STAT 135, Midterm exam, Spring 2016, H. Pitters

NAME (IN CAPS):____
SID number and SECTION:__________________________________

Show your work or provide a brief explanation for all answers. This quiz is closed books. You are allowed to use the notes you took during class, a calculator and extra scratch paper. Answers should be simplified as much as possible. Good luck!

1	2	3	4	5	Σ

Question 1. [16] Provide a concise answer to each of the following questions.

(1) Explain in your own words how the maximum likelihood method works. Is this method intuitively plausible? [2]

A: Given observations $x_1, \ldots, x_n$ from some random variables $X_1, \ldots, X_n$ with joint distribution $\mathbb{P}_\theta$ that depends on a parameter θ, the likelihood of θ is

$$\text{lik}(\theta) := f(x_1, \ldots, x_n|\theta),$$

where $f(x_1, \ldots, x_n|\theta)$ is the joint density of $X_1, \ldots, X_n$ under $\mathbb{P}_\theta$. The MLE for θ is the value $\hat{\theta}$ that maximizes lik(θ). This is a plausible estimator for θ, since it corresponds to the one model $\mathbb{P}_{\hat{\theta}}$ under which the observations are most likely to be observed.

(2) Explain what it means to have a 95% confidence interval for some parameter μ. [2]

A: A 95% confidence interval is a random interval $[a(X), b(X)]$ (usually) constructed from the sample $X = (X_1, \ldots, X_n)$. If one were to repeatedly and independently draw samples of size n and construct a 95% confidence interval for each of them, then about 95% of these intervals will contain the parameter θ.

(3) Based on a sample of size n, drawn from an unknown population, we are told that $[2.7, 5.2]$ is a 90% confidence interval for some unknown parameter μ. Does this mean the probability for μ to lie in $[2.7, 5.2]$ is 90%? Explain your answer. [2]

A: No. The parameter μ is deterministic, and so is the interval $[2.7, 5.2]$. It either contains μ or it does not.

(4) Suppose $\hat{\theta}$ is an estimator for some unknown parameter θ. What does it mean for $\hat{\theta}$ to be unbiased? [2]

A: It means that the mean value of $\hat{\theta}$ is θ, i.e. $\mathbb{E}\hat{\theta} = \theta$.

(5) Suppose T is a sufficient statistic for a parameter θ. [2]

(a) Explain in your own words what sufficiency means.
A: As far as inference (e.g. estimation) of θ is concerned, the statistic T contains all necessary information. Any information about the sample $X_1, \dots, X_n$ other than $T(X_1, \dots, X_n)$ does not improve the inference of θ.

(b) Why is the maximum likelihood estimator (MLE) for θ a function of T?
A: From the factorization theorem we have that

$$\mathrm{lik}(\theta) = f(x_1, \dots, x_n|\theta) = g(T(x_1, \dots, x_n), \theta)h(x).$$

The MLE $\hat{\theta}$ is the value that maxmizes $\mathrm{lik}(\theta)$, or equivalently $g(T(x_1, \dots, x_n), \theta)$. However, the latter depends on $x_1, \dots, x_n$ only via $T(x_1, \dots, x_n)$, hence $\hat{\theta}$ is a function of T.

(6) Explain the difference between the Central Limit Theorem (CLT) and the Law of Large Numbers (LLN). Assuming both theorems hold, can you derive one from the other? [2]

A: Consider the average of the first n items

$$\bar{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i$$

of a sequence $X_1, X_2, \dots$ of independent and identically distributed random variables with mean $\mu := \mathbb{E}X_1$ and variance $\sigma^2 := \mathrm{Var}(X_1)$. The Law of Large Numbers states that as n grows without bounds $\bar{X}_n$ converges towards μ.

Even if n is large, the average $\frac{1}{n}\sum_{i=1}^{n} X_i$ is going to fluctuate randomly about μ. The Central Limit Theorem states that these fluctuations have a Gaussian distribution, namely the standardized average

$$\frac{\frac{1}{n}\sum_{i=1}^{n} X_i - \mu}{\sigma/\sqrt{n}}$$

has approximately a standard normal distribution $\mathcal{N}(0, 1)$.

Assuming both theorems hold, for any $\epsilon > 0$ the CLT implies

$$\mathbb{P}\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| \geq \epsilon\right\} = \mathbb{P}\left\{\frac{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right|}{\sigma/\sqrt{n}} \geq \frac{\sqrt{n}}{\sigma}\epsilon\right\} \approx 1 - \Phi(\frac{\sqrt{n}}{\sigma}\epsilon) \to 0$$

as $n \to \infty$. Alternatively, Chebyshef's inequality implies for any $\epsilon > 0$

$$\mathbb{P}\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| \geq \epsilon\right\} \leq \frac{\sigma^2}{\epsilon^2 n} \to 0$$

as $n \to \infty$. Both statements show the so-called weak Law of Large Numbers.

(7) For random variables $X_1, \ldots, X_n$ is the statement

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} X_i$$

correct? Explain. [2]

A: The statement is false, since the left hand side is a (deterministic) number, whereas the right hand side is a random variable. For instance, with $n = 1$ the statement reads $\mathbb{E}X_1 = X_1$.

(8) You are advising a political candidate on opinion polls. Two organizations have submitted bids to do a survey to estimate the extent of his support. Both will use nationwide probability samples of 2,500 people, and both charge the same price. One organization only states that it has a 95% chance to get the estimate right to within 3%. The other absolutely guarantees that its estimate will be right to within 3%. Other things being equal, which bid should be accepted, and why? [2]

A: Simply because of chance error there is no way that an estimate constructed from a sample (whose size is a small fraction of the total population) should be guaranteed to be right within 3%. The modest bid should therefore be accepted.

Question 2. [4] In a simple random sample of 100 graduate students from a certain college, 48 were earning \$70.000 a year or more. Estimate the percentage of all graduates of that college earning \$70.000 a year or more. [2] Provide an (estimated) standard error (ignoring finite population correction). [2]

A: An (unbiased) estimator $\hat{p}$ for the percentage p of graduates earning \$70.000 a year and more is the fraction of graduates earning \$70.000 a year or more in the sample, i.e. $\hat{p} = 48\%$. An estimated standard error for this estimator is

$$\sigma_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \approx \frac{1}{20} = 0.5.$$

Question 3. [4] A sample $X_1, \ldots, X_n$ of n independent draws is taken with common distribution $\mathrm{Pn}(\alpha)$. We want to perform Bayesian inference, where the parameter α is modeled as random with prior distribution an exponential distribution with parameter $\lambda > 0$.

(1) Compute the joint distribution of $X_1, \ldots, X_n$ and α. [1]

A: Denoting by $\pi(\alpha) = \lambda e^{-\lambda\alpha}\mathbf{1}_{[0,\infty)}(\alpha)$ the density of the prior distribution of α, we have

$$f(x_1,\dots,x_n,\alpha) = f(x_1,\dots,x_n|\alpha)\pi(\alpha) = \prod_{i=1}^n e^{-\alpha}\frac{\alpha^{x_i}}{x_i!}\lambda e^{-\lambda\alpha}\mathbf{1}_{[0,\infty)}(\alpha)$$
$$= e^{-\alpha(n+\lambda)}\frac{\alpha^{\sum_i x_i}}{\prod_i x_i!}\lambda\mathbf{1}_{[0,\infty)}(\alpha).$$

(2) Compute the posterior distribution $f(\alpha|x_1,\dots,x_n)$ of α given the observations $X_1 = x_1,\dots,X_n = x_n$. (Hint: There is no need to compute the marginal distribution of $X_1,\dots,X_n$.) [3]

A: Denoting by $m(x_1,\dots,x_n)$ the probability mass function of the observations, we find

$$f(\alpha|x_1,\dots,x_n) = \frac{f(x_1,\dots,x_n,\alpha)}{m(x_1,\dots,x_n)} \propto e^{-\alpha(n+\lambda)}\alpha^{\sum_i x_i}\mathbf{1}_{[0,\infty)}(\alpha).$$

Consequently, the posterior distribution of α is a gamma$(\sum_i x_i + 1, n+\lambda)$ distribution.

Question 4. [16] In n independent experiments physicists measure the (unknown) mass μ of an object recording the result of the first experiment as X_1, the result of the second experiment as X_2, etc. up until X_n, the result of the nth experiment. The experimenters assume that measurements vary randomly about the true mass μ according to a normal distribution with variance 1, i.e. for each $1 \le i \le n$ they assume

$$X_i \sim \mathcal{N}(\mu, 1).$$

(1) Find the method of moments estimator for μ. [2]

A: Population mean: $\mu_1 := \mathbb{E}X_1 = \mu$, hence $\hat{\mu} = \hat{\mu}_1 = \bar{X}_n$ is the method of moments estimator for μ.

(2) Given the data $x_1,\dots,x_n$ of the experiments, write down the likelihood lik(μ) of μ. [2]

A:

$$\operatorname{lik}(\mu) := \prod_{i=1}^n \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x_i-\mu)^2} = (2\pi)^{-\frac{n}{2}}e^{-\frac{1}{2}\sum_{i=1}^n(x_i-\mu)^2}.$$

(3) Find an (simplified) expression for the log likelihood $l(\mu) := \log\operatorname{lik}(\mu)$. [2]

A:

$$l(\mu) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^n(x_i-\mu)^2.$$

(4) Find the maximum likelihood estimator for μ. [2]

A: The derivative

$$l'(\mu) = \sum_{i=1}^{n}(x_i - \mu) = n\bar{x}_n - n\mu$$

of the log likelihood has root $\hat{\mu} = \bar{x}$. Since for $\mu > 0$ we have $l''(\mu) < 0$, $\hat{\mu} = \bar{x}_n$ maximizes the likelihood and is therefore the MLE for μ.

(5) As an estimator for μ consider

$$\bar{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i.$$

(a) Is $\bar{X}_n$ an unbiased estimator? [2]

A: Yes, since $\mathbb{E}\bar{X}_n = \mathbb{E}X_i = \mu$.

(b) Is $\bar{X}_n$ a consistent estimator? [2]

A: Yes. The Law of Large Numbers implies the convergence (in probability)

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to \mathbb{E}X_1 = \mu,$$

and hence $\bar{X}_n$ is a consistent estimator for μ.

(c) What is the standard error of $\bar{X}_n$? [2]

A: The standard error is $\sqrt{\mathrm{Var}(\bar{X}_n)} = \sqrt{\frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i)} = \frac{1}{\sqrt{n}}$.

(d) What is the sampling distribution of $\bar{X}_n$? [2]

A: $\bar{X}_n$ is a convolution of n i.i.d. random variables each of which has distribution $\mathcal{N}(\mu/n, 1/n^2)$, hence $\bar{X}_n \sim \mathcal{N}(\mu, 1/n)$.

Question 5. [2] Suppose $\hat{\theta}$ is an unbiased estimator for θ. Starting from the definition of the MSE, derive a simple expression for $\mathrm{MSE}(\hat{\theta})$.

A:

$$\mathrm{MSE}(\hat{\theta}) := \mathbb{E}(\hat{\theta} - \theta)^2 = \mathbb{E}\hat{\theta}^2 - 2\theta\mathbb{E}\hat{\theta} + \theta^2 = \mathbb{E}\hat{\theta}^2 - \theta^2 = \mathbb{E}\hat{\theta}^2 - (\mathbb{E}\hat{\theta})^2 = \mathrm{Var}(\hat{\theta}),$$

since $\mathbb{E}\hat{\theta} = \theta$ by assumption.