MATH 3338 Probability

Wenjiang Fu: wfu2@central.uh.edu

Lecture 9 - MATH 3338
Ch 9 Central Limit Theorem

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Outline

1. Central Limit Theorem by examples

2. Central Limit Theorem for Continuous RVs

Central Limit Theorem

- Suppose A coin is tossed multiple times n with the indicator of head up $X_1, ..., X_n$, i.e. a Bernoulli Trial of n times with prob $p > 0$. Take the sum S_n, we have that S_n follows a Binomial distr with mean np, and variance $np(1-p)$.
  Now we take the standardized sums, by subtracting the mean np and dividing by its standard deviation $\sqrt{np(1-p)}$.
  Definition 9.1 The standardized sum of S_n i s given by

  $$S_n^* = \frac{S_n - np}{\sqrt{np(1-p)}}.$$

  S_n^* has mean 0 and variance 1. Because S_n takes values $0, 1, .., n$, there are $n+1$ values. We plot them with a height S_j^* for the j-th, we have a plot with a nice bell-curve as shown in the textbook p 327.

Central Limit Theorem

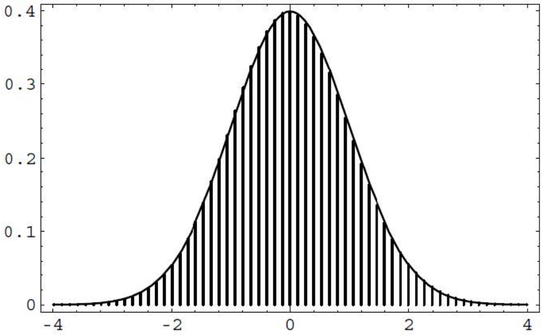

Figure 9.3: Corrected spike graph with standard normal density.

Central Limit Theorem

We have the following theorem.

- **Theorem 9.1** (Central Limit Theorem for Binomial Distr) For the binomial distr $Bin(n, p)$, we have

$$\sqrt{npq}\, b(n, p, \ll np + x\sqrt{npq} \gg) = \phi(x)$$

  where $\ll np + x\sqrt{npq} \gg$ is the nearest integer close to $np + x\sqrt{npq}$, $b(n, p, k)$ is the Binomial probability at the integer k, and $\phi(x)$ is the standard normal density function $\phi(x) = e^{-x^2/2}/\sqrt{2\pi}$.

- **Approximating Binomial Distributions** If we wish to find an approximation for binomial prob $b(n, p, j)$, we set $j = p + x\sqrt{npq}$, and solve for x, obtaining $x = (j - np)/\sqrt{npq}$. By Theorem 9.1, $\sqrt{npq}b(n, p, j)$ is approximately equal to $\phi(x)$. so

$$b(n, p, j) = \phi(x)/\sqrt{npq} = \frac{1}{\sqrt{npq}}\phi\left(\frac{j - np}{\sqrt{npq}}\right)$$

We have the following example.

- **Example 9.2** A coin is tossed 100 times. Estimate the probability the the number of heads lies between 40 and 60 (inclusive).
  Solution The expected number of heads is $100(1//2) = 50$ for fair coins. The standard deviation for the number of heads is $\sqrt{npq} = 5$. $n = 100$ is reasonably large, we have

$$P(40 \leq S_n \leq 60) = P\left(\frac{39.5 - 50}{5} \leq S_n^* \leq \frac{60.5 - 50}{5}\right)$$

$$= P(-2.1 \leq S_n^* \leq 2.1) = .9642$$

The actual number is .9648, is well approximated.

- **Normal Approximation to Binomial Prob**
  In general, the probability of binomial $S_n \sim Bin(n, p)$ is calculated by

$$P(i \leq S_n \leq j) = P\left(\frac{i - 1/2 - np}{\sqrt{np(1-p)}} \leq Z \leq \frac{j + 1/2 - np}{\sqrt{np(1-p)}}\right)$$

Central Limit Theorem for Continuous RVs

Consider two examples.

- **Uniform Distr** Let $X_1, ..., X_n$ be uniform RVs on [0,1]. Consider their sum $S_n = X_1 + ... + S_n$. $E(S_n) = n/2$, $Var(S_n) = n/12$. The standardized sum
$$S_n^* = \frac{S_n - n/2}{\sqrt{n/12}}$$
The standardized sum S_n^* has mean 0 and variance 1. $S_n^* \sim N(0,1)$, i.e. standard normal $N(0,1)$ can be used to calculate the prob of S_n^* and then S_n.
- **Exponential Distr** Let $X_1, ..., X_n$ be exponential RVs $Exp(\lambda)$ with parameter $\lambda > 0$. Consider their sum $S_n = X_1 + ... + S_n$. $E(S_n) = n/\lambda$, $Var(S_n) = n/\lambda^2$ The standardized sum
$$S_n^* = \frac{S_n - n/\lambda}{\sqrt{n}/\lambda}$$
The standardized sum S_n^* has mean 0 and variance 1.

Central Limit Theorem for Continuous RVs

- **Exponential Distr** Let $X_1, ..., X_n$ be exponential RVs $Exp(\lambda)$ with parameter $\lambda > 0$. Consider their sum $S_n = X_1 + ... + S_n$. $E(S_n) = n/\lambda$, $Var(S_n) = n/\lambda^2$ The standardized sum

$$S_n^* = \frac{S_n - n/\lambda}{\sqrt{n}/\lambda}$$

The density functions of the sum S_n and the standardized sum S_n^* are

$$f_{S_n}(x) = \frac{\lambda e^{-\lambda x}(\lambda x)^{n-1}}{(n-1)!},$$

$$f_{S_n^*}(x) = \frac{\sqrt{n}}{\lambda} f_{S_n}\left(\frac{\sqrt{n}x + n}{\lambda}\right).$$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Central Limit Theorem for Continuous RVs

- **Theorem 9.6** (Central Limit Theorem) Let $S_n = X_1 + ... + X_n$ be the sum of n indep continuous RVs with common density function p with expected value μ and variance σ^2. Let $S_n^* = (S_n - n\mu)/\sqrt{n}\sigma$. Then for all $a < b$,

$$\lim_{n \longrightarrow \infty} P(a < S_n^* < b) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx.$$

Note This theorem implies that for any distribution with finite mean and variance, this theorem applies, regardless of what distribution, symmetric or not, including Normal, uniform, exponential, chi-squares, etc.

Central Limit Theorem for Continuous RVs

- **Example 9.9** A surveyor wants to measure a known distance, say of 1 mile, using some method with *n* observations. He notices that the mean is $\mu = 1$, and the standard deviation $\sigma = 0.0002$. If *n* is large, the average S_n/n has a density function approximately normal, with mean $\mu = 1$ mile, and standard deviation $\sigma/\sqrt{n} = .0002/\sqrt{n}$.
  Q: How many measures need to be made to make sure that his average lies within .0001 of the true value?
  Solution. Take the error bound to be the distance $\varepsilon = .0001$. By the Chebyshev inequality,
  $$P\left(\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right) \leq \frac{.0002^2}{n\varepsilon^2} = \frac{4}{n},$$
  To make sure that his average will be within a distance of ε, we often need to make sure the above prob to be out of the distance ε smaller than .05, i.e. $4/n \leq .05$. Then $n \geq 4/.05 = 80$. But such an estimate by the Chebyshev inequality is too large.

- **Example 9.9**
  Solution. We need to solve it within the probability using the CLT, i.e.

$$P\left(\left|\frac{S_n}{n} - \mu\right| < \varepsilon\right) \geq .95$$

Now we need to calculate the probability using CLT.

$$P\left(\left|\frac{S_n}{n} - \mu\right| < \varepsilon\right) = P\left(\frac{\left|\frac{S_n}{n} - \mu\right|}{\sigma/\sqrt{n}} < \frac{\varepsilon}{\sigma/\sqrt{n}}\right) = .95$$

$$P(|Z| \leq \frac{\varepsilon}{\sigma/\sqrt{n}}) = .95 \rightarrow \frac{\varepsilon}{\sigma/\sqrt{n}} = 1.96$$

$$n = \left(\frac{1.96\sigma}{\varepsilon}\right)^2 = 16$$

Central Limit Theorem for Continuous RVs

- **Estimating population mean with sample mean**
  The sample mean $\overline{X} = (X_1 + ... + X_n)/n$ can be used to estimate the population mean μ.

  $$P\left(\left|\overline{X} - \mu\right| < \varepsilon\right) = .95$$

  It implies that

  $$P\left(\overline{X} - \varepsilon \leq \mu \leq \overline{X} + \varepsilon\right) = .95$$

  that is the random interval $[\overline{X} - \varepsilon,\ \overline{X} + \varepsilon]$ has a 95% probability to contain the population μ, and thus the interval $[\overline{X} - \varepsilon,\ \overline{X} + \varepsilon]$ is called a 95% confidence interval of the population mean μ.