

Due: Aug 3, 2023 11:59pm
Grace period until Aug 4, 2023 11:59pm

Sundry

Before you start writing your final homework submission, state briefly how you worked on it. Who else did you work with? List names and email addresses. (In case of homework party, you can just describe the group.)

1 Will I Get My Package?

Note 15
Note 16

A delivery guy in some company is out delivering n packages to n customers, where n is a natural number greater than 1. Not only does he hand each customer a package uniformly at random from the remaining packages, he opens the package before delivering it with probability $1/2$. Let X be the number of customers who receive their own packages unopened.

- (a) Compute the expectation $\mathbb{E}[X]$.
- (b) Compute the variance $\text{Var}(X)$.

2 Double-Check Your Intuition Again

Note 16

- (a) You roll a fair six-sided die and record the result X . You roll the die again and record the result Y .
 - (i) What is $\text{cov}(X + Y, X - Y)$?
 - (ii) Prove that $X + Y$ and $X - Y$ are not independent.

For each of the problems below, if you think the answer is "yes" then provide a proof. If you think the answer is "no", then provide a counterexample.

- (b) If X is a random variable and $\text{Var}(X) = 0$, then must X be a constant?
- (c) If X is a random variable and c is a constant, then is $\text{Var}(cX) = c \text{Var}(X)$?
- (d) If A and B are random variables with nonzero standard deviations and $\text{Corr}(A, B) = 0$, then are A and B independent?

- (e) If X and Y are not necessarily independent random variables, but $\text{Corr}(X, Y) = 0$, and X and Y have nonzero standard deviations, then is $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$?
- (f) If X and Y are random variables then is $\mathbb{E}[\max(X, Y) \min(X, Y)] = \mathbb{E}[XY]$?
- (g) If X and Y are independent random variables with nonzero standard deviations, then is

$$\text{Corr}(\max(X, Y), \min(X, Y)) = \text{Corr}(X, Y)?$$

3 Tellers

Note 17

Imagine that X is the number of customers that enter a bank at a given hour. To simplify everything, in order to serve n customers you need at least n tellers. One less teller and you won't finish serving all of the customers by the end of the hour. You are the manager of the bank and you need to decide how many tellers there should be in your bank so that you finish serving all of the customers in time. You need to be sure that you finish in time with probability at least 95%.

- (a) Assume that from historical data you have found out that $\mathbb{E}[X] = 5$. How many tellers should you have?
- (b) Now assume that you have also found out that $\text{Var}(X) = 5$. Now how many tellers do you need?

4 Those 3407 Votes

Note 17

In the aftermath of the 2000 US Presidential Election, many people have claimed that unusually large number of votes cast for Pat Buchanan in Palm Beach County are statistically highly significant, and thus of dubious validity. In this problem, we will examine this claim from a statistical viewpoint.

The total percentage votes cast for each presidential candidate in the entire state of Florida were as follows:

Gore	Bush	Buchanan	Nader	Browne	Others
48.8%	48.9%	0.3%	1.6%	0.3%	0.1%

In Palm Beach County, the actual votes cast (before the recounts began) were as follows:

Gore	Bush	Buchanan	Nader	Browne	Others	Total
268945	152846	3407	5564	743	781	432286

To model this situation probabilistically, we need to make some assumptions. Let's model the vote cast by each voter in Palm Beach County as a random variable X_i , where X_i takes on each of the six possible values (five candidates or "Others") with probabilities corresponding to the Florida percentages. (Thus, e.g., $\mathbb{P}[X_i = \text{Gore}] = 0.488$.) There are a total of $n = 432286$ voters, and their votes are assumed to be mutually independent. Let the r.v. B denote the total votes cast for Buchanan in Palm Beach County (i.e., the number of voters i for which $X_i = \text{Buchanan}$).

- (a) Compute the expectation $\mathbb{E}[B]$ and the variance $\text{Var}(B)$.
- (b) Use Chebyshev's inequality to compute an *upper bound* b on the probability that Buchanan receives at least 3407 votes, i.e., find a number b such that

$$\mathbb{P}[B \geq 3407] \leq b.$$

Based on this result, do you think Buchanan's vote is significant?

- (c) Suppose that your bound b in part (b) is exactly accurate, i.e., assume that $\mathbb{P}[X \geq 3407]$ is exactly equal to b . [*In fact the true value of this probability is much smaller.*] Suppose also that all 67 counties in Florida have the same number of voters as Palm Beach County, and that all behave independently according to the same statistical model as Palm Beach County. What is the probability that in *at least one* of the counties, Buchanan receives at least 3407 votes? How would this affect your judgment as to whether the Palm Beach tally is significant?
- (d) Our model assumes that all voters behave like the fabled "swing voters," in the sense that they are undecided when they go to the polls and end up making a random decision. A more realistic model would assume that only a fraction (say, about 20%) of voters are in this category, the others having already decided. Suppose then that 80% of the voters in Palm Beach County vote deterministically according to the state-wide proportions for Florida, while the remaining 20% behave randomly as described earlier. Does your bound b in part (b) increase, decrease or remain the same under this model? Justify your answer.

5 Markov Chains: Prove/Disprove

Note 22

Prove or disprove the following statements.

- (a) There exists an irreducible, finite Markov chain for which there exist initial distributions that converge to different distributions.
- (b) There exists an irreducible, aperiodic, finite Markov chain for which $\mathbb{P}(X_{n+1} = j | X_n = i) = 1$ or 0 for all i, j .
- (c) There exists an irreducible, non-aperiodic Markov chain for which $\mathbb{P}(X_{n+1} = j | X_n = i) \neq 1$ for all i, j .
- (d) For an irreducible, non-aperiodic Markov chain, any initial distribution not equal to the invariant distribution does not converge to any distribution.

6 Faulty Machines

Note 22

You are trying to use a machine that only works on some days. If on a given day the machine is working, it will break down the next day with probability $0 < b < 1$, and works on the next day $1 - b$. If it is not working on a given day, it will work on the next day with probability $0 < r < 1$, and not work on the next day with probability $1 - r$. Formulate this process as a Markov chain. As

$n \rightarrow \infty$, what does the probability that on a given day the machine is working converge to? What properties of the Markov chain allow us to conclude that the probability will actually converge?

7 Boba in a Straw

Note 22

Imagine that Gavin is drinking milk tea and he has a very short straw: it has enough room to fit two boba (see figure).

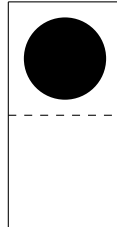


Figure 1: A straw with one boba currently inside. The straw only has enough room to fit two boba.

Here is a formal description of the drinking process: We model the straw as having two “components” (the top component and the bottom component). At any given time, a component can contain nothing, or one boba. As Gavin drinks from the straw, the following happens every second:

1. The contents of the top component enter Gavin’s mouth.
2. The contents of the bottom component move to the top component.
3. With probability p , a new boba enters the bottom component; otherwise the bottom component is now empty.

Help Gavin evaluate the consequences of his incessant drinking!

- (a) Draw the Markov chain that models this process, and show that it is both irreducible and aperiodic.
- (b) At the very start, the straw starts off completely empty. What is the expected number of seconds that elapse before the straw is completely filled with boba for the first time? [Write down the equations; you do not have to solve them.]
- (c) Consider a slight variant of the previous part: now the straw is narrower at the bottom than at the top. This affects the drinking speed: if either (i) a new boba is about to enter the bottom component or (ii) a boba from the bottom component is about to move to the top component, then the action takes two seconds. If both (i) and (ii) are about to happen, then the action takes three seconds. Otherwise, the action takes one second. Under these conditions, answer the previous part again. [Write down the equations; you do not have to solve them.]

- (d) Gavin was annoyed by the straw so he bought a fresh new straw (same as the straw from Figure 1). What is the long-run average rate of Gavin's calorie consumption? (Each boba is roughly 10 calories.)
- (e) What is the long-run average number of boba which can be found inside the straw? [Maybe you should first think about the long-run distribution of the number of boba.]
- (f) What is the long-run probability that the amount of boba in the straw doesn't change from one second to the next?