Stats 200 Midterm

Instructions:

- **Do not turn this page until instructed to do so.**
- Please write neatly: we can't grade what we can't read.
- To obtain full credit, you must justify your answers. We are primarily interested in your understanding of concepts, so show us what you know.
- No one is expected to answer all of the questions; but everyone is encouraged to try them.
- This midterm is open book. You may refer to your textbook, your class notes, the scribe notes, or the exercise sheets.
- At the end of this document, you will find a common distributions table of reference.

---

1 Problem 1: Probability

Consider a random variable X with cumulative distribution function (CDF)

$$F(x) = \begin{cases} 1 - \frac{1}{x^3} & \text{for } x \geq 1, \\ 0 & \text{else.} \end{cases}$$

(a) Derive the mean of X.

(b) Let X_i, $i = 1, \ldots, n$ be i.i.d. draws from F. Derive the asymptotic distribution of $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \mathbb{E}[X])$.

(c) Derive the density of $Y = X^3$.

Solution.

(a) A short calculation shows that the mean is $\int_1^\infty 3x^{-3} dx = \frac{3}{2}$. For this part of the question, we will not give points if students just state the mean.

(b) A similar calculation as above shows that the variance of X is $\int_1^\infty 3x^{-2} dx - \frac{3^2}{2^2} = \frac{3}{1} - \frac{9}{4} = \frac{3}{4}$. By the CLT, the asymptotic distribution is $\mathcal{N}(0, \frac{3}{4})$.

(c) $F_Y(y) = P(Y \leq y) = P(X^3 \leq y) = F(y^{1/3})$. Thus, the density of Y is

$$f_Y(y) = \begin{cases} \frac{1}{y^2} & \text{for } y \geq 1, \\ 0 & \text{else.} \end{cases}$$

2 Problem 2: Hypothesis testing

You observe the first symptom X of an illness and the goal is to distinguish between two possible illnesses (which have different distributions over X). The probabilities are given in the following table.

X	sneezing	fever	fainting	sore throat	runny nose
$H_0 : f_0$ (cold)	1/4	1/100	1/100	3/100	70/100
$H_A : f_A$ (flu)	1/2	10/100	2/100	5/100	33/100

.

(a) A doctor tells you that as a rule of thumb, you should consider *fainting* or *sore throat* as an indication of flu. Argue whether this is a good rule of thumb.

(b) Construct a most powerful test of H_0 vs H_A based on X with type I error $\alpha = .01$. Compute the power of this test.

(c) Now we take a Bayesian approach and consider a prior on the two illnesses with $P(H_0) = .8$. Compute the posterior odds $P(H_0|X)/P(H_A|X)$ for a friend who experiences $X =$ fainting as the first symptom of illness. Based on the posterior odds, do you believe that your friend has the flu?

Solution.

(a) This is not a good rule of thumb. There are other events with lower likelihood ratio such as having fever.

(b) The test that rejects the null if $X =$ fever is most powerful by the Neyman-Pearson Lemma at level $\alpha = .01$. Indeed it corresponds to a likelihood ratio strictly smaller than 0.101. This test has power $\mathbb{P}(\text{fever} \mid H_a) = .1$.

(c) The posterior odds are $.8/.2 \cdot .5 = 2$. Note that the posterior odds are larger than 1. Based on the posterior odds, it is likelier that your friend has a cold.

3 Problem 3: Maximum likelihood estimation and method of moments

Suppose $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} f(x;\sigma)$ where

$$f(x;\sigma) = \begin{cases} \frac{\sqrt{2}}{\sigma\sqrt{\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right) & \text{if } x > 0, \\ 0 & \text{if } x \le 0. \end{cases}$$

Here, $\sigma > 0$ is an unknown parameter that we want to estimate. You may also use the following facts without proof: $\mathbb{E}[X_i] = \frac{\sigma\sqrt{2}}{\sqrt{\pi}}$, $\mathbb{E}[X_i^2] = \sigma^2$ and $\mathbb{E}[X_i^4] = 3\sigma^2$.

(a) Find the MLE $\hat{\sigma}_{\text{mle}}$ for σ.

(b) Using the delta method (or another method), prove that $\sqrt{n}(\hat{\sigma}_{\text{mle}} - \sigma) \xrightarrow{\text{d}} \mathcal{N}(0, v^2)$ for some constant $v > 0$. Furthermore, find v^2 in terms of σ^2.

(c) Find a method of moments estimator $\hat{\sigma}_{\text{mom}}$ for σ based on $\mathbb{E}[X]$.

(d) Prove that for some constant v^2, $\sqrt{n}(\hat{\sigma}_{\text{mom}} - \sigma) \xrightarrow{\text{d}} \mathcal{N}(0, \kappa^2)$. Furthermore, find κ^2 in terms of σ^2.

(e) Derive the Cramér-Rao lower bound for σ. Compare it to the asymptotic distribution of the maximum likelihood estimator.

(f) Based on the results above (and not referring to the book or in-class results), argue which estimator you'd prefer in this setting. *Hint: recall that $\pi > 3$.*

Solution. Proof sketch:

(a) Note that the log likelihood is given by

$$l(\sigma) = n \log\left(\frac{\sqrt{2}}{\sqrt{\pi}}\right) - n\log(\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}^{n} X_i^2.$$

The MLE is the value of $\sigma > 0$ that maximizes $l(\sigma)$. Hence the MLE is $\hat{\sigma}_{\text{mle}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} X_i^2}$.

(b) We use the CLT and the delta method. Since $\mathbb{E}[X_i^2] = \sigma^2$ and $\text{Var}(X_i) = \mathbb{E}[X_i^4] - (\mathbb{E}[X_i^2])^2 = 2\sigma^4$, by the CLT,

$$\sqrt{n}\Big(\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \sigma^2\Big) \xrightarrow{\text{d}} \mathcal{N}(0, 2\sigma^4).$$

Taking $g(u) = \sqrt{u}$, by the delta method, the MLE satisfies

$$\sqrt{n}(\sigma_{\text{mle}} - \sigma) \xrightarrow{\text{d}} \mathcal{N}\Big(0, \left[g'(\sigma^2)\right]^2 \cdot 2\sigma^4\Big).$$

Hence the MLE is asymptotically Gaussian, with mean σ and asymptotic variance $v^2 = \frac{1}{(2\sqrt{\sigma^2})^2}(2\sigma^4) = \frac{\sigma^2}{2}$.

Note: There was a typo in the problem statement and actually $\mathbb{E}[X_i^4] = 3\sigma^4$. If you plugged in $\mathbb{E}[X_i^4] = 3\sigma^2$ to your calculations and your answer was that $v^2 = \frac{3-\sigma^2}{4}$, you will still get full credit.

(c) By partial integration we have $\mathbb{E}[X] = \frac{\sigma\sqrt{2}}{\sqrt{\pi}}$. Thus, the MoM estimator is $\frac{\sqrt{\pi}}{n\sqrt{2}} \sum_{i=1}^n X_i$.

(d) The log-likelihood for one sample is

$$l_1(\sigma) = \log\big(\frac{\sqrt{2}}{\sqrt{\pi}}\big) - \log(\sigma) - \frac{1}{2\sigma^2}X_1^2 \Rightarrow l_1''(\sigma) = \frac{1}{\sigma^2} - \frac{3X_1^2}{\sigma^4}.$$

Because the necessary smoothness conditions are met, the Fisher information for one sample is therefore

$$I_1(\sigma) = -\mathbb{E}[l_1''(\sigma)] = -\mathbb{E}\Big[\frac{1}{\sigma^2} - \frac{3X_1^2}{\sigma^4}\Big] = -\frac{1}{\sigma^2} + \frac{3\mathbb{E}[X_1^2]}{\sigma^4} = \frac{2}{\sigma^2}.$$

Because the samples are IID the Cramér-Rao lower bound is given by $\frac{1}{nI_1(\sigma)} = \frac{\sigma^2}{2n}$. It coincides with the asymptotic variance of the MLE.

(e) By the CLT, $\sqrt{n}(\hat{\sigma}_{\text{mom}} - \sigma) \xrightarrow{\text{d}} \mathcal{N}(0, \kappa^2)$. The variance is $\kappa^2 = \frac{\pi}{2}\text{Var}(X_1) = \frac{\pi}{2}(\sigma^2 - \frac{2}{\pi}\sigma^2) = (\frac{\pi}{2} - 1)\sigma^2$.

(f) Let's compare the two asymptotic variances found in parts (b) and (e). Since $\pi > 3$, we have $\frac{1}{2} < \frac{\pi}{2} - 1$. Thus, $v^2 < \kappa^2$.

4 Problem 4: Bayesian Estimation

You want to estimate the quality of coconuts in a container. You randomly take four independent coconuts from the container and open them. The first three coconuts are of great quality (type "g"), and the fourth coconut has mediocre quality (type "k"), but none of the drawn coconuts are of bad quality (type "b"). The quality of the coconuts can be expressed as a two-dimensional vector $\theta = (\theta_1, \theta_2)$ with nonnegative entries that satisfy $\theta_1 + \theta_2 \leq 1$. More specifically, the θ's correspond to $P(\text{Quality}_i = g|\theta) = \theta_1$, $P(\text{Quality}_i = k|\theta) = \theta_2$ and $P(\text{Quality}_i = b|\theta) = 1 - \theta_1 - \theta_2$, for $i = 1, \ldots, 4$. From past experience, you know that the quality of coconuts can vary drastically, so you decide to put a uniform prior on the quality of coconuts:

$$P(\Theta = \theta) = \begin{cases} \frac{1}{2} & \text{for } (\theta_1, \theta_2) \text{ with } \min(\theta_1, \theta_2) \geq 0 \text{ and } \theta_1 + \theta_2 \leq 1, \\ 0 & \text{else.} \end{cases}$$

(a) What is the posterior density of θ? Explicitly write the normalizing constant C as an integral, but you do not need to solve this integral.

(b) What is the posterior mean of θ_1? You may leave your answer in terms of the normalizing constant C and the function $\beta(a, b) := \int_0^1 x^a (1-x)^b dx$. For partial credit, express the posterior mean as an (unsimplified) multivariate integral.

(c) What is the posterior mode estimator of $\theta = (\theta_1, \theta_2)$? *Hint: recall that the mode of a continuous distribution is the value at which the density takes its maximum value.*

Solution.

(a) Since the prior is uniform, the posterior is proportional to the likelihood, which is $\theta_1^3 \theta_2$ since we observe three coconuts of quality "g" and one of type "k" (assuming θ_1, θ_2 are nonnegative and sum to less than one—otherwise they are not valid probabilities for this problem). Thus, the posterior distribution is

$$P(\Theta = \theta | \text{data}) = \begin{cases} C\theta_1^3 \theta_2 & \text{for } \theta_i \geq 0 \text{ and } \theta_1 + \theta_2 \leq 1, \\ 0 & \text{else} \end{cases}$$

where the normalizing constant satisfies $C = \left(\int_0^1 \int_0^{1-\theta_1} \theta_1^3 \theta_2 d\theta_2 d\theta_1 \right)^{-1} = 120$.

(b) The posterior mean of θ_1 is

$$\begin{aligned} \mathbb{E}[\theta_1] = C \int_0^1 \theta_1 \cdot \int_0^{1-\theta_1} \theta_1^3 \theta_2 d\theta_2 d\theta_1 &= C \int_0^1 \theta_1 \cdot \frac{1}{2} \theta_1^3 (1-\theta_1)^2 d\theta_1 \\ &= C \frac{1}{2} \int_0^1 \theta_1^4 (1-\theta_1)^2 d\theta_1 \\ &= C \frac{1}{2} \beta(4, 2) \\ &= \frac{4}{7}. \end{aligned}$$

(c) The posterior mode is achieved for $\theta_1 + \theta_2 = 1$ because $x \mapsto \theta_1^3 \theta_2$ is increasing in both θ_1 and θ_2. Thus we have to maximize the function

$\theta_1 \mapsto \theta_1^3(1-\theta_1)$. This is simply a binomial likelihood, which is maximized when $\theta_1 = 3/4$ and and $\theta_2 = 1/4$ (one could also take the log and take derivatives to maximize this).

Intuitively this makes sense since the posterior mode under a flat prior coincides with maximum likelihood.

Distribution on X	Support	PDF or PMF	$\mathbb{E}[X]$	$\mathrm{Var}(X)$
Bernoulli(p)	$\{0,1\}$	$p_X(k) = p^k(1-p)^{1-k}$	p	$p(1-p)$
Binomial(n,p)	$\{0,1,\ldots,n\}$	$p_X(k) = \binom{n}{k} p^k(1-p)^{n-k}$	np	$np(1-p)$
Poisson(λ)	$\{0,1,2,\ldots\}$	$p_X(k) = \frac{\lambda^k e^{-\lambda}}{k!}$	λ	λ
Geometric(p)	$\{1,2,3,\ldots\}$	$p_X(k) = p(1-p)^{k-1}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Normal(μ,σ^2)	$(-\infty,\infty)$	$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$	μ	σ^2
Exponential(λ)	$[0,\infty)$	$f_X(x) = \lambda e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Gamma(α,β)	$[0,\infty)$	$f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$	$\frac{\alpha}{\beta}$	$\frac{\alpha}{\beta^2}$
Beta(α,β)	$[0,1]$	$f_X(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

Table 1: Common distributions table of reference. The third column gives the probability mass function (in the case of discrete random variables) or the probability density function (in the case of continuous random variables). Note that the probability mass functions and probability density functions are defined to be equal to zero outside of the support.