Problemset 11

Albert Ye

November 11, 2022

1 Balls in Bins

Let the indicator variable I_k be 1 if bin k has exactly 1 ball inside. This gives us $\mathbb{E}[X] = \mathbb{E}[I_1] + \mathbb{E}[I_2] + \cdots + \mathbb{E}[I_m]$. We know that $\mathbb{E}[I_k] = \mathbb{P}[I_k] = \dfrac{(n-1)^{m-1} \cdot n}{n^m}$, as of the n^m total ways to put n balls in m bins, $n(n-1)^{m-1}$ of them have exactly one ball in the kth bin. There are $n-1$ balls remaining, $m-1$ slots remaining, and n ways to choose the ball that is put in the kth bin.

Therefore, $\mathbb{E}[x] = \sum_{k=0}^{n} \mathbb{E}[I_k] = n\left(\dfrac{n-1}{n}\right)^{m-1} = \dfrac{(n-1)^{m-1}}{n^{m-2}}$.

The formula for $\mathrm{Var}(x)$ is $\mathbb{E}[x^2] - \mathbb{E}[x]^2$, so we need to find $\mathbb{E}[x^2]$ as well. We find that this equals

$$\begin{aligned}
\mathbb{E}[x^2] &= \sum_{k=1}^{n} \mathbb{E}[I_k^2] + 2 \sum_{1 \le k_1 < k_2 \le n} \mathbb{E}[I_{k_1} I_{k_2}] \\
&= \sum_{k=1}^{n} \mathbb{E}[I_k] + 2 \sum_{1 \le k_1 < k_2 \le n} \mathbb{E}[I_{k_1} I_{k_2}] \\
&= \mathbb{E}[x] = 2 \sum_{1 \le k_1 < \le n} \mathbb{E}[I_{k_1} I_{k_2}].
\end{aligned}$$

We can evaluate $2\sum_{1 \le k_1 < k_2 \le n} \mathbb{E}[I_{k_1} I_{k_2}]$ through linearity of expectation on $\mathbb{E}[I_{k_1} I_{k_2}] = \mathbb{P}[I_{k_1} = 1 \wedge I_{k_2} = 1]$. For a given k_1, k_2, there are $n(n-1)$ ways to choose the two balls to put in bins k_1, k_2, and $(n-2)^{m-2}$ ways to arrange the remaining balls out of a total of n^m arrangements. Therefore, the probability equals $\dfrac{(n-2)^{m-2} \cdot n(n-1)}{n^m} = \dfrac{(n-1)(n-2)^{m-2}}{n^{m-1}}$. Multiplying this by the number of possible k_1, k_2, we get $2\sum_{1 \le k_1 < k_2 \le n} \mathbb{E}[I_{k_1} I_{k_2}] = 2\left(\dfrac{n(n-1)}{2}\right)\left(\dfrac{(n-2)^{m-2}(n-1)}{n^{m-1}}\right)$.

Hence,

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

and

$$\mathbb{E}[x]^2 = \frac{(n-1)^{2m-2}}{n^{2m-4}}.$$

Therefore,

$$\mathrm{Var}(x) = \mathbb{E}[x^2] - \mathbb{E}[x]^2 = \boxed{\frac{n^{m-2}(n-1)^m + n^{m-2}(n-1)^2(n-2)^{m-2} - (n-1)^{2m-2}}{n^{m-2}}}.$$

2 Will I Get My Package?

(a) Let X_i be a variable that equals 0 if person i gets their package, and Y_i be a variable that equals 0 if person i's package is opened. We see that X_i and Y_i are completely independent because whether or not person i gets their package does not influence whether or not the mailman is going to open the package. Therefore, we see that $X_i \cap Y_i = X_i Y_i$, and $\mathbb{E}[X] = \sum_{i=0}^{n} \mathbb{E}[X_i]\mathbb{E}[Y_i]$.

Because $\mathbb{E}[Y_i] = \frac{1}{2}$ for all i, we can ignore it for now and only look for $\mathbb{E}[X_i]$. For a given index i, there are $(n-1)!$ permutations that give person i the correct package out of $n!$ total, so $\mathbb{E}[X_i] = \frac{(n-1)!}{n!} = \frac{1}{n}$.

Plugging this back to our equation for $\mathbb{E}[X]$, we get $\mathbb{E}[X] = \sum_{i=0}^{n} \mathbb{E}[X_i]\mathbb{E}[Y_i] = \frac{1}{2}\mathbb{E}[X] \sum_{i=0}^{n} \mathbb{E}[X_i]\mathbb{E}[Y_i] = \sum_{i=0}^{n} \frac{1}{2n} = \boxed{\frac{1}{2}}$.

(b) Since $\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$, we need to find $\mathbb{E}[X^2]$.

This is equal to $(\mathbb{E}[X_1] + \mathbb{E}[X_2] + \cdots + \mathbb{E}[X_n])^2 = \sum_{i=1}^{n} \mathbb{E}[X_i^2] + 2\sum_{1 \le i < j \le n} \mathbb{E}[X_i X_j]$. We find that $\mathbb{E}[X_i X_j] = \mathbb{P}[X_i = 1 \wedge X_j = 1]$ To compute the number of ways to satisfy $X_i = 1 \wedge X_j = 1$, we give person i and person j their respective packages (with a $\frac{1}{4}$ chance that neither are open), and rearrange the other $(n-2)!$ packages randomly. This leads to a total of $(n-2)!$ ways out of $n!$ total, so $\mathbb{P}[X_i = 1 \wedge X_j = 1] = \frac{(n-2)!}{4n!} = \frac{1}{4n(n-1)}$.

Therefore, the second term evaluates to $2\sum_{1 \le i < j \le n} \mathbb{E}[X_i X_j] = 2\sum_{1 \le i < j \le n} \frac{1}{n(n-1)} = 2 \cdot \frac{n(n-1)}{2} \cdot \frac{1}{4n(n-1)} = \frac{1}{4}$. The first term evaluates to $\sum_{i=1}^{n} \mathbb{E}[X_i^2] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \frac{1}{2}$. Hence, $\mathbb{E}[X^2] = \frac{3}{4}$.

Finally, we just need to plug this into the variance formula to get $\mathrm{Var}(X) = \frac{3}{4} - \frac{1}{4} = \boxed{\frac{1}{2}}$.

3 Double-Check Your Intuition Again

(a) (i) From the covariance formula, $\text{Cov}(X+Y, X-Y) = \mathbb{E}[(X+Y)(X-Y)] - \mathbb{E}[X+Y]\mathbb{E}[X-Y]$. The second term can be calculated pretty direction: from linearity of expectation, $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y] = \frac{7}{2} + \frac{7}{2} = 7$ and $\mathbb{E}[X-Y] = \mathbb{E}[X] - \mathbb{E}[Y] = \frac{7}{2} - \frac{7}{2} = 0$, so $\mathbb{E}[X+Y]\mathbb{E}[X-Y] = 0$. The first term can be found by expanding $\mathbb{E}[(X+Y)(X-Y)] = \mathbb{E}[X^2 - Y^2]$. From linearity of expectation this equals $\mathbb{E}[X^2] - \mathbb{E}[Y^2] = 0$. Therefore, the final expected value is $\boxed{0}$.

(ii) Consider the case of $X + Y = 12$. Then, our probability distribution for $X - Y$ equals $\mathbb{P}[X - Y = 0] = 1$ and $\mathbb{P}[X - Y = i] = 0$ for remaining i. However, $\mathbb{P}[X - Y = 0]$ is not necessarily 1: examine the case of $X + Y = 11$; $\mathbb{P}[X - Y = 0] = 0$ because an odd $X + Y$ implies that $X \neq Y$. Therefore, the value of $X - Y$ depends on $X + Y$. □

(b) Yes. Assume for the sake of contradiction that there is a distribution with unequal value with variance of 0. The variance is the sum of squares, meaning that it must be nonnegative; the only way for the variance to equal 0 is for all the terms to be equal to 0, so all $x - \overline{x}$ must be 0. If we have another distinct value $y \neq x$, then $y - \overline{x} \neq 0$, which is a contradiction. □

(c) This is false. The formula for variance is $\sum_{x \in X}(x - \overline{x})^2$, so $\text{Var}(cx)$ is equivalent to $\sum_{x\in X}(cx - \overline{c}x)^2 = c^2 \sum_{x \in X}(x - \overline{x})^2 = c^2\text{Var}(x)$. □

(d) No, part (a) is a good example of this.

(e) $\text{Corr}(X, Y) = 0$ implies that $\text{Cov}(X, Y) = 0$, so

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y] \tag{1}$$

. Therefore,

$$\begin{aligned}
\text{Var}(X+Y) - (\text{Var}(X) + \text{Var}(Y)) &= \mathbb{E}[(X+Y)^2] - \mathbb{E}[X+Y]^2 - (\mathbb{E}[X^2] - \mathbb{E}[X]^2) - (\mathbb{E}[Y^2] - \mathbb{E}[Y]^2) \\
&= \mathbb{E}[X^2 + Y^2 + 2XY] - \mathbb{E}[X+Y]^2 - (\mathbb{E}[X^2] - \mathbb{E}[X]^2) - (\mathbb{E}[Y^2] - \mathbb{E}[Y]^2) \\
&= \mathbb{E}[X^2 + Y^2 + 2XY] - (\mathbb{E}[X] + \mathbb{E}[Y])^2 - (\mathbb{E}[X^2] - \mathbb{E}[X]^2) - (\mathbb{E}[Y^2] - \mathbb{E}[Y]^2) \\
&= \mathbb{E}[X^2] + \mathbb{E}[Y^2] + 2\mathbb{E}[XY] - \mathbb{E}[X]^2 - \mathbb{E}[Y]^2 - 2\mathbb{E}[X]\mathbb{E}[Y] \\
&\quad - \mathbb{E}[X^2] + \mathbb{E}[X]^2 - \mathbb{E}[Y^2] + \mathbb{E}[Y]^2. \\
&= 2\mathbb{E}[XY] - 2\mathbb{E}[X]\mathbb{E}[Y].
\end{aligned}$$

From (1), we see that this must equal 0, so $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$. □

(f) Note that $\max(X, Y)$ and $\min(X, Y)$ are both going to take on either X or Y, and both values are going to be considered. Therefore, $\max(X, Y)\min(X, Y) = XY$ so $\mathbb{E}[\max(X, Y)\min(X, Y)] = \mathbb{E}[XY]$. □

(g) Note that

$$\text{Corr}(\max(X, Y), \min(X, Y)) = \frac{\text{Cov}(\max(X, Y), \min(X, Y))}{\sigma(\max(X, Y))\sigma(\min(X, Y))}.$$

For the numerator, $\text{Cov}(\max(X, Y), \min(X, Y)) = \text{Cov}(X, Y)$ because the order of the terms does not matter, and one of $\max(X, Y), \min(X, Y)$ will equal X and the other will equal Y. For the same reason, the denominator $\sigma(\max(X, Y))\sigma(\min(X, Y)) = \sigma(X)\sigma(Y)$. Therefore,

$$\text{Corr}(\max(X, Y), \min(X, Y)) = \frac{\text{Cov}(X, Y)}{\sigma(X)\sigma(Y)} = \text{Corr}(X, Y).$$

□

4 Fishy On Me

(a) The distribution is seen to be poisson(20). Therefore, the probability $\mathbb{P}[X = 7] = \frac{20^7}{7!}e^{-20} = \boxed{\frac{20^7}{7!e^{-20}}}$.

(b) The distribution is poisson(2), so our probability is $\mathbb{P}[X \leq 1] = \frac{2^0}{0!}e^{-2} + \frac{2^1}{1!}e^{-2} = \boxed{\frac{3}{e^2}}$.

(c) The distribution is poisson(11.4), so we want to find $\mathbb{P}[X \geq 3]$, letting X be the number of boats in these two days. Note that the probability of $X \geq 3$ is the complement of $\mathbb{P}[X < 3]$, which is much easier to calculate. This sum is

$$\begin{aligned}
\mathbb{P}[X \geq 3] &= 1 - \mathbb{P}[X < 3] \\
&= 1 - \mathbb{P}[X = 0] - \mathbb{P}[X = 1] - \mathbb{P}[X = 2] \\
&= 1 - \frac{11.4^0}{0!}e^{-11.4} - \frac{11.4^1}{1!}e^{-11.4} - \frac{11.4^2}{2!}e^{-11.4} \\
&= 1 - e^{-11.4}(1 + 11.4 + 64.98) \\
&= \boxed{1 - 77.38e^{-11.4}}.
\end{aligned}$$

5 Geometric and Poisson

Using the Total Probability Rule, $\mathbb{P}[X > Y] = \sum_{i=1}^{\infty} \mathbb{P}[Y = i \cap X > i]$. As X and Y are independent, this is equal to

$$\begin{aligned}
\mathbb{P}[X > Y] &= \sum_{i=0}^{\infty} \mathbb{P}[Y = i]\mathbb{P}[X > i] = \sum_{i=1}^{\infty} \frac{\lambda^i}{i!} e^{-\lambda} \cdot (1-p)^i \\
&= \sum_{i=0}^{\infty} \frac{(\lambda - \lambda p)^i}{i!} e^{-\lambda}
\end{aligned}$$

We can now create another Poisson distribution poisson$(\lambda - \lambda p)$.

$$\begin{aligned}
\sum_{i=0}^{\infty} \frac{(\lambda - \lambda p)^i}{i!} e^{-\lambda} &= e^{-\lambda p} \sum_{i=0}^{\infty} \frac{(\lambda - \lambda p)^i}{i!} e^{-\lambda(\lambda p)} \\
&= \boxed{e^{-\lambda p}}.
\end{aligned}$$

6 Poisson Coupling

(a) Let's consider the case of $|\mathbb{P}[X = k] - \mathbb{P}[Y = k]|$. From the Total Probability Rule, we have $\mathbb{P}[X = k] = \mathbb{P}[X = k, Y = k] + \mathbb{P}[X = k, Y \neq k]$ and $\mathbb{P}[Y = k] = \mathbb{P}[Y = k, X = k] + \mathbb{P}[Y = k, X \neq k]$. Therefore,

$$\begin{aligned}|\mathbb{P}[X = k] - \mathbb{P}[Y = k]| &= |\mathbb{P}[X = k, Y = k] + \mathbb{P}[X = k, Y \neq k] - \mathbb{P}[Y = k, X = k] - \mathbb{P}[Y = k, X \neq k]| \\ &= |\mathbb{P}[X = k, Y \neq K] - \mathbb{P}[Y = k, X \neq K]| \\ &\leq \mathbb{P}[X = k, Y \neq K] + \mathbb{P}[Y = k, X \neq K].\end{aligned}$$

Taking the sum of this and then dividing by 2, we get

$$\begin{aligned}\frac{1}{2}\sum_{k=0}^{\infty} |\mathbb{P}[X = k] - \mathbb{P}[Y = k]| &\leq \frac{1}{2}\sum_{k=0}^{\infty} \mathbb{P}[X = k, Y \neq k] + \mathbb{P}[Y = k, X \neq k] \\ &= \frac{1}{2}\mathbb{P}[X \neq Y] \leq \mathbb{P}[X \neq Y].\end{aligned}$$

$\square$

(b) Using the union bound, we have that

$$\mathbb{P}\left[\bigcup_{i=1}^{n} X_i \neq Y_i\right] \leq \sum_{i=1}^{n} \mathbb{P}[X_i \neq Y_i].$$

Thus, all we have left to prove is

$$\mathbb{P}\left[\sum_{i=1}^{N} X_i \neq \sum_{i=1}^{N} Y_i\right] \leq \mathbb{P}\left[\bigcup_{i=1}^{n} X_i \neq Y_i\right].$$

If $\sum_{i=1}^{N} X_i \neq \sum_{i=1}^{N} Y_i$, then for some i, $X_i \neq Y_i$. However, the converse is not always true; if $X_i \neq Y_i$ then it is not always true that $\sum_{i=1}^{N} X_i \neq \sum_{i=1}^{N} Y_i$. For example, consider the case where $N = 2$, and $X_1 = X_2 = 1$ while $Y_0 = 0, Y_1 = 2$. Obviously $X_i \neq Y_i$ for some i, but the sum of all X_i equals the sum of all Y_i.

Therefore,

$$\begin{aligned}&(\sum_{i=0}^{n} X_i \neq \sum_{i=0}^{n} Y_i) \subseteq (\bigcup_{i=0}^{n} X_i \neq Y_i) \\ \implies &\mathbb{P}[\sum_{i=0}^{n} X_i \neq \sum_{i=0}^{n} Y_i] \leq \bigcup_{i=0}^{n} X_i \neq Y_i.\end{aligned}$$

$\square$

(c) We must check that all probabilities are between 0 and 1, and that the probabilities sum to 1.

Since the probabilities $\mathbb{P}[X_i = 0, Y_i = 0]$ and $\mathbb{P}[X_i = 1, Y_i = 0]$ are singular cases, we can just add them together. The first event has probability $1 - p_i \in [0, 1]$. The second has a probability of $e^{-p_i} - (1 - p_i) = \frac{1}{e^{p_i}} + p_i - 1$, which is less clearly between 0 and 1. Because $\frac{d}{dp_i}\frac{1}{e^{p_i}} + p_i - 1 = 1 - e^{-p_i}$, we know that $\frac{1}{e^{p_i}} + p_i - 1$ is monotonically increasing. As $\frac{1}{e^0} + 0 - 1 = 0$ and $\frac{1}{e^1} + 1 - 1 = \frac{1}{e}$, we know that all values of $\frac{1}{e^{p_i}} + p_i - 1$ are between 0 and $\frac{1}{e}$ for all $p_i \in [0, 1]$. Therefore, the first criterion of a valid distribution is satisfied.

The sum of probabilities is

$$1 - p_i + \frac{1}{e^{p_i}} + p_i - 1 + \sum_{y=1}^{\infty} \frac{e^{-p_i} p_i^y}{y!} = \frac{1}{e^{p_i}} + \sum_{y=1}^{\infty} \frac{e^{-p_i} p_i^y}{y!} = \sum_{y=0}^{\infty} \frac{e^{-p_i} p_i^y}{y!},$$

which is essentially the sum of an entire Poisson distribution. This must equal 1 by definition. $\square$

(d) $\mathbb{P}[X_i = 0, Y_i = 0] = 1 - p_i$ is the only equation ih tye system that includes the event of $X_i = 0$, so $\mathbb{P}[X_i = 0] = 1 - p_i$. Therefore, as $\mathbb{P}[X_i \geq 2] = 0$, the remaining p must belong to the case of $\mathbb{P}[X_i = 1]$. Therefore, X_i has the Bernoulli distribution with probability p_i. Thus, this joint distribution must be valid. $\square$

(e) For cases of $Y_i = 0$, we have the two cases of $X_i = 0, X_i = 1$. The first case has probabiltiy $1 - p_i$ and the second case has probability $e^{-p_i} - (1 - p_i)$, so the sum of these two cases is just $\mathbb{P}[Y_i = 0] = e^{-p_i}$. This is just the value of the Poisson distribution at 0.

For cases of $Y_i \neq 0$, the former case is equal to 0 so we only have to contend with $X_i = 1$. For each y, the value $\mathbb{P}[Y_i = y] = e^{-p_i}\dfrac{p_i^y}{y!}$, which equals the value of the Poisson distribution at y. $\square$

(f) We once again only need to consider two cases $X_i = 0, 1$. For the case of $X_i = 0$, the probability $\mathbb{P}[X_i \neq Y_i] = 0$ because $\mathbb{P}[X_i = 0, Y_i \neq 0] = 0$. And for the case of $X_i = 1$, we have

$$\begin{aligned}
\mathbb{P}[X_i \neq Y_i] &= \mathbb{P}[X_i = 1] - \mathbb{P}[X_i = 1, Y_i = 1] \\
&= \sum_{y=0}^{\infty} \frac{e^{-p_i} p_i^y}{y!} - (p_i - 1) - p_i e^{-p_i} \\
&= 1 - (1 - p_i) - p_i e^{-p_i} \\
&= p_i - p_i e^{-p_i} \\
&= p_i(1 - e^{-p_i}).
\end{aligned}$$

To prove that this probability is at most p_i^2, we need $1 - e^{-p_i} \leq p_i \implies e^{-p_i} + p_i - 1 \geq 0$, which we established in part (c). Therefore, $\mathbb{P}[X_i \neq Y_i] \leq p_i^2$. $\square$

(g) Let's look at individual terms of $d(X_i, Y_i)$.

$$\begin{aligned}
d(X_i, Y_i) &= \frac{1}{2}(\mathbb{P}[X_i = 0] + \mathbb{P}[X_i = 1] - \mathbb{P}[Y_i = 0] - \mathbb{P}[Y_i = 1] + \mathbb{P}[Y_i \geq 2]) \\
&= \frac{1}{2}(p_i + (1 - p_i) + 1 - 2(\mathbb{P}[Y_i = 0] + \mathbb{P}[Y_i = 1])) \\
&= 1 - \mathbb{P}[Y_i = 0] - \mathbb{P}[Y_i = 1] \\
&= 1 - \frac{e^{-p_i} p_i^0}{0!} - \frac{e^{-p_i} p_i^1}{1!} \\
&= 1 - e^{-p_i} - e^{-p_i} p_i \\
&= 1 - e^{-p_i}(1 + p_i).
\end{aligned}$$

We claim the difference $\mathbb{P}[X_i \neq Y_i] - d(X_i, Y_i) \geq 0$. Plugging in our results for the two values gives us $p_i(1 - e^{-p_i}) - (1 - e^{-p_i}(1 + p_i)) = p_i - p_i e^{-p_i} - 1 + e^{-p_i} + p_i e^{-p_i} = p_i - 1 + e^{-p_i}$. Again, from part (c), we get that this is nonnegative so $d(X_i, Y_i) \leq \mathbb{P}[X_i \neq Y_i]$.

Using this information, we have that $d(\sum_{i=1}^n X_i, \sum_{i=1}^n Y_i) \leq \sum_{i=1}^n \mathbb{P}[X_i \neq Y_i] \leq \sum_{i=1}^n p_i^2$ as desired. $\square$