Information Retrieval
WS 2012 / 2013

Lecture 14, Wednesday February 13th, 2013
(Course Evaluation, Exam, Work at our Chair)

Prof. Dr. Hannah Bast
Chair of Algorithms and Data Structures
Department of Computer Science
University of Freiburg

UNI
FREIBURG

Overview of this lecture

- **Organizational**

  - Your results + experiences with Ex. Sheet 13 (Z-test, T-test)

  - Results of the official **evaluation** of this course

- **Exam**

  - Types of tasks + grading scheme

  - Let's solve some tasks together live

- **Work at our chair**

  - How we work

  - Current projects + what's behind them

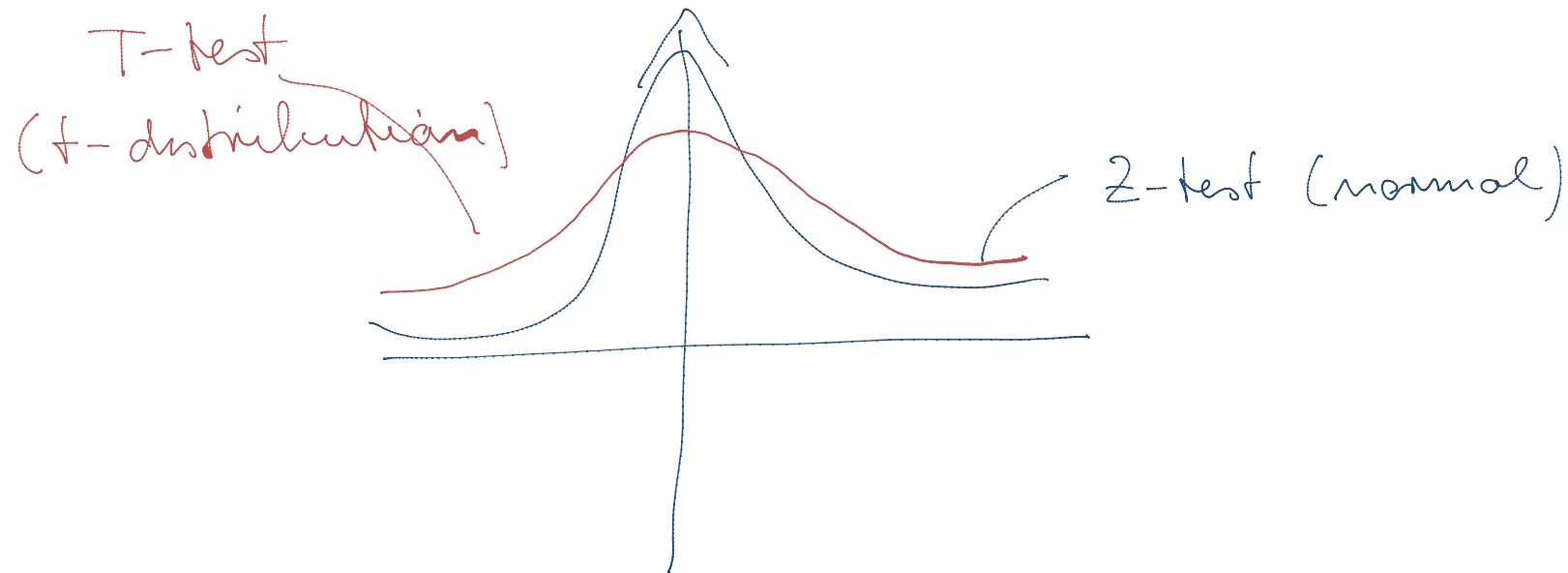 - Other lectures

Experiences with ES#13   (Z-test, T-test)

- **Summary / excerpts**   last checked February 13, 15:00

  - Easy enough, given the lecture files

  - Unfortunately, all p-values are very close to 0%

    This was not planned, but it's still interesting: even a relatively small difference in the means (0.49s vs 0.55s for our measurement) can be **very** significant if only the variance is small enough (0.01 for us)

  - R is a great tool for this kind of exercise

  - I guess you don't want me to upload the TSV / DB files?

  - "I have done the evaluation and now want my 20 points !"

  - "I don't want the points because it feels like a bribe"

  - Sacrificed exercise sheet for (other) exam preparation

  - Liked the shell hacking session at the end of last lecture

Z-test : μ, σ fixed, pick from $N(\mu, \sigma^2)$

T-test : μ fixed, pick σ^2 from $\chi^2(n)$ with
mean s^2

then pick from $N(\mu, \sigma^2)$

T-test
(t-distribution)

Z-test (normal)

Results Course Evaluation   1/7

- **Participants**

  - Still participating in the course: **58**

  - Registered for exam: **55** ... don't forget it !

  - Participated in the evaluation: **55** ... great !

    32 x MSc Inf, 15 x BSc Inf, 7 x ACS, 1 x ESE, 1 x Lehramt

  - Nominations for teaching award: **40** ... thank you !

    "Which committee oversees the award / will she award herself?"

  - In the following, a summary of your feedback

  - You find all the details linked on the course Wiki:

    http://ad-wiki.informatik.uni-freiburg.de/teaching/InformationRetrievalWS1213/Feedback

- **Style of the course**

  - Learned a lot: 39 x fully agree, 14 x agree, 2 x partly

  - Level of contents: 26 x high, 25 x appropriate, 4 x other

  - Well explained: 42 x fully agree, 12 x agree, 1 x partly

  - Responds to questions: 44 x fully agree, 11 x agree

  - Great mix of theory and practice / real data / cookies

  - Interesting / informative / comprehensive / well-structured

  - Also complex stuff explained + motivated well / not boring

  - Asks for student's feedback and considers it

  - "She really cares about the students and the teaching"

  - Many of you liked the live programming / shell hacking

Results Course Evaluation   3/7

- **Exercise Sheets**

  - Difficulty: 29 x appropriate, 20 x difficult, 5 x too difficult

  - Meaningful: 34 x fully agree, 16 x agree, 3 x partly

  - Effort: 29 x 9-12h, 12 x 13 – 16h, 10 x 1-8h, 4 x 17-20h

  - Very time-consuming / too much for several of you

  - Well designed practical exercises, make you really under-
    stand and remember the stuff + programming skills improve

  - Result tables on the Wiki, good for comparison and tuning

  - There were master solutions

  - Great / competent tutor, extensive feedback

  - "My tutor (Eugen) definitely earned a C++ Ninja award"

Results Course Evaluation   4/7

- **Materials / Online Support**

  - Helpful: 39 x fully agree, 14 x agree, 1 x partly, 1 x no

  - Consumed: 13 x present, 20 x recording, 22 x both

  - Active Forum, short answer times (usually)

  - Exercise submission system (Daphne, SVN) very good

  - Materials are "exemplary and unrivaled at the faculty"

  - Best recordings at the faculty, the HiWi is worth it !

    in particular: fast availability + camera picture = more lively

  - Materials are "exemplary and unrivaled at the faculty"

  - "Like the next episode of my favorite TV series"

- **Complaints**

  - Quite a lot of mistakes on the slides (for some lectures)

  - Second half of course: big picture less clear / harder to follow

  - Naïve Bayes, SVM, SPARQL belong into other lectures

  - Too much mathematics, esp. when only the result is needed

  - Less web app stuff, more about semantic search

  - "Requiring login for evaluation defeats the purpose"

  - Tutor feedback not only on the code, please

  - Unfair distribution of points over the sub-exercises per sheet

  - Awarding of points was too strict in the beginning

Results Course Evaluation   6/7

- **More complaints**

    - Colors are good / switching colors is bad

    - Improve space management when writing on slides

    - Don't modify slides shortly before the lecture

        same comment already came for Efficient Route Planning

    - Support usage of IDEs / Windows better

    - Alternative to SVN, allow languages beyond Java and C++

Results Course Evaluation   7/7

- **Planned improvements for next course**

  - Improve slides + explanations + time management

    In particular for: Web Stuff, LSI, k-means, SVM, SPARQL, T-test

  - Better split of web app stuff over two lectures

  - In general, there will be much fewer mistakes on slides

  - Improve specification for exercise sheets

    In particular, avoid the various pitfalls we are aware of now

  - Reduce time effort needed for sheets

  - Solve the pen color problem … maybe just switch rooms :-)

  - Maybe more about semantic search next time

  - And, of course, I will consider all the other feedback too …

Exam   1/2

- **Where, when, how**

  - Friday, **March 1, 2:00 – 4:00pm**, Kinohörsaal

  - Probably six tasks, out of which the five best will count

  - The exam is **open book** = you can bring books, paper, …

    But please be ecological when printing out slides

  - Electronic devices of any kind are not allowed

  - There will be a sub-forum for questions about the exam

  - Please bring: **student id**, **colored pens**, **brain**

Exam 2/2

- **Types of questions**

  - **Type 1:** Do the steps of an algorithm, or a variant thereof, like we did in the lecture … see colored pens

  - **Type 2:** Write a small program, or understand what a given small program does

  - **Type 3:** Small calculations or proofs … see brain

  - In general: the emphasis is on (basic) understanding, not on learning things by heart

  - If you have done the exercise sheets, all you have to do for exam preparation is refresh your memory a bit

  - Preparation for proofs: practice, practice, practice

Σ GRAPE

Σ		O	1	2	3	4	5
O							
R		2	2	1			
A		3					
N		4					
G		5					
E		6					

$\forall x_1, x_2, y_1, y_2$

$$ED(x_1 x_2, y_1 y_2) \leq \overbrace{ED(x_1, y_1)}^{k_1}$$

$$+ \underbrace{ED(x_2, y_2)}_{k_2}$$

$x_1 \xrightarrow{k_1} y_1$

$x_2 \xrightarrow{k_2} y_2$

$x_1 x_2 \xrightarrow{k_1} y_1 x_2 \xrightarrow{k_2} y_1 y_2$

$\Rightarrow \quad "\leq"$

$$ED(\underbrace{GR}_{x_1}\underbrace{AU}_{x_2}, \underbrace{RA}_{y_1}\underbrace{UM}_{y_2}) = 2$$

$ED(x_1, y_1) = 2$
GR RA

$ED(x_2, y_2) = 2$
AU UM

$\boxed{+} = 4$

Work at our Chair   1/3

- **How we work**
  - 1/3 Theory (new algorithms, performance analysis, etc.)

    E.g. an efficient index for semantic search, or for computing shortest paths in very large transportation networks

  - 1/3 Algorithm Engineering (good implementation)

    An idea that looks great in theory might not work that well, or even not at all, in practice

    On the other hand, hacking around without theoretical understanding often leads nowhere good either

  - 1/3 Software Engineering (good software)

    Writing a program for yourself which runs once now is one thing

    Writing software together with others that can still be used in five years from now is a totally different story

Work at our Chair   2/3

- **Current projects**

  - Multi-modal route planning

    Arbitrary combination of car, transit, bike, flights, ...

    Good models, efficient algorithms, a working system

  - Semantic search

    Search with "understanding" of the query and documents

    Show example of CompleteSearch and Broccoli

  - Research paper management

    Automatic metadata + reference extraction

    Show demo of IceCite prototype

Work at our Chair   3/3

- **Other courses**

  - **Information Retrieval** ... next time in WS 13/14

    Similar to this time, just better :-)

  - **Efficient Route Planning** ... next time only in SS 2014

    All you need to build a state-of-the-art route planner

    Many algorithms / heuristics + their implementation

  - **Algorithms and Data Structures** ... next time in SS 2013

    Basic course for 2nd semester BSc Informatik students

  - **Programming in C++** ... maybe again in SS 2014

    2nd semester BSc Info + 4th semester BSc ESE