# EE 325 Midterm-2 Exam, Autumn 2019

Prof. Animesh Kumar

- This is a closed book exam. You are allowed ONE SIDE of an A4-sheet with handwritten formulas.
- You have 60 minutes to finish this exam.
- **Show partial work to receive credit.**
- Calculators/gadgets/cellphones are **not allowed**.
- There are FOUR (4) questions in total. Please ensure that there are SEVEN (7) printed pages.
- **You may use the empty pages to do your work.**

NAME: Model Solutions

ROLL NO.: —

| Question No | 1 | 2 | 3 | 4 | | Total |
|---|---|---|---|---|---|---|
| Total | 7 | 3 | 5 | 3 | | 18 |
| Score | | | | | | |

1. Let $X$ be a random variable with the distribution

$$p_X(-1) = p_X(1) = p \text{ and } p_X(0) = 1 - 2p$$

where $0 < p < 1/2$ is an unknown parameter. Let $X_1, X_2, \ldots, X_n$ be IID samples of the random variable $X$. Answer the following questions. Explain your answers for credit.

(a) Write down a formula for the likelihood of $\mathbb{P}((X_1, \ldots, X_n) = (x_1, \ldots, x_n))$ in terms of $(x_1, \ldots, x_n)$ and $p$. *(2 marks)*

Observe that $p_X(b) = (b=-1)\cdot p + (b=1)\cdot p + (b=0)(1-2p)$

$= |b|\cdot p + (1-|b|)(1-2p)$

only for $b=-1,0,1$ $\left\{ = p^{|b|} \cdot (1-2p)^{(1-|b|)} \right.$ {this is more convenient.

So, by IID nature of $X_1, \ldots, X_n$,

$$p_{X_1,\ldots X_n}(x_1,\ldots,x_n) = \prod_{i=1}^{n} p_{X_i}(x_i)$$

$$= p^{\sum_{i=1}^{n} |x_i|} (1-2p)^{n - \sum_{i=1}^{n} |x_i|}$$

(b) Obtain the maximum-likelihood estimate $\widehat{p}$ of $p$ from the IID samples $(X_1, X_2, \ldots, X_n)$ of $X$. (*2 marks*)

From (a),

$$\log p_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = \sum_{i=1}^{n} |x_i| \cdot \log p + \left(n - \sum_{i=1}^{n} |x_i|\right)(\log(1-2p)).$$

Observe that $|X_i| \sim \text{Ber}(2p)$. Define $S_n = \sum_{i=1}^{n} |x_i|$.

Then, (log) likelihood maximization occurs when

$$\frac{d}{dp} \log p_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = 0$$

i.e. $$\frac{S_n}{p} + \frac{n-S_n}{1-2p} \cancel{(1-2p)} (-2) = 0$$

i.e., $\frac{S_n}{2n} = p$. Verify that $\frac{d^2}{dp^2} \log p_{X_1,\ldots,X_n}(x_1,\ldots,x_n)$

$$= \frac{-S_n}{p^2} + \frac{4(n-S_n)}{(1-2p)^2} < 0$$

at $p = \frac{S_n}{2n}$.

(c) Obtain a bound of the following form for the maximum likelihood estimate $\widehat{p}$ in part (b):

$$\mathbb{P}(|\widehat{p} - p| \geq \varepsilon) \leq C \exp(n\mu(\varepsilon))$$

for an arbitrary but small enough $\varepsilon > 0$. Specify the constants $C$ and $\mu(\varepsilon)$ in terms of $p, \varepsilon$. (More marks will be awarded for a $C$ which is smaller and $\mu(\varepsilon)$ which is more negative.) *(3 marks)*

Note that $\mathbb{E}(\hat{p}) = \mathbb{E}\left(\frac{S_n}{2n}\right) = p$.

By the chernoff formulation:

$$\mathbb{P}(|\hat{p} - p| \geq \varepsilon)$$

$$= \mathbb{P}\left(\left|\frac{1}{2n}\sum_{i=1}^{n}|X_i| - p\right| \geq \varepsilon\right)$$

$$= \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}|X_i| - 2p\right| \geq 2\varepsilon\right)$$

$$\leq \exp\left(n \cdot \mu_+(2p+2\varepsilon)\right) + \exp\left(n \cdot \mu_-(2p-2\varepsilon)\right).$$

where $\mu_+$ and $\mu_-$ are the Chernoff exponent for Ber$(2p)$ r.v. in $r \geq 0$ part of the ROC and $r \leq 0$ part of the ROC respectively.

From homeworks, $\mu_+(2p+2\varepsilon) = D(2p+2\varepsilon \| 2p)$ and $\mu_-(2p-2\varepsilon) = D(2p-2\varepsilon \| 2p)$.

So, $C = 2$ and $\mu(\varepsilon) = -\min\{D(2p+2\varepsilon \| 2p), D(2p-2\varepsilon \| 2p)\}$.

2. In a clocked router, at any clock instant, a packet arrives at a router with probability $p$ in an independent and identically distributed fashion. That is, at most one packet arrives at each clock instant independently, and probability of a packet arrival is $p$. Assume that the router serves for $N$ clock cycles, where $N$ is a Poisson$(n)$ random variable. The variable $N$ is independent of the packet arrivals. What is the mean and variance of total number of arrived packets at the router? Explain your answer for credit. *(3 marks)*

~~Let $Z_1, Z_2, \ldots$~~

Please check hwk solutions. Key steps are

- Defining the arrival process clearly and $S_N$ (arrivals till $N$). Indep. of arrivals has to be explicitly stated.
- Using Total Expectation Theorem. A common error was

$$E(S_N) = E\left(E(S_N \mid N=k)\right).$$

Also $N=k$ conditioning has to be removed with reasons.

- Rest of it was simple (but important!)

3. Let $f(x), 0 \leq x \leq 1$ with $0 \leq f(x) \leq b$ be a function of interest. It is desired to compute

$$I_f := \int_0^1 f(x)\mathrm{d}x.$$

As an outcome of an experiment, the samples $f(U_1), f(U_2), \ldots, f(U_n)$ are available where $U_1, U_2, \ldots, U_n$ are IID Uniform[0, 1] random variables. Answer the following questions. Explain your answers for credit.

(a) Find the expectation of $f(U)$, where $U \sim$ Uniform[0, 1]. *(1 mark)*

$$\mathbb{E}(f(U)) = \int_0^1 f(u) \cdot \underbrace{f_U(u)}_{=1} \, du$$

$$= \int_0^1 f(u)\, du = I_f.$$

(b) Is the random variable $f(U)$ subGaussian? If yes, what is its variance parameter? *(1 mark)*

Since $0 \leq f(U) \leq b$, i.e., $f(U)$ is bounded, so $f(U)$ is subGaussian with parameter $\left(\frac{b-0}{2}\right)^2 = \frac{b^2}{4}$.

(c) The puffy hair man argues that $f(U_1), \ldots, f(U_n)$ estimate $I_f$ well since there is a sequence of positive real numbers $c_n \downarrow 0, n \in \mathbb{N}$ such that

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} f(U_i) - I_f\right| > c_n\right) \downarrow 0 \text{ as } n \uparrow \infty.$$

Is the puffy hair man correct in his assertion? If yes, find and explain a sequence $c_n \downarrow 0$ supporting his assertion. If no, explain why not? (More marks will be awarded for a $c_n$ which is smaller under the given conditions.) *(3 marks)*

Since $f(U_1), \ldots, f(U_n)$ are all sub Gaussian with expectation $I_f$ and parameter $\frac{b^2}{4}$, so, by Hoeffding inequality:

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} f(U_i) - I_f\right| > c_n\right) \leq 2 \cdot \exp\left(-\frac{c_n^2}{2\sum_{i=1}^{n}\frac{b^2}{4n^2}}\right)$$

$$= 2\exp\left(-\frac{2n c_n^2}{b^2}\right).$$

To see that puffy hair man is correct, we need a $c_n \downarrow 0$ such that $n c_n^2 \uparrow \infty$.

Set $c_n = \sqrt{\frac{\log_e n}{n}}$ to get the result.

4. Let $Y$ be a lognormal random variables such that $\log_e Y \sim \mathcal{N}(0,1)$. For every integer $k \geq 1$, establish a probability inequality of the form

$$\mathbb{P}(Y \geq t) \leq \frac{C_k}{t^k}, \quad t \in [t_k, \infty).$$

Your answer should specify the constants $C_k$ and the values of $t_k$ for which these upper bounds are smaller than one. Explain your answer for credit. *(3 marks)*

Let $\log Y = Z$, where $Z \sim \mathcal{N}(0,1)$.

Then, $Y = e^Z \geq 0$.

For $t > 0$, By Markov's inequality on $Y^k$, $k \geq 1$ and integer:

$$\mathbb{P}(Y \geq t) = \mathbb{P}(Y^k \geq t^k) \quad \text{(1)}$$

$$\leq \frac{\mathbb{E}(Y^k)}{t^k} = \frac{\mathbb{E}(e^{kZ})}{t^k}.$$

$$= \frac{g_Z(k)}{t^k}.$$

Next, $g_Z(r) = e^{\mu r + \frac{\sigma^2 r^2}{2}} = e^{r^2/2}$ as $Z \sim \mathcal{N}(0,1)$.

So, $$\mathbb{P}(Y \geq t) \leq \frac{e^{\sigma^2 k^2/2}}{t^k} = \frac{e^{k^2/2}}{t^k}.$$

desired $\boxed{C_k = e^{k^2/2}}$

For upper bound to be $\leq 1$, $t \geq e^{k/2} = t_k$.

I.e., $\boxed{t_k = e^{k/2}}$.