Inferential Statistics
L6 - Confidence Sets

Erlis Ruli (erlis.ruli@unipd.it)

Department of Statistics, University of Padova

Contents

Recall this problem statement?

Suppose that the average energy consumption of our population of WMs, mounting a standard motor, is μ_0.

It's claimed that NG1 family motors would lead to more efficient WMs, i.e. would lead to average consumption μ, s.t. $\mu < \mu_0$.

There are two possibilities:

- the claim is false, so $\mu_0 \leq \mu$; this is called Null Hypothesis ("null" because it adds nothing to the current state of art)
- the claim is true, so $\mu < \mu_0$; it's called Alternative Hypothesis.

Problem statement

In L5 we equipped 10 WM's with the NG1 motor and measured their E consumption getting 19.1, 20.6, 17.3, 21.1, 19.5, 19.5, 21.4, 19.1, 20.5, 19.5.

$\bar{x} = 19.76$ is a good point estimate (L4) for the population E consumption μ. However, it's very unlikely that this estimates equals μ. Indeed,

$$P_\mu(\overline{X} = \mu) = 0.$$

Sometimes, however, it is desired to produce a set or an interval estimate, that includes μ with a pre-specified probability.

This set typically has infinite values, i.e. infinite estimates of μ, so it's less informative than the point estimate; however, the reward is that we have some guarantee that our assertion is correct.

Definition

An <u>interval estimate</u> of a scalar parameter θ is any pair of functions $L(\mathbf{x}), U(\mathbf{x})$ of a sample $\mathbf{x} = (x_1, \ldots, x_n)$ s.t. $L(\mathbf{x}) \leq U(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{X}$.

At the observed sample it is inferred that $L(\mathbf{x}) \leq \theta \leq U(\mathbf{x})$.

The random interval $[L(\mathbf{X}), U(\mathbf{X})]$ based on the random sample $\mathbf{X} = (X_1, \ldots, X_n)$ is called an <u>interval estimator</u>.

Interval estimators could also be lower or upper intervals, e.g. $(-\infty, U(\mathbf{X}))$ or $(L(\mathbf{X}), \infty)$, respectively.

Example 1

For an iid random sample X_1, X_2, X_3, X_4 from a $N(\mu, 1)$, we know that $\overline{X} \sim N(\mu, 1/4)$. Thus $[\overline{X} - 1, \overline{X} + 1]$ is an interval estimator for μ.

In L4 we saw that $\overline{X}$ was a good estimator for μ. Why on earth would we want the less precise estimator $\overline{X} \pm 1$?

The answer is that now we have a positive probability ($\approx .95$) that the interval contains the (unknown) parameter μ.

Definitions

For an interval estimator $[L(\mathbf{X}), U(\mathbf{X})]$ of a parameter θ, we define the coverage probability by

l'intervallo è la parte casuale

$$P_\theta(\theta \in [L(\mathbf{X}), U(\mathbf{X})]),$$

i.e. the probability that the random interval $[L(\mathbf{X}), U(\mathbf{X})]$ traps θ.

The smallest coverage probability among all θ, i.e.

$$\inf_{\theta \in \Theta} P_\theta(\theta \in [L(\mathbf{X}), U(\mathbf{X})]),$$

is called the confidence level.

An interval estimator $[L(\mathbf{X}), U(\mathbf{X})]$ with confidence level $1-\alpha$, (with $\alpha \in (0,1)$) is called confidence interval of level $1-\alpha$.

Method of inverting a test statistic

For a two tailed confidence interval, e.g. $[L(\mathbf{x}), U(\mathbf{x})]$, the method for constructing a $1-\alpha$-level confidence set consists in the following three steps:

(1) get R, the rejection region for $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$;

(2) get the acceptance region xR^c

(3) invert the acceptance region

Upper or lower confidence intervals can be built similarly; the shape of the rejection region determines the shape of the confidence interval.

Example 2 (See Example 9, L5)

Consider $X_1, \dots, X_n$ an iid random sample with $X_i \sim N(\mu, \sigma^2)$, σ^2 is known and $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$. The rejection region is

$$R = \{\mathbf{x} : |\overline{x} - \mu_0| > z_{1-\alpha/2}\sigma/\sqrt{n}\},$$

$$R^c = \{\mathbf{x} : |\bar{x} - \mu_0| < z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\}$$

so H_0 is accepted if $\mathbf{x} \in R^c$, or equivalently if

$$\overline{x} - z_{1-\alpha/2}\sigma/\sqrt{n} \leq \mu_0 \leq \overline{x} + z_{1-\alpha/2}\sigma/\sqrt{n}.$$

But,

probability of incorrectly rejecting H_0

$$P_{\mu_0}(X \in R) = \alpha \Longrightarrow$$

$$P_{\mu_0}(\mathbf{X} \in R^c) = P_{\mu_0}\left(\mu_0 \in [\overline{X} \pm z_{1-\alpha/2}\sigma/\sqrt{n}]\right)$$
$$= 1 - \alpha, \quad \forall \mu_0,$$

so $[\overline{X} \pm z_{1-\alpha/2}\sigma/\sqrt{n}]$ is a $1 - \alpha$ confidence interval for μ.

Example 2 (cont'd)

Suppose the observed sample is (as in L5)

19.1, 20.6, 17.3, 21.1, 19.5, 19.5, 21.4, 19.1, 20.5, 19.5

and let $\sigma^2 = 5$. Then $\bar{x} = 19.76$ and the 0.95 confidence interval for μ is

$$\left[19.76 - 1.96 \cdot \sqrt{\frac{5}{10}}, 19.76 + 1.96 \cdot \sqrt{\frac{5}{10}}\right] = [18.37, 21.14]$$

Caution!

[18.37, 21.14] is an observed interval and it's not correct to say "this interval contains the true mean μ_0 with probability 0.95". Indeed, μ_0 either is or is not inside this interval. We can only say that we are 0.95 confident that the interval contains μ_0.

Two sides of the same coin

Confidence sets of level $1 - \alpha$ are thus derived by inverting a given test of size (or level) α:

(i) Wald-type confidence sets are derived by inverting Wald tests

(ii) likelihood-based confidence sets are obtained inverting an LRT.

Inverting a Wald test

Let $R = \{\mathbf{X} : |\widehat{\theta} - \theta_0|/\widehat{se}) > z_{1-\alpha/2}\}$ be the rejection region of a Wald test of (approx.) size α for

$$H_0 : \theta = \theta_0 \text{ against} H_1 : \theta \neq \theta_0.$$

Then, the corresponding Wald-type confidence interval of (approx.)

because asymptotic

confidence level $1 - \alpha$ is

$$[\widehat{\theta} - z_{1-\alpha/2}\widehat{se}, \widehat{\theta} + z_{1-\alpha/2}\widehat{se}].$$

This immediately generalizes when θ is a vector and we are interested in a single component, say θ_i.

Wald test always gives
symmetrical confidence intervals

Inverting a LRT

For a scalar parameter θ, let again

$$H_0 : \theta = \theta_0 \text{ against } H_1 : \theta \neq \theta_0,$$

and for a fixed θ_0, consider the rejection region of size α of the LRT

$$R_\alpha(\theta_0) = \{\mathbf{X} : -2 \log \frac{L(\theta_0)}{L(\hat{\theta})} > \chi^2_{1,1-\alpha}\}$$

[theorem 5 says that $-2\log \lambda(Y) \sim \chi^2_{r-q}$

The likelihood-based confidence set of level $1 - \alpha$ is given by

$$\{\theta : \theta \in R_\alpha(\theta)^c\},$$

holding the data $\mathbf{X}$ fixed.

Here is an example.

Example 3

Let $X_1, \ldots, X_n$ be an iid random sample with $X_i \sim \text{Poi}(\theta)$, with θ unknown. Furthermore, let

$$x_1 = 0, \quad x_2 = 0, \quad x_3 = 3, \quad x_4 = 5, \quad x_5 = 7,$$

be an observed sample. The MLE is $\widehat{\theta} = 3$ and $-2\log$ of LRT statistic is

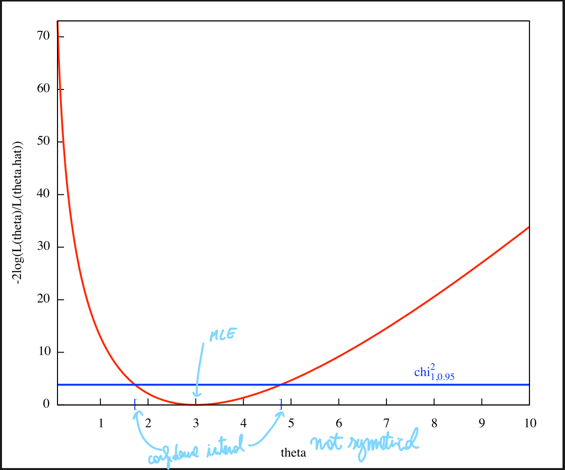
$$-2\log(L(\theta)/L(\widehat{\theta})) = -10(3-\theta) - 30\log(\theta/3).$$

The likelihood-based confidence set of level $1-\alpha$ is thus the set

$$\{\theta : -10(3-\theta) - 30\log(\theta/3) < \chi^2_{1,1-\alpha}\}.$$

Example 3 (cont'd): A déjà vu ?

The confidence interval in question is the set of values for θ that lie between the points of intersection of the two curves, here $[1.72, 4.78]$



Comments

A confidence set computed at an observed sample is a set of numbers and, in this case, the true parameter value either is or isn't inside the set.

For instance, in Example 3, if the true parameter happened to be $\theta = 1$, the probability that this 0.95 confidence set includes θ is 0; if $\theta = 2$, prob=1.

In practice, we'll never know θ, so we can only be 95% confident that the confidence set includes θ.

By "95% confident" we mean:

If we could collect a large number of samples, all of size n, and for each of them compute a 0.95 confidence set, then we expect that exactly 95% of these sets will include the true parameter value.

Choosing between confidence sets

By definition, a confidence region must cover the true parameter value with probability of at least $1 - \alpha$.

In practice, however, the test used to compute it is asymptotically of size α. Thus, for finite n the coverage may not be as desired.

Furthermore, the larger the confidence set the less informative it is.

We prefer confidence set that have:

(i) coverage probability as close as possible to $1 - \alpha$

(ii) length (or volume) as small as possible; applies only to bounded confidence set.
(smaller confidence sets are more informative)

Example 4 t confidence interval for the mean

t confidence interval Let $X_1, \ldots, X_n$ be an iid random sample from $N(\mu, \sigma^2)$, with both parameters unknown.

A confidence interval for μ of level $1 - \alpha$ can obtained by inverting the LRT test (see Example 12, L5). Indeed, given

$$R_\alpha(\mu_0) = \left\{ \mathbf{X} : \left| \frac{\sqrt{n}(\overline{X} - \mu_0)}{S} \right| > t_{n-1,1-\alpha/2} \right\},$$

the confidence set for fixed **X** is

$$\left\{ \mu : \overline{X} - t_{n-1,1-\alpha/2} \frac{S}{\sqrt{n}} < \mu < \overline{X} + t_{n-1,1-\alpha/2} \frac{S}{\sqrt{n}} \right\}.$$

in sto caso e' simmetrico anche se e' un LRT, perche' sono rv Normali

Example 5

Confidence interval for the variance Let $X_1, \ldots, X_n$ be an iid random sample from $N(\mu, \sigma^2)$, with both parameters unknown.

A confidence interval for σ^2 of level $1-\alpha$ can obtained by inverting the LRT test (see Example 13, L5). Indeed, given

version with a1 and a2 use the quantiles not the optimization

$$R_\alpha(\sigma_0^2) = \left\{ \mathbf{X} : \frac{n\hat{\sigma}^2}{\sigma_0^2} < \chi^2_{n-1,\frac{\alpha}{2}} \text{ or } \frac{n\hat{\sigma}^2}{\sigma_0^2} > \chi^2_{n-1,1-\frac{\alpha}{2}} \right\},$$

the confidence set for fixed $\mathbf{X}$ is $R^C = \left\{ \mathbf{X} : \chi^2_{n-1,\frac{\alpha}{2}} \leq \frac{n\hat{\sigma}^2}{\sigma^2} \leq \chi^2_{n-1,1-\frac{\alpha}{2}} \right\}$

Solve for σ^2

$$\left\{ \sigma^2 : \chi^2_{n-1,\alpha/2} < \frac{n\hat{\sigma}^2}{\sigma^2} < \chi^2_{n-1,1-\frac{\alpha}{2}} \right\} =$$

$$= \left\{ \sigma^2 : \frac{n\hat{\sigma}^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}} < \sigma^2 < \frac{n\hat{\sigma}^2}{\chi^2_{n-1,\frac{\alpha}{2}}} \right\}$$

Example 6

t confidence interval for the difference of means Let $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ are two iid random samples with $X_i \sim N(\mu_x, \sigma_x^2)$, $Y_j \sim N(\mu_y, \sigma_y^2)$ and X_i is independent from Y_j, all parameters unknown.

Assuming, $\sigma_x^2 = \sigma_y^2$, a confidence interval for $\mu_x - \mu_y$ is obtained by inverting the LRT (Example 14, L5). The confidence interval is

$$\mu_x - \mu_y \in \left[\overline{X} - \overline{Y} \pm t_{n+m-2,1-\alpha/2}\sqrt{S_p^2\left(\tfrac{1}{m} + \tfrac{1}{n}\right)}\right].$$

If $\sigma_x^2 \neq \sigma_y^2$, the $1-\alpha$ confidence interval becomes

$$\mu_x - \mu_y \in \left[\overline{X} - \overline{Y} \pm t_{\nu,1-\alpha/2}\sqrt{S_x^2/m + S_y^2/n}\right].$$