

STAT 110: Probability

Lecture Note

Harvard University

Last Edit Date: 01/10/2023

Disclaimer and Term of Use:

1. We do not guarantee the accuracy and completeness of the summary content. Some of the course material may not be included, and some of the content in the summary may not be correct. You should use this file properly and legally. We are not responsible for any results from using this file.
2. Although most of the content in this summary is originally written by the creator, there may be still some of the content that is adapted (derived) from the slides and codes from *Professor Joe Blitzstein*. We use those as references and quotes in this file. Please [contact us](#) to delete this file if you think your rights have been violated.
3. This work is licensed under a [Creative Commons Attribution 4.0 International License](#).

Content:

Lecture 1: Sample spaces, naive definition of probability, counting, sampling	3
Lecture 2: Bose-Einstein, story proofs, Vandermonde identity, axioms of probability.....	3
Lecture 3: birthday problem, properties of probability, inclusion-exclusion, matching problem.....	4
Lecture 4: Independence, Newton-Pepys, conditional probability, Bayes' rule.....	5
Lecture 5: Law of total probability, conditional probability examples, conditional independence.....	6
Lecture 6: Monty Hall problem, Simpson's paradox...	7
Lecture 7: Gambler's ruin, first step analysis, random variables, Bernoulli, Binomial.....	9
Lecture 8: Random variables, CDFs, PMFs, Hypergeometric.....	10
Lecture 9: Independence, Geometric, expected values, indicator r.v.s, linearity, symmetry, fundamental bridge.....	11
Lecture 10: Linearity, Putnam problem, Negative Binomial, St. Petersburg paradox.....	12
Lecture 11: Sympathetic magic, Poisson distribution, Poisson approximation.....	13
Lecture 12: Discrete vs. continuous, PDFs, variance, standard deviation, Uniform, universality.....	14
Lecture 13: Standard Normal, Normal normalizing constant.....	16
Lecture 14: Normal distribution, standardization, LOTUS.....	17
Lecture 15: Exponential distribution, memoryless property.....	18
Lecture 16: Moment generating functions (MGFs), hybrid Bayes' rule, Laplace's rule of succession.....	19
Lecture 17: MGFs to get moments of Expo and Normal, sums of Poissons, joint distributions.....	20
Lecture 18: Joint, conditional, and marginal distributions, 2-D LOTUS, chicken-egg.....	21
Lecture 19: Expected distance between Normals, Multinomial, Cauchy.....	22
Lecture 20: Covariance, correlation, variance of a sum, variance of Hypergeometric.....	23
Lecture 21: Transformations, LogNormal, convolutions, the probabilistic method.....	24
Lecture 22: Beta distribution, Bayes' billiards, finance preview and examples.....	25
Lecture 23: Gamma distribution, Poisson processes..	26
Lecture 24: Beta-Gamma (bank-post office), order statistics, conditional expectation, two envelope paradox	27
Lecture 25: Two envelope paradox (cont.), conditional expectation (cont.), waiting for HT vs. waiting for HH.....	28
Lecture 26: Conditional expectation (cont.), taking out what's known, Adam's law, Eve's law.....	29
Lecture 27: Sum of a random number of random variables, inequalities (Cauchy-Schwarz, Jensen, Markov, Chebyshev).....	30
Lecture 28: Law of large numbers, central limit theorem.....	31
Lecture 29: Chi-Square, Student-t, Multivariate Normal	
Lecture 30: Markov chains, transition matrix, stationary distribution.....	33
Lecture 31: Markov chains (cont.), irreducibility, reversibility, random walk on an undirected network.	34
Lecture 32: Markov chains (cont.), Google PageRank as a Markov chain.....	35

Lecture 1

A sample space is the set of all possible outcomes of an experiment.

An event is a subset of a sample space.



Naive definition of Probability: $P(A) = \frac{\text{# favorable outcomes}}{\text{# possible outcomes}}$

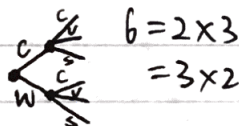
ex: Flip a coin twice: HH, HT, TH, TT. $P(TT) = \frac{1}{4}$

Assumption: All outcomes are equally likely, finite sample space.

Counting

1. Multiplication Rule: If have an experiment with n_1 possible outcomes, and for each outcome of 1st experiment, there are n_2 outcomes for 2nd experiment, ..., for each there are n_n outcomes for the n th experiment, the overall there are $n_1 \cdot n_2 \cdot \dots \cdot n_n$ possible outcomes.

ex: ice cream



ex: probability of full house in poker, 5 card hand

$$\binom{13}{1} \cdot \binom{4}{3} \cdot \binom{12}{1} \cdot \binom{4}{2}$$

$\binom{52}{5} \rightarrow$ Binomial coefficient $\binom{n}{k} = \frac{n!}{(n-k)!k!}$, 0 if $k > n$
subsets of k , of group of n people.

2. Combination

Binomial coefficient $\binom{n}{k} = \frac{n!}{k!(n-k)!}$, 0 if $k > n$

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

Sampling table: Choosing k objects out of n

	order	no order
replace	n^k	$\binom{n+k-1}{k}$
don't replace	$\frac{n(n-1) \dots (n-k+1)}{k!}$	$\binom{n}{k}$

↓ permutation combination

Lecture 2

ex: 10 people, split into team of 6 and team of 4 $\Rightarrow \binom{10}{4} = \binom{10}{6}$

two teams of 5 $\Rightarrow \binom{10}{5} / 2$

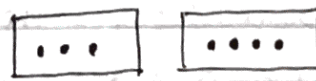
Pick k times from set of n objects, where order does not matter, with replacement
 $\binom{n+k-1}{k}$ ways.

Extreme cases: $k=0 \Rightarrow \binom{n-1}{0} = 1$

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

simplest
non-trivial
example

$$\rightarrow n=2 \Rightarrow \binom{k+1}{k} = \binom{k+1}{1} = k+1$$



dots is in
 $\{0, 1, \dots, k\}$

Equivalently, how many ways are there

to put k indistinguishable into n distinguishable boxes?



$n=4$
boxes

$k=4$
particles

all positions
 $\leftarrow k \text{ 's' }, n-1 \text{ |s|}$
 $\Rightarrow \binom{n+k-1}{k}$
 $k \text{ 's' positions}$

Story proof: Proof by interpretation

ex: $\binom{n}{k} = \binom{n}{n-k}$

ex: $n \binom{n-1}{k-1} = k \binom{n}{k}$, pick k people out of n , with 1 designated as President.
(Count the same thing in two ways)

ex: $\binom{m+n}{k} = \sum_{j=0}^k \binom{m}{j} \binom{n}{k-j}$ (Vandermonde)



Non-naive Definition: A probability sample consists of S and P , where S is a sample space, and P , a function which takes an event $A \subseteq S$ as input, return $P(A) \in [0, 1]$ as outputs such that:

(1) $P(\emptyset) = 0$, $P(S) = 1$

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

Lecture 3

Birthday Problem: k people, find probability that 2 have same birthday. Exclude Feb 29, assume the other 365 days are equally likely, assume independence of birth.

If $k > 365$, then probability is 1 (pigeonhole principle)

Let $k \leq 365$

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

$$P(\text{match}) \approx \begin{cases} 50.7\%, & \text{if } k=23 \\ 97\%, & \text{if } k=50 \\ >99.99\%, & \text{if } k=100 \end{cases} \quad \binom{k}{2} = \frac{k(k-1)}{2}, \quad \binom{23}{2} = \frac{23 \cdot 22}{2} = 253$$

$\uparrow$
 k choose a pair

(1) $P(\emptyset) = 0$, $P(S) = 1$; (2) $P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$ if $A_1, A_2, \dots, A_n$ are disjoint events



Properties

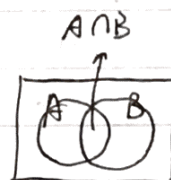
(1) $P(A^c) = 1 - P(A)$

Proof: $1 = P(S) = P(A \cup A^c) = P(A) + P(A^c)$
 Since $A \cap A^c = \emptyset$

(2) If $A \subseteq B$, then $P(A) \leq P(B)$

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

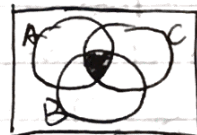
$$P(B) = P(A) + P(B \cap A^c) \geq P(A)$$



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

Proof: $P(A \cup B) = P(A \cup (A^c \cap B)) = P(A) + \overbrace{P(B \cap A^c)}^{(\text{disjoint})} = P(A) + P(B) - P(A \cap B)$
 $\Rightarrow P(A \cap B) + P(B \cap A^c) = P(B)$ True

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



since $P(A \cap B)$, $P(B \cap A^c)$ are disjoint
 Union is B

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

de Montmort's Problem (1713), matching problem.

n cards labeled $1, 2, \dots, n$. Let's call A_j be the event " j -th card matched".

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

$P(A_j) = \frac{1}{n}$ since all positions are equally likely for card labeled j

$$P(A_1 \cap A_2) = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)}$$

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

$$P(\text{no match}) = P\left(\bigcap_{j=1}^n A_j^c\right) = 1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \dots + (-1)^n \frac{1}{n!} \approx \left|1 - \frac{1}{e}\right|$$

$$\approx \frac{1}{e}$$

Lecture 4

Events A, B are independent if $P(A \cap B) = P(A) \cdot P(B)$.

Note: Completely different from disjointness.

A, B, C are independent if $P(A, B) = P(A)P(B)$, $P(A, C) = P(A)P(C)$, $P(B, C) = P(B)P(C)$,

$$P(A, B, C) = P(A)P(B)P(C).$$

Similarly, for events $A_1, \dots, A_n$ "independence means multiply".

ex: Newton-Pepys Problem: Have fair dice: which is the most likely?

(A) at least one 6 with 6 dice $\leftarrow$ truth.

(B) at least two 6's with 12 dice

(C) at least three 6's with 18 dice $\leftarrow$ Pepys's belief

$$P(A) = 1 - \left(\frac{5}{6}\right)^6, \quad P(B) = 1 - \left(\frac{5}{6}\right)^{12} - \binom{12}{1} \left(\frac{1}{6}\right) \left(\frac{5}{6}\right)^{11}, \quad P(C) = 1 - \left(\frac{5}{6}\right)^{18} - \binom{18}{1} \left(\frac{1}{6}\right) \left(\frac{5}{6}\right)^{17} - \dots$$

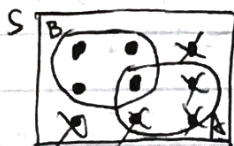
$$\approx 0.665 \quad \approx 0.619 \quad \approx 0.597 \text{ Binomial probability}$$

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

Conditioning is the soul of statistics.

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

Intuition 1: Pebble world

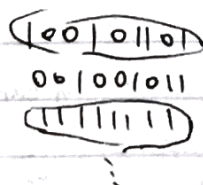


9 pebbles, total mass 1

$P(A|B)$: get rid of pebbles not in B

renormalize to make total mass 1 again.

Intuition 2: Frequentist world: Repeat experiment many times



Circle repetition where B occurs. Among those, what fraction of time did A occur?

Theorem 1: $P(A \cap B) = P(B)P(A|B) = P(A)P(B|A)$

Theorem 2: $P(A_1, \dots, A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1, A_2) \dots P(A_n|A_1, \dots, A_{n-1})$
[n! theorem]

Theorem 3: $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$ [Bayes' Rule]

Lecture 5

Thinking conditionally is a condition for thinking

How to solve a problem:

(1) Try simple and extreme cases.

(2) Break up problems into simpler pieces.



Let $A_1, \dots, A_n$ partition of S

Then $P(B) = P(B \cap A_1) + \dots + P(B \cap A_n)$

$$= P(B|A_1) \cdot P(A_1) + \dots + P(B|A_n) \cdot P(A_n) \quad \left. \begin{array}{l} \text{Law of total} \\ \text{probability} \end{array} \right\}$$

ex: Get random 2-card hand from standard deck.

Find $P(\text{both cards A} \mid \text{have A})$, $P(\text{both A} \mid \text{have A of spade})$

$$\begin{aligned} &= \frac{P(\text{both A} \cap \text{have A})}{P(\text{have A})} = \frac{P(\text{both A} \cap \text{A of spade})}{P(\text{A of spade})} = \frac{\frac{3}{51} \cdot \frac{1}{50}}{\frac{1}{13}} = \frac{3}{51} \\ &= \frac{\binom{4}{2} / \binom{52}{2}}{1 - P(\text{no A})} = \frac{\binom{4}{2} / \binom{52}{2}}{1 - \binom{48}{2} / \binom{52}{2}} = \frac{1}{17} = \frac{3}{51} \quad (\text{by symmetry}) \end{aligned}$$

ex: Patient gets tested for disease afflicts 1% of population, tests positive. Suppose test as advertised as "95%" accurate, suppose this means $P(T|D) = 0.95 = P(T^c|D^c)$. Find $P(D|T)$.

$$\begin{aligned} D: \text{patient has disease} \quad T: \text{patient tests positive} \\ P(D|T) &= \frac{P(T|D) \cdot P(D)}{P(T)} = \frac{P(T|D) \cdot P(D)}{P(T|D) \cdot P(D) + P(T|D^c) \cdot P(D^c)} = \frac{0.95 \cdot 0.01}{0.95 \cdot 0.01 + 0.05 \cdot 0.99} \\ &= 0.16 \end{aligned}$$

Biohazards

(1) Confusing $P(A|B)$, $P(B|A)$ ("prosecutor's fallacy")

ex: Sally Clark case, SIDS. $\frac{1}{8500} \cdot \frac{1}{8500} \approx \frac{1}{73 \times 10^5}$
Assume both independent

Want $P(\text{innocence} | \text{evidence})$

(2) confusing $P(A)$ "prior" with $P(A|B)$ "posterior". $P(A|A) = 1$

(3) confusing independence with conditional independence

Events A, B are conditionally independent given C if $P(A \cap B | C) = P(A|C) \cdot P(B|C)$.

Does conditional independence imply unconditional independence? NO

ex: chess opponent of unknown strength.

given strength of opponent.

May be that the game outcomes are conditionally independent but not independent.

Does unconditional independence imply conditional independence? NO unconditionally

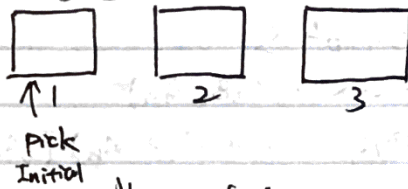
ex: A : fire alarm goes off

caused by: F : fire

C : popcorn

Suppose F, C independent. But $P(F|A, C^c) = 1$ not conditionally independent

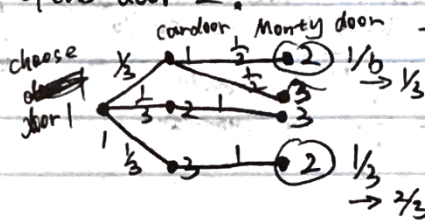
Lecture 6

Monty Hall

1 door has car, 2 doors have goats. Monty knows which. Monty always opens a goat door, if he has a choice, he picks with equal probabilities. Should you switch?

Note: if Monty opens door 2, we know door 2 has a goat, and Monty

opens door 2.



Method 1: Tree diagram

$$P(\text{success if switch} | \text{Monty opens door 2}) = \frac{2}{3}$$

Method 2: LOTP: We wish we know where the car is.

S: succeed (assuming switch)

D_j : Door j has the car ($j \in \{1, 2, 3\}$)

$$\begin{aligned} P(S) &= P(S|D_1) \cdot P(D_1) + P(S|D_2) \cdot P(D_2) + P(S|D_3) \cdot P(D_3) \\ &= P(S|D_1) \cdot \frac{1}{3} + P(S|D_2) \cdot \frac{1}{3} + P(S|D_3) \cdot \frac{1}{3} \\ &= 0 + 1 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} = \frac{2}{3} \end{aligned}$$

By symmetry $P(S | \text{Monty opens door 2}) = \frac{2}{3}$

Simpson's Paradox

	Hibert	bandaid
Success	70	10
Failure	20	0

	heart	bandaid
Success	2	81
Failure	8	9

Hibert: 80%
Nick: 83% } unconditionally

$$\frac{1}{3} + \frac{2}{3} \neq \frac{3}{8}$$

A: successful surgery B: treated by Dr. Nick C: heart surgery

$$P(A|B, C) \leq P(A|B^c, C)$$

$$P(A|B, C^c) \leq P(A|B^c, C^c)$$

but $P(A|B) \geq P(A|B^c)$

"C is a confounder"

WRONG PROOF:

$$P(A|B) = \underbrace{P(A|B, C)}_{< P(A|B^c, C)} \cdot \underbrace{P(C|B)}_{\text{weights}} + \underbrace{P(A|B, C^c)}_{< P(A|B^c, C^c)} \cdot \underbrace{P(C^c|B)}_{\text{weights}}$$



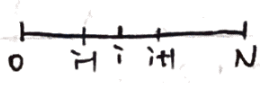
Lecture 7

(1) Conditioning: the soul of statistics

(2) Random variables and distributions

Gamblers Ruin

Two gamblers A, B with the sequence of rounds bet \$1, $p = P(A \text{ wins a certain round})$, $q = 1 - p$. Find probability that A wins the entire game (so B ruined), assuming that A starts with \$i and B starts with \$(N-i).

Random walk:  $p = \text{prob. moving right}$
absorbing states at 0, N

Strategy: condition on first step. Let $P_i = P(A \text{ wins the game} | A \text{ starts } \$i)$

$$P_i = pP_{i+1} + qP_{i-1}, \quad 1 \leq i \leq N-1 \quad \begin{cases} P_0 = 0 \\ P_N = 1 \end{cases}$$

(difference equation)

$$P_i = pP_{i+1} + qP_{i-1}$$

$$1 - 4pq = 1 - 4p(1-p) = 4p^2 - 4p + 1 = (2p-1)^2$$

Guess $P_i = x^i$

$$x^i = px^{i+1} + qx^{i-1}$$

$$Px^2 - x + q = 0$$

$$x = \frac{1 \pm \sqrt{1-4pq}}{2p} \in \{1, \frac{2-2p}{2p} = \frac{q}{p}\}$$

$$P_i = A \cdot 1^i + B \left(\frac{q}{p}\right)^i, \quad p \neq q$$

$$x = q/p$$

$$\lim_{x \rightarrow 1} \frac{1-x^i}{1-x^N} = \lim_{x \rightarrow 1} \frac{i x^{i-1}}{N x^{N-1}} = \frac{i}{N}$$

$$P_0 = 0 \Rightarrow B = -A, \quad P_N = 1 \Rightarrow 1 = A \left(1 - \frac{q}{p}\right)^N \Rightarrow P_i = \begin{cases} \frac{1 - \left(\frac{q}{p}\right)^i}{1 - \left(\frac{q}{p}\right)^N}, & \text{if } p \neq q \\ \frac{i}{N}, & \text{if } p = q \end{cases}$$

$$i = N-i, \quad p = 0.49$$

$$N=20 \Rightarrow 0.4 \text{ A wins}$$

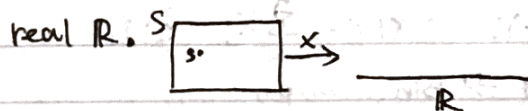
$$N=100 \Rightarrow 0.12 \text{ A wins}$$

$$N=200 \Rightarrow 0.02 \text{ A wins}$$

$$\frac{i}{N} + \frac{N-i}{N} = 1$$

$\Rightarrow$ prob. is 0 of game going on forever.

What is a random variable? It is a function from the sample space S to the



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

Bernoulli(p)

A random variable X is said to have Bern distribution ^(p) if X has only 2 possible values, 0 and 1, and if $P(X=1) = p$, $P(X=0) = 1-p$
event $\{s : X(s) = 1\}$

Binomial(np)

The distribution of # success ^(X) in n independent Bern(p) trials is called Bin(n,p), its distribution is given by $P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$ PMF (prob mass function)
 $0 \leq k \leq n$, all integers

$X \sim \text{Bin}(n, p)$, $Y \sim \text{Bin}(m, p)$ independent, then $X+Y \sim \text{Bin}(n+m, p)$.

Proof: consider n trials, m trials (both Bernoulli distributions).

Lecture 8

Binomial distribution $\text{Bin}(n, p)$ $X \sim \text{Bin}(n, p)$

prob. of success.

① Story: X is # successes in n independent Bern(p) trials

② Indicators sum r.v.s: $X = X_1 + X_2 + \dots + X_n$, $X_j = \begin{cases} 1, & \text{if } j\text{-th trial success} \\ 0, & \text{otherwise} \end{cases}$

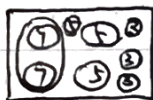
$X_1, \dots, X_n$ i.i.d. Bern(p)

independent & identically distributed.

③ PMF: $P(X=k) = \binom{n}{k} p^k q^{n-k}$, $q=1-p$, $k \in \{0, 1, \dots, n\}$ (Binomial theorem)

R.V.s

S $X=7$ is an event



CDF

$X \leq x$ is an event

$P(x) = P(X \leq x)$, then $F(x)$ is called CDF of x (cumulative distribution function).

PMF (for discrete r.v.s)

Discrete: Possible values $a_1, a_2, a_3, \dots, a_n$ or $a_1, a_2, \dots$

PMF is $P(X=a_j)$ for all j . $P_j \geq 0, \sum_j P_j = 1$

$X \sim \text{Bin}(n, p)$, $Y \sim \text{Bin}(m, p)$ indep. $\Rightarrow X+Y \sim \text{Bin}(n+m, p)$

① Immediate from story

② $X = X_1 + \dots + X_n$, $Y = Y_1 + \dots + Y_m \Rightarrow X+Y = \sum_{j=1}^n X_j + \sum_{i=1}^m Y_i$

Sum of $n+m$ i.i.d. Bern(p) $\Rightarrow \text{Bin}(n+m, p)$

③ $P(X+Y=k) = \sum_{j=0}^k P(X+Y=k | X=j) \cdot P(X=j)$

$$= \sum_{j=0}^k P(Y=k-j | X=j) \cdot P(X=j)$$

$$= \sum_{j=0}^k P(Y=k-j | \text{indep.}) \cdot \binom{n}{j} p^j q^{n-j}$$

$$= \sum_{j=0}^k \binom{m}{k-j} p^{k-j} q^{m-k+j} \binom{n}{j} p^j q^{n-j} = p^k \cdot q^{m+n-k} \left(\sum_{j=0}^k \binom{m}{k-j} \binom{n}{j} \right)$$

Vandermonde

PMF (or CDF)

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

ex: 5 card hand, find distribution of # aces. Let $X = (\# \text{ aces})$

$P(X=k)$ is 0 except if $k \in \{0, 1, 2, 3, 4\}$. Not Binomial, since trials are not indep.

$$P(X=k) = \frac{\binom{4}{k} \binom{48}{5-k}}{\binom{52}{5}} \quad \text{for } k \in \{0, 1, \dots, 4\}$$

aces non-aces

Have b black, w white marbles. Pick simple random sample of size n . Find dist. of (# white marble in the sample) X .

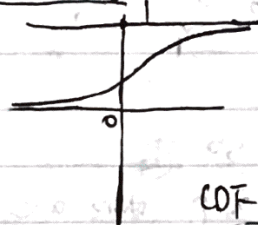
$$P(X=k) = \frac{\binom{w}{k} \binom{b}{n-k}}{\binom{w+b}{n}}, \quad 0 \leq k \leq w, \quad 0 \leq n-k \leq b. \quad \text{Hypergeometric, sampling without replacement}$$

Vandermonde

$$\sum_{k=0}^w \frac{\binom{w}{k} \binom{b}{n-k}}{\binom{w+b}{n}} = \frac{1}{\binom{w+b}{n}} \sum_{k=0}^w \binom{w}{k} \binom{b}{n-k} = 1$$

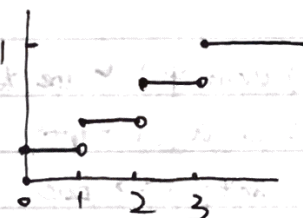
CDF $P(X \leq x)$

Continuous



CDF

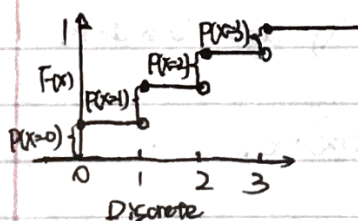
Discrete



PMF

Lecture 9

CDF: $F(x) = P(X \leq x)$ as a function of real x



$P(1 < X \leq 3)$ using CDF

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

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

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

Properties of CDF

① Increasing

② Right continuous

③ $F(x) \rightarrow 0$ as $x \rightarrow -\infty$; $F(x) \rightarrow 1$ as $x \rightarrow \infty$

This is "if and only if"

Independence

X, Y are indep. r.v.s if $P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y)$ for all x, y .

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

Averages (Means, Expected values)

$$1, 2, 3, 4, 5, 6 \rightarrow \frac{1+2+3+4+5+6}{6} = 3.5$$

$$\left[\frac{1}{n} \sum_{j=1}^n j = \frac{n+1}{2} \right] \text{ arithmetic series.}$$

(1, 1, 1, 1, 1) (3, 3) (5)

two ways: add, divide by 8, or $\frac{5}{8} \cdot 1 + \frac{2}{8} \cdot 3 + \frac{1}{8} \cdot 5$

Average of a discrete r.v. X

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

$X \sim \text{Bern}(p)$

$$E(X) = 1 \cdot P(X=1) + 0 \cdot P(X=0) = p$$

$$X = \begin{cases} 1, & \text{if } A \text{ occurs} \\ 0, & \text{otherwise} \end{cases}$$

(indicator r.v.)

$$E(X) = P(A)$$

Fundamental bridge.

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

$$E(X) = \sum_{k=0}^n k \binom{n}{k} p^k q^{n-k}$$

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

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

$$= np \underbrace{\sum_{j=0}^{n-1} \binom{n-1}{j} p^j q^{n-1-j}}_1 = np$$

Linearity

$E(X+Y) = E(X) + E(Y)$ even if X, Y are dependent.

$E(cX) = cE(X)$ if c is a constant.

so $\frac{np}{n}$

$X \sim \text{Bin}(n, p)$ using linearity $\Rightarrow$ Each k it is a p , there are n Bern(p)

ex: 5 card hand, $X = (\# \text{ aces})$, let X_j be indicator of j -th card being an Ace.

$$E(X) = E(X_1 + \dots + X_5) = E(X_1) + \dots + E(X_5) = 5E(X_1)$$

indicators

linearity

symmetry

$1 \leq j \leq 5$

$$= 5P(\text{1st card ace}) = \frac{5}{13}$$

fundamental bridge.

even though X_j are dependent

This gives expected value of any Hypergeometric.

Geometric Distribution $\text{Geom}(p)$

Independent Bern(p) trials, count # failures before 1st success. Let $X \sim \text{Geom}(p)$.

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

Valid since $\sum_{k=0}^{\infty} p q^k = \frac{p}{1-q} = 1$

$$E(X) = \sum_{k=0}^{\infty} k p q^k = p \sum_{k=1}^{\infty} k q^k = \frac{pq}{p^2} = \frac{q}{p}$$

Story proof

Let $c = E(X)$

$$c = 0 \cdot p + (1+c)q$$

$$= q + cq$$

$$\Rightarrow q/p$$

$$\begin{cases} \sum_{k=1}^{\infty} q^k = \frac{1}{1-q} \\ \sum_{k=1}^{\infty} k q^{k-1} = \frac{1}{(1-q)^2} \\ \sum_{k=1}^{\infty} k q^k = \frac{q}{1-q} \end{cases}$$

Lecture 10

Let $T = X + Y$, show $E(T) = E(X) + E(Y)$.

$$E(T) = \sum_t t P(T=t) \stackrel{?}{=} \sum_x x P(X=x) + \sum_y y P(Y=y)$$

$$P(T=t) = \sum_x P(T=t | X=x) \cdot P(X=x)$$

S

0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0

$x=0$ $x=1$ $x=2$ $x=3$

$$E(X) = \sum_x x P(X=x) = \underbrace{\sum_x x P(X=x)}_{\text{grouped}} = \underbrace{\sum_s X(s) \cdot P(\{s\})}_{\text{ungrouped}}$$

Proof of linearity (discrete case)

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

Similarly, $E(cX) = cE(X)$ if c is a constant.

If $X=Y$, then $E(X+Y) = E(2X) = 2E(X) = E(X) + E(Y)$

Negative Binomial parameters r, p

Story: Independent $\text{Bern}(p)$ trials, # of failures before the r th success.

$$\text{PMF: } P(X=n) = \binom{n+r-1}{r-1} p^r (1-p)^n, \text{ for } n=0, 1, 2, \dots$$

$$E(X) = E(X_1 + \dots + X_r) = E(X_1) + \dots + E(X_r) = \boxed{\frac{r}{p}}$$

X_j is # failures between $(j-1)$ st success and j th success, $X_j \sim \text{Geom}(p)$

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

$$\text{Let } Y = X - 1, \text{ Then } Y \sim \text{Geom}(p), E(X) = E(Y) + 1 = \frac{q}{p} + 1 = \boxed{\frac{1}{p}}$$

Putnam

Random permutation of $1, 2, \dots, n$, where $n \geq 2$. Find the expected # of local maxima. $(3) 214 (7) 56$

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

$$E(I_1) + \dots + E(I_n) = \frac{n-2}{3} + \frac{2}{2} = \boxed{\frac{n+1}{3}}$$

St. Petersburg Paradox

Get $\$2^x$, where x is # of flips of a coin until it lands first H, including the success.

$Y = 2^x$, find $E(Y)$

$$E(Y) = \sum_{k=1}^{\infty} 2^k \cdot \frac{1}{2^k} = \sum_{k=1}^{\infty} 1 = 1+1+1+\dots = \infty$$

Bound at $\$2^{40}$. Then $\sum_{k=1}^{40} 2^k \cdot \frac{1}{2^k} = 40$

$$\infty = E(2^x) \neq 2^{E(x)} = 4$$

Lecture 11

Sympathetic magic

Don't confuse a r.v. with its distribution.

$$P(X=x) + P(Y=y) \neq X$$

"Word is not the thing, the map is not the territory."

r.v. $\leftrightarrow$ house, distribution $\leftrightarrow$ blueprint

Poisson Distribution $X \sim \text{Pois}(\lambda)$

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

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

$$\text{Valid: } \sum_{k=0}^{\infty} e^{-\lambda} \lambda^k / k! = e^{-\lambda} e^{\lambda} = 1$$

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

Often used for applications where counting # of "successes", where we have a large number of trials, each with small prob. of "success".

① # emails in an hour

② # chips in chocolate chip cookie.

③ # earthquakes in a year in some regions.

Poisson Paradigm (Poisson approximation)

Events $A_1, A_2, \dots, A_n$. $P(A_j) = p_j$, n large. p_j 's small, events indep. or "weakly" dependent, then # of A_j 's that occur is approximately $\text{Pois}(\lambda)$, $\lambda = \sum_{j=1}^n p_j$.

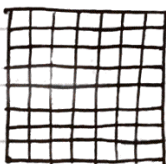
$X \sim \text{Bin}(n, p)$, let $n \rightarrow \infty$, $p \rightarrow 0$, $\lambda = np$ is held constant. $\rightarrow p = \lambda/n$ $(1 + \frac{x}{n})^n \rightarrow e^x$

Find what happens to $P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$, k fixed.

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

$$\rightarrow \frac{\lambda^k}{k!} e^{-\lambda}, \text{ Poisson PMF at } k. \rightarrow 1$$

ex: raindrops



ex: Have n people, find approximately prob. that there are 3 people with the same birthday.

$\binom{n}{3}$ triplets of people, indicator r.v. for each, I_{ijk} , $i < j < k$.

$$E(\# \text{ triple matches}) = \binom{n}{3} \frac{1}{365^2}$$

$X = \# \text{ triple matches}$ approximately $\text{Pois}(\lambda)$, $\lambda = \binom{n}{3} \cdot \frac{1}{365^2}$

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

Lecture 12

Discrete

X

PMF: $P(X=x)$

Continuous

X

PDF: $f_X(x) = F'_X(x)$
[$P(X=x) = 0$]

Discrete

CDF: $F_X(x) = P(X \leq x)$

$E(X) = \sum_x x P(X=x), \text{Var}(X) = E(X^2) - (E(X))^2$

Continuous

CDF: $F_X(x) = P(X \leq x)$

$E(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx, \text{Var}(X) = E(X^2) - (E(X))^2$

PDF (probability density function)Def: R.V. X has PDF $f(x)$ if $P(a \leq X \leq b) = \int_a^b f(x) dx$ for all $a, b, [a=b \Rightarrow \int_a^a f(x) dx = 0]$ To be valid, $f(x) \geq 0, \int_{-\infty}^{\infty} f(x) dx = 1$.

$f(x_0) \cdot \varepsilon \approx P(X \in (x_0 - \varepsilon/2, x_0 + \varepsilon/2))$ for ε very small but positive

If X has PDF f , the CDF is $F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt$. fundamental theorem of calculus.If X has CDF F (and X is continuous R.V.), then $f(x) = F'(x)$ by FTC.

$P(a < X < b) = \int_a^b f(x) dx = F(b) - F(a)$ by FTC.

Variance

$$\text{Var}(X) = E(X - E(X))^2 \quad \text{Standard Deviation: } SD(X) = \sqrt{\text{Var}(X)}$$

$$= E(X^2) - (E(X))^2$$

Uniform Distribution $\text{Unif}(a, b)$

"completely random point in $[a, b]$ "

Unit: Prob & length.

$$f(x) = \begin{cases} c, & \text{if } a \leq x \leq b \\ 0, & \text{otherwise} \end{cases} \Rightarrow 1 = \int_a^b c dx = c = \frac{1}{b-a}$$

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

$$E(X) = \int_a^b \frac{x}{b-a} dx = \left[\frac{x^2}{2(b-a)} \right]_a^b = \frac{1}{2(b-a)} (b-a)(b+a) = (a+b)/2$$

$E(X^2) = E(Y) = ?$ need PDF of Y

$$= \int_{-\infty}^{\infty} x^2 f(x) dx \quad (\text{Law of the unconscious of Statistician}) \quad \boxed{\text{LOTUS}}$$

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) f(x) dx \rightarrow \text{continuous}$$

$$E(g(X)) = \sum_x g(x) P(X=x) \rightarrow \text{discrete}$$

Let $U \sim \text{Unif}(0, 1)$, $E(U) = 1/2$, $E(U^2) = \int_0^1 u^2 f_u(u) du$

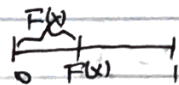
$$\text{Var}(U) = E(U^2) - (E(U))^2$$

$$= \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$$

$$= \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$$

Uniform is UniversalLet $U \sim \text{Uniform}(0, 1)$, F be a CDF (assume F is strictly increasing, continuous).

Theorem: Let $X = F^{-1}(U)$. Then $X \sim F$.



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

Lecture 13

Universality of Unif

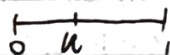
Let F be a continuous, strictly increasing CDF. Then $X = F^{-1}(U) \sim F$ if $U \sim \text{Unif}(0, 1)$.

Also: if $X \sim F$, then $F(X) \sim \text{Unif}(0, 1)$

$$F(x) = P(X \leq x), \quad \overline{F(x)} = P(X > x) = 1$$

ex: Let $F(x) = 1 - e^{-x}$, $x > 0$. (Exp(1)), $U \sim \text{Unif}(0, 1)$, Simulate $X \sim F$, $F^{-1}(U) = -\ln(1-U)$

$1-U \sim \text{Unif}(0, 1)$ (symmetry of uniform)



$a+bu$ is unit on some interval.

non-linear usually leads to non-uniform.

Indep. of R.V.s

Joint CDF

$X_1, \dots, X_n$ indep. if $P(X_1 \leq x_1, \dots, X_n \leq x_n) = P(X_1 \leq x_1) \cdot \dots \cdot P(X_n \leq x_n)$, for all $x_1, \dots, x_n$.

Joint PMF

Discrete case: $P(X_1 = x_1, \dots, X_n = x_n) = P(X_1 = x_1) \cdot \dots \cdot P(X_n = x_n)$

ex: $X_1, X_2 \sim \text{Bern}(\frac{1}{2})$ i.i.d., $X_3 = \begin{cases} 1, & \text{if } X_1 = X_2 \\ 0, & \text{otherwise} \end{cases}$

These are pairwise indep. but not indep.

sum of a lot of r.v.s, iid, looks normal

Normal Distribution (Central Limit Theorem)



$N(0, 1)$ has PDF: $f(z) = C \cdot e^{-z^2/2}$

normalizing function.

$$\int_{-\infty}^{\infty} e^{-z^2/2} dz \int_{-\infty}^{\infty} e^{-z^2/2} dz$$

$$= \int_{-\infty}^{\infty} e^{-x^2/2} dx \int_{-\infty}^{\infty} e^{-y^2/2} dy$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(x^2+y^2)/2} dx dy = \int_0^{2\pi} \int_0^{\infty} e^{-r^2/2} r dr d\theta = \int_0^{2\pi} \int_0^{\infty} e^{-u} du d\theta = 2\pi$$

$$C = \frac{1}{\sqrt{2\pi}}$$

$$\frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

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

Jacobian

$Z \sim N(0, 1)$

$E(Z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z e^{-z^2/2} dz = 0$ if $g(x)$ is an odd function, i.e. $g(-x) = -g(x)$

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

$\text{Var}(Z) = E(Z^2) - (E(Z))^2 = E(Z^2)$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^2 e^{-z^2/2} dz = \frac{2}{\sqrt{2\pi}} \int_0^{\infty} z^2 e^{-z^2/2} dz = \frac{2}{\sqrt{2\pi}} \left(\underbrace{[uv]_0^{\infty}}_0 + \int_0^{\infty} e^{-z^2/2} dz \right) = 1$$

even function

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

Notation: Φ is the standard Normal CDF

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-t^2/2} dt \quad \Phi(-z) = 1 - \Phi(z) \text{ by symmetry.}$$

Lecture 14

$$Z \sim N(0,1) \text{ , CDF: } \Phi, \overset{\text{first moment.}}{E(Z)=0}, \overset{\text{second moment}}{\text{Var}(Z)=E(Z^2)=1}, \overset{\text{third moment}}{E(Z^3)=0}$$

$$-Z \sim N(0,1) \text{ symmetry}$$

$$\int_{-\infty}^{\infty} z^3 \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz \quad \text{odd.}$$

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

Then we say $X \sim N(\mu, \sigma^2)$. $Z = \frac{X - \mu}{\sigma} \Rightarrow$ Standardization.

$$E(X) = \mu, \text{ Var}(X) = \sigma^2$$

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

$$\text{Var}(X+c) = \text{Var}(X)$$

$\text{Var}(X) \geq 0$. If $\text{Var}(X) = 0$, then $P(X=c) = 1$ always holds for some c .

$$\text{Var}(cX) = c^2 \text{Var}(X)$$

$\text{Var}(X+Y) \neq \text{Var}(X) + \text{Var}(Y)$ in general. It can be equal if X, Y indep.

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

$$\text{CDF: } P(X \leq x) = P\left(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right)$$

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

$$\text{PDF: } \frac{1}{\sigma} \cdot \frac{1}{\sqrt{2\pi}} e^{-\left(\frac{x - \mu}{\sigma}\right)^2/2}$$

$-X = -\mu + \sigma(-Z) \sim N(-\mu, \sigma^2)$ If $X_j \sim N(\mu_j, \sigma_j^2)$, and indep. then $X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$
 $X_1 - X_2 \sim N(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$

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

$$P(|X - \mu| \leq \sigma) \approx 68\%, \quad P(|X - \mu| \leq 2\sigma) = 95\%, \quad P(|X - \mu| \leq 3\sigma) = 99.7\%$$

$$X: 0, 1, 2, 3, \dots \quad X^2: 0, 1, 4, 9, \dots$$

$$\text{Prob: } p_1, p_2, p_3, \dots \quad E(X) = \sum_x x P(X=x)$$

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

$X \sim \text{Pois}(\lambda)$

$$E(X^2) = \sum_{k=0}^{\infty} k^2 \frac{e^{-\lambda} \lambda^k}{k!}$$

$$= e^{-\lambda} \cdot \lambda(\lambda+1)$$

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

$$\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{\lambda}$$

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

$$\sum_{k=1}^{\infty} \frac{k \lambda^k}{k!} = \lambda e^{\lambda}$$

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

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

Var(X) of Bin(n, p)

$$X = I_1 + \dots + I_n, \quad I_j \text{ iid Bern}(p) \quad X^2 = I_1^2 + \dots + I_n^2 + 2I_1I_2 + 2I_1I_3 + \dots + 2I_nI_{n-1}$$

$$E(X^2) = n E(I_1^2) + 2 \binom{n}{2} E(I_1I_2) \quad \text{indicator of success on both 1, 2}$$

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

$$\text{Var}(X) = E(X^2) - (E(X))^2 = np + n^2p^2 - np^2 - n^2p^2 = np - np^2 = np(1-p) = \boxed{npq}$$

Prove LOTUS for discrete sample space

$$\text{Show } E(g(x)) = \sum_x g(x) P(X=x)$$

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

grouped $g(x)$ ungrouped

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

Lecture 5

Exponential Distribution, rate parameter λ

$$X \sim \text{Expo}(\lambda)$$

$$\text{PDF: } \lambda e^{-\lambda x}, x > 0 \text{ (otherwise)} \quad \text{valid: } \int_0^\infty \lambda e^{-\lambda x} dx = 1$$

$$\text{CDF: } F(x) = \int_0^x \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x}, x > 0$$

$$\text{Let } Y = \lambda X, \text{ then } Y \sim \text{Expo}(1). \text{ Since } P(Y \leq y) = P(X \leq \frac{y}{\lambda}) = 1 - e^{-y}$$

Let $Y \sim \text{Expo}(1)$, find $E(Y)$ and $\text{Var}(Y)$

$$E(Y) = \int_0^\infty y e^{-y} dy = \underbrace{[-y e^{-y}]_0^\infty}_{\substack{u=y, \\ du=dy, \\ v=e^{-y}, \\ dv=-e^{-y}}} + \underbrace{\int_0^\infty e^{-y} dy}_1 = \boxed{1}$$

$$\text{Var}(Y) = E(Y^2) - (E(Y))^2 = \int_0^\infty y^2 e^{-y} dy - 1 = \boxed{1}$$

General Exponential

$$Y = \lambda X, \quad X = \frac{Y}{\lambda} \text{ has } E(X) = \frac{1}{\lambda}, \quad \text{Var}(X) = \frac{1}{\lambda^2}$$

Memoryless Property

$$P(X \geq s+t \mid X \geq s) = P(X \geq t)$$

$$\text{Here } P(X \geq s) = 1 - P(X \leq s) = e^{-\lambda s}$$

$$P(X \geq s+t \mid X \geq s) = \frac{P(X \geq s+t \text{ and } X \geq s)}{P(X \geq s)} \Rightarrow \frac{P(X \geq s+t)}{P(X \geq s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t}$$

$$X \sim \text{Expo}(\lambda)$$

$$E(X \mid X > a) = a + E(X - a \mid X > a) = a + \frac{1}{\lambda} \text{ by memoryless property.}$$

Lecture 16

$$E(T | T > 20) > E(T)$$

If memoryless, we would have $E(T | T > 20) = 20 + E(T)$

Theorem: If X is a continuous, positive r.v. with memoryless property, then $X \sim \text{Exp}(\lambda)$ for some λ .

Proof: Let F be the CDF of X , $G(x) = P(X > x) = 1 - F(x)$. Memoryless property is $G(s+t) = G(s)G(t)$. Solve for G .

$$S=t \Rightarrow G(2t) = (G(t))^2, G(3t) = (G(t))^3, \dots, G(kt) = (G(t))^k$$

$$G(\frac{t}{2}) = [G(t)]^{\frac{1}{2}}, G(\frac{t}{3}) = [G(t)]^{\frac{1}{3}}, \dots, G(\frac{t}{k}) = [G(t)]^{\frac{1}{k}}$$

$$G(\frac{m}{n}t) = G(t)^{\frac{m}{n}}. \text{ So } G(xt) = (G(t))^x \text{ for any real } x > 0.$$

$$t=1 \Rightarrow G(x) = (G(1))^x = e^{x \ln G(1)} = e^{-\lambda x}$$

Moment Generating Function (MGF)

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

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.

Why MGF important?

Let X have MGF $M(t)$.

① The n -th moment, $E(X^n)$, is coefficient of $\frac{t^n}{n!}$ in Taylor series of M , and $M^{(n)}(0) = E(X^n)$.

② The MGF determines the distributions, if X, Y have the same MGF, then they have the same CDF.

③ If X has MGF M_X , Y has MGF M_Y , X, Y indep, then the MGF of $X+Y$ is $E(e^{t(X+Y)}) = E(e^{tx})E(e^{ty}) = M_X(t)M_Y(t)$.

ex: $X \sim \text{Bern}(p)$, $M(t) = E(e^{tx}) = pe^t + q$

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

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

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

Laplace Rule of Succession

$X_1, \dots, X_n, \dots$ i.i.d Bern(p), given p

p is unknown. Bayesian: p is treated as r.v. Let $p \sim \text{Uniform}(0, 1)$. Let $S_n = X_1 + \dots + X_n$

So $S_n | p \sim \text{Bin}(n, p)$, $p \sim \text{Unit}(0, 1)$.

Find posterior $p | S_n$, and $P(X_{n+1} = 1 | S_n = n)$

$$f(p | S_n = k) = \frac{P(S_n = k | p) f(p)}{P(S_n = k)} = p^k (1-p)^{n-k}$$

$P(S_n = k)$ prior, 1 does not depend on p

$$f(p | S_n = n) = (n+1)p^n$$

$$P(X_{n+1} = 1 | S_n = n) = \int_0^1 (n+1)p p^n dp = \frac{n+1}{n+2}$$

$$P(S_n = k) = \int_0^1 P(S_n = k | p) f(p) dp$$

Lecture 17

Expo MGF

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

$$M(t) = E(e^{tx}) = \int_0^\infty e^{tx} e^{-x} dx = \int_0^\infty e^{-x(1-t)} dx = \frac{1}{1-t}, \quad t < 1$$

$$M'(0) = E(X), \quad M''(0) = E(X^2), \quad M'''(0) = E(X^3), \dots$$

$$\frac{1}{1-t} = \sum_{n=0}^{\infty} t^n, \quad |t| < 1$$

$$= \sum_{n=0}^{\infty} n! \frac{t^n}{n!} \Rightarrow E(X^n) = n!$$

$$Y \sim \text{Expo}(\lambda), \text{ let } X = \lambda Y \sim \text{Exp}(1), \text{ so } Y^n = \frac{X^n}{\lambda^n} \Rightarrow E(Y^n) = \frac{n!}{\lambda^n}$$

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

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

$$\text{MGF } M(t) = e^{t^2/2} = \sum_{n=0}^{\infty} \frac{(t^2/2)^n}{n!} = \sum_{n=0}^{\infty} \frac{t^{2n}}{2^n n!} = \sum_{n=0}^{\infty} \frac{(2n)! t^{2n}}{2^n n! (2n)!} \Rightarrow E(Z^{2n}) = \frac{(2n)!}{2^n n!}$$

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

$$n=2 \Rightarrow E(Z^4) = 3$$

!

ex: $X \sim \text{Pois}(\lambda)$

$$E(e^{tx}) = \sum_{k=0}^{\infty} e^{tk} e^{-\lambda} \lambda^k / k! = e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t - 1)}$$

$Y \sim \text{Pois}(\mu)$, X and Y indep. Find distribution of $X+Y$. Multiply MGFs:
 $e^{\lambda(e^t - 1)} e^{\mu(e^t - 1)} = e^{(\lambda+\mu)(e^t - 1)} \Rightarrow X+Y \sim \text{Pois}(\lambda+\mu)$

ex: counter) If X, Y dependent $X=Y \Rightarrow X+Y=2X$ is not Poisson.

$$E(2X) = 2\lambda, \quad \text{Var}(2X) = 4\lambda$$

even, cannot be Poisson.

Joint Distribution

X, Y Bernoulli

This example is independent

	$Y=0$	$Y=1$	
$X=0$	$\frac{2}{6}$	$\frac{1}{6}$	$\frac{3}{6}$
$X=1$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{2}{6}$
	$\frac{3}{6}$	$\frac{2}{6}$	

X, Y r.v.s

Joint CDF, $F(x, y) = P(X \leq x, Y \leq y)$

Joint PMF, $P(X=x, Y=y)$ (discrete)

Marginal CDF, $P(X \leq x)$ is marginal of X

Joint PDF $f(x, y)$ such that $P((x, y) \in B) = \iint_B f(x, y) dx dy$

Getting marginals

$$P(X=x) = \sum_y P(X=x, Y=y) \text{ (discrete)}$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx \text{ (continuous)}$$

ex: Uniform on square $\{(x, y) : x, y \in [0, 1]\}$

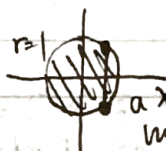


Joint PDF constant on that square, 0 outside.

$$\begin{cases} c, & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0, & \text{otherwise.} \end{cases}$$

Integral is area, $c = \frac{1}{\text{area}} = 1$. Marginally: X, Y are independent $\text{Unif}[0, 1]$.

ex: Unit in disk $x^2 + y^2 \leq 1$



a X can have multiple Y 's

Joint PDF: $\begin{cases} \frac{1}{\pi}, & x^2 + y^2 \leq 1 \\ 0, & \text{otherwise} \end{cases}$ X, Y Dependent. Given $X=x$, $-\sqrt{1-x^2} \leq Y \leq \sqrt{1-x^2}$

Lecture 18

Joint, conditional, marginal distribution

Joint CDF $F(x, y) = P(X \leq x, Y \leq y)$

Continuous case: Joint PDF: $f(x, y) = \frac{d^2}{dx dy} F(x, y)$

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

Marginal PDF of X : $\int_{-\infty}^{\infty} f(x, y) dy$ $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1$ always.

Conditional PDF of $Y|X$ is $f_{Y|X}(y|x) = \frac{f_{X,Y}(x, y)}{f_X(x)} = \frac{f_{X,Y}(x|y) f_Y(y)}{f_X(x)}$

X, Y are indep. if $f_{X,Y}(x, y) = f_X(x) f_Y(y)$ for all x, y .



Unif in disk $x^2 + y^2 \leq 1$

$$f(x, y) = \begin{cases} \frac{1}{\pi}, & x^2 + y^2 \leq 1 \\ 0, & \text{otherwise} \end{cases} \rightarrow \text{dependent } Y^2 \leq 1 - X^2$$

$$f_X(x) = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi} dy = \frac{2}{\pi} \sqrt{1-x^2}, \quad -1 \leq x \leq 1$$

$$f_{Y|X}(x) = \frac{\frac{1}{\pi}}{\frac{2}{\pi} \sqrt{1-x^2}} \text{ if } -\sqrt{1-x^2} \leq y \leq \sqrt{1-x^2}$$

$$Y|X \sim \text{Unif}(-\sqrt{1-x^2}, \sqrt{1-x^2})$$

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

$f_{X,Y}(x, y) \neq f_X(x) f_Y(y)$ so X, Y are dependent.

2-D LOTUS

Let (X, Y) have joint PDF $f(x, y)$ and let $g(x, y)$ be a real value of x, y .

Then $E(g(X,Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y) dx dy$.

Theorem: If X, Y are independent, then $E(XY) = E(X)E(Y)$. (Independent implies uncorrelated)

Proof: (continuous case)
$$E(XY) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_X(x) f_Y(y) dx dy$$

$$= \int_{-\infty}^{\infty} y f_Y(y) \underbrace{\int_{-\infty}^{\infty} x f_X(x) dx}_{E(X)} dy$$

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

ex: $X, Y \sim \text{Unif}(0,1)$, find $E(|X-Y|)$. LOTUS: $\int_0^1 \int_0^1 |x-y| dx dy$

$$= \iint_{x>y} (x-y) dx dy + \iint_{x<y} (y-x) dx dy$$

$$= 2 \iint_{x>y} (x-y) dx dy = 2 \int_0^1 \left(\frac{x^2}{2} - xy \right) \Big|_y^1 dy = 2 \int_0^1 \left(\frac{1}{2} - y - \frac{y^2}{2} + y^2 \right) dy = \boxed{\frac{1}{3}}$$

Let $M = \max(X,Y)$

$L = \min(X,Y)$

$|X-Y| = M-L$

$E(M-L) = \frac{1}{3}$

$E(M) - E(L) = \frac{1}{3}$

$E(M) = \frac{2}{3}, E(L) = \frac{1}{3} \Leftarrow E(M+L) = E(M) + E(L)$

Chick-egg N eggs $N \sim \text{Pois}(\lambda)$ eggs, each hatches with prob. p indep.

Let $X = \#$ that hatch, so $X|N \sim \text{Bin}(N, p)$. Let $Y = \#$ don't hatch.

So $[X+Y=N]$, Find joint PMF of X, Y .

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

$$= P(X=i, Y=j | N=i+j) P(N=i+j)$$

$$= \binom{i+j}{i} p^i q^j \cdot e^{-\lambda} \frac{\lambda^{i+j}}{(i+j)!} = \frac{p^i q^j}{i! j!} \lambda^{i+j} e^{-\lambda} = \left(e^{-\lambda p} \frac{(\lambda p)^i}{i!} \right) \left(e^{-\lambda q} \frac{(\lambda q)^j}{j!} \right)$$

$\Rightarrow X, Y$ are indep., $X \sim \text{Pois}(\lambda p)$, $Y \sim \text{Pois}(\lambda q)$.

Lecture 19

ex: Find $E(|Z_1 - Z_2|)$ with $Z_1, Z_2 \sim N(0,1)$

Theorem: $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$ independent. $X+Y \sim N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$.

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

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

Note $Z_1, -Z_2 \sim N(0,1)$. So $E(|Z_1 - Z_2|) = E(|Z_1|)$, $Z \sim N(0,1)$

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

even function

Multinomial

Def: $\text{Mult}(n, \vec{p})$, $\vec{p} = (p_1, \dots, p_k)$ is a prob. vector, $p_j \geq 0$, $\sum p_j = 1$. n objects in k categories.

$\vec{X} \sim \text{Mult}(n, \vec{p})$, $\vec{X} = (X_1, \dots, X_k)$ if $p_j = P(\text{category } j)$, $X_j = \#$ object in category j .

Joint PMF: $P(X_1=N_1, \dots, X_k=N_k) = \frac{n!}{n_1! n_2! \dots n_k!} p_1^{n_1} p_2^{n_2} \dots p_k^{n_k}$
 (if $n_1 + n_2 + \dots + n_k = n$)
 0 otherwise.

$\vec{X} \sim \text{Mult}_k(n, \vec{p})$. Find marginal distribution of X_j . Then $X_j \sim \text{Bin}(n, p_j)$.

$$E(X_j) = np_j, \text{Var}(X_j) = np_j(1-p_j)$$

Lumping Property

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

Let $\vec{Y} = (X_1, X_2, X_3 + \dots + X_{10})$. Then $\vec{Y} \sim \text{Mult}(n, (p_1, p_2, p_3 + \dots + p_{10}))$

Conditional Distribution

$\vec{X} \sim \text{Mult}(n, \vec{p})$. Then given $X_1 = n_1$, $(X_2, \dots, X_k) \sim \text{Mult}_{k-1}(n-n_1, (p'_2, \dots, p'_k))$ with

$$p'_2 = P(\text{being in category 2} \mid \text{not in category 1})$$

$$= \frac{p_2}{1-p_1} = \frac{p_2}{p_2 + \dots + p_k} \Rightarrow p'_j = \frac{p_j}{p_j + \dots + p_k}$$

Cauchy Interview Problem.

The Cauchy is distribution of $\frac{X}{Y} = T$ with $X, Y \text{ iid } N(0,1)$. Find PDF of T .

$$P\left(\frac{X}{Y} \leq t\right) \stackrel{\text{Symmetry of } N(0,1)}{=} P\left(\frac{X}{|Y|} \leq t\right) = P(X \leq t|Y|) = \int_{-\infty}^{\infty} \int_{-\infty}^{t|y|} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dx dy$$

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

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y e^{-y^2/2} e^{-t^2 y^2/2} dy$$

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

$$u = (1+t^2)y^2/2 \Rightarrow \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{1+t^2}$$

$$du = (1+t^2)y dy$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2} \int_{-\infty}^{t|y|} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx dy$$

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

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

$$P(X \leq t|Y|) = \int_{-\infty}^{\infty} P(X \leq t|Y| \mid Y=y) \underbrace{\varphi(y)}_{N(0,1) \text{ PDF}} dy = \int_{-\infty}^{\infty} \Phi(t|y|) \varphi(y) dy \quad (\text{law of total prob.})$$

Lecture 20

$$\boxed{\text{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))]} = \boxed{E(XY) - E(X)E(Y)} \quad \text{since } E(XY) - E(X)E(Y) - E(X)E(Y) = E(XY) - E(X)E(Y) + E(X)E(Y)$$

Properties: (1) $\text{Cov}(X, X) = \text{Var}(X)$ (2) $\text{Cov}(X, Y) = \text{Cov}(Y, X)$

(3) $\text{Cov}(X, c) = 0$ if c is a constant (4) $\text{Cov}(cX, Y) = c \cdot \text{Cov}(X, Y)$

(5) $\text{Cov}(X, Y+Z) = \text{Cov}(X, Y) + \text{Cov}(X, Z) \rightarrow \text{bilinearity}$

(6) $\text{Cov}(X+Y, Z+W) = \text{Cov}(X, Z) + \text{Cov}(X, W) + \text{Cov}(Y, Z) + \text{Cov}(Y, W)$

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

(7) $\text{Var}(X_1 + X_2) = \text{Var}(X_1) + \text{Var}(X_2) + 2\text{Cov}(X_1, X_2)$

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

Theorem: If X, Y are indep., then they are uncorrelated, i.e. $\text{Cov}(X, Y) = 0$. (Converse is false.)

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

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = E(Z^2) - E(Z)E(Z^2) = 0$$

But very dependent since Y is a function of X , and Y also determines magnitudes of X

$$\text{Def: } \text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\text{SD}(X)\text{SD}(Y)} = \frac{\text{Cov}(X - E(X), Y - E(Y))}{\text{SD}(X)\text{SD}(Y)}$$

Theorem: $-1 \leq \text{Corr}(X, Y) \leq 1$ (form of Cauchy-Schwarz)

Proof: WLOG (without the loss of generality) assume X, Y are standardized, let $\text{Corr}(X, Y) = \rho$.

$$\begin{aligned} \text{as } \text{Var}(X+Y) &= \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y) = 2 + 2\rho \\ \text{as } \text{Var}(X-Y) &= \text{Var}(X) + \text{Var}(Y) - 2\text{Cov}(X, Y) = 2 - 2\rho \end{aligned} \quad \left. \vphantom{\begin{aligned} \text{as } \text{Var}(X+Y) \\ \text{as } \text{Var}(X-Y) \end{aligned}} \right\} -1 \leq \rho \leq 1$$

ex: Cov in a Multinomial

$$(X_1, \dots, X_k) \sim \text{Mult}(n, \vec{p})$$

Find $\text{Cov}(X_i, X_j)$ for all i, j .

$$\text{If } i=j, \text{Cov}(X_i, X_j) = \text{Var}(X_i) = np_i(1-p_i).$$

$$\text{If } i \neq j, \text{Cov}(X_i, X_j) = -np_i p_j.$$

Find this first:

$$\text{Cov}(X_1, X_2) = c$$

$$\begin{aligned} \text{Var}(X_1 + X_2) &= np_1(1-p_1) + np_2(1-p_2) + 2c \\ &= n(p_1 + p_2)(1 - (p_1 + p_2)) \\ &\Rightarrow \text{Cov}(X_1, X_2) = -np_1 p_2 \end{aligned}$$

ex: $X \sim \text{Bin}(n, p)$, write as $X = X_1 + \dots + X_n$, $\{X_j\}$ are iid. $\text{Bern}(p)$.

$$\Rightarrow \text{Var}(X) = npq \text{ since } \text{Cov}(X_i, X_j) = 0 \text{ for } i \neq j$$

$$\begin{aligned} \text{Var}(X_j) &= E(X_j^2) - [E(X_j)]^2 \\ &= p - p^2 = p(1-p) \end{aligned}$$

Let I_A be indicator of event A .

$$I_A^2 = I_A$$

$$I_A^3 = I_A$$

$$I_A I_B = I_{A \cap B}$$

ex: $X \sim \text{HG}(w, b, n)$

$$X = X_1 + \dots + X_n, X_j = \begin{cases} 1, & \text{if } j\text{th ball white} \\ 0, & \text{otherwise} \end{cases}$$

$$\text{Var}(X) = n \text{Var}(X_1) + 2 \binom{n}{2} \text{Cov}(X_1, X_2)$$

symmetry

$$\text{Cov}(X_1, X_2) = E(X_1 X_2) - E(X_1)E(X_2)$$

$$= \frac{w}{w+b} \left(\frac{w-1}{w+b-1} \right) - \left(\frac{w}{w+b} \right)^2$$

Lecture 2

Var of Hypergeometric(w, b, n), $p = \frac{w}{w+b}$

$$\text{Var}\left(\sum_{j=1}^n X_j\right) = \text{Var}(X_1) + \dots + \text{Var}(X_n) + 2 \sum_{i < j} \text{Cov}(X_i, X_j) = np(1-p) + 2 \binom{n}{2} \left(\frac{w}{w+b} \left(\frac{w-1}{w+b-1} \right) - p^2 \right)$$

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

Extreme cases: $n=1$; N much much larger than n

Transformations

Theorem: Let X be a continuous r.v. with PDF f_X , $Y = g(X)$, where g is differentiable,

Strictly increasing. Then the PDF of Y is given by $f_Y(y) = f_X(x) \frac{dx}{dy}$, where $y = g(x)$, and this is written in terms of y . Also $\frac{dx}{dy} = \left(\frac{dy}{dx} \right)^{-1}$.

Proof: CDF of Y is $P(Y \leq y) = P(g(X) \leq y) = P(X \leq g^{-1}(y)) = F_X(g^{-1}(y)) = F_X(x)$

$$\Rightarrow f_Y(y) = f_X(x) \frac{dx}{dy} \quad (\text{Chain Rule})$$

(log Normal)

ex: $Y = e^Z$, $Z \sim N(0, 1)$

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(\ln(y))^2}{2}} \cdot \frac{1}{y}$$

$$\left[\frac{dy}{dz} = e^z = y \right]$$

transformations in $\mathbb{R}^n$

$\vec{Y} = g(\vec{X})$, $g: \mathbb{R}^n \rightarrow \mathbb{R}^n$; $\vec{X} = (x_1, \dots, x_n)$ continuous.

Joint PDF of $\vec{Y}$ is $f_{\vec{Y}}(\vec{y}) = f_{\vec{X}}(\vec{x}) \left| \frac{d\vec{x}}{d\vec{y}} \right|$ → Jacobian
abs. of determinant

$$\frac{d\vec{x}}{d\vec{y}} = \begin{bmatrix} \frac{dx_1}{dy_1} & \dots & \frac{dx_n}{dy_1} \\ \vdots & \ddots & \vdots \\ \frac{dx_1}{dy_n} & \dots & \frac{dx_n}{dy_n} \end{bmatrix}$$

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

Convolution (sums)

Let $T = X + Y$, 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 $F_T(t) = P(T \leq t) = \int_{-\infty}^{\infty} P(X+Y \leq t | X=x) f_X(x) dx$
 $= \int_{-\infty}^{\infty} F_Y(t-x) f_X(x) dx$

Idea: prove existence of object with desired properties A using prob.

Show $P(A) > 0$ for a random object.

Suppose each object has a "score". There is an object with "good" score. There is an object whose score is at least $E(X)$
 ← score of random object.

Ex: 100 people, 15 committees of 20; each person is on 3 committees. Show there exists 2 committees with overlap ≥ 3 . Idea: find average overlap of 2 random committees.

$$E(\text{overlap}) = 100 \cdot \frac{\binom{3}{2}}{\binom{100}{2}} = \frac{300}{\frac{15 \cdot 14}{2}} = \frac{40}{14} = \frac{20}{7} \Rightarrow \text{There exists pair of committees with overlap of } \geq \frac{20}{7} \approx \text{have overlap at least 3.}$$

Lecture 22

Beta(a, b), $a > 0, b > 0$

PDF: $f(x) = c x^{a-1} (1-x)^{b-1}$, $0 \leq x \leq 1$

- Flexible family of continuous distributions on $(0, 1)$
- Often used as prior for a parameter in $(0, 1)$
- "Conjugate prior to Binomial"
- Connections to other distributions

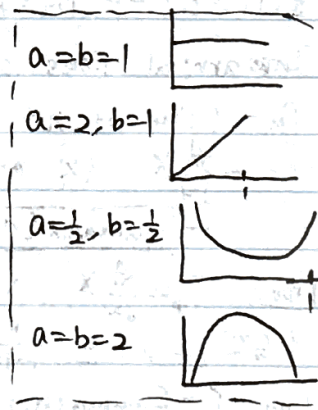
Conjugate prior for Bin

$X|p \sim \text{Bin}(n, p)$, $p \sim \text{Beta}(a, b)$ (prior)

Find posterior distribution $p|X$, $f(p|X=k)$

$$= \frac{P(X=k|p)f(p)}{P(X=k)}$$

→ does not depend on p



$$= \binom{n}{k} p^k (1-p)^{n-k} < p^{a-1} (1-p)^{b-1} / p(x=k)$$

$$\propto p^{a+k-1} (1-p)^{b+n-k-1} \Rightarrow p|X \sim \text{Beta}(a+X, b+n-X)$$

Find $\int_0^1 \binom{n}{k} x^k (1-x)^{n-k} dx$ without using calculus, using a story (Bayes' Billiards).

$n+1$ billiard balls, all white; paint one pink, throw them on $(0,1)$ indep.

Let $X = \#$ balls to left of pink. $P(X=k) = \int_0^1 (P(X=k|p)) f(p) dp$
 or: first throw, then paint one pink.

$$= \int_0^1 \binom{n}{k} p^k (1-p)^{n-k} dp$$

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

Lecture 23

$$n! \approx \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$$

Gamma function

$$\Gamma(a) = \int_0^\infty x^a e^{-x} \frac{dx}{x}, \text{ for real } a > 0$$

$$\Gamma(n) = (n-1)! \text{ for } n \text{ a positive integer}$$

$$\Gamma(x+1) = x\Gamma(x)$$

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}, \quad \Gamma\left(\frac{3}{2}\right) = \frac{1}{2}\Gamma\left(\frac{1}{2}\right) = \frac{\sqrt{\pi}}{2}$$

Gamma Distribution

$$1 = \int_0^\infty \left(\frac{1}{\Gamma(a)} x^a e^{-x} \right) \frac{dx}{x}$$

Gamma $(a, 1)$ PDF

$\sim$ Gamma (a, λ) PDF

Let $Y = X/\lambda$, $X \sim \text{Gamma}(a, 1)$

$$f_Y(y) = f_X(x) \frac{dx}{dy} = \frac{1}{\Gamma(a)} (\lambda y)^a e^{-\lambda y} \frac{1}{\lambda y} \lambda \quad (y > 0)$$

Gamma Exponential

$$\begin{matrix} T_1 & T_2 & T_3 & T_4 \\ | & | & | & | \\ 0 & X & X & X \end{matrix}$$

$N_t = \#$ emails up to time t

$\sim \text{Pois}(\lambda t)$ # arrivals in disjoint intervals are indep.

$$P(T_1 > t) = P(N_t = 0) = e^{-\lambda t} \frac{\lambda^0}{0!}$$

Interarrival times are iid $\text{Expo}(\lambda)$.

$T_n =$ (time of n th arrival)

$$= \sum_{j=1}^n X_j \quad (X_j \text{ are iid } \text{Expo}(\lambda))$$

$$\sim \boxed{\text{Gamma}(n, \lambda)}$$

Proof: $T_n = \sum_{j=1}^n X_j$, X_j iid $\text{Expo}(1)$ is $\text{Gamma}(n, 1)$.

MGF of X_1 is $\frac{1}{1-t}$, $t < 1 \Rightarrow$ MGF of T_n is $\left(\frac{1}{1-t}\right)^n$, $t < 1$.

Let $Y \sim \text{Gamma}(n, 1)$, $E(e^{ty}) = \frac{1}{\Gamma(n)} \int_0^\infty e^{ty} y^n e^{-y} \frac{dy}{y} = \frac{1}{\Gamma(n)} \int_0^\infty y^n e^{-(1-t)y} \frac{dy}{y}$, let $x = (1-t)y$

$$dx = (1-t)dy$$

$$\frac{(1-t)^n}{\Gamma(a)} \int_0^\infty x^n e^{-x} \frac{dx}{x} = (1-t)^n$$

Let $X \sim \text{Gamma}(a, 1)$, find $E(x^c)$.

$$\frac{1}{\Gamma(a)} \int_0^\infty x^c x^a e^{-x} \frac{dx}{x} = \frac{1}{\Gamma(a)} \int_0^\infty x^{a+c} e^{-x} \frac{dx}{x} = \frac{\Gamma(a+c)}{\Gamma(a)}, \text{ if } a+c > 0$$

$$E(x) = \frac{\Gamma(a+1)}{\Gamma(a)} = \frac{a \Gamma(a)}{\Gamma(a)} = a, \quad E(x^2) = \frac{\Gamma(a+2)}{\Gamma(a)} = \frac{(a+1)a \Gamma(a)}{\Gamma(a)} = (a+1)a = a^2 + a$$

$$\Rightarrow \text{Var}(x) = a$$

$$\text{Gamma}(a, \lambda) \Rightarrow \text{mean}\left(\frac{a}{\lambda}\right), \text{Var}\left(\frac{a}{\lambda^2}\right)$$

Lecture 24

ex: Bank - post office

$X \sim \text{Gamma}(a, \lambda)$, $Y \sim \text{Gamma}(b, \lambda)$, indep.
 $\nwarrow$ wait at bank $\nwarrow$ wait at post office

Find the ^{Joint} distribution of total waiting time $\Rightarrow T = X + Y$, $\frac{X}{X+Y} = W$. Let $\lambda = 1$ to simplify the notation.

$$\text{Joint PDF: } f_{T,W}(t,w) = f_{X,Y}(x,y) \left| \frac{d(x,y)}{d(t,w)} \right| = \frac{1}{\Gamma(a)} \cdot \frac{1}{\Gamma(b)} x^a e^{-x} y^b e^{-y} \frac{1}{xy} \left| -t \right|$$

$$\text{Jacobian: } x+y=t, \frac{x}{x+y}=w, \frac{y}{x+y}=1-w \quad = \frac{1}{\Gamma(a)\Gamma(b)} x^a e^{-x} y^b e^{-y} \frac{1}{xy} t$$

$$x = tw, y = t(1-w) = t - tw$$

$$\begin{vmatrix} w & t \\ 1-w & -t \end{vmatrix} = -tw - t(1-w) = -tw - t + tw = -t$$

$$= \frac{1}{\Gamma(a)\Gamma(b)} w^{a-1} (1-w)^{b-1} t^{a+b} e^{-t} \frac{1}{t}$$

$$f_W(w) = \int_0^\infty f_{T,W}(tw) dt = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} w^{a-1} (1-w)^{b-1} \underbrace{\int_0^\infty t^{a+b-1} e^{-t} dt}_{\Gamma(a+b)} = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} w^{a-1} (1-w)^{b-1}$$

$\Rightarrow$ know normalizing constant of Beta(a, b)

$T \sim \text{Gamma}(a+b, 1)$, $W \sim \text{Beta}(a, b)$ are indep. $\rightarrow$ in this case it is true

Find $E(W)$, $W \sim \text{Beta}(a, b)$. (1) LOTUS. (2) $E\left(\frac{x}{x+y}\right) = \frac{E(x)}{E(x+y)} = \frac{a}{a+b}$

Order Statistics

Let $X_1, \dots, X_n$ be iid. The order statistics are $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$, where $X_{(1)} = \min(X_1, \dots, X_n)$, $\dots$, $X_{(n)} = \max(X_1, \dots, X_n)$.

ex: if n is odd, the median is $X_{(\frac{n+1}{2})}$. Get other "quantiles".

Difficult since dependent.

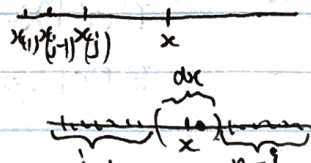
Tricky in discrete case because of "ties".

Let $x_1, \dots, x_n$ be iid with PDF f , CDF F . Find CDF and PDF of $X_{(j)}$.

$$\text{CDF: } P(X_{(j)} \leq x) = P(\text{at least } j \text{ of } X_i\text{'s are } \leq x)$$

$$= \sum_{k=j}^n \binom{n}{k} F(x)^k (1-F(x))^{n-k}$$

$$\text{PDF: } f_{X_{(j)}}(x) dx = n \binom{n-1}{j-1} f(x) dx F(x)^{j-1} (1-F(x))^{n-j}$$



ex: $U_1, \dots, U_n$ iid $\text{Unif}(0,1)$

$$f_{U_j}(x) = n \binom{n-1}{j-1} x^{j-1} (1-x)^{n-j}, \text{ for } 0 < x < 1$$

$$\Rightarrow U_j \sim \text{Beta}(j, n-j+1)$$

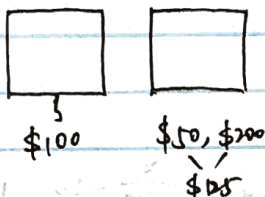
$$E(X|A) \xrightarrow{\text{event}}$$

$$E(X) = E(X|A) \cdot P(A) + E(X|A^c) \cdot P(A^c)$$

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

Expand with LOTP

Two envelope paradox



one envelope contains twice as much as other.

Lecture 25



one envelope has twice money as other.

Argument 1: $E(Y) = E(X)$ by symmetry.

Argument 2: $E(Y) = E(Y|Y=2X)P(Y=2X) + E(Y|Y=\frac{1}{2}X)P(Y=\frac{1}{2}X)$

$$E(Y|Y=2X) \neq E(2X)$$

$$\text{WRONG! } E(2X) \cdot \frac{1}{2} + E\left(\frac{X}{2}\right) \cdot \frac{1}{2}$$

Let I be indicator of $Y=2X$.

$$\neq \frac{5}{4} E(X)$$

$$= E(2X|Y=2X) \cdot \frac{1}{2} + E\left(\frac{X}{2} | Y=\frac{X}{2}\right) \cdot \frac{1}{2}$$

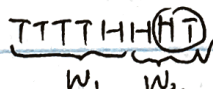
Then X, I are dependent.

ex: Patterns in coinflip

Repeated fair coin flips. How many flips until HT? $E(W_{HT}) = 4$

... HH? $E(W_{HH}) = 6$

$$E(W_{HT}) = E(W_1) + E(W_2) = 2 + 2 + 4$$



Since $W_j \sim \text{Geom}(\frac{1}{2})$

$$E(W_{HH}) = E(W_{HH} | \text{1st toss H}) \cdot \frac{1}{2} + E(W_{HH} | \text{1st toss T}) \cdot \frac{1}{2}$$

$$\text{TTTTHT} = (2 \cdot \frac{1}{2} + (2 + E(W_{HH})) \cdot \frac{1}{2}) + (1 + E(W_{HH})) \cdot \frac{1}{2} \Rightarrow E(W_{HH}) = 6$$

All wasted.



$$E(Y|X=x) = \sum_y y P(Y=y|X=x) \quad (\text{Discrete})$$

$$E(Y|X=x) = \int_{-\infty}^{\infty} y f_{Y|X}(y|x) dy = \int_{-\infty}^{\infty} y \frac{f_{X,Y}(x,y)}{f_X(x)} dy \quad (\text{Continuous})$$

Let $g(x) = E(Y|X=x)$. Then define $E(Y|X) = g(X)$.

ex: $X, Y \sim \text{Pois}(\lambda)$

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

$E(X|X+Y)$. Let $T=X+Y$, find conditional PMF.

$$P(X=k|T=n) = P(T=n|X=k)P(X=k)/P(T=n)$$

$$= \frac{P(Y=n-k)P(X=k)}{P(T=n)} = \frac{\frac{e^{-\lambda}\lambda^{n-k}}{(n-k)!} \cdot \frac{e^{-\lambda}\lambda^k}{k!}}{e^{-2\lambda}(2\lambda)^n/n!} = \binom{n}{k} \frac{1}{2^n}$$

$$X|T=n \sim \text{Bin}(n, \frac{1}{2})$$

$$E(X|T=n) = \frac{n}{2} \Rightarrow E(X|T) = \frac{T}{2}$$

$$E(X|X+Y) = E(Y|X+Y) \text{ by symmetry (since indep.)}$$

$$E(X|X+Y) + E(Y|X+Y) = E(X+Y|X+Y) = X+Y \Rightarrow E(X|T) = T/2$$

Iterated E (Adam's Law) $E(E(Y|X)) = E(Y)$.

Lecture 26

ex: Let $X \sim N(0,1)$, $Y=X^2$. Then $E(Y|X) = E(X^2|X) = X^2 = Y$.

$E(X|Y) = E(X|X^2) = 0$ since if observe $X^2=a$, then $X = \pm\sqrt{a}$, equally likely.

ex: $\text{stick} \rightarrow$ stick, break off random piece, break off another piece.

$X \sim \text{Unit}(0,1)$, $Y|X \sim \text{Unit}(0,X)$.

$$E(Y|X=X) = \frac{X}{2}, \text{ so } E(Y|X) = \frac{X}{2}, E(E(Y|X)) = \frac{1}{4} = E(Y)$$

Useful Properties

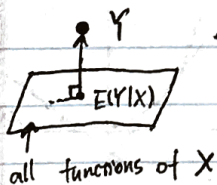
(1) $E(h(X)Y|X) = h(X)E(Y|X)$ (taking out what's known)

(2) $E(Y|X) = E(Y)$ if X, Y are independent.

(3) $E(E(Y|X)) = E(Y)$ (Iterated Expectation / Adam's Law)

(4) $E(Y - E(Y|X)) \cdot h(X) = 0$, i.e. $Y - E(Y|X)$ (residual) is uncorrelated with $h(X)$

$$\text{Cov}(Y - E(Y|X), h(X)) = E((Y - E(Y|X)) \cdot h(X)) - E(Y - E(Y|X))E(h(X))$$



$$\langle X, Y \rangle = E(XY)$$

Proof of (4): $E(Yh(X)) - E(E(Y|X)h(X))$

$$= E(Yh(X)) - E(E(h(X)Y|X)) \text{ (Adam's Law)}$$

$$= E(Yh(X)) - E(Yh(X)) = 0$$

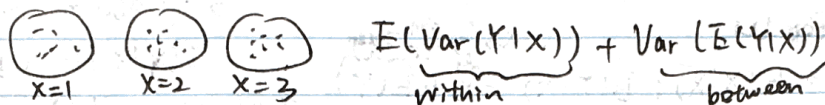
Proof of (3) Discrete case: Let $E(Y|X) = g(X)$.

$$E(g(X)) = \sum_x g(x)P(X=x) = \sum_x E(Y|X=x)P(X=x)$$

$$= \sum_x \left(\sum_y y P(Y=y|X=x) \right) P(X=x)$$

$$= \sum_y \sum_x y \cdot P(Y=y, X=x) = \sum_y y \sum_x P(Y=y) = E(Y)$$

$$\text{Def: } \text{Var}(Y|X) = E(Y^2|X) - (E(Y|X))^2 = E((Y - E(Y|X))^2|X)$$



ex: Pick random city, pick random sample of people in that city. $X = (\# \text{ with disease})$, $Q = (\text{proportion of people in the random city with disease})$. Find $E(X)$, $\text{Var}(X)$, assuming $Q \sim \text{Beta}(a, b)$, $X|Q \sim \text{Bin}(n, Q)$.

$$E(X) = E(E(X|Q)) = E(nQ) = n\alpha/(\alpha+\beta)$$

$$\text{Var}(X) = E(\text{Var}(X|Q)) + \text{Var}(E(X|Q)) = E(nQ(1-Q)) + n^2 \text{Var}(Q)$$

$$= n \frac{\alpha\beta}{(\alpha+\beta)(\alpha+\beta)} + \frac{n^2 \alpha(1-\alpha)}{\alpha+\beta+1}$$

$$\begin{aligned} E(Q(1-Q)) &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 q(1-q) d\pi(q) \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+1)\Gamma(\beta+1)}{\Gamma(\alpha+\beta+2)} \\ &= \frac{\alpha\beta}{(\alpha+\beta)(\alpha+\beta+1)} \end{aligned}$$

Lecture 2

ex: Store with a random # of customers N . Let X_j be the amount j th customer spends. X_j has mean μ , $\text{Var} \sigma^2$. Assume $N, X_1, \dots, X_N$ are indep. Find mean, var of $X = \sum_{j=1}^N X_j$.

~~$E(X) = N\mu$~~ category error.

$$E(X) = \sum_{n=0}^{\infty} E(X|N=n) P(N=n) = \sum_{n=0}^{\infty} n\mu P(N=n) = \mu \cdot E(N)$$

$$\text{Adam's Law: } E(X) = E(E(X|N)) = E(\mu N) = \mu \cdot E(N)$$

$$\text{Var}(X) = E(\text{Var}(X|N)) + \text{Var}(E(X|N)) = E(N\sigma^2) + \text{Var}(\mu N) = \sigma^2 E(N) + \mu^2 \text{Var}(N)$$

Statistical Inequalities

(1) Cauchy-Schwarz: $|E(XY)| \leq \sqrt{E(X^2)E(Y^2)}$ (if X, Y uncorrelated, $E(XY) = E(X)E(Y)$)

If X, Y have mean 0, then $|\text{Corr}(X, Y)| = \frac{|E(XY)|}{\sqrt{E(X^2)E(Y^2)}} \leq 1$.

(2) Jensen's Inequality: If g is convex then $E(g(X)) \geq g(E(X))$.

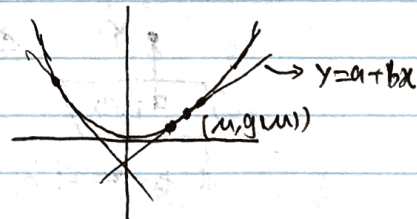
If it's concave $E(g(X)) \leq g(E(X))$, $g''(x) \geq 0$ $\cup$
 $g''(x) < 0 \cap$

$$\text{ex: } E\left(\frac{1}{X}\right) \geq \frac{1}{E(X)}, \quad E(\ln X) \leq \ln(E(X))$$

Proof: $g(x) \geq a + bx$

$$E(g(X)) \geq E(a + bX) = a + bE(X)$$

$$= a + b\mu = g(\mu) = g(E(X))$$



(3) Markov: $P(|X| \geq a) \leq \frac{E(|X|)}{a}$, for any $a > 0$.

Proof: $aE(I_{|X| \geq a}) \leq E(|X|)$ $aI_{|X| \geq a} \leq |X|$ is true. So $aE(I_{|X| \geq a}) \leq E(|X|)$

ex: 100 people. Is it possible that at least 95% people younger than avg. in group?
YES!

at least 50% people older than twice avg

(4) Chebyshev: $P(|X-\mu| > a) \leq \frac{\text{Var } X}{a^2}$ for $\mu = E(X)$, $a > 0$

$$P(|X-\mu| > \text{CSD}(X)) \leq \frac{1}{c^2} \rightarrow c > 0$$

Proof: $P(|X-\mu| > a) = P((X-\mu)^2 > a^2) \leq \frac{E(X-\mu)^2}{a^2} = \frac{\text{Var}(X)}{a^2}$

Lecture 28

Let $X_1, X_2, \dots$ be iid, Mean μ , Var σ^2 , let $\bar{X}_n = \frac{1}{n} \sum_{j=1}^n X_j$ (sample mean)

Law of large Numbers: $\bar{X}_n \rightarrow \mu$ as $n \rightarrow \infty$ with probability 1.

(strong law of large numbers) Sample mean converges to true mean

ex: $X_j \sim \text{Bern}(p)$, then $\frac{X_1 + \dots + X_n}{n} \rightarrow p$ with prob 1.

Weak law of large numbers: For any $c > 0$, $P(|\bar{X}_n - \mu| > c) \rightarrow 0$ as $n \rightarrow \infty$.

Proof: $P(|\bar{X}_n - \mu| > c) \leq \frac{\text{Var}(\bar{X}_n)}{c^2} = \frac{\frac{1}{n^2} \sum \sigma^2}{c^2} = \frac{\sigma^2}{nc^2} \rightarrow 0$

$\bar{X}_n - \mu \rightarrow 0$ with prob 1, but what does the dist. $n \rightarrow \infty$ look like?

$(\bar{X}_n - \mu) \sqrt{n} \Rightarrow$ Central Limit Theorem

$\downarrow$ Converges to

$N(0, 1)$ in distribution.

Equivalently: $\frac{\sum_{j=1}^n X_j - n\mu}{\sqrt{n} \cdot \sigma} \rightarrow N(0, 1)$ in distribution.

Proof (assume MGF $M(t)$ of X_j exists)

Can assume $\mu=0$, $\sigma=1$ since consider $\frac{1}{\sqrt{n}} \sum_{j=1}^n \frac{(X_j - \mu)}{\sigma}$. Let $S_n = \sum_{j=1}^n X_j$, show MGF of $\frac{S_n}{\sqrt{n}}$ goes to $N(0, 1)$ MGF.

$$E(e^{tS_n/\sqrt{n}}) = E(e^{tX_1/\sqrt{n}}) \dots E(e^{tX_n/\sqrt{n}})$$

$$= (M(\frac{t}{\sqrt{n}}))^n \quad (\text{take logs})$$

$$\lim_{n \rightarrow \infty} n \log M(\frac{t}{\sqrt{n}}) = \lim_{n \rightarrow \infty} \frac{\log M(\frac{t}{\sqrt{n}})}{1/n} = \lim_{y \rightarrow 0} \frac{\log M(y)}{y^2} = \lim_{y \rightarrow 0} \frac{t M'(y)}{2y M(y)} = \frac{t}{2} \lim_{y \rightarrow 0} \frac{M'(y)}{y} = \frac{t^2}{2} \lim_{n \rightarrow \infty} \frac{M'(y)}{1} = \frac{t^2}{2}$$

$\frac{t^2}{2}$ is the log of $e^{t^2/2}$ $N(0, 1)$ MGF

Binomial Approximation by Normal $\rightarrow$ prefer n large, p close to $\frac{1}{2}$ n is large, p is small, $\lambda=np$

Let $X \sim \text{Bin}(n, p)$, think of $X = \sum_{j=1}^n X_j$, $X_j \sim \text{Bern}(p)$ iid.

$$P(a \leq X \leq b) = P\left(\frac{a-np}{\sqrt{npq}} \leq \frac{X-np}{\sqrt{npq}} \leq \frac{b-np}{\sqrt{npq}}\right) \approx \Phi\left(\frac{b-np}{\sqrt{npq}}\right) - \Phi\left(\frac{a-np}{\sqrt{npq}}\right)$$

Contrast with Poisson

Continuity correction: $P(X = \overset{\text{integer}}{a}) = P(\underline{a - \frac{1}{2}} < X < \underline{a + \frac{1}{2}})$

Lecture 29

$\chi^2(n)$ (Chi-square)

Let $V = Z_1^2 + Z_2^2 + \dots + Z_n^2$, $Z_j \text{ iid } N(0,1)$. Then $V \sim \chi^2(n)$.

Fact: $\chi^2(1)$ is Gamma($\frac{1}{2}, \frac{1}{2}$).

So $\chi^2(n)$ is Gamma($\frac{n}{2}, \frac{1}{2}$).

Student-t (Gosset)

Let $T = \frac{Z}{\sqrt{V/n}}$, with $Z \sim N(0,1)$, $V \sim \chi^2(n)$ indep. Then $T \sim t_n$.

Properties: (1) Symmetric, i.e. $-T \sim t_n$

(2) $n=1 \Rightarrow$ Cauchy, mean doesn't exist.

(3) $n \geq 2 \Rightarrow E(T) = E(Z)E(\frac{1}{\sqrt{V/n}}) = 0$

(4) Heavier-tailed than normal

(5) For n large, t_n looks like very much like $N(0,1)$ distribution of t_n goes to $N(0,1)$ as $n \rightarrow \infty$.

Let $T_n = \frac{Z}{\sqrt{V_n/n}}$, with $Z_1, Z_2, \dots \text{ iid } N(0,1)$, $V_n = Z_1^2 + \dots + Z_n^2$

Then $\frac{V_n}{n} \rightarrow 1$ with prob 1 by Law of large numbers $Z \sim N(0,1)$ indep. Z_j 's.

Since $E(Z_i^2) = 1$, $\sqrt{\frac{V_n}{n}} \rightarrow 1$ with prob. 1. So $T_n \rightarrow Z$ with prob 1. So t_n converges to $N(0,1)$ distribution.

$$\begin{array}{l} \text{MGF} \\ E(Z^2) = 1 \\ E(Z^4) = 3 \\ E(Z^6) = 15 \\ \vdots \end{array} \quad \begin{array}{l} E(Z^{2n}) = E((Z^2)^n) \\ \downarrow \\ \text{Gamma}(\frac{n}{2}, \frac{1}{2}) \\ \chi^2(1) \end{array}$$

Multivariate Normal (MVN)

Def: Random vector $(X_1, X_2, \dots, X_k) = \vec{X}$ is multivariate normal if every linear combination $t_1 X_1 + t_2 X_2 + \dots + t_k X_k$ is normal.

ex: Let Z, W be iid $N(0,1)$. Then $(Z + 2W, 3Z + 5W)$ is MVN.

$S(Z + 2W) + t(3Z + 5W) = (5 + 3t)Z + (2S + 5t)W$ is normal.

non-ex: $Z \sim N(0,1)$, let S be random assignment indep. of Z . Then Z, SZ are marginally $N(0,1)$. But (Z, SZ) is not MVN. Look at $Z + SZ$.

MGF of MVN

$$\begin{aligned} \vec{X}^{MVN} &\Rightarrow E(e^{\vec{t}^T \vec{X}}) = E(e^{t_1 X_1 + \dots + t_k X_k}) \\ E(X_j) &= \mu_j &= e^{t_1 \mu_1 + \dots + t_k \mu_k + \frac{1}{2} \text{Var}(t_1 X_1 + \dots + t_k X_k)} \end{aligned}$$

Theorem: MVN with independent implies indep.

$\vec{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ MVN; if every component of $\vec{X}_1$ is uncorrelated, with every component of $\vec{X}_2$, then $\vec{X}_1$ is indep. of $\vec{X}_2$.

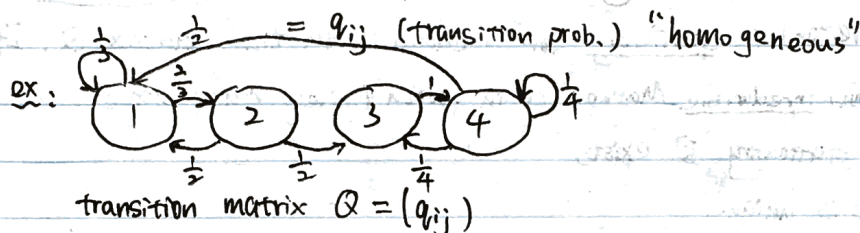
ex: Let X, Y be iid $N(0, 1)$. Then $(X+Y, X-Y)$ is MVN. Uncorrelated: $\text{Cov}(X+Y, X-Y) = \text{Var}(X) + \text{Cov}(X, Y) - \text{Cov}(X, Y) - \text{Var}(Y) = 0$. So $X+Y, X-Y$ are indep.,

Lecture 30

Markov Chains (example of Stochastic processes)

$X_0, X_1, X_2, \dots$ (think of X_n as state of system at (discrete) time n)

Markov properties: $P(X_{n+1}=j | X_n=i, X_{n-1}=i_{n-1}, X_{n-2}=i_{n-2}, \dots, X_0=i_0)$
 $= P(X_{n+1}=j | X_n=i)$ past and future are conditionally indep. given present.



$$= \begin{bmatrix} \frac{1}{3} & \frac{2}{3} & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 \\ \frac{1}{2} & 0 & \frac{1}{4} & \frac{3}{4} \end{bmatrix} \quad \text{Each row sums to 1}$$

Suppose at time n , X_n has distribution $\vec{z}$ (row vector) $1 \times M$ matrix.
 $\rightarrow$ PMF listed out

$$P(X_{n+1}=j) = \sum_i P(X_{n+1}=j | X_n=i) P(X_n=i) \\ = \sum_i s_i q_{ij} \quad \text{is } j\text{th entry of } \vec{z} \cdot Q$$

$1 \times M \quad M \times M$

So $\vec{z}Q$ is distribution at time $n+1$.

So $\vec{z}Q^2 \dots \dots \dots n+2, \vec{z}Q^3 \dots \dots \dots n+3.$

$$P(X_{n+1}=j | X_n=i) = q_{ij}$$

$$P(X_{n+2}=j | X_n=i) = \sum_k P(X_{n+2}=j | X_{n+1}=k, X_n=i) P(X_{n+1}=k | X_n=i) \\ = \sum_k q_{ik} q_{kj} \quad \text{is } (i,j) \text{ entry of } Q^2.$$

$\vdots$

$$P(X_{n+m}=j | X_n=i) = [(i,j) \text{ entry of } Q^m]$$

Stationary dist.

$\vec{z}$ (prob. vector $1 \times M$) is stationary for the chain if $\vec{z}Q = \vec{z}$.

$1 \times M \quad M \times M$

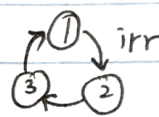
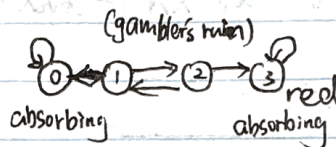
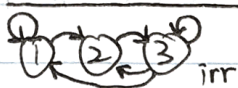
Questions: (1) Does a stationary dist. exist?

(2) Is it unique?

(3) Does chain converges to $\vec{z}$?

(4) How can we compute?

Lecture 31



Chain is irreducible if possible (with prob pos.) to get from anywhere $\rightarrow$ anywhere.

A state is recurrent if starting there, chain has prob. 1 of returning to that state. Otherwise transient.

$\vec{S}$ (prob. row vector) is stationary for a chain with trans. matrix Q . If $\vec{S}Q = \vec{S}$.

Theorem: For any irreducible Markov chain (with finitely many states):

(1) A stationary $\vec{S}$ exists

(2) It is unique

(3) $S_i = \frac{1}{r_i}$, where r_i is average return to time to i , starting at i

(4) If Q^m is strictly positive for some m , then $P(X_n = i) \rightarrow S_i$ as $n \rightarrow \infty$

$\vec{t}Q^n \rightarrow \vec{S}$ as $n \rightarrow \infty$

$\leftarrow$ any prob. vector.

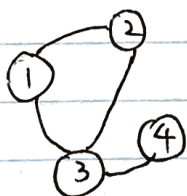
Reversible Markov Chain

Def: Markov chain with trans. matrix $Q = (q_{ij})$ is reversible if there is a prob. vector $\vec{S}$ such that $S_i q_{ij} = S_j q_{ji}$ for all states i, j .

Theorem: If reversible with respect to $\vec{S}$, then $\vec{S}$ is stationary

Let $S_i q_{ij} = S_j q_{ji}$ for all i, j , show $\vec{S}Q = \vec{S}$.

$$\sum_j S_i q_{ij} = \sum_j S_j q_{ji} = S_j \sum_i q_{ji} = S_j \Rightarrow \text{So } \vec{S}Q = \vec{S}$$

Random Walk on an (Undirected) Network

Let d_i be degree of i $\left[\begin{matrix} d_1=2, d_2=2 \\ d_3=3, d_4=1 \end{matrix} \right]$

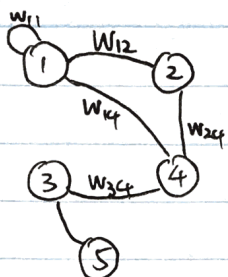
Then $d_i q_{ij} = d_j q_{ji}$ for all i, j

Let $i \neq j$, q_{ij}, q_{ji} are both 0 or both not 0

If $\{i, j\}$ is an edge, then $q_{ij} = 1/d_i$

So with M nodes, $1, 2, \dots, M$, degrees d_i , then $\vec{S}$ with $S_i = \frac{d_i}{\sum_j d_j}$ is stationary

Lecture 32



$w_{ij} \geq 0$ "weight", $w_{ij} = 0$ if no edge there $\{i, j\}$. $w_{ij} = w_{ji}$

random walk: from state i , go to j with prob $\propto w_{ij}$

If $\{i, j\}$ is an edge, then $q_{ij} = \frac{w_{ij}}{\sum_k w_{ik}}$

Note: $(\sum_k w_{ik}) q_{ij} = w_{ij} = w_{ji} = (\sum_k w_{jk}) q_{ji} \Rightarrow s_i \propto \sum_k w_{ik}$

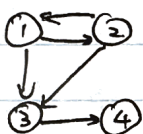
Conversely: Any reversible chain is of this form.

Let $w_{ij} = s_i q_{ij} = s_j q_{ji} = w_{ji}$

$$P(X_{n+1}=j | X_n=i) = w_{ij} / \sum_k w_{ik} = \frac{s_i q_{ij}}{\sum_k s_i q_{ik}} = \frac{s_i q_{ij}}{s_i \sum_k q_{ik}} = q_{ij}$$

Non-reversible example: Google PageRank

WWW



Importance of page based on pages linking to it and their importance.

$s_j = \sum_i s_i q_{ij} \Rightarrow \vec{s} = \vec{s} Q$, which says $\vec{s}$ is stationary dist. of random web-surfing chain.

$$Q = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{bmatrix} \quad \vec{s} \text{ is normalizing}$$

$$G = \alpha Q + (1-\alpha) \underbrace{\frac{J}{M}}_{\text{teleportation}}, \text{ where } M = \# \text{ pages}, J = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, 0 < \alpha < 1$$

$\alpha = 0.85$

G guarantees irreducibility, no zeros in transition matrix G

Use convergence to stationary.

Let $\vec{p}$ be initial prob. vector. $\vec{p} G = \alpha \vec{p} Q + (1-\alpha) \frac{\vec{p} J}{M}$ Very sparse (mostly 0's)

$$\vec{p} J = (1, 1, \dots, 1)$$

$$\vec{p} G^2 = (\vec{p} G) G, \vec{p} G^3, \dots, \vec{p} G^n, \dots$$