21/1/21

Probability - Lecture 1

Definition Suppose Ω is a set and $\mathcal{F}$ is a collection of subsets of Ω. $(\mathcal{F} \subset \mathbb{P}(\Omega))$ power set

$\mathcal{F}$ is a σ-algebra if

(i) $\Omega \in \mathcal{F}$ $\quad (\Rightarrow \emptyset \in \mathcal{F}$ by (ii))

(ii) if $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$

(iii) for any countable collection $(A_n)_{n \geq 1}$ with $A_n \in \mathcal{F}$ $\forall n$, we also have $\bigcup_n A_n \in \mathcal{F}$.

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

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

2) For any countable disjoint collection $(A_n)_{n \geq 1}$ in $\mathcal{F}$, then

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

Call $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space

$$(\underbrace{\Omega}_{\text{sample space}}, \underbrace{\mathcal{F}}_{\sigma\text{-algebra}}, \underbrace{\mathbb{P}}_{\text{prob. measure}})$$

Remark When Ω is countable, take $\mathcal{F}$ to be all subsets of Ω (then certainly σ-algebra conditions hold).

Elements of Ω are outcomes, elements of $\mathcal{F}$ are called events

Properties of $\mathbb{P}$ (immediate from defn)

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

Examples

1) Rolling a fair die: $\Omega = \{1,2,3,4,5,6\}$, $\mathcal{F}$ = all subsets of Ω

$\mathbb{P}(\{w\}) = \frac{1}{6}$ $\forall \omega \in \Omega$ and if $A \subseteq \Omega$ then $\mathbb{P}(A) = \frac{|A|}{6}$

2) Equally likely outcomes

let Ω be a finite set $\Omega = \{\omega_1, \dots, \omega_n\}$, $\mathcal{F}$ = all subsets

Define $\mathbb{P}: \mathcal{F} \to [0,1]$ by $\mathbb{P}(A) = \frac{|A|}{|\Omega|}$ (generalisation of (1))

3) Picking balls from a bag

n balls in bag with n labels $\{1, \dots, n\}$

Pick $k \leq n$ balls at $\underbrace{\text{random}}_{\text{means all outcomes equally likely}}$ without replacement

Take $\Omega = \{A \subseteq \{1, \dots, n\} : |A| = k\}$ "k-subsets"

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

so $\mathbb{P}(\{\omega\}) = \frac{1}{|\Omega|} = \frac{1}{\binom{n}{k}}$

4) Deck of 52 cards

$\Omega = \{\text{all permutations of 52 cards}\}$: $|\Omega| = 52!$

$\mathbb{P}(\text{top 2 are aces}) = \frac{4 \times 3 \times 50!}{52!} = \frac{1}{221}$

5) Largest digit

String of <u>random</u> digits from $0, 1, \dots, 9$, length n

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

$\mathbb{P}(\text{largest digit is } k) = ?$

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

$\mathbb{P}(B_k) = \frac{|B_k|}{|\Omega|}$; notice $B_k = A_k \setminus A_{k-1}$

$|A_k| = (k+1)^n \Rightarrow |B_k| = (k+1)^n - k^n$

so $\mathbb{P}(B_k) = \frac{(k+1)^n - k^n}{10^n}$

6) Birthday problem

n people - what is prob. that at least 2 share the same birthday? (Assume 365 day year)

$\Omega = \{1, 2, \dots, 365\}^n$ $\mathcal{F}$ = all subsets

$\mathbb{P}(\{\omega\}) = \frac{1}{365^n}$

$A = \mathbb{P}$(at least 2 people share same birthday)

Consider $A^c = \mathbb{P}$(all n birthdays are different)

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

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

$n = 22 \Rightarrow \mathbb{P}(A) \approx 0.476$

$n = 23 \Rightarrow \mathbb{P}(A) \approx 0.507$

Combinatorial Analysis

1. Ω finite set, $|\Omega| = n$

Want to partition Ω into k disjoint subsets $\Omega_1, \dots, \Omega_k$ with $|\Omega_i| = n_i$ and $\sum_{i=1}^{k} n_i = n$. How many ways?

$$\# \text{ ways } M = \binom{n}{n_1}\binom{n-n_1}{n_2} \dots \binom{n-(n_1+\dots+n_{k-1})}{n_k} = \frac{n!}{n_1! n_2! \dots n_k!}$$

Write
$$\binom{n}{n_1, \dots, n_k} = \frac{n!}{n_1! \dots n_k!}$$

2. How many strictly increasing functions are there from $\{1, 2, \dots, k\}$ to $\{1, 2, \dots, n\}$? $\quad k \leq n$

Any such function is uniquely determined by its range which is a subset of $\{1, \dots, n\}$ of size k. There are $\binom{n}{k}$ such subsets, so $\binom{n}{k}$ strictly increasing functions.

How many increasing functions?

Define bijection from $\{f : \{1, \dots, k\} \to \{1, \dots, n\} \text{ increasing}\}$ to $\{g : \{1, \dots, k\} \to \{1, \dots, n+k-1\} : \text{strictly increasing}\}$

$\forall f$ increasing, define $g(i) = f(i) + i - 1$. Then g is strictly increasing and takes values in $\{1, \dots, n+k-1\}$ and there are $\binom{n+k-1}{k}$ such functions, so $\binom{n+k-1}{k}$ increasing functions from $\{1, 2, \dots, k\}$ to $\{1, 2, \dots, n\}$. $\square$

Probability - Lecture 2

22/1/21

Stirling's Formula

Notation: Let (a_n) and (b_n) be sequences. We write $a_n \sim b_n$ if $\frac{a_n}{b_n} \to 1$ as $n \to \infty$.

Theorem (Stirling) $n! \sim n^n \sqrt{2\pi n}\, e^{-n}$ as $n \to \infty$.

Weaker statement $\log(n!) \sim n \log n$ as $n \to \infty$

Proof of weaker statement Define $l_n = \log(n!) = \log 2 + \dots + \log n$

For $x \in \mathbb{R}$, write $\lfloor x \rfloor$ for integer part of x. Then

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

Integrate from 1 to n.

$$\sum_{k=1}^{n-1} \log k \leq \int_1^n \log x \, dx \leq \sum_{k=1}^{n} \log k \Rightarrow l_{n-1} \leq n\log n - n + 1 \leq l_n$$

For all n, we get $l_n \leq (n+1)\log(n+1) - (n+1) + 1$

and $n \log n - n + 1 \leq l_n$

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

Divide through by $n \log n$ to get $\frac{l_n}{n \log n} \to 1$ as $n \to \infty$. $\square$

Proof (Stirling, non-examinable)

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

(check - IBP on 2nd integral twice)

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

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

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

Take the sum for $k=1,\dots,n-1$ of the equality above to get

$$\int_1^n \log x \, dx = \frac{\log((n-1)!) + \log(n!)}{2} + \frac{1}{2}\sum_{k=1}^{n-1}\int_0^1 \frac{x(1-x)}{(x+k)^2}\,dx$$

$$n\log n - n + 1 = \log(n!) - \frac{\log n}{2} + \sum_{k=1}^{n-1} a_k \quad \left(a_k = \frac{1}{2}\int_0^1 \frac{x(1-x)}{(x+k)^2}\,dx \right)$$

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

$$\Rightarrow \quad n! = n^n e^{-n}\sqrt{n}\, \exp\left(1 - \sum_{k=1}^{n-1} a_k\right)$$

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

so $\sum a_k$ coverges. Set $A = \exp\left(1 - \sum_{k=1}^{\infty} a_k\right)$

$$\Rightarrow \quad n! = n^n e^{-n}\sqrt{n}\, A \exp\Big(\underbrace{\sum_{k=n}^{\infty} a_k}_{\to 0 \text{ as } n\to\infty}\Big) \quad \text{so } \exp \to 1 \text{ as } n\to\infty$$

So $\frac{n!}{n^n e^{-n}\sqrt{n}} \to A$ as $n\to\infty$, so $\underline{n! \sim n^n e^{-n}\sqrt{n}\,A}$ as $n\to\infty$.

Find A. We have

$$2^{-2n}\binom{2n}{n} = \frac{2^{-2n}(2n)!}{n!\,n!} \sim \frac{2^{-2n}(2n)^{2n}\sqrt{2n}\cdot A e^{-2n}}{n^n e^{-n}\sqrt{n}A \cdot n^n e^{-n}\sqrt{n}A} = \underline{\frac{\sqrt{2}}{A\sqrt{n}}}$$

Consider $I_n = \int_0^{\pi/2} (\cos\theta)^n \, d\theta$, $n \geq 0$. $I_0 = \frac{\pi}{2}$, $I_1 = 1$

IBP: $I_n = \frac{n-1}{n} I_{n-2}$ (check) so $I_{2n} = \frac{2n-1}{2n} I_{2n-2}$

repeat recurrence: $I_{2n-2} = \frac{(2n-1)(2n-3)\dots 3\times 1}{2n(2n-2)\dots 2} I_0 = \frac{(2n)!}{2^{2n}\cdot(n!)^2}\cdot\frac{\pi}{2}$

so $I_{2n} = 2^{-2n}\binom{2n}{n}\frac{\pi}{2}$. Similarly $I_{2n+1} = \frac{2n\dots 4\times 2}{2n+1\dots 3\times 1} I_1$

By $I_n = \frac{n-1}{n} I_{n-2}$: $\frac{I_n}{I_{n-2}} \to 1$ as $n\to\infty$. $= \frac{1}{2n+1}\left(2^{-2n}\binom{2n}{n}\right)^{-1}$.

Note I_n is decreasing in n. Hence $\frac{I_{2n}}{I_{2n+1}} \le \frac{I_{2n-1}}{I_{2n+1}} \xrightarrow[(n\to\infty)]{} 1$

and $\dfrac{I_{2n}}{I_{2n+1}} \geq \dfrac{I_{2n}}{I_{2n-2}} \to 1$ as $n \to \infty$, again by initial recurrence.

So $\dfrac{I_{2n}}{I_{2n+1}} \to 1$ as $n \to \infty$, so $\dfrac{2^{-2n}\binom{2n}{n}\frac{\pi}{2}}{\left(2^{-2n}\binom{2n}{n}\right)^{-1}\frac{1}{2n+1}} \to 1$

$$\Rightarrow \left(2^{-2n}\binom{2n}{n}\right)^2 \frac{\pi}{2}(2n+1) \to 1 \text{ as } n \to \infty$$

$$\Rightarrow 2^{-2n}\binom{2n}{n} \sim \frac{1}{\sqrt{\pi n}} \quad \Rightarrow \quad \underline{A = \sqrt{2\pi}} \text{ as } \overset{2^{-2n}\binom{2n}{n}}{\text{it's}} \text{ also asymptotic to } \frac{\sqrt{2}}{A\sqrt{n}}.$$

So $\underline{n! \sim n^n e^{-n}\sqrt{2\pi n}}$ as required. $\square$

Properties of probability measures

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

countable additivity: $(A_n)_{n\geq 1}$ disjoint $\Rightarrow \mathbb{P}(\cup A_n) = \sum \mathbb{P}(A_n)$

1) Countable subadditivity

Let $(A_n)_{n\geq 1}$ be a sequence of events in $\mathcal{F}$ ($A_n \in \mathcal{F}\ \forall n$)

Then $\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mathbb{P}(A_n)$

<u>Proof</u> Define $B_1 = A_1$, $B_n = A_n \setminus (A_1 \cup \ldots \cup A_{n-1})$ $\forall n \geq 2$

Then B_n is a disjoint sequence of events in $\mathcal{F}$. (check: $B_n \in \mathcal{F}\ \forall n$)

Also, $\bigcup_{n\geq 1} B_n = \bigcup_{n\geq 1} A_n$ so $\mathbb{P}(\cup A_n) = \mathbb{P}(\cup B_n) = \sum_n \mathbb{P}(B_n)$

But $B_n \subseteq A_n$ so $\mathbb{P}(B_n) \leq \mathbb{P}(A_n)\ \forall n$.

Hence $\mathbb{P}(\cup A_n) = \sum \mathbb{P}(B_n) \leq \sum_{n\geq 1} \mathbb{P}(A_n)$. $\square$

Probability - Lecture 3 26/1/20

Continuity of Probability Measures

Let $(A_n)_{n \geq 1}$ be an increasing sequence in $\mathcal{F}$: $\forall n\ A_n \in \mathcal{F}$ and $A_n \subseteq A_{n+1}$. Then $\mathbb{P}(A_n) \leq \mathbb{P}(A_{n+1})$. So $\mathbb{P}(A_n)$ converges as $n \to \infty$.

Claim $\lim_{n \to \infty} \mathbb{P}(A_n) = \mathbb{P}(\bigcup_n A_n)$

Proof Set $B_1 = A_1$ and $\forall n \geq 2$: $B_n = A_n \setminus (A_1 \cup \dots \cup A_{n-1})$. ← disjoint sequence

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

$$\text{So } \mathbb{P}(A_n) = \mathbb{P}\left(\bigcup_{k=1}^{n} B_k\right) = \sum_{k=1}^{n} \mathbb{P}(B_k) \xrightarrow[n \to \infty]{} \sum_{k=1}^{\infty} \mathbb{P}(B_k)$$

so we need $\sum_{k=1}^{\infty} \mathbb{P}(B_k) = \mathbb{P}(\bigcup_n A_n)$.

Since $\bigcup_{k=1}^{\infty} B_k = \bigcup_{k=1}^{\infty} A_k$, get $\mathbb{P}(\bigcup A_n) = \underbrace{\mathbb{P}(\bigcup B_n)}_{\text{disjoint}} = \sum_n \mathbb{P}(B_n)$. □

Similarly if (A_n) is decreasing in $\mathcal{F}$, i.e. $\forall n\ A_n \in \mathcal{F}$ and $A_n \supseteq A_{n+1}$ then $\mathbb{P}(A_n) \to \mathbb{P}(\bigcap_n A_n)$ as $n \to \infty$. (consider complement)

Inclusion-Exclusion

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

Let $C \in \mathcal{F}$. Then $\cancel{\mathbb{P}(A \cup B)}$

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

Claim Let $A_1, \dots, A_n \in \mathcal{F}$. Then

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

$$= \sum_{k=1}^{n} (-1)^{k+1} \sum \mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}).$$

Proof by induction For $n=2$ we know it holds.

Assume it holds for $n-1$ events. We will prove it for n events.

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

$$= \mathbb{P}(A_1 \cup \ldots \cup A_{n-1}) + \mathbb{P}(A_n) - \underbrace{\mathbb{P}((A_1 \cup \ldots \cup A_{n-1}) \cap A_n)}_{= \mathbb{P}((A_1 \cap A_n) \cup \ldots \cup (A_{n-1} \cap A_n))} \quad (*)$$

(apply $n=2$ case)

Set $B_i = A_i \cap A_n$. By inductive hypothesis:

$$\mathbb{P}(A_i \cup \ldots \cup A_{n-1}) = \sum_{k=1}^{n-1} (-1)^{k+1} \sum_{1 \leq i_1 < \ldots < i_k \leq n-1} \mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_k})$$

also $$\mathbb{P}(B_1 \cup \ldots \cup B_{n-1}) = \sum_{k=1}^{n-1} (-1)^{k+1} \sum_{1 \leq i_1 < \ldots < i_k \leq n-1} \mathbb{P}(B_{i_1} \cap \ldots \cap B_{i_k})$$

by inductive hypothesis again.

Plugging back into $(*)$ gives result. (check) $\square$

Let $(\Omega, \mathcal{F}, \mathbb{P})$ with $|\Omega| < \infty$ and $\mathbb{P}(A) = \frac{|A|}{|\Omega|}$ $\forall A \in \mathcal{F}$

Let $A_1, \ldots, A_n \in \mathcal{F}$. Then $|A_1 \cup A_2 \cup \ldots \cup A_n|$ is

$$\sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \leq i_1 < \ldots < i_k \leq n} |A_{i_1} \cap \ldots \cap A_{i_k}|$$

Bonferroni Inequalities

Truncating the sum in the inclusion-exclusion formula at the r^{th} term gives an overestimate if r is odd and an underestimate if r is even.

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{k=1}^{r} (-1)^{k+1} \sum_{1 \leq i_1 < i_2 < \ldots < i_k \leq n} \mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_k}) \quad \text{if } r \text{ is odd}$$

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \geq \sum_{k=1}^{r} (-1)^{k+1} \sum_{1 \leq i_1 < i_2 < \ldots < i_k \leq n} \mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_k}) \quad \text{if } r \text{ is even.}$$

Proof Induction: for $n=2$, $\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$.

Suppose true for $n-1$ events.

Suppose r is odd. Then $\qquad B_i = A_i \cap A_n$

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

Since r is odd, apply inductive hypothesis to $\mathbb{P}(A_1 \cup \dots \cup A_{n-1})$ to get

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

$r-1$ is even: apply inductive hyp. to $\mathbb{P}(B_1 \cup \dots \cup B_{n-1})$:

$$\mathbb{P}(B_1 \cup \dots \cup B_{n-1}) \geq \sum_{k=1}^{r-1} (-1)^{k+1} \sum_{1 \leq i_1 < \dots < i_k \leq n-1} \mathbb{P}(B_{i_1} \cap \dots \cap B_{i_k})$$

~~Supp~~ Sub both upper bounds in $(*)$ to get an overestimate.

Exactly the same argument holds for r even (underestimate). $\square$

Counting

1) Number of surjections $f: \{1, \dots, n\} \to \{1, \dots, m\}$

Let $\Omega = \{f: \{1, \dots, n\} \to \{1, \dots, m\}\}$ and let

$A = \{f \in \Omega : f \text{ is surjective}\}$. What is $|A|$?

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

"elements not hit by f" i.e. A_i is set of functions that don't hit i

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

$$|A| = \underbrace{|\Omega|}_{m^n} - |A_1 \cup \dots \cup A_m|$$

by Inclusion-Exclusion

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

$= (m-k)^n$ as for each $1, \dots, n$ must hit something in $1, \dots, m$ not in $1, \dots, k$.

$$= m^n - \sum_{k=1}^{m} (-1)^{k+1} \binom{m}{k} (m-k)^n$$

$$= \sum_{k=0}^{m} (-1)^k \binom{m}{k} (m-k)^n. \qquad \square$$

Probability - Lecture 4 26/1/21

Counting derangements (permutations with no fixed points)

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

$A = \{\text{derangements}\} = \{f \in \Omega : f(i) \neq i \ \forall i = 1, \dots, n\}$

Define $A_i = \{f \in \Omega : f(i) = i\}$. Then $A = A_1^c \cap A_2^c \cap \dots \cap A_n^c = \left(\bigcup_{i=1}^n A_i\right)^c$

so $\mathbb{P}(A) = 1 - \mathbb{P}\left(\bigcup_{i=1}^n A_i\right)$. Use inclusion-exclusion:

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

$\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}) = \frac{(n-k)!}{n!}$ — fixing $f(1), f(2), \dots, f(k)$

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

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

As $n \to \infty$, this converges to $\frac{1}{e}$. $(\approx 0.3678\dots)$

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

Definition Let $A, B \in \mathcal{F}$. They are independent ($A \perp B$) if
$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B).$$

Similarly (A_n) independent if $\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}) = \prod_{j=1}^k \mathbb{P}(A_{i_j})$

Remark Pairwise independence $\not\Rightarrow$ independence.

Toss fair coin twice: $\Omega = \{(0,0), (0,1), (1,0), (1,1)\}$ $\mathbb{P}(\omega) = \frac{1}{4} \ \forall \omega \in \Omega$

Define $A = \{(0,0), (0,1)\}$ $B = \{(0,0), (1,0)\}$ $C = \{(1,0), (0,1)\}$

$\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = \frac{1}{2}$

$\mathbb{P}(A \cap B) = \mathbb{P}(\{(0,0)\}) = \frac{1}{4} = \frac{1}{2} \times \frac{1}{2} \Rightarrow A \perp B$

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

but $\mathbb{P}(A \cap B \cap C) = 0 \neq \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2}$.

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

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

$= \mathbb{P}(A) - \mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(A)\left(1 - \mathbb{P}(B)\right) = \mathbb{P}(A)\mathbb{P}(B^c)$. $\square$

Conditional Probability

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

Let $A \in \mathcal{F}$. Define conditional probability A given B $(A|B)$ by

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

If A and B are independent then $\mathbb{P}(A|B) = \mathbb{P}(A)$.

Suppose (A_n) is a disjoint sequence in $\mathcal{F}$. Then

$$\underline{\mathbb{P}(\cup A_n | B) = \sum_n \left(\mathbb{P}(A_n|B)\right)}$$ countable additivity for cond. prob.

Proof $\mathbb{P}(\cup A_n | B) = \dfrac{\mathbb{P}((\cup A_n) \cap B)}{\mathbb{P}(B)} = \dfrac{\mathbb{P}\left(\bigcup_n (A_n \cap B)\right)}{\mathbb{P}(B)}$

by countable additivity: $= \displaystyle\sum_n \frac{\mathbb{P}(A_n \cap B)}{\mathbb{P}(B)} = \sum_n \mathbb{P}(A_n|B)$. $\square$

Law of Total Probability

Suppose $(B_n)_{n \in \mathbb{N}}$ is a <u>disjoint</u> collection in $\mathcal{F}$ and $\cup B_n = \Omega$ and $\forall n$ $\mathbb{P}(B_n) > 0$.

Let $A \in \mathcal{F}$. Then $\underline{\mathbb{P}(A) = \sum_n \mathbb{P}(A|B_n) \cdot \mathbb{P}(B_n)}$

Proof $\mathbb{P}(A) = \mathbb{P}(A \cap \Omega) = \mathbb{P}\left(A \cap \left(\bigcup_n B_n\right)\right) = \mathbb{P}\left(\bigcup_n \underbrace{(A \cap B_n)}_{\text{disjoint}}\right)$

$\underbrace{=}_{\text{(countable additivity)}} \displaystyle\sum_n \mathbb{P}(A \cap B_n) = \underline{\sum_n \mathbb{P}(A|B_n) \cdot \mathbb{P}(B_n)}$. $\square$

Bayes' formula

Suppose that (B_n) are disjoint events, $\bigcup B_n = \Omega$, $\mathbb{P}(B_n) > 0 \ \forall n$.

Then
$$\mathbb{P}(B_n \mid A) = \frac{\mathbb{P}(A \mid B_n) \cdot \mathbb{P}(B_n)}{\sum_k \mathbb{P}(A \mid B_k)\, \mathbb{P}(B_k)}$$

Proof
$$\mathbb{P}(B_n \mid A) = \frac{\mathbb{P}(B_n \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A \mid B_n) \cdot \mathbb{P}(B_n)}{\mathbb{P}(A)}$$

$$= \frac{\mathbb{P}(A \mid B_n) \cdot \mathbb{P}(B_n)}{\sum_k \mathbb{P}(A \mid B_k)\, \mathbb{P}(B_k)}$$ by law of total probability.

This formula is the basis of Bayesian statistics.

We know the probabilities of the events (B_k) and we have a model which gives us $\mathbb{P}(A \mid B_n)$. Bayes' formula tells us how to find the posterior probabilities of B_n given A.

Probability - Lecture 5 31/1/21

Let (B_n) be a <u>partition</u> of Ω, i.e. (B_n) disjoint, $\bigcup B_n = \Omega$

$\forall A \in \mathcal{F}$: $$\mathbb{P}(B_n \mid A) = \frac{\mathbb{P}(A \mid B_n)\,\mathbb{P}(B_n)}{\sum_k \left(\mathbb{P}(A \mid B_k)\,\mathbb{P}(B_k)\right)}$$ Bayes' formula

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

Suppose condition A affects 0.1% of the population. We have a test which is positive for 98% of affected population and 1% of those unaffected. Pick individual at random - what is probability they suffer from A given they tested positive?

Define $A = \{$individual suffers from A$\}$, $P = \{$individual tests positive$\}$

Want $\mathbb{P}(A \mid P)$.

Have $\mathbb{P}(P \mid A) = 0.98$, $\mathbb{P}(P \mid A^c) = 0.01$, $\mathbb{P}(A) = 0.001$

$$\mathbb{P}(A \mid P) = \frac{\mathbb{P}(P \mid A) \cdot \mathbb{P}(A)}{\mathbb{P}(P \mid A)\,\mathbb{P}(A) + \mathbb{P}(P \mid A^c) \cdot \mathbb{P}(A^c)} = \frac{0.98 \times 0.001}{0.98 \times 0.001 + 0.01 \times 0.999} \approx 0.09$$

so $\mathbb{P}(A \mid P) \approx 0.09$. Given that they tested positive, the probability they have the condition is <u>low</u> - $\approx 9\%$.

This is because $\mathbb{P}(P \mid A^c) \gg \mathbb{P}(A)$.

<u>Example 2</u> Extra knowledge gives surprising results

3 statements:
(a) I have 2 children one of whom is a boy
(b) I have 2 children and the eldest is a boy
(c) I have 2 children one of whom is a boy born on a Thursday

$\mathbb{P}($have 2 boys $\mid a/b/c)$?

Define $BG = \{$elder boy, younger girl$\}$, $GB = \{$elder girl, younger boy$\}$
$BB = \{$both boys$\}$, $GG = \{$both girls$\}$

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

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

(c) define $GT = \{$ elder = girl, younger = boy born on Thursday $\}$

$TN = \{$ elder = boy born on Thursday, younger = boy not born on Thursday $\}$

etc. TT, TG, NT

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

no NN - we know 1 is born on Thursday

numerator $\frac{13}{49 \times 4}$

$$= \frac{\mathbb{P}(TT \cup TN \cup NT)}{\mathbb{P}(TT \cup TN \cup NT \cup GT \cup TG)} = \frac{\frac{1}{2}\cdot\frac{1}{7}\cdot\frac{1}{2}\cdot\frac{1}{7} + \frac{1}{2}\cdot\frac{1}{7}\cdot\frac{1}{2}\cdot\frac{6}{7} + \frac{1}{2}\cdot\frac{6}{7}\cdot\frac{1}{2}\cdot\frac{1}{7}}{\frac{13}{49\times 4} + \frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{7} + \frac{1}{2}\cdot\frac{1}{7}\cdot\frac{1}{2}}$$

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

Simpson's Paradox

accepted, rejected

All	✓	X	% ✓
State	25	25	50%
Indep	28	22	56%

Men	✓	X	% ✓
State	15	22	41%
Indep	5	8	38%

Women	✓	X	% ✓
State	10	3	77%
Indep	23	14	62%

Confounding in statistics. Let $A = \{$ individual admitted $\}$,
$B = \{$ ind. is a man $\}$, $C = \{$ ind. from state school $\}$

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

However $\mathbb{P}(A \mid C^c) > \mathbb{P}(A \mid C)$ i.e. state > indep success rate for both just men and just women

but indep > state success rate overall

$$\mathbb{P}(A \mid C) = \mathbb{P}(A \cap B \mid C) + \mathbb{P}(A \cap B^c \mid C) = \frac{\mathbb{P}(A \cap B \cap C)}{\mathbb{P}(C)} + \frac{\mathbb{P}(A \cap B^c \cap C)}{\mathbb{P}(C)}$$

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

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

but for this we also have $\mathbb{P}(B \mid C) \neq \mathbb{P}(B \mid C^c)$ $(*)$; if they were equal then we'd get
(not true here)

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

$\Rightarrow \mathbb{P}(A \mid C) > \mathbb{P}(A \mid C^c)$ as might seem intuitive. But $(*)$ does not hold, so we can't conclude this.

Probability - lecture 6 2/2/21

Discrete Probability Distributions

$(\Omega, \mathcal{F}, \mathbb{P})$ Ω finite/countable $\Omega = \{\omega_1, \omega_2, \dots\}$ $\mathcal{F} = 2^{\Omega}$

If we know $\mathbb{P}(\omega_i)$ $\forall i$ then this determines $\mathbb{P}$. (countable additivity)

We write $p_i = \mathbb{P}(\{\omega_i\})$ - call it a discrete prob. distribution.

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

1) **Bernoulli distribution**

Model outcome of coin toss. $\Omega = \{\overset{T}{0}, \overset{H}{1}\}$ $p_1 = \mathbb{P}(\{1\}) = p$

$p_2 = \mathbb{P}(\{0\}) = 1-p$

2) **Binomial distribution** $B(N, p)$ $N \in \mathbb{N}$, $p \in [0,1]$

Toss p-coin N times (independently) $\Omega = \{0, 1, \dots, N\}$

$\mathbb{P}(\text{see } k \text{ heads}) = \binom{N}{k} p^k (1-p)^{N-k}$

3) **Multinomial distribution** $M(N, p_1, \dots, p_k)$ $N \in \mathbb{N}$, $p_1, \dots, p_k \geq 0$, $\sum_{i=1}^{k} p_i = 1$

$\underset{1}{\sqcup}\ \underset{2}{\sqcup} \dots \underset{k}{\sqcup}$ k boxes, N balls, $\mathbb{P}(\text{pick box } i) = p_i$, independent

$\Omega = \{(n_1, \dots, n_k) \in \mathbb{N}^k : \sum_{i=1}^{k} n_i = N\}$ (ordered partitions)

$\mathbb{P}(n_1 \text{ balls in } 1, \dots, n_k \text{ balls in } k) = p_1^{n_1} p_2^{n_2} \dots p_k^{n_k} \cdot \binom{N}{n_1, \dots, n_k}$

4) **Geometric distribution** Toss p-coin until first H appears $\Omega = \mathbb{N}$

$p_k = \mathbb{P}(k \text{ tosses}) = (1-p)^{k-1} \cdot p$

5) **Poisson distribution** Model number of occurrences of event in given time interval

e.g. # customers entering shop in a day: $\Omega = \mathbb{N}$ $\lambda > 0$ (parameter)

$$p_k = \frac{e^{-\lambda} \lambda^k}{k!} \quad \forall k \in \Omega$$

Suppose customers arrive in shop during $[0,1]$. Discretise $[0,1]$ - subdivide into N intervals $\left[\frac{i-1}{N}, \frac{i}{N}\right]$. In each interval a customer arrives with probability p ($1-p$: nobody arrives).

$\mathbb{P}(k \text{ customers arrived}) = \binom{N}{k} p^k (1-p)^{N-k}$. Take $p = \frac{\lambda}{N}$, $\lambda > 0$.

$$\binom{N}{k} p^k (1-p)^{N-k} = \frac{N!}{k!(N-k)!}\left(\frac{\lambda}{N}\right)^k \cdot \left(1-\frac{\lambda}{N}\right)^{N-k} = \frac{\lambda^k}{k!}\frac{N!}{N^k(N-k)!}\left(1-\frac{\lambda}{N}\right)^{N-k}$$

fix k, let $n \to \infty$: $\quad \mathbb{P}(k) \to \dfrac{e^{-\lambda}\lambda^k}{k!}$. (limit of binomial distribution).

<u>Random variables</u>

$(\Omega, \mathcal{F}, \mathbb{P})$ <u>Random var X</u> is function $X : \Omega \to \mathbb{R}$ s.t. $\forall x \in \mathbb{R}$,

$\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F}$. $\quad (A \subseteq \mathbb{R} : \{X \in A\} = \{\omega : X(\omega) \in A\})$.

Given $A \in \mathcal{F}$, define indicator of A : $\quad 1_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{otherwise} \end{cases}$ $\quad (*)$

Since $A \in \mathcal{F}$, 1_A is a random variable.

Suppose X is a random variable. Define prob. distribution function (pdf) of X by <u>$F_X(x) = \mathbb{P}(X \leq x)$</u> $\quad F_X : \mathbb{R} \to [0,1]$

<u>Definition</u> $(X_1, \dots, X_n)$ is a random variable in $\mathbb{R}^n$ if

$(X_1, \dots, X_n) : \Omega \to \mathbb{R}^n$ and $\forall x_1, \dots, x_n \in \mathbb{R}$, we have

$\{X_1 \leq x_1, \dots, X_n \leq x_n\} \in \mathcal{F}$.
(notation $(*)$)

This defn is equivalent to saying that $X_1, \dots, X_n$ are all random variables in $\mathbb{R}$.

$$\{X_1 \leq x_1, \dots, X_n \leq x_n\} = \underset{\text{in } \mathcal{F}}{\{X_1 \leq x_1\}} \cap \dots \cap \underset{\text{in } \mathcal{F}}{\{X_n \leq x_n\}} \in \mathcal{F}.$$

Probability - Lecture 7 5/2/21

$(\Omega, \mathcal{F}, \mathbb{P})$ $X : \Omega \to \mathbb{R}$ s.t. $\forall x \in \mathbb{R}$, $\{X \leq x\} \in \mathcal{F}$

(X: random var)

Definition A random var X is discrete if it takes values in a countable set.

Set S : $\forall x \in S$, write $p_x = \mathbb{P}(X = x) = \mathbb{P}(\{\omega : X(\omega) = x\})$

Call $(p_x)_{x \in S}$ probability mass function of X (pmf) or distribution of X.

If (p_x) is Bernoulli: say X is a Bernoulli random var / Bernoulli distribution etc.

Definition Let $X_1, \dots, X_n$ be discrete RVs taking values in $S_1, \dots, S_n$.

Say $X_1, \dots, X_n$ are independent if

$$\mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = \mathbb{P}(X_1 = x_1) \dots \mathbb{P}(X_n = x_n) \quad \forall x_i \in S_i$$

Example Toss a p-biased coin N times independently.

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

$$\omega \in \Omega \quad p_\omega = \prod_{k=1}^{N} p^{\omega_k}(1-p)^{1-\omega_k} \qquad \omega = (\omega_1, \dots, \omega_N)$$

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

Then X_k gives outcome of k^{th} toss and is a discrete random var

$X_k : \Omega \to \{0,1\}$ $\mathbb{P}(X_k = 1) = \mathbb{P}(\omega_k = 1) = p$, $\mathbb{P}(X_k = 0) = \mathbb{P}(\omega_k = 0) = 1 - p$

Bernoulli distribution with parameter p.

Claim $X_1, \dots, X_n$ independent RVs

Proof Let $x_1, \dots, x_N \in \{0,1\}$, then

$$\mathbb{P}(X_1 = x_1, \dots, X_N = x_N) = \mathbb{P}(\omega = (x_1, \dots, x_N)) = p_{(x_1, x_2, \dots, x_n)} = \prod_{k=1}^{N} p^{x_k}(1-p)^{1-x_k}$$
$$= \prod_{k=1}^{N} \mathbb{P}(X_k = x_k)$$

$\square$

Define $S_N(\omega) = X_1(\omega) + \dots + X_N(\omega) = \#$ of H in N tosses

$S_N : \Omega \to \{0, \dots, N\}$ and $\mathbb{P}(S_N = k) = \binom{N}{k} p^k (1-p)^{N-k}$

so S_N has the binomial distribution of parameters N and p.

Expectation $(\Omega, \mathcal{F}, \mathbb{P})$ Assume Ω is finite or countable.

Let $X : \Omega \to \mathbb{R}$ be a DRV. X is non-negative if $X \geq 0$.

Define the expectation of $X \geq 0$ by

$$\mathbb{E}(X) = \sum_{\omega} X(\omega)\, \mathbb{P}(\{\omega\})$$

$$\Omega_X = \{ X(\omega) : \omega \in \Omega \} \quad \text{so} \quad \Omega = \bigcup_{x \in \Omega_X} \{X = x\}$$

$$\mathbb{E}(X) = \sum_{x \in \Omega_X} \sum_{\omega \in \{X = x\}} x \cdot \mathbb{P}(\{\omega\}) = \underline{\sum_{x \in \Omega_X} x\, \mathbb{P}(X = x)} = \sum_{x \in \Omega_X} x p_x$$

Examples $X \sim \mathrm{Bin}(N, p)$

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

$$\mathbb{E}(X) = \sum_{k=0}^{N} k\, \mathbb{P}(X = k) = \sum_{k=0}^{N} k \binom{N}{k} p^k (1-p)^{N-k}$$

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

$$= Np \sum_{k=0}^{N-1} \binom{N-1}{k} p^k (1-p)^{N-1-k} = Np\, (p + 1 - p)^{N-1} = \underline{Np}$$

Example 2 $X \sim \mathrm{Poi}(\lambda)$

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

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

Let X be a general DRV. Define $X_+ = \max(X, 0)$,

$X_- = \max(-X, 0)$. Then $X = X_+ - X_-$ and $|X| = X_+ + X_-$

We can define $\mathbb{E}(X_+)$, $\mathbb{E}(X_-)$ since both are non-negative.

If at least one of $\mathbb{E}(X_+)$, $\mathbb{E}(X_-)$ is finite: define

$\mathbb{E}(X) = \mathbb{E}(X_+) - \mathbb{E}(X_-)$. If both are ∞ then $\mathbb{E}(X)$ is undefined.

When we write $\mathbb{E}(X)$: assumed to be well-defined

If $\mathbb{E}(|X|) < \infty$: X is <u>integrable</u>.

When $\mathbb{E}(X)$ is well defined, have again $\sum_{x \in \Omega_X} x \, \mathbb{P}(X = x) = \mathbb{E}(X)$.

<u>Properties</u>

1) If $X \geq 0$ then $\mathbb{E}(X) \geq 0$.

2) If $X \geq 0$ and $\mathbb{E}(X) = 0$ then $\mathbb{P}(X = 0) = 1$

3) If $c \in \mathbb{R}$ then $\mathbb{E}(cX) = c\,\mathbb{E}(X)$ and $\mathbb{E}(c + X) = c + \mathbb{E}(X)$

4) If X and Y are 2 RVs: $\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$

5) Let $c_1, \dots, c_n \in \mathbb{R}$ and $X_1, \dots, X_n$ be RVs. Then (all integrable)

$$\mathbb{E}\left(\sum_{i=1}^{n} c_i X_i\right) = \sum_{i=1}^{n} c_i \, \mathbb{E}(X_i)$$

Probability - Lecture 8 6/2/21

More properties of expectation:

Suppose $X_1, X_2, \ldots$ are non-negative RVs. Then $\mathbb{E}\left(\sum_n X_n\right) = \sum_n \mathbb{E}(X_n)$

Proof (Ω countable) - countable additivity

$$\mathbb{E}\left(\sum_n X_n\right) = \sum_\omega \sum_n X_n(\omega)\,\mathbb{P}(\{\omega\}) \underset{\text{all} \geq 0}{=} \sum_n \sum_\omega X_n(\omega)\,\mathbb{P}(\{\omega\}) = \sum_n \mathbb{E}(X_n) \quad \square$$

- Now if $X = 1(A)$, $A \in \mathcal{F}$ (indicator), then $\mathbb{E}(X) = \mathbb{P}(A)$.

- If $g: \mathbb{R} \to \mathbb{R}$: define $g(X)$ to be the RV $g(X)(\omega) = g(X(\omega))$

then $\mathbb{E}(g(X)) = \sum_{x \in \Omega_X} g(x)\,\mathbb{P}(X = x)$

Proof Set $Y = g(X)$. Then $\mathbb{E}(Y) = \sum_{y \in \Omega_y} y\,\mathbb{P}(Y = y)$

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

so $$\mathbb{E}(Y) = \sum_{y \in \Omega_y} y\,\mathbb{P}(X \in g^{-1}(\{y\})) = \sum_{y \in \Omega_y} y \sum_{x \in g^{-1}(\{y\})} \mathbb{P}(X = x)$$
$$= \sum_{y \in \Omega_y} \sum_{x \in g^{-1}(\{y\})} g(x)\,\mathbb{P}(X = x) = \sum_{x \in \Omega_X} g(x)\,\mathbb{P}(X = x). \quad \square$$

If $X \geq 0$ and takes integer values, then
$$\mathbb{E}(X) = \sum_{k=1}^{\infty} \mathbb{P}(X \geq k) = \sum_{k=0}^{\infty} \mathbb{P}(X > k)$$

Proof Since $X \geq 0$, integer valued:
$$X = \sum_{k=1}^{\infty} 1(X \geq k) = \sum_{k=0}^{\infty} 1(X > k)$$

since e.g. if $X(\omega) = 4$ then $1(X \geq k) = 1$ for $k = 1,2,3,4$ so sum is $1+1+1+1 = 4$

Take $\mathbb{E}$ and use $\mathbb{E}(1(A)) = \mathbb{P}(A)$ and countable additivity.

Another proof of inclusion-exclusion:

Properties of indicator RVs: $1(A^c) = 1 - 1(A)$, $1(A \cap B) = 1(A)1(B)$

$1(A \cup B) = 1 - (1 - 1(A))(1 - 1(B))$

more generally: $1(A_1 \cup \ldots \cup A_n) = 1 - \prod_{i=1}^{n}(1 - 1(A_i)) = \sum_{i=1}^{n} 1(A_i) - \sum_{i_1 < i_2} 1(A_{i_1} \cap A_{i_2}) + \ldots$
$+ (-1)^{n+1}(1(A_1 \cap \ldots \cap A_n))$

Take $\mathbb{E}$ of both sides:

$$\mathbb{P}(A_1 \cup \ldots \cup A_n) = \sum_{i=1}^{n} \mathbb{P}(A_i) - \sum_{i_1 < i_2} \mathbb{P}(A_{i_1} \cap A_{i_2}) + \ldots + (-1)^{n+1} \mathbb{P}(A_2 \cap \ldots \cap A_n)$$

<u>Terminology</u> X a RV, $r \in \mathbb{N}$: call $\mathbb{E}(X^r)$ r^{th} moment of X.

<u>Definition</u> (Variance) $\mathrm{Var}(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right)$

$\mathrm{Var}(X) \geq 0$: $\mathrm{Var}(X) = 0 \Rightarrow \mathbb{P}(X = \mathbb{E}(X)) = 1$

$\mathrm{Var}(cX) = \underline{c^2}\,\mathrm{Var}(X) \qquad \mathrm{Var}(X + c) = \mathrm{Var}(X)$

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

$\boxed{\text{Properties of variance}}$

<u>Proof</u> $\mathrm{Var}(X) = \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2 - 2X\mathbb{E}(X) + (\mathbb{E}(X))^2)$

$= \mathbb{E}(X^2) - 2\mathbb{E}(X)\mathbb{E}(X) + \mathbb{E}(X)^2 = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 \quad \square$

$\mathrm{Var}(X) = \min_{c \in \mathbb{R}} \mathbb{E}((X-c)^2)$ achieved for $c = \mathbb{E}(X)$

<u>Proof</u> $f(c) = \mathbb{E}((X-c)^2) = \mathbb{E}(X^2) - 2c\,\mathbb{E}(X) + c^2$, minimised when $c = \mathbb{E}(X)$.

<u>Examples</u> (1) $X \sim \mathrm{Bin}(n, p)$, $\mathbb{E}(X) = np$

$\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$ and we have

$$\mathbb{E}(X(X-1)) = \sum_{k=2}^{n} k(k-1)\binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=2}^{n} \frac{k(k-1)\,n!\,p^k(1-p)^{n-k}}{(k-2)!(k-1)k((n-2)-(k-2))!}$$

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

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

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

(2) $X \sim \mathrm{Poi}(\lambda)$, $\lambda > 0$, $\mathbb{E}(X) = \lambda$ so $\mathrm{Var}(X) = \mathbb{E}(X^2) - \lambda^2$

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

so $\mathrm{Var}(X) = \lambda^2 + \mathbb{E}(X) - \lambda^2 = \lambda^2 + \lambda - \lambda^2 = \underline{\lambda}$.

Probability - Lecture 9 6/2/21

Definition Let X and Y be random variables. Their covariance is
$$\text{Cov}(X,Y) = \mathbb{E}\big((X-\mathbb{E}(X))(Y-\mathbb{E}(Y))\big)$$
Measures how dependent X and Y are

Properties 1) $\text{Cov}(X,Y) = \text{Cov}(Y,X)$

2) $\text{Cov}(X,X) = \text{Var}(X)$

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

Proof Expand $(X-\mathbb{E}(X))(Y-\mathbb{E}(Y))$, use properties of $\mathbb{E}$

4) $c \in \mathbb{R}$: $\text{Cov}(cX,Y) = c\,\text{Cov}(X,Y)$ and $\text{Cov}(c+X,Y) = \text{Cov}(X,Y)$

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

Proof $\text{Var}(X+Y) = \mathbb{E}((X+Y) - \mathbb{E}(X) - \mathbb{E}(Y))^2)$
$$= \mathbb{E}\left(((X-\mathbb{E}(X)) + (Y-\mathbb{E}(Y)))^2\right)$$
$$= \mathbb{E}((X-\mathbb{E}(X))^2) + \mathbb{E}((Y-\mathbb{E}(Y))^2) + 2\,\mathbb{E}\big((X-\mathbb{E}(X))(Y-\mathbb{E}(Y))\big)$$
$$= \text{Var}(X) + \text{Var}(Y) + 2\,\text{Cov}(X,Y). \qquad \square$$

6) $\forall c \in \mathbb{R}$ $\text{Cov}(c,X) = 0$

7) X,Y,Z RVs: $\text{Cov}(X+Y,Z) = \text{Cov}(X,Z) + \text{Cov}(Y,Z)$

Generally for $c_1, c_2, \dots, c_n, d_1, \dots, d_n \in \mathbb{R}$ and RVs X_i, Y_i $(i=1,\dots,n)$, have
$$\text{Cov}\left(\sum_{i=1}^n c_i X_i, \sum_{i=1}^n d_i Y_i\right) = \sum_{i=1}^n \sum_{j=1}^n c_i d_j \,\text{Cov}(X_i, Y_j)$$
In particular, $\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i) + \sum_{i \neq j} \text{Cov}(X_i, X_j)$

Claim X,Y independent RV, $f,g: \mathbb{R} \to \mathbb{R}^+$

Then $\mathbb{E}(f(X)g(Y)) = \mathbb{E}(f(X))\,\mathbb{E}(g(Y))$

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

indep: $= \sum_{x,y} f(x)g(y)\,\mathbb{P}(X=x)\,\mathbb{P}(Y=y) = \sum_x f(x)\,\mathbb{P}(X=x) \sum_y g(y)\,\mathbb{P}(Y=y)$
$$= \mathbb{E}(f(X))\,\mathbb{E}(g(Y)) \qquad \square$$

X, Y indep:

$\text{Cov}(X,Y) = 0$ since $\text{Cov}(X,Y) = \mathbb{E}\big((X-\mathbb{E}(X))(Y-\mathbb{E}(Y))\big) = 0$

so if X, Y indep, then $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$

But $\text{Cov}(X,Y) = 0 \not\Rightarrow$ independence:

<u>Example</u> Let X_1, X_2, X_3 be indep. $\text{Ber}\left(\frac{1}{2}\right)$

define $Y_1 = 2X_1 - 1$, $Y_2 = 2X_2 - 1$, $Z_1 = X_3 Y_1$, $Z_2 = X_3 Y_2$

$\mathbb{E}(Y_1) = \mathbb{E}(Y_2) = \mathbb{E}(Z_2) = \mathbb{E}(Z_1) = 0$

and $\text{Cov}(Z_1, Z_2) = \mathbb{E}(Z_1 \cdot Z_2) = \mathbb{E}(X_3^2 Y_1 Y_2) \overset{\text{indep}}{=} 0$

But Z_1, Z_2 are <u>not</u> independent.

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

but $\mathbb{P}(Z_1 = 0)\,\mathbb{P}(Z_2 = 0) = \mathbb{P}(X_3 = 0)^2 = \frac{1}{4}$ $\qquad \underline{\frac{1}{2} \neq \frac{1}{4}}$

<u>Inequalities</u>

1) <u>Markov's inequality</u> Let $X \geq 0$ be a RV. Then $\forall a > 0$,

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

<u>Proof</u> Observe that $X \geq a \mathbb{1}(X \geq a)$ (check: $\mathbb{1}(X \geq a)$ is 1 or 0)

Take $\mathbb{E}$: get $\mathbb{E}(X) \geq \mathbb{E}(a \mathbb{1}(X \geq a)) = a\mathbb{P}(X \geq a)$

so $\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a}$. $\square$

2) <u>Chebyshev's inequality</u> Let X be a RV with $\mathbb{E}(X) < \infty$.

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

<u>Proof</u> $\mathbb{P}(|X - \mathbb{E}(X)| \geq a) = \mathbb{P}(|X - \mathbb{E}(X)|^2 \geq a^2) \underset{\text{Markov}}{\leq} \frac{\mathbb{E}((X - \mathbb{E}(X))^2)}{a^2}$

$= \frac{\text{Var}(X)}{a^2}$ $\square$

3) <u>Cauchy - Schwarz inequality</u>

Let X and Y be RVs. Then

$$\mathbb{E}(|XY|) \leqslant \sqrt{\mathbb{E}(X^2)\mathbb{E}(Y^2)}$$

<u>Proof</u> suffices to prove for X, Y with $\mathbb{E}(X^2), \mathbb{E}(Y^2) < \infty$.

Also enough to prove for $X, Y \geqslant 0$:

$$XY \leqslant \tfrac{1}{2}(X^2+Y^2) \Rightarrow \mathbb{E}(XY) \leqslant \tfrac{1}{2}\left(\mathbb{E}(X^2) + \mathbb{E}(Y^2)\right) < \infty$$

so $\mathbb{E}(XY) < \infty$.

Assume $\mathbb{E}(X^2) > 0$, $\mathbb{E}(Y^2) > 0$ (else result is trivial)

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

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

$$\Rightarrow \underbrace{\mathbb{E}(X^2) - 2t\mathbb{E}(XY) + t^2\mathbb{E}(Y^2)}_{f(t)} \geqslant 0$$

Minimise f - gives $\bar{t} = \dfrac{\mathbb{E}(XY)}{\mathbb{E}(Y^2)} \Rightarrow f$ is minimised.

Then $f(\bar{t}) \geqslant 0 \Rightarrow \mathbb{E}(X^2) - \dfrac{2(\mathbb{E}(XY))^2}{\mathbb{E}(Y^2)} + \dfrac{(\mathbb{E}(XY))^2}{\mathbb{E}(Y^2)} \geqslant 0$

$$\Rightarrow (\mathbb{E}(XY))^2 \leqslant \mathbb{E}(X^2)\,\mathbb{E}(Y^2). \qquad \square$$

Probability - Lecture 10 $\qquad$ 11/2/21

X, Y RVs: then $\mathbb{E}(|XY|) \leq \sqrt{\mathbb{E}(X^2)\mathbb{E}(Y^2)}$

Cases of equality:

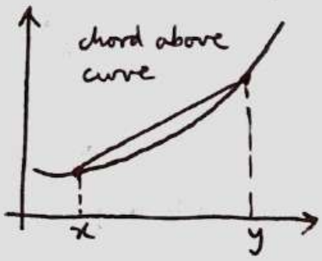
Recall $\mathbb{E}((X-tY)^2) \geq 0$ and $(X-tY)^2$ is minimised for $t = \frac{\mathbb{E}(XY)}{\mathbb{E}(Y^2)}$ where it's 0.

$\mathbb{E}((X-tY)^2) = 0 \implies \mathbb{P}(X = tY) = 1$

Jensen's inequality

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

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



Statement f convex function, X a RV. Then

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

Claim Let f be convex ($f: \mathbb{R} \to \mathbb{R}$). Then f is the supremum of all the lines lying below it. In other words

$\forall m \in \mathbb{R} \quad \exists a, b \in \mathbb{R}$ s.t. $f(m) = am + b$, $f(x) \geq ax + b \ \forall x$

Proof Let $m \in \mathbb{R}$. Let $x < m < y$. Then $m = tx + (1-t)y$ for some $t \in [0,1]$. By convexity $\underset{\downarrow}{f(m)} \leq tf(x) + (1-t)f(y)$

$tf(m) + (1-t)f(m)$

$\implies t(f(m) - f(x)) \leq (1-t)(f(y) - f(m))$

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

So $\exists a \in \mathbb{R}$ with $a = \sup_{x < m} \frac{f(m) - f(x)}{m - x}$ s.t.

$$\frac{f(m) - f(x)}{m - x} \leq a \leq \frac{f(y) - f(m)}{y - m} \quad \forall x < m < y$$

$\implies f(x) \geq a(x - m) + f(m) \quad \forall x.$ $\square$

Proof of ineq Set $m = \mathbb{E}(X)$ then from claim $\exists a, b \in \mathbb{R}$ s.t.

$f(m) = am + b \iff f(\mathbb{E}(X)) = a\mathbb{E}(X) + b$ and $\forall x \ f(x) \geq ax + b$ by claim. Apply to X to get $f(X) \geq aX + b$

$\implies \mathbb{E}(f(X)) \geq a\mathbb{E}(X) + b = f(\mathbb{E}(X)).$ $\square$

Cases of equality

X RV, f convex satisfying: if $m = \mathbb{E}(X)$ then $\exists a, b \in \mathbb{R}$ with $f(m) = am+b$ and $f(x) > ax+b \quad \forall x \neq m$.

What is the condition on X for equality? Want $\mathbb{E}(f(X)) = f(\mathbb{E}(X))$

$\mathbb{E}(X) = m$, ~~f(x)~~ $f(m) = am+b$, $f(x) > ax+b \quad \forall x \neq m$

Consider $f(X) \geq aX+b$. Then $f(X) - (aX+b) \geq 0 \Rightarrow \mathbb{E}\big(f(X) - (aX+b)\big) \geq 0$

We assumed $\mathbb{E}(f(X)) = f(\mathbb{E}(X)) \Rightarrow \mathbb{E}\big(f(X) - (aX+b)\big) = 0$ but $f(X) \geq aX+b$ forcing $f(X) = aX+b$ so $\underline{X = m}$ with prob. 1.

AM/GM Let ~~xxx~~ f be convex, $x_1, \dots, x_n \in \mathbb{R}$. Then

$$\frac{1}{n}\sum_{k=1}^{n} f(x_k) \geq f\left(\frac{1}{n}\sum_{k=1}^{n} x_k\right) \qquad \mathbb{E}(f(X)) \geq f(\mathbb{E}(X))$$

Define X a RV with $\mathbb{P}(X = x_i) = \frac{1}{n}$

Then $\mathbb{E}(f(X)) = \sum_{k=1}^{n} f(x_k)\cdot\frac{1}{n}$, $f(\mathbb{E}(X)) = f\left(\sum_{k=1}^{n} x_k \cdot \frac{1}{n}\right)$

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

Let $f(x) = -\log x$: convex. Then $-\frac{1}{n}\sum_{k=1}^{n}\log x_k \geq -\log\left(\frac{\sum_{k=1}^{n} x_k}{n}\right)$

$$\Rightarrow \left(\prod_{k=1}^{n} x_k\right)^{1/n} \leq \frac{1}{n}\sum_{k=1}^{n} x_k.$$ This is the AM-GM ineq. $\square$

Conditional expectation

Let $B \in \mathcal{F}$ with $\mathbb{P}(B) > 0$, X a RV. Define $$\mathbb{E}(X|B) = \frac{\mathbb{E}(X\cdot 1(B))}{\mathbb{P}(B)}$$

Law of total expectation Suppose $X \geq 0$ and let (Ω_n) be a partition of Ω into disjoint events. Then $\mathbb{E}(X) = \sum_n \mathbb{E}(X|\Omega_n)\cdot\mathbb{P}(\Omega_n)$

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

$\Rightarrow \mathbb{E}(X) = \mathbb{E}\left(\sum_n X\cdot 1(\Omega_n)\right) = \sum_n \mathbb{E}(X\cdot 1(\Omega_n))$ (countable additivity)

so $\mathbb{E}(X) = \sum_m \mathbb{E}(X\cdot 1(\Omega_n)) = \sum_m \mathbb{E}(X|\Omega_n)\cdot\mathbb{P}(\Omega_n)$. $\square$

Probability - Lecture 11

Definition Let $X_1, \dots, X_n$ be DRVs. Their joint distribution is

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

$$\mathbb{P}(X_1 = x_1) = \mathbb{P}\left(\{X_1 = x_1\} \cap \left(\bigcup_{i=2}^{n} \bigcup_{x_i} \{X_i = x_i\}\right)\right) = \sum_{x_2, \dots x_n} \mathbb{P}(X_1 = x_1, \dots, X_n = x_n)$$

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

Call $\left(\mathbb{P}(X_i = x_i)\right)_{x_i}$ the marginal distribution of X_i.

Let X and Y be 2 RVs. The conditional distribution of X given $Y = y$ $(y \in \Omega_Y)$ is $\mathbb{P}(X = x | Y = y)$, $x \in \Omega_X$.

$$\mathbb{P}(X = x | Y = y) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)}$$

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

(law of total prob.)

Distribution of the sum of indep. RVs

Let X and Y be 2 discrete RVs (indep.)

$$\mathbb{P}(X + Y = z) = \sum_y \mathbb{P}(X + Y = z, Y = y) = \sum_y \mathbb{P}(X = z - y, Y = y)$$

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

- this last sum is called the convolution of the distr. of X and Y.

Similarly $\mathbb{P}(X + Y = z) = \sum_x \mathbb{P}(X = x)\, \mathbb{P}(Y = z - x)$

Example Let $X \sim \mathrm{Poi}(\lambda)$, $Y \sim \mathrm{Poi}(\mu)$ independent

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

$$= \frac{e^{-(\lambda+\mu)}}{n!} \sum_{r=0}^{n} \lambda^r \mu^{n-r} \frac{n!}{r!(n-r)!} = \frac{(\lambda+\mu)^n}{n!} e^{-(\lambda+\mu)}$$

pmf of Poisson - parameter $\underline{\lambda + \mu}$

$X + Y \sim \mathrm{Poi}(\lambda + \mu)$

Definition Let X and Y be 2 DRVs. The <u>conditional expectation</u> of X given $Y=y$ is $\mathbb{E}(X \mid Y=y) = \dfrac{\mathbb{E}(X \cdot \mathbb{1}(Y=y))}{\mathbb{P}(Y=y)}$ (see Lecture 10)

$$\mathbb{E}(X \mid Y=y) = \frac{1}{\mathbb{P}(Y=y)} \mathbb{E}(X \cdot \mathbb{1}(Y=y)) = \frac{1}{\mathbb{P}(Y=y)} \sum_x x \cdot \mathbb{P}(X=x, Y=y)$$
$$= \sum_x x\, \mathbb{P}(X=x \mid Y=y)$$

Note that for every $y \in \Omega_Y$, $\mathbb{E}(X \mid Y=y)$ is a function of just y.
Set ~~$g(y) = \mathbb{E}(X$~~ $g(y) = \mathbb{E}(X \mid Y=y)$.

We define conditional expectation of X given Y and write as $\mathbb{E}(X \mid Y)$ for the RV $g(Y)$.
We emphasise that $\mathbb{E}(X \mid Y)$ is a random variable and it depends only on Y, as it is a function only of Y.

$$\mathbb{E}(X \mid Y) = g(Y) \cdot 1 = g(Y) \sum_y \mathbb{1}(Y=y) = \sum_y g(Y) \cdot \mathbb{1}(Y=y)$$
$$= \sum_y g(y) \cdot \mathbb{1}(Y=y) = \sum_y \mathbb{E}(X \mid Y=y) \cdot \mathbb{1}(Y=y)$$

<u>Example</u> Toss a p-biased coin n times independently.
Write $X_i = \mathbb{1}(i^{th} \text{ toss is a H})$ for $i=1,\dots,n$ and $Y_n = X_1 + \dots + X_n$
What is $\mathbb{E}(X_1 \mid Y_n)$?
Set $g(y) = \mathbb{E}(X_1 \mid Y_n = y)$ then $\mathbb{E}(X_1 \mid Y_n) = g(Y_n)$.
Need to find g.
Let $y \in \{0, \dots, n\}$. Then $g(y) = \mathbb{E}(X_1 \mid Y_n = y)$
$= \mathbb{P}(X_1 = 1 \mid Y_n = y)$

$y=0$: $\mathbb{P}(X_1 = 1 \mid Y_n = 0) = 0$

$y \neq 0$: $\mathbb{P}(X_1 = 1 \mid Y_n = y) = \dfrac{\mathbb{P}(X_1 = 1, Y_n = y)}{\mathbb{P}(Y_n = y)} = \dfrac{\mathbb{P}(X_1 = 1, X_2 + \dots + X_n = y-1)}{\mathbb{P}(Y_n = y)}$

Since the (X_i) are indep, identically distributed, we get

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

$$= p \binom{n-1}{y-1} p^{y-1} (1-p)^{n-y}$$

$$\mathbb{P}(Y_n = y) = \binom{n}{y} p^y (1-p)^{n-y}$$

so $$\mathbb{P}(X_1 = 1 \mid Y_n = y) = \frac{\cancel{p}\binom{n-1}{y-1} \cancel{p^{y-1}} \cancel{(1-p)^{n-y}}}{\binom{n}{y} \cancel{p^y} \cancel{(1-p)^{n-y}}} = \frac{y}{n}$$

Then $g(y) = \frac{y}{n}$. Therefore $\mathbb{E}(X_1 \mid Y_n) = g(Y_n) = \frac{Y_n}{n}$.

<u>Properties of conditional expectation</u>

- $\forall c \in \mathbb{R}$, $\mathbb{E}(cX \mid Y) = c\,\mathbb{E}(X \mid Y)$ and $\mathbb{E}(c \mid Y) = c$
- $X_1, \dots, X_n, Y$ RVs : $\mathbb{E}\left(\sum_{i=1}^n X_i \,\middle|\, Y\right) = \sum_{i=1}^n \mathbb{E}(X_i \mid Y)$
- $\mathbb{E}(\mathbb{E}(X \mid Y)) = \mathbb{E}(X)$

<u>Proof</u> $\mathbb{E}(X \mid Y) = \sum_y 1(Y=y)\, \mathbb{E}(X \mid Y=y)$

By standard properties $\mathbb{E}(\mathbb{E}(X \mid Y)) = \sum_y \mathbb{E}(X \mid Y=y)\, \mathbb{E}(1(Y=y))$

$$= \sum_y \mathbb{E}(X \mid Y=y)\, \mathbb{P}(Y=y) = \sum_y \frac{\mathbb{E}(X \cdot 1(Y=y))}{\cancel{\mathbb{P}(Y=y)}} \cancel{\mathbb{P}(Y=y)}$$

$$= \sum_y \mathbb{E}(X \cdot 1(Y=y)) = \cancel{\sum_y}\, \mathbb{E}\left(X \cdot \sum_y 1(Y=y)\right) = \mathbb{E}(X)$$

<u>Alternatively</u>

$$\sum_y \mathbb{E}(X \mid Y=y) \cdot \mathbb{P}(Y=y) = \sum_x \sum_y x\, \mathbb{P}(X=x \mid Y=y)\, \mathbb{P}(Y=y) = \mathbb{E}(X) \quad \square$$

- X, Y indep RVs, then $\mathbb{E}(X \mid Y) = \mathbb{E}(X)$

<u>Proof</u> $\mathbb{E}(X \mid Y) = \sum_y 1(Y=y)\, \mathbb{E}(X \mid Y=y)$

$$= \sum_y 1(Y=y) \sum_x x\, \mathbb{P}(X=x \mid Y=y) \overset{\text{indep}}{=} \underbrace{\sum_y 1(Y=y)}_{1} \underbrace{\sum_x x\, \mathbb{P}(X=x)}_{\mathbb{E}(X)} = \underline{\mathbb{E}(X)} \quad \square$$

Probability - Lecture 12 17/2/21

Let X and Y be 2 RVs.

$$\mathbb{E}(X|Y) = \sum_y 1(Y=y) \underbrace{\mathbb{E}(X|Y=y)}_{g(y)} \qquad \mathbb{E}(X|Y) = g(Y)$$

Suppose Y and Z are indep. RVs. Then

$$\mathbb{E}\big(\mathbb{E}(X|Y) \mid Z\big) = \mathbb{E}(X).$$

Proof We have $\mathbb{E}(X|Y) = g(Y)$ i.e. $\mathbb{E}(X|Y)$ depends only on Y.

Y, Z indep $\Rightarrow$ $g(Y)$ indep. of Z (in fact true for any function of Y)

So $g(Y)$ is indep. of Z. By last property from Lecture 11:

$$\mathbb{E}(g(Y)|Z) = \mathbb{E}(g(Y)) = \mathbb{E}\big(\mathbb{E}(X|Y)\big) \underset{\substack{\uparrow \\ \text{last time}}}{=} \mathbb{E}(X) \qquad \square$$

Suppose $h: \mathbb{R} \to \mathbb{R}$ is a function. Then $\mathbb{E}(h(Y)X|Y) = h(Y)\mathbb{E}(X|Y)$

Proof $\mathbb{E}(h(Y) \cdot X | Y=y) = \mathbb{E}(h(y) \cdot X | Y=y) = \underset{\substack{\uparrow \\ \text{lowercase}}}{h(y)} \mathbb{E}(X|Y=y)$

so $\mathbb{E}(h(Y) \cdot X | Y) = h(Y)\mathbb{E}(X|Y)$. $\square$

Corollary $\mathbb{E}(\mathbb{E}(X|Y)|Y) = \mathbb{E}(X|Y)$ and $\mathbb{E}(X|X) = X$.

Recall $X_i = 1(i^{\text{th}} \text{ toss is H})$ $\quad Y_n = X_1 + \dots + X_n$

$\mathbb{E}(X_1|Y_n) = \frac{Y_n}{n}$ $\quad$ by symmetry, $\forall i$ $\mathbb{E}(X_i|Y_n) = \mathbb{E}(X_1|Y_n)$

$$\mathbb{E}\Big(\underbrace{\sum_{i=1}^n X_i}_{=Y_n} \Big| Y_n\Big) = \sum_{i=1}^n \mathbb{E}(X_i|Y_n) = n\,\mathbb{E}(X_1|Y_n)$$

$$\therefore \mathbb{E}(X_1|Y_n) = \frac{1}{n}\mathbb{E}(Y_n|Y_n) = \frac{Y_n}{n}.$$

Random walks

A random/stochastic process is a sequence of RVs $(X_n)_{n \in \mathbb{N}}$.

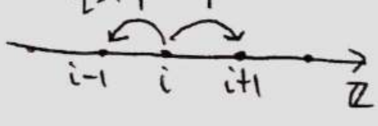
A random walk is a random process that can be expressed by

$X_n = x + Y_1 + \dots + Y_n$ $\quad$ (Y_i) independent, identically distributed

(iid) RVs and x is a deterministic number

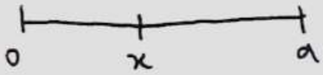
Let's focus on SRW (simple random walk) on $\mathbb{Z}$ defined by

$\mathbb{P}(Y_i = \pm 1) = p \qquad \mathbb{P}(Y_i = -1) = q = 1-p$



Can think of X_n as fortune of gambler who bets 1 at every step and either receives it back doubled (p) or loses it ($q = 1-p$)

Start with £x at time 0. What is $\mathbb{P}$(reaches a before going bankrupt)?



<u>Notation</u> $\mathbb{P}_x$ prob measure $\mathbb{P}(\cdot \mid X_0 = x)$

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

Define $h(x) = \mathbb{P}_x((X_n) \text{ hits } a \text{ before } 0)$.

By law of total prob: $h(x) = \mathbb{P}_x((X_n) \text{ hits } a \text{ bef. } 0 \mid Y_1 = +1)$ times $(\mathbb{P}_x(Y_1 = 1))$

$+ \mathbb{P}_x((X_n) \text{ hits } a \text{ bef. } 0 \mid Y_1 = -1) \cdot \mathbb{P}(Y_1 = -1)$

$\Rightarrow h(x) = p \cdot h(x+1) + q \cdot h(x-1) \qquad 0 < x < a \qquad h(0) = 0,\ h(a) = 1$

- Case $p = q = \frac{1}{2} \Rightarrow h(x) - h(x+1) = h(x-1) - h(x)$

  solve recurrence: $h(x) = \frac{x}{a}$

- Case $p \neq q \quad h(x) = p\,h(x+1) + q\,h(x-1)$ Try λ^x solution

  $\Rightarrow p\lambda^2 - \lambda + q = 0 \Rightarrow \lambda = 1$ or $\frac{q}{p}$: general sol^n $h(x) = A + B\left(\frac{q}{p}\right)^x$

  BCs : $h(a) = 1,\ h(0) = 0 \Rightarrow h(x) = \dfrac{\left(\frac{q}{p}\right)^x - 1}{\left(\frac{q}{p}\right)^a - 1}$

  "Gambler's ruin estimate"

<u>Expected time to absorption</u>

Define $T = \min\{n \geq 0 : X_n \in \{0, a\}\}$ i.e. T first time X hits 0 or a

Want $\mathbb{E}_x(T) = \tau_x = ?$ Conditioning on 1st step + law of tot. exp.:

$\tau_x = p\,\mathbb{E}_x(T \mid Y_1 = +1) + q\,\mathbb{E}_x(T \mid Y_1 = -1) \qquad 0 < x < a$

$\Rightarrow \tau_x = p(1 + \mathbb{E}_{x+1}(T)) + q(1 + \mathbb{E}_{x-1}(T))$

so $\tau_x = 1 + p\,\tau_{x+1} + q\,\tau_{x-1} \qquad 0 < x < a \qquad \tau_0 = \tau_a = 0$

- Case $p = \frac{1}{2}$ try soln $Ax^2 \Rightarrow Ax^2 = 1 + pA(x+1)^2 + qA(x-1)^2$

  $\Rightarrow A = -1$: general soln $\tau_x = Ax^2 + Bx + C = -x^2 + Bx + C$

  BCs $\tau_0 = \tau_a = 0 \Rightarrow \underline{\tau_x = x(a-x)}$

- Case $p \neq \frac{1}{2}$ try Cx as soln: $C = \frac{1}{q-p}$, general soln ~~τ_x~~

  $\tau_x = \frac{1}{q-p}x + A + B\left(\frac{q}{p}\right)^x \qquad$ BCs : $\tau_x = \frac{1}{q-p}x - \frac{a}{q-p}\,\dfrac{\left(\frac{q}{p}\right)^x - 1}{\left(\frac{q}{p}\right)^a - 1}$

  a

18/2/21

Probability - Lecture 13

Probability Generating Functions

X a RV, values in $\mathbb{N}$. $p_r = \mathbb{P}(X=r)$, $r \in \mathbb{N}$ be its pmf

Define pgf of X by

$$p(z) = \sum_{r=0}^{\infty} p_r z^r = \mathbb{E}(z^X) \quad \text{for } |z| \leq 1$$

When $|z| \leq 1$, pgf converges absolutely, so $p(z)$ is well-defined with radius of convergence at least 1.

Theorem The pgf uniquely determines the distribution of X.

Proof Suppose (p_r) and (q_r) are 2 pmfs with

$$\sum_{r=0}^{\infty} p_r z^r = \sum_{r=0}^{\infty} q_r z^r \quad \forall |z| \in 1. \quad \text{Will show } p_r = q_r \ \forall r.$$

Set $z=0$: then $p_0 = q_0$

Suppose $p_r = q_r \ \forall r \leq n$: want $p_{r+1} = q_{r+1}$

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

Divide through by z^{n+1} and let $z \to 0$: have $p_{n+1} = q_{n+1}$. $\square$

Theorem We have $\lim_{z \to 1^-} p'(z) = p'(1-) = \mathbb{E}(X)$.

Proof Assume first that $\mathbb{E}(X) < \infty$. Let $0 < z < 1$. We can differentiate $p(z)$ term by term (see Analysis):

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

Since $0<z<1$, $p'(z)$ is an increasing function of z.

Then $\lim_{z \to 1^-} p'(z) \leq \mathbb{E}(X)$. Let $\varepsilon > 0$.

Let N be large enough that $\sum_{r=0}^{N} r p_r \geq \mathbb{E}(X) - \varepsilon$.

Also $p'(z) \geq \sum_{r=1}^{N} r p_r z^{r-1}$ (as $z > 0$)

then $\lim_{z \to 1^-} p'(z) \geq \sum_{r=1}^{N} r p_r \geq \mathbb{E}(X) - \varepsilon \quad \forall \varepsilon > 0$

so $\lim_{z \to 1^-} p'(z) = p'(1-) = \mathbb{E}(X)$.

Now assume $\mathbb{E}(X) = \infty$. For any M take N large enough s.t. $\sum_{r=0}^{N} r p_r \geq M$. We know that $\lim_{z \to 1^-} p'(z) \geq \sum_{r=1} r p_r \geq M$

so as this is true $\forall M > 0$, we have $\lim_{z \to 1^-} p'(z) = p'(1-) = \mathbb{E}(X) = \infty$.

$\square$

Similarly, can prove

<u>Theorem</u> $p''(1-) = \lim_{z \to 1^-} p''(z) = \mathbb{E}\big(X(X-1)\big)$

$$p^{(k)}(1-) = \lim_{z \to 1^-} p^{(k)}(z) = \mathbb{E}\big(X(X-1)\cdots(X-k+1)\big)$$

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

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

Suppose that $X_1, \ldots, X_n$ are indep. RVs with pgfs $q_1, \ldots, q_n$.
i.e. $q_i(z) = \mathbb{E}(z^{X_i})$

$p(z) = \mathbb{E}(z^{X_1 + \cdots + X_n})$?

$$p(z) = \mathbb{E}(z^{X_1} z^{X_2} \cdots z^{X_n}) \overset{\text{indep}}{=} \mathbb{E}(z^{X_1}) \cdots \mathbb{E}(z^{X_n}) = \underline{q_1(z) \cdots q_n(z)}.$$

If X_i's are iid then $p(z) = q(z)^n$.

Examples

1) $X \sim Bin(n,p)$

$$p(z) = \mathbb{E}(z^X) = \sum_{r=0}^{n} z^r \binom{n}{r} p^r (1-p)^{n-r} = \sum_{r=0}^{n} \binom{n}{r} (pz)^r (1-p)^{n-r}$$

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

2) Let $X \sim Bin(n,p)$ and $Y \sim Bin(m,p)$ and $X \perp\!\!\!\perp Y$.

$$\mathbb{E}(z^{X+Y}) = \mathbb{E}(z^X)\,\mathbb{E}(z^Y) = (pz+1-p)^n (pz+1-p)^m = (pz+1-p)^{n+m}$$

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

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

$$\mathbb{E}(z^X) = \sum_{r=0}^{\infty} (1-p)^r p\, z^r = \frac{p}{1-z(1-p)}$$

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

$$\mathbb{E}(z^X) = \sum_{r=0}^{\infty} z^r e^{-\lambda} \frac{\lambda^r}{r!} = e^{-\lambda} e^{\lambda z} = e^{\lambda(z-1)}$$

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

$$\mathbb{E}(z^{X+Y}) = e^{\lambda(z-1)} e^{\mu(z-1)} = e^{(\lambda+\mu)(z-1)} \Rightarrow \underline{X+Y \sim Poi(\lambda+\mu)}$$

Sum of a random number of RVs $\quad (\Omega, \mathcal{F}, \mathbb{P})$

$X_1, X_2, \ldots$ iid and N be an indep. RV with values in $\mathbb{N}$

Define $S_n = X_1 + \ldots + X_n, \quad \forall n \geq 1$.

Then $S_N = X_1 + \ldots + X_N$. This means $\forall \omega \in \Omega$,

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

Let q be pgf of N, p pgf of X_1.

Then $r(z) = \mathbb{E}(z^{S_N}) = \mathbb{E}\left(z^{X_1+\ldots+X_N}\right) = \cancel{\sum_m \mathbb{E}(z^{X_1+\ldots+X_m})}$

$$= \sum_n \mathbb{E}\left(z^{X_1+\ldots+X_N} \cdot 1(N=m)\right)$$

$$= \sum_n \mathbb{E}\left(z^{X_1+\dots+X_n} \cdot 1(N=m)\right) \qquad \text{indep } (X_i) \text{ and } N:$$

$$= \sum_n \mathbb{E}\left(z^{X_1+\dots+X_n}\right) \mathbb{P}(N=n) \qquad \text{indep } X_i\text{'s}:$$

$$= \sum_n \left(\mathbb{E}(z^{X_1})\right)^n \mathbb{P}(N=n) = \sum_n (p(z))^n \mathbb{P}(N=n) = \underline{q(p(z))}$$

<u>Another conditional exp. proof</u>

$$r(z) = \mathbb{E}\left(z^{X_1+\dots+X_N}\right) = \mathbb{E}\left(\mathbb{E}\left(z^{X_1+\dots+X_N} \,\middle|\, N\right)\right) \quad \swarrow \text{given}$$

$$\mathbb{E}\left(z^{X_1+\dots+X_N} \mid N=n\right) = \cancel{\mathbb{E}\left(z^{X_1+\dots+X_n}(N=n)\right)}$$

$$\mathbb{E}\left(z^{X_1+\dots+X_n} \mid N=n\right) \underset{\text{indep}}{=} \left(\mathbb{E}(z^{X_1})\right)^n = p(z)^n$$

$$\text{so } r(z) = \mathbb{E}\left((p(z))^N\right) = q(p(z)).$$

$$\mathbb{E}(S_N) = \lim_{z \to 1^-} r'(z) = r'(1-)$$

$$r'(z) = q'(p(z))\, p'(z) \quad \text{so } \mathbb{E}(S_N) = q'(p(1-))\, p'(1-)$$

$$= \mathbb{E}(N) \cdot \mathbb{E}(X_1) = \mathbb{E}(S_N)$$

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

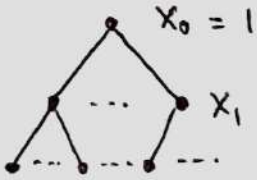
Probability - Lecture 14 $\qquad$ 18/2/21

Branching processes

$(X_n \quad n \geq 0)$ random process $\qquad X_n = \#$ of individuals in generation n

$X_0 = 1$ - the individual in gen. 0 produces a random # of offspring with distribution $g_k = \mathbb{P}(X_1 = k)$, $k = 0, 1, 2, \dots$

$\downarrow$ # of children of 1st individual



Every individual in gen 1 produces an indep. # of offspring, same distribution

Continue in same way: every new individual produces indep. # of offspring with same distribution as X_1.

Let $(Y_{k,n} : k \geq 1, n \geq 0)$ be an iid sequence with distribution $(g_k)_k$.

$Y_{k,n}$: # offspring of k^{th} indiv. in gen n.

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

Theorem $\quad \mathbb{E}(X_n) = (\mathbb{E}(X_1))^n \quad \forall n \geq 1$

Proof $\quad \mathbb{E}(X_{n+1}) = \mathbb{E}(\mathbb{E}(X_{n+1} \mid X_n))$

$\mathbb{E}(X_{n+1} \mid X_n = m) = \mathbb{E}(Y_{1,n} + \dots + Y_{X_n,n} \mid X_n = m)$

$= \mathbb{E}(Y_{1,n} + \dots + Y_{m,n} \mid X_n = m)$

$= m\,\mathbb{E}(X_1)$

so $\mathbb{E}(X_{n+1} \mid X_n) = X_n \cdot \mathbb{E}(X_1)$ so $\mathbb{E}(X_{n+1}) = \mathbb{E}(X_n \cdot \mathbb{E}(X_1))$
$= \mathbb{E}(X_1)\,\mathbb{E}(X_n)$

Iterating ~~and using~~ we get $\mathbb{E}(X_{n+1}) = (\mathbb{E}(X_1))^{n+1}$. $\qquad \square$

Theorem Set ~~gt~~ $G(z) = \mathbb{E}(z^{X_1})$ and $G_n(z) = \mathbb{E}(z^{X_n})$. Then
$$G_{n+1}(z) = G(G_n(z)) = G(G(\dots G(z))\dots) = G_n(G(z)).$$

Proof $\quad G_{n+1}(z) = \mathbb{E}(z^{X_{n+1}}) = \mathbb{E}(\mathbb{E}(z^{X_{n+1}} \mid X_n))$

$\mathbb{E}(z^{X_{n+1}} \mid X_n = m) = \mathbb{E}(z^{Y_{1,n} + \dots + Y_{m,n}} \mid X_n = m) = (\mathbb{E}(z^{X_1}))^m = (G(z))^m$

So $\mathbb{E}(\mathbb{E}(z^{X_{n+1}} \mid X_n)) = \mathbb{E}((G(z))^{X_n}) = G_n(G(z))$ $\square$

<u>Extinction probability</u> $\mathbb{P}(X_n = 0 \text{ for some } n \geq 1)$ call it q

$q_n = \mathbb{P}(X_n = 0)$ $\quad A_n = \{X_n = 0\} \subseteq \{X_{n+1} = 0\} = A_{n+1}$

(A_n) increasing sequence of events

So by continuity of prob. measure: $\mathbb{P}(A_n) \to \mathbb{P}(\bigcup_n A_n)$ as $n \to \infty$

But $\bigcup_n A_n = \{X_n = 0 \text{ for some } n \geq 1\}$

Hence $q_n \to q$ as $n \to \infty$.

<u>Claim</u> $q_{n+1} = G(q_n)$ $\quad (G(z) = \mathbb{E}(z^{X_1}))$ and also $q = G(q)$

<u>Proof</u> $q_{n+1} = \mathbb{P}(X_{n+1} = 0) = G_{n+1}(0) = G(G_n(0)) = G(q_n)$

Since G is continuous, taking limit as $n \to \infty$, using $q_n \to q$ get $\underline{G(q) = q}$ $\square$

Another proof that $q_{n+1} = G(q_n)$:

Conditional on $X_1 = m$ we get m independent branching processes.

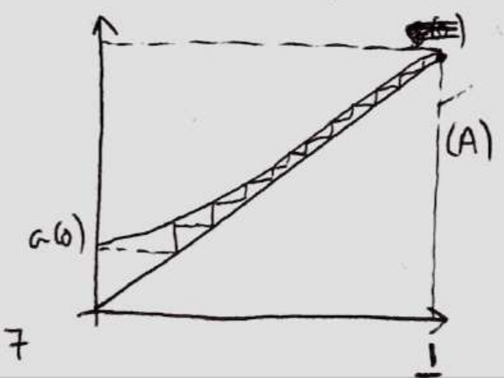
Can write $X_{n+1} = X_n^{(1)} + \dots + X_n^{(m)}$ where $X_i^{(j)}$ are iid BPs all with same offspring distribution

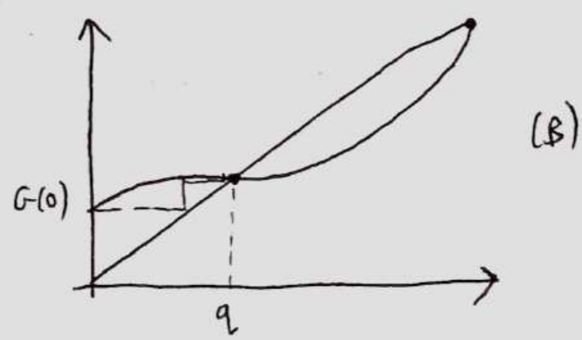
$$\text{So } q_{n+1} = \mathbb{P}(X_{n+1} = 0) \underset{\substack{\text{law of} \\ \text{total prob}}}{=} \sum_m \mathbb{P}(X_{n+1} = 0 \mid X_1 = m)\, \mathbb{P}(X_1 = m)$$

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

$$= \sum_m \underbrace{(\mathbb{P}(X_n^{(1)} = 0))}_{q_n}{}^m\, \mathbb{P}(X_1 = m) = \underline{G(q_n)}. \quad \square$$

So we have proved $q_{n+1} = G(q_n)$ and $q = G(q)$.





In (A) tangent to graph of G at 1 has slope < 1

Slope $= G'(1-) = \mathbb{E}(X_1) < 1$

In (B) the slope is $G'(1-) = \mathbb{E}(X_1) > 1$ and we see that $q < 1$ in this case.

Theorem Assume $\mathbb{P}(X_1 = 1) < 1$. Then extinction prob. is the minimal non-negative solution to $t = G(t)$. We also have $q < 1$ iff $\mathbb{E}(X_1) > 1$.

Proof of minimality Let t be the smallest non-negative solution to $x = G(x)$. We will show $q = t$.

We will prove by induction that $q_n \leq t$ $\forall n$ then taking limits as $n \to \infty$ gives $q \leq t$.

Since we know q is a solution, this implies $q = t$.

$q_0 = 0 = \mathbb{P}(X_0 = 0) \leq t$. Suppose $q_n \leq t$. Then

$q_{n+1} = G(q_n)$. G is increasing on $[0, 1]$ and since $q_n \leq t$ we get $q_{n+1} = G(q_n) \leq G(t) = t$. $\square$

Probability - Lecture 15 24/2/21

Continued proof of theorem from end of L4:

Assume $P(X_1 = 1) < 1$. Then $q < 1$ iff $\mathbb{E}(X_1) > 1$.

Proof Consider $H(z) = G(z) - z$. Assume $g_0 + g_1 < 1$.

Then $\mathbb{P}(X_1 \leq 1) < 1$. [(If not, then $\mathbb{P}(X_1 \leq 1) = 1 \Rightarrow \mathbb{E}(X_1) = \mathbb{P}(X_1 = 1) < 1$)

Then would have $G(z) = g_0 + g_1 z = 1 - \mathbb{E}(X_1) + \mathbb{E}(X_1) z$

Solving $G(z) = z$ would get $z = 1$ since $\mathbb{E}(X_1) < 1$.]

$$H''(z) = \sum r(r-1) g_r z^{r-2} > 0 \quad \forall z \in (0,1)$$

This implies $H'(z)$ is strictly increasing in $(0,1)$.

This implies H can have at most one root different from 1 in $(0,1)$

Case 1 H has no other root but 1

$$H(1) = 0 \text{ and } H(0) = g_0 \geq 0 \Rightarrow H(z) \geq 0 \quad \forall z \in (0,1)$$

$$H'(1-) = \lim_{z \uparrow 1} \frac{\overset{\geq 0}{H(z) - H(1)}}{\underset{\leq 0}{z - 1}} \leq 0$$

but $H'(1-) = G'(1-) - 1$ and $H'(1-) \leq 0 \Rightarrow G'(1-) \leq 1$

and $G'(1-) = \mathbb{E}(X_1)$

so we showed $q = 1 \Rightarrow \mathbb{E}(X_1) \leq 1$

Case 2 H has exactly one other root $r < 1$

$H(r) = 0$, $H(1) = 0$ so by Rolle's theorem $\exists z \in (r, 1)$ s.t. $H'(z) = 0$

But $H'(x) = G'(x) - 1 \Rightarrow G'(z) = 1$

$$G'(x) = \sum_{r=1}^{\infty} r g_r x^{r-1} \quad \text{and} \quad G''(x) = \sum_{r=2}^{\infty} r(r-1) g_r x^{r-2}$$

under $g_0 + g_1 < 1$, $G''(x) > 0 \ \forall x \in (0,1) \Rightarrow G'$ is strictly increasing

then $G'(1-) > G'(z) = 1 \Rightarrow \mathbb{E}(X_1) > 1$

so if $q < 1$ then $\mathbb{E}(X_1) > 1$. $\square$

Continuous random variables

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

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

Definition The prob. distribution function ~~(pmf)~~ is

$F : \mathbb{R} \to [0,1]$ with $F(x) = \mathbb{P}(X \leq x)$

Properties 1) If $x \leq y$ then $F(x) \leq F(y)$

as $\{X \leq x\} \subseteq \{X \leq y\}$

2) $\forall a < b$, $\mathbb{P}(a < X \leq b) = F(b) - F(a)$

Proof $\mathbb{P}(a < X \leq b) = \mathbb{P}(\{a < X\} \cap \{X \leq b\}) = \mathbb{P}(X \leq b)$

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

3) F is right continuous and left limits exist

$$F(x-) = \lim_{y \uparrow x} F(y) \leq F(x)$$

Proof Enough to show $\lim_{n \to \infty} F(x + \frac{1}{n}) = F(x)$

Define $A_n = \{X \leq x + \frac{1}{n}\}$ then (A_n) are decreasing events

$(A_{n+1} \subseteq A_n)$

and $\bigcap_n A_n = \phi$

so $\mathbb{P}(A_n) \to 0$ as $n \to \infty$

but $\mathbb{P}(A_n) = \mathbb{P}(x < X \leq x + \frac{1}{n}) = F(x + \frac{1}{n}) - F(x) \to 0$ as $n \to \infty$

Left limits exist by increasing property of F.

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

Proof $F(x-) = \lim_{n \to \infty} F\left(x - \frac{1}{n}\right)$

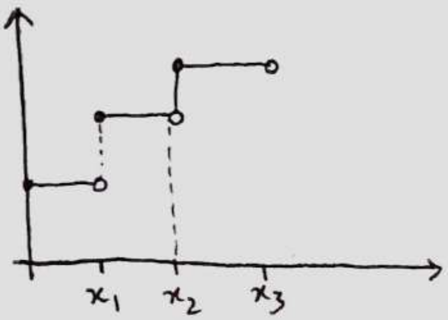
$F(x - \frac{1}{n}) = \mathbb{P}\left(X \leq x - \frac{1}{n}\right)$ Consider $B_n = \{X \leq x - \frac{1}{n}\}$

then $(B_n) \nearrow$ and $\bigcup_n B_n = \{X < x\}$

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

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

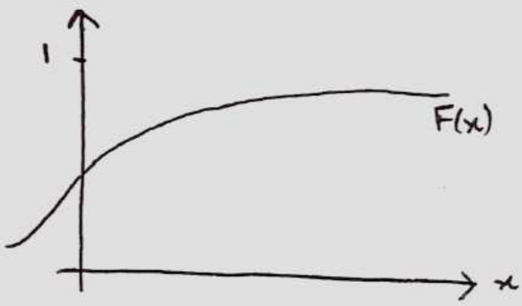
For a discrete variable



$F(x) = \mathbb{P}(X \le x)$

F is a step function (right continuous with left limits)

<u>Definition</u> A RV X is <u>continuous</u> if F is a continuous function which means $F(x) = F(x-)$ $\forall x \Rightarrow \mathbb{P}(X \le x) = \mathbb{P}(X < x)$ i.e. $\mathbb{P}(X = x) = 0$ $\forall x \in \mathbb{R}$.



(In this course, restrict to F continuous and differentiable - "absolutely continuous")

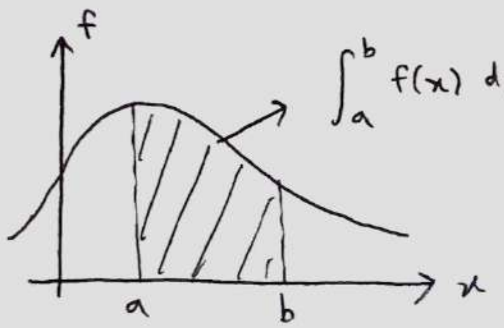
Set $\underline{F'(x) = f(x)}$ then f is <u>probability density function</u> of X

$f \geq 0$, $\int_{-\infty}^{\infty} f(x)\,dx = 1$, $F(x) = \int_{-\infty}^{x} f(y)\,dy$

<u>Intuitively</u>: Δx small, then $\mathbb{P}(x < X \le x + \Delta x) = \int_{x}^{x+\Delta x} f(y)\,dy$

$\approx \Delta x \cdot f(x)$ for Δx small

Think of as "probability X lies in interval close to x"



$\int_a^b f(x)\,dx = \mathbb{P}(a < X \le b)$

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

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

$f \geq 0$, $\int_{-\infty}^{\infty} f(x)\,dx = 1$, generally $\mathbb{P}(X \in A) = \int_A f(x)\,dx$

Expectation Let $X \geq 0$, density f. Then $\mathbb{E}(X) = \int_0^\infty x f(x)\,dx$

Suppose $g \geq 0$. Then $\mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(x) f(x)\,dx$

Let X be a RV. Define $X_+ = \max(X, 0)$, $X_- = \max(-X, 0)$

If at least one of $\mathbb{E}(X_+)$ or $\mathbb{E}(X_-)$ is finite, then we set

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

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

Can check linearity.

Claim Let $X \geq 0$. Then $\mathbb{E}(X) = \int_0^\infty \mathbb{P}(X \geq x)\,dx$.

Proof 1 $\mathbb{E}(X) = \int_0^\infty x f(x)\,dx = \int_0^\infty \left(\int_0^x 1\,dy\right) f(x)\,dx$

$= \int_0^\infty dy \int_y^\infty f(x)\,dx = \int_0^\infty dy\,(1 - F(y)) = \int_0^\infty \mathbb{P}(X \geq y)\,dy$ $\square$

Proof 2 $\forall \omega$ $X(\omega) = \int_0^\infty 1(X(\omega) \geq x)\,dx$

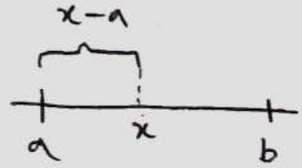
then $\underset{\substack{\uparrow \\ \text{not justifying}}}{\mathbb{E}(X) =} \int_0^\infty \mathbb{P}(X \geq x)\,dx$. $\square$

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

Examples (1) Uniform distribution $a < b$, $a, b \in \mathbb{R}$

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

$X \sim U(a,b)$ $\mathbb{P}(X \leq x) = \int_a^x f(y)\,dy = \frac{x-a}{b-a}$



$F(x) = 0$ $(x < a)$, $F(x) = \frac{x-a}{b-a}$ $(x \in [a,b])$, $F(x) = 1$ $(x > b)$

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

(2) Exponential distribution $f(x) = \lambda e^{-\lambda x}$, $\lambda > 0$, $x \geq 0$ $X \sim \mathrm{Exp}(\lambda)$

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

$$\mathbb{E}(X) = \int_0^\infty \lambda x e^{-\lambda x}\, dx = \frac{1}{\lambda}$$

<u>Limit of geometrics:</u> Let $T \sim \mathrm{Exp}(\lambda)$, $T_n = \lfloor nT \rfloor$ $\forall n \in \mathbb{N}$

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

so T_n is geometric parameter $p_n = 1 - e^{-\lambda/n} \sim \frac{\lambda}{n}$ as $n \to \infty$

So exp is limit of a rescaled geometric.

Memoryless property: $s, t > 0$ $\mathbb{P}(T > t+s \mid T > s)$ $T \sim \mathrm{Exp}(\lambda)$

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

<u>Proposition</u> Let T be a continuous positive RV not identically 0 or ∞. Then T has memoryless property iff T is exponential.

<u>Proof</u> $\forall s, t$ $\mathbb{P}(T > t+s) = \mathbb{P}(T > s)\,\mathbb{P}(T > t)$

Set $g(t) = \mathbb{P}(T > t)$. Need $g(t) = e^{-\lambda t}$ for some $\lambda > 0$.

$g(t+s) = g(t)\,g(s)$ $\forall s, t > 0$

$t \geq 0$, $m \in \mathbb{N}$: $g(mt) = (g(t))^m$ $t = 1 \Rightarrow \forall m \in \mathbb{N}$ $g(m) = g(1)^m$

$g\left(\frac{m}{n}\right)^n = g(m) = g\left(\frac{m}{n} \cdot n\right) \Rightarrow g\left(\frac{m}{n}\right) = g(1)^{m/n}$ $\forall m, n \in \mathbb{N}$

$g(1) = \mathbb{P}(T > 1) \in (0, 1)$. (not identically 0 or ∞) $\lambda = -\log \mathbb{P}(T > 1) > 0$

So we have proved $g(t) = \mathbb{P}(T > t) = e^{-\lambda t}$ $\forall t \in \mathbb{Q}_+$

Let $t \in \mathbb{R}_+$. Then $\exists r, s \in \mathbb{Q}$: $r \leq t \leq s$, $|r - s| \leq \varepsilon$

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

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

Theorem Let X be cont. RV density f. Let g be a cont. function which is either <u>strictly</u> increasing or decreasing and g^{-1} differentiable. Then $g(X)$ is a cont. RV with density $f(g^{-1}(x)) = \left|\frac{d}{dx} g^{-1}(x)\right|$.

<u>Proof</u> $g \uparrow$: $\mathbb{P}(g(X) \le x) = \mathbb{P}(X \le g^{-1}(x)) = F(g^{-1}(x))$

$$\frac{d}{dx}\mathbb{P}(g(X) \le x) = F'(g^{-1}(x)) \cdot \frac{d}{dx} g^{-1}(x) = f(g^{-1}(x)) \cdot \frac{d}{dx} g^{-1}(x)$$

(>0 as $g^{-1} \uparrow$)

$g \downarrow$: $\mathbb{P}(g(X) \le x) = \mathbb{P}(X \geq g^{-1}(x)) = 1 - P(X < g^{-1}(x))$
$= 1 - F(g^{-1}(x))$ since $\mathbb{P}(X = g^{-1}(x)) = 0$

$$\frac{d}{dx}\mathbb{P}(g(X) \le x) = -f(g^{-1}(x)) \cdot \frac{d}{dx} g^{-1}(x) = f(g^{-1}(x)) \left|\frac{d}{dx} g^{-1}(x)\right|$$

~~gk~~ $g^{-1} \downarrow$ so $\frac{d}{dx} g^{-1}(x) < 0$.

$\square$

3) <u>Normal distribution</u> $-\infty < \mu < \infty$, $\sigma > 0$

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \qquad x \in \mathbb{R}$$

Check f is a density: $\int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right) dx$

$u = \frac{x-\mu}{\sigma}$: $= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-u^2}{2}\right) du = 2\int_0^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du = I$

$$I^2 = \frac{2}{\pi} \int_0^{\infty} \int_0^{\infty} e^{-\frac{(u^2+v^2)}{2}} du\,dv$$

Polars: $u = r\cos\theta$, $v = r\sin\theta$

$$I^2 = \frac{2}{\pi} \int_0^{\infty} \int_0^{\pi/2} r e^{-r^2/2}\,dr\,d\theta = 1 \Rightarrow I = 1 \text{ as desired.}$$

So f is a density.

X density f

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} \frac{x}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx = \int_{-\infty}^{\infty} \frac{x-\mu}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx$$
$$+ \mu \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right) dx$$

So $\mathbb{E}(X) = \mu$ (first integral has odd function integrand : 0
2nd is density : 1)

$$\mathrm{Var}(X) = \mathbb{E}((X-\mu)^2) = \int_{-\infty}^{\infty} \frac{(x-\mu)^2}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx$$

$$= \sigma^2 \underbrace{\int_{-\infty}^{\infty} \frac{u^2}{\sqrt{2\pi}} e^{-u^2/2} \, du}_{1 \text{ by IBP}} = \sigma^2 \qquad \left(u = \frac{x-\mu}{\sigma}\right)$$

So $X \sim N(\mu, \sigma^2)$ mean μ, variance σ^2

$\mu = 0$, $\sigma^2 = 1$: standard normal $X \sim N(0,1)$

then write $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} \, du$ and

$\phi(x) = \Phi'(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$.

$\phi(x) = \phi(-x) \Rightarrow \Phi(x) + \Phi(-x) = 1$

$\Rightarrow \mathbb{P}(X \leq x) = 1 - \mathbb{P}(X \leq -x)$.

$X \sim N(\mu, \sigma^2)$ $\quad f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$

Let $a \neq 0$, $b \in \mathbb{R}$, set $g(x) = ax+b$

Define $Y = g(X)$. What is density of Y?

$Y = aX + b$ $\quad g^{-1}(x) = \frac{x-b}{a}$, $\quad \frac{d}{dx} g^{-1}(x) = \frac{1}{a}$

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\left(\frac{y-b}{a} - \mu\right)^2}{2\sigma^2}\right) \cdot \frac{1}{|a|}$$

$$= \frac{1}{\sqrt{2\pi a^2\sigma^2}} \exp\left(\frac{(y-(a\mu+b))^2}{2a^2\sigma^2}\right) \qquad \underline{Y \sim N(a\mu+b,\ a^2\sigma^2)}$$

linear combination

Suppose $X \sim N(\mu, \sigma^2)$, then $\frac{X-\mu}{\sigma} \sim N(0,1)$

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

and $\Phi(2) \geq 0.95$ (using tables)

__Definition__ X continuous RV. The __median__ m of X is the number satisfying $\mathbb{P}(X \leq m) = \mathbb{P}(X \geq m) = \frac{1}{2}$.

i.e. $\int_{-\infty}^{m} f(x)\,dx = \int_{m}^{\infty} f(x)\,dx = \frac{1}{2}$

If $X \sim N(\mu, \sigma^2)$ then $\mathbb{P}(X \leq \mu) = \mathbb{P}\left(\frac{X-\mu}{\sigma} \leq 0\right) = \Phi(0) = \frac{1}{2}$

__Multivariate density functions__

$X = (X_1, \dots, X_n) \in \mathbb{R}^n$ a RV: say X has density f if

$$\underbrace{\mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n)}_{F(x_1,\dots,x_n)} = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f(y_1,\dots,y_n)\,dy_1 \dots dy_n$$

Then $f(x_1,\dots,x_n) = \frac{\partial^n}{\partial x_1 \dots \partial x_n} F(x_1,\dots,x_n)$

This generalises "$\forall$" $B \subseteq \mathbb{R}^n$ $\quad \mathbb{P}((X_1,\dots,X_n) \in B) = \int_B f(y_1,\dots,y_n)\,dy_1 \dots dy_n$

<u>Independence</u> We say $X_1, \dots, X_n$ are independent if $\forall x_1, \dots, x_n \in \mathbb{R}$ have $\mathbb{P}(X_1 \le x_1, \dots, X_n \le x_n) = \mathbb{P}(X_1 \le x_1) \dots \mathbb{P}(X_n \le x_n)$

<u>Theorem</u> Let $X = (X_1, \dots, X_n)$ have density f.

(a) Suppose $X_1, \dots, X_n$ are indep., densities $f_1, \dots, f_n$

Then $f(x_1, \dots, x_n) = f_1(x_1) \dots f_n(x_n)$ $(*)$

(b) Suppose f factorises as in $(*)$ for some non-negative functions (f_i). Then $X_1, \dots, X_n$ are independent and have densities proportional to the f_is.

<u>Proof</u> (a) $\mathbb{P}(X_1 \le x_1, \dots, X_n \le x_n) = \mathbb{P}(X_1 \le x_1) \dots \mathbb{P}(X_n \le x_n)$

$$= \int_{-\infty}^{x_1} f_1(y_1)\, dy_1 \dots \int_{-\infty}^{x_n} f_n(y_n)\, dy_n = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} \prod_{i=1}^{n} f_i(y_i)\, dy_1 \dots dy_n$$

So the density of $(X_1, \dots, X_n)$ is $f = \prod f_i$.

(b) Let $B_1, \dots, B_n \subseteq \mathbb{R}$. Then

$$\mathbb{P}(X_1 \in B_1, \dots, X_n \in B_n) = \int_{B_1} \cdots \int_{B_n} f_1(x_1) \dots f_n(x_n)\, dx_1 \dots dx_n$$

Take $B_j = \mathbb{R}$ $\forall j \ne i$. Then

$$\mathbb{P}(X_i \in B_i) = \mathbb{P}(X_i \in B_i, X_j \in B_j \ \forall j \ne i)$$

$$= \int_{B_i} f_i(y_i)\, dy_i \prod_{j \ne i} \int_{\mathbb{R}} f_j(y)\, dy$$

f is a density so $\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \dots, x_n)\, dx_1 \dots dx_n = 1$

But $f = \prod f_i$ so $\prod_j \int_{-\infty}^{\infty} f_j(y)\, dy = 1$

$$\Rightarrow \prod_{j \ne i} \int_{\mathbb{R}} f_j(y)\, dy = \frac{1}{\int_{\mathbb{R}} f_i(y)\, dy} \Rightarrow \mathbb{P}(X_i \in B_i) = \frac{\int_{B_i} f_i(y)\, dy}{\int_{\mathbb{R}} f_i(y)\, dy}$$

so density of X_i is $\dfrac{f_i}{\int_{\mathbb{R}} f_i(y)\, dy}$.

The X_i's are independent since

$$\mathbb{P}(X_1 \le x_1, \dots, X_n \le x_n) = \frac{\int_{-\infty}^{x_1} f_1(y_1)\,dy_1 \dots \int_{-\infty}^{x_n} f_n(y_n)\,dy_n}{\int_{\mathbb{R}} f_1(y_1)\,dy_1 \dots \int_{\mathbb{R}} f_n(y_n)\,dy_n}$$

$$= \mathbb{P}(X_1 \le x_1) \dots \mathbb{P}(X_n \le x_n). \qquad \square$$

Suppose $(X_1, \dots, X_n)$ has density f.

$$\mathbb{P}(X_1 \le x) = \mathbb{P}(X_1 \le x, X_2 \in \mathbb{R}, \dots, X_n \in \mathbb{R})$$

$$= \int_{-\infty}^{x} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \dots, x_n)\, dx_1 \dots dx_n$$

$$= \int_{-\infty}^{x} \underbrace{\left(\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \dots, x_n)\, dx_2 \dots dx_n \right)}_{\text{density of } X_1 \;=\; \text{marginal density of } X_1} dx_1$$

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \dots, x_n)\, dx_2 \dots dx_n$$

<u>Density of sum of indep. RVs</u>

Let X, Y be indep. RVs with densities f_X and f_Y.

$$\mathbb{P}(X+Y \le z) = \iint_{\{x+y \le z\}} f_{X,Y}(x,y)\, dx\, dy = \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} f_X(x) f_Y(y)\, dx\, dy$$

$$\overset{y'=y+x}{=} \int_{-\infty}^{\infty} \left(\int_{-\infty}^{z} f_Y(y-x)\, \cancel{dy}\, f_X(x)\, dy \right) dx$$

$$= \int_{-\infty}^{z} dy \underbrace{\int_{-\infty}^{\infty} f_Y(y-x) f_X(x)\, dx}_{g(y)}$$

so density of $X+Y$ is $\underline{\int_{-\infty}^{\infty} f_Y(y-x) f_X(x)\, dx}$

We call this function the <u>convolution</u> of f_X and f_Y.

<u>Definition</u> f, g 2 densities:

$$f * g(x) = \int_{-\infty}^{\infty} f(x-y)\, g(y)\, dy$$ = <u>convolution</u> of f and g.

<u>Non-rigorous way</u>

$$\mathbb{P}(X+Y \leq z) = \int_{-\infty}^{\infty} \mathbb{P}(X+Y \leq z,\ Y \in dy)$$

$$= \int_{-\infty}^{\infty} \mathbb{P}(X \leq z-y)\, \mathbb{P}(Y \in dy)$$ by independence

$$= \int_{-\infty}^{\infty} F_X(z-y)\, f_Y(y)\, dy$$

Then $\frac{d}{dz} \mathbb{P}(X+Y \leq z) = \int_{-\infty}^{\infty} \frac{d}{dz} F_X(z-y) f_Y(y)\, dy$

$$= \underline{\int_{-\infty}^{\infty} f_X(z-y)\, f_Y(y)\, dy}$$ giving the same density.

Probability - Lecture 18

Conditional density

X, Y continuous variables with joint density $f_{X,Y}$, marginal densities f_X and f_Y. Then conditional density of X given $Y=y$ is defined by $f_{X|Y}(x|y) = \dfrac{f_{X,Y}(x,y)}{f_Y(y)}$

Law of total probability

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy = \int_{-\infty}^{\infty} f_{X|Y}(x|y)\,f_Y(y)\,dy$$

Want to define $\mathbb{E}(X|Y) = g(Y)$ for some function g

Define $g(y) = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\,dx$ and set $\underbrace{\mathbb{E}(X|Y) = g(Y)}_{\text{conditional expectation}}$

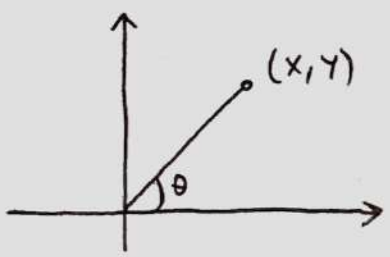
Transformation of a multidimensional RV

Theorem Let X be a RV with values in $D \subset \mathbb{R}^d$, density f_X. Let g be a bijection D to $g(D)$ with continuous derivative and $\det(g'(x)) \neq 0$ $\forall x \in D$. Then RV $Y = g(X)$ has density

$$f_Y(y) = f_X(x)\,|J| \quad \text{where } x = g^{-1}(y), \quad J = \det\left(\left(\frac{\partial x_i}{\partial y_j}\right)_{i,j=1}^{d}\right)$$

Examples

Let X, Y be indep. $N(0,1)$ RVs.



$R = \sqrt{X^2+Y^2}$

What is density of (R, θ)?

$X = R\cos\theta, \quad Y = R\sin\theta$

$$f_{R,\theta}(r,\theta) = f_{X,Y}(r\cos\theta, r\sin\theta) \cdot |J|$$

PTO

$$J = \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} = r$$

so $f_{r,\theta}(r,\theta) = f_X(r\cos\theta)\, f_Y(r\sin\theta)\cdot r$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{r^2\cos^2\theta}{2}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{r^2\sin^2\theta}{2}} \cdot r$$

$$\Rightarrow f_{R,\theta}(r,\theta) = \frac{1}{2\pi} r e^{-r^2/2} \qquad \forall r>0,\ \theta \in [0, 2\pi]$$

This shows that R, θ are independent with $\theta \sim U(0, 2\pi)$ and R has density $re^{-r^2/2}$ on $(0,\infty)$.

<u>Order statistics for a random sample</u>

$X_1, \dots, X_n$ iid, distr F, density f

Put in increasing order $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$, $Y_i = X_{(i)}$

Then (Y_i) are the <u>order statistics</u>.

$\mathbb{P}(Y_1 \leq x) = \mathbb{P}(\min(X_1, \dots, X_n) \leq x) = 1 - \mathbb{P}(\min(X_1, \dots, X_n) > x)$

$= 1 - (1-F(x))^n$.

$f_{Y_1}(x) = \frac{d}{dx}\left(1-(1-F(x))^n\right) = n\cdot(1-F(x))^{n-1}\cdot F(x)$

$\mathbb{P}(Y_n \leq x) = (F(x))^n, \quad f_{Y_n}(x) = n(F(x))^{n-1} f(x)$

Density of $Y_1, \dots, Y_n = ?$

Let $x_1 < x_2 < \dots < x_n$

$\mathbb{P}(Y_1 \leq x_1, \dots, Y_n \leq x_n) = n!\, \mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n, X_1 < X_2 < \dots < X_n)$

$$= n! \int_{-\infty}^{x_1} \int_{u_1}^{x_2} \cdots \int_{u_{n-1}}^{x_n} f(u_1)\dots f(u_n)\, du_1 \dots du_n$$

$f_{Y_1,\dots,Y_n}(x_1, \dots, x_n) = n!\, f(x_1)\dots f(x_n)$ <u>when</u> $x_1 < x_2 < \dots < x_n$

$= 0$ <u>otherwise</u>

Example Let $X \sim Exp(\lambda)$, $Y \sim Exp(\mu)$, X indep. of Y

Set $Z = \min(X, Y)$. $\mathbb{P}(Z \geq z) = \mathbb{P}(X \geq z, Y \geq z)$

$= e^{-\lambda z} e^{-\mu z} = e^{-(\lambda+\mu)z}$ so $Z \sim Exp(\lambda + \mu)$

More generally if $X_1, \dots, X_n$ are indep. with $X_i \sim Exp(\lambda_i)$

then $\min(X_1, \dots, X_n) \sim Exp\left(\sum_{i=1}^{n} \lambda_i\right)$

Now let $X_1, \dots, X_n$ be iid $Exp(\lambda)$, Y_i order statistics

$Z_1 = Y_1$, $Z_2 = Y_2 - Y_1, \dots, Z_n = Y_n - Y_{n-1}$ - what is density of $(Z_1, \dots, Z_n)$?

$$Z = \begin{pmatrix} Z_1 \\ \vdots \\ Z_n \end{pmatrix} = A \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \quad \text{where} \quad A = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & \dots & 0 \\ & & & \vdots & \end{pmatrix}$$

$\det A = 1$, $z = Ay$, then $y_j = \sum_{i=1}^{j} z_i$

$$f_{(Z_1, \dots, Z_n)}(z_1, \dots, z_n) = f_{(Y_1, \dots, Y_n)}(y_1, \dots, y_n) \, |J|^{\nearrow 1} \qquad z_i > 0 \ \forall i$$

$$= n! \, f(y_1) \cdots f(y_n) = n! \, \lambda e^{-\lambda y_1} \cdots \lambda e^{-\lambda y_n}$$

(applying earlier theorem)

$$= n! \, \lambda^n e^{-\lambda(nz_1 + (n-1)z_2 + \dots + z_n)}$$

$$= \prod_{i=1}^{n} (n-i+1) \, \lambda e^{-\lambda(n-i+1) z_i}$$

So $Z_1, \dots, Z_n$ are indep, and $Z_i \sim Exp(\lambda(n-i+1))$

Probability - Lecture 19 4/3/21

Moment generating functions (mgf)

X a RV, density f. The mgf of X is

$$m(\theta) = \mathbb{E}\left(e^{\theta X}\right) = \int_{-\infty}^{\infty} e^{\theta x} f(x)\, dx$$ whenever this is finite.

~~$m(0)$~~ $m(0) = 1$

Theorem statement The mgf uniquely determines the distribution of a RV provided that it is defined for an open interval of values of θ.

Theorem Suppose mgf is defined for an open interval of values of θ. Then $m^{(r)}(0) = \frac{d^r}{d\theta^r} m(\theta) \Big|_{\theta=0} = \mathbb{E}(X^r)$.

Examples

(1) Gamma distribution $f(x) = \dfrac{e^{-\lambda x}\lambda^n x^{n-1}}{(n-1)!}$ $\lambda > 0,\ n \in \mathbb{N},\ x \geq 0$

Denote X with density f as $X \sim \Gamma(n, \lambda)$.

Check f is a density: $I_n = \int_0^\infty f(x)\, dx = \int_0^\infty \lambda e^{-\lambda} \frac{\lambda^{n-1} x^{n-1}}{(n-1)!}\, dx$

IBP

$$= \int_0^\infty \frac{e^{-\lambda x}\lambda^{n-1}\cancel{(n-1)}\cdot x^{n-2}}{(\cancel{n-1})!\,(n-2)!}\, dx = I_{n-1} = \dots = I_1$$

for $n=1$: $f(x) = \lambda e^{-\lambda x}$, Exp$(\lambda)$ so $I_1 = 1$.

$$m(\theta) = \int_0^\infty e^{\theta x}\cdot e^{-\lambda x} \frac{\lambda^n x^{n-1}}{(n-1)!}\, dx$$ Take $\lambda > \theta$ so finite:

$$= \int_0^\infty e^{-(\lambda-\theta)x} \frac{\lambda^n x^{n-1}}{(n-1)!}\, dx = \left(\frac{\lambda}{\lambda-\theta}\right)^n \underbrace{\int_0^\infty \frac{e^{-(\lambda-\theta)x}(\lambda-\theta)^n x^{n-1}}{(n-1)!}\, dx}_{\substack{\Gamma(\lambda-\theta,\, n) \\ =1}}$$

so $m(\theta) = \left(\dfrac{\lambda}{\lambda-\theta}\right)^n$ for $\lambda > \theta$.

<u>Claim</u> $X_1, \dots, X_n$ indep. RVs

Then $m(\theta) = \mathbb{E}\left(e^{\theta(X_1 + \dots + X_n)}\right) = \prod_{i=1}^{n} \mathbb{E}\left(e^{\theta X_i}\right)$

Let $X \sim \Gamma(n, \lambda)$ and $Y \sim \Gamma(m, \lambda)$ and X indep. of Y.

Then $m(\theta) = \mathbb{E}\left(e^{\theta(X+Y)}\right) = \mathbb{E}(e^{\theta X})\,\mathbb{E}(e^{\theta Y}) = \left(\frac{\lambda}{\lambda - \theta}\right)^{n+m}$

So by uniqueness theorem, $X + Y \sim \Gamma(n+m, \lambda)$

In particular this implies $X_1, \dots, X_n$ iid $\mathrm{Exp}(1) = \Gamma(1, \lambda)$ implies $X_1 + \dots + X_n \sim \Gamma(n, \lambda)$.

Could also consider $\Gamma(\alpha, \lambda)$ $(\alpha > 0)$ by replacing $(n-1)!$ with $\Gamma(\alpha) = \int_0^\infty e^{-x} x^{\alpha - 1}\, dx$.

2) <u>Normal distribution</u>

$X \sim N(\mu, \sigma^2)$: $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ $\quad x \in \mathbb{R}$

$$m(\theta) = \int_{-\infty}^{\infty} e^{\theta x} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx$$

$$\theta x - \frac{(x-\mu)^2}{2\sigma^2} = \theta\mu + \frac{\theta^2\sigma^2}{2} - \frac{(x - (\mu + \theta\sigma^2))^2}{2\sigma^2}$$

so $$m(\theta) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\theta\mu + \frac{\theta^2\sigma^2}{2}} \underbrace{\exp\left(-\frac{(x - (\mu + \theta\sigma^2))^2}{2\sigma^2}\right)}_{N(\mu + \theta\sigma^2, \sigma^2} dx$$

$$= \underline{e^{\theta\mu + \frac{\theta^2\sigma^2}{2}}}$$

if $X \sim N(\mu, \sigma^2)$ then $aX + b \sim N(a\mu + b, a^2\sigma^2)$

so $\mathbb{E}\left(e^{\theta(aX+b)}\right) = e^{\theta(a\mu + b) + \frac{\theta^2 a^2 \sigma^2}{2}}$

Suppose $X \sim N(\mu, \sigma^2)$ and $Y \sim N(\nu, \tau^2)$ and $X \perp\!\!\!\perp Y$.

Then $\mathbb{E}\left(e^{\theta(X+Y)}\right) \underset{\text{indep}}{=} \mathbb{E}(e^{\theta X}) \cdot \mathbb{E}(e^{\theta Y}) = e^{\theta\mu + \frac{\theta^2\sigma^2}{2}} e^{\theta\nu + \frac{\theta^2\tau^2}{2}}$

$= e^{\theta(\mu+\nu) + \frac{\theta^2(\sigma^2 + \tau^2)}{2}}$

so $X + Y \sim N(\mu + \nu, \sigma^2 + \tau^2)$

(3) Cauchy distribution

$$f(x) = \frac{1}{\pi(1+x^2)} \qquad x \in \mathbb{R}$$

$$m(\theta) = \mathbb{E}(e^{\theta X}) = \int_{-\infty}^{\infty} \frac{e^{\theta x}}{\pi(1+x^2)}\, dx = \infty \qquad \forall\, \theta \neq 0$$
$$m(0) = 1$$

Suppose $X \sim f$, then $X, 2X, 3X, \ldots$ all have same mgf but not the same distribution. So assumption on $m(\theta)$ being finite for an open interval of values of θ is essential.

Multivariate mgf

Let $X = (X_1, \ldots, X_n)$ be a RV with values in $\mathbb{R}^n$. Then mgf of X is defined by

$$m(\theta) = \mathbb{E}\left(e^{\theta^T X}\right) = \mathbb{E}\left(e^{\theta_1 X_1 + \cdots + \theta_n X_n}\right), \qquad \theta = (\theta_1, \ldots, \theta_n)^*$$

like a vector

Theorem In this case provided mgf is finite for a range of values of θ, it uniquely determines distr. of X.

Also $$\left.\frac{\partial^r m}{\partial \theta_i^r}\right|_{\theta=0} = \mathbb{E}(X_i^r) \qquad \left.\frac{\partial^{r+s} m}{\partial \theta_i^r \partial \theta_j^s}\right|_{\theta=0} = \mathbb{E}(X_i^r X_j^s)$$

$$m(\theta) = \prod_{i=1}^{n} \mathbb{E}(e^{\theta_i X_i}) \quad \text{iff} \quad X_1, \ldots, X_n \text{ indep.}$$

Definition $(X_n : n \in \mathbb{N})$ sequence of RVs, X another RV. We say X_n converges to X in distribution ($X_n \xrightarrow{d} X$) if

$$F_{X_n}(x) \to F_X(x) \qquad \forall\, x \in \mathbb{R} \text{ that are continuity points of } F_X.$$

Theorem (Continuity property for mgf's)

Let X be a RV with $m(\theta)$ finite for some $\theta \neq 0$. Suppose that $m_n(\theta) \to m(\theta)$ $\forall \theta \in \mathbb{R}$ where $m_n(\theta) = \mathbb{E}(e^{\theta X_n})$ and $m(\theta) = \mathbb{E}(e^{\theta X})$. Then X_n converges to X in distribution.

Limit theorems for sums of iid RVs

Theorem (Weak law of large numbers)

Let $(X_n : n \in \mathbb{N})$ be a sequence of iid RVs, $\mu = \mathbb{E}(X_1) < \infty$
Set $S_n = X_1 + \dots + X_n$. Then $\forall \varepsilon > 0$:

$$\mathbb{P}\left(\left| \frac{S_n}{n} - \mu \right| > \varepsilon \right) \to 0 \quad \text{as} \quad n \to \infty$$

Proof Assuming σ^2 finite $= \mathrm{Var}(X_1)$:

$$\mathbb{P}\left(\left| \frac{S_n}{n} - \mu \right| > \varepsilon \right) = \mathbb{P}\left(|S_n - n\mu| > \varepsilon n \right) \overset{\text{Chebyshev}}{\leq} \frac{\mathrm{Var}(S_n)}{\varepsilon^2 n^2}$$

$S_n = X_1 + \dots + X_n \Rightarrow \mathrm{Var}(S_n) = n\sigma^2$: $\underline{\dfrac{n\sigma^2}{\varepsilon^2 n^2} \to 0}$
as $n \to \infty$.

□

Definition A sequence (X_n) converges to X in probability (write $X_n \xrightarrow{\mathbb{P}} X$ as $n \to \infty$) if $\forall \varepsilon > 0$,

$$\mathbb{P}(|X_n - X| > \varepsilon) \to 0 \quad \text{as} \quad n \to \infty.$$

Definition (X_n) converges to X with probability 1 / almost surely if $\mathbb{P}\left(\lim_{n \to \infty} X_n = X \right) = 1$.

Probability - Lecture 20 9/3/21

<u>Claim</u> Suppose $X_n \to 0$ almost surely as $n \to \infty$. Then $X_n \xrightarrow{\mathbb{P}} 0$ as $n \to \infty$.

<u>Proof</u> Need $\forall \varepsilon > 0$ $\mathbb{P}(|X_n| > \varepsilon) \to 0$ as $n \to \infty$ or equivalently $\mathbb{P}(|X_n| \leq \varepsilon) \to 1$ as $n \to \infty$.

$$\mathbb{P}(|X_n| \leq \varepsilon) \geq \mathbb{P}\Big(\underbrace{\bigcap_{m=n}^{\infty} \{|X_m| \leq \varepsilon\}}_{A_n}\Big) \qquad A_n \subseteq A_{n+1}$$

$\bigcup_n A_n = \{|X_m| \leq \varepsilon$ for all m sufficiently large $\}$.

So $\mathbb{P}(A_n) \nearrow \mathbb{P}(\bigcup_n A_n)$ as $n \to \infty$

so $\lim_{n\to\infty} \mathbb{P}(|X_n| \leq \varepsilon) \geq \lim_{n \to \infty} \mathbb{P}(A_n) = \mathbb{P}(\bigcup A_n) \geq \mathbb{P}(\lim_{n\to\infty} X_n = 0) = 1$ $\square$

<u>Theorem</u> (Strong law of large numbers)

Let $(X_n)_{n \in \mathbb{N}}$ be an iid sequence of RVs, $\mu = \mathbb{E}(X_1) < \infty$

Then setting $S_n = X_1 + \dots + X_n$, have $\frac{S_n}{n} \to \mu$ as $n \to \infty$ almost surely.

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

<u>Proof</u> (non-examinable) - assume further that $\mathbb{E}(X_1^4) < \infty$.

Set $Y_i = X_i - \mu$, then $\mathbb{E}(Y_i) = 0$, $\mathbb{E}(Y_1^4) \leq 2^4(\mathbb{E}(X_1^4) + \mu^4) < \infty$

It suffices to prove $\frac{S_n}{n} \to 0$ where $S_n = \sum_{i=1}^n X_i$ with $\mathbb{E}(X_i) = 0$, $\mathbb{E}(X_i^4) < \infty$

$$S_n^4 = \left(\sum_{i=1}^n X_i\right)^4 = \sum_{i=1}^n X_i^4 + \binom{4}{2} \sum_{1 \leq i < j \leq n} X_i^2 X_j^2 + R$$

where R is a sum of terms of form $X_i^2 X_j X_k$, $X_i^3 X_j$, $X_i X_j X_k X_l$

$$\mathbb{E}(S_n^4) = n\,\mathbb{E}(X_1^4) + \binom{4}{2} \frac{n(n-1)}{2} \mathbb{E}(X_1^2 X_2^2) + \underbrace{\cancel{\mathbb{E}(R)}}_{0} \quad \mathbb{E}(X_i) = 0$$

so $\mathbb{E}(S_n^4) \leq n\,\mathbb{E}(X_1^4) + 3n(n-1)\,\mathbb{E}(X_1^4)$ (Jensen)

$\mathbb{E}(S_n^4) \leq 3n^2\, \mathbb{E}(X_1^4)$

so $\mathbb{E}\left(\sum_{n=1}^\infty \left(\frac{S_n}{n}\right)^4\right) \leq \sum_{n=1}^\infty \frac{3}{n^2} \mathbb{E}(X_1^4) < \infty \Rightarrow \sum_{n=1}^\infty \left(\frac{S_n}{n}\right)^4 < \infty$ w.p.1

$\Rightarrow \frac{S_n}{n} \to 0$ as $n \to \infty$ w.p.1. $\square$

Suppose $\mathbb{E}(X_1) = \mu$, $Var(X_1) = \sigma^2 < \infty$

$$Var\left(\frac{S_n}{n} - \mu\right) = \frac{\sigma^2}{n} \quad \text{so} \quad \frac{\frac{S_n}{n} - \mu}{\sqrt{Var(\frac{S_n}{n} - \mu)}} = \frac{\frac{S_n}{n} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

<u>Central limit theorem</u> Let $(X_n)_{n \in \mathbb{N}}$ be iid RVs, $\mathbb{E}(X_1) = \mu$, $Var(X_1) = \sigma^2$

Set $S_n = X_1 + \dots + X_n$. Then

$$\forall x \in \mathbb{R}, \quad \mathbb{P}\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \le x\right) \to \Phi(x) = \int_{-\infty}^{x} \frac{e^{-y^2/2}}{\sqrt{2\pi}}\, dy \quad \text{as } n \to \infty$$

In other words $\dfrac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow[n\to\infty]{(d)} Z$ where $Z \sim N(0,1)$.

CLT says for n large enough, $\dfrac{S_n - n\mu}{\sigma\sqrt{n}} \approx Z$

$$\Rightarrow \quad S_n \approx n\mu + \sigma\sqrt{n}\, Z \sim N(n\mu, \sigma^2 n)$$

<u>Proof</u> Consider $Y_i = \dfrac{X_i - \mu}{\sigma}$ then $\mathbb{E}(Y_i) = 0$, $Var(Y_i) = 1$

Suffices to prove CLT with $S_n = X_1 + \dots + X_n$, $\mathbb{E}(X_i) = 0$, $Var(X_i) = 1$

Assume $\exists\, \delta > 0$ s.t. $\mathbb{E}(e^{\delta X_1}) < \infty$, $\mathbb{E}(e^{-\delta X_1}) < \infty$

Need to show $\dfrac{S_n}{\sqrt{n}} \xrightarrow{d} N(0,1)$ as $n \to \infty$.

By continuity property for mgf, suffices to show $\forall \theta \in \mathbb{R}$,

$$\mathbb{E}\left(e^{\theta S_n/\sqrt{n}}\right) \xrightarrow[n\to\infty]{} \mathbb{E}(e^{\theta Z}) = e^{\theta^2/2}$$

Set $m(\theta) = \mathbb{E}(e^{\theta X_1})$. Then $\mathbb{E}\left(e^{\theta S_n/\sqrt{n}}\right) = \mathbb{E}\left(e^{(\theta/\sqrt{n})X_1}\right)^n$

$= \left(m\left(\frac{\theta}{\sqrt{n}}\right)\right)^n$. Need to show $\left(m\left(\frac{\theta}{\sqrt{n}}\right)\right)^n \to e^{\theta^2/2}$ as $n \to \infty$.

$$|\theta| \le \frac{\delta}{2} \quad m(\theta) = \mathbb{E}\left(e^{\theta X_1}\right) = \mathbb{E}\left(1 + \theta X_1 + \frac{\theta^2 X_1^2}{2!} + \sum_{k=3}^{\infty} \frac{\theta^k X_1^k}{k!}\right)$$

$$\text{so} \quad m(\theta) = 1 + \frac{\theta^2}{2} + \underbrace{\mathbb{E}\left(\sum_{k \ge 3} \frac{\theta^k X_1^k}{k!}\right)}$$

<u>Claim</u> Suffices to prove that magnitude of this is $o(|\theta|^2)$ as $\theta \to 0$

Once we prove this, then $m\left(\frac{\theta}{\sqrt{n}}\right) = 1 + \frac{\theta^2}{2n} + o\left(\frac{|\theta|^2}{n}\right)$

$$\Rightarrow \left(m\left(\frac{\theta}{\sqrt{n}}\right)\right)^n \to e^{\theta^2/2} \quad \text{as } n \to \infty.$$

Proof that $\left| \mathbb{E}\left(\sum_{k \geq 3} \frac{\theta^k X_1^k}{k!} \right) \right| = o(|\theta|^2)$ as $\theta \to 0$:

$$\left| \mathbb{E}\left(\sum_{k\geq 3} \frac{\theta^k X_1^k}{k!} \right) \right| \leq \mathbb{E}\left(\sum_{k=3}^{\infty} \frac{|\theta|^k |X_1|^k}{k!} \right) = \cancel{|\theta X_1|^3} \; \mathbb{E}\left(|\theta X_1|^3 \sum_{k=0}^{\infty} \frac{|\theta X_1|^k}{(k+3)!} \right)$$

$$\leq \mathbb{E}\left(|\theta X_1|^3 \sum_{k=0}^{\infty} \frac{|\theta X_1|^k}{k!} \right) \leq \mathbb{E}\left(|\theta X_1|^3 e^{\frac{\delta}{2}|X_1|} \right) \quad |\theta| \leq \frac{\delta}{2}$$

$$|\theta X_1|^3 e^{\delta |X_1|/2} = |\theta|^3 \frac{\left(\frac{\delta}{2}|X_1|\right)^3}{3!} \cdot \frac{3!}{\left(\frac{\delta}{2}\right)^3} e^{\delta |X_1|/2}$$

$$\leq \frac{3!\,|\theta|^3}{\left(\frac{\delta}{2}\right)^3} e^{\delta |X_1|} = 3! \left(\frac{2|\theta|}{\delta} \right)^3 e^{\delta |X_1|}$$

$e^{\delta |X_1|} \leq e^{\delta X_1} + e^{-\delta X_1}$ so

$$\left| \mathbb{E}\left(\sum_{k\geq 3} \frac{\theta^k X_1^k}{k!} \right) \right| \leq 3! \left(\frac{2|\theta|}{\delta} \right)^3 \underbrace{\mathbb{E}\left(e^{\delta X_1} + e^{-\delta X_1} \right)}_{<\infty} = o(|\theta|^2) \text{ as } \theta \to 0$$

$\square$

<u>Applications</u> 1) Normal approximation to binomial

$S_n \sim \mathrm{Bin}(n,p)$: $S_n = \sum_{i=1}^{n} X_i$, (X_i) iid $\sim \mathrm{Ber}(p)$ $\quad \mathbb{E}(S_n) = np$, $\mathrm{Var}(S_n) = np(1-p)$

so by CLT $\quad \frac{S_n - np}{\sqrt{np(1-p)}} \xrightarrow{d} N(0,1)$ as $n \to \infty$

so $S_n \approx N(np, np(1-p))$ for n large.

$\mathrm{Bin}\left(n, \frac{\lambda}{n}\right) \to \mathrm{Poi}(\lambda)$

$\lambda > 0$ $\quad$ p depends on n here

2) Normal approximation to Poisson

Let $S_n \sim \mathrm{Poi}(n)$ $\quad S_n = \sum_{i=1}^{n} X_i$, (X_i) iid $\sim \mathrm{Poi}(\lambda)$

$\frac{S_n - n}{\sqrt{n}} \xrightarrow{d} N(0,1)$ as $n \to \infty$.

Probability - Lecture 21

13/3/21

Sampling error via CLT

Pick N random individuals, let $\hat{p}_N = \frac{S_N}{N}$, $S_N = \#$ of "yes" voters

How large should N be so that $|\hat{p}_N - p| \leq \frac{4}{100}$ with prob ≥ 0.99?

$S_N \sim \text{Bin}(N, p)$, by CLT $\quad S_N \simeq Np + \sqrt{Np(1-p)}\, Z \quad Z \sim N(0,1)$

So $\underbrace{\frac{S_N}{N}}_{\hat{p}_N} \simeq p + \sqrt{\frac{p(1-p)}{N}}\, Z \implies |\hat{p}_N - p| \simeq \sqrt{\frac{p(1-p)}{N}} \cdot |Z|$

Find N s.t. $\mathbb{P}(|\hat{p}_N - p| \leq 0.04) \geq 0.99$ or equivalently

$\mathbb{P}\left(\sqrt{\frac{p(1-p)}{N}}\, |Z| \leq 0.04\right) \geq 0.99$

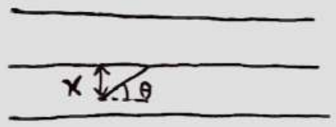
From tables $z = 2.58$: $\mathbb{P}(|Z| \geq 2.58) = 0.01$

$\left[\forall z \in \mathbb{R}, \quad \mathbb{P}(|Z| \geq z) = 2(1 - \Phi(z))\right]$

So we need $\quad 0.04 \sqrt{\frac{N}{p(1-p)}} \geq 2.58 \implies \underline{N \geq 1040}$

$\quad p = \frac{1}{2} \to \frac{1}{4}$ max $p(1-p)$

Buffon's needle



parallel lines, distance L apart

needle $\ell \leq L$

Throw needle at random - what is $\mathbb{P}$(it intersects ≥ 1 line) ?

$\theta \sim U[0, \pi]$, $X \sim U[0, L]$ indep.

It intersects a line iff $\quad X \leq \ell \sin\theta$

$\mathbb{P}(\text{intersection}) = \mathbb{P}(X \leq \ell \sin\theta) = \int_0^L \int_0^\pi \frac{1}{\pi L} 1(x \leq \ell \sin\theta)\, dx\, d\theta$

$= \underline{\frac{2\ell}{\pi L} = p}$

$\implies \pi = \frac{2\ell}{pL}$

Want to use this expt to approximate π.

Throw n needles indep, let $\hat{p}_n$ be proportion intersecting a line

Then $\hat{p}_n$ approximates p and $\hat{\pi}_n = \frac{2\ell}{\hat{p}_n L}$ approximates π.

Suppose $\mathbb{P}(|\hat{\pi}_n - \pi| \le 0.001) \geqslant 0.99$. How large should n be?

S_n = # needles intersecting a line, $S_n \sim Bin(n,p)$: CLT $S_n \approx np + \sqrt{np(1-p)}\, Z$

$\Rightarrow \frac{S_n}{n} \approx p + \sqrt{\frac{p(1-p)}{n}}\, Z$ so $\hat{p}_n - p \approx \sqrt{\frac{p(1-p)}{n}}\, Z$

$f(x) := \frac{2l}{xL} \Rightarrow f(p) = \pi,\ f'(p) = -\frac{\pi}{p},\quad \hat{\pi}_n = f(\hat{p}_n)$

$\hat{\pi}_n = f(\hat{p}_n) \approx f(p) + (\hat{p}_n - p) f'(p) \Rightarrow \hat{\pi}_n \approx \pi (\hat{p}_n - p) \cdot \frac{\pi}{p}$

$\Rightarrow \hat{\pi}_n - \pi \approx -\frac{\pi}{p}\sqrt{\frac{p(1-p)}{n}} = -\pi \sqrt{\frac{1-p}{pn}}\, Z$

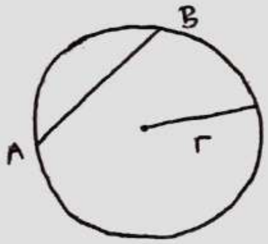
We want

$$\mathbb{P}\left(\pi \sqrt{\frac{1-p}{pn}} \cdot |Z| \leqslant 0.001\right) \geqslant 0.99$$

$\mathbb{P}(|Z| \geqslant 2.58) = 0.01$ and ~~so~~ we have $\frac{\pi^2(1-p)}{pn}$ decreasing in p – minimise by taking $l = L \Rightarrow p = \frac{2}{\pi}$, $Var = \frac{\pi^2}{n}\left(\frac{\pi}{2} - 1\right)$

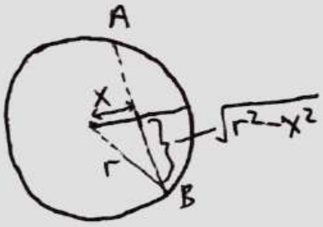
Taking $\sqrt{\frac{\pi^2}{n}\left(\frac{\pi}{2}-1\right)} \cdot 2.58 = 0.001 \Rightarrow \underline{n \approx 3.75 \times 10^7}$

Bertrand's paradox



Draw random chord. What is probability that $AB \leqslant r$?

Interpretation 1 Take $X \sim U[0, r]$



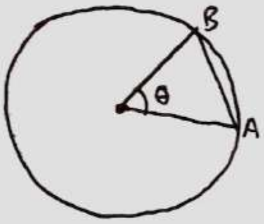
Let $C = |AB|$ $\quad C = 2\sqrt{r^2 - X^2}$

$\mathbb{P}(C \leqslant r) = \mathbb{P}(2\sqrt{r^2 - X^2} \leqslant r)$

$= \mathbb{P}(4(r^2 - X^2) \leq r^2)$

$= \mathbb{P}(4X^2 \geqslant 3r^2) = \mathbb{P}(X \geqslant \sqrt{3}r/2) = 1 - \frac{\sqrt{3}}{2} \approx 0.134$ (uniform X)

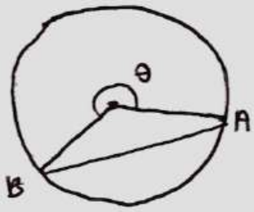
Interpretation 2



Let $\theta \sim U[0, 2\pi]$ and join A and B

$C = |AB|$

If $\theta \in [0, \pi] \Rightarrow C = 2r \sin \frac{\theta}{2}$



If $\theta \in [\pi, 2\pi]$:

$C = 2r \sin\left(\frac{2\pi - \theta}{2}\right) = 2r \sin \frac{\theta}{2}$

$$\mathbb{P}(C \leq r) = \mathbb{P}\left(2r \sin \tfrac{\theta}{2} \leq r\right) = \mathbb{P}\left(\sin \tfrac{\theta}{2} \leq \tfrac{1}{2}\right)$$
$$= \mathbb{P}\left(\theta \leq \tfrac{\pi}{3}\right) + \mathbb{P}\left(\theta \geq \tfrac{5\pi}{3}\right) = \tfrac{1}{6} + \tfrac{1}{6} = \underline{\tfrac{1}{3}}$$

Multidimensional Gaussian RVs

A RV with values in $\mathbb{R}$ is Gaussian/normal if $X = \mu + \sigma Z$ $(Z \sim N(0,1))$

The density of X is $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$, $x \in \mathbb{R}$

$X \sim N(\mu, \sigma^2)$. Let $X = (X_1, \dots, X_n)^T \in \mathbb{R}^n$. Then X is a <u>Gaussian vector</u> if $\forall u = (u_1, \dots, u_n)^T \in \mathbb{R}^n$,

$u^T X = \sum_{i=1}^{n} u_i X_i$ is a Gaussian RV in $\mathbb{R}$.

<u>Example</u> X Gaussian in $\mathbb{R}^n$, suppose A is an $m \times n$ matrix and $b \in \mathbb{R}^m$. Then $AX + b$ is also Gaussian in $\mathbb{R}^m$.

Let $u \in \mathbb{R}^m$. Then $u^T(Ax + b) = (u^T A)x + u^T b$

Set $v = A^T u$, then $u^T(AX + b) = v^T X + u^T b$

Since X is Gaussian, so is $v^T X$ and hence $\underline{v^T X + u^T b}$.

Let $X = (X_1, \dots, X_n) \in \mathbb{R}^n$. It is Gaussian if $\forall u \in \mathbb{R}^n$,

$$u^T X = \sum_{i=1}^n u_i X_i$$ has the normal distribution in $\mathbb{R}$.

$$\mu = \mathbb{E}(X) = \begin{pmatrix} \mathbb{E}(X_1) \\ \vdots \\ \mathbb{E}(X_n) \end{pmatrix} \qquad \mu_i = \mathbb{E}(X_i)$$

$$V = \mathrm{Var}(X) = \underset{(n \times n)}{\mathbb{E}\left((X-\mu) \cdot (X-\mu)^T\right)} = \begin{pmatrix} \mathbb{E}((X_i - \mu_i)(X_j - \mu_j)) = \mathrm{Cov}(X_i, X_j) \end{pmatrix}$$

V is a symmetric matrix.

$$\mathbb{E}(u^T X) = \mathbb{E}\left(\sum_{i=1}^n u_i X_i\right) = \sum_{i=1}^n u_i \mu_i = u^T \mu$$

$$\mathrm{Var}(u^T X) = \mathrm{Var}\left(\sum_{i=1}^n u_i X_i\right) = \sum_{i,j=1}^n u_i \mathrm{Cov}(X_i, X_j) u_j = u^T V u$$

So $\underline{u^T X \sim N(u^T \mu, u^T V u)}$

<u>Claim</u> V is a non-negative definite matrix ($\forall u \in \mathbb{R}^n, u^T V u \geq 0$)

<u>Proof</u> Let $u \in \mathbb{R}^n$. Then $\mathrm{Var}(u^T X) = u^T V u$. Since $\mathrm{Var}(u^T X) \geq 0$, $u^T V u \geq 0$. $\square$

<u>Mgf of X</u> is $\mathbb{E}(e^{\lambda^T X}) = m(\lambda) \quad \forall \lambda \in \mathbb{R}^n$

$m(\lambda) = \mathbb{E}(e^{\lambda^T X})$. We know $\lambda^T X \sim N(\lambda^T \mu, \lambda^T V \lambda)$

$m(\lambda) = e^{\lambda^T \mu + \frac{\lambda^T V \lambda}{2}}$ characterised by μ and V, write $X \sim N(\mu, V)$

<u>Construction</u> Let $Z_1, \dots, Z_n$ iid $N(0,1)$ RVs

Set $Z = (Z_1, \dots, Z_n)^T$. Then Z is a Gaussian vector.

<u>Proof</u> $\forall u \in \mathbb{R}^n$, $u^T Z$ is Gaussian.

$$u^T Z = \sum_{i=1}^n u_i Z_i$$ - need to show $\sum_{i=1}^n u_i Z_i$ is normal.

Let $\lambda \in \mathbb{R}$. $\mathbb{E}\left(e^{\lambda \sum_{i=1}^n u_i Z_i}\right) = \mathbb{E}\left(\prod_{i=1}^n e^{\lambda u_i Z_i}\right)$

$$\overset{\text{indep}}{=} \prod_{i=1}^{n} \mathbb{E}\left(e^{\lambda u_i z_i}\right) = \prod_{i=1}^{n} e^{(\lambda u_i)^2/2} = e^{\lambda^2 |u|^2/2}$$

So $u^T Z \sim N(0, |u|^2)$. $\square$

$\mathbb{E}(Z) = 0$ $\quad$ $\mathrm{Var}(Z) = I_n$ $\quad$ so $Z \sim N(0, I_n)$.

Let $\mu \in \mathbb{R}^n$ and V a non-negative definite matrix. Want to construct ~~a~~ Gaussian vector with mean μ, variance V using Z.

$\underline{n=1}$ μ, σ^2 If $Z \sim N(0,1)$ then $\mu + \sigma Z \sim N(\mu, \sigma^2)$

Since V is non-negative definite, $V = U^T D U$ with $U^{-1} = U^T$

and $D = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$ and $\lambda_i \geq 0$ $\forall i$

Define sqrt of V by $\sigma = U^T \sqrt{D}\, U$ where $\sqrt{D} = \begin{pmatrix} \sqrt{\lambda_1} & & 0 \\ & \ddots & \\ 0 & & \sqrt{\lambda_n} \end{pmatrix}$

Indeed $\sigma \cdot \sigma = U^T \sqrt{D}\, U \cdot U^T \sqrt{D}\, U = U^T D U = V$.

Let $Z = (z_1, \dots, z_n)$ with (z_i) iid $N(0,1)$ RVs. Set $X = \mu + \sigma Z$.

__Claim__ $X \sim N(\mu, V)$

__Proof__ X is Gaussian as it's a linear transformation of Z.

$\mathbb{E}(X) = \mu$ and $\mathrm{Var}(X) = \mathbb{E}\left((X-\mu)(X-\mu)^T\right)$

$= \mathbb{E}\left((\sigma Z)\cdot(\sigma Z)^T\right) = \mathbb{E}\left(\sigma Z \cdot Z^T \cdot \sigma^T\right) = \sigma \mathbb{E}(Z \cdot Z^T)\sigma$

$= \sigma \cdot I_n \cdot \sigma = \sigma \cdot \sigma = V$. $\square$

__Density of $X \sim N(\mu, V)$__ $\qquad$ $n=1$ $X \sim N(\mu, \sigma^2)$ $f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)$

V is positive definite $\quad \lambda_i > 0$ $\forall i$

$X = \mu + \sigma Z$, $\quad Z \sim N(0, I_n)$

$f_X(x) = f_Z(z) \cdot |J|$ $\qquad x = \mu + \sigma z$

Since V is positive definite, σ is invertible, so

$x = \mu + \sigma z \Rightarrow z = \sigma^{-1}(x - \mu)$ so $f_X(x) = f_Z(z) \cdot |J|$

$$= \prod_{i=1}^{n} \left(\frac{e^{-\frac{z_i^2}{2}}}{\sqrt{2\pi}} \right) \cdot |\det \sigma^{-1}|$$

$$\Rightarrow \quad f_X(x) = \frac{1}{(2\pi)^{n/2}} e^{-|z|^2/2} \frac{1}{\sqrt{\det V}} = \frac{1}{\sqrt{(2\pi)^n \det V}} e^{-z^T z/2}$$

$$z^T \cdot z = (\sigma^{-1}(x-\mu))^T (\sigma^{-1}(x-\mu))$$

$$= (x-\mu)^T \underbrace{(\sigma^{-1})^T}_{\sigma^{-1}} \sigma^{-1} \cdot (X-\mu) = (x-\mu)^T \cdot \sigma^{-1} \cdot \sigma^{-1} \cdot (x-\mu)$$

$$= (x-\mu)^T \cdot V^{-1} \cdot (x-\mu)$$

Hence we have $\quad f_X(x) = \dfrac{1}{\sqrt{(2\pi)^n \det V}} \cdot \exp\left(- \dfrac{(x-\mu)^T \cdot V^{-1}(x-\mu)}{2} \right)$

If V is non-negative definite (some eigenvalues could be 0).

By an orthogonal change of basis we can assume

$V = \begin{pmatrix} U & 0 \\ 0 & 0 \end{pmatrix}$, $\quad U$ $m \times n$ positive definite matrix, $\quad m < n$

and $\quad \mu = \begin{pmatrix} \lambda \\ \nu \end{pmatrix}$, $\quad \lambda \in \mathbb{R}^m$, $\nu \in \mathbb{R}^{n-m}$

We can write $\quad X = \begin{pmatrix} Y \\ \nu \end{pmatrix}$ where Y has density

$$f_Y(y) = \frac{1}{\sqrt{(2\pi)^m \det U}} \exp\left(- \frac{(y-\lambda)^T \cdot U^{-1}(y-\lambda)}{2} \right).$$

Probability - Lecture 23 22/3/21

Claim If X_i's independent, then V is a diagonal matrix.

Proof Since X_i's are indep., follows that $\mathrm{Cov}(X_i, X_j) = 0$ for $i \neq j$ so V is diagonal. $\square$

Lemma Suppose that X is a Gaussian vector. Then if V is diagonal, then X_i's are independent.

Proof If V is diagonal, then the density $f_X(x)$ factorises into product:

$$(x-\mu)^T V^{-1}(x-\mu) = \sum_{i=1}^n \frac{(x_i-\mu_i)^2}{\lambda_i}$$

so $$f_X(x) = \frac{1}{\sqrt{(2\pi)^n \det V}} \exp\left(-\sum_{i=1}^n \frac{(x_i-\mu_i)^2}{2\lambda_i}\right)$$ so X_is are indep.

Proof 2 $$m(\theta) = \mathbb{E}\left(e^{\theta^T X}\right) = e^{\theta^T \mu + \frac{\theta^T V \lambda}{2}} = e^{\sum \theta_i \mu_i} e^{\sum \theta_i^2 \lambda_i / 2}$$

$m(\theta)$ factorises into mgfs of Gaussian RVs in $\mathbb{R}$. $\square$

So for Gaussian vectors, have

$(X_1, \dots, X_n)$ are indep. iff $\mathrm{Cov}(X_i, X_j) = 0$ whenever $i \neq j$.

Bivariate Gaussian $n=2$:

Let $X = (X_1, X_2)$ be a Gaussian vector in $\mathbb{R}^2$.

Set $\mu_k = \mathbb{E}(X_k)$, $\sigma_k^2 = \mathrm{Var}(X_k)$, $\rho = \mathrm{Corr}(X_1, X_2) = \frac{\mathrm{Cov}(X_1, X_2)}{\sqrt{\mathrm{Var} X_1 \mathrm{Var} X_2}}$

Claim $\rho \in [-1, 1]$

Proof immediate from Cauchy-Schwarz inequality

$$V = \mathrm{Var}(X) = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

Claim For all $\sigma_k > 0$, $\rho \in [-1, 1]$, V is non-negative definite.

Proof Let $u \in \mathbb{R}^2$, $u^T V u = (1-\rho)(\sigma_1^2 u_1^2 + \sigma_2^2 u_2^2) + \rho(\sigma_1 u_1 + \sigma_2 u_2)^2$

$$= (1+\rho)(\sigma_1^2 u_1^2 + \sigma_2^2 u_2^2) - \rho(\sigma_1 u_1 - \sigma_2 u_2)^2$$

$$\geq 0 \quad \forall \rho \in [-1,1]. \qquad \square$$

When $\rho = 0$, $\sigma_1, \sigma_2 > 0$ then $f_{X_1,X_2}(x_1,x_2)$ is equal to

$$\prod_{k=1}^{2} \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left(-\frac{(x_k-\mu_k)^2}{2\sigma_k^2}\right)$$ so X_1, X_2 indep. in this case.

More generally, suppose (X_1, X_2) is a Gaussian vector.

Let $a \in \mathbb{R}$, consider $X_2 - aX_1$.

$$\mathrm{Cov}(X_2 - aX_1, X_1) = \mathrm{Cov}(X_2, X_1) - a\,\mathrm{Cov}(X_1, X_1)$$

$$= \mathrm{Cov}(X_1, X_2) - a\,\mathrm{Var}(X_1) = \rho\sigma_1\sigma_2 - a\sigma_1^2$$

Take $a = \frac{\rho\sigma_2}{\sigma_1}$, then $\mathrm{Cov}(X_2 - aX_1, X_1) = 0$

Set $Y = X_2 - aX_1$. Then (X_1, Y) is a Gaussian vector:

$$\begin{pmatrix} X_1 \\ Y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -a & 1 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \Rightarrow \begin{pmatrix} X_1 \\ Y \end{pmatrix} \text{ is of form } A \underbrace{\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}}_{\text{Gaussian.}} \qquad \square$$

From indep. criterion ~~of i~~ we get X_1 indep. of Y.

So we can express $X_2 = X_2 - aX_1 + aX_1 = Y + aX_1$

and $\mathbb{E}(X_2 | X_1) = \mathbb{E}(Y + aX_1 | X_1) = \mathbb{E}(Y|X_1) + a\,\mathbb{E}(X_1|X_1)$

$Y \perp\!\!\!\perp X_1$, so $\mathbb{E}(X_2|X_1) = \mathbb{E}(Y) + aX_1$

(X_1, X_2) Gaussian, $(X_2 - aX_1, X_1)$ Gaussian, $X_2 - aX_1 \perp\!\!\!\perp X_1$

$X_2 = X_2 - aX_1 + aX_1$ so given X_1, X_2 is normal with

$$X_2 \sim N(aX_1 + \mu_2 - a\mu_1, \mathrm{Var}(X_2 - aX_1))$$

where $\mathrm{Var}(X_2 - aX_1) = \mathrm{Var}(X_2) + a^2\,\mathrm{Var}(X_1) - 2a\,\mathrm{Cov}(X_1, X_2)$.

See lecture notes for non-examinable multivariate CLT.

Probability - Lecture 24

23/3/21

Example Let $U \sim U[0,1]$, set $X = -\log U$

$$\mathbb{P}(X \leq x) = \mathbb{P}(-\log U \leq x) = \mathbb{P}(U \geq e^{-x}) = 1 - e^{-x}$$

so $X \sim \mathrm{Exp}(1)$.

Theorem Let X be a cts RV with distribution function F. Then if U is uniform on $[0,1]$, then $F^{-1}(U) \sim F$.

Proof Set $Y = F^{-1}(U)$

$$\mathbb{P}(Y \leq x) = \mathbb{P}(F^{-1}(U) \leq x) = \mathbb{P}(U \leq F(x)) = F(x). \qquad \square$$

Rejection sampling Suppose $A \subset [0,1]^d$. Define $f(x) = \dfrac{1(x \in A)}{|A|}$
($|A|$ = volume of A)

Let X have density f. How can we simulate f?

Let $(U_n)_{n \in \mathbb{N}}$ be an iid sequence of d-dimensional uniforms:

$$U_n = (\underbrace{U_{k,n}}_{\text{both subscripted}} : k \in \{1, \dots, d\}) \qquad (U_{k,n})_{(k,n)} \text{ iid } \sim U[0,1]$$

Let $N = \min\{n \geq 1 : U_n \in A\}$

Claim $U_N \sim f$

Proof We want to show that $\forall B \subseteq [0,1]^d$,

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

$$\mathbb{P}(U_N \in B) = \sum_{n=1}^{\infty} \mathbb{P}(U_N \in B, N = n) = \sum_{n=1}^{\infty} \mathbb{P}(U_n \in A \cap B, U_{n-1} \notin A, \dots, U_1 \notin A)$$

$$\underset{\substack{\text{indep.}\\ U_i\text{s}}}{=} \sum_{n=1}^{\infty} \mathbb{P}(U_n \in A \cap B)\,\mathbb{P}(U_{n-1} \notin A) \dots \mathbb{P}(U_1 \notin A)$$

$$= \sum_{n=1}^{\infty} |A \cap B|(1 - |A|)^{n-1} = \frac{|A \cap B|}{|A|} = \int_A \frac{1(x \in B)}{|A|}\,dx = \int_B f(x)\,dx.$$

$\square$

Suppose f is a density on $[0,1]^{d-1}$ which is bounded.

Want to sample $X \sim f$.

Consider $A = \{ (x_1, \dots, x_d) \in [0,1]^d : x_d \leq f(x_1, \dots, x_{d-1}) / \lambda \}$

where $f(x) \leq \lambda$ (bounded) $\forall x \in [0,1]^{d-1}$

Set $X = (X_1, \dots, X_{d-1})$

<u>Claim</u> $X \sim f$

<u>Proof</u> We need to show that $\forall B \subseteq [0,1]^{d-1}$

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

$$\mathbb{P}(X \in B) = \mathbb{P}((X_1, \dots, X_{d-1}) \in B) = \mathbb{P}\big((X_1, \dots, X_d) \in (B \times [0,1] \cap A)\big)$$

$$= \frac{|(B \times [0,1]) \cap A|}{|A|} \quad \text{since } Y \text{ is uniform on } A$$

$$|(B \times [0,1]) \cap A| = \int \cdots \int 1\big((x_1, \dots, x_d) \in B \times [0,1] \cap A\big)\, dx_1 \dots dx_d$$

$$= \int \cdots \int 1\big((x_1, \dots, x_{d-1}) \in B\big)\, 1\left(x_d \leq \frac{f(x_1, \dots, x_{d-1})}{\lambda}\right) dx_1 \dots dx_d$$

$$= \int \cdots \int 1\big((x_1, \dots, x_{d-1}) \in B\big)\, \frac{f(x_1, \dots, x_{d-1})}{\lambda}\, dx_1 \dots dx_{d-1}$$

$$= \frac{1}{\lambda} \int_B f(x)\, dx$$

$$|A| = \frac{1}{\lambda} \int_{[0,1]^{d-1}} f(x)\, dx = \frac{1}{\lambda}$$

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