Harvard Stat 110

Mo Tu We Th Fr Sa Su

Memo No.

Date

it assumes that you have ~~split~~ broken up ~~resulting~~ problem ~~of~~ to equally likely outcomes

S1

* Sample Space : Set of all possible outcomes of some experiment

* experiment : can be anything, there are certain possible outcomes, before the experiment you don't know what is going to happen, you do the experiment, and you see an outcome

* An event is a subset of the sample space

* Naive Def of Prob. : $P(A) = \frac{\#\text{favorable outcomes}}{\#\text{possible outcomes}}$ (A: event)

→ Strong assumption (for naive def): all outcomes equally likely
finite sample space

→ we should get beyond naive bayes : 1) Prob is there a life on ~~a~~ neptune → 1/2    2) Prob inteligent life on neptune → also 1/2

* How do we find # possible ~~ob~~ outcomes?

* Counting : Multiplication rule

Binomial Coef : $\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$

# subset of size k, of group of n people (without order)

→ mult rule : $\frac{n \cdot (n-1) \cdot (n-2) \cdots (n-k+1)}{k!}$ ← because of order

* Story ~~Proof~~ : Proof by interpretation

Ex(1): $\binom{n}{k} = \binom{n}{n-k}$ / (2): $\underbrace{n\binom{n-1}{k-1}}_{\text{pick the boss first, then choose k-1 other for group}} = \underbrace{k\binom{n}{k}}_{\text{pick k people from a club, one of them should be boss}}$

~~it is a rigorous mathematical~~ proof / [illegible]

□ (3) $\binom{m+n}{k} = \sum_{j=0}^{k} \binom{m}{j}\binom{n}{k-j}$ (vandermonde)

□□ * non naive definition of probability, general def

(up until now we were assuming equally likely outcomes and we don't want that and we don't want that assume that there are finitely many possible outcomes so we want to go beyond that)

* ~~non-naive def~~: A probability space consists of S and P, where S is a sample space, and P, a function which takes an event $A \subseteq S$ as input, returns $P(A) \in [0, 1]$ as output

→ 2 axioms (rules) P should satisfy

such that: (1) $P(\phi) = 0$, $P(S) = 1$

(2) $P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$ if $A_1, A_2, \ldots, A_n$ are disjoint (non-overlapping)



* Sampling table: ~~We~~ choose k objects out of n

Sampling means we have some population of anything and we're drawing a sample

| | order matters | order doesn't matter |
|---|---|---|
| replace | $n^k$ | $\binom{n+k-1}{k}$ |
| don't replace | $n(n-1)(n-2) \cdots (n-k+1)$ | $\binom{n}{k}$ |

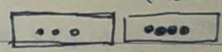
□ S2 * 10 people, divide to 2 5 player team → $\binom{10}{5} \div 2$

* Pick k times from set of n objects, order doesn't matter and with replacement. $\binom{n+k-1}{k}$ ways

~~Proof~~: extreme case: $k=0 \Rightarrow \binom{n-1}{0} = 1$

$k=1 \Rightarrow \binom{n}{1} = n$

simplest nontrivial: $n=2 \Rightarrow \binom{k+1}{k} = k+1$

example 

1 2

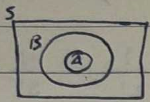
→ dot gives ~~h~~ us a hint → eqiv problem: how many ways are there to put k indistinguishable particles into n distinguishable boxes?

$n=4$
$K=6$

→ ···||··|· → k •'s, n-1 |'s ⇒ $\binom{n+k-1}{k} = \binom{n+k-1}{n-1}$

Memo No.
Date / /

(2) If $A \subseteq B$, then $P(A) \leq P(B)$, if A happens B also happens not vice versa




$B = A \cup (B \cap A^c)$, disjoint

$P(B) = P(A) + P(B \cap A^c) \Rightarrow P(B) \geq P(A)$

not disjoint

□ (3) exclusion inclusion $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

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

$\rightarrow P(A_1 \cup A_2 \cup \dots \cup A_n) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) - \dots + (-1)^{n+1} P(A_1 \cap A_2 \dots \cap A_n)$

* de Montmort's Problem (matching problem)

→ position of shuffled deck of cards equal to the card

n cards labeled 1, 2, ..., n

□ → let $A_j$ be the event "j-th card matches"

→ we want $P(A_1 \cup A_2 \cup \dots \cup A_n)$

$P(A_j) = \frac{1}{n}$, since all positions equally likely

$P(A_1 \cap A_2) = \frac{(n-2)!}{n!} = \frac{1}{(n-1)(n)}$ we can do this because of symmetry

$\vdots$

$P(A_1 \cap \dots \cap A_k) = \frac{(n-k)!}{n!}$ $\frac{n(n-1)}{2}$

$\rightarrow P(A_1 \cup A_2 \cup \dots \cup A_n) = n \times \frac{1}{n} - \binom{n}{2} \frac{1}{n(n-1)} + \frac{n(n-1)(n-2)}{3!} \times \frac{1}{n(n-1)(n-2)} - \dots$

$= 1 - \frac{1}{2} + \frac{1}{3!} - \dots + (-1)^{n+1} \frac{1}{n!}$

Memo No.
Date / /
53

□ Birthday Problem : a group of $\underbrace{\text{people}}_{\text{k people}}$ in a party, is there a pair that they have the same birthday

→ How many people do you need to have a 50-50 chance?

assumptions: 365 days equally likely, assume independence of births

R code ← people come from a uniform dist

1) $k > 365$ : $P = 1$ → Pigeon hole principle

2) $k \leq 365$ : $P(\text{no match}) = \frac{365 \cdot 364 \cdot 363 \dots (365-k+1)}{365^k}$

→ $P(\text{match}) = 1 - P(\text{no match})$

$\Rightarrow P(\text{match}) \begin{cases} 50.7\%, \text{ if } k = 23 \\ 97\%, \text{ if } k = 50 \\ 99.999\%, \text{ if } k = 100 \end{cases}$

an intuition : $\binom{23}{2} = \frac{23 \cdot 22}{2} = 253$

□ عدد ۲۳ به نظر عدد کوچکی می‌آید، اما می‌بینیم که تعداد زیادی pair را تشکیل می‌دهد

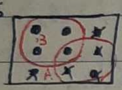
* harder : two people with an identical birthday or with birthdays one day apart → 14 for 50%.

* Properties of probability

(1) $P(A^c) = 1 - P(A)$ : $P(S) = 1 = P(A \cup A^c) = P(A) + P(A^c)$

Conditional Probability : How should you update prob / beliefs / uncertainty based on new evidence?

Def: $P(A|B) = \frac{P(A \cap B)}{P(B)}$ , if $P(B) > 0$

intuition 1: 

→ get rid of anything in $B^c$

→ our universe restricted to B

→ We have a new universe

□ → why is there division by $P(B)$ : we say get rid of pebbles not in B and then just do what we were doing before

→ the problem: things in B don't have total mass 1.

→ so renormalize

□ intuition 2 : frequentist world : repeat experiment many times

→ Circle reps where B occured, Among those, what fractions of time did A also occur?

* thm 1 $P(A \cap B) = P(B)\, P(A|B) = P(A)\, P(B|A)$

□ thm 2 $P(A_1, A_2, \ldots, A_n) = P(A_1)\, P(A_2, A_3, \ldots, A_n | A_1)$

we can think of different combination of it $= \;'' \; P(A_2|A_1)\, P(A_3 A_4 \ldots A_n | A_1 A_2)$

$= \ldots \ldots P(A_n | A_1, \ldots A_n)$

thm 3 $P(A|B) = \frac{P(B|A)\, P(A)}{P(B)}$

→ the implications are very deep → Bayesian Statistics



$P(\text{no match}) = P\left(\bigcap_{j=1}^{n} A_j^c\right)$ related to Poisson Approx.

$= 1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \cdots + (-1)^{n+1} \frac{1}{n!} \approx \frac{1}{e}$

□ ⇒ taylor series of $e^{-x}$ around zero and $x = 1$

* این مسئله جالب است. به خاطر inclusion exclusion | $\frac{1}{e}$ : مسئله انگار گسسته بود

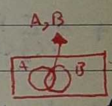
اما به یک جواب پیوسته رسیدیم. $\frac{1}{e}$ عدد مهمی است / symmetry در حل

مسئله کمک کرد

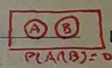
S4

* Independence : Def : Events A, B are independant if

$P(A \cap B) = P(A)\, P(B)$

* Note : Completely different from disjointness 


independence : A occured, tells nothing us about occur of B

disjoint : A occured, B can't possibly occur 


Newton - Pepys Problem : Have fair dice, which is more likely

1) At least one 6 with 6 dice ← truth

2) " " two 6's " 12 "

3) " " three 6's " 18 "

1): $P(A) = 1 - \left(\frac{5}{6}\right)^6 \approx 0.665$

2): $P(B) = 1 - \left(\frac{5}{6}\right)^{12} - 12\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^{11} \approx 0.619$

3): $P(C) = 1 - \underbrace{\sum_{k=0}^{2} \binom{18}{k} \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{18-k}}_{\text{Binomial}} \approx 0.597$

☐ S6 Monty Hall Problem → Sample Space → Not equally likely
can't use $\frac{n(A)}{n(S)}$
Simpson's Paradox
↳ An event more probable than another without conditioning & less with conditioning

Memo No.
Date

S7

Gambler's ruin: Two gamblers gamble on 1\$, they keep do that untill one of them became bankrupt

→ Two gamblers, sequence of rounds, bet 1\$

, $p = P(A \text{ wins a certain round})$, $q = 1-p$

→ Find ~~...~~ probability A wins entire game (so B is "ruined")

assuming A starts with \$$i$, B starts with \$$N-i$

Random Walk: 0 $i-1$ $i$ $i+1$ N

P: prob of going right, absorbing states at 0, and N

→

☐ Bayes rule in term of odds

Memo No.
Date / /

S5

* How to solve a problem 1) try simple and extreme cases
2) break problem into simpler pieces

Let $A_1, A_2, \ldots, A_n$ be partition of S

$P(B) = P(B, A_1) + P(B, A_2) + \ldots + P(B, A_n)$

$= P(B|A_1)P(A_1) + \ldots + P(B|A_n)P(A_n)$

Wishful thinking

law of total probability

☐ * fun problem of conditional probability

get random 2-card hand from deck

1) Find P( both aces | have ace)

2) Find P( " " | " " of spades)

1) $= \frac{P(\text{both aces, have ace})}{P(\text{have ace})} = \frac{\binom{4}{2} / \binom{52}{2}}{1 - \frac{\binom{48}{2}}{\binom{52}{2}}} = \frac{1}{33}$

→ We could use naive def. of Prob. because outcomes are equally likely

2) $\boxed{AS}$ $\boxed{?}$ → $P = \frac{3}{51} = \frac{1}{17}$

☐ hint: Second becomes double likely, why?

at second, ~~it is deterministically obvious~~ we know that a specific one exists, but at first we just know that at least there is an ace.

☐ * 2 conditionally independent events are not necessarily unconditionally indep. Chess opponent

☐ * 2 ~~indep events~~ are not necessarily conditionally independent Fire Alarm PopCorn

Memo No.
Date / /

its distribution is given by: $P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$

→ it is PMF: Prob. Mass Func.

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

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

S8

* Binomial Distribution $Bin(n,p)$ $X \sim Bin(n,p)$

(1) Story → distribution of number of success

→ X is # of successes in n independent $Bern(p)$ trials (p: prob. of success)

(2) sum of indicator r.v.s : $X = X_1 + X_2 + \cdots + X_n$,

$$X_j = \begin{cases} 1, & \text{if } j\text{th trial success} \\ 0 & \text{o.w} \end{cases}$$

indicator

breaking sth complicated into ~~simple~~ sum of very simple things

* $X_1, \dots, X_n$ $\underbrace{\text{i.i.d}}_{\text{used a lot}}$ $Bern(p)$

independent, identically distributed

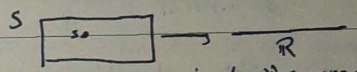
* Don't confuse RV with Distribution

→ RV is a function, a number $\in \mathbb{R}$

→ "the distribution is saying what are the prob. that X will behave in different ways" → So you can have lots of RV's with same distribution

Memo No.
Date / /

* Random Variable: A function from sample space S to $\mathbb{R}$.




→ Think of as a numerical "summary" of an aspect of the experiment

* Defn (Bernoulli) A r.v X is said to have Bernoulli (p) distribution if X has only 2 possible values, 0 and 1 / $P(X=1)=p$, $P(X=0)=1-p$

event, what event? it is intuitively the event that X actually takes the value 1

→ $\{s : X(s)=1\}$

* Binomial $(n,p)$ The distribution of # successes in n indep. $Bern(p)$ trials.

We didn't yet write formal defn. of independence for R.V.s. But the intuition is same as "events"

Memo No.
Date / /

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

(1) immediate from story

(2) $X = x_1 + x_2 + \cdots + x_n$, $Y = y_1 + y_2 + \cdots + y_m$

$\underbrace{x_1 + x_2 + \cdots + x_n}_{\text{i.i.d } Bern(p)}$, $\underbrace{y_1 + y_2 + \cdots + y_m}_{\text{i.i.d } Bern(p)}$

$$X + Y = \sum_{i=1}^{n} x_i + \sum_{j=1}^{m} y_j \sim Bin(n+m, p)$$

□ (3) $P(X+Y=k)$ → "convolution" → we will see that in future

□ → * wishful thinking: we wish that we knew value of $X$ (or $Y$) → $\underbrace{\text{LTP}}_{\text{law of Total probability}}$

$$P(X+Y=k) = \sum_{j=0}^{k} P(X+Y=k \mid X=j)\, P(X=j)$$

~~X and Y are indep. So: $\sum_{j=0}^{k}$~~

$$= \sum_{j=0}^{n} \underbrace{P(Y=k-j \mid X=j)}\binom{n}{j} p^j q^{n-j}$$

X and Y are indep. So: $\sum P(Y=k-j)\binom{n}{j} p^j q^{n-j}$

$$= \sum \binom{m}{k-j} p^{k-j} q^{m-k+j} \binom{n}{j} p^j q^{n-j}$$

$$= p^k q^{m+n-k} \underbrace{\sum_{j=0}^{k} \binom{m}{k-j}\binom{n}{j}}_{\text{Vandermonde}} = \binom{n+m}{k} p^k q^{n+m-k}$$

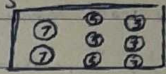
$= Bin(n+m, p)$ ☒

Memo No.
Date / /

(3) PMF: $P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$

* More about PMF

S



$X = 7$ → □ what does that mean → it is an event

(an even is a subset of sample space)

* CDF □ $X \leq x$ is an event → so we can talk about its probability → $F(x) = P(X \leq x)$, then $F$ is the CDF of $X$ (cummulative distribution func.)

→ CDF is a way to describe the distribution because it's telling us the probabilities of different possible values for X.

* PMF (for discrete r.v.s)

□ Discrete: in general, it should be sth that you can list, maybe a finite list, or an infinite list

$a_1, a_2, \ldots, a_n$ or $a_1, a_2, \ldots$

→ PMF: $p_j = P(X = a_j)$ for all $j$

□ $p_j \geq 0$, $\sum_j p_j = 1$

↳ check for binomial: $\sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k}$

□ $\xrightarrow[\text{theory}]{\text{binomial}}$ $(p+q)^n = 1^n = 1$

Memo No.
Date / /

→ in hypergeometric, we sample without replacement

→ if we did with replacement, it was binomial

important. Suppose that number of marbles is like a billion, and suppose that size of our sample is very small compared to a billion, let's say 10, it's extremely unlikely that 2 apples we pick from a billion are the same. So sampling with or without replacement should behave similarly. Then hypergeometric should be approximately binomial

prove that it is dist.: $\sum_{k=0}^{w} \frac{\binom{w}{k}\binom{b}{n-k}}{\binom{w+b}{n}}$

$= \frac{1}{\binom{w+b}{n}} \underbrace{\sum_{k=0}^{w} \binom{w}{k}\binom{b}{n-k}}_{\text{Vandermonde}} = \frac{1}{\binom{w+b}{n}} \times \binom{w+b}{n} = 1$

Memo No.
Date / /

* Common mistakes that taken as binomial

Ex. 5 card-hand, find distribution of # aces

→ for find distribution, we could find CDF, but it's going to be easier with PMF (it's discrete problem)

→ Let $X = (\#\text{ aces})$

↳ a function from $S$ to $\mathbb{R}$, it is easier to write it this way / → Find $P(X=k)$, this is 0 except if $k \in \{0, 1, 2, 3, 4\}$

→ The distribution is not binomial, because we can think each card as a trial, but trials are not indep.

for example, if the first two one are aces, then the 3rd is less likely to be ace.

→ PMF: $P(X=k) = \frac{\binom{4}{k}\binom{48}{5-k}}{\binom{52}{5}}$ → it is like elk problem or elf same

→ we have tagged and untagged elfs and we wanted $k$ tagged ones

→ $b$ black, $w$ white marbles. Pick simple random sample of size $n$. Find dist. of $(\#\text{ white marbles}) = X$

→ like elf → $P(X=k) = \frac{\binom{w}{k}\binom{b}{n-k}}{\binom{b+w}{n}}$, $0 \le k \le w$, $0 \le n-k \le b$

→ Hypergeometric → that's not binomial

Why $P(X=x)=0$ in continuous form?

→ Discrete case: $P(X=x, Y=y) = P(X=x) \cdot P(Y=y)$

→ knowing $X$ doesn't tell us anything about $Y$.

→ won't work for continuous case → $0=0$

* Averages (of R.V) (Means, Expected Value)

→ we do the exp. and we get R.V value, but beforehand we want to make some prediction

→ on average, what's going to happen

→ the average is just going to tell a one number summary of the center of the dist. in some sense, that's important but you have a complicated dist. and that would not be enough → Variance, SD → still need to calculate that.

→ $1, 2, 3, \ldots, 6 \rightarrow \frac{1+2+\ldots+6}{6} = 3.5$

$1, 2, 3, \ldots, n \rightarrow \frac{1}{n} \sum j$ → unweighted avg.

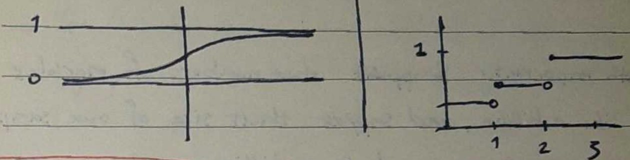
$1,1,1,1,1, 3, 3, 5 \rightarrow \begin{cases} 1. \text{ add, divide} \\ 2. \ \frac{5}{8} \cdot 1 + \frac{2}{8} \cdot 3 + \frac{1}{8} \cdot 5 \end{cases}$

weighted avg, → Because there are different weights

* Average of a discrete r.v. $X$

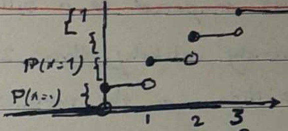
$E(X) = \sum_x \underbrace{x}_{\text{value}} \underbrace{P(X=x)}_{\text{PMF}}$, summed over $x$ with $P(X=x) > 0$

CDF: $P(X \leq x)$




59

* CDF, discrete




$F(x) = P(X \leq x)$, as a func of real $x$.

CDF Tells us what we want to know about dist.

→ Find $P(1 < X \leq 3)$ using $F$

$P(X \leq 1) + P(1 < X \leq 3) = P(X \leq 3)$

$\Rightarrow P(1 < X \leq 3) = F(3) - F(1)$

$\Rightarrow P(a < X \leq b) = F(b) - F(a)$

* Properties of CDF (1) increasing

(2) right continuous "از راست پیوسته"

(3) $F(x) \rightarrow 0$ as $x \rightarrow -\infty$

$F(x) \rightarrow 1$ as $x \rightarrow \infty$

This is "if and only if"

□ * indep. of r.v.s  $X, Y$ are indep. r.v.s if

$\underbrace{P(X \leq x, Y \leq y)}_{\text{joint}} = P(X \leq x) P(Y \leq y)$ for all $x, y$.

Mo Tu We Th Fr Sa Su

Memo No.
Date / /

* Redo Binomial : $X = \overbrace{X_1 + X_2 + \cdots + X_n}^{\text{i.i.d}}$

$X_j \sim Bern(p) \rightarrow$ by linearity $\rightarrow E(X) = E(X_1 + \cdots + X_n)$

$= E[X_1] + \cdots + E(X_n) = np$ / even if they were depen. that was true. (by linearity)

* Ex. 5 card hand, $X = (\#\text{ aces}) \rightarrow$ hypergeometric dist.

$E(X) = ?$

Let $X_j$ be indicator of j-th card being an ace, $1 \le j \le 5$

→ We can assume an order for cards.

$E(X) = E(X_1 + \cdots + X_5) \overset{\text{linearity}}{=} E(X_1) + \cdots + E[X_5]$

$\underset{\text{symmetry}}{=} 5E(X_1) \underset{\text{fundamental bridge}}{=} 5P(\text{1st card ace}) = 5 \cdot \frac{4}{52} = \frac{5}{13}$

→ even though $x_j$'s were dependent.

□ → tools we used : linearity, indicators, symmetry, fundamental bridge

* Next Famous Dist. Geometric $\overset{\text{Geom}(p)}{}$ : indep. Bern(p) trials, count # failures before 1st success. Let $X \sim Geom(p)$, $q = 1-p$

PMF : $P(X=k) = q^k p$, $k \in \{0, 1, 2, \cdots\}$

valid since $\sum_{k=0}^{\infty} q^k p = p \underbrace{\sum_{k=0}^{\infty} q^k = \frac{p}{1-q}}_{\text{geometric series}} = \frac{p}{p} = 1$

---

Mo Tu We Th Fr Sa Su

Memo No.
Date / /

* $X \sim Bern(p)$ : $E(X) = 1P(X=1) + 0P(X=0) = p$

$\rightarrow X = \begin{cases} 1 & \text{if A occur} \\ 0 & \text{o.w} \end{cases} \rightarrow$ indicator r.v.

□ Then $E(X) = P(A) \rightarrow$ fundamental bridge

پل میان احتمال و امید ریاضی

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

$E(X) = \sum_{k=0}^{n} k \binom{n}{k} p^k q^{n-k} \xrightarrow[\text{story proof □}]{\text{by}} E(X) = \sum_{k=0}^{n} n \binom{n-1}{k-1} p^k q^{n-k}$

□ choose $\underbrace{\binom{n}{k}}_{\text{committee}}$ then president $\equiv$ choose pre. then resting committe

→ we did this because k was annoying us

$= n \sum_{k=0}^{n} \binom{n-1}{k-1} p^k q^{n-k} = np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} q^{n-k}$

$\overset{j=k-1}{=} np \underbrace{\sum_{j=0}^{n-1} \binom{n-1}{j} p^j q^{n-1-j}}_{\substack{\text{Binomial Theorem} \\ 1}} = np$

* Linearity

$\rightarrow E(X+Y) = E(X) + E(Y)$, □ even if X, Y are dependant

$\rightarrow E(cX) = cE(X)$, if c is a constant

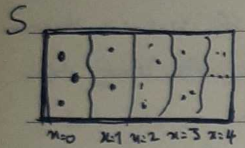
Mo Tu We Th Fr Sa Su

Memo No.

Date / /

* Another POV to prove linearity

S




$$E(X) = \underbrace{\sum x P(X=x)}_{\text{weighted Avg}} = \underbrace{\sum_s X(s) P(\{s\})}_{\text{unweighted Avg}}$$

we suppose each of pebbles one by one here

$\Rightarrow E(T) = \sum_s T(s) \; P(\{s\}) = \sum_s (X+Y)(s) \; P(\{s\})$

$= \sum (X(s) + Y(s)) P(\{s\}) = \sum_s X(s) P(\{s\}) + \sum_s Y(s) P(\{s\})$

$= E(X) + E(Y)$

* Similarly, $E(cX) = cE(X)$ if $c$ is const.

→ an interesting intuition: Extreme case of dependence: $X=Y$

Then $E(X+Y) = E(2X) = 2E(X) = E(X) + E(Y)$

* What makes a distribution important?

The story behind the distribution

* Negative Binomial → Generalization of Geometric

(parameters: $r, p$)

Story: independent Bern($p$) trials, # failures before the $r$-th success.

Mo Tu We Th Fr Sa Su

Memo No.

Date / /

* $X \sim Geom(p)$; $E(X) = \sum_{k=0}^{\infty} k p q^k = p \sum_{k=0}^{\infty} k q^k$

→ without, it is geometric series, with $k$ → it's annoying

→ $\sum_{k=0}^{\infty} q^k = \frac{1}{1-q}$ → now we want to appear $k$

→ derivative $\longrightarrow \sum_{k=1}^{\infty} k q^{k-1} = \frac{1}{(1-q)^2} \xrightarrow{\times q} \sum_{k=1}^{\infty} k q^k = \frac{q}{p^2}$

$\Rightarrow \frac{pq}{p^2} = \boxed{\frac{q}{p}}$

→ story proof method: Let $c = E(X)$

idea: first step Analysis (like Gambler's ruin)

$c = \underbrace{0 \times p}_{\text{the first coin is success}} + (\underbrace{1}_{\text{one failure}} + \underbrace{c}_{}) \underbrace{q}_{\text{failure the first time}}$

{ the same problem again; coin is memoryless, problem is restarted

$\Rightarrow c = q + cq \Rightarrow c = \frac{q}{1-q} = \frac{q}{p}$

S10

* Proof of linearity: Let $T = X+Y$, show

$E[T] = E[X] + E[Y]$

$\sum t P(T=t) \overset{?}{=} \sum x P(X=x) + \sum y P(Y=y)$

→: cannot be done

→: $P(T=t) = \sum P(T=t \mid X=x) P(X=x)$

→ if they were indep. that will work but they are not. So it doesn't work

□ * Putnum problem: Random Permutation of $1, 2, \ldots, n$, where $n \geq 2$. Find Expected # of local maxima

ex. 5 2 1 4 7 5 6

⇒ □ Let $I_j$, be indicator r.v. of position $j$ having a local maxima, $1 \leq j \leq n$

$$E(I_1 + \cdots + I_n) = \underbrace{E(I_1) + E(I_n)}\ + \underbrace{E(I_2) + \cdots + E(I_{n-1})}$$

local maxima

$2 \leq j \leq n-1$ ⇒ $\_ \uparrow \_$ □ $P(\text{middle is local maxima}) = \frac{1}{3}$ ✓

□ → wrong answer: $\underbrace{\frac{1}{2}}_{\text{greater than } i-1} \times \underbrace{\frac{1}{2}}_{\text{greater than } i+1} = \frac{1}{4}$ X → they are not indep.

$j = 1, n$
⇒ $\underset{1}{\cdot} - / - \underset{n}{\cdot} \rightarrow P = \frac{1}{2} \Longrightarrow E = \frac{2}{2} + \frac{n-2}{3} = \frac{n+1}{3}$

⇒ Let's check extreme cases, $n = 2 \rightarrow E = 1$ → consistent with intuition

□ → indicators, linearity, symmetry

■ * St. Petersburg Paradox: Get $\$2^X$, where $X$ is # flips of fair coin untill first H, including the success

$Y = 2^X$, find $EY$

$$E(Y) = E(2^X) = \sum_{k=1}^{\infty} 2^k \cdot \frac{1}{2^k} = \sum_{k=1}^{\infty} 1 = 1 + 1 + \cdots = \boxed{\infty}\ !$$

bound at $\underbrace{\$2^{40}}_{10^{12}}$, then $\sum_{k=1}^{40} 2^k \frac{1}{2^k} = 40$

$10^{12}$ ← تریلیون



→ PMF: $P(X = n) = \binom{n+r-1}{r-1} p^r (1-p)^n$

→ negative binomial, because if binomial theorem is powered to negative, we see sth like this

$\rightarrow E(X) = E(X_1 + \cdots + X_r)$

□ $X_j$ is # failures between $(j-1)$st and $j$th success; → $X_j \sim Geom(p)$ → $X_j$'s are indep.

→ but in this case we don't care

$$\Rightarrow E(X) = E(X_1) + \cdots + E(X_r) = r\frac{q}{p}$$

□ * یک برای توزیع هندسی، یک سری قرارداد خود success را هم بشمار می‌آورند و یک سری دیگر نه.

first Success

$X \sim FS(p)$, time until 1st success, counting the success.

Let $Y = X - 1$, Then $Y \sim Geom(p)$

$$EX = EY + 1 = \frac{q}{p} + 1 = \frac{1}{p}$$

→ intuition: $p = \frac{1}{10}$ → 10 trials to reach success → it is intuitive according to $p = \frac{1}{10}$

$X \sim Geom(p)$: $P(X = k) = q^k p$

$$P(X \leq x) = P(X = 0 \cup X = 1 \cup \cdots \cup X = x)$$
$$= \sum_{i=0}^{x} q^i p = p \sum_{i=0}^{x} q^i = p \times \frac{1(1-q^x)}{1-q} = 1 - (1-p)^x$$

→ used for applications, where counting # of "successes" (we can define it in different ways)
where there are a ☐ large number of things
each of which could lead to success or failure
☐ but the prob. of success for each one is small

* Poisson Paradigm (Approx.)

1. a lot of events $A_1, \dots, A_n$ with $P(A_i) = p$
2. number of trials $n$ is very large
3. $p$ is very small

events could be indep.
They could even be weakly dependent ($A_1$ occuring has some level of effect on the likelihood of $A_2$)

$A = A_1 + A_2 + \dots + A_n$

$E[A] = E[A_1] + \dots + E[A_n] = \sum_{j=1}^{n} p_j = \lambda$

$A \sim Pois(\lambda)$

* $X \sim Bin(n,p)$, $n \to \infty$, $p \to 0$, $\lambda = np$ constant

$$P(X=k) = \binom{n}{k} p^k q^{n-k} = \frac{n(n-1)\dots(n-k+1)}{(k!)} \left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^n \left(1-\frac{\lambda}{n}\right)^{-k}$$

$$\xrightarrow{n \to \infty} \frac{\lambda^k e^{-\lambda}}{k!}$$

in Binomial, we have to mak assumption about P being Identical, bet we don't have to do that in Poisson

---

* ☐ $\infty = E[2^X] \neq 2^{E[X]} = 4$

S11

* Sympathetic Magic: Confusing R.V with its dist.
→ Adding R.V.s is not the same adding PMFs.
$P(X=x) + P(Y=y)$ → no reason to be $< 1$
→ a function of $x$ and $y$, but $X+Y$ → function of r
→ R.V ↔ random house / distribution ↔ blueprint

→ from a blueprint, you can ~~make~~ build many houses.
→ i.i.d

* Poisson distribution

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

$k$ → non-negative value

$\lambda$ is the "rate" parameter, $\lambda > 0$

→ check valid PMF: $\sum_{k=0}^{\infty} \frac{e^{-\lambda}\lambda^k}{k!} = e^{-\lambda} \underbrace{\sum_{k=0}^{\infty} \frac{\lambda^k}{k!}}_{\text{☐ taylor series of } e^{\lambda}} = e^{-\lambda}e^{\lambda} = 1$

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

$= \lambda e^{-\lambda} \sum_{1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} e^{\lambda} = \boxed{\lambda}$

X is allowed to take on any real number value between zero and one, There are uncountably real numbers between 0 and 1, and any specific number has prob. 0 → So we define PDF

## Lecture 12: * The Big Picture!

| discrete $X$ | Continuous $X$ |
| --- | --- |
| | $[P(X=x)=0]$ |
| PMF $P(X=x)$ | PDF $f_X(x) = F_X'(x)$ |
| CDF $F_X(x)=P(X\leq x)$ | CDF $F_X(x)=P(X\leq x)$ |

**PDF: Probability Density Function** (Keyword): The most common way to specify a continuous dist.

→ it's not probability, it's probability density

**Defn** R.V. $X$ has PDF $f(x)$ if

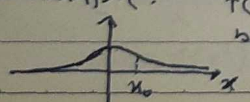
$$P(a\leq X\leq b) = \int_a^b f(x)\,dx \quad \text{for all } a,b$$

⇒ so $f(x)$ is not prob., it's what you integrate to get prob.

→ $\left[a=b \Rightarrow \int_a^b f(x)\,dx = 0\right]$.

→ To be valid, by analogy to Discrete, should be non-negative, and rather to summing to one, integratig to one, $f(x)\geq 0$, $\int_{-\infty}^{\infty} f(x) = 1$

⇒ for example: $f(x_0)$ → We can't say it's a Prob because it can be even greater than one




## Triple Matches · $\binom{n}{3}$ triplets

- Indicator R.V. $I_{i,j,k}$, $i<j<k$
- $E[\#\text{ triple matches}] = \binom{n}{3}\frac{1}{365^2}$

$= I_{123} + I_{124} + \cdots$

$= \cancel{\frac{365\times 1\times 1}{365^3}}\ \frac{1}{365^2} + \cdots = \binom{n}{3}\times\frac{1}{365^2}$

→ Why Poisson? 1. # of trials is very large $\left(\binom{n}{3}\right)$

2. $p$ is small $I_{ijk}$ 3. events are weakley dependend

→ $\text{Pois}(\lambda)$, $\lambda = E[\#\text{ triplete matches}] = \binom{n}{3}\frac{1}{365^2}$

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

Memo No.
Date / /

$\Rightarrow$ Variance $Var(X)$ → How far is $X$ from its mean

$\Rightarrow$ $Var(x) = E[(X - EX)]$ → but it is also zero by linearity

□ → now we want to fix it ⇒ absolute value is annoying → not differentiable

□ → $E[(X-EX)^2]$ → one annoying thing with this is
it's that it's not the same unit, if it was miles → now $miles^2$

□ ⇒ Standard deviation : $SD(X) = \sqrt{Var(X)}$

→ SD is something interpretable, but Variance is nicer to work with in math

⇒ Another way to express Var: $E[(X - EX)^2]$

$= E[X^2 - \underbrace{2X}_{\text{Constant}} \underbrace{E[X]}_{\text{Constant}} + \underbrace{(EX)^2}_{\text{Constant}}] = E[X^2] - 2E[X]E[X] + (EX)^2$

$= E[X^2] - (EX)^2$, by $E[(X - EX)^2]$, we can interpret that $E[X^2]$ is always greater than $(EX)^2$, except

□ when $X$ is constant, where $Var(X) = 0$, So $Var(X) \geq 0$

□ * The complexity of it, it's that how we can compute $E[X^2]$ ? → Later!

* Continue Big Picture

| discrete $X$ | continuous $X$ |
|---|---|
| $Var(X) = E[X^2] - (E[X])^2$ | $Var(X) = E[X^2] - (E[X])^2$ |

Memo No.
Date / /

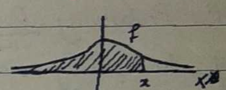
by multiplying this epsilon we're kinda converting it back to probability scale

□□ → What you can say : $f(x_0)\varepsilon \approx P(X \in (x_0 - \frac{\varepsilon}{2}, x_0 + \frac{\varepsilon}{2}))$
for $\varepsilon > 0$ very small ⇒ So you can think of it as like probability per unit of length.

* If $X$ has PDF $f$, the CDF is

□ (PDF is the thing you integrate to get prob.)

$F(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)\,dt$




if $X$ has CDF $F$ (and $X$ is continuous), then
$f_X(x) = F'(x)$ by FTC

⇒ $P(a < X < b) = \int_a^b f(x)\,dx = F(b) - F(a)$

* Continue The big picture

| Discrete $X$ | Continuous $X$ |
|---|---|
| $E(X) = \sum_x x P(X = x)$ | $E(X) = \int_{-\infty}^{\infty} x f_X(x)\,dx$ |

□ * Expectation is just a giving a one number summary of the average. ⇒ But it's not telling about the spread of distribution

Mo Tu We Th Fr Sa Su

Memo No. ______
Date / /

→ Variance, $Y = x^2$, a function of R.V is also a R.V

□ $E[x^2] = E[Y]$, need PDF of $Y$? How? Later!

$= \int_{-\infty}^{\infty} x^2 f_X(x)\,dx$, □ Law of the unconscious Statistician (LOTUS)

$$\boxed{E(g(x)) = \int_{-\infty}^{\infty} g(x) f_X(x)\,dx}$$

proof later

for discrete: $Eg(x) = \sum_x g(x) P(X=x)$

Let $U \sim \text{Uniform}(0,1)$

$E[U] = \frac{1}{2}$, $f(u)$

$E[U^2] = \int_0^1 u^2 \underbrace{\frac{1}{1-0}}_{f(u)} du = \frac{1}{3}$

$\Rightarrow Var(U) = \frac{1}{3} - \frac{1}{24} = \frac{1}{12}$

□□ * Uniform is Universal, Let $U \sim \text{Unif}(0,1)$, $F$ be a CDF (assume $F$ is strictly increasing continuous), Then Let $X = F^{-1}(U)$, Then $\underbrace{X \sim F}$

$X$ has CDF $F$, $X$ follows this distribution

Proof: $P(X \leq x) = P(F^{-1}(U) \leq x)$

$F$ is strictly increasing and continuous $\Longrightarrow$ $P(U \leq F(x)) = F(x)$

0 F(x) 1

⇒ If you give me one uniform R.V, and you're interested in some other distribution, there is a way to convert it and simulate that.

Mo Tu We Th Fr Sa Su

Memo No. ______
Date / /

* Uniform(a, b) a midpoint x b → we wanna take a "completely random" number in this interval $[a,b]$

→ but it's kinda vague, because $P(\overset{\text{point}}{x}) = 0$, it's not ~~for a better~~ interesting to say all the probabilities are the same, because all of them are Zero

→ The intuition now is, The probability of this half is equal to other half ⇒ Uniform means Probability □ $\propto$ length

→ $f(x) = \begin{cases} c, & \text{if } a \leq x \leq b \\ 0, & \text{o.w} \end{cases}$ should be a constant, o.w it is not uniform

□ $\Rightarrow 1 = \int_a^b c\,dx \Rightarrow c = \frac{1}{b-a}$

□ $F(x) = \int_{-\infty}^{x} f(t)\,dt \overset{x>a}{=} \int_a^x f(t)\,dt = \begin{cases} 0 & \text{if } x < a \\ \frac{x-a}{b-a}, & \text{if } a \leq x \leq b \\ 1 & \text{if } x \geq b \end{cases}$

it is linear Function of $x$, which makes sense, because, as more $x$, we expect more Cumulative

□ $\Rightarrow E[X] = \int_a^b \frac{x}{b-a}\,dx = \left.\frac{x^2}{2(b-a)}\right|_a^b = \frac{b+a}{2}$ midpoint

⇒ It is expectable because it is uniform

Memo No.
Date / /

## * Symmetry of The Standard Uniform Distribution.

□⇒ if $U \sim Unif(0,1)$, then also $1-U \sim Unif(0,1)$

$\underset{0 \quad u \quad\quad 1}{\vdash\!\!-\!\!+\!\!-\!\!-\!\!\dashv}$ , we just need one random point no ~~matter~~ difference that we measure from right or left.

→ $a+bU$ is uniform on some interval
non-linear → not Uniform

## * Indep. of R.V.s $X_1, \dots, X_n$

Defn $X_1, \dots, X_n$ are indep if

□ $\underbrace{P(X_1 \leq x_1, \dots, X_n \leq x_n)}_{\text{joint CDF}} = P(X_1 \leq x_1) \dots P(X_n \leq x_n)$ For all $x_1, \dots, x_n$

□⇒ In events, for example for 3 events, in addition to having all 3 independent, we needed to ~~have~~ write pairwise independence. But for R.V.s, it's that.

⇒ Discrete Case: $\underbrace{P(X_1 = x_1, \dots, X_n = x_n)}_{\text{joint PMF}} = P(X_1 = x_1) \dots P(X_n = x_n)$

in discrete case they are equivalent

□□ * Pairwise independence doesn't imply Independence

Ex. $X_1, X_2 \sim Bern(\frac{1}{2})$, $X_3 = \begin{cases} 1 & , X_1 = X_2 \\ 0 & , o.w \end{cases}$

Memo No.
Date / /

## Lecture 13: * Universality of Uniform

Also ; Moreover if $X \sim F$, then $\underbrace{F(X)}$ $\sim Unif(0,1)$
for any R.V. X

□ { X is a R.V., F(X) is a function of R.V., which is a R.V

proof: $P(F(X) \leq x) = P(X \leq F^{-1}(x))$

$= F(F^{-1}(x)) = x \Rightarrow$ $\underset{0 \quad x \quad 1}{\vdash\!\!-\!\!+\!\!-\!\!\dashv}$ $\Rightarrow F(X) \sim Unif(0,1)$

□ * Notational Confuse: $F(x) = P(X \leq x)$

~~$F(X) = \underbrace{P(X \leq X)}_{\text{An always True event}} = 1$~~ ✗

⇒ $F(X)$ seen in example: $F(x) = 1 - e^{-x}$, $x > 0$ (exp. dis)

$\Rightarrow F(X) = 1 - e^{-X}$

*Ex. Let $F(x) = 1 - e^{-x}$, $x > 0$ (Exp(1)), $U \sim Unif(0,1)$,

Simulate $X \sim F$. ⇒ $F^{-1}(u) = -\ln(1-u)$

⇒ $F^{-1}(U) = -\ln(1-U) \sim F$

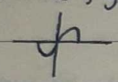
□ ⇒ if we want 10 random draws from this distribution, we just generate ten i.i.d uniform and then compute the function and finally we have 10 i.i.d samples from the target dist.

Memo No.
Date / /

$* \; N(0,1),\; EZ = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z e^{-\frac{z^2}{2}} = 0$ by Symmetry

( if $g(x)$ is an odd func., $g(-x) = -g(x)$

$\int_{-a}^{a} g(x)dx = 0$



$\rightarrow z e^{-\frac{z^2}{2}}$ is odd

$* \; Var(Z) = E[Z^2] - \underbrace{(EZ)^2}_{0}$

by LOTUS $\rightarrow$ $E[Z^2] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \underbrace{z^2 e^{-\frac{z^2}{2}}}_{\text{even function}} dz = \frac{2}{\sqrt{2\pi}} \int_{0}^{\infty} z^2 e^{-\frac{z^2}{2}} dz$

Integration by parts $\Rightarrow$ split the integrand into 2 pieces, one piece that's easy to integrate, and one piece that's easy to differentiate $\Rightarrow \frac{2}{\sqrt{2\pi}} \int_{0}^{\infty} \underbrace{z}_{u} \underbrace{z e^{-\frac{z^2}{2}}}_{dv} dz$

$u = z \rightarrow du = dz \;/\; dv = z e^{-\frac{z^2}{2}} \rightarrow v = -e^{-\frac{z^2}{2}}$

$= \frac{2}{\sqrt{2\pi}} \left( \underbrace{(uv)\Big|_{0}^{\infty}}_{0} + \underbrace{\int_{0}^{\infty} e^{-\frac{z^2}{2}} dz}_{\text{that's just we did}} \right) = 1!$

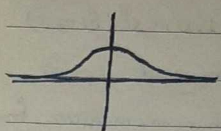
$\frac{1}{2}$ of that $\frac{\sqrt{2\pi}}{2}$

Taylor Series $f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots$

$dx\,dy = |J_d|\, dr\, d\theta = \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{vmatrix} dr\, d\theta$

Memo No.
Date / /

* Normal distribution (Central Limit Theorem)
sum of a lot of i.i.d r.v.s looks Normal



standard Normal: $N(\underset{\text{mean}}{0}, \underset{\text{Variance}}{1})$

$\rightarrow$ has PDF: $f(z) = \underbrace{c}_{\downarrow} \underbrace{e^{-\frac{z^2}{2}}}_{}$

tradition for standard Normal

normalizing constant

not PDF, because it doesn't integrate to 1

$\rightarrow$ decays to zero very fast, symmetric (زوج)

$* \; \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dz$, theorem: Indefinite Integral is impossible to solve!

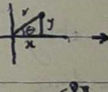
if we write taylor series, infinite sum!

$\Rightarrow$ but we can solve for definite

$\int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dz \xrightarrow{x \text{ itself}} \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dz \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dz$

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

$x^2 + y^2 \rightarrow$ فیثاغورث $\Rightarrow$



$\rightarrow$ تبدیل مختصات دکارتی به قطبی

$u = r^2/2$

$du = r\,dr$

$= \int_{0}^{2\pi} \int_{0}^{\infty} e^{-\frac{r^2}{2}} r\, dr\, d\theta = \int_{0}^{2\pi} \left( \int_{0}^{\infty} e^{-u} du \right) d\theta = 2\pi$

$\Rightarrow \sqrt{2\pi} \Rightarrow c = \frac{1}{\sqrt{2\pi}}$

Memo No. ______ Date / /

$\Rightarrow Var(\mu + \sigma Z) = \sigma^2 Var(Z) = \sigma^2$

→ We started with Standard Normal and made a general normal

→ but in the other way, a general normal which is Normal $\mu, \sigma^2$ ⇒ □ Standardization process

□□ $Z = \dfrac{X - \mu}{\sigma}$

□□* Find PDF of $X \sim N(\mu, \sigma^2)$

CDF : $P(X \leq x)$, A good trick is to standardize.

$P\Big(\underbrace{\frac{X-\mu}{\sigma}}_{\text{Standard Normal}} \leq \frac{x-\mu}{\sigma}\Big) = \Phi\Big(\frac{x-\mu}{\sigma}\Big)$

⇒ PDF → chain rule → $\frac{1}{\sigma} \times \frac{1}{\sqrt{2\pi}} e^{-\left(\frac{x-\mu}{\sigma}\right)^2/2}$

* What about $-X$? : $-X = -\mu + \sigma(-Z) \sim N(-\mu, \sigma^2)$

* Later we'll show: if $X_j \sim N(\mu_j, \sigma_j^2)$ indep

□ $X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

$X_1 - X_2 \sim N(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$

Memo No. ______ Date / /

Capital Phi فای □

* Notation : $\Phi$ is the standard Normal CDF

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}} dt$$

□□ $\Phi(-z) = 1 - \Phi(z)$ by symmetry

## Lecture 14

$Z \sim N(0,1)$, CDF $\Phi$, $\underbrace{E[Z] = 0}_{\text{first moment}}$, $Var(Z) = 1 = \underbrace{E[Z^2]}_{\text{second moment}}$, $\underbrace{E[Z^3] = 0}_{\text{third moment}}$

$\int_{-\infty}^{\infty} z^3 \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = 0$ odd function

* $-Z \sim N(0,1)$ (by symmetry). it's useful.

* Let □ $X = \mu + \sigma Z$, $\mu \in \mathbb{R}$ (mean, location), $\sigma > 0$ (SD, scale)

, Then we say $X \sim N(\mu, \sigma^2)$ — we are rescaling everything

→ $E[X] = E[\mu + \sigma Z] = \mu + \sigma E[Z] = \mu$

→ $Var(X) = E[(X - EX)^2] = EX^2 - (EX)^2$

$Var(X + c) = Var(X)$, immediate from and story

$Var(cX) = c^2 Var(X)$, immediate from and

also variance cannot be negative ($Var(X) \geq 0$,

$Var(X) = 0$ if and only if $P(X = a) = 1$ for some a)

$Var(X + Y) \neq Var(X) + Var(Y)$ in general

□ [equal if X, Y are indep.]

□ → Ex. $Var(\underbrace{X + X}_{\text{Extreme case of dependence}}) = Var(2X) = 4 Var(X)$

Mo Tu We Th Fr Sa Su

Memo No.
Date / /

$$\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{\lambda} \xrightarrow{\frac{d}{d\lambda}} \sum_{k=1}^{\infty} \frac{k\lambda^{k-1}}{k!} = e^{\lambda}$$

taylor series

$$\xrightarrow{\times\lambda} \sum_{k=1}^{\infty} \frac{k\lambda^k}{k!} = \lambda e^{\lambda} \xrightarrow{\frac{d}{d\lambda}} \sum_{k=1}^{\infty} \frac{k^2\lambda^{k-1}}{k!} = \lambda e^{\lambda} + e^{\lambda} = e^{\lambda}(\lambda+1)$$

$$\xrightarrow{\times\lambda} \sum_{k=1}^{\infty} \frac{k^2\lambda^k}{k!} = e^{\lambda}(\lambda+1)\lambda$$

$$\Rightarrow \sum_{k=0}^{\infty} k^2 \frac{\lambda^k e^{-\lambda}}{k!} = e^{-\lambda} e^{\lambda} \lambda(\lambda+1) = \lambda(\lambda+1) = \lambda^2 + \lambda$$

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

☐ → it's kinda strange that $E[X] = Var(X)$ because they are not in the same scale. But Poisson is all about counting and counting does not have scale.

☐ ⇒ Also about Normalization! Another reason that $\frac{X-\mu}{\sigma}$ is nice thing to work with in the normal, imagine that Normal is being a continuous measurement in some unit, for example time (s), then $\frac{(s)-(s)}{(s)}$ is a dimensionless quantity which makes this standardization more directly interpretable, instead makes you worried about measurement in seconds ~~and~~ or years

Mo Tu We Th Fr Sa Su

Memo No.
Date / /

* 68, 95, 99.7% Rule $X \sim N(\mu, \sigma^2)$

$P(|X-\mu| \leq \sigma) \approx 0.68$ The prob that X is within 1 Standard Deviation of its mean is about 68 %

$P(|X-\mu| \leq 2\sigma) \approx 0.95$ / $P(|X-\mu| \leq 3\sigma) \approx 0.997$

☐ → let's say you got a bunch of observations from this dist. independently ⇒ we would expect 95% of them are gonna be within 2 standard deviations of the mean

* LOTUS

| | prob | $P_0$ | $P_1$ | $P_2$ | $P_3$ | |
|---|---|---|---|---|---|---|
| $X$ : | | 0, | 1, | 2, | 3, | ... |
| $X^2$ : | | 0, | 1, | 4, | 9, | ... |

$E[X] = \sum_x x P(X=x)$

$E[X^2] = \sum_x x^2 P(X=x)$

the probability that X takes $3^2$ is still the same that it takes 3

when it's non-negative it's one to one, but LOTUS says also in non-negative form it is still true.

☐☐ Variance of Poisson

$\rightarrow X \sim Pois(\lambda)$ : $E[X^2] = \sum_{k=0}^{\infty} k^2 \frac{\lambda^k e^{-\lambda}}{k!}$

□ * Prove LOTUS for discrete  Show $E[g(x)] = \sum_x g(x) P(X=x)$

(Similar to prove of linearity)

$$\underbrace{\sum_x g(x) P(X=x)}_{\text{grouped}} = \underbrace{\sum_{s \in S} g(X(s)) P(\{s\})}_{\text{ungrouped}}$$

$X(s)$: R.V a function

$$\sum_x \sum_{s : X(s)=x} \overbrace{g(X(s))}^{g(x)} P(\{s\}) = \sum_x g(x) \sum_{s : X(s)=x} P\{s\}$$

$$= \boxed{\sum_x g(x) P(X=x)}$$

Lecture 15 / Midterm Review

□ Coupon collector (toy collector): n toy types, equally likely

Find expected time (i.e., # toys) untill have complete set

수

$T = T_1 + T_2 + \cdots + T_n$

$T_1$ = time untill 1st new toy = 1 constant

$T_2$ = (additional time untill second new toy)

$T_3$ = ( " " " 3rd " " )

⋮

$T_n$

$T_1 = 1$, $\underbrace{T_2 - 1}$ $\sim \text{Geom}\left(\frac{n-1}{n}\right)$, $T_j - 1 \sim \text{Geom}\left(\frac{n-(j-1)}{n}\right)$

because of convention of geometric starts at 0.
But of course we need at least one more new toy



□ $X \sim \text{Bin}(n, p)$, Find $Var(X)$  $X = I_1 + I_2 + \cdots + I_n$, $I_j \overset{iid}{\sim} \text{Bern}(p)$

$X^2 = I_1^2 + \cdots + I_n^2 + 2I_1I_2 + 2I_1I_3 + \cdots 2I_{n-1}I_n$

$E[X^2] = nE[I_1^2] + 2\binom{n}{2} E[I_1 I_2]$

symmetry

□ { Symmetry
indicator R.Vs
linearity

$= \underbrace{np} + n(n-1)p^2 = np + n^2p^2 - np^2$

$E[I_1^2] = E[I_1]$
because indicator *
$1^2 = 1$, $0^2 = 0$

$I_1 I_2$ : indicator of success on both trials 1, 2

$\Rightarrow Var(X) = np + n^2p^2 - np^2 - n^2p^2 = np(1-p)$
$= npq$

* $X \sim \text{Geom}(p)$, similar with poisson
except we use geometric series.

* □ Var(Hypergeometric) → Later!

$\sum k^2 q^k p$ . $\quad \sum q^k = \frac{1}{1-q} \xrightarrow{\prime} \sum k q^{k-1} = \frac{1}{(1-q)^2}$

$\xrightarrow{\times q} \sum k q^k = \frac{q}{(1-q)^2} \xrightarrow{\prime} \sum k^2 q^{k-1} = \frac{1-q^2}{(1-q)^4}$

$\xrightarrow{\times qp} \sum k^2 q^k p = \frac{(1+q)q}{(1-q)^2} = \frac{q+q^2}{p^2}$

$\rightarrow Var(X) = \frac{q+q^2}{p^2} - \frac{q^2}{p^2} = \frac{q}{p^2} = \frac{1-p}{p^2}$

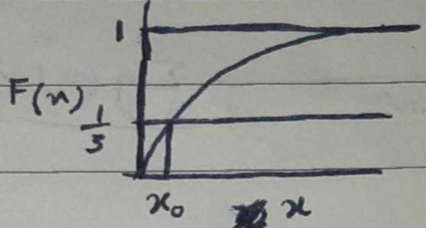
$\Rightarrow$ $T_j$'s are independent, but even if they were dependent, linearity of $E$ still Holds,

$$E[T] = E[T_1] + \cdots + E[T_n]$$

$$= 1 + \frac{n}{n-1} + \frac{n}{n-2} + \frac{n}{1} = n\underbrace{\left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n-1}\right)}_{\text{harmonic sum}}$$

$\approx$ $n \log n$ for Large $n$

* Universality




$X \sim F$

$F(x_0) = \frac{1}{3}$, $P(F(X) \le \frac{1}{3})$ ?

$= P(X \le x_0) = F(x_0) = \frac{1}{3}$

$\rightarrow F(X) \sim \text{Unif}(0,1)$

* Ex. Logistic Distribution: $F(x) = \frac{e^x}{1+e^x}$

$\rightarrow$ We want random draw from this distribution

$\Rightarrow u \sim \text{Unif}(0,1)$, consider

$F^{-1}(u) = \log \frac{u}{1-u}$

Memo No. ______
Date / /

* Ex. Let X, Y, Z be i.i.d positive r.v.s. Find $E\left[\frac{X}{X+Y+Z}\right]$

$$E\left[\frac{X}{X+Y+Z}\right] = E\left[\frac{Y}{X+Y+Z}\right] = E\left[\frac{Z}{X+Y+Z}\right] \quad \text{by symmetry}$$

$$E\left[\frac{X}{X+Y+Z}\right] + E\left[\frac{Y}{X+Y+Z}\right] + E\left[\frac{Z}{X+Y+Z}\right] \overset{\text{linearity}}{=} E\left[\frac{X+Y+Z}{X+Y+Z}\right] = 1$$

$\Rightarrow E\left[\frac{X}{X+Y+Z}\right] = \frac{1}{3}$, that makes sense because if we were asked that how much a R.V. contributes from 3 i.i.d R.V.s, we said $\frac{1}{3}$

* Ex. LOTUS $U \sim \text{unif}(0,1)$, $X = U^2$, $Y = e^X$, find $E[Y]$ as an integral

I) $E[Y] = \int_0^1 e^x f(x)\, dx$ PDF of X

$\xrightarrow{\text{CDF}} P(U^2 \leq x) = P(U \leq \sqrt{x}) = \sqrt{x}$ if $0<x<1$

$\rightarrow f(x) = \frac{1}{2\sqrt{x}}, \; x \in (0,1)$

II) $Y = e^X = e^{U^2} \Rightarrow E[Y] = \int_0^1 e^{u^2} \times 1\, du$

* Ex. Story Proof $X \sim \text{Bin}(n,p)$, $q = 1-p$, find distribution of $n-X$

PMF: $P(n-X=k) = P(X=n-k)$

$= \binom{n}{n-k} p^{n-k} q^k$

$= \binom{n}{k} q^k p^{n-k}$

Story Proof:

$n-X \sim \text{Bin}(n,q)$

by swapping "success"

Memo No.
Date / /

* Let $Y \sim Expo(1)$, find $E[Y]$, $Var(Y)$

$E[Y] = \int_0^\infty \underbrace{y}_{u} \underbrace{e^{-y}}_{dv} dy$, $u = y \to du = dy$

$dv = e^{-y} \to v = -e^{-y}$

$\Rightarrow \int_0^\infty ye^{-y}dy = (-ye^{-y})\Big|_0^\infty - \int_0^\infty -e^{-y}dy$

$= 0 - (e^{-y})\Big|_0^\infty = +1$

second moment

$Var(Y) = \underbrace{E[Y^2]}_{} - (EY)^2$

$\int_0^\infty \underbrace{y^2}_{u} \underbrace{e^{-y}}_{dv} dy - 1$, $u = y^2 \to du = 2y\,dy$

$dv = e^{-y} \to v = -e^{-y}$

$\Rightarrow \int_0^\infty y^2e^{-y}dy = (-y^2e^{-y})\Big|_0^\infty + 2\underbrace{\int_0^\infty e^{-y}y\,dy}_{1} = 2$

$\Rightarrow Var(Y) = 2 - 1 = 1$

* General Exponential: $X = \frac{Y}{\lambda}$ has $E[X] = \frac{1}{\lambda}$

$Var(X) = \frac{1}{\lambda^2}$

* The most important reason that Exp is important is the memorylessness property

$\Rightarrow$ no matter how long you waited, then it's like you're starting over from fresh

Memo No.
Date / /

for example: rate of emails in an hour

* Ex. Poisson # emails I get at time $t$ is $Pois(\lambda t)$

Find PDF of $T$, time of first got email

( connects discrete R.V to Continuous R.V )

Sometimes it's easier to ~~find~~ solve the complement problem

$P(T > t) = P(N_t = 0)$

where $N_t$ = (# emails in $[0, t]$)

$\frac{e^{-\lambda t}(\lambda t)^0}{(0)!} = e^{-\lambda t}$

$\Rightarrow$ CDF: $1 - e^{-\lambda t} \xrightarrow{} $ PDF: $\lambda e^{-\lambda t} \to$ Exponential Distribution

$t > 0$

Lecture 16

Exponential Distribution, parameter $\lambda$

rate parameter $\to$ a rate that some kind of event that occur

$X \sim Expo(\lambda)$ has PDF $\lambda e^{-\lambda x}$, $x > 0$ (0 otherwise).

valid PDF: $\int_{-\infty}^\infty \lambda e^{-\lambda x}dx = 1$

CDF: $F(x) = \int_{-\infty}^x \lambda e^{-\lambda t}dt = 1 - e^{-\lambda x}$, $x > 0$

modified to zero

* Let $Y = \lambda X$, then $Y \sim Expo(1) \Rightarrow$ Analogous to standardization in normal

Since: $P(Y \leq y) = P(\lambda X \leq y)$

$= P(X \leq \frac{y}{\lambda}) = 1 - e^{-\lambda \frac{y}{\lambda}} = 1 - e^{-y}$

* $X_i \sim Expo(\lambda_i)$ → $\min(X_1, \ldots, X_n) \sim Expo(\lambda_1 + \ldots + \lambda_n)$
$X_i$ independent — Itself is a hard problem

Memo No.
Date / /

$E[T \mid T>20] > E[T]$

* Life Expectancy, Let's say that it's 80
→ if memoryless, we would have $E[T \mid T>20]$
$= 20 + E[T]$, which is wrong
→ So memorylessness is not usually reasonable assumption, but it's sometimes and it's a good approximation

* If $X$ is a positive continuous r.v. with memoryless property, then it has to be $X \sim Expo(\lambda)$ for some $\lambda$

Proof Let $F$ be the CDF of $X$, $G(x) = P(X>x) = 1-F(x)$
memoryless property is : $G(s+t) = G(s)\,G(t)$
$: \underbrace{P(X \geq s+t \mid X>s)}_{P(X>t)} = \dfrac{P(X>s+t)}{P(X>s)}$
→ $P(X>s+t) = P(X>s)\,P(X>t)$
→ $G(s+t) = G(s)\,G(t)$ → not a usual equation, we want to solve it for $G$ → we want to show only exp satisfies
⇒ • Gradually solve the problem : $s=t$ → $G(2t) = G(t)^2$
, $G(3t) = G(t)^3$, ——, $G(kt) = G(t)^k$ if $k$ is positive integer
• $G(\frac{t}{2}) = G(t)^{1/2}$, $G(t/3) = G(t)^{1/3}$, ——, $G(t/k) = G(t)^{1/k}$
• $G(\frac{m}{n}t) = G(t)^{m/n}$
⇒ $G(xt) = G(t)^x$, $x \geq 0$
→ $t=1$ : $G(x) = G(1)^x = e^{x \ln G(1)} = e^{x \ln G(1)}$
$0 < G(1) < 1$ → $\ln G(1) < 0$ → $-\lambda$ → $e^{-\lambda x} = 1-F(x)$
→ $F(x) = 1 - e^{-\lambda x}$ → The only possibility

waiting R.v → exp in continuous
→ geometric in discrete

Memo No.
Date / /

* $P(X \geq s+t \mid X \geq s) = P(X \geq t)$
general defn of Memorylessness

→ $X \sim Expo(\lambda)$, Here $\underbrace{P(X \geq s)}$
$= 1 - P(X<s) = e^{-\lambda s}$

How long someone is gonna live
→ It's the prob that lives at least $s$ seconds
⇒ Survival ~~R~~ function

$P(X \geq s+t \mid X \geq s)$
$= \dfrac{P(X \geq s+t, \cancel{X \geq s})}{P(X \geq s)}$ redundant $= \dfrac{e^{-\lambda(s+t)}}{e^{-\lambda}} = e^{-\lambda t} = P(X \geq t)$

⇒ It turns out that Exponential is the only one that is memoryless in continuous time

* $X \sim Expo(\lambda)$ : $\underbrace{E[X \mid X>a]}$ $\xrightarrow{\text{linearity}}$ $a + \overset{\text{just a fresh R.v}}{E[X-a \mid X>a]}$
$= a + \dfrac{1}{\lambda}$

a brief introduction to conditional ~~Expectatio~~ Expectation

Lecture 17 * The fact that exp is the only one memoryless in continuous time. Geometric is the similar in Discrete time.

* Conditional Expectation

□(2) MGF determines the distribution, i.e., if X, Y have same MGF, then they have the same dist. (CDF, ...)

□(3) If X has MGF $M_X$, Y $M_Y$, X indep. of Y

MGF of X+Y is $E[e^{t(X+Y)}] \underset{\text{indep.}}{=} E[e^{tX}]\,E[e^{tY}]$

makes easier sums of r.v.s (convolution)

*. Ex. $X \sim Bern(p)$, $M(t) = E[e^{tX}] = pe^t + q$ ($q = 1-p$)

• $X \sim Bin(n,p)$, $\Rightarrow M(t) = (pe^t + q)^n$

• $Z \sim N(0,1)$ $M(t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{tz} e^{-z^2/2}\,dz$

$$= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{tz - z^2/2}\,dz = \frac{e^{t^2/2}}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(z-t)^2}\,dz$$

$$= e^{t^2/2}$$

* Laplace Rule of Succession: "What's the probability that the sun will rise tomorrow?"

or if we have observed the sun rising for the past n days in succession, then what is the prob. that the sun will rise tomorrow?

$X_1, X_2, \ldots$ i.i.d Bern(p), Given p,

Indicator Sun rises

given p, this is true

Conditional independence

* Moment Generating Function (MGF) Alternative way to describe a dist. like CDF, PDF

Defn A r.v. X has MGF □ $M(t) = E(e^{tX})$ as a function of t, if this is finite on some interval $(-a, a)$, $a > 0$

What's t variable: At first, letter t is a dummy variable, Secondly, what does it mean: any function of a r.v. is a r.v. → it makes sense that we can obtain the Expected value $E(e^{tX})$ → keeping track of the moments of a distribution

* Why Moment "Generating"?

□ $E[e^{tX}] = E\left[\sum_{n=0}^{\infty} \frac{X^n t^n}{n!}\right] = \sum_{n=0}^{\infty} \frac{E[X^n]\, t^n}{n!}$ n-th moment

if it was finite sum, that was clear by linearity but since it is infinite, it needs more justification → Real Analysis

* It would be useful if we were interested in the moments but what if we don't care about moments? mean and Variance OK but what about higher?

→ Why is MGF important? Let X have MGF M(t)

(1) The n-th moment $E[X^n]$, is the coef of $\frac{t^n}{n!}$ in taylor series of M, i.e., □ $M^{(n)}(0) = E[X^n]$

Bayes' rule for distribution

even though it's a r.v.

Memo No.
Date / /

this given p means we're just treating p as a constant

Q. $P(X_{n+1} = 1 \mid S_n = n)$

Solution: Bayes' rule: $f(p \mid S_n = k) = \dfrac{P(S_n = k \mid p)\, f(p)}{P(S_n = k)}$

does'nt depend on p

prior, 1 uniform

Law of total prob.
$P(S_n = k) = \int_0^1 P(S_n = k \mid p)\, f(p)\, dp$

we don't need it now, a constant doesn't depend on p

proportionality $\Longrightarrow \propto \underbrace{\cancel{\ }}_{\text{constant}} p^k (1-p)^{n-k}$ → to get the constant, then we have to integrate this thing ⇒ much later in the course

also remove $\binom{n}{k}$

⇒ $f(p \mid S_n = n) = (n+1)p^n$ → easy to integrate so we do and find constant

$P(X_{n+1} = 1 \mid S_n = n) = \int_0^1 p(n+1)p^n \, dp$

$= \frac{n+1}{n+2}$

fundamental bridge
$E[p \mid S_n = n]$

Lecture 18 Expo MGF $X \sim \text{Expo}(1)$, find MGF, moments

$M(t) = E(e^{tX}) = \int_0^\infty e^{tx} e^{-x} dx = \int_0^\infty e^{-x(1-t)} dx$

$= \frac{1}{1-t}$, $t < 1$ → exponential decay → this is okay because we talked about the fact that we wanted some interval it becomes finite

$t > 1$ → blow up

Memo No.
Date / /

→ ~~The twist is that Probability that the sun rises is unknown~~ → p is unknown → How to deal with unknown p → How to deal with unknowns?

??? → frequentist
→ bayesian: a r.v.: distribution is just a reflection of our uncertainty: treat p as a r.v.

→ and now we can use bayes' rule → start with some prior belief about p, before we have any data or any evidence then we collect data and update based on evidence

→ some new uncertainty ⇒ p as a r.v.

Structure of problem

→ let $p \sim \text{Unif}(0,1)$ (prior): why? complete uncertainty — uniform reflect

→ Let $S_n = X_1 + \cdots + X_n$

$\underbrace{S_n \mid p}_{}$ ~ $\text{Bin}(n, p)$, $p \sim \text{Unif}(0,1)$

treat p as a constant → if we know the prob. of sun rises (which coin we have)

then they're i.i.d (conditional Independence)

so and so sum of i.i.d Bern(p) → Bin(n, p)

Q. Find Posterior $p \mid \underbrace{S_n}$

we assume we observe S

⇒ You could also assume that you observe $X_1$ through $X_n$

⇒ You get the same thing ⇒ Sufficient Statistic ⇒ ???

→ number of ways to break two n people into end to end partnerships → $\frac{(2n)!}{2^n n!}$ or 1.3.5. ——

→ same as even moments of Normal → Not a coincidence → deep combinatorial explanation

* Poisson MGF, the other important reason to use MGF

$X \sim \text{Pois}(\lambda)$. $E(e^{tX}) = \sum_{k=0}^{\infty} e^{tk} \frac{e^{-\lambda}\lambda^k}{k!}$

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

→ We can start taking derivative, but we want to show other application of MGF

let $Y \sim \text{Pois}(\mu)$ indep of $X$. Find dist. of $\overbrace{X+Y}^{\text{Convoluting}}$

Multiply MGFs: $e^{\lambda(e^t-1)} \times e^{\mu(e^t-1)}$

$= e^{(\lambda+\mu)(e^t-1)} \Rightarrow X+Y \sim \text{Pois}(\lambda+\mu)$, $\square$

nice property of Poisson that sums of i.i.d, still Poisson

$X, Y$ should be independent

counter example: the most extreme case of dependence: $X = Y \Rightarrow X+Y = 2X$ is not Poisson

Why? This thing is always even, but Poisson: any non-negative integer value

mean → $2\lambda$
variance → $4\lambda$ } mean ≠ ~~variable~~ variance while mean = variance in Poisson

→ $M'(0) = E[X]$, $M''(0) = E[X^2]$, $M'''(0) = E[X^3]$, ——

→ But what if we don't want to find derivative?

$$\frac{1}{1-t} = \sum_{n=0}^{\infty} t^n = \sum_{n=0}^{\infty} n! \frac{t^n}{n!}$$

→ $\mathbb{E}[X^n] = n!$

* if you want to find the moment of some dist. by LOTUS → Maybe Incredibly hard Integral

→ but with MGF → derivatives

$Y \sim \text{Expo}(\lambda)$, $X = \lambda Y \sim \text{Expo}(1)$

, so $Y^n = \frac{X^n}{\lambda^n} \rightarrow E[Y^n] = \frac{n!}{\lambda^n}$

Easier to work with Expo(1), similarly like Standard Normal

* Normal Moments

let $Z \sim N(0,1)$, find <u>all</u> its moments

$E(Z^n) = 0$ for $n$ odd by symmetry

$M(t) = e^{t^2/2} = \sum_{n=0}^{\infty} \frac{(t^2/2)^n}{n!} = \sum_{n=0}^{\infty} \frac{t^{2n}}{2^n n!} \frac{(2n)!}{(2n)!}$

↳ don't want to take derivative

$E[Z^{2n}] = \frac{(2n)!}{2^n n!}$

$n=1 \Rightarrow E(Z^2) = 1$

$n=2 \Rightarrow E[Z^4] = 3$ $n=3 \rightarrow E[Z^6] = 1.3.5 = 15$

Mo Tu We Th Fr Sa Su    Memo No. ______ Date / /

→ number of ways to break two n people into end to end partnerships → $\frac{(2n)!}{2^n\, n!}$ or 1.3.5. ——

→ same as even moments of Normal → Not a coincidence → deep combinatorial explanation

* Poisson MGF, the other important reason to use MGF

$X \sim \text{Pois}(\lambda)$. $E(e^{tX}) = \sum_{k=0}^{\infty} e^{tk} \frac{e^{-\lambda}\lambda^k}{k!}$

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

→ We can start taking derivative, but we want to show other application of MGF

let $Y \sim \text{Pois}(\mu)$ indep of $X$. Find dist. of $\overbrace{X+Y}^{\text{Convolution}}$

Multiply MGFs: $e^{\lambda(e^t-1)} \times e^{\mu(e^t-1)}$

$= e^{(\lambda+\mu)(e^t-1)} \Rightarrow X+Y \sim \text{Pois}(\lambda+\mu)$ □

nice property of Poisson that sums of i.i.d, still Poisson

X, Y should be independent

counter example: the most extreme case of dependence: $X = Y \Rightarrow X+Y = 2X$ is not Poisson

Why? This thing is always even, but Poisson: any non-negative integer value

mean → $2\lambda$
variance → $4\lambda$ } mean ≠ variance while mean = variance in Poisson

---

Mo Tu We Th Fr Sa Su    Memo No. ______ Date / /

→ $M'(0) = E[X]$, $M''(0) = E[X^2]$, $M'''(0) = E[X^3]$

, ——

→ But what if we don't want to find derivative?

$\frac{1}{1-t} = \sum_{n=0}^{\infty} t^n = \sum_{n=0}^{\infty} n! \frac{t^n}{n!}$

→ $E[X^n] = n!$

* if you want to find the moment of some dist. by LOTUS → Maybe Incredibly hard Integral

→ but with MGF → derivatives

$Y \sim \text{Expo}(\lambda)$, $X = \lambda Y \sim \text{Expo}(1)$

, so $Y^n = \frac{X^n}{\lambda^n} \rightarrow E[Y^n] = \frac{n!}{\lambda^n}$

Easier to work with Expo(1), Similarly like Standard Normal

* Normal Moments

let $Z \sim N(0,1)$, find all its moments

$E(Z^n) = 0$ for n odd by symmetry

$M(t) = e^{t^2/2} = \sum_{n=0}^{\infty} \frac{(t^2/2)^n}{n!} = \sum_{n=0}^{\infty} \frac{t^{2n}}{2^n n!} \frac{(2n)!}{(2n)!}$

↳ don't want to take derivative

$E[Z^{2n}] = \frac{(2n)!}{2^n\, n!}$

$n=1 \Rightarrow E(Z^2) = 1$

$n=2 \Rightarrow E[Z^4] = 3$    $n=3 \rightarrow E[Z^6] = 1.3.5 = 15$

Mo Tu We Th Fr Sa Su

Memo No.

Date / /

~~Q. is X, Y in Ex. independent~~

→ **Getting Marginals** $P(X=x) = \sum_{y} P(X=x, Y=y)$

marginalizing over $y$

* continuous

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dx$$

⇒ we can't go othe direction : marginal $\not\to$ joint

→ Answer to Q:

| | Y=0 | Y=1 | |
|---|---|---|---|
| X=0 | 2/6 | 1/6 | 3/6 |
| X=1 | 2/6 | 1/6 | 3/6 |
| | 4/6 | 2/6 | |

▫ margin

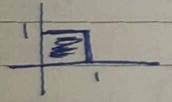
→ indep. because for every $x, y$ $P(X=x, Y=y) = P(X=x)\,P(Y=y)$

| | | | |
|---|---|---|---|
| 1/2 | 0 | 1/2 |
| 1/4 | 1/4 | 1/2 |
| 3/4 | 1/4 | |

→ dependent

* Continuous Ex. of Joint

Ex. Uniform on Square $\{(x,y) : x, y \in [0,1]\}$



joint PDF const on the square

, 0 outside of square

$$\begin{cases} c, & 0 \le x \le 1,\ 0 \le y \le 1 \\ 0, & \text{o.w} \end{cases}$$

▫ integral is area, $c = \frac{1}{\text{area}} = 1$

marginal : $X, Y$ are indep. Unif$(0,1)$

Mo Tu We Th Fr Sa Su

Memo No.

Date / /

**Joint Distributions** : How do we work with the dist. of more than one R.V.

indep R.V.s → joint just means multiply the individuals CDFs and PDFs

In general we need to have some tools and notation an so on to deal with dependent R.V.s

Simple Ex. → X, Y Bernoulli

| | Y=0 | Y=1 |
|---|---|---|
| X=0 | 2/6 | 1/6 |
| X=1 | 2/6 | 1/6 |

* X, Y r.v.s

joint CDF : $F(x,y) = P(X \le x, Y \le y)$

joint PMF : (discrete case) $P(X=x, Y=y)$

marginal CDF : means take them seperately : $P(X \le x)$ is marginal dist. of X

joint PDF : (cont.) $f(x,y)$ such that $P((x,y) \in B) = \iint_B f(x,y)\,dx\,dy$

The ~~PDP~~ PDF is what you integrate to get a probability

▫ independence $X, Y$ indep. iff $F(x,y) = \underbrace{F_X(x)}_{\text{marginal CDF of } X}\ \underbrace{F_Y(y)}_{\text{marginal CDF of } Y}$

conditional and Bayes rule for R.V.s in next session | Note the difference between Event form and R.V. form

Equiv. to $P(X=x, Y=y) = P(X=x)\,P(Y=y)$

$f(x,y) = f_X(x)\,f_Y(y)$

◇ all $x, y \in \mathbb{R}$

* conditional PDF of $Y|X$ is : We get to pretend that we know what $X$ is

$$\underbrace{f_{Y|X}}_{\text{still a R.V.}}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

joint

marginal

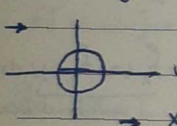
Bayes' rule $= \dfrac{f_{X|Y}(x|y)\, f_Y(y)}{f_X(x)}$

* $X, Y$ indep if $\underbrace{f_{X,Y}(x,y)}_{\text{joint}} = \underbrace{f_X(x)\, f_Y(y)}_{\text{marginal}}$ for all $x, y$

→ the same for CDF

* Continue Ex. from last time → Uniform means that probability of some region is proportional to area

→



unif in disc $x^2+y^2 \le 1$. $f(x,y) = \begin{cases} 1/\pi & x^2+y^2 \le 1 \\ 0 & \text{o.w} \end{cases}$

→ $X, Y$ are not independent

→ marginal: $f_X(x) = \int 1/\pi \, dy$ → *limits of integral

→ $y^2 \le 1 - x^2$ → think carefully about limits of integration

→ $\displaystyle\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} 1/\pi \, dy = \underbrace{\frac{2}{\pi}\sqrt{1-x^2}}$ $-1 \le x \le 1$

→ this should be integrated to one

□ not uniform anymore

→ conditional: $f_{Y|X}(y|x) = \dfrac{1/\pi}{\frac{2}{\pi}\sqrt{1-x^2}}$ if $-\sqrt{1-x^2} \le y \le \sqrt{1-x^2}$

just a fixed number

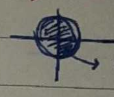
$\Rightarrow \underbrace{Y|X}_{\text{R.V}} \sim \text{Unif}\left(-\sqrt{1-X^2}, \sqrt{1-X^2}\right)$

△ We know $X$ is a R.V., but pretend $X$ is a known constant



→ Contrast with an ex. of dependence, instead of square, circle

Ex. Unif in disc $x^2+y^2 \le 1$



→ Unif in here

joint PDF: $\begin{cases} 1/\pi, & x^2+y^2 \le 1 \\ 0, & \text{o.w} \end{cases}$

□ but they are dependent : $X = 0 \to -1 \le y \le 1$

$x \to 1 \to$ $y$ can just be some tiny interval

□ Given $X = x \Rightarrow -\sqrt{1-x^2} \le y \le \sqrt{1-x^2}$

→ So we might guess that $y$ is uniform between here and here → next time we'll do integral to show ~~that~~,

Lecture 19

Joint, conditional, marginal dist.

→ we have tools for 1 R.V. → Now we need tools for more r.v.s

joint CDF $F(x,y) = P(X \le x, Y \le y)$ → $\begin{cases} \text{discrete} \\ \text{continuous} \\ \text{mix case} \end{cases}$

cont. case, joint PDF $f(x,y) = \dfrac{\partial}{\partial x \partial y} F(x,y)$

not prob., density → what you integrate to get prob.

$P((X,Y) \in A) = \iint_A f(x,y)\, dx\, dy$

marginal PDF of $X$ : $\displaystyle\int_{-\infty}^{\infty} f(x,y)\, dy$ → marginalization

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

Symmetry: $2\int_0^1\int_y^1 (x-y)\,dx\,dy$ (Limits of integration)

refers to y

outer limits have to just to be numbers

as you move inward, the limit can start to depend on other variables

$= 2\int_0^1 \left(\frac{x^2}{2} - yx\right)\Big|_y^1 dy = \frac{1}{3}$

0 — $\frac{1}{3}$ — $\frac{2}{3}$ — 1 $\rightarrow$ $M = \max(X, Y)$

$L = \min(X, Y)$

$|X-Y| = M - L \rightarrow E[M-L] = \frac{1}{3} = E[M] - E[L]$

$E[M+L] = \underbrace{E[M] + E[L]}_{\text{Uniform}} = \frac{1}{2} + \frac{1}{2} = 1$

$\Rightarrow E[M] = \frac{2}{3} \quad E[L] = \frac{1}{3}$

* Chicken-Egg. $N \sim \text{Pois}(\lambda)$ eggs, each hatches with prob. $p$ indep. Let $X$ = # that hatch so $\underbrace{X|N}_{\text{pretend } N \text{ is constant}} \sim \text{Bin}(N, p)$. Let $Y$ = # don't hatch

so $X + Y = N$. Find joint PMF of $X, Y$.

$\rightarrow P(X=i, Y=j) = \underbrace{\sum_{n=0}^{\infty} P(X=i, Y=j \mid N=n)\, P(N=n)}$

it looks intimitating infinite sum but it's actually just one term

seems dependent

Mo Tu We Th Fr Sa Su
Memo No.
Date / /

Which means $Y|X=x \sim \text{Unif}\left(-\sqrt{1-x^2}, \sqrt{1-x^2}\right)$

$\rightarrow$ So it's conditionally uniform not uniform

$\rightarrow$ They're not independent: $f_{X,Y}(x,y) \neq f_X(x) f_Y(y)$

*2-D LOTUS Let $(X,Y)$ have joint PDF $f(x,y)$, and let $g(x,y)$ be a real-valued function of $x, y$.

Then $E\,g(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\, f(x,y)\,dx\,dy$

*Thm If $X, Y$ are indep., then $E[XY] = E[X]E[Y]$

("Indep. implies uncorrelated.")

Proof (continuous case): $E[XY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, \underbrace{f_X(x) f_Y(y)}_{\text{independence}}\,dx\,dy$

$= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} xy\, f_X(x) f_Y(y)\,dx\right)dy = \int_{-\infty}^{\infty} y f_Y(y) \underbrace{\int_{-\infty}^{\infty} x f_X(x)\,dx}_{\text{constant}: E[X]}\,dy$

constant: E[X] (number) so take it out integral

$= E[X]\,E[Y]$

* Ex. $X, Y \overset{\text{i.i.d}}{\sim} \text{Unif}(0,1)$, find $E[|X-Y|]$

$f_{XY}(x,y) = 1$ (Uniform)

LOTUS $\int_0^1\int_0^1 |x-y| \times 1\,dx\,dy = \iint_{x>y} (x-y)\,dx\,dy + \iint_{x\leq y} (y-x)\,dx\,dy$

Memo No.
Date / /

* Multinomial : The most important discrete multivariate distribution

Multivariate Distribution: Just means a distribution that's a joint distribution → more than one r.v.

Multinomial : generalization of Binomial, Higher Dimension version

Defn / story of $Mult(n, \vec{p})$, $\vec{p} = (p_1, \ldots, p_k)$ prob. vector $\sum p_i = 1$, $p_i \geq 0$

The intuition: in Binomial we just talked about success and failure (2 possible outcome, 2 categories) → k categories

→ $\vec{X} \sim Mult(n, \vec{p})$, $\vec{X} = (X_1, \ldots, X_k)$ if have n objects, which we are indep. putting into k categories

$p_j = P(\text{category } j)$, $X_j$ = # objects in category j

→ Joint Distribution → Joint PMF

$$P(X_1 = n_1, \ldots, X_k = n_k) = \frac{n!}{n_1! n_2! \cdots n_k!} p_1^{n_1} \cdots p_k^{n_k}$$

if $n_1 + \cdots + n_k = n$
(0 o.w.)

2 3 3 1 1 1 1 2 2 2 1

* Marginal : $\vec{X} \sim Mult_k(N, \vec{p})$. Find Marginal Dist of $X_j$. Then $X_j \sim Bin(n, p_j)$

→ Immediate from story → n trial each trial either belongs to category j with $p_j$ and doesn't belong to $1 - p_j$

Memo No.
Date / /

$= P(X = i, \cancel{Y = j} \mid N = i+j) \, P(N = i+j)$

redundant info.

$= P(X = i \mid N = i+j) \, P(N = i+j)$

$= \frac{(i+j)!}{i! \, j!} p^i q^j \, e^{-\lambda} \frac{\lambda^{i+j}}{(i+j)!}$

$p+q=1$

$= \left(e^{-\lambda p} \frac{(\lambda p)^i}{i!}\right)\left(e^{-\lambda q} \frac{(\lambda q)^j}{j!}\right) \Rightarrow X, Y$ are indep!

$X \sim Pois(\lambda p)$, $Y \sim Pois(\lambda q)$

→ it's because of $N \sim Pois(\lambda)$

Lecture 20

Ex. Find $E[|Z_1 - Z_2|]$, with $Z_1, Z_2 \overset{i.i.d}{\sim} N(0,1)$

→ First we simplify the problem

**Thm** $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$ indep. Then $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

**Proof** Use MGFs: MGF of $X+Y$ is

$$e^{\mu_1 t + \frac{1}{2}\sigma_1^2 t^2} e^{\mu_2 t + \frac{1}{2}\sigma_2^2 t^2} = e^{(\mu_1 + \mu_2)t + \frac{1}{2}(\sigma_1^2 + \sigma_2^2)t^2}$$

→ MGF of $N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

$\Longrightarrow$ Note $Z_1 - Z_2 \sim N(0, 2)$

$\Rightarrow E[|Z_1 - Z_2|] = E[|\sqrt{2} Z|]$, $Z \sim N(0,1)$

$= \sqrt{2} \, E[|Z|]$

$$= \sqrt{2} \int_{-\infty}^{\infty} |z| \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz = 2\sqrt{2} \int_0^{\infty} z \frac{1}{2\pi} e^{-z^2/2} dz$$

$$= \sqrt{\frac{2}{\pi}}$$

Memo No.
Date / /

$F(t)$

$\rightarrow P(\frac{X}{Y} \leq t) \overset{\text{symmetry of } N(0,1)}{=} P(\frac{X}{|Y|} \leq t)$

$= P(X \leq t|y|) = \int_{-\infty}^{\infty}\int_{-\infty}^{t|y|} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} \, dx\,dy$

$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2} \, \Phi(t|y|) \, dy$

$= \sqrt{\frac{2}{\pi}} \int_0^{\infty} e^{-y^2/2} \, \Phi(ty) \, dy$

$\Phi$ is an intractable integral
we can't do it

PDF: $F'(t) = \sqrt{\frac{2}{\pi}} \int_0^{\infty} y e^{-y^2/2} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2y^2}{2}} dy$

$= \frac{1}{\pi} \int_0^{\infty} y e^{-\frac{(1+t^2)y^2}{2}} dy$

$= \frac{1}{\pi}\left(\frac{-1}{1+t^2} e^{-\frac{(1+t^2)y^2}{2}}\right)\Big|_0^{\infty} = \frac{1}{\pi(1+t^2)}$

$\Rightarrow$ CDF: $\int_{-\infty}^{t} \frac{1}{\pi(1+x^2)} dx = \frac{tg^{-1}(t)}{\pi}$

$\Rightarrow$ Alternative with $\iint$: Law of total probability

$P(X \leq t|Y|) \rightarrow$ what we would wish to know

Memo No.
Date / /

$\Rightarrow E[X_j] = np_j$ , $Var(X_j) = np_j(1-p_j)$

* lumping property $\vec{X} = (X_1, X_2, \dots, X_{10}) \sim Mult(n, (p_1, \dots, p_{10}))$

$\rightarrow$ 10 political parties, every n people belongs to one of those. $p_j$ is the prob. belonging to $X_j$. Two dominant party, and we want lump all other to a 3rd party

$\rightarrow$ Let $Y = (X_1, X_2, X_3 + \dots + X_{10})$

. Then $\vec{Y} \sim Mult(n, (p_1, p_2, p_3 + \dots + p_{10}))$

* Conditional: $\vec{X} \sim Mult(n, \vec{p})$. Then given $X_1 = n_1$,

$(X_2, \dots, X_k) \sim Mult_{k-1}(n - n_1, \cancel{(p_2, \dots, p_k)}$
$(p_2', \dots, p_k'))$

with $p_2' = P($being in category 2 | not in category 1$)$

$= \frac{p_2}{1-p_1} = \frac{p_2}{p_2 + \dots + p_k}$ , $\underbrace{p_j' = \frac{p_j}{p_2 + \dots + p_k}}_{\text{renormalize}}$

* Cauchy Interview Problem

The cauchy is dist. of $\frac{X}{Y}$ with $X, Y \overset{i.i.d}{\sim} N(0,1)$
Find PDF of this r.v.: $T = \frac{X}{Y}$

$\Rightarrow$ It doesn't have mean & Variance

$\Rightarrow$ But the dist. of average of million i.i.d ~~each~~ cauchys is still cauchy $\rightarrow$ doesn't change distribution

Law of Large number

Memo No.
Date / /

* More properties : (3) $Cov(X, c) = 0$ if c is constant

(4) $Cov(cX, Y) = c\,Cov(X,Y)$

(5) $Cov(X, Y+z)$

$E[X(Y+z)] - E[X]E[Y+z] = E[XY] + zE[X] - E[X]E[Y] - zE[X]$

$= Cov(X,Y)$ , if z is constant

(6) $Cov(X, Y+Z) = Cov(X,Y) + Cov(X,Z)$
bilinearity

(7) $Cov(X+Y, Z+W) = Cov(X,Z) + Cov(X,W) + Cov(Y,Z) + Cov(Y,W)$

→ $Cov\left(\sum_{i=1}^{n} a_i X_i , \sum_{j=1}^{n} b_j Y_j\right) = \sum_{i,j} a_i b_j Cov(X_i, Y_j)$

(8) $Var(X_1 + X_2) = Cov(X_1+X_2, X_1+X_2)$

$= Var(X_1) + Var(X_2) + 2Cov(X_1, X_2)$

→ if they are independent → Cov is 0 → linearity

$Var(X_1 + \dots + X_n) = Var(X_1) + \dots + Var(X_n) + 2\sum_{i<j} Cov(X_i, X_j)$

* Thm if X, Y are indep., then they're uncorrelated, i.e $Cov(X,Y) = 0$

→ Converse is False! : cov = 0 doesn't imply independence

Memo No.
Date / /

$$= \int_{-\infty}^{\infty} P(X \le t|Y| \mid Y=y) \underbrace{\varphi(y)}_{N(0,1)\ PDF} dy$$

$$\overset{\text{independent}}{=} \int_{-\infty}^{\infty} \Phi(t|y|)\, \varphi(y)\, dy \Rightarrow \text{the same integral}$$

## Lecture 21

Let us deal with variance of a sum.

→ it's what we need when we want to study two R.V.s

Defn $Cov(X,Y) = E[(X-EX)(Y-EY)]$

→ why define it this way ? first of all it's a product

→ we try to see how they vary together / $+ \times + = +$ ; $+ \times - = -$ ; $- \times - = +$ /

→ now suppose a lot of i.i.d pairs, within each pair $(x_i, y_i)$ they're not necessarily indep. / by indep. we have showed $E[XY] = E[X]E[Y]$

→ we're interested in what happens if they are not indep.

→ $\begin{cases} X > E[X] \to \text{tends to } Y \text{ above mean} \to \text{positively correlated} \\ X < E[X] \to \text{〃 〃 〃 below mean} \end{cases}$

→ $X > E[Y]$ → it doesn't imply that Y is below its mean but it has more tendency that $Y < E[Y]$ → negatively correlated

* Properties (1) $Cov(X,X) = Var(X)$

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

* $Cov(X,Y) = E[XY] - E[X]E[Y]$

Proof: $E(XY) - E[X]E[Y] - E[X]E[Y] + E[X]E[Y]$
by linearity

Mo Tu We Th Fr Sa Su    Memo No. ______ Date / /

→ Imagine sb say $Cov(X,Y) = 42$, what is that telling you?
is it big number or small?

⇒ Corr is dimensionless quantity which means unitless
→ major advantage of correlation

$$\to Cov\left(\frac{X - \overbrace{EX}^{\text{constant}}}{SD(X)}, \frac{Y - \overbrace{EY}^{\text{constant}}}{SD(Y)}\right) = Cov\left(\frac{X}{SD(X)}, \frac{Y}{SD(Y)}\right)$$

$$= \frac{Cov(X,Y)}{SD(X)\,SD(Y)}$$

* **Thm** $-1 \leq Corr(X,Y) \leq 1$ ( Form of Cauchy - Schwartz)

**Proof** WLOG (without loss of generality) assume $X, Y$ are ~~independent~~ standardized, let $Corr(X,Y) = \rho$

$Var(X+Y) = Var(X) + Var(Y) + 2\,Cov(X,Y)$

by standardized $= 1 + 1 + 2\rho = 2 + 2\rho$

$Var(X-Y) = Var(X + (-Y)) = Var(X) + Var(Y) - 2\,Cov(Y)$

$= 2 - 2\rho$

$$Var \geq 0 \to \left.\begin{array}{l} 2 - 2\rho \geq 0 \\ 2 + 2\rho \geq 0 \end{array}\right\} \to -1 \leq \rho \leq 1$$

**Ex. Cov in a Multinomial** $(X_1, \ldots, X_k) \sim Mult(n, \vec{p})$

number of objects in this category

Find $Cov(X_i, X_j)$ for all $i, j$

Mo Tu We Th Fr Sa Su    Memo No. ______ Date / /

$Z \sim N(0,1)$, $X = Z$, $Y = Z^2$.

$$Cov(X,Y) = E[XY] - E[X]E[Y] = \underbrace{E[Z^3]}_{\substack{\text{odd moment} \\ \text{of standard} \\ \text{normal}}} - \overbrace{E[Z]}^{\text{odd moment}}E[Z^2] = 0$$

uncorrelated, but very dependent

: $Y$ is a function of $X$

⇒ intuition: correlation is a measure of linear association

of why doesn't capture that

→ there is a thm which said if every function of $Y$ is uncorrelated of every function of $x$, then independent, but just having linear things uncorrelated, doesn't imply independent

* **Defn** ~~Corr~~ You think of it as a normalized version of covariance

$$\cancel{Corr}\; Corr(X,Y) = \frac{Cov(X,Y)}{STD(X)\,STD(Y)}$$

$$= Cov\left(\underbrace{\frac{X - E[X]}{STD(X)}}_{\text{Standardization}}, \underbrace{\frac{Y - E[Y]}{STD(Y)}}_{\text{Standardization}}\right]$$

→ So correlation means standardize them first then take the covariance → The reason this is a useful to do is that covariance has an annoying as far as interpretation in terms of unit and things like that → Imagine $X, Y$ are distance quantity

Imagine one in nanometers, one in light years

Mo Tu We Th Fr Sa Su

Memo No. ____
Date / /

Ex. $X \sim HGeom(w, b, n)$, we have a jar that has w white balls, b black balls, we take a sample of size n, dist of number of white balls in n samples

$X = X_1 + X_2 + \cdots + X_n$, $X_j = \begin{cases} 1 & \text{if white} \\ 0 & \text{o.w} \end{cases}$

drawing balls from a jar without replacement, in binomial with replacement ⇒ so they are dependent, because of replacement

$\Rightarrow Var(X) = n Var(X_1) + 2\binom{n}{2} Cov(X_1, X_2)$ think why symmetry hold

$\Rightarrow Cov(X_1, X_2) = \underbrace{E[X_1 X_2]}_{\text{by fact}} - \underbrace{E[X_1]}_{\frac{w}{w+b}} \underbrace{[X_2]}_{\frac{w}{w+b}}$ fundamental bridge

$I_{X_1 \cap X_2}$ → $\frac{w}{w+b} \times \frac{w-1}{w+b-1}$

## Lecture 22

* Quickly recap: Var of $HGeom(w, b, n)$,

$p = \frac{w}{w+b}$, $w+b = N$ → not a R.V, $\begin{cases} \text{Population size: } N \\ \text{Sample size: } n \end{cases}$

$Var\left(\sum_{j=1}^{n} X_j\right) = Var(X_1) + \cdots + Var(X_n) + 2 \sum_{i<j} Cov(X_i, X_j)$

→ the j-th ball before you draw any balls, that's equally likely any of the balls → symmetry

$= \underbrace{n Var(X_1)}_{np(1-p)} + 2\binom{n}{2}\Big(\underbrace{E[X_1 X_2]}_{\text{intersection}} - \underbrace{\underbrace{E[X_1]}_{Bern(p)} E[X_2]}_{p^2}\Big)$

intersection: $\frac{w}{w+b} \cdot \frac{w-1}{w+b-1}$

$= \underbrace{\frac{N-n}{N-1}}_{\text{finite population correction}} np(1-p)$

Mo Tu We Th Fr Sa Su

Memo No. ____
Date / /

If $i = j$, $Cov(X_i, X_i) = Var(X_i) = np_i(1-p_i)$

Now let $i \neq j$ → Find $Cov(X_1, X_2) = c$?

$\Rightarrow \underbrace{Var(X_1 + X_2)}_{\text{just this remained}} = np_1(1-p_1) + np_2(1-p_2) + 2c$

↳ this just says merge the first two categories into one bigger → lumping property

→ $Var(X_1 + X_2) = n(p_1 + p_2)(1 - (p_1 + p_2))$ ⇒ solve for c

$\Rightarrow Cov(X_1, X_2) = -np_1 p_2 \Rightarrow Cov(X_i, X_j) = -np_i p_j < 0 \quad i \neq j$

→ notice it's negative → fixed number of people, if you knew there were more people in the first category, there's fewer people left for second → competing for membership

* Ex. $X \sim Bin(n, p)$, $X = X_1 + \cdots + X_n$, $X_j \overset{i.i.d}{\sim} Bern(p)$

facts about Indicator r.v.

→ let $I_A$, be indicator r.v. of event A

$I_A^2 = I$, $I_A^3 = I_A$ → because it's zero or one

→ $I_A I_B = I_{A \cap B}$

$\Rightarrow Var\, X_j = E X_j^2 - (E X_j)^2$

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

$Var(X) = npq$ since $Cov(X_i, X_j) = 0 \quad i \neq j$ (indep.)

Mo Tu We Th Fr Sa Su

Memo No.

Date / /

→ CHECK the function first: $e^x$ : $\left\{ \begin{array}{l} \text{strictly increasing} \\ \text{differentiable} \end{array} \right.$

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-(\ln y)^2/2} \quad \frac{dz}{dy}$$

it should be written in terms of y

$$\frac{dy}{dz} = e^z = y \Rightarrow \frac{dz}{dy} = \frac{1}{y}$$

$$\Rightarrow f_Y(y) = \frac{1}{y}\frac{1}{\sqrt{2\pi}} e^{-(\ln y)^2/2}, \quad y > 0$$

* multidimensional version: transformation $\mathbb{R}^n$:

Y and X as $\underbrace{\text{random vectors}}_{\text{a list of r.v.s}}$: $\vec{Y} = g(\vec{X})$, $g: \mathbb{R}^n \to \mathbb{R}^n$

a list of r.v.s → n r.v listed in one vector

$\vec{X} = (X_1, \ldots, X_n)$, (it has joint PDF), what's the joint
continuous

PDF of Y? $f_{\vec{Y}}(\vec{y}) = f_{\vec{X}}(\vec{x}) \left| \frac{d\vec{x}}{d\vec{y}} \right|$ → How to interpret derivative of a vector based on a vector

Jacobian

absolute value of determinant

$$\frac{d\vec{x}}{d\vec{y}} = \begin{pmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \vdots & & & \\ \frac{\partial x_n}{\partial y_1} & \cdots & & \frac{\partial x_n}{\partial y_n} \end{pmatrix}$$

$$\left| \frac{d\vec{y}}{d\vec{x}} \right|^{-1} = \left| \frac{d\vec{x}}{d\vec{y}} \right|$$

Mo Tu We Th Fr Sa Su

Memo No.

Date / /

→ extreme cases: $n=1$ → $p(1-p)$ → variance of Bern(p) which make sense, because if you draw just one, → No difference with replacement or without replacement

→ $N \gg n$ → $\frac{N-n}{N-1} \to 1$ → Binomial Variance

→ make sense, very very unlikely you sample the same more than once

* Transformations

* Thm Let X be continuous r.v. with PDF $f_X$, $Y = g(X)$

→ Assumptions on g: g is differentiable, strictly increasing

→ Then the PDF of Y is given by:

$$f_Y(y) = f_X(x)\frac{dx}{dy} \quad \text{where } y = g(x),\ x = g^{-1}(y)$$

and this is written in terms of y

Also $\frac{dx}{dy} = \left(\frac{dy}{dx}\right)^{-1}$ ( Also if strictly decreasing, $\left| \frac{dx}{dy} \right|$
abs

Proof CDF of Y: $P(Y \le y) = P(g(X) \le y)$

$= P(X \le g^{-1}(y)) = F_X(g^{-1}(y)) = F_X(x)$

$\Rightarrow f_Y(y) = f_X(x)\frac{dx}{dy}$ [chain rule]

* Ex. $Y = e^Z$, $Z \sim N(0,1)$

log Normal → log is a normal, not log of normal our randomness is because of 2 random committees.

Idea: find average overlap of 2 random comittees.

E[overlap] = ?

$\rightarrow$ $E[\text{overlap}] = \underbrace{100}_{\text{indicator}} \cdot \frac{\binom{3}{2}}{\binom{15}{2}} = \frac{300}{\frac{15 \cdot 14}{2}} = \frac{40}{14} = \frac{20}{7}$

each person is on both of random committees

$\Rightarrow$ there exists a pair of committe with overlap of $\geq \frac{20}{7}$

$\rightarrow$ overlap is integer $\rightarrow$ have overlap of 3

Lecture 23

* Convolution (sums) Let $T = X+Y$, dist $T$ ? ($X, Y$ indep.)

discrete: $P(T=t) = \sum_x P(X=x)\,P(Y=t-x)$

continuous: $f_T(t) = \int_{-\infty}^{\infty} f_X(x)\, f_Y(t-x)\, dx$

Since: $F_T(t) = P(T \leq t) = \int_{-\infty}^{\infty} P(X+Y \leq t \mid X=x)\, f_X(x)\, dx$

$= \int_{-\infty}^{\infty} F_Y(t-x)\, f_X(x)$

$\xrightarrow{\frac{d}{dt}}$ $f_T(t) = \int_{-\infty}^{\infty} f_Y(t-x)\, f_X(x)$

* Idea: prove existence of objects with desired properting A using prob.

Show $P(A) > 0$ for a random object

suppose each object has a "score". Show there is an object with a good score. thm: there is an object whose score is at least $E[X]$

$\rightarrow$ Shannon, information theory $\rightarrow$ He showed a good code exist. The way he did it was to pick a random code. Probably took many to find one but he ~~didn't~~ couldn't, so picked a random

Ex. 100 people, 15 committees of 20; each person is on 3 committees. Show there exists 2 committees with overlap $\geq 3$