Professional Career Opportunities in Mathematics and Statistics

Kao-Tai Tsai, Ph.D.
Frontier Informatics Services, New Jersey, USA
(Email: kaotai.tsai@gmail.com / LINE: kaotaitsai)

Presentation at the
Department of Mathematics, National Taiwan Normal University

January 7, 2022

统計資料分析

蔡高太
師大數學系六二級畢業生

二五離系七十回,
鄉音無改視花花
(後兩字台語發音),
年輕學子不相識,
借問老翁何事來.

(西元2022年回母校母系與師長好友歡敍以資為記)

Outline of Presentation

1. The foundation of science: Mathematics and Statistics
   - Evolution of statistical data analysis
2. Organizations with great needs of statisticians and analysts
   - Primary activities for statisticians and analysts
3. Preparation for the opportunities and be successful
   - Analytical methodologies and tools
   - Selected useful supplemental readings
   - Big data source example for machine learning
4. Managing yourself
5. Advises for the future

Mathematical/Statistical Data Analysis

- Mathematical and statistical data analysis have a long history, actually, the Babylonia (c. 4500–1900 BC), is a famous example of where initial records of written mathematics and data analysis have been discovered.
- Through human history, data analysis had become the foundation for scientific discoveries, advances of medicine, and government policies.
- Due to the advances of sciences, business, and computing technologies, etc., mathematics science has been greatly advanced and data have become ubiquitous.
- With the emphasis of "everything evidence-based," how to extract valid, useful, and actionable information based on rigorous mathematical foundation and well-conducted data analysis is more critical than ever.

Evolution of mathematical/statistical data analysis

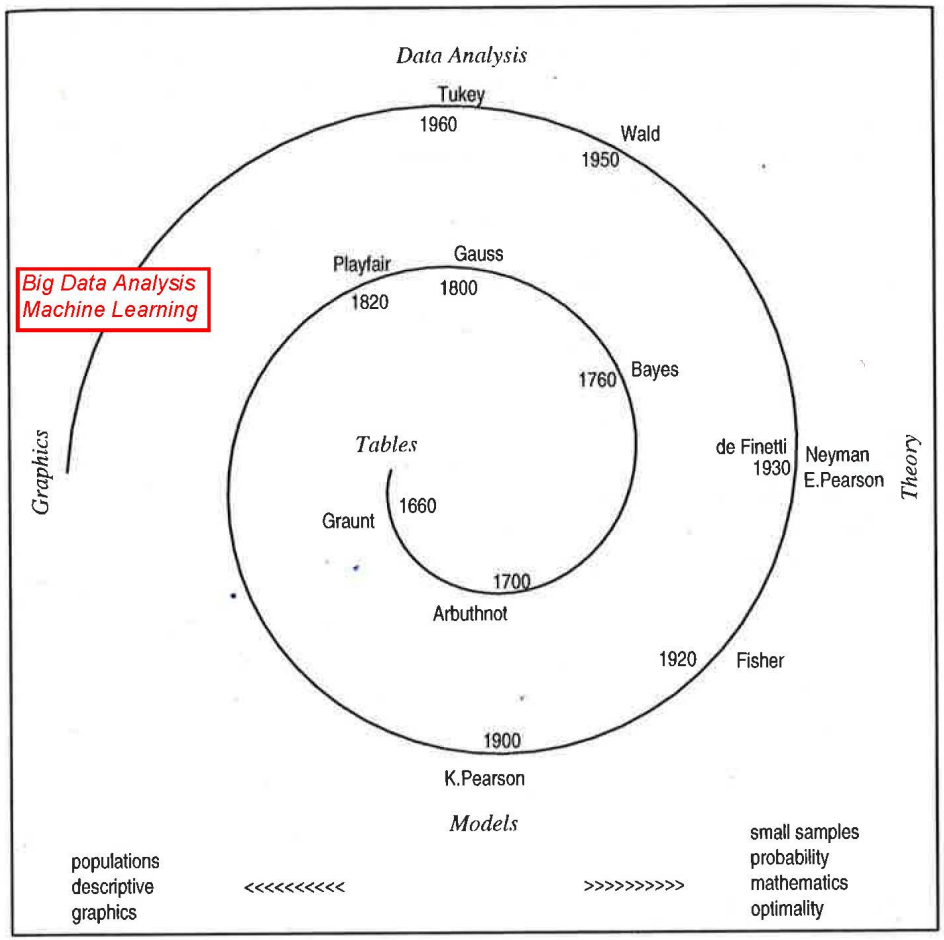


Figure 1: The evolving spiral path of statistics (Huber (1997))

Organizations with great needs of statisticians and analysts

- **Healthcare-related industries**:
  - Pharmaceutical industry (for drug or medical device development)
  - Health-oriented research institutes or hospitals
  - Insurance companies
  - etc.
- **Government**: NIH, FDA, Department of Labor, Census, Economics statistics, Medicare, National Defense, etc..
- **Financial**: banks, investment institutes, e-commerce (e.g., Alibaba, Amazon, Facebook, Google, Netflix, etc.).
- **Manufacturing** industries.
- **Services** industries.
- **Education** industries.
- etc.

Primary activities for statisticians and analysts - 1

As an example in Pharmaceutical industry:

Before clinical stage:

- Animal Toxicology/Pathology studies in variety of animals, such as mice, dogs, monkeys, etc. (work with toxicologists and pathologists to determine the relationship between the degree of toxicity and dose levels).

Primary activities for statisticians and analysts - 2

During clinical phases (in human):

- Pharmaco-kinetics and Pharmaco-dynamics (PK/PD): investigate relationship between efficacy/toxicity and dose levels of the chemical compounds under study in healthy volunteers (so-called dose-response).
- Phase 1: investigate relationship between toxicity and dose levels from PK/PD on patients. Usually not focus on efficacy yet.
- Phase 2: take the findings from Phase 1 and extending to larger studies. Efficacy is among one of the focuses.
- Phase 3: this usually is a larger study with more patients, with well-defined objectives. Intend to submit the results to the FDA, EMA, etc. for marketing approval, if the study is successful.

Primary activities for statisticians and analysts - 3

Note:

- From animal studies to marketing approval, it can take up to 7-10 or more years, tens to hundreds million of dollars investment, countless hours of many scientists in various disciplines and supporting staffs. It is a tremendous endeavor.
- In all phases, statisticians have to work with the study team to
  - decide the size of the study.
  - write *Statistical Analysis Plan* (SAP) to specify what will be analyzed and the statistical methods to be used.
  - join meeting with and answer questions from regulatory agencies.

Primary activities for statisticians and analysts - 4

During post marketing phase:

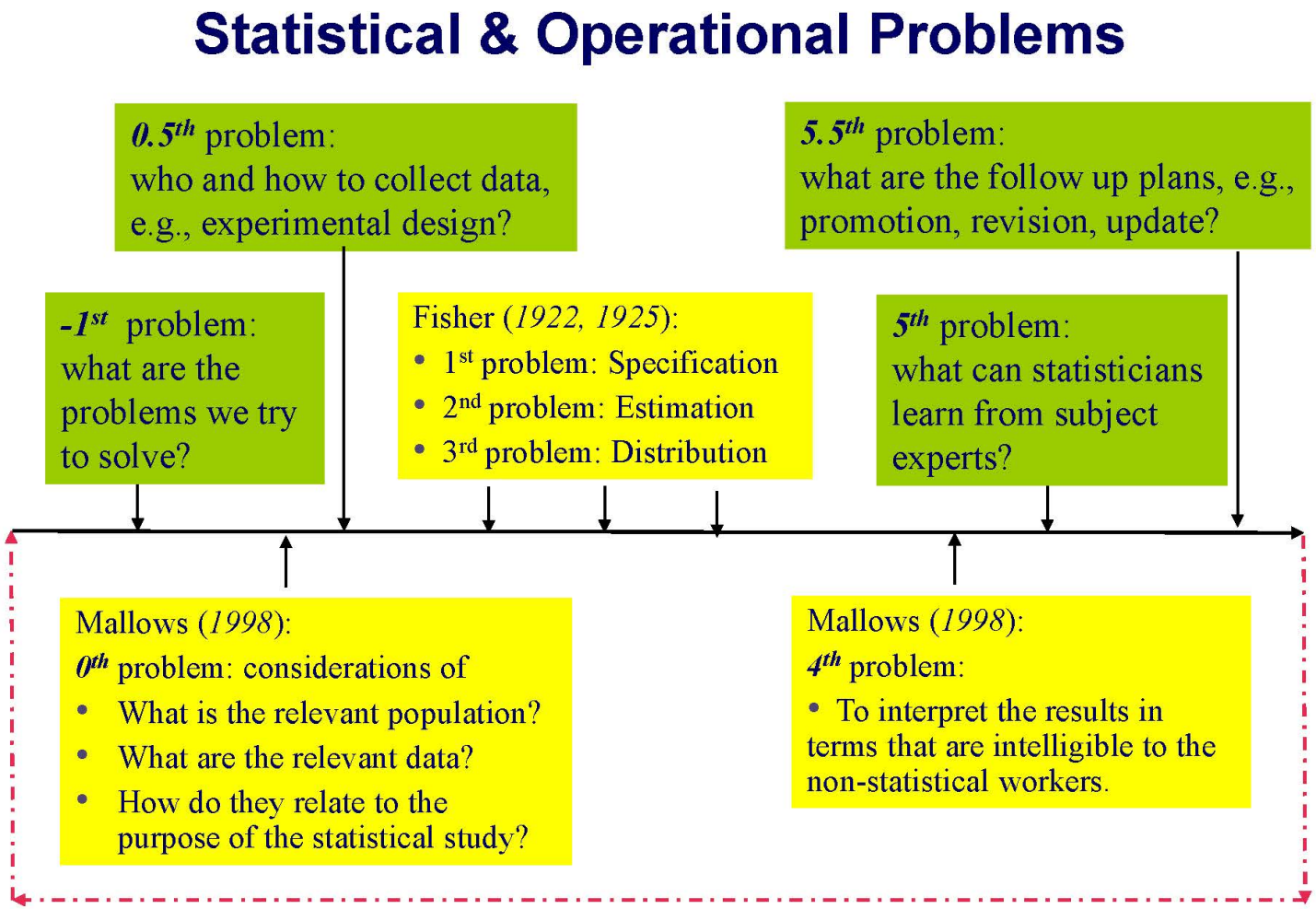
- Phase 4: this is the study design to generate more information for marketing purposes.
- This can be a registry study to collect more data about the drug effect and side-effects when the dung is used by general public.
- Statisticians will conduct statistical data analysis for marketing team to promote the products and to understand the broader aspects of the drug.

Primary activities for statisticians and analysts - 5

Support biology/genomics research:

- To have in-depth understanding of the drug effects, scientists (usually biologists, genomics, and clinicians, etc.) would conduct research based on DNA/RNA/Protein/etc. data.
- This is a very important step toward the goal of *Precision Medicine.*
- Most data sets are big data, use machine learning analytical methodologies more extensively.

Statistical analysis considerations



IV-14

Preparation for the opportunities and be successful

- Soft skills:
  - Inter-personal skills
  - Speak well: oral communication in English.
  - Write well: write good emails and reports in English.
  - Present yourself, your ideas, and your products well.
  - Leadership skills: if you have, good; if you don't, learn it!
  - Speak up of your ideas! Silence is not golden, it can be perceived as dumb.
  - Be adventurous, step out of your comfort zone! Don't just do the things you are very comfortable with, learn new knowledge/skills that is where the new opportunities coming from.
- Hard skills:
  - Mathematical and statistical knowledge
  - Programming coding in languages - R, SAS, Python, etc.
  - Proficient statistical computing skills.

Analytical methodologies and tools

Courses offered in NTNU curriculum:

- mathematical statistics, statistical methods, regression, survival data analysis, categorical data analysis, multivariate statistical methods, experimental design.
- statistical computing, Bayesian analysis, clustering analysis.

Some commonly used ML methodologies:

- Graphical methods,
- Regression methods (e.g., Lasso-types and extensions),
- Tree-based methods (recursive partitioning),
- Support Vector Machine,
- Network analysis,
- Casual-effect inference.

Selected useful supplemental readings - 1

Hard skills:

- *Biostatistics: Introductory Biostatistics*, by Chap T. Le.
- Introduction of control clinical trials (search Internet).
- For self-learning of R language and programming: go to Internet and look for tutorial of R (should have many).
- *An R Companion to Applied Regression*, by J. Fox and S. Weisberg. (intro data analysis.)
- *An Introduction to Statistical Learning with applications in R*, by James, Witten, Hastie, and Tibshirani. (entry level of ML)
- *Machine Learning for Knowledge Discovery with R*, by Kao-Tai Tsai, Chapmam & Hall (2021) (Note: More advanced ML, but the first 2 chapters are easy important reading.)

nasa

Selected useful supplemental readings - 2

Soft skills:

- Harvard Business Review magazine.
- Classical Chinese literature: e.g., Han-Fei Tze, Tz-wong Tze.
- The 7 habits of highly effective people.
- Getting to *YES* - *negotiating an agreement without giving in*.

7H

Big data source example for machine learning

UCI Machine Learning Repository

University of California, Irvine hosts this website. It has variety of big data from various fields.

Genomics and Clinical Data

US NCI and the National Human Genome Research Institute, as well as many other institutes contribute data to this website: https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics.

- The Cancer Genome Atlas (TCGA) had already generated over 2.5 petabytes of genomic, epigenomic, transcriptomic, and proteomic data, which already led to great improvements of diagnose, treat, and prevent cancer.

Managing yourself - work and life - 1

For work:

- Learn how to plan and prioritize.
- Under promise and over deliver.
- Ask for feedback.
- Sweat for the details. *Quality is Job One.*

Managing yourself - work and life - 2

For life:

- Values & Integrity.
- Be kind and generous: treat people in kind and people will treat you back in kind.
- Continuous learning: *Be Humble, Be Foolish, and Be Hungry.*
- Set quiet times for yourself from time to time, to reflect and to plan.
- Practice Chinese calligraphy, a very effective way to clear your thoughts and clam your mind.
- Out to the *Nature* and appreciate it: watch the clouds, listen to the birds, smell the flowers, and enjoy *The Beauty of Nature* with your loved ones.

Advises for the future

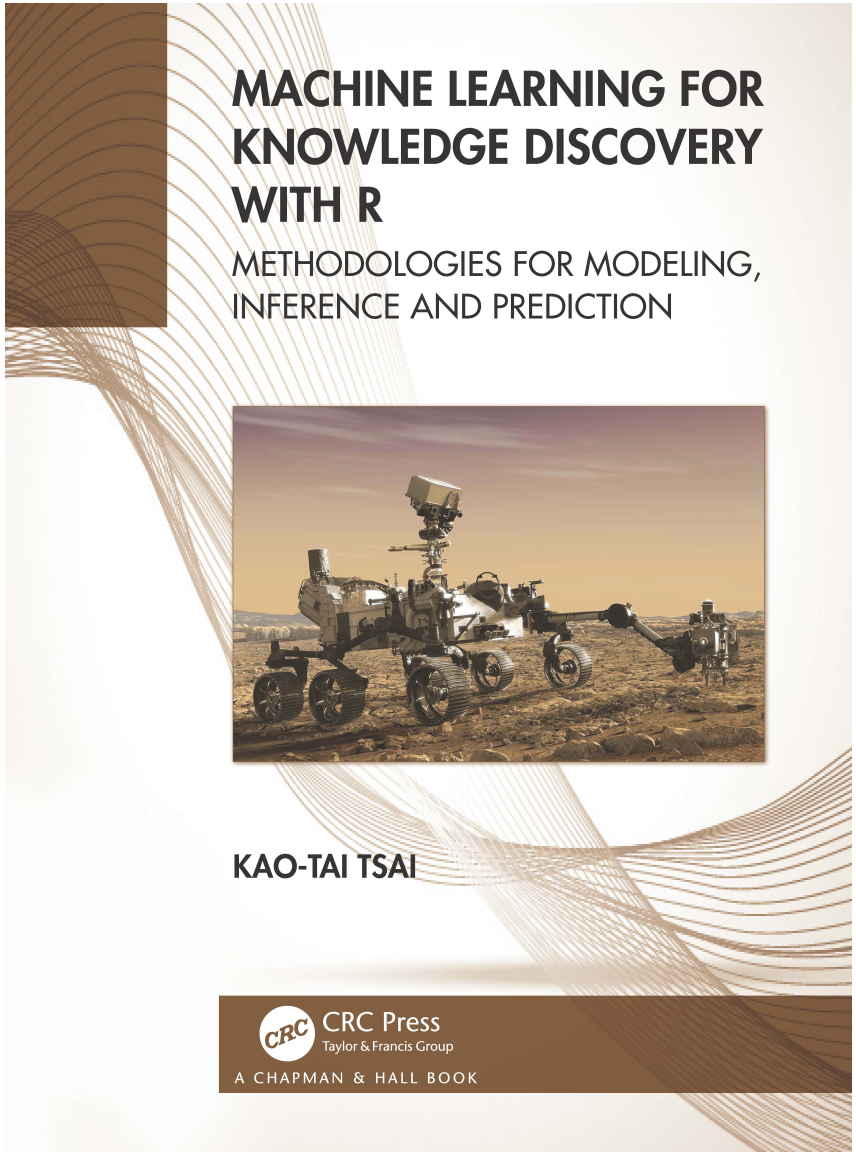
- In old America, the wise man advised "Young Man, Go West!"
- In old Polynesia, the wise man advised " Young Man, Go East!"
- In *The Sound of Music*, the wise Mother Abbess advised Maria:

Climb every mountain,
Ford every stream,
Follow every rainbow,
with All the love you can give,
Till you find your dream.

The 7 habits of highly effective people

- Be proactive,
- Begin with the end in mind,
- Put first things first,
- Think win-win,
- Seek first to understand, then to be understood,
- Synergize,
- Sharpen the saw.

2yes


MACHINE LEARNING FOR KNOWLEDGE DISCOVERY WITH R
METHODOLOGIES FOR MODELING, INFERENCE AND PREDICTION
KAO-TAI TSAI
CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK

hbr