STA2001 Probability and Statistics (I)

Lecture 16

Tianshi Chen

The Chinese University of Hong Kong, Shenzhen

Review of Last Lecture

Key concepts and/or techniques:

1. Sample mean: Let $X_1, X_2, \cdots, X_n$ be independent and identically distributed with mean μ. Then the sample mean is defined as

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

$\overline{X} = \sum_{i=1}^{n} X_i \cdot \frac{1}{n}$

and a statistic and also an estimator of mean μ.

2. Mgf technique: Mgf, if exists, uniquely determines the distribution of the RV. Therefore, the distribution of a RV can be equivalently found via its mgf.

Use the mgf technique to derive the distribution of

$$Y = \sum_{i=1}^{n} a_i X_i$$

$Y = \sum_{i=1}^{n} a_i X_i$

Review of Last Lecture

[Theorem 5.4-1]

If $X_1, X_2, \cdots, X_n$ are independent RVs with respective mgfs $M_{X_i}(t)$ where $|t| < h_i$ for positive number $h_i, i = 1, 2, \cdots, n$. Then the mgf of $Y = \sum_{i=1}^{n} a_i X_i$ is

$$M_Y(t) = \prod_{i=1}^{n} M_{X_i}(a_i t),$$

where $|a_i t| < h_i, i = 1, \cdots, n$. $\prod_{i=1}^{n} M_{X_i}(a_i t)$

Review of Last Lecture

[Theorem 5.4-2]

Let $X_1, X_2, \cdots, X_n$ be independent chi-square RVs with $r_1, r_2, \cdots, r_n$ degrees of freedom, respectively, i.e., $X_i \sim \chi^2(r_i), i = 1, \cdots, n$ Then

不同r　所有r相加

$$Y = X_1 + X_2 + \cdots + X_n \quad \text{is} \quad \chi^2(r_1 + r_2 + \cdots + r_n)$$

[Corollary 5.4-3]

If $X_1, X_2, \cdots, X_n$ are independent and have normal distributions $N(\mu_i, \sigma_i^2), i = 1, 2, ..., n$, respectively, then the distribution of

$$\sum_{i=1}^{n} \left(\frac{X_i - \mu_i}{\sigma_i} \right)^2 \sim \chi^2(n)$$

$\sum_{i=1}^{n} \left(\frac{X_i - u_i}{\sigma_i} \right)^2 \sim \chi^2(n)$

不同 $N(u_i, \sigma_i^2)$

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

$Y = \sum_{i=1}^{n} a_i X_i$

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

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

都是 $N(u, \sigma^2)$

e.g $P(|\overline{X}-u| \le \Delta) \geqslant 0.95$ confidence interval

$E[\overline{X}] = u$

$a_i = \frac{1}{n}$

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

$P\left(\frac{-\Delta}{\sigma/\sqrt{n}} \le \frac{\overline{X}-u}{\sigma/\sqrt{n}} \le \frac{\Delta}{\sigma/\sqrt{n}}\right)$

$\Rightarrow 2\Phi\left(\frac{\Delta}{\sigma/\sqrt{n}}\right) \geqslant 1.95$

Definition

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

i.i.d.

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

Theorem 5.5-2, page 202

[Theorem 5.5-2]

Let $X_1, X_2, \cdots, X_n$ be random sample of size n from the normal distribution $N(\mu, \sigma^2)$ with $\sigma^2 > 0$. 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

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

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

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

The independence of $\overline{X}$ and S^2 is not proved here but deferred to Section 6.7 on page 294, and we only prove the second part.

Proof of Theorem 5.5-2, page 202

Following the proof of $E(S^2) = \sigma^2$, we have

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

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

$\sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right)^2 \downarrow W$ $\quad Z \downarrow \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$

Now let

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

Then

$$W = \frac{(n-1)S^2}{\sigma^2} + Z^2$$

$W = \frac{(n-1)}{\sigma^2} S^2 + Z^2$

Further note that $W \sim \chi^2(n)$, $Z^2 \sim \chi^2(1)$, and moreover, S^2 and $\overline{X}$ are independent by assumption.

$W \sim \chi^2(n)$ $\quad Z^2 \sim \chi^2(1)$

Proof of Theorem 5.5-2, page 202

Note that $W \sim \chi^2(n), Z^2 \sim \chi^2(1)$, S^2 and $\overline{X}$ are independent

$$E[e^{tw}] = E[e^{t(\frac{(n-1)S^2}{\sigma^2}+Z^2)}] = E[e^{t\frac{(n-1)S^2}{\sigma^2}}]E[e^{tZ^2}]$$

$$(1-2t)^{-\frac{n}{2}} = E[e^{t\frac{(n-1)S^2}{\sigma^2}}]\cdot(1-2t)^{-\frac{1}{2}}, \quad t < \frac{1}{2}$$

$$\Rightarrow E[e^{t\frac{(n-1)S^2}{\sigma^2}}] = (1-2t)^{-\frac{n-1}{2}}, \quad t < \tfrac{1}{2} \Rightarrow \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

A remark

Combining Corollary 5.4-3 and Thm 5.5-2 leads to the observation:

If $X_1, X_2, \cdots, X_n$ is a random sample of size n from $N(\mu, \sigma^2)$, then

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

When the mean μ is replaced by the sample mean $\overline{X}$, one degree of freedom is lost.

This is because there is an additional constraint

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Example 5.5-3, page 204

Let X_1, X_2, X_3, X_4 be a random sample of size 4 from the normal distribution $N(76.4, 383)$.

$$\sum_{i=1}^{4} \frac{(X_i - 76.4)^2}{383} \sim \chi^2(4), \quad \sum_{i=1}^{4} \frac{(X_i - \overline{X})^2}{383} \sim \chi^2(3)$$

Student's t Distribution

[Theorem 5.5-3]

Let

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

$f(t) = \frac{\Gamma(\frac{r+1}{2})}{\sqrt{\pi r}\,\Gamma(\frac{r}{2})} \cdot \frac{1}{(1+\frac{t^2}{r})^{\frac{r+1}{2}}}$

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

$$f(t) = \frac{\Gamma(\frac{r+1}{2})}{\sqrt{\pi r}\Gamma(\frac{r}{2})} \frac{1}{(1+\frac{t^2}{r})^{\frac{r+1}{2}}}, \quad t \in (-\infty, \infty),$$

where r is called the degrees of freedom, and we simply write $T \sim t(r)$.

$T \sim t(r)$

Sketch of the Proof of Theorem 5.5-3, page 204

$$\frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \cdot \frac{1}{\Gamma(\frac{1}{2})2^{\frac{1}{2}}} u^{\frac{1}{2}-1} e^{-\frac{u}{2}}$$

Since Z and U are independent, their joint pdf $g(z,u)$ is

$$g(z,u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \cdot \frac{1}{\Gamma(\frac{r}{2})2^{\frac{r}{2}}} u^{\frac{r}{2}-1} e^{-\frac{u}{2}}, \quad z \in R, u \in [0,\infty)$$

1. The cdf of T, $F(t)$ is

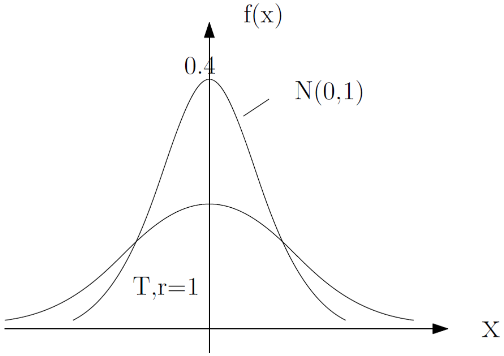
$$F(t) = P(T \leq t) = P\left(\frac{Z}{\sqrt{\frac{U}{r}}} \leq t\right) = P\left(Z \leq \sqrt{\frac{U}{r}}\, t\right)$$

2. The pdf of T,

$$f(t) = F'(t)$$

Student's t Distribution: Heavy-tailed Distribution

Student's t distribution is a heavy tailed distribution



Standard normal distribution :

$$f(x) = \frac{1}{\sqrt{2\pi}} \frac{1}{e^{\frac{1}{2}x^2}}, \quad x \in (-\infty, \infty)$$

Students' t distribution with $r = 1$:

$$f(x) = \frac{1}{\pi(1 + x^2)}, \quad x \in (-\infty, \infty)$$

Therefore, Student's t distribution is a better choice than the normal distribution when the data contains outliers.

A Student's t RV based on Random Samples from Normal Distribution

By using the result of Corollary 5.5-1 and Theorems 5.5-2 and 5.5-3, we can construct an important student's t random variable.

Assume that $X_1, X_2, \cdots, X_n$ is a random sample of size n from a normal distribution $N(\mu, \sigma^2)$.

A Student's t RV based on Random Samples from Normal Distribution

$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ $U = \frac{(n-1)S^2}{\sigma^2}$

Let

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

$Z \sim N(0,1)$ $U \sim \chi^2(n-1)$

Then $Z \sim N(0,1)$ and $U \sim \chi^2(n-1)$. Since Z and U are independent, then

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

A remark

If $X_1, \cdots, X_n$ is a random sample of size n from a normal distribution $N(\mu, \sigma^2)$, then

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1), \quad \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$$

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 $a_1, a_2, \ldots$ is said to converge to a limit a if

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

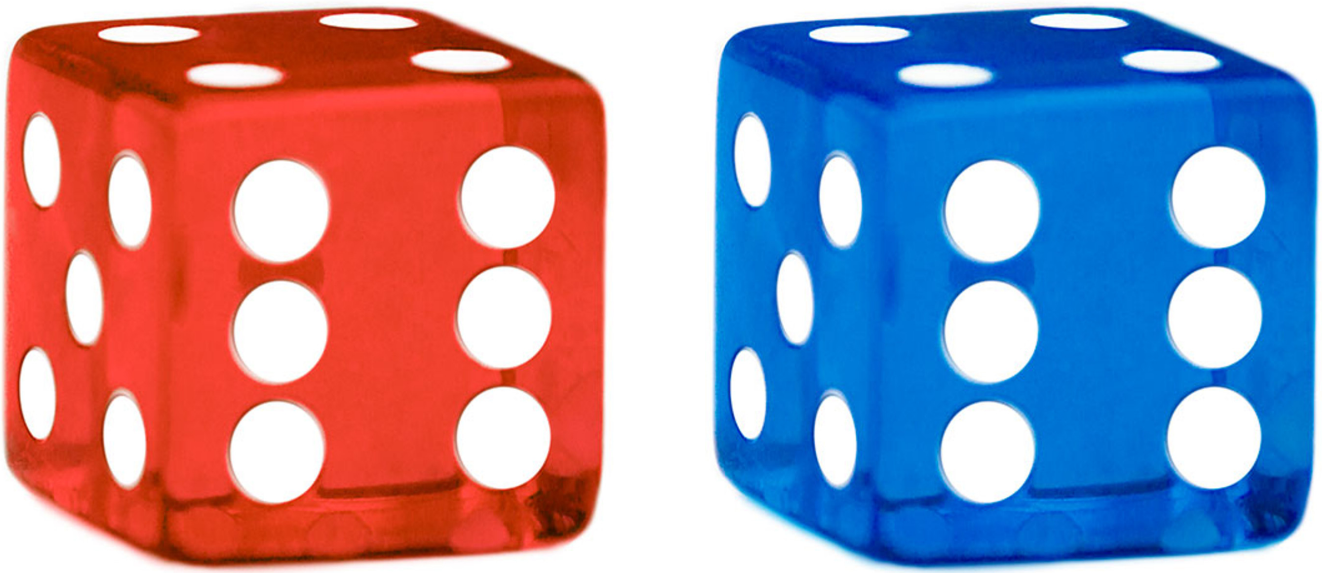
That is, for any $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that

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

How to define convergence of sequence of random variables?

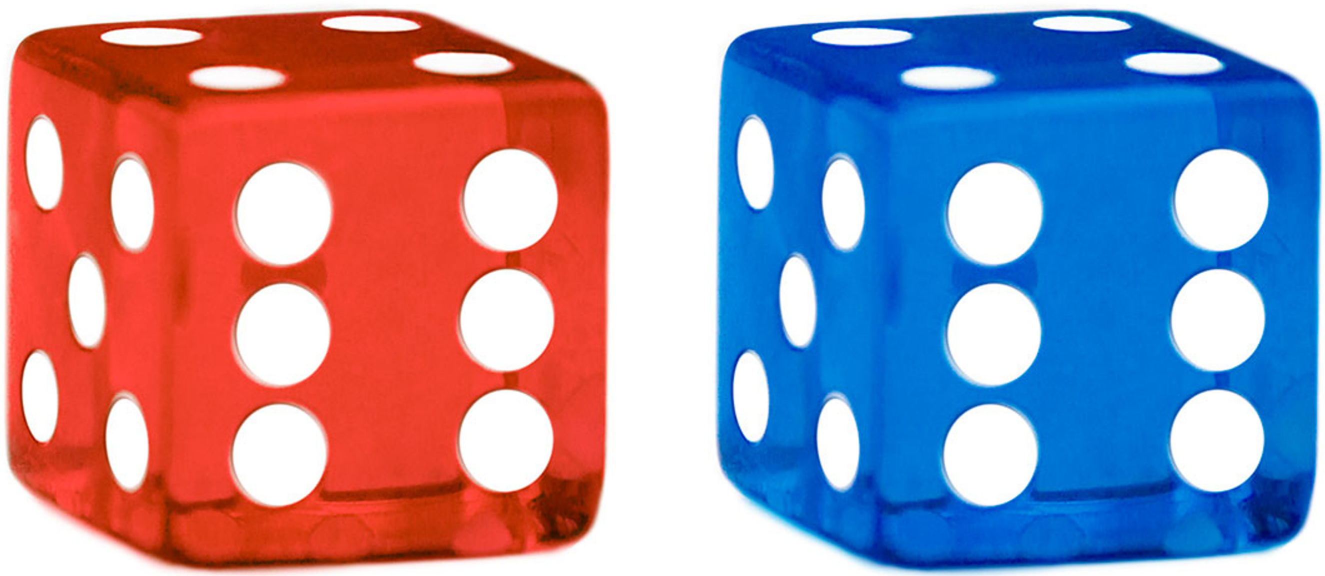
Convergence of Sequence of Random Variables

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

$$a_n = F_n(z) = P(Z_n \le z)$$

$$a = F(z) = P(Z \le z)$$

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

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

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 $z \le 0$, $F_{Z_n}(z) = F(z)$, 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 μ and a finite nonzero variance σ^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 μ and a finite nonzero variance σ^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, $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

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

$$P(a \le Y \le b) = P(\frac{a-\mu}{\sigma} \le \frac{Y-\mu}{\sigma} \le \frac{b-\mu}{\sigma})$$
$$= \Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})$$

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

$\overline{X} \sim N(\mu, \frac{\sigma^2}{n}) = N(15, (\frac{2}{5})^2)$

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 $U(0,1)$.

$X \sim U(0,1)$

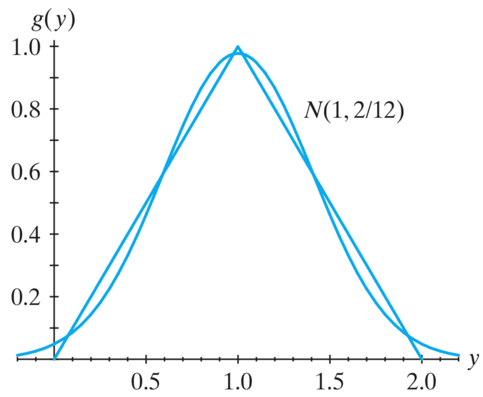
$\mu = \frac{a+b}{2}$ $\sigma^2 = \frac{(a+b)^2}{12}$

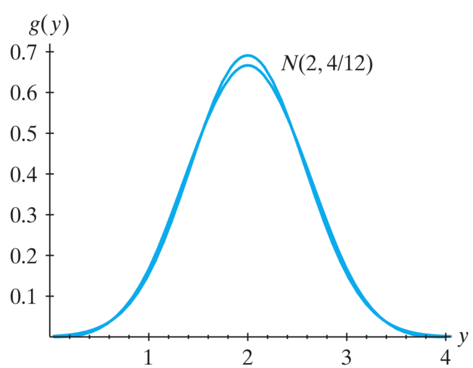
Recall its pdf, mean and variance are as follows:

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

Example 3, page 210

$Y = \sum_{i=1}^{n} X_i$





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

$X \sim U(0,1)$ $f(x_i)=1$

$Y = X_1 + X_2$

$\int_0^y dx_1 \int_0^{y-x_1} 1\, dx_2 = \frac{1}{2}y^2$

1. cdf of Y, $G(y) = P(Y \leq y) = P(X_1 + X_2 \leq y)$
   - $y \in (0,1)$, $G(y) = \int_0^y \int_0^{y-x_1} 1 dx_2 dx_1 = \frac{1}{2}y^2$
   - $y \in (1,2)$,

     $\int_0^{y-1} dx_1 \int_0^1 dx_2 + \int_{y-1}^1 dx_1 \int_0^{y-x_1} dx_2$

     $G(y) = \int_0^{y-1} \int_0^1 1 dx_2 dx_1 + \int_{y-1}^1 \int_0^{y-x_1} 1 dx_2 dx_1 = -1 + 2y - \frac{1}{2}y^2$

     积分区间!
2. pdf of Y, $g(y) = G'(y)$ at which $G(y)$ is differentiable $x \in (0,1)$
   - $y \in (0,1)$, $g(y) = y$
   - $y \in (1,2)$, $g(y) = 2 - y$

$y \in (0,1)$

$0 < x_1 < y$

$0 < x_2 < y - x_1$

① $y \in (1,2)$

$y-1 < x_1 < 1$ 为什么?

$0 < x_2 < y - x_1$

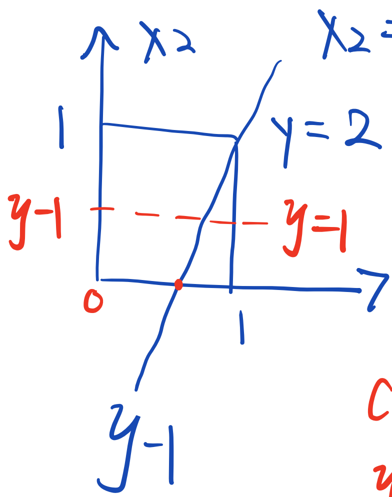
② $0 < x_1 < y-1$

$0 < x_2 < 1$

$x_1 + x_2 \le y$

$x + y \le z$ 将z看成 constant

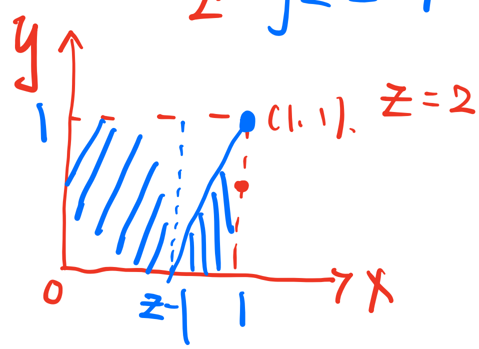
z $\quad y \le z - x$



Case ①

$0 \le x_1 \le y-1$

$0 \le x_2 \le 1$



Case ②

$y-1 \le x_1 < 1$

$0 \le x_2 \le y - x_1$

$f(x_1, x_2) = f(x_1) f(x_2) = 1$

$0 < x_1 < 1 \qquad x_1 + x_2 = Y$

$0 < x_2 < 1 \qquad S_Y \in (0, 2)$

$x_1 + x_2 \le Y$

Case ①. $0 < Y < 1$

$$G(Y) = P(x_1 + x_2 \le Y) = \int_0^Y dx_1 \int_0^{y - x_1} dx_2$$

Case ② $1 < y < 2$

$0 < x_1 < 1 \qquad x_1 + x_2 \le y$

$0 < x_2 < 1 \qquad x_2 \le y - x_1$

$$\int_0^{y-1} dx_1 \int_0^1 dx_2 + \int_{y-1}^1 dx_1 \int_0^{y - x_1} dx_2$$