Stat 245 HW1 Solutions

1. If $(X - np)/\sqrt{np(1-p)}$ is a pivotal, then for a fixed sample-size n, that random variable has the same distribution regardless of the value of the parameter p. In particular, the span of the support of these distributions should be the same. The support for a binomial X is $0, 1, 2, \cdots, n$ and its span is n. Therefore the span of the support of $(X - np)/\sqrt{np(1-p)}$ is $np/\sqrt{np(1-p)} = \sqrt{np/(1-p)}$, which is not a constant quantity for different values of p.

2. (Simulation part) The R code for the simulation is as follows:

```
> n=30
> alpha=0.05
> lambda=1
> N=1000
> z=qnorm(1-alpha/2)
>
> m=rep(0,N)
> for(i in 1:N)
+ m[i]=mean(rpois(n,lambda))
>
> l1=m+z^2/(2*n)-(4*m*z^2/n+z^4/n^2)^.5/2
> u1=m+z^2/(2*n)+(4*m*z^2/n+z^4/n^2)^.5/2
> l2=m-z*(m/n)^.5
> u2=m+z*(m/n)^.5
>
> mean((l1<1)&(1<u1))
[1] 0.942
> mean((l2<1)&(1<u2))
[1] 0.926
```

Here, we repeat the simulation study for $N = 1000$ times. In the code above, the 1000 means of 30 Poisson random variables are stored in the vector m. The vectors l1 and u1 are the lower and upper bounds of the confidence intervals by using the first method and l2 and u2 are those by the second method. The last two commands compute the proportions of the covering intervals using the first method and the second method, respectively. From the simulation results, both types of intervals have coverage probability lower than the specified level 0.95. But the first method performs slightly better than the second method. We can do this simulation again using more repetitions to obtain a more accurate estimate. Take, for example, $N = 100,000$. We reach a similar conclusion as before:

```
> N=100000
```

```
> m=rep(0,N)
> for(i in 1:N)
+ m[i]=mean(rpois(n,lambda))
>
> l1=m+z^2/(2*n)-(4*m*z^2/n+z^4/n^2)^.5/2
> u1=m+z^2/(2*n)+(4*m*z^2/n+z^4/n^2)^.5/2
> l2=m-z*(m/n)^.5
> u2=m+z*(m/n)^.5
>
> mean((l1<1)&(1<u1))
[1] 0.94595
> mean((l2<1)&(1<u2))
[1] 0.93072
```

2 (Sample size determination)

The maximum value of the polynomial $x(1-x)$ is obtained at $x = 1/2$. Therefore, the length of the confidence interval, which equals

$$2z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

is bounded by

$$\frac{z_{\alpha/2}}{\sqrt{n}},$$

and that is the smallest upper bound possible, since $\hat{p}$ can equal $1/2$. Since $z_{\alpha/2} = z_{0.975} = 1.96$, to obtain a length $\hat{U} - \hat{L} \leq 0.06$ it is enough to take $n \geq n_0 = \text{ceiling}((1.96/0.06)^2) = 1068$.

3 *(Approximate confidence intervals for Binomial distribution)*

(a) We follow the construction given in Brown *et al.*, page 114 (only that we will use the estimator $\hat{p} = X/n$ instead of the Anscombe estimator $\tilde{p} = \frac{X+3/8}{n+3/4}$ used in the paper). We have that

$$2n^{1/2}[\arcsin(\hat{p}^{1/2}) - \arcsin(p^{1/2})]$$

has a distribution that is close to the standard normal $N(0,1)$ when n is large. This leads to an approximate confidence interval given by

$$\left[\sin^2\left(\arcsin(\hat{p}^{1/2}) - \frac{z}{2\sqrt{n}}\right)\ ,\ \sin^2\left(\arcsin(\hat{p}^{1/2}) + \frac{z}{2\sqrt{n}}\right)\right],$$

where $z = z_{\alpha/2}$.

(b) If the approximations are good, we expect a proportion of approximately 0.95 of successful coverages.

We code the whole procedure in a function, with some standard values for the parameters. Changing these values we can run different simulations. Here is the R code:

```
interval.comparison <- function(n, k=100, p=0.1, alpha=0.05){

z <- qnorm(1-alpha/2)

# Generate the samples:
X <- rbinom(k,n,p)

# Compute the estimator
p.hat <- X/n

# Computations for the Wald CIs:
Wald.lower <- p.hat - z*sqrt(p.hat*(1-p.hat)/n)
Wald.upper <- p.hat + z*sqrt(p.hat*(1-p.hat)/n)
```

3(a) $g(\hat{p}) - g(p) \approx g'(p)(\hat{p}-p) + O(|\hat{p}-p|^2)$

$\mathrm{Var}(g(\hat{p})) = g'(p)^2 \frac{p(1-p)}{n}$.

$g'(p) = \frac{1}{\sqrt{p(1-p)}}$

$g(p) = \int_0^p \frac{1}{\sqrt{t(1-t)}}\,dt$. $t = \sin^2\theta$

$= \int_0^{\arcsin\sqrt{p}} 2\,d\theta = 2\arcsin\sqrt{p}$.

```
Wald.length <- Wald.upper - Wald.lower
Wald.covers <- Wald.lower < p & Wald.upper > p


# Computations for the Wilson CIs:
Wilson.lower <- ( p.hat + z^2/(2*n) -
      sqrt(p.hat*(1-p.hat)*z^2/n + z^4/(4*n^2)) ) / (1+z^2/n)
Wilson.upper <- ( p.hat + z^2/(2*n) +
      sqrt(p.hat*(1-p.hat)*z^2/n + z^4/(4*n^2)) ) / (1+z^2/n)
Wilson.length <- Wilson.upper - Wilson.lower
Wilson.covers <- Wilson.lower < p & Wilson.upper > p


# Computations for the Arcsine CIs:
Arc.lower <-sin( asin(sqrt(p.hat)) - z/(2*sqrt(n)) )^2
Arc.upper <-sin( asin(sqrt(p.hat)) + z/(2*sqrt(n)) )^2
Arc.length <- Arc.upper - Arc.lower
Arc.covers <- Arc.lower < p & Arc.upper > p

# Print out summaries of the results
print(paste("Proportion of successful coverage by the Wald CIs:",
                  mean(Wald.covers) ))
print(paste("Proportion of successful coverage by the Wilson CIs:",
                  mean(Wilson.covers) ))
print(paste("Proportion of successful coverage by the Arcsine CIs:",
                 mean(Arc.covers) ))

}
```

Running this program we obtain:

```
> interval.comparison(30)
[1] "Proportion of successful coverage by the Wald CIs: 0.83"
[1] "Proportion of successful coverage by the Wilson CIs: 0.98"
[1] "Proportion of successful coverage by the Arcsine CIs: 0.96"
```

Repeating the simulation we observed that for p near 0.5, all the intervals perform well; for moderately small values of p, the Wald intervals start to fail, and for very small values of p (say, 0.001) both the Wald and the Arcsine interval fail almost always, while the Wilson interval still performs well.

(c) We use again the function above:

```
> interval.comparison(150)
```

```
[1] "Proportion of successful coverage by the Wald CIs: 0.91"
[1] "Proportion of successful coverage by the Wilson CIs: 0.97"
[1] "Proportion of successful coverage by the Arcsine CIs: 0.95"
```

The behavior of the Wald interval improves, but it is not yet as good as the others.

Remark: Notice that all the operations in the function we coded are *vectorized*, that is, all the operations are performed "simultaneously" for all the 100 samples. This, in general, is much more efficient than using loops.

4 *(Distribution of a ratio)* Since X_1, X_2 are independent, their joint distribution has density:

$$f(x_1, x_2) = e^{-(x_1+x_2)}, \qquad \text{for } x_1, x_2 > 0.$$

Consider the transformation

$$T(X_1, X_2) = (U_1, U_2) = (X_1/X_2, X_2),$$

where we assume that $X_1, X_2 > 0$ with probability 1. The inverse transformation is given by:

$$T^{-1}(u_1, u_2) = (u_1 u_2, u_2),$$

so its Jacobian is:

$$J = \begin{vmatrix} u_2 & 0 \\ u_1 & 1 \end{vmatrix} = u_2, \qquad u_2 > 0.$$

This way, the joint distribution of (U_1, U_2) is given by the density:

$$g(u_1, u_2) = f(u_1 u_2, u_2) u_2 = e^{-(u_1+1)u_2} u_2.$$

Now we can compute the density for U_1, as a marginal density:

$$h(u_1) = \int_0^\infty e^{-(u_1+1)u_2} u_2 \, du_2 = (u_1 + 1)^{-2}.$$

This is the density of $F_{2,2}$, the F distribution with 2 degrees of freedom in both the numerator and the denominator.

5. *(F distribution)*

(a) Figure 1 shows the plot of the four densities. We can observe that the peak of the plots moves to the right, mostly under the influence of the numerator degrees of freedom.

Figure 1 was created using the following commands. I started with `pdf(...)` in order to create a pdf file with the plot, so that I could include it in this document (which was created using pdfLaTeX). You might want to use the commands `png(...)`, `postscript(...)`, etc. depending on what program you are using to write your homework).

5.

Prob 16. (a) (b) see below.

(c) FALSE. The CI makes no statement about the spread of the population, it makes a statement about the accuracy of our estimation of a parameter.

(d) FALSE. The statement is true only in average, in the long run. *Approximately* 95 of the intervals will contain μ.

Problem 17, pg. 241, Third Ed. (Problem 13, pg. 226, Second Ed.)

FALSE. The CI makes no statement about the spread of the population, it makes a statement about the accuracy of our estimation of a parameter. Confidence intervals become narrower as the sample size increases, so how could they still contain a fixed proportion of the population?

Problem 18, pg. 241, Third Ed. (Problem 14, pg. 226, Second Ed.)

Due to independence, we have:

$$\begin{aligned}\text{Prob[Neither CI contains the respective mean]} &= \\ \text{Prob[First CI does not contain first mean]} &\times \text{Prob[Second CI does not contain second mean]} \\ &= 0.1 \cdot 0.1 \\ &= \boxed{0.01}\end{aligned}$$

$$\begin{aligned}\text{Prob[Both CI's contain the respective means]} &= \\ \text{Prob[First CI contains first mean]} &\times \text{Prob[Second CI contains second mean]} \\ &= 0.9 \cdot 0.9 \\ &= \boxed{0.81}\end{aligned}$$

prob 16 (a). True: It is a value computed from the data, which is a random variable.

prob 16 (b): False. If built around the sample mean, it always contains the sample mean. Goal is to construct the CI s.t. with prob 0.95 it contains μ.

Problem 6

(a) Note that

$$\frac{\frac{1}{n}\sum_{i=1}^{n} x_i - \mu}{\mu/\sqrt{n}} \sim N(0,1)$$

So the pivotal for μ is $\sqrt{n}\left(\frac{\bar{x}}{\mu} - 1\right)$.

Remark: Pivotal function for μ is not unique. E.g. $\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{\mu}\right)^2$ is also pivotal (with distribution χ^2_{n-1}).

(b)

$$f(\mu|X_1, X_2, ...X_n) = \left(\frac{1}{\mu\sqrt{2\pi}}\right)^n e^{-\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{2\mu^2}},$$

$$l(\mu|X_1, X_2, ...X_n) = -n\log(\mu) - \frac{n}{2}\log(2\pi) - \sum_{i=1}^{n}\frac{(x_i-\mu)^2}{2\mu^2}.$$

The score function is

$$\frac{\partial l}{\partial \mu} = -\frac{n}{\mu} + \sum_{i=1}^{n}\frac{x_i^2}{\mu^3} - \sum_{i=1}^{n}\frac{x_i}{\mu^2}.$$

Set it to zero, we have

$$n\mu^2 + \mu\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i^2 = 0$$

So MLE $\hat{\mu}$ is the positive root of the equation above.

To derive the form of g, we try to use Taylor expansion in conjunction with asymptotic normality. If there is a function g such that $\sqrt{n}[g(\hat{\mu}) - g(\mu)] \to N(0,1)$, then by Taylor expansion, we have

$$\sqrt{n}g'(\mu)(\hat{\mu} - \mu) \approx \sqrt{n}(g(\hat{\mu}) - g(\mu)) \to N(0,1). \quad (1)$$

On the other hand, by asymptotic normality, we have

$$\sqrt{nI(\mu)}(\hat{\mu} - \mu) \to N(0,1) \quad (2)$$

where

$$\begin{aligned} I(\mu) &= -E\frac{\partial^2 l}{\partial \mu^2} \\ &= -\frac{1}{\mu^2} + \frac{3}{\mu^4}E(x^2) - \frac{2}{\mu^3}E(x) \\ &= -\frac{1}{\mu^2} + \frac{6}{\mu^2} - \frac{2}{\mu^2} \\ &= \frac{3}{\mu^2} \end{aligned}$$

is fisher information.

Compare LHS of (1) and (2) ,we have

$$\frac{dg}{du} = \frac{\sqrt{3}}{\mu},$$

which tells us

$$g(\mu) = \sqrt{3}\log\mu.$$

(c) From (a) we know that $\sqrt{n}(\frac{\bar{X}}{\mu} - 1) \sim N(0,1)$, so the CI with level α is

$$\begin{aligned} 1-\alpha &= P\left(-z_{\alpha/2} < \sqrt{n}(\frac{\bar{X}}{\mu} - 1) < z_{\alpha/2}\right) \\ &= P\left(\frac{\bar{X}^2}{(1+z_{\alpha/2}/\sqrt{n})^2} < \mu^2 < \frac{\bar{X}^2}{(1-z_{\alpha/2}/\sqrt{n})^2}\right). \end{aligned}$$

The corresponding length with level $\alpha = 0.05$ is

$$\bar{X}^2\left(\frac{1}{(1-1.96/\sqrt{n})^2} - \frac{1}{(1+1.96/\sqrt{n})^2}\right). \tag{3}$$

From (b) we know that $\sqrt{3n}\log(\frac{\hat{\mu}}{\mu}) \sim N(0,1)$, so the CI with level α is

$$\begin{aligned} 1-\alpha &= P\left(-z_{\alpha/2} < \sqrt{3n}\log(\frac{\hat{\mu}}{\mu}) < z_{\alpha/2}\right), \\ &= P\left(\hat{\mu}^2 e^{-2z_{\alpha/2}/\sqrt{3n}} < \mu^2 < \hat{\mu}^2 e^{2z_{\alpha/2}/\sqrt{3n}}\right). \end{aligned}$$

The corresponding length with $\alpha = 0.05$ is

$$\hat{\mu}^2\left(e^{2.26/\sqrt{n}} - e^{-2.26/\sqrt{n}}\right). \tag{4}$$

Compare (3) and (4). By consistency, $\bar{X} \to \mu$ and $\hat{\mu} \to \mu$ as $n \to \infty$. Moreover,

$$\frac{1}{(1-1.96/\sqrt{n})^2} - \frac{1}{(1+1.96/\sqrt{n})^2} \approx \frac{7.84}{\sqrt{n}},$$

$$e^{2.26/\sqrt{n}} - e^{-2.26/\sqrt{n}} \approx \frac{4.52}{\sqrt{n}}.$$

So the interval from part (b) has shorter length.