Obtaining an estimate of a parameter is not the final purpose of statistical inference because it is highly unlikely that the population value of a parameter is equal to the estimate. We wish instead to know how close the population value is likely to be to an estimate. To determine this, we must estimate an interval, which uses more of the information in the data (the variance of the estimator) than does point estimation.

Construction of a Confidence Interval

Let $\{Y_i\}_{i=1}^n$ be a sequence of independent identically distributed $N(\mu, \sigma^2)$ random variables.

Hence

$$\bar{Y}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right),$$

so

$$\frac{\bar{Y}_n - \mu}{\sigma/\sqrt{n}} \sim N(0, 1).$$

(Diagram density)

From tabulated values of $N(0, 1)$

$$P \left( -1.96 \leq \frac{(\bar{Y}_n - \mu)}{\sigma/\sqrt{n}} \leq 1.96 \right) = .95.$$ 

Do the algebra one step at a time on the board.

Step 1

$$P \left( -1.96 \frac{\sigma}{\sqrt{n}} \leq \bar{Y}_n - \mu \leq 1.96 \frac{\sigma}{\sqrt{n}} \right) = .95.$$ 

Step 2

$$P \left( -\bar{Y}_n - 1.96 \frac{\sigma}{\sqrt{n}} \leq -\mu \leq -\bar{Y}_n + 1.96 \frac{\sigma}{\sqrt{n}} \right) = .95.$$ 

Step 3

$$P \left( \bar{Y}_n - 1.96 \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{Y}_n + 1.96 \frac{\sigma}{\sqrt{n}} \right) = .95.$$
Thus, the random interval

$$\left( \bar{Y}_n - 1.96 \frac{\sigma}{\sqrt{n}}, \bar{Y}_n + 1.96 \frac{\sigma}{\sqrt{n}} \right)$$

contains \( \mu \) 95 percent of the time.

Remark: Note the interval is random and \( \mu \) is fixed. It is clearer to state 95 percent of all intervals contain \( \mu \) rather than \( \mu \) falls in the interval 95 percent of the time.

For a given sample, we have the estimate \( \bar{y}_n \). If we replace the estimator with the estimate we have

$$\left( \bar{y}_n - 1.96 \frac{\sigma}{\sqrt{n}}, \bar{y}_n + 1.96 \frac{\sigma}{\sqrt{n}} \right),$$

which is a fixed interval. Such an interval either contains \( \mu \) or does not, so we cannot refer to the probability that the interval contains \( \mu \). To overcome the difficulty, we introduce the word confidence, which has the same practical connotation as probability.

We say

the confidence that \( \mu \) lies in \( \left( \bar{y}_n - 1.96 \frac{\sigma}{\sqrt{n}}, \bar{y}_n + 1.96 \frac{\sigma}{\sqrt{n}} \right) \) is .95

or

a 95 percent confidence interval for \( \mu \) is \( \left( \bar{y}_n - 1.96 \frac{\sigma}{\sqrt{n}}, \bar{y}_n + 1.96 \frac{\sigma}{\sqrt{n}} \right) \).

While confidence and probability are similar, they are not identical.

The link

If there is a high probability that an estimator lies near the population parameter, then there is a high degree of confidence that the population parameter lies near the estimate.

Confidence satisfies several probability axioms:

i) the confidence for any interval is nonnegative
ii) the confidence for the entire parameter space is 1

Key difference:
To derive the probability interval, we begin with a probability distribution, which uniquely gives the probability for any interval.

For confidence, we begin with confidence interval but cannot uniquely determine a confidence density and so cannot assign confidence to other intervals from the information contained in one interval.

Hypothesis Testing

Statistical tests are not as definitive as mathematical proofs, because sample data are subject to sampling error. If we cannot be absolutely sure a theory is true or false, what do we do? We could make probability statements such as

“Based on the available data, there is a 90 percent probability that the theory is true.”

Bayesians are willing to make such statements, but most of us believe a theory is either true or false, rather than “true 90 percent of the time”.

Rather than estimate the probability that a theory is true based on the observed data, we calculate the probability that we would observe such data if the theory were true. If this probability is low, then the data are not consistent with the theory and we therefore reject it. This is proof by statistical contradiction. Notice, too, that for a theory not to be rejected, it need only be consistent with the data. Such a conclusion is relatively weak, as many other theories may be consistent with the data.

For proof by statistical contradiction, we first make an assumption, called the null hypothesis, about the population from which the sample is drawn. Typically, the null hypothesis is a “straw assumption” that we anticipate rejecting. (The term null hypothesis arises because in early development, the value put forward to reject was zero.)

The alternative hypothesis describes the population if the null hypothesis is not true.

Remark: There are two general classes of assumptions: 1) concerns the form of the probability distribution from which the population is drawn (i.e. normal
vs. lognormal); 2) concerns the parameters from a given probability distribution. We focus on 2).

Remark: The null and alternative hypotheses can be simple or composite. A simple hypothesis specifies the values of all parameters of the population distribution. The major theorem on hypothesis testing is proven only for simple hypotheses, although virtually all applications are of composite hypotheses.

Example: Let \( \{Y_i\}_{i=1}^n \) be a sequence of independent identically distributed \( N(\mu, 1) \) random variables. (The only unknown parameter is \( \mu \).) Simple null and alternative hypotheses are: \( H_0 : \mu = 0 \) and \( H_1 : \mu = .5 \). To see how easily composite hypotheses arise, note \( H_1 : \mu \neq 0 \) is composite, as would be both hypotheses above if the variance were unknown. In what follows we consider the simple null and composite alternative

\[
H_0 : \mu = 0 \quad \text{and} \quad H_1 : \mu \neq 0.
\]

The alternative hypothesis is two sided. If, before seeing the data, we could rule out that the population mean is positive (or negative), the alternative hypothesis would be one sided (\( H_1 : \mu < 0 \)).

Once we have specified our null and alternative hypotheses, we collect and examine our sample data. As we are concerned with the value of \( \mu \), the population mean, we calculate the mean of our random sample, because this is what we would use to estimate the value of the population mean. The probability that we obtain a sample mean identically equal to the value in \( H_0 \) is zero, so it would be unwise to reject \( H_0 \) simply because the estimate does not equal the hypothesized value.

The farther the sample mean is from the hypothesized value, the more persuasive is the evidence against the null hypothesis. How far is far enough to be statistically persuasive? Recall

\[
\frac{\bar{Y}_n - \mu}{1/\sqrt{n}} \sim N(0, 1),
\]

(remember, the variance is 1 here) so

\[
P\left(\bar{Y}_n - 1.96 \frac{1}{\sqrt{n}}, \bar{Y}_n + 1.96 \frac{1}{\sqrt{n}}\right) \text{ contains } \mu \text{ is 95 percent.}
\]

(We could choose other critical values to obtain a different significance level.)
Under the null hypothesis $\mu = 0$, so if the null hypothesis is true, 95 percent of all samples will yield an interval $\left( \bar{Y}_n - 1.96 \frac{1}{\sqrt{n}}, \bar{Y}_n + 1.96 \frac{1}{\sqrt{n}} \right)$ that contains 0. Thus we check to see if the estimated confidence interval contains 0. If it does not, we conclude that the observed difference between $\bar{y}_n$ and $H_0$ is too large to be attributed to chance alone and is statistically significant.

If the estimated confidence interval does contain 0, we cannot reject $H_0$. Not rejecting a null hypothesis is not at all the same as proving the null hypothesis to be true. An unrejected null hypothesis is but one of many parameter values that are consistent with the data. Thus we must be careful to say “the data do not reject the null hypothesis” rather than “the data allow us to accept the null hypothesis to be true”.

Given the inherent uncertainty of statistical proof by contradiction, what types of mistakes are we subject to? First, because the probability that our random interval contains the true parameter value is less than 1, we could reject the null hypothesis even though the null hypothesis is true. Because we reject the null hypothesis if the estimated confidence interval does not contain the hypothesized value

$$P(\text{reject } H_0 | H_0 \text{ is true}) = 5 \text{ percent},$$

which is termed the size of the test.

Second, it is possible that our random interval contains the hypothesized value even though the hypothesized value is false. Clearly, such an outcome is more likely if the population value is close to the hypothesized value.

$$P(\text{fail to reject } H_0 | H_0 \text{ is false}) = 1 - \text{ power}$$

We can reduce the size of the test by selecting a large interval (diagram) (a 99 percent confidence interval reduces the size to 1 percent)

Yet the confidence interval is now longer, so we are more likely to fail to reject the null when it is false. Such a result is the heart of the Neyman-Pearson lemma that states we cannot simultaneously reduce both errors.

(diagram power curve)
The above calculations all assume that the variance is known. In practice, the variance is unknown. We replace the known variance with an estimator of the variance, $S^2$. It can be shown that

$$\frac{(n - 1) S^2}{\sigma^2} \sim \chi^2_{n-1}$$

and that

$$\sqrt{n} \frac{\bar{Y}_n - \mu}{\sigma}$$

is independent of

$$\frac{(n - 1) S^2}{\sigma^2}.$$

We then call on a theorem, which states that

If $Y \sim N(0, 1)$, $Z \sim \chi^2_{n-1}$, and $Y$ is independent of $Z$, then

$$\frac{Y}{\sqrt{\frac{Z}{n-1}}} \sim t_{n-1}.$$

Thus

$$\frac{\sqrt{n} \frac{\bar{Y}_n - \mu}{\sigma}}{\sqrt{\frac{n-1}{n} s}} = \frac{\sqrt{n} (\bar{Y}_n - \mu)}{s} \sim t_{n-1}.$$