Stats100A **Summer 2024**

Week 5: Challenge Problems 5 - Solutions

Author: Andrew Lizarraga

Solutions

Problem 1 (A Blood Test): Suppose that a laboratory test on a blood sample yields one of two results: positive or negative. It is known that 95% of people with a particular disease produce a positive result. But 2% of people without the disease will also produce a positive result (a false positive). Suppose that 1% of the population has the disease. What is the probability that a person chosen at random from the population will have the disease, given that the person's blood yields a positive result.

Solution: Let D indicate a person has the disease, let $+$ denote a positive test, and $-$ denote a negative test. We are given $P(+|D) = \frac{95}{100}, P(+|D^c) = \frac{2}{100}, P(D) = \frac{1}{100}$. We are trying to find $P(D|+)$. This is a simple application of Baye's rule:

$$P(D|+) = \frac{P(+|D)P(D)}{P(+|D)P(D) + P(+|D^c)P(D^c)}$$

$$P(D|+) \approx 32\%$$

Problem 2 (Biased Runs): A biased coin is tossed n times, and heads shows with probability p on each toss. A run is a sequence of throws which result in the same outcome, so that, for example, the sequence $HHTHTTH$ contains five runs. Show that the expected number of runs is $1 + 2(n-1)p(1-p)$. Find the variance of the number of runs.

Solution: Let X_i be the indicator function of the event that the outcome of the $i+1$th toss is different from the outcome of the ith toss. Let X denote the number of distinct runs, which can be expressed in terms of indicators as follows: $X = 1 + \sum_{i=1}^{n-1} X_i$. Thus the expectation is given by:

$$E(X) = 1 + (n-1)E(X_i)$$

$$E(X) = 1 + 2(n-1)p(1-p)$$

Problem 3 (Urn Removal): An urn contains n balls numbered $1, \ldots, n$. We remove k balls at random (without replacement) and add up their numbers. Find the mean and variance of the total.

Solution: Let X_i denote the number shown on the ith ball and note that $E(X_i) = \frac{n+1}{2}$. Then the total for the k removed balls is given by $X = \sum_{i=1}^{k} X_i$, which has expectation given by $E(X) = \sum_{i=1}^{k} E(X_i) = \frac{k(n+1)}{2}$.
We use the following formula to compute the variance: $\mathrm{Var}(X) = E(X^2) - (E(X))^2$. Note that $(E(X))^2 = \frac{k^2(n+1)^2}{4}$. So all that remains is to compute $E(X^2)$:

$$
\begin{aligned}
E(X^2) &= E\left[\left(\sum_{i=1}^{k} X_i\right)^2\right] \\
&= E\left[\sum_{i=1}^{k} X_i^2 + \sum_{i \neq j} X_i X_j\right] \\
&= \sum_{i=1}^{k} E(X_i^2) + \sum_{i \neq j} E(X_i X_j) \\
&= kE(X_1^2) + 2\binom{k}{2} E(X_1 X_2) \\
&= k\left(\frac{1}{n}\sum_{j=1}^{n} j^2\right) + k(k-1)E(X_1 X_2) \\
&= \frac{k}{n}\left(\sum_{j=1}^{n} j^2\right) + k(k-1)E(X_1 X_2) \\
&= \frac{k}{n}\left(\sum_{j=1}^{n} j^2\right) + \frac{k(k-1)}{n(n-1)} 2 \sum_{i \neq j} ij \\
&= \frac{k}{n}\left(\frac{1}{3}n(n+1)(n+2) - \frac{1}{2}n(n+1)\right) + \frac{k(k-1)}{n(n-1)} 2 \sum_{i \neq j} ij \\
&= \frac{k}{n}\left(\frac{1}{3}n(n+1)(n+2) - \frac{1}{2}n(n+1)\right) + \frac{k(k-1)}{n(n-1)} 2 \sum_{j=1}^{n} j(n(n+1) - j(j+1)) \\
&= \frac{1}{6}k(n+1)(2n+1) + \frac{1}{12}k(k-1)(3n+2)(n+1)
\end{aligned}
$$

From here we have:

$$
\mathrm{Var}(X) = E(X^2) - (E(X))^2 = \frac{1}{12}(n+1)k(n-k)
$$

Problem 4 (Simpson's Game): You have two bags, each with a mixture of green and red marbles. You can choose a bag, then randomly pick a marble from the bag. If it's green you win \$100, but if it's red, you get nothing.

Game 1: Bag A contains 5 green marbles and 6 red marbles. Bag B has 3 green marbles and 4 red marbles. What bag should you choose to increase your odds of winning?

Game 2: Bag C contains 6 green marbles and 3 red marbles. Bag D has 9 green marbles and 5 red marbles. What bag should you choose to increase your odds of winning?

Game 3: We combine bags A and C into bag E and we combine bags B and D into bag F. What bag should you choose to increase your odds of winning?

Solution:

Game 1: Since there is a $\frac{5}{11} > \frac{3}{7}$ chance of winning with bag A, you should pick bag A.

Game 2: Since there is a $\frac{6}{9} > \frac{9}{14}$ chance of winning with bag C, you should pick bag C.

Game 3: Naively, you might think that bag E should have higher odds of winning than bag F. However this is wrong notice that the probability of winning with bag F is $\frac{12}{21} > \frac{11}{20}$, which seems counterintuitive, because bags A and C win with higher probability than bags B and D respectively, but when combined bag E wins with lower probability than bag F.