Stat510(Section001): Homework #6

Due on Dec.8, 2021 at 11:59pm

Instructor:Brandon Legried

Tiejin Chen **tiejin@umich.edu**

Problem 1

Part a
Let X_1 be the number of freshmen, X_2 be the number of sophomores,X_3 be the number of juniors, X_4 be the number of seniors..Then(X_1, X_2, X_3, X_4) is a mulitnomial distributiuon with $m = 15$. Hence, We can know that what we want is:

$$P(X_1 + X_2 \geq 8) = P(X_1 + X_2 > 7)$$

And we can know that $X_1 + X_2 \sim Bin(15, 0.26 + 0.24 = 0.5)$. Hence we have:

$$1 - X_1 - X_2 \sim Bin(15, 0.5)$$

$$P(X_1 + X_2 > 7) = 1 - P(X_1 + X_2 \leq 7) = 1 - P(1 - X_1 - X_2 > 7) = 1 - P(X_1 + X_2 > 7)$$

The last equation is because $1 - X_1 - X_2$ and $X_1 + X_2$ have same pdf, and have same cdf. Hence:

$$P(X_1 + X_2 \geq 8) = P(X_1 + X_2 > 7) = \frac{1}{2}$$

the probability that at least eight will be either freshmen or sophomores is 0.5. **Part b**
We have:

$$E(X_3 - X_4) = E(X_3) - E(X_4) = 1.5, Var(X_3 - X_4) = Var(X_3) + Var(X_4) - 2Cov(X_3, X_4) = 7.35$$

Problem 2

It is impossible. Let us prove by contradiction.
Assumed that:

$$P(S = s) = \frac{1}{11}, s \in \{2, .., 12\}$$

LetX_1, X_2 present the number of two dices. We have:

$$P(S = 2) = P(X_1 = 1, X_2 = 1) = p_{1,1}p_{2,1} = \frac{1}{11}$$

$$P(S = 12) = P(X_1 = 6, X_2 = 6) = p_{1,6}p_{2,6} = \frac{1}{11}$$

where $p_{i,j}$ present the probability $P(X_i = j)$. Hence we have:

$$P(X_1 = 1, X_2 = 6) + P(X_1 = 6, X_2 = 1) = \frac{1}{11}(\frac{p_{1,1}}{p_{1,6}} + \frac{p_{1,6}}{p_{1,1}}) \geq \frac{1}{11}$$

$$P(S = 7) = \sum_{X_1 + X_2 = 7} P(X_1, X_2) > P(X_1 = 1, X_2 = 6) + P(X_1 = 6, X_2 = 1)$$

$$P(S = 7) > \frac{1}{11}$$

Hence we get $P(S = 7) \neq \frac{1}{11}$, which is contradict with the assumption. Therefore, it is impossible for dices to be constructed so that $P(S = s) = \frac{1}{11}, s \in \{2, .., 12\}$.

Problem 3

Part a

We have:

$$(n-1)S_n^2 = \sum_{i=1}^{n} X_i^2 - n\bar{X}^2 = \sum_{i=1}^{n} X_i^2 - \frac{1}{n}(X_1 + ... + X_n)^2$$

$$(X_1 + ... + X_n)^2 = \sum_{i=1}^{n} X_i^2 + 2\sum_{i\neq j, i,j<n} X_iX_j$$

$$2n(n-1)S_n^2 = 2(n-1)\sum_{i=1}^{n} X_i^2 - 4\sum_{i\neq j, i,j<n} X_iX_j = \sum_{i,j}(X_i - X_j)^2$$

$$S_n^2 = \frac{1}{2n(n-1)}\sum_{i,j}(X_i - X_j)^2$$

Part b

We have:

$$\begin{aligned} Var(S_n^2) &= Var(\frac{1}{2n(n-1)}\sum_{i,j}(X_i - X_j)^2) \\ &= \frac{1}{n^2(n-1)^2}Var(\frac{1}{2}\sum_{i,j}(X_i - X_j)^2) \\ &= \frac{1}{n^2(n-1)^2}E(\frac{1}{2}\sum_{i,j}(X_i - X_j)^2 - \sigma^2)^2 \end{aligned} \tag{1}$$

For $(\sum_{i,j}(X_i - X_j)^2)^2$, we have three types of terms in this square form:

$$[\frac{1}{2}(X_i - X_j)^2 - \sigma^2][\frac{1}{2}(X_m - X_z)^2 - \sigma^2]$$

where $\{i,j\} \cap \{m,z\} \in \{0,1,2\}$.

If $\{i,j\} \cap \{m,z\} = 0$, We have $[(X_i - X_j)^2 - \sigma^2]$,$[(X_m - X_z)^2 - \sigma^2]$ are independent, and

$$E([\frac{1}{2}(X_i - X_j)^2 - \sigma^2][\frac{1}{2}(X_m - X_z)^2 - \sigma^2]) = E([\frac{1}{2}(X_i - X_j)^2 - \sigma^2])E([\frac{1}{2}(X_m - X_z)^2 - \sigma^2]) = 0$$

If $\{i,j\} \cap \{m,z\} = 1$, We have:

$$E([\frac{1}{2}(X_i - X_j)^2 - \sigma^2][\frac{1}{2}(X_m - X_z)^2 - \sigma^2]) = \frac{\theta_4 - \theta_2^2}{4}$$

If $\{i,j\} \cap \{m,z\} = 2$, We have:

$$E([\frac{1}{2}(X_i - X_j)^2 - \sigma^2][\frac{1}{2}(X_m - X_z)^2 - \sigma^2]) = \frac{\theta_4 + \theta_2^2}{2}$$

We have:

$$Var(S_n^2) = \frac{1}{n^2(n-1)^2}(\frac{4n(n-1)(n-2)(\theta_4 - \theta_2^2)}{4} + \frac{2n(n-1)(\theta_4 + \theta_2^2)}{2}) = \frac{1}{n}(\theta_4 - \frac{n-3}{n-1}\theta_2^2)$$

Part c

We have:

$$Cov(\bar{X}, S_n^2) = E(\bar{X}S_n^2) - E(\bar{X})E(S_n^2) = E((\bar{X} - \mu + mu)S_n^2) - \mu\sigma^2 = E((\bar{X} - \mu)S_n^2)$$

Problem 3 continued on next page. . .

Let $Y_i = X_i - \mu$. We $Var(Y_i) = Var(X_i)$,And we have:

$$\begin{aligned} Cov(\bar{X}, S_n^2) &= E(\bar{Y}S_Y^2) \\ &= \frac{1}{n(n-1)}E(\sum_{i=1}^{n} Y_i(\sum_{i=1}^{n} Y_i^2 - \frac{1}{n}(\sum_{i=1}^{n} Y_i)^2) \\ &= \frac{1}{n(n-1)}(E(\sum_{i=1}^{n} Y_i^3) - \frac{1}{n}E(\sum_{i=1}^{n} Y_i(\sum_{i=1}^{n} Y_i)^2)) \end{aligned} \tag{2}$$

We have:

$$E(\sum_{i=1}^{n} Y_i^3) = n\theta_3$$

$$E(\sum_{i=1}^{n} Y_i(\sum_{i=1}^{n} Y_i)^2) = E(\sum_{i=1}^{n} Y_i(\sum_{i=1}^{n} Y_i^2 + 2\sum i \neq j, ij < nY_iY_j)) = E(\sum_{i=1}^{n} Y_i^3) = n\theta_3$$

Hence:

$$Cov(\bar{X}, S_n^2) = \frac{\theta_3}{n}$$

Therefore, when$\theta_3 = 0$, we have $Cov(\bar{X}, S_n^2) = 0$.

Problem 4

First, we need to consider the joint pdf of $X_{(1)}, X_{(n)}$. Let U(u) be number of sample which is less than u. and V(v) be the number of sample which is less than v but greater than u. Hence, we have the joint cdf of $X_{(i)}, X_{(j)}$ is:

$$F_{X_{(i)}.X_{(j)}}(u, v) = P(U \geq i, U + V \geq j) = \sum_{k=i}^{j} \sum_{m=j-k}^{n-k} P(U = k, V = m)$$

Now, let us consider $P(U, V, n - U - V)$, we can know that, $(U, V, n - U - V)$ is a mulitnomial distribution. And we have:

$$P(U = k, V = m, n - U - V = n - j - m) = \frac{n!}{k!m!(n-k-m)!}(F_X(u))^k(F_X(v) - F_X(u))^m(1 - F_X(v))^{n-k-m}$$

And, we have:

$$P(U = k, V = m) = P(U = k, V = m, n - U - V = n - j - m)$$

Hence, we get the joint pdf of $X_{(i)}, X_{(j)}$ is:

$$f_{X_{(i)},X_{(j)}}(u, v) = \frac{n!}{(i-1)!(j-1-i)!(n-j)!}f_X(u)f_X(v)(F_X(u)))^{i-1}(F_X(v) - F_X(u))^{j-i-1}(1 - F_X(v))^{n-j}$$

And we can get the condition pdf:

$$f_{X_{(1)}|X_{(n)}}(u|y_n) = (n-1)\frac{f_X(u)}{F_X(y_n)}(1 - \frac{F_X(u)}{F_X(y_n)})^{n-2} = (n-1)\frac{e^{-u}}{1 - e^{-y_n}}(1 - \frac{1 - e^{-u}}{1 - e^{-y_n}})^{n-2}$$

Problem 5

Part a

We have:

$$P(Y_n \le y) = P(X_n \le y + logn) = [F_X(y + logn)]^n = (1 - e^{-(y+logn)})^n$$

Hence the cdf of Y_∞ is:

$$F_{Y_\infty}(y) = \lim_{n\to\infty}(1 - \frac{e^{-y}}{e^{logn}})^n = \lim_{n\to\infty}(1 - \frac{e^{-y}}{n})^n = e^{-(e^{-y})}$$

Now we prove this function is actually a cdf, then we can prove Y_n converges in distribution to some random variable. we have:

1. $\lim_{y\to\infty} F_{Y_\infty}(y) = \lim_{z\to 0} e^{-z} = 1$, $\lim_{y\to-\infty} F_{Y_\infty}(y) = \lim_{z\to\infty} e^{-z} = 0$
2. $F'_{Y_\infty}(y) = e^{-(e^{-y}+y)} > 0$ for any y. Hence it is a nondecreasing function of y.
3. it is a continuous function. Therefore it must be a right-continuous function.

We can know that this function is a cdf. We finish our proof. We get the pdf is:

$$f_{Y_\infty}(y) = e^{-(e^{-y}+y)}$$

Part b

To calculate mgf, we have:

$$E(e^{ty}) = \int_{y=-\infty}^{\infty} e^{-(e^{-y}+(1-t)y)}dy$$

Let$z = e^{-y}$, we have $dy = -\frac{dz}{z}, y = -logz$. Hence, we get:

$$\begin{aligned}
E(e^{ty}) &= \int_{y=-\infty}^{\infty} e^{-(e^{-y}+(1-t)y)}dy \\
&= -\int_{z=\infty}^{0} e^{-(z-(1-t)logz)}/zdz \\
&= \int_{z=0}^{\infty} \frac{e^{-z}z^{1-t}}{z}dz \\
&= \int_{z=0}^{\infty} e^{-z}z^{-t}dz
\end{aligned} \tag{3}$$

We can know that this integration is $\Gamma(1-t)$. Hence the mgf of Y_∞ is $\Gamma(1-t)$.
The expection is:

$$\Gamma'(1-t)|_{t=0} = \Gamma'(1) = -\gamma$$

Problem 6

Part a

Proof. Let $p = 1$,We can have:for any $\epsilon^2 > 0$, exists $N > 0$. for any $n > N$ so that $E(|X_n|) < \epsilon^2$. Now, consider $P(|X_n| > \epsilon)$. For any $\epsilon > 0$, and $n > N$,we have:

$$P(|X_n| > \epsilon) \le \frac{E(|X_n|)}{\epsilon} < \frac{\epsilon^2}{\epsilon} = \epsilon$$

Which shows that $P(|X_n|)$ converge to 0 as $n \to \infty$. Hence, we get X_n converges in probability to 0. □

Problem 6 continued on next page. . .

Part b

Proof. Using Chebyshev inequality to get:

$$P(|\bar{X_n} - E(\bar{X_n})| \geq \epsilon) = P(|\bar{X_n} - \mu| \geq \epsilon) \leq \frac{Var(\bar{X_n})}{\epsilon^2} \leq \frac{nC}{n^2\epsilon^2} = \frac{C}{n\epsilon^2}$$

we know that as $n \to \infty, \frac{C}{n\epsilon^2} \to 0$. Hence as $n \to \infty$, $P(|\bar{X_n} - \mu| \geq \epsilon) \to 0$, which proves that sample average converges in probability to μ. □

Problem 7

Part a

for any $s > 0$, we have:

$$P(X \geq t) = P(sX \geq st) = P(e^{sX} \geq e^{st})$$

e^{st} is always greater than 0,Using Markov inequality we can get:

$$P(X \geq t) = P(e^{sX} \geq e^{st}) \leq \frac{E(e^{st})}{e^{st}} = e^{-st}M_X(s)$$

This holds for every $s > 0$, even for $min_{s>0}e^{-st}M_X(s)$ Hence we get:

$$P(X \geq t) \leq min_{s>0}e^{-st}M_X(s)$$

Part b

Let $X_1, .., X_n$ be the iid RVs. and $X_i \sim Bernoulli(0.5)$, we can know that $X = \sum_{i=1}^{n} X_i \sim Bin(n, 0.5)$. Hence $\frac{X}{n}$ can been seen $\bar{X_n}$ here. Now we have:

$$P(|\frac{X}{n} - \frac{1}{2}| \geq \frac{1}{10}) = P(\frac{X}{n} - \frac{1}{2} \geq \frac{1}{10}) + P(\frac{1}{2} - \frac{X}{n} \geq \frac{1}{10})$$

Now we consider $P(X - \frac{n}{2} \geq \frac{1}{10})$, using the conclusion of part a we have:

$$P(X - \frac{n}{2} \geq \frac{1}{10}) \leq e^{-\frac{s}{10}} \prod_{i=1}^{n} E(e^{s(X_i - \frac{1}{2})})$$

$$E(e^{s(X_i - \frac{1}{2})}) = \frac{1}{2}(e^{s/2} + e^{-s/2})$$

Now consider $\frac{1}{2}e^{s/2} + e^{-s/2}$, we have:

$$\frac{1}{2}(e^{s/2} + e^{-s/2}) = e^{ln(\frac{1}{2} + \frac{1}{2}e^s)e^{-\frac{s}{2}}}$$

we let $f(s) = ln(\frac{1}{2} + \frac{1}{2}e^s) - \frac{s}{2}$, Using the Taylor's theorem, there exists $\epsilon \in [0, s]$ so that we can get:

$$f(s) = f(0) + sf'(0) + \frac{1}{2}s^2 f''(\epsilon) = \frac{1}{2}t(1-t)s^2 \leq \frac{1}{8}s^2$$

Hence:

$$\frac{1}{2}(e^{s/2} + e^{-s/2}) \leq e^{s^2/8}$$

We can get:

$$P(X - \frac{n}{2} \geq \frac{1}{10}) \leq e^{-s/10 + ns^2/8}$$

Problem 7 continued on next page. . .

the minimal value of RHS is when $s = \frac{4/10}{n}$,and its value is

$$P(X - \frac{n}{2} \geq \frac{1}{10}) \leq e^{-\frac{1}{50n}}$$

And we can have:

$$P(\frac{X}{n} - \frac{1}{2} \geq \frac{1}{10}) \leq e^{-\frac{n}{50}}$$

For $P(\frac{1}{2} - \frac{X}{n} \geq \frac{1}{10})$ we have the same logic. Hence we can get:

$$P(|\frac{X}{n} - \frac{1}{2}| \geq \frac{1}{10}) \leq 2(e^{-\frac{1}{50}})^n$$

That is to say $C = 2, b = e^{-\frac{1}{50}}$.

Problem 8

Part a

We will use X_i to present the i-th observation and μ to present the actual specific gravity. What we want is:

$$P(|\bar{X_n} - \mu| < \frac{\sigma}{4}) = 1 - P(|\bar{X_n} - \mu| \geq \frac{\sigma}{4})$$

Using Chebyshev inequality, we have:

$$P(|\bar{X_n} - \mu| \geq \frac{\sigma}{4}) \leq \frac{Var(\bar{X_n})}{\sigma^2/16} = \frac{16}{25}$$

Hence:

$$P(|\bar{X_n} - \mu| < \frac{\sigma}{4}) \geq 1 - \frac{16}{25} = \frac{9}{25}$$

the lower bound for the probability is $\frac{9}{25}$.

Part b

Using the central limit theorem, we can know $Z = \frac{5(\bar{X_n} - \mu)}{\sigma} \sim N(0,1)$. And we have:

$$P(|\bar{X_n} - \mu| < \frac{\sigma}{4}) = P(-\frac{\sigma}{4} < \bar{X_n} - \mu < \frac{\sigma}{4}) = P(-\frac{5}{4} < Z < \frac{5}{4}) \approx 0.7887$$

the estimated probability in part(a) is 0.7887.