LIMIT THEOREMS OF RANDOM VARIABLES

A LINEAR COMBINATION OF RANDOM VARIABLES

A.1 INTERPRETING LINEAR COMBINATIONS OF RANDOM VARIABLES

Ex 1: Let X_1 and X_2 be the result of a roll of first die and second die respectively.

1. What does $X_1 + X_2$ represent?
2. What does $2X_1$ represent?
3. What possible values can be taken by:
   (a) $X_1 + X_2$
   (b) $2X_1$

Ex 2: Let X_1, X_2, X_3 be the result of three coin flips, where $X_i = 1$ if the i-th coin lands on Head and 0 otherwise. Let $S_3 = X_1 + X_2 + X_3$.

1. What does S_3 represent?
2. What possible values can be taken by S_3?

Ex 3: Let X_1, X_2, X_3 be the result of three coin flips, where $X_i = 1$ if the i-th coin lands on Head and 0 otherwise. Let $\overline{X}_3 = \frac{X_1+X_2+X_3}{3}$.

1. What does $\overline{X}_3$ represent?
2. What possible values can be taken by $\overline{X}_3$?

Ex 4: Let X_1, X_2, X_3, X_4 be the responses of four people to the question "Do you plan to vote for Candidate A?", where $X_i = 1$ represents "Yes" and 0 represents "No". Let $\overline{X}_4 = \frac{X_1+X_2+X_3+X_4}{4}$.

1. What does $\overline{X}_4$ represent?
2. What possible values can be taken by $\overline{X}_4$?

A.2 FINDING THE PROBABILITY DISTRIBUTION

Ex 5: Let X_1 and X_2 be two independent random variables with the same probability distribution:

$$P(X_i = 10) = \frac{1}{100} \quad \text{and} \quad P(X_i = -1) = \frac{99}{100}$$

1. What are the possible values of $S_2 = X_1 + X_2$?
2. Find the probability distribution of S_2.

Ex 6: Let X_1 and X_2 be the result of two coin flips, where $X_i = 1$ if the i-th coin lands on Head and 0 otherwise (fair coin).Let $\overline{X}_2 = \frac{X_1+X_2}{2}$ be the sample mean.

1. What are the possible values of $\overline{X}_2$?

2. Find the probability distribution of $\overline{X}_2$.

A.3 CALCULATING EXPECTATION AND VARIANCE

Ex 7: Let X_1, X_2, X_3 be the results of flipping a fair coin three times.We define $X_i = 1$ if the coin lands on Heads, and $X_i = 0$ otherwise.This follows a Bernoulli distribution with $p = 0.5$.Let $S_3 = X_1 + X_2 + X_3$ be the total number of Heads.

1. Find the mean and the variance of a single coin flip X_1.

2. Deduce the expectation (mean) of S_3.

3. Deduce the variance of S_3.

4. Find the standard deviation of S_3.

Ex 8: Let X_1 and X_2 be the results of rolling two fair six-sided dice.Each die follows a discrete uniform distribution on $\{1, 2, 3, 4, 5, 6\}$.Let $S_2 = X_1 + X_2$ be the sum of the two dice.

1. Find the mean and the variance of a single die roll X_1.

2. Deduce the mean of S_2.

3. Deduce the variance of S_2.

4. Find the standard deviation of S_2.

Ex 9: Let $X_1, X_2, \dots, X_{20}$ be independent random variables, each with mean $\mu = 174$ and standard deviation $\sigma = 4$.Let $S_{20} = X_1 + X_2 + \cdots + X_{20}$.

1. Find the mean of S_{20}.

2. Find the variance of S_{20}.

3. Find the standard deviation of S_{20}.

Ex 10: The random variables $X_1, X_2, \ldots, X_6$ represent the weight of a cookie, each with a mean of 30 g and a standard deviation of 2 g. We assume the weights are independent.The random variable $S_6 = X_1 + X_2 + \cdots + X_6$ represents the total weight of a packet of 6 cookies.

1. Find the expected total weight (mean) of the packet, $E(S_6)$.
2. Find the standard deviation of the total weight of the packet, $\sigma(S_6)$.

Ex 11: Consider a population where a certain characteristic occurs with probability p (for example, a voter supporting a candidate).
We take a random sample of size n. Let $X_1, X_2, \ldots, X_n$ be independent random variables where $X_i = 1$ if the characteristic is present (success) and $X_i = 0$ otherwise (failure).
The **sample proportion**, denoted by $\hat{P}$, is defined as the sum of successes divided by the sample size:

$$\hat{P} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

1. Recall the mean $E(X_i)$ and variance $V(X_i)$ for a single Bernoulli variable.
2. Using the properties of expectation, determine the mean of the sample proportion, $E(\hat{P})$.
3. Using the properties of variance for independent variables, determine the variance of the sample proportion, $V(\hat{P})$.
4. Deduce the formula for the standard deviation (standard error) of the sample proportion, $\sigma(\hat{P})$.

B LAW OF LARGE NUMBERS

B.1 APPLYING THE LAW OF LARGE NUMBERS

Ex 12: A fair six-sided die is rolled n times. Let X_i be the outcome of the i-th roll, and let $\overline{X}_n$ be the average value of the rolls after n trials.

1. Calculate the theoretical mean μ of a single roll.
2. According to the Law of Large Numbers, what value does $\overline{X}_n$ approach as n becomes very large?
3. If you roll the die 10 times and get an average of 4.2, does this disprove the LLN? Explain.

Ex 13: A basketball player has a theoretical free-throw success rate of $p = 0.8$. We define a random variable $X_i = 1$ if the shot is made and $X_i = 0$ if it is missed.

1. If the player takes $n = 1000$ shots, what does the sample mean $\overline{X}_{1000}$ represent in this context?
2. Using the Law of Large Numbers, predict the value of $\overline{X}_n$ if the player takes an infinite number of shots.

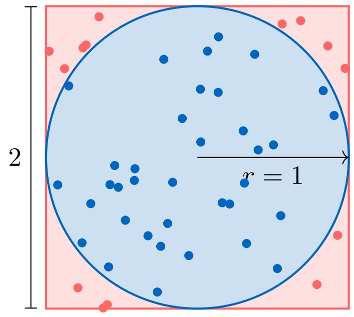
3. Explain why the player might make 5 shots in a row (100% success rate for that short sequence) despite the long-term average being 80%.

Ex 14: Let $X_1, X_2, \ldots$ be independent random variables with mean $\mu = 50$ and variance $\sigma^2 = 100$.Let $\overline{X}_n$ be the mean of the first n variables.

1. Calculate the standard deviation of the sample mean, $\sigma(\overline{X}_n)$, for $n = 1$, $n = 100$, and $n = 10,000$.
2. What happens to $\sigma(\overline{X}_n)$ as $n \to \infty$?
3. How does this result support the Law of Large Numbers?

Ex 15: We conduct a computer simulation to estimate the value of π.
Imagine a square with side length 2 (area = 4) enclosing a circle with radius 1 (area = π).We generate n random points uniformly inside the square. Let X_i be a random variable such that $X_i = 1$ if the point falls inside the circle, and $X_i = 0$ otherwise.

1. Calculate the theoretical probability p that a single random point falls inside the circle.
2. Let $\overline{X}_n$ be the proportion of points inside the circle after n trials. According to the Law of Large Numbers, what value does $\overline{X}_n$ approach as $n \to \infty$?
3. Propose a formula to estimate π using the experimental result $\overline{X}_n$.



Ex 16: An insurance company provides policies for a rare risk. For a single customer, the payout X is a random variable with:

- Mean $\mu = \$500$ (Expected loss per customer).
- Standard Deviation $\sigma = \$10,000$ (Very high volatility because most payouts are 0, but some are huge).

The company has n customers. Let $\overline{X}_n$ be the average payout per customer.

1. If the company has only $n = 1$ customer, what is the standard deviation of the average payout? Is this risky?
2. If the company has $n = 10,000$ customers, calculate the standard deviation of the average payout, $\sigma(\overline{X}_{10000})$.
3. Using the Law of Large Numbers, explain why the insurance business model works despite individual outcomes being unpredictable.

C CENTRAL LIMIT THEOREM

C.1 ANALYZING THE SUM OF NORMAL VARIABLES

Ex 17: The weight of an adult male is a **normally distributed** random variable with mean 80 kg and standard deviation 12 kg. An elevator has a maximum capacity of 850 kg. If 10 adult males enter the elevator, what is the probability that their total weight exceeds the maximum capacity? Assume weights are independent.

Ex 18: The weight of a package is a **normally distributed** random variable with mean 25 kg and standard deviation 3 kg. A small transport truck has a maximum weight limit of 1 300 kg.
If 50 packages are loaded onto the truck, what is the probability that the total weight exceeds the limit? Assume the weights of the packages are independent.

Ex 19: In a 4×400m relay race, the time taken by an individual runner is a **normally distributed** random variable with mean 52 seconds and standard deviation 1.5 seconds.
What is the probability that the team finishes the race in less than 205 seconds? Assume the runners' times are independent.

Ex 20: A coffee machine dispenses a quantity of coffee that is a **normally distributed** random variable with mean 150 ml and standard deviation 5 ml.
If a jug is filled by dispensing 10 cups one after another, what is the probability that the total volume of coffee exceeds 1 525 ml? Assume the cups are independent.

C.2 APPLYING THE CLT TO SUMS OF BERNOULLI TRIALS

Ex 21: An airline knows that passengers who buy a ticket do not always show up for the flight.
For a single ticket sold, let X_i be the number of people showing up (1 if they show up, 0 if they don't).
This variable X_i follows a **Bernoulli distribution** with probability of success $p = 0.9$ (representing a 90% show-up rate).
The airline sells 160 tickets for a flight that has only 150 seats.
Let S_{160} be the total number of passengers who show up.

1. Find the expected number of passengers who show up, $E(S_{160})$.
2. Calculate the standard deviation for a single passenger (σ), and then find the standard deviation of the total number of passengers, $\sigma(S_{160})$.
3. Can we assume the total number of passengers S_{160} is normally distributed? Explain.
4. Find the probability that the flight is overbooked (i.e., more than 150 passengers show up).

Ex 22: A hotel organizes a conference in a hall that has a maximum capacity of 200 seats.
To ensure the hall is full, they send out 220 invitations. Let X_i be the attendance of a single invited person (1 if they attend, 0 if they don't).
This variable X_i follows a **Bernoulli distribution** with probability of success $p = 0.85$ (representing an 85% attendance rate).
Let S_{220} be the total number of guests who attend the conference.

1. Find the expected number of guests, $E(S_{220})$.
2. Calculate the standard deviation for a single guest (σ), and then find the standard deviation of the total number of guests, $\sigma(S_{220})$.
3. Can we assume the total number of guests S_{220} is normally distributed? Explain.
4. Find the probability that there are not enough seats (i.e., more than 200 guests show up).

Ex 23: A farmer plants seeds to grow corn. Due to soil conditions and weather, not every seed germinates (sprouts).
For a single seed planted, let X_i be the result (1 if it germinates, 0 if it dies).
This variable X_i follows a **Bernoulli distribution** with probability of success $p = 0.8$ (an 80% germination rate).
The farmer plants 1000 seeds. Let S_{1000} be the total number of seeds that germinate.

1. Find the expected number of seeds that germinate, $E(S_{1000})$.

2. Calculate the standard deviation for a single seed (σ), and then find the standard deviation of the total number of germinated seeds, $\sigma(S_{1000})$.

3. Can we assume the total number of germinated seeds S_{1000} is normally distributed? Explain.

4. The farmer has a contract to deliver at least 780 corn stalks. Find the probability that he meets this requirement (i.e., $S_{1000} \geq 780$).

C.3 APPLYING THE CLT TO CONTINUOUS DISTRIBUTIONS

Ex 24: A coffee machine is supposed to dispense 200 ml of coffee per cup. However, the amount varies slightly from cup to cup.
Let X_i be the volume of a single cup. Based on quality control, the volume has a mean $\mu = 200$ ml and a standard deviation $\sigma = 15$ ml.
A health inspector takes a random sample of 50 cups. Let $\overline{X}_{50}$ be the average volume of these 50 cups.

1. Find the expected value of the sample mean, $E(\overline{X}_{50})$.

2. Calculate the standard deviation for a single cup (σ), and then find the standard deviation of the sample mean, $\sigma(\overline{X}_{50})$.

3. The distribution of a single cup is not necessarily normal. Can we assume the sample mean $\overline{X}_{50}$ is normally distributed? Explain.

4. Find the probability that the average volume of the sample is less than 195 ml.

Ex 25: A sociologist studies the income in a large city. The distribution of salaries is usually not normal (it is skewed).
Let X_i be the annual salary of a single person. Statistics show that the population mean is $\mu = \$45,000$ and the standard deviation is $\sigma = \$12,000$.
The sociologist surveys 100 random people. Let $\overline{X}_{100}$ be the average salary of this sample.

1. Find the expected average salary of the sample, $E(\overline{X}_{100})$.

2. Calculate the standard deviation for a single salary (σ), and then find the standard deviation of the sample mean salary, $\sigma(\overline{X}_{100})$.

3. Can we assume the average salary of the sample $\overline{X}_{100}$ is normally distributed? Explain.

4. Find the probability that the average salary of the sample exceeds $\$48,000$.

Ex 26: A large battery pack is built by connecting 50 individual cells in series. The pack fails as soon as one cell fails, but for capacity calculation, we consider the total energy.
Let X_i be the energy capacity of a single cell (in Wh). The manufacturer states that $\mu = 10$ Wh and $\sigma = 1$ Wh.
Let S_{50} be the total energy capacity of the pack.

1. Find the expected total capacity of the pack, $E(S_{50})$.
2. Find the standard deviation of the total capacity, $\sigma(S_{50})$.
3. Can we assume the total capacity S_{50} is normally distributed? Explain.
4. The pack is sold as a "500 Wh" battery. Find the probability that a pack actually has less than 490 Wh of capacity.

C.4 CALCULATING QUANTILES USING THE NORMAL APPROXIMATION

Ex 27: You have invited $n = 64$ guests to a party. You need to make sandwiches for the guests.You believe that the number of sandwiches a single guest needs, X_i, takes values $0, 1$, or 2 with probabilities $\frac{1}{4}$, $\frac{1}{2}$, and $\frac{1}{4}$ respectively.
Assume that the number of sandwiches each guest needs is independent of the others.Let S_{64} be the total number of sandwiches needed.

1. Find the mean μ and the variance σ^2 of the number of sandwiches for a **single** guest.
2. Find the expected total number of sandwiches $E(S_{64})$ and the standard deviation of the total $\sigma(S_{64})$.
3. Can we assume S_{64} is normally distributed? Explain.
4. How many sandwiches should you make so that you are 95% sure that there is no shortage (i.e., find k such that $P(S_{64} \leq k) \geq 0.95$)?

Ex 28: You are managing a moving company. You have $n = 100$ students moving out of a dormitory today. You need to provide boxes.
Based on past data, the number of boxes a single student needs, X_i, takes values $2, 3$, or 4 with probabilities $0.2, 0.5$, and 0.3 respectively.
Assume that the number of boxes each student needs is independent of the others.
Let S_{100} be the total number of boxes needed.

1. Find the mean μ and the variance σ^2 of the number of boxes for a **single** student.
2. Find the expected total number of boxes $E(S_{100})$ and the standard deviation of the total $\sigma(S_{100})$.
3. Can we assume S_{100} is normally distributed? Explain.

4. How many boxes should you bring so that you are 99% sure that there is no shortage (i.e., find k such that $P(S_{100} \leq k) \geq 0.99$)?

Ex 29: A shuttle bus service transports passengers from a hotel to the airport. The hotel has received $n = 50$ booking requests.
The number of passengers per booking, X_i, can be $1, 2$, or 3 people with probabilities $0.5, 0.3$, and 0.2 respectively.
Assume that the size of each booking group is independent.
Let S_{50} be the total number of passengers.

1. Find the mean μ and the variance σ^2 of the number of passengers for a **single** booking.

2. Find the expected total number of passengers $E(S_{50})$ and the standard deviation of the total $\sigma(S_{50})$.

3. Can we assume S_{50} is normally distributed? Explain.

4. The bus company wants to be 90% confident that they have enough seats. What should be the minimum capacity of the bus (find k)?

Ex 30: An Internet service provider sets up a local access point which serves $n = 5000$ subscribers.
At any given time, each subscriber has a probability of $p = 0.2$ of being connected. We assume that the behavior of each subscriber is independent of the others.
Let S_{5000} be the total number of active connections at a given time.

1. Find the expected number of active connections, $E(S_{5000})$.

2. Find the standard deviation of the number of active connections, $\sigma(S_{5000})$.

3. Using the Central Limit Theorem, approximate the distribution of S_{5000}.

4. Find the minimum number of simultaneous connections, k, that the access point must be able to handle so that the probability of the system **not** being saturated is greater than 99.99% (i.e., $P(S_{5000} \leq k) > 0.9999$).