

MATH 324: Statistics

Julian Lore

Last updated: January 26, 2018

Rough notes from Wackerly's Mathematical Statistics with Applications (7th edition).

Contents

8	Estimation	2
8.1	Introduction	2
8.2	The Bias and Mean Square Error of Point Estimators	3
8.3	Some Common Unbiased Point Estimators	4
8.4	Evaluating the Goodness of a Point Estimator	4

Information from Previous Chapters

Some formulas from probability:

$$Var(X) = E[(X - \mu)^2] = E[X^2] - (E[X])^2$$

Distributions:

Distribution	Probability Function	Mean	Variance	Moment-Generating Function
Binomial	$p(y) = \binom{n}{y} p^y (1-p)^{n-y}$	np	$np(1-p)$	$[pe^t + (1-p)]^n$

6.7 Order Statistics

Y_i , with $i = 1, \dots, n$ independent continuous random variables. Denote ordered random variables by: $Y_{(1)} \leq Y_{(2)} \leq \dots \leq Y_{(n)}$. We can then define:

$$Y_{(1)} = \min(Y_1, Y_2, \dots, Y_n)$$

$$Y_{(n)} = \max(Y_1, Y_2, \dots, Y_n)$$

In order to find the pdf of $Y_{(1)}$ and $Y_{(n)}$, we use the method of distribution functions.

$$\begin{aligned} F_{Y_{(n)}}(y) &= P(Y_{(n)} \leq y) = P(Y_1 \leq y, Y_2 \leq y, \dots, Y_n \leq y) \\ &\stackrel{\text{ind}}{=} P(Y_1 \leq y)P(Y_2 \leq y) \dots P(Y_n \leq y) = [F(y)]^n \end{aligned}$$

Derive for the density:

$$g_{(n)}(y) = n[F(y)]^{n-1}f(y)$$

$$\begin{aligned} F_{Y_{(1)}}(y) &= P(Y_{(1)} \leq y) = 1 - P(Y_{(1)} > y) = 1 - P(Y_1 > y, Y_2 > y, \dots, Y_n > y) \\ &\stackrel{\text{ind}}{=} 1 - [P(Y_1 > y)P(Y_2 > y) \dots P(Y_n > y)] = 1 - [1 - F(y)]^n \end{aligned}$$

Derive for the density:

$$g_{(1)}(y) = n[1 - F(y)]^{n-1}f(y)$$

Other Useful Information

$$E(\bar{Y}) = E(Y)$$

8 Estimation

8.1 Introduction

Point of statistics is to use sample information to infer data about the population. Populations are characterized by numbers (*parameters*) and we often want to estimate the value of parameter(s). Parameters include the proportion p , population mean μ , variance σ^2 and standard deviation σ .

Definition 1. The parameter of interest in an experiment is called the *target parameter*.

Definition 2. A *point estimate* is a type of estimate where we use a single value/point to estimate a parameter. If we estimate a parameter by saying that it might fall between two numbers, this is an *interval estimate*. We can use information from the sample to calculate these estimates, which are done using an estimator.

Definition 3. An *estimator* is a rule, often expressed as a formula, that tells how to calculate the value of an estimate based on the measurements contained in a sample.

Definition 4. *Sample mean:*

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

This is an example point estimator of μ .

There can be different estimators for the same population parameter. Some estimators are considered good and others are bad.

8.2 The Bias and Mean Square Error of Point Estimators

We cannot measure how good a point estimation procedure is with a single estimate, we need to observe the procedure many times. We create a frequency distribution to measure the goodness of a point estimator.

Point Estimators For a population parameter θ , the estimator of θ is called $\hat{\theta}$.

Definition 5. Ideally, we'd want $E(\hat{\theta}) = \theta$. Point estimators that satisfy this are called *unbiased*. Otherwise, they are called *biased*, where the *bias* is given by $B(\hat{\theta}) = E(\hat{\theta}) - \theta$

In addition, we'd also like the estimator $V(\hat{\theta})$ to be as small as possible, since a smaller variance guarantees a higher fraction of estimators to be “close” to θ . If two estimators are unbiased and everything else is equal other than variance, we prefer the one with smaller variance.

Definition 6. Another way to characterize goodness of a point estimator is via its *mean square error*,

$$MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2]$$

Which is the average of the square of the distance between the estimator and its target parameter. It can be shown that:

$$MSE(\hat{\theta}) = V(\hat{\theta}) + [B(\hat{\theta})]^2$$

8.3 Some Common Unbiased Point Estimators

Definition 7. $\sigma_{\hat{\theta}}^2$ denotes the variance of the sampling distribution of the estimator $\hat{\theta}$. The standard deviation $\sigma_{\hat{\theta}} = \sqrt{\sigma_{\hat{\theta}}^2}$ is called the *standard error* of the estimator $\hat{\theta}$.

Common Point Estimators If random samples are independent we have:

Target Parameter θ	Sample Size(s)	Point Estimator $\hat{\theta}$	$E(\hat{\theta})$	Standard Error $\sigma_{\hat{\theta}}$
μ	n	\bar{Y}	μ	$\frac{\sigma}{\sqrt{n}}$
p	n	$\hat{p} = \frac{Y}{n}$	p	$\sqrt{\frac{pq}{n}}$
$\mu_1 - \mu_2$	n_1, n_2	$\bar{Y}_1 - \bar{Y}_2$	$\mu_1 - \mu_2$	$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
$p_1 - p_2$	n_1, n_2	$\hat{p}_1 - \hat{p}_2$	$p_1 - p_2$	$\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$

Note, these are all **unbiased** estimators. Note that the expected values and standard errors for \bar{Y} and $\bar{Y}_1 - \bar{Y}_2$ are valid for any distribution of the population and all four estimators are approximately normal for large n (see CLT).

Definition 8. The sample variances,

$$S^2 = \frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n - 1}$$

$$S'^2 = \frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n}$$

are unbiased (S^2) and biased (S'^2) estimators for σ^2 .

8.4 Evaluating the Goodness of a Point Estimator

Definition 9. The *error of estimation* ε is the distance between an estimator and its target parameter:

$$\varepsilon = |\hat{\theta} - \theta|$$

Note that $\hat{\theta}$ is a random variable, so the error of estimation is a random quantity and thus we can make probability statements about it (rather than saying exactly how big it is for some estimate).

The probability $P(\varepsilon < b)$ represents the fraction of times, in repeated sampling that $\hat{\theta}$ falls within b units of θ .

Sometimes we want to find a value of b such that $P(\varepsilon < b) = \text{some value}$, say 0.90.

If $\hat{\theta}$ is unbiased we can find a bound on ε by expressing b as a multiple of the standard error. Take for example, $b = k\sigma_{\hat{\theta}}$, where $k \geq 1$, then by Tchebysheff's we know that

$P(\varepsilon < k\sigma_{\hat{\theta}}) \geq 1 - \frac{1}{k^2}$. $b = 2\sigma_{\hat{\theta}}$ is a good approximate bound on ε in practice. By Tchebysheff we know that $P(\varepsilon < 2\sigma_{\hat{\theta}}) \geq .75$.