Chapter 7
Parameter Estimation Using the Sample Mean

Chien-Kang Huang (黃乾綱)
臺大工科海洋系助理教授

Outline

7.1 Sample Mean: Expected Value and Variance

7.2 Deviation of a Random Variable from the Expected Value

7.3 Point Estimates of Model Parameters

7.4 Confidence Intervals

- In referring to applications of probability theory, we have assumed prior knowledge of the probability model that governs the outcomes of an experiment.
- In practice, however, we encounter many situations in which the probability model is not know in advance and experimenters collect data in order to learn about the model.
- In doing so, they apply principles of **statistical inference**, a body of knowledge that governs the use of measurements to discover the properties of a probability model.
- **Sample mean** of a set of data
  - The sample mean is simply the sum of the sample values divided by the number of trials.

7.1 Sample Mean: Expected Value and Variance

– In this section, we define the **sample mean** of a random variable and identify its expected value and variance. Later sections of this chapter show mathematically how the sample mean converges to a **constant** as the number of repetitions of an experiment increases. This chapter, therefore, provides the mathematical basis for the statement that although the result of a single experiment is unpredictable, predictable patterns emerge as we collect more and more data.

- To define the sample mean, consider repeated independent trials of an experiment. Each trial results in one observation of a random variable, X. After n trials, we have sample values of the n random variables $X_1, \ldots, X_n$, all with the same PDF as X. The sample mean is the numerical average of the observations:

Definition 7.1 Sample Mean

For iid random variables $X_1, \ldots, X_n$ with PDF $f_X(x)$, the sample mean *of X is the random variable*

$$M_n(X) = \frac{X_1 + \cdots + X_n}{n}.$$

- The first thing to notice is that $M_n(X)$ is a function of the random variables $X_1, \ldots, X_n$, and is therefore a random variable itself.
- It is important to distinguish the sample mean $M_n(X)$, from $E[X]$, which we sometimes refer to as the mean value of random variable X. While $M_n(X)$ is a random variable, $E[X]$ is a number.
- To avoid confusion when studying the sample mean, it is advisable to refer to $E[X]$ as the **expected value** of X, rather than the **mean** of X. The sample mean of X and the expected value of X are closely related.
- A major purpose of this chapter is to explore the fact that as n increases without bound, $M_n(X)$ predictably approaches $E[X]$. In everyday conversation, this phenomenon is often called the **law of averages**.

Theorem 7.1

The sample $M_n(X)$ has expected value and Variance

$$E[M_n(X)] = E[X], \qquad \mathrm{Var}[M_n(X)] = \frac{\mathrm{Var}[X]}{n}$$

Proof: Theorem 7.1

From Definition 7.1, Theorem 6.1 and the fact that $E[X_i] = E[X]$ for all i,

$$E[M_n(X)] = \frac{1}{n}(E[X_1] + \cdots + E[X_n]) = \frac{1}{n}(E[X] + \cdots + E[X]) = E[X].$$

Because $\text{Var}[aY] = a^2\,\text{Var}[Y]$ for any random variable Y (Theorem 2.14), $\text{Var}[M_n(X)] = \text{Var}[X_1 + \cdots + X_n]/n^2$. Since the X_i are iid, we can use Theorem 6.3 to show

$$\text{Var}[X_1 + \cdots + X_n] = \text{Var}[X_1] + \cdots + \text{Var}[X_n] = n\,\text{Var}[X].$$

Thus $\text{Var}[M_n(X)] = n\,\text{Var}[X]/n^2 = \text{Var}[X]/n$.

- Recall that in Section 2.5, we refer to the expected value of a random variable as a typical value. Theorem 7.1 demonstrates that $E[X]$ is a typical value of $M_n(X)$, regardless of n. Furthermore, Theorem 7.1 demonstrates that as n increases without bound, the variance of $M_n(X)$ goes to zero.
  - When we first met the variance, and its square root the standard deviation, we said that they indicate how far a random variable is likely to be form its expected value.
- Theorem 7.1 suggests that as n approaches infinity, it becomes highly likely that converges to the expected value as the number of sample n goes to infinity.
- The rest of this chapter contains the mathematically analysis that describes the nature of this convergence.

Quiz 7.1

- Let X be an exponential random variable with expected value 1. Let $M_n(X)$ denote the sample mean of n independent samples of X. How many samples n are needed to guarantee that the variance of the sample mean is no more than 0.01.

Quiz 7.1 Solution

An exponential random variable with expected value 1 also has variance 1. By Theorem 7.1, $M_n(X)$ has variance $\mathrm{Var}[M_n(X)] = 1/n$. Hence, we need $n = 100$ samples.

7.2 Deviation of a Random Variable from the Expected Value

- The analysis of the convergence of $M_n(X)$ to $E[X]$ begins with a study of the random variable $|Y - \mu_Y|$, the absolute different between an arbitrary random variable Y and its expected value. This study leads to the **Chebyshev inequality**, which states that the probability of a large deviation from the mean is inversely proportional to the square of the deviation. The deviation of the Chebyshev inequality begins with the **Markov inequality**, an upper bound on the probability that a sample value of a nonnegative random variable exceeds the expected value by an arbitrary factor.

Theorem 7.2 Markov Inequality

For a random variable X such that $P[X < 0] = 0$ and a constant c,

$$P\left[X \geq c^2\right] \leq \frac{E\left[X\right]}{c^2}.$$

Proof: Theorem 7.2

Since X is nonnegative, $f_X(x) = 0$ for $x < 0$ and

$$E[X] = \int_0^{c^2} x f_X(x)\, dx + \int_{c^2}^{\infty} x f_X(x)\, dx \geq \int_{c^2}^{\infty} x f_X(x)\, dx.$$

Since $x \geq c^2$ in the remaining integral,

$$E[X] \geq c^2 \int_{c^2}^{\infty} f_X(x)\, dx = c^2 P\left[X \geq c^2\right].$$

- Keep in mind that the Markov inequality is valid only for nonnegative random variables.
- The bound provided by the Markov inequality can be very loose.

Example 7.1

Let X represent the height (in feet) of a randomly chosen adult. If the expected height is $E[X] = 5.5$, then the Markov inequality states that the probability an adult is at least 11 feet tall satisfies

$$P[X \geq 11] \leq 5.5/11 = 1/2.$$

- We say the Markov inequality is a loose bound because the probability that a person is taller than 11 feet is essentially zero, while the inequality merely states that it is less than or equal to 1/2. Although the bound is extremely loose for many random variables, it is tight (in fact, an equation) with respect to some random variables.

Example 7.2

Suppose random variable Y takes on the value c^2 with probability p and the value 0 otherwise. In this case, $E[Y] = pc^2$ and the Markov inequality states

$$P\left[Y \geq c^2\right] \leq E\left[Y\right]/c^2 = p.$$

Since $P[Y \geq c^2] = p$, we observe that the Markov inequality is in fact an equality in this instance.

Theorem 7.3 Chebyshev Inequality

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

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

- The Chebyshev inequality applies the Marko inequality to the nonnegative random variable $(Y - \mu_Y)^2$, derived from any random variable Y.

Proof: Theorem 7.3

In the Markov inequality, Theorem 7.2, let $X = (Y - \mu_Y)^2$. The inequality states

$$P\left[X \geq c^2\right] = P\left[(Y - \mu_Y)^2 \geq c^2\right] \leq \frac{E\left[(Y - \mu_Y)^2\right]}{c^2} = \frac{\text{Var}[Y]}{c^2}.$$

The theorem follows from the fact that $\{(Y - \mu_Y)^2 \geq c^2\} = \{|Y - \mu_Y| \geq c\}$.

- Unlike the Markov inequality, the Chebyshev inequality is valid for all random variables.
- While the Markov inequality refers only to the expected value of a random variable, the Chebyshev inequality also refers to the variance. Because it uses more information about the random variable, the Chebyshev inequality generally provides a tighter bound than the Markov inequality.
- In particular, when the variance of Y is very small, the Chebyshev inequality says it is unlikely that Y is far away from $E[Y]$.

Example 7.3 Problem

If the height X of a randomly chosen adult has expected value $E[X] = 5.5$ feet and standard deviation $\sigma_X = 1$ foot, use the Chebyshev inequality to to find an upper bound on $P[X \geq 11]$.

Example 7.3 Solution

Since a height X is nonnegative, the probability that $X \geq 11$ can be written as

$$P[X \geq 11] = P[X - \mu_X \geq 11 - \mu_X] = P[|X - \mu_X| \geq 5.5].$$

Now we use the Chebyshev inequality to obtain

$$P[X \geq 11] = P[|X - \mu_X| \geq 5.5] \leq \mathrm{Var}[X]/(5.5)^2 = 0.033 \approx 1/30.$$

Although this bound is better than the Markov bound, it is also loose. In fact, $P[X \geq 11]$ is orders of magnitude lower than 1/30. Otherwise, we would expect often to see a person over 11 feet tall in a group of 30 or more people!

Quiz 7.2

Elevators arrive randomly at the ground floor of an office building. Because of a large crowd, a person will wait for time W in order to board the third arriving elevator. Let X_1 denote the time (in seconds) until the first elevator arrives and let X_i denote the time between the arrival of elevator $i-1$ and i. Suppose X_1, X_2, X_3 are independent uniform $(0, 30)$ random variables. Find upper bounds to the probability W exceeds 75 seconds using

(1) the Markov inequality,

(2) the Chebyshev inequality.

Quiz 7.2 Solution

The arrival time of the third elevator is $W = X_1 + X_2 + X_3$. Since each X_i is uniform $(0, 30)$,

$$E[X_i] = 15, \qquad \operatorname{Var}[X_i] = \frac{(30-0)^2}{12} = 75.$$

Thus $E[W] = 3E[X_i] = 45$, and $\operatorname{Var}[W] = 3\operatorname{Var}[X_i] = 225$.

(1) By the Markov inequality,

$$P[W > 75] \leq \frac{E[W]}{75} = \frac{45}{75} = \frac{3}{5}$$

(2) By the Chebyshev inequality,

$$\begin{aligned} P[W > 75] &= P[W - E[W] > 30] \\ &\leq P[|W - E[W]| > 30] \leq \frac{\operatorname{Var}[W]}{30^2} = \frac{225}{900} = \frac{1}{4} \end{aligned}$$

7.3 Point Estimates of Model Parameters

Definition 7.2 Consistent Estimator

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

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

Definition 7.3 Unbiased Estimator

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

Asymptotically Unbiased

Definition 7.4 Estimator

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

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

Definition 7.5 Mean Square Error

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

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

Theorem 7.4

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.

Proof: Theorem 7.4

Since $E[\hat{R}_n] = r$, we can apply the Chebyshev inequality to $\hat{R}_n$. For any constant $\epsilon > 0$,

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

In the limit of large n, we have

$$\lim_{n\to\infty} P\left[\left|\hat{R}_n - r\right| \geq \epsilon\right] \leq \lim_{n\to\infty} \frac{\mathrm{Var}[\hat{R}_n]}{\epsilon^2} = 0.$$

Example 7.4 Problem

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 $E[N_k] = kr$ packets. Let $\hat{R}_k = N_k/k$ denote an estimate of r. 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?

Example 7.4 Solution

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

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

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

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

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 $\lim_{k\to\infty} \text{Var}[\hat{R}_k] = 0$, the sequence of estimators $\hat{R}_k$ is consistent by Theorem 7.4.

Theorem 7.5

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

Theorem 7.6

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

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

Example 7.5 Problem

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

Example 7.5 Solution

Since the indicator X_A has variance $\text{Var}[X_A] = P[A](1 - P[A])$, Theorem 7.6 implies that the mean square error of $M_n(X_A)$ is

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

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

Theorem 7.7

- If X has finite variance, then the sample mean is a sequence of consistent estimates of $E[X]$

Proof: Theorem 7.7

By Theorem 7.6, the mean square error of $M_n(X)$ satisfies

$$\lim_{n\to\infty} \mathrm{Var}[M_n(X)] = \lim_{n\to\infty} \frac{\mathrm{Var}[X]}{n} = 0.$$

By Theorem 7.4, the sequence $M_n(X)$ is consistent.

Theorem 7.8 Weak Law of Large Numbers

If X has finite variance, then for any constant $c > 0$,

(a) $\lim_{n\to\infty} P[|M_n(X) - \mu_X| \geq c] = 0,$

(b) $\lim_{n\to\infty} P[|M_n(X) - \mu_X| < c] = 1.$

Theorem 7.9

As $n \to \infty$, the relative frequency $\hat{P}_n(A)$ converges to $P[A]$; for any constant $c > 0$,

$$\lim_{n\to\infty} P\left[\left|\hat{P}_n(A) - P\left[A\right]\right| \geq c\right] = 0.$$

Proof: Theorem 7.9

The proof follows from Theorem 7.4 since $\hat{P}_n(A) = M_n(X_A)$ is the sample mean of the indicator X_A, which has mean $E[X_A] = P[A]$ and finite variance $\mathrm{Var}[X_A] = P[A](1 - P[A])$.

Definition 7.6 Convergence in Probability

The random sequence Y_n *converges in probability to a constant* y *if for any* $\epsilon > 0$,

$$\lim_{n\to\infty} P\left[|Y_n - y| \geq \epsilon\right] = 0.$$

Definition 7.7 Sample Variance

The sample variance of a set 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 7.10

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

Proof: Theorem 7.10

Substituting Definition 7.1 of the sample mean $M_n(X)$ into Definition 7.7 of sample variance and expanding the sums, we derive

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

Because the X_i are iid, $E[X_i^2] = E[X^2]$ for all i, and $E[X_i]E[X_j] = \mu_X^2$. By Theorem 4.16(a), $E[X_i X_j] = \mathrm{Cov}[X_i, X_j] + E[X_i]E[X_j]$. Thus, $E[X_i X_j] = \mathrm{Cov}[X_i, X_j] + \mu_X^2$. Combining these facts, the expected value of V_n in Equation (7.22) is

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

Note that 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 $\mathrm{Cov}[X_i, X_i] = \mathrm{Var}[X]$, while the remaining covariance terms are all 0 because X_i and X_j are independent for $i \neq j$. This implies

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

Theorem 7.11

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 7.11

Using Definition 7.7, we have

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

and

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

Quiz 7.3

X is a uniform random variable between -1 and 1 with PDF

$$f_X(x) = \begin{cases} 0.5 & -1 \leq x \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

What is the mean square error of $V_{100}(X)$, the estimate of $\text{Var}[X]$ based on 100 independent observations of X?

Quiz 7.3 Solution

Define the random variable $W = (X - \mu_X)^2$. Observe that $V_{100}(X) = M_{100}(W)$. By Theorem 7.6, the mean square error is

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

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

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

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

Therefore $\text{Var}[W] = E[W^2] - \mu_W^2 = 1/5 - (1/3)^2 = 4/45$ and the mean square error is $4/4500 = 0.000889$.

7.4 Confidence Intervals

Theorem 7.12

For any constant $c > 0$,

(a) $P[|M_n(X) - \mu_X| \geq c] \leq \dfrac{\mathrm{Var}[X]}{nc^2} = \alpha$,

(b) $P[|M_n(X) - \mu_X| < c] \geq 1 - \dfrac{\mathrm{Var}[X]}{nc^2} = 1 - \alpha$.

Proof: Theorem 7.12

Let $Y = M_n(X)$. Theorem 7.1 states that

$$E[Y] = E[M_n(X)] = \mu_X \qquad \mathrm{Var}[Y] = \mathrm{Var}[M_n(X)] = \mathrm{Var}[X]/n.$$

Theorem 7.12(a) follows by applying the Chebyshev inequality (Theorem 7.3) to $Y = M_n(X)$. Theorem 7.12(b) is just a restatement of Theorem 7.12(a) since

$$P[|M_n(X) - \mu_X| \geq c] = 1 - P[|M_n(X) - \mu_X| < c].$$

Example 7.6 Problem

Suppose we perform n independent trials of an experiment and we use the relative frequency $\hat{P}_n(A)$ to estimate $P[A]$. Use the Chebyshev inequality to calculate the smallest n such that $\hat{P}_n(A)$ is in a confidence interval of length 0.02 with confidence 0.999.

Example 7.6 Solution

Recall that $\hat{P}_n(A)$ is the sample mean of the indicator random variable X_A. Since X_A is Bernoulli with success probability $P[A]$, $E[X_A] = P[A]$ and $\text{Var}[X_A] = P[A](1 - P[A])$. Since $E[\hat{P}_n(A)] = P[A]$, Theorem 7.12(b) says

$$P\left[\left|\hat{P}_n(A) - P\left[A\right]\right| < c\right] \geq 1 - \frac{P\left[A\right]\left(1 - P\left[A\right]\right)}{nc^2}.$$

In Example 7.8, we observed that $p(1 - p) \leq 0.25$ for $0 \leq p \leq 1$. Thus $P[A](1 - P[A]) \leq 1/4$ for any value of $P[A]$ and

$$P\left[\left|\hat{P}_n(A) - P\left[A\right]\right| < c\right] \geq 1 - \frac{1}{4nc^2}.$$

For a confidence interval of length 0.02, we choose $c = 0.01$. We are guaranteed to meet our constraint if

$$1 - \frac{1}{4n(0.01)^2} \geq 0.999.$$

Thus we need $n \geq 2.5 \times 10^6$ trials.

Example 7.7 Problem

Suppose we perform n independent trials of an experiment. For an event A of the experiment, use the Chebyshev inequality to calculate the number of trials needed to guarantee that the probability the relative frequency of A differs from $P[A]$ by more than 10% is less than 0.001.

Example 7.7 Solution

In Example 7.6, we were asked to guarantee that the relative frequency $\hat{P}_n(A)$ was within $c = 0.01$ of $P[A]$. This problem is different only in that we require $\hat{P}_n(A)$ to be within 10% of $P[A]$. As in Example 7.6, we can apply Theorem 7.12(a) and write

$$P\left[\left|\hat{P}_n(A) - P[A]\right| \geq c\right] \leq \frac{P[A](1 - P[A])}{nc^2}.$$

We can ensure that $\hat{P}_n(A)$ is within 10% of $P[A]$ by choosing $c = 0.1P[A]$. This yields

$$P\left[\left|\hat{P}_n(A) - P[A]\right| \geq 0.1P[A]\right] \leq \frac{(1 - P[A])}{n(0.1)^2 P[A]} \leq \frac{100}{nP[A]},$$

since $1 - P[A] \leq 1$. Thus the number of trials required for the relative frequency to be within a certain percent of the true probability is inversely proportional to that probability.

Example 7.8 Problem

Theorem 7.12(b) gives rise to statements we hear in the news, such as,

> Based on a sample of 1103 potential voters, the percentage of people supporting Candidate Jones is 58% with an accuracy of plus or minus 3 percentage points.

The experiment is to observe a voter at random and determine whether the voter supports Candidate Jones. We assign the value $X = 1$ if the voter supports Candidate Jones and $X = 0$ otherwise. The probability that a random voter supports Jones is $E[X] = p$. In this case, the data provides an estimate $M_n(X) = 0.58$ as an estimate of p. What is the confidence coefficient $1 - \alpha$ corresponding to this statement?

Example 7.8 Solution

Since X is a Bernoulli (p) random variable, $E[X] = p$ and $\text{Var}[X] = p(1-p)$. For $c = 0.03$, Theorem 7.12(b) says

$$P\left[|M_n(X) - p| < 0.03\right] \geq 1 - \frac{p(1-p)}{n(0.03)^2} = 1 - \alpha.$$

We see that

$$\alpha = \frac{p(1-p)}{n(0.03)^2}.$$

Keep in mind that we have great confidence in our result when α is small. However, since we don't know the actual value of p, we would like to have confidence in our results regardless of the actual value of p. If we use calculus to study the function $x(1-x)$ for x between 0 and 1, we learn that the maximum value of this function is $1/4$, corresponding to $x = 1/2$. Thus for all values of p between 0 and 1, $\text{Var}[X] = p(1-p) \leq 0.25$. We can conclude that

$$\alpha \leq \frac{0.25}{n(0.03)^2} = \frac{277.778}{n}.$$

Thus for $n = 1103$ samples, $\alpha \leq 0.25$, or in terms of the confidence coefficient, $1-\alpha \geq 0.75$. This says that our estimate of p is within 3 percentage points of p with a probability of at least $1 - \alpha = 0.75$.

Example 7.9 Problem

Suppose X_i is the ith independent measurement of the length (in cm) of a board whose actual length is b cm. Each measurement X_i has the form

$$X_i = b + Z_i,$$

where the measurement error Z_i is a random variable with expected value zero and standard deviation $\sigma_Z = 1$ cm. Since each measurement is fairly inaccurate, we would like to use $M_n(X)$ to get an accurate confidence interval estimate of the exact board length. How many measurements are needed for a confidence interval estimate of b of length $2c = 0.2$ cm to have confidence coefficient $1 - \alpha = 0.99$?

Example 7.9 Solution

Since $E[X_i] = b$ and $\text{Var}[X_i] = \text{Var}[Z] = 1$, Equation (7.42) states

$$P\left[M_n(X) - 0.1 < b < M_n(X) + 0.1\right] \geq 1 - \frac{1}{n(0.1)^2} = 1 - \frac{100}{n}.$$

Therefore, $P[M_n(X) - 0.1 < b < M_n(X) + 0.1] \geq 0.99$ if $100/n \leq 0.01$. This implies we need to make $n \geq 10{,}000$ measurements. We note that it is quite possible that $P[M_n(X) - 0.1 < b < M_n(X) + 0.1]$ is much less than 0.01. However, without knowing more about the probability model of the random errors Z_i, we need $10{,}000$ measurements to achieve the desired confidence.

Theorem 7.13

Let X be a Gaussian (μ, σ) random variable. A confidence interval estimate of μ of the form

$$M_n(X) - c \leq \mu \leq M_n(X) + c$$

has confidence coefficient $1 - \alpha$ where

$$\alpha/2 = Q\left(\frac{c\sqrt{n}}{\sigma}\right) = 1 - \Phi\left(\frac{c\sqrt{n}}{\sigma}\right).$$

Proof: Theorem 7.13

We observe that

$$\begin{aligned} P\left[M_n(X) - c \le \mu_X \le M_n(X) + c\right] &= P\left[\mu_X - c \le M_n(X) \le \mu_X + c\right] \\ &= P\left[-c \le M_n(X) - \mu_X \le c\right]. \end{aligned}$$

Since $M_n(X) - \mu_X$ is a zero mean Gaussian random variable with variance σ_X^2/n,

$$\begin{aligned} P\left[M_n(X) - c \le \mu_X \le M_n(X) + c\right] &= P\left[\frac{-c}{\sigma_X/\sqrt{n}} \le \frac{M_n(X) - \mu_X}{\sigma_X/\sqrt{n}} \le \frac{c}{\sigma_X/\sqrt{n}}\right] \\ &= 1 - 2Q\left(\frac{c\sqrt{n}}{\sigma_X}\right). \end{aligned}$$

Thus $1 - \alpha = 1 - 2Q(c\sqrt{n}/\sigma_X)$.

Example 7.10 Problem

In Example 7.9, suppose we know that the measurement errors Z_i are iid Gaussian random variables. How many measurements are needed to guarantee that our confidence interval estimate of length $2c = 0.2$ has confidence coefficient $1 - \alpha \geq 0.99$?

Example 7.10 Solution

As in Example 7.9, we form the interval estimate

$$M_n(X) - 0.1 < b < M_n(X) + 0.1.$$

The problem statement requires this interval estimate to have confidence coefficient $1 - \alpha \geq 0.99$, implying $\alpha \leq 0.01$. Since each measurement X_i is a Gaussian $(b, 1)$ random variable, Theorem 7.13 says that $\alpha = 2Q(0.1\sqrt{n}) \leq 0.01$, or equivalently,

$$Q(\sqrt{n}/10) = 1 - \Phi(\sqrt{n}/10) \leq 0.005.$$

In Table 3.1, we observe that $\Phi(x) \geq 0.995$ when $x \geq 2.58$. Therefore, our confidence coefficient condition is satisfied when $\sqrt{n}/10 \geq 2.58$, or $n \geq 666$.

Example 7.11 Problem

Y is a Gaussian random variable with unknown expected value μ but known variance σ_Y^2. Use $M_n(Y)$ to find a confidence interval estimate of μ_Y with confidence 0.99. If $\sigma_Y^2 = 10$ and $M_{100}(Y) = 33.2$, what is our interval estimate of μ formed from 100 independent samples?

Example 7.11 Solution

With $1-\alpha = 0.99$, Theorem 7.13 states that

$$P\left[M_n(Y) - c \le \mu \le M_n(Y) + c\right] = 1 - \alpha = 0.99$$

where

$$\alpha/2 = 0.005 = 1 - \Phi\left(\frac{c\sqrt{n}}{\sigma_Y}\right).$$

This implies $\Phi(c\sqrt{n}/\sigma_Y) = 0.995$. From Table 3.1, $c = 2.58\sigma_Y/\sqrt{n}$. Thus we have the confidence interval estimate

$$M_n(Y) - \frac{2.58\sigma_Y}{\sqrt{n}} \le \mu \le M_n(Y) + \frac{2.58\sigma_Y}{\sqrt{n}}.$$

If $\sigma_Y^2 = 10$ and $M_{100}(Y) = 33.2$, our interval estimate for the expected value μ is $32.384 \le \mu \le 34.016$.

Quiz 7.4

X is a Bernoulli random variable with unknown success probability p. Using n independent samples of X and a central limit theorem approximation, find confidence interval estimates of p with confidence levels 0.9 and 0.99. If $M_{100}(X) = 0.4$, what is our interval estimate?

Quiz 7.4 Solution

Assuming the number n of samples is large, we can use a Gaussian approximation for $M_n(X)$. Since $E[X] = p$ and $\text{Var}[X] = p(1-p)$, we apply Theorem 7.13 which says that the interval estimate

$$M_n(X) - c \le p \le M_n(X) + c$$

has confidence coefficient $1 - \alpha$ where

$$\alpha = 2 - 2\Phi\left(\frac{c\sqrt{n}}{p(1-p)}\right).$$

We must ensure for every value of p that $1 - \alpha \ge 0.9$ or $\alpha \le 0.1$. Equivalently, we must have

$$\Phi\left(\frac{c\sqrt{n}}{p(1-p)}\right) \ge 0.95$$