Problem Sheet 11 [1]

Based on Chapters 9.5 and 10.1-10.3 the course book.

Optional Revision Problems

Exercise 1. While Fred is sleeping one night, X legitimate emails and Y spam emails are sent to him. Suppose that X and Y are independent, with $X \sim Pois(10)$ and $Y \sim Pois(40)$. When he wakes up, he observes that he has 30 new emails in his inbox. Given this information, what is the expected value of how many new legitimate emails he has?

Hint: You might find Theorem 4.8.1 and 4.8.2 from the course book useful.

Solution 1. By Theorem 4.8.1, the total number of new emails sent to Fred overnight is $(X+Y) \sim Pois(50)$. By Theorem 4.8.2, the conditional distribution of X given $X+Y=30$ is a Binomial distribution with parameters $n=30$ and $p=\frac{10}{10+40}=\frac{1}{5}$. The expectation of a $Binom(n,p)$ is np, hence the expected number of valid emails that Fred receive overnight is $\frac{1}{5}\cdot 30=6$. This agrees with our intuition, as if he receives spam emails with a rate 4 times as the rate of legitimate emails, then the spam and non-spam emails should reflect this same ratio (6 to 24).

You can also solve this exercise using Definition 9.1.1 and by calculating the conditional PMF, however, that would require more calculation than multiplying two numbers.

Exercise 2. Let $X_1, X_2, \dots$ be i.i.d. r.v.s with mean 0, and let $S_n = X_1 + \cdots + X_n$. As shown in Example 9.3.6 in the book, the expected value of the first term given the sum of the first n terms is

$$E(X_1|S_n) = \frac{S_n}{n}$$

Generalize this result by finding $E(S_k|S_n)$ for all positive integers k and n.

Note: There is no restrictions which one of k and n is the greater number, so you should consider all cases.

Solution 2. We have to consider three cases here, when $k<n$, $k=n$, and $k>n$, the first two are easier, as when $k<n$ can use the linearity of expectation and the result from the example to get

$$E(S_k|S_n) = E(X_1|S_n) + \dots E(X_k|S_n) = \frac{S_n}{n} + \cdots + \frac{S_n}{n} = \frac{k}{n}S_n.$$

If $k=n$, we can either take the same approach as above, or even easier, by "taking out what is known" $E(S_k|S_n) = E(S_n|S_n) = S_n$.

For $k>n$, we have to realize that the terms X_i, $i>n$ in the sum are independent from S_n as the X_i-s are independent. Thus it would make sense to split the sum S_k to two parts, the sum until

[1] Exercises are based on the coursebook Statistics 110: Probability by Joe Blitzstein

n, that does depend on S_n and the sum after n, that does not. Hence by linearity of expectation and by the previous result

$$\begin{aligned}E(S_k|S_n) &= E(S_n + X_{n+1} + X_{n+2} + \dots X_k|S_n)\\ &= E(S_n|S_n) + E(X_{n+1} + X_{n+2} + \dots X_k|S_n)\\ &= S_n + E(X_{n+1}|S_n) + E(X_{n+2}|S_n) + \dots E(X_k|S_n).\end{aligned}$$

Now by the independence stated above, we can simplify all the conditional expectations left to unconditional ones i.e. $E(X_{n+1}|S_n) = E(X_{n+1}), \dots E(X_k|S_n) = E(X_k)$, and since all the X_i-s have mean zero, the sum of these expectations is zero, thus

$$E(S_k|S_n) = S_n,$$

for $k > n$.

In summary

$$E(S_k|S_n) = \frac{min(k,n)}{n} S_n,$$

for any positive integers k and n.

Week 11 exercises

Exercise 3. Joe will read $N \sim Pois(\lambda)$ books next year. Each book has a $G \sim Pois(\mu)$ number of pages, with book lengths independent of each other and independent of N. Find the variance of the number of book pages Joe will read next year.

Hint: Remember from last week that $E(T|N) = N\mu$

Solution 3. By Eve's law (law of total variance), $Var(T) = E(Var(T|N)) + Var(E(T|N))$ To calculate the first half of the sum, observe that

$$Var(T|N) = Var\left(\sum_{i=1}^{N} G_i \middle| N\right) = \sum_{i=1}^{N} Var(G_i|N),$$

where we can do the last step due to the independence of the G_i-s and by the property "taking out what's known". Since G_i is independent of N, and the G_i-s are identically distributed, the right-hand side simplifies to $NVar(G)$. Given that $G \sim Pois(\mu)$, we have that $Var(T|N) = N\mu$. Finally, as $N \sim Pois(\lambda)$, $E(N\mu) = \lambda\mu$.

For the second half of the sum, we can use that last week we established, that $E(T|N) = N\mu$. Then

$$Var(E(T|N)) = Var(N\mu) = \mu^2 Var(N) = \mu^2\lambda.$$

Putting everything together, the total variation is

$$Var(T) = \mu\lambda + \mu^2\lambda = \lambda(\mu + \mu^2).$$

Had we used **wrongly** the relationship $T = N \cdot G$, we would have gotten the same expectation, but it would be still incorrect to say that. To see that T and $N \cdot G$ do not have the same distribution, calculate the variance of $N \cdot G$, and compare to $Var(T)$.

Exercise 4. Let X_1, X_2, and Y be random variables, such that Y has finite variance. Let

$$A = E(Y|X_1) \text{ and } B = E(Y|X_1, X_2).$$

Show that

$$Var(A) \leq Var(B)$$

Also, check that this make sense in the extreme cases where Y is independent of X_1 and where $Y = h(X_2)$ for some function h.

Hint: Use Eve's law on B, with conditioning on X_1 (Theorem 9.3.8).

Solution 4. From the Hint, we can rewrite the variance of B as

$$Var(B) = E(Var(B|X_1)) + Var(E(B|X_1)).$$

Writing out B in the second term we have $Var(E(E(Y|X_1, X_2)|X_1))$, and by using the Theorem mentioned in the hint, this is equal to $Var(E(Y|X_1))$, that is by definition the variance of A.

Since the first term in the decomposition of $Var(B)$ is the expectation of a variance, i.e. the expectation of something non-negative, it follows that

$$Var(B) = E(Var(B|X_1)) + Var(A) \geq Var(A).$$

To consider the extreme cases mentioned above, note that if $Y \perp\!\!\!\perp X_1$, then $A = E(Y|X_1) = E(Y) = c$ for some constant $c \in \mathbb{R}$, hence $Var(A) = 0$, so the inequality trivially holds. If however $Y = h(X_2)$, then by taking out what is known $E(Y|X_1, X_2) = E(h(X_2)|X_1, X_2) = h(X_2) = Y$, thus $Var(B) = Var(Y)$. Applying Eve's law once again, we have by definition

$$Var(B) = Var(Y) = E(Var(Y|X_1)) + Var(E(Y|X_1)) = E(Var(Y|X_1)) + Var(A).$$

Arguing again that the variance is non-negative, and the expectation of a non-negative random variable is non-negative, it follows that

$$Var(B) \geq Var(A).$$

Exercise 5. One of two identical-looking coins is picked from a hat randomly, where one coin has probability p_1 of Heads and the other has probability p_2 of Heads. Let X be the number of Heads after flipping the chosen coin n times. Find the mean and variance of X.

Solution 5. The distribution of X is a *mixture* of two Binomials; this is **not** Binomial unless $p_1 = p_2$. Let I be the indicator of having the p_1 coin. Then

$$E(X) = E(X|I = 1)P(I = 1) + E(X|I = 0)P(I = 0) = \frac{1}{2}n(p_1 + p_2),$$

as the expectation of $Binom(n, p)$ is np.

Alternatively, we can represent X as

$$X = IX_1 + (1 - I)X_2$$

with $X_j \sim \text{Bin}(n, p_j)$, and I, X_1, X_2 independent. Then by Adam's law and using "taking out what's known"

$$E(X) = E(E(X|I)) = E(Inp_1 + (1 - I)np_2) = \frac{1}{2}n(p_1 + p_2).$$

For the variance, note that it is **not** valid to say

$$\mathrm{Var}(X) = \mathrm{Var}(X|I=1)P(I=1) + \mathrm{Var}(X|I=0)P(I=0);$$

an extreme example of this mistake would be claiming that "$\mathrm{Var}(I) = 0$ since $\mathrm{Var}(I|I=1)P(I=1) + \mathrm{Var}(I|I=0)P(I=0) = 0$;" of course, $\mathrm{Var}(I) = \frac{1}{4}$. Instead, we can use Eve's Law:

$$\mathrm{Var}(X) = E(\mathrm{Var}(X|I)) + \mathrm{Var}(E(X|I)),$$

where by using again that X is a mixture of binomials, $\mathrm{Var}(X|I) = Inp_1(1-p_1)+(1-I)np_2(1-p_2)$, and $E(X|I) = Inp_1+(1-I)np_2$. So, by taking the expectation over $Inp_1(1-p_1)+(1-I)np_2(1-p_2)$ and the variance over $Inp_1 + (1-I)np_2$ we get

$$\mathrm{Var}(X) = \frac{1}{2}\big(np_1(1-p_1) + np_2(1-p_2)\big) + \frac{1}{4}n^2(p_1-p_2)^2,$$

where it was used that the $I \sim Bern(1/2)$.

Exercise 6. Let X and Y be i.i.d. positive r.v.s, and let $c > 0$. For each part below, fill in the appropriate equality or inequality symbol: write $=$ if the two sides are always equal, $\leq$ if the left-hand side is less than or equal to the right-hand side (but they are not necessarily equal), and similarly for $\geq$. If no relation holds in general, write ?.

1. $E(\ln(X))$ ____ $\ln(E(X))$
2. $E(X)$ ____ $\sqrt{E(X^2)}$
3. $P(X > c)$ ____ $\frac{E(X^3)}{c^3}$
4. $P(X \leq Y)$ ____ $P(X \geq Y)$
5. $E(XY)$ ____ $\sqrt{E(X^2)E(Y^2)}$
6. $P(X+Y > 10)$ ____ $P(X > 5 \text{ or } Y > 5)$
7. $E(\min(X,Y))$ ____ $\min(E(X), E(Y))$
8. $E(X/Y)$ ____ $E(X)/E(Y)$
9. $E(X^2(X^2+1))$ ____ $E(X^2(Y^2+1))$
10. $E\left(\frac{X^3}{X^3+Y^3}\right)$ ____ $E\left(\frac{Y^3}{X^3+Y^3}\right)$

Solution 6. 1. $E(\ln(X)) \leq \ln(E(X))$

- By **Jensen's inequality**: logarithms are concave.

2. $E(X) \leq \sqrt{E(X^2)}$

- Follows from $\mathrm{Var}(X) \geq 0$ or by **Jensen's inequality**, after squaring both sides.

3. $P(X > c) \leq \frac{E(X^3)}{c^3}$

- Using **Markov's inequality** after cubing both sides, i.e. $P(X > c) = P(X^3 > c^3) \leq \frac{E(X^3)}{c^3}$

4. $P(X \leq Y) = P(X \geq Y)$
   - By **symmetry**, since X and Y are i.i.d.
5. $E(XY) \leq \sqrt{E(X^2)E(Y^2)}$
   - By the **Cauchy-Schwarz inequality**.
6. $P(X + Y > 10) \leq P(X > 5 \text{ or } Y > 5)$
   - If $X + Y > 10$, then either $X > 5$ or $Y > 5$, however, ($X > 5$ or $Y > 5$) does not imply $X + Y > 10$, e.g. take $X = 2$, $Y = 6$.
7. $E(\min(X, Y)) \leq \min(E(X), E(Y))$
   - Since $\min(X, Y) \leq X$ gives $E(\min(X, Y)) \leq E(X)$, and similarly for Y.
8. $E(X/Y) \geq E(X)/E(Y)$
   - From $E(X/Y) = E(X)E(1/Y)$ that follows from independence, with $E(1/Y) \geq 1/E(Y)$ by **Jensen's inequality**.
9. $E(X^2(X^2 + 1)) \geq E(X^2(Y^2 + 1))$
   - Since $E(X^4) \geq (E(X^2))^2 = E(X^2)E(Y^2)$, and X^2 because Y^2 are i.i.d., and independent implies uncorrelated.
10. $E\left(\frac{X^3}{X^3+Y^3}\right) = E\left(\frac{Y^3}{X^3+Y^3}\right)$
    - By symmetry, as X and Y are i.i.d.

Exercise 7. For i.i.d. r.v.s $X_1, \ldots, X_n$ with mean μ and variance σ^2, give a value of n (as a specific number) that will ensure that there is at least a 99% chance that the sample mean, defined as $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ will be within 2 standard deviations of the true mean μ.

Solution 7. We have to find n such that

$$P(|\overline{X}_n - \mu| > 2\sigma) \leq 0.01.$$

By Chebyshev's inequality (in the form $P(|Y - E(Y)| > c) \leq \frac{\mathrm{Var}(Y)}{c^2}$), we have

$$P(|\overline{X}_n - \mu| > 2\sigma) \leq \frac{\mathrm{Var}\,\overline{X}_n}{(2\sigma)^2}.$$

Using the independence of the X_i-s we have

$$\mathrm{Var}(\overline{X}_n) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}(X_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

Hence

$$P(|\overline{X}_n - \mu| > 2\sigma) \leq \frac{\frac{\sigma^2}{n}}{4\sigma^2} = \frac{1}{4n}.$$

So the desired inequality holds if $n \geq 25$.

Exercise 8. In a national survey, a random sample of people are chosen and asked whether they support a certain policy. Assume that everyone in the population is equally likely to be surveyed at each step, and that the sampling is with replacement (sampling without replacement is typically more realistic, but with replacement will be a good approximation if the sample size is small compared to the population size). Let n be the sample size, and let $\hat{p}$ and p be the proportion of people who support the policy in the sample and in the entire population, respectively. Show that for every $c > 0$,

$$P(|\hat{p} - p| > c) \leq \frac{1}{4nc^2}.$$

Solution 8. Let X be the number of people who were asked in the survey, who would support the policy. You can think about X as the number of successes, if a success is selecting a person who supports the policy from the total population to be included in the survey. Since the proportion of "supporters" is p in the total population, the probability of success is p, thus $X \sim \text{Bin}(n, p)$, where we also rely on the fact that we sample with replacement. We can write $\hat{p} = X/n$. So $E(\hat{p}) = p$, and $\text{Var}(\hat{p}) = p(1-p)/n$. Then by Chebyshev's inequality,

$$P(|\hat{p} - p| > c) \leq \frac{\text{Var}(\hat{p})}{c^2} = \frac{p(1-p)}{nc^2} \leq \frac{1}{4nc^2},$$

where the last inequality holds because $p(1-p)$ is maximized at $p = 1/2$.