

7.1: 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 \mathbf{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

$$\mathbf{X} = (X_1, X_2, \dots, X_n) \quad (7.1.1)$$

where X_i is the vector of measurements for the i th object. The most important special case is when (X_1, X_2, \dots, X_n) are independent and identically distributed (IID). In this case \mathbf{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: $\mathbf{U} = \mathbf{u}(\mathbf{X})$ where \mathbf{u} is a known function from S into another set T . Thus, a statistic is simply a random variable derived from the observation variable \mathbf{X} , with the assumption that \mathbf{U} is also observable. As the notation indicates, \mathbf{U} is typically also vector-valued. Note that the original data vector \mathbf{X} is itself a statistic, but usually we are interested in statistics derived from \mathbf{X} . A statistic \mathbf{U} may be computed to answer an inferential question. In this context, if the dimension of \mathbf{U} (as a vector) is smaller than the dimension of \mathbf{X} (as is usually the case), then we have achieved *data reduction*. Ideally, we would like to achieve significant data reduction with no loss of information about the inferential question at hand.

Parameters

In the technical sense, a *parameter* θ is a function of the *distribution* of \mathbf{X} , taking values in a *parameter space* T . Typically, the distribution of \mathbf{X} will have $k \in \mathbb{N}_+$ real parameters of interest, so that θ has the form $\theta = (\theta_1, \theta_2, \dots, \theta_k)$ and thus $T \subseteq \mathbb{R}^k$. In many cases, one or more of the parameters are unknown, and must be estimated from the data variable \mathbf{X} . This is one of the most important and basic of all statistical problems, and is the subject of this chapter. If \mathbf{U} is a statistic, then the distribution of \mathbf{U} will depend on the parameters of \mathbf{X} , and thus so will distributional constructs such as means, variances, covariances, probability density functions and so forth. We usually suppress this dependence notationally to keep our mathematical expressions from becoming too unwieldy, but it's very important to realize that the underlying dependence is present. Remember that the critical idea is that by observing a value \mathbf{u} of a statistic \mathbf{U} we (hopefully) gain information about the unknown parameters.

Estimators

Suppose now that we have an unknown real parameter θ taking values in a parameter space $T \subseteq \mathbb{R}$. A real-valued statistic $U = u(\mathbf{X})$ 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 (all of which, as noted above, will generally depend on θ). When we actually run the experiment and observe the data \mathbf{x} , the observed value $u = u(\mathbf{x})$ (a single number) is the *estimate* of the parameter θ . The following definitions are basic.

Suppose that U is a statistic used as an estimator of a parameter θ with values in $T \subseteq \mathbb{R}$. For $\theta \in T$,

1. $U - \theta$ is the *error*.
2. $\text{bias}(U) = E(U - \theta) = E(U) - \theta$ is the *bias* of U
3. $\text{mse}(U) = E[(U - \theta)^2]$ is the *mean square error* of U

Thus the error is the difference between the estimator and the parameter being estimated, so of course the error is a random variable. The bias of U is simply the expected error, and the mean square error (the name says it all) is the expected square of the error. Note that bias and mean square error are functions of $\theta \in T$. The following definitions are a natural complement to the definition of bias.

Suppose again that U is a statistic used as an estimator of a parameter θ with values in $T \subseteq \mathbb{R}$.

1. U is *unbiased* if $\text{bias}(U) = 0$, or equivalently $E(U) = \theta$, for all $\theta \in T$.

2. U is *negatively biased* if $\text{bias}(U) \leq 0$, or equivalently $\mathbb{E}(U) \leq \theta$, for all $\theta \in T$.
3. U is *positively biased* if $\text{bias}(U) \geq 0$, or equivalently $\mathbb{E}(U) \geq \theta$, for all $\theta \in T$.

Thus, for an unbiased estimator, the expected value of the estimator is the parameter being estimated, clearly a desirable property. On the other hand, a positively biased estimator overestimates the parameter, on average, while a negatively biased estimator underestimates the parameter on average. Our definitions of negative and positive bias are *weak* in the sense that the weak inequalities \leq and \geq are used. There are corresponding strong definitions, of course, using the strong inequalities $<$ and $>$. Note, however, that none of these definitions may apply. For example, it might be the case that $\text{bias}(U) < 0$ for some $\theta \in T$, $\text{bias}(U) = 0$ for other $\theta \in T$, and $\text{bias}(U) > 0$ for yet other $\theta \in T$.

$$\text{mse}(U) = \text{var}(U) + \text{bias}^2(U)$$

Proof

This follows from basic properties of expected value and variance:

$$\mathbb{E}[(U - \theta)^2] = \text{var}(U - \theta) + [\mathbb{E}(U - \theta)]^2 = \text{var}(U) + \text{bias}^2(U) \quad (7.1.2)$$

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

Ideally, we would like to have unbiased estimators with small mean square error. However, this is not always possible, and the result in (3) 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 , we naturally prefer the one with the smaller variance (mean square error).

Suppose that U and V are unbiased estimators of a parameter θ with values in $T \subseteq \mathbb{R}$.

1. U is *more efficient than* V if $\text{var}(U) \leq \text{var}(V)$.
2. The *relative efficiency* of U with respect to V is

$$\text{eff}(U, V) = \frac{\text{var}(V)}{\text{var}(U)} \quad (7.1.3)$$

Asymptotic Properties

Suppose again that we have a real parameter θ with possible values in a parameter space T . Often in a statistical experiment, we observe an infinite sequence of random variables over time, $\mathbf{X} = (X_1, X_2, \dots)$, so that at time n we have observed $\mathbf{X}_n = (X_1, X_2, \dots, X_n)$. In this setting we often have a general formula that defines an estimator of θ for each sample size n . Technically, this gives a *sequence* of real-valued estimators of θ : $\mathbf{U} = (U_1, U_2, \dots)$ where U_n is a real-valued function of \mathbf{X}_n for each $n \in \mathbb{N}_+$. In this case, we can discuss the asymptotic properties of the estimators as $n \rightarrow \infty$. Most of the definitions are natural generalizations of the ones above.

The sequence of estimators $\mathbf{U} = (U_1, U_2, \dots)$ is *asymptotically unbiased* if $\text{bias}(U_n) \rightarrow 0$ as $n \rightarrow \infty$ for every $\theta \in T$, or equivalently, $\mathbb{E}(U_n) \rightarrow \theta$ as $n \rightarrow \infty$ for every $\theta \in T$.

Suppose that $\mathbf{U} = (U_1, U_2, \dots)$ and $\mathbf{V} = (V_1, V_2, \dots)$ are two sequences of estimators that are asymptotically unbiased. The *asymptotic relative efficiency* of \mathbf{U} to \mathbf{V} is

$$\lim_{n \rightarrow \infty} \text{eff}(U_n, V_n) = \lim_{n \rightarrow \infty} \frac{\text{var}(V_n)}{\text{var}(U_n)} \quad (7.1.4)$$

assuming that the limit exists.

Naturally, we expect our estimators to improve, as the sample size n increases, and in some sense to converge to the parameter as $n \rightarrow \infty$. This general idea is known as *consistency*. Once again, for the remainder of this discussion, we assume that $\mathbf{U} = (U_1, U_2, \dots)$ is a sequence of estimators for a real-valued parameter θ , with values in the parameter space T .

Consistency

1. \mathbf{U} is consistent if $U_n \rightarrow \theta$ as $n \rightarrow \infty$ in probability for each $\theta \in T$. That is, $\mathbb{P}(|U_n - \theta| > \epsilon) \rightarrow 0$ as $n \rightarrow \infty$ for every $\epsilon > 0$ and $\theta \in T$.
2. \mathbf{U} is mean-square consistent if $\text{mse}(U_n) = \mathbb{E}[(U_n - \theta)^2] \rightarrow 0$ as $n \rightarrow \infty$ for $\theta \in T$.

Here is the connection between the two definitions:

If \mathbf{U} is mean-square consistent then \mathbf{U} is consistent.

Proof

From Markov's inequality,

$$\mathbb{P}(|U_n - \theta| > \epsilon) = \mathbb{P}[(U_n - \theta)^2 > \epsilon^2] \leq \frac{\mathbb{E}[(U_n - \theta)^2]}{\epsilon^2} \rightarrow 0 \text{ as } n \rightarrow \infty \quad (7.1.5)$$

That mean-square consistency implies simple consistency is simply a statistical version of the theorem that states that mean-square convergence implies convergence in probability. Here is another nice consequence of mean-square consistency.

If \mathbf{U} is mean-square consistent then \mathbf{U} is asymptotically unbiased.

Proof

This result follows from the fact that mean absolute error is smaller than root mean square error, which in turn is special case of a general result for norms. See the advanced section on vector spaces for more details. So, using this result and the ordinary triangle inequality for expected value we have

$$|\mathbb{E}(U_n - \theta)| \leq \mathbb{E}(|U_n - \theta|) \leq \sqrt{\mathbb{E}[(U_n - \theta)^2]} \rightarrow 0 \text{ as } n \rightarrow \infty \quad (7.1.6)$$

Hence $\mathbb{E}(U_n) \rightarrow \theta$ as $n \rightarrow \infty$ for $\theta \in T$.

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

Estimation in the Single Variable Model

Suppose that X is a basic real-valued random variable for an experiment, with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 \in (0, \infty)$. We sample from the distribution of X to produce a sequence $\mathbf{X} = (X_1, X_2, \dots)$ of independent variables, each with the distribution of X . For each $n \in \mathbb{N}_+$, $\mathbf{X}_n = (X_1, X_2, \dots, X_n)$ is a random sample of size n from the distribution of X .

Estimating the Mean

This subsection is a review of some results obtained in the section on the Law of Large Numbers in the chapter on Random Samples. Recall that a natural estimator of the distribution mean μ is the sample mean, defined by

$$M_n = \frac{1}{n} \sum_{i=1}^n X_i, \quad n \in \mathbb{N}_+ \quad (7.1.7)$$

Properties of $\mathbf{M} = (M_1, M_2, \dots)$ as a sequence of estimators of μ .

1. $\mathbb{E}(M_n) = \mu$ so M_n is unbiased for $n \in \mathbb{N}_+$
2. $\text{var}(M_n) = \sigma^2/n$ for $n \in \mathbb{N}_+$ so \mathbf{M} is consistent.

The consistency of \mathbf{M} is simply the weak law of large numbers. Moreover, there are a number of important special cases of the results in (10). See the section on Sample Mean for the details.

Special cases of the sample mean

1. Suppose that $X = \mathbf{1}_A$, the indicator variable for an event A that has probability $\mathbb{P}(A)$. Then the sample mean for a random sample of size $n \in \mathbb{N}_+$ from the distribution of X is the *relative frequency* or *empirical probability* of A , denoted $P_n(A)$. Hence $P_n(A)$ is an unbiased estimator of $\mathbb{P}(A)$ for $n \in \mathbb{N}_+$ and $(P_n(A) : n \in \mathbb{N}_+)$ is consistent..

- Suppose that F denotes the distribution function of a real-valued random variable Y . Then for fixed $y \in \mathbb{R}$, the *empirical distribution function* $F_n(y)$ is simply the sample mean for a random sample of size $n \in \mathbb{N}_+$ from the distribution of the indicator variable $X = \mathbf{1}(Y \leq y)$. Hence $F_n(y)$ is an unbiased estimator of $F(y)$ for $n \in \mathbb{N}_+$ and $(F_n(y) : n \in \mathbb{N}_+)$ is consistent.
- Suppose that U is a random variable with a discrete distribution on a countable set S and f denotes the probability density function of U . Then for fixed $u \in S$, the *empirical probability density function* $f_n(u)$ is simply the sample mean for a random sample of size $n \in \mathbb{N}_+$ from the distribution of the indicator variable $X = \mathbf{1}(U = u)$. Hence $f_n(u)$ is an unbiased estimator of $f(u)$ for $n \in \mathbb{N}_+$ and $(f_n(u) : n \in \mathbb{N}_+)$ is consistent.

Estimating the Variance

This subsection is a review of some results obtained in the section on the Sample Variance in the chapter on Random Samples. We also assume that the fourth central moment $\sigma_4 = \mathbb{E}[(X - \mu)^4]$ is finite. Recall that σ_4/σ^4 is the kurtosis of X . Recall first that if μ is known (almost always an artificial assumption), then a natural estimator of σ^2 is a special version of the sample variance, defined by

$$W_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2, \quad n \in \mathbb{N}_+ \quad (7.1.8)$$

Properties of $\mathbf{W}^2 = (W_1^2, W_2^2, \dots)$ as a sequence of estimators of σ^2 .

- $\mathbb{E}(W_n^2) = \sigma^2$ so W_n^2 is unbiased for $n \in \mathbb{N}_+$
- $\text{var}(W_n^2) = \frac{1}{n}(\sigma_4 - \sigma^4)$ for $n \in \mathbb{N}_+$ so \mathbf{W}^2 is consistent.

Proof

\mathbf{W}^2 corresponds to sampling from the distribution of $(X - \mu)^2$. This distribution has mean σ^2 and variance $\sigma_4 - \sigma^4$, so the results follow immediately from theorem (10).

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

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - M_n)^2, \quad n \in \{2, 3, \dots\} \quad (7.1.9)$$

Properties of $\mathbf{S}^2 = (S_2^2, S_3^2, \dots)$ as a sequence of estimators of σ^2

- $\mathbb{E}(S_n^2) = \sigma^2$ so S_n^2 is unbiased for $n \in \{2, 3, \dots\}$
- $\text{var}(S_n^2) = \frac{1}{n} \left(\sigma_4 - \frac{n-3}{n-1} \sigma^4 \right)$ for $n \in \{2, 3, \dots\}$ so \mathbf{S}^2 is consistent sequence.

Naturally, we would like to compare the sequences \mathbf{W}^2 and \mathbf{S}^2 as estimators of σ^2 . But again remember that \mathbf{W}^2 only makes sense if μ is known.

Comparison of \mathbf{W}^2 and \mathbf{S}^2

- $\text{var}(W_n^2) < \text{var}(S_n^2)$ for $n \in \{2, 3, \dots\}$.
- The asymptotic relative efficiency of \mathbf{W}^2 to \mathbf{S}^2 is 1.

So by (a) W_n^2 is better than S_n^2 for $n \in \{2, 3, \dots\}$, assuming that μ is known so that we can actually use W_n^2 . This is perhaps not surprising, but by (b) S_n^2 works just about as well as W_n^2 for a large sample size n . Of course, the sample standard deviation S_n is a natural estimator of the distribution standard deviation σ . Unfortunately, this estimator is biased. Here is a more general result:

Suppose that θ is a parameter with possible values in $T \subseteq (0, \infty)$ (with at least two points) and that U is a statistic with values in T . If U^2 is an unbiased estimator of θ^2 then U is a negatively biased estimator of θ .

Proof

Note that

$$\text{var}(U) = \mathbb{E}(U^2) - [\mathbb{E}(U)]^2 = \theta^2 - [\mathbb{E}(U)]^2, \quad \theta \in T \quad (7.1.10)$$

Since T has at least two points, U cannot be deterministic so $\text{var}(U) > 0$. It follows that $[\mathbb{E}(U)]^2 < \theta^2$ so $\mathbb{E}(U) < \theta$ for $\theta \in T$.

Thus, we should not be too obsessed with the unbiased property. For most sampling distributions, there will be no statistic U with the property that U is an unbiased estimator of σ and U^2 is an unbiased estimator of σ^2 .

Estimation in the Bivariate Model

In this subsection we review some of the results obtained in the section on the Correlation and Regression in the chapter on Random Samples

Suppose that X and Y are real-valued random variables for an experiment, so that (X, Y) has a bivariate distribution in \mathbb{R}^2 . Let $\mu = \mathbb{E}(X)$ and $\sigma^2 = \text{var}(X)$ denote the mean and variance of X , and let $\nu = \mathbb{E}(Y)$ and $\tau^2 = \text{var}(Y)$ denote the mean and variance of Y . For the bivariate parameters, let $\delta = \text{cov}(X, Y)$ denote the distribution covariance and $\rho = \text{cor}(X, Y)$ the distribution correlation. We need one higher-order moment as well: let $\delta_2 = \mathbb{E}[(X - \mu)^2(Y - \nu)^2]$, and as usual, we assume that all of the parameters exist. So the general parameter spaces are $\mu, \nu \in \mathbb{R}$, $\sigma^2, \tau^2 \in (0, \infty)$, $\delta \in \mathbb{R}$, and $\rho \in [0, 1]$. Suppose now that we sample from the distribution of (X, Y) to generate a sequence of independent variables $((X_1, Y_1), (X_2, Y_2), \dots)$, each with the distribution of (X, Y) . As usual, we will let $\mathbf{X}_n = (X_1, X_2, \dots, X_n)$ and $\mathbf{Y}_n = (Y_1, Y_2, \dots, Y_n)$; these are random samples of size n from the distributions of X and Y , respectively.

Since we now have two underlying variables, we need to enhance our notation somewhat. It will help to define the deterministic versions of our statistics. So if $\mathbf{x} = (x_1, x_2, \dots)$ and $\mathbf{y} = (y_1, y_2, \dots)$ are sequences of real numbers and $n \in \mathbb{N}_+$, we define the mean and special covariance functions by

$$m_n(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n x_i$$

$$w_n(\mathbf{x}, \mathbf{y}) = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)(y_i - \nu)$$

If $n \in \{2, 3, \dots\}$ we define the variance and standard covariance functions by

$$s_n^2(\mathbf{x}) = \frac{1}{n-1} \sum_{i=1}^n [x_i - m_n(\mathbf{x})]^2$$

$$s_n(\mathbf{x}, \mathbf{y}) = \frac{1}{n-1} \sum_{i=1}^n [x_i - m_n(\mathbf{x})][y_i - m_n(\mathbf{y})]$$

It should be clear from context whether we are using the one argument or two argument version of s_n . On this point, note that $s_n(\mathbf{x}, \mathbf{x}) = s_n^2(\mathbf{x})$.

Estimating the Covariance

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

$$W_n = w_n(\mathbf{X}, \mathbf{Y}) = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)(Y_i - \nu), \quad n \in \mathbb{N}_+ \quad (7.1.11)$$

Properties of $\mathbf{W} = (W_1, W_2, \dots)$ as a sequence of estimators of δ .

1. $\mathbb{E}(W_n) = \delta$ so W_n is unbiased for $n \in \mathbb{N}_+$.
2. $\text{var}(W_n) = \frac{1}{n}(\delta_2 - \delta^2)$ for $n \in \mathbb{N}_+$ so \mathbf{W} is consistent.

Proof

We've done this proof before, but it's so basic that it's worth repeating. Note that \mathbf{W} corresponds to sampling from the distribution of $(X - \mu)(Y - \nu)$. This distribution has mean δ and variance $\delta_2 - \delta^2$, so the results follow immediately from Theorem (10).

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

$$S_n = s_n(\mathbf{X}, \mathbf{Y}) = \frac{1}{n-1} \sum_{i=1}^n [X_i - m_n(\mathbf{X})][Y_i - m_n(\mathbf{Y})], \quad n \in \{2, 3, \dots\} \quad (7.1.12)$$

Properties of $\mathbf{S} = (S_2, S_3, \dots)$ as a sequence of estimators of δ .

1. $\mathbb{E}(S_n) = \delta$ so S_n is unbiased for $n \in \{2, 3, \dots\}$.
2. $\text{var}(S_n) = \frac{1}{n} \left(\delta_2 + \frac{1}{n-1} \sigma^2 \tau^2 - \frac{n-2}{n-1} \delta^2 \right)$ for $n \in \{2, 3, \dots\}$ so \mathbf{S} is consistent.

Once again, since we have two competing sequences of estimators of δ , we would like to compare them.

Comparison of \mathbf{W} and \mathbf{S} as estimators of δ :

1. $\text{var}(W_n) < \text{var}(S_n)$ for $n \in \{2, 3, \dots\}$.
2. The asymptotic relative efficiency of \mathbf{W} to \mathbf{S} is 1.

Thus, U_n is better than V_n for $n \in \{2, 3, \dots\}$, assuming that μ and ν are known so that we can actually use W_n . But for large n , V_n works just about as well as U_n .

Estimating the Correlation

A natural estimator of the distribution correlation ρ is the sample correlation

$$R_n = \frac{s_n(\mathbf{X}, \mathbf{Y})}{s_n(\mathbf{X})s_n(\mathbf{Y})}, \quad n \in \{2, 3, \dots\} \quad (7.1.13)$$

Note that this statistics is a nonlinear function of the sample covariance and the two sample standard deviations. For most distributions of (X, Y) , we have no hope of computing the bias or mean square error of this estimator. If we *could* compute the expected value, we would probably find that the estimator is biased. On the other hand, even though we cannot compute the mean square error, a simple application of the law of large numbers shows that $R_n \rightarrow \rho$ as $n \rightarrow \infty$ with probability 1. Thus, $\mathbf{R} = (R_2, R_3, \dots)$ is at least consistent.

Estimating the regression coefficients

Recall that the distribution regression line, with X as the predictor variable and Y as the response variable, is $y = a + bx$ where

$$a = \mathbb{E}(Y) - \frac{\text{cov}(X, Y)}{\text{var}(X)} \mathbb{E}(X), \quad b = \frac{\text{cov}(X, Y)}{\text{var}(X)} \quad (7.1.14)$$

On the other hand, the sample regression line, based on the sample of size $n \in \{2, 3, \dots\}$, is $y = A_n + B_n x$ where

$$A_n = m_n(\mathbf{Y}) - \frac{s_n(\mathbf{X}, \mathbf{Y})}{s_n^2(\mathbf{X})} m_n(\mathbf{X}), \quad B_n = \frac{s_n(\mathbf{X}, \mathbf{Y})}{s_n^2(\mathbf{X})} \quad (7.1.15)$$

Of course, the statistics A_n and B_n are natural estimators of the parameters a and b , respectively, and in a sense are derived from our previous estimators of the distribution mean, variance, and covariance. Once again, for most distributions of (X, Y) , it would be difficult to compute the bias and mean square errors of these estimators. But applications of the law of large numbers show that with probability 1, $A_n \rightarrow a$ and $B_n \rightarrow b$ as $n \rightarrow \infty$, so at least $\mathbf{A} = (A_2, A_3, \dots)$ and $\mathbf{B} = (B_2, B_3, \dots)$ are consistent.

Exercises and Special Cases

The Poisson Distribution

Let's consider a simple example that illustrates some of the ideas above. Recall that the *Poisson distribution* with parameter $\lambda \in (0, \infty)$ has probability density function g given by

$$g(x) = e^{-\lambda} \frac{\lambda^x}{x!}, \quad x \in \mathbb{N} \quad (7.1.16)$$

The Poisson distribution is often used to model the number of random “points” in a region of time or space, and is studied in more detail in the chapter on the Poisson process. The parameter λ is proportional to the size of the region of time or space; the proportionality constant is the average *rate* of the random points. The distribution is named for Simeon Poisson.

Suppose that X has the Poisson distribution with parameter λ . Hence

1. $\mu = \mathbb{E}(X) = \lambda$
2. $\sigma^2 = \text{var}(X) = \lambda$
3. $\sigma_4 = \mathbb{E}[(X - \lambda)^4] = 3\lambda^2 + \lambda$

Proof

Recall the permutation notation $x^{(n)} = x(x-1)\cdots(x-n+1)$ for $x \in \mathbb{R}$ and $n \in \mathbb{N}$. The expected value $\mathbb{E}[X^{(n)}]$ is the *factorial moment* of X of order n . It's easy to see that the factorial moments are $\mathbb{E}[X^{(n)}] = \lambda^n$ for $n \in \mathbb{N}$. The results follow from this.

Suppose now that we sample from the distribution of X to produce a sequence of independent random variables $\mathbf{X} = (X_1, X_2, \dots)$, each having the Poisson distribution with unknown parameter $\lambda \in (0, \infty)$. Again, $\mathbf{X}_n = (X_1, X_2, \dots, X_n)$ is a random sample of size $n \in \mathbb{N}_+$ from the distribution for each $n \in \mathbb{N}$. From the previous exercise, λ is both the mean and the variance of the distribution, so that we could use either the sample mean M_n or the sample variance S_n^2 as an estimator of λ . Both are unbiased, so which is better? Naturally, we use mean square error as our criterion.

Comparison of M to S^2 as estimators of λ .

1. $\text{var}(M_n) = \frac{\lambda}{n}$ for $n \in \mathbb{N}_+$.
2. $\text{var}(S_n^2) = \frac{\lambda}{n} \left(1 + 2\lambda \frac{n}{n-1}\right)$ for $n \in \{2, 3, \dots\}$.
3. $\text{var}(M_n) < \text{var}(S_n^2)$ so M_n for $n \in \{2, 3, \dots\}$.
4. The asymptotic relative efficiency of M to S^2 is $1 + 2\lambda$.

So our conclusion is that the sample mean M_n is a better estimator of the parameter λ than the sample variance S_n^2 for $n \in \{2, 3, \dots\}$, and the difference in quality increases with λ .

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

The emission of elementary particles from a sample of radioactive material in a time interval is often assumed to follow the Poisson distribution. Thus, suppose that the alpha emissions data set is a sample from a Poisson distribution. Estimate the rate parameter λ .

1. using the sample mean
2. using the sample variance

Answer

1. 8.367
2. 8.649

Simulation Exercises

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 and compare the sample mean to the distribution mean.

Run the normal estimation experiment 1000 times for several values of the parameters.

1. Compare the empirical bias and mean square error of M with the theoretical values.
2. 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?

In matching experiment, the random variable is the number of matches. Run the simulation 1000 times and compare

1. the sample mean to the distribution mean.
2. the empirical density function to the probability density function.

Run the exponential experiment 1000 times and compare the sample standard deviation to the distribution standard deviation.

Data Analysis Exercises

For Michelson's velocity of light data, compute the sample mean and sample variance.

Answer

852.4, 6242.67

For Cavendish's density of the earth data, compute the sample mean and sample variance.

Answer

5.448, 0.048817

For Short's parallax of the sun data, compute the sample mean and sample variance.

Answer

8.616, 0.561032

Consider the Cicada data.

1. Compute the sample mean and sample variance of the body length variable.
2. Compute the sample mean and sample variance of the body weight variable.
3. Compute the sample covariance and sample correlation between the body length and body weight variables.

Answer

1. 24.0, 3.92
2. 0.180, 0.003512
3. 0.0471, 0.4012

Consider the M&M data.

1. Compute the sample mean and sample variance of the net weight variable.
2. Compute the sample mean and sample variance of the total number of candies.
3. Compute the sample covariance and sample correlation between the number of candies and the net weight.

Answer

1. 57.1, 5.68
2. 49.215, 2.3163
3. 2.878, 0.794

Consider the Pearson data.

1. Compute the sample mean and sample variance of the height of the father.
2. Compute the sample mean and sample variance of the height of the son.
3. Compute the sample covariance and sample correlation between the height of the father and height of the son.

Answer

1. 67.69, 7.5396
2. 68.68, 7.9309
3. 3.875, 0.501

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.

This page titled [7.1: Estimators](#) is shared under a [CC BY 2.0](#) license and was authored, remixed, and/or curated by [Kyle Siegrist](#) ([Random Services](#)) via [source content](#) that was edited to the style and standards of the LibreTexts platform.