Mathematical statistics

February 27^{th}, 2018

Lecture 7: The distribution of the sample mean

Topics

Week 1	Probability reviews
Week 2	**Chapter 6: Statistics and Sampling Distributions**
Week 4	Chapter 7: Point Estimation
Week 7	Chapter 8: Confidence Intervals
Week 10	Chapter 9: Test of Hypothesis
Week 14	Regression

Overview

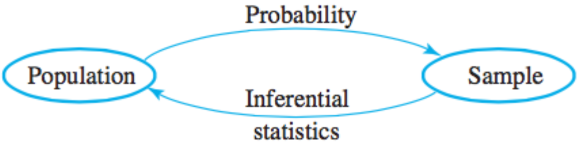
6.1 Statistics and their distributions

6.2 The distribution of the sample mean

6.3 The distribution of a linear combination

Order 6.1 $\rightarrow$ 6.3 $\rightarrow$ 6.2

Random sample



Definition

The random variables $X_1, X_2, ..., X_n$ are said to form a (simple) random sample of size n if

1. the X_i's are independent random variables
2. every X_i has the same probability distribution

Questions for this chapter

Given a random sample $X_1, X_2, \ldots, X_n$, and

$$T = a_1X_1 + a_2X_2 + \ldots + a_nX_n$$

- If we **know** the distribution of X_i's, can we obtain the distribution of T?
  - Simple cases
  - If $X_i's$ follow normal distribution, then so does T.
- If we **don't know** the distribution of X_i's, can we still obtain/approximate the distribution of T?
  - Can we at least compute the mean and the variance?
  - When T is the sample mean, i.e. $a_1 = a_2 = \ldots = \frac{1}{n}$

Theorem

Let $X_1, X_2, \ldots, X_n$ be independent normal random variables (with possibly different means and/or variances). Then

$$T = a_1 X_1 + a_2 X_2 + \ldots + a_n X_n$$

also follows the normal distribution.

What are the mean and the standard deviation of T?

- $E(T) = a_1 E(X_1) + a_2 E(X_2) + \ldots + a_n E(X_n)$
- $\sigma_T^2 = a_1^2 \sigma_{X_1}^2 + a_2^2 \sigma_{X_2}^2 + \ldots + a_n^2 \sigma_{X_n}^2$

Example 1*

Problem

Let $X_1, X_2, \ldots, X_n$ be a random sample from $\mathcal{N}(\mu, \sigma^2)$ (that is, normal distribution with mean μ and standard deviation σ).
Let $\bar{X}$ be the sample mean

$$\bar{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

What is the distribution of $\bar{X}$?

What if X_i's are not normal distributions?

Given a random sample $X_1, X_2, \ldots, X_n$, and

$$T = a_1X_1 + a_2X_2 + \ldots + a_nX_n$$

If we don't know the distribution of X_i's, can we obtain the distribution of T?

Linear combination of random variables

Theorem

Let $X_1, X_2, \ldots, X_n$ be independent random variables (with possibly different means and/or variances). Define

$$T = a_1X_1 + a_2X_2 + \ldots + a_nX_n,$$

then the mean and the standard deviation of T can be computed by

- $E(T) = a_1E(X_1) + a_2E(X_2) + \ldots + a_nE(X_n)$
- $\sigma_T^2 = a_1^2\sigma_{X_1}^2 + a_2^2\sigma_{X_2}^2 + \ldots + a_n^2\sigma_{X_n}^2$

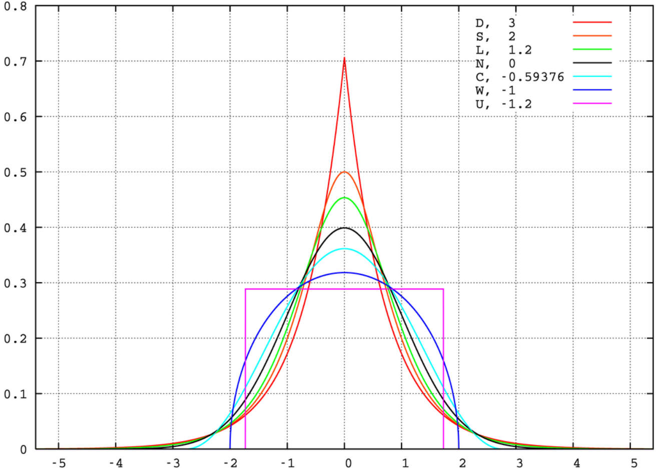
Example

Problem

A gas station sells three grades of gasoline: regular unleaded, extra unleaded, and super unleaded. These are priced at 2.20, 2.35, and 2.50 per gallon, respectively.
Let X_1, X_2, and X_3 denote the amounts of these grades purchased (gallons) on a particular day. Suppose the X_i's are independent with $\mu_1 = 1000$, $\mu_2 = 500$, $\mu_3 = 300$, $\sigma_1 = 100$, $\sigma_2 = 80$, $\sigma_3 = 50$. Compute the expected value and the standard deviation of the revenue from sales

$$Y = 2.2X_1 + 2.35X_2 + 2.5X_3.$$

Bad news



In general, the mean and the variance do not define a probability distribution.

Mean and variance of the sample mean

Problem

Given a random sample $X_1, X_2, ..., X_n$ from a distribution with mean μ and standard deviation σ, the mean is modeled by a random variable $\bar{X}$,

$$\bar{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

- *Compute $E(\bar{X})$*
- *Compute $Var(\bar{X})$*

Mean and variance of the sample mean

Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with mean value μ and standard deviation σ. Then

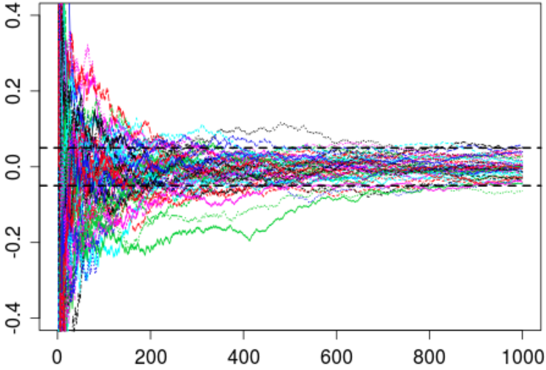
1. $E(\overline{X}) = \mu_{\overline{X}} = \mu$

2. $V(\overline{X}) = \sigma^2_{\overline{X}} = \sigma^2/n$ and $\sigma_{\overline{X}} = \sigma/\sqrt{n}$

THEOREM If $X_1, X_2, \ldots, X_n$ is a random sample from a distribution with mean μ and variance σ^2, then $\overline{X}$ converges to μ

a. In mean square $E[(\overline{X} - \mu)^2] \to 0$ as $n \to \infty$

b. In probability $P(|\overline{X} - \mu| \geq \varepsilon) \to 0$ as $n \to \infty$



The Central Limit Theorem

Theorem

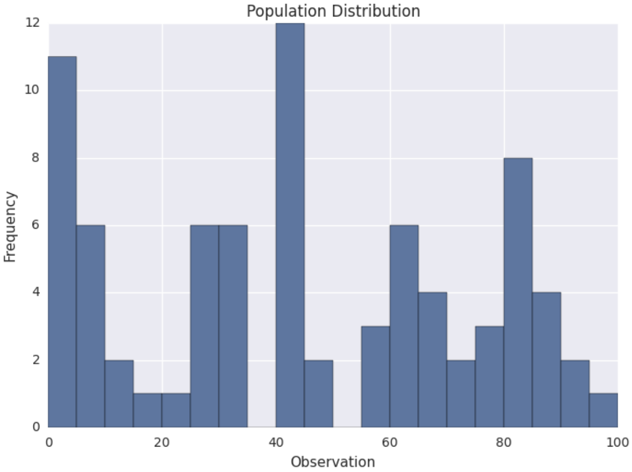
Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with mean μ and variance σ^2. Then, in the limit when $n \to \infty$, the standardized version of $\bar{X}$ have the standard normal distribution

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq z\right) = \mathbb{P}[Z \leq z] = \Phi(z)$$

Rule of Thumb:

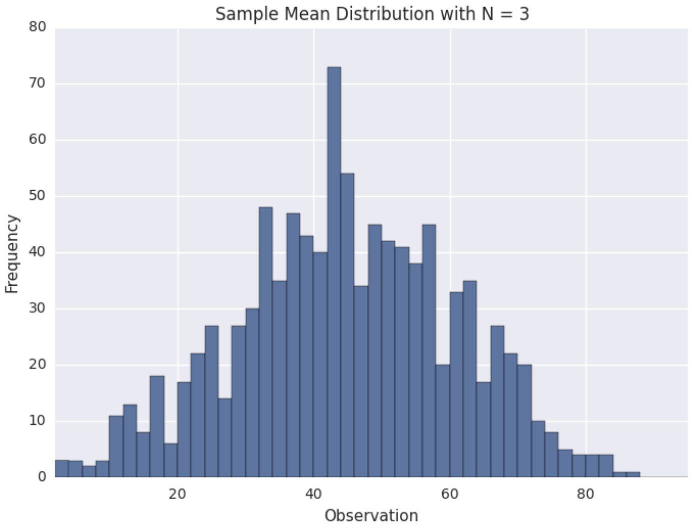
If $n > 30$, the Central Limit Theorem can be used for computation.

Example: population distribution

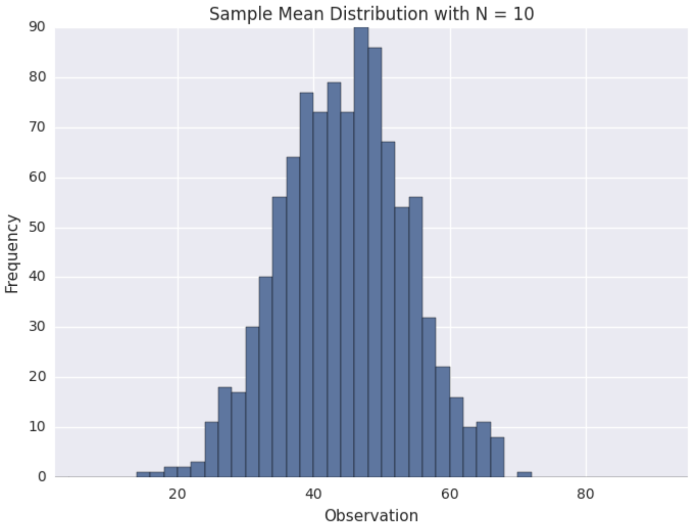


Matt Nedrick (2015).
http://github.com/mattnedrich/CentralLimitTheoremDemo

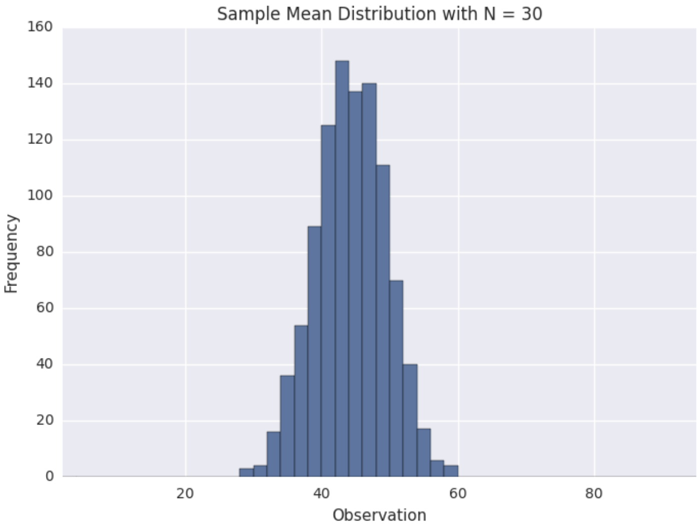
Sample distribution: $n = 3$



Sample distribution: $n = 10$



Sample distribution: $n = 30$



Example

Problem

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 value 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 X is between 3.5 and 3.8 g?

Hint:

- First, compute $\mu_{\bar{X}}$ and $\sigma_{\bar{X}}$
- Note that

$$\frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}}$$

is (approximately) standard normal.

Example

Problem

The tip percentage at a restaurant has a mean value of 18% and a standard deviation of 6%.

What is the approximate probability that the sample mean tip percentage for a random sample of 40 bills is between 16% and 19%?