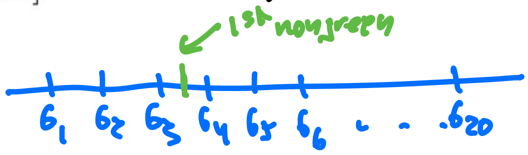
Stat 134 lec 17

Warmup 10:00 – 10:10

An urn contains 90 marbles, of which there are 20 greens, 20 blacks and 50 red marbles. Tom draws marble without replacement until the 6^{th} green marble. Let X = # of marbles drawn. Example: **GGG**BRB**GG**BR**G** with $x = 11$. Find $\mathbb{E}[X]$ and Var(X)



$X = I_1 + \dots + I_{70} + 6$

$E(X) = 70\left(\frac{6}{21}\right) + 6$

$p = 6/21$

$I_2 = \begin{cases} 1 & \text{if } 2^{nd} \text{ non G before } 6^{th} \text{ green} \\ 0 & \text{else} \end{cases}$

$p_{12} = 6/21 \cdot 7/22$

$I_{12} = \begin{cases} 1 & \text{if } 1^{st} \text{ and } 2^{nd} \text{ nongreen before } 6^{th} \text{ green} \\ 0 & \text{else} \end{cases}$

Today MT1 review

Rida Abdulwasay

7:35pm

Also, could you do a brief review of how to approach and solve max/min problems like P(max(X1, X2) ≥ x, P(min(X1, X2) ≤ x, etc. for different types of distributions?

And if there's time could you perhaps explain how to do the Quiz #2 expectation problem:

3. One day, the Stanford TAs are feeling lazy, so they decide to implement an alternative method for grading quizzes. For each student, the TA will roll 10 dice. For every "good" die that shows a number ≥ 3, the student gets 10 points on their quiz. There are 30 students in the class.

   (a) What distribution is the random variable $X =$ the number of "good dice" for a student?

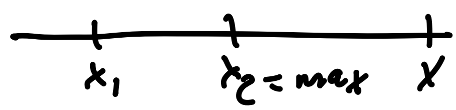
   (b) What is the expected score of a student?

   (c) Let N be the number of students who get at least 90 points on the quiz. What is $\mathbb{E}[N]$?

   (d) One TA suggests that it might be better for the students if instead of giving 10 points for each "good" die, the TAs take the number of "good" dice and square it to get the student's score. Will the average score of a student be higher or lower with this new scheme?

Thanks!

Edited by Rida Abdulwasay on Feb 27 at 7:48pm



$$P(\max(x_1, x_2) \geq x) = 1 - P(\max(x_1, x_2) < x)$$

$$= 1 - P(x_1 < x,\ x_2 < x)$$

This is useful when x_1, x_2 are indep and you know $P(x_1 < x)$

Similar for $\min(x_1, x_2)$.

ex

One day, the Stanford TAs are feeling lazy, so they decide to implement an alternative method for grading quizzes. For each student, the TA will roll 10 dice. For every "good" die that shows a number ≥ 3, the student gets 10 points on their quiz. There are 30 students in the class.

(a) What distribution is the random variable $X=$ the number of "good dice" for a student?

(b) What is the expected score of a student?

(c) Let N be the number of students who get at least 90 points on the quiz. What is $\mathbb{E}[N]$?

(d) One TA suggests that it might be better for the students if instead of giving 10 points for each "good" die, the TAs take the number of "good" dice and square it to get the student's score. Will the average score of a student be higher or lower with this new scheme?

(a) If X is the number of "good dice" for a student, X will have Binomial$(10, 2/3)$ distribution.

(b) The expected score of a student is

$$\mathbb{E}[10X] = 10\mathbb{E}[X] = 10 \cdot 10 \cdot \frac{2}{3} \approx 66.7.$$

(c) Let $I_1, \ldots, I_{300}$ be the indicators of each student's score being ≥ 90; that is,

$$I_k = \begin{cases} 1 & \text{if student } k \text{ gets} \geq 90 \text{ points} \\ 0 & \text{if student } k \text{ gets} < 90 \text{ points.} \end{cases}$$

Then $N = I_1 + \cdots + I_{30}$, and

$$\begin{aligned} \mathbb{E}[N] &= \mathbb{E}[I_1] + \cdots + \mathbb{E}[I_{30}] \\ &= 30\mathbb{E}[I_1] \\ &= 30\mathbb{P}(\text{student } k \text{ gets} \geq 90 \text{ points}) \\ &= 30\mathbb{P}(9 \text{ or } 10 \text{ dice are "good"}) \\ &= 30\left(\binom{10}{9}(2/3)^9(1/3) + \binom{10}{10}(2/3)^{10}\right). \end{aligned}$$

(d) As before, if X is the number of "good dice" for a student, then X will have Binomial$(10, 2/3)$ distribution. The average score of a student with this new scheme will be

$$\mathbb{E}[X^2] = \sum_{k=0}^{10} k^2 \binom{10}{k} (2/3)^k (1/3)^{10-k}.$$

On the other hand, the average score of a student with the old scheme is

$$\mathbb{E}[10X] = \sum_{k=0}^{10} 10k \binom{10}{k} (2/3)^k (1/3)^{10-k}.$$

This tells us that $\mathbb{E}[10X] \geq \mathbb{E}[X^2]$ because $10 \geq k$ in each term. So the new scheme makes the average student score lower.

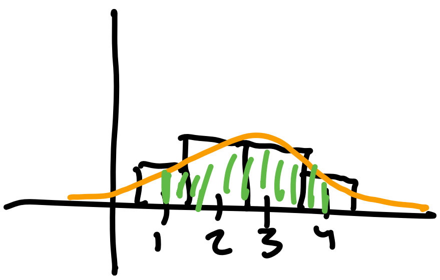
Xinyi Liu
8:09pm

Could you explain more about when we need to use the continuity correction factor and when we don't need? Also, is there any difference between CLT and normal approximation? Thank you very much in advance!

Reply (2 likes)

We need CC when we approximate a discrete distribution taking integer values by a continuous distribution

Ex approximate binomial with Normal



If there are many rectangles the benefit of correction is very small.

On a test I will ask you to explicitly use it if necessary.

Ekaterina Bushmareva
7:48pm

Could you please go over question 3 from the Fall 2018 Midterm (the one that begins with, "There are 30 marbles in a bag")? Additionally, do you mind providing some intuition on understanding when indicators are dependent vs. independent, and finding the probability of an indicator = 1 for these kinds of problems?

Reply (1 like)

1. (10 pts total) There are 30 marbles in a bag. 10 of them are black, 10 of them are white, and 10 of them are red. Let X denote the number of different colors appearing among 5 marbles selected at random from the bag. Find:

(a) (3 pts) $P(X = 2)$;

$$\mathbb{P}(X=2) = \binom{3}{2}\mathbb{P}(\text{red \& blue both appear})$$

$$= \binom{3}{2}\left(1 - \mathbb{P}(\text{no red} \cup \text{no blue})\right)$$

$= \mathbb{P}(\text{no red}) + \mathbb{P}(\text{no blue}) - \mathbb{P}(\text{no red and no blue})$

$$= 3\cdot\left(1 - 2\cdot\left(\frac{\binom{20}{5}\binom{10}{0}}{\binom{30}{5}}\right) + \frac{\binom{10}{5}}{\binom{30}{5}}\right)$$

For instance: ↑ all B or W ↑ all W

(b) (3 pts) $\mathbb{E}(X)$;

Let $\mathbb{1}_i$ be the indicator that color i appears in the sample. So $X = \mathbb{1}_b + \mathbb{1}_r + \mathbb{1}_w$. By symmetry,

$$\mathbb{E}(X) = \mathbb{E}(\mathbb{1}_b + \mathbb{1}_r + \mathbb{1}_w) = 3\mathbb{E}(\mathbb{1}_b)$$

$$= 3\left(1 - \frac{\binom{20}{5}}{\binom{30}{5}}\right)$$

see above.

(c) (4 pts) $Var(X)$.

$$Var(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2,$$

where $\mathbb{E}(X^2) = \mathbb{E}\left((\mathbb{1}_r + \mathbb{1}_b + \mathbb{1}_w)^2\right)$ r, b both appear.

$$= 3\mathbb{E}(\mathbb{1}_r^2) + 3\cdot 2\,\mathbb{E}(\mathbb{1}_r\mathbb{1}_b)$$

$$= \mathbb{E}(X) + 6\cdot\left(1 - \mathbb{P}(\text{no red} \cup \text{no blue})\right)$$

$$= \mathbb{E}(X) + 6\left(1 - \left(\frac{\binom{20}{5}}{\binom{30}{5}}\right)\cdot 2 + \left(\frac{\binom{10}{5}}{\binom{20}{5}}\right)\right)$$

If $X = I_1 + \dots + I_n$

$$Var(X) = \underbrace{np_1 + n(n-1)p_{12}}_{E(X^2)} - (E(X))^2$$

Indicators are independent iff

$X \sim Bin(n,p)$

Otherwise indicators are dependent

(ex $X \sim HG(n,N,G)$)

The probability an indicator is 1 is an unconditional probability

$1/20$

$X = \underbrace{I_1 + \cdots + I_n}_{\text{indep}}$

$X \sim Bin(n,p)$

$I_i = \begin{cases} 1 & \text{w/ prob } p \\ 0 & \text{else} \end{cases}$

X_1 and X_2 are indep iff

$P(X_1 | X_2) = P(X_1)$

Ekaterina Bushmareva

7:49pm

Also, can you please go over question 5 from the Fall 2018 Midterm ("Three couples attend a dinner") and how to apply the inclusion-exclusion formula to more than 2 or 3 events?

Reply

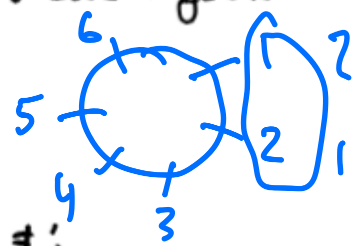
---

12. Inclusion–exclusion formula for n events. Derive the inclusion–exclusion formula for n events

$$P(\bigcup_{i=1}^{n} A_i) = \sum_i P(A_i) - \sum_{i<j} P(A_iA_j) + \sum_{i<j<k} P(A_iA_jA_k) - \cdots + (-1)^{n+1} P(A_1 \ldots A_n)$$

4. (5 pts) Three couples attend a dinner. Each of the six people chooses a seat randomly from a round table with six seats. What is the probability that no couple sits together?

Let A_i be the event that couple i sits together.



(1,2) couple.

So $\mathbb{P}(A_i) = \dfrac{2!\cdot 4!\cdot 6}{6!}$,

$\mathbb{P}(A_iA_j) = \dfrac{6\cdot 2!\cdot 3\cdot 2!\cdot 2!}{6!}$, $i \neq j$,

$\mathbb{P}(A_1A_2A_3) = \dfrac{6\cdot 2!^3\cdot 3!}{6!}$

By inclusion-exclusion, $\mathbb{P}(\text{no couple sits together}) = 1 - \mathbb{P}\left(\bigcup_{i=1}^{3} A_i\right)$

$= 1 - 3\mathbb{P}(A_i) + \binom{3}{2}\mathbb{P}(A_iA_j) - \mathbb{P}(A_1A_2A_3)$

Aaron Zaks

7:27pm

would you please go over an example problem that uses NegBin? Also, a problem that involves a sum of geometric distributions would be helpful.

Reply

ex A box contains 30 red marbles, 20 green marbles and 10 blue marbles.

You draw marbles with replacement.

let $X =$ # marbles until your sixth blue.

$X \sim NB(6, \frac{1}{6})$.

Find $P(X = 15)$

Review coupon collector problem for sum of geometric

Coupon Collector's problem

You have a collection of boxes each containing a coupon. There are n different coupons. Each box is equally likely to contain any coupon independent of the other boxes.

X = # boxes needed to get all n different coupons.

ex $n=3$ $\quad X = X_1 + X_2 + X_3$

$$\underbrace{C}_{X_1}\ \underbrace{C\ C\ C}_{X_2}\ \underbrace{C\ C\ C\ C}_{X_3}$$