STAT 415 — Final (Spring 2020)

Name: Solutions

Exam rules:

- You have ?? to complete the exam, scan it and upload onto Ecampus.
- There are **?** Questions.
- This exam is open book. You are free to use your class notes and HW solutions to solve the problems.

  Do not do a brain dump. Answers which are irrelevant will be penalized.

  Do not blindly copy answers from the HW solutions.
- State precisely in the derivations all the results that you use.
- Explain your arguments carefully, show all your work, and write out all your steps. Credit will be given for clarity. No credit will be given for a solution without working.
- If you are caught cheating or helping someone to cheat on this exam, you will both receive a grade of zero on the exam and you will be turned in.
- If a problem requires a numerical answer, you may express answer in terms of elementary functions or cdfs (e.g. $\Gamma(r+1)$ or $r!$).
- Presentation and scanning rules (20% will be knocked off if these rules are not adhered too).
  1. No pencils. Black pen is easiest to read.
  2. Write big; scrawls are painful to read.
  3. Name at top of exam paper (this makes it easier to identify them).
  4. Name the submitted file LASTNAME_FIRSTNAME.pdf
  5. Submit only **one** pdf file.
  6. The solutions should be on separate pages.
  7. Check clarity of scan before submission.

(1) Suppose the random variables $\underline{X} = (X_1, X_2, X_3, X_4, X_5)$ have the covariance matrix

$$\begin{pmatrix} 1 & 1/2 & 0 & 0 & 0 \\ 1/2 & 1 & 1/2 & 0 & 0 \\ 0 & 1/2 & 1 & 1/2 & 0 \\ 0 & 0 & 1/2 & 1 & 1/2 \\ 0 & 0 & 0 & 1/2 & 1 \end{pmatrix}$$

Obtain its covariance network. [3]

$$X_1 \text{ --- } X_2 \text{ --- } X_3 \text{ --- } X_4 \text{ --- } X_5$$

(2) Suppose $\{(X_t, Y_t)\}_{t=1}^n$ are independent bivariate random variables (meaning that over t they are independent), where

$$X_t = 2t + \varepsilon_t \qquad Y_t = \cos\left(\frac{2\pi t}{3}\right) + e_t$$

and $\{\varepsilon_t, e_t\}$ are random with $\mathrm{E}[\varepsilon_t] = \mathrm{E}[e_t] = 0$, $\mathrm{var}[\varepsilon_t] = \mathrm{var}[e_t] = \sigma^2$ and $\delta = \mathrm{cov}(\varepsilon_t, e_t)$. Let

$$S_n = \frac{1}{n}\sum_{t=1}^{n}(X_t - Y_t).$$

Obtain $\mathrm{E}[S_n]$ and $\mathrm{var}[S_n]$. [5]

(a) $$E[S_n] = \frac{1}{n}\sum_{t=1}^{n}\left\{E(X_t) - E(Y_t)\right\}$$

$$= \frac{1}{n}\sum_{t=1}^{n}\left\{2t - \cos\left(\frac{2\pi t}{3}\right)\right\}$$

(b) $$\mathrm{var}[S_n] = \frac{1}{n^2}\sum_{t=1}^{n}\left\{\mathrm{var}(X_t - Y_t)\right\}$$

$$= \frac{1}{n^2}\sum_{t=1}^{n}\left\{\mathrm{var}(X_t) - 2\mathrm{cov}(X_t, Y_t) + \mathrm{var}(Y_t)\right\}$$

$$= \frac{2\sigma^2 - 2\rho}{n}$$

(3) Suppose that $\{X_i\}$ are iid random variables with mean μ ($\mu \neq 0$) and variance σ^2. Let $\bar{X}$ denote the sample mean.

(i) Obtain the asymptotic distribution of $\log \bar{X}$.

(ii) Based on (ii) construct an approximate 95% confidence interval for $\log \mu$. The confidence interval should be computable, where unknown parameters are replaced with their estimators.

[5]

- By Theorem 1.1 $\quad \sqrt{n}(\bar{x} - \mu) \xrightarrow{D} N(0, \sigma^2)$
- By Lemma 1.1 $\quad \sqrt{n}[g(\bar{x}) - g(\mu)] \xrightarrow{D} N(0, g'(\mu)^2 \sigma^2)$

$g(\bar{x}) = \log \bar{x} \qquad g'(\mu) = \frac{1}{\mu}$

$\Rightarrow$ If $\mu \neq 0$

$$\sqrt{n}[\log \bar{x} - \log \mu] \xrightarrow{D} N\left(0, \frac{1}{\mu^2} \times \sigma^2\right)$$

The asymptotic standard error of $\log \bar{x}$ is

$$\sqrt{\frac{\sigma^2}{n \mu^2}}$$

To estimate the standard error we use

$\sigma^2 \mapsto s_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$

$\mu^2 \mapsto \bar{x}^2$. Thus approximate 95% CI for $\log \mu$ is

$$\left[\log \bar{x} \pm 1.96 \times \sqrt{\frac{s^2}{n \bar{x}^2}}\right]$$

4 Suppose that $\{X_i\}_{i=1}^n$ are iid Binomial random variables where $X_i \sim Bin(m,p)$.

(i) If $n = 3$ and $m = 10$, obtain the rejection region for the log-likelihood ratio test at the 5% level for

$$H_0 : p = 0.6 \text{ vs } H_A : p > 0.6. \quad (0.8)$$

Give all working (you may use R). → Explain why this test should be used for detection $p=0.8$

(ii) Calculate the power of the test for $p = 0.8$.

(iii) You observe the data $(x_1, x_2, x_3) = (8, 6, 7)$, based on the data what are the conclusions of the test at the 5% level.

[12]

(iv) Bonus question. Marks will only be given if you get full marks in the other parts of this question, this is only for those who want a small challange. If you do not do it you will not be penalized.

Obtain an alternative rejection region for (X_1, X_2, X_3), where the probability of a Type I error is 5% or less. Obtain the power for this new test when $p = 0.8$.

The likelihood is

$$L_n(p) = \prod_{i=1}^{n} \binom{m}{x_i} p^{x_i} (1-p)^{m-x_i}$$

The log-likelihood under the null is

$$\mathcal{L}_n(p_0) = \sum_{i=1}^{n} \left\{ \log \binom{n}{x_i} + x_i \log p + (m-x_i) \log(1-p) \right\}$$

The log-likelihood under the alternative is

$$\mathcal{L}_n(p_1) = \sum_{i=1}^{n} \left\{ \log \binom{n}{x_i} + x_i \log p_1 + (m-x_i) \log(1-p_1) \right\}$$

$$\log LR(\underline{X}) = \mathcal{L}_n(p_1) - \mathcal{L}_n(p_0)$$

$$= \sum_{i=1}^{n} x_i \log \frac{p_1}{p_0} + (m-x_i) \log \frac{(1-p_1)}{(1-p_0)}$$

$$= \sum_{i=1}^{n} x_i \left\{ \log \frac{p_1}{1-p_1} - \log \frac{p_0}{1-p_0} \right\} + mn \log\left(\frac{1-p_1}{1-p_0}\right) \Big\}$$

$$P\left\{ \log LR(\underline{x}) \geq k \;\middle|\; H_0 \right\}$$

$$= P\left\{ \sum_{i=1}^{n} x_i \cdot \underbrace{\left(\log \frac{p_1}{1-p_1} - \log \frac{p_0}{1-p_0} \right)}_{\gamma} \geq k - mn \log\left(\frac{1-p_1}{1-p_0}\right) \;\middle|\; H_0 \right\} \qquad \gamma > 0$$

$$= P\left\{ \sum_{i=1}^{n} x_i \geq T \;\middle|\; H_0 \right\}$$

Since $x_i \sim Bin(m, p_0)$ (under null)

$\sum_{i=1}^{n} x_i \sim Bin(mn, p_0)$ (under null)

$p_0 = 0.6$ $\quad n \times m = 30$. Using R the rejection region is $C_{0.05} = \{23, 24, \ldots, 30\}$

$$P\left\{ \sum x_i \in C_{0.05} \;\middle|\; p = 0.6 \right\} = 0.43$$

(ii) The power of the test for $p = 0.8$ is

$$\text{sum}\{\text{dbinom}(c(23:30), 30, 0.8)\} = 76\%$$

(*) There exist no other test (rejection region of (X_1, X_2, X_3)) with type 1 error $= 5\%$, with power greater than 76%.

(iii) Since $8+6+7 = 21 \notin C_{0.05}$. We cannot reject the null at the 5% level. There is no evidence in the data to suggest $p > 0.6$.

5 Suppose that $\{X_i\}_{i=1}^n$ are iid normal random variables with mean μ and variance σ^2. We will assume in this question that μ is known.

You may use that if $Z \sim N(0, \sigma^2)$, then $\text{var}[Z] = 2\sigma^4$.

(i) Obtain the maximum likelihood estimator of σ^2. Denote this as $\widehat{\sigma}^2_{MLE}$

(ii) Derive the exact distribution of $\widehat{\sigma}^2_{MLE}$.

(iii) Derive the asymptotic distribution of $\widehat{\sigma}^2_{MLE}$.

(iv) Derive the generalized log-likelihood ratio statistic for

$$H_0 : \sigma^2 = \sigma_0^2 \text{ vs } H_A : \sigma^2 \neq \sigma_0^2.$$

(v) Derive the (asymptotic) rejection region for the test statistic at the 5% level.

[15]

(i) The likelihood is

$$\mathcal{L}_n(\sigma^2) = -\frac{n}{2}\log \sigma^2 - \frac{1}{2n\sigma^2}\sum_{i=1}^n (x_i - \mu)^2$$

$$\frac{\partial \mathcal{L}_n}{\partial \sigma^2} = -\frac{n}{2\sigma^2} - \frac{1}{2n\sigma^4}\sum_{i=1}^n (x_i - \mu)^4 = 0$$

MLE $\quad \hat{\sigma}^2_n = \frac{1}{n}\sum_{i=1}^n (x_i - \mu)^2$

(ii) Since $\left(\frac{x_i - \mu}{\sigma}\right)^2 \sim \chi^2_1$ and $\{x_i\}$ are iid, then $\sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma}\right)^2 \rightarrow \chi^2_n$

$$\Rightarrow \quad s_n^2 \sim \frac{\sigma^2}{n}\chi^2_n$$

(iii) Since $Y_j = (X_i - \mu)^2$ are iid rvs.

with $E(Y_j) = \sigma^2$ and $var(Y_j) = 2\sigma^4$.

Then by the CLT

$$\sqrt{n}\left[S_n - \sigma^2\right] \xrightarrow{D} N(0, 2\sigma^4)$$

(iv) Under the null

$$\mathcal{L}_n(\sigma_0^2) = -\frac{n}{2}\log \sigma_0^2 - \frac{1}{2\sigma_0^2}\sum_{i=1}^{n}(X_i-\mu)^2$$

$$= \frac{n}{2}\log \sigma_0^2 - \frac{n}{2} \times \frac{\hat{\sigma}_n^2}{\sigma_0^2}$$

Under the alternative

$$\mathcal{L}_n(\hat{\sigma}_n^2) = -\frac{n}{2}\log \hat{\sigma}_n^2 - \frac{1}{2\hat{\sigma}_n^2}\sum_{i=1}^{n}(X_i-\mu)^2$$

$$= -\frac{n}{2}\log \hat{\sigma}_n^2 - \frac{n}{2\hat{\sigma}_n^2} \times \hat{\sigma}_n^2$$

$$= -\frac{n}{2}\log \hat{\sigma}_n^2 - \frac{n}{2}$$

Thus

$$\log \Lambda(\underline{x}) = \mathcal{L}_n(\hat{\sigma}_n^2) - \mathcal{L}_n(\sigma_0^2)$$

$$= -\frac{n}{2}\log\frac{\hat{\sigma}_n^2}{\sigma_0^2} + \frac{n}{2}\frac{\hat{\sigma}_n^2}{\sigma_0^2} - \frac{n}{2}$$

$$= \frac{n}{2}\left\{\frac{\hat{\sigma}_n^2}{\sigma_0^2} - \log\frac{\hat{\sigma}_n^2}{\sigma_0^2} - 1\right\}$$

$$2\log\Lambda(\underline{x}) = n\left\{\frac{\hat{\sigma}_n^2}{\sigma_0^2} - \log\frac{\hat{\sigma}_n^2}{\sigma_0^2} - 1\right\}$$

(v) Under the null hypothesis

$$2\log\Lambda(\underline{x}) \xrightarrow{D} \chi^2_1$$

Thus for a large n, we reject the null at the 5%-level if

$$n\left\{\frac{\hat{\sigma}_n^2}{\sigma_0^2} - \log\frac{\hat{\sigma}_n^2}{\sigma_0^2} - 1\right\} \geq 3.84$$

$\chi^2_{1-0.05}$

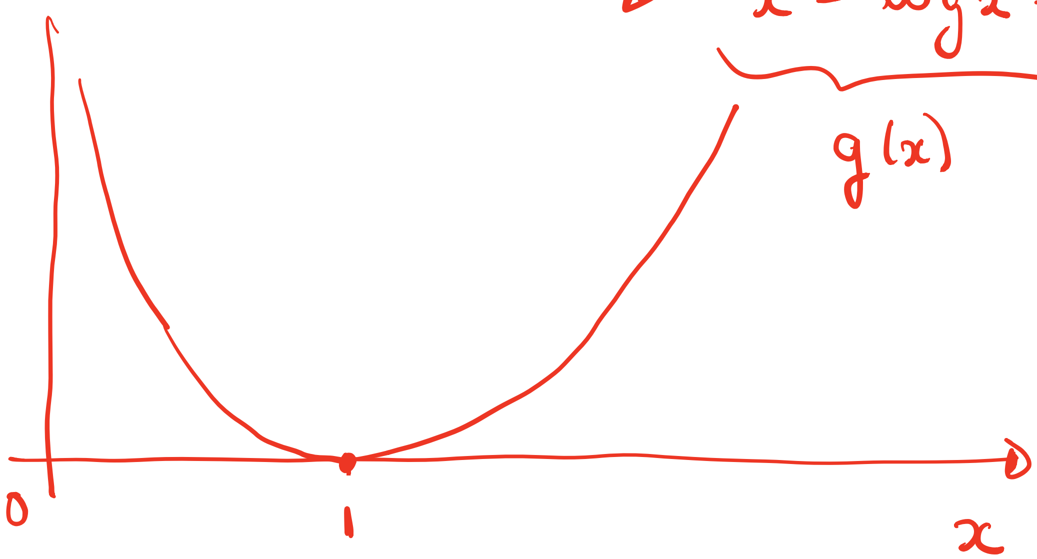
Bonus Question

Take a look at

$$n\left\{\frac{\hat{\sigma}_n^2}{\sigma_0^2} - \log\frac{\hat{\sigma}_n^2}{\sigma_0^2} - 1\right\}$$

$\underbrace{x - \log x - 1}_{g(x)}$, with $x = \dfrac{\hat{\sigma}_n^2}{\sigma_0^2}$

(1)

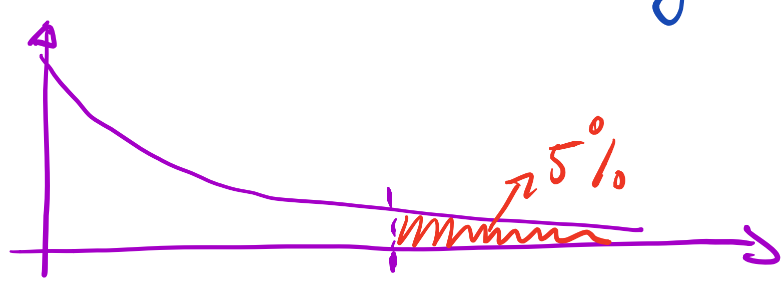


Thus

$$\left\{\frac{\hat{\sigma}_n^2}{\sigma_0^2} - \log\frac{\hat{\sigma}_n^2}{\sigma_0^2} - 1\right\}$$ is smallest

when $\hat{\sigma}_n^2 \approx \sigma_0^2$; which is true when H_0 is true. Thus the distribution of

$2\log\Lambda(\underline{X})$



3.84

However, if H_A is true we see from (1) that $\frac{2}{n} \log \Lambda(\underline{x})$ will be not be close to zero. Thus $2 \log \Lambda(\underline{x})$ will be "large". This means under the alternative