Performance Evaluation 42 (2000) 205–222

PERFORMANCE EVALUATION
An International Journal

www.elsevier.com/locate/peva

Nonparametric estimation of long-tailed density functions and its application to the analysis of World Wide Web traffic

Natalia M. Markovitch [a,*], Udo R. Krieger [b,c]

[a] *Institute of Control Sciences, Russian Academy of Sciences, Profsoyuznay 65, 117806 Moscow, Russia*
[b] *T-Nova Deutsche Telekom, Technologiezentrum, Am Kavalleriesand 3, D-64295 Darmstadt, Germany*
[c] *Department of Computer Science, J.W. Goethe-University, D-60054 Frankfurt, Germany*

Abstract

The study of WWW-traffic measurements has shown that different traffic characteristics can be modeled by long-tail distributed random variables (r.v.s). In this paper we discuss the nonparametric estimation of the probability density function of long-tailed distributions. Two nonparametric estimates, a Parzen–Rosenblatt kernel estimate and a histogram with variable bin width called polygram, are considered. The consistency of these estimates for heavy-tailed densities is discussed. To provide the consistency of the estimates in the metric space L_1, the transformation of the initial r.v. to a new r.v. distributed on the interval [0, 1] is proposed. Then the proposed estimates are applied to analyze real data of WWW-sessions. The latter are characterized by the sizes of the responses and inter-response intervals as well as the sizes and durations of sub-sessions. By these means the effectiveness of the nonparametric procedures in comparison to parametric models of the WWW-traffic characteristics is demonstrated. © 2000 Published by Elsevier Science B.V.

Keywords: World Wide Web; Nonparametric density estimation; Parzen–Rosenblatt estimate; Polygram; Long-tailed distribution; Consistency

1. Introduction

Considering the rapid growth of the Internet and the effective design of the underlying IP-based transport network, efficient data gathering and evaluation as well as a careful statistical analysis of the corresponding random processes and random variables (r.v.s) describing the World Wide Web (WWW) traffic characteristics are required. The analysis of existing measurements of WWW-traffic by statistical methods has shown that the characteristics can often be modeled by long-tailed distributions or follow mixtures of long-tailed distributions due to the heterogeneous sources of the information transfer (see [2,6,14,20] and references therein).

In the last few years, estimation methods for long-tailed probability density functions (p.d.f.s) have been developed (cf. [8]). The basic question is how to restore a long-tailed p.d.f. by empirical data of

* Corresponding author.
E-mail addresses: markovic@ipu.rssi.ru (N.M. Markovitch), udo.krieger@ieee.org (U.R. Krieger).

0166-5316/00/$ – see front matter © 2000 Published by Elsevier Science B.V.
PII: S0166-5316(00)00031-6

a limited number. First of all, standard nonparametric estimates, such as a histogram, a projection or a Parzen–Rosenblatt (P–R) kernel estimate, cannot describe the behavior of the p.d.f. on the tail due to the lack of information outside the closed interval determined by the range of an empirical sample. They operate just with the empirical samples of limited size which are not representative regarding the tail. However, for many applications it is important to know the probability of rare events. Hence, a parametric approach has been developed to describe the tails. Among these parametric tail estimates Hill's estimate and the kernel tail-estimates are popular (cf. [7,11]). Typically, there are not enough data to test the parametric form of the tail with sufficient confidence. Besides, parametric tail models do not reflect the behavior of the p.d.f. for relative small values of a r.v. The experiences of the restoration with parametric models have shown that some models describe the tails quite good and other models are better for the small-values area of the p.d.f. (cf. [14,20]).

Considering the mentioned difficulties, it is the aim of this paper to propose a reliable nonparametric estimate for a long-tailed p.d.f. arising from WWW-traffic characterization. To apply a nonparametric approach, we need only general information describing the p.d.f. We may know, e.g. that the p.d.f. is long-tailed, continuous or bounded, etc.

For a long time the nonparametric estimation of a long-tailed p.d.f. was based on the assumption that the p.d.f. has a compact support since all points of a fixed empirical sample are concentrated on a compact support, e.g. on some closed interval. Regarding, for instance, a Gaussian p.d.f. we may be convinced that 95% of the points are located within the interval $\mu \pm 3\sigma$. Then different methods for the restoration of a p.d.f. with compact support, such as projection and histogram estimates, have been applied to long-tailed p.d.f.s. However, in this case there is a source of an estimation error arising from the ignored tails. In contrast to that approach, a Parzen–Rosenblatt (P–R) estimate

$$f_{h,l}(t) = \frac{1}{lh}\sum_{i=1}^{l} K\left(\frac{t-x_i}{h}\right), \tag{1}$$

where $K(t)$ is a kernel function and h is a smoothing parameter ("a window width"), does not demand the assumption of a compact support. It is defined on the whole real axis $\mathbb{R}$ and may, therefore, be applied to a long-tailed p.d.f. In the following, C denotes the space of real-valued continuous functions. Considering the P–R estimate and its basic features, the asymptotic consistency and the limit lower bounds for the estimation of the risk in the metric spaces L_2 and C were proved in [9,10] for a smooth p.d.f. satisfying Hoelder's condition provided that $h \to 0$, $lh \to \infty$ for $l \to \infty$. This means that sufficiently accurate asymptotic estimates may be achieved even for a long-tailed p.d.f. The mathematical term consistency (or convergence) means that the error of the estimation which is determined by the metric of L_1, L_2 or C tends to zero in probability or almost surely (a.s.) if the sample size l goes to infinity.

If the sample size is limited, one selects the smoothing parameter h depending on the observations to get accurate estimates. One of these selection techniques is provided by the cross-validation method (c.v.m.) (cf. [5]).

It is the basic statistical problem of a long-tailed p.d.f. that the spacing between the extreme order statistics does not converge to zero. This feature is illustrated by any p.d.f. satisfying the condition

$$\lim_{x\to\infty} \frac{f(x)}{\int_x^\infty f(y)\,\mathrm{d}y} = 0, \qquad \int_x^\infty f(y)\,\mathrm{d}y > 0 \tag{2}$$

for any x, particularly those p.d.f.s whose tails decrease as $cx^{-\alpha}$, $\alpha > 1$. Cauchy, Pareto and Student distributions exhibit such a behavior. This leads to the L_1-nonconsistency of the P–R estimates (cf. [15]) and, therefore, the L_2- and C-nonconsistency for some long-tailed p.d.f.s if h is selected by the c.v.m.

In [8] the practical and theoretical importance of the L_1-consistency of the estimates is demonstrated. It seems that the convergence of the estimates in the L_1 metric is "weaker" than in the L_2 and C metric. Particularly, C-consistent estimates provide uniform reliability on the whole domain of definition. Nevertheless, according to Scheffe's theorem (cf. [3,8]) one half of the L_1 distance between two p.d.f.s f and g is equal to the total variation between two probabilities of any Borel set A, $\frac{1}{2}\int |f(x) - g(x)|\lambda(\mathrm{d}x) = \sup_A |\int_A f(x)\lambda(\mathrm{d}x) - \int_A g(x)\lambda(\mathrm{d}x)|$, where the supremum is taken over all measurable sets A and λ is a σ-finite measure, e.g. the Lebesgue measure on $\mathbb{R}^1$. Since in practice we are more often interested to estimate some probability functions, e.g. a distribution function (d.f.) $F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t$ or a tail $1 - F(x)$ than a p.d.f. $f(x)$, the L_1-consistency cannot be ignored. In [5] the L_1-consistency of the P–R and histogram estimates has been proved for p.d.f.s with compact support. It was assumed that the kernel function of (1) is bounded and has a compact support, i.e. $K(x) \neq 0$ for $x \in [a, b]$, $K(x) = 0$ for $x \notin [a, b]$, where $[a, b]$ is a closed interval, and h or the bin width of a histogram have been selected by the c.v.m. Generally, the L_1-consistency is not only satisfied for a p.d.f. with compact support. It seems that the borderline between the L_1-consistency and nonconsistency of the P–R estimate corresponds to the exponential distribution (then $h \to 0$ in probability) (cf. [8]). This means that for a p.d.f. with lighter tails than exponential, e.g. a Gaussian p.d.f. or any p.d.f. with compact support, there are L_1-consistent P–R estimates. For a p.d.f. with heavier tails than an exponential p.d.f., a so-called heavy-tailed p.d.f., the L_1-consistency of the P–R estimates is not guaranteed.

To provide the L_1-consistency of the estimate, we use in this paper a transformation function $T : [0, \infty) \to [0, 1]$ reflecting the positive half of the real axis to the interval $[0, 1]$. This transformation may also be extended to $T : \mathbb{R} \to [0, 1]$. This idea was first proposed in [5] and later investigated without implementation in [8]. Exploiting this concept, we propose a specific transformation function in our paper. This function T transforms any long-tailed r.v. with positive values to a new one whose p.d.f. has a compact support, namely the interval $[0, 1]$. Then the estimation of the p.d.f. of this new r.v. is provided by a P–R estimate with compact and noncompact kernels and by a polygram, i.e. a histogram with variable bin width based on statistically equi-probable cells (cf. [17]). The inverse transformation stretches the estimators on the tail. The visual effect is in a way similar to using a variable smoothing parameter: this parameter is larger on the tail and smaller near the mode. Then, due to the invariance of the L_1 metric regarding any monotone continuous transformation, the L_1-accuracy of the p.d.f. of the initial r.v. is the same as the L_1-accuracy of that p.d.f. with compact support arising from the transformed r.v. Generally, a polygram works like an adaptive P–R estimate (cf. [15]). In our paper we investigate the accuracy of a P–R estimate and a polygram for a long-tailed p.d.f. if the new transformation function and different smoothing methods are applied.

The paper is organized as follows. In Section 2, we present the nonparametric estimates and the transformation of a long-tail distributed r.v. to a new r.v. whose p.d.f. has a compact support. The choice of the smoothing parameters, i.e. h for the P–R estimate and the number of points in the equi-probable cells for the polygram, by the discrepancy method is proposed. We also discuss the accuracy of these estimates. Section 3 contains the results of a simulation study of the P–R estimates and the polygram combined with the discrepancy and cross-validation methods as smoothing procedures. In Section 4, we estimate the p.d.f.s arising from real data of WWW-traffic characteristics measured at the University of Würzburg. We conclude with a summary of our findings.

2. Mathematical framework of the estimation approach

We observe a sample $X^l = (x_1, \dots, x_l)$ of independent observations of a r.v., e.g. the size of WWW responses, where l denotes the sample size. They are assumed to be distributed with the p.d.f. $f(x)$ and the d.f. $F(x)$. For the purpose of the data analysis we use the P–R estimate with a Gaussian kernel

$$f^1_{h,l}(t) = \frac{1}{lh\sqrt{2\pi}} \sum_{i=1}^{l} \exp\left(-\frac{1}{2}\left(\frac{t-x_i}{h}\right)^2\right), \tag{3}$$

the P–R estimate with Epanechnikov's kernel, which has a compact support,

$$f^2_{h,l}(t) = \frac{3}{4lh} \sum_{i=1}^{l} \left(1 - \left(\frac{t-x_i}{h}\right)^2\right) \Theta(h + x_i - t), \tag{4}$$

where

$$\Theta(t) = \begin{cases} 1, & t \geq 0, \\ 0, & t < 0, \end{cases}$$

and a polygram, i.e. a histogram with variable bin width

$$f_{k,l}(t) = \frac{k}{(l+1)\lambda(\Delta_{rk})} \tag{5}$$

for $t \in \Delta_{rk}$ (cf. [17]). Here, we assume that λ is Lebesgue's measure, $\lambda(\Delta_{rk}) \to 0$ and $k = o(l)$, that $x_{(1)}, \dots, x_{(l)}$ is the order statistics corresponding to the sample X^l and that the number of points inside each interval $\Delta_{1k} = [x_{(1)}, x_{(k)}]$, $\Delta_{2k} = (x_{(k)}, x_{(2k)}]$, $\Delta_{3k} = (x_{(2k)}, x_{(3k)}], \dots$ is less than or equal to k. The estimate (5) can be rewritten in the form

$$f_{k,l}(t) = \frac{k}{l+1} \sum_{r=0}^{[l/k]-1} \frac{\Theta(t - x_{(rk+1)})\,\Theta(x_{(k(r+1))} - t)}{x_{(k(r+1))} - x_{(rk+1)}} + \frac{l - k[l/k]}{l+1} \mathcal{I}\left(\left[\frac{l}{k}\right] \neq \frac{l}{k}\right) \psi(t,k), \tag{6}$$

where $[r]$ denotes the integer part of $r \in \mathbb{R}$ and

$$\mathcal{I}\left(\left[\frac{l}{k}\right] \neq \frac{l}{k}\right) = \begin{cases} 1, & \left[\frac{l}{k}\right] \neq \frac{l}{k}, \\ 0, & \left[\frac{l}{k}\right] = \frac{l}{k}, \end{cases} \quad \psi(t,k) = \begin{cases} 1, & l - 1 = \left[\frac{l}{k}\right]k, \\ \dfrac{\Theta(t - x_{([l/k]k+1)})\,\Theta(x_{(l)} - t)}{x_{(l)} - x_{([l/k]k+1)}}, & l - 1 \neq \left[\frac{l}{k}\right]k. \end{cases}$$

Let us first consider the P–R estimate and its properties. For $l \to \infty$ the convergence of the P–R estimate to the p.d.f. $f(x)$ depends on the choice of h. It was shown by Parzen that for a uniformly continuous f a P–R estimate converges in the metric space C in probability if

$$h \to 0, \quad lh^2 \to \infty, \quad l \to \infty, \tag{7}$$

whereas for the convergence with probability one it is sufficient that for any positive μ the series

$$\sum_{l=1}^{\infty} \exp(-\mu h^2 l) < \infty \tag{8}$$

converges (Nadaraya's result — cf. [18]).

The P–R estimate converges in the L_1 metric almost surely if

$$h + (lh)^{-1} \to 0, \quad l \to \infty \tag{9}$$

holds (cf. [8]). Since a P–R estimate is defined on $(-\infty, \infty)$, it can be applied to restore a long-tailed p.d.f. However, it is the main disadvantage of a P–R estimate that h is constant and cannot be adapted locally. Therefore, the behavior of a P–R estimate became poor for a nonsmooth and heavy-tailed p.d.f. In [8] it was proved that for the P–R estimate with a bounded and compact kernel function $l^{2/5}\mathbb{E}\int |f_{h,l}(x) - f(x)|\,\mathrm{d}x \to \infty$ holds for $l \to \infty$ for any value of h if $\int\sqrt{f}$ or (and) $\int|f''|$ is unlimited.

Let us now consider the polygram estimate $f_{k,l}$ in (5) and its features. Its L_1-convergence was shown by the following Theorem (cf. [1]).

Theorem 1. *For a polygram $f_{k,l}$ the following three assertions are equivalent:*
1. $\int_{-\infty}^{\infty}|f_{k,l}(x) - f(x)|\,\mathrm{d}x \to 0$ *in probability for any Riemann integrable p.d.f. f;*
2. $\int_{-\infty}^{\infty}|f_{k,l}(x) - f(x)|\,\mathrm{d}x \to 0$ *a.s. for any Riemann integrable p.d.f. f;*
3. $$k \to \infty, \frac{k}{l} \to 0 \quad \textit{as } l \to \infty. \tag{10}$$

Many authors have pointed out that histograms with equi-probable cells generally achieve better results than those with equal-sized cells (cf. [8,17]). The asymptotic convergence rate of (5) in the L_1 metric reaches $l^{-2/5}$ for some f, the same as for a P–R estimate. A histogram with equal-sized cells cannot achieve a convergence rate better than $l^{-1/3}$ in L_1. Since histogram-type estimates are defined on closed intervals, they cannot be applied directly to the estimation of a long-tailed p.d.f.

For a long-tailed p.d.f. the accuracy of a P–R estimate may be improved by the transformation of the initial r.v. to a new one whose p.d.f. has a compact support. We first estimate the p.d.f. of the transformed r.v. and apply then the inverse transformation. Such an estimation procedure is derived from the following theoretical results. Let $T : [0, \infty) \to [0, 1]$ be a monotone increasing continuous "one-to-one" transformation. The inverse transformation T^{-1} and the derivatives T', $(T^{-1})'$ are assumed to be continuous. Then the transformed sequence of X^l is given by $Y^l = (y_1, \ldots, y_l)$, where $y_i = T(x_i)$ holds. Let $g(x)$ be the p.d.f. and $G(x)$ be the d.f. of the r.v. y_1. $g(x) = f(T^{-1}(x))(T^{-1}(x))'$ is a p.d.f. located on [0, 1] since $(T^{-1}(x))'$ exists and is continuous for any x.

If $g_l(x)$ is some estimate of this p.d.f. constructed by Y^l, then the estimate of the unknown p.d.f. $f(x)$ is given by

$$f_l(x) = g_l(T(x))T'(x). \tag{11}$$

The remarkable effect is that the estimation error in the metric space L_1 is invariant for any continuous transformation (cf. [8, p. 244]):

$$\int_0^{\infty} |f_l(x) - f(x)|\,\mathrm{d}x = \int_0^1 |g_l(x) - g(x)|\,\mathrm{d}x.$$

An optimal transformation providing $\min_g \mathbb{E}\int_0^1|g_l(x) - g(x)|\,\mathrm{d}x$ in the case of the P–R estimate is determined by

$$T(x)=\begin{cases}\left(\dfrac{F(x)}{2}\right)^{1/2}, & F(x)\le 0.5,\\ 1-\left(\dfrac{1-F(x)}{2}\right)^{1/2}, & F(x)>0.5,\end{cases}$$

and in the case of the histogram estimate by $T(x)=F(x)$ (cf. [8]). Since the d.f. $F(x)$ is unknown, one can construct just the estimate $T_l(x)$ of $T(x)$. An application of the empirical d.f. $F_l(x)$ is not recommended since the derivative of $F_l(x)$ does not exist everywhere and it is zero on the intervals of constancy. The difference between $T_l(x)$ and $T(x)$ does not influence the asymptotic L_1-accuracy. But the accuracy of the estimation may be sensitive to $T_l(x)$ if the sample size is limited. Using some parametric estimate of $F(x)$ causes all the difficulties of parametric tail modeling. For these reasons, we choose as $T(x)$ a fixed transformation

$$T(x)=\frac{2}{\pi}\arctan x,\qquad T'(x)=\frac{2}{\pi(1+x^2)}, \tag{12}$$

which does not depend on the empirical sample X^l and satisfies for $x\in[0,1)$ all previously mentioned conditions about the transformation. Such a transformation $T(x)$ generates a bounded $g(x)$ for some heavy-tailed $f(x)$ (see Fig. 1).

Usually, $g_l(x)$ is not a p.d.f. on [0, 1] since a part of the distribution is located outside [0, 1]. Taking this issue into account, we use the estimate

$$\tilde{g}_l(x)=\frac{g_l(x)}{\int_0^1 g_l(x)\,\mathrm{d}x}$$

instead of $g_l(x)$. Then

$$\tilde{f}_l(x)=\tilde{g}_l(T(x))T'(x) \tag{13}$$

holds, and

$$\int_0^\infty|\tilde{f}_l(x)-f(x)|\,\mathrm{d}x=\int_0^1|\tilde{g}_l(x)-g(x)|\,\mathrm{d}x\le\int_0^1|g_l(x)-g(x)|\,\mathrm{d}x$$

follows, i.e. the L_1 risk of $\tilde{g}_l$ is better than that of g_l (cf. [8, p. 245]).

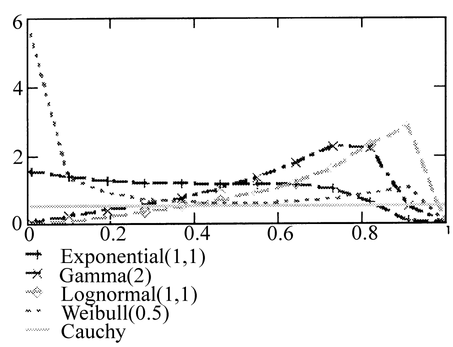

Fig. 1. Densities of a transformed r.v. generated by the transformation (12).

To overcome the boundary effects, which occur if the P–R estimate is fitted to a p.d.f. with compact support, the "mirror image" tool may be used (cf. [16]). In conclusion, the proposed algorithm to estimate a long-tailed p.d.f. looks as follows:

1. The nonparametric estimate g_l which is located on [0,1] is constructed by the transformed sample Y^l and normalized if it is necessary.
2. An estimate of the smoothing parameter of this estimate g_l is calculated.
3. To obtain the estimate of the p.d.f. $f(x)$, an inverse transformation is applied (see (11) and (13)).

For the transformation (12) the P–R estimates (3) and (4) of the transformed r.v. y_1 are determined by

$$g^1_{h,l}(x) = \frac{1}{lh\sqrt{2\pi}}\sum_{i=1}^{l}\exp\left(-\frac{1}{2}\left(\frac{x-y_i}{h}\right)^2\right), \tag{14}$$

$$g^2_{h,l}(x) = \frac{3}{4lh}\sum_{i=1}^{l}\left(1-\left(\frac{x-y_i}{h}\right)^2\right)\Theta(h+y_i-x), \tag{15}$$

respectively, where $y_i = (2/\pi)\arctan(x_i)$. By (13) we obtain after the normalization

$$\tilde{f}^1_{h,l}(x) = \frac{\sqrt{2}}{lh\pi^{3/2}I^1_{[0,1]}(h)(1+x^2)}\sum_{i=1}^{l}\exp\left(-\frac{1}{2}\left(\frac{(2/\pi)\arctan(x)-y_i}{h}\right)^2\right), \tag{16}$$

$$\tilde{f}^2_{h,l}(x) = \frac{3}{2\pi\, lhI^2_{[0,1]}(h)(1+x^2)}\sum_{i=1}^{l}\left(1-\left(\frac{(2/\pi)\arctan(x)-y_i}{h}\right)^2\right) \times\Theta\left(h+y_i-\frac{2}{\pi}\arctan(x)\right), \tag{17}$$

where

$$I^1_{[0,1]}(h) = \frac{1}{l}\sum_{i=1}^{l}\left(\Phi\left(\frac{1-y_i}{h}\right)-\Phi\left(-\frac{y_i}{h}\right)\right)$$

is the integral of $g^1_{h,l}(x)$ on [0, 1], $\Phi(x) = (1/\sqrt{2\pi})\int_{-\infty}^{x}\exp(-u^2/2)\,\mathrm{d}u$ is the Gaussian d.f. and

$$I^2_{[0,1]}(h) = \frac{3}{4lh}\sum_{i=1}^{l}\begin{cases}1-\dfrac{1}{3h^2}(1+3y_i(y_i-1)) & \text{for } h+y_i>1,\\[2ex] \dfrac{2}{3}h+y_i\left(1-\dfrac{y_i^2}{3h^2}\right) & \text{for } h+y_i\le 1\end{cases}$$

is the integral of $g^2_{h,l}(x)$ on [0, 1].

Let $g_{k,l}(x)$ be a polygram constructed on Y^l by the formula (6). We get after the inverse transformation (11) (since a normalization is not necessary):

$$f_{k,l}(x) = \frac{2}{\pi(1+x^2)}g_{k,l}\left(\frac{2}{\pi}\arctan(x)\right). \tag{18}$$

Let us now discuss the selection of the parameters h and k determining the accuracy of the P–R estimate and a polygram.

The conditions (7)–(9) and Theorem 1 recommend the a priori choice (before the calculations begin) of h and k as functions of the sample size l. The practice shows, however, that it is better to select the smoothing parameter depending on fixed sample points if the sample size is limited. According to the c.v.m. h or k are chosen in cross-validated density estimation as maximum of the likelihood-like expression

$$L_l = \prod_{i=1}^{l} g_{l-1}^{i}(y_i), \tag{19}$$

where g_{l-1}^{i} is the P–R estimate or a polygram based on the random sample Y^l excluding the ith observation (cf. [5]).

Here, we select h and k by an alternative based on the discrepancy principle, the so-called ω^2- and D-methods (cf. [12,13]). It is the essence of this principle to obtain h (or k) in the case of the ω^2-method from the equality

$$l\hat{\omega}_l^2 = \sum_{i=1}^{l} \left(G^h(y_{(i)}) - \frac{i-0.5}{l} \right)^2 + \frac{1}{12l} = 0.05, \tag{20}$$

or, in the case of the D-method, from the equality

$$\hat{D}_l = \max(\hat{D}_l^+, \hat{D}_l^-) = 0.5, \tag{21}$$

where

$$\hat{D}_l^+ = \sqrt{l} \max_{1\le i\le l} \left(\frac{i}{l} - G^h(y_{(i)}) \right), \qquad \hat{D}_l^- = \sqrt{l} \max_{1\le i\le l} \left(G^h(y_{(i)}) - \frac{i-1}{l} \right),$$

and $y_{(1)} \le y_{(2)} \le \cdots \le y_{(l)}$ is the order statistics of the transformed observations. For the normalized P–R estimate (14)

$$G_1^h(x) = \frac{1}{I_{[0,1]}^1(h)} \int_0^x g_{h,l}^1(t)\, \mathrm{d}t = \frac{1}{lI_{[0,1]}^1(h)} \sum_{i=1}^{l} \left(\Phi\left(\frac{x-y_i}{h} \right) - \Phi\left(-\frac{y_i}{h} \right) \right)$$

holds, for the normalized P–R estimate (15) we get

$$G_2^h(x) = \frac{1}{I_{[0,1]}^2(h)} \int_0^x g_{h,l}^2(t)\, \mathrm{d}t = \frac{3}{4lhI_{[0,1]}^2(h)} \sum_{i=1}^{l} \begin{cases} x - \dfrac{1}{3h^2}((x-y_i)^3 + y_i^3), & h + y_i \ge x, \\ x - \dfrac{1}{3h^2}(h^3 + y_i^3), & h + y_i < x. \end{cases}$$

Let $[x]$ denote now the smallest integer larger than x. For a polygram we have

$$\hat{D}_l = \sqrt{l} \left(\frac{k}{l+1} - \frac{1}{l} \right),$$

$$k = \left[\left(\frac{0.5}{\sqrt{l}} + \frac{1}{l} \right) (l+1) \right], \tag{22}$$

provides the solution of (21). It satisfies (10) and guarantees the L_1-convergence for a Riemann integrable p.d.f. The advantage is that the statistics $l\hat{\omega}_l^2$ and $\hat{D}_l$ are based on the observed (ungrouped) sample points. In [12] the ω^2- and D-methods have been investigated empirically by computer simulation for small samples and different distributions. Considering the L_2-distance as loss function, they have provided better results for P–R estimates and for nonsmooth distributions, e.g. triangle and uniform distributions, than the c.v.m. and the same results for smooth distributions. It was proved in [19] that if h is selected by the ω^2-method then the L_2-risk of the estimation is the best for the p.d.f. with a bounded variation of the kth derivative.

3. A simulation study of the estimates

Performing an experimental study presented in this section, we have compared the P–R estimates with the compact and noncompact kernel functions (16) and (17) and a polygram (18) for different long-tailed p.d.f.s. Regarding the selection of the smoothing parameters the ω^2- and D-methods have been compared with the c.v.m.

For the comparison we have generated samples of the known p.d.f.s

$$f_1(x) = \begin{cases} \dfrac{x^{s-1}\exp(-x)}{\Gamma(s)}, & x > 0, \\ 0, & x \le 0 \end{cases}$$

of a Gamma distribution with the parameter $s = 2$,

$$f_2(x) = \begin{cases} \dfrac{1}{\sqrt{2\pi}\sigma x}\exp\left(-\dfrac{1}{2\sigma^2}(\ln x - \mu)^2\right), & x > 0, \\ 0, & x < 0 \end{cases}$$

of a Lognormal distribution with $\mu = 1$, $\sigma = 1$ and

$$f_3(x) = \begin{cases} sx^{s-1}\exp(-x^s), & x > 0, \\ 0, & x < 0 \end{cases}$$

of a Weibull distribution with $s = 0.5$. The Gamma distribution is related to the light-tailed distributions, but the Lognormal and Weibull p.d.f. are heavy-tailed.

As characteristics of the estimates, we used the loss functions

$$\chi^1 = \int_0^\infty |f_l(x) - f_0(x)|\,\mathrm{d}x = \int_0^1 |g_l(x) - g_0(x)|\,\mathrm{d}x,$$
$$\chi^2 = \int_0^\infty (f_l(x) - f_0(x))^2\,\mathrm{d}x = \frac{2}{\pi}\int_0^1 \frac{(g_l(x) - g_0(x))^2}{1 + [\tan((\pi/2)\,x)]^2}\,\mathrm{d}x,$$
$$\chi^3 = \sup_{i=1,\dots,l} |f_l(x_i) - f_0(x_i)|,$$

where $f_l(x)$, $g_l(x)$ are the estimates of the p.d.f. and $f_0(x)$, $g_0(x)$ are the exact models of the p.d.f. arising from the initial and the transformed r.v. The normalized P–R estimates (14) and (15) and the polygram

Table 1
Comparison of the estimation methods for a Gamma distribution

Method		Size	$\bar{\rho}_1$	$\sigma_1^2 \times 10^3$	$\bar{\rho}_2$	$\sigma_2^2 \times 10^3$	$\bar{\rho}_3$	$\sigma_3^2 \times 10^3$	$\bar{h}/\bar{k}(\sigma_h^2 \times 10^4)$
(16)	ω^2	50	0.204	5.013	0.014	0.113	0.139	3.601	0.066 (3.793)
		100	0.158	2.016	0.008	0.035	0.114	2.309	0.064 (3.003)
(17)	ω^2	50	1.411	3.819	1.326	6.796	1.649	176	0.059 (4.365)
		100	1.413	3.633	1.331	4.916	1.851	150	0.067 (2.866)
	c.v.m.	50	1.4	3.814	1.303	6.558	1.634	169	0.03 (6.681)
		100	1.398	3.454	1.298	3.99	1.826	147	0.033 (4.137)
(18)	D	50	0.587	65	0.121	1.814	1.077	678	5
		100	0.508	28	0.099	1.512	0.61	101	7

(6) are used as $g_l(x)$. We have considered samples of the size $l = 50$, 100 and 300. For each size we have constructed 25 realizations. Then, we calculated the statistics

$$\bar{\rho}_j = \frac{1}{n}\sum_{i=1}^{n}\chi_i^j, \qquad \sigma_j^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\chi_i^j - \bar{\rho}_j)^2, \quad n = 25, \quad j = 1, 2, 3.$$

Based on the latter, we have compared the accuracy of the p.d.f. estimates and of the methods for selecting their smoothing parameters. We also calculated the means $\bar{h}, \bar{k}$ and the standard deviation σ_h^2 of the parameters h and k on the basis of these 25 realizations. The values of the statistic $l\hat{\omega}_l^2$ were chosen with an error of 2%.

Considering the results of the simulation study shown in Tables 1–3, the following observations can be made:

1. If the sample size increases, the polygram converges to f_1, f_2 in the metrics L_1, L_2 and C and to f_3 just in the metrics L_1 and L_2.
2. The P–R estimate (16) provides the convergence to f_1 and f_2 in L_1, L_2 and C and does not converge to f_3 in any metric.
3. The P–R estimate (17) converges to f_1 (for c.v.m.), f_2 and f_3 in the metrics L_1 and L_2 and does not converge in the metric C for both methods of smoothing.
4. The P–R estimate (16) is more accurate than a polygram and (17) for f_1 and f_2 and it is worse for f_3.
5. The P–R estimate (17) shows worse results than a polygram and the P–R estimate (16) for f_1 and f_2.
6. The ω^2-method provides the same results as a c.v.m.

Table 2
Comparison of the estimation methods for a Lognormal distribution

Method		Size	$\bar{\rho}_1$	$\sigma_1^2 \times 10^3$	$\bar{\rho}_2$	$\sigma_2^2 \times 10^3$	$\bar{\rho}_3$	$\sigma_3^2 \times 10^3$	$\bar{h}/\bar{k}(\sigma_h^2 \times 10^4)$
(16)	ω^2	50	0.337	7.956	0.022	0.1783	0.173	8.635	0.032 (0.9293)
		100	0.237	6.483	0.011	0.0888	0.15	15	0.033 (1)
(17)	ω^2	50	1.493	3.9	1.154	3.452	1.102	133	0.035 (0.9856)
		100	1.493	3.796	1.154	2.736	1.326	120	0.037 (0.7879)
	c.v.m.	50	1.494	3.951	1.156	3.557	1.103	134	0.031 (6.256)
		100	1.491	3.799	1.151	2.781	1.324	119	0.029 (1.564)
(18)	D	50	0.579	110	0.072	0.3685	0.621	484	5
		100	0.492	34	0.061	0.2429	0.298	34	7

Table 3
Comparison of the estimation methods for a Weibull distribution

Method		Size	$\bar{\rho}_1$	$\sigma_1^2 \times 10^3$	$\bar{\rho}_2$	$\sigma_2^2 \times 10^3$	$\bar{\rho}_3$	$\sigma_3^2 \times 10^3$	$\bar{h}/\bar{k}(\sigma_h^2 \times 10^4)$
(16)	ω^2	50	0.498	31	0.551	442	199.1	1.639×10^8	0.023 (1.448)
		100	0.579	41	0.966	544	319.1	9.506×10^8	0.011 (1.066)
(17)	ω^2	50	0.722	1.052	0.607	3.172	199.32	1.64×10^8	0.023 (1.422)
		100	0.72	0.9683	0.599	1.691	320.49	9.5×10^8	0.01 (0.409)
		300	0.718	0.8819	0.592	0.8622	474.88	1.015×10^9	0.003 (0.0266)
	c.v.m.	50	0.986	527	1.406	6557	198.8	1.64×10^8	0.181 (430)
		100	0.738	8.925	0.627	46	320.49	9.5×10^8	0.075 (180)
(18)	D	50	0.354	13	0.385	598	185.56	1.627×10^8	5
		100	0.337	11	0.307	533	306.16	9.365×10^8	7
		300	0.262	8.928	0.16	22	433.07	1.003×10^{10}	10

It follows from the simulation study that a polygram and the P–R estimate (16) are preferable for the application to real data if the true p.d.f. is not available. If one knows that the p.d.f. is heavy-tailed, then a polygram is recommended.

4. Data analysis of WWW-traffic characteristics

To illustrate the proposed nonparametric estimation approach, we have analyzed real data of WWW-traffic measured in the Ethernet segment of the Department of Computer Science at the University of Würzburg in 1997. For further details of the data gathering the readers are referred to [20].

4.1. Description of the WWW-traffic data

The data are described by a hierarchical model distinguishing a session and a page level. The first one is characterized by sub-sessions in [20]. Consequently, the data are described by two basic characteristics and four related r.v.s, namely, the characteristics of sub-sessions, i.e. the size of a sub-session (s.s.s.) in bytes and the duration of a sub-session (d.s.s.) in seconds, as well as the characteristics of the transferred WWW-pages, i.e. the size of the response (s.r.) in bytes and the inter-response time (i.r.t.) in seconds were measured (see Table 4). The sub-session has an average size of 1.283×10^6 byte and the average duration of 1.728×10^3 s, the variance of the s.s.s. is 1.664×10^{13} byte and the variance of the d.s.s. is 2.71×10^7 s, minimal and maximal s.s.s. are 128 byte and 5.884×10^7 byte, minimal and maximal d.s.s. are 2 and 9.058×10^4 s. The sample size l is 373 for both samples. For simplicity of the calculations, the data were scaled, namely, the s.s.s. was divided by 10^7 and the d.s.s. by 10^3.

Table 4
Modeling of WWW-sessions

Level	Characteristic	Definition
Sub-session	Duration	Time between beginning and termination of browsing a series of Web pages
	Size	Data volume of visited Web pages
Page	Inter-response time	Time between beginning of the old and of the new transfer of pages within a sub-session
	Size of response	Total amount of transferred data (HTML, images, sound, etc.)

The data of the WWW-page sizes contain the information about 7480 WWW-pages which have been down-loaded during 14 days by several TCP/IP connections. The size of a response is defined as the sum of the sizes of all packets which are down-loaded from a WWW-server to the client upon a request. To perform the analysis, we have used samples with the reduced sample size $l = 1000$ for the sizes and inter-response times of WWW requests, which have been observed in a shorter period within these two weeks.

The mean s.r. is 3.467×10^4 byte and the variance 2.853×10^{10} byte, the minimal and maximal s.r. are 1×10^{-3} byte and 4.351×10^6 byte. The mean and the variance of the i.r.t. are 73.282 and 4.184×10^4 s, the minimal and maximal i.r.t. are 0.03 and 3.277×10^3 s. For scaling the s.r. was divided by 10^6 and the i.r.t. by 10^3.

4.2. Results of the statistical data analysis

The statistical analysis of the underlying four r.v.s of the WWW-traffic characteristics is similar. First, we have checked whether an exponential d.f. $F_{\text{exp}}(t, \lambda) = 1 - \exp(-\lambda t)$ for $t > 0$ or a Pareto d.f. $F_{\text{p}}(t, \alpha) = 1 - t_0^{\alpha} t^{-\alpha}$ for $\alpha > 0$, $t \geq t_0 = 10^{-3}$ (for each sample) is an appropriate model for the d.f. of the corresponding r.v. Then we have estimated the p.d.f. of each r.v. by a P–R estimate (16) and a polygram (18) using the smoothing methods (20) and (21).

Maximum likelihood estimates were calculated by the formulas

$$\lambda = \left(\frac{1}{l} \sum_{i=1}^{l} x_i \right)^{-1}, \qquad \alpha = \left(\frac{1}{l} \sum_{i=1}^{l} \ln(x_i) - \ln(t_0) \right)^{-1}$$

for the exponential and the Pareto d.f., respectively, where $x_1, \ldots, x_l$ denotes the s.s.s., d.s.s., s.r. or i.r.t. sample. For the s.s.s. sample we obtained $\lambda = 7.795$, $\alpha = 0.305$, for the d.s.s. $\lambda = 0.579$, $\alpha = 0.161$, for the s.r. $\lambda = 28.846$, $\alpha = 0.483$, and for the i.r.t. $\lambda = 13.646$, $\alpha = 0.344$.

Let $F_l(t) = (1/l) \sum_{i=1}^{l} \Theta(t - x_i)$ denote the empirical d.f. In Figs. 2–5 the survival functions $1 - F_l(t)$, $1 - F_{\text{exp}}(t, \lambda)$ and $1 - F_{\text{p}}(t, \alpha)$ are shown for the s.s.s., d.s.s., s.r. and i.r.t. samples. The application of the Kolmogorov–Smirnov (K–S) test shows that no sample follows an exponential or Pareto distribution

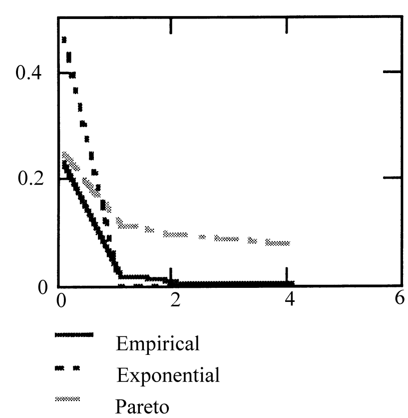

Fig. 2. Size of the sub-session sample: estimation of the survival functions.

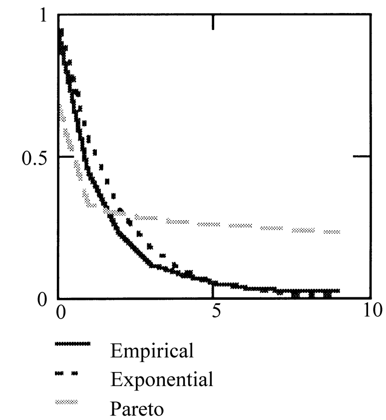

Fig. 3. Duration of the sub-session sample: estimation of the survival functions.

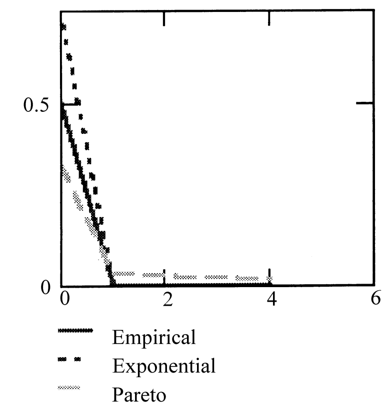

Fig. 4. Size of the response sample: estimation of the survival functions.

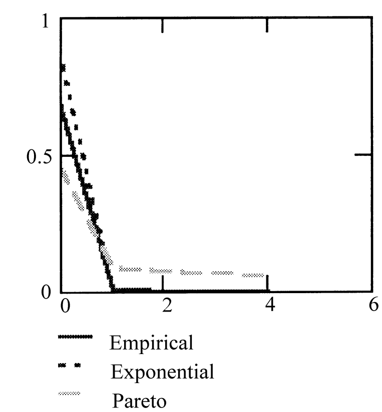

Fig. 5. Inter-response time sample: estimation of the survival functions.

despite of the visual similarity of these models. Since the samples contain more than 100 points, the quantiles of the K–S statistic have been estimated by the formula (cf. [4])

$$\tilde{D}_l(Q) = \sqrt{\frac{y}{2l}} - \frac{1}{6l} \quad \text{with} \quad y = -\ln(0.005Q),$$

where Q is the confidence level. For $Q = 5$, we get $\tilde{D}_l(Q) = 0.07$ for $l = 373$ and $\tilde{D}_l(Q) = 0.043$ for $l = 1000$. The values of the K–S statistic

$$\frac{D_l}{\sqrt{l}} = \max\left\{\sup_{1\le i\le l}\left(\frac{i}{l} - F(x_{(i)})\right), \sup_{1\le i\le l}\left(F(x_{(i)}) - \frac{i-1}{l}\right)\right\},$$

that were calculated for the exponential and Pareto d.f. $F(x)$ by the empirical samples are given by 0.281 and 0.229 for the s.s.s., by 0.157 and 0.344 for the d.s.s., by 0.276 and 0.217 for the s.r., and by 0.282 and 0.259 for the i.r.t., respectively. Since $(D_l/\sqrt{l}) > \tilde{D}_l(Q)$ holds in each case, the H_0-hypothesis that the empirical distribution coincides with the selected theoretical one should be rejected.

In Figs. 6–9 the polygram and the P–R estimate are presented for the s.s.s., d.s.s., s.r. and i.r.t. samples, respectively. Each figure depicts two graphs to demonstrate better the behavior on the tails and for small values. All graphs were constructed in the points $x_{(1)}, \dots, x_{(l)}$. Both estimates were first applied to the samples transformed by (12), i.e. $\{y_i = 2/\pi \arctan(x_i), i = 1, \dots, l\}$, and then the inverse transformations (11) for a polygram and (13) for a P–R estimate were used. The polygrams were calculated by the formulas (18) and (6) (applied to $g_{k,l}$), the P–R estimates by (16).

The parameter h of the P–R estimate was computed by the ω^2- and D-methods, i.e. by the Eqs. (20) and (21), called P–R estimate 1 and P–R estimate 2, respectively. The values $h \in \{7.5 \times 10^{-4}, 8.1 \times 10^{-3}, 1.75 \times 10^{-4}, 2.3 \times 10^{-4}\}$ are provided for $l\hat{\omega}_l^2 = 0.05$ for the s.s.s., d.s.s., s.r. and i.r.t. samples, respectively. The values $h \in \{3.6 \times 10^{-3}, 9.5 \times 10^{-5}, 2.6 \times 10^{-4}\}$ are provided for $\hat{D}_l = 0.5$, for the d.s.s., s.r. and i.r.t. samples, respectively. For s.s.s. $\hat{D}_l$ never reaches its maximum likelihood value for any h and we did not apply the D-method for the s.s.s. We see that the discrepancy methods ω^2 and D select similar values of h. The parameter k of the polygram was only calculated by the D-method (see (22)). k is equal to 11 for $l = 373$ and 17 for $l = 1000$.

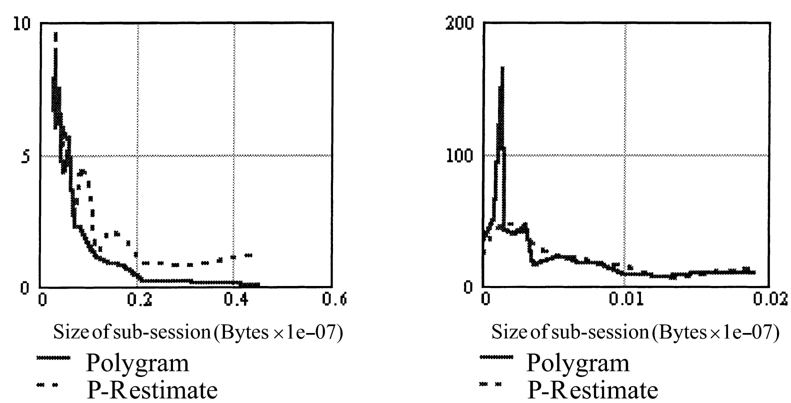

Fig. 6. Size of the sub-session sample: estimation of the probability density function.

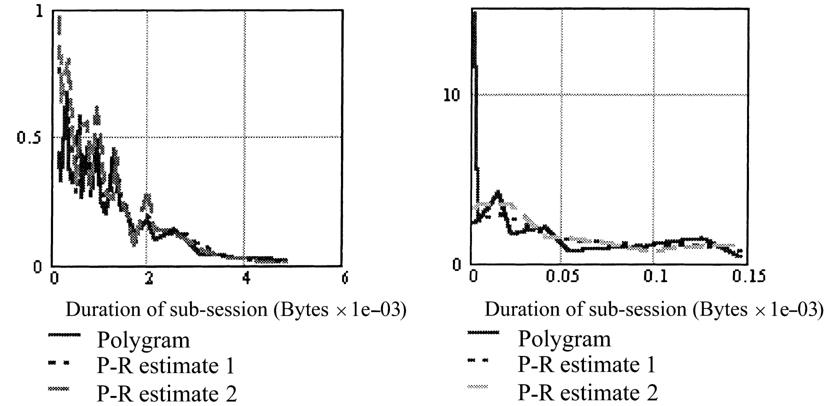

Fig. 7. Duration of the sub-session sample: estimation of the probability density function.

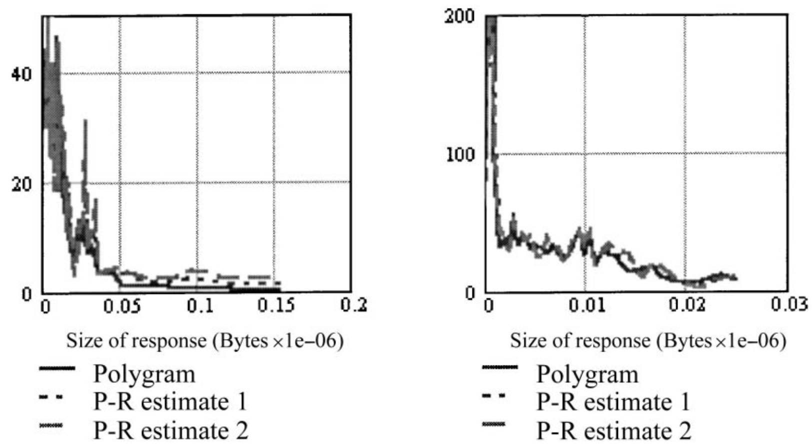

Fig. 8. Size of the response sample: estimation of the probability density function.

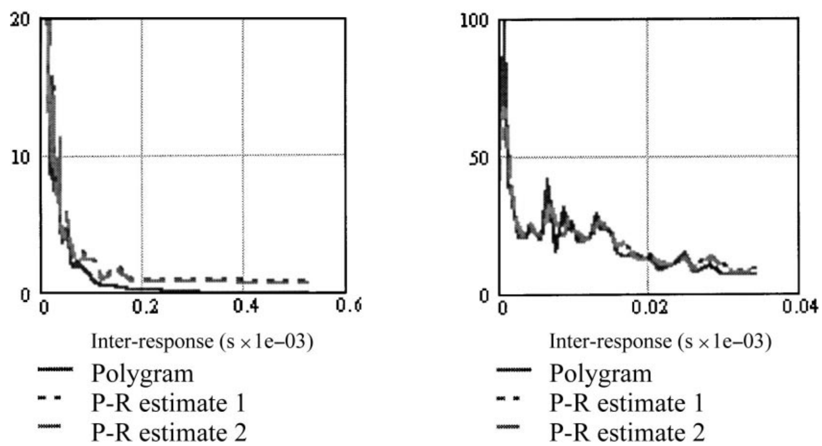

Fig. 9. Inter-response time sample: estimation of the probability density function.

The P–R estimate and the polygram restore the tail of the p.d.f. in a similar manner for each considered r.v. except the s.s.s. The difference between the estimates occurs for the small values. The maximal values of the polygram and the P–R estimate are given by 165.049 and 48.518 for s.s.s., 14.7 and 2.277 for d.s.s., 999.001 and 196.728 for r.s. and 98.605 and 70.557 for i.r.t., respectively. Due to the small distances between the order statistics near zero the polygrams may have big values. The P–R estimate is smoother. The difference became smaller for large sample sizes.

5. Conclusions

Considering the WWW-traffic characterization in the Internet, we have developed in this paper a new statistical methodology to analyze the corresponding measurements of limited size. We have proposed a nonparametric framework to estimate the underlying long-tailed p.d.f. of a relevant r.v. Following this approach, we have assumed that just general information about the kind of the distribution is available. To implement the proposed approach, a Parzen–Rosenblatt kernel estimate and a histogram with statistically equi-probable cells, called a polygram, are selected.

To improve the behavior of the P–R estimates on the tails and to get L_1-consistent estimates for the long-tailed p.d.f., the transformation of the initial r.v. to a new one having a p.d.f. with a compact support on the interval [0, 1] is proposed. Its introduction allows us to apply apart from the P–R estimate those estimates defined on a closed interval such as a histogram or projection estimates. Furthermore, an algorithm to construct L_1-consistent estimates has been described.

From a practical point of view, we are interested in the accuracy of the estimation for empirical samples of limited size. The reliability of the estimates is provided by the selection of smoothing parameters. In the paper two discrepancy-type methods, i.e. the ω^2- and D-method, are used to select these parameters. They provide the estimation based on the observed ungrouped sample points. To our best knowledge, the D-method is applied in this paper for the first time to smooth a polygram. The ω^2- and D-methods provide a sufficient accuracy and are simpler to apply than the cross-validation method. By a simulation study we have shown that for a heavy-tailed Weibull distribution a polygram and the P–R estimate with a compact kernel are reliable in the L_1 and L_2 metric.

To illustrate the power of the proposed estimation approach, we have finally applied it to analyze measurements arising from WWW-traffic. The latter were gathered at the Computer Science Department of the University of Würzburg. Using these real data, the p.d.f.s of relevant WWW-traffic characteristics have been estimated. We have shown that exponential and Pareto distributions are not appropriate models for the densities of the underlying r.v.s. It was demonstrated that the P–R estimate and a polygram work in a similar manner on the tails and are different for small values of a r.v. if the sample size is sufficiently small. The difference between these estimates became less for large sample sizes.

In conclusion, we have pointed out a new effective way to cope with the thorough statistical analysis of measured data of WWW-traffic characteristics. This sound data analysis is the first and one of the most decisive steps towards an effective design of the IP-based transport infrastructure in the extremely variable environment of the Internet.

Acknowledgements

We appreciate the kind assistance of Professor P. Tran-Gia and N. Vicari, Department of Computer Science of the University of Würzburg, who provided us with their measurements and supported us by their helpful explanations of the gathered data. Furthermore, we are grateful to the referees for their helpful suggestions that improved the presentation of the paper.

References

[1] S. Abou-Jaoude, Sur la convergence L_1 et L_∞ de l'estimateur de la partition aliatoire pour une densité, Ann. Inst. Henri Poincaré B 12 (1976) 299–317.
[2] P. Barford, M.E. Crovella, Measuring Web performance in the wide area, Perform. Eval. Rev. (August 1999).
[3] A.R. Barron, L. Györfi, E.C. van der Meulen, Distribution estimation consistent in total variation and in two types of information divergence, IEEE Trans. Inform. Theory 38 (1992) 1437–1454.
[4] L.N. Bol'shev, N.V. Smirnov, Tables of Mathematical Statistics, Nauka, Moscow, 1968 (in Russian).
[5] Y.-S. Chow, S. Geman, L.-D. Wu, Consistent cross-validated density estimation, Ann. Statist. 11 (1983) 25–38.
[6] M.E. Crovella, M.S. Taqqu, A. Bestavros, Heavy-tailed probability distributions in the World Wide Web, in: R. Adler et al. (Eds.), A Practical Guide to Heavy Tails: Statistical Techniques and Applications, Birkhauser, Boston, MA, 1998, pp. 3–25.
[7] S. Csörgö, Asymptotic methods in probability and statistics, in: B. Szyszkowicz (Ed.), A Volume in Honour of Miklos Csörgö, Elsevier, Amsterdam, 1998, pp. 833–881.
[8] L. Devroye, L. Györfi, Nonparametric Density Estimation, the L_1 View, Wiley, New York, 1985.
[9] R.H. Farell, On the best obtainable asymptotic rates of convergence in estimation of a density function at a point, Ann. Math. Statist. 43 (1972) 170–180.
[10] R.Z. Has'minskiĭ, A lower bound for risks of nonparametrical estimates of density in the uniform metrics, Prob. Theory Appl. 23 (4) (1978) 824–828.
[11] B.M. Hill, A simple general approach to inference about the tail of a distribution, Ann. Statist. 3 (1975) 1163–1174.
[12] N.M. Markovich, Experimental analysis of nonparametric probability density estimates and of methods for smoothing them, Auto. Remote Control 50 (7, Part 2) (1989) 941–948.
[13] V.A. Morozov, Methods for Solving Incorrectly Posed Problems, Springer, Berlin, 1984.
[14] M. Nabe, M. Murata, H. Miyahara, Analysis and modelling of World Wide Web traffic for capacity dimensioning of Internet access lines, Perform. Eval. 34 (1998) 249–271.
[15] E.F. Schuster, G.G. Gregory, On the nonconsistency of maximum likelihood nonparametric density estimates, in: W.F. Eddy (Ed.), Computer Science and Statistics, Proceedings of the 13th Symposium on the Interface, Springer, New York, 1981, pp. 295–298.
[16] E.F. Schuster, Incorporating support constraints into nonparametric estimators of densities, Commun. Statist. Theory Meth. 14 (5) (1985) 1123–1136.
[17] F.P. Tarasenko, On the evaluation of an unknown probability density function, the direct estimation of the entropy from independent observations of a continuous random variable and the distribution-free test of goodness-of-fit, Proc. IEEE 56 (11) (1968) 2052–2053.
[18] V.N. Vapnik, Estimation of Dependences Based on Empirical Data, Springer, New York, 1982.
[19] V.N. Vapnik, N.M. Markovich, A.R. Stephanyuk, Rate of convergence in L_2 of the projection estimator of the distribution density, Auto. Remote Control 53 (1992) 677–686.
[20] N. Vicari, Measurement and modelling of WWW-sessions, Technical Report No. 184, Institute of Computer Science, University of Würzburg, September 1997.

Natalia M. Markovitch received her M.S. degree in mathematics from the Applied Mathematics and Cybernetics Department of Tver University and the Ph.D. in statistics from the Institute of Control Sciences of the Russian Academy of Sciences in Moscow. She was affiliated with the All-Union Centre for Preventive Medicine and the Toxicology Information and Advisory Centre of the Russian Federation Ministry of Health in Moscow. Since 1995 she is working as a Senior Scientist at the Institute of Control Sciences of the Russian Academy of Sciences in Moscow. In 1995 and 1996 she was a visiting scientist at the Geomedical Research Unit of the Heidelberg Academy for the Humanities and Sciences in Heidelberg, in 1997 at the Max Planck Institute for Demographic Research in Rostock, Germany and since 1998 at the Research Center of Deutsche Telekom AG in Darmstadt, Germany. Her research interests include the statistical analysis of empirical data, simulation and stochastic optimization.

Udo R. Krieger received his M.S. degree in applied mathematics and the Ph.D. in computer science from the Technische Hochschule Darmstadt, Germany. In 1985 he joined the Research and Technology Center of Deutsche Telekom, now called T-Nova, in Darmstadt, Germany, as a Research Assistant of the group Digital Switching Systems. Since 1993, he is a member of the group Traffic Management. He is further affiliated with the Computer Science Department of J.W. Goethe-University in Frankfurt. His research interests include traffic management of ATM, IP and wireless networks, teletraffic theory, and numerical solution methods for Markov chains. He is a member of Gesellschaft für Informatik, Germany, and IEEE.