ORNL/TM-13468

Computer Science and Mathematics Division

Mathematical Sciences Section

POISSON TYPE MODELS AND DESCRIPTIVE STATISTICS OF COMPUTER NETWORK INFORMATION FLOWS

D. Downing and V. Fedorov
Statistics Group
T. Dunigan and S. Batsell
Network Research Group

Mathematical Sciences Section
Computer Science and Mathematics Division
Oak Ridge National Laboratory
P. O. Box 2008
Oak Ridge, Tennessee 37831-6367

Date Published: August 1997

Research sponsored by the Laboratory Directed Research and Development Program of Oak Ridge National Laboratory

Prepared by the
Oak Ridge National Laboratory
Oak Ridge, Tennessee 37831
managed by
Lockheed Martin Energy Research Corp.
for the
U.S. DEPARTMENT OF ENERGY
under Contract No. DE-AC05-96OR22464

Contents

List of Figures

POISSON TYPE MODELS AND DESCRIPTIVE STATISTICS OF COMPUTER NETWORK INFORMATION FLOWS

D. Downing and V. Fedorov

Statistics Group

T. Dunigan and S. Batsell

Network Research Group

Abstract

Many contemporary publications on network traffic gravitate to ideas of self-similarity and long-range dependence. The corresponding elegant and parsimonious mathematical techniques proved to be efficient for the description of a wide class of aggregated processes. Sharing the enthusiasm about the above ideas we also believe that whenever it is possible any problem must be considered at the most basic level in an attempt to understand the driving forces of the processes under analysis. We confine ourselves to a rather modest (in the sense of "physical size") network and show that some behavioral patterns of descriptive statistics which are typical for long-memory processes (a particular case of long-range dependence) can also be explained in the framework of the traditional Poisson process paradigm.

Applying the concepts of inhomogeneity, compoundness and double stochasticity we propose a simple and intuitively transparent approach of explaining the expected shape of the observed histograms of counts and the expected behavior of the sample covariance functions. Matching the images of these two descriptive statistics allows us to infer the presence of trends or double stochasticity in analyzed time series. We considered only statistics which are based on counts. A similar approach may be applied to waiting or inter-arrival time sequences and will be discussed in other publications. We hope that combining the reported results with the statistical methods based on aggregated models may lead to computationally affordable "on-line" techniques of compact and visualized data analysis of network flows.

1. Introduction

At present in computer network analysis there exists a rather strong tendency to depart from the traditional probabilistic models and related statistical methods, which are based on relatively simple "physical" assumptions about the background stochastic processes. The movement is directed towards models and ideas which provide a reasonable description of what is seen on the "surface" of the process without a detailed explanation of what is going on at the deeper levels. Two typical examples are closely related, self-similar processes and long-memory processes models and techniques (c.f. Beran (1992) Beran (1994), Gustafsson and Karlsson (1997), Samorodnitsky and Taqqu (1994), Willinger et al. (1995) Willinger et al. (1996)). Those approaches lead to simple and elegant methods of description of highly aggregated processes, in which the aggregation reaches the level when component-wise analysis may result in desperately complicated mathematical models or in ill-conditioned problems in data analysis.

There exist intermediate cases which can be studied in the framework of either direction: the choice is defined by the level of details which is needed for research (compare with Beran (1994), Chpt.1, Grossglauser and Bolot (1996)). In this particular study on open local area network flows, we gravitate to the traditional models and try to explain the unusual behavior of empirical statistics within the Poisson process paradigm. We have found that relatively simple models of the observed processes and sampling procedures explain effects which may be accredited by others to self-similarity or long-term memory. Only histograms and sample covariance functions were used in this study. However the approach may be extended to other popular descriptive statistics such as periodograms or statistics related to inter-arrival times.

It is clear from Section 2 that the available data provide an excellent test-ground for such models as inhomogeneous, compound, doubly stochastic Poisson processes or any mixture of them. For instance, applying the compound process model, we can use as markers the size of the transmitted packets, their type, or the length of packet trains when a file is partitioned in packets (compare with Jain and Routhier (1986)). We confine ourself to the simplest model (Section 3) which provides at least qualitative explanations of the observed histograms. Piece-wise linear approximation combined with simple stochastic models very satisfactorily explains histograms constructed on different time scales. It was found that under some simple assumptions these histograms must look very similar. This fact, to some extent, contradicts the intuitive guess that the smaller the time interval that is used to build a histogram the closer it will appear as a Poisson distribution. Simple formulae are derived to describe the expected shape of histograms (see, for instance, (3.7), (3.9)); we also use well known formulae in the theory of doubly stochastic Poisson processes (see, for instance, (3.12)) but in an unusual deterministic setting. We showed that either trend or double stochasticity may cause the "flattening" of histograms. To reduce the set of possible solutions additional statistics must be used. In particular, in Section 5 we proved that trend and double stochasticity may be separated when the histograms and the sampled covariance functions are viewed together.

Section 4 complements the analytical component of the paper with some data-oriented computations. We propose an estimation procedure based on the reiterated least squares method (c.f. Charnes et al. (1976), Fedorov (1974), Seber and Wild (1989)) and provide some theoretical background for it.

2. Data

The network traffic flows used in this analysis were captured by a dedicated Sun SPARC workstation attached to a 100 megabit/second FDDI (Fiber Distributed Data Interface) ring that interconnects the local network to the Internet (see Figure 2.1). The capture process was non-intrusive with respect to the data being collected. The collector program was written at ORNL using the promiscuous mode of the network interface. The collector program, similar to one described by Claffy (1994), writes a trace record of every IP packet to disk. Each trace record consists of a date-time stamp (microsecond resolution), and source and destination addresses and service port number. For this analysis, trace records were collected for a 24 hour period during the work week of January, 1997. The trace consisted of 40,937,472 packets (a 1.1 gigabyte file) representing typical Internet traffic (file transfer, email, telnet, and world-wide web) between over 40,000 hosts (4,222 were local hosts). There are two obvious features in this data set. Fig.2.2 shows a plot of the cumulative distribution for packet size for each hour of the day. The cumulative distribution for a given hour is weighted by the proportion of the total flow for the day by that hour of the day. This plot shows that there are steep increases in the cumulative distribution function when the packet size reaches 40, 576,

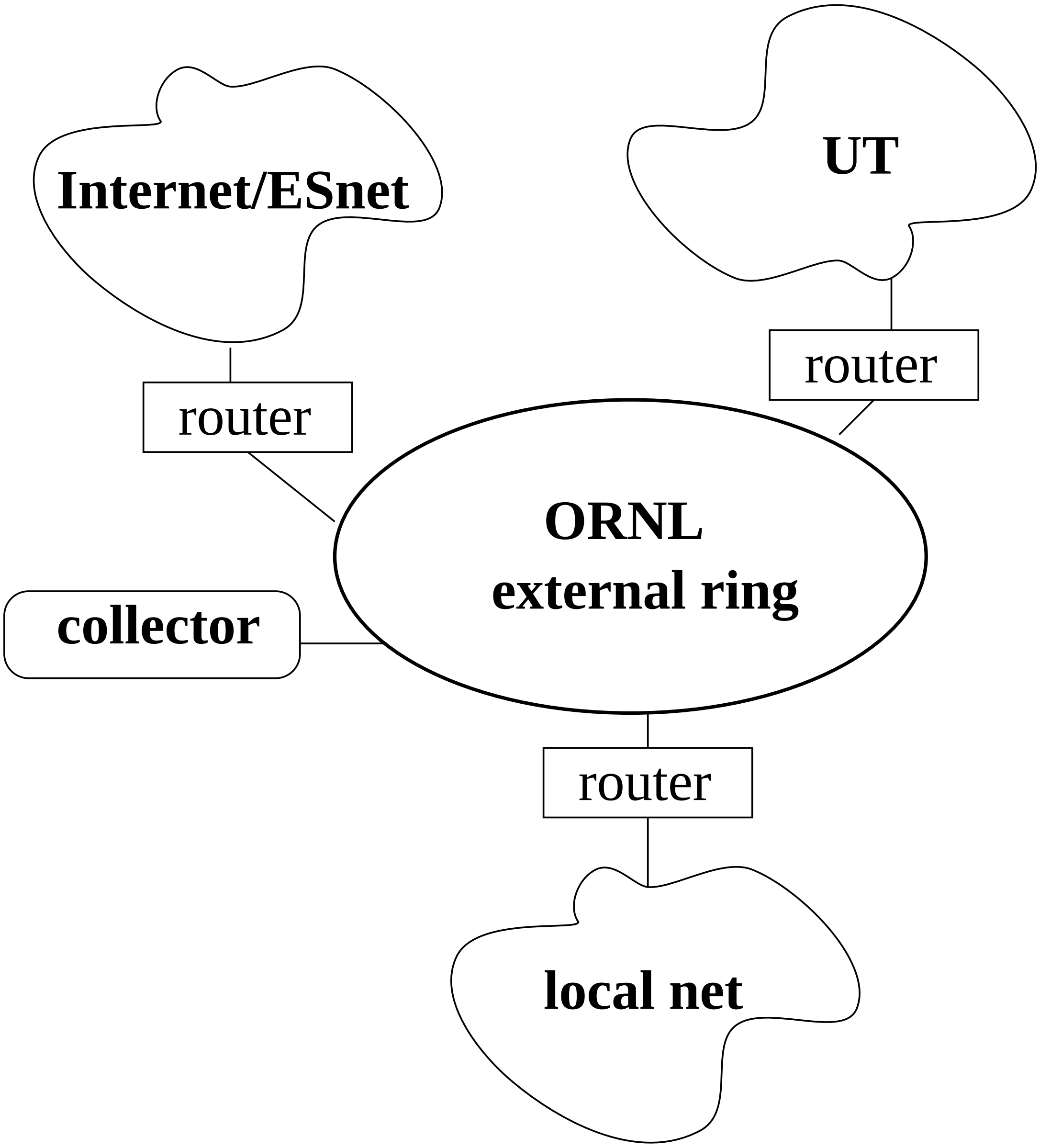

Figure 2.1: Network Configuration Used in Collecting Network Trace Data

and 1500 bytes. Also, there is a strong hourly component reflecting that peak flows occur around noon and early afternoon. Fig.2.3 is the weighted frequency histogram of packet size by hour of the day and shows the distinct peaks occurring at the 40, 576, and 1500 byte size packets. The peak at 40 bytes is from IP acknowledgment packets. Many operating systems limit the maximum size of Internet packets to 576 bytes. Newer operating systems use 1500 as the maximum IP packet size.

From these two plots we can immediately conclude that whatever background process we have on the local network, this process is inhomogeneous, i.e., its parameters are changing in time. In addition, from our knowledge about the mix of users on the local network and the applications being run, the process is also compound. Each occurrence time in the point process is associated with a vector of some markers or characteristics which takes random values in the corresponding space. For instance, one characteristic is the packet length and Fig.2.2 and 2.3 indicate that packet lengths occur with different probabilities.

3. Analysis of "averaged" histograms

Deterministic trend in time and flattening of histograms. As has been mentioned, Fig.2.2 and 2.3 show that there exists a very significant dependence of activities on the network upon time. The corresponding trend most likely can be modeled as a combination of deterministic and random components. To simplify the analysis we center attention on the packets of size 400 to 600 bytes which were examined over the day. Fig.3.1 is a time series plot of the number of packets per minute of size 400 to 600 bytes over the entire day. It is obvious from this plot that there exists a strong time-of-day dependence on the number of packets per minute. The reader can notice a few outliers. Some of them have a pure deterministic explanation: For instance, the outlier around midnight (00:09 a.m.) corresponds to one or more file copies (about 38 mbytes of data) from ORNL to the University of Tennessee (UT). The spike at approximately 9:00 p.m. was from an ftp between ORNL and Department of Physics of UT. The ORNL-UT connection is very high speed (150 megabit/second), so that data transfers with UT can achieve a very high rate. Fig.3.2 is the corresponding histogram of the number of packets per minute of size 400-600 bytes. The histogram corresponds to a distribution with a flattened central region and very heavy tail. This histogram has

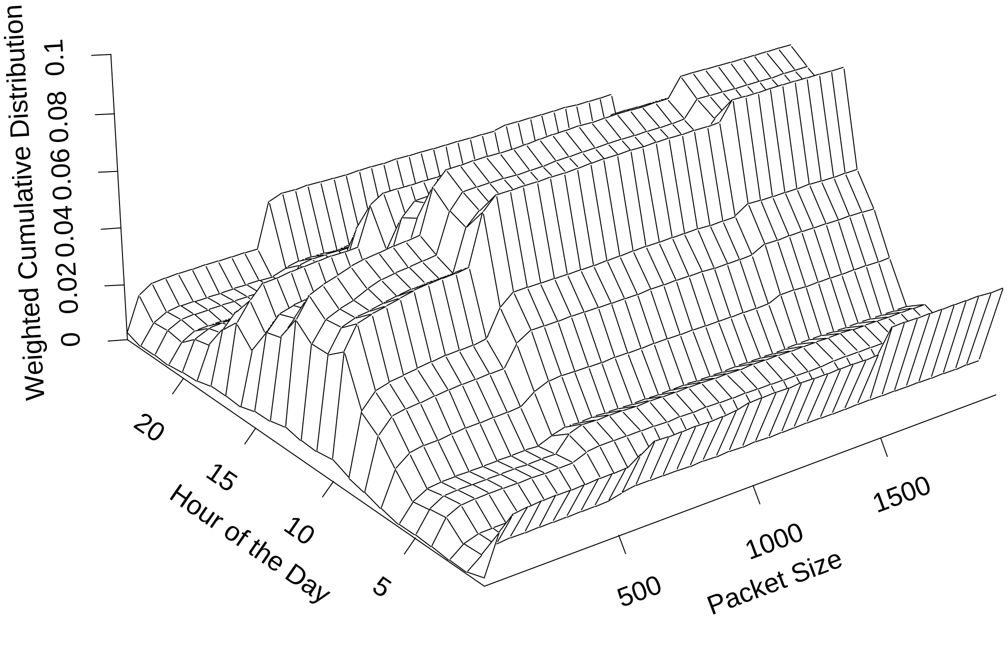

Figure 2.2: Weighted Cumulative Distributions for Each Hour of the Day

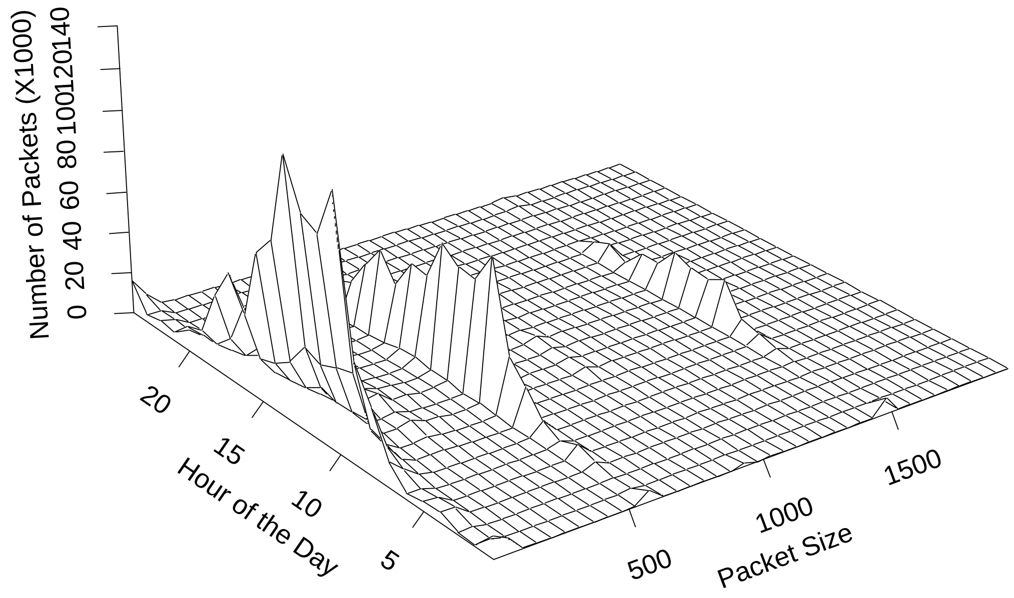

Figure 2.3: Frequency Histograms for Each Hour of the Day

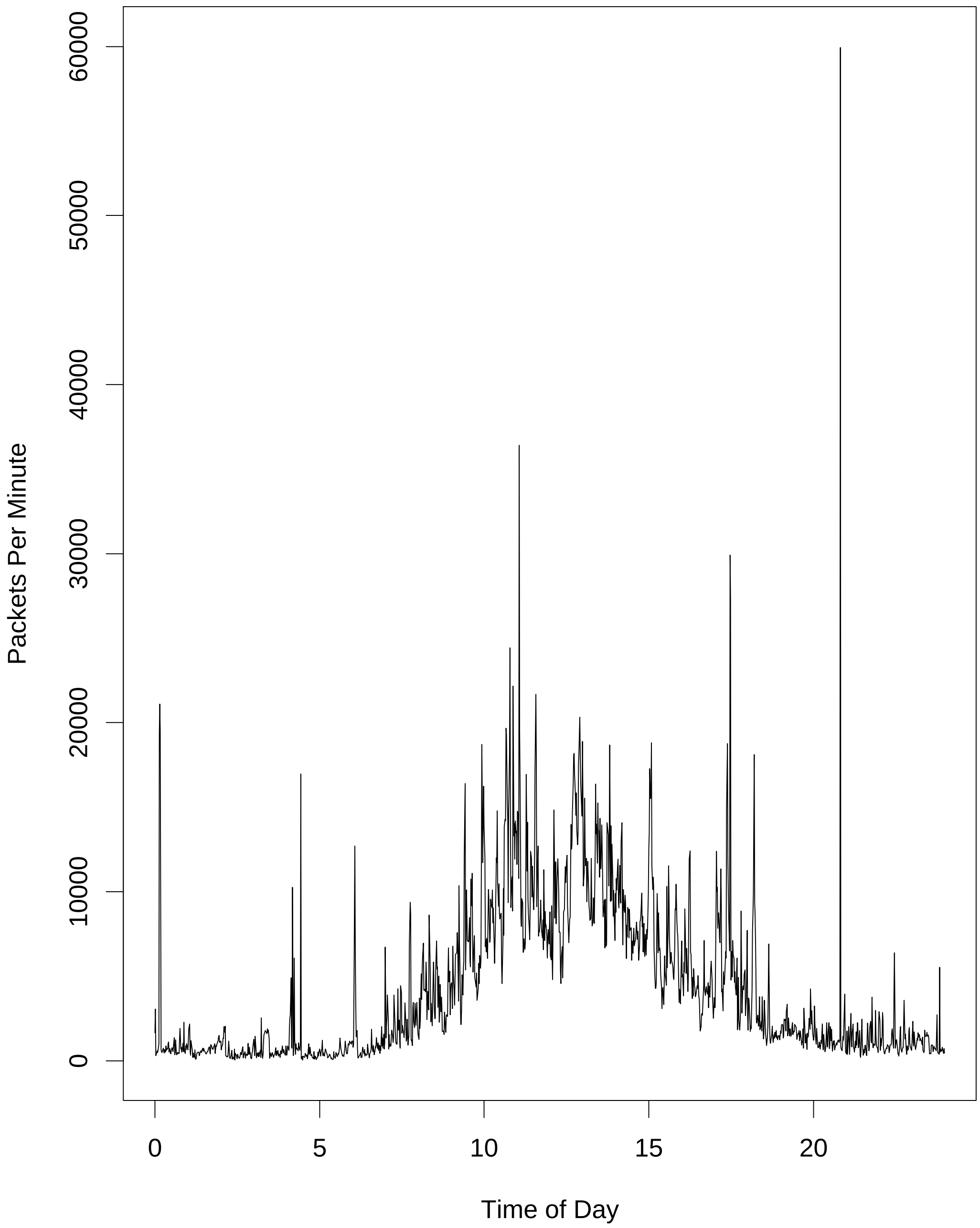

Figure 3.1: Time Series of Number of 400-600 Byte Packets Per Minute During Jan. 9, 1997

nothing in common with what one might expect for the standard point process, say the Poisson process. We show that the phenomenon is rooted in the presence of the strong time trend. Although not presented, the same was true for packets of size 40 and size 1500. From a "physical" point of view it is still natural to consider the Poisson process as a reasonable approximation of an observed process for very short intervals. Following the idea of looking at shorter intervals of time which was used by Paxson and Floyd (1995), we decided to construct histograms of counts per second for the first ten minute interval at the beginning of every hour. Fig.3.3 shows the fluctuation in number of packets per second at 9 AM, Noon, and 6 PM for 400-600 byte size packets. There are level changes within the hour (especially at noon) as well as level changes between the hours, again showing the time dependence. The smooth lines through the data are based on a moving average. Fig.3.4 plots the histograms based on the data shown in Fig.3.3. These histograms have long and heavy tails and are too flat to be generated by a Poisson process with a constant intensity function (the so called homogeneous Poisson process).

A possible explanation of that fact can be the following: let $\lambda(t)$ be the intensity function of the inhomogeneous Poisson process, then the probability to observe n packets in the interval $[\tau \pm \zeta/2]$ is (see, for instance, Snyder (1975), Chpt.2. Kovalenko et al. (1996), Chpt.7):

$$p\,(n|\tau,\zeta) = \frac{\Lambda^n(\tau,\zeta)}{n!}\, e^{-\Lambda(\tau,\zeta)}\ , \tag{3.1}$$

where

$$\Lambda(\tau,\zeta) = \int_{\tau-\zeta/2}^{\tau+\zeta/2} \lambda(t)dt\ . \tag{3.2}$$

Distribution (3.1) is the Poisson distribution with parameter $\Lambda(\tau,\zeta)$.

If we construct a histogram corresponding to distribution (3.1) using for the interval $[\tau \pm \zeta/2]$ the counts from different days, then (assuming that there is no day-wise trend) we may expect a histogram of the Poisson type. The information from contiguous intervals $[\tau_1 \pm \zeta/2], \ldots, [\tau_j \pm \zeta/2], \ldots$ will not produce the same picture if $\Lambda(\tau_j,\zeta)$ is not constant.

What kind of histograms can be generated with varying $\Lambda(\tau_j,\zeta)$? If all $\Lambda(\tau_j,\zeta)$ are

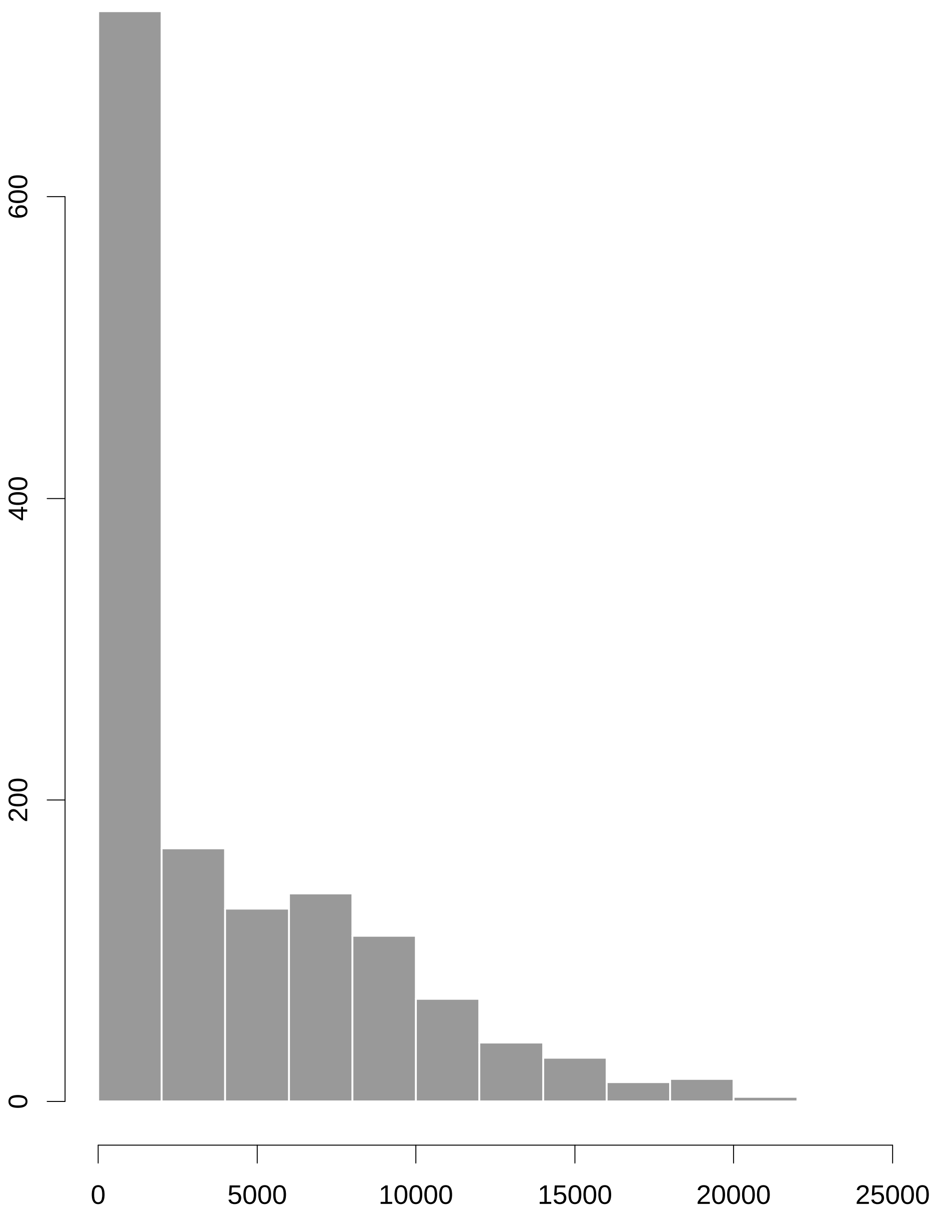

Figure 3.2: Histogram of Number of 400-600 Byte Packets Per Minute During Jan. 9, 1997

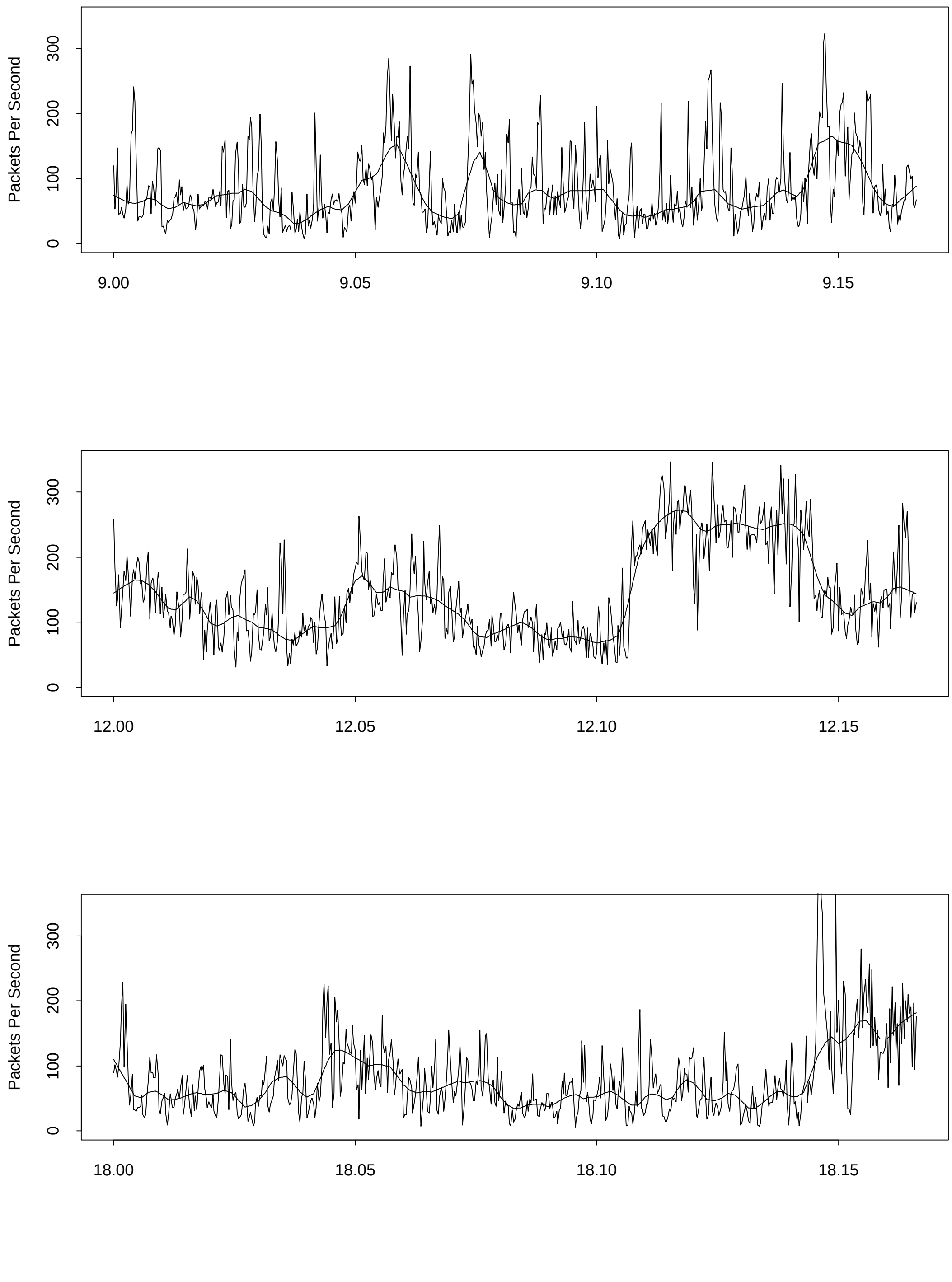

Figure 3.3: Time Series Plots of Number of 400-600 Byte Packets Per Second at 9AM, Noon, and 6PM

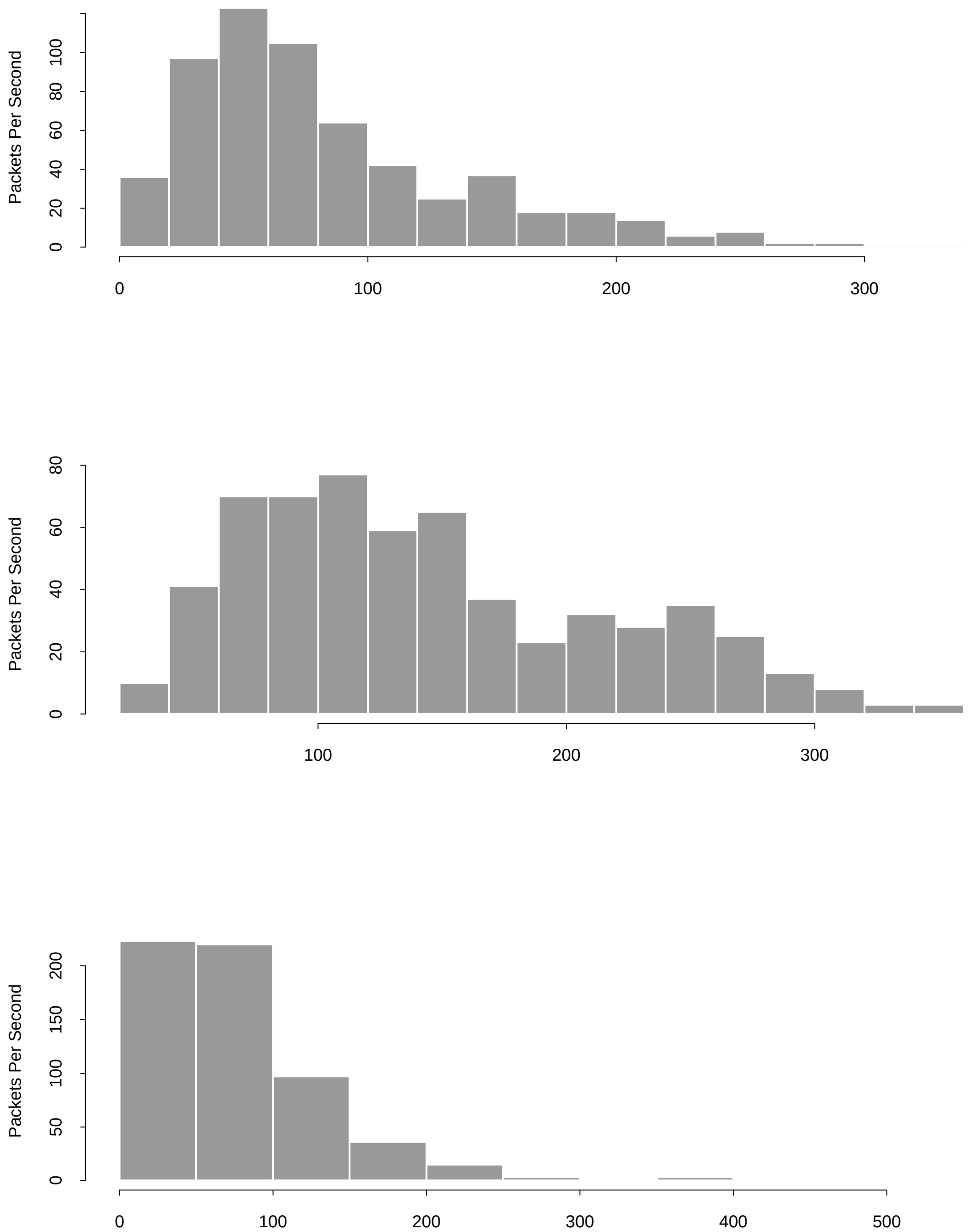

Figure 3.4: Histogram of Number of 400-600 Byte Packets Per Second at 9AM, Noon, and 6PM

known then the averaged over intervals distribution

$$p\,(n) = \frac{1}{L}\sum_{j=1}^{L} p\,(n|\tau_j,\zeta), \tag{3.3}$$

where the definition of $p\,(n|\tau_j,\zeta)$ is obvious from (3.1) and (3.2), defines the shape of a histogram.

To get a better image of the possible averaged distribution we consider the continuous analogue of (3.3)

$$p_a(n) = \frac{1}{\tau - \tau_0}\int_{\tau_0}^{\tau} \frac{\Lambda^n(t)}{n!}\, e^{-\Lambda(t)}dt, \tag{3.4}$$

where the following simplified notations were introduced: $\tau = \tau_L, \tau_0 = \tau_1$, $\Lambda(t) = \Lambda(t,\zeta)$. The subscript a indicates averaging. It must be emphasized that (3.4) is an approximation of (3.3). The precision of the approximation is $\mathrm{O}(\zeta^2)$ if $\lambda(t)$ has a continuous second derivative (c.f. Davis and Rabinowitz (1984)). This estimate was derived in the standard calculus setting, i.e., when (3.3) is used as numerical approximation of (3.4). Obviously, it stays valid for the inverse approximation.

If $\lambda(t)$ is a polynomial of degree k then $\Lambda(t)$ is a polynomial of the same degree. For instance, in the linear case when $\lambda(t) = \lambda_0 + \lambda_1 t$ we find from (3.2) that

$$\Lambda(t) = \lambda_0\zeta + \lambda_1\zeta t = \Lambda_0 + \Lambda_1 t. \tag{3.5}$$

Note that similar to (3.4) we use the shorter notation $\Lambda(t) = \Lambda(t,\zeta)$. Let $\Lambda(t)$ be defined by (3.5). Then applying (c.f. Gradsteyn and Ryzhik (1965)) the formula

$$\int e^{-bt}P_k(t)dt = -\frac{e^{-bt}}{b}\sum_{j=0}^{k} P_k^{(j)}(t)/b^j, \tag{3.6}$$

where $P_k(t)$ is a polynomial in t of degree k and $P_k^{(j)}(t)$ is its j-th derivative, we can derive that

$$p_a(n) = \frac{1}{\tau - \tau_0}\frac{F\,[n,\Lambda(\tau_0)] - F\,[n,\Lambda(\tau)]}{\Lambda_1}, \tag{3.7}$$

where $F(n,\Lambda)$ is the cumulative distribution function of the Poisson distribution. Of course, if the process is homogeneous then $\lambda_1 = \Lambda_1 = 0$ and

$$p_a\,(n) = \frac{1}{\tau - \tau_0}\lim_{\Lambda_1 \to 0}\frac{F(n, \Lambda(\tau_0)) - F(n, \Lambda(\tau))}{\Lambda_1} = \frac{\Lambda_0^n e^{-\Lambda_0}}{n!}\,. \tag{3.8}$$

Using (3.4) it is possible to find the average distribution for some other polynomials, for instance, for the second-order polynomial. However, the corresponding formulae are not as simple as (3.7). We think that in most cases piece-wise linear approximations of the intensity function $\Lambda(t)$ will be adequate, leaving the corresponding calculus still very affordable. For instance, let

$$\Lambda(t) = \begin{cases} \Lambda_0, & \tau_0 \le t < \tau_1\,, \\ \Lambda_0 + \Lambda_1(t - \tau_1), & \tau_1 \le t \le \tau_2 \end{cases}\,.$$

Then

$$p_a\,(n) = \frac{1}{\tau_2 - \tau_0}\left[(\tau_1 - \tau_0)\,\frac{\Lambda_0^n e^{-\Lambda_0}}{n!} + \frac{F(n, \Lambda(\tau_1)) - F(n, \Lambda(\tau_2))}{\Lambda_1}\right]. \tag{3.9}$$

Fig.3.5 shows the shape of the averaged distribution function, $p_a(n)$ for three special cases of $\Lambda(t)$: (a, a') $\Lambda(t) = t$ for $20 \le t \le 150$; (b, b') $\Lambda(t) = 50$ for $20 \le t < 100$; $\Lambda(t) = 2t - 150$ for $100 \le t \le 150$ and (c, c') $\Lambda(t) = t$ for $20 \le t \le 100$; $\Lambda(t) = 100$ for $100 \le t \le 150$. These distributions are much closer to what was observed (see Figure 3.4) than the standard Poisson distribution. Note that if in any of the considered examples the segment $[\tau_0, \tau_2]$ is partitioned in any given number of segments $[\tau_0, \tau_{(1)},], [\tau_{(1)}.\tau_{(2)}], \ldots, [\tau_{(k)}, \tau_2]$ and these segments are permuted together with the corresponding $\Lambda(t)$, then the function $p_a(n)$ does not change.

Curiously enough (3.7) and (3.9) may help to illuminate the concept of self-similarity, which is a popular subject of some recent publications on computer networks (see, for instance, Willinger etc. (1995), Gustafsson and Karlsson (1997)). Indeed, the interval length ζ (see (3.5)) may be considered as a scaling parameter which defines the function $p_a(n)$ given λ_0 and λ_1, changing $\Lambda_0 = \zeta\lambda_0$ and $\Lambda_1 = \zeta\lambda_1$. The variation of the two latter parameters changes positioning and spreadness of $p_a(n)$ leaving the shape (say, table-shape in case (3.7)) almost identical for different ζ. The tendency is better seen if, for instance, in (3.7) the ratio $\zeta/(\tau - \tau_0)$ is kept constant.

Comparison of deterministic trend with stochastic change of intensity. Another approximation of the averaged distribution $p_a(n)$ (or another form of (3.4)) can be

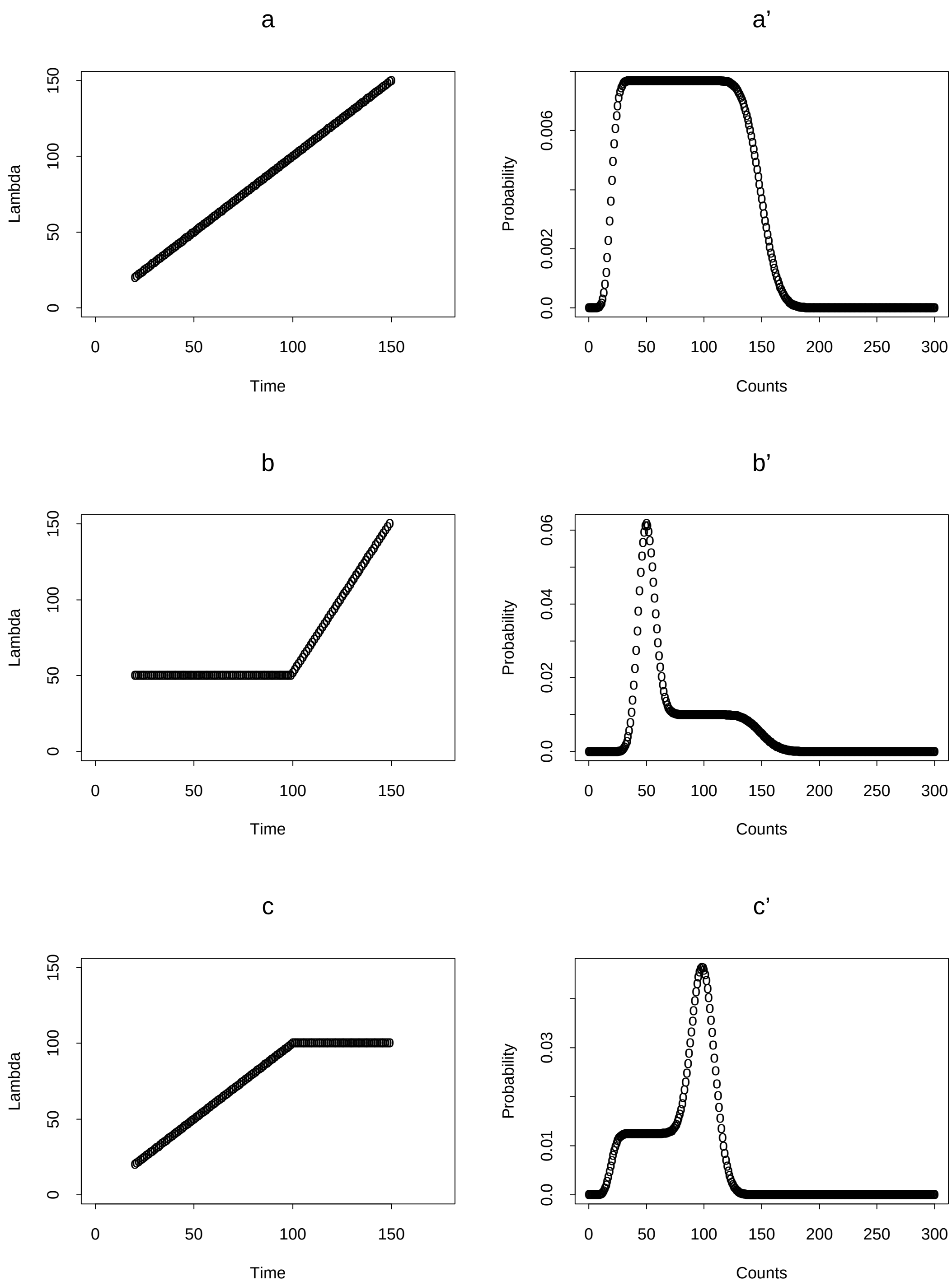

Figure 3.5: Examples of Averaged Distribution for Heterogeneous Poisson

obtained by a change of variable $\Lambda = \Lambda(t)$ in integral (3.4) which yields

$$p_a(n) = \int_{\Lambda(\tau_0)}^{\Lambda(\tau)} \frac{\Lambda^n}{n!} e^{-\Lambda} f(\Lambda)\, d\,\Lambda, \tag{3.10}$$

where

$$f(\Lambda) = \frac{1}{\tau - \tau_0} \frac{dt(\Lambda)}{d\Lambda}$$

and $t(\Lambda)$ is the inverse function of $\Lambda(t)$. We find that the average distribution (3.10) has the same form as the distribution of counts for the doubly stochastic Poisson process (see Snyder (1975), Chpt.6.2) in which the intensity Λ is random and has the probability density function $f(\Lambda)$. Intuitively (either in the deterministic or stochastic setting) the value $f(\Lambda)d\Lambda$ is a fraction of the total time $\tau - \tau_0$ when $\Lambda - d\Lambda/2 \leq \Lambda(t) \leq \Lambda + d\Lambda/2$.

A popular choice for $f(\Lambda)$ is the gamma density with parameters $\alpha > 0$ and $\beta > 0$ given by

$$f(\Lambda) = \begin{cases} \frac{(\Lambda/\alpha)^{\beta-1} e^{-\Lambda/\alpha}}{\alpha\Gamma(\beta)}, & \Lambda \geq 0, \\ 0, & \Lambda < 0. \end{cases} \tag{3.11}$$

Note that in the stochastic setting $\alpha\beta$ is the expected value of Λ (see, for instance, Johnson, et al. (1994)) when Λ is treated as a random variable. In the deterministic setting as above, it is simply a weighted arithmetic mean of Λ.

Evaluation of (3.10) with the intensity function defined by (3.11) where $\Lambda(\tau_0) = 0$ and $\Lambda(\tau) \to \infty$ results in

$$p_a(n) = \frac{\Gamma(n+\beta)}{\Gamma(n+1)\Gamma(\beta)} \left(\frac{1}{1+\alpha}\right)^{\beta} \left(\frac{\alpha}{1+\alpha}\right)^{n}, \tag{3.12}$$

which is the negative binomial distribution. The distribution $p_a(n)$ with $\alpha = 100$ and $\beta = 1$ is presented in Fig.3.6.

Assuming that $\alpha = \lambda_0\zeta$ (compare with (3.5)), one can analyze what kind of changes in histogram shapes may be observed with changing the scaling parameter ζ.

Similar to (3.6) and (3.7), we can show that under the assumption of continuity the function $\Lambda(t)$ corresponding to (3.11) may be found as a solution of the following equation

$$t = (\tau - \tau_0)\left[1 - F(\beta - 1, \Lambda/\alpha)\right], \tag{3.13}$$

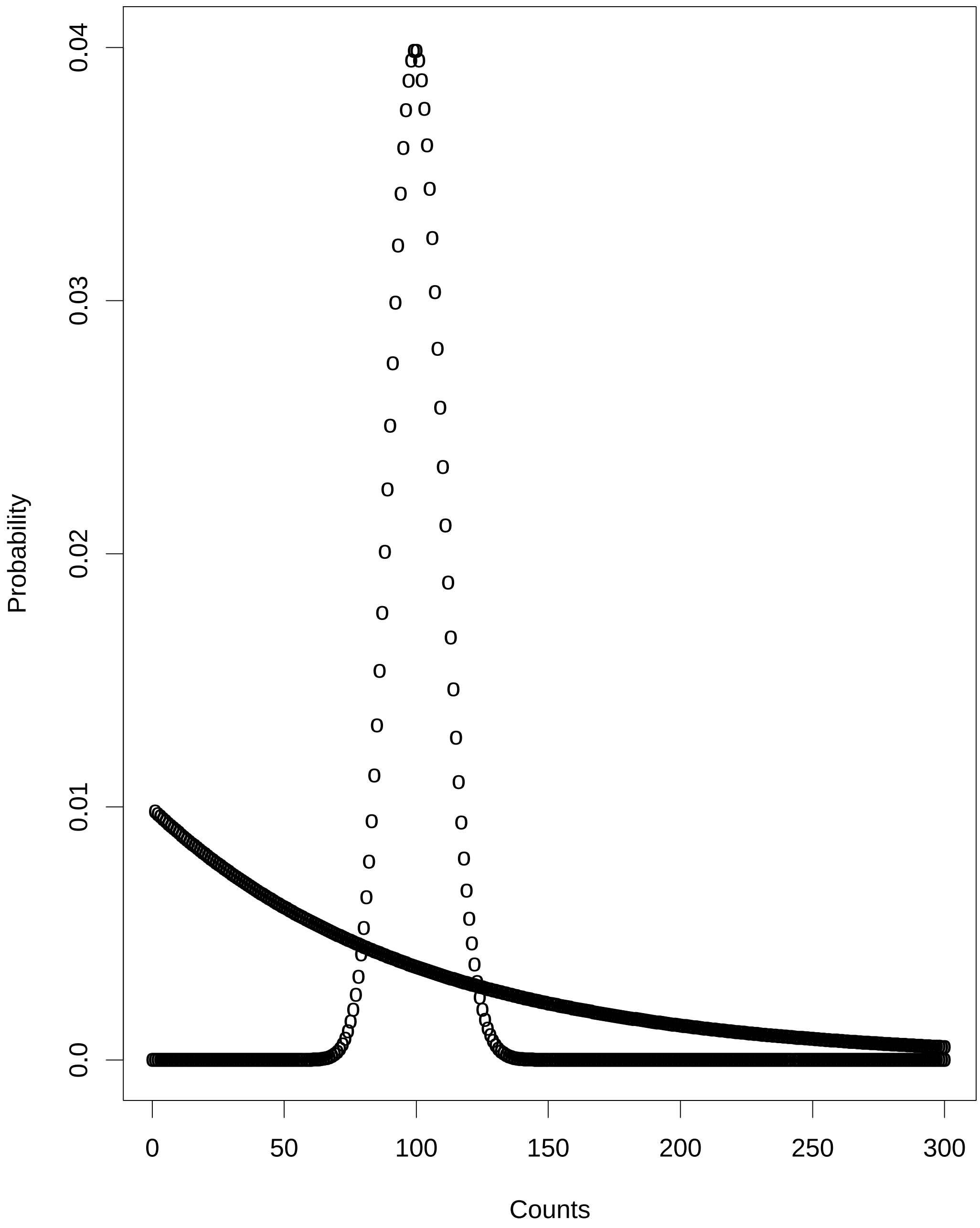

Figure 3.6: Negative Binomial with $\alpha = 100$ and $\beta = 1$ and Poisson Distribution with Mean = 100

where the function F is defined in comments to (3.7). In particular, when $\beta = 1$, then $f(\Lambda) = \frac{1}{\alpha} e^{-\Lambda/\alpha}$ and the continuous solution for $\Lambda(t)$ is

$$\Lambda(t) = -\, \alpha \, \ln\,(1 - \frac{t}{\tau});\; 0 \le t < \tau.$$

see also Fig.3.7.

Interestingly, for the linear trend

$$f(\Lambda) = \frac{1}{\Lambda_1(\tau - \tau_0)} = \frac{1}{\Lambda(\tau) - \Lambda(\tau_0)}, \tag{3.14}$$

i.e., averaged distribution (3.7) also corresponds to the case in which Λ is uniformly distributed on $[\Lambda(\tau_0), \Lambda(\tau)]$.

Thus, assuming either that Λ is random or deterministically changing in time, we come to the flattened and heavy tailed averaged distributions.

We may conclude that exactly the same histograms exist for completely different models describing variations of the intensity. Thus, the histograms being one of the simplest ways to understand the background random processes are not sufficient for a unique determination of the intensity behavior unless some additional assumptions are made.

4. Estimation of the intensity function

As soon as the preliminary analysis reveals flattened histograms, we have to develop a model which may describe evolution of the intensity λ in time. As mentioned before, histograms do not lead to a unique choice of that evolution in time. Flattening of histograms emphasizes variability of λ but does not explain the cause. To decide whether it is a smooth trend in time or a random variation or combination of both, we have to use additional information, for instance, to examine plots similar to Fig.3.1 or 3.3; see also Section 5.

When the choice is made, then the estimation problem can be considered with the help of well established statistical techniques after some modest modification. Let us start with the deterministic trend in time.

Maximum likelihood estimators and reweighted least squares method. Let the number of packets n_i be known for interval $[\tau_1 \pm \zeta/2], \ldots,$ $[\tau_L \pm \zeta/2]$. We selected equal

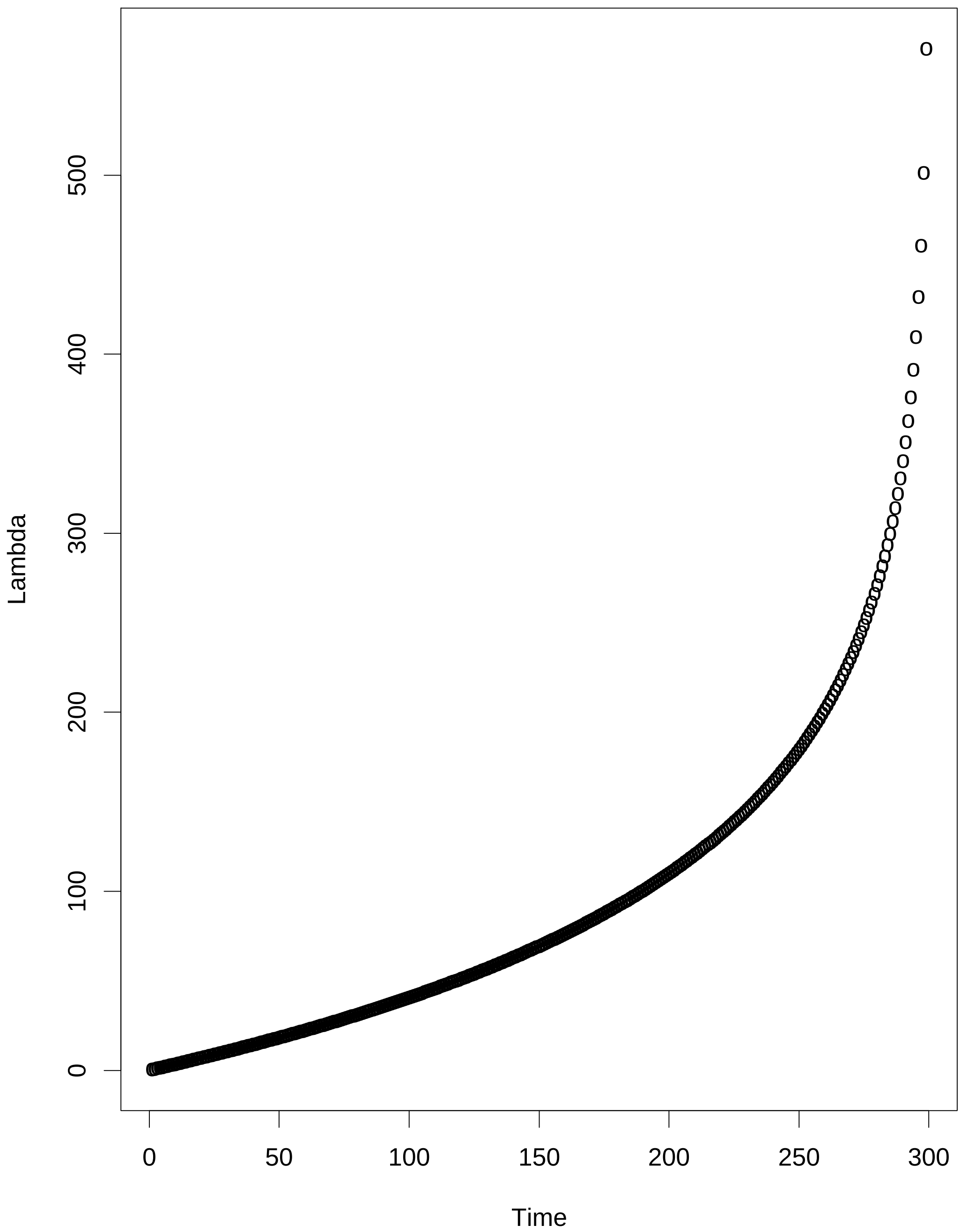

Figure 3.7: Plot of $\Lambda(t)$ when $\alpha = 100$ and $\beta = 1$

intervals only to simplify notations. Generalization to the case of intervals of different length will be obvious. It is known that for a inhomogeneous Poisson process (see, for instance, Snyder (1975), Chpt.2) the logarithm of the likelihood function defined by

$$\begin{aligned} \ln \mathcal{L}(\underline{\lambda}; n_1, \ldots, n_L) &= C(n_1, \ldots, n_L) \\ -\sum_{i=1}^{L} \int_{\tau_i - \zeta/2}^{\tau_i + \zeta/2} \lambda(\underline{\lambda}, t) dt &+ \sum_{i=1}^{L} n_i \ln \int_{\tau_i - \zeta/2}^{\tau_i + \zeta/2} \lambda(\underline{\lambda}, t) dt, \end{aligned} \tag{4.1}$$

where C is a function of $n_1, \ldots, n_L$, which does not depend upon unknown parameters $\underline{\lambda}$.

Assuming that

$$\lambda(\underline{\lambda}, t) = \underline{\lambda}^T \varphi(t) \tag{4.2}$$

and differentiating (4.1) we derive that

$$\sum_{i=1}^{L} n_i \frac{\phi(\tau_i)}{\underline{\lambda}^T \phi(\tau_i)} - \sum_{i=1}^{L} \phi(\tau_i) = \sum_{i=1}^{L} \frac{n_i \phi(\tau_i) - \phi(\tau_i) \phi^T(\tau_i) \underline{\lambda}}{\underline{\lambda}^T \phi(\tau_i)} = 0, \tag{4.3}$$

where

$$\phi(\tau) = \int_{\tau - \zeta/2}^{\tau + \zeta/2} \varphi(t) dt.$$

There exist a number of methods and recommendations either on direct minimization of (4.1) or on solving the system of equations (4.3). We recommend the iterative procedure which is based on the following presentation of (4.3):

$$\sum_{i=1}^{L} \omega_i(\underline{\lambda}) n_i \phi(\tau_i) - \sum_{i=1}^{L} \omega_i(\underline{\lambda}) \phi(\tau_i) \phi^T(\tau_i) \underline{\lambda} = 0, \tag{4.4}$$

where $\omega_i^{-1}(\lambda) = \underline{\lambda}^T \phi(\tau_i)$. Introducing

$$M(\underline{\lambda}) = \sum_{i=1}^{L} \omega_i(\underline{\lambda}) \phi(\tau_i) \phi^T(\tau_i)$$

and

$$Y(\underline{\lambda}) = \sum_{i=1}^{L} \omega_i(\underline{\lambda}) n_i \phi(\tau_i)$$

we have that

$$Y(\underline{\lambda}) - M(\underline{\lambda})\underline{\lambda} = 0 \tag{4.5}$$

and consequently the iterative procedure can be formulated as

$$\hat{\underline{\lambda}} = \lim_{s\to\infty} \underline{\lambda}_s, \tag{4.6}$$

$$\underline{\lambda}_s = M^{-1}(\underline{\lambda}_{s-1})Y(\underline{\lambda}_{s-1}). \tag{4.7}$$

Formula (4.7) coincides with the formula for the iterated least square estimates:

$$\underline{\lambda}_s = \arg\min_{\underline{\lambda}} \sum_{i=1}^{L} \omega_i(\underline{\lambda}_{s-1}) \left[n_i - \underline{\lambda}^T \phi(\tau_i)\right]^2. \tag{4.8}$$

It must be noted that in general

$$\hat{\underline{\lambda}} \neq \arg\min_{\underline{\lambda}} \sum_{i=1}^{L} \omega_i(\underline{\lambda}) \left[n_i - \underline{\lambda}^T \phi(\tau_i)\right]^2. \tag{4.9}$$

In most cases the iterative procedure (4.6), (4.7) converges to

$$\underline{\lambda}^* = \arg\min_{\underline{\lambda}} \mathcal{L}(\underline{\lambda}; n_1, \ldots, n_L), \tag{4.10}$$

but sometimes it does not. To justify this statement we have to recollect that $\{n_i\}_1^L$ are realizations of random variables. That is why, in general, we may talk about the probability $P(L)$ that the sequence $\{\underline{\lambda}_s\}$ converges to $\underline{\lambda}^*$, i.e. about the probability of the event that the limit $\lim_{s\to\infty} \underline{\lambda}_s = \hat{\underline{\lambda}}$ exists, and $\hat{\underline{\lambda}} = \underline{\lambda}^*$.

If ζ is fixed then the following statement holds:

$$\lim_{L\to\infty} P(L) = 1. \tag{4.11}$$

Appendix 1 contains the corresponding proof, which is based on the ideas used by Fedorov (1974), see also Harville (1977).

Asymptotically $\hat{\lambda}$ are normally distributed and $M^{-1}(\hat{\underline{\lambda}})$ coincides with the dispersion matrix of $\hat{\underline{\lambda}}$ (for more details see Appendix 1), i.e. for moderate $\sum_{i=1}^{L} n_i$ we have that

$$Var(\hat{\underline{\lambda}}) \simeq M^{-1}(\hat{\underline{\lambda}}). \tag{4.12}$$

In general, the use of statistics $n_1, \ldots, n_L$ leads to loss of information, i.e. these statistics are not sufficient, except for some trivial cases such as $\lambda(t) \equiv \lambda$. One of the ways to gain more from the data is reduction of the interval length ζ. Because the observed process is Poisson, we can select ζ such that in each interval the number of counts (transmitted packets) is either 1 or 0, and $\phi(\tau) \simeq \varphi(\tau)\zeta$ (see Andersen et al. (1992), Chpt.2.b and Snyder (1975), Chpt.2.4 for more rigorous mathematics). In this case (4.5) must be replaced by the equation

$$Y_c(\underline{\lambda}) - M_c(\underline{\lambda})\underline{\lambda} = 0, \tag{4.13}$$

where

$$\begin{aligned} M_c(\underline{\lambda}) &= \int_{\tau_1}^{\tau_L} \omega(\underline{\lambda}, t)\varphi(t)\varphi^T(t)dt, \\ Y_c(\underline{\lambda}) &= \sum_{j=1}^{N} \omega(\underline{\lambda}, t_j)\varphi(t_j), \end{aligned}$$

$\omega^{-1}(\underline{\lambda}, t) = \underline{\lambda}^T\varphi(t)$ and N is the total number of counts, $t_1, \ldots, t_N$ are the instants when the observed events have occurred, i.e. packets were registered. Changes which must be done in (4.7) and (4.8) to match (4.13) are evident. Estimators defined by (4.5) and (4.13) are significantly different only when N and L are relatively small.

Piece-wise estimation of trend and construction of averaged histogram. Combining the results of Section 3 and the iterated estimator (4.6), we may develop a simple method which gives us the opportunity to verify the validity of various assumptions about the structure of the analyzed process with the observed histograms. As it has been noted the histograms constructed for various time-scales look very similar for the considered data (compare Figure 3.2 and Fig.3.4). We have selected the more aggregated case (number of packets per minute) as a test example only because it gives a better opportunity to trace various peculiarities of the proposed approach. For instance, outliers are visually identifiable, the trend is well pronounced, and changes have easy intuitive explanations.

To keep the computations simple we partitioned 24 hour intervals in 15 minute intervals and used piece-wise linear approximation (compare with (3.5)) at each of them. We did not demand continuity of approximation at the boundary points for two reasons. First, it makes the computations more precision demanding (inversion of large ill-conditioned matrices is needed) even if we use some theoretically simple approach, for instance, based on

$$\begin{aligned}
\phi_1(\tau) &\equiv 1, \\
\phi_2(\tau) &= \tau, \\
\phi_j(\tau) &= \begin{cases} \tau - \tau_{j-2}, & \tau \geq \tau_{j-2} \\ 0, & \text{otherwise}, \end{cases} \\
j &= 3, \ldots, L+1.
\end{aligned} \tag{4.14}$$

Second, it is not obvious at all that the trend must be continuous. Note that the use of (4.14) makes the approximation model more parsimonious (it contains only $L+1$ unknown parameters instead of $2L$ as in the piece-wise case with possible jumps at the boundary points of intervals), but consequently less flexible.

The selection of coefficients for the first 24 intervals is presented at Table 1. From this table we may conclude that the iterations converge very rapidly and for practical needs 5 iterations are adequate. Initial weights were identically equal 1. Note that this choice guarantees unbiasedness of the first stage estimator $\underset{\sim}{\lambda}_1$. According to (3.7) we constructed

$$\hat{p}_a(n) = \frac{1}{L} \sum_{i=1}^{L} \frac{F\left(n, \hat{\Lambda}(\tau_{i-1})\right) - F\left(n, \hat{\Lambda}(\tau_i)\right)}{\hat{\Lambda}(\tau_i) - \hat{\Lambda}(\tau_{i-1})}$$

note that $\hat{\Lambda}(\tau_i) - \hat{\Lambda}(\tau_{i-1}) = (\tau_i - \tau_{i-1})\hat{\Lambda}_{1i}$ and $\hat{\Lambda}_{1i} = \hat{\lambda}_{1i}\zeta$, and in the considered case $\zeta =$ 1min. The function $\hat{p}_a(n)$ is plotted at Fig.4.1 together with the observed histogram, which is a finer version of the histogram presented at Fig.3.2. In general $\hat{p}_a(n)$ provides a very good explanation of the observed histogram confirming our assumption that the observed process may be approximated by inhomogeneous Poisson process. There are some obvious steps for the improvement. For instance, we may remove outliers, or introduce a mixed model containing deterministic trend with random parameters, or make partitioning more flexible and link boundary points with some daily events like

Table 4.1: Comparison of the Estimates for Different Number of Iteration

5 Iterations		**10 Iterations**		**20 Iterations**		**50 Iterations**	
Intercept	Slope	Intercept	Slope	Intercept	Slope	Intercept	Slope
4680.14	339.83	4679.90	340.01	4679.43	340.36	4678.04	341.40
748.36	-87.29	748.47	-87.24	748.68	-87.14	749.19	-86.88
925.12	-59.09	925.03	-59.17	924.84	-59.33	924.33	-59.79
686.34	6.08	686.34	6.05	686.35	6.01	686.36	5.89
2280.06	278.88	2279.97	278.92	2279.80	278.99	2279.33	279.21
515.20	-10.20	515.20	-10.20	515.20	-10.20	515.19	-10.21
2322.26	-281.66	2321.77	-281.88	2320.80	-282.32	2318.04	-283.56
1998.05	-203.34	1998.45	-203.13	1999.22	-202.72	2001.35	-201.55
4006.62	251.20	4006.56	251.25	4006.44	251.36	4006.09	251.66
5594.21	241.72	5594.22	241.70	5594.24	241.68	5594.30	241.60
8637.13	-89.91	8637.14	-89.90	8637.14	-89.89	8637.15	-89.85
12184.44	-874.89	12184.40	-874.91	12184.33	-874.97	12184.13	-875.12
9249.86	307.30	9249.86	307.30	9249.86	307.31	9249.84	307.34
11065.00	-390.88	11065.00	-390.88	11065.00	-390.89	11064.99	-390.90
10337.07	-24.84	10337.07	-24.84	10337.07	-24.84	10337.07	-24.85
10714.02	-938.83	10713.95	-938.87	10713.82	-938.95	10713.44	-939.18
6361.21	211.20	6361.22	211.18	6361.24	211.14	6361.31	211.03
7715.93	-12.75	7715.93	-12.75	7715.93	-12.76	7715.93	-12.77
5751.29	774.51	5753.57	773.60	5757.88	771.87	5769.26	767.25
2068.13	36.69	2068.13	36.69	2068.13	36.70	2068.12	36.72
1332.29	-80.92	1332.31	-80.90	1332.36	-80.86	1332.49	-80.73
1223.46	26.70	1223.45	26.72	1223.44	26.75	1223.40	26.85
1273.14	-118.95	1273.23	-118.90	1273.39	-118.80	1273.86	-118.52
1237.67	4.40	1237.67	4.40	1237.67	4.40	1237.67	4.38

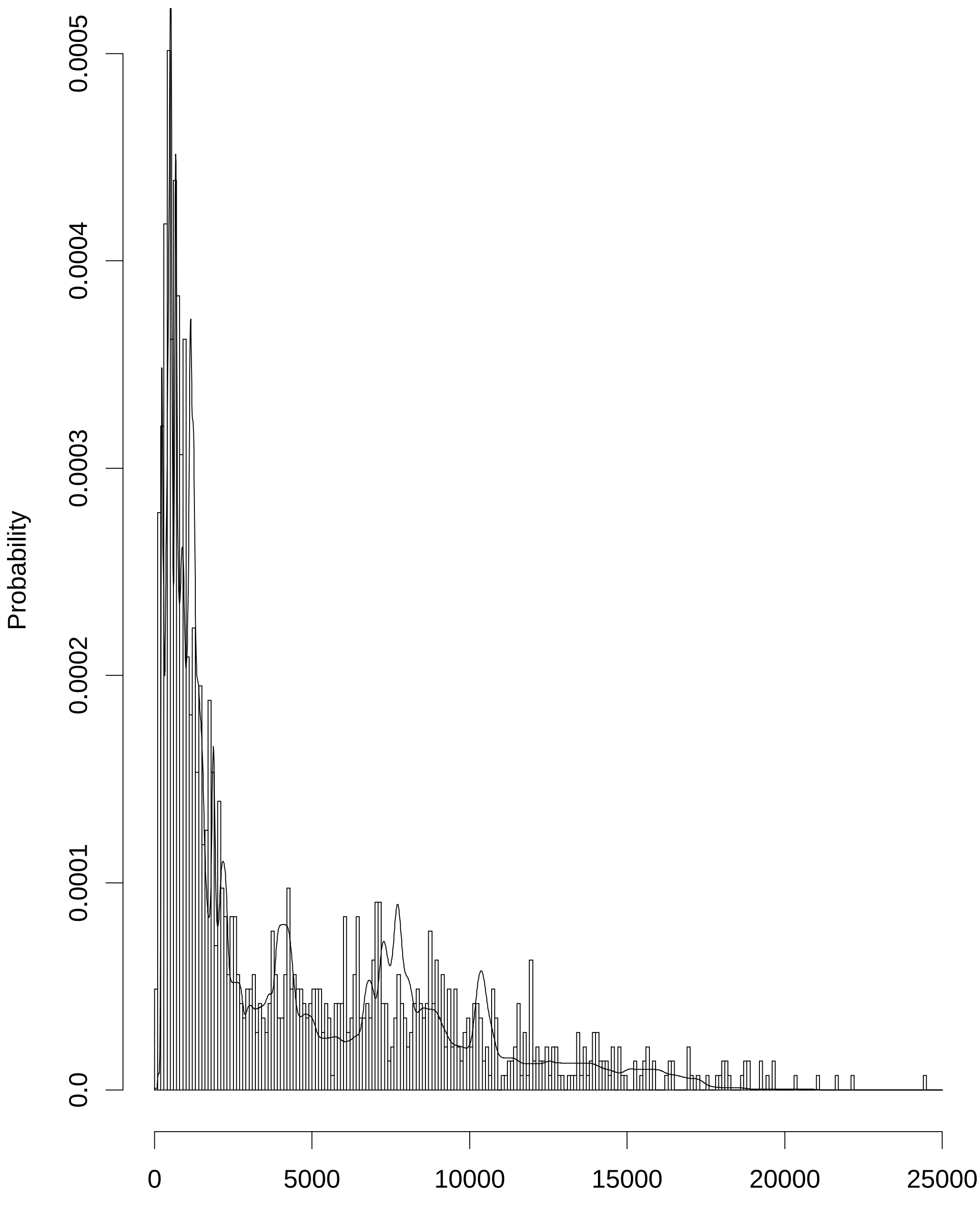

Figure 4.1: Average Distribution and Histogram for 400-600 Byte Packets Per Minute

lunch breaks or working hours.

5. Sample covariance function.

Covariance and trend. As it was noted in Section 3 the similar or even identical averaged histograms may appear for various background models. For instance, the averaged histogram exhibited at Fig.4.1 can be equally well modeled with a mixture of distributions of the intensity parameter λ. To minimize the set of computing models we have to use together with a histogram some other statistics. In particular, the sample covariance function (or more specifically, autocovariance/correlation function) allows one to determine whether the “flattening” of a histogram is caused by trend or by double stochasticity of the considered process.

To show this we will apply the approach similar to that used in Section 3.

Let L and k be given and

$$\hat{\rho}(k) = \frac{1}{L-k} \sum_{i=1}^{L-k} (n_i - \bar{n})(n_{i+k} - \bar{n}), i = 1, \ldots, \quad L, k = 0, 1, \ldots, L-1. \tag{5.1}$$

Other choices of a sample covariance function (or estimators of a covariance function) are possible (c.f. Anderson (1994), Chpt.8) and all of them result in similar inferences. We stay with (5.1) to avoid more extensive calculations.

In Fig.5.1 we plotted the function $\hat{\rho}(k)|\hat{\rho}(0)$ for the entire day; n_i is a number of packets per minute of size 400 to 600 bytes. Fig.5.2 contains the same functions for the 15 minute intervals at 9 a.m., noon and 6 p.m.; n_i is a number of the same packets per second. If the counts $n_1, \ldots, n_s$ are generated by a Poisson process either homogeneous or inhomogeneous, then (see Appendix 2):

$$\overline{\rho}(k) = E\left[\hat{\rho}(k)\right] = \frac{1}{L-k} \sum_{i=1}^{L-k} \Lambda_i \Lambda_{i+k} - \left(\overline{\Lambda} + L^{-1}\right)\left(\overline{\Lambda}_k + {}_k\overline{\Lambda} - \overline{\Lambda}\right) + \delta_{0k}\overline{\Lambda}, \tag{5.2}$$

where $\Lambda_i = \Lambda(t_i, \zeta)$, see (3.2), δ_{0k} is the Kronecker symbol and

$$\overline{\Lambda} \;=\; \frac{1}{L}\sum_{i=1}^{L} \Lambda_i,$$

$$
\begin{aligned}
\overline{\Lambda}_k &= \frac{1}{L-k}\sum_{i=k+1}^{L}\Lambda_i,\\
{}_k\overline{\Lambda} &= \frac{1}{L-k}\sum_{i=1}^{L-k}\Lambda_i.
\end{aligned}
\tag{5.3}
$$

In the homogeneous case when $\Lambda_i \equiv \Lambda$, and $\overline{\rho}(k) \equiv \Lambda(\delta_{0k} + L^{-1})$.

Similar to Section 4, let $\Lambda_i = \underline{\lambda}^T\phi(t_i) = \underline{\lambda}^T\phi_i$ where the vector function $\phi(t)$ is defined in comments to (4.3). Then the function $\overline{p}(k)$ may be presented as

$$
\begin{aligned}
\overline{\rho}(k) = \underline{\lambda}^T\Psi_k\underline{\lambda} \quad &- \quad \underline{\lambda}^T\Phi\left(\Phi_k + {}_k\Phi - \Phi\right)\underline{\lambda}\\
&- \quad L^{-1}\underline{\lambda}^T\left(\Phi_k + {}_k\Phi - \Phi\right) + \delta_{0k}\underline{\lambda}^T\Phi,
\end{aligned}
\tag{5.4}
$$

where

$$
\begin{aligned}
\Psi_k &= \tfrac{1}{L-k}\textstyle\sum_{i=1}^{L-k}\phi_i\phi_{i+k}^T, &\Phi_k &= \tfrac{1}{L-k}\textstyle\sum_{i=k+1}^{L}\phi_i,\\
{}_k\Phi &= \tfrac{1}{L-k}\textstyle\sum_{i=1}^{L-k}\phi_i, &\Phi &= \tfrac{1}{L}\textstyle\sum_{i=1}^{L}\phi_i.
\end{aligned}
$$

Note that for small ζ all terms in (5.4) are of order ζ^2 except the last one which is of order ζ. Thus, at $k = 0$ there is a "discontinuity" of the sample covariance function.

The averaged sample covariance function (5.4) exhibits long range dependence even for rapidly oscillating functions $\phi(t_i) = \phi_i$. To show this, let us, similar to Section 3, use the continuous approximation of $\underline{\rho}(k)$ on $[0, \tau_L]$:

$$
\overline{\rho}(t) = \zeta^2 c(t) + \delta_{0t}\underline{\lambda}^T\tilde{\Phi},
\tag{5.5}
$$

where

$$
\begin{aligned}
c(t) = \underline{\lambda}^T &\left[\tilde{\Psi}(t) - \tilde{\Phi}\left(\tilde{\Phi}_+(t) + \tilde{\Phi}_-(t) - \tilde{\Phi}\right)\right]\lambda\\
&+ O(\zeta^2\underline{\lambda}^T\underline{\lambda}) + O\left(\frac{\zeta\sqrt{\underline{\lambda}^T\underline{\lambda}}}{\tau_L}\right),\\
\tilde{\Psi}(t) &= \tfrac{1}{\tau_L - t}\textstyle\int_0^{\tau_L - t}\varphi(\tau)\varphi(\tau + t)d\tau,\ \tilde{\Phi} = \tfrac{1}{\tau_L}\textstyle\int_0^{\tau_L}\varphi(\tau)d\tau,\\
\tilde{\Phi}_+(t) &= \tfrac{1}{\tau_L - t}\textstyle\int_t^{\tau_L}\varphi(\tau)d\tau,\ \tilde{\Phi}_-(t) = \tfrac{1}{\tau_L - t}\textstyle\int_0^{\tau_L - t}\varphi(\tau)d\tau,
\end{aligned}
$$

functions $\varphi(t)$ are defined in (4.2), $0 \le t \le \tau_L, \delta_{0t} = 1$, if $t = 0$, and $\delta_{0t} = 0$ otherwise.

Typical trends. Let $\varphi(t) = \sin 2\pi t$. Then for $\tau_L \to \infty$

$$
c(t) \to \frac{\lambda^2}{2}\cos 2\pi t.
\tag{5.6}
$$

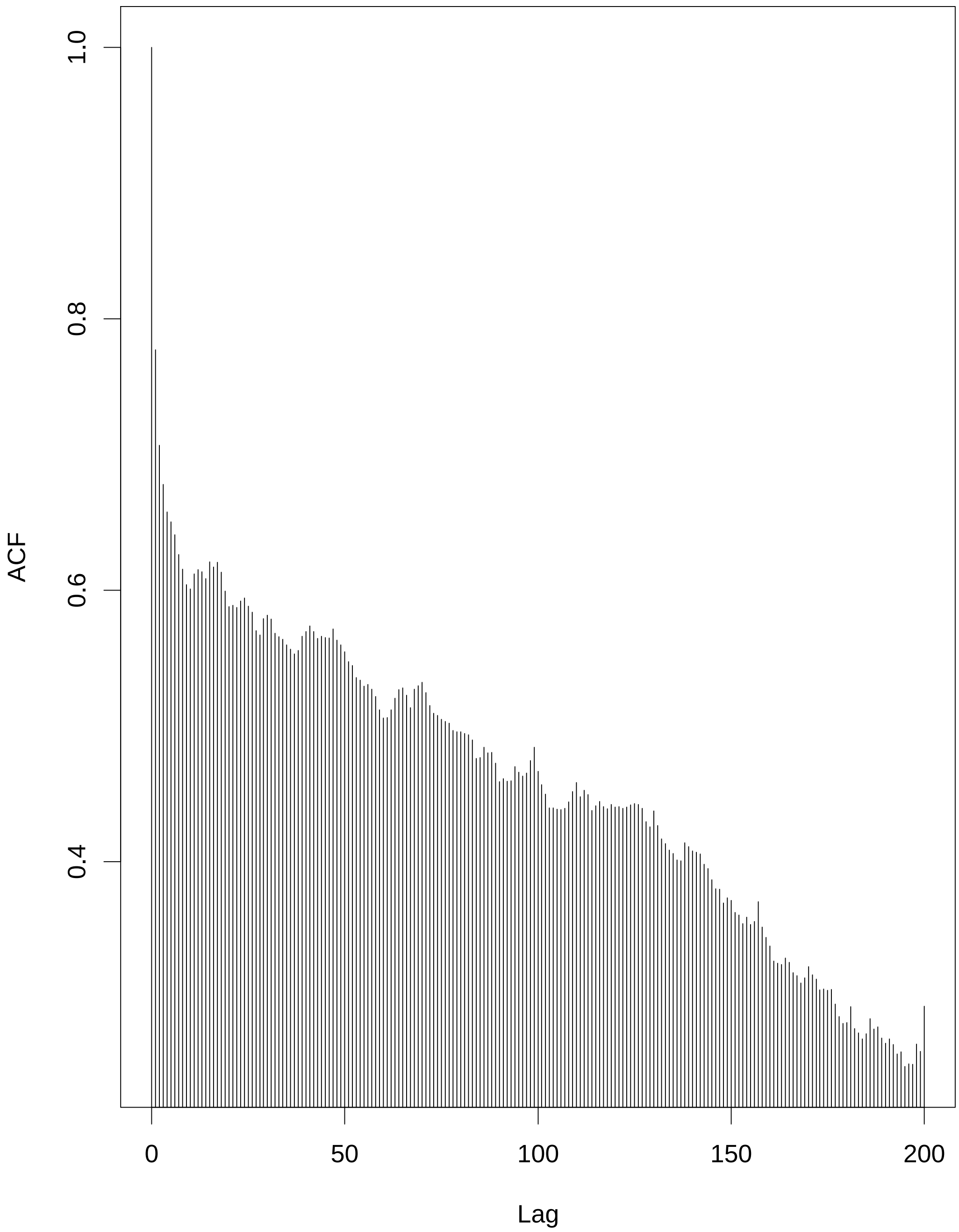

Figure 5.1: Autocorrelation Function for Number of Packets Per Minute of Size 400-600 Bytes

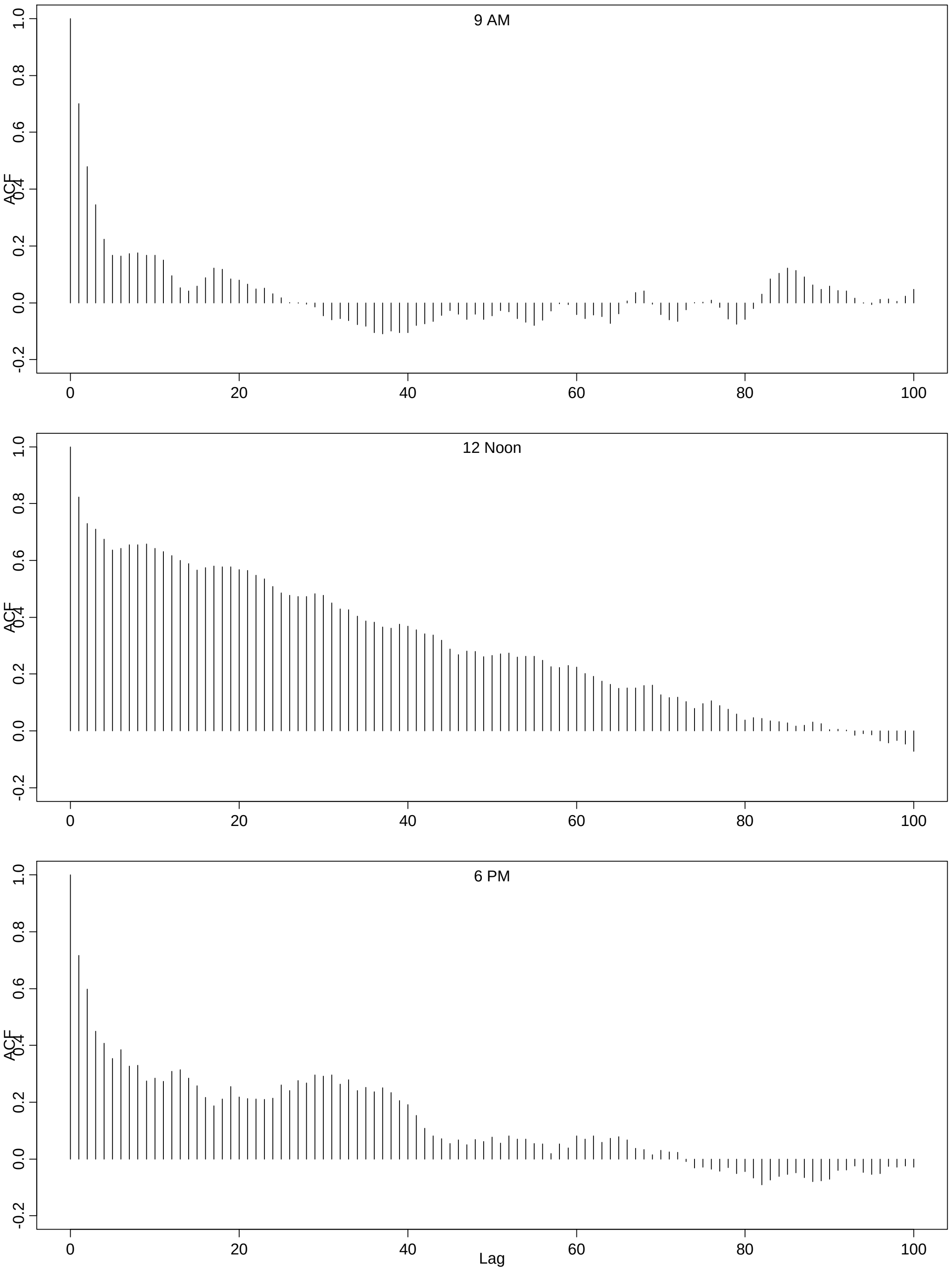

Figure 5.2: Autocorrelation Function for Number of Packets Per Second of Size 400-600 Bytes

To derive (5.6) we used the fact that

$$\int \sin 2\pi\tau \cdot \sin 2\pi(\tau + t) d\tau = \frac{\tau}{2} \cos 2\pi t - \frac{\sin 2\pi(2t + \tau)}{8\pi}.$$

Probably the reader will find it interesting to compare (5.5) and Fig.5.3 with Fig.1.8b from Beran (1994), Chpt.1.

If $\varphi_1(t) = \sin 2\pi\nu_1 t$ and $\varphi_2(t) = \sin 2\pi\nu_2 t$, then the function $c(t)$ contains the term proportional to

$$\cos 2\pi \frac{\nu_1 + \nu_2}{2} t \cos \frac{\nu_1 - \nu_2}{2} t,$$

which reminds "beat effect" in oscillation theory. When ν_1 and ν_2 are close then the covariance function may exhibit long wave oscillation.

For the linear trend with $\varphi(\tau) = \tau$

$$c(t) = \lambda^2 \left[\frac{(\tau_L - t)(2\tau_L + t)}{6} - \frac{\tau_L^2}{4} \right]; \tag{5.7}$$

see also Fig.5.3.

For the more general linear case with $\varphi_1(\tau) = 1$ and $\varphi_2(\tau) = \tau$ the formula becomes cumbersome but the function $c(t)$ is still a polynomial of the second order in t. From the definition of $c(\tau)$ it follows that if $\underline{\lambda}^T \varphi(\tau)$ is a polynomial of the k-th order then $c(t)$ is a polynomial of the $2k$-th order.

Summarizing (5.1) - (5.7) we can conclude that in the presence of a trend the sample covariance function exhibits symptoms of long range dependence.

Of course, whenever it is possible we have to subtract the trend from the analyzed processes. However, it is not an easy statistical problem, especially when the trend has a complicated structure or is intensively oscillating.

The typical examples of the continuous approximation $\overline{\rho}(t)$ of the averaged sample correlation function in the presence of the linear and sine type trend are exhibited in Fig.5.3 parts (a) and (b), respectively. We actually plotted the function $\overline{\rho}(t)|\rho(0)$ to make the comparison easier with the sampled correlation functions (see Fig.5.1 and 5.2). The second term in (5.5) was modeled by e^{at^2} with $a = 100$). At both plots $\zeta = 0.01$ and $0 \leq t \leq 1$. The reader can find the similarity between the observed and modeled functions.

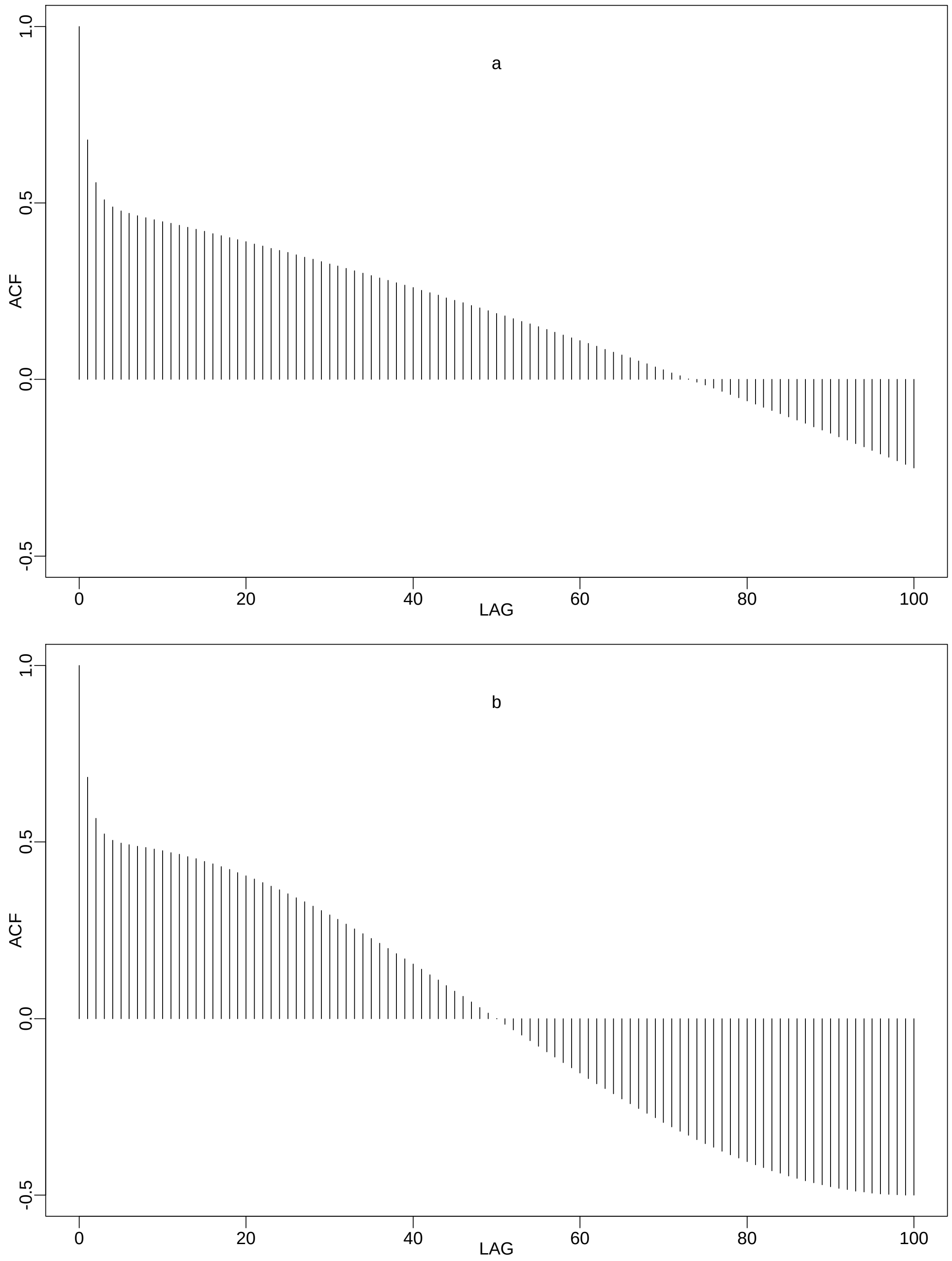

Figure 5.3: Examples of the Continuous Approximation of the Averaged Sample Correlation Function

The trend influence on other statistics was discussed in the statistical literature and most frequently in the context of long-range dependence and self-similarity; see, for instance, Beran (1994), Bhattacharya et al. (1983), Grossglauser and Bolot (1996) for examples and further references.

Random intensity. We consider here only a very simple version of double stochasticity. Namely, we assume that Λ is constant at each segment ζ and that Λ_i are independently sampled from a population with

$$E_\Lambda(\Lambda) = \Lambda_0 \text{ and } E_\Lambda\left[(\Lambda - \Lambda_0)^2\right] = \sigma_\Lambda^2.$$

The doubly averaged sample covariance function can be easily found by averaging (5.2) with respect to Λ:

$$\underline{\rho}(k) = E_\Lambda\left[\overline{\rho}(k)\right] = \sigma_{0k}\Lambda_0 - \frac{\Lambda_0 + \sigma^2}{L} \tag{5.8}$$

For the larger L the second term in (5.8) can be neglected, i.e. $\underline{\rho}(k)$ differs significantly from 0 only when $k = 0$. Thus, the use of sample covariance function helps to justify whether a histogram "flattening" caused by a trend or by a stochastic behavior of the intensity. In the first case the "flattening" or heavy tails are accompanied by the slowly decaying sample covariance (similar to what is observed for time series with long-range dependence). In the second one the sample covariance function is δ-shaped. For a short summary see Table 2.

Evidently, this table may be elaborated and extended. First, analyzing the behavior of descriptive statistics for other types of the Poisson processes such as compound and self-exciting, or for a mixture of various types will lead to an increase of a number of columns. Second, additional descriptive statistics such as various smoothers for the observed time series, residual plots, outlier diagrams, etc., can be analyzed in the manner similar to what was done above, and the corresponding rows will be a natural extension of the table.

Table 2: Behavior of two descriptive statistics under various assumptions

	Standard Poisson	Randomly Changing Intensity	Systematic Trend
"Flattened" Histogram	−	+	+
Heavy Tailed Covariogram	−	−	+

6. Conclusions

The results reported in Section 3 and 4 reveal that simple descriptive statistics such as histograms help to better understand network performance if they are interpreted properly within the framework of the reasonable set of assumptions on the "physics" of the processes under analysis.

The information which histograms carry in the case of either inhomogeneous, or compound or doubly Poisson processes (or any combination of the above), is obviously different from the information which usually might be extracted in the classical setting. The shape of the histogram gives us the opportunity to infer about the intensity of flows and their dynamics. For instance, the isolated peaks in the histogram (see Fig.4.1) tell us that the system spent some time in a relatively stable state with the intensity parameter corresponding to the "center" of that peak. Comparison of the width of the peak with the standard deviation of the Poisson distribution ($\sim$ square root of its center coordinate) allows one to conclude about the presence of randomness of parameters describing the intensity function (see (3.10) - (4.1)). The long plateaus indicate the presence of significant trends, for instance, linear changes of intensity in time. The hollows in the histogram (see area at the vicinity of 6000 packets/minute in Fig.4.1) may correspond to the intensities occurring only during some transitional periods: for example, the transition from the night hours activity to the day time hours activity.

It must be emphasized that the same patterns in the histogram shape may be explained differently (see comments to (4.1) and the introductory paragraph of Section 4) if no additional information is used. For instance, the plateau type pieces of the histogram may be generated by the linear trend as well as by the randomly uniformly distributed intensity parameter. Comparison of the observed histogram and the modeled average distribution as described in Section 4 allows us to verify the validity of various assumptions about the network performance. The simplicity of the proposed algorithm makes it possible to develop interactive procedures for performing the comparison in an "on-line" manner.

Matching of the sample covariance functions and histograms (see Section 5) allows us to discriminate between trends and double stochasticity and consequently to thin the set of possible solutions. However, as it was noted in the comments to (3.9) even if we know that the existing patterns in the analyzed histograms and the sample covariance indicate the presence of a trend the solution is still not unique unless we select some specific class of possible trends.

The generalization of the reported approach to other descriptive statistics is expected to be very straightforward. For example, the "averaged" empirical distributions of inter-arrival times or the shape of "averaged" correlograms can be explained with the use of the techniques similar to those from Section 3. Combining the corresponding findings with ideas developed and discussed, for instance, by Beran (1994), Chpt.1, Basu et al. (1995), Jain and Routhier (1986), and Willinger et al. (1995) may illuminate some results from self-similar and long-memory process theory.

7. Acknowledgment

Support from the Oak Ridge National Laboratory Directed Research and Development Fund is gratefully acknowledged.

8. References

[1] Andersen P., Brogan O., Gill R. and Keiding N. (1992). *Statistical Models Based on Counting Processes*, New York, Springer-Verlag.

[2] Basu S., Klivansky S. and Mukherjee A. (1995) "Time Series Models for Internet Traffic" GIT-CC-95-27, Georgia Institute of Technology, Atlanta.

[3] Beran J. (1992), "Statistical Methods for Data with Long-Range Dependence," Statistical Sciences, 7, 404-427.

[4] Beran J. (1994), *Statistics for Long-Memory Processes*, Chapman and Hall, New York.

[5] Bhattacharya R., Gupta V, Waymire E. (1983), "The Hurst Effects Under Trends", J. Appl. Prob. 20, 649-662.

[6] Charnes A., Frome E. and Yu P. (1976). "The Equivalence of Least Squares and Maximum Likelihood Estimates in the Exponential Family". JASA, 71, 216-222.

[7] Claffy, K. C. (1994), "Internet Traffic Characterization", UCSD.

[8] Davis P. and Rabinowitz P. (1984). *Methods of Numerical Integration* (2nd Edition), New York, Academic Press.

[9] Fedorov V. (1974). "Regression Problems with Controllable Variables Subject to Error", Biometrika, 61, 49-56.

[10] Grossglauser M. and Bolot J. C. (1996), "On Relevance of Long-Range Dependence in Network Traffic", RR-2830, INRIA Sophia Antipolis, France.

[11] Gustafsson E. and Karlsson G. (1997). "A Literature Survey on Traffic Dispersion", IEEE Network, 5, 28-29.

[12] Harville D. (1977), "Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems", JASA, 72 300-340.

[13] Jain R. and Routhier S. (1986). "Packet Trains - Measurements and New Model for Computer Network Traffic", IEEE Journal on Selected Areas in Communications. SAC-4, 986-995.

[14] Johnson N., Kotz S., Balakrishnan N. (1994). *Continuous Univariate Distributions*, 2nd Edition. J. Wiley, New York.

[15] Kovalenko I., Kuznetsov N. and Shurenkov V. (1996), *Models of Random Processes. A Handbook for Mathematicians and Engineers*, CRC Press, New York.

[16] Ortega J. and Rheinboldt W. (1970). *Iterative Solution of Nonlinear Equations in Several Variables.* Academic Press, New York.

[17] Paxson V. and S. Floyd (1995). "Wide Area Traffic: The Failure of Poisson Modeling", IEEE/ACM Transactions on Networking, 3, 226-244.

[18] Samorodnitsky G. and Taqqu M. (1994). *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance.* Chapman and Hall, New York.

[19] Seber G. and Wild C. (1989). *Nonlinear Regression*, J. Wiley, New York.

[20] Snyder D. (1975). *Random Point Processes.* J. Wiley, New York.

[21] Wilks S. (1962). *Mathematical Statistics.* J. Wiley, New York.

[22] Willinger W., Taqqu M., Leland W. and Wilson D. (1995). "Self-Similarity in High-Speed Packet Traffic: Analysis and Modeling of Ethernet Traffic Measurements", Statistical Science, 10, 67-85.

[23] Willinger W., Taqqu M. and Erramili A. (1996). "A bibliographical guide to self-similar traffic and performance modeling for modern high-speed networks", in Stochastic Networks: Theory and Applications, Kelly F., Zachary S., Ziedins I. (Eds.), Oxford, U.K., Clarendon Press.

A. Appendix 1 - Properties of the iterated estimator.

Let

1. $\underline{\lambda}_{true}$ be the vector of true values of the estimated parameters $\underline{\lambda}$, and let

$$\underline{\lambda}_{true} \in \Omega \subset R^m, \tag{A.1}$$

   where $m = \dim \underline{\lambda}$ and Ω is compact, and minimization in (4.9) be understood as minimization with respect to $\lambda \in \Omega$;

2. ζ is fixed and for any $\underline{\lambda} \in \Omega$ there exists the following limit

$$\lim_{L \to \infty} \frac{1}{L} M_L(\underline{\lambda}) = \mathcal{M}(\underline{\lambda}) \tag{A.2}$$

   where $M_L(\underline{\lambda}) = \sum_{i=1}^{L} \omega_i(\underline{\lambda}) \phi(\tau_i) \phi^T(\tau_i)$ and $\mathcal{M}(\underline{\lambda})$ is a positive definite matrix;

3. for all $\underline{\lambda} \in \Omega$ and all $\tau_i \quad 0 < \underline{\lambda}^T \phi(\tau_i) \leq \delta < \infty$.

Theorem 1. *If the above assumptions hold then*
1) The vector $\sqrt{L}(\underline{\lambda}_L^* - \underline{\lambda}_{true})$, where $\underline{\lambda}_L^* = \arg\max_{\underline{\lambda} \in \Omega} \mathcal{L}(\underline{\lambda}; n_1, \dots, n_L)$, is asymptotically normally distributed with zero mean components and with the dispersion matrix $\mathcal{M}^{-1}(\underline{\lambda}_{true})$.
2) The probability $P(L)$ that the iterative procedure converges to λ_L^* tends to 1 when $L \to \infty$, i.e.

$$\lim_{L \to \infty} P(L) = 1.$$

Proof. The first part of the theorem is a corollary of more general result about strong consistency and asymptotic normality of the maximum likelihood estimators (see, for instance, Wilks (1962), Chpt.12). We need assumption (b) to guarantee the regularity conditions of the maximum likelihood function.

To prove the second part of the theorem let us note that the iterative procedure (4.7) and (4.8) is a special case of the fixed point method (c.f. Ortega and Rheinboldt (1970), Chpt.5). To assure the convergence of (4.8) we have to verify that the mapping $\mathcal{M}_L^{-1}(\underline{\lambda}) Y_L(\underline{\lambda})$ is contractive, i.e.

$$\| M_L^{-1}(\lambda) Y_L(\lambda) - M_L^{-1}(\underline{\lambda}') Y_L(\underline{\lambda}') \| \leq \alpha \| \underline{\lambda} - \underline{\lambda}' \|, \tag{A.3}$$

where $0 < \alpha < 1$, for all $\underline{\lambda}, \underline{\lambda}' \in \Omega$. Introducing $\varepsilon_i = n_i - \underline{\lambda}_{true}^T \phi_i$ we can rewrite (A.3) as

$$\| M^{-1}(\underline{\lambda}) \sum_{i=1}^{L} \omega_i(\underline{\lambda}) \phi_i \varepsilon_i - M^{-1}(\underline{\lambda}') \sum_{i=1}^{L} \omega_i(\underline{\lambda}') \phi_i \varepsilon_i \| \leq \alpha \| \underline{\lambda} - \underline{\lambda}' \| \tag{A.4}$$

Noting that the random variables ε_i are independent, $E(\varepsilon_i) = 0$ and $0 < Var(\varepsilon_i) = E(\varepsilon_i^2) = \underline{\lambda}_{true}^T \phi_i \leq \delta$ and applying the strong law of large numbers (c.f. Wilks (1962), Chpt.4) we conclude that

$$\| M^{-1}(\lambda) \sum_{i=1}^{L} \omega_i(\lambda) \phi_i \varepsilon_i - M^{-1}(\lambda') \sum_{i=1}^{L} \omega_i(\lambda') \phi_i \varepsilon_i \| \leq \beta_L(\varepsilon_1, \ldots, \varepsilon_L) \| \underline{\lambda} - \underline{\lambda}' \|, \tag{A.5}$$

and β_L almost surely converges to 0 (c.f. Rao (1973), Chpt.2), i.e.

$$\beta_L(\varepsilon_1, \ldots, \varepsilon_L) \overrightarrow{a.s.} 0. \tag{A.6}$$

Combining (A.4) - (A.6) we conclude that the probability $P(L)$, that inequality (A.3) holds, converges to 1 and that proves the second part of the theorem.

Note. We formulated assumptions (b) and (c) in the form which allows to a reader who is familiar with nonlinear regression analysis to compare Theorem 2 (at least its first part) with many other similar results. Actually both of the assumptions may be replaced by the more compact assumption that all components of the vector function $\phi(\tau)$ are bounded. The proper choice of Ω must secure positiveness of $\underline{\lambda}^T \phi(t)$.

Interestingly enough that the simple linear model $\lambda_0 + \lambda_1 t$ does not satisfy (b) and (c) together with any other model containing component(s) $\phi_\alpha(t)$ such that

$$\lim_{t \to \infty} \phi_\alpha(t) = \infty.$$

The results, very similar to Theorem 1, can be proved for that type of models, if one introduces different (see A.2) normalization of the matrix $M_L(\underline{\lambda})$. Actually, the convergence of $\underline{\lambda}_L^*$ to $\underline{\lambda}_{true}$ will be faster than in the considered case.

B. Appendix 2 - Averaging the sample covariance function.

There are a few popular choices of the sample covariance function (c.f. Anderson (1994), Chpt.8). They are noticeably different for short data series or for large lags. In the computer network setting the discrepancies are negligible and we discuss the following sample covariance function (frequently called "sample auto-covariance function"):

$$\rho(k) = \frac{1}{L-k} \sum_{i=1}^{L-k} (n_i - \overline{n})(n_{i+k} - \overline{n}), \tag{B.1}$$

where $\overline{n} = \frac{1}{L} \sum_{i=1}^{L} n_i$.

We are interested in the Poisson distribution in which

$$E(n_i) = \Lambda_i \quad E(n_i^2) = \Lambda_i^2 + \Lambda_i.$$

Taking expectation of both parts of (B.1) we have for $k > 0$:

$$\begin{aligned} \overline{\rho}(k) &= E\left[\rho(k)\right] = \frac{1}{L-k} \sum_{i=1}^{L-k} E\left(n_i n_{i+k} - n_i \overline{n} - n_{i+k}\overline{n} + \overline{n}^2\right) \\ &= \frac{1}{L-k} \sum_{i=1}^{L-k} \left(\Lambda_i \Lambda_{i+k} - \Lambda_i \overline{\lambda} - L^{-1}\Lambda_i - \Lambda_{i+k}\overline{\Lambda} - L^{-1}\Lambda_{i+k} + \overline{\Lambda}^2 + L^{-1}\overline{\Lambda}\right) \\ &= \frac{1}{L-k} \sum_{i=1}^{L-k} \Lambda_i \Lambda_{i+k} - \left(\overline{\Lambda} + L^{-1}\right)\left(\overline{\Lambda}_k + {}_k\overline{\Lambda} + {}_k\overline{\Lambda} - \overline{\Lambda}\right), \end{aligned} \tag{B.2}$$

where $\overline{\Lambda}_k, {}_k\overline{\Lambda}$ and Λ are defined in the main text.

If $k = 0$, then the first term in the above expression splits in two parts, i.e.

$$\frac{1}{L} \sum_{i=1}^{L} E\left(n_i^2\right) = \frac{1}{L} \sum_{i=1}^{L} \Lambda_i^2 + \frac{1}{L} \sum_{i=1}^{L} \Lambda_i = \overline{\Lambda}^2 + \overline{\Lambda}.$$

Thus (see also (5.2)):

$$\overline{\rho}(k) = \frac{1}{L-k} \sum_{i=1}^{L-k} \Lambda_i \Lambda_{i+k} + \left(\overline{\Lambda} + L^{-1}\right)\left(\overline{\Lambda}_k + {}_k\overline{\Lambda} - \overline{\Lambda}\right) + \delta_{0k}\overline{\Lambda}. \tag{B.3}$$

INTERNAL DISTRIBUTION

1–4. S. G. Batsell
5. C. K. Bayne
6. W. R. Corwin
7. D. F. Craig
8. T. S. Darland
9. C. S. Daw
10–13. D. J. Downing
14–17. T. S. Dunigan
18–22. V. V. Fedorov
23. C. Glover
24. H. W. Hayden, Jr.
25. S. K. Iskander
26. P. Kanciruk
27. M. Kuliasha
28–32. M. R. Leuze
33. D. E. McCabe
34. R. C. Mann
35. J. G. Merkle
36. M. D. Morris
37. C. E. Oliver
38. G. Ostrouchov
39. R. F. Sincovec
40. P. T. Singley
41. D. A. Wolf
42. K-25 Applied Tech. Library
43. Y-12 Technical Library
44. Laboratory Records - RC
45–46. Laboratory Records Dept.
47. Central Research Library
48. M & C Records Office
49. ORNL Patent Office

EXTERNAL DISTRIBUTION

50. Prof. A. C. Atkinson, London School of Economics, London WC2A 2AE, United Kingdom

51. Dr. R. Beckman, Statistics Group Al, Los Alamos National Laboratory, MS F600, Los Alamos, NM 87545

52. Dr. J. Beran, Biostatistics Department, ISPM, University of Zurich, Sumatrastr. 30, 800b Zurich, Switzerland

53. Dr. K. C. Claffy, Computer Science and Engineering, University of California, San Diego, CA

54. Prof. R. D. Cook, School of Statistics, University of Minnesota, 1994 Bufford Avenue, St. Paul, MN 55108-6042

55. Prof. D. Cox, Department of Statistics, Rice University, Houston, TX 77251-1892

56. G. J. Davies, Department of Mathematics, Georgia State University, Atlanta, GA 30303-3083

57. Dr. M. Grossglauser, Unite′ de recherche INRLA, Sophia-Antipolis, 2004 route es Lucioles, BP93, 069002 SOPHIA-ANTIPOLIS Cedex, France

58. Prof. P. Hackl, Department of Statistics, Vienna University of Economics, Augasse 2-6, A-1090 WIEN, Austria

59. Dr. D. Hall, Statistics, Systems Department, Pacific Northwest Laboratory, P.O. Box 999, Richland, WA 99352

60. Prof. J. S. Hunter, 503 Lake Drive, Princeton, N.J. 08540
61. Dr. R. Jain, Digital Equipment Corporation, Littleton, MA 01460
62. Prof. M. Johnson, Department of Statistics, University of Central Florida, Orlando, FL 32816-0370
63. Prof. F. P. Kelly, Statistical Laboratory, University of Cambridge, 16 Mill Lane, Cambridge CB2 1SB, U.K.
64. Prof. S. Klivansky, College of Computing, Georgia Institute of Technology, Atlanta, GA 30332
65. Dr. A. M. Liebetrau, Computational Sciences Department, Battelle-Northwest, P. O. Box 999, Richland, WA 99352
66. Dr. M. McKay, Statistics Group Al, Los Alamos National Laboratory, MS F600, Los Alamos, NM 87545
67. B. K. Ryn, Department of Electrical Engineering and Center for Telecommunication Research, Columbia University, New York, N.Y. 10027.
68. Dr. J. Sacks, NISS, P. O. Box 14162, Research Triangle Park, NC 27709-4162
69. Prof. S. M. Stigler, Department of Statistics, University of Chicago, 5734 University Avenue, Chicago, IL 60637-1546
70. Prof. W. Welch, Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada
71. Dr. W. Willinger, Bellcore, 445 South Street, Morristown, NJ 07960
72. Prof. K. W. Weng, Biostatistics, UCLA School of Public Health, 10833 Le Conte Ave., Los Angeles, CA 90024-1722
73. Prof. J. Wu, Department of Statistics, Mason Hall, University of Michigan, Ann Arbor, MI 48109
74. Prof. D. Ylvisaker, Department of Mathematics, University of California, Los Angeles, CA 90024

75–76. Office of Scientific and Technical Information, P. O. Box 62, Oak Ridge, TN 37830