Massachusetts Institute of Technology
6.042J/18.062J, Fall '02: Mathematics for Computer Science
Prof. Albert Meyer and Dr. Radhika Nagpal

Solutions to In-Class Problems — Week 12, Wed

Problem 1. What are each of the following quantities when n independent Bernoulli trials are carried out with probability of success p_s? (Note: at most you are allowed n trials, not an infinite number of trials)

(a) The probability of no failures.

(b) The probability of at least one failure.

(c) The probability of at most one failure.

(d) The expected number of failures in n trials.

(e) The expected number of trials for the first failure.

Solution. Let F be a random variable equal to the number of failures in n trials.

$$\begin{aligned}
\Pr\{F = 0\} &= {p_s}^n \\
\Pr\{F > 0\} &= 1 - \Pr\{\text{No failure}\} = 1 - {p_s}^n \\
\Pr\{F \leq 1\} &= \Pr\{F = 0\} + \Pr\{F = 1\} \\
&= {p_s}^n + n(1 - p_s){p_s}^{n-1}
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}[F] &= \sum_{i=0}^{n} i \Pr\{F = i\} \\
&= \sum_{i=0}^{n} i \binom{n}{i} (1 - p_s)^i {p_s}^{n-i} \\
&= n(1 - p_s)
\end{aligned}$$

The expectation of a binomial distribution is worth remembering. But you can always derive it easily by using the linearity of expectations, *i.e.*, by assigning a random variable R_j to each trial j that has the value 1 if the trial is a success and 0 if not, and then summing those expectations. You can verify that you get the same answer.

(e) Let T be a random variable representing the number of trials before the first failure. There are several ways to solve this problem. One is to calculate the probability that it takes more than i trials before a failure occurs and use the corresponding expectation formula. This is similar to the Mean Time to Failure problem.

$$\begin{aligned} \mathrm{E}[T] &= \sum_{i=0}^{n} \Pr\{T > i\} \\ &= \sum_{i=0}^{n} {p_s}^i \\ &= \frac{1 - {p_s}^{n+1}}{1 - p_s} \end{aligned}$$

The last line uses the formula for the sum of a geometric series. Note that unlike the Mean Time to Failure example, the series is not from 0 to infinity.

Another method would be to calculate the probability that the first failure occurs at exactly i trials. This uses the usual definition of expectation.

$$\begin{aligned} \mathrm{E}[T] &= \sum_{i=0}^{n} i \Pr\{T = i\} \\ &= \sum_{i=0}^{n} i{p_s}^{i-1}(1 - p_s) \\ &= (1 - p_s) \sum_{i=0}^{n} i{p_s}^{i-1} \end{aligned}$$

We can derive the formula for the series $\sum_{i=0}^{n} ix^{i-1}$ by differentiating the formula for $\sum_{i=0}^{n} x^i$. The end result is the same. ■

Problem 2. Each bag of Doritos contains a cool sticker. There are n different kinds of sticker, and I want to collect at least one sticker of each kind. (Assume that sticker kinds in Dorito bags are uniformly random and mutually independent.)

(a) Suppose that I have already collected k kinds of sticker. What is the expected number of additional bags of Doritos that I must eat to collect one additional kind of sticker?

Solution. The probability that I must eat i or more bags is:

$$\begin{aligned}
&\text{E}\left[\text{number of bags to get new kind}\right] \\
&= \sum_{i=0}^{\infty} \Pr\left\{\text{number of bags to get new kind} > i\right\} \\
&= \sum_{i=0}^{\infty} \Pr\left\{\text{number of old kinds in first } i \text{ bags}\right\} \\
&= \sum_{i=0}^{\infty} \left(\frac{k}{n}\right)^i \\
&= \frac{n}{n-k}
\end{aligned}$$

Another way of thinking is that the probability of getting a new sticker in a bag is $(n-k)/n$, so what you want is the time until the first bag is opened that contains a new sticker which is just mean time to failure i.e. $E[T] = 1/p = n/(n-k)$. ■

(b) What is the expected number of bags of Doritos that I must eat to collect at least one sticker of each kind?

Solution. We sum the expected number of bags to get the first kind, the expected number to get the second kind, and so forth.

$$\begin{aligned}
&\text{E}\left[\text{number of bags to get all kinds}\right] \\
&= \sum_{k=0}^{n-1} \frac{n}{n-k} = n \cdot \sum_{k=0}^{n-1} \frac{1}{n-k} \\
&= n \cdot \sum_{j=1}^{n} \frac{1}{j} = nH_n.
\end{aligned}$$

■

Problem 3. **(a)** Suppose that I roll a 4-sided die, a 6-sided die, an 8-sided die, a 10-sided die, a 12-sided die, and a 20-sided die. What is the expected number of 6's that come up? (Assume that all of the dice are fair.)

Solution. Let X_i be the number of 6's on the i-sided die for $i \in \{4, 6, 8, 10, 12, 20\}$. Then the number of 6's on all dice is $X_4 + X_6 + X_8 + X_{10} + X_{12} + X_{20}$ and by linearity of expectation

$$\text{E}\left[X_4 + X_6 + X_8 + X_{10} + X_{12} + X_{20}\right] = \text{E}\left[X_4\right] + \text{E}\left[X_6\right] + \text{E}\left[X_8\right] + \text{E}\left[X_{10}\right] + \text{E}\left[X_{12}\right] + \text{E}\left[X_{20}\right].$$

Since every die has exactly one 6, $\mathrm{E}[X_i] = 1/i$, which gives

$$\mathrm{E}[X_4 + X_6 + X_8 + X_{10} + X_{12} + X_{20}] = \frac{0}{4} + \frac{1}{6} + \frac{1}{8} + \frac{1}{10} + \frac{1}{12} + \frac{1}{20} = \frac{21}{40}\,.$$

■

(b) Suppose that I roll n dice that are 6-sided, fair, and mutually independent. What is the expected value of the largest number that comes up?

Hint: You may want to use the "alternative" formula for expectation: $\mathrm{E}[M] = \sum_{i=0}^{\infty} \Pr\{M > i\}$. (This formula is valid since the random variables involved are non-negative.)

Solution. Let the random variable M be the largest number that comes up. To use the identity $\mathrm{E}[M] = \sum_{i=0}^{\infty} \Pr\{M > i\}$ we must compute the probability of the event $M > i$; that is, the event that some die shows a value greater than i.

$$\Pr\{M > i\} = 1 - \Pr\{\text{every die} \leq i\} = 1 - \prod_{k=1}^{n} \Pr\{k\text{th die} \leq i\} = 1 - (i/6)^n,$$

where the second equality holds since the dice are independent. Therefore,

$$\mathrm{E}[M] = \sum_{i=0}^{5} \Pr(M > i) = \sum_{i=0}^{5} 1 - (i/6)^n = 6 - \frac{1^n + 2^n + 3^n + 4^n + 5^n}{6^n}.$$

■