STAT 111

Recitation 4

Mo Huang

Email: mohuang@wharton.upenn.edu

Office Hours: Wednesdays 3:00 - 4:00 pm, JMHH F96

Slides: `github.com/mohuangx/STAT111-Spring2019`

February 15, 2019

Many Random Variables

- Let X and Y be two random variables. Then,

  $\text{mean}(X+Y) = \text{mean}(X) + \text{mean}(Y).$

  - If X, Y are also independent,

    $\text{variance}(X+Y) = \text{variance}(X) + \text{variance}(Y).$

  - For constants a, b, we have

    $\text{mean}(aX+bY) = a \times \text{mean}(X) + b \times \text{mean}(Y)$

    $\text{variance}(aX+bY) = a^2 \times \text{variance}(X) + b^2 \times \text{variance}(Y).$

- Let $D = X - Y$. What is the variance of D?

  $\text{variance}(D) = \text{variance}(X) + \text{variance}(Y).$

Many Random Variables

- Let $X_1, \ldots, X_n$ be i.i.d. random variables, each with mean μ and variance σ^2.

Many Random Variables

- Let $X_1, \ldots, X_n$ be i.i.d. random variables, each with mean μ and variance σ^2.
- Then for the sum, $T_n = X_1 + \cdots + X_n$:

$$\text{mean of } T_n = n\mu, \quad \text{variance of } T_n = n\sigma^2$$

Many Random Variables

- Let $X_1, \ldots, X_n$ be i.i.d. random variables, each with mean μ and variance σ^2.
- Then for the sum, $T_n = X_1 + \cdots + X_n$:

$$\text{mean of } T_n = n\mu, \quad \text{variance of } T_n = n\sigma^2$$

- For the average, $\bar{X} = \frac{X_1 + \cdots + X_n}{n}$:

$$\text{mean of } \bar{X} = \mu, \quad \text{variance of } \bar{X} = \frac{\sigma^2}{n}.$$

Questions

Q1: Suppose the company producing a medicine has different means and variances the amount produced on each day of the week:

Day	Mean	Variance
Monday (X_1)	450	1200
Tuesday (X_2)	550	800
Wednesday (X_3)	600	500
Thursday (X_4)	550	800
Friday (X_5)	350	1200

Questions

Q1: Suppose the company producing a medicine has different means and variances the amount produced on each day of the week:

Day	Mean	Variance
Monday (X_1)	450	1200
Tuesday (X_2)	550	800
Wednesday (X_3)	600	500
Thursday (X_4)	550	800
Friday (X_5)	350	1200

Find mean and variance of both the sum T_n and the average $\bar{X}$.

Questions

Q1: Suppose the company producing a medicine has different means and variances the amount produced on each day of the week:

Day	Mean	Variance
Monday (X_1)	450	1200
Tuesday (X_2)	550	800
Wednesday (X_3)	600	500
Thursday (X_4)	550	800
Friday (X_5)	350	1200

Find mean and variance of both the sum T_n and the average $\bar{X}$.

A1: X_1, X_2, X_3, X_4, and X_5 are no longer $i.i.d.$!

$$Mean(T_n) = 450 + 550 + 600 + 550 + 350 = 2500$$
$$Var(T_n) = 1200 + 800 + 500 + 800 + 1200 = 4500$$

Questions

Q1: Suppose the company producing a medicine has different means and variances the amount produced on each day of the week:

Day	Mean	Variance
Monday (X_1)	450	1200
Tuesday (X_2)	550	800
Wednesday (X_3)	600	500
Thursday (X_4)	550	800
Friday (X_5)	350	1200

Find mean and variance of both the sum T_n and the average $\bar{X}$.

A1: X_1, X_2, X_3, X_4, and X_5 are no longer $i.i.d.$!

$$Mean(T_n) = 450 + 550 + 600 + 550 + 350 = 2500$$
$$Var(T_n) = 1200 + 800 + 500 + 800 + 1200 = 4500$$

$$Mean(\bar{X}) = 1/n \times Mean(T_n) = 500$$
$$Var(\bar{X}) = 1/n^2 \times Var(T_n) = 4500/25 = 180$$

Questions

Q2: Let P_1 be the proportion of heads in 50 coin tosses, where $P(H) = 0.6$. Find $Mean(P_1)$ and $Var(P_1)$.

Questions

Q2: Let P_1 be the proportion of heads in 50 coin tosses, where $P(H) = 0.6$. Find $Mean(P_1)$ and $Var(P_1)$.

A2: $Mean(P_1) = 0.6$ and $Var(P_1) = 0.6 \times 0.4/50 = 0.0048$.

Questions

Q2: Let P_1 be the proportion of heads in 50 coin tosses, where $P(H) = 0.6$. Find $Mean(P_1)$ and $Var(P_1)$.

A2: $Mean(P_1) = 0.6$ and $Var(P_1) = 0.6 \times 0.4/50 = 0.0048$.

Q3: Let P_2 be the proportion of heads in 20 coin tosses, where $P(H) = 0.7$. From earlier, $Mean(P_2) = 0.7$ and $Var(P_2) = 0.0105$. Let $D = P_1 - P_2$. Find the mean and variance of D.

Questions

Q2: Let P_1 be the proportion of heads in 50 coin tosses, where $P(H) = 0.6$. Find $Mean(P_1)$ and $Var(P_1)$.

A2: $Mean(P_1) = 0.6$ and $Var(P_1) = 0.6 \times 0.4/50 = 0.0048$.

Q3: Let P_2 be the proportion of heads in 20 coin tosses, where $P(H) = 0.7$. From earlier, $Mean(P_2) = 0.7$ and $Var(P_2) = 0.0105$. Let $D = P_1 - P_2$. Find the mean and variance of D.

A3: $Mean(D) = 0.6 - 0.7 = -0.1$
$Var(D) = 0.0048 + 0.0105 = 0.0153$

Continuous Random Variables

- So far, we have just considered discrete random variables; those whose possible values are countable.
- A continuous random variable can take continuous values in a future experiment.
- Every continuous random variable X has an associated density function $f(x)$.

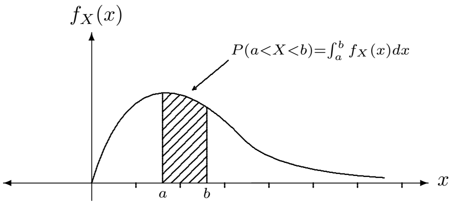
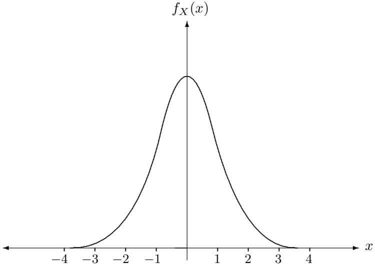


Figure 5: $P(a < X < b) = \int_a^b f_X(x)dx$.

The Normal Distribution

- A normal random variable is a continuous random variable.



The density function for the standard normal distribution with $\mu = 0$, $\sigma = 1$.

- We call a normal random variable with $\mu = 0$ and $\sigma^2 = 1$ a standard normal random variable.

- For standard normal random variables, we can use charts (or a computer) to find the area under the density function (i.e. the probabilities).

Questions

- $P(Z < -1.75)$

Questions

- $P(Z < -1.75)$

  $P(Z < -1.75) = 0.0401$

Questions

- $P(Z < -1.75)$

  $P(Z < -1.75) = 0.0401$

- $P(Z > 0.85)$

Questions

- $P(Z < -1.75)$

  $P(Z < -1.75) = 0.0401$

- $P(Z > 0.85)$

  $P(Z > 0.85) = 1 - 0.8023 = 0.1977$

Questions

- $P(Z < -1.75)$

  $P(Z < -1.75) = 0.0401$

- $P(Z > 0.85)$

  $P(Z > 0.85) = 1 - 0.8023 = 0.1977$

- $P(-1.43 < Z < 0.92)$

Questions

- $P(Z < -1.75)$

  $P(Z < -1.75) = 0.0401$

- $P(Z > 0.85)$

  $P(Z > 0.85) = 1 - 0.8023 = 0.1977$

- $P(-1.43 < Z < 0.92)$

  $P(-1.43 < Z < 0.92) = 0.8212 - .0764 = .7448$

Standardization

- What if a normal random variable has a different mean and variance?

Standardization

- What if a normal random variable has a different mean and variance?
- We need to *standardize* it.

Standardization

- What if a normal random variable has a different mean and variance?
- We need to *standardize* it.
- Let $X \sim N(\mu, \sigma^2)$: that is, X is a normal random variable with mean μ, variance σ^2. Let
$$Z = \frac{X - \mu}{\sigma}$$

Standardization

- What if a normal random variable has a different mean and variance?
- We need to *standardize* it.
- Let $X \sim N(\mu, \sigma^2)$: that is, X is a normal random variable with mean μ, variance σ^2. Let
$$Z = \frac{X - \mu}{\sigma}$$
- Then Z is a *standard* normal random variable.

Example

- X is a normal random variable with $\mu = 5$ and $\sigma^2 = 9$. Find $P(X > 8)$.

Example

- X is a normal random variable with $\mu = 5$ and $\sigma^2 = 9$. Find $P(X > 8)$.

$$\begin{aligned} P(X > 8) &= P\left(\frac{X-5}{3} > \frac{8-5}{3}\right) \\ &= P(Z > 1) \\ &= 0.1587 \end{aligned}$$

Questions

- $X \sim N(2, 16)$. Find $P(-1 < X < 6)$.

Questions

- $X \sim N(2, 16)$. Find $P(-1 < X < 6)$.

$$\begin{aligned} P(-1 < X < 6) &= P\left(\frac{-1-2}{4} < \frac{X-2}{4} < \frac{6-2}{4}\right) \\ &= P(-0.75 < Z < 1) \\ &= 0.8413 - 0.2266 \\ &= 0.6147 \end{aligned}$$

Questions

- $Y \sim N(7, 25)$. Find $P(8 < Y < 13)$.

Questions

- $Y \sim N(7, 25)$. Find $P(8 < Y < 13)$.

$$
\begin{aligned}
P(8 < Y < 13) &= P\left(\frac{8-7}{5} < \frac{Y-7}{5} < \frac{13-7}{5}\right) \\
&= P(0.2 < Z < 1.2) \\
&= 0.8849 - 0.5793 \\
&= 0.3056
\end{aligned}
$$

Two-Standard-Deviation Rule

- From the chart:

$$P(Z < -1.96) = 0.025, \quad P(Z > 1.96) = 0.025.$$

Two-Standard-Deviation Rule

- From the chart:

$$P(Z < -1.96) = 0.025, \quad P(Z > 1.96) = 0.025.$$

- Then:

$$P(-1.96 < Z < 1.96) = 0.95.$$

Two-Standard-Deviation Rule

- From the chart:

$$P(Z < -1.96) = 0.025, \quad P(Z > 1.96) = 0.025.$$

- Then:

$$P(-1.96 < Z < 1.96) = 0.95.$$

- Approximate $1.96 \approx 2$ and "unstandardize":

$$P\left(-2 < \frac{X - \mu}{\sigma} < 2\right) = 0.95$$

Two-Standard-Deviation Rule

- From the chart:

$$P(Z < -1.96) = 0.025, \quad P(Z > 1.96) = 0.025.$$

- Then:

$$P(-1.96 < Z < 1.96) = 0.95.$$

- Approximate $1.96 \approx 2$ and "unstandardize":

$$P\left(-2 < \frac{X-\mu}{\sigma} < 2\right) = 0.95$$

$\Rightarrow$

$$P(\mu - 2\sigma < X < \mu + 2\sigma) = 0.95.$$

Two-Standard-Deviation Rule

- From the chart:

$$P(Z < -1.96) = 0.025, \quad P(Z > 1.96) = 0.025.$$

- Then:

$$P(-1.96 < Z < 1.96) = 0.95.$$

- Approximate $1.96 \approx 2$ and "unstandardize":

$$P\left(-2 < \frac{X - \mu}{\sigma} < 2\right) = 0.95$$

$\Rightarrow$

$$P(\mu - 2\sigma < X < \mu + 2\sigma) = 0.95.$$

- The probability that a normal random variable is within 2 standard deviations of the mean is 95%.

Normal Distribution: Sums and Averages

- Let $X_1, \ldots, X_n$ be independent and *normally distributed*. Let

$$T_n = X_1 + \cdots + X_n, \qquad \bar{X} = \frac{X_1 + \cdots X_n}{n}, \qquad D = X_2 - X_1.$$

Normal Distribution: Sums and Averages

- Let $X_1, \ldots, X_n$ be independent and *normally distributed*. Let

$$T_n = X_1 + \cdots + X_n, \qquad \bar{X} = \frac{X_1 + \cdots X_n}{n}, \qquad D = X_2 - X_1.$$

- Then T_n, $\bar{X}$ and D are also normal random variables.

Normal Distribution: Sums and Averages

- Let $X_1, \ldots, X_n$ be independent and *normally distributed*. Let

$$T_n = X_1 + \cdots + X_n, \qquad \bar{X} = \frac{X_1 + \cdots X_n}{n}, \qquad D = X_2 - X_1.$$

- Then T_n, $\bar{X}$ and D are also normal random variables.

- Let $X_1, \ldots, X_n \stackrel{i.i.d}{\sim} N(\mu, \sigma^2)$. Then:

$$T_n \sim N(n\mu, n\sigma^2)$$

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$D \sim N(0, 2\sigma^2)$$

Example

- Suppose we know the weight X of an adult man chosen at random is normally distributed with mean 160 pounds and variance 64 pounds^2.

a) Find $P(156 < X < 164)$.

Example

- Suppose we know the weight X of an adult man chosen at random is normally distributed with mean 160 pounds and variance 64 pounds^2.

a) Find $P(156 < X < 164)$.

$$\begin{aligned} P(156 < X < 164) &= P\left(\frac{156-160}{8} < \frac{X-160}{8} < \frac{164-160}{8}\right) \\ &= P(-0.5 < Z < 0.5) \\ &= 0.3830 \end{aligned}$$

Example

- Suppose we know the weight X of an adult man chosen at random is normally distributed with mean 160 pounds and variance 64 pounds^2.

a) Find $P(156 < X < 164)$.

$$\begin{aligned} P(156 < X < 164) &= P\left(\frac{156-160}{8} < \frac{X-160}{8} < \frac{164-160}{8}\right) \\ &= P(-0.5 < Z < 0.5) \\ &= 0.3830 \end{aligned}$$

b) Find the probability that the average weight of 16 men chosen at random is between 156 and 164 pounds.

Example

- Suppose we know the weight X of an adult man chosen at random is normally distributed with mean 160 pounds and variance 64 pounds^2.

a) Find $P(156 < X < 164)$.

$$\begin{aligned} P(156 < X < 164) &= P\left(\frac{156-160}{8} < \frac{X-160}{8} < \frac{164-160}{8}\right) \\ &= P(-0.5 < Z < 0.5) \\ &= 0.3830 \end{aligned}$$

b) Find the probability that the average weight of 16 men chosen at random is between 156 and 164 pounds.

$$\overline{X} \sim N(160, 64/16 = 4)$$

$$\begin{aligned} P(156 < \overline{X} < 164) &= P(-2 < Z < 2) \\ &\approx 0.95 \end{aligned}$$

Example

- Suppose we know the weight X of an adult man chosen at random is normally distributed with mean 160 pounds and variance 64 pounds^2.

c) Calculate the numbers A and B such that $P(A < X < B) \approx 0.95$.

Example

- Suppose we know the weight X of an adult man chosen at random is normally distributed with mean 160 pounds and variance 64 pounds^2.

c) Calculate the numbers A and B such that $P(A < X < B) \approx 0.95$.

$$0.95 \approx P\left(-2 < \frac{X-\mu}{\sigma} < 2\right) = P(-2\sigma + \mu < X < 2\sigma + \mu)$$

$$A = -2(8) + 160 = 144, \quad B = 2(8) + 160 = 176$$

Example

- Suppose we know the weight X of an adult man chosen at random is normally distributed with mean 160 pounds and variance 64 pounds^2.

c) Calculate the numbers A and B such that $P(A < X < B) \approx 0.95$.

$$0.95 \approx P\left(-2 < \frac{X-\mu}{\sigma} < 2 \right) = P(-2\sigma + \mu < X < 2\sigma + \mu)$$
$$A = -2(8) + 160 = 144, \quad B = 2(8) + 160 = 176$$

d) Calculate the numbers C and D such that the average of 256 randomly chosen adults is between C and D with probability approximately 0.95.

Example

- Suppose we know the weight X of an adult man chosen at random is normally distributed with mean 160 pounds and variance 64 pounds^2.

c) Calculate the numbers A and B such that $P(A < X < B) \approx 0.95$.

$$0.95 \approx P\left(-2 < \frac{X-\mu}{\sigma} < 2\right) = P(-2\sigma + \mu < X < 2\sigma + \mu)$$
$$A = -2(8) + 160 = 144, \quad B = 2(8) + 160 = 176$$

d) Calculate the numbers C and D such that the average of 256 randomly chosen adults is between C and D with probability approximately 0.95.

$$\overline{X}_{256} \sim N(160, 64/256 = 1/4)$$
$$C = -2\sigma + \mu = -2(1/2) + 160 = 159$$
$$D = 2\sigma + \mu = 2(1/2) + 160 = 161$$

Central Limit Theorem

The Central Limit Theorem:

- Suppose $X_1, X_2, \ldots, X_n$ are *iid* with mean μ and variance σ^2.
- Then, for large n

$$T_n \sim N(n\mu, n\sigma^2) \quad \text{and} \quad \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

no matter the distribution of the individual X_i

- Allows approximation of all distributions using the normal distribution if you know the mean and variance.

Central Limit Theorem

The Central Limit Theorem:

- Suppose $X_1, X_2, \ldots, X_n$ are *iid* with mean μ and variance σ^2.
- Then, for large n

$$T_n \sim N(n\mu, n\sigma^2) \quad \text{and} \quad \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

no matter the distribution of the individual X_i

- Allows approximation of all distributions using the normal distribution if you know the mean and variance.

Note: if $X_1, \ldots, X_n$ are normally distributed, then this applies for *all* n, not just large n.

Central Limit Theorem: Example

- Let $X_1, X_2, \ldots, X_n \stackrel{iid}{\sim} Binomial(1, \theta)$.

Central Limit Theorem: Example

- Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} Binomial(1, \theta)$.

- For each X_i, $Mean(X_i) = \theta$ and $Var(X_i) = \theta(1 - \theta)$.

Central Limit Theorem: Example

- Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} Binomial(1, \theta)$.
- For each X_i, $Mean(X_i) = \theta$ and $Var(X_i) = \theta(1 - \theta)$.
- The sum is: $T_n = X_1 + X_2 + \cdots + X_n$.

Central Limit Theorem: Example

- Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} Binomial(1, \theta)$.
- For each X_i, $Mean(X_i) = \theta$ and $Var(X_i) = \theta(1 - \theta)$.
- The sum is: $T_n = X_1 + X_2 + \cdots + X_n$.
- The proportion is:

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

Central Limit Theorem: Example

- Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} Binomial(1, \theta)$.
- For each X_i, $Mean(X_i) = \theta$ and $Var(X_i) = \theta(1-\theta)$.
- The sum is: $T_n = X_1 + X_2 + \cdots + X_n$.
- The proportion is:

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

Central Limit Theorem: Example

- Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} Binomial(1, \theta)$.
- For each X_i, $Mean(X_i) = \theta$ and $Var(X_i) = \theta(1 - \theta)$.
- The sum is: $T_n = X_1 + X_2 + \cdots + X_n$.
- The proportion is:
$$P = \frac{X_1 + \cdots + X_n}{n}.$$
- For large n,
$$T_n \sim N\left(n\theta, n\theta[1 - \theta]\right)$$
$$P \sim N\left(\theta, \frac{\theta(1 - \theta)}{n}\right)$$

Central Limit Theorem: Problem

- Suppose you are rolling a fair die 1000 times. Calculate the numbers A and B such that the average of the 1000 rolls is between A and B with probability approximately 0.95. You may assume the mean of one roll is 3.5 and the variance is $35/12$.

Central Limit Theorem: Problem

- Suppose you are rolling a fair die 1000 times. Calculate the numbers A and B such that the average of the 1000 rolls is between A and B with probability approximately 0.95. You may assume the mean of one roll is 3.5 and the variance is 35/12.

$$Mean(X_i) = 3.5, \quad Var(X_i) = 35/12$$

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) = N\left(3.5, \frac{35}{12000}\right) \quad \text{by CLT}$$

$$A = -2\sigma + \mu = -2\sqrt{35/12000} + 3.5 \approx 3.392$$

$$B = 2\sigma + \mu = 2\sqrt{35/12000} + 3.5 \approx 3.608$$