

Crash Course in Statistics for Neuroscience Center Zurich University of Zurich

Dr. C.J. Luchsinger

6 Estimators and Confidence Intervals (CI)

Further readings: Chapters 7 and 9 in Stahel or Chapters 6 and 7 in Cartoon Guide

6.1 Estimators

6.1.1 Motivation

Below is an (ordered) sample of length $n = 11$. Suppose we know that this sample results from a sequence of iid $\mathcal{N}(\mu, 1)$ -RV's. We now want to estimate the unknown μ . What do you do?

-0.21, 0.11, 0.39, 0.64, 1.24, 1.46, 1.48, 1.51, 1.95, 2.89, 3.53

What characteristics do you expect from a *good* estimator?

6.1.2 Definition and desired Characteristics of an Estimator

Let us write $\hat{\mu}_n$ for an estimator of μ , if we have n data points and define:

Definition 6.1 [Estimator] *An Estimator $\hat{\mu}_n := \hat{\mu}_n(x_1, \dots, x_n)$ of an unknown parameter μ is an arbitrary function of the data.*

Definition 6.1 does not help us a lot. We want a good estimator (not just "some function of the data").

A technical detail: For example

$$\bar{x}$$

is the arithmetic mean of the data. If the data is given, this is simply a number, and not a random variable anymore. But to analyze whether the arithmetic mean of the data is a good estimator, we must be aware of the fact, that there is really a sequence of iid RV's behind, and this is really a

$$\bar{X}(\omega_1) = \bar{x},$$

if the outcome is some ω_1 . To find out whether \bar{x} is a good estimator, **we must analyze the random variable \bar{X}** . In other words: Don't believe too much in your *current* data: if we could sample again, the result could be quite different - it is random; that's why we model it with a RV. This is true and important for all statistical data analysis!!!

Amongst other possible characteristics, the following 4 are most widely used. We will illustrate the definitions with the problem from 6.1.1: a sample coming from iid $\mathcal{N}(\mu, 1)$ -RV's.

Characteristic 1: unbiased We call an estimator $\hat{\mu}_n$ for μ **unbiased** (GERMAN: erwartungstreu), if

$$E[\hat{\mu}_n] = \mu.$$

Otherwise, we define

$$b := E[\hat{\mu}_n] - \mu$$

to be the **bias** of this estimator.

The **arithmetic mean \bar{X} is an unbiased estimator**. Proof is a good homework for people who like Maths: from Chapter 3, Lemma 3.3: in $\mathcal{N}(\mu, 1)$, $\hat{\mu}_n := \bar{X}$ is unbiased:

$$E[\hat{\mu}_n] := E[\bar{X}] := E\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n} E\left[\sum_{i=1}^n X_i\right] = \frac{1}{n} \sum_{i=1}^n E[X_i] = \frac{1}{n} \sum_{i=1}^n \mu = \mu$$

□

Characteristic 2: consistent We call an estimator $\hat{\mu}_n$ for μ consistent, if

$$\hat{\mu}_n \rightarrow \mu$$

for $n \rightarrow \infty$. From Theorem 5.1 it follows that \bar{X} is consistent. This looks really great: \bar{X} seems to meet all requirements. Lets go to number 3:

Characteristic 3: small(est) variance Good homework for people who like Maths: Chapter 3, Lemma 3.4 and Lemma 3.5 b): in $\mathcal{N}(\mu, \sigma^2)$, $\hat{\mu}_n := \bar{X}$ has the following variance:

$$\begin{aligned} V[\hat{\mu}_n] &:= V[\bar{X}] := V\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n^2} V\left[\sum_{i=1}^n X_i\right] \\ &= \frac{1}{n^2} \sum_{i=1}^n V[X_i] = \frac{1}{n^2} \sum_{i=1}^n \sigma^2 = \frac{\sigma^2}{n}, \end{aligned}$$

and the so called **standard error of the sample mean** is therefore

$$\frac{\sigma}{\sqrt{n}},$$

with the holy \sqrt{n} in the denominator. This " \sqrt{n} " will stay with us throughout the whole rest of this course. It can be shown, that variance (and standard error) of unbiased estimators can not be smaller than σ^2/n (and σ/\sqrt{n} respectively), when data comes from a $\mathcal{N}(\mu, \sigma^2)$. So again, \bar{X} seems to meet all requirements so far. But:

Characteristic 4: Robustness Let us look at two competing estimators for estimating μ with data from a $\mathcal{N}(\mu, \sigma^2)$:

* arithmetic mean \bar{X} vs

* median

To summarize: **With data coming from $\mathcal{N}(\mu, \sigma^2)$, \bar{X} is the best estimator for μ** (unbiased, consistent, smallest variance). But: **if data quality is questionable (the usual case!), we use the median.** Median is unbiased and consistent too, but has a slightly larger variance than the arithmetic mean. That does not bother us. **So, in practice: use the median!**

6.1.3 Analogy of Theoretical Expressions and their Estimators

Remember from 3.3: we treat data such that we assign probability $1/n$ to each of the n data points. Then, due to LLN (at least for first line this is obvious, the other two results need difficult mathematics, which we omit in this course):

Theoretical Expression (<i>Model</i>) (discrete case)	Estimated by (<i>Data</i>)
$E[X] = \sum_x xP[X = x]$	$\sum_i x_i \frac{1}{n} \quad (= \bar{x})$
$V[X] = \sum_x (x - \mu)^2 P[X = x]$	$\sum_i (x_i - \bar{x})^2 \frac{1}{n}$
$Cor(X, Y) = \frac{\sum_{(x,y)} (x - \mu_X)(y - \mu_Y) P[X=x, Y=y]}{\sqrt{(\sum_x (x - \mu_X)^2 P[X=x])(\sum_y (y - \mu_Y)^2 P[Y=y])}}$	$\frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{(\sum_i (x_i - \bar{x})^2)(\sum_j (y_j - \bar{y})^2)}}$

The expressions on the right side (the estimators) are sometimes called empirical (or sample) mean, variance and correlation (the others are the theoretical (or population) mean, variance and correlation respectively).

6.1.4 n or $(n - 1)$, that's the question (it's not *the* question, by the way...)

Some people make a terrible fuss about whether you should divide

$$\sum_{i=1}^n (x_i - \bar{x})^2$$

by n or by $(n - 1)$ (best answer is $(n + 1)$ I say, but too complicated to justify in this course). Well, answer is simple:

$$\frac{1}{n - 1} \sum_{i=1}^n (x_i - \bar{x})^2 \tag{6.1}$$

is an unbiased estimator (see 6.1.2) of σ^2 , while

$$\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \tag{6.2}$$

is biased. This is shown by complicated mathematics (we omit this). But revisit 4.2.3 now to confirm these results. **We usually use (6.1)**, because statistical tables are built up with this expression (see chapter 7). Alternatively, the following animal

$$\hat{\sigma}_n^2 := (1.4826 \times \text{MAD})^2 := (1.4826 \times \text{median}(|x_i - \text{median}(x_1, \dots, x_n)|))^2 \tag{6.3}$$

is extremely robust. “MAD” is “Median of Absolut Deviations”. The factor ”1.4826” has been found to give better results than the virgin MAD for normal random variables.

By the way:

$$\sqrt{\frac{1}{n - 1} \sum_{j=1}^n (x_j - \bar{x})^2} \tag{6.4}$$

is a biased estimator for σ . Again, **we use (6.4) to estimate the standard deviation** σ in the so called t-test in chapter 7 and to construct confidence intervals in 6.2.

Both, (6.1) and (6.4) are consistent estimators.

In chapter 8 (Regression) other estimators than the ones treated might be used from time to time.

6.1.5 Estimation of "p" in the Be(p) (or Binomial) Situation

If data x_1, x_2, \dots, x_n are "0"'s (Death, Failure) or "1"'s (Survival, Success) in a situation where you count the number of successes, then the number of successes is simply

$$\sum_{i=1}^n x_i$$

and has by definition a Binomial distribution with parameters p, n : $\text{Bin}(n, p)$. We might be interested in estimating the "p" by some \hat{p} and obviously the canonical choice is

$$\hat{p} := \frac{1}{n} \sum_{i=1}^n x_i$$

(not even Mathematicians are able to complicate this solution: it can be shown that this is best choice). We might then be interested to know something about the accuracy of this estimate. We compute

$$\text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \text{Var}\left(\sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{1}{n^2} \sum_{i=1}^n p(1-p) = \frac{p(1-p)}{n}.$$

The standard error of the estimate of the p (sampling error) is therefore

$$\sqrt{\frac{p(1-p)}{n}}.$$

This is not quite satisfying. Why?

This is by the way best you can do. We will come back to this situation in 6.2.

Elections in the U.S.: Suppose $n = 10'000$ (people interviewed) and true (but unknown values) are $p = 0.52$ for Bush and $(1-p) = 0.48$ for Kerry. What can we now say if 5'058 of those 10'000 said they are going to vote for Bush? To be revisited in 6.2.

6.2 Confidence Intervals

... give us an idea of how precise our estimate is.

6.2.1 What is a confidence interval (CI) - and a common misunderstanding

The misunderstanding: News: “In a Gallup poll 50.6 % of US Citizens said they would re-elect President George W. Bush if elections were to be held tomorrow (10'000 interviews). Confidence interval for this poll is such that the true value is in the interval [49.6%, 51.6%] with 95 % probability.” Does this make sense?

See picture in Cartoon Guide, page 120.

Definition 6.2 [Confidence Interval CI] *A confidence interval CI for some parameter $\theta \in \mathbb{R}$ with coefficient $(1 - \alpha)$ is a random set [German: Menge] of \mathbb{R} with the characteristic that*

$$P[\theta \in CI] = 1 - \alpha$$

for any $\theta \in \mathbb{R}$.

Remark: Again, θ is unknown but fixed. CI is random! But once we have data points x_1, \dots, x_n , the CI becomes fixed too (for example these [49.6%, 51.6%]). We then (should) say: “[49.6%, 51.6%] is a realisation of a confidence interval with coefficient $(1 - \alpha)$ for θ .” Nobody does that (except mathematical statisticians), so we simply say: “[49.6%, 51.6%] is a $(1 - \alpha)$ -confidence interval for θ .”

6.2.2 How to construct CI

6.2.2.1 $\mathcal{N}(\mu, \sigma^2)$: CI for μ with σ^2 known to us

Given data x_1, \dots, x_n , compute arithmetic mean \bar{x} . Switch to RV's: look at \bar{X} . Distribution of \bar{X} is:

mean=

variance=

distribution type is:

together:

So

$$\frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}}$$

is a $\mathcal{N}(0, 1)$ due to the "Z-Transform" (see 4.2.2). But that's great: we know critical values of the $\mathcal{N}(0, 1)$ -distribution:

$$P\left[-1.96 \leq \frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \leq 1.96\right] = 95\%. \quad (6.5)$$

Obviously we are heading to a $(1 - \alpha) = 95\%$ -CI! But: we can't sell (6.5) as a CI. Let's do some algebra (or jump to (6.10) if you don't like maths, but give it a try):

$$P\left[-1.96\sqrt{\frac{\sigma^2}{n}} \leq \bar{X} - \mu \leq 1.96\sqrt{\frac{\sigma^2}{n}}\right] = 95\%. \quad (6.6)$$

Swop to absolute values:

$$P\left[\left|\bar{X} - \mu\right| \leq 1.96\sqrt{\frac{\sigma^2}{n}}\right] = 95\%, \quad (6.7)$$

write it more elegantly as

$$P\left[\left|\bar{X} - \mu\right| \leq \frac{1.96\sigma}{\sqrt{n}}\right] = 95\% \quad (6.8)$$

and finally

$$P\left[\mu \in \left[\bar{X} - \frac{1.96\sigma}{\sqrt{n}}, \bar{X} + \frac{1.96\sigma}{\sqrt{n}}\right]\right] = 95\%. \quad (6.9)$$

Compare (6.9) with Definition 6.2.

Therefore: 95 % CI for μ with known σ^2 and data x_1, \dots, x_n in model $\mathcal{N}(\mu, \sigma^2)$ is:

$$\left[\bar{x} - \frac{1.96\sigma}{\sqrt{n}}, \bar{x} + \frac{1.96\sigma}{\sqrt{n}} \right]. \quad (6.10)$$

Small exercises:

Give us the 99 % CI in this situation:

Observe our holy \sqrt{n} once more in (6.10)! What happens as n gets larger?

What happens as σ gets larger?

Observe that last two features make sense! Solve exercise 6.1.

6.2.2.2 $\mathcal{N}(\mu, \sigma^2)$: CI for μ with σ^2 unknown to us (usual case)

Given data x_1, \dots, x_n , compute arithmetic mean \bar{x} . Switch to RV's: look at \bar{X} . Distribution of \bar{X} is:

mean=

variance=

distribution type is:

together:

So

$$\frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}}$$

is a $\mathcal{N}(0, 1)$ due to the “Z-Transform” (see 4.2.2). Oooooops? We don't know σ^2 ! Don't give up: We're gonna estimate it:

$$\hat{\sigma}^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

The animal

$$\frac{\bar{X} - \mu}{\sqrt{\frac{\hat{\sigma}^2}{n}}} = \frac{\bar{X} - \mu}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}} \quad (6.11)$$

has a t_{n-1} -distribution (see 4.2.5, but be careful when comparing that T_n with our expression (6.11): it is the same if you use some algebra and some probability theory - we omit it here)!

Following the same arguments as in 6.2.2.1, analog to (6.10) is

$$\left[\bar{x} - \frac{CV \hat{\sigma}}{\sqrt{n}}, \bar{x} + \frac{CV \hat{\sigma}}{\sqrt{n}} \right], \quad (6.12)$$

where

$$\hat{\sigma} := \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}, \quad (6.13)$$

and Critical Value CV depends not only on our $(1 - \alpha)$ as in 6.2.2.1, but also on the n (see tables). For example if we choose $n = 20$, then the critical values for a 95%-CI are:

instead of the 1.96.

Let us solve exercise 6.1 once more, assuming we did not know σ and estimated it with (6.13) to be 1.12 mm. Compute a 0.95-CI again in this situation.

Observe our holy \sqrt{n} once more in (6.12)! What happens as n gets larger (2 things)?

What happens as σ gets larger?

Solve some exercises from 6.3.

Small miracle for people who like maths (proof is hard): in $\mathcal{N}(\mu, \sigma^2)$

$$\bar{X} \quad \text{and} \quad \sum_{i=1}^n (X_i - \bar{X})^2$$

are independent of each other!

6.2.2.3 $\text{Be}(p)$ and $\text{Bin}(n, p)$ respectively: CI for p

We repeat from 6.1.5: data x_1, x_2, \dots, x_n are "0"'s (Death, Failure) or "1"'s (Survival, Success) in a situation where you count the number of successes, then the number of successes is simply

$$\sum_{i=1}^n x_i$$

and has by definition a Binomial distribution with parameters p, n : $\text{Bin}(n, p)$. Canonical choice for \hat{p} is

$$\hat{p} := \frac{1}{n} \sum_{i=1}^n x_i$$

Variance of this estimator:

$$\text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{p(1-p)}{n}.$$

The standard error of the estimate of the p (sampling error) is therefore

$$\sqrt{\frac{p(1-p)}{n}}.$$

Best choice for 95 % CI is then (for n large)

$$\left[\hat{p} \pm 1.96 \sqrt{\hat{p}(1-\hat{p})/n} \right] = \left[\hat{p} - 1.96 \sqrt{\hat{p}(1-\hat{p})/n} , \hat{p} + 1.96 \sqrt{\hat{p}(1-\hat{p})/n} \right].$$

Elections in the U.S.: Suppose $n = 10'000$ (people interviewed) and true (but unknown values) are $p = 0.52$ for Bush and $(1-p) = 0.48$ for Kerry. What can we now say if 5'058 of those 10'000 said they are going to vote for Bush?

Now check whether example at beginning of 6.2.1 was correct. PS: (theoretically) needs an infinite population - but 10'000 out of entire U.S. population is so small that it's OK.

6.3 Exercises

σ^2 known

6.1 The diameter of steel balls produced is modelled with a normal distribution with $\sigma = 1.04$ mm. We take a sample of $n = 30$ and get the arithmetic mean of $\bar{x} = 12.14$ mm. Compute 0.95-CI and 0.99-CI for the mean diameter of these balls.

σ^2 unknown

6.2 Solve 6.1 (only 95 % case) with unknown σ having been estimated (by chance) completely correct as 1.04 mm with (6.13). Compare results from 6.1 and 6.2.

6.3 Researcher tries to grow nerves with some method. Assume length of growth can be modelled with a normal distribution. Using (6.13), σ has been estimated as $0.03 \mu m$. Sample size is $n = 51$, arithmetic mean is $\bar{x} = 0.13 \mu m$. Compute 0.95-CI for the mean growth of nerves with his/her method.