Modelling the citation process

Begeleider: Richard Gill

Scientific publication practices are currently under scrutiny (Peter Nijkamp affair; "Science in Transition" movement, http://www.scienceintransition.nl). In this project we will statistically model citation statistics of local mathematicians. Why do citation counts appear to follow a log-normal distribution? Is it a coincidence that μ and σ are almost equal? Is there a connection with the Zipf law? Is there a connection with the limiting distribution of population size in a Galton-Watson branching process? (Citations breed more citations).

It is well known that citation practices differ enormously over different fields of science. Moreover, the most important factor determining how many citations a particular researcher has obtained is their age. The infamous H-index (Hirsch 2005), is often refered to as the age-index. A researcher has H-index *n* if they have published *n* papers which each received *n* or more citations, while all their remaining papers got less than *n* citations.

Nils Lid Hjort (Oslo) recently proposed a new "*abc* model" which is a bit more general than the log normal. Workplan: fit Hjort's *abc* model to about ten different Google Scholar profiles (including those of Nils Lid Hjort and Richard Gill, and several other local colleagues), and then consider aspects like (1) is the model fitting OK, in each case, via the quantile equations fitting, (2) are there clear outliers or other discrepancies, (3) when we look at the (*a*,*b*,*c*) estimates, do the (*b*,*c*) ones look at least not unreasonably dissimilar (the a parameter being related to production volume, i.e., expected to be very different from profile to profile).

Xavier Gabaix: Zipf's Law for Cities: An Explanation
The Quarterly Journal of Economics, Vol. 114, No. 3 (Aug., 1999), pp. 739-767

J. E. Hirsch: An index to quantify an individual's scientific research output
PNAS 2005 102 (46) 16569-16572

Filippo Radicchi, Santo Fortunato, and Claudio Castellano: Universality of citation distributions: Toward an objective measure of scientific impact
PNAS 2008 105 (45) 17268-17272

Peter T. Breuer and Jonathan P. Bowen: Empirical Patterns in Google Scholar Citation Counts
Preprint, Academia.edu

http://rpubs.com/gill1109/11755; http://rpubs.com/gill1109/11774