# 501 Old Final Hints

December 8, 2025

## 1 Problem 1

See Chapter 13 page 122 on Hypothesis Testing in a Likelihood model in Joris's lecture notes.

## 2 Problem 2

### 2.1 a

The density is

$$f(x) = \frac{x}{\theta_0^2} \exp(-\frac{x^2}{2\theta_0^2})$$

The Likelihood is then

$$L(\theta; x_1, \cdots, x_n) = \Pi_{i=1}^n \frac{x_i}{\theta^2} \exp(-\frac{x_i^2}{2\theta^2})$$

The FOC for log-likelihood has

$$\sum_{i=1}^n \frac{x_i^2}{\hat{\theta}^3} = \sum_{i=1}^n \frac{2}{\hat{\theta}}$$

Hence

$$\hat{\theta} = \sqrt{\frac{\sum_{i=1}^n x_i^2}{2n}}$$

### 2.2 b

We first note that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n x_i^2 - E[x_i^2] \xrightarrow{d} \mathcal{N}(0, Var(x_i^2))$$

In particular, $Var(x_i^2) = E(x_i^4) - (E[x_i^2])^2$.

Using change of variables, we have $E(x_i^4) = 8\theta_0^4$ and $E(x_i^2) = 2\theta_0^2$. Therefore, the asymptotic variance is $4\theta_0^4$.

Now apply Delta method to conclude that

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}(0, \frac{\theta_0^2}{4})$$

One can also apply results on maximum likelihood estimator by computing $(E[\dot{l}_{\theta_0}\dot{l}'_{\theta_0}])^{-1}$

# 3 Problem 3

See old midterm 2 for ideas

# 4 Problem 4

This question is essentially about $\lim_{n\to\infty} E\hat{\theta}$ or about convergence theorems.

$\bar{x}$, $\bar{x}^2$, $\exp(-\bar{x}^2) - 1$, $\exp(-\frac{1}{\bar{x}^2})$ and $\mathbb{1}(|\bar{x}| \geq \epsilon)$ has limit of expectation being zero using direct computation or dominated convergence theorem.

$\exp(-\bar{x}) - 1$ cannot be concluded to have limit of expectation to be zero.

# 5 Problem 5

## 5.1 a

We verify the consistency theorem with compact parameter space:

1. The parameter space $\Theta = [-100, 100]^d$ is compact.
2. Assume that $\theta_0$ is the unique minimizer of $\Omega(\theta) = E[\ln(1 + (y - x'\theta)^2)]$ $\theta_0$ satisfies the FOC of the population objective given the symmetry assumption
3. The population objective $\Omega(\theta)$ is continuous
4. U-WCON: $\sup_{\theta\in\Theta} |\frac{1}{n}\sum_{i=1}^n \ln(1+(y_i-x_i'\theta)^2)-E(\ln(1+(y-x'\theta)^2)| = o_p(1)$

The last two conditions can be established by observing that

1. $(y_i, x_i')$ are iid samples
2. $f(y, x, \theta) = \ln(1+(y-x'\theta)^2)$ is a continuous function of $\theta$ for almost every $(y_i, x_i')$.
3. $\Theta$ is compact
4. $E(\sup_{\theta\in\Theta} \ln(1 + (y - x'\theta)^2) < \infty$

Under the above conditions, $\hat{\theta} \xrightarrow{p} \theta_0$

### 5.2 b

Here $\theta_0$ is in the interior of the parameter space, perform Taylor expansion around the approximate FOC:

$$\frac{1}{n}\sum_{i=1}^{n}\frac{y_i-\hat{\theta}'x_i}{1+(y_i-x_i'\hat{\theta})^2}x_i = o_p(n^{-1/2})$$

$$\frac{1}{n}\sum_{i=1}^{n}\frac{y_i-\theta_0'x_i}{1+(y_i-x_i'\theta_0)^2}x_i + \frac{1}{n}\sum_{i=1}^{n}\frac{(y_i-x_i'\theta^*)^2-1}{[1+(y_i-x_i'\theta^*)^2]^2}x_ix_i'(\hat{\theta}-\theta_0) = o_p(n^{-1/2})$$

where $\theta^*$ lies on the line segment between $\hat{\theta}$ and $\theta_0$.

Scale by $\sqrt{n}$, we have

$$\frac{1}{n}\sum_{i=1}^{n}\frac{(y_i-x_i'\theta^*)^2-1}{[1+(y_i-x_i'\theta^*)^2]^2}x_ix_i'\sqrt{n}(\hat{\theta}-\theta_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{y_i-\theta_0'x_i}{1+(y_i-x_i'\theta_0)^2}x_i + o_p(n^{-1/2})$$

As $y_i$ is distributed symmetrically around $\theta_0'x_i$, $\frac{y_i-\theta_0'x_i}{1+(y_i-x_i'\theta_0)^2}x_i$ has mean zero, and we can apply CLT (assuming existence of second moment for $\frac{y_i-\theta_0'x_i}{1+(y_i-x_i'\theta_0)^2}x_i$) to get

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{y_i-\theta_0'x_i}{1+(y_i-x_i'\theta_0)^2}x_i \xrightarrow{d} \mathcal{N}(0, E((\frac{y_i-\theta_0'x_i}{1+(y_i-x_i'\theta_0)^2})^2x_ix_i')$$

Abbreviate the limit variance as $\Sigma$.

In addition, assume that within a neighborhood of $\theta_0$ $E(\sup_{\theta\in N(\theta_0)}\|\frac{(y_i-x_i'\theta)^2-1}{[1+(y_i-x_i'\theta)^2]^2}x_ix_i'\|) < \infty$, then we have

$$\frac{1}{n}\sum_{i=1}^{n}\frac{(y_i-x_i'\theta^*)^2-1}{[1+(y_i-x_i'\theta^*)^2]^2}x_ix_i' \xrightarrow{p} E(\frac{(y_i-x_i'\theta_0)^2-1}{[1+(y_i-x_i'\theta_0)^2]^2}x_ix_i')$$

Define the RHS as $G$, assume that $G$ is full-rank, than we have

$$\sqrt{n}(\hat{\theta}-\theta_0) \xrightarrow{d} \mathcal{N}(0, G^{-1}\Sigma G^{-1})$$

# 6 Problem 6

### 6.1 a

We use Bernstein's inequality, here $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}Var(x_i) = \frac{1}{n}\sum_{i=1}^{n}2i = n+1$, the bound of the support of the random varibles is a function of $n$ and is given by $c_n = (n+1)b$. We are looking for $B_n$ such that for any $\epsilon > 0$ there exists a sufficiently large $K$,

$$\limsup_{n\to\infty} P(B_n|\sum_{i=1}^{n}(x_i - i)| \geq K) \leq \epsilon$$

We bound the above probability using Bernstein's inequality as:

$$\begin{aligned}
&P(|\sum_{i=1}^{n}(x_i - i)| \geq \frac{K}{B_n}) \\
&\leq 2\exp(-\frac{K^2/B_n^2}{2n(n+1) + \frac{2K}{3B_n}b(n+1)}) \\
&= 2\exp(-\frac{K^2}{2n^2B_n^2 + 2nB_n^2 + \frac{2K}{3}b(n+1)B_n})
\end{aligned}$$

It will be sufficient to choose $B_n$ such that the denominator has limit $\lim_{n\to\infty} n^2B_n^2 + nB_n^2 + nB_n < \infty$. Then by choosing $K$ sufficiently large the last line will be very small.

Therefore, $B_n$ should be decreasing faster than $\frac{1}{n}$. In other words, the slowest diverging sequence $C_n$ is $n$.

### 6.2 b

As $P(\min_i x_i \geq 1) = 1$, we can apply Markov: $P(\sum_{i=1}^{100} \ln(x_i) > 200) \leq \frac{\sum_{i=1}^{n} E(\ln(x_i))}{200} \leq \frac{\ln(2)}{2} \leq \frac{1}{2}$, where the second last step is Jensen's inequality.