Acta Applicandae Mathematicae **67**: 237–252, 2001.

Probabilistic Model for the Growth of Thesauri

MICHIEL HAZEWINKEL[1] and RIMANTAS RUDZKIS[2]
[1]*CWI, PO Box 94079, 1090 GB Amsterdam, The Netherlands. e-mail: mich@cwi.nl*
[2]*Institute of Mathematics and Informatics, Akademijos 4, 2600 Vilnius, Lithuania. e-mail: rudzkis@ktl.mii.lt*

(Received: 29 November 1999; in final form: 16 April 2001)

Abstract. The paper deals with a mathematical model which describes how the collection of key phrases (and key words) evolves as a field of science develops. The experimental material is based on statistical observations on the sets of key phrases which have been assigned to papers in representative major journals in the field in question. Asymptotic properties of the model are considered, as well as estimators for the parameters of the model with particular emphasis on estimators that can indicate at what stage a collection of key phrases can be assumed as complete for the field in question at a given moment in time.

Mathematics Subject Classifications (2000): 62Fxx, 68P10, 68P20.

Key words: thesaurus, index, growth of thesauri, probabilistic model, asymptotic property, statistical estimator, growth model.

1. Introduction

Let $\mathcal{F}$ be a field of science. Let K_t denote the set of key phrases that have been used in the field up to and including time t. The time t is thought of as a discrete time variable (though that is not necessary). At a time t_1 we start observing. Up to a time-scaling equivalence we can assume $t_1 = 1$. Denote the set of key phrases that has been observed up to (and including) time t by S_t, $S_t \subset K_t$. An important problem for practical aplications is to try to estimate K_t by the observations $S_1, S_2, \ldots, S_t$. This paper is concerned with this problem.

Here we introduce a definite probabilistic model for how the set of key phrases of a given field should develop; how the observation of key phrases takes place, and we consider the asymptotic properties of this model and ways to estimate its (asymptotic and other) parameters.

Let $X_t = \operatorname{card} K_t$, $Y_t = \operatorname{card} S_t$. In the model presented below, we give solutions to the following problems:

(1) Description of the asymptotic behaviour of the ratio Y_t / X_t as $t \to \infty$;
(2) Necessary and sufficient conditions for the probability $\mathbb{P}\{K_0 \subset S_t\}$ to converge to 1;
(3) Constructive determination or consistent estimates of the parameters that control the growth of K_t.

A very primitive version of the model was outlined in [5]. If in this model one replaces the stochastic variables by their averages, one obtains a differential equation that is actually explicitly solvable. This is, of course, an extremely crude and usually unjustifiable procedure. In this case, it turns out that it is asymptotically justified (even though the model is not linear).

The stochastic model seems to be of a new type and may have applications in other fields where one is dealing with an evolving population and/or evolving genera (the natural interpretation of a key phrase being a trait (as understood in biology and archeology)).

2. Mathematical Model

The basic assumptions of the model are as follows:

2.1. Assume that X_t, $t = 1, 2, \ldots$, is a random Poisson process, i.e., its increments $\Delta X_t = X_t - X_{t-1}$ are independent random variables (i.r.v.) distributed by the Poisson law. We use the notation: $P_\lambda(\cdot)$ for the Poisson distribution with the parameter λ, $N_t = \mathbb{E}X_t$, $\lambda_t = \Delta N_t$. Thus, the distribution of ΔX_t is P_{λ_t}. For simplicity of notation, we assume X_0 to be a determinate quantity, $X_0 = N_0 > 0$.

2.2. Let keywords be numbered according to their appearance time in increasing order, and let a keyword $w_k \in K_t$ and a nonnegative quantity W_k – the 'weight' of word w_k at the moment of its emergence (that reflects the scope of a denoted notion) – one-to-one correspondend to each natural number $k \leqslant X_t$. Suppose $W_1, W_2, \ldots$, are i.i.d. positive random variables with the distribution function F, independent of the sequence $X_{(\cdot)}$, $\mathbb{E}W_k \equiv 1$.

2.3. Let Q_t stand for the set of keywords that were observed at a time moment t ($S_t = \bigcup_{\tau=1}^{t} Q_\tau$), $A_{k,t} = \{w_k \in Q_t\}$. The probabilities of the random events $A_{k,t}$ depend on 'weights' $W_{(\cdot)}$ and the development of the system considered, i.e., on $I_t := \{K_1, \ldots, K_t, Q_1, \ldots, Q_{t-1}\}$. Assume that, for fixed $W(\cdot)$ and I_t, the events $A_{k,t}$, $k = 1, \ldots, X_t$, are conditionally independent and the equalities

$$\mathbb{P}\{A_{k,t} \mid I_t; W_{(\cdot)}\} = \min\left\{\frac{u_t W_k}{X_t}; 1\right\} \stackrel{\text{def}}{=} \pi_{k,t}, \quad k = 1, \ldots, X_t, \tag{1}$$

hold. (In fact, we assume that the keywords get into the observation samples independently with probabilities $\pi_{k,t}$, provided the history and weights for keywords are fixed. This is a quite weak assumption practically dictated by the actual way in which thesauri grow.) Here and in what follows, u_t is some deterministic function (that features the summary quantity of the journals observed at the time moment t) and the inequality $u_t < N_t$ holds. Note that $\mathbb{E}\,\mathrm{card}\, Q_t \cong u_t$.

In order to achieve greater modeling power, one could consider a more general model, where the quantities W_k (which are independent of time in formula (1))

are replaced by time functions $W_{k,t}$. For example, $W_{k,t}$ could be defined as certain functions of quantities W_k, k/X_t and $\gamma_{k,t} := \min\{\tau : w_k \in Q_{t-\tau},\ \tau > 0\}$ ($\gamma_{k,t} = 0$, if $w_k \notin S_{t-1}$). Proximity of k/X_t to 1 shows the novelty of word w_k, and a low value of $\gamma_{k,t}$ points to the fact that a small time period has elapsed since the last observation of w_k which increases the probability of a repeated observation of w_k in the nearest future relative to comments, references and the like that accompany scientific publications.

We restrict ourselves to model (1) in this paper. In addition, particular attention is paid to the special case where all weights are equal,

$$W_k \equiv 1. \tag{2}$$

At the first sight, model (2) does not reflect reality, since in scientific literature the frequency of observations of various keywords apparently differs. However, if K_t denotes not the whole totality of keywords of the considered field $\mathcal{F}$ of science but only the part which remains after eliminating the subset of the most popular words, then model (2) may be a sufficiently adequate approximation of reality. Its obvious advantage over the more general case (1) is that it enables us to use simpler statistical procedures for estimating the rate of development of the field $\mathcal{F}$.

3. Properties of the Model

3.1. ASYMPTOTIC PROPERTIES OF THE MODEL

Here and in what follows we denote in various places different positive finite constants by the letter C. To formulate results, we need the following concept.

DEFINITION 1. We call positive time series α_t and β_t asymptotically equivalent (in symbols $\alpha_t \sim \beta_t$), if there exists a function $f(x)$ such that $\lim_{x\to\infty} f(x) = 0$ and

$$\mathbb{E}|\alpha_t - \beta_t| \leqslant \mathbb{E}\big[f(\beta_t)\beta_t\big]. \tag{3}$$

Since $E(X_t - N_t)^2 = N_t - N_0$, we have

$$X_t \sim N_t. \tag{4}$$

In order to formulate similar statements and propositions for the sequence Y_t, we introduce additional constraints:

$$0 < C_1 \leqslant W \leqslant C_2 \tag{5}$$

$$\forall t = 1, 2, \ldots, \quad 1 \leqslant u_t \leqslant C_3 N_t^{\alpha} \quad \text{for some } \alpha < 1. \tag{6}$$

Here and below $W = W_1$. Let

$$\widetilde{M}_t = \mathbb{E}(Y_t \mid X_{(\cdot)} = N_{(\cdot)}), \qquad M_t = N_t - \sum_{\tau=1}^{t} d_\tau \mathbb{E}\exp\left\{-W\sum_{l=\tau}^{t}\frac{u_l}{N_l}\right\},$$

where $d_1 + \cdots + d_\tau = N_\tau$ for $\tau = 1, 2, \ldots$.

THEOREM 1. *Let condition* (5) *be fulfilled. Then*

$$Y_t \sim \mathbb{E}Y_t \sim \widetilde{M}_t \tag{7}$$

and under condition (6)

$$Y_t \sim M_t. \tag{8}$$

Note that conditions (5) and (6) are not necessary for relations (8), but they make it possible to simplify the proof.

Also note that the precise expressions of all the approximations, presented in Theorem 1, through the model characteristics $u_{(\cdot)}$, $N_{(\cdot)}$ and the distribution function F of W are rather complicated (especially of the mean EY_t). In the special case $\lambda_t \equiv$ const and $u_t \equiv$ const, however, there exists a very simple approximation. Write

$$h_t = \mathbb{E}\Delta Y_t, \qquad a = \mathbb{E}\frac{Wu\lambda}{Wu+\lambda}.$$

THEOREM 2. *Let* $\lambda_t \equiv \lambda$, $u_t \equiv u$ *and suppose that condition* (5) *holds. Then*

$$\lim_{t\to\infty} h_t = a, \tag{9}$$

$$Y_t \sim at, \qquad X_t \sim \lambda t \tag{10}$$

and

$$\lim_{t\to\infty} \mathbb{E}\left|\frac{Y_t}{X_t} - \frac{a}{\lambda}\right| = 0. \tag{11}$$

In the special case $W_{(\cdot)} = 1$, for Y_t/X_t, we have the limit $u/(u+\lambda)$.

Theorem 2 indicates that, if we know u and the distribution function F, then we can calculate the rate λ of development of $\mathcal{F}$ by the limit behaviour of the increments ΔY_t. Another urgent problem apart from the estimation of the rate of development is: can we state that, when observing the selected journals, all the keywords of field $\mathcal{F}$ which existed at the initial time moment will come into view sooner or later, i.e., whether $\mathbb{P}\{K_0 \subset S_t\} \to 1$ as $t \to \infty$

THEOREM 3. *There is the following equivalence*

$$\lim_{t\to\infty} \mathbb{P}\{K_0 \subset S_t\} = 1 \Leftrightarrow \sum_{t=1}^{\infty} \frac{u_t}{N_t} = \infty. \tag{12}$$

3.2. MARKOV PROPERTY

Let $W_{(\cdot)} = 1$. In this case, the two-dimensional sequence (X_t, Y_t) forms a Markov chain. Write

$$p_t(x, y) = \mathbb{P}\{X_t = x,\ Y_t = y\}.$$

Then, for all integer $x \geqslant N_0$, $x \geqslant y \geqslant 0$, we have

$$p_t(x, y) = \sum_{i=N_0}^{x} \sum_{j=0}^{\min(y;i)} p_{t-1}(i, j)\psi_t(i, j, x, y),$$

where the transition probabilities ψ satisfy the equalities

$$\begin{aligned}\psi_t(i, j, x, y) &= \mathbb{P}\{\Delta X_t = x - i\} \cdot \mathbb{P}\{\Delta Y_t = y - j \mid X_t = x, Y_{t-1} = j\} \\ &= P_{\lambda_t}(x - i) \cdot B_{x-j,u_t/x}(y - j).\end{aligned}$$

Here and in the sequel B denotes a binomial distribution, i.e.,

$$B_{n,q}(k) = \binom{n}{k} q^k (1 - q)^{n-k},$$

and P_λ, as above, is the Poisson distribution. In case $u_t \equiv u$, $\lambda_t \equiv \lambda$, we obtain a homogeneous Markov chain. Since at low values of u/x the binomial distribution is well approximated by the Poisson distribution law, in this case,

$$\psi(i, j, x, y) \cong P_\lambda(x - i) P_{u-uj/x}(y - j).$$

4. Statistical Estimation of the Parameters of the Model

The model presented is completely defined by the sequences $N_t = \mathbb{E}X_t$, $u_t \cong \mathbb{E}\,\mathrm{card}\, Q_t$ and by the distribution function F of the weights W_k. Applying this model in practical investigations, one can determine the sequence u_t a-priori by the volume of journals observed, while the sequence N_t and the function F should be estimated statistically. Let us discuss a parametric approach. Let $N_t = N_t(\theta)$ and $F(x) = F(x, \theta)$ be functions given a-priori, and θ be an unknown one- or multi-dimensional parameter to be estimated statistically, Θ is the set of possible values of θ. First we discuss the maximum likelihood estimate. We observe sets of words $Q_1, \ldots, Q_T$, $S_T = \bigcup_{t=1}^{T} Q_t$. Let us number the elements of S_T by their first observation moment in increasing order, for each $s \in S_T$ let $h(s)$ be the number of the word s. Write $H_t = \{h(s),\ s \in Q_t\}$. We have $H_t \subset \{1, \ldots, Y_t\}$, $t = 1, \ldots, T$. Obviously, the collection of sets $(H_1, \ldots, H_T)$ is a sufficient statistic for evaluating θ. For arbitrary sets of natural numbers $D_1, \ldots, D_T$, denote

$$G_\theta(D_1, \ldots, D_T) = \mathbb{P}_\theta\{H_1 = D_1, \ldots, H_T = D_T\},$$

where $\mathbb{P}_\theta$ stands for the probabilistic measure in case the value of an unknown parameter is θ. Then the likelihood function

$$L(\theta) = G_\theta(H_1, \ldots, H_T) \tag{13}$$

and the maximum likelihood estimate is defined by the equality

$$\theta_{\mathrm{ML}} = \arg\max_{\theta \in \Theta} L(\theta). \tag{14}$$

Regrettably, in case (13), it is very difficult to calculate the estimate θ_{ML}. We will consider more in detail the special case $W_k \equiv 1$, where this problem becomes simpler. The sufficient statistic with respect to θ is, in this case, the random vector

$$(Y_1, \ldots, Y_T, \zeta_2, \ldots, \zeta_T) \stackrel{\text{def}}{=} \Gamma, \quad \text{where } \zeta_t = \operatorname{card} Q_t \cap S_{t-1}.$$

Denote $\gamma = (n_1, \ldots, n_T, m_2, \ldots, m_T)$,

$$G_\theta(\gamma) = \mathbb{P}_\theta(\Gamma = \gamma).$$

In this case, the likelihood function $L(\theta) = G_\theta(\Gamma)$. For all nonnegative integers n_i and m_j satisfying the conditions

$$m_t \leqslant n_{t-1} \leqslant n_t, \quad t = 2, \ldots, T,$$

because of the above-mentioned Markov property, we obtain

$$\begin{aligned} G_\theta(\gamma) &= \mathbb{E}_\theta \prod_{t=1}^{T} \mathbb{P}\{Y_t = n_t,\ \zeta_t = m_t \mid Y_{t-1} = n_{t-1}, X_t\} \\ &\stackrel{\text{def}}{=} \mathbb{E}_\theta \Psi(\gamma, X_{(\cdot)}), \end{aligned} \tag{15}$$

where $Y_0 = n_0 = \zeta_1 = m_1 = 0$, and $\mathbb{E}_\theta$ is the mathematical expectation operator if value of the parameter under estimation is θ.

Next,

$$\begin{aligned} \mathbb{P}\{Y_1 &= n_1 \mid X_1\} = B_{X_1, u_1/X_1}(n_1), \\ \mathbb{P}\{Y_t &= n_t, \zeta_t = m_t \mid Y_{t-1} = n_{t-1}, X_t\} \\ &= B_{n_{t-1}, u_t/X_t}(m_t) \cdot B_{X_t - n_{t-1}, u_t/X_t}(n_t - n_{t-1}) \quad \text{if } X_t \geqslant n_t. \end{aligned}$$

We get the expression

$$\Psi(\gamma, x_{(\cdot)}) = \prod_{t=1}^{T} B_{x_t - m_{t-1}, u_t/x_t}(\Delta n_t) B_{n_{t-1}, u_t/x_t}(m_t), \tag{16}$$

where $B_{0,\cdot}(0) = 1$, $B_{s,\cdot}(r) = 0$ if $r > s$.

As we can see from (16), the calculation of the maximum likelihood estimate θ_{ML} is rather complicated even in the case of equal weights. We can modify this estimate by replacing $L(\theta)$ with a simpler function. When approximating binomial distributions in expression (19) by Poisson distributions, and ignoring the first $\tau - 1$ observations we obtain the following approximation of the function Ψ:

$$\widetilde{\Psi}(\gamma, x_{(\cdot)}) = \prod_{t=\tau}^{T} P_{u_t - u_t n_{t-1}/x_t}(\Delta n_t) P_{u_t n_{t-1}/x_t}(m_t), \tag{17}$$

where $P_0(0) = 1$, $P_\lambda(\cdot) = 0$ for $\lambda < 0$.

Based on the assumption $\mathbb{E}_\theta \widetilde{\Psi}(\gamma, X_{(\cdot)}) \cong \widetilde{\Psi}(\gamma, N_{(\cdot)}(\theta))$, we define a modification of the maximum likelihood estimate

$$\tilde{\theta}_{\mathrm{ML}} = \arg\max_{\theta} \widetilde{L}(\theta), \quad \widetilde{L}(\theta) \stackrel{\mathrm{def}}{=} \widetilde{\Psi}(\Gamma, N_{(\cdot)}(\theta)). \tag{18}$$

We take the maximum here over all $\theta \in \Theta$, for which

$$N_t(\theta) > Y_{t-1}, \quad t = \tau, \ldots, T. \tag{19}$$

In order that the estimate $\tilde{\theta}_{\mathrm{ML}}$ be consistent, the condition $\tau = \tau(T) \to \infty$, as $T \to \infty$ is necessary. Indeed, $\mathbb{P}\{Y_{t-1} > N_t\} > 0$. Therefore, in case $\tau(T) \leqslant C$, and $N_t(\theta)$ is a continuous function of θ, by (19) we have the inequality

$$\inf \lim_{T\to\infty} \mathbb{P}\{|\tilde{\theta}_{\mathrm{ML}} - \theta| > \varepsilon\} > 0 \quad \text{for some } \varepsilon > 0.$$

Striving for consistency of $\tilde{\theta}_{\mathrm{ML}}$, it is natural to select τ such that $\tau(T) \asymp T^\alpha$, as $T \to \infty, \alpha \in (0, 1)$.

In practical use, another modification, presented below, of the estimate θ_{ML} would be more convenient than $\tilde{\theta}_{\mathrm{ML}}$. Estimate (18) usually satisfies the equation

$$\frac{\mathrm{d}}{\mathrm{d}\theta} \log \widetilde{L}(\theta) = 0.$$

It is equivalent to the equation

$$\sum_{t=\tau}^{T} \frac{\Delta Y_t \cdot Y_{t-1} - \zeta_t(N_t(\theta) - Y_{t-1})}{N_t(\theta) \cdot (N_t(\theta) - Y_{t-1})} \cdot N_t'(\theta) = 0,$$

where $N_t'(\theta) = (\mathrm{d}/\mathrm{d}\theta) N_t(\theta)$. By rejecting Y_{t-1} in the denominator on the left-hand side of this equation (this can be substantiated using Theorems 1 and 2), we obtain the equation

$$\sum_{t=\tau}^{T} \frac{\Delta Y_t \cdot Y_{t-1} - \zeta_t(N_t(\theta) - Y_{t-1})}{N_t^2(\theta)} \cdot N_t'(\theta) = 0. \tag{20}$$

Choosing $\tau = 2$, define θ_{ML}^* as the solution to Equation (20) that belongs to a-priori given set Θ. In this case, the condition $\tau \to \infty$, as $T \to \infty$, is unnecessary. Let us consider the consistency of this estimate in the simplest case as $\lambda_t \equiv \lambda_0$. Then $N_t(\theta) = n + \lambda t$, $\theta = (n, \lambda)$. Let it be known a-priori that $\lambda \leqslant \lambda_{\max}$, $n \leqslant \underline{n_{\max}}$. In the case where Equation (20) has no solution (for example, $\zeta_t \equiv 0$, $t = \overline{1, T}$) we assume that $\theta_{\mathrm{ML}}^* = (n_{\max}, \lambda_{\max})$. If there exist some solutions of Equation (20) that satisfy a-priori constraints, θ_{ML}^* denotes the solution which corresponds to the highest value of λ.

THEOREM 4. *Let* $\lambda_t \equiv \lambda_0 \in (0, \lambda_{\max}]$, $u_t \equiv u < N_0$, $W_k \equiv 1$. *Then* λ_{ML}^* *is a consistent estimate of* λ *and has an optimal convergence rate*

$$\mathbb{E}(\lambda - \lambda_{\mathrm{ML}}^*)^2 = \mathrm{O}\left(\frac{1}{T}\right), \quad T \to \infty. \tag{21}$$

Equality (21) also holds under constraints weaker than the conditions of Theorem 4, however, these conditions simplify the proof a great deal.

We can get an even simpler consistent statistical estimate of the parameter than λ^*_{ML} under the conditions of Theorem 4, by making use of Theorems 1 and 2. Write

$$\hat{a} = \frac{Y_T - Y_L}{T - L}, \tag{22}$$

where $L = L(T)$ satisfies the condition

$$C_2 T \leqslant L \leqslant C_1 T,$$

for some constants $0 < C_2 < C_1 < 1$. If the conditions of Theorems 1 and 2 are fulfilled, then (22) is a consistent estimate of the parameter. In the case where $W_k \equiv 1$, if we know u we get the estimate of λ: $\hat{\lambda} = \hat{a}u/(u - \hat{a})$. If the weights W_k are not equal, then we can use the estimates of $\hat{a}$ for the comparison of rates of development. If in different fields or at different periods of time we have different unknown rates λ_1 and λ_2 but the same distribution of weights W_k, then

$$\frac{a_1}{u_1} < \frac{a_2}{u_2} \Rightarrow \frac{\lambda_1}{u_1} < \frac{\lambda_2}{u_2}.$$

Let us recall the general case: the weights W_k are random variables, and the model is completely defined by a parameter θ, i.e., $N_t = N_t(\theta)$, $F(x) = F(x, \theta)$. How can we statistically estimate θ in practice if, as mentioned, it is very difficult to calculate the estimate θ_{ML} defined by equalities (13) and (14)? Applying the maximum likelihood principle but using incomplete statistical information, we can obtain simpler (though less exact) statistical estimates. We present here a concrete modification of the maximum likelihood function, the calculation of which is rather simple. As above $(Q_1, \ldots, Q_T)$ is a sample. Let L be the integer part of the number T/c, where $c > 1$ is a selected constant, $q_i = \operatorname{card} S_L \cap Q_i$. In order to get a simpler estimation procedure, we shall use only the data vector $(q_{L+1}, \ldots, q_T) =_{\mathrm{def}} q$ to estimate the parameter θ. Evidently, q is not a sufficient statistic, therefore, in this case, we lose part of the statistical information on θ. Since the distribution of q is completely defined by parameter θ, we can apply the maximum likelihood method. Having fixed $W_{(\cdot)}$ and $X_{(\cdot)}$, the random variables q_t are relatively independent, therefore

$$\mathbb{P}\{q_t = m_t,\ L < t \leqslant T\} = \mathbb{E} \prod_{t=L+1}^{T} \mathbb{P}\{q_t = m_t \mid X_{(\cdot)}, W_{(\cdot)}\}. \tag{23}$$

Applying the Poisson approximation to the conditional distribution of the random variable q_t and denoting $Y_t^* = \sum_{k : w_k \in S_t} W_k$, $X_t^* = \sum_{k=1}^{X_t} W_k$, we have

$$\mathbb{P}\{q_t = m_t \mid X_{(\cdot)}, W_{(\cdot)}\} \cong P_{Y_L^* u_t / X_t^*}(m_t). \tag{24}$$

With a view to motivate further simplification, we need the following theorem. Denote $M_t^* = N_t - \sum_{\tau=1}^{t} d_\tau \mathbb{E}[W \exp\{-W \sum_{l=\tau}^{t} u_l / N_l\}]$.

THEOREM 5. *Let* (5) *and* (6) *hold. Then*

$$X_t^* \sim N_t, \qquad Y_t^* \sim M_t^*. \tag{25}$$

If we compare Theorems 1 and 5, by the results of Theorem 1 we can draw conclusions about the asymptotic behaviour of Y_t / X_t that determines which part of the set of keywords becomes known, while the results of Theorem 5 allow us to draw conclusions on the part of 'weight' of the observed words.

Making use of approximations (25), from (23) and (24) we obtain

$$\begin{aligned} \mathbb{P}_\theta\{q_t = m_t,\ L < t \leqslant T\} &\cong \prod_{t=L+1}^{\mathrm{T}} P_{u_t M_L^*(\theta)/N_t(\theta)}(m_t) \\ &\stackrel{\text{def}}{=} \Phi_\theta(m_{L+1}, \ldots, m_T), \end{aligned} \tag{26}$$

where $M_t^*(\theta)$ and $N_t(\theta)$ are defined analogously as M_t^* and N_t, by replacing the mathematical expectation operator $\mathbb{E}$ by $\mathbb{E}_\theta$. We define the proposed estimate by the equality

$$\theta^* = \arg\max_{\theta \in \Theta} \Phi_\theta(q). \tag{27}$$

Put

$$l(\theta) = \sum_{t=L+1}^{\mathrm{T}} \left[-r_t(\theta) + q_t \log r_t(\theta)\right], \quad \text{where} \quad r_t(\theta) = \frac{u_t M_L^*(\theta)}{N_t(\theta)}. \tag{28}$$

Definitions (28) and (27) yield the equality

$$\theta^* = \arg\max_{\theta \in \Theta} l(\theta). \tag{29}$$

It is desirable to investigate the accuracy of the estimates presented by means of simulation. The results of these investigations are to be presented in a following paper in the near future.

5. Proofs

For an arbitrary set D let $|D| = \operatorname{card} D$.

Proof of Theorem 1. Let (5) hold. We denote by symbols $\nu_{k,l}$ i.i.d. random variables such that do not depend on $X_{(\cdot)}$ and $W_{(\cdot)}$, have a uniform distribution in the interval (0, 1), and satisfy the equality $A_{k,l} = \{\nu_{k,l} < \pi_{k,l}\}$ (see (1)). Let us fix t and denote $\xi_{k,\tau} = 1 - \mathbb{1}_{\{\nu_{k,l} \geqslant \pi_{k,l},\, l = \overline{\tau,t}\}}$,

$$Z_\tau = \sum_{k \in D_\tau} \xi_{k,\tau}, \quad D_\tau = \{1, \ldots, X_\tau\} \setminus D_{\tau-1}, \quad D_0 = \phi.$$

Then $Y_t = \sum_{\tau=1}^{t} Z_\tau$. Let $\tilde{\pi}_{k,l} = W_k u_l / N_l$ and the quantities $\widetilde{Y}_{(\cdot)}, \widetilde{Z}_{(\cdot)}, \widetilde{\xi}_{(\cdot)}$ be defined analogously as $Y_{(\cdot)}, Z_{(\cdot)}, \xi_{(\cdot)}$, replacing $\pi_{(\cdot)}$ by $\tilde{\pi}_{(\cdot)}$. Note that

$$\mathbb{E}\widetilde{Y}_t = \widetilde{M}_t. \tag{30}$$

In order to prove (30), denote $\varphi_\tau(x_{(\cdot)}) = \mathbb{E}\prod_{l=\tau}^{t}(1 - W u_l / x_l)_+$, where $x_{(\cdot)}$ is any determined positive sequence, x_+ means $\max(x, 0)$. Then

$$\mathbb{E}(Y_t \mid X_{(\cdot)}) = \sum_{\tau=1}^{t} |D_\tau|(1 - \varphi_\tau(X_{(\cdot)})).$$

Since $\mathbb{E}\widetilde{\xi}_{k,\tau} = 1 - \varphi_\tau(N_{(\cdot)})$, we obtain (30). We show that

$$\operatorname{Var} \widetilde{Y}_t \leqslant C\widetilde{M}_t. \tag{31}$$

Denoting $\xi_\tau = \xi_{1,\tau}$ and $d_\tau = \mathbb{E}|D_\tau|$, due to the independence of sequence $\widetilde{Z}_\tau$ we get

$$\begin{aligned}\operatorname{Var} \widetilde{Y}_t &= \sum_{\tau=1}^{t} \operatorname{Var} \widetilde{Z}_\tau = \sum_{\tau=1}^{t}\left[\mathbb{E}\left(\sum_{k\in D_\tau} \widetilde{\xi}_{k,\tau}\right)^2 - (d_\tau \mathbb{E}\widetilde{\xi}_\tau)^2\right] \\ &= \sum_{\tau=1}^{t}\left[d_\tau \operatorname{Var}\widetilde{\xi}_\tau + (\mathbb{E}\widetilde{\xi}_\tau)^2 \operatorname{Var}|D_\tau|\right] \leqslant C\sum_{\tau=1}^{t} d_\tau \mathbb{E}\widetilde{\xi}_\tau = C\mathbb{E}\widetilde{Y}_t = C\widetilde{M}_t.\end{aligned}$$

In order to prove (7), it remains to estimate $\mathbb{E}|Y_t - \widetilde{Y}_t|$. Denote $\Pi_\tau = \sum_{l=\tau}^{t} u_l / N_l$. First, let us consider the case $\Pi_\tau > C \log N_\tau$. The equality

$$\mathbb{E}|Z_\tau - \widetilde{Z}_\tau| = \mathbb{E}\left[|D_\tau| \cdot \mathbb{P}\{\xi_\tau \neq \widetilde{\xi}_\tau | X_{(\cdot)}, W_{(\cdot)}\}\right]$$

is true. By virtue of (5) we have

$$\mathbb{P}\{\xi_\tau = 0 \mid X_{(\cdot)}, W_{(\cdot)}\} = \prod_{l=\tau}^{t}(1 - \pi_{1,l}) \leqslant \exp\left\{-\sum_{l=\tau}^{t} \frac{u_l}{X_l C}\right\}.$$

Analogously, $\mathbb{P}\{\tilde{\xi}_\tau = 0 \mid X_{(\cdot)}, W_{(\cdot)}\} \leqslant \exp\{-\Pi_\tau / C\}$, therefore,

$$\mathbb{E}|Z_\tau - \widetilde{Z}_\tau| \leqslant d_\tau\left(\exp\{-\Pi_\tau / C\} + p\right), \quad p = \mathbb{P}\left\{\sum_{l=\tau}^{t} \frac{u_l}{X_l} \leqslant \frac{\Pi_\tau}{2}\right\}.$$

Due to the Chebyshev inequality and the condition $X_l \geqslant 1$ we have

$$\begin{aligned} p &\leqslant \mathbb{P}\left\{\sum_{l=\tau}^{t} \frac{u_l(X_l - N_l)}{N_l X_l} \geqslant \Pi_\tau / 2\right\} \leqslant 2\sum_{l=\tau}^{t} \frac{u_l}{N_l}\mathbb{E}\left|\frac{X_l - N_l}{X_l}\right| / \Pi_\tau \\ &\leqslant 2\max_{l\geqslant\tau} \mathbb{E}\left|\frac{X_l - N_l}{X_l}\right| \leqslant C/\sqrt{N_\tau}. \end{aligned}$$

Thus,

$$\mathbb{E}|Z_\tau - \widetilde{Z}_\tau| \leqslant Cd_\tau\left(e^{-\Pi_\tau/C} + N_\tau^{-1/2}\right).$$

If $\Pi_\tau > 1$, then $\mathbb{E}\widetilde{Z}_\tau \geqslant d_\tau/C$, therefore

$$\mathbb{E}|Z_\tau - \widetilde{Z}_\tau| \leqslant C\mathbb{E}\widetilde{Z}_\tau\left(e^{-\Pi_\tau/C} + N_\tau^{-1/2}\right). \tag{32}$$

In the case $\Pi_\tau < C\log N_\tau$, are make use of the inequality

$$\mathbb{E}|Z_\tau - \widetilde{Z}_\tau| \leqslant \mathbb{E}\sum_{k\in D_\tau}\sum_{l=\tau}^{t}|\pi_{k,l} - \tilde{\pi}_{k,l}|.$$

By virtue of (5),

$$|\pi_{k,l} - \tilde{\pi}_{k,l}| \leqslant \frac{Cu_l|X_l - N_l|}{N_l X_l}.$$

Since

$$\mathbb{E}\left[|D_\tau|\cdot|1 - N_l/X_l|\right] \leqslant Cd_\tau/\sqrt{N_l},$$

we get

$$\mathbb{E}|Z_\tau - \widetilde{Z}_\tau| \leqslant Cd_\tau\Pi_\tau/\sqrt{N_\tau}. \tag{33}$$

As $\Pi_\tau < C\log N_\tau$, we have the estimate $\mathbb{E}\widetilde{\xi}_\tau \geqslant \Pi_\tau/(C\log N_\tau)$, therefore from (32) and (33) yield the inequality

$$\mathbb{E}|Y_t - \widetilde{Y}_t| \leqslant \sum_{\tau=1}^{t}\mathbb{E}|Z_\tau - \widetilde{Z}_\tau| \leqslant C\sum_{\tau=1}^{t}\frac{d_\tau\log N_\tau}{\sqrt{N_\tau}}\mathbb{E}\widetilde{\xi}_\tau. \tag{34}$$

Since $\widetilde{M}_t = \sum_{\tau=1}^{t} d_\tau\mathbb{E}\widetilde{\xi}_\tau$ and $\mathbb{E}\widetilde{\xi}_\tau \leqslant 1$, from (34) we obtain that

$$\mathbb{E}|Y_t - \widetilde{Y}_t| \leqslant C\widetilde{M}_t^{\beta} \quad \text{for some constant } \beta < 1. \tag{35}$$

(7) follows from (30), (31), and (35). Let (5) and (6) be valid. Without loss of generality, we can consider the case $C_2 W_{(\cdot)} < N_{(\cdot)}$. Then

$$\begin{aligned}\varphi_\tau(N_{(\cdot)}) &\overset{\text{def}}{=} \mathbb{E}\exp\left\{\sum_{l=\tau}^{t}\log\left(1 - \frac{Wu_l}{N_l}\right)\right\}\\ &= \mathbb{E}\exp\{-W\Pi_\tau\}\exp\left\{-rN_\tau^{\alpha-1}\Pi_\tau\right\}\end{aligned}$$

for some $r = r(\tau, t, u_{(\cdot)}, N_{(\cdot)}) \in [0, C]$. Consequently,

$$1 \geqslant \frac{1 - \varphi_\tau(N_{(\cdot)})}{1 - \mathbb{E}\exp\{-W\Pi_\tau\}} \geqslant 1 - \frac{C\log N_\tau}{N_\tau^{1-\alpha}}.$$

Since $\widetilde{M}_t = \sum_{\tau=1}^{t} d_\tau[1 - \varphi_\tau(N_{(\cdot)})]$, $M_t \sim \sum_{\tau=1}^{t} d_\tau(1 - \mathbb{E}\,\mathrm{e}^{-W\Pi_\tau})$, we have

$$\widetilde{M}_t \sim M_t. \tag{36}$$

(7) and (36) yield (8). □

To prove Theorem 2 we need (25), therefore, first let us discuss the proof of Theorem 5. We have

$$\mathbb{E}|X_t^* - N_t| \leqslant \mathbb{E}\big[X_t|\overline{W} - 1| + |X_t - N_t|\big], \quad \overline{W} = \frac{1}{X_t}\sum_{k=1}^{X_t} W_k.$$

By virtue of (5)

$$\mathbb{E}(\overline{W} - 1)^2 \leqslant \mathbb{E}\frac{C}{X_t} \leqslant \frac{C}{N_t},$$

therefore $\mathbb{E}[X_t|\overline{W} - 1|] \leqslant C\sqrt{N_t}$. Since $\mathbb{E}|X_t - N_t| \leqslant C\sqrt{N_t}$, we obtain

$$\mathbb{E}|X_t^* - N_t| \leqslant C\sqrt{N_t} \Rightarrow X_t^* \sim N_t.$$

The proof of the proposition $Y_t^* \sim M_t^*$ is analogous to that of (8).

Proof of Theorem 2. Trivially

$$h_t = \mathbb{E}|Q_t| - \mathbb{E}|Q_t \cap S_{t-1}|. \tag{37}$$

Next,

$$\mathbb{E}|Q_t| = \mathbb{E}\sum_{k=1}^{X_t} \pi_{k,t} = u - \rho_t, \tag{38}$$

where by virtue of (5) and the Chebyshev inequality

$$\rho_t = \mathbb{E}\sum_{k=1}^{X_t}\left(\frac{uW_k}{X_t} - 1\right)_+ \leqslant Cu\mathbb{P}\{X_t < Cu\} \leqslant \frac{Cu}{\lambda t}. \tag{39}$$

We obtain

$$\mathbb{E}|Q_t \cap S_{t-1}| = \mathbb{E}\sum_{k:w_k \in S_{t-1}} \frac{uW_k}{N_t} + \rho_t^* = \frac{u\mathbb{E}Y_{t-1}^*}{N_t} + \rho_t^*, \tag{40}$$

where

$$|\rho_t^*| \leqslant \mathbb{E}\sum_{k=1}^{X_t}\left|\pi_{k,t} - \frac{uW_k}{N_t}\right| \leqslant Cu\mathbb{E}\left|1 - \frac{X_t}{N_t}\right| \leqslant \frac{Cu}{\sqrt{\lambda t}}. \tag{41}$$

It follows from (25) and (37)–(41) that as $t \to \infty$

$$h_t = u - \frac{uM^*_{t-1}}{N_t} + o(1). \tag{42}$$

Under the conditions of the theorem

$$\begin{aligned} M^*_t &= N_0 + \lambda t - \sum_{\tau=1}^{t} \lambda \mathbb{E}\left[W \exp\left\{ -W \sum_{t=\tau}^{t} \frac{u}{N_0 + \lambda t} \right\} \right], \\ M^*_t &\sim \lambda t - \lambda \sum_{\tau=1}^{t} \mathbb{E}\left[W \exp\left\{ -\frac{Wu}{\lambda} \log \frac{t}{\tau} \right\} \right] \sim \lambda t - \lambda \mathbb{E} \sum_{\tau=1}^{t} W \left(\frac{\tau}{t} \right)^{\frac{Wu}{\lambda}}, \\ M^*_t &\sim \lambda t - \lambda t \mathbb{E} \frac{W}{1 + \frac{Wu}{\lambda}} = \lambda t \left(1 - \mathbb{E} \frac{W\lambda}{Wu + \lambda} \right). \end{aligned} \tag{43}$$

We have $N_t \sim \lambda t$ and from (42) and (43) we get that

$$\lim_{t \to \infty} h_t = u\lambda \mathbb{E} \frac{W}{Wu + \lambda} = a.$$

Relation (10) follows from (9) and (4) while equality (11) is obtained from the equivalences $Y_t \sim at$, $N_t \sim \lambda t$ and the inequality

$$\mathbb{E} \left| \frac{Y_t}{X_t} - \frac{Y_t}{N_t} \right| \leqslant \mathbb{E} \left| 1 - \frac{X_t}{N_t} \right| \leqslant C/\sqrt{N_t}. \qquad \square$$

Proof of Theorem 3. Denote $n = N_0 \geqslant 1$.

$$\begin{aligned} \forall k = 1, \ldots, n \quad \mathbb{P}\{w_k \in S_t\} &= 1 - \mathbb{P}\{\nu_{k,\tau} > \pi_{k,\tau}, \tau = \overline{1,t}\} \\ &= 1 - \mathbb{E} \prod_{\tau=1}^{t} (1 - \pi_{k,\tau}) \\ &\geqslant 1 - \mathbb{E} \exp\left\{ -W_k \sum_{\tau=1}^{t} \frac{u_\tau}{X_\tau} \right\}. \end{aligned} \tag{44}$$

It is not difficult to prove that

$$p(q) \stackrel{\text{def}}{=} \mathbb{P}\left\{ \sup_{t=1,2,\ldots} \frac{X_t}{N_t} > q \right\} \to 0 \quad \text{as } q \to \infty.$$

To this end, define the natural numbers $t_0, t_1, \ldots, t_r$ by the equalities $t_k = \min\{t : N_t > n \cdot 2^k\}$. Here $r = \infty$ if $\lim_{t \to \infty} N_t = \infty$. Due to monotonicity of the sequences X_t and N_t,

$$p(q) \leqslant \sum_{k=0}^{r} \mathbb{P}\{2X_{t_k} > qN_{t_k}\}.$$

Since X_t are Poisson random variables, $\mathbb{E}X_t = N_t$ and $N_{t_k} > 2^k$, we have $p(q) \leqslant C\,\mathrm{e}^{-q}$. Thus, if

$$\Pi_\infty \stackrel{\text{def}}{=} \sum_{\tau=1}^{\infty} \frac{u_\tau}{N_\tau} = \infty,$$

then

$$\sum_{\tau=1}^{t} \frac{u_\tau}{X_\tau} \xrightarrow{P} \infty \quad \text{as } t \to \infty.$$

Since W_k are positive random variables independent of $X_{(\cdot)}$, from (44) we get the implication

$$\Pi_\infty = \infty \Rightarrow \lim_{t\to\infty} \mathbb{P}\{K_0 \subset S_t\} = 1.$$

Let $\Pi_\infty < \infty$. Then for each positive number ε there exists a finite number $T = T(\varepsilon)$, such that $\Pi_\infty - \Pi_{T-1} < \varepsilon$. Set $B = \{W \leqslant 1,\ X_t \geqslant N_t\ \forall t < T\}$ we get

$$\mathbb{P}(B) > 0. \tag{45}$$

Denote

$$D = \left\{ \sum_{t=T}^{\infty} \frac{u_t}{X_t} < 1/2 \right\}, \qquad \eta_t = \sum_{\tau=T}^{t} (\Delta X_\tau - \Delta N_t).$$

Because of the independencies of ΔX_τ we get the estimate of conditional probability

$$\begin{aligned} 1 - \mathbb{P}(D \mid B) &\leqslant \mathbb{P}\left\{ 2\varepsilon \sum_{t=T}^{\infty} \frac{u_t}{N_t + \eta_t} \geqslant \sum_{t=T}^{\infty} \frac{u_t}{N_t} \right\} \\ &\leqslant \frac{2\varepsilon \sum_{t=T}^{\infty} \mathbb{E}\frac{u_t}{N_t+\eta_t}}{\sum_{t=T}^{\infty} u_t/N_t} \leqslant C\varepsilon, \end{aligned} \tag{46}$$

where C does not depend on T. Since

$$\prod_{t=T}^{\infty} \left(1 - \frac{u_t W}{X_t}\right) \geqslant 1 - \sum_{t=T}^{\infty} \frac{u_t}{X_t} \geqslant \frac{1}{2}, \quad \text{if the events } B \text{ and } D \text{ occur,}$$

for a sufficiently low value of ε (44)–(46) yield the inequalities

$$\lim_{t\to\infty} \{w_1 \notin S_t\} \geqslant \mathbb{P}(B) \prod_{\tau=1}^{T} \left(1 - \frac{u_t}{N_\tau}\right)(1 - C\varepsilon)/2 > 0.$$

Consequently,

$$\Pi_\infty < \infty \Rightarrow \lim_{t\to\infty} \mathbb{P}\{K_0 \subset S_t\} > 0. \qquad \square$$

Proof of Theorem 4. We restrict ourselves to a separate case, where the value of N_0 is known, $N_0 = n$. Thus, $\theta = \lambda$, $N_t(\lambda) = n + \lambda t$, $N_t = n + \lambda_0 t$. The proof of Theorem 4 in the general case is analogous to that given below, only more cumbersome.

Denoting $\xi_t = (\Delta Y_t + \zeta_t)Y_{t-1} - \zeta_t N_t$, $\beta(\lambda) = \sum \xi_t \cdot t/(n + \lambda t)^2$, we obtain from (20) the equation

$$\beta(\lambda) = (\lambda - \lambda_0) \sum \zeta_t t^2 (n + \lambda t)^2. \tag{47}$$

Here and in what follows, $\sum$ denotes summation over $t = 2, \ldots, T$. Let $\Lambda = [\lambda_0/2, \lambda_{\max}]$, $Z = \max_{\lambda \in \Lambda} \beta(\lambda)$, $Z^* = \min_{\lambda \in \Lambda} \sum \zeta_t t^2/(n + \lambda t)^2$. Since λ^*_{ML} is a bounded random variable and the inequality $|\lambda^*_{\mathrm{ML}} - \lambda_0| \leqslant Z/Z^*$ holds, to prove (21) it suffices to show that

$$\mathbb{E}Z^2 = \mathrm{O}(T) \tag{48}$$

and for a certain positive constant $C = C(n, \lambda_0, \lambda_{\max})$

$$\mathbb{P}\{Z^* \leqslant T/C\} = \mathrm{O}(1/T). \tag{49}$$

First we prove (48). Let $\xi_t^* := \mathbb{E}(\xi_t \mid I_t) = uY_{t-1}(1 - N_t/X_t)$, $\bar{\xi}_t = \xi_t - \xi_t^*$. We define random functions β^* and $\bar{\beta}$ analogously as β, by replacing ξ_t by ξ_t^* and $\bar{\xi}_t$, respectively. We get

$$\max_{\lambda \in \Lambda} \beta^*(\lambda) \leqslant C \sum |X_t - N_t|/t;$$

and

$$\mathbb{E} \max_{\lambda \in \Lambda} \beta^{*^2}(\lambda) \leqslant C\Big[\sum \big(\mathbb{E}[\zeta_t^2 (X_t - N_t)^2]\big)^{1/2}/t\Big]^2.$$

Since $\mathbb{E}(X_t - N_t)^2 \leqslant CN_t \leqslant Ct$, we have

$$\mathbb{E} \max_{\lambda \in \Lambda} \beta^{*^2}(\lambda) \leqslant CT. \tag{50}$$

Next,

$$\forall t \ \mathbb{E}\bar{\xi}_t^2 \leqslant Ct^2 \tag{51}$$

and

$$\mathbb{E}\bar{\xi}_t \bar{\xi}_s = 0, \quad t \neq s. \tag{52}$$

Therefore,

$$\mathbb{E}\bar{\beta}^2(\lambda) \leqslant CT, \quad \lambda \in \Lambda. \tag{53}$$

Denoting the derivative of function $\bar{\beta}$ by $\bar{\beta}'$, we have

$$\bar{\beta}'(\lambda) = -2 \sum \frac{\bar{\xi}_t t^2}{(n + \lambda t)^3}$$

and making use of (51) and (52) we obtain the inequality

$$\mathbb{E}\left|\bar{\beta}'(\lambda)\right|^2 \leqslant CT, \quad \lambda \in \Lambda. \tag{54}$$

By virtue of (53) and (54), we get

$$\begin{aligned}\mathbb{E}\max_{\lambda\in\Lambda}\bar{\beta}^2(\lambda) &\leqslant 2\mathbb{E}\bar{\beta}^2(\lambda_0/2) + 2\mathbb{E}\left(\int_\Lambda |\bar{\beta}'(\lambda)|\,\mathrm{d}\lambda\right)^2 \\ &\leqslant 2\mathbb{E}\bar{\beta}^2(\lambda_0/2) + C\max_{\lambda\in\Lambda}\mathbb{E}|\bar{\beta}'(\lambda)|^2 \leqslant CT.\end{aligned} \tag{55}$$

Since $Z^2 = \max_{\lambda\in\Lambda}[\beta^*(\lambda) + \bar{\beta}(\lambda)]$, from (50) and (55) we obtain (48).

It remains to prove (49). Let L be the integral part of number $T/2$, $G = |Q_1| + \cdots + |Q_L|$, $G^* = \zeta_2 + \cdots + \zeta_L$. Then

$$G \leqslant Y_L + G^*, \quad Z^* \geqslant G^*/C. \tag{56}$$

For simplicity, let $u \geqslant 1$. Since after fixing $X_{(\cdot)}$ the random variables $|Q_i|$ are conditionally independent, $\mathbb{P}\{|Q_i| \geqslant 1\} \geqslant 1/2$, we have

$$\mathbb{P}\{G < T/4\} = \mathrm{O}\left(\frac{1}{T}\right). \tag{57}$$

By (56) the implication

$$G \geqslant T/4,\ Y_L \leqslant T/8 \Rightarrow Z^* \geqslant T/C$$

holds, therefore, using (57) we obtain the inequality

$$\mathbb{P}\left\{Z^* \leqslant T/C\right\} \leqslant \mathbb{P}\left\{G^* \leqslant T/C \mid Y_L \geqslant \frac{T}{8}\right\} + \mathrm{O}\left(\frac{1}{T}\right). \tag{58}$$

Choosing a sufficiently large number C and approximating the binomial distribution by Poisson law, we have

$$\mathbb{P}\left\{G^* \leqslant \frac{T}{C}\middle| Y_L \geqslant \frac{T}{8},\ X_T \leqslant 2\lambda_0 T\right\} \leqslant 2\sum_{k\leqslant T/C} P_{\frac{uT}{32\lambda_0}}(k) = \mathrm{O}\left(\frac{1}{T}\right). \tag{59}$$

Since $\mathbb{P}\{X_T \geqslant 2\lambda_0 T\} = \mathrm{O}(1/T)$, (58) and (59) yield (49). □

References

1. Hazewinkel, M.: Index — Volumes 1–89 of 'Artificial Intelligence', *Artificial Intelligence* **96** (1997), 1–227.
2. Hazewinkel, M.: Index — Volumes 1–200 of 'Theoretical Computer Science', *Theoret. Comput. Sci.* **213/214** (1999), 1–699.
3. Hazewinkel, M.: Index — Volumes 1–91 of 'Discrete Applied Mathematics', *Discrete Appl. Math.* **106** (2000), 1–261.
4. Hazewinkel, M.: Index — Volumes 1–2000 of 'Discrete Mathematics', *Discrete Math.* **227/228** (2001), 1–648.
5. Hazewinkel, M.: Topologies and metrics on information spaces, *CWI Quarterly* **10**(2) (1999), 93–110.