Solutions to Practice Test 2

(1) For any $k$:

<table>
<thead>
<tr>
<th>$x$</th>
<th>$\mathbb{P}(X = x)$</th>
<th>$x^k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$1 - p$</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>$p$</td>
<td>1</td>
</tr>
</tbody>
</table>

So $\mathbb{E}(X^k) = (p \times 1) + ((1 - p) \times 0) = p$, $\mathbb{E}(X_i^{2k}) = (p \times 1) + ((1 - p) \times 0) = p$, $\mathbb{V}(X^k) = \mathbb{E}(X_i^{2k}) - (\mathbb{E}(X^k))^2 = p - p^2 = p(1 - p)$. Thus,

$$\mathbb{E} \left( \frac{1}{n} \sum_{i=1}^{n} X_i^k \right) = \mathbb{E}(X^k) = p$$

and

$$\mathbb{V} \left( \frac{1}{n} \sum_{i=1}^{n} X_i^k \right) = \frac{\mathbb{V}(X^k)}{n} = \frac{p(1 - p)}{n}.$$ 

Since $\mathbb{E} \left( \frac{1}{n} \sum_{i=1}^{n} X_i^k \right) = p$ and $\mathbb{V} \left( \frac{1}{n} \sum_{i=1}^{n} X_i^k \right) \to 0$ we have that

$$\frac{1}{n} \sum_{i=1}^{n} X_i^k \xrightarrow{qm} p.$$ 

Since convergence in quadratic mean implies convergence in probability, we also have that

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

(2) Let $W = \overline{X} - \overline{Y}$. Then $\mathbb{E}(W) = 68 - 64 = 4$ and

$$\mathbb{V}(W) = \mathbb{V}(\overline{X}) + \mathbb{V}(\overline{Y}) = \frac{4^2}{100} + \frac{3^2}{100} = \frac{25}{100} = \frac{1}{4}.$$ 

Hence, $rmsd(W) = 1/2$ and by the CLT,

$$W \approx N \left( 4, \frac{1}{4} \right).$$
Therefore,
\[ \mathbb{P}(X > Y) = \mathbb{P}(W > 0) = \mathbb{P} \left( \frac{W - 4}{2} > -\frac{4}{2} \right) = \mathbb{P} \left( \frac{W - 4}{2} > -8 \right) \approx \mathbb{P}(Z > -8) \approx 1. \]

(3) By Markov's inequality
\[ \mathbb{P}(|X_n| > \epsilon) = \mathbb{P}(X_n > \epsilon) \leq \frac{\mathbb{E}(X_n)}{\epsilon} = \frac{\lambda_n}{\epsilon} = \frac{1}{n\epsilon} \to 0 \]
and hence \( X_n \overset{p}{\to} 0 \). Now, for any \( \epsilon > 0 \),
\[
\mathbb{P}(|Y_n| \leq \epsilon) \geq \mathbb{P}(Y_n = 0) = \mathbb{P}(X_n = 0) = \frac{e^{-\lambda_n} \lambda_n^0}{0!} = e^{-\lambda/n} \to 1.
\]
Hence, \( \mathbb{P}(|Y_n| > \epsilon) \to 0 \) and so \( Y_n \overset{p}{\to} 0 \).

(4a) \( f(x; p) = p^x (1 - p)^{1-x} \).
\[ \mathcal{L}(p) = \prod_i f(X_i; p) = p^S (1 - p)^{n-S} \]
where \( S = \sum_i X_i \). Hence
\[ \ell(p) = S \log(p) + (n - S) \log(1 - p) \]
and
\[ \ell'(p) = \frac{S}{p} - \frac{n - S}{1 - p}. \]
So, \( \ell'(\hat{p}) = 0 \) yields \( \hat{p} = S/n \). Now
\[ I(p) = -\mathbb{E} \left( \frac{\partial^2 \log f(X: p)}{\partial p^2} \right) = -\mathbb{E} \left( -\frac{X}{p^2} - \frac{1 - X}{(1 - p)^2} \right) = \frac{1}{p(1 - p)}. \]

(4b) \( \text{se}(\hat{p}) = \sqrt{p(1 - p)/n} \) and \( g(p) = e^p, \ g'(p) = e^p \) so \( \text{se}(\hat{p}) = \text{se}(\hat{p}) |g'(\hat{p})| = \sqrt{p(1 - p)/n e^p} \). Now \( z_{0.025} = 1.96 \) so the confidence interval is
\[ e^{\hat{p}} \pm 1.96 \sqrt{\hat{p}(1 - \hat{p})/n e^{\hat{p}}}. \]
(4c)
Step 1: \( X_1^*, \ldots, X_n^* \sim \text{Bernoulli}(\hat{p}). \)
Step 2: \( \hat{p}^* = n^{-1} \sum_{i=1}^n X_i^*. \)
Step 3: Repeat steps 1 and 2 \( B \) times to get \( \hat{p}_{-B}^* \) where \( \hat{p}_{-B}^* = B^{-1} \sum_{j=1}^B \hat{p}_j^* \).
Step 4: \( \text{se}_{\text{boot}} = \sqrt{ \sum_{i=1}^n I(X_i \leq 0) \text{ on } 0 \text{ and } \hat{F}(1) = \sum_{i=1}^n I(X_i \leq 1) \text{ on } 1. \) So, \( \hat{F}(0) = 1 - \hat{p} \) and \( \hat{F}(1) = 1. \) Thus, the empirical puts mass \( 1 - \hat{p} \) on 0 and mass \( \hat{F}(1) - \hat{F}(0) = \hat{p} \) on 1. So when we draw \( X_i^* \) from \( \hat{F} \), it is the same as drawing from \( \text{Bernoulli}(\hat{p}). \)

(4d) The empirical CDF has \( \hat{F}(0) = \sum_{i=1}^n I(X_i \leq 0) \text{ on } 0 \text{ and } \hat{F}(1) = \sum_{i=1}^n I(X_i \leq 1) \text{ on } 1. \) So, \( \hat{F}(0) = 1 - \hat{p} \) and \( \hat{F}(1) = 1. \) Thus, the empirical puts mass \( 1 - \hat{p} \) on 0 and mass \( \hat{F}(1) - \hat{F}(0) = \hat{p} \) on 1. So when we draw \( X_i^* \) from \( \hat{F} \), it is the same as drawing from \( \text{Bernoulli}(\hat{p}). \)

(5a) Let \( X \) be the number of plants and let \( Y \) be the number that flower. Then \( X \sim \text{Binomial}(n, p) \) and \( Y|X = x \sim \text{Binomial}(x, q). \) Hence
\[
f(x, y; q) = f(x; p)f(y|x; q) = \binom{n}{x} p^x (1-p)^{n-x} \binom{x}{y} q^y (1-q)^{x-y}.
\]
So
\[
\mathcal{L}(p, q) \propto p^x (1-p)^{n-x} q^y (1-q)^{x-y}.
\]
(5b) The log-likelihood is
\[
\ell(p, q) = x \log(p) + (n - x) \log(1-p) + y \log(q) + (x - y) \log(1-q)
\]
and
\[
\frac{\partial \ell}{\partial p} = \frac{x}{p} - \frac{n - x}{1 - p}, \quad \frac{\partial \ell}{\partial q} = \frac{y}{q} - \frac{x - y}{1 - q}.
\]
Setting these equal to 0 yields
\[
\hat{p} = \frac{X}{n} \quad \text{and} \quad \hat{q} = \frac{Y}{X}.
\]
Now \( \mathbb{E}(X) = np \) and \( \mathbb{E}(Y) = \mathbb{E}(Y|X) = \mathbb{E}(qX) = q \mathbb{E}(X) = qnp \) so we solve
\[
\begin{align*}
\frac{n\hat{p}}{n\hat{q}} &= X \\
\frac{n\hat{q}}{n\hat{p}} &= Y
\end{align*}
\]
which yields the same estimator as the MLE.

(5c) The matrix of second derivatives is

\[
H = \begin{pmatrix}
-\frac{X}{p^2} - \frac{n-X}{(1-p)^2} & 0 \\
0 & -\frac{Y}{q^2} - \frac{X-Y}{(1-q)^2}
\end{pmatrix}
\]

and the Fisher information matrix is

\[
I(p, q) = -\mathbb{E}(H) = \begin{pmatrix}
\frac{n}{p(1-p)} & 0 \\
0 & \frac{np}{q(1-q)}
\end{pmatrix}
\]

(5d) With \( g(p, q) = pq \) we have

\[
\nabla g = \begin{pmatrix}
q \\
p
\end{pmatrix}
\]

. So

\[
\text{se}(\psi) = \sqrt{\nabla^T I^{-1} \nabla} = \sqrt{\frac{pq}{n} (1 - pq)}.
\]

The confidence interval is

\[
\hat{pq} \pm z_{1.\alpha} \text{se} = \hat{pq} \pm 1.28 \sqrt{\frac{\hat{pq}}{n} (1 - \hat{pq})}.
\]

(6) For any \( \epsilon > 0 \),

\[
P(|X_n| > \epsilon) = P(X_n^2 > \epsilon^2) \leq \frac{\mathbb{E}(X_n^2)}{\epsilon^2} = \frac{\frac{1}{n} + \frac{1}{n^2}}{\epsilon^2} \to 0
\]

so \( X_n \overset{p}{\to} 0 \).

(7) Proof by contradiction. Assume \( X_n \overset{p}{\to} X \) for some \( X \). Then \( X_n \Rightarrow X \).

Let \( F_n \) be the cdf of \( X_n \) and let \( Z \sim N(0,1) \). Then, for every \( c \),

\[
F_n(c) = \mathbb{P}(X_n \leq c) = \mathbb{P} \left( \frac{X_n - (1/n)}{\sqrt{n}} \leq \frac{1/n}{\sqrt{n}} \right) = \mathbb{P} \left( Z \leq \frac{1/n}{\sqrt{n}} \right) \to \mathbb{P}(Z \leq 0) = \frac{1}{2}.
\]
That is, $F_n(c) \to G(c)$ where $G(c) = 1/2$ for all $c$. But $G$ is not a CDF which contradicts the fact that $X_n$ converges in distribution.

(8) We’ve done this before.

(9) Suppose that $X_n \overset{d}{\to} X$. Let $F_n$ denote the cdf of $X_n$ and let $F$ denote the cdf of $X$. Every non-integer $x$ is a point of continuity of $F$. So, for every integer $k$, $F_n(k + \epsilon) \to F(k + \epsilon)$ for any $0 < \epsilon < 1$. Now,

$$P(X_n = k) = F_n(k + \epsilon) - F_n(k - \epsilon) \to F(k + \epsilon) - F(k - \epsilon) = P(X = k).$$

Now suppose that $P(X_n = k) \to P(X = k)$. Let $x$ be a point of continuity of $F$. Then $x$ is not an integer, so $x = k + \epsilon$ for some integer $k$ and some $0 < \epsilon < 1$.

$$F_n(x) = P(X_n \leq x) = \sum_{j=1}^{k} P(X_n = j) \to \sum_{j=1}^{k} P(X = j) = P(X \leq k) = P(X \leq x) = F(x).$$

Thus, $X_n \overset{d}{\to} X$.

(10) $p(x) = 0$ for all $x$ so $p$ is not a probability function. But $X_n \overset{d}{\to} 0$.

(11) The plug-in estimator is

$$\hat{\psi} = \frac{\left(\frac{1}{n} \sum_{i=1}^{n} X_i^2\right)^{k}}{\left(\frac{1}{n} \sum_{i=1}^{n} X_i\right)}.$$

Step 1: Sample $X_1^*, \ldots, X_n^* \sim \hat{F}_n$.
Step 2. Compute $\psi^*(X_1^*, \ldots, X_n^*)$.
Step 3. Repeat 1 and 2 $B$ times to get: $\hat{\psi}_1^*, \ldots, \hat{\psi}_B^*$. 

Step 4:

\[
\hat{\sigma} = \sqrt{\frac{1}{B} \sum_{j=1}^{B} (\hat{\psi}_j - \bar{\psi}^*)^2}.
\]

(12) Note that

\[
F(x) = \begin{cases} 
0 & x < 0 \\
1 - p & 0 \leq x < 1 \\
1 & x > 1 
\end{cases}
\]

and

\[
\hat{F}(x) = \begin{cases} 
0 & x < 0 \\
1 - \hat{p} & 0 \leq x < 1 \\
1 & x > 1 
\end{cases}
\]

where \(\hat{p} = n^{-1} \sum_{i=1}^{n} X_i\). Therefore,

\[
\max_x |\hat{F}(x) - F(x)| = |(1 - p) - (1 - \hat{p})| = |\hat{p} - p| \overset{p}{\to} 0.
\]

(13) Let \(p = (p_1, \ldots, p_k)\). The likelihood is

\[
\mathcal{L}(p) \propto \prod_{j=1}^{k} p_j^{X_j}
\]

and the log-likelihood is

\[
\ell(p) = \sum_{j=1}^{k} X_j \log p_j.
\]

To maximize this we need to take into account the constraint that \(\sum_{j=1}^{k} p_j = 1\) so we use the method of Lagrange multipliers. We maximize

\[
A(p) = \sum_{j=1}^{k} X_j \log p_j - \lambda (\sum_{j=1}^{k} p_j - 1).
\]
Taking \[
\frac{\partial A(p)}{\partial p_j} = 0
\]
gives \[
\hat{p}_j = \frac{X_j}{\lambda}
\]
Now, the constraint implies that \[
\sum_{j=1}^{n} \hat{p}_j = \frac{\sum_{j=1}^{k} X_j}{\lambda}
\]
hence \(\lambda = n\) and hence \[
\hat{p}_j = \frac{X_j}{n}
\]
To compute the Fisher information matrix, we need to remember that there are only \(k - 1\) free parameters so the matrix is \((k - 1) \times (k - 1)\). Keep in mind that \(p_k = 1 - p_1 - p_2 - \cdots - p_{k-1}\). Hence, for \(j = 1, \ldots, k - 1\),

\[
\frac{\partial \log f(X; p)}{\partial p_j} = \frac{X_j}{p_j} - \frac{X_k}{p_k}
\]

and hence

\[
\frac{\partial^2 \log f(X; p)}{\partial p_j^2} = -\frac{X_j}{p_j^2} - \frac{X_k}{p_k^2}
\]

and

\[
\frac{\partial^2 \log f(X; p)}{\partial p_j \partial p_x} = -\frac{X_k}{p_k^2}
\]

The Fisher information matrix is

\[
n \begin{pmatrix}
\frac{1}{p_1} + \frac{1}{p_k} & \frac{1}{p_1} + \frac{1}{p_k} & \cdots & \frac{1}{p_1} + \frac{1}{p_k} \\
\frac{1}{p_1} + \frac{1}{p_k} & \frac{1}{p_2} + \frac{1}{p_k} & \cdots & \frac{1}{p_2} + \frac{1}{p_k} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{1}{p_1} + \frac{1}{p_k} & \frac{1}{p_k} & \cdots & \frac{1}{p_{k-1}} + \frac{1}{p_k}
\end{pmatrix}.
\]