Confidence Intervals and Hypothesis Tests

H. Krieger, Mathematics 62, Harvey Mudd College

Fall, 2008

Overview: Recall that a point estimator for a parameter of interest is a statistic which is a function of the random variables from which our sampling method chooses values. The properties of such an estimator, such as having a small bias or a small mean square error, are usually dependent only upon certain features of the sampling distribution of the statistic, such as its mean or variance. However, knowing (or assuming) the nature of that sampling distribution will allow us to make probability statements about the accuracy of our point estimator or test hypotheses about the value of that parameter. These statements are often be described in terms of “confidence intervals”. Here’s the basic idea. Suppose we have a point estimator $\hat{\Theta} = H(X_1, X_2, \ldots, X_n)$ for a parameter $\theta$ of the common distribution of the random sample $X_1, X_2, \ldots, X_n$. Using the sampling distribution of this point estimator assuming a particular value of the parameter, we can construct a statistic with values which are the endpoints of an interval which will contain that value of the parameter (we often say “cover the parameter”) with a certain probability or confidence level that we can prescribe in advance. After we perform the experiment, we have a numerical interval which can be regarded as an interval estimate of the parameter with our specified confidence level. Alternatively, we can see whether or not this interval covers a hypothesized value for the parameter and decide on this basis whether or not to reject that hypothesis.

Let’s give an example of this using one of our most important estimators, the sample mean. Recall that the sample mean $\bar{X}_n = (X_1 + X_2 + \cdots + X_n)/n$ has expected value equal to the population mean $\mu$ and has variance $\sigma^2/n$. We can regard this statistic as an estimator for the population mean. If we assume our population is normally distributed with known variance or the sample size is so large (say, $n \geq 50$,) that the central limit theorem is applicable and the variance is accurately approximated by the sample variance, we can assume the sampling distribution of $\bar{X}_n$ is $N(\mu, \sigma^2/n)$, where $\mu$ is our parameter of interest. In this case, the standardized random variable

$$Z = \frac{\bar{X}_n - \mu}{\sqrt{\sigma^2/n}} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}$$
has a \( N(0, 1) \) distribution. Therefore, given some “small” \( \alpha \in (0, 1) \), e.g. \( \alpha = 0.05 \) or \( \alpha = 0.01 \), we know that

\[
P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = P \left( -z_{\alpha/2} \leq \frac{\sqrt{n}(X_n - \mu)}{\sigma} \leq z_{\alpha/2} \right) = 1 - \alpha.
\]

When we rewrite the inequality describing this event which has probability \( 1 - \alpha \), we get the result

\[
P \left( \bar{X}_n - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right) = 1 - \alpha.
\]

The way to read this last statement is: the random interval

\[
\left[ \bar{X}_n - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{X}_n + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right]
\]
covers \( \mu \) with probability \( 1 - \alpha \). Consequently, the interval produced when the end points are computed using the sample values \( x_1, x_2, \ldots, x_n \) is called a two-sided \( 100(1 - \alpha)\% \) confidence interval for \( \mu \).

To look at this result in terms of a hypothesis test, suppose that there is a hypothesis (usually called the null hypothesis) that we would like to refute on the basis of our sampling. In this case, the hypothesis would be a particular value, say \( \mu_0 \), for the population mean. We now know that with probability \( 1 - \alpha \), our confidence interval covers \( \mu_0 \) if the null hypothesis is true. Consequently, if the confidence interval produced by our sample does not cover \( \mu_0 \), we can reject this hypothesis with \( 100(1 - \alpha)\% \) confidence, i.e. we will only be making a mistake \( 100\alpha\% \) of the time.

**More examples of confidence intervals:**

1. Under exactly the same assumptions as in the above example, a one-sided \( 100(1 - \alpha)\% \) confidence interval estimator of the population mean \( \mu \) can be obtained by simply using a one-sided probability statement in place of the type used to construct the two-sided interval. For example,

\[
P \left( \frac{\sqrt{n}(X_n - \mu)}{\sigma} \leq z_{\alpha} \right) = 1 - \alpha
\]

is equivalent to

\[
P \left( \mu \geq \bar{X}_n - z_{\alpha} \frac{\sigma}{\sqrt{n}} \right) = 1 - \alpha.
\]

In other words, the one-sided random interval

\[
\left[ \bar{X}_n - z_{\alpha} \frac{\sigma}{\sqrt{n}}, +\infty \right)
\]
covers \( \mu \) with probability \( 1 - \alpha \) and the interval produced when the end point is computed using the sample values \( x_1, x_2, \ldots, x_n \) is a one-sided \( 100(1 - \alpha)\% \) confidence interval for \( \mu \).
2. If we assume our sample consists of i.i.d. normal random variables with both the mean and variance unknown, the estimator $S^2_n$ of the population variance $\sigma^2$ can be used to construct a confidence interval for this parameter. In fact, we know that $C^2_n = (n - 1)S^2_n / \sigma^2$ has a Chi-Square($n - 1$) distribution. Consequently, probability statements for this statistic can be rearranged to produce the random intervals which cover the true value of $\sigma^2$ at a given level of confidence. For example,

$$P \left( \chi^2_{1 - \alpha/2, n-1} \leq (n - 1)\frac{S^2_n}{\sigma^2} \leq \chi^2_{\alpha/2, n-1} \right) = 1 - \alpha$$

is equivalent to

$$P \left( (n - 1)\frac{S^2_n}{\chi^2_{\alpha/2, n-1}} \leq \sigma^2 \leq (n - 1)\frac{S^2_n}{\chi^2_{1 - \alpha/2, n-1}} \right) = 1 - \alpha.$$ 

In other words, the random interval

$$\left[ (n - 1)\frac{S^2_n}{\chi^2_{\alpha/2, n-1}}, (n - 1)\frac{S^2_n}{\chi^2_{1 - \alpha/2, n-1}} \right]$$

covers $\sigma^2$ with probability $1 - \alpha$ and the interval produced when the end point is computed using the sample values $x_1, x_2, \ldots, x_n$ is a two-sided $100(1 - \alpha)$% confidence interval for $\sigma^2$. We can also obtain one-sided confidence intervals for $\sigma^2$, as we did for $\mu$, by starting with one-sided probability statements involving $(n - 1)\frac{S^2_n}{\sigma^2}$.

3. Once again, we assume our sample consists of i.i.d. normal random variables with both the mean and variance unknown. We want to estimate the population mean when our sample size $n < 50$ and we can’t use the sample variance as a good estimate of the population variance. In this case, we know that the statistic

$$T_n = \frac{\bar{X}_n - \mu}{S_n / \sqrt{n}}$$

has a $t_{n-1}$ distribution. Therefore, as we’ve already seen, probability statements for this statistic can be used to construct a confidence interval for $\mu$. Proceeding much as we did in the case when the population variance $\sigma^2$ was known and the corresponding statistic was distributed $N(0, 1)$, we see that

$$1 - \alpha = P(-t_{\alpha/2, n-1} \leq T_n \leq t_{\alpha/2, n-1})$$

$$= P \left( -t_{\alpha/2, n-1} \leq \frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} \leq t_{\alpha/2, n-1} \right),$$

which is equivalent to

$$P \left( \bar{X}_n - t_{\alpha/2, n-1} \frac{S_n}{\sqrt{n}} \leq \mu \leq \bar{X}_n + t_{\alpha/2, n-1} \frac{S_n}{\sqrt{n}} \right) = 1 - \alpha.$$
In other words, the random interval

\[
\left[ \bar{X}_n - t_{\alpha/2,n-1} \frac{S_n}{\sqrt{n}}, \bar{X}_n + t_{\alpha/2,n-1} \frac{S_n}{\sqrt{n}} \right]
\]

covers \( \mu \) with probability \( 1 - \alpha \) and the interval produced when the end points are computed using the sample values \( x_1, x_2, \ldots, x_n \) is a two-sided 100(1 - \alpha)\% confidence interval for \( \mu \). Similarly, we can obtain one-sided confidence intervals for \( \mu \) by starting with one-sided probability statements involving \( T_n \).

4. Finally, suppose we want to compare the variances of two independent normal populations. We assume we have i.i.d. samples \( X_1, X_2, \ldots, X_m \) from the first \( N(\mu_X, \sigma^2_X) \) population and i.i.d. samples \( Y_1, Y_2, \ldots, Y_n \) from the second \( N(\mu_Y, \sigma^2_Y) \) population, where both population means and variances are unknown. In this case, we know that the statistic

\[
F = \frac{S^2_X}{\sigma^2_X} \frac{S^2_Y}{\sigma^2_Y}
\]

has an \( F_{m-1,n-1} \) distribution. Consequently,

\[
P(F_{1-\alpha/2,m-1,n-1} \leq F \leq F_{\alpha/2,m-1,n-1}) = 1 - \alpha.
\]

Since this probability statement is the same as

\[
P(F_{1-\alpha/2,m-1,n-1} \leq \frac{S^2_X}{\sigma^2_X} \frac{S^2_Y}{\sigma^2_Y} \leq F_{\alpha/2,m-1,n-1}) = 1 - \alpha,
\]

it can be rewritten as

\[
P\left( \frac{1}{F_{\alpha/2,m-1,n-1}} \frac{S^2_X}{S^2_Y} \leq \frac{\sigma^2_X}{\sigma^2_Y} \leq \frac{1}{F_{1-\alpha/2,m-1,n-1}} \frac{S^2_Y}{S^2_X} \right) = 1 - \alpha.
\]

Therefore, the random interval

\[
\left[ \frac{1}{F_{\alpha/2,m-1,n-1}} \frac{S^2_X}{S^2_Y}, \frac{1}{F_{1-\alpha/2,m-1,n-1}} \frac{S^2_Y}{S^2_X} \right]
\]

covers the ratio \( \sigma^2_X/\sigma^2_Y \) with probability \( 1 - \alpha \).

**Applications to hypothesis testing**

Often in science, medicine, or other human enterprise, there is an accepted hypothesis about the nature of a certain activity. However, new information based on observations of the activity or from controlled experiments may result in a challenge to this status quo hypothesis. Statistical hypothesis testing is a quantitative approach to this situation. Once again, suppose we have a point estimator \( \hat{\Theta} = H(X_1, X_2, \ldots, X_n) \) for a parameter \( \theta \) of the common distribution of the random sample \( X_1, X_2, \ldots, X_n \). The status quo or “null” hypothesis, which we denote by \( H_0 \), asserts that the value of this parameter is
a particular value $\theta_0$. Our objective is to use the value of our statistic as evidence for the rejection of $H_0$. One method is to reject $H_0$ if the $100(1 - \alpha)%$ confidence interval we obtained for $\theta$ does not cover $\theta_0$, i.e. does not cover the value of $\theta$ when $H_0$ is true. Consequently, any time we reject $H_0$ by this method, there is an error $\alpha$, referred to as the “type 1” error. Notice, however, that this method rejects $H_0$ whenever a particular statistic, based on $\Theta$, falls outside of a deterministic interval determined by the distribution of that statistic under the assumption that $\theta = \theta_0$. This complementary region is referred to as the “rejection” region and, hence, the probability that the value of this “test” statistic falls in the rejection region when $H_0$ holds is $\alpha$, the type 1 error.

On the other hand, we ordinarily have an alternative hypothesis in mind, call it $H_a$, something that seems more realistic or desirable. This alternative could be very general, such as $\theta \neq \theta_0$, one-sided, such as $\theta > \theta_0$, or very specific, such as $\theta = \theta_a \neq \theta_0$. In the latter case, if the alternative hypothesis is true, we can consider the probability that the test statistic falls in the rejection region, so that we, in fact, do reject the null hypothesis. This probability of rejecting the null hypothesis computed under the assumption that the alternative $\theta = \theta_a \neq \theta_0$ is true, is called the “power” of the test and we would like this to be as large as possible given the constraint put on the rejection region by the choice of $\alpha$. Equivalently, we would like the complementary probability of not rejecting the null hypothesis when $H_a$ is true to be small. This probability is called the “type 2” error and is denoted by $\beta$.

We now give three basic examples of such hypothesis tests, using the results we have already described for confidence intervals.

(a) For this example, we assume our population is normally distributed with *known* variance or the sample size is so large (say, $n \geq 50$,) that the central limit theorem is applicable and the variance is accurately approximated by the sample variance. Thus we can assume the sampling distribution of $\overline{X}_n$ is $N(\mu, \sigma^2/n)$, where $\mu$ is the parameter of interest. Let’s test the hypothesis $H_0$ that $\mu = \mu_0$ against the alternative $H_a$ that $\mu > \mu_0$ with type 1 error equal to $\alpha$. We know that under $H_0$,

$$P \left( \frac{\sqrt{n}(\overline{X}_n - \mu_0)}{\sigma} \leq z_\alpha \right) = 1 - \alpha$$

or, equivalently,

$$P \left( \overline{X}_n \leq \mu_0 + \frac{\sigma}{\sqrt{n}} z_\alpha \right) = 1 - \alpha.$$  

Consequently, our rejection region for $H_0$ in terms of $\overline{X}_n$ is the interval

$$\left( \mu_0 + \frac{\sigma}{\sqrt{n}} z_\alpha, +\infty \right).$$
In other words, if the value \( x_n \) of our sample mean statistic is greater than
\[
\mu_0 + \frac{\sigma}{\sqrt{n}} z_{\alpha},
\]
then we reject \( H_0 \); otherwise, we do not reject this hypothesis. By this choice, we will make an error only \( \alpha \)% of the time. Equivalently, we can write our rejection region in terms of the random variable
\[
Z = \left( \overline{X}_n - \mu_0 \right) / \sigma
\]
as the interval \((z_{\alpha}, +\infty)\). In other words, if the value of \( Z, z = \sqrt{n}(\overline{X}_n - \mu_0)/\sigma \), obtained from our sample is greater than \( z_{\alpha} \), then we reject \( H_0 \). If we define the “p value” of our hypothesis test as the probability that \( Z \) is greater than or equal to \( z \), the value we obtained from our sample, we see that rejecting \( H_0 \) is equivalent to saying that the p value of the test is less than or equal to \( \alpha \).

If we are given a specific value for the alternative hypothesis, say \( \mu = \mu_a > \mu_0 \), we can compute the power of the test against this alternative and, equivalently, its complement, the type 2 error, \( \beta \). So we now assume this alternative and compute
\[
P \left( \overline{X}_n > \mu_0 + \frac{\sigma}{\sqrt{n}} z_{\alpha} \right)
\]
under this assumption. Now, however, the standardized random variable
\[
Z = \frac{\overline{X}_n - \mu_a}{\sqrt{\sigma^2/n}} = \frac{\sqrt{n}(\overline{X}_n - \mu_0)}{\sigma}
\]
has a \( N(0,1) \) distribution. So, if we standardize \( \overline{X}_n \) in the above probability expression, we get
\[
P \left( Z > \frac{\sqrt{n}(\mu_0 - \mu_a)}{\sigma} + z_{\alpha} \right) = 1 - \beta.
\]

(b) Next, we assume our sample consists of i.i.d. normal random variables with both the mean and variance unknown. The estimator \( S^2_n \) of the population variance \( \sigma^2 \) that we used to construct a confidence interval for this parameter can be used to test the hypothesis \( H_0 \) that \( \sigma^2 = \sigma_0^2 \) against the alternative \( H_a \) that \( \sigma^2 > \sigma_0^2 \). We know that \( C^2_n = (n-1)S^2_n/\sigma_0^2 \) has a Chi-Square \((n-1)\) distribution under \( H_0 \). Therefore,
\[
P \left( S^2_n/\sigma_0^2 \leq \chi^2_{\alpha,n-1} \right) = 1 - \alpha
\]
or, equivalently,
\[
P \left( S^2_n \leq \sigma_0^2 \chi^2_{\alpha,n-1} \right) = 1 - \alpha.
\]
This shows that our rejection region is
\[ \left( \sigma_0^2 \chi^2_{\alpha,n-1}, +\infty \right). \]

In other words, if the value \( s_n^2 \) of our sample variance statistic is greater than
\[ \sigma_0^2 \chi^2_{\alpha,n-1}, \]
then we reject \( H_0 \); otherwise, we do not reject this hypothesis. By this choice, we will make an error only \( \alpha \)% of the time. Equivalently, we can write our rejection region in terms of the random variable \( C^2_n \) as the interval \( (\chi^2_{\alpha,n-1}, +\infty) \). In other words, if the value of \( C^2_n \), \( \chi^2_n = (n-1)s_n^2/\sigma^2 \), obtained from our sample is greater than \( \chi^2_{\alpha,n-1} \), then we reject \( H_0 \).

If we are given a specific value for the alternative hypothesis, say \( \sigma^2 = \sigma_a^2 > \sigma_0^2 \), we can compute the power of the test against this alternative and, equivalently, its complement, the type 2 error, \( \beta \). So we now assume this alternative and compute
\[ P(S^2_n > \sigma_0^2 \chi^2_{\alpha,n-1}) \]
under this assumption. Now, however, the random variable \( C^2_n = S^2_n/\sigma_a^2 \) has a Chi-Square\((n-1)\) distribution. So, if we divide both sides of the inequality in the above probability expression by \( \sigma_a^2 \), we get
\[ P\left( C^2_n > \frac{\sigma_0^2}{\sigma_a^2} \chi^2_{\alpha,n-1} \right) = 1 - \beta. \]

(c) Finally, let’s assume again that sample consists of i.i.d. normal random variables with both the mean and variance unknown. The estimator for \( \mu \) that we used to construct confidence intervals in this case, namely
\[ T_n = \frac{\overline{X}_n - \mu}{S_n/\sqrt{n}}, \]
can be used to test the hypothesis \( H_0 \) that \( \mu = \mu_0 \) against the alternative \( H_a \) that \( \mu = \mu_a > \mu_0 \). We know that \( T_n \) has a \( t_{n-1} \) distribution. Therefore,
\[ 1 - \alpha = P(T_n \leq t_{\alpha,n-1}) = P\left( \frac{\sqrt{n}(\overline{X}_n - \mu_0)}{S_n} \leq t_{\alpha,n-1} \right). \]
Consequently, in terms of the random variable \( T_n \), the rejection region is the interval \( (t_{\alpha,n-1}, +\infty) \). That is, we reject \( H_0 \) if the value of \( T_n \), \( t_n = \sqrt{n}(\overline{X}_n - \mu_0)/s_n \), obtained from our sample is greater than \( t_{\alpha,n-1} \). By this choice we make an error only \( \alpha \)% of the time.