Mathematical statistics

March 1^{st}, 2018

Lecture 8: The distribution of the sample mean

- 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 $\to$ 6.3 $\to$ 6.2

Next class (Monday, 03/04): QUIZ 1
Next Friday (03/08): Homework 2 due

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. $a_1 = a_2 = \ldots = \frac{1}{n}$

Linear combination of random variables

Theorem

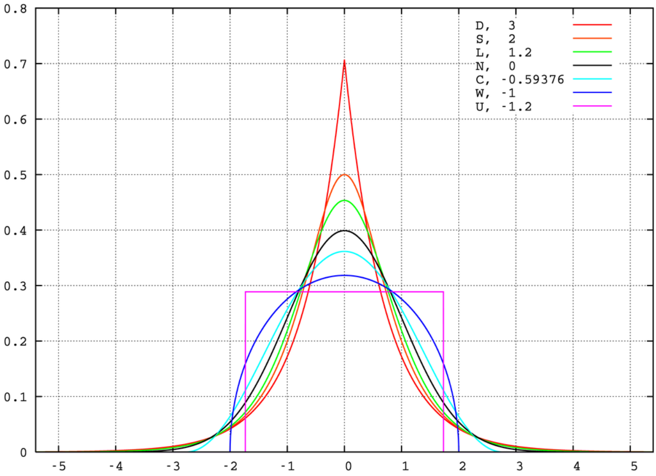
Let $X_1, X_2, \ldots, X_n$ be independent random variables (with possibly different means and/or variances). Define

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

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

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

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 $\sigma_{\overline{X}} = \sigma/\sqrt{n}$

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

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%?

Example

Problem

The weight of a certain crab is known to have an expected value of 1.4 kg and standard deviation of 0.4 kg. The distribution is unknown. Consider one net that contains 112 crabs.

What is the probability that the total weight of the crabs in the net is less than 150 kg?