Stat 134: Section 1

Adam Lucas

August 22, 2018

WELCOME TO STAT 134! Alongside lecture, discussion sections are a key way to clarify and reinforce the course materials. We hope to make discussions an engaging and welcoming environment!

Problem 1

Suppose a word is picked at random from this sentence.

a. Find the chance that the word contains at least 2 vowels (a,e,i,o,u).

b. Find the chance that the word contains at least 4 letters and at least 2 vowels.

c. What is the distribution of the length of the word picked?

What does a probability distribution consist of? What conditions must it satisfy?

From Ex 1.1.2 and 1.3.6 in Pitman's Probability

a) Words with at least 2 vowels: "suppose," "picked," "random," "sentence"

$\Rightarrow \frac{4}{10} = \frac{2}{5}$.

b) Similarly, $2/5$.

c) Let X = length of word picked. $\Rightarrow$

x	1	2	4	6	7	8
$P(X=x)$	$1/10$	$2/10$	$3/10$	$2/10$	$1/10$	$1/10$

Problem 2

Cards are dealt from a well-shuffled standard deck until the first heart appears.

a. What is the probability that exactly 5 deals are required?

b. What is the probability that 5 or fewer deals are required? Try to answer without using a summation term.

Ex 1.rev.8 in Pitman's Probability

a) First 4 cards not heart, then 5th card heart $\Rightarrow$ ~~$\frac{39}{13} \cdot \frac{38}{52} \cdot 37$~~ $\frac{39}{52} \cdot \frac{38}{51} \cdot \frac{37}{50} \cdot \frac{36}{49} \cdot \frac{13}{48}$

b) $P(\leq 5 \text{ deals}) = 1 - P(> 5 \text{ deals}) = 1 - P(\text{first 5 cards no heart}) =$

$= 1 - \frac{39}{52} \cdot \frac{38}{51} \cdot \frac{37}{50} \cdot \frac{36}{49} \cdot \frac{35}{48}$.

Problem 3: The Birthday Problem

CLASS ACTIVITY: In your discussion section, how many students do you think have the same birthday? As time permits, your GSI will go around the room and have students say their birthdays.

Suppose you are in a classroom of n students ($n \leq 365$). In the following calculations, ignore leap days and assume that students' birthdays are independent and distributed uniformly at random across the year. Find:

How are these assumptions violated in reality? How does this affect the true probability of these events?

a. the chance that at least one other student shares *your* birthday.

b. the chance that at least two students share the same birthday.

c. (continued from part b): Using your answer from part (b), derive a useful approximation for this expression, using the approximation $\log(1+x) \approx x$ for small x.

From Section 1.6, Example 5 (pg 62) in Pitman's Probability

a) $1 - P(\text{no one shares your birthday}) = 1 - \left(\frac{364}{365}\right)^{n-1}$.

b) $1 - P(\text{everyone has different birthdays}) =$

$$= 1 - \left(\frac{365}{365}\right)\left(\frac{364}{365}\right)\left(\frac{363}{365}\right)\cdots\left(\frac{365-(n-1)}{365}\right) = 1 - \prod_{k=0}^{n-1}\left(\frac{365-k}{365}\right).$$

first person can have any birthday

c) Note that $\frac{365-k}{365} = 1 - \frac{k}{365}$. We can reasonably say $\frac{k}{365}$ is small,

so ~~for $0 \leq k \leq n-1$,~~ $\log\left[\prod_{k=0}^{n-1}\left(1-\frac{k}{365}\right)\right] = \sum_{k=0}^{n-1}\left(\log\left(1-\frac{k}{365}\right)\right) \approx \sum_{k=0}^{n-1}\left(-\frac{k}{365}\right) =$

$$= -\frac{1}{365}\sum_{k=0}^{n-1} k = -\frac{1}{365}\cdot\frac{(n-1)n}{2} = -\frac{n(n-1)}{730}.$$

Therefore $1 - P(\text{everyone has different birthdays}) \approx 1 - e^{\frac{-n(n-1)}{730}}$.

Stat 134: Section 2

Adam Lucas

August 27, 2018

Problem 1

Events A, B, and C are defined in an outcome space. Find expressions for the following probabilities in terms of $P(A)$, $P(B)$, $P(C)$, $P(AB)$, $P(AC)$, $P(BC)$, and $P(ABC)$.

a. The probability that exactly two of A, B, C occur.

b. The probability that exactly one of these events occurs.

c. The probability that none of these events occur.

For these questions, it might prove helpful to first draw a Venn diagram.

In part c, use what you already know to avoid doing unnecessary work.

Ex 1.3.10 in Pitman's Probability

a) $P(A \cap B) + P(A \cap C) + P(B \cap C) - 3P(A \cap B \cap C)$.
$(= P(ABC^c) + P(AB^cC) + P(A^cBC) = \nearrow)$

b) $P(A \cap B^c \cap C^c) + P(A^c \cap B \cap C^c) + P(A^c \cap B^c \cap C) =$
$= P(A) + P(B) + P(C) - 2P(A \cap B) - 2P(A \cap C) - 2P(B \cap C) + 3P(A \cap B \cap C)$.

c) $1 - P(A \cup B \cup C) = 1 - [P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)]$

Problem 2

A hat contains $f + b$ coins, f of which are fair, b of which are biased to land heads with probability $2/3$. A coin is drawn from the hat and tossed twice. The first time it lands heads, and the second time it lands tails. Given this information, what is the probability that it is a fair coin?

(Bayes' Rule)

Again, it might prove helpful to first draw a diagram here; use a tree diagram this time.

Ex 1.rev.11 in Pitman's Probability

$$P(\text{fair} \mid HT) = \frac{P(HT \mid \text{fair})\, P(\text{fair})}{P(HT \mid \text{fair})\, P(\text{fair}) + P(HT \mid \text{biased})\, P(\text{biased})} =$$

$$= \frac{\frac{1}{4} \cdot \frac{f}{f+b}}{\frac{1}{4} \cdot \frac{f}{f+b} + \frac{2}{3} \cdot \frac{1}{3} \cdot \frac{b}{f+b}} = \frac{f/4}{\frac{f}{4} + \frac{2b}{9}} = \frac{\cdot 36}{\cdot 36}$$

$$= \cancel{\frac{f}{f + \frac{b}{18}}} = \frac{9f}{9f + 8b}.$$

Scanned by CamScanner

Problem 3

A box contains n tickets, labeled $1, 2, \ldots, n$. Two tickets are drawn at random from the box. Find the chance that the numbers on the two tickets differ by two or more if the draws are made:

a. with replacement;

b. without replacement.

a) $\frac{2}{n} \cdot \frac{n-2}{n} + \frac{n-2}{n} \cdot \frac{n-3}{n}$

b) $\frac{2}{n} \cdot \frac{n-2}{n-1} + \frac{n-2}{n} \cdot \frac{n-3}{n-1}$

Ex 1.rev.19 in Pitman's Probability

Problem 4

An experimenter observes an event A as the outcome of a particular experiment. There are three different hypotheses, H_1, H_2, and H_3, which the experimenter regards as the only possible explanations of event A. Under hypothesis H_1, the experiment should produce result A about 10% of the time, under H_2 about 1% of the time, and under H_3 about 39% of the time. Having observed A, the experimenter decides that H_3 is the most likely explanation, and that the probability that H_3 is true is

$$39\%/(10\% + 1\% + 39\%) = 78\%$$

a. What assumption is the experimenter implicitly making? All hypotheses equally likely. $\mathbb{P}(H_1) = \mathbb{P}(H_2) = \mathbb{P}(H_3) = 1/3$.

b. Does the probability 78% admit a long-run frequency interpretation? No, since we have an assumption, above.

c. Suppose the experiment is a laboratory test on a blood sample from an individual chosen at random from a particular population. The hypothesis H_i is that the individual's blood is of some particular type i. Over the whole population it is known that 50% of individuals have blood type 1, 45% have blood type 2, and the remaining proportion have type 3. Revise the experimenter's calculation of the probability of H_3 given A, so that it admits a long-run frequency interpretation. Is H_3 still the most likely hypothesis given A?

long-run frequency $\approx$ probability

Ex 1.5.6. in Pitman's Probability

Bayes' rule:

$$\Rightarrow \mathbb{P}(H_3|A) = \frac{\mathbb{P}(H_3)\,\mathbb{P}(A|H_3)}{\mathbb{P}(H_1)\,\mathbb{P}(A|H_1) + \mathbb{P}(H_2)\,\mathbb{P}(A|H_2) + \mathbb{P}(H_3)\,\mathbb{P}(A|H_3)} = \frac{(0.05)(0.39)}{(0.05)(0.39) + (0.45)(0.01) + (0.5)(0.1)} \approx$$

$\approx 0.2635.$

Prepared by Brian Thorsen and Yiming Shi

Posterior probabilities: Given A, posterior probability of H_i is given by $\mathbb{P}(H_i)\,\mathbb{P}(A|H_i)$.

$$\Rightarrow \begin{array}{l} \mathbb{P}(H_1)\,\mathbb{P}(A|H_1) = 0.05 \\ \mathbb{P}(H_2)\,\mathbb{P}(A|H_2) = 0.0045 \\ \mathbb{P}(H_3)\,\mathbb{P}(A|H_3) = 0.0195 \end{array}$$

$\Rightarrow H_1$ is now most likely.

Stat 134: Section 3

Adam Lucas

August 29, 2018

Problem 1

Given that there were 12 heads in 20 fair independent coin tosses, calculate the chance that

a. the first toss landed heads;

b. the first two tosses landed heads;

c. at least two of the first five tosses landed heads.

Try to do this problem with as little tedious work as possible.

Ex 2.1.5 in Pitman's Probability

a) $\mathbb{P}(H\cdots \mid 12\text{ H's in 20 tosses}) = \dfrac{\mathbb{P}(H\cdots,\ 12\text{ H's in 20 tosses})}{\mathbb{P}(12\text{ H's in 20 tosses})} = \dfrac{\frac{1}{2}\cdot \mathbb{P}(11\text{ H's in 19 tosses})}{\binom{20}{12}\left(\frac{1}{2}\right)^{12}\left(\frac{1}{2}\right)^{8}} =$

$= \dfrac{\frac{1}{2}\cdot\binom{19}{11}\left(\frac{1}{2}\right)^{19}}{\binom{20}{12}\left(\frac{1}{2}\right)^{20}} = \dfrac{19!}{11!\,8!}\cdot\dfrac{12!\,8!}{20!} = \dfrac{12}{20} = \dfrac{3}{5}.$

b) $\mathbb{P}(HH\cdots \mid 12\text{ H's in 20 tosses}) = \binom{18}{10}\Big/\binom{20}{12} = \dfrac{18!}{10!}\cdot\dfrac{12!}{20!} = \dfrac{12\cdot 11}{20\cdot 19} = \dfrac{33}{95}.$

c) $1-\mathbb{P}(\text{less than 2 H's in first 5 tosses} \mid 12\text{ H's in 20 tosses}) =$

$= 1-\mathbb{P}(0\text{ H in first }5 \mid \tfrac{12\text{ H}}{20}) - \mathbb{P}(1\text{ H in first }5 \mid \tfrac{12\text{ H}}{20}) = 1-\dfrac{\binom{15}{12}}{\binom{20}{12}} - 5\cdot\dfrac{\binom{15}{11}}{\binom{20}{12}}.$

Problem 2

A gambler decides to keep betting on red at roulette where there are 18 reds out of 38 tiles in total, and stop as soon as she has won a total of five bets.

a. What is the probability that she has to make exactly 8 bets before stopping?

b. What is the probability that she has to make at least 9 bets?

Ex 2.1.12 in Pitman's Probability

a) $\mathbb{P}(\text{won the 8th bet, 4 wins in first 7 bets}) = \binom{7}{4}\left(\frac{18}{38}\right)^4\left(\frac{20}{38}\right)^3\cdot\left(\frac{18}{38}\right).$

$\mathbb{P}(k\text{ bets})=0$ for $k<5$

~~b) $1-\mathbb{P}(\text{make at most 8 bets}) = 1-\mathbb{P}(8\text{ bets})-\mathbb{P}(7\text{ bets})-\mathbb{P}(6\text{ bets})-\mathbb{P}(5\text{ bets}) =$~~

~~$= 1-\binom{7}{4}\left(\frac{18}{38}\right)^5\left(\frac{20}{38}\right)^3-\binom{6}{4}\left(\frac{18}{38}\right)^5\left(\frac{20}{38}\right)^2-\binom{5}{4}\left(\frac{18}{38}\right)^5\left(\frac{20}{38}\right)-\left(\frac{18}{38}\right)^5.$~~

b) $\mathbb{P}(\text{wins at most 4 bets in first 8 bets}) =$

$= \left(\frac{20}{38}\right)^8 + \binom{8}{1}\left(\frac{18}{38}\right)\left(\frac{20}{38}\right)^7 + \cdots + \binom{8}{4}\left(\frac{18}{38}\right)^4\left(\frac{20}{38}\right)^4.$

Problem 3: The matching problem

There are n letters addressed to n people at n different addresses. The n addresses are typed on n envelopes. A disgruntled secretary shuffles the letters and puts them in the envelopes in random order, one letter per envelope.

a. What is the chance that the i_{th} letter is put in the correctly addressed envelope? How about both i_{th} letter and j_{th} letter ($i \neq j$)? And the chance that k different letters are put in correctly?

b. Find the probability that at least one letter is put in a correctly addressed envelope;

Hint: Use the inclusion-exclusion formula of Exercise 1.3.12

c. What is the probability in part b. approximately, for large n?

Ex 2.rev.28 in Pitman's Probability

a) $\mathbb{P}(\text{i-th letter in correct envelope}) = \frac{1}{n}$

$\mathbb{P}(\text{i-th \& j-th letter in correct envelopes}) = \frac{1}{n} \cdot \frac{1}{n-1}$.

$\mathbb{P}(\text{k different letters in correct envelopes}) = \frac{1}{n} \cdot \frac{1}{n-1} \cdot \ldots \cdot \frac{1}{n-k+1} = \frac{(n-k)!}{n!}$.

b) Let $A_i = \{\text{i-th letter put in correct envelope}\}$.

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

$= n \cdot \frac{1}{n} - \frac{n(n-1)}{2} \cdot \frac{1}{n(n-1)} + \frac{n(n-1)(n-2)}{3 \cdot 2} \cdot \frac{1}{(n-2)(n-1)n} - \cdots + (-1)^{k+1} \cdot \binom{n}{k} \cdot \frac{(n-k)!}{n!} + \cdots$

$+ \cdots + (-1)^{n+1} \frac{1}{n!} =$

$= \sum_{k=1}^{n} (-1)^{k+1} \binom{n}{k} \frac{(n-k)!}{n!} = \sum_{k=1}^{n} (-1)^{k+1} \frac{n!}{k!(n-k)!} \cdot \frac{(n-k)!}{n!} = \sum_{k=1}^{n} (-1)^{k+1} \cdot \frac{1}{k!} = -\sum_{k=1}^{n} \frac{(-1)^k}{k!}$.

c) Since $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$, we have $\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = -\sum_{k=1}^{n} \frac{(-1)^k}{k!} = -e^{-1} + 1 = 1 - e^{-1}$.

Large $n \Rightarrow$ take limit $n \to \infty$, i.e.

find $\lim_{n\to\infty} \sum_{k=1}^{n} (-1)^{k+1} \cdot \frac{1}{k!} = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k!}$.

Prepared by Brian Thorsen and Yiming Shi

Stat 134: Section 4

Adam Lucas

September 5, 2018

Note: You may leave your answers in terms of Φ or Φ^{-1} as necessary, where $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$, and Φ^{-1} is the inverse of Φ.

Problem 1

Let H be the number of heads in 400 tosses of a fair coin. Find normal approximations to

a. $P(190 \leq H \leq 210)$

b. $P(210 \leq H \leq 220)$

c. $P(H = 200)$

d. $P(H = 210)$

Ex 2.2.1 in Pitman's Probability

$H \sim \text{Binomial}(400, \frac{1}{2}) \Rightarrow \mu = 400 \cdot \frac{1}{2} = 200, \quad \sigma^2 = 400 \cdot \frac{1}{2} \cdot \frac{1}{2} = 100$

Since $n = 400$ is large, we can use the Normal approximation:

$\frac{H - \mu}{\sigma} \approx \cancel{\text{Normal}(\mu, \sigma^2)}$ Normal$(0,1)$. Denote $Z \sim \text{Normal}(0,1)$.

a) $P(190 \leq H \leq 210) = P\left(\frac{190-200}{10} \leq \frac{H-200}{10} \leq \frac{210-200}{10}\right) \approx$

$\approx P\left(\frac{-10-\frac{1}{2}}{10} \leq Z \leq \frac{10+\frac{1}{2}}{10}\right) = \Phi(1.05) - \Phi(-1.05) \approx 0.7062.$

b) $P(210 \leq H \leq 220) = P\left(\frac{210-200}{10} \leq \frac{H-200}{10} \leq \frac{220-200}{10}\right) \approx$

$\approx P\left(\frac{10-1/2}{10} \leq Z \leq \frac{20+1/2}{10}\right) = \cancel{\Phi(2.05) - \Phi(-1.05) \approx 0.15096}$

$\Phi(2.05) - \Phi(0.95) \approx 0.1509.$

c) $P(H=200) \approx P\left(\frac{200-200-1/2}{10} \leq Z \leq \frac{200-200+1/2}{10}\right) = \Phi(0.05) - \Phi(-0.05) \approx 0.0398.$

d) $P(H=210) \approx P\left(\frac{210-200-1/2}{10} \leq Z \leq \frac{210-200+1/2}{10}\right) = \Phi(1.05) - \Phi(0.95) \approx 0.0242.$

Problem 2

A fair coin is tossed repeatedly. Consider the following two possible outcomes: (i) 55 or more heads in the first 100 tosses, or (ii) 220 or more heads in the first 400 tosses.

a. Without calculation, say which of these outcomes is more likely. Why?

b. Confirm your answer to (a) by calculation.

Ex 2.2.3 in Pitman's Probability

a) By the Law of Large Numbers, (i) is more likely.

As $n \to \infty$, $P(|\bar{X}_n - \mu| > \varepsilon) \to 0$.

Let $\varepsilon > 0$.

$$\bar{X}_n = \frac{X_1 + \cdots + X_n}{n}.$$

Here, can take $X_i = \begin{cases} 1 & \text{if } i\text{th flip is heads} \\ 0 & \text{otherwise} \end{cases}$

$n\bar{X}_n \sim \text{Binomial}(n, \frac{1}{2}) \Rightarrow \mu = \frac{1}{2}$. (for all n, ie. 100 and 400)

Since $100 < 400$, (ii) must be less likely.

b) (i) $\text{Binomial}(100, \frac{1}{2})$

(ii) $\text{Binomial}(400, \frac{1}{2})$

Stat 134: Section 5

Adam Lucas

September 10, 2018

Problem 1

A cereal company advertises a prize in every box of its cereal. In fact, only about 95% of their boxes have prizes in them. If a family buys one box of this cereal every week for a year, estimate the chance that they will collect more than 45 prizes. What assumptions are you making? *Ex 2.4.9 in Pitman's Probability*

Assume each box of cereal has a prize w/ prob. 0.95, independently of the others.

52 weeks in a year $\Rightarrow \mu = (0.95)(52) = 49.4$, $\sigma = \sqrt{49.4(0.05)} \approx 1.57$. Since σ is small (< 3), use Poisson instead (w/ respect to failures):

$P(\geq 45$ prizes in 52 boxes$) = P(\leq 6$ boxes w/o a prize in 52 boxes$)$.

$\mu^* = 52 \cdot (0.05) = 2.6$

$$\Rightarrow \text{prob.} \approx \sum_{k=0}^{6} e^{-2.6} \cdot \frac{(2.6)^k}{k!} \approx 0.9828.$$

(Actual probability (binomial) is 0.9855
Normal approximation ≈ 0.9935)

Problem 2

A deck of cards is shuffled and dealt to four players, with each receiving 13 cards. Find:

← other players don't matter

a. the probability that the first player holds all the aces;

b. the probability that the first player holds all the aces given that she holds the ace of hearts;

c. the probability that the first player holds all the aces given that she holds at least one.

$$\boxed{P(A|B) = \frac{\#(AB)}{\#B}}$$

Ex 2.5.3 in Pitman's Probability

Bonus question:

(d) Prob. that second player has all the hearts given that the first player has all the aces.
Ans: 0

a) $$\frac{\binom{4}{4}\binom{48}{9}}{\binom{52}{13}}$$

b) $$\frac{\binom{3}{3}\binom{48}{9}}{\binom{51}{12}}$$

c) {Holding all the aces} $\cap$ {has at least one ace} = {Holding all the aces}.

$$\Rightarrow \frac{\binom{4}{4}\binom{48}{9}}{\underbrace{\binom{52}{13}}_{\text{all possible 13-card hands}} - \underbrace{\binom{48}{13}}_{\text{this many have no aces.}}}$$

Problem 3

Twelve cards are drawn from a well-shuffled deck of 52 cards. What is the probability the 12 cards contain

a. 4 aces;

b. 4 aces and 4 kings;

c. exactly 2 sets of four of a kind (any ranks).

Adapted from 2.rev.16 in Pitman's Probability

a) $$\frac{\binom{4}{4}\binom{48}{8}}{\binom{52}{12}} = \frac{\binom{48}{8}}{\binom{52}{12}}.$$

b) $$\frac{\binom{4}{4}\binom{4}{4}\binom{44}{4}}{\binom{52}{12}} = \frac{\binom{44}{4}}{\binom{52}{12}}.$$

c) $$\frac{\#(\text{hands w/ at least 2 sets of four of a kind}) - \#(\text{hands w/ 3 sets of four of a kind})}{\binom{52}{12}}$$

$$= \frac{\binom{13}{2}\binom{44}{4} - \binom{13}{3}\binom{4}{4}\binom{4}{4}\binom{4}{4}}{\binom{52}{12}} = \frac{\binom{13}{2}\binom{44}{4} - \binom{13}{3}}{\binom{52}{12}}.$$

Prepared by Brian Thorsen and Yiming Shi

Stat 134: Section 6

Adam Lucas

September 12, 2018

Problem 1

You flip a fair coin 10 times, and I flip a fair coin 10 times. What's the probability that we get the same number of heads? Write your answer without a summation.

$$\sum_{k=0}^{10} P(\text{I get } k \text{ H, You get } k \text{ H}) =$$

$$= \sum_{k=0}^{10} \binom{10}{k}\left(\frac{1}{2}\right)^{10}\binom{10}{k}\left(\frac{1}{2}\right)^{10} = \left(\frac{1}{20}\right)^{20}\sum_{k=0}^{10}\binom{10}{k}\binom{10}{10-k} =$$

$$= \left(\frac{1}{20}\right)^{20}\binom{20}{10}\sum_{k=0}^{10}\underbrace{\frac{\binom{10}{k}\binom{10}{10-k}}{\binom{20}{10}}} = \boxed{\binom{20}{10}\left(\frac{1}{20}\right)^{20}}$$

hypergeometric: from pop. size 20, sample 10 w/o replacement. 10 "good", 10 "bad" elements, pick k good elements.

Hint: Write down your answer using summation first. Now how can you relate your summation to a hypergeometric distribution?

*Alternate solution

Let X_H = # heads I get

Y_H = # heads you get

X_T = # tails I get, Y_T = # tails you get

$\Rightarrow$ *Note that # heads & # tails have the same distribution since coin is fair

$$\Rightarrow P(X_H = Y_H) = P(X_H - Y_H = 0) =$$

$$= P(X_H - (10 - Y_T) = 0) =$$

$$= P(X_H + Y_T = 10) = \boxed{\binom{20}{10}\left(\frac{1}{2}\right)^{20}}$$

$$= P(X_H + Y_H = 10)$$

*That is, $X_H \stackrel{d}{=} Y_T \stackrel{d}{=} Y_H \stackrel{d}{=} X_T$

Problem 2: Fisher's Exact Test

Suppose I am interested in determining whether or not taking a drug might cause a harmful side effect in its users. I sample 200 patients, and note whether they took the drug and whether they experienced the side effect. (The results are displayed below). Assuming there is no causal relationship, what is the chance that at least 16 of the 52 patients taking the drug experienced the side effect due to chance/random assortment?

	Side Effect	No Side Effect	
Taking Drugs	16	36	52
Not Taking Drugs	25	123	148
	41	159	$\boxed{200}$

$$\sum_{i=0}^{25}\frac{\binom{52}{16+i}\binom{148}{25-i}}{\binom{200}{41}} = \sum_{i=0}^{36}\frac{\binom{41}{16+i}\binom{159}{36-i}}{\binom{200}{52}}.$$

Problem 3

Draw five cards from a standard deck of cards to form a poker hand. Let E be the event of having no point. Find P(E).

$$\mathbb{P}(E) = \mathbb{P}(\text{singles, \& (not straight } \underline{\text{or}} \text{ flush)}) =$$

$$= \mathbb{P}(\text{singles}) - \mathbb{P}(\{\text{straight } \underline{\text{or}} \text{ flush}\}^c) =$$

$$= \mathbb{P}(\text{singles}) - \mathbb{P}(\text{straight}) - \mathbb{P}(\text{flush}) + \mathbb{P}(\text{straight \& flush}) =$$

$$= \frac{\binom{13}{5}\binom{4}{1}^5 - 10\binom{4}{1}^5 - \binom{4}{1}\binom{13}{5} + 10 \cdot 4}{\binom{52}{5}}.$$

Instead of subtracting the sum of probabilities of getting points, try to think how you can approach this directly.

Hint: For a poker hand to receive no point, they have to be all singles, but not a straight or flush.

Stat 134: Section 7

Adam Lucas

September 17th, 2018

Problem 1

Suppose the Stat department teaches 15 classes a semester: 2 have 60 students, 1 has 300 students, and 12 have 20 students. Each course is taught by a different professor, and each student only takes one class in the department.

a. For a randomly selected professor, what is the expected size of the class they teach?

b. For a randomly selected student, what is the expected size of the class they are in? How does this compare to part (a)?

a) 15 professors: 2 have 60 students,
1 has 300 students
12 have 20 students

$\Rightarrow$ Expected professors' class size is $\frac{2}{15} \cdot 60 + \frac{1}{15} \cdot 300 + \frac{12}{15} \cdot 20 = 44.$

b) There are $2 \cdot 60 + 300 + 12 \cdot 20 = 660$ students total.

120 students are in a class of size 60
300 students are in a class of size 300
240 students are in a class of size 20

$\Rightarrow$ Expected students' class size is $\frac{120}{660} \cdot 60 + \frac{300}{660} \cdot 300 + \frac{240}{660} \cdot 20 = \frac{1700}{11} = 154.545$

Problem 2

Let A and B be independent events, with indicator random variables I_A and I_B.

a. Describe the distribution of $(I_A + I_B)^2$ in terms of $P(A)$ and $P(B)$.

b. What is $\mathbb{E}\left[(I_A + I_B)^2\right]$?

c. Suppose we now have a set of identical but not necessarily independent indicators $I_1, I_2, \ldots, I_n$. Derive a useful formula for $\mathbb{E}\left[(I_1 + I_2 + \ldots + I_n)^2\right]$.

Hint: Expand the polynomial, then use linearity of expectations.

Ex 3.2.10 in Pitman's Probability

a) Possible values of $(I_A + I_B)^2$ are $\{0, 1, 4\}$

$$\mathbb{P}((I_A+I_B)^2 = 0) \underset{\uparrow}{=} \mathbb{P}(I_A = 0, I_B = 0) = \cancel{\mathbb{P}(A)\mathbb{P}(B)} = (1-\mathbb{P}(A))(1-\mathbb{P}(B)).$$

when does $(I_A+I_B)^2 = 0$?

$$\mathbb{P}((I_A+I_B)^2 = 1) = \mathbb{P}(I_A = 1, I_B = 0) + \mathbb{P}(I_A = 0, I_B = 1) =$$
$$= \mathbb{P}(A)(1-\mathbb{P}(B)) + \mathbb{P}(B)(1-\mathbb{P}(A)).$$

$$\mathbb{P}((I_A+I_B)^2 = 4) = \mathbb{P}(I_A = 1, I_B = 1) = \mathbb{P}(A)\,\mathbb{P}(B).$$

b) $$E[(I_A+I_B)^2] = E[I_A^2 + 2I_AI_B + I_B^2] = E[I_A^2] + 2\mathbb{E}[I_AI_B] + E[I_B^2] =$$
$$= \mathbb{P}(A) + 2\mathbb{P}(A)\mathbb{P}(B) + \mathbb{P}(B) =$$
$$= (\mathbb{P}(A) + \mathbb{P}(B))^2.$$

c) $$E\left[\left(\sum_{j=1}^{n} I_j\right)^2\right] = E\left[\sum_{j=1}^{n}\sum_{k=1}^{n} I_j I_k\right] = \sum_{j=1}^{n}\sum_{k=1}^{n} E[I_jI_k] =$$
$$= \sum_{j=1}^{n}\sum_{k=1}^{n} \mathbb{P}(I_j = 1, I_k = 1) =$$
$$= \sum_{j=1}^{n} \mathbb{P}(I_j = 1) + \sum_{j \neq k}^{n} \mathbb{P}(I_j = 1, I_k = 1).$$

Prepared by Brian Thorsen and Yiming Shi

Stat 134: Section 8

Adam Lucas

September 19th, 2018

Problem 1

In a well-shuffled standard deck of cards, we are interested in the number of adjacent pairs; i.e., cards which are the same rank as the card before or after them in the deck. Calculate the expected number of adjacent pairs.

Hint: consider the probability that a card is the same as the card before it.

For $j = 2, 3, \dots, 52$, define $I_j = \begin{cases} 1 & \text{if } j\text{-1-\& } j\text{-th cards same rank} \\ 0 & \text{otherwise} \end{cases}$

Let $X = \#$ adjacent pairs.

Then $X = I_2 + \dots + I_{52} \Rightarrow E(X) = \sum_{j=2}^{52} E[I_j] =$

$$= \sum_{j=2}^{52} P(I_j = 1) = \sum_{j=2}^{52} \frac{3}{51} = 51 \cdot \frac{3}{51} = 3.$$

Problem 2

Suppose the IQ scores of a million individuals have a mean of 100 and an SD of 10.

a. Without any further assumptions, find a bound for the proportion of individuals with an IQ over 130.

b. Now find a smaller upper bound, assuming the distribution is symmetric about 100.

c. Now suppose the scores follow a Normal curve. Find the proportion of individuals with an IQ over 130.

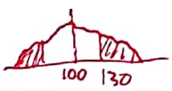
Ex 3.3.13 in Pitman's Probability

Let $X =$ IQ score: think of X as distribution

a) Markov's: $P(X \geq 130) \leq \frac{E(X)}{130} = \frac{100}{130} = \frac{10}{13}$.

Chebyshev's: $P(X \geq 130) = P(X - E(X) \geq 130 - E(X)) = P(X - E(X) \geq 30) \leq$

$\leq P(|X - E(X)| \geq 30) \leq \frac{Var(X)}{30^2} = \left(\frac{10}{30}\right)^2 = \frac{1}{9}$.



b) $P(X \geq 130) = P(X - 100 \geq 30) = \frac{1}{2} P(|X - 100| \geq 30) \leq \frac{1}{2 \cdot 9} = \frac{1}{18}$.

c) $1 - \Phi\left(\frac{130 - 100}{10}\right) = 1 - \Phi(3)$.

Problem 3

Suppose we have n unique pairs of chopsticks in a drawer (so $2n$ sticks in total). Hurrying to prepare for dinner, we grab k pairs of these at random from the drawer and try to make matching pairs from this pile of $2k$ chopsticks. Let X represent the number of matching pairs. Find $E(X)$ and $Var(X)$.

For $j = 1, 2, \ldots, n$, define $I_j = \begin{cases} 1 & \text{if pair } j \text{ was chosen} \\ 0 & \text{otherwise} \end{cases}$.

Then $X = \sum_{j=1}^{n} I_j$, so $E(X) = \sum_{j=1}^{n} E(I_j) = \sum_{j=1}^{n} \mathbb{P}(\text{pair } j \text{ was chosen}) =$

$$= n \cdot \frac{\binom{2}{2}\binom{2n-2}{2k-2}}{\binom{2n}{2k}}.$$

For variance, find $E(X^2)$ first.

$$E\left[\left(\sum_{j=1}^{n} I_j\right)^2\right] = E\left[\sum_{j=1}^{n}\sum_{k=1}^{n} I_j I_k\right] = \sum_{j=1}^{n} E[I_j^2] + \sum_{j \neq k}^{n} E[I_j I_k] =$$

$$= \sum_{j=1}^{n} \mathbb{P}(I_j = 1) + \sum_{j \neq k}^{n} \mathbb{P}(I_j = 1, I_k = 1) =$$

$$= n \cdot \frac{\binom{2n-2}{2k-2}}{\binom{2n}{2k}} + n(n-1) \cdot \frac{\binom{4}{4}\binom{2n-4}{2k-4}}{\binom{2n}{2k}},$$

$$\Rightarrow Var(X) = n \cdot \frac{\binom{2n-2}{2k-2}}{\binom{2n}{2k}} + n(n-1) \cdot \frac{\binom{2n-4}{2k-4}}{\binom{2n}{2k}} - \left[n \cdot \frac{\binom{2n-2}{2k-2}}{\binom{2n}{2k}}\right]^2.$$

Stat 134: Section 9

Adam Lucas

September 24, 2018

Problem 1

Suppose that in a particular application requiring a single battery, the mean lifetime of a battery is 4 weeks, with an SD of 1 week. The battery is replaced with a new one when it dies, and so on. Assume battery lifetimes are independent. Approximate the chance that more than 26 replacements will have to be made in a two year period, starting with a fresh battery and not counting that one as a replacement.

Should we use the continuity correction here? Why/why not?

Ex 3.3.23 in Pitman's Probability

Let X_i = lifetime of i-th battery, in weeks

Then more than 26 replacements in 2 yrs. $\Leftrightarrow X_1 + \cdots + X_{27} \le 2 \cdot 52 = 104$.

$\Rightarrow$ Approximate $P(X_1 + \cdots + X_{27} \le 104)$.

Note that $E(X_1 + \cdots + X_{27}) = 27 \cdot E(X_1) = 27 \cdot 4 = 108$,

$Var(X_1 + \cdots + X_{27}) = 27 \cdot Var(X_1) = 27$. Let $S = X_1 + \cdots + X_{27}$.

By normal approx., we get $P(S \le 104) = P\left(\frac{S-108}{\sqrt{27}} \le \frac{104-108}{\sqrt{27}}\right) \approx \Phi(-0.77) \approx 0.22$

Problem 2

Bill, Mary, and Tom have coins with respective probabilities p_1, p_2, p_3 of turning up heads. They toss their coins independently at the same times.

a. What is the probability that the first person to get a head has to toss more than n times? (What distribution does this follow?)

b. What is the probability that neither Bill nor Tom get a head before Mary?

Ex 3.4.5 in Pitman's Probability

Let $X_i \sim Geometric(p_i)$, $i = 1, 2, 3$.

- X_1 = # tosses for Bill until 1st heads
- X_2 = # tosses for Mary until 1st heads
- X_3 = # tosses for Tom until 1st heads

a) # tosses until first head is $\min\{X_1, X_2, X_3\}$. Let $X = \min\{X_1, X_2, X_3\}$.

$P(X > n) = P(X_1 > n, X_2 > n, X_3 > n) = (1-p_1)^n (1-p_2)^n (1-p_3)^n$.

b) This is $P(X_1 \ge X_2, X_3 \ge X_2) =$ (condition on $X_2 = k$)

$= \sum_{k=1}^{\infty} P(X_1 \ge X_2, X_3 \ge X_2, X_2 = k) =$

$= \sum_{k=1}^{\infty} P(X_2 = k) P(X_1 \ge k, X_3 \ge k) = \sum_{k=1}^{\infty} (1-p_2)^{k-1} p_2 \cdot (1-p_1)^{k-1} (1-p_3)^{k-1} =$

$= p_2 \sum_{k=0}^{\infty} [(1-p_1)(1-p_2)(1-p_3)]^k$.

Problem 3: Mean Absolute Deviation

One alternative method of measuring the spread of a distribution is the *mean absolute deviation*. For a random variable X, this is given by $E(|X-\mu|)$, where $\mu = E(X)$.

a. Let X be the result of a fair standard die. Calculate the mean absolute deviation of X. For comparison, $SD(X) \approx 1.71$.

b. Use the fact that $Var(|X-\mu|) \geq 0$ for all X to prove that $SD(X) \geq E(|X-\mu|)$, with equality if and only if $|X-\mu|$ is a constant.

Ex 3.3.26 in Pitman's Probability

Recall $E(X) = 7/2 = 3.5$

a) $E(|X-\mu|) = \sum_{x=1}^{6} P(X=x)\,|x-3.5| = \frac{1}{6}(2.5+1.5+0.5+0.5+1.5+2.5) = \frac{9}{6} = \frac{3}{2}$.

b) $Var(X) = Var(X-\mu) = E[(X-\mu)^2] - \cancel{(E[X-\mu])^2}^{\,0} = E[|X-\mu|^2] =$

$= Var(|X-\mu|) + (E[|X-\mu|])^2$.

Since $Var(|X-\mu|) \geq 0$, we have $Var(X) \geq (E|X-\mu|)^2$

$\Rightarrow SD(X) \geq E(|X-\mu|)$.

Prepared by Brian Thorsen and Yiming Shi

Stat 134: Section 10

Adam Lucas

September 26, 2018

Problem 1

In Bernoulli (p) trials let V_n be the number of trials required to produce either n successes or n failures, whichever comes first.

a. Write down the range of possible values of V_n.

b. Find the distribution of V_n.

Ex 3.4.14 in Pitman's Probability

a) $V_n \in \{n, n+1, n+2, \dots, 2n-1\}$.

b) $\mathbb{P}(V_n = k) = \mathbb{P}(n \text{ failures, } k\text{-th trial is failure}) +$

For $k \in \{n, n+1, \dots, 2n-1\}$,

$+ \mathbb{P}(n \text{ successes, } k\text{-th trial is success}) =$

$= \mathbb{P}(n-1 \text{ failures in } k-1 \text{ trials, then failure}) +$

$+ \mathbb{P}(n-1 \text{ successes, in } k-1 \text{ trials, then success}) =$

$$= \binom{k-1}{n-1}(1-p)^{n-1} p^{k-n} \cdot (1-p) + \binom{k-1}{n-1} p^{n-1} (1-p)^{k-n} \cdot p =$$

$$= \binom{k-1}{n-1}(1-p)^{n} p^{k-n} + \binom{k-1}{n-1} p^{n} (1-p)^{k-n} =$$

$$= \binom{k-1}{n-1}\left(p^{n}(1-p)^{k-n} + (1-p)^{n} p^{k-n}\right).$$

Problem 2

In Bernoulli(p) trials, let X = number of trials until there are 2 more successes than failures or 2 more failures than successes. Find $P(X = x)$.

First, find all possible values for X.

$\rightarrow X \in \{2, 4, 6, 8, \dots\}$.

Trick: View every 2 trials as one, i.e. SF FS SF SS
4 "trials"

View "success" as SS or FF.

Then probability of "success" is $p^2 + (1-p)^2$.

$$\Rightarrow P(X=x) = P(\text{first } \tfrac{x-2}{2} \text{ "trials" are "failure", then FF or SS}) =$$

$$= \left[2p(1-p)\right]^{\frac{x-2}{2}} \cdot \left[p^2 + (1-p)^2\right].$$