Comparative Study of Heuristic Hybrid of Markov Chain Monte Carlo and Dynamic Programming Methodologies for Network Fault Analysis

M. Jaudet†, N. Iqbal†, Nasir M. Mirza††, Sikander M. Mirza†† and Amir Hussain†††,

†Department of Electrical Engineering, Pakistan Institute of Engineering & Applied Sciences, Islamabad, Pakistan.
††Department of Physics & Applied Mathematics, Pakistan Institute of Engineering & Applied Sciences, Islamabad, Pakistan.
†††Department of Computing Science & Mathematics, Sterling University, Edinburgh, Scotland, UK.

Summary
Modeling of network-faults based time-sequence data by piecewise constant intensity function has been carried out using a heuristic approach that employs both Markov Chain Monte Carlo approach (MCMC) and Dynamic Programming algorithm (DPA) methodologies. The results for synthetic as well as for real data show that both MCMC and DPA have close agreement between predicted and actual values. Remarkable speedup (4 to 5 times) has been observed by augmentation of the heuristic method. Due to higher efficiency the proposed approach is well suited for cases with larger data sets requiring near-optimal solution.

Key words:
Data mining, Dynamic programming, Event sequence, Change-points, Maximum likelihood, MCMC.

1. Introduction

Random event sequences are encountered in a wide variety of applications, e.g., telecommunications networks, server logs and patient treatments. Events are normally tagged by event type and time of occurrence. In a typical situation, an event sequence may consist of hundreds of thousands of events. Telecommunications networks grow rapidly with the availability of devices based on newer technologies.

Modern communications devices send signals to management software over the network. These devices report their status and alarms to network management software for central monitoring. Alarms are indicators of various problems in the network [6]. This data often goes through correlation filters. It contains hundreds of different event types.

The characteristics of the data change as the result of changes in the network structure. Automatic methods to detect changes in intensities between stable periods are of interest for such data analysis. Detection of change points allows compression of very long sequences of alarm data, which can later be correlated to original alarm data to find interaction patterns.

Detection of change in the density of event occurrences as a function of time is of prime importance for network fault analysis. A sequence of conditionally independent events is based stochastically on some underlying process which may be described by an intensity or frequency distribution function. Such a process can be modeled as a Poisson or Normal distribution depending on real-time frequency distribution.

In this work a Bayesian model similar to [2, 8, 9] to find change-points in the sequence of telecommunications events is proposed. We have carried out several simulations to segment data in piecewise constant intensity functions. Based on Poisson likelihood criterion we chose the optimal piecewise constant intensity function. We compare the results of these simulations with those obtained with dynamic programming algorithm [12, 7] to obtain optimal piecewise constant intensity functions.

In this paper, first a Bayesian model has been developed with fixed dimensions for parametric estimation of change points. A Markov Chain Monte Carlo method is then used to sample from the target distribution. Later, a methodology to find change-points using dynamic programming as a non parametric approach is utilized. A likelihood criterion to choose optimal solution from available solutions is then applied along-with heuristic for speeding up the proposed methods. The results for a typical telecommunications network alarm data and synthetic data sets are then compared to establish best fitted data with models.

The paper is organized as following. Hierarchical Bayesian model and dynamic programming algorithm based approaches for change-point detection are described in Section 2. Poisson likelihood criterion to choose optimal solution is outlined in Section 3. Heuristic speedup of MCMC and DPA are discussed in Section 4. In Section 5, results from trails on data sets are given. Section 6 is a brief conclusion.

Manuscript received April 5, 2007
Manuscript revised April 25, 2007

2. Fault detection

To approximate piecewise constant intensity function [5], one needs to find time points where changes in intensity of events have occurred. Optimal intensity is then number of events divided by the length of the time period for discovered piece. Once these change points of an optimal piecewise constant intensity function are known we can compute intensity values giving likelihood solution. The change-points of the optimal piecewise constant function are always at occurrence times of the data sequence. Detection of changes in data has been done using variety of different methodologies [1]. Hierarchical Bayesian model and Dynamic programming have only been employed in this work to find change points.

2.1 Hierarchical Bayesian model

Arrival times of events can be modeled with a discrete time Poisson counting process [12]. The numbers of events counted in equal sized intervals (bins or time-points) obey Poisson distributions. So, we can write statistically:

$$y_i \sim \mathscr{P}(\lambda_k) \qquad (1)$$

where $i \in I_k = [c_{k-1}+1,...,c_k[, k=1,...,K$ are bins or time points in each segment k of the observed data with a Poisson parameter λ. c_k is the sample point after which the k^{th} change occurs in the event sequence data. By convention $c_0 = T_s$ and $c_K = T_e$, where T_s and T_e are start and end times of observation period $[T_s, T_e]$. In other words, the actual change locations are $t_k = c_k T + \tau$ with $0 \le \tau < T$, where T is the sampling period.

Segmenting an event sequence means the estimation of parameters K, $c = (c_1,...,c_{K-1})$ and $\lambda = (\lambda_1,...,\lambda_K)$ from the discretized data $y = (y_1,..., y_n)$. A hierarchical Bayesian model has been used in this work for segmentation of data similar to [2] for the estimation of these parameters. Bayesian estimation of these parameters requires prior information about the location of change-points and the Poisson parameters to be known.

The use of hyper-parameters improves the robustness in estimation of parameters [10]. This, however, increases difficulty in estimation of Maximum a posteriori (MAP) or minimum mean square (MMSE) estimators [2]. Markov Chain Monte Carlo (MCMC) methods can then be used to draw samples according to the posteriors of interest and the Bayesian estimators can be computed from these simulated samples.

The unknown parameters include K, $c = (c_1,...,c_{K-1})$ and $\lambda = (\lambda_1,...,\lambda_K)$. A standard procedure consists of introducing indicator function $r_i, i \in \{1,...,n\}$ such that $r_i = 1$ if there is a change-point at time point i and $r_i = 0$ otherwise. Using $r_n = 1$, the number of change-points and the number of steps become equal to $K = \sum_{i=1}^{n} r_i$. The unknown parameter vector resulting from this re-parameterization is $\theta = (r, \lambda, K)$ with $r = (r_1,...,r_n)$ and $\lambda = (\lambda_1,...,\lambda_K)$.

The unknown parameter vector θ belongs to a space whose dimension depends on K i.e. $\theta \in \Theta = \{0,1\}^n \times \mathrm{R}^K$. It is proposed that the estimation of unknown parameter vector θ should be carried out using Bayesian estimation theory. Bayesian inference on θ is based on the posterior distribution $f(\theta \mid y)$. The likelihood of the observed data vector y can be expressed as [2]:

$$f(y \mid \theta) = \frac{\prod_{k=1}^{K} \lambda_k^{s_k(r)} \exp(-\lambda_k n_k(r))}{\prod_{i=1}^{n} y_i!} \qquad (2)$$

where $s_k(r) = \sum_{i \in I_k} y_i$ (number of events in the k^{th} segment) and $n_k(r) = c_k - c_{k-1}$ (number of time points in the k^{th} segment). It is assumed that the probabilities of having $r_i = 0$ and $r_i = 1$ are *a priori* independent [8]. If P is the probability of $r_i = 1$ then indicator prior distribution of r being Binomial can be written:

$$f(r \mid P) = P^{\sum_{i=1}^{n-1} r_i} (1-P)^{n-1-\sum_{i=1}^{n-1} r_i} \qquad (3)$$

Using Gamma distributions $y_i \sim \mathscr{G}(\nu, \gamma)$ as priors for Poisson parameters λ_k, with ν being constant and γ as an adjustable hyper-parameter, the prior distribution of λ can be written as [2]:

$$f(\lambda \mid \gamma) = \left(\frac{\gamma^{\nu}}{\Gamma(\nu)} \right)^K e^{-\nu \sum_{k=1}^{K} \lambda_k} \prod_{k=1}^{K} \left(\lambda_k^{\nu-1} \mathrm{I}_{\mathrm{R}^+}(\lambda_k) \right) \qquad (4)$$

$\mathrm{I}_{\mathrm{R}^+}(\lambda_k)$ is the indicator function defined on R^+ such that $\mathrm{I}_{\mathrm{R}^+}(\lambda_k) = 1$ if $\lambda_k \ge 0$ and $\mathrm{I}_{\mathrm{R}^+}(\lambda_k) = 0$ otherwise.

The parameters P and γ needs to be fixed based on information about event sequence. But to increase the robustness of the algorithm these parameters can be considered as random variables with non-informative priors. So parameters P and γ need to be treated as hyper-parameters for assumed priors r and λ respectively. The hyper-parameter P can be assumed to be from a uniform distribution on $[0,1]$. The hyper-parameter γ has been taken as non-informative Jeffreys' prior. Both P and γ are assumed *a priori* independent [2].

The posterior distribution of the unknown parameter vector $\theta = (\lambda, r)$ and the hyper-parameter vector $\phi = (P, \gamma)$ can be computed from this hierarchical model:

$$f(\theta \mid y) = \int f(\theta, \phi \mid y) d\phi \tag{5}$$

The integration of parameters λ_k and P yields [2] the following mathematical expression:

$$\frac{f(r,\gamma \mid y)}{C(r \mid y)} \propto \frac{1}{\gamma}\left(\frac{\gamma^{\nu}}{\Gamma(\nu)}\right)^{K} \prod_{k=1}^{K} \frac{\Gamma(n_k(r)+\nu)}{(s_k(r)+\gamma)^{n_k(r)+\nu}} \tag{6}$$

where

$$C(r \mid y) = \Gamma\left(\sum_{i=1}^{n-1} r_i + 1\right)\Gamma\left(n - \sum_{i=1}^{n-1} r_i\right) \tag{7}$$

The posterior distribution as in Eq. 6 is quite complex. It is difficult to obtain closed-form expressions of the Bayesian estimators. A Gibbs sampling strategy, without reversible jumps can now be employed to estimate change-points. This can be done by sampling from conditional distributions $f(r \mid \gamma, y)$ and $f(\gamma \mid r, y)$ [2].

2.2 Standard dynamic programming algorithm

Standard Dynamic programming algorithm visits all the data points and describes the state of system based on maximum likelihood principle. It produces best known results however its execution cost is highest. It considers all the data points as potential change-points and number of maximum pieces is equal to size of the data set. Maximization of likelihood principle requires the knowledge about the origin of intensity and helps in selection of optimal number of pieces.

The time interval $[T_s, T_e]$ is divided into n independent parts and record time points for $n-1$ potential change-points. Let us denote the data sequence by E and the occurrence times of the events of E by $t_i : 1 \le i \le n$. Thus, $T_s < t_1 < t_2 < ... < t_n < T_e$, where T_s and T_e are the start and end points of the observation period. It is assumed that the change-points of the optimal piecewise constant intensity function with k pieces are known, and are denoted by a set $\{c_1, ..., c_k\}$.

The last change-point of the optimal k piece function with t_i is denoted as the end point of the time period by $C(k, t_i)$. It is assumed now that change-points of the optimal piecewise constant functions with k pieces in $[T_s, T_i]$ are known for all t_i with $i \le k$. Now the question is: how to find the change-points of the optimal function with $k+1$ pieces? The $L(k, t_i, t_j)$ is the (maximum) likelihood (of the data E) with an optimal piecewise constant intensity function having k number of pieces and the observation interval $[T_i, T_j]$. Then the maximum likelihood of the function with $k+1$ pieces is given by

$$L(k+1, T_s, t_j) = \max(L(k, T_s, t_i) + L(1, t_i, t_j)) \tag{8}$$

The standard algorithm computes the optimal division -into k pieces in the time interval $[T_s, t_i]$ for all t_i and uses these results when computing the best divisions into $k+1$ pieces. Then the change-points of the function are detected using backward substitutions. The standard dynamic programming algorithm finds both the optimal solution in the time range $[T_s, T_e]$ and also for all sub-ranges $[T_s, T_i]$. The time requirement of the of the algorithm is of the order of kn^2. The space required is of the order of kn due to storing values for all the sub-ranges and division sizes.

The dynamic programming algorithm uses a function L that calculates likelihood for varying length data segments or pieces. The calculation of this likelihood function depends on the probability distribution assumed for the process that generated the observed data. We assume only Poisson and Normal processes responsible intensity modeling and list their likelihood functions to be evaluated by standard dynamic programming algorithm.

For an event sequence S, which contains only events of a single event type $S = \{(e,t_1),(e,t_2),...,(e,t_n)\}$ with a time dependent intensity function $\lambda(t)$ the observation range being $[T_s, T_e]$. Let us split the observation range into "short" M intervals of length h (where $M = (T_s - T_e)/h$ and denote the intervals by $\Delta_k; k = 1,...,M$ and assume events in disjoint intervals are independent. Furthermore, we introduce the indicator function I such that $I(\Delta) = 1$ if an event occurred during the period Δ and $I(\Delta) = 0$ otherwise.

Assuming $\lambda(t)$ as rate parameter drawn from Poisson distribution, we can write the Poisson likelihood [3] as

$$L(S \mid \lambda) = \prod_{k=1}^{M} \frac{\exp(-\lambda(\Delta_k))}{I(\Delta_k)!} (\lambda(\Delta_k))^{I(\Delta_k)} \quad (9)$$

$$L(S \mid \lambda) = \exp\left(-\sum_{k=1}^{M} \lambda(\Delta_k)\right) \prod_{k=1}^{M} \lambda(t_k) \quad (10)$$

$$L(S \mid \lambda) = \exp\left(-\int_{T_s}^{T_e} \lambda(t)dt\right) \prod_{k=1}^{M} \lambda(t_k); (when\ h \to 0) \quad (11)$$

After by taking logarithms, we obtain log-likelihood for Poisson process as

$$l(S \mid \lambda) = -\int_{T_s}^{T_e} \lambda(t)dt + \sum_{k=1}^{M} \ln(\lambda(t_k)) \quad (12)$$

Assuming $\lambda(t)$ as rate parameter drawn from Normal distribution, the likelihood for the regular Normal distribution is calculated as following [4]:

$$L(\Delta; \lambda, \sigma) = \prod_{k=1}^{M} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(I(\Delta_k) - \lambda)^2}{2\sigma^2}\right) \quad (13)$$

After by taking logarithms, we obtain log-likelihood for Normal process as

$$l(\Delta; \lambda, \sigma) = -\frac{M}{2}\ln(2\pi) - M\ln(\sigma) - \frac{1}{2\sigma^2}\sum_{k=1}^{M}(I(\Delta_k) - \lambda)^2 \quad (14)$$

The evaluation of likelihood function is done by estimating the parameters of likelihood expression for each piece encountered by standard dynamic programming algorithm. Standard Dynamic programming algorithm is known to have complexity of the order $O(n^2)$ both in terms computations and memory requirements. Its performance does not depend on number of events or the length of monitoring interval. It depends on how the monitoring interval is discretized. The likelihood function calls increase monotonically as the discretized data set size is increased.

To estimate execution time of standard dynamic programming algorithm several simulations were carried out. Various discretized data set sizes ranging from 200 samples to 10,000 samples were used. The execution times are found to be between fractions of a second to twenty five hours. For further details about execution times of standard dynamic algorithm, please see [7].

3. Optimal piecewise constant intensity function

A piecewise constant frequency distribution has the following form:

$$\lambda(t) = \begin{cases} \lambda_i; & when\ c_{i-1} \le t \le c_i; i = 1, ..., n \\ 0; & otherwise \end{cases} \quad (15)$$

The values $c_0 = T_s$ and $c_n = T_e$ are the start and end times of the observation period, the values λ_i are the intensity values in i^{th} piece and $\{c_1, ..., c_{n-1}\}$ are the change-points of the function. For Poisson process, the log-likelihood [11] of data S given a piecewise constant intensity function with n pieces yields:

$$l(S \mid \lambda) = -\sum_{k=1}^{n} \int_{c_{k-1}}^{c_k} \lambda(t)dt + \sum_{k=1}^{n} \ln\{\lambda(t_k)\} \quad (16)$$

MCMC algorithm described in Section 2 partitions the data into piecewise constant intensity function with k segments and $k-1$ change-points. In each run a different value of k and different locations of change-points has been used. Similarly with the dynamic programming algorithm, the data can be partitioned into piecewise constant intensity function with $k = 2, ..., n-1$ segments and respective change-points. Eq. 16 has been used to select the optimal piecewise constant intensity function from various piecewise constant intensity functions computed from trials.

4. MCMC-DPA Hybrid Methodology

The standard MCMC and dynamic programming algorithms require huge runtimes for large discretized data sets. All time points in both algorithms are considered as potential change-point candidates for obtaining piecewise constant intensity functions. Both of these algorithms are modified to accept a reduced list of potential change-point candidates. This list stores pre-selected indices of discretized time points.

One of such possibility is logarithmic selection from with in the set of all potential candidates. The selection of the potential change-points is carried out by choosing every k^{th} time point of the discretized sequence to the candidate set. We first set k equal to 2^r where r is the greatest integer satisfying $2^r < n$ and n is the number of time points in the sequence. Then we decrease r by one, in steps. This procedure reduces the execution time of MCMC and dynamic programming algorithm irrespective of results of approximation.

Details regarding reduction in execution times of dynamic programming algorithm may be found in our previous work [7].

5. Empirical Results

Several constant intensity functions were generated with the MCMC and dynamic programming algorithms on both synthetic as well as real data sets used in this work. The results show differences between the computed piecewise constant intensity functions. A lot of change-points are however found common to both MCMC and dynamic programming algorithms when Poisson distribution is considered as underlying process. Results for instantaneous intensity data set made with Gamma distribution are shown in Fig. 1.

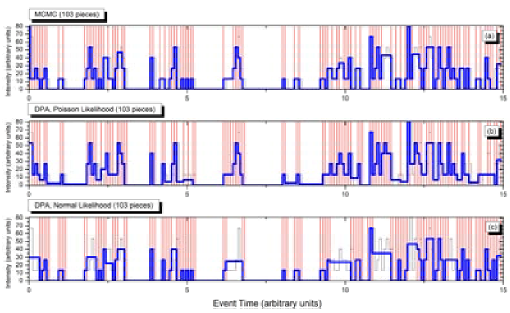


Fig. 1 Change-points of 103-piece division of instantaneous intensity data set made with Gamma distribution: MCMC (a) standard DPA Poisson (b) standard DPA Normal (c)

Both MCMC and dynamic programming algorithm segment a given data set to form piecewise constant intensity functions with variable number of pieces. Each run of MCMC algorithm can compute a different piecewise constant intensity function with different number of pieces. The range of number of pieces is found to short when compared to dynamic programming algorithm. The dynamic programming algorithm is capable of segmenting a data set in $k:2,...,n-1$ pieces.

To select optimal piecewise constant intensity function, each piecewise constant intensity function has to be evaluated with Eq. 16. The one giving the maximum value will be the optimal. Results show that piecewise constant intensity functions with MCMC algorithm fall between those obtained with dynamic programming algorithm using Poisson distribution and Normal distribution based likelihood functions.

We now present these results for our data sets generated with exponential distributions. We use all the time points of our data sets as potential candidates to compute piecewise constant intensity functions. Constant intensity functions using Bayesian modeling have been computed using MCMC simulation of Eq. 6. We use Eq. 12 and 14 with the dynamic programming algorithm for Poisson and Normal distribution based change-points. The values obtained from Eq. 16 for each piecewise constant intensity function are referred as likelihood curves.

Fig. 2 shows the likelihood curves for our constant intensity data set. Likelihood curve for MCMC algorithm gains maximum value for 163 piece division. MCMC algorithm is found to partition the data set between 167 and 198 pieces. Maximum gain in likelihood for Poisson distribution based change-points found using dynamic programming algorithm is found at 92 pieces. For Normal distribution based change-points found using dynamic programming algorithm, we get the maximum gain in likelihood when data is divided into 162 pieces.

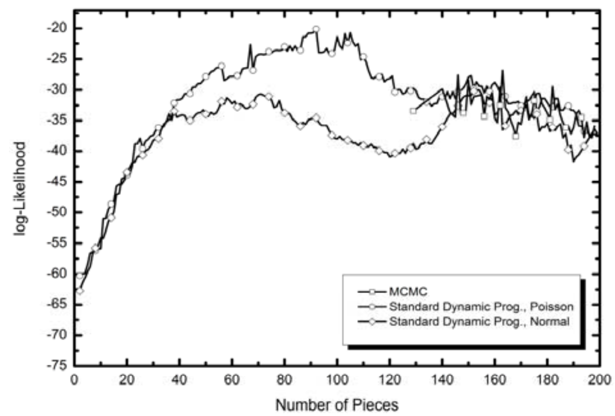


Fig. 2 Likelihood of piecewise constant intensity function: Exponential data set with constant intensity.

Fig. 3 provides results of trials related to increasing intensity data set produced with Exponential distribution. Likelihood curve for MCMC algorithm yields maximum value for 100 piece division. MCMC algorithm is found to partition the data set between 81 and 115 pieces. The maximum value for likelihood is found at 109 pieces for Poisson distribution based change-points (DPA). The maximum value of likelihood using Normal distribution based change-points (DPA) is achieved for division of 50 pieces.

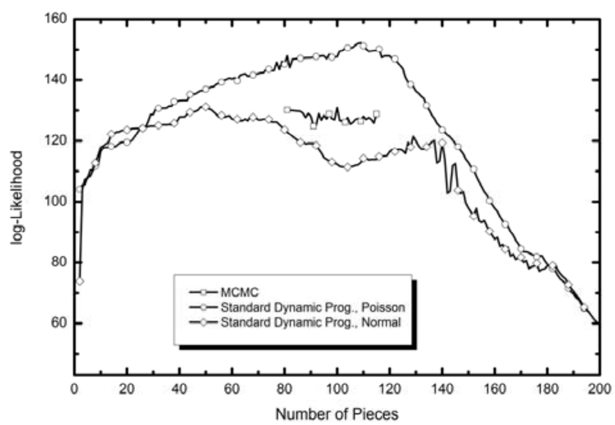


Fig. 3 Likelihood of piecewise constant intensity function: Exponential data set with increasing intensity.

Results for varying intensity data set constructed from Exponential distribution are shown in Fig 4. Likelihood curve for MCMC algorithm yields maximum value for 130 piece division. MCMC algorithm is found to partition the data set between 130 and 154 pieces. The maximum value for likelihood using Poisson distribution based change-points (DPA) is observed when 99 pieces are used. The maximum value of likelihood for Normal distribution based change-points (DPA) is achieved when 63 pieces are employed.

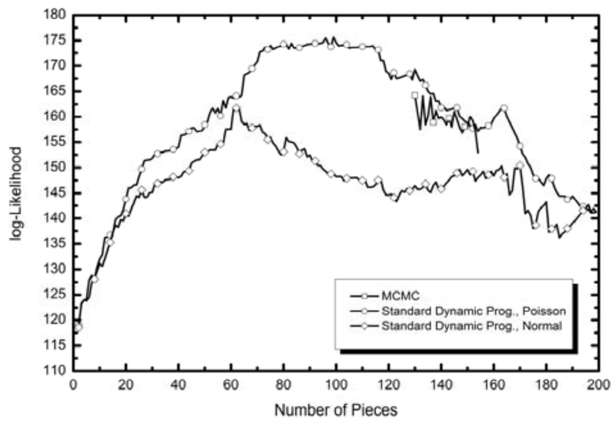


Fig. 4 Likelihood of piecewise constant intensity function: Exponential data set with varying intensity.

When instantaneous intensity data set is employed, likelihood curve for MCMC algorithm gains maximum value for 98 piece division. MCMC algorithm is found to partition the data set between 96 and 103 pieces. We find maximum value for likelihood at 99 pieces for Poisson distribution based change-points (DPA). The maximum value of likelihood using Normal distribution based change-points (DPA) is achieved at 65 pieces. These results are shown in Fig. 5.

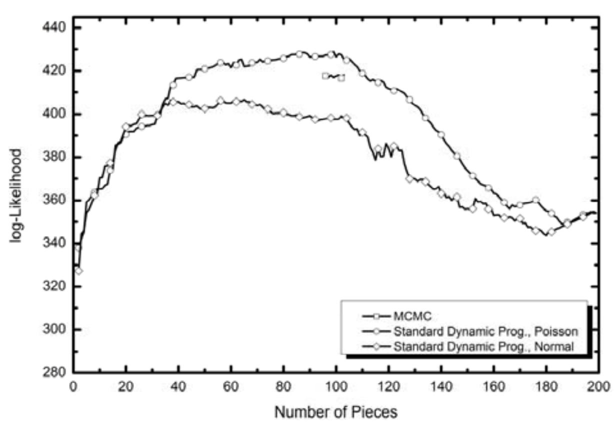


Fig. 5 Likelihood of piecewise constant intensity function: Exponential data set with instantaneous intensity.

In second part, the effect of reducing time points as potential change-points is observed. MCMC and dynamic programming algorithms are used with reduced number of potential change-point candidates based on logarithmic selection to find optimal constant intensity functions. Logarithmic selection decreases the required time to construct a piecewise constant intensity function exponentially with both MCMC and dynamic programming algorithms.

Fig. 6 shows likelihood values computed for optimal constant intensity functions employing the constant intensity data set using the Exponential distribution. The likelihood curves for MCMC and dynamic programming algorithm with Poisson and Normal distribution change points have increasing trend. The likelihood curve for MCMC stays below dynamic programming based change points with reduced candidates all the time.

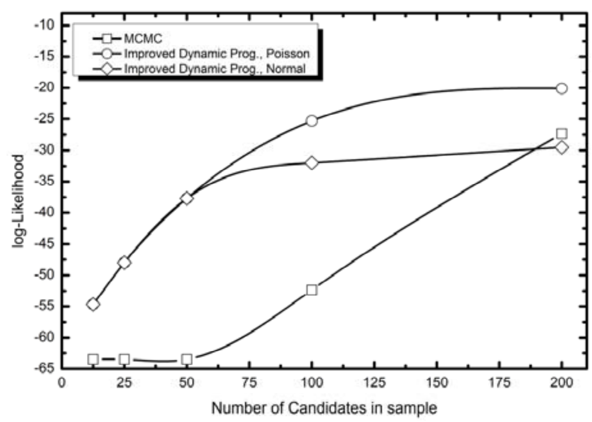


Fig. 6 Effect of candidate selection on likelihood: Exponential data set with constant intensity.

Likelihood values computed for optimal constant intensity functions employing the increasing intensity data set made from Exponential distribution are shown in Fig. 7. Again the likelihood curves for MCMC and dynamic programming algorithm with Poisson and Normal distribution change points have increasing trend. The likelihood curve for MCMC stays below dynamic programming based change points with reduced candidates all the time.

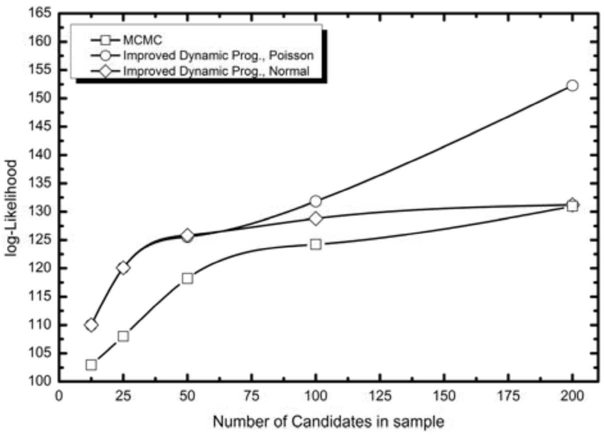


Fig. 7 Effect of candidate selection on likelihood: Exponential data set with increasing intensity.

Likelihood values computed for optimal constant intensity functions employing the varying intensity data set made from Exponential distribution are shown in Fig. 8. The likelihood curves for MCMC and dynamic programming algorithm with Poisson and Normal distribution change points have increasing trend. The likelihood curve for MCMC stays below dynamic programming based change points with reduced candidates all the time.

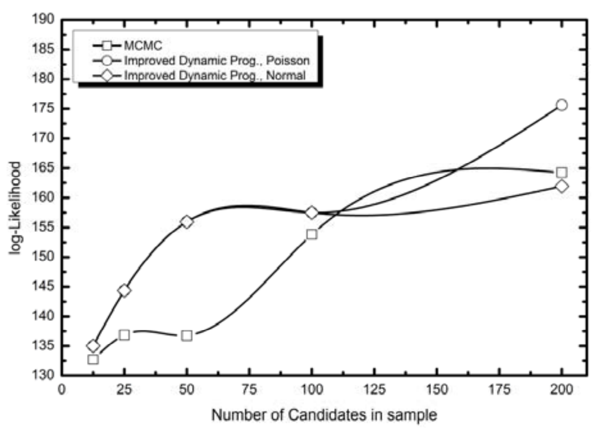


Fig. 8 Effect of candidate selection on likelihood: Exponential data set with varying intensity.

Results of instantaneous intensity data set obtained by using Exponential distribution are shown in Fig. 9. The likelihood curves for MCMC and dynamic programming algorithm with Poisson and Normal distribution change points have increasing trend.

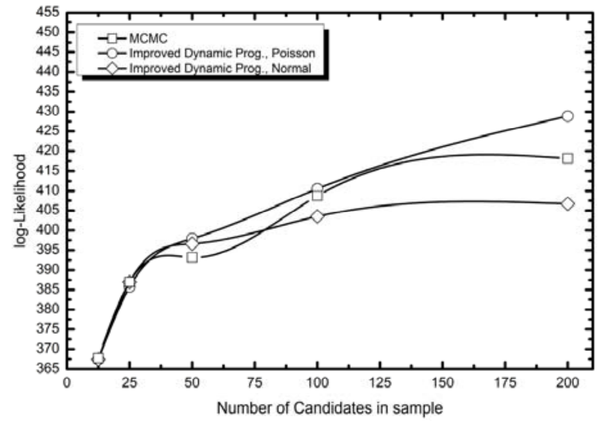


Fig. 9 Effect of candidate selection on likelihood: Exponential data set with instantaneous intensity.

For exponential samples with increasing intensity, the optimal piecewise intensity function has been computed using both standard dynamic programming and MCMC algorithms. Fig. 10 shows an optimal 109-piece constant intensity function computed using Poisson likelihood function using standard DPA. An optimal 50-piece constant intensity function computed with the Normal likelihood function using standard DPA is shown in Fig. 11. An optimal 100-piece constant intensity function computed with the MCMC algorithm is shown in Fig. 12.

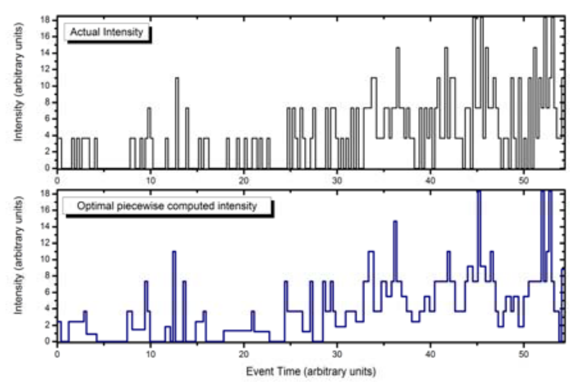


Fig. 10 Optimal piecewise constant intensity function computed using standard DPA Poisson likelihood.

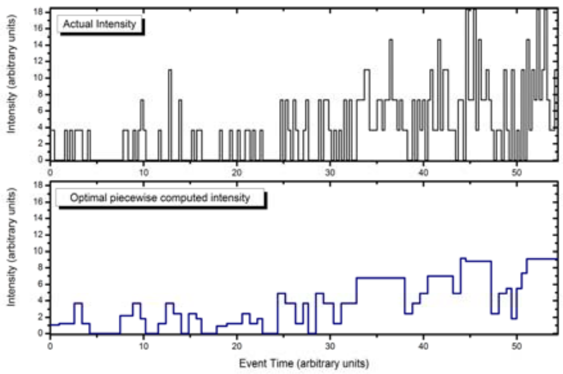


Fig. 11 Optimal piecewise constant intensity function computed using standard DPA Normal likelihood.

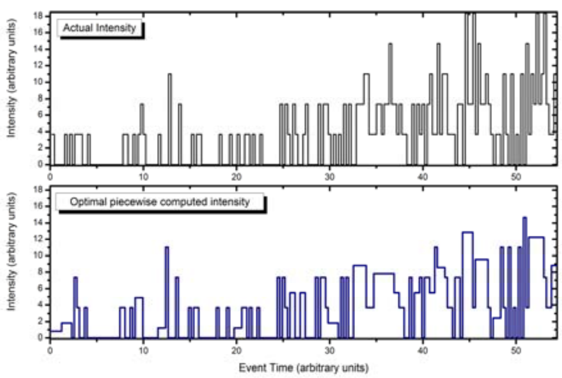


Fig. 12 Optimal piecewise constant intensity function computed using MCMC algorithm.

Table 1 through 2 show the results of optimal piecewise partitioning of the data sets with the help of Poisson and Normal distributions using dynamic programming and MCMC algorithms for finding change points. We have listed mean squared error values when an approximation is compared to the original discretized data. Optimal piecewise constant intensity functions are computed with all and varying number of change-point candidates.

Table 1: Comparison of intensity functions (using standard, improved DPA and MCMC) against real intensities using mean squared error.

Distribution/ Intensity	Approximation	Optimal 100% candi-dates	Optimal 50% candi-dates	Optimal 25% candi-dates	Optimal 12.5% candi-dates	Optimal 6.25% candi-dates
Exponential						
Constant	DPA, Poisson	2.020	1.756	1.835	1.870	1.908
Intensity	DPA, Normal	2.615	1.903	1.835	1.870	1.908
	MCMC	2.666	1.888	1.974	1.974	1.974
Increasing	DPA, Poisson	3.753	3.336	3.169	3.356	3.548
Intensity	DPA, Normal	3.166	3.213	3.165	3.356	3.548
	MCMC	4.683	3.363	3.390	3.586	3.666
Varying	DPA, Poisson	4.192	4.333	4.408	4.629	4.824
Intensity	DPA, Normal	4.177	4.333	4.408	4.629	4.824
	MCMC	5.053	4.561	4.736	4.764	4.870
Instantaneous	DPA, Poisson	17.059	14.515	13.848	14.324	15.449
Intensity	DPA, Normal	15.568	15.736	13.836	14.324	15.449
	MCMC	19.019	12.936	14.161	14.168	15.469
Normal						
Constant	DPA, Poisson	0.492	0.563	0.432	0.434	0.431
Intensity	DPA, Normal	0.580	0.563	0.426	0.434	0.431
	MCMC	0.427	0.427	0.427	0.427	0.427
Increasing	DPA, Poisson	1.883	1.791	1.641	1.655	1.656
Intensity	DPA, Normal	2.107	2.091	2.075	2.067	1.738
	MCMC	1.694	1.753	1.768	1.840	1.803
Varying	DPA, Poisson	2.944	2.635	2.581	2.620	2.680
Intensity	DPA, Normal	2.412	2.437	2.581	2.620	2.680
	MCMC	2.584	2.584	2.628	2.659	2.720
Instantaneous	DPA, Poisson	10.270	9.455	9.184	10.519	10.912
Intensity	DPA, Normal	9.283	9.585	9.186	10.519	10.912
	MCMC	13.693	9.782	10.008	10.415	11.002

5. Conclusion

Optimal modeling of time-sequence data by piecewise constant intensity functions by MCMC with fixed dimensions and dynamic programming have been performed. The conclusions from the current study are the following:

- Piecewise constant functions can be used in a flexible manner to represent time-sequence data.
- Augmentation of MCMC and standard dynamic programming algorithm by heuristics method has been found to yield remarkable speedup at the cost of minor loss in the accuracy of results.
- MCMC algorithm computes optimal piecewise intensity function with approximately same number of pieces as computed with dynamic programming

algorithm when Poisson distribution is considered as underlying process.

Table 2: Comparison of intensity functions (using standard, improved DPA and MCMC) against real intensities using mean squared error.

Distribution/ Intensity	Approximation	Optimal 100% candi-dates	Optimal 50% candi-dates	Optimal 25% candi-dates	Optimal 12.5% candi-dates	Optimal 6.25% candi-dates
Gamma						
Constant	DPA, Poisson	1.736	1.741	1.852	1.897	1.939
Intensity	DPA, Normal	1.799	1.839	1.852	1.904	1.939
	MCMC	2.499	1.985	1.985	1.985	1.985
Increasing	DPA, Poisson	4.131	3.417	3.151	3.208	3.277
Intensity	DPA, Normal	3.206	3.185	3.116	3.208	3.277
	MCMC	4.316	3.359	3.322	3.371	3.378
Varying	DPA, Poisson	4.195	4.026	4.319	4.577	4.646
Intensity	DPA, Normal	4.291	4.026	4.307	4.577	4.646
	MCMC	5.531	4.603	4.563	4.695	4.738
Instantaneous	DPA, Poisson	15.461	14.134	14.197	14.770	15.873
Intensity	Normal	14.302	15.094	14.130	14.770	15.873
	MCMC	18.636	13.783	14.380	15.307	15.932
Real						
data set I	DPA, Poisson	0.237	0.253	0.249	0.270	0.276
	DPA, Normal	0.236	0.256	0.248	0.270	0.276
	MCMC	0.283	0.283	0.283	0.283	0.283
data set II	DPA, Poisson	0.212	0.227	0.236	0.245	0.252
	DPA, Normal	0.212	0.227	0.236	0.245	0.252
	MCMC	0.260	0.260	0.260	0.260	0.260

Acknowledgments

M. Jaudet gratefully acknowledges the financial support from PIEAS Endowment Fund for the Higher Education and Research in IT and Telecom Sector, a scheme of Higher Education Commission (HEC), Pakistan for the Ph.D. fellowship.

References

[1] M. Basseville and I. V. Nikiforov, Detection of Abrupt Changes: Theory and Application, Prentice-Hall, Englewood Cliffs, NJ, 1993.

[2] N. Dobigeon, J. Tournert, D. J. Scargle, "Change-point detection in astronomical data by using hierarchical model and a Bayesian sampling approach," 13th Workshop on Statistical Signal Processing (IEEE/SP), 369-374, 2005.

[3] P. Guttorp, Stochastic modeling of scientific data. Stochastic modeling series, Chapman and Hall, London, 1995.

[4] J. W. Harris and H. Stocker, Handbook of Mathematics and Computational Science, Springer-Verlag, New York, 1998.

[5] D. Hawkins, "Point estimation of parameters of piecewise regression models," Journal Roy Stat Soc Ser C. 25,51-57, 1976.

[6] K. Hatonen, M. Klemettinen, H. Mannila, P. Ronkainen, and H. Toivonen, "TASA: Telecommunication alarm sequence analyzer, or how to enjoy faults in your network," in: Proceedings of the 1996 IEEE network operations and management symposium (NOMS'96). Kyoto, Japan, 520-529, 1996.

[7] M. Jaudet, N. Iqbal, N. M. Mirza, S. M. Mirza and A. Hussain, "Network fault analysis using hybrid heuristic-dynamic programming methodology," Computational Statistics and Data Analysis, in Process 2006.

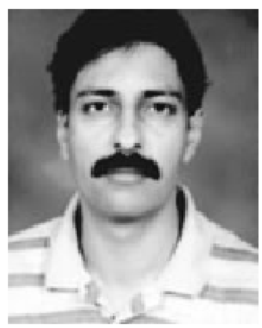
[8] M. Lavielle "Optimal segmentation of random processes," IEEE Trans. Signal Processing. 46,1365-1373, 1998.

[9] E. Punskaya, C. Andrieu, A. Doucet, and W. Fitzgerald, "Bayesian curve fitting using MCMC with applications to signal segmentation," IEEE Trans. Signal Processing. 50,747–758, 2002.

[10] C. P. Robert, The Bayesian Choice - A Decision Theoretic Motivation, Springer-Verlag, New York, 1994.

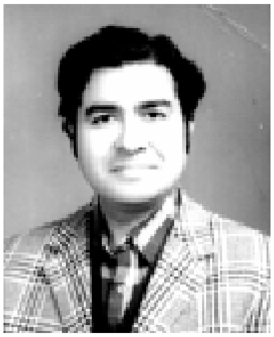
[11] M. Salmenkivi, Computational Methods for Intensity Models, Ph.D. thesis, Department of Computer Science, University of Helsinki, Series of Publications A-2001-2, 2001.

[12] M. Salmenkivi, H. Mannila, "Using Markov chain Monte Carlo and Dynamic programming for event sequence data," Knowledge and Information Systems. 7,267-288, 2004.



Mohammad Jaudet received an M.Sc. degree in Physics from Quaid-i-Azam University, Islamabad, Pakistan, in 1989. From 1991 to 2001, he remained involved in various R & D work as a faculty member in the Department of Electrical Engineering, PIEAS. He is currently a Ph.D. candidate at the Department of Electrical Engineering, PIEAS, Pakistan. His research interests include Telecommunication Systems Management, Fault Analysis & Predictions, Digital Signal Processing, and Neural Networks.

Naeem Iqbal holds Ph.D degree in Control & Automation from University of RENNES-I France. He is currently Associate Professor and Head of the Electrical Engineering Department at PIEAS and leads the Control and Automation group. His main research interests include linear and non-linear control systems, telecommunication systems, automation. He has authored numerous technical papers. Dr. Iqbal has keen interest in digital signal processing as well as in automation.



Nasir M. Mirza is currently Professor in the Department of Physics & Applied Mathematics and Head of the Computational Physics Group at the PIEAS. He has served as Dean of the Council for Graduate Studies at the PIEAS (2001-2006), and currently serving as the Head of the Department of Physics & Applied Mathematics at PIEAS. His main research interests include Modeling & Simulation, Computational Physics, and Stochastic Simulation. Professor Nasir Mirza has published numerous papers in various highly reputed international scientific and technical journals in several areas including Computational and Simulation Physics, Numerical Techniques and Applied Physics.

Sikander M. Mirza is a Professor in the Department of Physics & Applied Mathematics and member of the Computational Physics Group at the PIEAS. He is also member of the Computational Intelligence Group of the Department of Computer & Information Science at PIEAS. He has served as Head of the Physics & Applied Mathematics Department (2001-2006), and currently serving as Dean of Faculty of Applied Sciences at PIEAS with main research interest in Computational Physics, Modeling & Simulation and Advanced Numerical Techniques. He has published numerous papers in several areas including Numerical Methods, Computational & Simulation Physics and Applied Physics.

Amir Hussain is currently Senior Lecturer in the Department of Computing Science & Mathematics and Research Manager for European Commission IT&C Project. He holds visiting posts in several leading universities in Europe and Asia. With main research interests in modeling and control of complex systems, computational intelligence, telecommunications engineering, bio-informatics, and cognitive & computational neuroscience, he has authored over 80 technical papers and holds one patent. Dr. Hussain is serving as an Editor-in-Chief of the International J. of Natural and Artificial Intelligence Systems, associate editor of the International Journal of Robotics and Automation and of the International Journal of Neurocomputing. He has been invited keynote speaker at various international conferences. Also, he is full Member of the UK Higher Education Academy.