Confidence Intervals

Guy Lebanon

February 23, 2006

Confidence intervals (CI) is an important part of statistical inference. It refers to obtaining statements such as $P(a(X_1,\ldots,X_n) \le \theta \le b(X_1,\ldots,X_n)) = 1-\alpha$, where θ is the parameter of interest and a, b are quantities computed based on the iid sample $X_1,\ldots,X_n$. The probability $1-\alpha$ is called the confidence coefficient and $1-\alpha$ is typically taken to be $0.9, 0.95$ or 0.99. In contrast to point estimators $\hat{\theta}$ which give us a specific guess for θ, CIs provide an interval - which is less accurate than a specific number. The advantage of confidence intervals is that we can characterize the confidence of our statement $\theta \in [a,b]$. CIs of the form $(-\infty, b]$ or $[a,\infty)$ are called one-sided CIs (lower or upper).

In general, to construct a CI, we need to know some partial information concerning the unknown distribution - for example that it is a normal distribution. Such CIs are called small sample confidence intervals. If we can not make such an assumption we can still construct CIs by appealing to the central limit theorem. However, in this case, the CI will be only approximately correct - with the approximation improving in its quality as the sample size increases $n \to \infty$. Such CIs are called large sample CIs.

One of the most useful methods for constructing CIs is the method of pivotal quantities. This method constructs first CIs for an auxiliary quantity called a pivot, and then transforms the interval into a CI for the parameter θ.

Definition 1. *A pivot is a function of $\theta, X_1,\ldots,X_n$ whose distribution does not depend on θ.*

Typically, the chosen pivots $g(\theta, X_1,\ldots,X_n)$ have $N(0,1), \chi^2, t$ or F distributions. Since all of these distributions are well tabulated it is easy to obtain confidence intervals for the pivots

$$P(a \le g(\theta, X_1,\ldots,X_n) \le b) = 1-\alpha.$$

For example, if the pivot has a $N(0,1)$ distribution, $b = -a = z_{\alpha/2}$, which for $1-\alpha = 0.95$ is $z_{\alpha/2} = 1.96$. This last observation, together with the fact that for $N(\mu,\sigma^2), z_{\alpha/2} = \mu + 1.96\sigma$ is the source of the (not very good) practice of estimating the standard deviation of a RV by a quarter of the range of possible values (range-or-possible-values $\approx [\mu - 2\sigma, \mu + 2\sigma]$).

Transforming the pivot CI $P(a \le g(\theta, X_1,\ldots,X_n) \le b)$ to a θ CI $P(a(X_1,\ldots,X_n) \le \theta \le b(X_1,\ldots,X_n))$ may be done by

1. adding a real number to all three sides of the inequality
2. multiplying by a positive number all three sides of the inequality
3. multiplying by a negative number all three sides of the inequality (while reversing inequality signs)
4. taking the inverse $(\cdot)^{-1}$ of all three sides of the inequality (while reversing inequality signs).

Example: Suppose we have a single observation X from an exponential distribution whose expectation θ we are interested in. The transformation method may be used to show that X/θ is an exponential RV with parameter 1. That is X/θ is a pivot whose distribution does not depend on θ. We start by obtaining a confidence interval for the pivot (from tables of exponential distribution percentiles) $P(a \le X/\theta \le b) = 1-\alpha$ and proceed by dividing by X all three sides of the inequality and inverting to obtain a CI on θ:

$$P(a \le X/\theta \le b) = 1-\alpha \quad \Rightarrow \quad P(X/a \ge \theta \ge Y/b) = 1-\alpha.$$

Example: Suppose we have a single observation X from a uniform distribution $U([0,\theta])$ and we are interested in a confidence interval on θ. As before, the transformation method can be used to show that $X/\theta \sim U([0,1])$ and therefore a pivot. A lower 0.95 confidence interval for the pivot $0.95 = P(X/\theta \leq 0.95)$ transforms to a confidence interval on θ by dividing by X and taking the inverse of both sides $0.95 = P(X/\theta \leq 0.95) = P(\theta \geq X/0.95)$.

Large Sample Confidence Intervals for Means

Consider the case where we have an iid sample $X_1,\ldots,X_n$ (n is assumed to be large e.g., > 30) drawn from an unknown distribution with expectation μ. We are interested in constructing confidence intervals for μ using $\bar{X}$. Since we don't know the distribution of the sample we can't use the pivot method. The solution is to use the central limit theorem approximation to obtain a $N(0,1)$ pivot. More specifically, the CLT provides the following $N(0,1)$ pivot (approximately)

$$\sqrt{n}\frac{\bar{X}-\mu}{\sigma} = \frac{\sum_{i=1}^n (X_i-\mu)}{\sigma\sqrt{n}} \approx_{n\to\infty} Z \sim N(0,1).$$

We then first obtain confidence intervals for the pivot Z: $1-\alpha = P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2})$ and then transform it to an approximate CI on μ

$$1-\alpha = P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) \approx P\left(-z_{\alpha/2} \leq \sqrt{n}\frac{\bar{X}-\mu}{\sigma} \leq z_{\alpha/2}\right) = P\left(-\frac{\sigma z_{\alpha/2}}{\sqrt{n}} \leq \bar{X}-\mu \leq \frac{\sigma z_{\alpha/2}}{\sqrt{n}}\right)$$
$$= P\left(\bar{X} + \frac{\sigma z_{\alpha/2}}{\sqrt{n}} \geq \mu \geq \bar{X} - \frac{\sigma z_{\alpha/2}}{\sqrt{n}}\right).$$

If we don't know σ, the above CI may be approximated further using the estimator $S^2 \approx \sigma^2$ to yield

$$1-\alpha \approx P\left(\bar{X} + \frac{S z_{\alpha/2}}{\sqrt{n}} \geq \mu \geq \bar{X} - \frac{S z_{\alpha/2}}{\sqrt{n}}\right).$$

Above, we assumed that based on fixed α, n we calculated the resulting confidence interval. One could reverse the reasoning as follows. We may ask what is the sample size n that will provide a specific confidence interval $\theta \in [\bar{X}-a, \bar{X}+a]$ at a specific confidence level $1-\alpha$. In this case we should take

$$S z_{\alpha/2}/\sqrt{n} = a \Rightarrow \sqrt{n} = S z_{\alpha/2}/a \Rightarrow n \geq (S z_{\alpha/2}/a)^2,$$

where we use inequality since n has to be integer while $(S z_{\alpha/2}/a)^2$ is not necessarily an integer (If σ is known, it should replace S above).

Small Sample Confidence Intervals

If we know the distribution of the data we can do better than the large sample approximations based on the central limit theorem. Specifically, in this section we assume that $X_1,\ldots,X_n \sim N(\mu_1,\sigma^2), Y_1,\ldots,Y_m \sim N(\mu_2,\sigma^2)$. $\bar{X}$ and $\bar{Y}$ are as before and $S_1 = (n-1)^{-1}\sum_{i=1}^n (X_i-\bar{X})^2$ and $S_2 = (m-1)^{-1}\sum_{i=1}^m (Y_i-\bar{Y})^2$.

Confidence interval for μ_1: The pivot $\frac{\bar{X}-\mu_1}{S_1/\sqrt{n}}$ has a t distribution with $n-1$ dof. It leads to the CI $1-\alpha = P(-t_{\alpha/2} \leq \frac{\bar{X}-\mu_1}{S_1/\sqrt{n}} \leq t_{\alpha/2})$, which after simple manipulations yields

$$1-\alpha = P\left(\bar{X} - t_{\alpha/2}\frac{S_1}{\sqrt{n}} \leq \mu_1 \leq \bar{X} + t_{\alpha/2}\frac{S_1}{\sqrt{n}}\right).$$

Confidence interval for $\mu_1-\mu_2$: If $n=m$ the CI may be obtained by a simple derivation similar to the one above. However, if $n \neq m$ we need to be more careful. Recall that for $Z \sim N(0,1)$ and $W \sim \chi^2_\nu$, we have $\frac{Z}{\sqrt{W/\nu}} \sim t_{(\nu)}$. We will use the RV $\frac{Z}{\sqrt{W/\nu}} \sim t_{(\nu)}$ as a pivot with

$$Z = \frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\mathsf{Var}(\bar{X}-\bar{Y})}} = \frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\sigma^2/n+\sigma^2/m}} \sim N(0,1)$$

(this is a standartized normal RV since $\bar{X}-\bar{Y}$ is a linear combination of normal RVs and therefore is a normal RV, and we substract its mean and divide by its standard deviation). For W in the pivot $\frac{Z}{\sqrt{W/\nu}} \sim t_{(\nu)}$, we choose

$$W = \frac{(n-1)S_1^2}{\sigma^2} + \frac{(m-1)S_2^2}{\sigma^2} \sim \chi^2_{(n-1+m-1)} = \chi^2_{(n+m-2)}$$

(recall that a chi-squared RV $\chi^2_{(\nu)}$ is the same as a sum of ν squared standard normals $\sum_{j=1}^{\nu} Z_i$ and therefore the sum of $\frac{(n-1)S_1^2}{\sigma^2} \sim \chi^2_{(n-1)}$ and $\frac{(m-1)S_2^2}{\sigma^2} \sim \chi^2_{(m-1)}$ is the same as a sum of $n+m-2$ standard normal RVs which is $\chi^2_{(n+m-2)}$). Substituting Z and W above in the pivot $\frac{Z}{\sqrt{W/\nu}} \sim t_{(\nu)}$ gives the following CI

$$\begin{aligned} 1-\alpha =& P\left(-t_{\alpha/2} \leq \frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\sigma^2/n+\sigma^2/m}} / \sqrt{((n-1)S_1^2+(m-1)S_2^2)\sigma^{-2}(n+m-2)^{-1}} \leq t_{\alpha/2}\right) \\ =& P\left(-t_{\alpha/2} \leq \frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{1/n+1/m}} / \sqrt{((n-1)S_1^2+(m-1)S_2^2)(n+m-2)^{-1}} \leq t_{\alpha/2}\right) \\ =& P\left(-t_{\alpha/2} \leq \frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{S_p\sqrt{1/n+1/m}} \leq t_{\alpha/2}\right) \end{aligned}$$

using the notation $S_p = \sqrt{\frac{(n-1)S_1^2+(m-1)S_2^2}{n+m-2}}$ for the pooled (or weighted average) version of the two variance estimators. The above CI may be manipulated to obtain a CI for the desired parameter $\mu_1-\mu_2$

$$1-\alpha = P\left(\bar{X}-\bar{Y}-t_{\alpha/2}S_p\sqrt{1/n+1/m} \leq \mu_1-\mu_2 \leq \bar{X}-\bar{Y}+t_{\alpha/2}S_p\sqrt{1/n+1/m}\right).$$

Confidence Intervals for σ^2: We use the pivot $\frac{(n-1)S_1^2}{\sigma^2} \sim \chi^2_{(n-1)}$ to obtain the CI

$$1-\alpha = P\left(a \leq \frac{(n-1)S_1^2}{\sigma^2} \leq b\right) \quad \text{for appropriate } a,b \text{ chosen from the } \chi^2_{(n-1)} \text{ table}$$

(note that the pivot χ^2 distribution is not symmetric and is non-zero for positive numbers only; the resulting CI therefore is $[a,b]$ rather than a symmetric $[-a,a]$ as in the case of the t distribution pivots). Manipulating the above CI yields

$$1-\alpha = P\left(\frac{(n-1)S_1^2}{b} \leq \sigma^2 \leq \frac{(n-1)S_1^2}{a}\right).$$