CS 2810 Day 12
Feb 23

Day 1 of probability / statistics

Why probability? Coin flip discussion

Probability definitions
Uniform Distribution

Joint Distribution
Independence
Linearity of Expectation

Expectation
Linearity of expectation
Variance
Die scaling vs rolling
Linear fnc of random variables

Why Probability?

(modeling coin flip story)

Probability allows us to build a "black box" model of complex events and reason about the future by observing past outputs

Probability Language		EXAMPLE: COIN FLIP
Experiment	Event with uncertain outcome	COIN FLIP
Outcome	The result of an experiment	HEADS
Sample Space	Set of all possible outcomes	{HEADS, TAILS}
Event	subset of sample space	{HEADS}
Random Variable	Function of the outcome(s) of some experiments	GIVEN 2 COIN FLIPS LET $X=1$ IF BOTH HEADS ELSE $X=0$

Probability Language

EXAMPLE DIE ROLL

Experiment	Event with uncertain outcome	ROLLING 6 SIDED DIE
Outcome	The result of an experiment	3
Sample Space	Set of all possible outcomes	$\{1,2,3,4,5,6\}$
Event	subset of sample space	AN ODD OUTCOME $\{1,3,5\}$
Random Variable	Function of the outcome(s) of some experiments	GIVEN 1 DIE ROLL LET $X = 10 \cdot$ OUTCOME

Probability Language

		EXAMPLE WEATHER
Experiment	Event with uncertain outcome	WEATHER TODAY
Outcome	The result of an experiment	RAINY
Sample Space	Set of all possible outcomes	SUNNY RAINY SNOWY
Event	subset of sample space	PRECIPITATES = { RAINY SNOWY }
Random Variable	Function of the outcome(s) of some experiments	GIVEN 1 DAY'S WEATHER $G = 1$ IF RAINY OR SNOWY 0 OTHERWISE

GLOVES NEEDED

Probability Language

Experiment	Event with uncertain outcome
Outcome	The result of an experiment
Sample Space	Set of all possible outcomes
Event	subset of sample space
Random Variable	Function of the outcome(s) of some experiments

NOTATION

LOWERCASE LETTER x

CAPITAL LETTER X

RANDOM VAR

OUTCOME

$$P(X = x)$$

PROB

ICA 1

Let X be a random variable indicating a student's numeric grade on a quiz.

1. Define the sample space of the experiment

$[0, 100]$

2. Define the event of a student passing the quiz

$[60, 100]$

To describe a random event, we often use the term
"random variable"

Let X be the random variable which is the outcome a fair 6 sided die

Let Y be the random variable which is the % increase in stock market today

Let Z be the random variable which is Prof Higger's favorite integer.
(... not uncertain to him, but you can model it as such to encapsulate and manage uncertainty)

Probability Language

LET W BE RANDOM VARIABLE DESCRIBING WEATHER TODAY

Probability How likely a particular outcome is

$$P(W = w_0) = 40\%$$

PROB WEATHER TODAY IS CLOUDY IS 40%

Distribution A function which assigns a probability to each outcome in sample space

$P(\text{CLOUDY}) = 40\%$ $P(\text{RAINY}) = 20\%$ $P(\text{SUN}) = 40\%$

Probability Axioms

Axiom: a property or statement taken to be true without argument

1. The probability of every event is between 0 and 1 (including 0 and 1)

2. The probability of the entire sample space 1

DISTRIBUTION OVER WEATHER

$P(\text{cloudy}) = 40\%$ $\quad P(\text{rainy}) = 20\%$ $\quad P(\text{sun}) = 40\%$

$X_0 = \text{cloudy}$ $\quad X_1 = \text{rainy}$ $\quad X_2 = \text{sun}$

$$P(X_0) + P(X_1) + P(X_2) = 1$$

DISTRIBUTION

$$\sum_x P(X = x) = 1$$

RANDOM VARIABLE

PARTICULAR OUTCOME

SUM OVER ALL OUTCOMES IN SAMPLE SPACE

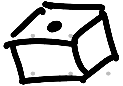
Uniform Distribution

Uniform distribution assigns equal prob to all outcomes in sample space

Fair coin

H	T
50%	50%

Fair die

	2	3	4	5	6
1/6	1/6	1/6	1/6	1/6	1/6

For uniform: $P(x) = 1/|S|$ ← S is sample space

$|S|$ is # elements in S

COMPUTING PROB FROM UNIFORM DISTRIBUTION

GIVEN A FAIR DIE COMPUTE PROB OF EACH EVENT

X = ROLL A 1	Y = ROLL AN EVEN #	Z = ROLL A PRIME #

ICA 2 On average, which lotto ticket yields most money?

(D)ouble Lotto

P(D)	D
1/2	\$2
1/2	\$0

$1/2 \cdot 2 + 1/2 \cdot 0 = 1$

(S)teady Lotto

P(S)	S
1/2	\$.9
1/2	\$1.1

$1/2 \cdot .9 + 1/2 \cdot 1.1 = 1$

Shoot for the (M)oon Lotto

P(M)	M
1/1000	\$1000
999/1000	\$0

$1/1000 \cdot 1000 + \frac{999}{1000} \cdot 0 = 1$

EXPECTED VALUE

EXPECTED VALUE REPRESENTS AN "AVERAGE" OUTCOME OF A RANDOM VARIABLE

"DOUBLE LOTTO"

$S = \{2, 0\}$

$P(w)$	w
$1/2$	\$2
$1/2$	\$0

HALF TIME WIN \$2

HALF TIME 'WIN' \$0

ELEMENTS IN SAMPLE SPACE

$\$1 = 2 \cdot 1/2 + 0 \cdot 1/2$

PROB OF EACH

X

$S = \{0, 1, 2\}$

EXPECTED VALUE OF RANDOM VARIABLE X

$$E[X] = 0 \cdot P(X=0) + 1 \cdot P(X=1) + 2 \cdot P(X=2)$$

$$E[X] = \sum_{x \in S} x \cdot P(x)$$

PROB OF OUTCOME

VALUE OF RANDOM VARIABLE

ALL OUTCOMES IN SAMPLE SPACE

$$E[10x] = \sum_{x} 10x \, P(X=x)$$

PROB OF EACH OUTCOME

SUM OVER OUTCOMES

ICA 3

ON AVERAGE, WHICH LOTTO TICKET YIELDS THE MOST CONSISTENT OUTPUT?

(D)OUBLE LOTTO

P(D)	D
1/2	\$2
1/2	\$0

(S)TEADY LOTTO

P(S)	S
1/2	\$.9
1/2	\$1.1

SHOOT FOR THE (M)OON LOTTO

P(M)	M
1/1000	\$1000
999/1000	\$0

PROB	MOON 2 PAYOUT
1/2	$1000
1/2	$0

MOST CONSISTENT RANDOM VARIABLE HAS OUTCOMES OFTEN CLOSE TO EXPECTED VALUE

How close, on average, is Double Lotto to its expected value?

$P(D)$	D	$D-E[D]$	$(D-E[D])^2$
$1/2$	0	-1	1
$1/2$	2	1	1

$E[D] = 1/2 \cdot 0 + 1/2 \cdot 2 = 1$

$E[(D-E[D])^2] = 1 \cdot 1/2 + 1 \cdot 1/2 = 1$

HOW CLOSE, ON AVERAGE, IS MOON LOTTO TO ITS EXPECTED VALUE?

$P(D)$	D	$D-E[D]$	$(D-E[D])^2$
$1/1000$	1000	999	999^2
$999/1000$	0	-1	1

$E[D]=1$

$$E[(D-E[D])^2]$$

$$= 999^2 \cdot \frac{1}{1000} + 1 \cdot \frac{999}{1000}$$

$$\cong 1000 + 1$$

VARIANCE (AND STANDARD DEVIATION)

VARIANCE (AND STD DEV) DESCRIBE HOW CLOSE OUTCOMES TYPICALLY ARE TO EXPECTED VALUE

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

$$\mathrm{STD\ DEV}(X) = \sqrt{\mathrm{VAR}(X)}$$

NOTATION / VOCAB

"Mu"

We often use μ to represent expected value $E[x]$

σ^2 to represent variance $E[(x-\mu)^2]$

"Sigma"

σ to represent std dev

ICA 4

Compute the expected value, variance and standard deviation of D where D is a random variable representing a fair (uniform) 6 sided die roll. (You may find a calculator helpful)

(+) How does the expected value / variance of D compare to another die which has 1000 sides? Justify your response algebraically or with a thoughtful appeal to one's intuition.

$$E[D] = \sum_{d} d\, P(D=d) = 1\cdot\frac{1}{6} + 2\cdot\frac{1}{6} + 3/6 + 4/6 + \frac{5}{6} + \frac{6}{6}$$

$$= 21/6$$

$P(D)$	D	$(D - E[D])^2$
$1/6$	1	$(1 - 21/6)^2$
$1/6$	2	$(2 - 21/6)^2$
$1/6$	3	$(3 - 21/6)^2$
$1/6$	4	$(4 - 21/6)^2$
$1/6$	5	$(5 - 21/6)^2$
$1/6$	6	$(6 - 21/6)^2$

$$E\left[(D - E[D])^2\right]$$

$$= 1/6 \left(1 - \frac{21}{6}\right)^2 +$$

$$1/6 \left(2 - \frac{21}{6}\right)^2 +$$

$$\cdots$$

LINEARITY OF EXPECTATION

LET X, Y BE TWO RANDOM VARIABLES WITH JOINT $P(x, y)$
AND α, β BE TWO SCALARS

$$E[\underbrace{\alpha X + \beta Y}] = \underbrace{\alpha E[X] + \beta E[Y]}$$

SCALE/ADD RANDOM VARS <u>BEFORE</u> EXPECTATION OPERATOR

SCALE/ADD RANDOM VARIABLES <u>AFTER</u> EXPECTATION OPERATOR

LINEARITY OF EXPECTATION (PROOF)

LET X, Y BE TWO RANDOM VARIABLES WITH JOINT $P(X, Y)$

AND α, β BE TWO SCALARS

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

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

MARGINALIZE!

MARG

(WE'LL COVER MARGINALIZATION IN NEXT LESSON)

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

$$= \alpha E[x] + \beta E[y]$$

Another way of expressing / computing variance

$$\sigma_x^2 = VAR(x) = E\left[(x - E[x])^2\right]$$ ← Old way

$$= E\left[x^2 - 2xE[x] + E[x]^2\right]$$

Apply linearity of expectation

$$= E[x^2] - 2E[x]E[x] + E[x]^2$$

$$= E[x^2] - 2E[x]^2 + E[x]^2$$

$$= E[x^2] - E[x]^2$$ ← New way

Validating New Variance Formula on Double Lotto

$P(D)$	D	$D - E[D]$	$(D - E[D])^2$	D^2
$1/2$	0	-1	1	0
$1/2$	2	1	1	4

$0 \cdot 1/2 + 4 \cdot 1/2 = 2$

$E[D] = 0 \cdot 1/2 + 2 \cdot 1/2 = 1$

$$\mathrm{VAR}(D) = E\left[(D - E[D])^2\right] = 1 \cdot 1/2 + 1 \cdot 1/2 = 1$$

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

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

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

$$E[CX] = CE[X]$$

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

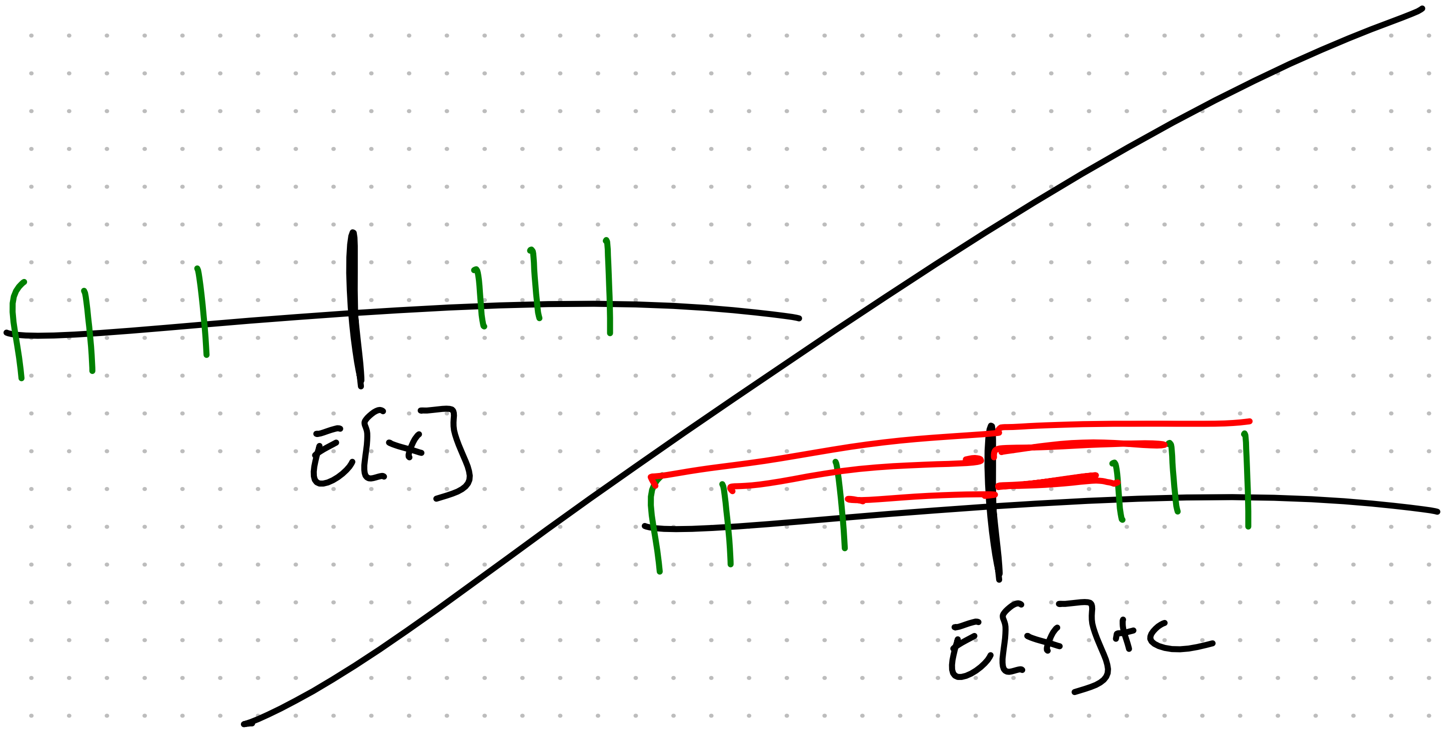
$$\mathrm{VAR}(CX) = C^2\,\mathrm{VAR}(X)$$

$$\mathrm{VAR}(cx) = E[(cx)^2] + E[cx]^2$$

$$= c^2 E[x^2] + c^2 E[x]^2$$

$$= c^2 \left(E[x^2] + E[x]^2 \right)$$

$$= c^2 \, \mathrm{VAR}(x)$$



E[x]
E[x]+c

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

$$E[cX] =$$

$$\mathrm{Var}(X+C)$$
$$= E[(X+C)^2] - E[X+C]^2$$
$$= E[X^2 + 2cE[X] + c^2] - (E[X]+c)^2$$
$$= E[X^2] + 2cE[X] + c^2$$
$$- E[X]^2 - 2cE[X] - c^2$$
$$= E[X^2] - E[X]^2 = \mathrm{Var}(X)$$

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

$$E[X+Y] = E[X] + E[Y] \qquad VAR(X+Y) = ?$$

NEXT LESSON

(NEED NOTION OF INDEPENDENCE)