Chapter 6 - Lecture 3
The distribution of the sample mean

Yuan Huang

January 14th, 2013

General Properties of Sample Mean

General: Mean and variance of Sample Mean

A random sample $X_1, X_2, ..., X_n$ from a distribution with mean μ and variance σ^2, then the distribution of the sample mean $\bar{X}$ has:

1. mean $E(\bar{X}) = \mu_{\bar{X}} = \mu$
2. variance $V(\bar{X}) = \sigma^2_{\bar{X}} = \frac{\sigma^2}{n}$

It means:

1. The sampling distribution of $\bar{X}$ is centered precisely at the mean of the population from which the sample has been selected.
2. The sampling distribution of $\bar{X}$ becomes more concentrated about μ as the sample size n increases.

Law of Large Numbers

$$\bar{X} \to ?, \text{ as } n \to \infty$$

Law of Large Numbers (LLN)

Thm:. If $X_1, \ldots, X_n$ is a random sample from a distribution with mean μ and variance σ^2, then as $n \to \infty$, $\bar{X}_n$ converges to μ :

- In mean square $E[(\bar{X} - \mu)^2] \to 0$
- In probability $P(|\bar{X} - \mu| \geq \epsilon) \to 0$

Remarks: LLN is one theoretical support of using large sample size - as sample size goes large the sample estimate becomes accurate.

For weak law:

Easy way to remember: The chance that $\bar{X}$ is far away from μ is going to 0 as sample size is growing!

[Proof]: By Chebyshev's Inequality,

$$\begin{aligned} P(|\bar{X}-\mu| \geq \epsilon) &= P\left(|\bar{X}-\mu| \geq \left(\epsilon\frac{\sqrt{n}}{\sigma}\right)\frac{\sigma}{\sqrt{n}}\right) \\ &\leq \frac{1}{\left(\epsilon\frac{\sqrt{n}}{\sigma}\right)^2} = \frac{\sigma^2}{n\epsilon^2} \end{aligned}$$

Distribution of Sample Mean for Normal distribution

Distribution of Sample Mean for Normal distribution

A random sample $X_1, X_2, ..., X_n$ from a normal distribution with mean μ and variance σ^2 $(X_i \sim N(\mu, \sigma^2))$ then the sample mean $\bar{X}$ has a sampling distribution which is :

- normally distributed (by prop 3 from Lec 6.2)
- with mean $\mu_{\bar{X}} = \mu$ (by prop 1 from Lec 6.2)
- with variance $\sigma^2_{\bar{X}} = \dfrac{\sigma^2}{n}$ (by prop 2 from Lec 6.2)

In short as: $\bar{X} \sim N\left(\mu, \dfrac{\sigma^2}{n}\right)$

Exercise 6.19

[6.19] Suppose the sediment density of a randomly selected specimen from a certain region is normally distributed with mean 2.65 and standard deviation 0.85. If a random sample of 25 specimen is selected, what is the probability that sample mean is at most 3.00 ?

Asymptotic distribution of $\bar{X}$ for general populations

Central Limit Theorem

Central Limit Theorem (CLT)

Thm:. If $X_1, \ldots, X_n$ is a random sample with mean μ and variance σ^2, then as $n \to \infty$, the limiting distribution of $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ is standard normal, written as

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \to_d N(0, 1).$$

Central Limit Theorem (CLT)

Thm:. If $X_1, \ldots, X_n$ is a random sample with mean μ and variance σ^2, then as $n \to \infty$, the limiting distribution of $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ is standard normal, written as

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \to_d N(0, 1).$$

Comments: converge in distribution as converge in cdf.

1. $\lim_{n\to\infty} P(\sqrt{n}(\bar{X}_n - \mu)/\sigma \leq z) \to \Phi(z)$ for any $z \in \Re$;
2. $\lim_{n\to\infty} P((T_o - n\mu)/(\sqrt{n}\sigma) \leq z) \to \Phi(z)$ for any $z \in \Re$;

Central Limit Theorem - CLT

A random sample $X_1, X_2, ..., X_n$ from ANY distribution. The sample mean $\bar{X}$ $\bar{X}$ is asymptotically normally distributed.

[Think:] Asymptotically means when the sample size n is large. But how large is large?

Central Limit Theorem - CLT

A random sample $X_1, X_2, ..., X_n$ from ANY distribution. The sample mean $\bar{X}$ $\bar{X}$ is asymptotically normally distributed.

[Think:] Asymptotically means when the sample size n is large. But how large is large?

[Rule of Thumb :] If $n > 30$ the Central Limit Theorem can be used.

Example 6.8 page 294

[e.g. 6.8] When a batch of a certain chemical product is prepared, the amount of a particular impurity in the batch is a random variable with mean 4.0 g and standard deviation 1.5 g. If 50 batches are independently prepared, what is the (approximate) probability that the sample average amount of impurity $\bar{X}$ is between 3.5g and 3.8g ?

Other Applications of CLT (1)

Justify normal approximation of Binomial Distribution (we did this in Stat 318 but we didn't justify it using CLT)

Let $X \sim \text{Binomial}(n, p)$, if $np \geq 10$ and $nq \geq 10$, then

$$P(X \leq x) = \Phi\left(\frac{x - np + 0.5}{\sqrt{npq}}\right)$$

Exercise 6.20

[6.20] The first assignment in a statistical computing class involves running a short program. If past experience indicates that 40% of all students will make no programming errors. compute the approximate probability that in a class of 50 students, at leas 25 will make no errors.

Other Applications of CLT (2)

Let $X_1, X_2, ..., X_n$ be random sample from a distribution for which only positive values are possible $(P(X_i > 0) = 1)$. Then if $n$is sufficiently large, the product $Y = X_1 X_2 \cdot \ldots \cdot X_n$ has approximately a lognormal distribution.

Homework for this session

Part of HW 2:

- Section 6.2 page 298 11, 12, 14
  - Exercises 11($E(\bar{X}), V(\bar{X})$)
  - Exercises 12 (distribution of $\bar{X}$ for normal distn)
  - Exercises 14 (CLT)