Probability

Measure or determine qualitatively the likelihood that an event or experiment will have a particular outcome.

0. Combinatorial analysis

Ex (basic principle of counting)

In Hungary license plates have 3 letters followed by 3 numbers.
a) how many different ones are there?
b) how many if repetition among letters or numbers is not allowed?

There are 26 possible letters, 10 possible numbers.

a) $26 \cdot 26 \cdot 26 \cdot 10 \cdot 10 \cdot 10 = 26^3 \cdot 10^3 = 17{,}576{,}000$ possibilities

b) $26 \cdot 25 \cdot 24 \cdot 10 \cdot 9 \cdot 8 = 11{,}232{,}000$ possibilities

[If there are k different experiments, each with $n_1, n_2, \dots, n_k$ possible outcomes (independent of each other), then the k experiments together have a total of $n_1 \cdot n_2 \cdot \ldots \cdot n_k$ different possibilities.]

In the Ex, there are 6 "experiments" with $(26, 26, 26, 10, 10, 10)$ possible outcomes in part a) and $(26, 25, 24, 10, 9, 8)$ in b).

0.1 Permutations

How many orders of n objects exist?

Notation $n! = n(n-1)(n-2) \cdot \ldots \cdot 2 \cdot 1$; $\binom{n}{k} = \frac{n!}{(n-k)!\,k!} = \binom{n}{n-k}$

Ex • without repetition (n different/distinguishable objects) $n!$

In a probability class there are 6 boys and 4 girls. In the midterm they all score different points. How many different rankings are there if
a) boys and girls are ranked together?
b) boys are ranked separately from the girls?

If they are ranked together, then doesn't matter if boy or girl
$\Rightarrow$ $10!$ possible rankings
If separately, then $6!$ for boys, $4!$ for girls $\xRightarrow[\text{of counting}]{\text{basic prin.}}$ $(6!)\cdot(4!)$ rankings

- with repetition (n objects, of which $n_1,\dots,n_k$ are indistinguishable) $\frac{n!}{n_1!\cdots n_k!}$

How many different signals, each consisting of 9 flags hung in a line, can be made from a set of 4 white flags, 3 red flags and 2 green flags if all flags of the same color are identical?

If all were different, then $9!$, but $4!$ orders of whites are indistinguishable similarly $3!$ red and $2!$ green $\Rightarrow$ $\frac{9!}{4!\,3!\,2!} = 1260$ different signals

0.2 Variations

How many different orders of k objects chosen out of n objects exist?

Ex
- without repetition (an object is chosen at most once) $\frac{n!}{(n-k)!}$

In an olympic final 8 athletes are competing for the medals. How many different podiums are possible? $\underset{\text{gold}}{8} \cdot \underset{\text{silver}}{7} \cdot \underset{\text{bronze}}{6} = \frac{8!}{(8-3)!}$

- with repetition (an object may be chosen several times) n^k

Recall part a) of first Ex with license plates.

Each letter can be used in all three places $\Rightarrow 26^3$; same for numbers 10^3

0.3 Combinations

How many ways k objects can be chosen out of n objects?

Ex
- without repetition (an object is chosen at most once) $k \le n$ $\binom{n}{k} = \frac{n!}{(n-k)!\,k!}$

From a group of 5 women and 7 men, how many different committees consisting of 2 women and 3 men can be formed? What if 2 of the men refuse to serve on the committee together?

To choose 2 women out of 5 is $\binom{5}{2} = \underbrace{\frac{5!}{(5-2)!}}_{\substack{\text{choose 2 with order}\\\text{(variation without repetition)}}} \cdot \underbrace{\frac{1}{2!}}_{\substack{\text{(permutation}\\\text{with repetition)}}} \Big\} \rightarrow$ order does not matter

Similarly $\binom{7}{3}$ for the men

Thus basic principle of counting $\Rightarrow \binom{5}{2}\binom{7}{3} = 350$ possible committees

In the second case we split the men into 5+2, out of the 2 we can choose only 0 or 1 of them $\Rightarrow \underbrace{\binom{5}{2}}_{\text{no change for women}} \cdot \Big(\underbrace{\binom{2}{0}\cdot\binom{5}{3} + \binom{2}{1}\cdot\binom{5}{2}}_{\text{still have to choose 3 men}} \Big) = 300$

- Out of n antennas m are defective. All defective ones and all functional ones are indistinguishable. How many linear orderings are there in which no two detective are consecutive?

functional ones ($n-m$ altogether)

The defective ones can be in any of the · places. So out of $n-m+1$ places we need to choose m $\Rightarrow \binom{n-m+1}{m}$ possible orderings

- with repetition (an object may be chosen several times) $\binom{k+n-1}{k}$ $k \in \mathbb{N}$

Among 10 students (n in general) we want to distribute 7 awards (k in general) so that a student can receive more than one. How many different ways can the awards be distributed?

↳ First explanation: let's ensure that all students get at least 1 award by distributing $7+10$ ($k+n$) awards.

To distribute the $7+10$ awards into 10 (n) pieces we need to place 9 ($n-1$) separators in the possible $7+10-1$ ($k+n-1$) · slots. This has $\binom{k+n-1}{n-1} = \binom{k+n-1}{k}$ possibilities. After this we simply take 1 award away from everyone and so we distributed the "real" 7 (k) awards amongst the 10 (n) students.

↳ Second explanation: let x_i be the number of awards the i-th student gets. Then $x_1+x_2+\ldots+x_{10}=7$ (in general $x_1+\ldots+x_n=k$). We are looking for the number of solutions of such an equation (x_i are non-negative integers). Such a solution can be represented by a string of dashes _ and bars |, where x_1 dashes is followed by a bar, then x_2 dashes by another bar, and so on, x_{n-1} dashes followed by a bar and finally x_n dashes (no dash is needed at the end). Thus the string has length

$$(x_1+1)+\ldots+(x_{n-1}+1)+x_n = k+n-1, \text{ where we put } k \text{ dashes}$$

$$\Rightarrow \binom{k+n-1}{k} \text{ possibilities}$$

0.4 Binomial / Multinomial coefficients

Binomial coefficients $\binom{n}{k} = \binom{n}{n-k}$, $\binom{0}{0} := 1$, $\binom{n}{0} = 1 = \binom{n}{n}$

Fact $\binom{n+1}{k} = \binom{n}{k-1} + \binom{n}{k}$, generally $\binom{n+m}{k} = \sum_{i=0}^{k} \binom{n}{i}\binom{m}{k-i}$

proof To choose k out of $n+1$, you can either choose $k-1$ from the first n and also choose the last one; or choose all k from the first n and not pick the last one. In general, you can choose i from the first n and $k-i$ from the last m (for all $i=0,1,\ldots,k$) + basic principle of counting. □

Using the first relation, the binomial coefficients can be arranged into triangle, called Pascal's triangle.

$$\binom{0}{0} = 2^0$$
$$\binom{1}{0} + \binom{1}{1} = 2^1$$
$$\binom{2}{0} + \binom{2}{1} + \binom{2}{2} = 2^2$$
$$\binom{3}{0} + \binom{3}{1} + \binom{3}{2} + \binom{3}{3} = 2^3$$

Thm (binomial theorem)

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

, as a consequence the sum of the n-th row in Pascal's Δ is $\sum_{k=0}^{n} \binom{n}{k} = (1+1)^n = 2^n$.

proof

$(x+y)^n = (x+y)(x+y) \cdot \ldots \cdot (x+y) =$ a very long sum of different $x^k y^{n-k}$, $k = 0, 1, \ldots, n$

how many $x^k y^{n-k}$ are there? exactly $\binom{n}{k}$

we just have to add all of these up $\square$

Multinomial coefficients

How many ways can n distinct items be divided into r distinct groups of size $n_1, \ldots, n_r$, respectively, where $\sum_{i=1}^{r} n_i = n$?

$$\binom{n}{n_1} \cdot \binom{n-n_1}{n_2} \cdot \ldots \cdot \binom{n-n_1-\ldots-n_{r-1}}{n_r} = \frac{n!}{n_1!\, n_2! \cdots n_r!} =: \binom{n}{n_1, n_2, \ldots, n_r}$$

(very similar to permutations with repetition, order counts there)

Ex To play basketball, 10 people divide themselves into two groups of 5. How many different ways can they do this?

$\binom{10}{5,5} = \frac{10!}{5!5!}$ ways into team A and team B. But because they are playing against each other, the two teams are indistinguishable, therefore we have to divide by another factor of $2!$ $\Rightarrow$ answer: $\frac{10!}{5!5!2!} = 126$

An analogous multinomial thm is also true:

$$(x_1 + x_2 + \ldots + x_r)^n = \sum_{(n_1, \ldots, n_r):\, n_1 + \ldots + n_r = n} \binom{n}{n_1, n_2, \ldots, n_r} x_1^{n_1} x_2^{n_2} \ldots x_r^{n_r}$$

Ex Balls and urns:

- n distinguishable balls into r distinguishable urns: each ball can be put into r different urns $\xRightarrow[\text{of count.}]{\text{basic princ.}}$ $r \cdot r \cdot \ldots \cdot r = r^n$
- n indistinguishable balls into r distinguishable urns and
  ↳ all urns are non-empty: o.o.oo.oo.oo.oo.o.o n balls, $n-1$ slots for urns

the $(r-1)$st urn determines what is left for the r-th one $\Rightarrow \binom{n-1}{r-1}$

Algebraicly: there are $\binom{n-1}{r-1}$ distinct positive integer-valued vectors $(x_1, \dots, x_r)$ satisfying $x_1 + \dots + x_r = n \quad x_i > 0 \quad i = 1, \dots, r$

↳ empty urns are allowed: this is the same as the second explanation for combinations with repetition $\binom{n+r-1}{r-1}$

1. Axioms of probability

1.1. Sample space and events

Ex · experiment: toss a (fair) dice

Sample space: all possible outcomes of the experiment
in the example $\Omega = \{1,2,3,4,5,6\}$

Event: any subset of the sample space, an outcome of the experiment is an elementary event

$\{\text{toss is prime}\} = \{2,3,5\}$, $\{\text{toss is even}\} = \{2,4,6\}$, $\{\text{toss} \leq 2\} = \{1,2\}$

events will usually be denoted with capital letters ex. $E, F, A, B, \ldots$

$\{\text{sure/certain event}\} = \Omega$; $\{\text{impossible/null event}\} = \emptyset$

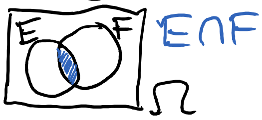
$\mathcal{A} = \{\text{set of all possible events}\} = \{\text{all subsets of } \Omega\}$

- Operations on events

Union / OR / $\cup$ $E \cup F$: All elementary events which are in at least one of E or F.
// $E+F$

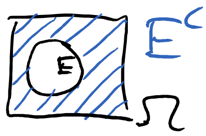
ex $\{\text{toss is prime}\} \cup \{\text{toss is even}\} = \{2,3,4,5,6\}$

Intersection / AND / $\cap$ $E \cap F \,(= EF)$: all elementary events which are in both E and F



ex $\{\text{toss is prime}\} \cap \{\text{toss is even}\} = \{2\}$

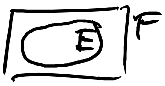
Complement / NOT / ^{c} or $\overline{}$ $E^c = \overline{E}$: all elementary events which are not in E



ex $\overline{\{\text{toss is prime}\}} = \{1,4,6\}$

E and F are mutually exclusive if $E \cap F = \emptyset$

E contains F if $E \subset F$



Good to know:
$$\left.\begin{aligned}(E \cup F) \cap G &= (E \cap G) \cup (F \cap G)\\ (E \cap F) \cup G &= (E \cup G) \cap (F \cup G)\end{aligned}\right\} \text{distributiv law}$$

$$\left.\left(\bigcup_{i=1}^{n} E_i\right)^c = \bigcap_{i=1}^{n} E_i^c \right\} \text{De Morgan's laws}$$

$$\left(\bigcup_{i=1}^{n} E_i\right)^c = \bigcap_{i=1}^{n} E_i^c$$
$$\left(\bigcap_{i=1}^{n} E_i\right)^c = \bigcup_{i=1}^{n} E_i^c$$
De Morgan's laws

1.2. Probability and simple facts

Def Probability $\mathbb{P}: \mathcal{A} \to [0,1]$ is a set-function which assigns a number to all possible events, satisfying the following axioms:

1) $0 \le P(E) \le 1$, for all $E \in \mathcal{A}$
2) $\mathbb{P}(\Omega) = 1$
3) For any mutually exclusive events $E_1, E_2, \dots$: $\mathbb{P}(\bigcup E_i) = \sum_i \mathbb{P}(E_i)$

$\mathbb{P}(E)$ is the probability of the event E. We will learn later the law of large numbers, which guarantees that if we repeat an experiment (ex. toss a dice) many-many times and record how many times the event E occured, say $N(E)$ times in the first N experiments, then the limit $\lim_{N\to\infty} \frac{N(E)}{N}$ always exists, this limit is what we call the probability of E.

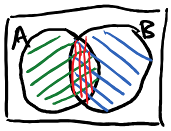
The triplet $(\Omega, \mathcal{A}, \mathbb{P})$ we call a probability space.

Fact
- $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$, since $A \cup A^c = \Omega$ and $A \cap A^c = \emptyset$
- If B contains A ($A \subseteq B$), then $\mathbb{P}(A) \le \mathbb{P}(B)$, (we say that probability is monotonous)

  since $\mathbb{P}(B) = \mathbb{P}(A) + \underbrace{\mathbb{P}(A^c \cap B)}_{\ge 0}$

- inclusion-exclusion formula

  For any events A, B : $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$



  In general $\mathbb{P}\left(\sum_{i=1}^{n} A_i\right) = \sum_{k=1}^{n} (-1)^{k-1} S_k$, where $S_k = \sum_{1 \le i_1 < i_2 < \dots < i_k \le n} \mathbb{P}(A_{i_1} A_{i_2} \dots A_{i_k})$

- A consequence is that $\mathbb{P}(\sum A_i) \le \sum \mathbb{P}(A_i)$.

Ex Remaining at the toss of the dice with events

A: result is a prime number B: result is even C: result ≤ 2

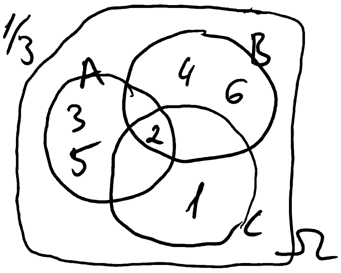
and suppose it is a fair dice, i.e. all elementary events have equal probability. In this case $\mathbb{P}(\{1\}) = \dots = \mathbb{P}(\{6\}) = 1/6$.

Since all elementary events are mutually exclusive, we have

$$\mathbb{P}(A) = \mathbb{P}(\{2,3,5\}) = \mathbb{P}(\{2\}) + \mathbb{P}(\{3\}) + \mathbb{P}(\{5\}) = 3 \cdot 1/6 = 1/2$$

$$\mathbb{P}(B) = \mathbb{P}(\{2,4,6\}) = 1/2 \qquad \mathbb{P}(C) = \mathbb{P}(\{1,2\}) = 1/3$$

$$\mathbb{P}(AB) = \mathbb{P}(\{2\}) = 1/6$$



Venn diagram

Using the inclusion-exclusion formula

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(AB) = \frac{1}{2} + \frac{1}{2} - \frac{1}{6} = \frac{5}{6}$$

which is the same of course as saying

$$\mathbb{P}(A \cup B) = \mathbb{P}(\{2,3,4,5,6\}) = 5 \cdot 1/6 \; ;$$

$$\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(AB) - \mathbb{P}(AC) - \mathbb{P}(BC) + \mathbb{P}(ABC)$$

$$= \frac{1}{2} + \frac{1}{2} + \cancel{\frac{1}{3}} \cancel{- \frac{1}{6}} \cancel{- \frac{1}{6}} \cancel{- \frac{1}{6}} \cancel{+ \frac{1}{6}} = \frac{1}{2} + \frac{1}{2} = 1$$

of course, because $A + B + C = \Omega$ and $\mathbb{P}(\Omega) = 1$.

$$\mathbb{P}(B) = 1 - \mathbb{P}(B^c) = 1 - \mathbb{P}(\text{odd}) \Rightarrow \mathbb{P}(\text{odd}) = 1 - \mathbb{P}(B) = 1/2.$$

1.3 Sample space having equally likely outcomes

This is the simplest type of probability space we can have, where probabilities are calculated with combinatorial analisys.

In general the sample space is a finite set $\Omega = \{1, 2, \ldots, N\}$, each elementary event has the same probability $\mathbb{P}(\{i\}) = 1/N$ for every $i \in \Omega$

Thus the probability of an event A is simply

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{\#(\text{elementary events in } A)}{\#(\text{elements in } \Omega)}$$

Ex Two dice are rolled, what is the probability that their sum $= 7$?

Sample space $\Omega = \{(i,j) : 1 \le i, j \le 6\}$ $\quad |\Omega| = 36$ $\leftarrow$ all possible outcomes

"Good" outcomes: $\{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\}$

$$\mathbb{P}(\text{sum} = 7) = \frac{\#\text{ good}}{\#\text{ all}} = \frac{6}{36} = 1/6.$$

Ex A bowl contains 6 white and 5 black balls. Choose 3 uniformly at random. What is the probability that 1 of them is white, other 2 black?

1st solution: regarding the balls as an ordered set $\Rightarrow$ use permutations & variations

$$|\Omega| = 11 \cdot 10 \cdot 9 = \frac{11!}{(11-3)!} = 990$$

good: the chosen white can be in 3 different positions, once that is fix, there are 6 possibilities and $5 \cdot 4$ for the two black. $\xrightarrow[\text{of counting}]{\text{basic princ.}}$ $|good| = 3 \cdot 6 \cdot (5 \cdot 4) = 360$

Thus $\mathbb{P}(1 \text{ white}, 2 \text{ black}) = \frac{360}{990} = \frac{4}{11}$

<u>2nd</u> solution: regarding the balls as an unordered set $\Rightarrow$ use combinations

$|\Omega| = \binom{11}{3}$ just chose 3 balls from a possible of 11

$|good| = \underbrace{\binom{6}{1}}_{\text{one white}} \cdot \underbrace{\binom{5}{2}}_{\text{two black}}$ $\quad \mathbb{P} = \frac{\binom{6}{1}\binom{5}{2}}{\binom{11}{3}} = \frac{6 \cdot \frac{5 \cdot 4}{2}}{\frac{11 \cdot 10 \cdot 9}{3 \cdot 2}} = \frac{4}{11}$

<u>Remark</u> The moral of the Ex is that it doesn't matter if you consider the outcome of the experiment as an ordered or unordered set in order to calculate the probability of an event. But do <u>NOT</u> mix the two, for example by calculating $|\Omega|$ as ordered and $|good|$ as unordered!! Usually counting it one way is easier than the other.

<u>Ex</u> An urn contains n balls of which one is special. k balls are withdrawn one at a time (allways evenly among the remaining ones). $A :=$ the special ball is chosen $\quad \mathbb{P}(A) = ?$

<u>1st</u> solution: regarding as unordered

simply choosing k from $n \Rightarrow |\Omega| = \binom{n}{k}$

good: special is chosen $\binom{1}{1}$; remaining $(k-1)$ from $(n-1) \Rightarrow \binom{n-1}{k-1}$

So $\quad \mathbb{P} = \frac{\binom{1}{1}\binom{n-1}{k-1}}{\binom{n}{k}} = \frac{k}{n}$

<u>2nd</u> solution: Let A_i be the event that the special ball is the i-th ball to be chosen $i = 1, \dots, k$. A_i and A_j $(i \neq j)$ are mutually exclusive, so $\mathbb{P}(A) = \mathbb{P}\left(\bigcup_{i=1}^{k} A_i\right) = \sum_{i=1}^{k} \mathbb{P}(A_i)$ $(*)$

What is $\mathbb{P}(A_i)$?

Each ball is equally likely to be the i-th ball to be chosen.

Thus $\mathbb{P}(A_i) = \frac{1}{n}$ independent of i.

More formally: we can choose k ordered balls $\frac{n!}{(n-k)!}$ ways, of which $\frac{(n-1)!}{(n-k)!}$ have the special ball as i-th.

Continuing (*) $\mathbb{P}(A) = k \cdot \frac{1}{n}$.

Remark The idea to break up an event into the union of mutually exclusive events can be very useful in situations where it is (much) easier to deal with the mutually exclusive events.

Ex In the game of bridge the entire deck of 52 cards is dealt out to 4 players (each get 13 cards). What is the probability that each player receives an A?

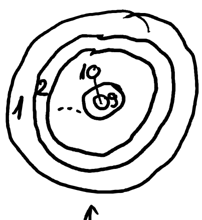
The 52 cards can be dealt in a total of $\binom{52}{13,13,13,13}$ ways.

Taking away the aces, the remaining 48 cards can be dealt $\binom{48}{12,12,12,12}$ ways.

The aces can be divided 4! different ways amongst the players.

$$\Rightarrow \quad \mathbb{P} = \frac{4!\binom{48}{12,12,12,12}}{\binom{52}{13,13,13,13}} \approx 0.105.$$

Ex Assume we are throwing at a target board like this:



regions are evenly placed

What is the probability that we throw a bull's-eye, if we throw the dart into any region of the board with probability proportional to its area?

Sample space Ω = closed unit disk

we have (uncountably) infinite elementary events!

$\mathbb{P}(\text{the dart lands at } (x,y)) = 0$ for every (x,y) $\rightsquigarrow$ delicacy between discrete and continuous

$$\mathbb{P}(\text{bull's eye}) = \frac{\text{area of bull's eye}}{\text{area of total board}} = \frac{0.1^2\pi}{1^2\pi} = 0.01$$

$$\mathbb{P}(\text{throw} \le 8) = \mathbb{P}(1) + \mathbb{P}(2) + \ldots + \mathbb{P}(8) = 1 - (\mathbb{P}(9) + \mathbb{P}(10)) = 1 - \frac{0.2^2\pi}{\pi} = 0.96$$

2. Conditional Probability

Ex We flip a fair coin twice. The sample space is $\Omega = \{(h,h),(h,t),(t,h),(t,t)\}$, all with probability 1/4. What is the probability that both are heads, if
(a) the first is heads? (b) at least one of them is heads?

(How do these informations change the original probability of 1/4)

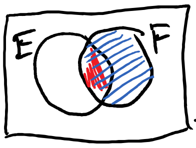
(a) if first is h then the possible outcomes are (h,h) and (h,t) with the same probability, of which only (h,h) is good, so $\mathbb{P} = 1/2$.

(b) at least on of them is h, then it can be $(h,h),(h,t),(t,h)$ and so $\mathbb{P} = 1/3$.
You could say that if at least one is heads, then either one of them or both of them are heads, so $\mathbb{P} = 1/2$. What's wrong? These two events <u>DO NOT</u> have the same probability! Be very careful with the sample space!

Def conditional probability
Given two events E, F such that $\mathbb{P}(F) > 0$, the conditional probability of E conditioned on F is
$$\mathbb{P}(E|F) = \frac{\mathbb{P}(EF)}{\mathbb{P}(F)}$$

Remark: In Ex E was both are heads and F was condition (a) or (b).

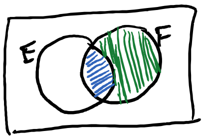
- For E to occur conditioned on F, both E <u>and</u> F have to occur on the reduced sample space of elementary events which are in E.
We are "zooming in" on F and see which points are in the intersection with E.



- If each outcome in (a finite) Ω is equally likely, then conditioned on $F \subset \Omega$, all outcomes in F become equally likely $\rightsquigarrow$ can use F as the sample space.

Fact. multiplication rule
$$\mathbb{P}(EF) = \mathbb{P}(F) \cdot \mathbb{P}(E|F) \quad \text{and in general}$$
$$\mathbb{P}(E_1 E_2 \dots E_n) = \mathbb{P}(E_1)\, \mathbb{P}(E_2|E_1)\, \mathbb{P}(E_3|E_1E_2) \dots \mathbb{P}(E_n|E_1 \dots E_{n-1}) \quad \text{because}$$
$$= \cancel{\mathbb{P}(E_1)} \quad \frac{\cancel{\mathbb{P}(E_2E_1)}}{\cancel{\mathbb{P}(E_1)}} \quad \frac{\cancel{\mathbb{P}(E_3E_2E_1)}}{\cancel{\mathbb{P}(E_2E_1)}} \dots \frac{\mathbb{P}(E_n \dots E_1)}{\cancel{\mathbb{P}(E_{n-1} \dots E_1)}}$$
only this remains

- $\mathbb{P}(E|F) + \mathbb{P}(E^c|F) = 1$, because



$$= \frac{\mathbb{P}(EF) + \mathbb{P}(E^cF)}{\mathbb{P}(F)} \underset{\text{exclusive}}{\overset{\text{mutually}}{=}}$$

$$= \frac{\mathbb{P}(EF) + \mathbb{P}(E^cF)}{\mathbb{P}(F)} \underset{\text{exclusive}}{\overset{\text{mutually}}{=}} \frac{\mathbb{P}(EF \cup E^cF)}{\mathbb{P}(F)} \overset{\text{distributive}}{=} \frac{\mathbb{P}((E \cup E^c) \cap F)}{\mathbb{P}(F)} = \frac{\mathbb{P}(F)}{\mathbb{P}(F)}.$$

Ex (same ex from game of bridge) 52 cards, distribute 13 to each of the four players. What is the probability that each player gets one Ace?

Now we calculate it with the multiplication rule:

Define 4 events: $E_1 = \{$Ace of spade is in any of the hands$\}$

$E_2 = \{$Ace of spade and hearts are in different hands$\}$

$E_3 = \{$Ace of spade, heart and diamond are in different hand$\}$

$E_4 = \{$All four aces are in different hands$\}$

Task is to determine $\mathbb{P}(E_4)$. Using that $E_4 \subset E_3 \subset E_2 \subset E_1$ we have

$$\mathbb{P}(E_4) = \mathbb{P}(E_4E_3E_2E_1) \overset{\text{mult. rule}}{=} \mathbb{P}(E_1)\,\mathbb{P}(E_2|E_1)\,\mathbb{P}(E_3|\underbrace{E_1E_2}_{=E_2})\,\mathbb{P}(E_4|\underbrace{E_1E_2E_3}_{=E_3})$$

$\mathbb{P}(E_1) = 1$ because E_1 is the whole sample space.

$\mathbb{P}(E_2|E_1) = \frac{\mathbb{P}(E_1E_2)}{\mathbb{P}(E_1)} = \mathbb{P}(E_2) = \frac{39}{51}$, because the ace of hearts can be any of the 39 cards in the hands of the other players and there are 51 cards remaining altogether.

$\mathbb{P}(E_3|E_1E_2) = \mathbb{P}(E_3|E_2) = \frac{26}{50}$, because Ace of diamond can not be in the piles in which Ace of spade or heart are $\rightsquigarrow$ 24 wrong places out of the possible 50.

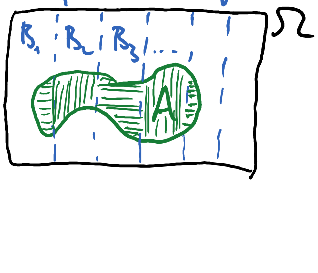
$\mathbb{P}(E_4|E_1E_2E_3) = \mathbb{P}(E_4|E_3) = \frac{13}{49} \quad \Rightarrow \mathbb{P}(E_4) = \frac{39 \cdot 26 \cdot 13}{51 \cdot 50 \cdot 49} \approx 0.105$

Thm (law of total probability)

Assume that $B_1, B_2, \ldots$ form a complete set of mutually exclusive events, i.e.

$B_i \cap B_j = \emptyset$ for every $i \neq j$ and $\sum_i \mathbb{P}(B_i) = 1$.

partition of Ω



Then for any event A

$$\mathbb{P}(A) = \sum_i \underbrace{\mathbb{P}(A|B_i)\,\mathbb{P}(B_i)}_{\mathbb{P}(AB_i)}$$

Ex We have two urns:

1st 2 black 1 white

2nd 2 black 3 white

First we toss a (fair) die to see which urn we choose from a single ball.

If toss ≤ 2, then choose from first; if toss ≥ 3 choose from second.

A: we choose a black ball $\quad \mathbb{P}(A) = ?$

Let B_1: we choose from 1st urn $\quad B_2$: we choose from second urn

then B_1 and B_2 form a complete set of mutually exclusive events with $\mathbb{P}(B_1) = \mathbb{P}(\text{toss} \leq 2) = \frac{1}{3}$ and $\mathbb{P}(B_2) = 2/3$.

We also know that $\mathbb{P}(A|B_1) = 2/3$ and $\mathbb{P}(A|B_2) = 2/5$. Thus the law of total probability implies $\underline{\mathbb{P}(A)} = \mathbb{P}(AB_1) + \mathbb{P}(AB_2) = \mathbb{P}(B_1)\mathbb{P}(A|B_1) + \mathbb{P}(B_2)\mathbb{P}(A|B_2) = \frac{2}{9} + \frac{4}{15} = \underline{\frac{22}{45}}$

Thm (Bayes theorem)

Let $B_1, B_2, \ldots$ be a complete set of mutually exclusive events, A an arbitrary event.

Then $$\mathbb{P}(B_k|A) = \frac{\mathbb{P}(A|B_k)\,\mathbb{P}(B_k)}{\sum_i \mathbb{P}(A|B_i)\,\mathbb{P}(B_i)}$$ for any $k = 1, 2, \ldots$

proof $\mathbb{P}(B_k|A) \overset{\text{def.}}{=} \frac{\mathbb{P}(AB_k)}{\mathbb{P}(A)}$ use multiplication rule in numerator and law of total probability in denominator $\square$

Ex Situation same as in previous example: 2 urns (●●○) & (●●○○○).

We choose a single ball from one of them the same way (toss a die, if ≤ 2 the first, if toss ≥ 3 the second). But now we toss the die in secret and only show the ball we chose. Assuming the chosen ball is white, which urn did we choose it from more likely?

Let A: chosen ball is white $\quad B_1$: choose from 1st urn $\quad B_2$: choose from second

The question is, which is bigger $\mathbb{P}(B_1|A)$ or $\mathbb{P}(B_2|A)$?

First of all $\mathbb{P}(B_1|A) = 1 - \mathbb{P}(B_2|A)$ (B_1 & B_2 complete, mutually exclusive)

Use Bayes thm!
$$\mathbb{P}(B_1|A) = \frac{\mathbb{P}(AB_1)}{\mathbb{P}(A)} = \frac{\mathbb{P}(B_1)\,\mathbb{P}(A|B_1)}{\mathbb{P}(B_1)\mathbb{P}(A|B_1) + \mathbb{P}(B_2)\mathbb{P}(A|B_2)} = \frac{1/3 \cdot 1/3}{1/3 \cdot 1/3 + 2/3 \cdot 3/5} = \frac{5}{23} \overset{0.217\ldots}{}$$

Thus $\mathbb{P}(B_2|A) = \frac{18}{23} > \mathbb{P}(B_1|A)$. So it is (much) more likely that the white ball was chosen from the second urn.

Def Given two events A and B such that $0 < \mathbb{P}(B) < 1$, we say that A and B are *independent*, if $\mathbb{P}(A) = \mathbb{P}(A|B) = \mathbb{P}(A|B^c)$.

If $\mathbb{P}(B) = 0$ or 1 then A and B are considered independent.

Thm. • A and B are independent $\iff$ $\mathbb{P}(AB) = \mathbb{P}(A)\,\mathbb{P}(B)$

proof "$\Rightarrow$" $\mathbb{P}(AB) \overset{\text{mult. rule}}{=} \mathbb{P}(A|B)\,\mathbb{P}(B) \overset{A,B \text{ indep.}}{=} \mathbb{P}(A)\,\mathbb{P}(B)$

"$\Leftarrow$" $\mathbb{P}(A|B) \overset{\text{def.}}{=} \dfrac{\mathbb{P}(AB)}{\mathbb{P}(B)} \overset{\text{cond.}}{=} \dfrac{\mathbb{P}(A)\,\cancel{\mathbb{P}(B)}}{\cancel{\mathbb{P}(B)}} = \mathbb{P}(A)$ $\square$

• If $A \cap B = \emptyset$ and $\mathbb{P}(A) \neq 0$, $\mathbb{P}(B) \neq 0$ $\Rightarrow$ A and B are <u>not</u> independent.

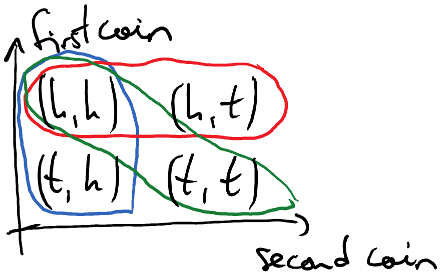
proof $\mathbb{P}(AB) = 0$ but $\mathbb{P}(A)\,\mathbb{P}(B) \neq 0$, so by prev. point $\not\Rightarrow$ $\square$

Def Events $A_1, A_2, A_3, \ldots$ are **completely independent**, if

$$\mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \cdots \mathbb{P}(A_{i_k}) \text{ for any } k\text{-tuple } A_{i_1}, \ldots, A_{i_k} \quad k \geq 2$$

For $k=2$ we say **pairwise independent**.

Ex We flip two fair coins. The sample space is



$A_1 = \{\text{flip with first is heads}\}$

$A_2 = \{\text{flip with second is heads}\}$

$A_3 = \{\text{both are heads or both are tails}\}$

A_1, A_2 and A_3 are pairwise independent, but not completely because

$\mathbb{P}(A_1 A_2) = \mathbb{P}((h,h)) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2)$, similarly

$\mathbb{P}(A_1 A_3) = \mathbb{P}((h,h)) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = \mathbb{P}(A_1) \cdot \mathbb{P}(A_3)$ and $\mathbb{P}(A_2 A_3) = \mathbb{P}(A_2)\,\mathbb{P}(A_3)$

But, since $A_3 = A_1 A_2 \cup A_1^c A_2^c$ A_3 is not independent of $\{A_1, A_2\}$:

$$\mathbb{P}(A_1 A_2 A_3) = \mathbb{P}((h,h)) = \frac{1}{4} \neq \frac{1}{8} = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \mathbb{P}(A_1)\,\mathbb{P}(A_2)\,\mathbb{P}(A_3).$$

3. Random variables

Def A random variable X is a real-valued function defined on the sample space Ω: for $\omega \in \Omega$ $X(\omega) \in \mathbb{R}$. Informally we think of a random variable (abbreviated r.v.) as some function of the possible outcomes of an experiment.

ex toss two dice, let X be the result of the first toss, Y the log of the second, and Z be the sum of the two tosses.

Sample space $\Omega = \{(i,j) : 1 \le i, j \le 6\}$

$$X(i,j) = i \; ; \quad Y(i,j) = \log j \; ; \quad Z(i,j) = i+j$$

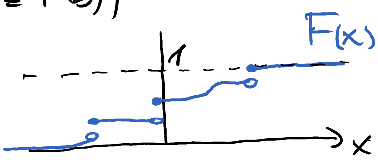
The distribution of X is the collection of probabilities $\mathbb{P}(\{\omega : X(\omega) \in B\})$ for subset B of $\mathbb{R}$. In particular, the function $F(x) := \mathbb{P}(X \le x)$ $-\infty < x < \infty$ is called the cumulative distribution function of X (c.d.f.)

Fact $F : \mathbb{R} \to \mathbb{R}$ c.d.f. is

non decreasing (i.e. if $a \le b$, then $F(a) \le F(b)$)

$\lim_{x \to \infty} F(x) = 1$, $\lim_{x \to -\infty} F(x) = 0$

F is right continuous.



3.1. Discrete random variables

Def A discrete random variable is a r.v. which can take a finite or countably infinite different values $x_1, x_2, x_3, \ldots$

The probability mass function (p.m.f.) of X is defined as

$$p(x_i) = p_i := \mathbb{P}(X = x_i), \text{ where } p_i \ge 0 \text{ and } \sum_i p_i = 1.$$

← also call it the distribution of X

The c.d.f. of a discrete r.v. is simply

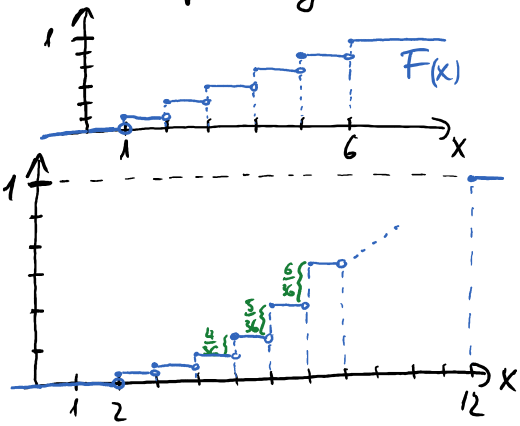
$$F(a) = \sum_{x \le a} p(x) \quad , \text{which is a step function:}$$

it is constant on the interval $[x_i, x_{i+1})$, and the "jumps" at x_{i+1} by p_{i+1}

Ex Remaining at toss of two dice X = result of first toss
Z = sum of the two tosses

The p.m.f. in the two cases:

The corresponding c.d.f.



X

x_i	1	2	3	4	5	6
p_i	1/6	1/6	1/6	1/6	1/6	1/6

Z

z_i	2	3	4	5	6	7	8	9	10	11	12
p_i	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

$$F_Z(a) = \begin{cases} 0, & \text{if } a<2 \\ 1/36, & 2\le a<3 \\ 3/36, & 3\le a<4 \\ 6/36, & 4\le a<5 \\ 10/36, & 5\le a<6 \\ 15/36, & 6\le a<7 \\ \vdots & \vdots \\ 1, & 12\le a \end{cases}$$

$$\mathbb{P}(Z\le 5) = F_Z(5) = \frac{10}{36}$$
$$\mathbb{P}(Z=5) = F_Z(5) - F_Z(4) = \frac{10-6}{36} = \frac{4}{36} (= p_5)$$
$$\mathbb{P}(Z>5) = 1-\mathbb{P}(Z\le 5) = 1-F_Z(5) = 1-\frac{10}{36}$$
$$\mathbb{P}(3<Z\le 5) = F_Z(5)-F_Z(3) = \frac{10-3}{36} = p_4+p_5$$

Def The *expected value* of a r.v. X is the weighted average
$$\mathbb{E}X = \sum_i p_i x_i$$, where X can take on the values x_i with probability p_i
(can be ∞)   (center of gravity of a distribution of mass)

Ex · flip a fair coin: $X=1$ if it is heads and $X=0$ if tails
Then $\mathbb{E}X = \frac{1}{2}\cdot 1 + \frac{1}{2}\cdot 0 = \frac{1}{2}$ (the average of 0 and 1)

flip a biased coin: say $\mathbb{P}(X=1) = 2/3$ and $\mathbb{P}(X=0)=1/3$
Then $\mathbb{E}X = \frac{2}{3}\cdot 1 + \frac{1}{3}\cdot 0 = \frac{2}{3}$ (the weighted average)

· *indicator random variable* X of an event A
$$X = \begin{cases} 1, & \text{if } A \text{ occurs} \\ 0, & \text{if } A^c \text{ occurs}\end{cases}$$ Then $\mathbb{E}X = \mathbb{P}(A)$.

· sum of the tosses of two dice:
$$\mathbb{E}Z = \frac{1}{36}\cdot 2 + \frac{2}{36}\cdot 3 + \ldots + \frac{2}{36}\cdot 11 + \frac{1}{36}\cdot 12 = 14\cdot\frac{(1+2+3+4+5)}{36} + \frac{6}{36}\cdot 7 = \frac{252}{36} = 7$$

Prop · *expectation of a function of a r.v.*
If X is a discrete r.v. and g is a real valued function, then
$$\mathbb{E}(g(X)) = \sum_i p(x_i)\cdot g(x_i)$$

In particular, for $g(x)=x^n$, we call
$$\mathbb{E}(X^n)=\sum_i p(x_i)\cdot x_i^n \text{ the } n\text{-th moment of } X.$$

- expectation is linear For any constants $a,b\in\mathbb{R}$
$$\mathbb{E}(aX+b)=a\,\mathbb{E}(X)+b.$$
- expectation is additive For any two r.v.-s X and Y
$$\mathbb{E}(X+Y)=\mathbb{E}(X)+\mathbb{E}(Y)$$
- If $X\geq 0$, then $\mathbb{E}X=\sum_{i=0}^{\infty}\mathbb{P}(X>i)$, because
$$\sum_{i=0}^{\infty}\mathbb{P}(X>i)=\sum_{i=0}^{\infty}\sum_{k=i+1}^{\infty}\mathbb{P}(X=k)\overset{\text{change}}{\underset{\text{order}}{=}}\sum_{k=1}^{\infty}\sum_{i=0}^{k-1}\mathbb{P}(X=k)=\sum_{k=1}^{\infty}k\,\mathbb{P}(X=k)=\mathbb{E}X.$$

ex sum of the toss of two dice Z can be written as $Z=X+Y$, where X is the result of the first toss and Y the result of the second. Then $\mathbb{E}X=\mathbb{E}Y=\frac{1}{6}\cdot(1+2+3+4+5+6)=3.5$. So $\mathbb{E}Z=2\cdot 3.5=7$.

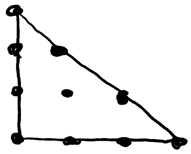
Def The variance of the r.v. X is defined
$$Var(X)=\mathbb{D}^2(X):=\mathbb{E}\left[(X-\mathbb{E}X)^2\right]$$
inertia with respect to the center of gravity of the mass distribution
measures the variation or spread of the values X can take

The standard deviation of X is $\mathbb{D}(X)=\sqrt{Var(X)}$

Prop
- $\mathbb{D}(X)\geq 0$ and $\mathbb{D}(X)=0 \iff X$ is a fix constant with probability 1
- For any constants $a,b\in\mathbb{R}$ $Var(aX+b)=a^2\,Var(X)$, because
$$Var(aX+b)=\mathbb{E}\big((aX+b)-\mathbb{E}(aX+b)\big)^2=\mathbb{E}(aX+\cancel{b}-a\mathbb{E}X-\cancel{b})^2=$$
$$=\mathbb{E}\big(a(X-\mathbb{E}X)\big)^2=a^2\,\mathbb{E}(X-\mathbb{E}X)^2=a^2\,Var(X)$$
- $Var(X)=\mathbb{E}X^2-(\mathbb{E}X)^2$, because using properties of $\mathbb{E}$
$$\mathbb{E}(X-\mathbb{E}X)^2=\mathbb{E}\big(X^2+\underbrace{(\mathbb{E}X)^2}_{\text{number}}-2X\,\mathbb{E}X\big)=\mathbb{E}X^2+(\mathbb{E}X)^2-2\,\mathbb{E}(X\,\overbrace{\mathbb{E}X}^{\text{number}})$$
$$=\mathbb{E}X^2+(\mathbb{E}X)^2-2(\mathbb{E}X)^2=\mathbb{E}X^2-(\mathbb{E}X)^2$$
- Steiner's theorem $\mathbb{E}(X-c)^2=\mathbb{E}(X-\mathbb{E}X)^2+(\mathbb{E}X-c)^2\geq Var(X)$ and minimal $\iff c=\mathbb{E}X$.

Def Random variables $X_1, X_2, \dots$ are independent, if for any subsets $B_1, B_2, \dots \subset \mathbb{R}$ the events $\{X_1 \in B_1\}, \{X_2 \in B_2\}, \dots$ are independent.

ex Consider the points $\{(i,j): i+j \le 3,\ i \ge 0,\ j \ge 0,\ i,j \text{ integers}\}$



Choose one of the 10 points uniformly at random.

$X :=$ first coordinate of the point

$Y :=$ second coordinate of the point

Are X and Y independent?

Let $B_1 = \{\text{first coordinate} = 3\}$, $B_2 = \{\text{second coordinate} = 3\}$

The $\mathbb{P}(X \in B_1) = \frac{1}{10} = \mathbb{P}(Y \in B_2)$

However, $\mathbb{P}(\{X \in B_1\} \cap \{Y \in B_2\}) = 0$

$\left.\right\} \frac{1}{100} \ne 0 \implies$ X and Y are not independent.

Prop If X and Y are independent r.v.-s, then

$$\mathbb{E}(XY) = \mathbb{E}(X)\,\mathbb{E}(Y) \quad \text{and} \quad \mathbb{D}^2(X+Y) = \mathbb{D}^2(X) + \mathbb{D}^2(Y)$$

3.2 Notable discrete r.v.-s

A| Bernoulli and binomial r.v.-s

Def A random variable X has Bernoulli distribution $\left(\begin{matrix} X \overset{d}{=} \mathrm{Ber}(p) \\ X \sim \mathrm{Ber}(p) \end{matrix}\right)$ if its p.m.f. is $\mathbb{P}(X=1) = p$ and $\mathbb{P}(X=0) = 1-p$ p is the parameter

X has Binomial distribution with parameters n and p ($X \sim \mathrm{Bin}(n,p)$) if its p.m.f. is $p_i = \mathbb{P}(X=i) = \binom{n}{i} p^i (1-p)^{n-i}$ $i = 0, 1, \dots, n$ and 0 otherwise

Remark Interpretation: Assume we conduct an experiment, wich is successful with probability p. The $\{X=1\}$ is the event that the experiment was a success. If we conduct n of these identical experiments independently of each other, the the number of successful experiments will have a $\mathrm{Bin}(n,p)$ distribution. In other words:

if $Y \sim \mathrm{Bin}(n,p)$, then $Y \overset{d}{=} X_1 + X_2 + \dots + X_n$, where

X_i are independent and identically distributed (i.i.d.) $\sim \mathrm{Ber}(p)$.

Ex · Choosing with replacement

In an urn there N balls, K of which are black. We choose n balls with replacement (i.e. after choosing one we put it back into the urn). What is the probability that we choose i-black ones?

At each stage, choosing a black ball has probability $p = \frac{K}{N}$.
If X_j is the indicator of the event that the j-th ball is black
then $Y := \#$ black balls chosen $= X_1 + \dots + X_n \sim Bin(n, \frac{K}{N})$

thus $\mathbb{P}(Y=i) = \underbrace{\binom{n}{i}}_{\text{choose which } i \text{ trials are a success}} \underbrace{\left(\frac{K}{N}\right)^i}_{\text{for those } i \text{ choose a black ball}} \underbrace{\left(1-\frac{K}{N}\right)^{n-i}}_{\text{for the rest of the trials choose a non-black ball}}$.

· A company produces screws. Each screw is defective with probability $p = 0.01$. The screws are sold in packs of 10. For any pack with at least two defective screws, the company gives a full refund. In what proportion of sold packages does the company have to refund?

Let $Y = \#$ defective screws in a given pack. A refund is given if $Y \geq 2$.
The distribution of Y is $Bin(10, 0.01)$. So

$$\mathbb{P}(Y \geq 2) = 1 - \mathbb{P}(Y=0) - \mathbb{P}(Y=1) = 1 - 0.99^{10} - 10 \times 0.01 \times 0.99^9 \approx 0.004.$$

So in about 1 out of every 250 cases they give a refund.

Prop · If $X \sim Ber(p)$, then

$$\left.\begin{aligned} \mathbb{E}X &= p = 1 \cdot p + 0 \cdot (1-p) \\ \mathbb{E}X^2 &= p = 1^2 \cdot p + 0^2 (1-p) \end{aligned}\right\} \Rightarrow \mathbb{D}^2(X) = p(1-p) = p - p^2 = \mathbb{E}X^2 - (\mathbb{E}X)^2$$

If $Y \sim Bin(n,p)$, then writing $X_1, \dots, X_n$ i.i.d. $\sim Ber(p)$

$\mathbb{E}Y = np$, because $\mathbb{E}(Y) = \mathbb{E}(X_1 + \dots + X_n) = n \cdot \mathbb{E}(X) = np$

$\mathbb{D}^2 Y = np(1-p) = \mathbb{D}^2(X_1 + \dots + X_n) \underset{\text{indep.}}{=} \mathbb{D}^2(X_1) + \dots + \mathbb{D}^2(X_n) = n \cdot \mathbb{D}^2(X)$.

and so $\mathbb{D}Y = \sqrt{np(1-p)}$.

· The p.m.f. of a $Bin(n,p)$ r.v. grows until it reaches its maximum at $k_0 = \lfloor (n+1)p \rfloor$ and then decreases

· For very large n and very small p s.t. $np =: \lambda$

Stirling formula

- For very large n and very small p s.t. $np = \lambda$

$$\binom{n}{k} p^k (1-p)^{n-k} \approx e^{-\lambda} \frac{\lambda^k}{k!}$$

Stirling formula

$$k! \sim k^{k+1/2} e^{-k} \sqrt{2\pi}$$

More precisely, $\lim_{\substack{n\to\infty,\ p\to 0 \\ np\to\lambda}} \binom{n}{k} p^k (1-p)^{n-k} = e^{-\lambda} \frac{\lambda^k}{k!}$

Calculating the binomial coefficients is very impractical for large n, small p. But this result says that a $\mathrm{Bin}(n,p)$ r.v. can be well approximated by another r.v. with p.m.f. given by $e^{-\lambda} \frac{\lambda^k}{k!}$.

- $\mathbb{P}(Y=k+1) = \frac{p}{1-p} \frac{n-k}{k+1} \mathbb{P}(Y=k)$

B) Poisson random variable

Def X has Poisson distribution with parameter λ ($X \stackrel{d}{=} \mathrm{Poi}(\lambda)$) if its p.m.f. is

$$p_k = \mathbb{P}(X=k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad \text{for } k = 0,1,2,\dots$$

Fact This is a valid distribution, because $p_k \geq 0$ for every k and

$$\sum_{k=0}^{\infty} p_k = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} \cdot e^{\lambda} = 1.$$

Ex · Previous example with defective screws. $n=10$ $p=0.01$

$\to$ on average the number of defective screws in a pack is $\lambda = 0.1$.

We approximate $\mathbb{P}(Y \geq 2)$ by

$$1 - e^{-\lambda} \frac{\lambda^0}{0!} - e^{-\lambda} \frac{\lambda}{1!} = 1 - e^{-0.1} - e^{-0.1} \cdot 0.1 = 0.00467\dots$$

- The number of typographical errors on a page (on a page there are many letters $\to$ large n, but each letter is mistyped with a very small probability p) on average (this is np) is $\lambda = 1/2$. What is the probability that there's at least one error on this page?

  Let $X = \#$errors on this page, then $X \sim \mathrm{Poi}(1/2)$.

  Question is $\mathbb{P}(X \geq 1) = 1 - \mathbb{P}(X=0) = 1 - e^{-1/2} \approx 0.393$.

- The number of people in a large community, who live to be 100 years old.

Prop · $\mathbb{E}X = \lambda$, because $\mathbb{E}X = \sum_{k=0}^{\infty} k \cdot e^{-\lambda} \frac{\lambda^k}{k!} = \sum_{k=1}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!} = \cancel{e^{-\lambda}} \cdot \lambda \underbrace{\sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!}}_{=e^{\lambda}} = \lambda$

Prop. · $\mathbb{E}X=\lambda$, because $\mathbb{E}X=\sum_{k=0} k\cdot e^{-\lambda}\frac{\lambda^k}{k!}=\sum_{k=1} k e^{-\lambda}\frac{\lambda^k}{k!}=\cancel{e^{-\lambda}}\cdot\lambda\underbrace{\sum_{k=1}\frac{\lambda^{k-1}}{(k-1)!}}_{=\cancel{e^{\lambda}}}=\lambda$

$\mathbb{E}X^2=\lambda(\lambda+1)$ (do as exercise!)

So $\mathbb{D}^2X=\lambda(\lambda+1)-\lambda^2=\lambda$

· $\mathbb{P}(X=i+1)=\frac{\lambda}{i+1}\mathbb{P}(X=i)$

· If λ is not an integer, then $\mathbb{P}(X=k)$ is maximal for $k=\lfloor\lambda\rfloor$, otherwise if λ is an integer, then maximal for both $k=\lambda$ and $\lambda-1$.

· If X and Y are two independent Poi r.v.-s with parameters λ and η then $X+Y\sim\text{Poi}(\lambda+\eta)$

Remark This last property gives rise to another type of application: number of "events" that occur at certain points in time

$N(t)$ be the number of these events in $[0,t)$, then $\mathbb{P}(N(t)=1)\approx\lambda t$ + small error — Poisson point process

Ex · The number of people that enter a given establishment (ex. post office) in one day

· The number of earthquakes during some fixed time span:

Assume that in the western USA there are on average two earthquakes per year (a) $\mathbb{P}$(there will be at least 3 earthquakes in the next two years) = ?

(b) $\mathbb{P}$(the next earthquake will happen before time t) = ?

X = # earthquakes the next year $\sim\text{Poi}(2)$, each year is indep. of each other

(a) Y = # earthquakes in the next two years $=X_1+X_2\overset{d}{=}\text{Poi}(2)+\text{Poi}(2)=\text{Poi}(4)$

$$\mathbb{P}(Y\geq3)=1-\big(\mathbb{P}(Y=0)+\mathbb{P}(Y=1)+\mathbb{P}(Y=2)\big)=1-e^{-4}\left(1+4+\frac{4^2}{2}\right)=1-13e^{-4}\approx0.76$$

(b) Z = time until next earthquake

$\mathbb{P}(Z>t)=\mathbb{P}(\text{no earthquakes in }[0,t])=e^{-\lambda t}$

$F(t)=\mathbb{P}(Z\leq t)=1-e^{-2t}$ (this will be called the exponential distribution)

· Assume that on average a fisherman does not catch anything 6 out of 100 times he goes fishing. How many fish does the fisherman catch most often?

Let X = # fish the fisherman catches during one fishing $\quad X\sim\text{Poi}(\lambda)$ ← unknown

What we do know is that

$\frac{6}{100}=\mathbb{P}(X=0)=e^{-\lambda}$, from which $\lambda=\ln\frac{100}{6}\approx2.8$

$\frac{6}{100} = \mathbb{P}(X=0) = e^{-\lambda}$, from which $\lambda = \ln \frac{100}{6} \approx 2.8$

thus most often the fisherman catches $\lfloor 2.8 \rfloor = 2$ fish.

C) Uniform distribution on a finite set

Def Given a finite set $\{1, 2, \dots, N\}$, X is uniformly distributed on $\{1, \dots, N\}$, if the p.m.f. is $\mathbb{P}(X=i) = 1/N$ for every $i = 1, \dots, N$

We have already seen many examples previously (fair die, coin ...)

Prop $\mathbb{E}X = \frac{N+1}{2}$ $\quad \mathbb{D}^2 X = \frac{N^2-1}{12}$

D) Geometric distribution (optimistic)

Def X has geometric distribution with parameter p ($X \stackrel{d}{=} \mathrm{Geo}(p)$), if the p.m.f is $\quad \mathbb{P}(X=i) = (1-p)^{i-1} \cdot p, \quad i = 1, 2, 3, \dots$

Interpretation: we repeat an experiment until the first time it is successful, then X counts on which trial we are first successful.

Remark It is called optimistic, because we are waiting until the first success. There is also a pessimistic version, where we count the number of failures until the first success. Then the p.m.f. is $\mathbb{P}(X=i) = (1-p)^i \cdot p, \; i = 0, 1, 2, \dots$

Prop · $\mathbb{E}X = 1/p$, $\mathbb{E}X^2 = \frac{2}{p^2} - \frac{1}{p}$, $\mathbb{D}^2 X = \frac{1-p}{p^2}$ (exercise to calculate)

· $\mathbb{P}(X > n) = \sum_{i=n+1}^{\infty} (1-p)^{i-1} \cdot p = (1-p)^n \, p \sum_{i=0}^{\infty} (1-p)^i = (1-p)^n \cdot \frac{\cancel{p}}{\cancel{1-(1-p)}}$

· memoryless property $\mathbb{P}(X > n+m \mid X > n) = \mathbb{P}(X > m)$, because

$$\mathbb{P}(X > n+m \mid X > n) = \frac{\mathbb{P}(X > n+m)}{\mathbb{P}(X > n)} = \frac{(1-p)^{n+m}}{(1-p)^n} = (1-p)^m = \mathbb{P}(X > m)$$

If we know that there was no success in the first n-trials, then the probability that it will not be in the next m trials neither is the same as counting the trials from 1 and no success in the first m trials.

The geometric distr. is the only discrete r.v. with this property.

Ex  Choose a ball u.a.r., write down its color and put it back.

Ex 

M black balls
N white balls

Choose a ball u.a.r., write down its color and put it back.
$X :=$ number of draws until first black ball is drawn.
$\mathbb{E}X = ?$

probability of success $p = \frac{M}{N+M}$, so $\mathbb{E}X = \frac{1}{p} = \frac{N+M}{M} = 1 + \frac{N}{M}$

if $N = M$, then $\mathbb{E}X = 2$; $N = kM \Rightarrow \mathbb{E}X = 1+k$

what about $Y :=$ number of draws until third black ball is drawn.

let Y_1 be the number of draws until first black, then $Y_1 \stackrel{d}{=} X$

Y_2 be the n of draws after first, until second $\quad Y_2 \stackrel{d}{=} X$

Y_3 — " — second until third $\quad Y_3 \stackrel{d}{=} X$

So $Y = Y_1 + Y_2 + Y_3$ is the sum of three i.i.d. $Geo\left(\frac{M}{N+M}\right)$ r.v.-s.

Def Y has negative binomial distribution with parameters r and p, if $Y = Y_1 + \ldots + Y_r$, where Y_i are i.i.d. $Geo(p)$ r.v.-s.

Prop. A negative binomial r.v. has p.m.f.

$$\mathbb{P}(Y = n) = \underbrace{\binom{n-1}{r-1}}_{\text{out of the first } n-1 \text{ trials choose which } r-1 \text{ are success}} \underbrace{p^r}_{\text{total } r \text{ success}} \underbrace{(1-p)^{n-r}}_{\text{rest are failures}} \qquad n = r, r+1, \ldots$$

- $\mathbb{E}Y \underset{\text{additive}}{=} r \cdot \frac{1}{p}$ and $Var(Y) \underset{\text{independence}}{=} r \cdot \frac{1-p}{p^2}$

E| Hypergeometric distribution

Def X has hypergeometric distribution with parameters n, N and m, if the p.m.f. is $\mathbb{P}(X = i) = \frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}}$ $\quad i = 0, 1, \ldots, n$

Interpretation: an urn contains N balls of which m are distinguished. We choose n balls without replacement, then X is the number of distinguished balls chosen.

3.3. Continuous random variables

Continuous random variables can take uncountably many different values.
For example, the lifetime of a light-bulb, or the amount of time we wait for the bus at the busstop.

Def We say that X is a continuous r.v. if there exists $f:\mathbb{R}\to\mathbb{R}$ non-negative function having the property that for any subset $B\subset\mathbb{R}$

$$\mathbb{P}(X\in B)=\int_B f(x)\,dx.$$

f is called the probability density function (p.d.f.) of X.

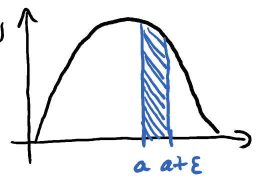
The distribution function of X is

$$F(a):=\mathbb{P}(X<a)=\mathbb{P}(X\le a)=\int_{-\infty}^{a} f(x)\,dx.$$

Prop · For every p.d.f. $\int_{-\infty}^{\infty} f(x)\,dx=1.$

· For $[a,b]$: $\mathbb{P}(X\in[a,b])=F(b)-F(a)=\int_a^b f(x)\,dx.$

interpretation of p.d.f.: $\frac{d}{da}F(a)=f(a)$

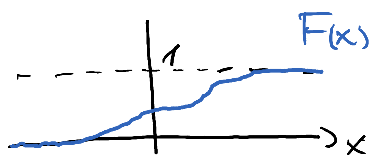


$\mathbb{P}(X\in[a,a+\varepsilon])=$
$=\int_a^{a+\varepsilon} f(x)\,dx\approx\varepsilon\cdot f(a)$

but for $b=a$ $\mathbb{P}(X=a)=\int_a^a f(x)\,dx=0.$ ($f(a)$ is a measure of how likely it is that X will be near 'a')

· $\mathbb{P}(X\ge a)=\mathbb{P}(X>a)=1-\mathbb{P}(X\le a)=1-F(a).$

· In the continuous case the distr. func $F(x)$ is continuous, non-decreasing and $\lim_{x\to\infty}F(x)=1$, $\lim_{x\to-\infty}F(x)=0$



Ex How should we choose C, so that the following function is a p.d.f.?

$$f(x):=\begin{cases}C(4x-2x^2), & 0<x<2\\ 0, & \text{otherwise}\end{cases}$$

We need $\int_{-\infty}^{\infty} f(x)\,dx=1$

$$\int_0^2 C(4x-2x^2)\,dx=C\cdot\left(4\left[\frac{x^2}{2}\right]_0^2-2\left[\frac{x^3}{3}\right]_0^2\right)=1\Rightarrow C=\frac{3}{8}$$

What is $\mathbb{P}(X>1)$? $\mathbb{P}(X>1)=\int^2 f(x)\,dx=\int^2\frac{3}{8}(4x-2x^2)\,dx=\ldots=\frac{1}{2}$

What is $\mathbb{P}(X>1)$? $\mathbb{P}(X>1) = \int_1^2 f(x)\,dx = \int_1^2 \frac{3}{8}(4x-2x^2)\,dx = \dots = \frac{1}{2}$.

Def The expectation of a cont. r.v. X is $\mathbb{E}X = \int x f(x)\,dx$,

variance is $\operatorname{Var} X = \mathbb{D}^2 X = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \int x^2 f(x)\,dx - \left(\int x f(x)\,dx\right)^2$

Remark

discrete r.v.	cont. r.v.
probability mass func.	prob. distribution func.
sum	integral
distr. func. with jumps	cont. distr. func.
finite/countable values	uncountably many values

Prop All properties of expectation for discrete r.v.-s are also true for cont. r.v.-s: linearity, additive prop., expectation of a func. of a r.v.

$$\mathbb{E}(g(X)) = \int g(x) f(x)\,dx$$

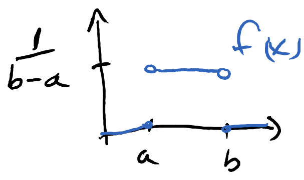
for a non-negative cont. r.v. $\mathbb{E}X = \int_0^\infty \mathbb{P}(X>x)\,dx$

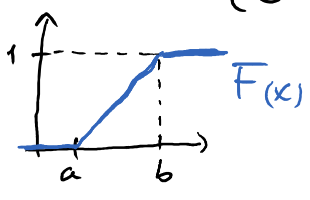
3.3.1. Some notable continuous r.v.-s

A) Uniform distribution

Def X is a uniform r.v. on the interval (a,b), if its p.d.f. & distr. func. are

$$f(x) = \begin{cases} \frac{1}{b-a}, & a<x<b \\ 0, & \text{otherwise} \end{cases} \qquad F(x) = \begin{cases} 1, & x\ge b \\ \frac{x-a}{b-a}, & a<x<b \\ 0, & x\le a \end{cases} \qquad X \stackrel{d}{=} \mathrm{Uni}(a,b)$$





$\mathbb{P}(X\in(\alpha,\beta)) = \beta-\alpha$

Fact $X \sim \mathrm{Uni}(a,b)$, then $\mathbb{E}X = \frac{a+b}{2}$, because

$$\mathbb{E}X = \int_a^b \frac{1}{b-a}\cdot x\,dx = \frac{1}{b-a}\,\frac{b^2-a^2}{2} = \frac{a+b}{2}$$

Similar easy calculation gives $\operatorname{Var} X = \frac{(b-a)^2}{12}$

Ex. Buses depart from the bus stop every 15 minutes starting at 7 a.m. A passenger arrives at the bus stop somewhere between 7 and 7:30 a.m. according to uniform distribution. What is the probability that the passenger (a) waits at most 5 minutes for a bus

according to uniform distribution. What is the probability that the passenger
(a) waits at most 5 minutes for a bus,
(b) waits at least 10 minutes for a bus? $X :=$ time that the passenger arrives

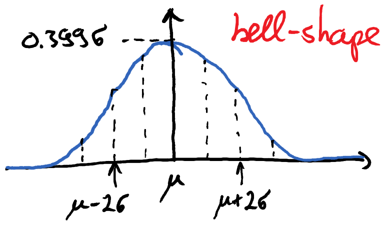
the passenger waits ≤ 5 minutes $\iff$ arrives between 7:10–7:15 or 7:25–7:30
≥ 10 $\iff$ 7:00–7:05 or 7:15–7:20

$$\mathbb{P}(\le 5 \text{ minute wait}) = \mathbb{P}(X\in 7{:}10\text{-}15) + \mathbb{P}(X \in 7{:}25\text{-}30) = \frac{5}{30}+\frac{5}{30} = \frac{1}{3}$$
Similarly $\mathbb{P}(\ge 10 \text{ minute wait}) = \frac{1}{3}$.

- recall the example with the dart board from Subsec. 1.3

B1 Normal distribution (also commonly called Gaussian distr.)

Def X is a normal r.v. with parameters μ & σ^2 ($X\sim N(\mu,\sigma^2)$) if its p.d.f. is
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad x\in\mathbb{R}$$



symmetric around μ,
decays very rapidly towards zero

Fact · $\int_{-\infty}^{\infty} f(x)\,dx = 1$. With change of variable $y := \frac{x-\mu}{\sigma}$ it is enough to show that $I := \int_{-\infty}^{\infty} e^{-y^2/2}\,dy = \sqrt{2\pi}$, this can be done by calculating I^2 and in the 2-dim integral change to polar coordinates...

- If $X\sim N(\mu,\sigma^2)$, then $Y = aX+b \sim N(a\mu+b, a^2\sigma^2)$, because
$$F_Y(x) = \mathbb{P}(Y\le x) = \mathbb{P}(aX+b\le x) = \mathbb{P}\left(X\le \frac{x-b}{a}\right) = F_X\left(\frac{x-b}{a}\right)$$
differentiating we get
$$f_Y(x) = \frac{d}{dx}F_Y(x) = \frac{d}{dx}F_X\left(\frac{x-b}{a}\right) = \frac{1}{a}f_X\left(\frac{x-b}{a}\right) = \frac{1}{\sqrt{2\pi a^2\sigma^2}}\,e^{-\frac{(x-a\mu-b)^2}{2a^2\sigma^2}}$$

- standardization of $X\sim N(\mu,\sigma^2)$ is the r.v. $Y := \frac{X-\mu}{\sigma}$

then $Y\sim N\left(\frac{\mu}{\sigma}-\frac{\mu}{\sigma}, \frac{1}{\sigma^2}\cdot\sigma^2\right)$, azaz $Y\sim N(0,1)$.

The special case $\mu=0$, $\sigma=1$ we call a standard normal r.v.

- expectation of a standard normal r.v.:
$$\mathbb{E}Y = \int_{-\infty}^{\infty} y\frac{1}{\sqrt{2\pi}}e^{-y^2/2}\,dy = \frac{1}{\sqrt{2\pi}}\left[-e^{-y^2/2}\right]_{-\infty}^{\infty} = 0$$

So in general, if $X = \sigma Y+\mu$, then
$\mathbb{E}X = \mu+\sigma\mathbb{E}Y = \mu$, i.e. the parameter $\mu = \mathbb{E}X$

So in general, if $X = \sigma Y + \mu$, then

$$\mathbb{E}\,X = \mu + \sigma\,\mathbb{E}\,Y = \mu, \text{ i.e. the parameter } \underline{\mu = \mathbb{E}\,X}$$

- variance of a standard normal r.v.: (need an extra integration by parts)

$$\mathrm{Var}\,Y = 1 \quad \text{and thus} \quad \mathrm{Var}\,X = \mathrm{Var}(\sigma Y + \mu) = \sigma^2$$

$$\text{i.e. the parameter } \underline{\sigma^2 = \mathrm{Var}\,X}.$$

Def The distribution function of $Y \sim N(0,1)$ is

$$\Phi(y) = \mathbb{P}(Y \le y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-x^2/2}\,dx.$$

Φ does not have a closed form. We look up the values from a table.

Important relationship: $\Phi(-x) = 1 - \Phi(x) \quad x \in \mathbb{R}$

To determine the values of the distr. func. of a non-standard normal r.v. we use the standardization technique: $X \sim N(\mu, \sigma^2)$

$$F_X(a) = \mathbb{P}(X \le a) = \mathbb{P}\left(\frac{X-\mu}{\sigma} \le \frac{a-\mu}{\sigma}\right) = \mathbb{P}\left(Y \le \frac{a-\mu}{\sigma}\right) = \Phi\left(\frac{a-\mu}{\sigma}\right).$$

Ex Let $X \sim N(3,9)$. Calculate (a) $\mathbb{P}(2 < X < 5)$ (b) $\mathbb{P}(|X-3| > 6)$

(a) $$\mathbb{P}(2 < X < 5) = \mathbb{P}\left(\frac{2-3}{3} < \frac{X-3}{3} < \frac{5-3}{3}\right) = \Phi\left(\frac{5-3}{3}\right) - \Phi\left(\frac{2-3}{3}\right)$$

$$= \Phi\left(\tfrac{2}{3}\right) - \Phi\left(\tfrac{-1}{3}\right) = \Phi\left(\tfrac{2}{3}\right) - \left(1 - \Phi\left(\tfrac{1}{3}\right)\right) \approx 0.3779$$

(b) $$\mathbb{P}(|X-\mu| > 2\sigma) = \mathbb{P}(X-\mu > 2\sigma) + \mathbb{P}(X-\mu < -2\sigma) = \mathbb{P}\left(\frac{X-\mu}{\sigma} > 2\right) + \mathbb{P}\left(\frac{X-\mu}{\sigma} < -2\right)$$

$$= (1 - \Phi(2)) + \Phi(-2) = 2\cdot(1 - \Phi(2)) \approx 0.0456$$

Similarly $\mathbb{P}(|X-\mu| > k\sigma) = 2\cdot(1-\Phi(k))$, for $k=1 \approx 0.3174$
$k=3 \approx 0.0026$

C1 Exponential distribution

Def X is exponentially distributed with parameter $\lambda > 0$ ($X \stackrel{d}{=} \mathrm{Exp}(\lambda)$) if its p.d.f. is

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \ge 0, \\ 0 & x < 0. \end{cases}$$

and the c.d.f. after integration is

$$F(x) = 1 - e^{-\lambda x}, \quad x \ge 0.$$

Prop $\cdot$ $\mathbb{E}\,X^n = \int_0^\infty \underbrace{x^n}_{g} \underbrace{\lambda e^{-\lambda x}}_{f'}\,dx$ can be calculated using integration by parts.

We get the recursion $\mathbb{E}\,X^n = \frac{n}{\lambda}\,\mathbb{E}(X^{n-1})$. From here

We get the recursion $\mathbb{E}X^n = \frac{n}{\lambda}\mathbb{E}(X^{n-1})$. From here $\mathbb{E}X = 1/\lambda$ and $\operatorname{Var} X = 1/\lambda^2$.

- memoryless property: $\mathbb{P}(X > s+t \mid X > t) = \mathbb{P}(X > s)$ for all $s,t \geq 0$.

$$\iff \underbrace{\mathbb{P}(X > s+t)}_{e^{-\lambda(s+t)}} = \mathbb{P}(X>t)\,\mathbb{P}(X>s)$$

$$e^{-\lambda(s+t)} = e^{-\lambda t} \cdot e^{-\lambda s} \checkmark$$

The exponential distribution is the only cont. distr. with this property.

In practice usually use exponential distr. for the elapsed time until a certain event happens.

Ex. At a public phone booth the average length of a call is 10 minutes. Someone arrives immediately before you. What is the probability that

(a) you have to wait at least 10 minutes?

(b) you wait between 10 to 20 minutes until you can make your call?

$X :=$ waiting time $\overset{d}{=} \operatorname{Exp}(\lambda)$, where $\lambda = 1/10 = 1/\mathbb{E}X$

So $\mathbb{P}(X > 10) = e^{-\lambda 10} = e^{-1} \approx 0.368$

$\mathbb{P}(10 < X < 20) = F(20) - F(10) = e^{-1} - e^{-2} \approx 0.233$

What if there are two phone booths? Both are occupied when you arrive. What is the probability that you will be the last one to finish your call?

You wait until the first one finishes and take his place. The memoryless property implies that the time until the other one will finish is the same $\operatorname{Exp}(\lambda)$ time that you will finish. Hence, by symmetry, you will finish last with probability $1/2$.

- Poisson point process

Let $N(t)$ be the number of occurences of an event in the interval $[0,t]$.

Assume that $N(t) \overset{d}{=} \operatorname{Poi}(\lambda t)$, and let X be the first occurence of the event.

Then $\mathbb{P}(X > t) = \mathbb{P}(N(t) = 0) = e^{-\lambda t}$, thus the c.d.f. of X is

$F(t) = \mathbb{P}(X \leq t) = 1 - e^{-\lambda t} \overset{d}{=} \operatorname{Exp}(\lambda)$ the waiting time is exponential distribution

In a restaurant an average of 2.5 glasses break each month. What is the probability that in the next 10 days (1/3 month) no glasses will break?

Let X be the time that the first glass breaks $\Rightarrow X \overset{d}{=} \operatorname{Exp}(2.5)$

$\mathbb{P}(X > 1/3) = e^{-2.5 \cdot \frac{1}{3}} \approx 0.4346$.

D1 Further continuous distributions

Def X has Laplace distribution (also called double exponential) if its p.d.f. is $f(x) = \frac{1}{2}\lambda e^{-\lambda|x|}$ $x \in \mathbb{R}$; $F(x) = \begin{cases} \frac{1}{2}e^{\lambda x} & x<0 \\ 1-\frac{1}{2}e^{-\lambda x} & x>0 \end{cases}$

Def The sum of k i.i.d. $\mathrm{Exp}(\lambda)$ is the Erlang distribution with param. (k,λ). The p.d.f. is $f(x;k,\lambda) = \frac{\lambda^k x^{k-1} e^{-\lambda x}}{(k-1)!}$

k: shape parameter , λ: rate parameter

The generalization of these distr.-s is the Gamma distribution, where k is allowed to be any positive real number

$(k-1)!$ is replaced by $\Gamma(k)$ (gamma function $\Gamma(z) = \int_0^\infty x^{z-1}e^{-x}\,dx$.

The χ^2-distribution ("chi-squared") is a special case with $\lambda = \frac{1}{2}$ and k an even natural number.

Weibull, Cauchy, Beta distributions are also important in applications.

4. Limit Theorems

Thm (Strong) Law of Large Numbers (LLN)

Let $X_1, X_2, \ldots$ be a sequence of i.i.d. r.v.-s with mean $\mu = \mathbb{E}X_i < \infty$.

Then, with probability 1 $\quad \dfrac{X_1 + \ldots + X_n}{n} \to \mu$ as $n \to \infty$.

other notation: $\quad \mathbb{P}\left(\lim\limits_{n\to\infty} \dfrac{X_1 + \ldots + X_n}{n} = \mu \right) = 1.$

for an i.i.d. sequence the average tends to the mean.

Ex X_i is a Bernoulli r.v., where $X_i = 1$ if an experiment E is successful. We repeat the experiment many times. How can we approximate the probability that E is a success?

Use the LLN! $\mathbb{E}X_i = \mathbb{P}(E_i)$

So just take the average of the number of successes $\Rightarrow$ that will always give a good approx.

Thm Central Limit Theorem (CLT)

Let $X_1, X_2, \ldots$ be a sequence of i.i.d. r.v.-s with

finite mean μ
& finite variance σ^2. Then the distribution of $\dfrac{X_1 + \ldots + X_n - n\mu}{\sqrt{n\sigma^2}}$

tends to the standard normal as $n \to \infty$. standardized r.v.

That is, for every $-\infty < a < \infty$

$$\mathbb{P}\left(\frac{X_1 + \ldots + X_n - n\mu}{\sqrt{n\sigma^2}} \le a \right) \xrightarrow[\text{as } n\to\infty]{} \Phi(a),$$ where $\Phi(a)$ is the c.d.f. of $N(0,1)$

$$\Phi(a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-x^2/2}\, dx.$$

Remark: it does not matter what distribution you use (as long as finite variance) the sum will always have a distribution that is approximately normal

- let $X_i \sim \text{Bern}(p)$ i.i.d., the $\sum_{i=1}^{n} X_i \sim \text{Bin}(n,p)$, thus in this special case the CLT is

$$\mathbb{P}\left(\frac{\text{Bin}(n,p) - np}{\sqrt{np(1-p)}} \leq a\right) \xrightarrow[n\to\infty]{} \Phi(a).$$

For historical reasons, this is know as the de Moivre-Laplace CLT

- more general forms are also true. For example

$X_1, X_2, \ldots$ are independent, but not identically distributed.

Let $\mu_i = \mathbb{E} X_i$ and $\sigma_i^2 = Var(X_i)$.

If a) X_i are uniformly bounded ($\exists M>0 : P(X_i \leq M) = 1 \ \forall i$)

AND (b) $\sum_{i=1}^{\infty} \sigma_i^2 = \infty$

then $$\mathbb{P}\left(\frac{\sum_{i=1}^n (X_i - \mu_i)}{\sqrt{\sum_{i=1}^n \sigma_i^2}} \leq a\right) \to \Phi(a) \text{ as } n\to\infty.$$

Ex $X_1, X_2, \ldots, X_{10}$ i.i.d. $\sim Uni(0,1)$ $\quad \mathbb{P}\left(\sum_{i=1}^{10} X_i > 6\right) \approx ?$

$$\mathbb{E} X_i = \frac{1}{2} \overset{\text{lin.}}{\Longrightarrow} \mathbb{E}\left(\sum_{i=1}^{10} X_i\right) = \frac{10}{2} = 5$$

$$Var\, X_i = \frac{1}{12} \overset{\text{indep.}}{\Longrightarrow} Var\left(\sum_{i=1}^{10} X_i\right) = 10 \cdot \frac{1}{12} = \frac{5}{6}$$

By the CLT:

$$\mathbb{P}\left(\sum_{i=1}^{10} X_i > 6\right) = \mathbb{P}\left(\frac{\sum_{i=1}^{10} X_i - 5}{\sqrt{5/6}} > \frac{6-5}{\sqrt{5/6}}\right) = 1 - \mathbb{P}\left(\frac{\sum_{i=1}^{10} X_i - 5}{\sqrt{5/6}} \leq \sqrt{\frac{6}{5}}\right)$$

$$\approx 1 - \Phi\left(\sqrt{6/5}\right) \approx 0.1367.$$

Thm Important inequalities in probability

- Markov's inequality If X is a r.v. that takes only non-neg. values, then for every $a>0$ $$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E} X}{a}.$$

- Chernoff bound / turbo Markov For any r.v. X and $a \in \mathbb{R}$

$$\mathbb{P}(X \geq a) \leq e^{-ta}\, \mathbb{E}\left(e^{tX}\right) \quad \text{for all } t>0,$$

$$\mathbb{P}(X \leq a) \leq e^{-ta}\, \mathbb{E}\left(e^{tX}\right) \quad \text{for all } t<0.$$

$\mathbb{P}(X \geq a) \leq e^{-ta}\,\mathbb{E}(e^{tX})$ for all $t>0$,

$\mathbb{P}(X \leq a) \leq e^{-ta}\,\mathbb{E}(e^{tX})$ for all $t<0$. moment generating function of X

because $\mathbb{P}(X \geq a) \overset{t>0}{=} \mathbb{P}(tX \geq ta) = \mathbb{P}(e^{tX} \geq e^{ta})$ now Markov $\leq$.

- Chebyshev's inequality If the r.v. X has finite mean μ and finite variance σ^2, then for every $a>0$ $\mathbb{P}(|X-\mu| \geq a) \leq \sigma^2/a^2$.
- One sided Chebyshev inequality If the r.v. X has mean $=0$ and finite variance σ^2, then for any $a>0$ $\mathbb{P}(X \geq a) \leq \dfrac{\sigma^2}{\sigma^2+a^2}$.

Ex Let $X \overset{d}{=} \mathrm{Poi}(100)$. Let use all the inequalities to get different bounds for the probability $\mathbb{P}(X \geq 116)$, then approximate the probab. with the CLT and finally determine the exact value!

$\mathbb{E}X = \mathrm{Var}\,X = 100$. Since $\mathrm{Poi}(a)+\mathrm{Poi}(b) \overset{d}{=} \mathrm{Poi}(a+b)$ (if they are indep.) we can write $X = Y_1 + \dots + Y_{100}$, where $Y_i \overset{d}{=} \mathrm{Poi}(1)$ or $Z_1 + \dots + Z_{10}$, where $Z_i \overset{d}{=} \mathrm{Poi}(10)$.

(a) Markov's inequality: $\mathbb{P}(X \geq 116) \leq \dfrac{\mathbb{E}X}{116} = \dfrac{100}{116} = 0.86207$

(b) Chernoff bound:

first calculate $\mathbb{E}(e^{tX}) = \sum_{k=0}^{\infty} e^{tk} e^{-\lambda} \dfrac{\lambda^k}{k!} = e^{-\lambda} \underbrace{\sum_{k=0}^{\infty} \dfrac{(e^t\lambda)^k}{k!}}_{e^{\lambda e^t}} = e^{\lambda(e^t-1)}$.

So $\mathbb{P}(X \geq a) \leq e^{\lambda(e^t-1)-at}$ this is minimal if exponent is minimal

Let us find the extreme point of the exponent:

$$\left((\lambda e^t-1)-at\right)'_t = \lambda e^t - a = 0 \Rightarrow e^{t^*} = a/\lambda \text{ and } t^* = \ln(a/\lambda).$$

substitute this back to obtain

$$\mathbb{P}(X \geq a) \leq e^{\lambda(e^{t^*}-1)-at^*} = e^{a-\lambda}\cdot\left(\frac{\lambda}{a}\right)^a.$$

In our case $a=116$, thus

$$\mathbb{P}(X \geq 116) \leq e^{116-100}\cdot\left(\frac{100}{116}\right)^{116} \approx 0.2962$$ (much better than simple Markov)

(simple Markov)

(c) one sided Chebyshev $\leq$:

$$P(X \geq 116) = P(\underbrace{X - 100}_{E(X-100)=0} \geq 16) \leq \frac{100}{100 + 16^2} \approx 0.2809$$

(d) CLT: $$P(X \geq 116) = 1 - P(X \leq 115) = 1 - P\left(\frac{X - 100.1}{\sqrt{100.1}} \leq \frac{115 - 100}{\sqrt{100}}\right)$$

$$= 1 - P\left(\frac{X - 100}{10} \leq \frac{3}{2}\right) \approx 1 - \Phi(3/2) = 0.0668.$$

(e) the exact value

$$P(X \geq 116) = 1 - P(X \leq 115) \overset{\text{computer software}}{=} 0.06318.$$

remark: as we choose 116 larger and larger (farther away from the expected value), then the Chernoff bound becomes better and better.

For example if $a = 126$, then already

$$P(X \geq 126) \leq \begin{cases} \dfrac{100}{100 + 26^2} \approx 0.1288 & \text{one sided Chebyshev} \\ e^{26} \cdot \left(\dfrac{100}{126}\right)^{126} \approx 0.044 & \text{for Chernoff bound} \end{cases}$$

CLT approx is $P(X \geq 126) \approx 1 - \Phi(5/2) \approx 0.0062.$