Discussion 6A

CS 70, Summer 2024

1 Balls and Bins

(a) By definition, $\Pr[X_i = 1] = \binom{n}{1} \cdot \left(\frac{1}{m}\right)^1 \left(1 - \frac{1}{m}\right)^{n-1} = \frac{n}{m}\left(1 - \frac{1}{m}\right)^{n-1}$.

(b) As $\mathrm{E}[X_i] = \frac{n}{m}\left(1 - \frac{1}{m}\right)^{n-1}$ and $X = X_1 + X_2 + \cdots + X_m$, by linearity of expectation, we get $\mathrm{E}[X] = \mathrm{E}[X_1] + \mathrm{E}[X_2] + \cdots + \mathrm{E}[X_m] = n \cdot \left(1 - \frac{1}{m}\right)^{n-1}$.

(c) For $i \neq j$, we can notice that $X_i X_j$ can only take on two values: 0 and 1. This means that $\mathrm{E}[X_i X_j] = \Pr[X_i X_j = 1] = \Pr[X_i = 1, X_j = 1]$, or the probability that exactly 1 ball falls into bin i and exactly one ball falls into bin j. This turns out to be

$$\begin{aligned}
\mathrm{E}[X_i X_j] &= \Pr[X_i X_j = 1] \\
&= \underbrace{\binom{n}{1}}_{\binom{\text{choose ball}}{\text{for bin } i}} \underbrace{\binom{n-1}{1}}_{\binom{\text{choose ball}}{\text{for bin } j}} \underbrace{\left(\frac{1}{m}\right)^1}_{\binom{\text{chosen ball}}{\text{goes in bin } i}} \underbrace{\left(\frac{1}{m}\right)^1}_{\binom{\text{chosen ball}}{\text{goes in bin } j}} \underbrace{\left(1 - \frac{2}{m}\right)^{n-2}}_{\binom{\text{other balls}}{\text{not in bins } i \text{ or } j}} \\
&= \frac{n(n-1)}{m^2}\left(1 - \frac{2}{m}\right)^{n-2}.
\end{aligned}$$

When $i = j$, $\mathrm{E}[X_i^2] = \mathrm{E}[X_i]$ because X_i is the indicator random variable which takes value 0/1.

(d) Since $X = \sum_i X_i$, we can use the computational formula for variance:

$$\begin{aligned}
\mathrm{Var}(X) &= \mathrm{E}[X^2] - \mathrm{E}[X]^2 \\
&= \mathrm{E}\left[\left(\sum_{i=1}^m X_i\right)^2\right] - \left(\mathrm{E}[\sum_{i=1}^m X_i]\right)^2 \\
&= \mathrm{E}\left[\sum_{i \neq j} X_i X_j + \sum_{i=1}^m X_i^2\right] - \left(\sum_{i=1}^m \mathrm{E}[X_i]\right)^2 \\
&= \sum_{i \neq j} \mathrm{E}[X_i X_j] + \sum_{i=1}^m \mathrm{E}[X_i] - \left(\sum_{i=1}^m \mathrm{E}[X_i]\right)^2,
\end{aligned}$$

where the last line followed from linearity of expectation and recognizing that $X_i^2 = X_i$, since it can only take on the values 0 or 1. Noting that $\sum_{i \neq j}$ has $m(m-1)$ terms, and the rest of the sums have m terms, we find

$$\mathrm{Var}(X) = m(m-1) \cdot \frac{n(n-1)}{m^2}\left(1 - \frac{2}{m}\right)^{n-2} + m \cdot \frac{n}{m}\left(1 - \frac{1}{m}\right)^{n-1} - m^2\left[\frac{n}{m}\left(1 - \frac{1}{m}\right)^{n-1}\right]^2.$$

2 Correlation and Independence

(a) Recall that for two random variables X and Y,

$$\mathrm{Cov}(X, Y) = \mathrm{E}[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])] = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y).$$

Two random variables are uncorrelated iff their covariance is equal to zero. If X and Y are uncorrelated, then there is no linear relationship between them.

(b) Recall that two random variables X and Y are independent if and only if the following criteria are met (the three criteria are equivalent and connected by Bayes rule):

$$\Pr(X = x \mid Y = y) = \Pr(X = x)$$

$$\Pr(Y = y \mid X = x) = \Pr(Y = y)$$

$$\Pr(X = x, Y = y) = \Pr(X = x)\Pr(Y = y)$$

for all x, y such that $\Pr(X = x), \Pr(Y = y) > 0$.

If X and Y are independent, any information about one variable offers no information whatsoever about the other variable.

(c) Note that if two random variables are independent, they must have no relationship whatsoever, including linear relationships; therefore they must be uncorrelated. The converse, however, is not true: two uncorrelated variables may not be independent. Consider two variables X and Y that follow a uniform joint distribution over the points $(1, 0), (0, 1), (-1, 0), (0, -1)$. See Figure 1. Then

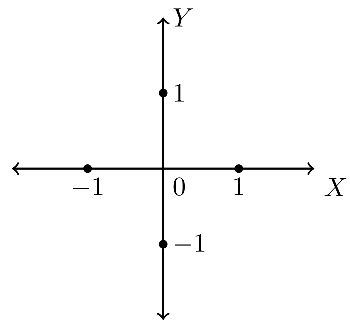

Figure 1: Choose one of the four points shown uniformly at random.

$$\text{Cov}(X, Y) = \text{E}(XY) - \text{E}(X)\text{E}(Y) = 0.$$

To see why, observe that $XY = 0$ always because at least one of X and Y is always 0, and furthermore $\text{E}[X] = \text{E}[Y] = 0$ because both X and Y are symmetric around 0. So, there is no linear relationship, but X and Y are not independent (for example, $\Pr(Y = 0) = 1/2$ but $\Pr(Y = 0 \mid X = 1) = 1$).

3 Diverse Hand

(a) Let X_i be the indicator of the ith value appearing in your hand. Then, $X = X_1 + X_2 + ... + X_{13}$. (Here we let 13 correspond to K, 12 correspond to Q, and 11 correspond to J.) By linearity of expectation, $\text{E}[X] = \sum_{i=1}^{13} \text{E}[X_i]$.

We can calculate $\Pr[X_i = 1]$ by taking the complement, $1 - \Pr[X_i = 0]$, or 1 minus the probability that the card does not appear in your hand. This is $1 - \frac{\binom{48}{5}}{\binom{52}{5}}$.

Then, $\text{E}[X] = 13\Pr[X_1 = 1] = 13\left(1 - \frac{\binom{48}{5}}{\binom{52}{5}}\right)$.

(b) To calculate variance, since the indicators are not independent, we have to use the formula $\text{E}[X^2] = \sum_{i=j} \text{E}[X_i^2] + \sum_{i \neq j} \text{E}[X_i X_j]$.

First, we have

$$\sum_{i=j} \text{E}[X_i^2] = \sum_{i=j} \text{E}[X_i] = 13\left(1 - \frac{\binom{48}{5}}{\binom{52}{5}}\right).$$

Next, we tackle $\sum_{i \neq j} \text{E}[X_i X_j]$. Note that $\text{E}[X_i X_j] = \Pr[X_i X_j = 1]$, as $X_i X_j$ is either 0 or 1.

To calculate $\Pr[X_i X_j = 1]$ (the probability we have both cards in our hand), we note that $\Pr[X_i X_j = 1] = 1 - \Pr[X_i = 0] - \Pr[X_j = 0] + \Pr[X_i = 0, X_j = 0]$. Then

$$\begin{aligned}
\sum_{i \neq j} \text{E}[X_i X_j] &= 13 \cdot 12 \Pr[X_i X_j = 1] \\
&= 13 \cdot 12(1 - \Pr[X_i = 0] - \Pr[X_j = 0] + \Pr[X_i = 0, X_j = 0]) \\
&= 156\left(1 - 2\frac{\binom{48}{5}}{\binom{52}{5}} + \frac{\binom{44}{5}}{\binom{52}{5}}\right)
\end{aligned}$$

Putting it all together, we have

$$\begin{aligned}\mathrm{Var}(X) &= \mathrm{E}[X^2] - \mathrm{E}[X]^2 \\ &= 13\left(1 - \frac{\binom{48}{5}}{\binom{52}{5}}\right) + 156\left(1 - 2\frac{\binom{48}{5}}{\binom{52}{5}} + \frac{\binom{44}{5}}{\binom{52}{5}}\right) - \left(13\left(1 - \frac{\binom{48}{5}}{\binom{52}{5}}\right)\right)^2.\end{aligned}$$

4 Covariance

(a) We can use the formula $\mathrm{Cov}(X_1, X_2) = \mathrm{E}[X_1X_2] - \mathrm{E}[X_1]\mathrm{E}[X_2]$.

$$\begin{aligned}\mathrm{E}[X_1] &= \frac{5}{10} \times 1 + \frac{5}{10} \times 0 = \frac{1}{2}, \\ \mathrm{E}[X_2] &= \frac{5}{10} \times 1 + \frac{5}{10} \times 0 = \frac{1}{2}, \\ \mathrm{E}[X_1X_2] &= \frac{5}{10} \cdot \frac{4}{9} \times 1 + \left(1 - \frac{5}{10} \cdot \frac{4}{9}\right) \times 0 = \frac{2}{9}.\end{aligned}$$

Therefore,

$$\mathrm{Cov}(X_1, X_2) = \mathrm{E}[X_1X_2] - \mathrm{E}[X_1]\mathrm{E}[X_2] = \frac{2}{9} - \frac{1}{2} \times \frac{1}{2} = -\frac{1}{36}.$$

(b) Again, we use the formula $\mathrm{Cov}(X_1, X_2) = \mathbb{E}[X_1X_2] - \mathbb{E}[X_1]\mathbb{E}[X_2]$.

$$\begin{aligned}\mathbb{E}[X_1] &= \frac{5}{10} \times 1 + \frac{5}{10} \times 0 = \frac{1}{2} \\ \mathbb{E}[X_2] &= \left(\frac{5}{10} \times \frac{6}{11} + \frac{5}{10} \times \frac{5}{11}\right) \times 1 + \left(\frac{5}{10} \times \frac{5}{11} + \frac{5}{10} \times \frac{6}{11}\right) \times 0 = \frac{1}{2} \\ \mathbb{E}[X_1X_2] &= \frac{5}{10} \times \frac{6}{11} \times 1 = \frac{30}{110}.\end{aligned}$$

Therefore,

$$\mathbb{E}[X_1X_2] - \mathbb{E}[X_1]\mathbb{E}[X_2] = \frac{30}{110} - \frac{1}{4} = \frac{1}{44}.$$

Note that in part (a), if one event happened, the other would be less likely to happen, and thus the covariance was negative. Similarly, in part (b), if one event happened, the other would be more likely to happen, and thus the covariance was positive.