Mathematical statistics

September 13^{st}, 2018

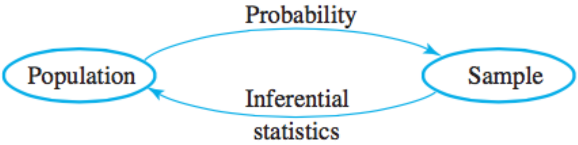
Lecture 6: Statistics and sampling distributions

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

Chapter 6: Summary

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

1. If the distribution and the statistic T is simple, try to construct the pmf of the statistic
2. If the probability density function $f_X(x)$ of X's is known, the
   - try to represent/compute the cumulative distribution (cdf) of T
     $$\mathbb{P}[T \leq t]$$
   - take the derivative of the function (with respect to t)

Section 6.3: Linear combination of normal random variables

Theorem

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

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

also follows the normal distribution.

If X has a normal distribution with mean μ and standard deviation σ, then

$$Z = \frac{X - \mu}{\sigma}$$

has a standard normal distribution. Thus

$$\begin{aligned} P(a \leq X \leq b) &= P\left(\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\right) \\ &= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right) \end{aligned}$$

$$P(X \leq a) = \Phi\left(\frac{a-\mu}{\sigma}\right) \quad P(X \geq b) = 1 - \Phi\left(\frac{b-\mu}{\sigma}\right)$$

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

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

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.

Practice problems

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 time that it takes a randomly selected rat of a certain subspecies to find its way through a maze is a normally distributed random variable with mean $\mu = 1.5$*(minutes) and standard deviation* $\sigma = .35$ *(minutes).*
Suppose five rats are selected. Let $X_1, X_2, \ldots, X_5$ *denote their times in the maze. Assuming the* X_i'*s to be a random sample from this normal distribution, what is the probability that the total time for the five*

$$T = X_1 + X_2 + X_3 + X_4 + X_5$$

is between 6 and 8 minutes.

Example

Problem

Let $X_1, X_2, \ldots, X_n$ *be random sample from a normally distribution with mean* 2.65 *and standard deviation* 0.85*.*

- *If* $n = 25$*, compute*
$$P[\bar{X} \leq 3]$$
- *Find* n *such that*
$$P[\bar{X} \leq 3] \geq 0.95$$

Example

Problem

Let X be a random sample from the exponential distribution

$$f(x) = \begin{cases} e^{-x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

Compute

$$E[e^{-X}]$$

Example

Problem

Let X be a random sample from the exponential distribution

$$f(x) = \begin{cases} e^{-x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

Compute

$$E[e^{X/2}]$$

Example

Problem

Let X be a random sample from the exponential distribution

$$f(x) = \begin{cases} e^{-x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

Compute

$$E[e^{tX}]$$

for $t < 1$.

Moment generating function

Definition

The moment generating function (mgf) of a continuous random variable X is

$$M_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f_X(x)dx$$

- Reading: 3.4 and 4.2
- In our previous example

$$M_X(t) = \frac{1}{1-t} \quad \text{if} \quad t < 1.$$