Chap. 5 Confidence intervals

- **Definition**:

  Confidence interval (CI) = interval supposed to contain true value of the parameter of interest with high probability

  → This definition is very vague

  → Purposes of this chapter:

  1. Understand what this means exactly
  2. Learn methods to find confidence intervals

- **Examples of CI**:

  - Fraction of left-handed people $\in [9.1\%; 10.2\%]$ @ 90% CL
  - Fraction of people with no access to clean water $\in [0.10; 0.12]$ @ 99% CL
  - Higgs mass $\in [124.94; 125.36]$ GeV @ 95% CL

- CI always have the following structure:

$$\boxed{\theta \quad \in \quad \underbrace{[\theta_{\text{min}}; \theta_{\text{max}}]}_{\text{interval}} \quad @ \quad \underbrace{\alpha}_{\substack{\text{value of} \\ \text{confidence level}}} \quad CL}$$

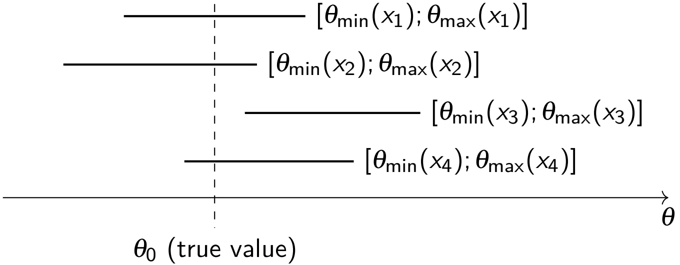
- θ_{min} and θ_{max} called **lower** and **upper bound**
- If either of the two bounds is equal to the parameter limit, interval said to be **one-sided**
- Otherwise, interval said to be **two-sided**
- **Confidence level** α reflects how confident we are that the true parameter is in the quoted interval
- Quoting a result as follows makes no sense

$$\theta \quad \in \quad [\theta_{\text{min}}; \theta_{\text{max}}]$$

- **CI are random objects** (bounds are functions of the data x):

$$[\theta_{\min}(\mathbf{x}); \theta_{\max}(\mathbf{x})]$$

- If you repeat the measurement, you'll get different intervals



  → CI can contain the true value or not

  → You're never 100% sure that the CI contains the true value

- Bayesian approach already described → Will focus on **frequentist approach**

Frequentist approach

- In frequentist approach, building of CI based on **coverage** notion
- Coverage = probability that CI contains true value

$$\boxed{\text{coverage} = P\left(\theta_0 \in [\theta_{\text{min}}(x); \theta_{\text{max}}(x)]\right)}$$

This probability should not be misunderstood:

- **It is not** the probability that θ_0 belongs to the interval you compute from your measurement
- **It is** the probability that, if you repeat the measurement many many times, the intervals you get contain θ_0

Frequentist approach

- CI built in order to have coverage as close as possible to a predefined value

  $\rightarrow$ This predefined value is called **confidence level (α)**

- Typical values of α: 68%, 90%, 95%, 99%

- Confidence level and coverage are not the same thing:
  - Confidence level = objective we try to attain
  - Coverage = what we actually reach when trying to attain the objective

$$\boxed{\textbf{Goal: coverage} \simeq \alpha}$$

More on coverage

- coverage $= \alpha$ can be difficult to achieve in realistic cases
- 3 cases must be distinguished:
  - $P(\theta_0 \in [\theta_{\rm min}(x); \theta_{\rm max}(x)]) = \alpha$: ideal case (perfect coverage)
  - $P(\theta_0 \in [\theta_{\rm min}(x); \theta_{\rm max}(x)]) > \alpha$: not ideal but acceptable (**overcoverage**)
  - $P(\theta_0 \in [\theta_{\rm min}(x); \theta_{\rm max}(x)]) < \alpha$: should be avoided (**undercoverage**)
- A good frequentist method is a method that has no undercoverage and minimal overcoverage

Approximate methods

Methods described here in general don't have known coverage

$\rightarrow$ Can undercover

$\rightarrow$ Use with caution !

- Simplest way to build CI is to start from an estimator $\hat{\theta}$:

$$\boxed{\mathsf{CI} = \left[\hat{\theta} - d\sqrt{\mathrm{var}\left[\hat{\theta}\right]}\,;\hat{\theta} + d\sqrt{\mathrm{var}\left[\hat{\theta}\right]}\right]}$$

where

- $\sqrt{\mathrm{var}\left[\hat{\theta}\right]}$ is the "uncertainty" on the estimate
- d is a real number used to adjust the size of the interval (and thus the coverage)
  - $\rightarrow$ If you want high confidence level, use large d (example: $d = 3$)
  - $\rightarrow$ If you want low confidence level, use small d (example: $d = 1$)

Unbiased normal case

- In this case, possible to achieve perfect coverage

$$\hat{\theta} \sim \mathcal{N}\left(\theta_0, \sqrt{\mathrm{var}\left[\hat{\theta}\right]}\right)$$

$$\begin{aligned}
\Rightarrow \text{coverage} &= P\left(\theta_0 \geq \hat{\theta} - d\sqrt{\mathrm{var}\left[\hat{\theta}\right]} \cap \theta_0 \leq \hat{\theta} + d\sqrt{\mathrm{var}\left[\hat{\theta}\right]}\right) \\
&= P\left(\theta_0 \geq \hat{\theta} - d\sqrt{\mathrm{var}\left[\hat{\theta}\right]}\right) + P\left(\theta_0 \leq \hat{\theta} + d\sqrt{\mathrm{var}\left[\hat{\theta}\right]}\right) - 1 \\
&= \underbrace{P\left(\hat{\theta} \leq \theta_0 + d\sqrt{\mathrm{var}\left[\hat{\theta}\right]}\right)}_{\Phi(d)} + \underbrace{P\left(\hat{\theta} \geq \theta_0 - d\sqrt{\mathrm{var}\left[\hat{\theta}\right]}\right)}_{1-\Phi(-d)=\Phi(d)} - 1
\end{aligned}$$

$$\Rightarrow \text{coverage} = 2\Phi(d) - 1 \quad \text{or} \quad d = \Phi^{-1}\left(\frac{1+\text{coverage}}{2}\right)$$

- **Conclusion:** coverage known once d fixed (we can thus achieve coverage$=\alpha$)

The "number of sigma" way of speaking

- d called in jargon the **"number of sigma"**

  $\rightarrow$ Example: a 2σ confidence interval is an interval with a confidence level of 95.45%

- **Note:** "number of sigma" terminology used even when problem is not normal

  $\rightarrow$ Quoting an interval with its corresponding "number of sigma" doesn't mean that underlying estimator is gaussian

  $\rightarrow$ Underlying estimator can have any distribution (e.g. gamma distribution)

  $\rightarrow$ The number of sigma is just a number that tells what the confidence level is

Approximate methods

- In cases other than the unbiased normal one, coverage in general not known
- However, the following holds in unbiased case:

$$\boxed{\text{coverage} \geq 1 - \frac{1}{d^2}}$$

Leading to:

d	1	2	3	4	5
coverage $\geq$	0	0,75	0,88	0,9375	0,96

- Note: this is true **in general** for unbiased estimators !

Bienaymé-Tchebichev inequality

- Bienaymé-Tchebichev inequality:

$$P\left(|X-\mathbb{E}[X]| \geq \varepsilon\right) \leq \frac{\mathrm{var}[X]}{\varepsilon^2} \qquad \forall \varepsilon > 0$$

  or equivalently

$$P\left(|X-\mathbb{E}[X]| \geq \varepsilon\sigma[X]\right) \leq \frac{1}{\varepsilon^2}$$

- 2 important consequences:
  1. Coverage bounded from below (result under discussion)
  2. Law of large numbers:

$$\mathbb{E}[M] = \mathbb{E}[X] \quad \text{and} \quad \mathrm{var}[M] = \frac{\mathrm{var}[X]}{n}$$

$$\Rightarrow P(|M-\mathbb{E}[X]| \geq \varepsilon) \leq \frac{\mathrm{var}[X]}{n\varepsilon^2}$$

coverage $\geq 1 - 1/d^2$: the proof

- Bienaymé-Tchebichev inequality:

$$P\left(|X - \mathbb{E}[X]| \geq \varepsilon \sigma[X]\right) \leq \frac{1}{\varepsilon^2}$$

- In unbiased case, this leads to:

$$P\left(\left|\hat{\theta} - \theta_0\right| \geq \varepsilon \sqrt{\mathrm{var}\left[\hat{\theta}\right]}\right) \leq \frac{1}{\varepsilon^2}$$

$$\Leftrightarrow P\left(\left|\hat{\theta} - \theta_0\right| \leq \varepsilon \sqrt{\mathrm{var}\left[\hat{\theta}\right]}\right) \geq 1 - \frac{1}{\varepsilon^2}$$

$$\Leftrightarrow P\left(\hat{\theta} - \varepsilon \sqrt{\mathrm{var}\left[\hat{\theta}\right]} \leq \theta_0 \leq \hat{\theta} + \varepsilon \sqrt{\mathrm{var}\left[\hat{\theta}\right]}\right) \geq 1 - \frac{1}{\varepsilon^2}$$

Approximate methods

- If variance not known, can use an estimator:

$$\boxed{\theta \in \left[\hat{\theta} - d\sqrt{\widehat{\mathrm{var}}\left[\hat{\theta}\right]}\,;\hat{\theta} + d\sqrt{\widehat{\mathrm{var}}\left[\hat{\theta}\right]}\right]}$$

- Can be **very risky** to do so:
  - → Coverage in general not known
  - → Can lead to large undercoverage
- Typical situations where such intervals are computed: calculation of proportions and efficiencies
  - → What is the fraction of left-handed people in population ?
  - → What percentage will such or such candidate get at the next presidential election ?
  - → What is the detection efficiency of your device ?

Confidence interval for the proportion

- **Statistical model:** binomial law of probability

$$P(k;N,p) = \binom{N}{k} p^k (1-p)^{N-k}$$

- **Goal:** find confidence interval for p
- **Possible solution:** start from ML estimator of p

$$\hat{p} = \frac{k}{N}$$

- Variance of $\hat{p}$ is:

$$\mathrm{var}[\hat{p}] = \frac{p(1-p)}{N}$$

- Problem: variance depends on unknown parameter of interest p

  $\rightarrow$ Rather than true variance, use an estimate:

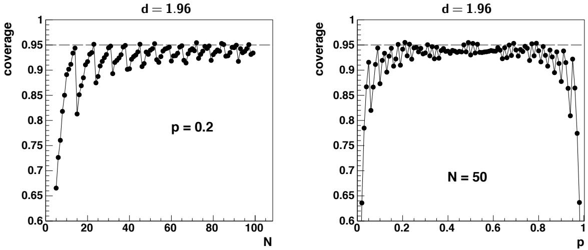
$$\widehat{\mathrm{var}}[\hat{p}] = \frac{\hat{p}(1-\hat{p})}{N}$$

Confidence interval for the proportion

- Following previous reasoning, the CI is:

$$p \in \left[\hat{p} - d\sqrt{\frac{\hat{p}(1-\hat{p})}{N}} ; \hat{p} + d\sqrt{\frac{\hat{p}(1-\hat{p})}{N}} \right] \quad \text{(Wald interval)}$$

- What is the coverage of this CI ?



→ Very large variations (in particular as a function of p)
→ As you don't know p, you don't know coverage

Confidence interval for the proportion

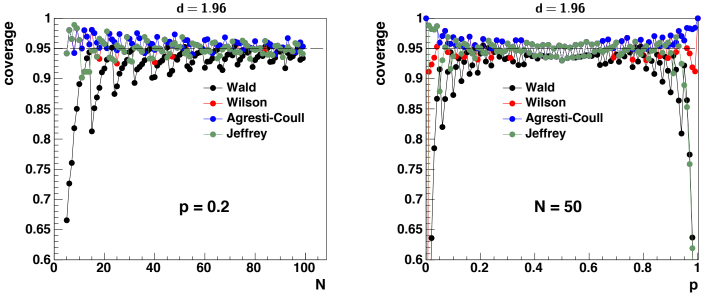
- **Other issue with Wald interval:** leads to empty set when $\hat{p} \to 0$ or 1

- **Example:** Suppose you ask $N = 2$ people whether they are left-handed or not

  - If both say no ($k = 0$), then the CI is: 0 ± 0
  - If both say yes ($k = 2$), then the CI is: 1 ± 0

$\Rightarrow$ Even though sample size very small ($N = 2$), we arrive at certain conclusions (uncertainty on estimated proportion is 0)

- Because of these issues, better intervals have been proposed over the years:



Exercice

Let's consider the following sample:

$$(20.4, 25.4, 25.6, 25.6, 26.6, 28.6, 28.7, 29, 29.8, 30.5, 30.9, 31.1)$$

We assume that these values are realizations of a normal random variable with mean μ and standard deviation σ (both unknown)

$\rightarrow$ Find a confidence interval for μ at 95% CL

Exercice

For the exercice of the previous slide, is it possible to find a better interval than the one you found ?

In order to address this question, consider the fact that

$$t = \frac{\sqrt{n-1}\,(M - \mathbb{E}[X])}{s}, \quad \text{with} \quad s^2 = \frac{\sum_i (X_i - M)^2}{n}$$

follows a Student distribution with $n-1$ degrees of freedom.

Quantiles of Student distribution

	γ										
k	0.25	0.20	0.15	0.10	0.05	0.025	0.010	0.005	0.0025	0.0010	0.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073

Neyman construction

- Method for building CI with good frequentist properties (i.e. no undercoverage, minimal overcoverage): **Neyman construction**

- **Principle:**
  1. For each θ, build acceptance region with probability α:

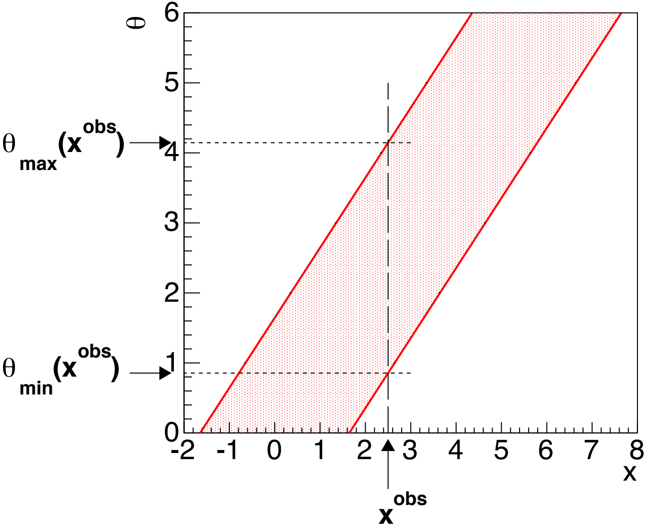
$$[x_{\mathrm{min}}(\theta); x_{\mathrm{max}}(\theta)]$$

  2. From all acceptance regions, build confidence belt
  3. Determine CI for θ from observed value x_{obs}

$$[\theta_{\mathrm{min}}(x_{\mathrm{obs}}); \theta_{\mathrm{max}}(x_{\mathrm{obs}})]$$

$\boxed{\rightarrow \textbf{Resulting interval has confidence level} = \alpha}$

Neyman construction



Remarks:

- Perfect coverage in continuous case
- Impossible to achieve perfect coverage in discrete case

  $\rightarrow$ Choose narrower CI that has coverage $> \alpha$ (i.e. minimal overcoverage)

- Choice of how to build acceptance region free
  - If $x_{\text{min}}(\theta) = -\infty$ or $x_{\text{max}}(\theta) = +\infty$: **one-sided** interval
  - If $x_{\text{min}}(\theta) \neq -\infty$ and $x_{\text{max}}(\theta) \neq +\infty$: **two-sided** interval

    $\rightarrow$ If $P(x < x_{\text{min}}(\theta); \theta) = P(x > x_{\text{max}}(\theta); \theta) = (1-\alpha)/2$: **central interval**

Neyman construction: discrete example

- Neyman construction for one-sided and two-sided intervals in Poisson case:

$$P(n;s) = \frac{(s+b)^n}{n!} e^{-(s+b)}$$