STAT310 - HWK Solution 5

1. (a)

$$|\bar{X}-\theta| \leqslant z\sqrt{\bar{X}(1-\bar{X})/n} \Longrightarrow (n^3-n^2z^2)X^2-(2n^2\theta+nz^2)X+n^3\theta^2$$

$$\Longrightarrow a = \frac{2n^2\theta+nz^2-\sqrt{(2n^2\theta+nz^2)^2-4n^3\theta^2(n+z^2)}}{2n+2z^2}$$

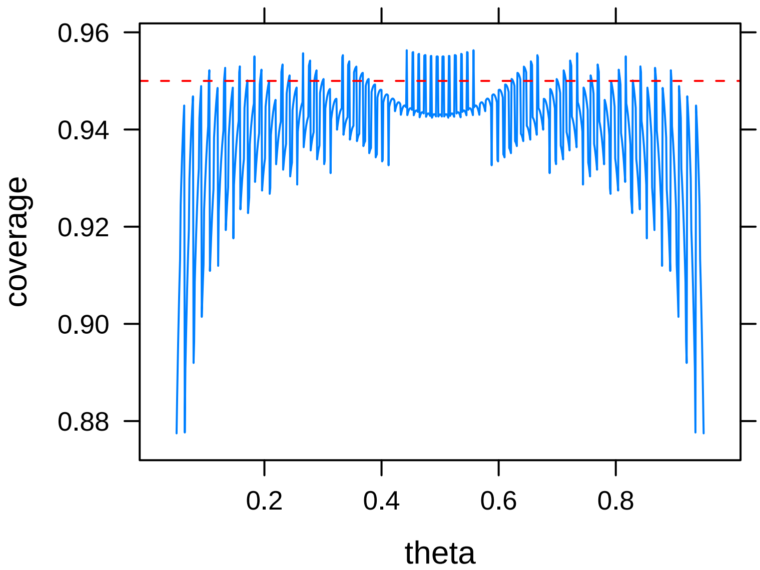
$$b = \frac{2n^2\theta+nz^2+\sqrt{(2n^2\theta+nz^2)^2-4n^3\theta^2(n+z^2)}}{2n+2z^2}$$

(b) a=2.9095, b=10.3795, coverage probability=0.9474

(c) a=3.1098, b=10.7892, coverage probability=0.8880

(d) Plot of true coverage probability

```
> source("http://www.stat.wisc.edu/courses/st310-larget/stat310.R")
> theta = seq(0.05,0.95,0.001)
> coverage = exact.binomial.coverage(100,theta)
> fig = xyplot(coverage ~ theta, type="l",
              panel = function(x,y,...){
                panel.xyplot(x,y,...)
                panel.abline(h=0.95,col="red",lty=2) } )
> print( fig )
```



For most θ values, the true coverage probabilities are below 0.95 and can be as far as to 0.88. When θ is close to 0 or 1, there is great variation existing in the coverage probability.

2. (a)

$$|\frac{X+2}{n+4}-\theta| \leqslant z\sqrt{\frac{\frac{X+2}{n+4}(1-\frac{X+2}{n+4})}{n+4}}$$

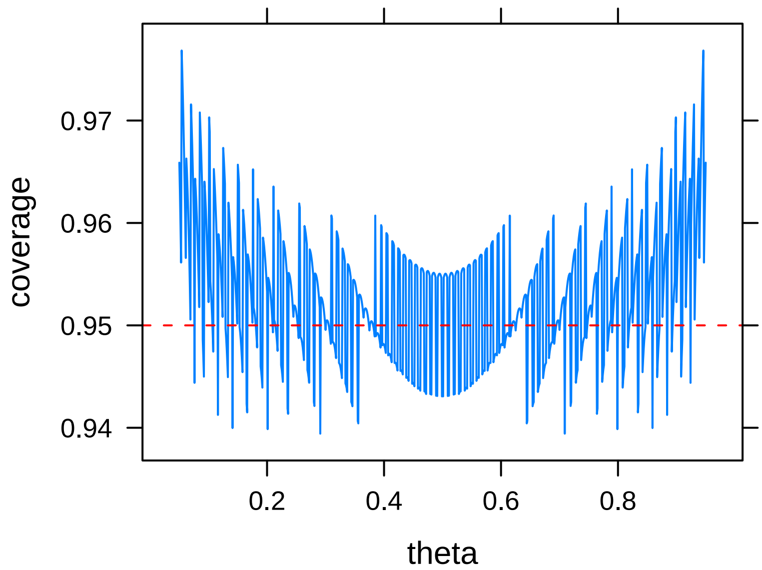
$$\Longrightarrow a = \frac{(n+4)[2\theta(n+4)+x^2]) - (n+4)\sqrt{[2\theta(n+4)+z^2]^2 - 4(n+4+z^2)(n+4)\theta^2}}{2(n+4+z^2)} - 2$$

$$b = \frac{(n+4)[2\theta(n+4)+x^2]) + (n+4)\sqrt{[2\theta(n+4)+z^2]^2 - 4(n+4+z^2)(n+4)\theta^2}}{2(n+4+z^2)} - 2$$

For n=20, $\theta = 0.3$, a=1.7236, b=10.001, coverage probability=0.9752
For n=21, $\theta = 0.3$, a=1.9321, b=10.3998, coverage probability=0.9680

(b) Plot of true coverage probability

```
> source("http://www.stat.wisc.edu/courses/st310-larget/stat310.R")
> theta = seq(0.05,0.95,0.001)
> coverage = exact.binomial.coverage(100,theta,adj=2)
> fig = xyplot(coverage ~ theta, type="l",
              panel = function(x,y,...){
                panel.xyplot(x,y,...)
                panel.abline(h=0.95,col="red",lty=2) } )
> print( fig )
```



(c) Compared to method in problem 1, the true coverage probabilities of the adjusted method are closer to 0.95. Most time if it is wrong, it is more conservative with coverage probability ¿0.95.

(d) The adjusted method is better for both large and small n.

3. To detect if these results are statistically significant or not we need to perform a z-test for testing $H_0 : \mu = 1$. The P-value is given by

$$P - value = P(|Z| \geqslant |\frac{1-1.05}{\sqrt{0.1/500}}|) = 0.000407$$

So these results are statistically significant at the 5% level, so we have evidence against $H_0 : \mu = 1$. Also, the observed difference of 1.05 - 1 = 0.05 is well within the range that the manufacturer thinks is of practical significance. So the test has detected a small difference that is not practically significant.

4. Based on the two-sided z-test to assess $H_0 : \theta = \frac{1}{6}$,

$$P - value = P(|Z| \geqslant |\frac{\sqrt{n}(\bar{x} - \theta)}{\sqrt{\theta(1-\theta)}}|) = P(|Z| \geqslant |\frac{\sqrt{30}(\frac{1}{3} - \frac{1}{6})}{\sqrt{\frac{1}{6}(1 - \frac{1}{6})}}|) = 0.014$$

So we can conclude that at the 5% significance level, there is evidence to conclude that the die is biased.

5. Let X be the number of tornadoes in a given year, then $X \sim Possion(\theta)$. The MLE of θ is $\hat{\theta}_{MLE} = \bar{X}$. We know that $\hat{\theta} \xrightarrow{D} N(\theta, \frac{1}{nI(\hat{\theta})})$. It can be found that the fisher information for Possion model is $I(\theta) = \frac{1}{\theta}$, thus $\hat{\theta} \xrightarrow{D} N(\theta, \frac{\hat{\theta}}{n})$. Replace $\hat{\theta}$ by $\bar{X}$, then we have $\bar{X} \xrightarrow{D} N(\theta, \frac{\bar{X}}{n})$; with some manipulation, we get $\frac{\bar{X}-\theta}{\sqrt{\bar{X}/n}} \xrightarrow{D} N(0,1)$.

$$P(-z_0.975 \leqslant \frac{\bar{X} - \theta}{\sqrt{\bar{X}/n}} \leqslant z_0.975) = 0.95$$

$$\Longrightarrow P(\bar{X} - z_0.975\sqrt{\frac{\bar{X}}{n}} \leqslant \theta \leqslant \bar{X} + z_0.975\sqrt{\frac{\bar{X}}{n}}) = 0.95$$

Thus, the 95% confidence interval is $(\bar{X} - z_0.975\sqrt{\frac{\bar{X}}{n}}, \bar{X} + z_0.975\sqrt{\frac{\bar{X}}{n}})$. Plug in $\bar{x} = 21.7826, n = 46, x_0.975 = 1.96$, the 95% confidence interval is (20.434, 23.131).

6. (a) The confidence interval is calculated by $\bar{x} \pm t_{0.975,9}\frac{s}{\sqrt{n}}$, where s is the sample standard deviation and $t_{0.975,9}$ is the 97.5% quantile of t distribution with degrees of freedom 9. The confidence interval for μ is (4.382, 5.377).
   (b) $P - value = P(|T| \geqslant |\frac{\bar{x}-\mu}{s/\sqrt{n}}|) = 0.0031$
   (c) $P - value > 0.05$
   (d) $P - value < 0.05$