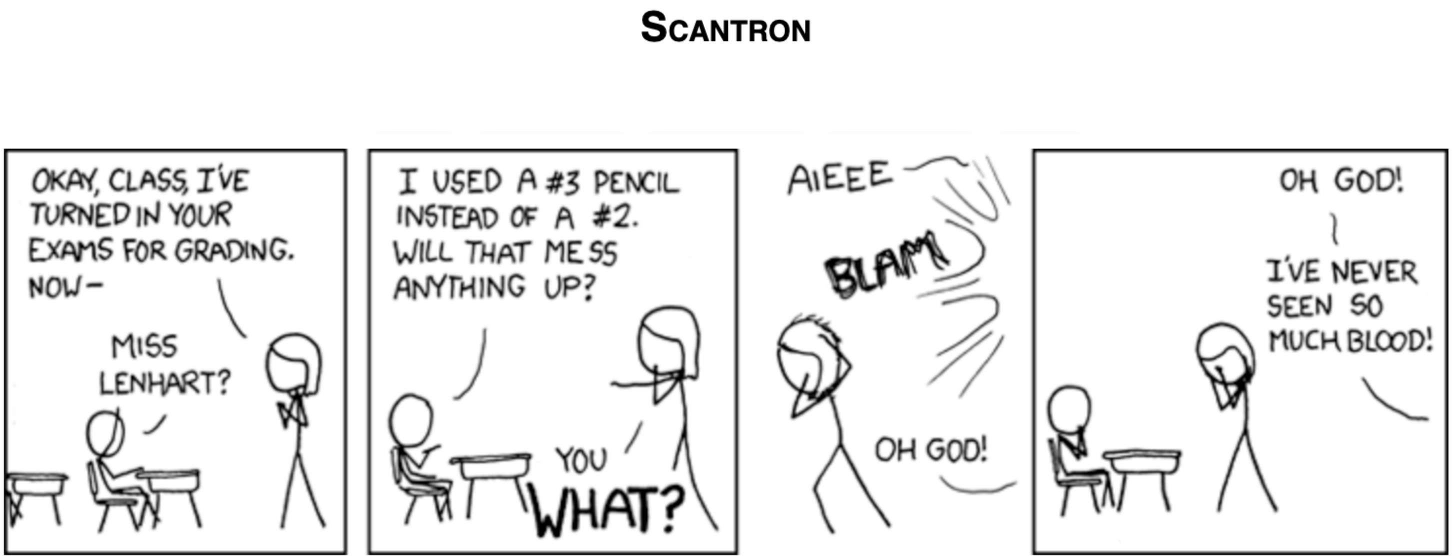
Stat 88: Probability & Math. Statistics in Data Science



https://xkcd.com/499/

Lecture 18: 3/3/2021

Talking about the midterm

All who wander might be lost…

A lost tourist arrives at a point with 2 roads. Road A brings him back to the same point after 1 hour of walking. 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 (in hours) to arrive at the city

$\mathbb{E}(T) = ?$

R = Road: A w/pr 1/2, B w/pr 1/2

$$\underbrace{\mathbb{E}(T)}_{\text{call this } t} = \mathbb{E}(\mathbb{E}(T|R)) = \underbrace{\mathbb{E}(T|R=A)}_{(\underbrace{\mathbb{E}(T)}_{t}+1)}\underbrace{P(R=A)}_{\frac{1}{2}} + \underbrace{\mathbb{E}(T|R=B)P(R=B)}_{2\cdot\frac{1}{2}}$$

$$t = (t+1)\cdot\frac{1}{2} + 1$$

$$\frac{1}{2}t = \frac{3}{2} \Rightarrow t = 3 \text{ hrs}$$

$\mathbb{E}(T) = 3$ hrs.

Expectation computation.

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

$$X \sim \text{Unif}\{1, 2, \ldots, 425\} \rightarrow \mathbb{E}(X) = \frac{425+1}{2} = \frac{426}{2}$$

$$Y \sim \text{unif}\{1, 2, \ldots X\}$$

What is the dsn of $Y|X$?

If I fix X to be n,

then $Y | X = n \sim \text{Unif}\{1, 2, \ldots n\}$

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

$$E(Y) = E(E(Y|X)) = \sum_n E(Y|X=n) \cdot P(X=n)$$

$$= \sum_n \frac{(n+1)}{2} \cdot P(X=n) = \frac{1}{2} \sum_n n P(X=n) + \frac{1}{2} \underbrace{\sum_n P(X=n)}_{=1}$$

$$\frac{1}{2} E(X) + 1 \cdot \frac{1}{2} \qquad \frac{1}{2} \cdot 1$$

$$= \frac{1}{2}\left(\frac{426}{2}\right) + \frac{1}{2} = \frac{1}{2}(214) = 107$$

Tamara, again

- Tamara chooses an integer *N* uniformly at random from *Pois(μ)*. She then picks *N* cards from a deck with replacement. Find the expected number of ace cards.

N is picked unif a rdm from a Poisson dsn.

X = # of aces among N cards picked

inside exp. summing over values of X, given n

$$E(X) = E(E(X|N)) = \sum_{n} \underbrace{E(X|N=n)} P(N=n)$$

outside expectation is summing over values of N

all possible values taken by N.

$X|N=n \sim Bin\left(n, \frac{4}{52}\right)$

$E(X|N=n) = \boxed{\frac{4}{52} \cdot n} \leftarrow np.$

$$E(X) = \sum_{n} \frac{4}{52} n \cdot P(N=n)$$

$$= \frac{4}{52} \underbrace{\sum_{n} n P(N=n)} = \frac{4}{52} \mu$$

is not np.

μ is some fixed exp. value

Midterm review

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

Note that if X has the Poisson(μ) distribution, and Y has the Poisson(λ) distribution, then $X + Y$ has the Poisson$(\mu + \lambda)$ distribution.

Note that:

- Unless otherwise stated, dice have six sides and are fair, and coins have two distinct ~~coins~~ sides and are fair, cards are dealt without replacement.
- "at random" means equally likely (uniformly at random)

- i.i.d: independent and identically distributed
- SRS: simple random sample or sampling without replacement
- pmf: probability mass function
- Cdf: cumulative distribution function

- When you study, verbalize. Try to describe the problem in words. Make sure you understand what the problem is asking for, and you understand what information you are given. For instance, the difference between waiting times and binomial rv
- Draw pictures!! Draw pictures!!
- Do NOT assume independence!! 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.

During the exam

- Start with the easy questions. When you open the midterm, skim through it quickly and mark the questions that are easy to answer and do those first. Not only will this help you maximize your scoring potential but it will also bolster your confidence. Once you are done with answering the easy questions, it's time to tackle the ones you skipped. Don't over-think straightforward questions.

- Read each question carefully. As you know, assumptions matter. If you have misread those then your solution will be off. For example, confusing with and without replacement is a big problem. Also: "the fourth head is on the 20th toss" and "four heads in 20 tosses" are quite different.

H 20th toss

3H in 19 tosses 4th Head

- Forcing youself to read slowly, underlining key assumptions as you read, is important for doing well. 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.

During the exam

- 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.

$X \sim Pois(5) \to P(X=4) = e^{-5} \frac{5^4}{4!} \cdot e^{-10}$

$P(\text{Full House}) = \frac{\binom{13}{1}\binom{4}{3}\binom{12}{1}\binom{4}{2}}{\binom{52}{5}}$

Triple + pair

- Don't waste time simplifying any arithmetic or algebra unless you are explicitly asked to. We need to see your thought process, and are confident that you are able to use a calculator – we do not need a demonstration of this knowledge.

- If an answer 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.

Suggestions for solving problems

- If multiple stages to a problem, try drawing a tree ~~diag~~ diagram.