DEPARTMENT OF MATHEMATICS, NORTHEASTERN UNIVERSITY

MTH G7241: Probability 1, Section 1 Fall 2022

Class: Hayden 221, together with some online classes
Instructor: Prof. Chris King
Office: 455 Lake; c.king@northeastern.edu
Office hours: TBA
Course Webpage: on Canvas

Recommended Texts and other Sources:

• "Introduction to Probability models", S. Ross, nth edition (published by Academic Press).

• "Introduction to Probability", C. M. Grinstead and J. L. Snell, second revised edition (published by American Mathematical Society). Also available online from the CHANCE project, can be downloaded (free!) at http://www.dartmouth/˜chance/.

• "Markov chains" from the Cambridge Stats Lab http://www.statslab.cam.ac.uk/ rrw1/markov/M.pdf

• Supplementary notes on Probability will be available on course webpage.

Grading:
Homework problems will be assigned each week or two. They will be partially graded, and solutions will be provided. There will be several quizzes or tests, and also one or more projects.

Grade breakdown will be approximately 60% for Homeworks/Projects, 40% for Tests/Quizzes/Final.

Course content: The course covers basic probability theory with emphasis on applications to modeling and data analysis. Topics will include finite state Markov chains, Bayesian inference, Markov Chain Monte Carlo, Poisson process, along with data-driven projects. No prior knowledge of probability theory will be assumed.

List of main topics:

1). Basics: Random variables, Independence, Conditioning, Law of Large Numbers, Central Limit Theorem.
2). Finite State Markov chains: Classification of chains, Stationary distribution, Time reversible chains, Absorbing chains, Sojourn times.
3) Bayesian inference: Parameter estimation, Hypothesis testing, multinomial estimation.
4) Sampling and Markov chain Monte Carlo: Rejection method, MCMC algorithm, sampling from multinomial, Uniform, Knapsack problem, application to Bayesian inference.
5) Continuous time Markov chains: Poisson process, Superposition and Thinning Theorems, birth-death model.
6) Selected Additional Topics (as time allows): Branching process, hidden Markov models and Viterbi algorithm.

University Academic Integrity Policy: The university's academic integrity policy at OSCCR (http://www.northeastern.edu/osccr/academic-integrity-policy) discusses actions regarded as violations and their consequences for students.

Title IX: The University strictly prohibits sex or gender discrimination in all university programs and activities. Information on how to report an incident of such discrimination (which includes sexual harassment and sexual assault) is located at http://www.northeastern.edu/titleix.

Students with Disabilities: Students who have disabilities who wish to receive academic services and accommodations should follow the standard Disabilities Resource Center (DRC) procedures, http://www.northeastern.edu/drc/getting-started-with-the-drc.

College of Science Policies: The current College of Science Academic Course Policies are available at https://cos.northeastern.edu/wp-content/uploads/2012/10/COS-teaching-policies-April-2017.pdf

MTH 7241 Fall 2022: Prof. C. King

Notes 1: Basics of probability

A number X generated by a random experiment is called a **random variable**. To understand X you must know its **range** $\mathrm{Ran}(X)$ (the set of possible values for X) and its **pdf** (the probabilities for those values).

Discrete Random Variables

If X is *discrete* then $\mathrm{Ran}(X)$ is finite or countably infinite, and the pdf p_X is the list of probabilities for each value. The pdf of a discrete r.v. X is a list of probabilities $\{p_X(x)\}$ where the index x runs over all possible values of X. If the range of X is small then this list can be written down explicitly. If the range of X is large or infinite then the pdf is given by a formula. The normalization condition says that the total probability for X must be one, that is

$$0 \le p_X(x) \le 1$$

impossible

certain

$$\sum_x p_X(x) = 1$$

where the sum runs over all possible values of X.

Example 1 (Bernoulli) *A Bernoulli random variable takes only two values, so it is determined by a single probability. The possible values are usually labeled* $\{0, 1\}$ *so the pmf is*

$$p_X(0) = \mathbb{P}(X = 0) = 1 - p, \quad p_X(1) = \mathbb{P}(X = 1) = p.$$

$$\mathrm{Ran}(X) = \{0, 1\}$$

p is a number between 0 and 1

$p \in [0, 1]$

pdf of X

X	0	1
prob.	$1-p$	p

← row sum $= 1$.

normalization condition

$$1 - p + p = 1 \quad ✓$$

Bernoulli r.v.

one-parameter family of r.v.'s.

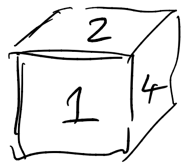
Example

X	0	1	2
prob.	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{3}$

$$\text{Ran}(X) = \{0, 1, 2\} \Rightarrow \text{discrete r.v.}$$

normalization : $\frac{1}{6} + \frac{1}{2} + \frac{1}{3} = 1$ ✓

Example 2 *Roll of a fair die, sum of two rolls, maximum of two rolls.*

Fair die: 

cube, numbers 1, .., 6 on faces.

Roll die: X = number on top face

$$\mathrm{Ran}(X) = \{1,2,3,4,5,6\}$$

$\downarrow$

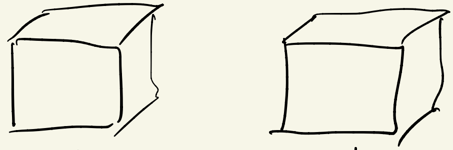
finite $\Rightarrow$ X is discrete r.v.

pdf:

X	1	2	3	4	5	6
prob.	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

"Fair die" $\Longleftrightarrow$ all outcomes equally likely.

Roll two dice:



top face $\downarrow$ X_1 $\downarrow$ top face. X_2

Sum of rolls:

$$S = X_1 + X_2$$

S	2	3	4	5	6	7	8	9	10	11	12
prob.	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{26}$	$\frac{2}{36}$	$\frac{1}{36}$

2: (1,1) $X_1 = 1$, $X_2 = 1$, prob. $= \frac{1}{36}$

3: (1,2) (2,1)

4: (1,3) (2,2) (3,1)

5: (1,4) (2,3) (3,2) (4,1)

8: (2,6) (3,5) (4,4) (5,3) (6,2)

all 36 ordered pairs are equally likely

<u>Note</u>: for a repeated trial, almost always easiest to count by using <u>ordered</u> sequences of outcomes. <u>Reason</u>? b/cause they will be equally likely.

Basic Rules for finding probabilities of combined events.

Sample space = set of all possible outcomes.

e.g. roll two dice,

sample space $= \{(1,1), (1,2), (1,3), \ldots$

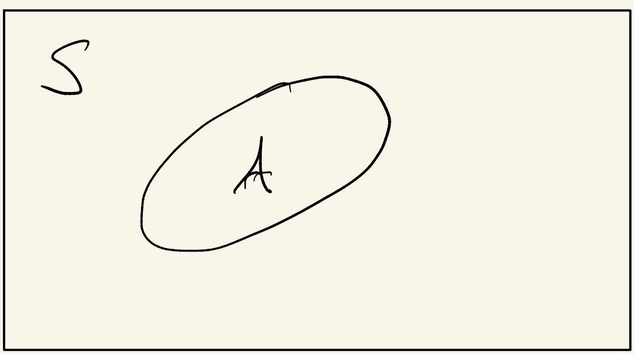
$(2,1), (2,2)$ —

$\vdots$

$(5,6), (6,6)\}$

$S = \{$all ordered pairs for two dice$\}$.

$|S|$ = number of outcomes $(= 36)$



A = event = subset of S.

e.g. roll two dice:

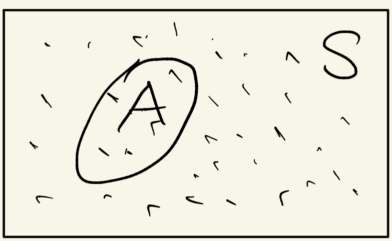
event $A = \{X_1 = 3\} = \{\text{first dice is } 3\}$.

$$= \{(3,1), (3,2), \dots, (3,6)\}$$

$|A|$ = size of A $(= 6)$

If all outcomes in S are equally likely, then

$$P(A) = \text{probability of } A = \frac{|A|}{|S|}$$



$$A \subset S$$

- randomly pick one outcome
- all equally likely.

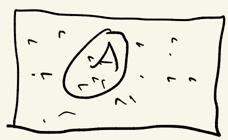
If all outcomes are <u>not</u> equally likely, then we have a list of probabilities for outcomes.

$$S = \{ s_1, s_2, s_3, s_4, \dots, s_n \}.$$

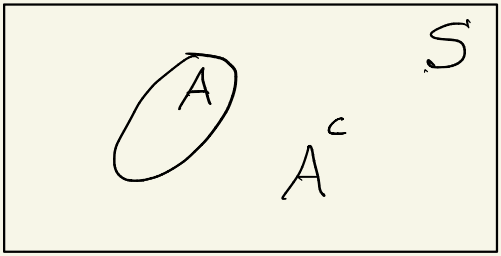
<u>probs</u>: $p(s_1), \ p(s_2), \ p(s_3), \ p(s_4), \dots, \ p(s_n)$.

Normalization: $\displaystyle\sum_{i=1}^{n} p(s_i) = 1.$

If $A \subset S$, then



$$P(A) = \sum_{\underbrace{i:\, s_i \in A}_{\text{sum over outcomes in } A}} p(s_i)$$



$A \subset S$

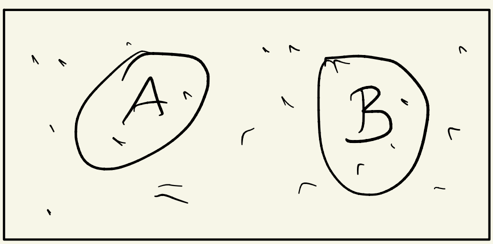
$$A^c = S \setminus A = \{ s \in S : s \notin A \}.$$

complement of A
aka
not A

$P(A)$ = prob. of A

$P(A^c)$ = prob. of A^c.

$$\Rightarrow \boxed{P(A) + P(A^c) = 1}$$



A, B disjoint events

$$\Rightarrow A \cap B = \emptyset$$

$A \cap B$ = intersection of A, B

$= \{ s \in A$ <u>and</u> $s \in B \}$.

ϕ = empty set.

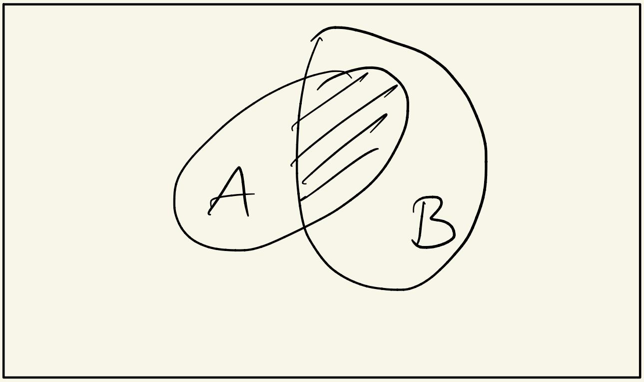
$A \cup B$ = union of A, B

$= \{ s \in A$ <u>or</u> $s \in B \}$.

$P(\phi) = 0$.

$$\boxed{P(A \cup B) = P(A) + P(B)} \, *$$

$*$ when $A \cap B = \phi$.



$A \cap B$ is not empty.

$$\boxed{P(A \cup B) = P(A) + P(B) - P(A \cap B)}$$

always holds.

3 operations: $\begin{cases} \text{complement} \\ \text{intersection} \\ \text{union} \end{cases}$

Roll two dice:

X_1 = first number

X_2 = second number

$$M = \max(X_1, X_2).$$

pdf of M?

M	1	2	3	4	5	6
prob	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$

(1,1)

(1,2)
(2,1)
(2,2)

(1,3)
(2,3)
(3,3)
(3,2)
(3,1)

<u>Normalization</u>: $\frac{1}{36} + \frac{3}{36} + \frac{5}{36} + \cdots + \frac{11}{36} = 1$ ✓

M is discrete, pdf is <u>not</u> uniform.

Example 3 (Binomial) *We write $X \sim \mathrm{Bin}(n, p)$ to indicate that X is a binomial r.v. where n is the number of independent trials and p is the probability of success on each trial. The formula for the pdf is*

$$p_X(k) = \mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n$$

The standard example comes from coin tossing. Suppose a coin is tossed n times, and p is the probability of Heads on each toss (if $p = 1/2$ the coin is fair, otherwise it is biased). Let X be the number of times the coin comes up Heads.

HHHH, THHH, HTHT, TTHT,
HHHT, HHTT, THHT, THTT,
HHTH, TTHH, HTTH, HTTT,
HTHH, THTH, TTTH, TTTT

$n=4$
toss 4 times
16 outcomes $= 2^4$

Then $X \sim \mathrm{Bin}(n, p)$. The explicit formula for p_X is useful if n and k are not too large. For example, an airline knows that 5% of people will not show up for a flight, so they overbook 52 people on a plane with 50 seats. What is the probability that nobody is bumped off the flight?

$X =$ number of people who do not show up

$p = 0.05$, $n = 52$

Blue outcomes = 2 Heads & 2 Tails.

6 Blue outcomes.

Why 6 ?

H T T H

How many outcomes with 2 Heads ?

Count the number of ways to select a subset of size 2 from the 4 coins.

4 objects, select a subgroup of 2

Number of ways

$$= \text{Binomial coefficient}$$

$$= \binom{4}{2} \qquad \text{"4 choose 2"} \quad {}^4C_2$$

$$= \frac{4!}{(4-2)!\,2!}$$

$$= \frac{(4)(3)(2)(1)}{(2)(1)(2)(1)} = 6$$

X = number of Heads in 4 tosses

$$P(X=2) = P(\text{Blue outcomes})$$

$$= 6\,p^2(1-p)^2$$

Airline problem: $n = 52$ (capacity of plane $= 50$ seats)

$p = 0.05$

$X =$ number who do not show up

$X \sim Bin(52, 0.05)$

P(nobody is bumped from the flight)

$$= P(X = 2, 3, 4, 5, \ldots, 52)$$

$$= P(X \geq 2)$$

$$= 1 - P((X \geq 2)^c)$$

$$= 1 - P(X < 2)$$

$$= 1 - P(X=0) - P(X=1)$$

$$= 1 - \binom{52}{0} p^0 (1-p)^{52-0} - \binom{52}{1} p^1 (1-p)^{52-1}$$

$$\binom{n}{0} = 1 \qquad \binom{n}{1} = n$$

$$= 1 - 1.(1-p)^{52} - 52.\, p.(1-p)^{51}$$

$$= 1 - (0.95)^{52} - 52(0.05)(0.95)^{51}$$

$$= 0.74$$

Example 4 (Poisson) *A Poisson random variable has infinite range* $0, 1, 2, \ldots$. *The pdf is given by the formula*

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

Discrete

where $\lambda > 0$ *is a fixed parameter. Note that the Poisson is a useful approximation for the binomial* $\mathrm{Bin}(n, p)$ *(and is much simpler) in the case where* n *is large and* p *is small. The binomial convergence result is*

$$\lim_{\substack{n\to\infty \\ p\to 0 \\ np=\lambda}} \binom{n}{k} p^k (1-p)^{n-k} = \frac{\lambda^k}{k!}\, e^{-\lambda}$$

Compare Poisson value with exact result for previous example.

$\mathrm{Ran}(X) = \{0, 1, 2, 3, 4, \ldots\}$

$\lambda =$ lambda positive constant,

pdf: $P(X=k) = \frac{\lambda^k}{k!} e^{-\lambda}$

$k = 0, 1, 3 \ldots$

Normalization:

$$\sum_{k=0}^{\infty} P(X=k) = 1 \quad ?$$

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

Taylor series of e^x

$$= e^{-\lambda} e^{\lambda}$$

$$= 1.$$

Check approximation for airline prob.

$n = 52$

$p = 0.05$

$\lambda = np = 2.6$

$$1 - P(X=0) - P(X=1)$$

$$= 1 - \frac{(2.6)^0}{0!} e^{-2.6} - \frac{2.6}{1!} e^{-2.6}$$

$$= 1 - e^{-2.6} - (2.6)\, e^{-2.6}$$

$$= 0.73$$

close to exact value 0.74.

- easier to compute.