CS2810 Day 13
Mar 2

Quiz Friday: Prof Higger Recitation style review weds @ 2pm
- recorded
- will build examples from popular topics, see piazza post

Joint Distribution
Marginalization
Independence

Law of Large Numbers

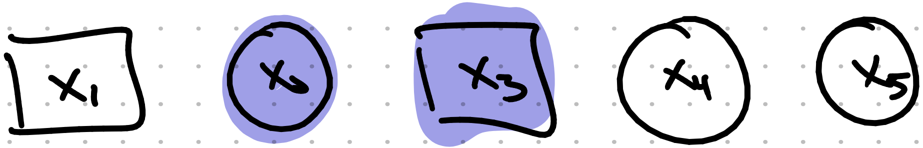
Poisson Distribution

Binomial Distribution

JOINT DISTRIBUTION

A joint distribution of two random variables gives the prob of pairs of outcomes, one from each experiment, occuring together

Experiment: choose one of 5 objects below (each has equal prob)



B=1 IF OBJECT BLUE ELSE 0

C=1 IF OBJECT CIRCLE ELSE 0

	B=0	B=1
C=0	1/5	1/5
C=1	2/5	1/5

Note about when joint distributions exist:

There must be some way of pairing observations in one random variable to the other

There is a natural pairing here (joint distribution defined):
X - temperature on a given day
Y - number of hats people where on that day

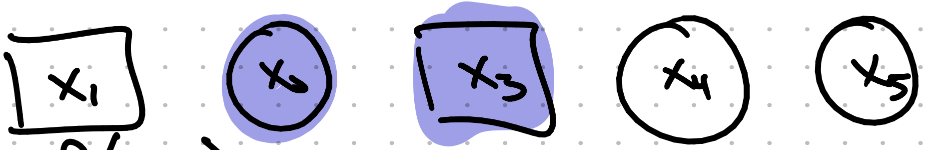
On each day I observe some outcome x and some outcome y.

There is no natural pairing here (no joint distribution defined):
X - temperature on a given day
Y - outcome of a 6 sided die roll

... not quite sure how to pair a temperature x with a six sided die roll y ... not well defined

MAGINALIZE

Marginalization is the process of removing one (or more) variables from the joint distribution.
It yields a distribution over the remaining variables.



JOINT $P(BC)$

Given the joint distribution below, what is the distribution of B, the event the chosen object is blue?

	B=0	B=1
C=0	1/5	1/5
C=1	2/5	1/5

MARGINALS

$$P(B=1) = P(B=1\ C=0) + P(B=1\ C=1)$$

$$= 1/5 + 1/5$$

$$P(B=0) = P(B=0\ C=0) + P(B=0\ C=1) = 3/5$$

MARGINALIZATION

$$P(X=x) = \sum_{y} P(X=x, Y=y)$$

INDEPENDENCE

Intuitive Definition:

We say that two Random Variables are Independent if observing the outcome of either doesn't inform us about the outcome of the other.

Independent Random Variables
X = stock market % increase on a given day
Y = how many people were wearing blue shoes @ 8AM in Boston on same day

Dependent Random Variables
X = number of points scored by basketball team in a game
Y = whether that team won the game

INDEPENDENCE

Algebraic Definition:

X, Y ARE INDEPENDENT RANDOM VARIABLES IF,

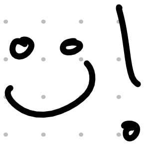
FOR ALL OUTCOMES x y:

$$P(X=x\ Y=y) = P(X=x)\,P(Y=y)$$

PROB OUTCOMES HAPPEN TOGETHER

PRODUCT OF PROB OF EACH OUTCOME HAPPENING

Linearity of Expectation



GOAL USE LINEARITY OF EXPECTATION TO FIND EXPECTED VALUE / VARIANCE OF ADDITION / MULTIPLICATION OF RANDOM VARIABLE X AND CONSTANT C

$$E[X+C] = E[X] + C$$

$$E[CX] = C\,E[X]$$

$$\mathrm{VAR}(X+C) = \mathrm{VAR}(X)$$

$$\mathrm{VAR}(CX) = C^2\,\mathrm{VAR}(X)$$

GOAL USE LINEARITY OF EXPECTATION TO FIND EXPECTED VALUE / VARIANCE OF ADDITION / MULTIPLICATION OF RANDOM VARIABLE X AND RANDOM VARIABLE Y

NEW!

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

$$VAR(X+Y) = VAR(X) + VAR(Y)$$

ASSUMES INDEPENDENCE

P	X	Y	X+Y
1/2	-1	1	0
1/2	1	-1	0

EXAMPLE DEPENDENCE IMPACTS VAR(X+Y)

$$\mathrm{VAR}(X+Y) = E[(X+Y)^2] - E[X+Y]^2$$

$$= E[X^2+2XY+Y^2] - (E[X]+E[Y])^2$$

$$= E[X^2] + 2E[XY] + E[Y^2] - E[X]^2 - 2E[X]E[Y] - E[Y]^2$$

$$= \mathrm{VAR}(X) + \mathrm{VAR}(Y) + 2(E[XY] - E[X]E[Y])$$

IF XY INDEP

$E[XY] = E[X]E[Y]$

(NEXT SLIDE)

ASSUME INDEPENDENCE

$$= \mathrm{VAR}(X) + \mathrm{VAR}(Y)$$

ASSUME X, Y INDEPENDENT

$$E[XY] = \sum_{x,y} x \cdot y \, P(X=x, Y=y)$$

HERE

$$= \sum_{x} x \, P(X=x) \left(\sum_{y} y \, P(Y=y) \right)$$

$$= \sum_{x} x \, P(X=x) \, E[Y]$$

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

ICA 1:

In terms of expectation and variance, explain how each of the following are similar / different. Which values are the same, which are bigger/smaller? Why?

let x be a "coin flip" random variable P(X=0) = .5, P(X=1) = .5

- the "average" of 1 coin flip
- the average of 10 independent coin flips
- the average of 100 independent coin flips

$$\frac{X_0 + X_1 + X_2 + \dots + X_9}{10}$$

First, build an intuition. If you get stuck (or feel confident in your intuition), use the linearity of expectation formulae to explicitly compute the expected val / variances below.

"VAR GETS SMALLER AS WE AVERAGE OVER MORE COIN FLIPS"

EXPECTED VALUE IS SAME FOR ALL AVERAGES

$$E[x] = \sum_x x\,P(x) = 0 \cdot \tfrac{1}{2} + 1 \cdot \tfrac{1}{2} = .5$$

$$E\left[\frac{x_0 + x_1 + \ldots + x_9}{10}\right] = E\left[\frac{x_0}{10}\right] + E\left[\frac{x_1}{10}\right] + \ldots + E\left[\frac{x_9}{10}\right]$$

$$= \frac{1}{10}E[x_0] + \frac{1}{10}E[x_1] + \ldots + \frac{1}{10}E[x_9]$$

$$= \frac{1}{10}(.5) + \frac{1}{10}(.5) + \ldots + \frac{1}{10}(.5) = .5$$

VARIANCE

$$\mathrm{VAR}(x) = E[x^2] - E[x]^2 = 1/2 - (1/2)^2 = 1/4$$

$$E[x^2] = \sum_x x^2 P(x) = 0^2 \cdot 1/2 + 1^2 \cdot 1/2 = 1/2$$

$$\mathrm{VAR}\left(\frac{X_0 + X_1}{2}\right) = \mathrm{VAR}(1/2\, X_0 + 1/2\, X_1) = \mathrm{VAR}(\tfrac{1}{2} X_0) + \mathrm{VAR}(\tfrac{1}{2} X_1)$$

$$= 1/4\, \mathrm{VAR}(X_0) + 1/4\, \mathrm{VAR}(X_1)$$

$$= 1/2\, \mathrm{VAR}(X)$$

PYTHON DEMO:

LAW OF LARGE NUMBERS

LAW OF LARGE NUMBERS

"iid"

GIVEN N INDEPENDENT IDENTICALLY DISTRIBUTED RANDOM VARIABLES X_i

ALL x_i HAVE SAME EXPECTED VALUE

CHANCES ARE $\frac{1}{N}\sum_{i=1}^{N} X_i$ GETS CLOSER TO $E[X]$ AS N INCREASES

LAW LARGE NUMBERS "PROOF"

GIVEN X Y INDEP

$VAR(X+Y) = VAR(X) + VAR(Y)$

$$VAR\left(\frac{1}{N}\sum X_i\right) = \frac{1}{N^2}\left(VAR\ X_1 + VAR\ X_2 + VAR\ X_3 + \ldots\right)$$

$VAR(CX) = C^2\ VAR(X)$

ALL X_i HAVE SAME VAR

$$= \frac{1}{N^2}\left(N\ VAR(X)\right)$$

$$= \frac{1}{N}\ VAR(X)$$

NOTICE: AS N INCREASES VARIANCE GETS CLOSER TO 0 → LAW OF LARGE NUMBERS

ALSO $E\left[\frac{1}{N}\sum X_i\right] = E[X]$

ICA 2: Build-a-nomial

INDEPENDENCE

A "bent" coin turns up tails 60% of the time. If it is flipped 10 times ...

1. What is the probability that it comes up heads 10 times?
2. What is the probability that it comes up tails 10 times?
3. What is the probability that it comes up tails 9 times in a row, then heads finally?

3.5 What is the probability that it comes up tails exactly 9 times (in any order)?

4. What is the probability that it comes up tails exactly 5 times?
5. What is the probability that it comes up tails exactly n times?
6. What is the probability that it comes up tails exactly 7 or more times?

1. $P(X_0=1, X_1=1, X_2=1 \ldots) = P(X_0=1)P(X_1=1)\ldots$

$= .4^{10}$

2. $.6^{10}$

③ $P(X_0=0\ X_1=0\ X_2=0\ \ldots\ X_9=1)$

$= P(X_0=0)P(X_1=0)P(X_2=0)\ \ldots\ P(X_9=1) = .6^9 \cdot .4$

TAILS HEADS

③.5 $(.6^9 \cdot .4) + (.6^9 \cdot .4) + (.6^9 \cdot .4) + \ldots = (10) .6^9 \cdot .4$

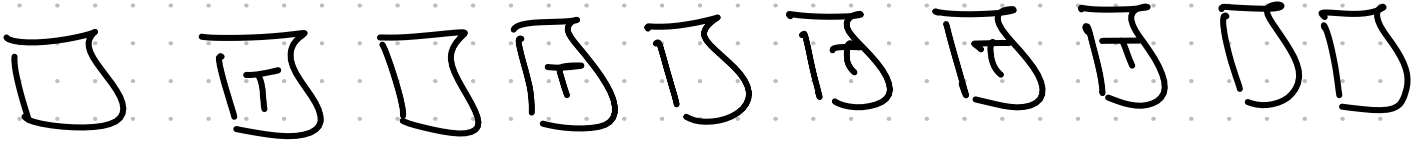
HEADS LAST

HEADS 2ND LAST

HEADS 3RD LAST

THERE ARE 10 POSITIONS PUT 1 HEADS IN 10 FLIPS

4 $\binom{10}{5} .6^5 \cdot .4^5$



$$\binom{10}{5}$$

5 $\binom{10}{n} .6^n \cdot .4^{10-n}$ ← BINOMIAL

"Parametric" distributions

... are "template" distributions. If you satisfy their assumptions you need
only define proper parameters and your problem can import their
well studied behavior to make quick analysis progress!

Big skill:
- match/evaluate assumptions of a parametric distribution to a given problem

BENOULLI "DISTRIBUTION"

HAS A BINARY SAMPLE SPACE

$$P(X=1) = p \qquad P(X=0) = 1-p$$

COIN FLIP, SPORTS SHOT MADE

EVENT HAPPENS | DOESN'T HAPPEN

Binomial Distribution

Sum of N independent identically distributed binary R.V.s

Discrete Positive

Distributed as

$$B \sim P_{Binom}(n, k, p) = \binom{n}{k} p^k (1-p)^{n-k}$$

Total trials

Total trials w/ outcome = 1

Prob each trial has outcome 1

$$E[B] = p \cdot n$$

$$Var(B) = p(1-p)n$$

Examples:
Flip a coin N times, how many coins are heads?

Take N shots, how many goals have you made?

Guess a new randomly drawn card N times, how many correct?

Poisson Distribution

Probability that N events occur in a given period of time. Assumes each event occurs independently of how recent the last event occured.

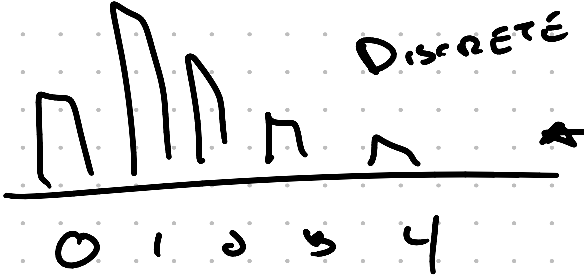
Examples:
- customer arrival in store
- cars arriving at traffic light
- failure rate of windshield wipers (in all of Boston)
- bike accidents in a typical day

$$X \sim Pois(\lambda, k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

RATE: TYPICAL NUMBER OF EVENTS IN A PERIOD OF TIME

OBSERVED NUMBER OF EVENTS IN GIVEN PERIOD OF TIME

DISCRETE, POSITIVE



$$E[x] = \lambda \qquad VAR(x) = \lambda$$

$$\text{Pois}(\lambda = 3 \quad k = 0) = \frac{\lambda^k e^{-\lambda}}{k!}$$

TYPICALLY

3 MUFFLERS / DAY

OBSERVED:

NO MUFFLERS

ICA 3:

For each of the problems below:
- Give the most appropriate parametric distribution for each scenario below
- State any assumptions in the context of the problem
- Evaluate the assumptions, are they reasonable? Do you trust the model?
- Answer the question using the distribution

A car shop typically repairs 3 mufflers a day. Whats the probability they repair no mufflers on a given day?

Poisson. Event somebody comes in with a muffler repair need, is independent of the next person arriving with another muffler repair job.

What is the probability that of 100 babies born in a maternity ward, all are male?

Binomial.

Sex at birth has two options (bernoulli) and we're counting total males (Binomial).

(Use prob / stats calcultor to review)