408 Midterm 2 Solutions[1]

1. Question 1

TRUE/FALSE

(a) A UMVU always exists.

FALSE. Suppose we want a UMVU for a binomial random variable X with known parameter n and unknown parameter $0 < \theta < 1$, and we want an estimator for $\theta/(1-\theta)$. No unbiased estimate exists for this function, since $\mathbf{E}_\theta t(X) = \sum_{j=0}^{n} \binom{n}{j} t(j)\theta^j(1-\theta)^{n-j}$ and this is a polynomial in θ, which in particular is bounded, but $\theta/(1-\theta)$ is not bounded as $\theta \to 1$.

(b) An MLE always exists.

FALSE. Let X be a Bernoulli random variable with unknown parameter $0 < p < 1$. Then the distribution of X is $f_p(x) = p^x(1-p)^{1-x}$. If $X = x = 1$, then $f_p(x) = p$, which is an increasing function of p. This function has no maximum value over all $0 < p < 1$, since $\sup_{0<p<1} p = 1$, but 1 is not in the range of allowed parameter values.

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.

(Your MLE should be a function of $X_1, \ldots, X_n$.)

Solution. The joint PMF of $X_1, \ldots, X_n$ is

$$\ell(\lambda) = \prod_{i=1}^{n} f_\lambda(x_i) = \prod_{i=1}^{n} e^{-\lambda}\lambda^{x_i}/x_i! = e^{-\lambda n}\lambda^{\sum_{i=1}^{n} x_i} \frac{1}{\prod_{i=1}^{n} x_i!}, \qquad \forall\, x_1, \ldots, x_n \in \mathbf{Z}_{\geq 0}.$$

We have $\ell'(\lambda) = 0$ only when

$$-ne^{-\lambda n}\lambda^{\sum_{i=1}^{n} x_i} + e^{-\lambda n}(\sum_{i=1}^{n} x_i)\lambda^{-1+\sum_{i=1}^{n} x_i} = 0.$$

That is,

$$-n + \lambda^{-1}\sum_{i=1}^{n} x_i = 0.$$

That is, (if $x_1, \ldots, x_n$ are not all zero),

$$\lambda = \frac{1}{n}\sum_{i=1}^{n} x_i. \qquad (*)$$

(If $x_1 = \cdots = x_n = 0$, then $\ell(\lambda) = e^{-\lambda n}$, and the maximum of $\ell(\lambda)$ occurs at $\lambda = 0$.) Moreover, the first derivative test implies that $(*)$ is the unique maximum of $\ell(\lambda)$, since $\ell'(\lambda) > 0$ when $\lambda < \frac{1}{n}\sum_{i=1}^{n} x_i$ and $\ell'(\lambda) < 0$ when $\lambda > \frac{1}{n}\sum_{i=1}^{n} x_i$. So, the MLE for λ is the statistic Y_n defined by

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

[1]November 8, 2023,

3. Question 3

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 $1/4$.

Hint: you can freely use the following facts about the PMF f_θ of X

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

Solution. The Neyman-Pearson Lemma says that a UMP test for the class of tests with an upper bound on the significance level must be a likelihood ratio test with significance level equal to $1/4$. That is, there is some $k > 0$ such that the UMP hypothesis test rejects only when $f_{3/4}(X) > k f_{1/2}(X)$.

Choosing e.g. $k = 1$, we see that H_0 is rejected if and only if $X = 2$, and this test is the unique UMP for tests with significance level at most $\mathbf{P}_{1/2}(X = 2) = 1/4$.

4. Question 4

Let $X := (X_1, \ldots, X_n)$ be a random sample of size n from a binomial distribution with parameters n and p. Here n is a positive (known) integer and $0 < p < 1$ is unknown. (That is, $X_1, \ldots, X_n$ are i.i.d. and X_1 is a binomial random variable with parameters n and p, so that $\mathbf{P}(X_1 = k) = \binom{n}{k} p^k (1-p)^{n-k}$ for all integers $0 \leq k \leq n$.)

You can freely use that $\mathbf{E}X_1 = np$ and $\mathrm{Var}X_1 = np(1-p)$.

- Computer the Fisher information $I_X(p)$ for any $0 < p < 1$.
  (Consider n to be fixed.)
- Let Z be an unbiased estimator of p (assume that Z is a function of $X_1, \ldots, X_n$). State the Cramér-Rao inequality for Z.
- Let W be an unbiased estimator of p^3 (assume that W is a function of $X_1, \ldots, X_n$). State the Cramér-Rao inequality for W.

Solution. Using that the information of independent random variables is the sum of the informations, using the alternate definition of Fisher information using the variance, and using that the variance is unchanged by adding a constant inside the variance,

$$\begin{aligned}
I_X(p) = nI_{X_1}(p) &= n\mathrm{Var}_p\Big(\frac{d}{dp}\Big[\log\Big(\binom{n}{X_1}p^{X_1}(1-p)^{n-X_1}\Big)\Big]\Big) \\
&= n\mathrm{Var}_p\Big(\frac{d}{dp}\Big[\log\binom{n}{X_1} + X_1\log p + (n-X_1)\log(1-p)\Big]\Big) \\
&= n\mathrm{Var}_p\Big(\frac{d}{dp}\Big[X_1\log p + (n-X_1)\log(1-p)\Big]\Big) \\
&= n\mathrm{Var}_p\Big(\frac{1}{p}X_1 - \frac{1}{1-p}(n-X_1)\Big) = n\mathrm{Var}_p\Big(\Big[\frac{1}{p}+\frac{1}{1-p}\Big]X_1\Big) \\
&= n\Big[\frac{1}{p}+\frac{1}{1-p}\Big]^2\mathrm{Var}_pX_1 = n\Big[\frac{1}{p(1-p)}\Big]^2 np(1-p) = \frac{n^2}{p(1-p)}
\end{aligned}$$

The Cramér-Rao inequality says, if $g(p) := \mathbf{E}_p Z$, then

$$\mathrm{Var}_p(Z) \geq \frac{|g'(p)|^2}{I_X(p)}.$$

If $g(p) = p$, then $g'(p) = 1$, so we get

$$\mathrm{Var}_p(Z) \geq \frac{1}{I_X(p)} = \frac{p(1-p)}{n^2}.$$

If $g(p) = p^3$, then $g'(p) = 3p^2$, so we get

$$\mathrm{Var}_p(Z) \geq \frac{9p^4}{I_X(p)} = 9p^4\frac{p(1-p)}{n^2}.$$

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 $W_n := \mathbf{E}_\lambda(Y \mid \sum_{i=1}^n X_i)$.

Simplify your answer to the best of your ability.

(Hint: a sum of i.i.d. Poisson random variables is also Poisson.)

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

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

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

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

That is,

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

That is,

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