.

6) Topic: **Random variables.** The definition of the σ-algebra should be $\mathcal{A} = \{\emptyset, \{1,2\}, \{3,4\}, \Omega\}$. (There was no other interpretation of the question possible which changed the sense of the question).
a) FALSE: $X = 1_{\{\omega \leq 3\}}$ is not a random variable because the set $\{X = 1\} = \{1,2,3\}$ is not in the Boolean algebra.
b) TRUE: $X = 1_{\{\omega \leq 2\}}$ is a random variable because the sets $\{X = 1\} = \{1,2\}$ and $\{X = 0\} = \{3,4\}$ are in $\mathcal{A}$.
c) TRUE: because every constant function is a random variable.
d) FALSE: Take $X, Y = 0$ and $Z \neq 0$.
e) FALSE: Take $X[1] = 1, X[3] = 0$ and $Y[1] = 0, Y[0] = 1$. Then $aY[1] = 0 \neq X[1] = 1$.

The printing mistake (Ω and a paranthesis were missing), turned out to be harmless. Nevertheless, it should not have occured.

7) Topic: **Expectation.** The probability space in roulette is defined by ($\Omega = \{0, 1, 2, \dots, 36\}, \mathcal{A} = \{A \subset \Omega\}, \mathrm{P}[\{i\}] = 1/37$).
a) We have

$$X(\omega) = \begin{cases} 4 & \omega \in \{5,10,15,20,25,30,35\} \\ -1 & \omega \in \{0,1,2,3,4,6,7,8,9,11,12,13, \\ & 14,16,17,18,19,21,22,23,24, \\ & 26,27,28,29,31,32,33,34,36\} \end{cases} .$$

b)

$$\mathrm{E}[X] = 4P[X = 4] + (-1)P[X = -1] = 28/37 - 30/37 = \boxed{-2/37} .$$

c)

$$\mathrm{E}[X^2] = 4^2 P[X=4] + (-1)^2 P[X=-1] = 16 \cdot 7/37 + 30/37 = 142/37\,.$$

$$\mathrm{Var}[X] = E[X^2] - E[X]^2 = \boxed{138/37}\,.$$

8) Topic: **Variance, Covariance.** Given random variables X, Y and constants $c, d \in \mathbf{R}$. In the following calculations, we use the linearity of the expectation and the fact that $E[c] = c$ and $E[d] = d$, since they are constants.
a) $\mathrm{Var}[X] = \mathrm{E}[(X - E[X])^2] = \mathrm{E}[(X + c - (E[X] + c))^2] = \mathrm{E}[(X + c - (E[X + c]))^2] = \mathrm{Var}[X + c]$.
b)

$$\begin{aligned}\mathrm{Cov}[X,Y] &= \mathrm{E}[(X - E[X])(Y - E[Y])] \\ &= \mathrm{E}[(X + c - E[X + c])(Y + d - E[Y + d])] \\ &= \mathrm{Cov}[X + c, Y + d]\,.\end{aligned}$$

c) Assume $\mathrm{Cov}[X, Y] = 1, \mathrm{Var}[X] = \mathrm{Var}[Y] = 1$. We use the basic properties of Var and E to get

$$\begin{aligned}\mathrm{Var}[3X + 5Y] &= \mathrm{Var}[3X] + \mathrm{Var}[5X] + 2\mathrm{Cov}[3X, 5Y] \\ &= 9 \cdot \mathrm{Var}[X] + 25\mathrm{Var}[Y] + 30\mathrm{Cov}[X, Y] \\ &= 9 + 25 + 30 = \boxed{64}\,.\end{aligned}$$

d) Assume $\mathrm{Cov}[X, Y] = 0, \mathrm{Var}[X] = \mathrm{Var}[Y] = 1$.

$$\mathrm{Cov}[5X+7Y, Y+X] = 5\mathrm{Cov}[X, X] + 7\mathrm{Cov}[Y, Y] = 5\mathrm{Var}[X] + 7\mathrm{Var}[Y] = \boxed{12}\,.$$

9) Topic: **Independent random variables.** To compute $\mathrm{E}[X^3], \mathrm{E}[X^4]$, we use the differentiation trick (as seen twice in class for $\mathrm{E}[X], \mathrm{E}[X^2]$): we get by differentiation with respect to p in a point $p + q = 1$

$$\begin{aligned}\sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} &= (p+q)^n = 1 \\ \sum_{k=1}^{n} k \binom{n}{k} p^{k-1} q^{n-k} &= n(p+q)^{n-1} = n \\ \sum_{k=1}^{n} k(k-1) \binom{n}{k} p^{k-2} q^{n-k} &= n(n-1)(p+q)^{n-2} = n(n-1) \\ \sum_{k=1}^{n} k(k-1)(k-2) \binom{n}{k} p^{k-3} q^{n-k} &= n(n-1)(n-2) \\ \sum_{k=1}^{n} k(k-1)(k-2)(k-3) \binom{n}{k} p^{k-4} q^{n-k} &= n(n-1)(n-2)(n-3)\end{aligned}$$

By multipling the second equation with p, the third equation with p^2, the fourth equation with p^3 the fifth equation with p^4 and using $E[X^j] =$

$\sum_{k=0}^{n} k^j \binom{n}{k} p^k q^{n-k}$, the second until fifth equations become

$$\begin{aligned} E[X] &= np \\ E[X^2] - E[X] &= n(n-1)p^2 \\ E[X^3] - 3E[X^2] + 2E[X] &= n(n-1)(n-2)p^3 \\ E[X^4] - 6E[X^3] + 11E[X^2] - 6E[X] &= n(n-1)(n-2)(n-3)p^4 \,. \end{aligned}$$

a) From this, we get

$$\begin{aligned} E[X^2] &= n(n-1)p^2 + np = (n^2-n)p^2 + np \\ E[X^3] &= n(n-1)(n-2)p^3 + 3E[X^2] - 2E[X] = p^3(n^3 - 3n^2 + 2n) + p^2(3n^2 - 3n) + np \\ E[X^4] &= n(n-1)(n-2)(n-3)p^4 + 6E[X^3] - 11E[X^2] - 6E[X] \\ &= (n^4 - 6n^3 + 11n^2 - 6n)p^4 + (-18n^2 + 12n + 6n^3)p^3 + (29n^2 - 29n)p^2 + 11np \,. \end{aligned}$$

b) $\mathrm{Cov}[X, X^2] = \mathrm{E}[X^3] - \mathrm{E}[X]E[X^2] = np(1-p)(1-2p+2np)$. $\sigma[X] = \sqrt{\mathrm{Var}[X]} = \sqrt{npq}$. $\sigma[X^2] = \sqrt{\mathrm{Var}[X^2]} = \sqrt{(\mathrm{E}[X^4] - E[X^2]^2)}$. Together with a), we get

This question was a bit unfortunate because it was possible to get a big mess. You will learn later a even better way to make the computations

$$\mathrm{Corr}[X, X^2] = \frac{\mathrm{Cov}[X, X^2]}{\sigma[X]\sigma[X^2]};.$$

c) $\mathrm{Cov}[X_1+X_2, X_3+X_4] = \boxed{0}$ since X_1+X_2 and X_3+X_4 are $\boxed{\text{independent}}$.

d)

$$\begin{aligned} \mathrm{Cov}[X_1+X_2, X] &= \mathrm{Cov}[X_1+X_2, X_1+X_2] + \mathrm{Cov}[X_1+X_2, X_3 + \ldots + X_n] \\ &= \mathrm{Var}[X_1+X_2] = \boxed{2\text{ p (1-p)}} \end{aligned}$$

because for $n = 2$, the random variable X is $X = X_1 + X_2$, where we know $\mathrm{Var}[X] = 2p(1-p)$.

10) Topic: **Independent events and random variables.**
Note that $A = \{X = 1\}, A^c = \{X = 0\}, B = \{Y = 1\}, B^c = \{Y = 0\}$.
a) • Assume X, Y are independent random variables. Then by theory, $E[XY] = E[X]E[Y]$. Since $E[X] = P[A], E[Y] = P[B]$ and $E[XY] = P[A \cap B]$, this means $P[A \cap B] = P[A]P[B]$ which means by definition that A, B are independent.

• Assume now A, B are independent events. This means $P[A \cap B] = P[A]P[B]$ and so $E[XY] = E[XY]$. With $P[A] = P[X = 1], P[B] = P[Y = 1]$, we get $P[X = 1; Y = 1] = P[X = 1]P[Y = 1]$. We have also to show other cases: $P[X = 0; Y = 1] = P[A^c \cap B] = E[(1-X)Y] = E[Y] - E[XY] = (1 - E[X])E[Y] = P[A^c]P[B] = P[X = 0]P[Y = 1]$.
$P[X = 1; Y = 0] = P[A \cap B^c] = E[X(1-Y)] = E[X] - E[XY] = E[X](1 - E[Y]) = P[A]P[B^c] = P[X = 1]P[Y = 0]$.
$P[X = 0; Y = 0] = P[A^c \cap B^c] = E[(1-X)(1-Y)] = 1 - E[X] - E[Y] - E[X]E[Y] = (1 - E[X])(1 - E[Y]) = P[A^c]P[B^c] = P[X = 0]P[Y = 0]$.

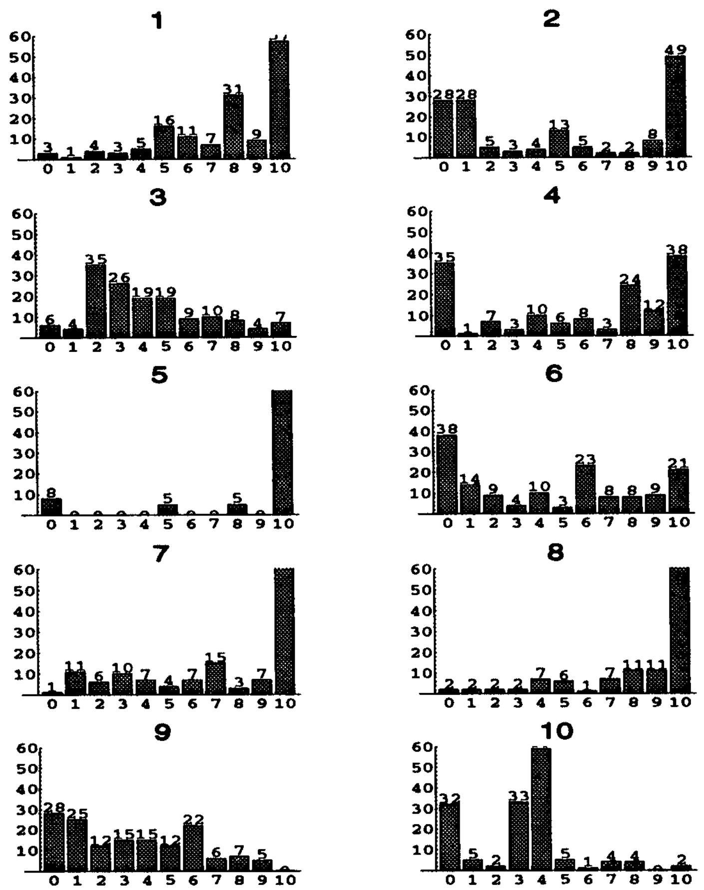
We need to prove two directions. The first direction is easier. In the second case, one has really to deal with all cases

b) If $X - Y$ and $X + Y$ are independent, then $0 = E[(X-Y)(X+Y)] = E[X^2 - Y^2]$. Since $\mathrm{E}[X] = \mathrm{E}[Y] = 0$, this means $E[(X - E[X])^2 - (Y -$

$E[Y]^2)] = 0$ or $\text{Var}[X] = \text{Var}[Y]$.
c) If $X = 1_A, Y = 1_B$ are uncorrelated, then $E[XY] = E[X]E[Y]$ which implies $P[A \cap B] = P[A]P[B]$ which means that A, B are independent and by the result proven in a) that X, Y are independent.

Referring to a) is necessary! It is NOT true in general that $E[XY] = E[X]E[Y]$ implies independence. Here, it is true only because we deal with random variables taking only two values.

REMARK: The large variety of questions and "pencil discrimiation" was an issue of critics. The later will probably be dropped in the final. The former will be kept. Advantages of many questions are that there is a minor risk of being stuck with a crucial question and that a finer and fairer grading is possible since most points are already prediscribed. I expected that you will be able to solve in 4 h about 8 of the 10 questions. Indeed, the average points were 59 points and multiplied with 5/4, this gives an average of 73 points which is reasonable. You find the scaled number of points also on your book. If you have questions/complaints about the grading, address it directly to me. O.K.



Ma2c Midterm Exam

Material: Open book, open notes, open homework, no computer.

Time: **4 hours in one sitting**. No credit is given for work done in **overtime**. Do the topics first with which you feel most comfortable with (since it is not sure that you will have the time to solve all the questions).

Form: A **blue book** is required. **No pencils**. Write Ma2c, name, section number and TA's name on the blue book. Use whenever possible a **new page** when starting a new problem and indicate on each page, at which problem you are working on there.

Due: Return the blue book to the slot marked Ma2C outside **Sloan 255 by 12:00 AM, Monday, May 8, 1995**. Late papers are not accepted.

Points: There are 10 problems, each problem gives up to 10 points and the number of points for each subquestion is indicated in each question. Write down, how you obtained the solution and which definitions/facts you need from the theory.

1) Topic: **Discrete probability spaces.** (12 points)

> We model a physical system which can be in different energy states and which is in contact with a heat reservoir of **inverse temperature** β. If the energies take discrete values $e_1, e_2, \ldots$, then the system will be in the **Bolzman distribution** which is given by $\mathrm{P}_\beta[\{j\}] = e^{-\beta e_j}/Z(\beta)$, where $Z(\beta) = \sum_{j=1}^{\infty} e^{-\beta e_j}$ is the normalization factor called **partition function**. The probabilities $\mathrm{P}_\beta[\{j\}]$ are called the **Bolzmann-factors**. The **energy** of the state j is $e_j = X(j)$. The case $e_j = j$ is essentally the situation of the **quantum mechanical oscillator**. The formula, you will compute in c) is **Planck's formula** for the **total energy** of a quantum mechanical oscillator. In suitable physical units, this is **Planck's blackbody radiation formula**, giving the average energy of the oscillator in dependence of the temperature.
> It is important to note that probability theory has its roots partly in the foundations of thermodynamics but the vocabulary is different Ω=phasespace, $(\Omega, \mathcal{A}, \mathrm{P})$= thermodynamic system, random variable=observable, probability density=thermodynamic state.

Given for every real number $\beta > 0$ the discrete probability space $(\Omega, \mathcal{A}, P_\beta)$ with

$$\Omega = \mathbb{N} = \{0, 1, 2, \ldots\}, \mathcal{A} = \{A \subset \Omega\}, \mathrm{P}_\beta[\{j\}] = \frac{e^{-\beta e_j}}{Z(\beta)} ,$$

where

$$Z(\beta) = \sum_{j=0}^{\infty} e^{-\beta e_j}$$

and the real numbers e_j are such that the sum $Z(\beta)$ is finite. Consider the random variable X on $(\Omega, \mathcal{A}, P_\beta)$ defined by $X(j) = e_j$.

a) (4) Show that

$$\mathrm{E}_\beta[X] = -\frac{\frac{d}{d\beta} Z(\beta)}{Z(\beta)} ,$$

where E_β is the expectation with respect to $(\Omega, \mathcal{A}, P_\beta)$.
b) (4) Compute $Z(\beta)$ in the case, when $e_j = j$ for all $j = 0, 1, \ldots$.
c) (4) Compute, using *a)* and *b)*, the value of $\mathrm{E}_\beta[X]$, in the case $e_j = j$.

2) Topic: **Finite probability spaces, discrete random variables.** (12 points)

> In many tables of physical constants and statistical data, the **leading digit** of the data is not uniformly distributed among the digits (as might naivly be expected). Rather, the lower digits appear much more frequently than the higher ones. **Benford's law** says that the probability that the first significant digit is k is given by $\log_{10}(1 + k^{-1})$. Benford derived this law in 1938 from some statistics he did from twenty different tables. It is today quite evident that Benford manipulated the round off errors to obtain a better fit. Dispite this fraud, the law is called after him. Apropos fraud: since the IRS is considering doing Bedford tests on the data obtained from the taxpayer and to audit the worst fits, one can now alrady read advises (seen in 1995) that a "creative" taxpayer who wants to outfox the IRS should fabricate his datas with first significant digits satisfying Benford's distribution ...

Consider the finite set $\Omega = \{1, 2, \dots, 9\}$ and the Boolean algebra $\mathcal{A} = \{A \subset \Omega\}$.
a) (6) Show that $\mathrm{P} : \mathcal{A} \to [0, 1]$

$$\mathrm{P}[A] = \sum_{k \in A} \log_{10}(\frac{k+1}{k})$$

is a probability measure, where $\log_{10}$ is the logarithm with respect to the base 10.
b) (6) Let X be the random variable, which gives the first digit $X(k) = k$. Compute $\mathrm{E}[X]$. (Simplify the result as much as possible. No numerical evaluation of the result is required).

3) Topic: **Continuous distribution, Arc-Sin distribution.** (12 points)

> A quantum mechanical particle moving freely in a one-dimensional discrete crystal has an energy density in $[-2, 2]$ given by $f(x) = \frac{1}{\pi}(4 - x^2)^{-1/2}$. This is called the **density of states** of the system. The probability distribution function $F(t) = \int_{-\infty}^{t} f(s)\, ds$ is called the **integrated density of states** and accessible to measurements in solid state physics.

Consider the function f which is 0 outside $(-2, 2)$ and $f(x) = 1/(\pi\sqrt{4 - x^2})$ for $x \in (-2, 2)$.

a) (3) Compute $F(t) = \int_{-\infty}^{t} f(s)\, ds$. Hint: see the title.
b) (3) Verify that f is a probability density function.
c) (3) Consider a random variable X which has the distribution function F. Compute $\mathrm{E}[X]$.
d) (3) Compute $\mathrm{Var}[X]$. (Hint: A table (or Mathematica) tells you that $\int x^2/\sqrt{4-x^2}\, dx = 2\arcsin(x/2) - x\sqrt{4-x^2}/2$.)

4) Topic: **General probability spaces, Law of large numbers.** (14 points)

Computing numbers by a random process is called a **Monte Carlo method.** It is often used in numerical investigations done in fields like statistical mechanics or elementary particle physics. For example, to study the question why quarks are confined (they have never been "seen" as simple particles) one uses path integral methods in lattice gauge theory. One goal in this business is to compute the potential of interacting particles or the determination of the masses of the quarks from basic principles. Even to compute expectation values of observables on a small universe consisting of 10^4 points, would require in the simplest case to sum over a probability space with say 2^{40000} elements which is ridiculous. Monte Carlo methods allow nevertheless to make computations.

Buffon's needle problem allows to compute the number π with a Monte Carlo method. The problem is: we are throwing a needle of length $L > 0$ onto a piece of paper on which there are parallel lines of distance 1. What is the probability that the needle will hit a line?
Assume that the angle ϕ of the needle is uniformly distributed in $[0, \pi)$ and that the center M of the needle has distance r from the line next to M and that this distance is uniformly distributed in $[0, 1/2]$. More precicely, the probability space is $(\Omega, \mathcal{A}, \mathrm{P})$, where $\Omega = [0, \pi] \times [0, 1/2]$, $\mathcal{A}$ is the Borel σ-algebra on Ω and P is the Lebesgue measure $\mathrm{P}[(a, b) \times (c, d)] = (b-a)(d-c) \cdot \frac{2}{\pi}$ for all $0 \leq a < b \leq \pi$ and $0 \leq c < d \leq 1/2$.
a) (5) Compute the event $A \subset \Omega$ consisting of all points $\omega = (\phi, r)$ such that the needle of length L with angle ϕ and having its center in distance r from the nearest line intersects a line.
b) (5) Compute the probability $\mathrm{P}[A]$ of this event.
c) (4) What law assures that for almost all sequences of experiments ω,

$$\mathrm{P}[A] = \lim_{n \to \infty} \frac{1}{n} \{\text{number of times, the needle hits a line in n experiments}\} \;?$$

5) Topic: **Borel-Cantelli lemma.** (12 points)

Borel-Cantelli's lemma tells about the probability that an experiment is in infinitely many events. It is often used in more advanced topics of probability theory like in the study of the law of large numbers, the law of iterated logarithm or stochastic processes like the random walker or Brownian motion.

Consider the probability space $(\Omega, \mathcal{A}, P)$, where $\Omega = [0,1]$, $\mathcal{A}$ is the Borel σ-algebra on Ω and P is determined by $\mathrm{P}[[a,b)] = b - a$ for all $0 \leq a < b \leq 1$. Consider the sets $A_n = [0, 1/n)$.
a) (4) Does the $\sum_n \mathrm{P}[A_n]$ converge to a finite value?
b) (4) Determine the set $A_\infty = \bigcap_{m=1}^{\infty} \bigcup_{n \geq m} A_n$.
c) (4) A statement in Borel-Cantelli's lemma reads that $\sum_{n \in \mathbb{N}} \mathrm{P}[A_n] = \infty \Rightarrow \mathrm{P}[A_\infty] = 1$. Does this agree with a) and b)? If not, what is wrong/missing?

6) Topic: **Independent random variables, Variance, sum of independent random variables, characteristic functions.** (12 points)

A central topic in probability theory is the study of sums of independent random variables. An example is the random walk, where we sum up random variables which take two values. Stock markets behave more as random walks with continuous distributions.

Let X_n be independent identically distributed random variables with uniform distribution on the interval $[-1, 1]$. Define $S_n = X_1 + X_2 + \ldots + X_n$.
a) (3) Compute $\mathrm{Var}[S_n]$.
b) (3) Compute the limit $\mathrm{Var}[S_n/n]$ for $n \to \infty$.
c) (3) Compute the limit $\mathrm{Var}[S_n/\sqrt{n}]$ for $n \to \infty$.
d) (3) Compute the characteristic function ϕ_{S_n} for $n \in \mathbb{N}$. Hint: use a result in the theory.

7) Topic: **Continuous distributions, transformation of densities.** (14 points)

The Gamma distribution is a continuous distribution which has been found useful to model things like the time needed for a diagnose and repair of a car engine or to find and repair a bug in a computer program. The Erlang and Exponential distribution are both special cases.

Consider a distribution which has the density with support on $(0, \infty)$ given by

$$f(x) = \frac{\lambda^{\beta} x^{(\beta-1)}}{\Gamma(\beta)} e^{-\lambda x} ,$$

where $\lambda > 0, \beta > 0$ are fixed real parameters. If a random variable is absolutely continuous and has this probability density, then it is called **Gamma**-distributed.

a) (4) Compute the characteristic function $\phi_X(t) = \mathrm{E}[e^{itX}]$ of a random variable X which has the Gamma distribution with the real parameters $\lambda > 0, \beta > 0$.
b) (3) Compute, using a), $\mathrm{E}[X]$.
c) (3) Compute, using a), $\mathrm{Var}[X]$.
d) (4) Compute the probability density $g(x)$ of the random variable $Y = X^3$.

8) Topic: **The weak law of large numbers.** (12 points)

Consider a population of a species (**not** dinosaurs) which is living on a planet suffering many meteor impacts. The probability that a meteor crashes into the planet in a 10 year period is p. In this case, 1/3 of the population is killed. With probability $1-p$ however, there is no meteor impact and the population is happily doubling in these 10 years. In which range of probabilities p does a population grow exponentially?

Let $X_1, X_2, \ldots,$ be a sequence of independent random variables which satisfy $\mathrm{P}_p[X_n = 2/3] = p$, $\mathrm{P}_p[X_n = 2] = 1-p$, where $p \in (0,1)$ is a parameter. Define for each n the random variable $T_n = (X_1 \cdot X_2 \cdots X_n)^{1/n}$.
a) (3) Fix $0 < p < 1$. Show that there exists a constant $c_p > 0$ (depending on p) such that for all $\epsilon > 0$

$$\mathrm{P}_p[|\log(T_n) - \log(c_p)| \geq \epsilon] \rightarrow 0 .$$

b) (3) Compute $\mathrm{E}_p[\log(X_n)]$, where E_p is the expectation with respect to the probability measure P_p.
c) (3) Compute the value of c_p for each p.
d) (3) For which p is $c_p > 1$?

Ma2c Final Exam (Solutions)

1) Topic: **Discrete probability spaces.** (12 points)
a) (4) We compute

$$\frac{d}{d\beta}Z(\beta) = -\sum_{j=0}^{\infty} e_j \cdot e^{-\beta e_j}$$

and so

$$-\frac{\frac{d}{d\beta}Z(\beta)}{Z(\beta)} = \sum_{j=0}^{\infty} e_j \cdot e^{-\beta e_j}/Z(\beta) = \sum_{j=0}^{\infty} X(j)P_\beta[\{j\}] = \mathrm{E}[X] .$$

b) (4) We have

$$Z(\beta) = \sum_{j=0}^{\infty} e^{-\beta j} = \sum_{j=0}^{\infty} (e^{-\beta})^j = \frac{1}{1-e^{-\beta}} .$$

c) (4) From b), we get

$$\frac{d}{d\beta}Z(\beta) = -\frac{e^{-\beta}}{(1-e^{-\beta})^2}$$

and

$$\mathrm{E}_\beta[X] = -\frac{\frac{d}{d\beta}Z(\beta)}{Z(\beta)} = \frac{e^{-\beta}}{1-e^{-\beta}} = \frac{1}{e^\beta - 1} .$$

2) Topic: **Finite probability spaces, discrete random variables.** (12 points)
a) (6) Since $\mathrm{P}[\{j\}] > 0$, we have also $\mathrm{P}[A] \geq 0$ for all $A \in \mathcal{A}$. The additivity is clear from the fact that we know the masses of the atoms, so that P is a measure. What have also to check is that P is normalized:

$$\begin{aligned} \mathrm{P}[\Omega] &= \sum_{k=1}^{9} \log_{10}(\frac{k+1}{k}) = \sum_{k=1}^{9} \log_{10}(k+1) - \sum_{k=0}^{9} \log_{10}(k) \\ &= \log_{10}(10) - \log_{10}(1) = 1 . \end{aligned}$$

If such cancellations as in the above calculation occurs, one says, **the sum is telescoping.**

b) (6) We have

$$\begin{aligned}
E[X] &= \sum_{k=1}^{9} k \cdot P[X=k] = \sum_{k=1}^{9} k \cdot \log_{10}(\frac{k+1}{k}) \\
&= \sum_{k=1}^{9} \log_{10}(k+1)^k - \log_{10} k^k \\
&= \sum_{k=1}^{9} \log_{10}(k+1)^{k+1} - \log_{10}(k+1) - \log_{10} k^k \\
&= \log_{10}(10^{10}) - \log_{10}(1^1) - \log_{10}(10!) = 10 - \log_{10}(10!) = \log_{10}(\frac{10^{10}}{10!}) .
\end{aligned}$$

Again, we had a telescopic sum in the second last step. Numerically, we get (this computation was not required) for the average: $E[X] = 3.44024$.

3) Topic: **Continuous distribution, Arc-Sin distribution.** (12 points)
a) (3) From the hint we get (make a substitution $y = 4x$) for $t \in (-2, 2)$

$$F(t) = \int_{-\infty}^{t} f(x)\,dx = \int_{-2}^{t} f(x)\,dx = \frac{1}{\pi}\arcsin(\frac{x}{2})|_{x=-2}^{t} = \frac{1}{2} + \frac{1}{\pi}\arcsin(\frac{t}{2})$$

and so $F(2) = 1/2 + 1/2 = 1$ and $F(t) = 1$ for $t \geq 2$.
b) (3) We either can use $a)$ to see that $\int_{-2}^{2} f(x)\,dx = 1$ or again compute

$$\int_{-2}^{2} f(x)\,dx = \frac{1}{\pi}\arcsin(\frac{x}{2})|_{-2}^{2} = 1 - 0 = 1 .$$

c) (3)

$$\begin{aligned}
E[X] &= \int_{-2}^{2} x f(x)\,dx = \frac{1}{\pi}\int_{-2}^{2} \frac{x}{\sqrt{4-x^2}}\,dx \\
&= \frac{1}{\pi}\sqrt{4-x^2}|_{x=-2}^{2} = 0
\end{aligned}$$

d) (3) Using the hint

$$\int x^2/\sqrt{4-x^2}\,dx = 2\arcsin(\frac{x}{2}) - x\sqrt{4-x^2}/2$$

we get

$$\begin{aligned}
\mathrm{Var}[X] &= E[X^2] - E[X]^2 = E[X^2] \\
&= \frac{1}{\pi}\int_{-2}^{2} \frac{x^2}{\sqrt{4-x^2}} \\
&= \frac{1}{\pi}\left(2\arcsin(\frac{x}{2}) - x\sqrt{4-x^2}/2\right)|_{-2}^{2} = 2 .
\end{aligned}$$

4) Topic: **General probability spaces, Law of large numbers.** (14 points)
a) (5) Assume the paper (=plane) $\mathbb{R}^2$ is is equiped with the lines $\{y = n\}$. The center $M = (x, y)$ of the needle can be assumed to be in the strip $0 \leq y \leq 1/2$. The needle consists of the line segment between the two points

$$(x \pm \frac{L}{2}\cos(\theta), y \pm \frac{L}{2}\sin(\theta))$$

and this segment intersects a line if $y - L/2\sin(\theta) \leq 0$. So

$$A = \{(\theta, y) \in [0, \pi) \times [0, 1/2] \mid y \leq \frac{L}{2}\sin(\theta)\}.$$

b) (5) We have

$$\mathrm{P}[A] = \frac{2}{\pi}\int_0^{\pi}(L/2)\sin(\theta) = \frac{2L}{\pi}.$$

c) (4) The **strong law of large numbers** assures that for independent random variables $X_n(\omega) = 1_A(\omega)$, S_n/n converges to $\mathrm{E}[X] = \mathrm{P}[A]$.

5) Topic: **Borel-Cantelli lemma.** (12 points)
a) (4) $\sum_n \mathrm{P}[A_n] = \sum_n 1/n$ diverges. The answer is NO.
b) (4) Since $A_{n+1} \subset A_n$, we have

$$A_\infty = \bigcap_{m=1}^{\infty}\bigcup_{n \geq m} A_n = \bigcap_{m=1}^{\infty}[0, 1/n) = \{0\}.$$

c) (4) If A_n would be independent and $\sum_{n \in \mathbb{N}} \mathrm{P}[A_n] = \infty$, then $\mathrm{P}[A_\infty] = 1$. Siunce $\mathrm{P}[A_\infty] = 0$, we conclude that the A_n are not independent.

6) Topic: **Independent random variables, Variance, sum of independent random variables, characteristic functions.** (12 points)
a) (3)

$$\mathrm{Var}[S_n] = \sum_{i=1}^{n}\mathrm{Var}[X_i] = n\int_{-1}^{1} x^2/2\, dx = \frac{n}{3}.$$

b) (3)

$$\mathrm{Var}[S_n/n] = \mathrm{Var}[S_n]/n^2 = 2/(3n) \to 0.$$

c) (3)

$$\mathrm{Var}[S_n/\sqrt{n}] = \mathrm{Var}[S_n]/n = 1/3.$$

d) (3)

$$\phi_{S_n}(t) = \prod_{i=1}^{n}\phi_{X_k}(t) = (\frac{\sin(t)}{t})^n.$$

7) Topic: **Continuous distributions, transformation of densities.** (14 points)

$$f(x) = \frac{\lambda^\beta x^{(\beta-1)}}{\Gamma(\beta)} e^{-\lambda x} ,$$

where $\lambda > 0, \beta > 0$ are fixed real parameters. If a random variable is absolutely continuous and has this probability density, then it is called **Gamma-distributed.**

a) (4) The computations of a) to c) are identical to the one done in homework Week 8 question 2), where β was assumed to take only integer values. We do it again. In order to compute

$$\mathrm{E}[e^{itX}] = \int_0^\infty e^{itx} \frac{\lambda^\beta \Gamma(\beta) x^{(\beta-1)}}{\Gamma(\beta)} e^{-\lambda x} \, dx = \frac{\lambda^\beta}{\Gamma(\beta)} \int_0^\infty x^{(\beta-1)} e^{-(\lambda - it)x} \, dx,$$

we make a change of variables $z = -(\lambda - it)x$ and get

$$\phi(t) = \frac{\lambda^\beta}{(\lambda - it)^\beta} \int_0^\infty z^{\beta-1} e^{-z} \, dz = \frac{\lambda^\beta}{(\lambda - it)^\beta} .$$

b) (3) $\mathrm{E}[X] = -i \cdot \phi'(0) = \frac{\beta\lambda^\beta}{\lambda^{\beta+1}} = \frac{\beta}{\lambda}$.
c) (3) We have $\phi''(0) = -\beta(\beta-1)/\lambda^2$ and so

$$\mathrm{Var}[X] = \mathrm{E}[X^2] - \mathrm{E}[X]^2 = -\phi''(0) + \frac{\beta^2}{\lambda^2} = \frac{\beta - \beta^2}{\lambda^2} + \frac{\beta^2}{\lambda^2} = \frac{\beta}{\lambda^2} .$$

d) (4) Take $\phi(x) = x^3$, $\psi(x) = x^{1/3}$. Then, by theory, the density of the random variable $Y = \phi(X)$ is

$$f(\psi(x)) \cdot |\psi'(x)| = \frac{1}{3x^{2/3}} f(x^{1/3}) .$$

8) Topic: **The weak law of large numbers.** (12 points)
a) (3) Fix $0 < p < 1$. Take $\log(c_p) = \mathrm{E}[\log(X_i)] = m$. The random variables $Y_i = \log(X_i)$ are independent too and $m = \mathrm{E}[Y_i]$. We have with the weak law of large numbers

$$\mathrm{P}_p[|\log(T_n) - \log(c_p)| \geq \epsilon] = \mathrm{P}_p[|\sum_{k=1}^n Y - m]| \geq \epsilon] \to 0 .$$

b) (3)

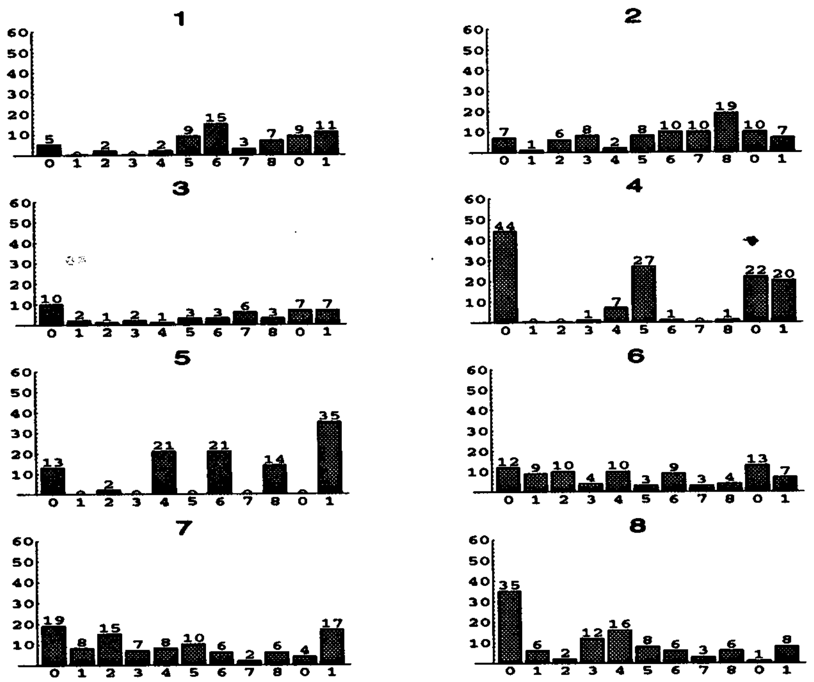
$$\begin{aligned} \mathrm{E}_p[\log(X_n)] &= \log(\frac{2}{3}) \mathrm{P}_p[X_n = 2/3] + \log(2) \mathrm{P}_p[X_n = 2] \\ &= \log(\frac{2}{3}) p + \log(2)(1-p) = \log(2) - p \log(3) . \end{aligned}$$

c) (3) $c_p = e^{\log(2) - p\log(3)} = \frac{2}{3^p}$.
d) (3) $2/3^p > 1$ if and only if $2 > 3^p$ and so $p < \frac{\log(2)}{3}$.

REMARK: The point average of this final is 61.5105. In order to get a similar average as in the midterm (average 74.0385), the number of points in the final is multplied with a factor 6/5 for the computation of the grade. The scaled average of the final is then 73.8126. The scaled number of points of the final is not displayed on your blue book. The homework average was 55.2261 points which gives, when multiplied with 10/7, an average of 78.8944, which is good and does not need an adjustement. The number of points for your grade will be computed (by Mathematica) as follows:

$$\textbf{Points} = \frac{3}{10} \cdot \textbf{Midterm Points} + \frac{3}{10} \cdot \frac{10}{7} \cdot \textbf{Homework Points} + \frac{4}{10} \cdot \frac{6}{5} \cdot \textbf{Final Points} .$$

If you have questions/complaints about the grading, address it directly to me (before friday). O.K.



Ma2c, O. Knill, **April 1995**

1. Week, (Summary of the theory)

NOTATION:

$r = (x, y, z)$ **point in** $\mathbf{R}^3$

$C = \alpha(I)$ **curve,** $\alpha : t \in I = [a, b] \to \alpha(t) \in \mathbf{R}^3$

$S = r(B)$ **surface,** $(u, v) \in B \mapsto r(u, v) \in \mathbf{R}^3$

U **region in** $\mathbf{R}^3$.

δC, δS, δU **boundaries of curve, surface, region**

$F = (P, Q, R)$ **vector field ,** f **scalar field**

DIFFERENTIAL OPERATORS:

$\partial_x f = \frac{\partial}{\partial x} f = f_x$ **(Partial derivative)**

$\nabla = (\partial_x, \partial_y, \partial_z)$ **(Nabla)**

$\text{grad}F = \nabla F = (\frac{\partial F}{\partial x}, \frac{\partial F}{\partial y}, \frac{\partial F}{\partial z})$

$$\text{curl}F = \nabla \wedge F = \begin{pmatrix} \partial_x \\ \partial_y \\ \partial_z \end{pmatrix} \wedge \begin{pmatrix} P \\ Q \\ R \end{pmatrix} = \begin{pmatrix} R_y - Q_z \\ P_z - R_x \\ Q_x - P_y \end{pmatrix}.$$

$\text{div}F = \nabla \cdot F = (\partial_x, \partial_y, \partial_z) \cdot (P, Q, R) = P_x + Q_y + R_z$.

IDENTITIES:

$\text{div curl}F = \nabla \cdot (\nabla \wedge F) = 0$

$\text{curl grad}f = \nabla \wedge \nabla f = 0$.

GEOMETRY OF SURFACE:

$S = r(B)$ surface.

r_u, r_v **tangent vectors**

$N = r_u \wedge r_v$ **normal vector**

$n = N/||N||$ **unit normal vector.**

DEFINITIONS:

line integral
$\int_C F \cdot ds = \int_a^b F(\alpha(t)) \cdot \alpha'(t)\, dt$

surface integral
$\int\int_S F \cdot dS = \int\int_B F(r(u, v)) \cdot N(r(u, v))\, dudv$

volume integral
$\int\int\int_U f\, dV = \int\int\int_U f(x, y, z)\, dxdydz$

FUNDAMENTAL THEOREM OF CALCULUS:

$\int_C \text{grad}f \cdot ds = f(\alpha(a)) - f(\alpha(b)) = \int_{\delta C} f$

THEOREM OF STOKES:

$\int\int_S \text{curl}F \cdot dS = \int_{\delta S} F \cdot ds$

THEOREM OF GAUSS:

$\int\int\int_U \text{div}F\, dV = \int\int_{\delta U} F\, dS$

2. Week, (Summary of the theory)

DEFINITION: Boolean algebra

Ω be a finite set. A set $\mathcal{A}$ of subsets of Ω is a **Boolean algebra**, if

$\Omega \in \mathcal{A}$,
$A \in \mathcal{A} \to A^c \in \mathcal{A}$,
$A, B \in \mathcal{A} \Rightarrow A \cup B \in \mathcal{A}$.

PROPERTIES: A Boolean algebra $(\Omega, \mathcal{A})$ is closed under all set theoretical operations: $A, B \in \mathcal{A}$, then

$\emptyset \in \mathcal{A}$ $\quad A \cap B \in \mathcal{A}$.
$A \setminus B \in \mathcal{A}$ $\quad A \Delta B \in \mathcal{A}$.

DEFINITION: Probability measure

A function $P : \mathcal{A} \to \mathbf{R}$ is a **probability measure** if

$\mathrm{P}[A] \geq 0$, (nonnegativity)
$\mathrm{P}[\Omega] = 1$, (normalisation)
$\mathrm{P}[\bigcup_{i=1}^n A_i = \sum_{i=1}^n \mathrm{P}[A_i], if A_i \cap A_j = \emptyset$, all i, j, (additivity)

DEFINITION: Finite probability space

We say $(\Omega, \mathcal{A}, P)$ is a finite probability space if $\mathcal{A}$ is a Boolean algebra on the finite set Ω and P is a probability measure on $(\Omega, \mathcal{A})$.

PROPERTIES:
$A \subset B \Rightarrow \mathrm{P}[A] \leq \mathrm{P}[B]$.
$\mathrm{P}[A^c] = 1 - \mathrm{P}[A]$
$\mathrm{P}[\emptyset] = 0$

SWITCH ON, SWITCH OFF formula:

$$\mathrm{P}[\bigcup_{i=1}^n A_i] = \sum_{k=1}^n (-1)^{k-1} \sum_{1 \leq i_1 < i_2 < \ldots < i_k \leq n} \mathrm{P}[A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}]$$

CONSTRUCTION OF NEW PROBABILITY SPACES:

Change of algebra: $(\Omega, \mathcal{B}, Q)$ is a probability space, where $\mathcal{B} \subset \mathcal{A}$ is a Boolean algebra, Q restriction of P to $\mathcal{B}$.
Product space:
$(\Omega, \mathcal{A}, P) = (\Omega_1, \mathcal{A}_1, P_1) \times (\Omega_2, \mathcal{A}_2, P_2)$, where $\Omega = \Omega_1 \times \Omega_2$, $\mathcal{A}$ is the smallest algebra containing $\mathcal{A}_1 \times \mathcal{A}_2 = \{A_1 \times A_2 \mid A_i \in \mathcal{A}_i\}$ and $\mathrm{P}[(A_1 \times A_2)] = P_1[A_1] \cdot P_2[A_2]$.
Conditional probability space:
$(B, \mathcal{A} \cap B, \mathrm{P}[\cdot|B])$, if $\Pr[B] > 0$ and $\mathcal{A} \cap A = \{A \cap B \mid A \in \mathcal{A}\}$

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

3. Week, (Summary of the theory)

DEFINITION:
$(\Omega, \mathcal{A}, P)$ finite probability space.
$A, B \in \mathcal{A}$ are **independent** if and only if

$$\mathrm{P}[A \cap B] = \mathrm{P}[A] \cdot \mathrm{P}[B] .$$

A finite set $\{A_i\}_{i \in I}$ of events is called **independent** if and only if for all $J \subset I$

$$\mathrm{P}[\bigcap_{i \in J} A_i] = \prod_{i \in J} \mathrm{P}[A_i] .$$

PROPERTIES:
$A, B \in \mathcal{A}$ are independent, if and only if either $\mathrm{P}[B] = 0$ or $\mathrm{P}[A|B] = \mathrm{P}[A]$.

$(\Omega, \mathcal{A}, P) = (\Delta, \mathcal{B}, Q)^n$ (product space). Given $B_i \in \mathcal{B}$ then

$$A_i = \{\omega = (\omega_1, \dots, \omega_n) \mid \omega_i \in B_i\}$$

are all independent.

DEFINITION: A **random variable** on a finite probability space $(\Omega, \mathcal{A}, P)$ is a map $X : \Omega \to \mathbf{R}$ such that for all $a \in \mathbf{R}$, we have $\{X = a\} \in \mathcal{A}$.

DEFINITION: The **expectation** of a random variable X is defined as

$$\mathrm{E}[X] = \sum_{a \in X(\Omega)} a \cdot \mathrm{P}[X = a] = \sum_{A \in \mathcal{A}, A \text{ atom}} X(A) \cdot \mathrm{P}[A] ,$$

where an **atom** is a set in $\mathcal{A}$ so that $B \subset A, B \in \mathcal{A} \Rightarrow B = A$ or $B = \emptyset$. If $\mathcal{A} = \{A \subset \Omega\}$, then the atoms are all of the form $\{\omega\}$ and

$$\mathrm{E}[X] = \sum_{\omega \in \Omega} X(\omega) \mathrm{P}[\{\omega\}] .$$

By the definition of a random variable, X must be constant on each atom A and $X(A)$ is defined as the common value, X takes on A. The two expressions for $\mathrm{E}[X]$ in the box to the left are seen to be the same using $a = X(A)$ and $\mathrm{P}[X = a] = \sum_{A \text{ atom } X(A) = a} \mathrm{P}[A]$.

PROPERTIES OF EXPECTATION: For random variables X, Y and $\lambda \in \mathbf{R}$

$\mathrm{E}[X + Y] = \mathrm{E}[X] + \mathrm{E}[Y]$	$\mathrm{E}[\lambda X] = \lambda \mathrm{E}[X]$
$X \leq Y \Rightarrow \mathrm{E}[X] \leq \mathrm{E}[Y]$	$\mathrm{E}[X^2] = 0 \Leftrightarrow X = 0$
$\mathrm{E}[X] = c$ if $X(\omega) = c$ is constant	$\mathrm{E}[X - \mathrm{E}[X]] = 0 .$

PROOF OF THE ABOVE PROPERTIES:

$\mathrm{E}[X + Y] = \sum_{A \text{ atom}} (X + Y)(A) \cdot \mathrm{P}[A] = \sum_{A \text{ atom}} (X(A) + Y(A)) \cdot \mathrm{P}[A] = \mathrm{E}[X] + \mathrm{E}[Y]$
$\mathrm{E}[\lambda X] = \sum_{A \text{ atom}} (\lambda X)(A) \mathrm{P}[A] = \lambda \sum_{A \text{ atom}} X(A) \mathrm{P}[A] = \lambda \mathrm{E}[X]$
$X \leq Y \Rightarrow X(A) \leq Y(A)$, for all atoms A and $\mathrm{E}[X] \leq \mathrm{E}[Y]$
$\mathrm{E}[X^2] = 0 \Leftrightarrow X^2(A) = 0$ for all atoms A $\Leftrightarrow X = 0$
$X(\omega) = c$ is constant $\Rightarrow \mathrm{E}[X] = c \cdot \mathrm{P}[X = c] = c \cdot 1 = c$
$\mathrm{E}[X - \mathrm{E}[X]] = \mathrm{E}[X] - \mathrm{E}[\mathrm{E}[X]] = \mathrm{E}[X] - \mathrm{E}[X] = 0$

4. Week, (Summary of the theory)

DEFINITION: $(\Omega, \mathcal{A}, P)$ probability space, X, Y random variables.

Variance

$$\text{Var}[X] = \text{E}[\,(X - \text{E}[X])^2\,]\,.$$

Standard deviation

$$\sigma[X] = \sqrt{\text{Var}[X]}\,.$$

Covariance

$$\text{Cov}[X,Y] = \text{E}[(X - \text{E}[X]) \cdot (Y - \text{E}[Y])]\,.$$

Correlation of $\text{Var}[X] \neq 0, \text{Var}[Y] \neq 0$

$$\text{Corr}[X,Y] = \frac{\text{Cov}[X,Y]}{\sigma[X]\sigma[Y]}\,.$$

$\text{Corr}[X,Y] = 0$: **uncorrelated** X and Y.

PROPERTIES of VAR, COV, and CORR:

$\text{Var}[X] \geq 0$.
$\text{Var}[X] = \text{E}[X^2] - \text{E}[X]^2$.
$\text{Var}[\lambda X] = \lambda^2 \text{Var}[X]$.

$\text{Cov}[X,Y] = \text{E}[XY] - \text{E}[X]\text{E}[Y]$.
$\text{Cov}[X,Y] \leq \sigma[X]\sigma[Y]$ (Schwarz inequality).

$-1 \leq \text{Corr}[X,Y] \leq 1$.
$\text{Corr}[X,Y] = 1$ if $X - \text{E}[X] = Y - \text{E}[Y]$
$\text{Corr}[X,Y] = -1$ if $X - \text{E}[X] = -(Y - \text{E}[Y])$.

BERNOULLI DISTRIBUTED RANDOM VARIABLES: $(\Omega = \{0,1\}^n, \mathcal{A}, P = Q^n)$, where $Q[\{1\}] = p, Q[\{0\}] = q = 1 - p$.

$$X(\omega) = \sum_{i=1}^n \omega_i$$

$$P[X = k] = \binom{n}{k} p^k q^{n-k}$$

$$E[X] = \sum_{k=1}^n k \binom{n}{k} p^k q^{n-k} = pn$$

$$\text{Var}[X] = \sum_{k=1}^n k^2 \binom{n}{k} p^k q^{n-k} - E[X]^2 = npq$$

DEFINITION: X, Y are **independent** if for all $a, b \in \mathbf{R}$

$$P[X = a; Y = b] = P[X = a] \cdot P[Y = b]\,.$$

A finite collection $\{X_i\}_{i \in I}$ of random variables are **independent**, if for all $J \subset I$ and $a_i \in \mathbf{R}$

$$P[X_i = a_i, i \in J] = \prod_{i \in J} P[X_i = a_i]\,.$$

PROPERTIES:

- If X and Y are independent, then $\text{E}[X \cdot Y] = \text{E}[X] \cdot \text{E}[Y]$.
- If X_i is a set of independent random variables, then $\text{E}[\prod_{i=1}^n X_i] = \prod_{i=1}^n \text{E}[X_i]$.
- If X, Y are independent then $\text{Cov}[X,Y] = 0$.
- A constant random variable is independent to any other random variable.

DEFINITION: The **regression line** of two random variables X, Y is defined as $y = ax + b$, where

$$a = \frac{\text{Cov}[X,Y]}{\text{Var}[X]}, b = \text{E}[Y] - a\text{E}[X]\,.$$

PROPERTY: Given $X, \text{Cov}[X,Y], \text{E}[Y]$, and the regression line $y = ax + b$ of X, Y. The random variable $\tilde{Y} = aX + b$ minimizes $\text{Var}[Y - \tilde{Y}]$ under the constraint $\text{E}[Y] = \text{E}[\tilde{Y}]$ and is the best guess for Y, when knowing only $\text{E}[Y]$ and $\text{Cov}[X,Y]$. We check $\text{Cov}[X,Y] = \text{Cov}[X,\tilde{Y}]$.

5. Week, (Summary of the theory)

DEFINITION: Important example: **One dimensional random walk**
$(\Omega = \{-1,1\}^N = \{(\omega = (\omega_1 < \dots, \omega_N) \mid \omega_i \in \{-1,1\}\}, \mathcal{A} = \{A \subset \Omega\}, \mathrm{P}[A] = |A|/|\Omega|)$.
The random variables $X_k(\omega) = \omega_k$ define the k'th step. The random variables $S_n = \sum_{k=1}^n X_k(\omega)$ describe the location of the random walker (drunken sailor) at time n. If X_k is the win or loss in a game at time k, then S_n is the total win or loss up to time n. Ω is the set of all possible trajectories up to time N.

PROPERTIES Random walk:

a) $I \subset \{1, \dots, N\}$, $x_i \in \{-1,1\}$

$$\mathrm{P}[X_i = x_i, i \in I] = 2^{-|I|}.$$

b) $\mathrm{E}[X_k] = 0$,
c) $\mathrm{E}[S_k] = 0$.
d) $n + x$ even, $\mathrm{P}[X_n = x] = 2^{-n} \binom{n}{\frac{n+x}{2}}$.
$n + x$ odd, $\mathrm{P}[X_n = x] = 0$.

DEFINITION: A **gambling system** attached to the random walk is sequence of random variables V_k such that every event $\{V_n = c\}$ is a union of sets of the form $\{\omega_1 = x_1, \dots, \omega_{n-1} = x_{n-1}\}$.
Let V_k be a gambling system, then

$$S_n^V = \sum_{i=1}^n V_k X_k$$

is the **total winnings** with this system.

PROPERTY of gambling systems:
You can't beat the system: $\mathrm{E}[S_N^V] = 0$.

DEFINITION: Cardinality.
$f : A \to B$ is $1:1$ or injective: $x \neq y \Rightarrow f(x) \neq f(y) \forall x, y \in A$.
$f : A \to B$ is **onto** or surjective): $f(A) = B$.
f is **bijectiv** $\Leftrightarrow$ f is 1:1 and onto.
A, B are called **equivalent**, if there exists a bijection $f : A \to B$.
A equivalent to $\mathbb{N}$: **countable infinite.**
A equivalent to finite set: **finite.**
A neither finite nor countable infinite: **uncountable.**

DEFINITION:

A **σ-algebra** on Ω is a set $\mathcal{A}$ of subsets of Ω satisfying

- $\Omega \in \mathcal{A}$,
- $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$,
- $\{A_1, A_2, \dots\} \subset \mathcal{A}$ countable $\Rightarrow \bigcup_{i=1}^\infty A_i \in \mathcal{A}$.

$P : \mathcal{A} \to \mathbb{R}$ is a **probability measure** if

- $\mathrm{P}[A] \geq 0$, (nonnegativity)
- $\mathrm{P}[\Omega] = 1$, (normalisation)
- $\{A_1, A_2, \dots\}$ countable set of disjoint sets $\Rightarrow$ $\mathrm{P}[\bigcup_{i=1}^\infty A_i] = \sum_{i=1}^\infty \mathrm{P}[A_i]$ (**σ-additivity**)

A **probability space** $(\Omega, \mathcal{A}, P)$ sonsists of a set Ω, a σ-algebra $\mathcal{A}$ on Ω and a probability measure P on $\mathcal{A}$.

If Ω is **finite**, the probability space is called a finite probability space. If Ω is countable, it is called **discrete**.

6. Week, (Summary of the theory)

DEFINITION: A function $d : \Omega \times \Omega \to \mathbf{R}$ is called a **metric** if

(i) $d(x,y) \geq 0$,
(ii) $d(x,z) \leq d(x,y) + d(y,z)$
(iii) $d(x,y) = 0 \Leftrightarrow x = y$

The pair (Ω, d) where Ω is a set and d is a metric is called a **metric space**. Examples. $(\mathbf{R}^n, d(x-y) = ||x-y||$, $(\{0,1\}^{\mathbf{N}}, d(x,y) = \sum_{i=1}^{\infty} |x_i - y_i| 2^{-i})$.

PROPOSITION:
Let $\{\mathcal{A}_i\}_{i\in I}$ be a collection of σ-algebras in Ω. Then $\bigcap_{i\in I} \mathcal{A}_i$ is a σ-algebra.
It follows that if $\mathcal{S}$ is a set of subsets of Ω, then there exists a **smallest σ-algebra**, which contains $\mathcal{S}$.

DEFINITION: Let (Ω, d) be a metric space and let $\mathcal{S}$ be the set of open balls $B_r(x) = \{y \in \Omega \mid d(x,y) < r\}$. The smallest σ-algebra which contains $\mathcal{S}$ is called the **Borel $\sigma-$ algebra** on Ω.

DEFINITION: Given a sequence of independent events in a probability space. Define $A_\infty := \limsup_{n\to\infty} A_n := \bigcap_{m=1}^{\infty} \bigcup_{n\geq m} A_n$. We have $A_\infty = \{\omega \mid \omega \text{ is in infinitely many } A_i\}$.

BOREL CANTELLI LEMMA: (Monkey typing Shakespeare)

a) If $\sum_n \mathrm{P}[A_n] < \infty$, then $\mathrm{P}[A_\infty] = 0$.
b) If $\sum_n \mathrm{P}[A_n] = \infty$, then $\mathrm{P}[A_\infty] = 1$.

DEFINITION: A random variable X on a probability space $(\Omega, \mathcal{A}, P)$ is called **discrete**, if $\Omega(X)$ is countable or finite. In this case, the expectation of X is defined as

$$\mathrm{E}[X] = \sum_{a \in X(\Omega)} a \cdot \mathrm{P}[X = a]$$

if the sum converges. We denote with $\mathcal{L}^1$ the set of random variables, for which $\mathrm{E}[|X|] < \infty$. Variance, Covariance etc. are defined as in the finite case (keep always an eye on convergence). Note that if $f(X) \in \mathcal{L}^1$, then

$$\mathrm{E}[f(X)] = \sum_{a \in X(\Omega)} f(a) \cdot \mathrm{P}[X = a] \,.$$

For example if $X^2 \in \mathcal{L}^1$, then

$$\mathrm{Var}[X] = \mathrm{E}[(X - \mathrm{E}[X])^2] = \sum_{a \in X(\Omega)} (a - m)^2 \mathrm{P}[X = a], \; m = \mathrm{E}[X] \,.$$

EXAMPLES:

$\mathrm{P}[X = k] = \frac{\lambda^k}{k!} e^{-\lambda}$	Poisson	$\mathrm{E}[X] = \lambda$	Electrons from cathode
$\mathrm{P}[X = k] = (1-p)^{k-1} p$	Geometric	$\mathrm{E}[X] = 1/p$	Waiting time for success
$\mathrm{P}[X = k] = \zeta(s)^{-1} k^{-s}$	Zeta	$\mathrm{E}[X] = \zeta(s+1)/\zeta(s)$	$\zeta(s) = \sum_{n=1}^{\infty} n^{-s}$

7. Week, (Summary of the theory)

DEFINITION: Integration, Expectation: Denote with $\mathcal{S}$ the set of random variables taking finitely many values: Define for $X \in \mathcal{S}$

$$\mathrm{E}[X] := \sum_{a \in X(\Omega)} a \cdot \mathrm{P}[X = a] .$$

Let $\mathcal{L}^1$ be the set of random variables X for which $\sup_{Y \in \mathcal{S}, Y \leq |X|} \mathrm{E}[Y] < \infty$. For $X \in \mathcal{L}^1$ and $X \geq 0$, the **integral or expectation** is defined as

$$\mathrm{E}[X] := \sup_{Y \in \mathcal{S}, Y \leq X} \mathrm{E}[Y] .$$

In general, we decompose X into $X = X^+ - X^-$ with $X^\pm \geq 0$ and put $\mathrm{E}[X] = \mathrm{E}[X^+] - \mathrm{E}[X^-]$. We write also $\int_\Omega X \, dP$ for $\mathrm{E}[X]$ since **expectation is integration**. Variance, Covariance etc. are defined as in the finite case: $\mathrm{Var}[X] = \mathrm{E}[(X - E[X])^2], \mathrm{Cov}[X, Y] = \mathrm{E}[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])]$.

DEFINITION: The **Distribution function** of a random variable X is $F(t) = \mathrm{P}[X \leq t]$. **Absolutely continuous random variable:** the **probability density function** $F' = f$ exists. **Discrete random variable:** F is piecewise constant with countably many jump discontinuities. The **expectation, variance** and $\mathrm{E}[g(X)]$ for $g(X) \in \mathcal{L}^1$ is in the continuous case

$$m = \mathrm{E}[X] = \int_{-\infty}^{\infty} x f(x) \, dx, \ \mathrm{Var}[X] = \int_{-\infty}^{\infty} (x - m)^2 f(x) \, dx \, , \mathrm{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) \, dx$$

For discrete random variables this is (repetition)

$$m = \mathrm{E}[X] = \sum_{a \in X(\Omega)} a \mathrm{P}[X = a], \ \mathrm{Var}[X] = \sum_{a \in X(\Omega)} (a - m)^2 \mathrm{P}[X = a], \mathrm{E}[g(X)] = \sum_{a \in X(\Omega)} g(a) \mathrm{P}[X = a]$$

Sometimes, one does not know the distribution of the random variable, then $\mathrm{E}[X], \mathrm{Var}[X]$ and $\mathrm{E}[g(X)]$ have to be computed by integrating (rsp. summing) over Ω.

EXAMPLES OF DISCRETE DISTRIBUTIONS:

Distribution	$\mathrm{P}[x = k] =$	Parameters	Domain	Mean	Variance
Binomial	$\binom{n}{k} p^k (1-p)^{n-k}$	$n \in \mathbb{N}, p \in [0,1]$	$\{0, \dots, n\}$	np	$np(1-p)$
Poisson	$\frac{\lambda^k}{k!} e^{-\lambda}$	$\lambda > 0$	$\{0, 1, \dots\}$	λ	λ
Geometric	$(1-p)^{k-1} p$	$p \in (0,1)$	$\{1, 2, \dots\}$	$1/p$	$1/p^2$

EXAMPLES OF ABSOLUTELY CONTINUOUS DISTRIBUTIONS:

Distribution	Density $f(x) =$	Parameters	Domain	Mean	Variance
Uniform	$1_{[a,b]} \cdot (b-a)^{-1}$	$a < b$	$[a, b]$	$(a+b)/2$	$(b-a)^2/12$
Exponential	$\lambda e^{-\lambda x}$	$\lambda > 0$	$\mathbb{R}^+$	$1/\lambda$	$1/\lambda^2$
Normal	$(2\pi\sigma^2)^{-1/2} e^{-\frac{(x-m)^2}{2\sigma^2}}$	$m \in \mathbb{R}, \sigma^2 > 0$	$\mathbb{R}$	m	σ^2
Erlang	$\frac{\lambda^k x^{k-1}}{(k-1)!} e^{-\lambda x}$	$\mathbb{R}^+$	$\lambda > 0, k \in \mathbb{N}$	k/λ	k/λ^2

8. Week, (Summary of the theory)

PROPERTIES OF DISTRIBUTION FUNCTIONS:

$$F(t) \in [0,1] \qquad \text{P}[a < X \leq b] = F(b) - F(a)$$
$$a \leq b \Rightarrow F(a) \leq F(b) \qquad \lim_{t \to -\infty} F(t) = 0, \lim_{t \to \infty} F(t) = 1$$
$$\lim_{\epsilon \searrow 0} F(a+\epsilon) = F(a) \qquad \lim_{\epsilon \searrow 0} F(a-\epsilon) = F(a) - \text{P}[X = a]$$

Every function $F : \mathbf{R} \to \mathbf{R}$ satisfying the above properties belongs to a random variable X: define the probability space $(\Omega, \mathcal{A}, P)$, where $\mathcal{A}$ is the Borel σ-algebra on $\Omega = \mathbf{R}$ and P is defined by $\text{P}[[a,b)] = F(b) - F(a) = \text{P}[X \in [a,b)]$.

DEFINITION: $X = (X_1, X_2, \dots, X_d)$ is called a **random vector** if X_i are random variables. The **distribution function** of X (also called **joint distribution** of $X_1, X_2, \dots, X_d$) is defined as

$$F(t_1, \dots, t_d) = \text{P}[X_1 \leq t_1, X_2 \leq t_2, \dots, X_d \leq t_d] \,.$$

The distribution is **continuous** if there exists a function $f : \mathbf{R}^d \to \mathbf{R}^+$ such that

$$F(t) = \int_{-\infty}^{t_1} \int_{-\infty}^{t_2} \cdots \int_{-\infty}^{t_d} f(y_1, y_2, \dots, y_d)\, dy_d \dots dy_2 dy_1 \,.$$

TRANSFORMATIION OF RANDOM VARIABLES:

- Let F be a continuous invertible distribution function. Let X be a random variable which is uniformly distributed in $[0,1]$. Then $Y = F^{-1}(X)$ gives random numbers with distribution F.
- Given a continuous random variable X with density f and a differentiable invertible function $\phi : \mathbf{R} \to \mathbf{R}$ with inverse ψ. The random variable $Y = \phi(X)$ has the density

$$g(t) = f(\psi(t))|\psi'(t)| \,.$$

- Given a continuous random vector X with density f and a differentiable invertible function $\phi : \mathbf{R}^d \to \mathbf{R}^d$ with inverse ψ. The random variable $Y = \phi(X)$ has the density

$$g(u) = f(\psi(u))|\text{Det}(D\psi(u))| \,.$$

DEFINITION: The **characteristic function** of X is defined as $\phi_X(t) = \text{E}[e^{itX}]$.
Discrete case: $\phi_X(t) = \sum_{a \in X(\Omega)} e^{ita}\text{P}[X = a]$. Continuous case: $\phi_X(t) = \int_{-\infty}^{\infty} e^{itx} f(x)\, dx$.
CALCULATION OF MOMENTS: $\text{E}[X^k] = (-i)^k \phi_X^{(k)}(0)$. Especially, $\text{E}[X] = -i\phi'_X(0)$.

SUMS OF INDEPNDENT RANDOM VARIABLES: X_i independent with distribution $\phi_i, S = \sum_{i=1}^n X_i$, then $\phi_S(t) = \phi_1(t) \cdot \phi_2(t) \cdots \phi_n(t)$.

THE GAMMA FUNCTION. Some distributions use the Gamma function:

$$\Gamma(n) = \int_0^\infty z^{n-1} e^{-z}\, dz \,.$$

For $n \in \mathbf{N}$, we have $(n-1)!$. Proof. $\Gamma(1) = 1$, $\Gamma(n) = (n-1)\Gamma(n-1)$ by partial integration. Computations like $\Gamma(1/2) = \sqrt{\pi}, \Gamma(3/2) = \sqrt{\pi}/2$ use $\int_{\mathbf{R}} e^{-x^2/2} = \sqrt{2\pi}$.

Distribution	Parameter	Charact. function
Normal	$m \in \mathbf{R}, \sigma^2 > 0$	$e^{mit - \sigma^2 t^2/2}$
Standard normal		$e^{-t^2/2}$
Uniform	$[-a, a]$	$\sin(at)/(at)$
Exponential	$\lambda > 0$	$\lambda/(\lambda - it)$
Binomial	$n \in \mathbf{N}, p \in [0,1]$	$(p + (1-p)e^{it})^n$
Poisson	$\lambda > 0$	$e^{\lambda(e^{it}-1)}$
Geometric	$p \in (0,1)$	$\frac{pe^{it}}{(1-(1-p)e^{it})}$

9. Week, (Summary of the theory)

CHEBYCHEV-MARKOV INEQUALITY.
Let $h : \mathbf{R}^+ \to \mathbf{R}^+$ be a monoton function and $X \geq 0$ a random variable with $h(X) \in \mathcal{L}^1$. Then for all $c > 0$

$$h(c) \cdot \mathrm{P}[X \geq c] \leq \mathrm{E}[h(X)] .$$

Proof. Take the expectation of $h(c) 1_{X \geq c}(\omega) \leq h(X)(\omega)$. Use the monotonicity and linearity of the expectation.
CHEBYCHEV INEQUALITY.
If $X \in \mathcal{L}^2$, then for all $c > 0$

$$\mathrm{P}[|X - \mathrm{E}[X]| \geq c] \leq \frac{\mathrm{Var}[X]}{c^2} .$$

Proof. Apply Chebychev-Markov's inequality to $Y = |X - \mathrm{E}[X]|$ and $h(x) = x^2$.

DEFINITION.
A sequence of random variables X_n **converges in probability** to a random variable X, if for all $\epsilon > 0$,

$$\lim_{n \to \infty} \mathrm{P}[|X_n - X| \geq \epsilon] = 0 .$$

WEAK LAW OF LARGE NUMBERS.
Assume X_i have common expectation $E[X_i] = m$ and satisfy $\sup_n \frac{1}{n} \sum_{i=1}^n \mathrm{Var}[X_i] < \infty$. If X_n are pairwise uncorrelated, then $\sum_{i=1}^n X_i/n \to m$ in probability.
Proof. Since in generl $\mathrm{Var}[X + Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2 \cdot \mathrm{Cov}[X, Y]$ and X_n are pairwise uncorrelated, we have $\mathrm{Var}[X_n + X_m] = \mathrm{Var}[X_n] + \mathrm{Var}[X_m]$ for $n \neq m$ and by induction $\mathrm{Var}[S_n] = \sum_{i=1}^n \mathrm{Var}[X_i]$. Using linearity, we obtain $E[S_n/n] = m$ and

$$\mathrm{Var}[S_n/n] = E[(S_n)^2/n^2] - E[S_n]^2/n^2 = \mathrm{Var}[S_n]/n^2 = \frac{1}{n^2} \sum_{i=1}^n \mathrm{Var}[X_n] \to 0 .$$

With Chebychev's inequality, we obtain

$$P[|S_n/n - m| \geq \epsilon] \leq \frac{\mathrm{Var}[S_n/n]}{\epsilon^2} \to 0 \; (n \to \infty) .$$

IMPORTANT SPECIAL CASE.
If X_i are independent random variables with the same distribution for which the mean m and variance exist both, then $\sum_{i=1}^n X_i/n \to m$ in probability.

EXISTENCE OF INDEPENDENT RANDOM VARIABLES: (don't read this!)
Given a distribution function F, there exists a probability space $(\Omega, \mathcal{A}, P)$ and independent random variables $X_1, X_2, \dots$ which have all the distribution F.
Proof. We know how to construct a single random variable X with distribution F on a probability space $(\mathbf{R}, \mathcal{B}, Q)$. Form the product space

$$(\Omega, \mathcal{A}, P) = (\mathbf{R}^{\mathbf{N}}, \mathcal{B}^{\mathbf{N}}, Q^{\mathbf{N}}) .$$

Ω contains sequences $\omega = (\omega_1, \omega_2, \dots)$ and the probability measure P is defined by

$$P[A_1 \times A_2 \times A_3 \cdots \times A_n \times \mathbf{R} \times \mathbf{R} \times \dots] = Q[A_1] \cdot Q[A_2] \dots Q[A_n] .$$

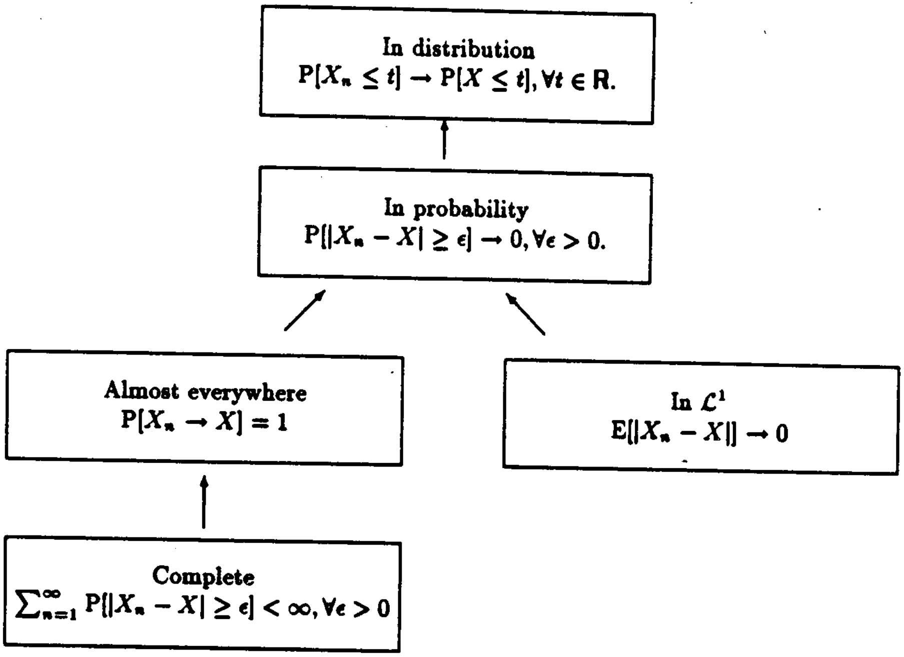
The σ-algebra $\mathcal{A}$ is the smallest σ-algebra containing all sets of the form $A_1 \times A_2 \times A_3 \cdots \times A_n \times \mathbf{R} \times \mathbf{R} \times \dots$ with $A_i \in \mathcal{B}$. The random variables $X_i(\omega) = \omega_i$ are independent and have all the distribution F.

Ma2c, O. Knill, **June 1995**

10. Week, (Summary of the theory)

DEFINITION. A sequence of random variables X_n converges **almost everywhere** to a random variable X, if $P[X_n \to X, n \to \infty] = 1$.

RELATION BETWEEN CONVERGENCE OF RANDOM VARIABLES: an arrow stands for "implies"



STRONGER WEAK LAW OF LARGE NUMBERS:
Assume X_i have common expectation $E[X_i] = m$ and satisfy $M = \sup_n E[X^4] < \infty, \sup_n E[X^2]^2 < \infty$. If X_i are independent, then $\sum_n P[|S_n/n - m| \geq \epsilon]$ converges for all $\epsilon > 0$.
Proof. Estimation of $E[X_n^4]$ with Chebychev-Markov's inequality gives $P[|S_n/n - m| \geq \epsilon] \leq C/n^2$ for some constant C.

STRONG LAW OF LARGE NUMBERS:
Assume X_n are independent random variables with $M = \sup_n E[X_n^4] < \infty$, $\sup_n E[X_n^2] < \infty$ with common expectation $E[X_n] = m$. Then $S_n/n \to m$ almost everywhere.
Proof. Direct consequence of the stronger weak law above since complete convergence implies convergence almost everywhere.

DEFINITION. A sequence of random variables X_n converges **in distribution** to a random variable X, if for all $t \in \mathbb{R}$, $P[X_n \leq t] \to P[X \leq t]$ for $n \to \infty$.

CENTRAL LIMIT THEOREM:
Given X_n which are independent with mean m and variance σ^2. Let X be a random variable with standard normal distribution. Then
$$\frac{S_n - nm}{\sigma\sqrt{n}} \to X$$
in distribution, where $S_n = X_1 + X_2 + \ldots + X_n$.
Proof. A calculation shows that the characteristic functions of $S_n^* = (S_n - E[S_n])/\sigma[S_n]$ converge to the characteristic function of X.