408 Final Exam Solutions[1]

1. Question 1

True/False

(a) The negation of the statement
"There exists an integer j such that $j^2 - j < 3$" is:
"For every integer j, we have $j^2 - j \geq 3$."

TRUE, by the rules of negation, "There exists" is negated to "For every," and the inequality $<$ is negated to $\geq$.

(b) Let $\mathbf{P}$ be the uniform probability law on $[0,1]$. Let $x_1, x_2, \ldots \in [0,1]$ be a countable set of distinct points. Then

$$\mathbf{P}\left(\cup_{n=1}^{\infty}\{x_n\}\right) = 0.$$

TRUE. By the definition of $\mathbf{P}$, $\mathbf{P}(\{x_n\}) = 0$ for all $n \geq 1$. So, from Axiom (ii) for probability laws,

(c) Let $X_1, \ldots, X_n$ be i.i.d random variables drawn from a family of probability density functions $\{f_\theta \colon \theta \in \mathbf{R}\}$ where $f_\theta \colon \mathbf{R} \to [0, \infty)$ for all $\theta \in \mathbf{R}$. Then there must exist some integer $k \geq 1$, $\exists$ some function $t \colon \mathbf{R}^n \to \mathbf{R}^k$ and there exists some statistic $Y = t(X_1, \ldots, X_n)$ such that Y is a sufficient statistic for θ.

TRUE. The statistic $(X_1, \ldots, X_n)$ is always sufficient for θ.

(d) Suppose $t(X)$ defined in the definition of p-value is a continuous random variable. Then the p-value satisfies

$$\mathbf{P}_\theta(p(X) \leq c) \leq c, \qquad \forall\, c \in (0,1).$$

TRUE by Remark 5.19 in the notes.

(e) Let $X_1, \ldots, X_n$ be positive random variables. Then Pearson's chi-squared statistic

$$S := \sum_{j=1}^{n} \frac{\Big(X_j - \mathbf{E}X_j\Big)^2}{\mathbf{E}X_j}$$

has a chi-squared distribution.

FALSE. Just let $n = 1$ and let X_1 be any non-Gaussian random variable.

2. Question 2

Let $X_1, \ldots, X_n$ be a random sample of size n from a Poisson distribution with unknown parameter $\lambda > 0$. (So, $\mathbf{P}(X_1 = k) = e^{-\lambda}\lambda^k/k!$ for all integers $k \geq 0$.)

- Find an MLE for λ. As usual, justify your answer.
- Is the MLE you found unique? That is, could there be more than one MLE for this problem? Justify your answer.

Solution. Denote $k = (k_1, \ldots, k_n) \in \mathbf{Z}_{\geq 0}^n$. Then

$$f_\lambda(k) = \prod_{i=1}^{n} e^{-\lambda}\lambda^{k_i}/k_i! = e^{-\lambda n}\lambda^{\sum_{i=1}^{n} k_i} \Big/ \prod_{i=1}^{n} k_i!.$$

[1]December 15, 2021,

$$\log f_\lambda(k) = -\lambda n + \log\lambda \cdot \sum_{i=1}^n k_i - \sum_{i=1}^n \log(k_i!).$$

Differentiating in λ, we get

$$\frac{d}{d\lambda}\log f_\lambda(k) = -n + \frac{1}{\lambda}\sum_{i=1}^n k_i$$

We get a single critical point for $\log f_\lambda$, when

$$\lambda = \frac{1}{n}\sum_{i=1}^n k_i.$$

Moreover, $\log f_\lambda$ is (possibly increasing) then decreasing, as λ increases. We conclude that the single critical point is then in fact a global maximum. So, the MLE is

$$\frac{1}{n}\sum_{i=1}^n X_i.$$

3. Question 3

Let $X_1, \ldots, X_n$ be a random sample of size n from the uniform distribution on $[\theta-1/2, \theta+1/2]$ where $\theta \in \mathbf{R}$ is unknown.

Show that

$$(X_{(1)}, X_{(n)})$$

is a sufficient statistic for θ.

(Here $X_{(1)} = \min_{1\leq i\leq n} X_i$ and $X_{(n)} = \max_{1\leq i\leq n} X_i$.)

Solution. Let $x = (x_1, \ldots, x_n) \in \mathbf{R}^n$. Using independence, we write the joint distribution of $X_1, \ldots, X_n$ as

$$f(x) = \prod_{i=1}^n 1_{x_i\in[\theta-1/2,\theta+1/2]}.$$

The quantity $\prod_{i=1}^n 1_{x_i\in[\theta-1/2,\theta+1/2]}$ is zero, except when $x_{(1)} \geq \theta - 1/2$ and $x_{(n)} \leq \theta + 1/2$. That is,

$$f(x) = 1_{x_{(1)}\geq\theta-1/2} 1_{x_{(n)}\leq\theta+1/2}.$$

So, defining $g_\theta(a,b) := 1_{a\geq\theta-1/2}1_{b\leq\theta+1/2}$, $h(x) := 1$, $t(x) := (x_{(1)}, x_{(n)})$, we have written

$$f(x) = g_\theta(t(x)) \cdot h(x), \qquad \forall\, x \in \mathbf{R}^n.$$

So, by the factorization theorem, $t(X) = (X_{(1)}, X_{(n)})$ is sufficient for θ.

4. Question 4

Suppose X is a binomial distributed random variable with parameters 2 and $\theta \in \{1/2, 3/4\}$. (So, X has the distribution of the number of heads that appears from flipping a coin twice, where θ is the probability that a heads appears in a single coin flip.)

We want to test the hypothesis H_0 that $\theta = 1/2$ versus the hypothesis H_1 that $\theta = 3/4$.

- Explicitly describe the rejection region C of the UMP (uniformly most powerful) test among all hypothesis tests with significance level at most 1/4.
- Suppose we observe that $X = 2$. Report a p-value for this observation, for the UMP test you found.

Solution. The Neyman-Pearson Lemma says that the UMP test for the class of tests with an upper bound on the significance level must be a likelihood ratio test. The likelihood ratio test has rejection region

$$C = \{x \in \mathbf{R}\colon f_{\theta_1}(x) \geq k f_{\theta_0}(x)\} = \{x \in \mathbf{R}\colon f_{3/4}(x) \geq k f_{1/2}(x)\}.$$

There are only three values that X can take, so we examine the likelihood ratios explicitly:

$$\frac{f_{3/4}(0)}{f_{1/2}(0)} = \frac{(1-3/4)^2}{(1-1/2)^2} = \frac{1}{4}, \qquad \frac{f_{3/4}(1)}{f_{1/2}(1)} = \frac{2(1-3/4)(3/4)}{2(1-1/2)(1/2)} = \frac{3}{4}, \qquad \frac{f_{3/4}(2)}{f_{1/2}(2)} = \frac{(3/4)^2}{(1/2)^2} = \frac{9}{4}.$$

We then get different likelihood ratio tests according to the choice of $k > 0$.

- If $3/4 < k \leq 9/4$, then H_0 is rejected if and only if $X = 2$, and this test is the unique UMP for tests with significance level at most $\mathbf{P}_{1/2}(X = 2) = 1/4$.
- If $1/4 < k \leq 3/4$, then H_0 is rejected if and only if $X = 1$ or 2, and this test is the unique UMP for tests with significance level at most $\mathbf{P}_{1/2}(X \in \{1, 2\}) = 3/4$.

In the case $k = 2$, we get (using the table of values of $f_{3/4}(X)/f_{1/2}(X)$),

$$p(2) = \mathbf{P}_{1/2}\Big(\frac{f_{3/4}(X)}{f_{1/2}(X)} \geq \frac{f_{3/4}(2)}{f_{1/2}(2)}\Big) = \mathbf{P}_{1/2}\Big(\frac{f_{3/4}(X)}{f_{1/2}(X)} \geq \frac{9}{4}\Big) = \mathbf{P}_{1/2}(X = 2) = (1/2)^2 = 1/4.$$

5. Question 5

Let $X_1, \ldots, X_n$ be a random sample of size n from a Poisson distribution with unknown parameter $\lambda > 0$. (So, $\mathbf{P}(X_1 = k) = e^{-\lambda}\lambda^k/k!$ for all integers $k \geq 0$.)

Let Y be the estimator $Y = 1_{\{X_1 = 0\}}$.

(That is, $Y = 1$ when $X_1 = 0$, and otherwise $Y = 0$.)

- Explicitly compute $W_n := \mathbf{E}_\lambda(Y \mid \sum_{i=1}^n X_i)$.
- State an inequality comparing $\mathrm{Var}_\lambda(Y)$ and $\mathrm{Var}_\lambda(W_n)$.
- What happens to W_n as $n \to \infty$? Does it converge to something? Justify your answer.

(Hint: a sum of n independent Poissons with parameter λ is a Poisson with parameter $n\lambda$.)

Solution. A sum of n independent Poisson random variables, each with parameter $\lambda > 0$, is a Poisson random variable with parameter $n\lambda$. That is,

$$\mathbf{P}(\sum_{i=1}^n X_i = x) = e^{-\lambda n}(\lambda n)^x/x!, \qquad \forall\, x \in \mathbf{Z}_{\geq 0}.$$

So, the conditional PMF satisfies

$$\mathbf{P}\Big(1_{\{X_1=0\}} = 1 \mid \sum_{i=1}^n X_i = x\Big) = \frac{\mathbf{P}(1_{\{X_1=0\}} = 1, \sum_{i=1}^n X_i = x)}{\mathbf{P}(\sum_{i=1}^n X_i = x)}$$

$$= \frac{\mathbf{P}(X_1 = 0, \sum_{i=2}^n X_i = x)}{e^{-\lambda n}(\lambda n)^x/x!} = \frac{e^{-\lambda}e^{-\lambda(n-1)}(\lambda(n-1))^x/x!}{e^{-\lambda n}(\lambda n)^x/x!} = \Big(1 - \frac{1}{n}\Big)^x.$$

Then, since $1_{\{X_1=0\}}$ only takes values 0 and 1, the conditional expectation is

$$\mathbf{E}\Big(1_{\{X_1=0\}} \mid \sum_{i=1}^n X_i = x\Big) = \mathbf{P}\Big(1_{\{X_1=0\}} = 1 \mid \sum_{i=1}^n X_i = x\Big) = \Big(1 - \frac{1}{n}\Big)^x.$$

That is,

$$\mathbf{E}\Big(1_{\{X_1=0\}}\mid \sum_{i=1}^{n} X_i\Big) = \Big(1-\frac{1}{n}\Big)^{\sum_{i=1}^{n} X_i}.$$

That is,

$$W_n = \Big(1-\frac{1}{n}\Big)^{n\frac{1}{n}\sum_{i=1}^{n} X_i}.$$

As $n\to\infty$, $\Big(1-\frac{1}{n}\Big)^n$ converges to e^{-1}, and $\frac{1}{n}\sum_{i=1}^n X_i$ converges in probability to the constant $\mathbf{E}X_1=\lambda$, by the weak law of large numbers. So, as $n\to\infty$, W_n converges in probability to the constant $e^{-\lambda}$. That is, $W_1, W_2, \ldots$ is consistent.

Finally, $\mathrm{Var}_\lambda(Y)\geq \mathrm{Var}_\lambda(W_n)$ by Rao-Blackwell, since Y is unbiased, as $\mathbf{E}Y=\mathbf{P}(X_1=0)=\lambda$.

6. Question 6

Suppose X_1, X_2 is a random sample from a Gaussian random variable X with unknown mean $\mu_X\in\mathbf{R}$ and unknown variance $\sigma^2>0$. Suppose Y_1, Y_2 is a random sample from a Gaussian random variable Y with unknown mean $\mu_Y\in\mathbf{R}$ and unknown variance $\sigma^2>0$. Assume that X_1, X_2 is independent of Y_1, Y_2, i.e. assume that X, Y are independent.

Suppose you find that $X_1=1$, $X_2=3$, $Y_1=2$ and $Y_2=4$.

Explicitly construct a confidence interval of the form $[a,b]$ for $\mu_X-\mu_Y$, so that

$$\mathbf{P}(a\leq \mu_X-\mu_Y\leq b) = \frac{1}{2\sqrt{2}}\int_{-3}^{3}\Big(1+\frac{s^2}{2}\Big)^{-3/2} ds.$$

Hint: $\Gamma(1/2)=\sqrt{\pi}$, $\Gamma(1)=1$, $\Gamma(3/2)=\sqrt{\pi}/2$, $\Gamma(2)=1$, $\Gamma(5/2)=3\sqrt{\pi}/4$, $\Gamma(3)=2$.

Hint: Recall that Student's t-distribution with p degrees of freedom has density

$$f_T(s) := \frac{\Gamma(\frac{p+1}{2})}{\sqrt{p}\sqrt{\pi}\Gamma(p/2)}\Big(1+\frac{s^2}{p}\Big)^{-(p+1)/2}, \qquad \forall\, s\in\mathbf{R}.$$

Solution. We have $n=m=2$,

$$\overline{X} := \frac{1}{n}\sum_{i=1}^{n} X_i = 2, \qquad \overline{Y} := \frac{1}{m}\sum_{i=1}^{m} Y_i = 3,$$

$$S_X^2 := \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2 = [(1-2)^2+(3-2)^2] = 2$$

$$S_Y^2 := \frac{1}{m-1}\sum_{i=1}^{m}(Y_i-\overline{Y})^2 = [(2-3)^2+(4-3)^2] = 2,$$

$$S^2 := \frac{(n-1)S_X^2+(m-1)S_Y^2}{n+m-2} = \frac{2+2}{2} = 2.$$

Then

$$\frac{\overline{X}-\overline{Y}-\mu_X+\mu_Y}{S\sqrt{\frac{1}{n}+\frac{1}{m}}}$$

has Student's t-distribution with $p = n + m - 2 = 2$ degrees of freedom. Therefore,

$$\mathbf{P}\Big(\overline{X} - \overline{Y} - tS\sqrt{\frac{1}{n} + \frac{1}{m}} < \mu_X - \mu_Y < \overline{X} - \overline{Y} + tS\sqrt{\frac{1}{n} + \frac{1}{m}}\Big)$$
$$= \frac{\Gamma(\frac{p+1}{2})}{\sqrt{p}\sqrt{\pi}\Gamma(p/2)} \int_{-t}^{t} \Big(1 + \frac{s^2}{p}\Big)^{-(p+1)/2} ds,$$

Choosing $t = 3$, we get

$$\mathbf{P}\Big(-1 - 3\sqrt{2} < \mu_X - \mu_Y < -1 + 3\sqrt{2}\Big) = \frac{\Gamma(\frac{3}{2})}{\sqrt{2}\sqrt{\pi}\Gamma(1)} \int_{-3}^{3} \Big(1 + \frac{s^2}{p}\Big)^{-(2+1)/2} ds$$
$$= \frac{\sqrt{\pi}/2}{\sqrt{2}\sqrt{\pi}} \int_{-3}^{3} \Big(1 + \frac{s^2}{p}\Big)^{-3/2} ds$$
$$= \frac{1}{2\sqrt{2}} \int_{-3}^{3} \Big(1 + \frac{s^2}{2}\Big)^{-3/2} ds$$

That is, we choose $[a, b] = [-1 - 3\sqrt{2}, -1 + 3\sqrt{2}]$.

7. Question 7

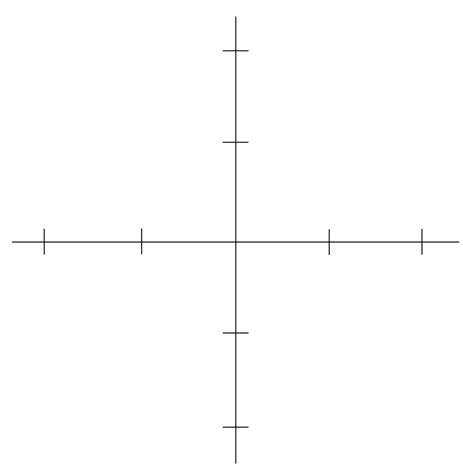
Suppose you are given the following three data points in (x, y) coordinates:

$$(x_1, y_1) = (-1, 0), \qquad (x_2, y_2) = (0, 0), \qquad (x_3, y_3) = (0, 1).$$

- Find the parabola of the form $y = mx^2 + b$ that best fits these three points. That is, find $m, b \in \mathbf{R}$ that minimizes the quantity.

$$h(m, b) := \frac{1}{2}\sum_{i=1}^{3} \Big(y_i - (mx_i^2 + b)\Big)^2.$$

- Make sure to prove that the minimal m, b that you find actually minimizes $h(m, b)$.
- Finally, plot the points (x_1, y_1), (x_2, y_2), (x_3, y_3) along with the parabola $y = mx^2 + b$ that best fits the points.



Solution. We have $h(m,b) = (1/2)[(m+b)^2 + b^2 + (1-b)^2]$. Then $h_m(m,b) = m+b$ and $h_b(m,b) = m+b+b+b-1 = 3b+m-1$. Solving for $h_m = h_b = 0$, we get $m = -b$ and $0 = 3b+m-1 = 2b-1$, so that $b = -m = 1/2$. So, the parameters $b = 1/2$, $m = -1/2$ are the only critical point of h. Since h is strictly convex, its critical point must be a global minimum.

The "best-fit" parabola is

$$y = -(1/2)x^2 + 1/2.$$

Note that $y(-1) = 0$, $y(0) = 1/2$. So, the parabola intersects $(-1,0)$, it lies above $(0,0)$ and it lies below $(0,1)$.

8. Question 8

Consider the following table with turkey data. We have 4 (vegetarian) turkeys, with various temperatures x (Fahrenheit), and the status y of each turkey is cooked (corresponding to a value of $y = 1$) or not cooked (corresponding to a value of $y = 0$). Using logistic regression, we would like to find $a, b \in \mathbf{R}$, i.e. find a function

$$h(ax+b)$$

that best fits your data, where $h(t) = 1/(1+e^{-t})$ for all $t \in \mathbf{R}$.

That is, given a temperature x, $h(ax+b)$ should be close to 1 when the turkey is cooked, and $h(ax+b)$ should be close to 0 when the turkey is not cooked.

Turkey	Temperature	Done? Yes or no.
1	150	no
2	155	yes
3	160	no
4	165	yes

Describe in detail how you would find the $a, b \in \mathbf{R}$ that best fit the data using a computer to do logistic regression.

Solution. Let $X_1, \ldots, X_4$ be i.i.d. real-valued random variables representing the temperatures of the turkeys. Let $g\colon \mathbf{R} \to \{0,1\}$ be an unknown function, and let $Y_i := g(X_i)$ for all $1 \leq i \leq n$, so that $g(X_i) = 0$ if turkey i is not cooked, and $g(X_i) = 1$ if turkey i is cooked, for all $1 \leq i \leq 4$.

By our assumptions, $Y_1, \ldots, Y_4$ are i.i.d. Bernoulli random variables with some unknown probability $0 \leq p \leq 1$ such that $p = \mathbf{P}(Y_1 = 1)$. Since the logistic function smoothly transitions from value 0 to value 1, we make the heuristic assumption that there are some unknown parameters $a, b \in \mathbf{R}$ such that

$$p \approx h(ax+b) \approx g(x).$$

The likelihood function is then

$$\ell(a,b) := \prod_{i=1}^{4} p^{y_i}(1-p)^{1-y_i} = \prod_{i=1}^{4} [h(ax_i+b)]^{y_i}[1-h(ax_i+b)]^{1-y_i},$$
$$\forall\, x_1, \ldots, x_4 \in \mathbf{R}, \quad \forall\, y_1, \ldots, y_4 \in \{0,1\}.$$

From a homework exercise, the log-likelihood function has at most one global maximum. So, if the MLE exists, it is unique.

To find the MLE, we start at some values of a, b (such as $a = b = 0$), and we perform the following iterative procedure many times

- Randomly perturb a, b. (For example, define $\widetilde{a} := a + X/100$, $\widetilde{b} := b + Y/100$, where X, Y are independent standard Gaussians.)
- If $\ell(\widetilde{a}, \widetilde{b}) > \ell(a, b)$, then replace (a, b) with $(\widetilde{a}, \widetilde{b})$, and perform the previous step again. Otherwise, keep the same a, b as before, the perform the previous step again.

Since the log likelihood has at most one global maximum, this stochastic gradient ascent procedure will eventually reach a value of ℓ that is close to its global maximum (if that maximum exists).