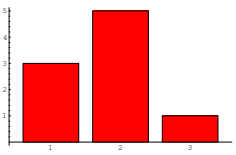
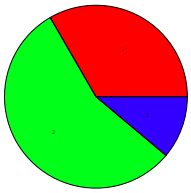
PROBABILITY FOR MULTIVARIABLE CALCULUS Math21a, O. Knill

AIM. We want to use our knowledge in multivariable calculus like vector geometry or integration to cover some basic probability theory and statistics.

PART 0, INTRODUCTION

Some topics in discrete and continuous probability theory use or rely on multivariable calculus. For example, random variables in finite probability spaces are **finite dimensional vectors**. With the dot product $(f,g) = \frac{1}{n}\sum_{i=1}^n f_i g_i$, statistical quantities like expectation, variance or covariance can be reformulated geometrically:



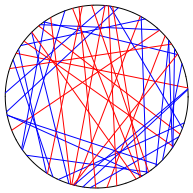


$\mathrm{E}[f] = (1, f)$	expectation	dot product with $1 = (1, ..., 1)$
$\mathrm{Var}[f] = (f - \mathrm{E}[f], f - \mathrm{E}[f])$	variance	square of the length of $f - \mathrm{E}[f]$.
$\mathrm{Cov}[f,g] = (f - \mathrm{E}[f], g - \mathrm{E}[g])$	covariance	dot product of $f - \mathrm{E}[f]$ and $g - \mathrm{E}[f]$.
$\sigma[f] = \sqrt{\mathrm{Var}[f]}$	standard deviation	length of $f - \mathrm{E}[f]$.
$\mathrm{Corr}[f,g] = (f - \mathrm{E}[f], g - \mathrm{E}[g])/(\sigma[f]\sigma[g])$	correlation	$\cos(\alpha)$, angle α

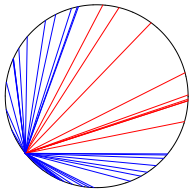
For example, two random variables f and g are **uncorrelated** if and only if $f - \mathrm{E}[f], g - \mathrm{E}[g]$ are **orthogonal**. Multivariable calculus can also be used to select out some probability distributions like for example the Bernoulli distribution on the probability space of all $0-1$ sequences of length n. Multiple integrals in multivariable calculus are used when computing expectations. For random vectors, one has to integrate functions of several variables.

PROBLEM: BERTRAND'S PARADOX (Bertrand 1889)

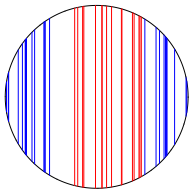
We throw randomly lines onto the unit disc. What is the probability that the intersection with the disc is smaller than the length $\sqrt{3}$ of the equilateral triangle inscribed in the unit circle?



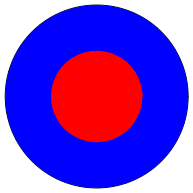
Answer Nr 1: take an arbitrary point P in the disc. The set of lines which pass through that point is parametrized by an angle ϕ. In order that the chord is longer than $\sqrt{3}$, the angle has to fall within an angle of $60°$ of a total of $180°$. The probability is $\boxed{1/3}$.



Answer Nr 2: consider all lines perpendicular to a fixed diameter. The chord is longer than $\sqrt{3}$, when the point of intersection is located on the middle half of the diameter. The probability is $\boxed{1/2}$.



Answer Nr. 3: if the midpoint of the intersection with the disc is located in the disc of radius $1/2$ with area $\pi/4$, then the chord is longer than $\sqrt{3}$. The probability is $\boxed{1/4}$.



REMARK. This example shows that a rigorous foundation probability is needed. Unsharp definitions lead to similar effects in political situations like with **voting systems**, where different voting systems can produce different winners. The following example is by Donald Saari:

"Consider 15 people deciding what beverage to serve at a party. Six prefer milk first, wine second, and beer third; five prefer beer first, wine second, and milk third; and four prefer wine first, beer second, and milk third. In a plurality vote, milk is the clear winner. But if the group decides instead to hold a runoff election between the two top contenders milk and beer, then beer wins, since nine people prefer it over milk. And if the group awards two points to a drink each time a voter ranks it first and one point each time a voter ranks it second, suddenly wine is the winner."

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 problem was discussed by Marilyn vos Savant in a "Parade" column in 1991 and provoked a big controversy. These disputes show that intuitive argumentation can easily lead to confusion.

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

PETERSBURG PARADOXON.

A casino offers you the following game. You enter a fee, 50 dollars the casino flips coins until the first time "tail"=0 appears. If before that n times "head"=1 showed up, you win 2^n dollars.

Example. You pay 50 dollars, the coins show $1, 1, 1, 1, 1, 0$, you win $2^6 - 50 = 64 - 50 = 14$ dollars in that game. In the next game the coins show $1, 1, 0$. You lose 42 dollars.

Question: what entry fee would you consider fair in that game?

Answer: every fixed entry fee would be unfair for the Casino. The probability to win $V_n = 2^n$ dollars is $p_n = 2^{-n}$. The expected win is $\sum_n p_n V_n = \infty$.
The paradoxon is that nobody would be willing to pay 50 dollars. This problem was discovered by the Swiss eighteenth-century mathematician Daniel Bernoulli (1738). Bernoulli "solved" the paradoxon by stating that instead of adding expected payoffs in dollars one should add the expected utilities of each consequence: because money would have a decreasing marginal utility, a realistic measure of the win would be $\log_2(V_n)$. with an expected win $\sum_{n=0}^{\infty} n 2^{-n} = 2$ or $\sqrt{V_n}$ with an expected win of $\sum_{n=0}^{n/2} 2^{-n} 2^{n/2} \sim 3.4$. These are rather philosophical explanations and the problem goes deeper.

ABOUT PROBABILITY THEORY.

Probability emerged in the 17th century as a systematic mathematical study. While during the 18th and 19th centuries the attitudes shifted for more clarity a solid foundation of probability theory was only laid in the 20'th century, when Kolmogorov published "General Theory of Measure and Probability Theory". Published in 1929, it gave the first description of an rigorous construction of probability theory based on integration. In 1933 Kolmogorov expanded the paper into the monograph "Grundbegriffe der Wahrscheinlichkeitsrechnung", which in 1950 was translated into English under the title "Foundations of the Theory of Probability".
Nowadays, probability theory is closely connected to other fields in Mathematics like combinatorics, dynamical systems theory or the theory of partial differential equations. It appears also in seemingly unrelated fields like number theory, logic, game theory or complex analysis. It is crucial in physical areas like

- quantum mechanics (the wave function ψ defines a probability distribution $|\psi|^2$).
- statistical mechanics (for example, percolation problems, spin systems).
- solid state physics (for example, random potentials).
- quantum field theory (quantization as an integration over all possible states).

ABOUT STATISTICS.

While statistics can be considered as a part of probability theory, the approach is is different. The starting point is usually a set of data. An important task is to analyze these data is to find the right mathematical model. Examples of statistical tasks are: **finding correlations** between data, to **give estimates** for the failure rate for equipment, or the **time predictions** for events. A quite recent application are **mail filters**. They use statistical analysis of mail messages to distinguish spam from legitimate mail.

PART I, DISCRETE DISTRIBUTIONS

FINITE PROBABILITY SPACES.

Let $\Omega = \{1, ..., n\}$ be a finite set. The elements of Ω are called **experiments**. The subsets of Ω are called **events**. If we assign weights p_i to each of the experiments in Ω so that $\sum_i p_i = 1$, we have a **probability space**. The **probability** of an event A is defined as $\mathrm{P}[A] = \sum_{i \in A} p_i$, the sum of the probabilities of each experiment in A.

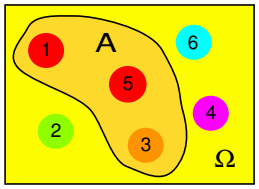
EQUAL PROBABILITY.
In most of the cases, an experiment has the same probability $p_i = 1/n$. This implies that the probability of an event A is the number of elements in A divided by the number of elements in Ω.

$$\mathrm{P}[A] = \frac{|A|}{|\Omega|} = \frac{\text{Number of experiments in A}}{\text{Number of all experiments}}$$

The task to **count** the elements in A often leads to **combinatorial problems**. Below, we will see that the probability distribution $p_i = 1/|\Omega|$ can characterized as the one which maximizes the "entropy" $-\sum_{i=1}^n p_i \log(p_i)$.

EXAMPLES.

1) With $\Omega = \{1, 2, 3, 4, 5, 6\}$ and $p_i = 1/6$, we model a fair dice, thrown once. The event $A = \{1, 3, 5\}$ for example is the event that "the dice produces an odd number". This event has the probability $1/2$.



2) What is the probability that a dice thrown twice shows a sum bigger than 10? To solve this problem, we take the set $\Omega = \{(i, j), i, j \in \{1, 2, 3, 4, 5, 6\}\}$ of all possible 36 experiments we can do. Each experiment has probability $p_{ij} = 1/36$. The event $\{(5, 6), (6, 5), (6, 6)\}$ consists of experiments which lead to a sum bigger than 10. Its probability is $3/36 = 1/12$.

3) The **Girl-Boy problem**:
"Dave has two child. One child is a boy. What is the probability that the other child is a girl"? Most people would say $1/2$.

Solution: $\Omega = \{BG, GB, BB\}$, $\mathrm{P}[\{BG\}] = \mathrm{P}[\{GB\}] = \mathrm{P}[\{BB\}] = 1/3$. Now $A = \{BG, GB\}$ is the event that the other child is a girl. $\mathrm{P}[A] = 2/3$.

SET THEORETICAL NOTATIONS.

$\emptyset$ denotes the **empty set**. A^c is the **complement** of A in Ω. Example: $\Omega = \{1, 2, 3, 4, 5, 6\}, A = \{1, 3, 5\}, A^c = \{2, 4, 6\}$. $A \cap B$ is the **intersection** of A and B. $A \cup B$ is the **union** of A and B. $\mathrm{P}[A|B] = \mathrm{P}[A \cap B]/\mathrm{P}[B]$ is the probability of A **under the condition** B. It is the probability that the event A happens if we know that the event B happens. (B has to have positive probability).

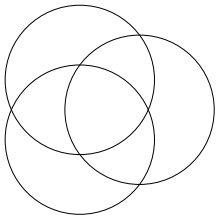
EXAMPLE. Solution of the boy-girl problem with conditional probability. $\Omega = \{BG, GB, BB, GG\}$ with $\mathrm{P}[\{BG\}] = \mathrm{P}[\{GB\}] = \mathrm{P}[\{BB\}] = \mathrm{P}[\{GG\}] = 1/4$. Let $B = \{BG, GB, BB\}$ be the event that there is at least one boy and $A = \{GB, BG, GG\}$ be the event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

PROPERTIES OF P[].

$\mathrm{P}[A^c] = 1 - \mathrm{P}[A]$,
$\mathrm{P}[\Omega] = 1$
$\mathrm{P}[\emptyset] = 0$

$A \subset B \Rightarrow \mathrm{P}[A] \leq \mathrm{P}[B]$
If $A \cap B = \emptyset$, then $\mathrm{P}[A \cup B] = \mathrm{P}[A] + \mathrm{P}[B]$.
$\sum_j P[B \cap A_j] = P[B]$ if $\bigcup_j A_j = B, A_j \cap A_i = \emptyset$.

SWITCH ON, SWITCH OFF FORMULA:
$\mathrm{P}[\bigcup_{i=1}^n A_i] = \sum_{k=1}^n (-1)^{k-1} \sum_{1\le i_1<i_2<\ldots<i_k\le n} \mathrm{P}[A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}]$.
Example:
$\mathrm{P}[A\cup B\cup C] = \mathrm{P}[A]+\mathrm{P}[B]+\mathrm{P}[C]-\mathrm{P}[A\cap B]-\mathrm{P}[A\cap C]-\mathrm{P}[B\cap C]+\mathrm{P}[A\cup B\cup C]$.



BAYES RULE:

If all possible experiments are divided into events A_j and the probabilities that B occurs is known under each of the conditions A_j, then we are able to compute the probability that A_i occurs under the condition that B happens:

$\boxed{\mathrm{P}[A_i|B] = \mathrm{P}[B|A_i]P[A_i]/\sum_j P[B|A_j]P[A_j]}$ if $\bigcup_j A_j = \Omega$ and $A_j \cap A_i = \emptyset$.

EXAMPLE. A fair dice is rolled and then a fair coin is tossed the number of times showing on the dice. Given that all coins are all heads, find the probability that the score of the dice was 5.

SOLUTION. Let B be the event that all coins are heads and let A_j be the event that the dice showed the number j. The problem is to find $\mathrm{P}[A_5|B]$.

We have $\mathrm{P}[B|A_j] = 2^{-j}$. Because the events A_j are disjoint $\mathrm{P}[B] = \sum_{j=1}^6 \mathrm{P}[B \cap A_j] = \sum_{j=1}^6 \mathrm{P}[B|A_j]\mathrm{P}[A_j] = \sum_{j=1}^6 2^{-j}/6 = (1 + 1/2 + 1/3 + 1/4 + 1/5 + 1/6)/6 = 49/120$. By Bayes rule, $\mathrm{P}[A_5|B] = \mathrm{P}[B|A_5]P[A_5]/(\sum_j \mathrm{P}[B|A_j]P[A_j]) = (2^{-5}/6)/(49/120) = 5/392$ which is about 1 percent.

NATURAL PROBABILITY DISTRIBUTION.

Consider a general probability distribution $p = (p_1, ..., p_n)$ on $\Omega = \{1, 2, 3, 4, 5, 6\}$. For example $p = (1/7, 1/7, 1/7, 1/7, 1/7, 2/7)$, models an **unfair dice** which has been rigged in such a way that 6 appears with larger probability. Define the **entropy** of the probability distribution as $H(p) = -\sum_{i=1}^6 p_i \log(p_i)$. For which $p = (p_1, ...p_n)$ is the entropy maximal? We have to extremize $H(p)$ under the constraint $G(p) = \sum_i p_i = 1$. This can be solved using Lagrange multipliers and leads to the solution $p_i = 1/6$.

INDEPENDENCE. Two events A, B are called **independent**, if $\mathrm{P}[A \cap B] = \mathrm{P}[A] \cdot \mathrm{P}[B]$.

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

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

where $\prod i = 1^n a_i = a_1 a_2 ... a_n$ is the product of numbers.

PROPERTIES:

- $A, B \in \mathcal{A}$ are independent, if and only if either $\mathrm{P}[B] = 0$ or $\mathrm{P}[A|B] = \mathrm{P}[A]$. "Knowing B does not influence the probability of the event A".
- Every event A is independent to the empty event $B = \emptyset$.

RANDOM VARIABLE.

A real valued function X on Ω is called a **random variable**. We can look at it as a **vector** $f = (f(1), f(2), ..., f(n))$. In multivariable calculus, we mostly encountered vectors in dimensions 2 or 3. In probability theory it is natural to look at vectors with arbitrary many coordinates.

EXPECTATION. The **expectation** of a random variable f is defined as $\mathrm{E}[f] = \sum_i p_i f(i)$. It is also called the **mean** or the **average value** of f.

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

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

PROOF OF THE ABOVE PROPERTIES:

$$\mathrm{E}[X+Y] = \sum_i p_i (X+Y)(i) \sum_i p_i (X(i)+Y(i)) = \mathrm{E}[X] + \mathrm{E}[Y]$$
$$\mathrm{E}[\lambda X] = \sum_i p_i (\lambda X)(i) = \lambda \sum_i p_i X(i) = \lambda \mathrm{E}[X]$$
$$X \leq Y \Rightarrow X(i) \leq Y(i), \text{ and } \mathrm{E}[X] \leq \mathrm{E}[Y]$$
$$\mathrm{E}[X^2] = 0 \Leftrightarrow X^2(i) = 0 \Leftrightarrow X = 0$$
$$X(\omega) = c \text{ is constant } \Rightarrow \mathrm{E}[X] = c \cdot \mathrm{P}[X = c] = c \cdot 1 = c$$
$$\mathrm{E}[X - \mathrm{E}[X]] = \mathrm{E}[X] - \mathrm{E}[\mathrm{E}[X]] = \mathrm{E}[X] - \mathrm{E}[X] = 0$$

VARIANCE, STANDARD DEVIATION

Variance

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

Standard deviation

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

Covariance

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

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

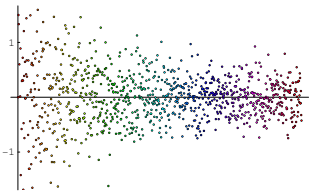
$$\mathrm{Corr}[X,Y] = \frac{\mathrm{Cov}[X,Y]}{\sigma[X]\sigma[Y]} \ .$$

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

STATISTICS. A set of data points $\{x_1, x_2, x_3, x_4, x_5\} = \{2.3, 1.4, 2.5, 0.6\}$ can be interpreted as a random variable X on $\{1, 2, \ldots, 5\}$, where $X(i) = x_i$. By default we assume a uniform distribution. The expectation of X is the average $\frac{1}{5}\sum_{i=1}^5 x_i = 6.8/5 = 1.7$. The variance is $\mathrm{Var}[X] = \frac{1}{5}\sum_{i=1}^5 (x_i - 1.7)^2 = 0.575$, the standard deviation $\sigma[X] = \sqrt{0.575} = 0.758...$.
Given a second set of data points $\{y_1, y_2, y_3, y_4, y_5\} = \{2.3, 1.7, 2.3, 0.7\}$, we can compute the covariance $\mathrm{Cov}[X,Y] = \frac{1}{5}\sum_{i=1}^5 (x_i - 1.7)(y_i - 1.75) = 0.485$.

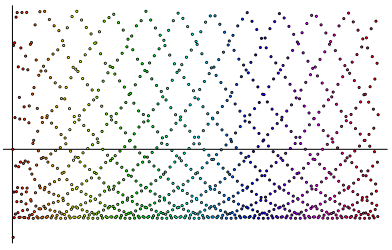
REMARK. In statistics, one usually takes $s^2 = \frac{n}{(n-1)}\sigma^2$ as estimate for the standard deviation of n data x_i.

EXAMPLE: DIGITS OF PI. The digits x_i of π can be used to produce random variables. For example, for any n, we can look at the random variables $X = (x_1, x_2, ..., x_n)$ and $Y = (x_{n+1}, y_{n+2}, ..., y_{2n})$ and look at their correlation. For $n = 6$ for example, we had the random variables $X = (3,1,4,1,5,9), Y = (2,6,5,3,5,8)$. The values of the correlation depending on n is plotted to the right.

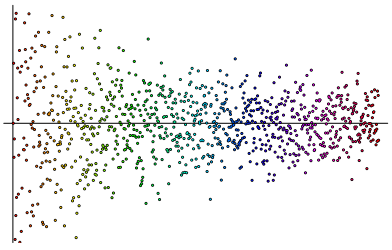


EXAMPLE: PSEUDO RANDOM NUMBER GENERATORS.

1) We produce random variables with the rule $x_{i+1} = x_i + \log(2) \bmod 1$, where $r \bmod 1$ is r minus the integer part of r (for example $1.323 \bmod 1 = 0.323$. The sequence x_i is called a **Sturmean sequence**. The picture to the right shows the correlation between the random variables $X = (x_1, x_2, ..., x_n)$ and $Y = (x_{n+1}, y_{n+2}, ..., y_{2n})$ in dependence of n.



2) We produce random variables with the rule $x_{i+1} = 4x_i(1 - x_i)$. The sequence x_i is called a **logistic sequence**. The picture to the right shows the correlation between the random variables $X = (x_1, x_2, ..., x_n)$ and $Y = (x_{n+1}, y_{n+2}, ..., y_{2n})$ in dependence of n. Unlike before, the correlation goes to zero. This is a sign of **"chaos"**. The recursion could indeed be used as a **pseudo random number generator**.



PROPERTIES of VAR, COV, and CORR:

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

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

FLIPPING COINS

If we flip coins n times, and $\{\text{head}, \text{tail}\}$ is encoded as $\{0,1\}$, we have the probability space $\Omega = \{0,1\}^n$ which contains 2^n experiments. For example, if $n=3$, then $\Omega = \{(0,0,0),(0,0,1),(0,1,0),(0,1,1),(1,0,0),(1,0,1),(1,1,0),(1,1,1)\}$.


If $Q[\{1\}] = p, Q[\{0\}] = q = 1-p$ is the probability distribution on $\{0,1\}$, then then $P[\{(\omega_1,\omega_2,\ldots,\omega_n)\}] = Q[\{\omega_1\}]\cdots Q[\{\omega_n\}]$ is the probability distribution on $\{0,1\}^6$.

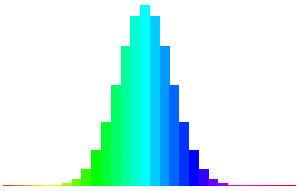
EXAMPLE. If $p = 1/3$ is the probability that "head" happens in one flips a coin, then the probability that $(1,1,0)$ happens is $pp(1-p) = p^2(1-p) = 4/27$.

BERNOULLI DISTRIBUTION. If $\omega_i \in \{0,1\}$ as in the last example, we are especially interested in the random variable $X(\omega) = \sum_{i=1}^n \omega_i$ which gives the number of times, that $\omega_i = 1$ appeared. Under the same assumption that 1 appears with probability p and 0 with probability $q = 1-p$ we have

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

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

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



COMBINATORICS I

Permutations To sort n elements, fix the position of the first, and sort the rest: $p(n) = np(n-1)$:

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

Example: There are $5! = 32$ ways to redistribute 5 coats of 5 people to these people.

Permutations with identification. Sort but identify groups of $n_1, ...n_k$ elements:

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

Example: With the letters A, A, B, B, B, B, E, U it is possible to form $p(8; 2, 4, 1, 1)$ different words.

COMBINATORICS II

Sampling. Choose k elements from n elements and distinguish the picking order.

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

Example. 6 persons can sit in $v(10,6)$ possible ways on 10 chairs.

Sampling with replacement. Choose k elements from n elements but put the element back each time:

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

Example: There are 6^7 possible ways to throw 7 distinguishable dices.

COMBINATORICS III

Combinatorics Pick k elements from n elements but do not distinguish the picking order. We have to divide $v(n,k)$ by $k!$:

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

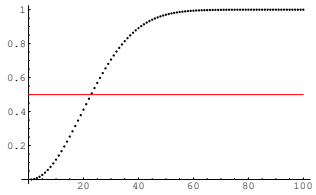
Example: We can choose in $c(52,10)$ possible ways a set of 10 cards from 52 cards.

Combinatorics with repetitions. As before, but we allow to pick several times the same element:

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

Example: We can throw 5 indistinguishable dices in $c^*(6,5)$ different ways.

BIRTHDAY PARADOX: The probability that among k persons nobody has the same birthday is $v(365,k)/365^k$. Therefore, the probability that at least 2 have the birthday at the same day $1-v(365,k)/365^k$. Already for 23 people, the probability of having the same birthday is bigger than $1/2$.



EXAMPLE. Random typing. There are 26^5 possibilities to write a word of 26 letters. In the exercice you will calculate the probability that random typing will produce this notes. To compare that probability, it is good to put the result into some framework of known large numbers (from R.E. Crandall, Scient. Amer., Feb. 1997):

10^4	One "myriad". The largest numbers, the Greeks were considering.
10^5	The largest number considered by the Romans.
10^{10}	The age of our universe in years.
10^{22}	Distance to our neighbor galaxy 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's estimate of number of sand grains in universe.
10^{52}	Mass of our universe in kg.
10^{80}	The number of atoms in our universe.
10^{100}	One "googol". (Name coined by 9 year old nephew of E. Kasner).
10^{153}	Number mentioned in a myth about Buddha.
10^{155}	Size of ninth Fermat number (factored in 1990).
10^{10^6}	Size of large prime number (Mersenne number, Nov 1996).
10^{10^7}	Years, ape needs to write "hound of Baskerville" (random typing).
$10^{10^{33}}$	Inverse is chance that a can of beer tips by quantum fluctuation.
$10^{10^{42}}$	Inverse is probability that a mouse survives on sun for a week.
$10^{10^{51}}$	Inverse is chance to find yourself on Mars (quantum fluctuations)
$10^{10^{100}}$	One "Gogoolplex", Decimal expansion can not exist in universe.

INDEPENDENT RANDOM VARIABLES:

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

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

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

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

PROPERTIES:

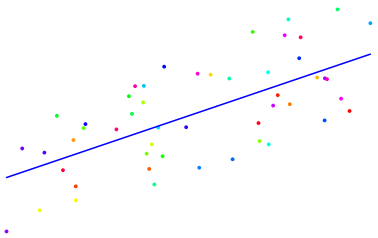
- If X and Y are independent, then $\mathrm{E}[X \cdot Y] = \mathrm{E}[X] \cdot \mathrm{E}[Y]$.
- If X_i is a set of independent random variables, then $\mathrm{E}[\prod_{i=1}^n X_i] = \prod_{i=1}^n \mathrm{E}[X_i]$.
- If X, Y are independent, then $\mathrm{Cov}[X, Y] = 0$.
- A constant random variable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

EXAMPLE. The random variable $X[\{x, y\}] = x$ (value of first dice) and $Y[\{x, y\}] = y$ (value of second dice) on $\Omega = \{1, 2, 3, 4, 5, 6\}$ are independent: $\mathrm{P}[X = a] = 1/6$, $\mathrm{P}[Y = b] = 1/6$, $\mathrm{P}[(X, Y) = (a, b)] = 1/36$.

EXAMPLE. The random variables $X[\{x, y\}] = x + y$ (sum of two dices) and $Y[\{x, y\}] = |x - y|$ (difference of two dices) are not independent: $\mathrm{E}[X \cdot Y] = \mathrm{E}[U^2 - V^2] = \mathrm{E}[U^2] - \mathrm{E}[V^2] = 0$, where $U[\{x, y\}] = x^2, V[\{x, y\}] = y^2$. But $\mathrm{E}[X] > 0$ and $\mathrm{E}[Y] > 0$ shows that $\mathrm{E}[XY] = \mathrm{E}[X]\mathrm{E}[Y]$ is not valid.

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

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



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

Examples: There are two extreme cases:
1) If X, Y are independent, then $a = 0$. It follows that $b = \mathrm{E}[Y]$. We can not guess Y better than replacing it by its mean.

2) If $X = Y$, then $a = 1$ and $b = 0$. The best guess for Y is X.

PROOF. To minimize $\mathrm{Var}[aX + b - Y]$ under the constraint $\mathrm{E}[aX + b - Y] = 0$ is equivalent to find (a, b) which minimizes $f(a, b) = \mathrm{E}[(aX + b - Y)^2]$ under the constraint $g(a, b) = \mathrm{E}[aX + b - Y] = 0$. This **least square** solution can be obtained with Lagrange or by solving $b = E[Y] - aE[X]$ and minimizing $h(a) = \mathrm{E}[(aX - Y - \mathrm{E}[aX - Y])^2] = a^2(\mathrm{E}[X^2] - E[X]^2) - 2a(E[XY] - E[X]E[Y]) = a^2\mathrm{Var}[X] - 2a\mathrm{Cov}[X, Y]$. Setting $h'(a) = 0$ gives $a = \mathrm{Cov}[X, Y]/\mathrm{Var}[X]$.

EXERCISES. (Due until Thursday December 12, 2002)

1) You throw three dice. What is the probability that the sum of the tree dices is smaller or equal than 5?

2) You throw two dice. If an experiment is $\omega = (\omega_1, \omega_2)$, consider the random variables $X = \omega_1$ and $Y = \omega_2$.
a) Find the expectation of X.
b) Find the variance of X.
c) Find the correlation of $X + Y$ and Y.
Hint: Write $\mathrm{Cov}[X + Y, Y]$ as a dot product and use orthogonality.

4) You throw two dice. What is the probability that the sum shows 8 under the condition that the difference is 0? **Hint.** Define two events A,B with which the quest is to compute $\mathrm{P}[A|B]$.

5) The data $(1, 5), (2, 2), (3, 8)$ define two random variables $X = (1, 2, 3), Y = (5, 2, 8)$. Find the variance of X, the expectations of X and Y and the covariance of X, Y. Use this to determine the regression line $y = ax + b$ for these data, where

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