1 Binomial Variance

Throw n balls into m bins uniformly at random. For a specific ball i, what is the variance of the number of roommates it has (i.e. the number of other balls that it shares its bin with)?

Solution:

When we concentrate on the bin that ball i is in, we care about how many of the other $n-1$ balls land in that same bin. Notice that when determining whether the balls land in this bin, each ball is independent of every other ball, which makes variance much easier to compute as we can take advantage of linearity of variance. This is a binomial distribution, with $n-1$ trials and probability $1/m$ of success for each trial, where success is defined as having the ball land in the same bin as ball i.

Therefore, the variance is

$$(n-1)\left(\frac{1}{m}\right)\left(1-\frac{1}{m}\right).$$

2 Will I Get My Package?

Sneaky delivery guy of some company is out delivering n packages to n customers. Not only does he hand a random package to each customer, 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 $\mathbf{E}(X)$.

(b) Compute the variance $\mathrm{Var}(X)$.

Solution:

(a) Define

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

By linearity of expectation, $\mathbf{E}(X) = \mathbf{E}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathbf{E}(X_i)$. We have

$$\mathbf{E}(X_i) = \Pr[X_i = 1] = \frac{1}{2n},$$

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,

$$\mathbf{E}(X) = n \cdot \frac{1}{2n} = \boxed{\frac{1}{2}}.$$

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

$$\mathbf{E}(X^2) = \mathbf{E}\big((X_1 + X_2 + \ldots + X_n)^2\big) = \mathbf{E}\left(\sum_{i,j} X_i X_j\right) = \sum_{i,j} \mathbf{E}(X_i X_j).$$

Then we consider two cases, either $i = j$ or $i \neq j$.
Hence $\sum_{i,j} \mathbf{E}(X_i X_j) = \sum_i \mathbf{E}(X_i^2) + \sum_{i \neq j} \mathbf{E}(X_i X_j)$.

$$\mathbf{E}(X_i^2) = \mathbf{E}(X_i) = \frac{1}{2n}$$

for all i. To find $\mathbf{E}(X_i X_j)$, we need to calculate $\Pr[X_i X_j = 1]$.

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

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

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

$$\text{Var}(X) = \mathbf{E}(X^2) - \mathbf{E}(X)^2 = \frac{3}{4} - \frac{1}{4} = \boxed{\frac{1}{2}}.$$

3 Variance

This problem will give you practice using the "standard method" to compute the variance of a sum of random variables that are not pairwise independent (so you cannot use "linearity" of variance).

(a) A building has n floors numbered $1, 2, \ldots, n$, plus a ground floor G. At the ground floor, m people get on the elevator together, and each person gets off at one of the n floors uniformly at random (independently of everybody else). What is the *variance* of the number of floors the elevator *does not* stop at? (In fact, the variance of the number of floors the elevator *does* stop at must be the same, but the former is a little easier to compute.)

(b) A group of three friends has n books they would all like to read. Each friend (independently of the other two) picks a random permutation of the books and reads them in that order, one book per week (for n consecutive weeks). Let X be the number of weeks in which all three friends are reading the same book. Compute $\text{Var}(X)$.

Solution:

(a) Let X be the number of floors the elevator does not stop at. We can represent X as the sum of the indicator variables $X_1, \ldots, X_n$, where $X_i = 1$ if no one gets off on floor i. Thus, we have

$$\mathbf{E}(X_i) = \Pr[X_i = 1] = \left(\frac{n-1}{n}\right)^m,$$

and from linearity of expectation,

$$\mathbf{E}(X) = \sum_{i=1}^{n} \mathbf{E}(X_i) = n\left(\frac{n-1}{n}\right)^m.$$

To find the variance, we cannot simply sum the variance of our indicator variables. However, we can still compute $\mathrm{Var}(X) = \mathbf{E}\left(X^2\right) - \mathbf{E}(X)^2$ directly using linearity of expectation, but now how can we find $\mathbf{E}\left(X^2\right)$? Recall that

$$\begin{aligned}
\mathbf{E}\left(X^2\right) &= \mathbf{E}\left((X_1 + \ldots + X_n)^2\right) \\
&= \mathbf{E}\left(\sum_{i,j} X_i X_j\right) \\
&= \sum_{i,j} \mathbf{E}\left(X_i X_j\right) \\
&= \sum_{i}^{n} \mathbf{E}\left(X_i^2\right) + \sum_{i \neq j} \mathbf{E}\left(X_i X_j\right).
\end{aligned}$$

The first term is simple to calculate:

$$\mathbf{E}\left(X_i^2\right) = 1^2 \Pr[X_i = 1] = \left(\frac{n-1}{n}\right)^m,$$

meaning that

$$\sum_{i=1}^{n} \mathbf{E}\left(X_i^2\right) = n\left(\frac{n-1}{n}\right)^m.$$

$X_i X_j = 1$ when both X_i and X_j are 1, which means no one gets off the elevator on floor i and floor j. This happens with probability

$$\Pr[X_i = X_j = 1] = \Pr[X_i = 1 \cap X_j = 1] = \left(\frac{n-2}{n}\right)^m.$$

Thus, we can now compute

$$\sum_{i \neq j} \mathbf{E}\left(X_i X_j\right) = n(n-1)\left(\frac{n-2}{n}\right)^m.$$

Finally, we plug in to see that

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

(b) Let $X_1, \ldots, X_n$ be indicator variables such that $X_i = 1$ if all three friends are reading the same book on week i. Thus, we have

$$\mathbf{E}\left(X_i\right) = \Pr[X_i = 1] = \left(\frac{1}{n}\right)^2,$$

and from linearity of expectation,

$$\mathbf{E}\left(X\right) = \sum_{i=1}^{n} \mathbf{E}\left(X_i\right) = n\left(\frac{1}{n}\right)^2 = \frac{1}{n}.$$

As before, we know that

$$\mathbf{E}\left(X^2\right) = \sum_{i}^{n} \mathbf{E}\left(X_i^2\right) + \sum_{i \neq j} \mathbf{E}\left(X_i X_j\right).$$

Furthermore, because X_i is an indicator variable, $\mathbf{E}\left(X_i^2\right) = 1^2 \Pr[X_i = 1] = 1/n^2$, and

$$\sum_{i}^{n} \mathbf{E}\left(X_i^2\right) = n\left(\frac{1}{n}\right)^2 = \frac{1}{n}.$$

Again, because X_i and X_j are indicator variables, we are interested in

$$\Pr[X_i = X_j = 1] = \Pr[X_i = 1 \cap X_j = 1] = \frac{1}{(n(n-1))^2},$$

the probability that all three friends pick the same book on week i and week j. Thus,

$$\sum_{i \neq j} \mathbf{E}\left(X_i X_j\right) = n(n-1)\left(\frac{1}{(n(n-1))^2}\right) = \frac{1}{n(n-1)}.$$

Finally, we compute

$$\mathrm{Var}(X) = \mathbf{E}\left(X^2\right) - \mathbf{E}\left(X\right)^2 = \frac{1}{n} + \frac{1}{n(n-1)} - \left(\frac{1}{n}\right)^2.$$