1. An article in the *Los Angeles Times* reports that 1 in 200 people carry the gene that causes colon cancer. Suppose you randomly sample 1000 individuals.

   a) What is the distribution of the number of sampled individuals who have the gene?

   The distribution is Binomial with $n = 1000$ and $q = 1/200 = 0.005$.

   b) What is a reasonable approximate distribution?

   The Poisson distribution can approximate the Binomial when n is large. Then, $\lambda = nq = 1000/200 = 5$.

   c) Use the exact and the approximate distributions to calculate the probability that between 5 and 8 people have the gene. Compare your results.

   Using the Binomial,

$$\binom{P(5 \le X \le 8) = \sum_{k=5}^{8} 1000}{k(0.005)^k(0.995)^{1000-k}}$$

$$= .1759 + .1466 + .1046 + .0652 = .49$$

   Using Poisson,

$$P(5 \le X \le 8) = \sum_{k=5}^{8} \frac{5^k\, e^{-5}}{k!}$$

$$= .1755 + .1462 + .1044 + .0653 = 0.49$$

   The approximation is very accurate!

2. A toll bridge charges \$1.00 for passenger vehicles and \$2.50 for any other vehicle. Suppose that during daytime hours 60% of vehicles using the

toll bridge are passenger vehicles. If 25 vehicles pass over the toll bridge on a given day, what is the expected toll revenue?

Let X be the number of passenger vehicles passing over the bridge out of the 25 vehicles. Then, the revenue can be calculated as $1.00 * X + 2.50 * (25 - X)$. It may not surprise you that the expectation operator is *linear*, that is, $E(aX + b) = aE(X) + b$ for constants a and b. Since X has a binomial $n = 25$, $q = .6$ distribution, its expectation is $E(X) = 15$. Then, the expected revenue is $1.00 * 15 + 2.50 * (25 - 15) = 40$.

3. Customers at a gas station pay with a credit card, a debit card, or cash. Assume that successive customers make independent choices, with $P($ credit card $) = .5$, $P($ debit card $) = .2$, $P($ cash $) = .3$. Among the next 100 customers, what are the mean and variance of the number who pay with a credit card?

   Even though there are three options, credit, debit, and cash, since we are only interested in credit vs. non-credit, we have a binomial distribution with $n = 100$ and $q = 0.5$. Then, the mean and variance are $100*0.5 = 50$ and $100 * 0.5 * 0.5 = 25$.

4. A geologist has collected 10 samples of basaltic rock and 10 samples of granite. The geologist instructs his graduate assistant to randomly pick 6 rocks for analysis.

   a) What is the probability that all samples selected for analysis are of the same type?

   Since we know the whole population of rocks, the hypergeometric distribution applies in this problem. The probability that we have all rocks of the same type is the probability we get all basalt or all granite. This is

$$2 * \frac{\binom{10}{6}}{\binom{20}{6}} = 0.0108$$

b) What is the probability distribution for the number of granite samples selected for analysis?

Hypergeometric with $M = 10$, $n = 6$, $N = 20$. c) What is the probability that the number of granite samples selected is within one standard deviation of its mean? The mean is $\frac{nM}{N} = 3$. The variance is $\frac{N-n}{N-1} n \frac{M}{N} \frac{N-M}{N} = 1.105$ and the standard deviation is about 1.05. Then the probability we are interested in is the probability we get either 2, 3, or 4 granite rocks in a sample of 6. This is

$$\frac{\binom{10}{2}\binom{10}{4}}{\binom{20}{6}} + \frac{\binom{10}{3}\binom{10}{3}}{\binom{20}{6}} + \frac{\binom{10}{4}\binom{10}{2}}{\binom{20}{6}} = .859$$

5. Suppose that $q = 0.5 = P($ female birth $)$. A couple wishes to have exactly two daughters in their family. They will keep having children until they have two daughters.

   a) What is the probability that the family has x children?

   Since the negative binomial distribution applies with $q = .5$ and $k = 2$, the expected number of children is the expected number of "trials" which is $2/q = 2/0.5 = 4$ b) What is the probability that the family has x male children?

   The probability that the family has x male children is the same as the probability the family has $x + 2$ children, which is the negative binomial above. c) What is the probability that the family has at most 4 children?

$$P(2, 3, 4) = .5^2 + \binom{2}{1}(.5)^3 + \binom{3}{1}.5^4 = .6875.$$

d) How many male children would you expect this family to have?

If they expect to have 4 children, and 2 of those must be girls, then they expect to have 2 boys.