Inferential Statistics **First semester, 2021**

Lecture 5: Confidence sets

Instructor: Erlis Ruli (ruli@stat.unipd.it), Department of Statistical Sciences

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

Let us start again with a problem inspired by a real-life application.

For the problem of measuring energy consumption of WM's under the cotton 40°C washing program, by means of an estimator we can get an estimate for the mean of the population distribution. For instance, if we are willing to assume a parametric model, we can use $\widehat{\mu}$, the MLE of μ. However, announcing a single value as the estimate of a parameter calls for some auxiliary indication of how much that value can be trusted – a measure of reliability. Thus, we need to ask an additional question: how much variable is the estimate $\widehat{\mu}$? The notion of standard error we introduced in Lecture 4, is a commonly used measure, and sometimes estimates are given as `estimate` $\pm$ `3se`. The standard error is not a maximum error, of course, so one could not expect the true value of the parameter to be within three standard errors of the estimate in every instance. Sometimes it will and sometimes it won't.

The essence of a confidence interval is thus to produce a lower limit and an upper limit which contain the true value θ_0 with a pre-specified probability. These lower and upper limit constitute a confidence interval and are the topic of the present lecture. An alternative name for confidence interval is <u>interval estimation</u>, since the aim of confidence interval theory is to <u>estimate</u> lower and upper limits for the true parameter value θ_0.

First we give some definitions.

Definition 5.1 *A random interval is a finite or infinite interval with at least one endpoint being a r.v.*

Definition 5.2 *Given a random sample $Y_1, \dots, Y_n$ from some distribution F_θ, let $L_n = L(Y_1, \dots, Y_n)$ and $U_n = U(Y_1, \dots, Y_n)$ be two statistics such that $L_n \leq U_n$. The random* *interval* $[L_n, U_n]$ *is a confidence interval for* θ *with confidence level* $1-\alpha$, $0 < \alpha < 1$ *if*

$$P_\theta(L_n \leq \theta \leq U_n) \geq 1-\alpha, \quad \textit{for all } \theta \in \Theta.$$

The notation $P_\theta(\cdot)$ *is to remind us that the probability has to be computed with respect to the distribution of the sample and under the true parameter* θ*. Also we say that* $[U_n, \infty)$ *and* $(-\infty, L_n]$ *is an upper and a lower confidence limit for* θ*, respectively, with confidence level* $1-\alpha$*, if for all* $\theta \in \Theta$

$$P_\theta(-\infty < \theta \leq U_n) \geq 1-\alpha$$

and

$$P_\theta(L_n \leq \theta < \infty) \geq 1-\alpha.$$

Thus $[L_n, U_n]$ traps θ with probability at least $1-\alpha$, no matter what $\theta \in \Theta$. We call $P_\theta(L_n \leq \theta \leq U_n)$ the *coverage* of the confidence interval and $1-\alpha$ the *confidence level*. In other words, the confidence level is the smallest coverage probability over all possible parameter values θ.

Be careful! $[L_n, U_n]$ is a random interval whereas θ is fixed.

Commonly, people use 95 or 99 percent confidence intervals, which correspond to choosing $\alpha = .05$ or $\alpha = .01$, respectively. If θ is a vector then we use a *confidence set*, such as an ellipse or an ellipsoid, instead of an interval.

5.1 Properties of confidence intervals

5.1.1 Expected length

It is quite possible that there exist more than one confidence interval for θ with the same confidence level $1-\alpha$. In such a case, it is obvious that we would be interested in finding the shortest confidence interval on average and within a certain class of confidence intervals.

More formally, we define the *length* of the random confidence interval $[L_n, U_n]$ by

$$D_n = D(Y_1, \ldots, Y_n) = U_n - L_n,$$

and the expected length is defined by $\Delta = E(D_n)$. Thus the length of a random confidence interval, if it exists, is also a statistic. The *length criterion* then says that:

> Among all possible confidence intervals considered, all having the same confidence level, we should prefer the shortest on average.

This is because, the shorter is the confidence interval, the more precise is the inference about the true parameter θ.

5.2 Methods for confidence intervals

The general procedure for constructing confidence intervals is as follows. We start out with a r.v. $T_n(\theta) = T(Y_1, \ldots, Y_n; \theta)$ which depends on θ and on Y_i's only through a hopefully sufficient statistic, and whose distribution is completely determined. By "completely determined" we mean that the distribution of $T_n(\theta)$ does not depend on θ. Then L_n and U_n are some simple functions of $T_n(\theta)$.

A r.v. $T_n(\theta)$ which has a completely determined distribution is called *pivot* or *pivotal quantity*. If the r.v. $T_n(\theta)$ has a completely determined distribution in the limit as $n \to \infty$, then $T_n(\theta)$ is an *asymptotic pivot*.

Given a set B such that $P_\theta(T_n(\theta) \in B) \geq 1 - \alpha$, for all $\theta \in \Theta$, then the set

$$\{\theta \in \Theta : T_n(\theta) \in B\}$$

is a confidence region for θ of confidence level $1 - \alpha$. In many practical cases, this region is actually an interval. The essence of finding a confidence interval for θ is thus to find a set B which satisfies the above inequality for all θ.

Thus confidence intervals are constructed from a given pivotal quantity. Furthermore, the result below shows that when $g(\theta)$ is a strictly monotone function, a confidence interval for $g(\theta)$ can be readily derived from a confidence interval for θ.

Theorem 5.1 *If $[L_n, U_n]$ is a $(1 - \alpha)$ confidence interval for θ and $g(\theta)$ is a monotone increasing function, then* *(i)* $[g(L_n), g(U_n)]$ *is a* $(1-\alpha)$ *confidence interval for* $g(\theta)$ *whenever* $g(\cdot)$ *is an increasing function.*

(ii) $[g(U_n), g(L_n)]$ *is a* $(1-\alpha)$ *confidence interval for* $g(\theta)$ *whenever* $g(\cdot)$ *is a decreasing function.*

The following examples help illustrating the point.

5.2.1 Exact pivots

Example 5.1 *Let* $X_i \overset{\text{iid}}{\sim} \sim \text{Exp}(1/\theta)$ *be a random sample of size* n*. We want to build a* $1-\alpha$ *level confidence interval for* θ*. To build a pivot for this problem first note that the MLE of* θ *is* $\widehat{\theta} = \overline{X}$*. Furthermore,*

$$\begin{aligned} n\overline{X} &= \sum_{i=1}^{n} X_i \sim \text{Ga}(n, 1/\theta) \iff 2n\overline{X} \sim \text{Ga}(n, 1/2\theta) \\ &\qquad\qquad\qquad\qquad\qquad\quad \iff \tfrac{2n\overline{X}}{\theta} \sim \text{Ga}(n, 1/2). \end{aligned}$$

But $\text{Ga}(n, 1/2) = \text{Ga}(2n/2, 1/2) = \chi^2_{2n}$ *(see L1), thus the quantity* $2n\overline{X}/\theta$ *is a pivot with distribution* χ^2_{2n}*. Note that* $2n\overline{X}$ *is not a pivot because its distribution depends on* θ*, which is unknown.*

Let $0 < c_1 \leq c_2$ *be two constants such that*

$$P(\chi^2_{2n} \in [c_1, c_2]) = 1 - \alpha.$$

There are infinitely many possible choices for $c_1 \leq c_2$ *that satisfy this inequality, but two of them are typically used. These are shown in Figure 5.1. In case (i), we restrict* c_1, c_2 *to have highest density. In this case* c_1 *and* c_2 *are found by cutting the density at the point which leads to* c_1 *and* c_2 *such that the area in the tails sums to* α*. In case (ii) we split* α *in two equal parts and choose* c_1 *and* c_2 *such that* $P(\chi^2_{2n} \leq c_1) = P(\chi^2_{2n} \geq c_2) = \alpha/2$*.*

Option (i) typically leads to shorter intervals. Nevertheless, most practitioners prefer the equi-tailed option (ii) since it is much simpler to implement. Indeed, here $c_1 = \chi^2_{2n,1-\alpha/2}$ *and* $c_2 = \chi^2_{2n,\alpha/2}$*, where* $\chi^2_{2n,\alpha}$ *is the upper* α*th quantile of the* χ^2_{2n} *distribution. In the case of the*

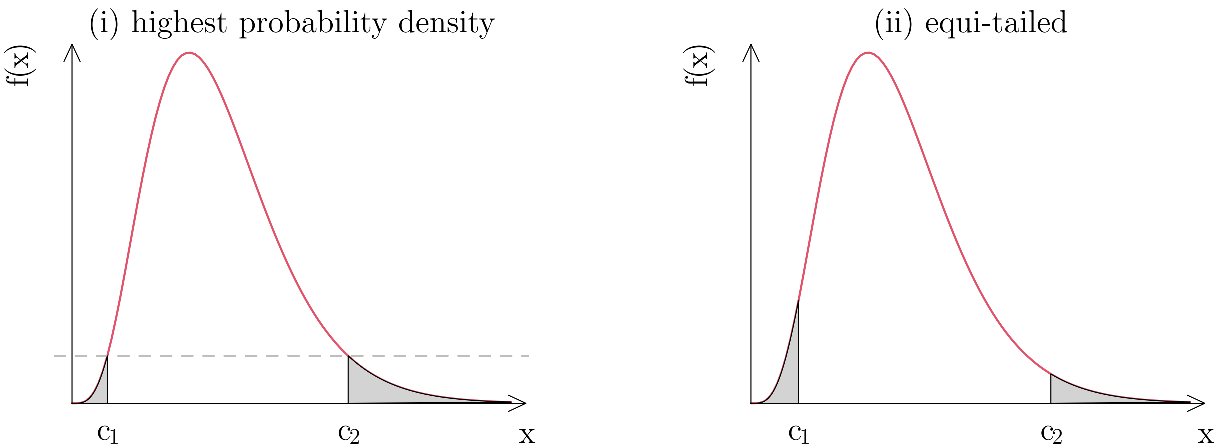

Figure 5.1: Two different approaches for defining thresholds c_1, c_2 which leave in the tails an overall probability equal to α. In (i) the thresholds have highest density but lead to tails of different probabilities. In case (ii), thresholds have different density values but they lead to equal-probability tails, i.e. tails of probability $\alpha/2$.

equi-tailed thresholds we have

$$\begin{aligned}
1-\alpha &= P(\chi^2_{2n,1-\alpha/2} \le \chi^2_{2n} \le \chi^2_{2n,\alpha/2}) \\
&= P_\theta\left(\chi^2_{2n,1-\alpha/2} \le \tfrac{2n\overline{X}}{\theta} \le \chi^2_{2n,\alpha/2}\right) \\
&= P_\theta\left(\frac{1}{\chi^2_{2n,\alpha/2}} \le \frac{\theta}{2n\overline{X}} \le \frac{1}{\chi^2_{2n,1-\alpha/2}}\right) \\
&= P_\theta\left(\frac{2n\overline{X}}{\chi^2_{2n,\alpha/2}} \le \theta \le \frac{2n\overline{X}}{\chi^2_{2n,1-\alpha/2}}\right).
\end{aligned}$$

So $\left[\frac{2n\overline{X}}{\chi^2_{2n,\alpha/2}}, \frac{2n\overline{X}}{\chi^2_{2n,1-\alpha/2}}\right]$ *is confidence interval for* θ *with confidence level* $1-\alpha$.

As a numerical example, suppose X_i*'s are time in seconds taken by a HPC server to reboot in* $n = 10$ *reboots, and let the observed sample be (48.386, 65.418, 26.510, 15.830, 22.381, 21.882, 30.121, 18.714, 10.874, 1.759). Then* $\overline{x} = 26.1875$. *With* $\alpha = .10$ *we have that* $\chi^2_{20,.05} = 31.41$ *and* $\chi^2_{20,.95} = 10.85$. *Thus the 90 percent confidence interval for* θ *is* $[16.67, 48.27]$.

We interpret this result by saying that we are 90 percent confident that the interval $[16.67, 48.27]$ *will contain the true parameter value* θ*. "90 percent confident" means the following. If we could draw a large number of samples of size* $n = 10$ *from the same distribution and compute, for each of them, a 0.90 confidence interval, then we expect that 90 percent of these intervals will contain the true parameter value.*

Be careful! There is much confusion about how to interpret a confidence interval. Given the observed data, a confidence interval is not a probability statement about θ since the latter is a fixed quantity (i.e. chosen by Nature and never revealed to us), not a random variable. A 95 percent confidence interval is interpreted as follows: if we repeat the experiment over and over again under the same conditions, only 95 percent of the confidence intervals obtained in each sample will actually contain the true parameter θ.

Example 5.2 *Let* $Y_i \overset{\text{iid}}{\sim} \mathrm{N}(\mu, \sigma^2)$ *be a random sample of size* n*. First, suppose that* σ *is known, so that* μ *is the only unknown parameter. Consider the r.v.* $\sqrt{n}(\overline{Y} - \mu)/\sigma$*. This depends on the sample only through the sufficient statistic* $\overline{Y}$ *of* μ *and its distribution is the same as the r.v.* $Z \sim N(0, 1)$*, for all* μ*.*

define the MLE of sigma^2

Next determine two numbers $c_1 \le c_2$ *such that*

$$\begin{aligned} 1 - \alpha &= P(c_1 \le Z \le c_2) \\ &= P_\mu \left(c_1 \le \tfrac{\sqrt{n}(\overline{Y} - \mu)}{\sigma} \le c_2 \right) \\ &= P_\mu \left(\overline{Y} - c_2 \tfrac{\sigma}{\sqrt{n}} \le \mu \le \overline{Y} - c_1 \tfrac{\sigma}{\sqrt{n}} \right), \end{aligned}$$

is a confidence interval for μ *with coverage* $1 - \alpha$*. Its length is equal to* $(c_2 - c_1)\sigma/\sqrt{n}$ *which is also equal to the expected length. Form this it follows that, among all confidence intervals with coverage* $1 - \alpha$ *which have the above form, the shortest one is that for which* $c_2 - c_1$ *is smallest. This happens if* $c_2 = z_{\alpha/2}$ *and* $c_1 = -c_2$*,* $z_{\alpha/2}$ *is the upper* $(\alpha/2)$*th quantile of the* $\mathrm{N}(0, 1)$ *distribution. Therefore, the shortest confidence interval for* μ *with coverage* $1 - \alpha$ *(and which is of the form above) is given by*

$$\left[\overline{Y} - z_{\alpha/2} \tfrac{\sigma}{\sqrt{n}}, \overline{Y} + z_{\alpha/2} \tfrac{\sigma}{\sqrt{n}} \right].$$

Example 5.3 *Let* $Y_i \overset{\text{iid}}{\sim} \mathrm{N}(\mu, \sigma^2)$ *be a random sample of size* n*. Now suppose that* μ *is* *known, so that σ^2 is the unknown parameter and consider the r.v.*

$$\frac{n\widehat{\sigma^2}_\mu}{\sigma^2}, \quad \textit{where} \quad \widehat{\sigma^2}_\mu = \frac{1}{n}\sum_{i=1}^{n}(Y_i-\mu)^2.$$

This r.v. depends on the sample only through the sufficient statistic $\widehat{\sigma^2_\mu}$ for σ^2 and its distribution is χ^2_n for all σ^2.

Now determine two numbers c_1 and c_2, $0 < c_1 \leq c_2$ such that

$$\begin{aligned} 1-\alpha &= P(c_1 \leq \chi^2_n < c_2) \\ &= P_{\sigma^2}\left(c_1 \leq \frac{n\widehat{\sigma^2}_\mu}{\sigma^2} \leq c_2\right) \\ &= P_{\sigma^2}\left(\frac{n\widehat{\sigma^2}_\mu}{c_2} \leq \sigma^2 \leq \frac{n\widehat{\sigma^2}_\mu}{c_1}\right), \end{aligned}$$

Thus

$$\left[\frac{n\widehat{\sigma^2}_\mu}{c_2}, \frac{n\widehat{\sigma^2}_\mu}{c_1}\right]$$

is a confidence interval for σ^2 with coverage $(1-\alpha)$ and length equal to $(1/c_1 - 1/c_2)n\widehat{\sigma^2}_\mu$. The expected length is equal to $(1/a-1/b)n\sigma^2$. As in Example 5.1, it is possible to determine
not a,b c1, c2
c_1 and c_2 such that the resulting confidence interval of the form above has the shortest length. This can be achieved by thresholds which have highest density (option (i) in Figure 5.1).

However, in practice c_1 and c_2 are often chosen by assigning probability $\alpha/2$ to both tails of the χ^2_n distribution. The resulting equi-tailed confidence interval is then

$$\left[\frac{n\widehat{\sigma^2}_\mu}{\chi^2_{n,\alpha/2}}, \frac{n\widehat{\sigma^2}_\mu}{\chi_{n,1-\alpha/2}}\right].$$

Note that this is not the best choice because the corresponding interval is not the shortest one.

Furthermore, thanks to Theorem 5.1, a $(1-\alpha)$ confidence interval for $\sigma = \sqrt{\sigma^2}$ is

$$\left[\sqrt{\frac{n\widehat{\sigma^2}_\mu}{\chi^2_{n,\alpha/2}}}, \sqrt{\frac{n\widehat{\sigma^2}_\mu}{\chi_{n,1-\alpha/2}}}\right],$$

Example 5.4 *Let* $Y_i \overset{\text{iid}}{\sim} \mathrm{N}(\mu, \sigma^2)$ *be a random sample of size* n*, where both* μ *and* σ^2 *are unknown. We consider confidence intervals for* μ *and* σ^2 *separately. Consider the two pivots*

$$\tfrac{\sqrt{n}(\overline{Y}-\mu)}{\sqrt{S^2}} \sim t_{n-1}, \qquad \tfrac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1};$$

see Lecture 4. Thus a confidence interval of level $1-\alpha$ *for* μ *and a confidence interval of level* $1-\alpha$ *for* σ^2 *are*

$$\left[\overline{Y} - t_{n-1,\alpha/2}\sqrt{\tfrac{S^2}{n}}, \overline{Y} + t_{n-1,\alpha/2}\sqrt{\tfrac{S^2}{n}}\right]$$

and

$$\left[\tfrac{(n-1)S^2}{\chi^2_{n-1,\alpha/2}}, \tfrac{(n-1)S^2}{\chi_{n-1,1-\alpha/2}}\right],$$

respectively; here $t_{n-1,\alpha}$ *denotes the upper* α*th quantile of the* t*-Student distribution with* $n-1$ *degrees of freedom.*

Here are two two-sample problems. In the first one we wish to make inference about the difference between the means of the two normal populations, whereas in the second we wish to make inference about the ratio of their variances.

Example 5.5 *(Example 2.5 revised) Let* $Y_i \overset{\text{iid}}{\sim} N(\mu_1, \sigma_1^2)$ *be a random sample of size* n *and* $X_j \overset{\text{iid}}{\sim} N(\mu_2, \sigma_2^2)$*, be an i.i.d. random sample of size* m *where we assume* $\sigma_1^2 = \sigma_2^2 = \sigma^2$ *and* μ_1, μ_2, σ^2 *are unknown. We wish to build a confidence interval for* $\mu_1 - \mu_2$*. Consider the r.v.*

$$\tfrac{(\overline{Y}-\overline{X})-(\mu_1-\mu_2)}{\sqrt{S^2_{pool}\left(\frac{1}{m}+\frac{1}{n}\right)}}, \tag{5.1}$$

where

$$S^2_{pool} = \tfrac{(n-1)S_Y^2+(m-1)S_X^2}{n+m-2}$$

is sometimes referred to as the pooled variance estimate. It can be shown that the r.v. in (5.1) *has distribution* t_{n+m-2}*, thus it is a pivotal quantity. A confidence interval of level*

$1-\alpha$ *for* $\mu_1 - \mu_2$ *is then*

$$\left[(\overline{Y}-\overline{X}) - t_{n+m-2,\alpha/2}\sqrt{S^2_{pool}\left(\tfrac{1}{m}+\tfrac{1}{n}\right)}, (\overline{Y}-\overline{X}) + t_{n+m-2,\alpha/2}\sqrt{S^2_{pool}\left(\tfrac{1}{m}+\tfrac{1}{n}\right)}\right].$$

Example 5.6 *Let* $Y_i \overset{\text{iid}}{\sim} N(\mu_1, \sigma_1^2)$ *be a random sample of size* n *and* $X_j \overset{\text{iid}}{\sim} N(\mu_2, \sigma_2^2)$, *be an i.i.d. random sample of size* m *and* $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$ *are all unknown. We wish to build a confidence interval for* σ_2^2/σ_1^2. *The r.v.*

$$\frac{S_Y^2/\sigma_1^2}{S_X^2/\sigma_2} \sim F_{n-1,m-1}$$

provides a pivotal quantity for σ_1^2/σ_2^2. *Thus a* $(1-\alpha)$ *confidence interval for* σ_2^2/σ_1^2 *can be obtained by setting the quantiles for the* F *distribution with degrees of freedom* $n-1$ *and* $m-1$ *such that*

$$1-\alpha \;=\; P_\theta\left(F_{n-1,m-1,1-\alpha/2} \le \frac{S_Y^2/\sigma_1^2}{S_X^2/\sigma_2^2} \le F_{n-1,m-1,\alpha/2}\right)$$

and thus the confidence interval is

$$\left[\frac{S_X^2}{S_Y^2}F_{n-1,m-1,1-\alpha/2}, \frac{S_X^2}{S_Y^2}F_{n-1,m-1,\alpha/2}\right],$$

where $F_{n,m,\alpha}$ *denotes the upper* α*th quantile of the* $F_{n,m}$ *distribution.*

Example 5.7 *(Paired samples). In all examples above we assumed independent samples. In some cases, such as test-retest experiments, dependent samples are appropriate. For example, to measure the effect of new type of motor for WM and compare it with the old motor, we would select* n *WM's at random from the production line and measure their energy consumption both with the old motor and with the new motor. The observations would be independent between pairs, but the observations within a pair would not be independent because they were taken on the same WM. Other examples are: effectiveness of a diet plan on weight loss for a sample of people, in which the weight is measure before and after the diet, measuring the performance of two different prediction algorithms on the same validation set, etc.*

We have a random sample made of n *pairs* (X_i, Y_i) *and we assume that the differences* $D_i = Y_i - X_i$, *for* $i = 1, \ldots, n$ *are normally distributed with mean* $\delta = \mu_1 - \mu_2$ *and variance* $\sigma_D^2 = \sigma_1^2 + \sigma_2^2 - 2\sigma_{12}$, *or*

$$D_i \sim \mathrm{N}(\delta, \sigma_D^2).$$

All parameters $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \sigma_{12}$ *are unknown. Now let* $\overline{D} = \sum_{i=1}^n D_i/n = \overline{Y} - \overline{X}$, *and*

$$S_D^2 = \frac{\sum_{i=1}^n (D_i - \overline{D})^2}{n-1}.$$

It follows that $\overline{D} \sim \mathrm{N}(\delta, \sigma_D^2/n)$ *and thus*

$$\frac{\overline{D} - \delta}{\sqrt{S_D^2/n}} \sim t_{n-1},$$

is a pivotal quantity. A $(1-\alpha)$ *confidence interval for* $\delta = \mu_1 - \mu_2$ *is given by*

$$\left[\overline{D} - t_{n-1,\alpha/2}\sqrt{\frac{S_D^2}{n}}, \overline{D} + t_{n-1,\alpha/2}\sqrt{\frac{S_D^2}{n}}\right].$$

5.2.2 Asymptotic pivots

All pivots considered in the previous section have exact distributions for any sample size n. When a pivotal quantity is not available, it may still be possible to determine a confidence region for a parameter θ if a statistic exists with an asymptotic distribution that depends on θ but not on any other unknown nuisance parameters. Specifically, let $Y_1, \ldots, Y_n$ have joint p.d.f. $f(y_1, \ldots, y_n; \theta)$ and let $T_n = T(Y_1, \ldots, Y_n) \sim g(t; \theta)$ as n diverges; thus $g(\cdot; \theta)$ is some distribution that depends only on θ. T_n may be a sufficient statistic for θ, or possibly some reasonable estimator such as an MLE. The point here is that, as $n \to \infty$, the distribution $g(\cdot; \theta)$ will have a simple form for which it is possible to eliminate the dependence on θ, and thus get an asymptotic pivot.

We saw in Lecture 4 that under suitable regularity condition the MLE is asymptotically normally distributed, i.e.

$$\frac{\widehat{\theta}_n - \theta}{\mathrm{se}(\widehat{\theta}_n)} \dot{\sim} \mathrm{N}(0,1).$$

The quantity on the right hand side is thus an asymptotic pivot for the parameter θ, and

$$\left[\widehat{\theta} - z_{\alpha/2}\mathrm{se}(\widehat{\theta}_n), \widehat{\theta} + z_{\alpha/2}\mathrm{se}(\widehat{\theta}_n)\right],$$

is an approximate confidence interval for θ with coverage level approximately $(1-\alpha)$; here "approximately" means that

$$P_\theta\left[\widehat{\theta} - z_{\alpha/2}\mathrm{se}(\widehat{\theta}_n) \le \theta \le \widehat{\theta} + z_{\alpha/2}\mathrm{se}(\widehat{\theta}_n)\right] \to 1-\alpha, \quad \text{as} \quad n \to \infty.$$

These type of confidence intervals are called Wald intervals, are always symmetric and are easy to compute since only the MLE and its standard error are required. However, in problems of small sample size in which the actual distribution of the MLE could be far from the normal, these type of intervals may be inaccurate. That is, their actual coverage probability may be far from the confidence level $1-\alpha$. Another issue with Wald confidence intervals is that they do not take care of the constraints on the parameter space Θ. For instance, we may get $[-1.2, 1.6]$ as the confidence interval for the mean λ of a Poisson distribution. Since negative values for λ are not allowed, in such a case, it is customary to discard the negative values and set the lower bound of the interval to the lowest value of the parameter space, i.e. zero in this case.

Under the maximum likelihood framework, it is possible to build likelihood-based confidence intervals which are more accurate than Wald intervals and solve many of the shortcomings of the latters. We will see this in the next lecture.

For the time being here are some examples.

Example 5.8 *Let $Y_1, \ldots, Y_n$ be an i.i.d. sample from the* Poi(λ) *distribution and we wish to build a confidence interval for λ. Consider the sufficient statistic $T_n = \sum_{i=1}^n Y_i$, for which we know that $T_n \sim \mathrm{Poi}(n\lambda)$. Note that this time we do not have a pivot since T_n does not depend on λ.*

However, we know that the MLE of λ is $\widehat{\lambda} = \overline{Y}$ and thus by the consistency of the MLE we know that

$$\frac{\sqrt{n}(\overline{Y}-\lambda)}{\sqrt{\overline{Y}}} \xrightarrow{d} \mathrm{N}(0,1).$$

Thus an approximate $(1-\alpha)$ *Wald confidence interval for* λ *is*

$$\left[\overline{Y} - z_{\alpha/2}\sqrt{\frac{\overline{Y}}{n}}, \overline{Y} + z_{\alpha/2}\sqrt{\frac{\overline{Y}}{n}}\right].$$

Example 5.9 *Let* $Y_1, \ldots, Y_n$ *be an i.i.d. sample from the* $\mathrm{Ber}(\theta)$ *distribution and we wish to build a confidence interval for* θ*. Note again that the sufficient statistic* $T_n = \sum_{i=1}^n Y_i$*, for which we know that* $T_n \sim \mathrm{Bin}(n,\theta)$*, is not a pivot since* T_n *does not depend on* θ*.*

However, we know that the MLE of θ *is* $\widehat{\theta} = \overline{Y}$ *and thus by the consistency of the MLE we know that*

$$\frac{\sqrt{n}(\overline{Y}-\theta)}{\sqrt{\overline{Y}(1-\overline{Y})}} \xrightarrow{d} \mathrm{N}(0,1).$$

Thus an approximate $(1-\alpha)$ *Wald confidence interval for* θ *is*

$$\left[\overline{Y} - z_{\alpha/2}\sqrt{\frac{\overline{Y}(1-\overline{Y})}{n}}, \overline{Y} + z_{\alpha/2}\sqrt{\frac{\overline{Y}(1-\overline{Y})}{n}}\right].$$

Example 5.10 *Let* $Y_1, \ldots, Y_n$ *be an i.i.d. sample from* $\mathrm{Ber}(\theta_1)$ *and let* $X_1, \ldots, X_m$ *be an i.i.d. sample from* $\mathrm{Ber}(\theta_2)$*. We wish to build a confidence interval for* $\theta_2 - \theta_1$*, the difference between the two success probabilities.*

Let $\widehat{\theta}_1 = \overline{Y}$ *and the* $\widehat{\theta}_2 = \overline{X}$ *be the MLE of* θ_1 *and* θ_2*, respectively. By the consistency of the MLE we know that*

$$\frac{\sqrt{n}(\overline{Y}-\theta_1)}{\sqrt{\overline{Y}(1-\overline{Y})}} \xrightarrow{d} \mathrm{N}(0,1) \quad \textit{and} \quad \frac{\sqrt{m}(\overline{X}-\theta_2)}{\sqrt{\overline{X}(1-\overline{X})}} \xrightarrow{d} \mathrm{N}(0,1)$$

Furthermore, since the X_i*'s and* Y_i*'s are independent, then also* $\widehat{\theta}_1$ *is independent from* $\widehat{\theta}_2$*. By the properties of the normal distribution we have that*

$$\widehat{\theta}_2 - \widehat{\theta}_1 \sim \mathrm{N}\left(\theta_2 - \theta_1, \frac{\overline{Y}(1-\overline{Y})}{n} + \frac{\overline{X}(1-\overline{X})}{m}\right)$$

Thus an approximate $(1-\alpha)$ Wald confidence interval for $\theta_2 - \theta_1$ is

$$\left[\overline{Y} - \overline{X} \pm z_{\alpha/2}\sqrt{\tfrac{\overline{Y}(1-\overline{Y})}{n} + \tfrac{\overline{X}(1-\overline{X})}{m}}\right].$$

Example 5.11 *Let $Y_1, \ldots, Y_n$ and $X_1, \ldots, X_m$ be as in Example 5.5 but without assuming variance equality. When variances are not equal it is not easy to eliminate them to obtain a pivotal quantity for $\mu_1 - \mu_2$. One possible approach would be by noticing that, as n and m both diverge to ∞,*

$$\tfrac{\overline{Y}-\overline{X}-(\mu_1-\mu_2)}{\sqrt{S_Y^2/n+S_X^2/m}} \xrightarrow{d} \mathrm{N}(0,1).$$

Thus for large sample sizes, approximate confidence limits for $\mu_1 - \mu_2$ may be easily obtained from this expression. Note that the above limiting result also holds if the samples are not from normal distributions, so this provides a general large-sample result for differences of means. How good the limiting approximation will be in a particular sample depend somewhat on the form of the densities. A better approximation can be obtained by using the t-Student distribution as limiting distribution. This is

$$\tfrac{\overline{Y}-\overline{X}-(\mu_1-\mu_2)}{\sqrt{S_Y^2/n+S_X^2/m}} \sim t_\nu,$$

where the degrees of freedom are estimated by the formula

$$\nu = \tfrac{(S_Y^2/n+S_X^2/m)^2}{\frac{(S_Y^2/n)^2}{n-1}+\frac{(S_X^2/m)^2}{m-1}},$$

where this time ν may not necessarily be an integer. The general problem of making inferences about $\mu_1 - \mu_2$ with unequal variances is known as the Behrens-Fisher problem.

References

[HMC20] Hogg, R. V., McKeen, J. W. and Craig, A. T. (2018) *Introduction to Mathematical Statistics*, 8th edition, global ed., Pearson Education, Chapp. 4.