DISCUSSION 2

Practice Problems

1. Consider a discrete random variable X with $E(X) = 2.5$. The probability distribution of X is given by the following table.

k	0	?	2	3	4
P(X=k)	0.1	0.2	0.1	?	0.3

Fill in the missing values of the table.

2. Consider two random variables Y_1 and Y_2 where $E(Y_1) = 3$ and $E(Y_2) = -1$. In addition, $E(Y_1^2) = 10$ and $E(Y_2^2) = 2$. Calculate the following.

   (a) $E(3Y_1 - 2Y_2)$

   (b) $E(Y_1 - 2Y_1^2 - Y_2 + 6Y_2^2)$

   (c) $Var(Y_1)$ and $Var(Y_2)$

   (d) If additionally assuming Y_1 and Y_2 are independent, calculate $Var(2Y_1 - 3Y_2)$.

3. In 1995, John Wayne played Genghis Khan in a movie called *The Conqueror*. Unfortunately the movie was filmed downwind of the site of 11 above-ground nuclear bomb tests. Of the 220 people who worked on this movie, 91 had been diagnosed with cancer by the early 1980s., including Wayne, his co-stars and the director. According to large-scale epidemiological data, only about 14% of people of this age group, on average, should have been stricken with cancer within this time frame. We want to know whether there is evidence for an increased cancer risk of people associated with this film.

   (a) What is the best estimate of the probability of a member of the cast or crew getting cancer within the study interval? Assume that this probability is the same for each member of the cast.

   (b) What is the standard error of your estimate? What does this quantity measure?

   (c) What is the 95% confidence interval for this probabilty estimate? Does this interval bracket the typical cancer rate of 14% for people of the same age group? Interpret the result.

4. To infer the number of red balls in a bin of a mixture of 600 red and white balls, Tom randomly select 60 balls with replacement and 12 of them are red.

   (a) What probability distribution would we use to calculate the number of red balls in a sample of size 60?

   (b) Write the likelihood of p and the log likelihood of p.

   (c) Evaluate the log-likelihood at the value $p = 0.5$ and $p = 0.2$ respectively, given the data.

Solution

1. (a) $P(X=3) = 1 - 0.1 - 0.2 - 0.1 - 0.3 = 0.3$.

   (b) Denote the missing possible value of X by y. Since $E(X) = 2.5$, we have $0*0.1 + y*0.2 + 2*0.1 + 3*0.3 + 4*0.3 = 1.9$. By solving the equation, we get $y = 1$.

2. (a) $E(3Y_1 - 2Y_2) = 3*E(Y_1) - 2*E(Y_2) = 3*3 - 2*(-1) = 11$

   (b) $E(Y_1 - 2Y_1^2 - Y_2 + 6Y_2^2) = E(Y_1) - 2*E(Y_1^2) - E(Y_2) + 6*E(Y_2^2) = 3 - 2*10 - (-1) + 6*2 = -4$

   (c) $Var(Y_1) = E(Y_1^2) - (E(Y_1))^2 = 10 - 3^2 = 1$ and similarly, $Var(Y_2) = 2 - (-1)^2 = 1$.

   (d) If additionally assuming Y_1 and Y_2 are independent,
   $Var(2Y_1 - 3Y_2) = 2^2 * Var(Y_1) + (-3)^2 * Var(Y_2) = 4*1 + 9*1 = 13$.

3. (a) If assuming the same probability of being diagnosed as cancer, the number of people out of 220 getting the disease is a binomial distributed random variable. The best estimate of such a probability is the sample proportion $\frac{91}{220} = 0.414$

   (b) The standard error of the estimate of the proportion is approximated by $\sqrt{\frac{0.414(1-0.414)}{220}} = 0.033$, which is an estimate of the size of the difference between p and $\hat{p}$.

   (c) i. $p' = \frac{91+2}{220+4} = 0.415$

   ii. The estimated standard error is $\sqrt{\frac{0.415(1-0.415)}{220+4}} = 0.033$.

   iii. The margin of error is $1.96 * 0.033 = 0.065$.

   iv. We then construct 95% confidence interval for p.

$$0.415 - 0.065 < p < 0.415 + 0.065$$
$$0.350 < p < 0.480$$

   We are 95% confident that the probabilty of a member of the cast geeting cancer is between 0.350 and 0.480. It does not bracket the typical cancer rate of 14% for people of the same age group. Thus we are 95% confident that the the probability of getting cancer in the cast is higher than the typical cancer rate 14%.

4. (a) The number of red balls that Tom draws X is binomial distributed random variable, where $X \sim Binomial(60, p)$ and p is the proportion of red balls in the bin.

   (b)

$$L(p) = P(X = 12|p) = \binom{60}{12} p^{12}(1-p)^{48}$$
$$l(p) = logL(p) = log\binom{60}{12} + 12 * log(p) + 48 * log(1-p)$$

$$l(0.5) = log\binom{60}{12} + 12 * log(0.5) + 48 * log(1-0.5) = -13.622$$
$$l(0.2) = log\binom{60}{12} + 12 * log(0.2) + 48 * log(1-0.2) = -2.057$$

```
#Rcode:
> log(dbinom(12,60,0.5))
[1] -13.62180
> log(dbinom(12,60,0.2))
[1] -2.05711
> dbinom(12,60,0.5,log=T)
[1] -13.62180
> dbinom(12,60,0.2,log=T)
[1] -2.05711
```