Mathematical statistics

February 22^{nd}, 2018

Lecture 5: Statistics and sampling distribution (cont.)

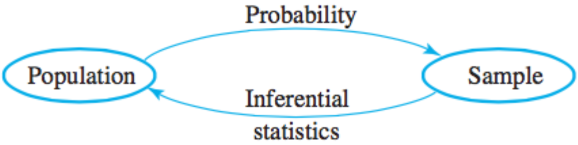
Topics

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

Order 6.1 $\to$ 6.3 $\to$ 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

Definition

A statistic is any quantity whose value can be calculated from sample data

- prior to obtaining data, there is uncertainty as to what value of any particular statistic will result $\rightarrow$ a statistic is a random variable
- the probability distribution of a statistic is referred to as its *sampling distribution*

Example of a statistic

- Let $X_1, X_2, \ldots, X_n$ be a random sample of size n
- The sample mean of $X_1, X_2, \ldots, X_n$, defined by

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

is a statistic

- When the values of $x_1, x_2, \ldots, x_n$ are collected,

$$\bar{x} = \frac{x_1 + x_2 + \ldots x_n}{n},$$

is a realization of the statistic $\bar{X}$

Example of a statistic

- Let $X_1, X_2, \ldots, X_n$ be a random sample of size n
- The random variable

$$T = X_1 + 2X_2^2 + 3X_5^3$$

  is a statistic
- When the values of $x_1, x_2, \ldots, x_n$ are collected,

$$t = x_1 + 2x_2^2 + 3x_5^3,$$

  is a realization of the statistic T

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?
- If we **don't know** the distribution of X_i's, can we still obtain/approximate the distribution of T?

Example 1

Problem

Consider the distribution P

x	10	15	20
$p(x)$	0.2	0.3	0.5

Let $\{X_1, X_2\}$ *be a random sample of size 2 from* P, *and* $T = X_1 + X_2$.

1. *Compute* $P[T = 100]$
2. *Derive the probability mass function of* T
3. *Compute the expected value and the standard deviation of* T

Example 2

Problem

Let $\{X_1, X_2\}$ be a random sample of size 2 from the exponential distribution with parameter λ

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

and $T = X_1 + X_2$.
What is the distribution of T?

For continuous random variable:

$$F_X(t) = P(X \leq t) = \int_{-\infty}^{t} f(x)\, dx$$

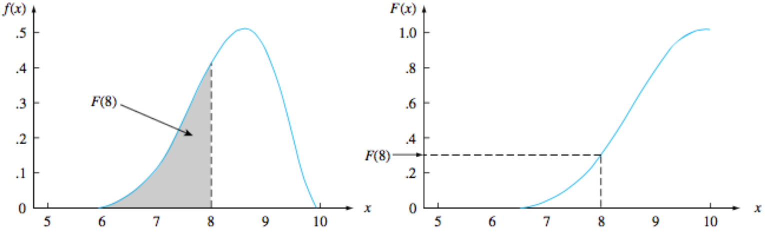


Figure 4.5 A pdf and associated cdf

Moreover:

$$f(x) = F'(x)$$

Example 2

Problem

Let $\{X_1, X_2\}$ *be a random sample of size 2 from the exponential distribution with parameter* λ

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

and $T = X_1 + X_2$.

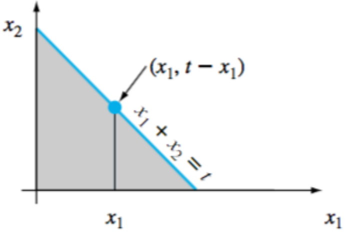
1. *Compute the cumulative density function (cdf) of* T

Example 2

$$F_{T_o}(t) = P(X_1 + X_2 \leq t) = \iint\limits_{\{(x_1, x_2): x_1 + x_2 \leq t\}} f(x_1, x_2)\, dx_1\, dx_2$$

$$= \int_0^t \int_0^{t-x_1} \lambda e^{-\lambda x_1} \cdot \lambda e^{-\lambda x_2}\, dx_2\, dx_1 = \int_0^t (\lambda e^{-\lambda x_1} - \lambda e^{-\lambda t})\, dx_1$$

$$= 1 - e^{-\lambda t} - \lambda t e^{-\lambda t}$$



Example 2b

Problem

Let $\{X_1, X_2\}$ *be a random sample of size 2 from the exponential distribution with parameter* $\lambda = 2$

$$f(x) = \begin{cases} 2e^{-2x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

and $T = X_1 + X_2$.

1. *Compute the cumulative density function (cdf) of* T
2. *Compute the probability density function (pdf) of* T

Summary

1. If the distribution and the statistic T is simple, try to construct the pmf of the statistic (as in Example 1)
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)

Example 1*

Problem

Consider the distribution P

x	40	45	50
$p(x)$	0.2	0.3	0.5

Let $\{X_1, X_2\}$ *be a random sample of size 2 from* P, *and* $T = X_1 - X_2$.

1. *Derive the probability mass function of T*
2. *Compute the expected value and the standard deviation of T*

Example 2*

Problem

Let $\{X_1, X_2\}$ *be a random sample of size 2 from the exponential distribution with parameter* λ

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

and $T = X_1 + 2X_2$.

1. *Compute the cumulative density function (cdf) of* T
2. *Compute the probability density function (pdf) of* T

Linear combination of normal random variables

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

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

also follows the normal distribution.

What are the mean and the standard deviation of T?

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

Moment generating function

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

Moment generating function

Property

Two distributions have the same pdf if and only if they have the same moment generating function

Moment generating function

Distribution	Moment-generating function $M_X(t)$
Bernoulli $P(X=1)=p$	$1-p+pe^t$
Geometric $(1-p)^{k-1}\,p$	$\dfrac{pe^t}{1-(1-p)e^t}$ $\forall t<-\ln(1-p)$
Binomial B(n, p)	$\left(1-p+pe^t\right)^n$
Poisson Pois(λ)	$e^{\lambda(e^t-1)}$
Uniform (continuous) U(a, b)	$\dfrac{e^{tb}-e^{ta}}{t(b-a)}$
Uniform (discrete) U(a, b)	$\dfrac{e^{at}-e^{(b+1)t}}{(b-a+1)(1-e^t)}$
Normal $N(\mu, \sigma^2)$	$e^{t\mu+\frac{1}{2}\sigma^2t^2}$
Chi-squared χ^2_k	$(1-2t)^{-\frac{k}{2}}$
Gamma $\Gamma(k, \theta)$	$(1-t\theta)^{-k}; \forall t<\frac{1}{\theta}$
Exponential Exp(λ)	$\left(1-t\lambda^{-1}\right)^{-1},\ (t<\lambda)$
[illegible]	[illegible]

Moment generating function

Definition

Let X_1, X_2 be a 2 independent random variables and $T = X_1 + X_2$, then

$$M_T(t) = M_{X_1}(t)M_{X_2}(t)$$

Hint:

$$M_T(t) = E(e^{tT}) = E(e^{t(X_1+X_2)}) = E(e^{tX_1} \cdot e^{tX_2})$$

Example 3

Problem

Given that the mgf of a Poisson variables with mean λ is

$$e^{\lambda(e^t-1)}$$

Suppose X and Y are independent Poisson random variables, where X has mean a and Y has mean b. Show that $T = X + Y$ also follows the Poisson distribution.