IEOR 165 – Lecture 17
Confidence Intervals

1 Definition of Confidence Interval

Suppose we have a distribution $f(u; \nu, \phi)$ that depends on some parameters $\nu \in \mathbb{R}^p, \phi \in \mathbb{R}^q$. The distinction between ν and ϕ is that ν are parameters whose values we care about, while ϕ are parameters whose values we do not care about. The ϕ are sometimes called *nuisance parameters*. The distinction can be situational: An example is a normal distribution $f(u; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-(x-\mu)/2\sigma^2)$; in some cases we might care about only the mean, only the variance, or both. We make the following definitions:

- A statistic $\underline{\nu}$ is called a level $(1-\alpha)$ *lower confidence bound* for ν if for all ν, ϕ,
$$\mathbb{P}(\underline{\nu} \leq \nu) \geq 1 - \alpha.$$
- A statistic $\overline{\nu}$ is called a level $(1-\alpha)$ *upper confidence bound* for ν if for all ν, ϕ,
$$\mathbb{P}(\overline{\nu} \geq \nu) \geq 1 - \alpha.$$
- The random interval $[\underline{\nu}, \overline{\nu}]$ formed by a pair of statistics $\underline{\nu}, \overline{\nu}$ is a level $(1-\alpha)$ *confidence interval* for ν if for all ν, ϕ,
$$\mathbb{P}(\underline{\nu} \leq \nu \leq \overline{\nu}) \geq 1 - \alpha.$$

These definitions should be interpreted carefully. A confidence interval does *not* mean that there is a $(1-\alpha)$ chance that the true parameter value lies in this range. Instead, a confidence interval at level $(1-\alpha)$ means that $100 \cdot (1-\alpha)\%$ of the time the computed interval contains the true parameter. The probabilistic statement should be interpreted within the frequentist framework of the frequency of correct answers.

Another point to note about confidence intervals is that they are in some sense dual to null hypothesis testing. Suppose we are interested in performing a one-sample location test. So, for instance, we would like to evaluate the null hypothesis
$$H_0 : X_i \sim \mathcal{N}(\mu_0, \sigma^2),$$
where μ_0 is known, and we would like to use a two-sided test. If we compute the $(1-\alpha)$ confidence interval and $\mu_0 \notin [\underline{\mu}, \overline{\mu}]$, then $p \leq \alpha$. Similar relationships hold between one-sided tests and upper/lower confidence bounds. If we compute the $(1-\alpha)$ lower confidence bound and $\mu_0 < \underline{\mu}$, then this is equivalent to a right-tailed test with $p \leq \alpha$. Additionally, if we compute the $(1-\alpha)$ upper confidence bound and $\mu_0 > \overline{\mu}$, then this is equivalent to a left-tailed test with $p \leq \alpha$.

2 Confidence Intervals for Gaussian

There are two key cases. In the first, the variance is known; in contrast, the variance is unknown in the second case.

2.1 Gaussian with Known Variance

Suppose $X_i \sim \mathcal{N}(\mu, \sigma^2)$, where σ^2 is known. Let $Z \sim \mathcal{N}(0, 1)$, and define the notation $z(1-\alpha)$ to be the value θ such that $\mathbb{P}(Z \leq \theta) = 1 - \alpha$. Note that by the properties of a Gaussian distribution, we have that $-z(\alpha) = z(1-\alpha)$. Equivalently, we have $\mathbb{P}(Z \geq \theta) = 1 - \alpha$ if $\mathbb{P}(Z \leq \theta) = 1 - \mathbb{P}(Z \geq \theta) = \alpha$; this can be rewritten as $\mathbb{P}(Z \geq z(\alpha)) = 1 - \alpha$. Then the three types of confidence intervals are

- Lower Confidence Bound:

$$\begin{aligned}
&\mathbb{P}\left(\sqrt{n} \cdot \Big(\frac{\overline{X} - \mu}{\sigma}\Big) \leq z(1-\alpha)\right) = 1 - \alpha \\
\Rightarrow &\mathbb{P}\left(\overline{X} - \sigma \cdot z(1-\alpha)/\sqrt{n} \leq \mu\right) = 1 - \alpha \\
\Rightarrow &\underline{\mu} = \overline{X} - \sigma \cdot z(1-\alpha)/\sqrt{n}.
\end{aligned}$$

  When rounding, we should round this value down.

- Upper Confidence Bound:

$$\begin{aligned}
&\mathbb{P}\left(\sqrt{n} \cdot \Big(\frac{\overline{X} - \mu}{\sigma}\Big) \geq z(\alpha)\right) = 1 - \alpha \\
\Rightarrow &\mathbb{P}\left(\overline{X} - \sigma \cdot z(\alpha)/\sqrt{n} \geq \mu\right) = 1 - \alpha \\
\Rightarrow &\overline{\mu} = \overline{X} + \sigma \cdot z(1-\alpha)/\sqrt{n}.
\end{aligned}$$

  When rounding, we should round this value up.

- Confidence Interval:

$$\begin{aligned}
&\mathbb{P}\left(\sqrt{n} \cdot \Big|\frac{\overline{X} - \mu}{\sigma}\Big| \leq z(1-\alpha/2)\right) = 1 - \alpha \\
\Rightarrow &\mathbb{P}\left(\overline{X} - \sigma \cdot z(1-\alpha/2)/\sqrt{n} \leq \mu \leq \overline{X} + \sigma \cdot z(1-\alpha/2)/\sqrt{n}\right) = 1 - \alpha \\
\Rightarrow &\underline{\mu} = \overline{X} - \sigma \cdot z(1-\alpha/2)/\sqrt{n}, \quad \overline{\mu} = \overline{X} + \sigma \cdot z(1-\alpha/2)/\sqrt{n}.
\end{aligned}$$

  When rounding, we should round $\underline{\mu}$ down and $\overline{\mu}$ up.

2.2 Example: Chemical pH Testing

Q: In one chemical process, it is very important that a particular solution used as a reactant have a pH of exactly 8.15. A method for determining pH available for solutions of this type is known to give measurements that are normally distributed with a mean equal to the actual pH and a standard deviation of 0.02. Suppose the sample size is 5, and we measure $\overline{X} = 8.26$. What is the level 0.9 confidence interval for the pH of the solution? If we conduct a hypothesis test to check if the pH is equal to 8.15, then what is the conclusion?

A: Since we want a level 0.9 confidence interval, this means we have $(1-\alpha) = 0.9$ or equivalently that $\alpha = 0.1$. The formula for $\underline{\mu}$ is

$$\begin{aligned}\underline{\mu} &= \overline{X} - \sigma \cdot z(1-\alpha/2)/\sqrt{n} = 8.26 - 0.02 \cdot z(1-0.1/2)/\sqrt{5} \\ &= 8.26 - 0.02 \cdot z(0.95)/\sqrt{5} = 8.26 - 0.02 \cdot 1.64/\sqrt{5} = 8.24.\end{aligned}$$

The formula for $\overline{\mu}$ is

$$\begin{aligned}\overline{\mu} &= \overline{X} + \sigma \cdot z(1-\alpha/2)/\sqrt{n} = 8.26 + 0.02 \cdot z(1-0.1/2)/\sqrt{5} \\ &= 8.26 + 0.02 \cdot z(0.95)/\sqrt{5} = 8.26 + 0.02 \cdot 1.64/\sqrt{5} = 8.28.\end{aligned}$$

Thus, the level 0.9 confidence interval for the pH is $[8.24, 8.27]$. For the hypothesis test, we reject the null hypothesis that the pH is exactly 8.15 because $8.15 \notin [8.24, 8.27]$.

2.3 Example: Toothpaste Effectiveness

Q: An advertisement for a new toothpaste claims that it reduces cavities of children in their cavity-prone years. Cavities per year for this age group are normal with mean 3 and standard deviation 1. A study of 2,500 children who used this toothpaste found an average of 2.95 cavities per child. Assume that the standard deviation of the number of cavities of a child using this new toothpaste remains equal to 1. What is the level 0.95 upper confidence bound? Is this data strong enough, at the 5 percent level of significance, to establish the claim of the toothpaste advertisement?

A: The formula for the upper confidence bound $\overline{\mu}$ is

$$\begin{aligned}\overline{\mu} &= \overline{X} + \sigma \cdot z(1-\alpha)/\sqrt{n} = 2.95 + 1 \cdot z(1-0.05)/\sqrt{2500} \\ &= 2.95 + 1 \cdot z(0.95)/\sqrt{2500} = 2.95 + 1 \cdot 1.62/\sqrt{2500} = 2.99.\end{aligned}$$

Thus, we should reject the null hypothesis that the new toothpaste is as good or worse than 3 cavities a year because $3 > 2.99$.

2.4 Gaussian with Unknown Variance

Suppose $X_i \sim \mathcal{N}(\mu, \sigma^2)$, where σ^2 is unknown. Let T_{n-1} be a random variable with the Student's t-distribution with $n-1$ degrees of freedom, and define the notation $t_{n-1}(1-\alpha)$ to be the value θ such that $\mathbb{P}(T_{n-1} \le \theta) = 1 - \alpha$. Note that by the properties of a t-distribution, we have that $-t_{n-1}(\alpha) = t_{n-1}(1-\alpha)$. Equivalently, we have $\mathbb{P}(T_{n-1} \ge \theta) = 1 - \alpha$ if $\mathbb{P}(T_{n-1} \le \theta) = 1 - \mathbb{P}(T_{n-1} \ge \theta) = \alpha$; this can be rewritten as $\mathbb{P}(T_{n-1} \ge t_{n-1}(\alpha)) = 1 - \alpha$.

Define the value

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2,$$

Then the three types of confidence intervals are

- Lower Confidence Bound:

  $$\underline{\mu} = \overline{X} - s \cdot t_{n-1}(1-\alpha)/\sqrt{n}.$$

  When rounding, we should round this value down.

- Upper Confidence Bound:

  $$\overline{\mu} = \overline{X} + s \cdot t_{n-1}(1-\alpha)/\sqrt{n}.$$

  When rounding, we should round this value up.

- Confidence Interval:

  $$\underline{\mu} = \overline{X} - s \cdot t_{n-1}(1-\alpha/2)/\sqrt{n}, \quad \overline{\mu} = \overline{X} + s \cdot t_{n-1}(1-\alpha/2)/\sqrt{n}.$$

  When rounding, we should round $\underline{\mu}$ down and $\overline{\mu}$ up.

2.5 Example: Diet and Lifespan

Q: The following data relate to the ages at death of a certain species of rats that were fed 1 of 3 types of diets. Thirty rats of a type having a short life span of an average of 17.9 months were randomly divided into 3 groups of 10 each. The sample means and variances of ages at death (in months) are:

	Very Low Calorie	Moderate Calorie	High Calorie
Sample mean	22.4	16.8	13.7
Sample variance	24.0	23.2	17.1

Compute a level 0.99 lower confidence bound for the life-span under a very low calorie diet. Would you accept a null hypothesis that life-span does not increase under a very low calorie diet at a significance level of 0.01?

A: Since we want a level 0.99 confidence interval, this means we have $(1-\alpha) = 0.99$ or equivalently that $\alpha = 0.01$. The formula for the lower confidence bound $\underline{\mu}$ is

$$\begin{aligned}\underline{\mu} &= \overline{X} - s \cdot t_{n-1}(1-\alpha)/\sqrt{n} = 22.4 - \sqrt{24} \cdot t_{10-1}(0.99)/\sqrt{10} \\ &= 22.4 - \sqrt{24} \cdot 2.82/\sqrt{10} = 18.03.\end{aligned}$$

We should reject the null hypothesis since $17.9 < 18.03$, and hence conclude that life-span increases under a very low calorie diet.

3 Bootstrap Confidence Intervals

One of the challenges with confidence intervals is that closed-form expressions may not be available for random variables with complicated distributions. Furthermore, in many cases, we may not even *a priori* know the distribution of the random variables. The *bootstrap* is a nonparametric technique from machine learning and statistics, and one of its possible applications is to computing confidence intervals. The idea of the bootstrap is to pretend that the sample distribution

$$\hat{F}(u) = \frac{1}{n}\sum_{i=1}^{n} 1(X_i \leq u)$$

is the real distribution, and then use this distribution to compute the quantities of interest. This might seem like an odd idea, but it does lead to correct answers in many cases. However, it is worth noting that there are cases where the bootstrap will fail and will give incorrect answers. However, for the cases where the bootstrap works, the intuition is that the sample distribution will be an accurate approximation of the true distribution when we have measured a large amount of data.

There are many variants of the bootstrap, and one popular version uses the Monte Carlo algorithm. To make the discussion more concrete, consider the Monte Carlo algorithm to estimate the mean:

1. For $j = 1, \ldots, n$
   (a) Randomly choose $i \in \{1, \ldots, n\}$;
   (b) Set $Y_j = X_i$;
2. Set the mean estimate to $\hat{\mu} = \frac{1}{n}\sum_{j=1}^{n} Y_j$.

Even when using a Monte Carlo algorithm, there are many variants to compute the confidence interval. One of the simpler methods is the *percentile bootstrap*. Let M be a large number, say $M = 1000$. The algorithm is as follows:

1. For $k = 1, \ldots, M$
   (a) For $j = 1, \ldots, n$
      i. Randomly choose $i \in \{1, \ldots, n\}$;
      ii. Set $Y_j = X_i$;
   (b) Set the k-th mean estimate to $\hat{\mu}_k = \frac{1}{n}\sum_{j=1}^{n} Y_j$.
2. Compute the $\alpha/2$-percentile of the $\hat{\mu}$ values, and call this $\underline{\mu}$;
3. Compute the $(1-\alpha/2)$-percentile of the $\hat{\mu}$ values, and call this $\overline{\mu}$.

Recall that the γ-percentile of a random variable V is the value θ such that $\mathbb{P}(V < \theta) \leq \gamma$ and $\mathbb{P}(V \leq \theta) \geq \gamma$. The range $[\underline{\mu}, \overline{\mu}]$ is a $(1-\alpha)$ confidence interval.