Test I Solution

<table>
<thead>
<tr>
<th>Question</th>
<th>Score</th>
<th>out of:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>90</strong></td>
<td></td>
</tr>
</tbody>
</table>

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

NAME: __________________________________________________________
1. (a) [5 points] Suppose that we have $X_1, \ldots, X_n$ which are each independent, and have distribution $X_i \sim \text{Bernoulli}(p_i)$ (each $X_i$ has a different distribution). A Bernoulli($p_i$) takes value 1 with probability $p_i$ and 0 with probability $1 - p_i$.

Propose a (shift and) re-scaling of the average:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} X_i,$$

which might converge in distribution to a $N(0, 1)$ distribution.

**Solution.**

Since $X_1, \ldots, X_n$ are independent, the mean and the variance of $\hat{\mu}$ can be computed as

$$\mathbb{E}[\hat{\mu}] = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[X_i] = \frac{1}{n} \sum_{i=1}^{n} p_i \quad \text{and}$$

$$\text{Var}[\hat{\mu}] = \frac{1}{n^2} \sum_{i=1}^{n} \text{Var}[X_i] = \frac{1}{n^2} \sum_{i=1}^{n} p_i(1 - p_i).$$

We then propose a centered and scaled statistic defined by

$$T_n = \frac{\hat{\mu} - \mathbb{E}[\hat{\mu}]}{\sqrt{\text{Var}[\hat{\mu}]}} = \frac{n(\hat{\mu} - \frac{1}{n} \sum_{i=1}^{n} p_i)}{\sqrt{\sum_{i=1}^{n} p_i(1 - p_i)}}.$$

The limiting behavior of $T_n$ depends on $p_1, \ldots, p_n$ but under the setting where $p_1 = \ldots = p_n = p \in (0, 1)$, the usual central limit theorem shows that

$$T_n = \frac{\sqrt{n}(\hat{\mu} - p)}{\sqrt{p(1 - p)}} \xrightarrow{d} N(0, 1).$$
(b) [10 points] Now we would like to appeal to the Lyapunov CLT to give some sufficient conditions for the (re-scaled) average to in fact converge in distribution to a $N(0, 1)$.

Let us recall the version of the Lyapunov CLT that we discussed in lecture. Suppose that,

$$\lim_{n \to \infty} \frac{\sum_{i=1}^{n} E|X_i - E[X_i]|^3}{s_n^3} \to 0,$$

where $s_n^2 = \sum_{i=1}^{n} E(X_i - E[X_i])^2$, then the appropriately re-scaled sum converges to a $N(0, 1)$.

Prove that in our setting $E|X_i - E[X_i]|^3 \leq E(X_i - E[X_i])^2$.

Solution.

Since $X_i \in \{0, 1\}$ and $p_i \in [0, 1]$, it can be seen that $|X_i - p_i| \leq 1$. This implies

$$|X_i - p_i|^3 \leq |X_i - p_i|^2.$$

By taking the expectation on both sides, we obtain the result.

Use this to argue that if $\lim_{n \to \infty} \frac{1}{s_n} = 0$ then the CLT is valid. Prove that it is sufficient for each $p_i$ to be bounded away from 0 and 1 for this condition to hold.

Solution.

From the previous problem,

$$\frac{\sum_{i=1}^{n} E|X_i - E[X_i]|^3}{s_n^3} \leq \frac{\sum_{i=1}^{n} E|X_i - E[X_i]|^2}{s_n^3} = \frac{1}{s_n}.$$

Therefore if $\lim_{n \to \infty} s_n^{-1} = 0$, the Lyapunov condition is satisfied. Assume that each $p_i$ is bounded away from 0 and 1. Then

$$\lim_{n \to \infty} \frac{1}{s_n} = \lim_{n \to \infty} \frac{1}{\sqrt{\sum_{i=1}^{n} p_i(1 - p_i)}} = 0.$$
(c) [5 points] Describe a setting of \( p_1, \ldots, p_n \) for which the CLT does not hold.

**Solution.**

Let \( p_1 = p \in (0, 1) \) and \( p_2 = p_3 = \ldots = p_n = 0 \). In this case, we have the scaled and centered statistic as

\[
T_n = \frac{n(\hat{\mu} - \frac{1}{n} \sum_{i=1}^{n} p_i)}{\sqrt{\sum_{i=1}^{n} p_i(1-p_i)}} = \frac{X_1 - p_1}{\sqrt{p_1(1-p_1)}}.
\]

Since \( X_i \sim \text{Binomial}(1, p) \),

\[
\mathbb{P}(T_n = t) = \begin{cases} 
  p, & \text{if } t = (1-p_1)/\sqrt{p_1(1-p_1)}, \\
  1-p, & \text{if } t = -p_1/\sqrt{p_1(1-p_1)}, \\
  0, & \text{otherwise}.
\end{cases}
\]

This holds for every \( n \). Therefore \( T_n \) does not converge to a normal distribution.

**Alternate solution.**

Assume that \( p_1 = \ldots = p_n = p \) and \( np \to \lambda \in (0, \infty) \) as \( n \to \infty \). In this case, \( \sum_{i=1}^{n} X_i \sim \text{Binomial}(n, p) \). By Poisson limit theorem,

\[
\sum_{i=1}^{n} X_i \xrightarrow{d} \text{Poisson}(\lambda).
\]

Therefore

\[
T_n = \frac{\hat{\mu} - \mathbb{E}[\hat{\mu}]}{\sqrt{\text{Var}[\hat{\mu}]}} \xrightarrow{d} \frac{\text{Poisson}(\lambda) - \lambda}{\sqrt{\lambda}}.
\]
2. Suppose that $X_1, \ldots, X_n \sim N(0, \sigma^2)$, and define the random variable 

$$Z = \max_{i=1}^{n} |X_i|.$$ 

(a) [5 points] Fix an index $i$. Use the (exponential) Gaussian tail bound we derived in class find a value $t$ (ideally the smallest possible value) such that, 

$$\mathbb{P}(|X_i| \geq t) \leq \delta.$$ 

t should be a function of $\delta$ and $\sigma$. 

Solution. 

We know that: 

$$\mathbb{P}(|X_i| \geq t) \leq 2e^{-\frac{t^2}{2\sigma^2}}$$ 

Then by setting the quantity on the right to $\delta$, we get: 

$$2e^{-\frac{t^2}{2\sigma^2}} = \delta \iff t = \sigma \sqrt{2 \ln \left( \frac{2}{\delta} \right)}$$
(b) [10 points] The inequality you gave in the previous part shows that a single Gaussian draw is unlikely to be too large. Use the union bound to find a value \( t \) (ideally the smallest possible value) such that,

\[
P\left( \max_{i=1}^{n} |X_i| \geq t \right) \leq \delta.
\]

\( t \) in this case should be a function of \( \delta, \sigma \) and \( n \).

Comment particularly on how fast \( t \) grows as a function of \( n \). Does it grow logarithmically, polynomially or exponentially?

**Solution.**

Observe that:

\[
P\left( \max_{i=1}^{n} |X_i| \geq t \right) = P(\exists i \in [n]: |X_i| \geq t) = P\left( \bigcup_{i=1}^{n} \{ |X_i| \geq t \} \right) \leq
\]

\[
\leq n \cdot \max_{i=1}^{n} P(|X_i| \geq t) = nP(|X_1| \geq t) \leq 2ne^{-\frac{t^2}{2\sigma^2}}
\]

Then by setting the quantity on the right to \( \delta \), one get:

\[
t = \sigma \sqrt{2 \ln \left( \frac{2n}{\delta} \right)}
\]

Observe that \( t \) depends on number of variables logarithmically.
(c) [10 points] Repeat the previous calculation, except now use Chebyshev’s inequality for the individual $X_i$ (instead of the Gaussian concentration inequality) together with the union bound to find a value $t$ such that,

$$\mathbb{P}(\max_{i=1}^{n} |X_i| \geq t) \leq \delta.$$  

You can use the fact that $X_i$ and $-X_i$ have the same distribution. Comment once again on how fast $t$ grows as a function of $n$. Does it grow logarithmically, polynomially or exponentially?

**Solution.**

Chebyshev inequality gives:

$$\mathbb{P}(|X_1| \geq t) = \mathbb{P}(|X_1 - \mathbb{E}X_1| \geq t) \leq \frac{\sigma^2}{t^2}.$$  

Hence, this way one gets:

$$\mathbb{P}(\max_{i=1}^{n} |X_i| \geq t) \leq n\mathbb{P}(|X_1| \geq t) \leq \frac{n\sigma^2}{t^2}.$$  

Solving for $\delta$ gives:

$$\frac{n\sigma^2}{t^2} = \delta \iff t = \sigma \sqrt{\frac{n}{\delta}}.$$  

Observe that now dependence on $n$ and $\frac{1}{\delta}$ is now polynomial instead of logarithmic.
3. Suppose we observe \( X_1, \ldots, X_n \) which are i.i.d. with a Bernoulli distribution with parameter \( p \). We define the log-odds ratio as the parameter:

\[
\mu = \log \left( \frac{p}{1-p} \right).
\]

Suppose we want to estimate the parameter \( \mu \) and we use the estimator:

\[
\hat{\mu} = \log \left( \frac{\hat{p}}{1-\hat{p}} \right),
\]

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

(a) [10 points] Use the delta method to give a limiting distribution for the estimator \( \hat{\mu} \).

**Solution.**

For this problem, we assume that \( p \) is a fixed constant within \((0, 1)\). First the central limit theorem yields

\[
\sqrt{n}(\hat{p} - p) \xrightarrow{d} N(0, p(1-p)).
\]

Let \( g(x) = \log(x/(1-x)) \) with \( g'(x) = x^{-1}(1-x)^{-1} \). We then apply the delta method to observe

\[
\sqrt{n}(g(\hat{p}) - g(p)) \xrightarrow{d} N(0, p(1-p)\{g'(p)\}^2).
\]

This directly implies that

\[
\sqrt{n}(\hat{\mu} - \mu) \xrightarrow{d} N(0, p^{-1}(1-p)^{-1}).
\]
(b) [15 points] Propose an estimator for the limiting variance of the estimator and use the limiting distribution you have provided to construct an approximate $1 - \alpha$ confidence interval for the parameter $\mu$.

**Solution.**

From the previous problem, we have the limiting variance of $\sqrt{n}(\hat{\mu} - \mu)$ as $p^{-1}(1-p)^{-1}$. To estimate this variance, we propose the plug-in estimator given by

$$\frac{1}{\hat{p}(1 - \hat{p})} = \frac{1}{n} \sum_{i=1}^{n} X_i (1 - \frac{1}{n} \sum_{i=1}^{n} X_i) \overset{p}{\to} \frac{1}{p(1 - p)}.$$

Note that the convergence in probability follows by the weak law of large number and the continuous mapping theorem. Applying Slutsky’s theorem to the previous problem yields

$$\sqrt{n\hat{p}(1 - \hat{p})}(\hat{\mu} - \mu) \overset{d}{\to} N(0, 1).$$

Based on this limiting distribution, an approximate $1 - \alpha$ confidence interval can be given as

$$\mu \in \left( \hat{\mu} - \frac{z_{\alpha/2}}{\sqrt{n\hat{p}(1 - \hat{p})}}, \hat{\mu} + \frac{z_{\alpha/2}}{\sqrt{n\hat{p}(1 - \hat{p})}} \right),$$

where $z_{\alpha/2}$ is the $1 - \alpha/2$ quantile of $N(0, 1)$. 

4. Suppose that we have a sequence of random variables such that

\[ X_n \sim \text{Unif} \left[ \frac{1}{2} - \frac{1}{n}, \frac{1}{2} + \frac{1}{n} \right]. \]

Define the random variable \( X \) to be deterministically equal to 1/2.

(a) \([5 \text{ points}]\) Compute the CDF \( F_{X_n}(1/2) \). Compare it with \( F_X(1/2) \).

**Solution.**

Observe that:

\[
F_{X_n}(1/2) = \mathbb{P}(X_n \leq \frac{1}{2}) = \frac{\frac{1}{2} - (\frac{1}{2} - \frac{1}{n})}{\frac{1}{2} + \frac{1}{n} - (\frac{1}{2} - \frac{1}{n})} = \frac{1}{2} \frac{2}{n} = \frac{1}{2}
\]

However, for \( X = \frac{1}{2} \), we get:

\[
F_X(1/2) = \mathbb{P}(X \leq \frac{1}{2}) = 1
\]

Hence, \( \lim_{n \to \infty} F_{X_n}(1/2) \neq F_X(1/2) \). Note that point 1/2 is not a continuity point of the CDF of \( X \).

(b) \([5 \text{ points}]\) Prove that \( X_n \) converges to \( X \) in quadratic mean.

**Solution.**

One way to show that is to spot that \( X_n \) converges to \( X \) in distribution and, since \( X \) is a constant, in probability. \( X_n \) and \( X \) are bounded random variables and, hence, \( X_n \) converges to \( X \) in quadratic mean. Another way to prove the fact is to show it by definition:

\[
\mathbb{E}(X_n - X)^2 = \mathbb{E}(X_n - \frac{1}{2})^2 = \mathbb{E}(X_n - \mathbb{E}X_n)^2 = \mathbb{V}(X_n) =
\]

\[
= \frac{1}{12} \left( \left( \frac{1}{2} + \frac{1}{n} \right) - \left( \frac{1}{2} - \frac{1}{n} \right) \right)^2 = \frac{1}{3n^2} \xrightarrow{n \to \infty} 0
\]
(c) **[10 points]** Suppose that $X$ is a Bernoulli$(1/2)$ random variable, and we construct a sequence of random variables $X_1, \ldots, X_n$ each equal to $X$. Consider an additional random variable $Z = 1 - X$. Does $X_1, \ldots, X_n$ converge to $Z$ in:

i. Quadratic Mean?

ii. Probability?

iii. Distribution?

Briefly justify your answers.

**Solution.**

Observe that:

$$
E (X_n - Z)^2 = E (X - 1 + X)^2 = E (2X - 1)^2 = \frac{1}{2} (2 - 1)^2 + \frac{1}{2} (1)^2 = 1^n \to 0
$$

If $X_n$ was to converge to $X$ in probability and random variables are bounded, then it would imply that $X_n$ must converge to $X$ in quadratic mean, but this is not the case. Random variable $Z$ clearly has the same distribution as $X_n$’s since $X$ is Bernoulli$(1/2)$, so $X_n$ converge to $Z$ in distribution (their CDF, hence, match).