Statistics from samples and Limit theorems

Andrew Thangaraj

IIT Madras

Subsection 1

Statistics from *iid* samples

Where have we seen *iid* samples?

- Benoulli trials
- Monte carlo simulations
- Computing histograms

Bernoulli trials

- Experiment and an event of interest A
  - Occurance of an Event A is considered success
  - What is $p = P(A)$?

Bernoulli trials

- Experiment and an event of interest A
  - Occurance of an Event A is considered success
  - What is $p = P(A)$?
- Bernoulli trails
  - Independent repetitions or trials of experiment, say, n times
  - $X_i = 1$ if A occurs in the i-th trial, and $X_i = 0$ otherwise

iid Bernoulli samples: $X_1, X_2, \ldots, X_n$

Bernoulli trials

- Experiment and an event of interest A
  - Occurance of an Event A is considered success
  - What is $p = P(A)$?
- Bernoulli trails
  - Independent repetitions or trials of experiment, say, n times
  - $X_i = 1$ if A occurs in the i-th trial, and $X_i = 0$ otherwise

iid Bernoulli samples: $X_1, X_2, \ldots, X_n$

Goal: Try to estimate $P(A) = P(X_i = 1)$

- Useful in finding prevalence of a disease in a population etc.

Monte carlo simulations

- Experiment and event of interest A
  - Too complex for modeling and computation
  - Can be simulated on a computer

Monte carlo simulations

- Experiment and event of interest A
  - Too complex for modeling and computation
  - Can be simulated on a computer
- Repeat simulation n times independently
  - Record number of occurances of A as n_A

$$P(A) \approx \frac{n_A}{n}$$

  - Why is the above true? What should be n?

Monte carlo simulations

- Experiment and event of interest A
  - Too complex for modeling and computation
  - Can be simulated on a computer
- Repeat simulation n times independently
  - Record number of occurances of A as n_A

$$P(A) \approx \frac{n_A}{n}$$

  - Why is the above true? What should be n?

iid samples: $X_1, X_2, \ldots, X_n$

- $X_i = 1$ if A occurs in the i-th trial, and $X_i = 0$ otherwise
- Estimate $P(A) = P(X_i = 1)$ (similar to Bernoulli trial)

Computing histograms

- n data points of some variable of interest
  - $x_1, x_2, \dots, x_n$
- Bin: $[a, b]$
  - n_b: number of x_i that fall inside $[a, b]$

Computing histograms

- n data points of some variable of interest
  - $x_1, x_2, \ldots, x_n$
- Bin: $[a, b]$
  - n_b: number of x_i that fall inside $[a, b]$

Model

- X: continuous random variable with density $f_X(x)$
- Event $A = (a < X < b)$

Computing histograms

- n data points of some variable of interest
  - $x_1, x_2, \ldots, x_n$
- Bin: $[a, b]$
  - n_b: number of x_i that fall inside $[a, b]$

Model

- X: continuous random variable with density $f_X(x)$
- Event $A = (a < X < b)$

Histogram count

- Data points: *iid* samples $X_1, X_2, \ldots, X_n \sim f_X(x)$
- Estimate $P(a < X < b) \approx n_b/n$

iid samples hold information on distribution

- What is common to all 3 of the previous scenarios?
  - Given: *iid* samples
  - Goal: get some partial information about the distribution
    - Procedures to gather the information needed

iid samples hold information on distribution

- What is common to all 3 of the previous scenarios?
  - Given: *iid* samples
  - Goal: get some partial information about the distribution
    - Procedures to gather the information needed
- *iid* samples of an unknown or partially known distribution form the input for statistical procedures
  - Data: modelled as observations from *iid* repetitions of an experiment
  - Example: Iris data
    - Data from every iris is considered to be *iid* observations from the distribution of the 4 lengths

iid samples hold information on distribution

- What is common to all 3 of the previous scenarios?
  - Given: *iid* samples
  - Goal: get some partial information about the distribution
    - Procedures to gather the information needed
- *iid* samples of an unknown or partially known distribution form the input for statistical procedures
  - Data: modelled as observations from *iid* repetitions of an experiment
  - Example: Iris data
    - Data from every iris is considered to be *iid* observations from the distribution of the 4 lengths

Analysis

- How to decide if the statistical procedure is "good"?
- How many samples are needed for a "goodness" guarantee?

Example: 20 *iid* Bernoulli(p) samples with p unknown

- **Goal**
  - Find p from *iid* samples

Example: 20 *iid* Bernoulli(p) samples with p unknown

- **Goal**
  - Find p from *iid* samples
- Sampling 1
  - 1, 1, 0, 1, 0, 0, 0, 1, 0, 1, 1, 0, 1, 1, 0, 1, 0, 1, 1, 1
- Sampling 2
  - 0, 0, 1, 1, 1, 0, 0, 1, 1, 0, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1
- Sampling 3
  - 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1
- and so on. . . .

Example: 20 *iid* Bernoulli(p) samples with p unknown

- **Goal**
  - Find p from *iid* samples
- Sampling 1
  - 1, 1, 0, 1, 0, 0, 0, 1, 0, 1, 1, 0, 1, 1, 0, 1, 0, 1, 1, 1
- Sampling 2
  - 0, 0, 1, 1, 1, 0, 0, 1, 1, 0, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1
- Sampling 3
  - 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1
- and so on. . . .
- **Important**
  - p is the same for all samplings
  - However, samples do not remain the same
  - Each sample in each sampling is an observation of a random variable

Example: 20 *iid* Bernoulli(p) samples with p unknown

- **Goal**
  - Find p from *iid* samples
- Sampling 1
  - 1, 1, 0, 1, 0, 0, 0, 1, 0, 1, 1, 0, 1, 1, 0, 1, 0, 1, 1, 1
- Sampling 2
  - 0, 0, 1, 1, 1, 0, 0, 1, 1, 0, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1
- Sampling 3
  - 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1
- and so on. . . .
- **Important**
  - p is the same for all samplings
  - However, samples do not remain the same
  - Each sample in each sampling is an observation of a random variable
- **Requirement on statistical procedure**
  - In spite of variations in samples, provide p with some guarantee

What is a typical statistical problem?

X: PMF $f_X(x)$
Each X_i: PMF $f_X(x)$
X_i are independent

- **Model for Samples**: $X_1, X_2, \ldots, X_n \sim$ iid X
- **Given "data"**: $x_1, x_2, \ldots, x_n$ from one sampling instance
- Distribution of X is partially known or unknown
  - What is partially known? Know distribution but parameters unknown
  - Example: Bernoulli(p) with p unknown, Normal(μ, σ^2) with μ and σ unknown
- **Goal**: Procedures to find information about the distribution of X

What is a typical statistical problem?

- **Model for Samples**: $X_1, X_2, \ldots, X_n \sim$ iid X
- **Given "data"**: $x_1, x_2, \ldots, x_n$ from one sampling instance
- Distribution of X is partially known or unknown
  - What is partially known? Know distribution but parameters unknown
  - Example: Bernoulli(p) with p unknown, Normal(μ, σ^2) with μ and σ unknown
- **Goal**: Procedures to find information about the distribution of X
- **What information?**
  - What is the mean of X? What is the variance of X?
  - What is $P(X > t)$? What is $P(a < X < b)$?
  - What is the distribution of X? What is the size of T_X?

Subsection 2

Empirical distribution and descriptive statistics

Empirical distribution

Definition (Empirical distribution)

Let $X_1, X_2, \ldots, X_n \sim X$ be iid samples. Let $\#(X_i = t)$ denote the number of times t occurs in the samples. The empirical distribution is the discrete distribution with PMF

$$p(t) = \frac{\#(X_i = t)}{n}.$$

Empirical distribution

Definition (Empirical distribution)

Let $X_1, X_2, \ldots, X_n \sim X$ be iid samples. Let $\#(X_i = t)$ denote the number of times t occurs in the samples. The empirical distribution is the discrete distribution with PMF

$$p(t) = \frac{\#(X_i = t)}{n}.$$

Example: $n = 20$

- 1, 1, 0, 1, 0, 0, 0, 1, 0, 1, 1, 0, 1, 1, 0, 1, 0, 1, 1, 1

Empirical distribution

Definition (Empirical distribution)

Let $X_1, X_2, \ldots, X_n \sim X$ be iid samples. Let $\#(X_i = t)$ denote the number of times t occurs in the samples. The empirical distribution is the discrete distribution with PMF

$$p(t) = \frac{\#(X_i = t)}{n}.$$

Example: $n = 20$

- 1, 1, 0, 1, 0, 0, 0, 1, 0, 1, 1, 0, 1, 1, 0, 1, 0, 1, 1, 1
  - $p(0) = 8/20$, $p(1) = 12/20$

Empirical distribution

Definition (Empirical distribution)

Let $X_1, X_2, \ldots, X_n \sim X$ be iid samples. Let $\#(X_i = t)$ denote the number of times t occurs in the samples. The empirical distribution is the discrete distribution with PMF

$$p(t) = \frac{\#(X_i = t)}{n}.$$

Example: $n = 20$

Empirical Distribution $\{\overset{8/20}{0}, \overset{12/20}{1}\}$

- 1, 1, 0, 1, 0, 0, 0, 1, 0, 1, 1, 0, 1, 1, 0, 1, 0, 1, 1, 1
  - $p(0) = 8/20$, $p(1) = 12/20$
- 0, 0, 1, 1, 1, 0, 0, 1, 1, 0, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1

Empirical distribution

Definition (Empirical distribution)

Let $X_1, X_2, \ldots, X_n \sim X$ be iid samples. Let $\#(X_i = t)$ denote the number of times t occurs in the samples. The empirical distribution is the discrete distribution with PMF

$$p(t) = \frac{\#(X_i = t)}{n}.$$

Example: $n = 20$

- 1, 1, 0, 1, 0, 0, 0, 1, 0, 1, 1, 0, 1, 1, 0, 1, 0, 1, 1, 1
  - $p(0) = 8/20$, $p(1) = 12/20$
- 0, 0, 1, 1, 1, 0, 0, 1, 1, 0, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1
  - $p(0) = 7/20$, $p(1) = 13/20$

Empirical Distribution $\overset{7/20 \quad 13/20}{\{0, 1\}}$

Empirical distribution

Definition (Empirical distribution)

Let $X_1, X_2, \ldots, X_n \sim X$ be iid samples. Let $\#(X_i = t)$ denote the number of times t occurs in the samples. The empirical distribution is the discrete distribution with PMF

$$p(t) = \frac{\#(X_i = t)}{n}.$$

Example: $n = 20$

- 1, 1, 0, 1, 0, 0, 0, 1, 0, 1, 1, 0, 1, 1, 0, 1, 0, 1, 1, 1
  - $p(0) = 8/20$, $p(1) = 12/20$
- 0, 0, 1, 1, 1, 0, 0, 1, 1, 0, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1
  - $p(0) = 7/20$, $p(1) = 13/20$
- 1, 2, 0, 3, 0, 0, 1, 2, 0, 1, 3, 2, 1, 1, 0, 3, 0, 2, 2, 1 $\{0, 1, 2, 3\}$

Empirical distribution

Definition (Empirical distribution)

Let $X_1, X_2, \ldots, X_n \sim X$ be iid samples. Let $\#(X_i = t)$ denote the number of times t occurs in the samples. The empirical distribution is the discrete distribution with PMF

$$p(t) = \frac{\#(X_i = t)}{n}.$$

Example: $n = 20$

- 1, 1, 0, 1, 0, 0, 0, 1, 0, 1, 1, 0, 1, 1, 0, 1, 0, 1, 1, 1
  - $p(0) = 8/20$, $p(1) = 12/20$
- 0, 0, 1, 1, 1, 0, 0, 1, 1, 0, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1
  - $p(0) = 7/20$, $p(1) = 13/20$
- 1, 2, 0, 3, 0, 0, 1, 2, 0, 1, 3, 2, 1, 1, 0, 3, 0, 2, 2, 1
  - $p(0) = 6/20$, $p(1) = 6/20$, $p(2) = 5/20$, $p(3) = 3/20$

$6/20 \quad 6/20 \quad 5/20 \quad 3/20$

$\{0, 1, 2, 3\}$

Observations about empirical distribution

- Is the empirical distribution random?

Observations about empirical distribution

- Is the empirical distribution random?
  - Yes, it depends on the actual sample instances

Observations about empirical distribution

- Is the empirical distribution random?
  - Yes, it depends on the actual sample instances
  - t and $p(t)$ may change from one sampling to another

Observations about empirical distribution

- Is the empirical distribution random?
  - Yes, it depends on the actual sample instances
  - t and $p(t)$ may change from one sampling to another
  - Example: 20 Bernoulli(p) samples

Observations about empirical distribution

- Is the empirical distribution random?
  - Yes, it depends on the actual sample instances
  - t and $p(t)$ may change from one sampling to another
  - Example: 20 Bernoulli(p) samples
- **Descriptive statistics**: Properties of empirical distribution

Observations about empirical distribution

- Is the empirical distribution random?
  - Yes, it depends on the actual sample instances
  - t and $p(t)$ may change from one sampling to another
  - Example: 20 Bernoulli(p) samples
- **Descriptive statistics**: Properties of empirical distribution
  - Mean of the distribution

Observations about empirical distribution

- Is the empirical distribution random?
  - Yes, it depends on the actual sample instances
  - t and $p(t)$ may change from one sampling to another
  - Example: 20 Bernoulli(p) samples
- **Descriptive statistics**: Properties of empirical distribution
  - Mean of the distribution
  - Variance of the distribution

Observations about empirical distribution

- Is the empirical distribution random?
  - Yes, it depends on the actual sample instances
  - t and $p(t)$ may change from one sampling to another
  - Example: 20 Bernoulli(p) samples
- **Descriptive statistics**: Properties of empirical distribution
  - Mean of the distribution
  - Variance of the distribution
  - Probability of an event

Observations about empirical distribution

- Is the empirical distribution random?
  - Yes, it depends on the actual sample instances
  - t and $p(t)$ may change from one sampling to another
  - Example: 20 Bernoulli(p) samples
- **Descriptive statistics**: Properties of empirical distribution
  - Mean of the distribution
  - Variance of the distribution
  - Probability of an event
- As number of samples increases, the properties of empirical distribution *should* become close to that of the original distribution

Sample mean

Definition (Sample mean)

Let $X_1, X_2, \dots, X_n$ be iid samples. The sample mean, denoted $\overline{X}$, is defined to be the random variable

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}.$$

- Given a sampling $x_1, \dots, x_n$, the value taken by the sample mean $\overline{X}$ is $\overline{x} = (x_1 + \cdots + x_n)/n$. Often, $\overline{X}$ and $\overline{x}$ are both called sample mean.

Sample mean

Definition (Sample mean)

Let $X_1, X_2, \ldots, X_n$ be iid samples. The sample mean, denoted $\overline{X}$, is defined to be the random variable

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}.$$

- Given a sampling $x_1, \ldots, x_n$, the value taken by the sample mean $\overline{X}$ is $\overline{x} = (x_1 + \cdots + x_n)/n$. Often, $\overline{X}$ and $\overline{x}$ are both called sample mean.

Example: $n = 20$

- 1, 1, 0, 1, 0, 0, 0, 1, 0, 1, 1, 0, 1, 1, 0, 1, 0, 1, 1, 1

Sample mean

Definition (Sample mean)

Let $X_1, X_2, \ldots, X_n$ be iid samples. The sample mean, denoted $\overline{X}$, is defined to be the random variable

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}.$$

- Given a sampling $x_1, \ldots, x_n$, the value taken by the sample mean $\overline{X}$ is $\overline{x} = (x_1 + \cdots + x_n)/n$. Often, $\overline{X}$ and $\overline{x}$ are both called sample mean.

Example: $n = 20$

- 1, 1, 0, 1, 0, 0, 0, 1, 0, 1, 1, 0, 1, 1, 0, 1, 0, 1, 1, 1
  - Value taken by $\overline{X}$: 12/20

Sample mean

Definition (Sample mean)

Let $X_1, X_2, \dots, X_n$ be iid samples. The sample mean, denoted $\overline{X}$, is defined to be the random variable

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}.$$

- Given a sampling $x_1, \dots, x_n$, the value taken by the sample mean $\overline{X}$ is $\overline{x} = (x_1 + \cdots + x_n)/n$. Often, $\overline{X}$ and $\overline{x}$ are both called sample mean.

Example: $n = 20$

- 1, 1, 0, 1, 0, 0, 0, 1, 0, 1, 1, 0, 1, 1, 0, 1, 0, 1, 1, 1
  - Value taken by $\overline{X}$: 12/20
- 0, 0, 1, 1, 1, 0, 0, 1, 1, 0, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1

Sample mean

Definition (Sample mean)

Let $X_1, X_2, \dots, X_n$ be iid samples. The sample mean, denoted $\overline{X}$, is defined to be the random variable

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}.$$

- Given a sampling $x_1, \dots, x_n$, the value taken by the sample mean $\overline{X}$ is $\overline{x} = (x_1 + \cdots + x_n)/n$. Often, $\overline{X}$ and $\overline{x}$ are both called sample mean.

Example: $n = 20$

- 1, 1, 0, 1, 0, 0, 0, 1, 0, 1, 1, 0, 1, 1, 0, 1, 0, 1, 1, 1
  - Value taken by $\overline{X}$: $12/20$
- 0, 0, 1, 1, 1, 0, 0, 1, 1, 0, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1
  - Value taken by $\overline{X}$: $13/20$

Sample mean

Definition (Sample mean)

Let $X_1, X_2, \ldots, X_n$ be iid samples. The sample mean, denoted $\overline{X}$, is defined to be the random variable

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}.$$

- Given a sampling $x_1, \ldots, x_n$, the value taken by the sample mean $\overline{X}$ is $\overline{x} = (x_1 + \cdots + x_n)/n$. Often, $\overline{X}$ and $\overline{x}$ are both called sample mean.

Example: $n = 20$

- 1, 1, 0, 1, 0, 0, 0, 1, 0, 1, 1, 0, 1, 1, 0, 1, 0, 1, 1, 1
  - Value taken by $\overline{X}$: 12/20
- 0, 0, 1, 1, 1, 0, 0, 1, 1, 0, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1
  - Value taken by $\overline{X}$: 13/20
- 1, 2, 0, 3, 0, 0, 1, 2, 0, 1, 3, 2, 1, 1, 0, 3, 0, 2, 2, 1

Sample mean

Definition (Sample mean)

Let $X_1, X_2, \ldots, X_n$ be iid samples. The sample mean, denoted $\overline{X}$, is defined to be the random variable

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}.$$

- Given a sampling $x_1, \ldots, x_n$, the value taken by the sample mean $\overline{X}$ is $\overline{x} = (x_1 + \cdots + x_n)/n$. Often, $\overline{X}$ and $\overline{x}$ are both called sample mean.

Example: $n = 20$

- 1, 1, 0, 1, 0, 0, 0, 1, 0, 1, 1, 0, 1, 1, 0, 1, 0, 1, 1, 1
  - Value taken by $\overline{X}$: 12/20
- 0, 0, 1, 1, 1, 0, 0, 1, 1, 0, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1
  - Value taken by $\overline{X}$: 13/20
- 1, 2, 0, 3, 0, 0, 1, 2, 0, 1, 3, 2, 1, 1, 0, 3, 0, 2, 2, 1
  - Sample mean: 25/20

Illustration 1: Bernoulli(0.5) samples

$$X_1, \ldots, X_n \sim \text{ iid } \{\overset{1/2}{0}, \overset{1/2}{1}\}$$

- $n = 5$

Illustration 1: Bernoulli(0.5) samples

$$X_1, \ldots, X_n \sim \text{ iid } \{ \overset{1/2}{0}, \overset{1/2}{1} \}$$

- $n = 5$
  - Samples: 0, 0, 1, 1, 1; Sample mean: $3/5$

Illustration 1: Bernoulli(0.5) samples

$$X_1, \ldots, X_n \sim \text{ iid } \{\overset{1/2}{0}, \overset{1/2}{1}\}$$

- $n = 5$
  - Samples: 0, 0, 1, 1, 1; Sample mean: 3/5
  - Samples: 1, 1, 1, 0, 1; Sample mean: 4/5

Illustration 1: Bernoulli(0.5) samples

$$X_1, \ldots, X_n \sim \text{ iid } \{ \overset{1/2}{0}, \overset{1/2}{1} \}$$

- $n = 5$
  - Samples: 0, 0, 1, 1, 1; Sample mean: 3/5
  - Samples: 1, 1, 1, 0, 1; Sample mean: 4/5
  - Samples: 0, 1, 1, 1, 0; Sample mean: 3/5

Illustration 1: Bernoulli(0.5) samples

$$X_1, \ldots, X_n \sim \text{ iid } \{ \overset{1/2}{0}, \overset{1/2}{1} \}$$

- $n = 5$
  - Samples: 0, 0, 1, 1, 1; Sample mean: $3/5$
  - Samples: 1, 1, 1, 0, 1; Sample mean: $4/5$
  - Samples: 0, 1, 1, 1, 0; Sample mean: $3/5$
- $n = 20$

Illustration 1: Bernoulli(0.5) samples

$$X_1, \ldots, X_n \sim \text{ iid } \{ \overset{1/2}{0}, \overset{1/2}{1} \}$$

- $n = 5$
  - Samples: 0, 0, 1, 1, 1; Sample mean: 3/5
  - Samples: 1, 1, 1, 0, 1; Sample mean: 4/5
  - Samples: 0, 1, 1, 1, 0; Sample mean: 3/5
- $n = 20$
  - 1, 1, 0, 1, 0, 0, 0, 1, 0, 1, 1, 0, 1, 1, 0, 1, 0, 1, 1, 1; 12/20

Illustration 1: Bernoulli(0.5) samples

$$X_1, \ldots, X_n \sim \text{ iid } \{ \overset{1/2}{0}, \overset{1/2}{1} \}$$

- $n = 5$
  - Samples: 0, 0, 1, 1, 1; Sample mean: 3/5
  - Samples: 1, 1, 1, 0, 1; Sample mean: 4/5
  - Samples: 0, 1, 1, 1, 0; Sample mean: 3/5
- $n = 20$
  - 1, 1, 0, 1, 0, 0, 0, 1, 0, 1, 1, 0, 1, 1, 0, 1, 0, 1, 1, 1; 12/20
  - 0, 0, 1, 1, 1, 0, 0, 1, 1, 0, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1; 13/20

Illustration 1: Bernoulli(0.5) samples

$$X_1, \ldots, X_n \sim \text{ iid } \{ \overset{1/2}{0}, \overset{1/2}{1} \}$$

- $n = 5$
  - Samples: 0, 0, 1, 1, 1; Sample mean: 3/5
  - Samples: 1, 1, 1, 0, 1; Sample mean: 4/5
  - Samples: 0, 1, 1, 1, 0; Sample mean: 3/5
- $n = 20$
  - 1, 1, 0, 1, 0, 0, 0, 1, 0, 1, 1, 0, 1, 1, 0, 1, 0, 1, 1, 1; 12/20
  - 0, 0, 1, 1, 1, 0, 0, 1, 1, 0, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1; 13/20
  - 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1; 13/20

Illustration 1: Bernoulli(0.5) samples

$$X_1, \ldots, X_n \sim \text{ iid } \{ \overset{1/2}{0}, \overset{1/2}{1} \}$$

- $n = 5$
  - Samples: 0, 0, 1, 1, 1; Sample mean: 3/5
  - Samples: 1, 1, 1, 0, 1; Sample mean: 4/5
  - Samples: 0, 1, 1, 1, 0; Sample mean: 3/5
- $n = 20$
  - 1, 1, 0, 1, 0, 0, 0, 1, 0, 1, 1, 0, 1, 1, 0, 1, 0, 1, 1, 1; 12/20
  - 0, 0, 1, 1, 1, 0, 0, 1, 1, 0, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1; 13/20
  - 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1; 13/20
- $n = 200$

Illustration 1: Bernoulli(0.5) samples

$$X_1, \ldots, X_n \sim \text{ iid } \{ \overset{1/2}{0}, \overset{1/2}{1} \}$$

- $n = 5$
  - Samples: 0, 0, 1, 1, 1; Sample mean: 3/5
  - Samples: 1, 1, 1, 0, 1; Sample mean: 4/5
  - Samples: 0, 1, 1, 1, 0; Sample mean: 3/5
- $n = 20$
  - 1, 1, 0, 1, 0, 0, 0, 1, 0, 1, 1, 0, 1, 1, 0, 1, 0, 1, 1, 1; 12/20
  - 0, 0, 1, 1, 1, 0, 0, 1, 1, 0, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1; 13/20
  - 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1; 13/20
- $n = 200$
  - Sampling 1: 95/200, Sampling 2: 102/200, Sampling 3: 98/200

Illustration 1: Bernoulli(0.5) samples

$$X_1, \ldots, X_n \sim \text{ iid } \{ \overset{1/2}{0} , \overset{1/2}{1} \}$$

- $n = 5$
  - Samples: 0, 0, 1, 1, 1; Sample mean: 3/5
  - Samples: 1, 1, 1, 0, 1; Sample mean: 4/5
  - Samples: 0, 1, 1, 1, 0; Sample mean: 3/5
- $n = 20$
  - 1, 1, 0, 1, 0, 0, 0, 1, 0, 1, 1, 0, 1, 1, 0, 1, 0, 1, 1, 1; 12/20
  - 0, 0, 1, 1, 1, 0, 0, 1, 1, 0, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1; 13/20
  - 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1; 13/20
- $n = 200$
  - Sampling 1: 95/200, Sampling 2: 102/200, Sampling 3: 98/200
- $n = 1000$

Illustration 1: Bernoulli(0.5) samples

$$X_1, \ldots, X_n \sim \text{ iid } \{ \overset{1/2}{0} , \overset{1/2}{1} \}$$

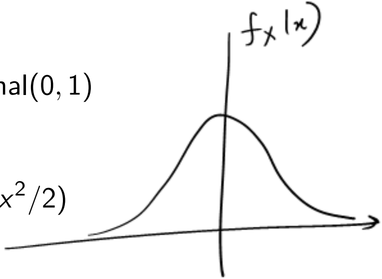
Distribution mean $= 1/2$

- $n = 5$
  - Samples: 0, 0, 1, 1, 1; Sample mean: 3/5
  - Samples: 1, 1, 1, 0, 1; Sample mean: 4/5
  - Samples: 0, 1, 1, 1, 0; Sample mean: 3/5
- $n = 20$
  - 1, 1, 0, 1, 0, 0, 0, 1, 0, 1, 1, 0, 1, 1, 0, 1, 0, 1, 1, 1; 12/20
  - 0, 0, 1, 1, 1, 0, 0, 1, 1, 0, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1; 13/20
  - 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1; 13/20
- $n = 200$
  - Sampling 1: 95/200, Sampling 2: 102/200, Sampling 3: 98/200
- $n = 1000$
  - Sampling 1: 495/~~200~~ 1000, Sampling 2: 490/~~200~~ 1000, Sampling 3: 504/~~200~~ 1000

Illustration 2: Normal(0, 1) samples

$$X_1, \ldots, X_n \sim \text{ iid Normal}(0, 1)$$

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$$



- $n = 5$

Illustration 2: Normal(0, 1) samples

$$X_1, \ldots, X_n \sim \text{ iid Normal}(0, 1)$$

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$$

- $n = 5$
  - Samples: 2.17, 0.10, -0.75, -1.05 , -1.72; Sample mean: -0.25

Illustration 2: Normal(0, 1) samples

$$X_1, \ldots, X_n \sim \text{ iid Normal}(0, 1)$$

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$$

- $n = 5$
  - Samples: 2.17, 0.10, -0.75, -1.05 , -1.72; Sample mean: -0.25
  - Samples: -0.26, 0.12, -0.31, -0.07, 1.35; Sample mean: 0.17

Illustration 2: Normal(0, 1) samples

$$X_1, \ldots, X_n \sim \text{ iid Normal}(0, 1)$$

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$$

- $n = 5$
  - Samples: 2.17, 0.10, -0.75, -1.05 , -1.72; Sample mean: -0.25
  - Samples: -0.26, 0.12, -0.31, -0.07, 1.35; Sample mean: 0.17
  - Samples: -0.20, 0.37, 1.00, -0.41, -0.21; Sample mean: 0.11

Illustration 2: Normal(0, 1) samples

$$X_1, \dots, X_n \sim \text{ iid Normal}(0, 1)$$

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$$

- $n = 5$
  - Samples: 2.17, 0.10, -0.75, -1.05 , -1.72; Sample mean: -0.25
  - Samples: -0.26, 0.12, -0.31, -0.07, 1.35; Sample mean: 0.17
  - Samples: -0.20, 0.37, 1.00, -0.41, -0.21; Sample mean: 0.11
- $n = 20$

Illustration 2: Normal(0, 1) samples

$$X_1, \ldots, X_n \sim \text{ iid Normal}(0, 1)$$

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$$

- $n = 5$
  - Samples: 2.17, 0.10, -0.75, -1.05 , -1.72; Sample mean: -0.25
  - Samples: -0.26, 0.12, -0.31, -0.07, 1.35; Sample mean: 0.17
  - Samples: -0.20, 0.37, 1.00, -0.41, -0.21; Sample mean: 0.11
- $n = 20$
  - Sampling 1: 0.08, Sampling 2: -0.24, Sampling 3: 0.41

Illustration 2: Normal$(0, 1)$ samples

$$X_1, \ldots, X_n \sim \text{ iid Normal}(0, 1)$$

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$$

- $n = 5$
  - Samples: 2.17, 0.10, -0.75, -1.05 , -1.72; Sample mean: -0.25
  - Samples: -0.26, 0.12, -0.31, -0.07, 1.35; Sample mean: 0.17
  - Samples: -0.20, 0.37, 1.00, -0.41, -0.21; Sample mean: 0.11
- $n = 20$
  - Sampling 1: 0.08, Sampling 2: -0.24, Sampling 3: 0.41
- $n = 200$

Illustration 2: Normal(0, 1) samples

$$X_1, \ldots, X_n \sim \text{ iid Normal}(0, 1)$$

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$$

- $n = 5$
  - Samples: 2.17, 0.10, -0.75, -1.05 , -1.72; Sample mean: -0.25
  - Samples: -0.26, 0.12, -0.31, -0.07, 1.35; Sample mean: 0.17
  - Samples: -0.20, 0.37, 1.00, -0.41, -0.21; Sample mean: 0.11
- $n = 20$
  - Sampling 1: 0.08, Sampling 2: -0.24, Sampling 3: 0.41
- $n = 200$
  - Sampling 1: -0.01, Sampling 2: 0.11, Sampling 3: -0.12

Illustration 2: Normal(0, 1) samples

$$X_1, \ldots, X_n \sim \text{ iid Normal}(0, 1)$$

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$$

- $n = 5$
  - Samples: 2.17, 0.10, -0.75, -1.05 , -1.72; Sample mean: -0.25
  - Samples: -0.26, 0.12, -0.31, -0.07, 1.35; Sample mean: 0.17
  - Samples: -0.20, 0.37, 1.00, -0.41, -0.21; Sample mean: 0.11
- $n = 20$
  - Sampling 1: 0.08, Sampling 2: -0.24, Sampling 3: 0.41
- $n = 200$
  - Sampling 1: -0.01, Sampling 2: 0.11, Sampling 3: -0.12
- $n = 1000$

Illustration 2: Normal(0, 1) samples

$$X_1, \ldots, X_n \sim \text{ iid Normal}(0, 1)$$

Distribution mean

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$$

- $n = 5$
  - Samples: 2.17, 0.10, -0.75, -1.05 , -1.72; Sample mean: -0.25
  - Samples: -0.26, 0.12, -0.31, -0.07, 1.35; Sample mean: 0.17
  - Samples: -0.20, 0.37, 1.00, -0.41, -0.21; Sample mean: 0.11
- $n = 20$
  - Sampling 1: 0.08, Sampling 2: -0.24, Sampling 3: 0.41
- $n = 200$
  - Sampling 1: -0.01, Sampling 2: 0.11, Sampling 3: -0.12
- $n = 1000$
  - Sampling 1: 0.04, Sampling 2: -0.04, Sampling 3: -0.02

Expected value and variance of sample mean

Theorem

Let $X_1, X_2, \ldots, X_n$ be iid samples whose distribution has a finite mean μ and variance σ^2. The sample mean $\overline{X} = \frac{X_1+X_2+\cdots+X_n}{n}$ has expected value and variance given by

$$E[\overline{X}] = \mu, Var(\overline{X}) = \frac{\sigma^2}{n}.$$

$$E[\bar{X}] = \frac{\overset{\mu}{E[X_1]} + \overset{\mu}{E[X_2]} + \cdots + \overset{\mu}{E[X_n]}}{n} = \frac{n\mu}{n} = \mu$$

$$E[\bar{X}^2] = \frac{E[X_1^2] + \cdots + E[X_n^2] + 2E[X_1X_2] + \cdots + 2E[X_{n-1}X_n]}{n^2}$$

$$= \frac{nE[X^2] + n(n-1)\mu^2}{n^2} = \frac{n\sigma^2 + \cancel{n\mu^2} + n^2\mu^2 - \cancel{n\mu^2}}{n^2}$$

$$= \frac{\sigma^2}{n} + \mu^2$$

$$Var(\bar{X}) = \frac{\sigma^2}{n}$$

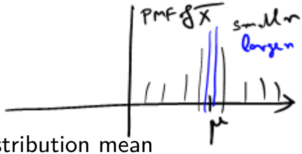
Expected value and variance of sample mean

Theorem

Let $X_1, X_2, \ldots, X_n$ be iid samples whose distribution has a finite mean μ and variance σ^2. The sample mean $\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$ has expected value and variance given by

$$E[\overline{X}] = \mu, Var(\overline{X}) = \frac{\sigma^2}{n}.$$

- Expected value of sample mean equals the expected value or mean of the distribution
  - Mean of distribution: constant real number and not random
  - Sample mean: random variable with mean equal to distribution mean
- Variance of sample mean decreases with n
  - As n increases. . .
    - variance of sample mean tends to zero
    - the spread of sample mean will decrease
    - sample mean will take values close to the distribution mean



Sample variance

Definition (Sample variance)

Let $X_1, X_2, \ldots, X_n$ be iid samples. The sample variance, denoted S^2, is defined to be the random variable

$$S^2 = \frac{(X_1 - \overline{X})^2 + (X_2 - \overline{X})^2 + \cdots + (X_n - \overline{X})^2}{n-1},$$

where $\overline{X}$ is the sample mean.

- Given a sampling $x_1, \ldots, x_n$, the value taken by the sample variance S^2 is $s^2 = ((x_1 - \overline{x})^2 + \cdots + (x_n - \overline{x})^2)/(n-1)$. Often, S^2 and s^2 are both called sample variance.

Sample variance

Definition (Sample variance)

Let $X_1, X_2, \ldots, X_n$ be iid samples. The sample variance, denoted S^2, is defined to be the random variable

$$S^2 = \frac{(X_1 - \overline{X})^2 + (X_2 - \overline{X})^2 + \cdots + (X_n - \overline{X})^2}{n-1},$$

where $\overline{X}$ is the sample mean.

- Given a sampling $x_1, \ldots, x_n$, the value taken by the sample variance S^2 is $s^2 = ((x_1 - \overline{x})^2 + \cdots + (x_n - \overline{x})^2)/(n-1)$. Often, S^2 and s^2 are both called sample variance.
- Why $n-1$ in the denominator instead of n?
  - Some books use n (this causes confusion)
  - Expected value of sample variance is simple in this case

Expected value of sample variance

Theorem

Let $X_1, X_2, \ldots, X_n$ be iid samples whose distribution has a finite variance σ^2. The sample variance $S^2 = \frac{(X_1-\overline{X})^2+\cdots+(X_n-\overline{X})^2}{n}$ has expected value given by

$$E[S^2] = \sigma^2.$$

Expected value of sample variance

Theorem

Let $X_1, X_2, \ldots, X_n$ be iid samples whose distribution has a finite variance σ^2. The sample variance $S^2 = \frac{(X_1-\overline{X})^2+\cdots+(X_n-\overline{X})^2}{n}$ has expected value given by

$$E[S^2] = \sigma^2.$$

- Expected value of sample variance equals the variance of the distribution
  - Variance of distribution: constant real number and not random
  - Sample variance: random variable with mean equal to distribution variance
- Values of sample variance, on average, give the variance of distribution
  - Variance of sample variance will decrease with number of samples (in most cases)
  - As n increases, sample variance takes values close to distribution variance

Illustration

- Bernoulli(1/2), mean = 0.5, variance = 0.25
  - Sample variance values: $n = 20$
    - 0.26, 0.26, 0.26, 0.25, 0.26
  - Sample variance values: $n = 200$
    - 0.2500, 0.2487, 0.2496, 0.2456, 0.2476
  - Sample variance values: $n = 1000$
    - 0.2498, 0.2490, 0.2499, 0.2501, 0.2502

Illustration

- Bernoulli(1/2), mean $= 0.5$, variance $= 0.25$
  - Sample variance values: $n = 20$
    - 0.26, 0.26, 0.26, 0.25, 0.26
  - Sample variance values: $n = 200$
    - 0.2500, 0.2487, 0.2496, 0.2456, 0.2476
  - Sample variance values: $n = 1000$
    - 0.2498, 0.2490, 0.2499, 0.2501, 0.2502
- Normal(0,1), mean $= 0$, variance $= 1$
  - Sample variance values: $n = 20$
    - 0.89, 0.57, 1.19, 1.01, 1.41
  - Sample variance values: $n = 200$
    - 0.93, 1.07, 0.85, 0.83, 1.09
  - Sample variance values: $n = 1000$
    - 1.0268, 0.9535, 0.9781, 0.9766, 0.9831

Sample proportion

$$X_1, X_2, \ldots, X_n \overset{iid}{\sim} X$$

- *iid* samples from the distribution of X
- Let A be an event defined using X
  - Example: $A = (X > t)$, $A = (a < X < b)$ etc

Sample proportion

$$X_1, X_2, \ldots, X_n \sim X$$

- *iid* samples from the distribution of X
- Let A be an event defined using X
  - Example: $A = (X > t)$, $A = (a < X < b)$ etc

Definition (Sample proportion)

The sample proportion of A, denoted $S(A)$, is defined as

$$S(A) = \frac{\#(X_i \text{ for which } A \text{ is true})}{n}.$$

Sample proportion

$$X_1, X_2, \ldots, X_n \sim X$$

- *iid* samples from the distribution of X
- Let A be an event defined using X
  - Example: $A = (X > t)$, $A = (a < X < b)$ etc

Definition (Sample proportion)

The sample proportion of A, denoted $S(A)$, is defined as

$$S(A) = \frac{\#(X_i \text{ for which } A \text{ is true})}{n}.$$

- Samples: 0, 1, 1, 1, 0
  - $S(X = 1) = 3/5$
- Samples: -0.2, 1.1, 0.3, -1.2, 0.7
  - $S(X \leq 0) = 2/5$, $S(0 < X < 1) = 2/5$, $S(X > 1) = 1/5$

Expected value and variance of sample proportion

Theorem

Let $X_1, X_2, \ldots, X_n$ be iid samples from the distribution of X. Let A be an event defined using X and let $P(A)$ be the probability of A. The sample proportion of A, denoted $S(A)$, has expected value and variance given by

$$E[S(A)] = P(A), Var(S(A)) = \frac{P(A)(1 - P(A))}{n}.$$

Proof

mean $= P(A)$
variance $= P(A)(1-P(A))$

- Convert samples into Bernoulli($P(A)$) samples $Y_1, \ldots, Y_n$
  - $Y_i = 1$ if A is true for X_i, and $Y_i = 0$ otherwise
- $S(A)$ is the sample mean of $Y_1, \ldots, Y_n$ ~ iid Bernoulli (A)

Expected value and variance of sample proportion

Theorem

Let $X_1, X_2, \ldots, X_n$ be iid samples from the distribution of X. Let A be an event defined using X and let $P(A)$ be the probability of A. The sample proportion of A, denoted $S(A)$, has expected value and variance given by

$$E[S(A)] = P(A), Var(S(A)) = \frac{P(A)(1 - P(A))}{n}.$$

Proof

- Convert samples into Bernoulli($P(A)$) samples $Y_1, \ldots, Y_n$
  - $Y_i = 1$ if A is true for X_i, and $Y_i = 0$ otherwise
- $S(A)$ is the sample mean of $Y_1, \ldots, Y_n$
- As n increases, values of $S(A)$ will be close to $P(A)$
  - Mean of $S(A)$ equals $P(A)$
  - Variance of $S(A)$ tends to 0

Illustration

$$X_1, \ldots, X_n \sim \text{Normal}(0,1)$$

$\int_{-\infty}^{-1} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = 0.159\ldots$

- $P(X \leq -1) = 0.159$ ← from distribution
  - Sample proportion values: $n = 20$
    - ⋆ 0.15, 0.20, 0.15, 0.15, 0.15 (5 samplings)
  - Sample proportion values: $n = 200$
    - ⋆ 0.170, 0.140, 0.150, 0.155, 0.165 (5 samplings)
  - Sample proportion values: $n = 1000$
    - ⋆ 0.160, 0.180, 0.162, 0.135, 0.153 (5 samplings)
- $P(-1 < X < 1) = 0.683$
  - Sample proportion values: $n = 20$
    - ⋆ 0.75, 0.70, 0.55, 0.45, 0.70
  - Sample proportion values: $n = 200$
    - ⋆ 0.705, 0.690, 0.705, 0.670, 0.720
  - Sample proportion values: $n = 1000$
    - ⋆ 0.678, 0.678, 0.686, 0.679, 0.681

Where have we seen *iid* samples?

- Benoulli trials: Sample mean tends to distribution mean
  - Bernoulli(p) samples
  - Distribution mean $= p$
  - Sample mean $=$ fraction of successes
- Monte carlo simulations
  - Sample proportion tends to actual probability
- Computing histograms
  - Sample proportion tends to actual probability

Subsection 3

Illustrations with data

Iris data

- 3 classes of irises: 0, 1, 2
  - 50 instances of data for each class
  - Each instance: [sepal length, sepal width, petal length, petal width] (cm)

Iris data

- 3 classes of irises: 0, 1, 2
  - 50 instances of data for each class
  - Each instance: [sepal length, sepal width, petal length, petal width] (cm)
- Sepal length of Class 0
  - Model: *iid* samples according to some unknown distribution
  - Data: 5.1, 4.9, 4.7, ..., 5.3, 5
  - Sample mean: 5.006, Sample variance: $0.1242 = 0.3524^2$
  - S(Sepal length $> 5) = 22/50$, S($4.8 <$ Sepal length $< 5.2) = 20/50$

Iris data

- 3 classes of irises: 0, 1, 2
  - 50 instances of data for each class
  - Each instance: [sepal length, sepal width, petal length, petal width] (cm)
- Sepal length of Class 0
  - Model: *iid* samples according to some unknown distribution
  - Data: 5.1, 4.9, 4.7, $\ldots$, 5.3, 5
  - Sample mean: 5.006, Sample variance: $0.1242 = 0.3524^2$
  - $S(\text{Sepal length} > 5) = 22/50$, $S(4.8 < \text{Sepal length} < 5.2) = 20/50$
- Petal width of Class 3
  - Model: *iid* samples according to some unknown distribution
  - Data: 2.5, 1.9, 2.1, $\ldots$, 2.3, 1.8
  - Sample mean: 2.026, Sample variance: $0.0754 = 0.2746^2$
  - $S(\text{Petal width} > 2) = 23/50$, $S(1.8 < \text{Petal length} < 2.2) = 17/50$

Iris data

- 3 classes of irises: 0, 1, 2
  - 50 instances of data for each class
  - Each instance: [sepal length, sepal width, petal length, petal width] (cm)
- Sepal length of Class 0
  - Model: *iid* samples according to some unknown distribution
  - Data: 5.1, 4.9, 4.7, . . . , 5.3, 5
  - Sample mean: 5.006, Sample variance: $0.1242 = 0.3524^2$
  - S(Sepal length $> 5) = 22/50$, S($4.8 <$ Sepal length $< 5.2) = 20/50$
- Petal width of Class 3
  - Model: *iid* samples according to some unknown distribution
  - Data: 2.5, 1.9, 2.1, . . . , 2.3, 1.8
  - Sample mean: 2.026, Sample variance: $0.0754 = 0.2746^2$
  - S(Petal width $> 2) = 23/50$, S($1.8 <$ Petal length $< 2.2) = 17/50$
- Model: how good is the iid samples model?

Taj Mahal air quality

Date	SO2	NO2	PM2.5	PM10
12/4	4	60	77	185
13/4	4	53	65	196
11/4	4	57	72	223
10/4	4	45	68	200
8/4	5	33	52	250
7/4	4	27	67	266
6/4	4	12	60	219
5/4	7	27	70	207
4/4	4	58	100	282
3/4	4	17	55	158
1/4	4	31	37	465
Max	80	80	60	100

- 24-hour average of particles in air, units: micrograms/cubic metre

Taj Mahal air quality sample statistics

- Sample means
  - SO2: 4.36, NO2: 38.18, PM2.5: 65.72, PM10: 241
- Sample variance/ standard deviations
  - SO2: 0.9244^2, NO2: 17.1803^2, PM2.5: 15.9002^2, PM10: 82.6184^2
- S(max exceeded)
  - SO2: 0, NO2: 0, PM2.5: 7/11, PM10: 11/11

Taj Mahal air quality sample statistics

- Sample means
  - SO2: 4.36, NO2: 38.18, PM2.5: 65.72, PM10: 241
- Sample standard deviations
  - SO2: 0.9244^2, NO2: 17.1803^2, PM2.5: 15.9002^2, PM10: 82.6184^2
- S(max exceeded)
  - SO2: 0, NO2: 0, PM2.5: 7/11, PM10: 11/11
- Model
  - Do you like the iid samples model for this data?
  - Is the Taj in trouble or not? How do we answer such questions?
  - How confident are our conclusions when we have looked at just 11 data points?

IPL: Runs scored in Deliveries 0.1, 0.2, 0.3

- Data from 1598 innings
  - See shared spreadsheet
  - Download csv from cricsheet.org
- All calculations done using spreadsheets or other computing tools

IPL: Runs scored in Deliveries 0.1, 0.2, 0.3

- Data from 1598 innings
  - See shared spreadsheet
  - Download csv from cricsheet.org
- All calculations done using spreadsheets or other computing tools

- Sample means
  - 0.1: 0.7347, 0.2: 0.8686, 0.3: 0.9524
- Sample variances
  - 0.1: 1.4975, 0.2: 1.7961, 0.3: 2.0666
- Sample proportions
  - S(dot ball) - 0.1: 0.5989, 0.2: 0.5551, 0.3: 0.5338
  - S(4 or 6) - 0.1: 0.0914, 0.2: 0.1145, 0.3: 0.1302

IPL: Runs scored in Deliveries 0.1, 0.2, 0.3

- Data from 1598 innings
  - See shared spreadsheet
  - Download csv from cricsheet.org
- All calculations done using spreadsheets or other computing tools

- Sample means
  - 0.1: 0.7347, 0.2: 0.8686, 0.3: 0.9524
- Sample variances
  - 0.1: 1.4975, 0.2: 1.7961, 0.3: 2.0666
- Sample proportions
  - S(dot ball) - 0.1: 0.5989, 0.2: 0.5551, 0.3: 0.5338
  - S(4 or 6) - 0.1: 0.0914, 0.2: 0.1145, 0.3: 0.1302

- Clear trend from samples
  - Runs scored increases from 0.1 to 0.3
- Enough data points to be confident in the trend
  - Agrees with intuition
- Model: Do you like the iid samples model for each delivery?

Subsection 4

Sum of independent random variables

Expected value and variance

Theorem

Let $X_1, X_2, \ldots, X_n$ be random variables. Let $S = X_1 + \cdots + X_n$ be their sum. Then,

$$E[S] = E[X_1] + \cdots + E[X_n].$$

If $X_1, \ldots, X_n$ are pairwise uncorrelated, then

$$Var(S) = Var(X_1) + \cdots + Var(X_n).$$

- What is pairwise uncorrelated? $E[X_iX_j] = E[X_i]E[X_j]$ for all i, j, $i \neq j$

$$Var(S) = E[(S - E[S])^2]$$

$$\big((X_1 - E[X_1]) + (X_2 - E[X_2]) + \cdots + (X_n - E[X_n])\big)^2$$

$$= (X_1 - E[X_1])^2 + (X_2 - E[X_2])^2 + \cdots + (X_n - E[X_n])^2$$

$$+ 2(X_1 - E[X_1])(X_2 - E[X_2]) + \cdots \quad \Big\{ E[(X_i - E[X_i])(X_j - E[X_j])] = 0 \Big\}$$

Expected value and variance

Theorem

Let $X_1, X_2, \ldots, X_n$ be random variables. Let $S = X_1 + \cdots + X_n$ be their sum. Then,

$$E[S] = E[X_1] + \cdots + E[X_n].$$

If $X_1, \ldots, X_n$ are pairwise uncorrelated, then

$$Var(S) = Var(X_1) + \cdots + Var(X_n).$$

- What is pairwise uncorrelated? $E[X_i X_j] = E[X_i]E[X_j]$ for all i, j, $i \neq j$
- Mean of sum is sum of means
- If uncorrelated, variance of sum is sum of variances

Expected value and variance

Theorem

Let $X_1, X_2, \ldots, X_n$ be random variables. Let $S = X_1 + \cdots + X_n$ be their sum. Then,

$$E[S] = E[X_1] + \cdots + E[X_n].$$

If $X_1, \ldots, X_n$ are pairwise uncorrelated, then

$$Var(S) = Var(X_1) + \cdots + Var(X_n).$$

- What is pairwise uncorrelated? $E[X_i X_j] = E[X_i]E[X_j]$ for all i, j, $i \neq j$
- Mean of sum is sum of means
- If uncorrelated, variance of sum is sum of variances
- If the X_i are independent, they are also uncorrelated
  - So, above result holds for independent random variables

Extensions of previous result

- Scaling and summing
  - Suppose $S = a_1X_1 + \cdots + a_nX_n$, where a_i are constants
  - $E[S] = a_1E[X_1] + \cdots + a_nE[X_n]$
  - $\text{Var}(S) = a_1^2\text{Var}(X_1) + \cdots + a_n^2\text{Var}(X_n)$, if pairwise uncorrelated

Extensions of previous result

- Scaling and summing
  - Suppose $S = a_1X_1 + \cdots + a_nX_n$, where a_i are constants
  - $E[S] = a_1E[X_1] + \cdots + a_nE[X_n]$
  - $\text{Var}(S) = a_1^2\text{Var}(X_1) + \cdots + a_n^2\text{Var}(X_n)$, if uncorrelated
- *iid* samples: $X_1, \ldots, X_n \sim X$, iid
  - Suppose $S = a_1X_1 + \cdots + a_nX_n$, where a_i are constants
  - $E[S] = (a_1 + \cdots + a_n)E[X]$
  - $\text{Var}(S) = (a_1^2 + \cdots + a_n^2)\text{Var}(X)$

Extensions of previous result

- Scaling and summing
  - Suppose $S = a_1X_1 + \cdots + a_nX_n$, where a_i are constants
  - $E[S] = a_1E[X_1] + \cdots + a_nE[X_n]$
  - $\text{Var}(S) = a_1^2\text{Var}(X_1) + \cdots + a_n^2\text{Var}(X_n)$, if uncorrelated
- *iid* samples: $X_1, \ldots, X_n \sim X$, iid
  - Suppose $S = a_1X_1 + \cdots + a_nX_n$, where a_i are constants
  - $E[S] = (a_1 + \cdots + a_n)E[X]$
  - $\text{Var}(S) = (a_1^2 + \cdots + a_n^2)\text{Var}(X)$
- Sample mean: $X_1, \ldots, X_n \sim X$, iid
  - $\overline{X} = (X_1 + \cdots + X_n)/n$, $a_i = 1/n$
  - $E[(\overline{X})] = E[X]$
  - $\text{Var}((\overline{X})) = \text{Var}(X)/n$

Sample mean versus distribution mean

$$X_1, \ldots, X_n \sim \text{ iid } X$$

- Let $\mu = E[X]$, $\sigma^2 = \mathrm{Var}(X)$

Sample mean versus distribution mean

$$X_1, \ldots, X_n \sim \text{ iid } X$$

- Let $\mu = E[X]$, $\sigma^2 = \text{Var}(X)$
- Sample mean: $\overline{X} = (X_1 + \cdots + X_n)/n$
  - Expected value: μ, Variance: σ^2/n
  - Variance (or spread) goes to 0 as n grows

Can we say something more precise about $\overline{X}$ and μ?

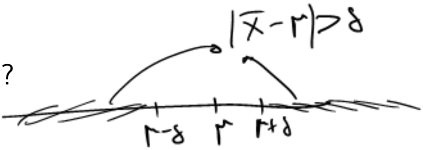
Sample mean versus distribution mean

$$X_1, \ldots, X_n \sim \text{ iid } X$$

- Let $\mu = E[X]$, $\sigma^2 = \text{Var}(X)$
- Sample mean: $\overline{X} = (X_1 + \cdots + X_n)/n$
  - Expected value: μ, Variance: σ^2/n
  - Variance (or spread) goes to 0 as n grows

Can we say something more precise about $\overline{X}$ and μ?

- What is $P(\overline{X} > \mu + \delta)$?
- What is $P(\overline{X} < \mu - \delta)$?
- What is $P(|\overline{X} - \mu| > \delta)$?



Weak law of large numbers

$$X_1, \ldots, X_n \sim \text{ iid } X$$

- Let $\mu = E[X]$, $\sigma^2 = \text{Var}(X)$
- Sample mean: $\overline{X} = (X_1 + \cdots + X_n)/n$
    - Expected value: μ, Variance: σ^2/n

Weak law of large numbers

$$X_1, \ldots, X_n \sim \text{ iid } X$$

- Let $\mu = E[X]$, $\sigma^2 = \text{Var}(X)$
- Sample mean: $\overline{X} = (X_1 + \cdots + X_n)/n$
  - Expected value: μ, Variance: σ^2/n

Theorem (Weak law of large numbers)

$$P(|\overline{X} - \mu| > \delta) \leq \frac{\sigma^2}{n\delta^2} \to 0.$$

Chebyshev: $P\left((\bar{X}-\mu)^2 > \delta^2\right) \leq \frac{E[(\bar{X}-\mu)^2]}{\delta^2} = \frac{\sigma^2}{n\delta^2} \xrightarrow[\substack{n\to\infty \\ \text{fixed } \delta}]{} 0$

Weak law of large numbers

$$X_1, \ldots, X_n \sim \text{ iid } X$$

- Let $\mu = E[X]$, $\sigma^2 = \text{Var}(X)$
- Sample mean: $\overline{X} = (X_1 + \cdots + X_n)/n$
  - Expected value: μ, Variance: σ^2/n

Theorem (Weak law of large numbers)

$$P(|\overline{X} - \mu| > \delta) \leq \frac{\sigma^2}{n\delta^2} \to 0.$$

- With probability more than $1 - \frac{\sigma^2}{n\delta^2}$, sample mean lies in $[\mu - \delta, \mu + \delta]$
  - **What is the meaning of this probability?**

Weak law of large numbers

$$X_1, \ldots, X_n \sim \text{ iid } X$$

$\overline{X}$: converges to μ "in probability"

- Let $\mu = E[X]$, $\sigma^2 = \text{Var}(X)$
- Sample mean: $\overline{X} = (X_1 + \cdots + X_n)/n$
  - Expected value: μ, Variance: σ^2/n

Theorem (Weak law of large numbers)

$$P(|\overline{X} - \mu| > \delta) \leq \frac{\sigma^2}{n\delta^2} \to 0.$$

- With probability more than $1 - \frac{\sigma^2}{n\delta^2}$, sample mean lies in $[\mu - \delta, \mu + \delta]$
  - **What is the meaning of this probability?**
- Chebyshev is usually a very "weak" bound, and we will see sharper bounds soon

Examples: n iid samples

- Bernoulli(p) samples $\mu = p, \sigma^2 = p(1-p)$
  - With probability more than $1 - \frac{p(1-p)}{n\delta^2}$, sample mean lies in $[p-\delta, p+\delta]$
- Uniform$\{-M, \ldots, M\}$ samples $\mu = 0, \sigma^2 = \frac{M(M+1)}{3}$
  - With probability more than $1 - \frac{M(M+1)}{3n\delta^2}$, sample mean lies in $[-\delta, \delta]$
- Normal$(0, \sigma^2)$ samples
  - With probability more than $1 - \frac{\sigma^2}{n\delta^2}$, sample mean lies in $[-\delta, \delta]$
- Uniform$[-A, A]$ samples Continuous $\mu = 0, \sigma^2 = A^2/3$
  - With probability more than $1 - \frac{A^2}{3n\delta^2}$, sample mean lies in $[-\delta, \delta]$

Examples: n iid samples

- Bernoulli(p) samples
  - With probability more than $1 - \frac{p(1-p)}{n\delta^2}$, sample mean lies in $[p - \delta, p + \delta]$
- Uniform$\{-M, \ldots, M\}$ samples
  - With probability more than $1 - \frac{M(M+1)}{3n\delta^2}$, sample mean lies in $[-\delta, \delta]$
- Normal$(0, \sigma^2)$ samples
  - With probability more than $1 - \frac{\sigma^2}{n\delta^2}$, sample mean lies in $[-\delta, \delta]$
- Uniform$[-A, A]$ samples
  - With probability more than $1 - \frac{A^2}{3n\delta^2}$, sample mean lies in $[-\delta, \delta]$

- When distribution is known, a precise statement is possible about "confidence" of finding sample mean within a certain precise interval
  - Improvement in bound will improve precision

Examples: Iris, Taj Mahal and IPL

- Iris data: Sepal length
  - $n = 50$, Sample mean: 5.006, Sample variance: 0.1242
  - With probability more than $1 - \sigma^2/50\delta^2$, sample mean lies in $[\mu - \delta, \mu + \delta]$
    - Works for $\delta > \sigma/\sqrt{50}$

Examples: Iris, Taj Mahal and IPL

- Iris data: Sepal length
  - $n = 50$, Sample mean: 5.006, Sample variance: 0.1242
  - With probability more than $1 - \sigma^2/50\delta^2$, sample mean lies in $[\mu - \delta, \mu + \delta]$
    - ★ Works for $\delta > \sigma/\sqrt{50}$

$1 - \frac{\sigma^2}{50\delta^2} > 0.95 \Rightarrow \delta^2 > \frac{\sigma^2}{50 \times 0.05}$

$\delta > \sqrt{\frac{2}{5}}\,\sigma$

- Taj Mahal air quality: PM2.5
  - $n = 11$, Sample mean $= 65.72$, Sample variance $= 15.9^2$
  - With probability more than $1 - \sigma^2/11\delta^2$, sample mean lies in $[\mu - \delta, \mu + \delta]$
    - ★ Works for $\delta > \sigma/\sqrt{11}$

$1 - \frac{\sigma^2}{11\delta^2} > 0.95 \Rightarrow \delta^2 > \frac{\sigma^2}{11 \times 0.05}$

$\Downarrow$

$\delta^2 > \frac{20\,\sigma^2}{11}$

$\Downarrow$

$\delta > \sqrt{\frac{20}{11}}\,\sigma$

Examples: Iris, Taj Mahal and IPL

- Iris data: Sepal length
  - $n = 50$, Sample mean: 5.006, Sample variance: 0.1242
  - With probability more than $1 - \sigma^2/50\delta^2$, sample mean lies in $[\mu - \delta, \mu + \delta]$
    - ★ Works for $\delta > \sigma/\sqrt{50}$
- Taj Mahal air quality: PM2.5
  - $n = 11$, Sample mean $= 65.72$, Sample variance $= 15.9^2$
  - With probability more than $1 - \sigma^2/11\delta^2$, sample mean lies in $[\mu - \delta, \mu + \delta]$
    - ★ Works for $\delta > \sigma/\sqrt{11}$
- IPL: Runs scored in Delivery 0.3
  - $n = 1598$, Sample mean: 0.9524, Sample variance: 2.0666
  - With probability more than $1 - \sigma^2/1598\delta^2$, sample mean lies in $[\mu - \delta, \mu + \delta]$
    - ★ Works for $\delta > \sigma/\sqrt{1598}$

Examples: Iris, Taj Mahal and IPL

- Iris data: Sepal length
  - $n = 50$, Sample mean: 5.006, Sample variance: 0.1242
  - With probability more than $1 - \sigma^2/50\delta^2$, sample mean lies in $[\mu - \delta, \mu + \delta]$
    - ★ Works for $\delta > \sigma/\sqrt{50}$
- Taj Mahal air quality: PM2.5
  - $n = 11$, Sample mean $= 65.72$, Sample variance $= 15.9^2$
  - With probability more than $1 - \sigma^2/11\delta^2$, sample mean lies in $[\mu - \delta, \mu + \delta]$
    - ★ Works for $\delta > \sigma/\sqrt{11}$
- IPL: Runs scored in Delivery 0.3
  - $n = 1598$, Sample mean: 0.9524, Sample variance: 2.0666
  - With probability more than $1 - \sigma^2/1598\delta^2$, sample mean lies in $[\mu - \delta, \mu + \delta]$
    - ★ Works for $\delta > \sigma/\sqrt{1598}$
- What to do when distribution is unknown? Have to assume something

Subsection 5

Sum of independent random variables II

Subsection 6

Concentration phenomenon

Bounding $P(|\overline{X} - \mu| > t)$

$$X_1, \ldots, X_n \sim \text{iid } X$$

$\mu = E[X]$
$\sigma^2 = Var(X)$

- Sample mean $\overline{X} = (X_1 + \cdots + X_n)/n$
- Chebyshev bound

$$P(|\overline{X} - \mu| > \delta) \leq \frac{\sigma^2}{n\delta^2}$$

Bounding $P(|\overline{X} - \mu| > t)$

$$X_1, \ldots, X_n \sim \text{iid } X$$

- Sample mean $\overline{X} = (X_1 + \cdots + X_n)/n$
- Chebyshev bound

$$P(|\overline{X} - \mu| > \delta) \leq \frac{\sigma^2}{n\delta^2}$$

How weak is Chebyshev?

$X_1 + \cdots + X_n \sim \text{Binom}(n, 1/2)$

$\{0, 1\}$ with $1/2$, $1/2$

- Let $X \sim \text{Bernoulli}(1/2)$, $\mu = 0.5$, $\sigma^2 = 0.25$
  - $n = 10$: $P(|\overline{X} - 0.5| > 0.3) = 0.0215 \leq 0.278$
  - $n = 50$: $P(|\overline{X} - 0.5| > 0.3) = 5.61 \times 10^{-6} \leq 0.056$

$n=10$: $P(|\overline{X} - 1/2| > 0.3)$

$= P(X_1 + \cdots + X_{10} > 8$ (or) $X_1 + \cdots + X_{10} < 2)$

$\text{Binom}(10, 1/2)$

Chebyshev

$n = 50$

$\text{Binom}(50, 1/2)$

$P(X_1 + \cdots + X_{50} > 40$ (or) $X_1 + \cdots + X_{50} < 10)$

Bounding $P(|\overline{X} - \mu| > t)$

$$X_1, \ldots, X_n \sim \text{iid } X$$

- Sample mean $\overline{X} = (X_1 + \cdots + X_n)/n$
- Chebyshev bound

$$P(|\overline{X} - \mu| > \delta) \leq \frac{\sigma^2}{n\delta^2}$$

How weak is Chebyshev?

- Let $X \sim \text{Bernoulli}(1/2)$, $\mu = 0.5$, $\sigma^2 = 0.25$
  - $n = 10$: $P(|\overline{X} - 0.5| > 0.3) = 0.0215 \leq 0.278$
  - $n = 50$: $P(|\overline{X} - 0.5| > 0.3) = 5.61 \times 10^{-6} \leq 0.056$

- Chebyshev falls as $1/n$
- In many cases, we can have e^{-cn}
  - Exponential fall with n
  - Much much faster than $1/n$

Markov, Chebyshev and Chernoff

- Markov inequality: X takes positive values

$$P(X > t) \leq \frac{E[X]}{t}$$

- Chebyshev inequality: X could take positive/negative values with finite variance

$$P(|X - E[X]| > t) = P((X - E[X])^2 > t^2) \overset{\text{Markov}}{\leq} \frac{\overset{= E[(X-E[X])^2]}{\mathrm{Var}(X)}}{t^2}$$

Markov, Chebyshev and Chernoff

- Markov inequality: X takes positive values

$$P(X > t) \leq \frac{E[X]}{t}$$

- Chebyshev inequality: X could take positive/negative values with finite variance

$$P(|X - E[X]| > t) = P((X - E[X])^2 > t^2) \leq \frac{\text{Var}(X)}{t^2}$$

- Chernoff inequality: X could take positive/negative values, $E[X] = 0$

+ve values   markov

$$P(X > t) = P(e^{\lambda X} > e^{\lambda t}) \leq \frac{E[e^{\lambda X}]}{e^{\lambda t}}, \ \underline{\lambda > 0}$$

Markov, Chebyshev and Chernoff

- Markov inequality: X takes positive values

$$P(X > t) \leq \frac{E[X]}{t}$$

- Chebyshev inequality: X could take positive/negative values with finite variance

$$P(|X - E[X]| > t) = P((X - E[X])^2 > t^2) \leq \frac{\mathrm{Var}(X)}{t^2}$$

- Chernoff inequality: X could take positive/negative values, $E[X] = 0$

$$P(X > t) = P(e^{\lambda X} > e^{\lambda t}) \leq \frac{E[e^{\lambda X}]}{e^{\lambda t}},\ \lambda > 0$$

- **Moment generating function (MGF)** of X: $E[e^{\lambda X}]$ (for $E[X] = 0$)
  - Pick λ that provides best bound
  - MGF too unwieldy? Use upper bound on MGF

MGF of Centralised Bernoulli(1/2)

- What is centralising?
  - X: random variable with mean $E[X]$
  - Centralised version of X: $X - E[X]$

MGF of Centralised Bernoulli(1/2)

- What is centralising?
  - X: random variable with mean $E[X]$
  - Centralised version of X: $X - E[X]$
- Centralised Bernoulli(1/2)
  - Bernoulli(1/2): $\{\overset{1/2}{0}, \overset{1/2}{1}\}$, mean $= 1/2$
  - Centralised: $X \sim \{\overset{1/2}{-1/2}, \overset{1/2}{1/2}\}$, mean $= 0$

MGF of Centralised Bernoulli(1/2)

- What is centralising?
  - X: random variable with mean $E[X]$
  - Centralised version of X: $X - E[X]$
- Centralised Bernoulli(1/2)
  - Bernoulli(1/2): $\{\overset{1/2}{0}, \overset{1/2}{1}\}$, mean $= 1/2$
  - Centralised: $X \sim \{\overset{1/2}{-1/2}, \overset{1/2}{1/2}\}$, mean $= 0$

$$E[e^{\lambda X}] = P(X{=}-1/2)e^{-\lambda/2} + P(X{=}1/2)e^{\lambda/2} = \frac{e^{\lambda/2} + e^{-\lambda/2}}{2}$$

MGF of Centralised Bernoulli(1/2)

- What is centralising?
  - X: random variable with mean $E[X]$
  - Centralised version of X: $X - E[X]$
- Centralised Bernoulli(1/2)
  - Bernoulli(1/2): $\{\overset{1/2}{0}, \overset{1/2}{1}\}$, mean $= 1/2$
  - Centralised: $X \sim \{\overset{1/2}{-1/2}, \overset{1/2}{1/2}\}$, mean $= 0$

$$E[e^{\lambda X}] = P(X{=}-1/2)e^{-\lambda/2} + P(X{=}1/2)e^{\lambda/2} = \frac{e^{\lambda/2} + e^{-\lambda/2}}{2}$$

Bound on MGF

$$E[e^{\lambda X}] = \frac{e^{\lambda/2} + e^{-\lambda/2}}{2} \leq e^{\lambda^2/4}$$

MGF and sum of *iid* random variables

$$X_1, \ldots, X_n \sim \text{iid } X$$

- Let $S = X_1 + \cdots + X_n$. What is MGF of S?
- Why is this question important?
  - MGF gives bounds on $P(S > t)$
  - Sample mean $\overline{X} = S/n$

MGF and sum of *iid* random variables

$$X_1, \ldots, X_n \sim \text{iid } X$$

- Let $S = X_1 + \cdots + X_n$. What is MGF of S?
- Why is this question important?
  - MGF gives bounds on $P(S > t)$
  - Sample mean $\overline{X} = S/n$

$$E[e^{\lambda S}] = E[e^{\lambda X_1} \cdots e^{\lambda X_n}] = E[e^{\lambda X_1}] \cdots E[e^{\lambda X_n}] = E[e^{\lambda X}]^n$$

- **MGF of sum of independent random variables is product of the individual MGFs**

Example: Sum of centralised Benoulli

- $X \sim$ Centralised Bernoulli(1/2), i.e. $\{\overset{1/2}{-1/2}, \overset{1/2}{1/2}\}$

$$X_1, \ldots, X_n \sim \text{iid } X$$

- $S = X_1 + \cdots + X_n$

$$E[e^{\lambda S}] = \left(\frac{e^{\lambda/2} + e^{-\lambda/2}}{2}\right)^n \leq e^{n\lambda^2/4}$$

- **Upper bound is so much easier than MGF!**

Chernoff bound for Binomial(n,1/2)

$$X_1, \ldots, X_n \sim \text{iid Centralised Bernoulli}(1/2)$$

Chernoff bound for Binomial(n,1/2)

$$X_1, \ldots, X_n \sim \text{iid Centralised Bernoulli}(1/2)$$

- $S = X_1 + \cdots + X_n$

$$\text{Chernoff} \quad P(S > t) \le \frac{E[e^{\lambda S}]}{e^{\lambda t}} \le e^{\underbrace{n\lambda^2/4 - \lambda t}_{\text{quadratic in } \lambda}}$$

Chernoff bound for Binomial(n,1/2)

$$X_1, \ldots, X_n \sim \text{iid Centralised Bernoulli}(1/2)$$

- $S = X_1 + \cdots + X_n$

$$P(S > t) \leq \frac{E[e^{\lambda S}]}{e^{\lambda t}} \leq e^{n\lambda^2/4 - \lambda t}$$

- Pick $\lambda = 2t/n$: $P(S > t) \leq e^{-t^2/n}$

Chernoff bound for Binomial(n,1/2)

$$X_1, \ldots, X_n \sim \text{iid Centralised Bernoulli}(1/2)$$

- $S = X_1 + \cdots + X_n$

$$P(S > t) \le \frac{E[e^{\lambda S}]}{e^{\lambda t}} \le e^{n\lambda^2/4 - \lambda t}$$

- Pick $\lambda = 2t/n$: $P(S > t) \le e^{-t^2/n}$

$t = \frac{n\delta}{2}$

- Now, $Y_1 = X_1 + 1/2 \sim$ Bernoulli(1/2). So, $Y = Y_1 + \cdots + Y_n = S + n/2 \sim$ Binomial(n,1/2)

$$P(Y > \overbrace{n/2}^{E[Y]} + \overbrace{n\delta/2}^{\delta E[Y]}) = P(S > n\delta/2) \le e^{-n\delta^2/4}$$

$$\frac{Y}{n} - \frac{1}{2} > \frac{1}{2} \cdot \delta$$

Chernoff bound for Binomial(n,1/2)

$$X_1, \ldots, X_n \sim \text{iid Centralised Bernoulli}(1/2)$$

- $S = X_1 + \cdots + X_n$

$$P(S > t) \leq \frac{E[e^{\lambda S}]}{e^{\lambda t}} \leq e^{n\lambda^2/4 - \lambda t}$$

- Pick $\lambda = 2t/n$: $P(S > t) \leq e^{-t^2/n}$
- Now, $Y_1 = X_1 + 1/2 \sim$ Bernoulli$(1/2)$. So, $Y = Y_1 + \cdots + Y_n = S + n/2 \sim$ Binomial(n,1/2)

$$P(Y > n/2 + n\delta/2) = P(S > n\delta/2) \leq e^{-n\delta^2/4}$$

- Chebyshev: $P(Y > n/2 + n\delta/2) \leq \frac{1}{n\delta^2}$

Chebyshev vs Chernoff: $Y \sim \text{Binomial}(n, 1/2)$

n	Event, $\delta = 0.6$	Prob	$1/n\delta^2$	$e^{-n\delta^2/4}$
10	$Y - 5 > 5 \times 0.6$	0.0107	0.278	0.407
50	$Y - 25 > 25 \times 0.6$	2.81×10^{-6}	0.056	0.011
100	$Y - 50 > 50 \times 0.6$	1.35×10^{-10}	0.028	1.23×10^{-4}
200	$Y - 100 > 100 \times 0.6$	4.16×10^{-19}	0.014	1.52×10^{-8}
400	$Y - 200 > 200 \times 0.6$	5.40×10^{-36}	0.007	2.32×10^{-16}

Chebyshev vs Chernoff: $Y \sim \text{Binomial}(n, 1/2)$

n	Event, $\delta = 0.6$	Prob	$1/n\delta^2$	$e^{-n\delta^2/4}$
10	$Y - 5 > 5 \times 0.6$	0.0107	0.278	0.407
50	$Y - 25 > 25 \times 0.6$	2.81×10^{-6}	0.056	0.011
100	$Y - 50 > 50 \times 0.6$	1.35×10^{-10}	0.028	1.23×10^{-4}
200	$Y - 100 > 100 \times 0.6$	4.16×10^{-19}	0.014	1.52×10^{-8}
400	$Y - 200 > 200 \times 0.6$	5.40×10^{-36}	0.007	2.32×10^{-16}

- $1/n$ vs e^{-cn}: difference is clearly seen as n increases
  - $1/n$ is giving a very wrong idea about the magnitude of the probability

Remarks on concentration phenomenon

$X_1, \ldots, X_n \sim$ iid X, $Y = X_1 + \cdots + X_n$

- Concentration phenomenon

Remarks on concentration phenomenon

$X_1, \ldots, X_n \sim$ iid X, $Y = X_1 + \cdots + X_n$

- Concentration phenomenon
  - Exponential bounds for $P(Y > E[Y] + t)$ by bounding $E[e^{\lambda Y}]$ and using Chernoff

Remarks on concentration phenomenon

$X_1, \ldots, X_n \sim$ iid X, $Y = X_1 + \cdots + X_n$

- Concentration phenomenon
  - Exponential bounds for $P(Y > E[Y] + t)$ by bounding $E[e^{\lambda Y}]$ and using Chernoff
  - We saw this when X is Bernoulli

Remarks on concentration phenomenon

$X_1, \ldots, X_n \sim$ iid X, $Y = X_1 + \cdots + X_n$

- Concentration phenomenon
  - Exponential bounds for $P(Y > E[Y] + t)$ by bounding $E[e^{\lambda Y}]$ and using Chernoff
  - We saw this when X is Bernoulli
- What about $P(Y < E[Y] - t)$?

Remarks on concentration phenomenon

$X_1, \ldots, X_n \sim$ iid X, $Y = X_1 + \cdots + X_n$

- Concentration phenomenon
  - Exponential bounds for $P(Y > E[Y] + t)$ by bounding $E[e^{\lambda Y}]$ and using Chernoff
  - We saw this when X is Bernoulli
- What about $P(Y < E[Y] - t)$?
  - Replace Y with $-Y$ on $P(Y > E[Y] + t)$

Remarks on concentration phenomenon

$X_1, \ldots, X_n \sim$ iid X, $Y = X_1 + \cdots + X_n$

- Concentration phenomenon
  - Exponential bounds for $P(Y > E[Y] + t)$ by bounding $E[e^{\lambda Y}]$ and using Chernoff
  - We saw this when X is Bernoulli
- What about $P(Y < E[Y] - t)$?
  - Replace Y with $-Y$ on $P(Y > E[Y] + t)$
- What about other distributions? Several extensions exist

Remarks on concentration phenomenon

$X_1, \ldots, X_n \sim$ iid X, $Y = X_1 + \cdots + X_n$

- Concentration phenomenon
  - Exponential bounds for $P(Y > E[Y] + t)$ by bounding $E[e^{\lambda Y}]$ and using Chernoff
  - We saw this when X is Bernoulli
- What about $P(Y < E[Y] - t)$?
  - Replace Y with $-Y$ on $P(Y > E[Y] + t)$
- What about other distributions? Several extensions exist
  - Hoeffding's inequality: X is bounded within an interval $[-M, M]$

Remarks on concentration phenomenon

$X_1, \ldots, X_n \sim$ iid X, $Y = X_1 + \cdots + X_n$

- Concentration phenomenon
  - Exponential bounds for $P(Y > E[Y] + t)$ by bounding $E[e^{\lambda Y}]$ and using Chernoff
  - We saw this when X is Bernoulli
- What about $P(Y < E[Y] - t)$?
  - Replace Y with $-Y$ on $P(Y > E[Y] + t)$
- What about other distributions? Several extensions exist
  - Hoeffding's inequality: X is bounded within an interval $[-M, M]$
  - Bennett's inequality: X is bounded in $[-M, M]$ and has finite variance

Remarks on concentration phenomenon

$X_1, \ldots, X_n \sim$ iid X, $Y = X_1 + \cdots + X_n$

- Concentration phenomenon
  - Exponential bounds for $P(Y > E[Y] + t)$ by bounding $E[e^{\lambda Y}]$ and using Chernoff
  - We saw this when X is Bernoulli
- What about $P(Y < E[Y] - t)$?
  - Replace Y with $-Y$ on $P(Y > E[Y] + t)$
- What about other distributions? Several extensions exist
  - Hoeffding's inequality: X is bounded within an interval $[-M, M]$
  - Bennett's inequality: X is bounded in $[-M, M]$ and has finite variance
- Y is the sum of the iid samples. What about other functions $f(X_1, \ldots, X_n)$?

Remarks on concentration phenomenon

$X_1, \ldots, X_n \sim$ iid X, $Y = X_1 + \cdots + X_n$

- Concentration phenomenon
  - Exponential bounds for $P(Y > E[Y] + t)$ by bounding $E[e^{\lambda Y}]$ and using Chernoff
  - We saw this when X is Bernoulli
- What about $P(Y < E[Y] - t)$?
  - Replace Y with $-Y$ on $P(Y > E[Y] + t)$
- What about other distributions? Several extensions exist
  - Hoeffding's inequality: X is bounded within an interval $[-M, M]$
  - Bennett's inequality: X is bounded in $[-M, M]$ and has finite variance
- Y is the sum of the iid samples. What about other functions $f(X_1, \ldots, X_n)$?
  - Many extensions: f should depend "equally" on all variables

Subsection 7

Central Limit Theorem

Moment generating function (MGF)

Definition (MGF)

Let X be a zero-mean random variable. The MGF of X, denoted $M_X(\lambda)$, is a function from $\mathcal{R}$ to $\mathcal{R}$ defined as

$$M_X(\lambda) = E[e^{\lambda X}].$$

Moment generating function (MGF)

Definition (MGF)

Let X be a zero-mean random variable. The MGF of X, denoted $M_X(\lambda)$, is a function from R to R defined as

$$M_X(\lambda) = E[e^{\lambda X}].$$

- X: Discrete with PMF f_X
  - X takes values $\{x_1, x_2, \ldots\}$

$$M_X(\lambda) = f_X(x_1)e^{\lambda x_1} + f_X(x_2)e^{\lambda x_2} + \cdots$$

Moment generating function (MGF)

Definition (MGF)

Let X be a zero-mean random variable. The MGF of X, denoted $M_X(\lambda)$, is a function from R to R defined as

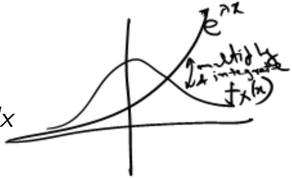
$$M_X(\lambda) = E[e^{\lambda X}].$$

- X: Discrete with PMF f_X
  - X takes values $\{x_1, x_2, \ldots\}$

$$M_X(\lambda) = f_X(x_1)e^{\lambda x_1} + f_X(x_2)e^{\lambda x_2} + \cdots$$

- X: continuous with PDF f_X and support T_X

$$M_X(\lambda) = \int_{x \in T_X} f_X(x) e^{\lambda x} dx$$



Examples

- $X \sim \overset{1}{\{0\}}$ (X is 0 with probability 1)
  - $M_X(\lambda) = 1 \times e^0 = 1$

Examples

- $X \sim \{\overset{1}{0}\}$ (X is 0 with probability 1)
  - $M_X(\lambda) = 1 \times e^0 = 1$
- $X \sim \{\overset{1-p}{-p}, \overset{p}{1-p}\}$ ← Centralised
  - $M_X(\lambda) = (1-p)e^{-p\lambda} + pe^{(1-p)\lambda}$

Examples

- $X \sim \{\overset{1}{0}\}$ (X is 0 with probability 1)
  - $M_X(\lambda) = 1 \times e^0 = 1$
- $X \sim \{\overset{1-p}{-p}, \overset{p}{1-p}\}$
  - $M_X(\lambda) = (1-p)e^{-p\lambda} + pe^{(1-p)\lambda}$
- $X \in \{\overset{1/2}{-1}, \overset{1/4}{0}, \overset{1/4}{2}\}$
  - $M_X(\lambda) = 0.5e^{-\lambda} + 0.25 + 0.25e^{2\lambda}$

Examples

- $X \sim \{\overset{1}{0}\}$ (X is 0 with probability 1)
  - $M_X(\lambda) = 1 \times e^0 = 1$
- $X \sim \{\overset{1-p}{-p}, \overset{p}{1-p}\}$
  - $M_X(\lambda) = (1-p)e^{-p\lambda} + pe^{(1-p)\lambda}$
- $X \in \{\overset{1/2}{-1}, \overset{1/4}{0}, \overset{1/4}{2}\}$
  - $M_X(\lambda) = 0.5e^{-\lambda} + 0.25 + 0.25e^{2\lambda}$
- $M_X(\lambda) = (1/3)e^{3\lambda/2} + (1/6)e^{-3\lambda} + (1/8)e^{-\lambda} + (1/8)e^{\lambda} + 1/8$
  - $X \sim \{\overset{1/6}{-3}, \overset{1/8}{-1}, \overset{1/8}{0}, \overset{1/8}{1}, \overset{1/3}{3/2}\}$

Examples

- $X \sim \{\overset{1}{0}\}$ (X is 0 with probability 1)
  - $M_X(\lambda) = 1 \times e^0 = 1$
- $X \sim \{\overset{1-p}{-p}, \overset{p}{1-p}\}$
  - $M_X(\lambda) = (1-p)e^{-p\lambda} + pe^{(1-p)\lambda}$
- $X \in \{\overset{1/2}{-1}, \overset{1/4}{0}, \overset{1/4}{2}\}$
  - $M_X(\lambda) = 0.5e^{-\lambda} + 0.25 + 0.25e^{2\lambda}$
- $M_X(\lambda) = (1/3)e^{3\lambda/2} + (1/6)e^{-3\lambda} + (1/8)e^{-\lambda} + (1/8)e^{\lambda} + \underbrace{1/2\, e^{\lambda \cdot 0}}$
  - $X \sim \{\overset{1/6}{-3}, \overset{1/8}{-1}, \overset{1/2}{0}, \overset{1/8}{1}, \overset{1/3}{3/2}\}$
- ***** $X \sim \text{Normal}(0, \sigma^2)$ *****
  - $M_X(\lambda) = e^{\lambda^2\sigma^2/2}$

Exercise

$$\int_{-\infty}^{\infty} e^{\lambda x} \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2}\, dx$$

$E[e^{\lambda X}] = \cdots + \underbrace{f_X(x_i)e^{\lambda x_i}} + \cdots$

$X = x_i$ w.p. $f_X(x_i)$

$X = 0$ w.p. $1/4$

Why Moment Generating Function?

$$E[e^{\lambda X}] = E[1 + \lambda X + \frac{\lambda^2}{2!}X^2 + \frac{\lambda^3}{3!}X^3 + \cdots]$$

MGF $$= 1 + \lambda \underbrace{E[X]}_{1^{st}\text{ moment}} + \frac{\lambda^2}{2!}\underbrace{E[X^2]}_{2^{nd}\text{ moment}} + \frac{\lambda^3}{3!}\underbrace{E[X^3]}_{3^{rd}\text{ moment}} + \cdots$$

Why Moment Generating Function?

$$E[e^{\lambda X}] = E[1 + \lambda X + \frac{\lambda^2}{2!}X^2 + \frac{\lambda^3}{3!}X^3 + \cdots]$$
$$= 1 + \lambda E[X] + \frac{\lambda^2}{2!}E[X^2] + \frac{\lambda^3}{3!}E[X^3] + \cdots$$

- $X \sim \text{Normal}(0, \sigma^2)$, $M_X(\lambda) = e^{\lambda^2\sigma^2/2}$

$$1 + \lambda E[X] + \frac{\lambda^2}{2!}E[X^2] + \frac{\lambda^3}{3!}E[X^3] + \cdots = 1 + \frac{\lambda^2}{2!}\sigma^2 + \overbrace{\frac{\lambda^4}{4!}(3\sigma^4)}^{=\frac{(\lambda^2\sigma^2/2)^2}{2!}} + \cdots$$

- $E[X] = 0$, $E[X^2] = \sigma^2$, $E[X^3] = 0$, $E[X^4] = 3\sigma^4$ and so on

Examples: Sum of two independent random variables

$$X_1, X_2 \sim \text{iid } X, Y = X_1 + X_2$$

- $X \sim$ iid Centralised Bernoulli(p) $\{\overset{1-p}{-p}, \overset{p}{1-p}\}$
  - $M_X(\lambda) = (1-p)e^{-p\lambda} + pe^{(1-p)\lambda}$ square
  - $M_Y(\lambda) = M_X(\lambda)^2 = (1-p)^2 e^{-2p\lambda} + 2p(1-p)e^{(1-2p)\lambda} + p^2 e^{2(1-p)\lambda}$
    - ★ $Y \sim \{\overset{(1-p)^2}{-2p}, \overset{2p(1-p)}{1-2p}, \overset{p^2}{2(1-p)}\}$ read distribution

Examples: Sum of two independent random variables

$$X_1, X_2 \sim \text{iid } X, Y = X_1 + X_2$$

- $X \sim$ iid Bernoulli(p)
  - $M_X(\lambda) = (1-p)e^{-p\lambda} + pe^{(1-p)\lambda}$
  - $M_Y(\lambda) = M_X(\lambda)^2 = (1-p)^2 e^{-2p\lambda} + 2p(1-p)e^{(1-2p)\lambda} + p^2 e^{2(1-p)\lambda}$
    - ★ $Y \sim \{\overset{(1-p)^2}{-2p}, \overset{2p(1-p)}{1-2p}, \overset{p^2}{2(1-p)}\}$
- $X \in \{\overset{1/2}{-1}, \overset{1/4}{0}, \overset{1/4}{2}\}$
  - $M_X(\lambda) = 0.5e^{-\lambda} + 0.25 + 0.25e^{2\lambda}$
  - $M_Y(\lambda) = 0.25e^{-2\lambda} + 0.25e^{-\lambda} + 0.0625 + 0.25e^{\lambda} + 0.125e^{2\lambda} + 0.0625e^{4\lambda}$

Examples: Sum of two independent random variables

$$X_1, X_2 \sim \text{iid } X, \, Y = X_1 + X_2$$

- $X \sim$ iid Bernoulli(p)
  - $M_X(\lambda) = (1-p)e^{-p\lambda} + pe^{(1-p)\lambda}$
  - $M_Y(\lambda) = M_X(\lambda)^2 = (1-p)^2 e^{-2p\lambda} + 2p(1-p)e^{(1-2p)\lambda} + p^2 e^{2(1-p)\lambda}$
    - ★ $Y \sim \{\overset{(1-p)^2}{-2p}, \overset{2p(1-p)}{1-2p}, \overset{p^2}{2(1-p)}\}$
- $X \in \{\overset{1/2}{-1}, \overset{1/4}{0}, \overset{1/4}{2}\}$
  - $M_X(\lambda) = 0.5e^{-\lambda} + 0.25 + 0.25e^{2\lambda}$
  - $M_Y(\lambda) = 0.25e^{-2\lambda} + 0.25e^{-\lambda} + 0.0625 + 0.25e^{\lambda} + 0.125e^{2\lambda} + 0.0625e^{4\lambda}$
- $M_X(\lambda) = (1/3)e^{3\lambda/2} + (1/6)e^{-3\lambda} + (1/8)e^{-\lambda} + (1/8)e^{\lambda} + 1/4$
  - $M_Y(\lambda) = \frac{e^{-6t}}{36} + \frac{e^{-4t}}{24} + \frac{e^{-3t}}{12} + \frac{11e^{-2t}}{192} + \frac{1}{9}e^{-3t/2} + \frac{e^{-t}}{16} + \frac{3}{32} + \frac{e^{t/2}}{12} + \frac{e^{t}}{16} + \frac{1}{6}e^{3t/2} + \frac{e^{2t}}{64} + \frac{1}{12}e^{5t/2} + \frac{e^{3t}}{9}$

Example: MGF of sample mean

- Samples: $X_1, \ldots, X_n \sim$ iid X, $M_X(\lambda) = \frac{e^{\lambda/2}+e^{-\lambda/2}}{2}$
  - Sample mean: $\overline{X} = (X_1 + \cdots + X_n)/n$
  - $M_{X/n}(\lambda) = \frac{e^{\lambda/2n}+e^{-\lambda/2n}}{2}$

$\hookrightarrow \frac{X_1}{n} + \frac{X_2}{n} + \cdots + \frac{X_n}{n}$

$\frac{1}{2} e^{\lambda \cdot -1/2n} + \frac{1}{2} \cdot e^{\lambda \cdot 1/2n}$

$$\underline{M_{\overline{X}}(\lambda)} = \left(\frac{e^{\frac{\lambda}{2n}} + e^{-\frac{\lambda}{2n}}}{2} \right)^n$$

fn of λ

X: Centralised Bernoulli(1/2)

$\{-1/2, 1/2\}$ with $1/2$, $1/2$

$\frac{X}{n} \sim \left\{ \frac{-1}{2n}, \frac{1}{2n} \right\}$ with $1/2$, $1/2$

$P_r = 1/4$, $1/2$, P_r

$-1/2$, $-1/2n$, $1/2n$, $1/2$

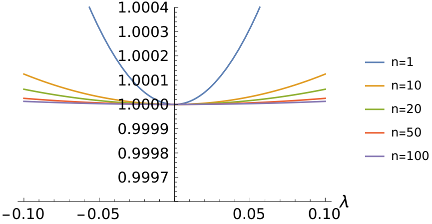
$\frac{X_1}{n}, \frac{X_2}{n}, \ldots, \frac{X_n}{n} \sim$ iid

Example: MGF of sample mean

- Samples: $X_1, \ldots, X_n \sim$ iid X, $M_X(\lambda) = \frac{e^{\lambda/2} + e^{-\lambda/2}}{2}$
  - Sample mean: $\overline{X} = (X_1 + \cdots + X_n)/n$
  - $M_{X/n}(\lambda) = \frac{e^{\lambda/2n} + e^{-\lambda/2n}}{2}$

$$M_{\overline{X}}(\lambda) = \left(\frac{e^{\frac{\lambda}{2n}} + e^{-\frac{\lambda}{2n}}}{2} \right)^n$$

MGF of sample mean for different n



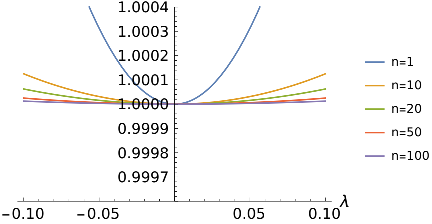
$M_{\overline{X}}(\lambda) \to 1$ as n increases

Example: MGF of sample mean

- Samples: $X_1, \ldots, X_n \sim$ iid X, $M_X(\lambda) = \frac{e^{\lambda/2}+e^{-\lambda/2}}{2}$
  - Sample mean: $\overline{X} = (X_1 + \cdots + X_n)/n$
  - $M_{X/n}(\lambda) = \frac{e^{\lambda/2n}+e^{-\lambda/2n}}{2}$

$$M_{\overline{X}}(\lambda) = \left(\frac{e^{\frac{\lambda}{2n}} + e^{-\frac{\lambda}{2n}}}{2} \right)^n$$

MGF of sample mean for different n



$M_{\overline{X}}(\lambda) \to 1$ as n increases

- WLLN: $\overline{X} \to E[X] = 0$
- Constant 0 has MGF $= 1$

MGF convergence at $1/\sqrt{n}$ scaling

- Samples: $X_1, \ldots, X_n \sim$ iid X, $M_X(\lambda) = \frac{e^{\lambda/2} + e^{-\lambda/2}}{2}$
  - $E[X] = 0$, $\text{Var}(X) = 1/4$
  - Consider $Y = (X_1 + \cdots + X_n)/\sqrt{n}$
  - $M_{X/\sqrt{n}}(\lambda) = \frac{e^{\lambda/2\sqrt{n}} + e^{-\lambda/2\sqrt{n}}}{2}$

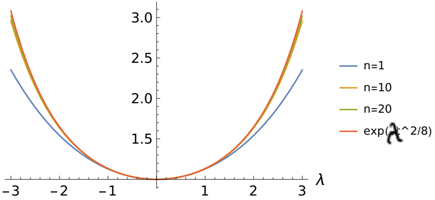
$$M_Y(\lambda) = \left(\frac{e^{\frac{\lambda}{2\sqrt{n}}} + e^{-\frac{\lambda}{2\sqrt{n}}}}{2} \right)^n$$

MGF convergence at $1/\sqrt{n}$ scaling

- Samples: $X_1, \ldots, X_n \sim$ iid X, $M_X(\lambda) = \frac{e^{\lambda/2}+e^{-\lambda/2}}{2}$
  - $E[X] = 0$, $\text{Var}(X) = 1/4 \leftarrow \sigma^2$
  - Consider $Y = (X_1 + \cdots + X_n)/\sqrt{n}$
  - $M_{X/\sqrt{n}}(\lambda) = \frac{e^{\lambda/2\sqrt{n}}+e^{-\lambda/2\sqrt{n}}}{2}$

$$M_Y(\lambda) = \left(\frac{e^{\frac{\lambda}{2\sqrt{n}}} + e^{-\frac{\lambda}{2\sqrt{n}}}}{2} \right)^n$$

MGF of Y for different n



$M_{\overline{X}}(\lambda) \to e^{\lambda^2\sigma^2/2}$ as n increases

$\updownarrow$

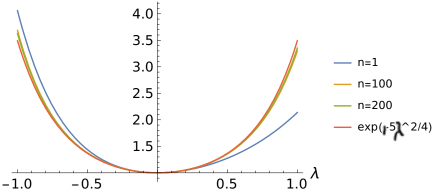
Normal$(0, \sigma^2)$

Another example: MGF convergence at $1/\sqrt{n}$ scaling

- Samples: $X_1, \ldots, X_n \sim$ iid X,
  $M_X(\lambda) = (1/3)e^{3\lambda/2} + (1/6)e^{-3\lambda} + (1/8)e^{-\lambda} + (1/8)e^{\lambda} + 1/4$
  - $E[X] = 0$, $\text{Var}(X) = 5/2$ $\sim \sigma^2$
  - Consider $Y = (X_1 + \cdots + X_n)/\sqrt{n}$

$$M_Y(\lambda) = \left((1/3)e^{\frac{3\lambda}{2\sqrt{n}}} + (1/6)e^{-\frac{3\lambda}{\sqrt{n}}} + (1/8)e^{-\frac{\lambda}{\sqrt{n}}} + (1/8)e^{\frac{\lambda}{\sqrt{n}}} + 1/4 \right)^n$$

MGF of Y for different n



$M_{\overline{X}}(\lambda) \to e^{\lambda^2\sigma^2/2}$ as n increases

$\updownarrow$

Normal$(0, \sigma^2)$

Central Limit Theorem (CLT)

Theorem (CLT)

Let $X_1, \ldots, X_n \sim$ *iid* X *with* $E[X] = 0$, $Var(X) = \sigma^2$. *Let* $Y = (X_1 + \cdots + X_n)/\sqrt{n}$. *Then,*

$$M_Y(\lambda) \to e^{\lambda^2\sigma^2/2}$$

- MGF of Normal$(0, \sigma^2)$: $e^{\lambda^2\sigma^2/2}$
- Y is said to converge in distribution to Normal$(0, \sigma^2)$

Central Limit Theorem (CLT)

Theorem (CLT)

Let $X_1, \ldots, X_n \sim$ iid X with $E[X] = 0$, $Var(X) = \sigma^2$. Let $Y = (X_1 + \cdots + X_n)/\sqrt{n}$. Then,

$$M_Y(\lambda) \to e^{\lambda^2\sigma^2/2}$$

- MGF of Normal$(0, \sigma^2)$: $e^{\lambda^2\sigma^2/2}$
- Y is said to converge in distribution to Normal$(0, \sigma^2)$

Observations

- Contrast with WLLN
  - $\overline{X} = (X_1 + \cdots + X_n)/n$ converges in distribution to $E[X]$.
- Sum of iid random variables tends to be normal
  - Do I like above statement? Not entirely.
  - Scaling is important: $1/n$ constant, $1/\sqrt{n}$ normal

Using CLT to approximate probability

$$X_1, \ldots, X_n \overset{iid}{\sim} X$$

- Let $\mu = E[X]$, $\sigma^2 =$Var(X)
- $Y = X_1 + \cdots + X_n$, $E[Y] = n\mu$
- What is $P(Y - n\mu > \delta n\mu)$?

Using CLT to approximate probability

$$X_1, \ldots, X_n \overset{iid}{\sim} X$$

- Let $\mu = E[X]$, $\sigma^2 =$Var(X)
- $Y = X_1 + \cdots + X_n$
- What is $P(Y - n\mu > \delta n\mu)$?

Approximating using CLT

- $(Y - n\mu)/\sqrt{n}$: approximately Normal$(0, \sigma^2)$

$$\frac{Y - n\mu}{\sqrt{n}\sigma} \approx \text{Normal}(0, 1)$$

Using CLT to approximate probability

$$X_1, \ldots, X_n \sim X$$

- Let $\mu = E[X]$, $\sigma^2 =$Var(X)
- $Y = X_1 + \cdots + X_n$
- What is $P(Y - n\mu > \delta n\mu)$?

Approximating using CLT

- $(Y - n\mu)/\sqrt{n}$: approximately Normal$(0, \sigma^2)$

$$\frac{Y - n\mu}{\sqrt{n}\sigma} \approx \text{Normal}(0, 1)$$

- $F(z)$: CDF of Normal(0,1) (known)
- $P(Y - n\mu > \delta n\mu) = P(\underbrace{\tfrac{Y-n\mu}{\sqrt{n}\sigma}}_{\approx \text{Normal}(0,1)} > \tfrac{\delta\sqrt{n}\mu}{\sigma}) \approx 1 - F(\tfrac{\delta\sqrt{n}\mu}{\sigma})$

Approximating Binomial$(n, 1/2)$

$X_1, \dots, X_n \sim$ iid Bernoulli$(1/2)$

F: CDF of Normal$(0,1)$

- $\mu = \sigma = 1/2$
- $P(Y - n/2 > 0.6n/2) \approx 1 - F(0.6\sqrt{n})$

n	Event, $\delta = 0.6$	Prob	$1 - F(\sqrt{n}\delta)$
10	$Y - 5 > 5 \times 0.6$	0.0107	0.0289
50	$Y - 25 > 25 \times 0.6$	2.81×10^{-6}	1.10×10^{-5}
100	$Y - 50 > 50 \times 0.6$	1.35×10^{-10}	9.87×10^{-10}

- Approximation is really close!
- Normal approximation is quite good for Binomial

Examples

$$X_1, \ldots, X_n \sim \text{iid } X, Y = X_1 + \cdots + X_n$$

- $X \sim \{\overset{1/6}{-3}, \overset{1/8}{-1}, \overset{1/4}{0}, \overset{1/8}{1}, \overset{1/3}{3/2}\}$
  - $\mu = 0$, $\sigma^2 = 5/2$
  - CLT: $\frac{Y}{\sqrt{5n/2}} \approx \text{Normal}(0, 1)$
  - $P(Y > \delta n) = P(\frac{Y}{\sqrt{5n/2}} > \delta\sqrt{2n/5}) \approx 1 - F(\delta\sqrt{2n/5})$
    - $n = 10$, $\delta = 1$: ≈ 0.0228
    - $n = 100$, $\delta = 1$: $\approx 1.27 \times 10^{-10}$

Examples

$$X_1, \ldots, X_n \sim \text{iid } X,\, Y = X_1 + \cdots + X_n$$

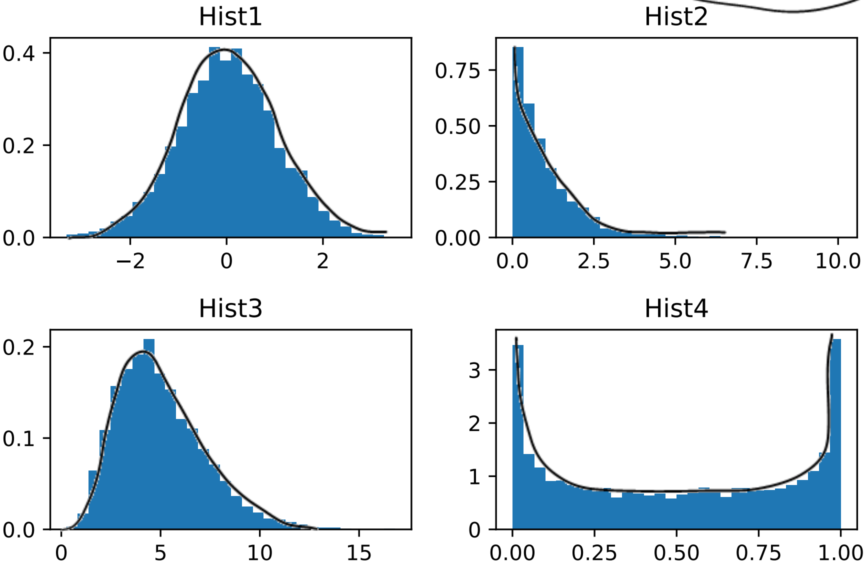
- $X \sim \{\overset{1/6}{-3}, \overset{1/8}{-1}, \overset{1/4}{0}, \overset{1/8}{1}, \overset{1/3}{3/2}\}$
  - $\mu = 0$, $\sigma^2 = 5/2$
  - CLT: $\frac{Y}{\sqrt{5n/2}} \approx \text{Normal}(0, 1)$
  - $P(Y > \delta n) = P(\frac{Y}{\sqrt{5n/2}} > \delta\sqrt{2n/5}) \approx 1 - F(\delta\sqrt{2n/5})$
    - ★ $n = 10$, $\delta = 1$: ≈ 0.0228
    - ★ $n = 100$, $\delta = 1$: $\approx 1.27 \times 10^{-10}$
- $X \sim \text{Uniform}[-1, 1]$ (continuous)
  - $\mu = 0$, $\sigma^2 = 1/3$
  - CLT: $\sqrt{3}Y \approx \text{Normal}(0, 1)$
  - $P(Y > 0.1\sqrt{n}) = P(\sqrt{3}Y > 0.1\sqrt{3n}) \approx 1 - F(0.1\sqrt{3n})$
    - ★ $n = 10$: ≈ 0.2919
    - ★ $n = 100$: ≈ 0.0416

Subsection 8

Distributions, properties and connections

Shapes of histograms: What distribution?

shape, location, rate, scale



Linear combination of iid normals

$$X_1, \ldots, X_n \sim \cancel{\text{iid}}\ \text{indep Normal}$$

- Let $X_i \sim \text{Normal}(\mu_i, \sigma_i^2)$
- Suppose $Y = a_1X_1 + \cdots + a_nX_n$
  - Linear combination of ~~iid~~ indep normals
- Then,

$$Y \sim \text{Normal}(\mu, \sigma^2),$$

$$e^{\frac{\lambda^2(a_1^2\sigma_1^2+\cdots+a_n^2\sigma_n^2)}{2}}$$

where $\mu = E[Y] = a_1\mu_1 + \cdots + a_n\mu_n$, $\sigma^2 = a_1^2\sigma_1^2 + \cdots + a_n^2\sigma_n^2$.

$$e^{\frac{\lambda^2\sigma_1^2a_1^2}{2}} \cdots e^{\frac{\lambda^2a_n^2\sigma_n^2}{2}}$$

- Linear combinations of ~~iid~~ indep normals is normal
  - Proof: Use moment generating functions

$\mu_1 = \cdots = \mu_n = 0$

$$E[e^{\lambda Y}] = E[e^{\lambda(a_1X_1+\cdots+a_nX_n)}] = E[e^{\lambda a_1X_1}]\ldots E[e^{\lambda a_nX_n}]$$

Gamma distribution

$$f_X(x) = \frac{x^{\alpha-1}e^{-\beta x}}{\int_0^\infty x^{\alpha-1}e^{-\beta x}\,dx}$$

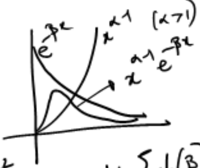
how the distribution looks?

overall scaling

$X \sim \text{Gamma}(\alpha, \cancel{beta}\ \beta)$ if PDF $f_X(x) \propto \underline{x^{\alpha-1}e^{-\beta x}}, \underline{x > 0}$

- $\alpha > 0$: shape parameter, $\beta > 0$: rate parameter, $\theta = 1/\beta$: scale parameter
- Mean: α/β, Variance: α/β^2, MGF: $(1 - \lambda/\beta)^{-\alpha}$, $\lambda < \beta$
- Sum of n iid Exp(β) is Gamma(n, β)
  - Proof: Use MGF (mostly)
- Square of Normal$(0, \sigma^2)$ is Gamma$(1/2, 1/2\sigma^2)$
  - Proof: Use CDF method

pdf $\propto x^{-1/2} e^{-x/2\sigma^2}$



$\underline{\alpha = 1}$: Exp(β)

Cauchy distribution

$$X \sim \text{Cauchy}(\theta, \alpha^2) \text{ if PDF } f_X(x) = \frac{1}{\pi}\frac{\alpha}{\alpha^2 + (x-\theta)^2}$$

- θ: location parameter, $\alpha > 0$: scale parameter
- Mean: undefined, Variance: undefined, MGF: undefined
- Suppose $X, Y \sim$ iid Normal$(0, \sigma^2)$. Then,

$$\frac{X}{Y} \sim \text{Cauchy}(0, 1)$$

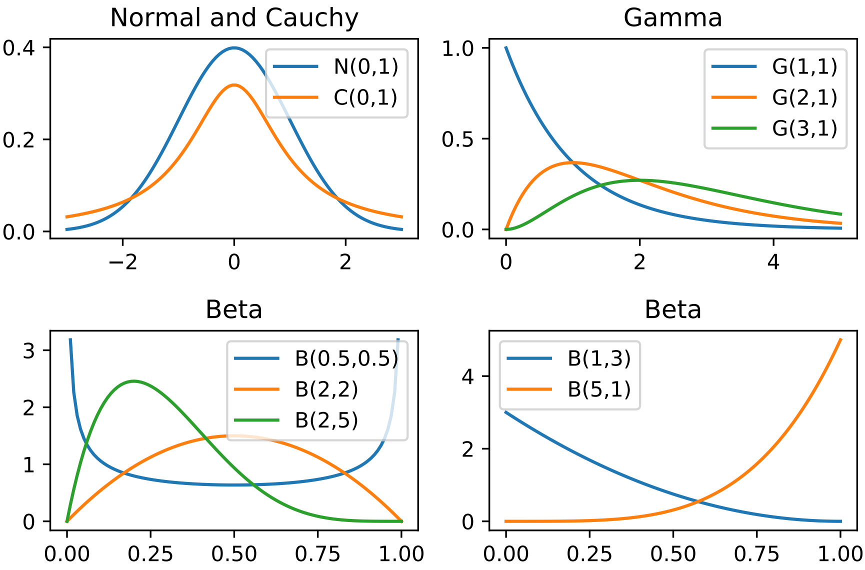
Beta distribution

$$X \sim \text{Beta}(\alpha, \beta) \text{ if PDF } f_X(x) \propto x^{\alpha-1}(1-x)^{\beta-1}, \underline{\underline{0 < x < 1}}$$

- $\alpha > 0, \beta > 0$: shape parameters
- Mean: $\alpha/(\alpha+\beta)$, Variance: $\alpha\beta/((\alpha+\beta)^2(\alpha+\beta+1))$
- $\text{Beta}(\alpha, 1)$ has PDF $\propto x^{\alpha-1}$: power function distribution
- Suppose $X \sim \text{Gamma}(\alpha, 1/\theta)$, $Y \underset{\text{indep}}{\sim} \text{Gamma}(\beta, 1/\theta)$, then

$$\frac{X}{X+Y} \sim \text{Beta}(\alpha, \beta)$$

Plots of PDFs



Subsection 9

Descriptive statistics of normal samples

Normal samples

$$X_1, \ldots, X_n \sim \text{iid Normal}(\mu, \sigma^2)$$

- Very common assumption in many situations
  - CLT is used as justification, often

Normal samples

$$X_1, \ldots, X_n \sim \text{iid Normal}(\mu, \sigma^2)$$

- Very common assumption in many situations
  - CLT is used as justification, often
- Sample mean

$$\overline{X} = \frac{X_1 + \cdots + X_n}{n}$$

- Sample variance

$$S^2 = \frac{(X_1 - \overline{X})^2 + \cdots + (X_n - \overline{X})^2}{n-1}$$

- Recall: Sample mean and sample variance are random variables
- For normal samples, the distribution of the sample mean and variance can be characterised in more detail

Distribution of Sample Mean

$$X_1, \ldots, X_n \sim \text{iid Normal}(\mu, \sigma^2)$$

- $\overline{X} = \frac{1}{n}X_1 + \cdots + \frac{1}{n}X_n$
- Sample mean is a linear combination of iid normal random variables
    - So, Sample mean has a normal distribution

Distribution of Sample Mean

$$X_1, \ldots, X_n \sim \text{iid Normal}(\mu, \sigma^2)$$

- $\overline{X} = \frac{1}{n}X_1 + \cdots + \frac{1}{n}X_n$
- Sample mean is a linear combination of iid normal random variables
  - So, Sample mean has a normal distribution

$$\overline{X} \sim \text{Normal}(\mu, \sigma^2/n)$$

- $E[\overline{X}] = \mu$
- $\text{Var}(\overline{X}) = \frac{1}{n^2}\sigma^2 + \cdots + \frac{1}{n^2}\sigma^2 = \sigma^2/n$

Sum of squares of normal samples: Chi-square

$$X_1, \ldots, X_n \sim \text{iid Normal}(0, \sigma^2)$$

- X_i^2: Gamma$(1/2, 1/2\sigma^2)$, independent
- **Result**: Sum of n independent Gamma(α, β) is Gamma$(n\alpha, \beta)$

Sum of squares of normal samples: Chi-square

$$X_1, \ldots, X_n \sim \text{iid Normal}(0, \sigma^2)$$

- X_i^2: Gamma$(1/2, 1/2\sigma^2)$, independent
- **Result**: Sum of n independent Gamma(α, β) is Gamma$(n\alpha, \beta)$

$$X_1^2 + \cdots + X_n^2 \sim \text{Gamma}(\frac{n}{2}, \frac{1}{2\sigma^2})$$

- Gamma$(n/2, 1/2)$: called Chi-square distribution with n degrees of freedom, denoted χ_n^2

Sample mean and variance of normal samples

Theorem

Suppose $X_1, \ldots, X_n \overset{iid}{\sim} \textit{Normal}(\mu, \sigma^2)$. *Then,*

1. $\overline{X} \sim \textit{Normal}(\mu, \sigma^2/n)$
2. $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$, *Chi-square with* $n-1$ *degrees of freedom.*
3. $\overline{X}$ *and* S^2 *are independent.*

- For normal samples, the joint distribution of sample mean and variance is precisely known.

$$S^2 = \frac{(X_1-\overline{X})^2 + \cdots + (X_n-\overline{X})^2}{n-1}$$

$$\sum_{i=1}^{n}\left(\frac{X_i-\overline{X}}{n}\right) = \overline{X} - \overline{X} = 0$$

$$\frac{(n-1)S^2}{\sigma^2} = \left(\frac{X_1-\overline{X}}{\sigma}\right)^2 + \cdots + \left(\frac{X_n-\overline{X}}{\sigma}\right)^2 \sim \chi^2_{n-1}$$