Test II Solution

<table>
<thead>
<tr>
<th>Question</th>
<th>Score</th>
<th>out of:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>100</strong></td>
<td></td>
</tr>
</tbody>
</table>

There are extra blank pages at the end of the test.

NAME: __________________________________________________________
1. Suppose that $Y_1, \ldots, Y_k$ are drawn independently where each $Y_i \sim \text{Poisson}(\beta N_i)$, where $\beta$ is the unknown parameter, while $N_1, \ldots, N_k$ are known (fixed) constants.

Here are some potentially useful facts: the Poisson($\lambda$) distribution has pmf for $k$ an integer $\geq 0$:

$$P(X = k) = \frac{\exp(-\lambda)\lambda^k}{k!},$$

and has mean and variance $\lambda$.

(a) [10 points] Compute the MLE for $\beta$. Compute the bias and the variance of the MLE.

Solution.

Since Poisson distribution belongs to exponential family (and hence, the log-likelihood is a concave function of $\beta$), the MLE is obtained via setting the derivative to zero:

$$\mathcal{L}(\beta) = \sum_{i=1}^{k} (-\beta N_i + Y_i \log(\beta N_i) - \log(Y_i!))$$

$$\mathcal{L}'(\beta) = \sum_{i=1}^{k} \left( -N_i + Y_i \frac{N_i}{\beta N_i} \right) = 0 \iff \hat{\beta}_{\text{MLE}} = \frac{\sum_{i=1}^{k} Y_i}{\sum_{i=1}^{k} N_i}$$

For the mean we use linearity of expectation and that $\mathbb{E}Y_i = \beta N_i$:

$$\mathbb{E} \left( \frac{\sum_{i=1}^{k} Y_i}{\sum_{i=1}^{k} N_i} \right) = \beta$$

and, hence, the estimator is unbiased. For the variance we use standard rules:

$$\mathbb{V} \left( \frac{\sum_{i=1}^{k} Y_i}{\sum_{i=1}^{k} N_i} \right) = \mathbb{V} \left( \frac{\sum_{i=1}^{k} Y_i}{(\sum_{i=1}^{k} N_i)^2} \right) = \frac{\sum_{i=1}^{k} \mathbb{V}(Y_i)}{(\sum_{i=1}^{k} N_i)^2} = \frac{\beta}{\sum_{i=1}^{k} N_i}$$
(b) **[5 points]** An alternative estimator is \( \hat{\beta} = \frac{1}{k} \sum_{i=1}^{k} (Y_i/N_i) \). Compute the bias and the variance of this estimator.

**Solution.**

For the mean we use linearity of expectation and that \( \mathbb{E}Y_i = \beta N_i \):

\[
\mathbb{E} \left( \frac{1}{k} \sum_{i=1}^{k} \frac{Y_i}{N_i} \right) = \frac{1}{k} \sum_{i=1}^{k} \frac{\mathbb{E}Y_i}{N_i} = \frac{1}{k} \sum_{i=1}^{k} \frac{\beta N_i}{N_i} = \frac{1}{k} \sum_{i=1}^{k} \beta = \beta
\]

and, hence, the estimator is unbiased. For the variance we use standard rules:

\[
\text{Var} \left( \frac{1}{k} \sum_{i=1}^{k} \frac{Y_i}{N_i} \right) = \frac{1}{k^2} \sum_{i=1}^{k} \text{Var}(Y_i) \frac{1}{N_i^2} = \frac{\beta}{k^2} \sum_{i=1}^{k} \frac{1}{N_i}
\]

(c) **[10 points]** Compare the two estimators in terms of their mean squared errors, and describe which one you would prefer. If you find this difficult in general, you will get full credit if you try a couple of reasonable, different values for \( \{N_1, \ldots, N_k\} \) and compare the bias and variance of the two estimators for these.

**Solution.**

Recall representation of the MSE in terms of bias and variance. Since both estimators are unbiased, we prefer the one with lower variance. Consider the variance of the MLE:

\[
\frac{\beta}{\sum_{i=1}^{k} N_i}
\]

and observe that since \( f(x) = 1/x \) is a convex function, then we can apply Jensen’s inequality for:

\[
\frac{1}{\sum_{i=1}^{k} N_i} = \frac{1}{k \sum_{i=1}^{k} \frac{1}{N_i}} = \frac{1}{k} f(\mathbb{E}Z_i) \leq \frac{1}{k} \mathbb{E} f(Z_i) = \frac{1}{k^2} \sum_{i=1}^{k} \frac{1}{N_i}
\]

and, hence, the MLE is better in a sense of having lower MSE.
2. (a) **[15 points]** We conduct a multiple testing experiment and would like to test \(d\) null hypotheses \(H_0, \ldots, H_{0d}\). Throughout this question you should assume that we have access to the p-values for each hypothesis, and that these p-values satisfy the usual condition, i.e. under the null

\[
P_0\text{(p-value} \leq t) \leq t, \quad \text{for } 0 \leq t \leq 1.
\]

Our goal is to control the \(k\)-Family Wise Error Rate (\(k\)-FWER) – which is the probability of making \(k\) or more false rejections (imagine \(k > 1\)).

Consider, the following Bonferroni-type procedure.

- We reject each null for which the corresponding p-value \(\leq k\alpha/d\).

Prove that this procedure controls the \(k\)-FWER at \(\alpha\).

**Hint:** As a first step – use Markov’s inequality to relate the \(k\)-FWER to the expected number of false rejections.

**Solution.**

Let \(X_i\) be a binary random variable (indicator) describing whether \(H_i\) is falsely rejected and let \(X\) be a random variable describing the number of false rejections. We have:

\[
X = \sum_{i=1}^{d} X_i
\]

Then by Markov’s inequality:

\[
k\text{-FWER} = \mathbb{P}(X \geq k) \leq \frac{\mathbb{E}X}{k} = \frac{\sum_{i=1}^{d} \mathbb{E}X_i}{k} = \frac{\sum_{i=1}^{d} \mathbb{P}(X_i = 1)}{k} \leq \frac{\sum_{i=1}^{d} \frac{k\alpha}{d}}{k} = \alpha
\]
(b) [10 points] Provide a very short answer (one short sentence) to these questions:

- Compare the procedure you analyzed to the classic Bonferroni procedure (i.e. do you expect the new procedure to make fewer rejections or more?).

Solution.
We would reject more since we reject at higher level \( \frac{k\alpha}{d} \) as opposed to \( \frac{\alpha}{d} \)

- For a given procedure which of the following inequalities do you expect to be true in general?
  - FWER ≤ k-FWER.
  - k-FWER ≤ FWER.

Solution.
Since rejecting at least \( k \) hypotheses yields rejecting at least one, we have:
\[
P(X \geq k) \leq P(X \geq 1)
\]
then the second statement is true.

- Under the global null which of the following inequalities do you expect to be true in general?
  - FDR ≤ k-FWER.
  - k-FWER ≤ FDR.

Solution.
The result follows from equivalence of FDR and FWER under the global null.
3. (a) [15 points] Suppose that $R_{\text{minimax}}$ denotes the minimax risk for estimating some parameter $\theta \in \Theta$, and that for some prior $\pi$ we let $\hat{\theta}_\pi$ denote the corresponding Bayes estimator. You can also assume for convenience that there is some estimator $\hat{\theta}_{\text{minimax}}$ which attains the minimax risk.

Prove the following two claims (explain each step carefully):

$$B_\pi(\hat{\theta}_\pi) \leq R_{\text{minimax}}$$

$$R_{\text{minimax}} \leq \sup_{\theta \in \Theta} R(\hat{\theta}_\pi, \theta).$$

Solution.

For the first inequality,

$$B_\pi(\hat{\theta}_\pi) = \int_{\theta \in \Theta} R(\hat{\theta}_\pi, \theta) \pi(\theta) d\theta$$

$$\leq \int_{\theta \in \Theta} R(\hat{\theta}_{\text{minimax}}, \theta) \pi(\theta) d\theta$$

$$\leq \sup_{\theta \in \Theta} R(\hat{\theta}_{\text{minimax}}, \theta) \quad (iii) \quad = R_{\text{minimax}}$$

where step (i) follows since $\hat{\theta}_\pi$ is a minimizer of the Bayes risk and step (ii) uses the fact that the supremum is always greater than or equal to the average. For step (iii), we use the assumption that $\hat{\theta}_{\text{minimax}}$ achieves the minimax risk.

The second inequality follows directly from the definition of the minimax estimator, that is

$$R_{\text{minimax}} = \inf_{\hat{\theta}} \sup_{\theta \in \Theta} R(\hat{\theta}, \theta) \leq \sup_{\theta \in \Theta} R(\hat{\theta}_\pi, \theta).$$
(b) Suppose we consider a Bernoulli estimation problem, where we toss a coin with bias \( p \), \( n \) times, where \( 0 \leq p \leq 1 \), and consider the estimator that is always 0.5, i.e. \( p = 0.5 \) irrespective of the outcomes. Suppose we are using the squared loss.

- **[5 points]** Show that the estimator above is Bayes under the point mass prior, \( \pi(\theta) = I(\theta = 0.5) \) which puts all the prior mass at 0.5.

**Solution.**

Note that the Bayes risk of \( \hat{\theta} = 0.5 \) is zero under the point mass prior as

\[
B_\pi(0.5) = \int_0^1 (0.5 - \theta)^2 I(\theta = 0.5) d\theta = 0.
\]

In other words, \( \hat{\theta} \) minimizes the Bayes risk; hence it is Bayes.

- **[10 points]** Furthermore, its risk is constant under the prior, i.e. \( \text{R}(0.5, p) = 0 \). So we have an estimator which is Bayes, whose risk is constant. Is the estimator minimax? If yes, explain why and if not prove this (by constructing an estimator with lower worst-case risk).

**Solution.**

No. This is not a minimax estimator. To prove this, consider another estimator

\[
\tilde{p} = \frac{\sum_{i=1}^n X_i + \sqrt{n/4}}{n + \sqrt{n}}.
\]

The maximum risk of \( \tilde{p} \) is always smaller than that of \( \hat{\theta} = 0.5 \) as

\[
\sup_{0 \leq p \leq 1} (p - \tilde{p})^2 = \frac{n}{4(n + \sqrt{n})^2} < \sup_{0 \leq p \leq 1} (p - 1/2)^2 = 1/4.
\]

This proves the claim.
4. Suppose we observe a sample $X \sim p$, where $p$ is a multinomial on $d$ categories. We would like to distinguish the hypotheses:

\[
H_0 : p = p_0 \\
H_1 : p = p_1,
\]

where $p_0$ and $p_1$ are two given (distinct) multinomials. We decide to use the following LRT-based testing procedure. We compute:

\[
T = \frac{\mathcal{L}(p_0)}{\mathcal{L}(p_1)},
\]

and reject the null if $T \leq 1$, and accept the null otherwise.

(a) [20 points] Compute the Type I and Type II errors (individually) of this procedure (this should be a function of the probabilities \{\(p_{01}, \ldots, p_{0d}, p_{11}, \ldots, p_{1d}\)\}).

Solution.

Note that the LRT statistic is

\[
T = \frac{\prod_{i=1}^{d} p_{0i}^{I(X=i)}}{\prod_{i=1}^{d} p_{1i}^{I(X=i)}} = \prod_{i=1}^{d} \left(\frac{p_{0i}}{p_{1i}}\right)^{I(X=i)}.
\]

Based on this representation, the Type I error is

\[
P_{X \sim p_0} (T \leq 1) = \sum_{i=1}^{d} p_{0i} I(p_{0i} \leq p_{1i}).
\]

On the other hand, the Type II error is

\[
P_{X \sim p_1} (T > 1) = \sum_{i=1}^{d} p_{1i} I(p_{0i} > p_{1i}).
\]