Biostat 202B Homework 2

Due April 18, 2024 @ 11:59PM

AUTHOR

Hanbei Xiong 605257780

Question 1

Let $X_1, \ldots, X_n$ be iid random variables with mean μ and variance σ^2. Derive the asymptotic distribution of the appropriately normalized sample variance $S_n^2 = \sum_{i=1}^n (X_i - \bar{X})^2/n$ in terms of the kurtosis of the distribution of the X_i's, $\gamma = E(X_1-\mu)^4/\sigma^4$. Carefully and clearly justify each step.

Hint: Start by showing that $nS_n^2 = \sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n (X_i-\mu)^2 - n(\bar{X}-\mu)^2$.

Answer:

$$\begin{aligned}
nS_n^2 &= \sum_{i=1}^n (X_i - \bar{X})^2 \\
&= \sum_{i=1}^n ((X_i-\mu) - (\bar{X}-\mu))^2 \\
&= \sum_{i=1}^n ((X_i-\mu)^2 - 2(X_i-\mu)(\bar{X}-\mu) + (\bar{X}-\mu)^2) \\
&= \sum_{i=1}^n (X_i-\mu)^2 - 2(\bar{X}-\mu)\sum_{i=1}^n (X_i-\mu) + n(\bar{X}-\mu)^2 \\
&= \sum_{i=1}^n (X_i-\mu)^2 - 2n(\bar{X}-\mu)^2 + n(\bar{X}-\mu)^2 \\
&= \sum_{i=1}^n (X_i-\mu)^2 - n(\bar{X}-\mu)^2
\end{aligned}$$

Hence, $S_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i-\mu)^2 - (\bar{X}-\mu)^2$

Let $Y_i = (X_i-\mu)^2$, then $E(Y_i) = \mu_y = \sigma^2$

$Var(Y_i) = E(X_i-\mu)^4 - \{E(X_i-\mu)^2\}^2 = E(X_i-\mu_x)^4 - \sigma^4$

Since $\gamma = \frac{E(X_i-\mu_x)^4}{\sigma^4}$

$Var(Y_i) = \sigma^4(\gamma-1)$

$S_n^2 = \frac{1}{n}\sum_{i=1}^n Y_i - (\bar{X}-\mu)^2 = \bar{Y} - (\bar{X}-\mu)^2$

Consider
$\sqrt{n}(S_n^2 - \sigma_x^2) = \sqrt{n}\{\bar{Y} - (\bar{X} - \mu)^2 - \sigma_x^2\} = \sqrt{n}\{\bar{Y} - (\bar{X} - \mu)^2 - \mu_y\} = \sqrt{n}(\bar{Y} - \mu_y) - \sqrt{n}(\bar{X} - \mu)^2$

By CLT, $\sqrt{n}(\bar{Y} - \mu_y) \xrightarrow{D} \sigma_y Z \sim N(0, \sigma_y^2)$

By CLT, $\sqrt{n}(\bar{X} - \mu) \xrightarrow{D} N(0, \sigma^2)$

By slutsky, $\bar{X} - \mu = \frac{1}{\sqrt{n}}\sqrt{n}(\bar{X} - \mu) \xrightarrow{P} 0$

By C.M.T, $\sqrt{n}(S_n^2 - \sigma_x^2) \xrightarrow{D} N(0, \sigma^4(\gamma - 1))$

Question 2

Let Y_n be a sequence of random variables that satisfies $\sqrt{n}(Y_n - \theta) \to N(0, \sigma^2)$ in distribution. Show that this implies $Y_n \to \mu$ in probability.

Answer:

Since $\frac{1}{\sqrt{n}} \to 0$ as $n \to \infty$, we have

$Y_n - \theta \to 0$ as $n \to \infty$.

$Y_n \to \theta$ in probability. (Note: There is a typo in the question)

Question 3

Let $X_1, \ldots, X_n$ be a random sample from Bernoulli(p). Derive a 95% asymptotic CI for p (based on $\hat{p} = \bar{X}$). Justify your steps completely and carefully.

Answer:

By WLLN, $\bar{X} \xrightarrow{P} p$

$E(X) = p$ and $Var(X) = p(1-p)$

By CLT,

$\frac{\sqrt{n}(\bar{X} - p)}{\sqrt{p(1-p)}} \xrightarrow{D} N(0, 1)$

Since $\frac{p(1-p)}{\sqrt{\bar{X}(1-\bar{X})}} \xrightarrow{P} 1$

$\frac{\sqrt{n}(\bar{X} - p)}{\sqrt{\bar{X}(1-\bar{X})}} \xrightarrow{D} N(0, 1)$

$P(|\frac{\sqrt{n}(\bar{X}-p)}{\sqrt{\bar{X}(1-\bar{X})}}| \leq 1.96) = 0.95$

$P(-1.96 \leq \frac{\sqrt{n}(\bar{X}-p)}{\sqrt{\bar{X}(1-\bar{X})}} \leq 1.96) = 0.95$

$P(\frac{-1.96\sqrt{\bar{X}(1-\bar{X})}}{\sqrt{n}} \leq \bar{X} - p \leq \frac{1.96\sqrt{\bar{X}(1-\bar{X})}}{\sqrt{n}}) = 0.95$

$P(\bar{X} - \frac{1.96\sqrt{\bar{X}(1-\bar{X})}}{\sqrt{n}} \leq p \leq \bar{X} + \frac{1.96\sqrt{\bar{X}(1-\bar{X})}}{\sqrt{n}}) = 0.95$

Hence, 95% asymptotic CI for p is $(\bar{X} - \frac{1.96\sqrt{\bar{X}(1-\bar{X})}}{\sqrt{n}}, \bar{X} + \frac{1.96\sqrt{\bar{X}(1-\bar{X})}}{\sqrt{n}})$

Question 4

Consider a clinical trial comparing a clinical outcome for two independent groups of patients, say treatment and control group. The treatment group yields $X_1, X_2, \ldots, X_n$ samples with mean μ_x and variance σ^2. The control group, yields $Y_1, Y_2, \ldots, Y_m$ samples, with mean μ_y and the variance σ^2 (same as the variance in the treatment group). Using the appropriate theorems characterize the limiting distribution of

$$\frac{(\bar{X}_n - \bar{Y}_n) - (\mu_x - \mu_y)}{\sqrt{S_x^2/n + S_y^2/m}}$$

as n and m go to infinity.[1]

[1] You may assume $\frac{n}{m} \to \gamma \in (0, 1)$.

Answer:

Let $N = n + m, \gamma$ is Kurtosis of X

By WLLN and C.M.T,

$\sqrt{N}\sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}} = \sqrt{\frac{S_X^2}{n/N} + \frac{S_Y^2}{m/N}} \xrightarrow{P} \sqrt{\frac{\sigma_X^2}{\gamma} + \frac{\sigma_Y^2}{1-\gamma}}$

$\sqrt{N}\{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)\} = \sqrt{\frac{N}{n}}\sqrt{n}(\bar{X} - \mu_X) - \sqrt{\frac{N}{m}}\sqrt{m}(\bar{Y} - \mu_Y)$

Then,

$$\begin{aligned} M(t) &= E(e^{t\sqrt{\frac{N}{n}}\sqrt{n}(\bar{X}-\mu_X)-t\sqrt{\frac{N}{m}}\sqrt{m}(\bar{Y}-\mu_Y)}) \\ &= E\{e^{t\sqrt{\frac{N}{n}}\sigma_X\frac{\sqrt{n}(\bar{X}-\mu_X)}{\sigma_X}}\}E\{e^{t\sqrt{\frac{N}{m}}\sigma_Y\frac{\sqrt{m}(\bar{Y}-\mu_Y)}{\sigma_Y}}\} \\ &= M_Z(t\sqrt{\frac{1}{\gamma}}\sigma_X)M_Z(t\sqrt{\frac{1}{1-\gamma}}\sigma_Y) \\ &= e^{\frac{t^2}{2}(\frac{\sigma_X^2}{\gamma}+\frac{\sigma_Y^2}{1-\gamma})}, \forall t \end{aligned}$$

Hence,

$$\sqrt{N}\{\bar{X}-\bar{Y}-(\mu_X-\mu_Y)\} \xrightarrow{D} N(0, \frac{\sigma_X^2}{\gamma}+\frac{\sigma_Y^2}{1-\gamma})$$

By combining the above results,

$$\frac{\bar{X}-\bar{Y}-(\mu_X-\mu_Y)}{\sqrt{\frac{S_X^2}{n}+\frac{S_Y^2}{m}}} \xrightarrow{D} Z \sim N(0,1)$$

Question 5

Suppose $X_1, \ldots, X_n$ are iid from a Poisson distribution with parameter θ.

a. Find the limiting distribution of $\sqrt{n}(\bar{X}-\theta)$.

b. State the limiting distribution of $\sqrt{n}(g(\bar{X})-g(\theta))$, assuming that $g(u)$ is sufficiently smooth.

(a) Answer:

Since $X_i \sim Poisson(\theta)$,

BY WLLN,

$$\frac{\sqrt{n}(\bar{X}-\theta)}{\sqrt{\theta}} \xrightarrow{D} N(0,1)$$

$$\sqrt{n}\{\bar{X}-\theta\} \xrightarrow{D} N(0,\theta)$$

(b) Answer:

By delta method,

$$\sqrt{n}\{g(\bar{X})-g(\theta)\} \xrightarrow{D} N(0, \theta^2(g'(\theta))^2)$$

Question 6

Let X_n be a consistent estimator of θ, with Normal limiting distribution and variance depending only on θ, s.t.

$$\sqrt{n}(X_n - \theta) \xrightarrow{d} N\{0, \sigma^2(\theta)\}.$$

This situation occurs in a number sampling models and defines a couple of drawbacks. Namely, $\sigma(\theta)$ is not constant, inducing additional variability in the limit, and the accuracy of the normal approximation may vary depending on the unknown parameter θ.

These problems can be solved by a *variance stabilizing transformation* $g(X_n)$, where the main task is to find a function $g(t)$ differentiable at $t = \theta$, s.t. $[g\prime(\theta)]^2\sigma^2(\theta) = 1$.

(a) Find a variance stabilizing transformation for $X_1, X_1, \ldots, X_n$ iid Exponential$(1/\theta)$.
(b) Find a variance stabilizing transformation for $X_1, X_1, \ldots, X_n$ iid Poisson(θ).
(c) Find a variance stabilizing transformation for $X_1, X_1, \ldots, X_n$ iid Bernoulli(θ).
(d) Use the result in part b to construct an approximate 95% CI for θ.

(a) Answer:

To ensure $[g'(\theta)]^2\sigma^2(\theta) = 1$, we need $g'(\theta) = \frac{1}{\sqrt{\sigma^2}}$

Since $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} Exponential(\frac{1}{\theta})$

Then, $\sigma^2 = \theta^2$

$g'(\theta) = \frac{1}{\sqrt{\theta^2}} = \frac{1}{\theta}$

$g(\theta) = \ln(\theta)$

Hence, $g(X_n) = \ln(X_n)$ is the variance stabilizing transformation.

(b) Answer:

Similarly, we need $g'(\theta) = \frac{1}{\sqrt{\sigma^2}}$

Since $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} Poisson(\theta)$

Then, $\sigma^2 = \theta$

$g'(\theta) = \frac{1}{\sqrt{\theta}}$

$g(\theta) = 2\sqrt{\theta}$

Hence, $g(X_n) = 2\sqrt{X_n}$ is the variance stabilizing transformation.

(c) Answer:

Similarly, we need $g'(\theta) = \frac{1}{\sqrt{\sigma^2}}$

Since $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} Bernoulli(\theta)$

Then, $\sigma^2 = \theta(1-\theta)$

$g'(\theta) = \frac{1}{\sqrt{\theta(1-\theta)}}$

$g(\theta) = \arcsin(2\theta - 1)$

Hence, $g(X_n) = \arcsin(2X_n - 1)$ is the variance stabilizing transformation.

(d) Answer:

$\sqrt{n}(X_n - \theta) \overset{D}{\to} N(0, \sigma^2)$

After transformation, we have

$\sqrt{n}(2\sqrt{X_n} - 2\sqrt{\theta}) \overset{D}{\to} N(0, 1)$

$P(-1.96 \leq \sqrt{n}(2\sqrt{X_n} - 2\sqrt{\theta}) \leq 1.96) = 0.95$

$P(\frac{-1.96}{\sqrt{n}} \leq 2\sqrt{X_n} - 2\sqrt{\theta} \leq \frac{1.96}{\sqrt{n}}) = 0.95$

$P(\frac{-1.96}{2\sqrt{n}} \leq \sqrt{X_n} - \sqrt{\theta} \leq \frac{1.96}{2\sqrt{n}}) = 0.95$

$P(\sqrt{X_n} - \frac{1.96}{2\sqrt{n}} \leq \sqrt{\theta} \leq \sqrt{X_n} + \frac{1.96}{2\sqrt{n}}) = 0.95$

$P((\sqrt{X_n} - \frac{1.96}{2\sqrt{n}})^2 \leq \theta \leq (\sqrt{X_n} + \frac{1.96}{2\sqrt{n}})^2) = 0.95$