Journal of Informetrics 4 (2010) 313–319

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

On reliability and robustness of scientometrics indicators based on stochastic models. An evidence-based opinion paper[☆]

Wolfgang Glänzel[a,b]

[a] Katholieke Universiteit Leuven, Centre for R&D Monitoring and Dept. MSI, Leuven, Belgium
[b] Hungarian Academy of Sciences, Institute for Research Policy Studies, Budapest, Hungary

ARTICLE INFO

Article history:
Received 10 December 2009
Received in revised form 26 January 2010
Accepted 26 January 2010

Keywords:
Stochastic models
Scientometric indicators
Statistical reliability
Ranking
h-Index

ABSTRACT

In the present paper we give an overview over the opportunities of probabilistic models in scientometrics. Four examples from different topics are used to shed light on some important aspects of reliability and robustness of indicators based on stochastic models. Limitations and future tasks are discussed as well.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

The terms scientometrics, bibliometrics and informetrics are usually explained as *the application of mathematical and statistical methods to information and communication processes* in different contexts (cf., Gorkova, 1988; Nacke, 1979; Nalimov & Mulchenko, 1969; Pritchard, 1969; Tague-Sutcliffe, 1992). The creation and application of mathematical, notably stochastic models to scientometrics and related fields seems therefore quite obvious. In particular, the links created by co-authorship relations, by received and given citations form complex networks of scientific communication which can best be described and analysed with the help of mathematical tools.

The application of stochastic models and probability distributions has the following important advantages (Glänzel, 2008).

1. It provides mathematical interpretations beside the scientometric ones. This means a more general notion of phenomena with the opportunity of extensions and generalisations through the choice of appropriate models, even beyond our field. Mathematical meaning and interpretation of scientometric measures can be given by parameters and statistical functions.
2. It helps understand complex structures such as communication networks. Although deterministic network models also allow randomness, the use of probabilistic network models such as Bayesian networks opens new perspectives, above all, concerning inference and learning.

☆ This paper is based on invited lectures delivered at the Workshop "Modelling science—Understanding, forecasting, and communicating the science system" held in Amsterdam, the Netherlands, on 6–9 October, 2009 and the "Fifth International Conference on Webometrics, Informetrics, and Scientometrics & Tenth COLLNET Meeting" held in Dalian, China, on 13–16 September, 2009.

E-mail address: wolfgang.glanzel@econ.kuleuven.ac.be.

1751-1577/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.joi.2010.01.005

3. It provides information about statistical reliability, random errors and confidence intervals for indicators.
4. It allows predictions concerning the expectation and probability of future events.

Scientometric phenomena to be quantified and measured can often be expressed by non-negative integer- or real-valued random variables. Most scientometric indicators can thus be defined as statistical functions such as mean values, quantiles, relative frequencies, rank statistics, Hirsch-type statistics. Therefore, we will focus on the last two issues in the above list since the other two questions are too general to be discussed here.

2. Reliability and robustness of scientometric indicators based on stochastic models

Deterministic as well as probabilistic models are used to describe patterns and processes of scholarly communication. First, in its pioneering days, scientometrics has adopted laws and models from other, often not even related fields (e.g., the model of *radioactive decay* for obsolescence of literature by Gosnell, 1944, models from quantitative linguistics for bibliometric rank frequencies by Zipf, 1949, or the theory of *intellectual epidemics* as a model of scientific communication by Goffman & Newill, 1964). Most of them were deterministic ones. However, scientometricians have soon recognised that own, specific models are needed to describe the observed phenomena in an adequate manner. The presence of latent variables and dependency relations among variables go far beyond the possibilities of the initially used deterministic models. The various factors influencing publication activity, citation impact or collaboration affinity might just serve as an example. The interdependency among the variables such as the relation between collaboration and citation impact or between the choice of communication channels and visibility have become common places in our field. Nevertheless, it is not at all obvious if these dependencies also imply causality relations. In the first subsection we will, therefore, have a closer look at correlation and (in-)dependence issues in scientometrics.

2.1. Correlation between variables

One of the crucial issues in the analysis of empirical data and in modelling scientometric processes is the question of independence or (partial) interdependence of the underlying variables. This question is important for the possible separation of variables in practice or for finding appropriate interrelations between in building the model. Of course, the definition itself of variables can already imply certain interdependence, which can result in build-in biases. There are several methods to analyse (partial) interdependence in a set of variables. Regression analysis based on statistical correlation is beyond doubt one of the most popular methods. However, this method is often source of misinterpretations.

The following misinterpretation is perhaps the most fatal one. Correlation, which is from the mathematical viewpoint a symmetric relation, is sometimes confused with causality.

The second misunderstanding concerns the fact that independent variables are uncorrelated but the reverse statement does not hold. Uncorrelatedness does not imply independence. In practice, it is often difficult to capture regularities behind weak correlation. The following example illustrates how relatively weak correlation can be mathematically expressed by a "weak" law.

Example. Several years ago, we analysed the possible interdependence between author self-citations (ξ) and foreign citations (ζ), where the two citation rates are considered random variables (Glänzel, Thijs, & Schlemmer, 2004). A linear regression analysis was used to test the hypothesis of independence, namely H_0: $P(\xi(t)=i,\ \zeta(t)=k)=P(\xi(t)=i)\cdot P(\zeta(t)=k)$ for all pairs i and $k \geq 0$. We know that the random variable $t = \sqrt{n-2}(r/\sqrt{1-r^2})$ has a Student distribution with $n-2$ degrees of freedom, where n is the number of publications and r is the correlation coefficient. For different citation windows and different regression models we have observed weak correlations with $r^2 \sim 0.20$. Nevertheless, we had to reject H_0, that is, self-citations and foreign citations cannot be considered independent. On the other hand, the weak correlation suggested that neither the linear nor the power-function model can be accepted. In other words, individual self-citation rates cannot explicitly be expressed with the help of foreign citations alone.

We could conclude that although both variables cannot be considered independent, there is no unique functional relationship of linear, power, logarithmic, exponential or polynomial or any other form between the two types of citations. Therefore, we decided to analyse a weaker type of relationship, namely between the conditional expectation $E(\xi(t)|\zeta(t))$ and $\zeta(t)$. The condition $E(\xi(t)|\zeta(t)) \equiv E\xi(t)$ (i.e., $r=0$) is necessary but not sufficient for independence. Thus we have supplemented the first analysis by a second regression analysis where we expected to find a explicit relation between the two variables. Because of the properties of conditional expectations there exists an appropriate function f, such that $E(\xi(t)|\zeta(t))=f(\zeta(t))$. We have assumed a power model and have chosen $f(x)=C\cdot(x+d)^{\beta}$ with C, d and β being appropriate positive real parameters. The correlation of $r^2=0.992$ proved to be strong for the assumed model with parameters $C=1$, $d=0.25$ and $\beta=0.547$ (cf. Fig. 1). In verbal terms, the first regression model based on the two variables self-citations and foreign citation suggested to reject independence but did not support any explicit relationship, either. Under weaker assumptions we have found such a unique relation based on conditional expectations, namely $E(\xi(t)|\zeta(t)) \sim (\zeta(t)+1/4)^{1/2}$. This case is quite typical of scientometric phenomena as it expresses *weak but measurable interdependence without any causal statements*. We will see in a later subsection that this form of regression is of great importance in scientometrics.

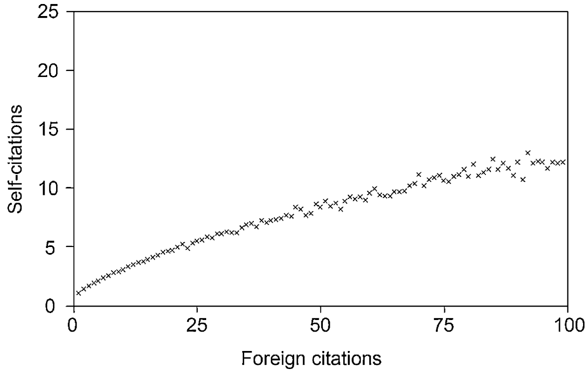

Fig. 1. The plot of foreign citations vs. mean self-citation rate (1992–2001). Recreated from Figure 3 bottom (Glänzel et al., 2004) based on the original data with permission by Scientometrics.

2.2. Statistical reliability of comparisons

2.2.1. Asymptotic normality of scientometric indicators

Authorship, publication activities, references, citations and other links as well as the co-occurrences of these links can be expressed by random variables and/or corresponding dependence relations. Most scientometric phenomena can be modelled using non-negative integer-valued random variables but derivatives can take real values as well. The application of most statistical methods are approximate solutions. One reason is ambiguity and uncertainty (cf. Bookstein, 1997). For instance, the independence of the random variables under study is often not guaranteed. Scientific collaboration, author co-affiliation, co-citations or subject co-assignment may in fact distort independence. The other reason is that most statistical techniques require underlying normal distributions to obtain 'exact' solutions.

Indeed, the Gaussian normal distribution arises in many areas of statistics. 'Normality' is the basis and the condition of many statistical tests. If a statistical sample follows a normal distribution, then the observations should be symmetrically distributed around the sample mean and the standard deviation can be used to determine a tolerance threshold for observations. However, this is often not the case in scientometrics. Most distributions here are discrete and extremely skewed so that the majority of the observations are found below the sample mean and the rest of the sample elements are located in the long tail of the distributions.

Nevertheless, the *central limit theory* guarantees the asymptotic normality of sample means even if the underlying distribution is discrete and skewed, provided the distribution belongs to the *domain of attraction* of the Gaussian distribution.

Definition. Let $X_1, X_2, \ldots, X_n$ be a sequence of *n independent and identically distributed random variables* and a_n and $b_n > 0$ suitable sequences of constants such that the distribution of the normalised sequence of partial sums $Z_n = \left(\sum_{i=1}^{n} X_i - a_n\right)/b_n$ converges weakly, as n tends to infinity, to a non-degenerate distribution function $F(x)$. We say that the common distribution of the random variables X_i belongs to the *domain of attraction* of the (stable) distribution F.

Theorem (Central limit theorem). *Let $X_1, X_2, \ldots, X_n$ be a sequence of n independent and identically distributed random variables with finite expectation μ and variance $\sigma^2 > 0$. Then the distribution of the random variable $Z_n = \left(\sum_{i=1}^{n} X_i - n\mu\right)/\sigma\sqrt{n}$ converges weakly to the standard normal distribution $N(0,1)$ as n tends to infinity*.

Under certain conditions (e.g., Lindeberg condition), a weaker form of the central limit theorem, where identical distribution is not required, holds. However, the following approximation holds as well.

Theorem. *Let $X_1, X_2, \ldots, X_n$ be a sequence of n independent random variables with finite expectations μ_i and variances ${\sigma_i}^2 > 0$. Then $Z_n = \left(\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mu_i\right)/\sqrt{\sum_{i=1}^{n} \sigma_i^2}$ has a limiting distribution function which approaches a normal distribution*.

Further information about attraction domains of distributions and weaker conditions of the central limit theorem can be found, e.g., in Rényi (1962), Hazewinkel (2002) or, more recently, Gut (2009).

As a consequence of the central limit theorem, the sample mean of random variables $\bar{x} = \left(\sum_{i=1}^{n} X_i/n\right)$ with *any distribution* belonging to the attraction domain of the normal distribution is approximately normally distributed. The standard deviation of the sample mean equals the standard deviation of the common distribution of the sample elements divided by the square root of the sample size. Furthermore, confidence intervals can be given for sample means and statistical tests originally designed for statistics taken from normally distributed populations can thus be applied, provided that the sample size is large enough. A *t*-test should not be applied to the comparison of two sample means since the standard deviation of the random variables is not known and is not assumed to be identical in the case of different samples. However, a simple *Welch* test can be applied instead, provided the sample size amounts to about 30 or more (see, e.g., Sawilowsky, 2002).

In particular,

$$w = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$$

has *approximately* a Student distribution, where $\bar{x}_i$ are the sample means, n_i the sample sizes, s_i are estimators of the standard deviations that are *independent* of $\bar{x}_i$. For sample sizes above 50 or even 100, the Student distribution can in practice be approximated by a normal distribution. This is very often the case in scientometric practice. If one of the two variables is actually constant, that is, a given fixed value, the standard deviation of the remaining random variable appears in the denominator.

This approach can be extended to other statistics as well. According to the Glivenko–Cantelli theorem, the empirical distribution converges to the underlying theoretical one with probability 1. The relative frequency $\hat{p}$ is an unbiased estimator of the corresponding probability p. For instance, the standard error of the share (e.g., of cited or uncited papers) $\hat{p}$ can be calculated analogously to the sample mean. In particular, we have $E(\hat{p}) = p$ and $D(\hat{p}) = \{p(1-p)/n\}^{-1/2}$. The standard deviation of the relative frequency is a decreasing function of the sample size.

In this context we have to stress that mean values and relative frequencies are unbiased estimators for the expectation and the corresponding probabilities. Their use in scientometrics is therefore correct and not just a 'workaround'. Whenever possible, their standard deviation should be given in order to be able to judge their tolerance and to allow approximate significance tests as has done first by Schubert and Glänzel (1983) for ISI Impact Factors.

Another consequence of the above considerations is the effect that seemingly large deviations between indicators sometimes prove to be not significant. In terms of *ranking* according to scientometric indicators, ties might occur where indicators otherwise take different values. Specific problems of ranking will be discussed in the following section.

2.3. Ranking

Before we tackle the question to which extent reliable and reproducible ranking lists are at all possible, we attempt to clarify the notion of ranking by presenting the following comprehensible but nevertheless precise definition (see, Glänzel & Debackere, 2009).

Definition. Ranking is positioning comparable objects on an ordinal scale based on a non-strict order relation among (statistical) functions of measures or scores associated with those objects.

Thus ranking can, in particular, be considered a multivariate comparison of variables in a given set by defining a non-strict order relation on this set.

These functions or variables, usually based on evaluation, are called indicators. Different indicators X_i representing different aspects usually form components of a composite indicator Y which is used as the basis of the ranking, particularly,

$$Y = \sum_{i=1}^{n} \lambda_i X_i \text{ with } \lambda_i \text{ being weights and } \sum_{i=1}^{n} \lambda_i = 1. \tag{1}$$

Generalisation

If some k subsets consisting of n_j indicators X_i each form a partition of the entire space, a set of (composite) indicators Y_j (j = 1, 2, ... k) can be defined for a given k (with $1 \le k \le n$) in a similar manner.

$$Y = \sum_{i=n_{j-1}+1}^{n_j} \lambda_i X_i \text{ with } n_0 = 0,\ n_k = n \text{ and } \sum_{i=n_{j-1}+1}^{n_j} \lambda_i = 1 \tag{2}$$

Remark. For $k = 1$, the previous case is obtained with $Y_1 = Y$ according to Eq. (1), while for $k = n$ the partition is formed by the original individual variables with $j = i$, where each subset contains exactly one element, i.e., we have $Y_i = X_i$.

General problems in using composite indicators

Problems might occur if several variables are (at least partially) bundled into a single measure. These problems are related to the following issues.

- "Random errors" of statistical functions;
- Possible interdependence of components;
- Altering weights can result in different ranking outcomes;
- Results might be obscure and irreproducible;
- Information loss by crashing the multi-dimensional space into linearity.

Assume that we have solved the latter four problems and have obtained valid and robust indicators for ranking, we are still faced with the problem of ties which is clearly a disadvantage of linear scoring. Instead of scalar scoring (i.e., positioning objects on an ordinal scale according to their indicator values Y_j) objects with similar scores can be grouped into classes.

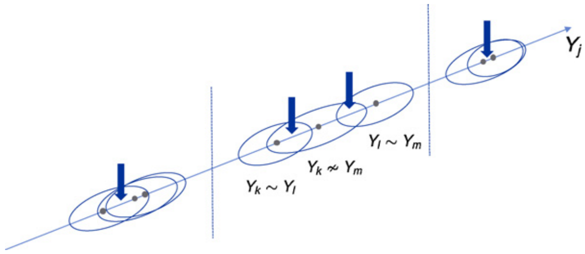

Fig. 2. Pinprick and separation problem (sketch).

Classes might be pre-defined according to some criteria (e.g., poor, fair, good, very good, excellent, etc.), or be self-adjusting according to some mathematical rules. The most obvious advantage of pre-defined classes is that the number of classes can be chosen according to the needs and that the same categories can be used for all dimensions if a "multi-dimensional" approach is applied to the ranking exercise. By contrast, the arbitrariness and difficulty in finding common criteria for classification.

Self-adjusting classes are less arbitrary since they adjust themselves whenever the underlying system changes. This property is, for instance, important for longitudinal approaches. On the other hand, difficult algorithms form the disadvantage of this approach. Transforming linear scoring into self-adjusting groups remains a challenge since the underlying relations are reflexive, symmetric but not transitive (see 'pinprick and separation problem' presented in Fig. 2). The task in Fig. 2 is to fix objects represented by dots on a line based on tolerance intervals (represented by ellipses) using the minimum number of pins (represented by arrows). To fix the three objects in the centre, one needs two pins; however if the variance of object Y_l were somewhat greater the problem could be solved with one pin less.

2.3.1. The special case of the h-index

Ranking according to Hirsch-related indices is based on one single indicator and forms a special case of ranking exercises. Therefore most of the problems occurring in the context of composite indicators do not apply here. On the other hand, the mathematical–statistical properties of the *h*-index and its derivatives have been studied only recently. The most recent results prove the robustness of this indicator but also its low discriminatory power (cf. Glänzel, 2009).

The *asymptotic normality* of the *h*-index has been shown by Beirlant and Einmahl (2007). This property has been proved under very weak conditions. In the following I give the version for *Paretian* distributions, the class of distributions, which are of the most practical importance in our field.

Proposition (Beirlant & Einmahl, 2007). *Assume that if the elements of the original sample* $\{X_i\}_{i=1,...,n}$ *have a Paretian distribution, i.e.,* $G(x) := 1-F(x) = x^{-\alpha}\ell(x)$*, where F is the common probability distribution and* ℓ *is a slowly varying function such as*

$$\lim_{x\to\infty}\frac{\ell(ux)}{\ell(x)} = 1 \text{ for all } u > 0.$$

If *F* satisfies the *von Mises condition*, i.e., if $\lim_{x\to\infty} -xG'(x)/G(x) = \alpha$ and $\hat{\alpha}$ is a consistent estimator for α, then

$$\frac{1+\hat{\alpha}}{\sqrt{\hat{h}}}(\hat{h}-h)\xrightarrow{D}N(0,1),$$

where h and $\hat{h}$ are the theoretical and empirical *h*-index, respectively.

Note that the distribution of the empirical *h*-index may belong to the domain of attraction of the normal distribution even if the underlying Paretian distribution does not, i.e., if $\alpha \le 2$ (see Barcza & Telcs, 2009).

Critical values and confidence intervals can thus readily be given for any confidence level p and any parameter α. The confidence interval for the empirical *h*-index is obtained from the previous proposition.

$$\left(h + \frac{c_p^{*2}}{2} - c_p^*\sqrt{h+\frac{c_p^{*2}}{4}},\, h + \frac{c_p^{*2}}{2} + c_p^*\sqrt{h+\frac{c_p^{*2}}{4}}\right), \text{ with } c_p^* = \frac{c_p}{(1+\alpha)},$$

where $p = 2\Phi(c_p) - 1$ and Φ is the cumulative distribution function of the standard normal distribution. Table 1 presents the lower and upper bounds for selected *h*-indices and three different α values. The first case ($\alpha = 1$) refers to the original Lotka-type distribution, the second one ($\alpha = 2$) is frequent in scientometric practice (see Pao, 1986; Schubert & Glänzel, 2007) and the third case ($\alpha = 10$) can already be approximated by an exponential distribution. The low discriminative power of this indicator for small α values is blatantly obvious.

Table 1
Some confidence intervals for empirical h-indices with different α parameters ($p = 0.95$).

h-index	Lower bound	Upper bound	Lower bound	Upper bound	Lower bound	Upper bound
	$\alpha = 1$		$\alpha = 2$		$\alpha = 10$	
20	17	24	18	23	20	20
25	21	30	22	28	25	25
30	26	35	27	33	30	30
35	30	41	32	39	34	36
40	35	46	37	44	39	41
50	44	57	46	54	49	51
100	91	110	94	106	99	101

2.4. Predictive aspects

Scholarly communication patterns are subject to changes in time. Changing scientific "productivity", dissemination and ageing of information, in general, and of scientific literature, in particular, are typical examples for these processes. Stochastic processes are the preferred models for describing scientometric processes (e.g., Burrell, 2003, 2005; Glänzel & Schoepflin, 1994; Schubert & Glänzel, 1984). Birth-processes can be used for the analysis of the regression function of the increments on the process. The empirical version of this function was also called mean value function of a process for simplicity. This approach was used by Glänzel and Schubert (1995) to characterise future citation rates. In principle, this model is appropriate to describe an author's future publication activity as well. It was shown that the above-mentioned regression function $M(s,t)$ for a negative-binomial birth-processes is linear. In other words,

$$M(s, t) = E(X(t) - X(s)|X(s) = i) = u \cdot i + v,$$

where $X(t)$ denotes the citation process and $i \geq 0$ the number of citations and u and v two positive real parameters. The standard deviation of its empirical value, denoted by $M^*(s,t)$, can be approximated by the following expression (Glänzel, 1997).

$$D(M_i^*(s, t)) \approx \frac{\sqrt{(v/u + 1)\{1 - (E(X(s)) + v/u)/(E(X(t)) + v/u)\}}}{\sqrt{n \cdot P(X(s) = i)} \cdot (E(X(s)) + v/u))/(E(X(t)) + v/u)}$$

According to this equation, reliable predictions of future citation rates are possible, if the initial reference period is close to the prediction period. The "goodness-of-prediction" increases with the sample size and the length of the initial period and decreases with the length of the interval to be predicted. Beyond providing "predictive indicators", this model also sheds light on important validity aspects of traditional citation indicators. Thus the choice of three- to five-year citation window still allows the evaluation of recent research results, and is usually long enough to determine future citation impact.

Furthermore, several life-time related indicators can be derived from stochastic models as well. Citation half-life, Price-type indices or obsolescence indicators are just some examples. More opportunities are offered by the application of Markov stopping times (Burrell, 2001, 2002; Glänzel, 1992). This model is of special interest in the context of the analysis of first-citation distributions (cf. Rousseau, 1994 and Egghe and Rao, 2001). Furthermore, stopping times can be used to construct response indicators (Schubert & Glänzel, 1986) as well.

Let T_i ($i = 1, 2, 3, \ldots$) denote the shortest time t_n ($n = 0, 1, 2, \ldots$) during which the papers have received exactly i citations (i.e., $T_i = \min\{t_n: X(t_n) \geq i\}$). Random variables of this type are called "stopping times". From the definition we obtain the following property.

$$P(T_i = t_n) = P(X(t_n) \geq i) - P(X(t_{n-1}) \geq i) \quad \text{and} \quad P(T_i = t_0) = P(X(t_0) \geq i)$$

T_i can take the value $+\infty$ with positive probability: $P(T_i = +\infty) = P(X(\infty) < i)$. Hence $E(T_i) = \infty$ follows. This implies that the mean value cannot be used as indicator. Moreover, citations are only known within a finite citation window. However, there is an alternative solution, namely the application of an appropriate transformation as, for instance, was suggested by Schubert and Glänzel (1986) for first-citations. This can be generalised as follows

$$T_i^e = -\ln E(e^{-T_i}).$$

This transformation has the advantage that citations beyond a citation window of about five years can be neglected since e^{-x} quickly converges to 0 with growing x.

Although Burrell (2002) has studied the case of an underlying Poisson process, the mathematical–statistical properties of stopping-time indicators and their relationship with more general models for citation processes are not yet sufficiently analysed. It is certainly one of the future tasks to validate and improve the reliability of reception indicators and to make them fit for comparative studies and ranking.

3. Conclusions

In its pioneering days, scientometrics has adopted laws and models from other, often not even related fields to describe observed phenomena. The exponential and logistic growth model to describe the growth of literature, the model of radioactive decay for the ageing of information or epidemic models for dissemination of information may just serve as an example. These models have been supplemented by generic scientometric approaches but most of them remained deterministic ones (e.g., Lotka's and Bradford's Law). Their disadvantages were that most of them were formulated as "natural laws" not allowing for the influences of those factors which play an important part in scholarly communication. Probabilistic models have the potential to solve this problem. They proved also to be useful to measure indirect or latent effects which can often be expressed by conditional measures, that is, conditional probabilities or moments. Good examples are the interdependence between author self-citations and foreign citations or the prediction of future citation rates.

When scientometrics shifted from a tool in scientific information towards a tool for research evaluation, a second issue became crucial: the statistical reliability of indicators. Only the probabilistic approach could meet this requirement, however, sometimes only approximately. In turn, new problems arose for the application of indicators for ranking entities according to these measures.

We could also see that robustness does not necessarily go with discriminative power. The example of Hirsch-type indices convincingly illustrates that although these indicators are relatively insensitive to changes in the underlying distributions, their discriminative power is disappointingly low.

We should also keep in mind that scientometrics has its own peculiarities which have to be taken into account when creating and applying appropriate models. We are still at the very beginning of the process of the creation of such models and its translation into reality.

Acknowledgement

Herewith I would like to thank Ronald Rousseau for valuable comments on a previous version of this paper.

References

Barcza, K., & Telcs, A. (2009). Paretian publication patterns imply Paretian Hirsch index. *Scientometrics*, *81*(2), 513–519.
Beirlant, J., Einmahl, J. H. J. (2007). *Asymptotics for the Hirsch Index*. CentER Discussion Paper #2007-86. Accessible at: http://arno.uvt.nl/show.cgi?fid=65780.
Bookstein, A. (1997). Informetric distributions. 3. Ambiguity and randomness. *Journal of the American Society for Information Science*, *48*(1), 2–10.
Burrell, Q. L. (2001). Stochastic modelling of the first-citation distribution. *Scientometrics*, *52*(1), 3–12.
Burrell, Q. L. (2002). The *n*th-citation distribution and obsolescence. *Scientometrics*, *53*(3), 309–323.
Burrell, Q. L. (2003). Predicting future citation behavior. *Journal of the American Society for Information Science and Technology*, *54*(5), 372–378.
Burrell, Q. L. (2005). The use of the generalized Waring process in modelling informetric data. *Scientometrics*, *64*(3), 247–270.
Egghe, L., & Rao, I. K. R. (2001). Theory of first-citation distributions and applications. *Mathematical and Computer Modelling*, *34*(1–2), 81–90.
Glänzel, W. (1992). On some stopping times on citation processes. From theory to indicators. *Information Processing & Management*, *28*(1), 53–60.
Glänzel, W. (1997). On the reliability of predictions based on stochastic citation processes. *Scientometrics*, *40*(3), 481–492.
Glänzel, W. (2008). *Bibliometrics as a research field. Course script*. Belgium: KU Leuven.
Glänzel, W. (2009). The role of the h-index and the characteristic scores and scales in testing the tail properties of scientometric distributions. In B. Larsen, & J. Leta (Eds.), *Proceedings of ISSI 2009—the 12th international conference on scientometrics and informetrics* Rio de Janeiro, (pp. 120–130).
Glänzel, W., & Debackere, K. (2009). In C. Dehon, D. Jacobs, & C. Vermandele (Eds.), *On the "multi-dimensionality" of ranking and the role of bibliometrics in university assessment* (pp. 65–75). Ranking Universities, Editions de l' Université de Bruxelles.
Glänzel, W., & Schoepflin, U. (1994). A Stochastic Model for the ageing analyses of scientific literature. *Scientometrics*, *30*(1), 49–64.
Glänzel, W., & Schubert, A. (1995). Predictive aspects of a stochastic model for citation processes. *Information Processing & Management*, *31*(1), 69–80.
Glänzel, W., Thijs, B., & Schlemmer, B. (2004). A bibliometric approach to the role of author self-citations in scientific communication. *Scientometrics*, *59*(1), 63–77.
Goffman, W., & Newill, V. A. (1964). Generalization of epidemic theory. *Nature*, *204*(4955), 225–228.
Gorkova, V. I. (1988). Informetrics. In *Informatics, Vol. 10*. Moscow: VINITI.
Gosnell, C. F. (1944). Obsolescence of books in college libraries. *College and Research Libraries*, *5*, 115–125.
Gut, A. (2009). *An intermediate course in probability. Series: Springer texts in statistics* (2nd edition). Springer.
Hazewinkel, M. (Ed.). (2002). *Encyclopaedia of mathematics*. Berlin, Heidelberg, New York: Springer-Verlag. Accessible via:. http://eom.springer.de/default.htm
Nacke, O. (1979). Informetrie: Eine neuer Name für eine neue Disziplin. *Nachrichten für Dokumentation*, *30*(6), 219–226.
Nalimov, V. V., & Mulchenko, Z. M. (1969). Наукометрия. Moscow: Nauka.
Pao, M. L. (1986). An empirical examination of Lotka's law. *Journal of the American Society for Information Science*, *37*(1), 26–33.
Pritchard, A. (1969). Statistical bibliography or bibliometrics? *Journal of Documentation*, *24*, 348–349.
Rényi, A. (1962). *Wahrscheinlichkeitsrechnung*. Berlin: VEB Deutscher Verlag der Wissenschaften.
Rousseau, R. (1994). Double exponential models for first-citation processes. *Scientometrics*, *30*(1), 213–227.
Sawilowsky, S. S. (2002). Fermat, Schubert, Einstein, and Behrens–Fisher: The probable difference between two means when $\sigma_1 \neq \sigma_2$. *Journal of Modern Applied Statistical Methods*, *1*(2), 461–472.
Schubert, A., & Glänzel, W. (1983). Statistical reliability of comparisons based on the citation impact of scientific publications. *Scientometrics*, *5*(1), 59–74.
Schubert, A., & Glänzel, W. (1984). A dynamic look at a class of skew distributions. A model with scientometric applications. *Scientometrics*, *6*(3), 149–167.
Schubert, A., & Glänzel, W. (1986). Mean response time. A new indicator of journal citation speed with application to physics journals. *Czechoslovakian Journal of Physics*, *36*(1), 121–125.
Schubert, A., & Glänzel, W. (2007). A systematic analysis of Hirsch-type indices for journals. *Journal of Informetrics*, *1*(3), 179–184.
Tague-Sutcliffe, J. (1992). An introduction to informetrics. *Information Processing & Management*, *28*(1), 1–3.
Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Addison-Wesley.