
Course:	Mathematical Statistics
Term:	Fall 2017
Instructor:	Gordan Žitković

Lecture 10

Confidence intervals

10.1 Point vs. interval estimators

We talked about *point* estimators in the previous lectures, and now we move on to *interval* estimators. Their main advantage over point estimators is that they give us an idea of how accurate our estimate is.

Example 10.1.1. Consider the following two data sets, both consisting of measurements of the same quantity (say the distance to Proxima Centauri) and in the same units, but made with two different methods.

method 1: 4.51, 4.52, 4.48, 4.49, 4.47, 4.53

method 2: 14.12, 1.30, 0.40, 2.50, 1.00, 3.18

If we use the sample mean \bar{Y} as the (point) estimator for the “true” mean μ , both of these data sets yield the same result, namely $\bar{Y} = 4.5$. It is clear, however, that the first method is more accurate and that, in general, one should trust the results produced by method 1 more than those obtained by method 2.

An interval estimator, i.e., a pair $\hat{\theta}_L \leq \hat{\theta}_R$ of point estimators, helps us convey this missing information. We usually think of $\hat{\theta}_L$ and $\hat{\theta}_R$ as end-points of an interval $[\hat{\theta}_L, \hat{\theta}_R]$, called a **confidence interval**, with the property that the “true” parameter θ lies in $[\hat{\theta}_L, \hat{\theta}_R]$, “with high probability”.

To be useful, confidence intervals should have the following properties

1. they should be *short*, and
2. they should contain θ with *high probability*.

These two, like in many other situations arising in statistics, cannot be optimized simultaneously. Indeed, as we make the interval shorter and shorter, the probability that it contains θ will typically get smaller and smaller. On the other hand, we can make this probability get as close to 1 as we like by making the interval larger and larger. The usual way out of this (and we will see the same approach in the lecture on testing later) is to decide on the

least level of one of these criteria we can tolerate, and then optimize the other under this constraint.

In this case, we first pick a number $\alpha \in (0, 1)$, called the **significance level**, and require that the interval contains θ with the probability at least $1 - \alpha$, i.e.,

$$\mathbb{P}[\theta_L \leq \theta \leq \theta_R] \geq 1 - \alpha. \quad (10.1.1)$$

The number $1 - \alpha$ is called the **confidence¹ level** or (**confidence coefficient**). As before, when we defined unbiasedness, the inequality in (10.1.1) is assumed to hold for *all* values of the parameter θ . Therefore, the goal is to construct the shortest interval such that (10.1.1) still holds.

Sometimes, you only care about not underestimating (or not overestimating) the parameter, without worrying about overestimating (or underestimating) it. In those cases we talk about **one-sided confidence intervals** and set either $\theta_L = -\infty$ or $\theta_R = \infty$. In those cases, the requirement (10.1.1) is either $\mathbb{P}[\theta \leq \theta_R] \geq 1 - \alpha$ in the former, and $\mathbb{P}[\theta_L \leq \theta] \geq 1 - \alpha$ in the later. If the probabilities corresponding to both tails are the same, (and, typically, equal to $1 - \alpha/2$) we say that the interval is **symmetric**.

10.2 Pivotal quantities

Unfortunately, we do not have enough mathematics at our disposal to give an algorithm that will produce a confidence interval in every conceivable model. What we can do is describe one particular construction which, when it works (and it works in many cases of interest) yields very good results. The concept we depend on is similar to the concept of an estimator, but with one notable difference:

Definition 10.2.1. Given a random sample (Y_1, \dots, Y_n) from a distribution D with an unknown parameter θ , a **pivotal quantity** is a function of the data (Y_1, \dots, Y_n) and the parameter θ , whose distribution *does not* depend on θ .

So, unlike in the case of an estimator, the dependence on parameters is allowed, but at the expense of a different requirement. This time, the *distribution* of the pivotal quantity cannot depend on the parameter at all.

Example 10.2.2. The normal model:

1. $N(\mu, 1)$: Let (Y_1, \dots, Y_n) be a random sample from $N(\mu, 1)$, with an unknown mean μ , but known variance 1. The sample mean \bar{Y} is an estimator, but it is not a pivotal quantity. Indeed, we have seen before, that its distribution is normal with mean μ and variance $1/n$. This clearly depends on μ .

¹confidence+significance=1

On the other hand,

$$\bar{Y} - \mu$$

is not an estimator, but it is a pivotal quantity. Indeed, it is normally distributed with mean 0 and variance $1/n$ - a distribution which does not depend on μ .

If we multiply a pivotal quantity by a constant (which depends neither on the unknown parameter μ nor on the data) we still get a pivotal quantity. This way, $\sqrt{n}(\bar{Y} - \mu)$ is also a pivotal quantity. What is its distribution?

2. $N(\mu, \sigma)$: This time, the parameter θ is two-dimensional, i.e., $\theta = (\mu, \sigma)$, i.e., both μ and σ are unknown. To construct a pivotal quantity, we start with $\bar{Y} - \mu$ from above, but, since σ is unknown, $\bar{Y} - \mu$ is no longer a pivotal quantity in this new model. Indeed, its distribution is $N(0, \frac{1}{\sqrt{n}}\sigma)$, which clearly depends on σ . To make it independent of θ , we need to get rid of σ , as well. This is easy in this case - we simply need to normalize by σ :

$$\frac{\bar{Y} - \mu}{\sigma}$$

is a pivotal quantity because its distribution is $N(0, \frac{1}{\sqrt{n}})$. As above, it remains a pivotal quantity if we multiply it by \sqrt{n} . As an added benefit, this way we get rid of the dependence on n in the distribution since

$$\sqrt{n} \frac{\bar{Y} - \mu}{\sigma} \sim N(0, 1).$$

10.3 Confidence intervals from pivotal quantities

We are now ready for the construction of confidence intervals with the confidence level $1 - \alpha$ (significance level α). It takes three easy steps:

1. Pick a pivotal quantity U (it will always be given to you in this course).
2. Define a and b by

$$a = q_U(\alpha/2) \text{ and } b = q_U(1 - \alpha/2).$$

Since the distribution of U does not depend on θ , its quantiles do not either. Therefore, no matter what θ happens to be, we have

$$\mathbb{P}[a \leq U \leq b] = 1 - \mathbb{P}[U < a \text{ or } U > b] = 1 - (\alpha/2 + \alpha/2) = 1 - \alpha.$$

3. “Solve” the inequality $a \leq U \leq b$ for θ , i.e., rearrange it so that the only thing left in the middle is θ . Give names $\hat{\theta}_L$ and $\hat{\theta}_R$ to whatever you get on the left and on the right, respectively.

Two comments need to be made here. First, if we set $a = -\infty$ and $b = q_U(1 - \alpha)$ in 2. above, we would get a one-sided interval. Same for $a = q_U(\alpha)$ and $b = +\infty$. In fact, by picking $a = q_U(\alpha_1)$ and $b = q_U(1 - \alpha_2)$ with $\alpha_1 + \alpha_2 = 1$, we get a whole family of confidence intervals (but we will not consider them in this course).

Second, all of this works for one-dimensional θ only. If, like in the normal model with unknown μ and σ , the parameter $\theta = (\mu, \sigma)$ is two-dimensional (or higher), “solving” for θ in 3. above makes no sense. What could be constructed, but we do not pursue it here, is a so-called **confidence region**, i.e. a region in \mathbb{R}^2 (or \mathbb{R}^n if you have n components of θ) which contains θ with probability at least $1 - \alpha$.

Example 10.3.1.

1. $N(\mu, 1)$: We use the pivotal quantity $U = \sqrt{n}(\bar{Y} - \mu)$, as in Example 10.2.2 above. U has the standard normal distribution $N(0, 1)$, and, therefore, we can use the quantiles a and b of the standard normal. For concreteness, let us pick $1 - \alpha = 95\%$ and using tables (if you must) or software (the R-command is `qnorm`) compute the values of a and b as

$$a = -1.96 \text{ and } b = 1.96.$$

This way,

$$\mathbb{P}[-1.96 \leq \sqrt{n}(\bar{Y} - \mu) \leq 1.96] = 0.95$$

Since

$$\mathbb{P}[-1.96 \leq \sqrt{n}(\bar{Y} - \mu) \leq 1.96] = 0.95 \quad \text{if and only if}$$

$$\mathbb{P}[-1.96/\sqrt{n} \leq \bar{Y} - \mu \leq 1.96/\sqrt{n}] = 0.95 \quad \text{if and only if}$$

$$\mathbb{P}[-\bar{Y} - 1.96/\sqrt{n} \leq -\mu \leq -\bar{Y} + 1.96/\sqrt{n}] = 0.95, \quad \text{if and only if}$$

$$\mathbb{P}[\bar{Y} - 1.96/\sqrt{n} \leq \mu \leq \bar{Y} + 1.96/\sqrt{n}] = 0.95.$$

and we can set

$$\hat{\theta}_L = \bar{Y} - 1.96/\sqrt{n} \text{ and } \hat{\theta}_R = \bar{Y} + 1.96/\sqrt{n},$$

so that, as required,

$$\mathbb{P}[\hat{\theta}_L \leq \mu \leq \hat{\theta}_R] = 1 - \alpha.$$

If we wanted a one-sided interval, we would pick $\theta_L = -\infty$ and θ_R would be the same as above, except that now the $1 - \alpha/2 = 0.975$ -quantile 1.96 of $N(0, 1)$ should be replaced by the $1 - \alpha = 0.95$ -quantile 1.64, i.e., $\theta_R = \bar{Y} + 1.64/\sqrt{n}$.

Finally, let us consider a toy data set from $N(\mu, 1)$:

7.54, 10.32, 8.34, 10.46, 10.69, 10.61.

We compute the average to get $\bar{Y} = 9.66$. Therefore, with $n = 6$, the formula above yields the interval $[8.86, 10.46]$. The value of the parameter μ used to simulate the data above was $\mu = 10$, which was nicely captured by the interval (as we expect it happen about 19 out of 20 times).

2. $N(\mu, \sigma_0)$: This example is similar to the one above; the only difference is that the *known* value of the standard deviation σ is now σ_0 instead of 1. We could use the same pivotal quantity as above, but its distribution would not be $N(0, 1)$ anymore - it would be $N(0, \sigma_0)$. Therefore, if we divide by σ_0 (which is now a known constant, and not an unknown parameter), we obtain the pivotal quantity $U = \sqrt{n} \frac{\bar{Y} - \mu}{\sigma_0}$. Going through the same steps as above, we find $\theta_L = \bar{Y} - 1.96 \frac{\sigma_0}{\sqrt{n}}$ and $\theta_R = \bar{Y} + 1.96 \frac{\sigma_0}{\sqrt{n}}$. The center of the interval is still \bar{Y} , but its size (length) gets multiplied by σ_0 .

Here is another example from a different distribution

Example 10.3.2. Let (Y_1, \dots, Y_n) be a random sample from the exponential distribution with an unknown parameter $\tau > 0$. A good pivotal quantity to use is

$$U = \frac{1}{\tau} \bar{Y} = \frac{1}{n\tau} (Y_1 + \dots + Y_n).$$

We know that the exponential distribution is a special case of a gamma distribution (with the shape parameter $k = 1$), so $Y_1 + \dots + Y_n$ is a gamma $\Gamma(n, \tau)$. Using mgfs, we can easily see that the division by $n\tau$ leaves us in the gamma family, but changes the shape parameter to $1/n$, so that $U \sim \Gamma(n, 1/n)$, a quantity which does not depend on τ . We pick $\alpha = 0.1$ in this case and $n = 6$ as above and compute the values a, b (remember, R requires $1/\tau$ not τ for exponential and gamma distributions!)

$a = \text{qgamma}(0.05, 6, 6) = 0.44$ and $b = \text{qgamma}(0.95, 6, 6) = 1.75$,

so that

$$\mathbb{P}[0.44 \leq U \leq 1.75] = 1 - \alpha = 0.9$$

Since $U = \tilde{Y}/\tau$, we have

$$\begin{aligned}\mathbb{P}[0.44 \leq U \leq 1.75] &= 0.95 && \text{if and only if} \\ \mathbb{P}[0.44 \leq \tilde{Y}/\tau \text{ and } \tilde{Y}/\tau \leq 1.75] &= 0.95 && \text{if and only if} \\ \mathbb{P}[\tau \leq \tilde{Y}/0.44 \text{ and } \tilde{Y}/1.75 \leq \tau] &= 0.95 && \text{if and only if} \\ \mathbb{P}[\tilde{Y}/1.75 \leq \tau \leq \tilde{Y}/0.44] &= 0.95,\end{aligned}$$

so that $\hat{\theta}_L = \tilde{Y}/1.75$ and $\hat{\theta}_R = \tilde{Y}/0.44$ define the left and the right limits of a $1 - \alpha = 0.9$ -confidence interval. For the following toy data set coming from $E(\tau)$ with $\tau = 2$:

2.89, 2.45, 2.10, 9.92, 5.48, 0.91

and the mean $\tilde{Y} = 3.96$. Therefore, the 90%-confidence interval is given by $[2.26, 9]$. Note how the true parameter, $\tau = 2$, is not in the interval. That is not so unexpected - in fact, it would be strange if it did not happen in about one case out of ten. Given another sample from the same distribution;

1.43, 0.14, 1.26, 7.15, 0.57, 0.34,

the mean is $\tilde{Y} = 1.81$ and the resulting confidence interval is $[1.04, 4.12]$.

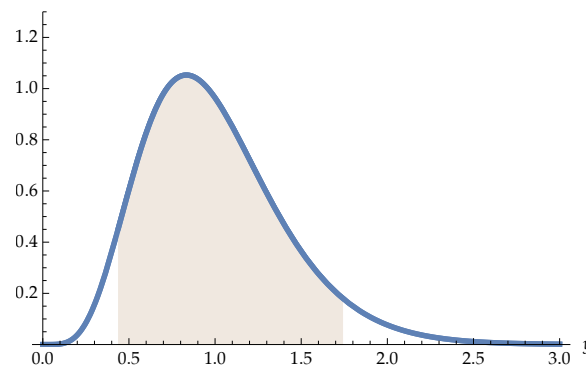


Figure 1. The pdf of the $\Gamma(6, 1/6)$ -distribution. The shaded area is 90% of the total area under the graph (which is 1), and the two unshaded parts are 5% each.

A feature of this example is that the natural estimator \tilde{Y} for τ (an UMVUE, in this case) is *not* the center of the confidence interval. This is due to the fact that τ enters the distribution $E(\tau)$ in a *multiplicative*

way. More precisely, if $Y \sim E(\tau)$ then $\alpha Y \sim E(\alpha\tau)$ for any $\alpha > 0$. The relative error in estimation does not depend on the value of τ , and, thus, it is only natural that the confidence interval reflects that, too.

10.4 Approximate confidence intervals

Consider the elections example, i.e., a random sample (Y_1, \dots, Y_n) from the Bernoulli $B(p)$ distribution, with an unknown $p \in (0, 1)$. To construct a confidence interval for p , we need a pivotal quantity, but it turns out that the distribution of the usual suspect, namely

$$U = \frac{\bar{Y} - p}{\sqrt{p(1-p)/n}}$$

which we get when we normalize the sample proportion \bar{Y} (subtract $p = \mathbb{E}[\bar{Y}]$ and divide by $\sqrt{p(1-p)/n} = \text{sd}[\bar{Y}]$), *still depends on p* even though $\mathbb{E}[U] = 0$ and $\text{Var}[U] = 1$ (can you argue why?). Therefore, U is simply *not a pivotal quantity*. While there are other methods that can help us construct (exact) confidence intervals in this case, we note that U is exactly of the form which appears in Central Limit Theorem. Therefore, its distribution is close to $N(0, 1)$ when n is large enough. Therefore, for large n , U is an **approximate (or asymptotic) pivotal quantity** and if we use it to construct a confidence interval, we will get an **approximate confidence interval**. If n is not too small, this procedure yields very good results.

So, if we pretend that U had a normal $N(0, 1)$ -distribution, it is not hard to find the quantiles $a = \text{qnorm}(\alpha/2, 0, 1)$ and $b = \text{qnorm}(1 - \alpha/2, 0, 1)$ so that

$$\mathbb{P}[a \leq U \leq b] \approx 1 - \alpha$$

The “solving for p ”-step should be similar to the normal case, except that the parameter p now appears both in the numerator and the denominator and one gets a quadratic equation in p . To simplify the calculations (and since we are approximating already anyway) we usually replace every occurrence of the unknown parameter p in the denominator $\sqrt{p(1-p)/n}$ by its estimate $\hat{p} = \bar{Y}$. This way the inequality $a \leq U \leq b$ can be easily rearranged to

$$\hat{p} + a\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + b\sqrt{\frac{\hat{p}(1-\hat{p})}{n}},$$

making

$$\left[\hat{p} + a\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + b\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right]$$

an $(1 - \alpha)$ -confidence interval for p .

It should be noticed that one gets the same interval in the case of a normal model with known $\sigma = \sigma_0$ if one takes $\sigma_0 = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$. Also, the above

discussion does not apply to the Bernoulli model only. You can use an approximate pivotal quantity and construct approximate confidence intervals whenever you can use the central limit theorem.

10.5 Choosing the sample size

In many situations the statistician is consulted before the experiment is even performed. It turns out to be a very good idea, as the proper design of an experiment can significantly enhance the amount of useful information it yields, and, hence, its scientific value. We do not talk about (the entire field of) **experimental design** in these notes, but do give a simple example of its flavor.

Whether you are measuring the distance to Proxima centauri, polling the electorate, or conducting research on the effectiveness of a new drug, your research will not have much credibility unless you can report your results with sufficient accuracy. In the language of statistics, this can be phrased as the following question

How do we have to design the experiment so that, once the data are collected, the confidence interval for the parameter of interest does not exceed a prescribed size?

The size of the confidence interval is usually specified by prescribing its **margin of error**, i.e., the distance $\frac{1}{2}(\hat{\theta}_R - \hat{\theta}_L)$ from the center of the interval to either of its endpoints. As for the experiment design itself, we do not have much freedom in these notes, because all our models are built on a random sample from a given distribution. The only value we get to control is the *sample size*. Therefore, the question above simplifies into:

Given $m > 0$ and a confidence level $1 - \alpha$, how large a sample size (i.e., n) do we need to be able to guarantee the margin of error of at most m ?

Example 10.5.1. The question above is simple to answer in the case when Y_1, \dots, Y_n is a random sample from $N(\mu, \sigma)$ with *known* σ . In that case, as derived in 10.3.1 above (in the case $\sigma = 1$), the $1 - \alpha$ -confidence for μ is given by

$$\left[\bar{Y} - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{Y} + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \right],$$

where $z_{1-\alpha/2}$ is the $1 - \alpha/2$ -quantile of the standard normal $N(0, 1)$. Therefore, the margin of error is $z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$, and, in order for it to be at most m , we need

$$z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \leq m.$$

We solve this inequality for n to obtain

$$n \geq \left(\frac{z_{1-\alpha/2}\sigma}{m} \right)^2.$$

The computation in the example above depends crucially on the fact that σ was known. The situation is more complicated when it is unknown. In some cases, we need to perform a preliminary small-scale study in order to estimate σ (we will see how to do that a bit later) and then use the estimated value as above. In other cases, like in our next example, we can bound the size of the confidence interval from above without ever constructing it explicitly.

Example 10.5.2. As in the elections example, let Y_1, \dots, Y_n be a random sample from the $B(p)$ -distribution with an unknown $p \in (0, 1)$. We have seen in the previous example that, since the parameter p determines both the mean and the variance of the sample proportion \hat{p} , it is difficult to construct an exact confidence interval. This can be overcome in the large-sample case by employing an approximation based on the central limit theorem. Even so, the value we use for σ is

$$\sigma = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}},$$

which is, itself, a function of the observed data. Not everything is lost, though, since

$$\hat{p}(1-\hat{p}) \leq \frac{1}{4} \text{ no matter what the value of } \hat{p} \text{ is.}$$

Indeed, the function $x \mapsto x(1-x)$ achieves its maximum $\frac{1}{2} \times (1 - \frac{1}{2}) = \frac{1}{4}$ at $x = \frac{1}{2}$. Therefore, no matter what the value of \hat{p} turns out to be, we know that the margin of error will admit the following upper bound (assuming a symmetric interval, so that $a = -b = z_{1-\alpha/2}$):

$$m \leq \frac{z_{1-\alpha/2}}{2\sqrt{n}},$$

and the sample size of $n \geq \left(\frac{z_{1-\alpha/2}}{2m} \right)^2$ guarantees the prescribed margin of error. For example, for $1 - \alpha = 0.95$ and $m = 1\%$, we get $n \geq 9604$.

10.6 Confidence intervals for the sample variance

The unbiased estimator we used to estimate the variance parameter σ^2 in the $N(\mu, \sigma)$ -model with μ and σ unknown is the sample variance

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2.$$

We argued that $\mathbb{E}[S^2] = \sigma^2$ but did not say much about its distribution. In order to base a pivotal quantity on S^2 , we need to know more about it than only its expectation - we need the entire distribution. For that, in turn, we need an important result from probability, which we only partially prove here:

Theorem 10.6.1. *Let Y_1, \dots, Y_n be a sample from the normal $N(\mu, \sigma)$ -distribution, and let*

$$\bar{Y} = \frac{1}{n}(Y_1 + \dots + Y_n) \text{ and } Q^2 = \frac{1}{\sigma^2} \sum_{i=1}^n (Y_i - \bar{Y})^2.$$

Then

1. \bar{Y} is normal with mean μ and variance $\frac{1}{n}\sigma^2$,
2. Q^2 is a pivotal quantity with the chi-squared distribution $\chi^2(n-1)$ with $n-1$ degrees of freedom, and
3. \bar{Y} and Q^2 are independent.

Proof/Explanation.

1. We already know this. Indeed, \bar{Y} is a linear combination of independent normal random variables, so it is, itself, normally distributed. Its mean is μ and its variance is $\frac{1}{n}\sigma^2$.
2. We cannot prove this in full, but we need to observe, first, that

$$\begin{aligned} Q^2 &= \sum_{i=1}^n (Y_i/\sigma - \bar{Y}/\sigma)^2 = \sum_{i=1}^n (Y_i/\sigma - \mu - (\bar{Y}/\sigma - \mu))^2 \\ &= \sum_{i=1}^n (Z_i - \bar{Z})^2, \text{ where } Z_i = (Y_i - \mu)/\sigma \sim N(0, 1). \end{aligned}$$

The random variables Z_1, \dots, Z_m are also independent, so Q^2 can be written as a function of n independent standard normal random variables. It follows that its distribution, whatever it is, does not depend on either μ or σ .

The fact that the distribution of Q^2 is $\chi^2(n-1)$ is harder to argue. It is true that Q^2 can be written as a sum of squares of normal random variables, but these random variables are neither independent, nor is it a sum of $n-1$ terms, as required by the definition of $\chi^2(n-1)$. The trick is to rearrange the terms in $\sum_{i=1}^n (Z_i - \bar{Z})^2$ to write it as a sum of squares of $n-1$ independent normals. These normals will no longer be Z_1, \dots, Z_n , but some linear combinations of them. We do this for $n=2$ only:

$$\begin{aligned} (Z_1 - \bar{Z})^2 + (Z_2 - \bar{Z})^2 &= (\tfrac{1}{2}Z_1 - \tfrac{1}{2}Z_2)^2 + (\tfrac{1}{2}Z_2 - \tfrac{1}{2}Z_1)^2 \\ &= \tfrac{1}{2}(Z_1 - Z_2)^2 = \left(\tfrac{Z_1 - Z_2}{\sqrt{2}}\right)^2. \end{aligned}$$

The random variable $(Z_1 - Z_2)/\sqrt{2}$ is normal with mean 0 and variance

$$\text{Var}\left[\frac{Z_1 - Z_2}{\sqrt{2}}\right] = \tfrac{1}{2} \text{Var}[Z_1 - Z_2] = \tfrac{1}{2}(\text{Var}[Z_1] + \text{Var}[Z_2]) = 1.$$

The independence of \bar{Y} and Q^2 is also beyond the scope of these notes. To give at least a little bit of an explanation, we just mention this independence follows from the fact that the linear combinations we use in 2. above to write Q^2 are a sum of squares of $n-1$ independent normals happen to be independent of \bar{Z} . \square

We use the first two statements of Theorem 10.6.1 right away, and save the third one for a little bit later. The second statement gets us most of the way to the construction of a confidence interval for the variance parameter σ^2 given a random sample Y_1, \dots, Y_n from $N(\mu, \sigma)$ with unknown μ and σ . Indeed, Q^2 is pivotal quantity with $\chi^2(n-1)$ -distribution, so for a significance level α , we can find positive constants $a < b$ such that

$$\mathbb{P}[a \leq Q^2 \leq b] = 1 - \alpha,$$

e.g., $a = \text{qchisq}(\alpha/2, n-1)$ and $b = \text{qchisq}(1 - \alpha/2, n-1)$. All we need to do now is remember that $Q^2 = (n-1)S^2/\sigma^2$, so that

$$a \leq Q^2 \leq b \text{ if and only if } \frac{n-1}{b} S^2 \leq \sigma^2 \leq \frac{n-1}{a} S^2.$$

Therefore, $\hat{\theta}_L = \frac{n-1}{b} S^2$ and $\hat{\theta}_R = \frac{n-1}{a} S^2$ are the end-points of a symmetric $1 - \alpha$ -confidence interval for σ^2 .

10.7 Confidence intervals when the variance is unknown

We have already constructed a confidence interval in the normal model for μ when σ is known, and for σ itself, when it is unknown. It remains to see what happens when both μ and σ are unknown. Theorem 10.6.1 tells us that the

random variable $Q^2 = (n-1)S^2/\sigma^2$ has a $\chi^2(n-1)$ -distribution and that it is independent of \bar{Y} . Moreover, $\bar{Y} \sim N(\mu, \sigma/\sqrt{n})$, so that, after standardization,

$$\frac{\bar{Y} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1).$$

It follows that the distribution of the quantity T , given by,

$$T = \frac{\bar{Y} - \mu}{\sqrt{S^2/n}} = \frac{\frac{\bar{Y} - \mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{S^2}{\sigma^2}}} = \frac{Z}{\sqrt{\frac{1}{n-1}Q^2}}$$

is the same as the distribution of a standard normal (i.e., Z) divided by the square root of an independent $\chi^2(n-1)$ -distribution divided by $n-1$ (i.e., $\sqrt{\frac{1}{n-1}Q^2}$). This distribution is so important it has a name:

Definition 10.7.1. A **Student t -distribution** with k degrees of freedom - denoted by $t(k)$ - is the distribution of the random variable

$$\frac{Z}{\sqrt{Q^2/k}},$$

where $Z \sim N(0, 1)$ and $Q^2 \sim \chi^2(k)$, with Z and Q^2 independent.

The R notation for the t distribution is simply `t` and the only parameter is k , the number of degrees of freedom, so that, e.g., the 97.5%-quantile of the t distribution with 5 degrees of freedom is `qt(0.975, 5)`, which evaluates to 2.5706 (compare that to the corresponding normal quantile, which is 1.96).

The discussion above can now be summarized in the following proposition:

Proposition 10.7.2. Let Y_1, \dots, Y_n be a random sample from $N(\mu, \sigma)$ with μ and $\sigma > 0$ unknown. Then

$$T = \frac{\bar{Y} - \mu}{\sqrt{\frac{S^2}{n}}},$$

where \bar{Y} and S^2 are the sample mean and the (unbiased) sample variance, is a pivotal quantity with the Student t -distribution $t(n-1)$ with $n-1$ degrees of freedom.

It is now easy to construct a confidence interval when both μ and σ are unknown. We pick two quantiles a, b from the $t(n-1)$ -distribution, say $a =$

$-t_{\alpha/2} := q_{t(n-1)}(\alpha/2)$ and $b = t_{\alpha/2} = q_{t(n-1)}(1 - \alpha/2)$, so that

$$\mathbb{P} \left[-t_{\alpha/2} \leq \frac{\bar{Y} - \mu}{\sqrt{\frac{S^2}{n}}} \leq t_{\alpha/2} \right] = 1 - \alpha.$$

This translates to

$$\mathbb{P} \left[\bar{Y} - t_{\alpha/2} \frac{\sqrt{S^2}}{\sqrt{n}} \leq \mu \leq \bar{Y} + t_{\alpha/2} \frac{\sqrt{S^2}}{\sqrt{n}} \right] = 1 - \alpha,$$

which, in turn, tells us that $[\hat{\theta}_L, \hat{\theta}_R]$, where

$$\hat{\theta}_L = \bar{Y} - t_{\alpha/2} \frac{\sqrt{S^2}}{\sqrt{n}} \text{ and } \hat{\theta}_R = \bar{Y} + t_{\alpha/2} \frac{\sqrt{S^2}}{\sqrt{n}}$$

is an $1 - \alpha$ -confidence interval for μ .

10.8 Problems

Problem 10.8.1. Let Y_1, Y_2, \dots, Y_n be a random sample from the exponential distribution with parameter $\tau > 0$.

1. What is the distribution of $U = \frac{2 \sum_{i=1}^n Y_i}{\tau}$? Is it a pivotal quantity?
2. Construct a 95%-confidence interval for τ when $n = 2$ and the observed values of Y_1 and Y_2 are 0.4 and 0.1.

(Note: Use R to get the exact numbers, or write your answer in terms of R commands as if they were numbers.)

Problem 10.8.2. Let Y_1, \dots, Y_5 be a random sample from the normal distribution $N(\mu, 2)$, with an unknown mean μ and the known standard deviation $\sigma = 2$. The collected data turn out to be

$$y_1 = 2, y_2 = 5, y_3 = 1, y_4 = 4, y_5 = 3.$$

The right end-point of the one-sided 90%-confidence interval $(-\infty, \hat{\mu}_R]$ for μ is

- (a) $3 + \frac{2}{\sqrt{5}} \text{qnorm}(0.9, 0, 1)$.
- (b) $3 + \frac{2}{5} \text{qnorm}(0.9, 0, 1)$.
- (c) $3 + \frac{1}{\sqrt{5}} \text{qt}(0.9, 4)$.
- (d) $3 + \frac{1}{5} \text{qnorm}(0.9, 5)$.

(e) none of the above

Problem 10.8.3. A pollster is trying to estimate the proportion of the population in favor of candidate A (in a two-way race between A and B). The quality of her sample is such that it can be safely assumed to be a random sample from the Bernoulli distribution with the unknown parameter p . She is interested in the smallest sample size she will need in order to be able to pinpoint the value of p with $\pm 1\%$ accuracy, with 95% confidence.

Basing your analysis on the estimator $\hat{p} = Y/n$, where Y is the number of supporters of candidate A in the sample, find the smallest such n under the assumption that the sampling distribution of \hat{p} is well approximated by a normal distribution (of appropriate mean and variance left for you to figure out).

Problem 10.8.4. A sample of $n = 10000$ voters is chosen and their preferences between candidates A and B are recorded. It turns out that 5800 prefer A and the rest prefer B . Construct an approximate 95%-confidence interval for the proportion p of the voters in the entire population who prefer A .

Problem 10.8.5. In a sample Y_1, \dots, Y_n from the exponential distribution $E(\tau)$ with parameter $\tau > 0$, $U = c\bar{Y}$ is a pivotal quantity if the value of the constant c is

(a) 1 (b) 2 (c) $2/\tau$ (d) τ (e) none of the above

Problem 10.8.6. A random sample of size $n = 5$ from the normal distribution with unknown mean μ and an unknown standard deviation σ yielded the values y_1, \dots, y_5 such that

$$\sum y_i = 10 \text{ and } \sum_{i=1}^5 (y_i - 2)^2 = 4.$$

The value of $\hat{\mu}_L$ such that $(\hat{\mu}_L, \infty)$ is (an asymmetric) 95%-confidence interval for μ is

(a) $2 - \frac{1}{\sqrt{5}} \text{qt}(0.95, 4)$

(b) $2 - \frac{1}{5} \text{qnorm}(0.95)$

(c) $2 - \frac{1}{5} \text{qt}(0.95, 4)$

(d) $2 - \frac{1}{\sqrt{5}} \text{qnorm}(0.95)$

(e) $2 - \frac{1}{\sqrt{5}} \text{qt}(0.975, 4)$

Problem 10.8.7. A sample of size $n = 2$ from normal distribution with unknown μ and σ is collected and the data are

$$y_1 = 1 \text{ and } y_2 = 5.$$

The left end-point of a symmetric 95% confidence interval for σ^2 is

- (a) $8/\text{qchisq}(0.975, 2)$
- (b) $16/\text{qchisq}(0.975, 1)$
- (c) $8/\text{qchisq}(0.975, 1)$
- (d) $16/\text{qchisq}(0.975, 2)$
- (e) none of the above

Problem 10.8.8. 5 astronomy teams from across the world measured the distance to Proxima Centauri using a new method. It is reasonable to assume that the error of this method is normally distributed, but, since it is new, there is no information about its standard variation. Find a 95%-confidence interval for the distance if the obtained measurements are (in light years)

$$4.20, 4.16, 4.04, 4.01, 4.20.$$

What would your confidence interval look like if they used an established method whose standard deviation of the measurement error is 0.1?

Problem 10.8.9. Heights of trees in a certain forest are uniformly distributed between 0 and θ , where θ is the height of the tallest tree in that forest. Having measured the heights of the trees in a random sample of size n , our task is to estimate the height of the tallest tree.

We model the measured heights (Y_1, \dots, Y_n) as a random sample from a uniform distribution $U(0, \theta)$ on the interval $[0, \theta]$, where the right endpoint $\theta > 0$ is an unknown parameter. It is intuitively clear that once the n measurements are taken, the only “relevant” one is the largest one, as the others do not give us any additional information about θ . With this in mind, we define $M = \max(Y_1, \dots, Y_n)$ and base our inference on it.

1. The first order of business is to get a grip on the sampling distribution of M . It turns out that the cdf-method works best. Using the fact that $M \leq y$ if and only if $Y_i \leq y$ for each i , derive an expression for the cdf of M and then for the pdf of M .
2. If you got the correct result above, you would realize that M is not a pivotal quantity. There is a simple fix, though. How does one need to modify M to get a pivotal quantity? Denote the modified version of M by U and write down its cdf F_U and pdf f_U .

- Suppose that we measured $n = 20$ trees and the largest one in the sample was 10 ft tall. Based on the pivotal quantity obtained in 2. above, give a symmetric 95%-confidence interval for the height of the tallest tree in the forest.

Problem 10.8.10 ((*) The pdf of the Student t distribution).

- Derive the pdf of the random variable $Y = \sqrt{X/k}$, where $X \sim \chi^2(k)$ (where $k \in \mathbb{N}$). (Note: use the fact that the pdf of a $\chi^2(k)$ -distribution is given by

$$f_X(x) = c_k x^{k/2-1} e^{-x/2} \mathbf{1}_{(0,\infty)}(x),$$

where c_k is a constant chosen so that f_X integrates to 1 (it is not important for this exercise what its exact value is, but if you are curious, $c_k = (2^{k/2} \Gamma(k/2))^{-1}$, where Γ is the Gamma function.))

- Let Z be a standard normal, independent of Y . Derive an expression for the pdf of the quotient Z/Y in terms of a single integral from 0 to ∞ . (Hint: Derive an expression for the cdf for Z/Y using a double integral. Observe that the integral in the z variable can be written using the cdf F_Z of the standard normal. Then differentiate inside the integral to turn F_Z into the normal pdf f_Z .)
- Evaluate this integral. (Hint: Use the identity $\int_0^\infty y^{k+2} e^{-\frac{1}{2}\beta y^2} dy = C_k \beta^{-k-3}$ for some constant C_k , when $\beta > 0$.)