STAT 415 — Midterm 2 (Spring 2021)

Name: Solutions

Exam rules:

- You have 90 minutes to complete the exam.
- There are **4** Questions.
- You may use the provided formula sheet (this is a closed book exam).
- You are only allowed to use a pen and pencil (or ipad for writing). You cannot use the internet to find anwers and you cannot confer with class mates.
- 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.

1. Define the bivariate random vector

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}\right).$$

Let $Y_1 = X_1 - X_2$ and $Y_2 = X_1 + X_2$.

Obtain (necessary and sufficient) conditions that ensure that Y_1 and Y_2 are independent random variables. [2]

We use the following two results;

(a) If (X_1, X_2) are jointly normal, then any linear combination is normal.

(b) If (Y_1, Y_2) are jointly normal, then $\text{cov}[Y_1, Y_2] = 0$ if and only if Y_1 and Y_2 are independent.

From (a) $Y_1 = (X_1 - X_2)$ and $Y_2 = (X_1 + X_2)$ are jointly normal.

To find necessary and sufficient conditions for independence we evaluate $\text{cov}[Y_1, Y_2]$:

$$\text{cov}[Y_1, Y_2] = \text{cov}[X_1 - X_2, X_1 + X_2] = \text{var}(X_1) - \text{var}(X_2)$$
$$= \sigma_1^2 - \sigma_2^2.$$

$\text{cov}[Y_1, Y_2] = 0$ (and thus independent) iff $\sigma_1^2 = \sigma_2^2$.

(the means μ_1 and μ_2 play no role).

2. Suppose that $\{X_t\}_{t=1}^n$ are iid (independent, identically distributed) random variables with $\mathrm{E}[X_t] = \mu$ and $\mathrm{var}[X_t] = \sigma^2$. Suppose that U and $\{X_t\}_{t=1}^n$ are independent of each other. At the time points $t = 2^k$ $(k = 1, 2, 3 \ldots)$ X_t is corrupted with the noise U where $\mathrm{E}[U] = 0$ and $\mathrm{var}[U] = \sigma_U^2$. We observe the random variables $\{Y_t\}$ where

$$Y_t = \begin{cases} X_t & \text{if } t \neq 2^k \text{ where } k = 1, 2, 3, \ldots \\ X_t + U & \text{if } t = 2^k. \end{cases}$$

Let $\bar{Y} = n^{-1} \sum_{t=1}^n Y_t$.

(i) Show that $\bar{Y}$ is an unbiased estimator of μ.

(ii) Obtain the Mean Squared error of $\bar{Y}$, $\mathrm{E}[(\bar{Y} - \mu)^2]$.

(iii) Is $\bar{Y}$ an asymptotically consistent estimator of μ? Prove your claim (yes or no is not enough). [4]

We define some notation. Let $\lfloor a \rfloor$ denote the largest integer less than or equal to a.

Let $\log x$ denote log to the base 2.

Then we split the sum over Y into its respective X and U components.

$$\bar{Y} = \frac{1}{n}\sum_{t=1}^{n} Y_t = \frac{1}{n}\sum_{t=1}^{n} X_t + \frac{1}{n}\sum_{k=1}^{\lfloor \log n \rfloor} Y$$

(since $2^{\lfloor \log n \rfloor}$ = largest 2^k less than n).

$$= \bar{X} + \frac{1}{n}\sum_{k=1}^{\lfloor \log n \rfloor} Y .$$

(i) $$E[\bar{Y}] = \underbrace{E[\bar{X}]}_{\mu} + \underbrace{E\left[\frac{1}{n}\sum_{k=1}^{\lfloor \log n \rfloor} Y\right]}_{= \frac{1}{n}\sum_{k=1}^{\lfloor \log n \rfloor} E(Y)}$$

$$= \frac{1}{n} \cdot 0 = 0$$

$$= \mu .$$

Thus $\overline{Y}$ is an unbiased estimator of μ.

(ii) $$E\left[(\overline{Y} - \mu)^2\right] = \mathrm{var}\left[\overline{Y}\right]$$ (since $\overline{Y}$ is unbiased)

$$= \mathrm{var}\left[\overline{X} + \frac{1}{n}\sum_{k=1}^{\lfloor \log n \rfloor} u\right]$$

$$= \mathrm{var}\left[\overline{X}\right] + \mathrm{var}\left[\frac{1}{n}\sum_{k=1}^{\lfloor \log n \rfloor} u\right]$$ (by independence)

$$= \underbrace{\frac{\sigma_x^2}{n}}_{\text{By iid } X_i\text{s}} + \frac{1}{n^2}\sum_{k_1,k_2=1}^{\lfloor \log n \rfloor} \mathrm{cov}[u, u]$$

$$= \frac{\sigma_x^2}{n} + \frac{1}{n^2}\sum_{k_1,k_2=1}^{\lfloor \log n \rfloor} \mathrm{var}(u)$$

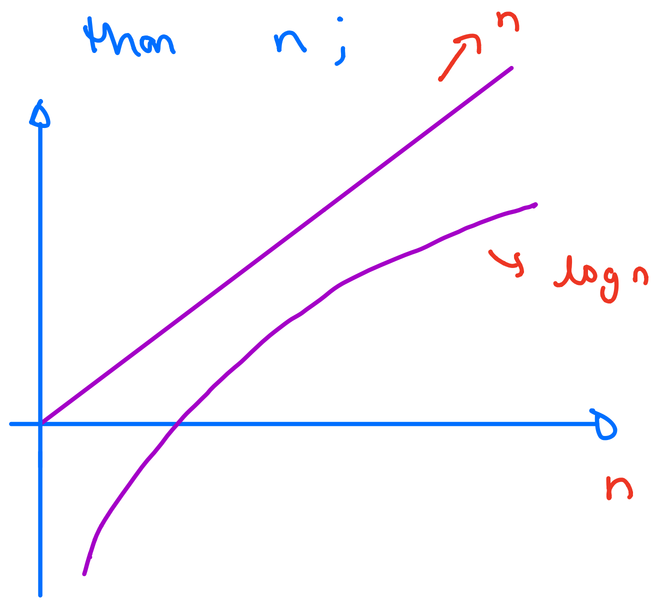
$$= \frac{\sigma_x^2}{n} + \frac{(\lfloor \log n \rfloor)^2}{n^2}\sigma_u^2$$

(iii) $$E\left[(\overline{Y} - \mu)^2\right] = \frac{\sigma_x^2}{n} + \left(\frac{\lfloor \log n \rfloor}{n}\right)^2 \sigma_u^2$$

- $\dfrac{\sigma_x^2}{n} \longrightarrow 0$ as $n \to \infty$

- $\dfrac{\lfloor \log n \rfloor}{n} \leq \dfrac{\log n}{n} \longrightarrow 0$ as $n \to \infty$

(It is well known that $\log n \to \infty$ but at a very slow pace, far slower than n;



Thus $E\left[(\bar{Y} - \mu)^2\right] \longrightarrow 0$ as $n \to \infty$.

Hence $\bar{Y}$ is asymptotically consistent.

3. Suppose that $\{X_i\}_{i=1}^n$ are iid random variables with the exponential distribution

$$f(x;\theta) = \frac{1}{\theta}\exp(-x/\theta), \quad x \geq 0.$$

You can use that $\mathrm{E}[X_i] = \theta$.

(i) Obtain the maximum likelihood estimator of θ and the Fisher information $I(\theta)$.

(ii) For a fixed sample size n, make a sketch of the log-likelihood function corresponding to a large Fisher information and a small Fisher information (and the corresponding θ). Indicate on your plots which log-likelihood will have an MLE with a small variance and which will have a large variance. [4]

(i) The likelihood is

$$L_n(\theta) = \prod_{i=1}^{n} \frac{1}{\theta} \exp\left\{-\frac{X_i}{\theta}\right\}$$

$$= \frac{1}{\theta^n} \exp\left\{-\frac{1}{\theta}\sum_{i=1}^{n} X_i\right\}$$

$$= \frac{1}{\theta^n} \exp\left\{-\frac{1}{\theta} \cdot n\bar{X}\right\}$$

The log-likelihood is

$$\mathcal{L}_n(\theta) = -n\log\theta - \frac{1}{\theta} n\bar{X}$$

$$\frac{\partial \mathcal{L}_n(\theta)}{\partial\theta} = -\frac{n}{\theta} + \frac{1}{\theta^2} n\bar{X}$$

The MLE solves $\dfrac{\partial \mathcal{L}_n}{\partial\theta} = 0.$

$$-\frac{n}{\theta} + \frac{1}{\theta^2} n \bar{X} = 0 \quad \text{when}$$

$$\hat{\theta} = \bar{X} \qquad \text{Hence} \quad \text{MLE} =$$

$$\hat{\theta}_{MLE} = \bar{X}.$$

(ii) The Fisher Information is

$$I(\theta) = -E\left[\frac{\partial^2 \log f}{\partial \theta^2}\right]$$

$$\log f(x;\theta) = -\log\theta - \frac{1}{\theta} x$$

$$\frac{\partial \log f}{\partial \theta} = -\frac{1}{\theta} + \frac{1}{\theta^2} x$$

$$\frac{\partial^2 \log f}{\partial \theta^2} = +\frac{1}{\theta^2} - \frac{2}{\theta^3} x$$

$$\Rightarrow \quad I(\theta) = -E\left[\frac{\partial^2 \log f}{\partial \theta^2}\right] = -\frac{1}{\theta^2} + \frac{2}{\theta^3} \underbrace{E[X]}_{=\theta \text{ (given)}}$$

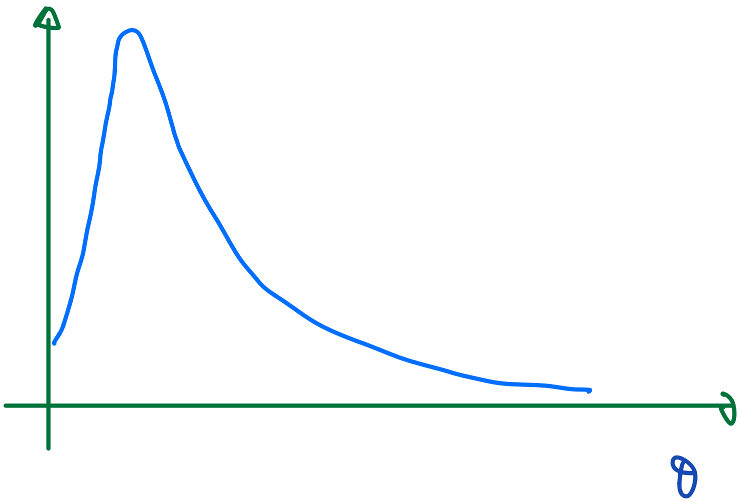
$$= -\frac{1}{\theta^2} + \frac{2}{\theta^3} \times \theta = \frac{1}{\theta^2}$$

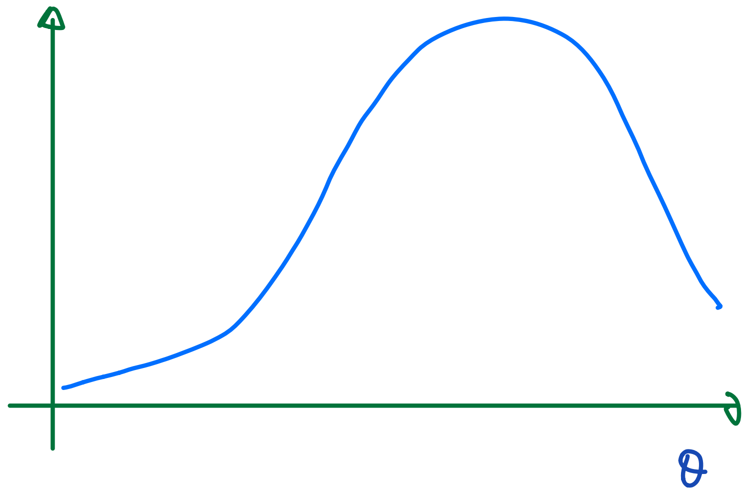
(*) Observe $I(\theta)$ large when θ small

$I(\theta)$ small when θ large.

We make a rough plot of the log-likelihood

$\ell_n(\theta)$ below





Left: θ small, peaky likelihood small variance for MLE

Right: θ large, flattish likelihood large variance for MLE

Asymptotic variance of $\hat{\theta}_{MLE}$ is

$$\frac{1}{n\,I(\theta)} = \frac{\theta^2}{n}.$$

4. Suppose $\{X_i\}_{i=1}^{n}$ are iid geometrically distributed random variables with probability mass function

$$P(X = k; \pi) = \pi(1-\pi)^{k-1} \qquad k = 1, 2, \ldots$$

You can use that $\mathrm{E}[X_i] = 1/\pi$ and $\mathrm{var}[X_i] = (1-\pi)/\pi^2$.

(i) Define the new random variable

$$Y_i = \begin{cases} 1 & \text{if } X_i = 1 \\ 0 & \text{if } X_i > 1 \end{cases}$$

Obtain $\mathrm{E}[Y_i]$ and derive the method of moments estimator of $\overset{\pi}{\cancel{p}}$ based on $\{Y_i\}$. Call this estimator $\widehat{\pi}_{MoM}$

(ii) Derive the asymptotic distribution of $\widehat{\pi}_{MoM}$. [3]

(iii) Using $\{X_i\}$ derive the maximum likelihood estimator of π. Call this $\widehat{\pi}_{MLE}$

(iv) Derive the asymptotic distribution of $\widehat{\pi}_{MLE}$ (using any of the results we derived in class). [3]

(v) Construct asymptotic 95% confidence intervals for π using $\widehat{\theta}_{MoM}$ and $\widehat{\theta}_{MLE}$ (this should be two intervals).

(vi) Compare the asymptotic variance of $\widehat{\pi}_{MoM}$ and asymptotic variance of $\widehat{\pi}_{MLE}$. In terms of variance which estimator is better and how does this impact the confidence interval in part (v)?

(vii) Which of the random variables $\{Y_i\}$ and $\{X_i\}$ contain more "information" about the parameter π. [4]

Only a heuristic explanation is required you do not need to calculate the Fisher information of both estimators.

(i) Y_i = Bernoulli random variable where

$$P\{Y_i = 1\} = E[Y_i] = \pi$$

$$\hat{\pi}_{mom} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$ (= proportion out of n with $Y=1$)

(ii) $\mathrm{var}(Y) = \pi(1-\pi)$. By classical CLT, then for a "large" enough n

$$\hat{\pi}_{mom} \xrightarrow{D} N\left(\pi, \frac{\pi(1-\pi)}{n}\right)$$

(iii) The likelihood of π based on $\{X_i\}$ is

$$L_n(\pi) = \prod_{i=1}^{n} \pi (1-\pi)^{x_i - 1}$$

The log-likelihood is

$$\mathcal{L}_n(\pi) = \sum_{i=1}^{n} \left[\log \pi + (x_i - 1) \log (1-\pi) \right]$$

The MLE is found by differentiating $\mathcal{L}_n(\pi)$ wrt π and solving;

$$\frac{\partial \mathcal{L}_n(\pi)}{\partial \pi} = \sum_{i=1}^{n} \left[\frac{1}{\pi} - \frac{(x_i - 1)}{1-\pi} \right] = 0$$

Now solve

$$\frac{n}{\pi} - \frac{1}{(1-\pi)} \sum_i x_i + \frac{n}{(1-\pi)} = 0$$

$$\Rightarrow \quad \frac{n[(1-\cancel{\pi}) + \cancel{\pi}]}{\pi \cancel{(1-\pi)}} = \frac{1}{\cancel{(1-\pi)}} \sum_i x_i$$

$$\hat{\pi}_{MLE} = \left[\frac{1}{n} \sum_{i=1}^{n} x_i \right]^{-1} = \left[\bar{X} \right]^{-1}$$

(iv) The above is just an average. So by the classical CLT we have that for a sufficiently large n that

$$\bar{X} \xrightarrow{D} N\left(\pi, \frac{var(X)}{n}\right)$$

$$= N\left(\pi, \frac{(1-\pi)/\pi^2}{n}\right).$$

By using Lemma 1.2 we have

$$\hat{\pi}_{MLE} = g(\bar{X}) = \frac{1}{\bar{X}}, \quad \text{thus} \quad g'(\bar{X}) = -\frac{1}{\bar{X}^2}$$

$$\Rightarrow g'(\mu) = -\frac{1}{\mu^2}$$

Thus $[g'(\mu)]^2 = \frac{1}{\mu^4} = \pi^4$ (since $\mu = \frac{1}{\pi}$)

$$\Rightarrow \quad \hat{\pi}_{MLE} \longrightarrow N\left(\pi, \pi^4 \times \frac{(1-\pi)}{\pi^2 n}\right)$$

$$= N\left(\pi, \frac{(1-\pi)\pi^2}{n}\right).$$

Alternatively Evaluate the Fisher information;

$$\frac{\partial \log f(x_i;\pi)}{\partial \pi} = \left[\frac{1}{\pi} - \frac{(x_i - 1)}{1-\pi} \right]$$

$$\frac{\partial^2 \log f(x_i;\pi)}{\partial \pi^2} = \left[-\frac{1}{\pi^2} - \frac{(x_i - 1)}{(1-\pi)^2} \right]$$

$$E\left[-\frac{\partial^2 \log f(x_i;\pi)}{\partial \pi^2} \right] = \frac{1}{\pi^2} + \frac{E(x_i) - 1}{(1-\pi)^2}$$

$$= \frac{1}{\pi^2} + \frac{1/\pi - 1}{(1-\pi)^2} = \frac{1}{\pi} + \frac{(1-\pi)}{\pi(1-\pi)^2}$$

$$= \frac{1}{\pi^2} + \frac{1}{\pi(1-\pi)} = \frac{(1-\pi) + \pi}{\pi^2(1-\pi)}$$

$$= \boxed{\frac{1}{\pi^2(1-\pi)} = I(\pi)}$$ Thus using Theorem 3.2

$$\hat{\pi}_{MLE} \rightarrow N\left(\pi, \frac{\pi^2(1-\pi)}{n}\right)$$

(*) Of course its the same as above!

(v) The (asymptotic) 95% CI for π based on $\hat{\pi}_{MoM}$ is

$$\left[\hat{\pi}_{MoM} \pm 1.96 \times \sqrt{\frac{\pi(1-\pi)}{n}}\right]$$ Replace π with $\hat{\pi}_{MoM}$

The (asymptotic) 95% for π based on $\hat{\pi}_{MLE}$ is

$$\left[\hat{\pi}_{MLE} \pm 1.96 \times \sqrt{\frac{\pi^2(1-\pi)}{n}}\right]$$ Replace π with $\hat{\pi}_{MLE}$.

(vi) The variance of $\hat{\pi}_{MoM}$ is $\frac{\pi(1-\pi)}{n}$

The asymptotic variance of $\hat{\pi}_{MLE}$ is $\frac{\pi^2(1-\pi)}{n}$.

Observe since $0 \le \pi \le 1$, then

$\frac{\pi(1-\pi)}{n} \ge \frac{\pi^2(1-\pi)}{n}$, thus the asymptotic variance of $\hat{\pi}_{MoM}$ is greater than $\hat{\pi}_{MLE}$.

Based on the variance $\hat{\pi}_{MLE}$ is a better estimator than $\hat{\pi}_{mom}$ and leads to narrower confidence intervals.

(vii) Observe that

$$Y_i = \begin{cases} X_i & \text{if } X_i = 1 \\ 0 & \text{if } X_i \neq 1 \end{cases}$$

Thus Y_i is effectively "putting" together all information on X_i when $X_i \neq 1$. This means we are loosing information on the distribution when we change from X_i to Y_i.

Conclusion $\{Y_i\}$ contains less information on π than $\{X_i\}$. This is why the variance of the estimator based on of $\{Y_i\}$ is greater than the variance of the estimator based on $\{X_i\}$.