# Review of Last Lecture

Key concepts and/or techniques:

1. Distribution of functions of $X_1, \cdots, X_n$, which are independent and normally distributed

- mgf technique
- first cdf and then pdf

# Review of Last Lecture

[Theorem 5.5-1]

If $X_1, X_2, \cdots, X_n$ are $n$ independent normal variables with means $\mu_1, \mu_2, \cdots, \mu_n$ and variances $\sigma_1^2$, $\sigma_2^2$, $\cdots$, $\sigma_n^2$ , respectively, then $Y = \sum_{i=1}^{n} a_i X_i$ has the normal distribution

$$Y \sim N\left(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right)$$

# Review of Last Lecture

[Corollary 5.5-1]

If $X_1, X_2, \cdots, X_n$ is a random sample of size $n$ from the normal distribution $N(\mu, \sigma^2)$, then the sample mean $\overline{X}$ has the following distribution

$$\overline{X} \sim N(\mu, \frac{\sigma^2}{n}) \Leftrightarrow \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

Definition

Let $X_1, X_2, \cdots, X_n$ be independent and identically distributed with mean $\mu$ and $\sigma^2$. Then the sample variance is defined as

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2, \qquad E(S^2) = \sigma^2.$$

# Review of Last Lecture

[Theorem 5.5-2]

Let $X_1, X_2, \cdots, X_n$ be random sample of size $n$ from the normal distribution $N(\mu, \sigma^2)$. Then the sample mean $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and the sample variance $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ are independent, and

$$\frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^{n}\left(\frac{X_i - \overline{X}}{\sigma}\right)^2 \sim \chi^2(n-1)$$

# Review of Last Lecture

**[Student's t distribution]**

Let

$$T = \frac{Z}{\sqrt{U/r}}$$

where $Z \sim N(0,1), U \sim \chi^2(r)$, and $Z$ and $U$ are independent. Then $T$ has a student's $t$ distribution, i.e., $T \sim t(r)$, where $r$ is called the degrees of freedom. Let

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}, \quad U = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{(n-1)S^2}{\sigma^2}$$

$$T = \frac{Z}{\sqrt{U/(n-1)}} = \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$$

## Review of Last Lecture

Let $X_1, \cdots, X_n$ be a random sample of size $n$ from a normal distribution $N(\mu, \sigma^2)$. Then we have

- $$\sum_{i=1}^{n} \left( \frac{X_i - \mu}{\sigma} \right)^2 \sim \chi^2(n), \quad \sum_{i=1}^{n} \left( \frac{X_i - \overline{X}}{\sigma} \right)^2 \sim \chi^2(n-1)$$
- $$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1), \quad \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$$

# STA2001 Probability and Statistics I

## Lecture 23

Tianshi Chen

The Chinese University of Hong Kong, Shenzhen

# Section 5.6 The Central Limit Theorem

# Motivation

Let $\overline{X}$ be the sample mean of a random sample $X_1, X_2, \cdots, X_n$

of size $n$ from $N(\mu, \sigma^2)$. Then for any $n$,

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1) \Longleftrightarrow \overline{X} \sim N(\mu, \frac{\sigma^2}{n}) \Longleftrightarrow \sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2)$$

# Motivation

Let $\overline{X}$ be the sample mean of a random sample $X_1, X_2, \cdots, X_n$ of size $n$ from $N(\mu, \sigma^2)$. Then for any $n$,

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1) \Longleftrightarrow \overline{X} \sim N(\mu, \frac{\sigma^2}{n}) \Longleftrightarrow \sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2)$$

The result can be extended to more general random distributions: as $n \to \infty$, the sequence $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ converges to $N(0,1)$ in some sense, which concerns the topic of convergence of sequence of random variables!

# Convergence of Sequence of Numbers

**Definition**

A sequence of numbers $a_1$, $a_2$, ... is said to converge to a limit $a$ if

$$\lim_{n \to \infty} a_n = a.$$

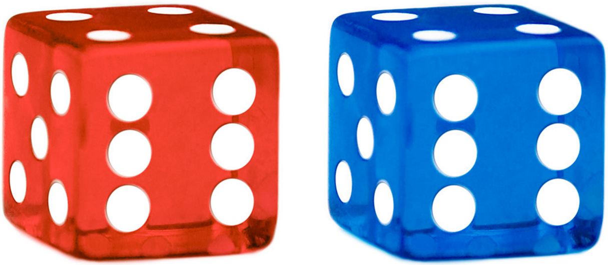
That is, for any $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that

$$|a_n - a| < \epsilon, \qquad \text{for all } n > N.$$

How to define convergence of sequence of random variables?

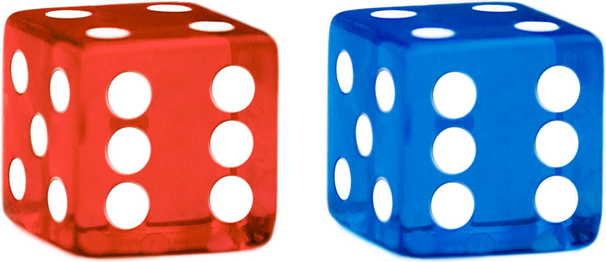
# Convergence of Sequence of Random Variables

Key: How to measure the closeness between two random variables?



# Convergence of Sequence of Random Variables

Key: How to measure the closeness between two random variables?



- probability
- mathematical expectation

# Convergence in Distribution

**Definition**

A sequence of random variables $Z_1$, $Z_2$, ... is said to converge in distribution, or converge weakly, or converge in law to a random variable $Z$, denoted by $Z_n \xrightarrow{d} Z$, if

$$\lim_{n\to\infty} F_n(z) = F(z),$$

for every number $z \in R$ at which $F(z)$ is continuous, where $F_n(z)$ and $F(z)$ are the cdfs of random variables $Z_n$ and $Z$, respectively.

# Remark

For a given $z$ at which $F(z)$ is continuous, let

$$a_n = F_n(z) = P(Z_n \leq z)$$
$$a = F(z) = P(Z \leq z)$$

The convergence in distribution of sequence of random variables

$$\lim_{n\to\infty} F_n(z) = F(z),$$

can be interpreted as the convergence of sequence of numbers

$$\lim_{n\to\infty} a_n = a,$$

that is, for any $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that

$$|P(Z_n \leq z) - P(Z \leq z)| < \epsilon, \qquad \text{for all } n > N.$$

# Example 1

Let $Z_2, Z_3 \cdots$ be a sequence of random variables such that

$$F_{Z_n}(z) = \begin{cases} 1 - \left(1 - \frac{1}{n}\right)^{nz}, & z > 0 \\ \\ 0, & z \leq 0 \end{cases}$$

Then prove that $Z_n$ converges in distribution to exponential distribution with $\theta = 1$, whose cdf $F(z) = 0$ for $z \leq 0$ and $F(z) = 1 - e^{-z}$ for $z > 0$.

# Example 1

For $z \leq 0$, $F_{Z_n}(z) = F(z)$, for $n = 2, \cdots$.

For $z > 0$, we have

$$\lim_{n\to\infty} F_{Z_n}(z) = 1 - \lim_{n\to\infty} \left(1 - \frac{1}{n}\right)^{nz} = 1 - e^{-z} = F(z)$$

# Central Limit Theorem (CLT), page 208

**CLT**

Let $\overline{X}$ be the sample mean of the random sample of size $n$, $X_1$,$X_2$ ,$\cdots$,$X_n$ from a distribution with a finite mean $\mu$ and a finite nonzero variance $\sigma^2$, then as $n \to \infty$, the random variable $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ converges in distribution to $N(0, 1)$.

# Central Limit Theorem (CLT), page 208

**CLT**

Let $\overline{X}$ be the sample mean of the random sample of size $n$, $X_1, X_2, \cdots, X_n$ from a distribution with a finite mean $\mu$ and a finite nonzero variance $\sigma^2$, then as $n \to \infty$, the random variable $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ converges in distribution to $N(0,1)$.

Practical use of CLT: for large $n$,

- $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ can be approximated by $N(0,1)$.
- $\overline{X}$ can be approximated by $N(\mu, \frac{\sigma^2}{n})$.
- $\sum_{i=1}^{n} X_i$ can be approximated by $N(n\mu, n\sigma^2)$.

# Practical Use of CLT

For large $n$, the probabilities of events of $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$, $\overline{X}$ and $\sum_{i=1}^{n} X_i$ can be calculated approximately by treating them as if they are $N(0,1)$, $N(\mu, \frac{\sigma^2}{n})$, and $N(n\mu, n\sigma^2)$, respectively, and by looking up tables of normal distributions.

# Practical Use of CLT

For large $n$, the probabilities of events of $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$, $\overline{X}$ and $\sum_{i=1}^{n} X_i$ can be calculated approximately by treating them as if they are $N(0,1)$, $N(\mu, \frac{\sigma^2}{n})$, and $N(n\mu, n\sigma^2)$, respectively, and by looking up tables of normal distributions.

Recall that if $Y \sim N(\mu, \sigma^2)$

$$
\begin{aligned}
P(a \leq Y \leq b) &= P(\frac{a-\mu}{\sigma} \leq \frac{Y-\mu}{\sigma} \leq \frac{b-\mu}{\sigma}) \\
&= \Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})
\end{aligned}
$$

where $\Phi(\cdot)$ is the cdf of $N(0,1)$

# Example 2, page 209

**Question**

Let $X_1, \cdots, X_{25}$ be a random sample of size $n = 25$ from a distribution with mean 15 and variance 4.

Q1: Compute $P(14.4 < \overline{X} < 15.6)$ approximately ?

# Example 2, page 209

Q1: By CLT, $\overline{X}$ approximately have $N(\mu, \frac{\sigma^2}{n}) = N(15, \frac{4}{25} = 0.4^2)$

$$P(14.4 < \overline{X} < 15.6) = P(\frac{14.4 - 15}{0.4} < \frac{\overline{X} - 15}{0.4} < \frac{15.6 - 15}{0.4})$$

$$= \Phi(1.5) - \Phi(-1.5) = 0.9332 - (1 - 0.9332)$$

$$= 0.8664$$

# Example 3, page 210

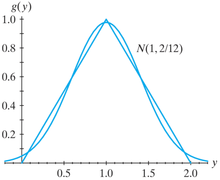
Let $X_1, \cdots, X_n$ be a random sample of size $n$ from the uniform

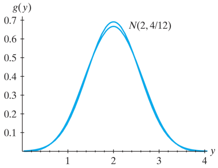
distribution $U(0, 1)$.

Recall its pdf, mean and variance are as follows:

$$f(x) = 1, \quad x \in [0, 1]. \quad E(X) = \mu = \frac{1}{2}, \quad Var(X) = \sigma^2 = \frac{1}{12}.$$

# Example 3, page 210




Consider $Y = \sum_{i=1}^{n} X_i$. Our goal is to check the difference between the pdf of $Y$ and the pdf of its approximation $N(n\mu, n\sigma^2)$ from CLT.

- check $n = 2$, pdf of $Y$,
  $$g(y) = \begin{cases} y, & y \in [0,1] \\ 2 - y, & y \in [1,2] \end{cases}$$
  pdf of $N(2 \cdot \frac{1}{2}, 2 \cdot \frac{1}{12}) = N(1, \frac{1}{6})$
- check $n = 4$.

# Example 3, page 210

We sketch the derivation of the pdf of $Y$ for $n = 2$.

Clearly, the joint pdf of $(X_1, X_2)$ is

$$f(x_1, x_2) = 1, \quad 0 < x_1 < 1,\ 0 < x_2 < 1.$$

1. cdf of $Y$, $G(y) = P(Y \leq y) = P(X_1 + X_2 \leq y)$
2. pdf of $Y$, $g(y) = G'(y)$ at which $G(y)$ is differentiable

# Example 3, page 210

1. cdf of $Y$, $G(y) = P(Y \leq y) = P(X_1 + X_2 \leq y)$
   - $y \in (0,1)$, $G(y) = \int_0^y \int_0^{y-x_1} 1 dx_2 dx_1 = \frac{1}{2}y^2$
   - $y \in (1,2)$,
     $G(y) = \int_0^{y-1} \int_0^1 1 dx_2 dx_1 + \int_{y-1}^1 \int_0^{y-x_1} 1 dx_2 dx_1 = -1 + 2y - \frac{1}{2}y^2$
2. pdf of $Y$, $g(y) = G'(y)$ at which $G(y)$ is differentiable
   - $y \in (0,1)$, $g(y) = y$
   - $y \in (1,2)$, $g(y) = 2 - y$

# Section 5.7 Approximations for Discrete Distributions

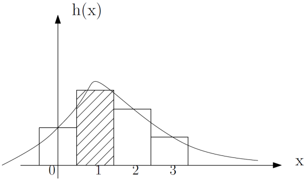
# Motivation

By CLT, we will use normal distributions to approximate the discrete distribution of $\overline{X}$ or $\sum_{i=1}^{n} X_i$, where $X_1, \cdots, X_n$ is a random sample of size $n$ from discrete distributions, in the sense that the pdf of the normal distribution is close to the histogram of the discrete distribution of $\overline{X}$ or $\sum_{i=1}^{n} X_i$.

# Histogram for Discrete Distribution

Consider a discrete RV $Y$ with pmf $f(y) : \overline{S} \to (0,1]$ with $\overline{S} = \{0, 1, \cdots, n\}$.Then the histogram for $Y$ is

$$h(y) = f(k), y \in (k - \frac{1}{2}, k + \frac{1}{2}), k = 0, 1, \cdots, n$$




For $k = 0, 1, \cdots, n$, $P(Y = k) = f(k)$ corresponds to the area of the rectangle with a height of $P(Y = k)$ and a base of length 1 centered at $k$.

# Approximate Discrete Distribution by Continuous Distribution

**Key idea:** The area below the histogram corresponds to probability, which make the histogram has similar property as the pdf of continuous distribution.

# Approximate Discrete Distribution by Continuous Distribution

**Key idea:** The area below the histogram corresponds to probability, which make the histogram has similar property as the pdf of continuous distribution.

**Key usage:** If it is possible to find a continuous distribution with pdf "close" to the histogram of the discrete distribution, then we can compute the probability of discrete distribution approximately by using the continuous distribution.

However, there is a catch, which is called the half-unit correction!