'Models of Data' Fifty Years On

Paul Humphreys

University of Virginia

March 9 2012

Some Main Themes of 'Models of Data'

•The comparison of empirical theories with data requires a hierarchy of models of different logical type
•'Theoretical notions are used in the theory which have no directly observable analogue in the experimental data'
•Models of the data are possible realizations of the data constrained by theories of the experiment

For greater generality, I shall call models of data 'models of data generating conditions', where the generating conditions can be traditional experiments, randomized trials, computer simulations, or some other source.

Types of Data

Question 1: What are the differences between data gained from empirical observation and experiment on the one hand, and data from computer simulations on the other?
Question 2: What is the nature of the relations between the models of data and the actual data?
Question 3: What do the answers to Questions 1 and 2 say about the nature of applied mathematics?

Suppes' Running Example: Linear Response Theory

Two Axioms

If $P(E_{i,n}A_{i',n}x_{n-1}) > 0$, then
$P(A_{i,n+1}|E_{i,n}A_{i',n}x_{n-1}) = (1-\theta)P(A_{i,n}|x_{n-1}) + \theta$

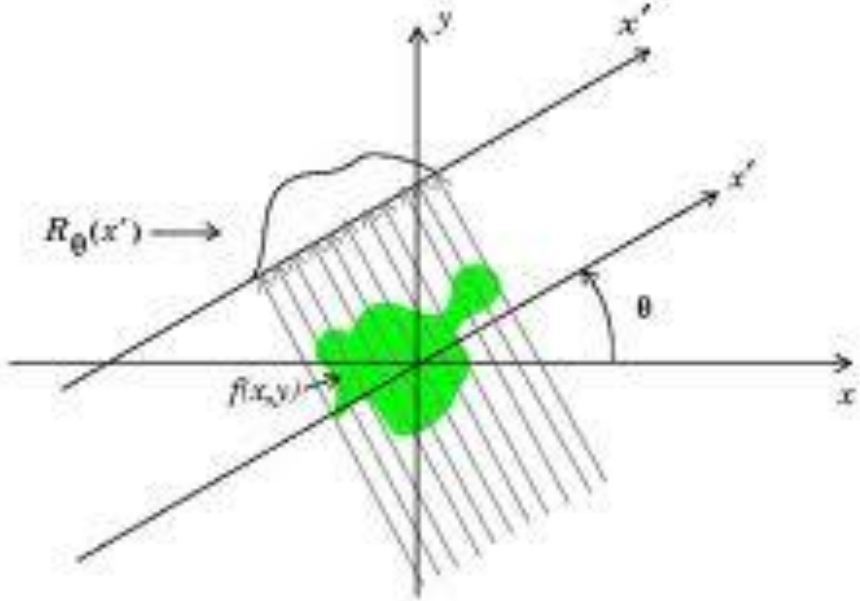
If $P(E_{j,n}A_{i',n}x_{n-1}) > 0$ and $i \neq j$, then
$P(A_{i,n+1}|E_{j,n}A_{i',n}x_{n-1}) = (1-\theta)P(A_{i,n}|x_{n-1})$

Central question: Given a model of the data, is there a linear response model satisfying goodness of fit relations?

Computed Tomography Instruments

An image is a function g defined on a plane where $g(x, y)$ represents the value of some property possessed by the target.

Here the property of interest is the X-ray attenuation coefficient.



y
x'
x'
$R_\theta(x')$
θ
$f(x,y)$
x

Filtered Backprojection

Consider the detector frame oriented at an angle θ to the target frame

$p(r, \theta) = \int_{y=-\infty}^{\infty} \int_{x=-\infty}^{\infty} g(x, y)\delta(xcos(\theta) + ysin(\theta) - r)dxdy$ is the Radon transform of g over the projection ray at r and

$g(x, y) = \int_0^{\pi} p(r, \theta) * h(r)d\theta$
is the inverse Radon transform applied to the convolution of the projections with a filter h.

Comptational Processes

Using discrete versions of the relevant functions:

- Convolve the spatial projections with a filter
- Fourier transform into frequency domain
- Compute convolutions as products in frequency domain
- Inverse Fourier transform back to spatial domain
- Compute inverse Radon transforms

Question 1

Question 1: What are the differences between data gained from empirical observation and experiment on the one hand, and data from computer simulations on the other?

Background: Models of the instrument include those for physical processes (Compton scattering, detector inefficiencies, beam hardening, noise, and many others), mathematical approximations (discrete functions, finite Fourier transforms, filter truncations, etc), and computational approximations (truncation error, inverse Fast Fourier transform approximations, trade-offs between computational time and accuracy, etc)

Questions 2 and 3

Question 2: What is the nature of the relations between the models of data and the actual data?
Question 3: What do the answers to Questions 1 and 2 say about the nature of applied mathematics?

Applied Mathematics and Applying Mathematics

'From a conceptual standpoint the distinction between pure and applied mathematics is spurious – both deal with set-theoretical entities, and the same is true of theory and experiment'. (Suppes, 'Models of Data', p. 260)

Physical Data and Physical Computations

CT instruments are a hybrid of physical data generation and automated data processing. More importantly, the computational processes after the detector counts have been made are both purely automatic and (arguably in some cases) concrete, not abstract.

Thanks Pat and Happy Birthday!