# Statistical Intervals Based on a Single Sample (Devore Chapter Seven)

MATH-252-01: Probability and Statistics II*

Spring 2019

## Contents





### Tuesday 22 January 2019

# 0 Introduction

## 0.1 Motivation

We continue our study of inferential statistics, which allow us to define procedures where the content of a random sample tells us something about the unknown parameters of the underlying probability distribution. For example, if we know that a random sample is drawn from a normal distribution with a known standard deviation $\sigma$ and an unknown mean $\mu$, we can make a

rule that says for this random sample, construct an interval of a specified width centered on the sample mean, and this interval will have a 95% probability of containing the true population mean $\mu$. This is known as a *confidence interval* for $\mu$.

As a specific example where this might be useful, when GPS was introduced back in the 20th century, the government added some random errors into the reported position, for national security purposes. So for example, if you put a GPS receiver on top of your house and attempted to measure the elevation above sea level, the GPS would return the true value plus a random error with a standard deviation of about 50 meters. This random error varied from day to day, so you could get a more accurate measurement by averaging together multiple measurements. If you took, say, 25 measurements, and used their sample mean to construct a 95% confidence interval on your elevation, there would be a 95% chance your interval included the true elevation, a 2.5% chance the whole interval was above the true elevation, and a 2.5% chance that it was below.

## 0.2 Reminder of Notation

If $Z$ is a standard normal random variable,

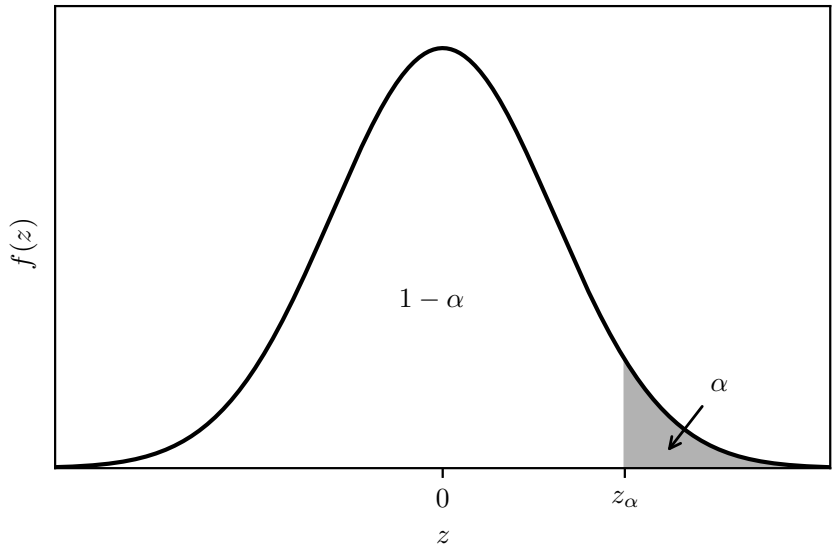
$$P(Z > z_\alpha) = \alpha \tag{0.1}$$

so

$$1 - \alpha = P(Z \leq z_\alpha) = \Phi(z_\alpha) \tag{0.2}$$

Note that because the standard normal distribution is symmetric,

$$\Phi(-z_\alpha) = 1 - \Phi(z_\alpha) = \alpha \tag{0.3}$$




# 1 Confidence Intervals Using Standard Normal Percentiles

## 1.1 Confidence Interval for the Mean of a Normal Population with Known Variance

Suppose $\{X_i\}$ is a random sample of size $n$ drawn from a distribution of mean $\mu$ and standard deviation $\sigma$. (I.e., for each $i$, $E(X_i) = \mu$ and $V(X_i) = \sigma^2$.) You showed in MATH 251 that the sample mean

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{1.1}$$

has mean $E(\overline{X}) = \mu$ and variance $V(\overline{X}) = \sigma^2/n$. You also saw that it is normally distributed under the following conditions:

1. Exactly, if the underlying probability distribution from which each $X_i$ is drawn is a normal distribution.
2. Approximately, for large $n$, by the Central Limit Theorem.

If one of those conditions holds, we can define the (approximately, in the latter case) standard normal random variable

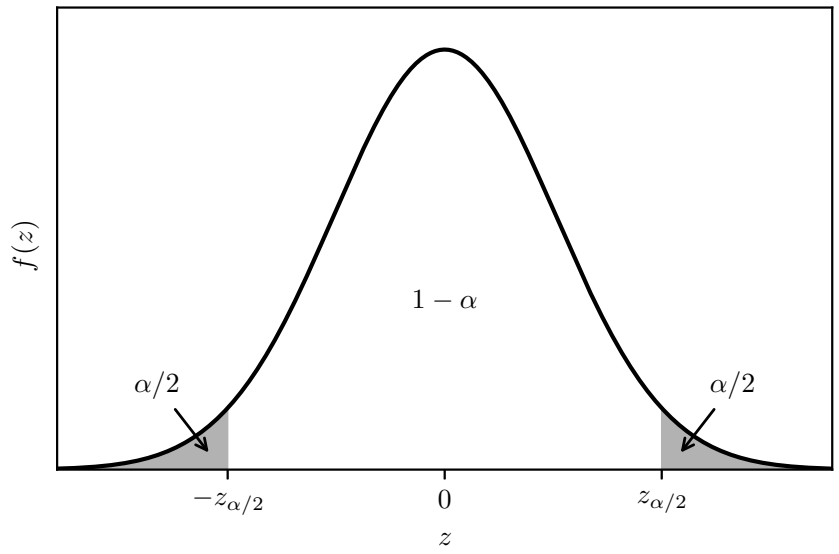
$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \tag{1.2}$$

Given a value $\alpha$ between 0 and 1, we can construct an interval that $Z$ has a probability $1 - \alpha$ of landing in. If we put the boundaries of the interval at $\pm z_{\alpha/2}$ we find

$$P(Z < -z_{\alpha/2}) = \Phi(-z_{\alpha/2}) = \alpha/2 \tag{1.3a}$$

$$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = \Phi(z_{\alpha/2}) - \Phi(-z_{\alpha/2}) = \left(1 - \frac{\alpha}{2}\right) - \frac{\alpha}{2} = 1 - \alpha \tag{1.3b}$$

$$P(z_{\alpha/2} < Z) = 1 - \Phi(z_{\alpha/2}) = 1 - \left(1 - \frac{\alpha}{2}\right) = \alpha/2 \tag{1.3c}$$




For example, taking $\alpha = 0.05$, since $\Phi(1.96) \approx 0.9750 \approx 0.975$ and therefore $z_{0.025} \approx 1.96$,

- $Z$ has a 2.5% chance of lying below $-1.96$
- $Z$ has a 95% chance of lying between $-1.96$ and 1.96
- $Z$ has a 2.5% chance of lying above 1.96

This has an interesting application to the situation where we know $\sigma$ but not $\mu$. We already know that $\overline{X}$ can be used to estimate $\mu$, but now we can construct an interval which has a good chance of containing $\mu$ as follows:

$$\begin{aligned}
1 - \alpha = P(-z_{\alpha/2} < Z < z_{\alpha/2}) &= P\left(-z_{\alpha/2} < \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} < z_{\alpha/2}\right) \\
&= P\left(-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \overline{X} - \mu < z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) \\
&= P\left(-\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < -\mu < -\overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) \\
&= P\left(\overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} > \mu > \overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) \\
&= P\left(\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)
\end{aligned} \tag{1.4}$$

The interpretation of this is a bit tricky: it's tempting to look at

$$P\left(\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha \tag{1.5}$$

or, specifically

$$P\left(\overline{X} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + 1.96\frac{\sigma}{\sqrt{n}}\right) \approx 0.95 \tag{1.6}$$

and think that if I take a sample, find $\overline{x}$, and construct the interval $(\overline{x} - 1.96\,\frac{\sigma}{\sqrt{n}}, \overline{x} + 1.96\,\frac{\sigma}{\sqrt{n}})$, that I can say there's a 95% probability that $\mu$ lies in that interval. This is wrong, though, since we constructed the interval in a frequentist picture in which $\overline{X}$ is a random variable, not $\mu$. From this perspective, the value of $\mu$, while it may not be known to us, is not random, and we're really not talking about the probability for $\mu$ to take on a particular value. Rather, we're saying that whatever the true, unknown value of $\mu$ is, if we collect a sample $\{x_i\}$ of size $n$ and construct an interval $(\overline{x} - 1.96\,\frac{\sigma}{\sqrt{n}}, \overline{x} + 1.96\,\frac{\sigma}{\sqrt{n}})$, it has a 95% chance of bracketing that true value.

## 1.2 Aside: Bayesian Statistics and Probability Theory

As an aside, this is not the only possible way to do things. There are situations (e.g., observational rather than experimental sciences) where you can't necessarily repeat the collection of a sample and take another shot at your confidence interval. In that case, there really is one set of values $\{x_i\}$ and you would like to state something about your degree of belief in different possible values of $\mu$. We can still cast all of this in the language of random variables if we want, as follows. Suppose we can assign a probability distribution to the value of $\mu$, and represent it by a random variable $\mathrm{M}$ ($\mathrm{M}$ is an uppercase $\mu$.) Given a value $\mu$ for $\mathrm{M}$, we construct a random sample $\{X_i\}$ of size $n$ using a normal distribution with that $\mu$, and then construct $\overline{X}$ from that sample (so it's doubly-random in that case). We've seen that

$$P\left(\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \,\middle|\, \mathrm{M} = \mu\right) = 1 - \alpha \quad (1.7)$$

We could ask instead, suppose we do the whole experiment and find that $\overline{X}$ equals some specific value $\overline{x}$. We could then ask for an interval $\mathcal{I}(\overline{x})$ such that[1]

$$P\left(\mathrm{M} \in \mathcal{I}(\overline{x}) | \overline{X} = \overline{x}\right) = 1 - \alpha \quad (1.8)$$

The details are beyond the scope of this course, but you can get a flavor of it by thinking about a simpler case where $\overline{X}$ and $\mathrm{M}$ are discrete random variables. Then, the usual frequentist probabilities are written in terms of some specific value $\mu$ and correspond to

$$P(\overline{X} = \overline{x} | \mathrm{M} = \mu) \quad (1.9)$$

but if we've done the experiment and have a specific $\overline{x}$, we'd really like to talk about

$$P(\mathrm{M} = \mu | \overline{X} = \overline{x}) \quad (1.10)$$

But this is just the sort of situation that Bayes's theorem was meant to handle. We can write

$$P(\mathrm{M} = \mu | \overline{X} = \overline{x}) = \frac{P(\overline{X} = \overline{x} | \mathrm{M} = \mu)\, P(\mathrm{M} = \mu)}{P(\overline{X} = \overline{x})} \quad (1.11)$$

This is the bread and butter of Bayesian probability (which Devore rather dismissively refers to as "subjective probability"). There are complications, notably in deciding what to assign as the prior probability distribution $P(\mathrm{M} = \mu)$, but it does allow you to actually talk about the probabilities of things you're interested in.

## 1.3 Sample Size Determination

Given a sample of size $n$, we see that the width of the interval we construct at confidence level $(1 - \alpha)$ will be

$$w = 2z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \quad (1.12)$$

[1] Actually, since you want to use all the information in the data, what you're looking for is $P\left(\mathrm{M} \in \mathcal{I}(\{x_i\}) | \{X_i = x_i\}\right) = 1 - \alpha$.

Often, though, we'll be in a position of being able to collect a bigger sample (possibly at greater expense) and therefore we'd like to know how big the sample should be to correspond to a certain width. If we know $\sigma$, and our desired confidence level, we require

$$\frac{\sigma}{\sqrt{n}} = \frac{w}{2z_{\alpha/2}} \tag{1.13}$$

i.e.,

$$\sqrt{n} = 2z_{\alpha/2}\,\frac{\sigma}{w} \tag{1.14}$$

or

$$n = \left(2z_{\alpha/2}\,\frac{\sigma}{w}\right)^2 \tag{1.15}$$

## 1.4 One-sided Intervals

The confidence interval

$$P\left(\overline{X} - z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha \tag{1.16}$$

is called a *two-sided interval* because the probability $\alpha$ for $\mu$ to lie inside the interval is split up, with a probability of $\alpha/2$ that it lies below and $\alpha/2$ that it lies above. It's also possible to define a *one-sided interval* with all of the probability on one side. For example,

$$P\left(\mu < \overline{X} + z_{\alpha}\,\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha \tag{1.17}$$

is an upper limit at confidence level $1-\alpha$ and

$$P\left(\overline{X} - z_{\alpha}\,\frac{\sigma}{\sqrt{n}} < \mu\right) = 1-\alpha \tag{1.18}$$

is a lower limit at confidence level $1-\alpha$.

## 1.5 General Formalism

A bit of notation is needed to generalize this prescription. We said that $\overline{X}$ could be used to estimate $\mu$. It is sometimes called an *estimator* $\hat{\mu}$ where the hat means "estimator" and we write it in blue to emphasize that it's a random variable.

If we want to talk about a general parameter, the convention is to call that parameter $\theta$. (In this case $\theta$ is just $\mu$).

Given a random sample $\{X_i\}$, we construct a random variable (statistic) $h(\{X_i\};\theta)$ whose probability distribution doesn't depend on $\mu$. (In this case $h(\{X_i\};\theta)$ is $h(\{X_i\};\mu) = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$, which obeys a standard normal distribution independent of the value of $\mu$.)

The ends of the confidence interval on $h(\{X_i\};\theta)$ are defined by

$$P\big(h(\{X_i\};\theta) < a\big) = \alpha/2 \tag{1.19a}$$
$$P\big(b < h(\{X_i\};\theta)\big) = \alpha/2 \tag{1.19b}$$

so that

$$P\big(a < h(\{X_i\};\theta) < b\big) = 1-\alpha \tag{1.20}$$

(In this case, $a$ is $-z_{\alpha/2}$ and $b$ is $z_{\alpha/2}$.)

The next step is to convert (if possible) the bounds on $h(\{X_i\};\theta)$ to bounds on $\theta$ itself,

$$P\big(\theta < l(\{X_i\})\big) = \alpha/2 \tag{1.21a}$$
$$P\big(u(\{X_i\}) < \theta\big) = \alpha/2 \tag{1.21b}$$

so that the lower and upper bounds $l(\{X_i\})$ and $u(\{X_i\})$ have a probability $\alpha$ of bracketing the true parameter value $\theta$. (In this case, $l(\{X_i\})$ is $\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ and $u(\{X_i\})$ is $\overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$.)

## 1.6 Confidence Interval for the Population Proportion

We can apply this method to estimation of the proportion of a large population satisfying some property, i.e., the parameter $p$ of a binomial distribution. If $X \sim \text{Bin}(n,p)$, we can also think of $X$ as being the sum $X = \sum_{i=1}^{n} B_i$ of $n$ Bernoulli random variables $\{B_i\}$ obeying $P(B_i = 1) = p$; $P(B_i = 0) = 1 - p \equiv q$. We know that $X$ has mean $E(X) = np$ and variance $V(X) = npq$ and that for $np \gtrsim 10$ and $nq \gtrsim 10$, the binomial distribution associated with $X$ is reasonably approximated by a normal distribution. This means that if we construct the random variable

$$\frac{X - np}{\sqrt{npq}} \tag{1.22}$$

then

$$\begin{aligned} 1 - \alpha &\approx P\left(-z_{\alpha/2} < \frac{X - np}{\sqrt{npq}} < z_{\alpha/2}\right) \\ &= P\left(-z_{\alpha/2} < \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} < z_{\alpha/2}\right) \end{aligned} \tag{1.23}$$

where we've defined the estimator $\hat{p} = X/n$. To get an interval for the parameter $p$, we want to solve for $p$ at the endpoints. Squaring the relation at either end of the interval gives

$$\frac{(\hat{p} - p)^2}{p(1-p)/n} = z_{\alpha/2}^2 \tag{1.24}$$

which can be converted into the quadratic equation

$$\left(1 - \frac{z_{\alpha/2}^2}{n}\right) p^2 - 2\left(\hat{p} + \frac{z_{\alpha/2}^2}{n}\right) + \frac{\hat{p}^2}{n} = 0 \tag{1.25}$$

Using the quadratic formula and a bit of algebra, we can find the roots

$$p_\pm = \frac{\hat{p} + \frac{z_{\alpha/2}^2}{2n} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z_{\alpha/2}^2}{4n^2}}}{1 + \frac{z_{\alpha/2}^2}{n}} \tag{1.26}$$

which gives the confidence interval on the parameter $p$ (the population proportion)

$$P(p_- < p < p_+) \approx 1 - \alpha \ . \tag{1.27}$$

You might have asked, if we're approximating the binomial distribution by a normal distribution anyway, why can't we just take $\hat{p} \pm z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$ as the CI limits? Unfortunately, that requires $n$ to be rather large, and so it's best to use the more complicated limits given in (1.26). (You don't need to memorize the form of this, but you should write it on your formula sheet for the exams.)

### Practice Problems

7.1, 7.5, 7.7, 7.15, 7.19, 7.23, 7.26

### Thursday 24 January 2019

# 2 Confidence Intervals When the Variance is Estimated

## 2.1 Large-Sample Intervals

Recall that last time we saw that given a random sample $\{X_i\}$ with known variance $\sigma^2 = V(X_i)$, we could construct the random variable

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \ ; \tag{2.1}$$

If we can state that this is a standard normal random variable, either because the underlying $\{X_i\}$ are known to be normally-distributed random variables, or approximately by virtue of the Central Limit Theorem, we can use it to set a confidence interval on the unknown value of $\mu$:

$$\begin{aligned} 1-\alpha &= P(-z_{\alpha/2} < Z < z_{\alpha/2}) \\ &= P\left(\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) \end{aligned} \tag{2.2}$$

What do we do if $\sigma$ is also unknown? Well, we know that the sample variance

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 \tag{2.3}$$

is an estimator for the variance $\sigma^2$, so what if we construct

$$\frac{\overline{X}-\mu}{S/\sqrt{n}} \; ? \tag{2.4}$$

If we're relying on the central limit theorem to make this approximately Gaussian for large $n$, it all basically works as before, and you can set a confidence interval $(\overline{X} - z_{\alpha/2}\frac{S}{\sqrt{n}}, \overline{X} + z_{\alpha/2}\frac{S}{\sqrt{n}})$ which has a $1-\alpha$ chance of containing $\mu$. The catch is that, since we're adding randomness by using $S$ instead of $\sigma$, we require $n \gtrsim 40$ rather than 30.

## 2.2 The $t$-Distribution

If $n$ is not large, but $\{X_i\}$ are iid normally-distributed random variables with unknown $\mu$ and $\sigma$, we can construct a random variable

$$T = \frac{\overline{X}-\mu}{S/\sqrt{n}} \tag{2.5}$$

This is not a standard normal random variable; it obeys something called a $t$-distribution, also known as a Student's[2] $t$-distribution, with $n-1$ degrees of freedom. Devore never actually writes down the probability distribution function; you don't need to memorize it, but in case you want to play around with plotting it, it's

$$f_T(t; n-1) \propto \left(1 + \frac{t^2}{n-1}\right)^{-n/2} \tag{2.6}$$

If you really want to know, with the proportionality constant written out it's

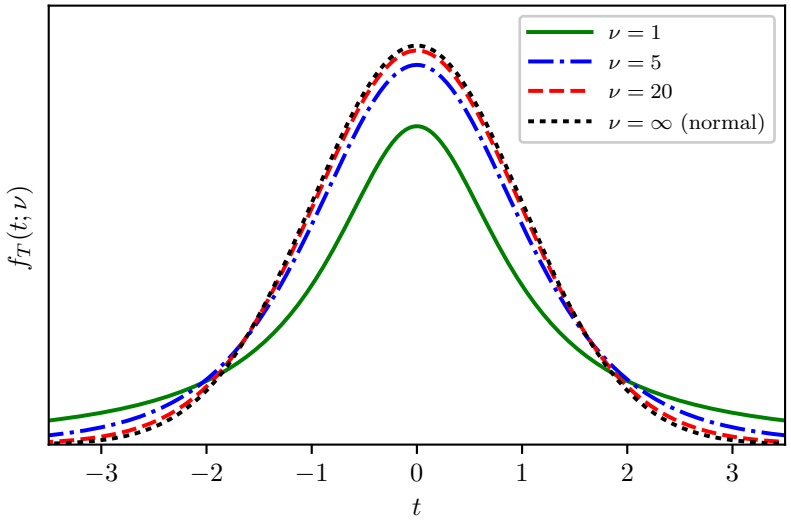
$$f_T(t;\nu) = \frac{\Gamma([\nu+1]/2)}{\sqrt{\nu\pi}\Gamma(\nu/2)}\left(1+\frac{t^2}{\nu}\right)^{-[\nu+1]/2} \tag{2.7}$$

Note that the number of degrees of freedom (df) is $\nu = n-1$; this sort of makes sense, because when $n = 1$, we can't define $S$, and the whole thing goes haywire. Note also that as $\nu$ or equivalently $n$ becomes large, the distribution does tend towards a standard normal distribution:

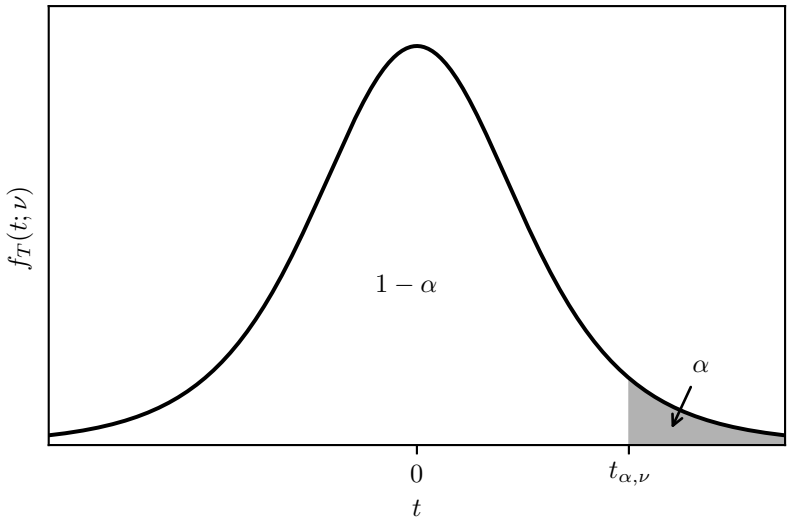
$$\begin{aligned} \lim_{n\to\infty} f_T(t; n-1) &\propto \lim_{n\to\infty}\left(1+\frac{t^2}{n-1}\right)^{-n/2} \\ &= \lim_{n\to\infty}\left[\left(1+\frac{t^2}{n}\right)^{n}\right]^{-1/2} = \left(e^{t^2}\right)^{-1/2} = e^{-t^2/2} \end{aligned} \tag{2.8}$$

Here's what the $t$ distribution looks like for various values of $\nu = n-1$:

[2] "Student" was the pen name of statistician William Sealy Gosset.

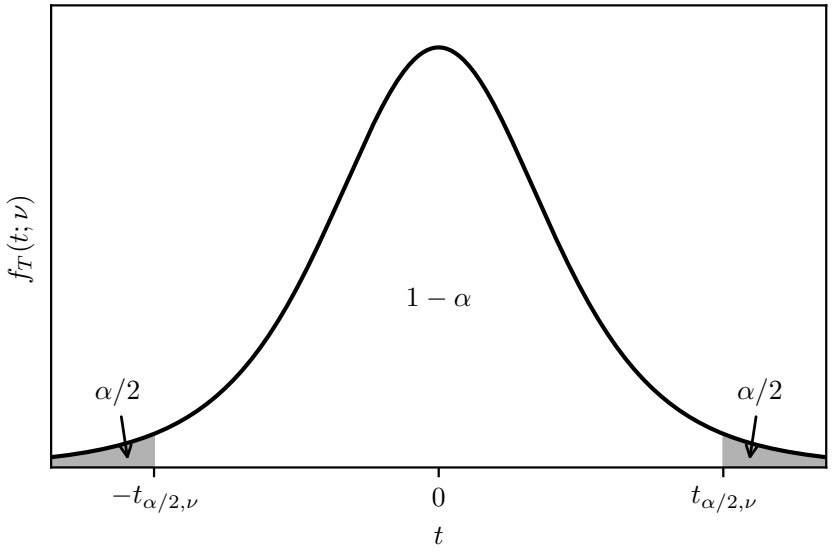



By analogy to the definition of $z_\alpha$, we define $t_{\alpha,\nu}$ (known as the *t critical value*) by $P(T > t_{\alpha,\nu}) = 1 - F_T(t_{\alpha,\nu};\nu) = \alpha$




This means that

$$\begin{aligned} 1-\alpha &= P(-t_{\alpha/2,n-1} < T < t_{\alpha/2,n-1}) \\ &= P\left(-t_{\alpha/2,n-1} < \frac{\overline{X}-\mu}{S/\sqrt{n}} < t_{\alpha/2,n-1}\right) \end{aligned} \tag{2.9}$$




and by the same manipulation as before

$$P\left(\overline{X} - t_{\alpha/2,n-1}\frac{S}{\sqrt{n}} < \mu < \overline{X} + t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}\right) = 1-\alpha \tag{2.10}$$

## 2.3 Prediction Intervals

So far we've used confidence intervals to make statements about the parameters of the underlying distribution, given properties of a sample drawn from that distribution. Instead, we could try to construct an interval with, say, a 95% chance of containing *another* value drawn from that distribution. Specifically, if

$X_1, \ldots, X_n, X_{n+1}$ are iid random variables, we can try to make predictions about the $n+1$st variable based on the first $n$. Of course, the best estimator for $X_{n+1}$ which we can construct from $X_1, \ldots, X_n$ is

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{2.11}$$

The error we make with that guess has mean

$$E(\overline{X} - X_{n+1}) = E(\overline{X}) - E(X_{n+1}) = \mu - \mu = 0 \tag{2.12}$$

and variance

$$V(\overline{X} - X_{n+1}) = V(\overline{X}) + V(X_{n+1}) = \sigma^2/n + \sigma^2 = \left(1 + \frac{1}{n}\right)\sigma^2 \tag{2.13}$$

If the underlying distribution for each $X_i$ is normal, then $\overline{X} - X_{n+1}$, being a linear combination of normal random variables, is normal, and

$$Z = \frac{\overline{X} - X_{n+1} - 0}{\sigma\sqrt{1 + \frac{1}{n}}} \tag{2.14}$$

is a standard normal random variable.

If we don't know $\sigma$ ahead of time, we can use the sample variance

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 \tag{2.15}$$

calculated from the first $n$ random variables and construct

$$T = \frac{\overline{X} - X_{n+1}}{S\sqrt{1 + \frac{1}{n}}}\ . \tag{2.16}$$

This turns out to once again obey a $t$ distribution with $\nu = n-1$ degrees of freedom, and therefore we can say

$$P\left(\overline{X} - t_{\alpha/2,n-1}\, S\sqrt{1 + \frac{1}{n}} < X_{n+1} < \overline{X} + t_{\alpha/2,n-1}\, S\sqrt{1 + \frac{1}{n}}\right) = 1 - \alpha \tag{2.17}$$

and, given a sample $\{x_i\}$ of size $n$, construct the *prediction interval*

$$\overline{x} \pm t_{\alpha/2,n-1}\, s\sqrt{1 + \frac{1}{n}} \tag{2.18}$$

for $x_{n+1}$ at "prediction level" $(1-\alpha)\times 100\%$.

Note that prediction intervals are wider than confidence intervals in general; in particular when $n$ becomes large, the width of a confidence interval goes to zero, but the width of the prediction interval goes to $2z_{\alpha/2}s$.

## 2.4 Tolerance Intervals

One more type of interval to consider is a tolerance interval. This is constructed, based on a sample with mean $\overline{x}$ and standard deviation $s$, to have a $(1-\alpha)\times 100\%$ chance of containing $k\%$ of the area of the underlying probability distribution. This is a complicated construction, and for our purposes, it's simply treated as a black box with an associated table in the appendix. It's mostly a matter of understanding the definition.

## Practice Problems

7.29, 7.35, 7.37, 7.41

**Tuesday 29 January 2019**

# 3 Confidence Interval for the Variance of a Normal Population

So far, we've considered confidence intervals on the mean (or in one case, the proportion) of a distribution. Finally, let's consider how to set a confidence interval on the variance of the normal distribution from which we've drawn a sample. We know that the sample variance

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 \tag{3.1}$$

can be used to estimate the variance, since

$$E(S^2) = \sigma^2 \tag{3.2}$$

We can set up a confidence range for $\sigma^2$ by considering the ratio

$$\frac{S^2}{\sigma^2} = \frac{1}{n-1}\sum_{i=1}^{n}\frac{(X_i - \overline{X})^2}{\sigma^2} \tag{3.3}$$

Now, since

$$\frac{X_i - \mu}{\sigma} \tag{3.4}$$

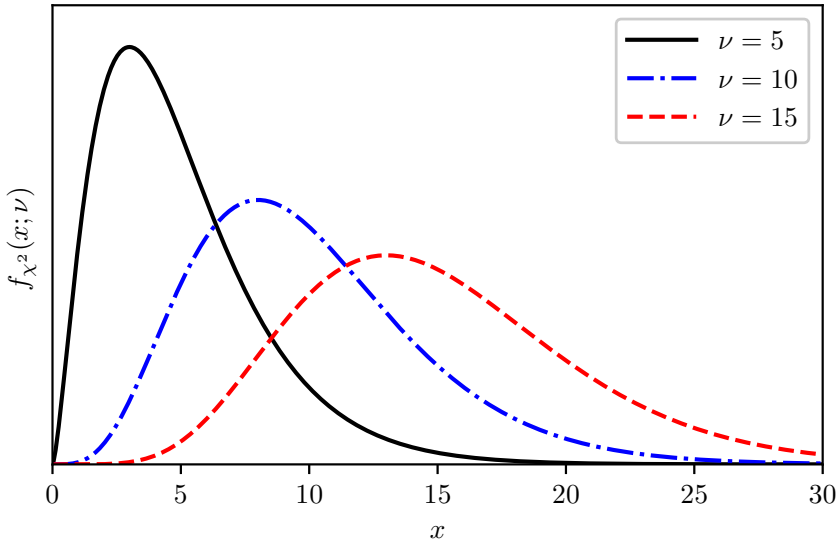
is a standard normal random variable, we know

$$\sum_{i=1}^{n}\frac{(X_i - \mu)^2}{\sigma^2} \tag{3.5}$$

obeys a chi-square distribution with $n$ degrees of freedom (df). If we use the sample mean $\overline{X}$ instead of the population mean $\mu$, we get

$$\sum_{i=1}^{n}\frac{(X_i - \overline{X})^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2} \tag{3.6}$$

which turns out to obey a $\chi^2$ distribution with $\nu = n - 1$ df.



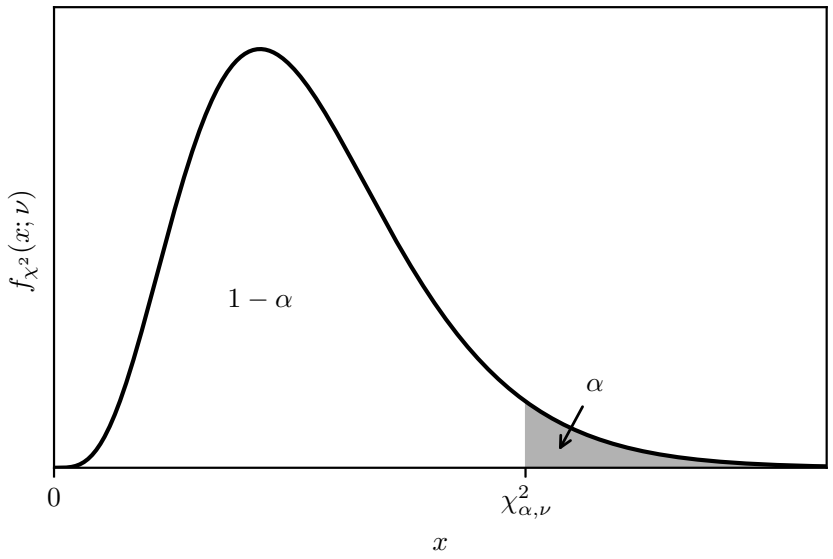

So we need to set intervals for a chi-square random variable. Just as we defined $z_\alpha$ and $t_{\alpha,\nu}$ as the left edge of an area $\alpha$ under the pdf for a standard normal rv and a $t$-distribution with $\nu$ df, respectively, we define $\chi^2_{\alpha,\nu}$ as the corresponding quantity for a chi-square distribution with $\nu$ degrees of freedom:

$$\frac{1}{\sqrt{2\pi}}\int_{z_\alpha}^{\infty} e^{-z^2/2}\,dz = 1 - \Phi(z_\alpha) = \alpha \tag{3.7a}$$

$$\frac{\Gamma([\nu+1]/2)}{\sqrt{\nu\pi}\,\Gamma(\nu/2)}\int_{t_{\alpha,\nu}}^{\infty}\left(1 + \frac{t^2}{\nu}\right)^{-[\nu+1]/2} dt = 1 - F_T(t_{\alpha,\nu};\nu) = \alpha \tag{3.7b}$$

$$\frac{1}{2^{\nu/2}\Gamma(\nu/2)}\int_{\chi^2_{\alpha,\nu}}^{\infty} x^{(\nu/2)-1}e^{-x/2}\,dx = 1 - F\left(\frac{\chi^2_{\alpha,\nu}}{2};\frac{\nu}{2}\right) = \alpha \tag{3.7c}$$

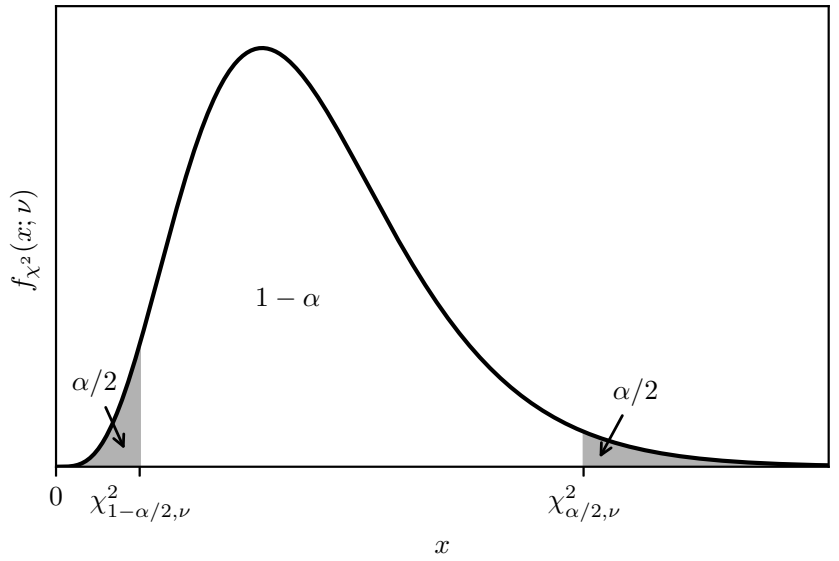
where we've written the cdf for the chi-square distribution in terms of the standard Gamma cdf $F(x;\alpha)$. In practice each of these three values–$z_\alpha$, $t_{\alpha,\nu}$, $\chi^2_{\alpha,\nu}$–is tabulated.




So, having defined the *chi-squared critical value* $\chi^2_{\alpha,\nu}$, we want to split up the area under the chi-squared distribution into the first $\alpha/2$, the middle $1-\alpha$, and the last $\alpha/2$. The one complication is that, since the chi-squared distribution is not symmetric like the standard normal and $t$ distributions were, we have to handle the lower limit more carefully. If we want a value so that $\alpha/2$ of the area under the curve lies to the left of it, then $1-\alpha/2$ lies to the right, and we have $\chi^2_{1-\alpha/2,\nu}$ and $\chi^2_{\alpha/2,\nu}$ as our $\alpha\times 100$th and $(1-\alpha)\times 100$th and percentiles. The lack of symmetry means

$$\chi^2_{1-\alpha/2,\nu} \neq -\chi^2_{\alpha/2,\nu} \tag{3.8}$$

(In fact, both $\chi^2_{1-\alpha/2,\nu}$ and $\chi^2_{\alpha/2,\nu}$ must be positive.)




The interval containing the middle $1-\alpha$ of probability is thus

$$\begin{aligned} 1-\alpha &= P\left(\chi^2_{1-\alpha/2,n-1} < \sum_{i=1}^{n}\frac{(X_i-\overline{X})^2}{\sigma^2} < \chi^2_{\alpha/2,n-1}\right) \\ &= P\left(\chi^2_{1-\alpha/2,n-1} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{\alpha/2,n-1}\right) \\ &= P\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}\right) \end{aligned} \tag{3.9}$$

and, if we prefer to write a confidence interval for the standard deviation,

$$P\left(\sqrt{\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}}} < \sigma < \sqrt{\frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}}\right) = 1-\alpha \tag{3.10}$$

## Practice Problems

7.43, 7.45, 7.55

# 4 Summary of Confidence Intervals

## 4.1 For Normal Random Samples

At confidence level $(1-\alpha)\times 100\%$:

| known | unknown | variable | lower | upper |
|---|---|---|---|---|
| $\sigma$ | | $\mu$ | $\overline{x}-z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ | $\overline{x}+z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ |
| $\sigma$ | | $\mu$ | $\overline{x}-z_{\alpha}\frac{\sigma}{\sqrt{n}}$ | $\infty$ |
| $\sigma$ | | $\mu$ | $-\infty$ | $\overline{x}+z_{\alpha}\frac{\sigma}{\sqrt{n}}$ |
| | $\sigma$ | $\mu$ | $\overline{x}-t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}$ | $\overline{x}+t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}$ |
| | $\sigma$ | $\mu$ | $\overline{x}-t_{\alpha,n-1}\frac{s}{\sqrt{n}}$ | $\infty$ |
| | $\sigma$ | $\mu$ | $-\infty$ | $\overline{x}+t_{\alpha,n-1}\frac{s}{\sqrt{n}}$ |
| | $\mu$ | $\sigma$ | $s\sqrt{\frac{(n-1)}{\chi^2_{\alpha/2,n-1}}}$ | $s\sqrt{\frac{(n-1)}{\chi^2_{1-\alpha/2,n-1}}}$ |
| | $\mu$ | $\sigma$ | $s\sqrt{\frac{(n-1)}{\chi^2_{\alpha,n-1}}}$ | $\infty$ |
| | $\mu$ | $\sigma$ | $0$ | $s\sqrt{\frac{(n-1)}{\chi^2_{1-\alpha,n-1}}}$ |

## 4.2 Approximate Confidence Intervals for Large Samples

At approximate level $(1-\alpha)\times 100\%$:

| Distribution | variable | lower | upper |
|---|---|---|---|
| general | $\mu$ | $\overline{x}-z_{\alpha/2}\frac{s}{\sqrt{n}}$ | $\overline{x}+z_{\alpha/2}\frac{s}{\sqrt{n}}$ |
| general | $\mu$ | $\overline{x}-z_{\alpha}\frac{s}{\sqrt{n}}$ | $\infty$ |
| general | $\mu$ | $-\infty$ | $\overline{x}+z_{\alpha}\frac{s}{\sqrt{n}}$ |
| Bernoulli/Binomial | $p$ | $\frac{\hat{p}+\frac{z^2_{\alpha/2}}{2n}-z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}+\frac{z^2_{\alpha/2}}{4n^2}}}{1+\frac{z^2_{\alpha/2}}{n}}$ | $\frac{\hat{p}+\frac{z^2_{\alpha/2}}{2n}+z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}+\frac{z^2_{\alpha/2}}{4n^2}}}{1+\frac{z^2_{\alpha/2}}{n}}$ |