Homework 5

1. Consider a set of N people, each of whom has an opinion about whether Brian should be the chair of a new biostatistics department committee. The opinion of individual i is represented by an indicator: $b_i = 1$ if individual i is for Brian and $b_i = 0$ if individual i is against Brian. $B = \sum_{i=1}^{N} b_i/N$ is the proportion of the population in favor of Brian for chair. To estimate the population proportion, a random sample of n is chosen and polled. The sample is random in the sense that all $\binom{N}{n}$ samples are equally likely. The proportion of those polled who support Brian, $\bar{b} = \sum_{i=1}^{n} b_i/n$, is used to estimate the true population proportion. What is the mean and variance of $\bar{b}$? (Note: this is sampling without replacement from a finite population, different from the usual sampling scheme where we select n observations from an infinite pool. Hint: Define a random variable $I_i = 1$ if individual i is selected into the sample, $I_i = 0$ otherwise.)

2. Suppose $X_1, \cdots, X_n$ are i.i.d. with mean ξ, and suppose that $E|X_i|^k < \infty$, so that the k^{th} central moment $\mu_k = E(X_i - \xi)^k$ exists. Show that the k^{th} sample moment $M_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^k$ converges in probability to μ_k.

3. Let $X_1, ..., X_n$ be a random sample from a population with pdf $f_X(x) = \frac{1}{\theta} I\{0 < x < \theta\}$. Show that $\frac{X_{(1)}}{X_{(n)}} \perp X_{(n)}$.

4. A researcher measures the pulse of n subjects while the subjects are resting and again while the subjects are exercising. For subject i, R_i is the resting pulse and E_i is the pulse measured while exercising. Assume that the pairs (R_i, E_i), $i = 1, ..., n$ are iid draws from the same underlying joint distribution. [Assume that the CLT holds for $\left(\bar{R}, \bar{E}\right)$.]

   (a) The researcher speculates that E is a location-scale transformation of R. If this is true, what is the limiting distribution of the ratio $\bar{R}/\bar{E}$?

   (b) The researcher collects data on the pulse while sleeping, S, for the same n patients. This time the researcher thinks that, for $Y_1, ..., Y_n$ iid $N(0, 1)$, $E_i = R_i + Y_i$ and $S_i = R_i - Y_i$. What is the limiting distribution of $\bar{S}/\bar{E}$?

5. Assuming $\bar{X}$ has a limiting normal distribution, what is the limiting distribution of $\bar{X}^3 - \bar{X}^2$?

Statistical Theory Homework 5

Bohao Tang

September 27, 2017

1. Here B, N, n are all known. Define random variables I_i: $I_i = 1$ if individual i is selected into bthe sample and $I_i = 0$ otherwise.

   N should be large, or we don't need to choose a sample, and the sample number should be reasonable large. So we here don't consider the extreme condition with $N \leq 2$ or $n \leq 2$. We suppose $N > 2$ and $n > 2$.

   Then we have that $\bar{b} = \frac{\sum_{i=1}^{N} b_i I_i}{n}$. Therefore:

$$\mathbb{E}\left[\bar{b}\right] = \frac{1}{n}\sum_{i=1}^{N} b_i \mathbb{E}\left[I_i\right] = \frac{1}{n}\sum_{i=1}^{N}\frac{nb_i}{N} = B$$

   And

$$\begin{aligned}
var\left(\bar{b}\right) &= \mathbb{E}\left[\bar{b}^2\right] - \mathbb{E}\left[\bar{b}\right]^2 = \frac{1}{n^2}\sum_{i,j}^{N} b_i b_j \mathbb{E}\left[I_i I_j\right] - B^2 && (1)\\
&= \frac{1}{n^2}\sum_{i\neq j}^{N} b_i b_j \frac{\binom{N-2}{n-2}}{\binom{N}{n}} + \frac{1}{n^2}\sum_{i}^{N}\frac{nb_i^2}{N} - B^2 && (2)\\
&= \frac{1}{N}\frac{n}{N}\frac{n-1}{N-1}\left(\sum_{i,j}^{N} b_i b_j - \sum_{i}^{N} b_i^2\right) + \frac{B}{n} - B^2 && (3)\\
&= \frac{(n-1)(NB^2 - B)}{n(N-1)} + \frac{B - nB^2}{n} && (4)\\
&= \frac{(N-n)B(1-B)}{n(N-1)} && (5)
\end{aligned}$$

2. *Proof.* Suppose k is a real number bigger than 1, then by mean value theorem:

$$\begin{aligned}
|(X_i - \bar{X})^k - (X_i - \xi)^k| &\leq |\xi - \bar{X}|\left|\max_{t \text{ between } X_i - \bar{X}, X_i - \xi} kt^{k-1}\right| && (6)\\
&\leq |\xi - \bar{X}|k(|X_i - \bar{X}|^{k-1} + |X_i - \xi|^{k-1}) && (7)\\
&\leq |\xi - \bar{X}|k\max(1, 2^{k-2})(2|X_i|^{k-1} + |\bar{X}|^{k-1} + |\xi|^{k-1}) && (8)
\end{aligned}$$

   Here we use the C_r inequality: $|a+b|^r \leq \max(1, 2^{r-1})(|a|^r + |b|^r)$ for $a, b \in \mathbb{R}$ and

$r > 0$. Set $C = k\max(1, 2^{k-2})$, then we have:

$$(\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^k - \frac{1}{n}\sum_{i=1}^{n}(X_i - \xi)^k) \le |\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^k - \frac{1}{n}\sum_{i=1}^{n}(X_i - \xi)^k| \quad (9)$$

$$\le \frac{1}{n}\sum_{i=1}^{n}|(X_i - \bar{X})^k - (X_i - \xi)^k| \quad (10)$$

$$\le C|\bar{X} - \xi|(|\xi|^{k-1} + |\bar{X}|^{k-1} + 2\frac{\sum_{i=1}^{n}|X_i|^{k-1}}{n}) \quad (11)$$

Since $\mathbb{E}\left[|X|^k\right] < \infty$, by Hölder inequality, we have $\mathbb{E}\left[|X|^{k-1}\right] < \infty$. Therefore we have $\frac{\sum_{i=1}^{n}|X_i|^{k-1}}{n} \xrightarrow{p} \mathrm{Ma}_{k-1} = \mathbb{E}\left[|X|^{k-1}\right]$. By continuous mapping theory, we have:

$$|\bar{X} - \xi| \xrightarrow{p} 0 \quad (12)$$

$$(|\xi|^{k-1} + |\bar{X}|^{k-1} + 2\frac{\sum_{i=1}^{n}|X_i|^{k-1}}{n}) \xrightarrow{p} \text{constant.}. \quad (13)$$

By Slutsky Theorem (notice that when convergent to a constant, converge in probability is equvilant to converge in distribution), we have $C|\bar{X} - \xi|(|\xi|^{k-1} + |\bar{X}|^{k-1} + 2\frac{\sum_{i=1}^{n}|X_i|^{k-1}}{n}) \xrightarrow{p} 0$. And then by definition $\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^k - \frac{1}{n}\sum_{i=1}^{n}(X_i - \xi)^k \xrightarrow{p} 0$. Together with $\frac{1}{n}\sum_{i=1}^{n}(X_i - \xi)^k \xrightarrow{p} \mu_k$(weak law of large number) and Slutsky Theorem, we have that $\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^k \xrightarrow{p} \mu_k$. □

3. *Proof.* First, we write out the joint pdf of $X_{(1)}, X_{(n)}$:

$$f_{X_{(1)},X_{(n)}}(x_1, x_n) = n(n-1)\frac{(x_n - x_1)^{n-2}}{\theta^n}\mathbf{1}_{0<x_1\le x_2<\theta}$$

And then we do the transform $U = \frac{X_{(1)}}{X_{(n)}};\ V = X_{(n)}$. We have $X_{(1)} = UV$, $X_{(n)} = V$ and $|\frac{\partial(X_{(1)},X_{(n)})}{\partial(U,V)}| = V$. So that the joint pdf of U, V is:

$$f_{U,V}(u, v) = \frac{n}{\theta^n}v^{n-1}\mathbf{1}_{0<v<\theta} \times (n-1)(1-u)^{n-2}\mathbf{1}_{0<u<1}$$

Notice that $f_{U,V}$ can be factorized in production of function of u and v. Therefore $U \perp V$, which means $\frac{X_{(1)}}{X_{(n)}} \perp X_{(n)}$. □

4. (a) Here, we have $E = aR + b$ for some constant $a > 0, b$.(For specific $a = \sqrt{\frac{var(E)}{var(R)}}$ and $b = \mathbb{E}[E] - a\mathbb{E}[R]$) Suppose $\mathbb{E}[R] = R_e$ and $var(R) = \sigma^2$(of course $aR_e + b$ will not be zero !!). Then we have $\frac{\bar{R}}{\bar{E}} = \frac{\bar{R}}{a\bar{R}+b}$. By delta method, we have $\sqrt{n}(\frac{\bar{R}}{\bar{E}} - \frac{R_e}{aR_e+b}) = \sqrt{n}(\frac{\bar{R}}{a\bar{R}+b} - \frac{R_e}{aR_e+b}) \xrightarrow{D} N(0, \sigma^2\frac{b^2}{(aR_e+b)^4})$.

   (b) Suppose $\mathbb{E}[R] = R_e$, $var(R) = \sigma^2$ and $cov(R, Y) = \rho\sigma$. Then by CLT, we have $\sqrt{n}[(\bar{R}, \bar{Y}) - (R_e, 0)] \xrightarrow{D} N(0, \begin{pmatrix}\sigma^2 & \rho\sigma \\ \rho\sigma & 1\end{pmatrix})$.

   Therefore by delta method, we have:

$$\sqrt{n}(\frac{\bar{S}}{\bar{E}} - 1) = \sqrt{n}(\frac{\bar{R} - \bar{Y}}{\bar{R} + \bar{Y}} - \frac{R_e - 0}{R_e + 0}) \quad (14)$$

$$\xrightarrow{D} N(0, \begin{pmatrix}0 & -2/R_e\end{pmatrix}\begin{pmatrix}\sigma^2 & \rho\sigma \\ \rho\sigma & 1\end{pmatrix}\begin{pmatrix}0 \\ -2/R_e\end{pmatrix}) \quad (15)$$

$$\xrightarrow{D} N(0, \frac{4}{R_e^2}) \quad (16)$$

5. Suppose $\sqrt{n}(\bar{X} - \mu) \xrightarrow{D} N(0, \sigma^2)$. Then by delta method, we have that $(x^3 - x^2)'(\mu) = \mu(3\mu - 2)$ and:

$$\sqrt{n}[(\bar{X}^3 - \bar{X}^2) - (\mu^3 - \mu^2)] \xrightarrow{D} N(0, \sigma^2\mu^2(3\mu - 2)^2)$$