Stat 88 lec 17

warmup: 2:00- 2:10

A lost tourist arrives at a point with 2 roads. Road A brings him back to the same point after 1 hour of walk. Road B leads to the city in 2 hours. Assuming the tourist randomly chooses a road at all times, what is the expected time until the tourist arrives to the city?

T = time to arrive in city

$E(T)$

$R = \text{road} \begin{cases} A \\ B \end{cases}$

$$E(T) = \underbrace{E(T|R=A)}_{=1+E(T)}\underbrace{P(R=A)}_{=1/2} + \underbrace{E(T|R=B)}_{=2}\underbrace{P(R=B)}_{=1/2}$$

$$E(T) = \tfrac{1}{2}\left(1 + E(T)\right) + 1$$

$$2E(T) = 1 + E(T) + 2$$

$$\boxed{E(T) = 3 \text{ hrs}}$$

Last time

Conditional expectation:

	$M=2$	$M=3$
$S=2$	0	$\frac{1}{3}$
$S=1$	$\frac{1}{3}$	$\frac{1}{3}$

$$\boxed{E(S \mid M=m) = \sum_{\text{all } s} s \cdot P(S=s \mid M=m)}$$

and

$$\boxed{E(S) = \sum_{\text{all } m} E(S \mid M=m) P(M=m)}$$

or equivalently:

$$\boxed{E(S) = E(E(S \mid M))}$$

Today
1. Sec 5.6 more practice with conditional expectation
2. Midterm review

Next time Enter specific questions on b-courses discussion board by 8pm Tuesday to go over Wednesday in class.

① Sec 5.6 More practice with conditional expectation

Tamara chooses an integer X uniformly at random from 1 to 425. She then chooses an integer Y uniformly at random from 1,...,X. Find E(Y).

Hint: $E(Y) = E(E(Y|X))$

$Y|X \sim \text{Unif}\{1,2,\dots,X\}$, $X \sim \text{Unif}\{1,\dots,425\}$

$Y|X=x \sim \text{Unif}\{1,2,\dots,x\}$

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

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

$= \frac{1}{2} + \frac{E(X)}{2}$ $E(X) = \frac{425+1}{2} = 213$

$= \frac{214}{2} = \boxed{107}$

2) Midterm review

a) reference sheet (provided on exam)

DISTRIBUTION FACTS

Name and Parameters	Values	$P(X=k)$	$E(X)$
Bernoulli (p) (also Indicator)	$0,1$	$P(X=1)=p$	p
Uniform on 1 through N	$1,2,\dots,N$	$\frac{1}{N}$	$\frac{N+1}{2}$
Binomial (n,p)	$0,1,2,\dots,n$	$\binom{n}{k}p^k(1-p)^{n-k}$	np
Hypergeometric (N,G,n)	$0,1,2,\dots,n$	$\frac{\binom{G}{k}\binom{N-G}{n-k}}{\binom{N}{n}}$	$n\frac{G}{N}$
Poisson (μ)	$0,1,2,3,\dots$	$e^{-\mu}\frac{\mu^k}{k!}$	μ
Geometric (p)	$1,2,3,\dots$	$(1-p)^{k-1}p$	$\frac{1}{p}$

- If X has the Poisson (μ) distribution, Y has the Poisson (λ) distribution, and X and Y are independent, then $X+Y$ has the Poisson $(\mu+\lambda)$ distribution.

CLASS CONVENTIONS

- Unless otherwise stated, coins are two-sided and fair, and dice are six-sided and fair.
- "At random" means uniformly at random; all outcomes equally likely.
- "The number of trials till an event occurs" means the number of trials up to and including the trial at which the event occurs for the first time.

ACRONYMS

- i.i.d.: independent and identically distributed
- SRS: simple random sample (drawn at random without replacement)
- cdf: The cumulative distribution function of X is the function F defined by $F(x)=P(X\le x)$, $-\infty<x<\infty$

MATH FACTS

- Exponents: $x^ax^b=x^{a+b}$, $x^ay^a=(xy)^a$
- Pascal's triangle: For integers $0\le k\le n$, $\binom{n}{k} = \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!} = \frac{n!}{k!(n-k)!}$
- Geometric series: For $0<r<1$, $\sum_{k=0}^{\infty} r^k = 1+r+r^2+r^3+\cdots = \frac{1}{1-r}$
- Exponential series: For all x, $\sum_{k=0}^{\infty}\frac{x^k}{k!} = 1+x+\frac{x^2}{2!}+\frac{x^3}{3!}+\cdots = e^x$
- Logs: $\log(x^n)=n\log(x)$, $x=e^{\log(x)}$
- For exponential approximations: $\log(1-x)\approx -x$ for small x

Note: the midterm isn't very mathy even though this reference sheet is.

Preparation

Study the textbook (including the Exercises at the end of each chapter), homework, and quizzes. As I keep saying till you are bored to death, know the conditions under which you can do the different kinds of calculations. I am confident you will agree after the exam that nothing more was needed.

ex Binomial(n,p) formula is used for
successes out of n independent p - trials

Geometric (p) formula is used for #
independent p-trials until a success.

During the exam

• I expect that the vast majority of you will move swiftly through the early questions in the exam and therefore have time to think about the later ones. Please don't over-think straightforward questions.I once had a Stat 2 student give me something like 1682.01733 as the answer to, "If you toss a coin 200 times, how many heads do you expect?"

• **Read each question carefully.** As you know, assumptions matter. If you have misread those then your solution will be off. For example, confusing "with replacement" and "without replacement" can have a massive effect. "The fourth head is on the 20th toss" is not the same as, "There are four heads in 20 tosses." Forcing yourself to read slowly, underlining key assumptions as you read, is important for doing well. There are no prizes for finishing early. If you are done with the test, check your work by reading each question afresh and solving it again instead of just reading over your answer.

• Provide reasoning or a calculation in all questions. If you did a calculation in your head, write out the calculation you did in your head.

• Don't simplify any arithmetic or algebra in your answers unless a question explicitly asks you to. The unsimplified version shows us your thought process. Simplification isn't worth the time on the test, and besides, you might mess up the simplification.

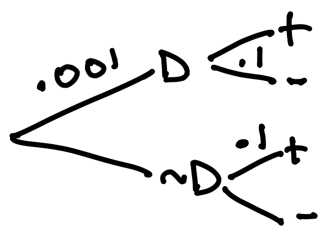
• I'm very conscious that it's a 45-minute exam and have designed questions accordingly. If an asnwer is taking you numerous lines of calculation, or complicated algebra/calculus, you've probably missed something. Rethink, or move to a different problem.

• If you don't know how to do a problem, try not to leave it blank. Almost always, you will have an idea of what might be relevant. If you write that, and it is indeed important for the problem, you might get some partial credit. That said, you shouldn't expect partial credit for everything you write. We'll be looking for substantive ideas and progress towards a solution.

Problem Solving Suggestions

• If the experiment involves two stages, draw a tree diagram.

ex
D = event you have disease
+ = positive test for disease.
$P(D) = .001$
Chance of false pos and false negative are .1
To find $P(D|+)$ make a tree



• If you find yourself writing something like 0.3^5, check that you have independence. You can't just assume independence, and it's wrong for situations like dealing cards etc. Either the problem has to provide an assumption of independence, or the assumption has to follow from the conditions of the experiment, e.g. sampling with replacement, or events based on separate sets of tosses, etc.

When faced with a complicated probability, try and describe the outcome space in simple English

ex

Draw cards from a standard deck until 3 aces have appeared.
Let X = number of cards drawn.
Find P (X>x)

drawing 0, 1, 2 aces in x draws.

$$P(X>x) = \underbrace{\frac{\binom{4}{0}\binom{48}{x}}{\binom{52}{x}}}_{0 \text{ aces}} + \underbrace{\frac{\binom{4}{1}\binom{48}{x-1}}{\binom{52}{x}}}_{1 \text{ ace}} + \underbrace{\frac{\binom{4}{2}\binom{48}{x-2}}{\binom{52}{x}}}_{2 \text{ aces}}$$

• If you are using the probability formula for one of the famous distributions on the reference sheet, pause for a second and check the possible values. For example, this might prevent you from using the geometric (possible values 1, 2, 3, . . .) when you should be using the binomial (possible values 0 through n).

ex X = # of blueberries in a bite of a blueberry muffin.
What are the possible values of X — 0, 1, 2, ...
What kind of RV is X? — Poisson

• If a question asks for a distribution, start by listing all the possible values of the random variable. Take the time to do this carefully, identifying the minimum possible value (if there is one) and the maximum possible value (if there is one). Not only will it help focus your calculation of the probabilities, it might also get you partial credit for having understood the variable even if you didn't get the probabilities correctly.

ex

A deck consists of 12 cards, 5 of which are blue and 7 green. I draw from the deck at random without replacement till both colors have appeared among the draws. Let D be the number of draws. Find the distribution of D.

Possible values of D. — 2, 3, 4, ..., 8

Find P(D) for a couple values of D.

$P(D=3)$ bbg or ggb $\frac{5}{12}\cdot\frac{4}{11}\cdot\frac{7}{10} + \frac{7}{12}\cdot\frac{6}{11}\cdot\frac{5}{10}$

$P(D=8)$ $\frac{\binom{5}{2}\binom{7}{0}}{\binom{12}{2}}\cdot\frac{7}{10} + \frac{\binom{7}{2}\binom{5}{0}}{\binom{12}{2}}\cdot\frac{5}{10}$

ggggggg b

$\frac{\binom{7}{7}\binom{5}{0}}{\binom{12}{7}}\cdot\frac{5}{5}$

• If a question asks for an expectation, don't immediately try to find the distribution of the random variable and then apply the defintion of expectation. Try properties of expectation first. The most common are additivity (see if you can write the random variable as a sum of simpler ones) and conditioning (see if you would know the expectation if you were given the result of an early stage of the experiment).

ex

A fair die is rolled 14 times. Let X be the number of faces that appear exactly 2 times. Find E(X).

X = # of faces (out of 6) that appear 2 times

p

$$I_2 = \begin{cases} 1 & \text{if 2nd face appears twice} \\ 0 & \end{cases}$$

$$p = \binom{14}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^{12}$$

$$X = I_1 + \cdots + I_6$$

$$E(X) = 6p$$

- The conditional distribution of Y given $X = x$ is just an ordinary distribution. You have to first recognize that "given $X = x$" means you can treat the random variable X as the constant x. You therefore have to provide the possible values of Y (under the condition that $X = x$) and the corresponding conditional probabilities given $X = x$. How you calculate those conditional probabilities depends on the setting. Sometimes you can just see what they are because the condition $X = x$ simplifies your outcome space, and sometimes you have to use the division rule $P(Y = y \mid X = x) = P(X = x, Y = y)/P(X = x)$. This is not particular to finding conditional distributions. It's a feature of finding conditional probabilities in general.

ex

Let $\left.\begin{array}{l} X_1 \sim \text{Pois}(a) \\ X_2 \sim \text{Pois}(b) \end{array}\right\}$ independent.

let m be a fixed positive integer.

Find the distribution of

$X_1 \mid X_1 + X_2 = m$

Recognize this as one of the famous ones and provide its parameters.