Probability IA (Lent 2020)

Please report errors and comments to Perla Sousi (`ps422@cam.ac.uk`).

March 20, 2020

Probability theory is the mathematical formulation of randomness.

Examples of random experiments:

throw a die, pick a ball from a bag, shuffle a deck of cards, ...

Need to develop a mathematical framework to study randomness.

<u>Def.</u> <u>Probability space</u>

Let Ω be a set, $\mathcal{F}$ a set of subsets of Ω. We call $\mathcal{F}$ a <u>σ-algebra</u> if

- $\Omega \in \mathcal{F}$
- if $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$
- for every countable sequence

$(A_n)_{n\geq 1}$ in $\mathcal{F}$, also $\bigcup_{n\geq 1} A_n \in \mathcal{F}$.
$(A_n \in \mathcal{F}, \forall n)$,

Suppose $\mathcal{F}$ is indeed a σ-algebra.

A function $\mathbb{P} : \mathcal{F} \to [0,1]$ is called a probability measure if

- $\mathbb{P}(\Omega) = 1$
- for every sequence of disjoint sets $(A_n)_{n\geq 1}$, $\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n)$

We call $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space.

<u>Remark</u> When Ω is countable, we take $\mathcal{F}$ to be all subsets of Ω.

<u>Def.</u> The elements of Ω are called outcomes and the elements of $\mathcal{F}$

are called events.

If $A \in \mathcal{F}$, then we interpret $\mathbb{P}(A)$ as the probab. of the event A.

We talk about probab. of events and not of outcomes.

Later we will see that if we pick a uniform point from $[0,1]$, then any number in $[0,1]$ has 0 prob.

Properties of $\mathbb{P}$ (easy to check from def)

- $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
- $\mathbb{P}(\emptyset) = 0$
- if $A \subseteq B$, then $\mathbb{P}(A) \leq \mathbb{P}(B)$
- $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

Examples of prob. spaces

1) Rolling a fair die

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

$\mathcal{F}$ = all subsets of Ω.

$\mathbb{P}(\{\omega\}) = \frac{1}{6}$ and $A \subseteq \Omega$
$\omega \in \Omega$ $\qquad \mathbb{P}(A) = \frac{|A|}{6}$.

because all outcomes are equally likely.

2) <u>Equally likely outcomes</u>

Let Ω be a finite set

$\Omega = \{\omega_1, \dots, \omega_n\}$, $\mathcal{F}$ = all subsets.

Define $\mathbb{P} : \mathcal{F} \to [0,1]$ via

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|}.$$

In classical prob. this is a model for a randomly chosen point of Ω.

Indeed, $\mathbb{P}(\{\omega\}) = \frac{1}{|\Omega|}$, $\omega \in \Omega$

so all outcomes are equally likely.

3) Balls from a bag

Suppose we have a bag with n labelled balls $\{1, \dots, n\}$, indistinguishable by touch.

We pick $k \leq n$ balls at once at random without looking. By saying at random we mean all outcomes are equally likely. So we take

$$\Omega = \{\text{sets of size } k, \text{ subsets of } \{1, \dots, n\}\}$$

$$|\Omega| = \binom{n}{k}.$$

$$\mathbb{P}(\{\omega\}) = \frac{1}{|\Omega|}.$$
$\omega \in \Omega$

4) Deck of cards

Suppose we have a well-shuffled deck of 52 cards.

well-shuffled = all possible orderings are equally likely.

$\Omega = \{\text{permutations of } 52 \text{ elements}\}$
label cards $\{1, \dots, 52\}$

$|\Omega| = 52!$

$$\mathbb{P}(\text{top 2 cards are aces}) = \frac{\overset{\to 1^{st}\text{ card}}{4} \times \overset{\to 2^{nd}}{3} \times 50!}{52!} = \frac{1}{221}$$

5) Largest digit

Consider a string of random digits $0, 1, \dots, 9$ of length n.

Take $\Omega = \{0, 1, \dots, 9\}^n$, $|\Omega| = 10^n$.

"random digits" = all outcomes are equally likely.

Let $A_k = \{\text{no digit exceeds } k\}$ and

$B_k = \{\text{largest digit is } k\}$.

$|A_k| = (k+1)^n$

$B_k = A_k \setminus A_{k-1} \Rightarrow |B_k| = (k+1)^n - k^n$.

So $\mathbb{P}(B_k) = \dfrac{|B_k|}{|\Omega|} = \dfrac{(k+1)^n - k^n}{10^n}$.

6) <u>Birthday problem</u>

Suppose there are n people in the room. What is the probab. that at least 2 share the same birthday?

Assume nobody is born on the 29^{th} Feb.

So $\Omega = \{1, \dots, 365\}^n$.

Assume all outcomes are equally likely.

Let $A = \{\text{all } n \text{ birthdays are different}\}$

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{365 \times 364 \times \dots \times (365-n+1)}{365^n}$$

So $\mathbb{P}(\text{at least 2 share same birthday})$

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

$n = 22 \rightsquigarrow \mathbb{P}(2 \rightsquigarrow \text{same birthday}) \approx 0.476$

$n = 23 \rightsquigarrow$ — " — ≈ 0.507

Combinatorial analysis

Subsets Set Ω finite, $|\Omega| = n$
Let M be the number of ways of partitioning Ω into k subsets $S_1, \dots, S_k$ with $|S_1| = n_1, \dots, |S_k| = n_k$ s.t.
$n_1 + \dots + n_k = |\Omega| = n$

Then $M = \binom{n}{n_1, \dots, n_k} = \dfrac{n!}{n_1! \dots n_k!}$.

2) Strictly increasing and increasing functions

$x < y \Rightarrow f(x) < f(y)$ $\qquad$ $x < y \Rightarrow f(x) \leq f(y)$

$\{1, \ldots, k\} \longrightarrow \{1, \ldots, n\}$

How many strictly increas. functions are there?

$f : \{1, \ldots, k\} \longrightarrow \{1, \ldots, n\}$

A str. incr. f. is uniquely characterised by its range, which is a subset of $\{1, \ldots, n\}$ of size k.
So # of such functions $= \binom{n}{k}$.

We define a bijection
$\{\text{strictly incr. } f : \{1, \ldots, k\} \to \{1, \ldots, n+k-1\}\}$
to $\{\text{incr. } f : \{1, \ldots, k\} \to \{1, \ldots, n\}\}$

Let f be incr. $f: \{1, \dots, k\} \to \{1, \dots, n\}$

Define $g(i) = f(i) + i - 1$.

So g is a strictly incr. function

$$g: \{1, \dots, k\} \to \{1, \dots, n+k-1\}.$$

So # increas. functions $= \binom{n+k-1}{k}$.

Stirling's formula

For 2 sequences (a_n) and (b_n) we write $a_n \sim b_n$ as $n \to \infty$ if

$$\frac{a_n}{b_n} \to 1 \text{ as } n \to \infty.$$

Theorem (Stirling)

$$n! \sim n^n \sqrt{2\pi n} \cdot e^{-n} \quad \text{as } n \to \infty.$$

Weaker statement

$$\log(n!) \sim n \log n \quad \text{as } n \to \infty$$

Proof $\lfloor x \rfloor$: integer part of x.

$l_n = \log(n!)$

$\log(n!) = \log 2 + \dots + \log n.$

$$\log \lfloor x \rfloor \leq \log x \leq \log \lfloor x + 1 \rfloor$$

Integrate from 1 to n to get

$$l_{n-1} \leq \underbrace{\int_1^n \log x \, dx}_{= n \log n - n + 1} \leq l_n$$

$$n \log n - n + 1 \leq l_n \leq (n+1)\log(n+1) - n$$

So $\dfrac{l_n}{n \log n} \longrightarrow 1$ as $n \to \infty$. $\square$

Proof of Stirling (non-examinable)

For any f

$$\int_a^b f(x)\,dx = \frac{f(a)+f(b)}{2}(b-a) - \frac{1}{2}\int_a^b (x-a)(b-x)f''(x)\,dx$$

Check by integrating by parts twice the right hand side.

Take $f(x) = \log x$, $a = k$, $b = k+1$

$$\int_k^{k+1} \log x\,dx = \frac{\log k + \log(k+1)}{2} + \frac{1}{2}\int_k^{k+1} \frac{(x-k)(k+1-x)}{x^2}\,dx$$

$$= \frac{\log k + \log(k+1)}{2} + \frac{1}{2}\int_0^1 \frac{x(1-x)}{(x+k)^2}\,dx$$

Sum over $k = 1, \dots, n-1$ to get

$$\int_1^n \log x\,dx = \frac{\log((n-1)!) + \log(n!)}{2} + \sum_{k=1}^{n-1} a_k,$$

where $a_k = \frac{1}{2}\int_0^1 \frac{x(1-x)}{(x+k)^2}\,dx.$

$$n\log n - n + 1 = \log(n!) - \frac{\log n}{2} + \sum_{k=1}^{n-1} a_k \quad (\star)$$

Note $a_k \leqslant \frac{1}{2}\,\frac{1}{k^2}\int_0^1 x(1-x)\,dx = \frac{1}{12k^2}$.

So $\sum_{k=1}^{\infty} a_k < \infty$.

Define $A = \exp\left(1 - \sum_{k=1}^{\infty} a_k\right)$

Rearranging $(\star)$ gives

$$\log(n!) = n\log n + \frac{\log n}{2} - n + 1 - \sum_{k=1}^{n-1} a_k .$$

Exponentiating gives

$$n! = n^n \cdot \sqrt{n} \cdot e^{-n} \cdot A \cdot \underbrace{\exp\left(\sum_{k=n}^{\infty} a_k\right)}_{\to 1 \text{ as } n\to\infty} .$$

So we showed $n! \sim n^n \sqrt{n}\, e^{-n} \cdot A$.

Remains to prove $A = \sqrt{2\pi}$.

$$2^{-2n} \cdot \binom{2n}{n} \sim \frac{\sqrt{2}}{A\sqrt{n}} \quad \text{as} \quad n \to \infty$$

We will prove $2^{-2n} \cdot \binom{2n}{n} \sim \frac{1}{\sqrt{\pi n}}$.

So this will show $A = \sqrt{2\pi}$.

Consider $I_n = \int_0^{\frac{\pi}{2}} \cos^n \theta \, d\theta \quad n \geq 0.$

$I_0 = \frac{\pi}{2}$ and $I_1 = 1$.

By integration by parts

$$I_n = \frac{n-1}{n} \cdot I_{n-2}.$$

So $I_{2n} = \frac{2n-1}{2n} \cdots \frac{3}{4} \cdot \frac{1}{2} \underbrace{I_0}_{= \frac{\pi}{2}} =$

$$= \frac{(2n)\cdot(2n-1)\cdot(2n-2)\cdots 1}{2^{2n} \cdot (n!)^2} \cdot \frac{\pi}{2}$$

$$= \frac{1}{2^{2n}} \binom{2n}{n} \cdot \frac{\pi}{2}$$

$$I_{2n+1} = \frac{2n}{2n+1} \cdots \frac{4}{5} \cdot \frac{2}{3} \cdot \overset{=1}{I_1} =$$

$$= \left(2^{-2n} \cdot \binom{2n}{n}\right)^{-1} \cdot \frac{1}{2n+1} \cdot$$

$$\frac{I_n}{I_{n-2}} = \frac{n-1}{n} \longrightarrow 1 \text{ as } n \to \infty.$$

The sequence (I_n) is decreasing in n.

So $\frac{I_{2n}}{I_{2n+1}} \longrightarrow 1$ as $n \to \infty$.

This shows

$$\left(2^{-2n} \cdot \binom{2n}{n}\right)^2 \sim \frac{2}{\pi(2n+1)} \sim \frac{1}{n\pi}$$

as $n \to \infty$.

This proves $A = \sqrt{2\pi}$. $\square$

Properties of prob. measures

Recall def. of prob. space $(\Omega, \mathcal{F}, \mathbb{P})$.

Ω is a set

$\mathcal{F}$ is a σ-algebra

$$\mathbb{P} : \mathcal{F} \to [0,1] \quad \begin{cases} \mathbb{P}(\Omega) = 1 \\ \mathbb{P}\Big(\bigcup_{n=1}^{\infty} \underbrace{A_n}_{\text{disjoint}}\Big) = \sum_{n=1}^{\infty} \mathbb{P}(A_n) \end{cases}$$

[countable additivity]

1) Countable subadditivity

$A_n \in \mathcal{F}$, $\forall n$, then

$$\mathbb{P}\Big(\bigcup_{n=1}^{\infty} A_n\Big) \leq \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

Proof Define $B_1 = A_1$ and $\forall\, n \geq 2$

$$B_n = A_n \setminus (A_1 \cup \dots \cup A_{n-1}).$$

(B_n) is a disjoint sequence

and $\bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} A_n$

So $\mathbb{P}(\bigcup_n A_n) = \mathbb{P}(\bigcup_n B_n) \overset{\text{countable addit.}}{=} \sum_n \mathbb{P}(B_n)$

$\forall n$, $B_n \subseteq A_n \Rightarrow \mathbb{P}(B_n) \leq \mathbb{P}(A_n)$.

Hence $\mathbb{P}(\bigcup_n A_n) \leq \sum_n \mathbb{P}(A_n)$. $\square$

2) <u>Continuity of prob. meas.</u>

Let (A_n) be an increas. seq. in $\mathcal{F}$,

i.e. $A_n \subseteq A_{n+1}$ $\forall n$.

Then we know $\mathbb{P}(A_n) \nearrow$ and converges.

We will prove $\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(\bigcup_n A_n)$.

<u>Proof</u> Set $B_1 = A_1$ and $\forall n \geq 2$

$$B_n = A_n \setminus (A_1 \cup \ldots \cup A_{n-1})$$

Then $\bigcup_{k=1}^{n} B_k = A_n$. and $\bigcup_{k=1}^{\infty} B_k = \bigcup_{k=1}^{\infty} A_k$.

So $\mathbb{P}(A_n) = \mathbb{P}\left(\bigcup_{k=1}^{n} B_k\right) \underset{\substack{\downarrow \\ \text{disjoint events}}}{=} \sum_{k=1}^{n} \mathbb{P}(B_k)$

and $\sum_{k=1}^{n} \mathbb{P}(B_k) \xrightarrow[n\to\infty]{} \sum_{k=1}^{\infty} \mathbb{P}(B_k)$.

But by countable additivity

$$\sum_{k=1}^{\infty} \mathbb{P}(B_k) = \mathbb{P}\left(\bigcup_{k=1}^{\infty} B_k\right) = \mathbb{P}\left(\bigcup_{k=1}^{\infty} A_k\right). \quad \square$$

Similarly if (A_n) is a decreasing sequence, i.e.

$$A_n \supseteq A_{n+1} \quad \forall n$$

then $\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right)$.

<u>Proof</u> Take complements and apply previous.

4) Inclusion-exclusion formula

$(\Omega, \mathcal{F}, \mathbb{P})$

$A, B \in \mathcal{F}$ $\quad \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$

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

For n events $A_1, \dots, A_n$ we have

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

$$1 \leq i_1 < \dots < i_k \leq n$$

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathbb{P}(A_i) - \sum_{1 \leq i_1 < i_2 \leq n} \mathbb{P}(A_{i_1} \cap A_{i_2})$$

$$+ \sum_{1 \leq i_1 < i_2 < i_3 \leq n} \mathbb{P}(A_{i_1} \cap A_{i_2} \cap A_{i_3}) - \dots$$

$$\dots + (-1)^{n+1} \mathbb{P}(A_1 \cap \dots \cap A_n)$$

Proof For $n=2$ it holds.

Suppose it holds for $n-1$ events.

Then $\mathbb{P}(A_1 \cup \dots \cup A_n) = \mathbb{P}(A_1 \cup \dots \cup A_{n-1}) + \mathbb{P}(A_n) -$

$$- \mathbb{P}(\underbrace{A_n \cap (A_1 \cup \dots \cup A_{n-1})}_{}) .$$

$$\mathbb{P}((A_1 \cap A_n) \cup \dots \cup (A_{n-1} \cap A_n))$$

Write $B_k = A_k \cap A_n$. Then

$$\mathbb{P}(A_1 \cup \dots \cup A_n) = \mathbb{P}(A_1 \cup \dots \cup A_{n-1}) + \mathbb{P}(A_n) - \mathbb{P}(B_1 \cup \dots \cup B_{n-1})$$

Apply the inductive hypothesis to

$\mathbb{P}(A_1 \cup \dots \cup A_{n-1})$ and $\mathbb{P}(B_1 \cup \dots \cup B_{n-1})$. $\square$

Specializing to equally likely outcomes
i.e. Ω is a finite set and

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} \qquad \text{gives}$$

$$|A_1 \cup \ldots \cup A_n| = \sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \le i_1 < \ldots < i_k \le n} |A_{i_1} \cap \ldots \cap A_{i_k}|.$$

Bonferroni inequalities

Truncating the sum in incl.-excl. ~~up~~ at the r-th term gives an overestimate if r is odd and an underestimate if r is even.

Proof $n=2$ ✓ $(\mathbb{P}(A \cup B) \le \mathbb{P}(A) + \mathbb{P}(B))$

Induction on n.
Suppose r is odd.

$\mathbb{P}(A_1 \cup \ldots \cup A_n) = \mathbb{P}(A_1 \cup \ldots \cup A_{n-1}) + \mathbb{P}(A_n) - \mathbb{P}(B_1 \cup \ldots \cup B_{n-1})$,
where $B_i = A_i \cap A_n$.
Since r is odd, by the induction hyp. we get an overestimate for $\mathbb{P}(A_1 \cup \ldots \cup A_{n-1})$ when truncating at r-th term.
$r-1$ is even, so truncating at $r-1$ in $\mathbb{P}(B_1 \cup \ldots \cup B_{n-1})$ gives an underestimate.
Putting them together gives an overestimate.

Similarly, if r is even, then $r-1$ is odd $\square$

Counting using incl.-excl.

1) Want the number of surjections

$f : \{1, \dots, n\} \to \{1, \dots, m\}$.

Take $\Omega = \{f : \{1, \dots, n\} \to \{1, \dots, m\}\}$
and $A = \{\text{surjections}\}$.

For every $i \in \{1, \dots, m\}$ define

$$A_i = \{f \in \Omega : i \notin \{f(1), \dots, f(n)\}\}$$

Then $A = A_1^c \cap A_2^c \cap \dots \cap A_m^c = (A_1 \cup \dots \cup A_m)^c$.

By incl.-excl.

$$|A_1 \cup \dots \cup A_m| = \sum_{k=1}^{m} (-1)^{k+1} \sum_{1 \leq i_1 < \dots < i_k \leq m} |A_{i_1} \cap \dots \cap A_{i_k}|$$

Let $i_1 < i_2 < \dots < i_k$
$|A_{i_1} \cap \dots \cap A_{i_k}| = (m-k)^n$.

So

$$|A_1 \cup \dots \cup A_m| = \sum_{k=1}^{m} (-1)^{k+1} \binom{m}{k} \cdot (m-k)^n$$

and

$$|A| = |(A_1 \cup \dots \cup A_m)^c| = \sum_{k=0}^{m} (-1)^k \cdot \binom{m}{k} \cdot (m-k)^n.$$

2) Derangements = permutations with no fixed points.

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

$A = \{\text{derangements}\} = \{f \in \Omega : f(i) \neq i \ \forall i\}.$

Pick a permutation at random, i.e. all perm. are equally likely.

Want $\mathbb{P}(\text{random perm. is a derangement})$

Let $A_i = \{f \in \Omega : f(i) = i\}$

Then $A = (A_1 \cup \dots \cup A_n)^c$.

By incl.-excl.

$$\mathbb{P}(A_1 \cup \dots \cup A_n) = \sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \le i_1 < \dots < i_k \le n} \mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k})$$

$\mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_k}) = \frac{(n-k)!}{n!}$

So $\mathbb{P}(A_1 \cup \ldots \cup A_n) = \sum_{k=1}^{n} (-1)^{k+1} \cdot \binom{n}{k} \cdot \frac{(n-k)!}{n!}$

$\Rightarrow \mathbb{P}(A) = \sum_{k=0}^{n} (-1)^k \cdot \frac{1}{k!} \xrightarrow[n\to\infty]{} e^{-1} \approx 0.3678.$

<u>Independence</u> $(\Omega, \mathcal{F}, \mathbb{P})$

Let $A, B \in \mathcal{F}$. They are said to be <u>independent</u> if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B).$$

A countable seq. $(A_n)_{n\in\mathbb{N}}$ of events is indep. if $\forall$ $k \geq 2$ and $\forall$ distinct indices $i_1, \ldots, i_k$

$$\mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \ldots \mathbb{P}(A_{i_k}).$$

<u>Rem.</u> Pairwise indep. $\not\Rightarrow$ indep.

Toss a fair coin twice.

Take $\Omega = \{(0,0), (0,1), (1,0), (1,1)\}$.

and

$$A = \{(0,0),(0,1)\}, \quad B = \{(0,0),(1,0)\} \text{ and}$$

$$C = \{(0,1),(1,0)\}.$$

$$\mathbb{P}(A \cap B) = \mathbb{P}(A \cap C) = \mathbb{P}(B \cap C) = \frac{1}{4}$$

and $\mathbb{P}(A) \cdot \mathbb{P}(B) = \mathbb{P}(B) \cdot \mathbb{P}(C) = \mathbb{P}(A) \cdot \mathbb{P}(C) = \frac{1}{4}$

so pairwise they are indep.

However, $\mathbb{P}(A \cap B \cap C) = 0 \neq \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2}$.

Rem. If A is indep. of B, then

A is also indep. of B^c.

Indeed, $\mathbb{P}(A \cap B^c) = \mathbb{P}(A) - \mathbb{P}(A \cap B)$

$$\overset{\text{indep.}}{=} \mathbb{P}(A) - \mathbb{P}(A) \cdot \mathbb{P}(B)$$

$$= \mathbb{P}(A) \cdot \mathbb{P}(B^c).$$

Conditional prob. $(\Omega, \mathcal{F}, \mathbb{P})$

$A, B \in \mathcal{F}$ with $\mathbb{P}(B) > 0$.

The cond. prob. of A given B is defined

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

If A is indep. of B, then $\mathbb{P}(A \mid B) = \mathbb{P}(A)$.

Suppose $(A_n)_{n \in \mathbb{N}}$ a disjoint seq. in $\mathcal{F}$.

Then $\mathbb{P}(\bigcup_n A_n \mid B) = \dfrac{\mathbb{P}((\bigcup A_n) \cap B)}{\mathbb{P}(B)} =$

$$= \frac{\mathbb{P}(\bigcup (A_n \cap B))}{\mathbb{P}(B)} \underset{\substack{\uparrow \\ \text{countable} \\ \text{additivity}}}{=} \sum_n \frac{\mathbb{P}(A_n \cap B)}{\mathbb{P}(B)}$$

$$= \sum_n \mathbb{P}(A_n \mid B)$$

Law of total probability

Suppose $(B_n)_{n\in\mathbb{N}}$ is a disjoint sequence in $\mathcal{F}$ with $\bigcup B_n = \Omega$ and $\mathbb{P}(B_n)>0$ $\forall n$.

Let $A \in \mathcal{F}$. Then

$$\mathbb{P}(A) = \sum_n \mathbb{P}(A|B_n)\cdot\mathbb{P}(B_n).$$

Indeed, $\mathbb{P}(A) = \mathbb{P}(A\cap\Omega) =$

$$= \mathbb{P}(A\cap(\bigcup_n B_n)) = \mathbb{P}(\bigcup_n (A\cap B_n))$$

$$\overset{\text{count. addit.}}{=} \sum_n \mathbb{P}(A\cap B_n) = \sum_n \mathbb{P}(A|B_n)\cdot\mathbb{P}(B_n)$$

Bayes' formula (B_n) disjoint and $\bigcup B_n = \Omega$
$\mathbb{P}(A)>0$

$$\mathbb{P}(B_n|A) = \frac{\mathbb{P}(A|B_n)\cdot\mathbb{P}(B_n)}{\sum_k \mathbb{P}(A|B_k)\cdot\mathbb{P}(B_k)}.$$

This is the basis of Bayesian statistics. We have a prior on $\mathbb{P}(B_n)$ and a model giving us $\mathbb{P}(A|B_n)$.

Then Bayes' formula gives us the posterior prob. of B_n given A.

Recall Bayes' formula

$\bigcup B_n = \Omega$, (B_n) disjoint

$$\mathbb{P}(B_n|A) = \frac{\mathbb{P}(A|B_n)\cdot \mathbb{P}(B_n)}{\sum_k \mathbb{P}(A|B_k)\cdot \mathbb{P}(B_k)}$$

<u>Example</u> (False positives for a rare condition)

Suppose there is a rare medical condition A that affects 0.1 % of the population. We have a medical test which is positive for 98 % of the people affected and 1% of those unaffected by the condition. Pick a random individual. What is the prob. he has the condition A given that he was tested positive.

Let $A = \{$ indiv. suffers from A $\}$

$P = \{$ indiv. was tested positive $\}$

$$\mathbb{P}(A|P) = \frac{\mathbb{P}(P|A)\cdot \mathbb{P}(A)}{\mathbb{P}(P|A^c)\cdot \mathbb{P}(A^c) + \mathbb{P}(P|A)\cdot \mathbb{P}(A)}$$

So $\mathbb{P}(A|P) = \dfrac{0.98 \times 0.001}{0.98 \times 0.001 + 0.01 \times 0.999} = 0.089\ldots \approx 0.09$

This might seem counter-intuitive (that this prob. seems low).
But $\mathbb{P}(P|A^c) >> \mathbb{P}(A)$

$$\mathbb{P}(A|P) = \frac{1}{1 + \dfrac{\mathbb{P}(P|A^c) \cdot \mathbb{P}(A^c)}{\mathbb{P}(P|A) \cdot \mathbb{P}(A)}}$$

typically $\mathbb{P}(A^c)$ and $\mathbb{P}(P|A)$ are close to 1

$$\approx \frac{1}{1 + \dfrac{\mathbb{P}(P|A^c)}{\mathbb{P}(A)}}$$

Suppose in 1000 people only 1 is suffering from the condition.
Among 999 not suffering about 10 will test positive. So in total about 11 will test positive.
So prob. a random positive indiv. has A is $\frac{1}{11}$.

Example Extra knowledge changes prob. in surprising ways.

a) I have 2 children, the elder of whom is a boy.

b) I have 2 children, one of them is a boy

c) I have 2 children, one of them is a boy born on a Tuesday.

What is the prob. I have 2 boys when I condition on a / b / c?

~~a)~~ Since no further info, we take all possible outcomes to be equally likely.

a) Write $BG = \{\text{elder} = \text{boy}, \text{younger} = \text{girl}\}$

$GB = \{ \text{—"—} = \text{girl}, \text{—"—} = \text{boy}\}$

BB, GG

$$P(BB \mid BG \cup BB) = \frac{1}{2}.$$

b) $P(BB \mid BG \cup GB \cup BB) = \frac{1}{3}$

c) $TT = \{$elder = boy on Tuesday, younger = boy on T$\}$

$TN = \{$ — " — , younger = boy not on T$\}$

NT, GT, TG

$$\mathbb{P}(TT \cup TN \cup NT \mid TT \cup TN \cup NT \cup GT \cup TG)$$

$$= \frac{\mathbb{P}(TT \cup TN \cup NT)}{\mathbb{P}(TT \cup TN \cup NT \cup GT \cup TG)}$$

$$P(TT \cup TN \cup NT) = \frac{1}{2}\cdot\frac{1}{7}\cdot\frac{1}{2}\cdot\frac{1}{7} + \frac{1}{2}\cdot\frac{1}{7}\cdot\frac{1}{2}\cdot\frac{6}{7} + \frac{1}{2}\cdot\frac{6}{7}\cdot\frac{1}{2}\cdot\frac{1}{7}$$

$$P(TT \cup TN \cup NT \cup GT \cup TG) = \ldots + \frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{7} + \frac{1}{2}\cdot\frac{1}{7}\cdot\frac{1}{2}$$

$$= \frac{13}{27}.$$

Simpson's paradox

There are 50 men and 50 women applying to a college.

All applicants	Admitted	Rejected	% Adm.
State	25	25	50%
Indep.	28	22	56%

Overall prob. of being accepted is 0.53

Men only	Adm.	Rej.	% Adm.
State	15	22	41%
Indep.	5	8	38%

Women only	Adm.	Rej.	% Adm.
State	10	3	77%
Indep.	23	14	62%

This is called <u>confounding</u> in statistics. It happens when one aggregates data from 2 different populations.

Overall, women have 66% acceptance rate and men have 40% accept. rate.
But proportion of men from state schools is 74% and 26% from indep.
This is reversed for women.

Another way to see it: A, B, a, b, C, D, c, d positive numbers.

Suppose $\frac{A}{B} > \frac{a}{b}$ and $\frac{C}{D} > \frac{c}{d}$

$$\not\Rightarrow \frac{A+C}{B+D} > \frac{a+c}{b+d}.$$

Discrete distributions

$(\Omega, \mathcal{F}, \mathbb{P})$, Ω is a finite/countable set

$$\Omega = \{\omega_1, \omega_2, \dots\}$$

$\mathcal{F} = \{$all possible subsets of $\Omega\}$.

In order to determine $\mathbb{P}$, it suffices to specify $\mathbb{P}(\{\omega_i\})$ $\forall i$ $(\{\omega_i\} \in \mathcal{F})$
and $\forall A \subset \Omega$ $\mathbb{P}(A) = \sum_{i : \omega_i \in A} \mathbb{P}(\{\omega_i\})$.

We write $p_i = \mathbb{P}(\{\omega_i\})$ and we call (p_i) a [discrete] distribution
[Prob.]

- $p_i \geq 0$ $\forall i$
- $\sum_i p_i = 1$

1) Bernoulli distr.

$\Omega = \{0, 1\}$

Toss a coin once - prob. p of H
$1-p$ of T.

The Bernoulli distr. models the number of H.

$\mathbb{P}(1\ H) = P_1 = p$

$\mathbb{P}(0\ H) = P_0 = 1-p$

2) Binomial distr.

Toss a p-coin N times independently.
Count the number of H.

So $\Omega = \{0, 1, \dots, N\}$

and $P_k = \mathbb{P}(k \text{ Heads}) = \binom{N}{k} p^k \cdot (1-p)^{N-k}$

$$\sum_{k=0}^{N} P_k = \sum_{k=0}^{N} \binom{N}{k} \cdot p^k \cdot (1-p)^{N-k} = (p + (1-p))^N = 1.$$

3) Multinomial distr.

$\underset{1}{\sqcup}\ \underset{2}{\sqcup}\ \dots\ \underset{k}{\sqcup}$ boxes

Throw N balls indep. Prob. a ball falls in box i is p_i, so $\sum_{i=1}^{k} p_i = 1$

$\Omega = \{$ordered partitions of $N\} =$

$$= \left\{ (n_1, \dots, n_k) \in \mathbb{N}_0^k : \sum_{i=1}^{k} n_i = N \right\}.$$

$P(n_1$ balls fell in box $1, \dots, n_k$ in box $k)$

$$= \binom{N}{n_1, \dots, n_k} p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k}.$$

4) Geometric distr.

Toss a p-coin until first H.

$\Omega = \{1, 2, \dots\}$

$p_k = P($tossed k times until $H)$

$$= (1-p)^{k-1} \cdot p$$

$$\sum_{k=1}^{\infty} p_k = 1$$

Let $\Omega = \{0, 1, 2, \dots\}$

$p_k = \mathbb{P}(k\ T$ appeared before first $H)$

$$= (1-p)^k \cdot p.$$

Poisson distribution

It is used to model the number of occurrences of events in a given period of time, i.e. number of customers entering a shop in a day.

Take $\Omega = \{0, 1, 2, \dots\}$

and $p_k = e^{-\lambda} \cdot \dfrac{\lambda^k}{k!}$, $k \in \Omega$, where

λ is a positive real.

We call this the Poisson distribution with parameter λ.

Suppose customers arrive in a shop during $[0,1]$. Take $n \in \mathbb{N}$ and subdivide $[0,1]$ into $\left[\frac{i-1}{n}, \frac{i}{n}\right]$ $i = 1, \dots, n$. In each interval a customer arrives with probability p (indep. for dif. intervals)

$$\mathbb{P}(k \text{ customers arrived}) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}.$$

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

Take $p = \dfrac{\lambda}{n}$, $\lambda \geq 0$.

Then $\binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} = \dfrac{n!}{k!(n-k)!} \dfrac{\lambda^k}{n^k} \left(1 - \dfrac{\lambda}{n}\right)^{n-k} =$

$$= \frac{\lambda^k}{k!} \times \frac{n!}{n^k (n-k)!} \cdot \left(1 - \frac{\lambda}{n}\right)^{n-k}.$$

Keep k fixed and take the limit as $n \to \infty$

$$\lim_{n\to\infty} \frac{\lambda^k}{k!} \cdot \underbrace{\frac{n!}{n^k (n-k)!}}_{\downarrow\, n\to\infty \atop 1} \cdot \underbrace{\left(1 - \frac{\lambda}{n}\right)^{n-k}}_{\downarrow \atop e^{-\lambda}} = \frac{\lambda^k}{k!} \cdot e^{-\lambda}$$

which is the Poisson distr. with par. λ.

So we proved that the Binomial distr. with parameters ~~λ~~ n and $\frac{\lambda}{n}$ converges to the Poisson(λ).

$$\left[\sum_{k=0}^{\infty} p_k = \sum_{k=0}^{\infty} e^{-\lambda} \cdot \frac{\lambda^k}{k!} = e^{-\lambda} \cdot e^{\lambda} = 1, \text{ so indeed a distr.}\right]$$

Random variables

Def. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a prob. space. X is called a random variable if $X : \Omega \to \mathbb{R}$ and $\forall x \in \mathbb{R}$

$$\{X \leq x\} = \{\omega : X(\omega) \leq x\} \in \mathcal{F}.$$

More generally we write

$$\{X \in A\} = \{\omega : X(\omega) \in A\}.$$

Given $A \in \mathcal{F}$, define the indicator of A to be

$$1_A(\omega) = 1(\omega \in A) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \in A^c \end{cases}.$$

Then 1_A is a random variable.

Let X be a r.v. Define the prob. distrib. function of X to be $F_X : \mathbb{R} \to [0,1]$ given by $F_X(x) = \mathbb{P}(X \leq x)$.

Properties
- $F_X(x) \to 1$ as $x \to \infty$.
  $F_X(x) \to 0$ as $x \to -\infty$
- F_X is increasing (obvious)
- F_X is right continuous, i.e.

$$\lim_{h \downarrow 0} F_X(x+h) = F_X(x)$$

All these properties follow from the continuity of prob. measure property.

<u>Def.</u> $(\Omega, \mathcal{F}, \mathbb{P})$. $(X_1, \dots, X_n)$ is called a random variable in $\mathbb{R}^n$ if $(X_1, \dots, X_n) : \Omega \to \mathbb{R}^n$ and $\forall\, x_1, \dots, x_n \in \mathbb{R}$

$$\{X_1 \le x_1, \dots, X_n \le x_n\} \in \mathcal{F}.$$

This is equivalent to $X_1, \dots, X_n$ all being random variables (in $\mathbb{R}$).

$$= \{\omega \in \Omega : X_1(\omega) \le x_1, \dots, X_n(\omega) \le x_n\}.$$

$$\{X_1 \le x_1, \dots, X_n \le x_n\} = \{X_1 \le x_1\} \cap \dots \cap \{X_n \le x_n\}$$

Focus on the case where $n=1$ and X takes values in a countable set. We call X a discrete random variable. Suppose X takes values in a countable set S. Then
$p_x = \mathbb{P}(X=x) = \mathbb{P}(\{\omega : X(\omega) = x\})$ is called the probability mass function or the distribution of X.

Def. Let $X_1, \dots, X_n$ be random variables (discrete). We call them <u>independent</u> if

$\forall\ x_1, \dots, x_n$
$$\mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = \mathbb{P}(X_1 = x_1) \cdots \mathbb{P}(X_n = x_n).$$

Example Toss a p-biased coin N times indep.

Take $\Omega = \{0,1\}^N$.

$\omega \in \Omega \quad p_\omega = \prod_{k=1}^{N} p^{\omega_k} \cdot (1-p)^{1-\omega_k}$.

$\omega = (\omega_1, \dots, \omega_N) \qquad \omega_i \in \{0,1\}$

Define $X_k(\omega) = \omega_k$, $k = 1, \dots, N$

gives the outcome of the k-th toss.

Then (X_k) have the Bernoulli distribution and they are independent.

Indeed $\mathbb{P}(X_1 = x_1, \dots, X_N = x_N) = p_{(x_1, \dots, x_N)} = \prod_{k=1}^{N} p^{x_k} \cdot (1-p)^{1-x_k}$.

$(x_i \in \{0,1\})$
$$= \prod_{k=1}^{N} \mathbb{P}(X_k = x_k).$$

So indep.

[X r.v. with mass function (p_x). If (p_x) is Bernoulli we say X has the Bernoulli distr. If (p_x) is Geom, ...]

Define $S_N(\omega) = X_1(\omega) + \dots + X_N(\omega)$.

S_N counts the number of H.

$$\underset{k=0,\dots,N}{\mathbb{P}(S_N = k)} = \binom{N}{k} \cdot p^k \cdot (1-p)^{n-k}$$

$$\{S_N = k\} = \{\omega : S_N(\omega) = k\}.$$

So S_N has the Binomial distr. with parameters N and p.

Expectation (discrete)

$(\Omega, \mathcal{F}, \mathbb{P}) \quad X: \Omega \to \mathbb{R}$

X is called non-negative if $X \geq 0$.

Define the expectation to be

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

Write $\Omega_X = \{X(\omega) : \omega \in \Omega\}$

$$\Omega = \bigcup_{x \in \Omega_X} \{X = x\}$$

Recall $\{X = x\} = \{\omega : X(\omega) = x\}$

So $$\mathbb{E}[X] = \sum_{x \in \Omega_X} \sum_{\omega \in \{X = x\}} X(\omega) \cdot \mathbb{P}(\{\omega\}) =$$

$$= \sum_{x \in \Omega_X} \sum_{\omega \in \{X = x\}} x \cdot \mathbb{P}(\{\omega\}) = \sum_{x \in \Omega_X} x \cdot \underbrace{\sum_{\omega \in \{X = x\}} \mathbb{P}(\{\omega\})}_{\mathbb{P}(X = x)}$$

$$= \sum_{x \in \Omega_X} x \cdot \mathbb{P}(X = x).$$

So the expectation is an average of the values taken by X with weights given by the corresponding probabilities.

__Example__ Suppose $X \sim \mathrm{Bin}(n,p)$, i.e.

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

$$\mathbb{E}[X] = \sum_{k=0}^{n} k \cdot \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} =$$

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

$$= np \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!\,((n-1)-(k-1))!}\, p^{k-1} \cdot (1-p)^{(n-1)-(k-1)}$$

$$= np \sum_{k=0}^{n-1} \underbrace{\binom{n-1}{k} \cdot p^k \cdot (1-p)^{n-1-k}}_{(p+(1-p))^{n-1}} = np$$

__Example__ $X \sim \mathrm{Poi}(\lambda)$, $\lambda > 0$, i.e.

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

$$\mathbb{E}[X] = \sum_{k=0}^{\infty} k\, e^{-\lambda} \cdot \frac{\lambda^k}{k!} = \lambda \sum_{k=1}^{\infty} e^{-\lambda} \cdot \frac{\lambda^{k-1}}{(k-1)!} = \lambda$$

Let X be a general r.v. (discrete)

Define $X_+ = \max(X, 0)$ and $X_- = \max(-X, 0)$

Then $X = X_+ - X_-$ and $|X| = X_+ + X_-$

If not both $\mathbb{E}[X_+]$, $\mathbb{E}[X_-]$ are equal to ∞, then we define $\mathbb{E}[X] = \mathbb{E}[X_+] - \mathbb{E}[X_-]$

If both $\mathbb{E}[X_+] = \infty$ and $\mathbb{E}[X_-] = \infty$, then the expectation of X is not defined.

If $\mathbb{E}[|X|] < \infty$, then we call X integrable.

Whenever we write $\mathbb{E}[X]$, it is assumed to be well-defined.

Like before $\mathbb{E}[X] = \sum_{x \in \Omega_X} x \cdot \mathbb{P}(X = x)$.

<u>Properties</u> 1) $X \geq 0$, then $\mathbb{E}[X] \geq 0$ and if $X \geq 0$ and $\mathbb{E}[X] = 0$, then $\mathbb{P}(X = 0) = 1$.

2) If c is a real constant, then $\mathbb{E}[cX] = c\,\mathbb{E}[X]$ and $\mathbb{E}[c + X] = c + \mathbb{E}[X]$.

and $\mathbb{E}[c] = c$.

3) Let X and Y be r.v.'s.

Then $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$.

$\mathbb{E}[\cdot]$: linear operator, i.e.
$\forall\, c_1, \dots, c_n \in \mathbb{R}$ and $X_1, \dots, X_n$ r.v.'s

$$\mathbb{E}\Big[\sum_{i=1}^{n} c_i X_i\Big] = \sum_{i=1}^{n} c_i\, \mathbb{E}[X_i].$$

Let $(X_n)_n$ be a sequence of non-negative r.v.'s. Then

$$\mathbb{E}\Big[\sum_{n=1}^{\infty} X_n\Big] = \sum_{n=1}^{\infty} \mathbb{E}[X_n].$$

<u>Proof</u> $\mathbb{E}\Big[\sum_{n=1}^{\infty} X_n\Big] = \sum_{\omega} \sum_{n=1}^{\infty} X_n(\omega) \cdot \mathbb{P}(\{\omega\}) =$

$$\overset{\geq 0}{=} \sum_{n=1}^{\infty} \underbrace{\sum_{\omega} X_n(\omega) \cdot \mathbb{P}(\{\omega\})}_{\mathbb{E}[X_n]} = \sum_{n=1}^{\infty} \mathbb{E}[X_n].$$

4) Let $A \in \mathcal{F}$ and consider $X = \mathbb{1}(A)$.
Then $\mathbb{E}[X] = \mathbb{P}(A)$.

5) Let $g : \mathbb{R} \to \mathbb{R}$ and X a r.v.
Then $g(X)$ is another r.v. defined via
$g(X)(\omega) = g(X(\omega))$.

$$\mathbb{E}[g(X)] = \sum_{x \in \Omega_X} g(x) \cdot \mathbb{P}(X = x)$$

Proof Let $Y = g(X)$. Then

$$\mathbb{E}[Y] = \sum_{y \in \Omega_Y} y \cdot \mathbb{P}(Y = y)$$

$$\{Y = y\} = \{\omega : Y(\omega) = y\} = \{\omega : g(X(\omega)) = y\} =$$
$$= \{\omega : X(\omega) \in g^{-1}(\{y\})\} = \{X \in g^{-1}(\{y\})\}$$

$$\mathbb{E}[Y] = \sum_{y \in \Omega_Y} y \cdot \mathbb{P}(X \in g^{-1}(\{y\})) =$$

$$= \sum_{y \in \Omega_Y} y \cdot \sum_{x \in g^{-1}(\{y\})} \mathbb{P}(X = x) = \sum_{y \in \Omega_Y} \sum_{x \in g^{-1}(\{y\})} y \cdot \mathbb{P}(X = x)$$

$$= \sum_{y \in \Omega_Y} \sum_{x \in g^{-1}(\{y\})} g(x) \cdot \mathbb{P}(X = x) =$$

$$= \sum_{x \in \Omega_X} g(x) \cdot \mathbb{P}(X = x).$$

6) Let $X \geq 0$ and take integer values. Then

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} \mathbb{P}(X \geq k) = \sum_{k=0}^{\infty} \mathbb{P}(X > k).$$

Proof We can write

$$X = \sum_{k=1}^{\infty} 1(X \geq k) = \sum_{k=0}^{\infty} 1(X > k)$$

Use $\mathbb{P}(A) = \mathbb{E}[1(A)]$ and linearity of expectation

Another proof of the inclusion-exclusion formula

Prop. of indicators

- $1(A^c) = 1 - 1(A)$
- $1(A \cap B) = 1(A) \cdot 1(B)$
- $1(A \cup B) = 1 - (1 - 1(A))(1 - 1(B))$.

More generally

$$1(A_1 \cap \dots \cap A_n) = \prod_{i=1}^{n} 1(A_i)$$

$$1(A_1 \cup \dots \cup A_n) = 1 - \prod_{i=1}^{n} (1 - 1(A_i)) =$$

$$= \sum_i 1(A_i) - \sum_{i_1 < i_2} 1(A_{i_1} \cap A_{i_2}) + \dots + (-1)^{n+1} \mathbb{P}(A_1 \cap \dots \cap A_n)$$

Take expectations to get

$$\mathbb{P}(A_1 \cup \dots \cup A_n) = \sum \mathbb{P}(A_i) - \sum_{i_1 < i_2} \mathbb{P}(A_{i_1} \cap A_{i_2}) + \dots + (-1)^{n+1} \mathbb{P}(A_1 \cap \dots \cap A_n)$$

Terminology X r.v., $r \in \mathbb{N}$

We call $\mathbb{E}[X^r]$ the r-th moment of X.
(assuming it is well-defined)

Variance is defined to be

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

It is a measure of how spread out X is around $\mathbb{E}[X]$. The smaller the variance, the more concentrated the distr. of X is around $\mathbb{E}[X]$.

$\sqrt{Var(X)}$ = standard deviation.

Prop. of Var

1) $Var(X) \geq 0$ and if $Var(X) = 0$, then $\exists$ a constant c s.t. $\mathbb{P}(X = c) = 1$. $(c = \mathbb{E}[X])$

2) $c \in \mathbb{R}$ a constant, then

$$Var(cX) = c^2 Var(X) \quad \text{and} \quad Var(X+c) = Var(X).$$

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

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

$$= \mathbb{E}[X^2 - 2X \cdot \mathbb{E}[X] + (\mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$

4) $\forall c \in \mathbb{R}$

$$\mathbb{E}[(X-c)^2] \geq Var(X) \text{ with equality when } c = \mathbb{E}[X].$$

<u>Proof</u> Expand in $\mathbb{E}[(X-c)^2] =$

$$= \mathbb{E}[X^2] - 2c\,\mathbb{E}[X] + c^2 = f(c)$$

Then f has a minimum at $c = \mathbb{E}[X]$ which is equal to $Var(X)$.

<u>Example</u> $X \sim Bin(n,p)$ $\quad \mathbb{E}[X] = np$

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

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

$$= \sum_{k=0}^{n} k\cdot(k-1)\cdot\binom{n}{k}\cdot p^k \cdot (1-p)^{n-k} =$$

$$= p^2 \sum_{k=2}^{n} \frac{n\cdot(n-1)\,(n-2)!}{(k-2)!\,((n-2)-(k-2))!}\, p^{k-2}\cdot(1-p)^{(n-2)-(k-2)}$$

$$= p^2 \sum_{k=0}^{n-2} n(n-1)\cdot\binom{n-2}{k} p^k \cdot (1-p)^{n-2-k}$$

$$= p^2 n(n-1)$$

<u>Example</u> Let $X \sim Poi(\lambda)$, $\lambda > 0$

$$\mathbb{P}(X=k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!} \quad , \quad k = 0, 1, 2, \ldots$$

$$Var(X) = \mathbb{E}[X(X-1)] + \mathbb{E}[X] - (\mathbb{E}[X])^2$$

$$\mathbb{E}[X(X-1)] = \sum_{k=2}^{\infty} k(k-1) e^{-\lambda} \cdot \frac{\lambda^k}{k!} =$$

$$= \sum_{k=2}^{\infty} e^{-\lambda} \cdot \frac{\lambda^{k-2}}{(k-2)!} \cdot \lambda^2 = \lambda^2$$

We calculated $\mathbb{E}[X] = \lambda$

So $Var(X) = \lambda^2 + \lambda - \lambda^2 = \lambda$.

<u>Def.</u> Let X and Y be 2 r.v.'s.
Define the covariance to be

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

It is a "measure" of how dependent X and Y are.

<u>Prop.</u> 1) $Cov(X,Y) = Cov(Y,X)$

2) $Cov(X,X) = Var(X)$

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

Proof Expand $(X-\mathbb{E}X)(Y-\mathbb{E}Y)$ and use basic prop. of expectation.

4) Let $c \in \mathbb{R}$ be a constant. Then $\mathrm{Cov}(cX,Y) = c\,\mathrm{Cov}(X,Y)$ and $\mathrm{Cov}(X+c,Y) = \mathrm{Cov}(X,Y)$

5) $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X,Y)$

Proof $\mathrm{Var}(X+Y) = \mathbb{E}\left[\left(X+Y-(\mathbb{E}(X)+\mathbb{E}(Y))\right)^2\right] =$

$= \mathbb{E}\left[\left((X-\mathbb{E}[X]) + (Y-\mathbb{E}[Y])\right)^2\right] =$

$= \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X,Y)$.

6) Let $c \in \mathbb{R}$. Then $\mathrm{Cov}(c,X) = 0$

7) $\mathrm{Cov}(X+Z, Y) = \mathrm{Cov}(X,Y) + \mathrm{Cov}(Z,Y)$

More generally, $c_1,\dots,c_n$, $d_1,\dots,d_n \in \mathbb{R}$ and $X_1,\dots,X_n$ and $Y_1,\dots,Y_n$ r.v.'s, then

$$\mathrm{Cov}\left(\sum_{i=1}^n c_i X_i, \sum_{i=1}^n d_i Y_i\right) = \sum_{i=1}^n \sum_{j=1}^n c_i d_j \mathrm{Cov}(X_i, Y_j).$$

In particular

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + \sum_{i \neq j} \mathrm{Cov}(X_i, X_j).$$

Recall X and Y are <u>indep.</u> if

$$\forall x,y \quad \mathbb{P}(X=x, Y=y) = \mathbb{P}(X=x)\,\mathbb{P}(Y=y)$$

Let X and Y be indep. Let f, g be non-negative functions, $f, g : \mathbb{R} \to \mathbb{R}_+$.

Then $$\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)] \cdot \mathbb{E}[g(Y)]$$

<u>Proof</u> $$\mathbb{E}[f(X)g(Y)] = \sum_{x,y} f(x)\,g(y)\,\mathbb{P}(X=x, Y=y)$$

$$\overset{\text{indep.}}{=} \sum_{x,y} f(x)\,g(y)\,\mathbb{P}(X=x)\,\mathbb{P}(Y=y) =$$

$$= \sum_{x} f(x)\,\mathbb{P}(X=x) \sum_{y} g(y)\,\mathbb{P}(Y=y) =$$

$$= \mathbb{E}[f(X)] \cdot \mathbb{E}[g(Y)].$$

In particular, if X and Y are indep., then $\mathrm{Cov}(X,Y) = 0$, since

$$\mathrm{Cov}(X,Y) = \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)] = 0$$

$\mathrm{Cov}(X,Y)=0 \not\Rightarrow$ independence.

Indeed, let X_1, X_2, X_3 be indep. $\mathrm{Ber}(\frac{1}{2})$.

Define $Y_1 = 2X_1 - 1$, $Y_2 = 2X_2 - 1$ and

$$Z_1 = Y_1 X_3 \quad \text{and} \quad Z_2 = Y_2 X_3.$$

$\mathbb{E}[Y_1] = \mathbb{E}[Y_2] = \mathbb{E}[Z_1] = \mathbb{E}[Z_2] = 0$

$\mathrm{Cov}(Z_1, Z_2) = 0$

However, $\mathbb{P}(Z_1 = 0, Z_2 = 0) = \mathbb{P}(X_3 = 0) = \frac{1}{2} \neq \frac{1}{4} = \mathbb{P}(Z_1 = 0) \cdot \mathbb{P}(Z_2 = 0)$

so not indep.

Inequalities

Markov's ineq. Let X be a non-negative r.v. Then $\forall\, a>0$ we have

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}.$$

Proof Observe $\quad X \geq a \cdot \mathbb{1}(X \geq a)$. $(X \geq 0)$

Take expectation of both sides to get

$$\mathbb{E}[X] \geq a \cdot \mathbb{P}(X \geq a). \qquad \square$$

Chebyshev's ineq. Let X be a r.v. with $\mathbb{E}[X] < \infty$.

Then $\forall\, a>0$, $\mathbb{P}(|X-\mathbb{E}[X]| \geq a) \leq \dfrac{\mathrm{Var}(X)}{a^2}$.

Proof Notice

$$\mathbb{P}(|X-\mathbb{E}[X]| \geq a) = \mathbb{P}((X-\mathbb{E}[X])^2 \geq a^2)$$

Now apply Markov's ineq. to the non-negat. r.v. $(X-\mathbb{E}[X])^2$.

Cauchy-Schwartz ineq.

Let X and Y be r.v.'s. Then

$$\mathbb{E}[|XY|] \le \sqrt{\mathbb{E}[X^2]}\sqrt{\mathbb{E}[Y^2]}.$$

Suffices to prove it for $X, Y \geqslant 0$ and $\mathbb{E}[X^2] < \infty$ $\mathbb{E}[Y^2] < \infty$.

Since $XY \le \frac{1}{2}(X^2 + Y^2)$, take expect. to get

$$\mathbb{E}[XY] \le \frac{1}{2}\left(\mathbb{E}[X^2] + \mathbb{E}[Y^2]\right) < \infty.$$

Assume $\mathbb{E}[X^2] > 0$ and $\mathbb{E}[Y^2] > 0$. Otherwise result is trivial.

Let $t \in \mathbb{R}$ and consider

$$0 \le (X - tY)^2 = X^2 - 2tXY + t^2Y^2$$

Take expect. $\rightsquigarrow$ $\underbrace{\mathbb{E}[X^2] - 2t\,\mathbb{E}[XY] + t^2\,\mathbb{E}[Y^2]}_{f(t)} \geqslant 0.$

Minimising f gives that for $t = \dfrac{\mathbb{E}[XY]}{\mathbb{E}[Y^2]}$

We get that the minimum is achieved. Plugging in this value and rearranging gives

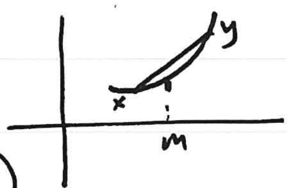
$$\mathbb{E}[XY] \leq \sqrt{\mathbb{E}[X^2]\mathbb{E}[Y^2]}. \qquad \square$$

Let's examine cases of equality in C-S.

In this case we have $\mathbb{E}[(X-tY)^2]=0$ for t as above. But this forces the ≥ 0 r.v. $(X-tY)^2$ to be equal to 0.
So this gives $\mathbb{P}(X=\lambda Y)=1$ for $\lambda \in \mathbb{R}$.

<u>Jensen's ineq</u>. Let f be a convex function defined on $\mathbb{R}$, i.e. $\forall\, x,y \in \mathbb{R}$, $\forall\, t \in [0,1]$

$$f(tx+(1-t)y) \leq t f(x) + (1-t) f(y).$$



<u>Jensen</u> f convex, X a r.v. Then

$$\mathbb{E}[f(X)] \geq f(\mathbb{E}[X]).$$

[Apply to $f(x)=x^2$ then $\mathbb{E}[X^2] \geq (\mathbb{E}[X])^2$ which is true since $\mathrm{Var}(X) \geq 0$.]

(Rule to remember $\geq$ or $\leq$)

Proof A convex function is equal to the sup of all lines lying below it, i.e.
$\forall\ m \in \mathbb{R} \quad \exists\ a, b \in \mathbb{R}$ s.t.

$$f(m) = am + b \quad \text{and} \quad ax + b \leq f(x) \quad \forall x.$$

Pf of this claim: Let $m \in \mathbb{R}$ and let $x < m < y$
Then $\exists\ t \in (0,1)$ s.t.

$$m = tx + (1-t)y.$$

By convexity $f(m) \leq t f(x) + (1-t) f(y)$.

Rearranging this gives

$$\frac{f(m) - f(x)}{m - x} \leq \frac{f(y) - f(m)}{y - m}$$

So $\exists\ a \in \mathbb{R}$ s.t. $\dfrac{f(m) - f(x)}{m - x} \leq a \leq \dfrac{f(y) - f(m)}{y - m}$

So $\forall\ x$ this gives $f(x) \geq a(x - m) + f(m)$.

Let $m = \mathbb{E}[X]$. Then $\exists\ a, b \in \mathbb{R}$

$$f(m) = am + b \quad \text{and} \quad aX + b \leq f(X) \quad \star$$

Take exp. in $\star$

$$a\,\mathbb{E}[X] + b \leq \mathbb{E}[f(X)]. \text{ But } \underbrace{f(m)}_{f(\mathbb{E}(X))} = a\,\mathbb{E}[X] + b.$$

$\square$

Cases of <u>equality</u> in <u>Jensen</u>

Let f be a convex function.
Assume $\exists\ m \in \mathbb{R}$ s.t. $f(m) = am+b$ and
$f(x) > ax+b \quad \forall\ x \neq m$ for some $a, b \in \mathbb{R}$.

Suppose $m = \mathbb{E}[X]$.

Consider $f(X) - (aX+b) \geqslant 0$.

Assume $\mathbb{E}[f(X)] = f(\mathbb{E}[X])$.

Also $\mathbb{E}[f(X) - (aX+b)] = 0$ since it is equal to
$\mathbb{E}[f(X)] - (a\,\mathbb{E}[X] + b) = \mathbb{E}[f(X)] - f(\mathbb{E}[X]) = 0$
by assumption.

So $f(X) - (aX+b) = 0$ with prob. 1, i.e.

$\mathbb{P}(f(X) = aX+b) = 1$.

This forces $\mathbb{P}(X = m) = 1$.

AM/GM ineq.

Let f be convex and let $x_1, \dots, x_n \in \mathbb{R}$. Then

$$\frac{1}{n} \sum_{k=1}^{n} f(x_k) \geq f\left(\frac{1}{n}\sum_{k=1}^{n} x_k\right).$$

Indeed, take X to be a r.v. taking values $x_1, \dots, x_n$ each with prob. $\frac{1}{n}$ and apply Jensen.

Let $f = -\log$. Then

$$\left(\prod_{k=1}^{n} x_k\right)^{\frac{1}{n}} \leq \frac{1}{n} \sum_{k=1}^{n} x_k$$

$$GM \leq AM.$$

Conditional expect. Let B be an event, $\mathbb{P}(B) > 0$. Let X be a r.v.

Define $\mathbb{E}[X \mid B] = \dfrac{\mathbb{E}[X \cdot 1(B)]}{\mathbb{P}(B)}$.

Law of total expect. $(\Omega_n)_n$ disjoint, $\bigcup \Omega_n = \Omega$

$$\mathbb{E}[X] = \sum_n \mathbb{E}[X \mid \Omega_n] \cdot \mathbb{P}(\Omega_n) \qquad X \geq 0$$

Proof Write $X = \sum_n X \cdot 1(\Omega_n)$

Use countable additivity for expect., i.e.

$$\mathbb{E}[X] = \sum_n \mathbb{E}[X \cdot 1(\Omega_n)] = \sum_n \mathbb{E}[X \mid \Omega_n] \cdot \mathbb{P}(\Omega_n).$$

<u>Terminology</u> $X_1, \dots, X_n$ r.v.'s

Their joint distribution is defined to be

$$\mathbb{P}(X_1 = x_1, \dots, X_n = x_n), \quad x_1 \in \Omega_{X_1}, \dots, x_n \in \Omega_{X_n}.$$

The marginal distr. of X_i is

$$\mathbb{P}(X_i = x_i) = \sum_{x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n} \mathbb{P}(X_1 = x_1, \dots, X_n = x_n)$$

by the law of total prob.

Let X & Y be r.v.'s.

The conditional distr. of X given $Y = y$ $(y \in \Omega_Y)$ is defined to be

$$\mathbb{P}(X = x \mid Y = y), \quad x \in \Omega_X.$$
$$= \mathbb{P}(X = x, Y = y) / \mathbb{P}(Y = y)$$

Law of total prob.

$$\mathbb{P}(X = x) = \sum_y \mathbb{P}(X = x, Y = y) = \sum_y \mathbb{P}(X = x \mid Y = y) \mathbb{P}(Y = y)$$

Let X and Y be indep.

$$P(X+Y=z) = \sum_y P(X+Y=z, Y=y) =$$

$$= \sum_y P(X=z-y, Y=y) \overset{\text{indep.}}{=} \sum_y P(X=z-y) \cdot P(Y=y)$$

This is called the <u>convolution</u> of the distr. of X and Y.

Similarly $P(X+Y=z) = \sum_x P(X=x) \cdot P(Y=z-x)$.

<u>Ex.</u> $X \sim \mathrm{Poi}(\lambda)$, $Y \sim \mathrm{Poi}(\mu)$, $X \perp\!\!\!\perp Y$ ↳ independent

$$P(X+Y=n) = \sum_{r=0}^{n} P(X=r, Y=n-r) =$$

$$= \sum_{r=0}^{n} P(X=r) \cdot P(Y=n-r) = \sum_{r=0}^{n} e^{-\lambda} \cdot \frac{\lambda^r}{r!} \, e^{-\mu} \cdot \frac{\mu^{n-r}}{(n-r)!} =$$

$$= e^{-(\lambda+\mu)} \frac{\mu^n}{n!} \underbrace{\sum_{r=0}^{n} \left(\frac{\lambda}{\mu}\right)^r \binom{n}{r}}_{\left(1+\frac{\lambda}{\mu}\right)^n} =$$

$$= e^{-(\lambda+\mu)} \frac{(\lambda+\mu)^n}{n!}$$, indeed the distr. function of a $\mathrm{Poi}(\lambda+\mu)$.

The cond. expectation of X given $Y=y$ is defined to be

$$\mathbb{E}[X|Y=y] = \frac{\mathbb{E}[X \cdot 1(Y=y)]}{\mathbb{P}(Y=y)} = \sum_x x \cdot \mathbb{P}(X=x|Y=y).$$

This cond. expect. is only a function of y.
Define $g(y) = \mathbb{E}[X|Y=y]$.

So $g: \mathbb{R} \to \mathbb{R}$ is a function.

Define the conditional expect. of X given Y to be $g(Y)$. We denote it by $\mathbb{E}[X|Y]$.

This is a random variable and it depends entirely on Y (it is a function of Y).

$$\mathbb{E}[X|Y] = g(Y) = g(Y) \sum_y 1(Y=y) =$$

$$= \sum_y g(y) \cdot 1(Y=y) = \sum_y 1(Y=y)\, \mathbb{E}[X|Y=y].$$

Ex. Toss a ~~fair~~ p-coin n times indep.

let $X_i = 1$ if i-th toss is H,
$\quad\quad\;\; 0$ o-w.

Define $Y_n = X_1 + \dots + X_n$. Want $\mathbb{E}[X_1 | Y_n] = ?$

Let $0 \le r \le n$. Need $\mathbb{E}[X_1 | Y_n = r]$.

$$\mathbb{E}[X_1 | Y_n = r] = 1 \cdot P(X_1 = 1 | Y_n = r)$$

$\underline{r=0}$ $\quad \mathbb{P}(X_1 = 1 | Y_n = 0) = 0$

$r \neq 0$ $\quad \mathbb{P}(X_1 = 1 | Y_n = r) = \dfrac{\mathbb{P}(X_1 = 1, Y_n = r)}{\mathbb{P}(Y_n = r)} =$

$$= \frac{\mathbb{P}(X_1 = 1, X_2 + \dots + X_n = r-1)}{\mathbb{P}(Y_n = r)} =$$

$$= \frac{\mathbb{P}(X_1 = 1) \cdot \mathbb{P}(X_2 + \dots + X_n = r-1)}{\mathbb{P}(Y_n = r)} =$$

$$= p \cdot \frac{\binom{n-1}{r-1} p^{r-1} (1-p)^{n-r}}{\binom{n}{r} p^r (1-p)^{n-r}} = \frac{r}{n}.$$

So $\mathbb{E}[X_1 | Y_n = r] = \frac{r}{n}$, which means that

$$\mathbb{E}[X_1 | Y_n] = \frac{Y_n}{n}.$$

$\left(g(r) = \frac{r}{n} \right)$

<u>Prop. of cond. exp.</u>

1) $\forall c \in \mathbb{R}$ $\quad \mathbb{E}[cX | Y] = c\,\mathbb{E}[X|Y]$ and $\mathbb{E}[c|Y] = c$.

2) $X_1, \dots, X_n$ r.v.'s

$$\mathbb{E}\left[\sum_{i=1}^{n} X_i \,\middle|\, Y\right] = \sum_{i=1}^{n} \mathbb{E}[X_i | Y].$$

3) $\mathbb{E}[\,\mathbb{E}[X|Y]\,] = \mathbb{E}[X]$.

<u>Proof</u> $\mathbb{E}[\,\mathbb{E}[X|Y]\,] = \sum_y \mathbb{P}(Y=y)\,\mathbb{E}[X|Y=y] =$

$$= \sum_y \mathbb{P}(Y=y) \sum_x x \cdot \mathbb{P}(X=x | Y=y) =$$

$$= \sum_x x \underbrace{\sum_y \mathbb{P}(X=x | Y=y) \cdot \mathbb{P}(Y=y)}_{\substack{\text{law of total prob.} \\ = \mathbb{P}(X=x)}}$$

$$= \sum_x x \cdot \mathbb{P}(X=x) = \mathbb{E}[X].$$

Properties of cond. expectation

4) If X and Y are indep., then

$$\mathbb{E}[X|Y] = \mathbb{E}[X].$$

Proof $\mathbb{E}[X|Y=y] = \sum_x x \cdot \mathbb{P}(X=x|Y=y) \overset{\text{indep.}}{=}$

$$= \sum_x x \cdot \mathbb{P}(X=x) = \mathbb{E}[X].$$

So $\mathbb{E}[X|Y] = \mathbb{E}[X]$.

5) Suppose Y and Z are indep. Then

$$\mathbb{E}[\,\mathbb{E}[X|Y]\,|Z] = \mathbb{E}[X].$$

Proof The condit. expect. $\mathbb{E}[X|Y]$ is a r.v. which is a function of Y, call it $g(Y)$.
Since Y and Z are indep., it follows that $g(Y)$ and Z are indep.
Apply (4) to get

$$\mathbb{E}[g(Y)|Z] = \mathbb{E}[g(Y)] = \mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}[X]$$

from prop. (3).

6) Let $h: \mathbb{R} \to \mathbb{R}$. Then

$$\mathbb{E}[h(Y)X \mid Y] = h(Y)\,\mathbb{E}[X \mid Y].$$

<u>Proof</u> $\mathbb{E}[h(Y) \cdot X \mid Y=y] = h(y) \cdot \mathbb{E}[X \mid Y=y]$.

<u>Corollary</u> $\mathbb{E}[\mathbb{E}[X \mid Y] \mid Y] = \mathbb{E}[X \mid Y]$.

<u>Random walks</u> A random/stochastic process $(X_n)_{n \in \mathbb{N}}$ is a sequence of random variables.

A <u>random walk</u> is a random process that can be expressed as

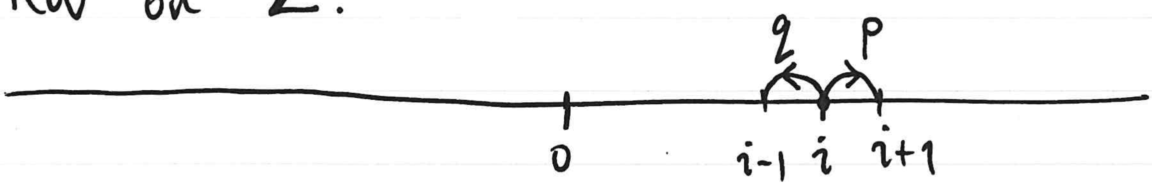
$$X_n = x + Y_1 + \dots + Y_n,$$ where

the Y_i's are independent, identically distr. random variables.
<u>iid</u>

Now we focus on the case where

$$Y_i = \begin{cases} 1 & \text{w.p. } p \\ -1 & \text{w.p. } q \end{cases}.$$

SRW on $\mathbb{Z}$.



Think of X_n as the fortune of a gambler who bets 1 at every step and he gets it back doubled w.p. p and he loses it w.p. q. His goal is to reach a before going bankrupt

Write $\mathbb{P}_x$ for $\mathbb{P}(\cdot \mid X_0 = x)$.

$\mathbb{P}_x(A) = \mathbb{P}(A \mid X_0 = x)$.

Want $\mathbb{P}_x(X_n \text{ hits } a \text{ before } 0) = h_x$.

Conditional on $Y_1 = +1$, X becomes a RW starting from $x+1$. Similarly with $Y_1 = -1$.

By the law of total prob.

$$h_x = \mathbb{P}_x(X \text{ hits } a \text{ bef. } 0,\ Y_1 = +1) + \mathbb{P}_x(X \text{ hits } a \text{ bef. } 0,\ Y_1 = -1)$$

$$= p \cdot h_{x+1} + q\, h_{x-1}.$$

So $\underset{0<x<a}{h_x = p\, h_{x+1} + q\, h_{x-1}}$, $h_0 = 0$, $h_a = 1$

• Case $p = \frac{1}{2} \Rightarrow h_x - h_{x-1} = h_{x+1} - h_x \quad \forall x$

$\Rightarrow h_x = \frac{x}{a}$.

• Case $p \neq \frac{1}{2}$. Try a solution of the form λ^x.

Then $p\lambda^2 - \lambda + q = 0 \Rightarrow \lambda = \begin{cases} 1 \\ \frac{q}{p} \end{cases}$

So the general solution will be of the form

$$h_x = A + B\left(\frac{q}{p}\right)^x.$$

$h_0 = 0$ and $h_a = 1 \Rightarrow \begin{array}{l} A + B = 0 \\ A + B\left(\frac{q}{p}\right)^a = 1 \end{array}$

$$\Rightarrow h_x = \frac{\left(\frac{q}{p}\right)^x - 1}{\left(\frac{q}{p}\right)^a - 1}.$$

Gambler's ruin estimates.

Expected time to absorption

Let T be the first time X hits $\{0,a\}$, i.e.

$$T = \min\{n \geq 0 : X_n \in \{0,a\}\}.$$

Want $\mathbb{E}_x[T] = \tau_x$

By the law of total expect. we get

$$\tau_x = 1 + p\,\tau_{x+1} + q\,\tau_{x-1}. \qquad 0<x<a$$

$$\tau_0 = \tau_a = 0$$

- Case $p=\frac{1}{2}$. Try Ax^2 as a solution. Then

$$\cancel{Ax^2} = 1 + pA(x+1)^2 + qA(x-1)^2 =$$
$$= 1 + \cancel{pAx^2} + \cancel{2pAx} + pA + \cancel{qAx^2} - \cancel{2qAx} + qA$$

$$\Rightarrow A = -1.$$

General sol. will be of the form

$$\tau_x = -x^2 + Bx + C \qquad \tau_0 = \tau_a = 0$$

$\Rightarrow \quad \tau_x = x(a-x).$

- Case $p \neq \frac{1}{2}$. Try cx as a solution

$$\Rightarrow \quad C = \frac{1}{q-p}.$$

General sol. : $\tau_x = \frac{1}{q-p} x + A + B \cdot \left(\frac{q}{p}\right)^x.$

$\tau_0 = \tau_a = 0$

$$\Rightarrow \quad \tau_x = \frac{1}{q-p} x - \frac{a}{q-p} \frac{\left(\frac{q}{p}\right)^x - 1}{\left(\frac{q}{p}\right)^a - 1}.$$

Probability generating functions

Let X be an integer valued r.v., i.e. $X \in \mathbb{N}$.

The prob. distr. of X is $p_r = \mathbb{P}(X=r), r \in \mathbb{N}$.

The prob. gen. fun. (pgf) is defined to be

$$p(z) = \sum_{r=0}^{\infty} p_r \cdot z^r = \mathbb{E}[z^X], \quad |z| \leq 1.$$

When $|z| \leq 1$, then $p(z)$ converges absolutely, because $|\sum p_r z^r| \leq \sum p_r |z|^r \leq \sum p_r = 1$. for $|z| \leq 1$.
So $p(z)$ is well-defined and has radius of converg. at least 1.

<u>Theorem</u> The pgf uniquely determines the distr. of X.

<u>Proof</u> ~~Prove by indu~~ Suppose

$$\sum_{r=0}^{\infty} p_r z^r = \sum_{r=0}^{\infty} q_r z^r.$$

<u>NTS</u> $p_r = q_r \quad \forall\, r \geq 0.$

Plug in $z=0$ to get $p_0 = q_0$.

Proceed by induction. Suppose $p_r = q_r \quad \forall\, r \leq n.$

Then $$\sum_{r=n+1}^{\infty} p_r z^r = \sum_{r=n+1}^{\infty} q_r z^r.$$

Divide by z^{n+1} and then take limit as $z \to 0$ to get
$$p_{n+1} = q_{n+1} .$$ $\square$

Probability generating functions

X takes values in $\mathbb{N}$

$p_r = \mathbb{P}(X = r)$, $r \in \mathbb{N}$

pgf $\quad p(z) = \mathbb{E}[z^X] = \sum_{r=0}^{\infty} p_r z^r, \quad |z| \le 1.$

Thm $\quad \lim_{z \uparrow 1} p'(z) = p'(1-) = \mathbb{E}[X].$

Pf Assume first that $\mathbb{E}[X] < \infty$.

Let $0 < z < 1$ we can differentiate term by term to get

$$p'(z) = \sum_{r=0}^{\infty} r p_r z^{r-1} \le \sum_{r=0}^{\infty} r p_r = \mathbb{E}[X].$$

Since $p'(z)$ is increasing in z, we can take $\lim_{z \uparrow 1}$ to get

$$\lim_{z \uparrow 1} p'(z) \le \mathbb{E}[X].$$

Let $\varepsilon > 0$ and choose N large enough s.t.

$$\sum_{r=0}^{N} r p_r \geqslant \mathbb{E}[X] - \varepsilon.$$

We have $p'(z) \geqslant \sum_{r=0}^{N} r p_r z^{r-1}$.

$$\lim_{z \uparrow 1} p'(z) \geqslant \sum_{r=0}^{N} r p_r \geqslant \mathbb{E}[X] - \varepsilon .$$

This is true $\forall\, \varepsilon > 0$. So $\lim_{z \uparrow 1} p'(z) = p'(1-) = \mathbb{E}[X]$.

If $\mathbb{E}[X] = \infty$, then $\forall M$ ~~take~~ N large s.t.

$$\sum_{r=0}^{N} r p_r \geqslant M .$$

As before $\lim_{z \uparrow 1} p'(z) \geqslant M \quad \forall M$, so

$$\lim_{z \uparrow 1} p'(z) = p'(1-) = \infty . \qquad \square$$

Similarly, $p''(1-) = \lim_{z \uparrow 1} p''(z) = \mathbb{E}[X(X-1)]$.

$\forall k$, $p^{(k)}(1-) = \mathbb{E}[X(X-1) \cdots (X-k+1)]$.

In particular, $\mathrm{Var}(X) = p''(1-) + p'(1-) - (p'(1-))^2$.

We also have

$$\mathbb{P}(X=n) = \frac{1}{n!}\left(\frac{d}{dz}\right)^n \Big|_{z=0} p(z)$$

~~Examples~~

Let $X_1, \dots, X_n$ independent r.v.'s with pgf's q_i, $i=1,\dots,n$.

Let $p(z) = \mathbb{E}\left[z^{X_1+\dots+X_n}\right] \underset{\downarrow}{=} \mathbb{E}\left[z^{X_1}\right]\dots\mathbb{E}\left[z^{X_n}\right]$

$\left[\text{Recall } \mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)]\cdot\mathbb{E}[g(Y)] \text{ for } X \perp\!\!\!\perp Y\right]$

So $p(z) = \prod_{i=1}^{n} q_i(z)$.

If the X_i's all have the same distr., i.e. same pgf q, then $p(z) = (q(z))^n$.

Examples

1) Let $X \sim \text{Bin}(n,p)$.

$$p(z) = \mathbb{E}\left[z^X\right] = \sum_{r=0}^{n} \binom{n}{r} z^r p^r (1-p)^{n-r} = (1-p+zp)^n.$$

Let $X \sim Bin(n,p)$ and $Y \sim Bin(m,p)$ indep.

Then $\mathbb{E}[z^{X+Y}] = \mathbb{E}[z^X] \cdot \mathbb{E}[z^Y] =$

$$= (1-p+pz)^n \cdot (1-p+pz)^m = (1-p+pz)^{n+m}.$$

So $X+Y \sim Bin(n+m, p)$.

2) Let $X \sim Geo(p)$

Then $p(z) = \mathbb{E}[z^X] = \sum_{r=0}^{\infty} (1-p)^r p \cdot z^r = \dfrac{p}{1-z(1-p)}$.

3) Let $X \sim Poi(\lambda)$

$$p(z) = \sum_{r=0}^{\infty} e^{-\lambda} \cdot \frac{\lambda^r}{r!} \cdot z^r = e^{-\lambda} \cdot e^{\lambda z} = e^{-\lambda(1-z)}$$

$X \sim Poi(\lambda)$, $Y \sim Poi(\mu)$, $X \perp\!\!\!\perp Y$

$$\mathbb{E}[z^{X+Y}] = e^{-\lambda(1-z)} \cdot e^{-\mu(1-z)} = e^{-(\lambda+\mu)(1-z)}.$$

So $X+Y \sim Poi(\lambda+\mu)$.

Sum of a random number of r.v.'s

Let $X_1, X_2, \ldots$ be iid and let N be an indep. r.v. taking values in $\mathbb{N}$.

Define $S_n = X_1 + \ldots + X_n$.

Now $S_N = X_1 + \ldots + X_N$ means

$$\forall \omega \in \Omega \quad S_N(\omega) = X_1(\omega) + \ldots + X_{N(\omega)}(\omega) = \sum_{i=1}^{N(\omega)} X_i(\omega).$$

Let q be the pgf of ~~X~~ N and p the pgf of X_1.

Then $r(z) = \mathbb{E}[z^{S_N}] = \mathbb{E}[z^{X_1 + \ldots + X_N}] =$

$$= \sum_{n=0}^{\infty} \mathbb{E}\left[z^{X_1 + \ldots + X_N} \cdot 1(N=n)\right] =$$

$$= \sum_{n=0}^{\infty} \mathbb{E}\left[z^{X_1 + \ldots + X_n} \cdot 1(N=n)\right] \overset{\text{indep. } N \perp\!\!\!\perp (X_i)}{=}$$

$$= \sum_{n=0}^{\infty} \mathbb{E}\left[z^{X_1 + \ldots + X_n}\right] \cdot \mathbb{P}(N=n) \overset{\text{indep. of } X_i\text{'s}}{=}$$

$$= \sum_{n=0}^{\infty} (p(z))^n \, \mathbb{P}(N=n) = \mathbb{E}\left[(p(z))^N\right] = q(p(z)).$$

Another way using condit. expect.

$$r(z) = \mathbb{E}[z^{S_N}] = \mathbb{E}\left[\mathbb{E}[z^{S_N} \mid N]\right] =$$

$$= \mathbb{E}\left[\mathbb{E}[z^{X_1+\dots+X_N} \mid N]\right]$$

$$\forall n \quad \mathbb{E}[z^{X_1+\dots+X_N} \mid N=n] = \left(\mathbb{E}[z^{X_1}]\right)^n = (p(z))^n$$

So $\mathbb{E}[z^{X_1+\dots+X_N} \mid N] = (p(z))^N$.

So $r(z) = \mathbb{E}\left[(p(z))^N\right] = q(p(z))$.

So $\mathbb{E}[S_N] = \lim\limits_{z \uparrow 1} r'(z) = r'(1-)$

$r'(z) = q'(p(z)) \cdot p'(z)$. So $\mathbb{E}[S_N] = q'(p(1-)) \cdot p'(1-) \Rightarrow$

$$\mathbb{E}[S_N] = \mathbb{E}[N] \cdot \mathbb{E}[X_1].$$

Similarly, $\mathrm{Var}(S_N) = \mathbb{E}[N] \cdot \mathrm{Var}(X_1) + \mathrm{Var}(N) \cdot \left(\mathbb{E}[X_1]\right)^2$.

Branching processes

Bienaymé – Galton-Watson processes
independ. 1874
earlier.

$(X_n : n \geqslant 0)$ random process

$X_0 = 1$, X_n = number of individuals in generation n.

The individual at time 0 produces a random number of offspring with prob. distr.

$$\mathbb{P}(X_1 = k) = g_k, \quad k = 0, 1, 2, \ldots$$

Every individual in the 1^{st} gener. produces an indep. number of offspring (indep. for dif. indiv. in gen. n) with the same distr.
We call the distr. of X_1 the offspring distr. We continue in the same way.

$X_0 = 1$

$$X_{n+1} = \begin{cases} Y_{1,n} + \dots + Y_{X_n,n} & \text{if } X_n \geq 1 \\ 0 & \text{if } X_n = 0 \end{cases}$$

where $(Y_{k,n}, k \geq 1, n \geq 0)$ are iid with the same distr. as X_1.

$Y_{k,n}$ = # of offspring of k-th indiv. in gener. n.

<u>Theorem</u> $\mathbb{E}[X_n] = (\mathbb{E}[X_1])^n$. $\forall n \geq 1$.

<u>Proof</u> $n=1$ ✓ Set $\mu = \mathbb{E}[X_1]$.

Suppose $\mathbb{E}[X_n] = \mu^n$. <u>NTS</u> for $n+1$.

$\mathbb{E}[X_{n+1}] = \mathbb{E}[\mathbb{E}[X_{n+1} \mid X_n]]$.

$\forall m \quad \mathbb{E}[X_{n+1} \mid X_n = m] =$

$$= \mathbb{E}[Y_{1,n} + \dots + Y_{m,n} \mid X_n = m] = m \cdot \mathbb{E}[X_1]$$

So $\mathbb{E}[X_{n+1} \mid X_n] = X_n \cdot \mathbb{E}[X_1] = \mu \cdot X_n.$

Hence $\mathbb{E}[X_{n+1}] = \mathbb{E}[\mu \cdot X_n] = \mu \cdot \mathbb{E}[X_n] = \mu^{n+1}.$ $\square$

<u>Theorem</u> Set $G(z) = \mathbb{E}[z^{X_1}]$ and

$G_n(z) = \mathbb{E}[z^{X_n}]$. Then

$$\cancel{G_{n+1}(z)} \; G_{n+1}(z) = G(G_n(z)) =$$

$$= G(G(\dots G(z))\dots) = G_n(G(z)).$$

<u>Proof</u> $n = 1$, $G_1(z) = \mathbb{E}[z^{X_1}] = G(z)$.

~~n ⇒ n+1~~.

$$G_{n+1}(z) = \mathbb{E}[z^{X_{n+1}}] = \mathbb{E}[\underbrace{\mathbb{E}[z^{X_{n+1}} \mid X_n]}].$$

$$\mathbb{E}[z^{X_{n+1}} \mid X_n = m] = \mathbb{E}[z^{Y_{1,n} + \dots + Y_{m,n}} \mid X_n = m] =$$

$\overset{\text{indep.}}{=} \left(\mathbb{E}[z^{X_1}]\right)^m$.

So $\mathbb{E}[z^{X_{n+1}} \mid X_n] = \left(\mathbb{E}[z^{X_1}]\right)^{X_n} = \left(G(z)\right)^{X_n}$.

Hence $\mathbb{E}[z^{X_{n+1}}] = \mathbb{E}\left[\left(G(z)\right)^{X_n}\right] = G_n(G(z))$. $\square$

<u>Prob. of extinction</u>

Write $q = \mathbb{P}(X_n = 0 \text{ for some } n \geq 1)$.

and define $q_n = \mathbb{P}(X_n = 0)$.

let $A_n = \{X_n = 0\}$. Then $A_n \subseteq A_{n+1}$ and
$A_n \uparrow \bigcup A_n = \{X_n = 0 \text{ for some } n \geq 1\}$.

So by continuity of prob. meas. we get

$$q_n = \mathbb{P}(A_n) \uparrow \mathbb{P}(\bigcup A_n) = q \quad \text{as} \quad n \to \infty.$$

So $q = \lim_{n \to \infty} q_n$.

$$q_{n+1} = \mathbb{P}(X_{n+1} = 0) = G_{n+1}(0) = G(G_n(0)) = G(q_n).$$

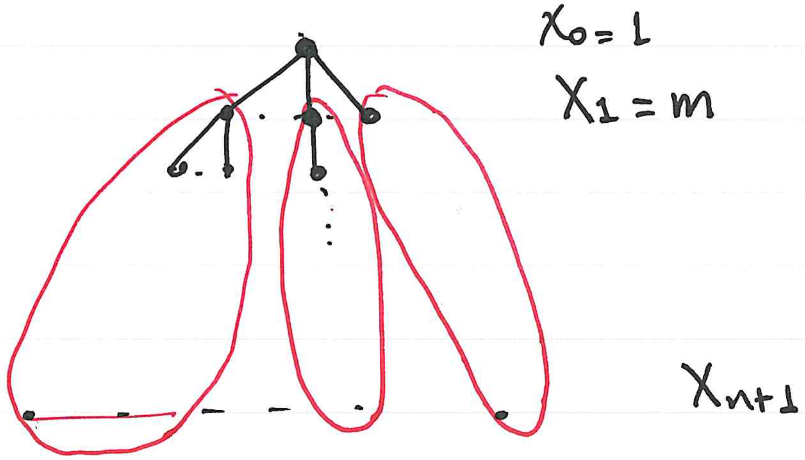
So $\quad q_{n+1} = G(q_n) \quad \forall\, n.$

We know G is continuous [and increasing in $[0,1]$], so we can take limits as $n \to \infty$ to get

$$q = G(q).$$

~~Note~~ <u>Another way</u> to get $q_{n+1} = G(q_n)$.

Instead of conditioning on X_n we are going to condition on X_1.



Condition on $X_1 = m$. Then we can write

$X_{n+1} = X_n^{(1)} + \dots + X_n^{(m)}$, where

$(X_j^{(i)})$ are iid branching processes with the same offspring distr.

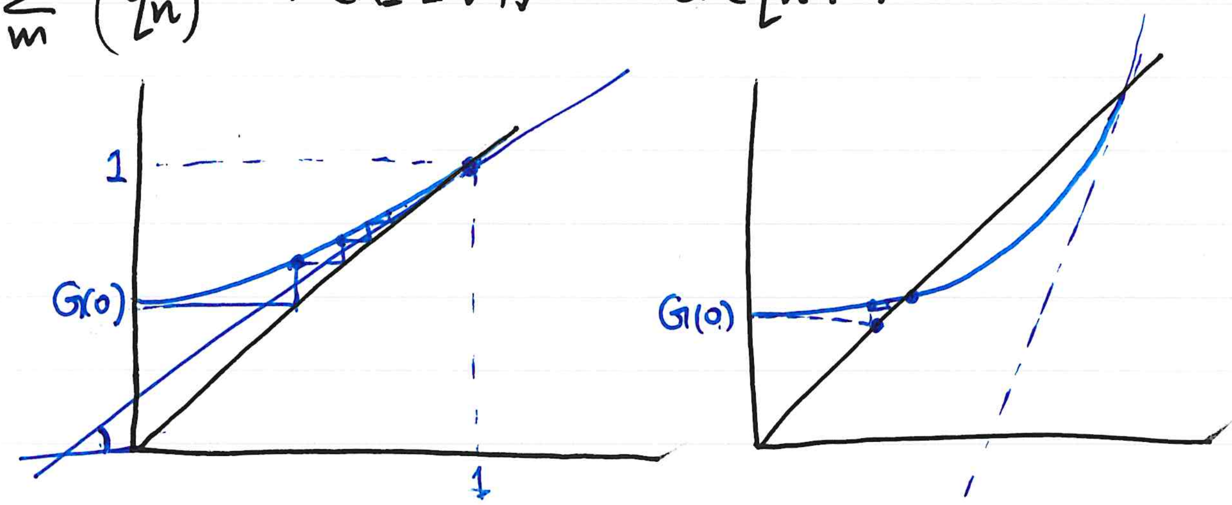
So $q_{n+1} = \mathbb{P}(X_{n+1} = 0) =$

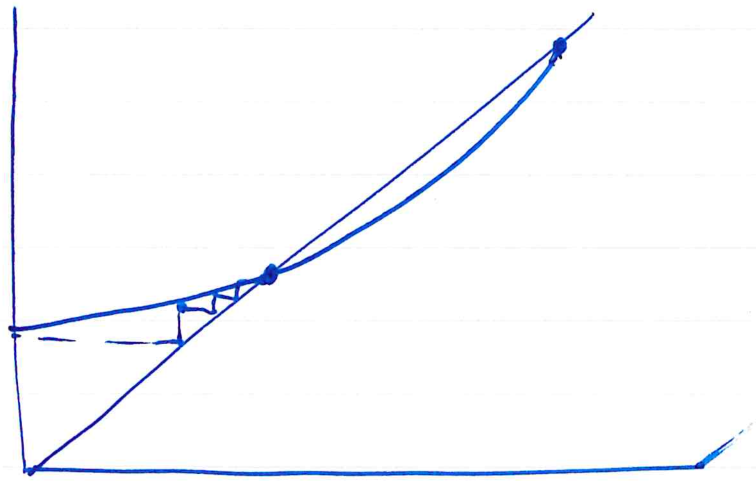
$$= \sum_m \mathbb{P}(X_{n+1} = 0 \mid X_1 = m) \cdot \mathbb{P}(X_1 = m) =$$

$$= \sum_m \mathbb{P}(X_n^{(1)} + \dots + X_n^{(m)} = 0 \mid X_1 = m) \cdot \mathbb{P}(X_1 = m)$$

$$= \sum_m \mathbb{P}(X_n^{(1)} = \dots = X_n^{(m)} = 0 \mid X_1 = m) \cdot \mathbb{P}(X_1 = m)$$

$$= \sum_m (q_n)^m \cdot \mathbb{P}(X_1 = m) = G(q_n).$$





<u>Theorem</u> The extinction prob. q is the smallest non-negative solution to $q = G(q)$. Provided $\mathbb{P}(X_1 = 1) < 1$ we have

$$q < 1 \quad \text{iff} \quad \mu > 1 . \quad (\mu = \mathbb{E}[X_1]).$$

<u>Proof</u> Let t be the smallest non-negative sol. to $q = G(q)$. We will show $q = t$.

$q_0 = 0 \leq t$. Suppose $q_n \leq t$. <u>NTS</u> $n+1$.

$q_{n+1} = G(q_n) \leq G(t)$ because G is incr. in $[0,1]$ & $q_n \leq t$.

So $q_{n+1} \leq G(t) = t$. Taking limits as $n \to \infty$

shows $q \leq t \Rightarrow q = t$.

Consider $H(z) = G(z) - z$.

Then $H''(z) = \sum_{r=2}^{\infty} r(r-1) g_r z^{r-2}$

Assume $g_0 + g_1 < 1$. [If not, then $\mathbb{P}(X_1 \leq 1) = 1$ and we can exclude $\mu = 1$ because $\cancel{\mathbb{P}(X_1 = 1) < 1}$.

$\mathbb{P}(X_1 = 1) < 1$. Then $G(t) = 1 - \mu + \mu t$ and $\mu < 1 \Rightarrow t = 1$ is the unique solution to $G(t) = t$!]

So $H''(z) > 0 \quad \forall z \in (0,1)$.

This means that $H'(z)$ is strictly increas. in $[0,1]$ which by Rolle's theorem implies that H can have at most one solution different from 1 in $[0,1]$.

- H has no root in $[0,1)$. Since $H(0) = G(0) \geq 0 \Rightarrow H(z) \geq 0 \quad \forall z \in [0,1]$

$$\Rightarrow H'(1-) = \lim_{z \uparrow 1} \frac{\overset{=0}{H(1)} - H(z)}{1-z} \leq 0, \text{ because } H(1) = 0.$$

$$H'(1-) = G'(1-) - 1 \Rightarrow G'(1-) \leq 1 \Rightarrow \mu \leq 1.$$

Branching processes

$X_0 = 1$ $\quad X_n$ = # of indiv. in gen. n.

X_1 : offspring distr. $\quad g_r = \mathbb{P}(X_1 = r)$

$G(z) = \mathbb{E}[z^{X_1}]$.

$q = \mathbb{P}(X_n = 0$ for some $n \geq 1)$

Theorem q is the minimal non-neg. sol. to $z = G(z)$. $q \lneq 1$ iff $\mu = \mathbb{E}[X_1] \leq 1$, provided $\mathbb{P}(X_1 = 1) < 1$.

Pf of 2nd part $\quad H(z) = G(z) - z$

$$H''(z) = \sum_{r=2}^{\infty} r(r-1) g_r z^{r-2}$$

Assume $g_0 + g_1 < 1 \Rightarrow H'' > 0$ for $z \in (0,1)$.
$\Rightarrow H'$ strictly incr. $\Rightarrow$ (Rolle's thm) H can have at most 1 root different from 1 in $[0,1]$.

- no other root. $\Rightarrow \mu \leq 1$. (last time)

- let r be another root in $[0,1)$ of H.
So $G(r) = r$.

We assumed $g_0 + g_1 < 1 \Rightarrow G'$ is strictly incr.

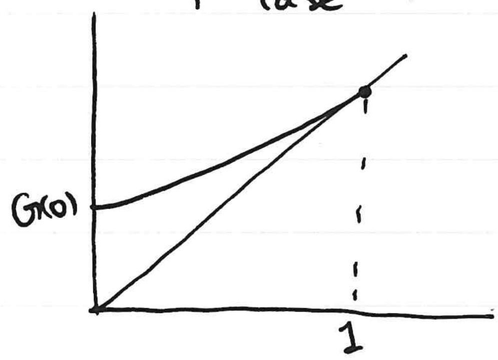
$\Rightarrow H'$ must have ~~another~~ a root in $(r, 1)$ (Rolle's)

Let it be $z \Rightarrow G'(z) = 1$.

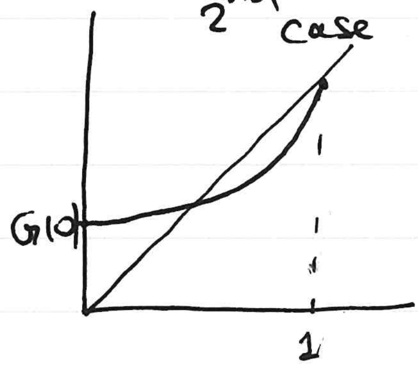
$G' \uparrow$ and $z < 1 \Rightarrow \underset{\mu}{\underset{\|}{G'(1-)}} > G'(z) = 1$.

$\Rightarrow \mu > 1$.

1st case



2nd case



$\square$

Continuous random variables

$(\Omega, \mathcal{F}, \mathbb{P})$ $\quad X : \Omega \to \mathbb{R}$ $\quad \forall x \in \mathbb{R}$

$$\{X \le x\} = \{\omega : X(\omega) \le x\} \in \mathcal{F}.$$

Recall $\forall x \in \mathbb{R}$ we defined

$F(x) = \mathbb{P}(X \le x)$. prob. distr. function

Prop. 1) $x \le y \Rightarrow F(x) \le F(y)$

since $\{X \le x\} \subseteq \{X \le y\}$.

2) $\forall\ a < b$, $a, b \in \mathbb{R}$

$$\mathbb{P}(a < X \le b) = F(b) - F(a).$$

Pf $\mathbb{P}(a < X \le b) = \mathbb{P}(\{X \le b\} \cap \{X > a\}) =$

$\mathbb{P}(\{X \le b\}) - \mathbb{P}(\{X \le b\} \cap \{X \le a\}) =$

$= \mathbb{P}(X \le b) - \mathbb{P}(X \le a) = F(b) - F(a).$

3) F is right continuous and the left limits always exist

$$F(x-) = \lim_{y \uparrow x} F(y) \le F(x) \quad (1).$$

Pf NTS. $\lim_{n\to\infty} F(x + \frac{1}{n}) = F(x)$. (right ct)

$$A_n = \{ x < X \le x + \tfrac{1}{n} \}.$$

$A_n \downarrow$, i.e. $A_{n+1} \subseteq A_n$ and $\bigcap_n A_n = \emptyset$

ct of prob. meas. we get

$$\mathbb{P}(A_n) \searrow 0 \Rightarrow F(x + \tfrac{1}{n}) \xrightarrow[n\to\infty]{} F(x) .$$

$$F(x-) = \lim_{n\to\infty} F(x - \tfrac{1}{n})$$

$$F(x - \tfrac{1}{n}) = \mathbb{P}(X \le x - \tfrac{1}{n})$$

$$\bigcup_n \{ X \le x - \tfrac{1}{n} \} = \{X < x\}.$$

So $F(x - \frac{1}{n}) \to \mathbb{P}(X < x)$

$\Rightarrow$ $F(x-) = \mathbb{P}(X < x)$.

4) $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$.

<u>Def.</u> X is a <u>continuous</u> r.v. if F is a continuous function, which means

$$F(x-) = F(x) \Rightarrow$$

$$\mathbb{P}(X < x) = \mathbb{P}(X \leq x) \quad \forall x \in \mathbb{R} \Rightarrow$$

$$\mathbb{P}(X = x) = 0 \qquad \forall x \in \mathbb{R}.$$

In this course we will restrict to the case where F is not only continuous but also differentiable.

Set $f(x) = F'(x)$. Call f the prob. density function of X.

<u>Prop. of f</u>

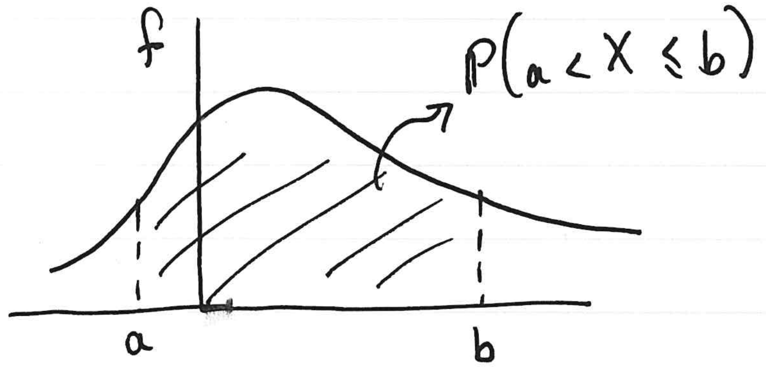
i) $f(x) \geq 0$ ii) $\int_{-\infty}^{\infty} f(x)\,dx = 1$.

We have $F(x) = \int_{-\infty}^{x} f(y)\,dy$

<u>Intuitively</u> Δx small

$$\mathbb{P}(x < X \leq x + \Delta x) = \int_x^{x+\Delta x} f(y)\,dy \approx f(x)\,\Delta x .$$

$f(x)$ is no longer a prob. as in the discrete world, but we can think of it as the prob. that X lies in a small interval around x.



Let X be a cts r.v. with a density f.
Let X_+ and X_- as before

$$X_+ = \max(X, 0) \quad \text{and} \quad X_- = \max(-X, 0)$$

~~Define~~

$\cancel{\mathbb{E}[X_+] = \int}$ Let $Y \geq 0$. Define

$$\mathbb{E}[Y] = \int_0^\infty y\, f(y)\,dy . \quad \text{and} \quad g \geq 0$$

define $\mathbb{E}[g(Y)] = \int_{-\infty}^{\infty} g(y)\, f(y)\,dy$ for any Y.

If at least 1 of $\mathbb{E}[X_+]$ or $\mathbb{E}[X_-]$ is finite, define

$$\mathbb{E}[X] = \mathbb{E}[X_+] - \mathbb{E}[X_-] = \int_{-\infty}^{\infty} x\, f(x)\, dx$$

because
$$\mathbb{E}[X_+] = \int_0^{\infty} x\, f(x)\, dx$$
$$\mathbb{E}[X_-] = \int_{-\infty}^{0} (-x) f(x)\, dx$$

The expectation is again a linear function.

Let $X \geq 0$. Then

$$\mathbb{E}[X] = \int_0^{\infty} \mathbb{P}(X \geq x)\, dx .$$

<u>1st pf</u> Write $X = \int_0^{\infty} 1(X \geq x) dx$

$$\mathbb{E}[X] \overset{\substack{\text{not justified}\\ \uparrow}}{=} \int_0^{\infty} \mathbb{P}(X \geq x) dx = \int_0^{\infty} (1 - F(x)) dx .$$

<u>2nd pf</u> $\mathbb{E}[X] = \int_0^{\infty} x\, f(x)\, dx =$

$$= \int_0^{\infty} \left(\int_0^{x} dy \right) f(x) dx = \int_0^{\infty} dy \int_y^{\infty} f(x) dx = \int_0^{\infty} \mathbb{P}(X \geq y) dy.$$

Continuous random variables

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

Recall $F(x) = \mathbb{P}(X \leq x)$, $f(x) = F'(x)$

$\mathbb{P}(X = x) = 0 \quad \forall x.$

Examples

1) Uniform distribution

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

Define $f(x) = \begin{cases} \frac{1}{b-a} & x \in [a,b] \\ 0 & \text{otherwise} \end{cases}$

Let X have density f.

Then $\underset{x \in [a,b]}{\mathbb{P}(X \leq x)} = \int_a^x f(u)\,du = \frac{x-a}{b-a}.$

We write $X \sim U[a,b]$

$$\mathbb{E}[X] = \int x f(x)\,dx = \int_a^b \frac{x}{b-a}\,dx = \frac{a+b}{2}.$$

2) Exponential distr.

$f(x) = \lambda e^{-\lambda x}$, $x > 0$, $\lambda > 0$.

$X \sim \mathrm{Exp}(\lambda)$.

$F(x) = \mathbb{P}(X \leq x) = \int_0^x \lambda e^{-\lambda y} dy = 1 - e^{-\lambda x}$.

$\mathbb{E}[X] = \frac{1}{\lambda}$.

Exponential as a limit of geometrics

Let $T \sim \mathrm{Exp}(\lambda)$. Define $T_n = \lfloor nT \rfloor$.

$\mathbb{P}(T_n \geq k) = \mathbb{P}(T \geq k/n) = e^{-\frac{\lambda k}{n}} = \left(e^{-\frac{\lambda}{n}}\right)^k$.

So T_n is a geometric r.v. with parameter $p_n = 1 - e^{-\frac{\lambda}{n}}$.

As $n \to \infty$ $p_n \sim \frac{\lambda}{n}$ and $\frac{T_n}{n} \to T$.

So the exponential arises as the limit of a rescaled geometric.

Memoryless property

Let $T \sim \mathrm{Exp}(\lambda)$. Then $\forall\, s,t \geq 0$

$$\mathbb{P}(T > s+t \mid T > s) = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = \mathbb{P}(T>t)$$

Let T be a positive r.v. not identically equal to 0 or ∞. Then T has the memoryless property iff T is exponential.

Pf of $\Rightarrow$ Assumption $\mathbb{P}(T>s+t) = \mathbb{P}(T>s)\,\mathbb{P}(T>t)$ $\forall\, s,t$

Set $g(t) = \mathbb{P}(T>t)$

Then $g(t+s) = g(t)\cdot g(s)$ $\forall\, s,t \geq 0$

$\forall\, t \geq 0,\ \forall\, m \in \mathbb{N}$ $g(mt) = (g(t))^m$.

$$g\left(\frac{m}{n}\right)^n = g(m). \quad \forall\, m,n \in \mathbb{N}.$$

and $g(m) = (g(1))^m$

~~Set~~ $g(1) = \mathbb{P}(T>1) > 0$, because $T \neq 0, \infty$

We can set $\lambda = -\log \mathbb{P}(T>1)$.

We have thus proved that

$$\mathbb{P}(T>t) = e^{-\lambda t} \quad \forall\, t \in \mathbb{Q}_+.$$

$\forall\, t>0 \quad \exists\, r,s \in \mathbb{Q}$ s.t. $|r-s| \leq \varepsilon$, $r \leq t \leq s$.

and by the non-increas. property of $\mathbb{P}(T>t)$ and $e^{-\lambda t}$ we get

$$e^{-\lambda s} \leq \mathbb{P}(T>t) \leq e^{-\lambda r}$$

Letting $\varepsilon \to 0$ finishes the proof. $\square$

Theorem Let X be a cts r.v. with density f. Let g be a cts function, ~~st~~ g is either strictly increasing or strictly decreasing with g^{-1} differentiable. Then $g(X)$ is a cts r.v. with density

$$f(g^{-1}(x)) \cdot \left| \frac{d}{dx} g^{-1}(x) \right| .$$

Proof $\mathbb{P}(g(X) \le x)$ and differ.

Assume $g \nearrow$. Then

$$\mathbb{P}(g(X) \le x) = \mathbb{P}(X \le g^{-1}(x)) = F(g^{-1}(x))$$

$$\frac{d}{dx} F(g^{-1}(x)) = f(g^{-1}(x)) \cdot \frac{d}{dx} g^{-1}(x) .$$

Assume $g \swarrow$.

$$\mathbb{P}(g(X) \le x) = \mathbb{P}(X \ge g^{-1}(x)) = 1 - F(g^{-1}(x)),$$

because $\mathbb{P}(X = g^{-1}(x)) = 0$ (X is cts)

$$\frac{d}{dx}\left(1 - F(g^{-1}(x))\right) = -f(g^{-1}(x)) \cdot \frac{d}{dx} g^{-1}(x). \qquad \square$$

3) <u>Normal</u> distr.

Let μ and σ be 2 parameters
$-\infty < \mu < \infty$, $\sigma > 0$.

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

Check f is a density.

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

$u = \frac{x-\mu}{\sigma}$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du = 2\int_0^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du = I.$$

$$I^2 = \frac{2}{\pi} \int_0^{\infty}\int_0^{\infty} e^{-\frac{(u^2+v^2)}{2}} du\, dv$$

Polar coordinates. Set $u = r\cos\vartheta$, $v = r\sin\vartheta$

$$\Rightarrow I^2 = \frac{2}{\pi}\int_0^\infty \int_0^{\frac{\pi}{2}} e^{-\frac{r^2}{2}}\, r\, dr\, d\vartheta = 1$$

$$\Rightarrow I = 1.$$

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

$$= \underbrace{\int_{-\infty}^{\infty} \frac{x-\mu}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx}_{\sigma \underbrace{\int_{-\infty}^{\infty} \frac{u\, e^{-\frac{u^2}{2}}}{\sqrt{2\pi}}\, du}_{0 \text{ because integrand is odd}}} + \mu \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx}_{1}$$

So $\mathbb{E}[X] = \mu$.

$$\mathrm{Var}(X) = \mathbb{E}\left[(X-\mu)^2\right] = \int_{-\infty}^{\infty} \frac{(x-\mu)^2}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx$$

$u = \frac{x-\mu}{\sigma}$

$$= \sigma^2 \int_{-\infty}^{\infty} \frac{u^2}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du = \sigma^2$$ (integration by parts)

So $Var(X) = \sigma^2$.

We denote the normal distr. with these parameters $N(\mu, \sigma^2)$.

When $\mu = 0$ and $\sigma^2 = 1$ we call

$N(0,1)$ the standard normal.

$X \sim N(0,1)$ we write

$$\Phi(x) = \int_{-\infty}^{x} \frac{e^{-\frac{u^2}{2}}}{\sqrt{2\pi}} du \quad \text{and} \quad \varphi(x) = \Phi'(x) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}.$$

Since $\varphi(x) = \varphi(-x) \Rightarrow$

$$\Phi(x) + \Phi(-x) = 1.$$

i.e. $\mathbb{P}(X \leq x) = 1 - \mathbb{P}(X \leq -x)$.

Let $X \sim \mathcal{N}(\mu, \sigma^2)$ and let $a, b \in \mathbb{R}$, $a \neq 0$

Set $Y = aX + b$. Then $\mathbb{E}[Y] = a\mu + b$

$\mathrm{Var}(Y) = a^2\sigma^2$.

We will prove $Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$.

Normal distribution

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

$X \sim f$, $X \sim \mathcal{N}(\mu, \sigma^2)$

Let $a \neq 0, b \in \mathbb{R}$. Set $Y = aX + b$

NTS $Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$.

Let $g(x) = ax + b$. Then $Y = g(X)$

From last time

$$f_Y(y) = f_X(g^{-1}(y)) \cdot \left| \frac{d}{dy} g^{-1}(y) \right| =$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(g^{-1}(y) - \mu)^2}{2\sigma^2}\right) \left| \frac{d}{dy} g^{-1}(y) \right|$$

$$g^{-1}(y) = \frac{y-b}{a} \Rightarrow f_Y(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-b-\mu a)^2}{2a^2\sigma^2}\right) \times$$

$$\times \frac{1}{|a|}$$

So $f_Y(y) = \frac{1}{\sqrt{2\pi a^2 \sigma^2}} \exp\left(-\frac{(y-b-a\mu)^2}{2a^2\sigma^2}\right)$.

So if $X \sim \mathcal{N}(\mu, \sigma^2)$, then

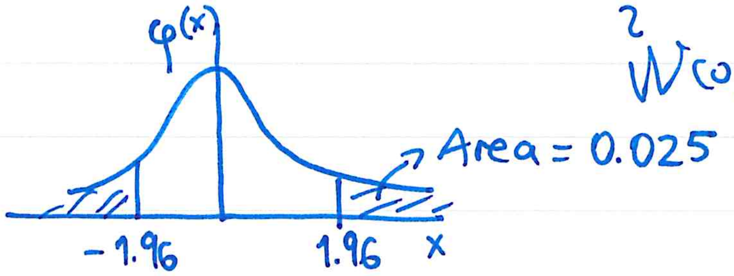
$$\frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1).$$

more than 95% of the normal distribution lies within 2 standard deviations of the mean.

Let $X \sim \mathcal{N}(\mu, \sigma^2)$

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

$$= \mathbb{P}\left(\left|\underbrace{\frac{X-\mu}{\sigma}}\right| < 2\right) \geq \mathbb{P}\left(\left|\underbrace{\frac{X-\mu}{\sigma}}_{\sim \mathcal{N}(0,1)}\right| < 1.96\right) = 0.95.$$



Def. Let X be a cts r.v. with density f. The median m of X is the number satisfying

$$\mathbb{P}(X \geq m) = \mathbb{P}(X \leq m) = \frac{1}{2}.$$

In other words $\displaystyle\int_m^{\infty} f(x)dx = \int_{-\infty}^{m} f(x)dx = \frac{1}{2}.$

If $X \sim \mathcal{N}(\mu, \sigma^2)$, then the median is equal to μ.

Let X have density f. Then we know $\forall x \in \mathbb{R}$

$$\mathbb{P}(X \leq x) = \int_{-\infty}^{x} f(y)dy.$$

It can be proved that this generalises to an "arbitrary" set $B \subseteq \mathbb{R}$, i.e.

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

Multivariate density functions

$X = (X_1, \dots, X_n) \in \mathbb{R}^n$ X_i are r.v.'s.

X is said to have density f if

$$F(x_1, \dots, x_n) = \mathbb{P}(X_1 \le x_1, \dots, X_n \le x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f(y_1, \dots, y_n)\, dy_1 \cdots dy_n .$$

Then $f(x_1, \dots, x_n) = \dfrac{\partial^n F(x_1, \dots, x_n)}{\partial x_1 \cdots \partial x_n}$.

It generalises to "arbitrary" sets $B \subseteq \mathbb{R}^n$

$$\mathbb{P}(X \in B) = \int_B f(x_1, \dots, x_n)\, dx_1 \cdots dx_n .$$

Let $g : \mathbb{R}^n \to \mathbb{R}_+$. Define

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}^n} g(x) f(x)\, dx$$

Independence

We say that $X_1, \dots, X_n$ are indep. if $\forall x_1, \dots, x_n \in \mathbb{R}$

$$\mathbb{P}(X_1 \le x_1, \dots, X_n \le x_n) = \mathbb{P}(X_1 \le x_1) \cdots \mathbb{P}(X_n \le x_n).$$

Theorem Let $X = (X_1, \dots, X_n)$ have density f.

(a) Suppose $X_1, \dots, X_n$ are indep. with densities (f_i). Then

$$f(x_1, \dots, x_n) = \prod_{i=1}^{n} f_i(x_i) \quad . \quad (*)$$

(b) Suppose that f factorises as in $(*)$ for some non-negative functions (f_i). Then $X_1, \dots, X_n$ are indep. and have density functions proportional to the f_i's.

Proof (a) $\mathbb{P}(X_1 \le x_1, \dots, X_n \le x_n) \overset{\text{ind.}}{=} \prod_{i=1}^{n} \mathbb{P}(X_i \le x_i) =$

$$= \prod_{i=1}^{n} \int_{-\infty}^{x_i} f_i(y_i)\,dy_i = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} \prod f_i(y_i)\,dy_1 \dots dy_n .$$

(b) By moving constants among the f_i's in the product we can assume that $\forall i$

$$\int_{-\infty}^{\infty} f_i(x)dx = 1.$$

$$\mathbb{P}(X \in B_1 \times \dots \times B_n) = \int_{B_1} \dots \int_{B_n} \prod_{i=1}^{n} f_i(x_i)\, dx_1 \dots dx_n$$

Fix i. For all $j \neq i$, let $B_j = \mathbb{R}$.

Then $\mathbb{P}(X_i \in B_i) = \mathbb{P}(X_i \in B_i,\ X_j \in B_j\ \ \forall j \neq i)$

$$= \int_{-\infty}^{\infty} \cdots \int_{B_i} \cdots \int_{-\infty}^{\infty} \prod f_j(x_j)\, dx = \int_{B_i} f_i(x_i)\, dx_i$$

So the density of X_i is f_i and

$$\mathbb{P}(X_1 \in B_1, \dots, X_n \in B_n) = \prod_{i=1}^{n} \int_{B_i} f_i(x_i)\, dx_i =$$

$$= \prod_{i=1}^{n} \mathbb{P}(X_i \in B_i) \quad . \quad \square$$

Let $X = (X_1, \dots, X_n)$ have density f.

Then $\mathbb{P}(X_1 \leq x) = \mathbb{P}(X_1 \leq x, X_2 \in \mathbb{R}, \dots, X_n \in \mathbb{R})$

$$= \int_{-\infty}^{x} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, x_2, \dots, x_n)\, dx_1 \dots dx_n$$

$$= \int_{-\infty}^{x} \left(\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, x_2, \dots, x_n)\, dx_2 \cdots dx_n \right) dx_1 .$$

So the density of X_1 is

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \dots, x_n)\, dx_2 \dots dx_n .$$

It is called the marginal density.

Let f and g be 2 densities on $\mathbb{R}$.
Their convolution is defined to be

$$f * g(x) = \int_{-\infty}^{\infty} f(x-y) g(y)\, dy .$$

Let $X \sim f_X$, $Y \sim f_Y$ be 2 indep. r.v.'s. Want the density of $X+Y$.

$$P(X+Y \leq z) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} 1(x+y \leq z)\, f_{X,Y}(x,y)\,dx\,dy$$

$$\overset{\text{indep.}}{=} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} 1(x+y \leq z)\, f_X(x)\, f_Y(y)\,dx\,dy$$

$$= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{z-x} f_X(x) f_Y(y)\,dy \right) dx =$$

$$= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{z} f_Y(y-x) f_X(x)\,dy \right) dx =$$

$$= \int_{-\infty}^{z} \left(\int_{-\infty}^{\infty} f_X(x) f_Y(y-x)\,dx \right) dy$$

So the density ~~is~~ of $X+Y$ is

$$\int_{-\infty}^{\infty} f_X(x) f_Y(y-x)\,dx = \int_{-\infty}^{\infty} f_X(x-y) f_Y(y)\,dy.$$

Non-rigorous

$$\mathbb{P}(X+Y\le z) = \int_{-\infty}^{\infty} \mathbb{P}(X+Y\le z,\ Y\in dy) =$$

$$= \int_{-\infty}^{\infty} \mathbb{P}(X+y\le z,\ Y\in dy) \overset{\text{indep.}}{=}$$

$$= \int_{-\infty}^{\infty} \mathbb{P}(X+y\le z)\cdot \underbrace{\mathbb{P}(Y\in dy)}_{= f_Y(y)dy}$$

$$= \int_{-\infty}^{\infty} F_X(z-y)\cdot f_Y(y)\,dy$$

$$\frac{d}{dz}\mathbb{P}(X+Y\le z) = \int_{-\infty}^{\infty} f_X(z-y) f_Y(y)\,dy.$$

So $\quad f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(z-y) f_Y(y)\,dy.$

Conditional density

Let X and Y be 2 cts r.v.'s with joint density $f_{X,Y}$. Then the conditional

density of X given $Y=y$

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

The conditional expectation of X given Y is $g(Y)$ where

$$g(y) = \int_{-\infty}^{\infty} x\, f_{X|Y}(x|y)\, dx \qquad (= \mathbb{E}[X|Y=y])$$

We denote it by $\mathbb{E}[X|Y] = g(Y)$

Law of total probability

$$f_X(x) = \int_{-\infty}^{\infty} f_{X|Y}(x|y) f_Y(y)\, dy$$

$$\left(f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy \right).$$

Example Let $X \sim Exp(\lambda)$ and $Y \sim Exp(\mu)$ be indep. Set $Z = \min(X, Y)$.

$\mathbb{P}(Z \leq z) = 1 - \mathbb{P}(Z > z) = 1 - \mathbb{P}(X > z, Y > z) =$

$= 1 - \mathbb{P}(X > z)\mathbb{P}(Y > z) = 1 - e^{-\lambda z} \cdot e^{-\mu z} = 1 - e^{-(\lambda+\mu)z}$.

So $Z \sim Exp(\lambda + \mu)$.

More generally, if $X_1, \dots, X_n$ are indep. $X_i \sim Exp(\lambda_i)$. Then

$\min(X_1, \dots, X_n) \sim Exp\left(\sum_{i=1}^{n} \lambda_i\right)$.

Transformation of a multidimensional r.v.

Theorem Let X be a r.v. with values in $D \subseteq \mathbb{R}^d$ and with density f_X. Let g be a bijection from D into $g(D)$ with continuous derivative on D and $\det g'(x) \neq 0 \ \forall x \in D$. Set $y = g(x)$ and $Y = g(X)$, then the density of Y is given by

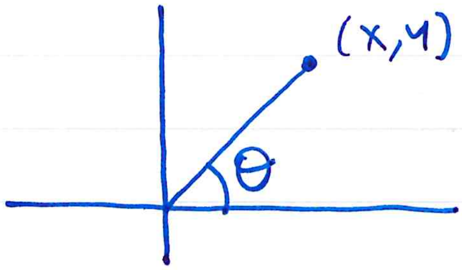
$$f_Y(y) = f_X(x) \cdot |J| \quad , \text{ where}$$

J is the Jacobian

$$J = \det\left(\left(\frac{\partial x_i}{\partial y_j} \right)_{i,j=1}^{d} \right) .$$

No proof.

<u>Example</u> Let X and Y be indep. $\sim \mathcal{N}(0,1)$



$R = |(X,Y)| = \sqrt{X^2+Y^2}$

Want joint density of (R,Θ).

Set $X = R\cos\Theta$ and $Y = R\sin\Theta$.

$x = r\cos\vartheta$ and $y = r\sin\vartheta$

$$f_{(R,\Theta)}(r,\theta) = f_{(X,Y)}(x,y) \cdot |J|$$

$$|J| = \det\begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix} = r$$

So $f_{(R,\Theta)}(r,\theta) \underset{\text{indep.}}{=} f_X(x) \cdot f_Y(y) \cdot r =$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{r^2\cos^2\theta}{2}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{r^2\sin^2\theta}{2}} \cdot r$$

$$= \frac{1}{2\pi} r e^{-\frac{r^2}{2}}.$$

So R and Θ are indep. with

$\Theta \sim U[0,2\pi]$ and $R \sim re^{-\frac{r^2}{2}}$ on $(0,\infty)$.

Order statistics of a random sample

Let $X_1, X_2, \dots, X_n$ be iid with distr. function F and density f.
If we put them in order from smallest to biggest

$$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$$

then $Y_i = X_{(i)}$ are the order statistics.

$\underline{\underline{Y_1}}$: $\mathbb{P}(Y_1 \leq x) = 1 - \mathbb{P}(Y_1 > x) = 1 - (1-F(x))^n$.

density of Y_1 : $\frac{d}{dx}\left(1-(1-F(x))^n\right) = n(1-F(x))^{n-1} \cdot f(x)$

$\underline{\underline{Y_n}}$: $\mathbb{P}(Y_n \leq x) = (F(x))^n$

density of Y_n : $\frac{d}{dx}(F(x))^n = n \cdot (F(x))^{n-1} \cdot f(x)$.

What is the density of $(Y_1, \dots, Y_n)$?

Let $x_1 < x_2 < \dots < x_n$

$\mathbb{P}(Y_1 \leq x_1, Y_2 \leq x_2, \dots, Y_n \leq x_n) =$

$= n!\, \mathbb{P}(X_1 \le x_1, X_2 \le x_2, \dots, X_n \le x_n, X_1 < X_2 < \dots < X_n)$

$$= n! \int_{-\infty}^{x_1} f(u_1) \int_{u_1}^{x_2} f(u_2) \cdots \int_{u_{n-1}}^{x_n} f(u_n)\, du_n \dots du_1 .$$

Differentiating

$$f(y_1, \dots, y_n) = \begin{cases} n!\, f(y_1) \cdots f(y_n) & y_1 < y_2 < \dots < y_n \\ 0 & \text{otherwise} \end{cases}.$$

<u>Example</u> Let $X_1, \dots, X_n$ be iid $\mathrm{Exp}(\lambda)$ and Y_i their order statistics.
Set $Z_1 = Y_1,\ Z_2 = Y_2 - Y_1, \dots, Z_n = Y_n - Y_{n-1}$.

Then $Z = \begin{pmatrix} Z_1 \\ \vdots \\ Z_n \end{pmatrix} = A \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}$

where $A = \begin{pmatrix} 1 & 0 & \dots & & 0 \\ -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & 0 & \dots \\ & & & & \\ & & & -1 & 1 \end{pmatrix}$

and $\det A = 1$ and $y_j = \sum_{i=1}^{j} z_i$

$$f_{(Z_1,\dots,Z_n)}(z_1,\dots,z_n) = f_{(Y_1,\dots,Y_n)}(y_1,\dots,y_n)\cdot \underset{1}{\overset{}{|J|}}$$

$\cancel{= n!\, f_{Y_1}(y_1)\cdots f_{Y_n}(y_n)}$ Let $f(x) = \lambda e^{-\lambda x}$.

$$= n!\, e^{-\lambda y_1} \cdots e^{-\lambda y_n}\cdot \lambda^n = n!\,\lambda^n e^{-\lambda(n z_1 + (n-1) z_2 + \dots + 2 z_{n-1} + z_n)}$$

$$= \prod_{i=1}^{n} (n-i+1)\lambda\, e^{-\lambda(n-i+1) z_i}$$

So $(Z_1,\dots,Z_n)$ are indep. exponentials with $Z_i \sim \text{Exp}(\lambda(n-i+1))$.

<u>Example</u> $U \sim U[0,1]$. Set $Y = -\log U$.

$$\mathbb{P}(Y \le x) = \mathbb{P}(-\log U \le x) = \mathbb{P}(U \ge e^{-x}) = 1 - e^{-x}.$$

So $Y \sim \text{Exp}(1)$.

<u>Theorem</u> Let X be a cts r.v. with ~~continu~~ distr. function F. Then if $U \sim U[0,1]$, then $F^{-1}(U)$ has the same distr. as X.

<u>Proof</u> $\mathbb{P}(F^{-1}(U) \le x) = \mathbb{P}(U \le F(x)) = F(x)$. $\square$

Rejection sampling

Suppose $A \subset [0,1]^d$ and

$$f(x) = \frac{\mathbb{1}(x \in A)}{|A|}, \quad |A| = \text{volume of } A.$$

Let X have density f.

Let $(U_n)_{n\in\mathbb{N}}$ be an iid sequence of d-dim. uniforms, i.e.

$$U_n = (U_{k,n} : k \in \{1,\dots,d\})$$

and $(U_{k,n})$ are iid with $U_{1,n} \sim U[0,1]$.

$N = \min\{n \geq 1 : U_n \in A\}$.

Set $X = U_N$. Let $B \subset [0,1]^d$.

NTS $\mathbb{P}(X \in B) = \dfrac{|B \cap A|}{|A|}$.

Proof $\mathbb{P}(X \in B) = \sum_{n=1}^{\infty} \mathbb{P}(X \in B, N = n) =$

$$= \sum_{n=1}^{\infty} \mathbb{P}(U_n \in A\cap B, U_{n-1} \notin A, \dots, U_1 \notin A) \overset{\text{indep.}}{=\!=}$$

$$= \sum_{n=1}^{\infty} \mathbb{P}(U_n \in A\cap B) \cdot \big(\mathbb{P}(U_1 \notin A)\big)^{n-1} =$$

$$= \sum_{n=1}^{\infty} |A\cap B| \cdot (1-|A|)^{n-1} = \frac{|A\cap B|}{|A|} = \int_B f(x)\,dx \,.$$

Let f be a bounded density on $[0,1]^{d-1}$,
i.e. $\exists\, \lambda$ s.t. $f(x) \le \lambda \quad \forall\, x \in [0,1]^{d-1}$.
$X \sim f$.
Consider $A = \left\{ (x_1, \dots, x_d) \in [0,1]^d : x_d \le \dfrac{f(x_1,\dots,x_{d-1})}{\lambda} \right\}$

Let $Y = (X_1, \dots, X_d)$ be uniform on A generated as before.
Set $X = (X_1, \dots, X_{d-1})$. NTS $X \sim f$.
Let $B \subseteq [0,1]^{d-1}$.

$$\mathbb{P}\big((X_1,\dots,X_{d-1}) \in B\big) = \mathbb{P}\big((X_1,\dots,X_d) \in (B\times[0,1]) \cap A\big)$$

$$= \frac{|(B\times[0,1]) \cap A|}{|A|}$$

$$|(B\times[0,1])\cap A| = \int\cdots\int 1\big((x_1,\dots,x_d)\in B\times[0,1]\cap A\big)\,dx_1\cdots dx_d$$

$$= \int\cdots\int 1\big((x_1,\dots,x_{d-1})\in B\big)\cdot 1\Big(x_d \le \frac{f(x_1,\dots,\cancel{x}_{d-1})}{\lambda}\Big)\,dx_1\cdots dx_d$$

$$= \int\cdots\int 1\big((x_1,\dots,x_{d-1})\in B\big)\,\frac{f(x_1,\dots,x_{d-1})}{\lambda}\,dx_1\cdots dx_{d-1}$$

$$= \int_B \frac{f(x)}{\lambda}\,dx$$

$$|A| = \int \frac{f(x)}{\lambda}\,dx = \frac{1}{\lambda}$$

So $\mathbb{P}\big((X_1,\dots,X_{d-1})\in B\big) = \int_B f(x)\,dx$. $\square$

Moment generating functions

Let X have density f. The moment generating function (mgf)

$$m(\theta) = \mathbb{E}\left[e^{\theta X}\right] = \int_{-\infty}^{\infty} e^{\theta x} f(x)\,dx$$

$m(0) = 1$. We define $m(\theta)$ only if it is finite.

<u>Theorem</u> The mgf uniquely determines the distribution of a r.v. provided it is defined for an open interval of values of θ.

<u>Theorem</u> Suppose the mgf is defined for some open interval of values of θ, then

$$m^{(r)}(0) = \frac{d^r}{d\theta^r} m(\theta)\Big|_{\theta=0} = \mathbb{E}[X^r].$$

__Examples__ 1) Gamma distribution

Let $f(x) = e^{-\lambda x} \lambda^n \cdot \dfrac{x^{n-1}}{(n-1)!}$, $\lambda \geq 0$, $n \in \mathbb{N}$.
$x \geq 0$

__$n=1$__ $\rightsquigarrow$ $\mathrm{Exp}(\lambda)$.

Why is f a density?

$$I_n = \int_0^\infty f(x)\,dx = \int_0^\infty \lambda e^{-\lambda x} \cdot \lambda^{n-1} \frac{x^{n-1}}{(n-1)!}\,dx$$

$$= \int_0^\infty (n-1)\, e^{-\lambda x} \cdot \lambda^{n-1} \cdot \frac{x^{n-2}}{(n-1)!}\,dx = I_{n-1} = \dots = I_1 = 1$$

$$m(\theta) = \mathbb{E}[e^{\theta X}] = \int_0^\infty e^{\theta x} e^{-\lambda x} \cdot \lambda^n \frac{x^{n-1}}{(n-1)!}\,dx$$

__$\theta < \lambda$__

$$= \int_0^\infty e^{-(\lambda-\theta)x} \frac{\lambda^n}{(\lambda-\theta)^n} \cdot (\lambda-\theta)^n \cdot \frac{x^{n-1}}{(n-1)!}\,dx$$

$$= \left(\frac{\lambda}{\lambda-\theta}\right)^n \underbrace{\int_0^\infty e^{-(\lambda-\theta)x} (\lambda-\theta)^n \cdot \frac{x^{n-1}}{(n-1)!}\,dx}_{\Gamma(\lambda-\theta,\, n) \quad = 1}$$

$$= \left(\frac{\lambda}{\lambda-\theta}\right)^n.$$

For $n=1 \rightsquigarrow$ mgf of $\mathrm{Exp}(\lambda)$.

Suppose $X_1, \dots, X_n$ are indep. r.v.'s.

Then $m(\theta) = \mathbb{E}[e^{\theta(X_1+\dots+X_n)}] = \prod_{i=1}^{n} \mathbb{E}[e^{\theta X_i}]$

Let $X \sim \Gamma(n,\lambda)$ and $Y \sim \Gamma(m,\lambda)$ and

$\downarrow$

Gamma with λ, n

X and Y are indep. $\theta < \lambda$

$$\mathbb{E}[e^{\theta(X+Y)}] = \mathbb{E}[e^{\theta X}] \cdot \mathbb{E}[e^{\theta Y}] = \left(\frac{\lambda}{\lambda-\theta}\right)^n \left(\frac{\lambda}{\lambda-\theta}\right)^m =$$

$$= \left(\frac{\lambda}{\lambda-\theta}\right)^{n+m}.$$

So $X+Y \sim \Gamma(n+m, \lambda)$ since mgf uniquely characterises the distr.

In particular, if $X_1, \dots, X_n$ are indep. $\mathrm{Exp}(\lambda)$ then

$$X_1 + \dots + X_n \sim \Gamma(n,\lambda).$$

One can also consider $\Gamma(\alpha, \lambda)$ $(\alpha > 0)$ by replacing $(n-1)!$ by

$$\Gamma(\alpha) = \int_0^\infty e^{-x} \cdot x^{\alpha-1} dx.$$

2) <u>Normal distribution</u>

Let $X \sim \mathcal{N}(\mu, \sigma^2)$

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

$$\theta x - \frac{(x-\mu)^2}{2\sigma^2} = \theta\mu + \frac{\theta^2\sigma^2}{2} - \frac{(x-\mu-\theta\sigma^2)^2}{2\sigma^2}$$

$$\mathbb{E}[e^{\theta X}] = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \underbrace{e^{\theta\mu + \frac{\theta^2\sigma^2}{2}}} \exp\left(-\frac{(x-\mu-\theta\sigma^2)^2}{2\sigma^2}\right) dx$$

$\mathcal{N}(\mu+\theta\sigma^2, \sigma^2)$

$$= e^{\theta\mu + \frac{\theta^2\sigma^2}{2}}$$

If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $aX+b \sim \mathcal{N}(a\mu+b, a^2\sigma^2)$

$$\mathbb{E}[e^{\theta(aX+b)}] = e^{\theta b} \mathbb{E}[e^{a\theta X}] = e^{\theta b} \cdot e^{a\theta\mu + \frac{a^2\theta^2\sigma^2}{2}}$$

$$= e^{\theta(b+a\mu)+\theta^2\frac{a^2\sigma^2}{2}} .$$

Let $X \sim \mathcal{N}(\mu,\sigma^2)$ and $Y \sim \mathcal{N}(\nu,\tau^2)$, $X \underset{\text{indep.}}{\perp\!\!\!\perp} Y$.

$$\mathbb{E}\left[e^{\theta(X+Y)}\right] = e^{\theta\mu+\frac{\theta^2\sigma^2}{2}} \cdot e^{\theta\nu+\frac{\theta^2\tau^2}{2}}$$

$$= e^{\theta(\mu+\nu)+\frac{\theta^2}{2}(\sigma^2+\tau^2)}$$

So $X+Y \sim \mathcal{N}(\mu+\nu, \sigma^2+\tau^2)$.

3) Cauchy distribution

$$f(x) = \frac{1}{\pi(1+x^2)} \qquad x \in \mathbb{R}.$$

$$m(0) = 1$$

$$m(\theta) = \mathbb{E}\left[e^{\theta X}\right] = \int_{-\infty}^{\infty} \frac{e^{\theta x}}{\pi(1+x^2)}\,dx = \infty \qquad \forall\, \theta \neq 0.$$

Let $X \sim f$. Then $X, 2X, \dots$ have the same mgf but not the same distr.

So assumption on $m(\theta)$ being finite for an open interval of values of θ is <u>essential</u>.

Multivariate moment generating function

Let $X = (X_1, \dots, X_n)$ a random variable in $\mathbb{R}^n$. The mgf of X is defined to be

$$m(\theta) = \mathbb{E}\left[e^{\theta^T X}\right] = \mathbb{E}\left[e^{\theta_1 X_1 + \dots + \theta_n X_n}\right],$$

where $\theta = (\theta_1, \dots, \theta_n)^T$.

- Provided $m(\theta)$ is finite for a range of values of θ, it uniquely characterises the distribution of X.

- $$\left.\frac{\partial^r m}{\partial \theta_i^r}\right|_{\theta=0} = \mathbb{E}\left[X_i^r\right]$$

  $$\left.\frac{\partial^{r+s} m}{\partial \theta_i^r \, \partial \theta_j^s}\right|_{\theta=0} = \mathbb{E}\left[X_i^r X_j^s\right].$$

- $m(\theta) = \prod_{i=1}^{n} \mathbb{E}\left[e^{\theta_i X_i}\right]$ iff $X_1, \dots, X_n$ are indep.

Definition Let $(X_n, n \in \mathbb{N})$ be a sequence of random variables and let X be another r.v. We say X_n converges to X in distribution, $X_n \xrightarrow{d} X$ if

$$F_{X_n}(x) \longrightarrow F_X(x),$$

$\forall\, x \in \mathbb{R}$ that are continuity points of F_X.

Theorem (Continuity thm for mgf's)

Let X be a r.v. with $m(\theta) < \infty$ for some $\theta \neq 0$. Suppose that writing

$m_n(\theta) = \mathbb{E}[e^{\theta X_n}]$ we have

$$m_n(\theta) \to m(\theta) \quad \forall\ \theta \in \mathbb{R}.$$

Then X_n converges to X in distribution.

Limit theorems for sums of iid r.v's

Thm Weak law of large numbers

Let $(X_n : n \in \mathbb{N})$ be an iid sequence of r.v.'s with finite expectation $\mu = \mathbb{E}[X_1]$. Set $S_n = X_1 + \ldots + X_n$.
Then $\forall\, \varepsilon > 0$ we have

$$\cancel{\mathbb{P}\left(\left|\frac{S_n}{n}\right| > \varepsilon\right)}$$

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) \xrightarrow[n\to\infty]{} 0 .$$

Proof Assume further that $\mathrm{Var}(X_1) = \sigma^2 < \infty$.

$$\mathbb{E}\left[\frac{S_n}{n}\right] = \mu \quad \text{and} \quad \mathrm{Var}\left(\frac{S_n}{n}\right) = \frac{\sigma^2}{n} .$$

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) \underbrace{\leq}_{\text{Chebyshev's ineq.}} \frac{1}{\varepsilon^2} \cdot \mathrm{Var}\left(\frac{S_n}{n}\right) = \frac{\sigma^2}{\varepsilon^2 n} \xrightarrow[n\to\infty]{} 0$$

$\square$

Def. A sequence (X_n) converges to X in probability and we write

$$X_n \xrightarrow{\mathbb{P}} X \quad \text{as } n \to \infty$$

if $\forall \varepsilon > 0$ $$\mathbb{P}(|X_n - X| > \varepsilon) \xrightarrow[n\to\infty]{} 0 .$$

Def. (X_n) converges to X with probability 1 / almost surely if

$$\mathbb{P}\left(\lim_{n\to\infty} X_n = X\right) = 1$$

Thm Strong law of large numbers

Let $(X_n : n \in \mathbb{N})$ be an iid sequence of r.v.'s with finite expectation $\mu = \mathbb{E}[X_1]$. Set $S_n = X_1 + \dots + X_n$.

Then $$\mathbb{P}\left(\frac{S_n}{n} \to \mu \text{ as } n \to \infty\right) = 1 .$$

Proof non-examinable

Assume further that $\mathbb{E}[X_1^4] < \infty$.

By considering $Y_i = X_i - \mu$ we can reduce

to the case of 0 expectation.

We have $\mathbb{E}[Y_1^4] < \infty$ $(\mathbb{E}[X_1^4] < \infty)$.

Let $S_n = \sum_{i=1}^n X_i$ with $\mu = 0$ and $\mathbb{E}[X_1^4] < \infty$.

$$S_n^4 = \left(\sum_{i=1}^n X_i\right)^4 = \sum_{i=1}^n X_i^4 + \binom{4}{2} \sum_{1 \le i < j \le n} X_i^2 X_j^2 + R$$

where R is a sum of terms of

the form $X_i^2 X_j X_k$, $X_i^3 X_j$, $X_i X_j X_k X_\ell$

with i, j, k, ℓ distinct.

Since the X_i's are indep. and of 0 mean

$$\Rightarrow \quad \mathbb{E}[X_i^2 X_j X_k] = 0 = \mathbb{E}[X_i^3 X_j] = \mathbb{E}[X_i X_j X_k X_l]$$

So $\quad \mathbb{E}[R] = 0.$

$$\mathbb{E}[S_n^4] = \sum_{i=1}^{n} \mathbb{E}[X_i^4] + 3n(n-1)\, \overset{\text{indep.}}{\mathbb{E}[X_1^2]\,\mathbb{E}[X_2^2]}$$

$$= n\,\mathbb{E}[X_1^4] + 3n(n-1)\underbrace{\left(\mathbb{E}[X_1^2]\right)^2}_{\leq\ \mathbb{E}[X_1^4]}$$

$$\leq \left(n + 3n(n-1)\right) \mathbb{E}[X_1^4] \leq 3n^2\, \mathbb{E}[X_1^4].$$

So $\quad \mathbb{E}\left[\sum_{n=1}^{\infty} \left(\frac{S_n}{n}\right)^4\right] = \sum_{n=1}^{\infty} \mathbb{E}\left[\left(\frac{S_n}{n}\right)^4\right] \leq$

$$\leq 3\,\mathbb{E}[X_1^4] \cdot \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty.$$

So this means that

$$\cancel{\mathbb{E}}\ \mathbb{P}\left(\sum_{n=1}^{\infty} \left(\frac{S_n}{n}\right)^4 < \infty \right) = 1$$

$$\Rightarrow \quad \mathbb{P}\left(\frac{S_n}{n} \to 0 \text{ as } n \to \infty \right) = 1. \qquad \square$$

SLLN $\Rightarrow$ WLLN.

Suppose $X_n \to 0$ almost surely

$\left(\mathbb{P}(\lim X_n = 0) = 1 \right)$

then $X_n \xrightarrow{\mathbb{P}} 0$.

<u>NTS</u> $\forall \varepsilon > 0$ $\mathbb{P}(|X_n| > \varepsilon) \to 0$ as $n \to \infty$.

equivalently $\mathbb{P}(|X_n| \leq \varepsilon) \to 1$ as $n \to \infty$.

$$\mathbb{P}(|X_n| \leq \varepsilon) \geq \mathbb{P}\Big(\underbrace{\bigcap_{m=n}^{\infty} \{|X_m| \leq \varepsilon\}}_{= A_n} \Big)$$

$A_n \subseteq A_{n+1}$ and $\bigcup A_n = \left\{ |X_m| \leq \varepsilon \text{ for all } m \text{ suf. large} \right\}$.

$\Rightarrow$ So $\lim_{n \to \infty} \mathbb{P}(|X_n| \leq \varepsilon) \geq \mathbb{P}(\bigcup A_n) \geq \mathbb{P}(X_n \to 0)$
$= 1$ $\square$

We saw $\frac{S_n}{n} - \mu \to 0$ as $n \to \infty$

$\mu = \mathbb{E}[X_1]$

$S_n = X_1 + \dots + X_n$

$$\operatorname{Var}\left(\frac{S_n}{n} - \mu\right) = \frac{\sigma^2}{n}$$

$$\frac{\frac{S_n}{n} - \mu}{\sqrt{\operatorname{Var}\left(\frac{S_n}{n} - \mu\right)}} = \frac{\frac{S_n}{n} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

Theorem Central limit theorem

Let $(X_n : n \in \mathbb{N})$ be an iid sequence of r.v.'s with finite epectation and variance $\mu = \mathbb{E}[X_1]$ and $\sigma^2 = \operatorname{Var}(X_1)$. Set $S_n = X_1 + \dots + X_n$. Then

$$\mathbb{P}\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x\right) \xrightarrow[n\to\infty]{} \Phi(x) = \int_{-\infty}^{x} \frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}}\, dy .$$

$\forall\, x \in \mathbb{R}$.

In other words,

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} Z \quad \text{as } n \to \infty,$$

where $Z \sim N(0, 1)$.

This says for n large enough

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \approx Z \quad , Z \sim N(0,1)$$

$$\Rightarrow S_n \approx n\mu + \sigma\sqrt{n}\, Z$$

So for n large $S_n \approx N(n\mu, \sigma^2 n)$.

<u>Proof</u> By considering $Y_i = \frac{X_i - \mu}{\sigma}$

it suffices to prove the theorem under 0 mean and variance 1 assumption.

$S_n = X_1 + \dots + X_n$ $\qquad \mathbb{E}[X_i] = 0, \ \mathrm{Var}(X_i) = 1.$

Assume further that $\exists\, \delta > 0$ s.t.

$$\mathbb{E}[e^{\delta X_1}] < \infty \quad \text{and} \quad \mathbb{E}[e^{-\delta X_1}] < \infty.$$

<u>NTS</u> $\dfrac{S_n}{\sqrt{n}} \xrightarrow{d} N(0,1)$ as $n \to \infty$.

By the continuity property of mgf's we need to show $\forall\, \theta \in \mathbb{R}$.

$$\mathbb{E}\left[e^{\theta \frac{S_n}{\sqrt{n}}}\right] \xrightarrow[n\to\infty]{} \mathbb{E}[e^{\theta Z}] = e^{\theta^2/2}, \quad Z \sim N(0,1)$$

$$\mathbb{E}\left[e^{\frac{\theta S_n}{\sqrt{n}}}\right] = \mathbb{E}\left[e^{\frac{\theta}{\sqrt{n}}(X_1 + \dots + X_n)}\right] = \left(\mathbb{E}\left[e^{\frac{\theta}{\sqrt{n}} X_1}\right]\right)^n$$

Set $m(\theta) = \mathbb{E}[e^{\theta X_1}]$

NTS $\left(m\left(\frac{\theta}{\sqrt{n}}\right)\right)^n \longrightarrow e^{\theta^2/2}$ as $n \to \infty$.

$$m(\theta) = \mathbb{E}[e^{\theta X_1}] = \mathbb{E}\left[1 + \theta X_1 + \frac{\theta^2}{2}X_1^2 + \sum_{k=3}^{\infty}\frac{\theta^k X_1^k}{k!}\right]$$

$|\theta| \leq \frac{\delta}{2}$

So $$m(\theta) = 1 + \frac{\theta^2}{2} + \mathbb{E}\left[\sum_{k=3}^{\infty}\frac{\theta^k X_1^k}{k!}\right]$$

We will prove

$$\left|\mathbb{E}\left[\sum_{k=3}^{\infty}\frac{\theta^k X_1^k}{k!}\right]\right| = o(|\theta|^2) \quad \text{as } \theta \to 0. \quad (\star)$$

Then we will get

$$m\left(\frac{\theta}{\sqrt{n}}\right) = 1 + \frac{\theta^2}{2n} + o\left(\frac{\theta^2}{n}\right) = 1 + \frac{\theta^2}{2n}\left(1 + o(1)\right).$$

and $m\left(\frac{\theta}{\sqrt{n}}\right)^n \longrightarrow e^{\frac{\theta^2}{2}}$ as $n \to \infty$.

We prove $(\star)$.

$$\left|\mathbb{E}\left[\sum_{k=3}^{\infty}\frac{\theta^k X_1^k}{k!}\right]\right| \leq \mathbb{E}\left[\sum_{k=3}^{\infty}\frac{|\theta X_1|^k}{k!}\right]$$

$$\sum_{k=3}^{\infty} \frac{|\theta X_1|^k}{k!} = |\theta X_1|^3 \sum_{k=0}^{\infty} \frac{|\theta X_1|^k}{(k+3)!} \leq$$

$$\leq |\theta X_1|^3 \cdot \sum_{k=0}^{\infty} \frac{|\theta X_1|^k}{k!}$$

$$\leq |\theta X_1|^3 \, e^{\frac{\delta}{2}|X_1|} \qquad \text{because } |\theta| \leq \frac{\delta}{2} .$$

$$\leq |\theta|^3 \cdot \underbrace{\frac{\left|\frac{\delta}{2} X_1\right|^3}{3!} e^{\frac{\delta}{2}|X_1|}}_{\leq e^{\frac{\delta}{2}|X_1|}} \cdot \frac{3!}{\left(\frac{\delta}{2}\right)^3}$$

$$\cancel{\leq 3! \left(\frac{2\theta}{\delta}\right)^3 e^{\delta|X_1|}} \; . \quad \leq 3! \left(\frac{2\theta}{\delta}\right)^3 e^{\delta|X_1|} .$$

Taking expectation we get

$$\left| \mathbb{E}\left[\sum_{k=3}^{\infty} \frac{|\theta X_1|^k}{k!} \right] \right| \leq 3! \left(\frac{2|\theta|}{\delta}\right)^3 \mathbb{E}\left[e^{\delta|X_1|}\right]$$

$$\leq 3! \left(\frac{2|\theta|}{\delta}\right)^3 \left(\mathbb{E}\left[e^{\delta X_1}\right] + \mathbb{E}\left[e^{-\delta X_1}\right] \right)$$

So $\left| \mathbb{E}\left[\sum_{k=3}^{\infty} \cdots \right] \right| = o(|\theta|^2)$. $\square$

<u>Applications</u>

1) Normal approximation to the <u>binomial</u> distribution.

Let $S_n \sim \mathrm{Bin}(n,p)$.

$S_n = \sum_{i=1}^{n} X_i$ with (X_i) iid $\sim \mathrm{Ber}(p)$.

$\mathbb{E}[S_n] = np$ $\quad \mathrm{Var}(S_n) = np(1-p)$.

So $\dfrac{S_n - np}{\sqrt{np(1-p)}} \xrightarrow{d} \mathcal{N}(0,1)$ as $n \to \infty$

$S_n \approx \mathcal{N}(np, np(1-p))$ for n large.

In the Poisson approximation to the binomial, p scales as $\frac{\lambda}{n}$, $\lambda > 0$.

$\text{Bin}\left(n, \frac{\lambda}{n}\right) \longrightarrow \text{Poi}(\lambda)$.

In the normal approx. p is kept fixed i.e. not depending on n.

2) Normal approx. to the Poisson distr.

Let $S_n \sim \text{Poi}(n)$. Then S_n can be realised as the sum of n iid Poi(1).

So $S_n = \sum_{i=1}^{n} X_i \qquad X_i \sim \text{Poi}(1)$

$$\frac{S_n - n}{\sqrt{n}} \longrightarrow N(0,1) \quad \text{as } n \to \infty$$

Sampling error via the CLT

A proportion p of a population votes Yes and $1-p$ No in a referendum. Want to estimate p with an error of at most $\pm 4\%$ w. prob. at least 0.99.

We pick N individuals at random. Let S_N be the number who vote Yes.
We are going to approximate p by

$$\hat{p}_N = \frac{S_N}{N}.$$

We want $\mathbb{P}\left(|\hat{p}_N - p| \leq \frac{4}{100} \right) \geq 0.99.$

$S_N \sim \mathrm{Bin}(N, p)$

$$\frac{S_N - Np}{\sqrt{Np(1-p)}} \approx \mathcal{N}(0,1).$$

$$S_N \approx Np + \sqrt{Np(1-p)}\, Z \quad , \; Z \sim \mathcal{N}(0,1)$$

$$\frac{S_N}{N} \approx p + \sqrt{\frac{p(1-p)}{N}}\, Z \, .$$

Take N large enough

$$\hat{p}_N = \frac{S_N}{N} = p + \sqrt{\frac{p(1-p)}{N}}\, Z$$

$$\mathbb{P}\left(|\hat{p}_N - p| \leq \frac{4}{100} \right) = 0.99$$

$$\mathbb{P}\left(\sqrt{\frac{p(1-p)}{N}}\, |Z| \leq \frac{4}{100} \right) = 0.99$$

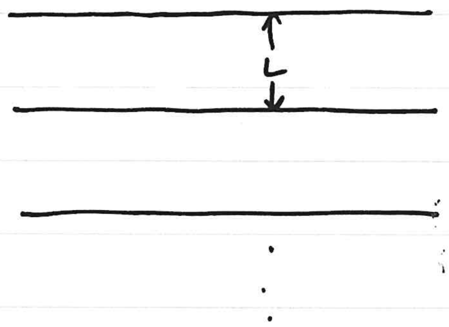
$z \in \mathbb{R} \qquad \mathbb{P}(|Z| \geq z) = 2(1 - \Phi(z))$

$z = 2.58$, then $\mathbb{P}(|Z| \geq 2.58) = 0.01$.

$$\left.\begin{array}{l} \text{Want } \dfrac{4}{100} \sqrt{\dfrac{N}{p(1-p)}} \geq 2.58 \\ \text{Worst variance when } p = \frac{1}{2} \end{array}\right\} \Rightarrow N \geq 1040 \, .$$

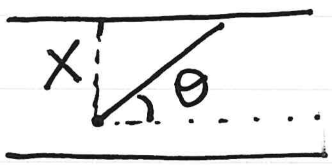
Buffon's needle

Parallel lines on the plane at distance L apart.



$\ell \leq L$

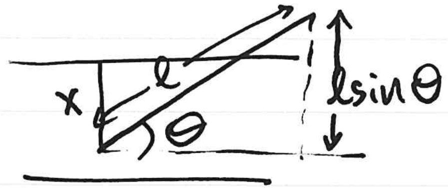
Throw the needle at random.
What is the probability it intersects at least 1 line?



X = distance of left end to line above

Θ = angle with horizontal line.

Let's consider the model, where $X \sim U[0, L]$ and $\Theta \sim U[0, \pi]$ and X and Θ are independent.



The needle will intersect a line iff

$X \leq l\sin\Theta$.

So $\mathbb{P}$(needle intersects a line) =

$$= \mathbb{P}(X \leq l\sin\Theta) =$$

$$= \int_0^L \int_0^\pi 1(x \leq l\sin\theta)\, f_{X,\Theta}(x,\theta)\, dx\, d\theta$$

$$\overset{\text{indep.}}{=} \int_0^L \int_0^\pi 1(x \leq l\sin\theta)\, \frac{1}{\pi L}\, dx\, d\theta$$

$$= \frac{1}{\pi L} \int_0^\pi d\theta \underbrace{\int_0^L 1(x \leq l\sin\theta)\, dx}_{l\sin\theta} = \frac{1}{\pi L}\int_0^\pi l\sin\theta\, d\theta$$

$$= \frac{2l}{\pi L}.$$

$p = \mathbb{P}(\text{intersection})$

We found $p = \dfrac{2l}{\pi L}$.

$\Leftrightarrow \quad \pi = \dfrac{2l}{pL}$.

– Want to approximate π.

Throw n needles independently and

let $\hat{p}_n$ be the proportion of them intersecting a line.

This approximates p and we approximate π by $\hat{\pi}_n = \dfrac{2l}{\hat{p}_n L}$.

Suppose we want

$$\mathbb{P}\left(|\hat{\pi}_n - \pi| \leq 0.001\right) \geq 0.99.$$

How large should n be?

Define $f(x) = \dfrac{2l}{xL}$.

Then $f(p) = \pi$ and $f'(p) = -\dfrac{2l}{p^2 L} = -\dfrac{\pi}{p}$.

Then $\hat{\pi}_n = f(\hat{p}_n)$.

Let S_n = # of needles intersecting a line.

Then $S_n \sim \mathrm{Bin}(n, p)$.

So $$\frac{S_n - np}{\sqrt{np(1-p)}} \xrightarrow{d} N(0,1) \quad \text{as } n \to \infty \quad \underline{\text{CLT}}$$

$$\Rightarrow S_n \approx np + \sqrt{np(1-p)}\, Z, \text{ where } Z \sim N(0,1)$$

$$\Rightarrow \hat{p}_n \approx p + \sqrt{\frac{p(1-p)}{n}}\, Z.$$

By Taylor's thm

$$\hat{\pi}_n = f(\hat{p}_n) \approx f(p) + (\hat{p}_n - p) f'(p)$$

Substitute

So $$\hat{\pi}_n \approx \pi - (\hat{p}_n - p)\frac{\pi}{p}$$

$\Rightarrow \quad \hat{\pi}_n - \pi \approx - \pi \sqrt{\frac{1-p}{pn}} \, Z$

$\mathbb{P}\left(|\hat{\pi}_n - \pi| \leq 0.001 \right) = \mathbb{P}\left(\pi \sqrt{\frac{1-p}{pn}} \, |Z| \leq 0.001 \right)$

$\mathbb{P}(|Z| \geq 2.58) = 0.01$

~~If~~ ~~we~~ ~~take~~ The variance of $\pi \sqrt{\frac{1-p}{pn}} Z$

is $\pi^2 \cdot \frac{1-p}{pn}$; decreasing in p

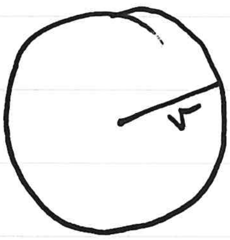
Minimize variance by taking $l = L$

$\Rightarrow \quad p = \frac{2}{\pi}$. and $\mathrm{Var} = \frac{\pi^2}{n}\left(\frac{\pi}{2} - 1 \right)$

Taking $\sqrt{\frac{\pi^2}{n}\left(\frac{\pi}{2} - 1 \right)} \; 2.58 = 0.001$

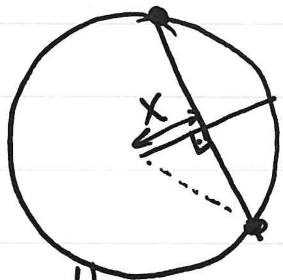
$\Rightarrow \quad n \approx 3.75 \times 10^7$.

Bertrand's paradox



Draw a chord at random. What is the probability it has length $\leq r$?

1st approach Let $X \sim U[0,r]$



Draw the chord perpendicular to the radius at the point X.

Let C be the length.

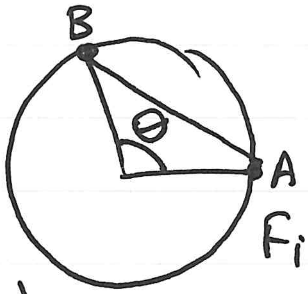
Then $C = 2\sqrt{r^2 - X^2}$

$$\mathbb{P}(C \leq r) = \mathbb{P}(4(r^2 - X^2) \leq r^2) =$$

$$= \mathbb{P}(4X^2 \geq 3r^2) = \mathbb{P}(X \geq \sqrt{3}r/2) =$$

$$= \frac{r - \frac{\sqrt{3}r}{2}}{r} = 1 - \frac{\sqrt{3}}{2} \approx 0.134.$$

2^{nd} approach

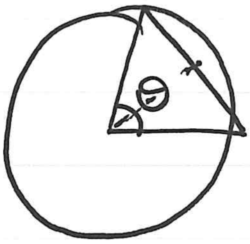


Let $\theta \sim U[0, 2\pi]$

Fix A one endpoint and $C = |AB|$.

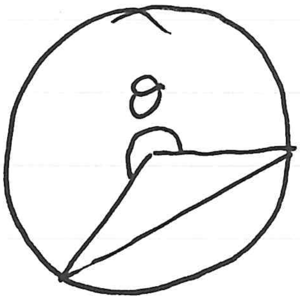
$\mathbb{P}(C \leq r)$

$\theta \in [0, \pi]$



$\rightarrow C = 2r \sin \frac{\theta}{2}$

$\theta \in [\pi, 2\pi]$



$\rightarrow C = 2r \sin\left(\frac{2\pi - \theta}{2}\right)$

$\Rightarrow C = 2r \sin\left(\pi - \frac{\theta}{2}\right)$

$\Rightarrow C = 2r \sin \frac{\theta}{2}$

$$\mathbb{P}(C \leq r) = \mathbb{P}(C \leq r, \theta \leq \pi) +$$

$$+ \mathbb{P}(C \leq r, \theta \in [\pi, 2\pi]) =$$

$$= \mathbb{P}\left(2\cancel{r}\sin\frac{\theta}{2} \leq \cancel{r},\ \theta \leq \pi\right) + \mathbb{P}\left(2r\sin\frac{\theta}{2} \leq r,\ \theta \in [\pi, 2\pi]\right)$$

$$= \mathbb{P}\left(\sin\frac{\theta}{2} \leq \frac{1}{2},\ \theta \leq \pi\right) + \mathbb{P}\left(\sin\frac{\theta}{2} \leq \frac{1}{2},\ \pi \leq \theta \leq 2\pi\right)$$

$$= \mathbb{P}\left(\frac{\theta}{2} \leq \frac{\pi}{6}\right) + \mathbb{P}\left(\frac{\theta}{2} \in \left[\frac{5\pi}{6}, \pi\right]\right)$$

$$= \frac{1}{3} \approx 0.333$$

Multidimensional Gaussian r.v.'s.

A r.v. X in $\mathbb{R}$ is called Gaussian/Normal if

$$X = \mu + \sigma Z \, , \, \mu \in \mathbb{R}, \, \sigma \in (0, \infty) \text{ and}$$

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

The density function of X is

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}$$

We denoted $X \sim \mathcal{N}(\mu, \sigma^2)$

A r.v. $X = (X_1, \dots, X_n)^T$ in $\mathbb{R}^n$ has the Gaussian distribution / is called Gaussian if $\forall \; u = (u_1, \dots, u_n)^T \in \mathbb{R}^n$

$u^T X = \sum_{i=1}^{n} u_i X_i$ is a Gaussian variable in $\mathbb{R}$.

We call X a Gaussian vector.

Suppose A is $m \times n$ matrix and $b \in \mathbb{R}^m$

Then $AX + b$ is ~~a~~ Gaussian in $\mathbb{R}^m$.

<u>Why?</u> Let $u \in \mathbb{R}^m$. Then

$$u^T(AX+b) = (u^T A)X + u^T b.$$

Set $v = A^T u$

$$u^T(AX+b) = \underbrace{v^T X}_{\text{Gaussian}} + u^T b \quad \text{is Gaussian.}$$

Set $\mu = \mathbb{E}[X] = \begin{pmatrix} \mathbb{E}[X_1] \\ \vdots \\ \mathbb{E}[X_n] \end{pmatrix}$

and $V = \mathrm{var}(X) = \mathbb{E}\left[(X-\mu)(X-\mu)^T\right]$

$$= \begin{pmatrix} & & \\ & \mathrm{Cov}(X_i, X_j) & \\ & & \end{pmatrix}$$

Let $X = (X_1, \dots, X_n)$ be a Gaussian vector, i.e.

$\forall u \in \mathbb{R}^n \quad u^T X = \sum_{i=1}^n u_i X_i$ is Gaussian in $\mathbb{R}$

$$\mu = \mathbb{E}[X] = \begin{pmatrix} \mathbb{E}[X_1] \\ \vdots \\ \mathbb{E}[X_n] \end{pmatrix} \qquad \mathbb{E}[X_i] = \mu_i$$

$$V = \operatorname{Var}(X) = \mathbb{E}\left[(X-\mu)(X-\mu)^T \right] = \left(\operatorname{Cov}(X_i, X_j) \right)$$

$$\mathbb{E}[u^T X] = \sum_{i=1}^n u_i \mathbb{E}[X_i] = u^T \mu$$

$$\operatorname{Var}(u^T X) = \operatorname{Var}\left(\sum_{i=1}^n u_i X_i \right) = \sum_{i,j=1}^n u_i \operatorname{Cov}(X_i, X_j) u_j$$

$$\Rightarrow \operatorname{Var}(u^T X) = u^T V u$$

So $\quad u^T X \sim \mathcal{N}(u^T \mu, u^T V u)$

V is a symmetric matrix and since

$$\operatorname{Var}(u^T X) \geq 0 \Rightarrow u^T V u \geq 0$$

which means that V is non-negative definite.

mgf of X $\lambda = (\lambda_1, \dots, \lambda_n)^T$

$$m(\lambda) = \mathbb{E}[e^{\lambda^T X}] = e^{\lambda^T \mu + \frac{\lambda^T V \lambda}{2}}$$

By uniqueness of mgf's we see that the distribution of a Gaussian vector is characterised uniquely by its mean μ and Covariance matrix V.

We write $X \sim \mathcal{N}(\mu, V)$

We can write V

$$V = U^T D U \quad \text{where} \quad U^{-1} = U^T$$

and D is a diagonal matrix

$$D = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} \quad \text{with} \quad \lambda_i \geq 0.$$

Define the square root matrix of V to be

$$\sigma = U^T \sqrt{D} U, \text{ where } \sqrt{D} = \begin{pmatrix} \sqrt{\lambda_1} & & 0 \\ & \ddots & \\ 0 & & \sqrt{\lambda_n} \end{pmatrix}.$$

Then $\quad \sigma \cdot \sigma = U^T \sqrt{D} U \cdot U^T \sqrt{D} U = U^T D U = V.$

Let $Z_1, \dots, Z_n$ be iid $\mathcal{N}(0,1)$ r.v.'s.
Set $Z = (Z_1, \dots, Z_n)^T$.

<u>Z is Gaussian</u> Let $u \in \mathbb{R}^n$

Then $u^T Z = \sum u_i Z_i$. NTS $u^T Z \sim$ normal.

$$\mathbb{E}\left[e^{\lambda u^T Z}\right] = \mathbb{E}\left[e^{\lambda \sum u_i Z_i}\right] = \mathbb{E}\left[e^{\sum \lambda u_i Z_i}\right]$$

$\lambda \in \mathbb{R}$

$$\overset{\text{indep.}}{=} \prod_{i=1}^{n} e^{\frac{\lambda^2 u_i^2}{2}} = e^{\lambda^2 |u|^2/2}.$$

So $\quad u^T Z \sim \mathcal{N}(0, |u|^2)$.

$\cancel{\mathbb{E}[Z]} = \quad \mathbb{E}[Z] = 0 \qquad Var(Z) = I_n = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}$

So $\quad Z \sim \mathcal{N}(0, I_n)$.

Let $\quad X = \mu + \sigma Z$, where $\mu \in \mathbb{R}^n$.

X is a Gaussian vector as a linear

transformation of a Gaussian variable.

$\mathbb{E}[X] = \mu$

$$\mathrm{Var}(X) = \mathbb{E}[(X-\mu)(X-\mu)^T] = \mathbb{E}[\sigma Z \cdot (\sigma Z)^T]$$

$$= \mathbb{E}[\sigma Z Z^T \sigma^T] = \sigma \, \mathbb{E}[Z \cdot Z^T] \cdot \sigma$$

$$= \sigma \, \mathrm{Var}(Z) \cdot \sigma = \sigma \cdot I_n \cdot \sigma = \sigma \cdot \sigma = V.$$

So $X \sim \mathcal{N}(\mu, V)$.

<u>Density of X</u> $\sim \mathcal{N}(\mu, V)$

- V is positive definite, so all eigenvalues are strictly positive.

$$x = \mu + \sigma z \Rightarrow z = \sigma^{-1}(x-\mu)$$

$$f_X(x) = f_Z(z) \cdot |J| = \prod_{i=1}^{n} \frac{e^{-\frac{z_i^2}{2}}}{(2\pi)^{1/2}} \, |\det \sigma^{-1}|$$

$$f_X(x) = \frac{1}{(2\pi)^{n/2}} e^{-\frac{|z|^2}{2}} \cdot \frac{1}{|\det V|^{1/2}}$$

$$= \frac{1}{\sqrt{(2\pi)^n \det V}} e^{-\frac{z^T z}{2}}$$

$$\cancel{e^{-\frac{z^T z}{2}}} = \cancel{e} \qquad z^T \cdot z = \left(\sigma^{-1}(x-\mu)\right)^T \cdot \sigma^{-1}(x-\mu) =$$

$$= (x-\mu)^T \sigma^{-1} \cdot \sigma^{-1} \cdot (x-\mu) = (x-\mu)^T \cdot V^{-1} \cdot (x-\mu).$$

So $$f_X(x) = \frac{1}{\sqrt{(2\pi)^n \det V}} e^{-\frac{(x-\mu)^T V^{-1}(x-\mu)}{2}}$$

- V is non-negative definite, so could have some 0 eigenvalues.

By an orthogonal change of basis we can assume that

$$V = \begin{pmatrix} U & 0 \\ 0 & 0 \end{pmatrix}$$ where U is an $m \times m$ ~~$m \times m$~~ positive definite matrix.

and $\mu = \begin{pmatrix} \lambda \\ \nu \end{pmatrix}$, $\lambda \in \mathbb{R}^m$ and $\nu \in \mathbb{R}^{n-m}$.

We can write $X = \begin{pmatrix} Y \\ \nu \end{pmatrix}$ where Y has density $f_Y(y) = \frac{1}{\sqrt{(2\pi)^m \det U}} e^{-\frac{(y-\lambda)^T U^{-1} (y-\lambda)}{2}}$.

Let $X = (X_1, \dots, X_n)$ be a Gaussian vector

Suppose $\mathrm{Cov}(X_i, X_j) = 0$ when $i \neq j$.

Then (X_i) are independent Gaussian r.v.'s in $\mathbb{R}$.

<u>Proof</u> The covariance matrix is diagonal so the density of X factorises in the product of the densities of the X_i's.

Another way to see it is with the mgf that again factorises. $\square$

$(X_1, \dots, X_n)$ are indep. iff $\mathrm{Cov}(X_i, X_j) = 0$ whenever $i \neq j$.

for <u>Gaussian</u> vectors.

Bivariate Gaussian

$X = (X_1, X_2)$ $(n=2)$ is a Gaussian vector.

Set $\mu_k = \mathbb{E}[X_k]$, $k = 1,2$, $\sigma_k^2 = \mathrm{Var}(X_k)$

$$\rho = \mathrm{corr}(X_1, X_2) = \frac{\mathrm{Cov}(X_1, X_2)}{\sqrt{\mathrm{Var}(X_1)\,\mathrm{Var}(X_2)}}$$

By Cauchy - Schwartz $\rho \in [-1, 1]$.

$\mu_k \in \mathbb{R}$, $\sigma_k \in [0, \infty)$.

$$V = \mathrm{Var}(X) = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

$x = (x_1, x_2)^T$

$$x^T V x = (1-\rho)(\sigma_1^2 x_1^2 + \sigma_2^2 x_2^2) + \rho(\sigma_1 x_1 + \sigma_2 x_2)^2$$

$$= (1+\rho)(\sigma_1^2 x_1^2 + \sigma_2^2 x_2^2) - \rho(\sigma_1 x_1 - \sigma_2 x_2)^2$$

$\forall\ \rho \in [-1, 1]$ we have $x^T V x \geqslant 0$.

so V is always non-negative definite for all choices of parameters in the given range.

When $\rho = 0$ and $\sigma_1, \sigma_2 > 0$, then

$$f_{X_1,X_2}(x_1,x_2) = \prod_{k=1}^{2} \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left(-\frac{(x_k-\mu_k)^2}{2\sigma_k^2}\right).$$

i.e. as we also saw before X_1 and X_2 are indep.

Let $a \in \mathbb{R}$, then

$$\mathrm{Cov}(X_2 - aX_1, X_1) = \mathrm{Cov}(X_1, X_2) - a\,\mathrm{Var}(X_1) =$$
$$= \rho\,\sigma_1\sigma_2 - a\sigma_1^2.$$

Take $a = \frac{\rho\sigma_2}{\sigma_1}$, then if $Y = X_2 - aX_1$

$\mathrm{Cov}(X_1, Y) = 0$. and we can write

$$\begin{pmatrix} X_1 \\ Y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -a & 1 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

so $\begin{pmatrix} X_1 \\ Y \end{pmatrix}$ is another Gaussian vector.

$\begin{pmatrix} X_1 \\ Y \end{pmatrix}$ is Gaussian and $\mathrm{Cov}(X_1, Y) = 0$

it follows that X_1 is indep. of Y.

We can write

$$X_2 = X_2 - aX_1 + aX_1 = Y + aX_1$$

So $\mathbb{E}[X_2 | X_1] = \mathbb{E}[Y | X_1] + a\, \mathbb{E}[X_1 | X_1] =$

$$= \underset{(Y \perp\!\!\!\perp X_1)}{\mathbb{E}[Y]} + a \cdot X_2 .$$

Multivariate CLT (non-examinable)

Let X be a Gaussian vector in $\mathbb{R}^k$ with $\sigma_i^2 < \infty$ $\forall\, i = 1, \dots, k$. Suppose that X has covariance matrix Σ. Let $X_1, \dots$ be iid copies of X. Then

$$S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \mathbb{E}[X_i]) \xrightarrow[n \to \infty]{(d)} \mathcal{N}(0, \Sigma)$$

"$\forall$" $B \subseteq \mathbb{R}^k$ $\quad \mathbb{P}(S_n \in B) \longrightarrow \mathbb{P}(\mathcal{N}(0, \Sigma) \in B)$

Balls in bins

n balls indistinguishable.

For every ball we pick a bin uniformly at random and place it there independently for different balls.

Let X_i = # of balls in bin i

Define the maximum load

$$M_n = \max_{i \leq n} X_i .$$

$\forall i \quad X_i \sim \text{Bin}\left(n, \frac{1}{n}\right).$

<u>Heuristically</u> : $\mathbb{P}(M_n \geq x) \leq n \cdot \mathbb{P}(X_1 \geq x)$

$$\approx n \cdot \mathbb{P}(\text{Poi}(1) \geq x)$$

$\underline{x > \lambda}$ $\quad \mathbb{P}(\text{Poi}(\lambda) \geq x) \leq \exp\left(-x \log \frac{x}{\lambda} - \lambda + x\right)$

$\underline{x < \lambda}$ $\quad \mathbb{P}(\text{Poi}(\lambda) < x) \leq \exp\left(x \log \frac{\lambda}{x} + x - \lambda\right)$. Check!

$\mathbb{P}(\mathrm{Poi}(1) \geq x) \leq \exp\left(-x\log x - 1 + x\right)$

$\ast$ Need $\quad n \cdot e^{-x\log x - 1 + x} \longrightarrow 0 \quad n \to \infty$

$x = (1+\varepsilon) \dfrac{\log n}{\log\log n}$.

<u>Theorem</u> $\quad \dfrac{M_n}{\frac{\log n}{\log\log n}} \xrightarrow[\text{as } n\to\infty]{\mathbb{P}} 1$

i.e. $\quad \forall\, \varepsilon > 0$

$$\mathbb{P}\left(\left| \frac{M_n}{\frac{\log n}{\log\log n}} - 1 \right| > \varepsilon \right) \xrightarrow[n\to\infty]{} 0$$

Let $\quad N \sim \mathrm{Poi}(\lambda)$

Set $\quad X = \sum_{k=1}^{N} \xi_k \quad$ where (ξ_k) iid $\sim \mathrm{Ber}(p)$

Then $\quad X \sim \mathrm{Poi}(\lambda p)\,, \quad N - X \sim \mathrm{Poi}(\lambda(1-p))$
and $\quad X \perp\!\!\!\perp N - X$.

$$\mathbb{P}(X=x, N-X=y) = \mathbb{P}(N=x+y, X=x) =$$

$$= e^{-\lambda} \cdot \frac{\lambda^{x+y}}{(x+y)!} \binom{x+y}{x} p^{x} \cdot (1-p)^{y}$$

<u>Poissonisation</u> Suppose we instead throw

$\mathrm{Poi}(n(1+\varepsilon))$ balls. Let Y_i = load of bin i

Then (Y_i) are iid $\sim \mathrm{Poi}(1+\varepsilon)$.

Call $\tilde{M}_n = \max_{i \le n} Y_i$.

$$\mathbb{P}(M_n \geq x) \leq \mathbb{P}(\tilde{M}_n \geq x, \mathrm{Poi}(n(1+\varepsilon)) \geq n) +$$

$$+ \mathbb{P}(\mathrm{Poi}(n(1+\varepsilon)) < n)$$

$$\leq \mathbb{P}(\tilde{M}_n \geq x) + \mathbb{P}(\mathrm{Poi}(n(1+\varepsilon)) < n)$$

$$\mathbb{P}(\mathrm{Poi}(n(1+\varepsilon)) \leq n) \leq \exp\left(n \cdot \log(1+\varepsilon) - \varepsilon n\right)$$

$$\leq \exp\left(-\frac{n\varepsilon^2}{10}\right) \text{ for } \varepsilon \in (0,1).$$

$$\mathbb{P}\left(M_n \geq (1+\varepsilon)\frac{\log n}{\log\log n}\right) \leq$$

$$\leq \mathbb{P}\left(\tilde{M}_n \geq (1+\varepsilon)\frac{\log n}{\log\log n}\right) + \underbrace{e^{-\frac{n\varepsilon^2}{10}}}_{\xrightarrow[n\to\infty]{} 0}$$

$\tilde{M}_n = \max_{i \leq n} Y_i$ where

$Y_i \sim \mathrm{Poi}(1+\varepsilon)$ iid

Balls in bins

$X_i = \#$ of balls in bin i

$$M_n = \max_{i \le n} X_i$$

$$\mathbb{P}\left(M_n > (1+\varepsilon)\frac{\log n}{\log\log n}\right) \xrightarrow[n\to\infty]{} 0 .$$

$$\mathbb{P}\left(\max_{i \le n} Y_i > (1+\varepsilon)\frac{\log n}{\log\log n}\right) \to 0$$

where (Y_i) iid $\sim \mathrm{Poi}(1+\varepsilon)$.

$$\mathbb{P}\left(\max_{i \le n} Y_i > (1+\varepsilon)\frac{\log n}{\log\log n}\right) \le n \cdot \mathbb{P}\left(Y_1 > (1+\varepsilon)\frac{\log n}{\log\log n}\right)$$

$$\le n \cdot \exp\left(-(1+\varepsilon)\frac{\log n}{\log\log n}\log\left(\frac{\log n}{\log\log n}\right) - (1+\varepsilon) + (1+\varepsilon)\frac{\log n}{\log\log n}\right)$$

$$\le n \exp\left(-(1+\varepsilon)\log n + 10\,\frac{\log n \cdot \log\log\log n}{\log\log n}\right)$$

$$\xrightarrow[n\to\infty]{} 0$$

NTS $\quad \mathbb{P}\left(M_n < (1-\varepsilon)\dfrac{\log n}{\log\log n} \right) \to 0 \quad n\to\infty$

Throw instead $\mathrm{Poi}(n(1-\varepsilon))$ balls.

$$\mathbb{P}\left(\mathrm{Poi}(n(1-\varepsilon)) > n \right) \leq e^{-\frac{n\varepsilon^2}{10}}.$$

$$\mathbb{P}\left(M_n < (1-\varepsilon)\frac{\log n}{\log\log n} \right) \leq e^{-n\varepsilon^2/10} +$$

$$+ \mathbb{P}\left(\tilde{M}_n < (1-\varepsilon)\frac{\log n}{\log\log n} \right),$$

where $\tilde{M}_n = \max \tilde{Y}_i$

$\tilde{Y}_i \sim \mathrm{Poi}(1-\varepsilon)$ iid.

$$\mathbb{P}\left(\tilde{M}_n < (1-\varepsilon)\frac{\log n}{\log\log n} \right) = \left(\mathbb{P}\left(\tilde{Y}_1 < (1-\varepsilon)\frac{\log n}{\log\log n} \right) \right)^n$$

$$\mathbb{P}\left(\tilde{Y}_1 \geq \underbrace{(1-\varepsilon)\frac{\log n}{\log\log n}}_{M} \right) \geq e^{-(1-\varepsilon)} \cdot \frac{(1-\varepsilon)^M}{M!}$$

$$\mathbb{P}(\tilde{M}_n < M) \leq \left(1 - e^{-(1-\varepsilon)} \cdot \frac{(1-\varepsilon)^M}{M!}\right)^n$$

$$(1-x \leq e^{-x}) \qquad \leq \exp\left(-n\, e^{-(1-\varepsilon)} \cdot \frac{(1-\varepsilon)^M}{M!}\right)$$

We proved $M! \sim \sqrt{2\pi M}\, e^{-M} \cdot M^M$

For M large enough

$$M! \leq M \cdot \left(\frac{M}{e}\right)^M .$$

Using this and the value of M

we get $\mathbb{P}(\tilde{M}_n < M) = o(1)$ as $n \to \infty$.

The power of 2 choices

Azar, Karlin, ...

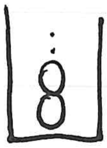
Probability and computing
  Mitzenmacher and Upfal.

Throw n balls into n bins. with replacement
Every time pick $d \geq 2$ bins at random
and place the ball in the least
loaded bin.

After all balls are placed, the maximum
load is $\frac{\log\log n}{\log d} + O(1)$ w.p. $1-o(1)$.

Chernoff for Binomial

$$\mathbb{P}(\mathrm{Bin}(n,p) \geq 2np) \leq e^{-np/3}.$$



height of a ball = # of balls already in the bin it is placed + 1.

Let ν_i = # of bins of load $\geq i$
and μ_i = # of balls of height $\geq i$

Then $\nu_i \leq \mu_i$

We want to find a sequence (β_i) so that

w.h.p. $\nu_i \leq \beta_i$, $\forall i \leq i^*$, where i^* is

to be determined and we will show that $i^* = \frac{\log\log n}{\log d}$. Then at this time

β_{i^*} will be of order $\log n$ and we can finish the proof easily from there.

Suppose we condition on $\nu_i \leq \beta_i$.

$\mathbb{P}(\text{a ball has height} \geq i+1 \mid \nu_i \leq \beta_i)$ "$\leq$" $\left(\frac{\beta_i}{n}\right)^d$

because all the d choices have to come from bins with load $\geq i$.

So $\mathbb{P}(\nu_{i+1} > k) \leq \mathbb{P}\left(\mathrm{Bin}\left(n, \left(\frac{\beta_i}{n}\right)^d\right) > k\right)$.

So if we take $\beta_{i+1} = 2n \cdot \left(\frac{\beta_i}{n}\right)^d$ we get

$\mathbb{P}(\nu_{i+1} > \beta_{i+1}) = o(1)$ by the Chernoff bound for the Binomial.

The proof rigorises the conditioning.

<u>Lemma</u> Let $X_1, X_2, \ldots$ be a sequence of variables and let $Y_i = Y_i(X_1, \ldots, X_i)$ be binary variables.

If $\mathbb{P}(Y_i = 1 \mid X_1, \ldots, X_{i-1}) \leq p$, then

$$\mathbb{P}\left(\sum_{i=1}^{n} Y_i > k\right) \leq \mathbb{P}(\mathrm{Bin}(n, p) > k).$$

<u>Proof</u> Each Y_i is upper bounded by $\mathrm{Ber}(p)$. and for the sum we use induction. $\square$

<u>Proof</u> Azar, Broder, Karlin, Upfal
Balanced Allocations

$h(t)$ = height of t-th ball

$\nu_i(t)$ = # of bins of load $\geq i$ at time t

(at time t = after the t-th ball is placed)

$\mu_i(t)$ = # of balls of height $\geq i$ at time t.

$\nu_i(n) = \nu_i$ and $\mu_i(n) = \mu_i$.

$\nu_i(t) \leq \mu_i(t) \quad \forall\, i, t.$

Want to define a sequence (β_i) s.t.
$\nu_i \leq \beta_i \quad \forall\, i < i^*$ with high prob.

Let $\beta_4 = \frac{n}{4}$ and $\beta_{i+1} = 2n \cdot \left(\frac{\beta_i}{n}\right)^d$

Define $\mathcal{E}_i = \{\nu_i \leq \beta_i\}$. $\mathbb{P}(\mathcal{E}_4) = 1$.

<u>Claim</u> $\forall\ 4 \le i < i^*$, where i^* is to be determined

$$\mathbb{P}(\mathcal{E}_{i+1}^c) \le \mathbb{P}(\mathcal{E}_i^c) + \frac{1}{n^2}.$$

<u>Proof</u> Define

$$Y_t = 1\left(h(t) \geq i+1 \text{ and } \nu_i(t-1) \le \beta_i\right).$$

Let w_j be the bins selected by the j-th ball. Then

$$\mathbb{P}(Y_t = 1 \mid w_1, \dots, w_{t-1}) \le \left(\frac{\beta_i}{n}\right)^d = p_i$$

By lemma from last time

$$\mathbb{P}\left(\sum_{t=1}^{n} Y_t > k\right) \le \mathbb{P}\left(\mathrm{Bin}(n, p_i) > k\right) \ \forall k.$$

$$\mathbb{P}(\mathcal{E}_{i+1}^c \mid \mathcal{E}_i) = \mathbb{P}(\nu_{i+1} > \beta_{i+1} \mid \mathcal{E}_i)$$

$$\le \mathbb{P}(\mu_{i+1} > \beta_{i+1} \mid \mathcal{E}_i)$$

Conditioned on $\mathcal{E}_i$, $Y_t = \mathbb{1}(h(t) \geq i+1)$,

and hence $\sum_{t=1}^{n} Y_t = \mu_{i+1}$.

So $\mathbb{P}(\mu_{i+1} > \beta_{i+1} \mid \mathcal{E}_i) = \mathbb{P}(\sum_{t=1}^{n} Y_t > \beta_{i+1} \mid \mathcal{E}_i) =$

$$= \frac{\mathbb{P}(\sum_{t=1}^{n} Y_t > \beta_{i+1}, \mathcal{E}_i)}{\mathbb{P}(\mathcal{E}_i)} \leq \frac{\mathbb{P}(\sum_{t=1}^{n} Y_t > \beta_{i+1})}{\mathbb{P}(\mathcal{E}_i)}.$$

$$\leq \frac{\mathbb{P}(\mathrm{Bin}(n, p_i) > \beta_{i+1})}{\mathbb{P}(\mathcal{E}_i)}$$

$$\beta_{i+1} = 2n \cdot \left(\frac{\beta_i}{n}\right)^d = 2np_i$$

$$\mathbb{P}(\mathrm{Bin}(n, p_i) > \underbrace{\beta_{i+1}}^{= 2np_i}) \leq e^{-\frac{np_i}{3}}$$

For all i s.t. $np_i \geq 6 \log n$ we have

$$\mathbb{P}(\mathcal{E}_{i+1}^c \mid \mathcal{E}_i) \leq \frac{1}{n^2 \mathbb{P}(\mathcal{E}_i)}.$$

So $\mathbb{P}(\mathcal{E}_{i+1}^c) = \mathbb{P}(\mathcal{E}_{i+1}^c \mid \mathcal{E}_i) \cdot \mathbb{P}(\mathcal{E}_i) + \mathbb{P}(\mathcal{E}_{i+1}^c \mid \mathcal{E}_i^c) \cdot \mathbb{P}(\mathcal{E}_i^c)$

$$\Rightarrow \mathbb{P}(\mathcal{E}_{i+1}^c) \le \frac{1}{n^2} + \mathbb{P}(\mathcal{E}_i^c).$$

as long as $p_i n \ge 6 \log n$.

Define $i^* = \min\{i \ge 0 : np_i < 6 \log n\}$.

<u>Claim</u> $i^* \le \dfrac{\log\log n}{\log d} + O(1)$.

Suffices to show by induction that

$$\beta_{i+4} = \frac{n}{2^{2d^i - \sum_{j=0}^{i-1} d^j}}$$

$i=0$ ✓ $i \rightarrow$ induction hyp. & def. of β_i.

$$\beta_{i+4} \le \frac{n}{2^{d^i}}$$

So $\mathbb{P}(\mathcal{E}_{i^*}^c) \le \dfrac{i^*}{n^2}$

$$\beta_{i^*+1} = 2n \cdot p_{i^*} < 2\cancel{n} \cdot \frac{6 \log n}{\cancel{n}} = 12 \log n.$$

$$\mathbb{P}(\nu_{i^*+1} \geq 18 \log n \mid \mathcal{E}_{i^*}) \leq$$

$$\leq \mathbb{P}(\mu_{i^*+1} \geq 18 \log n \mid \mathcal{E}_{i^*}) \leq$$

$$\leq \frac{\mathbb{P}\left(\mathrm{Bin}\left(n, \frac{6 \log n}{n}\right) \geq 18 \log n\right)}{\mathbb{P}(\mathcal{E}_{i^*})}$$

$$\underset{\text{Chernoff}}{\leq} \frac{e^{-2\log n}}{\mathbb{P}(\mathcal{E}_{i^*})} = \frac{1}{n^2 \mathbb{P}(\mathcal{E}_{i^*})}.$$

$$\mathbb{P}(\nu_{i^*+1} \geq 18 \log n) \leq \frac{1}{n^2} + \mathbb{P}(\mathcal{E}_{i^*}^c)$$

$$\leq \frac{i^*+1}{n^2}.$$

$$\{\nu_{i^*+3} \geq 1\} \subseteq \{\mu_{i^*+3} \geq 1\} \subseteq \{\mu_{i^*+2} \geq 2\}.$$

$$\mathbb{P}(\mu_{i^*+2} \geq 2 \mid \nu_{i^*+1} < 18 \log n) \leq$$

$$\leq \mathbb{P}\left(\mathrm{Bin}\left(n, \left(\frac{18 \log n}{n}\right)^d\right) \geq 2\right) / \mathbb{P}(\nu_{i^*+1} < 18 \log n)$$

$$\mathbb{P}(\mu_{i^*+2} \geq 2) \leq \mathbb{P}\left(\mathrm{Bin}\left(n, \left(\frac{18\log n}{n}\right)^d\right) \geq 2\right) + \mathbb{P}(\nu_{i^*+1} \geq 18 \log n)$$

$$\leq \binom{n}{2} \cdot \left(\frac{18\log n}{n}\right)^{2d} + \frac{i^*+1}{n^2}$$

$$\leq \frac{n^2}{n^{2d}}(18\log n)^{2d} + \frac{i^*+1}{n^2} = o\left(\frac{1}{n}\right)$$

So $\mathbb{P}(\nu_{i^*+3} \geq 1) \leq \mathbb{P}(\mu_{i^*+2} \geq 2) = o(\frac{1}{n})$. $\square$