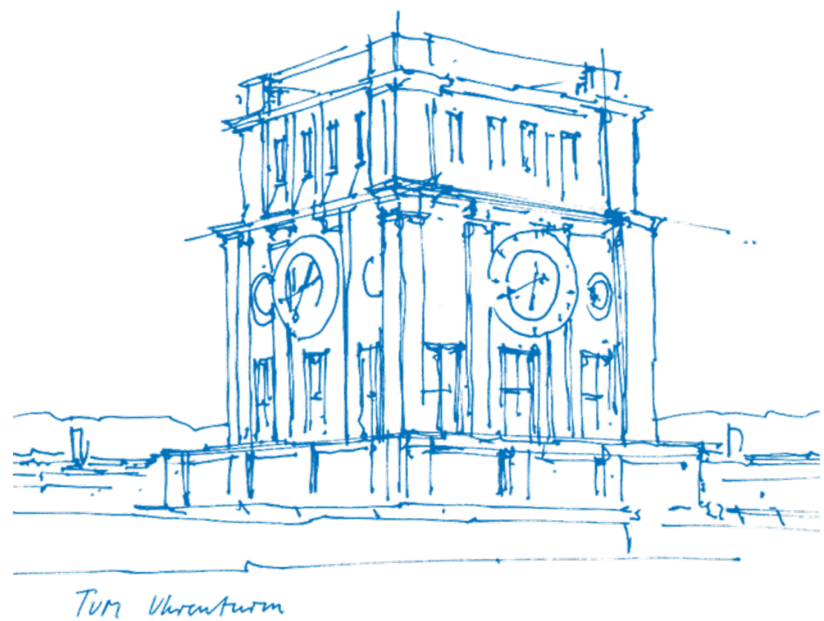
Lecture 6: Statistics Part II

Dr. Shu Okabe
Technische Universität München
shu.okabe@tum.de

Original slides by
Dr. Jakob Piribauer & Prof. Dr. Susanne Albers



Today’s outline: Statistics

1. Estimators
   - Estimators
   - Maximum likelihood estimators

2. Confidence intervals
   - Confidence intervals
   - Estimating the mean
   - Unknown variance

3. Testing hypotheses
   - Testing hypotheses
   - Construction of a test
   - Types of statistical tests

Outline

Testing hypotheses

- Testing hypotheses
- Construction of a test
- Types of statistical tests

Conclusion

Testing hypotheses

So far, we investigated how to estimate certain parameters of a distribution.

Often, one is not interested in the exact parameters, but rather wants to check whether statements related to these parameters are correct.

We will begin by illustrating the idea how to test whether a *hypothesis* about a probability distribution is correct.

Testing hypotheses

Let X be a Bernoulli random variable with success probability p. So, $\Pr(X = 1) = p$ and $\Pr(X = 0) = 1 - p$.

We want to test whether $p < 1/3$ or $p \geq 1/3$.

To this end, we consider a random sample $X_1, \ldots, X_n$ from the distribution of X. For a resulting sample vector $\boldsymbol{x}$, we have to decide whether we accept or reject the hypothesis "$p \geq 1/3$".

The *critical region* of our test is the set

$$K := \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{x} \text{ leads to rejection of the hypothesis}\}.$$

Testing hypotheses

To construct K, we usually construct a new random variable T from $X_1, \dots, X_n$ called the *test statistic*.

Then, the value range of T is divided into regions that lead to the rejection of the hypothesis or not.

Let $\tilde{K} \subseteq \mathbb{R}$ be the region of the value range of T that leads to rejection of the hypothesis. Then, often also $\tilde{K}$ (instead of K itself) is called critical region.

We have

$$K = T^{-1}(\tilde{K}) \subseteq \mathbb{R}^n.$$

Testing hypotheses

The hypothesis we want to test (and reject) is usually denoted by H_0. Hence it is called the *null hypothesis*.

In our earlier example, we could choose

$$H_0: \; p \geq 1/3.$$

If we reject H_0, we accept the *alternative hypothesis* H_1, which in this case is

$$H_1: \; p < 1/3.$$

Often, H_1 is not explicitly stated and implicitly states that H_0 does not hold. In this case, H_1 is called the *trivial* alternative hypothesis.

A non-trivial alternative hypothesis would be

$$H_1': \; p \leq 1/6.$$

Testing hypotheses

Example

We investigate a hard drive, for which we know that it is of one of two possible types. The average access time for type 1 is $9\mathrm{ms}$, while for type 2 it is $12\mathrm{ms}$.

We want to determine the type by measuring the access time of n independent accesses. In this case, we could formulate

$$H_0: \ \mu \leq 9$$

and

$$H_1: \ \mu \geq 12$$

where μ is the expected access time in ms.

Possible errors in hypothesis testing

It could happen that we reject or do not reject the null hypothesis wrongfully based on the random sample, we obtain.

Definition

Rejection of the null hypothesis when it is true is called a *type I error*.
Non-rejection of the null hypothesis when it is false is called a *type II error*.

	H_0 is true	H_0 is false
do not reject H_0	correct decision	type II error
reject H_0	type I error	correct decision

Possible errors in hypothesis testing

Of course, we want to keep the probability of errors of both types small.

However, keeping the probability of type I and of type II errors small are contrary objectives.

If we choose $K = \emptyset$, i.e., if we never reject H_0, the probability for type I errors is 0. Clearly, this is a useless test.

Possible errors in hypothesis testing

The probability of type I errors (rejecting H_0 although it is true) is typically denoted by α. Sometimes, such an error is hence also called α-error.

The probability α is called the *level of significance*.

In practice, it is common to define a level of significance α. The critical region K is then constructed such that the probability of a type I error is α.

Construction of a test

We construct a test for the parameter p of a Bernoulli random variable.

The hypotheses are

$$H_0: \; p \geq p_0 \qquad \text{and} \qquad H_1: \; p < p_0.$$

The test statistic we use is

$$T := X_1 + \ldots + X_n.$$

In this case, we want to reject the null hypothesis if T returns a small value. The critical region for T will hence be of the form $[0, k]$ for a $k \in \mathbb{R}$ to be determined.

This is a *one-sided* test. If the null hypothesis were $p = p_0$, a *two-sided* test where we reject the null hypothesis for small and large values of T would be appropriate.

Construction of a test

The test statistic $T \sim \mathrm{Bin}(n, p)$ is binomially distributed with parameters n and p.

As we assume the size n of the random sample to be large, we can approximate the distribution of T by a normal distribution.

Let

$$\tilde{T} := \frac{T - np}{\sqrt{np(1-p)}}.$$

Then, $\tilde{T}$ is approximately standard normally distributed.

Construction of a test

For the critical region $K = [0, k]$ for T, we now compute the (worst-case) probability of a type I error, i.e., the level of significance α.

We get

$$\begin{aligned}
\Pr(\text{type I error}) &= \max_{p \geq p_0} \Pr_p(T \in K) = \max_{p \geq p_0} \Pr_p(T \leq k) = \Pr_{p=p_0}(T \leq k) = \alpha \\
&= \Pr_{p=p_0}\left(\tilde{T} \leq \frac{k - np}{\sqrt{np(1-p)}}\right) \\
&= \Pr\left(\tilde{T} \leq \frac{k - np_0}{\sqrt{np_0(1-p_0)}}\right) \\
&\approx \Phi\left(\frac{k - np_0}{\sqrt{np_0(1-p_0)}}\right)
\end{aligned}$$

Construction of a test

Using the quantiles z_q of the standard normal distribution, we obtain the following requirement:

The bound k should be chosen such that

$$\frac{k - np_0}{\sqrt{np_0(1-p_0)}} = z_\alpha.$$

This means that k (depending on n) should be chosen as

$$k = z_\alpha \sqrt{np_0(1-p_0)} + np_0.$$

For smaller values of k, the level of significance would be higher. However, this would increase the probability of type II errors. Hence, we choose k such that the level of significance is (approximately) equal to the specified α.

Construction of a test – Type II error

What can we say about the probability of type II errors?

Problem:

$$H_0:\ p \geq p_0 \qquad \text{and} \qquad H_1:\ p < p_0.$$

So, if H_1 is the case, the probability p can still be arbitrarily close to p_0.

For the worst-case probability of a type II error, we obtain

$$\Pr(\text{type II error}) = \sup_{p<p_0} \Pr_p(T > k) = \Pr_{p=p_0}(T > k) \approx 1 - \alpha.$$

If the true situation only slightly deviates from the null hypothesis, with very high probability, we will not reject the null hypothesis although it is false.

For test with a non-trivial alternative hypothesis, however, we can choose the critical region such that both error probabilities are small.

Construction of a test – Type II error

Example

Suppose in a message processing system, two different components may be used. One leads to a probability of message loss of $1/3$; the other to a probability of $1/6$. In this case, we could choose

$$H_0: \ p \geq 1/3 \qquad \text{and} \qquad H_1: \ p \leq 1/6.$$

For the type II error probability, we now obtain

$$\begin{aligned}
\Pr(\text{type II error}) &= \sup_{p \leq 1/6} \Pr_p(T > k) = \Pr_{p=1/6}(T > k) \\
&= \Pr\left(\tilde{T} > \frac{k - n/6}{\sqrt{(1/6)\cdot(5/6)\cdot n}}\right) \\
&\approx 1 - \Phi\left(\frac{k - n/6}{\sqrt{(1/6)\cdot(5/6)\cdot n}}\right).
\end{aligned}$$

The probability of a type II error is often denoted by β.

Construction of a test

Put together, for the hypothesis

$$H_0: \; p \geq 1/3 \qquad \text{and} \qquad H_1: \; p \leq 1/6,$$

we have seen the following:

To achieve a level of significance of α, we have to choose

$$k = z_\alpha \sqrt{n \cdot (1/3) \cdot (1-1/3)} + n/3.$$

For $\alpha = 0.05$, we obtain

$$k \approx 0.33 \cdot n - 0.77 \cdot \sqrt{n}.$$

The resulting probability of a type II error is

$$\beta = \Pr(\text{type II error}) \approx 1 - \Phi\left(\frac{k - n/6}{\sqrt{(1/6) \cdot (5/6) \cdot n}}\right) \approx 1 - \Phi(0.45\sqrt{n} - 2.07).$$

Construction of a test

$$H_0:\ p \geq 1/3 \qquad \text{and} \qquad H_1 \leq 1/6.$$

For $\alpha = 0.05$, we get the following values for different sample sizes n:

n	k	$\beta = \Pr(\text{type II error})$
25	4.4	$1 - \Phi(0.18) \approx 0.43$
100	25.3	$1 - \Phi(2.43) \approx 0.0075$
900	276.9	$1 - \Phi(11.43) \approx 0$

Application of statistical tests

The test we constructed above is called *approximate binomial test*.

Assumptions: $X_1, \ldots, X_n$ are independent and identically distributed with $\Pr(X_i = 1) = p$ and $\Pr(X_i = 0) = 1 - p$ for all i and an unknown p.

Further, n is sufficiently large; we assume we can approximate the binomial distribution $\mathrm{Bin}(n, p)$ with a normal distribution.

Hypotheses:

1. $H_0: p = p_0$ and $H_1: p \neq p_0$ (two-sided).
2. $H_0: p \geq p_0$ and $H_1: p < p_0$ (one-sided).
3. $H_0: p \leq p_0$ and $H_1: p > p_0$ (one-sided).

Test statistic:

$$Z := \frac{\sum_{i=1}^{n} X_i - np_0}{\sqrt{np_0(1-p_0)}}.$$

Critical region (rejection of H_0) for level of significance α:

1. $|Z| > z_{1-\alpha/2}$.
2. $Z < z_{\alpha}$.
3. $Z > z_{1-\alpha}$.

Application of statistical tests

Typical values of α are 0.05 and 0.01.

For the two-sided test, H_0 is rejected if $|Z| > z_{1-\alpha/2}$.

In the one-sided tests, H_0 is rejected if $Z < z_\alpha$ or $Z > z_{1-\alpha}$, respectively.

α	0.05	0.01
$z_{1-\alpha/2}$	1.96	2.58
z_α	-1.64	-2.33
$z_{1-\alpha}$	1.64	2.33

The quantiles in the table specify by how many standard deviations the observed number of successes has to deviate from what is to be expected for $p = p_0$ in order to reject the null hypothesis.

Application of statistical tests

In general, one can follow the following procedure to test hypotheses while ensuring a fixed probability of type I errors:

1. State the null hypothesis and the alternative hypothesis.
2. Choose a fixed level of significance α.
3. Choose an appropriate test statistic and establish the critical region based on α.
4. Compute the test statistic from a random sample. Reject H_0 if the computed test statistic lies in the critical region. Otherwise, do not reject.
5. Draw conclusions.

P-values

Instead of fixing a level of significance α and computing a critical region in which H_0 is rejected, one can compute a so called *p-value*.

Example

Consider an approximate binomial test with the hypotheses

$$H_0:\ p = 1/2 \qquad \text{and} \qquad H_1:\ p \neq 1/2.$$

Assume, we consider a random sample of size $n = 100$. Suppose we observe 59 successes in the random sample.

We compute

$$Z = \frac{59 - 100 \cdot 1/2}{\sqrt{100 \cdot (1/2) \cdot (1/2)}} = \frac{9}{5} = 1.8.$$

P-values

Example (continued)

For the fixed level of significance $\alpha = 0.05$, we cannot reject the null hypothesis:

$$1.8 = |Z| < z_{1-\alpha/2} \approx 1.96.$$

However, we see that the test statistic is close to its critical region. Hence, we might wonder: For which level of significance α would we reject the null hypothesis based on the observed random sample?

We are interested in

$$\inf\{\alpha \mid 1.8 < z_{1-\alpha/2}\} \approx 0.072.$$

Definition

The *p-value* of an observed test statistic is the lowest level of significance for which the null hypothesis would be rejected.

Types of statistical tests

There are several criteria according to which statistical tests can be classified.

Number of involved random variables

If there is only one distribution that is analysed, tests are called *one-sample tests*.

If there are multiple distributions that are compared, tests are called *two-sample tests*.

Example

One sample tests can be used to answer questions like “Is the average access time to the server at most $10\mathrm{ms}$?”.

Two sample tests can be used to answer questions like “Is the average access time of server A lower than that of server B?”.

Types of statistical tests

In the case of two involved random variables, it is important to know whether the random variables are independent.

To detect (in-)dependence, there are various ways of *correlation analysis*.

In order to provide a quantitative connection between random variables, *regression analysis* can be used.

Types of statistical tests

Formulation of the null hypothesis

Often, the null hypothesis is a statement about a *location parameter* of a distribution.

These parameters could be the *expected value*, the *variance*, or the *median*, i.e., the least value x such that $F(x) = 1/2$ where F is the cumulative distribution function of the distribution.

A two-sample test could, e.g., address the question whether the median of two distributions is the same.

Besides hypotheses about location parameters, one can also formulate hypotheses like “The analysed distribution has the distribution function F”.

Types of statistical tests

Assumptions on the analysed distributions

Usually, some assumptions about the distribution are made when constructing a test.

E.g., a certain type of distribution is assumed or the expected value or the variance is assumed to be a certain value.

Selection of some statistical tests

We have already seen the **approximate binomial test**:

Assumptions: $X_1, \dots, X_n$ are independent and identically distributed with $\Pr(X_i = 1) = p$ and $\Pr(X_i = 0) = 1 - p$ for all i and an unknown p.
Further, n is sufficiently large; we assume we can approximate the binomial distribution $\mathrm{Bin}(n, p)$ with a normal distribution.

Hypotheses:

1. $H_0: \ p = p_0$ and $H_1: \ p \neq p_0$ (two-sided).
2. $H_0: \ p \geq p_0$ and $H_1: \ p < p_0$ (one-sided).
3. $H_0: \ p \leq p_0$ and $H_1: \ p > p_0$ (one-sided).

Test statistic:

$$Z := \frac{\sum_{i=1}^{n} X_i - np_0}{\sqrt{np_0(1-p_0)}}.$$

Critical region (rejection of H_0) for level of significance α:

1. $|Z| > z_{1-\alpha/2}$.
2. $Z < z_{\alpha}$.
3. $Z > z_{1-\alpha}$.

Selection of some statistical tests – Z-test

As we used the central limit theorem, to assume that the test statistic is approximately normally distributed, the approximate binomial test is a special case of Z**-test**.

Assumptions: $X_1, \dots, X_n$ are independent and identically distributed with $X_i \sim \mathcal{N}(\mu, \sigma^2)$, where σ^2 is known. Alternatively, $\mathbb{E}(X_i) = \mu$ and $\mathrm{Var}(X_i) = \sigma^2$ and the sample size n is large.

Hypotheses:

1. $H_0: \ \mu = \mu_0$ and $H_1: \ \mu \neq \mu_0$ (two-sided).
2. $H_0: \ \mu \geq \mu_0$ and $H_1: \ \mu < \mu_0$ (one-sided).
3. $H_0: \ \mu \leq \mu_0$ and $H_1: \ \mu > \mu_0$ (one-sided).

Test statistic:

$$Z := \frac{\bar{X} - \mu_0}{\sigma} \sqrt{n}.$$

Critical region (rejection of H_0) for level of significance α:

1. $|Z| > z_{1-\alpha/2}$.
2. $Z < z_{\alpha}$.
3. $Z > z_{1-\alpha}$.

Selection of some statistical tests – Z-test

The downside of the Z-test is that the variance σ^2 has to be known.

As we have seen before, it is an obvious attempt to replace σ^2 with the sample variance S^2 if the variance is not known.

When using S^2 instead of the true σ^2, the test statistic has a t-distribution (with $n-1$ degrees of freedom). For $n \geq 30$, this distribution is very similar to the standard normal distribution.

Selection of some statistical tests – t-test

Assumptions: $X_1, \ldots, X_n$ are independent and identically distributed with $X_i \sim \mathcal{N}(\mu, \sigma^2)$.
Alternatively, $\mathbb{E}(X_i) = \mu$ and $\mathrm{Var}(X_i) = \sigma^2$ and the sample size n is large.

Hypotheses:

1. $H_0: \ \mu = \mu_0$ and $H_1: \ \mu \neq \mu_0$ (two-sided).
2. $H_0: \ \mu \geq \mu_0$ and $H_1: \ \mu < \mu_0$ (one-sided).
3. $H_0: \ \mu \leq \mu_0$ and $H_1: \ \mu > \mu_0$ (one-sided).

Test statistic:

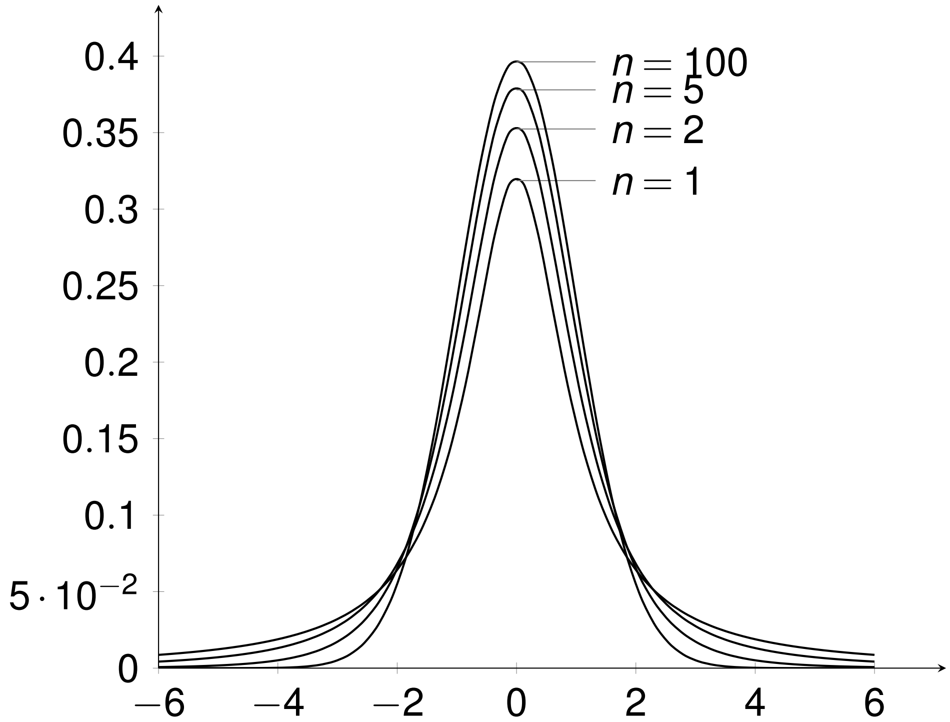
$$T := \frac{\bar{X} - \mu_0}{S} \sqrt{n}.$$

Critical region (rejection of H_0) for level of significance α:

1. $|T| > t_{1-\alpha/2}$.
2. $T < t_{\alpha}$.
3. $T > t_{1-\alpha}$.

Selection of some statistical tests – t-test

The probability density function of the t-distribution with n degrees of freedom:



For $n \to \infty$, the t-distribution converges to the standard normal distribution.

Selection of some statistical tests – t-test

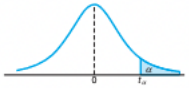


Table A.4 Critical Values of the t-Distribution

	α						
v	**0.40**	**0.30**	**0.20**	**0.15**	**0.10**	**0.05**	**0.025**
1	0.325	0.727	1.376	1.963	3.078	6.314	12.706
2	0.289	0.617	1.061	1.386	1.886	2.920	4.303
3	0.277	0.584	0.978	1.250	1.638	2.353	3.182
4	0.271	0.569	0.941	1.190	1.533	2.132	2.776
5	0.267	0.559	0.920	1.156	1.476	2.015	2.571
6	0.265	0.553	0.906	1.134	1.440	1.943	2.447
7	0.263	0.549	0.896	1.119	1.415	1.895	2.365
8	0.262	0.546	0.889	1.108	1.397	1.860	2.306
9	0.261	0.543	0.883	1.100	1.383	1.833	2.262
10	0.260	0.542	0.879	1.093	1.372	1.812	2.228
11	0.260	0.540	0.876	1.088	1.363	1.796	2.201
12	0.259	0.539	0.873	1.083	1.356	1.782	2.179
13	0.259	0.538	0.870	1.079	1.350	1.771	2.160
14	0.258	0.537	0.868	1.076	1.345	1.761	2.145
15	0.258	0.536	0.866	1.074	1.341	1.753	2.131
16	0.258	0.535	0.865	1.071	1.337	1.746	2.120
17	0.257	0.534	0.863	1.069	1.333	1.740	2.110
18	0.257	0.534	0.862	1.067	1.330	1.734	2.101
19	0.257	0.533	0.861	1.066	1.328	1.729	2.093
20	0.257	0.533	0.860	1.064	1.325	1.725	2.086
21	0.257	0.532	0.859	1.063	1.323	1.721	2.080
22	0.256	0.532	0.858	1.061	1.321	1.717	2.074
23	0.256	0.532	0.858	1.060	1.319	1.714	2.069
24	0.256	0.531	0.857	1.059	1.318	1.711	2.064
25	0.256	0.531	0.856	1.058	1.316	1.708	2.060
26	0.256	0.531	0.856	1.058	1.315	1.706	2.056
27	0.256	0.531	0.855	1.057	1.314	1.703	2.052
28	0.256	0.530	0.855	1.056	1.313	1.701	2.048
29	0.256	0.530	0.854	1.055	1.311	1.699	2.045
30	0.256	0.530	0.854	1.055	1.310	1.697	2.042
40	0.255	0.529	0.851	1.050	1.303	1.684	2.021
60	0.254	0.527	0.848	1.045	1.296	1.671	2.000
120	0.254	0.526	0.845	1.041	1.289	1.658	1.980
∞	0.253	0.524	0.842	1.036	1.282	1.645	1.960

Picture: Walpole, Myers, Myers, Ye: Probability and Statistics for Engineers and Scientists (Ninth edition)

Selection of some statistical tests – Two-sample test

Now, we take a quick look at a two-sample test.

For two random variables X and Y, we want to test whether the expected values $\mu_X = \mathbb{E}(X)$ and $\mu_Y = \mathbb{E}(Y)$ are equal.

A random sample now consists of independent copies $X_1, \ldots, X_n$ of X and $Y_1, \ldots, Y_m$ of Y.

There are several variants of the two-sample-t-test in the literature.

These variants can be applied if the two distributions do not have the same variance.

Two-sample-t-test

Assumptions: $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$ are independent and identically distributed with $X_i \sim \mathcal{N}(\mu_X, \sigma_X^2)$ and $Y_i \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$.
Furthermore, the variances are equal, i.e., $\sigma_X^2 = \sigma_Y^2$.

Hypotheses:

1. $H_0: \ \mu_X = \mu_Y$ and $H_1: \ \mu_X \neq \mu_Y$ (two-sided).
2. $H_0: \ \mu_X \geq \mu_Y$ and $H_1: \ \mu_X < \mu_Y$ (one-sided).
3. $H_0: \ \mu_X \leq \mu_Y$ and $H_1: \ \mu_X > \mu_Y$ (one-sided).

Test statistic:

$$T := \frac{\bar{X} - \bar{Y}}{\sqrt{(n-1)S_X^2 + (m-1)S_Y^2}} \sqrt{\frac{n+m-2}{1/m + 1/n}}.$$

Critical region (rejection of H_0) for level of significance α:

- $|T| > t_{1-\alpha/2}$.
- $T < t_\alpha$.
- $T > t_{1-\alpha}$.

Selection of some statistical tests

Finally, we take a look at a test that is not concerned with a location parameter.

Example

We want to test a die and find out whether it is fair.

So, our hypothesis is that the outcome has a uniform distribution with probability mass function

$$f(x) = 1/6 \quad \text{for } x = 1, \ldots, 6.$$

We toss the die 120 times and observe the following:

Outcome:	1	2	3	4	5	6
Expected:	20	20	20	20	20	20
Observed:	20	22	17	18	19	24

χ^2-goodness-of-fit test

For this hypothesis, we can use a **goodness-of-fit test**.

This goodness-of-fit test makes use of the test statistic

$$\chi^2 = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i}$$

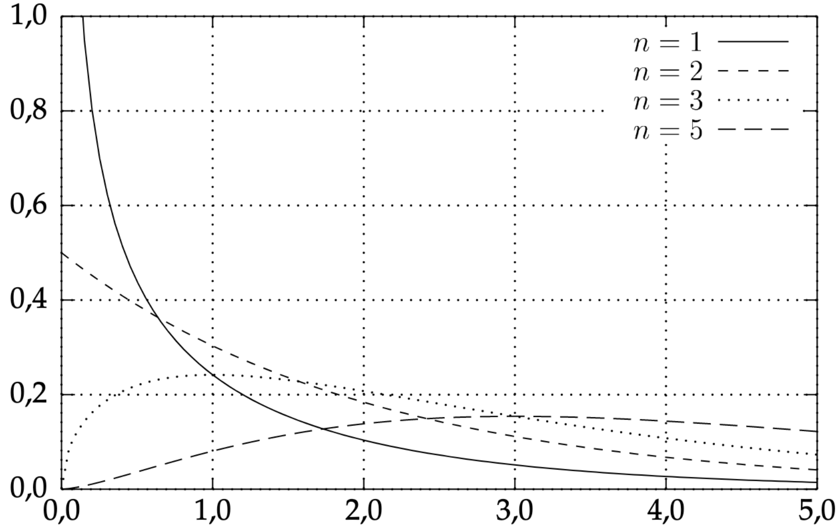
where o_i are the observed frequencies for $i = 1, \ldots, k$ and e_i are the corresponding expected frequencies.

The distribution of the test statistic χ^2 is approximated very closely by the so-called *chi-squared distribution* with $k - 1$ degrees of freedom.

If χ^2 is large, the observed frequencies are a poor fit for the expected frequencies and the null hypothesis should be rejected.

The critical region for a level of significance $1 - \alpha$ can be obtained via the quantiles χ^2_α of the chi-squared distribution. (In the following table χ^2_α is what we would denote by $\chi^2_{1-\alpha}$.)

χ^2-goodness-of-fit test



Plot of the probability density function of the chi-squared distribution with *n* degrees of freedom.

χ^2-goodness-of-fit test

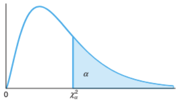


Table A.5 Critical Values of the Chi-Squared Distribution

	α									
v	**0.995**	**0.99**	**0.98**	**0.975**	**0.95**	**0.90**	**0.80**	**0.75**	**0.70**	**0.50**
1	0.0^4393	0.0^3157	0.0^3628	0.0^3982	0.00393	0.0158	0.0642	0.102	0.148	0.455
2	0.0100	0.0201	0.0404	0.0506	0.103	0.211	0.446	0.575	0.713	1.386
3	0.0717	0.115	0.185	0.216	0.352	0.584	1.005	1.213	1.424	2.366
4	0.207	0.297	0.429	0.484	0.711	1.064	1.649	1.923	2.195	3.357
5	0.412	0.554	0.752	0.831	1.145	1.610	2.343	2.675	3.000	4.351
6	0.676	0.872	1.134	1.237	1.635	2.204	3.070	3.455	3.828	5.348
7	0.989	1.239	1.564	1.690	2.167	2.833	3.822	4.255	4.671	6.346
8	1.344	1.647	2.032	2.180	2.733	3.490	4.594	5.071	5.527	7.344
9	1.735	2.088	2.532	2.700	3.325	4.168	5.380	5.899	6.393	8.343
10	2.156	2.558	3.059	3.247	3.940	4.865	6.179	6.737	7.267	9.342
11	2.603	3.053	3.609	3.816	4.575	5.578	6.989	7.584	8.148	10.341
12	3.074	3.571	4.178	4.404	5.226	6.304	7.807	8.438	9.034	11.340
13	3.565	4.107	4.765	5.009	5.892	7.041	8.634	9.299	9.926	12.340
14	4.075	4.660	5.368	5.629	6.571	7.790	9.467	10.165	10.821	13.339
15	4.601	5.229	5.985	6.262	7.261	8.547	10.307	11.037	11.721	14.339
16	5.142	5.812	6.614	6.908	7.962	9.312	11.152	11.912	12.624	15.338
17	5.697	6.408	7.255	7.564	8.672	10.085	12.002	12.792	13.531	16.338
18	6.265	7.015	7.906	8.231	9.390	10.865	12.857	13.675	14.440	17.338
19	6.844	7.633	8.567	8.907	10.117	11.651	13.716	14.562	15.352	18.338
20	7.434	8.260	9.237	9.591	10.851	12.443	14.578	15.452	16.266	19.337
21	8.034	8.897	9.915	10.283	11.591	13.240	15.445	16.344	17.182	20.337
22	8.643	9.542	10.600	10.982	12.338	14.041	16.314	17.240	18.101	21.337
23	9.260	10.196	11.293	11.689	13.091	14.848	17.187	18.137	19.021	22.337
24	9.886	10.856	11.992	12.401	13.848	15.659	18.062	19.037	19.943	23.337
25	10.520	11.524	12.697	13.120	14.611	16.473	18.940	19.939	20.867	24.337
26	11.160	12.198	13.409	13.844	15.379	17.292	19.820	20.843	21.792	25.336
27	11.808	12.878	14.125	14.573	16.151	18.114	20.703	21.749	22.719	26.336
28	12.461	13.565	14.847	15.308	16.928	18.939	21.588	22.657	23.647	27.336
29	13.121	14.256	15.574	16.047	17.708	19.768	22.475	23.567	24.577	28.336
30	13.787	14.953	16.306	16.791	18.493	20.599	23.364	24.478	25.508	29.336
40	20.707	22.164	23.838	24.433	26.509	29.051	32.345	33.66	34.872	39.335
50	27.991	29.707	31.664	32.357	34.764	37.689	41.449	42.942	44.313	49.335
60	35.534	37.485	39.699	40.482	43.188	46.459	50.641	52.294	53.809	59.335

Picture: Walpole, Myers, Myers, Ye: Probability and Statistics for Engineers and Scientists (Ninth edition)

χ^2-goodness-of-fit test

Table A.5 (continued) Critical Values of the Chi-Squared Distribution

	α									
v	**0.30**	**0.25**	**0.20**	**0.10**	**0.05**	**0.025**	**0.02**	**0.01**	**0.005**	**0.001**
1	1.074	1.323	1.642	2.706	3.841	5.024	5.412	6.635	7.879	10.827
2	2.408	2.773	3.219	4.605	5.991	7.378	7.824	9.210	10.597	13.815
3	3.665	4.108	4.642	6.251	7.815	9.348	9.837	11.345	12.838	16.266
4	4.878	5.385	5.989	7.779	9.488	11.143	11.668	13.277	14.860	18.466
5	6.064	6.626	7.289	9.236	11.070	12.832	13.388	15.086	16.750	20.515
6	7.231	7.841	8.558	10.645	12.592	14.449	15.033	16.812	18.548	22.457
7	8.383	9.037	9.803	12.017	14.067	16.013	16.622	18.475	20.278	24.321
8	9.524	10.219	11.030	13.362	15.507	17.535	18.168	20.090	21.955	26.124
9	10.656	11.389	12.242	14.684	16.919	19.023	19.679	21.666	23.589	27.877
10	11.781	12.549	13.442	15.987	18.307	20.483	21.161	23.209	25.188	29.588
11	12.899	13.701	14.631	17.275	19.675	21.920	22.618	24.725	26.757	31.264
12	14.011	14.845	15.812	18.549	21.026	23.337	24.054	26.217	28.300	32.909
13	15.119	15.984	16.985	19.812	22.362	24.736	25.471	27.688	29.819	34.527
14	16.222	17.117	18.151	21.064	23.685	26.119	26.873	29.141	31.319	36.124
15	17.322	18.245	19.311	22.307	24.996	27.488	28.259	30.578	32.801	37.698
16	18.418	19.369	20.465	23.542	26.296	28.845	29.633	32.000	34.267	39.252
17	19.511	20.489	21.615	24.769	27.587	30.191	30.995	33.409	35.718	40.791
18	20.601	21.605	22.760	25.989	28.869	31.526	32.346	34.805	37.156	42.312
19	21.689	22.718	23.900	27.204	30.144	32.852	33.687	36.191	38.582	43.819
20	22.775	23.828	25.038	28.412	31.410	34.170	35.020	37.566	39.997	45.314
21	23.858	24.935	26.171	29.615	32.671	35.479	36.343	38.932	41.401	46.796
22	24.939	26.039	27.301	30.813	33.924	36.781	37.659	40.289	42.796	48.268
23	26.018	27.141	28.429	32.007	35.172	38.076	38.968	41.638	44.181	49.728
24	27.096	28.241	29.553	33.196	36.415	39.364	40.270	42.980	45.558	51.179
25	28.172	29.339	30.675	34.382	37.652	40.646	41.566	44.314	46.928	52.619
26	29.246	30.435	31.795	35.563	38.885	41.923	42.856	45.642	48.290	54.051
27	30.319	31.528	32.912	36.741	40.113	43.195	44.140	46.963	49.645	55.475
28	31.391	32.620	34.027	37.916	41.337	44.461	45.419	48.278	50.994	56.892
29	32.461	33.711	35.139	39.087	42.557	45.722	46.693	49.588	52.335	58.301
30	33.530	34.800	36.250	40.256	43.773	46.979	47.962	50.892	53.672	59.702
40	44.165	45.616	47.269	51.805	55.758	59.342	60.436	63.691	66.766	73.403
50	54.723	56.334	58.164	63.167	67.505	71.420	72.613	76.154	79.490	86.660
60	65.226	66.981	68.972	74.397	79.082	83.298	84.58	88.379	91.952	99.608

Picture: Walpole, Myers, Myers, Ye: Probability and Statistics for Engineers and Scientists (Ninth edition)

χ^2-goodness-of-fit test

In our example, we had

Outcome:	1	2	3	4	5	6
Expected:	20	20	20	20	20	20
Observed:	20	22	17	18	19	24

We compute the value

$$\chi^2 = \sum_{i=1}^{6} \frac{(o_i - e_i)^2}{e_i} \approx 1.7.$$

For a level of significance of 95%, we get from the table that $\chi^2_{0.05}$ with five degrees of freedom is 11.070. So, there is no reason to reject the null hypothesis.

χ^2-goodness-of-fit test

Assumptions: $X_1, \ldots, X_n$ are independent and identically distributed with $W_{X_i} = \{1, \ldots, k\}$.

Hypotheses:

- $H_0:\ \Pr(X_i = i) = p_i$ for $i = 1, \ldots, k$.
- $H_1:\ $ trivial alternative, i.e., H_0 does not hold.

Test statistic:

$$\chi^2 = \sum_{i=1}^{k} \frac{(o_i - n \cdot p_i)^2}{n \cdot p_i}$$

where o_i is the observed frequency of value i in the random sample.

Critical region (rejection of H_0) for level of significance α:
$\chi^2 > \chi^2_\alpha$ for $k-1$ degrees of freedom. (Or in our notation for quantiles $\chi^2 > \chi^2_{1-\alpha}$.)

The expected frequencies of all outcomes should be sufficiently large, e.g., all at least 5.

A more detailed rule of thumb asks all expected frequencies to be at least 1 and 80% of the expected frequencies to be at least 5.

Approximate Binomial Test

Example

A small proportion $p \in [0,1]$ of the displays produces in a factory is defective. It should be guaranteed that this does not exceed 1%. To test this, $N = 10100$ randomly selected displays were scanned for their quality out of which 151 were defective.

Is the results in accordance with the quality criterion on a level of 95% confidence?

To check this, we test the null hypothesis versus the alternative hypothesis:

$$H_0\colon p \leq 0.01 \text{ vs } H_1\colon p > 0.01$$

Thus, we need a one-sided test.

Our test will be based on a statistical model for the random sample. We assume that these are N=10,100 independent realisations of a Bernoulli experiment with parameter p.

Equivalently, we can consider the number of defective displays X in the batch which is binomially distributed with parameters N and p.

Approximate Binomial Test

Example (continued)

For large sample size (10,100 certainly is "large") we may use the approximate binomial test which is based on assuming

$$X \sim \mathcal{N}(np, np(1-p)).$$

The test statistic here is

$$\begin{aligned} Z &= \frac{X-\mu}{\sqrt{\mathrm{Var}(X)}} = \frac{X - N \cdot d}{\sqrt{N \cdot d \cdot (d-1)}} \\ &= \frac{151 - 10100 \cdot 0.01}{\sqrt{\underbrace{10100 \cdot 0.99}_{\approx 10000} \cdot 0.01}} \\ &\approx \frac{50}{10} = 5. \end{aligned}$$

Approximate Binomial Test

Example (continued)

We have to compare this with the critical value (or quantile)

$$z_{1-\alpha} = z_{0.95} = 1.65.$$

Since it holds that

$$Z = 5.0 > 1.65 = z_{0.95},$$

we are in the critical region of H_0.

We have to change our belief and reject H_0. Based on the observation, the quality criterion is not kept with the probability of 95%.

Approximate Binomial Test

Example (continued)

Now assume that the number of broken pixels on a single display is Poisson distributed. What is the corresponding rate λ if the true proportion of defective displays is $p = 1 - e^{-0.01} \approx 0.01$? Here, all displays with at least one broken pixel are treated as defective. The complement event is that a display does not contain a defective pixel. This has the probability

$$\text{Poi}(\lambda, 0) = \frac{\lambda^0 e^{-\lambda}}{0!} = e^{-\lambda}.$$

This corresponds to the proportion of proper displays $1 - p$. So,

$$e^{-\lambda} = 1 - p = e^{-0.01} \quad \Rightarrow \lambda = 0.01.$$

Approximate Binomial Test

Example (continued)

Lastly, we are interested in the p-value of our observation. We ask: up to which confidence level does the rejection hold?

$$z_{1-\alpha} = 5 \quad \Rightarrow \quad \Phi(z_{1-\alpha}) = \Phi(5.)$$
$$1-\alpha \approx 0.9999997 \quad \Rightarrow \quad \alpha \approx 3 \cdot 10^{-7}.$$

So, only on a super-safe confidence level of more than 99.99%, we could rise some doubt about the rejection.

What we have covered

- Estimators and maximum likelihood estimation
- Confidence intervals
- Statistical tests

Acknowledgements

- This lecture is a slightly modified version from Dr. Jakob Piribauer and Prof. Dr. Susanne Albers's slides on discrete probability theory.
- Some material are extracted from the book: Stanley H. Chan, *Introduction to Probability for Data Science*, Michigan Publishing, 2021.

Questions?

Thank you for your attention!