4th problem set,

1. Suppose $X_1, \ldots, X_n \sim \text{i.i.d. } N(\mu, \sigma^2)$.

   (a) Find the mean squared error of

   $$\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \quad \text{and of} \quad \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

   as estimators of σ^2, and state which is smaller.

   (b) Show that among estimators of σ^2 of the form $c\sum_{i=1}^{n}(X_i - \overline{X})^2$ the one with the smallest mean squared error is neither the one in which $c = 1/(n-1)$ nor the one in which $c = 1/n$. Find the value of c for that estimator.

2. Suppose $X_1, \ldots, X_n \sim \text{i.i.d. } N(\mu, \sigma^2)$. To say that $\overline{X}_n = (X_1 + \cdots + X_n)/n$ is a sufficient statistic for μ when σ is known, means that the conditional distribution of $X_1, \ldots, X_n$ given $\overline{X}_n$ is the same regardless of the value of μ. That implies in particular that the conditional distribution of X_1 given $\overline{X}_n$ is the same regardless of the value of μ. By following the steps below, you will show that that is true by explicitly finding the conditional distribution of X_1 given $\overline{X}$. You may assume that the pair $(X_1, \overline{X}_n)$ has a bivariate normal distribution.

   (a) Fill in these blanks: $\begin{bmatrix} X_1 \\ \overline{X}_n \end{bmatrix} \sim N_2\left(\begin{bmatrix} ? \\ ? \end{bmatrix}, \begin{bmatrix} ? & ? \\ ? & ? \end{bmatrix}\right)$.

   (b) Use your results from (a) to fill in these blanks: $X_1 \mid \overline{X}_n \sim N_1(?, ?)$. (Of course, if you see the letter "μ" in your bottom-line answer to this one, then you know something's wrong; that's the point!)

3. Suppose the probability distribution of X depends on θ, and $T(X)$ is a sufficient statistic for θ. Suppose $\delta(X)$ is an estimator of θ.

   (a) Explain why the random variable $\delta_0(X) = E(\delta(X)|T(X))$ that emerges from the Rao-Blackwell process does not depend on θ (if it did depend on θ, then it would not be a statistic, and so could not be used as an estimator). (Although $\delta_0(X)$ does not depend on θ, its probability distribution does generally depend on θ; if that were not true then $\delta_0(X)$ would be useless as an estimator of θ.) (This should be easy since you've already seen this problem done in class. But your being required to write it explicitly, may serve as a useful reminder.)

   (b) Show that if $\delta(X)$ is unbiased, then so is $\delta_0(X)$. (Do this by citing a standard result; a very quick one-line answer suffices.)

4. Suppose $X_1, \ldots, X_n \sim \text{i.i.d. Geo}(P)$. (Recall that X_1 has a geometric distribution with parameter P iff X_1 is distributed as the number of number of independent trials needed to get one success, with probability P of success on each trial, e.g., the number of times you need to throw a die to get a "1".)

CONTINUED⟶

(a) Show that $X_1 + \cdots + X_n$ is sufficient for P.

(b) Let $U = \left\{ \begin{array}{ll} 1 & \text{if } X_1 = 1 \\ 0 & \text{if } X_1 > 1 \end{array} \right\}$. Show that U is an unbiased estimator of P.

(c) Use the results of (a) and (b) and the Rao-Blackwell process to get a better estimator of P.

(d) Find the maximum-likelihood estimator of P.

(e) Show that the sufficient statistic in that problem is complete.

HINTS: If everyone here were majoring in math, I would expect you to recall what I will tell you here explicitly:

(1) The expected value you need to consider is a power series in $q = 1 - p$, i.e., it is of the form $\sum_y \{(y\text{th coefficient}) \cdot q^y\}$.
(2) In any <u>term</u> in the series, any <u>factor</u> that does not depend on q is part of the coefficient.
(3) A power series in q can add up to zero for <u>every</u> value of q only if all of the coefficients are zero.

(f) Draw (and explain!) a conclusion about the <u>best</u> unbiased estimator of P.

(g) Occasionally a statistician will use a bizarre creature called an "improper prior distribution," an idea that we illustrate here. Consider the problem of finding a constant c such that

$$f_P(p) = \begin{cases} c \cdot \dfrac{1}{p(1-p)} & \text{if } 0 < p < 1 \\ \\ 0 & \text{if } p < 0 \text{ or } p > 1 \end{cases}$$

can be used as a prior density of P.

Explain why this is nonsense.

Then pretend you haven't noticed this is nonsense, and you want to use the posterior expectation $\mathbf{E}(P \mid X_1, \ldots, X_n)$ as an estimator of P. Evaluate this estimator.

HINT: You don't need to do any integrals, since you can rely on other things that we have already done. We have a <u>name</u> for the posterior distribution in this case.

5. Suppose $X_1, \ldots, X_n \sim$ i. i. d. and each is uniformly distributed on the finite set $\{1, 2, 3, \ldots, k\}$.

(a) Find the maximum-likelihood estimator of k.

(b) Show that the maximum-likelihood estimator is sufficient.

(c) Show that the maximum-likelihood estimator is complete.

(d) Find a constants a and b such that $aX_1 + b$ is an unbiased estimator of k.

CONTINUED⟶

(e) Use the Rao-Blackwell process to find a better unbiased estimator of k.

(f) Use the reasoning done in parts (a) through (d) to show that the answer to (e) is the best unbiased estimator of k.

6. Suppose $X_1, \ldots, X_n \sim \text{i.i.d.}\,\text{Poisson}(\lambda)$. As observed in class on March 4th,

$$U = \left\{ \begin{array}{ll} 1 & \text{if } X_1 = 0 \\ 0 & \text{if } X_1 > 0 \end{array} \right\}$$

is a very crude unbiased estimator of $P(X_1 = 0) = e^{-\lambda}$. As observed in class on March 1st, $T = X_1 + \cdots + X_n$ is a sufficient statistic for λ. It was asserted near the end of the hour on March 4th, that the Rao-Blackwell estimator $\mathbf{E}(U \mid T)$ is

$$\left(1 - \frac{1}{n}\right)^{X_1 + \cdots + X_n}.$$

Derive this result.

7. See the **CLARIFICATION** at ⟨**http://web.mit.edu/18.441/assignments.html**⟩.