

STK2100- Machine Learning and Statistical Methods for Prediction and Classification

Spring 2026

Short introduction to confidence intervals and
hypothesis testing

Ingrid Hobæk Haff
Department of Mathematics
University of Oslo

Confidence intervals

- Assume that we have observations $x_1, \dots, x_n$ of the random variables $X_1, \dots, X_n$, that follow some distribution with parameter θ
- Then a confidence interval for θ is an interval $(L(x_1, \dots, x_n), U(x_1, \dots, x_n))$, that with large probability contains the true value of θ
- A general way to construct such an interval is to find a statistic $h(X_1, \dots, X_n; \theta)$ (i.e. a random variable that is a function of $X_1, \dots, X_n$), that depends on θ , men whose distribution does not depend on θ
- Let a og b be the $\alpha/2$ - og $(1 - \alpha/2)$ -percentiles of the distribution of $h(X_1, \dots, X_n; \theta)$, meaning that $P(h(X_1, \dots, X_n; \theta) \leq a) = \alpha/2$ and $P(h(X_1, \dots, X_n; \theta) \leq b) = 1 - \alpha/2$
- Then we have

$$P(a \leq h(X_1, \dots, X_n; \theta) \leq b) = 1 - \alpha$$

Confidence intervals

- Then we manipulate the inequalities in the probability to obtain

$$P(L(X_1, \dots, X_n) \leq \theta \leq U(X_1, \dots, X_n)) = 1 - \alpha$$

- A $100 \cdot (1 - \alpha)\%$ confidence interval for θ (where we typically have $\alpha = 0.05$) is obtained by plugging the observed values $x_1, \dots, x_n$ into the limits in the above probability:

$$(L(x_1, \dots, x_n), U(x_1, \dots, x_n))$$

- Often we have an estimate $\hat{\theta}$ of θ , and a confidence interval around $\hat{\theta}$ is a way to quantify our uncertainty as to how close $\hat{\theta}$ should be to the true θ

Hypothesis tests

- In a statistical hypothesis test, one test one hypothesis against another
- The null hypothesis H_0 is the statement that to begin with is assumed to be true, while the alternative hypothesis H_a contradicts H_0
- If the data give strong indications that H_0 is not true, one rejects H_0 in favour of H_a , thus concluding with reasonable certainty that H_a is true
- If one does not reject H_0 , this is not the same as concluding that H_0 is true; the reason could be either that H_0 is true, or that the available data are not sufficient to be able to reject H_0 , even if H_a is true

Hypothesis tests

- In a hypothesis test, we have the following possible outcomes

	H_0 is true	H_a is true
Not reject H_0	Correct	False
Reject H_0	False	Correct

- Hence, one can make two types of error
 - **type I error:** reject H_0 when H_0 is true
 - **type II-feil:** not reject H_0 when H_a is true
- In hypothesis testing, the problem is phrased in such a way that type I error is more serious than type II error
- A statistical test is a rule for when to reject H_0
- It consists in specifying
 - 1 a test statistic, which is a random variable that is a function of the available data, and on which the test is based
 - 2 a rejection area, which are the values of the test statistics for which H_0 is rejected

Hypothesis tests

- The rejection area is a trade-off between the risks of making type I and type II errors, but with the strongest emphasis on avoiding type I error
- The rejection area is specified through the significance level α of the test, which is the maximum probability of making a type I error, and is a parameter that must be chosen (typically 5% or 1%)
- An alternative to defining a rejection region is to use the P-value
- The P-value is the probability of getting a value of the test statistic that contradicts H_0 at least as much as the observed value, given that H_0 is true
- H_0 is then rejected if the P-value is lower than the chosen significance level α