INTRODUCTION

A quick disclaimer before reading the notes: they were taken by me during the lectures, they do not replace the professor's work and are not sufficient for passing the exams. Moreover they might contain mistakes, so please double check all that you read. The notes are freely readable and can be shared (always remembering to credit me and to not obscure this page), but **can't** be modified. Thank you and hope these notes are useful!

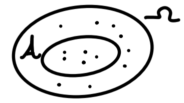
-Francesca Cinelli

PROBABILITY

Probabilistic scheme 3.10

3 ingredients

→ sample space - what are the possible outcomes



$\Omega = \{$ possible outcomes of the experiment $\}$

ex.

- throwing a dice $\Omega = \{1,2,3,4,5,6\}$
- toss a coin $\Omega = \{H,T\}$
- pick 5 balls out of urn with 90 balls $\Omega = \{12345, 21345, 60\,61\,62\,63\,70\}$
- measure position and velocity of all molecules of air in room
  $\Lambda \times \mathbb{R}^3 \rightarrow N$ molecules: $\Omega = (\Lambda \times \mathbb{R}^3)^N$

mathematically Ω is a set (finite) without any additional structure

→ algebra of events - a class/family of subsets of Ω closed w.r.t. $\cap, \cup, *^c$

↳ means if $A, B \in \mathcal{A}$ then $A \cup B, A \cap B, A^c \in \mathcal{A}$

↳ binary question on the result of the random experiment

ex.

- coin tossing: $\mathcal{A}$ { does the coin give head? } Yes/No
- dice throwing: $\mathcal{A}$ { " " dice give a # ≥ 4? } Yes/No
- lotto : $\mathcal{A}$ { have the # 1 and 2 been picked? } Yes/No
- molecules: $\mathcal{A}$ { all the molecules at your side of room? } Yes/No

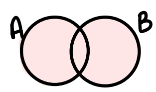
mathematically : event corresponds to a subset of the sample space

$\Omega = \{1,2,3,4,5,6\}$ $\mathcal{A} = \{4,5,6\} \subset \Omega$

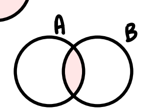
Operations on events

A, B are events

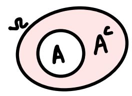
→ $A \cup B = \{A$ or $B\} = \{$either A or B occured$\} = \{\omega \in A$ or $\omega \in B\}$



→ $A \cap B = \{A$ and $B\} = \{$both A and B occured$\} = \{\omega \in \Omega : \omega \in A$ and $\omega \in B\}$



→ $A^c = \Omega \setminus A = \{$not $A\} = \{A$ has not occured$\}$



Properties of $\cup, \cap, *^c$

→ $A \cup B = B \cup A$
→ $(A \cup B) \cup C = A \cup (B \cup C) \Rightarrow A \cup B \cup C$
→ $A \cap B = B \cap A$
→ $(A \cap B) \cap C = A \cap (B \cap C) \Rightarrow A \cap B \cap C$
→ $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
→ $(A \cap B)^c = A^c \cup B^c$

Elementary events

↳ subsets of Ω by considering the singleton $\{\omega\}$. ω is a point in Ω

ex.

$\Omega = \{1,2,3,4\}$ $3 \in \Omega$ el. event
$\{3\} \subset \Omega$ event

→ Probability

$\mathbb{P}: \mathcal{A} \to [0,1]$ if $A \in \mathcal{A}$ is an event $\mathbb{P}(A) \in [0,1]$ is the probability of A such that

- $\mathbb{P}(\emptyset) = 0$ • $\mathbb{P}(\Omega) = 1$ (normalization)
- $\mathbb{P}$ is additive → that is if $A, B \in \mathcal{A}$ are disjoint ($A \cap B = \emptyset$) then $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$

Uniform probability

↳ the elementary events have the same probability (maximal ignorance) then

$\mathbb{P}(A) = \frac{|A|}{|\Omega|}$ ↳ cardinality of A = # elements in A

Consequences of these axioms

- $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$

  P: this follows from additivity $\Omega = A \cup A^c$ $A \cap A^c = \emptyset$

  by additivity $\mathbb{P}(\Omega) = \mathbb{P}(A) + \mathbb{P}(A^c) = 1 \Rightarrow \mathbb{P}(A^c) = 1 - \mathbb{P}(A)$

- A, B events, not necessarily disjoint $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$

  P: from additivity $B' := B \setminus (A \cap B)$ then $A \cap B' = \emptyset$

  * $\mathbb{P}(A \cup B') = \mathbb{P}(A) + \mathbb{P}(B')$ $B = B' \cup (A \cap B)$ → disjoint union

  $\mathbb{P}(B) = \mathbb{P}(B') + \mathbb{P}(A \cap B) \to$ * $\mathbb{P}(B') = \mathbb{P}(B) - \mathbb{P}(A \cap B)$

  → $\mathbb{P}(A \cup B') = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$ $[A \cup B = A \cup B']$

  $\Rightarrow \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$

How to construct a probability on a given (finite) sample space Ω?

note. if $|\Omega| = n \to |\mathcal{A}| = 2^n$

infact it is enough to construct the probability of elementary events

let p → function of a single element $p: \Omega \to [0,1]$ s.t. $\sum_{w \in \Omega} p(w) = 1$

then I define $\mathbb{P}: \mathcal{A} \to [0,1]$ by $\mathbb{P}(A) := \sum_{w \in A} p(w)$

claim: $\mathbb{P}$ is a probability on $\mathcal{A}$, namely, $\mathbb{P}$ is additive

$\mathbb{P}(\emptyset) = \sum_{w \in \emptyset} p(w) = 0$, $\mathbb{P}(\Omega) = \sum_{w \in \Omega} p(w) = 1$, $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) = \sum_{w \in A \cup B} p(w) = \sum_{w \in A} p(w) + \sum_{w \in B} p(w)$

↳ disjoint

Uniform probability thus corresponds to the case $p(w) = \frac{1}{|\Omega|}$ so that

$\mathbb{P}(A) = \sum_{w \in A} p(w) = \sum_{w \in A} \frac{1}{|\Omega|} = \frac{|A|}{|\Omega|}$

Conversely, by restricting the probability $\mathbb{P}$ to elementary events we get a function

$p: \Omega \to [0,1]$ s.t. $\sum_{w \in \Omega} p(w) = 1$ $p(w) = \mathbb{P}(\{w\})$ → $\subset \Omega$ (subset of one singleton)

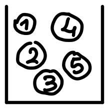
↳ point $\in \Omega$ (singleton)

Infact $\sum_{w \in \Omega} p(w) = \sum_{w \in \Omega} \mathbb{P}(\{w\}) = \mathbb{P}(\bigcup_{w \in \Omega} \{w\}) = \mathbb{P}(\Omega) = 1$

↳ union of all the subsets of one singleton to reconstruct the entire set

Elements of combinatorial calculus 10.10

ex.



$W = 4, 5 \quad B = 1, 2, 3$

$\Omega = \{12, 13, 14, 15, 21, 23, 24, 25, 31, 32, 34, 35, 41, 42, 43, 45, 51, 52, 53, 54\}$

$|\Omega| = 20$

$A = \{2^{nd}$ drawn is white$\} = \{14, 15, 24, 25, 34, 35, 45, 54\}$

$|A| = 8 \qquad \mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{8}{20} = \frac{2}{5}$

or

$|\Omega| = \underbrace{5}_{\text{choices } 1^{st} \text{ draw}} \cdot \underbrace{4}_{\text{choices } 2^{nd} \text{ draw}} = 20 \qquad |A| = \underbrace{2}_{\text{choices } 2^{nd} \text{ draw}} \cdot \underbrace{4}_{\text{choices } 1^{st} \text{ draw}} = 8$

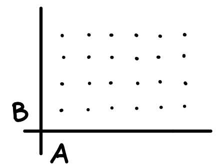
Basic principle of combinatorial calculus

A, B two finite sets

$A \times B = \{(a,b) \mid a \in A,\ b \in B\}$ -> Cartesian product

$|A \times B| = \#$ of ordered pairs with the first in A and the second in B

$= |A| \cdot |B|$



Permutations

$S =$ set $\quad |S| = n$

is a choice of an order between the elements of S, in other words, a permutation is a bijection between S and $\{1, 2, 3, \dots, n\}$

$\gamma : \{1, 2 \dots n\} \to S$ (bijection)

$\gamma(1) =$ first $\quad \gamma(2) =$ second $\quad \gamma(n) =$ last

$\#$ of permutations on $S = n \cdot (n-1) \cdot (n-2) \dots 2 \cdot 1 = n!$

Probabilistic models of drawing balls at random

- box with n balls numbered 1-n, pick k balls

1. ordered drawing with replacement
   - pick the first, look at #, put it back in box, repeat k times

   $\Omega = \{(w_1, w_2, \dots, w_k) \ \ w_i = 1, 2, \dots, n\} = \{1, 2, \dots n\}^k \qquad w_i = \#$ ball on i^{th} draw

   $|\Omega| = n^k$

   ex. k distinct objects, n boxes, distribute k objects among the n boxes. How many choices? n^k

   ex. # anagrams of "banana" $= \frac{6!}{3!\,2!}$

   first distinguish between the 3As and the 2Ns: $B A_1 N_1 A_2 N_2 A_3$

   if so # anagrams = # permutations of the set $\{B, A_1, N_1, A_2, N_2, A_3\} = 6!$

   to correct the overcounting we take away (divide) by the possible orderings of the Ns ($= 2!$) and of the As ($= 3!$)

2. ordered drawing without replacement
   - draw first and put aside, repeat k times ($k \leq n$)

   $\tilde{\Omega} = \{(w_1, w_2 \dots w_k) \ \ w_i = 1, 2, \dots, n \quad i \neq j \Rightarrow w_i \neq w_j\}$

   $|\tilde{\Omega}| = n \cdot (n-1)(n-2) \dots (n-k+1) = \frac{n!}{(n-k)!}$

   remark: if $k = n$ $\ w \in \Omega$ is infact a permutation

   ex. building with 10 floors, lift with 7 people. $\mathbb{P}$(people all get off at different floors)

   $\Omega = \{(w_1, \dots w_7) \ w_i = 1 \dots 10\}$ $\ w_1 =$ floor at which passenger 1 exits ecc. $\ |\Omega| = 10^7 \quad A = \{\uparrow\} \quad |A| = 10 \cdot 9 \cdot 8 \dots 4 = \frac{10!}{3!}$

   $\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{10!}{3!\,10^7} \simeq 0.06$

3. unordered drawing/sampling without replacement

→ pick k balls together ($k \le n$)

let $\tilde{\Omega}$ be the sample space relative to ordered sampling, now we want to forget about the order. Introduce the binary relationship on $\tilde{\Omega}$ defined by $\omega \sim \omega'$ iff ω and ω' differ only for the order. For example $(3,1,7) \sim (1,3,7) \sim (7,1,3)$ ecc. Readily $\sim$ is an equivalence relationship

$\omega \sim \omega$, $\omega \sim \omega' \iff \omega' \sim \omega$, $\omega \sim \omega'$ $\omega' \sim \omega'' \Rightarrow \omega \sim \omega''$

set $\underline{\Omega = \tilde{\Omega}/\sim}$: family of equivalence classes (wrt $\sim$) in $\tilde{\Omega}$

↳ sample space

equivalently I can select an element in each equivalence class by choosing the one in which $\omega_1 < \omega_2 < \ldots < \omega_k$. $\Omega = \{(\omega_1 \ldots \omega_k),\ 1 \le \omega_1 < \omega_2 < \ldots < \omega_k \le n\}$

#smallest extracted ↳ # biggest extracted

remark: each equivalence (wrt $\sim$) in $\tilde{\Omega}$ has exactly $k!$ elements hence $|\Omega| = |\tilde{\Omega}/\sim| = \frac{|\tilde{\Omega}|}{k!}$

$|\tilde{\Omega}| = |\tilde{\Omega}/\sim| \cdot k!$

$|\Omega| = \frac{n!}{k!(n-k)!} =: \binom{n}{k}$ "choose k out of n"

↳ # of ways of selecting k objects among n

Properties

$\binom{n}{k} = \binom{n}{n-k}$ ↳ unchosen

* $\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1}$ → # of ways we can choose $k+1$ elements among $n+1$

* proof

$1 \ldots n+1$ elements → if 1 is chosen $\binom{n}{k}$

how many ways can we choose $k+1$? → if 1 is not chosen $\binom{n}{k+1}$

→ they are exclusive so we can add them

↳ $\binom{n+1}{k+1} = \binom{n}{k} + \binom{n}{k+1}$

ex. 4 boxes, 10 (distinct) balls are placed in these boxes

$\mathbb{P}$(in box A there are 5 balls)?

$\Omega = \{(\omega_1 \ldots \omega_{10})\ \omega_i \in \{A,B,C,D\}\}$ "in which box I put each ball" $|\Omega| = 4^{10}$

$A = \{$box A contains 5 balls$\} = \{\omega \in \Omega : \#i$ for which $\omega_i = A = 5\}$ (= {10 letter words in alphabet A,B,C,D with 5 As})

$|A| = 3^5 \binom{10}{5}$ $\qquad \mathbb{P}(A) = \frac{\binom{10}{5} 3^5}{4^{10}}$

↳ choices of non A letters

Newton binomial

$a, b \in \mathbb{R}$

$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$ infact $(a+b)^n = \overbrace{(a+b)\ldots(a+b)}^{n\text{-times}} = b^n + nab^{n-1} + \binom{n}{2} a^2 b^{n-2} + \ldots + \binom{n}{k} a^k b^{n-k}$

Multinomial coefficient

4 players, each is delt 13 cards. How many hands of cards for the 4 players?

$$\binom{52}{13}\binom{39}{13}\binom{26}{13}\binom{13}{13} = \frac{52!}{13!39!} \cdot \frac{39!}{13!26!} \cdot \frac{26!}{13!13!} \cdot \frac{13!}{13!0!} = \frac{52!}{13!13!13!13!} = \binom{52}{13\ 13\ 13\ 13}$$

→ number of ways of dividing n objects into k distinct groups of cardinality $n_1, n_2, \ldots, n_k$

$$\frac{n!}{n_1! n_2! \ldots n_k!} = \binom{n}{n_1, n_2, \ldots, n_k}, \quad \sum_{i=1}^{k} n_i = n$$

Inclusion/exclusion principle

- A_1, A_2 events (not necessarily disjoint)
  $\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2) - \mathbb{P}(A_1 \cap A_2)$
- A_1, A_2, A_3 events (not necessarily disjoint)
  $\mathbb{P}(A_1 \cup A_2 \cup A_3) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \mathbb{P}(A_3) - \mathbb{P}(A_1 \cap A_2) - \mathbb{P}(A_2 \cap A_3) - \mathbb{P}(A_3 \cap A_1) + \mathbb{P}(A_1 \cap A_2 \cap A_3)$
- $A_1, A_2, \dots, A_n$ events
  $\mathbb{P}(A_1 \cup A_2 \cup \dots \cup A_n) = \mathbb{P}(A_1) + \dots + \mathbb{P}(A_n) - \underbrace{[\mathbb{P}(A_1 \cap A_2) + \mathbb{P}(A_1 \cap A_3) + \dots + \mathbb{P}(A_{n-1} \cap A_n)]}_{\binom{n}{2} \text{ terms}} + [\mathbb{P}(A_1 \cap A_2 \cap A_3) + \dots + \mathbb{P}(A_{n-2} \cap A_{n-1} \cap A_n)] + \dots + (-1)^{n-1} \mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_n)$

$\rightarrow \mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \le i_1 < \dots < i_k \le n} \mathbb{P}(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k})$

ex. n guests, each with an umbrella, party ends and everyone picks an umbrella at random. What is the probability that at least one picks his own umbrella?

$A = \{$at least one picks his own umbrella$\}$

$\Omega = \{\omega_1, \omega_2 \dots \omega_n \mid \omega_i \neq \omega_j, \ i \neq j \le n\}$ $\quad |\Omega| = n(n-1)(n-2)\dots 2 \cdot 1 = n!$

$\Omega = \{$set of permutations of $\{1, \dots, n\}\}$

$A = \{\omega \in \Omega : \exists i \in \{1 \dots n\}$ s.t. $\omega_i = i\} = \{$the permutation has at least one fixed point$\} = A_1 \cup A_2 \cup \dots \cup A_n = \bigcup_{i=1}^{n} A_i$

$A_n = \{n$ picks his own umbrella$\} = \{\omega \in \Omega : \omega_n = n\}$

$\mathbb{P}(A_1) = \frac{|A_1|}{|\Omega|} = \frac{(n-1)!}{n!} = \frac{1}{n} \quad \mathbb{P}(A_2) = \frac{1}{n} \ \dots \ \mathbb{P}(A_n) = \frac{1}{n}$

$\mathbb{P}(A) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \dots + \mathbb{P}(A_n) = n \cdot \frac{1}{n} = 1$ X $\quad A_1 \dots A_n$ are not disjoint, so inclusion/exclusion principle

already have prob. of single events so compute prob. of intersections

$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(\text{both } 1 \text{ and } 2 \text{ picked their own umbrella}) = \frac{|\{\omega \in \Omega : \omega_1 = 1 \ \omega_2 = 2\}|}{|\Omega|} = \frac{(n-2)!}{n!}$

infact if $i_1 \neq i_2$ $\mathbb{P}(A_{i_1} \cap A_{i_2}) = \frac{(n-2)!}{n!}$

$\mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_k) = \mathbb{P}(1 \dots k \text{ pick their own umbrella}) = \frac{|\{\omega \in \Omega : \omega_1 = 1 \dots \omega_k = k\}|}{|\Omega|} = \frac{(n-k)!}{n!}$

$\mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_n) = \frac{1}{n!}$

$\mathbb{P}(\text{at least one guest picks his own umbrella}) = \mathbb{P}\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \le i_1 < \dots < i_k \le n} \frac{(n-k)!}{n!}$

$= \sum_{k=1}^{n} (-1)^{k+1} \binom{n}{k} \frac{(n-k)!}{n!} = \sum_{k=1}^{n} (-1)^{k+1} \frac{\cancel{n!}}{k!\cancel{(n-k)!}} \frac{\cancel{(n-k)!}}{\cancel{n!}} = \sum_{k=1}^{n} (-1)^{k+1} \frac{1}{k!}$

what happens when $n \to \infty$?

$\mathbb{P}(\text{nobody picks his own umbrella}) = \mathbb{P}(\text{permutation has no fixed points}) = 1 - \mathbb{P}(\text{at least one} \dots)$

$= 1 - \sum_{k=1}^{n} (-1)^{k+1} \frac{1}{k!} = \sum_{k=0}^{n} (-1)^k \frac{1}{k!}$

in particular

$\lim_{n \to \infty} \mathbb{P}(\text{nobody} \dots) = \sum_{k=0}^{\infty} (-1)^k \frac{1}{k!} = \frac{1}{e}$

proof

$e = \sum_{k=0}^{\infty} \frac{1}{k!}$

$e^x = \sum_{k=0}^{\infty} x^k \frac{1}{k!}$ $\quad$ in this case $x = -1$ so $e^{-1} = \frac{1}{e}$

Product probability spaces and independence of events 17.10

two probabilistic schemes: $(A_1, \mathbb{P}_1)$ and $(A_2, \mathbb{P}_2)$
construct a scheme for the joint experiment: $\Omega = \Omega_1 \times \Omega_2 = \{(\omega_1,\omega_2) \; \omega_1 \in \Omega_1, \omega_2 \in \Omega_2\}$
many possibilities of introducing a probability on Ω: when the two random experiments do not affect eachother, a natural choice is the product probability $\mathbb{P}(\{(\omega_1,\omega_2)\}) := \mathbb{P}_1(\{\omega_1\}) \cdot \mathbb{P}_2(\{\omega_2\})$
check that it defines a probability: $\sum_{\omega\in\Omega} \mathbb{P}(\{\omega\}) = 1$

$$\hookrightarrow \sum_{\omega_1\in\Omega_1}\sum_{\omega_2\in\Omega_2} \mathbb{P}_1(\{\omega_1\})\mathbb{P}_2(\{\omega_2\}) \overset{\text{because independent}}{=} \underbrace{\sum_{\omega_1\in\Omega_1} \mathbb{P}_1(\{\omega_1\})}_{1} \underbrace{\sum_{\omega_2\in\Omega_2} \mathbb{P}_2(\{\omega_2\})}_{1}$$

ex. toss a coin (A_1), pick a card from a deck of 52 (A_2)
$\Omega = \Omega_1 \times \Omega_2 = \{H1, \dots H52, T1, \dots T52\}$

Compatibility condition

consider an event $A \subset \Omega$ that I can decide only looking at ω_1. that is A has the form $A = A_1 \times \Omega_2$ for some $A_1 \subset \Omega_1$
ex. $A = \{\text{the coin is H}\} = \{H1, H2, \dots, H52\} = \{H\} \times \Omega_2$
$\mathbb{P}(A) = \mathbb{P}_1(A_1)$
proof: $\mathbb{P}(A) = \mathbb{P}(A_1 \times \Omega_2) = \sum_{(\omega_1,\omega_2)\in A_1\times\Omega_2} \mathbb{P}(\{(\omega_1,\omega_2)\}) = \sum_{\omega_1\in A_1}\sum_{\omega_2\in\Omega_2} \mathbb{P}_1(\{\omega_1\})\mathbb{P}_2(\{\omega_2\}) \underset{\text{independent}}{\overset{=1}{=}} \sum_{\omega_1\in A_1} \mathbb{P}_1(\{\omega_1\}) = \mathbb{P}_1(A_1)$

likewise let B an event ... ω_2 that is B has the form $B = \Omega_1 \times B_2$ for some $B_2 \subset \Omega_2 \to \mathbb{P}(B) = \mathbb{P}_2(B_2)$
Rem: let A, B be as before $A = A_1 \times \Omega_2$, $B = \Omega_1 \times B_2$. $\mathbb{P}(A\cap B) = \mathbb{P}((A_1\times\Omega_2)\cap(\Omega_1\times B_2)) = \mathbb{P}(A_1 \times B_2)$
$= \sum_{(\omega_1,\omega_2)\in A_1\times B_2} \mathbb{P}(\{(\omega_1,\omega_2)\}) = \sum_{\omega_1\in A_1}\sum_{\omega_2\in B_2} \mathbb{P}_1(\{\omega_1\})\mathbb{P}_2(\{\omega_2\}) = \mathbb{P}_1(A_1)\cdot\mathbb{P}_2(B_2) = \mathbb{P}(A)\cdot\mathbb{P}(B)$

Consider a probabilistic scheme $(\Omega, \mathbb{P})$
def: A, B events are independent iff $\mathbb{P}(A\cap B) = \mathbb{P}(A)\cdot\mathbb{P}(B)$

independence for three events

$A, B, C \subset \Omega$ are 3 events
2 possibilities
I) pairwise independence
$$\mathbb{P}(A\cap B) = \mathbb{P}(A)\cdot\mathbb{P}(B), \quad \mathbb{P}(A\cap C) = \mathbb{P}(A)\cdot\mathbb{P}(C), \quad \mathbb{P}(B\cap C) = \mathbb{P}(B)\cdot\mathbb{P}(C)$$
II) three-wise independence
$$\mathbb{P}(A\cap B\cap C) = \mathbb{P}(A)\cdot\mathbb{P}(B)\cdot\mathbb{P}(C)$$
$\mathrm{I} \not\Rightarrow \mathrm{II}$
find a probability scheme and three events for which I holds but II fails
$\hookrightarrow$ fair dice 4 faces: R, B, G, RBG (striped)

$\mathcal{R} = \{\text{outcome shows R}\} = \{R, RBG\}$ $\quad \mathbb{P}(\mathcal{R}) = \mathbb{P}(\mathcal{B}) = \mathbb{P}(\mathcal{G}) = \frac{2}{4}$
$\mathcal{B} = \{\text{" " B}\} = \{B, RBG\}$ $\quad \mathbb{P}(\mathcal{R}\cap\mathcal{B}) = \mathbb{P}(\mathcal{B}\cap\mathcal{G}) = \mathbb{P}(\mathcal{R}\cap\mathcal{G}) = \frac{1}{4}$
$\mathcal{G} = \{\text{" " G}\} = \{G, RBG\}$ $\quad \mathbb{P}(\mathcal{R}\cap\mathcal{B}\cap\mathcal{G}) = \mathbb{P}(RBG) = \frac{1}{4} \neq \frac{1}{8} = \mathbb{P}(\mathcal{R})\cdot\mathbb{P}(\mathcal{B})\cdot\mathbb{P}(\mathcal{G})$ (II doesn't hold)

} $\mathcal{R}, \mathcal{B}, \mathcal{G}$ are pairwise independent (I holds)

Def: A, B, C are independent when both I and II hold

Bernoulli scheme 24.10

binary experiment, we repeat n times (n tosses of a coin, not necessarily fair) $n \in \mathbb{N}$ # experiments

single experiment:

$\Omega_1 = \{0, 1\}$ (0 = Tail ⇔ failure, 1 = Head ⇔ success) $\mathbb{P}_1(\{1\}) = p \in [0,1]$ parameter of the model

$\mathbb{P}_1(\{0\}) = 1-p$ $p \in [0,1]$ = probability of success in a single experiment

repeat n times:

$\Omega = \Omega_1 \times \ldots \times \Omega_1 = \{0,1\}^n = \{(\omega_1, \ldots, \omega_n) \ \ \omega_i \in \{0,1\}\}$

$\mathbb{P}$ = product probability of the probability $\mathbb{P}_1$ on the single experiments

$n = 1 \rightarrow \Omega = \{\underset{1-p}{0}, \underset{p}{1}\}$

$n = 2 \rightarrow \Omega = \{\underset{(1-p)^2}{00}, \underset{(1-p)p}{01}, \underset{p(1-p)}{10}, \underset{p^2}{11}\}$

$n = 3 \rightarrow \Omega = \{\underset{(1-p)^3}{000}, \underbrace{001, 010, 100}_{p(1-p)^2}, \underbrace{011, 101, 110}_{p^2(1-p)}, \underset{p^3}{111}\}$

in general:

$$\mathbb{P}(\{\omega_1, \ldots, \omega_n\}) = p^{\#1}(1-p)^{\#0} = p^{\sum_{i=1}^n \omega_i}(1-p)^{n-\sum_{i=1}^n \omega_i}$$

remark:

→ fair dice $p = \frac{1}{2}$ so $\mathbb{P}$ is the uniform probability on Ω

$|\Omega| = 2^n$

$$\mathbb{P}(\{\omega\}) = \left(\tfrac{1}{2}\right)^{\sum_{i=1}^n \omega_i}\left(\tfrac{1}{2}\right)^{n-\sum_{i=1}^n \omega_i} = \left(\tfrac{1}{2}\right)^n = \frac{1}{|\Omega|}$$

Binomial distribution

$A_k = \{\omega \in \Omega : \#1 = k\} = \{\omega \in \Omega : \sum_{i=1}^n \omega_i = k\}$ $k = 0, 1, \ldots, n$

$k = 0 \rightarrow \mathbb{P}(A_0) = \mathbb{P}(\{\underbrace{0, \ldots, 0}_{n\text{-times}}\}) = (1-p)^n$

$k = 1 \rightarrow \mathbb{P}(A_1) = \mathbb{P}(\{\overbrace{(1,0\ldots0), (0,1,\ldots0), (0,0,1,\ldots0), \ldots, (0\ldots0,1)}^{n \text{ tuples}}\}) = |A_1|\, p(1-p)^{n-1} = n\,p(1-p)^{n-1}$

$k = 2 \rightarrow \mathbb{P}(A_2) = \mathbb{P}(\{(1,1,0\ldots0), (0,1,1,0,\ldots0), \ldots (0\ldots0,1,1)\}) = |A_2|\, p^2(1-p)^{n-2} = \binom{n}{2} p^2(1-p)^{n-2}$

→ $k = k \rightarrow \mathbb{P}(A_k) = |A_k|\, p^k(1-p)^{n-k} = \binom{n}{k} p^k (1-p)^{n-k}$

remark: $A_0, A_1 \ldots A_k$ form a partition of Ω

$\bigcup_{k=0}^n A_k = \Omega$ $i \neq j$ $A_i \cap A_j = \emptyset$

in particular $\sum_{k=0}^n \mathbb{P}(A_k) = 1$ infact $\sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} = [p + (1-p)]^n = 1$

Conditional probability

ex. 

sample 2 balls without replacement (ordered)

$\Omega = \{WW, WB, BW, BB\}$

1. number balls $1, \ldots, 5$ → uniform probability → computed probability

$\mathbb{P}(WW) = \frac{6}{20}$ $\mathbb{P}(BW) = \mathbb{P}(WB) = \frac{6}{20}$ $\mathbb{P}(BB) = \frac{2}{20}$

2.

$\mathbb{P}(WW) = \frac{3}{5} \cdot \frac{2}{4}$ $\mathbb{P}(WB) = \frac{3}{5} \cdot \frac{2}{4} = \mathbb{P}(BW) = \frac{2}{5} \cdot \frac{3}{4}$ $\mathbb{P}(BB) = \frac{2}{5} \cdot \frac{1}{4}$

probability conditional probability

Def: given a probabilistic scheme $(\Omega, \mathbb{P})$ and $A, B \subset \Omega$ events
→ $\mathbb{P}(A|B)$ = probability of B given that A occurred $= \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$

in the example:

$$\mathbb{P}(2^{nd}W \mid 1^{st}W) = \frac{\mathbb{P}(WW)}{\mathbb{P}(1^{st}W)} = \frac{6/20}{12/20} = \frac{1}{2} \quad \text{(conditional probability)}$$

$$\mathbb{P}(2^{nd}W \mid 1^{st}B) = \frac{\mathbb{P}(BW)}{\mathbb{P}(1^{st}B)} = \frac{6/20}{8/20} = 3/4$$

$$\mathbb{P}(2^{nd}B \mid 1^{st}W) = \frac{\mathbb{P}(WB)}{\mathbb{P}(1^{st}W)} = \frac{6/20}{12/20} = \frac{1}{2}$$

$$\mathbb{P}(2^{nd}B \mid 2^{nd}B) = \frac{\mathbb{P}(BB)}{\mathbb{P}(1^{st}B)} = \frac{2/20}{8/20} = 1/4$$

Remark: in $\mathbb{P}(A|B)$ the role of the events A, B is not symmetric
A → conditional event, known
B → unknown

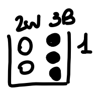
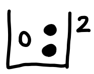
Remark: relation between independence and conditional probability
A, B (independent) $\iff \mathbb{P}(B|A) = \mathbb{P}(B)$
$\iff \mathbb{P}(A|B) = \mathbb{P}(A)$

indeed $\mathbb{P}(B|A) = \mathbb{P}(B)$

$$\frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \mathbb{P}(B) \to \mathbb{P}(A \cap B) = \mathbb{P}(B) \cdot \mathbb{P}(A) \to \text{A and B are independent}$$

same goes for $\mathbb{P}(A|B) = \mathbb{P}(A)$
in general $\mathbb{P}(A|B) \neq \mathbb{P}(B|A)$ but in special cases it might be true

Total probability formula

ex.
trow a fair coin
→ if H pick a ball from 
→ if T " " " " 

$$\mathbb{P}(W) = \mathbb{P}(H)\,\mathbb{P}(W \text{ from } 1) + \mathbb{P}(T)\,\mathbb{P}(W \text{ from } 2)$$
$$= \mathbb{P}(H)\,\mathbb{P}(W|H) + \mathbb{P}(T)\,\mathbb{P}(W|T)$$
$$= \tfrac{1}{2} \cdot \tfrac{2}{5} + \tfrac{1}{2} \cdot \tfrac{1}{3}$$

in general:
let $D_1, \dots, D_n$ be a partition of Ω
→ $\mathbb{P}(A) = \mathbb{P}(D_1)\mathbb{P}(A|D_1) + \dots + \mathbb{P}(D_n)\mathbb{P}(A|D_n) = \sum_{i=1}^{n} \mathbb{P}(D_i)\mathbb{P}(A|D_i)$

Bayes formula

ex.

$$\mathbb{P}(T|W) = \frac{\mathbb{P}(T \cap W)}{\mathbb{P}(W)} = \frac{\mathbb{P}(T)\,\mathbb{P}(W|T)}{\mathbb{P}(T)\mathbb{P}(W|T) + \mathbb{P}(H)\,\mathbb{P}(W|H)} = \frac{1/2 \cdot 1/3}{1/2 \cdot 1/3 + \frac{1}{2} \cdot 2/5}$$

in general:

$$\to \mathbb{P}(D_i|A) = \frac{\overbrace{\mathbb{P}(D_i \cap A)}^{\text{conditional probability}}}{\underbrace{\mathbb{P}(A)}_{\text{total probability}}} = \frac{\mathbb{P}(D_i)\,\mathbb{P}(A|D_i)}{\mathbb{P}(D_1)\mathbb{P}(A|D_1) + \dots + \mathbb{P}(D_n)\mathbb{P}(A|D_n)}$$

Conditional probability of subsets 31.10

ex. deck 52 cards. Pick one at random. $\mathbb{P}(\text{ace of hearts} \mid \text{hearts}) = \dfrac{\mathbb{P}(\{\text{ace } H\} \cap \{H\})}{\mathbb{P}\{H\}} = \dfrac{\overbrace{\mathbb{P}(\text{ace of } H)}^{\frac{1}{52}}}{\underbrace{\mathbb{P}(H)}_{\frac{13}{52}}} = \dfrac{1}{13}$

one can argue directly that this $\mathbb{P}$ is $\frac{1}{13}$.

general situation

fix $A \subset \Omega$. Consider the map $B \longmapsto \mathbb{P}(B \mid A)$ it is a probability on the set A. Check that it is additive $B_1 \cap B_2 = \emptyset$ $B_1, B_2 \subset A$, we want to show that $\mathbb{P}(B_1 \cup B_2 \mid A) = \mathbb{P}(B_1 \mid A) + \mathbb{P}(B_2 \mid A)$

$\rightarrow \dfrac{\mathbb{P}(B_1 \cup B_2 \mid \cap A)}{\mathbb{P}(A)} = \dfrac{\mathbb{P}(B_1 \cap A) \cup \mathbb{P}(B_2 \cap A)}{\mathbb{P}(A)} = \dfrac{\mathbb{P}(B_1 \cap A)}{\mathbb{P}(A)} + \dfrac{\mathbb{P}(B_2 \cap A)}{\mathbb{P}(A)} = \mathbb{P}(B_1 \mid A) + \mathbb{P}(B_2 \mid A)$

check that $\mathbb{P}(A \mid A) = 1$

$\rightarrow \dfrac{\mathbb{P}(A \cap A)}{\mathbb{P}(A)} = \dfrac{\mathbb{P}(A)}{\mathbb{P}(A)} = 1$

remark: if $\mathbb{P}$ is the uniform probability on Ω and fix $A \subset \Omega$, then $A \supset B \longmapsto \mathbb{P}(B \mid A)$ is the uniform probability on A.

$\rightarrow$ use conditional probability and total probability formula in a recursive way to compute (absolute) probability of interesting events

ex. n tosses of biased coin ($p = \mathbb{P}\text{rob}(H)$), $\mathbb{P}(\#H \text{ is even})$.

Direct computation

$\mathbb{P}(\#H \text{ is even}) = \mathbb{P}(\#H = 0) + \mathbb{P}(\#H = 2) + \ldots + \mathbb{P}(\#H = 2h) + \ldots \rightarrow$ binomial distribution $\rightarrow$

$\binom{n}{0} p^0 (1-p)^n + \binom{n}{2} p^2 (1-p)^{n-2} + \ldots + \binom{n}{2h} p^{2h} (1-p)^{n-2h} + \ldots \rightarrow \sum_{h=0}^{(\frac{n}{2})} \binom{n}{2h} p^{2h} (1-p)^{n-2h}$

Recursive arguing

$q_n = \mathbb{P}(\text{in } n \text{ tosses } \#H \text{ is even}) = \overbrace{\mathbb{P}(1^{st} \text{ toss} = H)}^{p} \cdot \underbrace{\mathbb{P}\left(\begin{smallmatrix}\text{in } n \text{ tosses} \\ \#H \text{ is even}\end{smallmatrix} \,\middle|\, \begin{smallmatrix}1^{st} \text{ toss} \\ = H\end{smallmatrix}\right)}_{\mathbb{P}(\text{in } n-1 \text{ tosses } \#H \text{ is odd})} + \overbrace{\mathbb{P}(1^{st} \text{ toss} = T)}^{p-1} \underbrace{\mathbb{P}\left(\begin{smallmatrix}\text{in } n \text{ tosses} \\ \#H \text{ is even}\end{smallmatrix} \,\middle|\, \begin{smallmatrix}1^{st} \text{ toss} \\ = T\end{smallmatrix}\right)}_{\mathbb{P}(\text{in } n-1 \text{ tosses } \#H \text{ is even})}$

condition on the first toss and use tot. prob. $D_1 = \{1^{st} \text{ toss} = H\}$ $D_2 = \{1^{st} \text{ toss} = T\}$

$\rightarrow q_n = p(1 - q_{n-1}) + (1-p) q_{n-1}$

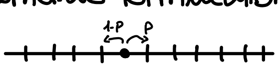
$\begin{cases} q_n = p + (1-2p) q_{n-1} \\ q_1 = 1 - p \end{cases}$ (linear 1 step recursion)

Gambler's ruin problem

two players A and B are addicted betters. In a single bet A wins 1€ with probability p, B wins 1€ with probability $1-p$. the initial capital is A has a€ and B has b€. They continue betting until one of two players is ruined (he loses all his capital). $\mathbb{P}(A \text{ is ruined}) = ? = \mathbb{P}(B \text{ is ruined})$.

$\mathbb{P}(\text{the betting lasts forever}) \overset{\text{by guess}}{=} 0$

remark $\rightarrow$ geometric formulation in terms of random walks



$S_0 = 0$

$S_1 = \begin{cases} 1 & \text{with probability } p \\ -1 & \text{" " } 1-p \end{cases}$

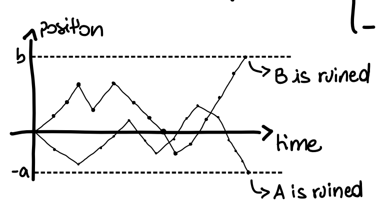
$S_2 = S_1 + \begin{cases} +1 & \text{with prob. } p \\ -1 & \text{" " } 1-p \end{cases}$

$S_{n+1} = S_n + \begin{cases} +1 & \text{" " } p \\ -1 & \text{" " } 1-p \end{cases}$

this corresponds with the gamblers

S_n = sum won by A in the past n bets

$S_0 = 0 \ldots S_{n+1} = S_n + \begin{cases} + & \text{with prob. } p \\ - & \text{" " } 1-p \end{cases}$



$\mathbb{P}(A \text{ is ruined}) = \mathbb{P}\big((S_n)_{n \geq 0} \text{ hits } -a \text{ before } b\big)$

$\mathbb{P}(B \text{ " " }) = \mathbb{P}\big((S_n)_{n \geq 0} \text{ hits } b \text{ before } -a\big)$

$\mathbb{P}(\text{betting lasts forever}) = \mathbb{P}\big((S_n)_{n \geq 0} \text{ does not exit the strip } (-a; b)\big)$

idea: as in previous example, condition on the outcome of the first bet and use Total probability conditioned on this event we are in a slightly different situation as the original random walk started from 0 but now we are either -1 or 1. To cover this we consider the random walk with an arbitrary starting point.

Given $x \in \mathbb{Z}$ let $(S_n^{(x)})_{n\geq 0}$ be the random walk starting from x.

$$S_0^{(x)} = x \quad S_1^{(x)} = x + \begin{cases} 1 & \text{with prob. } p \\ -1 & \text{" " } 1-p \end{cases} \qquad S_{n+1}^{(x)} = S_n^{(x)} + \begin{cases} 1 & \text{with prob. } p \\ -1 & \text{" " } 1-p \end{cases}$$

$\alpha(x) = \mathbb{P}((S_n^{(x)})_{n\geq 0} \text{ hits } -a \text{ before } b) \quad \beta(x) = \mathbb{P}((S_n^{(x)})_{n\geq 0} \text{ hits } b \text{ before } -a)$

$\alpha(0) = \mathbb{P}(A \text{ is ruined}) \quad \beta(0) = \mathbb{P}(B \text{ is ruined})$

-> condition on first step and use Tot. prob. formula

$$\alpha(x) = \mathbb{P}((S_n^{(x)})_{n\geq 0} \text{ hits } -a \text{ before } b) = \underbrace{\mathbb{P}(\text{first step to the right})}_{=p}\overbrace{\mathbb{P}((S_n^{(x)})_{n\geq 0} \text{ hits } -a \text{ before } b \mid \text{first step to right})}^{=\alpha(x+1)} +$$
$$+ \underbrace{\mathbb{P}(\text{first step to the left})}_{=1-p} \underbrace{\mathbb{P}((S_n^{(x)})_{n\geq 0} \text{ hits } -a \text{ before } b \mid \text{first step to left})}_{=\alpha(x-1)}$$

$$\begin{cases} \alpha(x) = p\alpha(x+1) + (1-p)\alpha(x-1) & -a<x<b \\ \alpha(b) = 0 \\ \alpha(-a) = 1 \end{cases} \qquad \begin{cases} \beta(x) = p\beta(x+1) + (1-p)\beta(x-1) & -a<x<b \\ \beta(b) = 1 \\ \beta(-a) = 0 \end{cases}$$

Computing ruin problem solving the equations

- $p = \frac{1}{2}$ but $a \neq b$

$$\left(\tfrac{1}{2}+\tfrac{1}{2}\right)\beta(x) = \tfrac{1}{2}\beta(x+1) + \tfrac{1}{2}\beta(x-1) \rightarrow \tfrac{1}{2}[\beta(x) - \beta(x+1)] = \tfrac{1}{2}[\beta(x+1) - \beta(x)] \rightarrow \beta(x+1) - \beta(x) = \beta(x) - \beta(x-1)$$

β has constant increments*

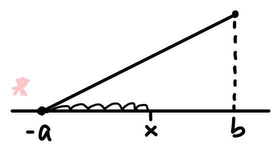
$\beta(-a) = 0$

$\beta(-a+1) = \overset{=0}{\beta(-a)} + \beta(-a+1) - \beta(-a) = \gamma$

$\beta(-a+2) = \beta(-a+1) + \underbrace{\beta(-a+1) - \beta(-a)}_{\gamma}$

$\beta(x) = \beta(-a) + \gamma(x+a)$

$\overset{=1}{\beta(b)} = \gamma(b+a) \Rightarrow \gamma = \dfrac{1}{b+a} \longrightarrow \beta(x) = \dfrac{x+a}{b+a}$



-> $\beta(0) = \mathbb{P}(B \text{ is ruined}) = \dfrac{a}{b+a} \qquad \alpha(0) = \mathbb{P}(A \text{ is ruined}) = \dfrac{b}{b+a}$ → the 2 probabilities sum up to 1 so $\mathbb{P}(\text{betting goes on forever}) = 0$

$\mathbb{P}(A \text{ is ruined}) : \mathbb{P}(B \text{ is ruined}) = b : a$

- $p \neq \frac{1}{2}$

$$(p + (1-p))\beta(x) = p\beta(x+1) + (1-p)\beta(x-1) \rightarrow (1-p)[\beta(x) - \beta(x-1)] = p[\beta(x+1) - \beta(x)] \rightarrow \beta(x+1) - \beta(x) = \tfrac{1-p}{p}[\beta(x) - \beta(x-1)]$$

increments of β are proportional

$\beta(-a) = 0$

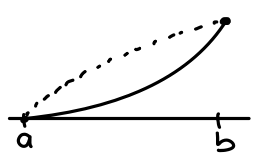
$\beta(-a+1) = \overset{0}{\beta(-a)} + [\beta(-a+1) - \beta(-a)] = \gamma$

$\beta(-a+2) = \beta(-a+1) + [\beta(-a+2) - \beta(-a+1)] = \gamma + \frac{1-p}{p}\overbrace{[\beta(-a+1) - \beta(-a)]}^{\gamma} = \gamma + \gamma\frac{1-p}{p}$

$\beta(x) = \beta(-a) + [\beta(-a+1) - \beta(-a)] + \ldots + [\beta(x) - \beta(x-1)] = \displaystyle\sum_{i=0}^{a+x-1} \gamma\left(\frac{1-p}{p}\right)^i = \gamma\,\frac{1-\left(\frac{1-p}{p}\right)^{a+x}}{1-\frac{1-p}{p}}$

since $\beta(b) = 1$ I deduce: $\gamma\,\dfrac{1-\left(\frac{1-p}{p}\right)^{a+b}}{1-\frac{1-p}{p}} = 1$

$\beta(x) = \dfrac{1-\left(\frac{1-p}{p}\right)^{a+x}}{1-\left(\frac{1-p}{p}\right)^{a+b}}$



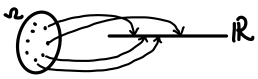
in particular:

$\mathbb{P}(B \text{ is ruined}) = \beta(0) = \dfrac{1-\left(\frac{1-p}{p}\right)^{a}}{1-\left(\frac{1-p}{p}\right)^{a+b}}$

Random variables 7.11

-> X random variable (r.v.) is a real valued function on Ω

$X : \Omega \longrightarrow \mathbb{R}$



X_1 is a bijection, knowing X_1 has the same content of knowing the result ω of the experiment
X_2 is not a bijection, its value has less information than the outcome of the dice
If we are interested only on X_2 we can "restrict" the probabilistic space

interpretation:
is a bet on the result t of the random experiment modeled by $(\Omega, \mathbb{P})$

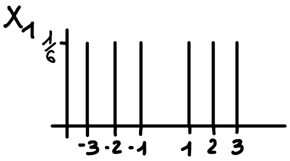
example:
Toss a fair dice -> $\Omega = \{1 \dots 6\}$
$\mathbb{P}$ uniform probability

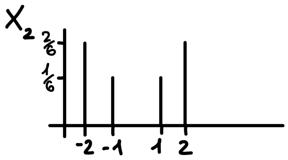
$$X_1 = \begin{cases} -3 & w=1 \\ -2 & w=2 \\ -1 & w=3 \\ 1 & w=4 \\ 2 & w=5 \\ 3 & w=6 \end{cases} \qquad X_2 = \begin{cases} -2 & w=1,2 \\ -1 & w=3 \\ 1 & w=4 \\ 2 & w=5,6 \end{cases}$$

Toss a coin 20 times -> $|\Omega| = 2^{20}$
if interested only in # of heads there are only 21 values (0…20)

Distribution of a random variable

- draw an istogram of $X_{1,2}$ with the relative probability





these have all the relevant info about the random variable

$$X_3 = \begin{cases} -1 & w=1,2 \\ 0 & w=3 \\ 1 & w=4,5 \\ 2 & w=6 \end{cases} \qquad X_{1,2} \text{ are "fair"} \quad X_{3,4} \text{ are not fair}$$

$$X_4 = \begin{cases} -100 & w=1 \\ 0 & w=2,3,4 \\ 1 & w=5,6 \end{cases}$$

how to compensate it with a "fair" entrance fee? $\frac{2}{6}$, $\frac{-98}{6}$ *

- in general

$X : \Omega \longrightarrow \mathbb{R}$

$Im(X) = \{x \in \mathbb{R} : \exists\, w \in \Omega \text{ for which } X(w) = x\} \subset \mathbb{R}$
$= \{X(w),\ w \in \Omega\}$

the distribution of the r.v. is the probability on $Im(X)$ defined by:

$x \in Im(X)$, $\mu_X(\{x\}) = \mathbb{P}(\underbrace{\{w \in \Omega : X(w) = x\}}_{\subset \Omega}) = \mathbb{P}(X = x)$ [value on singletons] (height of the bars)

$B \subset Im(X)$, $\mu_X(B) = \sum_{x \in B} \mu_X(\{x\})$ [general value]

X^{-1} is not well defined on points because X may be not injective, but it is well defined on sets $B \subset \mathbb{R}$

$X^{-1}(B)$ = pre-image of $B = \{w \in \Omega : X(w) \in B\}$

therefore

$\mu_X(B) = \mathbb{P}(X^{-1}(B)) \qquad B \subset Im(X)$

$\mu_X = \mathbb{P} \circ X^{-1}$
↳ composite

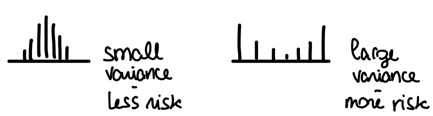
Expectation value of a r.v.

$X : \Omega \to \mathbb{R}$ r.v. * example

$\mathbb{E}(X) = \sum_{x \in Im(X)} x\, \mathbb{P}(X = x)$ "average" value of X

Variance of a r.v.

X_1 and X_2 are "fair" bets so $\mathbb{E}(X_1) = \mathbb{E}(X_2) = 0$ "riskyness" of X
but X_1 is more risky then X_2, how to quantify?



-> so $\mathbb{V}(X_1) > \mathbb{V}(X_2)$

$$\mathbb{V}(X) = \sum_{x \in Im(X)} \overbrace{[x - \mathbb{E}(X)]}^{\text{displacement from expectation value}} \mathbb{P}(X = x) = \sum \underbrace{[x - \mathbb{E}(X)]^2}_{\text{to make positive}} \mathbb{P}(X = x)$$

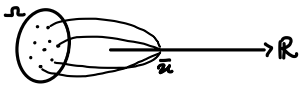
$= \mathbb{E}([X - \mathbb{E}(X)]^2)$ = mean square displacement $= \mathbb{E}(X^2) - (\mathbb{E}(X))^2$

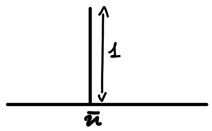
Catalog of popular random variables

1. certain r.v.

$X:\Omega \longrightarrow \mathbb{R}$ is constant

$\exists\, \bar{x} \in \mathbb{R} : X(\omega) = \bar{x} \quad \forall \omega \in \Omega$





$\mathbb{E}(X) = \bar{x}\, \underbrace{\mathbb{P}(X=\bar{x})}_{1} = \bar{x}$ $\qquad \mathbb{V}(X) = (\bar{x} - \underbrace{\mathbb{E}(X)}_{\bar{x}})^2\, \underbrace{\mathbb{P}(X=\bar{x})}_{1} = 0$

remark: if $\mathbb{V}(X) = 0 \rightarrow \sum [x - \mathbb{E}(X)]^2 \mathbb{P}(X=x) = 0 \Rightarrow x - \mathbb{E}(X) = 0 \quad \forall x \in Im(X) \Rightarrow X = \bar{x}$ for some $\bar{x} \in \mathbb{R}$

↳ X is a certain r.v.

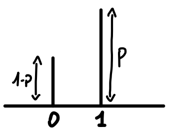
2. X bernoulli r.v.

$Im(X) = \{0,1\}$

$X = \begin{cases} 0 & \text{with prob. } 1-p \\ 1 & \text{" " } p \end{cases}$ $\qquad p$ is a parameter, $p \in [0,1]$

$X \sim Bern(p)$

X describes the result of a binary experiment



$\mathbb{E}(X) = 0 \cdot \mathbb{P}(X=0) + 1 \cdot \mathbb{P}(X=1) = 0(1-p) + 1 \cdot p = p$

$\mathbb{V}(X) = \mathbb{E}([X - \mathbb{E}(X)]^2) = \sum_{x \in Im(X)} (x - \mathbb{E}(X))^2 \mathbb{P}(X=x) = (0-p)^2 \mathbb{P}(X=0) + (1-p)^2 \mathbb{P}(X=1) =$

$p^2(1-p) + (1-p)^2 p = p(1-p)[p + 1 - p] = p(1-p)$

remark: $\mathbb{V}(X)$ is invariant w.r.t. $p \leftrightarrow 1-p$

: $\mathbb{V}(X)$ is maximal for $p = \frac{1}{2}$ [a fair coin is the least predictable one]

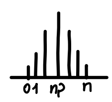
3. binomial r.v.

n tosses of a coin with $p = \mathbb{P}(H)$

$X = \# H$

$\Omega = \{0,1\}^n = \{(\omega_1, \omega_2 .. \omega_n) \ \omega_i \in \{0,1\}\}$

$X(\omega) = \sum_{i=1}^{n} \omega_i \qquad Im(X) = \{0, 1, \dots, n\}$



$\mathbb{P}(X=k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad k = 0, \dots, n$

$$\mathbb{E}(X) = \sum_{k=0}^{n} k\, \mathbb{P}(X=k) = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=1}^{n} k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} = \sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!} p^k (1-p)^{n-k} \overset{h=k-1}{=}$$

$$= \sum_{h=0}^{n-1} \frac{n(n-1)!}{h!(n-1-h)!} p^{h+1} (1-p)^{n-1-h} = np \sum_{h=0}^{n-1} \binom{n-1}{h} p^h (1-p)^{n-1-h} = np\, [\underbrace{p + (1-p)}_{1}]^{n-1} = np$$

→ in the bernoulli scheme, let

$X_i(n) = \begin{cases} 1 & \text{if } \omega_i = 1 \quad (i^{th} \text{ toss is head}) \\ 0 & \text{if } \omega_i = 0 \quad (i^{th} \text{ toss is tail}) \end{cases} = \omega_i$

$X = X_1 + X_2 + \dots + X_n = \sum_{i=1}^{n} X_i$

$X_i \sim Bern(p) \qquad \mathbb{E}(X_i) = p$

→ $\mathbb{E}$ is linear: if X and Y are r.v. then $\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$ and if $\alpha \in \mathbb{R}$ $\mathbb{E}(\alpha X) = \alpha \mathbb{E}(X)$

I deduce $\mathbb{E}(X) = \mathbb{E}(X_1 + \dots + X_n) = \mathbb{E}(X_1) + \dots + \mathbb{E}(X_n) = \sum_{i=1}^{n} \mathbb{E}(X_i) = np$

Remark: $\mathbb{E}(X) = \sum_{\omega \in \Omega} X(\omega) \cdot \mathbb{P}(\{\omega\}) = \sum_{x \in Im(X)} \sum_{\omega : X(\omega) = x} \overbrace{X(\omega)}^{x} \mathbb{P}(\{\omega\}) = \sum_{x \in Im(X)} x \underbrace{\sum_{\omega : X(\omega) = x} \mathbb{P}(\{\omega\})}_{\mathbb{P}(\{\omega \in \mathbb{R} : X(\omega) = x\}) = \mathbb{P}(X = x)} = \sum_{x \in Im(X)} x\, \mathbb{P}(X=x) = \mathbb{E}(X)$

Proof 1: $X+Y$ is the r.v. defined by $(X+Y)(\omega) := X(\omega) + Y(\omega)$

then $\mathbb{E}(X+Y) = \sum_{\omega \in \Omega} (X+Y)(\omega)\, \mathbb{P}(\{\omega\}) = \sum_{\omega \in \Omega} X(\omega) \mathbb{P}(\{\omega\}) + \sum_{\omega \in \Omega} Y(\omega) \mathbb{P}(\{\omega\}) \overset{\text{by remark!}}{=} \mathbb{E}(X) + \mathbb{E}(Y)$

Proof 2: $(\alpha X)(\omega) := \alpha X(\omega)$ then

$\mathbb{E}(\alpha X) = \sum_{\omega \in \Omega} (\alpha X)(\omega) \mathbb{P}(\{\omega\}) = \sum_{\omega \in \Omega} \alpha X(\omega) \mathbb{P}(\{\omega\}) = \alpha \sum_{\omega \in \Omega} X(\omega) \mathbb{P}(\{\omega\}) \overset{\text{by remark!}}{=} \alpha \mathbb{E}(X)$

Variance of binomial r.v.

$\mathbb{V}(X)=\mathbb{E}(X^2)-(\mathbb{E}(X))^2=$ (computation similar to one for $\mathbb{E}(X)$) $=np(1-p)$

$X=\sum_{i=1}^{n} X_i \qquad X_i \sim \text{Bern}(p) \rightarrow \mathbb{E}(X_i)=p \quad \mathbb{V}(X_i)=p(1-p)$

it happens that $\underbrace{\mathbb{V}(X)}_{np(1-p)}=\sum_{i=1}^{n}\underbrace{\mathbb{V}(X_i)}_{p(1-p)}$ why $\mathbb{V}(X+Y)=\mathbb{V}(X)+\mathbb{V}(Y)$ cannot be linear...

this will happen because the r.v. X_i refer to different tosses and therefore independent

Def: X,Y are independent iff $x\in Im(X)$, $y \in Im(Y)$ the events $\{X=x\}$ and $\{Y=y\}$ are independent

equivalently $\mathbb{P}(X=x, Y=y)=\mathbb{P}(X=x)\mathbb{P}(Y=y)\ \forall x \in Im(X)\ \forall y\in Im(Y)$

$\rightarrow$ if X and Y are independent, then $\mathbb{V}(X+Y)=\mathbb{V}(X)+\mathbb{V}(Y)$

Proof: $\mathbb{V}(X+Y)=\mathbb{E}[((X+Y)-\mathbb{E}(X+Y))^2]=\mathbb{E}[[X-\mathbb{E}(X)+Y-\mathbb{E}(Y)]^2]=\mathbb{E}[[X-\mathbb{E}(X)]^2+[Y-\mathbb{E}(Y)]^2+2[X-\mathbb{E}(X)][Y-\mathbb{E}(Y)]]$

$=\mathbb{V}(X)+\mathbb{V}(Y)+2\underbrace{\mathbb{E}[[X-\mathbb{E}(X)][Y-\mathbb{E}(Y)]]}_{cov(X,Y)=\text{covariance between r.v. } X \text{ and } Y} \quad =\mathbb{E}(XY)-\mathbb{E}(X)\cdot\mathbb{E}(Y)$

remark: $cov(X,X)=\mathbb{V}(X)\geq 0$

$\mathbb{V}(X+Y)=\mathbb{V}(X)+\mathbb{V}(Y)+2cov(X,Y)$ [if not independent]

Remark: if X and Y are independent then $cov(X,Y)=0$ in particular $\mathbb{V}(X+Y)=\mathbb{V}(X)+\mathbb{V}(Y)$

Proof: need to show $\mathbb{E}(XY)=\mathbb{E}(X)\cdot\mathbb{E}(Y)$ for X,Y independent

$$\mathbb{E}(XY)=\sum_{\omega\in\Omega}X(\omega)Y(\omega)\mathbb{P}(\{\omega\})=\sum_{x\in Im(X)}\sum_{y\in Im(Y)}\sum_{\substack{\omega\in\Omega:\\ X(\omega)=x\\ X(\omega)=y}}X(\omega)Y(\omega)\mathbb{P}(\{\omega\})=\sum_{x\in Im(X)}\sum_{y\in Im(Y)}xy\sum_{\substack{\omega\in\Omega:\\ X(\omega)=x\\ X(\omega)=y}}\mathbb{P}(\{\omega\})$$

$$=\sum_{x\in Im(X)}\sum_{y\in Im(Y)}xy\,\mathbb{P}(X=x,Y=y)\underset{\text{by independence}}{=}\sum_{x\in Im(X)}x\mathbb{P}(X=x)\sum_{y\in Im(Y)}y\mathbb{P}(Y=y)=\mathbb{E}(X)\cdot\mathbb{E}(Y)$$

$$\mathbb{V}(X)=\mathbb{V}(X_1+\ldots+X_n)=\underbrace{\mathbb{V}(X_1)}_{p(1-p)}+\ldots+\mathbb{V}(X_n)=np(1-p)$$

Remark: if the $cov(X,Y)=0$ it is not always true that X,Y are independent

Heuristic meaning of covariance

make experiment and measure X,Y: do this n times

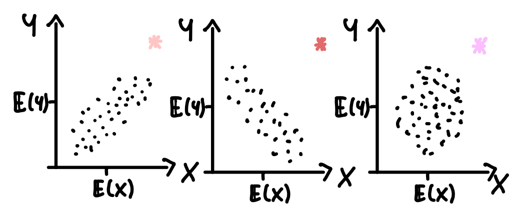
X	Y
x_1	y_1
x_2	y_2
$\vdots$	$\vdots$
x_n	y_n

$cov(X,Y)>0 \rightarrow X,Y$ are positively correlated $\leftarrow$

$cov(X,Y)<0 \rightarrow X,Y$ are negatively correlated $\leftarrow$

$cov(X,Y)\approx 0 \rightarrow X,Y$ are not correlated $\leftarrow$

$\hookrightarrow$ indication of independence between X,Y



How to compute the distribution of $X+Y$?

no simple answer in general. Doable when X and Y are independent

$Z:=X+Y \rightarrow$ aim: $\mathbb{P}(Z=z)=\ldots$ for $z\in Im(Z)$

$$\mathbb{P}(Z=z)=\mathbb{P}(X+Y=z)\underset{\text{additivity of }\mathbb{P}}{=}\sum_{x\in Im(X)}\mathbb{P}(\{X+Y=z\}\cap\{X=x\})=\sum_{x\in Im(X)}\mathbb{P}(Y=z-x, X=x)\underset{\text{independence}}{=}\sum_{x\in Im(X)}\mathbb{P}(Y=z-x)\mathbb{P}(X=x)$$

$$\mathbb{P}(Z=z)=\sum_{x\in Im(X)}\mathbb{P}(X=x)\mathbb{P}(Y=z-x)$$

$X\sim \text{Bin}(n,p) \quad Y\sim\text{Bin}(m,p) \rightarrow$ independent $\rightarrow X+Y\sim\text{Bin}(n+m,p)$

$$\rightarrow \mathbb{P}(X+Y=k)=\sum_{h=0}^{n}\mathbb{P}(X=h)\mathbb{P}(Y=k-h)=\sum_{h=0}^{\min(n,k)}\binom{n}{h}p^h(1-p)^{n-h}\binom{m}{k-h}p^{k-h}(1-p)^{m-(k-h)}=\sum_{h=0}^{\min(n,k)}\frac{n!}{h!(n-h)!}\frac{m!}{(k-h)!(m-k-h)!}p^k(1-p)^{n+m-k}$$

$$=p^k(1-p)^{n+m-k}\underbrace{\sum_{h=0}^{\min(n,k)}\binom{n}{h}\binom{m}{k-h}}_{\binom{n+m}{k}}=\binom{n+m}{k}p^k(1-p)^{m+n-k}$$

Geometric random variable 21.11

motivation:

from reliability theory -> machine subject to faults

-> model the time at which the first fault happens, no aging effect, machine operates on cycles

1^{st} cycle has a fault with probability p (p is a parameter $\in [0,1]$)

2^{nd} cycle independently of the first, machine breaks with probability p

X = time of the first fault

Bernoulli scheme with infinitely many tosses

$\Omega = \{0,1\}^{\mathbb{N}} = \{(\omega_1, \omega_2, \omega_3 \dots \omega_n \dots) \ \omega_i \in \{0,1\}\}$

product probability with $p = \mathbb{P}_0(\{\omega_1 = 1\})$

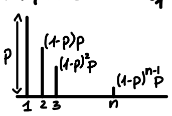
$$X(\omega) = \begin{cases} 1 & \omega_1 = 1 \\ 2 & \omega_1 = 0, \omega_2 = 1 \\ \vdots & \\ n & \omega_1 = \omega_2 = \dots = \omega_{n-1} = 0, \omega_n = 1 \end{cases}$$

[X = Toss in which the first head occurred]

Distribution of $X \sim Geom(p)$

$Im(X) = \{1, 2, 3, \dots, n, \dots\}$

$\mathbb{P}(X=n) = \mathbb{P}(\{\omega \in \Omega : \omega_1 = \dots = \omega_{n-1} = 0, \omega_n = 1\}) = (1-p)^{n-1} p \qquad n \in \mathbb{N}$

istogram 

geometric series: $|q| < 1 \to \sum_{n=0}^{\infty} q^n = \frac{1}{1-q}$

$$\mathbb{P}(\text{sooner or later machine has a fault}) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} \{X=n\}\right) \overset{\text{additivity}}{=} \sum_{n=1}^{\infty} \mathbb{P}(X=n) = \sum_{n=1}^{\infty} (1-p)^{n-1} p \overset{q=(1-p)\in(0,1)}{=} p \sum_{n=1}^{\infty} q^{n-1} \overset{m=n-1}{=} p \sum_{m=0}^{\infty} q^m = p \frac{1}{1-q} = \frac{p}{p} = 1$$

Expectation value

$$\mathbb{E}(X) = \sum_{n=1}^{\infty} n \mathbb{P}(X=n) = \sum_{n=1}^{\infty} n (1-p)^{n-1} p \overset{\text{guess}}{=} \frac{1}{p}$$

$$\text{computation} \to = p \sum_{n=0}^{\infty} n q^{n-1} \overset{nq^{n-1} = \frac{d}{dq} q^n}{=} p \sum_{n=0}^{\infty} \frac{d}{dq} q^n \overset{\text{linearity of derivative}}{=} p \frac{d}{dq} \sum_{n=0}^{\infty} q^n \overset{\text{geometric series}}{=} p \frac{d}{dq} \frac{1}{1-q} = p \frac{1}{(1-q)^2} \overset{q=1-p}{=} \frac{p}{p^2} = \frac{1}{p}$$

Variance

$\mathbb{V}(X) = \mathbb{E}([X - \mathbb{E}(X)]^2) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 = \frac{1-p}{p^2}$

$\mathbb{E}(X^2)$ = try the same trick as in $\mathbb{E}(X)$

Survival probability

$$G(n) = \mathbb{P}(X > n) = \mathbb{P}(\text{machine did not break in the first } n \text{ cycles}) = \sum_{i=n+1}^{\infty} \mathbb{P}(X=i) = \sum_{i=n+1}^{\infty} (1-p)^{i-1} p = p \sum_{i=n+1}^{\infty} q^{i-1}$$

$$j = i-1-n \to = p \sum_{j=0}^{\infty} q^{j+n} = pq^n \sum_{j=0}^{\infty} q^j = pq^n \frac{1}{1-q} = q^n$$

Loss of memory of geometric r.v.

$\mathbb{P}(X = n+i \mid X > n) = \mathbb{P}(\text{breaking at time } n+i \mid \text{it survived the first } n) = \mathbb{P}(X=i)$

Proof: $$\mathbb{P}(X=n+i \mid X>n) = \frac{\mathbb{P}(X=n+i, \cancel{X>n})}{\mathbb{P}(X>n)} \underset{n+i>n}{=} \frac{\mathbb{P}(X=n+i)}{\mathbb{P}(X>n)} = \frac{(1-p)^{n+i-1} p}{(1-p)^n} = (1-p)^{i-1} p = \mathbb{P}(X=i)$$

intersection will only be $n+i \to \{X=n+i\} \cap \{X>n\} = \{X=n+i\}$

Poisson random variable

motivation:

queeing theory

how To dimensionate a server?

model: how many clients arrive in a unit time

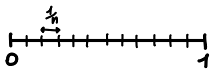
$X = \#$ clients

assumptions:

- in the time interval Δt (Δt small) will arrive a client with probability proportional To Δt

  $\mathbb{P}(\text{a client arrives at the queue in the interval } \Delta t) = \lambda \Delta t + o(\Delta t)$ $\quad \lambda =$ rate of arrivals $\in (0;+\infty)$
- arrivals at disjoint intervals are independent ("clients do not cooperate")

mathematical:



unit interval

$[0,1] = \bigcup_{i=1}^{n} [\frac{i-1}{n}, \frac{i}{n})$ $\quad$ (n large)

$X_i = \#$ clients arriving at time interval $[\frac{i-1}{n}; \frac{i}{n})$ $(i=1 \dots n) = \begin{cases} 0 & \text{with prob. } 1-\frac{\lambda}{n} \\ 1 & \quad " \quad " \quad \frac{\lambda}{n} \end{cases}$ $\quad$ (fine when $n \to \infty$)

$X_i \sim Bern(\frac{\lambda}{n})$

$X^{(n)} = \#$ clients in the line $= X_1 + X_2 + \dots + X_n = \sum_{i=1}^{n} X_i \sim Bin(n; \frac{\lambda}{n})$

we want To understand $n \to \infty$

Thm: (poisson)

Fix $k = 0, 1, \dots, \dots$

$$\lim_{n\to\infty} \mathbb{P}(X^{(n)} = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

Proof:

$X^{(n)} \sim Bin(n; \frac{\lambda}{n})$

$$\mathbb{P}(X^{(n)} = k) = \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k} = \frac{n!}{k!(n-k)!} \left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k} = \frac{\lambda^k}{k!} \frac{\overbrace{n(n-1)\dots(n-k+1)}^{k \text{ terms}}}{n^k} \frac{(1-\frac{\lambda}{n})^n}{(1-\frac{\lambda}{n})^k} =$$

$$= \frac{\lambda^k}{k!} \underbrace{\frac{n}{n}}_{=1} \underbrace{\frac{(n-1)}{n}}_{\lim \to 1} \dots \underbrace{\frac{(n-k+1)}{n}}_{\lim \to 1} \underbrace{\frac{(1-\frac{\lambda}{n})^n}{(1-\frac{\lambda}{n})^k}}_{\lim \to 1} = e^{-\lambda} \frac{\lambda^k}{k!}$$

$\hookrightarrow \lim_{n\to\infty} (1-\frac{\lambda}{n})^n = e^{-\lambda}$ notable limit

Def:

$X \sim Poisson(\lambda)$ poisson r.v. with parameter $\lambda \in (0, +\infty)$

$X = \#$ clients arriving at queue in a unit time

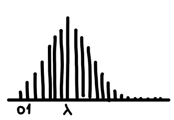
Distribution

$Im(X) = \{0, 1, 2, \dots, k, \dots\} = \mathbb{Z}_+$ $\qquad$ $\lambda = \text{rate of arrival} = \lim_{\Delta t \to 0} \frac{\mathbb{P}(\text{one arrival in } \Delta t)}{\Delta t}$

$\mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$ $\quad k \in \mathbb{Z}_+$

istogram

$X^{(n)} \sim Bin(n; \frac{\lambda}{n})$



Expectation value

$X \sim Poisson(\lambda)$

$$\mathbb{E}(X) = \sum_{k=0}^{\infty} k \mathbb{P}(X=k) = \sum_{k=1}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} \overset{h=k-1}{=} e^{-\lambda} \sum_{h=0}^{\infty} \frac{\lambda^h \lambda}{h!} = \lambda e^{-\lambda} \sum_{h=0}^{\infty} \frac{\lambda^h}{h!} \overset{\text{exponential series}}{=} \lambda$$

infact we knew already:

$X^{(n)} \sim Bin(n; \frac{\lambda}{n}) \to \mathbb{E}(X^{(n)}) = n \cdot \frac{\lambda}{n} = \lambda$ $\quad$ now $n \to \infty \Rightarrow \mathbb{E}(X) = \lambda$

Variance

$\mathbb{V}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2$

$\mathbb{E}(X^2)$ = computed as we did for $\mathbb{E}(X)$

$\mathbb{V}(X^{(n)}) = n\frac{\lambda}{n}\left[1-\frac{\lambda}{n}\right] \xrightarrow[n\to\infty]{} \lambda$

by poisson thm we conclude

$X \sim \text{poisson}(\lambda)$

$\mathbb{V}(X) = \lambda$

Distribution of the sum of two Poisson r.v.

two types of clients (red, blue)

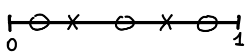
$\left.\begin{array}{l} X_1 = \#\text{red clients} \sim \text{Poisson}(\lambda_1) \\ X_2 = \#\text{blue} \;\;" \;\; \sim \text{Poisson}(\lambda_2) \end{array}\right\}$ independent

• rem: $\mathbb{P}(X+Y=z) = \sum_{x \in Im(X)} \mathbb{P}(X=x)\mathbb{P}(Y=z-x) \quad z \in Im(X+Y)$

X = # clients (either red or blue) $= X_1 + X_2$

$$\mathbb{P}(X=k) = \sum_{h=0}^{\infty} \mathbb{P}(X_1=h)\mathbb{P}(X_2=k-h) \overset{k \geq h}{=} \sum_{h=0}^{\infty} e^{-\lambda_1}\frac{\lambda_1^h}{h!} e^{-\lambda_2}\frac{\lambda_2^{k-h}}{(k-h)!}\frac{k!}{k!} = e^{-(\lambda_1+\lambda_2)} \sum_{h=0}^{\infty} \binom{k}{h} \lambda_1^k \lambda_2^{k-h} \overset{\text{binomial}}{=} \frac{e^{-(\lambda_1+\lambda_2)}}{k!}(\lambda_1+\lambda_2)^k \quad (X \sim \text{Poisson}(\lambda_1+\lambda_2))$$

$\overset{k<h}{=} 0$



O = arrival of red clients
X = " " blue clients

how a daltonic person sees the picture?
clients arriving with rate $\lambda_1+\lambda_2$

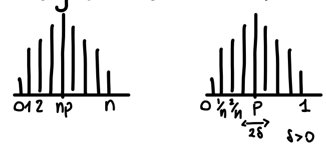
Law of large numbers 28.11

· Bernoulli scheme

$\Omega = \{(w_1, \dots, w_n) \quad w_i = 0,1\} \quad p = \text{Prob}(H) \in [0,1]$

$S_n = \#H = \sum_{i=1}^{n} w_i \sim Bin(n,p) \quad n \to \infty$ (with p fixed)

-> istogram distribution of S_n



$\frac{S_n}{n}$ = fraction of #H in n tosses

$0 \leq \frac{S_n}{n} \leq 1$

$\frac{S_n}{n} \to p$ as $n \to \infty$

Rem: $S_n = X_1 + \dots + X_n$ where $X_i \sim Bern(p)$ independent

$\mathbb{P}\left(p-\delta < \frac{S_n}{n} < p+\delta\right) \xrightarrow{n\to\infty} 1$

Thm: $\forall \delta > 0 \quad \lim_{n\to\infty} \mathbb{P}\left(p-\delta < \frac{S_n}{n} < p+\delta\right) = 1$

General setting

let $X_1, X_2, \dots, X_n$ be independent and identically distributed r.v.

$\mu = \mathbb{E}(X_1) = \mathbb{E}(X_2) = \dots = \mathbb{E}(X_n)$

$\sigma^2 = \mathbb{V}(X_1) = \mathbb{V}(X_2) = \dots = \mathbb{V}(X_n)$

$S_n := X_1 + X_2 + \dots + X_n = \sum_{i=1}^{n} X_i$

Theorem

$\frac{S_n}{n} \xrightarrow[n\to\infty]{} \mu$ in the following sense $\forall \delta > 0 \quad \lim_{n\to\infty} \mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| < \delta\right) = 1$

Proof: based on two ingredients

-> computation of $\mathbb{E}\left(\frac{S_n}{n}\right)$ and $\mathbb{V}\left(\frac{S_n}{n}\right)$

• $\mathbb{E}\left(\frac{S_n}{n}\right) \underset{\text{def}}{=} \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \underset{\text{linearity}}{=} \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_i) \underset{\text{identically distributed}}{=} \frac{1}{n}\sum_{i=1}^{n}\mu = \mu$

• $\mathbb{V}\left(\frac{S_n}{n}\right) \underset{\substack{\text{variance is}\\ \text{quadratic}}}{=} \frac{1}{n^2}\mathbb{V}(S_n) = \frac{1}{n^2}\mathbb{V}\left(\sum_{i=1}^{n} X_i\right) \underset{\text{independence}}{=} \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}(X_i) \underset{\text{identically distributed}}{=} \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{1}{n}\sigma^2$

in particular: $\mathbb{V}\left(\frac{S_n}{n}\right) \xrightarrow[n\to\infty]{} 0$

-> quantitative bound

if a r.v. X has small variance then its distribution is concentrated around the expectation value

Lemma 1 Chebyshev's inequality

let X be a r.v. and $\lambda > 0$ then $\mathbb{P}(|X - \mathbb{E}(X)| \geq \lambda) \leq \frac{1}{\lambda^2} \mathbb{V}(X)$

need to show $\forall \delta > 0$ $\mathbb{P}(|\frac{S_n}{n} - \mu| \geq \delta) \xrightarrow[n \to \infty]{} 0$ we know that $\mu = \mathbb{E}(\frac{S_n}{n})$ so by the lemma 1

$\mathbb{P}(|\frac{S_n}{n} - \mathbb{E}(\frac{S_n}{n})| \geq \delta) \leq \frac{1}{\delta^2} \mathbb{V}(\frac{S_n}{n}) = \frac{1}{\delta^2} \frac{\sigma^2}{n} \xrightarrow[n \to \infty]{} 0$

Lemma 2 Markov's inequality

let Y be a positive r.v., $Y \geq 0$, and $\lambda > 0$ then $\mathbb{P}(Y \geq \lambda) \leq \frac{1}{\lambda} \mathbb{E}(Y)$

Proof 2:

$$\mathbb{P}(Y \geq \lambda) = \sum_{\substack{y \in Im(Y): \\ y \geq \lambda}} \mathbb{P}(Y = y) \underset{\frac{y}{\lambda} \geq 1}{\leq} \sum_{\substack{y \in Im(Y): \\ y \geq \lambda}} \frac{y}{\lambda} \mathbb{P}(Y = y) \leq \frac{1}{\lambda} \sum_{y \in Im(Y)} y\, \mathbb{P}(Y = y) = \frac{1}{\lambda} \mathbb{E}(Y)$$

Proof 1:

set $Y = [X - \mathbb{E}(X)]^2$ $Y \geq 0$

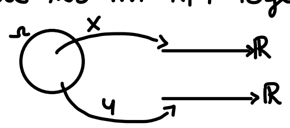
$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \lambda) \underset{\text{both sides positive}}{=} \mathbb{P}(|X - \mathbb{E}(X)|^2 \geq \lambda^2) = \mathbb{P}(Y \geq \lambda^2) \underset{\text{Lemma 2}}{\leq} \frac{1}{\lambda^2} \mathbb{E}(Y) = \frac{1}{\lambda^2} \mathbb{E}([X - \mathbb{E}(X)]^2) = \frac{1}{\lambda^2} \mathbb{V}(X)$$

Joint distribution of random variables

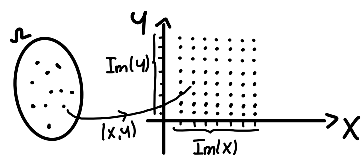
Recall: for a single r.v. the distribution of X is the probability on $Im(X)$ defined by

$\mu_X(\{x\}) = \mathbb{P}(X = x) = \mathbb{P}(\{\omega \in \Omega : X(\omega) = x\})$ $x \in Im(X)$

now two r.v. X, Y together



$(X, Y) : \Omega \to \mathbb{R} \times \mathbb{R} = \mathbb{R}^2$



Def: the joint distribution of X, Y is the probability on $Im(X) \times Im(Y)$ defined by $\mu_{X,Y}(\{(x,y)\}) = \mathbb{P}(X = x, Y = y) = \mathbb{P}(\{\omega \in \Omega : X(\omega) = x, Y(\omega) = y\})$ $x \in Im(X)$, $y \in Im(Y)$

ex.

toss a fair dice $\Omega = \{1, 2, \dots, 6\}$ with uniform probability

$$X = \begin{cases} -1 & \omega = 1,2 \\ 0 & \omega = 3,4 \\ 1 & \omega = 5,6 \end{cases} \qquad Y = \begin{cases} -2 & \omega = 1,2,3 \\ 2 & \omega = 4,5,6 \end{cases}$$

joint distribution of X and Y

Y \ X	-1	0	1	
-2	2/6	1/6	0	3/6
2	0	1/6	2/6	3/6
	2/6	2/6	2/6	

the contents of the cells

$\mathbb{P}(X = -1, Y = -2) = \mathbb{P}(\{1, 2\}) = 2/6$ $\mathbb{P}(X = 1, Y = 2) = \mathbb{P}(\{5, 6\}) = 2/6$

$\mathbb{P}(X = 0, Y = -2) = \mathbb{P}(\{3\}) = 1/6$ $\mathbb{P}(X = 0, Y = 2) = \mathbb{P}(\{4\}) = 1/6$

$\mathbb{P}(X = 1, Y = -2) = \mathbb{P}(X = -1, Y = 2) = 0$

How to obtain the marginal distribution of X or Y from the joint distribution of X and Y?

-> $\mathbb{P}(X = x) = \sum_{y \in Im(Y)} \mathbb{P}(X = x, Y = y)$ $\forall x \in Im(X)$

-> $\mathbb{P}(Y = y) = \sum_{x \in Im(X)} \mathbb{P}(X = x, Y = y)$ $\forall y \in Im(Y)$

Conditional distribution

Bob (bet Y) does not know the outcome of the dice but knows the outcome X (bet by Alice) upon this information Bob recomputes his victory/loss probability

-> if $X=-1$:
$\mathbb{P}(Y=-2|X=-1)=1$
$\mathbb{P}(Y=2|X=-1)=0$

-> if $X=0$:
$\mathbb{P}(Y=-2|X=0)=1/2$
$\mathbb{P}(Y=2|X=0)=1/2$

-> if $X=1$:
$\mathbb{P}(Y=-2|X=1)=0$
$\mathbb{P}(Y=2|X=1)=1$

Def: fix $x\in Im(X)$ the conditional probability of Y given $X=x$ is

$$\mu_{Y|X}(\{y\}|x)=\mathbb{P}(Y=y|X=x)=\frac{\mathbb{P}(X=x,Y=y)}{\mathbb{P}(X=x)}=\frac{\mu_{X,Y}(\{(x,y)\})}{\mu_X(\{x\})}\quad y\in Im(Y)$$

in the same way given $y\in Im(Y)$ the conditional probability of X given $Y=y$ is

$$\mu_{X|Y}(\{x\}|y)=\mathbb{P}(X=x|Y=y)=\frac{\mathbb{P}(X=x,Y=y)}{\mathbb{P}(Y=y)}=\frac{\mu_{X,Y}(\{(x,y)\})}{\mu_Y(\{y\})}\quad x\in Im(X)$$

Conditional expectation

fix $x\in Im(X)$, $\mathbb{E}(Y|X=x)$ is the expectation of Y according to the conditional distribution

$$\mathbb{E}(Y|X=x)=\sum_{y\in Im(Y)} y\,\mathbb{P}(Y=y|X=x)$$

Multinomial random variable

an experiment with k possible outcomes, $1,2,3,\dots k$, with $k\geq 2$, that occur with probabilities $p_1,p_2\dots p_k$ $[p_1+p_2+\dots+p_k=1]$. Repeat independently n times $X_1=\#1, X_2=\#2,\dots, X_k=\#k$ $[X_1+X_2+\dots+X_k=n]$

$(X_1,X_2,\dots,X_k)\sim Multi(n;p_1,p_2,\dots p_k)$

joint distribution of $(X_1,\dots,X_k)$

$Im(X_1)=\{0,1,\dots,n\}\dots\ Im(X_k)=\{0,1\dots,n\}$

$$\mathbb{P}(X_1=n_1,X_2=n_2,\dots,X_k=n_k)=\frac{n!}{n_1!n_2!\dots n_k!}p_1^{n_1}p_2^{n_2}\dots p_k^{n_k}\quad \text{when } n_1+n_2+\dots+n_k=n$$

Conditional distribution of X_3 given $X_1=n_1$

answers the following question: toss a fair dice 10 times. we know $\#1=3$. What is the distribution of $\#3$?

$$\mathbb{P}(X_3=n_3|X_1=n_1)=\frac{\mathbb{P}(X_1=n_1,X_3=n_3)}{\mathbb{P}(X_1=n_1)}=\frac{\mathbb{P}(X_1=n_1,X_2=n-n_1-n_3,X_3=n_3)}{\mathbb{P}(X_1=n_1)}=\frac{\frac{\cancel{n!}}{\cancel{n_1!}(n-n_1-n_3)!\,n_3!}\cancel{p_1^{n_1}}p_2^{n-n_1-n_3}p_3^{n_3}}{\frac{\cancel{n!}}{\cancel{n_1!}(n-n_1)!}\cancel{p_1^{n_1}}(p_2+p_3)^{n-n_1}}=\frac{\binom{n-n_1}{n_3}p_2^{n-n_1-n_3}p_3^{n_3}}{(p_2+p_3)^{n-n_1+n_3-n_3}}$$

$$=\binom{n-n_1}{n_3}\left(\frac{p_3}{p_2+p_3}\right)^{n_3}\left(\frac{p_2}{p_2+p_3}\right)^{n-n_1-n_3}$$

conditionally on $X_1=n_1$ $X_3\sim Bin\left(n-n_1,\frac{p_3}{p_2+p_3}\right)$ infact $\underbrace{\frac{p_3}{p_2+p_3}}_{\text{in a single toss}}=\mathbb{P}(3|\text{not }1)$

answer to question is $Bin(7,1/2)$

Continuous random variable 5.12

ex. call your random number generator, it outputs a pseudo random number in the interval $[0,1]$

$X \sim Unif([0,1])$ $\mathbb{P}(X = \frac{1}{\sqrt{2}}) = 0$

$\forall x \in [0,1]$ $\mathbb{P}(X = x) = 0$

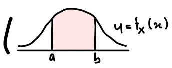
X is infact a continuous r.v.

Up to now what we have seen are discrete r.v. that is $Im(X)$ is either finite $(Im(X) = \{0,1,\dots,n\})$ or countable $(Im(X) = \{0,1,\dots n,\dots\})$. In this case $Im(X) = [0,1]$ is not countable.

A relevant question is $\mathbb{P}(a \leq x \leq b)$ with $a<b$. $\mathbb{P}(X \in [0,\frac{1}{2}]) = \frac{1}{2}$, $\mathbb{P}(X \in [\frac{3}{8},\frac{4}{8}]) = \mathbb{P}(X \in [0,\frac{1}{8}]) = \frac{1}{8}$

Let x be a continuous r.v. ($Im(X)$ is not countable), its distribution is specified by the probability density, that is $\mathbb{P}(X \in [x, x+\Delta x]) = f_X(x)\Delta x + o(\Delta x)$

Probability density

$f_X : \mathbb{R} \to [0;+\infty)$ so that $a<b$: $\mathbb{P}(a \leq X \leq b) = \int_a^b f_X(x)\,dx = Area($  $)$

$\mathbb{P}(-\infty < X < +\infty) = \int_{-\infty}^{+\infty} f_X(x)\,dx = 1$

ex. $X \sim Unif([0,1])$

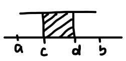
its density $\to f_X(x) = \begin{cases} 1 & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$

more generally:

pick $a<b$, $X \sim Unif([a,b])$

its density is $f_X(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$

so that given $a<c<d<b$

$\mathbb{P}(c \leq X \leq d) = Area($  $) = \frac{(d-c)}{b-a}$

Expectation of a continuous r.v.

we approximate it by a discrete one

$\mathbb{E}(X) \approx \sum_i x_i \underbrace{\mathbb{P}(X \in [x_i, x_i+\Delta x])}_{f_X(x_i)\Delta x} \xrightarrow[\Delta x \to 0]{} \int_{-\infty}^{\infty} x f_X(x)\,dx$

$\to \mathbb{E}(X) = \int_{-\infty}^{\infty} x f_X(x)\,dx$

ex. $X \sim Unif([0,1])$ $f_X(x) = \begin{cases} 1 & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$

$\mathbb{E}(X) = \frac{1}{2}$ by symmetry

$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f_X(x)dx = \int_0^1 x \, 1 \, dx = \frac{x^2}{2}\Big|_0^1 = \frac{1}{2} - 0 = \frac{1}{2}$

in general: $X \sim Unif([a,b])$ $f_X = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$

$\mathbb{E}(X) = \frac{a+b}{2}$ by symmetry

$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f_X(x)\,dx = \int_a^b x \frac{1}{b-a}\,dx = \frac{1}{b-a} \frac{x^2}{2}\Big|_a^b = \frac{1}{b-a}(\frac{b^2}{2} - \frac{a^2}{2})$

$= \frac{1}{2}\frac{b^2-a^2}{b-a} = \frac{a+b}{2}$

Let $Y \sim Unif([0,1])$, set $X = a + (b-a)Y$ then $X \sim Unif([a,b])$

Variance of a continuous r.v.

$\mathbb{V}(X) = \mathbb{E}([x - \mathbb{E}(X)]^2)$ if X is continuous $\mathbb{V}(X) = \int_{-\infty}^{\infty} [x - \mathbb{E}(X)]^2 f_X(x)\,dx$

$\mathbb{V}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2$ if X is continuous $\mathbb{E}(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x)\,dx$

ex. $X \sim Unif([0,1])$

$\mathbb{E}(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x)\,dx = \int_0^1 x^2 \cdot 1 \, dx = \frac{x^3}{3}\Big|_0^1 = \frac{1}{3}$

$\mathbb{V}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 = \frac{1}{3} - (\frac{1}{2})^2 = \frac{1}{12}$

in general: $X \sim Unif([a,b])$

$X = a + (b-a)Y$ where $Y \sim Unif([0,1])$

$\mathbb{V}(X) = \mathbb{V}(a + (b-a)Y) = \mathbb{V}((b-a)Y) = (b-a)^2\mathbb{V}(Y) = \frac{(b-a)^2}{12}$

Distribution function

X be a r.v. (either discrete or continuous)

$F_X(x) = \mathbb{P}(X \le x) \quad x \in \mathbb{R}$

it identifies the distribution of the r.v. X

$a<b \to \mathbb{P}(a<X\le b) = \mathbb{P}(X \le b) - \mathbb{P}(X \le a) = F_X(b) - F_X(a)$

if X is discrete, say $\mathrm{Im}(X) = \{x_1, \dots, x_n\}$ with $p_1 = \mathbb{P}(X=x_1), p_2 = \mathbb{P}(X=x_2), \dots, p_n = \mathbb{P}(X=x_n)$ then

$$F_X(x) = \begin{cases} 0 & x<x_1 \\ p_1 & x_1 \le x < x_2 \\ p_1+p_2 & x_2 \le x < x_3 \\ \vdots & \\ p_1+\dots+p_{n-1} & x_{n-1} \le x < x_n \\ 1 & x \ge x_n \end{cases}$$

F_X is piece wise constant, increasing, with jumps at the points x_i $i=1,\dots,n$ and the value of the jump at point x_i is $p_i = \mathbb{P}(X=x_i)$

if X is continuous, with density f_X

$F_X(x) = \mathbb{P}(X \le x) = \text{Area}(\ \ ) = \int_{-\infty}^{x} f_X(y)\,dy$

in particular, $\frac{d}{dx} F_X(x) = f_X(x)$

ex.

let X, Y be independent r.v. (either cont. or discrete), set $Z = \max\{X,Y\}$, find the distribution of Z

we compute the distribution function of Z, $z \in \mathbb{R}$

$$F_Z(z) = \mathbb{P}(Z \le z) = \mathbb{P}(\max\{X,Y\} \le z) = \mathbb{P}(X \le z, Y \le z) \overset{\text{independence}}{=} \mathbb{P}(X \le z)\mathbb{P}(Y \le z) = F_X(z)F_Y(z)$$

→ we found $F_{\max\{X,Y\}}(z) = F_X(z)F_Y(z)$ if X,Y independent

sum of independent r.v.

let X,Y be independent r.v., set $Z = X+Y$, find distribution of Z

if X,Y are discrete $\to \mathbb{P}(Z=z) = \sum_{x \in \mathrm{Im}(X)} \mathbb{P}(X=x)\mathbb{P}(Y=z-x)$

if X,Y are continuous $\to f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\,dx$

ex. $X,Y \sim \mathrm{Unif}([0,1])$ independent, $Z = X+Y$, compute the probability density of Z

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\,dx = \int_0^1 1 \cdot f_Y(z-x)\,dx$$

$\llcorner\ f_Y(z-x) = 1 \iff 0 \le z-x \le 1 \iff \begin{cases} x \le z \\ x \ge z-1 \end{cases}$

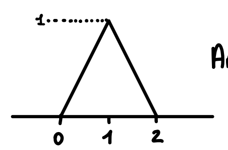
1. $0 \le z \le 1$ then $x \ge z-1$ always

$\int_0^1 f_Y(z-x)\,dx = \int_0^z 1\,dx = z$

2. $1 \le z \le 2$ then $x \le z$ always

$\int_0^1 f_Y(z-x)\,dx = \int_{z-1}^{1} 1\,dx = 2-z$

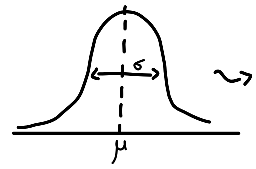
$$f_Z(z) = \begin{cases} 0 & z<0 \\ z & 0 \le z \le 1 \\ 2-z & 1 \le z \le 2 \\ 0 & z>2 \end{cases}$$



$\text{Area}(\triangle) = 1$

Gaussian (or normal) random variables 12.12

- theory of errors in measurement -> measure a physical observable, repeat n times
- -> random fluctuations due either to intrinsic fluctuation, or to measurement apparatus is not totally precise
- -> istogram of empirical data



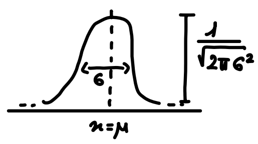
~> bell shape

μ = "real" value of the observal (istogram symmetric w.r.t. μ)
σ = "typical" value of fluctuations

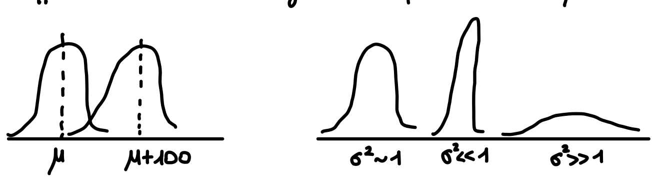
-> Gauss proposed a formula for the curve of the istogram
X is the r.v. describing the measured value, continuous with probability density

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right\}$$

-> plot the graph



-> what happens if we change the parameters μ and σ^2?



$\mathbb{E}(X) = \mu$
$\mathbb{V}(X) = \sigma^2$

-> f_X is a probability density?

1. $f_X \geq 0$ —> exponential is bigger than zero
2. $\int_{-\infty}^{\infty} f_X(x)\, dx = \text{Area}(\ ) = 1$

$\hookrightarrow \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz \underset{\text{Gauss}}{=} 1$

$z = \frac{x-\mu}{\sigma}$ $\quad \sigma = \sqrt{\sigma^2}$
$dz = \frac{dx}{\sigma}$ $\quad z^2 = \frac{(x-\mu)^2}{\sigma^2}$

Area under graph does not depend on parameters

-> $X \sim \mathcal{N}(\mu, \sigma^2)$; X is a Normal (Gaussian) r.v. with parameters μ, σ^2

Expectation

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f_X(x)\, dx = \int_{-\infty}^{\infty} \underset{\substack{\text{odd w.r.t.}\\ x=\mu}}{x} \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}}_{\text{even w.r.t. } x=\mu} dx \underset{\text{by symmetry}}{=} \mu$$

$$\underset{\text{by change of variable}}{=} \int_{-\infty}^{\infty} (\mu + \sigma z) \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz = \mu + \sigma \underbrace{\int_{-\infty}^{\infty} z \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz}_{=0 \text{ by symmetry}} = \mu$$

Variance

$$\mathbb{V}(X) = \mathbb{E}([X - \mathbb{E}(X)]^2) = \int_{-\infty}^{\infty} [x-\mu]^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} dx = \int_{-\infty}^{\infty} \sigma^2 z^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz = \sigma^2 \int_{-\infty}^{\infty} z^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz$$

it remains: $\int_{-\infty}^{\infty} z^2 \frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}} dz = 1$ $\qquad \frac{d}{dz}\left(\frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}}\right) = \frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}}(-z)$

$$= \int_{-\infty}^{\infty} z \cdot \frac{z e^{-z^2/2}}{\sqrt{2\pi}} dz = -\int_{-\infty}^{\infty} z \frac{d}{dz}\left(\frac{e^{-z^2/2}}{\sqrt{2\pi}}\right) dz \underset{\text{by parts}}{=} -z \frac{e^{-\frac{1}{2}z^2}}{\sqrt{2\pi}}\Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \frac{d}{dz}(z) \frac{e^{-\frac{1}{2}z^2}}{\sqrt{2\pi}} = 0 + \int_{-\infty}^{\infty} 1 \frac{e^{-\frac{1}{2}z^2}}{\sqrt{2\pi}} dz \underset{\text{gauss}}{=} 1$$

$\sigma = \sqrt{\sigma^2}$ => standard deviation

Probabilities

fix $a<b$

$\mathbb{P}(a<X<b) = \text{Area}(\ \ ) = \int_a^b \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right\} dx$

reduce to the case $\mu=0$, $\sigma^2=1$:

set $Z = \frac{X-\mu}{\sigma}$ then by the change of variable $Z \sim \mathcal{N}(0,1)$ => standard Gauss r.v.

Standardization of gauss r.v.

$\mathbb{P}(a<X<b) = \mathbb{P}(a-\mu < X-\mu < b-\mu) = \mathbb{P}\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right) = \mathbb{P}\left(Z < \frac{b-\mu}{\sigma}\right) - \mathbb{P}\left(Z < \frac{a-\mu}{\sigma}\right) = \phi\left(\frac{b-\mu}{\sigma}\right) - \phi\left(\frac{a-\mu}{\sigma}\right)$

$Z \sim \mathcal{N}(0,1)$

Table of Gauss integral

given $z \in \mathbb{R}$ it gives the numerical value of $\mathbb{P}(Z \le z) = \text{Area}(\ \ ) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy =: \phi(z)$

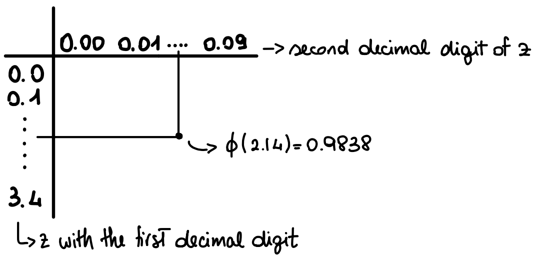


Table 1 Area $\Phi(x)$ Under the Standard Normal Curve to the Left of X.

X	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

ex. let $Z \sim \mathcal{N}(0,1)$, compute using table of gaussian integral

$\mathbb{P}(-2<Z<0.5) = \text{Area}(\ \ )_{-2\ \ 0.5} = \int_{-2}^{0.5} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz = \underbrace{\phi(0.5)}_{\cong 0.6915} - \phi(-2) \underset{\text{table}}{=} 0.6915 - (1-0.9772) \cong 0.67$

$\phi(-2)$: not in table, do the symmetry

↳ $\text{Area}(\ \ ) = 1 - \text{Area}(\ \ ) = 1 - \phi(2) = 1 - 0.9772$

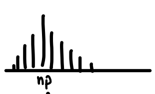
Central limit theorem

-> explains the reason of universality of Gaussian r.v.

Bernoulli scheme, n tosses of biased coin, $S_n = \#\text{heads} \sim Bin(n,p)$

istogram of the distribution:

Law of large #: $\mathbb{P}(n(p-\delta) \le S_n \le n(p+\delta)) \xrightarrow[n\to\infty]{} 1$ given $\delta>0$



-> looks like a gaussian density for n large

-> which gaussian? $\mu = \mathbb{E}(S_n) = np$; $\sigma^2 = \mathbb{V}(S_n) = np(1-p)$

crucial ingredient: S_n is sum of independent r.v.

-> general statement

$X_1, X_2, \dots, X_n$ independent and identically distributed r.v. ($X_i \sim Bern(p)$ in the Binomial case)

let $\mu = \mathbb{E}(X_i)$ and $\sigma^2 = \mathbb{V}(X_i)$ do not depend on i because X_i are identically distributed

$S_n = X_1 + X_2 + \dots + X_n = \sum_{i=1}^{n} X_i$ $\quad \underbrace{\mathbb{E}(S_n) = n\mu}_{\text{linearity of } \mathbb{E}(X)}$ $\quad \underbrace{\mathbb{V}(S_n) = n\sigma^2}_{\text{independence}}$

Thm:

As $n \to \infty$ $\frac{S_n - n\mu}{\sigma\sqrt{n}}$ approaches $\mathcal{N}(0,1)$ in the sense that given $a<b$:

$\lim_{n\to\infty} \mathbb{P}\left(a \le \frac{S_n - n\mu}{\sigma\sqrt{n}} \le b\right) = \mathbb{P}(a<Z<b)$ with $Z \sim \mathcal{N}(0,1)$

-> in the previous example $X_i \sim Bern(p)$: $\mathbb{P}\left(a \le \frac{S_n - np}{p(1-p)\sqrt{n}} \le b\right) \approx \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz$

ex. toss a coin and want to be sure that its fair, we toss it 10^4 times and we get 5500 heads, we are skeptical that is fair, seller says its typical fluctuations. To test it, assume coin is fair,

$S_n = \#\text{heads} \sim Bin(n, \frac{1}{2}) \quad n = 10^4$

$\mathbb{P}(|S_n - n\frac{1}{2}| \geq 500)$ if too small we refuse the hypothesis that the coin is fair, if not too small, the data are actually compatible with a fair coin.

By chebyshev inequality: $\mathbb{P}(|S_n - n\frac{1}{2}| \geq 500) \leq \frac{1}{(500)^2} V(S_n) = \frac{1}{5^2 10^4} 10^4 \frac{1}{4} = \frac{1}{25} \cdot \frac{1}{4} = \frac{1}{100} = 1\%$

by chance 1% is the confidence level of the guarantee of the seller

To be more accurate we use the central limit thm: $\mathbb{P}(|S_n - n\frac{1}{2}| \geq 500) = \mathbb{P}\left(\frac{|S_n - n\frac{1}{2}|}{\sqrt{V(S_n)}} \geq \frac{500}{\sqrt{V(S_n)}}\right) =$

$= \mathbb{P}\left(\frac{|S_n - n\frac{1}{2}|}{\sqrt{V(S_n)}} \geq \frac{500}{\frac{1}{2}10^2}\right) = \mathbb{P}\left(\frac{|S_n - n\frac{1}{2}|}{\sqrt{V(S_n)}} \geq 10\right) \approx \mathbb{P}(|Z| > 10) \sim 10^{-6}$

↳ gaussian approximation

How to simulate a random variable? 19.12

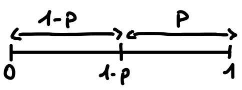
Aim: by using the r.n.g. we can construct an arbitrary r.v.

we have $U \sim Unif(0,1)$

Simple examples

→ let $X \sim Bern(p)$ $\quad X = \begin{cases} 0 & \text{with prob. } 1-p \\ 1 & \text{" " } p \end{cases}$

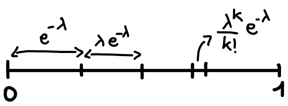
how to "simulate" X by using U?



set $X = \begin{cases} 0 & \text{if } 0 < U < 1-p \\ 1 & \text{if } 1-p < U < 1 \end{cases}$ $\quad \mathbb{P}(X=0) = \mathbb{P}(0 < U < 1-p) = 1-p$

$\mathbb{P}(X=1) = \mathbb{P}(1-p < U < 1) = p$

→ let $X \sim Poisson(\lambda)$ $\quad Im(X) = \mathbb{Z}_+ \quad \mathbb{P}(X=k) = e^{-\lambda}\frac{\lambda^k}{k!} \quad k \in \mathbb{Z}_+$



$\mathbb{P}(X=0) = e^{-\lambda}$

$\mathbb{P}(X=1) = e^{-\lambda}\lambda$

$\vdots$

$\mathbb{P}(X=k) = e^{-\lambda}\frac{\lambda^k}{k!}$

$\phi : (0,1) \to \mathbb{Z}_+$

$$\phi(u) = \begin{cases} 0 & 0 < u \leq e^{-\lambda} \\ 1 & e^{-\lambda} < u \leq \lambda e^{-\lambda} \\ 2 & e^{-\lambda}(1+\lambda) < u < e^{-\lambda}(1+\lambda+\frac{\lambda^2}{2}) \\ \vdots & \\ k & e^{-\lambda}(1+\lambda+\ldots+\frac{\lambda^{k-1}}{(k-1)!}) < u \leq e^{-\lambda}(1+\lambda+\ldots+\frac{\lambda^k}{k!}) \\ \vdots & \end{cases}$$

setting $X = \phi(U)$ then $X \sim Poisson(\lambda)$

$\mathbb{P}(X=0) = \mathbb{P}(\phi(U)=0) = \mathbb{P}(0 < U \leq e^{-\lambda}) = e^{-\lambda}$

$\vdots$

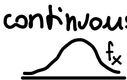
$\mathbb{P}(X=k) = \mathbb{P}(\phi(U)=k) = \mathbb{P}(e^{-\lambda}(1+\ldots+\frac{\lambda^{k-1}}{(k-1)!}) < U \leq e^{-\lambda}(1+\ldots+\frac{\lambda^k}{k!})) = e^{-\lambda}\frac{\lambda^k}{k!}$

Thm: let X be an arbitrary r.v. and $U \sim Unif(0,1)$ there exists a function $\phi_X : (0,1) \to \mathbb{R}$ (depending on the r.v. X) such that $\phi(U) \sim X$ (has the same distribution)

Morale: in the r.n.g. there is all the randomness you'll ever need

Rem: the ϕ_X will be explicitly constructed

Proof: discrete r.v. in the previous examples, so we focus on continuous r.v.

let X be a continuous r.v. with probability density $f_X > 0$ 

let F_X be the distribution function $F_X : \mathbb{R} \to (0,1)$ and defined by $F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f_X(y)\,dy$

since $f_X > 0$ in particular F_X is a bijection from (0,1) to $\mathbb{R}$, since also F_X is stricly increasing

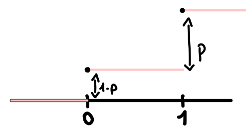
set $\phi : (0,1) \to \mathbb{R}$ $\quad \phi_X := (F_X)^{-1}$ need to show $\phi_X(U) \sim X$ for this we show that the distribution function of $\phi(U)$ is F_X, fix $x \in \mathbb{R}$

→ $\underbrace{\mathbb{P}(\phi_X(U) \leq x) = \mathbb{P}((F_X)^{-1}(U) \leq x)}_{\text{by definition}} \underbrace{=}_{F_X \text{ strictly increasing}} \mathbb{P}([F_X \circ (F_X)^{-1}](U) \leq F_X(x)) = \mathbb{P}(U \leq F_X(x)) \underbrace{=}_{U \sim Unif(0,1)} F_X(x)$

infact the examples of discrete r.v. discussed before are special cases

-> $X \sim Bern(p)$

$$\phi(u) = \begin{cases} 0 & 0 < u \leq 1-p \\ 1 & 1-p < u < 1 \end{cases}$$



$F_X(x) = \mathbb{P}(X \leq x)$

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1-p & 0 \leq x < 1 \\ p & x \geq 1 \end{cases}$$

- can put because $\mathbb{P}(X = x)$ for r.n.g. is 0

-> in a fair coin there is enough randomness to simulate any r.v.

it suffices to simulate $U \sim Unif(0,1)$

toss the coin ∞ times $X_i = \begin{cases} 0 & \text{"T" with probability } \frac{1}{2} \\ 1 & \text{"H"} \quad " \quad " \quad \frac{1}{2} \end{cases}$ independent

produce $U \in (0,1)$ by letting X_i's be the digits of the expansion in base 2 of X

in other words set $U = \sum_{i=1}^{\infty} \frac{1}{2^i} X_i$

claim: $U \sim Unif(0,1)$

enough to show that if $x \in (0,1)$ $\mathbb{P}(U \leq x) = x$

for $x = \frac{1}{2}$ $\mathbb{P}(\sum_{i=1}^{\infty} \frac{1}{2^i} X_i < \frac{1}{2}) = \mathbb{P}(X_1 = 0) = \frac{1}{2}$

for $x = \frac{3}{4}$ $\mathbb{P}(\sum_{i=1}^{\infty} \frac{1}{2^i} X_i \leq \frac{3}{4}) = \mathbb{P}(10\ldots \cup 01\ldots \cup 00\ldots) = 1 - \mathbb{P}(\sum_{i=1}^{\infty} \frac{1}{2^i} X_i > \frac{3}{4}) = 1 - \mathbb{P}(X_1 = X_2 = 1) = 1 - \frac{1}{4} = \frac{3}{4}$