BST 676 — Spring 2010 — Dr. Charnigo

Written Assignment 5 Solutions

1a. We have

$$\begin{aligned}
& \sum_{i=1}^{n}(x_i - \bar{x})^2 \\
= & \sum_{i=1}^{n} x_i^2 \\
= & \sum_{i=1}^{n} \left(\frac{2i-1}{n} - 1\right)^2 \\
= & \sum_{i=1}^{n} \left(\frac{(2i-1)^2}{n^2} - \frac{4i-2}{n} + 1\right) \\
= & \sum_{i=1}^{n} \left(\frac{4i^2-4i+1}{n^2} - \frac{4i-2}{n} + 1\right) \\
= & n^{-2}\sum_{i=1}^{n}(4i^2-4i+1) - n^{-1}\sum_{i=1}^{n}(4i-2) + \sum_{i=1}^{n} 1 \\
= & n^{-2}4n(n+1)(2n+1)/6 - n^{-2}4n(n+1)/2 + n^{-2}n - n^{-1}4n(n+1)/2 + n^{-1}2n + n \\
= & n^{-2}\{(4/3)n^3 + 2n^2 + (2/3)n - 2n^2 - 2n + n - 2n^3 - 2n^2 + 2n^2 + n^3\} \\
= & n^{-2}\{(1/3)n^3 - (1/3)n\} \\
= & \frac{n^2-1}{3n}.
\end{aligned}$$

1b. From Written Assignment 4, we know that

$$\hat{\beta} \sim N\left(\beta, 1/\sum_{i=1}^{n}(x_i - x)^2\right) = \beta + N\left(0, 1/\sum_{i=1}^{n}(x_i - x)^2\right)$$

and hence

$$\sqrt{\sum_{i=1}^{n}(x_i - x)^2}\ \hat{\beta} \sim \sqrt{\sum_{i=1}^{n}(x_i - x)^2}\ \beta + N(0,1).$$

Thus,

$$-2\log\lambda = \hat{\beta}^2\sum_{i=1}^{n}(x_i - x)^2 \sim \left(\sqrt{\sum_{i=1}^{n}(x_i - x)^2}\ \beta + N(0,1)\right)^2,$$

which is chi-square on one degree of freedom with noncentrality parameter $\sqrt{\sum_{i=1}^{n}(x_i - x)^2}\ \beta$. Part a affords the further simplification that the noncentrality parameter is $\beta\sqrt{(n^2-1)/(3n)}$. Note that the noncentrality parameter is nonzero if and only if H_1 is true.

Also, if we wish to adhere to a convention that the noncentrality parameter should be nonnegative, then we may take the absolute value, $|\beta|\sqrt{(n^2-1)/(3n)}$. This is justified because $N(0,1)$ and $-N(0,1)$ have the same distribution, hence so do $N(0,1) + a$ and $-N(0,1) + a$, hence so do $(N(0,1) + a)^2$ and $(-N(0,1) + a)^2 = (N(0,1) - a)^2$.

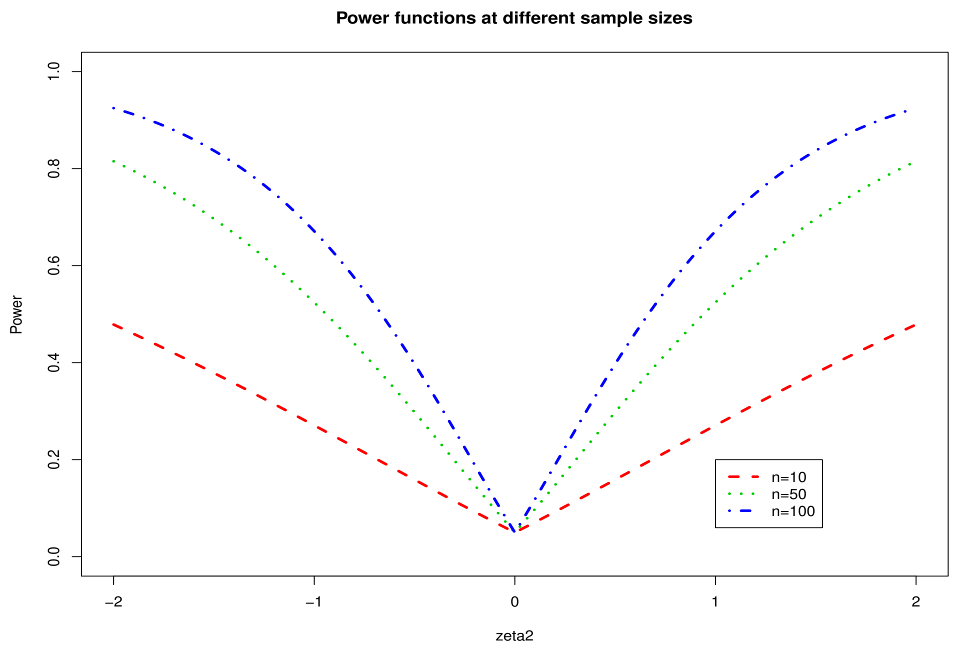
1c. For the likelihood ratio test of $H_0 : \beta = 0$ against $H_1 : \beta \neq 0$ at significance level α, the power function is

$$G(\zeta) = P(C > \chi^2_{1,0.95}),$$

where C has the chi-square distribution on one degree of freedom with noncentrality parameter $|\zeta_2|\sqrt{(n^2-1)/(3n)}$. I used the following R code to generate the graphic below.

```
n1<-10
n2<-50
n3<-100
zeta2<-(-200:200)/100
power1<- 1 - pchisq( qchisq(.95,df=1), df=1, ncp=abs(zeta2)*sqrt((n1^2-1)/(3*n1)))
power2<- 1 - pchisq( qchisq(.95,df=1), df=1, ncp=abs(zeta2)*sqrt((n2^2-1)/(3*n2)))
power3<- 1 - pchisq( qchisq(.95,df=1), df=1, ncp=abs(zeta2)*sqrt((n3^2-1)/(3*n3)))
plot(zeta2,power1,type="n", lwd=3, ylab="Power",ylim=c(0,1))
lines(zeta2,power1,col=2,lty=2,lwd=3)
lines(zeta2,power2,col=3,lty=3,lwd=3)
lines(zeta2,power3,col=4,lty=4,lwd=3)
title("Power functions at different sample sizes")
legend(1,0.2,c("n=10","n=50","n=100"),col=c(2,3,4),lty=c(2,3,4),lwd=c(3,3,3))
```

Figure 1:



1d. Each curve has its lowest point at $\zeta_2 = 0$; this suggests unbiasedness. At any fixed $\zeta_2 \neq 0$, the curves move up toward 1 as the sample size increases; this suggests consistency.

2. We have $f(x;\theta) = a(x)b(\theta)\exp[c(x)d(\theta)]$ with $a(x) = x^{-1}1_{\{x\in(0,1)\}}$, $b(\theta) = \theta$, $c(x) = -\log x$, and $d(\theta) = -\theta$. Put $T := \sum_{i=1}^n c(X_i) = \sum_{i=1}^n -\log X_i$. From Written Assignment 3, we know that T has the gamma distribution with shape n and scale $1/\theta$. So, for $\theta_1 > \theta_2$ and $t \in (0,\infty)$ we have

$$\frac{h(t;\theta_1)}{h(t;\theta_2)} = \frac{\theta_1^n t^{n-1}\exp[-\theta_1 t]/\Gamma[n]}{\theta_2^n t^{n-1}\exp[-\theta_2 t]/\Gamma[n]} = (\theta_1/\theta_2)^n \exp[(\theta_2-\theta_1)t],$$

which is a decreasing function of t. As such, we obtain a level α uniformly most powerful test of $H_0 : \theta \geq \theta_0$ against $H_1 : \theta < \theta_0$ by rejecting H_0 when T is greater than the $1-\alpha$ quantile of its distribution under θ_0. This is $g_{n,1/\theta_0,1-\alpha} = g_{n,1,1-\alpha}/\theta_0$, where $g_{a,b,c}$ denotes the c quantile of the gamma distribution with shape a and scale b.

To see whether this result squares with our intuition, we may note that $E_\theta[X] = \int_0^1 \theta x^\theta\, dx = \theta/(\theta+1)$ is an increasing function of θ. So, a false null hypothesis suggests that the X_i should be small, hence the $-\log X_i$ should be large. Indeed, we end up rejecting H_0 precisely when the sum of the $-\log X_i$ is too large.

3. The score test rejects $H_0 : \theta = 4$ in favor of $H_1 : \theta \neq 4$ when $|\bar{X} - 4| > 1.96\sqrt{4/5} = 1.753$, which occurs when either $\sum_{i=1}^5 X_i \geq 29$ or $\sum_{i=1}^5 X_i \leq 11$. If H_0 is true, then $\sum_{i=1}^5 X_i \sim Pois(20)$ and the actual significance level is

$$\sum_{k=0}^{11}\exp[-20]20^k/k! + \sum_{k=29}^{\infty}\exp[-20]20^k/k! = \sum_{k=0}^{11}\exp[-20]20^k/k! + 1 - \sum_{k=0}^{28}\exp[-20]20^k/k! = 0.05572.$$

Noting that

$$\sum_{k=0}^{10}\exp[-20]20^k/k! + \sum_{k=30}^{\infty}\exp[-20]20^k/k! = \sum_{k=0}^{10}\exp[-20]20^k/k! + 1 - \sum_{k=0}^{29}\exp[-20]20^k/k! = 0.03263$$

and that

$$\exp[-20]20^{11}/11! + \exp[-20]20^{29}/29! = 0.02309,$$

we can make the following modification to force the actual significance level to 0.05:

- Reject H_0 when either $\sum_{i=1}^5 X_i \geq 30$ or $\sum_{i=1}^5 X_i \leq 10$.
- Accept H_0 when $12 \leq \sum_{i=1}^5 X_i \leq 28$.
- When $\sum_{i=1}^n X_i = 11$ or 29, draw a random number from the uniform distribution on $(0,1)$. Reject H_0 if that random number is less than $(0.05 - 0.03263)/(0.05572 - 0.03263) = 0.7523$; otherwise accept H_0.