Probability

April 13, 2022

Contents

Start of lecture 1

Example 0. Dice: outcomes $1, 2, \ldots, 6$.

- $\mathbb{P}(2) = \frac{1}{6}$
- $\mathbb{P}(\text{multiple of } 3) = \frac{2}{6} = \frac{1}{3}$.
- $$\begin{aligned} \mathbb{P}(\text{prime or a multiple of } 3) &= \frac{1}{3} + \frac{1}{2} = \frac{5}{6} \\ &= \frac{4}{6} = \frac{2}{3} \\ &= \frac{1}{3} + \frac{1}{2} - \mathbb{P}(\text{prime } \textit{and} \text{ a multiple of } 3) \\ &= \frac{1}{3} + \frac{1}{2} - \frac{1}{6} = \frac{2}{3} \end{aligned}$$
- $\mathbb{P}(\text{not a multiple of } 3) = \frac{2}{3}$.

1 Formal Setup

Definition.

- *Sample space* Ω, a set of *outcomes*.
- $\mathcal{F}$ a collection of subsets of Ω (called *events*).
- $\mathcal{F}$ is a σ-*algebra* ("sigma-algebra") if:
  - F1 $\Omega \in \mathcal{F}$
  - F2 if $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$ ($A^c := \Omega \setminus A$)
  - F3 $\forall$ countable collections $(A_n)_{n \ge 1}$ in $\mathcal{F}$ the union

    $$\bigcup_{n \ge 1} A_n \in \mathcal{F}$$

    also.

Given σ-algebra $\mathcal{F}$ on Ω, function $\mathbb{P} : \mathcal{F} \to [0, 1]$ is a *probability measure* if

P2 $\mathbb{P}(\Omega) = 1$

P3 $\forall$ countable collections $(A_n)_{n \ge 1}$ of disjoint events in $\mathcal{F}$:

$$\mathbb{P}\left(\bigcup_{n \ge 1} A_n \right) = \sum_{n \ge 1} \mathbb{P}(A_n).$$

(P1 was historically taken to state that $\mathbb{P}(A) \ge 0$, but this is already captured by the notation $\mathbb{P} : \mathcal{F} \to [0, 1]$).

Then $(\Omega, \mathcal{F}, \mathbb{P})$ is a *probability space.*

Revisiting dice example

For a dice we have:

$$\Omega = \{1, 2, \dots, 6\}$$

$$\mathbb{P}(1 \text{ or } 2 \text{ or } 3 \text{ or } 4 \text{ or } 5 \text{ or } 6) = 1.$$

$$\mathcal{F} = \mathcal{P}(\Omega)$$

Question: Why $\mathbb{P} : \mathcal{F} \to [0,1]$ not $\mathbb{P} : \Omega \to [0,1]$?

Ω finite / countable

- In general: $\mathcal{F}$ = all subsets of Ω. $(\mathbb{P}(\Omega))$.
- $\mathbb{P}(2)$ is shorthand for $\mathbb{P}(\{2\})$.
- $\mathbb{P}$ is determined by $(\mathbb{P}(\{\omega\}), \forall\, \omega \in \Omega)$. (eg unfair dice)

Ω uncountable

- For example $\Omega = [0,1]$. Want to choose a real number, all equally likely.
- If $\mathbb{P}(\{0\}) = \alpha > 0$, then

$$\mathbb{P}\left(\left\{0, 1, \frac{1}{2}, \dots, \frac{1}{n}\right\}\right) = (n+1)\alpha$$

  ⨳ if n large as $\mathbb{P} > 1$.

- So $\mathbb{P}(\{0\}) = 0$, or $\mathbb{P}(\{0\})$ is undefined.
- What about $\mathbb{P}(\{x : x \le \frac{1}{3}\})$?
  - ? "Add up" all $\mathbb{P}(\{x\})$ for $x \le \frac{1}{3}$.

Example. $\Omega = \{f : \text{continuous on } [0,1] \to \mathbb{R},\ f(0) = 1\}$. What is $\mathbb{P}$(differentiable)?

1.1 From the axioms

- $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$. *Proof.* A, A^c are disjoint. $A \cup A^c = \Omega$ and hence

$$\mathbb{P}(A) + \mathbb{P}(A^c) \overset{P3}{=} \mathbb{P}(\Omega) \overset{P2}{=} 1$$

□

- $\mathbb{P}(\emptyset) = 0$.
- If $A \subseteq B$ then $\mathbb{P}(A) \le \mathbb{P}(B)$.
- $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

1.2 Examples of Probability Spaces

Ω finite, $\Omega = \{\omega_1, \dots, \omega_n\}$, $\mathcal{F}$ = all subsets uniform choice (equally likely).

$$\mathbb{P} : \mathcal{F} \to [0,1], \quad \mathbb{P}(A) = \frac{|A|}{|\Omega|}.$$

In particular:

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

Example 1. Choosing without replacement n indistinguishable marbles labelled $\{1, \dots, n\}$. Pick $k \le n$ marbles uniformly at random. Here:

$$\Omega = \{A \subseteq \{1, \dots, n\} : |A| = k\} \qquad |\Omega| = \binom{n}{k}$$

Example 2. Well-shuffled deck of cards. Uniformly chosen *permutation* of 52 cards.

$$\Omega = \{\text{all permutation of 52 cards}\} \qquad |\Omega| = 52!$$

$$\mathbb{P}(\text{first three cards have the same suit}) = \frac{52 \times 12 \times 11 \times 49!}{52!} = \frac{22}{425}$$

Note: $= \frac{12}{51} \times \frac{11}{50}$.

Start of lecture 2

Example 3 (Coincident Birthdays). n people. What is the probability that at least two share a birthday?
Assumptions:

- No leap years! (365 days)
- All birthdays equally likely.

Now note that

$$\Omega = \{1, \dots, 365\}^n \qquad \mathcal{F} = \mathcal{P}(\Omega)$$

$$A = \{\text{at least 2 people share a birthday}\}$$

$$A^c = \{\text{all } n \text{ birthdays different}\}$$

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

so

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

$$\begin{cases} n = 22: & \mathbb{P}(A) \approx 0.476 \\ n = 23: & \mathbb{P}(A) \approx 0.507 \end{cases}$$

$n \geq 366$: $\mathbb{P}(A) = 1$.

1.3 Choosing uniformly from infinite countable set

(For example $\Omega = \mathbb{N}$ or $\Omega = \mathbb{Q} \cap [0, 1]$) Suppose possible, then

- $\mathbb{P}(\{\omega\}) = \alpha > 0 \ \forall \ \omega \in \Omega$. Then

$$\mathbb{P}(\Omega) = \sum_{\omega \in \Omega} \mathbb{P}(\{\omega\}) = \sum_{\omega \in \Omega} \alpha = \infty \qquad ※$$

- $\mathbb{P}(\{\omega\}) = 0 \ \forall \ \omega \in \Omega$. Then

$$\mathbb{P}(\Omega) = \sum_{\omega \in \Omega} \mathbb{P}(\{\omega\}) = \sum_{\omega \in \Omega} 0 = 0 \qquad ※$$

Note possible, but still, there exist lots of interesting probability measures of $\mathbb{N}$!

1.4 Combinatorial Analysis

Subsets: Ω finite. $|\Omega| = n$.
Question: How many ways to *partition* Ω into k disjoint subsets $\Omega_1, \dots, \Omega_k$ with $|\Omega_i| = n_i$

(with $\sum_{i=1}^{k} n_i = n$)?

$$
\begin{aligned}
M &= \binom{n}{1}\binom{n-n_1}{n_2}\binom{n-n_1-n_2}{n_3}\cdots\binom{n-(n_1+\cdots+n_{k-1}}{n_k} \\
&= \frac{n!}{n_1!\cancel{(n-n_1)!}} \times \frac{\cancel{(n-n_1)!}}{n_2!\cancel{(n-n_1-n_2)!}} \times \cdots \times \frac{\cancel{[n-(n_1+\cdots+n_{k-1})]!}}{n_k!0!} \\
&= \frac{n!}{n_1!n_2!\cdots n_k!} \\
&=: \binom{n}{n_1, n_2, \ldots, n_k}
\end{aligned}
$$

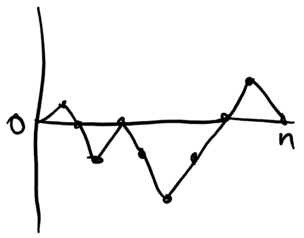
Key sanity check: Does ordering of subsets matter? For example, do we have

$$\big[\Omega_2 = \{3,4,7\}, \Omega_3 = \{1,5,8\}\big] \overset{\text{different}}{=} \big[\Omega_2 = \{1,5,8\}, \Omega_3 = \{3,4,7\}\big]?$$

Yes!

Random Walks

$$\Omega = \{(X_0, X_1, \ldots, X_n) : X_0 = 0, |X_k - X_{k-1}| = 1, k = 1, \ldots, n\} \qquad |\Omega| = 2^n.$$



Could ask: $\mathbb{P}(X_n = 0)$?

$$\mathbb{P}(X_n = n) = \frac{1}{2^n}$$

$$\mathbb{P}(X_n = 0) = 0 \qquad \text{if } n \text{ is odd}$$

If n is even?
Idea - Choose $\frac{n}{2}$ ks for $X_k = X_{k-1} + 1$ and the rest $X_k = X_{k-1} - 1$. So

$$
\begin{aligned}
\mathbb{P}(X_n = 0) &= 2^{-n}\binom{n}{n/2} \\
&= \frac{n!}{2^n\left[\left(\frac{n}{2}\right)!\right]^2}
\end{aligned}
$$

Question: What happens when n is large?

Stirling's Formula

Notation. (a_n), $(b_n 0$ two sequences.

Say $a_n \sim b_n$ as $n \to \infty$ if $\frac{a_n}{b_n} \to 1$ as $n \to \infty$. For example, $n^2 + 5n + \frac{6}{n} \sim n^2$. Non-example: $\exp\left(n^2 + 5n + \frac{6}{n}\right) \not\sim \exp(n^2)$.

Theorem (Stirling)**.**

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

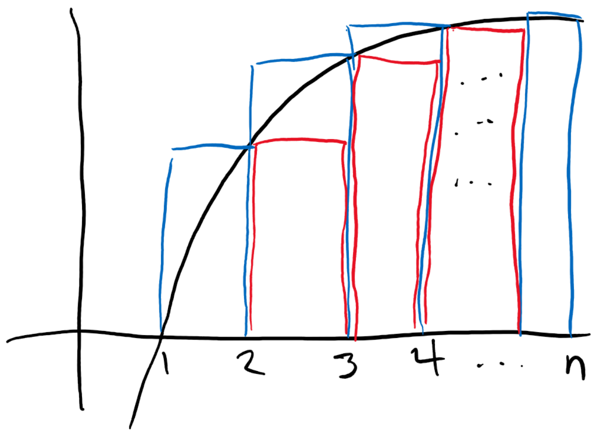
as $n \to \infty$.
Weaker version:

$$\log(n!) \sim n \log n.$$

Start of lecture 3

Proof (weaker version).

$$\log(n!) = \log 2 + \log 3 + \cdots + \log n.$$



$$\underbrace{\int_1^n \log x \mathrm{d}x}_{\text{“Upper integral”}} \le \log(n!) \le \underbrace{\int_1^{n+1} \log x \mathrm{d}x}_{\text{“Lower integral”}}$$

$$\underbrace{n \log n - n = 1}_{\sim n \log n} \le \log(n!) \le \underbrace{(n+1)\log(n+1) - n}_{\sim n \log n}$$

Hence $\log(n!) \sim n \log n$. □

Key idea: *Sandwiching* between lower/upper integrals.
Useful:

- $\log x$ is increasing
- $\log x$ has nice integral!

(Ordered) Compositions

A *composition* of m with k parts is sequence $(m_1, \dots, m_k)$ of non-negative integers with

$$m_1 + \cdots + m_k = m.$$

For example, $3+0+1+2=6$. Bijection between compositions and sequences of m stars and $k-1$ dividers (stars and bars). So number of compositions is $\binom{m+k-1}{m}$.
Comments: Q11 on example sheet 1.

Properties of Probability Measures

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

- P1:
$$\mathbb{P} : \mathcal{F} \to [0, 1]$$
- P2: $\mathbb{P}(\Omega) = 1$.
- P3:
$$\mathbb{P}\left(\bigcup_{n \ge 1} A_n\right) = \sum_{n \ge 1} \mathbb{P}(A_n)$$
$(A_n)_{n \ge 1}$ disjoint. "Countable additivity".

(1) Countable sub-additivity

$(A_n)_{n \ge 1}$ sequence of events in $\mathcal{F}$. Then

$$\mathbb{P}\left(\bigcup_{n \ge 1} A_n\right) \le \sum_{n \ge 1} \mathbb{P}(A_n).$$

Intuition: this sum can "double count" some sub-events.
Proof. Idea: rewrite $\bigcup_{n \ge 1} A_n$ as a *disjoint* union. Define $B_1 = A_1$ and $B_n = A_n \setminus (A_1 \cup \cdots \cup A_{n-1})$ for $n \ge 2$ (which is in $\mathcal{F}$ by example sheet). So

- $\bigcup_{n \ge 1} B_n = \bigcup_{n \ge 1} A_n$
- $(B_n)_{n \ge 1}$ disjoint (by construction)
- $B_n \subseteq A_n \implies \mathbb{P}(B_n) \le \mathbb{P}(A_n)$ (by example sheet)

Hence

$$\mathbb{P}\left(\bigcup_{n \ge 1} A_n\right) = \mathbb{P}\left(\bigcup_{n \ge 1} B_n\right) = \sum_{n \ge 1} \mathbb{P}(B_n) \le \sum_{n \ge 1} \mathbb{P}(A_n).$$

□

(1) Continuity

$(A_n)_{n\geq 1}$ is increasing sequence of events in $\mathcal{F}$ i.e. $A_n \subseteq A_{n+1}$. Then $\mathbb{P}(A_n) \leq \mathbb{P}(A_{n+1})$. So $\mathbb{P}(A_n)$ converges as $n \to \infty$. (Because bounded and increasing.) In fact, $\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}\left(\bigcup_{n\geq 1} A_n\right)$.

Proof. Re-use the B_ns!

- $\bigcup_{k=1}^{n} B_k = A_n$ (disjoint union)
- $\bigcup_{n\geq 1} B_n = \bigcup_{n\geq 1} A_n$

$$\mathbb{P}(A_n) = \sum_{k=1}^{n} \mathbb{P}(B_k) \to \sum_{k\geq 1} \mathbb{P}(B_k)$$

$$\mathbb{P}\left(\bigcup_{n\geq 1} A_n\right) = \mathbb{P}\left(\bigcup_{n\geq 1} B_n\right) = \sum_{n\geq 1} \mathbb{P}(B_n)$$

□

Try Q6.

(3) Inclusion-Exclusion Principle

Background: $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.
Similarly: for $A, B, C \in \mathcal{F}$

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

Theorem (Inclusion Exclusion Principle)**.** Let $A_1, A_2, \dots, A_n \in \mathcal{F}$. Then:

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathbb{P}(A_i) - \sum_{1\leq i_1 < i_2 \leq n} \mathbb{P}(A_{i_1} \cap A_{i_2}) + \sum_{1\leq i_1 < i_2 < i_3 \leq n} \mathbb{P}(A_{i_1} \cap A_{i_2} \cap A_{i_3}) - \cdots + (-1)^{n+1}\mathbb{P}(A_1 \cap \cdots \cap A_n)$$

Or, abbreviated:

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{\substack{I\subset\{1,\dots,n\} \\ I\neq\emptyset}} (-1)^{|I|+1}\mathbb{P}\left(\bigcap_{i\in I} A_i\right)$$

Start of lecture 4

Proof. Use induction $n^{-1} \mapsto n$. For $n = 2$, check Example Sheet 1, Q4(e). For the inductive step:

$$\begin{aligned}
\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) &= \mathbb{P}\left(\left(\bigcup_{i=1}^{n-1} A_i\right) \cup A_n\right) \\
&= \mathbb{P}\left(\bigcup_{i=1}^{n-1} A_i\right) + \mathbb{P}(A_n) - \mathbb{P}\left(\left(\bigcup_{i=1}^{n-1} A_i\right) \cap A_n\right)
\end{aligned}$$

Idea:

$$\begin{aligned}
\left(\bigcup_{i=1}^{n-1} A_i\right) \cap A_n &= \bigcup_{i=1}^{n-1} (A_i \cap A_n) \\
\implies \bigcap_{i \in J} (A_i \cap A_n) &= \bigcap_{i \in J \cup \{n\}} A_i
\end{aligned}$$

$(J \subset \{1, \ldots, n-1\})$.

$$\begin{aligned}
\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) &= \sum_{\substack{J \subset \{1,\ldots,n-1\} \\ J \neq \emptyset}} (-1)^{|J|+1} \mathbb{P}\left(\bigcap_{i \in J} A_i\right) + \mathbb{P}(A_n) - \sum_{\substack{J \subset \{1,\ldots,n-1\} \\ J \neq \emptyset}} (-1)^{|J|+1} \mathbb{P}\left(\bigcap_{i \in J \cup \{n\}} A_i\right) \\
&= \sum_{\substack{I \subset \{1,\ldots,n-1\} \\ I \neq \emptyset}} (-1)^{|I|+1} \mathbb{P}\left(\bigcap_{i \in I} A_i\right) + \mathbb{P}(A_n) + \sum_{\substack{I \subset \{1,\ldots,n\} \\ n \in I, |I| \geq 2}} (-1)^{|I|+1} \mathbb{P}\left(\bigcap_{i \in I} A_i\right) \\
&= \sum_{\substack{I \subset \{1,\ldots,n\} \\ I \neq \emptyset}} (-1)^{|I|+1} \mathbb{P}\left(\bigcap_{i \in I} A_i\right)
\end{aligned}$$

Where $J \cup \{n\} \mapsto I$, so $-(-1)^{|J|+1} \mapsto (-1)^{|I|}$. □

Bonferroni Inequalities

Question: What if you *truncate* Inclusion-Exclusion Principle?
Recall: $\mathbb{P}(\cup A_i) \leq \sum \mathbb{P}(A_i)$ (*union bound*).

- When r is even:

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

- When r is odd:

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

Question: When is it good to truncate at for example $r = 2$?
Proof. Induction on r and n. For r odd:

$$\begin{aligned}
\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) &= \mathbb{P}\left(\bigcup_{i=1}^{n-1} A_i\right) + \mathbb{P}(A_n) - \mathbb{P}\left(\bigcup_{i=1}^{n-1} (A_i \cap A_n)\right) \\
&\le \sum_{\substack{J \subset \{1,\dots,n-1\} \\ I \le |J| \le r}} (-1)^{|J|+1} \mathbb{P}\left(\bigcap_{i \in J} A_i\right) + \mathbb{P}(A_n) - \sum_{\substack{J \subset \{1,\dots,n-1\} \\ 1 \le |J| \le r-1}} (-1)^{|J|+1} \mathbb{P}\left(\bigcap_{i \in J \cup \{n\}} A_i\right) \\
&\le \sum_{\substack{I \subset \{1,\dots,n\} \\ 1 \le |I| \le r}} (-1)^{|I|+1} \mathbb{P}\left(\bigcap_{i \in I} A_i\right)
\end{aligned}$$

r even is similar. □

Counting with Inclusion-Exclusion Principle

Uniform probability measure on Ω, $|\Omega| < \infty$.

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

Then $\forall \ A_1, \dots, A_n \subseteq \Omega$.

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

(and similar for Bonferroni Inequalities).

Example 1. *Surjections* $f : \{1, \dots, n\} \to \{1, \dots, n\}$

$$\Omega = \{f : \{1, \dots, n\} \to \{1, \dots, m\}\} \qquad \text{all functions}$$

$$A = \{f : \mathrm{Im}(f) = \{1, \dots, m\}\} \qquad \text{all surjections}$$

$\forall\ i \in \{1, \dots, m\}$. Define

$$B_i = \{f \in \Omega : i \notin \mathrm{Im}(f)\}.$$

Key observations:

- $A = B_1^c \cap \dots \cap B_m^c = (B_1 \cup \dots \cup B_m)^c$.
- $|B_{i_1} \cap \dots \cap B_{i_k}|$ is nice to calculate! In particular, it is

$$|\{f \in \Omega : i_1, \dots, i_k \notin \mathrm{Im}(f)\}| = (m-k)^n.$$

Inclusion-Exclusion Principle implies:

$$\begin{aligned} |B_1 \cup \dots \cup B_m| &= \sum_{k=1}^m (-1)^{k+1} \sum_{i_1 < \dots < i_k} |B_{i_1} \cap \dots \cap B_{i_k}| \\ &= \sum_{k=1}^m (-1)^{k+1} \binom{m}{k} (m-k)^n \end{aligned}$$

$$\begin{aligned} |A| &= m^n - \text{previous expression} \\ &= \sum_{k=0}^m (-1)^k \binom{m}{k} (m-k)^n \end{aligned}$$

Start of lecture 5

Example 2. Derangements (Permutation with no fixed points)

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

$$D = \{\sigma \in \Omega : \sigma(i) \neq i \; \forall \; i = 1, \dots, n\}$$

Question: Is $\mathbb{P}(D) = \frac{|D|}{|\Omega|}$ large or small (when $n \to \infty$)?

$$\forall \; i \in \{1, \dots, n\} : A_i = \{\sigma \in \Omega : \sigma(i) = i\}.$$

- $D = A_1^c \cap \dots \cap A_n^c = \left(\bigcup_{i=1}^n A_i\right)^c$.
- $\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}) = \frac{(n-k)!}{n!}$

Now Inclusion-Exclusion Principle implies:

$$\begin{aligned} \mathbb{P}\left(\bigcup_{i=1}^n A_i\right) &= \sum_{k=1}^n (-1)^{k+1} \sum_{i_1 < \dots i_k} \mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}) \\ &= \sum_{k=1}^n (-1)^{k+1} \binom{n}{k} \frac{(n-k)!}{n!} \end{aligned}$$

So

$$\begin{aligned} \mathbb{P}(D) &= 1 - \mathbb{P}\left(\bigcup_{i=1}^n A_i\right) \\ &= 1 - \sum_{k=1}^n \frac{(-1)^{k+1}}{k!} \\ &= \sum_{k=1}^n \frac{(-1)^k}{k!} \end{aligned}$$

And as $n \to \infty$,

$$\mathbb{P}(D) \to \sum_{k=0}^\infty \frac{(-1)^k}{k!} = e^{-1} \approx 0.37$$

Comments

What if instead we have

$$\Omega' = \{f : \{1, \dots, n\} \to \{1, \dots, n\}\}.$$

$$D = \{f \in \Omega' : f(i) \neq i \; \forall \; i = 1, \dots, n\}.$$

Then

$$\mathbb{P}(D) = \frac{(n-1)^n}{n^n} = \left(1 - \frac{1}{n}\right)^n$$

which also approaches e^{-1} as $n \to \infty$.

- Would be nice to write as a product of probabilities, i.e. $\left(\frac{n-1}{n}\right)^n$, and we will be allowed to do this soon.
- $f(i)$ is a random quantity associated to Ω. (Will be allowed to study $f(i)$ as a *random variable*.)
- Are allowed to toss a fair coin n times.

$$\Omega = \{H, T\}^n$$

Independence

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

Definition.

- Events $A, B \in \mathcal{F}$ are *independent* if

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

(denoted $A \perp\!\!\!\perp B$).

- A *countable* collection of events (A_n) is *independent* if $\forall$ distinct $i_1, \dots, i_k$ we have:

$$\mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_k}) = \prod_{j=1}^{k} \mathbb{P}(A_{i_j}).$$

Note. “Pairwise independence” does not imply independence.

Example. $\Omega = \{(H,H),(H,T),(T,H),(T,T)\}$, $\mathbb{P}(\{\omega\}) = \frac{1}{4}\ \forall\, \omega \in \Omega$. Now define

$$A = \text{first coin in } H = \{(H,H),(H,T)\}$$

$$B = \text{second coin } H = \{(H,H),(T,H)\}$$

$$C = \text{same outcome} = \{(H,H),(T,T)\}.$$

Then we have that

$$\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = \frac{1}{2} \qquad A \cap B = A \cap C = B \cap C = \{(H,H)\}$$

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

so pairwise independent, however

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

so the events are not independent.

Example(s) of Independence

- Define

$$\Omega' = \{f : \{1,\dots,n\} \to \{1,\dots,n\}\}.$$
$$A_i := \{f \in \Omega' : f(i) = i\}.$$
$$\mathbb{P}(A_i) = \frac{n^{n-1}}{n^n} = \frac{1}{n}$$
$$\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}) = \frac{n^{n-k}}{n^n} = \frac{1}{n^k} = \prod_{j=1}^{k} \mathbb{P}(A_{i_j})$$

  Here: (A_i) independent events.

- Define

$$\Omega = \{\sigma : \text{permutation of}\{1,\dots,n\}\}$$
$$A_i = \{\sigma \in \Omega : \sigma(i) = i\}$$

  For $i \neq j$,

$$\mathbb{P}(A_i \cap A_j) = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)} \neq \mathbb{P}(A_i)\mathbb{P}(A_j)$$

  So here, (A_i) are not independent.

Properties

Claim 1. If A is independent of B, then A is also independent of B^c.

Proof.

$$\begin{aligned}\mathbb{P}(A \cap B^c) + \mathbb{P}(A) - \mathbb{P}(A \cap B) \\ &= \mathbb{P}(A) - \mathbb{P}(A)\mathbb{P}(B) \\ &= \mathbb{P}(A)[1 - \mathbb{P}(B)] \\ &= \mathbb{P}(A)\mathbb{P}(B^c)\end{aligned}$$

□

Claim 2. A is independent of $B = \Omega$ and of $C = \phi$.

Proof.

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

And by claim 1, this implies that $A \perp\!\!\!\perp \emptyset$. □

As an exercise, one can further prove that if $\mathbb{P}(B) = 0$ or 1, then A is independent of B.

Conditional Probability

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

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

Definition. The *conditional probability of A given B* is

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

"The probability of A if we know B happened". (for example revealing info in succession).

Example. If A, B independent,

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

"Knowing whether B happened doesn't affect the probability of A."

Start of lecture 6

Properties

- $\mathbb{P}(A \mid B) \geq 0$
- $\mathbb{P}(B \mid B) = \mathbb{P}(\Omega \mid B) = 1$.
- (A_n) disjoint events $\in \mathcal{F}$.

Claim. $\mathbb{P}\left(\bigcup_{n\geq 1} A_n \mid B\right) = \sum_{n\geq 1} \mathbb{P}(A_n \mid B)$.

Proof.

$$\begin{aligned}\mathbb{P}\left(\bigcup_{n\geq 1} A_n \mid B\right) &= \frac{\mathbb{P}\left(\left(\bigcup_{n\geq 1} A_n\right) \cap B\right)}{\mathbb{P}(B)} \\ &= \frac{\mathbb{P}\left(\bigcup_{n\geq 1} (A_n \cap B)\right)}{\mathbb{P}(B)} \\ &= \frac{\sum_{n\geq 1} \mathbb{P}(A_n \cap B)}{\mathbb{P}(B)} \\ &= \sum_{n\geq 1} \mathbb{P}(A \mid B)\end{aligned}$$

□

$\mathbb{P}(\bullet \mid B)$ is a function from $\mathcal{F} \to [0,1]$ that satisfies the rules to be a probability measure Ω. Consider $\Omega' = B$ (especially in finite / countable setting), $\mathcal{F}' = \mathcal{P}(B)$. Then $(\Omega', \mathcal{F}', \mathbb{P}(\bullet \mid B))$ also satisfies the rules to be a probability measure on Ω'.

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

$$\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2 \mid A_1)\mathbb{P}(A_3 \mid A_1 \cap A_2) \cdots \mathbb{P}(A_n \mid A_1 \cap \cdots \cap A_{n-1}).$$

Example. Uniform permutation $(\sigma(1), \sigma(2), \dots, \sigma(n)) \in \sum_n$.

Claim.

$$\mathbb{P}(\sigma(k) = i_k \mid \sigma(i) = i, \dots, \sigma(k-1) = i_{k-1}) = \begin{cases} 0 & \text{if } i_k \in \{i_1, \dots, i_{k-1}\} \\ \frac{1}{n-k+1} & \text{if } i_k \notin \{i_1, \dots, i_{k-1}\} \end{cases}$$

Proof.

$$\begin{aligned} \mathbb{P}(\sigma(k) = i_k \mid \sigma(i) = i, \dots, \sigma(k-1) = i_{k-1}) &= \frac{\mathbb{P}(\sigma(i) = i, \dots, \sigma(k) = i_k)}{\mathbb{P}(\sigma(i) = i_1, \dots, \sigma(k-1) = i_{k-1})} \\ &= \frac{0 \text{ or } \frac{(n-k)!}{n!}}{\frac{(n-k+1)!}{n!}} \\ &= \frac{(n-k)!}{(n-k+1)!} \\ &= \frac{1}{n-k+1} \end{aligned}$$

□

Law of Total Probability and Bayes' Formula

Definition. $(B_1, B_2, \dots) \subset \Omega$ is a *partition* of Ω if:

- $\Omega = \bigcup_{n \geq 1} B_n$
- (B_n) are disjoint

Theorem. (B_n) a finite countable partition of Ω with $B_n \in \mathcal{F}$ and for all n $\mathbb{P}(B_n) > 0$, then for all $A \in \mathcal{F}$:

$$\mathbb{P}(A) = \sum_{n \geq 1} \mathbb{P}(A \mid B_n)\mathbb{P}(B_n).$$

(Sometimes known as "Partition Theorem").

Proof. Note that $\bigcup_{n \geq 1}(A \cap B_n) = A$.

$$\mathbb{P}(A) = \sum_{n \geq 1} \mathbb{P}(A \cap B_n) = \sum_{n \geq 1} \mathbb{P}(A \mid B_n)\mathbb{P}(B_n).$$

□

Theorem (Bayes' Formula)**.**

$$\mathbb{P}(B_n \mid A) = \frac{\mathbb{P}(A \cap B_n)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A \mid B_n)\mathbb{P}(B_n)}{\sum_{m \geq 1} \mathbb{P}(A \mid B_m)\mathbb{P}(B_m)}.$$

Rephrasing for $n = 2$:

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

This allows us for example to calculate $\mathbb{P}(B \mid A)$ given $\mathbb{P}(A)$, $\mathbb{P}(A \mid B)$ and $\mathbb{P}(B)$.

Example 1. Lecture course: $\frac{2}{3}$ probability that it is a weekday, and $\frac{1}{3}$ probability that it is a weekend.

$$\mathbb{P}(\text{forget notes} \mid \text{weekday}) = \frac{1}{8}$$

$$\mathbb{P}(\text{forget notes} \mid \text{weekend}) = \frac{1}{2}.$$

What is $\mathbb{P}(\text{weekend} \mid \text{forget notes})$?

$$B_1 = \{\text{weekend}\}, \qquad B_2 = \{\text{weekend}\}, \qquad A - \{\text{forget notes}\}.$$

Law of Total Probability:

$$\mathbb{P}(= \frac{2}{3} \times \frac{1}{8} + \frac{1}{3} \times \frac{1}{2} = \frac{1}{12} + \frac{1}{6} = \frac{1}{4}.$$

Bayes':

$$\mathbb{P}(B_2 \mid A) = \frac{\frac{1}{3} \times \frac{1}{2}}{\frac{1}{4}} = \frac{2}{3}.$$

Example 2. Disease testing: probability p that you are infected, probability $1 - p$ that you are not.

$$\mathbb{P}(\text{tests positive} \mid \text{infected}) = 1 - \alpha$$

$$\mathbb{P}(\text{test positive} \mid \text{not infected}) = \beta$$

Ideally both α, β are small (and ideally p is small).

$$\mathbb{P}(\text{infected} \mid \text{test positive}).$$

Law of Total Probability:

$$\mathbb{P}(\text{test positive}) = p(1 - \alpha) + (1 - p)\beta.$$

Bayes':

$$\mathbb{P}(\text{infected} \mid \text{positive}) = \frac{p(1 - \alpha)}{p(1 - \alpha) + (1 - p)\beta}.$$

Suppose $p \ll \beta$. Then

$$p(1 - \alpha) \ll (1 - p)\beta$$

Then

$$\mathbb{P}(\text{infected} \mid \text{positive}) \sim \frac{p(1 - \alpha)}{(1 - p)\beta}$$

Start of lecture 7

Example 3 (Simpson's Paradox)**.**

$$A = \{\text{change colour}\}, \qquad B = \{\text{blue}\} \qquad B^c = \{\text{green}\}$$

$$C = \{\text{Cambridge}\} \qquad C^c = \{\text{Oxford}\}$$

$$\mathbb{P}(A \mid B \cap C) > \mathbb{P}(A \mid B^c \cap C)$$

$$\mathbb{P}(A \cap B \cap C^c) > \mathbb{P}(A \mid B^c \cap C^c)$$

$$\not\Longrightarrow \mathbb{P}(A \mid B) > \mathbb{P}(A \mid B^c)$$

Law of Total Probability for Conditional Probabilities

Suppose $C_1, C_2, \dots$ a partition of B.

$$\begin{aligned}
\mathbb{P}(A \mid B) &= \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \\
&= \frac{\mathbb{P}\left(A \cap \left(\bigcup_n C_n\right)\right)}{\mathbb{P}(B)} \\
&= \frac{\mathbb{P}\left(\bigcup_n \left(A \cap C_n\right)\right)}{\mathbb{P}(B)} \\
&= \frac{\sum_n \mathbb{P}(A \cap C_n)}{\mathbb{P}(B)} \\
&= \frac{\sum_n \mathbb{P}(A \mid C_n)\mathbb{P}(C_n)}{\mathbb{P}(B)} \\
&= \sum_n \mathbb{P}(A \mid C_N) \frac{\mathbb{P}(B \cap C_n)}{\mathbb{P}(B)} \\
&= \sum_n \mathbb{P}(A \mid C_n) \frac{\mathbb{P}(C_n)}{\mathbb{P}(B)}
\end{aligned}$$

Conclusion:

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

Special case:

- If all $\mathbb{P}(C_n)$ are equal, then all $\mathbb{P}(C_n \mid B)$ are equal too.
- If $\mathbb{P}(A \mid C_n)$s all equal, then $\mathbb{P}(A \mid B) = \mathbb{P}(A \mid C_n)$ also.

Example. Uniform permutation $(\sigma(1), \dots, \sigma(52)) \in \sum_{52}$ ("well-shuffled cards"). $\{1,2,3,4\}$ are *aces*. What is $\mathbb{P}(\{\sigma(1), \sigma(2) \text{ both aces}\})$?

$$A = \{\sigma(1), \sigma(2) \text{ aces}\}, \qquad B = \{\sigma(1) \text{ is ace}\} = \{\sigma(1) \leq 4\}$$

$$C_1 = \{\sigma(1) = 1\}, \dots, C_4 = \{\sigma(1) = 4\}$$

Note.

- $$\begin{aligned} \mathbb{P}(A \mid C_i) &= \mathbb{P}(\sigma(2) \in \{1,2,3,4\} \mid \sigma(1) = i) \qquad i \leq 4 \\ &= \frac{3}{51} \end{aligned}$$

- $\mathbb{P}(C_1) = \cdots = \mathbb{P}(C_4) = \frac{1}{52}$

So conclude:

$$\mathbb{P}(A \mid B) = \frac{3}{51}$$

$$\mathbb{P}(A) = \mathbb{P}(B) \times \mathbb{P}(A \mid B) = \frac{4}{52} \times \frac{3}{51}$$

2 Discrete Random Variables

Motivation: Roll two dice.

$$\Omega = \{1, \dots, 6\}^2 = \{(i,j) : 1 \le i,j \le 6\}$$

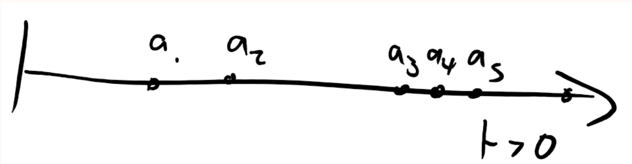
Restrict attention to first dice, for example $\{(i,j) : i = 3\}$, or sum of dice values for example $\{(i,j) : i + j = 8\}$, or max of dice, for example $\{(i,j) : i,j \le 4, i \text{ or } j = 4\}$.
Goal: "Random real-valued measurements".

Definition. A *discrete random variable* X on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a function $X : \Omega \to \mathbb{R}$ such that

- $\{\omega \in \Omega : X(\omega) = x\} \in \mathcal{F}$
- $\mathrm{Im}(X)$ is finite or countable (subset of $\mathbb{R}$)

If Ω finite or countable and $\mathcal{F} = \mathcal{P}(\Omega)$ then both bullet points hold automatically.

Example (Part II Applied Probability)**.**



$$\Omega = \{\text{countable subsets } (a_1, a_2, \dots) \text{ of } (0, \infty)\}$$

$$\begin{aligned} N_t &= \text{number of arrivals by time } t \\ &= |\{a_i : a_i \le t\}| \in \{0, 1, 2, \dots\} \end{aligned}$$

is a discrete random variable for each time t.

Definition. The *probability mass function* of discrete random variable X is the function $p_X : \mathbb{R} \to [0,1]$ given by

$$p_X(x) = \mathbb{P}(X = x)\ \forall\ x \in \mathbb{R}$$

Note.

- if $x \notin \operatorname{Im}(X)$ then
$$p_X(x) = \mathbb{P}(\{\omega \in \Omega : X(\omega) = x\}) = \mathbb{P}(\emptyset) = 0$$
- $$\begin{aligned}\sum_{x \in \operatorname{Im}(X)} P_X(x) &= \sum_{x \in \operatorname{Im}(X)} \mathbb{P}(\{\omega \in \Omega : X(\omega) = x\}) \\ &= \mathbb{P}\left(\bigcup_{x \in \operatorname{Im}(x)} \{\omega \in \Omega : X(\omega) = x\}\right) \\ &= \mathbb{P}(\Omega) \\ &= 1\end{aligned}$$

Example. Event $A \in \mathcal{F}$, define $\mathbb{1}_A : \Omega \to \mathbb{R}$ by

$$\mathbb{1}_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A \end{cases}$$

("*Indicator function of* A") $\mathbb{1}_A$ is a discrete random variable with $\operatorname{Im} = \{0, 1\}$.
Probability mass function:

$$\mathbb{P}_{\mathbb{1}_A}(1) = \mathbb{P}(\mathbb{1}_A = 1) = \mathbb{P}(A)$$

$$\mathbb{P}_{\mathbb{1}_A}(0) = \mathbb{P}(\mathbb{1}_A = 0) = 1 - \mathbb{P}(A)$$

$$\mathbb{P}_{\mathbb{1}_A}(x) = 0\ \forall x \notin \{0, 1\}.$$

This encodes "did A happen?" as a real number.

Remark. Given a probability mass function p_X, we can always construct a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a random variable defined on it with this probability mass function.

- $\Omega = \text{Im}(X)$ i.e. $\{x \in \mathbb{R} : p_X(x) > 0\}$.
- $\mathcal{F} = \mathcal{P}(\Omega)$
- $\mathbb{P}(\{x\}) = p_X(x)$ and extend to all $A \in \mathcal{F}$.

Start of lecture 8

Discrete Probability Distributions

Ω finite.

1. Bernoulli Distribution

("(biased) coin toss").
$X \sim \text{Bern}(p)$, $p \in [0, 1]$.

$$\text{Im}(x) = \{0, 1\}$$
$$p_X(1) = \mathbb{P}(X = 1) = p$$
$$p_X(0) = \mathbb{P}(X = 0) = 1 - p.$$

Key example: $\mathbb{1}_A \sim \text{Bern}(p)$ with $p = \mathbb{P}(A)$.

2. Binomial Distribution

$X \sim \text{Bin}(n, p)$, $n \in \mathbb{Z}^+$, $p \in [0, 1]$.
('Toss coin n times, count number of heads".)

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

check:

$$\sum_{k=0}^{n} p_X(k) = (p + (1 - p))^n = 1$$

More than one Random Variable

Motivation: Doll a dice. Outcome $X \in \{1, 2, \dots, 6\}$. Events:

$$A = \{1 \text{ or } 2\}, \qquad B = \{1 \text{ or } 2 \text{ or } 3\}, \qquad C = \{1 \text{ or } 3 \text{ or } 5\}.$$
$$\mathbb{1}_A \sim \text{Bern}\left(\frac{1}{3}\right), \qquad \mathbb{1}_B \sim \text{Bern}\left(\frac{1}{2}\right), \qquad \mathbb{1}_C \sim \text{Bern}\left(\frac{1}{2}\right)$$

Note. $\mathbb{1}_A \leq \mathbb{1}_B$ for all outcomes, but $\mathbb{1}_A \leq \mathbb{1}_C$ for outcomes is *false*.

Definition. $X_1, \dots, X_n$ discrete random variables. Say $X_1, \dots, X_n$ are *independent* if

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

(suffices to check $\forall\ x_i \in \mathrm{Im}(X_i)$).

Example. $X_1, \dots, X_n$ independent random variables each with the Bernoulli(p) distribution. Study $S_n = X_1 + \cdots + X_n$. Then

$$\begin{aligned} \mathbb{P}(S_n = k) &= \sum_{\substack{X_1 + \cdots X_n = k \\ X_i \in \{0,1\}}} \mathbb{P}(X_1 = x_1, \dots, X_n = x_n) \\ &= \sum_{X_1 + \cdots + X_n = k} \mathbb{P}(X_1 = x_1) \cdots \mathbb{P}(X_n = x_n) \\ &= \sum_{X_1 + \cdots + X_n = k} p^{|\{i : x_i = 1\}|} (1-p)^{|\{i : x_i = 0\}|} \\ &= \sum_{X_1 + \cdots + X_n = k} p^k (1-p)^{n-k} \\ &= \binom{n}{k} p^k (1-p)^{n-k} \end{aligned}$$

so $S_n \sim \mathrm{Bin}(n, k)$.

Example (Non-example)**.** $(\sigma(1), \sigma(2), \dots, \sigma(n))$ uniform in $\sum_n$.

Claim. $\sigma(1)$ and $\sigma(2)$ are *not* independent.

Suffices to find i_1, i_2 such that

$$\mathbb{P}(\sigma(1) = i, \sigma(2) = i_2) \neq \mathbb{P}(\sigma(1) = i_1)\mathbb{P}(\sigma(2) = i_2)$$

for example

$$\mathbb{P}(\sigma(1) = 1, \sigma(2) = 1) = 0 \neq \frac{1}{n} \times \frac{1}{n} = \mathbb{P}(\sigma(1) = 1)\mathbb{P}(\sigma(2) = 1)$$

Consequence of definition

$X_1, \dots, X_n$ independent then $\forall\, A_1, \dots, A_n \subset \mathbb{R}$ countable, then

$$\mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n) = \mathbb{P}(X_1 \in A_1) \cdots \mathbb{P}(X_n \in A_n)$$

$\Omega = \mathbb{N}$

"Ways of choosing a random integer"

3. Geometric distribution

("waiting for success")
$X \sim \mathrm{Geom}(p)$, $p \in (0, 1]$.
("Toss a coin with $\mathbb{P}(\text{heads}) = p$ until a head appears. Count how many trials were needed.")

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

$$p_X(k) = \mathbb{P}((k-1) \text{ failures, then success on } k\text{-th}) = (1-p)^{k-1}p$$

Check:

$$\sum_{k \ge 1} (1-p)^{k-1} p = p \sum_{l \ge 0} (1-p)^l = \frac{p}{1 - (1-p)} = 1$$

> **Note.** We could alternatively "count how many failures before a success".
>
> $$\mathrm{Im}(Y) = \{0, 1, 2, \dots\}$$
>
> $$p_Y(k) = \mathbb{P}(k \text{ failures, then success on } (k+1)\text{-th}) = (1-p)^k p$$
>
> Check:
>
> $$\sum_{k \ge 0} (1-p)^k p = 1$$

4. Poisson Distribution

$\lambda \in (0, \infty)$.

$$X \sim \mathrm{Po}(\lambda)$$

$$\mathrm{Im}(X) = \{0, 1, 2, \dots\}$$

$$\mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!} \qquad \forall\, k \ge 0$$

Note.

$$\sum_{k\geq 0}\mathbb{P}(X=k)=e^{-k}\sum_{k\geq 0}\frac{\lambda^k}{k!}=e^{-\lambda}e^{\lambda}=1$$

Motivation: Consider $X_n \sim \text{Bin}\left(n, \frac{\lambda}{n}\right)$. .image

- Probability of an arrival in each interval is p, independently across intervals.
- Total arrivals is X_n.

$$\mathbb{P}(X_n=k)=\binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k}$$

Fix k, let $n\to\infty$:

$$\mathbb{P}(X_n=k)=\underbrace{\frac{n!}{n^k(n-k)!}}_{\to 1}\times\frac{\lambda^k}{k!}\times\underbrace{\left(1-\frac{\lambda}{n}\right)^n}_{\to e^{-\lambda}}\times\underbrace{\left(1-\frac{1}{n}\right)^{-k}}_{\to 1}$$

so

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

Start of lecture 9

"Bin $\left(n, \frac{\lambda}{n}\right)$ converges to Po(λ)". (note the "converges" is not very meaningful).

Expectation

$(\Omega, \mathcal{F}, \mathbb{P})$ and X a discrete random variable. For now: X only takes non-negative values. "$X \geq 0$"

Definition. The *expectation of* X (or *expected value* of *mean*) is

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

"average of values taken by X, weighted by p_X".

Example 1. X uniform on $\{1,2,\dots,6\}$ (i.e. dice) then

$$\mathbb{E}[X]=\frac{1}{6}\times 1+\frac{1}{6}\times 2+\cdots+\frac{1}{6}\times 6=3.5$$

Note. $\mathbb{E}[X]\notin\text{Im}(X)$.

Example 2. $X \sim \text{Binomial}(n, p)$.

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

Trick:

$$\begin{aligned}
k\binom{n}{k} &= \frac{k \times n!}{k! \times (n-k)!} \\
&= \frac{n!}{(k-1)!(n-k)!} \\
&= \frac{n \times (n-1)!}{(k-1)! \times (n-k)!} \\
&= n\binom{n-1}{k-1} \\
\mathbb{E}[X] &= n \sum_{k=1}^{n} \binom{n-1}{k-1} p^k (1-p)^{n-k} \\
&= np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} (1-p)^{(n-1)-(k-1)} \\
&= np \sum_{l=0}^{n-1} \binom{n-1}{l} p^{l} (1-p)^{(n-1)-l} \\
&= np(p + (1-p))^{n-1} \\
&= np
\end{aligned}$$

Note. Would like to say:

$$\mathbb{E}[\text{Bin}(n,p)] = \mathbb{E}[\text{Bern}(p)] + \cdots + \mathbb{E}[\text{Bern}(p)]$$

Example 3. $X \sim \text{Poisson}(\lambda)$.

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{k\geq 0} k\mathbb{P}(X = k) \\
&= \sum_{k\geq 0} k \cdot e^{-\lambda}\frac{\lambda^k}{k!} \\
&= \sum_{k\geq 1} e^{-\lambda}\frac{\lambda^k}{(k-1)!} \\
&= \lambda\sum_{k\geq 0} e^{-\lambda}\frac{\lambda^{k-1}}{(k-1)!} \\
&= \lambda\sum_{l\geq 0} e^{-\lambda}\frac{\lambda^l}{l!} \\
&= \lambda
\end{aligned}$$

Note. Would like to say

$$\mathbb{E}[\text{Poisson}(\lambda)] \approx \mathbb{E}\left[\text{Bin}\left(n, \frac{\lambda}{n}\right)\right] = \lambda$$

Can't say this: not true in general that

$$\mathbb{P}(X_n = k) \approx \mathbb{P}(\lambda = k) \implies \mathbb{E}[X_n] \approx \mathbb{E}[X]$$

Example 4. $X \sim \text{Geometric}(p)$. Exercise.

Positive and negative: General X (not necessarily $X \geq 0$).

$$\mathbb{E}[X] = \sum_{x\in \text{Im}(X)} x\mathbb{P}(X = x)$$

unless

$$\sum_{\substack{x>0 \\ x\in\text{Im}(x)}} x\mathbb{P}(X = x) = +\infty$$

and

$$\sum_{\substack{x<0 \\ x\in\text{Im}(x)}} x\mathbb{P}(X = x) = -\infty$$

then we say that $\mathbb{E}[X]$ is not defined.
<u>Summary</u>:

- both infinite: not defined
- first infinite, second not: $\mathbb{E}[X] = +\infty$
- second infinite, first not: $\mathbb{E}[X] = -\infty$
- neither infinite: X is *integrable*, i.e.

$$\sum_{x \in \mathrm{Im}(X)} |x| \mathbb{P}(X = x)$$

  converges.

Note that some people say that in cases 2 and 3, the expectation is undefined.

Example 5. Most examples in the course are integrable *except*:

- $\mathbb{P}(X = n) = \frac{6}{\pi^2} \times \frac{1}{n^2}$ for $n \geq 1$. (Note $\sum \mathbb{P}(X = n) = 1$). Then

$$\mathbb{E}[X] = \sum \frac{6}{\pi^2} \times \frac{1}{n} = +\infty$$

- $\mathbb{P}(X = n) = \frac{3}{\pi^2} \times \frac{1}{n^2}$ for $n \in \mathbb{Z} \setminus \{0\}$, then $\mathbb{E}[X]$ is not defined. ("It's symmetric so $\mathbb{E}[X] = 0$" is considered wrong for us).

Example. $\mathbb{E}[\mathbb{1}_A = \mathbb{P}(A)$ Important!

Properties of Expectation

(X discrete).

(1) If $X \geq 0$, then $\mathbb{E}[X] \geq 0$ with equality if and only $\mathbb{P}(X = 0) = 1$. Why?

$$\mathbb{E}[X] = \sum_{\substack{x \in \mathrm{Im}(X) \\ x \neq 0}} x \mathbb{P}(X = x)$$

(2) If $\lambda, c \in \mathbb{R}$ then:

   (i) $\mathbb{E}[X + c] = \mathbb{E}[X] + c$

   (ii) $\mathbb{E}[\lambda X] = \lambda \mathbb{E}[X]$

(3) (i) X, Y random variables (both integrable) on same probability space.

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

(ii) In fact $\lambda, \mu \in \mathbb{R}$

$$\mathbb{E}[\lambda X + \mu Y] = \lambda \mathbb{E}[X] + \mu \mathbb{E}[Y]$$

similarly:

$$\mathbb{E}[\lambda_1 X_1 + \cdots + \lambda_n X_n] = \lambda_1 \mathbb{E}[X_1] + \cdots + \lambda_n \mathbb{E}[X_n]$$

Proof of (3)(ii).

$$\begin{aligned}\mathbb{E}[\lambda X + \mu Y] &= \sum_{\omega \in \Omega} (\lambda X(\omega) + \mu Y(\omega)) \mathbb{P}(\{\omega\}) \\ &= \lambda \sum_{\omega \in \Omega} X(\omega) \mathbb{P}(\{\omega\}) + \mu \sum_{\omega \in \Omega} Y(\omega) \mathbb{P}(\{\omega\}) \\ &= \lambda \mathbb{E}[X] + \mu \mathbb{E}[Y]\end{aligned}$$

Note that this proof only works for countable Ω, but there is also a proof for general Ω. □

Note. Independence is *not* required for linearity of expectation to hold. (This is the name for property (3)(ii)).

Start of lecture 10

Corollary. $X \geq Y$ (meaning $X(\omega) \geq Y(\omega)$ for all $\omega \in \mathbb{R}$) then $\mathbb{E}[X] \geq \mathbb{E}[Y]$.

Proof. $X = (X - Y) + Y$ hence

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

but $X - Y \geq 0$ hence $\mathbb{E}[X - Y] \geq 0$. □

Key Application: Counting problems.
$(\sigma(1), \ldots, \sigma(n))$ uniform on σ_n.

$$Z = |\{i : \sigma(i) = i\}| = \text{number of fixed points}$$

Let $A_i = \{\sigma(i) = i\}$. (Recall A_is are *not* independent)
Key step:

$$Z = \mathbb{1}_{A_1} + \cdots + \mathbb{1}_{A_n}$$

so

$$\begin{aligned}\mathbb{E}[Z] &= \mathbb{E}[\mathbb{1}_{A_1} + \cdots + \mathbb{1}_{A_n}] \\ &= \mathbb{E}[\mathbb{1}_{A_1}] + \cdots + \mathbb{E}[\mathbb{1}_{A_n}] \\ &= \mathbb{P}(A_1) + \cdots \mathbb{P}(A_n) \\ &= \frac{1}{n} \times n \\ &= 1\end{aligned}$$

Note. Same answer as Bin $\left(n, \frac{1}{n}\right)$.

Application: X takes values in $\{0, 1, 2, \dots\}$.
Fact: $\mathbb{E}[X] = \sum_{k \geq 1} \mathbb{P}(X \geq k)$. *Proof 1.* Write

$$X = \sum_{k \geq 1} \mathbb{1}_{(X \geq k)}$$

Then

$$\begin{aligned} \mathbb{E}[X] &= \mathbb{E}\left[\sum \mathbb{1}_{(X \geq k)}\right] \\ &= \sum \mathbb{E}[\mathbb{1}_{(X \geq k)}] \\ &= \sum \mathbb{P}(X \geq k) \end{aligned}$$

□

Sanity Check: for example if $X = 7$ then

$$\mathbb{1}_{(X \geq 1)} = \cdots = \mathbb{1}_{(X \geq 7)} = 1$$

$$\mathbb{1}_{(X \geq 8)} = \mathbb{1}_{(X \geq 9)} = \cdots = 0$$

Markov's Inequality

$X \geq 0$ a random variable. Then $\forall\ a > 0$:

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

Comment:
Is $a = \frac{\mathbb{E}[X]}{2}$ useful? Definitely not.
Is a is large useful? Maybe.
Proof. Observe: $X \geq a\mathbb{1}_{(X \geq a)}$. Then

$$\mathbb{E}[X] \geq a\mathbb{E}[\mathbb{1}_{X \geq a}] = a\mathbb{P}(X \geq a)$$

now just rearrange. □

Note that $\mathbb{1}_{(X \geq a)}$ means $X(\omega) \geq a\mathbb{1}_{(X \geq a)}(\omega)$.
Check: if $X \in [0, a)$ then RHS $= 0$, if $X \in [a, \infty)$ then RHS $= a$.

Note. Also true for continuous random variables (later).

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

Let $f : \mathbb{R} \to \mathbb{R}$ be a function. Then $f(X)$ is also a *random variable.*

Claim. $\mathbb{E}[f(X)] = \sum_{x \in \text{Im}(X)} f(x)\mathbb{P}(X = x)$.

Proof. Let

$$A = \text{Im}(f(X)) = \{y : y = f(x), x \in \text{Im}(X)\} = \{f(x) : x \in \text{Im}(X)\}$$

Start with RHS:

$$\sum_{x \in \text{Im}(X)} f(x)\mathbb{P}(X = x) = \sum_{y \in A} \sum_{\substack{x \in \text{Im}(X) \\ f(x) = y}} f(x)\mathbb{P}(X = x)$$

$$\sum_{y \in A} y \sum_{\substack{x \in \text{Im}(X) \\ f(x) = y}} \mathbb{P}(X = x)$$

$$= \sum_{y \in A} y\mathbb{P}(f(X) = y)$$

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

□

Motivation

$$U_n \sim \text{Uniform}(\{-n, -n+1, \dots, n\})$$

$$V_n \sim \text{Univorm}(\{-n, +n\})$$

$$Z_n = 0$$

$$S_n = \text{random walk for } n \text{ steps} \sim n - 2\text{Bin}\left(n, \frac{1}{2}\right)$$

All of these have $\mathbb{E} = 0$.

Variance

"Measure how concentrated a random variable is around its mean".

Definition. The *variance* of X is:

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

Property:

$$\text{Var}(X) \geq 0$$

with equality $\iff \mathbb{P}(X = \mathbb{E}[X]) = 1$.

<u>Alternative Characterisation</u>:

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

Proof. Write $\mu = \mathbb{E}[X]$. Then

$$\begin{aligned}
\text{Var}(X) &= \mathbb{E}[(X-\mu)^2] \\
&= \mathbb{E}[X^2 - 2\mu X + \mu^2] \\
&= \mathbb{E}[X^2] - 2\mu \underbrace{\mathbb{E}[X]}_{\mu} + \mu^2 \\
&= \mathbb{E}[X^2] - \mu^2
\end{aligned}$$

□

Properties

If $\lambda, c \in \mathbb{R}$:

- $\text{Var}(\lambda X) = \lambda^2 \text{Var}(X)$
- $\text{Var}(X + c) = \text{Var}(X)$.

Proof. $\mathbb{E}[X + c] = \mu + c$

$$\begin{aligned}
\text{Var}(X+c) &= \mathbb{E}[(X + c - (\mu + c))^2] \\
&= \mathbb{E}[(X-\mu)^2] \\
&= \text{Var}(X)
\end{aligned}$$

□

Start of lecture 11

Example 1. $X \sim \text{Poisson}(\lambda)$, $\mathbb{E}[X] = \lambda$.

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

"Falling factorial trick": sometimes $\mathbb{E}[X(X-1)]$ is easier than $\mathbb{E}[X^2]$. Here:

$$\begin{aligned}
\mathbb{E}[X(X-1)] &= \sum_{k\ge 2} k(k-1)e^{-\lambda}\frac{\lambda^k}{k!} \\
&= \lambda^2 e^{-\lambda}\sum_{k\ge 2}\frac{\lambda^{k-2}}{(k-2)!} \\
&= \lambda^2 \\
\mathbb{E}[X^2] = \mathbb{E}[X(X-1)+X] \\
&= \mathbb{E}[X(X-1)] + \mathbb{E}[X] \\
&= \lambda^2 + \lambda \\
\implies \text{Var}(x) = \lambda
\end{aligned}$$

Example 2. $Y \sim \text{Geom}(p) \in \{1, 2, 3, \dots\}$. $\mathbb{E}[Y] = \frac{1}{p}$. $\text{Var}(y) = \dots = \frac{1-p}{p^2}$. (left as an exercise)

Note. λ large: $\text{Var}(X) = \mathbb{E}[X]$.
p small (so Y large): $\text{Var}(Y) \approx \frac{1}{p^2} = (\mathbb{E}[X])^2$.

Example 3. $X \sim \text{Bern}(p)$. $\mathbb{E}[X] = 1 \times p = p$. $\mathbb{E}[X^2] = 1^2 \times p = p$.

$$\text{Var}(X) = p - p^2 = p(1-p)$$

Example 4. $X \sim \text{Bin}(n, p)$, $\mathbb{E}[X] = np$.

$$\mathbb{E}[X^2] = \text{ugly}\dots$$

Goal: Study $\text{Var}(X_1 + \dots + X_n)$ for not independent.

Preliminary: $\mathbb{E}$[Products of RVs]. Setting: X, Y *independent* random variables and f,

f functions $\mathbb{R} \to \mathbb{R}$. Then:

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

"splits as a product"
Key example 1: $f, g : f(x) = g(x) = x$. Then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.
Key example 2: $f(x) = g(x) = z^x$ (or e^{tx}).
Proof.

$$\begin{aligned}
LHS &= \sum_{x,y \in \mathrm{Im}} f(x)g(y)\mathbb{P}(X = x, Y = y) \\
&= \sum_{x,y \in \mathrm{Im}} f(x)g(y)\mathbb{P}(X = x)\mathbb{P}(Y = y) \\
&= \left[\sum_{x \in \mathrm{Im}\ X} f(x)\mathbb{P}(X = x)\right]\left[\sum_{y \in \mathrm{Im}\ Y} g(y)\mathbb{P}(Y = y)\right] \\
&= \mathbb{E}[f(X)]\mathbb{E}[g(Y)]
\end{aligned}$$

□

Sums of Independent Random Variables

$X_1, \dots, X_n$ independent. Then

$$\mathrm{Var}(X_1 + \cdots X_n) = \mathrm{Var}(X_1) + \cdots \mathrm{Var}(X_n)$$

Proof. (Suffices to prove $n = 2$ by induction). Say $\mathbb{E}[X] = \mu$, $\mathbb{E}[Y] = \nu$. Then $\mathbb{E}[X+Y] = \mu + \nu$.

$$\begin{aligned}
\mathrm{Var}(X + Y) &= \mathbb{E}[(X + Y - \mu - \nu)^2] \\
&= \mathbb{E}[(X - \mu)^2] + \mathbb{E}[(Y - \mu)^2] + 2\mathbb{E}[(X - \mu)(Y - \nu)] \\
&= \mathrm{Var}(X) + \mathrm{Var}(Y) + \mathbb{E}[X - \mu]\mathbb{E}[Y - \nu]
\end{aligned}$$

$$\mathrm{Var}(X) + \mathrm{Var}(Y)$$

□

Example 4. $\mathrm{Var}(\mathrm{Bin}(n, p)) = np(1 - p)$.

Goal: Study $\mathrm{Var}(X + Y)$ when X, Y are not independent.

Definition. X, Y two random variables. Their *covariance* is

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

"Measures how dependent X, Y are, and in which *direction*": If $\mathrm{Cov} > 0$ then X bigger means Y bigger, and if $\mathrm{Cov} < 0$ then X bigger means Y smaller.

Properties

- $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
- $\text{Cov}(X, X) = \text{Var}(X)$.
- Alternative characterisation:

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

  (often more useful, and particularly nice if $\mathbb{E}[X] = 0$) *Proof.*

  $$\text{Cov}(X, Y) = \mathbb{E}[(X - \mu)(Y - \nu)]$$

  $$\mathbb{E}[XY] - \mu \underbrace{\mathbb{E}[Y]}_{\nu} - \nu \underbrace{\mathbb{E}[X]}_{\mu} + \mu\nu$$

  $$= \mathbb{E}[XY] - \mu\nu$$

  □

- $c, \lambda \in \mathbb{R}$:
  - $\text{Cov}(c, X) = 0$
  - $\text{Cov}(X + c, Y) = \text{Cov}(X, Y)$
  - $\text{Cov}(\lambda X, \lambda Y) = \lambda^2 \text{Cov}(X, Y)$
- $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$
- Covariance is *linear* in each argument, i.e.

  $$\text{Cov}\left(\sum \lambda_i X_i, Y\right) = \sum \lambda_i \text{Cov}(X_i, Y)$$

  and (applying in two stages)

  $$\text{Cov}\left(\sum \lambda_i X_i, \sum \mu_j Y_j\right) = \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \mu_j \text{Cov}(X_i, Y_j)$$

  "Special case":

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

  (for an example, see Q11 on sheet 3)

Note. We have already seen that X, Y independent implies $\mathrm{Cov}(X,Y) = 0$, but it is not the case the zero covariance implies independence.

Start of lecture 12

Example 0. $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$ for independent X, Y. Consider $Y = -X$. Then

$$\mathrm{Var}(Y) = \mathrm{Var}(-X) = (-1)^2\mathrm{Var}(x) = \mathrm{Var}(X)$$

$$0 = \mathrm{Var}(0) = \mathrm{Var}(X+Y) \neq \mathrm{Var}(X) + \mathrm{Var}(Y) = 2\mathrm{Var}(X)$$

Example 1. $(\sigma(1), \dots, \sigma(n))$ uniform on $\sum_n$. $A_i = \{\sigma(i) = i\}$.

$$N = \mathbb{1}_{A_1} + \cdots + \mathbb{1}_{A_n} = \text{number of fixed points}$$

Already seen: $\mathbb{E}[N] = n \times \frac{1}{n} = 1$. Goal: $\mathrm{Var}(N)$.

Note. A_i and A_j are *not* independent.

$$\begin{aligned}
\mathrm{Var}(\mathbb{1}_{A_i}) &= \frac{1}{n}\left(1 - \frac{1}{n}\right) \\
\mathrm{Cov}(\mathbb{1}_{A_i}, \mathbb{1}_{A_j}) &= \mathbb{E}[\mathbb{1}_{A_i}\mathbb{1}_{A_j}] - \mathbb{E}[\mathbb{1}_{A_i}]\mathbb{E}[\mathbb{1}_{A_j}] \\
&= \mathbb{E}[\mathbb{1}_{A_i \cap A_j}] - \mathbb{E}[\mathbb{1}_{A_i}]\mathbb{E}[\mathbb{1}_{A_j}] \\
&= \mathbb{P}(A_i \cap A_j) - \mathbb{P}(A_i)\mathbb{P}(A_j) \\
&= \frac{1}{n(n-1)} - \frac{1}{n} \times \frac{1}{n} \\
&= \frac{1}{n^2(n-1)} \\
&> 0 \\
\implies \mathrm{Var}(N) &= \sum_{i=1}^n \mathrm{Var}(\mathbb{1}_{A_i}) + \sum_{i \neq j} \mathrm{Cov}(\mathbb{1}_{A_i}, \mathbb{1}_{A_j}) \\
&= n \times \frac{1}{n}\left(1 - \frac{1}{n}\right) + n(n-1) \times \frac{1}{n^2(n-1)} \\
&= 1 - \frac{1}{n} + \frac{1}{n} \\
&= 1
\end{aligned}$$

Compare with Bin $\left(n, \frac{1}{n}\right)$:

$$\mathbb{E} = 1, \qquad \mathrm{Var} = n \times \frac{1}{n}\left(1 - \frac{1}{n}\right) = 1 - \frac{1}{n}$$

Chebyshev's Inequality

Theorem (Chebyshev's Inequality)**.** X a random variable, $\mathbb{E}[X] = \mu$, $\mathrm{Var}(X) = \sigma^2 < \infty$. Then:

$$\mathbb{P}(|X - \mu| \geq \lambda) \leq \frac{\mathrm{Var}(X)}{\lambda^2}$$

Comment: Remember the proof, not the statement!

Proof. Idea: Apply Markov's Inequality to

$$(X - \mu)^2$$

(which is non-negative as required). Then:

$$\begin{aligned}\mathbb{P}(|X - \mu| \geq \lambda) &= \mathbb{P}((X - \mu)^2 \geq \lambda^2) \\ &\leq \frac{\mathbb{E}[(X - \mu)^2}{\lambda^2} \\ &= \frac{\mathrm{Var}(X)}{\lambda^2}\end{aligned}$$

□

Comments

- Chebyshev's Inequality gives better bounds than Markov's inequality.
- Note can apply to all Random Variables, not just ≥ 0.
- However, $\mathrm{Var}(X) < \infty$ is a stronger condition than $\mathbb{E}[X] < \infty$.

Definition.
- Quantity $\sqrt{\mathrm{Var}(X)} = \sigma$ is called the *standard deviation* of X.
- Same "units" as X. (Scales linearly)
- (Not many nice properties).
- Rewriting Chebyshev; use $\lambda = k\sqrt{\sigma^2}$, then

$$\mathbb{P}(|X - \mu| \geq \sigma) \leq \frac{1}{k^2}$$

- Nice uniform statement

Conditional Expectation

Setting: $(\Omega, \mathcal{F}, \mathbb{P})$.
Recall: $B \in \mathcal{F}$ with $\mathbb{P}(B) > 0$ we defined

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

Definition. $B \in \mathcal{F}$ with $\mathbb{P}(B) > 0$, X a random variable. The conditional expectation is

$$\mathbb{E}[X \mid B] = \frac{\mathbb{E}[X \mathbb{1}_B]}{\mathbb{P}(B)}$$

Example. X dice, uniform on $\{1, \dots, 6\}$.

$$\begin{aligned}\mathbb{E}[X \mid X \text{ prime}] &= \frac{\frac{1}{6}[0+2+3+0+5+0]}{\frac{1}{2}} \\ &= \frac{1}{3}(2+3+5 \\ &= \frac{10}{3}\end{aligned}$$

Alternative Characterisation:

$$\mathbb{E}[X \mid B] = \sum_{x \in \text{Im } X} \mathbb{P}(X = x \mid B)$$

Proof.

$$\begin{aligned}RHS &= \sum \frac{x\mathbb{P}(\{X = x\} \cap B)}{\mathbb{P}(B)} \\ &= \sum_{\substack{x \neq 0 \\ x \in \text{Im } X}} \frac{x\mathbb{P}(X \mathbb{1}_B = x}{\mathbb{P}(B)}\end{aligned}$$

and note

$$\mathbb{E}[X \mathbb{1}_B] = \sum_{\substack{x \neq 0 \\ x \in \text{Im } X}} x\mathbb{P}(X \mathbb{1}_B = x)$$

□

Law of Total Expectation

$(B_1, B_2, \dots)$ a finite or countably infinite partition of Ω with $B_n \in \mathcal{F}$ for all n such that $\mathbb{P}(B_n) > 0$. X is a random variable. Then:

$$\mathbb{E}[X] = \sum_n \mathbb{E}[X \mid B_n]\mathbb{P}(B_n)$$

For example, $X = \mathbb{1}_A$ recovers the law of total probability.
Proof.

$$\begin{aligned} RHS &= \sum_n \mathbb{E}[X \mathbb{1}_{B_n}] \\ &= \mathbb{E}[X \cdot (\mathbb{1}_{B_1} + \cdots + \mathbb{1}_{B_n})] \\ &= \mathbb{E}[X \cdot 1] \\ &= \mathbb{E}[X] \end{aligned}$$

□

Application: Two stage randomness where (B_n) describes what happens in stage 1.
Application 1: "random sums" (random number of terms).
$(X_n)_{n\geq 1}$ independent and identically distributed random variables. $N \in \{0, 1, 2, \dots\}$ random index independent of (X_n).

$$S_n = X_1 + \cdots + X_n$$

with $\mathbb{E}[X_n] = \mu$ so $\mathbb{E}[S_n] = n\mu$. Then

$$\begin{aligned} \mathbb{E}[S_N] &= \sum_{n\geq 0} \mathbb{E}[S_N \mid N = n]\mathbb{P}(N = n) \\ &= \sum \mathbb{E}[S_n]\mathbb{P}(N = n) \\ &= \sum_{n\geq 0} n\mu\mathbb{P}(N = n) \\ &= \mu\mathbb{E}[N] \end{aligned}$$

Start of lecture 13

Random Walks

Setting: $(X_n)_{n\geq 1}$ independent and identically distributed random variables

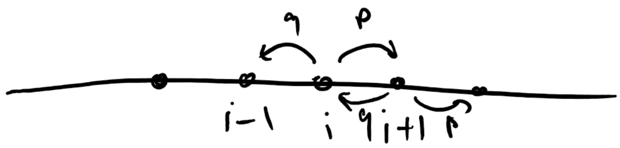
$$S_n = x_0 + X_1 + \cdots + X_n$$

$(S_0, S_1, S_2, \dots)$ is a random process called *Random Walk* started from x_0.

Main example in our course:
Simple Random Walk (SRW) on $\mathbb{Z}$.

$$\mathbb{P}(X_i = +1) = p \qquad \mathbb{P}(X_i = -1) = q = 1 - p$$

$x_0 \in \mathbb{Z}$ (often $x_0 = 0$).
Special case: $p = q = \frac{1}{2}$. ("symmetric"):



For example, $\mathbb{P}(S_2 = x_0) = pq + qp = 2pq$.

Useful interpretation: A gambler repeatedly plays a game where he wins £1 with $\mathbb{P} = p$ and losses £1 with $\mathbb{P} = q$.
Often we stop if we ever reach £0.

Question: Suppose we start with £x at time 0. What is the probability he reaches £a before £0?

Notation.

$$\mathbb{P}_X(\bullet) \text{``}=\text{''} \mathbb{P}(\bullet \mid x_0 = x)$$

"measure of RW started from x_0".

Key Idea: Conditional on $S_1 = z$, $(S_1, S_2, \dots)$ is a random walk started from z.
Now we apply the Law of Total Probability:

$$\begin{aligned}\mathbb{P}_X(S \text{ hits } a \text{ before } 0 &= \sum \mathbb{P}_X(S \text{ hits } a \text{ before } 0 \mid S_1 = z)\mathbb{P}_X(S_1 = z) \\ &= \sum_z \mathbb{P}_Z(S \text{ hits } a \text{ before } 0)\mathbb{P}_Z(S_1 = z)\end{aligned}$$

so $h_X = \mathbb{P}_X(S \text{ hit } a \text{ before } 0)$. $S_1 = x \pm 1$.

$$h_X = px_{x+1} + qh_{x-1}$$

Important to specify boundary conditions:

$$h_0 = 0, \qquad h_a = 1.$$

Now we apply law of total expected value. Expected absorption time:

$$T = \min\{n \geq 0 : S_n = 0 \text{ or } S_n = a\}$$

"first time S hits $\{0, a\}$". Want: $\mathbb{E}_x[T] = \tau_x$.

$$\begin{aligned}\tau_x = \mathbb{E}_x[T] &= p\mathbb{E}_x[T \mid S_1 = x+1] + q\mathbb{E}_x[T \mid S_1 = x-1] \\ &= p\mathbb{E}_{x+1}[T+1] + q\mathbb{E}_{x-1}[T+1] \\ &= p(1 + \mathbb{E}_{x+1}[T]) + q(1 + \mathbb{E}_{x-1}[T]) \\ &= 1 + p\tau_{x+1} + q\tau_{x-1}\end{aligned}$$

Boundary conditions:

$$\tau_0 = \tau_a = 0$$

"we're already there"

Solving Linear Recurrence Equations

Homogeneous case (boundary conditions: h_0, h_a):

$$ph_{x+1} - h_x + qh_{x-1} = 0$$

- Analagous to DEs
- Solutions form a vector space.

Plan: (homogeneous case):

- Find two solutions (linearly independent)

Guess $h_x = \lambda^x$, so

$$p\lambda^{x+1} - \lambda^x + q\lambda^{x-1} = 0$$

$$p\lambda^2 - \lambda + q = 0$$

Quadratic in $\lambda \implies \lambda = 1$ or $\frac{p}{q}$.
Case $q \neq p$: $h_x = A + B\left(\frac{q}{p}\right)^x$.

- Use boundary conditions to find A, B: i.e.

$$x = 0: \qquad h_0 = 0 = A + B$$

$$x = a: \qquad h_a = 1 = A + B\left(\frac{q}{p}\right)^a$$

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

Case $p = q = \frac{1}{2}$: (symmetric random walk)

- Note $h_x = x$ "x is the average of $x+1$ and $x-1$".

General solution: $h_x = A + Bx$. Boundary conditions:

$$h_0 = 0 = A$$

$$h_a = 1 = A + Ba$$

so $A = 0$, $B = \frac{1}{a}$. Hence

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

Probability sanity check: $p = q = \frac{1}{2}$.
Study Expected profit if you start from £x and play until time T.

$$\mathbb{E}_x[S_T] = a\mathbb{P}_x(S_T = a) + 0 \times \mathbb{P}_x(S_T = 0) = a \cdot \frac{x}{a} = x$$

fits intuition for fair games.

Inhomogeneous Case

$$ph_{x+1} - h_x + qh_{x-1} = f(x) = -1$$

Plan:

- Find a *particular solution* Guess: "one level more complicated than general solution".
- Add on general solution
- Solve for boundary conditions

For $p \neq q$: Guess $h_x = \frac{x}{q-p}$ works as a particular solution.
For $p = q = \frac{1}{2}$: Guess $h_x = Cx^2$ *might* work. Sub in:

$$\frac{C}{2}(x+1)^2 - Cx^2 + \frac{C}{2}(x-1)^2 = -1 \implies C = -1$$

So

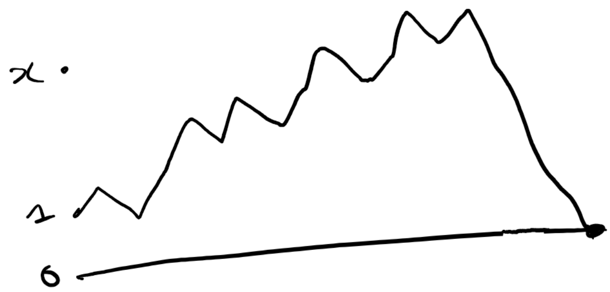
$$h_x = A + Bx - x^2$$

then find A, B with boundary conditions: roots are 0 and a, so

$$h_x = x(a - x)$$

Start of lecture 14

Unbounded Random Walk: "Gambler's Ruin"



$$\mathbb{P}_x(\text{hit } 0) = \lim_{a\to\infty}(\text{hit } 0 \text{ before } a)$$
$$= \begin{cases} 1 - \left(\frac{q}{p}\right)^x & p > q \\ 1 & p < q \\ 1 & p = q = \frac{1}{2} \end{cases}$$

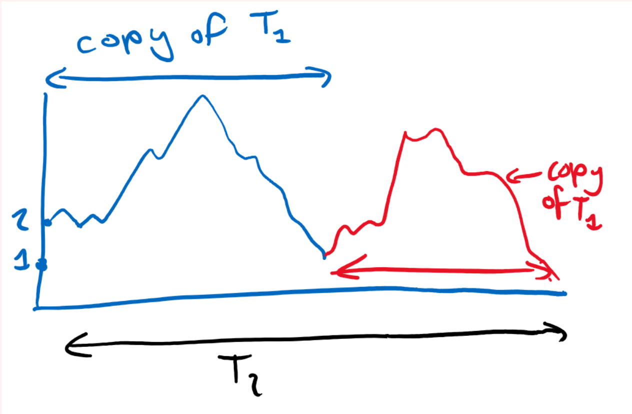
$$p = \frac{1}{2} : \mathbb{E}_x[\text{time to hit } 0] \geq \mathbb{E}_x[\text{time to hit } 0 \text{ or } a] = x(a-x)$$

which $\to \infty$ as $a \to \infty$.

Key conclusion: T_x (time to hit 0 from x) is for $p = \frac{1}{2}$:

- finite with probability $= 1$
- infinite expectation

Note (non-examinable)**.** Alternative derivation of $\mathbb{E}[T_1] = \infty$.



"Random Walk $2 \mapsto 1$" = "Random Walk $1 \mapsto 0$" + 1

$$\mathbb{E}[T_1] = \frac{1}{2} \times 1 + \frac{1}{2}(1 + \underbrace{\mathbb{E}[T_2]}_{2\mathbb{E}[T_1]})$$

$$\mathbb{E}[T_1] = 1 + \mathbb{E}[T_1]$$

so $\mathbb{E}[T_1] = \infty$.

Generating Functions

Setting: X is a random variable taking values in $\{0, 1, 2, \dots\}$.

Definition. The *Probability Generating Function* of X is

$$G_X(z) = \mathbb{E}[z^X] = \sum_{k \geq 0} z^k \mathbb{P}(X = k).$$

Analytic comment: $G_X : (-1, 1) \overset{k \geq 0}{\Rightarrow} \mathbb{R}$.

Idea: "To *encode* the distribution of X as a function with nice analytic properties".

Example 1. $X \sim \text{Bern}(p)$

$$G_X(z) = z^0 \mathbb{P}(X = 0) + z^1 \mathbb{P}(X = 1) = (1 - p) + pz$$

Example. $X \sim \text{Bin}(n, p)$ we will save for later.

Example 2. $X \sim \text{Poisson}(\lambda)$

$$\begin{aligned} G_X(z) &= \sum_{k \ge 0} z^k e^{-\lambda} \frac{\lambda^k}{k!} \\ &= e^{-\lambda} \sum_{k \ge 0} \frac{(\lambda z)^k}{k!} \\ &= e^{-\lambda} e^{\lambda z} \\ &= e^{\lambda(z-1)} \end{aligned}$$

Recovering PMF (mass function) from PGF

Note. $G_X(0) = 0^0 \mathbb{P}(X = 0) = \mathbb{P}(X = 0)$.

Idea: Differentiate n times.

$$\begin{aligned} \frac{\mathrm{d}^n}{\mathrm{d}z^n} G_X(z) &= \sum_{k \ge 0} \frac{\mathrm{d}^n}{\mathrm{d}z^n}(z^k) \mathbb{P}(X = k) \\ &= \sum_{k \ge 0} k(k-1)\cdots(k-n+1) z^{k-n} \mathbb{P}(X = k) \\ &= \sum_{k \ge n} k(k-1)\cdots(k-n+1) z^{k-n} \mathbb{P}(X = k) \\ &= \sum_{l \ge 0} (l+1)(l+2)\cdots(l+n) z^l \mathbb{P}(X = l+n) \end{aligned}$$

Evaluate at 0:

$$\frac{\mathrm{d}^n}{\mathrm{d}z^n} G_X(0) = n! \mathbb{P}(X = n).$$

$$\mathbb{P}(X = n) = \frac{1}{n!} G_X^{(n)}(0)$$

Key fact: PGF *determines* PMF / distribution exactly.

Recovering other probabilistic quantities

Note. $G_X(1) = \sum_{k \ge 0} \mathbb{P}(X = k) = 1$.

Technical comment: $G_X(1)$ means $\lim_{z \to 1} G_X(z)$ if the domain is $(-1, 1)$ (the limit is from below).

- What about $G'_X(1)$?

$$G'_X(z) = \sum_{k \geq 1} k z^{k-1} \mathbb{P}(X = k)$$

$$G'_X(1) = \sum_{k \geq 1} k \mathbb{P}(X = k) = \mathbb{E}[X]$$

- What about $G_X^{(n)}(1)$?

$$\begin{aligned} G_X^{(n)}(1) &= \sum_{k \geq n} k(k-1) \cdots (k-n+1) \mathbb{P}(X = k) \\ &= \mathbb{E}[X(x-1) \cdots (X-n+1)] \end{aligned}$$

- Other expectations:

$$\begin{aligned} \mathbb{E}[X^2] &= \mathbb{E}[X(X-1)] + \mathbb{E}[X] \\ &= G''_X(1) + G'_X(1) \end{aligned}$$

$$\mathrm{Var}(X) = G''_X(1) + G'_X(1) - [G'_X(1)]^2$$

Idea: Find in general $\mathbb{E}[P(X)]$ using $\mathbb{E}$[falling factorials of X.

Note (Linear Algebra Aside)**.** The falling factorials

$$1, X, X(X-1), X(X-1)(X-2)$$

form a *basis* for $\mathbb{R}[X]$ (the set of polynomials with real coefficients).

PGFs for sums of Independent Random Variables

$X_1, \ldots, X_n$ independent random variables.
$G_{X_1}, \ldots, G_{X_n}$ are the PGFs.
Let $X = X_1 + \cdots + X_n$.
Question: What's the PGF of X? (Is it nice)?

$$\begin{aligned} G_X(z) &= \mathbb{E}[Z^X] \\ &= \mathbb{E}[z^{X_1 + \cdots + X_n}] \\ &= \mathbb{E}[z^{X_1} z^{X_2} \cdots z^{X_n}] \\ &= \mathbb{E}[z^{X_1}] \cdots \mathbb{E}[z^{X_n}] \\ &= G_{X_1}(z) \cdots G_{X_n}(z) \end{aligned}$$

Special case: $X_i = X_1 \to G_X(z) = (G_{X_1}(z))^n$.

Note.

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

for independent random variables X, Y.

Start of lecture 15

Note. PGF is much nicer than PMF of X!

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

$$X = X_1 + \cdots X_n$$

(Identical independently distributed Bern(p))

$$G_X(z) = (1 - p + pz)^n$$

Example. $X \sim \text{Poi}(\lambda)$, $Y \sim \text{Poi}(\mu)$ independent.

$$G_X(z) = e^{\lambda(z-1)}, \qquad G_Y(z) = e^{\mu(z-1)}$$

We will study $Z = X + Y$.

$$\begin{aligned} G_{X+Y}(z) &= G_X(y)G_Y(z) \\ &= e^{\lambda(z-1)}e^{\mu(z-1)} \\ &= e^{(\lambda+\mu)(z-1)} \\ &= \text{PGF of Poi}(\lambda + \mu) \end{aligned}$$

So $X + Y \sim \text{Poisson}(\lambda + \mu)$.

PGF for Random Sums

Setting: $X_1, X_2, \dots$ IID with same distribution as X. X takes values in $\{0, 1, 2, \dots\}$ and N is a random value taking values in $\{0, 1, 2, \dots\}$ *independent* of (X_n).

Remark. Perfect pairing with PGFs.

$$\begin{aligned}\mathbb{E}[z^{X_1+\cdots+X_n}] &= \sum_{n\geq 0}\mathbb{E}[z^{X_1+\cdots+X_N} \mid N=n]\mathbb{P}(N=n)\\ &= \sum_{n\geq 0}\mathbb{E}[z^{X_1+\cdots+X_n} \mid N=n]\mathbb{P}(N=n)\\ &= \sum_{n\geq 0}\mathbb{E}[z^{X_1+\cdots+X_n}]\mathbb{P}(N=n)\\ &= \sum_{n\geq 0}\mathbb{E}[z^{X_1}]\cdots\mathbb{E}[z^{X_n}]\mathbb{P}(N=n)\\ &= \sum_{n\geq 0}(G_X(z))^n\mathbb{P}(N=n)\\ &= G_N(G_X(z))\end{aligned}$$

Example. $X_i \sim \text{Bern}(p)$, $N \sim \text{Poisson}(\lambda)$.

$$G_{X_i}(z) = (1-p) + pz$$

$$G_N(s) = e^{\lambda(s-1)}$$

Interpretation: "Poisson thinning", for example "Poi(λ) misprints, each gets found with $\mathbb{P} = 1-p$." (see Q7 on Example sheet)

$$Y = X_1 + \cdots + X_N$$

$$\begin{aligned}G_Y(z) &= G_N(G_{X_i}(z))\\ &= e^{\lambda[1-p+pz-1]}\\ &= e^{\lambda p(z-1)}\\ &= \text{PGF of Poi}(\lambda p)\end{aligned}$$

In general: PMF of $X_1 + \cdots + X_n$ is horrible, $G_N(G_X(z))$ is nice.

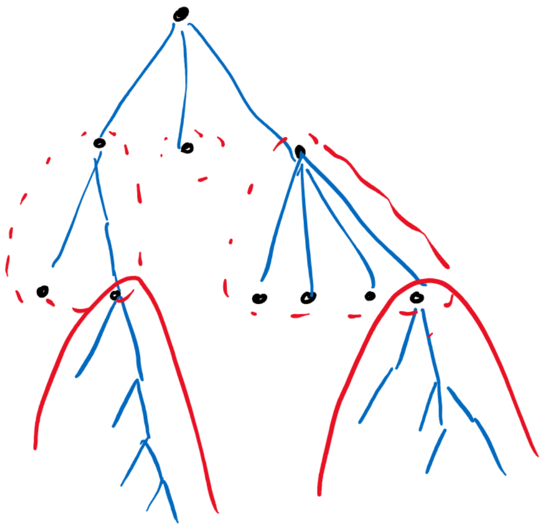
Branching Process

"Modelling growth of a population".
History:

- Bienaymé (1840s)
- Galton-Watson (1870s)

Setting: Random branching tree.
Let X be a random variable on $\{0, 1, 2, \dots\}$.

- One individual at generation 0
- has a random number of children, with distribution X. If 0, end. Each child independently has some children, each with distribution X.
- Continue.



Goal:

- Study number of individuals in each generation
- Total population size: is it *finite* of *infinite*.

Reduction: Write Z_n = number of individuals in generation n.

$$Z_0 = 1, \qquad Z_1 \sim X, \qquad Z_{n+1} = Z_1^{(n)} + \cdots + X_{Z_n}^{(n)}$$

"$X_k^{(n)}$ = number of children of k-th individual in generation n".

> **Note.** If $Z_n = 0$ then $Z_{n+1} = Z_{n+2} = \cdots = 0$.

Key Observation: Z_{n+1} is a random sum,

$$\mathbb{E}[Z_{n+1}] = \mathbb{E}[X]\mathbb{E}[Z_n]$$

Induction:

$$\mathbb{E}[Z_n] = (\mathbb{E}[X])^n$$

Notation:

$$\mu = \mathbb{E}[X] \implies \mathbb{E}[Z_n] = \mu^n.$$

Using PGFs: Let G be the PGF of X, G_n the PGF of Z_n.
Random sums:

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

Induct:

$$G_n(z) = \underbrace{G(\cdots G(}_{n\ Gs} z) \cdots)$$

Key event of interest:

$$\{Z_n = 0\}, \qquad q_n = \mathbb{P}(Z_n = 0)$$

"extinct by generation n".

Definition (Extinction Probability)**.**

$$q = \mathbb{P}(Z_n = 0 \text{ for } n \geq 1)$$

(which is the probability that the population size is finite)

Note. $\{Z_n = 0\} \subseteq \{Z_{n+1} =\}$. Why? Because $Z_n = 0 \implies Z_{n+1} = 0$, and

$$\{Z_n \text{ for some } n \geq 1\} = \bigcup_{n \geq 1} \{Z_n = 0\}$$

So continuity gives

$$\mathbb{P}(Z_n = 0) \uparrow \mathbb{P}\left(\bigcup_{n \geq 1} \{Z_n = 0\}\right)$$

so

$$q_n \uparrow q$$

as $n \to \infty$.

Classification:

- $\mu < 1$ *subcritical*
- $\mu = 1$ *critical*
- $\mu > 1$ *supercritical*

Degenerate case: $\mathbb{P}(X = 1) = 1$. Boring $\to$ exercise.

Theorem. Assume $\mathbb{P}(X = 1) \neq 1$. Then $q = 1$ (i.e. "always finite / dies out") if and only if $\mu = \mathbb{E}[X] \le 1$.

Remark. Interesting that depends on X only through $\mathbb{E}$.

Start of lecture 16

Interpretation: "Finite" eg 100 out of a large population, "Infinite" $\to$ affects positive proportion of population.

Proof (baby proof). (subcritical) $\mu < 1$

$$\mathbb{P}(Z_n \ge 1) \le \frac{\mathbb{E}[Z_n]}{1} = \mu^n \to 0$$

(Markov's Inequality)
(supercritical):

Note. $\mathbb{E}[Z_n] \to \infty$ does *not* imply $\mathbb{P}(Z_n = 0) \not\approx 1$.

Reminder: G the PGF of X, G_n the PGF of Z_n. We care about $\{Z_n = 0\}$, $q_n = \mathbb{P}(Z_n = 0)$. Also $q_n = G_n(0)$.

Claim. q the extinction probability, then $G(q) = q$.

Proof 1. G continuous. Note $q_{n+1} = G(q_n)$ and $q_{n+1} \to q$, and $G(q_n) \to G(q)$ so $q = G(q)$. $\square$

Proof 2. LTP (revision of random sums)

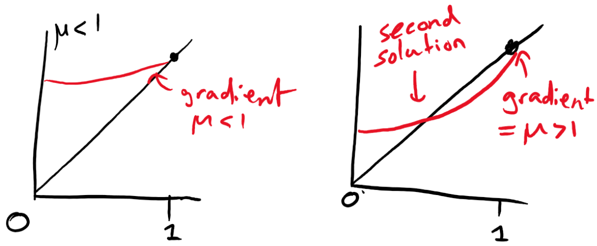
$$\text{Total finite} \iff \text{ALl subtrees of 1st generation are finite}$$

$$\begin{aligned} q &= \mathbb{P}(\text{finite}) \\ &= \sum_{k \ge 0} \mathbb{P}(\text{all finite} \mid Z_1 = l)\mathbb{P}(Z_1 = k) \\ &= \sum_{k \ge 0} [\mathbb{P}(\text{finite})]^k \mathbb{P}(Z_1 = k) \\ &= \sum_{k \ge 0} q^k \mathbb{P}(Z_1 = k) \\ &= G(q) \end{aligned}$$

$\square$

Facts about G:

- $G(0) = \mathbb{P}(X = 0) \geq 0$
- $G(1) = 1$
- $G'(1) = \mathbb{E}[X] = \mu$
- G is *smooth*, all derivatives ≥ 0 on $[0, 1)$.



Remark.

- Exactly one solution on $[0, 1)$
- By IVT / Rolle on $G(z) - z$.

Theorem. Assume $\mathbb{P}(X = 1) \neq 1$. Then q is the *minimal* solution to $z = G(z)$ in $[0, 1]$.

Corollary. $q = 1 \iff \mu \leq 1$.

Proof. Let t be the minimal solution. Reminder: G is increasing,

$$
\begin{aligned}
&t \geq 0 \\
\implies &G(t) \geq G(0) \\
\implies &G(G(t)) \geq G(G(0)) \\
\implies &G_n(t) \geq G_n(0) \\
\implies &t \geq q_n \\
\implies &t \geq q
\end{aligned}
$$

Note q *is* a solution, so we must have $q = t$. □ □

Continuous Probability

Focus now: Case where $\text{Im}(X)$ is an *interval* in $\mathbb{R}$.
Why?

- Natural for measuring, for example physical quantity, for example proportions
- "Limits" of discrete random variable
- Calculus tools for nice calculations

Redefinition:

Definition. A random variable X on $(\omega, \mathcal{F}, \mathbb{P})$ is a function $X : \Omega \to \mathbb{R}$ such that $\{X \leq x\} \in \mathcal{F}$.

Check: consistent with previous definition when Ω countable (or $\text{Im}(X)$ is countable).

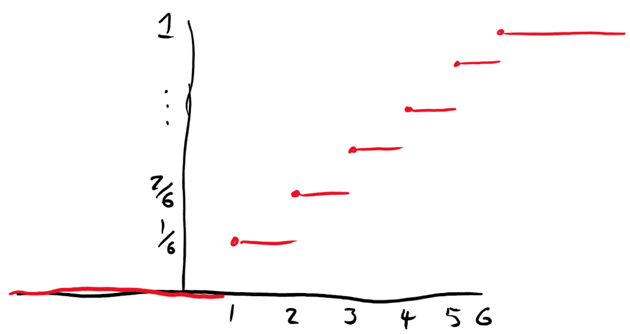
Drawback: Can't take $\mathcal{F} = \mathcal{P}(\mathbb{R})$.

Definition. The *cumulative distribution function* (CDF) of RV X is $F_X : \mathbb{R} \to [0, 1]$

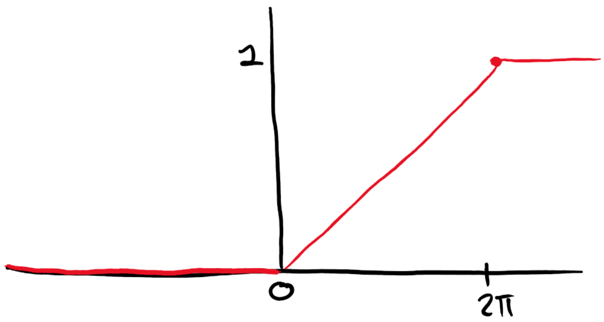
$$F_X(x) = \mathbb{P}(X \leq x)$$

Examples

X a dive on $\{1, \ldots, 6\}$.



Angle of ludo spinner:



Properties of CDF

- F_X increasing, i.e. $x \le y \implies F_X(x) \le F_X(y)$. Why? $F_X(x) = \mathbb{P}(X \le x) \le \mathbb{P}(X \le y) = F_X(y)$.
- $\mathbb{P}(X > x) = 1 - F_X(x)$
- $\mathbb{P}(a < x \le b) = F_X(b) - F_X(a)$. Why? $\mathbb{P}(a < X \le b) = \mathbb{P}(X \le b) - \mathbb{P}(X \le a)$.
- F_X is right-continuous and left limits exist, i.e.

  $$\lim_{y \downarrow x} F_X(y) = F_X(x)$$

  and

  $$\lim_{y \uparrow x} F_X(y) = F_X(x^-) = \mathbb{P}(X < x)$$

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

Start of lecture 17

Proof.

- (right-continuous) Sufficient to prove

  $$F_X\left(x + \frac{1}{n}\right) \to F_X(x)$$

  as $n \to \infty$.

  $$A_n = \left\{x < X \le x + \frac{1}{n}\right\}$$

  decreasing events, with

  $$\bigcap_{n \ge 1} A_n = \emptyset$$

  so

  $$\mathbb{P}(A_n) = F_X\left(x + \frac{1}{n}\right) - F_X(x) \to 0$$

- (left-limits) $F_X\left(x - \frac{1}{n}\right)$ is a sequence *increasing* bounded above by $F_X(x)$. $\left\{X_n \leq x - \frac{1}{n}\right\}$ is a *increasing* sequence of events with

$$\bigcup_{n \geq 1} \left\{X \leq x - \frac{1}{n}\right\} = \{X < x\}$$

so

$$F_X\left(x - \frac{1}{n}\right) = \mathbb{P}\left(X \leq x - \frac{1}{n}\right) \to \mathbb{P}(X < x)$$

- ($\lim_{x\to\infty} F_X(x) = 1$) $\{X \leq n\}$ increasing events,

$$\bigcup_{n \geq 1} \{X \leq n\} = \Omega$$

so

$$F_X(n) = \mathbb{P}(X \leq n) \to \mathbb{P}(\Omega) = 1$$

- Similar for $\lim_{x\to-\infty} F_X(x) = 0$.

□

Definition. • A random variable is *continuous* if F is continuous. This implies that

- $F_X(x) = F_X(x^-) \iff \mathbb{P}(X \leq x) = \mathbb{P}(X < x) \iff \mathbb{P}(X = x) = 0 \quad \forall x$
- and *in this course* F is also differentiable so that

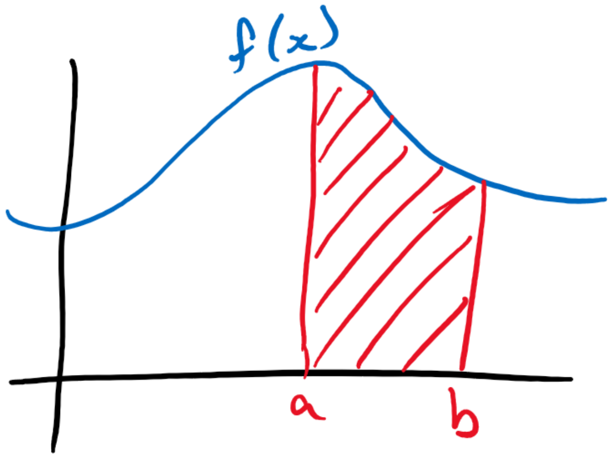
$$F_X(x) = \mathbb{P}(X \leq x) = \int_{u=-\infty}^{x} f_X(u)\mathrm{d}u$$

(cf Part II P & M) where $f_X : \mathbb{R} \to \mathbb{R}$ has the properties:

- $f_X(x) \geq 0$ for all x
- $\int_{-\infty}^{\infty} f_X(x)\mathrm{d}x = 1$

f_X is the *probability density function* of X (*PDF* or "*density*").

Intuitive Meaning:



$$\mathbb{P}(x < X \leq x + \delta x) = \int_x^{x+\delta x} f_X(u)\mathrm{d}u \approx \delta x \cdot f(x)$$

$$\mathbb{P}(a < X \leq b) = \int_a^b f_X(x)\mathrm{d}x = \mathbb{P}(a \leq X < b)$$

So for $S \subset \mathbb{R}$ (S "nice" for example interval or countable union of intervals).

$$\mathbb{P}(X \in S) = \int_S f_X(u)\mathrm{d}u$$

Key Takeaways

- The CDF is a collection of probabilities
- PDF is *not* a probability. How to use? Integrate it to get a probability.

Examples

(1) Uniform distribution $X \sim U[a, b]$ ($a, b \in \mathbb{R}$, $a < b$).

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

$$F_X(x) = \int_a^x f_X(u)\mathrm{d}u \frac{x-a}{b-a}$$

for $a \leq x \leq b$.
Question: "Limit of discrete uniform random variables?"

(2) Exponential distribution $\lambda > 0$.

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

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

Check:

(i) ≥ 0? Yes

(ii) $\int_0^\infty f_X(x) = [-e^{-\lambda x}]_0^\infty = 1$.

$$F_X(x) = \mathbb{P}(X \geq x) = \int_0^x \lambda e^{-\lambda u} \mathrm{d}u = 1 - e^{-\lambda x}$$

Remember:

$$\mathbb{P}(X \geq x) = 1 - F_X(x) + \mathbb{P}(X = x) = e^{-\lambda x}$$

"Limit of (rescaled) geometric distribution". Good way to model *arrival times* "how long to wait before something happens" $\to$ link to Poisson usage $\leftrightarrow$ Part II Applied Probability.

Memoryless Probability

(Conditional $\mathbb{P}$ works as before). $X \sim \mathrm{Exp}(\lambda)$, $s, t > 0$.

$$\begin{aligned} \mathbb{P}(X \geq s + t \mid X \geq s) &= \frac{\mathbb{P}(X \geq s + t)}{\mathbb{P}(X \geq s)} \\ &= \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} \\ &= e^{-\lambda t} \\ &= \mathbb{P}(X \geq t) \end{aligned}$$

Exercise: X memoryless $\iff$ $X \sim \mathrm{Exp}(\lambda)$. (continuous random variable with a density).

Expectation of Continuous Random Variables

Definition. X has density f_X. The *expectation* is

$$\mathbb{E}[X] := \int_{-\infty}^{\infty} x f_X(x) \mathrm{d}x$$

and

$$\mathbb{E}[g(X)] := \int_{-\infty}^{\infty} g(x) f_X(x) \mathrm{d}x$$

<u>Technical Comment</u>: assumes at most one of

$$\int_{-\infty}^{0} |x| f_X(x) \mathrm{d}x$$

and

$$\int_0^\infty x f_X(x)\mathrm{d}x$$

is infinite.

Linearity of expectation:

$$\mathbb{E}[\lambda X + \mu Y] = \lambda \mathbb{E}[X] + \mu \mathbb{E}[Y]$$

as before.

Claim. $X \ge 0$. Then

$$\mathbb{E}[X] = \int_0^\infty \mathbb{P}(X \ge x)\mathrm{d}x$$

Proof.

$$\begin{aligned}
\mathbb{E}[X] &= \int_0^\infty x f_X(x)\mathrm{d}x \\
&= \int_0^\infty \left(\int_0^x 1\mathrm{d}u \right) f_X(x)\mathrm{d}x \\
&= \int_0^\infty \mathrm{d}u \int_u^\infty \mathrm{d}x f_X(x) \\
&= \int_0^\infty \mathrm{d}u \mathbb{P}(X \ge u)
\end{aligned}$$

□

Start of lecture 18

Variance:

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

Examples

Uniform: $U \sim U[a, b]$.

$$\mathbb{E}[U] = \int_a^b x \frac{\mathrm{d}x}{b-a} = \frac{\frac{1}{2}b^2 - \frac{1}{2}a^2}{b-a} = \frac{a+b}{2}$$

$$\mathbb{E}[U^2] = \int_a^b x^2 \frac{\mathrm{d}x}{b-a} = \frac{\frac{1}{3}b^3 - \frac{1}{3}a^3}{b-a} = \frac{1}{3}(a^2 + ab + b^2)$$

$$\begin{aligned}
\mathrm{Var}(U) &= \frac{1}{3}(a^2 + ab + b^2) - \left(\frac{a+b}{2} \right)^2 \\
&= \frac{(b-a)^2}{12}
\end{aligned}$$

Exponential: $X \sim \text{Exp}(\lambda)$.

$$\begin{aligned}
\mathbb{E}[X] &= \int_0^\infty \lambda x e^{-\lambda x} \mathrm{d}x \\
&= [-xe^{-\lambda x}]_0^\infty + \int_0^\infty e^{-\lambda x} \mathrm{d}x \\
&= \frac{1}{\lambda} \\
E[X^2] &= \int_0^\infty \lambda x^2 e^{-\lambda x} \mathrm{d}x \\
&= [-x^2 e^{-\lambda x}]_0^\infty + 2\int_0^\infty x e^{-\lambda x} \mathrm{d}x \\
&= 0 + \frac{2}{\lambda^2} \\
\text{Var}(X) &= \frac{2}{\lambda^2} - \frac{1}{\lambda^2} \\
= \frac{1}{\lambda^2}
\end{aligned}$$

Goal: $U \sim \text{Unif}[a,b]$, $\tilde{U} \sim \text{Unif}[0,1]$. Write $U = (b-a)\tilde{U} + a$, and carry all calculations over.

Transformations of Continuous Random Variables

Goal: View $g(X)$ as a continuous random variable with its own density.

Theorem.
- X continuous random variable with density f
- $g : \mathbb{R} \to \mathbb{R}$ continuous such that
  (i) g is either strictly increasing or decreasing
  (ii) g^{-1} is differentiable

Then $g(X)$ is a continuous random variable with density

$$\hat{f}(x) = f(g^{-1}(x)) \underbrace{\left| \frac{\mathrm{d}}{\mathrm{d}x} g^{-1}(x) \right|}_{(\dagger)} \qquad (*)$$

($\dagger$ is ≥ 0 if g is strictly increasing).

Comments

- Density is? Something to integrate over to get a probability
- $(*)$ *is* integration by substitution

- Proof use CDFs (which *are* probabilities).

Proof.

$$\begin{aligned} F_{g(X)}(x) &= \mathbb{P}(g(X) \leq x) \\ &= \mathbb{P}(X \leq g^{-1}(x)) \\ &= F_X(g^{-1}(X)) \end{aligned}$$

Differentiate:

$$\begin{aligned} F'_{g(X)}(x) &= F'_X(g^{-1}(x))\frac{\mathrm{d}}{\mathrm{d}x}g^{-1}(x) \\ &= f(g^{-1}(x))\frac{\mathrm{d}}{\mathrm{d}x}g^{-1}(x) \end{aligned}$$

(g strictly decreasing is similar $\rightarrow$ exercise (revision!)) □

Sanity check: We've got two expressions for $\mathbb{E}[g(x)]$ (assume: $\mathrm{Im}(X) = \mathrm{Im}(g(X)) = (-\infty, \infty)$) new expression:

$$\begin{aligned} \mathbb{E}[g(X)] &= \int_{-\infty}^{\infty} x\hat{f}(x)\mathrm{d}x \\ &= \int_{-\infty}^{\infty} xf(g^{-1}(x))\frac{\mathrm{d}}{\mathrm{d}x}g^{-1}(x)\mathrm{d}x \end{aligned}$$

Substitute: $g^{-1}(x) = u$. So $\mathrm{d}u = \mathrm{d}x\frac{\mathrm{d}}{\mathrm{d}x}g^{-1}(x)$.

$$= \int_{u=-\infty}^{\infty} g(u)f(u)\mathrm{d}u$$

Example.

- $X \sim \mathrm{Exp}(\lambda)$, $Y = cX$.

$$\mathbb{P}(Y \leq x) = \mathbb{P}\left(X \leq \frac{X}{c}\right) = 1 - e^{-\lambda\frac{x}{c}} = 1 - e^{-\frac{\lambda}{c}x} = \text{CDF of Exp}\left(\frac{\lambda}{c}\right)$$

- $\hat{f}(x) = \frac{1}{c}f\left(\frac{x}{c}\right) = \frac{1}{c}\lambda e^{-\lambda\frac{x}{c}} = \frac{\lambda}{c}e^{-\frac{\lambda}{c}x}$.

Example. The *Normal* Distribution (also *Gaussian*). Range: $(-\infty, \infty)$. Two parameters: $\mu \in (-\infty, \infty), \sigma^2 \in (0, \infty)$. (the mean and variance).

$$X \sim \mathrm{N}(\mu, \sigma^2)$$

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

Special case: "Standard normal": $Z \sim \mathrm{N}(0, 1)$

$$f_Z(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} =: \varphi(x)$$

Comments

- $\frac{1}{\sqrt{2\pi}}$ is a "*normalising constant*". (Recall we need $\int f \mathrm{d}x = 1$).
- $e^{-\frac{x^2}{2}}$ = very rapid decay as $x \to \pm\infty$.
- $\mathrm{N}(\mu, \sigma^2)$ used for modelling non-negative quantity. (because if μ is large $\mathbb{P}(\mathrm{N}(\mu, \sigma^2) < 0)$ is *very* small).

Checklist

(Z, standard normal)

(i) f_Z is a density. *Proof.*

$$I = \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} \mathrm{d}x$$

Clever idea: use I^2 instead

$$I^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{u^2}{2}} e^{-\frac{v^2}{2}} \mathrm{d}u\mathrm{d}v = \iint e^{-\frac{u^2+v^2}{2}} \mathrm{d}u\mathrm{d}v$$

Polar coordinates: $u = r\cos\theta$, $v = r\sin\theta$:

$$= \int_{r=0}^{\infty} \int_{\theta=0}^{2\pi} re^{-\frac{r^2}{2}} \mathrm{d}r\mathrm{d}\theta = 2\pi \int_{r=0}^{\infty} re^{-\frac{r^2}{2}} \mathrm{d}r = 2\pi$$

□

(ii) $\mathbb{E}[Z] = 0$ by symmetry.

(iii) $\mathrm{Var}(Z) = 1$. *Proof.* Sufficient to prove $\mathbb{E}[Z^2] = 1$.

$$\begin{aligned}
\mathbb{E}[Z^2] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2}} \, \mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x \cdot x e^{-\frac{x^2}{2}} \, \mathrm{d}x \\
&= \left[-x \cdot \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} \right]_{-\infty}^{\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} \, \mathrm{d}x \\
&= 1
\end{aligned}$$

□

Start of lecture 19

Studying $\mathrm{N}(\mu, \sigma^2)$ via linear transformations

Facts about $X \sim \mathrm{N}(\mu, \sigma^2)$:

(i) X has the same distribution as $\mu + \sigma Z$ where $Z \sim \mathrm{N}(0, 1)$.

(ii) X has CDF

$$F_X(x) = \Phi\left(\frac{x - \mu}{\sigma}\right)$$

Notation. Φ is the CDF of $\mathrm{N}(0, 1)$

(iii) $\mathbb{E}[X] = \mu$, $\mathrm{Var}(X) = \sigma^2$.

Proof.

(i) $g(z) = \mu + \sigma z$ so $g^{-1}(x) = \frac{x-\mu}{\sigma}$. Then $g(Z)$ has density

$$\begin{aligned}
&= \frac{1}{\sigma} f_Z\left(\frac{x - \mu}{\sigma}\right) \\
&= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}
\end{aligned}$$

(ii) $F_{g(Z)}(x) = \mathbb{P}(g(Z) \le x) = \mathbb{P}\left(Z \le \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right)$.

(iii) Use part (i):

$$\mathbb{E}[X] = \mathbb{E}[\mu + \sigma Z] = \mu + \sigma \mathbb{E}[Z] = \mu$$

$$\mathrm{Var}(\mu + \sigma Z) = \sigma^2 \mathrm{Var}(Z) = \sigma^2$$

□

Remark. Reduces to Φ: lookup in book / table / Wolfram Alpha.

Usage: $X \sim \mathrm{N}(\mu, \sigma^2)$

$$\begin{aligned}\mathbb{P}(a \le X \le b) &= \mathbb{P}\left(\frac{a-\mu}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{b-\mu}{\sigma}\right) \\ &= \mathbb{P}\left(\frac{a-\mu}{\sigma} \le Z \le \frac{b-\mu}{\sigma}\right) \\ &= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)\end{aligned}$$

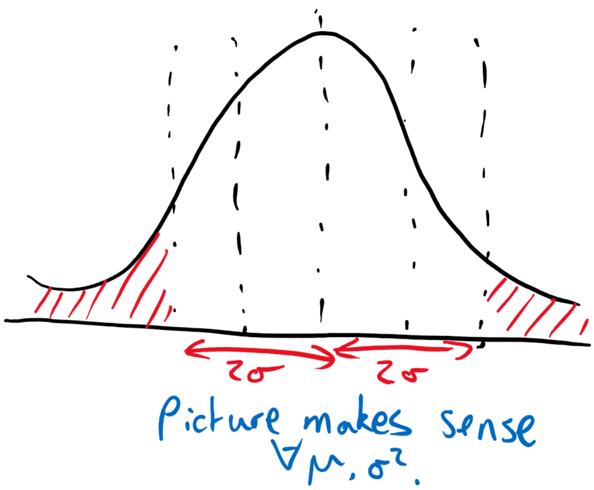
Special case:

$$a = \mu - k\sigma, \qquad b = \mu + k\sigma$$

($k \in \{1, 2, \dots\}$). Recall: σ is the *standard deviation*.

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

"within k standard deviations of the mean".



Definition. X a continuous random variable. The *median* of X is the number m such that $\mathbb{P}(X \le m) = \mathbb{P}(X \ge m) = \frac{1}{2}$, i.e.

$$\int_{-\infty}^{m} f_X(x)\mathrm{d}x = \int_{m}^{\infty} f_X(x)\mathrm{d}x = \frac{1}{2}$$

Comments

- For $X \sim \mathrm{N}(\mu, \sigma^2)$ and other distributions symmetric about mean, we have median $m = \mathbb{E}[X]$.
- Sometimes $|X - m|$ better than $|X - \mu|$ for interpretation.

More than one continuous Random Variables

Allow random variables to take values in $\mathbb{R}^n$. For example

$$X = (X_1, \dots, X_n) \in \mathbb{R}^n$$

is a random variable. Say X has density $f : \mathbb{R}^n \to [0, \infty)$ if

$$\mathbb{P}(X_1 \le x_1, \dots, x_n \le x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f(u_1, \dots, u_n) \prod_i \mathrm{d}u_i$$

(integrate over $(-\infty, x_1] \times \cdots \times (-\infty, x_n]$)

Consequence:

$$\mathbb{P}((X_1, \dots, X_n) \in A) = \int_A f(u)\mathrm{d}u$$

for all "measurable" $A \subset \mathbb{R}^n$.

Definition. f is called a *multivariate density function* or (especially $n = 2$) a *joint density.*

Definition. Random variables $X_1, \dots, X_n$ *independent* if

$$\mathbb{P}(X_1 \le x_1, \dots, X_n \le x_n) = \mathbb{P}(X_1 \le x_1) \cdots \mathbb{P}(X_n \le x_n) \qquad (*)$$

Goal: convert to statement about densities.

Definition. $X = (X_1, \dots, X_n)$ has density f. The *marginal density* f_{X_i} of X_i is

$$f_{X_i}(x_i) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \dots, x_n) \prod_{j \ne i} \mathrm{d}x_j$$

"density of X_i viewed as a random variable by itself".

Theorem 1. $X = (X_1, \dots, X_n)$ has density f.

(a) if $X_1, \dots, X_n$ independent, with marginals $f_{X_1}, \dots, x_{X_n}$. Then

$$f(X_1, \dots, X_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n)$$

(b) Suppose f factorises as

$$f(X_1, \dots, X_n) = g_1(x_1) \cdots g_n(x_n)$$

for non-negative functions (g_i). Then $X_1, \dots, X_n$ are independent and marginal $f_{X_i} \propto g_i$.

Proof.

(a)

$$\begin{aligned}
\mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n) &= \mathbb{P}(X_1 \leq x_1) \cdots \mathbb{P}(X_n \leq x_n) \\
&= \left[\int -\infty^{\infty} f_{X_1}(u_1) \mathrm{d}u_1 \right] \cdots \left[\int_{-\infty}^{\infty} f_{X_n}(u_n) \mathrm{d}u_n \right] \\
&= \int_{-\infty}^{x_1} \int_{-\infty}^{x_n} \prod f_{X_i}(u_i) \prod \mathrm{d}u_i
\end{aligned}$$

which matches with definition of f.

(b) Idea:

- Replace $g_i(x)$ with $h_i(x) = \frac{g_i(x)}{\int g_i(u)\mathrm{d}u}$. h_i *is* a density.
- compute integral at $(*)$

□

Transformation of Multiple Random Variables

Key Example 1: X, Y independent with densities f_X, f_Y.
Goal: density of $Z = X + Y$.

Step 1: Declare the joint density

$$f_{X,Y}(x, y) = f_X(x) f_Y(y).$$

Step 2: CDF of Z:

$$\begin{aligned}\mathbb{P}(X+Y \le z) &= \iint_{\{x+y\le z\}} f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y \\ &= \int_{x=-\infty}^{\infty}\int_{y=-\infty}^{z-x} f_X(x)f_Y(y)\mathrm{d}x\mathrm{d}y \\ &= \int_{x=-\infty}^{\infty}\int_{y'=-\infty}^{z} f_Y(y'-x)f_X(x)\mathrm{d}y'\mathrm{d}x \qquad \text{substitute } y' = y+x \\ &= \int_{y=-\infty}^{x} \mathrm{d}y \left(\int x=-\infty^{\infty} f_Y(y-x)f_X(x)\mathrm{d}x\right)\end{aligned}$$

So density of Z:

$$f_Z(z) = \underbrace{\int_{x=-\infty}^{\infty} f_Y(z-x)f_X(x)\mathrm{d}x}_{\text{Convolution of } f_X \text{ and } f_Y}$$

Start of lecture 20

Note. The discrete equivalent is $X, Y \ge 0$ independent,

$$\mathbb{P}(X+Y) = k) = \sum_{l=0}^{k} \mathbb{P}(X=l)\mathbb{P}(Y=k-l)$$

Example. $X, Y \overset{\text{IID}}{\sim} \mathrm{Exp}(\lambda)$. $Z = X+Y$.

$$\begin{aligned}f_Z(z) &= \int_{x=0}^{z} \lambda^2 e^{-\lambda x}e^{-\lambda(z-x)}\mathrm{d}x \\ &= \lambda^2 \int_{x=0}^{z} e^{-\lambda z}\mathrm{d}z \\ &= \lambda^2 z e^{-\lambda z}\end{aligned}$$

Definition. $X \sim \mathrm{J}(n,\lambda)$ *Gamma distribution.* $\lambda > 0$, $n \in \{1, 2, \dots\}$. Range is $[0,\infty)$. Density:

$$f_X(x) = e^{-\lambda x}\frac{\lambda^n x^{n-1}}{(n-1)!}$$

$$n = 1 \mapsto \mathrm{Exp}(\lambda)$$

$$n = 2 \mapsto \lambda^2 x e^{-\lambda x}$$

So $X+Y \sim J(2,\lambda)$. (and in fact: $X_1 + \cdots + X_n \sim J(n,\lambda)$).

Example. $X_1 \sim \mathrm{N}(\sigma_1, \sigma_1^2)$, $X_2 \sim \mathrm{N}(\mu_2, \sigma_2^2)$ independent. Then: $X_1 + X_2 \sim \mathrm{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

Note. Already know that

$$\mathbb{E}[X_1 + X_2] = \mu_1 + \mu_2 \qquad \mathrm{Var}(X_1 + X_2) = \sigma_1^2 + \sigma_2^2$$

Proof.

- Calculation exercise
- Generating functions?? Coming up.

□

Theorem. Let $X = (X_1, \ldots, X_n)$ on D. $g : \mathbb{R}^n \to \mathbb{R}^n$ well-behaved.

$$U = g(X) = (U_1, \ldots, U_n)$$

Joint density $f_X(x)$ is continuous. Then joint density

$$f_U(\mathbf{u}) = f_X(g^{-1}(\mathbf{u}))|J(\mathbf{u})|$$

where

$$J = \det\left(\left(\frac{\partial [g^{-1}]_i}{\partial u_j}\right)_{i,j=1}^{n}\right)$$

"Jacobean" ($d \times d$ matrix)

"Proof" Definition of multivariate integration by substitution. □

Example (Radial Symmetry). $X, Y \overset{\text{IID}}{\sim} \mathrm{N}(0,1)$. Write $(X,Y) = (R\cos\theta, R\sin\theta)$. Range: $R > 0$, $\theta \in [0, 2\pi)$.

$$\begin{aligned} f_{X,Y}(x,y) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \\ &= \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}} \end{aligned}$$

Note.

$$|\text{Jacobean of } g^{-1}| = \frac{1}{|\text{Jacobean of } g|}$$

$$J = \begin{vmatrix} \cos\theta & \sin\theta \\ -R\sin\theta & R\cos\theta \end{vmatrix} = R(\cos^2\theta + \sin^2\theta) = R$$

So $f_{R,\theta}(r,\theta) = \frac{1}{2\pi} e^{-\frac{r^2}{2}} \times r$. Marginal:

$$f_\theta(\theta) = \frac{1}{2\pi}$$

$$f_R(r) = e^{-\frac{r^2}{2}} \times r$$

Conclusion: θ, R are independent. θ is uniform on $[0, 2\pi)$.

Note. Change of range: for example $X, Y \geq 0$, $Z = X + Y$.

$$f_{X,Z}(x,z) = ?(x,z)\mathbb{1}_{(Z \geq x)}$$

so X, Z *not* independent, even if ? splits as a product.

Moment Generating Function

Definition. Let X have density f. The *MGF* of X is:

$$m_X(\theta) := \mathbb{E}[e^{\theta X}] = \int_{-\infty}^{\infty} e^{\theta x} f(x) \mathrm{d}x$$

whenever this is finite.

Note. $m_X(0) = 1$.

Theorem. The MGF uniquely determines distribution of a random variable whenever it exists for all $\theta \in (-\varepsilon, \varepsilon)$ for some $\varepsilon > 0$.

Theorem. Suppose $m(\theta)$ exists for all $\theta \in (-\varepsilon, \varepsilon)$. Then

$$m^{(n)}(0) = \frac{\mathrm{d}^n}{\mathrm{d}\theta^n} m(\theta)\big|_{m=0} = \mathbb{E}[X^n]$$

($\mathbb{E}[X^n]$ is the "n-th moment")

Proof comment: $\frac{\partial e^{\theta x}}{\partial \theta} = x^n e^{\theta x}$.

Claim. $X_1, \dots, X_n$ independent.

$$X = X_1 + \cdots + X_n$$

Then

$$\begin{aligned} m_X(\theta) &= \mathbb{E}[e^{\theta(X_1 + \cdots + X_n)}] \\ &= \mathbb{E}[e^{\theta X_1}] \cdots \mathbb{E}[e^{\theta X_n}] \\ &= \prod m_{X_i}(\theta) \end{aligned}$$

Example. Gamma distribution: $X \sim \text{J}(n, \lambda)$.

$$f_X(x) = e^{-\lambda x} \frac{\lambda^n x^{n-1}}{(n-1)!}$$

$$\begin{aligned} m(\theta) &= \int_0^\infty e^{\theta x} e^{-\lambda x} \frac{\lambda^n x^{n-1}}{(n-1)!} \mathrm{d}x \\ &= \int_0^\infty e^{-(\lambda-\theta)x} x^{n-1} \frac{\lambda^n}{(n-1)!} \mathrm{d}x \\ &= \left(\frac{\lambda}{\lambda-\theta}\right)^n \int_0^\infty e^{-(\lambda-\theta)x} x^{n-1} \frac{(\lambda-\theta)^n}{(n-1)!} \mathrm{d}x \\ &= \left(\frac{\lambda}{\lambda-\theta}\right)^n \end{aligned}$$

($\theta < \lambda$ (and infinite if $\theta \ge \lambda$))

$$\text{Exp}(\lambda) \to \left(\frac{\lambda}{\lambda-\theta}\right) \text{MGF}$$

We've proved

$$X_1 + \cdots + X_n \sim J(n, \lambda)$$

Start of lecture 21

Example. $X \sim \text{N}(\mu, \sigma^2)$

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

$$m_X(\theta) = \exp\left(\theta\mu + \frac{\theta^2\mu^2}{2}\right)$$

So $X_1 \sim \text{N}(\mu_1, \sigma_1^2)$, $X_2 \sim \text{N}(\mu_2, \sigma_2^2)$ independent.

$$\begin{aligned} m_{X_1+X_2}(\theta) &= \exp\left(\theta\mu_1 + \frac{\theta^2\mu_1^2}{2}\right) \exp\left(\theta\mu_2 + \frac{\theta^2\sigma^2}{2}\right) \\ &= \underbrace{\exp\left(\theta(\mu_1+\mu_2) + \frac{\theta^2}{2}(\sigma_1^2+\sigma_2^2)\right)}_{\text{MGF of N}(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)} \end{aligned}$$

Convergence of Random Variables

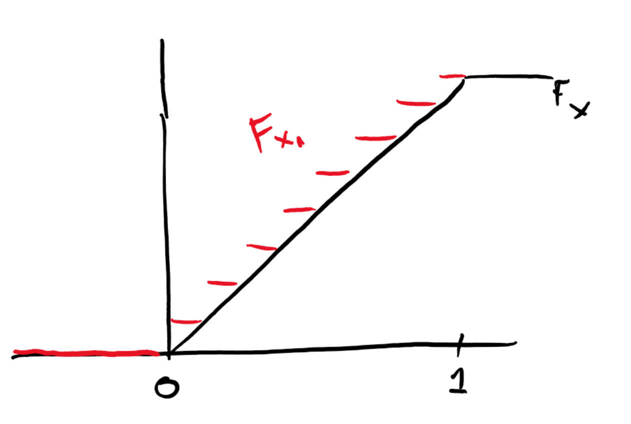
Definition. Let $(X_n)_{n\geq 1}$ and X be random variables. We say X_n *converges to* X *in distribution* and write $X_n \xrightarrow{d} X$ if

$$F_{X_n}(x) \to F_X(x) \tag{$*$}$$

for all $x \in \mathbb{R}$ which are continuity points of F_x.

Example 1.

$$X_n = \frac{1}{n}\text{Unif}(\{1, \dots, n\}) \qquad X \sim \text{Uni}[0, 1]$$



F_x continuous

- $(*)$ holds for all $x \in [0, 1]$.

Example 2.

$$X_n = \begin{cases} 0 & \text{with } \mathbb{P} = \frac{1}{2} \\ 1 + \frac{1}{n} & \text{with } \mathbb{P} = \frac{1}{2} \end{cases}$$

$$X_n \to \text{Bern}\left(\frac{1}{2}\right)$$

since $F_{X_n}(x) = \frac{1}{2}$ for all $x \in (0, 1)$, $F_{X_n}(x) = 1$ for all $x > 1$. When n is large

$$F_{X_n}(1) = \frac{1}{2} \qquad F_X(1) = 1$$

But $F_X(\bullet)$ has a discontinuity at $x = 1$. (i.e. deterministic convergence of reals)

Consequences

(1) If X is a constant c, then equivalent to:

$$\forall \varepsilon > 0 \qquad \mathbb{P}(|X_n - c| > \varepsilon) \to 0$$

as $n \to \infty$. "convergence in probability to constant".

(2) If X is a continuous random variable: $X_n \xrightarrow{d} X$. Usage:

$$\mathbb{P}(a \le X_n \le b) \to \mathbb{P}(a \le X \le b)$$

for all $a, b \in [-\infty, \infty]$.

> **Note.** Does not say that *densities* converge. For example, in Example 1 no density.

Laws of Large Numbers

$\frac{S_n}{n}$ "$\to$" μ.

> **Theorem** (Weak LLN)**.** Setup: $(X_n)_{n \ge 1}$ IID with $\mu = \mathbb{E}[X_1] < \infty$. Set
>
> $$S_n = X_1 + \cdots + X_n \qquad \forall n \ge 0$$
>
> Then $\forall \varepsilon > 0$:
>
> $$\mathbb{P}\left(\left| \frac{S_n}{n} - \mu \right| > \varepsilon \right) \to 0$$
>
> as $n \to \infty$.

Proof. (assume $\mathrm{Var}(X_1) = \sigma^2 < \infty$)

$$\begin{aligned} \mathbb{P}\left(\left| \frac{S_n}{n} - \mu \right| > \varepsilon \right) &= \mathbb{P}(|S_n - n\mu| > \varepsilon n) \\ &\le \frac{\mathrm{Var}(S_n)}{\varepsilon^2 n^2} \\ &= \frac{n\sigma^2}{\varepsilon^2 n^2} \\ &\to 0 \end{aligned}$$

as $n \to \infty$. (Note that ε is fixed, not $\varepsilon \to 0$!) □

Central Limit Theorem

Theorem (CLT)**.** Same setup as previous. Demand $\sigma^2 < \infty$. Then

$$\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \xrightarrow{d} \mathrm{N}(0,1)$$

as $n \to \infty$.

Discussion: three stage summary

(1) Distribution of S_n concentrated on $n\mu$ (WLLN)

(2) Fluctuations around $n\mu$ have order $\sqrt{n}$ (New and important)

(3) Shape is normal (Detail)

Usage:

(i) $S_n \overset{d}{\approx} \mathrm{N}(n\mu, n\sigma^2)$

(ii)

$$\begin{aligned}\mathbb{P}(a \le S_n \le b) &= \mathbb{P}\left(\frac{a-n\mu}{\sqrt{n\sigma^2}} \le \frac{S_n - n\mu}{\sqrt{n\sigma^2}} \le \frac{b-n\mu}{\sqrt{n\sigma^2}}\right) \\ &\approx \mathbb{P}\left(\frac{a-n\mu}{\sqrt{n\sigma^2}} \le Z \le \frac{b-n\mu}{\sqrt{n\sigma^2}}\right)\end{aligned}$$

Get a nice answer if $a = n\mu + z_a\sqrt{n}$ and $b = n\mu + z_b\sqrt{n}$.

Theorem (Continuity theorem for MGFs)**.** $(X_n), X$ have MGFs $m_{X_n}(\bullet), m_X(\bullet)$

- $m_X(\theta) < \infty$ for $\theta \in (-\varepsilon, \varepsilon)$
- if $m_{X_n}(\theta) \to m_X(\theta)$ for all θ such that $m_X(\theta) < \infty$.

Then $X_n \xrightarrow{d} X$.

Proof. Part II Probability and Measure. □

Idea: Expand $m_X(\theta)$ as Taylor series around 0.

$$\begin{aligned}m_X(\theta) &= 1 + m_X'(0)\theta + \frac{m_X''(0)}{2!}\theta^2 + \cdots \\ &= 1 + \theta\mathbb{E}[X] + \frac{\theta^2}{2}\mathbb{E}[X^2] + o(\theta^2)\end{aligned}$$

Proof: (WLLN via MGFs).

Remark. Know MGF of S_n. Want to study the MGF of $\frac{S_n}{n}$.

$$\begin{aligned} m_{\frac{S_n}{n}}(\theta) &= \mathbb{E}[e^{\theta \frac{S_n}{n}}] \\ &= \mathbb{E}[e^{\frac{\theta}{n} S_n}] \\ &= m_{S_n}\left(\frac{\theta}{n}\right) \\ &= m_{X_1}\left(\frac{\theta}{n}\right) \cdots m_{X_n}\left(\frac{\theta}{n}\right) \\ &= \left(1 + \mu\frac{\theta}{n} + o(\theta)\right)^n \\ &\to e^{\mu\theta} \end{aligned}$$

MGF of the random variable $X = \mu$ with $\mathbb{P} = 1$. So $\frac{S_n}{n} \xrightarrow{d} \mu$ by the continuity theorem.

Theorem (Strong LLN)**.** Same setup: Then

$$\mathbb{P}\left(\frac{S_n}{n} \to \mu \text{ as } n \to \infty\right) = 1.$$

"almost sure convergence" or "convergence with probability 1".

Start of lecture 22

Proof (CLT with MGFs). Assume WLOG $\mu = 0$ and $\sigma^2 = 1$. (So $\mathbb{E}[X_i^2] = 1$). (In general $X \mapsto \frac{X-\mu}{\sqrt{\sigma^2}}$).

Goal:

$$\frac{S_n}{\sqrt{n}} \xrightarrow{d} \mathrm{N}(0,1)$$

Study MGF of $\frac{S_n}{\sqrt{n}}$.

$$\begin{aligned} m_{X_i}(\theta) &= 1 + \frac{\theta^2}{2} + o\left(\frac{1}{n}\right) \\ m_{\frac{S_n}{\sqrt{n}}}(\theta) &= \mathbb{E}[e^{\theta \frac{S_n}{\sqrt{n}}}] \\ &= \mathbb{E}[e^{\frac{\theta}{\sqrt{n}} S_n}] \\ &= m_{S_n}\left(\frac{\theta}{\sqrt{n}}\right) \\ &= \left(m_{X_1}\left(\frac{\theta}{\sqrt{n}}\right)\right)^n \\ &= \left(1 + \frac{\theta^2}{2n} + o\left(\frac{1}{n}\right)\right)^n \\ &\to e^{\frac{\theta^2}{2}} \end{aligned}$$

□

Inequalities for $\mathbb{E}[f(X)]$

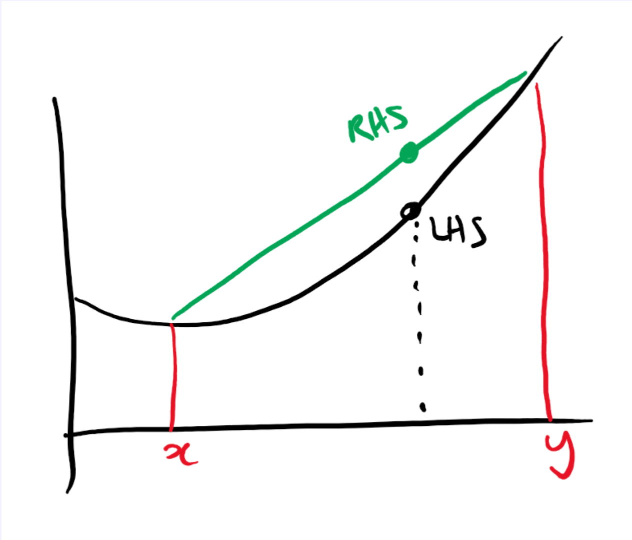
Motivation: $f(x) = x^2$. We know

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

via $\mathrm{Var}(X) \geq 0$. What about general f?

Definition. A function $f : \mathbb{R} \to \mathbb{R}$ is *convex* if $\forall\, x, y \in \mathbb{R}$ and $t \in [0,1]$,

$$f(tx + (1-t)y) \le tf(x) + (1-t)f(y)$$



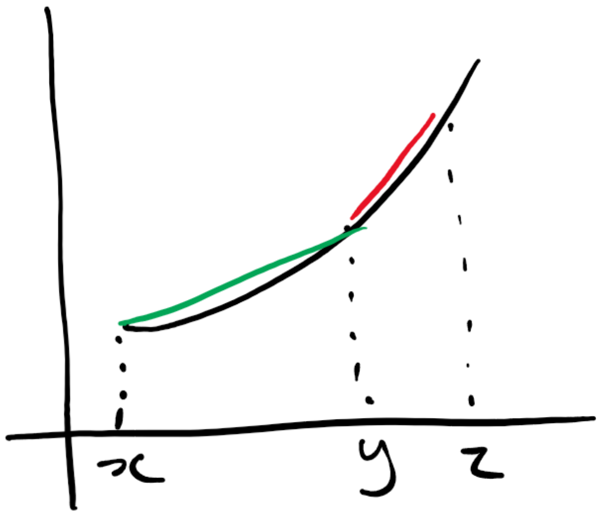
(Aside: region above f is *convex* in $\mathbb{R}^2$.)

Consequence: $\forall\, y$ there exists a line $l(x) = mx + c$ such that

- $l(x) \le f(x)$ for all x
- $l(y) = f(y)$

Proof. Convexity implies that for all $x < y < z$,

$$\frac{f(y) - f(x)}{y - x} \le \frac{f(z) - f(y)}{z - y}$$

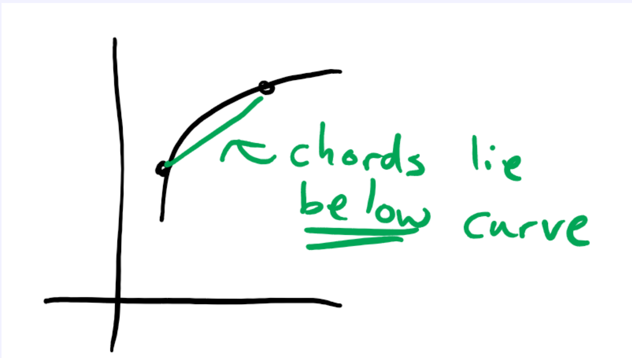


hence

$$M^- := \sup_{x<y} \frac{f(y)-f(x)}{y-x} \leq \inf_{z>y} \frac{f(z)-f(y)}{z-y} =: M^+$$

any value $m \in [M^-, M^+]$ works as the gradient of $l(\bullet)$. □

Definition. f is concave if and only if $-f$ is convex.



Fact: if f is twice differentiable then

$$f \text{ convex} \iff f''(x) \geq 0 \;\forall\; x$$

for example $f(x) = \frac{1}{x}$ is convex on $(0, \infty)$ and concave on $(-\infty, 0)$.

Jensen's Inequality

Theorem (Jensen's Inequality)**.** X a random variable, f convex:
Then $\mathbb{E}[f(X)] \geq f(\mathbb{E}[X])$. (reverse if f concave)

Proof. Set $y = \mathbb{E}[X]$ as in $(*)$, $l(x) = mx + c$, such that $l(y) = f(y) = f(\mathbb{E}[X])$ and $f \geq l$.

$$\begin{aligned} \mathbb{E}[f(X)] &\geq \mathbb{E}[l(X)] \\ &= \mathbb{E}[mX + c] \\ &= m\mathbb{E}[X] + c \\ &= my + c \\ &= f(\mathbb{E}[X]) \end{aligned}$$

If f *strictly convex*, then $\forall t \in (0,1)$, $\forall x \neq y$,

$$f(tx + (1-t)y) < tf(x) + (1-t)f(y)$$

Then equality in Jensen's inequality only if $X = \mathbb{E}[X]$ with $\mathbb{P} = 1$ (for example constant random variable). □

Informal comment:

Jensen's Inequality $\geq$ Most other inequalities!

Application to Sequences

AM-GM inequality: $x_1, \ldots, x_n \in (0, \infty)$

$$\frac{x_1 + \cdots + x_n}{n} \geq \left(\prod_{i=1}^{n} x_i \right)^{1/n}$$

Case $n = 2$:

$$\frac{x+y}{2} \geq \sqrt{xy}$$

Proof. Rearrange to get $(x-y)^2 \geq 0$. □

General proof:
Let X be a random variable taking values $\{x_1, \ldots, x_n\}$ each with probability $\frac{1}{n}$.
Take: $f(x) = -\log x$. Check convex: second derivative ≥ 0.
Jensen:

$$\begin{aligned} \mathbb{E}[f(X)] &\geq f(\mathbb{E}[X]) \\ -\frac{\log x_1 + \cdots + \log x_n}{n} &\geq -\log\left(\frac{x_1 + \cdots + x_n}{n}\right) \\ \log((x_1 \cdots x_n)^{1/n}) &\leq \log\left(\frac{x_1 + \cdots + x_n}{n}\right) \end{aligned}$$

$\log x$ and e^x are increasing so

$$\left(\prod_i x_i \right)^{1/n} \leq \frac{x_1 + \cdots + x_n}{n}$$

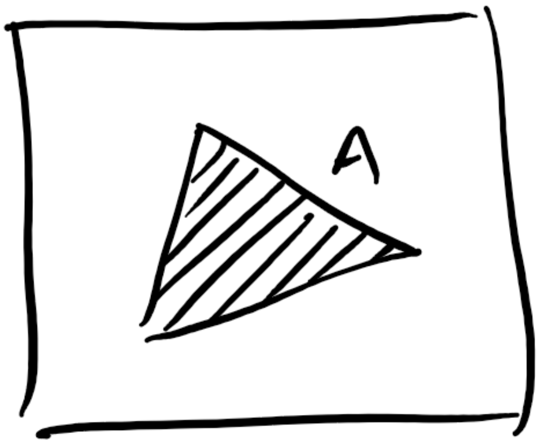
Sampling a Continuous Random Variable

Theorem. X a continuous random variable with CDF F. Then if $U \sim U[0,1]$, we have
$$Y = F^{-1}(U) \sim X$$

Proof. Goal: find CDF of Y.
$$\begin{aligned}\mathbb{P}(Y \le x) &= \mathbb{P}(F^{-1}(U) \le x) \\ &= \mathbb{P}(U \le F(x)) \\ &= F(x)\end{aligned}$$
so CDF of Y = CDF of X. So $Y \sim X$. □

Rejection Sampling

Idea: Uniform on $[0,1]^d$ is easy. (take $(U^{(1)}, \dots, U^{(d)})$ IID on $U[0,1]$.)



What about uniform on A?

Start of lecture 23

Goal:
$$f(x) = \begin{cases} \frac{1}{\text{area}(A)} & x \in A \\ 0 & x \notin A \end{cases}$$
(in higher dimensions, $\text{volume}(A)^{-1}$)

Rewrite as
$$f(x) = \frac{\mathbb{1}_A}{\text{area}(A)}$$
Let $U_1, U_2, \dots$ IID uniform on $[0,1]^d$ and let $N = \min\{n : U_n \in A\}$.

Claim. U_N is uniform on A. (i.e. has density f)

Proof. Note $\mathbb{P}(N < \infty) = 1$ if $\text{area}(A) > 0$.
Goal:

$$\mathbb{P}(U_n \in B) = \int_B f(x)\mathrm{d}x = \frac{\text{area}(B)}{\text{area}(A)}$$

for all $B \subset A$ with a well-defined area.

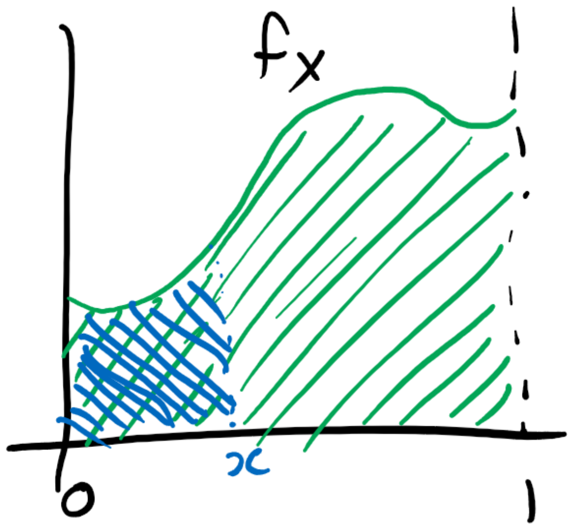
$$\begin{aligned}
\mathbb{P}(U_n \in B) &= \sum_{n\geq 1} \mathbb{P}(U_n \in B, N = n) \\
&= \sum_{n\geq 1} \mathbb{P}(U_1 \notin A, \dots, U_{n-1} \notin A, U_n \in B) \\
&= \sum_{n\geq 1} \mathbb{P}(U_1 \notin A)^{n-1}\mathbb{P}(U_n \in B) \\
&= \sum_{n\geq 1} (1 - \text{area}(A))^{n-1} \times \text{area}(B) \\
&= \frac{\text{area}(B)}{1 - (1 - \text{area}(A))} \\
&= \frac{\text{area}(B)}{\text{area}(A)}
\end{aligned}$$

□

Idea: X a continuous random variable on $[0, 1]$, density f is *bounded*. Let

$$A = \{(x, y) : x \in [0, 1], y \leq f_X(x)\}$$

i.e. shaded region



Let $U = (U^{(1)}, U^{(2)})$ be uniform on A. Then claim: $U^{(1)} \sim X$.
Why?

$$\begin{aligned}\mathbb{P}(U^{(1)} \leq u) &= \mathbb{P}(\text{in relevant area}) \\ &= \text{area}(\{x, y\} : x \leq u, y \leq f_X(x)) \\ &= \int_0^u f_X(x)\mathrm{d}x \\ &= F_X(u)\end{aligned}$$

(note that the first and last expressions are the CDFs of $U^{(1)}$ and X respectively)
Usage: in higher dimension.
X a continuous random variable on $[-K, K]^d$ with density bounded. Let

$$A = \{(\mathbf{x}, y) : x \in [-K, K]^d, y \leq f_X(x)\} \subset \mathbb{R}^{d+1}$$

Let $U = (\mathbf{U}, U^+)$. Then $\mathbf{U} \sim X$. (the proof is similar).

Multivariate Normals / Gaussians

Definition. A random variable is *Gaussian* if $X \sim \mathrm{N}(\mu, \sigma^2)$.

Motivation: X, Y independent Gaussian. Then $bX + cY$ is Gaussian $(*)$.
Exercise: there exist joint random variables (X, Y) such that both X, Y are Gaussian, but $X + Y$ not Gaussian.

Question: Can we have dependent X, Y such that $(*)$ still holds?

Definition. Random vector (X, Y) is *Gaussian* if $bX + cY$ are Gaussian for all $b, c \in \mathbb{R}$, i.e. $bX + cY \sim \mathrm{N}(??, ??)$.

Consequences:

$$\mathbb{E}[bX + cY] = b\mathbb{E}[X] + c\mathbb{E}[Y]$$
$$\mathrm{Var}(bX + cY) = b^2\mathrm{Var}(X) + c^2\mathrm{Var}(Y) + 2bc\mathrm{Cov}(X, Y)$$

Linear Algebra Rewrite

Random vector $X = (X_1, \ldots, X_n) \in \mathbb{R}^n$ is *Gaussian* if $u^\top X$ is Gaussian $\forall\ u \in \mathbb{R}^n$. Write $\mu = \mathbb{E}[X] \in \mathbb{R}^n$.
Covariance matrix:

$$V = (\mathrm{Cov}(X_i, X_j))_{i,j=1}^n \in \mathbb{R}^n \times \mathbb{R}^n$$

i.e. for $n = 2$:

$$V = \begin{pmatrix} \mathrm{Var}(X) & \mathrm{Cov}(X, Y) \\ \mathrm{Cov}(Y, X) & \mathrm{Var}(Y) \end{pmatrix}$$

(note V is symmetric). In fact $u^\top X \sim \mathrm{N}(u^\top \mu, u^\top V u)$.

MGFs in One Direction (Recap)

Distribution of $X \in \mathbb{R}$ determined by function $m_X(\theta) = \mathbb{E}[e^{\theta X}]$, $\theta \in (-\varepsilon, \varepsilon)$.

MGFs in $\mathbb{R}^n$

Distribution of $X \in \mathbb{R}^n$ determined by

$$m_X(u) = \mathbb{E}[e^{u^\top X} \qquad u \in (-\varepsilon, \varepsilon)^n$$

If X Gaussian, then

$$m_X(u) = \exp\left(u^\top \mu + \frac{1}{2} u^\top V u\right)$$

Logical overview: $X \in \mathbb{R}^n$ Gaussian

- distribution defined by MGF
- MGF defined by μ and V

$$\implies \text{distribution of } X \text{ defined by } \mu \text{ and } V$$

Remark. Density:

$$f_X(x) = \frac{1}{(2\pi)^{n/2}} \frac{1}{\sqrt{\det(V)}} \exp\left(-\frac{1}{2}(x-\mu)^\top V(x-\mu)\right)$$

Return to $n = 2$: For a Gaussian vector (X_1, X_2)

$$\text{Independent} \iff \mathrm{Cov}(X_1, X_2) = 0$$

(Note that the backwards direction is not true in general!)

Why useful? Imagine X_1, X_2 describe real-world parameters, for example height vs 1km running time.

- Independence would be an interesting conclusion
- Cov(?, ?) can be sampled.

Start of lecture 24

Proof. $X = (X_1, X_2)$ independent. If $m_X((u_1, u_2))$ splits as a product $f_1(u_1)f_2(u_2)$. In our setting:

$$\exp(u^\top \mu) = \exp(u_1\mu_2)\exp(u_2\mu_2)$$
$$\exp\left(\frac{1}{2}u^\top V u\right) = \exp(u_1^2\sigma_1^2)\exp(u_2^2\sigma_2^2)\exp(2u_1u_2\mathrm{Cov}(X_1, x_2))$$

So it splits as a product if and only if Cov $= 0$. □

Motivation: $\mathrm{Cov}(100X_1, X_2) = 100\mathrm{Cov}(X_1, X_2)$ so "*large* covariance" doesn't imply "*very* dependent".

Definition. *Correlation* of X, Y is

$$\text{Corr}(X,Y) = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$

(It is a fact that this is always $\in [-1,1]$)

Proposition. If (X,Y) Gaussian, then $Y = aX + Z$ where Z is Gaussian, and (X,Z) independent.

Proof. Define $Z = Y - aX$ for $a \in \mathbb{R}$.

Claim. (X,Z) is Gaussian.

Proof.

$$u_1X + u_2Z = u_1X + u_2(Y - aX) = (u_1 - au_2)X + u_2Y.$$

□

Goal: find a such that $\text{Cov}(X,Z) = 0$.

$$\text{Cov}(X,Z) = \text{Cov}(X, Y - aX) = \text{Cov}(X,Y) - a\text{Var}(X)$$

so take

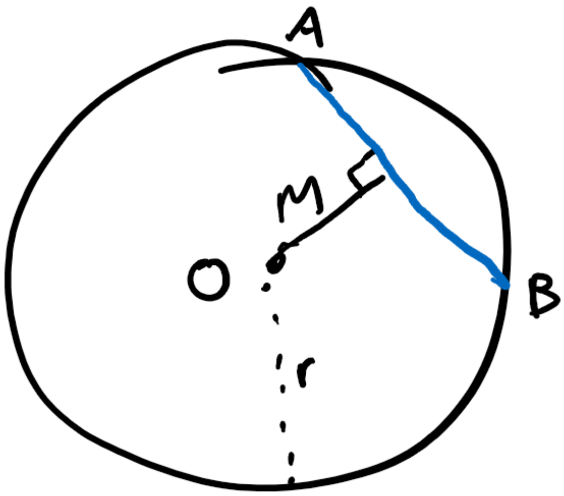
$$a = \frac{\text{Cov}(X,Y)}{\text{Var}(X)}$$

Then $\text{Cov}(X,Z) = 0$ so X, Z independent. □

2.1 Two Historical Models

Bertrand's Paradox

Goal: choose a uniform chord of circle. Two methods:

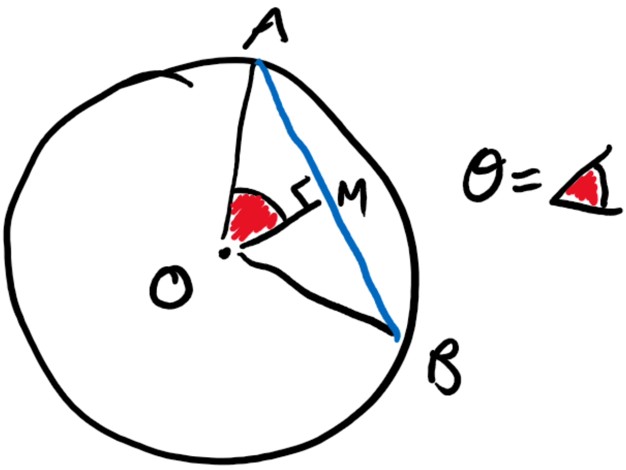


(i) A, B uniform on circumference.

(ii) midpoint M uniform on disc.

Conclusion: Gives different distributions. (Completely unsurprising?)

Method (i)

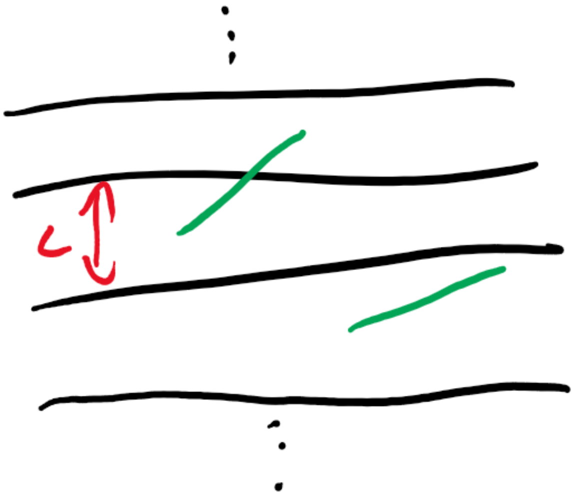


$\theta \sim \text{Unif}\left[0, \frac{\pi}{2}\right]$ then $|AB| = 2r \sin\theta$. Note $|OM| = r\cos\theta$, so $\mathbb{P}(|OM| \le \varepsilon r) \approx r\varepsilon$ when $\varepsilon \to 0$.

Method (ii)

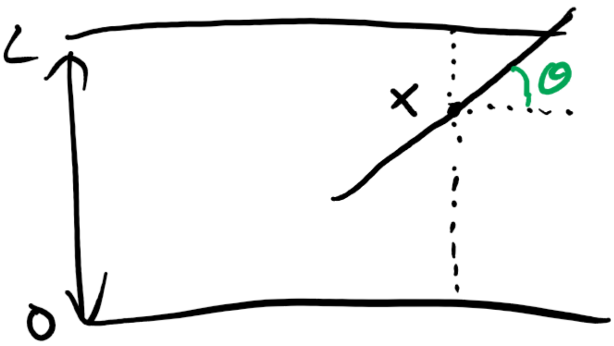
$\mathbb{P}(|OM| \le \varepsilon r) = \frac{\pi(\varepsilon r)^2}{\pi r^2} = \varepsilon^2$.

Buffon's Needle



- Lines spaced L apart.
- Needle length L dropped "uniformly"
- Observe whether intersects a line.

We work "modulo L":



$$X \text{ centre} \sim \text{Unif}[0, L)$$

$$\text{Angle } \theta \sim \text{Unif}[0, \pi)$$

Density of (X, θ) constant $= \frac{1}{L\pi}$. Crosses line if

$$X \leq \frac{L}{2} \sin \theta$$

or

$$L - X \le \frac{L}{2}\sin\theta$$

$$\begin{aligned}
\mathbb{P}(\text{crosses line}) &= \mathbb{P}\left(\min(X, L-X) \le \frac{L}{2}\sin\theta\right) \\
&= 2\mathbb{P}\left(X \le \frac{L}{2}\sin\theta\right) \\
&= 2\int_{\theta=0}^{\pi}\int_{x=0}^{\frac{L}{2}\sin\theta} \frac{1}{L\pi}\mathrm{d}x\mathrm{d}\theta \\
&= 2\int_{\theta=0}^{\pi} \frac{1}{2\pi}\sin\theta\mathrm{d}\theta \\
&= \frac{2}{\pi} \\
&\approx 0.64
\end{aligned}$$

What's the point? Calculate π experimentally.
Efficiency? Try n times. Number of intersections: $S_n \sim \text{Bin}\left(n, \frac{\pi}{2}\right)$.
Proportion $\hat{p}_n$ of intersections $= \frac{S_n}{n}$. By CLT:

$$\hat{p}_n = p + \sqrt{\frac{p(1-p)}{n}}Z$$

so

$$\hat{p}_n - p \approx \sqrt{\frac{p(1-p)}{n}}Z.$$

Estimate:

$$\hat{\pi}_n = \frac{2}{\hat{p}_n}$$

Taylor expanding:

$$\begin{aligned}
\hat{\pi}_n &= \frac{2}{\hat{p}_n} \\
&\approx \frac{2}{p} - (\hat{p}_n - p)\frac{2}{p^2}
\end{aligned}$$

so

$$\hat{\pi}_n - \pi \approx -\frac{\pi^2}{2}\sqrt{\frac{p(1-p)}{n}}Z \approx \frac{-2.4}{\sqrt{n}}Z$$

So if you seek

$$\hat{\pi}_n - \pi \approx O(10^{-k})$$

(correct to k decimal places) then we need $n \approx 10^{2k}$.

- Historical interest.
- Not computationally efficient.
- Detailed calculation of *sampling errors* in other settings on problem sheet.