Hanze Yao
3033093286

STAT 201B Homework 3

4. $\delta(X_1, \cdots, X_n) := E(\bar{X}_n \mid X_{(1)}, X_{(n)})$

$T(X) = \{X_{(1)}, X_{(n)}\}$

Now I need to show that $T(X) = \{X_{(1)}, X_{(n)}\}$ is a sufficient statistic for X.

$$P_{(\alpha,\beta)}(X) = \prod_{i=1}^{n} \frac{1}{\beta-\alpha} I(\alpha < X_i < \beta) = \left(\frac{1}{\beta-\alpha}\right)^n \prod_{i=1}^{n} I(\alpha < X_i < \beta)$$

$$= \left(\frac{1}{\beta-\alpha}\right)^n I(X_{(1)} > \alpha)\, I(X_{(n)} < \beta)$$

According to Neyman Factorization Theorem $\left(\frac{1}{\beta-\alpha}\right)^n$ has nothing to do with θ ($\theta = \frac{\alpha+\beta}{2}$), $I(X_{(1)} > \alpha)\, I(X_{(n)} < \beta)$ is $g(T(X), \theta)$, so $T(X) = \{X_{(1)}, X_{(n)}\}$ is a sufficient statistic.

$\therefore$ According to Rao-Blackwell Theorem $\delta(X_1, \cdots, X_n)$ is better than $\bar{X}_n$ because $R(\theta, \delta(X_1, \cdots, X_n)) \le R(\theta, \bar{X}_n)$.

$\delta(X_1, \cdots, X_n) = E(\bar{X}_n \mid X_{(1)}, X_{(n)}) = \frac{1}{n}\sum_{i=1}^{n} E(X_i \mid X_{(1)}, X_{(n)})$

$\because X_1 \cdots X_n$ are iid $\therefore = E(X_i \mid X_{(1)}, X_{(n)})$

$\because X \sim$ Uniform distribution $\therefore$ given $X_{(1)}$ and $X_{(n)}$, X_i takes any value between $X_{(1)}$ and $X_{(n)}$ is the same: $\frac{1}{n}$

$$\therefore E(X_i \mid X_{(1)}, X_{(n)}) = \frac{1}{n}X_{(1)} + \frac{1}{n}X_{(n)} + \frac{n-2}{n} \times \frac{X_{(1)}+X_{(n)}}{2}$$

$$= \frac{X_{(1)}+X_{(n)}}{n} + \frac{(n-2)(X_{(1)}+X_{(n)})}{2n} = \frac{n(X_{(1)}+X_{(n)})}{2n} = \frac{X_{(1)}+X_{(n)}}{2}$$

5. (a)

$$W = \frac{\hat{\theta}_n - \theta}{\hat{se}(\hat{\theta}_n)} = \frac{\hat{\lambda}_n - \lambda_0}{\hat{se}(\hat{\lambda}_n)} = \frac{\bar{X}_n - \lambda_0}{\hat{se}(\bar{X}_n)} = \frac{\bar{X}_n - \lambda_0}{\sqrt{\frac{\hat{\lambda}_n}{n}}}$$

$$= \frac{\bar{X}_n - \lambda_0}{\sqrt{\bar{X}_n / n}}$$

(b) See the R code on next page.

After running this R code, in the total 10000 Wald tests, 495 time rejections. Now I generate Wald test from $H_0: \lambda = \lambda_0$ $\therefore$ These rejections are Type I error. $\therefore$ The rate of type I error is $\frac{495}{10000} = 0.0495$ which is very close to $\alpha = 0.05$

STAT 201B Homework 3 by Hanze Yao

Q5 (b)

```
lambda0 <- 1
n <- 20
alpha <- 0.05
B <- 10000 # Run Wald test 10000 times.
W <- rep(0, B)
for (i in 1:B) {X <- rpois(n, lambda0) # Generate n random variables iid poisson distribution.
  W[i] <- (mean(X) - lambda0) / sqrt(mean(X) / n)
}
num_rejection <- sum(abs(W) > qnorm(1 - alpha / 2))
type_one_error_rate <- num_rejection / B
```

```
> num_rejection
[1] 495
> type_one_error_rate
[1] 0.0495
```

6. LRT : $H_0: p = p_0$ $H_1: p \neq p_0$

MLE $\hat{p} = \frac{X}{n}$ for $Bin(n, p)$

$$\lambda = 2\log T(X) = 2\log\left(\frac{L_n(\hat{p})}{L_n(p_0)}\right) = 2\left(\ell_n(\hat{p}) - \ell_n(p_0)\right)$$

$$= 2\left(\log\binom{n}{x} + X\log\hat{p} + (n-X)\log(1-\hat{p}) - \log\binom{n}{x} - X\log p_0 - (n-x)\log(1-p_0)\right)$$

$$= 2\left(X\log(\hat{p}/p_0) + (n-X)\log(1-\hat{p}/1-p_0)\right)$$

$$= 2X\log\left(\frac{X}{np_0}\right) + 2(n-X)\log\left(\frac{n-X}{n(1-p_0)}\right) \longrightarrow \chi_1^2$$

$\because r=1 \quad q=0 \quad \therefore r-q=1$

$\therefore$ rejection region is $\{\lambda : \lambda > \chi^2_{1,\alpha}\}$

Wald test : $W = \dfrac{\hat{\theta}_n - \theta_0}{\hat{se}(\hat{\theta}_n)} = \dfrac{\hat{p} - p_0}{\hat{se}(\hat{p})}$

$\because$ the MLE of $\hat{p}$ is $\frac{X}{n}$

$\therefore se(\hat{p}) = \sqrt{\hat{p}(1-\hat{p})/n}$

Under null hypothesis : $W = \dfrac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$

$\therefore$ rejection region is $\{|W| > Z_{\frac{\alpha}{2}}\}$

Compare LRT and Wald test

Use Taylor expansion to approximate $\ell_n(p_0)$

$$\ell_n(p_0) \simeq \ell_n(\hat{p}) + \frac{d\,\ell_n(p)}{dp}\bigg|_{p=\hat{p}}(p_0-\hat{p}) + \frac{1}{2!}\frac{d^2\ell_n(p)}{dp^2}\bigg|_{p=\hat{p}}(p_0-\hat{p})^2$$

$\because \frac{X}{n} = \hat{p} \therefore$

$$= \ell_n(\hat{p}) + \underbrace{\frac{X(1-\hat{p})+(X-n)\hat{p}}{\hat{p}(1-\hat{p})}}_{=0}(p_0-\hat{p}) + \frac{-n\hat{p}(1-\hat{p}) + \overbrace{(X-n\hat{p})}^{=0}(1-2\hat{p})}{\hat{p}^2(1-\hat{p})^2} \times \frac{1}{2} \times (p_0-\hat{p})^2$$

$$= \ell_n(\hat{p}) - \frac{n}{2\hat{p}(1-\hat{p})}(p_0-\hat{p})^2$$

$$\therefore \lambda = 2[\ell_n(\hat{p}) - \ell_n(p_0)] \approx \frac{n(p_0-\hat{p})^2}{\hat{p}(1-\hat{p})}$$

$\because W = \dfrac{\hat{p}-p_0}{\sqrt{p_0(1-p_0)/n}} \quad \therefore W^2 = \dfrac{n(\hat{p}-p_0)^2}{p_0(1-p_0)}$

$\because$ For MLE as $n \to \infty$ $\hat{p} \xrightarrow{P} p_0$ under null hypothesis

$\therefore \lambda \xrightarrow{P} W^2 \quad \therefore W^2$ is asymptotically equivalent to λ.

7. For this question, I need to test whether the holiday can postpone people's death.

So define θ as the probability of people die after the holiday.

$H_0: \theta = \frac{1}{2}$ (holiday no effects) $H_1: \theta > \frac{1}{2}$ (holiday has effect, then people die more after holiday $\Rightarrow \theta > \frac{1}{2}$)

Because now consider this as a binomial distribution

$\therefore$ Use Wald test

$$W = \frac{\hat{\theta}_n - \theta_0}{se(\hat{\theta}_n)} = \frac{\frac{X}{n} - \frac{1}{2}}{\sqrt{\frac{1}{2}(1-\frac{1}{2})/n}} \text{ (under null hypothesis)}$$

$\Rightarrow$ should be $se(\theta_0)$

Where X is random variable that represents the number of people die after the holiday.

Then the observed wald test value is:

$$w = \frac{\frac{997}{1919} - \frac{1}{2}}{\sqrt{\frac{1}{2}(1-\frac{1}{2})/1919}} \approx 1.712$$

$\because H_1: \theta > \frac{1}{2}$ but not $\theta \neq \frac{1}{2}$ $\therefore W > 0$ This is a one-side problem

$\therefore$ p-value $= P_{\theta_0=\frac{1}{2}}(|W| > |w|) = P_{\theta_0=\frac{1}{2}}(W > w)$

$\because$ wald test is asymptotically normal

$\therefore = 1 - \Phi(1.712) \approx 0.043$ (Φ is the CDF of $N(0,1)$)

$\because$ 0.043 is the smallest level at which we can reject H_0 with w observed $\therefore$ The null hypothesis can be rejected at level 0.05.

8. $\because X_1 \cdots X_n \overset{iid}{\sim} f(x;\beta) = \beta e^{-\beta x}$ $(x>0, \beta>0)$

$\therefore X_1 \cdots X_n \overset{iid}{\sim} Exp(\beta)$

$\because E(X) = \frac{1}{\beta}$ $\because$ MLE for $E(X)$ is $\bar{X}_n$ $\therefore$ MLE for β is $\frac{1}{\bar{X}_n}$

$$\therefore \lambda = 2\log T(X) = 2\log\frac{L_n(\hat{\beta})}{L_n(\beta_0)} = 2(\ell_n(\hat{\beta}) - \ell_n(\beta_0))$$

$$\because \ell_n(\beta) = \log\left(\prod_{i=1}^{n} \beta e^{-\beta X_i}\right) = n\log\beta - \beta\sum_{i=1}^{n} X_i$$

$$\therefore \lambda = 2\left(-n\log\bar{X}_n - n - n\log\beta_0 + \beta_0\sum_{i=1}^{n} X_i\right)$$

$$= 2n(\beta_0\bar{X}_n - \log\beta_0\bar{X}_n - 1) \longrightarrow \chi^2_1$$

when $\lambda > \chi^2_{1,\alpha}$

Find $\chi^2_{1,\alpha}$ such that $P(\lambda \leq \chi^2_{1,\alpha}) = 1-\alpha$, then reject null hypothesis