Parametric Statistics

Large Random Samples

Sofia Triantafillou

sof.triantafillou@gmail.com

University of Crete
Department of Mathematics and Applied Mathematics

October 26, 2023

Lecture Summary

6.3 The central limit theorem

6.4 Correction for continuity

Last time

Let $X_1, X_2, \ldots$ be a sequence of random variables, let X be a random variable.

Convergence in Probability

$X_1, X_2, \ldots$ converges in probability to X if

$$\forall\epsilon > 0, lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0$$

Convergence in Distribution

$X_1, X_2, \ldots$ converges in distribution to X if

$$\forall\epsilon > 0, lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

for all x where F_X is continuous.

Last time

Markov Inequality

Let X be a random variable such that $P(X \geq 0) = 1$. Then for every real number t,

$$P(X \geq t) \leq \frac{E(X)}{t}.$$

Chebysev Inequality

Let X be a random variable for which $Var(X)$ exists. Then for every number $t > 0$,

$$P(|X - E(X)| \geq t) \leq \frac{Var(X)}{t^2}.$$

Compute the probability of Uniform

Example

Suppose that a random sample of size $n = 12$ is taken from the uniform distribution on the interval $[0, 1]$. We shall approximate the value of $\Pr\left(\left|\overline{X}_n - \frac{1}{2}\right| \leq 0.1\right)$

- $E(X_i) = \frac{1}{2}, E(\overline{X}_n) = \frac{1}{2}$
- $Var(X_i) = \frac{1}{12}, Var(\overline{X}_n) = \frac{1}{12n}$

Central Limit Theorem

Lindberg and Lévy

If the random variables $X_1, \ldots, X_n$ form a random sample of size n from a given distribution with mean μ and variance σ^2 $\left(0 < \sigma^2 < \infty\right)$, then for each fixed number x,

$$\lim_{n \to \infty} \Pr\left[\frac{\overline{X}_n - \mu}{\sigma / n^{1/2}} \leq x\right] = \Phi(x),$$

where Φ denotes the c.d.f. of the standard normal distribution.

Stable variance

- $S_n = \sum_{i=1}^{n} X_i$, mean $n\mu$, variance $n\sigma^2$.
- $\overline{X_n} = \frac{S_n}{n}$, mean μ variance $\frac{\sigma^2}{n}$.
- $\frac{S_n}{\sqrt{n}}$, mean $\mu\sqrt{n}$, variance σ^2.
- $Z_n = \frac{S_n - n\mu}{\sqrt{n}\sigma} = \frac{\sqrt{n}(\overline{X_n} - \mu)}{\sigma}$, mean 0, variance 1.

The Delta Method

Sometimes we are interested in the asymptotic behavior of a function of the sample mean.

Delta Method for Average of a Random Sample.

Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables from a distribution with mean μ and finite variance σ^2. Let α be a function with continuous derivative such that $\alpha(\mu)' \neq 0$. Then the asymptotic distribution of

$$\frac{n^{1/2}}{\sigma\alpha'(\mu)}[\alpha(\overline{X}_n) - \alpha(\mu)]$$

is the standard normal distribution.

Example

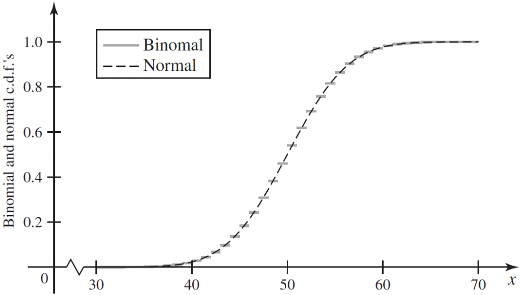
What is the asymptotic limit of $\sqrt{n}\left(\overline{X}_n^2 - \mu^2\right)$?

Correction for continuity

- $Y \sim \text{Binomial}\left(n = 100, p = \frac{1}{2}\right)$. We are interested in $P(Y \leq 50)$.
- We know that a Binomial $\left(n = 100, p = \frac{1}{2}\right)$ can be written as the sum of n i.i.d. Bernoulli (p) random variables.
- Using CLT:

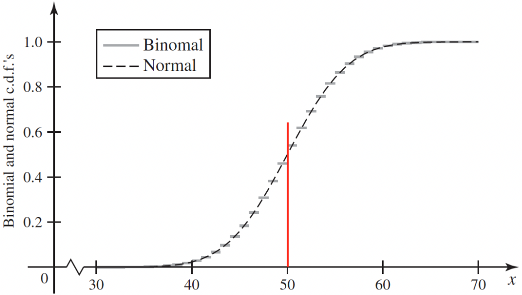
Correction for continuity

- $Y \sim \text{Binomial}\left(n = 100, p = \frac{1}{2}\right)$. We are interested in $P(Y \le 50)$.
- We know that a Binomial $\left(n = 100, p = \frac{1}{2}\right)$ can be written as the sum of n i.i.d. Bernoulli (p) random variables.
- Using CLT: $P(Y \le 50) = \Phi(0) = 0.5$
- Using Binomial: $P(Y \le 50) = 0.539$



Correction for continuity

- $Y \sim \text{Binomial}\left(n = 100, p = \frac{1}{2}\right)$. We are interested in $P(Y \leq 50)$.
- We know that a $\text{Binomial}\left(n = 100, p = \frac{1}{2}\right)$ can be written as the sum of n i.i.d. Bernoulli (p) random variables.
- Using CLT: $P(Y \leq 50) = \Phi(0) = 0.5$
- Using Binomial: $P(Y \leq 50) = 0.539$



Correction for Continuity

- $Y \sim \text{Binomial}\left(n = 100, p = \frac{1}{2}\right)$. We are interested in $P(Y \leq 50)$.
- We know that a $\text{Binomial}\left(n = 100, p = \frac{1}{2}\right)$ can be written as the sum of n i.i.d. Bernoulli (p) random variables.
- Using CLT: $P(Y \leq 50) = \Phi(0) = 0.5$
- Using Binomial: $P(Y \leq 50) = 0.539$
- Instead of $\Pr(Y \leq x)$ use $\Pr(Y \leq x + 0.5)$, which is larger and usually closer to $\Pr(X \leq x)$.
- $\Pr(Y = x) = P(x - 0.5 \leq Y \leq x + 0.5)$

Application of the CLT

- You are doing a poll on "ratio of people who believe in climate change".
- True ratio: p, estimate $\overline{X}_n$.
- Your boss wants a guarantee that your prediction will be "off" by more than 1% with probability at most 5%

Application of the CLT

- You are doing a poll on "ratio of people who believe in climate change".
- True ratio: p, estimate $\overline{X}_n$.
- Your boss wants a guarantee that your prediction will be "off" by more than 1% with probability at most 5%
- No guarantee for finding exactly p, so

$$P(|\overline{X}_n - p| \geq 0.01) \leq 0.05$$

- Apply Chebysev inequality with $t = 0.01$:
- Apply CLT:

Group Exercise

- Form teams of two.
- Go to https://rolladie.net/roll-4-dice.
- Complete the exercise sheet given to you and answer the questions.