Chapter 3: Basics from Probability Theory and Statistics

3.1 Probability Theory

Events, Probabilities, Bayes‘ Theorem,

Random Variables, Distributions, Moments, Tail Bounds,

Central Limit Theorem, Entropy Measures

3.2 Statistical Inference

Sampling, Parameter Estimation, Maximum Likelihood,

Confidence Intervals, Hypothesis Testing, p-Values,

Chi-Square Test, Linear and Logistic Regression

mostly following L. Wasserman Chapters 6, 9, 10, 13

3.2 Statistical Inference

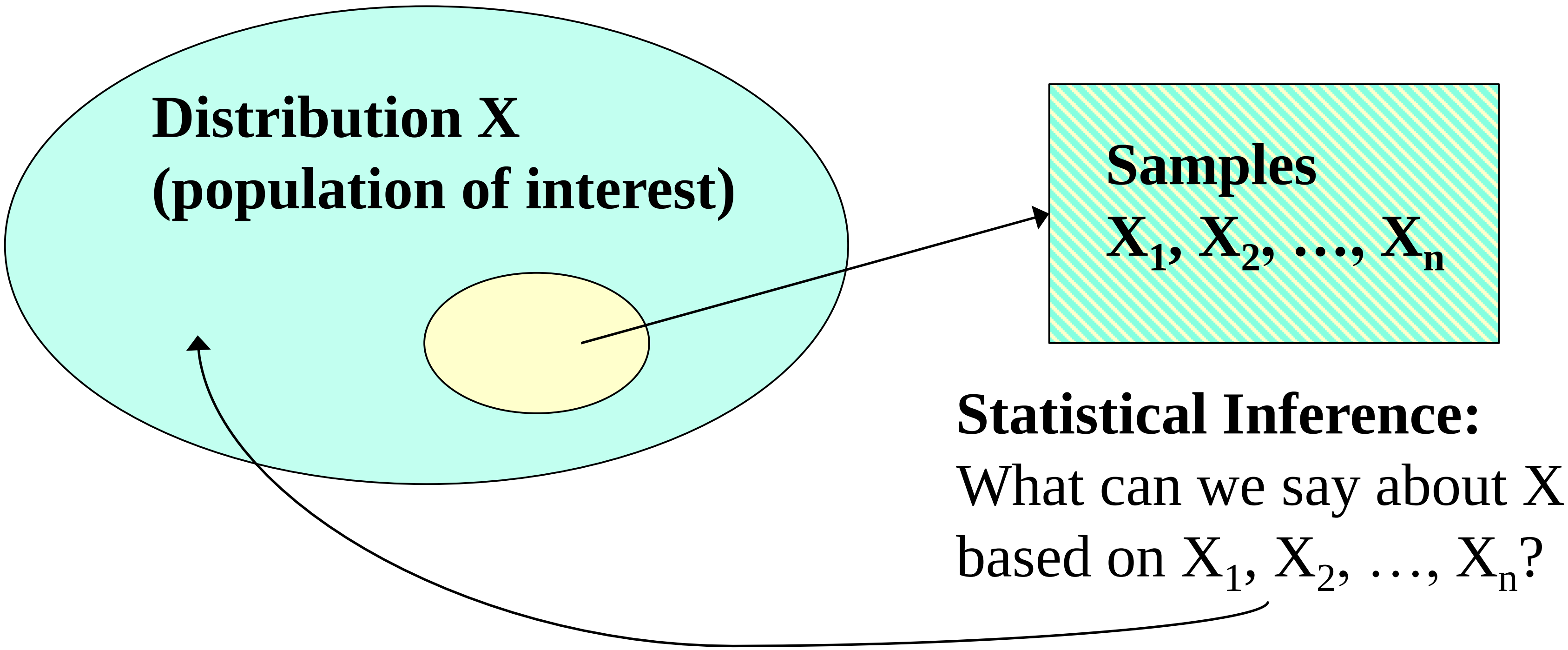
A ***statistical model*** is a set of distributions (or regression functions), e.g., all unimodal, smooth distributions.
A ***parametric model*** is a set that is completely described by a finite number of parameters,
(e.g., the family of Normal distributions).

Statistical inference: given a sample $X_1, \ldots, X_n$ how do we infer the distribution or its parameters within a given model.

For multivariate models with one specific ***„outcome (response)" variable Y***, this is called ***prediction*** or ***regression,***
for discrete outcome variable also ***classification***.
$r(x) = E[Y \mid X=x]$ is called the ***regression function***.

Example for classification: biomedical markers $\rightarrow$ cancer or not
Example for regression: business indicators $\rightarrow$ stock price

Sampling Illustrated



Example: estimate the average salary in Germany?

Approach 1: ask your 10 neighbors
Approach 2: ask 100 random people you spot on the Internet
Approach 2: ask all 1000 living Germans in Wikipedia
Approach 4: ask 1000 random people from all age groups, jobs, …

Basic Types of Statistical Inference

Given: independent and identically distributed (iid) samples
$X_1, X_2, \ldots, X_n$ from (unknown) distribution X

- **Parameter estimation:**
  What is the parameter p of a Bernoulli coin?
  What are the values of μ and σ of a Normal distribution?
  What are $\alpha_1, \alpha_2, \lambda_1, \lambda_2$ of a Poisson mixture?

- **Confidence intervals:**
  What is the interval [mean $\pm$ tolerance] s.t. the expectation of my observations or measurements falls into the interval with high confidence?

- **Hypothesis testing:**
  H_0: p=1/2 (fair coin) vs. H_1: p $\neq$1/2
  H0: p1 = p2 (methods have same precision) vs. H1: p1 $\neq$ p2

- **Regression (for parameter fitting)**

3.2.1 Statistical Parameter Estimation

A ***point* estimator** for a parameter θ of a prob. distribution is a random variable X derived from a random sample $X_1, \ldots, X_n$.

<u>Examples:</u>

Sample mean: $$\bar{X} := \frac{1}{n} \sum_{i=1}^{n} X_i$$

Sample variance: $$S^2 := \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

An estimator T for parameter θ is ***unbiased***

if $E[T] = \theta$;

otherwise the estimator has ***bias*** $E[T] - \theta$.

An estimator on a sample of size n is ***consistent***

if $lim_{n \to \infty} P[|T - \theta| < \varepsilon] = 1 \;\; for\ each\ \varepsilon > 0$

Sample mean and sample variance
are unbiased, consistent estimators with minimal variance.

Estimation Error

Let $\hat{\theta}_n$ = T(θ) be an estimator for parameter θ over sample $X_1, ..., X_n$.

The distribution of $\hat{\theta}_n$ is called the sampling distribution.

The ***standard error*** for $\hat{\theta}_n$ is: $se(\hat{\theta}) = \sqrt{Var(\hat{\theta}_n)}$

The ***mean squared error (MSE)*** for $\hat{\theta}_n$ is:

$$
\begin{aligned}
MSE(\hat{\theta}) &= E[(\hat{\theta}_n - \theta)^2] \\
&= bias^2(\hat{\theta}_n) + Var[\hat{\theta}_n]
\end{aligned}
$$

If bias → 0 and se → 0 then the estimator is consistent.

The estimator $\hat{\theta}_n$ is ***asymptotically Normal*** if

$(\hat{\theta}_n - \theta)/se$ converges in distribution to standard Normal N(0,1)

Nonparametric Estimation

The ***empirical distribution function*** $\hat{F}_n$ is the cdf that puts prob. mass 1/n at each data point X_i: $\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \le x)$

where indicator function $I(X_i \le x)$

is 1 if $X_i \le x$ and 0 otherwise

A ***statistical functional*** T(F) is any function of F,
e.g., mean, variance, skewness, median, quantiles, correlation

The ***plug-in estimator*** of θ = T(F) is: $\hat{\theta}_n = T(\hat{F}_n)$

Nonparametric Estimation: Histograms

Instead of the full empirical distribution, often compact data synopses may be used, such as ***histograms*** where $X_1, \ldots, X_n$ are grouped into m cells (buckets or bins) $c_1, \ldots, c_m$ with

bucket boundaries $lb(c_i)$ and $ub(c_i)$ s.t.

$lb(c_1) = -\infty$, $ub(c_m) = \infty$, $ub(c_i) = lb(c_{i+1})$ for $1 \le i < m$, and

freq(c_i) = $\hat{F}_n(x) = \frac{1}{n} \sum_{\nu=1}^{n} I(lb(c_i) \le X_\nu < ub(c_i))$

Histograms provide a (discontinuous) ***density estimator***.

<u>Example:</u>

***X*₁** —

$\boldsymbol{X_1 = X_2 = 1}$

$\boldsymbol{X_3 = X_4 = X_5 = 2}$

$\boldsymbol{X_6 = \ldots X_{10} = 3}$

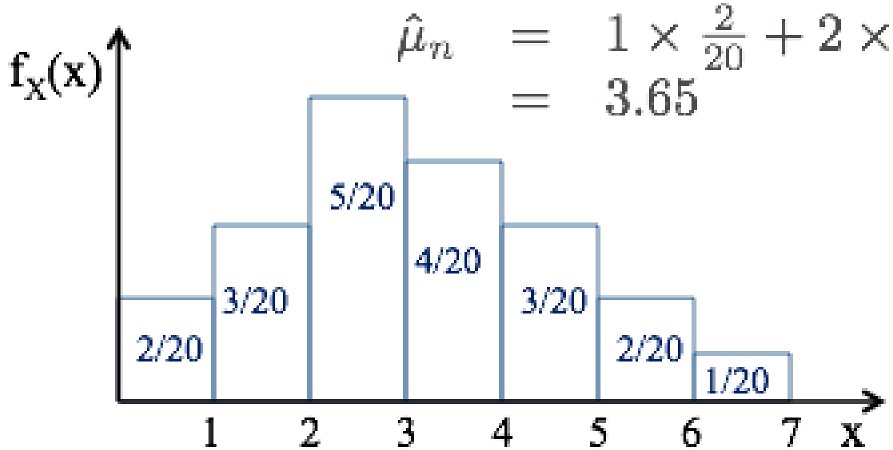
$\boldsymbol{X_{11} = \ldots X_{14} = 4}$

$\boldsymbol{X_{15} = \ldots X_{17} = 5}$

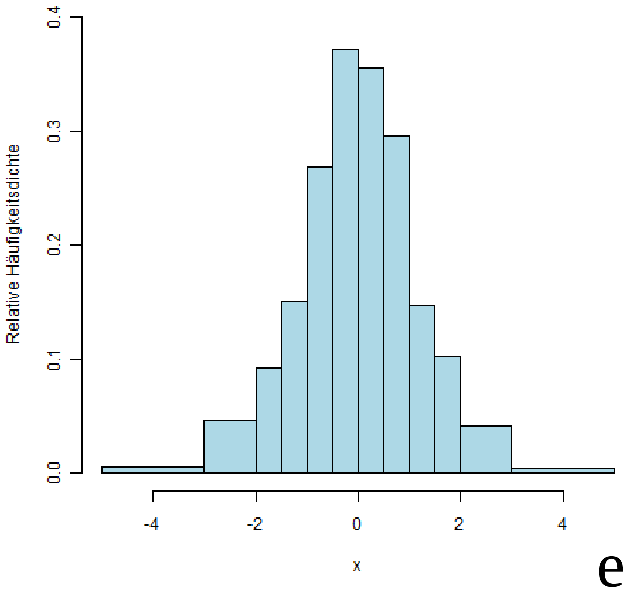
$\boldsymbol{X_{18} = X_{19} = 6}$

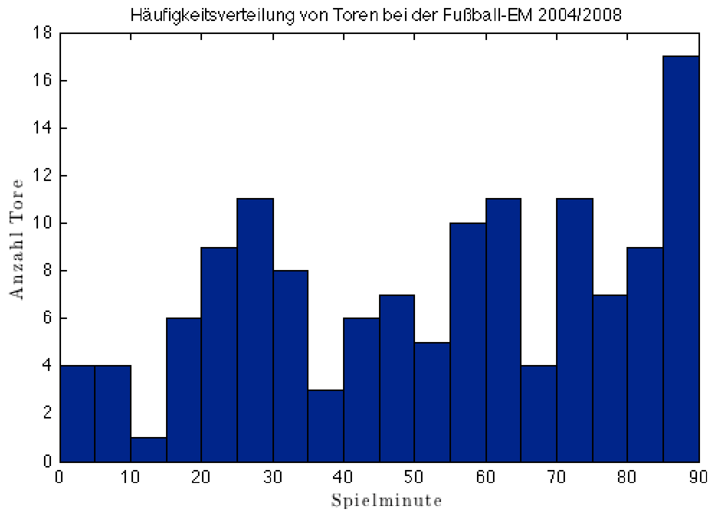
$\boldsymbol{X_{20} = 7}$

$$\hat{\mu}_n = 1 \times \frac{2}{20} + 2 \times \frac{3}{20} + \ldots + 7 \times \frac{1}{20} = 3.65$$

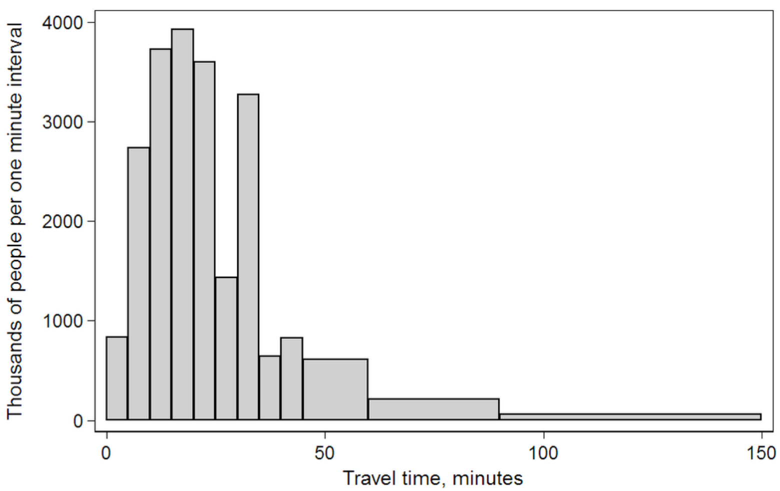


Different Kinds of Histograms





equidistant buckets



Sources:
en.wikipedia.org
de.wikipedia.org

non-equidistant buckets

Method of Moments

- Suppose parameter $\theta = (\theta_1, \ldots, \theta_k)$ has k components
- Compute ***j*-th moment** for $1 \le j \le k$:

$$\alpha_j = \alpha_j(\theta) = E_\theta[X^j] = \int_{-\infty}^{+\infty} x^j \, f_X(x) \, dx$$

- Compute ***j*-th sample moment** for $1 \le j \le k$: $\hat{\alpha}_j = \frac{1}{n} \sum_{i=1}^{n} X_i^j$

- **Method-of-moments** estimate of θ is obtained by solving a system of k equations in k unknowns:

$$\alpha_1(\hat{\theta}_n) = \hat{\alpha}_1$$
$$\vdots$$
$$\alpha_k(\hat{\theta}_n) = \hat{\alpha}_k$$

Method-of-moments estimators are usually consistent and asympotically Normal, but may be biased

Example: Method of Moments

Let $X_1, \ldots, X_n \sim \text{Normal}(\mu,\sigma^2)$

$$\alpha_1 = E_\theta[X] = \mu$$

$$\alpha_2 = E_\theta[X^2] = Var[X] + E[X]^2 = \sigma^2 + \mu^2$$

Solve the equation system:

$$\mu = \alpha_1 = \widehat{\alpha_1} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \sigma^2 + \mu^2 = \alpha_2 = \widehat{\alpha_2} = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

Solution: $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$ $\qquad \widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$

Parametric Inference: Maximum Likelihood Estimators (MLE)

Estimate parameter θ of a postulated distribution f(θ,x) such that the probability that the data of the sample are generated by this distribution is maximized.

→ **Maximum likelihood estimation:**

Maximize $L(x_1,...,x_n, \theta) = P[x_1, ..., x_n$ originate from $f(\theta,x)]$

often written as

$$\hat{\theta}_{MLE} = argmax_{\theta}\ L(\theta, x_1,...,x_n)$$
$$= argmax_{\theta}\ \prod_{i=1}^{n} f(x_i, , \theta)$$

or maximize *log L*

if analytically untractable → use numerical iteration methods

MLE Properties

Maximum Likelihood Estimators are
consistent, asymptotically Normal, and
asymptotically optimal in the following sense:

Consider two estimators U and T which are asymptotically Normal.
Let u^2 and t^2 denote the variances of the two Normal distributions to which U and T converge in probability.
The asymptotic relative efficiency of U to T is $ARE(U,T) = t^2/u^2$.

Theorem: For an MLE $\hat{\theta}_n$ and any other estimator $\tilde{\theta}_n$ the following inequality holds:

$$ARE(\tilde{\theta}_n, \hat{\theta}_n) \leq 1$$

Simple Example for Maximum Likelihood Estimator

given:

- coin with Bernoulli distribution with unknown parameter p für head, 1-p for tail
- sample (data): k times head with n coin tosses

needed: maximum likelihood estimation of p

Let L(k, n, p) = P[sample is generated from distr. with param. p]

$$= \binom{n}{k} p^k \, (1-p)^{n-k}$$

Maximize log-likelihood function log L (k, n, p):

$$\log \quad \mathrm{L} = \log \binom{\mathrm{n}}{\mathrm{k}} \quad + \quad \mathrm{k} \quad \log \mathrm{p} \quad + \quad (\mathrm{n}-\mathrm{k}) \quad \log \quad (1-\mathrm{p})$$

$$\frac{\partial \log L}{\partial p} = \frac{k}{p} - \frac{n-k}{1-p} = 0 \qquad \Rightarrow p = \frac{k}{n}$$

Advanced Example for Maximum Likelihood Estimator

given:

- Poisson distribution with parameter λ (expectation)
- sample (data): numbers $x_1, ..., x_n \in N_0$

needed: maximum likelihood estimation of λ

Let r be the largest among these numbers,
and let $f_0, ..., f_r$ be the absolute frequencies of numbers 0, ..., r.

$$L(x_1,\ldots,x_n,\lambda) = \prod_{i=0}^{r} \left(e^{-\lambda} \frac{\lambda^i}{i!} \right)^{f_i}$$

$$\Rightarrow \frac{\partial \ln L}{\partial \lambda} = \sum_{i=0}^{r} f_i \left(\frac{i}{\lambda} - 1 \right) = 0 \qquad \Rightarrow \hat{\lambda} = \frac{\sum_{i=0}^{r} i\, f_i}{\sum_{i=0}^{r} f_i} = \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x}$$

Sophisticated Example for Maximum Likelihood Estimator

given:

- discrete uniform distribution over $[1,\theta] \subseteq N_0$ and density $f(x) = 1/\theta$
- sample (data): numbers $x_1, ..., x_n \in N_0$

MLE for θ is $\max\{x_1, ..., x_n\}$ (see Wasserman p. 124)

MLE for Parameters of Normal Distributions

$$L(x_1,\ldots,x_n,\mu,\sigma^2) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \prod_{i=1}^{n} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

$$\frac{\partial\, ln(L)}{\partial\,\mu} = \frac{-1}{2\sigma^2}\sum_{i=1}^{n} 2(x_i-\mu) \;=0$$

$$\frac{\partial\, ln(L)}{\partial\,\sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2 = 0$$

$$\Rightarrow \quad \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad\qquad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\hat{\mu})^2$$

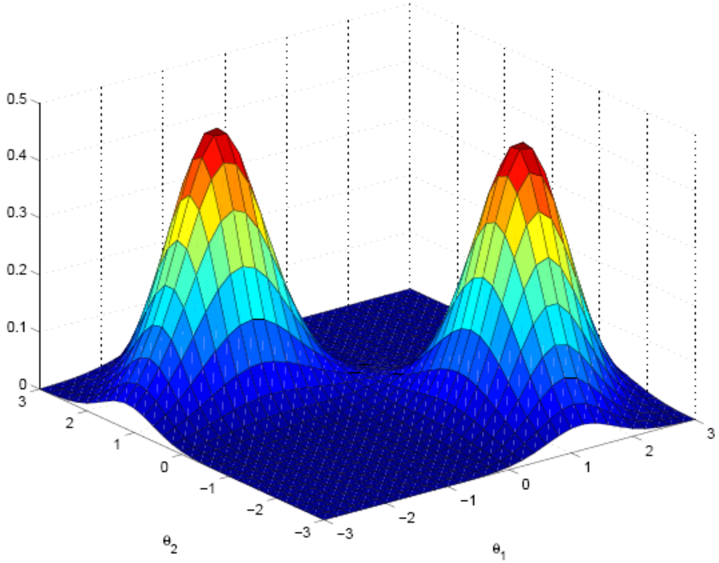
Analytically Non-tractable MLE for parameters of Multivariate Normal Mixture

consider samples from a mixture of multivariate Normal distributions with the density (e.g. height and weight of males and females):

$$f(\vec{x}, \pi_1, \ldots, \pi_k, \vec{\mu}_1, \ldots, \vec{\mu}_k, \Sigma_1, \ldots, \Sigma_k)$$

$$= \sum_{j=1}^{k} \pi_j \; n(\vec{x}, \vec{\mu}_j, \Sigma_j) = \sum_{j=1}^{k} \pi_j \; \frac{1}{\sqrt{(2\pi)^m \left|\Sigma_j\right|}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu}_j)^T \Sigma_j^{-1} (\vec{x}-\vec{\mu}_j)}$$

with expectation values $\vec{\mu}_j$
and invertible, positive definite, symmetric m×m covariance matrices Σ_j



$\rightarrow$ maximize log-likelihood function:

$$\log L(\vec{x}_1, \ldots, \vec{x}_n, \theta) := \log \prod_{i=1}^{n} P[\vec{x}_i \mid \theta] = \sum_{i=1}^{n} \left(\log \sum_{j=1}^{k} \pi_j \; n(\vec{x}_i, \vec{\mu}_j, \Sigma_j) \right)$$

Expectation-Maximization Method (EM)

When $L(\theta, X_1, ..., X_n)$ is analytically intractable then

- introduce ***latent (non-observable) random variable(s) Z*** such that:
  joint distribution $J(X_1, ..., X_n, Z, \theta)$ of „complete" data is tractable
- iteratively compute:
  - **Expectation (E Step):**

    compute expected complete data likelihood

    $E_Z[\log J(X_1, \ldots, X_n, Z \mid \theta^{(t)})]$ given a previous estimate of θ
  - **Maximization (M Step):**

    estimate $\theta^{(t+1)}$ that maximizes $E_Z[\log J(X_1, \ldots, X_n, Z \mid \theta^{(t)})]$

details depend on distribution at hand (often mixture models)
convergence guaranteed, but problem is non-convex $\rightarrow$ numerical methods

Bayesian Viewpoint of Parameter Estimation

- assume prior distribution $g(\theta)$ of parameter θ
- choose statistical model (generative model) $f(x \mid \theta)$ that reflects our beliefs about RV X
- given RVs $X_1, \ldots, X_n$ for observed data, the posterior distribution is $h(\theta \mid x_1, \ldots, x_n)$

for $X_1=x_1, \ldots, X_n=x_n$ the likelihood is

$$L(x_1 \ldots x_n, \theta) = \prod_{i=1}^{n} f(x_i \mid \theta) = \prod_{i=1}^{n} \frac{h(\theta \mid x_i) \cdot \sum_{\theta'} f(x_i \mid \theta') g(\theta')}{g(\theta)}$$

which implies

$$h(\theta \mid x_1 \ldots x_n) \sim L(x_1 \ldots x_n, \theta) \cdot g(\theta)$$

(posterior is proportional to likelihood times prior)

MAP estimator (maximum a posteriori):

compute θ that maximizes $h(\theta \mid x_1, \ldots, x_n)$ *given a prior for* θ

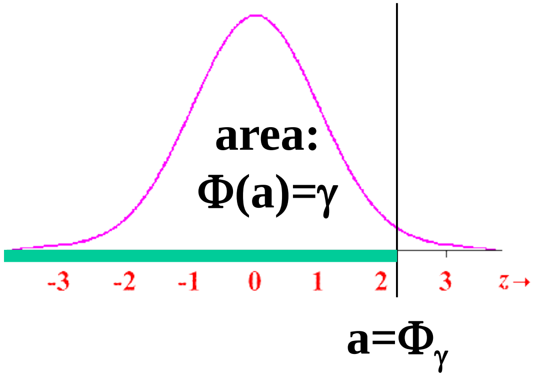
3.2.2 Confidence Intervals

Estimator T for an interval for parameter θ such that

$$P[T - a \le \theta \le T + a] = 1 - \alpha$$

[T-a, T+a] is the **confidence interval** and 1-α is the **confidence level.**

For the distribution of random variable X a value
x_γ (0< γ <1) with $P[X \le x_\gamma] \ge \gamma \wedge P[X \ge x_\gamma] \ge 1-\gamma$
is called a **γ quantile**; the 0.5 quantile is called the **median**.
For the normal distribution N(0,1) the γ quantile is denoted Φ_γ .



Confidence Intervals for Expectations (1)

Let $x_1, \ldots, x_n$ be a sample from a distribution with unknown expectation μ and known variance σ^2.

For sufficiently large n the sample mean $\bar{X}$ is $N(\mu,\sigma^2/n)$ distributed and $\frac{(\bar{X}-\mu)\sqrt{n}}{\sigma}$ is N(0,1) distributed:

$$P[-z \le \frac{(\bar{X}-\mu)\sqrt{n}}{\sigma} \le z] = \Phi(z) - \Phi(-z) = \Phi(z) - (1-\Phi(z)) = 2\Phi(z) - 1$$

$$= P[\bar{X} - \frac{z\sigma}{\sqrt{n}} \le \mu \le \bar{X} + \frac{z\sigma}{\sqrt{n}}]$$

$$\Rightarrow P[\bar{X} - \frac{\Phi_{1-\alpha/2}\,\sigma}{\sqrt{n}} \le \mu \le \bar{X} + \frac{\Phi_{1-\alpha/2}\,\sigma}{\sqrt{n}}] = 1-\alpha$$

For required confidence interval $[\bar{X}-a, \bar{X}+a]$ or confidence level 1-α set

$z := \frac{a\sqrt{n}}{\sigma}$

then look up $\Phi(z)$

to determine $1-\alpha/2$

or

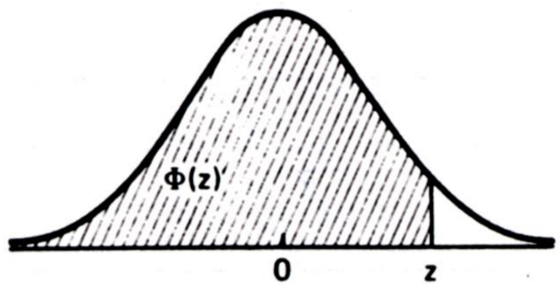
$z := (1-\frac{\alpha}{2})$ *quantile of* $N(0,1)$

then set $a := \frac{z\sigma}{\sqrt{n}}$

to determine interval

Normal Distribution Table

The Normal Distribution Functions $\Phi(z) = \int_{-\infty}^{z} \frac{e^{-t^2/2}}{\sqrt{2\pi}} dt$



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188	.53586
0.1	.53983	.54380	.54776	.55172	.55567	.55962	.56356	.56749	.57142	.57535
0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026	.61409
0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803	.65173
0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082	.68439	.68793
0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904	.72240
0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175	.75490
0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230	.78524
0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785	.81057	.81327
0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398	.83646	.83891
1.0	.84134	.84375	.84614	.84849	.85083	.85314	.85543	.85769	.85993	.86214
1.1	.86433	.86650	.86864	.87076	.87286	.87493	.87698	.87900	.88100	.88298
1.2	.88493	.88686	.88877	.89065	.89251	.89435	.89617	.89796	.89973	.90147
1.3	.90320	.90490	.90658	.90824	.90988	.91149	.91308	.91466	.91621	.91774
1.4	.91924	.92073	.92220	.92364	.92507	.92647	.92785	.92922	.93056	.93189
1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062	.94179	.94295	.94408
1.6	.94520	.94630	.94738	.94845	.94950	.95053	.95154	.95254	.95352	.95449
1.7	.95543	.95637	.95728	.95818	.95907	.95994	.96080	.96164	.96246	.96327
1.8	.96407	.96485	.96562	.96638	.96712	.96784	.96856	.96926	.96995	.97062
1.9	.97128	.97193	.97257	.97320	.97381	.97441	.97500	.97558	.97615	.97670
2.0	.97725	.97778	.97831	.97882	.97932	.97982	.98030	.98077	.98124	.98169
2.1	.98214	.98257	.98300	.98341	.98382	.98422	.98461	.98500	.98537	.98574
2.2	.98610	.98645	.98679	.98713	.98745	.98778	.98809	.98840	.98870	.98899
2.3	.98928	.98956	.98983	.99010	.99036	.99061	.99086	.99111	.99134	.99158

Confidence Intervals for Expectations (2)

Let $x_1, \ldots, x_n$ be a sample from a distribution with unknown expectation μ and *unknown variance* σ^2 and sample variance S^2.
For sufficiently large n the random variable

$$T := \frac{(\bar{X} - \mu)\sqrt{n}}{S}$$

has a t distribution (Student distribution) with n-1 degrees of freedom:

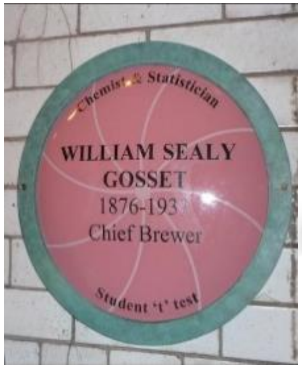
$$f_{T,n}(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \frac{1}{\sqrt{n\pi}\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}}$$

with the Gamma function: $\Gamma(x) = \int_0^\infty e^{-t}\, t^{x-1}\, dt \quad \textit{für } x > 0$

$$(\textit{with the properties } \Gamma(1) = 1 \textit{ and } \Gamma(x+1) = x\Gamma(x))$$

$$\Rightarrow P\left[\, \bar{X} - \frac{t_{n-1,1-\alpha/2}\; S}{\sqrt{n}} \le \mu \le \bar{X} + \frac{t_{n-1,1-\alpha/2}\; S}{\sqrt{n}} \,\right] = 1 - \alpha$$

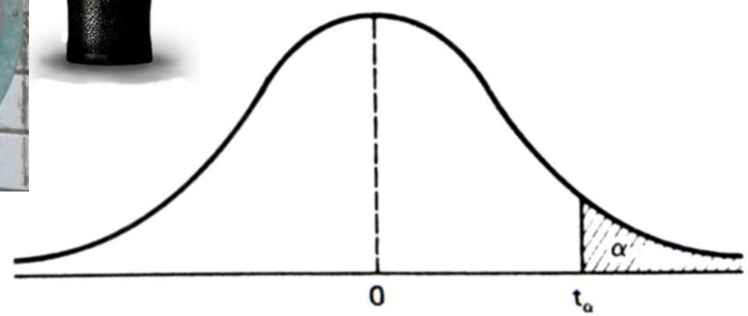
Student‘s t Distribution Table



William Gosset
(1876-1937)

Table 5

Critical Values of the Student-t Distribution[a]



A. Student:
The Probable Error of a Mean,
Biometrika 6(1), 1908

n \ α	0.10	0.05	0.025	0.01	0.005
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898

Example: Confidence Interval for Expectation

X: time for student to solve exercise

n=16 samples, $\bar{X} = 2.5, S^2 = 0.25$

$$P[-z \le \frac{(\bar{X}-\mu)\sqrt{n}}{\sigma} \le z] = \Phi(z) - \Phi(-z) = \Phi(z) - (1-\Phi(z)) = 2\Phi(z) - 1$$

$$= P[\bar{X} - \frac{z\sigma}{\sqrt{n}} \le \mu \le \bar{X} + \frac{z\sigma}{\sqrt{n}}]$$

$$\Rightarrow P[\bar{X} - \frac{\Phi_{1-\alpha/2}\,\sigma}{\sqrt{n}} \le \mu \le \bar{X} + \frac{\Phi_{1-\alpha/2}\,\sigma}{\sqrt{n}}] = 1-\alpha$$

A) Assume σ^2 is known: σ^2=0.25

A1) Estimate μ±0.2

A2) Estimate μ with 1–α=0.9 confidence

for interval $[\bar{X}-a, \bar{X}+a]$:

$$z := \frac{a\sqrt{n}}{\sigma}$$

then look up Φ(z)
to determine 1–α/2

B) Assume σ^2 is unknown

B1) Estimate μ±0.2

B2) Estimate μ with 1–α=0.9 confidence

for confidence 1–α:

$$z := (1-\frac{\alpha}{2}) \text{ quantile of } N(0,1)$$

then set $a := \frac{z\sigma}{\sqrt{n}}$
to determine interval

3.2.3 Hypothesis Testing

Hypothesis testing:

- aims to falsify some hypothesis by lack of statistical evidence
- design of **test RV (test statistic)** and its (approx. / limit) distribution

Example:

- Toss a coin n times and judge if the coin is fair
  $X_1, \ldots, X_n \sim$ Bernoulli(p), coin is fair if $p = 0.5$
- Let the **null hypothesis** H_0 be "the coin is fair"
- The **alternative hypothesis** H_1 is then "the coin is not fair"
- Intuitively, if $|\bar{X} - 0.5|$ is large, we should **reject H_0**

H_0 is default, interest is in H_1: aim to reject H_0
(e.g. suspecting that the coin is unfair)

Hypothesis Testing Terminology (1)

A hypothesis test determines a probability 1-α (***test level*** $\boldsymbol{\alpha}$***, significance level***) that a sample $X_1, ..., X_n$ from some unknown probability distribution has a certain property.

Examples:

1) The sample originates from a normal distribution.
2) Under the assumption of a normal distribution the sample originates from a $N(\mu, \sigma^2)$ distribution.
3) Two random variables are independent.
4) Two random variables are identically distributed.
5) Parameter λ of a Poisson distribution from which the sample stems has value 5.

H_0 is default, interest is in H_1

General form:

null hypothesis $\boldsymbol{H_0}$ ***vs. alternative hypothesis*** $\boldsymbol{H_1}$

needs ***test variable (test statistic) X*** (derived from $X_1, ..., X_n, H_0, H_1$) and ***test region R*** with $X \in R$ for rejecting H_0 and $X \notin R$ for retaining H_0

	Retain H_0	Reject H_0
H_0 true	✓	type I error
H_1 true	type II error	✓

Hypothesis Testing Terminology (2)

- $\theta = \theta_0$ is called a **simple hypothesis**
- $\theta > \theta_0$ or $\theta < \theta_0$ is called a **composite hypothesis**
- $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$ is called a **two-sided test**
- $H_0 : \theta \leq \theta_0$ vs. $H_1 : \theta > \theta_0$ and $H_0 : \theta \geq \theta_0$ vs. $H_1 : \theta < \theta_0$ are called a **one-sided test**
- **Rejection region** R : if $X \in R$, reject H_0 otherwise retain H_0
- The rejection region is typically defined using a **test statistic** T and a **critical value** c: $R = \{ X : T(X) > c \}$

p-Value

Suppose that for every level $\alpha \in (0,1)$ there is a test with rejection region R_α. Then the ***p-value*** is the smallest level at which we can reject H_0: $p\text{-}value = inf\{ \alpha \mid T(X_1,...,X_n) \in R_\alpha \}$

small p-value means strong evidence against H_0

typical interpretation of p-values:

- < 0.01 **very strong** evidence against H_0
- 0.01 – 0.05: **strong** evidence against H_0
- 0.05 – 0.10: **weak** evidence against H_0
- > 0.1: **little or no** evidence against H_0

p-value: prob. of test statistic (sample) as extreme as the observed data under H_0

Caution: p-value $\neq P[H_0|data]$

Hypothesis Testing Example

Null hypothesis for n coin tosses: coin is fair or has head probability $p = p_0$; ***alternative hypothesis***: $p \neq p_0$

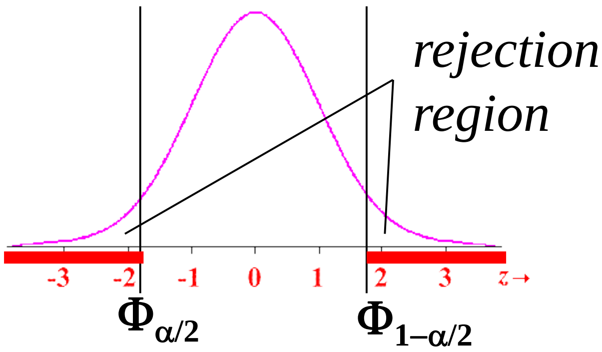
Test variable: X, the #heads, is

N(pn, p(1-p)n) distributed (by the Central Limit Theorem),

thus $$Z := \frac{(X/n - p)\sqrt{n}}{\sqrt{p(1-p)}}$$ is N(0, 1) distributed

Rejection of null hypothesis at test level α (e.g. 0.05) if

$$Z > \Phi_{1-\alpha/2} \vee Z < \Phi_{\alpha/2}$$



Wald Test

for testing H_0: $\theta = \theta_0$ vs. H_1: $\theta \neq \theta_0$ use the test variable $W = \frac{\hat{\theta} - \theta_0}{se(\hat{\theta})}$

with sample estimate $\hat{\theta}$ and standard error $se(\hat{\theta}) = \sqrt{Var[\hat{\theta}]}$

W converges in probability to N(0,1)

$\rightarrow$ reject H_0 at level α when $W > \Phi_{1-\alpha/2}$ or $W < \Phi_{\alpha/2}$

the p-value for the Wald test is $2\Phi(-|w|)$
where w is the value of the test variable W

generalization (for unknown variance):
t-test (based on Student‘s t distribution)

Example: Wald Test

n=20 coin tosses $X_1, \ldots, X_n$ with 15 times heads

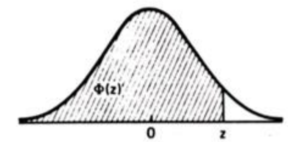
H_0: p=0.5 (coin is fair) vs. $_{H1}$: $p \neq 0.5$

not variance, but sample variance

sample mean: $\hat{p} = 0.75$, variance $\mathrm{Var}[\hat{p}] = n\,\hat{p}\,(1-\hat{p}) \,/\, n^2 = \frac{3}{320}$

Test statistic $W = \frac{\hat{p}-p}{se(p)} \approx \frac{0.25}{\sqrt{1/100}} \approx 2.5$

The Normal Distribution Functions $\Phi(z) = \int_{-\infty}^{z} \frac{e^{-t^2/2}}{\sqrt{2\pi}}\,dt$



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188	.53586
0.1	.53983	.54380	.54776	.55172	.55567	.55962	.56356	.56749	.57142	.57535
0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026	.61409
0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803	.65173
0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082	.68439	.68793
0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904	.72240
0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175	.75490
0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230	.78524
0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785	.81057	.81327
0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398	.83646	.83891
1.0	.84134	.84375	.84614	.84849	.85083	.85314	.85543	.85769	.85993	.86214
1.1	.86433	.86650	.86864	.87076	.87286	.87493	.87698	.87900	.88100	.88298
1.2	.88493	.88686	.88877	.89065	.89251	.89435	.89617	.89796	.89973	.90147
1.3	.90320	.90490	.90658	.90824	.90988	.91149	.91308	.91466	.91621	.91774
1.4	.91924	.92073	.92220	.92364	.92507	.92647	.92785	.92922	.93056	.93189
1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062	.94179	.94295	.94408
1.6	.94520	.94630	.94738	.94845	.94950	.95053	.95154	.95254	.95352	.95449
1.7	.95543	.95637	.95728	.95818	.95907	.95994	.96080	.96164	.96246	.96327
1.8	.96407	.96485	.96562	.96638	.96712	.96784	.96856	.96926	.96995	.97062
1.9	.97128	.97193	.97257	.97320	.97381	.97441	.97500	.97558	.97615	.97670
2.0	.97725	.97778	.97831	.97882	.97932	.97982	.98030	.98077	.98124	.98169
2.1	.98214	.98257	.98300	.98341	.98382	.98422	.98461	.98500	.98537	.98574
2.2	.98610	.98645	.98679	.98713	.98745	.98778	.98809	.98840	.98870	.98899
2.3	.98928	.98956	.98983	.99010	.99036	.99061	.99086	.99111	.99134	.99158
2.4	.99180	.99202	.99224	.99245	.99266	.99286	.99305	.99324	.99343	.99361
2.5	.99379	.99396	.99413	.99430	.99446	.99461	.99477	.99492	.99506	.99520
2.6	.99534	.99547	.99560	.99573	.99585	.99598	.99609	.99621	.99632	.99643
2.7	.99653	.99664	.99674	.99683	.99693	.99702	.99711	.99720	.99728	.99736
2.8	.99744	.99752	.99760	.99767	.99774	.99781	.99788	.99795	.99801	.99807
2.9	.99813	.99819	.99825	.99831	.99836	.99841	.99846	.99851	.99856	.99861
3.0	.99865	.99869	.99874	.99878	.99882	.99886	.99889	.99893	.99896	.99900
3.1	.99903	.99906	.99910	.99913	.99916	.99918	.99921	.99924	.99926	.99929
3.2	.99931	.99934	.99936	.99938	.99940	.99942	.99944	.99946	.99948	.99950
3.3	.99952	.99953	.99955	.99957	.99958	.99960	.99961	.99962	.99964	.99965
3.4	.99966	.99968	.99969	.99970	.99971	.99972	.99973	.99974	.99975	.99976
3.5	.99977	.99978	.99978	.99979	.99980	.99981	.99981	.99982	.99983	.99983
3.6	.99984	.99985	.99985	.99986	.99986	.99987	.99987	.99988	.99988	.99989
3.7	.99989	.99990	.99990	.99990	.99991	.99991	.99992	.99992	.99992	.99992
3.8	.99993	.99993	.99993	.99994	.99994	.99994	.99994	.99995	.99995	.99995

Test level $\alpha=0.1$:

$W > \Phi_{1-\alpha/2} = \Phi_{0.95}$ or $W < \Phi_{\alpha/2} = \Phi_{0.05}$

Test: $2.5 > 1.65 \;\; \rightarrow$ reject H0

Test level $\alpha=0.01$:

$W > \Phi_{1-\alpha/2} = \Phi_{0.995}$ or $W < \Phi_{\alpha/2} = \Phi_{0.005}$

Test: $2.5 < 2.58 \;\; \rightarrow$ retain H0

p-value in between

t-Test

Given: n samples for θ with sample mean $\hat{\theta}$
and ***sample standard deviation*** $\boldsymbol{S(\hat{\theta})}$

for testing H_0: θ = θ_0 vs. H_1: θ ≠ θ_0 use the test variable $T = \frac{\hat{\theta} - \theta_0}{se(\hat{\theta})}$

with sample estimate $\hat{\theta}$ and standard error $se(\hat{\theta}) = \sqrt{S^2(\hat{\theta})}$

T converges in probability to a ***t-distribution with n-1 degrees***

→ reject H_0 at level α when T > $t_{n-1,1-\alpha/2}$ or T < $t_{n-1,\alpha/2}$

Extensions for
- two-sample tests – comparing two independent samples
- paired two-sample tests – for testing differences (ordering) of RVs

t-test is most widely used test for statistical significance of experimental data

Paired t-Test Tools

Usable∑Stats
because statistics shouldn't be greek

https://www.usablestats.com/calcs/2samplet

Home | Products | Tutorials | Calculators

2 Sample t-test Calculator

View all Calculators

Test the mean difference between two samples of continuous data using the 2-sample t-test. The calculator uses the probabilities from the student t distribution.
For all t-tests see the easyT Excel Calculator : : Sample data is available.

Fore more information on 2-Sample t-tests View the Comparing Two Means: 2 Sample t-test tutorial

use software like Matlab, R, etc.

Data

Enter Summarized Data

Sample 1	Sample 2
10	7
12	6
15	11
5	13
12	8
13	8
8	5
9	8
10	9
10	11

Submit

Descriptive Statistics

	N	Mean	StDev	SE Mean
Sample 1	10	10.4	2.7968	0.884
Sample 2	10	8.6	2.4585	0.777

2 Sample t Tutorial

Observed difference (Sample 1 - Sample 2): 1.8
Standard Deviation of Difference : 1.1776

Unequal Variances
DF : 17
95% Confidence Interval for the Difference (-0.6845 , 4.2845)
T-Value 1.5285
Population 1 ≠ Population 2: P-Value = 0.1448
Population 1 > Population 2: P-Value = 0.9276
Population 1 < Population 2: P-Value = 0.0724

Equal Variances
Pooled Standard Deviation: 2.6331
Pooled DF: 18
95% Confidence Interval for the Difference (-0.674 , 4.274)
T-Value 1.5286
Population 1 ≠ Population 2: P-Value = 0.1438
Population 1 > Population 2: P-Value = 0.9281
Population 1 < Population 2: P-Value = 0.0719

Download the 2-Sample t-test Excel Calculator

Chi-Square Distribution

Let $X_1, \ldots, X_n$ be independent, N(0,1) distributed random variables.

Then the random variable $\chi_n^2 := X_1^2 + \ldots + X_n^2$

is chi-square distributed with n degrees of freedom:

$$f_{\chi_n^2}(x) = \frac{x^{\frac{n-2}{2}} e^{-\frac{x}{2}}}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} \ for \ x > 0, \ 0 \ otherwise$$

Let n be a natural number, let X be N(0,1) distributed and

Y χ^2 distributed with n degrees of freedom.

Then the random variable $T_n := \sqrt{n} \frac{X}{\sqrt{Y}}$

is t distributed with n degrees of freedom.

Chi-Square Goodness-of-Fit-Test

Given:

n sample values $X_1, ..., X_n$ of random variable X
with absolute frequencies $H_1, ..., H_k$ for k value classes v_i
(e.g. value intervals) of random variable X

<u>Null hypothesis:</u>

the values X_i are f distributed (e.g. uniformly distributed),
where f has expectation μ and variance σ^2

Approach: $Y_k := \sum_{i=1}^{k} (H_i - E(v_i)) \sqrt{n} / \sigma$ and $Z_k := \sum_{i=1}^{k} \frac{(H_i - E(v_i))^2}{E(v_i)}$

with $E(v_i) := n\, P[X \text{ is in class } v_i \text{ according to } f\,]$

are both approximately $\chi 2$ distributed with k-1 degrees of freedom

Rejection of null hypothesis at test level α (e.g. 0.05) if $Z_k > \chi^2_{k-1,1-\alpha}$

Chi-Square Independence Test

Given:

n samples of two random variables X, Y or, equivalently,
a twodimensional random variable
with absolute frequencies $H_{11}, \ldots, H_{rc}$ for $r \times c$ value classes,
where X has r and Y has c distinct classes.
(This is called a ***contingency table***.)

Null hypothesis:

X und Y are independent; then the
expectations for the absolute frequencies of the value classes would be

$$E_{ij} = \frac{R_i C_j}{n} \quad \text{with } R_i := \sum_{j=1}^{c} H_{ij} \text{ and } \quad C_j := \sum_{i=1}^{r} H_{ij}$$

Approach: $Z := \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(H_{ij} - E_{ij})^2}{E_{ij}}$ is approximately χ^2 distributed with (r-1)(c-1) degrees of freedom

Rejection of null hypothesis at test level α (e.g. 0.05) if

$$Z > \chi^2_{(r-1)(c-1),1-\alpha}$$

Example: Chi-Square Independence Test

women and men seem to prefer different study subjects
$\rightarrow$ we compiled enrollment data in a **contingency table**

Gender / *Subject*	*Male*	*Female*	*Total*
CS	*80*	*20*	*100*
Math	*40*	*20*	*60*
Bioinf	*20*	*20*	*40*
Total	*140*	*60*	*200*

Hypothesis H_0: Gender and Subject are independent

Test statistic $Z = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(H_{ij} - E_{ij})^2}{E_{ij}} \sim \chi^2((r-1)(c-1)) \sim \chi 2\ (2)$

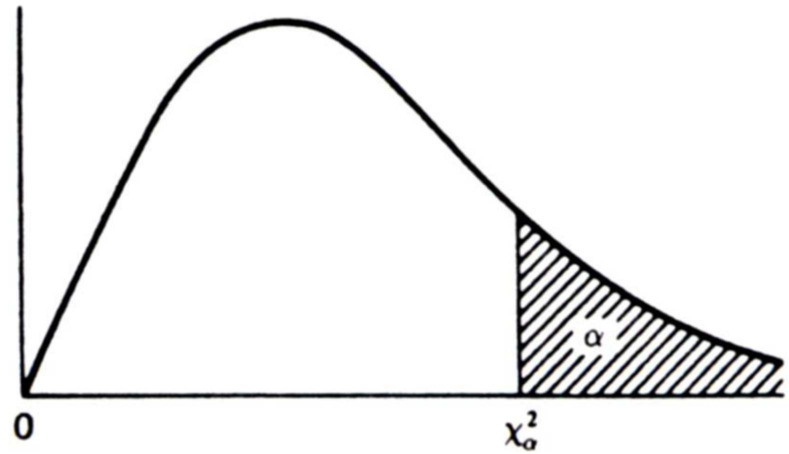
$$Z = \frac{10^2}{70} + \frac{(-10)^2}{30} + \frac{(-2)^2}{42} + \frac{2^2}{18} + \frac{(-8)^2}{28} + \frac{8^2}{12} \approx 12.6$$

Test level $1-\alpha=0.95 \rightarrow \chi^2_{2,0.95} \approx 5.99 \rightarrow$ reject H_0

Chi-Square Distribution Table

Table 4

Critical Values of the Chi-Square Distribution[a]



n^b \ α	0.995	0.990	0.975	0.950	0.05	0.025	0.010	0.005
1	**0.0^4393**[c]	**0.0^3157**[c]	**0.0^3982**[c]	**0.0^2393**[c]	3.8415	5.0239	6.6349	7.8794
2	0.0100	0.0201	0.0506	0.1026	5.9915	7.3778	9.2103	10.597
3	0.0717	0.1148	0.2158	0.3518	7.8147	9.3484	11.345	12.838
4	0.2070	0.2971	0.4844	0.7107	9.4877	11.143	13.277	14.860
5	0.4117	0.5543	0.8312	1.1455	11.071	12.833	15.086	16.750
6	0.6757	0.8721	1.2373	1.6354	12.592	14.449	16.812	18.548
7	0.9893	1.2390	1.6899	2.1674	14.067	16.013	18.475	20.278
8	1.3444	1.6465	2.1797	2.7326	15.507	17.535	20.090	21.955
9	1.7350	2.0879	2.7004	3.3251	16.920	19.023	21.666	23.589
10	2.1559	2.5582	3.2470	3.9403	18.307	20.483	23.209	25.188
11	2.6032	3.0535	3.8158	4.5748	19.675	21.920	24.725	26.757
12	3.0738	3.5706	4.4038	5.2260	21.026	23.337	26.217	28.300
13	3.5650	4.1069	5.0087	5.8919	22.362	24.736	27.688	29.819

Chi-Square Distribution Table

1.1.2.10. **Obere 100α-prozentige Werte χ^2_α der χ^2-Verteilung** (s. 5.2.3.)

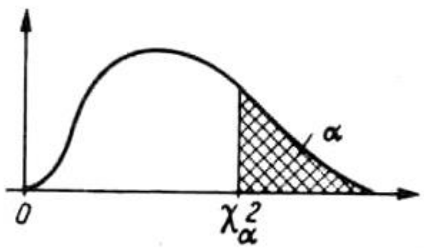


Abb. 1.4

Anzahl der Freiheits-grade m	Wahrscheinlichkeit $p = \alpha$															
	0,99	0,98	0,95	0,90	0,80	0,70	0,50	0,30	0,20	0,10	0,05	0,02	0,01	0,005	0,002	0,001
1	0,00016	0,0006	0,0039	0,016	0,064	0,148	0,455	1,07	1,64	2,7	3,8	5,4	6,6	7,9	9,5	10,83
2	0,020	0,040	0,103	0,211	0,446	0,713	1,386	2,41	3,22	4,6	6,0	7,8	9,2	10,6	12,4	13,8
3	0,115	0,185	0,352	0,584	1,005	1,424	2,366	3,67	4,64	6,3	7,8	9,8	11,3	12,8	14,8	16,3
4	0,30	0,43	0,71	1,06	1,65	2,19	3,36	4,9	6,0	7,8	9,5	11,7	13,3	14,9	16,9	18,5
5	0,55	0,75	1,14	1,61	2,34	3,00	4,35	6,1	7,3	9,2	11,1	13,4	15,1	16,8	18,9	20,5
6	0,87	1,13	1,63	2,20	3,07	3,83	5,35	7,2	8,6	10,6	12,6	15,0	16,8	18,5	20,7	22,5
7	1,24	1,56	2,17	2,83	3,82	4,67	6,35	8,4	9,8	12,0	14,1	16,6	18,5	20,3	22,6	24,3
8	1,65	2,03	2,73	3,49	4,59	5,53	7,34	9,5	11,0	13,4	15,5	18,2	20,1	22,0	24,3	26,1
9	2,09	2,53	3,32	4,17	5,38	6,39	8,34	10,7	12,2	14,7	16,9	19,7	21,7	23,6	26,1	27,9
10	2,56	3,06	3,94	4,86	6,18	7,27	9,34	11,8	13,4	16,0	18,3	21,2	23,2	25,2	27,7	29,6
11	3,1	3,6	4,6	5,6	7,0	8,1	10,3	12,9	14,6	17,3	19,7	22,6	24,7	26,8	29,4	31,3
12	3,6	4,2	5,2	6,3	7,8	9,0	11,3	14,0	15,8	18,5	21,0	24,1	26,2	28,3	30,9	32,9
13	4,1	4,8	5,9	7,0	8,6	9,9	12,3	15,1	17,0	19,8	22,4	25,5	27,7	29,8	32,5	34,5
14	4,7	5,4	6,6	7,8	9,5	10,8	13,3	16,2	18,2	21,1	23,7	26,9	29,1	31,3	34,0	36,1
15	5,2	6,0	7,3	8,5	10,3	11,7	14,3	17,3	19,3	22,3	25,0	28,3	30,6	32,8	35,6	37,7
16	5,8	6,6	8,0	9,3	11,2	12,6	15,3	18,4	20,5	23,5	26,3	29,6	32,0	34,3	37,1	39,3
17	6,4	7,3	8,7	10,1	12,0	13,5	16,3	19,5	21,6	24,8	27,6	31,0	33,4	35,7	38,6	40,8
18	7,0	7,9	9,4	10,9	12,9	14,4	17,3	20,6	22,8	26,0	28,9	32,3	34,8	37,2	40,1	42,3
19	7,6	8,6	10,1	11,7	13,7	15,4	18,3	21,7	23,9	27,2	30,1	33,7	36,2	38,6	41,6	43,8
20	8,3	9,2	10,9	12,4	14,6	16,3	19,3	22,8	25,0	28,4	31,4	35,0	37,6	40,0	43,0	45,3
21	8,9	9,9	11,6	13,2	15,4	17,2	20,3	23,9	26,2	29,6	32,7	36,3	38,9	41,4	44,5	46,8
22	9,5	10,6	12,3	14,0	16,3	18,1	21,3	24,9	27,3	30,8	33,9	37,7	40,3	42,8	45,9	48,3
23	10,2	11,3	13,1	14,8	17,2	19,0	22,3	26,0	28,4	32,0	35,2	39,0	41,6	44,2	47,3	49,7
24	10,9	12,0	13,8	15,7	18,1	19,9	23,3	27,1	29,6	33,2	36,4	40,3	43,0	45,6	48,7	51,2
25	11,5	12,7	14,6	16,5	18,9	20,9	24,3	28,2	30,7	34,4	37,7	41,6	44,3	46,9	50,1	52,6
26	12,2	13,4	15,4	17,3	19,8	21,8	25,3	29,2	31,8	35,6	38,9	42,9	45,6	48,3	51,6	54,1
27	12,9	14,1	16,2	18,1	20,7	22,7	26,3	30,3	32,9	36,7	40,1	44,1	47,0	49,6	52,9	[illegible]5,5
28	13,6	14,8	16,9	18,9	21,6	23,6	27,3	31,4	34,0	37,9	41,3	45,4	48,3	51,0	54,4	56,9

3.2.4 Regression for Parameter Fitting
Linear Regression

Estimate $r(x) = E[Y \mid X_1=x_1 \wedge \ldots \wedge X_m=x_m]$ using a linear model

$Y = r(x) + \varepsilon = \beta_0 + \sum_{i=1}^{m} \beta_i x_i + \varepsilon$ with error ε with $E[\varepsilon]=0$

given n sample points $(x_1^{(i)}, \ldots, x_m^{(i)}, y^{(i)})$, i=1..n, the least-squares estimator (LSE) minimizes the quadratic error:

$$\sum_{i=1..n} \left(\left(\sum_{k=0..m} \beta_k x_k^{(i)} \right) - y^{(i)} \right)^2 =: E(\beta_0, \ldots, \beta_m) \qquad (\text{with } x_o^{(i)}=1)$$

Solve linear equation system: $\frac{\partial E}{\partial \beta_k} = 0$ for k=0, ..., m

equivalent to MLE $\vec{\beta} = (X^T X)^{-1} X^T Y$

with $Y = (y^{(1)} \ldots y^{(n)})^T$ and $X = \begin{pmatrix} 1 \; x_1^{(1)} \; x_2^{(1)} \; \ldots \; x_m^{(1)} \\ 1 \; x_1^{(2)} \; x_2^{(2)} \; \ldots \; x_m^{(2)} \\ \ldots \\ 1 \; x_1^{(n)} \; x_2^{(n)} \; \ldots \; x_m^{(n)} \end{pmatrix}$

Logistic Regression

Estimate r(x) = E[Y | X=x] for Bernoulli Y using a logistic model

$$Y = r(x) + \varepsilon = \frac{e^{\beta_0 + \sum_{i=1}^{m} \beta_i x_i}}{1 + e^{\beta_0 + \sum_{i=1}^{m} \beta_i x_i}} + \varepsilon$$ loglinear

with error ε with E[ε]=0

→ solution for MLE for β_i values
based on numerical gradient-descent methods

Summary of Section 3.2

- **Samples** and **Estimators** are RVs
- Estimators should be **unbiased**
- **MLE** is canonical estimator for parameters
- **Confidence intervals** based on Normal and t distributions
- **Hypothesis testing**: reject or retain H_0 at level α
- **p-value**: smallest level α for rejecting H_0
- **Wald test** and **t-test** for (in)equality of parameters
- **Chi-Square test** for independence or goodness-of-fit
- Linear **regression** for predicting continuous variables

Additional Literature for Section 3.2

- A. Allen: Probability, Statistics, and Queueing Theory With Computer Science Applications, Wiley 1978
- G. Casella, R. Berger: Statistical Inference, Duxbury 2002
- M. Greiner, G. Tinhofer: Stochastik für Studienanfänger der Informatik, Carl Hanser Verlag, 1996
- G. Hübner: Stochastik: Eine Anwendungsorientierte Einführung für Informatiker, Ingenieure und Mathematiker, Vieweg & Teubner 2009