Virtual Laboratories > 7. Point Estimation > 1 2 3 4 5 6

1. Estimators

The Basic Statistical Model

As usual, our starting point is a random experiment with an underlying sample space and a probability measure $\mathbb{P}$. In the basic statistical model, we have an observable random variable $\boldsymbol{X}$ taking values in a set S. Recall that in general, this variable can have quite a complicated structure. For example, if the experiment is to sample n objects from a population and record various measurements of interest, then the data vector has the form

$$\boldsymbol{X} = (X_1, X_2, ..., X_n)$$

where X_i is the vector of measurements for the i^{th} object. The most important special case is when $(X_1, X_2, ..., X_n)$ are independent and identically distributed (IID). In this case $\boldsymbol{X}$ is a random sample of size n from the distribution of an underlying measurement variable X.

Statistics

Recall also that a **statistic** is an observable function of the outcome variable of the random experiment:

$$\boldsymbol{W} = \boldsymbol{W}(\boldsymbol{X})$$

Thus, a statistic is simply a random variable derived from the observation variable $\boldsymbol{X}$, with the assumption that $\boldsymbol{W}$ is also observable. As the notation indicates, $\boldsymbol{W}$ is typically also vector-valued.

Parameters

In the general sense, a **parameter** $\boldsymbol{\theta}$ is a function of the distribution of $\boldsymbol{X}$, taking values in a **parameter space** Θ. Typically, the distribution of $\boldsymbol{X}$ will have k real parameters of interest, so that $\boldsymbol{\theta}$ has the form

$$\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_k)$$

and thus $\Theta \subseteq \mathbb{R}^k$. In many cases, one or more of the parameters are unknown, and must be estimated from the data variable $\boldsymbol{X}$. This is one of the of the most important and basic of all statistical problems, and is the subject of this chapter.

Estimators

Suppose now that we have an unknown real parameter θ taking values in a parameter space $\Theta \subseteq \mathbb{R}$. A real-valued statistic $W = W(\boldsymbol{X})$ that is used to estimate θ is called, appropriately enough, an **estimator** of θ.

Thus, the estimator is a random variable and hence has a distribution, a mean, a variance, and so on. When we actually run the experiment and observe the data $\boldsymbol{x}$, the observed value $w = W(\boldsymbol{x})$ (a single number) is the **estimate** of the parameter θ.

Basic Properties

The (random) **error** is the difference between the estimator and the parameter: $W - \theta$. The expected value of the error is known as the **bias**:

$$\operatorname{bias}(W) = \mathbb{E}(W - \theta)$$

1. Use basic properties of expected value to show that $\operatorname{bias}(W) = \mathbb{E}(W) - \theta$

Thus, the estimator is said to be **unbiased** if the bias is 0 for all $\theta \in \Theta$, equivalently if the expected value of the estimator is the parameter being estimated: $\mathbb{E}(W) = \theta$ for all $\theta \in \Theta$. The quality of the estimator is usually measured by computing the **mean square error**:

$$\operatorname{MSE}(W) = \mathbb{E}\left((W - \theta)^2\right)$$

2. Use basic properties of expected value and variance to show that

$$\operatorname{MSE}(W) = \operatorname{var}(W) + \operatorname{bias}^2(W)$$

In particular, if the estimator is unbiased, then the mean square error of W is simply the variance of W.

Ideally, we would like to have unbiased estimators with small mean square error. However, this is not always possible, and Exercise 2 shows the delicate relationship between bias and mean square error. In the next section we will see an example with two estimators of a parameter that are multiples of each other; one is unbiased, but the other has smaller mean square error. However, if we have two unbiased estimators of θ, denoted U and V, we naturally prefer the one with the smaller variance (mean square error). The **relative efficiency** of V to U is simply the ratio of the variances:

$$\operatorname{eff}(U, V) = \frac{\operatorname{var}(U)}{\operatorname{var}(V)}$$

Asymptotic Properties

Often we have a general formula that defines an estimator of θ for any sample size n. Technically, this gives a *sequence* of real-valued estimators of θ:

$$W_n = W_n(X_1, X_2, \ldots, X_n), \quad n \in \mathbb{N}_+$$

In this case, we can discuss the asymptotic properties of the estimators as $n \to \infty$. Most of the definitions are natural generalizations of the ones above. First, the sequence of estimators W_n is said to be **asymptotically unbiased** if

$$\text{bias}(W_n) \to 0 \text{ as } n \to \infty \text{ for any } \theta \in \Theta$$

3. Show that W_n is asymptotically unbiased if and only if $\mathbb{E}(W_n) \to \theta$ as $n \to \infty$ for any $\theta \in \Theta$.

Suppose now that U_n and V_n are two sequences of estimators that are asymptotically unbiased for θ. The **asymptotic relative efficiency** of V_n to U_n is the following limit, if it exists:

$$\lim_{n \to \infty} \frac{\text{var}(U_n)}{\text{var}(V_n)}$$

Naturally, we expect our estimators to improve, in some sense, as the sample size n increases. Specifically, the sequence of estimators W_n is said to be **consistent** for θ if $W_n \to \theta$ as $n \to \infty$ in probability.:

$$\mathbb{P}(|W_n - \theta| > \varepsilon) \to 0 \text{ as } n \to \infty \text{ for every } \varepsilon > 0 \text{ and every } \theta \in \Theta$$

4. Suppose that $\text{MSE}(W_n) \to 0$ as $n \to \infty$ for any $\theta \in \Theta$. Show that W_n is consistent for θ. *Hint*: Use Markov's inequality.

The condition in Exercise 4 is known as **mean-square consistency**. Thus, mean-square consistency implies simple consistency. This is simply a statistical version of the theorem that states that mean-square convergence implies convergence in probability.

Estimation Problems

In the next several subsections, we will review several basic estimation problems that were studied in the chapter on Random Samples.

Estimating the Mean

Suppose that $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ is a random sample of size n from the distribution of a real-valued random variable X that has mean μ and standard deviation σ. A natural estimator of the distribution mean μ is the sample mean, defined by

$$M(\boldsymbol{X}) = \frac{1}{n} \sum_{i=1}^{n} X_i$$

5. Show or recall that

a. $\mathbb{E}(M) = \mu$ so M is an unbiased estimator of μ.
b. $\text{var}(M) = \frac{\sigma^2}{n}$ so M is a consistent estimator of μ.

6. In the sample mean experiment, set the sampling distribution to gamma. Increase the sample size with the scroll bar and note graphically and numerically the unbiased and consistent properties. Run the experiment 1000 times updating every 10.

7. Run the normal estimation experiment 1000 times, updating every 10 runs, for several values of the parameters. In each case, compare the empirical bias and mean square error of M_n with the theoretical values.

The consistency of the sample mean M_n as an estimator of the distribution mean μ is simply the weak law of large numbers. Moreover, there are a number of important special cases of the results in Exercise 5. See the section on Sample Mean for the details.

- Suppose that $X = \mathbf{1}(A)$, the indicator variable for an event A that has probability $\mathbb{P}(A)$. Then the sample mean of the random sample $\boldsymbol{X}$ is the relative frequency or empirical probability of A, denoted $P_n(A)$. Hence $P_n(A)$ is an unbiased and consistent estimator of $\mathbb{P}(A)$.
- Suppose that F denotes the distribution function of a real-valued random variable X. Then for fixed x, the empirical distribution function $F_n(x)$ is simply the sample mean for a random sample of size n from the distribution of the indicator variable $\mathbf{1}(X \le x)$. Hence $F_n(x)$ is an unbiased and consistent estimator of $F(x)$.
- Suppose that X is a random variable with a discrete distribution on a countable set S and f denotes the probability density function of X. Then for fixed $x \in S$, the empirical probability density function $f_n(x)$ is simply the sample mean for a random sample of size n from the distribution of the indicator variable $\mathbf{1}(X = x)$. Hence $f_n(x)$ is an unbiased and consistent estimator of $f(x)$.

8. In matching experiment, the random variable is the number of matches. Run the simulation 1000 times updating every 10 runs and note the apparent convergence of

a. the sample mean to the distribution mean.
b. the empirical density function to the distribution density function.

Estimating the Variance

As in the last subsection, suppose that $\boldsymbol{X} = (X_1, X_2, ..., X_n)$ is a random sample of size n from the distribution of a real-valued random variable X that has mean μ and standard deviation σ. We will also assume that the fourth central moment $d_4 = \mathbb{E}\left((X - \mu)^4\right)$ is finite.

If μ is known (usually an artificial assumption), then a natural estimator of σ^2 is a special version of the sample variance, defined by

$$W^2(\boldsymbol{X}) = \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2$$

9. Show or recall that

a. $\mathbb{E}(W^2) = \sigma^2$ so W^2 is an unbiased estimator of σ^2
b. $\text{var}(W^2) = \frac{1}{n}(d_4 - \sigma^4)$ so W^2 is a consistent estimator of σ^2

If μ is unknown (the more reasonable assumption), then a natural estimator of the distribution variance is the standard version of the sample variance, defined by

$$S^2(\boldsymbol{X}) = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - M(\boldsymbol{X}))^2$$

10. Show or recall that

a. $\mathbb{E}(S^2) = \sigma^2$ so S^2 is an unbiased estimator of σ^2
b. $\text{var}(S^2) = \frac{1}{n}\left(d_4 - \frac{n-3}{n-1}\sigma^4\right)$ so S^2 is a consistent estimator of σ^2

11. Run the exponential experiment 1000 times with an update frequency of 10. Note the apparent convergence of the sample standard deviation to the distribution standard deviation.

12. Show that

a. $\text{var}(W^2) < \text{var}(S^2)$. Thus, W^2 is better than S^2, assuming that μ is known so that we can actually *use* W^2.
b. The asymptotic relative efficiency of S^2 to W^2 is 1.

13. Run the normal estimation experiment 1000 times, updating every 10 runs, for several values of the parameters. In each case, compare the empirical bias and mean square error of S^2 and of W^2 to their theoretical values. Which estimator seems to work better?

The Poisson Distribution

For an example of the ideas in the last two subsections, suppose that X has the Poisson distribution with unknown parameter $a > 0$. Then $\mathbb{E}(X) = \text{var}(X) = a$, so that we could use either the sample mean M or the sample variance S^2 as an estimator of a. Both are unbiased, so which is better? Naturally, we use mean square error as our criterion.

14. Show that

a. $\mathbb{E}(X) = a$
b. $\mathbb{E}(X^2) = a^2 + a$
c. $\mathbb{E}(X^3) = a^3 + 3a^2 + a$

d. $\mathbb{E}(X^4) = a^4 + 6a^3 + 7a^2 + a$
e. $d_4 = 3a^2 + a$

15. Show that

a. $\text{var}(M) = \frac{a}{n}$
b. $\text{var}(S^2) = \frac{a}{n}\left(1 + 2a\frac{n}{n-1}\right)$
c. $\text{var}(M) < \text{var}(S^2)$, so the sample mean M is a better estimator of the parameter a than the sample variance S^2.
d. The asymptotic relative efficiency of M to S^2 is $1 + 2a$

16. Run the Poisson experiment 100 times, updating every run, for several values of the parameter. In each case, compute the estimators M and S^2. Which estimator seems to work better?

Estimating the Covariance

Suppose that $((X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n))$ is a random sample of size n from the distribution of (X, Y), where X is a real-valued random variable with mean μ and standard deviation σ, and where Y is a real-valued random variable with mean ν and standard deviation τ. Let δ denote the covariance of (X, Y). As usual, we will let $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ and $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_n)$; these are random samples of size n from the distributions of X and Y, respectively.

If μ and ν are known (usually an artificial assumption), then a natural estimator of the distribution covariance δ is a special version of the sample covariance, defined by

$$W(\boldsymbol{X}, \boldsymbol{Y}) = \frac{1}{n}\sum_{i=1}^{n} (X_i - \mu)(Y_i - \nu)$$

17. Show or recall that

a. $\mathbb{E}(W) = \delta$ so W is an unbiased estimator of δ.
b. W is consistent estimator of δ.

If μ and ν are unknown (usually the more reasonable assumption), then a natural estimator of the distribution covariance δ is the usual version of the sample covariance, defined by

$$S(\boldsymbol{X}, \boldsymbol{Y}) = \frac{1}{n-1}\sum_{i=1}^{n} (X_i - M(\boldsymbol{X}))(Y_i - M(\boldsymbol{Y}))$$

18. Show or recall that

a. $\mathbb{E}(S) = \delta$ so S is an unbiased estimator of δ.
b. S is consistent estimator of δ.

Chapter Topics

The estimators of the mean, variance, and covariance that we have considered in this section have been natural in a sense. However, for other parameters, it is not clear how to even find a reasonable estimator in the first place. In the next several sections, we will consider the problem of constructing estimators. Then we return to the study of the mathematical properties of estimators, and consider the question of when we can know that an estimator is the best possible, given the data.

Virtual Laboratories > 7. Point Estimation > 1 2 3 4 5 6
Contents | Applets | Data Sets | Biographies | External Resources | Keywords | Feedback | ©