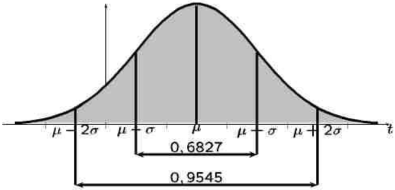
Statistics

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2}$$



Bachelor Studiengang Informatik

Prof. Dr. Egbert Falkenberg

Fachbereich Informatik & Ingenieurwissenschaften

Wintersemester 21/22

Section 1

Inferential Statistics

Introduction I

Example: Average height of all adults (over 18 years old) in the U.S.

- Population: all adults over 18 years of age in the U.S.
- Census: measure every adult and then compute the average $\rightarrow$ time-consuming and cost-intensive

Introduction II

- Using statistics:
  - Take a random sample and measure the heights
  - Conjecture that the average (**point estimation**) of the total population is “close to” the average of our sample
  - Calculate an interval containing the true value in for example 95% of all cases (**confidence interval**)
  - According to the Centers for Disease Control and Prevention Trusted Source [1], the average is 175.4 centimeters. Based on the sample value can we state that this value has been changed? (**hypothesis testing**)
- Goal of **inferential statistics**: use sample statistics to make inference about population parameters
- **Estimation** and **Hypothesis Testing** will be discussed in the following

[1] https://www.cdc.gov/nchs/data/nhsr/nhsr122-508.pdf, published in December 2018 based on data collected between 1999 and 2016

Characteristics of Estimators I

compare: Heumann, Schomaker 9.2

Notations: Let $x = \{x_1, x_2, ..., x_n\}$ be observations of a random sample from a population.

- Random sample $x = \{x_1, x_2, ..., x_n\}$ = realized values of a random variable X.
- More formally: x_i are realisiations of independent and identically distributed (i.i.d) random variables X_i.
- Statistic = function of random variables. For example: $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
- A statistic $T(X)$ is used to estimate a parameter ϑ.
- $T(X)$ is called an estimator of ϑ.
- $\hat{\vartheta} = T(X)$ denote the estimate of ϑ using $T(X)$.
- $T(X)$ is a random variable but $T(x)$ is its observed value calculated from the sample $x = (x_1, x_2, ..., x_n)$.

Characteristics of Estimators II

Important characteristics of estimators:

bias and sampling variability

- Bias refers to whether an estimator tends to either over or underestimate the parameter.
- Sampling variability refers to how much the estimate varies from sample to sample.
- An estimator is biased if the long-term average value of the statistic is not the parameter it is estimating.
- More technically: biased if the expected value is not equal to the parameter to be estimated.

Example: A stop watch that is a little bit fast gives biased estimates of elapsed time.

Characteristics of Estimators III

Definition: An estimator T(X) is unbiased if

$$E_{\vartheta}(T(X)) = \vartheta$$

The bias of an estimator $T(X)$ is defined as

$$Bias_{\vartheta}(T(X)) = E_{\vartheta}(T(X)) - \vartheta.$$

Remark: The index ϑ denotes that the expectated value is calculated with respect to the distribution whose parameter is ϑ.

Characteristics of Estimators IV

Remark: An unbiased estimator is not necessarily an accurate statistic.

- If a statistic is sometimes much too high and sometimes much too low, it can still be unbiased. It would be very imprecise, however.
- A slightly biased statistic that systematically results in very small overestimates of a parameter could be quite efficient.

Measure for the quality of an estimator: $E((T(X) - \vartheta)^2)$

Remark:

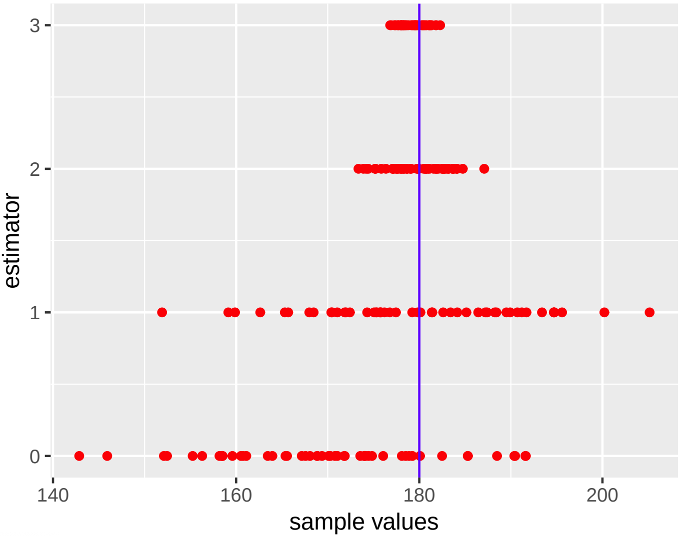
$$
\begin{aligned}
E_\vartheta((T(X) - \vartheta)^2) &= Var_\vartheta(T(X)) + (E_\vartheta(T(X)) - \vartheta)^2 \\
&= Var_\vartheta(T(X)) + Bias_\vartheta(T(X))^2
\end{aligned}
$$

$Var_\vartheta(T(X)) =$ Variance of the Estimator; a measure for the sampling variability

Characteristics of Estimators V

Example: different estimators for the paramter $\theta = 180$

estimator with different properties



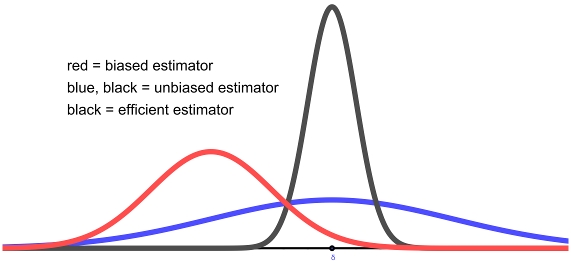
Characteristics of Estimators VI

Remark:

- Since an estimator is a random variable usually a new sample leads to a new estimate of θ.
- The estimators 0 and 3 are biased. The others are unbiased.
- The variability of estimator is 2 is much lower than the variability of estimator 1.
- Since the variability of estimator 3 is very low and the bias is not to high, it is the most accurate estimator of these 4 estimators.

Characteristics of Estimators VII

Objective: Find an unbiased estimator with smallest variance! Such an estimator is called efficient.



- Efficiency of a statistic decribes the precision of the estimate.
- The more efficient the statistic, the more precise the statistic is as an estimator of the parameter.

Characteristics of Estimators VIII

Consistency of Estimators: If the values of an estimator get closer to the parameter estimated if the sample size is increased, we call the estimator consistent.

Definition: Let $T_i = T_i(X_1, X_2, ..., X_i), i \in \mathbb{N}$ a sequence of estimators for the parameter ϑ. The sequence is a consistent sequence of estimators for ϑ if for every $\varepsilon > 0$

$$\lim_{n \to \infty} P(|T_n - \vartheta| < \varepsilon) = 1$$

As the sample size increases, the probability that T_n is getting closer to ϑ is approaching 1.

Characteristics of Estimators IX

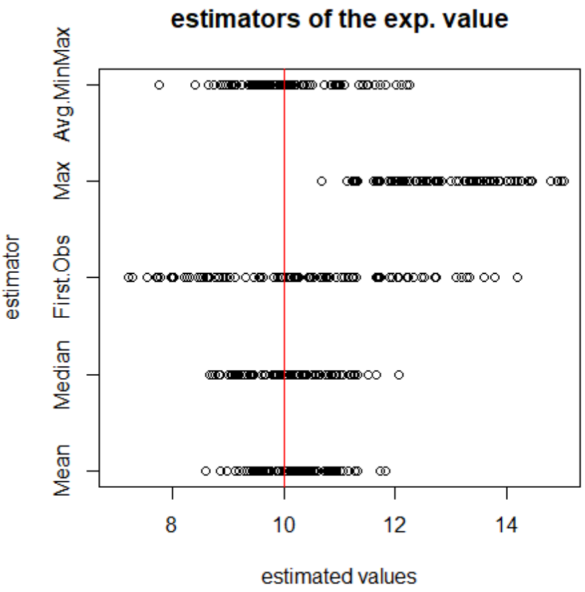
Example: A Bookseller operates a large number of stores. The expected monthly profit in 1000 Euro of a store is to be estimated. To do this, the monthly profit of 10 randomly selected stores is taken into account.

Observation of different estimators for expected monthly profit over the time: mean, median, first store, max value, average of the min and max value

The first 6 samples:

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	Mean	Median	First.Obs	Max	Avg.MinMax
1	13.51	9.92	15.15	10.27	13.94	10.73	10.73	12.54	13.14	10.42	12.03	11.64	13.51	15.15	12.53
2	14.05	14.00	12.61	12.43	9.10	11.78	10.35	11.73	8.37	10.23	11.47	11.75	14.05	14.05	11.21
3	10.38	9.67	11.91	9.74	9.60	9.01	10.74	11.42	8.39	7.43	9.83	9.70	10.38	11.91	9.67
4	9.04	9.67	11.44	14.03	8.21	11.00	10.77	12.19	8.02	11.87	10.62	10.89	9.04	14.03	11.03
5	11.90	11.64	10.76	10.75	11.13	11.70	9.77	11.13	12.31	9.03	11.01	11.13	11.90	12.31	10.67
6	7.84	8.40	10.13	10.49	8.98	8.63	7.96	7.02	9.48	11.33	9.03	8.81	7.84	11.33	9.18

Characteristics of Estimators IX



Questions

Which of the following statements are true or false?

t	f	
☐	☐	If you have estimated an unknown parameter in a sample it will not change in a new sample.
☐	☐	If an estimator deviates on average from the true value to be estimated, he has a bias.
☐	☐	The mean square deviation of an estimator from its true value is the bias of the estimator.
☐	☐	An efficient estimator is an unbiased estimator with smallest variance.
☐	☐	In the case of a consistent estimator, the estimates become with probability 1 more accurate as the sample size increases.

Maximum Likelihood Method I

Let $f(x;\theta)$ be the density function of $X_i, i = 1, 2, ..., n$.

$$L(x_1, x_2, ..., x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta)$$

density function of the random sample $X_1, X_2, ..., X_n$ is called **Likelihood-Function.**

Idea: Choose as an estimator $\hat{\theta}$ for θ the value which maximizes the Likelihood-Function, i.e.

$$L(x_1, x_2, ..., x_n; \theta) \leq L(x_1, x_2, ..., x_n; \hat{\theta}) \quad \text{for all } \theta$$

Remark: A maximum likelihood estimator is not necessary unbiased!

Maximum Likelihood Method II

Example: Urn with 10 black and white marbles. The number θ of black marbles is unknown.

1: 3 marbles are randomly drawn without replacement, X=2 marbles are black

θ	$P_\theta(X=2)$
0	0
1	0
2	0.067
3	0.175
4	0.4
5	0.417
6	0.5
7	**0.525**
8	0.467
9	0.3
10	0

$$L(x_1;\theta) = P_\theta(X=20) = \frac{\binom{\theta}{2}\binom{10-\theta}{1}}{\binom{10}{3}}$$

$$\hat{\theta} = 7$$

Maximum Likelihood Method III

2: The 3 drawn marbles are replaced and again 3 marbles are randomly drawn without replacement, X=0 marbles are black

θ	$P_\theta(X_2=0)$	$P_\theta(X_1=2)\cdot P_\theta(X_2=0)$
0	1	0
1	0.7	0
2	0.0467	0.031
3	0.292	**0.051**
4	0.167	0.050
5	0.083	0.035
6	0.033	0.017
7	0.008	0.004
8	0	0
9	0	0
10	0	0

$$
\begin{aligned}
L(x_1,x_2;\theta) &= P_\theta(X_1=2,X_2=0)=P_\theta(X_1=2)\cdot P_\theta(X_2=0)\\
&= \frac{\binom{\theta}{2}\binom{10-\theta}{1}}{\binom{10}{3}}\cdot\frac{\binom{\theta}{0}\binom{10-\theta}{3}}{\binom{10}{3}}\\
\hat{\theta} &= 3
\end{aligned}
$$

Maximum Likelihood Method IV

Example: A bus is coming exactly after θ minutes. But θ is unknown. You are coming at a random time to the bus stop. Let T be the waiting time to the bus. T is uniformly distributed on $[0, \theta]$, i.e. the density of T is

$$f(t;\theta) = \begin{cases} \frac{1}{\theta} & 0 \leq t \leq \theta \\ 0 & \text{else} \end{cases}$$

IF $T_1, T_2, ..., T_n$ are the waiting times at n days [2] the Likelihood-Function is

$$\begin{aligned} L(t_1, t_2, ..., t_n; \theta) &= \prod_{i=1}^{n} f(t_i, \theta) = \begin{cases} 0 & \text{if one } t_i > \theta \\ \frac{1}{\theta^n} & \text{all } t_i \leq \theta \end{cases} \\ &= \begin{cases} 0 & \max_i(t_i) > \theta \\ \frac{1}{\theta^n} & \text{else} \end{cases} \end{aligned}$$

---

[2] we can assume that these are independent of each other

Maximum Likelihood Method V

- $L(t_1, t_2, ..., t_n; \theta)$ maximal if $\theta = \max_i(t_i)$, i.e. $T_{(n)} = \max(T_1, ..., T_n)$ is a Maximum Likelihood Estimator.
- Since $E(T_{(n)}) = \frac{n}{n+1}\theta$ and $\mathrm{Var}(T_{(n)}) = \frac{n}{(n+2)(n+1)^2}\theta^2$ the estimator $T_{(n)} = \max(T_1, ..., T_n)$ is biased but consistent.
- $\hat{T}_{(n)} = \frac{n+1}{n} T_{(n)}$ is an unbiased and consistent estimator.

Maximum Likelihood Method VI

Example: Simulation experiment

- Experiment: Measure the waiting times $T_1, T_2, ..., T_i$ at i days for i=1,2,...,10 and $\theta = 5$
- estimator of θ: $E_i = \max(T_1, ..., T_i)$
- Sample of a simulation

i	T_i	E_i
1	1.12	1.12
2	0.0964	1.12
3	4.18	4.18
4	3.73	4.18
5	0.894	4.18
6	4.61	4.61
7	4.90	4.90
8	2.87	4.90
9	1.81	4.90
10	3.54	4.90

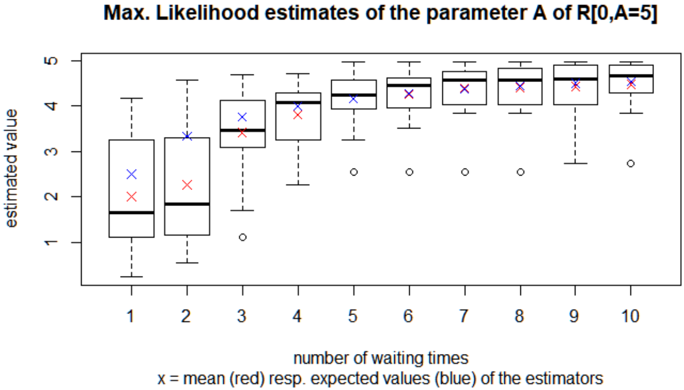
- Repeat this experiment 20 times:
  values of E_{10}: 4.90, 4.55, 4.70, 4.78, 4.28, 3.91, 4.63, 4.91, 4.93, 4.72, 4.30, 4.91, 4.89, 4.43, 4.97, 3.97, 2.74, 4.47, 3.84, 4.76

Maximum Likelihood Method VII

- some characteristic numbers of E_i

i	Min	q1	q2	Mean	q3	Max
1	0.247	1.12	1.66	2.00	3.26	4.18
3	1.10	3.14	3.46	3.42	4.11	4.68
5	2.55	3.94	4.25	4.15	4.56	4.97
7	2.55	4.06	4.57	4.40	4.77	4.97
10	2.74	4.29	4.66	4.48	4.89	4.97

- Boxplots over i=1,2,...,10



- Mention the biased but consistent estimator E_i

Estimating the Mean I

Let $(X_1, X_2, ..., X_n)$ be a random sample of size n from the distribution of a real-valued random variable X that has mean μ and standard deviation σ.

Estimator of μ: sample mean, defined by

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

- $E(\bar{X}_{(n)}(X)) = \mu$, i.e. $\bar{X}_{(n)}$ is an unbiased estimator of μ.
- $\mathrm{Var}(\bar{X}_{(n)}(X)) = \frac{\sigma^2}{n}$, i.e. the estimator tends to get closer to the parameter it is estimating as the sample size increases. Therefore the estimator is a **consistent** estimator.
- $T_n = \sum_{i=1}^{n} \alpha_i X_i$ with $\sum_{i=1}^{n} \alpha_i = 1$ is an unbiased estimator for E(X). Var(T_n) =Var(X) $\sum_{i=1}^{n} \alpha_i^2$ is minimal if $\alpha_i = \frac{1}{n}$, i.e. $\bar{X}_{(n)}$ is an efficient estimator of E(X) in the set of linear estimators.

Estimating the Mean II

Some important special cases:

- $X(\omega) = I_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \text{else} \end{cases}$ indicator variable for an event A with probability $P(A)$.

  $\bar{X}_{(n)}(X)$: relative frequency $P_n(A)$ of A

  $\Rightarrow P_n(A)$ unbiased and consistent estimator of $P(A)$.
- F distribution function of a real-valued random variable X

  For fixed x, the value $F_n(x)$ of the empirical distribution function is the sample mean for a random sample of size n from the distribution of the indicator variable $I_{X \leq x}$.

  $\Rightarrow F_n(x)$ is a unbiased and consistent estimator of $F(x)$.

Estimating the Mean III

- X random variable with a discrete distribution on a countable set S

  f probability density function of X

  For fixed $x \in S$, the empirical probability density function $f_n(x)$ is the sample mean for a random sample of size n from the distribution of the indicator variable $I_{X=x}$.

  $\Rightarrow f_n(x)$ unbiased and consistent estimator of $f(x)$.

Estimating the Variance I

Let $X = (X_1, X_2, ..., X_n)$ be a random sample of size n from the distribution of a real-valued random variable X that has mean μ and standard deviation σ.

- μ is known (usually an artificial assumption)

  $$W_n^2(X) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$$ is an unbiased and consistent estimator of σ^2.

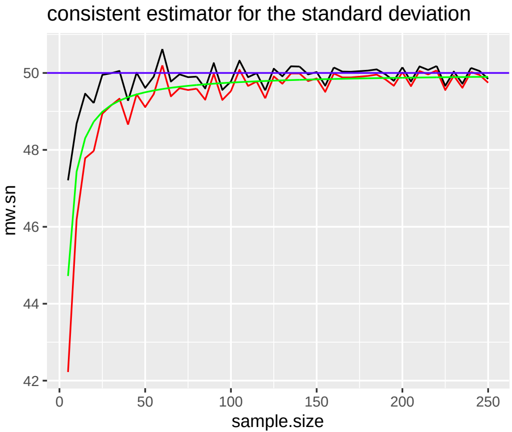
- μ is unknown (the more realistic assumption)

  $$S_n^2(X) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_{(n)}(X))^2$$ is an unbiased and consistent estimator of σ^2.

Remark: $\hat{W}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_{(n)}(X))^2$ is a biased maximum likelihood estimator of σ^2, it tends to underestimate σ^2.

Estimating the Variance II

- S_n unbiased and consistent estimator for σ
- $W_n = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$ is a biased but consistent estimator for σ with $E(W_n^2) = \sigma \cdot (1 - \frac{1}{n})$



- blue line: true value of σ
- black line: mean value of S_n in 250 random samples
- green line: expected value of W_n
- red line: mean value of W_n in 250 random samples

Questions

Which of the following statements are true or false?

t	f	
☐	☐	The arithmetic mean of a random sample $(X_1, ..., X_n)$ with i.i.d random variable X_i is an unbiased and consistent estimator for $E(X)$.
☐	☐	The arithmetic mean of a random sample $(X_1, ..., X_n)$ with i.i.d random variable X_i is an efficient estimator for $E(X)$.
☐	☐	The mean square deviation of the observed values from the sample mean in random sample $(X_1, ..., X_n)$ with i.i.d random variable X_i is an unbiased estimator of the variance.
☐	☐	The empirical distribution is an unbiased estimator of the distribution function.

Introduction to Confidence Intervals I

- A confidence interval gives an estimated range of values which is likely to include an unknown population parameter.
- It is being calculated from a given set of sample data.

Example: Mean heigt of male students at FHF

- It is impractical to weigh all male students.
- Sample of 25: mean height 177,52 cm
  177,52 is a point estimate of the population mean.
- A point estimate does not reveal the uncertainty associated with the estimate.
- Can you be confident that the population mean is within 5 cm of 177,52?

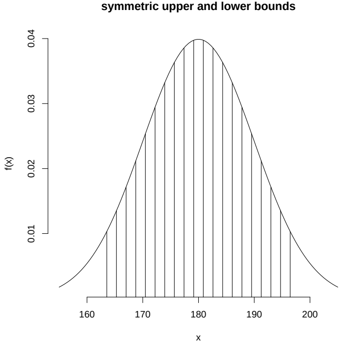
Introduction to Confidence Intervals II

Confidence intervals provide more information than point estimates:

- They are constructed using a procedure that will contain the unknown population parameter a specified proportion of the time, typically either 95% or 99% of the time.
- In case of i.i.d. samples a 95% confidence interval calculated for each sample 95% of the intervals will include the unknown population parameter.
- Width of the confidence interval gives us some idea about how uncertain we are about the unknown parameter.

Confidence Intervals for Normal Distributions: parameter μ I

Example: Assume the height of the male students is normally distributed with mean $\mu = 180$ and standard deviation $\sigma = 10$.



$$\bar{X}_{(n)} \sim N(\mu, \frac{\sigma^2}{n})$$

We compute for a given α an upper bound o and lower bound u for the possible values of $\bar{X}_{(n)}$ thuch that:

$$P(u \leq \bar{X}_{(n)} \leq o) = 1 - \alpha$$

Confidence Intervals for Normal Distributions: parameter μ II

$$P(u \leq \bar{X}_{(n)} \leq o) = 1-\alpha \Rightarrow \Phi(\frac{o-\mu}{\sigma/\sqrt{n}}) - \Phi(\frac{u-\mu}{\sigma/\sqrt{n}}) = 1-\alpha$$

Let: $[u, o] = [\mu - \delta, \mu + \delta]$; u resp. o is the 2.5% resp. 97.5% quantil of $\bar{X}_{(n)}$.

$$\Phi(\frac{\delta}{\sigma/\sqrt{n}}) - \Phi(\frac{-\delta}{\sigma/\sqrt{n}}) = \Phi(\frac{\delta}{\sigma/\sqrt{n}}) - (1 - \Phi(\frac{\delta}{\sigma/\sqrt{n}})) = 1-\alpha \Rightarrow$$

$$\Phi(\frac{\delta}{\sigma/\sqrt{n}}) = 1 - \alpha/2 \Rightarrow \delta = \frac{\sigma}{\sqrt{n}} \cdot u_{1-\alpha/2}$$

$$\text{i.e. } u = \mu - \frac{\sigma}{\sqrt{n}} \cdot u_{1-\alpha/2}, o = \mu + \frac{\sigma}{\sqrt{n}} \cdot u_{1-\alpha/2}$$

With $\mu = 180, \sigma = 10, n = 25, \alpha = 0,05$ we get $u_{1-\alpha/2} = u_{0.975} = 1.96$ and $u = 176.08, o = 183.92$

Confidence Intervals for Normal Distributions: parameter μ III

$$P\left(\mu - \frac{\sigma}{\sqrt{n}} \cdot u_{1-\alpha/2} \leq \bar{X}_{(n)} \leq \mu + \frac{\sigma}{\sqrt{n}} \cdot u_{1-\alpha/2}\right) = 1 - \alpha \Rightarrow$$

$$P\left(\bar{X}_{(n)} - \frac{\sigma}{\sqrt{n}} \cdot u_{1-\alpha/2} \leq \mu \leq \bar{X}_{(n)} + \frac{\sigma}{\sqrt{n}} \cdot u_{1-\alpha/2}\right) = 1 - \alpha$$

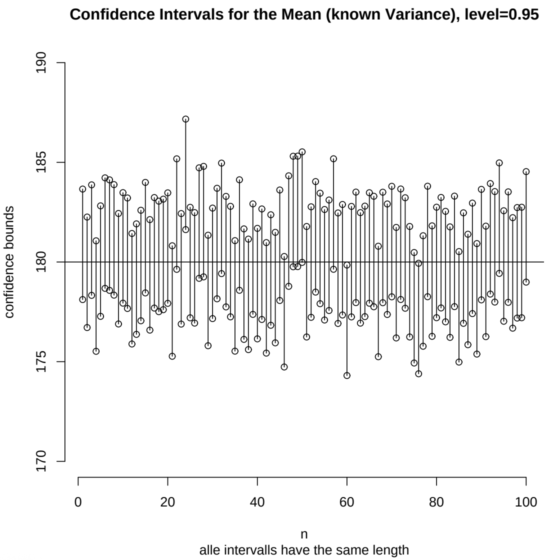
Thus we have the following confidence interval:

$$\left[\bar{X}_{(n)} - \frac{\sigma}{\sqrt{n}} \cdot u_{1-\alpha/2}, \bar{X}_{(n)} + \frac{\sigma}{\sqrt{n}} \cdot u_{1-\alpha/2}\right] = [173.61, 181.44]$$

Remark:

- $1 - \alpha$ is called the confidence level.
- The lower and upper bounds are random variables.

Confidence Intervals for Normal Distributions: parameter μ IV



If repeated samples were taken and the 95% confidence interval computed for each sample, 95% of the intervals would contain the population mean. Naturally, 5% of the intervals would not contain the population mean.

Confidence Intervals for Normal Distributions: parameter μ V

Remark:

- The formula of the confidence interval is derived using

$$\frac{\bar{X}_{(n)} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

  in the case of i.i.d. $N(\mu, \sigma^2)-$ distributed random variables $X_1, ..., X_n$.

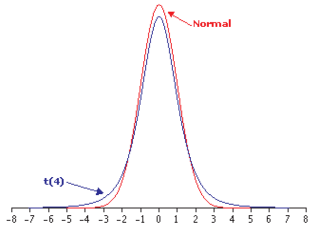
- In the case of unknown standard deviation σ, it must be estimated.

$$\frac{\bar{X}_{(n)} - \mu}{S_{(n)}/\sqrt{n}} \sim t_{n-1}$$

  i.e. t-distribution with $n-1$ degrees of freedom instead of the standard normal distribution.

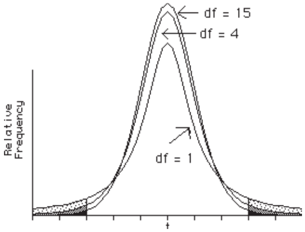
Confidence Intervals for Normal Distributions: parameter μ VI

t-distribution:



HyperStat Online Chapter 8

- The shape depends on the degrees of freedom (df) that went into the estimate of the standard deviation.
- It has relatively more scores in its tails than does the normal distribution.
- As the degrees of freedom increases, the t distribution approaches the standard normal distribution.



Stat Online Chapter 8

t distributions with 1, 4, and 15 degrees of freedom: Areas greater than +2 and less than -2 are shaded. This figure shows that the t distribution with 1 df has the least area in the middle of the distribution and the greatest area in the tails.

Confidence Intervals for Normal Distributions: parameter μ VII

Remark: Usually the variance is unknown. Therefore, the construction of a confidence interval involves the estimation of both μ and σ and the t-distribution is to be used instead of the normal distribution.

Confidence interval for μ and unknown σ:

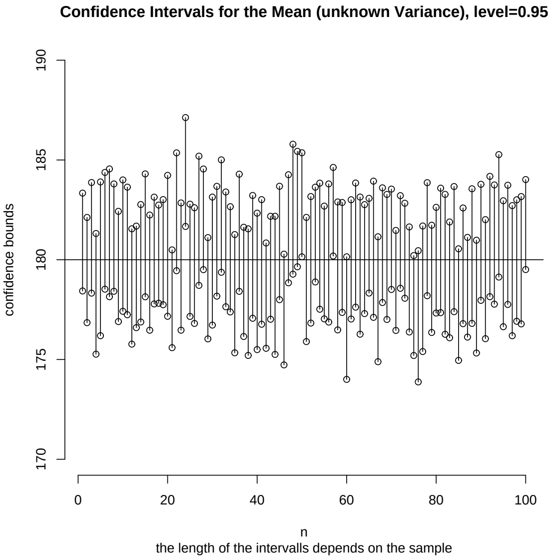
$$\left[\bar{X}_{(n)} - t_{n-1;1-\frac{\alpha}{2}} \cdot \frac{S_{(n)}}{\sqrt{n}}, \bar{X}_{(n)} + t_{n-1;1-\frac{\alpha}{2}} \cdot \frac{S_{(n)}}{\sqrt{n}}\right]$$

Example: Sample of 25 male students (i.i.d. normally distributed)

- $\bar{x}_{(25)} = 177.52, s_{(25)} = 8.227$
- confidence level $1 - \alpha = 0.95 \Rightarrow t_{24,0.975} = 2.064$

$\Rightarrow$ confidence interval for μ: $[174.12, 180.92]$

Confidence Intervals for Normal Distributions: parameter μ VIII



Note that the location and the length of the confidence interval are varying from sample to sample.

Confidence Intervals for Normal Distributions: parameter σ^2 I

In the case of i.i.d. $N(\mu, \sigma^2)-$distributed random variable $X_1, ..., X_n$ it can be shown that

$$\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi^2_n \text{ and } \sum_{i=1}^{n}\left(\frac{X_i-\bar{X}_{(n)}}{\sigma}\right)^2 \sim \chi^2_{n-1}$$

Using these results the following formulas for the confidence interval for σ^2 can be derived.

1. mean μ_0 known:

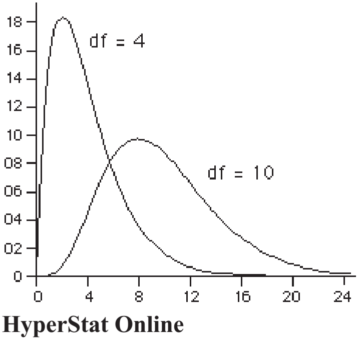
$$\left[\frac{Q_{(n)}}{\chi^2_{n;1-\frac{\alpha}{2}}}, \frac{Q_{(n)}}{\chi^2_{n;\frac{\alpha}{2}}}\right] \quad \text{with} \quad Q_{(n)} = \sum_{i=1}^{n}(X_i-\mu_0)^2$$

2. mean unknown:

$$\left[\frac{(n-1)S^2_{(n)}}{\chi^2_{n-1;1-\frac{\alpha}{2}}}, \frac{(n-1)S^2_{(n)}}{\chi^2_{n-1;\frac{\alpha}{2}}}\right]$$

Confidence Intervals for Normal Distributions: parameter σ^2 II

Chi Square Distribution:



- One parameter: degrees of freedom (df)
- A positive skewness; the skewness is less with more degrees of freedom
- As the df increase, the chi square distribution approaches a normal distribution.
- The mean of a chi square distribution is its df. The mode is df - 2 and the median is approximately df - 0.7.

Tail Areas und Chi-Squared Distributions

Source: "Tail Areas under Chi-Squared Distributions"from the Wolfram Demonstrations Project

http://demonstrations.wolfram.com/TailAreasUnderChiSquaredDistributions/

Confidence Intervals for Normal Distributions: parameter σ^2 III

Example: Sample of 25 male students (i.i.d. normally distributed)

- $\mu_0 = 180$ known
  - $Q_{(n)} = 1778$
  - confidence level $1 - \alpha = 0.95 \Rightarrow$ $\chi^2_{25,0.975} = 40.65, \chi^2_{25,0.025} = 13.12$

  $\Rightarrow$ confidence interval for σ^2: $[43.75, 135.53]$
- μ unknown
  - $S^2_{(n)} = 67.68$
  - confidence level $1 - \alpha = 0.95 \Rightarrow$ $\chi^2_{24,0.975} = 39.36, \chi^2_{24,0.025} = 12.40$

  $\Rightarrow$ confidence interval for σ^2: $[41.27, 130.98]$

Questions

Which of the following statements are true or false?

t	f	
☐	☐	The confidence level $1-\alpha$ describes the probability that the unknown parameter is contained in the confidence interval.
☐	☐	If you increase the confidence level the length of the confidence interval decreases.
☐	☐	The length of the confidence interval depends on the confidence level and the variability of the sample values.
☐	☐	Increasing the sample sizes will (in the average) improve the precision of the intervall estimation.

Confidence Intervals for Normal Distributions: Summary I

Let $\bar{X}_{(n)} = \frac{1}{n}\sum_{i=1}^{n} X_i, S^2_{(n)} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_{(n)})^2$ and $1-\alpha$ be the level of confidence.

Assumptions: Normal distribution $N(\mu, \sigma^2)$, scores are sampled randomly and are independent

1. Confidence interval for μ, standard deviation known:
$$\left[\bar{X}_{(n)} - u_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \bar{X}_{(n)} + u_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right]$$
2. Confidence interval for μ, standard deviation unknown:
$$\left[\bar{X}_{(n)} - t_{n-1;1-\frac{\alpha}{2}} \cdot \frac{S_{(n)}}{\sqrt{n}}, \bar{X}_{(n)} + t_{n-1;1-\frac{\alpha}{2}} \cdot \frac{S_{(n)}}{\sqrt{n}}\right]$$

Confidence Intervals for Normal Distributions: Summary II

3. Confidence interval for σ^2, mean μ_0 known:
$$\left[\frac{Q_{(n)}}{\chi^2_{n;1-\frac{\alpha}{2}}}, \frac{Q_{(n)}}{\chi^2_{n;\frac{\alpha}{2}}}\right] \quad \text{with} \quad Q_{(n)} = \sum_{i=1}^{n}(X_i - \mu_0)^2$$
4. Confidence interval for σ^2, mean unknown:
$$\left[\frac{(n-1)S^2_{(n)}}{\chi^2_{n-1;1-\frac{\alpha}{2}}}, \frac{(n-1)S^2_{(n)}}{\chi^2_{n-1;\frac{\alpha}{2}}}\right]$$

Remark: You get a lower resp. an upper confidence bound if you change in the corresponding bound $\alpha/2$ by α.

Confidence Intervals on the Proportion I

Example: Proportion p of X-ray machines that malfunction and produce excess radiation

- A random sample of 40 machines is taken and 12 of the machines malfunction.
- Although the point estimate of the proportion $\hat{p} = \frac{12}{40}$ is informative, it is important to also compute a confidence interval.

Confidence Intervals on the Proportion II

Assumption: Observations are sampled randomly and independently.

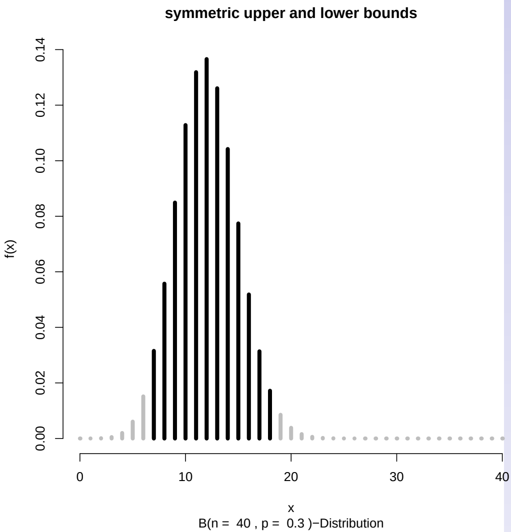
Let $X_i = \begin{cases} 1 & \text{machine i malfunctions} \\ 0 & \text{else} \end{cases}$ then $X_1, \ldots, X_n$ are independent identically $B(1,p)$-distributed with unknown p

$$\Rightarrow X = \sum_{i=1}^{n} X_i \sim B(n,p)$$

$$\Rightarrow \hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{with} \quad E(\hat{p}) = p, \mathrm{Var}(\hat{p}) = \frac{p(1-p)}{n}$$

Confidence Intervals on the Proportion III

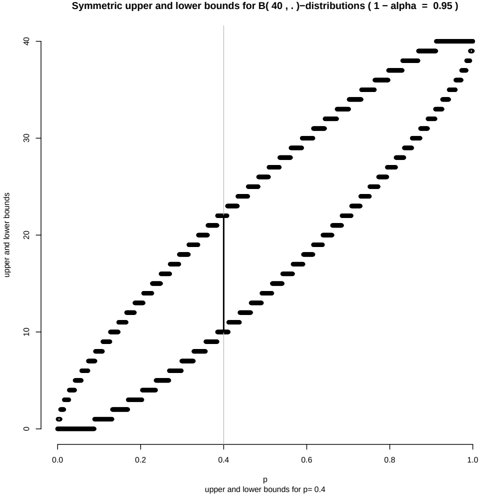
We first discuss symmetric $1 - \alpha$ upper $u_{n,\alpha}()$ and lower bounds $l_{n,\alpha}(p)$ of $X \sim B(n, p)$.



Confidence Intervals on the Proportion IV

n and α fix; $p \in [0, 1]$

- $l_{n,\alpha}(p)$, $u_{n,\alpha}(p)$ uniquely defined for p
- both functions monotonously increasing in p

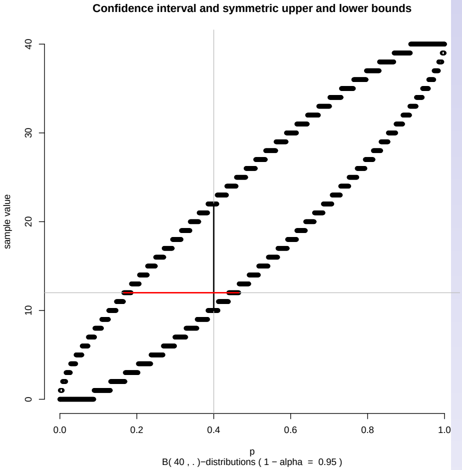


Confidence Intervals on the Proportion V

consider the sets $u_{n,\alpha}^{-1}(x) = \{p \,|\, u_{n,\alpha}(p) = x\}$, $l_{n,\alpha}^{-1}(x) = \{p \,|\, l_{n,\alpha}(p) = x\}$ and take the min and max values of these sets

sample value $x \in [0, n]$:

- $pu_{n,\alpha}(x) = \min_p u_{n,\alpha}^{-1}(x)$
- $pl_{n,\alpha}(x) = \max_p l_{n,\alpha}^{-1}(x)$.
- $p \in [pu_{n,\alpha}(x), pl_{n,\alpha}(x)]$ $(*) \Rightarrow x \in [l_{n,\alpha}(x), u_{n,\alpha}(x)]$
- $(*)$ is a $1 - \alpha$ confidence interval for p.



Confidence Intervals on the Proportion VI

Approximate the distribution of $\hat{p}$ by $(N(p, \frac{p(1-p)}{n}))$ we get with $\frac{X}{n} = \hat{p}$

$$1 - \alpha = P\left(-c \leq \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \leq c\right) \approx 2\Phi(c) - 1$$

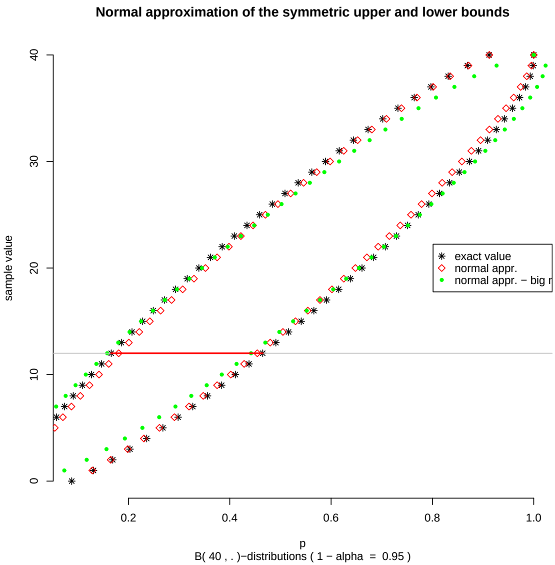
$\Rightarrow$ Approximate confidence bounds for p for the level $1 - \alpha$:

$$\left[\frac{X + \frac{c^2}{2} - c\sqrt{\frac{X(n-X)}{n} + \frac{c^2}{4}}}{c^2 + n}, \frac{X + \frac{c^2}{2} + c\sqrt{\frac{X(n-X)}{n} + \frac{c^2}{4}}}{c^2 + n}\right]$$

For big n we get approximately

$$\left[\hat{p} - u_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + u_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$

Confidence Intervals on the Proportion VII



Confidence Intervals on the Proportion VIII

Example: $X = 12, n = 40$: $\hat{p} = 12/40 = 0.30$. The estimated value of $s = \sqrt{\frac{0.3 \cdot 0.7}{40}}$. For $1 - \alpha = 0.95$ we get the following bounds:

bound	exact value	normal appr.	normal appr. - big n
lower	0.1657	0.1807	0.1580
upper	0.4653	0.4543	0.4420

Confidence Intervals on the Proportion IX

Remark: The adequacy of the normal approximation depends on the sample size n and p. Although there are no hard and fast rules, the following is a guide to needed sample size:

- If p is between 0.4 and 0.6 then an n of 10 is adequate. If p is as low as 0.2 or as high as 0.8 then n should be at least 25. For p as low as 0.1 or as high as 0.9, n should be at least 30.
- A more conservative rule of thumb that is often recommended is that np and $n(1-p)$ should both be at least 10.

Hypothesis Testing I

compare: Online Statistics IX

Hypothesis testing: statistical procedure for testing whether chance is a plausible explanation of an experimental finding.

Logic of Hypothesis Testing I

compare: Online Statistics IX

Example: Experiment to determine whether Mr. Bond is better than chance at determining whether a martini is shaken or stirred.

- 16 tests:
  - A fair coin is flipped to determine whether to stir or shake the martini.
  - Mr. Bond decides whether it was shaken or stirred.
- Correct 13 / 16 times
- Is this proof Mr. Bond can determine a stirred martini?

Logic of Hypothesis Testing II

How plausible is the explanation that Mr. Bond was just lucky?

- Probability of getting 13 or more if just guessing?

  X = number of right decisions,
  π probability of a right decision.
  $\Rightarrow X \sim B(16, \pi)$.
- $\pi \leq 0.5$: Mr. Bond is just guessing or tends to make a wrong decision
- $P_{\pi=0.5}(X > 12) = 0,0106$ and $P_{\pi=0.5}(X > x) \geq P_{\pi \leq 0.5}(X > x)$
- Strong evidence for not just guessing

x	$P_{\pi=0.5}(X > x)$
0	0.99998
1	0.99974
2	0.99791
3	0.98936
4	0.96159
5	0.89494
6	0.77275
7	0.59819
8	0.40181
9	0.22725
10	0.10506
11	0.03841
12	0.01064
13	0.00209
14	0.00026
15	0.00002
16	0.00000

Logic of Hypothesis Testing III

Probability Value:

- In the James Bond example, the computed probability of 0.0106 is the probability he would be correct on 13 or more taste tests (out of 16) if he were just guessing or tends to make a wrong decision.
- The probability of 0.016 is the probability of a certain outcome (13 or more out of 16) assuming a certain state of the world.
- If the probability of recognizing a stirred martini is less equal than 0,5 we get a probability value less than 0,0106.
- Thus the probability value is
  - not the probability he cannot tell the difference,
  - it is the probability of a certain outcome assuming a state of the world

Logic of Hypothesis Testing IV

- **Null Hypothesis** hypothesis about a state of the world (here about a population parameter)
  Typically a hypothesis of no differences, i.e. that an apparent effect is due to chance.
- Purpose of hypothesis testing: test the viability of the null hypothesis in the light of experimental data.
- Depending on the data, the null hypothesis either will or will not be rejected.

James Bond Example: Is Mr. Bond better at chance at tasting a stirred martini?
Null Hypothesis: $\pi \leq 0.5$

Logic of Hypothesis Testing V

Hypothesis could not be proven by statistical hypothesis testing.

- If the sample data could only be explained with a rather low probability assuming the hypothesis, this is used as an evidence of the opposite of the hypothesis.
- **Therefore researchers null hypothesis is the opposite of the researcher's hypothesis and they hope to reject the null hypothesis to support his hypothesis.**

Logic of Hypothesis Testing VI

Significance Testing:

- Low probability values cast doubt on the null hypothesis
- The probablility value below which the null hypothesis is rejected is called significance level or simply α level
- Conventional significance levels are 0,05 and 0,01
- A result is called statistically significant if the null hypothesis is rejected

Logic of Hypothesis Testing VII

Type I and II Errors: Two kinds of errors:

- A true null hypothesis can be incorrectly rejected (type I error).
- A false null hypothesis can fail to be rejected (type II error).

Statistical decision	True state of the Null Hypothesis H_0: H_0 True	H_0 False
Reject H_0	Type I Error	Correct
Do not reject H_0	Correct	Type II Error

α probability of rejecting H_0 given that H_0 is true

β probability of not rejecting H_0 given that H_0 is false

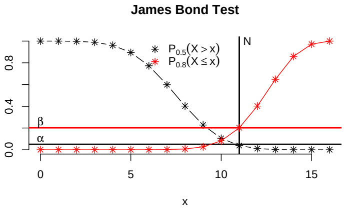
Logic of Hypothesis Testing VIII

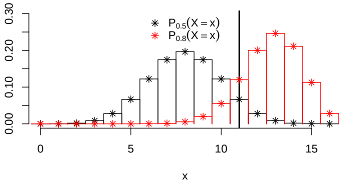
Example James Bond: $H_0 : \pi \leq 0.5 \qquad H_1 : \pi > 0.5$

X number of right decisions $\sim B(n, \pi)$

significance level $\alpha = 0.05$

Decision Rule: Reject H_0 if $X > N$





- $P_\pi(X > x)$ is a monotonously increasing function in π. Thus N is the $(1 - \alpha)$-quantil of the $B(n, \pi = 0.5)$-distribution.

  The probability of a type I error is $P_{\pi \leq 0.5}(X > N) \leq \alpha$
- The probability β of a type II error is $\beta(\pi) = P_\pi(X \leq N)$ for $\pi > 0.5$. Since $P_\pi(X < N)$ is a monotonuously decreasing function in π, β increases if π decreases to 0.5.

Logic of Hypothesis Testing IX

- **A Type II error is not really an error.**
  When a statistical test is not significant, it means that the data do not provide strong evidence that the null hypothesis is false. Lack of significance does not support the conclusion that the null hypothesis is true.
  As in a court of law: “In doubt for the accused”
- **A Type I error is really an error.**
  In case of a Type I Error the researcher erroneously concludes that the null hypothesis is false when, in fact, it is true.

Therefore, Type I errors are generally considered more serious than Type II errors. The probability of a Type I error is set by the experimenter.

Logic of Hypothesis Testing X

Tradeoff between Type I and Type II errors:

- The more an experimenter protects himself or herself against Type I errors by choosing a low level, the greater the chance of a Type II error.
- Requiring very strong evidence to reject the null hypothesis makes it very unlikely that a true null hypothesis will be rejected.
- However, it increases the chance that a false null hypothesis will not be rejected.

$\rightarrow$ For any given set of data, type I and type II errors are inversely related:

the smaller the risk of one, the higher the risk of the other

Logic of Hypothesis Testing XI

One and Two Tailed Tests:

- Depending on the form of the rejection area we distinguish one and two tailed tests.
- In the James Bond Example our question is whether Mr. Bond is better than chance at determining whether a martini is stirred or not, i.e.

$$H_0 : \pi \leq 0.5 \qquad H_1 : \pi > 0.5$$

Rejection, if Mr. Bond does very well, i.e. X is bigger a certain value $\rightarrow$ **one tailed test**

Logic of Hypothesis Testing XII

two tailed test:

- If we are asking whether Mr. Bond can tell the difference between shaken or stirred martinis, then we would conclude he could if he performed either much better than chance or much worse than chance, i.e.

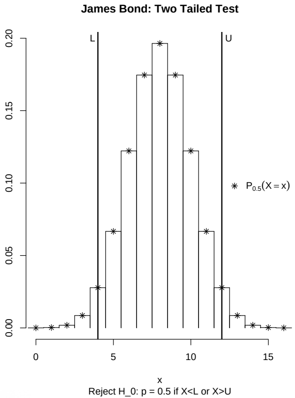
$$H_0 : \pi = 0.5 \qquad H_1 : \pi \neq 0.5$$

- If he performed much worse than chance, we would conclude that he can tell the difference, but he does not know which is which. Therefore, since we are going to reject if Mr. Bond does either very well or very poorly, i.e. rejection if X is bigger a certain value or X is lower than another certain value

Logic of Hypothesis Testing XIII

Example: James Bond

- $H_0 : \pi = 0.5 \qquad H_1 : \pi \neq 0.5$ with level $\alpha = 0.05$
- Decision Rule: Reject H_0 if $X > U$ or $X < L$
- U is the $(1 - \alpha/2)$-quantil and L is the $\alpha/2$-quantil of the $B(n, 0.5)$-distribution.



x	$P_{0.5}(X \leq x)$
0	0.00002
1	0.00026
2	0.00209
3	0.01064
4	0.03841
5	0.10506
6	0.22725
7	0.40181
8	0.59819
9	0.77275
10	0.89494
11	0.96159
12	0.98936
13	0.99791
14	0.99974
15	0.99998
16	1.00000

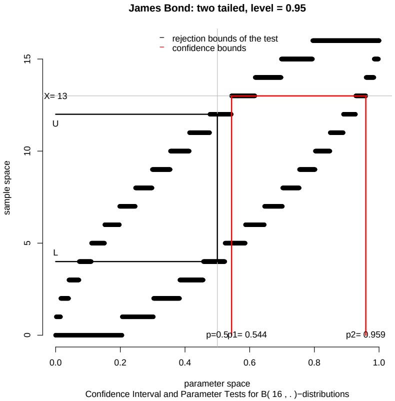
Logic of Hypothesis Testing XIV

Relationship between confidence intervals and hypothesis testing:

- A 95% confidence interval is constructed.
- Values in the interval are considered as plausible values for the parameter being estimated.
- Values outside the interval are rejected as relatively implausible.
- If the value of the parameter specified by the null hypothesis is contained in the 95% interval then the null hypothesis cannot be rejected at the 0.05 level.
- If the value specified by the null hypothesis is not in the interval then the null hypothesis can be rejected at the 0.05 level.
- Be careful if the value is close to the bounds of the confidence interval.

Logic of Hypothesis Testing XV

Example: James Bond



- A two tailed test with level $\alpha = 0.05$

  $H_0 : \pi = 0.5 \quad H_1 : \pi \neq 0.5$

  will be rejected if $X \notin [L, U]$.
- For the sample value X we get the confidence interval $[p1, p2]$ for π with level $1 - \alpha = 0.95$
- If $\pi = 0.5 \in [p1, p2]$ the null hypothesis can not be rejected at the 0.05 level.

Questions

Which of the following statements are true or false?

t	f	
☐	☐	With the help of a statistical test you can check the correctness of a hypothesis.
☐	☐	If one decides for the correctness of the hypothesis, although it is wrong, then one makes an error of 1st kind.
☐	☐	The significance level alpha of a test is the max. probability of an error 1st kind.
☐	☐	The probability of a 1st kind error and the probability of a 2nd kind error are always the same.
☐	☐	A test is used to check whether a value of interest can be brought into conformity with the data by taking into account a certain probability of error. If the value of interest lies within a confidence interval, this will be affirmed and a test will be obtained from a confidence interval. This also works the other way round.

Basic Model I

Starting Point:

- Random experiment: outcome is a sequence of n observable random variables taking values in a sample space S:

$$X = (X_1, X_2, ..., X_n).$$

- A particular outcome $x = (x_1, x_2, ..., x_n)$ of the experiment forms our data.
- Most important special: a random sample of size n from the distribution of X, i.e. $X_1, X_2, ..., X_n$ are n independent, identicaly distributed variables

Statistical Hypothesis: A statement about the distribution of the data variable X

Basic Model II

Question: Is there sufficient statistical evidence to reject a presumed null hypothesis H_0 in favor of a conjectured alternative hypothesis H_1.

Hypothesis Test $=$ Statistical Decision:

- Conclusion: reject H_0 in favor of H_1, or fail to reject H_0
- Decision is based on the data vector X

$$R \subset S : \text{reject } H_0 \text{ if and only if } X \in R$$

- Usually, the critical region R is defined in terms of a statistic $W(X)$ (test statistic).
  Asymmetry between H_0 and H_1: We assume H_0 and then see if there is sufficient evidence in X to overturn this assumption in favor of the alternative.

Basic Model III

Types of errors:

1. Type 1 error: reject the null hypothesis when it is true.
2. Type 2 error: fail to reject the null hypothesis when it is false.

- H_0 is true: $P(X \in R)$ is the probability of a type 1 error
  - Maximum probability of a type 1 error: significance level α of the test
  - R is constructed so that the significance level is a prescribed, small value (typically 0.1, 0.05, 0.01).
- H_1 is true: $P(X \notin R)$ is the probability of a type 2 error

Basic Model IV

- **Tradeoff between the type 1 and type 2 error probabilities:** If we reduce the probability of a type 1 error, by making the rejection region R smaller, we necessarily increase the probability of a type 2 error because the complementary region $S \setminus R$ is larger.
- **p-value:** The p-value of the data variable X, denoted $p(X)$ is defined to be the smallest α for which $X \in R_\alpha$; that is, the smallest significance level for which H_0 is rejected, given X.
  - If $p(X) \leq \alpha$ then we would reject H_0 at significance level α.
  - If $p(X) > \alpha$ then we fail to reject H_0 at significance level α.

Parameter Tests in the Normal Model I

Suppose that

$$X = (X_1, X_2, ..., X_n)$$

is a random sample of size n from the normal distribution with mean $\mu \in \mathbb{R}$ and standard deviation $\sigma \in (0, \infty)$.

Objective: *Hypothesis tests for μ and σ*

Parameter Tests in the Normal Model II

Example: The length of a certain part is supposed to be 70 centimeters.

- Due to imperfections in the manufacturing process, the actual length is a random variable.
- The standard deviation remains relatively stable over time due to inherent factors in the process.
  From historical data, the standard deviation is known to be 4.
- The mean may be set by adjusting various parameters in the process and hence may change to an unknown value fairly frequently.
- Sample of n parts: X_i measured length of part i $(1 \leq i \leq n)$

$$X = (X_1, X_2, ..., X_n) \quad \text{with } X_i \text{ i.i.d. } N(\mu, 16)$$

Parameter Tests in the Normal Model III

Test Statistic: $\bar{X}_{(n)} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is normally distributed with mean μ and standard deviation $\sigma/\sqrt{n}$.

$H_0 : \mu_0 = 70 \quad$ versus $\quad H_1 : \mu \neq \mu_0 \quad$ (significance level α)

Decision Rule: Reject H_0 if $\bar{x}_{(n)} \notin [\mu_0 - c, \mu_0 + c]$ where

$$P_{\mu=\mu_0}(\mu_0 - c \leq \bar{X}_{(n)} \leq \mu_0 + c) = 1 - \alpha$$

Therefore we get $c = u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$ and the rejection region

$$R_\alpha = \left[\mu_0 - u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \mu_0 + u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right]^c$$

Using the test statistic $\dfrac{\bar{X}_{(n)} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$ we get the decision rule:

Rejection of $H_0 \Leftrightarrow \dfrac{\bar{x}_{(n)} - \mu_0}{\sigma/\sqrt{n}} \notin [-u_{1-\alpha/2}, u_{1-\alpha/2}]$

Parameter Tests in the Normal Model IV

- H_0 will be rejected if

$$\bar{X}_{(n)} < \mu_0 - u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \quad \text{or} \quad \bar{X}_{(n)} > \mu_0 + u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$\iff \mu_0 \notin \left[\bar{X}_{(n)} - u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X}_{(n)} + u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right],$$

  i.e. if μ_0 is outside the corresponding confidence interval.

- p-value: with $c = |\bar{x} - \mu_0|$, $\bar{x}$ = sample mean

$$p(\bar{x}) = 1 - P_{\mu=\mu_0}(\mu_0 - c \leq \bar{X}_{(n)} \leq \mu_0 + c) = 2(1 - \Phi(\frac{c}{\sigma/\sqrt{n}}))$$

- probability of type 2 error:

$$\begin{aligned}\beta(\mu) &= P_\mu(\bar{X}_{(n)} \notin R_\alpha) \\ &= P_\mu(\mu_0 - u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \bar{X}_{(n)} \leq \mu_0 + u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}})\end{aligned}$$

  If α decreases then $\beta(\mu)$ increases.

Parameter Tests in the Normal Model V

Example: Sample of 100 parts with mean 71: $\alpha = 0,05$ and $\sigma = 0.4$

Decision Rule: Reject H_0 if $\bar{X}_{(n)} \notin [69,216; 70,784]$

- Since $\bar{X}_{(n)} = 71$ H_0 should be rejected.
- p-value $= 1 - P_{\mu=70}(69 \leq \bar{X}_{(n)} \leq 71) = 0,0124$
- probability of type 2 error:

$$\begin{aligned}\beta(\mu) &= P_{\mu}(69,216 \leq \bar{X}_{(n)} \leq 70,784) \Rightarrow \\ \beta(71) &= \Phi(\frac{70,784 - 71}{0,4}) - \Phi(\frac{69,216 - 71}{0,4}) = 0,295\end{aligned}$$

Example: Hypothesis Tests About A Population Mean
Samples are drawn from a normally distributed population with mean zero and variance given.

Source: "Hypothesis Tests about a Population Mean"from the Wolfram Demonstrations Project

http://demonstrations.wolfram.com/HypothesisTestsAboutAPopulationMean/

Parameter Tests in the Normal Model VI

Summary parameter tests in the normal model:
Analogously we get using the relationship between confidence intervals and hypothesis testing the following results for the significance level α.

I) **Gauß-Test:** $N(\mu, \sigma_0^2)$ with μ unknown and σ_0 known

Teststatistic: $\dfrac{\bar{X}_{(n)} - \mu}{\sigma_0}\sqrt{n} \sim N(0,1)$

Decision Rule: $\dfrac{\bar{x} - \mu_0}{\sigma_0}\sqrt{n} \in R \Rightarrow$ reject H_0

H_0	rejection region R
$\mu = \mu_0$	$(-\infty, -u_{1-\frac{\alpha}{2}}) \cup (u_{1-\frac{\alpha}{2}}, \infty)$
$\mu \leq \mu_0$	$(u_{1-\alpha}, \infty)$
$\mu \geq \mu_0$	$(-\infty, -u_{1-\alpha})$

Parameter Tests in the Normal Model VII

II) **t-Test:** $N(\mu, \sigma^2)$ with μ and σ_0 unknown

Teststatistic: $\frac{\bar{X}_{(n)} - \mu}{s_{(n)}}\sqrt{n} \sim t_{n-1}$ with

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

Decision Rule: $\frac{\bar{x} - \mu_0}{s_{(n)}}\sqrt{n} \in R \Rightarrow$ reject H_0

H_0	rejection region R
$\mu = \mu_0$	$(-\infty, -t_{n-1,1-\frac{\alpha}{2}}) \cup (t_{n-1,1-\frac{\alpha}{2}}, \infty)$
$\mu \leq \mu_0$	$(t_{n-1,1-\alpha}, \infty)$
$\mu \geq \mu_0$	$(-\infty, -t_{n-1,1-\alpha})$

Parameter Tests in the Normal Model VII

III) $N(\mu, \sigma^2)$ with μ and σ unknown

Teststatistic: $\dfrac{(n-1)S^2_{(n)}}{\sigma^2} \sim \chi^2_{n-1}$ with

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

Decision Rule: $\dfrac{(n-1)s^2_{(n)}}{\sigma_0^2} \in R \Rightarrow$ reject H_0

H_0	rejection region R
$\sigma^2 = \sigma_0^2$	$(0, \chi^2_{n-1,\frac{\alpha}{2}}) \cup (\chi^2_{n-1,1-\frac{\alpha}{2}}, \infty)$
$\sigma^2 \leq \sigma_0^2$	$(\chi^2_{n-1,1-\alpha}, \infty)$
$\sigma^2 \geq \sigma_0^2$	$(0, \chi^2_{n-1,\alpha})$

Parameter Tests in the Normal Model VIII

Remark: Calculation of the p-value in these cases:

- Exchange the quantile in the rejection region R by the value of the teststatistic for the given data. $\rightarrow \tilde{R}$
- p-value$= P_{H_0}(\text{teststatistic } \in \tilde{R})$

Questions

A sample of the size n=25 from a normal distribution resulted in $\bar{x} = 9$ and $s = 2$. You want to conduct a hypothesis on μ at the $\alpha = 0.05$-level.
Which of the following statements are true or false?

t	f	
☐	☐	The value of the test statistic T is 2.5.
☐	☐	$H_0 : \mu = 10$ can be rejected if the absolute value of the test statistic T is bigger $u_{0.975}$.
☐	☐	$H_0 : \mu > 10$ can be rejected if the value of the test statistic $T < -t_{24,0.95}$.
☐	☐	The p-value for $H_0 : \mu = 10$ is ca. 0.02.
☐	☐	The rejection region for $H_0 : \sigma^2 \leq 1.5$ is $\frac{24 * s^2}{1.5^2} > \chi^2_{24,0.95}$.

Tests in the Bernoulli Model I

- $X_i, 1 \leq i \leq n$ are independent random variables taking the values 1 and 0 with probabilities p and $1 - p$ respectively.
- $X = (X_1, X_2, ..., X_n)$ is a random sample from the Bernoulli distribution with unknown success parameter $p \in (0, 1)$.

Tests in the Bernoulli Model II

Applications:

- Event of interest in a basic experiment with unknown probability p.
- Replicate the experiment n times and define $X_i = 1$ if and only if the event occurred on run i.

Tests in the Bernoulli Model III

Example:

- Population of objects of several different types.
- p unknown proportion of objects of a particular type of interest.
- Select n objects at random from the population and let $X_i = 1$ if and only if object i is of the type of interest.
- When the sampling is with replacement, these variables really do form a random sample from the Bernoulli distribution.
- When the sampling is without replacement, the variables are dependent, but the Bernoulli model may still be approximately valid.

Tests in the Bernoulli Model IV

- Number of successes $X = \sum_{i=1}^{n} X_i \sim B(n,p)$, i.e.
  $E(X) = np$ and $\text{Var(X)} = np(1-p)$.
- In case of large n the distribution of X is approximately normal, by the central limit theorem. An approximate normal test can be constructed using the test statistic $\frac{r_n - p}{\sqrt{p(1-p)/n}}$ with $r_n = \frac{X}{n}$.[3]
- **Decision Rule:** $\frac{r_n - p_0}{\sqrt{p_0(1-p_0)/n}} \in R \Rightarrow$ reject H_0

H_0	rejection area R
$p = p_0$	$(-\infty, -u_{1-\frac{\alpha}{2}}) \cup (u_{1-\frac{\alpha}{2}}, \infty)$
$p \leq p_0$	$(u_{1-\alpha}, \infty)$
$p \geq p_0$	$(-\infty, -u_{1-\alpha})$

[3] The adequacy of the normal approximation depends on n and p. A rule of thumb is that np and n(1 - p) should both be at least 10.

Tests in the Bernoulli Model V

Example: (compare Heumann, Schomaker, p 228) A party wants to know whether the proportion of votes will exceed 30%. In a representative sample of size n=2000 of eligible voters 700 have voted the party.

- test statistic $T = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n} = \frac{0.35-03}{\sqrt{0.3(1-0.3)}}\sqrt{2000} = 4.8795$
- If $\alpha = 0.05$, $T = 4.8795 > u_{1-\alpha} = 1.64$, $H_0 = p \leq 0.3$ can be rejected
- p-value: $P(T \geq 4.8795) = 5.318e - 07$

Tests in the Bernoulli Model VI

Exact binomial test: above example with $\alpha = 0.05$

- Teststatistic $T = X \sim B(n = 2000, p_0)$
- Critical region: find c with $P_{p=0.3}(T \geq c) \leq 0.05$, i.e. $P_{p=0.3}(T < c) \geq 0.95$
  From R (qbinom(p=0.95, size=2000, prob=0.3)) we get c = 634. Since $T = 700 > 634$ we reject $H_0 : p \leq 0.3$.
- p-value: $P_{p=0.3}(T \geq 700) = 1 - P_{p=0.3}(T \leq 699) = 8.395e - 07$ (pbinom(699,size=2000,prob=0.3))
- The result can be easily get by R binom.test(): binom.test(700,2000, p=0.3, alternative="greater")

Questions

In a big city, 20% of all households have subscribed to a certain magazine so far. In a random sample of 100 households, 16 households have subscribed to the magazine. Let $H_0 : \pi >= 0.2$ and $\alpha = 0.05$.
Which of the following statements are true or false?

t	f	
☐	☐	Based on a normal approximation the value of the test statistic T is -0.875.
☐	☐	The rejection region of the approximation of the binomial test is $T < -u_{0.95}$ if T is the test statistic.
☐	☐	The p-value of the approximation of the binomial test is $\Phi(t)$, if t is value of the test statistic.
☐	☐	The p-value of the exact binomial test is $P(X <= 16)$ if $X \sim B(100, 0.2)$.

Content