1 Three loooong minutes on interval estimation, with a bit on hypothesis testing

erm: intervalestimation_new.tex and .pdf, Dec 2, 2010

In a previous section, we discussed point estimation: coming up with a specific estimate for a parameter or parameters. Here we discuss estimating an interval that is likely to include the true value of the parameter: interval estimation.

Our main concern will be estimating intervals for an unknown parameter or parameters.

As you might suspect the estimated intervals can be used to test hypotheses, so hypothesis testing and interval estimation are highly related topics. In fact, when we identified the likelihood ratio test in the max lik section, we identified a confidence interval.

Rather that starting by investigating the estimation an interval for an unknown parameter, I will, for simplicity, introduce intervals on a random variable where the distribution of that rv is known completely, including the parameter values, a case where there is no need to estimate an interval on any parameter, but one might still want to test a hypothesis

1.1 The interval on the rv X, or a statistic on X, when the distribution of X is known completely

Assume that the rv X has some known distribution, $f_X(x; \mu_x, \sigma_x^2)$, and that we **know** μ_x and σ_x^2. (Or more generally, we know $f_X(x : \theta)$ and θ)

1.1.1 A confidence interval on X

Given the above assumption that we know $f_X(x; \mu_x, \sigma_x^2)$ and know μ_x and σ_x^2, one can determine

$$\begin{aligned} & P\left[(\mu_x - w) \leq X \leq (\mu_x + w)\right] \\ = & \int_{\mu_x - w}^{\mu_x + w} f_X(u; \mu_x, \sigma_x^2) du \end{aligned}$$

which can, without loss of generality, be written

$$\begin{aligned} & P\left[(\mu_x - m\sigma_x) \le X \le (\mu_x + m\sigma_x)\right] \\ = & \int_{\mu_x - m\sigma_x}^{\mu_x + m\sigma_x} f_X(u; \mu_x, \sigma_x^2) du \end{aligned}$$

Given $f_X(u; \mu_x, \sigma_x^2)$, this is a function of only m: it is not statistic or rv.

Now choose m such that

$$\int_{\mu_x - m\sigma_x}^{\mu_x + m\sigma_x} f_X(u; \mu_x, \sigma_x^2) du = .95$$

and call the resulting m, $m_{.95}$

In which case

$$
\begin{aligned}
& P\left[(\mu_x - m_{.95}\sigma_x) \leq X \leq (\mu_x + m_{.95}\sigma_x)\right] \\
= & \int_{\mu_x - m_{.95}\sigma_x}^{\mu_x + m_{.95}\sigma_x} f_X(u; \mu_x, \sigma_x^2) du = .95
\end{aligned}
$$

X is a rv but the interval $(\mu_x - m_{.95}\sigma_x)$ to $(\mu_x + m_{.95}\sigma_x)$ is **not** random: no data or estimation was involved in determining this interval; it is not a statistic and it is not an estimated interval.

How would one interpret the interval $(\mu_x - m_{.95}\sigma_x)$ to $(\mu_x + m_{.95}\sigma_x)$? If X is randomly drawn from $f_X(x; \mu_x, \sigma_x^2)$, before it is drawn it has a 95% chance of being in this fixed interval - once drawn it either is, or is not, in the interval.

The interval, $(\mu_x - m_{.95}\sigma_x) \leq X \leq (\mu_x + m_{.95}\sigma_x)$, is called a *confidence* interval - it gives one confidence about where x will lie, but differs from many of the confidence intervals that you will encounter because this interval is not a random–it does not vary.

Note that by construction I constrained the interval to be centered on μ_x.
We choose the critical level of m, say $m_{.95}$ or $m_{.99}$.

Note a few things before we proceed One could rearrange the above interval to get

$$
\begin{aligned}
& .95 = P\left[(\mu_x - m_{.95}\sigma_x) \leq X \leq (\mu_x + m_{.95}\sigma_x)\right] \\
= & \Pr\left[-m_{.95} \leq Q \leq m_{.95}\right]
\end{aligned}
$$

where $Q = \frac{X - \mu_x}{\sigma_x}$ is a statistic, rv, with some density, $f_Q(q)$.

To figure out the appropriate value for $m_{.95}$, one can calculate $m_{.95}$ using one's knowledge of $f_X(x; \mu_x, \sigma_x^2)$, μ_x and, σ_x^2, like we did above.

Or one can determine it given knowledge of $f_Q(q)$ - it might be difficult to determine $f_Q(q)$. In terms of $f_Q(q)$, $m_{.95}$ is that m for which $\int_{-m}^{m} f_Q(q) dq = .95$.[1]

[1] Both ways can be useful. Note that if one knows $f_X(x; \mu_x, \sigma_x^2)$, μ_x, σ_x^2 and $Q = \frac{X - \mu_x}{\sigma_x}$, one can, in theory, determine $f_Q(q)$ – it is a derived distribution. And, if one knows $f_Q(q)$, μ_x, σ_x^2 and $Q = \frac{X - \mu_x}{\sigma_x}$ $(\Longrightarrow X = \sigma_x Q + \mu_x)$, one can, in theory, determine $f_X(x; \mu_x, \sigma_x^2)$.

Note that, by construction I constrained the interval to be centered on μ_x: μ_x plus and minus the same term I also constrained the interval to be gapless.

For our X, density known, there are many 95% confidence intervals, only one of which is centered on μ_x

We will mostly want to center our confidence intervals on the expected value of the variable, or parameter, of interest, but not always.

We will typically want to work with the shortest confidence interval because it is most informative interval–more on this in a bit.

The shortest confidence interval will often by centered on the expected value of the variable or parameter. For example if $f_X(u; \mu_x, \sigma_x^2)$ is symmetric and unimodal with the mode equal to μ_x, like the Normal, then the shortest X interval that spans 95% of the density is the one centered on μ_x.

In contrast, consider an asymmetric distribution or a U-shaped distribution. See the review question about Weird Shirley.[2]

[2] Note that the shortest interval might consist of segments that are not contiguous.

What I have said to here is quite general in that I did not assume a particular density for X or $f_Q(q)$ where $Q = \frac{X - \mu_x}{\sigma_x}$ There is nothing said above that requires that X is normally distributed.

For "fun", calculate these 95% and 99% confidence intervals for X for a few different specific specifications of $f_X(x; \mu_x, \sigma_x^2)$. For example, do it assuming X has a normal distribution, assuming X has a Poisson distribution with a mean of 3 (and a mean of 6), and assuming X has some t distribution. (Since the Poisson is a discrete distribution the range will consist of only a finite number of points. The fact that the Poisson is not a symmetric distribution brings to light the possibilty that we might sometimes prefer an interval that is not symmetric around μ)

The interval for other statistics Above we have specificed an interval

for X but we could have specified an interval for lots of other things as well. For example we could have assumed a random sample $X_1, X_2, ..., X_n$ and considered an interval for $g(X_1, X_2, ..., X_n) = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, the interval for the mean, this being only one of many examples.

Can you identify a confidence interval for $\overline{X}$ for random samples of size n, continuing to assume that the density of X is known completely? Doing this is a special case of some of the things we do below.

a hypothesis test If we randomly draw an x from $f_X(x; \mu_x, \sigma_x^2)$ we know it is a member of X.

However, if we observe a value y, where y is maybe a value of X and maybe not, how likely is it that y is a random draw from the population X?

The answer depends on the value of y.

If y is unlikely to have come from $f_X(x; \mu_x, \sigma_x^2)$ we might reasonably conclude that it did not. Two questions arise: how to determine the probability that it came from $f_X(x; \mu_x, \sigma_x^2)$? and how low does this probability have to be to judge that it did not come from $f_X(x; \mu_x, \sigma_x^2)$–reject that it came from $f_X(x; \mu_x, \sigma_x^2)$?

You choose the rejection probability. Typicall choices include .10, .05 and .01.

If, for example, you choose a rejection probability of .05, one tests the hypothesis null hypothesis that y is a draw from X by calculating the shortest .95 confidence interval for X. If y is in this interval, fails to reject the null hypothesis that y is a draw from X. If y is not in this interval one rejects the hypothesis that y is a draw from X.

You should conjecture that specifying an interval and doing a hypothesis test both require the specification of an interval.

Keep in mind that the calculation of the above interval, and the corresponding test required no estimation: there were no parameters to estimate.

Develope an argument as to why this test makes sense.

1.2 Interval estimation of unknown parameters

Turn now to our main concern: interval estimates for unknown parameters.

Assume that the rv X has some distribution, $f_X(x; \mu_x, \sigma_x^2)$, and we know its functional form, but we **not know either** μ_x, σ_x^2 **or both**.

The goal is to come up with interval estimates for μ and σ_x^2.

Assume $X_1, X_2, ..., X_n$ is a random sample

1.2.1 Consider the case where we know σ_x but only have an estimate of μ_x, $\widehat{\mu}_x = \overline{X}$.

Consider some rv R with density function $f_R(r)$ where R is a function of μ_x, σ_x, $\overline{X}$ and the sample size, n. We will obtain a confidence interval for R and then convert it ("pivot it") into a confidence interval for μ_x.

This process imposes two requirements on R: (1) one must choose a functional form for R such that its **density** is not a function of the unknown μ_x, and (2) it has to be "convertible."

> R is what is called a "pivotal quantity." Quoting Wikipedia, "In statistics, a pivotal quantity or pivot is a function of observations and unobservable parameters whose probability distribution does not depend on unknown parameters. Note that a pivot quantity need not be a statistic—the function and its value can depend on parameters of the model, but its distribution must not. If it is a statistic, then it is known as an ancillary statistic.
>
> More formally, given an independent and identically distributed sample $X = (X_1, X_2, ..., X_n)$ from a distribution with parameter θ, a function g is a pivotal quantity if the **distribution** of $g(X, \theta)$ is independent of θ.

Assume your chosen R is a pivotal quantity and you have figured out $f_R(r)$– typically difficult.[3] In which one you can determine

[3] This derived density will depend on the the functional form of R, the functional form of $f_X(x; \mu_x, \sigma_x^2)$, σ_x^2 and often have a parameter that is a function of the sample size, n. Remember that R was chosen so its density function does not have a parameter that depends on μ_x.

$$.95 = \Pr\left[r_{l.95} \leq R \leq \ r_{u.95}\right]$$

where $r_{l.95}$ and $r_{u.95}$ are the lower and upper bound on the shortest .95 interval for R. One then coverts this to a confidence interval on μ_x.

For example, imagine that $R = \frac{\overline{X}-\mu_x}{\frac{\sigma_x}{\sqrt{n}}}$, which would be an ideal choice is X is normally distributed. Here, I simply assume this R meets the above requirements. In which case

$$\begin{aligned} .95 &= \Pr\left[r_{l.95} \leq R \leq \ r_{u.95}\right] \\ &= \Pr\left[r_{l.95} \leq \frac{\overline{X}-\mu_x}{\frac{\sigma_x}{\sqrt{n}}} \leq \ r_{u.95}\right] \end{aligned}$$

Note that everything in R is known except for μ_x. Once you have determine the value for $r_{l.95}$ and $r_{u.95}$, this expression can be converted/pivoted as follows, to get a confidenc interval on μ_x. "Pivoting" around μ_x

$$\begin{aligned} .95 &= \Pr\left[r_{l.95} < \frac{\overline{X}-\mu_x}{\frac{\sigma_x}{\sqrt{n}}} < r_{u.95}\right] \\ &= \Pr\left[r_{l.95}\frac{\sigma_x}{\sqrt{n}} < \overline{X}-\mu_x < r_{u.95}\frac{\sigma_x}{\sqrt{n}}\right] \\ &= \Pr\left[-\overline{X}+r_{l.95}\frac{\sigma_x}{\sqrt{n}} < -\mu_x < -\overline{X}+r_{u.95}\frac{\sigma_x}{\sqrt{n}}\right] \\ &= \Pr\left[\overline{X}-r_{l.95}\frac{\sigma_x}{\sqrt{n}} > \mu_x > \overline{X}-r_{u.95}\frac{\sigma_x}{\sqrt{n}}\right] \\ &= \Pr\left[\overline{X}-r_{u.95}\frac{\sigma_x}{\sqrt{n}} < \mu_x < \overline{X}+r_{l.95}\frac{\sigma_x}{\sqrt{n}}\right] \end{aligned}$$

making the .95 confidence interval

$$\overline{X}-r_{u.95}\frac{\sigma_x}{\sqrt{n}} < \mu_x < \overline{X}+r_{l.95}\frac{\sigma_x}{\sqrt{n}}$$

So, what does this mean? The interval $\overline{X}-r_{u.95}\frac{\sigma_x}{\sqrt{n}}$ to $\overline{X}+r_{l.95}\frac{\sigma_x}{\sqrt{n}}$ varies from sample to sample and 95% of these intervals will contain μ. For a given sample, $\overline{x}-r_{u.95}\frac{\sigma_x}{\sqrt{n}}$ to $\overline{x}+r_{l.95}\frac{\sigma_x}{\sqrt{n}}$ is a numerical range (e.g. 2 to 8) and it highly likely that μ will fall in this range, but not for sure.

$\overline{X}-r_{u.95}\frac{\sigma_x}{\sqrt{n}}$ to $\overline{X}+r_{l.95}\frac{\sigma_x}{\sqrt{n}}$is our interval (.95) estimate for μ_x.

(As an aside, note the distinction between $\frac{\overline{X}-\mu}{\frac{\sigma_x}{\sqrt{n}}}$ and $\frac{X-\mu}{\sigma_x}$and note that both have a standard normal distribution if X is normally distributed.)

If one is, in addition, willing to assume X is normally distributed If X is normally distributed then $\frac{\overline{X}-\mu}{\frac{\sigma_x}{\sqrt{n}}}$ has a standard normal distribution, where $\frac{\sigma_x}{\sqrt{n}}$ is the standard deviation of $\overline{X}$ (See MGB page 374 and page 381). So $f_R(r)$ is not a function of μ_x, the unknown parameter. And, -1.96 and 1.96 are the critical values for the 95% confidence interval: $r_{l.95} = -1.96$ and $r_{u.95} = 1.96$. So

$$\Pr\left[-1.96 < \frac{\overline{X}-\mu}{\frac{\sigma_x}{\sqrt{n}}} < 1.96\right] = .95$$

So, the C.I. is

$$\overline{X} - 1.96\frac{\sigma_x}{\sqrt{n}} < \mu_x < \overline{X} + 1.96\frac{\sigma_x}{\sqrt{n}}$$

1.2.2 Consider the case where we know μ_x, but only have an estimate of σ_x, $\hat{\sigma}_x$.

Notationally, let $\hat{\sigma}_x$ denote an estimate of σ_x and let $\hat{\boldsymbol{\sigma}}_x$ denote the corresponding estimat**or** of σ_x.[4]

Find a 95% confidence interval for σ^2 and interpret it.

Here we are looking for a rv, call it Q, such that Q is a function of μ_x, σ_x, n, and the sample $(X_1, X_2, ...X_n)$ but where the parameters of the density function for Q are not a function of the unknown σ_x

Assume your chosen Q meets the requirements, it is a pivotal quantity, and you have figured out $f_Q(q)$. In which case, you can determine

$$.95 = \Pr\left[q_{l.95} \leq Q \leq \; q_{u.95}\right]$$

where $q_{l.95}$ and $q_{u.95}$ are the lower and upper bound on the shortest .95 interval for Q.

One then coverts this to a confidence interval on σ_x.

For example, imagine that $Q = \frac{\sum_{i=1}^{5}(X_i-\mu)^2}{\sigma_x^2} = \sum_{i=1}^{5}\left(\frac{X_i-\mu}{\sigma}\right)^2$, which would be an ideal choice is X is normally distributed. Here, I simply assume this Q meets the above requirements. In which case

$$\begin{aligned} .95 &= \Pr\left[q_{l.95} \leq Q \leq \; q_{u.95}\right] \\ &= \Pr\left[q_{l.95} \leq \frac{\sum_{i=1}^{5}(X_i-\mu)^2}{\sigma_x^2} \leq \; q_{u.95}\right] \end{aligned}$$

Once you have determined $f_Q(q)$, which is not a function σ_q^2, determine the numerical values of $q_{l.95}$ and $q_{u.95}$. The above expression can be pivoted to obtain.

[4] Note that $\hat{\sigma}_x$is an estimate of σ_x and that this is not the same thing as the estimated standard deviation of the sample.

$$\begin{aligned}
.95 &= \Pr\left[q_{l.95} < \frac{\sum_{i=1}^{5}(X_i-\mu)^2}{\sigma_x^2} < q_{u.95} \right] \\
&= \Pr\left[\frac{1}{q_{l.95}} > \frac{\sigma_x^2}{\sum_{i=1}^{5}(X_i-\mu)^2} > \frac{1}{q_{u.95}} \right] \\
&= \Pr\left[\frac{1}{q_{l.95}} \sum_{i=1}^{5}(X_i-\mu)^2 > \sigma_x^2 > \frac{1}{q_{u.95}} \sum_{i=1}^{5}(X_i-\mu)^2 \right] \\
&= \Pr\left[\frac{1}{q_{u.95}} \sum_{i=1}^{5}(X_i-\mu)^2 < \sigma_x^2 < \frac{1}{q_{l.95}} \sum_{i=1}^{5}(X_i-\mu)^2 \right]
\end{aligned}$$

making the .95 confidence interval for σ_x^2

$$\frac{1}{q_{u.95}} \sum_{i=1}^{5}(X_i-\mu)^2 \text{ to } \frac{1}{q_{l.95}} \sum_{i=1}^{5}(X_i-\mu)^2$$

This interval is random, varying from sample to sample, and 95% of them will contain σ_x^2.

If one is, in addition, willing to assume X is normally distributed

If X is normally distribution Q has a Chi-squared distribution with parameter (degrees for freedom) n–see MGB page 243–so is a pivotal quantitiy. Therefore

$$.95 = \Pr\left[\chi^2_{(q<Q=.025:n)} < \frac{\sum_{i=1}^{5}(X_i-\mu)^2}{\sigma_x^2} < \chi^2_{(q>Q=.025:n)} \right]$$

where $\chi^2_{(q<Q=.025:n)}$ is the value of Q such that 2.5% of the distribution is less than q. And the c.i. for σ_x^2 is

$$\frac{1}{\chi^2_{(q>Q=.25:n)}} \sum_{i=1}^{5}(X_i-\mu)^2 \text{ to } \frac{1}{\chi^2_{(q<Q=.025:n)}} \sum_{i=1}^{5}(X_i-\mu)^2$$

For example, with 5 degrees of freedom, 2.5% of the values of Q are less than .831 and 2.5% are greater than 12.8.

$$.95 = \Pr\left[.831 < \frac{\sum_{i=1}^{5}(X_i-\mu)^2}{\sigma_x^2} < 12.8\right]$$

and the c.i, is

$$.07\,8123\sum_{i=1}^{5}(X_i-\mu)^2 < \sigma_x^2 < 1.\,203\,4\sum_{i=1}^{5}(X_i-\mu)^2$$

1.2.3 Now consider the case where both μ_x and σ_x^2 are unknown.

First find a 95% confidence interval for μ_x in terms of $\overline{X}$ and $\widehat{\sigma}_x$, and interpret this interval.

Here we are looking for a rv, call it H, such that H is a function of μ_x, $\overline{X}$, n and $\widehat{\sigma}_x$, but where the parameters of the density function for H are not a function of μ_x nor σ_x.

Assume your chosen H meets the requirements and you have figured out $f_H(h)$. In which one you can determine

$$.95 = \Pr\left[h_{l.95} \leq H \leq\ h_{u.95}\right]$$

where $h_{l.95}$ and $h_{u.95}$ are the lower and upper bound on the shortest .95 interval for H.

One then coverts/pivots this to a confidence interval on μ_x.

For example, imagine that $H = \frac{\overline{X}-\mu_x}{\frac{\widehat{\sigma}_x}{\sqrt{n}}}$ where $\widehat{\sigma}_x^2 = \frac{\sum_{i=1}^{5}(X_i-\overline{X})^2}{(n-1)}$ is an estimate of σ_x^2. This would be an ideal choice for H if X is normally distributed. Here, I simply assume this H meets the above requirements. In which case

$$\begin{aligned} .95 &= \Pr\left[h_{l.95} \leq H \leq\ h_{u.95}\right] \\ &= \Pr\left[h_{l.95} \leq \frac{\overline{X}-\mu_x}{\frac{\widehat{\sigma}_x}{\sqrt{n}}} \leq\ h_{u.95}\right] \end{aligned}$$

Pivoting this around μ_x

$$\begin{aligned}
.95 &= \Pr\left[h_{l.95} < \frac{\overline{X} - \mu_x}{\frac{\widehat{\sigma}_x}{\sqrt{n}}} < h_{u.95}\right] \\
&= \Pr\left[h_{l.95}\frac{\widehat{\sigma}_x}{\sqrt{n}} < \overline{X} - \mu_x < h_{u.95}\frac{\widehat{\sigma}_x}{\sqrt{n}}\right] \\
&= \Pr\left[-\overline{X} + h_{l.95}\frac{\widehat{\sigma}_x}{\sqrt{n}} < -\mu_x < -\overline{X} + h_{u.95}\frac{\widehat{\sigma}_x}{\sqrt{n}}\right] \\
&= \Pr\left[\overline{X} - h_{l.95}\frac{\widehat{\sigma}_x}{\sqrt{n}} > \mu_x > \overline{X} - h_{u.95}\frac{\widehat{\sigma}_x}{\sqrt{n}}\right] \\
&= \Pr\left[\overline{X} - h_{u.95}\frac{\widehat{\sigma}_x}{\sqrt{n}} < \mu_x < \overline{X} - h_{l.95}\frac{\widehat{\sigma}_x}{\sqrt{n}}\right]
\end{aligned}$$

And 95% of the intervals, $\overline{X} - h_{u.95}\frac{\widehat{\sigma}_x}{\sqrt{n}}$ to $\overline{X} - h_{l.95}\frac{\widehat{\sigma}_x}{\sqrt{n}}$ will contain μ_x.

If one is, in addition, willing to assume X is normally distributed In this case, it can be shown that the statistic H has a t distribution with parameter $(n-1)$ – see MGB page ??? Since the t distribution is symmetrical $h_{l.95} = -h_{u.95} = t_{(h>H=.025:n-1)}$. So,

$$.95 = \Pr\left[-t_{(h>H=.025:n-1)} \leq \frac{\overline{X} - \mu_x}{\frac{\widehat{\sigma}_x}{\sqrt{n}}} \leq t_{(h>H=.025:n-1)}\right]$$

And 95% of the intervals, $\overline{X} - t_{(h>H:n-1)}\frac{\widehat{\sigma}_x}{\sqrt{n}}$ to $\overline{X} + t_{(h>H=.025:n-1)}\frac{\widehat{\sigma}_x}{\sqrt{n}}$ will contain μ_x.

For example, if $n = 5$, 2.766 is the critical t value for .025 when the parameter of the t is 4. Making

$$.95 = \Pr\left[-2.766 \leq \frac{\overline{X} - \mu_x}{\frac{\widehat{\sigma}_x}{\sqrt{n}}} \leq 2.766\right]$$

And 95% of the intervals, $\overline{X} - 2.766\frac{\widehat{\sigma}_x}{\sqrt{n}}$ to $\overline{X} + 2.766\frac{\widehat{\sigma}_x}{\sqrt{n}}$ will contain μ_x.

Now find a 95% **confidence interval for** σ_x^2 **in terms of** $\overline{X}$ **and** $\widehat{\sigma}_x$**, and interpret this interval.** Here we are looking for a rv, call it W, such that W is a function of σ_x^2, $\overline{X}$, n and $\widehat{\sigma}_x$, but where the parameters of the density function for W are not a function of μ_x nor σ_x.

Assume your chosen W meets the requirements and you have figured out $f_W(w)$. In which one you can determine

$$.95 = \Pr\left[w_{l.95} \leq W \leq \ w_{u.95}\right]$$

where $w_{l.95}$ and $w_{u.95}$ are the lower and upper bound on the shortest .95 interval for W.

One then coverts/pivots this to a confidence interval on σ_x^2.

Imagine, for example, that $W = \frac{\sum_{i=1}^{5}(X_i-\overline{X})^2}{\sigma_x^2} = \frac{(n-1)\widehat{\sigma}_x^2}{\sigma_x^2}$ where $\widehat{\sigma}_x^2 = \frac{\sum_{i=1}^{5}(X_i-\overline{X})^2}{(n-1)}$ is an estimate of σ_x^2.

If one is willing to assume that X **is normally distributed,**

W has a Chi-square distribution with parameter (degrees for freedom) $n-1$–see MGB page ???–so W is a pivotal quantity.

$$.95 = \Pr\left[\chi^2_{(q<Q=.025:n-1)} < \frac{(n-1)\widehat{\sigma}_x^2}{\sigma_x^2} < \chi^2_{(q>Q=.025:n-1)}\right]$$

Pivoting this around σ_x^2

$$\begin{aligned}
.95 &= \Pr\left[\chi^2_{(q<Q=.025:n-1)} < \frac{(n-1)\widehat{\sigma}_x^2}{\sigma_x^2} < \chi^2_{(q>Q=.025:n-1)}\right] \\
&= \Pr\left[\frac{1}{\chi^2_{(q<Q=.025:n-1)}} > \frac{\sigma_x^2}{(n-1)\widehat{\sigma}_x^2} > \frac{1}{\chi^2_{(q>Q=.025:n-1)}}\right] \\
&= \Pr\left[\frac{1}{\chi^2_{(q<Q=.025:n-1)}}(n-1)\widehat{\sigma}_x^2 > \sigma_x^2 > \frac{1}{\chi^2_{(q>Q=.025:n-1)}}(n-1)\widehat{\sigma}_x^2\right] \\
&= \Pr\left[\frac{1}{\chi^2_{(q>Q=.025:n-1)}}(n-1)\widehat{\sigma}_x^2 < \sigma_x^2 < \frac{1}{\chi^2_{(q<Q=.025:n-1)}}(n-1)\widehat{\sigma}_x^2\right]
\end{aligned}$$

and the c.i. for σ_x^2 is $\frac{1}{\chi^2_{(q>Q=.025:n-1)}}(n-1)\widehat{\sigma}_x^2$ to $\frac{1}{\chi^2_{(q<Q=.025:n-1)}}(n-1)\widehat{\sigma}_x^2$

If for example there are 5 observations, the parameter of the Chi-squared distribution is 4, and 2.5% of the values of Q are less than .484 and 2.5% are greater than 11.1, so

$$\begin{aligned} .95 &= \Pr\left[.484 < \frac{(n-1)\widehat{\sigma}_x^2}{\sigma_x^2} < 11.1\right] \\ &= \Pr\left[0.090\,09(n-1)\widehat{\sigma}_x^2 < \sigma_x^2 < 2.066\,1(n-1)\widehat{\sigma}_x^2\right] \\ &= \Pr[.45045\widehat{\sigma}_x^2 < \sigma_x^2 < 10.331\widehat{\sigma}_x^2] \end{aligned}$$

and the 95% confidence interval for σ_x^2 is $0.45045\widehat{\sigma}_x^2$ to $10.331\widehat{\sigma}_x^2$

1.2.4 Consider another estimated interval

Assume $y_i = \alpha + \beta x_i + \varepsilon_i$ where $\varepsilon \sim N(0, \sigma_\varepsilon^2)$. Consider the rv $T = \frac{\widehat{\beta}-\beta}{\hat{\sigma}_{\widehat{\beta}}}$ where $\widehat{\beta}$ is the least-squares estimate of β, and $\hat{\sigma}_{\widehat{\beta}}$ is the corresponding point estimate of the standard deviation of $\widehat{\beta}$.[5]

At issue is the density function of $T = \frac{\widehat{\beta}-\beta}{\hat{\sigma}_{\widehat{\beta}}}$, $f_T(t)$. It can be shown that T has a t-distribution with parameter (degrees of freedom) $n-2$, so T is a pivotal quantity. Therefore,

$$.95 = \Pr\left[-t_{.95(n-2)} \leq \frac{\widehat{\beta}-\beta}{\hat{\sigma}_{\widehat{\beta}}} \leq t_{.95(n-2)}\right]$$

where the critical t values are chosen from the t Table with parameter $n-2$ so that there is 2.5% of the density left of $-t_{.95(n-2)}$ and 2.5% right of $t_{.95(n-2)}$.[6]

Rearranging:

$$\begin{aligned}
.95 &= \Pr\left[-t_{.95(n-2)} \leq \frac{\widehat{\beta}-\beta}{\hat{\sigma}_{\widehat{\beta}}} \leq t_{.95(n-2)}\right] \\
&= \Pr\left[-t_{.95(n-2)}\hat{\sigma}_{\widehat{\beta}} \leq \widehat{\beta}-\beta \leq t_{.95(n-2)}\hat{\sigma}_{\widehat{\beta}}\right] \\
&= \Pr\left[-\widehat{\beta} - t_{.95(n-2)}\hat{\sigma}_{\widehat{\beta}} \leq -\beta \leq -\widehat{\beta} + t_{.95(n-2)}\hat{\sigma}_{\widehat{\beta}}\right] \\
&= \Pr\left[\widehat{\beta} + t_{.95(n-2)}\hat{\sigma}_{\widehat{\beta}} \geq \beta \geq \widehat{\beta} - t_{.95(n-2)}\hat{\sigma}_{\widehat{\beta}}\right] \\
&= \Pr\left[\widehat{\beta} - t_{.95(n-2)}\hat{\sigma}_{\widehat{\beta}} \leq \beta \leq \widehat{\beta} + t_{.95(n-2)}\hat{\sigma}_{\widehat{\beta}}\right]
\end{aligned}$$

The interval $\widehat{\beta} - t_{.95(n-2)}\hat{\sigma}_{\widehat{\beta}}$ to $\widehat{\beta} + t_{.95(n-2)}\hat{\sigma}_{\widehat{\beta}}$ has a sampling variation. 95% of these intervals will contain β.

---

[5] Just to make sure we are all clear on the notation, $\sigma^2_{\widehat{\beta}}$ is the "true" sampling variance of the estimator of β, the $\widehat{\beta}$. And $\widehat{\sigma}^2_{\widehat{\beta}}$ is our estimate of this sampling variance.

[6] $t_{.95(n-2)}$ is the t such that $\int_{-t}^{t} f_T(t : n-2) = .95$ where $f_T(t : n-2)$ is the student t density with parameter $n-2$. Recollect that the t distribution has a mean of zero.

1.3 Another method for finding a confidence interval: the "statistical method" (MGB page 389)

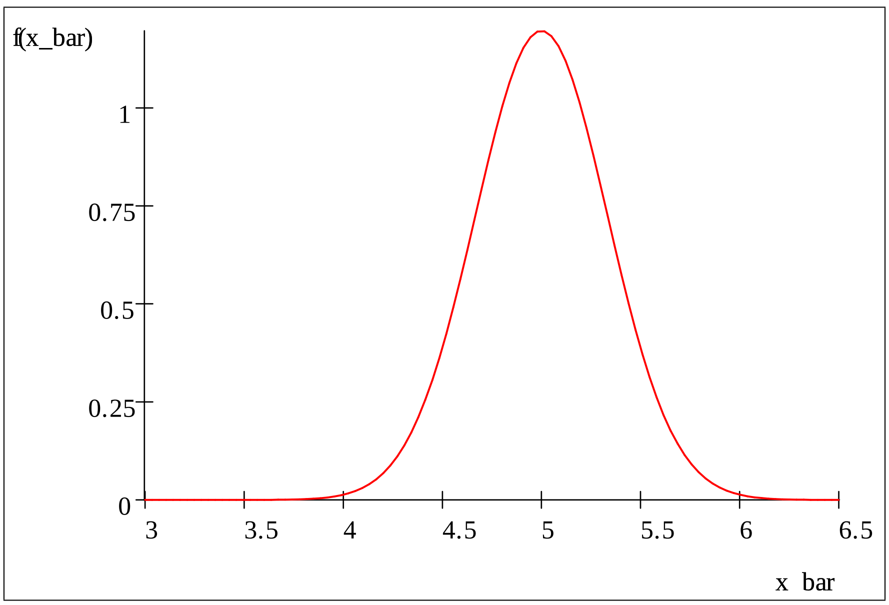
Only an example, but not the one in the book:

Assume the RV X is normally distributed with unknown mean, μ_x, and σ_x^2, known, equal to 1, $\phi_X(x:\mu_x,1)$. One has a random sample of size n, and one wants a c.i. on μ_x based on $\overline{x}$

One point estimates μ_x with $\overline{x}$, the numerical sample mean.

Let $\overline{x}$ denote the numerical sample mean and $\overline{X}$ the rv.

$\overline{X}$ has some density $f_{\overline{X}}(\overline{x})$ which in this case is simply $\phi_{\overline{X}}(\overline{x}:\mu_x,\frac{1}{\sqrt{n}})$. For example if $\mu_x=5$ and $n=9$ is NormalDen $\left(u;5,\frac{1}{3}\right)$



One can use $\overline{x}$, the numerical sample mean, and $\phi_{\overline{X}}(\overline{x}:\mu_x,\frac{1}{\sqrt{n}})$ to derive a .95 confidence interval for μ_x. The method is called the "statistical method".

Put simply, we ask the following question, "What values of μ_x would make it highly unlikely that we would observe our $\overline{x}$?"

Specifically what value of μ_x would make the probability of observing a $\overline{X}$ of $\overline{x}$ or less .025. One determines this by solving $.025=\int_{-\infty}^{\overline{x}}\phi(t:\mu_x,\frac{1}{\sqrt{n}})dt$ for μ_x

This value of μ_x, call it $\overleftarrow{\mu}_x$, is one of the limits on the .95 confidence interval on μ_x

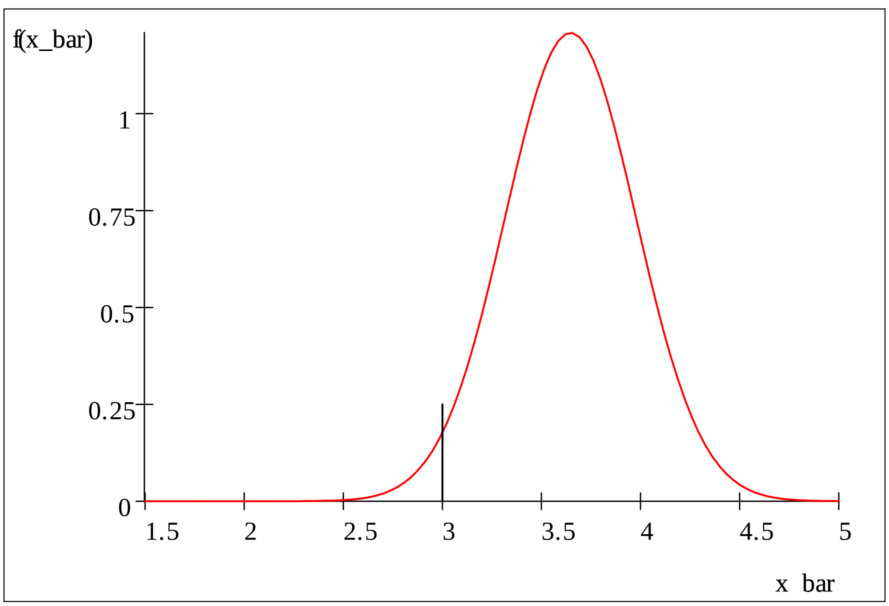
Then determine the value of μ_x would make the probability of observing a $\overline{X}$ of $\overline{x}$ or more .025. One determines this by solving $.025 = \int_{\overline{x}}^{\infty} \phi(t : \mu_x, \frac{1}{\sqrt{n}})dt$ for μ_x

This value of μ_x, call it $\overrightarrow{\mu}_x$,is the other limit on the .95 confidence interval on μ_x

A numerical example: assume $\overline{x} = 3$ and $n = 9$. In which case $.025 = \int_{-\infty}^{3} \phi(t : \mu_x, \frac{1}{3})dt$

Asking my software to solve this $.025 = \int_{-5}^{3} \text{NormalDen}\,(t; \mu, .33)\,dt$, Solution is: $\{[\mu = 3.646\,8]\}$

Checking that makes sense, if $\mu = 3.646\,8$, NormalDen $(x; 3.6468, .33)$

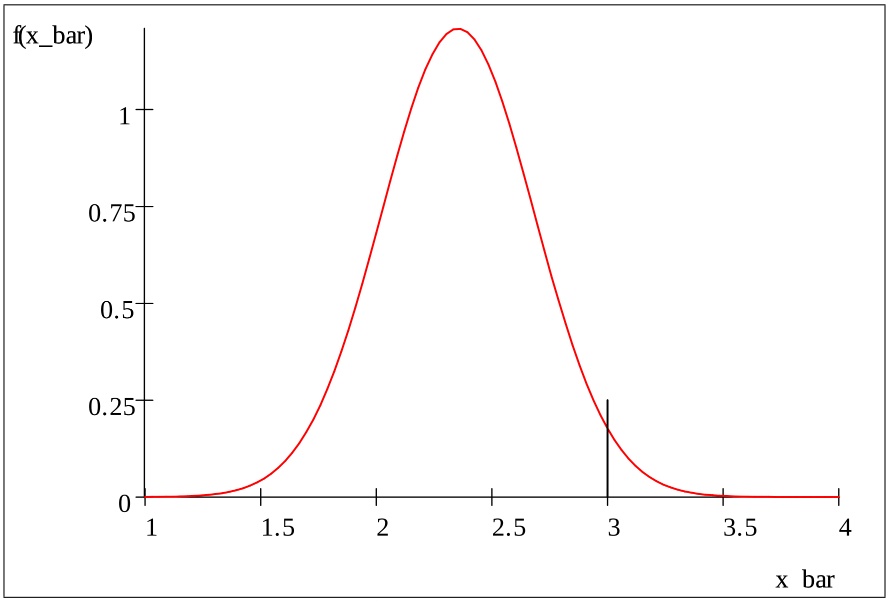


if $\mu_x = 3.6484$

And the critical value of $\overline{X}$ for $p = .025$ is NormalInv $(.025; 3.6468, .33) = 3.0$. That is, if $\mu = 3.646\,8$ the chances that $\overline{X}$ is less than or equal to 3.0 is 2.5%. So, one of the bounds on the c.i. is 3.6468.

For the other bound, $.025 = \int_{3}^{\infty} \phi(t : \mu_x, \frac{1}{3})dt$

Asking my software to solve this $.025 = \int_{3}^{6} \text{NormalDen}\,(t; \mu, .33)\,dt$, Solution is: $\{[\mu = 2.353\,2]\}$. Are you surprised? Let's investigate,



if $\mu_x = 2.3532$

The other critical value of $\overline{X}$ for $p = .025$ is NormalInv $(.975; 2.3532, .33) = 3.0$. That is, if $\mu = 2.3532$ the chances that $\overline{X}$ is greater than or equal to 3.0 is 2.5% (97.5% is less than 3)

So, for this numerical example, our estimated 95% confidence interval for μ_x is 2.3532 to 3.6468. Wow.

Note that we got the same answer above using the pivotal method.

The pivotal method identified the c.i. for μ_x based on $\overline{X}$, and assuming knowledge of σ_x^2, as

$$\overline{X} - 1.96\frac{\sigma_x}{\sqrt{n}} < \mu_x < \overline{X} + 1.96\frac{\sigma_x}{\sqrt{n}}$$

Plugging $\overline{X} = 3$, $n = 9$ and $\sigma_x^2 = 1$, one gets, $2.3467 < \mu_x < 3.653$, which is the same, except for rounding errors. WOW.

You should contemplate on the difference(s) between the pivotal method and the statistical method for finding a c.i.

1.4 Consider the following simulation exercise:

Assume the rv X is normally distributed with parameters μ_x and σ_x^2.

Assume specific numerical values for μ_x and σ_x^2, and based on these values draw a random sample of n values of X–you choose n.

Now pretend you do not know the values of μ_x and σ_x^2 that produced your sample

For this sample estimate μ_x and σ_x^2 and then use these estimates to get the estimated .95 confidence interval for u_x.

For the same specific numerical values for μ_x and σ_x^2, draw S additional random samples of size n. You choose S.

For each of these S additional samples, as above, estimate a .95 confidence interval for μ_x

Graph, all on the same graph, these $S+1$ confidence intervals. I might try a graph with a stack of $S+1$ horizontal lines or $S+1$ vertical lines. Each line representing the length and position of a different estimated c.i.

What did you notice and what did you learn from this simulation exercise? What proportion of your estimated c.i. include your assumed value for μ_x?