8.5 Confidence Interval

Suppose $X_1, \ldots, X_n$ are a sample from a Normal population with mean μ and variance σ^2 (both unknown). Consider the following estimators $\bar{X}$ and S for μ and σ:

$$\bar{X} = \frac{1}{n}\sum X_i \qquad S = \sqrt{\frac{1}{n-1}\sum (X_i - \bar{X})^2}$$

Previously, we have shown than $\bar{X}$ is the MLE for μ, and that S^2 is an unbiased estimator for σ^2 (it turns out that S is neither the MLE, nor an unbiased estimator, for σ).

So if we are interested in estimating the value of μ given a sample, there are good reasons to use $\bar{X}$ to do so.

However, there is one drawback to using $\bar{X}$ to estimate μ (or really, using any estimator to estimate a paramter).

If $\hat{\theta}(\mathbf{X})$ has a continuous joint density function, then

$$P(\hat{\theta} = \theta) = 0.$$

This is the case if $\hat{\theta} = \bar{X}$, for example, since $\bar{X} \sim N(\mu, \sigma^2/n)$. That is, we know with absolute certainty that our estimated value of the parameter is incorrect.

To rectify this, we may instead consider **interval estimators** for a parameter. These interval estimators represent a pair of statistics $A(\mathbf{X}), B(\mathbf{X})$ with $A < B$ that we can use to bound the parameter θ. That is, we will produce an estimate of the form (A, B) and say that θ is likely to be in the interval (A, B). If choose the endpoints of interval wisely, then there will be a positive probability that the interval actually contains the unknown parameter:

$$P(A(\mathbf{X}) < \theta < B(\mathbf{X}))) > 0$$

Moreover, we can control the probability that the interval contains the parameter by manipulating the **length** $B - A$ and location of the interval.

We'll first see how to do this in the case of the Normal distribution. We know that the variable T given by

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

has t-distribution with $n-1$ degrees of freedom. For each $0 < \gamma < 1$, let c_γ be the constant so that

$$P(-c_\gamma < T < c_\gamma) = 1 - \gamma$$

Since T is symmetric around 0, then

$$\gamma = P(-c_\gamma < T < c_\gamma) = 1 - P(T \leq -c_\gamma) - P(T \geq c_\gamma) = 1 - 2P(T \geq c_\gamma)$$

and so

$$P(T \geq c_\gamma) = \frac{1-\gamma}{2} \iff 1 - P(T < c_\gamma) = \frac{1-\gamma}{2} \iff P(T < c_\gamma) = \frac{1+\gamma}{2}$$

Hence, c_γ is the $\frac{1+\gamma}{2}$ quantile of the $t(n-1)$ distribution; that is, $c_\gamma = F_{n-1}^{-1}\left(\frac{1+\gamma}{2}\right)$ where F_{n-1} is the CDF and F_{n-1}^{-1} is the quantile function of the $t(n-1)$ distribution.

Why are we interested in the quantiles of the t distribution?

Note that by arithmetic operations on inequalities,

$$\begin{aligned} -c_\gamma < T < c_\gamma \iff -c_\gamma < \frac{\bar{X} - \mu}{S/\sqrt{n}} < c_\gamma &\iff -\bar{X} - c_\gamma \frac{S}{\sqrt{n}} < -\mu < -\bar{X} + c_\gamma \frac{S}{\sqrt{n}} \\ &\iff \bar{X} + c_\gamma \frac{S}{\sqrt{n}} > \mu > \bar{X} - c_\gamma \frac{S}{\sqrt{n}} \end{aligned}$$

and so

$$\gamma = P(-c_\gamma < T < c_\gamma) = P\left(\bar{X} - c_\gamma \frac{S}{\sqrt{n}} < \mu < \bar{X} + c_\gamma \frac{S}{\sqrt{n}}\right)$$

Let A and B denote the random variables:

$$A = \bar{X} - c_\gamma \frac{S}{\sqrt{n}} \qquad B = \bar{X} + c_\gamma \frac{S}{\sqrt{n}}$$

Observe that since $\bar{X}$ and S are statistics, and since c_γ is a constant that only depends on the $t(n-1)$ distribution (and hence, depends only on the sample size), then A and B are statistics.

Moreover, with probability γ, the random variables A and B satisfy

$$A < \mu < B$$

Or, alternatively, the fixed parameter μ is in the random **interval** (A, B) with probability γ.

Def: Let $\mathbf{X}$ be a random sample from a distribution with parameter θ. Let $g(\theta)$ be a real-valued function of θ. Let $A \leq B$ be two statistics with the property that for all values of θ,

$$P(A < g(\theta) < B) \geq \gamma$$

Then the random interval (A, B) is called a confidence γ confidence interval for $g(\theta)$. If the inequality for γ is actually an equality, then the confidence interval is called **exact**. After the values of the random sample $\mathbf{X} = \mathbf{x}$ have been observed, and the values $A = a$ and $B = b$ are computed, the interval (a, b) is called the **observed value** of the confidence interval.

Thm: Let $\mathbf{X}$ be a random sample from $N(\mu, \sigma^2)$. Consider estimators $\bar{X}$ and S for μ and σ:

$$\bar{X} = \frac{1}{n}\sum X_i \qquad S = \sqrt{\frac{1}{n-1}\sum (X_i - \bar{X})^2}$$

For each $0 < \gamma < 1$, then interval (A, B) with the following endpoints is an exact confidence interval for μ:

$$A = \bar{X} - F_{n-1}^{-1}\left(\frac{1+\gamma}{2}\right)\frac{S}{\sqrt{n}} \qquad B = \bar{X} + F_{n-1}^{-1}\left(\frac{1+\gamma}{2}\right)\frac{S}{\sqrt{n}}$$

where F_{n-1}^{-1} is the quantile function for the t-distribution with $n-1$ degrees of freedom.

Ex 1: A batch of stout beer is best when it has an original gravity (OG) close to 1.071. The particular OG of a batch depends on a number factors (like temperature, rest time, recipe, etc.) but is (approximately) Normally distributed. Suppose we sample 5 OG measurements from a batch of beer:

1.067 1.060 1.077 1.072 1.067 with $\bar{x} = 1.0686$ and $s = 0.0064$

Using this sample of 5 OG measurements from a batch of beer, construct a 95% confidence interval for μ.

For any given confidence level γ, it is possible to construct infinitely many confidence intervals (A, B) so that $P(A < \mu < B) = \gamma$. However, among all such intervals, the symmetric interval has the shortest length.

But there are cases we may be interested in an asymmetric interval.

Def: A **one-sided γ confidence interval** for $g(\theta)$ is a random interval of the form (A, ∞) of $(-\infty, B)$ so that

$$P(A < g(\theta)) \geq \gamma \qquad \text{or } P(g(\theta) < B) \geq \gamma$$

That is, one-sided confidence intervals provide lower or upper bounds (but not both) for the parameter.

Ex 2: Using the sample of 5 OG measurements from a batch of beer, construct a 90% lower confidence interval for μ.

We must be careful interpreting confidence intervals. **Before** a sample is taken, we can make statements like "There is a 95% probability that the confidence interval (A, B) contains the unknown parameter μ". But **after** the sample $\mathbf{X} = x$ is observed and the confidence interval (a, b) computed, we cannot say "There is a 95% chance the mean μ is in the interval (a, b)."

Why? The latter statement contains no source of randomness. The only thing that is unknown is μ. We could, of course, adopt a Bayesian perspective and treat μ as random. But we would then need a prior for μ.

And moreover, it would still not likely be the case that there is 95% probability that μ is in (a, b), since this statement wouldn't reflect the update to our prior based on the observed data. For example, if we have reason to believe that μ is relatively small (for example, between -10^6 and 10^6) and we obtain a confidence interval of the form $(1.5 \cdot 10^9, 1.6 \cdot 10^9)$, we are unlikely to say this interval has a 95% chance of containing μ.