PROBLEM 1. With your group, roll a pair of dice twelve times. Record the first roll on which you roll doubles and also the total number of doubles that you roll and report these numbers to the instructor. What is the expected number of doubles in twelve rolls? How long should it take to roll doubles? How do these numbers compare with the class's statistics?

SOLUTION: We can model the sample space as $[6] \times [6]$, in which case the event of doubles is the diagonal $\Delta = \{(a, a) \mid a \in [6]\}$. Then under the uniform distribution, $P(\Delta) = 6/36 = 1/6$. Let X be the number of doubles out of 12 rolls. Let I_j denote the indicator variable for the j-th roll being a double. Then $E(I_j) = P(I_j = 1) = P(\Delta) = 1/6$. Since $X = I_1 + \cdots + I_{12}$, $E(X) = 12 \cdot 1/6 = 2$. We expect two doubles to be rolled.

The number of rolls expected until the first double is modeled by a geometric random variable with $p = 1/6$. Our text shows the expected value is $1/p = 6$.

PROBLEM 2. An airline has sold 205 tickets for a flight that can hold 200 passengers. Each ticketed person, independently, has a 5% chance of not showing up for the flight. What is the probability that more than 200 people will show up for the flight?

SOLUTION: Let X be the number of people who show up for the flight. We are looking for $P(X > 200) = P(X = 201) + P(X = 202) + \cdots + P(X = 205)$. Since this is a binomial random variable, $P(X = k) = \binom{205}{k}(0.95)^k(0.05)^{205-k}$. Thus

$$P(X > 200) = \sum_{k=201}^{205} \binom{205}{k} (0.95)^k (0.05)^{205-k} \approx 0.02236.$$

We conclude that the flight will be oversold about 2.2% of the time.

PROBLEM 3. If the same airline consistently oversells the flight from Problem 2 at the same rate, how many flights until we expect more ticketed passengers to show up than there are seats.

SOLUTION: This is a geometric random variable with $p = P(X > 200) \approx 0.02236$. As such, the expected number of flights until an oversold one is $1/p \approx 44.7$.

PROBLEM 4. With a binomial random variable, we run experiments independently, but there are many circumstances of interest that do not follow this pattern. One such is *sampling without replacement*: suppose we have a basket of N lottery tickets, K of which are winners. Consider a process in which you draw n of the tickets from the basket. Let X denote the number of winning tickets drawn; this is called a *hypergeometric random variable*.

Assume in this problem that $0 \leq K, n \leq N$.

(a) Prove that
$$P(X=k)=\frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}.$$

(b) In an election audit, a sample of machine-counted precincts are recounted by hand to check if the machine and hand audits match. Suppose there are N precincts, K of them have counting errors, we sample n precincts, and X counts the number of precincts in which errors are detected. In what sense is X a hypergeometric random variable, and what is the significance of the quantity $P(X=0)$?

(c) Suppose that there are machine-counting errors in 7 of 200 precincts. How many precincts must one sample in order to guarantee that there is at most a 5% chance of detecting no errors?

SOLUTION:

(a) We are choosing one of the $\binom{N}{n}$ subsets of the initial set. In order to have exactly k winners, we must choose one of the $\binom{K}{k}$ sets of winners with k elements, and choose one of the $\binom{N-K}{n-k}$ sets of $n-k$ losers. By multiplicative counting,
$$P(X=k)=\frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}.$$

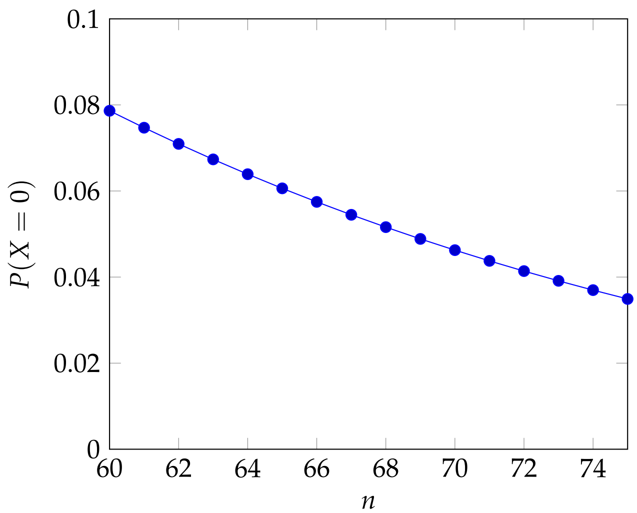
(b) This is exactly the setup for a hypergeometric random variable where "success" is the existence of a counting error in a given precinct. The quantity $P(X=0)$ is the probability that our sample detects no errors. (We would like for this number to be small!)

(c) Let n denote the number of precincts samples. We have
$$P(X=0)=\frac{\binom{7}{0}\binom{193}{n}}{\binom{200}{n}}=\frac{193!/(193-n)!}{200!/(200-n)!}=\frac{193!(200-n)!}{200!(193-n)!}.$$

Examining the following graph and playing with a calculator,* we see that the probability first dips below 0.05 when $n=69$. This is a fairly large sample of precincts! Hopefully the elections board has hired a mathematician.†

* Or doing some heinous algebra?

† Some states *have* hired mathematicians because they are required by statute to conduct risk-limiting audits. These are incremental systems in which fewer ballots are reviewed when the margin of an election is wide, and they don't fall under the rubric to which hypergeometric random variables apply. See the National Conference of State Legislatures website for more information.


0.1
0.08
0.06
0.04
0.02
0
60
62
64
66
68
70
72
74
$P(X = 0)$
n