STAT 135 Spring 2015, Midterm Exam

March 13, 2015

If you are stuck in one question of a problem, you can move to the following questions. Partial credit will be given. Try to answer short and to the point. Good luck!

Problem 1 [50%]

The following are reasoning questions that require very little derivations.

1. [2 pts] If $\widehat{\theta}$ is an efficient estimator of θ, can you find an efficient estimator of 2θ? Justify your answer.

   Since $\widehat{\theta}$ is efficient, it satisfies

   $$\mathbf{E}\left(\widehat{\theta}\right) = \theta \quad , \quad \mathrm{var}\left(\widehat{\theta}\right) = \frac{1}{\mathrm{I}(\theta)} \ .$$

   We pick $2\widehat{\theta}$ as estimator for 2θ. It is unbiased since $\mathbf{E}\left(2\widehat{\theta}\right) = 2\mathbf{E}\left(\widehat{\theta}\right) = 2\theta$, and $\mathrm{var}\left(2\widehat{\theta}\right) = 4\mathrm{var}\left(\widehat{\theta}\right)$. Now, applying the change of variables $g(\theta) = 2\theta$ to the Fisher information (available in the cheat sheet), we obtain that $\mathrm{I}(2\theta) = \frac{\mathrm{I}(\theta)}{|g(\theta)'|^2} = \frac{\mathrm{I}(\theta)}{4}$, and hence $\mathrm{var}\left(2\widehat{\theta}\right) = \frac{1}{\mathrm{I}(2\theta)}$ so $2\widehat{\theta}$ is also efficient.

   Intuitively, this means that the quality of an estimator does not depend on the measurement units.

2. [2 pts] Explain the difference between the Central Limit Theorem and the Law of Large Numbers. Can you use one to derive the other? (assuming both theorems hold). The Law of Large Numbers tells us that the sample mean of iid random variables $X_1, \ldots, X_n$ converges in probability towards a constant, which is $\mathbf{E}(X)$. The Central Limit theorem tells us more than that: it also specifies the asymptotic distribution of the sample mean around its limit. In other words, LLN only tells us that $\bar{X}_n \to \mathbf{E}(X)$, whereas the CLT tells us that $\bar{X}_n \sim \mathcal{N}(\mathbf{E}(X), \frac{\mathrm{var}(X)}{n})$. So, since $\frac{\mathrm{var}(X)}{n} \to 0$, it results that $P(|\bar{X}_n - \mathbf{E}(X)| > \epsilon) \leq \frac{\mathrm{var}(X)}{n\epsilon^2} \to 0$, and therefore the CLT implies the LLN.

   For example, when constructing a confidence interval for $\mathbf{E}(X)$ using the CLT, it takes the form $(\bar{X}_n \pm q(1-\alpha/2)\frac{\sigma}{\sqrt{n}})$, which tells us in particular that $\bar{X}_n \to \mathbf{E}(X)$ as $n \to \infty$.

3. [2 pts] Imagine you have to decide whether someone has special power to predict coin tosses. Assuming a null hypothesis that no such power exists, which experiment has smaller p-value,

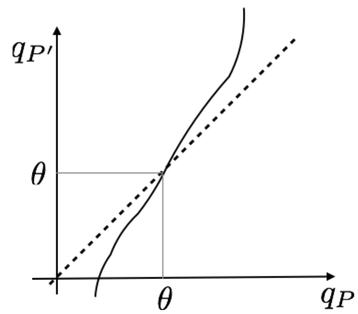
70 correct guesses out of 100, or 700 correct guesses out of 1000? (*Note: You don't need to compute the p-values; Assume Normal approximation.*).

Under the null, and assuming normal approximation, the number of correct guesses s is distributed as

$$s \sim \mathcal{N}(\frac{n}{2}, \frac{n}{4}) \ , \ \text{or} \ \ \frac{2}{\sqrt{n}}(s - n/2) \sim \mathcal{N}(0, 1)$$

so the p-value depends monotonically on $\frac{2s-n}{\sqrt{n}}$. When n and s are scaled by a factor 10, the numerator is scaled by a factor 10, but the denominator by $\sqrt{10}$, and therefore the second experiment has smaller p-value.

4. [2 pts] This figure shows a Q-Q plot of two distributions P and P':



Which distribution has heavier tails? If $X \sim P$ and $Y \sim P'$, what is more likely, $X \geq \theta$ or $Y \geq \theta$?

P' has heavier tails, since for large quantiles we have $q_{P'}(\alpha) > q_P(\alpha)$. By definition, for any point $(q_P(\alpha), q_{P'}(\alpha))$ in the QQ-plot, we have $Pr(X \geq q_P(\alpha)) = \alpha = Pr(Y \geq q_{P'}(\alpha))$, so in particular $Pr(X \geq \theta) = Pr(Y \geq \theta)$.

5. [2 pts] We observe $X \sim \text{Unif}[0, \theta]$ and we want to test whether

$$H_0 \ : \ \theta = 1 \ \ \text{or} \ \ H_1 \ : \ \theta = 2 \ .$$

Design a test with significance 0. What is the largest power you can get?
Design a test with power 1. What is the smallest significance you can get?

We first consider the test that rejects when $x > 1$. Its significance is $P(X > 1|H_0) = 0$ and its power is $P(X > 1|H_1) = \int_1^2 \frac{1}{2} dx = \frac{1}{2}$. This is the largest power we can obtain, since $P(\text{reject}|H_1) > 1/2$ necessarily implies that the test will reject with non-zero probability when $x < 1$, which implies a positive significance.

A test with power 1 means that $P(\text{reject}|H_1) = 1$, or, in other words, that the test must reject for any $x \in (0, 2)$. In particular, we reject for any $x \in (0, 1)$, which means that such test has significance 1.

Problem 2 [50%]

Let $X_1, \ldots, X_n \sim \text{Poisson}(\lambda)$ iid, where λ is an unknown parameter, and let $\theta = P(X_1 = 0)$. We are interested in estimating θ.

1. [1 pts] Let us first consider the non-parametric estimator of θ. Let us define

$$Y_i = \begin{cases} 1 & \text{if } X_i = 0 \ , \\ 0 & \text{otherwise.} \end{cases}$$

   Show that Y_i are iid Bernoulli variables with parameter θ. Show that $\overline{Y} = \frac{1}{n} \sum Y_i$ is an unbiased estimator of θ and compute its MSE.

   By the definition, we have that $P(Y = 1) = P(X = 0) = \theta$, so Y is a Bernoulli with parameter θ. By the linearity of the Expected value, we have that $\mathbf{E}\left(\overline{Y}\right) = \mathbf{E}\left(Y_i\right) = \theta$, and by independence we have that $\text{var}\left(\overline{Y}\right) = \frac{\theta(1-\theta)}{n}$. Since $\overline{Y}$ is unbiased, we have that $MSE(\overline{Y}) = \frac{\theta(1-\theta)}{n}$.

Let us now construct a parametric estimator for θ.

2. [1 pt] Show that $\theta = e^{-\lambda}$. By applying the definition of Poisson distribution we have that $\theta = P(X = 0) = e^{-\lambda}$.

3. [1 pt] Show that $\widehat{\lambda}_{MLE} = \frac{\sum_{i=1}^n X_i}{n}$.

   The log-likelihood of $X_1, \ldots, X_n$ is

$$l(\lambda) = \sum X_i(\log \lambda) - n\lambda + C ,$$

   and hence $l'(\lambda) = \frac{\sum_i X_i}{\lambda} - n$. By setting to 0 we obtain $\widehat{\lambda}_{MLE} = \frac{\sum_{i=1}^n X_i}{n}$. We verify that $l'(\lambda) > 0$ for $\lambda < \widehat{\lambda}_{MLE}$ and $l'(\lambda) < 0$ for $\lambda > \widehat{\lambda}_{MLE}$, so we conclude that $\widehat{\lambda}_{MLE}$ is the MLE.

4. [2 pts] Show that if $y = g(\lambda)$ is an invertible function of λ and $\widehat{\lambda}_{MLE}$ is the MLE of λ, then $\hat{y} = g(\widehat{\lambda}_{MLE})$ is the MLE of y. *(Hint: Consider the change of variables $\lambda = g^{-1}(y)$)* .

   Since g is invertible, the likelihood of the data with respect to y is

$$(1) \quad , \forall y \ , \ \widetilde{lik}(y) = lik(g^{-1}(y)) \leq lik(\hat{\lambda}_{MLE})$$

   by definition of MLE of θ. But we have that $\hat{y} = g(\hat{\lambda}_{MLE})$ satisfies

$$(2) \quad \widetilde{lik}(\hat{y}) = lik(g^{-1}(\hat{y})) = lik(g^{-1}(g(\hat{\lambda}_{MLE}))) = lik(\hat{\lambda}_{MLE})$$

   so from (1) and (2) we have that

$$\forall y \ , \ \widetilde{lik}(y) \leq \widetilde{lik}(\hat{y}) ,$$

   which means that $\hat{y}$ is the maximum likelihood estimator.

5. [1 pt] Use the previous result to show that the Maximum Likelihood estimator $\widehat{\theta}_{MLE}$ of θ is

$$\widehat{\theta}_{MLE} = e^{\left(-\frac{\sum X_i}{n}\right)} .$$

   Since $\widehat{\lambda}_{MLE} = \frac{\sum_{i=1}^n X_i}{n}$ and $\theta = g(\lambda)$ with $g(x) = e^{-x}$, we apply the previous result to obtain $\widehat{\theta}_{MLE} = e^{-\widehat{\lambda}_{MLE}}$.

6. [2 pts] The δ-Method says that if $\sqrt{n}(X_n - \theta) \to \mathcal{N}(0, \sigma^2)$ and g is a smooth function, then $\sqrt{n}(g(X_n) - g(\theta)) \to \mathcal{N}(0, |g(\theta)'|^2\sigma^2)$. Use the Central Limit Theorem on $\widehat{\lambda}_{MLE}$ and the δ-method to show that the asymptotic MSE of $\widehat{\theta}_{MLE}$ is $\frac{\lambda}{n}e^{-2\lambda}$.

   The CLT applied to $\widehat{\lambda}_{MLE}$ says that for large n, $\sqrt{n}(\widehat{\lambda}_{MLE} - \lambda) \sim \mathcal{N}(0, \lambda)$, since it is an average of independent Poisson random variables with variance λ. Now, the δ-method applied with $g(x) = e^{-x}$ gives

$$\sqrt{n}(\widehat{\theta}_{MLE} - \theta) \sim \mathcal{N}(0, e^{-2\lambda}\lambda) ,$$

   which means that

$$MSE(\widehat{\theta}_{MLE}) = \text{var}\left(\widehat{\theta}_{MLE}\right) + (\mathbf{E}\left(\widehat{\theta}_{MLE}\right) - \theta)^2 = \frac{e^{-2\lambda}\lambda}{n} .$$

7. [2 pts] Finally, show that the efficiency of $\overline{Y}$ relative to $\widehat{\theta}_{MLE}$ is

$$\frac{MSE(\overline{Y})}{MSE(\widehat{\theta}_{MLE})} = \frac{e^{\lambda} - 1}{\lambda} > 1 \ .$$

How do you interpret this fact? Which estimator is using more information? By inserting the previous results we obtain

$$\frac{MSE(\overline{Y})}{MSE(\widehat{\theta}_{MLE})} = \frac{e^{2\lambda} e^{-\lambda}(1 - e^{-\lambda})}{\lambda} = \frac{e^{\lambda} - 1}{\lambda} \ ,$$

which is always larger than 1 since

$$\frac{e^{\lambda} - 1}{\lambda} = 1 + \sum_{k>0} \lambda^k/(k+1)! \ .$$

The parametric estimator thus is more efficient, since it is using information for all possible values of X, rather than just remembering when $X > 0$ or not.

Useful Formulas

- If $X \sim \text{Poisson}(\lambda)$ then

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \ , \ k = 0, 1, 2 \ldots \ .$$

  We have

$$\mathbf{E}\left(X\right) = \lambda \ , \ \operatorname{var}\left(X\right) = \lambda \ .$$

- The Fisher Information of a parameter θ with respect to a random variable X with density f_θ and parameter $\theta = \theta_0$ is

$$\mathrm{I}(\theta) = \mathbf{E}\left(\left(\left.\frac{\partial \log f_\theta(X)}{\partial \theta}\right|_{\theta=\theta_0}\right)^2\right) = -\mathbf{E}\left(\left.\frac{\partial^2 \log f_\theta(X)}{\partial \theta^2}\right|_{\theta=\theta_0}\right) \ .$$

  If $\beta = g(\theta)$ is another parameter and g is invertible and smooth, then $\mathrm{I}(\beta) = \frac{\mathrm{I}(\theta)}{|g'(\theta)|^2}$.

- The Central Limit Theorem says that if $X_1, \ldots, X_n$ are iid random variables with $\mathbf{E}\left(X\right) = \mu$ and $\operatorname{var}\left(X\right) = \sigma^2 < \infty$, then

$$\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu) \xrightarrow{d} \mathcal{N}(0, 1) \ .$$

- The pdf of a normal distribution with mean μ and variance σ^2 is

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \ .$$