CS 70 Summer 2023 — Discrete Mathematics and Probability Theory — Huang, Suzani, and Tausik — **HW 6**

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 $\text{Var}(X)$.

Solution:

(a) Define

$$X_i = \begin{cases} 1, & \text{if the } i\text{-th customer gets his/her package unopened,} \\ 0, & \text{otherwise.} \end{cases}$$

By linearity of expectation, $\mathbb{E}[X] = \mathbb{E}[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} \mathbb{E}[X_i]$. We have

$$\mathbb{E}[X_i] = \mathbb{P}[X_i = 1] = \frac{1}{2n},$$

since the ith customer will get his/her own package with probability $1/n$ and it will be unopened with probability $1/2$ and the delivery guy opens the packages independently.
Hence,

$$\mathbb{E}[X] = n \cdot \frac{1}{2n} = \boxed{\frac{1}{2}}.$$

(b) To calculate $\text{Var}(X)$, we need to know $\mathbb{E}[X^2]$.
By linearity of expectation,

$$\mathbb{E}[X^2] = \mathbb{E}\left[(X_1 + X_2 + \cdots + X_n)^2\right] = \mathbb{E}\left[\sum_{i,j} X_i X_j\right] = \sum_{i,j} \mathbb{E}[X_i X_j].$$

Then we consider two cases, either $i = j$ or $i \neq j$.
Hence $\sum_{i,j} \mathbb{E}[X_i X_j] = \sum_{i} \mathbb{E}[X_i^2] + \sum_{i \neq j} \mathbb{E}[X_i X_j]$.

$$\mathbb{E}[X_i^2] = \mathbb{E}[X_i] = \frac{1}{2n}$$

for all i. To find $\mathbb{E}[X_i X_j]$, we need to calculate $\mathbb{P}[X_i X_j = 1]$:

$$\mathbb{P}[X_i X_j = 1] = \mathbb{P}[X_i = 1]\mathbb{P}[X_j = 1 \mid X_i = 1] = \frac{1}{2n} \cdot \frac{1}{2(n-1)}$$

since if customer i has received his/her own package, customer j has $n-1$ choices left. Hence,

$$\mathbb{E}[X^2] = n \cdot \frac{1}{2n} + n \cdot (n-1) \cdot \frac{1}{2n} \cdot \frac{1}{2(n-1)} = \frac{3}{4},$$

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{3}{4} - \frac{1}{4} = \boxed{\frac{1}{2}}.$$

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 $\mathrm{cov}(X+Y, X-Y)$?

(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 $\mathrm{Var}(X) = 0$, then must X be a constant?

(c) If X is a random variable and c is a constant, then is $\mathrm{Var}(cX) = c\,\mathrm{Var}(X)$?

(d) If A and B are random variables with nonzero standard deviations and $\mathrm{Corr}(A,B) = 0$, then are A and B independent?

(e) If X and Y are not necessarily independent random variables, but $\mathrm{Corr}(X,Y) = 0$, and X and Y have nonzero standard deviations, then is $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$?

(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

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

Solution:

(a) (i) Using bilinearity of covariance, we have

$$\begin{aligned}
\mathrm{cov}(X+Y, X-Y) &= \mathrm{cov}(X,X) + \mathrm{cov}(X,Y) - \mathrm{cov}(Y,X) - \mathrm{cov}(Y,Y) \\
&= \mathrm{cov}(X,X) - \mathrm{cov}(Y,Y), \\
&= 0
\end{aligned}$$

where we use that $\mathrm{cov}(X,Y) = \mathrm{cov}(Y,X)$ to get the second equality.

(ii) Observe that $\mathbb{P}[X+Y=7, X-Y=0]=0$ because if $X-Y=0$, then the sum of our two dice rolls must be even. However, both $\mathbb{P}[X+Y=7]$ and $\mathbb{P}[X-Y=0]$ are nonzero, so $\mathbb{P}[X+Y=7, X-Y=0] \neq \mathbb{P}[X+Y=7] \cdot \mathbb{P}[X-Y=0]$.

(b) Yes. If we write $\mu = \mathbb{E}[X]$, then $0 = \operatorname{Var}(X) = \mathbb{E}[(X-\mu)^2]$ so $(X-\mu)^2$ must be identically 0 since perfect squares are non-negative. Thus $X = \mu$.

(c) No. We have $\operatorname{Var}(cX) = \mathbb{E}[(cX - \mathbb{E}[cX])^2] = c^2\,\mathbb{E}[(X-\mathbb{E}[X])^2] = c^2 \operatorname{Var}(X)$ so if $\operatorname{Var}(X) \neq 0$ and $c \neq 0$ or $c \neq 1$ then $\operatorname{Var}(cX) \neq c\operatorname{Var}(X)$. This does prove that $\sigma(cX) = c\sigma(X)$ though.

(d) No. Let $A = X+Y$ and $B = X-Y$ from part (a). Since A and B are not constants then part (b) says they must have nonzero variances which means they also have nonzero standard deviations. Part (a) says that their covariance is 0 which means they are uncorrelated, and that they are not independent.

Recall from lecture that the converse is true though.

(e) Yes. If $\operatorname{Corr}(X,Y)=0$, then $\operatorname{cov}(X,Y)=0$. We have $\operatorname{Var}(X+Y) = \operatorname{cov}(X+Y, X+Y) = \operatorname{Var}(X) + \operatorname{Var}(Y) + 2\operatorname{cov}(X,Y) = \operatorname{Var}(X) + \operatorname{Var}(Y)$.

(f) Yes. For any values x,y we have $\max(x,y)\min(x,y) = xy$. Thus, $\mathbb{E}[\max(X,Y)\min(X,Y)] = \mathbb{E}[XY]$.

(g) No. You may be tempted to think that because $(\max(x,y), \min(x,y))$ is either (x,y) or (y,x), then $\operatorname{Corr}(\max(X,Y), \min(X,Y)) = \operatorname{Corr}(X,Y)$ because $\operatorname{Corr}(X,Y) = \operatorname{Corr}(Y,X)$. That reasoning is flawed because $(\max(X,Y), \min(X,Y))$ is not always equal to (X,Y) or always equal to (Y,X) and the inconsistency affects the correlation. It is possible for X and Y to be independent while $\max(X,Y)$ and $\min(X,Y)$ are not.

For a concrete example, suppose X is either 0 or 1 with probability $1/2$ each and Y is independently drawn from the same distribution. Then $\operatorname{Corr}(X,Y)=0$ because X and Y are independent. Even though X never gives information about Y, if you know $\max(X,Y)=0$ then you know for sure $\min(X,Y)=0$.

More formally, $\max(X,Y)=1$ with probability $3/4$ and 0 with probability $1/4$, and $\min(X,Y) = 1$ with probability $1/4$ and 0 with probability $3/4$. This means

$$\mathbb{E}[\max(X,Y)] = 1 \cdot \frac{3}{4} + 0 \cdot \frac{1}{4} = \frac{3}{4}$$

and

$$\mathbb{E}[\min(X,Y)] = 1 \cdot \frac{1}{4} + 0 \cdot \frac{3}{4} = \frac{1}{4}.$$

Thus,

$$\begin{aligned}\operatorname{cov}(\max(X,Y), \min(X,Y)) &= \mathbb{E}[\max(X,Y)\min(X,Y)] - \frac{3}{16} \\ &= \frac{1}{4} - \frac{3}{16} = \frac{1}{16} \neq 0\end{aligned}$$

We conclude that $\operatorname{Corr}(\max(X,Y), \min(X,Y)) \neq 0 = \operatorname{Corr}(X,Y)$.

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?

Solution:

(a) You should apply Markov's. Aiming for the probability of finishing t in time to be least 95% is equivalent to aiming to limit the probability of not finishing (or in other words, taking more time to finish t customers) as 5%. So, you want

$$\mathbb{P}[X \geq t] \leq 0.05$$

Using Markov's, we know $\mathbb{P}[X \geq t] \leq \frac{\mathbb{E}[X]}{t}$. Therefore you want $\frac{\mathbb{E}[X]}{t} = 0.05$ for the above inequality to hold. Therefore, $t = 100$ to be able to limit the probability of not finishing under 5%. The inequality becomes $\mathbb{P}[X \geq 100] \leq 0.05$ as wanted. So you need 99 tellers.

(b) You should apply Chebyshev's. Same as part a, aiming for the probability of finishing t in time to be least 95% is equivalent to aiming to limit the probability of not finishing (or in other words, taking more time to finish t customers) as 5%. So, you want

$$\mathbb{P}[|X - E[X]| \geq t] \leq 0.05$$

Using Chebyshev's we know $\mathbb{P}[|X - E[X]| \geq t] \leq \frac{Var(X)}{t^2}$. Plugging in $E[X] = 5$ and $Var(X) = 5$, you get $\mathbb{P}[|X-5| \geq t] \leq \frac{5}{t^2}$. Since you want to limit $\mathbb{P}[|X-5| \geq t] \leq 0.05$ you get $\frac{5}{t^2} = 0.05$. Thus $t^2 = 100$ and $t = 10$. Now plugging $t = 10$:

$$\mathbb{P}[|X-5| \geq 10] = \mathbb{P}[X \geq 5 + 10] = \mathbb{P}[X \geq 15] \leq 0.05$$

as wanted. So you need 14 tellers this time.

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.

Solution:

1. Let B_i be a random variable representing whether the ith person voted for Buchanan. Then $B_i = 1$ if and only if $X_i = \text{Buchanan}$, so $B_i \sim \text{Bernoulli}(0.003)$. Note that the B_i's are independently and identically distributed, with $\mathbb{E}[B_i] = 0.003$ and $\text{Var}(B_i) = 0.003 \times (1 - 0.003) =$

0.002991. Moreover, by linearity of expectation and independence, we find that $\mathbb{E}[B] = \sum_{i=1}^{n} \mathbb{E}[B_i] = 432286 \times 0.003 \approx 1297$ and $\operatorname{Var}(B) = \sum_{i=1}^{n} \operatorname{Var}(B_i) = 432286 \times 0.002991 \approx 1293$.

2. Chebyshev's inequality says that

$$\mathbb{P}[|B - \mathbb{E}[B]| \geq a] \leq \frac{\operatorname{Var}(B)}{a^2}.$$

In our case $\mathbb{E}[B] = 1297$ and $\operatorname{Var}(B) = 1293$, so if we take $a = 2110$, we find that $\mathbb{P}[|B - 1297| \geq 2110] \leq 1293/2110^2 \approx 0.0003$. Now note that the condition $|B - 1297| < 2110$ is equivalent to the condition $-813 < B < 3407$, and since B is non-negative, we find that $\mathbb{P}[B > 3407] \leq 0.0003$ (roughly), so we can take $b \approx 0.0003$. In other words, receiving 3407 votes for Buchanan in Palm Beach County seems very unlikely to happen by chance, under this simple model. So yes, this is statistically significant.

3. Let p_j be the probability that the jth county does not receive 3407 votes for Buchanan. We have from part (b) that $p_j = 1 - b \approx 0.9997$. Note that the probability that no county yields at least 3407 votes for Buchanan is $p_1 \times \cdots \times p_{67}$, since the voters in each county behave independently. Thus, the probability that Buchanan does not receive 3407 votes in any county is about $(0.9997)^{67} \approx 0.98$. Consequently, the probability that Buchanan *does* receive at least 3407 votes in some county is about $1 - 0.98 \approx 0.02$. In other words, this seems unlikely to happen by chance.

4. In the modified model, b would decrease, since an even larger portion of the "random" voters would have to vote for Buchanan to yield such an unusually large proportion of Buchanan-votes.

   Note: We can be more precise. Chebyshev says that $\mathbb{P}[|B - \mu| \geq \alpha] \leq \operatorname{Var}(B)/\alpha^2$. We know that $\operatorname{Var}(B) = n \operatorname{Var}(B_i)$ where n is the total number of random "swing" voters in the county; so by reducing the number of swing voters, we reduce the variance linearly. However, the new $\alpha' = 3407 - 432286 \cdot 0.8 \cdot 0.003 - 432286 \cdot 0.2 \cdot 0.003$ does not change (i.e., $\alpha' = \alpha$). Hence, the likelihood decreases.

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 $\mathbb{P}(X_{n+1} = j \mid X_n = i) = 1$ or 0 for all i, j.

(c) There exists an irreducible, non-aperiodic Markov chain for which $\mathbb{P}(X_{n+1} = j \mid X_n = i) \neq 1$ 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.

Solution:

(a) False. Every finite irreducible Markov chain has a unique stationary distribution. If it's possible for the Markov chain to converge to two different distributions given different starting distributions, it implies there are two stationary distributions. To elaborate further, we know in the long run the fraction of time spent in each state converges to the stationary distribution. So if the distribution converges, the long-run fraction of time will be whatever distribution it converges to, which we see must be the stationary distribution.

(b) True, you can have one state pointing to itself. However for number of states > 1 it is false. Consider the initial distribution of having a probability of 1 of being in an arbitrary state. After a transition, the resulting distribution must be a probability 1 of being in a different state (if it were the same state, this would immediately imply that the Markov chain is reducible). Further transitions have the same effect. Therefore this initial distribution does not converge. Therefore this Markov chain cannot be aperiodic and irreducible (since it would converge in that case).

(c) True. Consider the states $\{0,1,2,3\}$. Set $P(i,j) = 1/2$ if $i \equiv j \pm 1 \mod 4$ and 0 otherwise. In other words, the Markov chain is a square with each side replaced with two links pointing in opposite directions with probabilities of $1/2$. Consider the period of state 0. Any path from 0 back to itself, such as $0-1-2-1-0$, alternates in parity of each consecutive state since each state only points to the state above or below it mod 4. Therefore state 0 has period 2. Therefore this Markov chain is not aperiodic (and all states have period 2).

(d) False. Take the initial distribution $\begin{bmatrix} 0.25 & 0.30 & 0.25 & 0.20 \end{bmatrix}$ for the above Markov chain. After one transition it goes to the invariant distribution, $\begin{bmatrix} 0.25 & 0.25 & 0.25 & 0.25 \end{bmatrix}$.

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 $n \to \infty$, 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?

Solution: We define the following states $\chi = \{W, B\}$ where W is the state that represents the machine working on a given day, and B is the state that represents the machine broken on a given day. The following are the transition probabilities.

$$\mathbb{P}(W,B) = b; \mathbb{P}(W,W) = 1-b; \mathbb{P}(B,W) = r; \mathbb{P}(B,B) = 1-r$$

We know that the Markov chain is finite and irreducible. Hence, it has a unique invariant distribution π

Furthermore, since the Markov chain has a self-loop of nonzero probability, it is aperiodic.

Hence, for any probability distribution of states at time n, $\pi_n : \lim_{n\to\infty} \pi_n = \pi$

We use the balance equations to find the invariant distribution.

$$\pi = \pi P$$
$$\pi(W) = (1-b)\pi(W) + r\pi(B)$$
$$\pi(B) = b\pi(W) + (1-r)\pi(B)$$
$$\pi(B) + \pi(W) = 1$$
$$\implies \pi(W) = \frac{r}{b+r}, \pi(B) = \frac{b}{b+r}$$

$\implies$ As $n \to \infty$, the probability that on a given day the machine is working is $\pi(W) = \dfrac{r}{b+r}$

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).

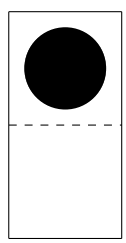

Figure 1: A straw with one boba currently inside. The straw only has enough room to fit two boba.

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 Jonathan'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?

Solution:

(a) We model the straw as a four-state Markov chain. The states are $\{(0,0),(0,1),(1,0),(1,1)\}$, where the first component of a state represents whether the top component is empty (0) or full (1); similarly, the second component represents whether the bottom component is empty or full. See Figure 2. This chain is irreducible as we can get from any state to any other with the cycle $(0,0) \to (0,1) \to (1,0) \to (0,1) \to (1,1) \to (1,0) \to (0,0)$. Furthermore, this chain contains a self-loop, so it is aperiodic.

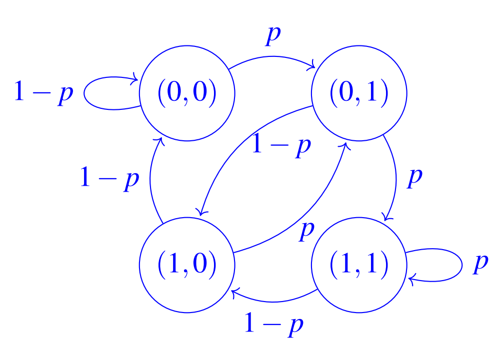

Figure 2: Transition diagram for the Markov chain.

(b) We set up the hitting time equations. Let T denote the time it takes to reach state $(1,1)$, i.e. $T = \min\{n > 0 : X_n = (1,1)\}$. Let $\mathbb{E}_i[\cdot] = \mathbb{E}[\cdot \mid X_0 = i]$ denote the expectation starting from state i (for convenience of notation). The hitting-time equations are

$$\begin{aligned}
\mathbb{E}_{(0,0)}[T] &= 1 + (1-p)\,\mathbb{E}_{(0,0)}[T] + p\,\mathbb{E}_{(0,1)}[T], \\
\mathbb{E}_{(0,1)}[T] &= 1 + (1-p)\,\mathbb{E}_{(1,0)}[T] + p\,\mathbb{E}_{(1,1)}[T], \\
\mathbb{E}_{(1,0)}[T] &= 1 + (1-p)\,\mathbb{E}_{(0,0)}[T] + p\,\mathbb{E}_{(0,1)}[T], \\
\mathbb{E}_{(1,1)}[T] &= 0.
\end{aligned}$$

The question did not ask you to solve the equations. If you solved the equations anyway and would like to check your work, the hitting time is $\mathbb{E}_{(0,0)}[T] = (1+p)/p^2$.

(c) The new hitting-time equations are

$$\begin{aligned}
\mathbb{E}_{(0,0)}[T] &= (1-p)(1+\mathbb{E}_{(0,0)}[T]) + p(2+\mathbb{E}_{(0,1)}[T]), \\
\mathbb{E}_{(0,1)}[T] &= (1-p)(2+\mathbb{E}_{(1,0)}[T]) + p(3+\mathbb{E}_{(1,1)}[T]), \\
\mathbb{E}_{(1,0)}[T] &= (1-p)(1+\mathbb{E}_{(0,0)}[T]) + p(2+\mathbb{E}_{(0,1)}[T]), \\
\mathbb{E}_{(1,1)}[T] &= 0.
\end{aligned}$$

You did not have to solve the equations, but to get a sense for what the solution is like, solving the equations and plugging in $p = 1/2$ yields (after some tedious algebra) $\mathbb{E}_{(0,0)}[T] = 11$.

(d) This part is actually more straightforward than it might initially seem: the average rate at which Gavin consumes boba must equal the average rate at which boba enters the straw, which is p per second. Hence, his long-run average calorie consumption rate is $10p$ per second.

(e) We compute the stationary distribution. The balance equations are

$$\begin{aligned}
\pi(0,0) &= (1-p)\pi(0,0) + (1-p)\pi(1,0), \\
\pi(0,1) &= p\pi(0,0) + p\pi(1,0), \\
\pi(1,0) &= (1-p)\pi(0,1) + (1-p)\pi(1,1), \\
\pi(1,1) &= p\pi(0,1) + p\pi(1,1).
\end{aligned}$$

Expressing everything in terms of $\pi(0,0)$, we find

$$\begin{aligned}
\pi(0,1) = \pi(1,0) &= \frac{p}{1-p}\pi(0,0), \\
\pi(1,1) &= \frac{p^2}{(1-p)^2}\pi(0,0).
\end{aligned}$$

From the normalization condition we have

$$\pi(0,0)\left(1 + \frac{2p}{1-p} + \frac{p^2}{(1-p)^2}\right) = 1,$$

so $\pi(0,0) = (1-p)^2$. Hence, the stationary distribution is

$$\begin{aligned}
\pi(0,0) &= (1-p)^2, \\
\pi(0,1) &= \pi(1,0) = p(1-p), \\
\pi(1,1) &= p^2.
\end{aligned}$$

In states $(0,1)$ and $(1,0)$, there is one boba in the straw; in state $(1,1)$, there are two boba in the straw. Therefore, the long-run average number of boba in the straw is

$$\pi(0,1) + \pi(1,0) + 2\pi(1,1) = 2p(1-p) + 2p^2 = 2p.$$

Alternate Solution: The goal of the question was to have you work through the balance equations, but there is a simple solution. Observe that at any given time after at least two seconds have passed, each component has probability p of being filled with boba. Therefore, the number of boba in the straw is like a binomial distribution with 2 independent trials and success probability p, so the average number of boba in the straw is $2p$.

(f) The long run probability that the amount of boba doesn't change is the probability that either (a) we are in state $(0,1)$ and transition to $(1,0)$, (b) we are in state $(1,0)$ and transition to $(0,1)$, (c) we are in state $(1,1)$ and transition to $(1,1)$, or (d) we are in state $(0,0)$ and transition to $(0,0)$. In the long run, the probability we are in a particular state is given by the stationary distribution, so we have

$$\begin{aligned}
\mathbb{P}[(0,0) \to (0,0)] &= \pi(0,0)\mathbb{P}[X_{n+1} = (0,0) \mid X_n = (0,0)] = (1-p)^3 \\
\mathbb{P}[(0,1) \to (1,0)] &= \pi(0,1)\mathbb{P}[X_{n+1} = (1,0) \mid X_n = (0,1)] = p(1-p)^2 \\
\mathbb{P}[(1,0) \to (0,1)] &= \pi(1,0)\mathbb{P}[X_{n+1} = (0,1) \mid X_n = (1,0)] = p^2(1-p) \\
\mathbb{P}[(1,1) \to (1,1)] &= \pi(1,1)\mathbb{P}[X_{n+1} = (1,1) \mid X_n = (1,1)] = p^3.
\end{aligned}$$

Thus, our overall probability is $\boxed{p^3 + p^2(1-p) + p(1-p)^2 + (1-p)^3}$.