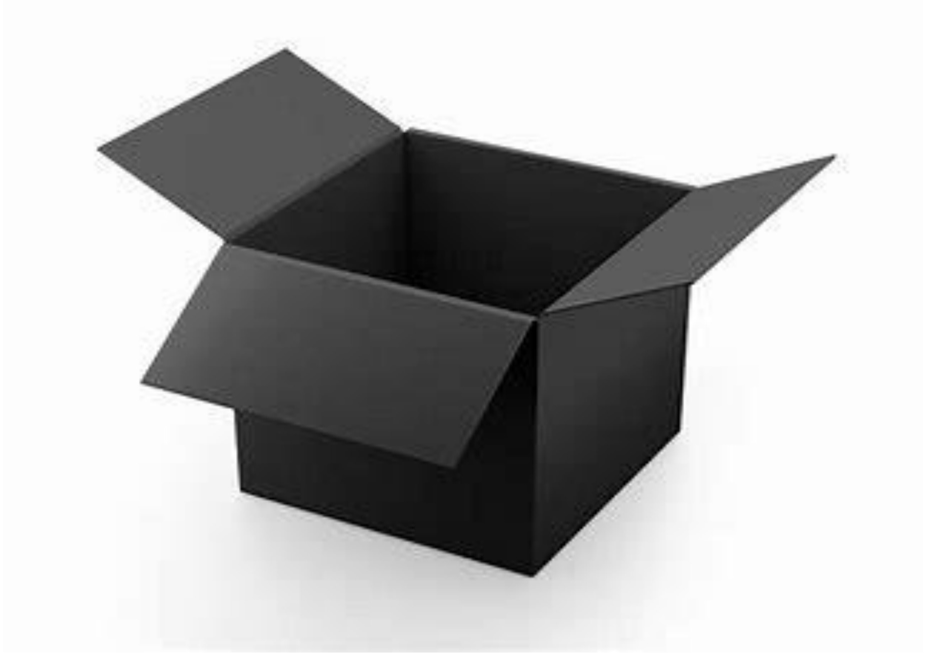
Section 10.4

Point Estimates of Model Parameters

Model Parameters

- The general problem is estimation of a *parameter* of a probability model.
- A parameter is any number that can be calculated from the probability model.
- For example, for an arbitrary event A, $\mathsf{P}[A]$ is a model parameter.



Estimates of Model Parameters

- Consider an experiment that produces observations of sample values of the random variable X.
- The observations are sample values of the random variables $X_1, X_2, \ldots$, all with the same probability model as X.
- Assume that r is a parameter of the probability model.
- We use the observations $X_1, X_2, \ldots$ to produce a sequence of estimates of r.
- The estimates $\hat{R}_1, \hat{R}_2, \ldots$ are all random variables.
- $\hat{R}_1$ is a function of X_1.
- $\hat{R}_2$ is a function of X_1 and X_2, and in general $\hat{R}_n$ is a function of $X_1, X_2, \ldots, X_n$.

Definition 10.3 Consistent Estimator

The sequence of estimates $\hat{R}_1, \hat{R}_2, \ldots$ *of parameter* r *is* consistent *if for any* $\epsilon > 0$,

$$\lim_{n\to\infty} \mathrm{P}\left[\left|\hat{R}_n - r\right| \geq \epsilon\right] = 0.$$

“一致估计”

Definition 10.4 Unbiased Estimator

An estimate, $\widehat{R}$*, of parameter* r *is* unbiased *if* $\mathrm{E}[\widehat{R}] = r$*; otherwise,* $\widehat{R}$ *is* biased.

“无偏估计”

Definition 10.3 Consistent Estimator

The sequence of estimates $\widehat{R}_1, \widehat{R}_2, \ldots$ *of parameter* r *is* consistent *if for any* $\epsilon > 0$,

$$\lim_{n\to\infty} \mathrm{P}\left[\left|\widehat{R}_n - r\right| \geq \epsilon\right] = 0.$$

“一致估计”

相等？区别？

Definition 10.4 Unbiased Estimator

An estimate, $\widehat{R}$, *of parameter* r *is* unbiased *if* $\mathrm{E}[\widehat{R}] = r$; *otherwise,* $\widehat{R}$ *is* biased.

“无偏估计”

Definition 10.3 Consistent Estimator

The sequence of estimates $\hat{R}_1, \hat{R}_2, \ldots$ *of parameter* r *is* consistent *if for any* $\epsilon > 0$,

$$\lim_{n\to\infty} \mathrm{P}\left[\left|\hat{R}_n - r\right| \geq \epsilon\right] = 0.$$

"对 估计值与真实值的距离概率分布 的限制"

"一致估计"

Definition 10.4 Unbiased Estimator

An estimate, $\hat{R}$*, of parameter* r *is* unbiased *if* $\mathrm{E}[\hat{R}] = r$*; otherwise,* $\hat{R}$ *is* biased.

"对 估计值的期望（概率分布加权和） 的限制"

"无偏估计"

Asymptotically Unbiased

Definition 10.5 Estimator

The sequence of estimators $\hat{R}_n$ *of parameter* r *is* asymptotically unbiased *if*

$$\lim_{n\to\infty} \mathsf{E}[\hat{R}_n] = r.$$

“渐进无偏估计”

$$\lim_{n\to\infty} Var[\hat{R}_n] = 0$$

“无偏估计”

$$\mathsf{E}[\hat{R}] = r$$

“一致估计”

$$\lim_{n\to\infty} \mathsf{P}\left[\left|\hat{R}_n - r\right| \geq \epsilon\right] = 0.$$

Definition 10.6 Mean Square Error

The mean square error *of estimator* $\hat{R}$ *of parameter* r *is*

$$e = \mathrm{E}\left[(\hat{R} - r)^2\right].$$

“均方误差”

Note：不是均方差=标准差

Theorem 10.8

If a sequence of unbiased estimates $\hat{R}_1, \hat{R}_2, \ldots$ of parameter r has mean square error $e_n = \mathrm{Var}[\hat{R}_n]$ satisfying $\lim_{n\to\infty} e_n = 0$, then the sequence $\hat{R}_n$ is consistent.

$$\boxed{\mathrm{E}[\hat{R}] = r} \longrightarrow \boxed{\lim_{n\to\infty} \mathrm{P}\left[\left|\hat{R}_n - r\right| \geq \epsilon\right] = 0.}$$

For an arbitrary random variable Y and constant $c > 0$,

$$\mathrm{P}\left[|Y - \mu_Y| \geq c\right] \leq \frac{\mathrm{Var}[Y]}{c^2}.$$

$$\Rightarrow \quad \boxed{\mathrm{P}\left[\left|\hat{R}_n - r\right| \geq \epsilon\right] \leq \frac{\mathrm{Var}[\hat{R}_n]}{\epsilon^2}.}$$

$$\left.\begin{array}{l} \boxed{e_n = \mathrm{Var}[\hat{R}_n]} \\ \boxed{\lim_{n\to\infty} e_n = 0} \end{array}\right\}$$

$$\Rightarrow \quad \boxed{\lim_{n\to\infty} \frac{\mathrm{Var}[\hat{R}_n]}{\epsilon^2} = 0.}$$

Example 10.5 Problem

$\boxed{\mathrm{Var}[N_k] = kr}$

In any interval of k seconds, the number N_k of packets passing through an Internet router is a Poisson random variable with expected value $\mathrm{E}[N_k] = kr$ packets. Let $\hat{R}_k = N_k/k$ denote an estimate of the parameter r packets/second. Is each estimate $\hat{R}_k$ an unbiased estimate of r? What is the mean square error e_k of the estimate $\hat{R}_k$? Is the sequence of estimates $\hat{R}_1, \hat{R}_2, \ldots$ consistent?

$$\boxed{\mathrm{E}[\hat{R}] = r}$$

$$\boxed{e = \mathrm{E}\left[(\hat{R} - r)^2\right].}$$

$$\boxed{\lim_{n\to\infty} \mathrm{P}\left[\left|\hat{R}_n - r\right| \geq \epsilon\right] = 0.}$$

Example 10.5 Solution

First, we observe that $\hat{R}_k$ is an unbiased estimator since

$$\mathrm{E}[\hat{R}_k] = \mathrm{E}\left[N_k/k\right] = \mathrm{E}\left[N_k\right]/k = r. \tag{1}$$

Next, we recall that since N_k is Poisson, $\mathrm{Var}[N_k] = kr$. This implies

$$\mathrm{Var}[\hat{R}_k] = \mathrm{Var}\left[\frac{N_k}{k}\right] = \frac{\mathrm{Var}\left[N_k\right]}{k^2} = \frac{r}{k}. \tag{2}$$

Because $\hat{R}_k$ is unbiased, the mean square error of the estimate is the same as its variance: $e_k = r/k$. In addition, since $\lim_{k\to\infty} \mathrm{Var}[\hat{R}_k] = 0$, the sequence of estimators $\hat{R}_k$ is consistent by Theorem 10.8.

Theorem 10.9

The sample mean $M_n(X)$ is an unbiased estimate of $\mathsf{E}[X]$.

$$\boxed{\mathsf{E}[M_n(X)] = \mathsf{E}[X]} \qquad \boxed{\mathsf{E}[\hat{R}] = r}$$

Theorem 10.10

The sample mean estimator $M_n(X)$ has mean square error

$$e_n = \mathsf{E}\left[(M_n(X) - \mathsf{E}[X])^2\right] = \mathsf{Var}[M_n(X)] = \frac{\mathsf{Var}[X]}{n}.$$

Standard Error

- In the terminology of statistical inference, $\sqrt{e_n}$, the standard deviation of the sample mean, is referred to as the *standard error* of the estimate.

- The standard error gives an indication of how far we should expect the sample mean to deviate from the expected value.

- In particular, when X is a Gaussian random variable (and $M_n(X)$ is also Gaussian),

$$\mathsf{P}\left[\mathsf{E}[X] - \sqrt{e_n} \le M_n(X) \le \mathsf{E}[X] + \sqrt{e_n}\right] = 2\Phi(1) - 1 \approx 0.68. \tag{1}$$

  In words, Equation (10.24) says there is roughly a two-thirds probability that the sample mean is within one standard error of the expected value.

- This same conclusion is approximately true when n is large and the central limit theorem says that $M_n(X)$ is approximately Gaussian.

Example 10.6 Problem

How many independent trials n are needed to guarantee that $\hat{P}_n(A)$, the relative frequency estimate of $\mathrm{P}[A]$, has standard error ≤ 0.1?

$$\boxed{\mathrm{Var}[X_A] = \mathrm{P}[A](1 - \mathrm{P}[A])}$$

The sample mean estimator $M_n(X)$ has mean square error

$$e_n = \mathrm{E}\left[(M_n(X) - \mathrm{E}[X])^2\right] = \mathrm{Var}[M_n(X)] = \frac{\mathrm{Var}[X]}{n}.$$

$$\boxed{p(1-p) \leq 0.25 \quad \text{for all } 0 \leq p \leq 1}$$

In the terminology of statistical inference, $\sqrt{e_n}$, the standard deviation of the sample mean, is referred to as the *standard error* of the estimate.

Example 10.6 Solution

Since the indicator X_A has variance $\text{Var}[X_A] = P[A](1 - P[A])$, Theorem 10.10 implies that the mean square error of $M_n(X_A)$ is

$$e_n = \frac{\text{Var}[X]}{n} = \frac{P[A](1 - P[A])}{n}. \qquad (1)$$

We need to choose n large enough to guarantee $\sqrt{e_n} \le 0.1$ ($e_n \le= 0.01$) even though we don't know $P[A]$. We use the fact that $p(1-p) \le 0.25$ for all $0 \le p \le 1$. Thus, $e_n \le 0.25/n$. To guarantee $e_n \le 0.01$, we choose $n = 0.25/0.01 = 25$ trials.

Theorem 10.11

If X has finite variance, then the sample mean $M_n(X)$ is a sequence of consistent estimates of $\mathsf{E}[X]$.

The sample mean estimator $M_n(X)$ has mean square error

$$e_n = \mathsf{E}\left[(M_n(X) - \mathsf{E}[X])^2\right] = \mathrm{Var}[M_n(X)] = \frac{\mathrm{Var}[X]}{n}.$$

For an arbitrary random variable Y and constant $c > 0$,

$$\mathsf{P}\left[|Y - \mu_Y| \geq c\right] \leq \frac{\mathrm{Var}[Y]}{c^2}.$$

$$\mathsf{P}\left[\left|\hat{R}_n - r\right| \geq \epsilon\right] \leq \frac{\mathrm{Var}[\hat{R}_n]}{\epsilon^2}.$$

$$\lim_{n\to\infty} \mathsf{P}\left[\left|\hat{R}_n - r\right| \geq \epsilon\right] = 0.$$

Estimating the Variance

- When $\mathsf{E}[X]$ is a known quantity μ_X, we know $\mathsf{Var}[X] = \mathsf{E}[(X - \mu_X)^2]$.
- In this case, we can use the sample mean of $W = (X - \mu_X)^2$ to estimate $\mathsf{Var}[X]$.,

$$M_n(W) = \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu_X)^2 . \tag{1}$$

  If $\mathsf{Var}[W]$ exists, $M_n(W)$ is a consistent, unbiased estimate of $\mathsf{Var}[X]$.
- When the expected value μ_X is unknown, the situation is more complicated because the variance of X depends on μ_X.
- We cannot use Equation (10.28) if μ_X is unknown.
- In this case, we replace the expected value μ_X by the sample mean $M_n(X)$.

Definition 10.7 Sample Variance

The sample variance of n independent observations of random variable X is

$$V_n(X) = \frac{1}{n} \sum_{i=1}^{n} (X_i - M_n(X))^2 .$$

Theorem 10.12

$$\mathrm{E}\left[V_n(X)\right] = \frac{n-1}{n}\mathrm{Var}[X].$$

Intuitive Explanation:

The observed values fall, on average, closer to the sample mean than to the population mean, the variance which is calculated using variances from the sample mean ***underestimates*** the desired variance of the population.

Hence, using n-1 instead of n as the divisor corrects for that by making the result a little bit bigger.

Theorem 10.12

$$\mathsf{E}\left[V_n(X)\right] = \frac{n-1}{n}\mathsf{Var}[X].$$

$$\boxed{\mathsf{E}\left[M_n(X)\right] = \mathsf{E}\left[X\right]}$$

$$\boxed{V_n(X) = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - M_n(X)\right)^2.}$$

$$\boxed{V_n = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} X_i X_j.}$$

$$\boxed{\mathsf{E}[X_i^2] = \mathsf{E}[X^2]}$$

$$\boxed{\mathsf{E}[X_i X_j] = \mathsf{Cov}[X_i, X_j] + \mathsf{E}[X_i]\,\mathsf{E}[X_j]}$$

$$\boxed{\begin{aligned}\mathsf{E}\left[V_n\right] &= \mathsf{E}\left[X^2\right] - \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\mathsf{Cov}\left[X_i, X_j\right] + \mu_X^2\right) \\ &= \mathsf{Var}[X] - \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathsf{Cov}\left[X_i, X_j\right].\end{aligned}}$$

Proof: Theorem 10.12

Substituting Definition 10.1 of the sample mean $M_n(X)$ into Definition 10.7 of sample variance and expanding the sums, we derive

$$V_n = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} X_i X_j. \tag{1}$$

Because the X_i are iid, $\mathsf{E}[X_i^2] = \mathsf{E}[X^2]$ for all i, and $\mathsf{E}[X_i]\,\mathsf{E}[X_j] = \mu_X^2$. By Theorem 5.16(a), $\mathsf{E}[X_i X_j] = \mathsf{Cov}[X_i, X_j] + \mathsf{E}[X_i]\,\mathsf{E}[X_j]$. Thus, $\mathsf{E}[X_i X_j] = \mathsf{Cov}[X_i, X_j] + \mu_X^2$. Combining these facts, the expected value of V_n in Equation (10.29) is

$$\begin{aligned}
\mathsf{E}[V_n] &= \mathsf{E}\left[X^2\right] - \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\mathsf{Cov}[X_i, X_j] + \mu_X^2\right) \\
&= \mathsf{Var}[X] - \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathsf{Cov}[X_i, X_j]\,.
\end{aligned} \tag{2}$$

Since the double sum has n^2 terms, $\sum_{i=1}^{n}\sum_{j=1}^{n}\mu_X^2 = n^2\mu_X^2$. Of the n^2 covariance terms, there are n terms of the form $\mathsf{Cov}[X_i, X_i] = \mathsf{Var}[X]$, while the remaining covariance terms are all 0 because X_i and X_j are independent for $i \neq j$. This implies

$$\mathsf{E}[V_n] = \mathsf{Var}[X] - \frac{1}{n^2}\left(n\,\mathsf{Var}[X]\right) = \frac{n-1}{n}\mathsf{Var}[X]. \tag{3}$$

Theorem 10.13

The estimate

$$V_n'(X) = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - M_n(X))^2$$

is an unbiased estimate of $\mathrm{Var}[X]$.

Proof: Theorem 10.13

Using Definition 10.7, we have

$$V_n'(X) = \frac{n}{n-1} V_n(X), \tag{1}$$

and

$$\mathsf{E}\left[V_n'(X)\right] = \frac{n}{n-1}\,\mathsf{E}\left[V_n(X)\right] = \mathsf{Var}[X]. \tag{2}$$

Quiz 10.4

X is the continuous uniform $(-1, 1)$ random variable. Find the mean square error, $\mathsf{E}[(\mathsf{Var}[X] - V_{100}(X))^2]$, of the sample variance estimate of $\mathsf{Var}[X]$, based on 100 independent observations of X.

Theorem 10.12

$$\mathsf{E}[V_n(X)] = \frac{n-1}{n}\mathsf{Var}[X].$$

Quiz 10.4 Solution

Define the random variable $W = (X - \mu_X)^2$. Observe that $V_{100}(X) = M_{100}(W)$. By Theorem 10.10, the mean square error is

$$\mathsf{E}\left[(M_{100}(W) - \mu_W)^2\right] = \frac{\mathsf{Var}[W]}{100}. \tag{1}$$

Observe that $\mu_X = 0$ so that $W = X^2$. Thus,

$$\mu_W = \mathsf{E}\left[X^2\right] = \int_{-1}^{1} x^2 f_X(x)\,dx = 1/3, \tag{2}$$

$$\mathsf{E}\left[W^2\right] = \mathsf{E}\left[X^4\right] = \int_{-1}^{1} x^4 f_X(x)\,dx = 1/5. \tag{3}$$

Therefore $\mathsf{Var}[W] = \mathsf{E}[W^2] - \mu_W^2 = 1/5 - (1/3)^2 = 4/45$ and the mean square error is $4/4500 = 0.0009$.