STA2001 Probability and Statistics (I)

Lecture 17

Tianshi Chen

The Chinese University of Hong Kong, Shenzhen

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

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

Review of Last Lecture

Convergence in distribution

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 $Z_n \overset{d}{\to} Z$, 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.

Note: convergence of sequence of numbers.

Review of Last Lecture

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 $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ converge 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 $N(n\mu, n\sigma^2)$.

Review of Last Lecture

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

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

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

Section 5.7 Approximations for Discrete Distributions

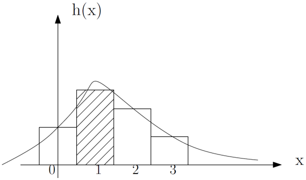
Motivation

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!

Half-unit correction for continuity

Now, let $Y = \sum_{i=1}^{n} X_i$, where $X_1, \cdots, X_n$ are i.i.d. random sample drawn from discrete distribution with mean μ and variance σ^2, then

$P(Y = k)$	$\approx$	$P(k - \frac{1}{2} < Y < k + \frac{1}{2})$
discrete RV		approximated by continuous RV
pmf $f(y)$		by CLT for large n, $Y = \sum_{i=1}^{n} X_i$ can be approximated by $N(n\mu, n\sigma^2)$ in the sense that the pdf of the normal distribution is close to the histogram of Y
hard to calculate		easy to calculate

Binomial distribution

Let $X_1, \cdots, X_n$ be a random sample of size n from Bernoulli distribution $b(1, p)$, whose mean is p and variance $p(1-p)$.

Binomial distribution

Let $X_1, \cdots, X_n$ be a random sample of size n from Bernoulli distribution $b(1, p)$, whose mean is p and variance $p(1-p)$.

Then

$$Y = \sum_{i=1}^{n} X_i \sim b(n, p)$$

with mean np and variance $np(1-p)$.

Binomial distribution

Let $X_1, \cdots, X_n$ be a random sample of size n from Bernoulli distribution $b(1, p)$, whose mean is p and variance $p(1-p)$.

Then

$$Y = \sum_{i=1}^{n} X_i \sim b(n, p)$$

with mean np and variance $np(1-p)$.

To calculate $P(Y = k)$ by definition, i.e.,
$P(Y = k) = \binom{n}{k} p^k (1-p)^{n-k}$ is complicated.

Binomial distribution

Now we try to calculate $P(Y = k)$ by CLT,

$$\frac{Y/n - p}{\sqrt{p(1-p)/n}} \xrightarrow{d} N(0,1)$$

For sufficiently large n , Y can be approximated by $N(np, np(1-p))$ and thus probability for $b(n,p)$ can be approximated by that for $N(np, np(1-p))$.

Binomial distribution

$$P(Y = k) \quad \approxeq \quad P(k - \tfrac{1}{2} < Y < k + \tfrac{1}{2})$$

$$\uparrow \qquad\qquad\qquad \uparrow$$

$$f(k) = \tfrac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \qquad Y \sim N(np, np(1-p)) \text{ for large n}$$

$$\begin{aligned} P(k - \frac{1}{2} < Y < k + \frac{1}{2}) &= P\left(\frac{k - \frac{1}{2} - np}{\sqrt{np(1-p)}} < \frac{Y - np}{\sqrt{np(1-p)}} < \frac{k + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right) \\ &= \Phi\left(\frac{k + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k - \frac{1}{2} - np}{\sqrt{np(1-p)}}\right), \end{aligned}$$

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

Example 1, page 216

Question

Assume $Y \sim b(10, 0.5)$. $\quad Q : P(3 \leq Y < 6)$?

Example 1, page 216

Question

Assume $Y \sim b(10, 0.5)$. $\quad Q : P(3 \le Y < 6)$?

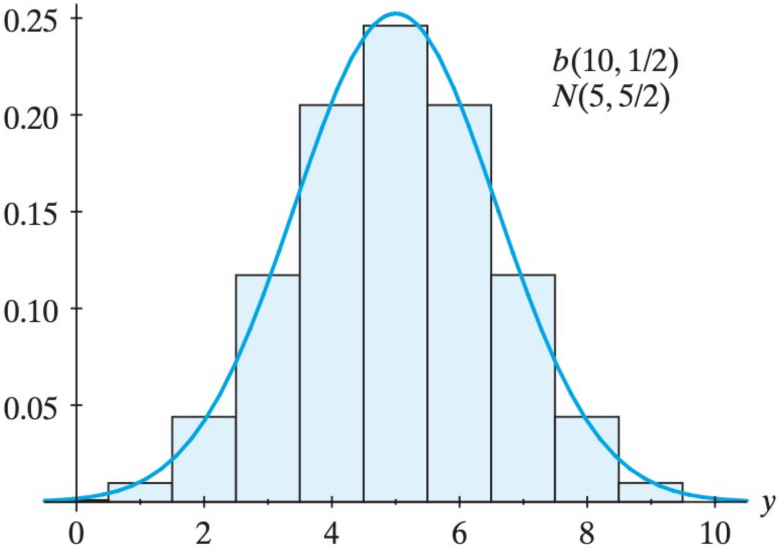
1. By definition,

$$P(3 \le Y < 6) = \sum_{k=3}^{5} P(Y = k) = \sum_{k=3}^{5} f(k) = 0.5683$$

2. By CLT, $Y = \sum_{i=1}^{10} X_i, \quad X_1, \cdots, X_{10}$ are i.i.d. from $b(1, \frac{1}{2})$

$$Y \quad \text{approximately} \quad N(np, np(1-p)) = N(5, 2.5)$$

Example 1, page 216



Example 1, page 216

$$P(3 \leq Y < 6) = \sum_{k=3}^{5} P(Y = k) \approxeq \sum_{k=3}^{5} P(k - \frac{1}{2} < Y < k + \frac{1}{2})$$

$$= P(2.5 < Y < 5.5) = P(\frac{2.5 - 5}{\sqrt{2.5}} < \frac{Y - 5}{\sqrt{2.5}} < \frac{5.5 - 5}{\sqrt{2.5}})$$

$$= \Phi(0.316) - \Phi(-1.581)$$

$$\approxeq 0.6240 - 0.0570 = 0.5670.$$

Example 2, page 217

Let $X_1, X_2, \cdots, X_{20}$ be a random sample of size 20 drawn from Poisson distribution with mean $\lambda = 1$. Then

Q1. what is the distribution of $Y = \sum_{i=1}^{20} X_i$?

Q2. find $P(16 < Y \leq 21)$ approximately?

Example 2, page 217

Let $X_1, X_2, \cdots, X_{20}$ be a random sample of size 20 drawn from Poisson distribution with mean $\lambda = 1$. Then

Q1. what is the distribution of $Y = \sum_{i=1}^{20} X_i$?

Q2. find $P(16 < Y \leq 21)$ approximately?

Q1: Recall that the mgf of Poisson distribution with mean $\lambda > 0$ is

$$M(t) = e^{\lambda(e^t - 1)}, \quad t \in \mathbb{R}$$

Then by Theorem 5.4-1, the mgf of Y, $M_Y(t)$ takes the form of

$$M_Y(t) = \prod_{i=1}^{20} e^{(e^t-1)} = e^{20(e^t-1)},$$

implying that Y has a Poisson distribution with mean and variance both equal to 20.

Example 2, page 217

Q2: Let $X_1, X_2, \cdots, X_n$ be a random sample of size n from the Poisson distribution with mean $\lambda = 1$. Then by CLT,

$$\frac{Y/n - 1}{1/\sqrt{n}} \xrightarrow{d} N(0,1).$$

Therefore, for large n, Y can be approximated by $N(n, n)$.

When $n = 20$,

$$\begin{aligned} P(16 < Y \leq 21) &= P(17 \leq Y \leq 21) \\ &\approxeq P(16.5 \leq Y \leq 21.5) \\ &= P\left(\frac{16.5 - 20}{\sqrt{20}} \leq \frac{Y - 20}{\sqrt{20}} \leq \frac{21.5 - 20}{\sqrt{20}}\right) \\ &\approxeq \Phi(0.335) - \Phi(-0.783) = 0.4142 \end{aligned}$$

One may also try $P(16 < Y \leq 21) = \sum_{x=17}^{21} \frac{20^x e^{-20}}{x!} = 0.4226$

Section 5.8 Chebyshev's Inequality and Convergence in Probability

Motivation

CLT: given a random sample of size n, say $X_1, \cdots, X_n$, from a random distribution (the distribution may be unknown) with its mean and variance known, it is possible to compute approximately the probability of events for $\overline{X}$ or $\sum_{i=1}^{n} X_i$.

Motivation

CLT: given a random sample of size n, say $X_1, \cdots, X_n$, from a random distribution (the distribution may be unknown) with its mean and variance known, it is possible to compute approximately the probability of events for $\overline{X}$ or $\sum_{i=1}^{n} X_i$.

Chebyshev's inequality: given a random distribution (the distribution may be unknown) with its mean and variance known, it is possible to compute approximately the probability of events for a random variable X whose distribution is the given distribution.

Theorem 5.8-1 [Chebyshev's inequality]

Theorem 5.8-1(Chebyshev's inequality)

If a RV X has a finite mean μ and finite nonzero variance σ^2, then for every $k \geq 1$,

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

Proof of Theorem 5.8-1

Consider the discrete RV case. Let $f(x) : \overline{S} \to (0, 1]$ be the pmf.

$$\sigma^2 = E[(X - \mu)^2] = \sum_{x \in \overline{S}} (x - \mu)^2 f(x)$$
$$= \sum_{x \in A} (x - \mu)^2 f(x) + \sum_{x \in A'} (x - \mu)^2 f(x)$$

where

$$A = \{x \mid |x - \mu| \geq k\sigma\}$$

Proof of Theorem 5.8-1

Since

$$\sum_{x \in A'} (x-\mu)^2 f(x) \geq 0$$

$$\sigma^2 \geq \sum_{x \in A} (x-\mu)^2 f(x) \geq k^2\sigma^2 \sum_{x \in A} f(x) = k^2\sigma^2 P(X \in A)$$

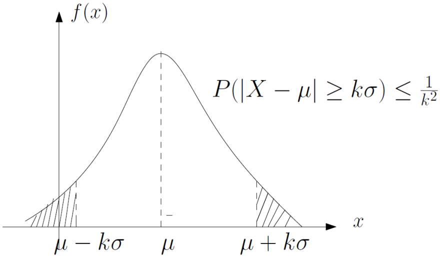
Corollary 5.8-1, Page 222

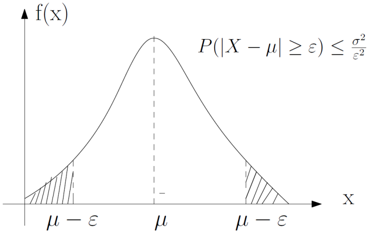
Corollary 5.8-1(Chebyshev's inequality)

If a RV X has a finite mean μ and finite nonzero variance σ^2, then for any $\varepsilon > 0$,

$$P(|X - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

Graphical Interpretation of Cheybshev's Inequality





This links to the interpretation of σ^2: a measure of dispersion of the values that X can take with respect to its mean μ.

Example 1, page 222

Let X be a RV with mean $\mu = 25$ and variance $\sigma^2 = 16$.

Question: Find a lower bound for $P(17 < X < 33)$ and an upper bound for $P(|X - 25| \geq 12)$.

Example 1, page 222

Let X be a RV with mean $\mu = 25$ and variance $\sigma^2 = 16$.

Question: Find a lower bound for $P(17 < X < 33)$ and an upper bound for $P(|X - 25| \geq 12)$.

Lower bound for:

$$P(17 < X < 33) = 1 - P(|X - \mu| \geq 2\sigma) \geq 1 - \frac{1}{4}$$

Upper bound for:

$$P(|X - 25| \geq 12) = P(|X - \mu| \geq 3\sigma) \leq \frac{1}{9}$$

Note: X is arbitrary!