HOMEWORK N° 10

Chapter 5

A-1

$\forall X_i, i \in \{1, ..., n\}$,i.i.d. $X_i \sim \text{Exp}(2).n \to +\infty, \frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow{P} ?$

Solution:

According to the deduction of Khinchin's large number theorem:

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow{P} \mathbb{E}[X_i^2], n \to +\infty$$

And since $\mathbb{E}[X_i^2]$ is known:

$$\mathbb{E}[X_i^2] = \mathbb{E}[X_i]^2 + Var[X_i] = \frac{1}{2}$$

It is confident to say that:

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow{P} \frac{1}{2}, n \to +\infty \qquad \checkmark$$

A-2

Let $\{X_n, n \geq 1\}$ be a sequence of random variables such that $X_n \xrightarrow{P} 3$ as $n \to \infty$. Determine the limit in probability of the following sequences:

1. X_n^2

2. $2X_n - 3$

Solution:

We use the following property:
Theorem (Continuous Mapping Theorem for Convergence in Probability):
If $X_n \xrightarrow{P} a$, and g is a continuous function at a, then

$$g(X_n) \xrightarrow{P} g(a)$$

Let $g(x) = x^2$, which is continuous at $x = 3$. Then:

$$X_n^2 = g(X_n) \xrightarrow{P} g(3) = 9$$

Let $g(x) = 2x - 3$, which is also continuous at $x = 3$. Then:

$$2X_n - 3 = g(X_n) \xrightarrow{P} g(3) = 3$$

Therefore:

$$X_n^2 \xrightarrow{P} 9 \quad ✓$$

$$2X_n - 3 \xrightarrow{P} 3$$

✓

A-3

Let X_1, X_2 be two independent variables with $\mathbb{E}[X_i] = 2, Var[X_i] = 4$. Determine the upper bound of $P(|X_1 - X_2| \geq 4)$ using Chebyshev's inequality.

Solution:

The characteristics of the new variable $Y = X_1 - X_2$ are:

$$\mathbb{E}[Y] = \mathbb{E}[X_1] - \mathbb{E}[X_2] = 0$$

$$Var[Y] = Var[X_1] + Var[X_2] = 8$$

Therefore, according to Chebyshev's inequality:

$$P(|Y| \geq 4) = P(|Y - \mathbb{E}[Y]| \geq 4) \leq \frac{Var[Y]}{4^2} = \frac{1}{2}$$

The upper bound of the given probability is $\frac{1}{2}$.

✓

A-4

Let $X_1, X_2, \cdots, X_{315}$ be independent and identically distributed random variables, with the density function of X_1 given by:

$$f(x) = \begin{cases} \frac{2}{3}x, & 1 < x < 2, \\ 0, & \text{otherwise.} \end{cases}$$

Let Y denote the number of times $\{X_i < 1.5\}$ occurs for $i = 1, 2, \cdots, 315$. Find the approximate value of $P\{Y < 140\}$.

Solution:

Compute the integral:

$$p = P(X_i < 1.5) = \int_1^{1.5} f(x)\,dx = \int_1^{1.5} \frac{2}{3}x\,dx = \frac{5}{12}$$

Y follows a binomial distribution with parameters $n = 315$ and $p = \frac{5}{12}$. For large n, we can approximate Y by a normal distribution:

$$\hat{Y} \sim N(\mu, \sigma^2)$$

where:

$$\mu = np = 315 \cdot \frac{5}{12} = 131.25$$

$$\sigma^2 = np(1-p) = 315 \cdot \frac{5}{12} \cdot \frac{7}{12} = 315 \cdot \frac{35}{144} = \frac{11025}{144} = 76.5625$$

$$\sigma = \sqrt{76.5625} = 8.75$$

Therefore:

$$P(\hat{Y} < 140) = \Phi(\frac{\hat{Y} - \mu}{\sigma}) = \Phi(1) \approx 0.84$$

✓

B-2

A certain genetic disease has an inter-generational incidence rate of 10%. In a study of 500 affected families, use Chebyshev's inequality to estimate the lower bound of the probability that the absolute difference between the observed inter-generational incidence proportion and the true incidence rate is less than 5%.

Solution:

Let X_i be an indicator variable for the i-th family, where $X_i = 1$ if the disease occurs in the next generation, and $X_i = 0$ otherwise.

- The total number of families where the disease occurs in the next generation is $S = \sum_{i=1}^{500} X_i$.

- The observed proportion is $\hat{p} = \frac{S}{500}$.

Each X_i is a Bernoulli random variable with

- $p = 0.10$.
- $\mu = E[S] = 500 \times 0.10 = 50$.
- $\text{Var}(S) = 500 \times 0.10 \times 0.90 = 45$.

We want:

$$P\left(|\hat{p} - p| < 0.05\right) = P\left(|S - \mu| < 25\right)$$

Chebyshev's inequality states:

$$P\left(|S - \mu| \geq k\right) \leq \frac{\text{Var}(S)}{k^2}.$$

For the complementary event:

$$P\left(|S - \mu| < k\right) \geq 1 - \frac{\text{Var}(S)}{k^2}.$$

Here, $k = 25$:

$$P\left(|S - \mu| < 25\right) \geq 1 - \frac{45}{25^2} = 1 - \frac{45}{625} = 1 - 0.072 = 0.928.$$

B-6

Let $\{X_i, i \geq 1\}$ be a sequence of i.i.d. normal random variables, where $X_1 \sim N(\mu, \sigma^2)$ and $\sigma > 0$. Determine whether the following sequences of random variables converge in probability as $n \to +\infty$. If they converge, provide the limit; otherwise, explain why not:

$$\frac{X_1 + X_2 + \cdots + X_n}{\sqrt{n \sum_{i=1}^{n} (X_i - \mu)^2}}$$

Solution:

It is evident that the variable series is a combination of:

$$Y = \frac{\dfrac{1}{n} \displaystyle\sum_{i=1}^{n} X_i}{\sqrt{\dfrac{1}{n} \displaystyle\sum_{i=1}^{n} (X_i - \mu)^2}}$$

The separations of which is known:

$$\frac{1}{n} \sum_{i=1}^{n} X_i \xrightarrow{P} \mu, n \to +\infty$$

$$\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 \xrightarrow{P} \sigma^2, n \to +\infty$$

Hence according to the Continuous Mapping Theorem for Convergence in Probability:

$$Y \xrightarrow{P} \frac{\mu}{\sqrt{\sigma^2}} = \frac{\mu}{\sigma}, n \to +\infty$$

B-7

Let the sequence of random variables $\{X_i, i \geq 1\}$ be independent and identically distributed (i.i.d.), following an exponential distribution with mean $\frac{1}{\lambda}$, where $\lambda > 0$.

(1) If for any $\varepsilon > 0$, the following holds:

$$\lim_{n\to+\infty} P\left\{\left|\frac{X_1^2 + X_2^2 + \cdots + X_n^2}{n} - a\right| < \varepsilon\right\} = 1$$

find the value of a.

(2) Provide the approximate distribution of:

$$\frac{1}{50}\sum_{i=1}^{100} X_i$$

(3) Find the approximate value of:

$$P\left\{\frac{1}{100}\sum_{i=1}^{100} X_i^2 \leq \frac{2}{\lambda^2}\right\}$$

Solution:

1. By the Law of Large Numbers, the sample mean converges in probability to the expectation of $\frac{1}{n}\sum_{i=1}^{n} X_i^2$. Which is:

$$a = \mathbb{E}[\frac{1}{n}\sum_{i=1}^{n} X_i^2] = \mathbb{E}[X_i^2] = \mathbb{E}[X_i]^2 + Var[X_i] = \frac{2}{\lambda^2}$$

2. By the Central Limit theorem, the sum approximate a normal distribution:

$$\frac{1}{50}\sum_{i=1}^{\hat{100}} X_i \sim N(\frac{2}{\lambda}, \frac{1}{25\lambda^2})$$

3. By the Central Limit theorem, the sum approximate an another normal distribution:

$$\frac{1}{100}\sum_{i=1}^{\hat{100}} X_i^2 \sim N(\frac{2}{\lambda^2}, \frac{1}{5\lambda^4})$$

Therefore:

$$P\left\{\frac{1}{100}\sum_{i=1}^{100} X_i^2 \le \frac{2}{\lambda^2}\right\} \approx \Phi(0) = 0.5$$

B-10

A company is celebrating its centennial anniversary and has invited some public figures and relevant individuals to attend the celebration. Each invitee may choose to:

(1) Attend alone (probability: 0.3),

(2) Attend with one companion (probability: 0.5), or

(3) Decline the invitation (probability: 0.2).

If the company sent out 800 invitations, and the attendance decisions of all invitees are independent, what is the probability that the total number of attendees exceeds 1,000?

Solution:

The response attendance numberX_i of any individual letters is i.i.d. and has such a characteristic:

$$\mathbb{E}[X_i] = 1 \times 0.3 + 2 \times 0.5 + 0 = 1.3$$

$$Var[X_i] = 0.61$$

By the Central Limit theorem, their sum approximate a normal distribution:

$$\sum_{i=1}^{800\hat{}} X_i \sim N(1040, 488)$$

Therefore, the probability of the number of attendance exceeding 1000 is:

$$P\left\{\sum_{i=1}^{800} X_i \ge 1000\right\} \approx 1 - \Phi(\frac{1000 - 1040}{\sqrt{488}}) = 0.965$$

B-11

A "Knowledge Competition" has the following rules:

- Each contestant can attempt up to 3 mutually independent questions.
- If a question is answered incorrectly, the contestant is eliminated and cannot proceed to the next question.
- Each correct answer earns 1 point.
- If all 3 questions are answered correctly, an additional bonus point is awarded (total of 4 points).

Now, 100 contestants participate, each answering questions independently.

(1) Probability of "At Most 35 Contestants Score 0 Points"

Given:

- The probability that a contestant scores at least 1 point is 0.7.

- Use the Central Limit Theorem (CLT) to calculate the probability that no more than 35 contestants score 0 points.

(2) Probability of "Total Score Exceeds 220 Points"

Given:

- Each question has a success probability of 0.8 (similar difficulty).

- Calculate the probability that the total score of all 100 contestants exceeds 220 points.

Solution:

1. By the Central Limit Theorem, the number of contestants scoring at least 1 points approximates a normal distribution:

$$\sum_{i=1}^{100} X_i \sim N(100 \times 0.7, 100 \times 0.21)$$

Therefore, the probability that no more than 35 contestants score 0 points is:

$$P\left\{\sum_{i=1}^{100} X_i \geq 65\right\} \approx 1 - \Phi(\frac{65-70}{\sqrt{21}}) = 0.862$$

2. The score Y_i an individual obtained complies:

$$Y_i = \begin{cases} 0, p = 0.2 \\ 1, p = 0.16 \\ 2, p = 0.128 \\ 4, p = 0.512 \end{cases}$$

$$\mathbb{E}[Y_i] = 2.464, Var[Y_i] = 2.793$$

By the Central Limit Theorem, the total score of contestants approximates an another normal distribution:

$$\sum_{i=1}^{100} Y_i \sim N(100 \times 2.464, 100 \times 2.793)$$

Therefore, the probability of the total score exceeding 220 is:

$$P\left\{\sum_{i=1}^{100} Y_i \geq 220\right\} \approx 1 - \Phi(\frac{220-246.4}{\sqrt{279.3}}) = 0.943$$