Discussion 3

Practice Probelm

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. 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 whehter there is evidence for an increased cancer risk of people assoicated 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 probablity estimate? Does this interval barcket the typical cancer rate of 14% for people of the same age group? Interpret the result.

3. 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.

4. A group of researchers tested whether snakes tend to choose a warm resting site when both a warm site and a cool site are presented to them. Their hypotheses were H_0: Snakes do not prefer the warmer site. H_A: Snakes prefer the warmer site. They carried out the experiment and with their data calculated a one-tailed P-value of $P = 0.03$. They rejected their null hypothesis and concluded that snakes prefer the warmer sites.

   a. Is one-tailed test appropriate here? Explain.

   b. What would their hypothesis statements have been had they use a two-tailed test instead?

   c. What would their P-value have been had they used a two-tailed test instead?

5. In a test of Murphy's Law, pieces of toast were buttered on one side and then dropped. Out of 9821 total slices of toast dropped, 6101 landed butter-side down. Is it reasonable to believe that there is a 50:50 chance of the toast landing butter-side down or butter-side up? State hypotheses, find a test statistic, and compute the p-value. What is your conclusion?

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) 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%.

3. (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
```

4. a. Yes, because in H_A, they said snakes prefer WARMER site.

   b. H_0: Snakes do not have preference on their sites' temperature. H_A: Snakes do have preference on their sites' temperature.

   c. $0.03 \times 2 = 0.06$.

5. Denote p as the probability that the toast lands butter-side down.

$$H_0 : p = 0.5 \qquad H_A : p \neq 0.5$$

See the following R output for the results of hypothesis test.

```
> binom.test(6101,9821,p=0.5,alternative="two.sided")

        Exact binomial test

data:  6101 and 9821
number of successes = 6101, number of trials = 9821, p-value < 2.2e-16
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
 0.6115404 0.6308273
sample estimates:
probability of success
             0.6212198
```

CONCLUSION: There is very strong evidence to reject the hypothesis that the chance of the toast landing butter-side down is 50:50.