Hypothesis tests

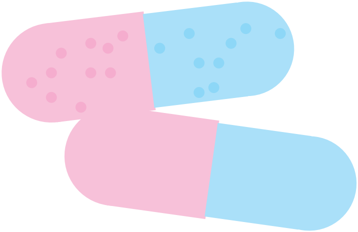
Jana de Wiljes

jana@dewiljes.de

www.dewiljes-lab.com

28. November 2022

Clinical trial: test efficacy a new drugs



Aspekte:

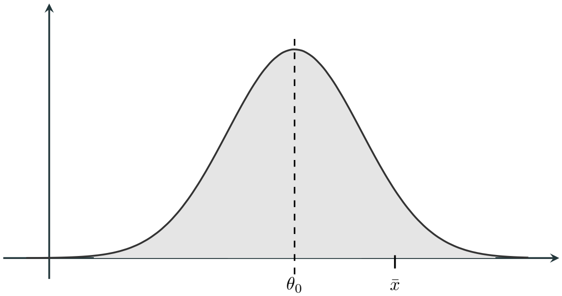
- Given Score-Function mapping to $(-\infty, \infty)$ acting as a measure for the efficacy of new drug
- Know mean efficacy θ_0 of placebos
- Want to know, if average efficacy of new drug is going being placebo effect
- Given samples $x_1, \ldots, x_n$; need decision tool

Parametric Hypotheses Tests

Consider: Family of densities $\{f(x|\theta) : \theta \in \Theta\}$

Definition: A hypothesis is a statement about a population parameter θ.

Example: Consider Gaußdistirbution $\mathcal{N}(\theta, 1)$; hypothesis H_0: efficacy of new drug has expected value θ_0

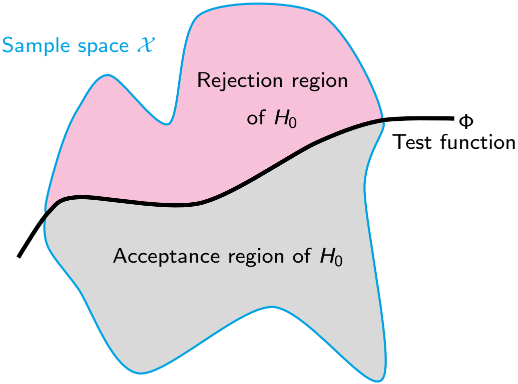


Hypothesis test

Problem setting:

$$\begin{cases} H_0: & \theta \in \Theta_0 \quad \text{Null hypothesis} \\ H_1: & \theta \in \Theta_1 \quad \text{Alternativ hypothesis} \end{cases}$$

where $\Theta_0 \cap \Theta_1 = \emptyset$



Likelihood Ratio Tests

Def: The likelihood ratio test (LRT) statistic for testing

$$H_0 : \theta \in \Theta_0 \text{ gegen } H_1 : \theta \in \Theta_1$$

is

$$\lambda(x_1, \ldots, x_n) = \frac{\sup_{\Theta_0} L(\theta|x_1, \ldots, x_n)}{\sup_{\Theta} L(\theta|x_1, \ldots, x_n)}.$$

A likelihood ratio test (LRT) is any test that has a rejection region of the form

$$\{(x_1, \ldots, x_n) \in \mathcal{X} : \lambda(x_1, \ldots, x_n) \leq c\},$$

where c is any number satisfying $0 \leq c \leq 1$.

Example: Normal LRT

Example: For a given set of iid samples $x_1, \ldots, x_n \sim \mathcal{N}(\theta, 1)$

$$\lambda(x_1, \ldots, x_n) = \frac{(2\pi)^{-n/2} \exp(-\sum_{i=1}^{n}(x_i - \theta_0)^2/2)}{(2\pi)^{-n/2} \exp(-\sum_{i=1}^{n}(x_i - \bar{x})^2/2)} \tag{1}$$

$$= \exp\left[\Big(-\sum_{i=1}^{n}(x_i - \theta_0)^2 + \sum_{i=1}^{n}(x_i - \bar{x})^2\Big)/2\right] \tag{2}$$

$$\begin{aligned}
\sum_{i=1}^{n}(x_i - \theta_0)^2 &= \sum_{i=1}^{n}(x_i - \bar{x} + \bar{x} - \theta_0)^2 \\
&= \sum_{i=1}^{n}(x_i - \bar{x})^2 + 2(\bar{x} - \theta_0)\sum_{i=1}^{n}(x_i - \bar{x}) + \sum_{i=1}^{n}(\bar{x} - \theta_0)^2 \\
&= \sum_{i=1}^{n}(x_i - \bar{x})^2 + 2(\bar{x} - \theta_0)\underbrace{\Big(\sum_{i=1}^{n}(x_i) - n\bar{x}\Big)}_{=0} + \sum_{i=1}^{n}(\bar{x} - \theta_0)^2 \\
&= \sum_{i=1}^{n}(x_i - \bar{x})^2 + n(\bar{x} - \theta_0)^2
\end{aligned}$$

Example: Normal LRT

Inserting

$$\sum_{i=1}^{n}(x_i - \theta_0)^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2 + n(\bar{x} - \theta_0)^2$$

in

$$\begin{aligned}\lambda(x_1, \dots, x_n) &= \exp\left[\Big(-\sum_{i=1}^{n}(x_i - \theta_0)^2 + \sum_{i=1}^{n}(x_i - \bar{x})^2\Big)/2\right] \\ &= \exp\left[\Big(-n(\bar{x} - \theta_0)^2\Big)/2\right]\end{aligned}$$

Ansatz: An LRT test rejects H_0 for small values of $\lambda(x_1, \dots, x_n)$. Using the rejection region

$$\{x_1, \dots, x_n : \lambda(x) \geq c\} = \{x_1, \dots, x_n : |\bar{x} - \theta_0| \geq \sqrt{-2(\log c)/n}\} \tag{3}$$

Error Probabilities

Truth \ Decision	Acceptance H_0	Rejection H_0
H_0	correct decision	Type-I-Error
H_1	Type-II-Error	correct decision

Note that

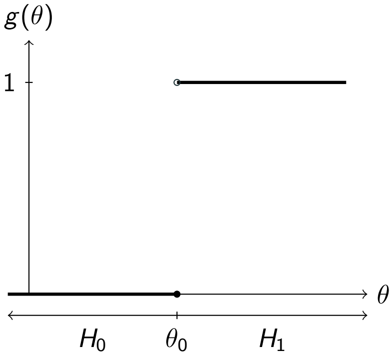
$$\mathbb{P}((X_1, \ldots, X_n) \in R) = \begin{cases} \text{probability of a Type I Error} & \text{if } \theta \in \Theta_0 \\ \text{1 - the probability of a Type II Error} & \text{if } \theta \in \Theta_0^c \end{cases}$$

Power function

Def: The power function of a hypothesis test with rejection region R is the function of θ defined by

$$g(\theta) = \mathbb{P}_\theta((X_1, \ldots, X_n) \in R)$$

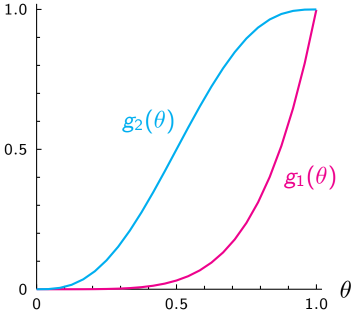
.



Example

Let $X \sim binomial(5, \theta)$

$H_0 : \theta \leq \frac{1}{2}$ versus $H_1 : \theta > \frac{1}{2}$



Zwei Beispieltests:

- Consider test that rejects H_0 if and only if all *successes* are observed. The power function for this test is $g_1(\theta) = \mathbb{P}_\theta(X \in R) = \mathbb{P}_\theta(X = 5) = \theta^5$
- Consider test that rejects H_0 if $X = 3, 4$ or 5 are observed. The power function for this test is

$$g_2(\theta) = \mathbb{P}_\theta(X = 3, 4, 5) = \binom{5}{3}\theta^3(1-\theta)^2 + \binom{5}{4}\theta^4(1-\theta)^1 + \binom{5}{5}\theta^5(1-\theta)^0$$

Significance level α

Ansatz: Trying to fix error of Typ-I first

Def: For $0 \leq \alpha \leq 1$, a test with power function $\beta(\theta)$ is a level α test if

$$\sup_{\theta \in \Theta_0} g(\theta) \leq \alpha.$$

Idea for construction of tests with significance level α :

- Set level α (in applications typical values: $\alpha \in \{0{,}05; 0{,}01; 0{,}001\}$)
- Choose c of LRTs, so that the test a level α test is, i.e., determine c with

$$\sup_{\theta \in \Theta_0} \mathbb{P}_\theta(\lambda(X_1, \ldots, X_n) \leq c) = \alpha$$

Example

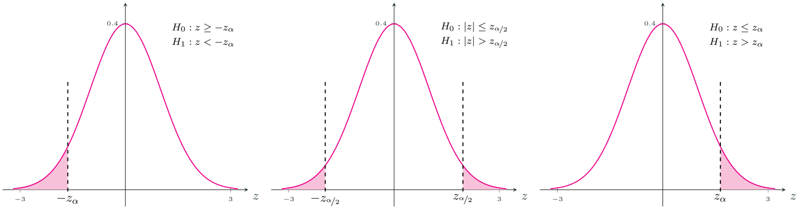
Let $X_1, \ldots, X_n$ be a RVs distributed according to $\mathcal{N}(\theta, \sigma^2)$ with known σ^2. An LRT of $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$ is a test that rejects H_0 if $(\bar{X} - \theta_0)/(\sigma/\sqrt{n}) > c$. The constant c can be any positive number. The power function of this is

$$
\begin{aligned}
g(\theta) &= \mathbb{P}_\theta\Big(\frac{\bar{X} - \theta_0}{\sigma/\sqrt{n}} > c\Big) \\
&= \mathbb{P}_\theta\Big(\frac{\bar{X} - \theta}{\sigma/\sqrt{n}} > c + \frac{\theta_0 - \theta}{\sigma/\sqrt{n}}\Big) \\
&= \mathbb{P}\Big(Z > c + \frac{\theta_0 - \theta}{\sigma/\sqrt{n}}\Big)
\end{aligned}
$$

where Z is a standard normal random variable, since $\frac{\bar{X} - \theta_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$
In general: Construct size α test by choosing c such that we obtain α, e.g. for LRTs

$$
\sup_{\theta \in \Theta} \mathbb{P}_\theta(\lambda(X_1, \ldots, X_n) \leq c) = \alpha \tag{4}
$$

Z-Test/Gaußtest



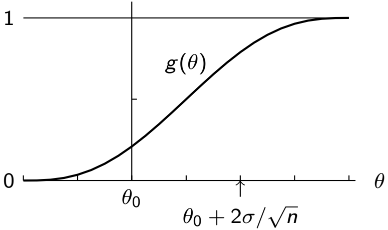
Example: $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ with $\sigma = 1$

- $\theta = \theta_0$, since $\Theta_0 = \{\theta_0\}$ und $Z = \sqrt{n}(\bar{X} - \theta_0) \sim \mathcal{N}(0, 1)$
- the test rejects H_0, if $\sqrt{n}|\bar{X} - \theta_0| \geq z_{\alpha/2}$
- for $z_{\alpha/2}$ holds $\mathbb{P}(Z > z_{\alpha/2}) = \alpha/2$ with $Z \sim \mathcal{N}(0, 1)$

New drug: $\bar{x} = 3{,}7; \theta_0 = 3{,}1; n = 16; \alpha = 0{,}05; z_{\alpha/2} \approx 1{,}96$

$$\sqrt{n} \cdot |\bar{x} - \theta_0| = \sqrt{16} \cdot |3{,}7 - 3{,}1| = 4 \cdot 0{,}6 = 2{,}4 > z_{\alpha/2}$$

Typ-II-Error



Consider: $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$ mit $g(\theta) = \mathbb{P}\Big(Z > c' + \frac{\theta_0 - \theta}{\sigma/\sqrt{n}}\Big)$
Goal:

- Typ-I-Error not larger than $\alpha = 0{,}1$, choose $c' = 1{,}28$ and obtain $g(\theta_0) = \mathbb{P}\Big(Z > 1{,}28\Big) = 0.1$
- Typ-II-Error not larger than 0,2 for $\theta \geq \theta_0 + \sigma$

$$g(\theta_0 + \sigma) = \mathbb{P}\Big(Z > \underbrace{c' - \sqrt{n}}_{-0{,}84}\Big) = 0{,}8.$$

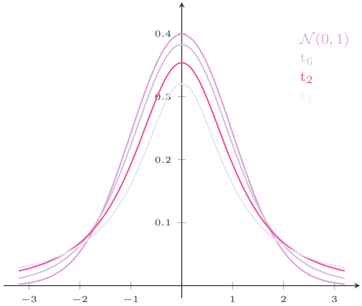
For $n = 4{,}49$ the desired value can be obtained.

Student t-distribution

Def: A continuous random variable X with density

$$f(x) = \frac{\Gamma(n+1)/2}{\sqrt{n\pi}\Gamma(n/2)(1+x^2/n)^{(n+1)/2}} \tag{5}$$

is called t-distributed with n degrees of freedom, abbreviated by the notation $t \sim t_n$ where $\Gamma(x) = \int\limits_0^{+\infty} t^{x-1} e^{-t} \, \mathrm{d}\, t$.



t-Test

Lemma: Let $X_1, \ldots, X_n$ be iid RVs with $X_i \sim \mathcal{N}(\mu, \sigma^2)$, then

$$\frac{\bar{X} - \mu}{S} \sqrt{n} \sim t_{n-1},$$

where

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

Example: $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$, where σ unknown,

- it holds that $\sqrt{n}(\bar{X} - \theta_0)/\sqrt{S^2} \sim t_{n-1}$
- the test rejects H_0, if $|\bar{X} - \theta_0|/\sqrt{S^2} \geq t_{n-1,\alpha/2}/\sqrt{n}$
- for $t_{n-1,\alpha/2}$ holds $\mathbb{P}(T_{n-1} \geq t_{n-1,\alpha/2}) = \alpha$ with $T_{n-1} \sim t_{n-1}$

Unbiased Tests

Want: a test to be more likely rejecting H_0 if $\theta \in \Theta_0^c$ than if $\theta \in \Theta_0$

Def: A test with power function $\beta(\theta)$ is unbiased if $\beta(\theta') \leq \beta(\theta'')$ for every $\theta' \in \Theta_0^c$ and $\theta'' \in \Theta_0$.

Example: Consider LRT with Gauß-RV for $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$ with power function

$$g(\theta) = \mathbb{P}\left(Z > c' + \frac{\theta_0 - \theta}{\sigma/\sqrt{n}}\right).$$

Since $g(\theta)$ is increasing with θ the following holds for fixed θ_0:

$$g(\theta) > g(\theta_0) = \max_{t \leq \theta_0} g(t) \quad \text{für alle } \theta > \theta_0$$

and therefore the test unbiased.

Uniformly most powerful

Def: Let $\mathcal{C}$ be class of tests for testing $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_0^c$. A test in class $\mathcal{C}$, with power function $g(\theta)$, is a uniformly most powerful (UMP) class $\mathcal{C}$ test if $g(\theta) \geq \beta'(\theta)$ for every $\theta \in \Theta_0^c$ and every $\beta'(\theta)$ that is a power function of a test in class $\mathcal{C}$.

Neyman-Pearson Lemma

Lemma: Consider testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$, where the pdf or pmf corresponding to θ_0 and θ_1 are $f(x_1, \ldots, x_n|\theta_0)$ and $f(x_1, \ldots, x_n|\theta_1)$ using a test with rejection region R that satisfies

$$x_1, \ldots, x_n \in R \text{ if } f(x_1, \ldots, x_n|\theta_1) > kf(x_1, \ldots, x_n|\theta_0) \tag{6}$$

and

$$x_1, \ldots, x_n \in R^c \text{ if } f(x_1, \ldots, x_n|\theta_1) < kf(x_1, \ldots, x_n|\theta_0) \tag{7}$$

for some $k \geq 0$ and $\alpha = \mathbb{P}_{\theta_0}((X_1, \ldots, X_n) \in R)$. Then

1. Any test that satisfies the above conditions is an UMP level α test.
2. If there exists a test satisfying the above conditions with $k > 0$ then every UMP level α test is a size α test and every UMP level test satisfies expect perhaps on a set A satisfying $\mathbb{P}_{\theta_0}((X_1, \ldots, X_n) \in A) = \mathbb{P}_{\theta_1}((X_1, \ldots, X_n) \in A) = 0$

χ^2-distribution

Definition: A continuous, non-negative random variable X with density

$$f(x) = \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} \exp\Big(-\frac{1}{2}x\Big), \quad x > 0 \tag{8}$$

is called χ^2-distributed with n degrees of freedom, abbreviated by the notation $X \sim \chi^2_n$.

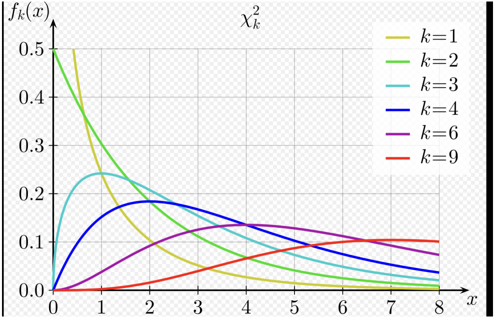
Lemma: Let $X \sim \chi^2_n$ be a continuous, non-negative random variable. Then its expectation and variance are given by:

- $\mathbb{E}[X] = n$
- $Var(X) = 2n$

Lemma: Let $X_1, ..., X_n$ be independent and identically standard normally distributed, then

$$Y_n = \sum_{i=1}^{n} X_i^2 \tag{9}$$

is χ^2- distributed with n degrees of freedom.



Some References

- **Original Paper:**

1. J. Neyman and E. S. Pearson (1933) *On the Problem of the Most Efficient Tests of Statistical Hypotheses*, Philosophical Transactions of the Royal Society of London. Series A, Vol. 231, pp. 289–337.

- **Introductions:**

1. D. R. Cox, D. V. Hinkley (1974) *Theoretical Statistics*, Springer US.
2. Van der Vaart, A. W. (2000) *Asymptotic Statistics*, Cambridge University Press.
3. George Casella und Roger L. Berger (2002) *Statistical Inference*, Cengage Learning.
4. Caputo, Fahrmeir, Künstler, Pigeot, Tutz (2011) *Statistik: Der Weg zur Datenanalyse*, Springer.