# MTL108: Problem Set-6 with hints/solutions

IIT Delhi

1. **Coin Tossing.** A fair coin is tossed 10,000 times. Let $\overline{X}$ be the proportion of heads. Use Chebyshev's inequality to bound the probability that $\overline{X}$ differs from 0.5 by more than 0.01.

   **Solution:** Let $X_i$ be 1 if heads, 0 if tails. Then $\mathbb{E}[X_i] = 0.5$, $\mathrm{Var}(X_i) = 0.25$. By Chebyshev's inequality,
   $$P(|\overline{X} - 0.5| \geq 0.01) \leq \frac{\mathrm{Var}(\overline{X})}{0.01^2} = \frac{0.25/n}{0.0001} = \frac{2500}{n}.$$
   For $n = 10000$, this bound is $2500/10000 = 0.25$. So the probability is at most 0.25. This guarantees that this probability goes to 0 as $n$ increases.

2. **Dice Rolling.** A fair six-sided die is rolled $n$ times. The average of the rolls is $\overline{X}_n$. Use the WLLN to show that $\overline{X}_n$ converges in probability to 3.5. For $n = 1000$, find an upper bound for $P(|\overline{X}_n - 3.5| > 0.1)$ using Chebyshev.

   **Solution:** For a fair die, $\mathbb{E}[X_i] = 3.5$, $\mathrm{Var}(X_i) = \frac{35}{12} \approx 2.9167$. Here $\overline{X}_n = \frac{X_1 + \cdots + X_n}{n}$, where $X_1, \dots, X_n$ are IID RVs with mean 3.5 and variance 35/12. So, using WLLN for IID case, we have as $n \to \infty$
   $$\overline{X}_n = \frac{X_1 + \cdots + X_n}{n} \xrightarrow{p} \mathbb{E}[X_i] = 3.5.$$
   Next, by Chebyshev's inequality,
   $$P(|\overline{X}_n - 3.5| > 0.1) \leq \frac{\mathrm{Var}(\overline{X}_n)}{0.1^2} = \frac{2.9167/n}{0.01} = \frac{291.67}{n}.$$
   For $n = 1000$, bound $\approx 0.2917$.

3. **Polling.** In a large population, 60% support candidate A. A pollster interviews $n$ randomly chosen voters. Let $\hat{p}$ be the sample proportion supporting A. Use Chebyshev to find how large $n$ must be so that $P(|\hat{p} - 0.6| \geq 0.03) \leq 0.05$.

   **Solution:** We have $\mathbb{E}[\hat{p}] = 0.6$, $\mathrm{Var}(\hat{p}) = \frac{0.6 \times 0.4}{n} = \frac{0.24}{n}$. Chebyshev's inequality gives
   $$P(|\hat{p} - 0.6| \geq 0.03) \leq \frac{0.24/n}{0.03^2} = \frac{0.24}{0.0009\, n} = \frac{266.67}{n}.$$
   Set $\frac{266.67}{n} \leq 0.05 \Rightarrow n \geq \frac{266.67}{0.05} = 5333.4$. So $n \geq 5334$.

4. **Insurance.** An insurance company sells policies with expected claim \$500 and standard deviation \$800. They sell $n$ independent policies. Let $\overline{X}$ be the average claim. The WLLN assures that the average claim is close to \$500 for large $n$. For $n = 10,000$, find a Chebyshev's bound on $P(|\overline{X} - 500| > 50)$.

   **Solution:** $\text{Var}(\overline{X}) = \frac{800^2}{n} = \frac{640000}{n}$. For $n = 10000$, $\text{Var}(\overline{X}) = 64$. Chebyshev's inequality gives

   $$P(|\overline{X} - 500| > 50) \leq \frac{64}{50^2} = \frac{64}{2500} = 0.0256.$$

   So probability is at most 2.56%.

5. **Manufacturing.** A factory produces items with a defect rate of 2%. Each day, they inspect 5000 items. Let $\hat{p}$ be the sample defect rate. Find an upper bound for $P(|\hat{p} - 0.02| > 0.005)$ using Chebyshev's inequality.

   **Solution:** $\text{Var}(\hat{p}) = \frac{0.02 \times 0.98}{5000} = \frac{0.0196}{5000} = 3.92 \times 10^{-6}$. Chebyshev's inequality gives

   $$P(|\hat{p} - 0.02| > 0.005) \leq \frac{3.92 \times 10^{-6}}{(0.005)^2} = \frac{3.92 \times 10^{-6}}{2.5 \times 10^{-5}} = 0.1568.$$

   So the probability is at most about 15.7%.

6. **Stock Returns.** A stock has expected daily return 0.1% and standard deviation 1.5%. An investor holds the stock for 250 trading days. Let $\overline{R}$ be the average daily return over these days. Assuming daily returns are IID, use Chebyshev's inequality to bound $P(|\overline{R} - 0.001| > 0.002)$.

   **Solution:** Here $\mu = 0.001$, $\sigma = 0.015$, $n = 250$. $\text{Var}(\overline{R}) = \frac{0.015^2}{250} = \frac{0.000225}{250} = 9 \times 10^{-7}$. Chebyshev's inequality gives

   $$P(|\overline{R} - 0.001| > 0.002) \leq \frac{9 \times 10^{-7}}{(0.002)^2} = \frac{9 \times 10^{-7}}{4 \times 10^{-6}} = 0.225.$$

   So the probability is at most 22.5%.

7. **Light Bulb Lifetimes.** Let light bulbs have an exponentially distributed lifetime with mean 1000 hours. A room has 100 such bulbs. Let $\overline{X}$ be the average lifetime of these bulbs. Use Chebyshev's inequality to bound $P(|\overline{X} - 1000| > 200)$.

   **Solution:** For exponential, mean $\mu = 1000$, variance $\sigma^2 = 1000^2 = 10^6$. For $n = 100$, $\text{Var}(\overline{X}) = 10^6/100 = 10000$. Chebyshev's inequality gives

   $$P(|\overline{X} - 1000| > 200) \leq \frac{10000}{200^2} = \frac{10000}{40000} = 0.25.$$

   So bound is 25%.

8. **Election Poll (different proportion).** In a two-candidate race, suppose the true support for candidate B is 52%. How many voters must be polled so that, with Chebyshev, the probability that the sample proportion differs from 0.52 by more than 0.02 is at most 0.10?

**Solution:** $\mathrm{Var}(\hat{p}) = \frac{0.52\times 0.48}{n} = \frac{0.2496}{n}$. Chebyshev's inequality gives

$$P(|\hat{p} - 0.52| > 0.02) \le \frac{0.2496/n}{0.0004} = \frac{624}{n}.$$

Set $\frac{624}{n} \le 0.10 \Rightarrow n \ge 6240$. So at least 6240 voters.

9. **Measurement Error.** A scale has random measurement error with mean 0 and standard deviation 2 grams. To estimate the true weight of an object, an experimenter takes 36 independent measurements and averages them. Use Chebyshev's inequality to bound the probability that the average error exceeds 0.5 grams in absolute value.

   **Solution:** Here $\mu = 0$, $\sigma = 2$, $n = 36$. $\mathrm{Var}(\overline{X}) = 4/36 = 1/9 \approx 0.1111$. Chebyshev's inequality gives

   $$P(|\overline{X} - 0| > 0.5) \le \frac{1/9}{0.25} = \frac{1}{2.25} \approx 0.4444.$$

   So probability is at most 44.4%.

10. **Call Center.** A call center receives an average of 200 calls per hour with a standard deviation of 30 calls. Over a 10-hour day, let $\overline{X}$ be the average number of calls per hour. Use Chebyshev to bound $P(|\overline{X} - 200| > 20)$.

    **Solution:** $\mathrm{Var}(\overline{X}) = \frac{30^2}{10} = \frac{900}{10} = 90$. Chebyshev:

    $$P(|\overline{X} - 200| > 20) \le \frac{90}{400} = 0.225.$$

    So bound is 22.5%.

11. **Sum of Uniform Random Variables.** Let $X_1, \ldots, X_{100}$ be i.i.d. uniform on $[0, 1]$. Using CLT, approximate $P\left(\sum_{i=1}^{100} X_i > 55\right)$.

    **Solution:** For $U[0, 1]$, $\mu = 0.5$, $\sigma^2 = 1/12$. Then $S = \sum X_i$ has mean $100 \times 0.5 = 50$, variance $100 \times 1/12 = 100/12 \approx 8.3333$, so $\sigma_S = \sqrt{100/12} = 10/\sqrt{12} \approx 2.8868$. By CLT,

    $$P(S > 55) = P\left(\frac{S - 50}{\sigma_S} > \frac{55 - 50}{2.8868}\right) \approx P(Z > 1.732) = 1 - \Phi(1.732).$$

    From normal tables, $\Phi(1.73) \approx 0.9582$, so probability $\approx 0.0418$.

12. **Binomial Approximation.** A fair coin is tossed 400 times. Using CLT approximate the probability of getting between 180 and 220 heads inclusive. Use continuity correction.

    **Solution:** Let $X \sim \mathrm{Binomial}(400, 0.5)$. Mean $\mu = 200$, variance $\sigma^2 = 400 \times 0.5 \times 0.5 = 100$, $\sigma = 10$. Using continuity correction, $P(180 \le X \le 220)$ becomes $P(179.5 \le X \le 220.5)$. Standardizing:

    $$z_1 = \frac{179.5 - 200}{10} = -2.05, \quad z_2 = \frac{220.5 - 200}{10} = 2.05.$$

    Then $P \approx \Phi(2.05) - \Phi(-2.05) = 2\Phi(2.05) - 1$. $\Phi(2.05) \approx 0.9798$, so probability $\approx 0.9596$.

13. **Average of Exponential Variables.** The lifetime of a certain electronic component is exponentially distributed with mean 100 hours. A system uses 64 such components independently. Using CLT approximate the probability that the average lifetime exceeds 110 hours.

    **Solution:** For exponential, mean $\mu = 100$, variance $\sigma^2 = 100^2 = 10000$. For $n = 64$, $\overline{X}$ has mean 100, standard deviation $\sigma/\sqrt{n} = 100/8 = 12.5$. Then

    $$P(\overline{X} > 110) = P\left(Z > \frac{110 - 100}{12.5}\right) = P(Z > 0.8) = 1 - \Phi(0.8) \approx 1 - 0.7881 = 0.2119.$$

14. **Polling Margin of Error.** A pollster wants to estimate the proportion of voters favoring a candidate. How many voters should be sampled so that, with probability 0.95, the sample proportion is within 3% of the true proportion? Use the CLT and the conservative bound $\sigma \leq 1/2$.

    **Solution:** We want $P(|\hat{p} - p| \leq 0.03) = 0.95$. By CLT, $\hat{p} \approx N(p, p(1-p)/n)$. The margin of error is $z_{0.025} \times \sqrt{p(1-p)/n} = 0.03$. For 95% confidence, $z_{0.025} = 1.96$. The worst-case variance is when $p = 0.5$, giving $p(1-p) = 0.25$. Then

    $$1.96 \times \sqrt{\frac{0.25}{n}} = 0.03 \quad \Rightarrow \quad \sqrt{\frac{0.25}{n}} = \frac{0.03}{1.96} \approx 0.015306.$$

    Square both sides: $\frac{0.25}{n} = (0.015306)^2 \approx 0.0002343 \Rightarrow n = \frac{0.25}{0.0002343} \approx 1067.1$. So sample size at least 1068.

15. **Sum of Random Variables.** A grocery store sells bags of apples. The weight of each apple is a random variable with mean 150 g and standard deviation 30 g. A bag contains 50 apples. Using CLT approximate the probability that the total weight of a bag exceeds 7.7 kg (7700 g).

    **Solution:** Total weight $S = \sum_{i=1}^{50} X_i$. Mean $\mathbb{E}[S] = 50 \times 150 = 7500$ g, variance $\text{Var}(S) = 50 \times 30^2 = 50 \times 900 = 45000$, so $\sigma_S = \sqrt{45000} \approx 212.132$ g. Then

    $$P(S > 7700) = P\left(Z > \frac{7700 - 7500}{212.132}\right) = P(Z > 0.943) \approx 1 - \Phi(0.94) = 1 - 0.8264 = 0.1736.$$

    (Using $\Phi(0.94) \approx 0.8264$ from tables.)

16. **Sum of Bernoulli.** A multiple-choice exam has 100 questions, each with 5 choices. A student guesses randomly on each question. Approximate the probability that the student gets at least 30 correct answers. Use continuity correction.

    **Solution:** Let $X \sim \text{Binomial}(100, 0.2)$. Mean $\mu = 20$, variance $\sigma^2 = 100 \times 0.2 \times 0.8 = 16$, $\sigma = 4$. $P(X \geq 30)$ with continuity correction: $P(X \geq 29.5)$. Standardize:

    $$z = \frac{29.5 - 20}{4} = 2.375.$$

    Then $P \approx 1 - \Phi(2.375)$. From tables, $\Phi(2.37) \approx 0.9911$, $\Phi(2.38) \approx 0.9913$; interpolate: $\approx 0.9912$. So probability $\approx 0.0088$.

17. **Average of Poisson.** A website gets an average of 5 hits per minute. Over a 2-hour period (120 minutes), using CLT, approximate the probability that the average number of hits per minute exceeds 5.5.

    **Solution:** Let $X_i$ be hits in minute $i$. Then $X_i \sim \text{Poisson}(5)$, so $\mu = 5$, $\sigma^2 = 5$. For $n = 120$, $\overline{X}$ has mean 5, standard deviation $\sqrt{5/120} = \sqrt{1/24} \approx 0.2041$. Then

$$P(\overline{X} > 5.5) = P\left(Z > \frac{5.5 - 5}{0.2041}\right) = P(Z > 2.45) \approx 1 - \Phi(2.45) = 1 - 0.9929 = 0.0071.$$

18. **Quality Control.** A manufacturing process produces items with a defect rate of 3%. A random sample of 500 items is taken. Using CLT approximate the probability that the sample defect rate exceeds 4%.

    **Solution:** Let $\hat{p}$ be sample proportion. Mean $p = 0.03$, variance $p(1-p)/n = 0.03 \times 0.97/500 \approx 0.0291/500 = 5.82 \times 10^{-5}$, so $\sigma_{\hat{p}} \approx \sqrt{5.82 \times 10^{-5}} \approx 0.00763$. Then

$$P(\hat{p} > 0.04) = P\left(Z > \frac{0.04 - 0.03}{0.00763}\right) = P(Z > 1.31) \approx 1 - \Phi(1.31) = 1 - 0.9049 = 0.0951.$$

19. **Fish Lengths.** The lengths of a certain fish species are normally distributed with mean 30 cm and standard deviation 4 cm. A fisherman catches 25 fish. Using CLT approximate the probability that the average length exceeds 31 cm.

    **Solution:** Since the population is normal, the sample mean is exactly normal, not just approximate. But we still use CLT reasoning. $\overline{X} \sim N(30, 4^2/25) = N(30, 0.64)$, so $\sigma_{\overline{X}} = 0.8$. Then

$$P(\overline{X} > 31) = P\left(Z > \frac{31 - 30}{0.8}\right) = P(Z > 1.25) = 1 - \Phi(1.25) = 1 - 0.8944 = 0.1056.$$

20. **Portfolio Returns.** An investment portfolio consists of 200 independent stocks. Each stock has an expected annual return of 8% and standard deviation 25%. Using CLT approximate the probability that the average return of the portfolio exceeds 10%.

    **Solution:** For each stock, $\mu = 0.08$, $\sigma = 0.25$. For $n = 200$, $\overline{X}$ has mean 0.08, standard deviation $0.25/\sqrt{200} \approx 0.25/14.142 \approx 0.01768$. Then

$$P(\overline{X} > 0.10) = P\left(Z > \frac{0.10 - 0.08}{0.01768}\right) = P(Z > 1.131) \approx 1 - \Phi(1.13) = 1 - 0.8708 = 0.1292.$$

21. Let $X_1, X_2, \ldots, X_n$ be IID with mean $\mu$ and variance $\sigma^2$.

    (a) Use Chebyshev to bound $\mathbb{P}(|\bar{X}_n - \mu| > \varepsilon)$ and find $n$ to make the bound $< 0.01$.

    (b) Use CLT to find minimum value of $n$ to such that $\mathbb{P}(|\bar{X}_n - \mu| > \varepsilon) < 0.01$.

    (c) If the underlying distribution is Binomial$(10, 0.4)$ and $\epsilon = 0.05$, then what are minimum possible value of $n$ in (a) and (b). Comment on your observations.

**Solution.** For $X \sim \text{Binomial}(10, 0.4)$ (this is the distribution of each $X_i$):

$$\mu = 10 \cdot 0.4 = 4, \qquad \sigma^2 = 10 \cdot 0.4 \cdot 0.6 = 2.4, \qquad \sigma = \sqrt{2.4} \approx 1.549193338.$$

**(a) Chebyshev bound:**

$$n > \frac{100\,\sigma^2}{\varepsilon^2} = \frac{100 \times 2.4}{(0.05)^2} = \frac{240}{0.0025} = 96{,}000.$$

So Chebyshev requires

$$n > 96{,}000 = 96{,}000.$$

**(b) CLT approximation:**

$$n \geq \left(\frac{2.575829 \cdot \sigma}{\varepsilon}\right)^2 = \left(\frac{2.575829 \cdot 1.549193338}{0.05}\right)^2.$$

Compute numerator: $2.575829 \times 1.549193338 \approx 3.9899$. Divide by 0.05 gives $\approx 79.798$, square gives

$$n \gtrsim 79.798^2 \approx 6{,}367.9.$$

Hence take

$$n \geq \lceil 6{,}368 \rceil = 6{,}368.$$

**Comment.** The CLT-based required $n$ (approx. 6,368) is much smaller than the distribution-free Chebyshev bound (96,000). This illustrates a general fact:

- Chebyshev's inequality is *very conservative* because it makes no distributional assumptions beyond finite variance; it must hold for worst-case distributions with that variance.
- The CLT uses the actual distributional behaviour of averages (approximate normality for moderate-to-large $n$) and therefore yields much sharper (smaller) sample-size requirements in practice.

However, CLT-based answers are asymptotic/approximate: for moderate $n$ the normal approximation may be imperfect (especially for highly skewed or heavy-tailed distributions).