How mathematicians deal with new trends like big data, etc.

Peter Bühlmann

Seminar für Statistik, ETH Zürich

New trends: some history

- Felix Klein (1872): “Erlangen program” (Vergleichende Betrachtungen über neuere geometrische Forschungen)
- David Hilbert (ICM Paris, 1900): “The Hilbert problems” (10 problems in 1900; 23 in total)
- Henri Poincaré (1908): “L’avenir des mathématiques”
- Millennium Prize Problems (Clay Math. Institute, 2000) (7 problems; 6 are still open)

did the trends initiate

new activities for (specific) problem solving ?
new frameworks, new theory, new connections ?

Felix Browder, AMS President (1999-2000),
Reflections on the Future of Mathematics (2002):

"The most important area of bisociation lies in the interaction between mathematics and the sciences. One major external influence on mathematics is, of course, the computer."*

*bisociation: the creative act, which ... always operates on more than one plane (Koestler, 1964)

the theme today is slightly different:

how mathematicians deal with new trends?

I am not qualified to discuss the question in general, with a many mathematicians working in diverse exciting fields...

Felix Browder, AMS President (1999-2000),
Reflections on the Future of Mathematics (2002):

"The most important area of bisociation lies in the interaction between mathematics and the sciences. One major external influence on mathematics is, of course, the computer."*

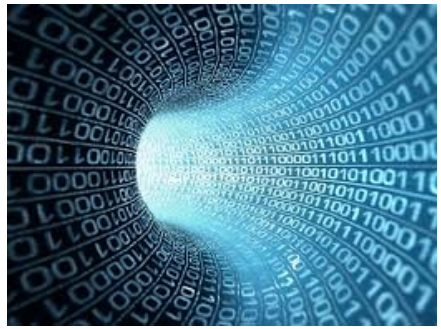
*bisociation: the creative act, which ... always operates on more than one plane (Koestler, 1964)

the theme today is slightly different:

how mathematicians deal with new trends?

I am not qualified to discuss the question in general, with a many mathematicians working in diverse exciting fields...

Big Data: mathematics, applied mathematics & statistics



data data data Big Data: massive amounts of data

- because we have better measurement devices covering a huge range of fields, beyond engineering and natural sciences (e.g. “social networks”, ...)
- a hope for high societal value and business value

“modern” (exploratory) data analysis: Tukey (1960’s): statistics
huge data sets: Huber (1985), Wegman (1995), Donoho (2000): statistics
new data analysis: Gromov (1998), Carlsson (2009): mathematics
large-scale data: von Mehring et al. (2002), Kanehisa (2011): comp. biology
...
Big Data: ... Sutton (1998): comp. science/machine learning

$\rightsquigarrow$ the creation of Data Science !

... the “trend created” Data Science !

in the past: demand in applications broadened mathematics, leading to new fields such as

- Applied Mathematics
- Statistics
- Computational Sciences

is Data Science really of the same order of magnitude?

at least:
a major current trend, involving a huge number of scientists!

... the “trend created” Data Science !

in the past: demand in applications broadened mathematics, leading to new fields such as

- Applied Mathematics
- Statistics
- Computational Sciences

is Data Science really of the same order of magnitude?

at least:
a major current trend, involving a huge number of scientists!

main aspect for mathematics and statistics:

analyzing the data; efficient and optimal information extraction

how do mathematicians deal with this?

"Tukey (1960s) recognized Data Analysis as an emerging discipline, distinct from Mathematical Statistics and requiring its own literature"

(Donoho; 2000, AMS "Math. Challenges of the 21st Cent."):

$\rightsquigarrow$ a separation:

Anglo-Saxon "culture": statistics separated from mathematics

Anglo-Saxon "culture": biostatistics separated from statistics

(but reverse happens as well: comp. science and statistics move towards each other)

a call for mathematics: *"... Data Analysis movement has insinuated itself into every branch of society ... And the missing ingredient in facing those problems, it seems to me, is mathematics."* (Donoho, 2000)

a call for inter-“disciplinary” communication:

because interaction with others opens new paths for exciting mathematics/statistics

examples (w.r.t. “Big Data”)

- Fast Fourier transform (Cooley and Tukey, 1965)
- Gibbs sampling (Geman and Geman, 1984)
- Compressed sensing (Donoho and Huo, 2001; Candes and Tao, 2005)
- many interesting mathematical works on high-dimensional analysis and approximation, topological and geometric invariants, computational-statistical trade-off,..

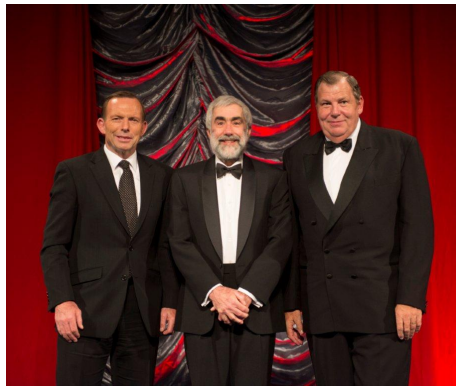
a call for inter-“disciplinary” communication:

because mathematicians/statisticians want to contribute to societal values

examples (w.r.t. “Big Data”):

- Michael Waterman

  for gene sequencing (Smith and Waterman, 1981) which was crucial for the Human Genome Project (finished 2003)



- Terry Speed

  Australian Prime Minister’s Prize for Science 2013:
  “For his contribution to making sense of genomics and related technologies (using statistics)”

Lots of other trends (arising from “external sources”)

see e.g. Browder (2000):

- the computer: numerical methods and simulation
- the computer: mathematical optimization
- the computer: algorithms, graph theory and discrete mathematics
- interaction between mathematics and the sciences: mathematical physics, mathematical biology, mathematical chemistry,...
- interaction between mathematics and the sciences: mathematical modeling
- ...

Summarizing thoughts

- historically, especially in
  “neighboring areas of” applied mathematics and statistics:
  fields and communities (departments) have split
  (recently, parts of Comp. Science and Statistics join forces)
  - can/should it be avoided?
- interaction and communication with external disciplines:
  - is it creating new conceptual frameworks?
  - a central part of mathematics?
- in need of mathematics ... in a huge amount of applications
  - how to respond to all the needs,
    given the small size of the mathematics community?
  - should we lead/co-lead collaborative efforts?
    should we “grab the opportunity” and become more highly visible?
- interaction opens new paths for exciting mathematics
  - can we improve the communication
    within and across the border of mathematics?