Properties of Estimators

We study estimators as random variables. In this setting we suppose $X_1, X_2, \ldots, X_n$ are random variables observed from a statistical model $\mathcal{F}$ with parameter space Θ.

In our usual setting we also then assume that X_i are iid with pdf (or pmf) $f(\cdot;\theta)$ for some $\theta \in \Theta$. Also in our usual setting $\Theta \subset R^d$ for some finite d, that is a finite dimensional parameter model. In this case then $X_1, X_2, \ldots, X_n$ has joint pdf (or pmf) given by the function

$$f_n(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f(x_i;\theta) \ .$$

In our statistical inference setting the specific value of the parameter θ is not known and is the object to be estimated form observable data. If the statistical model is *correct* then there is one special value of the parameter that is the *true* value of the parameter, say θ_0, so that

$$f_n(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f(x_i;\theta_0) \ .$$

Aside : Unless we need this extra notation we do not use the subscript 0 to designate the true value θ_0 of the parameter.

Definition 1 *Consider an experiment with random data (r.v.s) $X_1, X_2, \ldots, X_n$ from a statistical model $\mathcal{F}$ and parameter space Θ. Consider a random variable*

$$T = h(X_1, X_2, \ldots, X_n)$$

for some function h. We say T is a statistic if T can be calculated from the observable data only, and does not require knowing which parameter value θ is the ***true*** *value of the parameter.*

Examples of Statistics: X_i are iid from a distribution f. In some the examples below we need $n \geq 2$.

The following are statistics

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

2. $$\hat{\mu}_{k,n} = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

3.
$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X}_n\right)^2$$

4. If $P(X_i > 0) = 1$, let
$$T_n = \frac{1}{n} \sum_{i=1}^{n} \log(X_i)$$
$$W_n = \frac{1}{\bar{X}_n}$$

5.
$$T_n = \prod_{i=1}^{n} X_i$$

6.
$$\text{median}(X_1, X_2, \ldots, X_n)$$

End of Example

Example of a RV that is not a statistic

The following is not a statistic : Suppose X_i are iid exponential, parameter λ. Notice that the population median is given by x which solves
$$F(x) = \frac{1}{2}$$
Therefore x solves
$$\frac{1}{2} = 1 - e^{-\lambda x}$$
and hence $x = \frac{1}{\lambda}\log(2)$. For convenience of notation let m be the population median. Notice it changes depending on the value of the parameter.

Let $T =$ the number of the random variables X_i greater than the median m.

T is not a statistic, since its value cannot be calculated from the observed data $x_1, x_2, \ldots, x_n$ from an experiment, without knowing the value of the parameter (or number in this example) λ. However it is a random variable, just not a statistic.

Notice that if X_i are iid exponential λ then the distribution of T is easy to obtain. It is Binomial$(n, \frac{1}{2})$. This is because
$$T = \sum_{i=1}^{n} \mathrm{I}(X_i > m)$$
and $\mathrm{I}(X_i > m)$ are iid Bernoulli($\frac{1}{2}$).

End of Example

Estimators of a parameter θ are of the form $\hat{\theta}_n = T(X_1, \ldots, X_n)$ so it is a function of r.v.s $X_1, \ldots, X_n$ and is a statistic. Hence an estimator is a r.v. As such it has a distribution. This distribution of course is determined the distribution of $X_1, \ldots, X_n$. If these r.v.s are iid, say distribution o$f(\cdot;\theta_0)$, then the distribution of $\hat{\theta}_n$ is determined by the distribution $f(\cdot,\theta_0)$. Here we are using θ_0 to represent the true parameter value $\theta\ in \Theta$, the parameter space. Generally in this course we will just use θ instead of θ_0.

Sometimes we also wish to find an estimator of a function of a parameter. For example for a Poisson, parameter λ, we might wish to estimate λ, be we also might wish to estimate a function of λ such as $h(\lambda) = e^{-\lambda}$, which happens to be the probability of a Poisson r.v. taking the value 0.

Terminology. Since T is a function, it gives a number for given values $x_1, \ldots, x_n$. We refer to the r.v. $\hat{\theta}_n = T(X_1, \ldots, X_n)$ as an *estimator* and refer to the observed value $\hat{\theta}_n = T(x_1, \ldots, x_n)$ (for observed data $x_1, \ldots, x_n$) as the *estimate* or the observed value of the estimator. This is the same role as in earlier courses where we refer to Y as a random variable and also need to consider the observed value y from a given experiment. We also refer to an estimator as an *estimator of* θ when this estimator is chosen for the purpose of estimating a parameter θ. In principle any statistic can be used to estimate any parameter, or a function of the parameter, although in general these would not be good estimators of some parameters. For example the sample variance is generally not a very estimator of the population mean parameter.

The distribution of the estimator, that is the distribution of $\hat{\theta}_n = T(X_1, \ldots, X_n)$ will play a key role in statistical inference.

Sample Variance

Suppose $X_i, i = 1, \ldots, n$ are iid r.v.s with finite mean μ and finite variance σ^2.

Consider the statistic

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2$$

We now calculate $\mathrm{E}\left(\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2\right)$.

$$\mathrm{E}\left(\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2\right) = \sum_{i=1}^{n}\mathrm{E}\left[\left(X_i - \bar{X}_n\right)^2\right]$$

For a given i

$$\begin{aligned}
\mathrm{E}\left[\left(X_i - \bar{X}_n\right)^2\right] &= \mathrm{E}\left[\left(X_i - \mu - (\bar{X}_n - \mu)\right)^2\right] \\
&= \mathrm{E}\left[(X_i - \mu)^2\right] - 2\mathrm{E}\left[(X_i - \mu)(\bar{X}_n - \mu)\right] + \mathrm{E}\left[\left((\bar{X}_n - \mu)\right)^2\right] \\
&= \sigma^2 - \frac{2}{n}\sum_{j=1}^{n}\mathrm{E}\left[(X_i - \mu)(X_j - \mu)\right] + \frac{\sigma^2}{n} \\
&= \sigma^2 - \frac{2}{n}\sigma^2 + \frac{\sigma^2}{n} \\
&= \sigma^2\left(1 - \frac{1}{n}\right) \\
&= \sigma^2\frac{n-1}{n}\ .
\end{aligned}$$

Notice we are using the fact that independent r.v.s with finite variances have covariance equal to 0. We are also using linearity properties of expectation.

Thus we can now complete the calculation of the expected value of s_n^2

$$\begin{aligned}
\mathrm{E}(S_n^2) &= \frac{1}{n-1}\mathrm{E}\left(\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2\right) \\
&= \frac{1}{n-1}\sum_{i=1}^{n}\sigma^2\frac{n-1}{n} \\
&= \sigma^2
\end{aligned}$$

End of Example

Gaussian (or Normal) Example

Suppose that X_i, $i = 1, \ldots, n$ are iid $N(\mu, \sigma^2)$. The parameter space is

$$\Theta = \{(\theta_1, \theta_2) = (\mu, \sigma^2) \mid \theta_1 \in R\ ,\ \theta_2 > 0\ \}\ .$$

Consider the statistic

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2$$

$$\mathrm{E}_\theta(S_n^2) = \sigma^2\ .$$

The student should also verify that

$$\mathrm{E}_\theta(\bar{X}_n) = \mu$$

Remark : These two properties hold for any iid random variables with mean μ and variance σ^2, no just normal r.v.s. The student should verify this. There is a subscript θ for the expectation operator. This is used with the meaning that the expectation is with respect to the normal distribution with this particular parameter θ, that corresponds to the parameter value for the (specific) distribution of the X_i's.

To help to clarify this, if $X \sim N(2, 9)$ then $E(X) = \mu = 2$, $E(X^2) = \sigma^2 + \mu^2 = 13$, and these expectations are not given by any other value of $\theta = \theta \neq (2, 9)$.

Since this property in our example holds for all θ we say that $\bar{X}_n$ is an unbiased estimator of the *parameter* μ. To be more precise it is an unbiased estimator of $\mu = h(\theta) = h(\mu, \sigma^2)$ where h is the function that maps the pair of arguments to the first element of this pair, that is $h(x, y) = x$.

Similarly S_n^2 is an unbiased estimator of σ^2.

End of Example

The notation E_θ means to calculate the expectation with respect to the distribution with parameter value θ. Notice that the expectation is different for each θ. We use this notation to emphasize the dependence of these calculations on the *true* parameter θ. When this is not needed then typically we do not use the subscript.

Definition 2 (Unbiased Estimator) *Consider a statistical model. Let T be a statistic. T is said to be an unbiased estimator of θ if and only if $E_\theta(T) = \theta$ for all θ in the parameter space.*

More generally we say T is an unbiased estimator of $h(\theta)$ if and only if $E_\theta(T) = h(\theta)$ for all θ in the parameter space.

This property may apply only to an estimator of one of the components of vector valued parameter. Thus we also want to consider the part of the definition h, so for example

$$\mu = h(\mu, \sigma^2)$$

where $h(x, y) = x$.

Notice that the property of unbiasedness is a property determined by the distribution of the statistic T and the statistical model. The distribution of the statistic T, which is a function of the observable r.v.'s, is one of the main topics of study in our previous course. It is referred to as the sampling distribution of T and is obtained for each possible parameter θ in the parameter space Θ. Sometimes these properties can be verified generally, for example sample means. Other times it requires specific and special properties of the sampling distribution of T; see for example the Poisson example below and the estimator of $e^{-\theta}$.

Suppose X_i are iid from a distribution with finite k-th moment, that is $\mu_k = \mathrm{E}(X^k)$ is finite.

For any positive integer k consider the sample k-th

$$\hat{\mu}_{k,n} = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

Verify that $\mathrm{E}(\hat{\mu}_{k,n}) = \mu_k$. Thus for any statistical model, and appropriate finite moments, we obtain for iid samples that sample k-th moments are unbiased estimators of the population k-th moments.

Since $\mathrm{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2$, one might consider as an estimator of population variance

$$\widehat{\mathrm{Var}}(X) = \hat{\mu}_{2,n} - (\hat{\mu}_{1,n})^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}_n^2$$

Show this estimator is not unbiased (show it is biased) for $\sigma^2 = \mathrm{Var}(X)$. How can you modify this estimator to produce a new estimator that is an unbiased estimator of σ^2?

Aside : Calculate the expectation. You will find that you can multiply this by a constant, which does involve n, to obtain σ^2. Call this constant a_n. Then the estimator

$$\frac{1}{a_n}\widehat{\mathrm{Var}}(X)$$

will have expectation σ^2 and hence be an unbiased estimator of σ^2. This method uses the linearity of expectation, and does not work very generally.

Poisson Example

Consider the Poisson model with parameter space $\Theta = R^+ = (0, \infty)$. Suppose X_i are iid Poisson with parameter θ. $\bar{X}_n$ is an unbiased estimator of θ.

There are various functions of θ that are also interesting. For $X \sim \text{Poisson}, \theta$, we have $P_\theta(X = 0) = e^{-\theta}$. If we consider the function $h(\theta) = e^{-\theta}$ and the statistic $T_n = e^{-\bar{X}_n}$, is T_n unbiased for $h(\theta)$?

To answer this we need to calculate $\mathrm{E}_\theta(T_n)$ and determine if this is equal to $e^{-\theta}$ for every possible $\theta \in \Theta$. In general this may not be easy to calculate but we can take advantage of some special properties of sums of iid Poisson random variables.

Let $Y = \sum_{i=1}^n X_i$. Using MGFs we can then determine that $Y \sim \text{Poisson}, n\theta$. Again using MGFs we then for any number t

$$\mathrm{E}_\theta(e^{tY}) = M_Y(t) = e^{n\theta(e^t - 1)}$$

Using this we then note that

$$E_\theta(e^{-\bar{X}_n}) = E_\theta(e^{-\frac{1}{n}Y}) = M_Y(-\frac{1}{n})$$

Thus

$$E_\theta(e^{-\bar{X}_n}) = \exp\left\{n\theta\left(e^{-\frac{1}{n}} - 1\right)\right\} \ .$$

Notice that this is not equal to $e^{-\theta}$. Therefore for any n the random variable T_n is not an unbiased estimator of $e^{-\theta}$.

This example is interesting in that we may have an unbiased estimator of a parameter but functions of this estimator might not be unbiased estimators of the corresponding function of the parameter.

End of Example

Terminology [This is also given earlier in this note.] An estimator is a random variable that is a statistic which is a function of the observable data, that is of the form

$$T = h(X_1, X_2, \ldots, X_n) \ .$$

We study the properties of the random variable T and use these to make *statistical inference* rules or procedures. On the other hand, after we have conducted an experiment or observational study we have observed data $x_1, x_2, \ldots, x_n$. In this case we have an observed value of T, namely $h(x_1, x_2, \ldots, x_n)$. This observed value of T is called the *estimate* or *observed value of the estimator*. The distinction is that the *estimator* is a random variable and the *estimate* is the observed value of the random variable.

Definition 3 (Consistent Estimator) *Consider a statistical model. Let T_n be an estimator of θ. (more specifically T_n is a sequence of estimators indexed by n the sample size.)*

We say that T_n (more specifically the sequence T_n, $n \geq 1$) is a consistent estimator of θ iff $T_n \to \theta$ in probability as $n \to \infty$.

Recall the definition of convergence in probability to a constant. Thus an estimator is consistent for θ iff and only if for every $\epsilon > 0$

$$P(|T_n - \theta| > \epsilon) \to 0$$

as $n \to \infty$.

Comments

1. In our earlier courses we studied a result of this type, namely convergence in probability. In the case of a sequence of iid r.v.s X_i, $i \geq 1$, we used the Law of Large Number to show, under some conditions (the student should review the LLN and these conditions) to show

   $$\frac{1}{n}\sum_{i=1}^{n} X_i \to \mathrm{E}(X) = \mu$$

   in probability as $n \to \infty$.

2. This property of estimators is perhaps one of the most important properties. If an estimator does not have this property it is not a very good estimator.

   An estimator with this property is guaranteed to be close to the population parameter θ provided there is a big enough sample size.

   A related notion is of course to know something about how big the sample size should be.

3. Sample moments are consistent estimators of the corresponding population moments.

   A method of moments estimator for the (vector valued) parameter θ is of the form

   $$\hat{\theta}_n = g(\hat{\mu}_{1,n}, \ldots, \hat{\mu}_{K,n})$$

   where g is a continuous function.

   It is then the case that

   $$\hat{\theta}_n = g(\hat{\mu}_{1,n}, \ldots, \hat{\mu}_{K,n}) \to g(\mu_1, \ldots, \mu_K) = \theta$$

   where the convergence is in probability as $n \to \infty$. Recall that g is chosen so that $g(\mu_1, \ldots, \mu_K) = \theta$ holds.

   This property of convergence is studied in Problem 5.4.7 in Rice, one of the problems from Stat 3657.

Can one obtain estimators that are not consistent? This is actually quite easy, but generally these are not sensible estimators. As an example consider the following :

Consider $X_i, i \geq 1$ iid with mean μ and finite variance $\sigma^2 > 0$. This last part is just to rule out a simple trivial example of $X_i = \mu$ with probability 1. Also to rule out another non interesting trivial case we suppose that X has a positive density in the neighbourhood of μ. For example and to be specific we can take the example of X_i iid exponential mean μ.

$X_i, i = 1, \ldots, n$ is the sample of the first n of these. Let

$$T_n(X_1, \ldots, X_n) = X_n$$

that is the last of these n r.v.s. Then

$$\mathrm{E}(T_n) = \mathrm{E}(X_n) = \mu$$

so this is an unbiased estimator. However, except in the trivial case when X is constant with probability 1, ie $\mathrm{Var}(X) = 0$,

$$P(|T_n - \mu| > \epsilon) = P(|X_1 - \mu| > \epsilon) > 0$$

for any n greater than or equal to 1. (Technical aside : If X is a continuous r.v. this will hold for all $\epsilon > 0$. If X is discrete this will hold, but with some minor constraints on $\epsilon > 0$.) Thus this probability is the same positive number for every n and hence the sequence of numbers $P(|T_n - \mu| > \epsilon)$ cannot converge to 0.

This example, while not particularly interesting gives a sequence of unbiased estimators that are not consistent.