Discussion 7A

Tarang Srivastava - CS 70 Summer 2020

Mini Review

In light of Continuous Probability

EDIT: are you teaching continuous probability right now? because if so, no one's going to get it anyway, so tell people as much and to just stick to the formulas

thanks! | 0

Updated 5 days ago by Victor Huang

followup discussions *for lingering questions and comments*

Resolved Unresolved

Tarang Srivastava 5 days ago

Yes it is the continuous probability portion so sounds good! My students are very hungry for intuition for some reason, I feel like I can't present a formula without them immediately asking me to explain how it makes sense. I guess I can just tell them to read the notes for that?

helpful! | 0

Victor Huang 5 days ago

call me jaded, but I tell them straight up that what 70 has taught them up until this point can't explain continuous probability, and to just go with it :-p

"it's like discrete probability, but with integrals... kind of"

undo helpful | 1

Reply to this followup discussion

For context this is a Piazza thread between my instructor and I for CS 375, the class on how to teach computer science. Also Victor has been like a 3 time TA and 2 time Head TA for CS 70 so I'm not doubting he knows how to best teach CS70 for success. .

Maybe I'm just naive, but there's more to the class than the grade...

note @1116

68 views

Resource for Discussion 6D!

Hi, everyone! I'm an AI for this class (despite showing up as a student on Piazza).

As you might recall from Discussion 6D, we used the fact that $f_{Y|X}(y \mid x) = \frac{f(x,y)}{f_X(x)}$ without proof in our answer to part (c) of question 1. A student during discussion asked why this is valid, and I thought this was a very interesting question. A relevant proverb says "what can be asserted without proof can be dismissed without proof," so I thought I'd write up a semi-formal outline explaining why this statement is true.

I'm attaching the outline here: BayesForContinuousRVs.pdf. Hopefully it at least provides some intuition behind the broad concepts in probability for continuous random variables.

Shoutout to @Shahzar for his feedback on this guide :)

Wishing you guys all the best as we approach the end of the semester!

other discussion

~ An instructor (Kevin Zhu) thinks this is a good note ~

edit · undo good note | 7

Updated 5 hours ago by Ayush Pancholy (Anon. Helix to classmates)

Stay curious and question everything, not every rabbit hole will end up benefiting your CS70 grade. In fact, most probably won't, but there's more to all this than just the class. Here's to the final stretch!

Lecture Highlights

If X is a nonnegative random variable with finite mean and $a > 0$, then the probability that X is at least a is at most the expectation of X divided by a.

Markov

$$P(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$$

mean := expectation

$\frac{1}{n}$

Consider tossing a fair coin n times and let X denote the number of heads observed.
What is the probability of observing more than 0.75*n heads? (Note 17 pg. 3)

$$P[X \geq \tfrac{3}{4}n] \leq \frac{\frac{n}{2}}{\frac{3}{4}n} = \frac{2}{3}$$

$X \sim \text{Binomial}(n, \frac{1}{2})$

$E[X] = np = \frac{n}{2}$

$\text{Var}[X] = np(1-p) = n\frac{1}{2}(1-\frac{1}{2}) = \frac{n}{4}$

Weakest "loose" inequality

If X is **any** random variable with finite mean and $a > 0$, then for any $r > 0$:

$$P(|X| \geq a) \leq \frac{\mathbb{E}[|X|^r]}{a^r}$$

generalized Markov

Lecture Highlights

If X is a random variable with finite mean μ and $a > 0$, then the probability that X is at least c away from its mean is at most the variance of X divided by c^2.

$$P(|X - \underset{E[X]}{\mu}| \geq c) \leq \frac{\mathrm{Var}[X]}{c^2}$$

← Chebyshev

LLN

We make n samples as $n \to \infty$ our distribution will be closer and closer to the true mean

For any random variable X with finite expectation $\mathbb{E}[X] = \mu$ and finite standard deviation $\sigma = \sqrt{\mathrm{Var}[X]}$,

$$\Pr[|X - \mu| \geq k\sigma] \leq \frac{1}{k^2},$$

for any constant $k > 0$.

Consider tossing a fair coin n times and let X denote the number of heads observed.
What is the probability of observing more than 0.75*n heads? (Note 17 pg. 3)

$$P[X \geq \tfrac{3}{4}n] \Rightarrow P[\underbrace{X - \tfrac{1}{2}n}_{E[X]} \geq \tfrac{1}{4}n] \leq P[|X - \tfrac{1}{2}n| \geq \tfrac{1}{4}n]$$

★ trick

$$\leq \frac{\mathrm{Var}(X)}{(\frac{1}{4}n)^2} \qquad \frac{\frac{1}{4}n}{(\frac{1}{4}n)^2} = \frac{n}{4}$$

Question 3

Question 3 (Zoom Poll)

(a) A fair coin is tossed multiple times and you win a prize if there are more than 60% heads. Which number of tosses would you prefer: 10 tosses or 100 tosses? Explain.

10. ✓

n samples

$n \to \infty$

(b) A fair coin is tossed multiple times and you win a prize if there are more than 40% heads. Which number of tosses would you prefer: 10 tosses or 100 tosses? Explain.

100 ✓

5 %

(c) A fair coin is tossed multiple times and you win a prize if there are between 40% and 60% heads. Which number of tosses would you prefer: 10 tosses or 100 tosses? Explain.

100 ✓

2

1000 coins

exactly

(d) A fair coin is tossed multiple times and you win a prize if there are exactly 50% heads. Which number of tosses would you prefer: 10 tosses or 100 tosses? Explain.

$\boxed{10}$ ✓

$$\binom{10}{5}\frac{1}{2^5} > \binom{100}{50}\frac{1}{2^{100}}$$

0.246 0.079

500 ⟷ 500

Question 3

(a) 10 tosses. By LLN, the sample mean should have higher probability to be close to the population mean as n increases. Therefore the average proportion of coins that are heads should be closer to 0.50, and has a lower chance of being greater than 0.60 if there are 100 tosses (compared with 10 tosses).

(b) 100 tosses. Again, by LLN, the sample mean should have higher probability to be close to the population mean as n increases. Therefore the average proportion of coins that are heads should be closer to 0.50, and has a lower chance of being smaller than 0.40 if there are 100 tosses. A lower chance of being smaller than 0.40 is the desired result.

(c) 100 tosses. Again, by LLN, the average proportion of coins that are heads should be closer to 0.50, and has a lower chance of being both smaller than 0.40 if there are 100 tosses. Similarly, there is a lower chance of being larger than 0.60 if there are 100 tosses. Lower chances of both of these events is desired if we want the fraction of heads to be between 0.4 and 0.6.

(d) 10 tosses. Compare the probability of getting equal number of heads and tails between $2n$ and $2n+2$ tosses.

$$\mathbb{P}[n \text{ heads in } 2n \text{ tosses}] = \binom{2n}{n}\frac{1}{2^{2n}}$$

$$\begin{aligned}
\mathbb{P}[n+1 \text{ heads in } 2n+2 \text{ tosses}] &= \binom{2n+2}{n+1}\frac{1}{2^{2n+2}} = \frac{(2n+2)!}{(n+1)!(n+1)!}\cdot\frac{1}{2^{2n+2}} \\
&= \frac{(2n+2)(2n+1)2n!}{(n+1)(n+1)n!n!}\cdot\frac{1}{2^{2n+2}} \\
&= \frac{2n+2}{n+1}\cdot\frac{2n+1}{n+1}\binom{2n}{n}\cdot\frac{1}{2^{2n+2}} < \left(\frac{2n+2}{n+1}\right)^2\binom{2n}{n}\cdot\frac{1}{2^{2n+2}} \\
&= 4\binom{2n}{n}\cdot\frac{1}{2^{2n+2}} = \binom{2n}{n}\frac{1}{2^{2n}} = \mathbb{P}[n \text{ heads in } 2n \text{ tosses}]
\end{aligned}$$

As we increment n, the probability will always decrease. Therefore, the larger n is, the less probability we'll get exactly 50% heads. □

Question 1

Question 1

A random variable X has variance $\mathrm{Var}(X) = 9$ and expectation $\mathbb{E}[X] = 2$. Furthermore, the value of X is never greater than 10. Given this information, provide either a proof or a counterexample for the following statements.

(a) $\mathbb{E}[X^2] = 13$.

(b) $\mathbb{P}[X = 2] > 0$. Discrete

(c) $\mathbb{P}[X \geq 2] = \mathbb{P}[X \leq 2]$.

(d) $\mathbb{P}[X \leq 1] \leq 8/9$.

(e) $\mathbb{P}[X \geq 6] \leq 9/16$.

Question 1

(a) $\mathbb{E}[X^2] = 13$.

True

$$Var(X) = E[X^2] - E[X]^2$$
$$9 = \quad ? \quad - 2^2$$
$$E[X^2] = 9 + 4 = 13$$

(b) $\mathbb{P}[X = 2] > 0$.

False.

$P[X = 5] = \frac{1}{2}$

$P[X = -1] = \frac{1}{2}$

arbitrarily

$P[X = a] = \frac{1}{2}$

$P[X = b] = \frac{1}{2}$

$E[X] = \frac{a}{2} + \frac{b}{2} = 2$

$\boxed{4 = a + b}$

$E[X^2] = 13 = \frac{a^2}{2} + \frac{b^2}{2}$

$\boxed{26 = a^2 + b^2}$

$a = 5 \quad b = -1$

(c) $\mathbb{P}[X \geq 2] = \mathbb{P}[X \leq 2]$.

False:

dist doesn't have to be symmetric about the mean

$$\boxed{\begin{aligned} P[X = -7] &= \tfrac{9}{10} \\ P[X = 3] &= \tfrac{1}{10} \end{aligned}}$$

Question 1

$1 - P[X > 1] \le \frac{8}{9}$

(d) $\mathbb{P}[X \le 1] \le 8/9$.

True:

denominator

$Y = 10 - X$

is non-negative

$P[Y \ge 9] \le \frac{E[Y]}{9} = \frac{10 - E[X]}{9}$

$\Downarrow$

$P[10 - X \ge 9]$

$\Downarrow$

$P[X \le 1] \le \frac{8}{9} = \frac{8}{9}$

(e) $\mathbb{P}[X \ge 6] \le 9/16$.

True:

Chebyshev's

$E[X]$

$|X - \mu|$

$P[X - 2 \ge 4] \le P[|X - 2| \ge 4] \le \frac{Var(X)}{4^2}$

$P[X \ge 6] \le \frac{9}{16}$

easy

$|x-2| \ge 4$ $\quad x - 2 \ge 4$

$x \geq 6$ or

$-(x-2) \leq 4$

$-x+2 \leq 4$

$-2 \quad -2$

Question 2

Question 2

A friend tells you about a course called "Laziness in Modern Society" that requires almost no work. You hope to take this course next semester to give yourself a well-deserved break after working hard in CS 70. At the first lecture, the professor announces that grades will depend only on two homework assignments. Homework 1 will consist of three questions, each worth 10 points, and Homework 2 will consist of four questions, also each worth 10 points. He will give an A to any student who gets at least 60 of the possible 70 points.

However, speaking with the professor in office hours you hear some very disturbing news. He tells you that, in the spirit of the class, the GSIs are very lazy, and to save time the grading will be done as follows. For each student's Homework 1, the GSIs will choose an integer randomly from a distribution with mean $\mu = 5$ and variance $\sigma^2 = 1$. They'll mark each of the three questions with that score. To grade Homework 2, they'll again choose a random number from the same distribution, independently of the first number, and will mark all four questions with that score.

If you take the class, what will the mean and variance of your total class score be? Use Chebyshev's inequality to conclude that you have less than a 5% chance of getting an A when the grades are randomly chosen this way.

Question 2

A friend tells you about a course called "Laziness in Modern Society" that requires almost no work. You hope to take this course next semester to give yourself a well-deserved break after working hard in CS 70. At the first lecture, the professor announces that grades will depend only on two homework assignments. Homework 1 will consist of three questions, each worth 10 points, and Homework 2 will consist of four questions, also each worth 10 points. He will give an A to any student who gets at least 60 of the possible 70 points.

However, speaking with the professor in office hours you hear some very disturbing news. He tells you that, in the spirit of the class, the GSIs are very lazy, and to save time the grading will be done as follows. For each student's Homework 1, the GSIs will choose an integer randomly from a distribution with mean $\mu = 5$ and variance $\sigma^2 = 1$. They'll mark each of the three questions with that score. To grade Homework 2, they'll again choose a random number from the same distribution, independently of the first number, and will mark all four questions with that score.

If you take the class, what will the mean and variance of your total class score be? Use Chebyshev's inequality to conclude that you have less than a 5% chance of getting an A when the grades are randomly chosen this way.

Question 2