Unit 3

Team 8

April 18, 2018

Contents

Q67

We have a simple linear regression model: $Y_i = \beta x_i + e_i$, $i = 1, ..., n$ with e_i i.i.d. $N(0, \sigma^2)$.

a) Since the errors are distributed independently according to a normal distribution, we can write:

$$\epsilon_i \sim N(0, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(Y_i - x_i\beta)^2}{2\sigma^2}}$$

Then we have the vector of x_i as data, so the observations Y_i have density functions $Y_i \sim N(x_i\beta, \sigma^2)$. The likelihood function is:

$$L = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(Y_i - x_i\beta)^2}{2\sigma^2}} = (2\pi\sigma^2)^{-\frac{n}{2}} e^{\frac{-1}{2\sigma^2}(y - X\beta)'(y - X\beta)}$$

The log-likelihood function will be:

$$l = -\frac{n}{2} ln 2\pi - \frac{n}{2} ln\sigma^2 - \frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)$$

now we take the partial derivatives and equalize to zero to find the different estimates:

$$\frac{\delta l}{\delta \beta} = \frac{1}{\sigma^2}(y - X\beta)'X = 0$$

$$\hat{\beta} = (X'X)^{-1}X'y$$

Considering a simple linear regression model, $E(\hat{\beta}) = \beta$, β is unbiased. Since it is unbiased, the MSE is equal to the variance:

$$MSE(\hat{\beta}) = var(\hat{\beta}) = E[(\hat{\beta} - \beta)'(\hat{\beta} - \beta)] = \sigma^2(X'X)^{-1}$$

b) MLE for σ^2:

$$\frac{\delta l}{\delta \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}(y - X\beta)'(y - X\beta) = 0$$

$$\hat{\sigma}^2 = \frac{1}{n}(y - X\hat{\beta})'(y - X\hat{\beta}) = \frac{\sum_{i=1}^{n}(Y_i - \hat{\beta}x_i)^2}{n}$$

c) MLE of $\frac{\beta}{\sigma}$:

$$\frac{\hat{\beta}}{\hat{\sigma}} = \frac{(X'X)^{-1}X'y}{\sqrt{\frac{\sum_{i=1}^{n}(Y_i - \hat{\beta}x_i)^2}{n}}}$$

d)

The Fisher information matrix is just the expected value of the negative of the Hessian matrix.

$$\begin{bmatrix} \frac{1}{\sigma^2}\sum_{i=1}^{n} x_i^2 & 0 \\ 0 & \frac{2n}{\sigma^2} \end{bmatrix}$$

e)

$$var(\beta/\sigma) \geq \left(\frac{1}{\sigma}, \frac{-\beta}{\sigma^2}\right)^T \begin{bmatrix} \frac{\sigma^2}{\sum_{i=1}^{n} x_i^2} & 0 \\ 0 & \frac{\sigma^2}{2n} \end{bmatrix} \left(\frac{1}{\sigma}, \frac{-\beta}{\sigma^2}\right) = \frac{1}{\sum_{i=1}^{n} x_i^2} + \frac{\beta^2}{2n\sigma^2}$$

Q78

In the example we are given: $p(x) = p(1-p)^x, \quad x = 0, 1, ...$
$p(H_0) = 0.5 \; p(H_1) = 0.7$

a)

$$\frac{P(H_0|x)}{P(H_1|x)} = \frac{\frac{P(H_0,x)}{P(x)}}{\frac{P(H_1,x)}{P(x)}} = \frac{P(H_0)}{P(H_1)}\frac{P(x|H_0)}{P(x|H_1)} = \frac{P(x|H_0)}{P(x|H_1)} = \frac{0.5}{0.7}\frac{0.5^x}{0.3^x}$$

$$\frac{0.5}{0.7}\frac{0.5^x}{0.3^x} > 1 \implies x \geq 1$$

This means for every $x \geq 1$ we will favor H_0, else H_1.

b)

$$\frac{P(H_0)}{P(H_1)} = 10$$

Substituting this into the formula from question (a) we have:

$$10\ \frac{0.5}{0.7}\frac{0.5^x}{0.3^x} > 1$$

$$\left(\frac{5}{3}\right)^x > \frac{0.7}{5}.$$

This is true, if $x \geq 0$. So we will always favor H_0.

c)

The significance level α is the probability that one makes the type I error. This can be calculated as follows (here is $P = 0.5$ since H_0 is true):

$$\alpha = P(H_1|H_0) = P(X \geq 8|H_0) = 1 - P(X < 8|H_0) =$$

$$= 1 - \sum_{x=0}^{7} p(1-p)^x = 1 - p\frac{1-(1-p)^8}{1-(1-p)} = 0.0039$$

d)

The power of the test is the probability not to do the type two error. It can be calculated as follows (here is $P = 0.7$ since H_1 is true):

$$1 - \beta = 1 - P(H_0|H_1) = 1 - P(x \geq 8|H_1) = 1 - (1 - P(x < 8|H_1)) =$$

$$= \sum_{x=0}^{7} p(1-p)^x = p\frac{1-(1-p)^8}{1-(1-p)} = 1 - 0.3^8 = 0.9999344$$

Q79

We have $X_1 \dots X_n$ i.i.d. $\sim Poisson(\lambda)$

$$H_0 : \lambda = \lambda_0$$
$$H_A : \lambda = \lambda_A$$
$$LR = \frac{f(x_1, \dots, x_n|\lambda_0)}{f(x_1, \dots, x_n|\lambda_A)}$$

We know that the sum of independent Poisson random variables follows a Poisson distribution. Hence:

$$LR = \frac{\frac{\lambda_0^{(x_1+\dots+x_n)} e^{-n\lambda_0}}{x_1! * \dots * x_n!}}{\frac{\lambda_A^{(x_1+\dots+x_n)} e^{-n\lambda_A}}{x_1! * \dots * x_n!}} = \frac{\lambda_0^{(x_1+\dots+x_n)} e^{-\lambda_0 n}}{\lambda_A^{(x_1+\dots+x_n)} e^{-\lambda_A n}}$$

Since $\frac{\lambda_0}{\lambda_A} < 1$, large values of $\sum_{i=1}^n X_i$ corresponds to small values of LR, which in turn favors H_A.

According to the Neyman - Pearson Lemma: suppose that H_0 and H_A are simple hypotheses and that the test that rejects H_0 whenever the likelihood ratio is less than c and significance level α. Then any other test for which the significance level is less than or equal to α has power less than or equal to that of the likelihood ratio test.

So, the likelihood ratio test is the most powerful of level α when likelihood ratio is small.

$$reject\, H_0\, when \sum_{i=1}^n x_i > c,$$

where c is chosen s.t. level $\alpha = P(rejectH_0|H_0) = P(\sum_{i=1}^n x_i > c|\lambda_0)$

Using the Central Limit Theorem (recall: sum of independent Poisson random variables follows $\sim Poisson(n\lambda) => Mean = n\lambda, Var = n\lambda$):

$\alpha = P(\frac{\sum_{i=1}^n x_i - n\lambda_0}{\sqrt{\lambda_0 n}} > \frac{c - n\lambda_0}{\sqrt{\lambda_0 n}}|\lambda_0) = 1 - \Phi(\frac{c-n\lambda_0}{\sqrt{\lambda_0 n}})$

For the large n this $\frac{\sum_{i=1}^n x_i - n\lambda_0}{\sqrt{\lambda_0 n}}$ will be approximated with $\sim N(0,1)$. Thus, $\frac{c-n\lambda_0}{\sqrt{\lambda_0 n}} = Z_{1-\alpha}$, quantile of standard normal distribution.

$=> c = n\lambda_0 + \sqrt{\lambda_0 n} Z_{1-\alpha}$

$$reject\, H_0 \;<=> \sum_{i=1}^{n} x_i > n\lambda_0 + \sqrt{\lambda_0 n} Z_{1-\alpha} \tag{1}$$

$$<=> \bar{x} > \lambda_0 + \sqrt{\frac{\lambda_0}{n}} Z_{1-\alpha} \tag{2}$$

$$<=> \frac{\bar{x} - \lambda_0}{\sqrt{\lambda_0/n}} > Z_{1-\alpha} \tag{3}$$

We got the rejection region: if the Z - statistics $\frac{\bar{x}-\lambda_0}{\sqrt{\lambda_0/n}}$ will be higher than critical value $Z_{1-\alpha}$, we reject H_0.

Test is is said to be uniformly most powerful (UMP), If H_A is composite and this test is the most powerful for every simple alternative H_A.

We can notice that Z - statistics $\frac{\bar{x}-\lambda_0}{\sqrt{\lambda_0/n}}$ is not depend on λ_A, it means that it will be also the most powerful test for any λ_A. Hence, we can write:

$$H_0 : \lambda = \lambda_0; \;\; H_A : \lambda > \lambda_0.$$

And we have that H_A is composite, therefore Likelihood ratio test will be uniformly most powerful test.

Q80

a)

We have $f(x|\theta, \gamma) = \frac{\theta\gamma^\theta}{x^{\theta+1}}$, $x \geq \gamma$ and γ is known.
First let's calculate likelihood:

$$LR = \frac{f_0(x)}{f_1(x)} = \prod_{i=1}^{n} \frac{\frac{\theta_0\gamma^{\theta_0}}{x_i^{\theta_0+1}}}{\frac{\theta_A\gamma^{\theta_A}}{x_i^{\theta_A+1}}}$$

$$= \prod_{i=1}^{n} \frac{\theta_0\gamma^{\theta_0} x_i^{\theta_A+1}}{\theta_A\gamma^{\theta_A} x_i^{\theta_0+1}}$$

$$= \prod_{i=1}^{n} \frac{\theta_0}{\theta_A} \frac{\gamma^{\theta_0-\theta_A}}{x_i^{\theta_0+1-\theta_A-1}}$$

$$= (\frac{\theta_0}{\theta_A})^n \prod_{i=1}^{n} (\frac{x_i}{\gamma})^{\theta_A - \theta_0}$$

The LR is large when the second part ot the expression is large. Let's define:

$$\hat{x} = \prod_{i=1}^{n} (\frac{x_i}{\gamma})$$

Then we can conclude that $\hat{x} < c$ is our condition to reject H_0 where

$$\alpha = P(\hat{X} > c|H_0) = P(\prod_{i=1}^{n} (\frac{\hat{X}_i}{\gamma}) > c|H_0) = P(log \prod_{i=1}^{n} (\frac{\hat{X}_i}{\gamma}) > log(c)|H_0) =$$

$$= P(\sum_{i=1}^{n} log(\frac{\hat{X}_i}{\gamma}) > log(c)|H_0) = (*)$$

Let's find distribution of $log(\frac{\hat{X}_i}{\gamma})$ where $X_i \sim Pareto(\gamma, \theta)$ We can show that $log(\frac{\hat{X}_i}{\gamma})$ is exponentially distributed:

$$P(log(\frac{\hat{X}_i}{\gamma}) \leq x) = P(\hat{X}_i \leq \gamma e^x) = F_{Pareto}(\gamma e^x) = 1 - e^{-\theta x}$$

which is exponential distribution. The sum of n i.i.d. random variables from exponential distribution has gamma distribution with parameters n- shape, θ - rate.

Now we can continue calculations:

$$(*) = P(G < log(c)), G \sim \Gamma(n, \theta_0)$$

Now we can determine c as:

$$c = e^{z_\alpha}$$

,where z_α - α - quantile of this Gamma distribution.

Thus, to find the critical value we are looking for

$$\alpha = P(T \leq c|H_o) = F_{\Gamma(n,\theta_0)}(c)$$

And $c = F^{-1}_{\Gamma(n,\theta_0)}(\alpha)$ but we now need to convert it back to not logarithmized version:

$$log(\frac{X}{\gamma}) \leq F^{-1}_{\Gamma(n,\theta_0)}(\alpha)$$

$$\prod_{1}^{n} x \leq \gamma^n exp(F^{-1}_{\Gamma(n,\theta_0)}(\alpha))$$

And this is the rule for rejection of H_0.

b)

We have that , the LRT is the most powerful in testing simple against simple hypothesys, while, according to N-P, we can apply this to the composite hypotheses via the ratio of supremums of the functions. But, as can be noticed, we had no specific information of θ_0 and θ_A, except for the fact that $\theta_A > \theta_0$, thus one can dedict that the result holds for testing $H_0 : \theta \leq \theta_0$ against $H_A : \theta > \theta_A$.

Q81

By definition :
A necessary and sufficient condition for $T(X_1, X_2, ..., X_n)$ to be a sufficient for a parameter θ is that the joint probability function factors in the form :

$$f(x_1, ..., x_n|\theta) = g[T(x_1, ..., x_n), \theta]h(x_1, ...x_n)$$

The likelihood ratio test for $H_0 : \theta = \theta_0, H_A : \theta = \theta_A$ therefore becomes:

$$L(x) = \frac{f(x|\theta_0)}{f(x|\theta_A)} = \frac{g(T(x), \theta_0)}{g(T(x), \theta_A)}$$

Which, of course is a function of T; and the maximum likelihood estimate is found by maximizing $g[T(x_1, ..., x_n), \theta_0]$.

Now, the rejection region is when: $L(T(x)) < c$, for a given constant c. Since it depends on $T(x)$, we can rewrite it as: $T(x) < c_0$ fro a given constant c_0 such that also the previous inequality is verified.
The test level is defined as :

$$\mathbb{P}\left(L(T(x)) < c|H_0\right) = \mathbb{P}\left(T(x) < c_0|H_0\right) = \alpha$$

Define the distribution of T under H_0 as $F(x)$ and its quantile function F^{-1}. We can rewrite the above equation as follow:

$$\mathbb{P}\left(F(x) < c_0\right) = \alpha$$

Using the quantile function we finally have:

$$c_0 = F^{-1}(\alpha)$$

One remark is that a sufficient statistic does not depend on θ, therefore if the distribution of T is known under H_0 it will be known also under H_A. That's why we can reduce the likelihood ratio to a function known under H_0.

Q82:

Part a:

We have $X \sim N(0, \sigma^2)$, $H_0 : \sigma = \sigma_0$ and $H_A : \sigma = \sigma_A$, where $\sigma_A > \sigma_0$. And we know that we need to make Likelihood ratio small.

$$\begin{aligned} LR &= \frac{\frac{1}{\sigma_0\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma_0^2}}}{\frac{1}{\sigma_A\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma_A^2}}} \longrightarrow small & (4) \\ &= \frac{\sigma_A}{\sigma_0} e^{-\frac{x^2}{2\sigma_0^2}+\frac{x^2}{2\sigma_A^2}} & (5) \\ &=> -\frac{x^2}{2\sigma_0^2} + \frac{x^2}{2\sigma_A^2} \longrightarrow small & (6) \\ &=> x^2 \longrightarrow large & (7) \\ & & (8) \end{aligned}$$

Hence, $\alpha = P(rejectH_0|H_0) = P(X^2 > c|\sigma_0)$.

We need to standardize:

$\alpha = P(X^2/\sigma_0^2 > c/\sigma_0^2)$

$=> 1 - \alpha = P(X^2/\sigma_0^2 \leq c/\sigma_0^2)$

We know that $X/\sigma \sim N(0, 1)$, so $X^2/\sigma^2 \sim \chi_1^2$ where 1 is degree of freedom.

Hence $c/\sigma_0^2 = F_{1-\alpha}^{-1}$ (quantile of Chi - square distribution).

$=> c = F_{1-\alpha}^{-1}\sigma_0^2$

$X^2 > F_{1-\alpha}^{-1}\sigma_0^2$ this is the rejection region for H_0 for given α.

Part b:

Now we have $X_1 \dots X_n$ i.i.d. $\sim N(0, \sigma^2)$. We will have the same case like in Part a, but in stead of X^2, we will have $\sum_{i=1}^{n} X^2$. Just shortly we wil show it:

$$LR = \frac{\frac{1}{(\sigma_0\sqrt{2\pi})^n}\ e^{-\sum_{i=1}^{n} X_i^2/2\sigma_0^2}}{\frac{1}{(\sigma_A\sqrt{2\pi})^n}\ e^{-\sum_{i=1}^{n} X_i^2/2\sigma_A^2}} \longrightarrow small$$

$$(\frac{\sigma_A}{\sigma_0})^n e^{\frac{1}{2}\sum_{i=1}^{n} X_i^2(1/\sigma_A^2 - 1/\sigma_0^2)}$$

$=> \sum_{i=1}^{n} X_i^2 \longrightarrow large$ because $(1/\sigma_A^2 - 1/\sigma_0^2)$ is negative.

Hence, $\alpha = P(rejectH_0|H_0) = P(\sum_{i=1}^{n} X_i^2 > c|\sigma_0)$.

$\alpha = P(\sum_{i=1}^{n} X_i^2/\sigma_0^2 > c/\sigma_0^2)$
$=> 1 - \alpha = P(\sum_{i=1}^{n} X_i^2/\sigma_0^2 \leq c/\sigma_0^2)$

$=> c/\sigma_0^2 = F_{1-\alpha}^{-1}$ (quantile of Chi - square distribution with n degree of freedom).

Basically, the main difference between this part and part A is that in part A we had degree of freedom 1 (because we had one X) and now we have degree of freedom n (number of i.i.d. X_i). $=> c = F_{1-\alpha}^{-1}\sigma_0^2$

$\sum_{i=1}^{n} X^2 > F_{1-\alpha}^{-1}\sigma_0^2$ this is the rejection region for H_0 for given α.

Part c:

Yes, in this case the Likelihood ratio rest will be uniformly the most powerful because this test does not depend on σ_A, so basically this test will work for any σ_A. In this case we have hypothesis:

$$H_0 : \sigma = \sigma_0; \ \ H_A : \sigma > \sigma_0,$$

where H_A is composite, therefore Likelihood ratio test will be uniformly most powerful test.

Q83:

Part a:

We have $X \sim U(0, \theta)$ and

$$H_0 : \theta = 1; \ \ H_A : \theta = 2$$

$$\alpha = P(rejectH_0|H_0) = P(rejectH_0|\theta = 1) = 0$$

So, we need to find the situation which will have probability zero. When H_0 is true, we have that $X \sim U(0,1)$. It means that if we test $X > 1$ the probability of this will be zero. Hence $P(X > 1|\theta = 1) = 0$. The power of this test is $1 - \beta$:

$$\beta = P(AcceptH_0|H_A) = P(X \leq 1|\theta = 2) = 0.5$$

$$=> 1 - \beta = 0.5$$

Part b:

Now we have the test:

$$\alpha = P(rejectH_0|H_0) = P(X \leq c|\theta = 1) = \frac{c-0}{\theta-0} = c$$

where $0 < c < 1$. Remark: $X \sim U(0,\theta) \implies F_X(x) = \frac{x-0}{\theta-0}$.

Power of this test:

$$1-\beta = 1-P(AcceptH_0|H_A) = 1-P(X > c|\theta = 2) = 1-1+P(X \leq c|\theta = 2) = \frac{c-0}{2-0} = \frac{c}{2}$$

Part c:

$$\alpha = P(rejectH_0|H_0) = P(1-c \leq X \leq c|\theta = 1) = F_{X\sim U}(1) - F_{X\sim U}(1-c) =$$
$$= \frac{1-0}{1-0} - \frac{1-c-0}{1-0} = c$$
$$1 - \beta = 1 - P(AcceptH_0|H_A) = 1 - P(X \leq 1 - c \cup X \geq 1|\theta = 2) =$$
$$= 1 - (F(1-c) + 1 - F(1)) = 1 - (1-c)/2 - 1 + 1/2 = c/2$$

Part d:

Let's find the likelihood ratio for Uniform distribution:

$$LR = \frac{f(x|\theta_0)}{f(x|\theta_A)} = \frac{1I_{x\in[0,1]}}{1/2I_{x\in[0,2]}}$$

This gives us:
1) LR = 2, when $x \in [0,1]$

2) LR $= 0$, when $x \in (1, 2]$
3) LR is undefined, if else

Hence, we conclude that Likelihood ratio will not give us the unique rejection region!

Part e:

Suppose hypothesis interchanged.

$$LR = \frac{f(x|\theta_0)}{f(x|\theta_A)} = \frac{1/2 I_{x\in[0,2]}}{1 I_{x\in[0,1]}}$$

This will give us the following:
1) LR $= 1/2$, when $x \in [0, 1]$
2) LR is undefined, if else!

Given $0 < \alpha < 1/2$, a likelihood ratio test is: we reject H_0 if and only if $X < 2\alpha$ since

$$P(0 < X < 2\alpha|H_0) = 2\alpha(1/2) = \alpha$$

This test have power:

$$P(0 < X < 2\alpha|H_1) = 2\alpha$$

We again see that Likelihood ratio does not determine the unique rejection region for H_0. Another likelihood ratio test with the same significance level and power is: Reject H_0 iff $1 - 2\alpha < X < 1$

Q84

a

The uniform distribution on $[0, 1]$ has the probability density function:

$$f(x) = \begin{cases} 1 & if \quad x \in [0,1] \\ 0 & else \end{cases}$$

Substituting $\theta = 1$, we get:

$$f(x|1) = 1x^{1-1} = 1, \quad 0 \le x \le 1$$

which is equivalent to the uniform distribution. So, the uniform distribution on $[0, 1]$ is a special case of the Beta distribution with parameters 1 and 1.

b

We want to test $H_0 : \theta = 1$ and $H_A : \theta \neq 1$. The likelihood of our density is:

$$L(\theta, x) = \theta^n (\prod_{i=1}^{n} x_i)^{\theta-1}$$

And the MLE is:

$$\bar{\theta} = -\frac{n}{\sum_{i=1}^{n} lnx_i}$$

The GLR is then:

$$\Lambda = \frac{L(\bar{\theta}, x)}{L(1, x)}$$

$$L(1, x) = 1$$

$$\Lambda = \left(-\frac{n}{\sum_{i=1}^{n} ln(x_i)}\right)^n \prod_{i=1}^{n} x_i^{-\frac{n}{\sum_{i=1}^{n} ln(x_i)}-1}$$

$$= \left(-\frac{n}{\sum_{i=1}^{n} ln(x_i)}\right)^n e^{-\sum_{i=1}^{n} ln(x_i)(n/\sum_{i=1}^{n} ln(x_i)+1)}$$

$$= \frac{n}{e}^n \left(-\frac{n}{\sum_{i=1}^{n} ln(x_i)}\right)^{-n} e^{-\sum_{i=1}^{n} ln(x_i)}$$

The test will reject H_0 iff $\Lambda \geq C$ or in terms of statistics $T(\Lambda) = -2\sum_{i=1}^{n} ln(x_i) \overset{H_0}{\sim} \chi^2_{2n}$ iff $T(\Lambda) \leq C_1$ or $T(\Lambda) \geq C_2$, where:

$$\chi^2_{2_n}(C_2) - \chi^2_{2_n}(C_1) = 1 - \alpha, \quad C_2 - C_1 = 2nln\left(\frac{C_2}{C_1}\right)$$

There are no closed solution to that, but we can commonly used quantiles $\chi^2_{2n;1-\alpha/2}$ and $\chi^2_{2n;\alpha/2}$. The GLRT will not reject the null hypothesis of the uniform distribution iff:

$$\chi^2_{2n;1-\alpha/2} \leq -2\sum_{i=1}^{n} ln(x_i) \leq \chi^2_{2n;\alpha/2}$$

.