Econ 704 Discussion Section-Handout 4

Emilio Cuilty,

Handout desingned by: Jian Zhang

cardenascuil@wisc.edu . Office: 7143, Office Hours: Thursdays 9:00-11:00 A.M.

---

1. Review:

- Chebyshev's inequality:

$$\Pr(|X - \mu_X| > \epsilon) \leq \frac{\sigma_X^2}{\epsilon^2} \text{ for } \epsilon > 0$$

  This is a very useful method to estimate the upper bound of a random variable's tail probability and order of a statistics. A more general inequality is the so-called Markov inequality, which states that if $X > 0$ is a random variable, then

$$\Pr(X > \epsilon) \leq \frac{E(X)}{\epsilon} \text{ for } \epsilon > 0.$$

  Proof. $\epsilon \Pr(X > \epsilon) = \epsilon E(1\{X > \epsilon\}) \leq E(X \cdot 1\{X > \epsilon\}) \leq E(X)$

- Law of Large Number: Assume a sequence of i.i.d. random variables $X_1, X_2, \cdots$, if we know $EX_1 = \mu$ and $Var(X_1) = \sigma^2 < \infty$, then $\bar{X}_n = \sum_{i=1}^n X_i/n$ converges in probability to μ.

- Moment Generating Function: $M_X(t) = E(e^{tX})$. Here, we always assume that $M_X(t)$ exists and has finite value.
  1. If $X \sim N(\mu, \sigma^2)$, then $M_X(t) = \exp(\mu t + \frac{1}{2}\sigma^2 t^2)$.
  2. $M_X^{(k)}(0) = EX^k$.
  3. $M_X(t) = M_Y(t)$ for t in a neighbor of zero, then X and Y have the same distribution.
  4. If $M_{X_n}(t) \to M_X(t)$ on a neighbor of zero, then X_n converges to X in distribution.

- Central Limit Theorem: $\{X_i\}$ defined as above. Then $(S_n - n\mu)/\sqrt{n}\sigma$ converges in distribution to $N(0, 1)$. Here, $S_n = X_1 + X_2 + \cdots + X_n$.
  Remark: If S_n takes only integer value, the event $\{S_n < 15\}$, $\{S_n \leq 15\}$, $\{S_n \geq 15\}$ and $\{S_n > 15\}$ often convert to $\{S_n \leq 14.5\}$, $\{S_n \leq 15.5\}$, $\{S_n \geq 14.5\}$ and $\{S_n \geq 15.5\}$ respectively before applying CLT.

2. Problems

1. **(Convolution)** If $X, Y, Z \sim exponential(\lambda)$ are independent, try to calculate the density function of $W = X + Y + Z$.

   For $w > 0$,

$$\begin{aligned} f_W(w) &= \int_{\infty}^{\infty} dz \int_{-\infty}^{\infty} f_X(w - y - z) f_Y(y) f_Z(z) dy \\ &= \int_0^w dz \int_0^{w-z} \lambda^3 e^{-\lambda(w-y-z)} e^{-\lambda y} e^{-\lambda z} dy \\ &= \lambda^3 e^{-\lambda w} \int_0^w (w - z) dz = \frac{\lambda^3 w^2}{2} e^{-\lambda w}. \end{aligned}$$

   For $w \leq 0$, $f_W(w) = 0$.

2. **(Chebyshev's inequalities)** Are the following statements correct or not? (Assume that all random variables below have finite variance)

   (a) The chance that the distance between a random variable and its mean is larger than 3 standard deviations is less than 10%.

   For $w > 0$,
   True. By the Chebyshev's inequality, $\Pr(|X - \mu_X| > 3\sigma) \leq \sigma^2/(3\sigma)^2 = 1/9 \approx 0.11 < 12\%$.

   (b) If a random variable X has mean $EX = 2$ and variance $Var(X) = 1$, then the probability that X is positive is higher than 0.75.

   True. Note that $\{|X-2| < 2\} = \{0 < X < 4\} \subseteq \{X > 0\}$. By Chebyshev's inequality,

$$\Pr(X > 0) \geq \Pr(|X-2| < 2) = 1 - \Pr(|X-2| \geq 2) \geq 1 - \frac{1}{2^2} = 0.75 > 0.70.$$

3. **(Moment Generating Functions)** Prove the following statements:

   (a) X is a random variable with moment generating function $M_X(t)$ and $x > 0$, show that for all $t > 0$,

$$\Pr(X > x) \leq e^{-xt} M_X(t).$$

   Since $x > 0$ and $t > 0$, we have $\{X > x\} = \{tX > tx\} = \{e^{tX} > e^{tx}\}$. Apply the Markov inequality to e^{tX}, it shows that

$$\Pr(X > x) = \Pr(e^{tX} > e^{tx}) \leq \frac{E(e^{tX})}{e^{tx}} = e^{-tx} M_X(t).$$

   (b) If X is a standard normal random variable and $x > 0$, show that

$$\Pr(|X| > x) \leq 2e^{-x^2/2}.$$

   Since standard normal distribution is symmetry, then $\Pr(|X| > x) = \Pr(X < -x) + \Pr(X > x) = 2\Pr(X > x)$. Use the conclusion in Part (a), it can be derived that

$$\Pr(X > x) \leq e^{-xt} M_X(t) = e^{-xt + \frac{1}{2}t^2}$$

   holds for all positive value of t. Specifically, we choose $t = x$, then above inequality becomes that $\Pr(X > x) \leq e^{-1/2x^2}$. Therefore,

$$\Pr(|X| > x) \leq 2\Pr(X > x) \leq 2e^{-\frac{1}{2}x^2}.$$

4. **(Moment Generating Functions)** If standard normal random variables $X_1, X_2, \cdots, X_n$ are independent, Use moment generating functions to show that $\bar{X}_n = \sum_{i=1}^n X_i/n$ follows a normal distribution with mean 0 and variance $1/n$.

   Calculate $M_{\bar{X}_n}(t)$:

$$M_{\bar{X}_n}(t) = E(e^{t\bar{X}_n}) = E(e^{t\frac{X_1+X_2+\cdots+X_n}{n}}) = \prod_{i=1}^n E(e^{\frac{t}{n}X_i}) = \prod_{i=1}^n M_{X_i}(t/n) = \left(e^{\frac{1}{2}\frac{t^2}{n^2}}\right)^n = e^{\frac{1}{2}\frac{1}{n}t^2}$$

   Therefore, $\bar{X}_n \sim N(0, 1/n)$.

5. **(Moment Generating Functions)** If $X \sim Binomial(n,p)$, calculate its moment generating functions $M_X(t)$ and $E(X^3)$.

   X can be view as the sums of Bernoulli trials, say, $X = Z_1 + Z_2 + \cdots + Z_n$ with i.i.d. $\{Z_i\} \sim Bernoulli(p)$. Therefore,

$$M_X(t) = \prod_{i=1}^n M_{Z_i}(t) = (1-p+pe^t)^n.$$

   Then we calculate its derivatives:

$$M'_X(t) = np(1-p+pe^t)^{n-1}e^t$$

$$M''_X(t) = n(n-1)p^2(1-p+pe^t)^{n-2}e^t + np(1-p+pe^t)^{n-1}e^t$$

$$M'''_X(t) = n(n-1)(n-2)p^3(1-p+pe^t)^{n-3}e^t + 2n(n-1)p^2(1-p+pe^t)^{n-2}e^t + np(1-p+pe^t)^{n-1}e^t$$

   Therefore, $E(X^3) = M'''_X(0) = n(n-1)(n-2)p^3 + 2n(n-1)p^2 + np$

6. **(LLN, Challenging)** Suppose that $\{X_i\}$ is a sequence of Bernoulli trials with distribution $Bernoulli(3/4)$ and $\{Y_i\}$ is another sequence of Bernoulli trials with distribution $Bernoulli(1/4)$. Denote $S_n = \sum_{i=1}^n X_i$ and $T_n = \sum_{i=1}^n Y_i$. Show that $\Pr(S_n \neq T_n) \to 1$ as $n \to \infty$.

   Note that

$$\{S_n \neq T_n\} = \{\frac{S_n}{n} \neq \frac{T_n}{n}\} \supseteq \{\left|\frac{S_n}{n} - \frac{3}{4}\right| < \frac{1}{4}, \left|\frac{T_n}{n} - \frac{1}{4}\right| < \frac{1}{4}\}$$

   Therefore, it can be derived that

$$\begin{aligned} 1 \geq \Pr(S_n \neq T_n) &\geq \Pr\left(\left|\frac{S_n}{n} - \frac{3}{4}\right| < \frac{1}{4}, \left|\frac{T_n}{n} - \frac{1}{4}\right| < \frac{1}{4}\right) \\ &\geq \Pr\left(\left|\frac{S_n}{n} - \frac{3}{4}\right| < \frac{1}{4}\right) + \Pr\left(\left|\frac{T_n}{n} - \frac{1}{4}\right| < \frac{1}{4}\right) - 1 \\ &\to 1 + 1 - 1 = 1 \end{aligned}$$

   The limitation of the last line of above inequality follows from Law of Large Number. The second inequality is due to the inequality $\Pr(A \cap B) \geq \Pr(A) + \Pr(B) - 1$. Therefore, $\Pr(S_n \neq T_n) \to 1$.

7. **(CLT)** Suppose there is 50 multiple choice questions in 704 midterm, each of these questions has 4 choices and the correct answer has 2 points. If Ash just randomly choose his answer, then what is the probability that he get more than 50 points in this exam?

Denote the number of questions that Ash have correctly solved by X. Then $X \sim Binomial(50, 0.25)$. Therefore, the event $\{$Ash get more than 50 points$\} = \{X > 25\} = \{X \geq 25.5\}$. By CLT, we have

$$\Pr(X \geq 25.5) = \Pr(\frac{X - 0.25 \cdot 50}{\sqrt{50 \cdot 0.25(1-0.25)}} \geq \frac{25.5 - 0.25 \cdot 50}{\sqrt{50 \cdot 0.25(1-0.25)}}) \approx \Pr(Z \geq \frac{52}{\sqrt{150}}) = 0.001\%$$

Here, $Z \sim N(0,1)$. Thus, this probability is just about 0.001%.

8. **(LLN, CLT)** Suppose that you roll a fair die n times, the results are X_1, X_2, X_3, $\cdots X_n$. Let $S_n = X_1 + X_2 + X_3 + \cdots + X_n$. Show that

(a) $\Pr(3n < S_n < 4n) \to 1$ as $n \to \infty$.

Note that $\{3n < S_n < 4n\} = \{|S_n/n - 3.5| < 0.5\}$ and $EX_i = \frac{1+2+\cdots+6}{6} = 3.5$. By LLN, it has

$$\Pr(3n < S_n < 4n) = \Pr(|S_n/n - 3.5| < 0.5) \to 1.$$

(b) $\lim_{n\to\infty} \Pr(3.5n - \sqrt{n} < S_n < 3.5n + \sqrt{n}) < 1$.

Note that $\mu = 3.5$ and $\sigma^2 = \frac{1}{6}(1-3.5)^2 + \cdots + \frac{1}{6}(6-3.5)^2 = \frac{35}{12}$. By CLT, it can be derived

$$\Pr(3.5n - \sqrt{n} < S_n < 3.5n + \sqrt{n}) = \Pr(-\sigma^{-1} < \frac{S_n - 3.5n}{\sqrt{n}\sigma} < \sigma^{-1}) \to \Pr(-\sigma^{-1} < Z < \sigma^{-1}) < 1.$$

Here, $Z \sim N(0,1)$.