MATH 464/564

THEORY OF PROBABILITY

Spring Semester 1997

INSTRUCTOR:	Oliver Knill
OFFICE:	Room 510 Math Building,
PHONE:	(520) 621-4835
E-MAIL:	knill@math.arizona.edu,
WWW :	http://www.math.arizona.edu/~knill
COURSE URL:	http://www.math.arizona.edu/~knill/MA464/ma464.html
OFFICE HOUR:	Monday 15:00-16:00, Tuesday, Thursday 13:00-14:00
MIDTERMS:	Thursdays, February 20, 1997 and March 27, 1997
FINAL:	Thursday, May 15, 1997, 14:00-16:00
HOMEWORK:	Due every Tuesday in class.
FIRST DUE:	January 28, 1997
ROOM, TIME:	NEW!! PAS 414, TTH 14:00-15:15
FIRST CLASS:	January 16, 1997
EXAM POLICY:	No make-up exams.
ATTENDENCE:	Regular attendance is nessesary.
GRADING:	HW 40%, MT each $15\%W$, F 30%. + Activity bonus.
INCOMPLETES:	Only exceptionally.

MATH 464/564

THEORY OF PROBABILITY

Spring Semester 1997

INSTRUCTOR:	Oliver Knill
OFFICE:	Room 510 Math Building,
PHONE:	(520) 621-4835
E-MAIL:	knill@math.arizona.edu,
WWW :	http://www.math.arizona.edu/~knill
COURSE URL:	http://www.math.arizona.edu/~knill/MA464/ma464.html
OFFICE HOUR:	Monday 15:00-16:00, Tuesday, Thursday 13:00-14:00
MIDTERM:	Thursdays, February 20, 1997 and March 27, 1997
FINAL:	Thursday, May 15, 1997, 14:00-16:00
HOMEWORK:	Due every Tuesday in class.
ROOM, TIME:	PAS 414, TTH 14:00-15:15
FIRST CLASS:	January 16, 1997
FIRST HOMEWORK DUE:	January 28, 1997

Math 464/564, O. Knill, **January 1997**

Why is a solid foundation of probability theory necessary?

Colloquial language is not precice enough to treat many probabilistic problems. Paradoxons or difficulties appear, when the definition of objects allows different interpretations. This leads to headache, to funny or wrong conclusions even by smart people.

BERTRAND'S PARADOX (Bertrand 1889)
We 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?

Answer Nr 1: take an arbitrary point P in the disc. The set of all lines passing through that point is parametrized by an angle ϕ. In order that the chord is longer than $\sqrt{3}$, the line has to lie within an angle of 60° out of 180°. The probability is thus $1/3$.
Answer Nr 2: consider 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 $1/2$.
Answer Nr. 3: if the intersections of the line with the disc lies in the disc of radius $1/2$ which has area $1/4$ times the area of the whole disc, the chord is longer than $\sqrt{3}$. The probability is thus $1/4$.

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

Fact Nr. 1: Fair would be an entrence fee of

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

which is the expected win.

Fact Nr. 2: Nobody would agree to pay $c = 10$ dollars, would you? Try it out!

THE THREE DOOR 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, the host 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?

The paradox is that most people analyze this simple problem in a wrong way. 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 confusion.

What do you think? Check out the web links on the course websites.

Also real life problems can be complicated: "How to get my own grandfather" (A joke)

I married a widow, who had an adult stepdaughter. My father, a widow and who often visited us, fell in love with my stepdoughter and married her. So, my father became my son-in-law and my stepdaughter became my stepmother. But my wife became the mother-in-law of her father-in-law. My stepmother, stepdaughter of my wife, became a son and I therefore a brother, because he is the son of my father and my stepmother. But since he was in the same time the son of our stepdaughter, my wife became his grandmother and I became the grandfather of my stepbrother. My wife gave me also a son. My stepmother, the stepsister of my son, is in the same time his grandmother, because he is the son of her stepson and my father is the brother-in-law of my child, because his sister is his wife. My wife, who is the mother of my stepmother, is therefore my grandmother. My son, who is the child of my grandmother, is the grandchild of my father. But I'm the husband of my wife and in the same time the grandson of my wife. This means: I'm my own grandfather.

Math 464/564, O. Knill, **January 1997**

Some History and Motivation

SOME HISTORICAL MARKS IN PROBABILITY.

- 17'th century: Pascal and Ferat begin to discuss combinatorical problems connected with dice throwing.
- Inputs by J. Bernoulli A. de Moivre, T. Bayes, L. de Bufon, D. Bernoulli, A. Legendre and J. Lagrange.
- (1812) Laplace book "Théorie analytique des probabilités: new methods like difference equations or generating functions.
- (1909) E. Borel: Strong law of large numbers.
- (1923) N. Wiener Rigorous definition of Brownian motion.
- (1933) A. Kolmogorov. Axiomatic foundation of probability theory.
- Until now, an explosion of results and new directions: stochastic processes, ergodic theory and dynamical systems, Martingales, stochastic differential equations statistical mechanics, percolation, random walks, game theory or quantum field theory.

WHERE IS PROBABILITY THEORY USEFUL?

Probability theory has close relations with other fields like computer science, ergodic theory and dynamical systems, cryptology, game theory, analysis, partial differential equation, mathematical physics, economical sciences or statistical mechanics.

1) **Random walks** (statistical mechanics, gambling, stock markets, quantum field theory).
We walk randomly through a lattice by choosing at each vertex a direction at random. A problem is to determine the probability that the walk returns back to the origin.

2) **Percolation problems** (model of a porous medium, statistical mechanics).
To each bond in a lattice is attached the number 0 (closed) with probability $(1-p)$ or 1 (open) with probability p. Two sites x, y in the lattice are in the same cluster, if there is a path from x to y going though closed bonds. One says that percolation occurs if there is a positive probability that an infinite cluster appears. One problem is to find the critical probabily p_c, which is the infimum over all p, for which percolation occurs.

3) **Random Schrödinger operators**. (quantum mechanics, functional analysis, disordered systems, solid state physics)
Consider the linear map $Lu(n) = \sum_n u(n) + V(n)u(n)$, where $V(n)$ takes randomly values in $\{0, 1\}$. The problem is to determine the spectrum or spectral type of the infinite matrix L. The map L describes an electron in a one dimensional disordered crystal. The spectral properties of L have a relation with the conductivity of the crystal.

4) **Classical dynamical systems** (celestial mechanics, fluid dynamics, population models)
The study of deterministic dynamical systems like the map $x \mapsto 4ax(1-x)$ on the interval $[0, 1]$ or the three body problem in celestial mechancs has shown that such systems can behave like random systems. Many effects can be described by ergodic theory, which can be viewed as a brother of probability theory.

5) **Cryptology** (computer science, coding theory, data encryption)
Data encryption like the DES are used in most computers or phones. Coding theory tries to find good codes which can repair loss of information due to bad channels. Public key systems use trapdoor algorithms like the problem to factor a large integer $N = pq$ with prime p, q. The number N can be public but only the one, who knows the factorisation can read the mails. Many algorithms need pseudo random number generators like the sequence generated by $m \mapsto m^2 + c(modp)$. Much probability theory is involved in designing, investigating and attacking data encryption, codes or random number generators.

Math 464/564, O. Knill, **January 1997**

Kolmogorov axioms and the foundation of probability

DEFINITION: Set theoretical operations.

Ω any set	”The set of all experiments”
$A, B, A_n \subset \Omega$	”Events”
$A \cap B = \{\omega \in \Omega \mid \omega \in A \; and \; \omega \in B\}$	”Both events A and B happen”
$A \cup B = \{\omega \in \Omega \mid \omega \in A \; or \; \omega \in B\}$	”Either A or B happens”
$A \Delta B = \{\omega \in \Omega \mid \omega \in A \; or \; \omega \in B \; but \; not \; in \; both\}$	”One of the events A or B happens”
$A \setminus B = \{\omega \in \Omega \mid \omega \in A \; but \; not \; \omega \in B\}$	”A but not B happens”
$A^c = \{\omega \in \Omega \mid \omega \notin A\}$	”A does not happen”
$\bigcap_n A_n = \{\omega \in \Omega \mid \omega \in A_n, \; for \; all \; n\}$	”All events A_n happen”
$\bigcup_n A_n = \{\omega \in \Omega \mid \omega \in A_n, \; for \; at \; least \; one \; n\}$	”At least one event A_n happens”

DEFINITION: Boolean algebra

Ω be a set. A set $\mathcal{A}$ of subsets of Ω is a **Boolean algebra**, if

$$\begin{array}{l} \Omega \in \mathcal{A}, \\ A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}, \\ A, B \in \mathcal{A} \Rightarrow A \cup B \in \mathcal{A}. \end{array}$$

DEFINITION: σ-**field**

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

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

PROPERTIES: A Boolean algebra $(\Omega, \mathcal{A})$ is closed under all set theoretical operations: $A, B \in \mathcal{A}$, then

$$\begin{array}{ll} \emptyset \in \mathcal{A} & A \cap B \in \mathcal{A}. \\ A \setminus B \in \mathcal{A} & A \Delta B \in \mathcal{A}. \end{array}$$

A σ-field $(\Omega, \mathcal{A})$ is closed under all set theoretical operations which can be enumerated.

DEFINITION: Probability measure

A function $P : \mathcal{A} \to \mathbb{R}$ on a σ-field $(\Omega, \mathcal{A})$ is a **probability measure** if

$$\begin{array}{ll} P[A] \geq 0, & \text{(nonnegativity)} \\ P[\Omega] = 1, & \text{(normalisation)} \\ P[\bigcup_{i=1}^n A_i] = \sum_{i=1}^n P[A_i], \text{ if } A_i \cap A_j = \emptyset, \text{ for all } i, j, & \text{(additivity)} \end{array}$$

”$P[A]$ is the probability that the event A happens.”

PROPERTIES:
$A \subset B \Rightarrow P[A] \leq P[B]$.
$P[A^c] = 1 - P[A]$, $\quad P[\emptyset] = 0$
$P[B] = P[A \cap B] + P[A^c \cap B]$
$P[\bigcup_n A_n] = 1 - P[\bigcap_n A_n^c]$.

SWITCH ON, SWITCH OFF formula:

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

Math 464/564, O. Knill, **January 1997**

KOLMOGOROV AXIOMS

A probability space $(\Omega, \mathcal{A}, P)$ consists of a set Ω (the set of exeriments), a σ-algebra $\mathcal{A}$ (the set of events) and a function $P : \mathcal{A} \to [0,1]$ (the probability measure).

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

$$\begin{array}{|l|}\hline \Omega \in \mathcal{A}, \\ A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}, \\ \{A_1, A_2, \dots\} \subset \mathcal{A} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}. \\ \hline \end{array}$$

DEFINITION: Probability measure

A function $P : \mathcal{A} \to \mathbb{R}$ on a σ-field $(\Omega, \mathcal{A})$ is a **probability measure** if

$$\begin{array}{|lr|}\hline P[A] \geq 0, & \text{(nonnegativity)} \\ P[\Omega] = 1, & \text{(normalisation)} \\ P[\bigcup_{i=1}^{n} A_i] = \sum_{i=1}^{n} P[A_i], \text{ if } A_i \cap A_j = \emptyset, \text{ for all } i,j, & (\sigma\text{-additivity}) \\ \hline \end{array}$$

Math 464/564, O. Knill, **January 1997**

Two controversial examples

The Girl-Boy problem

QUESTION. "Dave has two childs. One child is a boy. What is the probability that the other child is a girl".

SOLUTION. $\Omega = \{BG, GB, BB\}$, $\mathcal{A}$ = set of subsets of Ω and $P[\{BG\}] = P[\{GB\}] = P[\{BB\}] = 1/3$. Now $A = \{BG, GB\}$ is the event that the other child is a girl. $P[A] = 2/3$.

Solution with conditional probability. $\Omega = \{BG, GB, BB, GG\}$ with $P[\{BG\}] = P[\{GB\}] = P[\{BB\}] = P[\{GG\}] = 1/4$. Let $B = \{BG, GB, BB\}$ be the event that there is at least one boy and $A = \{GB, BG, GG\}$ be the event that there is at least one girl. We have $P[A|B] = P[A \cap B]/P[B] = (1/2)/(3/4) = 2/3$.

Check out:

`http://www.wiskit.com/marilyn.boys.html`

The Monty-Hall Problem

QUESTION. "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, the host 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?"

SOLUTION. The probability space is $\Omega = \{goat1, goat2, car\}$ with $P[\{goat1\}] = P[\{goat2\}] = P[\{car\}] = 1/3$.
First case: No switching: The winning event is $A = \{car\}$ which has probability 1/3.
Second case: Switching: The winning event is $A = \{goat1, goat2\}$ which has probability 2/3.

Check out:

`http://www.wiskit.com/marilyn.gameshow.html`

I recommend the cite

`http://http://www.cut-the-knot.com`

for more Mathematics puzzles.

Math 464/564, O. Knill, **February 1997**

What is randomness?

Randomness in the distribution of Primenumbers?

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? Experimets suggest that $n \mapsto (q_{n+1} - q_n)/\log(n)$ behaves asymptotically like an exponential distributed random variable X. It is an unsolved mathematical problem if such a relation can be proven asymptotically.

```
ListPlot[Table[Prime[n+1]/Log[Prime[n]],{n,1000}]]
```

```
q[n_]:=Prime[n+1]/Log[Prime[n]];
h[n_]:=(q[n+1]-q[n])/Log[n];
ListPlot[Table[h[n],{n,1000,3000}]]
```

Randomness in the digits of π ?

Consider the the numbers $\pi * 10^k \ mod 1$, obtained by cutting away a bunch of digits in π. We can define Ω as the set of all numbers, which are limits of numbers π_k. Do the π_k behave as independent random variables over the probability space $\Omega = [0,1], \mathcal{A} = \textit{measurable sets}, P[a,b] = b - a$)?

```
IntegerDigits[Floor[10^1000*N[Pi,1000]]];
```

Math 464/564, O. Knill, **February 1997**

Basic formulas in combinatorics

PERMUTATIONS. Let $X = \{1, \dots, n\}$ be a finite set. A permutation π of X is a bijective function $\pi : X \to X$. There are

$$p(n) = n! = n(n-1) \cdots 1$$

such permutations. Proof: by induction, if one of the n possible values $\pi(1)$ is fixed, there are $p(n-1)$ possibilities to fix the other values $\pi(k)$.
EXAMPLE: There are $n!$ possible ways to redistribute n coats of n people.

PERMUTATIONS WITH IDENTIFICATION. Take $X = \{1_1, \dots, 1_{n_1}, 2_1, \dots, 2_{n_2}, \dots, k_1, \dots, k_{n_k}\}$. Elements like 2_3 and 2_5 are identified. There are

$$p(n; n_1, \dots, n_k) = \frac{n!}{n_1! \cdots n_k!}$$

permutations without distinguishing identified elements. Among all $n!$ permutations, there are $n_1!$ which are distinguished only by a permutation of identified elements 1_k so that $n!/n_1!$ permutations have 1-elments not distinguished. Proceed inductively.
EXAMPLE: From the letters A, A, B, B, B, B, E, U one can make $p(8; 2, 4, 1, 1)$ different words.

SAMPLING. We choose k elements from $X = \{1, \dots, n\}$. we look for k elements and do distinguish the picking order.

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

if the picking order is distinguished.
EXAMPLE. 6 persons can in $v(10,6)$ possible ways take place on 10 chairs. See also the birthday paradox from class.

SAMPLING WITH REPLACEMENT. We choose k elements out of a finite set $X = \{1, \dots, n\}$ but put the element back each time. The number of such elements is

$$v^*(n,k) = n^k$$

EXAMPLE: There are 6^{10} possible ways to throw 10 distinguishible dices.

COMBINATIONS. We pick k elements from n elements not distinguishing the picking order. We have to divide $v(n,k)$ by $k!$ to get

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

possibilities to choose k elements out of n elements.
EXAMPLE: We can choose in $c(52,10)$ possible ways a set of 10 cards from 52 cards.

COMBINATIONS WITH REPETITIONS. If in the combinations, we allow to pick several times the same elements, the number of possibilities is

$$c^*(n,k) = \frac{(n+k-1)!}{k!(n-1)!}$$

Math 464/564, O. Knill, **February 1997**

Basic problems in combinatorics (II)

PERMUTATIONS. • There are $n!$ bijective maps on a set of n elements.

PERMUTATIONS WITH IDENTIFICATION. • Using the digits $1, 1, 2, 5$, one can form the $p(4; 2, 1, 1) = 12$ different numbers

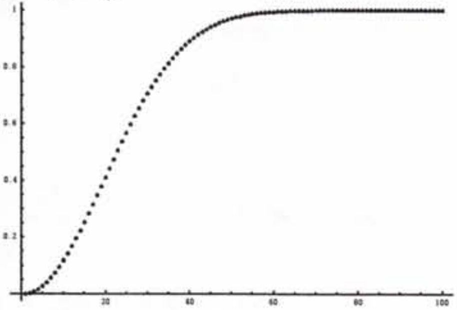
$$1125, 1152, 1215, 1251, 1512, 1521, 2115, 2151, 2511, 5112, 5121, 5211$$

• The relation

$$\left(\begin{array}{c} n \\ n_1, n_2, \dots, n_k \end{array} \right) = \left(\begin{array}{c} n \\ n_1 \end{array} \right) \left(\begin{array}{c} n - n_1 \\ n_2 \end{array} \right) \left(\begin{array}{c} n - n_1 - n_2 \\ n_3 \end{array} \right) \cdots \left(\begin{array}{c} n_k \\ n_k \end{array} \right)$$

holds because we can first choose n_1 elements from n slots, fill them with the n_1 elements of the frist type, then choose n_2 elements from the remaining $n - n_1$ slots, etc.

SAMPLING. • 6 persons can in $v(10, 6) = 10 \cdot 9 \cdots 4$ possible ways take place on 10 chairs. (The first person has 10 possibilites, the second person 9, the last one only 4).



• Birthday paradox: The probability that amoung k persons at least 2 have the birthday at the same day in the year is $1 - v(365, k)/365^k$.

SAMPLING WITH REPLACEMENT. • There are 6^{10} possible ways to throw 10 distinguishible dices.
• There are n^n possible maps from a set X with n elements into itself.

COMBINATIONS. • Arizona Lotto. The probability to win the jackpot (choosing 6 correct numbers from 42 numbers) is one to

$$\left(\begin{array}{c} 42 \\ 6 \end{array} \right) = 5'245'786$$

If k of n elements have some winning property and we choose k elements, the probability to get m winning elements is

$$\left(\begin{array}{c} n - k \\ k - m \end{array} \right) \left(\begin{array}{c} k \\ m \end{array} \right) / \left(\begin{array}{c} n \\ k \end{array} \right)$$

• Arizona Lotto again. The probability to have 3 numbers of the 6 numbers right (and win 2 dollars) is: $\left(\begin{array}{c} 6 \\ 3 \end{array} \right) \left(\begin{array}{c} 36 \\ 3 \end{array} \right) / \left(\begin{array}{c} 42 \\ 6 \end{array} \right) = 1/36.74$.
Arizona Power ball. Choose 5 white balls from $\{1, \dots, 45\}$ and choose 1 red ball from $\{1, \dots, 45\}$. What is the probability to have 6 right? Answer : $1/124926$.

COMBINATIONS WITH REPETITIONS.
• The recursion $c^*(n, k) = c^*(n, k-1) + c^*(n-1, k)$ proves the formula and can be seen as follows: mark one element and divide the possible combinations into two classes. In the class with the distinguished element, there are $c^*(n, k-1)$ elements. In the class without, there are $c^*(n-1, k)$ elements.
• Music accord. There are 12 different tunes in an octave. An accord is a set of tunes played simultaneously. A jazz music group having three saxophons wants to know, how many accords three saxophons can play. (It is allowed that two saxophons play the same tune). There are $c^*(12, 3) = 364$

Math 464/564, O. Knill, **February 1997**

Combinatorics in Music

Are there combinatorical restrictions in creating new music? How come that one plays today also music from say 200 years ago while in mathematics for example, it is unthinkable to read Gauss or Eulers work? Maybe there are not so many melodies to invent and stealing of tunes is a necessity? Evenso, we can not answer the above question, this brings us to **combinatorial music theory**.

A first interesting question is what are the choices to using frequencies, rythms etc? Let's stick to frequencies.

The western music divides the octave into 12 equal steps in the sense that going up one steps corresponds to a multiplication of the frequency by a factor $2^{1/12}$. The frequencies form a **geometric sequence.** Why does one do that? The mulitplicative structure is because our ear listens logarithmically (which is both physically and physiologically reasonable). Frequency relations, which are rational are considered "harmonic".

Euler defined in his music theory for a frequency ration p/q the **"gradus suavitatis"** (degree of pleasure) $\Gamma(p/q) = (1 + \sum_i (p_i - 1))$, where p_i are the prime factors of $p * q$ (with multiplicity). For example, the little decime 12/5 has the degree of pleasure $(1+(2-1)(2-1)(3-1)(5-1)) = 9$. Now, choosing a geometric scale of 12 tunes interpolating a frequency doubling allows to approximate quite well the most harmonic relations.

$$\begin{aligned}
2 &= \Gamma(1/2)\\
3 &= \Gamma(1/3) = \Gamma(1/4)\\
4 &= \Gamma(2/3) = \Gamma(1/6) = \Gamma(1/8)\\
5 &= \Gamma(1/5) = \Gamma(1/9) = \Gamma(1/12) = \Gamma(3/4) = \Gamma(1/16)
\end{aligned}$$

As a matter of fact, the square of $2^{-1/12}$ which is $2^{-1/6} = 1.122462048...$ is close to $9/8 = 1.125$. The "antique doric tune" is obtained by taking from $c-d$ and $d-e$ the frequency relation 9/8 and from $e-f$ the relation 128/117, ($c-f$ has then a relation 4/3) then again 9/8 from $f-g$, $g-a$ and $a-h$ and again 128/117 from h to c. Other tunes are the "diatonic tune", $4/3 = (9/8)*(8/7)*(28/27)$ the "chromatic tune" $4/3 = (32/27)*(36/35)*(28/27)$, or the "enharmonic tune" using $4/3 = (5/4)*(36/35)*28/27)$. The frequency relation $2^{1/12}$ for the smaller steps has the advantage that it is "translational invariant", that is one can compose music pieces translated. This is changing the possibilities for a composer of piano muscic because the scale on a piano is not translational invariant (black and white keys).

From the building block of 12 frequencies, one can choose sub scales, which have better "degrees of pleasure". Examples are the scale $c-d-e-f, g, a, h$ of children songs consisting of 7 elements. Taking the same scale but starting with an other initial point gives so called "church-scales" or the scale "minor".

Our ear does not like frequency relations which large "degree of pleasure". That's why one has a smaller set of frequencies 7 instead of 12, with which one can play simpler melodies.

Now, we look at some problems to play accords (tunes played together), and melodies (tunes played in a sequence):

Question 1: A piano has 88 tunes. How many accords consisting of 3 tunes can one play? Answer: $c(88,3) = 109736$.

Some of the accords are translational invariant and the tunes have to be taken modulo 12.
Question 2: How many accords of 3 tunes can one play modulo 12. Answer: put the first tune at c. Now, can choose 2 elements of 11. So, we have only 55 such accords.
Question 3 (to combinations with symmetry): How many accords can one play on the stair of 7 tunes $\{c, d, e, f, g, a, h\} = \{1, 2, 3, 4, 5, 6, 7\}$? Answer 2^7. To make the question more difficult, we ask for translational invariant different accords.
This question is now more difficult bug can be answered with the

LEMMA OF BURNSIDE: Let a group G act on a set H. Then the number of cycles of the group action is
$$|G|^{-1} \sum_{g \in G} Fix(g) .$$

To apply this lemma to the problem of accords: let G be the group of translations consisting of 7 elements. The number of fixed points is 2^7 for the identity, 2 for every other element. The answer is $(2^7 + 2 * 6)/20$. Most of these accords are used in music: $\{"pause","one-tune","second","tertz","quart","second-quart","second-tertz","second-quint","three-accord","tertz-quart","quarte-ajoutee","second-tertz-quarte","second-quart-quint","sixth-ajoutee","second-ajoutee","second-quart-ajoutee","quart-sixth-ajoutee","second-sith-ajoutee","second-quart-sixth-ajoutee","septime-second-quarte-sixth-ajoutee"\}$.

Question 4 (to permutations): How many pieces can one play in the "twelf-tone music"? Answer $12! = 479'001'600$.
Question 5 (to sampling) Anton Webern's Sinphonie op 21 the second "satz" consisists of a "theme" with 8 "variations" and a "coda". Each, "theme", "variation" and "coda" is a "twelf-tone music" and they have to be different. How many "sinphonies" can one write like this? Answer: we have to choose 11 permutations out of the 12! permutations. Therefore

$c(12!, 11) = 7630462054792415927705549450172273571842040394984563168904316741381986181085708708377600$.

I remind you that our universe is a about 10^{17} seconds old! So even if we could play 10^{10} themes in a second (with future 1000x CD playes using future 10 GHz Pentium processors!) and every person of the soon 10^{10} people in the world would play with 10^{10} computers. One would still have to play 10^{40} ages of universes long to enjoy all possibile sinphonies. A reason to switch the music taste?

Question 6 (to sampling without replacement): How many melodies of length n can one play on a piano with 88 keys? 88^n.

Question 7 (to combinations with repetition): Assume three saxophones play together. How many accords can one form in a scale of 12 frequencies? $c^*(12, 3) = c(14, 3) = 364$.
Question 8 (to permutations with identification): an orchestra consists of 20 strings, 10 cellos and 5 bass. We permute them randomly and let one after the other play the tune c. How many "songs" can one hear like that? Answer: $p(35; 20, 10, 5) = 9753573309480$.

Math 464/564, O. Knill, **February 1997**

Large numbers and small probabilities

10^4	A "myriad". The largest numbers, the greeks were considering.
10^5	The largest number considered by the romans.
10^{10}	The age of the universe in years.
10^{22}	Distance to our neighbour galaxie Andromeda in meters.
10^{23}	Number of atoms in two gram Carbon (Avogadro).
10^{26}	Size of universe in meters.
10^{41}	Mass of our home galaxy "milky way" in kg.
10^{51}	Archimedes (300 A.D.) estimeate of the number of sand grains in the universe.
10^{52}	Mass of our universe in kg.
10^{80}	The number of atoms in our universe.
10^{100}	One "googol". (Name by 9 year old nephew of the mathematician E. Kasner).
10^{153}	Number mentiond in a myth about Buddha.
10^{155}	Size of the ninth Fermat number (factored in 1990).
$2^{1'398'269} - 1$	Largest known prime number (This Mersenne number was found in November 1996).
10^{10^7}	Years for a parrot to write "the hound of Baskerville" by random typing.
$1 : 10^{10^{33}}$	The odds that a can of beer tips due to quantum fluctuations (Richard Crandall)
$1 : 10^{10^{42}}$	Probability that a mouse survives on the the sun for a week (John Littlewood).
$1 : 10^{10^{51}}$	Odds for finding yourself on Mars reassembled by quantum fluctuations (R. Crandall).
$10^{10^{100}}$	One "gogoolplex", does not exist in decimal expansion in our universe.

See the recent article of R.E. Crandall in

```
Scientific American, February 1997
```

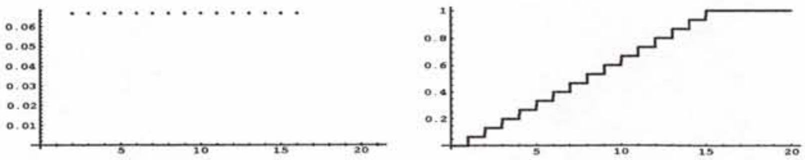
Math 464/564, O. Knill, **February 1997**

Discrete probability distributions

UNIFORM DISTRIBUTION. $\Omega = \{1, \ldots, n\}$.

$$\boxed{P[\{X(\omega) = k\}] = n^{-1}}$$

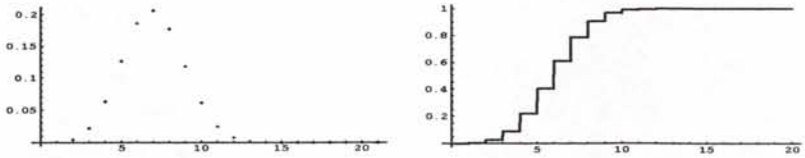
Equal districution on finite probability space.



BERNOULLI DISTRIBUTION. $\Omega = \{0, \ldots, n\}$.

$$\boxed{P[\{X(\omega) = k\}] = \begin{pmatrix} n \\ k \end{pmatrix} p^k (1-p)^{n-k}}$$

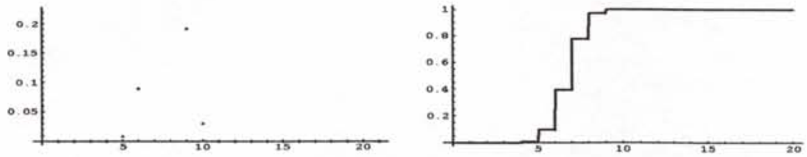
Make n experiments with success rate p. Have k times success.



HYPERGEOMETRIC DISTRIBUTION. $\Omega = \{0, 1, \ldots, n\}$.

$$\boxed{P[\{X(\omega) = k\}] = \frac{\begin{pmatrix} r_1 \\ k \end{pmatrix} \cdot \begin{pmatrix} r - r_1 \\ n - k \end{pmatrix}}{\begin{pmatrix} r \\ n \end{pmatrix}}}$$

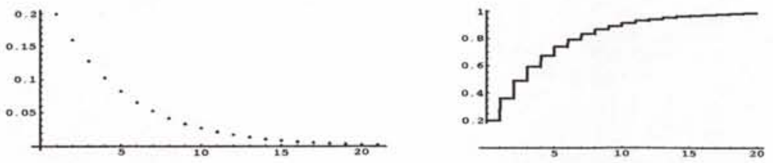
Win k from n tries by choosing from r elements amoung which r_1 have winning property. (Lotto).



GEOMETRIC DISTRIBUTION. $\Omega = \{0, 1, 2, \ldots, \} = \mathbb{N}$.

$$\boxed{P[\{X(\omega) = k\}] = p(1-p)^k}$$

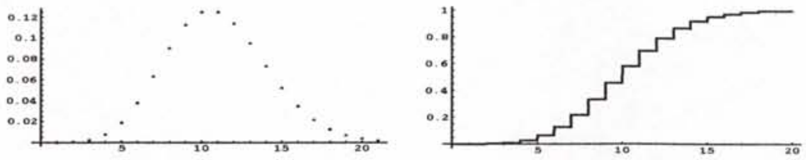
Waiting time for failing of equipment.



POISSON DISTRIBUTION. $\Omega = \{0, 1, 2, \ldots, \} = \mathbb{N}$.

$$\boxed{P[\{X(\omega) = k\}] = \frac{\lambda^k e^{-\lambda}}{k!}}$$

Used for counting pheneomena.



The illustrations for this page were done by running the following program.

```
<<Statistics‘DiscreteDistributions‘
PDFandCDF[distr_]:=Show[GraphicsArray[
      {ListPlot[Table[PDF[distr, n],{n,0,20}],
          DisplayFunction->Identity],
       Plot[CDF[distr, x],{x,0,20},
          DisplayFunction->Identity,
          AxesOrigin->{0.0,0.0}] }],
      DisplayFunction->$DisplayFunction];

Display["!psfix -land>dist01.ps",PDFandCDF[DiscreteUniformDistribution[15]]];
Display["!psfix -land>dist02.ps",PDFandCDF[BinomialDistribution[15,0.4]]];
Display["!psfix -land>dist03.ps",PDFandCDF[GeometricDistribution[0.2]]];
Display["!psfix -land>dist04.ps",PDFandCDF[PoissonDistribution[10.0]]];
Display["!psfix -land>dist05.ps",PDFandCDF[HypergeometricDistribution[15,9,20]]];
```

Math 464/564, O. Knill, **February 1997**

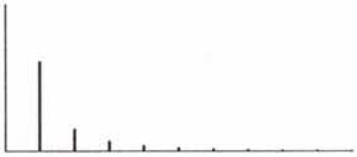
Discrete probability distributions (II)

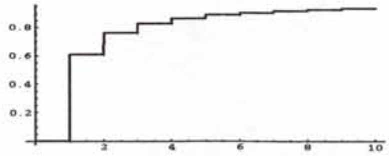
ZETA DISTRIBUTION. $\Omega = \mathbb{N} \setminus \{0\} = \{1, \dots, \}$.

$$\boxed{P[\{X(\omega) = k\}] = \zeta(s)^{-1} n^{-s}}$$

For fun. Useful in number theory, where it gives an elegant proof of the Euler formula

$$\zeta(s)^{-1} = \prod_{p \in \mathbb{N}\ p\ is\ a\ prime} (1 - p^{-s}) .$$

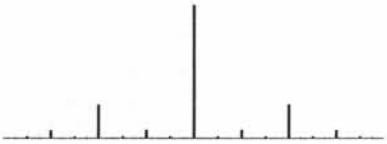


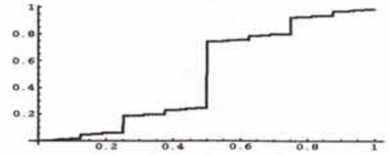


DEVILISH DISTRIBUTION. $\Omega = \{k/2^n, 1 \leq k \leq 2^n, k\ odd\}$.

$$\boxed{P[\{X(\omega) = k2^{-n}\}] = 2/4^n.}$$

Such distributions occur in quantum mechanics in crystallic alloys or random materials The function F is there called the density of states. The spikes of the density correspond to the location of the possible energy values of the electron.

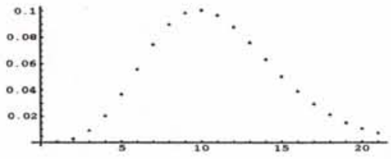


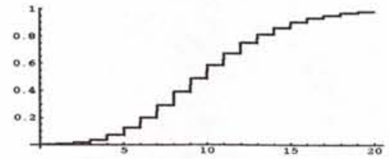


NEGATIVE BINOMIAL DISTRIBUTION. $\Omega = \{0, 1, \dots, \}$. Parameters $0 \leq p \leq 1$, $n \in \mathbb{N}$.

$$\boxed{P[\{X(\omega) = k\}] = \left(\begin{array}{c} n+k-1 \\ n-1 \end{array} \right) p^n (1-p)^k}$$

Distribution of number of failures that occur prior to the n'th success, when iterating experiments which succeed with probability p.





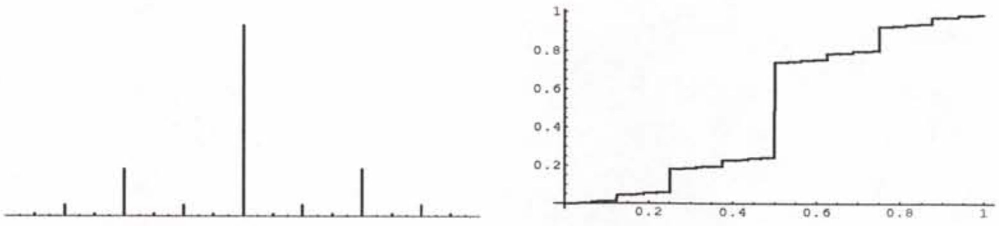
EXAMPLE 3): A DEVILISH DISTRIBUTION. The random variable

$$\begin{aligned}
X(\frac{1}{2}) &= \frac{1}{2} \\
X(\frac{1}{4}) &= X(\frac{3}{4}) = \frac{1}{8} \\
X(\frac{1}{8}) &= X(\frac{3}{8}) = X(\frac{5}{8}) = X(\frac{7}{8}) = \frac{1}{32} \\
&\;\;\cdot \\
&\;\;\cdot \\
X(\frac{1}{2^n}) &= X(\frac{3}{2^n}) = \ldots = X(\frac{2^n-1}{2^n}) = \frac{2}{4^n} \\
&\;\;\cdot \\
&\;\;\cdot
\end{aligned}$$

takes the values $(2k-1)/2^n, k = 1, \ldots, 2^{n-1}$ which is a dense set in the interval $[0,1]$.

The distribution function of X is a **devilish stair case**: you have to take in every time interval through infinitely many steps.

```
DevilSkin := {Line[{{0,0},{1,0}}]};
Hair[n_]:=Table[Line[{{(2k-1)/2^n,0},{(2k-1)/2^n,2/4^n}}],{k,2^(n-1)}];
DevilHair:=Show[Graphics[{DevilSkin,Table[Hair[n],{n,DeepnesLevel}]}],
  PlotRange->{{0,1},{0,0.55}},DisplayFunction->Identity];
LevelValues[n_]:=Table[{(2k-1)/2^n,2/4^n},{k,2^(n-1)}];
UpToLevel[nn_]:=Sort[Module[{s={}}, Do[s=Union[s,LevelValues[n]],{n,nn}];s]];
DevilStair[x_]:=Module[{v=.0},Do[If[u[[k,1]]<=x,v=v+u[[k,2]]],{k,Length[u]}];v];
DevilStairCase :=Plot[DevilStair[x],{x,0,1},DisplayFunction->Identity];
HellsBells[DL_]:=Module[{}, DeepnesLevel=DL; u=UpToLevel[DeepnesLevel];
  Show[GraphicsArray[{DevilHair, DevilStairCase}],
  DisplayFunction->$DisplayFunction]];
Display["!psfix -land>devilstair.ps",HellsBells[6]];
```



Math 464/564, O. Knill, **February 1997**

A drunken Sailor, Euler's formula and a Devilish staircase.

DEFINITION. A discrete random variable is a function $X : \Omega \to \mathbb{R}$ with a discrete image, such that $X^{-1}(x) = P[X = x]$ is an event. The function $f_X(x) = P[\{X = x\}]$ is the discrete density function of X, while $F_X(x) = P[X \leq x]$ is the discrete distribution function of X.
INTERPRETATION. Random variables describe measurements. Every experiment $\omega \in \Omega$ gives the outcome $X(\omega)$, like the value of a thrown dice.
MANIPULATION OF RANDOM VARIABLES. One can add, multiply and manipulate discrete random variables: if X, Y are random variables and λ is a real number, then $X + Y, XY, \lambda X$ are discrete random variables. Also, if $f : \mathbb{R} \to \mathbb{R}$ is a function, then $f \circ X$ is a random variable. For example $\sin(X)$ is a random variable.

EXAMPLE 1): RANDOM WALK (DRUNKEN SAILOR) The one-dimensional **random walk** models the random motion of a particle on the lattice $\mathbb{Z}$. The particle starts at 0 and makes in each period a time step $+1$ or -1 with probability $1/2$. It is also used for discribing the balance development of a **gambler in a fair game** or the **value of a stock** in the market or the **steps of a drunken sailor**.
THE PROBABILITY SPACE OF THE RANDOM WALK is the set of all possible steps

$$\Omega = \{-1, 1\}^N = \{(\omega = (\omega_1, \dots, \omega_N) \mid \omega_i \in \{-1, 1\}\}$$

the set of events $\mathcal{A} = \{A \subset \Omega\}$ and P is the probability measure $P[A] = |A|/|\Omega|$.
THE LOCATION OF THE WALKER. Consider the random variables $X_k(\omega) = \omega_k$ and $S_n = \sum_{k=1}^n X_k(\omega)$. The value $S_n(\omega)$ is the location of the walker at time n having done the sequence of steps ω. In a gambling interpretation X_k is the win rsp. loss at the k'th step and S_n is the total win until the n'th step. For each $\omega \in \Omega$, one gets a trajectory $\{S_n(\omega)\}_{n=1}^N$.
THE DISTRIBUTION OF THE RANDOM VARIABLE S_n is a **Bernoulli distribution**

$$P[S_n = k] = \left(\begin{array}{c} n \\ \frac{n+k}{2} \end{array} \right) p^k (1-p)^{n-k}$$

for $p = 1/2$. The walker can only reach places k at time n, if $n + k$ is even.

EXAMPLE 2): A FORMULA OF EULER ABOUT PRIMES. Probability theory is useful in number theory.
THE ZETA DISTRIBUTION. Consider a random variable with distribution $P[X = n] = \zeta(s)^{-1} n^{-s}$, where ζ is the **Rieman zeta function**

$$s \mapsto \zeta(s) = \sum_{n=1}^{\infty} n^{-s} \; .$$

The famous open **Riemann Hypothesis** predicts that all zeros of $\zeta(s)$ in the complex line are on the line $Re(s) = 1/2$. We know from theory that the random variable X defines a probability space $\Omega = \{1, 2, \dots, \}, \mathcal{A} = \{A \subset \Omega\}, P[\{n\}] = X(\omega))$. We have only to check that $P[\Omega] = \sum_n P[\{n\}] = \zeta(s)^{-1} \sum_n n^{-1} = 1$.
SOME INDEPENDENT EVENTS. The sets $A_p = \{n \in \Omega \mid p \; divides \; n\}$ with prime p are independent: we have $P[A_p] = p^{-s}$ and if $p_i, i = 1, \dots, k$ are primes, then

$$P[A_{p_1} \cap A_{p_2} \cap \dots \cap A_{p_k}] = \zeta(s)^{-1} \sum_n (np_1 \cdots p_k)^{-s} == (p_1 \cdots p_k)^{-s} \zeta(s)^{-1} \sum_n n^{-s} = P[A_{p_1}] \cdots P[A_{p_k}] \; .$$

EULER'S FORMULA. Consider the event $A = \{n \mid n \text{ is not divisible by any prime number }\}$. Clearly, $A = \{1\}$. And $P[A] = \zeta(s)^{-1}$. On the other hand, A is the interesection of all events $A_p^c = \Omega \setminus A_p$, where p runs over the prime numbers: from $A = \prod_{p \; prime} A_p^c$, the independence of A_p^c, we get the **formula of Euler**

$$\frac{1}{\zeta(s)} = \prod_{p \in \mathbb{N} \; p \; is \; a \; prime} (1 - p^{-s}) \; .$$

3/4/97

2/18/97

1) (5-10 minutes) Repetetion: Expectation, Variance, Properties.
Maybe compute the expectation and variance of the uniform distribution using the Probability generating function.

2) (5-10 minutes) Reminder: A random variable does not need to have a finite Expectation or finite variance (or other finite moments). See Petersburg paradox.
Example: $P[X = k] = 1/k - 1/(k+1)$. Brick paradox, numerical convergence paradox.

3) (5-10 minutes) Definition of covariance. Formula for $Var[X+Y] = Var[X]+2Cov[X,Y]+Var[Y]$.
What is a uncorrelated random variable.

4) (10 minutes) Independent random variables are uncorrelated. Thm: $E[XY] = E[X]E[Y]$ if X, Y are independent. Direct proof.
Application: independent $\Rightarrow$ uncorrelated.

Example. Binomial: $Var[X_{n,p}] = nVar[X_{1,p}] = np(1-p)$.
Example. Poisson: $Var[X_\lambda]$ is linear in λ.

5) (5-10 minutes) A useful formula for the expectation. If X is integer valued, then $E[X] = \sum_{n=1}^\infty P[X \geq n]$.
Proof. $E[X] = \sum_n \sum_{k=1}^n 1P[X = n]$. Change summation.
Example. $P[X \geq n] = \sum_{k=n}^\infty p(1-p)^k = p(1/p-(1-(1-p)^n)/p) = (1-p)^n$ for geometric distribution.
So $E[X] = \sum_{n=1}^\infty (1-p)^n = 1/p - 1$.

6) (15 minutes) The Correlation coefficient. $\rho(X,Y)$. Think of it as a Cosine of an angle. It is zero, if the random variables are uncorrelated. They can be positively correlated or negatively correlated.

Schwartz inequality $E[XY]^2 \leq E[X^2]^2 E[Y]^2$. Proof: $0 \leq E[(X+\lambda Y)^2] = E[X^2] + \lambda^2 E[Y^2] + 2\lambda E[XY]$. The minimum of the RHS is achieved for $\lambda = E[XY]/E[Y^2]$ for the RHS is $E[X^2] - E[XY]^2/E[Y^2]$.
Implies $\rho(X,Y) \leq 1$.

Example. Compute the Correlation coefficient of X and $3X$, where X is the random variable of throwing a dice once. We have $E[X] = 3.5$, $Var[X] = 35/12$ (see table). $Var[3X] = 9Var[X]$. $E[XY] = 3E[X^2] = 3(Var[X] + E[X]^2) = 3(35/12 + 70/100)$.

The correlation coefficient as well as the regression line is more useful in statistics.

The **regression line** of two random variables X, Y is the line $y = ax + b$, where

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

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

Example. If X, Y are independent, then $a = 0$ and $b = E[Y]$. Then $\tilde{Y} = E[Y]$. We can not guess better Y than replacing it by the mean.

8) (15 minutes) The Chebychev inequality.
Take a random variable X and define $Y = 0$ if $X \leq t$ as well as $Y = t$ if $X \geq t$. Then $X \geq Y$ and so $E[X] \geq E[Y]$. That is $E[X] \geq tP[X \geq t]$. Applying this to $(X - E[X])$ and $t = \epsilon$, we get

$$E[(X - E[X]) \geq \epsilon] \leq Var[X]/\epsilon^2$$

which is called the **Chebychev inequality**.

9) (15 minutes) The weak law of large numbers. $Var[S_n/n] = Var[X_j]/n^2$ so that

$$E[(S_n/n - E[X]) \geq \epsilon] \leq Var[X]/(n\epsilon 0 \to 0 .$$

Application: Bernoulli ¶$[S_n/n - p] \geq \epsilon] \leq p(1-p)/(n\epsilon^2)$.

Math 464/564, O. Knill, **February 1997**

Expectation and variance for discrete random variables

DEFINITION: Let X be a discrete random variable.
The **expectation** of X is $E[X] = \sum_{x \in X(\Omega)} x \cdot P[X = x]$.
The **variance** of X is $Var[X] = E[(X - E[X])^2]$.

SOME PROPERTIES OF EXPECTATION AND VARIANCE For random variables X, Y and $\lambda \in \mathbb{R}$

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) $E[X] = c\ if\ X(\omega) = c\ is\ constant$

f) $E[X - E[X]] = 0$

g) $Var[X] \geq 0$

h) $Var[X] = E[X^2] - E[X]^2$

i) $Var[\lambda X] = \lambda^2 Var[X]$

PROOF OF THE ABOVE PROPERTIES:

a) $E[X+Y] = \sum_{z \in (X+Y)(\Omega)} z \cdot P[X+Y = z] = \sum_{x \in X(\Omega), y \in Y(\Omega)} (x+y) P[X = x, Y = y]$
$= \sum_{x \in X(\Omega)} x P[X = x] \sum_{y \in Y(\Omega)} y P[Y = y] = E[X] + E[Y]$
b) $E[\lambda X] = \sum_{x \in X(\Omega)} x P[X = x] = \sum_{x \in X(\Omega)} \lambda x P[X = x] = \lambda E[X]$
c) $X \leq Y \Rightarrow X(\omega) \leq Y(\omega) \Rightarrow E[X] \leq E[Y]$
d) $E[X^2] = 0 \Leftrightarrow P[X^2 = x] = 0\ \forall x \Leftrightarrow X = 0$
e) $X(\omega) = c\ is\ constant \Rightarrow E[X] = c \cdot P[X = c] = c \cdot 1 = c$
f) $E[X - E[X]] = E[X] - E[E[X]] = E[X] - E[X] = 0$
g) $(X - E[X]) \geq 0 \Rightarrow Var[X] \geq 0$
h) $E[(X - E[X])^2] = E[X^2 - 2XE[X] + E[X]^2] = E[X^2] - 2E[X]E[X] + E[X]^2 = E[X^2] - E[X]^2$.
i) $E[\lambda^2 X^2] = \lambda^2 E[X^2]\ and\ E[\lambda X]^2 = \lambda^2 E[X]^2$

EXPRESSIONS FROM THE DISCRETE DENSITY FUNCTION. Let $f_X(x)$ be the discrete density function of X, then

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

$$Var[X] = \sum_{x \in X(\Omega)} x^2 f_X(x) - (\sum_{x \in X(\Omega)} x f_X(x))^2 .$$

EXPECTATION OF $h(X)$ like for example $h(X) = \sin(X)$.

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

EXAMPLES OF DISCRETE DISTRIBUTIONS:

Distribution	$f_X(k) = P[x = k] =$	Parameters	Domain	Expectation	Variance
Binomial	$\binom{n}{k} p^k (1-p)^{n-k}$	$n \in \mathbb{N}, p \in [0,1]$	$\{0, \ldots, n\}$	np	$np(1-p)$
Uniform	$1/n$	$n \in \mathbb{N}$	$\{1, \ldots, n\}$	$(1+n)/2$	$(n^2-1)/12$
Poisson	$\frac{\lambda^k}{k!} e^{-\lambda}$	$\lambda > 0$	$\{0, 1, \ldots\}$	λ	λ
Geometric	$(1-p)^k p$	$p \in (0,1)$	$\{0, 1, \ldots\}$	$(1-p)/p$	$(1-p)/p^2$

Math 464/564, O. Knill, **March 1997**

Probability generating functions

Probability generating functions are useful for computing distributions of sums of independent random variables as well as to compute expectation and variance or higher moments $E[X^n]$. Useful are also the moment generating function $E[e^{-tX}]$ or the characteristic function $E[e^{itX}]$ which we will see later.

DEFINITION: Let X be a discrete random variable taking values in $\mathbb{N} = \{0, 1, \ldots\}$. The **probability generating function** of X is $\phi_X(t) = \sum_{n=0}^{\infty} P[X = n]t^n = \sum_{n=0}^{\infty} f_X(n)t^n$.

THE SUM OF TWO INDEPENDENT RANDOM VARIABLES.
If X, Y are independent, then $\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t)$.

PROOF OF THIS FACT.

$$f_{X+Y}(n) = P[X+Y = n] = \sum_k P[X = k, Y = n-k] = \sum_k P[X = k]P[Y = n-k] = \sum_k f_X(k)f_Y(n-k)$$

and we obtain

$$\phi_X(t)\phi_Y(t) = (\sum_k f_X(k)t^k)(\sum_m f_Y(m)t^m) = \sum_n (\sum_k f_X(k)f_Y(n-k))\, t^n = \sum_n f_{X+Y}(n)\, t^n$$

COMPUTATION OF EXPECTATION AND VARIANCE.

$$E[X] = \phi_X'(1) .$$

$$Var[X] = \phi_X''(1) + \phi_X'(1)(1 - \phi_X'(1)) .$$

PROOF OF THESE FACTS.

$$\phi_X(t) = \sum_n P[X = n]\, t^n .$$

Differentiation gives

$$\phi_X'(t) = \sum_n nP[X = n]\, t^{n-1} .$$

Evaluate at $t = 1$ gives the formula for the expectation.
Again differentiation gives

$$\phi_X''(t) = \sum_n n(n-1)P[X = n]\, t^{n-2} .$$

For $t = 1$, we obtain

$$\phi_X''(1) = E[X^2] - E[X] .$$

Adding $E[X] - E[X]^2$ gives the variance.

PROBABILITY GENERATING FUNCTIONS OF SOME DISCRETE DISTRIBUTIONS:

Distribution	$f_X(k) = P[X = k] =$	Parameters	Domain	Probability generating function
Binomial	$\binom{n}{k} p^k(1-p)^{n-k}$	$n \in \mathbb{N}, p \in [0,1]$	$\{0, \ldots, n\}$	$(pt + 1 - p)^n$
Uniform	$1/n$	$n \in \mathbb{N}$	$\{1, \ldots, n\}$	$\frac{(t-t^{n+1})}{n(1-t)}$
Poisson	$\frac{\lambda^k}{k!}e^{-\lambda}$	$\lambda > 0$	$\{0, 1, \ldots\}$	$e^{\lambda(t-1)}$
Geometric	$(1-p)^k p$	$p \in (0,1)$	$\{0, 1, \ldots\}$	$p/(1 - (1-p)t)$

Math 464/564, O. Knill, **March 1997**

An application of the weak law of large numbers

DEFINITION.
Denote by $C[0,1]$ the set of all continuous functions on the interval $[0,1]$. For two functions f,g in $C[0,1]$ we define the distance $d(f,g) = \sup_{x\in[0,1]} |f(x)-g(x)|$. The set $C[0,1]$ with this distance is a metric space. Given a sequence f_n in $C[0,1]$ satisfying $d(f_n,f) \to 0$, one says f_n **converges uniformly** to f and writes $f_n \to f$. A set $A \subset C[0,1]$ is called dense, if for every $f \in C[0,1]$ there exists a sequence $f_n \in A$ with $f_n \to f$.

The weak law of large numbers provides an elegant and constructive proof of the theorem of Weierstrass.

THEOREM OF WEIERSTRASS.
Polynomials are dense in $C[0,1]$: more precicely, given $f \in C[0,1]$, the **Bernstein polynomials**

$$B_n(x) = \sum_{k=1}^{n} f(\frac{k}{n}) \left(\begin{array}{c} n \\ k \end{array} \right) x^k (1-x)^{n-k}$$

converge uniformly to f. If $f(x) \geq 0$ for all x, then also $B_n(x) \geq 0$ for all x.

PROOF OF WEIERSTRASS.
For $x \in [0,1]$, let X_n be a sequence of independent $\{0,1\}$- valued random variables with expectation x. In other words, we take the probability space $(\{0,1\}^{\mathbb{N}}, \mathcal{A}, P)$ defined by $P[\omega_n = 1] = x$. Since $P[S_n = k] = \left(\begin{array}{c} n \\ k \end{array} \right) x^k (1-x)^{n-k}$, one has $B_n(x) = E[f(S_n/n)]$. Now

$$\begin{aligned} |B_n(x) - f(x)| &= |E[f(\frac{S_n}{n})] - f(x)]| \leq E[|f(\frac{S_n}{n}) - f(x)|] \\ &\leq 2||f|| \cdot P[|\frac{S_n}{n} - x| \geq \delta] + \sup_{|x-y|\leq\delta} |f(x)-f(y)| \cdot P[|\frac{S_n}{n} - x| < \delta] \\ &\leq 2||f|| \cdot P[|\frac{S_n}{n} - x| \geq \delta] + \sup_{|x-y|\leq\delta} |f(x)-f(y)| \quad . \end{aligned}$$

The second term on the right hand side goes clearly to zero for $\delta \to 0$. By the weak law of large numbers, the first term to the right is $\leq 2||f|| Var[X_i]/(n\delta^2)$ which goes to zero with $n \to \infty$ because $Var[X_i] = x(1-x) \leq 1/4$. Given $\epsilon > 0$, we make δ so small that $\sup_{|x-y|\leq\delta} |f(x)-f(y)| \leq \epsilon/2$ and then so n large that $||f||/(2n\delta^2) < \epsilon/2$ so that $|B_n(x) - f(x)| \leq \epsilon$ for all $x \in [0,1]$.

WEAK LAW OF LARGE NUMBERS.
Assume X_i have common expectation $E[X_i] = m$ and common variance $M = Var[X_i] < \infty$. If X_n are pairwise uncorrelated, then $P[|S_n/n - m| \geq \epsilon] \to 0$.

PROOF OF THE WEAK LAW.
Since $Var[X+Y] = Var[X] + Var[Y] + 2 \cdot Cov[X,Y]$ and X_n are pairwise uncorrelated, we get $Var[X_n + X_m] = Var[X_n] + Var[X_m]$ and by induction $Var[S_n] = \sum_{i=1}^{n} Var[X_n]$. Using the linearity of expectation, we obtain $E[S_n/n] = m$ and

$$Var[\frac{S_n}{n}] = E[\frac{S_n^2}{n^2}] - \frac{E[S_n]^2}{n^2} = \frac{Var[S_n]}{n^2} = \frac{1}{n^2} \sum_{i=1}^{n} Var[X_n] \leq \frac{M}{n} \; .$$

With this and Chebychev's inequality, we obtain

$$P[|\frac{S_n}{n} - m| \geq \epsilon] \leq \frac{Var[S_n/n]}{\epsilon^2} \leq \frac{M}{n\epsilon^2} \to 0 \; .$$

Math 464/564, O. Knill, **March 1997**

The Banach-Tarsky paradox

QUESTION.
Let $\mathcal{A}$ be the σ-algebra consisting of all subsets of a cube $\Omega \subset \mathbb{R}^3$. Is there a function $P : \mathcal{A} \to [0,1]$ such that $(\Omega, \mathcal{A}, P)$ is a probability space such that if two sets $Y, Z \subset \Omega$ which are congruent satisfy $P[Y] = P[Z]$?

There is a dramatic theorem showing that the answer to this question is no:

BANACH-TARSKY THEOREM (1924).
One can decompose a ball $Y \subset \Omega$ into 5 disjoint pieces Y_1, Y_2, Y_3, Y_4, Y_5 satisfying $\cup_{i=1}^5 Y_i = Y$ and rotate and move the 5 pieces in Ω in such a way that two of the moved pieces Y_1', Y_2' fit together to a ball with the same radius than the ball Y and that the remaining three moved pieces Y_3', Y_4', Y_5' fit together to a second ball with the same radius than Y.

This gives the answer to the question because it follows from the axioms for P that $P[Y_1 \cup Y_2]Y_3 + Y_4 + Y_5] = P[Y_1 + Y_2 + Y_3 + Y_4 + Y_5] = P[Y]$ and $P[Y_1 \cup Y_2] = P[Y_1] + P[Y_2] = P[Y_1'] + P[Y_2'] = P[Y]$ as well as $P[Y_3 \cup Y_4 \cup Y_5] = P[Y_3] + P[Y_4] + P[Y_5] = P[Y_3'] + P[Y_4'] = P[Y_5'] = P[Y]$. However $P[Y] + P[Y] = P[Y]$ implying $P[Y] = 0$. Since Ω can be covered with finitely many translated pieces Y or part of such pieces also $1 = P[\Omega] = 0$. This contradiction shows and the assumption on the existence of P can not hold.
Remark. 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 $\mathbb{R}^d$ and such that $P[R(A)] = P[A]$ for any rotation or translation R and $P[\prod_{i=1}^n [a_i, b_i]] = \prod_i (b_i - a_i)$. So, there exists no Banach-Tarsky paradox in one or two dimensions.

TO THE PROOF.
The proof uses the axiom of choice which is part of the axiom system ZFC (Zermelo-Fraenkel with Axiom of Choice) all conventional mathematics relies on. (There is a branch of mathematical logic where the game is to figure out, what one can do without the axiom of choice. However, since most models for nature we have use heavily the axiom of choice in its foundations (for example in quantum mechanics through probability theory or functional analysis) nobody would like to live without this axiom.)

A proof of the paradox can be found in the article K. Stromberg, Americ. Math. Mongthly 86 (1979) p. 151-161. There is also a book about the paradox: S. Wagon, The Banach-Tarski Paradox, Cambridge 1985

VITALI'S EXAMPLE (1905).
Let $\Omega = [0,1]$. Call $x, y \in \Omega$ equivalent if $x - y$ is a rational number. Let $\mathcal{K}$ be the set of equivalence classes with respect to this relation. The axiom of Choice assures that one can choose for each equivalence class K an element k. That is, there exists a "choice function" $f : \mathcal{K} \to \Omega$ with $f(K) \in K$. Define $Y = f(\mathcal{K})$. Let $R = \{r_n\}$ be an enumeration of the rational numbers in Ω and define $Y_n = (r_n + Y)$. Then $[0,1] = (\bigcup_n Y_n) \cap [0,1]$.

Also this result shows that the σ-algebra has to be smaller. Vitali's result is less spectacular because of a decomposition into countabley many elements. Because the sets Y_n are all translations to each other, $P[Y_n] = a$ is independent of a. $\sum_n a = \sum_n P[Y_n] = P[\bigcup_n Y_n] = P[\Omega] = 1$ implies $a = 0$ in which case also $1 = P[\Omega] = 0$ which is not possible.

CONCLUSION.
In probability theory, if Ω is uncountable, we have to take the σ-algebra smaller than the set of all subsets.

Math 464/564, O. Knill, **March 1997**

A paradox in computing a surface integral

One might believe that one can compute the surface area by triangulating the surface into finer and finer pieces and take the limit of these areas. This is wrong and there is a famous **example of Schwarz**:

Take the cylinder S of radius 1 and height 1. Assume S is parametrized by $[0, 2\pi] \times [0, 1]$. A subdivision of B into n^4 squares Q_i is obtained by subdividing for every n the interval $[0, 1]$ by n^3 intervals and the interval $[0, 2\pi]$ into n intervals. Construct a triangulation with $4n^4$ triangles by taking also the images of the midpoints of Q_i. The area of each triangle is $\geq 2\sin(\pi/n)\sin^2(\pi/2n) \geq 4/n^3$ and the area of the triangulation goes to ∞ of the order n.

CARATHEODORY EXTENSION THEOREM

The construction of general probability spaces needs some measure theory. If you are interested into the foundations of probability theory, here is the tool which is necessary for constructing general probability spaces like the unit interval with the Lebesgue measure. The idea is to define the measure first on a Boolean algebra one extend this to a σ-algebra. For example, one first defines the Lebesgue measure on sets which are finite unions of disjoint intervals $[a_i, b_i)$ by $\mu([a_i, b_i)) = b_i - a_i$. The Carathéodory extension theorem (proven here) assures then that this can be extended consistently to a σ-algebra of sets containing these intervals.

REPETITION DEFINITION.
A set of subsets of Ω $\mathcal{A}$ is an **Boolean algebra** if $\Omega \in \mathcal{A}$,
$A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$,
$A, B \in \mathcal{A} \Rightarrow A \cup B \in \mathcal{A}$.
If $A_n \in \mathcal{A} \Rightarrow \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{A}$, then $\mathcal{A}$ is called a **σ-algebra** and the pair $(\Omega, \mathcal{A})$ is called a **measurable space.**

DEFINITION. A σ-additive map from a Boolean algebra $\mathcal{A}$ to $[0, \infty)$ is called a **measure**.

CARATHEODORY CONTINUATION THEOREM. Any measure on a Boolean algebra $\mathcal{R}$ can be continued uniquely to a measure on $\sigma(\mathcal{R})$, the smallest σ-algebra containing $\mathcal{R}$.

DEFINITION. Let $\mathcal{A}$ be a Boolean algebra and let $\lambda : \mathcal{A} \mapsto [0, \infty]$ be a measure with $\lambda(\emptyset) = 0$. A set $A \in \mathcal{A}$ is called a **λ-set**, if $\lambda(A \cap G) + \lambda(A^c \cap G) = \lambda(G)$ for all $G \in \mathcal{A}$.

LEMMA. The set $\mathcal{L}$ of λ-sets of a Boolean algebra $\mathcal{A}$ is again a Boolean algebra and satisfies $\sum_{k=1}^n \lambda(A_k \cap G) = \lambda((\bigcup_{k=1}^n A_k) \cap G)$ for all finite disjoint families $\{A_k\}_{k=1}^n$ and all $G \in \mathcal{A}$.

PROOF OF THE LEMMA. From the definition, it is clear that $\Omega \in \mathcal{L}$ and that if $B \in \mathcal{L}$, then $B^c \in \mathcal{L}$. Given $B, C \in \mathcal{L}$. Then $A = B \cap C \in \mathcal{L}$. Proof. Since $C \in \mathcal{L}$, we get

$$\lambda(C \cap A^c \cap G) + \lambda(C^c \cap A^c \cap G)u = \lambda(A^c \cap G)$$

This can be rewritten with $C \cap A^c = C \cap B^c$ and $C^c \cap A^c = C^c$ as

$$\lambda(A^c \cap G) = \lambda(C \cap B^c \cap G) + \lambda(C^c \cap G) \ . \tag{1}$$

Since B is a λ-set, we get using $B \cap C = A$.

$$\lambda(A \cap G) + \lambda(B^c \cap C \cap G) = \lambda(C \cap G) \ . \tag{2}$$

Since C is a λ set, we have

$$\lambda(C \cap G) + \lambda(C^c \cap G) = \lambda(G) \ . \tag{3}$$

Adding up these three equations gives that $B \cap C$ is a λ set. We have shown that Λ is a Boolean algebra. If B and C are disjoint in $\mathcal{L}$ we get since B is a λ set

$$\lambda(B \cap (B \cup C) \cap G) + \lambda(B^c \cap (B \cup C) \cap G) = \lambda((B \cup C) \cap G) \ .$$

This can be rewritten as $\lambda(B \cap G) + \lambda(C \cap G) = \lambda((B \cup C) \cap G)$. The analogue statement for finitely many sets is obtained by induction.

DEFINITION. Let $\mathcal{A}$ be a σ-algebra. A map $\lambda : \mathcal{A} \Rightarrow [0, \infty]$ is called an **outer measure**, if

$\lambda(\emptyset) = 1$,
$A, B \in \mathcal{A}$ with $A \subset B \Rightarrow \lambda(A) \leq \lambda(B)$.
$A_n \in \mathcal{A} \ \Rightarrow \lambda(\bigcup_n A_n) \leq \sum_n P(A_n)$ (σ subadditivity)

CARATHEODORY LEMMA. If λ is an outer measure on a measurable space $(\Omega, \mathcal{A})$, then the λ sets $\mathcal{L} \subset \mathcal{A}$ form a σ-algebra on which λ is countably additive.

PROOF OF THE CARATHEODORY LEMMA. Given a disjoint sequence $A_n \in \mathcal{L}$. We have to show that $A = \bigcup_n A_n \in \mathcal{L}$ and $\lambda(A) = \sum_n \lambda(A_n)$. By the above lemma, $B_n = \bigcup_{k=1}^n A_k$ is in $\mathcal{L}$. We have therefore using the monotonicity, the additivity proved in the above lemma and the σ-subadditivity

$$\begin{aligned}\lambda(G) &= \lambda(B_n \cap G) + \lambda(B_n^c \cap G) \geq \lambda(B_n \cap G) + \lambda(A^c \cap G) \\ &= \sum_{k=1}^n \lambda(A_k \cap G) + \lambda(A^c \cap G) \geq \lambda(A \cap G) + \lambda(A^c \cap G) .\end{aligned}$$

Subadditivity for λ gives $\lambda(G) \leq \lambda(A \cap G) + \lambda(A^c \cap G)$. All the inequalities in this proof are therefore equalities. We conclude that $A \in \mathcal{L}$ and that λ is σ additive on $\mathcal{L}$.

PROOF OF THE CARATHEODORY CONTINUATION THEOREM.
Given an algebra $\mathcal{R}$ with a measure μ. Define $\mathcal{A} = \sigma(\mathcal{R})$ and the σ-algebra $\mathcal{P}$ consisting of all subsets of Ω. Define on $\mathcal{P}$ the function

$$\lambda(A) = \inf\{\sum_{n \in \mathbb{N}} \mu(A_n) \mid \{A_n\}_{n \in \mathbb{N}} \; sequence \; in \mathcal{R} \; satisfying \; A \subset \bigcup_n A_n\} .$$

(i) λ is an outer measure on $\mathcal{P}$.
$\lambda(\emptyset) = 0$ and $\lambda(A) \geq \lambda(B)$ for $B \geq A$ are obvious. To see the σ-subadditivity take a sequence $A_n \in \mathcal{P}$ with $\lambda(A_n) < \infty$ and fix $\epsilon > 0$. For all $n \in \mathbb{N}$, one can by the definition of λ find a sequence $\{B_{n,k}\}_{k \in \mathbb{N}}$ in $\mathcal{R}$ such that $A_n \subset \bigcup_{k \in \mathbb{N}} B_{n,k}$ and $\sum_{k \in \mathbb{N}} \mu(B_{n,k}) \leq \lambda(A_n) + \epsilon 2^{-n}$. Define $A = \bigcup_{n \in \mathbb{N}} A_n \subset \bigcup_{n,k \in \mathbb{N}} B_{n,k}$, so that $\lambda(A) \leq \sum_{n,k} \mu(B_{n,k}) \leq \sum_n \lambda(A_n) + \epsilon$. Since ϵ was arbitrary, the σ-subadditivity is proven.

(ii) $\lambda = \mu$ on $\mathcal{R}$.
Given $A \in \mathcal{R}$. Clearly $\lambda(A) \leq \mu(A)$. Suppose that $A \subset \bigcup_n A_n$, with $A_n \in \mathcal{R}$. Define a sequence $\{B_n\}_{n \in \mathbb{N}}$ of disjoint sets in $\mathcal{R}$ defined inductivly by $B_1 = A_1$, $B_n = A_n \cap (\bigcup_{k<n} A_k)^c$ such that $B_n \subset A_n$ and $\bigcup_n B_n = \bigcup_n A_n \supset A$. From the σ-additivity of μ on $\mathcal{R}$, we get

$$\mu(A) \leq \bigcup_n \mu(B_n) \leq \bigcup_n \mu(A_n)$$

so that $\mu(A) \geq \lambda(A)$.

(iii) Let $\mathcal{L}$ be the set of λ-sets in $\mathcal{P}$. Then $\mathcal{R} \subset \mathcal{L}$.
Given $A \in \mathcal{R}$ and $G \in \mathcal{P}$. There exists a sequence $\{B_n\}_{n \in \mathbb{N}}$ in $\mathcal{R}$ such that $G \subset \bigcup_n B_n$ and $\sum_n \mu(B_n) \leq \lambda(G) + \epsilon$. By the definition of λ

$$\sum_n \mu(B_n) = \sum_n \mu(A \cap B_n) + \sum_n \mu(A^c \cap B_n) \geq \lambda(A \cap G) + \lambda(A^c \cap G)$$

because $A \cap G \subset \bigcup_n A \cap B_n$ and $A^c \cap G \subset \bigcup_n A^c \cap B_n$. Since ϵ is arbitary, we get $\lambda(A) \geq \lambda(A \cap G) + \lambda(A^c \cap G)$. On the other hand, since λ is subadditive, we have also $\lambda(A) \leq \lambda(A \cap G) + \lambda(A^c \cap G)$ and A is a λ-set.

(iv) By Carathéodory's lemma, λ is a measure on $(\Omega, \mathcal{L})$. Since by step *(iii)* $\mathcal{R} \subset \mathcal{L}$, we know by Caratheodories lemma that $\mathcal{A} \subset \mathcal{L}$ so that we can define μ on $\mathcal{A}$ as the restriction of λ to $\mathcal{A}$. By step *(ii)*, this is an extension of the measure μ on $\mathcal{R}$.

Math 464/564, O. Knill, **March 1997**

Integration

In order to define the expectation for general random variables, one needs integration $E[X] = \int_\Omega X(\omega) dP(\omega)$. For $\Omega = \mathbb{R}^n$ and $dP = \prod_i dx_i$, this is the usual (multiple) integral in calculus.

DEFINITION. Let $(\Omega, \mathcal{A}, P)$ be a proability space. Let $\mathcal{B}$ be $\sigma-$ algebra of all measurable sets on $\mathbb{R}$. A function $X : \Omega \to \mathbb{R}$ 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 $\Omega = \mathbb{R}$ and $\mathcal{A} = \mathcal{B}$ the Borel σ-algebra. A continuous function $X : \Omega \to \mathbb{R}$ is a random variable.

DEFINITION. A **step function** is a random variable which is of the form $X = \sum_{i=1}^n \alpha_i \cdot 1_{A_i}$ with $\alpha_i \in \mathbb{R}$ 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*

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

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**

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

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

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

3) Let $\Omega = [a, b]$ with $a < b$, $\mathcal{A}$ the Borel σ algebra and $P[(c, d)] = (d - c)/(b - a)$.

Math 464/564, O. Knill, **April 1997**

Absolutely continuous distributions

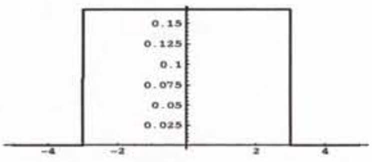
REMINDER: The **Distribution function** of a random variable X is $F_X(t) = P[X \leq t]$. A random variable is **absolutely continuous** if the density $F'_X = f_X$ exists. The **expectation**, **variance** and $E[g(X)]$ for $g(X)$ is in the continuous case

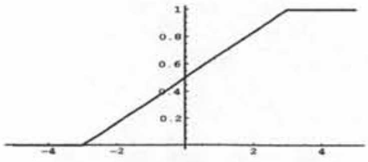
$$m = E[X] = \int_{-\infty}^{\infty} x f(x)\, dx,\ Var[X] = \int_{-\infty}^{\infty} (x-m)^2 f(x)\, dx\ , E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x)\, dx$$

UNIFORM DISTRIBUTION. $\Omega = [a,b]$.

$$\boxed{f(x) = (b-a)^{-1}}$$

The natural distribution on a finite interval $[a,b]$. Example: angle of a weel of a parking bike on campus is uniformly distributed on $[0, 2\pi)$.

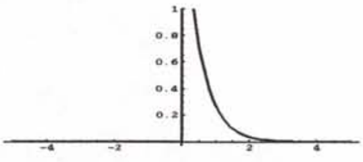


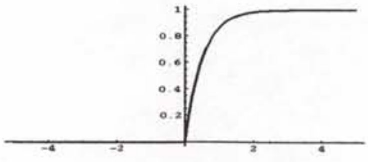


EXPONENTIAL DISTRIBUTION. $\Omega = [0, \infty)$.

$$\boxed{f(x) = \lambda e^{-\lambda x}}$$

The natural distribution on the positive real axes. Example: the decay times of radioactive isotopes is exponentially distributed

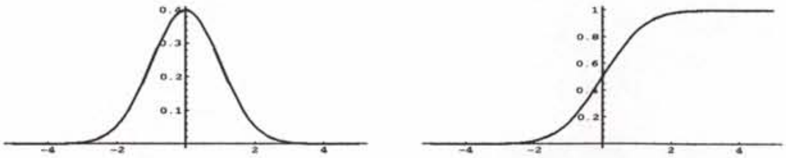




NORMAL DISTRIBUTION. $\Omega = \mathbb{R}$

$$\boxed{f(x) = (2\pi\sigma^2)^{-1/2} e^{-(x-\mu)^2/2\sigma^2}.}$$

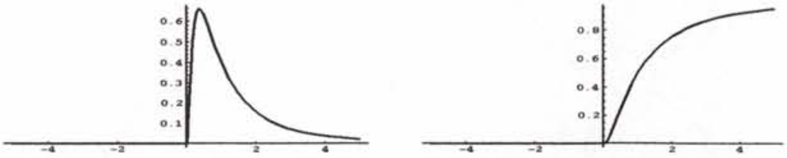
The natural distribution on the real line. Example: measurement errors in scientific experiments.



LOG-NORMAL DISTRIBUTION. $\Omega = [0, \infty)$.

$$\boxed{f(x) = \qquad\qquad\qquad \text{(see HW 10)}}$$

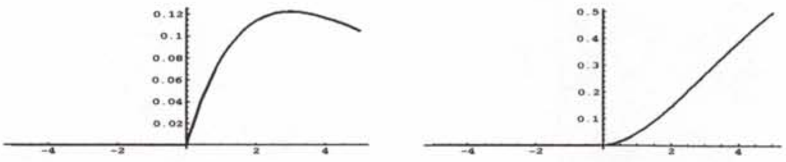
The exponential of a normal distribued random variable is log-Normal distributed. (The terminology is because the LOGarithm of a log-normal distribued random variable has a NORMAL distribution).



GAMMA DISTRIBUTION. $\Omega = [0, \infty)$. The function $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x}\, dx$ is called the **Gamma function**. For integers $\Gamma(n) = (n-1)!$.

$$\boxed{f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}.}$$

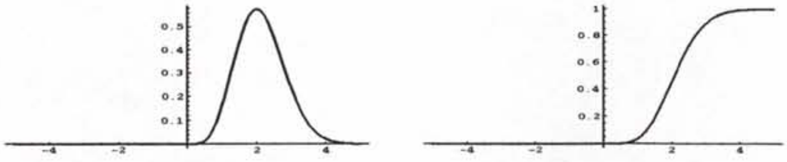
Useful in population statistics. The square of a standard normal distributed random variable is an example of a Gamma distributed random variable.



χ^2-DISTRIBUTION. $\Omega = [0, \infty)$.

$$\boxed{f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}.}$$

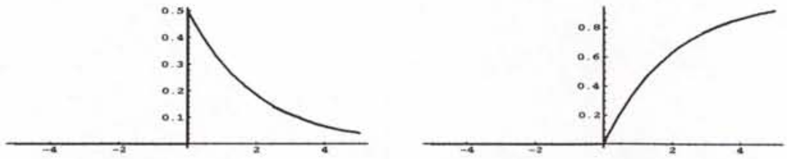
Special case of Gamma distribution. The sum of n squares of independent standard normal distributed random variable is χ^2-distributed. A Γ distributed random variables with parameter $\alpha = n/2, \lambda = 1/2$ is χ^2 distributed.



WEIBULL-DISTRIBUTION. $\Omega = [0, \infty)$.

$$\boxed{f(x) = \lambda \alpha x^{\alpha-1} e^{-\lambda x^\alpha}.}$$

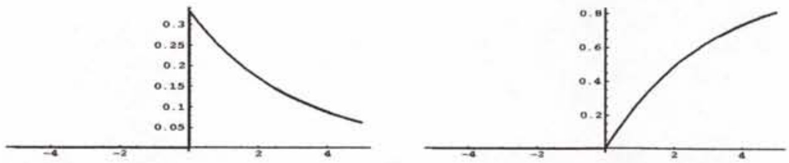
Generalisation of the exponential distribution which is the special case $\alpha = 1$. Is useful to model life-times of items.



ERLANG-DISTRIBUTION. $\Omega = [0, \infty)$.

$$\boxed{f(x) = \frac{\lambda^k}{(k-1)!} x^{k-1} e^{-\lambda x} \;,}$$

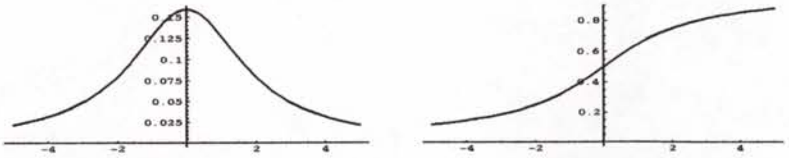
Special case of Γ distribution for integer values of α. The sum of n independent exponential distribued random variables is Erlang distributed.



CAUCHY DISTRIBUTION. $\Omega = \mathbb{R}$.

$$\boxed{f(x) = \frac{1}{\pi(1+x^2)}.}$$

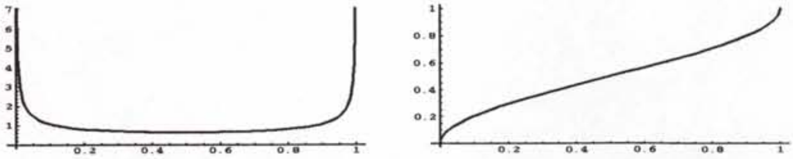
The tangent of a uniformly distributed angle is Cauchy distributed.



ARC-SIN DISTRIBUTION. $\Omega = \mathbb{R}$.

$$\boxed{f(x) = \frac{1}{\pi\sqrt{x(1-x)}}.}$$

The spectral density of a free particle in a pure crystal has the arc-sin distribution.



Mathematica program which generated all the pictures on these pages:

```
<<Statistics`ContinuousDistributions`;
PlotPDFandCDF[distr_]:=Show[GraphicsArray[
      {Plot[PDF[distr, x],{x,-5,5},
          DisplayFunction->Identity],
       Plot[CDF[distr, x],{x,-5,5},
          DisplayFunction->Identity]},
      DisplayFunction->$DisplayFunction]];
PlotPDFandCDFArcSin:=
Show[GraphicsArray[
      {Plot[1/(N[Pi] Sqrt[x(1-x)]),{x,0.0000001,1-0.0000001},
          DisplayFunction->Identity],
       Plot[2/N[Pi] ArcSin[Sqrt[x]],{x,0,1},
          DisplayFunction->Identity]},
      DisplayFunction->$DisplayFunction]];
ErlangDistribution=GammaDistribution;

Display["!psfix -land>cdist01.ps",PlotPDFandCDF[UniformDistribution[-3,3]]];
Display["!psfix -land>cdist02.ps",PlotPDFandCDF[ExponentialDistribution[2]]];
Display["!psfix -land>cdist03.ps",PlotPDFandCDF[NormalDistribution[0,1]]];
Display["!psfix -land>cdist04.ps",PlotPDFandCDF[LogNormalDistribution[0,1]]];
Display["!psfix -land>cdist05.ps",PlotPDFandCDF[GammaDistribution[2,3]]];
Display["!psfix -land>cdist06.ps",PlotPDFandCDF[ChiDistribution[5]]];
Display["!psfix -land>cdist07.ps",PlotPDFandCDF[WeibullDistribution[1,2]]];
Display["!psfix -land>cdist08.ps",PlotPDFandCDF[ErlangDistribution[1,3]]];
Display["!psfix -land>cdist09.ps",PlotPDFandCDF[CauchyDistribution[0,2]]];
Display["!psfix -land>cdist10.ps",PlotPDFandCDFArcSin];
```

Math 464/564, O. Knill, **April 1997**

Multidimensional distributions (Summary I)

DEFINITION. Given n random variables $X_1, \dots, X_n$. The map $X = (X_1, \dots, X_n) : \Omega \to \mathbb{R}^n$ is called a **random vector**.

DEFINITION. The **joint distribution function** of $X = (X_1, X_2, \dots, X_n)$ is the function F on $\mathbb{R}^n$:

$$F(x_1, \dots, x_n) = P[X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n] .$$

DEFINITION. Like in the one dimensional case, a distribution function is called **discrete** if there exists a countable set $C \subset \mathbb{R}^d$ such that $P[X \in C] = 1$. It is called **absolutely continuous** if there exists a function $f : \mathbb{R}^n \to \mathbb{R}^+$ such that

$$F(t_1, \dots t_n) = \int_{-\infty}^{t_1} \int_{-\infty}^{t_1} \dots \int_{-\infty}^{t_n} f(y_1, y_2, \dots, y_n) \, dy .$$

The function $f : \mathbb{R}^n \to \mathbb{R}$ is called the **joint density**.

LEMMA. One can recover a probability space $(\mathbb{R}^n, \mathcal{A}, P)$ given a joint distribution function F. One has for example in the case $n = 2$ the formula $P[(a, b] \times (c, d]] = F(b, d) + F(a, c) - F(a, d) - F(a, c)$.

LEMMA. If F is a absolutely continuous distribution function. The density is

$$f(x) = F_{12,\dots,n}(x) = \frac{\partial}{\partial x_1} \cdots \frac{\partial}{\partial x_n} F(x) .$$

For $n = 2$ one has especially $f(x, y) = \frac{\partial}{\partial x} \frac{\partial}{\partial y} F(x, y)$.

DEFINITION. Given a joint distribution function F of the random vector, then $F_{X_i}(x) = F(\infty, \infty, \dots, x, \dots, \infty)$ is called the $i'th$ **marginal distribution** of X. It is the distribution function of the random variable X_i because $P[X_i \leq x, X_j \leq \infty, \forall j \neq i] = P[X_i \leq x]$.

REMARK. Given an absolutely continuous random vector with joint density f, then the density of X_i is

$$f_{X_i}(x) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{x} \dots \int_{-\infty}^{\infty} f(x_1, x_2, \dots, x_{i-1}, x, x_{i+1}, \dots, x_d) \, dx_1, \dots dx_n .$$

(Integrate over all variables except over the variable x_i). For example, for $n = 2$ one has for the random vector (X, Y) the formulas $F_X(x) = \int_{-\infty}^{\infty} f(x, y) \, dy$ and $F_Y(y) = \int_{-\infty}^{\infty} f(x, y) \, dx$.

DEFINITION. Two random variables are called **independent** if

$$(*) \qquad P[X \in [a, b], Y \in [c, d]] = P[X \in [a, b]] \cdot P[Y \in [c, d]] .$$

THEOREM. X and Y are independent if and only if

$$(**) \qquad F(x, y) = F_X(x) F_Y(y) .$$

(**) follows from (*) for $a = c = -\infty$. Using $F(a, b) = F(-\infty, b) - F(-\infty, a)$ and $F(c, d) = F(-\infty, d) - F(-\infty, c)$, one obtains the other direction $(**) \Rightarrow (*)$.

COROLLARY. Assume X and Y are absolutely continuous. By differentiation of (**), one obtains the important fact:
X and Y are independent if and only if $f(x, y) = f_X(x) f_Y(y)$.

Math 464/564, O. Knill, **April 1997**

Stirling's formula

The Stirling's formula can be derived using the central limit theorem.

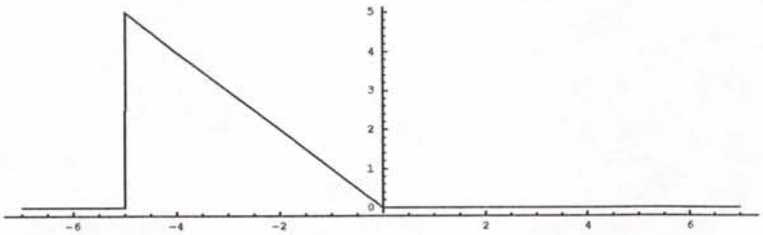
STIRLING'S FORMULA:

$$\boxed{n! \sim n^{n+\frac{1}{2}} e^{-n} \sqrt{2\pi}}$$

PROOF OF STIRLING'S FORMULA.

Let X_i be IID random variables, which are Poisson distributed with parameter $\lambda = 1$. That is $P[X_i = k] = \frac{1}{k!} e^{-1}$. We know that the sum $S_n = \sum_{j=1}^{n} X_j$ is Poisson distributed with parameter $\lambda = k$. That is $P[S_n = k] = \frac{n^k}{k!} e^{-n}$.

Let $g_M(x)$ be the function on $\mathbb{R}$ which is $g_M(x) = -x$ for $x \in [-M, 0]$ and 0 else.



We compute

$$\begin{aligned}
E[g_M(\frac{S_n - n}{\sqrt{n}})] &= \sum_{n-M\sqrt{n} \leq j \leq n} g(\frac{j-n}{\sqrt{n}}) \frac{n^j e^{-n}}{j!} \\
&= \sum_{n-M\sqrt{n} \leq j \leq n} \frac{n-j}{\sqrt{n}} \frac{n^j e^{-n}}{j!} \\
&= \sum_{n-M\sqrt{n} \leq j \leq n} \frac{e^{-n}}{\sqrt{n}} \left(\frac{n^{j+1}}{j!} - \frac{n^j}{(j-1)!} \right) \\
&= \frac{e^{-n}}{\sqrt{n}} \left(\frac{n^{n+1}}{n!} - \frac{n^{[n-M\sqrt{n}+1]}}{[n - M\sqrt{n} + 1]} \right) .
\end{aligned}$$

(The square bracket $[r]$ for a real number detnotes the largest integer smaller or equal to r.).
We see that for $M \to \infty$,

$$E[g_M(S_n^*)] = E[g_M(\frac{S_n - n}{\sqrt{n}})] \to \frac{e^{-n}}{\sqrt{n}} \frac{n^{n+1}}{n!} = n^{n+\frac{1}{2}} e^{-n} \frac{1}{n!} . \qquad (1)$$

The central limit theorem implies that $E[g_M(S_n^*)] \to E[g_M(X)]$ for $n \to \infty$, where X is a standard normal distributed random variable. Because

$$E[g_M(X)] = \int_{-M}^{0} x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \, dx = \frac{1}{\sqrt{2\pi}} (1 - e^{-\frac{1}{2}M^2}) ,$$

we see that for $M \to \infty$,

$$E[g_M(X)] \to \frac{1}{\sqrt{2\pi}} \qquad (2)$$

Putting the facts (??) and (??) together with $E[g_M(S_n^*)] \to E[g_M(X)]$ gives the Stirling formula.

Math 464/564, O. Knill, **April 1997**

Characteristic functions

DEFINITION. The **characteristic function** of a random variable X is defined as $\phi_X(t) = E[e^{itX}] = \int_\Omega e^{itx}\, dP(x)$. Mathematicians call it the **Fourier transform** of the measure P.

CASE: If X takes integer values, then $\phi_X(t) = \sum_{n \in \mathbb{Z}} P[X = n] e^{inx}$ is a 2π-periodic function. The numbers $a_n = P[X = n]$ are the **Fourier coefficients** of the function $\phi_X(t)$.
CASE: If X is absolutely continuous with density f_X, then $\phi_X(t) = \int_{-\infty}^{\infty} e^{itx} f_X(x)\, dx$ is the **Fourier transform** of the density function f_X. Outside probability theory, one uses the notation $\hat{f}_X$ for the Fourier transform.

FOURIER INVERSION: The characteristic function determines P: if a, b are continuity points of F_X, then

$$F_X(b) - F_X(a) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ita} - e^{-itb} .$$

CASE: If X takes integer values, then

$$P[X = k] = \frac{1}{2\pi} \int_0^{2\pi} e^{-itk} \phi_X(t)\, dt .$$

CASE: If X is absolutely continuous with density f_X, then

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \phi_X(t)\, dt .$$

PROPERTIES:
SUM OF INDEPENDENT RANDOM VARIABLES: If X and Y are independent, then $\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t)$.
(The Fourier transform renders convolutions into multiplications: $\hat{f} \cdot \hat{g} = \widehat{f \star g}$.)
CALCULATION OF MOMENTS:

$$E[X^k] = (-i)^k \phi_X^{(k)}(0) .$$

LINEAR CHANGE OF VARIABLES:

$$\phi_{aX+b}(t) = e^{ib} \phi_X(at) .$$

RIEMANN-LEBESGUE: If X is absolutely continuous, then $\phi_X(t) \to 0$ for $|t| \to \infty$.
FOURIER-SERIES: If X takes integer values, then $\phi_X(t)$ is 2π-periodic.

EXAMPLES.

Distribution	Parameter	Characteristic function
Normal	$m \in \mathbb{R}, \sigma^2 > 0$	$e^{mit - \sigma^2 t^2/2}$
Standard normal	$m = 0, \sigma = 1$	$e^{-t^2/2}$
Uniform	$[-a, a]$	$\sin(at)/(at)$
Exponential	$\lambda > 0$	$\lambda/(\lambda - it)$
Gamma	$\beta > 0, \lambda$	$\lambda^\beta/(\lambda - it)^\beta$
Cauchy	$f_X(x) = 1/\pi(1 + x^2)$	$\exp^{-\lvert\theta\rvert}$.
Binomial	$n \in \mathbb{N}, p \in [0, 1]$	$(1 - p + pe^{it})^n$
Poisson	$\lambda > 0,\ \lambda$	$e^{\lambda(e^{it} - 1)}$
Geometric	$p \in (0, 1)$	$\frac{pe^{it}}{1-(1-p)e^{it}}$
Uniform	$\{1, 2, \ldots, n\}$	$\frac{1}{n} \frac{e^{i(n+1)t} - e^{it}}{e^{it} - 1}$

Math 464/564, O. Knill, **April 1997**

Proof of the CLT using characteristic functions

DEFINITIONS. A sequence of random variables X_n **converges in distribution** to X if $P[X_n \leq t] \to P[X \leq t]$ for all $t \in \mathbb{R}$.

DEFINITION. Recall that the **characteristic function** of X is defined as $\phi_X(t) = E[e^{itX}]$.
Discrete case:

$$\phi_X(t) = \sum_{a \in X(\Omega)} e^{ita} P[X = a] .$$

Continuous case:

$$\phi_X(t) = \int_{-\infty}^{\infty} e^{itx} f_X(x)\, dx .$$

THE CENTRAL LIMIT THEOREM FOR IID RANDOM VARIABLES. Assume X_n are IID random variables with nonzero finite variance. Then S_n^* converges in distribution to a standard normal distributed random variable.

LEMMA. X_n converges in distribution to X if $\phi_{X_n}(t) \to \phi_X(t)$ for all t.

PROOF. This follows from the fact that the distribution function F_X is determined from ϕ_X by

$$F_X(b) - F_X(a) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-ita} - e^{-itb}}{it} \phi_X(t)\, dt$$

and that most a, b are continuity points for all X_n and X.

PROOF OF THE CENTRAL LIMIT THEOREM.
After a replacement of X_n by $X_n - E[X_n]$ we can assume that X_n have zero mean.

Because the characteristic function of $N(0,1)$ is $e^{-t^2/2}$ and becaause of the above lemma, we have only to show that for all $t \in \mathbb{R}$

$$E[e^{it\frac{S_n}{\sigma\sqrt{n}}}] \to e^{-t^2/2} .$$

Denote by ϕ the characteristic function of X_n (it is independent of n because of the IID assumption). Since $\phi'(t) = iE[X] = 0, \phi''(t) = -E[X^2] = -Var[X] = \sigma^2$, the Taylor expansion of $\phi(t)$ at $t = 0$ gives

$$\phi(t) = 1 - \frac{\sigma^2}{2} t^2 + o(t^2) .$$

Using the two properties $\phi(S_n)(t) = \phi(t)^n$ and $\phi_{aX}(t) = \phi_X(at)$ of characteristic functions, we have

$$\begin{aligned} E[e^{it\frac{S_n}{\sigma\sqrt{n}}}] &= \phi(\frac{t}{\sigma\sqrt{n}})^n \\ &= (1 - \frac{1}{2}\frac{t^2}{n} + o(1/n))^n \\ &\to e^{-t^2/2} , \end{aligned}$$

where we have used the fact from calculus that $(1 + a/n + o(1/n))^n \to e^a$ for $n \to \infty$.

Math 464/564, O. Knill, **April 1997**

Some Inequalities

DEFINITION. A function $h : \mathbb{R} \to \mathbb{R}$ is called *convex*, if there exists for all $x_0 \in \mathbb{R}$ a linear map $l(x) = ax + b$ such that $l(x_0) = h(x_0)$ and for all $x \in \mathbb{R}$ the inequality $l(x) \leq h(x)$ holds.

DEFINITION. Let $\mathcal{L}^p$ be the set of random variables which satisfy $E[|X|^p] < \infty$ for $p \in [1, \infty)$ and $|X| \leq K$ almost everywhere for $p = \infty$. On the vector space $\mathcal{L}^p$ define $||X||_p = E[|X|^p]^{1/p}$ rsp. $||X||_\infty = \inf\{K \in \mathbb{R} \mid |X| \leq K \textit{ almost everywhere } \}$.

JENSEN INEQUALITY. Given $X \in \mathcal{L}^1$. For any convex function $h : \mathbb{R} \to \mathbb{R}$, one has

$$E[h(X)] \geq h(E[X]) .$$

PROOF. Let l be the linear map at $x_0 = E[X]$. By the linearity and monoticity of the expectation, we get

$$h(E[X]) = l(E[X]) = E[l(X)] \leq E[h(X)] .$$

HOELDER INEQUALITY. Given $p, q \in [1, \infty]$ with $p^{-1} + q^{-1} = 1$ and $X \in \mathcal{L}^p$ and $Y \in \mathcal{L}^q$. Then $XY \in \mathcal{L}^1$ and

$$||XY||_1 \leq ||X||_p ||Y||_q .$$

PROOF. We can restrict to $X, Y \geq 0$ and $||X||_p > 0$. Define the probabilty measure $Q = \frac{X^p P}{E[X^p]}$ and define $u = 1_{\{X>0\}} Y / X^{p-1}$. Jensen gives $Q(u)^q \leq Q(u^q)$ so that

$$E[|XY|] \leq ||X||_p ||1_{\{Z>0\}} Y||_q \leq ||X||_p ||Y||_q .$$

CAUCHY BUNYAKOWSKY-SCHWARZ INEQUALITY. $||XY||_1 \leq ||X||_2 ||Y||_2$.

MINKOWSKI INEQUALITY. Given $p \in [1, \infty]$ $X, Y \in \mathcal{L}^p$. Then $||X + Y||_p \leq ||X||_p + ||Y||_p$.

PROOF. We use Hölder's inequality to get

$$E[|X + Y|^p] = E[|X||X + Y|^{p-1}] + E[|Y||X + Y|^{p-1}] \leq ||X||_p C + ||Y||_p C ,$$

where $C = |||X + Y|^{p-1}||_q = E[|X + Y|^p]^{1/q}$ which leads to the claim.

PROOF. Given $p \leq q$. Define $h(x) = |x|^{q/p}$. Jensen's inequality gives $E[|X|^q] = E[h(|X|^p)] \leq h(E[|X|^p] = E[|X|^p]^{q/p}$. This implies that $||X||_q := E[|X|^q]^{1/q} \leq E[|X|^p]^{1/p} = ||X||_p$ for $p \leq q$. In other words, if $\mathcal{L}^p$ is the set of random variables with $E[|X|^p] < \infty$, then $\mathcal{L}^p \subset \mathcal{L}^q$ for $p \geq q$.

CHEBYCHEV-MARKOV INEQUALITY Let h be a monotone function on $\mathbb{R}$ with $h \geq 0$. For every $c > 0$, and $h(X) \in \mathcal{L}^1$ we have

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

PROOF. Integrate the inequality $h(c) 1_{X \geq c} \leq h(X)$ using the monotonicity and linearity of the expectation.

Math 464/564, O. Knill, **April 1997**

The central limit theorem

DEFINITIONS. A sequence of random variables X_n **converges in distribution** to X if $P[X_n \leq t] \to P[X \leq t]$ for all $t \in \mathbb{R}$. Denote by $Y^* = (Y - E[Y])/\sigma(Y)$ the **normalized random variable** to Y.

THE CENTRAL LIMIT THEOREM FOR INDEPENDENT RANDOM VARIABLES. Assume X_n are independent random variables satisfying $\sup_n E[|X_n|^3] < \infty, \liminf_{n\to\infty} \frac{1}{n} \sum_{i=1}^n Var[X_i] > 0$. Then S_n^* converges in distribution to a standard normal distributed random variable. The same conclusion holds if X_n are IID with nonzero finite variance.

PROOF. Define for fixed $n \geq 1$ the random variables

$$Y_i = (X_i - E[X_i])/\sigma(S_n),\ 1 \leq i \leq n$$

so that $S_n^* = \sum_{i=1}^n Y_i$. Define $N(0, \sigma^2 = Var[Y_i])$-distributed random variables $\tilde{Y}_i$ having the property that $\{Y_1, \dots, Y_n, \tilde{Y}_1, \dots \tilde{Y}_n\}$ are independent. The distribution of $\tilde{S}_n = \sum_{i=1}^n \tilde{Y}_i$ is just the normal distribution $N(0,1)$. The convergence in distribution is equivalent to $E[f(S_n^*)] - E[f(\tilde{S}_n)] \to 0$ for any smooth function f which is zero outside some interval. Define

$$Z_k = \tilde{Y}_1 + \dots \tilde{Y}_{k-1} + Y_{k+1} + \dots + Y_n \ .$$

Note that $Z_1 + Y_1 = S_n^*$ and $Z_n + \tilde{Y}_n = \tilde{S}_n$. Using a telescopic sum and Taylor's theorem, we write

$$\begin{aligned} f(S_n^*) - f(\tilde{S}_n) &= \sum_{k=1}^n [f(Z_k + Y_k) - f(Z_k + \tilde{Y}_k)] \\ &= \sum_{k=1}^n [f'(Z_k)(Y_k - \tilde{Y}_k) + \frac{1}{2} f''(Z_k)(Y_k^2 - \tilde{Y}_k^2) \\ &\quad + R(Z_k, Y_k) + R(Z_k, \tilde{Y}_k)] \end{aligned}$$

with a Taylor rest term $R(Z,Y)$ depending on f. We get therefore

$$|E[f(S_n^*)] - E[f(\tilde{S}_n)]| \leq \sum_{k=1}^n E[|R(Z_k, Y_k)|] + E[|R(Z_k, \tilde{Y}_k)|] \ . \tag{1}$$

Since $\tilde{Y}_k$ are $N(0, \sigma^2)$-distributed we get using the computation of the moments of normal random variables and Jensen's inequality

$$E[|\tilde{Y}_k|^3] = \sqrt{\frac{8}{\pi}} \sigma^3 = \sqrt{\frac{8}{\pi}} E[|Y_k|^2]^{3/2} \leq \sqrt{\frac{8}{\pi}} E[|Y_k|^3] \ .$$

The Taylor rest can be estimated as $|R(Z_k, Y_k)| \leq const \cdot |Y_k|^3$ so that

$$\begin{aligned} \sum_{k=1}^n E[|R(Z_k, Y_k)|] + E[|R(Z_k, \tilde{Y}_k)|] &\leq const \cdot \sum_{k=1}^n E[|Y_k|^3] \\ &\leq const \cdot n \cdot \sup_i E[|X_i|^3]/Var[S_n]^{3/2} \\ &= const \cdot \frac{\sup_i E||X_i|]^3}{(Var[S_n]/n)^{3/2}} \cdot \frac{1}{\sqrt{n}} \to 0 \ . \end{aligned}$$

If we know only the finite nonzero variance of X_n but instead that the X_n are IID, we estimate the Taylor rest as $|R(z,y)| \leq \delta(y) \cdot y^2$, where $\delta(y) \to 0$ for $|y| \to 0$. The IID property and the dominated convergence gives

$$\sum_{k=1}^n E[|R(Z_k, Y_k)|] + E[|R(Z_k, \tilde{Y}_k)|] \leq \sum_{k=1}^n E[\delta(Y_k) Y_k^2] + E[\delta(\tilde{Y}_k) \tilde{Y}_k^2]$$

Math 464/564, O. Knill, **April 1997**

Multidimensional distributions (Summary II)

The distribution of a sum

PROPOSITION. If X and Y are independent and absolutely continuous, then $f_{X+Y}(z) = f_X \star f_Y(y) := \int_{-\infty}^{\infty} f_X(x) f_Y(y-x)\, dx$.

EXAMPLES.
1) The sum of independent $\Gamma(\alpha_1, \lambda)$ and $\Gamma(\alpha_2, \lambda)$ distributed random variables is $\Gamma(\alpha_1+\alpha_2, \lambda)$ distributed.
2) If X and Y are independent normal $N(m_1, \sigma_1^2)$ and $N(m_2, \sigma_2^2)$ distributed random variables, then $X+Y$ is $N(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$ distributed.

Transformations of densities

The usual transformation rule for volume integrals gives:

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

$$g(y) = f(\psi(y)) |Det(D\psi(y))| \ ,$$

where $[D\psi(y)]_{ij} = \frac{\partial}{\partial x_j} \psi_i(y)$ is the Jacobian matrix.

EXAMPLES.
1) The formula generalizes the formula for $n = 1$, where $D\psi(y) = \psi'(y)$.
2) Given a random vector (X, Y) with joint density f. Problem: 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 $Det(D\psi) = Det(D\phi^{-1}) = 1/Det(D\phi)$. Therefore

$$f_{(U,V)}(u, v) = \frac{1}{2} f_{(X,Y)}(\frac{u+v}{2}, \frac{u-v}{2}) \ .$$

Conditional densities

DEFINITION. Given an absolutely continuous random vector (X, Y), The **conditional probability density function** for X given Y is defined as

$$f_{X|Y}(x|y) = \frac{f_{(X,Y)}(x, y)}{f_Y(y)} \ .$$

It is a density function of a random variable.

EXAMPLE. If X and Y are independent, then $f_{X|Y}(x|y) = f_X(x)$ because then $f_{(X,Y)}(x, y) = f_X(x) f_Y(y)$.

CONTINUOUS BAYES RULE.

$$f_{X|Y}(x|y) = \frac{f_X(x) f_{Y|X}(y|x)}{\int_{-\infty}^{\infty} f_X(x) f_{Y|X}(y|x)\, dx} \ .$$

Proof. Plug in $f(x, y) = f_X(x) f_{Y|X}(y|x)$ into $f_Y(x) = \int_{-\infty}^{\infty} f(x, y)\, dx$ and put this into the formula $f_X(x) f_{Y|X}(y|x) = f_{X|Y}(x|y) f_Y(x)$.

ALMOST EVERYWHERE $\Rightarrow$ IN PROBABILITY:

Since

$$\{X_n \to X\} = \bigcap_k \bigcup_m \bigcap_{n \geq m} \{|X_n - X| \leq 1/k\} ,$$

"almost everywhere convergence" is equivalent to

$$1 = P[\bigcup_m \bigcap_{n \geq m} \{|X_n - X| \leq 1/k\}] = \lim_{n \to \infty} P[\bigcap_{n \geq m} \{|X_n - X| \leq 1/k\}]$$

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 $\epsilon > 0$.

$\mathcal{L}^1$ CONVERGENCE $\Rightarrow$ IN PROBABILITY:

With Chebychev-Markov inequality for $g(\epsilon) = \epsilon^p$, one obtains

$$P[|X_n - X| \geq \epsilon] \leq E[|X_n - X|^p]/\epsilon^p .$$

COMPLETE $\Rightarrow$ ALMOST EVERYWHERE:

The Borel-Cantelli lemma (see seperate page) implies that for all $\epsilon > 0$

$$P[|X_n - X| \geq \epsilon, \ infinitely\ often] = 0 .$$

We get so for $\epsilon_n \to 0$

$$P[\bigcup_n |X_n - X| \geq \epsilon_k, \ infinitely\ often] \leq \sum_n P[|X_n - X| \geq \epsilon_k, \ infinitely\ often] = 0$$

from which we obtain $P[X_n \to X] = 1$.

IN PROBABILITY $\Rightarrow$ IN DISTRIBUTION:

Let t be a continuity point of F_X. For $\epsilon > 0$ and $n \in \mathbb{N}$, one has

$$\begin{array}{rcl} F_{X_n}(t) & = & P[X_n \leq t] \\ & = & P[X_n \leq t, |X_n - X| \leq \epsilon] + P[X_n \leq t, |X_n - X| > \epsilon] \\ & \leq & P[X \leq t + \epsilon] + P[|X_n - X| > \epsilon] \\ & = & F_X(t + \epsilon) + P[X_n - X] > \epsilon] \end{array}$$

which implies $\limsup_{n \to \infty} F_{X_n}(t) \leq F_X(t + \epsilon)$ and using the continuity of F at t also $\limsup_{n \to \infty} F_{X_n}(t) \leq F_X(t)$. Similarly, one gets

$$\begin{array}{rcl} F_X(t - \epsilon) & = & P[X \leq t - \epsilon] \\ & = & P[X \leq t - \epsilon, |X_n - X| \leq \epsilon] + P[X \leq t - \epsilon, |X_n - X| > \epsilon] \\ & \leq & P[X_n \leq t] + P[|X_n - X| > \epsilon] \\ & = & F_{X_n}(t) + P[|X_n - X| > \epsilon] . \end{array}$$

For $n \to \infty$, one gets $F_X(t - \epsilon) \leq \liminf_{n \to \infty} F_{X_n}(t)$. Using the continuity of F at t one has $F_X(t) \leq \liminf_{n \to \infty} F_{X_n}(t)$.

Math 464/564, O. Knill, **April 1997**

Convergence of random variables

There are different definitions of convergence for random variables. Examples: The strong law of large numbers assures **almost everywhere convergence** of S_n/n to $E[X]$. The weak law of large numbers assures **convergence in probability** of S_n/n to $E[X]$. The central limit theorem states the **convergence in distribution** of S_n^* to a standard normal distributed random variable.

DEFINITIONS: (only convergence in probability and in distribution occur in the book).

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

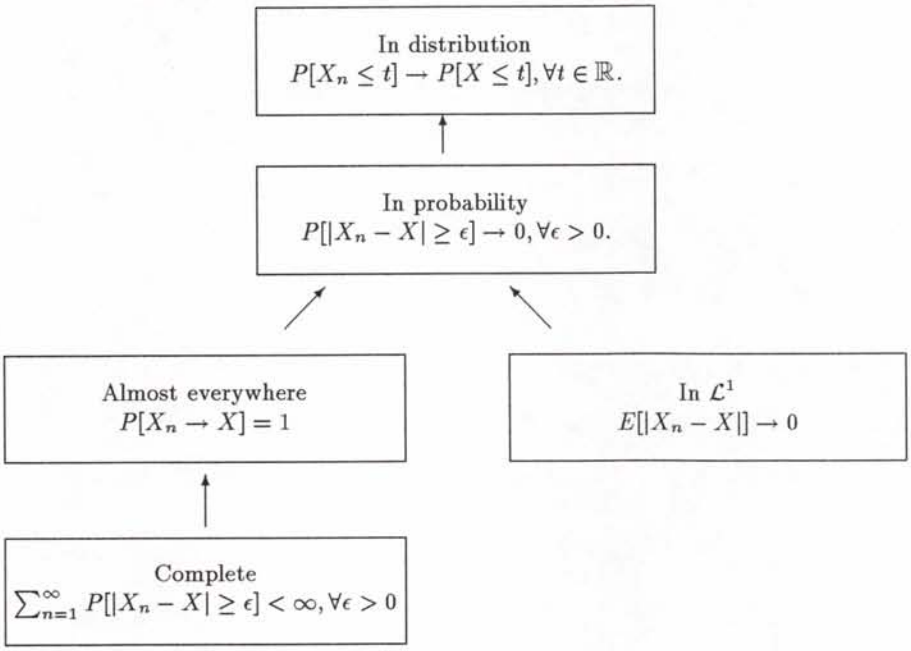
A sequence of random variables X_n converges **in probability** to a random variable X, if $P[|X_n - X| \geq \epsilon] \to 0$ for all $\epsilon > 0$.

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

A sequence of random variables X_n **converges in** $\mathcal{L}^1$ to a random variable X, if $E[|X_n - X|] \to 0$.

A sequence of random variables X_n converges **completely** if $\sum_n P[|X_n - X| \geq \epsilon] < \infty$ for all $\epsilon > 0$.

RELATIONS BETWEEN CONVERGENCE: (an arrow stands for "implies")


In distribution
$P[X_n \leq t] \to P[X \leq t], \forall t \in \mathbb{R}$.
In probability
$P[|X_n - X| \geq \epsilon] \to 0, \forall \epsilon > 0$.
Almost everywhere
$P[X_n \to X] = 1$
In $\mathcal{L}^1$
$E[|X_n - X|] \to 0$
Complete
$\sum_{n=1}^{\infty} P[|X_n - X| \geq \epsilon] < \infty, \forall \epsilon > 0$

The convergence in distribution is the only type of convergence which does not assume that X_n, X are defined on the same probability space. The $\mathcal{L}^1$ convergence (and more generally $\mathcal{L}^p$ convergence) is used especially in more advanced topics of probability theory like stochastic processes or martingale theory. If the random variables are bounded $X_n \leq M$, the $\mathcal{L}^1$ convergence is equivalent to convergence in probability. The complete convergence can be useful to prove almost everywhere convergence. For example, the Chebychev-Markov inequality with the function $g(\epsilon) = \epsilon^4$ allows to prove that for a sequence of IID random variables with finite forth moment, S_n/n converges completely to $E[X]$. This implies then the strong law of large numbers.

Math 464/564, O. Knill, **April 1997**

Recurrence of random walks

QUESTION. A walker starts at the origin 0 of the d dimensional lattice $\mathbb{Z}^d$ and makes each second a random step into one of the $2d$ directions. Does the walker return back to the origin again and again?

THEOREM. The walker returns to the origin infinitely often with probability one in one or two dimensions. In dimensions $d \geq 3$, the walker returns only finitely many times to zero and escapes to infinity with probability one.

SETUP. Consider IID random vectors X_n taking values in $\{e \in \mathbb{Z}^d \mid |e| = \sum_{i=1}^d |e_i| = 1\}$ with uniform distribution $P[X_n = e] = (2d)^{-1}$. The random vector $S_n = \sum_{i=1}^n X_i$ with $S_0 = 0$ is the position of the walker at time n. The random variable $Y_n = 1_{A_n}$ with $A_n = \{X_n = 0\}$ tells, whether the walker is at the origin at time n or not. The sum $B_n = \sum_{i=0}^n Y_n$ counts the number of visits of the walker up to time n and $B = \sum_{i=0}^\infty Y_n$ is the random variable which gives the total number of visits during the evolution. The expectation $E[B] = \sum_{n=0}^\infty P[S_n = 0]$ is the expected number of returns to the origin.

LEMMA OF POLYA. $E[B] = \infty$ for $d = 1, 2$ and $E[B] < \infty$ for $d > 2$.

PROOF.

(i) For $n \in \mathbb{N}$, we know that $P[S_n = k]$ for $k \in \mathbb{Z}^d$. vanishes for large $|k|$ because the walker can only reach a finite region in finite time. The characteristic function $\phi_{S_n} = \sum_k P[S_n = k]e^{ixk}$ is therefore a smooth function on the torus $\mathbb{T}^d := \mathbb{R}^d/(2\pi\mathbb{Z}^d)$ (each coordinate of $\phi_{S_n}(x)$ is a 2π-periodic function).

(ii) $\phi_{X_k}(x) = \frac{1}{2d}\sum_{|j|=1} e^{ix_j} = \frac{1}{d}\sum_{i=1}^d \cos(x_i)$. Because the S_n is a sum of n IID random variables we get $\phi_{S_n} = \phi_{X_k}^n = \frac{1}{d^n}(\sum_{i=1}^d \cos(x_i))^n$.

(iii) By the Fourier inversion formula, we have $P[S_n = 0] = (2\pi)^{-d}\int_0^{2\pi} \dots \int_0^{2\pi} \phi_{S_n}(x)\, dx_1 \dots dx_d$. Therefore

$$E[B] = \int_{\mathbb{T}^d} \sum_n \phi_X^n(x)\, dx = \int_{\mathbb{T}^d} \frac{1}{1 - \phi_X(x)}\, dx \ .$$

With a Taylor expansion $\phi_{X_k}(x) = \frac{1}{d}\sum_{i=1}^d \cos(x_i) = 1 - \sum_j x_j^2/(4d) + \dots$ we can estimate $\frac{1}{8d}|x|^2 \leq 1 - \phi_X(x) \leq \frac{1}{2d}|x|^2$. The lemma follows because $\int_{\{|x|<\epsilon\}} \frac{1}{|x|^2}\, dx$ is finite if and only if $d \geq 3$. (For the later fact see the Homework).

PROOF OF THE THEOREM.

(i) $d > 2$: define $A_n = \{S_n = 0\}$. The event $A_\infty = \limsup_n A_n$ is the subset of Ω, for which the walker returns to 0 infinitely many times. Since by the lemma $E[B] = \sum_{n=0}^\infty P[A_n] < \infty$, the Borel-Cantelli lemma gives $P[A_\infty] = 0$ for $d > 2$. The walker returns therefore back to 0 only finitely many times and in the same way it visits each lattice point only finitely many times. The walker eventually leaves every bounded set and escapes to infinity.

(ii) $d = 1$ or $d = 2$: if $p = P[\bigcup_n A_n]$ is the probability that the walker returns to 0 at least once, then p^{m-1} is the probability that there are at least m visits of the origin and $p^{m-1} - p^m = p^{m-1}(1-p)$ is the probability that there are exactly m visits. We can write

$$E[B] = \sum_{m \geq 1} m p^{m-1}(1-p) = 1/(1-p) \ .$$

Since by Polya's lemma $E[B] = \infty$ we see that $p = 1$.

Math 464/564, O. Knill, **April 1997**

1D walk (1): gamblers ruin, beating the system, Wald identities

DEFINITION: Consider IID $\{-1,1\}$-valued random variables X_k. The random variables $S_n = \sum_{k=1}^n X_k(\omega)$ is the location of the random walker 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.

SIMPLE PROPERTIES: $E[X_k] = 0$ and so $E[S_k] = 0$. If $n+x$ even, $P[X_n = x] = 2^{-n} \left(\begin{array}{c} n \\ \frac{n+x}{2} \end{array} \right)$, if $n+x$ odd then $P[X_n = x] = 0$.

DEFINITION: A **gambling system** or betting strategy to the random walk X_k is sequence of random variables V_k such that every event $\{V_n = c\}$ depends only on the values of X_j for $j < k$ (one says V is a previsible process with respect to the "martingale" S_n). Define

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

the **winning** with this system. Remark. S_n^V is also written as $V \cdot S$ or $\int V dS$. You can also use the fancy name like "martingale transform" or (totally impressive) "discrete stochastic integral".

YOU CAN'T BEAT THE SYSTEM.
If $E[X_k] = 0$, then $E[S_N^V] = 0$. PROOF. Because V_k is independent of X_k, we have $E[V_k X_k] = E[V_k X_k] = E[V_k] E[X_k]$.

DEFINITION. A **stopping time** is a random variable T taking values in $\overline{\mathbb{N}} := \{0, 1, \dots, \infty\}$ such that the event $\{T \leq k\}$ can be defined using the random variables $X_1, \dots, X_k$. Examples are

$$T_a = \min\{n \in \mathbb{N} \setminus \{0\} \mid S_n = a\} .$$

or

$$T_{a,b} = \min\{n \in \mathbb{N} \,| S_n = b \; or \; S_n = -a \; \} .$$

Gambling interpretation: T_a is the time until a player gets broke or the time, when the player decides to stop the game. Whether or not to stop immediatly after the n'th game depends only on the history up to time n. For example, $T_{a,b}$ is the time until one of two competing players gets broke. $P[X_T = -a]$ is the probability that the first player with capital a gets broke, $P[X_T = b]$ is the probability that the second player gets broke.

The proofs of the following theorems are in the book.

FIRST WALD'S IDENTITY. If T be a stopping time, then

$$E[S_T] = E[T] \cdot E[X] .$$

Especially, for a symmetric random walk, where $E[X] = 0$, one has $E[S_T] = 0$.

SECOND WALD IDENTITY. If T is a stopping time, then

$$Var[S_T] = E[T] \cdot Var[X] .$$

APPLICATION OF FIRST WALD IDENTITY: FORMULA FOR THE RUIN PROBABILITY.

$$P[X_{T_{a,b}} = -a] = 1 - P[X_{T_{a,b}} = b] = b/(a+b) .$$

PROOF. $0 = E[S_T] = -aP[X_T = -a] + bP[X_T = b] = -aP[X_T = -a] + b(1 - P[X_T = -a])$.

Math 464/564, O. Knill, **April 1997**

The Borel-Cantelli lemma

The Borel-Cantelli lemma appeared in two places: in a proof of an advanced central limit theorem and for establishing the recurrence or transience of the random walk.

DEFINITION: Given a sequence of events A_n in a probability space $(\Omega, \mathcal{A}, P)$. Define $A_\infty := \limsup_{n\to\infty} A_n := \bigcap_{m=1}^\infty \bigcup_{n\geq m} A_n$. We have $A_\infty = \{\omega \mid \omega \textit{ is in infinitely many } A_i\}$.
INTERPRETATION. The set A_∞ is is the event that infinitely many of the events A_n happen.

BOREL CANTELLI LEMMA:

(1) If $\sum_n P[A_n] < \infty$, then $P[A_\infty] = 0$.

(2) If A_n are independent and $\sum_n P[A_n] = \infty$, then $P[A_\infty] = 1$.

PROOF.

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

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

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

(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

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

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

MONKEY TYPING SHAKESPEARE:
Sit a monkey at a typewriter. The probability that he will write Shakespeares collected work is one. Actually, according to the Borel-Cantelli lemma, he will write it infinitely often. The following Mathematica procedures allow to simulate the typing of the Monkey typing Shakespeare: We let him type 700 letters on the computer and save it in a file Sheakspeare. Enjoy reading.

```
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}]

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

Math 464/564, O. Knill, **April 1997**

1D walk (2), Ballot and Arc Sin theorem.

REFLECTION PRINCIPLE. For $a, b \in \mathbb{N} \setminus \{0\}$, one has

$$P[a + S_n = b \mid T_{-a} \leq n] = P[S_n = a + b] \, .$$

PROOF. The number of paths from a to b passing zero is equal to the number of paths from $-a$ to b which is the number of paths from zero to $a + b$.

DISTRIBUTION OF THE RUIN TIME. $P[T_{-a} \leq n] = P[S_n \leq -a] + P[S_n > a]$.
$P[T_{-a} = n] = \frac{a}{n} P[S_n = a]$.

PROOF. a) Use the reflection principle

$$\begin{aligned} P[T_{-a} \leq n] &= \sum_{b \in \mathbb{Z}} P[T_{-a} \leq n | a + S_n = b] \\ &= \sum_{b \leq 0} P[a + S_n = b] + \sum_{b > 0} P[S_n = a + b] \\ &= P[S_n \leq -a] + P[S_n > a] \end{aligned}$$

b) From $P[S_n = a] = \left(\begin{array}{c} n \\ \frac{a+n}{2} \end{array} \right)$ we get

$$\frac{a}{n} P[S_n = a] = \frac{1}{2} (P[S_{n-1} = a - 1] - P[S_{n-1} = a + 1]) \, .$$

$$\begin{aligned} P[S_n > a] - P[S_{n-1} > a] &= P[S_n > a \mid S_{n-1} \leq a] + P[S_n > a \mid S_{n-1} > a] - P[S_{n-1} > a] \\ &= \frac{1}{2} (P[S_{n-1} = a] - P[S_{n-1} = a + 1]) \end{aligned}$$

and analoguously

$$P[S_n \leq -a] - P[S_{n-1} \leq -a] = \frac{1}{2} (P[S_{n-1} = a - 1] - P[S_{n-1} = a]) \, .$$

Therefore, using a)

$$\begin{aligned} P[T_{-a} = n] &= P[T_{-a} \leq n] - P[T_{-a} \leq n - 1] \\ &= P[S_n \leq -a] - P[S_{n-1} \leq -a] + P[S_n > a] - P[S_{n-1} > a] \\ &= \frac{1}{2} (P[S_{n-1} = a] - P[S_{n-1} = a + 1]) + \frac{1}{2} (P[S_{n-1} = a - 1] - P[S_{n-1} = a]) \\ &= \frac{1}{2} (P[S_{n-1} = a - 1] - P[S_{n-1} = a + 1]) = \frac{a}{n} P[S_n = a] \end{aligned}$$

BALLOT THEOREM. $P[S_n = a \mid S_1 > 0, \dots, S_{n-1} > 0] = \frac{a}{n} \cdot P[S_n = a]$.
PROOF. Reverse time: the number of paths from 0 to a of length n which do no more hit 0 is the number of paths of length n which start in a and for which $T_{-a} = n$.

DISTRIBUTION OF FIRST RETURN TIME.

$$P[T_0 > 2n] = P[S_{2n} = 0] \, .$$

PROOF.

$$\begin{aligned} P[T_0 > 2n] &= \frac{1}{2}P[T_{-1} > 2n-1] + \frac{1}{2}P[T_1 > 2n-1] \\ &= P[T_{-1} > 2n-1] \qquad\qquad (\textit{ by symmetry}) \\ &= P[S_{2n-1} > -1 \textit{ and } S_{2n-1} \leq 1] \\ &= P[S_{2n-1} \in \{0,1\}] \\ &= P[S_{2n-1} = 1] = P[S_{2n} = 0] \end{aligned}$$

REMARK. We see that $\lim_{n\to\infty} P[T_0 > 2n] = 0$ which restates that the random walk is recurrent. However, the expected return time is very long:

$$E[T_0] = \sum_{n=0}^{\infty} nP[T_0 = n] = \sum_{n=0}^{\infty} P[T_0 > n] = \sum_{n=0}^{\infty} P[S_n = 0] = \infty$$

since by Stirling's formula

$$P[S_{2n} = 0] \sim (\pi n)^{-1/2} .$$

DEFINITION. The stopping time

$$L = \max\{0 \leq n \leq 2N \mid S_n = 0\}$$

is the **last visit of the random walk in 0 before time** $2N$. Interpretation. If the random walk describes a game between two players, who play over a time $2N$, then L is the time when one of the two players does no more give up his or her leadership.

ARC-SIN LAW. a) L has the discrete arc-sin distribution:

$$P[L = 2n] = 2^{-2N} \binom{2n}{n} \binom{2N-2n}{N-n}$$

b) For $N \to \infty$, we have

$$P[\frac{L}{2N} \leq z] \to \frac{2}{\pi} \arcsin(\sqrt{z}) .$$

PROOF.

$$P[L = 2n] = P[S_{2n} = 0] \cdot P[T_0 > 2N - 2n] = P[S_{2n} = 0] \cdot P[S_{2N-2n} = 0]$$

which gives a).
Stirling gives $P[S_{2k} = 0] \sim \frac{1}{\sqrt{\pi k}}$ so that

$$P[L = 2k] = \frac{1}{\pi} \frac{1}{\sqrt{k(N-k)}} = \frac{1}{N} f(k/N)$$

with

$$f(x) = \frac{1}{\pi\sqrt{x(1-x)}} .$$

It follows that

$$P[L/2N \leq z] \to \int_0^z f(x)\, dx = \frac{2}{\pi} \arcsin(\sqrt{z}) .$$

INTERPRETATION. From the shape of the arc-sin distribution, One has to expect that the winner takes the final leading position either very early or very late.

Math 464/564, O. Knill, **May 1997**

Definition and existence of Poisson processes

DEFINITION. Let S be a region of $\mathbb{R}^n$, $\mathcal{B}$ the set of measurable sets and μ a measure (called **mean measure**) on S. A **Poisson process on** S (S, μ, N) is a family of random variables N_B, $B \in \mathcal{B}$ such that
(i) N_B is Poisson distributed with parameter $\mu(B)$.
(ii) If $B_j, j = 1, \dots k$ are disjoint, then N_{B_j} are independent.

INTERPRETATION. The random variables N_B can be defined over a probability space Ω where each point ω is a random collection of points in S. $N_B(\omega)$ is the number of points of ω which ly in B.
EXAMPLES. a) If B is a sky region in sky and N_B is the number of stars in that region. b) If B is a region in the Arizona desert and N_B is the number of Saguari in that region. c) If B is a time interval then N_B, the number of cars passing a certain spot on the $I10$ is a Poisson process.

SUPERPOSITION LEMMA. If $(S, \mu^{(i)}, N^{(i)})$ are independent Poisson processes on S, then $(S, \sum_i \mu^{(i)}, \sum_i N^{(i)})$ is a Poisson process on S.
PROOF. We check the conditions.
(i) The sum of independent Poisson distributed random variables is Poisson distributed.
(ii) If $B_k, k = 1, \dots, n$ are disjoint, then $\{N_{B_k}^{(i)}\}_{i \in \mathbb{N}, k=1,\dots n}$ are all independent and so are therefore the random variables $N_{B_k} = \sum_{i=1}^\infty N_{B_k}^{(i)}, k = 1, \dots n$.

DEFINITION. A measure μ is called **nonatomic** if $\mu(\{x\}) = 0$ for all $x \in S$. A measure μ is called **σ-finite** if $\mu = \sum_n \mu^{(n)}$ with $\mu^{(n)}(S) < \infty$.

EXISTENCE THEOREM. If μ is a nonatomic σ-finite measure on S, then there is a Poisson process (S, μ, N) with mean measure μ.

PROOF. Write $\mu = \sum_n \mu^{(n)}$, where $\mu^{(n)}$ are nonatomic measures with $0 < \mu^{(n)}(S) < \infty$.
Take IID N_n, $n = 1, 2, \dots$ which are Poisson-$(\mu^{(n)}(S))$ distributed. Independent from them take IID $\mathbb{R}^d$-valued random vectors X_{nr}, $r = 1, 2, \dots$ which have joint distribution of a vector $X(x) = x$ on the probability space $(S, \mathcal{B}, \mu^{(n)}(\cdot)/\mu^{(n)}(S))$. All these independent random variables N_n, X_{nr} are defined on some probability space Δ with elements δ. Define $N_B^{(n)}(\delta)$ as the number of points in the finite set $\{X_{n1}(\delta), \dots, X_{nN(\delta)}(\delta)\} \cap B$. Claim: $(S, \mu^{(n)}, N^{(n)})$ is a Poisson process with mean measure $\mu^{(n)}$.
(i) $N_B^{(n)}$ is $\mu^{(n)}(B) := \mu^{(n)}(B)$-Poisson distributed because $P[N_B^{(n)} = k] = P[N_n = k]$.
(ii) Let $B_k, k = 1, \dots, m$ be disjoint. For $l = l_0 + \dots + l_m$ and $B_0 := S \setminus \bigcup_{k=1}^m B_k$, one has $P[N_{B_1}^{(n)} = l_1, \dots, N_{B_m}^{(n)} = l_m \mid N_n = l = l_0 + \sum_{j=1}^m l_j] = \frac{l!}{l_0! \cdots l_m!} \prod_{j=1}^m (\mu^{(n)}(B_j)/\mu^{(n)}(S))^{l_j}$. Therefore

$$\begin{aligned} P[N_{B_1}^{(n)} = l_1, \dots, N_{B_m}^{(n)} = l_m] &= \sum_{l=\sum_j l_j}^{\infty} \frac{e^{-\mu^{(n)}(S)} \mu^{(n)}(S)^l}{l!} \frac{l!}{l_0! \cdots l_m!} \prod_{j=1}^m (\mu^{(n)}(B_j)/\mu^{(n)}(S))^{l_j} \\ &= [\sum_{l_0=0}^{\infty} \frac{e^{-\mu^{(n)}(B_0)} \mu^{(n)}(B_0)^{l_0}}{l_0!}] \prod_{j=1}^m \frac{e^{-\mu^{(n)}(B_j)} \mu^{(n)}(B_j)^{l_j}}{l_j!} \\ &= \prod_{j=1}^m \frac{e^{-\mu^{(n)}(B_j)} \mu^{(n)}(B_j)^{l_j}}{l_j!} = \prod_{j=1}^m P[N_{B_j}^{(n)} = l_j] \,. \end{aligned}$$

The superposition lemma shows now that $(S, \mu = \sum_i \mu^{(i)}, N = \sum_i N^{(i)})$ is a Poisson process with mean measure μ.

Math 464/564, O. Knill, **May 1997**

Summary and further directions

SUMMARY. Probability theory is the study of random variables (or random vectors) on a probability space. A distribution of a variable is determined in the discrete case by $P[X = x]$ and in the continuous case by the (joint) density $f_X(x)$. The distribution is determined by the characteristic function $\Phi_X(x) = E[\exp(it \cdot X)]$. It determines properties like the (vector) mean $E[X]$ the covariance matrix $Cov[X_i, X_j]$ etc. For independent X, Y, one has $f_{X+Y}(z) = f_X \star f_Y)(z)$ and $\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t)$.
Calculations for discrete distributions need some combinatorial tools, for continuous distributions, we rely on analytic (calculus) tools. Combinatorics (e.g. the Monty-Hall problem) or problems using continuous distributions (e.g. the Bertrand paradox) can be puzzling without a solid foundation of probability theory.
A bunch of distributions with their properties, characteristic functions (make a comprehensive list!) help to model and solve practical problems.
Some important theorems in probability theory are the theorem of Caratheodory, the central limit theorem, the (weak) law of large numbers. The Chebychev-Markov inequalities as well as other inequalities (Jensen, Hölder, Cauchy-Schwartz) are useful for theory or practical estimate.
We have seen that probabilistic tools can give easy access to analytic results like the theorem of Weierstrass or the Stirling formula.
Probability theory hits at several places the foundations of mathematics. Examples are the foundation of integration (the Lebesgue integral) or the impossibility to define a probability space on the interval $[0, 1]$ using all subsets as σ-algebra (Banach-Tarsky paradox).
Studiing the sum of independent random variables (like the random walk) lead to interesting questions and beautyful results (like e.g. limit theorems, Polya's theorem) and applications (e.g. ruin problems, Wald identities). Beside the random walk, other processes like the Poisson process or the Brownian motion (not treated here) are important.

FURTHER DIRECTIONS. A continuation of the theory involves the study of stochastic processes. The later is a set of random variables which are labeled by time. For continuous time, the setup needs some care. An important study is the theory of Brownian motion or more general martingales. With such processes, one can define stochastic integrals and therefore stochastic PdE's. This is useful in mathematical physics and some problems in quantum mechanics can be solved quite nicely using path integrals. Stochastic differential equations appear also in Economics (e.g. the Black-Scholl option pricing formula).
Some directions in mathematical physics like statistical physics, quantum mechanics or transport theory involve and motivate research in probability theory. Examples are percolation problems, the statistical theory of lattice gases, gauge theory in particle physics, the theory of random Schrödinger operators, the statistical mechanics of hyperbolic dynamical systems or problems in nonequilibrium statistical mechanics and fluid dynamics. A branch of dynamical system theory called ergodic theory can be viewed as an extension of probability theory: there, a measure preserving map T on the probability space and a random variable X define a sequence of random variables $\omega \mapsto X_n(\omega) = X(T^n\omega)$ which are in general not independent.
An other direction important in applications is statistics which has a relation to probability theory similar than numerical mathematics to analysis. The problem is to model data by distributions and get information about the reliablity of these models. This is in applications mostly used in descriptive statistics, where one wants to describe, illustrate and interpret data obtained for example in a laboratory, determine correlations between different quantities, determine the parameters of the distributions, the confidence intervals, the error probability etc.

FOR THE FINAL: we went through all relevant sections all except 6.5, 6.6 in the book. The comprehensive final will be on topics, which we have treated in class and can include topics about distributed material (not the proofs of theorems like Caratheodory or the advanced central limit theorem or Polya's theorem). A good preparation includes a personal summary of the material, especially to make a reliable list of all distributions with properties and a list of definitions and results as well as an organization of the notes, a review of the homework and the midterms. As in the midterms, you can use all your notes but no book.

Math 464/564, O. Knill, **May 1997**

Definition and existence of Brownian motion

Brownian motion is an important object in mathematics. Not part of the current course we give here as a preview the definition and existence of Brownian motion.

DEFINITION. A collection of random vectors X_t, $t \in T \subset \mathbb{R}$ is called a **stochastic process.** If T is a discrete set then it is a discrete time stochastic process. An example is the random walk, where $\mathbb{T} = \mathbb{N}$. If T is an interval, then it is a continuous time stochastic process. The process is called **continuous** if $(t, \omega) \mapsto X_t(\omega)$ is continuous for almost all ω.

RECALL. A $\mathbb{R}^n$-valued random variable X is called **Gaussian** (= normal) if it has the characteristic function $E[e^{it \cdot X}] = e^{-(t,Vt)/2+imt}$, where V is the covariance matrix and $m = E[X]$ is the mean vector. A **Gaussian processs** is a stochastic process such that $(X_{t_0}, X_{t_1}, \dots X_{t_n})$ is a Gaussian random vector for any $t_0 \leq t_1 < \dots < t_n$. It is called centered, if $m = 0$. The importance of the Gaussian distribution is that two Gaussian random vectors X, Y are independent if and only if they are uncorrelated.

DEFINITION. A continuous Gaussian process X_t with values in $\mathbb{R}^d$ having the mean vector $m_t = E[X_t]$ and the covariance matrix $V(s,t) = Cov[X_s, X_t] = E[(X_s - m_s)(X_t - m_t)^*]$ is called **Brownian motion** if for any $0 \leq t_0 < t_1 < \dots < t_n$, the random variables $X_{t_0}, X_{t_{i+1}} - X_{t_i}$ are independent and the covariance matrix V satisfies $V(s,t) = V(r,r)$, where $r = \min(s,t)$ and $s \mapsto V(s,s)$ is increasing. It is called the **standard Brownian motion** if $m_t = 0$ for all t and $V(s,t) = \min\{s,t\}$.

LEMMA. A Gaussian process with covariance $V(s,t) = V(r,r)$ with $r = \min(s,t)$ is a Brownian motion.
PROOF. For independence it is enough to check that $Cov[X_{t_0}, X_{t_{j+1}} - X_{t_j}] = 0$, $Cov[X_{t_{i+1}} - X_{t_i}, X_{t_{j+1}} - X_{t_j}] = 0$ which follows from $Cov[X_{t_0}, X_{t_{j+1}} - X_{t_j}] = V(t_0, t_{j+1}) - V(t_0, t_j) = V(t_0, t_0) - V(t_0, t_0) = 0$ and $Cov[X_{t_{i+1}} - X_{t_i}, X_{t_{j+1}} - X_{t_j}] = V(t_{i+1}, t_{j+1}) - V(t_{i+1}, t_j) - V(t_i, t_{j+1}) + V(t_i, t_j) = V(t_{i+1}, t_{i+1}) - V(t_{i+1}, t_{i+1}) - V(t_i, t_i) + V(t_i, t_i) = 0$.

THEOREM. Brownian motion exists!

SOME HISTORY. Water under a microscope contains little things moving around. One first thought that these particles were alive. **Brown** was studying the fertilization process in a species of flowers. Looking at the pollen in water through a microscope, he observed small particles in "rapid oscillatory motion". Brown's explanation to this was that matter is composed of small particles, which he calls active molecules, which exhibit a rapid, irregular motion having its origin in the particles themselves and not in the surrounding fluid. Brown's contribution was to establish Brownian motion as an important phenomenon, to demonstrate its presence in inorganic as well as organic matter and to refute by experiment wrong explanations of the phenomenon.

The topic was neglected in the first part of the 19'th century but awareness of the pheomenon remained widespread. From 1860 on, many scientists worked on the phenomenon. The first one, to express a notion close to the modern theory of Brownian motion was **N. Wiener** in 1863. Careful experiments and arguments lead to the kinetic theory that Brownian motion is caused by bombardment by the molecules of fluid. But the results failed the theory by a factor of about 100'000. The difficulty was the fact that the motion is very irregular composed of translations and rotations and that the trajectory appears to have no tangent. So, any attempt to determine the velocity of the particles failed. The success of **Einstein**'s theory of Brownian motion was largely due to his go around this question.

Einstein himself was unaware of the phenomenon Brownian motion. He predicted it on theoretical grounds and formulated a correct quantitative theory of it. Einsteins's arguments do not give a dynamical theory of Brownian motion. It only determines the nature of the motion and the value of the diffusion coefficients on the basis of some assumptions.

A modern probabilistic treatment of Brownian motion became possible in this century with an axiomatically developed theory of probability and stochastic processes.

PROOF. CONSTRUCTION OF BROWNIAN MOTION.

I) LEMMA. Given a Hilbert space $(H, ||\cdot||)$ that is a vector space on which there is a scalar product. There exists a probability space $(\Omega, \mathcal{A}, P)$ and a family $X(h), h \in H$ of random variables such $h \mapsto X(h)$ is linear, and $X(h)$ is Gaussian with mean zero and $E[X(h)^2] = ||h||^2$.
PROOF. Pick an orthonormal basis e_n in H and attach to each e_n a centered Gaussian IID random variable g_n satisfying $||g_n||_2 = 1$. Given a general $h = \sum h_n e_n \in H$, define $X(h) = \sum_n h_n g_n$ which converges in $\mathcal{L}^2$. Since g_n are independent, they are orthonormal in $\mathcal{L}^2$ so that $||X(h)||_2^2 = \sum_n h_n^2 ||g_n||_2 = \sum_n h_n^2 = ||h||_2^2$.

II) Take $H = L^2(\mathbb{R}^+, dx)$. For a measurable set $A \subset \mathbb{R}^+$, define $X(A) = X(1_A)$. The vector space $X(H) \subset \mathcal{L}^2$ is a Hilbert space isomorphic to H and in particular $E[X(h)X(h')] = (h, h')$. From I), we know that h and h' are orthogonal if and only if $X(h)$ and $X(h')$ are independent and that $E[X(A)X(B)] = Cov[X(A), X(B)] = (1_A, 1_B) = |A \cap B|$. Especially $X(A)$ and $X(B)$ are independent if and only if A and B are disjoint.

III) DEFINITION OF B_t. Define $B_t = X([0,t])$. This process has independent increments $B_{t_i} - B_{t_{i-1}}$ and is a Gaussian process. For each t, we have $E[B_t^2] = t$ and for $s < t$, the increment $B_t - B_s$ has variance $t - s$ so that $E[B_s B_t] = E[B_s^2] + E[B_s(B_t - B_s)] = E[B_s^2] = s$. This model of Brownian motion has therefore everything except continuity!

IV) KOLMOGOROV LEMMA. Given a process $X_t, t \in [a,b]$ for which there exists $p > r, K$ such that $E[|X_{t+h} - X_t|^p] \leq K \cdot h^{1+r}$ for every $t, t+h \in [a,b]$. Then X_t has a modification Y_t which is continuous: $|Y_t(\omega) - Y_s(\omega)| \leq C(\omega)\ |t-s|^\alpha$, where $0 < \alpha < r/p$
PROOF. We can assume $a = 0, b = 1$. Define $\epsilon = r - \alpha p$. By Chebychev-Markov inequality $P[|X_{t+h} - X_t|] \geq |h|^\alpha] \leq |h|^{-\alpha p} E[|X_{t+h} - X_t|^p] \leq K|h|^{1+\epsilon}$, so that $P[|X_{(k+1)/2^n} - X_{k/2^n}| \geq 2^{-n\alpha}] \leq K 2^{-n(1+\epsilon)}$. Therefore

$$\sum_{n=1}^{\infty} \sum_{k=0}^{2^n-1} P[|X_{(k+1)/2^n} - X_{k/2^n}| \geq 2^{-n\alpha}] < \infty .$$

By Borel-Cantelli's lemma, there exists $n(\omega) < \infty$ almost everywhere such that for all $n \geq n(\omega)$ and $k = 0, \dots, 2^n - 1$

$$|X_{(k+1)/2^n}(\omega) - X_{k/2^n}(\omega)| < 2^{-n\alpha} .$$

Let $n \geq n(\omega)$ and $t \in [k/2^n, (k+1)/2^n]$ of the form $t = k/2^n + \sum_{i=1}^m \gamma_i / 2^{n+i}$ with $\gamma_i \in \{0,1\}$. Then $|X_t(\omega) - X_{k2^{-n}}(\omega)| \leq \sum_{i=1}^m \gamma_i 2^{-\alpha(n+i)} \leq d2^{-n\alpha}$ with $d = (1 - 2^{-\alpha})^{-1}$. Similarly

$$|X_t - X_{(k+1)2^{-n}}| \leq d2^{-n\alpha} .$$

Given $t, t+h \in D = \{k2^{-n} \mid n \in \mathbb{N}, k = 0, \dots n-1\}$. Take n so that $2^{-n-1} \leq h < 2^{-n}$ and k so that $k/2^{n+1} \leq t < (k+1)/2^{n+1}$. Then $(k+1)/2^{n+1} \leq t+h \leq (k+3)/2^{n+1}$ and $|X_{t+h} - X_t| \leq 2d2^{-(n+1)\alpha} \leq 2dh^\alpha$. For almost all ω, this holds for sufficiently small h. We know now that for almost all ω, the path $X_t(\omega)$ is uniformly continuous on the dense set of dyadic numbers D. Such a function can be extended to a continuous function on $[0,1]$ by defining

$$Y_t(\omega) = \lim_{s \in D \to t} X_s(\omega) .$$

Since the inequality in the assumption of the theorem implies $E[X_t(\omega) - \lim_{s \in D \to t} X_s(\omega)] = 0$ and by Fatou's lemma in measure theory $E[Y_t(\omega) - \lim_{s \in D \to t} X_s(\omega)] = 0$ we know that $X_t = Y_t$ almost everywhere. Y is therefore a modification of X. Moreover, Y satisfies for all s, t,

$$|Y_t(\omega) - Y_s(\omega)| \leq C(\omega)\ |t-s|^\alpha .$$

V) PROOF OF THE CLAIM. In one dimensions, take B_t from above. Since $X_h = B_{t+h} - B_t$ is centered with variance h, we have $E[X_h^4] = \frac{d^4}{dx^4} \exp(-x^2 h/2)_{|x=0} = 3h^2$, so that $E[(B_{t+h} - B_t)^4] = 3h^2$. By Kolmogorov's lemma there is a continuous modification of B. To define standard Brownian motion in n-dimension, we take the joint motion $B_t = (B_t^{(1)}, \dots, B_t^{(n)})$ of n independent one- dimensional Brownian motions.

Math 464/564, O. Knill, **May 1997**

Option pricing: Black-Scholes formula

The theory of stochastic processes both with discrete or continuous time can be applied to problems in financial economics. A celebrated example is the Black-Scholes option pricing formula (1973) which is usually formulated in terms of continuous time stochastic processes.

Given $a < r < b < \infty$, define $p = (r-a)/(b-a)$. Let E_n be a random walk, that is a sequence of IID random variables satisfying $P[E_n = 1] = p, P[E_n = -1] = 1-p$.

DEFINITION OF AN ECONOMY OF STOCKS AND BONDS. We define two processes B_n, S_n, where B stands for **bonds** with fixed **interest rate** r, and S for **stocks** with **fluctuating interest rates** $R_n = (a+b)/2 + E_n(a-b)/2$. The definition is

$$\begin{aligned} B_n &= (1+r)B_{n-1}, B_0 = 1 \\ S_n &= (1+R_n)S_{n-1}, S_0 = 1 \ . \end{aligned}$$

We see that B_n satisfies the **difference equations** $B_n - B_{n-1} = rB_n$ and S_n satisfies the **stochastic difference equation** $S_n - S_{n-1} = R_n S_{n-1}$.

DEFINITION OF A PORTFOLIO AND FORTUNE. A with respect to R_n previsible sequence of pairs of random variables A_n, V_n is called a **portfolio**. Just after time n, you have A_n units of stock and V_n units of bonds. (A negative value of A_n means "short selling" of stocks, a negative value of V means borrowing with a fixed interest rate r.) We define our **fortune** X_n by $X_0 = x$ and $X_n - X_{n-1} = A_n(S_n - S_{n-1}) + V_n(B_n - B_{n-1})$. Using $B_n - B_{n-1} = rB_n$ and $S_n - S_{n-1} = R_n S_{n-1}$, we have the recursion

$$X_n = (1+r)X_{n-1} + A_n S_{n-1}(R_n - r) \ .$$

COMPARISON WITH THE GAMBLING SYSTEM MET EARLIER. To relate this with a gambling system, we write $R_n - r = \frac{1}{2}(b-a)(Z_n - Z_{n-1})$ with $Z_n = \sum_{k=1}^n (E_k - 2p + 1)$ having expectation $E[Z_n] = 0$. The process $Y_n := (1+r)^{-n} X_n$ satisfies then

$$\begin{aligned} Y_n - Y_{n-1} &= (1+r)^{-n} A_n S_{n-1}(R_n - r) \\ &= \frac{1}{2}(b-a)(1+r)^{-n} A_n S_{n-1}(Z_n - Z_{n-1}) \\ &= C_n(Z_n - Z_{n-1}) \end{aligned}$$

showing that $Y = \int C \, dZ$ is your winning in a fair random walk Z_n with **gambling system** C_n.

DEFINITION OF THE EUROPEAN OPTION. The **European option** is a contract made at time 0 which will allow you to buy one unit of stock later at time N with prize K. If you buy such an option, then you will make the buy of the stock if the value S_N of the stock at time N satisfies $S_N > K$ and not if $S_N < K$. Your win is $(S_N - K)^+$.

DEFINITION OF A HEDGING STRATEGY. A **hedging strategy** with initial fortune x, **strike time** N and **prize** K is a **portfolio management scheme** $\{(A_n, V_n)\}_{1 \leq n \leq N}$ defined above, where $X_0 = x$, $X_n \geq 0$ (you never go bankrupt) and $X_N = (S_N - K)^+$, where $z^+ = \max\{0, z\}$.

BLACK-SCHOLES FORMULA. For $x = E[(1+r)^{-N}(S_N - K)^+]$, a unique hedging strategy exists.

PROOF. Define Y_n be the expectation of $(1-r)^{-N}(S_N - K)^+$ under the condition that $R_1, \dots, R_n$ are known. The recursion

$$Y_n - Y_{n-1} = (1-r)^{-n} A_n S_{n-1}(R_n - r)$$

defines the previsible process A_n. With $X_n = (1+r)^n Y_n$, define $V_n = (X_n - A_n S_n)/B_n$. One can actually show $A_n \geq 0$ so that no short selling is necessary.

Math 464/564, O. Knill, **May 1997**

Feynman path integrals

THE PROBLEM. In quantum mechanics, the Schrödinger equation $i\hbar\dot{u} = Hu$ defines the evolution of the wave function $f(t) = U^t f = e^{-itH/\hbar} f(0)$ in a Hilbert space $\mathcal{H}$. The operator H is the **Hamiltonian**. If $\mathcal{H} = L^2(\mathbb{R}^d)$, $H = H_0 + V$, where $H_0 = -\Delta/2$ is the Hamiltonian of a free particle and $V : \mathbb{R}^d \to \mathbb{R}$ is the potential the operator is called a **Schrödinger operator**. How do we compute $(f, U^t g)$?

DEFINITION. A linear operator $A : D(A) \subset \mathcal{H} \to \mathcal{H}$ is **symmetric** if $(Au, v) = (u, Av)$ for all $u, v \in D(A)$ and **selfadjoint**, if it is symmetric and $D(A) = D(A^*)$.

BACKGROUND. One has to restrict the opeator H to a vector space $D(H) \subset \mathcal{H}$ called **domain** containing the dense set $C_0^\infty(\mathbb{R}^d)$ of all smooth functions vanishing at infinity. Define $D(A^*) = \{u \in \mathcal{H} \mid v \mapsto (Av, u) \textit{ is a bounded linear functional on } D(A)\}$. If $u \in D(A^*)$, then there exists a unique function $w = A^*u \in \mathcal{H}$ such that $(Av, u) = (v, w)$ for all $u \in D(A)$. This defines the **adjoint** A^* of A with domain $D(A^*)$.

FEYNMAN'S IDEA. Assume $H = H_0 + V$ is selfadjoint. Then

$$e^{-itH} u(x_0) = \lim_{n \to \infty} (\frac{2\pi it}{n})^{-d/2} \int_{(\mathbb{R}^d)^n} e^{iS_n(x_0, x_1, x_2, \dots, x_n, t)} u(x_n) \, dx_1 \dots dx_n$$

where

$$S_n(x_0, x_1, \dots, x_n, t) = \frac{t}{n} \sum_{i=1}^{n} \frac{1}{2} (\frac{|x_i - x_{i-1}|}{t/n})^2 - V(x_i) \ .$$

This is essentially a consequence of the following formula which generalizes the Lie product formula

$$\lim_{n \to \infty} (\exp(A/n) \exp(B/n))^n = \exp(A + B)$$

for finite dimensional matrices A, B. One says X_n **converges strongly** to X if $X_n f \to Xf$ for all $f \in \mathcal{H}$ and writes $X = s - \lim_{n \to \infty} X_n$.

TROTTER PRODUCT FORMULA. Given selfadjoint operators A, B defined on $D(A), D(B) \subset \mathcal{H}$. Assume $A + B$ is selfadjoint on $D = D(A) \cap D(B)$, then $e^{it(A+B)} = s - \lim_{n \to \infty} (e^{itA/n} e^{itB/n})^n$. If A, B are bounded from below, then $e^{-t(A+B)} = s - \lim_{n \to \infty} (e^{-tA/n} e^{-tB/n})^n$.

Unfortunately, there are problems to perform this summation in general over all paths in the limit when the time intervals $[x_n, x_{n-1}]$ goes to zero. However, when replacing t by t/i, one gets $(f, e^{-tH} g)$ which can be computed using Brownian motion B_t defined over a probability space $(\Omega, \mathcal{A}, P)$. If $V \in C_0^\infty(\mathbb{R}^n)$, the integral $\int_0^t V(B_s(\omega)) \, ds$ can be taken for each ω as a limit of Rieman sums. Then, $\int_0^t V(B_s) \, ds$ is a random variable. The existence of Brownian motion defines a probability measure on the space of all continuous paths on $\mathbb{R}^n$. Integration with respect to this measures is denoted by $\int \cdot dB$.

FEYNMAN-KAC FORMULA.

$$(f, e^{-tH} g) = \int \overline{f}(B_0) g(B_t) e^{-\int_0^t V(B_s)) \, ds} \, dB \ .$$

This is an integral over all continuous paths B_s starting at zero.

The Feynman-Kac formula is useful in mathematical physics: it allows to treat operators with magnetic fields or to compute groundstates and groundstate energies perturbatively. The concept of functional integration is a way of quantisation which generalizes to more situations, where canonical quantisation (replacing classical variables by operators) is not available.

1) PROBLEM.

> 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 $e_1, e_2, \dots$, then the system will be in the **Bolzman distribution** which is given by $P_\beta[\{j\}] = e^{-\beta e_j}/Z(\beta)$, where $Z(\beta) = \sum_{j=1}^{\infty} e^{-\beta e_j}$ is the normalization factor called **partition function**. The probabilities $P_\beta[\{j\}]$ 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, $(\Omega, \mathcal{A}, P)$= 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

$$\Omega = \mathbb{N} = \{0, 1, 2, \ldots\}, \mathcal{A} = \{A \subset \Omega\}, P_\beta[\{j\}] = \frac{e^{-\beta e_j}}{Z(\beta)} ,$$

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

$$E_\beta[X] = -\frac{\frac{d}{d\beta} Z(\beta)}{Z(\beta)} ,$$

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 $E_\beta[X]$, in the case $e_j = j$.

Additional information: Planck's law of black body radiation was crucial for the development of quantum mechanics. No continuous energy distribution could explain this law. So, the first time, it was established that for some systems, the energy can only take discrete ("quantized") values.
The assumption $e_j = j$ in this exercice should in a physical situation be replaced by $e_j = e_0 + j\hbar\omega$, where ω is the frequency of the oscillator, e_0 is the ground

state energy and $\hbar$ is the Planck constant. The inverse temperature β stands for $(kT)^{-1}$, where k is the Bolzman factor and T is the temperature.

SOLUTION a) (4) We compute

$$\frac{d}{d\beta}Z(\beta) = -\sum_{j=0}^{\infty} e_j \cdot e^{-\beta e_j}$$

and so

$$-\frac{\frac{d}{d\beta}Z(\beta)}{Z(\beta)} = \sum_{j=0}^{\infty} e_j \cdot e^{-\beta e_j}/Z(\beta) = \sum_{j=0}^{\infty} X(j)P_\beta[\{j\}] = E[X] \, .$$

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

$$E_\beta[X] = -\frac{\frac{d}{d\beta}Z(\beta)}{Z(\beta)} = \frac{e^{-\beta}}{1-e^{-\beta}} = \frac{1}{e^\beta - 1} \, .$$

2. PROBLEM.

> 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 $P : \mathcal{A} \to [0, 1]$

$$P[A] = \sum_{k \in A} \log_{10}(\frac{k+1}{k})$$

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 $E[X]$. (Simplify the result as much as possible. No numerical evaluation of the result is required).

SOLUTION. a) (6) Since $P[\{j\}] > 0$, we have also $P[A] \geq 0$ 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:

$$\begin{aligned} P[\Omega] &= \sum_{k=1}^{9} \log_{10}(\frac{k+1}{k}) = \sum_{k=1}^{9} \log_{10}(k+1) - \sum_{k=0}^{9} \log_{10}(k) \\ &= \log_{10}(10) - \log_{10}(1) = 1 \ . \end{aligned}$$

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: $E[X] = 3.44024$.

Math464/564 O. Knill, **January 1997**

Homework 1

Due: Tuesday, January 28, 1997 in class

Topics: Boolean algebras, Probability spaces, Modelling of probability spaces.

Remark. This HW contains 5 questions. Each question can give 10 points, so that you can get a maximal credit of 50 points for this HW.

1) (10 points) Topic: **Boolean algebras: a group**. Let $\mathcal{A}$ be a Boolean algebra. 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) $A, B \in \mathcal{A} \Rightarrow A \Delta B \in \mathcal{A}$.
b) For all A, B, C the **associativity law** $(A \Delta B) \Delta C = A \Delta (B \Delta C)$ holds.
c) There exists a **zero element** $A_{zero} \in \mathcal{A}$ such that $A \Delta A_{zero} = A$ for all $A \in \mathcal{A}$.
d) For all A, there exists an **inverse** B such that $A \Delta B = A_{zero}$.
e) For all A, B the **commutativity law** holds $A \Delta B = B \Delta A$.

2) (10 points) Topic: **Probability spaces: internet**. Consider a part of the internet computer network consisting of 3 nodes (computers). Assume that two of the three computers are connected with probability p and that with probability $q = 1 - p$, the connection is broken.
a) Construct the probability space $(\Omega, \mathcal{A}, P)$ and find the event that all nodes have connection to all other nodes.
b) What is the probability that one can reach from any of the computers any other computer.

3) (10 points) Topic: **Probability spaces: dices**. Suppose two dice are rolled once and that each of the 36 possible outcomes are equally likely. What is the probability that the product of the two numbers on the two faces is even?

4) (10 points) Topic: **Probability spaces: darts**. We are throwing darts onto a disc of radius 1 and assume that the dart hits each region in the disc equally likely. What is the probability that two darts hit both a point in distance $\leq 1/2$ from the center.

5) (10 points) Topic: **Probability spaces: roulette**. A weel of circumferecne 2π is made into a probability space by assigning the probability that a ball lands in an arc of length s is $s/2\pi$. The weel is divided into 37 zones numbered $0, \dots 36$. The zones with even numbers except are painted black, while the odd numbers are painted red, The zone with number 0 is painted green. What is the probability that in two games, both red and black occur?

Math464/564 O. Knill, **January 1997**

Homework 2

Due: Tuesday, February 4, 1997 in class

Topics: Modelling of probability spaces, conditional probability, independence.

Remark. This HW contains 5 questions. Each question can give 10 points, so that you can get a maximal credit of 50 points for this HW.

1) (10 points) Topic: **Modeling probability spaces, a political problem.** The president of a country comes from a family of two children. What is the probability that the other child is his sister?

2) (10 points) Topic: **Conditional probability spaces, Polya's urn scheme.** An urn has 16 balls, 12 red and 4 black balls. Without putting the balls back, one draws two balls, one after the other mixing the balls in urn before each drawing. Assume the second ball is red. What is the probability that the first ball is red.

3) (10 points) Topic: **Independence, the complement events.**
Let A and B be two independent events.
a) (5 points) Prove that A and B^c are independent.
b) (5 points) Prove that A^c and B^c are independent.

4) (10 points) Topic: **Independence, overbooking of flights**
Experience shows that 5 % of the people reserving flights do not show up. Airlines therefore decide to overbook the flights and take the risk to pay out people with no seat. Assume a plane has 230 seats and that the airline sells 233 tickets. What is the probability that all people will be accommodated?

5) Topic: (10 points) Topic: **Probability spaces, combinatorics, the wardrobe problem.**
We distribute randomly $n \in \mathbb{N}$ coats of n people and wonder what is the probabilihy that no person gets his or her coat.
a) (2 poits) Set up the probability space $(\Omega, \mathcal{A}, P)$ for the situation.
b) (2 points) 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 $1 \leq i_1 < i_2 < \ldots < i_k \leq n$.
c) (3 points) 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 proven in class

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

d) (3 points) What happens with the probability $p(n)$ for $n \to \infty$?

Math464/564 O. Knill, **February 1997**

Homework 3

Due: Tuesday, February 11, 1997 in class

Topics: Combinatorics and finite probability spaces.

Remark. This HW contains 5 questions. Each question can give 10 points, so that you can get a maximal credit of 50 points for this HW.

1) (10 points) Topic: **Permutations and maps**.
A permutation of a set $X = \{1, 2, \dots, n\}$ is a map on this set which is bijective. A general map from X to X is not a permutation. What is the probability that a random map on X is a permutation?

2) (10 points) Topic: **Combinations. Random walk**.
A random walker on the line moves in each step with probability 1/2 to the right and with probability 1/2 to the left. Assume the walker starts at 0.
a) What is the probability that the walker is returning back to the origin after n steps.
b) Each step has length 1. What is the probability that the walker is in distance $k \in \mathbb{N}$ from the origin after n steps.

3) (10 points) Topic: **Scrabble**.
While playing scrabble, you get the letters $F, E, R, I, T, R, C, F, I$. How many words can you write with these letters? An example is $TERRIFFIC$.

4) (10 points) Topic: **Amino acids**.
The genetic code of all living cells is built by nucleotides. There are four kind of those T(hymine),A(denine),C(ytosine),G(uanine) which are used to build up the DNA (DeoxyriboNucleicAcid). An amino acid is specified by a sequence of three nucleotides. How many amino acids can be coded in this manner?

5) (10 points) Topic: **RSA and factoring integers by a probabilistic method**. Let n be a number and p a factor of n. Given m random numbers out of the set $X = \{1, \dots, n\}$. What is the probability that two of them are congruent modulo p? Hint: the number n is irrelevant, you can as well take random numbers modulo p.

MOTIVATING STORY FOR PROBLEM 5) The RSA scheme is a widely used encryption scheme for example used on the internet. Its security is based on the fact that it is hard to factor integers with large prime factors and that this problem is considered by so many people that nobody could resist publishing a better method. The best current factoring algorithms can handle numbers with say up 150 integers (currently nobody is able to factor 200 digit numbers which are products of two 100 digit primes). There are current challenges on the internet where prizes are given out for successes (this is also politically motivated, the last week a prize was won, see the article below).
One method for factoring integers is the Pollard ρ method. This method was used in 1980 to factor the first time the eight's Fermat number $n = F_8 = 2^{(2^8)} + 1$. The idea to find a prime factor p like here $p = 1238926361552897$ of $n = F_8$ is to produce randomly numbers $1 \leq x_k \leq n$ for example with a pseudo random number generator $x_{k+1} = x_k^2 + 3 \; mod \; n$ and to check whether any pair $x_k - x_l$ has a common prime factor with n by forming the product $Q_m = \prod_{j=1}^{m} (x_{2j} - x_j) \; mod \; n$ finding occasionally the greatest common divisor $GCD(Q_m, n)$. If two numbers x_k, x_l have a common prime factor p among the m elements $x_1, \dots, x_m$, then it shows up in $GCD(Q_m, n)$.

(*) If you have time and if you are interested in the extremely fascinating story of computational number theory (which has a lot of probabilistic stuff inside), try this:
How big does m have to be, in order to factor $n = 2^{(2^8)} + 1$ with probability 1/2 with the

Pollard rho method?

Here is my Mathematica implementation of the simplest Pollard -ρ method (you can find the source code on the web site of the course). The six'th Fermat number has already 20 digits and a usual "Baby method"="trying out all possible factors" up to $\sqrt{n}$ would be hopeless already. So, this primitive algorithm is already quite effective. One can improve the method.

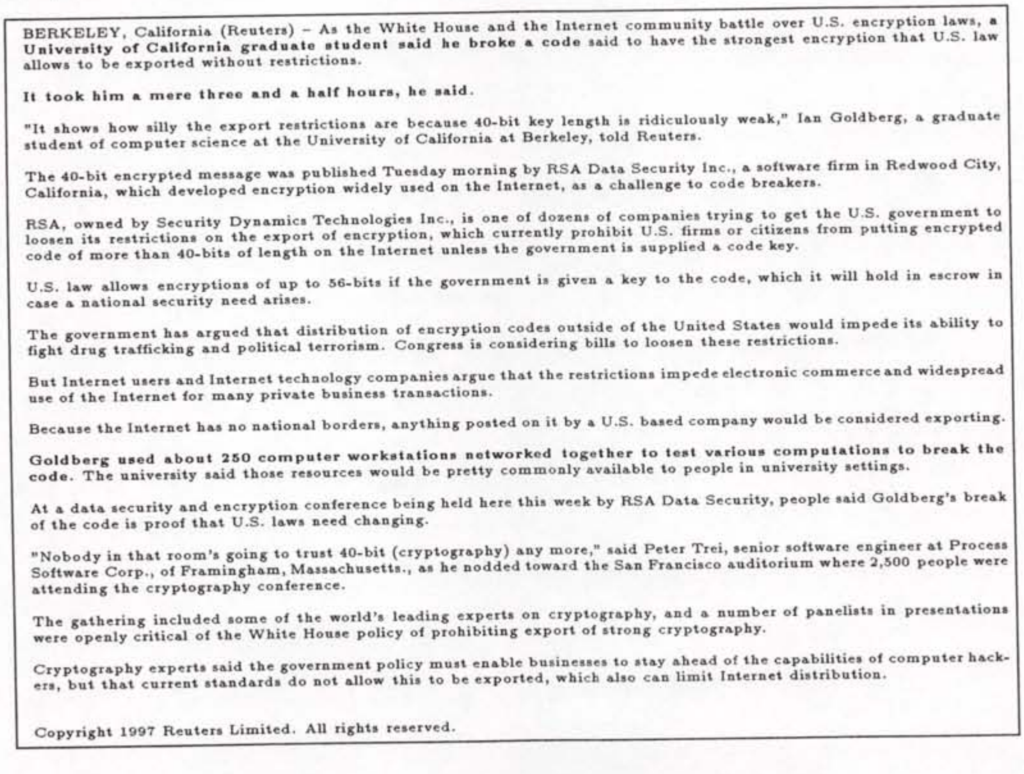
```
FactorWithPollard[n_] := Module[{a=3, x=17, y=17, q=1},
   While[q<2,
     Do[x=Mod[x*x-a,n];
        y=Mod[y*y-a,n];
        y=Mod[y*y-a,n];
        q=Mod[q*(x-y),n],{i,20}];
     q=Mod[GCD[n,x-y],n] ];q];

FermatNumber[n_]:=2^(2^n)+1;
Example=FactorWithPollard[FermatNumber[6]]
```

Bonus question: Why is the method called ρ? Hint: look at the form of the letter ρ and relate it with what the algorithm does.

See the cite:

```
http://www.cnn.com/TECH/9701/30/encryption.reut/index.html
```

about the story of the "California student unscrambling the internet code" (January 30, 1997 in CNN) from last week!



BERKELEY, California (Reuters) – As the White House and the Internet community battle over U.S. encryption laws, a University of California graduate student said he broke a code said to have the strongest encryption that U.S. law allows to be exported without restrictions.

It took him a mere three and a half hours, he said.

"It shows how silly the export restrictions are because 40-bit key length is ridiculously weak," Ian Goldberg, a graduate student of computer science at the University of California at Berkeley, told Reuters.

The 40-bit encrypted message was published Tuesday morning by RSA Data Security Inc., a software firm in Redwood City, California, which developed encryption widely used on the Internet, as a challenge to code breakers.

RSA, owned by Security Dynamics Technologies Inc., is one of dozens of companies trying to get the U.S. government to loosen its restrictions on the export of encryption, which currently prohibit U.S. firms or citizens from putting encrypted code of more than 40-bits of length on the Internet unless the government is supplied a code key.

U.S. law allows encryptions of up to 56-bits if the government is given a key to the code, which it will hold in escrow in case a national security need arises.

The government has argued that distribution of encryption codes outside of the United States would impede its ability to fight drug trafficking and political terrorism. Congress is considering bills to loosen these restrictions.

But Internet users and Internet technology companies argue that the restrictions impede electronic commerce and widespread use of the Internet for many private business transactions.

Because the Internet has no national borders, anything posted on it by a U.S. based company would be considered exporting.

Goldberg used about 250 computer workstations networked together to test various computations to break the code. The university said those resources would be pretty commonly available to people in university settings.

At a data security and encryption conference being held here this week by RSA Data Security, people said Goldberg's break of the code is proof that U.S. laws need changing.

"Nobody in that room's going to trust 40-bit (cryptography) any more," said Peter Trei, senior software engineer at Process Software Corp., of Framingham, Massachusetts., as he nodded toward the San Francisco auditorium where 2,500 people were attending the cryptography conference.

The gathering included some of the world's leading experts on cryptography, and a number of panelists in presentations were openly critical of the White House policy of prohibiting export of strong cryptography.

Cryptography experts said the government policy must enable businesses to stay ahead of the capabilities of computer hackers, but that current standards do not allow this to be exported, which also can limit Internet distribution.

Math464/564 O. Knill, **February 1997**

Homework 4

Due: Tuesday, February 18, 1997 in class

Topics: Discrete random variables, discrete densities, distribution functions, independence, geometric densities.

Remark. This HW contains 5 questions. Each question can give 10 points, so that you can get a maximal credit of 50 points for this HW.

> REMINDER OF SOME DEFINITIONS:
> **A discrete random variable** is a real-valued function X on a probability space $(\Omega, \mathcal{A}, P)$, which has the property that $X(\Omega)$ is a discrete set $\{x_1, x_2, \dots, \}$ in the real line $\mathbb{R}$ and such that the sets $A_j = X^{-1}(x_j) = \{\omega \in \Omega \mid X(\omega) = x_j\}$ are in $\mathcal{A}$. If we replace $\mathbb{R}$ by $\mathbb{R}^d$, we speak of a **random vector** $X = (X_1, \dots, X_d)$.
> The **discrete density function** f of a random variable X is the function on $\mathbb{R}$ given by $f(x) = P[X = x]$. For a random vector $X = (X_1, \dots, X_d)$ the **discrete density function** is the function $f(x_1, \dots, x_d) = P[X_i = x_i]$ on $\mathbb{R}^d$.
> The **distribution function** of a random variable X is the function $F_X : \mathbb{R} \to \mathbb{R}$ given by $F_X(y) = P[X \leq y]$. For a random vector, the distribution function is the function $F_X(x_1, \dots, x_d) = P[X_i \leq x_i]$ on $\mathbb{R}^d$.
> Two random variables or random vectors X, Y are **independent** if the sets $A_i = X^{-1}(x_i)$ and $B_j = Y^{-1}(y_j)$ are independent events for all pairs $i \neq j$.

1) (10 points) Topic: **Definition of discrete random variables** .
a) How many random variables do there exist on the probability space

$$(\Omega = \{1, \dots, 37\}, \mathcal{A} = \{A \subset \Omega\}, P[A] = |A|/|\Omega|]) .$$

b) Find all random variables on the probability space $(\Omega = \{1, 2, \dots, 1000\}, \mathcal{A} = \{\emptyset, \Omega\}, P[\emptyset] = 0, P[\Omega] = 1)$.
c) Is it true that on every finite probability space, where $\mathcal{A}$ is the set of subsets of Ω, and P is an arbitrary probability measure, every real function is a random variable?
d) Let $(\Omega = [0, 1], \mathcal{A} = \{measurable\ sets\}, P[[a, b]] = b - a)$.
Is $X(x) = x$ a discrete random variable?
e) Find a discrete random variable X on $(\Omega = [0, 1], \mathcal{A} = \{measurable\ sets\}, P[[a, b]] = b - a)$ such that $X(\Omega) = \mathbb{N}$.

2) (10 points) Topic: **Density and distribution functions**. Let $\Omega = \{\omega = (\omega_1, \omega_2, \omega_3, \omega_4) \mid \omega_i \in \{0, 1\}\ \}$ be the probability space of throwing a dime 4 times. Let $X(\omega) = \sum_{i=1}^{4} \omega_i$. Draw the discrete density function and the distribution function of X.

3) (10 points) Topic: **Independent random variables**. Let $\Omega = \{\omega = (\omega_1, \omega_2) \mid \omega_i, \omega_2 \in \{1, \dots, 6\}\}, \mathcal{A} = \{A \subset \Omega\}, P[A] = |A|/|\Omega|\}$ be the probability space of throwing two dices. Consider the two random variables $X(\omega) = \omega_1, Y(\omega) = \omega_2$. Determine explicitely the events $X^{-1}(3) = \{$ the first dice is 3 $\}$ and $Y^{-1}(5) = \{$ the second dice is 5 $\}$ and verify that they are indeed independent events.

4) (10 points) Topic: **Indendent random variables** Show that if two random varables X, Y are independent, then also X^2 and Y^2 are independent.

5) (10 points) Topic: **Geometric density**. Let X, Y be independent random variables having geometric densities with parameters p and q. Find
a) $P[X \geq Y]$.
c) The density of $X + Y$.

Math464/564 O. Knill, **February 1997**

Homework 5

Due: Tuesday, February 25, 1997 in class

Topics: Discrete random variables, independence of random variables, discrete random vectors.

Remark. This HW contains 3 questions. The number of questions is reduced because of possible midterm stress this week. You can here get a maximal credit of 30 points.

1) (10 points) Topic: **Independent random variables** . What is the distribution of the sum of two independent random variables X, Y where X is Poisson distributed with parameter $\lambda = 2$ and Y is Poisson distributed with parameter $\lambda = 3$?

2) (10 points) Topic: **Poisson approximation of Bernoulli**. The production of microchips becomes more and more delicate. What is the probability that 2 chips of a silicon waffer containing 1000 chips are defect, if a chip is defect with probability 0.01. Approximate the Bernoulli distribution by a suitable Poisson distribution.

3) (10 points) Topic: **Random vectors, multinomial distribution**. If $X = (X_1, \dots, X_r)$ is a random vector, then the distribution of X_i is called a marginal distribution of X. What is the marginal distribution of X_1 of the random vector $X = (X_1, X_2, X_3)$, if X has the multinomial distribution $P[X = k] = \left(\begin{array}{c} 10 \\ k_1, k_2, k_3 \end{array} \right) 2^{-10}$ if $k_1 + k_2 + k_3 = 10$ and $P[X = k] = 0$ if $k_1 + k_2 + k_3 \neq 10$.

Checklist for first midterm.

The material is about the topics of the first three chapters of the book which we covered as weill as the material which appeared in class. There will be a short theory part (multiple choice) and some problems in the style of the homework. You will be allowed to use all your notes, all in class distributed material, all homwork, but no printed books (the later restriction is to prevent you from "reading " and losing time during the exam and to encourage to do a private collection of notes. My suggestion for a preparation: look again at the homework problems HW1-HW4 (can I solve all problems now ?), make a short private list of important definitions, formulas, results and examples. Topics,

PROBABILITY SPACES.

- Definition of σ-algebra, probability measure, probability spaces.
- Examples of probability spaces.
- Working in the algebra of sets.
- Basic formulas like $P[A] = 1 - P[A^c]$ or the switch on switch off Formula of Sylvester.

COMBINATORICS.

- Knowledge of all basic formulas in combinatorics.
- Some formula deduced from basic formulas like Bernoulli or the formula in lotto.
- Know at least one example for each basic formula.

DISCRETE RANDOM VARIABLES.

- Definition of a discrete random variable.
- Definition of density and distribution function.
- Know from the basic distributions what is set of values, what are the parameters and an example and where it is used.

Math464/564 O. Knill, **February 1997**

Homework 6

Due: Tuesday, March 4, 1997 in class

Topic: Expectation and Variance of discrete random variables.

Remark. This HW contains 5 questions with 50 points.

1) (10 points) Topic **Expectation and Variance** Consider the probability space of throwing a dice once, that is $\Omega = \{1,2,3,4,5,6\}$ and $P[A] = |A|/|\Omega|$. a) What is the expectation and variance of the random variable $X(\omega) = \omega$?
b) What is the expectation and variance of the random variable $X(\omega) = \omega^2$?

2) (20 points) Topic: **Expectation of discrete random variables.**

> A physical system 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 $e_1, e_2, \dots$, the system will be in the **Bolzman distribution** $P_\beta[\{j\}] = e^{-\beta e_j}/Z(\beta)$, where $Z(\beta) = \sum_{j=1}^\infty e^{-\beta e_j}$ is called the **partition function.** The probabilities $P_\beta[\{j\}]$ are called the **Bolzmann-factors.** The **energy** of the state j is $e_j = X(j)$. The case $e_j = j$ is the situation of the **quantum mechanical oscillator.** The formula defined in d) is **Planck's formula** for the **total energy** of a quantum mechanical oscillator. In suitable physical units, this is **Planck's blackbody radiation formula,** which gives the average energy of the oscillator in dependence of the temperature.

Given for every real number $\beta > 0$ the discrete probability space $(\Omega, \mathcal{A}, P_\beta)$ with

$$\Omega = \mathbb{N} = \{1,2,\dots\}, \mathcal{A} = \{A \subset \Omega\}, P_\beta[\{j\}] = \frac{e^{-\beta e_j}}{Z(\beta)} ,$$

where $Z(\beta) = \sum_{j=1}^\infty e^{-\beta e_j}$. 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) (5 points) Verify that P is a probability measure. (In the midterm, the Ω contained 0).
b) (5 points) Show that

$$E_\beta[X] = -\frac{\frac{d}{d\beta} Z(\beta)}{Z(\beta)} ,$$

where E_β is the expectation with respect to $(\Omega, \mathcal{A}, P_\beta)$.
c) (5 points) Compute $Z(\beta)$ in the case, when $e_j = j$ for $j = 1, \dots$. Simplify as much as possible.
d) (5 points) Compute, using b) and c), the value of $E_\beta[X]$, in the case $e_j = j$.

3) Topic: **Expectation of discrete random variables.** (10 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. The IRS is considering doing Bedford tests on the data obtained from the taxpayer.

Consider the finite set $\Omega = \{1,2,\dots,9\}$ and the Boolean algebra $\mathcal{A} = \{A \subset \Omega\}$.
a) (5) Show that $P : \mathcal{A} \to [0,1]$

$$P[A] = \sum_{k \in A} \log_{10}(\frac{k+1}{k})$$

is a probability measure, where $\log_{10}$ is the logarithm with respect to the base 10.
b) (5) Let X be the random variable, which gives the first digit $X(k) = k$. Compute $E[X]$. (Simplify the result as much as possible).

4) (10 points). Assume X is geometrically distributed with parameter λ and Y is Poisson distributed with parameter μ.
a) What is the expectation of $X + Y$?

REMINDER OF SOME DEFINITIONS

A **random variable** on a probability space $(\Omega, \mathcal{A}, P)$ is a function $X : \Omega \to \mathbb{R}$ such that for all $x \in \mathbb{R}$, the event $\{X = x\}$ is in $\mathcal{A}$. The random variable X is **discrete** if $X(\Omega)$ is a discrete set.

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

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

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

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

Convenient is the formula $Var[X] = E[X^2] - E[X]^2$.
The **standard deviation** of X is

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

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

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

We know $Cov[X, Y] = E[XY] - E[X]E[Y]$.

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

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

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

The **correlation coefficient** $\rho(X, Y)$ of two random variables with $Var[X] > 0, Var[Y] > 0$ is defined as

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sigma[X]\sigma[Y]} \, .$$

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

$$P[X = x; Y = y] = P[X = x] \cdot P[Y = y] \, .$$

The **regression line** of two random variables X, Y is the line $y = ax + b$, where

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

REMINDER OF SOME FACTS

$E[XY] = E[X]E[Y]$ if X and Y are independent.

If X and Y are independent, then $Cov[X, Y] = 0$.

Schwartz inequality $E[XY]^2 \leq E[X^2]E[Y^2]$. It implies that $|\rho(X, Y)| \leq 1$.

Chebychev inequality. $P[|X - E[X]| \geq \epsilon] \leq Var[X]/\epsilon^2$

If $y = ax + b$ is the regression line of two random variables X and Y, then the random variable $\tilde{Y} = aX + b$ minimizes $Var[Y - \tilde{Y}]$ under the constraint $E[Y] = E[\tilde{Y}]$. It is the best guess for Y,

Math464/564 O. Knill, **March 1997**

Homework 7

Due: Tuesday, March 11, 1997 in class

Topic: Expectation and Variance, Correlation Probability generating function, Correlation coefficient.

Remark. This HW contains 5 questions. Each question can give 10 points, so that you can get a maximal credit of 50 points for this HW.

1) (10 points) Topic: **Probability generating function**. a) (5 points) Compute the expectation and variance of a (n, p)-Bernoulli distributed random variable using the probability generating function. b) (5 points) Compute the expectation and variance of a Poisson distributed random variable using the probability generating function.

2) (10 points) Topic: **Covariance and independence**. Given two independent random variables X, Y satisfying $Var[X] = Var[Y] = 1$.

a) Compute $Var[3X + 5Y]$.

b) Compute the correlation coefficient $\rho(2X + Y, X)$.

3) (10 points) Topic: **Expectation and Variance**. The probability space in roulette is defined by $(\Omega = \{0, 1, 2, \dots, 36\}, \mathcal{A} = \{A \subset \Omega\}, P[\{i\}] = 1/37)$. 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) (5 points) Determine $E[X]$, which is the expected win or loss.

b) (5 points) Compute $Var[X]$ which is a measure for the risk of this option.

4) (10 points) Topic: **Correlation.** Given two discrete random variables X, Y which have both positive variance.
a) For each $\theta \in [0, 2\pi]$, define

$$\begin{aligned} X_\theta &= X\cos(\theta) - Y\sin(\theta) \ , \\ Y_\theta &= X\sin(\theta) + Y\cos(\theta) \ . \end{aligned}$$

Verify that there exists a value of θ for which X_θ, Y_θ are uncorrelated.

5) (10 points) Topic: **Chebychev inequality**. We throw a dime 100 times. Let X be the number of heads.

Give with Chebychev's inequality an upper bound for the probability that $|X - 50| \geq 10$.

Math464/564 O. Knill, **March 1997**

Homework 8

Due: Tuesday, March 25, 1997 in class

Topic: Review of the first half of the course.

Remark. This HW contains 5 questions. Each question can give 10 points, so that you can get a maximal credit of 50 points for this HW. This problem set is a review and covers also older topics. The aim is to improve your routine before we go into the next half of the course after the spring break.

1) (10 points) Topic: **Probability space.**

a) Prove $P[A] = P[A \cap B] + P[A \cap B^c]$.
b) Prove $P[A \Delta A] = 0$.
c) Find $P[A \cap B]$ if A and B are independent events of probability $1/4$.
d) Prove $1 = P[B|A] + P[B^c|A]$.
e) Prove the formula $P[A \cap B] \geq P[A] + P[B] - 1$.

2) (10 points) Topic: **Combinatorics.**

a) What is the probability to have 4 kings in a poker hand. (5 cards out of 52)?
b) A family has three childs. What is the probability that two of the childs are boys?
c) A Morse code consists of a sequence of dots and dashes. A letter is encoded by a sequence of dots and dashes with length 1 to 5. (Repetitions are allowed). How many letters can be encoded with 1 to 5 symbols?

(See http://www.soton.ac.uk/ scp93ch/refer/alphabet.html for the Morse code alphabet on the web.)

3) (10 points) Topic: **Probability distributions** Let X and Y be independent random variables with Poisson distribution $\lambda = 1$ and $\mu = 5$.

a) Find $P[X \geq Y]$.
b) Find $P[X = Y]$.

4) (10 points) Topic: **Expectation, Variance etc.** .

a) Compute the probability generating function ϕ_X of a $\mathbb{N}$-valued random variable X with probability density function $f_X(n) = 2^{-n}$, $n = 1, 2, \ldots$. (there should be no sum in the end).
b) Use a) to compute $E[X]$.
c) Use a) to compute $Var[X]$.
d) Assume X and Y are independent. We know $E[X] = 2, E[Y] = 3, Var[X] = 4, Var[Y] = 2$. Find $Var[7X - 2Y]$.
e) Let X,Y be as in d). Find $\rho[X + Y, Y] = Corr[X + Y, Y]$.

5) (10 points) Topic: **Chebychev-Markov inequality**. Prove Cantelli's inequality

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

Have a relaxing spring break!

Math464/564 O. Knill, **March 1997**

Homework 9

Due: Tuesday, April 1, 1997 in class

Topic: Absolutely continuous random variables.

Remark. This HW contains 5 questions. Each question can give 10 points, so that you can get a maximal credit of 50 points for this HW.

1) (10 points) Topic: **Discrete and continuous random variables.**
Decide from each of the following random variables on $[0,1]$ whether it is discrete or absolutely continuous:

a) $X(\omega) = \sin(\omega)$.
b) $X(\omega) = \pi/5$.
c) $X(\omega) = [1000 \cdot \omega]$, where $[x]$ is the largest integer smaller or equal to x (for example $[3.1415926...] = 3$).
d) $X(\omega) = 1_{\omega=0.5}$, where $1_A(\omega) = 1$ if $\omega \in A$ and $1_A(\omega) = 0$ if $\omega \notin A$.
e) $X(\omega) = 1_{[0.2,0.3]}(\omega)$.

2) (10 points) Topic: **Normal distribution**. Consider the normal distribution with probability density

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

a) Show that $f(x)$ is maximal for $x = m$.
b) For $x = m \pm \sigma$, the second derivative of f is vanishing.

3) (10 points) Topic: **The Erlang distribution**. Consider a random variable X with probability density function

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

where λ, k are parameters. This is called the Erlang distribution. It arises naturally because as we will see later, the sum of independent exponential distributed random variables is Erlang distributed.

a) Show that the exponential distribution is a special case.
b) Verify that f is indeed a density functions, that means, verify that $\int_0^\infty f(t)\, dt = 1$.
c) Compute $E[X]$.
d) Compute $Var[X]$.

4) (10 points) Topic: **The 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?

5) (10 points) Topic: **General probability spaces.**
It might surprise that there exist open dense sets on the interval $[0,1]$ which have not full probabiliy. Enumate the rational numbers $x_1, x_2, \dots$ and define $A = \bigcup_n \{ x \mid |x - x_n| < 4^{-n} \}$. The set A is dense because it contains the rational numbers. The set A is open because it is a union of open intervals. What is the measure of A?

Math464/564 O. Knill, **April 1997**

Homework 10

Due: Tuesday, April 8, 1997 in class

Topic: Transformation of absolutely continuous random variables.

Remark. This HW contains 4 questions. Two questions give 10 points, the other two 15 points. You can get a maximal credit of 50 points for this HW.

1) (15 points) Topic: **The Arc-Sin distribution**. Background: 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 that a particle has its energy in an interval $[a, b]$ is $\int_a^b f(x)\, dx$. The probability distribution function $F(t) = \int_{-\infty}^t f(s)\, ds$ is called the integrated density of states and is accessible to measurements in solid state physics.

   Consider a random variable X with probability density function f_X which vanishes outside $(-2, 2)$ and which is $f_X(x) = 1/(\pi\sqrt{4 - x^2})$ for $x \in (-2, 2)$.

   a) (5 points) Verify that f is a probability density function and compute the distribution function $F_X(t) = \int_{-\infty}^t f(s)\, ds$. Hint: remember the name of the distribution.
   b) (5 points) Compute $E[X]$, the average energy of a particle.
   c) (5 points) Compute $Var[X]$. (Hint: your symbolic computing software tells you that $\int x^2/\sqrt{4 - x^2}\, dx = 2\arcsin(x/2) - x\sqrt{4 - x^2}/2$.)

2) (10 points) Topic: **The Log-Normal distribution**. It appears often in applied statistics that logarithms of observed datas are normal distributed. In this case, one deals with log-Normal distributed random variables.

   Assume X is normal distributed with mean m and variance σ^2. Compute the density of the random variable $Y = e^X$. This is called the **log-Normal** density.

3) (10 points) Topic: **Exponential distributed random variable**. The time X with which an isotop decays is an exponential distribued random variable with parameter $\lambda = 10$. You look up λ in a table where it is given in the case, when a time unit is 1 second. For some reason, you want to know λ, when the time unit is 1 year = 31'536'000 seconds. In other words, determine the density of the random variable cX for $c = 1/31'536'000$.

4) (15 points) Topic: **Bertrand paradox revisited**. Remember the Bertrand paradox? There Ω was the set of all possible ways with which one can put a line intersecting the unit disc $\mathbb{D} = \{x^2 + y^2 \leq 1\}$. We considered the random variable $X(\omega)$="length of the segment cut by a random line in Ω". The problem was to determine $P[X \leq \sqrt{3}]$. The probability measure P on Ω and so the distribution of X was not determined and lead to different answers.

   a) (5 points) In the first case, $\omega \in \Omega$ was represented by the distance of the line to the origin, assuming $P[\{\omega \in [0, 1] \mid \omega \in [a, b]\}\}] = b - a$. Determine the probability density function f_X of the random variable $X : [0, 1] \to \mathbb{R}$ in this case and compute $E[X]$.
   b) (5 points) In the second case, $\omega\Omega$ was represented by the angle it hits the circle assuming $P[\{\omega \in [0, \pi]\, \omega \in [a, b]\}] = b - a$. Determine the proability density function f_X random variable $X : [0, \pi] \to \mathbb{R}$ in this case and compute $E[X]$.
   c) (5 points) In the third case, $\omega \in \Omega$ was represented by that point in the disc, which is the center of the segment cut by the line and P was the normalized area measure on the disc $\mathbb{D}$. Determine the probability density function of the random variable $X : \mathbb{D} \to \mathbb{R}\}$ and compute $E[X]$.

Math464/564 O. Knill, **April 1997**

Homework 11

Due: Tuesday, April 15, 1997 in class

Topic: Random vectors, joint distributions, marginal density, distributions of a sum of two independent random variables, transformation of densities, conditional density, Bayes rule.

Remark. This HW contains 5 questions each giving 10 points so that you can get a maximal credit of 50 points for this HW.

1) (10 points) Topic: Marginal density. Let X, Y be absolutely continous with density $f(x, y) = \lambda^2 e^{-\lambda y}$ for $0 \leq x \leq y$ and $f(x, y) = 0$ else.

   a) Sketch the level curves of f.
   b) Find the marginal density of X.
   c) Find the marginal density of Y.

2) (10 points) Topic: Covariance matrix, normal distribution. Assume X and Y are $N(0, 1)$-distributed random variables satisfying $Cov[X, Y] = 2$. Find the density of the random vector $(X + 2Y, X + Y)$.

   Hint: you can use the fact that a Gaussian random vector $X = (X_1, \dots, X_n)$ with covariance matrix $K_{ij} = Cov(X_i, Y_j)$ and vector mean $m = (m_1, \dots, m_n)$ with $m_i = E[X_i]$ has the density

$$f(x) = (2\pi)^{-n/2} |\det K|^{-1/2} e^{-\frac{1}{2}(x-m)^T K^{-1}(x-m)} .$$

   Compute first the 2×2 covariance matrix K of $(X + 2Y, X + Y)$.

3) (10 points) Topic: Distribution of a sum of independent random variables. Let X, Y be independent random variables both with continuous distribution f. What is the density of the random variable $Z = X - Y$.

4) (10 points) Topic: Transformation of joint densities. Let (X, Y) be a random vector with density $f(x, y) = 6e^{-2x-3y}$. What is the joint density of the random vector $(3X - Y, X)$.

5) (10 points) Topic: Conditional density, Bayes rule.
   Suppose you know that $f_{Y|X}(x|y)$ is $N(0, 4)$-distributed (normal distributed with mean $m = 0$ and variance $\sigma^2 = 4$) and that X is $N(0, 1)$-distributed. Compute $f_{X|Y}(x|y)$.

Math464/564 O. Knill, **April 1997**

Homework 12

Due: Tuesday, April 22, 1997 in class

Topic: Convergence of random variables. The central limit theorem.

Remark. This HW contains a question with 20 points and three questions giving 10 points so that you can get a maximal credit of 50 points for this HW.

1) (20 points) Topic: Convergence of random variables. Recall that X_n converges in distribution to X if and only if $P[X_n \leq t] \to P[X \leq t]$ for all t. Find in each of the following cases a random variable X to which X_n converges in distribution.

a) (4 points) X_n is Normal distributed with parameters $N(1/(1+n^2), 1+1/n)$.
b) (4 points) X_n is Binomial distributed with parameter $B(n, 1/n)$.
c) (4 points) $X_n = (S_n - np)/\sqrt{np(1-p)}$, where S_n is a sum of n independent $B(n,p)$-distributed random variables.
d) (4 points) X_n has a exponential distribution with parameter $\lambda = n$.
e) (4 points) X_n is uniformly distributed on the interval $[-n, n]$.

2) (10 points) Topic: Convergence of random variables. Assume X_n is uniformly distributed on $\{0, 1, \ldots, n\}$ satisfying $P[X_n = k/n] = 1/(n+1)$. Prove that X_n converges in distribution to the uniform distribution on $[0,1]$.

3) (10 points) Topic: Convergence of random variables. Assume X_n is a sequence of pairwise uncorrelated random variables with mean $E[X_n] = 0$ and $Var[X_n] \leq 5$.

a) Prove that S_n/n^α converges to 0 in distribution for all $\alpha > 1/2$.

b) Under the additional condition that $Var[X_n] \geq 1$, prove that S_n/n^α does not converge in distribution to a random variable with finite variance if $\alpha < 1/2$.

4) (10 points) Topic: Central limit theorem. Let X_n be a sequence of independent random variables and let $\Phi(t)$ is the distribution function of a standard normal distributed random variable. The central limit theorem implies that $P[S_n \leq t]$ is for large n close to $\Phi(\frac{x-E[S_n]}{\sigma(S_n)})$.

Denote by X_n the number of strokes, a golf player has to do in order to drive the ball into one hole and assume X_n are independent. Assume the course has 18 teeing areas and the performance of a golf player called Tiger Woods is $E[X_n] = 3$ and $Var[X_n] = 2$. What is the probability that Woods needs more than 50 strokes in order to finish a course having 18 holes.

Hint. If the function Φ is not on your computer, there is a table on page 252 in the book.

Math464/564 O. Knill, **April 1997**

Homework 13

Due: Tuesday, April 29, 1997 in class

Topic: Characteristic functions, moment generating functions.

Remark. This HW contains 5 questions with 10 points giving a total of 50 points.

1) (10 points) Topic: **Characteristic functions.**
Reminder: The Gamma distribution is 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. The density of a $\Gamma(\lambda, \beta)$-distributed random variable is

$$f(x) = \frac{\lambda^\beta x^{(\beta-1)}}{\Gamma(\beta)} e^{-\lambda x} \ , x \geq 0$$

where $\lambda > 0, \beta > 0$ are fixed parameters.

a) (4) What is the characteristic function $\phi_X(t) = E[e^{itX}]$ of a $\Gamma(\lambda, \beta)$ distributed random variable?
b) (3) Determine, using a), $E[X]$.
c) (3) Determine, using a), $Var[X]$.

2) (10 points) Topic: **Characteristic functions**
Find the characteristic function of a sum $S_n = X_1 + \ldots + X_n$ of n independent λ-exponentially distributed random variables X_i. Compare this function with the characteristic function of the (k, λ)-Erlang distribution which has the density

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

and for which the characteristic function has been determined in the previous problem already.

3) (10 points) Topic: **Characteristic functions**
In class, we have computed the characteristic function of a discrete Poisson distributed random variable

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

Find $E[X^4]$ using this characteristic function.

4) (10 points) Topic: **Characteristic functions**
A random variable is called **symmetric** if X and $-X$ have the same distribution.

a) (5 points) Prove that X is symmetric if and only if the characteristic function ϕ_X is real.
b) (5 points) Assume X and Y are IID random variables. Show that $\phi_{X-Y}(t) = |\phi_X|^2$ and conclude that $X - Y$ is symmetric.
c) (*) What would it mean that ϕ_X is purely imaginary. Is it possible?

5) (10 points) Topic: **Moment generating functions** Let X be a random variable such that $M_X(t)$ is finite for all t. Prove that

$$P[X \geq x] \leq e^{-tx} M_X(t)$$

for $t \geq 0$, where $M_X(t)$ is the moment generating function of X. Hint: Scan backwards for a theorem which appeared in the theory.

Math464/564 O. Knill, **May 1997**

Homework 14

Due: Tuesday, May 6, 1997 in class

Topic: Random walks, gambling

Remark. This HW contains 5 questions with 10 points giving a total of 50 points.

1) (10 points) Topic: **Polya's theorem** .
You have seen that the reason why the random walk in two or less dimensions returns and not in higher dimensions is that

$$\int_{x \in \mathbb{R}^d \mid |x| \leq r} |x|^{-2} \, dx$$

is finite if and only if $d \geq 3$. Verify this fact.

2) (10 points) Topic: **Can you beat the system with insider information?** .
We have proven that for a gambling system V and a fair random walk S_n, one can not beat the system. That is, the expectation for the winning $E[S_N^V] = 0$. It was assumed that V_k was previsible that is it depends only on $X_1, \dots X_{k-1}$. Assume you drop this information and you allow V_k to depend also on X_k (you have some insider information). Design a concrete gambling system, which provides a win.

3) (10 points) Topic: **Gambling system for random walk with drift.**
Assume you have a random walk S_n with drift, that is X_k are IID random variables with expectation $E[X_k] = p$. Assume you have a gambling system V_k. Give a formula for the expected win $E[S_n^V]$ after time n.

4) (10 points) Topic: **Breaking the bank.**
Assume you play in a fair casino which is a random walk with $E[X_k] = 0$. Assume the bank has $a = 1000$ dollars available and you have $b = 2000$ dollars available. If $S_n = X_1 + \dots X_n = -a$, then you broke the bank, if $S_n = b$, then you lost all your money and you are broke. What is the probability that you break the bank?

5) (10 points) Topic: **The variance of the winning.**
You play the game as in the previous question but this time, the casino is no more fair; $E[X] = const$. The bank knows empirically $E[S_T]$, the average win in the game, where the bank has initially a dollars and a player has b dollars and $T = T_{a,b}$ is the ruin time for one of the players. What is the variance of S_T?

Math464/564 O. Knill, **February 1997**

First Midterm

Time: Thursday, February 20, 1997, 14:00-15:15

Topics: Probability spaces (chapter 1), Combinatorics (chapter 2), Discrete random variables (chapter 3).

<u>**Material**</u>: you can use all your notes (as many as you have), your homework, my distributed text. During this midterm, we don't consult the text book nor any other book.

<u>**Points**</u>: There are maximal 100 points. Points for each questions and subquestion are indicated.

<u>**Form**</u>: The answer to question 1 has to be answered on this page. Use seperate pages for the rest of the questions. Please sign these additional pages and indicate the problem number on each of the seperate pages.

NAME: □

1) (20 points) Random (true/false) questions which are not necessarily sorted according to topics. A correct answer gives 2 points. Check the true boxes with a cross.

01) □ In a σ-algebra, the union of an arbitrary number of sets is also in the σ-algebra.

02) □ The set of all subsets of a set is a σ-algebra.

03) □ The probability of a union of two events is always the sum of the probabilies of the single events.

04) □ $P[A\Delta B] = P[A \cup B] - P[A \cap B]$.

05) □ If $A \subset B$, then $P[A] \leq P[B]$.

06) □ Five people can sit in 32 different ways on five chairs.

07) □ Throw a dime 5 times. There are 10 different possibilities to have excatly 2 times head.

08) □ A discrete random variable takes only finitely many values.

09) □ It is always true that $P[A|B] \leq P[A]$ if $P[B] > 0$.

2) (10 points) Consider $\Omega = \{0, 1, 2, \ldots\}$, $\mathcal{A} = \{A \subset \Omega\}$ and $P[\{n\}] = e^{-\beta n}(1 - e^{-\beta})$, where β is a real constant.
a) (5 points) Check the axioms of Kolmogorov to verify that $(\Omega, \mathcal{A}, P)$ is a probability space.
b) (5 points) What is the probability of the event $A = \{2, 4, 6, \ldots\}$ of all even numbers?

("This probability space is important for Planck's law of black body radiation which was crucial for the development of quantum mechanics.")

3) (10 points) Topic: Conditional probability. In the book "uniformity in the world", the german philosopher Marbe expresses the believe that in a repeated toss of a fair coins, a run of heads makes a tail more likely in the next toss. He thinks that this is implied by the "law of averages".
For four tosses of a fair coin, find the conditional probability that the fourth toss comes up is tail, given that the first three tosses give heads. Solve this problem carefully along the following steps:

a) (3 points) Construct the probability space $(\Omega, \mathcal{A}, P)$.
b) (3 points) Find the event B that the first three tosses give head and the event A that the fourth toss gives tail.
c) (4 points) Determine the conditional probability using the definition.

4) (10 points) Topic: Probability spaces, independence of events.
a) (5 points) Prove that two events A and B with $P[B] > 0$ are independent if and only if $P[A|B] = P[A]$.
b) (5 points) You know that $P[A|B] = 0.1$ and $P[B|A] = 0.3$. What is $P[A]/P[B]$?

5) (10 points) Topic: Combinatorics. In the "California super lotto", there is this February a jackpot of 33 Millions. In this game, one chooses 6 numbers out of 51 numbers (The Arizona lotto analyzed in class has a smaller Jackpot however better chances to win the jackpot).

a) (5 points) What is the probability to have 6 right in this game?
b) (5 points) What is the probability to have 2 right?

6) (10 points) Topic: Combinatorics. A standard piano has 88 keys. A common figure for a pianist is to play two accords in sequence where the first accord consists of three tunes (the pianist presses three different keys simultaneously) and the second accord of two tunes (the pianist presses two different keys simultanously). The two accords are played after each other and do not need to be different. A jazz pianist wonders how many such figures can be played. (Actually, the pianist will need both hands and a nose to play some of those accords, but as mathematicians we do not care).

7) (10 points) Topic: Independence. The chances of having rain during a day in the summer in Tucson is 1/100. Assume the events to have rain at different mondays are independent.

a) (5 points) What is the probabiliy to have no rain during 15 mondays in summer?
b) (5 points) What is the probabilty to have no rain during 13 of 15 mondays in summer?

8) (10 points) Topic: Discrete random variables. The desert laboratory (located on the nearby hill with the little telescope and the antennas in Tucson at a 10 minute drive from the university) measures that the number X of Saguaros in one acre is Poisson distributed with parameter $\lambda = 2$. Assume, you want to buy an acre of beautyful land on a nearby hill. What is the probability to become with this buy the owner of maximal 5 Saguaros (that is you will

Math464/564 O. Knill, **March 1997**

Second Midterm

Time: Thursday, March 28, 1997, 14:00-15:15

Topics: Chapter 3-4: Discrete random variables, expectation and variance.

Material: your notes (as much you have), homework, distributed text, no books (to prevent loosing time with reading and encourage good personal summaries of the material)

Points: There are 100 points for 5 questions of 20 points. There is a "super-choice type" question on the fourth page with 20 points. The points achieved in this six'th problem will replace the lowest score of the first five questions. You have the choice to either replace one of the first five questions with the six'th, or (if your time permits) to do all 6 questions and have the minimal number among the 6 subtotals cancelled.

Form: The answer to the first question has to be answered on this page, the answer to the six'th question, (if you make use of it) on the fourth page. You can use the space in the boxes on these pages as well as additional pages for the rest of the questions. Please sign those additional pages and indicate the problem number on each of them. For Problems 2-5, the derivations for the result must be provided.

NAME: □

1) (20 points) Not necessarily sorted (True/False) questions. Each correct answer gives two points.

01) □ The variance of a random variable which is not identically zero is positive.

02) □ Chebychev's inequality allows to give an upper bound on probabilities $P[|X - E[X]| \geq \epsilon]$.

03) □ The probability generating function of a random variable is a polynomial if and only if the random variable takes only finitely many values.

04) □ The covariance of two independent random variables vanishes.

05) □ The sum of two independent Poisson distributed random variables is Poisson distributed.

06) □ If X takes the value 6 with probability 1/3 and the value 3 with probability 2/3, then $E[X] = 4$.

07) □ If $Var[X + Y] = Var[X] + Var[Y]$, then X and Y are uncorrelated.

08) □ There are random variables taking finitely many values for which the expectation is not defined even so the variance is finite.

09) □ Given a sequence of random variables, then $(X_1 + X_2 + \ldots + X_n)/n$ converges.

10) □ $E[(X - E[X])Y] = 0$ if X and Y are independent

2) (20 points) Topic: **The Cramer-Bernoulli solution to the Petersburg paradox by the idea of utility.**

Recall the Petersburg paradox (suggested by Daniel Bernoulli in 1713): you enter the game with an entrance fee of c dollars. Then a sequence of dimes are thrown and you win in that game $X = 2^Y$ dollars if Y subsequent heads occur. For example, for the outcome $\omega = (HHHT\ldots)$, you would win $X(\omega) = 2^{Y(\omega)} = 2^3 = 8$ dollars. The total profit per game is $2^Y - c$ and the game is fair if the expectation of $2^Y - c$ is zero. The paradox was that no matter which entrance fee c you pay, the game will be favorable to you because the expectation of $X = 2^Y$ is infinite. This contradicts the the fact that after having done some games, nobody would agree to pay an entrance fee of say $c = 20$. Gabriel Cramers suggested in 1728 (similarly than James Bernoulli (an oncle of a cousin of Daniel Bernoulli) the following "solution" to the problem. Because for "a person with a lot of money a dollar is less worth), the effective win in a game should be replaced by $E[\sqrt{X}]^2$. (This suggestion makes sense because for example a win of say one gogool (which is more than the number of atoms in the universe) dollars is certainly useless).

Problem: Compute $E[\sqrt{X}]^2$ which is the value of a new fair entrance fee c. (You should get a number).

3) (20 points) Topic: **The Lotka model for population statistics.**

Based on experimental data, Lotka proposed the following model for the number X of children in a random family: $P[X = k] = \beta p q^k$ if $k \neq 0$ and $P[X = k] = 1 - \beta q$ if $k = 0$, where $q = 1 - p$. The case $k = 0$ is teated as a special case because by biological or personal reasons some families are not able or do not want to have children. Based on data from 1930 for American families, one obtains a good fit for $\beta = 9/10$ and $p = 1/4$ (and therefore $q = 3/4$).

Problem. Compute the expected number $E[X]$ of children with these parameters. (The answer which is a number).

4) (20 points) Topic: **A casino buys an assurance.**
In roulette, while betting on one number, a customer wins 70′000 dollars with a probability 1/37 and loses 2′000 dollars with probability 36/37. To soften the loss in the case of a win of the customer, the manager of the casino plans to sign a contract with an assurance company which proposes the following option: if the casino loses to a customer, then assurance pays 40′000 dollars to the casino, each time, the casino wins, it has to pay the assurance company 1′000 dollars.

a) What is the expected win of the casino in one game with assurance. Compare it with the expected win in one game without assurance.
b) What is the expected win of the assurance company in one game?

5) (20 points) Topic: **Chevalier de Méré's problem with mathematics.**
Let X be the number of 1's in 4 rolls of a fair dice and Y the number of double 1's in 24 rolls of two fair dice.

a) Find the probability that X is positive. b) Find the probability that Y is positive.
(Hint to a) and b): compute the probability of the complementary event).

Chevalier de Méré was a high roller who thought that the answers to part a) and b) should be the same. Accordingly, he usually bet on Y being positive and blamed mathematics when he lost money over a long sequence of bets.

What might have led Méré to think wrongly? Maybe the answers to the following question c) and d) explain:

c) What is the expectation of X? d) What is the expectation of Y?

NAME:

6) (20 points) The points in this question will replace the lowest score of the previous 5 questions. Each correct gives 5 points and you can choose 4 of the five following questions (or do all and get the lowest of the scores in a)-e) dropped. Write the answers into the boxes.

a) (5 points)

If X is a $(5, 1/3)$-Bernoulli distributed random variable, find using Chebychev's inequality a number C such that $P[|X - E[X]| \geq 2] \leq C$.

b) (5 points)

What is the probability generating function of $X + Y$, where X and Y are independent and X is $\lambda = 3$ has a geometric distribution and Y is $\lambda = 5$ Poisson distributed.

c) (5 points)

Compute $E[(X+Y)^4]$, where X and Y independent random variables with $E[X^2] = E[Y^2] = 1, E[X^3] = E[Y^3] = 0$ and $E[X^4] = E[Y^4] = 2$.

d) (5 points)

Write down $Var[X]$ for a random variable, for which you know that the probability generating function is $\phi_X(t) = (t^4 + t + 1)/3$.

e) (5 points)

Surely, everybody of you has seen the **Hale-Bopp comet** which is now nicely visible even in a lighted street near campus. There is an understandable concern of humans to calculate the risk that a comet of this size would hit the earth (remember the dino-doom 65 million years ago). One knows of about 100 ring structures which earth which come from meteorites. An example is the Arizona crater which was created 50'000 years ago. One thinks that impact of such a 60 meter meteorite occurs in average every 500-1000 years. One believes that impacts with considerable change of the earth climate can occur several times per million years. Even so this sounds pretty scary, no human in the past 1000 years is known to have been killed by a meteorite or by the effects of an impact. (For more information on comets or meteorites surf the site $http://www.jpl.nasa.gov/$.)

Problem. Assume, the number of years one has to wait until a ~ 60 meter meteorite impacts the earth is a Poisson distributed random variable, with an expected value of 700 years. What is the probability to get a hit in the next 2 years? (Hint: the probability in the answer is smaller than $1/(1gogool)$.)

Math464/564 O. Knill, **March 1997**

Test:

Time: Thuesday, March 11, 1997, 14:00-14:45

> <u>**Material**</u>: This test will be discussed during the second part of class and eventually during part of the next class. This test is on working techniques and will not be part of your grade

1) **Old mistakes** The following questions appeared in the first Midterm of this course and are only slightly changed. Do not look up the answer if you want to profit from this question.

01) □ The union of a countable number of sets in a σ-algebra is in the σ-algebra.

02) □ The set of all subsets of a set is a σ-algebra.

03) □ A discrete random variable takes only finitely many values.

04) □ It is always true that $P[A|B] \leq P[A]$ if $P[B] > 0$.

Having one or more mistakes here indicates that you should learn more from old mistakes.

2) **In class** The following things were mentioned only aside during class.

a) How old is probability theory?
b) What is the Banach-Tarsky paradox?
c) Why is the number 10^{100} called a "gogool"?
d) How does probability theory enter in quantum mechanics?
e) How did Euler define the "gradus suavitatis" for a frequency ratio.
f) Where do random variables with devilish distribution function appear in nature?
g) In which sense can one say that the harmonic series $\sum_{n=1}^{\infty} n^{-1}$ converges?
h) What is the geometric interpretation of the correlation coefficient?
i) Who found Chebychev's inequality before Chebychev?

If you should feel weak in this group of questions, try to do more careful notes during class (even if you know the topic already) or ask more questions during class. Aquire a technique to write fast and review each class for at least a half an hour before the next class.

3) **Intuition** Intuition helps to grasp formal definitions much easier. Having the right "picture" of an object in the mind allows you to figure out the definition without remembering it. That's one reason why computers are not yet good mathematicians.

How do you think about the following objects intuitively?

a) Ω

c) P
d) A random variable X
e) $E[X]$
f) $\sigma(X)$
g) The covariance of two random variables X and Y.
h) The correlation coefficient
i) Independence.
k) Uncorrelated.

Feeling uncomfortable with this question might mean that you can improve your understanding by more thinking about the material or discussing the topics.

4) **Routine.**

a) 20 people are in a new year party. At midnight, they wish each other a happy good year. How many times, does one hear the clink of glasses?
b) Let X be the random variable which takes the value n with probability 2^{-n}. Compute the expectation of X.
c) Let X and Y be independent random variables. Express $E[(X+Y)^3]$ in terms of moments of X and Y.
d) Let X be a random variable which takes the value 1 with probability 1/4 and the value -2 with probability 3/4. Compute its expectation.

If you are slow with answering these questions, you need more routine for solving easier problems. A good series of books to improve routine are "Schaum's outlines", where many problems with solutions are provided. Usually however, routine comes automatically.

5) **Yahoo**. In this question, your look-up performance is addressed.

a) Write down Bayes formula.
b) Find the probability generating function of the geometric distribution.
c) What is the switch on-switch off formula of Sylvester?
d) List the properties of the distribution function F_X?
e) What is the expectation of $X+Y+Z$, where X is Bernoulli(n=5,p=0.3) distributed, Y has geometric distribution with parameter 2 and Z is Poisson distributed with parameter 5?

If you are slow here, you can better organize the course material for example by concentrating the material onto a few pages, putting things into an organized binder.

6) **Big picture.**

a) Describe in three sentences what is "probability theory".
b) Which branches of mathematics do you know which are used in probability theory?
c) Which branches of science do you know which rely on probability theory?
d) Explain in three sentences to a interested layman what you are learning in probability theory.

Looking for the "big picture" is not everything and does not replace a detailed understanding. However, having the subject in the right "folder" helps considerably learning it. So it is a good idea to think from time to time about the big picture.

7) **Solving techniques**. Of how many of the following points are you aware of.

01) □ Can I solve a specific case of the problem (for example, if a problem has to be solved for general n, can I reformulate or solve it for $n = 1, 2, 3$?

02) □ Can I solve a more general problem? (Often, a specific case is psychologically harder than the general case, which might occur in the theory already).

03) □ I can reformulate the problem in my own words. (It might be that the obstacle is that the problem is not properly understood.

04) □ I ask myself actively: did I see a similar problem before? (This process usually runs automatically passively. It is worth giving it some more priority for a second.

05) □ I aks myself actively: to which topic does the problem belong? (Knowing where the problem belongs to is extremely helpful.).

06) □ Did I make at least two attempts to solve the problem? (If the problem can not be solved immediately, a second look after doing something else or sleeping over it can help.)

07) □ Did I sit over the problem long enough? Sometimes, it needs a while (a half an hour maybe) until the right "flash" of idea comes. Sitting over a problem without beeing able to solve it is an extremely important exercice (this actually happens most of the time when doing research).

08) □ Do I need a hint? Can I find the answer in a book or my notes? Can I formulate the question (for asking in an e-mail or in an office hour or in class). Can I guess a hint?

09) □ Are my working conditions good for me? (Light, table, enough sleep, coffeine, distraction by TV or Newspaper or background noise, enough paper and scratch paper for trying out ideas. Finding ideal working conditons need experimentation because because it is individual. It is worth to experiment with it.

10) □ Did I start early to think about the problem set. Starting late produces stress and can make it difficult to find the solution (starting an hour before having to turn in the problem set is too late for most people. A better idea is to solve a problem set early but with a watch. The later can help to improve performance under test conditions.

11) □ Do I think from time to time about improving my learning and solving techniques? For example, having bad physical working conditions can be an obstacle to find a solution. Evenso trivial, most people actually are not aware that techniques exist or that not being able to solve a problem immediately is very natural.

Final word: If you have spotted a point (or several points) which you feel needs improvement, try to work on this during the next weeks as well as in other classes. More stuff and material about problem solving techniques can be found in the library. A good book for mathematicians is

> "Polya, George, 1887-1985. "How to solve it; a new aspect of mathematical method", Princeton University Press, 1971

If you are short of time: most people have actually a lot of free CPU time awailable. One can think about mathematics (and other things) for example while biking or walking, trying to sleep, eating lunch, doing dishes, waiting in a line etc. So, keep always one or two "batch jobs" ready for processing during such "dead times".