LECTURE 1

Brief Summary:

- **Basic Objects of Probability:**

① Outcome space Ω: (should be clearly defined in each situation)
   $\omega \in \Omega$

② Event: $A \subseteq \Omega$ is a subset of Ω

③ Probability Measure $\mathbb{P}$: a function from the space of A to $[0,1]$
   (different from probability mass function $p: \Omega \to [0,1]$).
   ☆ $\mathbb{P}(A) = \sum_{\omega \in A} p(\omega)$.

In particular, $\mathbb{P}$ must satisfy three properties:

① $0 \le \mathbb{P}(A) \le 1$ for all events $A \subseteq \Omega$

② If $A \subseteq \Omega, B \subseteq \Omega, A \cap B = \phi$, then $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$

③ $\mathbb{P}(\Omega) = 1$

=> A stronger version: countable additivity:

If $A \subseteq \Omega$, $B \subseteq \Omega$, then $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$

$\Downarrow$ further strengthen

$\forall \{A_k\}_{k=1}^{\infty}$, $A_i \cap A_j = \phi$ $\forall i \ne j$:

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

Some other equalities can be derived using these basic properties.

e.g. $\mathbb{P}(\Omega^c)$: $\Omega \cap \Omega^c = \phi$, so

$$\mathbb{P}(\Omega^c \cup \Omega) = \mathbb{P}(\Omega) = 1 = \mathbb{P}(\Omega) + \mathbb{P}(\Omega^c)$$
$$1 = 1 + \mathbb{P}(\Omega^c)$$
$$\mathbb{P}(\Omega^c) = 0$$

Aside: important counting principles:

①. the # of sequences of length k with n symbols: n^k.

②. the # of ordered subsets of k elements from a set with n elements: $\frac{n!}{(n-k)!}$

③. the # of ways to order a set of n objects: $n!$

④ the # of subsets of k elements from a set with n elements: $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ (n choose k).

Questions:

Regarding the game of cycles, I came up with a hypothesis that:
if the two players are both rational and choose the best strategy under any circumstances, then: when the # of nodes is odd, the first mover wins;
when the # of nodes is even, the second mover wins,
regardless of the different connections (edges). (This is true for games with 2, 3, and 4 nodes as I tried).

Is it correct? How can I prove or disprove it?

LECTURE 2

Brief Summary:

- Discrete Probability: Ω is finite or countably infinite

- probability mass function: $p: \Omega \to [0,1]$
  $p(\omega) = \mathbb{P}(\{\omega\})$

- For any event $A \subseteq \Omega$: $\mathbb{P}(A) = \sum_{\omega \in A} p(\omega)$
  pf: $\mathbb{P}(A) = \mathbb{P}(\bigcup_{k=1}^{\infty} \{\omega_k\}) = \sum_{k=1}^{\infty} \mathbb{P}(\{\omega_k\}) = \sum_{k=1}^{\infty} p(\omega)$

- "Equally likely outcomes" Ω. for any $A \subseteq \Omega$, $\mathbb{P}(A) = \frac{\#A}{\#\Omega}$; for any $\omega \in \Omega$, $p(\omega) = \frac{1}{\#\Omega}$
  pf: $\mathbb{P}(\Omega) = 1$
  $\mathbb{P}(\Omega) = \mathbb{P}(\bigcup_{k=1}^{\infty} \{\omega_k\}) = \sum_{k=1}^{\infty} \mathbb{P}(\{\omega_k\}) = \sum_{k=1}^{\infty} p(\omega_k)$
  let $p(\omega_k) = c$
  then: $1 = \#\Omega \cdot c$
  $c = \frac{1}{\#\Omega}$

Important Example

Transforming the intuitive idea to mathematical model:

Background: Toss coin N times.

$\Omega = \{H, T\}^N$ $\quad \#\Omega = 2^N$

Let B_j = event that H occurs on j^{th} toss.

ω is the form: $\omega = (\cdots, \underset{j}{H}, \cdots)$ $\quad \omega \in B_j$.

Case 1: Two headed Coin.

let $\bar{\omega} = (H, H, \cdots, H)$

Define $p(\omega) = \begin{cases} 1 & \text{if } \omega = \bar{\omega} \\ 0 & \text{otherwise} \end{cases}$: (this is 100% surely to happen, intuitively)

Then: $\forall A \subseteq \Omega$: $\mathbb{P}(A) = \sum_{\omega \in A} p(\omega) = \begin{cases} 1 & \text{if } \bar{\omega} \in A \\ 0 & \text{if } \bar{\omega} \notin A \end{cases}$

$$\mathbb{P}(B_j) = \begin{cases} 1 & \text{if } \bar{\omega} \in B_j \\ 0 & \text{if } \bar{\omega} \notin B_j \end{cases}$$

since $\bar{\omega} = (H, H \cdots H) \in B_j$, $\mathbb{P}(B_j) = 1$

Case 2: Fair Toss.

Assume that all outcomes are equally likely.

Define $p(\omega) = \frac{1}{\#\Omega}$, $\forall \omega \in \Omega$

In this case, $p(\omega) = \frac{1}{2^N}$

For any event $A \subseteq \Omega$: $\mathbb{P}(A) = \sum_{\omega \in A} p(\omega) = \frac{\#A}{\#\Omega} = \frac{\#A}{2^N}$.

then: $\mathbb{P}(B_j) = \frac{\#B_j}{2^N} = \frac{2^{N-1}}{2^N} = \frac{1}{2}$, align with the intuition.

Another event: for $k \in \{0, 1, 2, \cdots, N\}$, Let A_k = event that a total of k heads are tossed among N tosses.

then: $\#A_j$ = # of ways of selecting k objects from a total of N = $\binom{N}{k} = \frac{N!}{k!(N-k)!}$

$\mathbb{P}(A_j) = \frac{\#A_j}{\#\Omega} = \frac{\binom{N}{k}}{2^N} = \frac{N!}{2^N k!(N-k)!}$.

Questions:

I don't have questions for this lecture. Everything is clear to me.

LECTURE 3

Brief Summary

- Conditional Probability:

  Suppose $A, B \in \Omega$ are two events. Assume $\mathbb{P}(B) > 0$

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

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

  since $A \cap B = B \cap A$

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

  suppose $\mathbb{P}(A), \mathbb{P}(B) > 0$, then we have:

- Bayes' Rule:

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

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

- Partition Rule:

  Events $\{B_k\}_{k=1}^N$ are a partition of Ω if: i) B_i are disjoint $B_i \cap B_j = \emptyset \ \forall i, j$
  ii) $\bigcup_{i=1}^N B_i = \Omega$

  Likewise, Let $A \subseteq \Omega$ be any event. Then, the sets $\{A \cap B_k\}_{k=1}^N$ are partitions of A if
  i) $\{A \cap B_k\}_{k=1}^N$ are disjoint ; ii) $\bigcup_{k=1}^N A \cap B_k = A$

  Then, by additivity property:

  $$\mathbb{P}(A) = \mathbb{P}\left(\bigcup_{k=1}^N A \cap B_k\right) = \sum_{k=1}^N \mathbb{P}(A \cap B_k) = \sum_{k=1}^N \mathbb{P}(A|B_k)\mathbb{P}(B_k)$$

  Rewrite Bayes' Rule as:

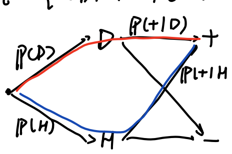
  $$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(A|B_i)\mathbb{P}(B_i)}{\sum_{k=1}^N \mathbb{P}(A|B_k)\mathbb{P}(B_k)}$$

Important Example:

illustrating Bayes' Rule using a tree diagram.

Suppose test is 90% accurate. Deseases occur randomly in 1% of population.
Suppose test positive. What is the probability of having the desease?

$\Omega = \{(D, +), (H, +), (D, -), (H, -)\}$.



$$\mathbb{P}(D|+) = \frac{\mathbb{P}(+|D)\mathbb{P}(D)}{\mathbb{P}(+|D)\mathbb{P}(D) + \mathbb{P}(+|H)\mathbb{P}(H)}$$

weight on the path

Questions:

Optimal strategy for the game before class: A says the opposite color he sees on B's hat, then B says the opposite color that A says.
By this way, B is sure to say the right color.

LECTURE 4

Basic Summary

- Independence

  Let $A_1 \cdots A_n$ be some events. The collection is independent if for any subset of indices $J \subseteq \{1, \cdots, n\}$

  $$\mathbb{P}\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in A} \mathbb{P}(A_j)$$

  If $A_1 \cdots A_n$ are independent, then any pair A_i, A_j are independent.

- If A and B are independent

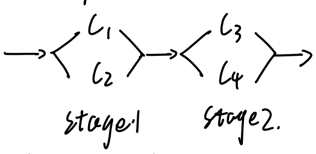
  $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$

  $\mathbb{P}(A \mid B) = \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \dfrac{\mathbb{P}(A)\mathbb{P}(B)}{\mathbb{P}(B)}$ assuming $\mathbb{P}(B) \neq 0$ $= \mathbb{P}(A)$

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

Important Examples.

1. Workflow



Stage i fails if only both component fail. Process fail if either stages fails.

Suppose $f_k = \mathbb{P}$(component C_k fails). What is $\mathbb{P}$(Process fails)?

Assume failures are independent.

Let A_k = event that component k fails

B = event that the process fails

Then $B = (A_1 \cap A_2) \cup (A_3 \cap A_4)$

$\mathbb{P}(B) = \mathbb{P}((A_1 \cap A_2) \cup (A_3 \cap A_4))$

$= \mathbb{P}((A_1 \cap A_2)) + \mathbb{P}(A_3 \cap A_4) - \mathbb{P}(A_1 \cap A_2 \cap A_3 \cap A_4)$

by independence:

$= \mathbb{P}(A_1)\mathbb{P}(A_2) + \mathbb{P}(A_3)\mathbb{P}(A_4) - \mathbb{P}(A_1)\mathbb{P}(A_2)\mathbb{P}(A_3)\mathbb{P}(A_4)$

$= f_1 f_2 + f_3 f_4 - f_1 f_2 f_3 f_4$

2. Biased Coin:

Toss a coin k times. The coin lands in Head with probability q and Tail with probability $1-q$

Let A_k = event that k out of n tosses are Head.

A short-cut way: # that this happen $= \binom{n}{k}$

each happens with probability $q^k(1-q)^{n-k}$

then: $\mathbb{P}(A_k) = \binom{n}{k} q^k (1-q)^{n-k}$.

And we can see that $\mathbb{P}(\Omega) = \mathbb{P}\left(\bigcup_{k=0}^{n} A_k\right) = \sum_{k=0}^{n} \mathbb{P}(A_k) = \sum_{k=0}^{n} \binom{n}{k} q^k (1-q)^k = 1$

Question

I don't have any for this lecture.

LECTURE 5

Important Example.

1. Birthday Problem:

Let $\Omega = \{1, 2, \cdots, 365\}^N$ where $N = \#$ of people.

$\omega \in \Omega$, $\omega = (d_1, d_2, \cdots, d_N)$, $d_k =$ birthday of k^{th} people

Consider event: $A_{N,365}$ = event that at least one pair with same birthday.

$\mathbb{P}$: by symmetry, all outcomes are equally likely.

For any A, $\mathbb{P}(A) = \sum_{\omega \in A} p(\omega) = \frac{\#A}{\#\Omega}$ $\quad \#\Omega = 365^N$

$A^C_{N,365} \subseteq \Omega$: "event that all birthdays are distinct" $= \{(d_1 \cdots d_N) \mid d_k \neq d_j \text{ for all } k \neq j\}$

For $k = 1, 2, \cdots, N$, D_k = event that there is no match among first k people, i.e. birthdays $d_1 \cdots d_k$ are distinct.

$\mathbb{P}(D_1) = 1 \quad \mathbb{P}(D_2) = \frac{365-1}{365} \quad .. \quad \mathbb{P}(D_{k+1}) = \mathbb{P}(D_{k+1} \mid D_k)\mathbb{P}(D_k) = \frac{365-k}{365}\mathbb{P}(D_k)$

So $\mathbb{P}(D_{k+1}) = \prod_{j=1}^{k}\left(\frac{365-j}{365}\right)$

$\mathbb{P}(D_k) < \frac{1}{2}$ when $k \geq 1$, $\mathbb{P}(A_N) = 1 - \mathbb{P}(A_N^c) = 1 - \mathbb{P}(D_N) \geq \frac{1}{2}$ for $N > 21$.

2. There are M distinct marbles in the box $1, 2, \cdots, M$. Draw marble independently with replacement k times. Compute the prob. of a match among k times.

$\Omega = \{1, 2, \cdots, M\}^k$

$\mathbb{P}(\text{No match among first } k \text{ trials}) = \mathbb{P}(D_{k,M})$

$\mathbb{P}(D_{k+1,M}) = \prod_{j=1}^{k}\left(\frac{M-j}{M}\right)$. When is $\mathbb{P}(D_{k+1,M}) < \frac{1}{2}$?

Consider $M \gg 1$ very large:

$\ln \mathbb{P}(D_{k+1,M}) = \sum_{j=1}^{k} \ln \frac{M-j}{M} = \sum_{j=1}^{k} \ln\left(1 - \frac{j}{M}\right)$

$\mathbb{P}(D_{k+1,M}) = e^{\sum_{j=1}^{k} \ln\left(1-\frac{j}{M}\right)}$ $\qquad$ For $|x|$ small: $\ln(1-x) \approx -x$

If $\frac{k}{M}$ is small, then $\frac{j}{M}$ is small for each j.

So when $\frac{k}{M}$ is small, $\sum_{j=1}^{k} \ln\left(1-\frac{j}{M}\right) \approx \sum_{j=1}^{k} \frac{-j}{M} = -\frac{1}{M}\sum_{j=1}^{k} j = -\frac{1}{M}\frac{k(k+1)}{2}$

$\mathbb{P}(D_{k+1,M}) \approx e^{-\frac{1}{M}\frac{k(k+1)}{2}}$.

So, for M large enough, $k > k^* = \sqrt{2M\ln 2}$, $\mathbb{P}(D_{k+1,M}) < \frac{1}{2}$.

$\frac{k^*}{M} = \frac{\sqrt{2M\ln 2}}{M} = \frac{\sqrt{2\ln 2}}{\sqrt{M}} \to 0$ as $M \to \infty$.

So, we only need to draw surprisingly few samples to see a match, when M is large.

Brief Summary

- Law of Small Numbers

Toss a p-coin n times independently

$\Omega = \{H, T\}^N$.

For any $\omega \in \Omega$, $\mathbb{P}(\{\omega\}) = p(\omega) = p^k(1-p)^{n-k}$ if there are k heads in ω.

Let A_k = event that k heads are tossed.

$\mathbb{P}(A_k) = \binom{N}{k}p^k(1-p)^{N-k}$. When N is large, p is small, pick $\lambda > 0$, suppose $p = \frac{\lambda}{N}$ and let $N \to \infty$.

$\mathbb{P}(A_k) = \frac{N!}{k!(N-k)!}\left(\frac{\lambda}{N}\right)^k\left(1-\frac{\lambda}{N}\right)^{N-k} = \frac{\lambda^k}{k!}\frac{N!}{(N-k)!N^k}\left(1-\frac{\lambda}{N}\right)^N\left(1-\frac{\lambda}{N}\right)^{-k}$

$\frac{N!}{(N-k)!N^k} = \frac{N(N-1)(N-2)\cdots(N-k+1)}{N^k} = \left(\frac{k \text{ terms of } N}{N^k}\right) \to 1$, $\left(1-\frac{\lambda}{N}\right)^{-k} \to 1$, $\left(1-\frac{\lambda}{N}\right)^N \to e^{-\lambda}$

so $\mathbb{P}(A_k) \to e^{-\lambda}\frac{\lambda^k}{k!}$ Poisson distribution. Extremely useful when draws N is large and each probability p is small.

LECTURE 6

• Important Examples.

Toss a fair coin $n = 10000$ times.

Consider fraction of heads: f

What is the probability that $f \geq 53\%$?

Theorem: $\forall \varepsilon > 0,\ n \geq 1,\ \mathbb{P}(\underbrace{\bigcup_{k \geq (\frac{1}{2}+\varepsilon)n} A_k}_{\text{event that } f \geq \frac{1}{2}+\varepsilon}) \leq e^{-\varepsilon^2 n}$

proof:

$$\mathbb{P}\Big(\bigcup_{k \geq (\frac{1}{2}+\varepsilon)n} A_k\Big) = \sum_{k \geq (\frac{1}{2}+\varepsilon)n} \mathbb{P}(A_k)$$

$$= \sum_{k \geq (\frac{1}{2}+\varepsilon)n} \binom{n}{k}\Big(\frac{1}{2}\Big)^n$$

$$\leq \sum_{k \geq (\frac{1}{2}+\varepsilon)n} \binom{n}{k}\Big(\frac{1}{2}\Big)^n e^{\alpha(k-(\frac{1}{2}+\varepsilon)n)} \quad \forall \alpha \geq 0$$

$$\leq \sum_{k \geq 0} \binom{n}{k}\Big(\frac{1}{2}\Big)^n e^{\alpha(k-(\frac{1}{2}+\varepsilon)n)}$$

$$= 2^{-n} e^{-\alpha\varepsilon n} \sum_{k \geq 0} \binom{n}{k} e^{\alpha(k-\frac{1}{2}n)}$$

$$= e^{-\alpha\varepsilon n}\, 2^{-n} \sum_{k \geq 0} \binom{n}{k} e^{\alpha(\frac{k}{2}-\frac{n}{2})} e^{\alpha(\frac{k}{2})}$$

$$= e^{-\alpha\varepsilon n}\, 2^{-n} \sum_{k \geq 0} \binom{n}{k} (e^{-\frac{\alpha}{2}})^{n-k} (e^{\frac{\alpha}{2}})^{k}$$

$$= e^{-\alpha\varepsilon n}\, 2^{-n} (e^{-\frac{\alpha}{2}} + e^{\frac{\alpha}{2}})^n$$

$$= e^{-\alpha\varepsilon n} \Big(\frac{e^{-\frac{\alpha}{2}} + e^{\frac{\alpha}{2}}}{2}\Big)^n \quad \text{by } \frac{e^{x}+e^{-x}}{2} \leq e^{x^2}$$

$$\leq e^{-\alpha\varepsilon n} e^{\alpha^2 n}$$

$$= e^{-\alpha\varepsilon n + \frac{\alpha^2 n}{4}}$$

minimize it with respect to α:

$\alpha^* = 2\varepsilon$

$$\mathbb{P}\Big(\bigcup_{k \geq (\frac{1}{2}+\varepsilon)n} A_k\Big) \leq e^{-\varepsilon^2 n}$$

• Brief Summary:

by symmetry: $\forall \varepsilon > 0, n \geq 1, \mathbb{P}(\bigcup_{k \leq (\frac{1}{2}-\varepsilon)n} A_k) \leq e^{-\varepsilon^2 n}$

Corollary: $\mathbb{P}(\bigcup_{(\frac{1}{2}-\varepsilon)n \leq k \leq (\frac{1}{2}+\varepsilon)n} A_k) \geq 1 - 2e^{-\varepsilon^2 n}$

This is assuming: n tosses are independent tosses

$$\mathbb{P}\Big(\tfrac{1}{2} - \delta \leq \text{fraction of heads} \leq \tfrac{1}{2} + \beta\Big) = 1 - \mathbb{P}\Big(\bigcup_{k \geq (\frac{1}{2}+\beta)n} A_k \cup \bigcup_{k \leq (\frac{1}{2}-\delta)n} A_k\Big)$$

$$= 1 - \mathbb{P}\Big(\bigcup_{k \geq (\frac{1}{2}+\beta)n} A_k\Big) - \mathbb{P}\Big(\bigcup_{k \leq (\frac{1}{2}-\delta)n} A_k\Big)$$

$$\leq 1 - e^{-\beta^2 n} - e^{-\delta^2 n}$$

$$\to 1 \text{ as } n \to \infty.$$

Questions. I don't have any questions for this lecture.

LECTURE 7

Brief Summary

- Random Variable $X(\omega)$ is a function $X: \Omega \to \mathbb{R}$
- Distribution of a random variable is the assignment of probabilities that X can take.

☆ A random variable is not the same as "the distribution of it".
It is possible that for two r.v.s to have the same distribution.

- Cumulative Distribution Function CDF.
  If X is a r.v., its CDF is the function $F(y) = \mathbb{P}(X \le y)$ for $y \in \mathbb{R}$. $F: \mathbb{R} \to [0,1]$

- Properties of F:

① F is non-decreasing

If $y_1 < y_2$, then $\mathbb{P}(X \le y_1) \le \mathbb{P}(X \le y_2)$ because $\{\omega | X(\omega) \le y_1\} \subseteq \{\omega | X(\omega) \le y_2\}$

② F can have jump, but it is right-continuous.

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

$\mathbb{P}(X = y) = F(y) - \underbrace{F(y^-)}_{= \lim_{z \uparrow y} F(z)}$

Important Examples

Let $A \in \Omega$ be some event. Define $\mathbb{1}_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A \end{cases}$

This r.v. is the characteristic function; or the "indicator" of A.

Distribution of $\mathbb{1}_A(\omega)$: Bernoulli Distribution.

$\mathbb{P}(\mathbb{1}_A(\omega) = 1) = \mathbb{P}(\{\omega | \omega \in A\}) = \mathbb{P}(A)$

$\mathbb{P}(\mathbb{1}_A(\omega) = 0) = \mathbb{P}(\{\omega | \omega \notin A\}) = \mathbb{P}(A^c) = 1 - \mathbb{P}(A)$

① Bernoulli (p) distribution:

$\mathbb{P}(X=1) = p$ $\quad$ $\mathbb{P}(X = y) = 0$ if $y \ne 0, 1$

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

$\mathbb{1}_A$ has the Bernoulli (p) distribution if $p = \mathbb{P}(A)$

② Bernoulli (n, p) distribution:

$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ for $k \in \{0, 1, \cdots, n\}$

and $\mathbb{P}(X \notin \{0, 1, \cdots, k\}) = 0$

③ Poisson (λ) distribution: LSN: Binomial $\xrightarrow{d}$ Poisson $\quad n \to \infty,\ p = \frac{\lambda}{n}$

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

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

LECTURE 8

④ Geometric (p) Distribution

$X(\omega): \Omega \to \mathbb{R}$ has the Geo(p) distribution if $\mathbb{P}(X=k) = p(1-p)^{k-1}$, $k=1,2,\cdots$

e.g. Toss a p-coin many times until head appears.

Let X = # of tosses until head. Then, $X \sim \text{Geo}(p)$

Also, notice that $\sum_{k=1}^{\infty} \mathbb{P}(X=k) = \sum_{k=1}^{\infty} p(1-p)^{k-1} = p\sum_{k=1}^{\infty}(1-p)^{k-1} = 1.$

Brief Summary

- Independence of Random Variables.

Two discrete r.v. X & Y are independent if $\mathbb{P}(X=a, Y=b) = \mathbb{P}(X=a)\,\mathbb{P}(Y=b)$ $\forall\, a \in R(X),\ b \in R(Y)$

$\Leftrightarrow \mathbb{P}(X=a \mid Y=b) = \mathbb{P}(X=a)$.

- Suppose $X_1, X_2, \cdots, X_n$ are discrete r.v. They are independent if

$\mathbb{P}(X_1=a_1, X_2=a_2, \cdots, X_n=a_n) = \mathbb{P}(X_1=a_1)\,\mathbb{P}(X_2=a_2)\cdots\mathbb{P}(X_n=a_n)$ for all $a_1, a_2, \cdots, a_n$

$\sum_{a_1 \in R(X_1)} \mathbb{P}(X_1=a_1, X_2=a_2, \cdots, X_n=a_n) = \mathbb{P}(X_2=a_2, \cdots, X_n=a_n)$

Equivalently:

1) For any intervals $I_1, I_2, \cdots, I_n \subset \mathbb{R}$, $\mathbb{P}(X_1\in I_1, X_2\in I_2, \cdots, X_n \in I_n) = \mathbb{P}(X_1\in I_1)\,\mathbb{P}(X_2\in I_2)\cdots\mathbb{P}(X_n\in I_n)$

2) For any $a_1, a_2, \cdots a_n$, $\mathbb{P}(X_1\le a_1, X_2\le a_2, \cdots, X_n\le a_n) = \mathbb{P}(X_1 \in a_1)\,\mathbb{P}(X_2\le a_2)\cdots\mathbb{P}(X_n\le a_n)$

Important Example

1. Toss a p-coin 5 times independently, $\Omega = \{H, T\}^5$.

Let X = # of heads among tosses 1,2,3; Y = # of tails among tosses 3,4,5.

X & Y are not independent! Check: let $a=3$. Then $\mathbb{P}(X=3, Y=3) = 0 \neq \mathbb{P}(X=3)\,\mathbb{P}(Y=3)$.

Let Z = # of tails in tosses 4,5.

Claim: X & Z are independent. Check: Let H_k = event that heads occur on k^{th} tosses. $T_k = H_k^c$ = tail on k^{th} tosses.

By construction: $H_1\ H_2\ H_3\ H_4\ H_5$ are independent.

$X(\omega) = \mathbb{1}_{H_1}(\omega) + \mathbb{1}_{H_2}(\omega) + \mathbb{1}_{H_3}(\omega)$; $Z(\omega) = \mathbb{1}_{T_4}(\omega) + \mathbb{1}_{T_5}(\omega) = \mathbb{1}_{H_4^c}(\omega) + \mathbb{1}_{H_5^c}(\omega)$

Check: $\forall\ a = 1,2,3,4,5$, event $\{X=a\}$ (and $\{Z=a\}$) can be expressed using H_1, H_2, H_3, (T_4, T_5).

then we can verify that for each values X and Z can take, they are independent.

e.g. $\{X=0\} = H_1^c \cap H_2^c \cap H_3^c$ $\quad \{Z=2\} = T_4 \cap T_5 = H_4^c \cap H_5^c$. Then $\mathbb{P}(X=0, Z=2) = \mathbb{P}((H_1^c \cap H_2^c \cap H_3^c) \cap (H_4^c \cap H_5^c)) = \mathbb{P}(X=0)\,\mathbb{P}(Z=2)$

2. Toss a p-coin N times. Let X = # heads. Y = # of tails.

i) If N is fixed, then X & Y not independent. Check: $\mathbb{P}(X=N, Y=N) = 0 \neq \mathbb{P}(X=N)\,\mathbb{P}(Y=N)$.

ii) If $N \sim \text{Poisson}(\lambda)$, $\mathbb{P}(N=k) = e^{-\lambda}\frac{\lambda^k}{k!}$ for $k=0,1,2,\cdots$

Let X = # of heads Y = # of tails. Then: 1) X & Y are independent 2). $X \sim \text{Poisson}(\lambda p)$, $Y \sim \text{Poisson}(\lambda(1-p))$

proof. 2): $\mathbb{P}(X=k) = \sum_{n=0}^{\infty}\mathbb{P}(X=k\mid N=n)\,\mathbb{P}(N=n) = \sum_{n=k}^{\infty}\mathbb{P}(X=k\mid N=n)\,\mathbb{P}(N=n) = \sum_{n=k}^{\infty}\binom{n}{k}p^k(1-p)^{n-k}e^{-\lambda}\frac{\lambda^n}{n!} = \sum_{n=k}^{\infty}\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}e^{-\lambda}\frac{\lambda^n}{n!}$

$$= \frac{e^{-\lambda}p^k}{k!}\sum_{n=k}^{\infty}\frac{(1-p)^{n-k}\lambda^n}{(n-k)!} = \frac{e^{-\lambda}p^k}{k!}\sum_{j=0}^{\infty}\frac{(1-p)^j\lambda^{j+k}}{j!}\ (n=j+k) = \frac{e^{-\lambda}(\lambda p)^k}{k!}\sum_{j=0}^{\infty}\frac{(1-p)^j\lambda^j}{j!} = \frac{e^{-\lambda}(\lambda p)^k}{k!}e^{\lambda(1-p)} = e^{-\lambda p}\frac{(\lambda p)^k}{k!}$$

similarly, using the partition technique, $\mathbb{P}(Y=k) = e^{-\lambda(1-p)}\frac{[\lambda(1-p)]^k}{k!}$.

1): verify $\mathbb{P}(X=k, Y=j) \overset{?}{=} \mathbb{P}(X=k)\,\mathbb{P}(Y=j)$

$$\mathbb{P}(X=k, Y=j) = \sum_{n=0}^{\infty}\underbrace{\mathbb{P}(X=k, Y=j\mid N=n)}_{=0 \text{ if } k+j\neq n}\mathbb{P}(N=n) = \mathbb{P}(X=k, Y=j\mid N=k+j)\,\mathbb{P}(N=k+j) = \binom{k+j}{k}p^k(1-p)^j e^{-\lambda}\frac{\lambda^{j+k}}{(j+k)!}$$

$$= \frac{(j+k)!}{j!\,k!}p^k(1-p)^j e^{-\lambda}\frac{\lambda^{j+k}}{(j+k)!} = e^{-\lambda p}\frac{(\lambda p)^k}{k!}\,e^{-\lambda(1-p)}\frac{[\lambda(1-p)]^j}{j!} = \mathbb{P}(X=k)\,\mathbb{P}(Y=j).$$ Independent!

LECTURE 9

Brief Summary

- Expectation: If X is a discrete r.v., then its expectation is $\mathbb{E}[X] = \sum_{x \in R(X)} x \, \mathbb{P}(X = x)$
- ① $\mathbb{E}[X]$ is a number ② $\mathbb{E}[X]$ is determined completely by the distribution of X.
  ③ If X & Y are two different r.v.s but have the same dist., then $\mathbb{E}[X] = \mathbb{E}[Y]$
- Interpretation: ① $\mathbb{E}[X]$ is weighted average, the center of mass. (Possibly $\mathbb{E}[X] \notin R(X)$)
  ② Long-run average. LLN. ③ "Fair Price" for a random prize

e.g. Let $A \subseteq \Omega$ be any event. Consider $X(\omega) = \mathbb{1}_A(\omega) = \begin{cases} 1, & \omega \in A \\ 0, & \omega \notin A \end{cases}$

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{1}_A] = \sum_{x \in R(\mathbb{1}_A)} x \cdot \mathbb{P}(\mathbb{1}_A = x) = \mathbb{P}(\mathbb{1}_A = 1) = \mathbb{P}(A)$$

- Properties of Expectation:
  ① Linearity: $\forall$ r.v. $X \neq Y$, $\alpha, \beta \in \mathbb{R}$, $\mathbb{E}[\alpha X + \beta Y] = \alpha \mathbb{E}[X] + \beta \mathbb{E}[Y]$. $\mathbb{E}$: abstract integration on Ω.

e.g. Let $X \sim \text{Binomial}(n, p)$

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

Easier Way: Tossing n p-coins. Let A_j = event that head occurs on j^{th} toss.

Define $Y(\omega) = \mathbb{1}_{A_1}(\omega) + \mathbb{1}_{A_2}(\omega) + \cdots + \mathbb{1}_{A_n}(\omega) = \sum_{j=1}^{n} \mathbb{1}_{A_j}(\omega)$ = # of heads in n tosses.

Then: $Y(\omega) \sim \text{Binomial}(n, p)$

$$\mathbb{E}[X] = \mathbb{E}[Y] = \mathbb{E}[\mathbb{1}_{A_1}(\omega) + \mathbb{1}_{A_2}(\omega) + \cdots + \mathbb{1}_{A_n}(\omega)] = \sum_{j=1}^{n} \mathbb{E}[\mathbb{1}_{A_j}] = \sum_{j=1}^{n} \mathbb{P}(A_j) = \sum_{j=1}^{n} p = np.$$

② For functions: Let X be a r.v. $X: \Omega \to \mathbb{R}$. $g(x): \mathbb{R} \to \mathbb{R}$. Consider $g(X(\omega))$. then: $\mathbb{E}[g(X)] = \sum_{x \in R(X)} g(x) \mathbb{P}(X = x)$.

③ Tail Sum Formula: Suppose X has range $R(X) = \{0, 1, 2, \cdots\}$. non-negative integers.

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

proof. Let $A_k = \{\omega \mid X(\omega) = k\}$

Then: $\mathbb{E}[X] = \sum_{k=1}^{\infty} k \, \mathbb{P}(X = k) = \sum_{k=1}^{\infty} k \, \mathbb{P}(X = k) = \sum_{k=1}^{\infty} k \, \mathbb{P}(A_k) = \sum_{k=1}^{\infty} k \, \mathbb{E}[\mathbb{1}_{A_k}]$

by linearity: $= \mathbb{E}\left[\sum_{k=1}^{\infty} k \, \mathbb{1}_{A_k}(\omega)\right]$

Define $B_j = \{\omega \mid X(\omega) > j\}$

For $\omega \in A_k$: $\mathbb{1}_{B_j}(\omega) = \begin{cases} 1 & \text{for } j = 0, 1, \cdots, k-1 \\ 0 & \text{for } j = k, k+1, \cdots \end{cases}$

$\{X = k\}$

If $\omega \in A_k$, $\sum_{j=0}^{\infty} \mathbb{1}_{B_j}(\omega) = \underbrace{\mathbb{1}_{B_0}(\omega) + \mathbb{1}_{B_1}(\omega) + \cdots + \mathbb{1}_{B_{k-1}}(\omega)}_{=1} + \underbrace{\mathbb{1}_{B_k}(\omega) + \cdots}_{=0}$

So, $k \, \mathbb{1}_{A_k}(\omega) = \sum_{j=0}^{\infty} \mathbb{1}_{B_j}(\omega)$ if $\omega \in A_k$

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{k=1}^{\infty} k \, \mathbb{1}_{A_k}\right] = \mathbb{E}\left[\sum_{k=1}^{\infty} \sum_{j=0}^{\infty} \mathbb{1}_{B_j}(\omega)\right] = \sum_{k=1}^{\infty} \sum_{j=0}^{\infty} \mathbb{E}[\mathbb{1}_{B_j}(\omega)]$$

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

If $X \geq 0$, then: $\mathbb{E}[X] = \int_0^{\infty} \mathbb{P}(X \geq s) \, ds$.

Way Two:

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} k \, \mathbb{P}(X = k)$$

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

$+ \mathbb{P}(X = 2) + \mathbb{P}(X = 2)$

$+ \mathbb{P}(X = 3) + \mathbb{P}(X = 3) + \mathbb{P}(X = 3)$

$+ \cdots\cdots$

$= \mathbb{P}(X > 0) + \mathbb{P}(X > 1) + \mathbb{P}(X > 2) + \cdots$

$= \sum_{j=0}^{\infty} \mathbb{P}(X > j)$

e.g. Let $X \sim \text{Geo}(p)$, then: $\mathbb{E}[X] = \sum_{k=0}^{\infty} k \, p(1-p)^{k-1}$

$$= \sum_{j=0}^{\infty} \mathbb{P}(X > j) = \sum_{j=0}^{\infty} (1-p)^j = \frac{1}{1-(1-p)} = \frac{1}{p}.$$

For continuous r.v. $X \geq 0$: or tail sum:

$$\mathbb{E}[X] = \int_0^{\infty} (1 - F(x)) \, dx$$

LECTURE 10

Brief Summary

- Variance: $\forall$ r.v. X, $var(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$. Characterizes the spread of distribution.

  If $\mathbb{E}[X] = \mu$, then $var(X) = \mathbb{E}[(X-\mu)^2] = \mathbb{E}[X^2 - 2\mu X + \mu^2] = \mathbb{E}[X^2] - 2\mathbb{E}[X]\mathbb{E}[X] + \mathbb{E}[X]^2$
  $= \mathbb{E}[X^2] - \mathbb{E}[X]^2$

- Standard Deviation $SD(X) = var(X)^{\frac{1}{2}}$

- Properties: 0). $var(X) \geq 0$

  1). Possibly $var(X) = +\infty$ even if $\mathbb{E}[X] < +\infty$

  But $var(X) < +\infty$ iff $\mathbb{E}[X^2] < +\infty$

  2). Scaling. $\forall \alpha, \beta \in \mathbb{R}$, $var(\alpha X + \beta) = \alpha^2 var(X)$

  $SD(\alpha X + \beta) = \alpha\, SD(X)$

  3). Variation of Sums. Suppose $\{X_i\}_{i=1}^n$ are r.v. with finite variance

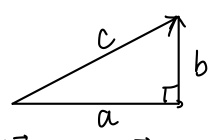
$$
\begin{aligned}
var(X_1 + \cdots + X_n) &= \mathbb{E}[(\sum_{i=1}^n X_i)^2] - \mathbb{E}[\sum_{i=1}^n X_i]^2 \\
&= \mathbb{E}[X_1^2 + X_2^2 + \cdots + X_n^2 + 2\sum_{j>k} X_j X_k] - (\sum_{i=1}^n \mathbb{E}[X_i])^2 \\
&= \sum_{i=1}^n \mathbb{E}[X_i^2] + 2\sum_{j>k}\mathbb{E}[X_j X_k] - \sum_{i=1}^n \mathbb{E}[X_i]^2 - 2\sum_{j>k}\mathbb{E}[X_j]\mathbb{E}[X_k] \\
&= \sum_{i=1}^n \mathbb{E}[X_i^2] - \sum_{i=1}^n \mathbb{E}[X_i]^2 + 2\sum_{j>k}\mathbb{E}[X_j X_k] - 2\sum_{j>k}\mathbb{E}[X_j]\mathbb{E}[X_k] \\
&= \sum_{i=1}^n var(X_i) + 2\sum_{j>k} cov(X_j, X_k)
\end{aligned}
$$

- $cov(X,Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$
- Theorem: if X & Y are independent, then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ ①   ① => uncorrelated

  $\therefore$ Independence => Uncorrelated.

  proof:
$$
\begin{aligned}
\mathbb{E}[XY] &= \sum_{l \in R(XY)} l\, \mathbb{P}(XY = l) \\
&= \sum_{l\in R(XY)} l \sum_{k \in R(Y)} \mathbb{P}(XY = l, Y = k) \quad \text{since the events } \{Y=k\}_{k\in R(Y)} \text{ partition } \Omega. \\
&= \sum_{l\in R(XY)} l \sum_{k\in R(Y)} \mathbb{P}(X = \tfrac{l}{k}, Y = k) \\
&= \sum_{l\in R(XY)} l \sum_{k\in R(Y)} \mathbb{P}(X = \tfrac{l}{k})\mathbb{P}(Y=k) \quad \text{since independent} \\
&= \sum_{k\in R(Y)} \Big(\sum_{l\in R(XY)} l\, \mathbb{P}(X = \tfrac{l}{k})\Big)\mathbb{P}(Y=k) \quad x = \tfrac{l}{k} \quad l = kx \\
&= \sum_{k\in R(Y)} \Big(\sum_{x\in R(X)} kx\, \mathbb{P}(X=x)\Big)\mathbb{P}(Y=k) \\
&= \sum_{k\in R(Y)} k \Big(\sum_{x\in R(X)} x\, \mathbb{P}(X=x)\Big)\mathbb{P}(Y=k) \\
&= \Big(\sum_{x\in R(X)} x\, \mathbb{P}(X=x)\Big)\cdot\Big(\sum_{k\in R(Y)} k\, \mathbb{P}(Y=k)\Big) \\
&= \mathbb{E}[X]\,\mathbb{E}[Y]
\end{aligned}
$$

If we say $var(X) = (\text{"length"})^2$



$var(\vec{c}) = var(\vec{a}) + var(\vec{b})$

$\|\vec{c}\|^2 = \|\vec{a}\|^2 + \|\vec{b}\|^2$

independent: $\vec{a} \perp \vec{b}$.

square: a nice geometric property better than $|\cdot|$.

- If $X_1, X_2 \cdots X_n$ are (pairwise) independent, then $var(\sum_{i=1}^n X_i) = \sum_{i=1}^n var(X_i)$.

Important Example

Let $A \subset \Omega$ be any event. Consider $X = \mathbb{1}_A$.

$\mathbb{E}[X] = \mathbb{E}[\mathbb{1}_A] = \mathbb{P}(A)$

$var(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathbb{E}[\mathbb{1}_A^2] - \mathbb{E}[\mathbb{1}_A]^2 = \mathbb{E}[\mathbb{1}_A] - \mathbb{E}[\mathbb{1}_A]^2 = \mathbb{P}(A) - \mathbb{P}(A)^2 \leq \frac{1}{4}$ when $\mathbb{P}(A) = \frac{1}{2}$.

notice $\mathbb{1}_A = \mathbb{1}_A^2$.

LECTURE 11

Brief Summary

- Variance: $\{X_i\}_{i=1}^n$ independent, $var(X_i)=\sigma^2\ \forall i$. Then $var(\sum_{i=1}^n X_i) = \sum_{i=1}^n var(X_i) = n\sigma^2$. $SD(\sum_{i=1}^n X_i) = \sqrt{n}\sigma$

- Chebyshev's Inequality.
  Suppose Y is an r.v., $Y \geq 0$. Then $\forall t$, $\mathbb{P}(Y > t) \leq \frac{\mathbb{E}[Y]}{t}$ (tail)
  pf: Y discrete: $\mathbb{E}[Y] = \sum_{z \in R(Y)} z\mathbb{P}(Y=z) \geq \sum_{\substack{z\in R(Y) \\ z \geq t}} z\mathbb{P}(Y=z) \geq \sum_{\substack{z\in R(Y) \\ z\geq t}} t\mathbb{P}(Y=z) = t\sum_{\substack{z\in R(Y) \\ z\geq t}} \mathbb{P}(Y=z) = t\,\mathbb{P}(Z\geq t)$
  $\therefore \mathbb{P}(Y\geq t) \leq \frac{\mathbb{E}[Y]}{t}$

- Markov's Inequality (Variation of ↓)
  Suppose X r.v. $var(X) < \infty$, then $\mathbb{P}(|X-\mu| \geq \varepsilon) \leq \frac{var(X)}{\varepsilon^2}$.
  pf: change Y to $(X-\mu)^2$.

- Suppose $\{X_i\}_{i=1}^n$ i.i.d $\mathbb{E}[X_i]=\mu$, $var(X_i)=\sigma^2$. Then $var(\sum_{i=1}^n X_i - n\mu) = var(\sum_{i=1}^n X_i) = n\sigma^2$
  $\mathbb{E}[\sum_{i=1}^n X_i] = n\mu$

  then $\mathbb{P}(|\sum_{i=1}^n X_i - n\mu| \geq t) \leq \frac{n\sigma^2}{t^2}$
  $\mathbb{P}(|\frac{1}{n}\sum_{i=1}^n X_i - \mu| \geq \frac{t}{n}) \leq \frac{n\sigma^2}{t^2}$
  $\frac{t}{n} = \varepsilon$. $t = n\varepsilon$: $\mathbb{P}(|\bar{X}_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}$. $\forall \varepsilon > 0$

  $\forall \varepsilon > 0$, $\lim_{n\to\infty} \mathbb{P}(|\bar{X}_n - \mu| \geq \varepsilon) = 0$ $\quad\Rightarrow \bar{X}_n \xrightarrow{P} \mu$.

- WLLN: Let $\{X_i\}_{i=1}^n$ i.i.d r.v. $\mathbb{E}[X_i]=\mu<\infty$. Then $\forall \varepsilon>0$, $\mathbb{P}(|\bar{X}_n-\mu|\geq\varepsilon) \to 0$.
  pf. $\mathbb{P}(|\bar{X}_n-\mu|\geq\varepsilon) \leq \frac{\mathbb{E}[|\bar{X}_n-\mu|]}{\varepsilon} = \frac{\mathbb{E}[|\sum_{i=1}^n X_i - \mu|]}{n\varepsilon} \leq \frac{\mathbb{E}[\sum_{i=1}^n |X_i-\mu|]}{n\varepsilon} = \frac{\sum_{i=1}^n \mathbb{E}[|X_i-\mu|]}{n\varepsilon}$ ... ?

- Suppose X is a r.v. with distribution $\mathbb{P}(X=k) = \frac{c}{k^\alpha}$ $k=1,2,3,\cdots$ for some $\alpha>0$
  Need: $\sum_{k\in R(X)} \mathbb{P}(X=k) = 1 \Rightarrow c\sum_{k\in R(X)} \frac{1}{k^\alpha} = 1$ Choose $c = \frac{1}{\sum_{k\in R(X)} 1/k^\alpha}$ Suppose $\alpha>1$, then: valid $\mathbb{P}$ distribution.

  $\mathbb{E}[X] = \sum_{k\in R(X)} k\mathbb{P}(X=k) = \sum_{k\in R(X)} \frac{c}{k^{\alpha-1}} < \infty$ if $\alpha > 2$

  Then: for $\alpha>1$, $\mathbb{P}$ well defined. For $1<\alpha\leq 2$: $\mathbb{E}[X]=\infty$, $var(X)=\infty$.
  $2<\alpha\leq 3$: $\mathbb{E}[X]<\infty$, $var(X)=\infty$
  $\alpha\geq 3$: $\mathbb{E}[X]<\infty$, $var(X)<\infty$.

- Joint Distribution: Given 2 r.v. X, Y on some outcome space, the joint dist is the distribution of the pair (X,Y) $\mathbb{P}(X=x, Y=y)$ $\forall x\in R(X)$, $y\in R(Y)$
  $\sum_{x\in R(X)}\sum_{y\in R(X)} \mathbb{P}(X=x, Y=y) = 1$

Important Example

$X \sim Binom(n,p)$, $Var(X)$? Suppose X = # of heads in n tosses of a p-coin.
Assume $X(\omega) = \sum_{i=1}^n \mathbb{1}_{H_i}(\omega)$ H_i = event that head occurs on toss i. $\{\mathbb{1}_{H_i}(\omega)\}$ i.i.d. $var(\mathbb{1}_{H_i}(\omega)) = p(1-p)$
$var(X) = \sum_{i=1}^n p(1-p) = np(1-p)$ $\quad\therefore X\sim Binom(n,p)$, then $\mathbb{E}[X]=np$, $Var[X]=np(1-p)$

Two experiments: ① Toss p-coin n times independently, X = # heads. $Var(\sum_{i=1}^n X_i) = np(1-p)$
☆ ② Toss p-coin 1 time, define $\tilde{X}_k = \tilde{X}_1$ $\forall k=1,2,\cdots,n$. $var(\sum_{i=1}^n \tilde{X}_i) = var(n\tilde{X}_1) = n^2 var(\tilde{X}_1) = n^2 p(1-p)$.
$\tilde{X}_1 = \begin{cases} 1 & H \\ 0 & T \end{cases}$

LECTURE 12.

Brief Summary.

Let X & Y be two r.v. on the same sample space.

- Joint Distribution: $\mathbb{P}(X=k, Y=j) = p_{XY}(k,j) \quad \forall\ k \in R(X),\ j \in R(Y)$.
- Marginal Distribution: $\mathbb{P}(X=k) = p_X(k) \quad , k \in R(X)$

$$\mathbb{P}(X=k) = \sum_{j\in R(Y)} \mathbb{P}(X=k, Y=j) \quad \text{since } \{Y=j\}_{j\in R(Y)} \text{ is a partition}$$

$$\mathbb{P}(Y=j) = \sum_{k\in R(X)} \mathbb{P}(X=k, Y=j)$$

Joint dist. determines marginal distribution. Marginal distribution cannot determine joint dist. without more info.

- Conditional Distribution: $\mathbb{P}(X=k \mid Y=j) = p_{X|Y}(k\mid j) = \dfrac{\mathbb{P}(X=k, Y=j)}{\mathbb{P}(Y=j)}$
- Independence: $X \perp\!\!\!\perp Y$ if $\mathbb{P}(X=k, Y=j) = \mathbb{P}(X=k)\mathbb{P}(Y=j) \quad , \forall (k,j) \in R(X,Y)$

  i.e. $(k,j) \mapsto \mathbb{P}(X=k, Y=j)$ has a product structure.

  $p_{XY}(k,j) = p_X(k)\, p_Y(j)$

- Given two r.v. X & Y: $\mathbb{E}[X \mid Y=j] = \sum_{k\in R(X)} k\, \mathbb{P}(X=k \mid Y=j)$
- Discrete Version of LIE. $\mathbb{E}[X] = \sum_{j\in R(Y)} \mathbb{E}[X\mid Y=j]\, \mathbb{P}(Y=j)$

Pf:

$$RHS = \sum_{j\in R(Y)} \left(\sum_{k\in R(X)} k\, \mathbb{P}(X=k \mid Y=j) \right) \mathbb{P}(Y=j) = \sum_{j\in R(Y)} \left(\sum_{k\in R(X)} k\, \mathbb{P}(X=k\mid Y=j)\, \mathbb{P}(Y=j) \right)$$

$$= \sum_{j\in R(Y)} \left(\sum_{k\in R(X)} k\, \mathbb{P}(X=k, Y=j) \right) = \sum_{j\in R(Y)} \sum_{k\in R(Y)} k\, \mathbb{P}(X=k, Y=j) = \mathbb{E}[X] = LHS$$

Important Example:

Sea turtle lay N eggs. $N \sim \text{Poisson}(\lambda)$. Given N eggs, they each hatch independently w.p. p.

Let X = # of eggs that hatched. Then: $X \mid N=m \sim \text{Binom}(m, p)$

$$\mathbb{P}(X=k, N=m) = \mathbb{P}(X=k \mid N=m)\, \mathbb{P}(N=m)$$

$$\mathbb{P}(N=m) = e^{-\lambda}\frac{\lambda^m}{m!} \quad m \in \{0,1,2,\cdots\}$$

$$\mathbb{P}(X=k \mid N=m) = \binom{m}{k} p^k (1-p)^{m-k}, \quad k \in \{0,1,\cdots,m\}$$

$$\mathbb{P}(X=k, N=m) = \begin{cases} \binom{m}{k} p^k (1-p)^{m-k} e^{-\lambda}\frac{\lambda^m}{m!}, & \text{if } m \in \{0,1,2,\cdots\} \quad k\in\{0,1,2,\cdots,m\} \\ 0 & \text{, otherwise.} \end{cases}$$

$$\mathbb{P}(X=k) = \sum_{m=0}^{\infty} \mathbb{P}(X=k, N=m) = \sum_{m=0}^{\infty} \binom{m}{k} p^k (1-p)^{m-k} e^{-\lambda}\frac{\lambda^m}{m!} = e^{-\lambda} p^k \frac{1}{k!} \lambda^k \sum_{m=k}^{\infty} \frac{m!}{(m-k)!} \frac{[\lambda(1-p)]^{m-k}}{m!}$$

$$= e^{-\lambda}\frac{(\lambda p)^k}{k!} \sum_{i=0}^{\infty} \frac{[\lambda(1-p)]^i}{i!} = e^{-\lambda}\frac{(\lambda p)^k}{k!} e^{\lambda-\lambda p} = e^{-\lambda p}\frac{(\lambda p)^k}{k!}$$

$\therefore X \sim \text{Poisson}(\lambda p) \implies \mathbb{E}[X] = \lambda p$

Alternative way to calculate $\mathbb{E}[X]$:

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X\mid N]] = \sum_{m\in R} \mathbb{E}[X\mid N=m]\, \mathbb{P}(N=m)$$

$$= \sum_{m\in R(N)} pm \cdot \mathbb{P}(N=m)$$

$$= p \sum_{m\in R(N)} m\, \mathbb{P}(N=m)$$

$$= p \cdot \mathbb{E}[N]$$

$$= p\lambda.$$

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

$$= \mathbb{E}[pN]$$

$$= p\,\mathbb{E}[N] \quad N \sim \text{Poisson}(\lambda)$$

$$= p\lambda$$

LECTURE 13

Brief Summary.

- Convolution: from marginal to joint.

Suppose X, Y are independent, discrete r.v. $Z = X+Y$. Given marginal distribution $p_X(x)$ $p_Y(y)$

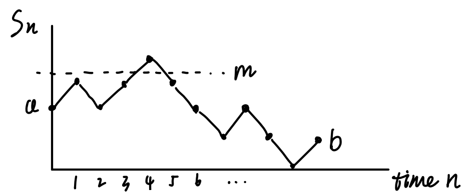
then $p_Z(k) = \sum_{y \in R(Y)} p_X(k-y)\,p_Y(y) = \sum_{x \in R(X)} p_X(x)\,p_Y(k-x) = p_X * p_Y(k)$

proof: $\mathbb{P}(Z=k) = \mathbb{P}(X+Y=k) = \sum_{y \in R(Y)} \mathbb{P}(X+Y=k \mid Y=y)\,\mathbb{P}(Y=y) = \sum_{y \in R(Y)} \mathbb{P}(X+y=k \mid Y=y)\,\mathbb{P}(Y=y)$

$= \sum_{y \in R(Y)} \mathbb{P}(X+y=k)\,\mathbb{P}(Y=y) = \sum_{y \in R(Y)} p_X(k-y)\,p_Y(y)$ since independent

General Version: suppose $X_1 \cdots X_n$ i.i.d, $p_{X_k}(\cdot) = p(\cdot)$. Then $\mathbb{P}(X_1+X_2+\cdots+X_n=k) = p*(p(*\cdots*p))$
n-fold convolution of p with itself.

- Random Walk:



Path on $\mathbb{Z}$. S_n = position of time n. $a \in \mathbb{Z}$: starting point

At each time, flip a fair coin $\begin{cases} H: \text{ step up } +1 \\ T: \text{ step down } -1. \end{cases}$

Consider path $\{S_k\}_{k=1}^{n}$

Fix n. $\Omega = \{w=(w_1, w_2, \cdots, w_n) : w_k \in \{-1,+1\}\}$. w_k: step of time k. $\forall\, k \in \{1,2,\cdots,n\}$.

$S_k(w) = a + \sum_{j=1}^{k} w_j$. Then, path of a random walk is $(a, S_1, S_2, \cdots, S_n)$ $\quad S_{k+1} - S_k = w_k \in \{-1, 1\}$

If the coin is fair, indep. toss. => all outcomes w are equally likely $\mathbb{P}(A) = \frac{\# A}{\# \Omega} = 2^{-n}$

Fix $a \in \mathbb{Z}$. Let $b \in \mathbb{Z}$. $\mathbb{P}^a(S_n = b)$ (pr. of starting from a and ending at b)

first, observe that when n is odd, then the parity of b changes from a. $b-a$ is odd.

when n is even, then the parity of b doesn't change from a. $b-a$ is even.

then, assume n & $(b-a)$ have the same parity. (If different parity, then $\mathbb{P}^a(S_n=b) = 0$.)

$\mathbb{P}^a(S_n=b) = \#\text{ paths} \cdot 2^{-n}$

$= N_n(a,b) \cdot 2^{-n}$ where $N_n(a,b) = \#$ of paths from a to b ① with steps n ②.

$N_n(a,b)$: Given $w = (w_1, w_2, \cdots, w_n)$ where $w_k \in \{-1,+1\}$,

let $\begin{cases} X(w) = \#\ 1\text{'s in } w \\ Y(w) = \#\ {-1}\text{'s in } w \end{cases}$. In order that $S_n = b$, $\begin{cases} X - Y = b-a \ ① \\ X+Y = n \ ② \end{cases} \Rightarrow \begin{cases} X = \frac{1}{2}(n+b-a) \leftarrow \text{need } X\text{'s } 1 \text{ in } w \\ Y = \frac{1}{2}(n-b+a) \end{cases}$

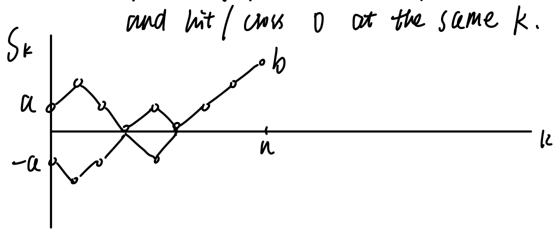
Then: $N_n(a,b) = \binom{n}{X} = \binom{n}{\frac{1}{2}(n+b-a)} = \#$ of sequences in Ω for which $X(w) = \frac{1}{2}(n+b-a)$

$\mathbb{P}^a(S_n=b) = \mathbb{P}^a(X = \frac{1}{2}(n+b-a)) = \begin{cases} \binom{n}{\frac{1}{2}(n+b-a)} 2^{-n} & \text{if } n \,\&\, (b-a) \text{ has the same parity.} \\ 0 & \text{otherwise} \end{cases}$

- Now assume $a > 0$, $b > 0$.

$N_n^0(a,b) = \#$ of paths of length n from a to b that hit/cross 0 at some point.

- Reflection Principle: for every path that hit/cross 0, there is a corresponding path starting from $-a$ and hit/cross 0 at the same k. (1-1 correspondence).



So, $N_n^0(a,b) = N_n(-a,b)$

$\mathbb{P}^a(S_n=b,\ S_k=0 \text{ for some } k) = \begin{cases} N_n^0(a,b) \cdot 2^{-n} & \text{if } a>0, b>0, n \,\&\, (a-b) \text{ have the same parity.} \\ 0 & \text{otherwise.} \end{cases}$

LECTURE 14.

Brief Summary.

- Ballot Theorem

Consider $N_n^+(0,b) =$ paths that from $S_0=0$ to $S_n=b$ for which $S_k>0$ $\forall k\in\{1,2,\cdots,n\}$

For any $b\neq 0$, $N_n^+(0,b)=\frac{|b|}{n}N_n(0,b)$, $\frac{|b|}{n}$: fraction of paths that hit x-axis.

proof: Suppose $b>0$. $N_n^+(0,b) = N_{n-1}^+(1,b)$ first step must step up

$$= N_{n-1}(1,b) - N_{n-1}^0(1,b) \quad \text{all paths - paths that touch/cross 0}$$

$$= N_{n-1}(1,b) - N_{n-1}(-1,b)$$

$$= N_{n-1}(1,b) - N_{n-1}(0,b+1)$$

$$= \binom{n-1}{\frac{n-1+b-1}{2}} - \binom{n-1}{\frac{n-1+b+1}{2}}$$

$$= \frac{b}{n}N_n(0,b) = \frac{b}{n}\binom{n}{\frac{n+b}{2}}$$

$$\mathbb{P}_n^0(S_k>0\ \forall k\in\{1,2,\cdots,n\} \mid S_n=b\} = \frac{\mathbb{P}_n^0(S_n>0\ \forall k\in\{1,\cdots,n\}, S_n=b)}{\mathbb{P}_n^0(S_n=b)}$$

$$= \frac{N_n^+(0,b)\cdot 2^{-n}}{N_n(0,b)\cdot 2^{-n}} = \frac{b}{n}$$

- Central Limit Theorem.

Suppose $\{X_i\}_{i=1}^\infty$ i.i.d $\mathbb{E}[X_i]=\mu$, $var(X_i)=\sigma^2$. $\mathbb{E}[X_k^4]<\infty$.

Then $\forall \alpha,\beta\in\mathbb{R}$, $\lim\limits_{n\to\infty}\mathbb{P}\left(\alpha\le\sqrt{n}\frac{\bar{X}_n-\mu}{\sigma}\le\beta\right)=\int_\alpha^\beta\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{y^2}{2}\right)dy$

$$\int_{-\infty}^{\infty}e^{-\frac{y^2}{2}}dy = I \qquad I^2 = \left(\int_{-\infty}^{\infty}e^{-\frac{y^2}{2}}dy\right)\left(\int_{-\infty}^{\infty}e^{-\frac{x^2}{2}}dx\right)$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{-\frac{(x^2+y^2)}{2}}dxdy$$

$$= \int_0^{\infty}e^{-\frac{r^2}{2}}2\pi r\,dr \quad \text{polar coordinates.}$$

$$= 2\pi$$

$$\sqrt{n}\frac{\bar{X}_n-\mu}{\sigma}\xrightarrow{d}N(0,1)$$

$$\lim_{n\to\infty}\mathbb{P}\left(\alpha\le\sqrt{n}\frac{\bar{X}_n-\mu}{\sigma}\le\beta\right)=\Phi(\beta)-\Phi(\alpha)$$

Important Example

Toss p-coin repeatedly, $Z_n=\#$ of heads.

$Z_n\sim Binom(n,p)$. $\mathbb{P}(Z_n=k)=\binom{n}{k}p^k(1-p)^{n-k}$

Think that $X_k\in\{0,1\}$

$$\mathbb{P}\left(\alpha\le\sqrt{n}\frac{\bar{X}_n-p}{\sqrt{p(1-p)}}\le\beta\right)=\int_\alpha^\beta\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{y^2}{2}\right)dy \quad \text{as } n\to\infty$$

$$\sqrt{n}\frac{\bar{X}_n-p}{\sqrt{p(1-p)}} = \frac{(X_1+\cdots+X_n)-np}{\sqrt{np(1-p)}}$$

$$=\mathbb{P}\left(\alpha\sqrt{np(1-p)}+np\le Z_n\le\beta\sqrt{np(1-p)}+np\right).$$

LECTURE 15

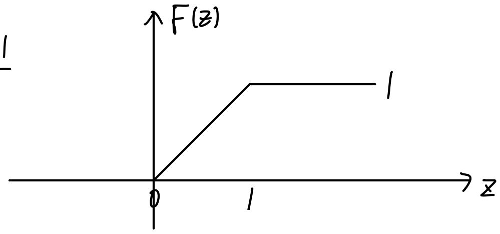
Brief Summary.

Def. Let X be any $\mathbb{R}$ valued r.v. The CDF of X is $F(t) = \mathbb{P}[X \le t]$. $F: \mathbb{R} \to \Omega$, non-decreasing

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

e.g. $X \sim \text{Unif}(0,1)$, $\forall I \subseteq \mathbb{R}$, $\mathbb{P}(X \in I) = \frac{|I \cap (0,1)|}{|(0,1)|}$

$$F(z) = \begin{cases} 0 & , z < 0 \\ z & , 0 \le z \le 1 \\ 1 & , z > 1 \end{cases}$$



Def. If the CDF of X is continuous then X is continuously distributed.

Proposition: For any continuously distributed r.v. X. $\mathbb{P}(X = z) = 0$ for all z.

$$F(b) - F(a) = \mathbb{P}(X \le b) - \mathbb{P}(X \le a)$$
$$= \mathbb{P}(X \in (a, b])$$

Fix z, let $\varepsilon > 0$, $\mathbb{P}(X = z) \le \mathbb{P}(X \in (z - \varepsilon, z])$
$$= F(z) - F(z - \varepsilon)$$
$\to 0$ as $\varepsilon \to 0$ if F is continuous.

Instead we say that a continuous dist. r.v. X has a density $f(x)$

if $\forall a, b$, $\mathbb{P}(X \in (a, b)) = \int_a^b f(x)\,dx = \mathbb{P}(X \in [a, b])$

Notice: $\mathbb{P}(X \in (a, b)) = F(b) - F(a) = \int_a^b f(x)\,dx$

$$\frac{F(z+\varepsilon) - F(z)}{\varepsilon} = \frac{\int_z^{z+\varepsilon} f(z)\,dz}{\varepsilon} \qquad \varepsilon > 0$$

as $\varepsilon \to 0$: $F'(z) = f(z)$

If $f(x)$ is any function on $\mathbb{R}$ such that 1) $f(x) \ge 0$ 2) $\int_{\mathbb{R}} f(x)\,dx = 1$, then it can be the density for a r.v.

Important Examples

① $X \sim \text{Unif}(0,1)$. Then the density of X is $f(x) = \begin{cases} 1 & x \in (0,1) \\ 0 & x \notin (0,1) \end{cases}$

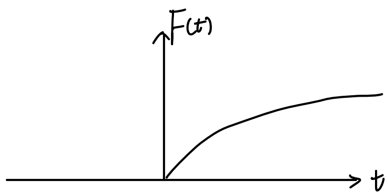
$f(x) \ne \mathbb{P}(X = x)$

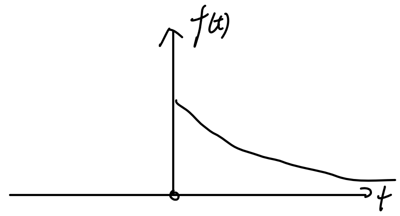
② Fix $\lambda > 0$. $X \sim \text{Exp}(\lambda)$ Exponential Distribution if

$$\mathbb{P}(X \ge t) = \begin{cases} e^{-\lambda t} & \text{if } t \ge 0 \\ 0 & \text{if } t < 0 \end{cases}$$

$$F(t) = \mathbb{P}(X \le t) = \begin{cases} 1 - e^{-\lambda t} & \text{if } t \ge 0 \\ 0 & \text{if } t < 0 \end{cases}$$

$$f(t) = F'(t) = \begin{cases} \lambda e^{-\lambda t} & \text{if } t > 0 \\ 0 & \text{if } t < 0 \end{cases}$$





$\mathbb{E}[X] = \frac{1}{\lambda}$ $\quad \text{Var}(X) = \frac{1}{\lambda^2}$

Intuition: Toss a coin until head. Z = # tosses until first head occurs $\sim \text{Geo}(p)$

if $p = \frac{\lambda}{n}$ $\quad \mathbb{P}[\frac{Z}{n} > t] = \mathbb{P}[Z > nt] \to e^{-\lambda t}$ as $n \to \infty$ $\qquad \lambda = \frac{1}{\mathbb{E}[Z]}$

③. let $\mu \in \mathbb{R}$, $\sigma^2 > 0$. $X \sim N(\mu, \sigma^2)$ if

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

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

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

LECTURE 16

Brief Summary:

- Expectation. Suppose X is continuously distributed & has density $f_X(x)$.
  Then $\mathbb{E}[X] = \int_{\mathbb{R}} x f_X(x)\,dx = \int_{-\infty}^{\infty} z f_X(z)\,dz$, assuming $\int_{-\infty}^{\infty} |z|\, f_X(z)\,dz < \infty$

- Properties: ① Linearity: $\alpha, \beta \in \mathbb{R}$: $\mathbb{E}[\alpha X + \beta Y] = \alpha \mathbb{E}[X] + \beta \mathbb{E}[Y]$
  ② if $g: \mathbb{R} \to \mathbb{R}$. Consider $Y = g(X)$. then $\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(z) f(z)\,dz$.

- Variance: $\text{var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathbb{E}[(X - \mathbb{E}[X])^2]$
- Tail Sum formula: If $X \geq 0$, then $\mathbb{E}[X] = \underbrace{\int_0^{\infty} \mathbb{P}(X > t)\,dt}_{(*)} = \int_0^{\infty} (1 - F_X(t))\,dt$

- Markov & Chebyshev Inequality: $\mathbb{E}[X] = \int_0^{\infty} \mathbb{P}(X>t)\,dt \geq \underbrace{\int_0^{r} \mathbb{P}(X>t)\,dt}_{\forall r > 0} \geq \int_0^{r} \mathbb{P}(X>r)\,dt = r\,\mathbb{P}(X>r)$
  $\mathbb{P}(X>t)$ non increasing.
  $\therefore\ \mathbb{P}(X>r) \leq \dfrac{\mathbb{E}[X]}{r}$, $\forall\ r > 0$.

Important Examples

- $X \sim \text{Unif}(a, b)$, fix $a < b$.

$$f_X(z) = \begin{cases} \frac{1}{b-a} & \text{if } z \in (a, b) \\ 0 & \text{otherwise} \end{cases} \qquad F_X(z) = \begin{cases} 0, & \text{if } z \in (-\infty, a] \\ \frac{z-a}{b-a}, & \text{if } z \in (a, b) \\ 1, & \text{if } z \in [b, +\infty) \end{cases}$$

$$\mathbb{E}[X] = \int_{\mathbb{R}} z f_X(z)\,dz = \int_a^b \frac{z}{b-a}\,dz = \frac{1}{b-a}\left(\frac{1}{2}z^2\Big|_a^b\right) = \frac{b+a}{2}$$

$$\mathbb{E}[X^2] = \int_a^b z^2 f_X(z)\,dz = \frac{1}{b-a}\left(\frac{1}{3}z^3\Big|_a^b\right)$$

Or: If $X \sim \text{Unif}(a, b)$, $\text{var}(X) = \mathbb{E}[(X-\mu)^2]$, $\boxed{\mu = \mathbb{E}[X] = \frac{b-a}{2}}$

$Z = \frac{X-a}{b-a} \sim \text{Unif}(0, 1)$, so $X = a + (b-a)Z$

Consider $W = X - \left(\frac{b+a}{2}\right)$. shift the r.v. to be symmetric around 0.

$W \sim \text{Unif}\left(\frac{-(b-a)}{2}, \frac{(b-a)}{2}\right)$

compute $\text{var}(W)$ where $W \sim \text{Unif}(-\ell, \ell)$:

$$\text{var}(W) = \mathbb{E}[W^2] - \mathbb{E}[W]^2 = \mathbb{E}[W^2] = \int_{\mathbb{R}} z^2 f_W(z)\,dz = \int_{-\ell}^{\ell} z^2 \frac{1}{2\ell}\,dz = \frac{z^3}{6\ell}\Big|_{-\ell}^{\ell} = \frac{\ell^3 - (-\ell)^3}{6\ell} = \frac{\ell^2}{3}.$$

when $\ell = \frac{a-b}{2}$: $\boxed{\text{var}(X) = \text{var}(W) = \frac{(b-a)^2}{12}}$

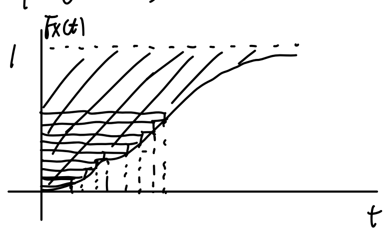
- If $W \sim \text{Unif}(-\ell, \ell)$, then $\text{var}(W) = \mathbb{E}[W^2] = \frac{\ell^2}{3}$.

- $X \sim \text{Exp}(\lambda)$ $\lambda > 0$. $f_X(t) = \begin{cases} \lambda e^{-\lambda t}, & t \geq 0 \\ 0, & t \leq 0 \end{cases}$ $\quad \mathbb{P}(X>t) = \begin{cases} e^{-\lambda t}, & t \geq 0 \\ 1, & t \leq 0 \end{cases}$ $\quad F_X(t) = \begin{cases} 1 - e^{-\lambda t}, & t \geq 0 \\ 0, & t \leq 0 \end{cases}$

$\lambda\uparrow$: more concentrated at 0.

$\mathbb{E}[X] = \int_{\mathbb{R}} z f_X(z)\,dz = \int_0^{\infty} z \lambda e^{-\lambda z}\,dz$ (intergration by parts) $= \frac{1}{\lambda}$.

or: $\mathbb{E}[X] = \int_0^{\infty} \mathbb{P}(X>t)\,dt = \int_0^{\infty} (1 - F_X(t))\,dt = \int_0^{\infty} e^{-\lambda t}\,dt = \left(-\frac{1}{\lambda}e^{-\lambda t}\Big|_0^{\infty}\right) = \frac{1}{\lambda}$

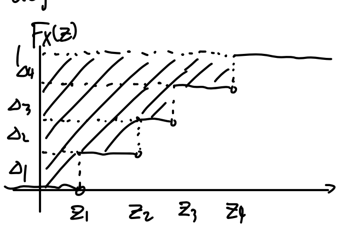


Compare tail sum to other definition:

Discrete: $\mathbb{E}[X] = \sum_{z \in R(X)} z\,\mathbb{P}(X=z)$.

Assume $X \geq 0$:

Claim: $\mathbb{E}[X] = \int_0^{\infty} \mathbb{P}(X>t)\,dt$
$= \int_0^{\infty} (1 - F_X(t))\,dt$
$= \sum_{z \in R(Z)} z\,\mathbb{P}(X=z)$



$\Delta_1 = \mathbb{P}(X = z_1)$
$\Delta_2 = \mathbb{P}(X = z_2)$
$\Delta_3 = \mathbb{P}(X = z_3)$
$\Delta_4 = \mathbb{P}(X = z_4)$

Continuous: $\mathbb{E}[X] = \int_{\mathbb{R}} z f_X(z)\,dz$

Assume $X \geq 0$: $\mathbb{E}[X] = \int_0^{\infty} \mathbb{P}(X>t)\,dt$
$= \int_0^{\infty} (1 - F_X(t))\,dt$
$= \int_0^{\infty} t\,dF_X(t) = \int_0^{\infty} t f_X(t)\,dt$

LECTURE 17

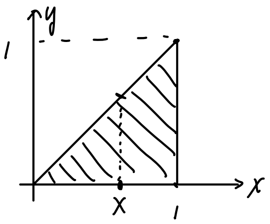
Brief Summary.

- Continuous Joint Distributions: X & Y have a joint density $f_{XY}(x,y)$ if $\mathbb{P}((X,Y)\in B) = \iint_B f_{XY}(x,y)\,dx\,dy$ for all open set $B \subset \mathbb{R}^2$
- Continuous Marginal Density of X is $f_X(x) = \int_{\mathbb{R}} f_{XY}(x,y)\,dy$
  $f_Y(y) = \int_{\mathbb{R}} f_{XY}(x,y)\,dx$

- Independence: Suppose X and Y have densities f_X and f_Y, they are independent iff joint density is $f_{XY}(x,y) = f_X(x) f_Y(y)$ for all $x, y \in \mathbb{R}$.

Important Examples.

1. Let T be the triangle. Pick $X \sim \text{unif}(0,1)$. Given X, choose $Y \sim \text{unif}(0,X)$



integrate y [0, x] first

$$\mathbb{P}(X<a) = \iint_B f_{XY}(x,y)\,dx\,dy \quad , B = \{(x,y) \mid x<a\}$$

Let $(X,Y) \sim \text{Unif}(T)$, then $f_{XY}(x,y) = \begin{cases} c & , (x,y)\in T \\ 0 & , (x,y)\notin T \end{cases}$

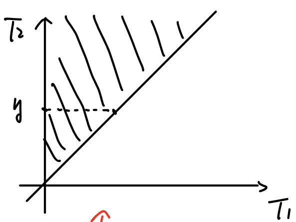
Since $\iint_{\mathbb{R}^2} f_{XY}(x,y)\,dx\,dy = \iint_T c\,dx\,dy = c \cdot \text{Area}(T) = 1$
$c = \text{Area}(T)^{-1}$.

$$\mathbb{P}(X<a) = \iint_B f_{XY}(x,y)\,dx\,dy \qquad B=\{(x,y) \mid x<a\}$$
$$= \iint_{T\cap\{x<a\} = \Delta a} f_{XY}(x,y)\,dx\,dy = \frac{\text{Area}(\Delta a)}{\text{Area}(T)} = \frac{\left(\frac{a^2}{2}\right)}{\left(\frac{1}{2}\right)} = a^2 \quad \text{for } a\in[0,1]$$

X is not uniformly distributed.

Generally, $\mathbb{P}(X\le a) = \begin{cases} 0 & a<0 \\ a^2 & a\in[0,1] \\ 1 & a>1 \end{cases}$ $\qquad f_X(a) = \begin{cases} 0 & a<0 \\ 2a & a\in[0,1] \\ 0 & a>1 \end{cases}$

2. Consider 2 independent arrivals. $T_1 \sim \text{Exp}(\lambda_1)$, $T_2 \sim \text{Exp}(\lambda_2)$
   What is $\mathbb{P}(T_1<T_2)$? What is the dist. of $T_* = \min\{T_1, T_2\}$?



integrate x [0, y] first.

Marginal density. T_1: $f_1(x) = \begin{cases} \lambda e^{-\lambda x} & , x\ge 0 \\ 0 & \text{otherwise} \end{cases}$ $\qquad T_2$: $f_2(y) = \begin{cases} \lambda e^{-\lambda y} & , y\ge 0 \\ 0 & , \text{otherwise} \end{cases}$

$$\mathbb{P}((T_1,T_2)\in B) = \iint_B f_{12}(x,y)\,dx\,dy \qquad B = \{(x,y) \mid x<y\}$$

$$f_{12}(x,y) = f_1(x) f_2(y) = \begin{cases} \lambda_1\lambda_2 e^{-\lambda_1 x} e^{-\lambda_2 y} & , x>0,\ y>0 \\ 0 & , \text{otherwise} \end{cases}$$

$$\mathbb{P}(T_1<T_2) = \int_0^\infty \int_0^y \lambda_1\lambda_2 e^{-\lambda_1 x} e^{-\lambda_2 y}\,dx\,dy$$
$$= \int_0^\infty \lambda_1\lambda_2 e^{-\lambda_2 y} \left(\int_0^y e^{-\lambda_1 x}\,dx \right) dy$$
$$= -\frac{1}{\lambda_1} e^{-\lambda_1 x}\Big|_0^y$$
$$= -\frac{1}{\lambda_1}\left(e^{-\lambda_1 y} - 1\right)$$
$$= -\frac{1}{\lambda_1} e^{-\lambda_1 y} + \frac{1}{\lambda_1}$$
$$= \int_0^\infty \lambda_1\lambda_2 e^{-\lambda_2 y} \left(-\frac{1}{\lambda_1}\right)\left(e^{-\lambda_1 y} - 1\right) dy$$
$$= -\lambda_2 \int_0^\infty \left(e^{-(\lambda_1+\lambda_2)y} - e^{-\lambda_2 y} \right) dy$$
$$= -\lambda_2 \left(\frac{1}{-(\lambda_1+\lambda_2)} e^{-(\lambda_1+\lambda_2)y}\Big|_0^\infty - \frac{1}{-\lambda_2} e^{-\lambda_2 y}\Big|_0^\infty \right)$$
$$= -\lambda_2 \left(\frac{1}{\lambda_1+\lambda_2} - \frac{1}{\lambda_2} \right)$$
$$= -\lambda_2 \, \frac{\lambda_2 - (\lambda_1+\lambda_2)}{(\lambda_1+\lambda_2)\lambda_2}$$
$$= \frac{\lambda_1}{\lambda_1+\lambda_2}.$$

Similarly, $\mathbb{P}(T_1 > T_2) = \dfrac{\lambda_2}{\lambda_1+\lambda_2}$.

LECTURE 18

- Joint Distribution: X, Y has a joint distribution in $\mathbb{R}^2$ if $\mathbb{P}((X,Y)\in B) = \iint_B f_{XY}(x,y)\,dx\,dy$
- Marginals: $f_X(x) = \int_{\mathbb{R}} f_{XY}(x,y)\,dy$
- If X & Y are independent, then $f_{XY}(x,y) = f_X(x) f_Y(y)$. In general, marginal $\not\Rightarrow$ joint density.
- Conditional densities:

Discrete case: $\mathbb{P}(X=k \mid Y=j) = \dfrac{\mathbb{P}(X=k, Y=j)}{\mathbb{P}(Y=j)} = \dfrac{\mathbb{P}(X=k, Y=j)}{\sum_{l\in R(X)} \mathbb{P}(X=l, Y=j)}$

Density case: If X & Y has a joint density $f_{XY}(x,y)$, the conditional density of X given $Y=y$ is:

$$f_{X|Y}(x \mid Y=y) = \frac{f_{XY}(x,y)}{\int_{\mathbb{R}} f_{XY}(l,y)\,dl} = \frac{f_{XY}(x,y)}{f_Y(y)}$$

$$\int_{\mathbb{R}} f_{X|Y}(x \mid Y=y)\,dx = \int_{\mathbb{R}} \frac{f_{XY}(x,y)}{f_Y(y)}\,dx = f_Y(y)/f_Y(y) = 1$$

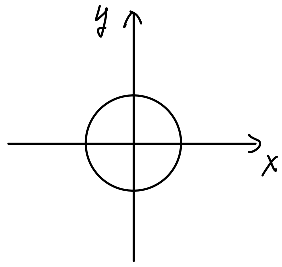
Discrete: $\mathbb{P}(X=k, Y=j) = \mathbb{P}(X=k \mid Y=j)\,\mathbb{P}(Y=j)$

$\mathbb{P}(X=k) = \sum_{j\in R(Y)} \mathbb{P}(X=k, Y=j) = \sum_{j\in R(Y)} \mathbb{P}(X=k \mid Y=j)\,\mathbb{P}(Y=j)$

Density. $f_{XY}(x,y) = f_{X|Y}(x \mid Y=y)\, f_Y(y)$

$f_X(x) = \int_{\mathbb{R}} f_{XY}(x,y)\,dy = \int_{\mathbb{R}} f_{X|Y}(x \mid Y=y)\, f_Y(y)\,dy.$

Important Examples.



e.g. let D_1 = unit disc $= \{(x,y)\in\mathbb{R}^2 \mid \sqrt{x^2+y^2} \le 1\}$

$(X,Y) \sim \mathrm{Unif}(D_1)$ $\quad f_{XY}(x,y) = \begin{cases} 0 & (x,y)\notin D_1 \\ \frac{1}{\mathrm{Area}(D_1)} & (x,y)\in D_1 \end{cases} = \frac{1}{\pi}$

Let $R = \sqrt{X^2+Y^2}$, Range$(R) = [0,1]$

$\mathbb{P}(R<t) = \begin{cases} 1 & \text{if } t \ge 1 \\ 0 & \text{if } t \le 0 \\ \frac{\mathrm{Area}(D_t)}{\mathrm{Area}(D_1)} & \text{if } t\in(0,1) = \frac{\pi t^2}{\pi 1} = t^2 \end{cases}$

Density for R: $f_R(t) = \frac{d}{dt}\mathbb{P}(R<t) = \begin{cases} 0 & t \le 0 \\ 2t & t\in(0,1) \\ 0 & t \ge 1 \end{cases}$

$\mathbb{E}[R] = \mathbb{E}[\sqrt{X^2+Y^2}]$

2 ways to calculate. ① $\mathbb{E}[R] = \int_{\mathbb{R}} t\, f_R(t)\,dt = \int_0^1 t\cdot 2t\,dt = \frac{2}{3}t^3\Big|_0^1 = \frac{2}{3}$

② $\mathbb{E}[\sqrt{X^2+Y^2}] = \iint_{D_1} \sqrt{x^2+y^2}\, f_{XY}(x,y) = \int_0^1\int_0^1 \sqrt{x^2+y^2}\,\frac{1}{\pi}\,dx\,dy.$

$f_{XY}(x,y) = \begin{cases} 0 & (x,y)\notin D_1 \\ \frac{1}{\pi} & (x,y)\in D_1 \end{cases}$ $\quad \sqrt{x^2+y^2} \le 1 \Rightarrow -\sqrt{1-x^2} \le y \le \sqrt{1-x^2}$

$f_X(x) = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} f_{XY}(x,y)\,dy = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi}\,dy = \frac{2\sqrt{1-x^2}}{\pi}$. for $-1 \le x \le 1$, 0 otherwise.

e.g. $T_1 \sim \mathrm{Exp}(\lambda)$ $T_2 \sim \mathrm{Exp}(\lambda)$. $T_{min} = \min\{T_1, T_2\}$. $T_{max} = \max\{T_1, T_2\}$

$\mathbb{P}(T_1 < T_2) = \frac{\lambda_1}{\lambda_1+\lambda_2} = \frac{\lambda}{2\lambda} = 2$. $\mathbb{P}(\min\{T_1,T_2\} > s) = \mathbb{P}(T_1 > s, T_2 > s) = \mathbb{P}(T_1 > s)\mathbb{P}(T_2 > s) = e^{-\lambda s}e^{-\lambda s} = e^{-2\lambda s}$ if $s > 0$.

$\therefore \min\{T_1, T_2\} \sim \mathrm{Exp}(2\lambda)$. Induction: $T_1 \cdots T_n$ indep. $\mathrm{Exp}(\lambda) \Rightarrow \min(T_1 \cdots T_n) \sim \mathrm{Exp}(n\lambda)$

$\mathbb{P}(T_{max} \le s) = \mathbb{P}(T_1 \le s, T_2 \le s) = \mathbb{P}(T_1 < s)\mathbb{P}(T_2 < s) = (1-e^{-\lambda s})(1-e^{-\lambda s})$ if $s > 0$

$f_{T_{max}}(s) = \frac{d}{ds}\mathbb{P}(T_{max} \le s) = 2(1-e^{-\lambda s})\cdot\lambda = 2\lambda(1-e^{-\lambda s})$ if $s > 0$. 0 otherwise

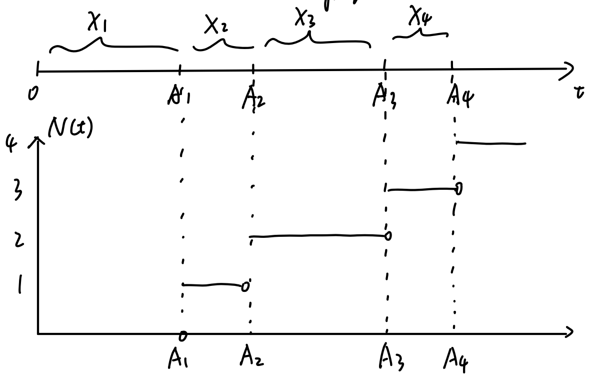
$f_{T_{min}}(s) = \frac{d}{ds}\mathbb{P}(T_{min} \le s) = \frac{d}{dt}(1-e^{-2\lambda s}) = 2\lambda e^{-2\lambda s}$ if $y \ge 0$. 0 otherwise.

LECTURE 19

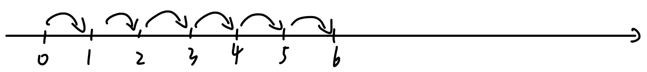
Brief Summary:

Poisson Arrival Process: N

- Let $\lambda > 0$. Let $X_1, X_2, \cdots\cdots$ be independent r.v. with $X_i \sim Exp(\lambda)$. Interprete X_i as the i^{th} "inter-arrival" or "waiting time".
- Define $A_n = \sum_{k=0}^{n} X_k$ = time of n^{th} arrival, $A_{n+1} - A_n = X_{n+1}$
- $N(0) = 0$, $N(t) = \max\{n \geq 0 \mid A_n \leq t\} = \max\{n \geq 0 \mid \sum_{k=0}^{n} X_k \leq t\}$, $t \in [0, +\infty)$: # of arrivals in time $\in [0, t]$
  $N(t) \geq 0$, non-decreasing, piecewise constant. $N(t) - N(t^-) = 1$ if $t = A_n$ for some n.

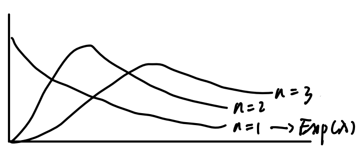


Another View: Jumping process on integer $\mathbb{Z}$.



- For an interval $I \in (a, b] \subset (0, +\infty)$, define $N(I)$ = # of arrivals in I = # of arrivals in I $= N(b) - N(a)$
  $N(t) = N((0, t]) = N(t) - N(0)$
- Theorem: Let N be a Poisson Arrival Process (PAP) with parameter λ
  (i). If $I = (a, b]$ is any interval, then $N(I) = N(b) - N(a)$ has a Poisson$(\lambda|I|)$ distribution. $\mathbb{P}(N(I) = k) = \frac{(\lambda|I|)^k}{k!} e^{-\lambda|I|}$
  In particular, $N(t) \sim Poisson(\lambda t)$ $\mathbb{P}(N(t) = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$, $k = 0, 1, \cdots$
  $\mathbb{E}[N(I)] = \lambda|I|$. As $\lambda \uparrow$, $\mathbb{E}[X_i] = \frac{1}{\lambda} \downarrow$, inter arrival $\downarrow$, more arrivals.
  (ii) For any disjoint intervals $I_j \in (a_j, b_j]$ $j = 1, \cdots, n$, the r.v.s $N(I_1), N(I_2), \cdots N(I_n)$ are independent.
  Fix times, $0 = t_1 < t_2 < \cdots < t_n$, increments $\{(N(t_{k+1}) - N(t_k))\}_{k=0}^{n-1} = \{N((t_k, t_{k+1}])\}_{k=0}^{n-1}$ are independent.
- Proposition: the n^{th} arrival time A_n has a Gamma(n, λ) distribution. Gamma(n, λ) is a continuous dist. on $[0, +\infty)$ with density $g_n(t) = \begin{cases} \frac{(\lambda t)^{n-1}}{(n-1)!} \lambda e^{-\lambda t} & , t \geq 0 \\ 0 & , t < 0. \end{cases}$

  Gamma(n, λ) is the dist of a sum of n independ. Exp(λ) r.v.s.



proof. $A_n = \sum_{k=1}^{n} X_k$.

$A_1 = X_1 \sim Exp(\lambda) = Gamma(1, \lambda)$

Induction: Suppose $A_n \sim Gamma(n, \lambda)$.

Then: $A_{n+1} = A_n + X_{n+1}$. f_{n+1} = density of A_{n+1}; $\underset{\downarrow Gamma(n,\lambda)}{g_n}$ = density for A_n; $g_1 = g$ = density for X_{n+1} $\;\downarrow Gamma(1,\lambda) = Exp(\lambda)$

$f_{n+1}(t) = g_n * g(t) = \int_{\mathbb{R}} g_n(t-s) g(s) ds$ by convolution

$= \int_0^\infty \underline{g_n(t-s)} g(s) ds$ $\left(g(t-s) = 0 \text{ if } s > t. \text{ So, } 0 < s < t \right)$ ☆

$= \int_0^t \frac{(\lambda(t-s))^{n-1}}{(n-1)!} \lambda e^{-\lambda(t-s)} \cdot \lambda e^{-\lambda s} ds$

$= \frac{\lambda^{n-1}}{(n-1)!} \lambda^2 \int_0^t (t-s)^{n-1} e^{-\lambda t + \lambda s - \lambda s} ds$

$= \frac{\lambda^{n-1}}{(n-1)!} \lambda^2 e^{-\lambda t} \underbrace{\int_0^t (t-s)^{n-1} ds}_{I}$ $\left(I = -\frac{1}{n}(t-s)^n \Big|_0^t = \frac{t^n}{n} \right)$

$= \frac{(\lambda t)^n}{n!} \lambda e^{-\lambda t}$, Gamma$(n+1, \lambda)$

- Memoryless: If $X \sim Exp(\lambda)$, then: $\mathbb{P}(X > t+s \mid X > t) = \mathbb{P}(X - t > s \mid X > t) = \mathbb{P}(X > s)$. $\forall t > 0, s > 0$.

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

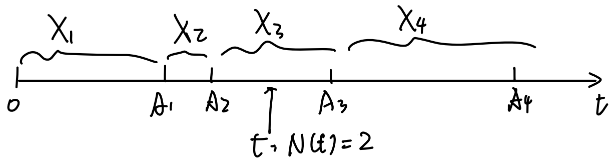
LECTURE 20

- Poisson Arrival Process

Two equivalent description:

(I). $X_1, X_2, X_3, \cdots$ be independent r.v. $X_i \sim Exp(\lambda)$

Let $A_n = \sum_{k=1}^{n} X_k = n^{th}$ arrival time. $N(t) = \max\{n \mid A_n \leq t\} = \#$ of arrival up to time t.



$N(t)$ is a jumping process. $N(t) - N(s) = \#$ arrivals in $(s, t]$

(II) $N(t)$ is a random process (function) on $[0, +\infty)$ with 2 properties:

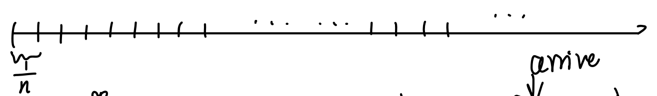
(i) $N(0) = 0$

(ii) For any $0 = t_0 < t_1 < \cdots < t_n$, $\{N(t_i) - N(t_{i-1})\}_{i=1}^{n}$ are independent.

(iii). $N(t) - N(s) \sim Poisson(\lambda|t-s|)$

Discrete Approximation:

Chop time into small intervals with length $\frac{1}{n}$.



Let $\{C_k\}_{k=1}^{\infty} \sim$ indep. Bernoulli$(\frac{\lambda}{n})$: $\mathbb{P}(C_k = 1) = \frac{\lambda}{n}$ (arrive), $\mathbb{P}(C_k = 0) = 1 - \frac{\lambda}{n}$ (not arrive). for each time interval

then: let $\tilde{N}(t) = \#$ of arrivals up to time t where heads occurs at time $t_k = \frac{k}{n}$, if $C_k = 1$

then: $\tilde{N}(t) = Binom(nt, \frac{\lambda}{n})$ since there are $t/\frac{1}{n}$ intervals in $[0, t]$.

let $m = nt$, then: $\frac{\lambda}{n} = \frac{\lambda t}{m}$, $\tilde{N}(t) \sim Binom(m, \frac{\lambda t}{m})$.

$$\mathbb{P}(\tilde{N}(t) = k) = \binom{m}{k}\left(\frac{\lambda t}{m}\right)^k\left(1 - \frac{\lambda t}{m}\right)^{m-k}$$

$$= \frac{m!}{k!(m-k)!}\frac{(\lambda t)^k}{m^k}\cdot\left(1 - \frac{\lambda t}{m}\right)^{m-k}$$

$$= \frac{(\lambda t)^k}{k!}\frac{m!}{(m-k)!}\cdot\frac{1}{m^k}\left(1 - \frac{\lambda t}{m}\right)^{m}\cdot\left(1 - \frac{\lambda t}{m}\right)^{-k}$$

$$= \frac{(\lambda t)^k}{k!}\frac{m(m-1)\cdots(m-k+1)}{m^k}\left(1 - \frac{\lambda t}{m}\right)^{-k}\left(1 - \frac{\lambda t}{m}\right)^{m}$$

$$\longrightarrow \frac{(\lambda t)^k}{k!}e^{-\lambda t}.$$

$\therefore \tilde{N}(t) \xrightarrow{d} Poisson(\lambda t)$ as $n \to \infty$.

Let $Z_n = \#$ tosses till heads, then. $Z_n \sim Geo(\frac{\lambda}{n})$.

The time lapse until first heads $= \frac{1}{n}Z_n$

Let $\tilde{X} = \frac{1}{n}Z_n$ = time lapse until first arrival.

$$\mathbb{P}(\tilde{X} > t) = \mathbb{P}\left(\frac{1}{n}Z_n > t\right) \qquad e^{-\lambda t}$$

$$= \mathbb{P}(Z_n > nt)$$

$$= \sum_{i=nt+1}^{\infty}\left(1 - \frac{\lambda}{n}\right)^{i-1}\frac{\lambda}{n} = \frac{\lambda}{n}\frac{\left(1 - \frac{\lambda}{n}\right)^{nt+1}}{1 - \left(1 - \frac{\lambda}{n}\right)} = \left(1 - \frac{\lambda}{n}\right)^{nt+1} = \left(1 - \frac{\lambda t}{nt}\right)^{nt}\left(1 - \frac{\lambda}{n}\right) \longrightarrow e^{-\lambda t} \quad \text{as } n \to \infty.$$

$\therefore \tilde{X} = \frac{1}{n}Z_n \xrightarrow{d} Exp(\lambda)$

- Let $X_1, X_2, \cdots, X_n$ iid $\mathbb{E}[X_i] = \mu$. $var(X_i) = \sigma^2$ for all i.

Weak LLN: $\lim_{n\to\infty}\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}X_i - \mu\right| > \varepsilon\right) = 0$

Strong LLN: $\mathbb{P}\left(\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}X_i(\omega) = \mu\right) = 1$.

LECTURE 21st.

- Markov chain: $\{X_n\}_{n\in\mathbb{N}}$, takes value in $\mathcal{S}$: state space. Have Markov Property:
- Markov Property: the future is independent of the past, condition on the present.
$$\mathbb{P}(X_{n+1}=y \mid X_n=x_n, X_{n-1}=x_{n-1}, \cdots X_0=x_0) = \mathbb{P}(X_{n+1}=y \mid X_n=x_n)$$
- Time homogeneous: $\mathbb{P}(X_{n+1}=y \mid X_n=x)$ does not depend on n.
  ↳ then, define Transition Probability Matrix as:
$$P(x,y) = \mathbb{P}(X_{n+1}=y \mid X_n=x) = \mathbb{P}(X_1=y \mid X_0=x) \quad \text{for any } n.$$
- Property of P: ① $0 \le P(x,y) \le 1$
  ② for any x: $\sum_{y\in S} P(x,y) = 1$

- $$\mathbb{P}(X_2=x_2 \mid X_0=x_0) = \sum_{x_1\in\mathcal{S}} \mathbb{P}(X_2=x_2, X_1=x_1 \mid X_0=x_0) = \sum_{x_1\in\mathcal{S}} \mathbb{P}(X_2=x_2 \mid X_1=x_1, X_0=x_0)\,\mathbb{P}(X_1=x_1 \mid X_0=x_0)$$
$$= \sum_{x_1\in\mathcal{S}} \mathbb{P}(X_2=x_2 \mid X_1=x_1)\,\mathbb{P}(X_1=x_1 \mid X_0=x_0) = \sum_{x_1\in\mathcal{S}} P(x_0,x_1)P(x_1,x_2) = P^{(2)}(x_0,x_2)$$

- Suppose X_0 is random. $\mathbb{P}(X_0=x_0)=V(x_0)$
  then: $\mathbb{P}(X_1=x_1) = \sum_{x_0\in\mathcal{S}} \mathbb{P}(X_1=x_1 \mid X_0=x_0)\,\mathbb{P}(X_0=x_0) = \sum_{x_0\in\mathcal{S}} V(x_0)P(x_0,x_1)$ row vector times matrix column x_1.

e.g. Urn model

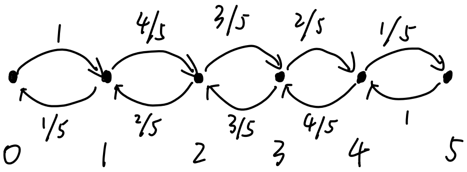
Consider an urn full of N red & blue marbels. At each step, we draw a marbel randomly from the urn and switch its color.

$\mathcal{S}=\{0,\dots,N\}$, numbers of red marbles.

If $i,j\in\mathcal{S}$, $P(i,j) = \begin{cases} 0 & \text{if } |i-j|\neq 1 \\ \frac{N-i}{N} & \text{if } j=i+1 \\ \frac{i}{N} & \text{if } j=i-1 \end{cases}$

Property of P: ① $0\le P(x,y)\le 1$

② for any x: $\sum_{y\in S} P(x,y)=1$.



Discrete time Moran model:

Consider a population of size N with two types: long and short.
At each step, a random individuals dies and another independent random individual gets cloned.

$\mathcal{S}=\{0,1,\dots,N\}$ numbers of long individuals

$$P(i,j) = \begin{cases} \frac{i}{N}\left(\frac{N-i}{N}\right) & \text{if } j=i+1,\ i\neq N \\ \left(\frac{N-i}{N}\right)\frac{i}{N} & \text{if } j=i-1,\ i\neq 0 \\ 1-2\frac{i}{N}\left(\frac{N-i}{N}\right) & \text{if } j=i \quad \text{(since the sum is 1).} \\ 0 & \text{otherwise} \end{cases}$$

$\left(\frac{N-i}{N}\right)^2+\left(\frac{i}{N}\right)^2$? $\quad \mathbb{P}(\{R \text{ clone}\}\cap\{R \text{ die}\} \mid X_n=i)$

In the example: 0 and N are called absorbing state since $P(0,0)=P(N,N)=1$.

A Markov Chain is not time homogeneous for a random walk on $\mathbb{Z}$: at time n, ... See note P4.

LECTURE 22.

- Time-homogeneous: $P(x,y) = \mathbb{P}(X_1=y \mid X_0=x) = \mathbb{P}(X_{n+1}=y \mid X_n=x)$
- n-step transition probability: $\mathbb{P}(X_n=y \mid X_0=x) = P^{(n)}(x,y)$

$\mathbb{P}(X_2=y \mid X_0=x) = \sum_{x_1\in S} \mathbb{P}(X_2=y, X_1=x_1 \mid X_0=x) = \sum_{x_1\in S} \mathbb{P}(X_2=y \mid X_1=x_1, X_0=x)\mathbb{P}(X_1=x_1 \mid X_0=x) = \sum_{x_1\in S}\mathbb{P}(X_2=y \mid X_1=x_1)\mathbb{P}(X_1=x_1 \mid X_0=x)$

$= \sum_{x_1\in S} P(x,x_1)P(x_1,y) = P^{(2)}(x,y)$

- If X_0 is random, $X_0 \sim \nu$, a probability distribution on S, that satisfies $\mathbb{P}(X_0=x) = \nu(x)$, then

$\mathbb{P}(X_n=y \mid X_0\sim\nu) = \sum_{x\in S}\mathbb{P}(X_n=y \mid X_0=x)\mathbb{P}(X_n=x) = \sum_{x\in S}\nu(x)P^{(n)}(x,y) = (\nu P^{(n)})(y)$

- Invariant / Stationary Distribution π on S: $\pi P = \pi$.

$\sum_{x\in S}\pi(x)P(x,y) = \pi(y)$ or $\mathbb{P}(X_1=y \mid X_0\sim\pi) = \pi(y)$

$\mathbb{P}(X_n=y \mid X_0\sim\pi) = \sum_{x\in S}\pi(x)P^{(n)}(x,y) = (\pi P^{(n)})(y) = (\pi P P^{(n-1)})(y) = (\pi P^{(n-1)})(y) = \cdots = \pi(y).$

π is the left eigenvector of P with eigenvalue 1.

Important Examples:

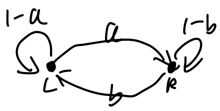
① R.V. on Graph $G=(V,E)$. Define $\pi(x) = \frac{\deg(x)}{D}$, $x\in V$. $d(x)$ is the degree of x. $D = \sum_{x\in V} d(x)$

$$\mathbb{P}(X_{n+1}=y \mid X_n=x) = \begin{cases} 0 & \text{if } (x,y)\notin E \\ \frac{1}{\deg(x)} & \text{if } (x,y)\in E \text{ or } x\sim y \end{cases}$$

Check that π is a stationary distribution.

$\forall y$, $(\pi P)(y) = \sum_{x\in S}\pi(x)P(x,y) = \sum_{\substack{x\in S\\ y\sim x}}\pi(x)P(x,y) = \sum_{\substack{x\in S\\ y\sim x}}\frac{\deg(x)}{D}\cdot\frac{1}{\deg(x)} = \sum_{\substack{x\in S\\ y\sim x}}\frac{1}{D} = \frac{\deg(y)}{D} = \pi(y)$

thus, $\pi P = \pi$.

② $S=\{L,R\}$, $P = \begin{pmatrix} 1-a & a \\ b & 1-b\end{pmatrix}$, 

then: $\pi = \left(\frac{b}{a+b}, \frac{a}{a+b}\right)$ is a stationary probability. Check: $\pi P = \left(\frac{b}{a+b}, \frac{a}{a+b}\right)\begin{pmatrix} 1-a & a \\ b & 1-b\end{pmatrix} = \left(\frac{b}{a+b}, \frac{a}{a+b}\right) = \pi$.

Claim. For any initial distribution ν_0, $\nu P^{(n)} \to \pi$ as $n\to\infty$, if $|1-a-b|<1$

Given ν_0, Let $\nu_n = \nu_0 P^{(n)}$, then $\nu_{n+1} = \nu_n P$

$\nu_{n+1}(L) = \nu_n(L)P(L,L) + \nu_n(R)P(R,L)$

$= \nu_n(L)(1-a) + (1-\nu_n(L))b$

$= \nu_n(L)(1-a-b) + b$

$\nu_{n+1}(L) - \frac{b}{a+b} = \nu_n(L)(1-a-b) + b - \frac{b}{a+b}$

$\nu_{n+1}(L) - \frac{b}{a+b} = \left(\nu_n(L) - \frac{b}{a+b}\right)(1-a-b)$

So, $|\nu_{n+1}(L) - \pi(L)| \le |1-a-b|\,|\nu_n(L)-\pi(L)| \le \cdots \le |1-a-b|^n|\nu_0(L)-\pi(L)|$, $n\ge 1$

When $|1-a-b|<1$, $\nu_n = \nu_0 P^{(n)} \to \pi$.

LECTURE 23

- Recurrence: A state $x \in S$ is recurrent if $\mathbb{P}(X_n = x \text{ for some } n \geq 1 \mid X_0 = x) = 1$

  Transience: A state $x \in S$ is transient if $\mathbb{P}(X_n = x \text{ for some } n \geq 1 \mid X_0 = x) = 1 - a < 1$

  or $\mathbb{P}(X_n \neq x \text{ for all } n \geq 1 \mid X_0 = x) = a > 0$. There is a chance of never coming back.

- Communicate: two states $x, y \in S$ communicate if $\exists\ n \geq 1,\ m \geq 1$, s.t. $P^{(n)}(x,y) > 0$ and $P^{(m)}(y,x) > 0$

  $x \leftrightarrow y$

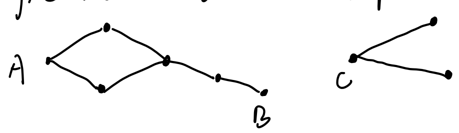
- Irreducible: the chain is irreducible if all pairs $x, y \in S$ communicate, that is, for any $x, y \in S$,

  $P^{(n)}(x,y) > 0$ for some n, $P^{(m)}(y,x) > 0$ for some m.

- For any chain, the states can be partitioned uniquely into $S = T \cup C_1 \cup C_2 \cup \cdots$, where T is the set of transient states and C_k $k = 1, 2, \cdots$ are closed communication classes of recurrent states.
  - Closed communication class C_k: 1) for all $x, y \in C_k$ $x \leftrightarrow y$. 2). $P(x,z) = 0$ whenever $x \in C_k$ but $z \notin C_k$.

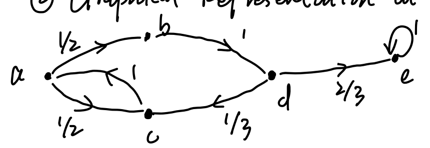
- Lemma 1.1. If $|S| < \infty$, then there is at least one recurrent state.
- Lemma 1.2. If the chain is irreducible, then either i) all states are recurrent or ii) all states are transient.

  e.g. ① Random Walk on Graph



not irreducible because B and C not communicate.

② Graphical Representation at a Chain

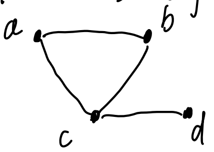


not irreducible $P^{(n)}(e, d) = 0$ for all $n \geq 1$

- Lemma. If there is $n \geq 1$ such that $P^{(n)}(x,y) > 0$ for all x, y then the chain is irreducible

- Period: $\forall x \in S$, define the period of x to be $d(x) = \gcd\{n \geq 1 \mid P^{(n)}(x,x) > 0\}$ greatest common divider.

  e.g.



$d(a) = \gcd\{2, 4, 6, \cdots$

$3, 6, 9, \cdots$

$5, 10, 15, \cdots$

$\cdots\cdots\cdots\}$

- If $x \leftrightarrow y$, then $d(x) = d(y)$. If the chain is irreducible, then all states have the same period.

  Define the period of the chain to be $d(x)$, $\forall x \in S$.
- The chain is aperiodic if its period $= 1$.
- Suppose $|S| < \infty$. If the chain is irreducible, then there is a unique invariant probability distribution π.

  If the chain is also aperiodic, then for any initial distribution ν, $\lim_{n \to \infty} \nu P^{(n)} = \pi$.

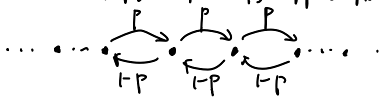
  Hence: $\lim_{n \to \infty} \nu(x) P^{(n)}(x,y) = \pi(y)$ for all $x, y \in S$.

  Furthermore, for any $F: S \to \mathbb{R}$. $\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} F(X_n) = \sum_{x \in S} F(x)\pi(x) = \mathbb{E}_\pi[F(x)]$ w.p.1. (does not depend on the initial distribution).

LECTURE 24th

- First Visit: for each $x \in S$, the first visit to x is $T_x = \min\{n \geq 1 \mid X_n = x\}$
  If X_n never reaches x, then $T_x = +\infty$.
- Mean Return Time to x: $\mu_x = \mathbb{E}[T_x \mid X_0 = x]$
- If x is transient, then $\mu_x = +\infty$ since $\mathbb{P}(T_x = +\infty \mid X_0 = x) > 0$.
  If x is recurrent, and $\begin{cases} \mu_x = +\infty: x \text{ is null recurrent} \\ \mu_x < +\infty: x \text{ is positive recurrent.} \end{cases}$

e.g. r.w. on $\mathbb{Z}$: For $k >> 1$, $\mathbb{P}(T_x = k \mid X_0 = k) \sim \frac{C}{k^{3/2}}$, $\sum_{k=1}^{\infty} k \mathbb{P}(T_x = k \mid X_0 = k) = +\infty$.



$p = \frac{1}{2}$: all states are null recurrent
$p \neq \frac{1}{2}$: all states are transient.

- Theorem: An irreducible chain has a stationary distribution π. Iff all states are positive recurrent.
  for all $x \in S$: $\pi(x) = \frac{1}{\mu_x} = \frac{1}{\mathbb{E}[T_x \mid X_0 = x]}$, π is unique.

- $|S| < \infty$ + irreducible $\Rightarrow$ all states are positive recurrent $\Rightarrow \exists$ stationary π: $\forall x \in S$. $\pi(x) = \frac{1}{\mu_x}$.
- $|S| < \infty$ + irreducible + aperiodic $\Rightarrow \forall$ initial dist ν, $\lim_{n \to \infty} \nu P^{(n)} = \pi$

e.g. Let $B \subset S$, $g: B \to \mathbb{R}$ some function. (General Case)
Consider $h(x) = \mathbb{E}[g(X_\tau) \mid X_0 = x]$
$\tau = \min\{n \geq 0 \mid X_n \in B\}$
If $x \in B$, then: $\tau = 0$, $h(x) = \mathbb{E}[g(X_0) \mid X_0 = x]$
$= g(x)$

If $x \notin B$, then: $\tau > 0$

$$h(x) = \mathbb{E}[g(X_\tau) \mid X_0 = x] = \sum_y \mathbb{E}[g(X_\tau) \mid X_1 = y, X_0 = x] \mathbb{P}(X_1 = y \mid X_0 = x)$$
$$= \sum_y \mathbb{E}[g(X_\tau) \mid X_1 = y] P(x, y)$$
$$= \sum_y P(x, y) h(y)$$

Therefore, h satisfies the linear equation systems:
$$\begin{cases} h(x) = \sum_y P(x, y) h(y) & \forall x \in S \setminus B \\ h(x) = g(x) & \forall x \in B \end{cases}$$

e.g. Let $B = \{a, b\}$, $g(y) = \mathbb{1}(y = b)$
$h(x) = \mathbb{E}[g(X_\tau) \mid X_0 = x] = \mathbb{P}(X_\tau = b \mid X_0 = x)$ since $g(X_\tau) = \mathbb{1}(X_\tau = b)$
since $\tau = \min\{n \geq 0 \mid X_n \in B\}$
$= \mathbb{P}(X_n \text{ reaches } b \text{ before } a \mid X_0 = x)$
then: $h(a) = g(a) = 0$, $h(b) = g(b) = 1$.

LECTURE 25th

Occupation Times & Absorbing States

Suppose a chain $\{X_n\}$ on an S ($|S| < \infty$) is irreducible. Let $B \subset S$, $A = S \setminus B$.

☆ For $x \in A$, how many steps will the chain take before reaching a state in the set B?

Define $\tau_B = \min\{n \geq 0 \mid X_n \in B\}$. the first time that X is in B.

What is $\mathbb{E}[\tau_B \mid X_0 = x]$?

Method 1: let $h(x) = \mathbb{E}[\tau_B \mid X_0 = x]$

$$h(x) = \sum_{y \in S} \mathbb{E}[\tau_B \mid X_1 = y, X_0 = x]\, \mathbb{P}(X_1 = y \mid X_0 = x) \qquad \tau_B = 1$$

$$= \sum_{y \in A} \mathbb{E}[\tau_B \mid X_1 = y, X_0 = x]\, \mathbb{P}(X_1 = y \mid X_0 = x) + \sum_{y \in B} \mathbb{E}[\tau_B \mid X_1 = y, X_0 = x]\, \mathbb{P}(X_1 = y \mid X_0 = x)$$

$$= \sum_{y \in A} P(x, y)\, \mathbb{E}[\tau_B \mid X_1 = y] + \underbrace{\sum_{y \in B} \mathbb{P}(X_1 = y \mid X_0 = x)}_{\text{this does not equals } 1.}$$

How can we rigorously deduce that $h(x) = \sum_{y \in A} P(x, y)\, \mathbb{E}[\tau_B \mid X_1 = y] + 1$?

Method 2: modify the chain: $X \to \tilde{X}$, $P \to \tilde{P}$ where $\tilde{P}(x, y) = \begin{cases} P(x, y) & \text{if } x \in A,\ y \in S \\ 1 & \text{if } x = y = B \\ 0 & \text{otherwise} \end{cases}$

$$\tilde{P} = \begin{array}{c|c} & \begin{array}{cc} A & B \end{array} \\ \begin{array}{c} A \\ B \end{array} & \left(\begin{array}{c} P \\ \hline \begin{array}{cc} 0 & \begin{array}{cc} 1 & 0 \\ 0 & \ddots\ 1 \end{array} \end{array} \end{array} \right) \end{array}$$

Then: state B becomes the absorbing state: once $\tilde{X}_n$ reaches a state $z \in B$, it will not leave.

Using tail sum formula:

$$\mathbb{E}[\tau_B \mid X_0 = x] = \sum_{k=0}^{\infty} \mathbb{P}(\tau_B > k \mid X_0 = x)$$

$$\mathbb{P}(\tau_B > k \mid X_0 = x) = \mathbb{P}(X_0 \in A, X_1 \in A, \cdots, X_k \in A \mid X_0 = x)$$

$= \mathbb{P}($all the states before k (including k) are not in the absorbing $B \mid X_0 = x)$

$$= \mathbb{P}(\tilde{X}_k \in A \mid X_0 = x)$$

$$= \sum_{y \in A} \tilde{P}^{(k)}(x, y)$$

Let M be an $|A| \times |A|$ submatrix of P, containing $P(x, y)$ where $x, y \in A$.

Then, for $x, y \in A$: $P^{(k)}(x, y) = M^{(k)}(x, y)$

$$\mathbb{E}[\tau_B \mid X_0 = x] = \sum_{k=0}^{\infty} \sum_{y \in A} \tilde{P}^{(k)}(x, y) = \sum_{k=0}^{\infty} \sum_{y \in A} M^{(k)}(x, y) = \sum_{y \in A} \left(\sum_{k=0}^{\infty} M^{(k)} \right)(x, y)$$

if all the eigenvalues of M has $|\cdot| < 1$, then $\sum_{k=0}^{\infty} M^{(k)} = (I - M)^{-1}$

$$\mathbb{E}[\tau_B \mid X_0 = x] = \sum_{y \in A} (I - M)^{-1}(x, y).$$