POINT PROCESS ESTIMATION DERIVED FROM STATISTICAL DESCRIPTION OF VEHICLE HEADWAYS

by

Arthur J. Dorsey

Thesis submitted to the Faculty of the Graduate School
of the University of Maryland in partial fulfillment
of the requirements for the degree of
Master of Science
1978

ABSTRACT

Title of Thesis: Point Process Estimation Derived From Statistical Description of Vehicle Headways

Arthur Joseph Dorsey, Master of Science, 1978

Thesis directed by: John S. Baras
Associate Professor
Department of Electrical Engineering

This thesis applies recent developments of the martingale approach to point processes to obtain a recursive nonlinear filter in the context of urban traffic. Beginning with a review of the statistical description of vehicle headways, a convex probability density of headways is proposed. Techniques are then discussed to determine the four parameters necessary to specify this density. Subsequently, a complete description of the interarrival times is given which incorporates the entire past statistics of an observed counting process and leads to the derivation of its local description. The results are then utilized to formulate and solve the disorder problem for the switch rate point process involved. The utility of these results to traffic estimation/detection problems is discussed and a series of evaluations are performed.

DEDICATION

This thesis is dedicated to my parents and my Lord through the following passage that served as a guide throughout these years.

> "When the perfect comes, the imperfect will pass away. When I was a child I used to talk like a child, think like a child, reason like a child. When I became a man I put childish ways aside. Now we see indistinctly, as in a mirror; then we shall see face to face. My knowledge is imperfect now; then I shall know even as I am known. There are in the end three things that last: faith, hope, and love, and the greatest of these is love".
>
> 1 Corinthians 13, 8-13.

ACKNOWLEDGEMENTS

I would like to gratefully acknowledge the encouragement, advice and guidance of my thesis advisor, Professor John S. Baras. In the early phase of this research, it was his innovative ideas that simulated further progress. Also, I would like to thank Professor William S. Levine for his valuable suggestions and insight during the evaluation phase of this work. The help of the following individuals during the course of this thesis is also acknowledged: William Roberts, Lester Lin, Lynn Runt, and Dr. Steven Cohen. These latter two individuals provided indispensable assistance with the UTCS-1 simulator. Special acknowledgement is given to IBM Manassas in providing me with a Leave of Absence to complete this degree. Finally, this thesis was done in conjuction with research supported by the U.S. Department of Transportation under contract DOT-OS-60134 and in part by the Computer Science Center of the University of Maryland.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. STATISTICAL DESCRIPTION OF VEHICLE HEADWAYS

1.1 Introduction

With the emphasis on increased fuel efficiency of vehicles, it is apparent that in urban traffic inefficient traffic light control results in increased fuel consumption. Consequently in more than 100 cities, there are plans to alleviate this problem by using computerized urban traffic control. For residents of the Washington, D.C. area the most recent development in this regard was the Urban Traffic Control System (UTCS) where detectorized network link data was processed by a central computer to update traffic light patterns. However, these traffic activated controllers using only current traffic volumes obtained only a marginal improvement over previously implemented systems based on time of day and historical data [1,2]. The limiting factor in its performance resulted from poor estimates of traffic volume, occupancy, queue length and average speed. Therefore improvements in traffic control require more effective filtering and prediction algorithms.

Current predictions and estimators model detector data as Markov diffusion processes, i.e. the output of a dynamical system driven by white noise. However, detector observations are not a continuous random process but rather constitute a mark point process. In particular, the observation process has highly localized events distributed randomly in a continum and can be characterized by:

i) a sequence of activation times $W_1, W_2, W_3 \ldots\ldots$ where $W_i < W_{i+1}$.

ii) an associated sequence of pulse widths characterizing vehicle velocity.

Realizing that improper modeling of the physical phenomenon effects the performance of the estimator, it is reasonable to expect that pre-whitening techniques and Kalman filtering estimation yields unsatisfactory results [2]. Therefore, the investigation considered here models the detector activations as a random point process.

By this approach, we consider the microscopic aspects of urban traffic in a network link. To be effective though, one must incorporate the statistics of an underlying process universal to all traffic conditions and traffic environments. It has been found that one such parameter is the headway between two vehicles [3, p. 20]. Because of the ease of acquiring headway measurenments from detector data and because of its implication to point process theory, it serves as the basic tool in our subsequent analysis.

Therefore the purpose of this thesis is to apply point process theory, derived from the statistical description of vehicle headways, in the development of an urban traffic estimator. In particular, our discussion can be divided into four areas:

1. development of a complete statistical description of vehicle headways.
2. development of an urban traffic estimator using the martingale approach to point processes.

3. investigation of the resulting filter to variation in parameters.

4. evaluation of the filter against simulated urban traffic.

In this chapter, we review previous headway probability density models, derived from frequency data. It is shown that a more complex, four paramenter model yields the best performance in all traffic conditions; a necessary requirement for any filter considered. The proposed headway model is extended by a series of reasonable assumptions to obtain a complete statistical description. In the next chapter, the rates of the observed point process are derived from this statistical description. (In this thesis, we shall not consider the associated mark process, vehicle velocity.) Utilizing a recent result in point process theory, a nonlinear filter of a traffic parameter queue length, is developed. The recursive estimator is shown to depend on both the historical data and current detector outputs to obtain this estimate. The filter is then evaluated against an urban traffic simulator, UTCS-1 whose validity is well-documented [47, 48]. In Chapter 3, the filter's sensitivity to parameter variations is studied. Since the filter is a differentiable function of three unknown parameters, small scale sensitivity analysis is performed on the conditional error variance. Parameter estimation techniques and associated confidence intervals are discussed. The results reveal the robustness of the estimate to parameter error.

1.2 Models of Headway Probability Density

Headway is a time and space measurement. At a fixed location, the interarrival time of two consecutive vehicles is defined as time headway, which we call simply headway. Time headway as obtained from presence detectors is simply the time difference of the leading edge of two consecutive pulses. The space headway which we call spacing is the distance from the front of one vehicle to the front of the following vehicle. To obtain space headway measurements, one can use aerial photogrammatic techniques which measure the location of vehicles at different instants in time. A generally accepted technique for obtaining space headway is to assume a uniform velocity and multiply the measured time headway by this velocity.

The inherent nature of headways implies some degree of interaction between vehicles. Since in general traffic vehicles are randomly placed in space and time, headway is a random variable. As Buckley noted [4 , p. 156]

> "No matter how homogeneous traffic conditions may be, headway observed from the roadside, even during a short period of observation can be extremely dissimilar."

The only satisfactory way to analyze headways is therefore the determination of the statistics of headway measurements.

For the remainder of this section, the probability density function for headways is discussed. It is conjectured that the underlying headway probability is of the same type for all different types of traffic (rural, freeway, urban, etc.) but the parameters may vary. The reader should pay close attention to the sampling technique and the criterion used to verify the results. The development follows earlier reviews done by Edie [5] and Gerlough et al [6].

One of the first statistical analyses of headways was done by Adams [7 , p. 122]. Observing the counting process of vehicles showed "under normal conditions, freely-flowing traffic correspondes very closely to a random series of events ....

(a) each event, for example , the moment of arrival of a a given point is completely independent of any other event.

(b) equal intervals of time are equally likely to contain equal numbers of events."

Subsequently, by counting the number of vehicles passing a point in every 10 second observation period, for flows ranging from 70 to 1400 vph, a Poisson density was observed:

$$Pr(N=k) = \frac{\alpha^k}{k!} e^{-\alpha} \qquad k = 0, 1, 2, \ldots \tag{1.2.1}$$

where α = mean number of vehicles expected in a given period. For such a counting process, the interval of time elapsing between arrivals obeys a negative exponential density:

$$p(h) = \begin{cases} \alpha e^{-\alpha h} & h \geq 0 \\ 0 & h < 0 \end{cases} \tag{1.2.2}$$

Therefore, by time sampling, the headway distribution for street traffic in London was described by a negative exponential density.

However, this model broke down when traffic was no longer freely flowing. Adams observed that a sudden increase or decrease in traffic, the presence of external control such as a police officer or a traffic light or the difficulty in passing other vehicles resulted in alternating periods of greater or smaller traffic flows. The result of such platooning caused the headways to be longer or shorter than that observed in normal freely-flowing traffic. Also

as traffic approached a saturation flow of 1000 vph, Adams noted that headways became shorter than expected.

This indicated that there are inherent features in the negative exponential density which limit its ability to describe headways in various traffic flows. Because of the assumed independence of events, knowledge of a vehicle's headway gives no information about the likely headway of the next vehicle. If such a hypothesis were true, the car-following models would be useless. Even in moderate volume traffic, a certain degree of interaction between successive vehicles is apparent. Although Adams conjectured that the negative exponential density was applicable for flows up to 1000 vph., his data was obtained for flow only up to 500 vph. A second shortcoming of the negative exponential occurs for very short headways. The frequency of short headways predicted by the negative exponential density is significantly greater than is actually observed (Fig. 1.1). Hence, a more elaborate probability density is required for headway in heavier traffic flows.

The work of Adams and his contemporaries has been extended in several directions. There has been extensive work concerning the Poisson counting process and the negative exponential density. A review of this work can be found in [8]. The independence of events has been utilized to define headways exceeding a certain critical headway value. The criteria for defining a platoon of traffic rest on this property of the negative exponential. The inadequacy for short headways has resulted in a series of refinements of the negative exponential, which are presently discussed.

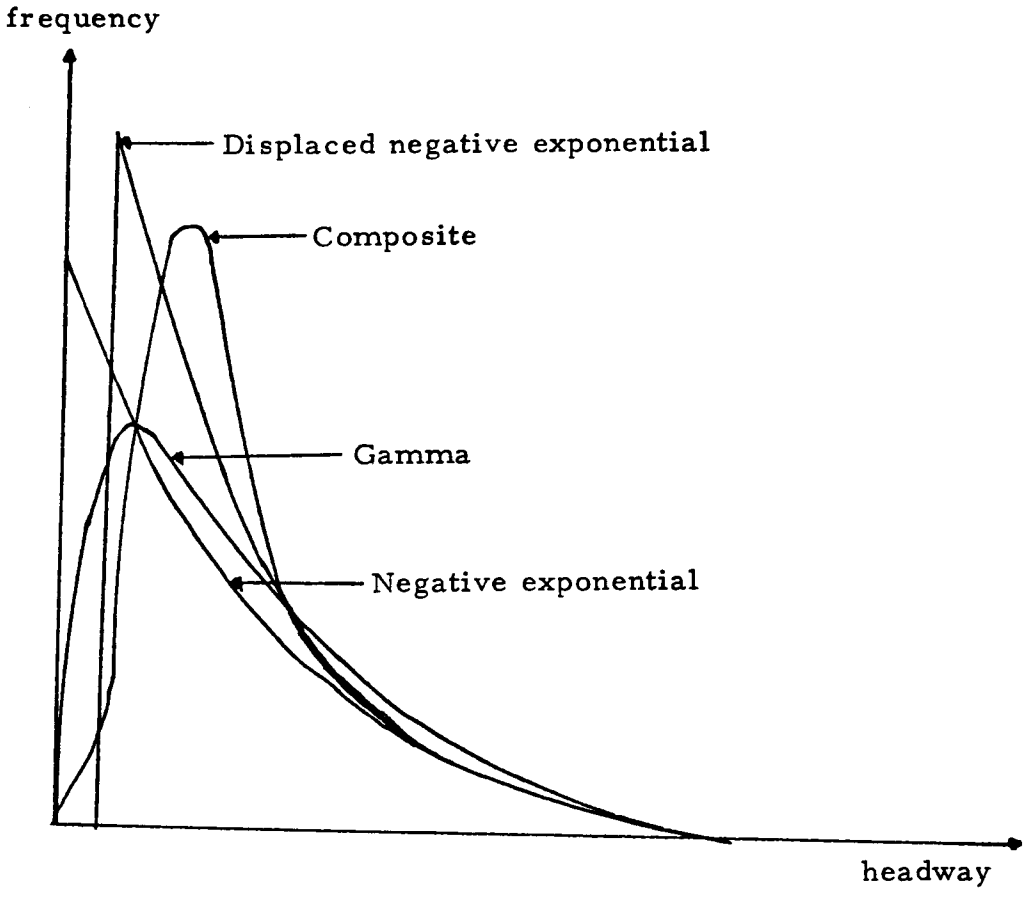


Fig. 1.1 Illustrating differences between various headway probability densities.

The deficiency near the origin of the negative exponential is corrected by displacing the density to the right (Fig. 1.1). From the definition of spacing, it can easily be seen that there always exists a minimum distance between two consecutive vehicles. Consequently, the displacement factor α is related to this minimum distance. The resulting displaced exponential density has the form:

$$p(h) = \begin{cases} \frac{1}{\beta} e^{-(h-\alpha)/\beta} & h \geq \alpha \\ 0 & h < \alpha \end{cases} \qquad (1.2.3)$$

Daou [9, p. 10], by linear regression techniques, determined this buffer distance for tunnel traffic to be 34 ft. Several researchers have obtained satisfactory agreement with the displaced negative exponential. For street traffic in San Francisco and Boston, Oliver [10, p. 810] claimed good agreement with the displaced negative exponential. However, since no criterion for goodness-of-fit was defined, it is assumed only a visual check was made.

While specializing his results for traffic through a signalized intersection, Newell [11, p. 366] compared the displaced negative exponential to the measured time headway relationship obtained by Clayton [12]. By generating a recursive relationship of the density of the displaced negative exponential, a first-order approximation yielded an equivalent formula to Clayton. Newell also found for light traffic, a uniform probability density of headway agreed with Clayton's formula. In other words, in light traffic the headway probability density is not significant. However, Newell [11, p. 369] claimed as traffic flow increased, headways become less uniformly distributed; thus also suggesting a dependency of headway statistics on traffic flow.

With the displaced negative exponential, one obtains a discontinuity for headway measurements, which is unrealistic. The family of Erlang densities which appear frequently in queuing theory, rectifies this dilemma.

$$p(h) = \begin{cases} \dfrac{h^{k-1}\, e^{-h/\beta}}{\beta^{k}(k-1)!} & h \geq 0 \qquad k=1, 2, 3 \ldots \\ 0 & h < 0 \end{cases} \tag{1.2.4}$$

This density represents the sum of k negative exponentials, (in classical queuing theory, the k equals the number of servers in a queue). For k=1, the density is a negative exponential and as k increases the density becomes more peaked, indicating a higher correlation between headways. At $k=\infty$, the headways become uniformly distributed. Hence, the parameter k is a measure of interaction between vehicles and is referred to as the "Erlang number". The advantage of this density is that for small headways, p(h) is small. An application of the queuing aspect of the Erlang distribution was studied by Ancker et al [13]. The headway probability density of vehicles with respect to their queue position was determined using a photo-detector located two feet after the stopline at a signalized intersection. Assuming no turns and no bottlenecks downstream, the headway data was grouped into pairs for eleven queue positions: (1, 2), (2, 3), (3, 4), ... (10, 11). Each pair was found to be statistically independent using five different criteria: contingency table, regression test, Corner test, Spearman rank correlation and correlation coefficient. The results strongly suggest that headway pairs and more importantly, individual headways <u>are mutually independent upon leaving a signalized intersection</u> [13, p. 350]. Furthermore, the 3-11 positions had the same variance and a shifted mean that

decreased monotonically with increasing queue position. By using a modified maximum likelihood estimate for the three parameters, the resulting Erlang density function for all positions agreed well with the empirical data. The χ^2 goodness-of-fit test showed non-significance at the 5% level.

An undesirable feature of the Erlang density is the integer values of the Erlang number. From the traffic standpoint, the degree of interaction k, is not discrete but rather continuous. Also, to determine integer values for a density's parameters by maximum likelihood techniques involves computational difficulties [13, p. 358]. Generalizing the Erlang density to include non-integer values of k results in the gamma probability density for describing headways. While studying rural traffic, Miller [14, p. 68] found individual vehicles had headways described moderately well (no criterion denoted) by the gamma density. Also an unpublished report by Tindal for $1 < k < 2$, showed bridge headway data yielded good curve fits to the gamma density [4, p. 162].

There are several common properties to the models of headway densities discussed thus far. First, no internal structure of traffic is assumed except that vehicles are random in time and space. Numerous studies [8] can substantiate this fact. Second, from the empirical data, the headway frequency function shows a distinct peak; headways exceeding this value tend to follow a negative exponential. From this observation, a close examination shows all densities discussed are special cases of the Pearson Type III probability density:

$$p(h) = \begin{cases} \frac{1}{\beta\,\Gamma(k)} \left(\frac{h-\alpha}{\beta} \right)^{k-1} e^{-(h-\alpha)/\beta} & , h \geq \alpha \\ 0 & , h < \alpha \end{cases} \qquad (1.2.5)$$

where $\Gamma(k)$ is the gamma function. Third, the differences in these densities clearly exist primarily at shorter headways. This suggests that we may consider a model for headways composed of two sub-populations defined by a composite density:

$$p(h) = \psi\, p_1(h) + (1-\psi)\, p_2(h) , \qquad (1.2.6)$$

where

$p_1(h)$ = density describing short headways

$p_2(h)$ = displaced negative exponential for longer headways

ψ = degree of interaction

Since headway is dependent on traffic flow, the degree of interaction incorporates this dependency. For example for light traffic, ψ equals zero yielding a composite density that is a displaced negative exponential.

By the composite density model, a certain structure of traffic is assumed and the emphasis is placed on defining the density for short headways. The first theoretical work in this direction was done by Schuhl [15,16]. By assuming the spacing between successive vehicles consisted of two subsets, each having a distinct mean value and each obeying some Poisson-type density, a composite negative-exponential was obtained. The justification for such a density rested on geometrical arguments which divided traffic into constrained vehicles and free-flowing vehicles. A free-flowing vehicle is defined as one that can travel without modifying its desired time-space trajectory. Noting that constrained vehicles had some minimum

spacing. Schuhl incorporated the displacement factor α in the composite density:

$$p(h) = \psi\, e^{-d_1(h-\alpha)} + (1-\psi)\, e^{-d_2 h} \qquad (1.2.7)$$

To validate these results, Kell [17] used 589 headway measurements from an urban street with flows ranging from 100 to 1200 vph. By segmenting the data into 18 constant flow groups, the χ^2 goodness-of-fit test was performed with satisfactory results (no significance level noted). Similar results were obtained by Sword [18].

Following the development of Schuhl, Buckley [4] considered the Gaussian density for shorter headways:

$$p_1(h) = \frac{e^{-h^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}} \quad , \quad p_2(h) \Rightarrow \begin{cases} \text{convolution of shifted} \\ \text{Gaussian density and} \\ \text{negative exponential} \\ \text{density} \end{cases} \qquad (1.2.8)$$

where σ^2 is the variance of headways. Physically, if one assumes a driver attempts to maintain some desired spacing, his ability to track the leading vehicle could be described by such a density, $p_1(h)$. Since earlier work showed headways are dependent on traffic flow, Buckley segmented freeway traffic data into several minute flow groups and analyzed the 33, 34 and 35 vpm headways. For the 607 measurements, a comparison was made of the curve fits of five densities: negative and displaced negative exponential, gamma, Pearson Type III and the composite semi-Poisson Gaussian given in (1.2.8). By the method of moments, the parameter estimates were obtained. The results showed that the composite semi-Poisson Gaussian density yielded the best χ^2 goodness-of-fit.

Unfortunately, this composite density assigned a probability to negative headway values and numerical techniques were required to obtain estimates of the four parameters.

However following a suggestion by Buckley, Dawson et al [19] considered the displaced Erlang density for describing short headways:

$$p_1(h) = \begin{cases} \frac{1}{\beta\,(k-1)!}\,(h-\alpha)^{k-1}\,e^{-(h-\alpha)/\beta} & , \; h \geq \alpha \quad k=1,2,3,\ldots \\ 0 & , \; h < \alpha \end{cases} \qquad (1.2.9)$$

As noted earlier, by varying the paramter, k of this density, headways from the most uniform to the most random can be described. Headway data was obtained from the 1965 Highway Capacity Manual [20] and a Purdue University study [18]. The measurements were stratified by flow, ranging from 100 to 1050 vph in 100 vph increments. By plotting the cumulative frequency function as probabilistic log vs time, the flow classes had nearly parallel curves. However because of poor flow monitoring, several curves in the Purdue data intersected. The parameter estimates were obtained by graphical techniques (for rough values) which were refined by non-linear least-squares methods. The comparison of the empirical and theoretical densities was done by determining the multiple correlation coefficient which varied from 0.9959 to 0.9998 (exact agreement yields 1.0).

As a continuation of his earlier work, Buckley [21] considered the composite semi-Poisson gamma density:

$$p(h) = \psi\, p_1(h) + (1-\psi)\, p_2(h) \qquad (1.2.10)$$

with $p_1(h)$= gamma density

$p_2(h)$= convolution of shifted gamma density and negative exponential density

Using 10,000 headway measurements from free-flowing freeway traffic, seven different flow classes from 1 to 30 vpm, in 5 vpm increments, were analyzed. The following densities were considered: negative and displaced negative exponential, gamma and displaced gamma, composite semi-Poisson gamma and composite semi-Poisson Gaussian with parameter estimates obtained by the method of moments. A comparison of these densities showed that the semi-Poisson gamma and the semi-Poisson Gaussian yielded the best χ^2 goodness-of-fit curves over all flow classes (significance level - 5%). It was observed that all densities performed best for intermediate headway values. Only the displaced gamma, the semi-Poisson Gaussian and the semi-Poisson gamma densities exhibited good fits for high flows [21, p. 118]. From these results, one can conclude that composite densities describe headways over all traffic flows better than Pearson Type III densities.

It can be seen that by strengthening the assumptions on the structure of traffic flow, we obtain better agreement with empirical data. Under the intuitive assumption of random vehicles, the resulting headway densities were from the Pearson Type III family. By insisting on two subpopulations, the composite density yielded better agreement between the theoretical and empirical data for higher traffic flow. As an extension, one can consider traffic as being composed of random platoons. A platoon is defined as a group of consecutive vehicles travelling at similar velocity separated by relatively short gaps.

In order to validate any model resulting from such an assumption,

criteria that can be effectively applied to experimental data for separating platoons are necessary. A study of random platoons was done by Miller [14] on rural traffic in Sweden and England using time series data containing velocity and counting information. The resulting headway frequency function showed that the 4-6 second range fitted a negative exponential. Also it was observed that the 6-8 second range contained a higher than expected number of occurrences. A close examination revealed that vehicles in this range had a slightly higher velocity than the leading vehicle. Thus, assuming a displaced negative exponential between gaps, Miller [14 , p. 67] defined vehicles with relative velocities less than 6 mph or with headways below 6 seconds as being in the same platoon and those exceeding these limits as being in different platoons. Given this criterion, and unless traffic density was too high to permit gaps in the traffic stream, a test was performed to insure independence of platoons. Grouping consecutive counts into sets of four, eight, etc. vehicles and analyzing the variance of each pair by Pacey's test, it was found that only 12 of 182 pairs exceeded the 5% significance level. Applying the same test to individual vehicles, however yielded a significance level of 45%. Thus, there is strong evidence to suggest that platoons are independent and follow an exponential density.

By assuming a random platoon model of traffic, one can study properties of the traffic structure. For evaluating the model a significant parameter becomes the criterion for determining whether vehicles are within the same platoon. Also, since platoons are

random, the critical part of the model becomes the model of headway density within a platoon. This development is similar to that in the composite density model. The length, formation and dispersion of a platoon are also of interest. Finally, fine structural points such as the effect of the platoon position and platoon length on the headway density can be studied.

A study was performed by Pahl and Sands [22] to determine whether headway was a valid parameter to differentiate platoons. Using velocity and counting time series data from an urban freeway for flows varying from 600 to 2300 vph, a vehicle interaction process was defined:

i) process of a vehicle slowing down when it approaches a slower vehicle, resulting in a decrease in headway and speed of the approaching vehicle.

ii) process of vehicle queuing where relative speed oscillates around zero and headway varies within small limits.

Using a similar approach to Buckley, over 10,000 measurements were divided into constant flow groups. The relative velocity between vehicles was found to be statistically dependent. However by conditioning the relative velocity by headway, it was determined that relative vehicle velocities were statistically independent when a certain critical headway was exceeded. Similarly, for smaller headway values, the relative velocities became dependent. These results were validated by a modified χ^2 test (significance level - 1%) over all flow classes. It was observed that the critical headway value decreased as flow increased. Therefore, it is concluded that headway is a valid parameter for distinguishing platoons.

Since headway measurements differentiate platoons, they can be incorporated into the platoon criterion (similar to Miller's definition). The platoon criterion is used to test whether vehicles are within the same platoon. Several different criteria have been developed. Edie et al [23] defined vehicles in the same platoon under the following conditions:

i) if space headway is less than 200 ft. for speed 40 ft./sec. or greater.

ii) if space headway is less than 170 ft. for speed 35 to 40 ft./sec.

iii) if space headway is less than 150 ft. for speed less than 35 ft./sec.

A different approach was taken by Athol [24] who defined a platoon as the stable portion of traffic throughout the entire spectrum of traffic behavior and its complement as a group. Then by segmenting data into platoons and groups using different headway definitions, the interarrival time of platoons was described by a negative exponential for a headway definition of 2.1 seconds [24]. In other words, headways of less than 2.1 seconds represented vehicles in the same platoon. Other platoon criteria were formulated by Greenshields [25] and Underwood [26].

Once the platoon criterion is specified, the headway probability density within a platoon can be experimentally determined. From velocity and counting data from tunnel traffic, Daou [9] obtained a lognormal density of headway within platoons:

$$p(h) = \begin{cases} \dfrac{1}{\sigma h \sqrt{2\pi}} \exp\left(- \dfrac{(\ell n\, h - \mu)^2}{2\sigma^2} \right) & , \; h \geq 0 \\ 0 & , \; h < 0 \end{cases} \qquad (1.2.11)$$

where μ and σ^2 are the mean and variance of $\ell n\, h$. Using Greenshields' 200 ft. spacing between vehicles as the platoon criterion, the data was stratified by velocity into seven classes from 15 to 50 ft./sec., in 5 ft./sec. increments. The χ^2 goodness-of-fit test was performed and a non-significance at the 5% level was achieved for the lognormal hypothesis. An interesting feature of Daou's work was that the variance to mean ratio overall velocity classes was approximately constant [9 , p. 9]. More recently, Tolle [27] utilized a shifted lognormal density to describe headways:

$$p(h) = \begin{cases} \dfrac{1}{\sigma(h-\alpha) \sqrt{2\pi}} \exp\left(- \dfrac{(\ell n\,(h-\alpha) - \mu)^2}{2\sigma^2} \right) & , \; h \geq \alpha \geq 0 \\ 0 & , \; h < \alpha \end{cases} \qquad (1.2.12)$$

Space headway measurements of freeway traffic were segmented into 11 flow classes and using the property that $\ell n\, h$ is gaussian, maximum likelihood parameter estimates were obtained. The Kolmogorov-Smirnov (KS) goodness-of-fit test showed good agreement with the lognormal hypothesis with $\alpha = 0.3$ (significance level - 5%). There are various justification for these findings about the lognormal density. The primary reason is that multiplicative, independent identically distributed errors by various drivers attempting to follow each other combine to give a lognormal density.

One next needs to investigate the dependency of the headway

density on platoon length and relative position in the platoon. Athol [24], while studying freeway traffic, found the headway density within a platoon independent of platoon length by comparing the relative frequency function of a 2, 4 and 6-vehicle platoon. Also, by grouping headways by queue position and computing the mean and variance of each position, a significance test showed the mean headway was not significantly different for relative position in the platoon.

From the discussion above, we can now formulate a statistical model of headway. We first observe that the model must accomodate wide traffic flow conditions. Even as early as Adams, the headway model proposed broke down with heavy volume. We observed that the model was not crucial for light traffic but as traffic flow increased, the model became more significant. Only in the comparison tests done by Buckley [4 , 21] did the merits of the composite model over other Pearson Type III densities become apparent in high traffic flows. Therefore, the statistical model is of the composite structure to accomodate wide traffic flow. A second feature of any statistical model is its ability to fit the empirical data. From the numerous densities considered using the χ^2 or KS curve fitness test, the shorter headways were more influential on the model chosen. With regard to the composite models, only the gamma and the lognormal densities provide good agreement with the following headways [21, p. 131]. Since the former can be translated into the latter [28], and because the lognormal has certain satisfying statistical and physical properties we assume this density for describing the following headways.

Finally, the statistical model must incorporate the structural properties inherent in urban traffic. Clearly traffic lights induce a periodicity in traffic flow and cause a decoupling of vehicle platoons. This results in a certain independence in the interplatoon gaps and these gaps canbe described by an exponential density [14]. The model must incorporate the independence of headways with respect to platoon position and platoon length and must merge the following and non-following headway processes. Hence by considering a non-following headway as resulting from the sum of two independent variables, one with lognormal density and the other with a negative exponential (to represent the gap between platoons), one can embody all these features. In summary then, we propose the following statistical model of headways:

$$p(h) = \psi \; p_f(h) + (1-\psi) \; p_{nf}(h) \qquad , \; h \geq 0 \tag{1.2.13}$$

where

$$p_f(h) = \frac{1}{\sigma h \sqrt{2\pi}} \exp\left(- \frac{(\ell n\, h - \mu)^2}{2\sigma^2} \right) = g(h) \qquad , \; h \geq 0$$

$$p_{nf}(h) = \lambda \exp(-\lambda h) \int_0^h g(x) \exp(\lambda x) \, dx \qquad , \qquad , \; h \geq 0$$

where ψ is the perentage of following vehicles in the overall traffic flow and $1/\lambda$ is the mean interbunch (or interplatoon) gap. According to this model, each following headway within a platoon belongs to a population with a lognormal probability density with the same parameter μ and σ, irrespective of the length of the platoon or vehicle position in the platoon.

From the filtering standpoint there are several reasons for choosing this model. The parameters introduced are physically reasonable and are important parameters for filtering/prediction. The model can accomodate all traffic conditions (light, moderate, heavy) and is valid for practically all ranges of traffic flow and speed. The densities involved imply underlying stochastic processes that can be completely described by a finite number of moments (at most two). This is an important fact for the development of simple but effective filter/predictors.

The model formulated rests upon two assumptions:

i) a following headway has a lognormal density

ii) interplatoon gaps are independent and exponentially distributed.

To validate these assumptions and therefore the model, Branston [29] made a comparison of the proposed model (1.2.13) to the composite semi-Poisson Gaussian and the composite semi-Poisson gamma. The space headway measurements were stratified into constant flow groups and parameter estimates were obtained by a combination of the method of moments and the minimum chi-squared. The resulting χ^2 goodness-of-fit test (significance level - 5%) showed that the model considered gave the best overall performance. Moreover, the parameter variations to traffic flow were investigated. It was observed that μ and σ both decreased as traffic flow increased, although often the variation was small enough to allow a very good fit with constant μ and σ. The dependence on speed was more drastic. Both μ and σ tended to increase with speed, but real data indicated that μ varied widely for the same speed for different traffic locations

and/or times, while σ did not show a similar wide variation. These established facts are encouraging, and actually imply that a periodic estimation of μ and σ is likely to be an effective way of obtaining values of μ and σ from on-line data. This adaptation of the parameters can be done at a much slower pace than the actual filter/predictor.

Finally for the determination of ψ and λ and their relation to traffic flow rate, the following models were found to be in good agreement with real data [29]. Let γ denote the traffic intensity

$$\gamma = \frac{\text{mean following headway}}{\text{mean headway}} = \frac{\exp(\mu + \frac{1}{2}\sigma^2)}{E\{h\}} \qquad (1.2.14)$$

and λ^* the flow rate. Then

$$\lambda = \lambda^* - \tfrac{1}{2}\lambda^{*3/2}$$
$$\psi = \gamma - \tfrac{1}{2}(\gamma-1)\lambda^{*3/2} \qquad (1.2.15)$$

Although, the above formulas are the results of curve fitting real data from specific traffic locations, they can be used as a first approximation to the relations between these parameters, because experimental evidence indicates low sensitivity to traffic location. In conclusion the model proposed above (1.2.13) provides an acceptable model for headway distribution with many desirable properties.

1.3 Complete Model for Headway Statistics

The statistical model of headway given in the previous section does not provide a complete description of the stochastic process. Headway statistics (or interarrival time statistics) need higher order probability density functions for their complete description. This is due, as was observed before, to the considerable correlation between successive headways. On the other hand from point process theory, we know that interarrival time statistics completely characterize the process and in particular can be used in determining the rate of the process, which plays a central role in estimation problems.

The extension of the statistical model of headways to obtain a complete description develops under two assumptions:

(i) the following and non-following headway processes are independent.

(ii) there is only a dependency between consecutive pairs of following headways.

Assumption (i) is reasonable since the statistical model for the convex combination of two probability densities is obtained by assuming we have two processes, one consisting of nonfollowing headways and the other consisting of following headways, together with the relation

$$\begin{aligned} &\Pr\{\text{headway is following}\} = \psi \\ &\Pr\{\text{headway is nonfollowing}\} = 1-\psi \end{aligned} \tag{1.3.1}$$

Assumption (ii) means that a driver reacts only to the vehicle directly ahead of him, i.e. any other vehicle ahead of his does not affect his desired trajectory. Since following headways are independent

of platoon length and platoon position, this is also a fair assumption. The remainder of this section utilizes these assumptions to characterize the multidimensional statistics for each process separately and thus obtain a complete model. In particular, the joint densities of each of the following cases are treated separately: two consecutive following headways, two nonfollowing headways, a following headway preceded by a nonfollowing headway and vice versa.

For the following headway process, the first order probability density is lognormal and as observed by Buckley [4 , p. 171], there is negative correlation between successive headways. Furthermore, it is reasonable to assume that the correlation between two non-consecutive following headways is negligible. This is an approximation but we believe it is well justified. We thus use a process with memory one to describe the following headway process:

$$E\{(h_i - E\{h_i\})(h_{i+k} - E\{h_{i+k}\})\} = \begin{cases} \rho & , k=1 \\ 0 & , k \geq 2 \end{cases} \qquad (1.3.2)$$

where $\rho \leq 0$ usually and the index refers to the order of occurence of the headways. This is the same as saying that the headway sequence is a Markov sequence. Then, a reasonable model for the following headway sequence is described fully by a lognormal Markov sequence. That is

a) the first order probability density is given by

$$h_i(\xi) = \begin{cases} \dfrac{1}{\xi\sigma\sqrt{2\pi}} \exp\left(-\dfrac{(\ell n\,\xi - \mu)^2}{2\sigma^2}\right) & , \xi \geq 0 \\ 0 & , \xi < 0 \end{cases} \qquad (1.3.3)$$

where $\mu = E\{h_i\}$, $\sigma^2 = E\{(\ell n\, h_i - \mu)^2\}$

b) the transition probability density is given by:

$$p_{h_{i+1}|h_i}(\zeta|\xi) = \begin{cases} \dfrac{1}{\zeta\sigma\sqrt{2\pi(1-r^2)}} \exp\left(-\dfrac{(\ell n\,\zeta-\mu-r\,(\ell n\,\xi-\mu))^2}{2\sigma^2(1-r^2)}\right) \\ \qquad\qquad , \quad \zeta \geq 0, \xi \geq 0 \\ 0 \qquad \text{elsewhere} \end{cases} \tag{1.3.4}$$

where

$$r = \frac{E\{(\ell n\, h_i-\mu)(\ell n\, h_{i+1}-\mu)\}}{\sigma^2} \tag{1.3.5}$$

and is related to ρ of (1.3.2) via

$$r = \frac{1}{\sigma^2}\,\ell n\,[1+\rho\,(\exp\sigma^2-1)] \tag{1.3.6}$$

From elementary facts about Markov processes, the following headway sequence is completely characterized by (1.3.3) (1.3.4).

Obviously we could have modified our assumption(ii) on consecutive pairs of following headways and used a higher order lognormal Markov sequence to model following headways. The advantages however of the more complex resulting model are at least ambiguous at this stage. The new parameter ρ introduced depends among other things on traffic volume as observed by Buckley [4], having larger absolute values for higher volumes and becoming zero for very light traffic. Also if the average headway is large there is obviously no correlation and $\rho = 0$. In view of (1.3.5), (1.3.6) this parameter can be easily (and is usually) estimated from data.

For the non-following process, each nonfollowing headway is the sum of a following headway and an exponential gap which are mutually independent. We can obtain a complete statistical description of the nonfollowing headway sequence. By definition,

a nonfollowing headway is typically larger than a following headway and corresponds to lighter traffic. But in light traffic headways are independent, so the nonfollowing headways are an independent, identically distributed sequence of random variables with common density:

$$p_{h_i}(\xi) = \begin{cases} \lambda \exp(-\lambda \xi) \int_0^{\xi} g(x) \exp(\lambda x)\, dx & , \xi \geq 0 \\ 0 & , \xi < 0 \end{cases} \tag{1.3.7}$$

where $g(\cdot)$ is given by (1.2.13). If

$$h_i = h_i' + t_i \quad , \; i = 1, 2, 3, \ldots$$

represents the decomposition of each nonfollowing headway to a lognormal and negative exponential random variable, respectively then by the independence of $\{h_i'\}$ and $\{t_i\}$

$$\begin{aligned} p_{h_i, h_j}(\xi_1, \xi_2) &= \int_0^{\xi_1} \int_0^{\xi_2} p_e(\xi_1 - x)\, p_e(\xi_2 - y)\, p_{h_i', h_j'}(x, y)\, dx\, dy \\ &\qquad , \; \xi_1 \geq 0, \; \xi_2 \geq 0 \\ &= p_{h_i}(\xi_1)\, p_{h_j}(\xi_2) \end{aligned} \tag{1.3.8}$$

where the last equality follows from the independence of $\{h_i\}$. But

$$\begin{aligned} p_{h_i}(\xi_1)\, p_{h_j}(\xi_2) &= \int_0^{\xi_1} \int_0^{\xi_2} p_e(\xi_1 - x)\, p_{h_i'}(x)\, p_e(\xi_2 - y)\, p_{h_j'}(y)\, dx\, dy \\ &\qquad , \; \xi_1 \geq 0, \xi_2 \geq 0 \end{aligned} \tag{1.3.9}$$

where p_e = negative exponential density. In other words, the following components of the nonfollowing headway sequence are independent, also.

From the completed descriptions of these two processes, the entire statistical model (1.2.13) of headways can be obtained from

the transition probability density:

$$p_{h_{i+1}|h_i}(\zeta|\xi) = \frac{p_{h_{i+1},h_i}(\zeta,\xi)}{p_{h_i}(\xi)} \tag{1.3.10}$$

Let

$p_{f,f}$ = joint probability density of two consecutive following headways

$p_{f,nf}$ = joint probability density of a nonfollowing headway preceded by a following headway

$p_{nf,f}$ = joint probability density of a following headway preceded by a nonfollowing headway

$p_{nf,nf}$ = joint probability density of two consecutive nonfollowing headways.

Then we have that

$$p_{f,f}(\zeta,\xi) = p_{h_{i+1}|h_i}(\zeta|\xi)\, p_{h_i}(\xi)$$

$$p_{nf,nf}(\zeta,\xi) = p_{h_{i+1}}(\zeta)\, p_{h_i}(\xi)$$

from (1.3.3), (1.3.4) and (1.3.8) respectively. The two other densities can be obtained from the definition of a nonfollowing headway and (1.2.13)

$$p_{f,nf}(\zeta,\xi) = \lambda \exp(-\lambda\xi) \int_0^{\xi} \exp(\lambda x)\, p_{f,f}(\zeta,x)\, dx \tag{1.3.11}$$

$$p_{nf,f}(\zeta,\xi) = \lambda \exp(-\lambda\zeta) \int_0^{\zeta} \exp(\lambda x)\, p_{f,f}(x,\xi)\, dx \tag{1.3.12}$$

Then the transition density is:

$$p_{h_{i+1}|h_i}(\zeta|\xi) = \{\psi^2\, p_{f,f}(\zeta,\xi) + \psi(1-\psi)\, p_{f,nf}(\zeta,\xi) + \psi(1-\psi)\, p_{nf,f}(\zeta,\xi) + (1-\psi)^2\, p_{nf,nf}(\zeta,\xi)\} \,/\, \{\psi\, p_f(\xi) + (1-\psi)\, p_{nf}(\xi)\}$$
$$, \zeta \geq 0, \quad \xi \geq 0 \tag{1.3.13}$$

where p_{h_i} is provided by (1.2.13). We now have a complete

description of the headway (interarrival time) statistics and can completely characterize the associated point process [30].

As a consequence of the headway model adopted, a model can now be developed for urban traffic flow. Each link is divided in sections in accordance with the detectorization of the link. For each section of the link the input and output traffic flows will have headway distributions as described above. Notice that the headway distribution model can vary (and it should) from lane to lane. The required parameters of the model will be estimated at appropriate intervals from actual data, or from historical data as required. The effect of the link will be to alter the parameter values as traffic moves downstream.

The versatility of the proposed model is now briefly indicated along with the ability to incorporate all desired situations. If the next downstream section provides greater congestion than the current section of the roadway, this will appear as an increase in ψ for the next section, followed by a decrease in μ and σ. Often, this change in μ and σ will not be necessary. In case the current section is in front of a traffic light which just turned red, then the incoming flow parameters will be adjusted so that ψ will increase and μ and σ will decrease gradually (according to time required to form the queue). In addition, for properly located detectors, the number of cars in this particular section can serve as a measure of the queue in front of the red traffic light. Similarly when the light turns green, the ψ will decrease and μ and σ will increase accordingly to reflect the transition from stopped queue to the level of traffic flow. By appropriate variation of ψ, one can thus create platoons

or disperse platoons and thus realistically emulate traffic flow. To complete the model, a distribution for the mark of the underlying point process, that is pulse length, is also needed. As a first approximation however, this is omitted here. To summarize, the model requires the determination of four (4) parameters for each section, namely $\psi, \lambda, \mu, \sigma$, which will depend on traffic flow, speed time and location, in general. This provides a "local description" of the underlying point process which is realistic and allows the use of the modern theory [31] to obtain filter/ predictors.

2. DISORDER PROBLEM

2.1 Introduction

Having developed a complete statistical description of interarrival times (headways) in the preceding chapter, we now incorporate it into the framework of point process theory to answer an important traffic flow question: By using only detector data, can one determine when a change has occurred in the traffic flow? Two commonly occurring phenomena bring this question into perspective:

(i) By observing the detector activation times at a location on a freeway, determine whether or not an accident has occurred downstream.

(ii) At a detector located downstream from a traffic light, determine when the last vehicle in a queue has passed (or when a platoon starts and ends).

The motivation for considering (i) is apparent since in freeway traffic, by knowing the time and location of an accident enables emergency vehicles to be dispatched to rectify the blockage. Recent research in this area [32] attempts to solve the "incident detection problem" by generalized likelihood ratio methods. We feel that a more natural modelling can be obtained within the context of the statistical model proposed. As for situation (ii), many urban traffic systems are being considered using local computer-oriented controllers to estimate the queues within links. One shortcoming of these queue estimators is that they accumulate errors from previous light cycles because of the estimator's inability to know exactly whether a vehicle gets caught by an amber-red light change. By utilizing a detector downstream from the traffic light and determining the last vehicle in the queue, the queue estimator can be re-initialized each cycle to

correct this cumulative error.

The answer to either question posed can be related to different interpretations of the mixing parameter ψ in model (1.3.13). In particular, for slower estimation of traffic flow parameters, ψ can be used to estimate gross traffic patterns. For situation (i), ψ can be interpreted as the probability of a particular headway being a following headway and when the estimate of ψ exceeds a certain threshold, an emergency vehicle can be dispatched. On the other hand, if the model is intended for use in urban nets with small link lengths and for sections where queues occur, in order to compute estimates or predictions for short time intervals, a different interpretation for ψ is needed. In such cases, it is crucial to model the periodic formation and propagation of platoons or queues as modulated by traffic lights. Then ψ is more abrupt and can be modelled as a time function with value 1 corresponding to passing of a platoon or a queue discharge and 0 corresponding to nonfollowing, freely-flowing traffic as shown in the figure below. This latter interpretation of ψ is developed in this chapter.

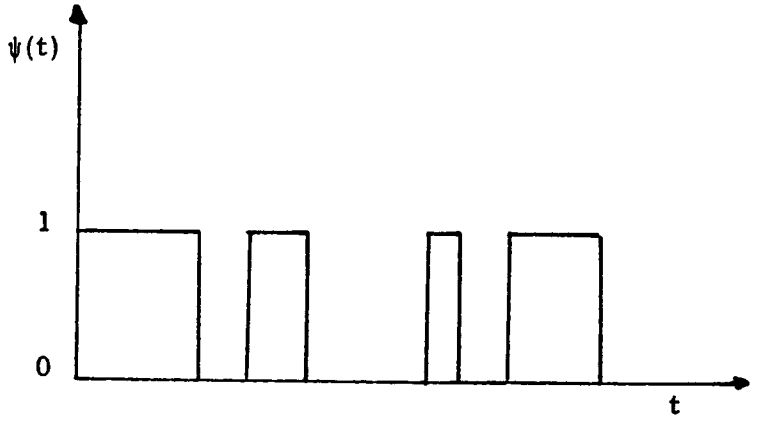


Figure 2.1. Illustrating time variation of ψ for accurate platoon or queue modelling.

In more precise terms, the above problem can be posed: given a stochastic point process regulated by a probability measure P_o until some random time T, and then governed by a different probability measure P_1, estimate the switch time T using the observations. The solution is based on the martingale approach of point processes and results in a nonlinear recursive filter representation. By a series of simplifying assumptions, a closed form estimator is obtained. The discussion which follows converts the statistical description of headways into a local description of a counting process, formulates the "disorder problem" by the martingale approach, specializes the results for several cases and evaluates the resulting estimator against simulated urban traffic.

2.2 Point Processes and their rates

A point process is a mathematical model for describing a physical phenomenon such that at random times, highly localized events occur. A point or jump process, as it is commonly called, is stochastic with the property that each sample path is piecewise constant, right-continuous and having usually a finite number of discontinuities on any finite interval. A jump process having only jumps of unit magnitude is called a counting process. A common example of such a process is the homogeneous Poisson process where interarrival times are independent, identically distributed random variables with a negative exponential probability density.

One chooses the point process framework over alternative techniques on the basis of agreement to the physical phenomenon and for the mathematical structure it affords the particular problem.

As discussed in the review, as early as 1939, Adams [7] modelled vehicle flow by a homogeneous Poisson process. The reason that more sophisticated modelling was not utilized was the lack of extensive mathematical treatment of statistical inference based on point process observations. On the other hand, the treatment of estimation problems based on continuous observations was heavily developed. In particular for certain cases, recursive estimators were developed (Kalman filter, extended Kalman filter) which were applied in traffic flow problems [33, 34, 35]. Recently, embedding point processes into martigale theory resulted in both a rich mathematical structure and in several generalizations of earlier results [31, 36-38].

Underlying the martingale approach in point process theory is the fact that every stochastic process $(N_t, \mathfrak{F}_t, P)$ can be decomposed into a predictable process and a local martingale; analogous to signal in white, additive Gaussian noise. The predictable process, called the local description has the following intuitive interpretation: the probability that N_t has a jump in $(t, t+\Delta t]$ given the entire past information $\mathfrak{F}_t$, depends only on the past observations of N_t up to time t and is denoted by:

$$\Delta\tilde{N}_t = \Pr\{N_{t+\Delta t} - N_t = 1 \mid \mathfrak{F}_t\} \qquad (2.2.1)$$

Moreover, if $\Delta\tilde{N}_t$ is differentiable with respect to time, then the associated counting process is called a regular point process [39, p. 548] and the corresponding derivative, λ_t is called the rate of the process. The differentiability of $\Delta\tilde{N}_t$ depends on the conditional orderliness of the process [30, p. 50]. Qualitatively conditional orderliness means

that the probability of two or more occurrences in an interval can be made arbitrarily small by choosing Δt sufficiently small. Within the context of Bremaud [36] and other early researchers [40] for a counting process N_t, a square integrable martingale was defined:

$$w_t = N_t - \int_0^t \lambda_s \, ds \tag{2.2.2}$$

where the integral is the local description. Heuristically these two descriptions are equivalent:

$$E\left[N_{t+\Delta t} - N_t \,|\, \mathcal{B}_t\right] \approx \lambda_t(\omega)\, \Delta t$$

$$E\left[N_{t+\Delta t} - N_t \,|\, \mathcal{B}_t\right] \approx E\left[\lambda_t(\omega)\, \Delta t \,|\, \mathcal{B}_t\right] = \lambda_t(\omega)\, \Delta t$$

where $\mathcal{B}_t$ is a σ-algebra containing $\mathcal{F}_t$ and $\lambda_t(\omega)$ is $\mathcal{B}_t$ measurable. For the remainder of this text, we shall assume $\Delta\tilde{N}_t$ is differentiable.

A homogeneous Poisson process can be described by its occurrence time statistics $\{w_n\}$ as:

$$\Pr\left\{N_{t,\, t+\Delta t} = n\right\} = \frac{(\lambda\, \Delta t)^n}{n!} \exp\left(-\lambda\, \Delta t\right)$$

or can equivalently be described by its interarrival time statistics, $\{t_n\}$. In general, jump processes can be completely described by their occurrence time statistics or their interarrival time statistics [31]. We shall show that the rate defined above can be explicitly written in terms of their interarrival times [also see 30, p. 246]. The rate associated with the predictable process in (2.2.1) is explicitly defined as:

$$\lambda_t = \lambda_t(N_t;\, W_1, W_2, \ldots, W_{N_t})$$

$$= \lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} \Pr\left\{N_{t+\Delta t} - N_t = 1 \,|\, N_t = n, w_1 = W_1, w_2 = W_2, \ldots, w_n = W_n\right\} \tag{2.2.3}$$

Since $\{N_{t,\,t+\Delta t}=1\}$ and $\{t< w_{n+1}\leq t+\Delta t\}$ are equivalent events,

$$\Pr\{N_{t+\Delta t}-N_t=1 \mid N_t=n, w_1=W_1, w_2=W_2, \ldots, w_n=W_n\}$$

$$= \Pr\{t< w_{n+1}\leq t+\Delta t \mid N_t=n, w_1=W_1, w_2=W_2, \ldots, w_n=W_n\}$$

$$= \frac{\Pr\{t< w_{n+1}\leq t+\Delta t \mid w_1=W_1, w_2=W_2, \ldots, w_n=W_n\}}{\Pr\{t< w_{n+1} \mid w_1=W_1, w_2=W_2, \ldots, w_n=W_n\}}$$

where the last equality results because for $t> W_n$, $\{N_t=n\}$ and $\{w_{n+1}> t\}$ are equivalent events. Then taking the limit as $\Delta t \downarrow 0$ yields:

$$\lambda_t = \frac{P_{w_{n+1}\mid w_1, w_2, \ldots, w_n}(t \mid W_1, W_2, \ldots, W_n)}{P_{w_{n+1}\mid w_1, w_2, \ldots, w_n}(t \mid W_1, W_2, \ldots, W_n)} \tag{2.2.4}$$

where

$$P_{w_{n+1}\mid w_1, w_2, \ldots, w_n}(t \mid W_1, W_2, \ldots, W_n)$$

$$= \int_t^\infty P_{w_{n+1}\mid w_1, w_2, \ldots, w_n}(x \mid W_1, W_2, \ldots, W_n)\, dx$$

and

$P_{w_{n+1}\mid w_1, w_2, \ldots, w_n}(t \mid W_1, W_2, \ldots, W_n)$ is the conditional probability density function of w_{n+1} given $\{w_1=W_1, w_2=W_2, \ldots, w_n=W_n\}$ evaluated at time t. We also have by (2.2.4)

$$\lambda_t(N_t=n; W_1, W_2, \ldots, W_n) = -\frac{d}{dt}\left[\ell n\, P_{w_{n+1}\mid w_1, w_2, \ldots, w_n}(t \mid W_1, \ldots, W_n)\right] \tag{2.2.5}$$

and

$$P_{w_{n+1}\mid w_1, \ldots, w_n}(t \mid W_1, \ldots, W_n) = \exp\left[-\int_{W_n}^{t} \lambda_s(N_s=n; W_1, W_2, \ldots, W_n)\, ds\right] \tag{2.2.6}$$

Since interarrival time statistics are related to the occurrence time statistics by the obvious:

$$\Pr\{t_{n+1} > t \mid t_1=T_1, t_2=T_2, \ldots, t_n=T_n\}$$

$$= P_{w_{n+1}|w_1,w_2,\ldots,w_n}(T_1+T_2+\ldots+T_n+t \mid T_1, T_1+T_2, \ldots, T_1+T_2+\ldots+T_n) \tag{2.2.7}$$

they can be used as the basis for the statistical description of a point process. In particular, these statistics can be used to compute the rate of a counting process. An example is provided by the following from Snyder [30, p. 265].

> Theorem: A self-exciting point process is 1-step memory with rate
>
> $$\lambda_t(N_t=n; W_n) = F(N_t=n; t-W_n)$$
>
> for some function $F(\cdot)$ if and only if the interarrival times, $t_1=w_1, t_2=w_2-w_1, \ldots, t_n=w_n-w_{n-1}, \ldots$ form a sequence of statistically independent random variables.

By combining (2.2.5), (2.2.7) for a 1-step memory process

$$\lambda_t(N_t=n; t-W_n) = -\frac{d}{dt}\left[\ell n \Pr\{t_{n+1} > t-W_n\}\right]$$

$$= \frac{p_{t_{n+1}}(t-W_n)}{\int_{t-W_n}^{\infty} p_{t_{n+1}}(x)\,dx} = h(t-W_n) \qquad , \; t \geq W_n \tag{2.2.8}$$

where $p_{t_{n+1}}(\cdot)$ is the (n+1)th interarrival time probability density.

The function $h(\cdot)$ is called the hazard function in birth or renewal processes. From (2.2.3), (2.2.8) it is seen that to order Δt the quantity $h(t-\tau)\Delta t$ is the conditional probability that the (n+1)th detector activation occurs in $(t, t+\Delta t]$ given that the (n)th occurred at time τ. The hazard function is easily measured from experimental data and has important implications in the performance of the filter developed below.

Similarly, for a 2-step memory process

$$\lambda_t(N_t=n;t-W_n, W_n-W_{n-1}) = -\frac{d}{dt}\left[\ell n \Pr\left\{t_{n+1}\geq t-W_n \,|\, t_n=W_n-W_{n-1}\right\}\right]$$

$$= \frac{p_{t_{n+1}|t_n}(t-W_n|T_n)}{\int_{t-W_n}^{\infty} p_{t_{n+1}|t_n}(x|T_n)} \qquad (2.2.9)$$

where $T_n=W_n-W_{n-1}$. Several useful examples are now considered.

Example 1. Given a homogeneous Poisson process, N_t with rate λ, then the interarrival times $\{t_n\}$ are independent, identically distributed (iid) with common pdf

$$p_{t_n}(T_n) = \lambda \exp(-T_n\lambda) \qquad , \; n=1,2,3,\ldots$$

and rate

$$\lambda_t^1 \equiv \lambda_t(N_t=n;t-W_n) = \frac{\lambda \exp(-\lambda(t-W_n))}{\int_{t-W_n}^{\infty} \lambda \exp(-\lambda x)\, dx} = \lambda \qquad (2.2.10)$$

The fact that in this case the hazard function is constant is one criterion usually employed to determine whether empirical interarrival times result from a homogeneous Poisson process [30, p. 269].

Example 2. Suppose interarrival times $\{t_n\}$ are lognormal variables as specified by (1.3.3) and notationally denoted by $L(\mu,\sigma)$ with transition probability density given by (1.3.4), then by (2.2.9)

$$\lambda_t(N_t=n;t-W_n, W_n-W_{n-1})$$

$$= \frac{\dfrac{1}{(t-W_n)\sigma\sqrt{2\pi(1-r^2)}} \exp\left[-\dfrac{(\ell n(t-W_n)-\mu-r(\ell n(W_n-W_{n-1})-\mu))^2}{2\sigma^2(1-r^2)}\right]}{1-\operatorname{erf}\left[\dfrac{\ell n(t-W_n)-\mu-r(\ell n(W_n-W_{n-1})-\mu)}{\sigma\sqrt{1-r^2}}\right]} \qquad (2.2.11)$$

where

$$\text{erf } x = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp(-t^2/2)\, dt$$

When $\rho=0$ in (1.3.6), then $r=0$, i.e. the interarrival times are iid with common density $L(\mu,\sigma)$ and (2.2.11) reduces to

$$\lambda_t^0 \equiv \lambda_t(N_t=n;t-W_n) = \frac{g(t-W_n)}{1 - \text{erf}\left[\frac{\ell n\,(t-W_n)-\mu}{\sigma}\right]} \tag{2.2.12}$$

where $g(\cdot)$ denotes $L(\mu,\sigma)$ probability density (see Fig. 2.11 for plot of λ_t^0 vs. time for different parameter values).

Example 3. Suppose interarrival times $\{t_n\}$ are iid with common pdf given by (1.3.7)

$$p_{t_n}(\xi) = \begin{cases} \lambda \exp(-\lambda\xi) \int_0^{\xi} g(x) \exp(\lambda x)\, dx & , \xi \geq 0 \\ 0 & , \xi < 0 \end{cases} \tag{2.2.13}$$

where $g(\cdot)$ is $L(\mu,\sigma)$ density and then the rate is

$$\lambda_{nf} \equiv \lambda_t(N_t=n;t-W_n) = \frac{p_{t_n}(t-W_n)}{1 - \int_0^{t-W_n} p_{t_n}(x)\, dx} \tag{2.2.14}$$

which is shown in Fig. 2.2 below.

From the complete statistical description of headways given by (1.3.10)-(1.3.13), it can be observed that the rate of the associated point process has a 2-step memory and in particular

$$\lambda_t(N_t=n;W_n, W_{n-1}) = \frac{p_{t_{n+1}|t_n}(t-W_n|W_n-W_{n-1})}{\int_{t-W_n}^{\infty} p_{t_{n+1}|t_n}(x|W_n-W_{n-1})\, dx} \tag{2.2.15}$$

where $p_{t_{n+1}|t_n}(\cdot|\cdot)$ is provided by (1.3.13). This expression for the rate is complicated and certain computational difficulties arise in further development. We obtain simpler expressions by considering special cases of the above. If $\rho=0$ in (1.3.6), then the following headways are independent and the conditional density in (1.3.13) reduces to (1.2.13), i.e. the headways are iid with common density $p_{t_{n+1}}(\cdot)$ provided in (1.2.13).

$$p_{t_{n+1}}(x) = \psi\, p_f(x) + (1-\psi)\, p_{nf}(x)$$

Defining the event and using the relation (1.3.1)

$$A = \{\text{headway is following}\} \quad , \Pr\{A\} = \psi$$

$$A^c = \{\text{headway is nonfollowing}\} \quad , \Pr\{A^c\} = 1-\psi$$

then by the definition of rate (2.2.3)

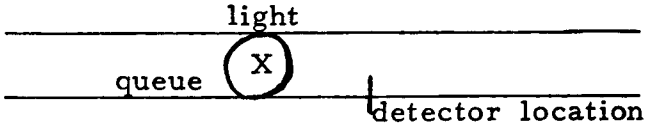
$$\begin{aligned}\lambda_t(N_t=n;t-W_n) &= \lim_{\Delta t\downarrow 0} \Pr\{N_{t+\Delta t}-N_t=1\,|\,\mathfrak{F}_t\} \\ &= \psi \lim_{\Delta t\downarrow 0} \Pr\{N_{t+\Delta t}-N_t=1\,|\,\mathfrak{F}_t, A\} \\ &\quad + (1-\psi) \lim_{\Delta t\downarrow 0} \Pr\{N_{t+\Delta t}-N_t=1\,|\,\mathfrak{F}_t, A^c\} \\ &= \psi\, \lambda_t^0 + (1-\psi)\, \lambda_{nf} \end{aligned} \tag{2.2.16}$$

where $\mathfrak{F}_t = \{N_t=n; w_1=W_1, w_2=W_2, \ldots, w_n=W_n\}$, λ_t^0 and λ_{nf} given by (2.2.12) and (2.2.14), respectively. Clearly this decomposition results because of the independence of the following and non-following processes [assumption(i), p.23]. Since the development of a filter/predictor in this chapter follows from the fact that ψ takes on only one of two values (Fig. 2.1), we determine the rates for these two cases. If $\psi=1$ in (1.3.13), i.e. all headways are following, then the conditional

density (1.3.13) has the form of (1.3.4) and the rate computed in (2.2.11). If $\psi = 0$ in (1.3.13), i.e. all headways are nonfollowing, then the conditional density (1.3.13) reduces to (2.2.13) with rate provided by (2.2.14). These latter two rates (with slight modifications) are utilized in the development of the "disorder problem" presented below.

2.3 Disorder Problem Formulation

In this section, we discuss the formulation and solution of the so called "disorder problem" as it relates to urban traffic problems. To motivate the development, consider the following physical traffic situation. Assume traffic downstream from a traffic light is free-flowing and the time the light changes from red to green is known. After the light changes, one observes at



the detector a series of following vehicles whose rate λ_t^0 is specified by (2.2.12). After the last following vehicle passes, a different rate λ_t^1, describing nonfollowing vehicles is observed. The relevant question is: when did the last vehicle in the queue pass the detector location? By having an estimate of the convex parameter ψ using some cost criteria, the time of disorder can be determined. More precisely, one observes a counting process N_t that is governed by a rate until some random "disorder" time T at which point the process is governed by a different rate. The problem is to estimate the switch time T using only the observations, N_t.

In a general context, one attempts to estimate a signal process, x_t given the observations of a counting process, N_t which is influenced by x_t. Let $(\Omega, \mathcal{B}, P)$ be a fixed probability space, $\mathcal{F}_t$ be an increasing σ-algebra generated by $\{N_t; t \geq 0\}$ and $\mathcal{B}_t$ be an increasing σ-algebra such that all processes discussed below are adapted to it. Consider the signal process to be modelled by:

$$dx_t = f_t\, dt + dv_t \quad , \; x(0) = x_0 \tag{2.3.1}$$

where v_t is a square-integrable martingale with respect to $\mathcal{B}_t$ and the process f_t satisfies:

$$E \int_0^t |f_s|\, ds < \infty \quad \text{for all } t > 0$$

Let the observation process, N_t with rate λ_t be defined by:

$$dN_t = \lambda_t\, dt + dw_t \quad , \; E[N_t] < \infty \tag{2.3.2}$$

where w_t is a square-integrable martingale with respect to $\mathcal{B}_t$. Finally assume the cross-quadratic variation process defined as:

$$d\langle v, w\rangle_t = E[dv_t^c\, dw_t^c \,|\, \mathcal{B}_t] + E[\Delta v_t \Delta w_t \,|\, \mathcal{B}_t]$$

where $\Delta v_t = v_t - v_{t-}$ and v_t^c, w_t^c are continuous processes, is absolutely continuous with respect to the Lebesque measure. Then a nonlinear recursive estimate

$$\hat{x}_t \equiv E[x_t \,|\, \mathcal{F}_t] \tag{2.3.3}$$

has been obtained by Bremaud [36], Boel-Varaiya-Wong [37, 38] and Segall-Davis-Kailath [41] using the martingale approach:

$$d\hat{x}_t = \hat{f}_t\, dt + (\hat{\lambda}_t)^{-1}\, E\{[x_t(\lambda_t - \hat{\lambda}_t) + \frac{d}{dt}\langle v, w\rangle_t] \,|\, \mathcal{F}_t\} \Big|_{t-} (dN_t - \hat{\lambda}_t\, dt)$$

$$\hat{x}_0 = E[x(0)]. \tag{2.3.4}$$

where " ^ " denotes conditional expectation with respect to $\mathfrak{F}_t$.

The disorder problem has been considered by several researchers[42 ,43 ,44]. Assuming the a priori probability density of switch time T is an exponential and using properties of Markov processes, a switch time estimate can be obtained [42]. Recently, these results have been generalized by Davis within the martingale approach for point processes [43,44]. We follow this latter reference in the development presented here.

We first need to establish the structure of the problem as in (2.3.1), (2.3.2). Let us define

$$x_t = I_{\{t \geq T\}} \tag{2.3.5}$$

$$B_t = \sigma\{N_s, x_s; s \leq t\} \quad , \quad \mathfrak{F}_t = \sigma\{N_s; s \leq t\}$$

where $I_{\{t \geq T\}}$ is the characteristic function of the set $\{t \geq T\}$ and $\sigma(\cdot)$ denotes an increasing σ-algebra generated by variables specified. From the discussion above and by (2.3.5), it follows that the counting process N_t has a rate

$$\lambda_t = (1-x_t)\,\lambda_t^0 + x_t\,\lambda_t^1 \tag{2.3.6}$$

where λ_t^0 , λ_t^1 are rates measurable with respect to $\mathfrak{F}_t$. By the definition of rates (2.2.2), then N_t has the form of (2.3.2). Notice that since λ_t and N_t are sample discontinuous, w_t is sample discontinuous. We now consider several cases in which x_t is represented in the form of (2.3.1) with an explicit filter equation derived.

Case A: Assume T is independent of N_t and has an a priori distribution function

$$P_T(t) = \Pr\{T \leq t\} \quad , \quad F_t = 1 - P_T(t) \tag{2.3.7}$$

Writing x_t into form of (2.3.1) by a $\mathcal{B}_t$-decomposition:

$$E\,[dx_t|\mathcal{B}_t] = E\,[x_{t+dt}-x_t|\mathcal{B}_t]$$

$$= E\,[I_{\{t+dt \geq T\}}-I_{\{t \geq T\}}|\mathcal{B}_t]$$

The random variable

$$I_{\{t+dt \geq T\}}-I_{\{t \geq T\}} = \begin{cases} 1 & \text{for } t < T \leq t+dt \\ 0 & \text{otherwise} \end{cases}$$

and by the independence of N_t and T, then

$$E\,[dx_t|\mathcal{B}_t] = I_{\{t< T\}}\ \Pr\ \{t<T\leq t+dt|\mathcal{B}_t\} \tag{2.3.8}$$

$$= I_{\{t< T\}}\ \Pr\ \{t< T\leq t+dt|t< T\} \quad \text{for } t\geq 0$$

In terms of F_t

$$E\,[dx_t|\mathcal{B}_t] = I_{\{t< T\}}\left(-\frac{dF_t}{F_{t-}}\right) = -(1-x_t)\left(\frac{dF_t}{F_{t-}}\right)$$

Hence

$$dv_t = dx_t - E\,[dx_t|\mathcal{B}_t] = dx_t + (1-x_t)\left(\frac{dF_t}{F_{t-}}\right) \tag{2.3.9}$$

is a square-integrable $\mathcal{B}_t$ martingale of the form (2.3.1). Computing the other terms in the filter representation (2.3.4):

$$\hat{f}_t\,dt = E\,[f_t|\mathcal{F}_t]\,dt = E\,\left[-(1-x_t)\left(\frac{dF_t}{F_{t-}}\right)|\mathcal{F}_t\right]$$

$$= -(1-\hat{x}_t)\left(\frac{dF_t}{F_{t-}}\right) \tag{2.3.10}$$

where $\hat{x}_t$ is defined in (2.3.3) and because F_t is deterministic.

$$d\langle v, w\rangle_t = 0 \tag{2.3.11}$$

because N_t and T are independent.

$$E\,[x_t(\lambda_t-\hat{\lambda}_t)|\mathcal{F}_t] = E\,[x_t\lambda_t|\mathcal{F}_t] - \hat{x}_t\hat{\lambda}_t$$

since $\hat{\lambda}_t$ is adapted to $\mathcal{F}_t$.

But

$$E\,[x_t \lambda_t | \mathcal{F}_t] = E\,[x_t ((1-x_t)\,\lambda_t^0 + x_t \lambda_t^1) | \mathcal{F}_t]$$

$$= \lambda_t^0\,(\hat{x}_t - E\,[x_t^2 | \mathcal{F}_t]\,) + \lambda_t^1\,E\,[x_t^2 | \mathcal{F}_t]$$

since λ_t^0 , λ_t^1 are adapted to $\mathcal{F}_t$. Because $x_t = I_{\{t \geq T\}}$, $x_t = x_t^2$ and

$$E\,[x_t \lambda_t | \mathcal{F}_t] = \lambda_t^0\,(\hat{x}_t - \hat{x}_t) + \lambda_t^1\,\hat{x}_t = \lambda_t^1\,\hat{x}_t$$

Combining

$$E\,[x_t(\lambda_t - \lambda_t) | \mathcal{F}_t] = (\lambda_t^1 - \hat{\lambda}_t)\,\hat{x}_t$$

$$= \{\lambda_t^1 - [(1-\hat{x}_t)\,\lambda_t^0 + \hat{x}_t\,\lambda_t^1]\}\,\hat{x}_t$$

$$= \hat{x}_t\,\lambda_t^1\,(1-\hat{x}_t) - \hat{x}_t\,(1-\hat{x}_t)\,\lambda_t^0$$

$$= (\lambda_t^1 - \lambda_t^0)\,\hat{x}_t(1-\hat{x}_t) \qquad (2.3.12)$$

and combining (2.3.10), (2.3.11) and (2.3.12) in (2.3.4) yields:

$$d\hat{x}_t = -\,(1-\hat{x}_t)\,(\frac{dF_t}{F_{t-}}) + (\hat{\lambda}_t)^{-1}(\lambda_t^1 - \lambda_t^0)\,\hat{x}_t(1-\hat{x}_t)\;\;(dN_t - \hat{\lambda}_t\,dt) \qquad (2.3.13)$$

where $\hat{\lambda}_t = (1-\hat{x}_t)\,\lambda_t^0 + \hat{x}_t \lambda_t$ and $\hat{x}_0 = E\,[x_0] = 1 - F_0$

Case B: Assume T is equal to $\{W_n\}$, i.e. some occurrence time and the events $\{T=W_n\}$ are independent of $\{N_t ; t \geq 0\}$.

Let

$$p_n = Pr\,\{T=W_n\}$$

and define

$$q_n = Pr\,\{T=W_n | T > W_{n-1}\} = \frac{p_n}{\sum\limits_{i \geq n} p_i} \qquad (2.3.14)$$

Consider the $\mathcal{B}_t$-decomposition of the signal process. Clearly,

$$N_t = \sum_n I_{\{t \geq W_n\}}$$

Therefore we have for $W_{n-1} \leq t < W_n$

$$\Pr\{\text{the switch occurs at the next detector activation time and in the interval } (t, t+dt], \text{ given the past observations}\}$$

$$= \Pr\{N_{t,t+dt}=1, T=W_n \mid T>t, N_t=n-1; w_1=W_1, w_2=W_2, \ldots, w_{n-1}=W_{n-1}\}$$

$$= \Pr\{t< W_n \leq t+dt, T=W_n \mid T>t, N_t=n-1; w_1=W_1, w_2=W_2, \ldots, w_{n-1}=W_{n-1}\}$$

(by the independence of $\{T=W_n\}$ and $\{N_t;\ t\geq 0\}$)

$$= \Pr\{t< W_n \leq t+dt \mid N_t=n-1; w_1=W_1, w_2=W_2, \ldots, w_{n-1}=W_{n-1}\} \times$$

$$\Pr\{T=W_n \mid T>t, N_t=n-1; w_1=W_1, w_2=W_2, \ldots, w_{n-1}=W_{n-1}\}$$

$$= (\lambda_t^0\ dt)\ q_n$$

where we used (2.3.14) and the definition of the rate of point processes (2.2.3). Summing over all occurrence times $\{W_n\}$

$$\Pr\{t< T \leq t+dt \mid \mathcal{B}_t\}$$

$$= \sum_n \Pr\{N_{t,t+dt}=1, T=W_n \mid T>t, N_t=n-1; w_1=W_1, \ldots, w_{n-1}=W_{n-1}\}\ I_{\{W_{n-1}\leq t< W_n\}}$$

$$= \sum_n q_n \lambda_t^0\ I_{\{W_{n-1}\leq t< W_n\}}\ dt \qquad (2.3.15)$$

Define the function

$$q_t = \sum_n q_n\ I_{\{W_{n-1}\leq t< W_n\}} \qquad (2.3.16)$$

Then by (2.3.8)

$$E[dx_t \mid \mathcal{B}_t] = I_{\{t< T\}} \Pr\{t< T \leq t+dt \mid \mathcal{B}_t\}$$

$$= I_{\{t< T\}}\ q_t\ \lambda_t^0\ dt = (1-x_t)\ q_t\ \lambda_t^0\ dt = f_t\ dt \qquad (2.3.17)$$

which has the form of (2.3.1)

$$dx_t = (1-x_t)\ q_t\ \lambda_t^0\ dt + dv_t \qquad (2.3.18)$$

where v_t is a $\mathcal{B}_t$-martingale. Since x_t and q_t are sample discontinuous then v_t is sample discontinuous with

$$dv_t = v_{t+dt} - v_t = \begin{cases} 0 & t \neq T \\ 1 & t = T \end{cases}$$

Since

$$\Delta v_t \, \Delta w_t = \begin{cases} 1 & t = T \\ 0 & \text{otherwise} \end{cases}$$

and

$$\Pr\{\Delta v_t \, \Delta w_t = 1 \mid \mathcal{B}_t\} = I_{\{t < T\}} \Pr\{t < T \leq t+dt \mid \mathcal{B}_t\}$$

$$= (1-x_t)\, q_t\, \lambda_t^0\, dt$$

by (2.3.17). Then

$$d\langle v, w\rangle_{t-} = (1-x_t)\, q_t\, \lambda_t^0\, dt \Big|_{t-}$$

and

$$E\left[\frac{d}{dt}\langle v, w\rangle_{t-} \mid \mathcal{F}_t\right] = (1-\hat{x}_t)\, q_t\, \lambda_t^0 \Big|_{t-} = \hat{f}_{t-} \tag{2.3.19}$$

By the previous calculation (2.3.12)

$$E\left[x_t(\lambda_t - \hat{\lambda}_t) \mid \mathcal{F}_t\right] = (\lambda_t^1 - \lambda_t^0)\, \hat{x}_t (1-\hat{x}_t) \tag{2.3.20}$$

and combining (2.3.17), (2.3.19) and (2.3.20) in (2.3.4) yields:

$$d\hat{x}_t = (1-\hat{x}_t)\, q_t\, \lambda_t^0\, dt + (\hat{\lambda}_t)^{-1}\left[(\hat{\lambda}_t - \lambda_t^0)\, \hat{x}_t (1-\hat{x}_t) + (1-\hat{x}_t) q_t \lambda_t^0\right]\Big|_{t-} \cdot (dN_t - \hat{\lambda}_t\, dt)$$

Rearranging

$$d\hat{x}_t = -(\lambda_t^1 - \lambda_t^0)\, \hat{x}_t (1-\hat{x}_t)\, dt$$

$$+ \frac{(\lambda_t^1 - \lambda_t^0)\, \hat{x}_t (1-\hat{x}_t) + q_t \lambda_t^0\, (1-\hat{x}_t)}{(\lambda_t^1 - \lambda_t^0)\, \hat{x}_t + \lambda_t^0}\, dN_t \tag{2.3.21}$$

with $\hat{x}_0 = E[x_0] = p_0$

Case C: Assume T is equal to $\{W_n\}$, i.e. some occurrence time and the events $\{T=W_n\}$ are dependent on $\{N_t; t \geq 0\}$.

Consider again the $\mathcal{B}_t$-decomposition of the signal process.

$$\Pr\{N_{t,t+dt}=1, T=W_n \mid T>t, N_t=n-1; w_1=W_1, \ldots, w_{n-1}=W_{n-1}\}$$

$$= \Pr\{t<W_n\leq t+dt, T=W_n \mid T>t, N_t=n-1; w_1=W_1, \ldots, w_{n-1}=W_{n-1}\}$$

$$= \Pr\{T=W_n \mid t<W_n\leq t+dt, T>t, N_t=n-1; w_1=W_1, \ldots, w_{n-1}=W_{n-1}\}$$

$$\cdot \Pr\{t<W_n\leq t+dt \mid T>t, N_t=n-1; w_1=W_1, \ldots, w_{n-1}=W_{n-1}\}$$

$$= q_n(t, W_1, \ldots, W_{n-1})\ \lambda_t^0\, dt$$

where we have used the definition of rate of a point process(2.2.3) and have defined:

$$q_n(t, W_1, \ldots, W_{n-1}) = \Pr\{T=W_n \mid t<W_n\leq t+dt, T>t, N_t=n-1; w_1=W_1, \ldots w_{n-1}=W_{n-1}\} \tag{2.3.22}$$

Summing over all occurrence times $\{W_n\}$:

$$\Pr\{t<T\leq t+dt \mid \mathcal{B}_t\}$$

$$= \sum_n \Pr\{N_{t,t+dt}=1, T=W_n \mid T>t, N_t=n-1; w_1=W_1, \ldots, w_{n-1}=W_{n-1}\}\ I_{\{W_{n-1}\leq t<W_n\}}$$

$$= \sum_n (\lambda_t^0\, dt)\ q_n(t, W_1, \ldots, W_{n-1})\ I_{\{W_{n-1}\leq t<W_n\}}$$

$$= q_t\ \lambda_t^0\, dt \tag{2.3.23}$$

where

$$q_t = \Sigma\, q_n(t, W_1, \ldots, W_{n-1})\ I_{\{W_{n-1}\leq t<W_n\}} \tag{2.3.24}$$

The function q_t is adapted to $\mathcal{F}_t$ and generalizes the results of Case B by making the a priori probability density dependent on the sample path $\{N_t; t\geq 0\}$. The resulting filter has the same form as (2.3.21) with (2.3.22) and (2.3.24) replacing (2.3.14) and (2.3.16) respectively.

2.4 Evaluation

From the development of the recursive filter, it is apparent that once the rates and the a priori density are specified, the resulting switch time estimate can be obtained. Considering the traffic problem posed in the previous section, several assumptions are made to simplify the evaluation. First we compare the trade-off between a higher complexity model for the two rates which will better represent the sensed data, (2.2.11) (2.2.14) and the resulting computational complexity of the filter. We use the 1-memory self-exciting point process model for following headways, i.e. λ_t^0 is provided by (2.2.12). As will be seen from the evaluation below, this approximation does not reduce the performance of the estimator as it applies to this specific traffic problem. The other rate should be that for the nonfollowing headways, i.e. $\lambda_t^1 = \lambda_{nf}$ as given in (2.2.14). In view of the complexity of (2.2.13), the nonfollowing headway probability density, it is difficult to compute λ_t^1 and an approximation is needed in order to produce a simple filter. From the nonfollowing hazard function (Fig. 2.2), it is seen that for headway ranges where the nonfollowing hazard function should dominate, this function is constant to a fairly good approximation. Therefore we assume,

$$\lambda_t^1 = \lambda = \text{rate of free-flowing traffic}$$

This indeed is very well justified since a Poisson model for nonfollowing vehicles is rather widely accepted (see Section 1.2). The evaluation of the filter's performance under this assumption proves that this approximation also does not harm the accuracy of the estimator.

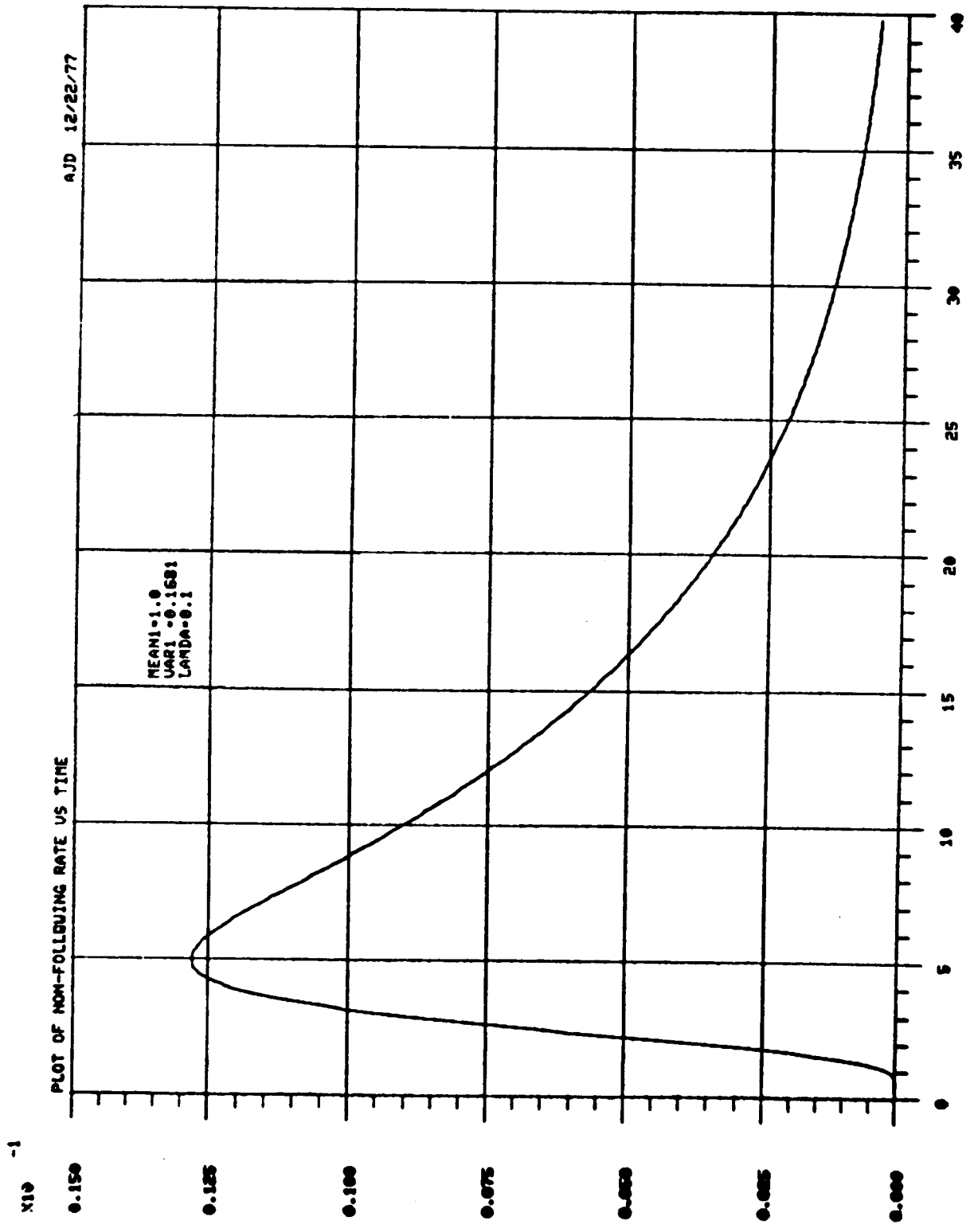


Fig. 2.2: Plot of Non-following Rate vs time

Besides the rates, the a priori density affects the response of the filter to detector activation times. Clearly, the a priori density should be related to these activation times because the estimate desired is the number of vehicles passing a detector location. From the physical standpoint though these times should also influence the a priori density. But the degree of such influence is unknown and the added complexity makes the assumption that the events $\{T=W_n\}$ are independent of $\{N_t; t \geq 0\}$ acceptable. In summary then, we shall evaluate the disorder problem formulated in Case B above with rate $\lambda_t^1 = \lambda$, a constant.

In the following discussion, we consider the implementation of the filter equation (2.3.21) and the associated simulation techniques. Afterwards the filter's sensitivity to changes in the a priori distribution, to different hazard functions of λ_t^0 and parameter variations of a particular hazard function are considered. Finally the filter is tested against urban traffic simulated data and a summary of results are presented.

Before turning to the implementation of (2.3.21), we consider an acceptable range in which the a priori distribution and the following headway rate may vary. For a long queue (10 or more vehicles) leaving a traffic light, a downstream detector observes headways as modelled by the lognormal rate (2.2.12). However, for short queues (3 or less vehicles), the counting process is more random and a Poisson rate may be more acceptable. As for the a priori density, it is clear from the definition of the p_n's that the information carried by them is identical to the probability density for queue length. In other words, one could bias the estimator around a

particular activation time if one knew the size of the queue or alternatively, if no knowledge is known about the queue, a uniform density would be more appropriate. Therefore we shall evaluate the filter for the following cases:

Case I Two Poisson rates with a discrete 3-valued a priori density.

Case II Two Poisson rates with a uniform a priori density.

Case III A lognormal rate switching to a Poisson rate with a discrete 3-valued a priori density.

Case IV A lognormal rate switching to a Poisson rate with a uniform a priori density.

Finally, for this phase of the evaluation, the generation of the sample path $\{N_t ; t \geq 0\}$ with the specified rates follows Knuth [45]. Define

$$\pi_t \equiv \hat{x}_t = E\,[x_t | \mathfrak{F}_t] \tag{2.4.1}$$

Case I, II: For $\lambda_t^0 = \lambda_0$, $\lambda_t^1 = \lambda_1$, and $W_{n-1} < t < W_n$, then by (2.3.21)

$$d\pi_s = -(\lambda_1 - \lambda_0)\,\pi_s(1-\pi_s)\,ds$$

$$\int_{W_{n-1}}^{t} \frac{d\pi_s}{\pi_s(1-\pi_s)} = -\int_{W_{n-1}}^{t} (\lambda_1 - \lambda_0)\,ds$$

$$-\ell n \left(\frac{1-\pi_s}{\pi_s} \right) \Bigg|_{W_{n-1}}^{t} = -(t - W_{n-1})\,(\lambda_1 - \lambda_0)$$

$$\left(\frac{1-\pi_t}{\pi_t} \right) \cdot \left(\frac{1-\pi_{W_{n-1}}}{\pi_{W_{n-1}}} \right)^{-1} = \exp\left[(t - W_{n-1})\,(\lambda_1 - \lambda_0) \right]$$

$$\pi_t = \left[\left(\frac{1-\pi_{W_{n-1}}}{\pi_{W_{n-1}}} \right) \exp\left[(t-W_{n-1})(\lambda_1-\lambda_0) \right] + 1 \right]^{-1} \tag{2.4.2}$$

and at $t = W_n$ (i.e. at detector activation times)

$$\pi_t - \pi_{t-} = (1-\pi_{t-}) \left[\frac{(\lambda^1_{t-} - \lambda^0_{t-})\, \pi_{t-} + q_{t-}\, \lambda^0_{t-}}{(\lambda^1_{t-} - \lambda^0_{t-})\, \pi_{t-} + \lambda^0_{t-}} \right] \tag{2.4.3}$$

Case III, IV: For $\lambda^1_t = \lambda$, $W_{n-1} < t < W_n$ and by (2.2.12)

$$\lambda^0_t \equiv \frac{g(t-W_{n-1})}{1 - \operatorname{erf}\left| \dfrac{\ell n\,(t-W_{n-1}) - \mu}{\sigma} \right|}$$

where $g(\cdot)$ denotes $L(\mu, \sigma)$ probability density. Then by (2.3.21)

$$d\pi_s = -(\lambda - \lambda^0_s)\, \pi_s (1-\pi_s)\, ds$$

$$\int_{W_{n-1}}^{t} \frac{d\pi_s}{\pi_s (1-\pi_s)} = -\int_{W_{n-1}}^{t} (\lambda - \lambda^0_s)\, ds$$

$$-\ell n \left(\frac{1-\pi_s}{\pi_s} \right) \Bigg|_{W_{n-1}}^{t} = -\lambda (t-W_{n-1}) + \int_{W_{n-1}}^{t} \lambda^0_s\, ds \tag{2.4.4}$$

By (2.2.6), (2.2.7)

$$\int_{W_{n-1}}^{t} \lambda^0_s\, ds = -\ell n \left[\Pr \{ t_n > t - W_{n-1} \} \right]$$

$$= -\ell n \left[\int_{t-W_{n-1}}^{\infty} g(x)\, dx \right] = -\ell n \left[1 - \operatorname{erf}\left(\frac{\ell n\,(t-W_{n-1}) - \mu}{\sigma} \right) \right] \tag{2.4.5}$$

(since interarrival times for following process are modelled by lognormal density $L(\mu, \sigma)$). Combining (2.4.4) and (2.4.5)

$$\left(\frac{1-\pi_t}{\pi_t}\right)\left(\frac{1-\pi_{W_{n-1}}}{\pi_{W_{n-1}}}\right)^{-1} = \exp\left[\lambda(t-W_{n-1})\right]\cdot\left[1-\operatorname{erf}\left(\frac{\ell n(t-W_{n-1})-\mu}{\sigma}\right)\right]$$

$$\pi_t = \left[\left(\frac{1-\pi_{W_{n-1}}}{\pi_{W_{n-1}}}\right)\exp\left[(\lambda(t-W_{n-1})\right]\left[1-\operatorname{erf}\left(\frac{\ell n(t-W_{n-1})-\mu}{\sigma}\right)\right]+1\right]^{-1}$$

$$\text{for } W_{n-1} < t < W_n \qquad (2.4.6)$$

Hence given the sample path $\{N_t; t \geq 0\}$, then between occurrence times, π_t is computed by using either (2.4.2) or (2.4.6) and at jump times $t = W_n$, the discontinuity is computed using (2.4.3). The evaluation of the erf function is performed by a five (5) term series expansion [46].

The above equations are implementable once the a priori density and the rates are specified. For Case I and III, a discrete 3-valued a priori density centered around the 10th activation time was used

$$p_n = \begin{cases} 0 & n=0,1,2,\ldots,8 \\ 0.0125 & n=9,11 \\ 0.975 & n=10 \end{cases} \qquad (2.4.7)$$

with the associated q_t shown in the figure below.

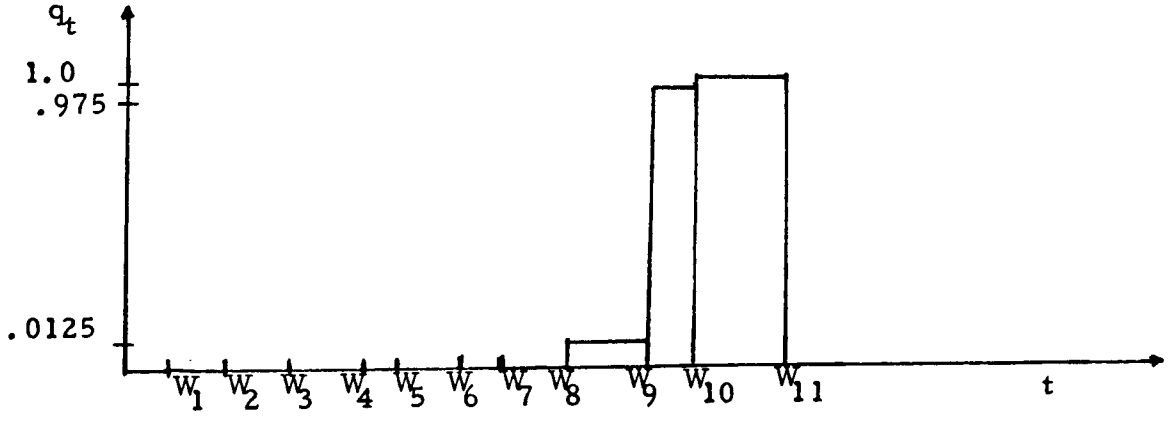


Figure 2.3 The function q_t for nonuniform a priori density of switching time T.

For Cases II and IV, the uniform density was selected as:

$$p_n = \frac{1}{20} \qquad n = 0, 1, 2, \ldots, 19 \tag{2.4.8}$$

with the associated q_t shown in the figure below.

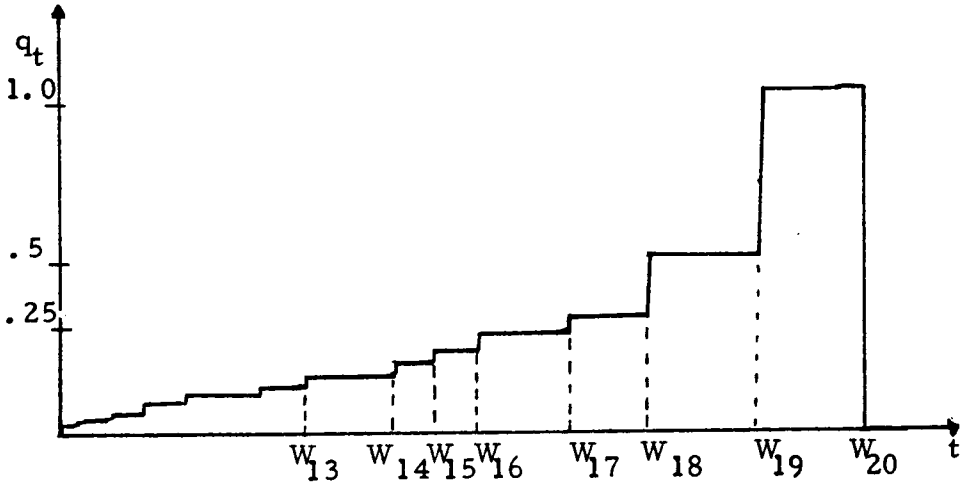


Figure 2.4 The function q_t for uniform a priori distribution of switching time T.

The parameter values for specifying the rates were chosen arbitrarily but are reasonable values given the empirical results of Branston [29]. The variations considered attempt to show the sensitivity of the filter. Several lognormal rates are shown in Fig. 2.11. The sample path $\{N_t ; t \geq 0\}$ generated for the different rates specified, had the switch time set arbitrarily to occur after W_{10}.

Having presented the necessary assumptions to implement the filter (2.3.21), we now consider certain general properties. Re-writing

$$d\pi_t = -(\lambda_t^1 - \lambda_t^0)\, \pi_t (1-\pi_t)\, dt + f(\pi_{t-})\, dN_t \tag{2.4.9}$$

where

$$f(\pi_{t-}) = (1-\pi_t) \left[\frac{\lambda_t^1\, \pi_{t-} + \lambda_t^0\, (q_{t-} - \pi_{t-})}{\lambda_t^1\, \pi_{t-} + \lambda_t^0\, (1-\pi_{t-})} \right]$$

Suppose $\lambda_t^0 > \lambda_t^1$ and $dN_t = 0$, then π_t is strictly increasing, approaching one asymptotically. For $dN_t \neq 0$, the discontinuities of π_t are positive or negative depending on the rates and the associated a priori density. Defining

$$A = \{ t \in W_n : \lambda_t^0 (\pi_{t-} - q_{t-}) \geq \lambda_t^1 \pi_{t-} \}$$
$$\tau = \text{imf } (A) \tag{2.4.10}$$

then $\Delta \pi_\tau = \pi_\tau - \pi_{\tau-} \leq 0$. The propagation of the minimum error variance estimate of the signal x_t given in (2.4.9) is a conditional distribution. Therefore to obtain a scalar estimate of x_t one generally accepted cost criterion is optimizing the expectation of some integral functional of x_t, i.e.

$$E \int_0^{t_n} f(x_s)\, ds$$

over acceptable values of the integrand. By the properties of conditional expectation, this optimization can sometimes be simplified. A particular Poisson rate disorder problem using the above criterion was shown to reduce to a simple threshold test [42,43]. In our situation, we felt a more relevant criterion was to choose the activation time which showed the largest increase in the estimated probability that a platoon of vehicles had passed the detector. This estimate is called the maximum jump estimate. From the physical standpoint such a single, scalar estimate could be directly related to the number of vehicles passing the detector in any particular traffic light cycle. The conditional error variance of x_t can be shown to achieve its maximum using this estimate (see Section 3.2). Finally, π_t being a functional of μ, λ, σ makes the filter's performance sensitive to parameter variations. In a subsequent chapter, this sensitivity of π_t is examined and bounds on the error performance are discussed.

We consider first the effect of the a priori density on the filter's response. For the highly localized 3-valued a priori density, the filter should be biased around the 10^{th} activation, W_{10}. For the resulting q_t (Fig. 2.3), the set A in (2.4.10) is the null set. Therefore (π_t=0 for $t< W_9$), all sebsequent activation times cause $\Delta\, \pi_t$ to be positive. For the two types of rate processes,

Fig. 2.5*: Case I, Rates $\lambda_t^0 = 0.9 \quad \lambda_t^1 = 0.1$

Fig. 2.6: Case III, Rates $\lambda_t^0 \sim L(-0.2412, 0.832) \quad \lambda_t^1 = 0.2$

the filter responds only around $t=W_{10}$. The filter is independent of the lognormal or the Poisson rate for λ_t^0. Therefore (as other simulations showed) variation of rate parameters for either Case I or III did not alter the filter's behavior. Obviously

$$\pi_t = \begin{cases} 0 & t< W_9 \\ 1 & t> W_{11} \end{cases}$$

causes the influence of the filter to different rates to be minimal. However once the a priori density is uniform, the filter responds to each activation time and variation of rate parameters is more pronounced. By having q_t (Fig. 2.4) perturb π_t at each activation time, the filter weighs more highly upon the difference between the two rates to determine the switch time.

Fig. 2.7: Case II, Rates $\lambda_0 = 0.9 \quad \lambda_1 = 0.1$

Fig. 2.8: Case II, Rates $\lambda_0 = 0.3 \quad \lambda_1 = 0.2$

---

* This series of figures shows the evolution of π_t vs. time computed by equations (2.4.2), (2.4.3) and (2.4.6). The associated table of numbers are the randomly generated activation times $\{W_n\}$ and interarrival times $\{T_n\}$.

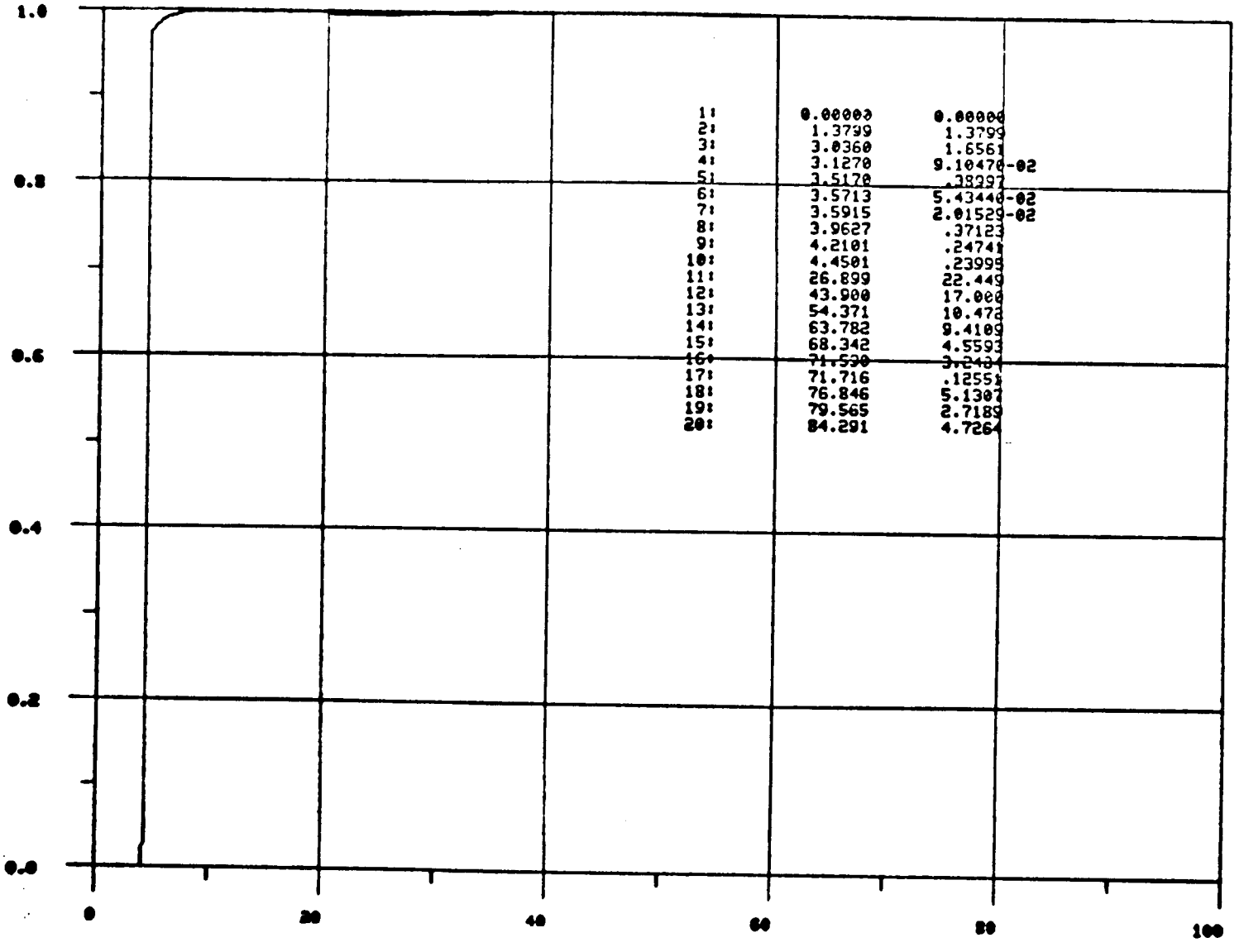


Fig. 2.5: Conditional Probability Distribution π_t vs. time

Case I, Rates $\lambda_t^0 = 0.9 \quad \lambda_t^1 = 0.1$

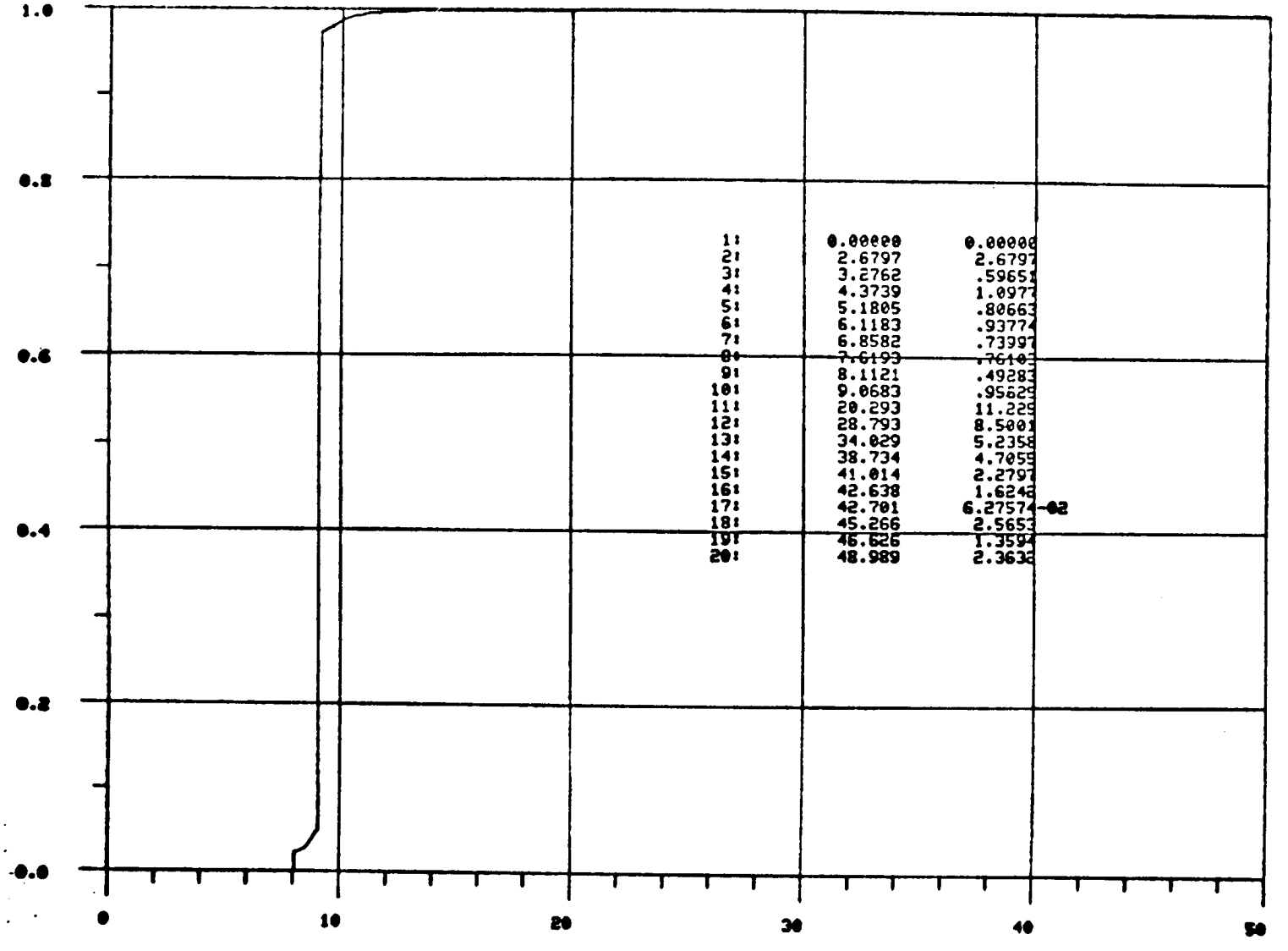


Fig. 2.6: Conditional Probability Distribution π_t vs. time

Case III, Rates $\lambda_t^0 \sim L(-0.2412, 0.832)$ $\lambda_t^1 = 0.2$

Fig. 2.9*: Case IV, Rates $\lambda_t^0 \sim L(-0.2412, 0.832)$ $\lambda_t^1 = 0.2$

Fig. 2.10: Case IV, Rates $\lambda_t^0 \sim L(0.857, 0.832)$ $\lambda_t^1 = 0.2$

In Fig. 2.7, 2.9 the filter's response is almost identical. The smoothness of the latter π_t results because λ_t^0 is time varying (Fig. 2.11) while in the former λ_t^0 is constant. In both figures though for widely separated rates, the filter readily distinguishes the switch time. When the rates are nearly equal (Fig. 2.8, 2.10) the filter is unable to determine the switch time and π_t follows the a priori distribution (a ramp). In summary, a highly localized a priori density makes the filter insensitive to past observations of the counting process and becomes biased around the associated p_n. Alternatively for the uniform a priori density, the filter has a greater difficulty in determining the switch time but clearly utilizes the entire past observations in the process.

From the above (Fig. 2.7-2.10), the filter's sensitivity to parameter changes in the rate can be seen. For the widely separated rates (Fig. 2.7, 2.9) and before the 10^{th} activation time, whatever increase determined by the continuous part π_t^c, is negated by the discontinuous part π_t^d. In other words uniformity in the p_n's causes a balancing effect between π_t^c and π_t^d on the filter's response. Notice that π_t^d can be positive and even zero (Fig. 2.8, 2.10). The influence of this nonlinear term in the rates is obviously related to the set A in (2.4.10). However, an explanation relating the quantities involved to the filter's response is unknown.

---

* The parameters for the lognormal rates in Fig. 2.9, 2.10 were chosen such that their mean and variance, given by (2.4.11) are equal to the mean and variance for a Poisson rate of 0.9 and 0.3 respectively.

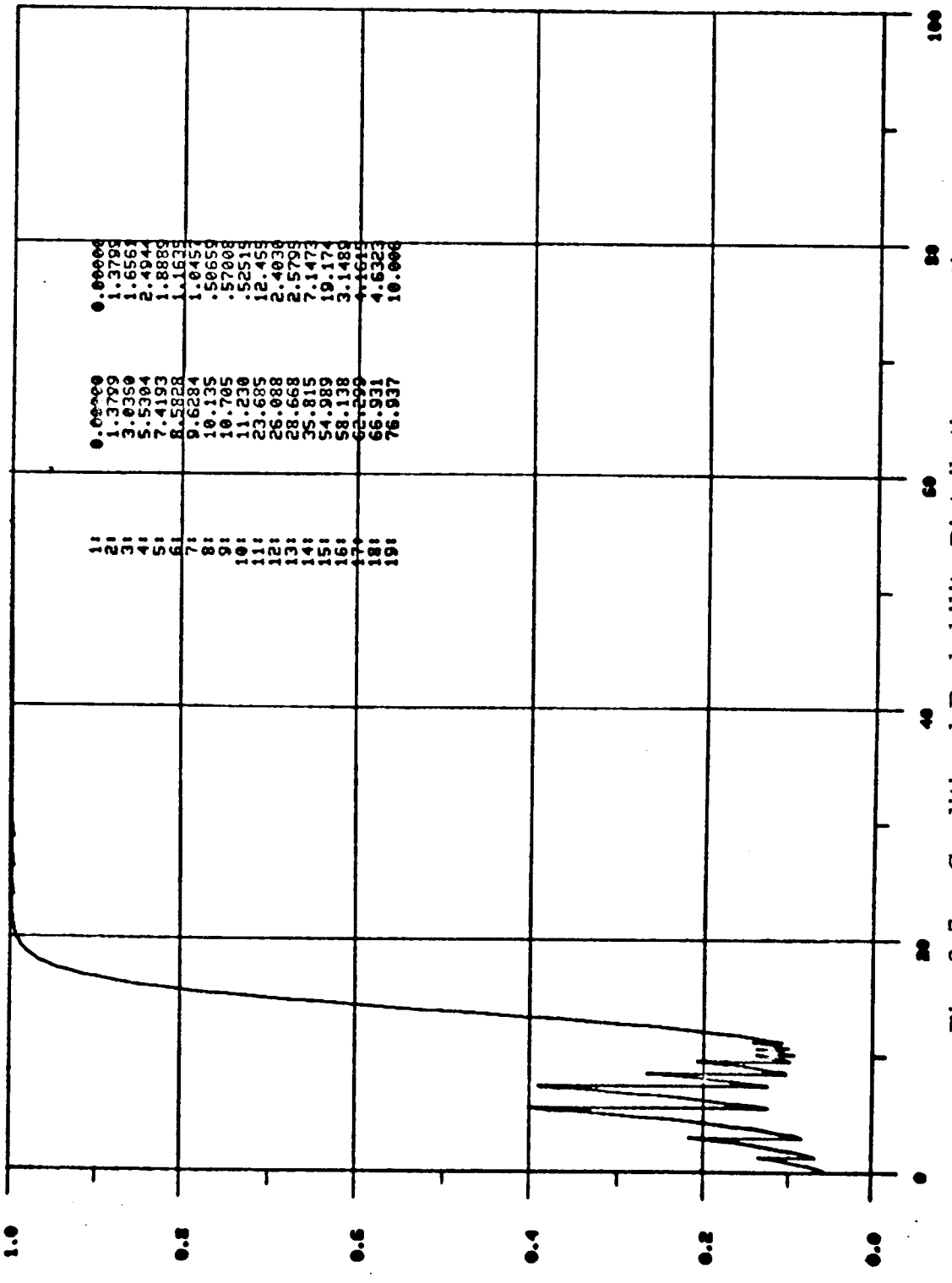


Fig. 2.7: Conditional Probability Distribution π_t vs. time
Case II, Rates $\lambda_0 = 0.9$ $\lambda_1 = 0.1$

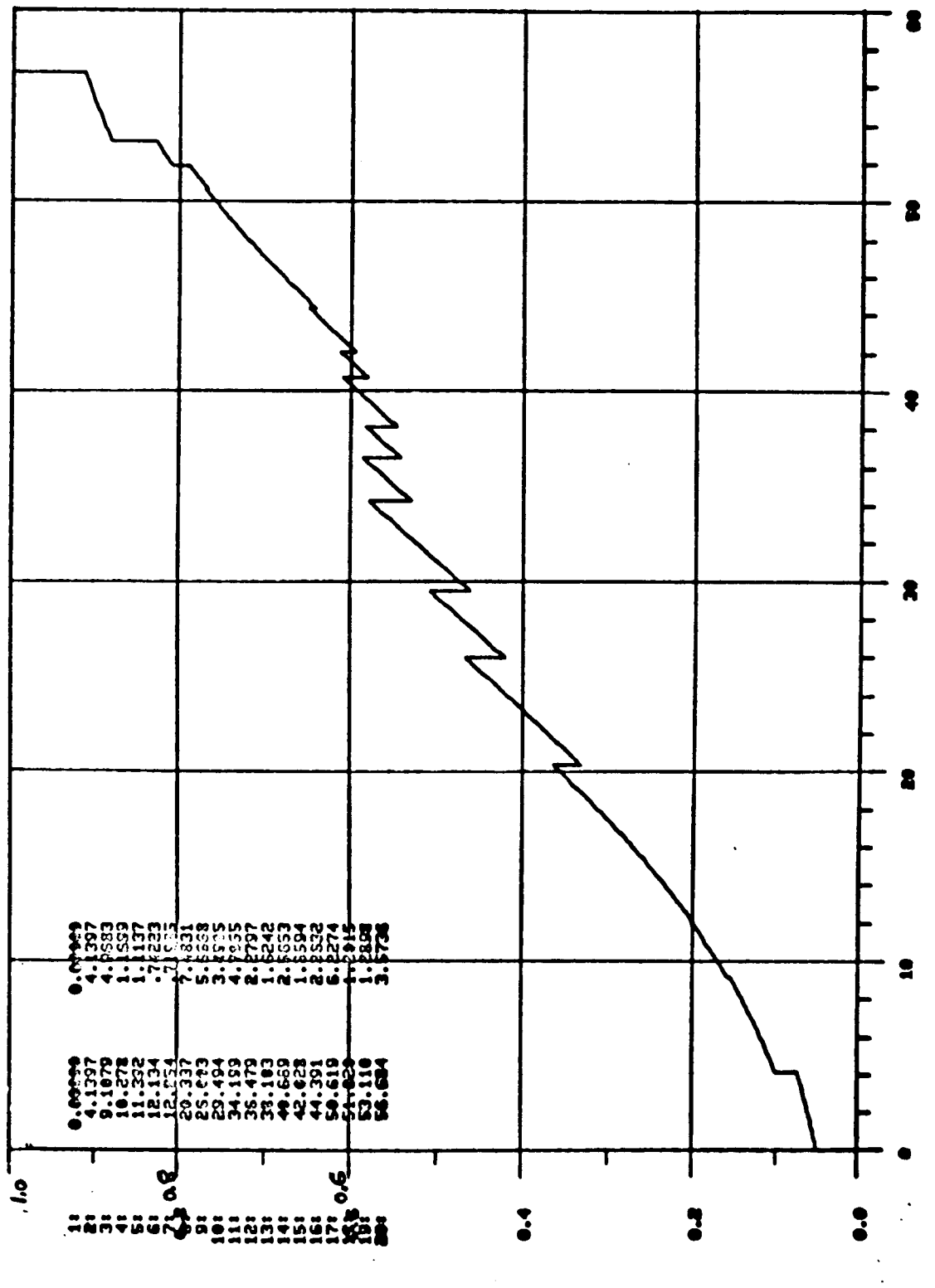


Fig. 2.8: Conditional Probability Distribution π_t vs. time
Case II, Rates $\lambda_0 = 0.3$ $\lambda_1 = 0.2$

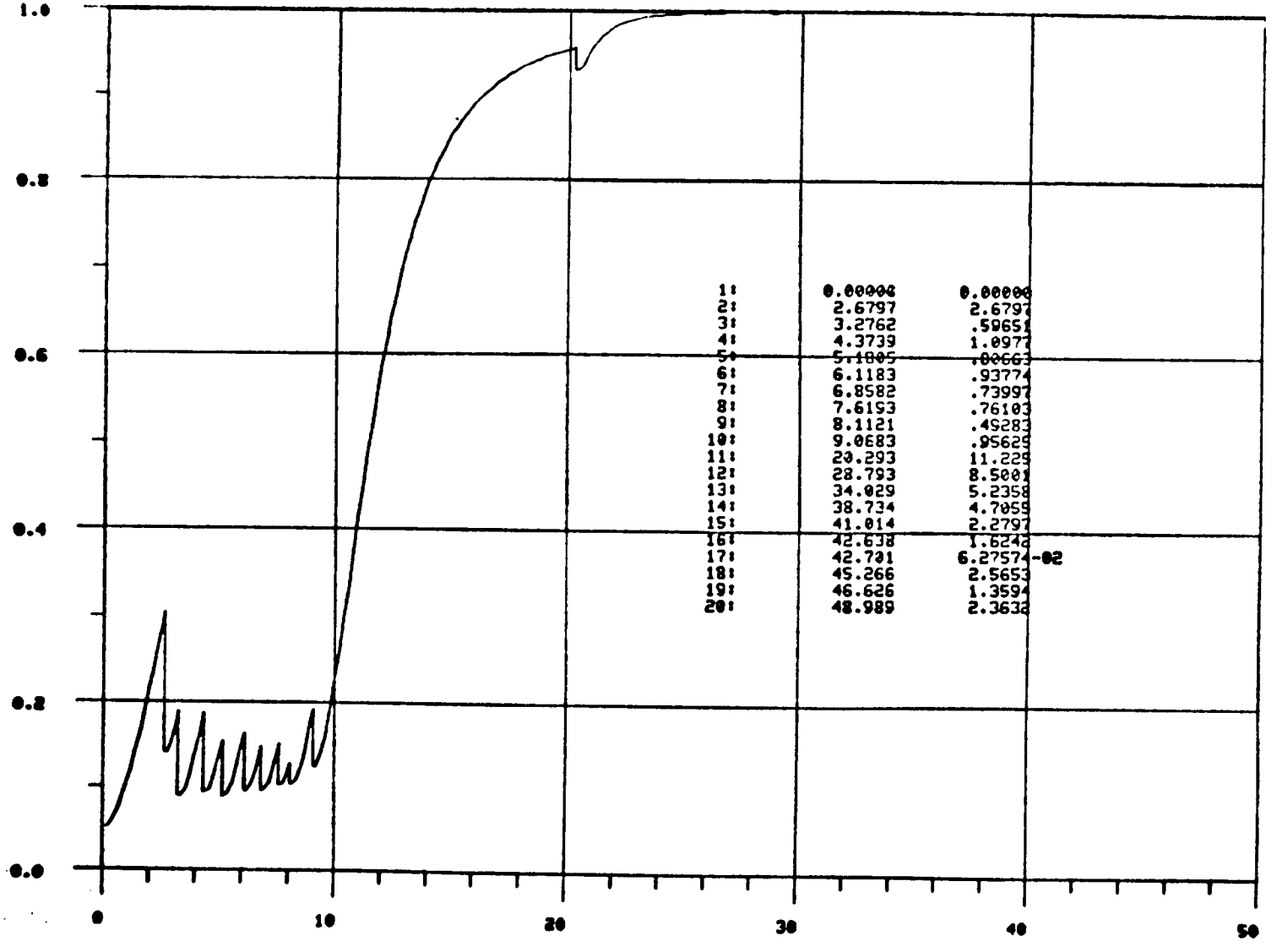


Fig. 2.9: Conditional Probability Distribution π_t vs. time

Case IV, Rates $\lambda_t^0 \sim L(-0.2412, 0.832)$ $\lambda_t^1 = 0.2$

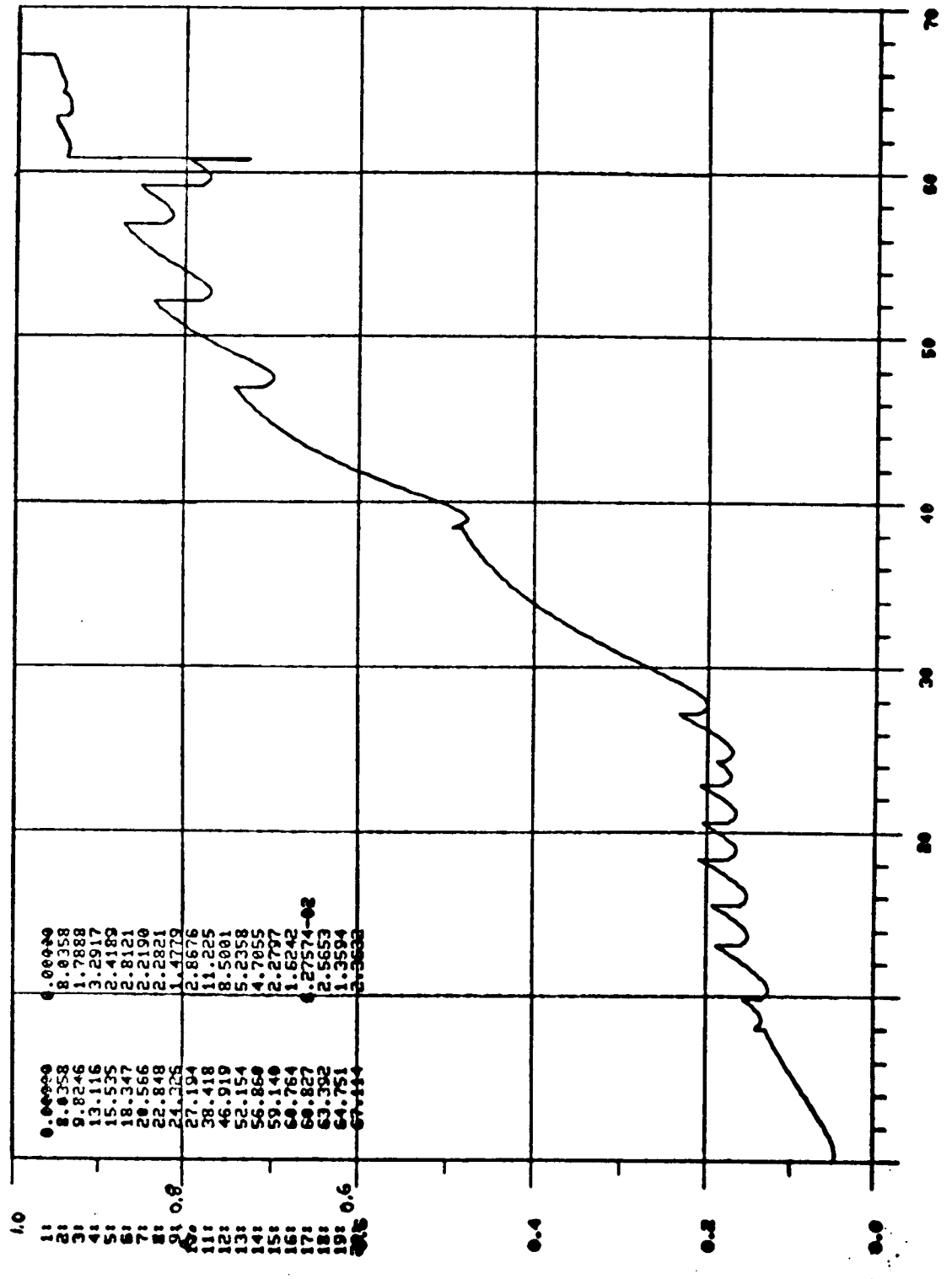


Fig. 2.10: Conditional Probability Distribution π_t vs. time
Case IV, Rates $\lambda_t^0 \sim L(0.857, 0.832)$ $\lambda_t^1 = 0.2$

To understand the effect of the different types of rates on the filter's behavior requires an examination of their associated hazard functions. As noted above, for a homogeneous Poisson process, the hazard function is constant and equal to the rate parameter, λ. The lognormal hazard function (2.2.12) is time varying and concave with respect to time (Fig. 2.11). Because of this concavity, the lognormal rate effects π_t^c more for shorter headways and less for longer headways. In Fig. 2.10, this fact is apparent for the time interval between W_9 through W_{11} where the slope of π_t^c decreases with larger interarrival times. But for a homogeneous Poisson rate, π_t^c increases linearily (Fig. 2.7, 2.8). Although from empirical data, following headways are modeled better by a lognormal density, certain considerations became apparent in the simulations. In Fig. 2.11 observe for short headways, it is possible for λ_t^1 to be greater than λ_t^0. This explains why after each activation time in Fig. 2.10, π_t^c momentarily decreases before increasing. However the major impact of these short headways occurs in π_t^d. For W_{16}-W_{17} in Fig. 2.10, the filter responds with a positive jump, i.e. increases the conditional probability that the switch has occurred. In other words, the filter is misled by this short headway. The discrepancy results from the assumption that the nonfollowing rate can be approximated by a constant, λ. More correctly, the nonfollowing headways can be approximated by a displaced negative exponential, i.e. $\lambda_t^1 = 0$ for short headways. For urban traffic (velocity $\approx$ 25 mph), it would be unrealistic to have headways less than 0.5 seconds because of time and space constraints between vehicles. Thus this would suffice as the displacement factor in λ_t^1 above. This modification was made in the rate and the associated discrepancy in the filter was removed. A

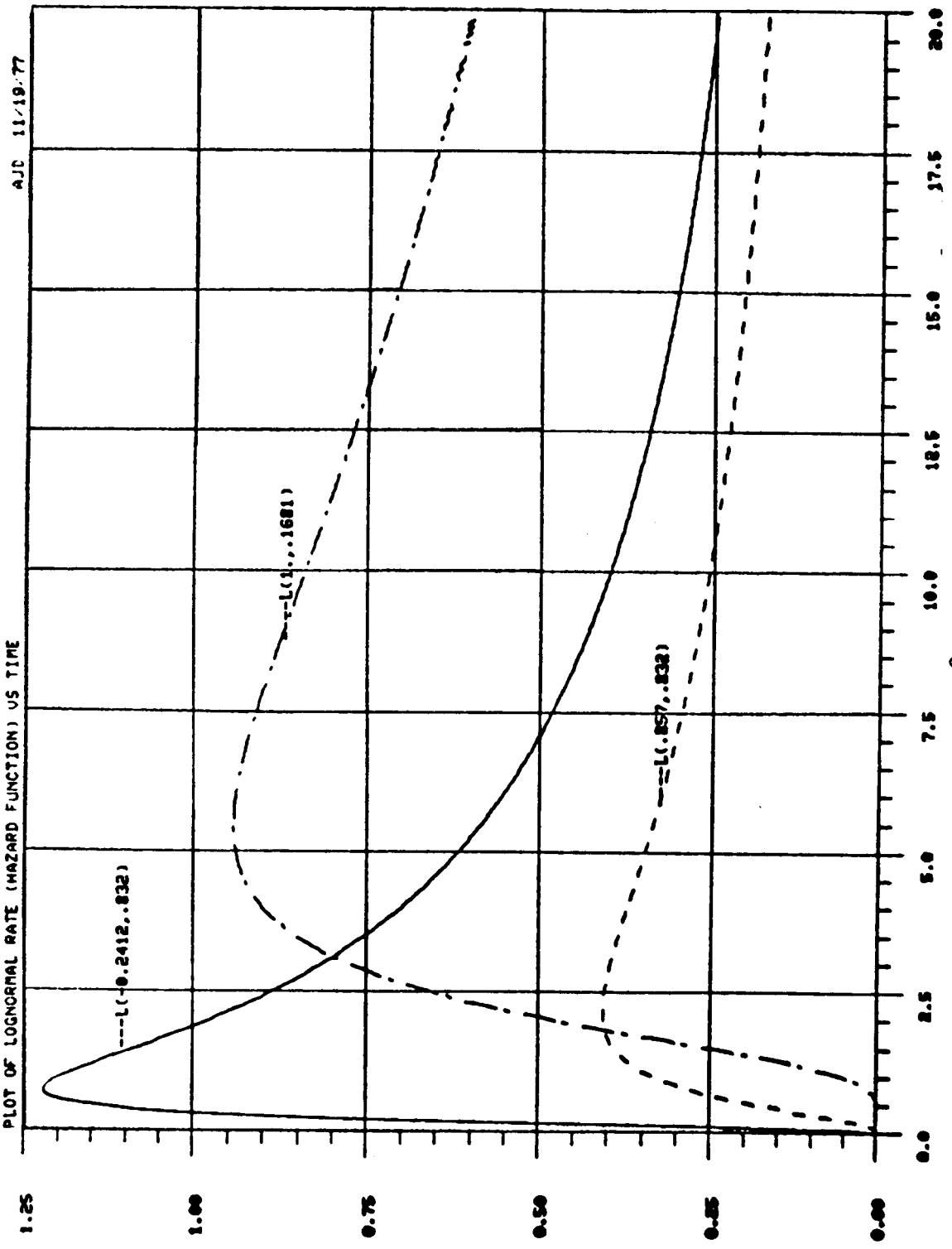


Fig. 2.11: Plot of Lognormal rate, λ_t^0 vs. time for different parameter values

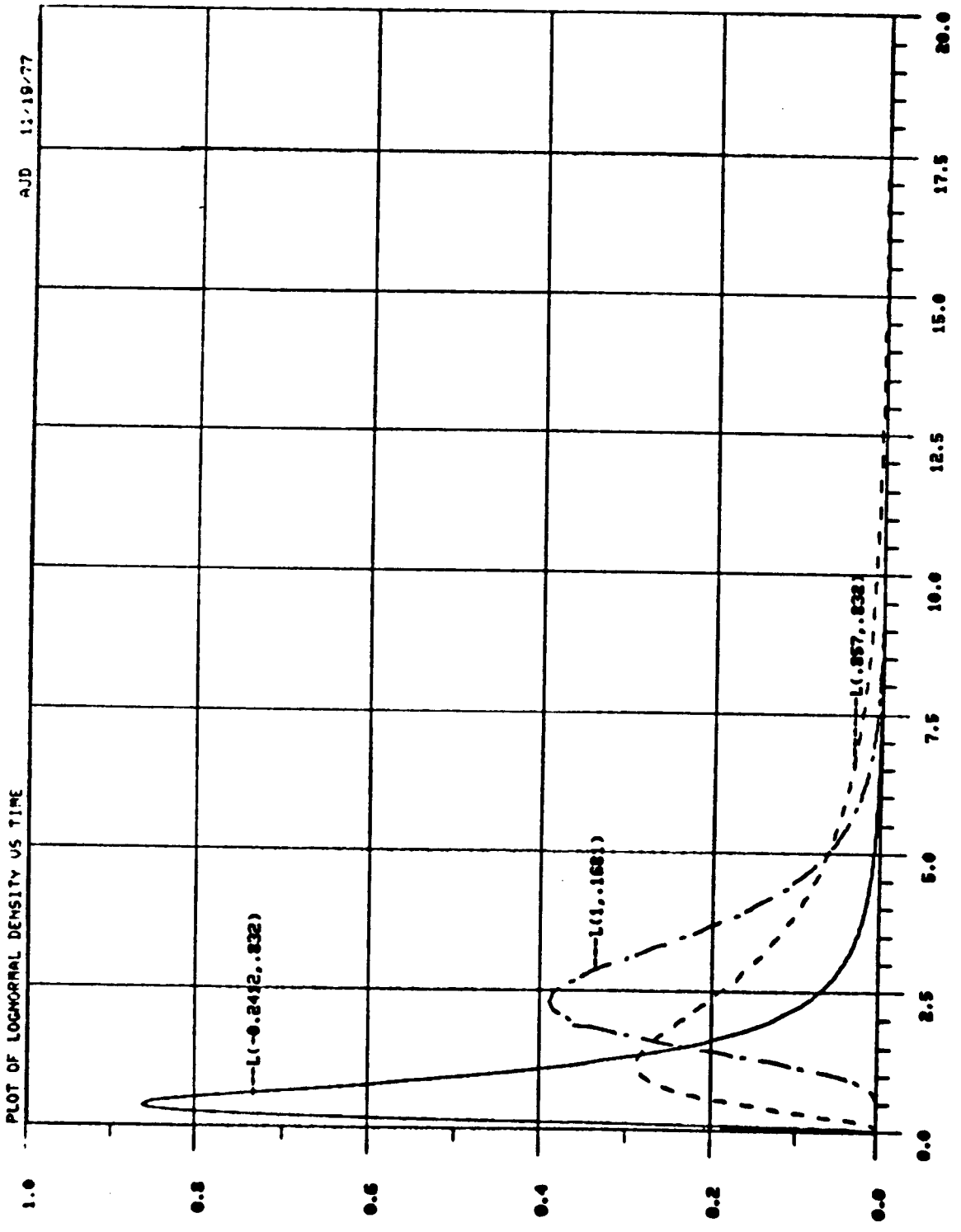


Fig. 2.12: Plot of Lognormal density vs. time for different parameter values

second observation can be made about the evaluation of λ_t^0 in (2.2.12). Because the numerator and denominator approach zero as $t \to \infty$ and the erf function is evaluated by a five (5) term series expansion, oscillations in λ_t^0 are observable (Fig. 2.11, L(1.0, 0.41) for t = 20). In particular, this hazard function has analytical difficulties for large interarrival times and the rate computed is only valid for headways less than these oscillations. Fortunately, this results in no shortcomings in the statistical model of headways nor the filter's response because for either the following or non-following processes, it is clear that only the neighborhoods of the means of each process is modelled reasonability correct; the outliers are generally discarded (see Section 3.2). These facts are brought to the attention of the reader to show the physical insight and the mathematical structure afforded the problem by point process theory.

Finally, the parameter variation of the lognormal rate is discussed. The convexity of the lognormal rate makes it possible to have a highly localized rate in one case while in the other a less peaked function can be obtained (Fig. 2.11). Consequently, the filter is more sensitive to parameter variations of the lognormal rate than the Poisson rate. A highly localized lognormal rate results in a greater difference in the two rates and therefore the slope of π_t^c and the variation of π_t^d are both greater (Fig. 2.9). Alternatively, for a less peaked lognormal rate, the difference in the two rates is more uniform and the slope of π_t^c and the variation of π_t^d are less drastic (Fig. 2.10). In other words, when the following headways have a small variance, the hazard function is more peaked and the filter distinguishes the switch time more readily. The "peakness" of the lognormal density is related to the ratio of its

mean and variance defined as:

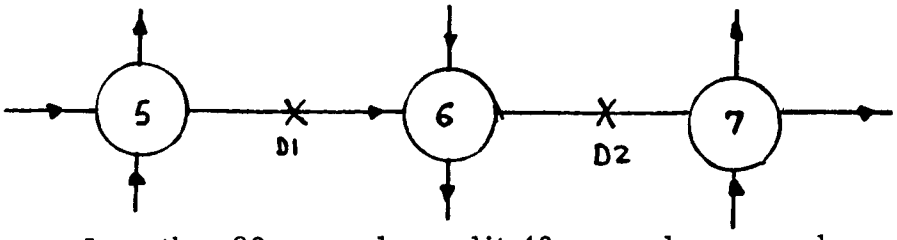
$$\text{Mean} = \exp\left(\mu + \tfrac{1}{2}\sigma^2\right)$$
$$\text{Variance} = \exp\left(2\mu + \sigma^2\right)\left[\exp(\sigma^2) - 1\right] \qquad (2.4.11)$$

For the same σ (=0.832) and different μ, the density is more localized for the negative μ (Fig. 2.12) and the resulting hazard function is more peaked (Fig. 2.11). Similarly for the same μ and different σ, the rate is more peaked for the smaller σ (Fig. 2.11). In summary, a more localized interarrival time probability density gives a more localized hazard function and thereby increases the performance of the filter.

Having developed an understanding of the filter's dependence on its a priori density and the rates, we shall present the results of the filter to simulated urban traffic. The Urban Traffic Control System (UTCS), developed for the U.S. Federal Highway Administration, simulates various urban traffic networks and its use and validity are well-documented [47, 48]. From the UTCS, we obtained the queue length every second, the signal code of the traffic light, and the activation time and velocity of each vehicle crossing a detector location. Two different street configurations were considered (see Fig. 2.13, 2.14).

The filter discussed above with the switch time a priori density independent of the observations and λ_t^1 equal to a constant, λ was evaluated. For any given cycle, an average vehicle discharge rate of 3 seconds and a green phase of 40 seconds, meant that no more than 15 queued vehicles could pass a detector loaction. Thus the a priori density was uniformly distributed between 0 to 15 occurences. The parameters λ, μ and σ were held fixed during this phase of the evaluation. The variance (σ =0.41) in the lognormal was chosen to match Branston's empirical data [29, p. 143].

Fig. 2.13 Characteristics of Test Network #1

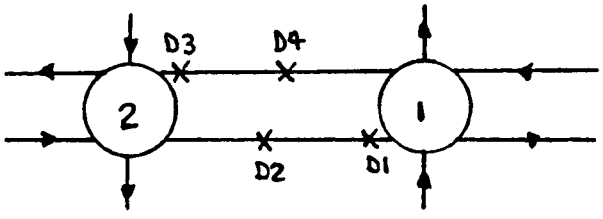


Cycle: Length = 80 seconds, split 40 seconds, no amber
Node 6 offset 20 seconds relative to Node 5
Node 7 offset 40 seconds relative to Node 5

Link: Length = 500 feet, one-way
Headway discharge rate = 3.0 seconds

Detector: D1 (D2) are located 210 feet from the stopline of Node 5 (6).

Fig. 2.14 Characteristics of Test Network #2



Cycle: Length = 80 seconds, split 40 seconds, no amber
Node 2 offset 20 seconds relative to Node 1

Link: Length = 800 feet, two lanes, two-way
Headway discharge rate = 3.0 seconds

Detector: D1 (D3) are located 26 feet from stopline of Node 1 (2)
D2 (D4) are located 210 feet from stopline of Node 1 (2)
(Each detector location denotes two separate detectors, one for each lane).

The associated mean ($\mu = 1.0$) was selected so that the average headway discharge rate and the average following headway (eq. 2.4.11) were equal. The Poisson rate ($\lambda = 0.1$) was selected to yield an average nonfollowing headway of 10 seconds. These parameter selections were arbitrary and could have been replaced by using the moment estimates of the headway data. However, because of the small sample size these parameter estimates would not have been reliable. Finally, a parameter not mentioned previously, the detector location was selected following the recommendation of a study utilizing the UTCS, which found that the detector placement can critically effect the performance of a traffic controller strategy [49]. Clearly at a stopline location, we should observe a greater rate change than at a detector located downstream (e.g. D2 in Fig. 2.13) where the traffic flow is more regular.

For the ease of comparing the filter's maximum jump estimate to the actual queue size, we tabulate the results below:

Table 2.1 Link(2, 1) - Lane 2 - Stopline detector D1

Cycle Number	Actual Queue	Estimate	Deciding T_n (sec)
1	6	7	24.2
2	8	8	24.4
3	7	7	8.0
4*	7	7	24.2
5	9	9	20.4
6	7	6	**

* Switch time condition distribution is shown.

Table 2.2 Link(6, 7)-Lane 1-Detector D2

Cycle Number	Actual Queue	Estimate	Deciding T_n (sec)
1	2	11	19.9
2	2	2	6.5
3	3	3	9.9
4*	1	10	18.7
5	1	11	**
6	2	3	8.8

Table 2.3 Link(1, 2)-Lane 1-Stopline detector D3

Cycle Number	Actual Queue	Estimate
1*	0	6
2	1	4
3	1	4
4	1	8
5	1	6
6	0	4

For a large and distinct queue, the maximum jump estimate gives generally the exact estimate of the queue size (Table 2.1). The definition of a distinct queue is done by way of a typical cycle, e.g. the fourth. In Fig. 2.15, observe that at a stopline detector the vehicles leave a queue

** The interarrival time is either the default value because it is the end of the cycle or the end of the simulation.

with a fairly constant headway before the switch time estimate. As we conjectured, a lognormal density for describing headways implies this degree of regularity. When the first T_n is abnormal ($T_5 = 24.2$), the filter interprets a rate chage has occurred and the estimate obtained is related to the subsequent activation time and vehicle position. For a small queue, the filter gives a less satisfactory estimate (Table 2.2). Since the observations are 290 feet dowstream from the light, the switch time is more difficult to determine. The estimate for the first, fourth and fifth cycles show a significant error. Considering the fourth cycle, for a one vehicle queue, one would expect a large interarrival time after the first activation. In Fig. 2.16, observe that the headways are nearly the same until W_9 whereupon with $T_{10} = 18.7$ seconds, the filter estimates a queue of 10 vehicles. Although this appears as an incorrect estimate of the queue, it results because after the 1^{st} vehicle leaves the street, a platoon of nine (9) vehicles joins the first by the time it reachs detector D2. This situation occurs frequently in urban traffic; the queue is not empty before an upcoming platoon of vehicles joins it. One can improve the filter's performance in this regard by making the a priori density dependent on the past observations. Then the associated q_t provided by (2.3.22), (2.3.24) would depend on not only a particular activation but also on the amount of time since the previous activation. However for a stopline detector with a small queue, the above would not necessarily improve the filter (Table 2.3). Since detector D1 is 26 feet from the stopline at least three (3) vehicles are required in the queue (one interarrival time) before the filter can be utilized. This results because the filter must be governed by a lognormal rate λ_t^0 before the switch time and governed by a Poisson rate λ_t^1 afterwards.

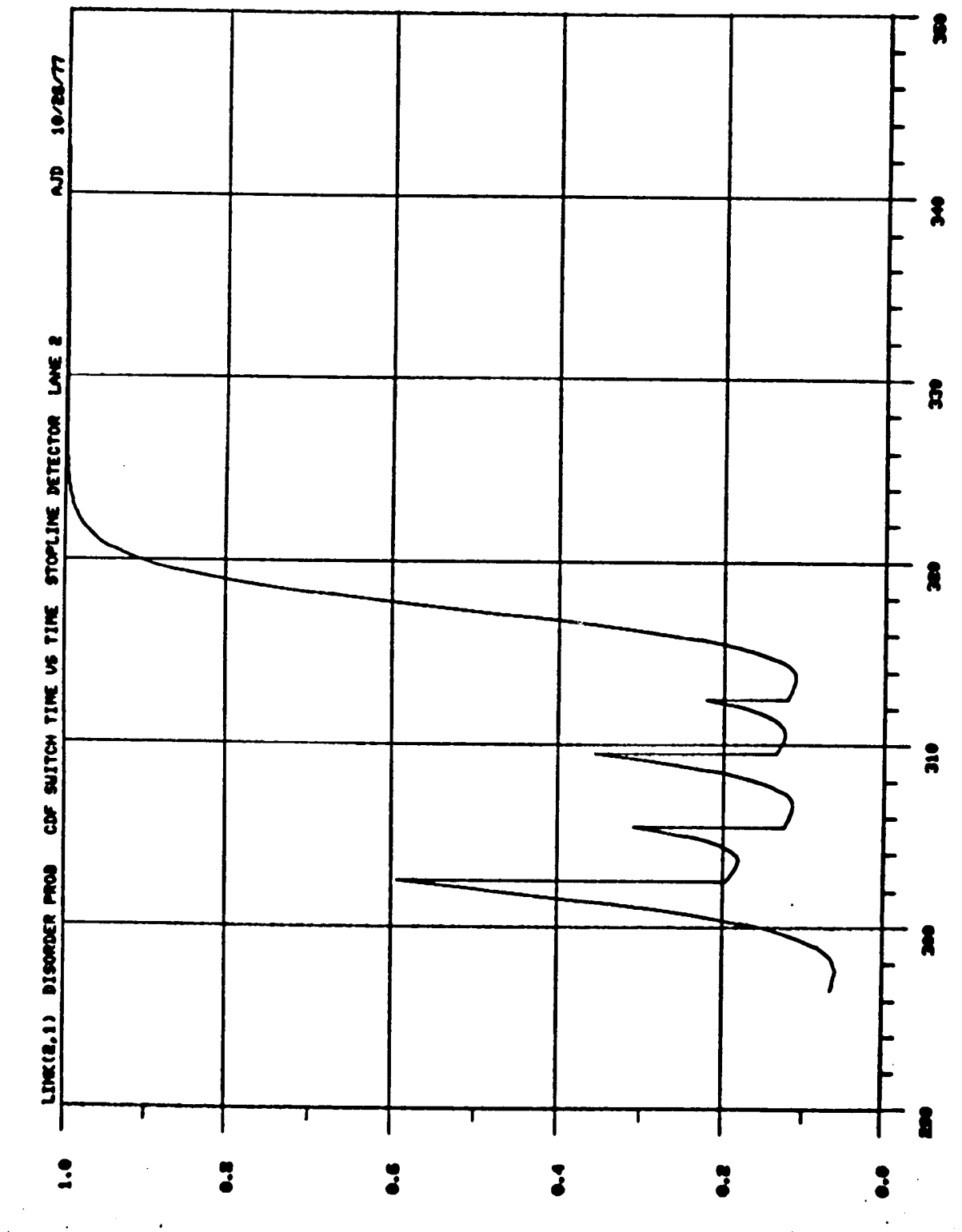


Fig. 2.15: Estimate of the Conditional Probability that Platoon has Passed Link(2, 1)-Lane 2-Stopline detector D1-Cycle 4.

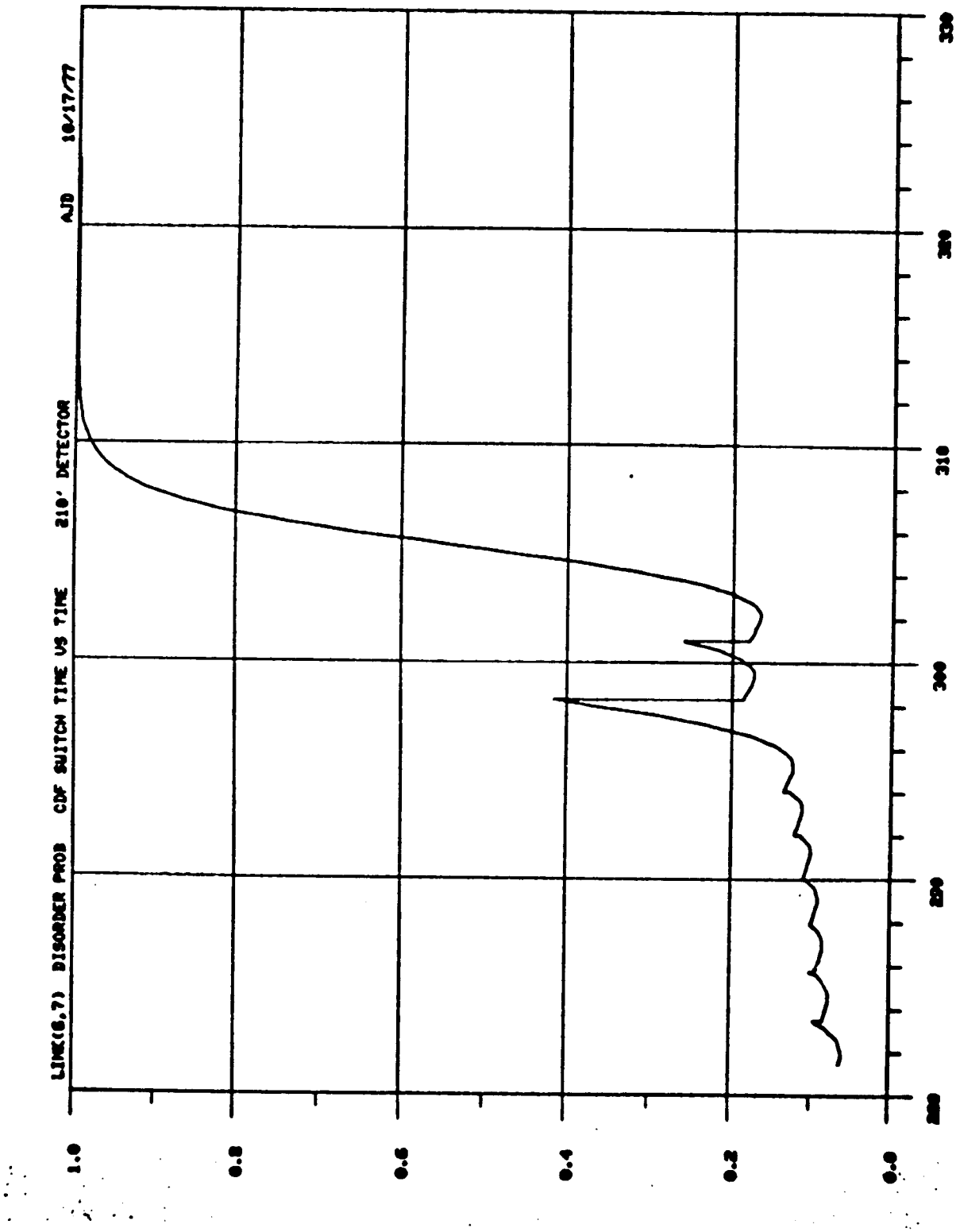


Fig. 2.16: Estimate of the Conditional Probability that Platoon has Passed Link(6, 7)-Lane 1-Detector D2-Cycle 4.

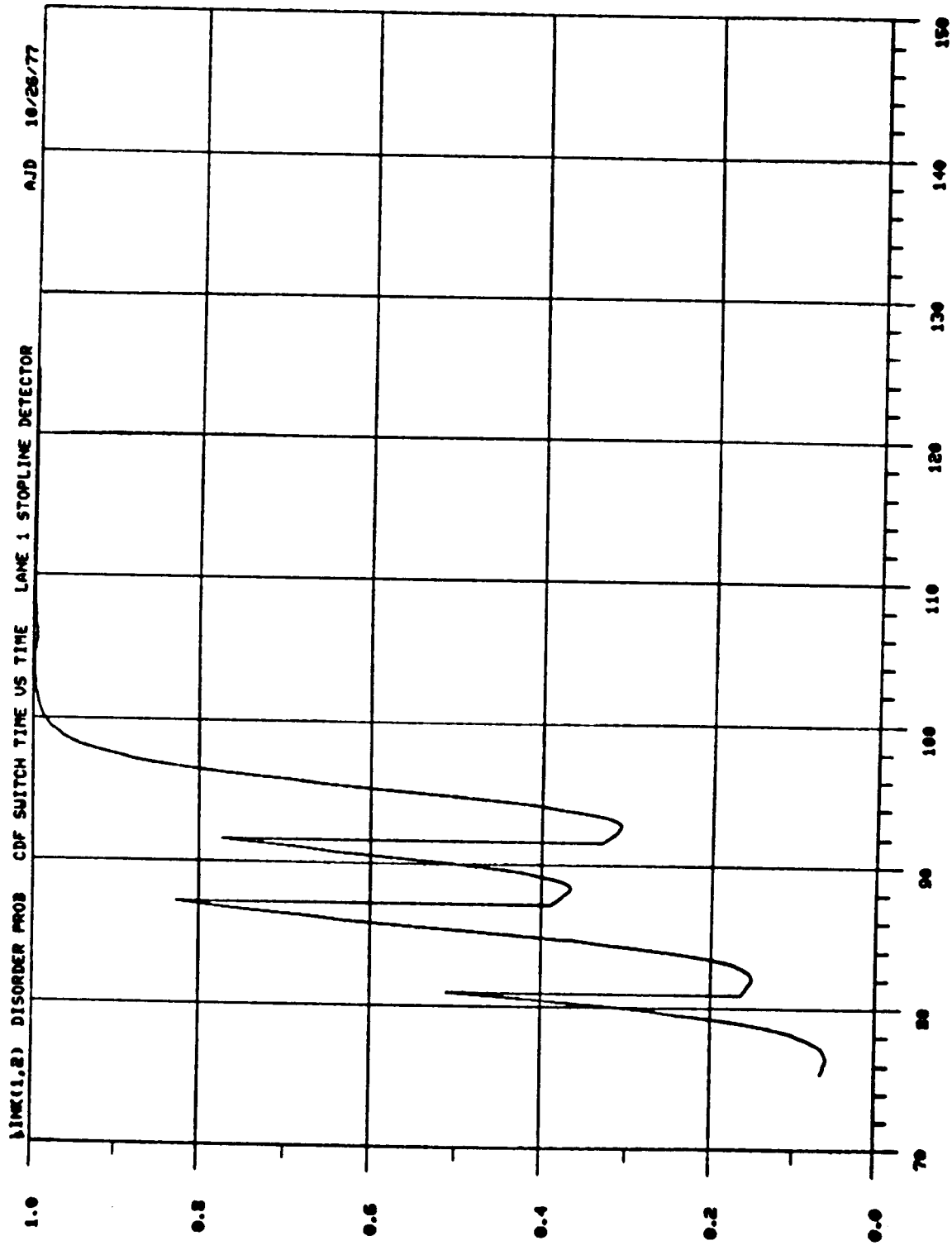


Fig. 2.17: Estimate of the Conditional Probability that Platoon has Passed Link(1,2)-Lane 1-Stopline detector D3-Cycle 1.

Considering the first cycle (Fig. 2.17) with no vehicles in the queue, the estimation problem is ill-posed. Therefore to avoid this situation, one must position the detector closer to the stopline and utilize some knowledge about the queue in the a priori density.

In conclusion, the effectiveness of the disorder problem to this particular traffic problem is somewhat guarded. Clearly, the assumptions concerning the two rates did not adversely affect the filter's performance. Naturally by "fine-tuning" the filter to these parameters would have improved the resulting estimates. The extend of such techniques are discussed in a subsequent chapter. The a priori density and the associated q_t did affect the filter's accuracy. Since p_n was uniform, the filter depended more heavily on the counting process rates. For a distinct queue, the change in the rate was easily distinguished. When the queue was small or when vehicles override a queue, q_t provided by (2.3.16) did not incorporate the necessary influence to the filter. Therefore, the associated estimates were unreliable. Improvements in this regard can be obtained by using (2.3.24) for q_t and making the p_n less uniform. One interesting feature of the filter was an improved criteria for separating platoons. In Table 3.1 and 3.2, the deciding interarrival time varied with each traffic light cycle.

3. PARAMETER SENSITIVITY AND ESTIMATION

3.1 Introduction

The estimation problem considered here can be divided into two areas: parameter estimation and sequential estimation. From the historical data, in the sense that all vehicular traffic follows the same characteristic pattern, the statistical model of headways was developed. The model is complete once the estimates of the four parameters are specified. Utilizing this model a sequential traffic flow estimator was developed which used current detector outputs to update the estimator, π_t . This latter estimator specializes the historical data to a particular link in a traffic network. Yet, to be addressed is the affect of uncertainty in the statistical headway model, i.e., parameter estimates, on the performance of the traffic estimate.

By assuming the following headways have a lognormal density one obtains the robustness of the associated gaussian process, $\ell n\ h_i$. Reducing the nonfollowing density from a convolution to a negative exponential eliminated both computational and statistical difficulties. These assumptions have simplified the determination of the parameter estimators. However, these estimates are probabilistic and, consequently, there is a margin of uncertainty between the estimate and the parameter. The amount of error in any given parameter can affect the filter's performance differently.

In the discussion which follows we shall assume that the parameters are known exactly. In the first section we discuss the affect of a nominal change in a parameter on the filter's performance. The sensitivity analysis of the conditional variance is shown to be an important tool in this regard. Subsequently, an error bound for one activation interval $(W_{n-1}\ ,\ W_n]$ to

parameter error is developed. In the next section, parameter estimation is discussed. The degree of uncertainty is determined from the sufficient statistics of these parameter estimators. The emphasis here being not on the exact estimate, but rather on the range in which the parameter estimate may lie within a certain confidence level. These results are then interpreted against nomial parameter variation to the maximum jump estimate defined above. To conclude this chapter, areas of future research and extension of these results are considered.

3.2 Sensitivity Analysis

In the development of the traffic estimator, we desired a model for the signal (2.3.18) and sufficiently simple so that the resulting algorithm was computationally feasible (2.3.21). The rate processes essential in the formulation were desired to be indicative of the actual processes involved. Though developed (2.2.11), (2.2.14), the complete probalistic description of these processes are computationally difficult to determine. Consequently, unaccountable errors are introduced into the signal and observation models. A major tool in determining the affect of these errors on the filter's performance is a sensitivity analysis of the conditional error covariance.

The conditional mean, π_t , defined in (2.4.1) is the minimum mean square error estimate of the signal, x_t . In particular π_t was chosen to

$$\min \quad E[(x_t - \hat{x}_t)^2]$$

where the minimization is over the set of estimators of x_t. Let the actual signal and observation processes be modeled by (2.3.1.) and (2.3.2), and suppose we <u>assume</u> the signal and observation processes are modeled by

$$d\bar{x}_t = \bar{f}_t\, dt + d\bar{v}_t$$

$$dN_t = \bar{\lambda}_t\, dt + d\bar{w}_t \qquad (3.2.1)$$

Define the actual and assumed conditional error variance

$$V_t = E[(x_t - \hat{x}_t)^2 | F_t]$$

$$\bar{V}_t = E[(x_t - \hat{\bar{x}}_t)^2 | F_t] \qquad (3.2.2.)$$

The algorithm presented for the optimal estimate $\hat{x}_t$ yields the minimum cost function only if the model represents the actual process. Otherwise the actual and assumed conditional errors variance are different. Consequently, when the incorrect model is utilized, the conditional error variance gives an indication of the error, and therefore is incorporated in an error analysis and sensitivity study. The analysis development using the conditional variance is dependent though on the observation algebra, F_t .

Because of our state equation (2.3.5), the conditional error variance reduced to a simple expression

$$V_t = E[(x_t - \hat{x}_t)^2 | F_t] = E[x_t^2 | \bar{F}_t] - \hat{x}_t^2 = \pi_t(1-\pi_t) \qquad (3.2.3)$$

where we have used π_t defined in (2.4.1). Large-scale sensitivity analysis of the conditional error variance can be defined

$$\frac{V_t - \bar{V}_t}{\Delta\theta} = \frac{(1-\pi_t)\pi_t - (1-\bar{\pi}_t)\bar{\pi}_t}{\Delta\theta} \qquad (3.2.4)$$

where $\Delta\theta$ represents the difference between a parameter in the actual process and the same parameter in the assumed process. Similar small-scale analysis can be defined:

$$\frac{\delta V_t}{\delta\theta}\Big|_\theta = \frac{\delta}{\delta\theta}[(1-\pi_t)\pi_t]\Big|_\theta \tag{3.2.5}$$

where $\delta\theta$ has a similar interpretation as $\Delta\theta$ above. The advantage and results of either method can be different. In small-scale analysis, the evaluation may be easier than large-scale analysis because typically V_t is an involved expression. For the model provided in (2.3.1), (2.3.2), the conditional error variance has the representation [30, p. 349]

$$\begin{aligned} dV_t &= E[2(f_t-\hat{f}_t)(x_t-\hat{x}_t) + 1 \mid F_t]dt \\ &\quad + (\hat{\lambda}_t)^{-1} E[(x_t-\hat{x}_t)^2(\lambda_t-\hat{\lambda}_t) \mid F_t]\cdot(dN_t-\hat{\lambda}_t dt) \\ &\quad - (\hat{\lambda}_t)^{-2} E^2[(x_t-\hat{x}_t)(\lambda_t-\hat{\lambda}_t) \mid F_t]dN_t \end{aligned} \tag{3.2.6}$$

$$V_o = E(x_o^2) - E^2(x_o) = 0$$

However, large-scale sensitivity results are more useful [50, p. 386] especially when V_t is not differentiable in the parameter under study. Since π_t is a differentiable function of μ, λ, σ, we shall examine only small-scale analyses below.

The dependency of π_t on these parameters is through the rates (2.4.2), (2.4.3)

$$\pi_t^c = \left(\frac{1-\pi_{W_{n-1}}}{\pi_{W_{n-1}}}\right) \exp[\lambda_t^1(t-W_{n-1})]\left[1-\operatorname{erf}\left(\frac{\ell n(t-W_{n-1})-\mu}{\sigma}\right)\right] + 1^{-1}$$

$$\text{for } W_{n-1} < t < W_n \tag{3.2.7}$$

$$\pi_t^d = (1-\pi_{t-}) \frac{(\lambda_t^1-\lambda_t^0)\pi_{t-} + q_{t-}\lambda_{t-}^0}{(\lambda_t^1-\lambda_t^0)\pi_{t-} + \lambda_t^0} \tag{3.2.8}$$

where

$$\lambda_t^1 = \lambda$$

$$\lambda_t^0 = \frac{g(t-W_{n-1})}{1 - \text{erf}\left[\frac{\ell n(t-W_{n-1}) - \mu}{\sigma}\right]}$$

$$g(\cdot) = L(\mu, \sigma) \text{ density} \quad .$$

From (3.2.5), the parameter variation in the conditional error variance

$$\frac{\delta V_t}{\delta \theta} = (1-2\pi_t) \frac{\delta \pi_t}{\delta \theta} \tag{3.2.10}$$

is directly proportional to the variation in the conditional mean. Consequently, the sensitivity analysis has been reduced to studying the variation in the filter equations (3.2.7) - (3.2.9).

The approach we follow is to examine the continuous part and the discontinuous part of the estimate separately. Suppose on the interval $(W_{n-1}, W_n]$, the actual and assumed estimates π_t, $\bar{\pi}_t$ are identical at $t = W_{n-1}$. At $t = W_n$, we desire to know the amount of change

$$\Delta\pi_t = \pi_t - \bar{\pi}_t \tag{3.2.11}$$

that occurs in the estimate due to errors in the parameter estimates.

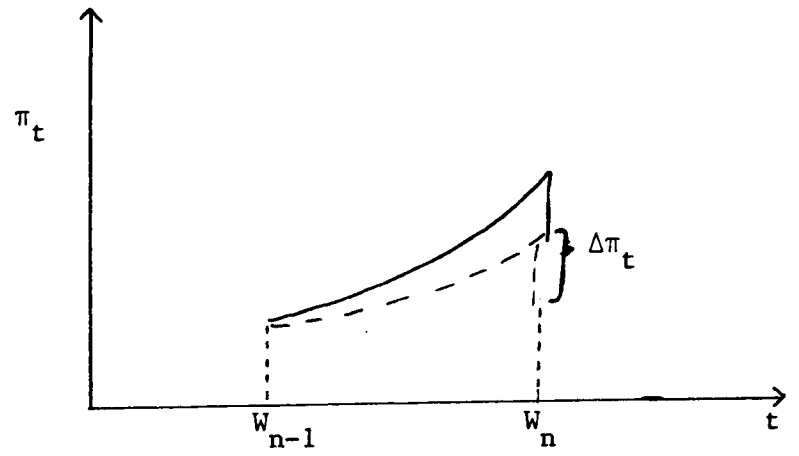


Fig. 3.1. Variation of Conditional Mean Due to Variation in Parameter

First we consider the variation of the continuous part π_t^c .

Variation wrt λ

Define

$$C_1 = \left(\frac{1-\pi_{W_{n-1}}}{\pi_{W_{n-1}}}\right) \left[1 - \text{erf}\left(\frac{\ell n(t-W_{n-1}) - \mu}{\sigma}\right)\right] \tag{3.2.12}$$

then by (3.2.7)

$$\pi_t^c = [C_1 \exp[\lambda(t-W_{n-1})] + 1]^{-1}$$

and

$$\frac{\delta\pi_t^c}{\delta\lambda} = \frac{-(t-W_{n-1})C_1 \exp[\lambda(t-W_{n-1})]}{[C_1 \exp[\lambda(t-W_{n-1})] + 1]^2} \tag{3.2.13}$$

$$\text{for } W_{n-1} < t < W_n$$

Before continuing, we develop two useful equalities.

Define

$$h(\mu, \sigma) = 1 - \text{erf}\left(\frac{\ell n(t-W_{n-1}) - \mu}{\sigma}\right) = \int_{\frac{\ell n(t-W_{n-1}) - \mu}{\sigma}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt \tag{3.2.14}$$

so that

$$\frac{\delta h(\mu, \sigma)}{\delta \mu} = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{-(\ell n(t-W_{n-1}) - \mu)^2}{2\sigma^2}\right]$$

$$= (t-W_{n-1})\, g(t-W_{n-1}) \tag{3.2.15}$$

$$\frac{\delta h(\mu, \sigma)}{\delta \sigma} = \frac{-1}{\sqrt{2\pi}} \exp\left[\frac{-(\ell n(t-W_{n-1}) - \mu)^2}{2\sigma^2}\right] \cdot \left[\frac{-(\ell n(t-W_{n-1}) - \mu)}{\sigma^2}\right]$$

$$= \left[\frac{\ell n(t-W_{n-1}) - \mu}{\sigma}\right] \frac{\delta}{\delta \mu} h(\mu, \sigma) \tag{3.2.16}$$

where $g(\cdot)$ is $L(\mu, \sigma)$ density provided in (1.2.13)

Variation wrt μ, σ

Define

$$C_2 = \left(\frac{1-\pi_{W_{n-1}}}{\pi_{W_{n-1}}}\right) \exp[\lambda_t^1 (t-W_{n-1})] \tag{3.2.17}$$

then by (3.2.7), (3.2.14)

$$\pi_t^c = [C_2 h(\mu, \sigma) + 1]^{-1}$$

and

$$\frac{\delta\pi_t}{\delta\mu} = \frac{-C_2 \frac{\delta}{\delta\mu}[h(\mu,\sigma)]}{[C_2\, h(\mu,\sigma)+1]^2} = \frac{-C_2(t-W_{n-1})g(t-W_{n-1})}{[C_2\, h(\mu,\sigma)+1]^2} \tag{3.2.18}$$

by (3.2.15)

$$\frac{\delta\pi_t}{\delta\sigma} = \frac{-C_2 \frac{\delta}{\delta\sigma}[h(\mu,\sigma)]}{[C_2\, h(\mu,\sigma)+1]^2}$$

$$= \frac{\ell n(t-W_{n-1}) - \mu}{\sigma} \left(\frac{\delta\pi_t}{\delta\mu}\right) \tag{3.2.19}$$

by (3.2.16) .

The variations above all have the same denominator and since C_1 , C_2 are positive, π_t^c is a decreasing function of λ , μ (3.2.13), (3.2.18). Therefore, if $\pi_t^c < \frac{1}{2}$, then overestimating λ or μ causes the conditional error variance (continuous part) V_t^c to be smaller (3.2.10) and, conversely, if $\pi_t^c > \frac{1}{2}$. By (3.2.19), π_t^c is a decreasing function of σ for $\mu < \ell n(t-W_{n-1})$ and an increasing function for the reverse inequality. Similar comments can be made about the affect of σ on the conditional error variance V_t^c . If

$$h(\mu,\sigma) = \int_{t-W_{n-1}}^{\infty} g(x)dx \geq g(t-W_{n-1}) \tag{3.2.20}$$

then from (3.2.12), (3.2.13) and (3.2.17), (3.2.18)

$$\frac{\delta\pi_t^c}{\delta\lambda} - \frac{\delta\pi_t^c}{\delta\mu} = -(t-W_{n-1}) \frac{[h(\mu,\sigma) - g(t-W_{n-1})]}{[C_2\, h(\mu,\sigma) + 1]^2} \leq 0$$

Since $\frac{\delta\pi_t^c}{\delta\lambda}$, $\frac{\delta\pi_t^c}{\delta\mu}$ are negative

$$\left|\frac{\delta\pi_t^c}{\delta\lambda}\right| \geq \left|\frac{\delta\pi_t^c}{\delta\mu}\right| \qquad \text{for all} \quad t \geq 0 \tag{3.2.21}$$

By (3.2.19), $\frac{\delta\pi_t^c}{\delta\sigma}$ is negative for $\ell n(t-W_{n-1}) \geq \mu$ and positive for the reverse inequality, so by (3.2.21)

$$\left|\frac{\delta\pi_t^c}{\delta\lambda}\right| \geq \left|\frac{\delta\pi_t^c}{\delta\mu}\right| \geq \left|\frac{\delta\pi_t^c}{\delta\sigma}\right| \qquad \text{if } [\ell n(t-W_{n-1}) - \mu] < \sigma$$

$$\left|\frac{\delta\pi_t^c}{\delta\mu}\right| \leq \left|\frac{\delta\pi_t^c}{\delta\lambda}\right| \leq \left|\frac{\delta\pi_t^c}{\delta\sigma}\right| \qquad \text{if } [\ell n(t-W_{n-1}) - \mu] \geq \sigma \tag{3.2.22}$$

From (3.2.22) observe that the filter is <u>not</u> dominated by any one parameter for all time. Instead, the filter's sensitivity is divided into two regions. For long headways, the filter is more sensitive to variations in σ . This fact can be deduced from the lognormal rate (Fig. 2.11). For the same variance (σ = .832) and different mean, the rate has the same asympotic shape. But for the same mean (approx.) and different variance, there is a more drastic change. Recall, the localization of the lognormal rate depends on the ratio of its mean and variance (2.4.11) which can clearly be seen to be more sensitive to σ than μ . For short headways, variations in λ dominate the filter's sensitivity provided (3.2.20) is satisfied. This is reasonable since in this region, the filter attempts to distinguish the following and nonfollowing headways. Representative curves of (3.2.13), (3.2.18) and (3.2.19) are shown in Fig. 3.2, 3.3, and 3.4, respectively. The initial condition $\pi_{W_{n-1}}$ (= PIN($\cdot$)) in (3.2.12) and

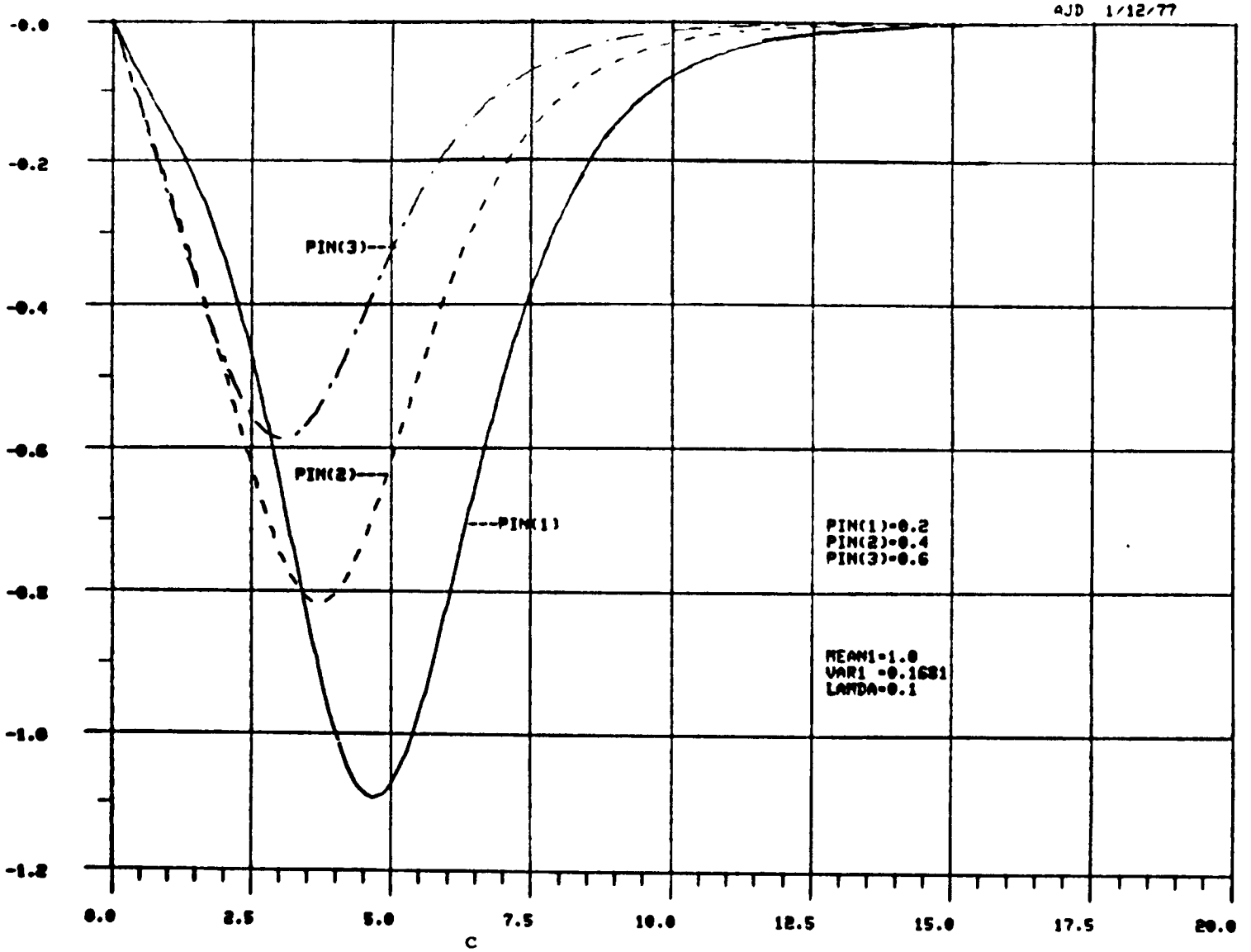


Fig. 3.2 : Plot of $\left(\frac{\delta \pi_t^c}{\delta \lambda}\right)$ vs. time

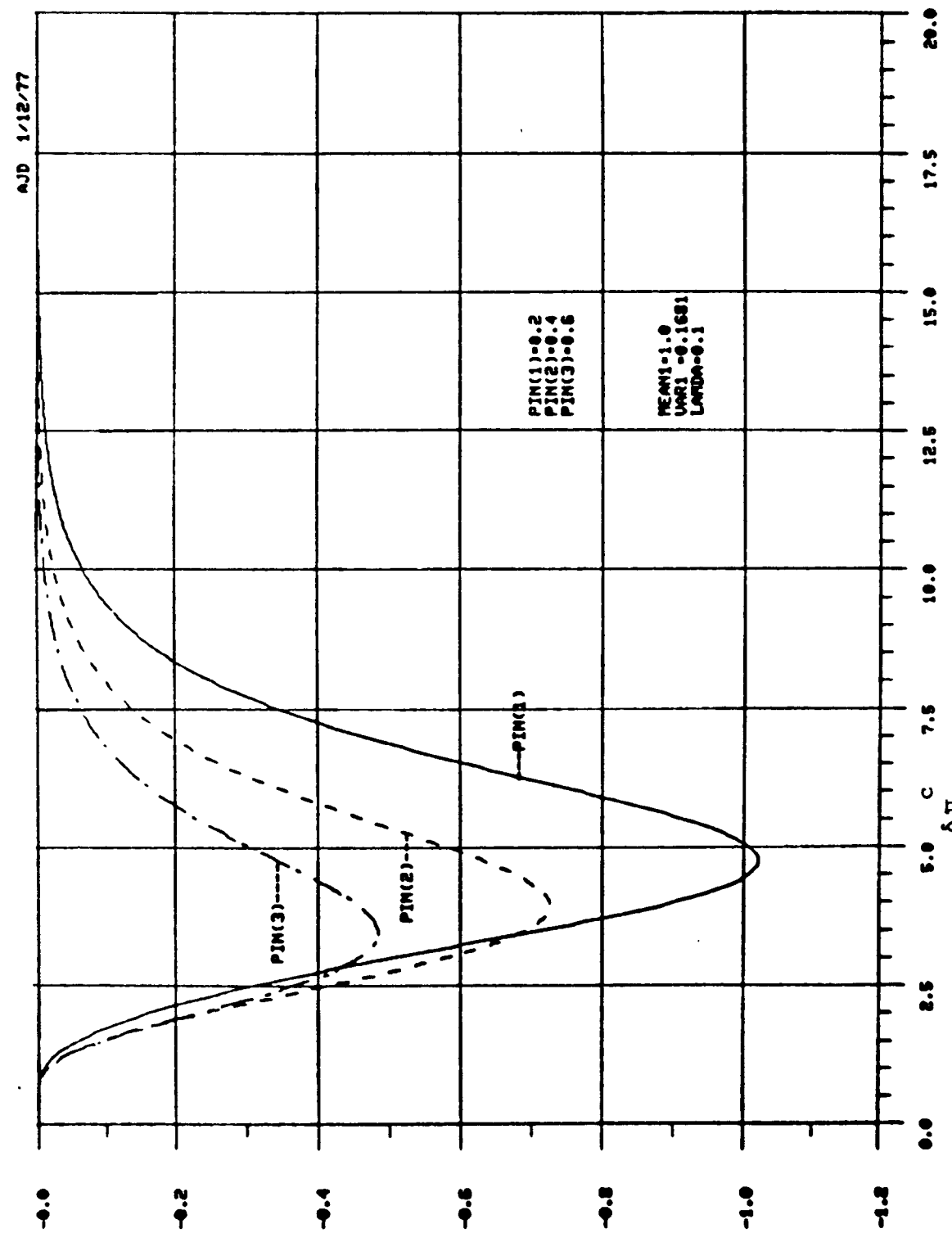


Fig. 3.3: Plot of $\left(\frac{\delta \pi_t}{\delta \mu}\right)$ vs time

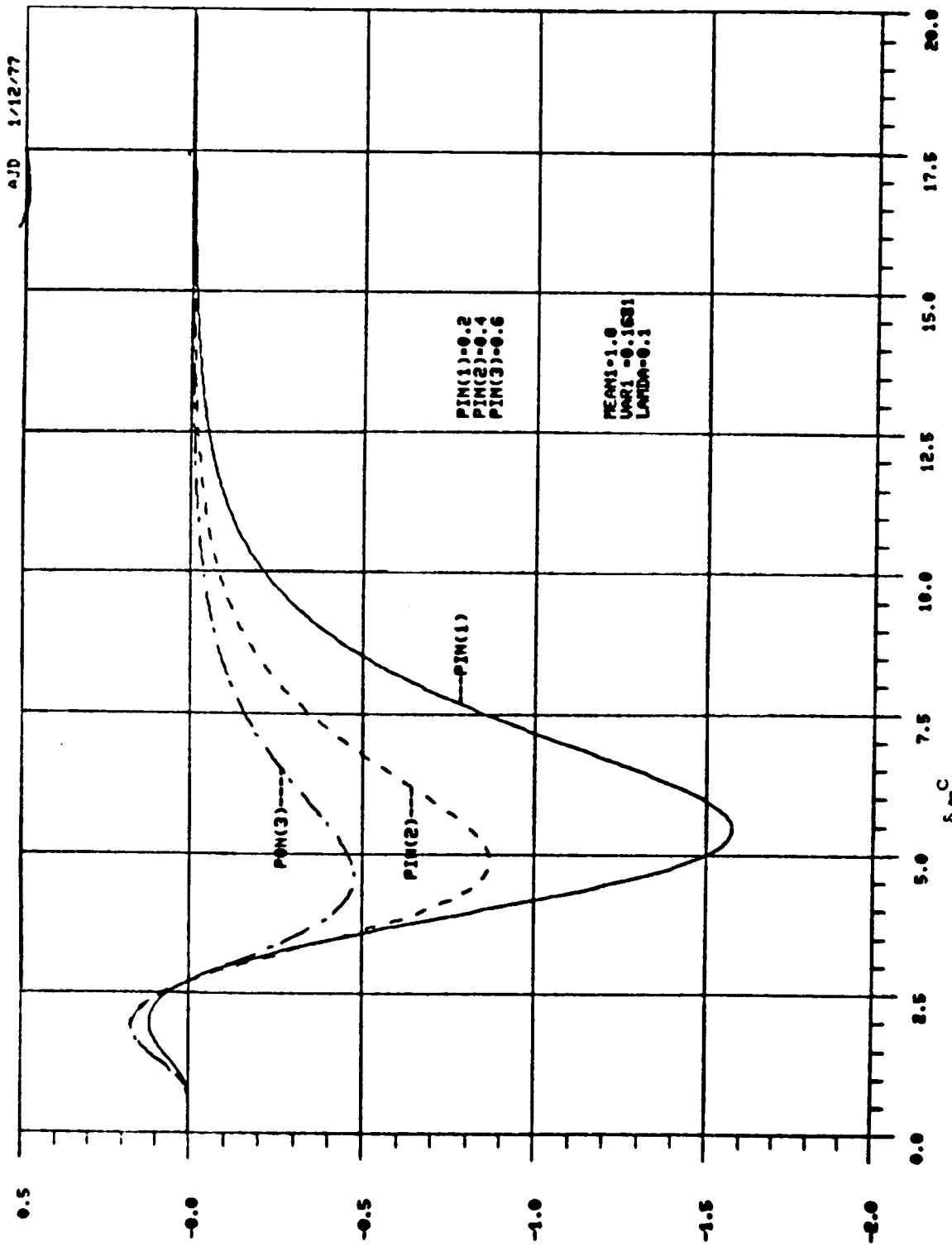


Fig. 3.4: Plot of $\left(\frac{\delta \pi_t^c}{\delta \sigma}\right)$ vs. time

(3.2.17) is varied as shown. The relation (3.2.20) plotted in Fig. 3.5 simplifies the results (3.2.22). However, if not satisfied, (e.g., L(-0.2412 , 0.832)), then (3.2.21) is no longer true and, consequently, the filter's sensitivity can be divided into three regions. In particular, for short headways variations in μ dominate, for medium headways variations in λ dominate, and for long headways variations in σ dominate. In summary though, if (3.2.20) is satisfied since σ , λ are always less than 0.5 in real traffic, the error introduced by the mean is insignificant.

The variation of the continuous part of the conditional error variance, V_t^c can be easily determined by (3.2.10). By (3.2.22) for $\pi_t < \frac{1}{2}$

$$\left|\frac{\delta V_t^c}{\delta \lambda}\right| \geq \left|\frac{\delta V_t^c}{\delta \mu}\right| \geq \left|\frac{\delta V_t^c}{\delta \sigma}\right| \qquad \text{if} \qquad \sigma > [\ell n(t-W_{n-1}) - \mu]$$

$$\left|\frac{\delta V_t^c}{\delta \mu}\right| \leq \left|\frac{\delta V_t^c}{\delta \lambda}\right| \leq \left|\frac{\delta V_t^c}{\delta \sigma}\right| \qquad \text{if} \qquad \sigma < [\ell n(t-W_{n-1}) - \mu] \tag{3.2.23}$$

and conversely for $\pi_t \geq \frac{1}{2}$ (assuming (3.2.20) satisfied). The variation of V_t^c with respect to λ , μ , σ are shown in Fig. 3.6--3.8, respectively. The curves are of the same shape as the variation in π_t but are symmetric about the time axis. For $\pi_t < \frac{1}{2}$ the sensitivity of V_t^c to parameter variation decreases for increasing π_t^c and the maximum for any particular $\pi_{W_{n-1}}$ moves to the left. It is apparent from Fig. 3.2--3.4 that for increasing π_t , the conditional mean has less sentivity to parameter variations. However, because of the convex relationship (3.2.10), variations of the conditional variance are smallest in the neighborhood of $\pi_t = \frac{1}{2}$ and greatest in the neighborhood of $\pi_t = 0$, $\pi_t = 1$.

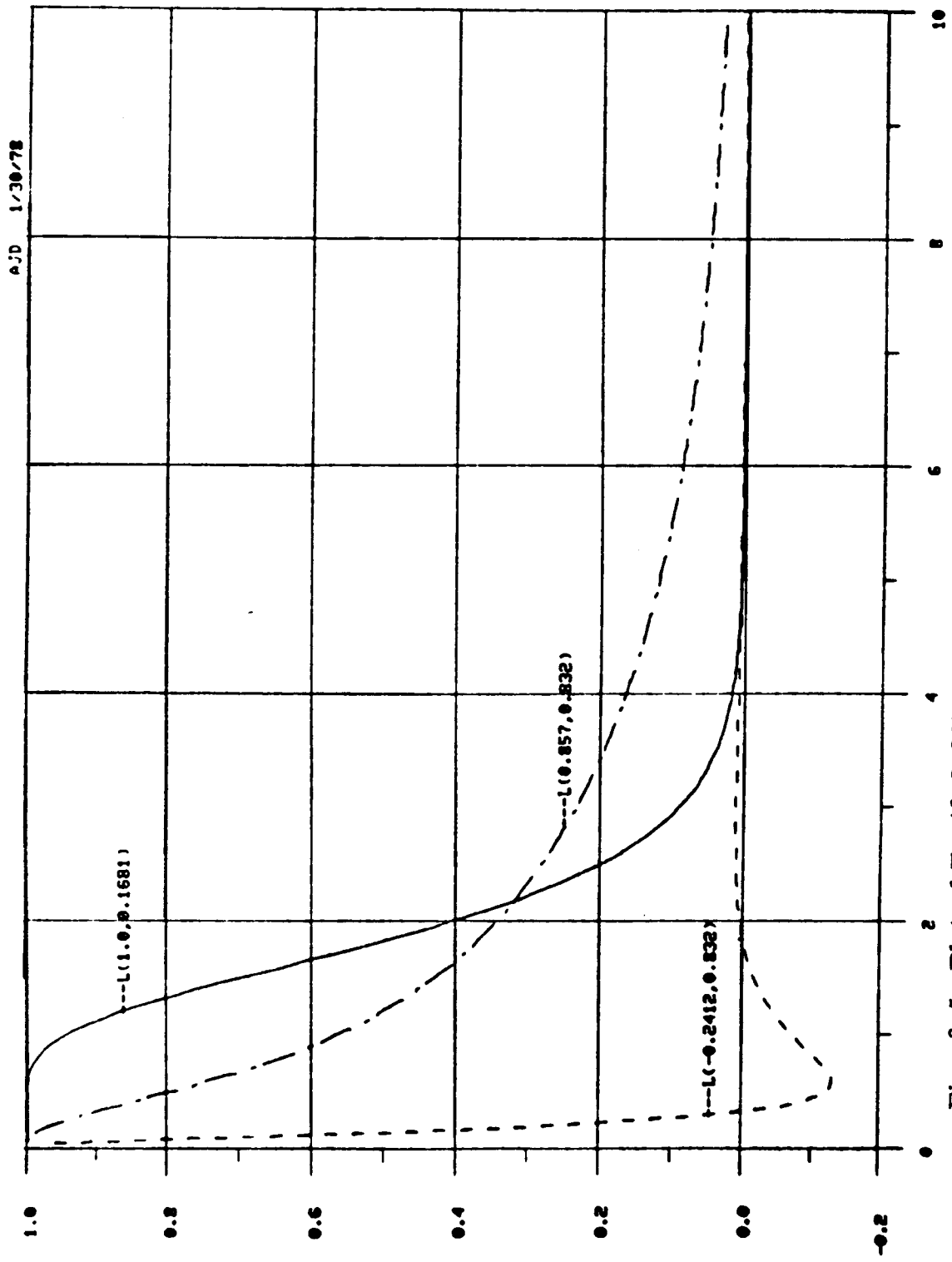


Fig. 3.5: Plot of Eq. (3.2.20) vs. time for different parameter values

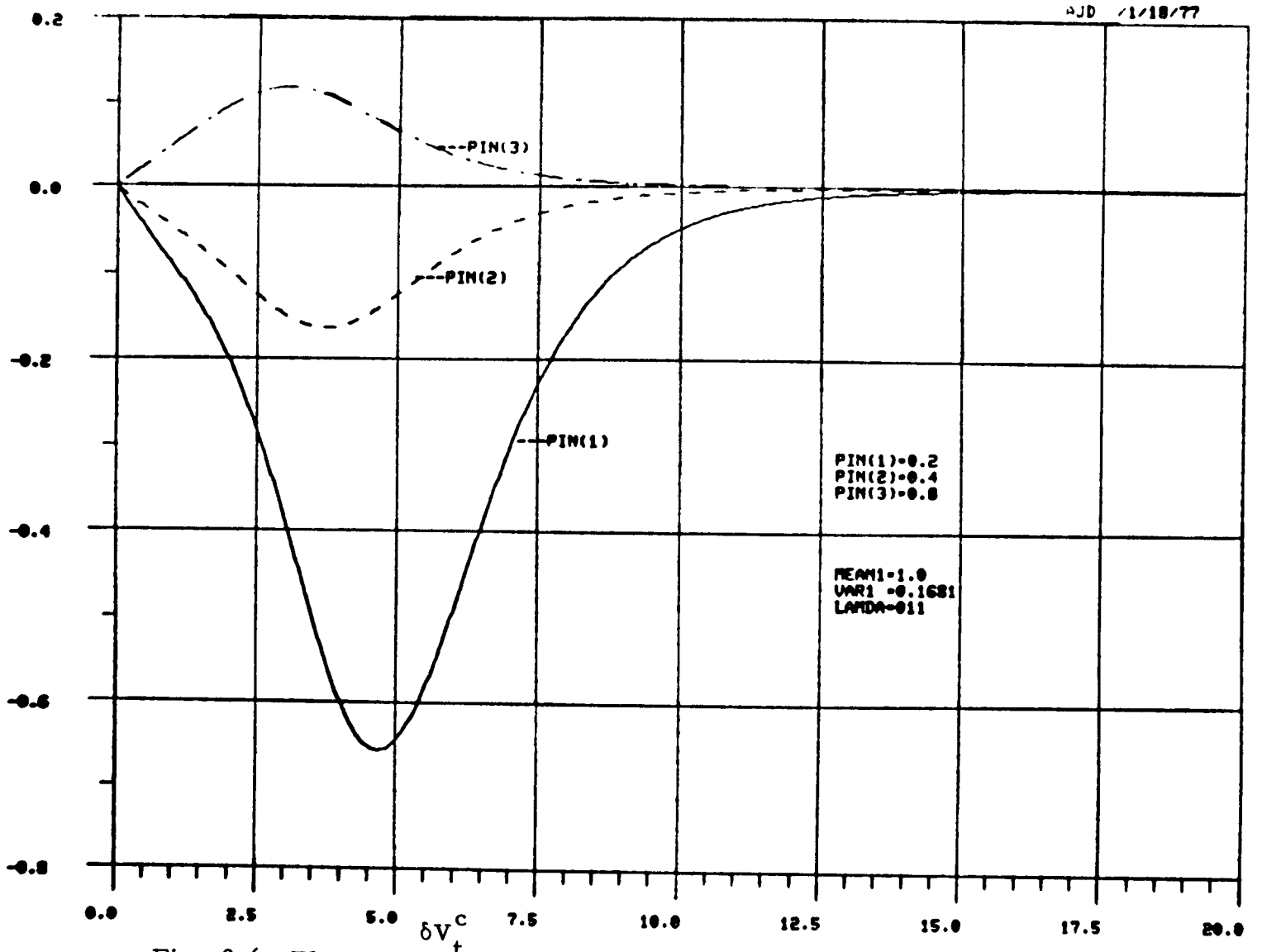


Fig. 3.6: Plot of $\left(\frac{\delta V_t^c}{\delta \lambda}\right)$ vs. time

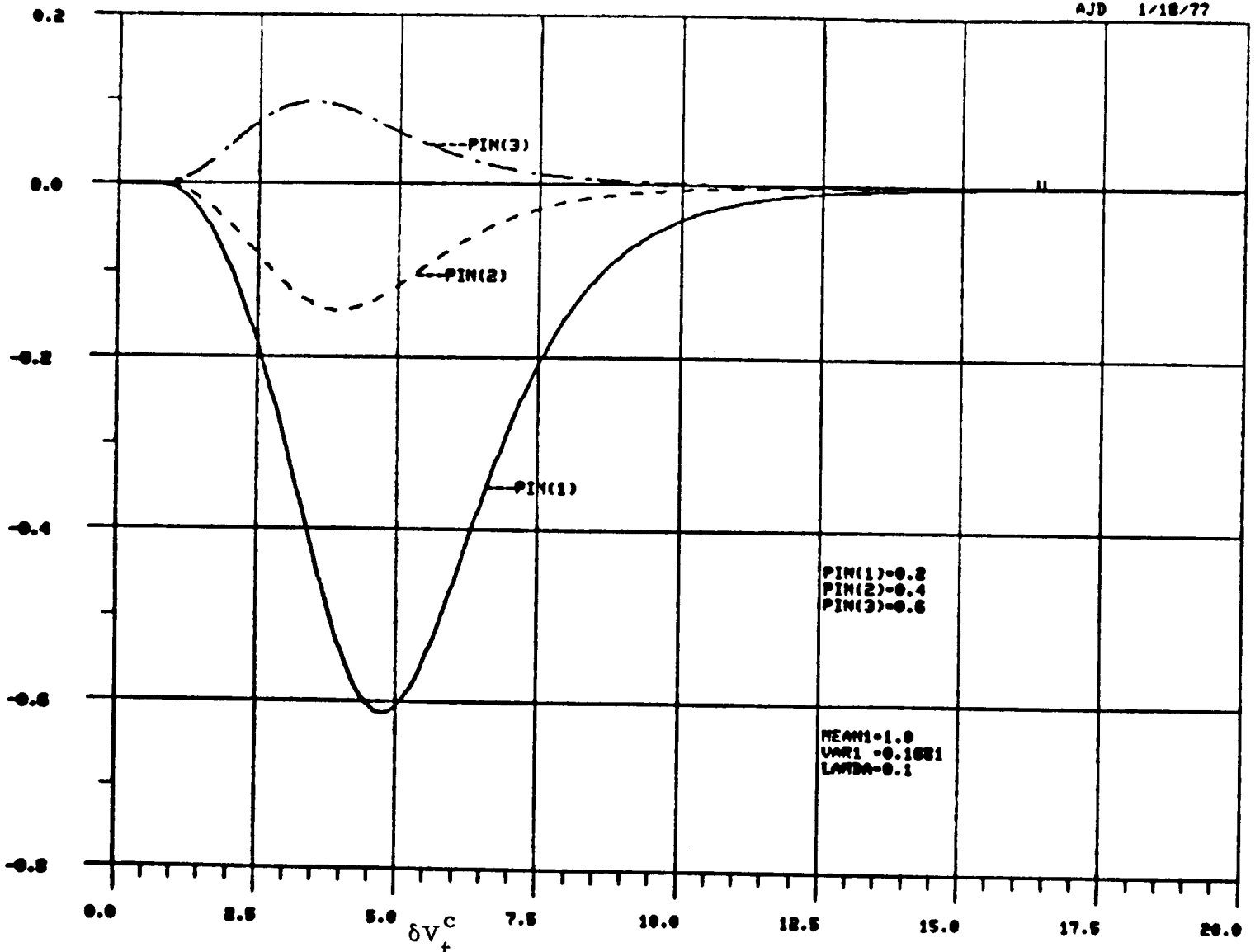


Fig. 3.7: Plot of $\left(\frac{\delta V_t^c}{\delta \mu}\right)$ vs. time

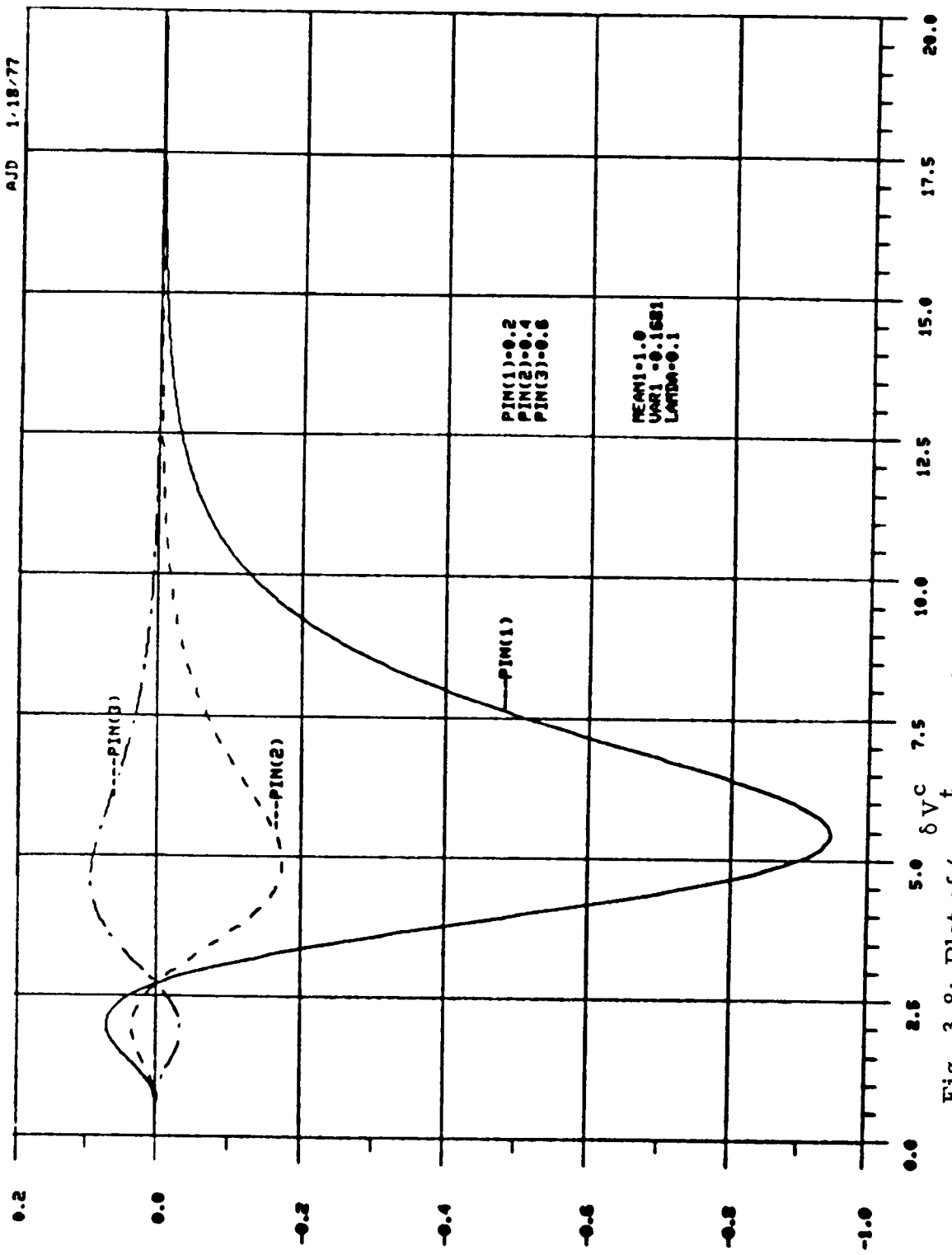


Fig. 3.8: Plot of $\left(\frac{\delta V_t^c}{\delta \sigma}\right)$ vs. time

For the discontinuous part π_t^d , we define

$$C_3 = q_{t-} - \pi_{t-} \tag{3.2.24}$$

$$C_4 = 1 - \pi_{t-} \tag{3.2.25}$$

$$C_5 = \frac{(1-\pi_{t-})\pi_t(1-q_{t-})}{[(\lambda_t^1-\lambda_t^0)\pi_{t-} + \lambda_t^0]^2} \tag{3.2.26}$$

then by (3.2.8)

$$\pi_t^d = C_4 \frac{(\lambda_t^1\pi_{t-} + C_3\lambda_t^0)}{(\lambda_t^1\pi_{t-} + C_4\lambda_t^0)}$$

Variation wrt λ

$$\frac{\delta\pi_t^d}{\delta\lambda} = C_4\frac{\pi_{t-}\lambda_t^0(C_4 - C_3)\lambda_t^0}{[\lambda_t^1\pi_{t-} + C_4\lambda_t^0]^2} = C_5\lambda_t^0 \tag{3.2.27}$$

by applying (3.2.26).

Since π_t^d is more involved, several useful equalities are developed. By (1.2.13)

$$g(t-W_{n-1}) = \frac{1}{\sigma(t-W_{n-1})\sqrt{2\pi}} \exp\left[\frac{-(\ell n(t-W_{n-1}) - \mu)^2}{2\sigma^2}\right]$$

we have

$$\frac{\delta g(t-W_{n-1})}{\delta \mu} = \frac{g(t-W_{n-1})}{\sigma} \left[\frac{\ell n(t-W_{n-1}) - \mu}{\sigma}\right] \tag{3.2.28}$$

$$\frac{\delta g(t-W_{n-1})}{\delta \sigma} = g(t-W_{n-1}) \left[\frac{(\ell n(t-W_{n-1}) - \mu)^2}{\sigma^3} - \frac{1}{\sigma}\right]$$

$$= \frac{g(t-W_{n-1})}{\sigma} \left(\frac{\ell n(t-W_{n-1}) - \mu}{\sigma}\right)^2 \left[1-\left(\frac{\ell n(t-W_{n-1}) - \mu}{\sigma}\right)^{-2}\right]$$

$$= \frac{\delta g(t-W_{n-1})}{\delta \mu} \left(\frac{\ell n(t-W_{n-1}) - \mu}{\sigma}\right) \left[1-\left(\frac{\ell n(t-W_{n-1}) - \mu}{\sigma}\right)^{-2}\right] \tag{3.2.29}$$

and by (3.2.9), (3.2.14)

$$\frac{\delta \lambda_t^0}{\delta \mu} = \frac{\frac{\delta g(t-W_{n-1})}{\delta \mu} h(\mu, \sigma) - g(t-W_{n-1}) \frac{\delta}{\delta \mu} L(\mu, \sigma)}{h^2(\mu, \sigma)}$$

$$= \frac{g(t-W_{n-1})}{[\sigma h(\mu, \sigma)]^2} [\ell n(t-W_{n-1}) - \mu) h(\mu, \sigma)$$

$$- (t-W_{n-1})\sigma^2 g(t-W_{n-1})] \tag{3.2.30}$$

by applying (3.2.15).

$$\frac{\delta \lambda_t^0}{\delta \sigma} = \frac{\frac{\delta g(t-W_{n-1})}{\delta \sigma} h(\mu, \sigma) - g(t-W_{n-1}) \frac{\delta h(\mu, \sigma)}{\delta \sigma}}{h^2(\mu, \sigma)}$$

$$= \left[\frac{\left(\frac{\ell n(t-W_{n-1}) - \mu}{\sigma}\right)}{h^2(\mu,\sigma)}\right]\left\{\frac{\delta g(t-W_{n-1})}{\delta\mu}\left[1-\left(\frac{\ell n(t-W_{n-1}) - \mu - 2}{\sigma}\right)\right]h(\mu,\sigma)\right.$$
$$\left. - g(t-W_{n-1})\frac{\delta}{\delta\mu}h(\mu,\sigma)\right\}$$

by applying (3.2.16), (3.2.29)

$$\frac{\delta\lambda_t^0}{\delta\sigma} = \left[\frac{\ell n(t-W_{n-1}) - \mu}{\sigma}\right]\left(\frac{\delta\lambda_t^0}{\delta\mu}\right) - \frac{\lambda_t^0}{\sigma} \tag{3.2.31}$$

by applying (3.2.28), (3.2.30) and (3.2.14), (3.2.9)

Variation wrt μ, σ

$$\frac{\delta\pi_t^d}{\delta\mu} = -C_4\,\frac{\pi_{t-}\lambda_t^1\,(C_4-C_3)}{[\lambda_t^1\pi_{t-} + C_4\lambda_t^0]^2}\left(\frac{\delta\lambda_t^0}{\delta\mu}\right) = -C_5\,\lambda_t^1\left(\frac{\delta\lambda_t^0}{\delta\mu}\right) \tag{3.2.32}$$

by (3.2.24) - (3.2.26). Similarly

$$\frac{\delta\pi_t^d}{\delta\sigma} = -C_5\,\lambda_t^1\left(\frac{\delta\lambda_t^0}{\delta\sigma}\right) \tag{3.2.33}$$

From this development, it is apparent that the parameter sensitivity of the nonlinear terms in the rates, π_t^d, is more complicated and bounds similar to (3.2.22) are not feasible. Although analytically difficult to prove, it can be shown computationally that $\frac{\delta\lambda_t^0}{\delta\mu}$ in (3.2.30) is negative (Fig. 3.9). Consequently, since C_5 is positive (3.2.26), π_t^d is an increasing function of λ, μ (3.2.27), (3.2.32).

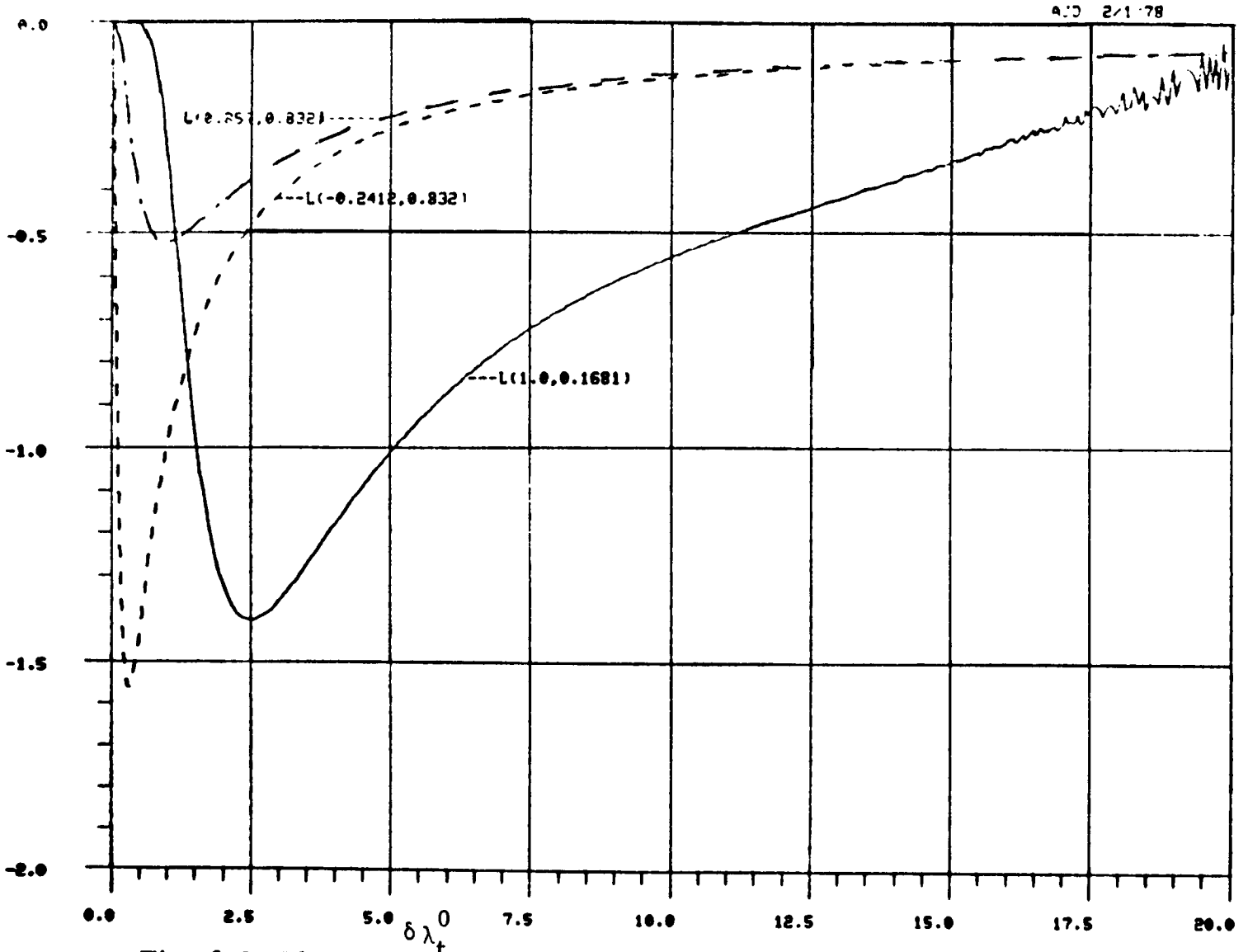


Fig. 3.9: Plot of $\left(\frac{\delta \lambda_t^0}{\delta \mu}\right)$ vs. time for different parameter values

The functional dependency of σ on π_t^d is not as explicit because of the involved relationship (3.2.31) and thus no comment is made. From the results of Fig. 3.9 and by (3.2.31) - (3.2.33),

$$\frac{\delta\pi_t}{\delta\sigma} - \frac{\delta\pi_t}{\delta\mu} = C_5\,\lambda_t^1\,[\frac{\delta\lambda_t^0}{\delta\mu} - \frac{\delta\lambda_t^0}{\delta\sigma}]$$

$$= (\frac{C_5\lambda_t^0}{\sigma})\{(\frac{\delta\lambda_t^0}{\delta\mu})[\sigma - (\ell n(t-W_{n-1}) - \mu)] + \lambda_t^0\} \qquad (3.2.34)$$

is positive if $\ell n(t-W_{n-1}) - \mu \geq \sigma$. In other words, on this interval $\frac{\delta\pi_t^d}{\delta\sigma}$, $\frac{\delta\pi_t^d}{\delta\mu}$ are positive and hence

$$|\frac{\delta\pi_t^d}{\delta\sigma}| \geq |\frac{\delta\pi_t^d}{\delta\mu}| \qquad \text{if} \qquad \sigma \leq \ell n(t-W_{n-1}) - \mu \qquad (3.2.35)$$

On the complement interval, no such simple relationship can be obtained as observed from the right most factor in (3.2.34) and by (3.2.30). A plot of this factor for various parameters is shown in Fig. 3.10. By (3.2.27), (3.2.32)

$$\frac{\delta\pi_t^d}{\delta\lambda} - \frac{\delta\pi_t^d}{\delta\mu} = C_5(\lambda_t^0 + \lambda_t^1\ \frac{\delta\lambda_t^0}{\delta\mu}) \qquad (3.2.36)$$

Observe a similar difficulty results because of $\frac{\delta\lambda_t^1}{\delta\mu}$ given in (3.2.30) and hence no conclusive statement can be made. A plot of the right most factor in (3.2.36) is given in Fig. 3.11. Since the bounds similar to (3.2.22) are not obtainable, the relationship between (3.2.27), (3.2.32), (3.2.33), shown respectively in Fig. 3.12-3.14, are dependent on the specific parameter

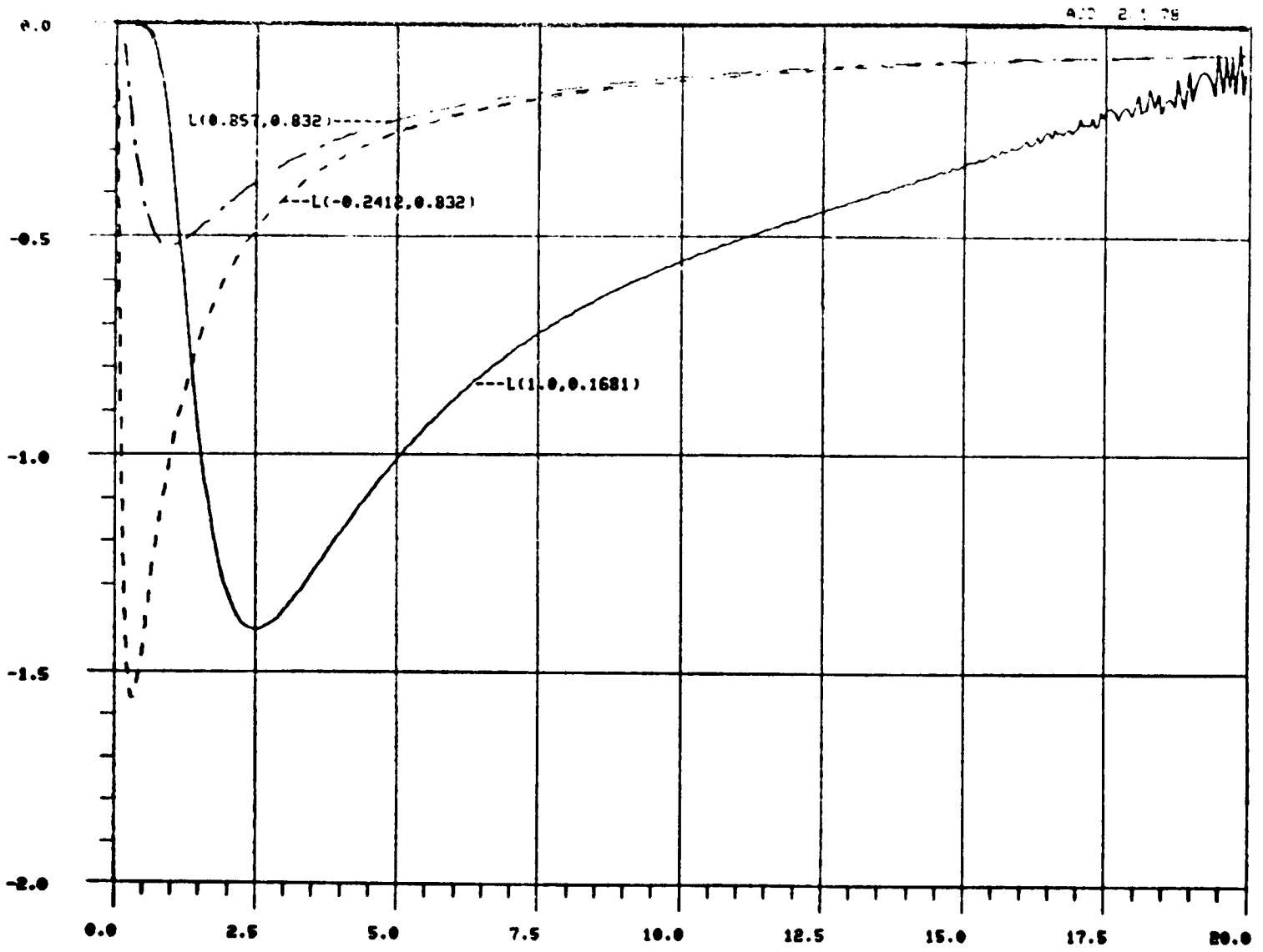


Fig. 3.10: Plot of $\{ (\frac{\delta\lambda_t^0}{\delta\mu}) [\sigma - (\ell n(t) - \mu)] + \lambda_t^0 \}$ in Eq. (3.2.34) vs. time

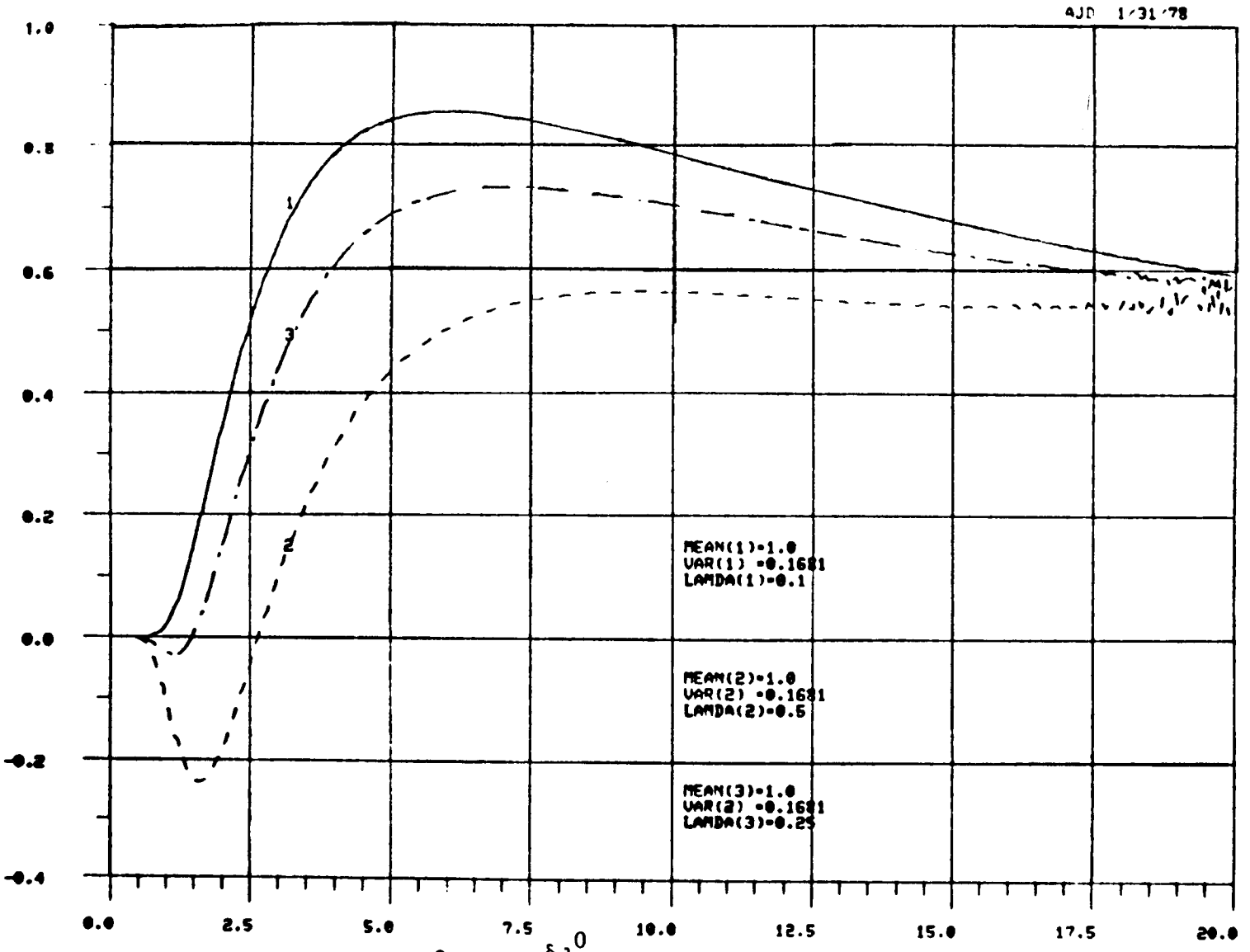


Fig. 3.11: Plot of $\{ \lambda_t^0 + \lambda_t^1 \frac{\delta \lambda_t^0}{\delta \mu} \}$ in Eq. (3.2.36) vs. time

values.* By Fig. 3.11

$$\lambda_t^0 + \lambda_t^1 \frac{\sigma\lambda_t^0}{\delta\mu} \geq 0 \qquad \text{for all } t \geq 0$$

and thus by (3.2.36)

$$\left|\frac{\delta\pi_t^d}{\delta\lambda}\right| \geq \left|\frac{\delta\pi_t^d}{\delta\mu}\right| \qquad \text{for all } t \geq 0 \tag{3.2.37}$$

By Fig. 3.10

$$\left(\frac{\delta\lambda_t^0}{\delta\mu}\right)\left[\sigma - \ell n(t-W_{n-1}) - \mu\right] + \lambda_t^0 \begin{cases} > 0 & \text{for } t \geq 2.550 \\ \leq 0 & \text{for } t < 2.550 \end{cases}$$

and by comparing Fig. 3.12, 3.14

$$\left|\frac{\delta\pi_t^d}{\delta\lambda}\right| \geq \left|\frac{\delta\pi_t^d}{\delta\sigma}\right| \qquad \text{for all } t \geq 0$$

hence, by (3.2.34), (3.2.37)

$$\begin{aligned} &\left|\frac{\delta\pi_t^d}{\delta\lambda}\right| \geq \left|\frac{\delta\pi_t^d}{\delta\mu}\right| \geq \left|\frac{\delta\pi_t^d}{\delta\sigma}\right| && \text{if} \quad t < 2.550 \\ &\left|\frac{\delta\pi_t^d}{\delta\lambda}\right| \geq \left|\frac{\delta\pi_t^d}{\delta\sigma}\right| \geq \left|\frac{\delta\pi_t^d}{\delta\mu}\right| && \text{if} \quad t \geq 2.550 \end{aligned} \tag{3.2.38}$$

From (3.2.38), the discontinuous part of the filter is more sensitive to one parameter, λ, over all time. This result is reasonable for two

---

*Discussion concerns parameter values $\mu = 1.0$, $\sigma^2 = 0.1681$, $\lambda = 0.1$ $q_t = 0.2$.

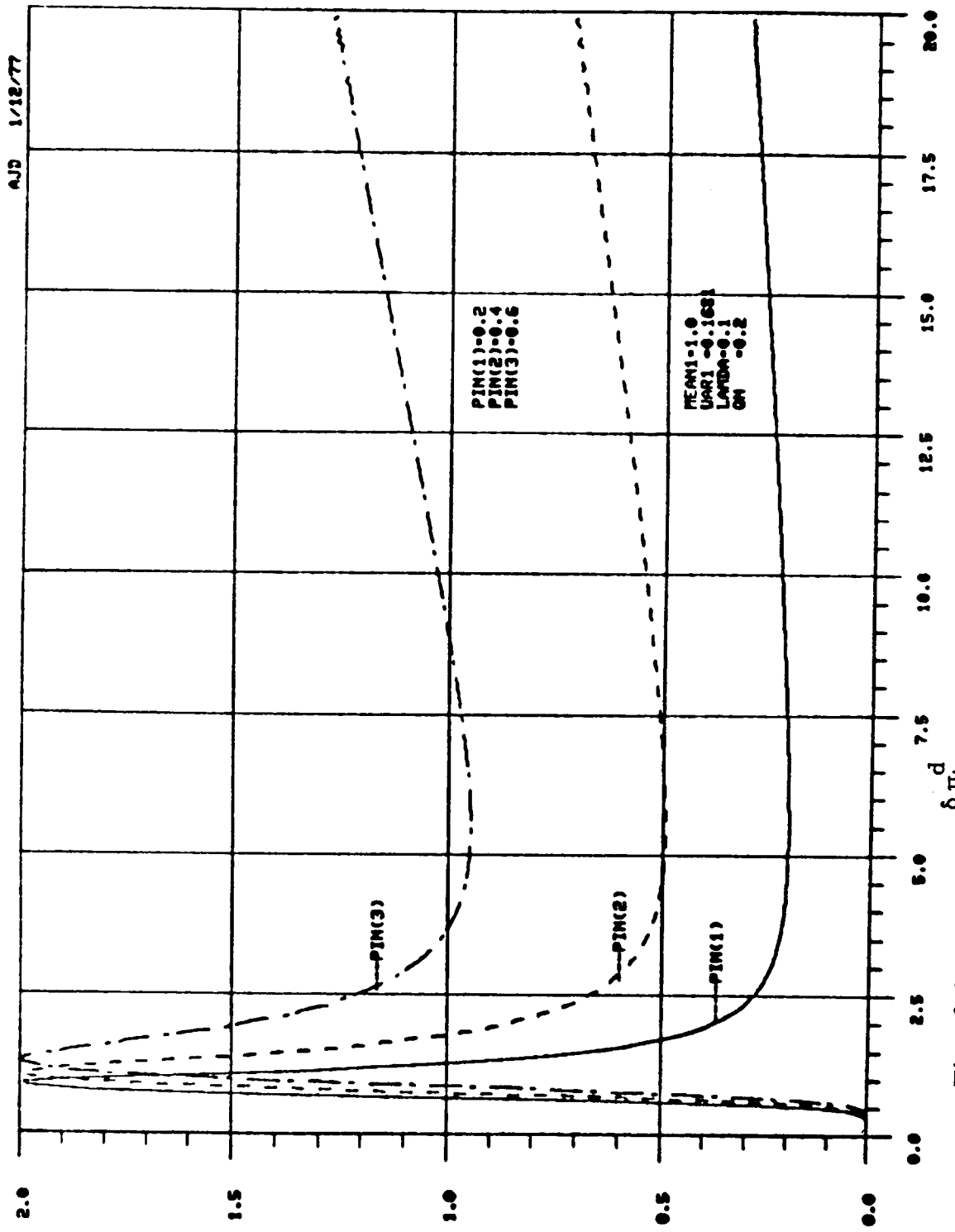


Fig. 3.12: Plot of $\left(\frac{\delta \pi_f^d}{\delta \lambda}\right)$ vs. time

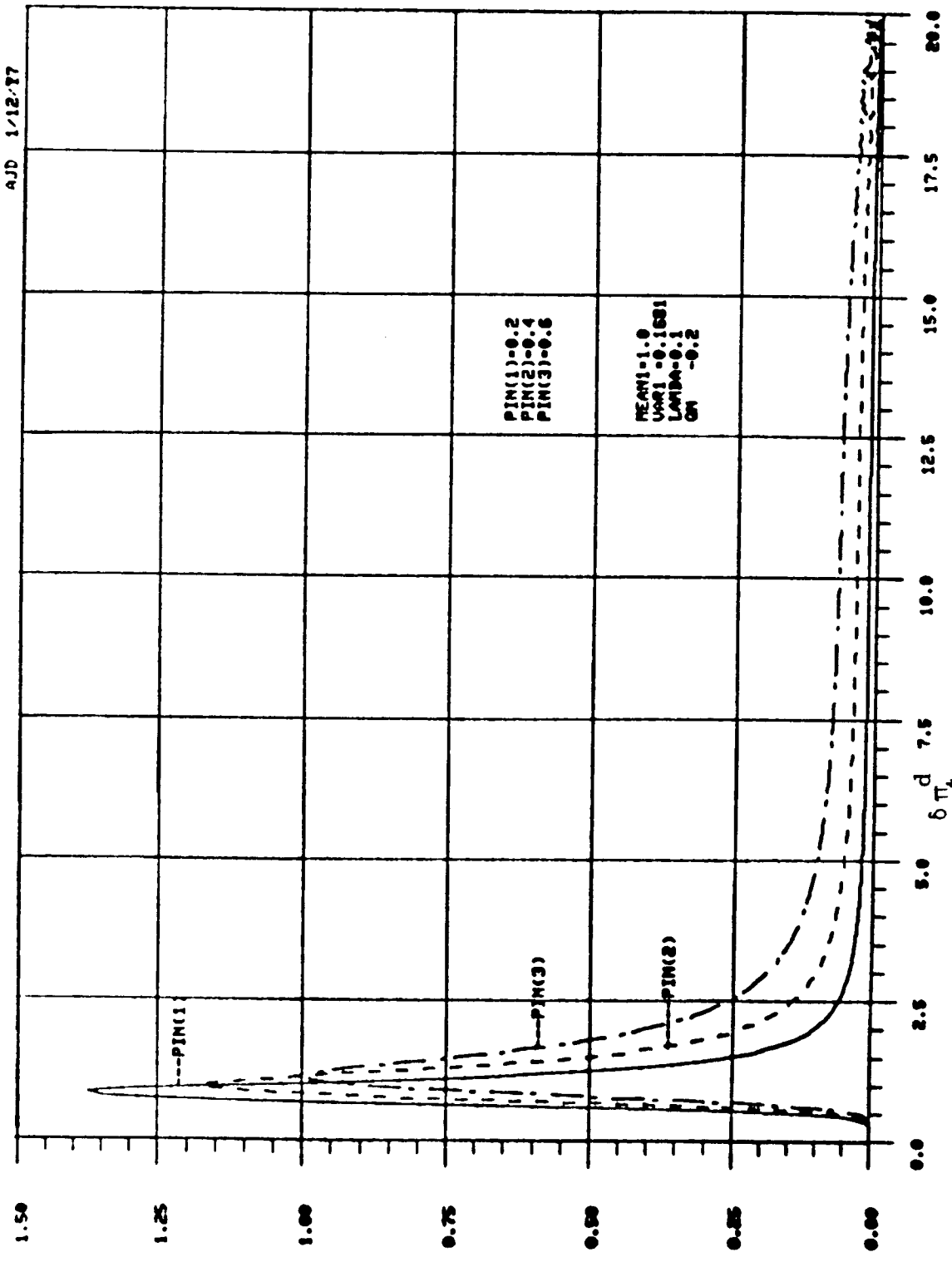


Fig. 3.13: Plot of $\left(\frac{\delta \pi_t^d}{\delta \mu}\right)$ vs. time

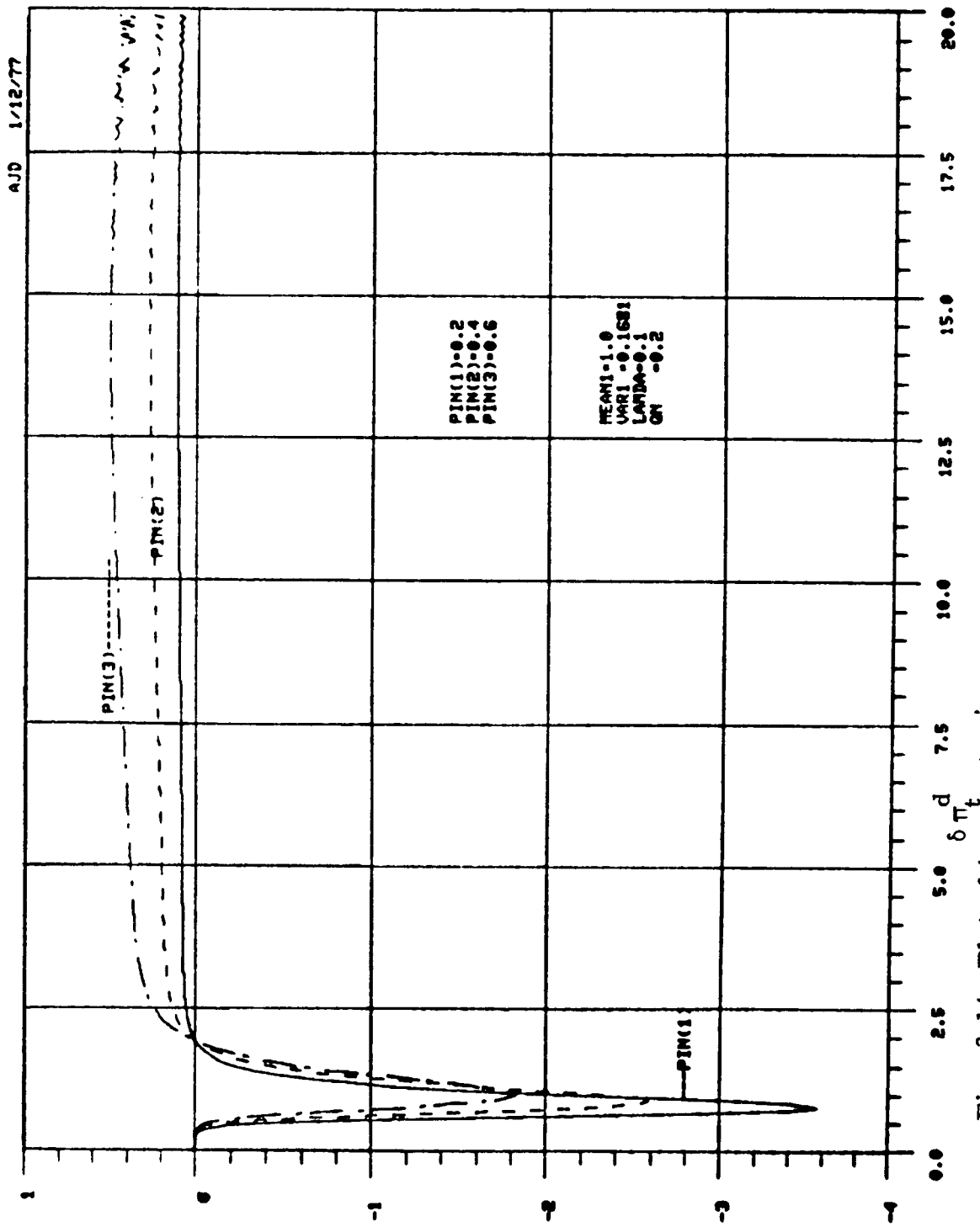


Fig. 3.14: Plot of $\left(\frac{\delta \pi_t^d}{\delta \sigma}\right)$ vs. time

reasons. First, π_t^d is explicitly dependent on λ in (3.2.8), while μ, σ are implicitly dependent on π_t^d via λ_t^0. Second, in the evaluation (Fig. 2.7-2.10) for almost all activation times, π_t^d is negative, i.e., the probability that the switch had occurred decreases. In other words, the second rate dominates the behavior of the discontinuity. The interesting aspect of (3.2.38), though, is the relationship between μ and σ. As in (3.2.22) for short headways, μ influences the filter more and for long headways, σ influences the filter more. This result can be readily related to the lognormal rate as discussed above.

The variation of the discontinuous part of the conditional error variance, V_t^d, is similarly deduced from (3.2.10), (3.2.38). For $\pi_t < \frac{1}{2}$

$$\begin{aligned} &\left|\frac{\delta V_t^d}{\delta \lambda}\right| \geq \left|\frac{\delta V_t^d}{\delta \mu}\right| \geq \left|\frac{\delta V_t^d}{\delta \sigma}\right| \qquad \text{if} \quad t < 2.550 \\ &\left|\frac{\delta V_t^d}{\delta \lambda}\right| \geq \left|\frac{\delta V_t^d}{\delta \sigma}\right| \geq \left|\frac{\delta V_t^d}{\delta \mu}\right| \qquad \text{if} \quad t > 2.550 \end{aligned} \tag{3.2.39}$$

and conversely for $\pi_t \geq \frac{1}{2}$. Comments similar to those concerning V_t^0 can be made except if $\pi_t < \frac{1}{2}$, the maximum value of $\frac{\delta V_t^d}{\delta \theta}$ for increasing π_{t-} moves to the right (instead of the left). The variation of V_t^d with respect to λ, μ, σ are shown in Fig. 3.15-3.17, respectively.

From the analysis of the continuous and discontinuous parts of the conditional variance, several comments are made. First, since

$$\max_t \frac{\delta V_t^d}{\delta \theta} \geq \max_t \frac{\delta V_t^c}{\delta \theta} \tag{3.2.40}$$

the discontinuous part π_t^d is more sensitive to parameter variation given any arbitrary interval time. From (3.2.40), let $t_d(t_c)$ denote the time when the

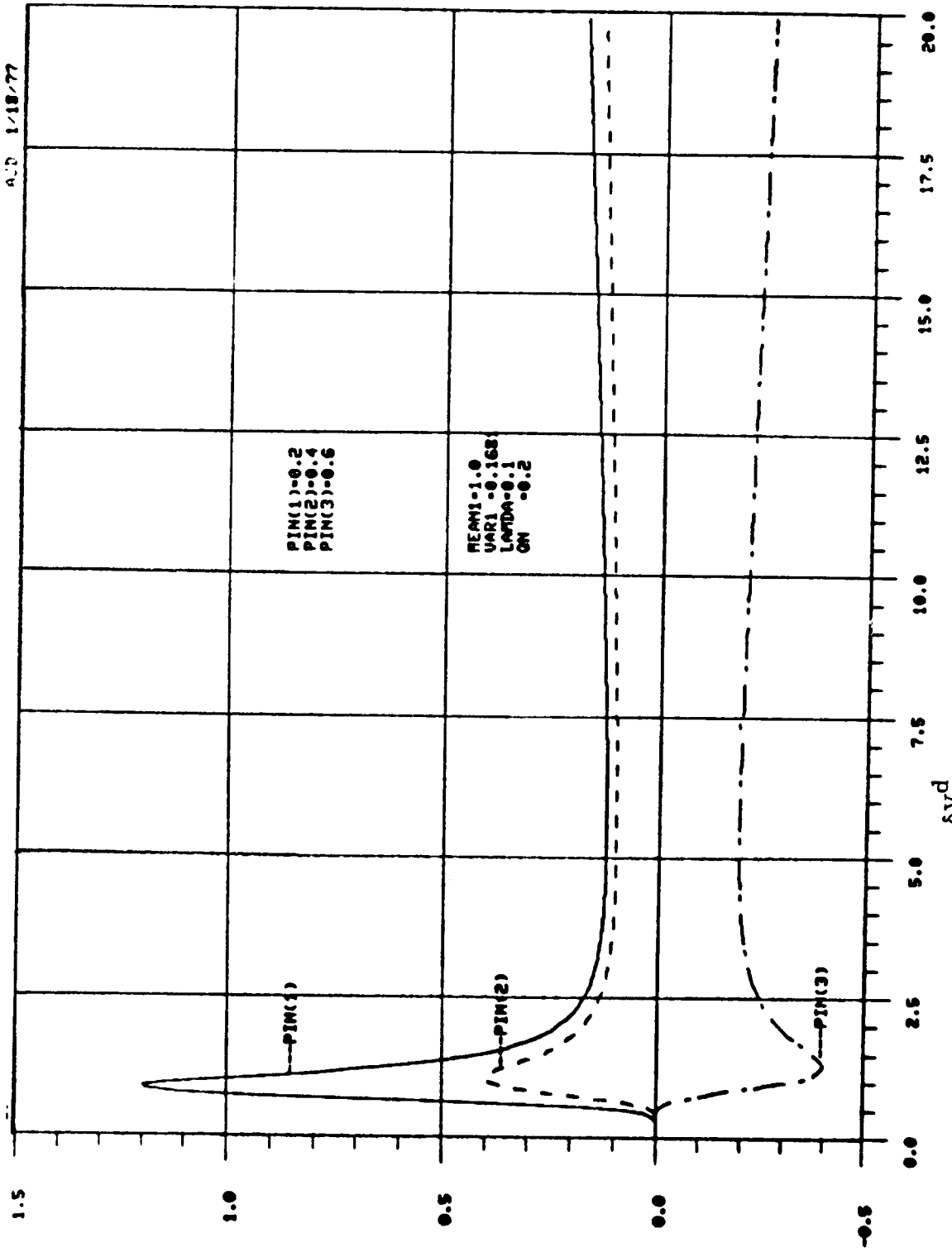


Fig. 3.15: Plot of $\left(\frac{\delta V_t^d}{\delta \lambda}\right)$ vs. time

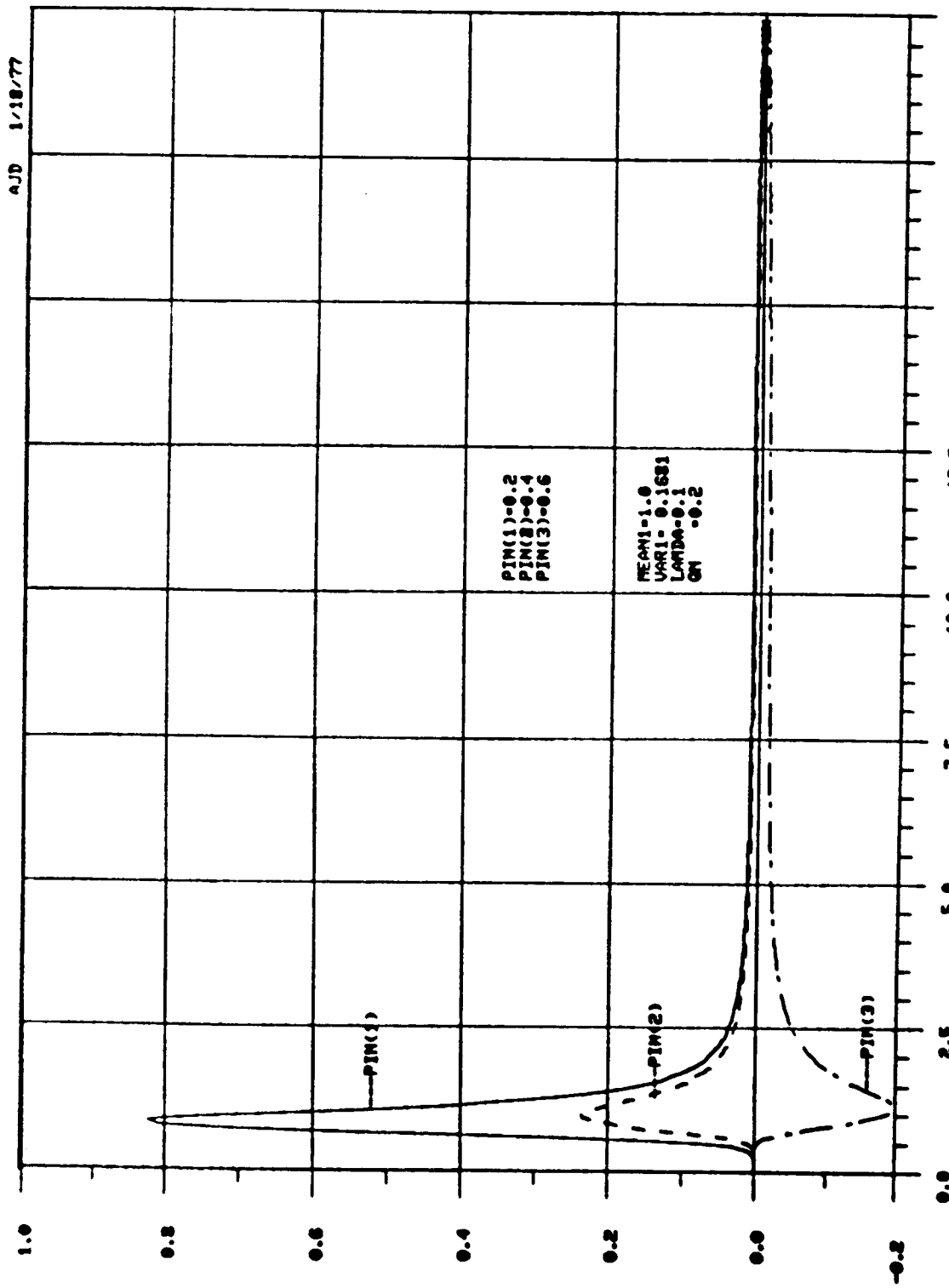


Fig. 3.16: Plot of $\left(\frac{\delta V_t^d}{\delta \mu}\right)$ vs. time

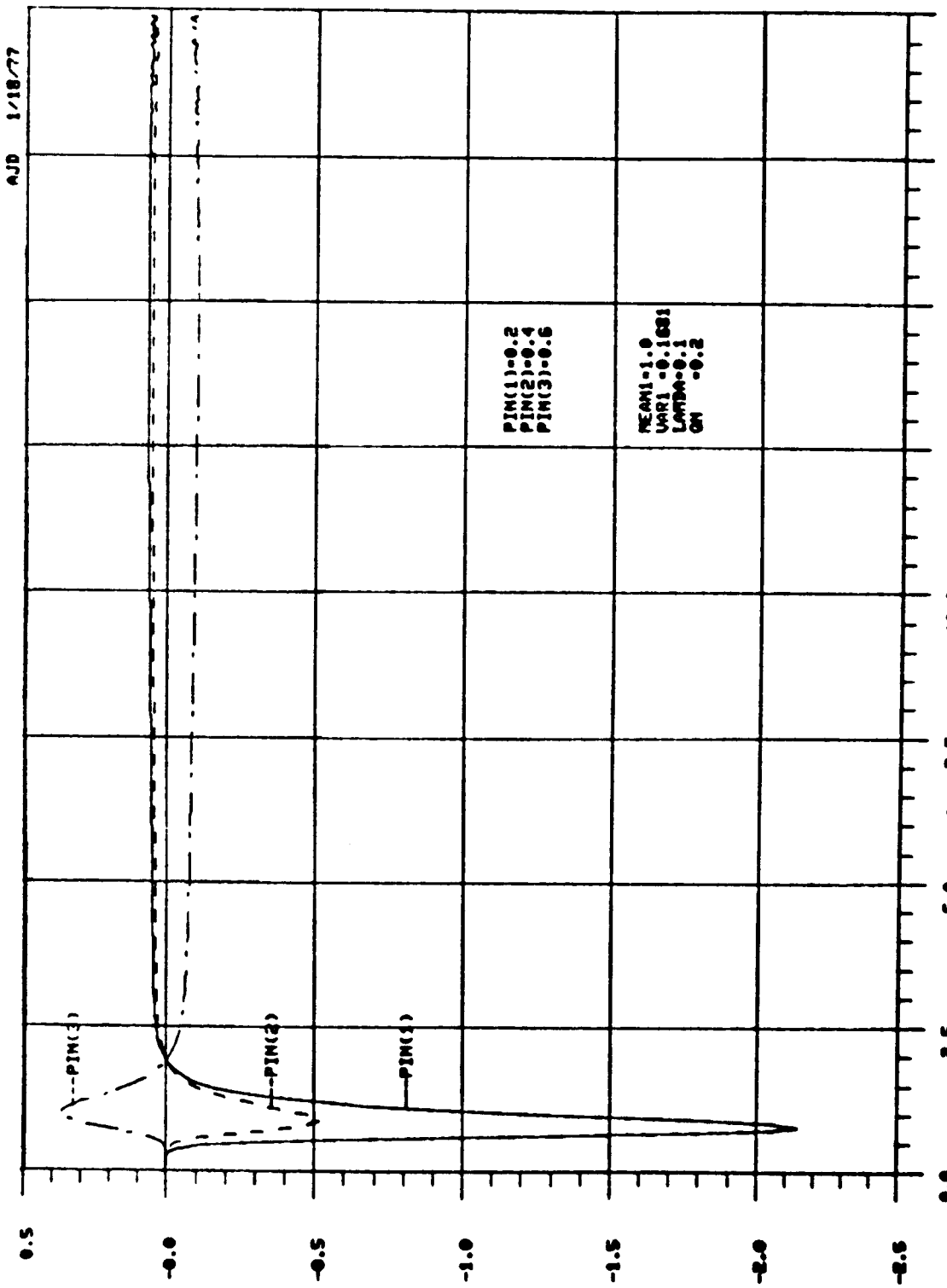


Fig. 3.17: Plot of $\left(\frac{\delta V_t^d}{\delta \sigma}\right)$ vs. time

left (right)-hand side achieves its maximum. Then from Fig. 3.6-3.8, 3.15-3.17, observe that t_d is always less than t_c for any particular parameter, θ . Recall from Fig. 2.10 that the short headway between $W_{16} - W_{17}$ misled the filter. In this neighborhood in which t_d is included, the difference in the rates λ_t^0 , λ_t^1 has a sign change (Fig. 2.12). Similarly (although not shown), V_t^d has a second peak when this difference in the rates reverses sign. Thus, for short headways and quite long headways, π_t^d is more sensitive to any given parameter variation, while for intermediate headways π_t^c is more sensitive. This latter fact can be deduced from (2.3.21) for $W_{n-1} < t < W_n$ where the greatest influence of $\frac{\delta V_t^c}{\delta \theta}$ occurs when the difference between λ_t^0 , λ_t^1 is greatest. Observe (Fig. 2.11), t_c defined above is in this neighborhood. One desirable property of any estimator is that the error variance decreases with increasing observations. The conditional variance curves show that parameter error introduced into the signal model affect the filter differently. Parameter uncertainty in π_t^c approaches zero for increasing (larger interarrival) time, while in π_t^d , the inaccuracy decreases to a nonzero value. In the case of λ , error in π_t^d even increases with increasing time. Finally, the symmetry of the conditional variance results because of (3.2.10). Naturally, parameter uncertainty has its greatest influence when $\pi_t = 0$ or 1 . Notice, though, that π_t^c and π_t^d have a complementary affect to this uncertainty. The variation with respect to any particular parameter is of opposite sign and for increasing initial conditions, $\pi_{W_{n-1}}$, the maximum defined in (3.2.40) move in the exact opposite direction. Recall this same complementary effect of π_t^c and π_t^d was noted when the a prior density, p_n was uniform.

To conclude this section, we return to the discussion of (3.2.11). Any uncertainty in a parameter can be approximated to first order by

$$\Delta\pi_t = \pi_t - \pi_t \cong \frac{\delta\pi_t}{\delta\theta}(\theta-\bar{\theta}) \leq \left|\frac{\delta\pi_t^c}{\delta\theta} + \frac{\delta\pi_t^d}{\delta\theta}\right| (\theta-\bar{\theta}) \tag{3.2.41}$$

The dependency on time of this left-hand side is explicit. By removing it, one can obtain a sample independent error bound (maximum number of inter-arrival times are specified by a prior density)

$$\Delta\pi_t \leq \max_t \left[\left|\frac{\delta\pi_t^c}{\delta\theta} + \frac{\delta\pi_t^d}{\delta\theta}\right|\right] \cdot (\theta-\bar{\theta}) \tag{3.2.42}$$

Unfortunately analytical expressions for this extremization were not obtainable and, consequently, computational techniques are necessary. Fortunately, such optimums exist by the convexity of the functions involved. Hence, given any interval $(W_{n-1}, W_n]$, except in the neighborhood of $\pi_t = 0$ or 1, variation of the filter to parameter uncertainty can be bounded.

3.3 Parameter Estimate Uncertainty

For the bound in (3.2.42) to be meaningful, one must specify the neighborhood of the parameter, θ. However, any estimator $\hat{\theta}$ of this unknown constant is probabilistic. (For a particular sample, we shall denote the estimate by $\bar{\theta}$.) From its statistics, the uncertainty between the parameter and the estimate can be related by the probability measure:

$$\Pr\{\,|\theta - \bar{\theta}| < \varepsilon\,\} \leq \delta \tag{3.3.1}$$

In other words, given a certain confidence coefficient δ, one can insure (in probability) that

$$|\theta - \bar{\theta}| \leq \varepsilon$$

for $\varepsilon > 0$. By (3.2.42), one can similarly insure this same confidence level for the bound on $\Delta\pi_t$. Before preceding in this line of discussion, we first examine techniques to estimate these parameters.

In our initial development, we conjectured that the statistical model of headways was the same for all traffic flows (light, moderate, heavy) and different traffic environments (urban, rural, freeway, etc.), but the parameters would vary. Therefore, the model proposed (1.2.12)

$$p(h) = \psi p_f(h) + (1-\psi)\ p_{nf}(h) \tag{3.3.2}$$

$$p_f(h) = \frac{1}{\sqrt{2\pi\sigma^2 h}} \exp\ [\frac{-(\ell n\ h-\mu)^2}{2\sigma^2}] = g(h) \qquad , \qquad h \geq 0$$

$$p_{nf}(h) = \lambda \exp\ (-\lambda h) \int_o^h g(x) \exp\ (\lambda x) dx \qquad , \qquad h \geq 0$$

was specialized to a particular traffic network link by the parameters ψ , λ , μ , σ . Then the obvious question arises: Given traffic detector data and traffic signal code, determine estimates of these parameters? This question appears somewhat similar to the one posed in the introduction of the disorder problem. However, in that context, we were concerned with estimating a traffic parameter, queue length, while here we are concerned with estimating parameters in our model. Also the observation period of the former was smaller compared to that for the latter. As mentioned previously, the parameter estimates have a certain periodicity defined by the lights.

Fortunately, there are a number of techniques available for determining parameter estimates from traffic data. The method of moments, whereby

one equals the sample moments $\{m_i\}$ to the population moments determined by (3.3.2) results in a set of four equations in four unknowns. Using this technique, Buckley [4 ,21] found numerical methods necessary to solve the composite model equations. The population moments about the origin of (3.3.2) can be shown to satisfy

$$\mu_n = \mu_{fn} + (\tfrac{n}{\lambda})[\ \mu_{n-1} - \psi\mu_{f(n-1)}] \tag{3.3.3}$$

where μ_n (μ_{fn}) is the n^{th} moment about the origin of the overall (following) headway density. These parameter estimates, by simple calculation can seem to be complex expressions and Branston found poor overall χ^2 goodness-of-fit results [29, p. 129]. Only when m_3 , m_4 are small compared to m_2 are these estimators efficient.

Consequently, variations of the method of moments were necessary. Branston utilized a combination of minimum chi-squared and method of moments. By varying μ and σ , the parameters, λ and ψ were determined from the first two sample moments. Then, the resulting chi-squared was minimized using a numerical routine. The resulting parameter estimates were obtained from suburban traffic data. In urban traffic, since the maximum headway is dependent on the light cycle, the model (3.3.2) is a truncated density. Consequently, a different approach is necessary.

Before continuing, we consider the model proposed (3.3.2) in the urban traffic environment. Since there is a maximum (finite) headway, the observed headways are truncated in the sense that only a certain range of the variate is defined, e.g., $0 < h_i <$ maximum light cycle. Alternatively, though, given n observations obeying (3.3.2), those determined by p_f and p_{nf} are not easily identifiable. Instead, for parameter estimates, the sample set must be

censored. In particular, from these n observations, we may consider only $n - j$ ($j > 0$) of them whose value lies within a certain range. For example, following headways are (on the average) always less than nonfollowing headways and, consequently, only variates between $0 < h_i <$ preselected threshold are used in determining estimates of μ and σ. The distinction between truncating and censoring is that the latter is a property of the sample while the former is a property of the distribution.

To clarify these remarks and develop an alternative estimation technique, we consider a lognormal density $L(\mu, \sigma)$ and a shifted negative exponential density with parameter λ. Given n observations $\{x_i\}_{i=1}^{n}$ of which $n - j$ obey a lognormal density, we wish to determine estimates of μ and σ. By the robustness of the gaussian process, we first take the ℓn of every data point which enables us to use the so-called outlier test [51,§ 32.23] to remove the remaining j impurities. In particular, let $\bar{x}(\bar{s}^2)$ denote the sample mean (variance) and it can be seen that the following algorithm converges.

ALGH1: Censored Normal Variate Estimates

(i) Let $k = 1$ and compute $\bar{x}_1$ and $\bar{s}_1^2$ for the entire sample of n truncated observations.

(ii) By a two-sided outlier test, discard those observations outside the $2\bar{s}_k^2$ neighborhood of $\bar{x}_k$

(iii) Let $k = k + 1$ and compute $\bar{x}_k$ and $\bar{s}_k^2$ of the remaining sample

(iv) If $|\bar{x}_k - \bar{x}_{k-1}| < \varepsilon_1$ and $|\bar{s}_k^2 - \bar{s}_{k-1}^2| < \varepsilon_2$, then set $\mu = \bar{x}_k$ and $\sigma^2 = \bar{s}_k^2$. Otherwise go to step (ii).

Clearly the outlier test censors the observations and the $2\bar{s}_k^2$ criteria removes both the high and low variates and retains 99.6% of the remaining

(normal) observations. Suppose the j impurities obey the displaced negative exponential. Then by the law of large numbers a similar technique can be developed with the difference that only a one-sided outlier test is meaningful. Since the smallest observation x_{min} is a sufficient statistic of the displacement factor, [51 ; §17.40] then

$$\lambda = [\bar{x} - x_{min}]^{-1} \tag{3.3.4}$$

ALGH2: Censored Shift Negative Exponential Variate Estimates

(i) Let k = 1 and compute $\bar{x}_1$ and $\bar{s}_1^2$ for the entire sample of n truncated observations and determine λ_1 from (3.3.4)

(ii) By a one-sided outlier test, discard those observations less than $2\bar{s}_k^2$ neighborhood of $\bar{x}_k$

(iii) Let k = k + 1 and compute $\bar{x}_k$, $\bar{s}_k^2$ and λ_k of the remaining sample

(iv) If $|\lambda_k - \lambda_{k-1}| < \varepsilon_1$ and $|\bar{s}_k^2 - \bar{s}_{k-1}^2| < \varepsilon_2$, then set $\lambda = \lambda_k$. Otherwise go to step (ii).

Since the variance of these j observations is large (for $\lambda \approx 0.1$), the resulting estimate may not be reliable.

The parameter estimates required in the disorder problem can be determined by using ALGH1 and/or 2. However, the observations must be translated by ℓn before ALGH1 is used. To test the validity of this technique, headway sample paths were generated (as in Section 2.4) with the displacement factor, $\alpha = \exp(\mu + \frac{1}{2}\sigma^2)$ and sample size of 500 observations. Only one parameter set was examined: $\mu = 1.0$, $\sigma^2 = 0.1681$, $\lambda = 0.1$ with $\varepsilon_1 = \varepsilon_2 = 10^{-3}$. The reason $\alpha(=3.0)$ was selected to be the mean of the

lognormal density was to insure sufficient separation of the two populations under study. The results of ALGH1 for different values of ψ is shown in Table 3.1.

Table 3.1. Censored Lognormal Parameter Estimates

ψ	$\bar{\mu}$	$\bar{\sigma}^2$	$\bar{\lambda}$	Number of Following Headways
1.0	1.017	.0605	2.101	382
0.5	1.355	.3485	.048	407
0.25	1.976	.6749	.041	479

The parameter estimate $\bar{\lambda}$ was determined from (3.3.4) using that part of the population discarded by the two-sided outlier test (translated by exponential though). These preliminary results show that for ψ near 1.0 , $\bar{\mu}$ is quite near the exact value, but deteriorates as ψ approaches zero; similarly for $\bar{\sigma}^2$. For ALGH1 , these results are reasonable. With decreasing ψ , the sample variance increases and, consequently, $\bar{\sigma}^2$ increases. But with an increasing $\bar{s}_k^2$, the two-sided outlier test discards more of the lower value variates and causes $\bar{\mu}$ to increase. The estimate $\bar{\lambda}$ is unreliable for $\psi = 1.0$ but increases in reliability as ψ decreases. The censored shifted negative exponential technique performed even worse (ALGH2). Because the variance was so large ($\approx$70), the outlier test did not discard any observations. A slight variation of ALGH1 and 2 was investigated where instead of using $2\bar{s}_k^2$ in the outlier criteria, $\bar{s}_k^2$ and $3\bar{s}_k^2$ were used. The effect was a deterioration in the results.

The accuracy of the parameter estimate, however, cannot be measured simply by comparing the difference of θ and $\bar{\theta}$. Having statistics, the estimator $\hat{\theta}$ defines a certain level of confidence as specified by (3.3.1). Let $\mathfrak{2}$ be a variable dependent on the observations $\{x_i\}_{i=1}^{n}$ and on a parameter θ , i.e. $\mathfrak{2} = f(\theta , x_1 , x_2,..,x_n)$ whose sample distribution is independent of θ , e.g., let $\mathfrak{2}$ be a sufficient statistic of the observations. Then given δ , we can find a z_1 , z_2 such that

$$\Pr \{z_1 \leq \mathfrak{2} \leq z_2\} \leq \delta$$

Consequently, by the inverse mapping $f^{-1}(\mathfrak{2})$ and the independence of θ , limits θ_1 , θ_2 can be determined such that

$$\Pr \{\theta_1 \leq \theta \leq \theta_2\} \leq \delta \tag{3.3.5}$$

This statement must be interpreted very carefully. It does not mean with probability δ that θ is within the specific limits of the interval, but rather says that with probability δ , the random interval $[\theta_1 , \theta_2]$ contains the parameter θ . In other words, on the average, in a proportion of δ cases, we are insured that θ lies within this interval.

For the parameters μ and θ , there are sufficient statistics having well-known distributions. For an unknown σ , the "student" statistic

$$t = \frac{\bar{x} - \mu}{s} \tag{3.3.6}$$

with

$$s^2 = \frac{1}{n - 1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

is a sufficient statistic of μ having a Student t- distribution with (n-1) degrees of freedom [52, p. 305]. Since this distribution is symmetric, given

$\delta = 1 - \alpha > 0$, there exist a $t_{n-1\ ,\ 1-\alpha/2}$ such that

$$\Pr\ \{-\ t_{n-1\ ,\ 1-\alpha/2} \leq t \leq t_{n-1\ ,\ 1-\alpha/2}\ \} = 1 - \alpha$$

or by (3.3.6)

$$\Pr\ \{\bar{x} - \frac{\sqrt{s}}{h}\ t_{n-1\ ,\ 1-\alpha/2} \leq \mu \leq \bar{x} + \frac{\sqrt{s}}{n}\ t_{n-1\ ,\ 1-\alpha/2}\ \} = 1 - \alpha \quad (3.3.7)$$

where $t_{n-1\ ,\ 1-\alpha/2}$ is determined from a table of the t - distribution [53]. For the variance of a normal population, the statistic

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} \quad (3.3.8)$$

is distributed as a chi-square distribution with (n-1) degrees of freedom. For large n (greater than 30), the chi-square distribution can be approximated by the normal distribution [52, p. 198] by

$$\Pr\ \{\chi^2 \leq \chi_1^2\} \approx \Phi(\sqrt{2\chi_1^2} - \sqrt{2n-3})$$

so that for a confidence coefficient of $1 - \alpha$

$$1 - \alpha = \Pr\ \{\chi_0^2 \leq \frac{(n-1)s^2}{2} \leq \chi_1^2\}$$

$$= \Pr\ \{\frac{2(n-1)s^2}{[\Phi^{-1}(1-\alpha/2) + \sqrt{2n-3}]^2} \leq \sigma^2 \leq \frac{2(n-1)s^2}{[\Phi^{-1}(\alpha/2) + \sqrt{2n-3}]^2}\ \}$$

where $\Phi(\cdot)$ is the evaluation of the standard normal distribution, N(0 , 1).

From the results in Table 3, 1, the confidence interval of μ and σ^2 can be computed using (3.3.7) and (3.3.9), respectively.

Table 3.2. Confidence Intervals of μ and σ^2

ψ	Confidence Coefficient	Confidence Interval of μ	Confidence Interval of σ^2
1.0	0.999	(.9745, 1.0594)	(.0484, .0781)
	0.90	(.9962, 1.0377)	(.0540, .0685)
	0.80	(1.000, 1.0331)	(.0553, .0666)
0.5	0.999	(1.2569, 1.4543)	(.2805, .4461)
	0.90	(1.3074, 1.4038)	(.3119, .3931)
	0.80	(1.3180, 1.3932)	(.3196, .3827)
0.25	0.999	(1.8493, 2.1025)	(.5519, .8462)
	0.90	(1.9141, 2.0378)	(.6091, .7537)
	0.80	(1.9277, 2.0241)	(.6229, .7353)

Knowing the exact value of μ and σ^2 for the sample population, we can assess the merits of the associated estimates using these Confidence Intervals (CI). Since, for 99.9% of the time, μ is outside the CI for ψ = 0.25 and ψ = 0.5, the associated estimates are poor. Only for ψ = 1.0 is the estimate $\bar{\mu}$ accuracy with a confidence coefficient of 80%. The variance estimates are overall poor with the inaccuracy increasing with decreasing ψ .

Although the estimation technique obtained generally poor estimates, other techniques have not yet been examined (or tested). From these methods, the CI can serve as a useful tool in insuring the accuracy between the estimate and the parameter and thereby obtain a bound necessary in (3.2.42).

In this regard, a qualitative investigation was made of the conditional mean π_t and variance V_t when a known parameter error was introduced into the signal model (2.3.1). In particular, for the same parameter set used in Table 3.1, an interarrival time sample path obeying (3.3.2) was generated (as in Section 2.4). The conditional mean (variance) was computed using this set of parameters and using the same set with one parameter perturbed. In Fig. 3.18, π_t and $\bar{\pi}_t$ (dotted line) have a 50% variation in μ. Similarly, a 50% variation in λ, μ, σ on the conditional variance V_t are shown respectively in Fig. 3.19-3.21.

These results reveal the robustness of the filter to parameter variation. Clearly the maximum jump estimate (defined in Section 2.4) is not affected by these perturbations for the parameter set considered. From Fig. 3.19, the variation to λ is quite small due to the fact that $\Delta\lambda = 0.01$. However, variations to μ and σ are not significant. Notice in Fig. 3.18, the parameter variation does cause different effects on π_t^c and π_t^d are discussed in the previous section.

3.4 Future Research

The results of the maximum jump or queue length estimate are promising for large queue. The insensitivity of the filter to parameter variations in λ, μ, σ is satisfying. However, certain areas still need to be investigated.

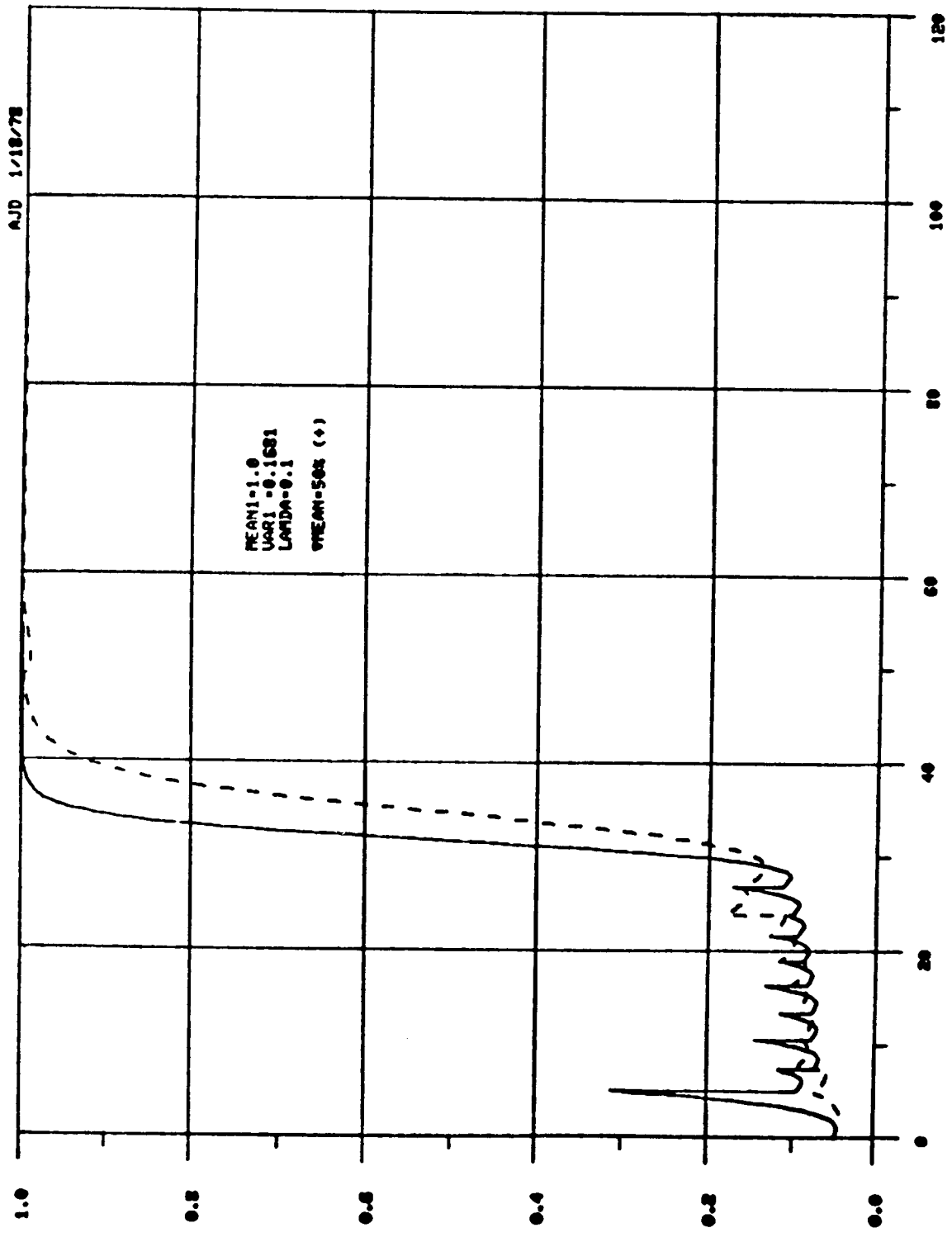


Fig. 3.18: Variation od Conditional Mean with respect to μ vs. time

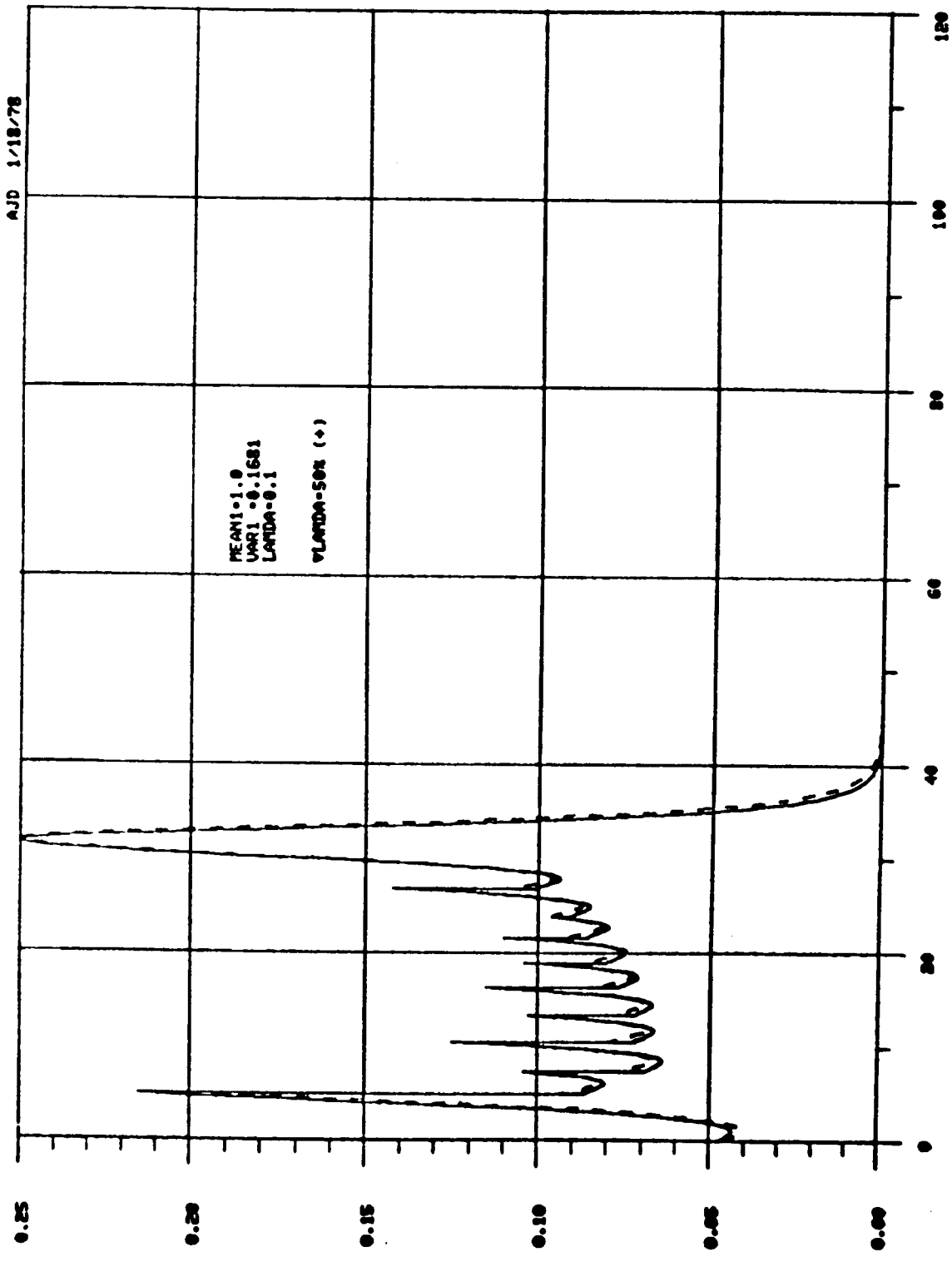


Fig. 3.19: Variation of Conditional Variance with respect to λ vs. time

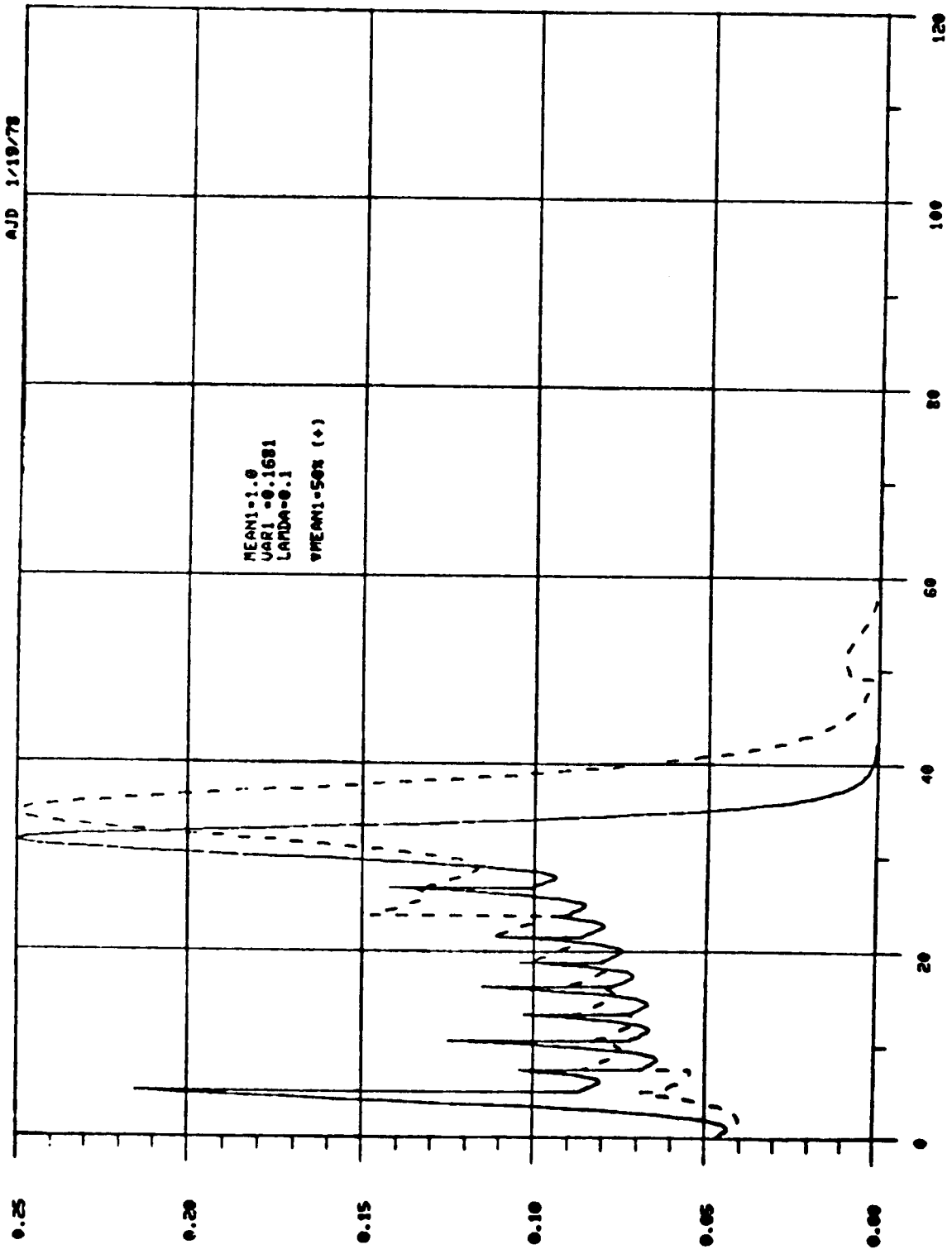


Fig. 3.20: Variation of Conditional Variance with respect to u_1 vs. time

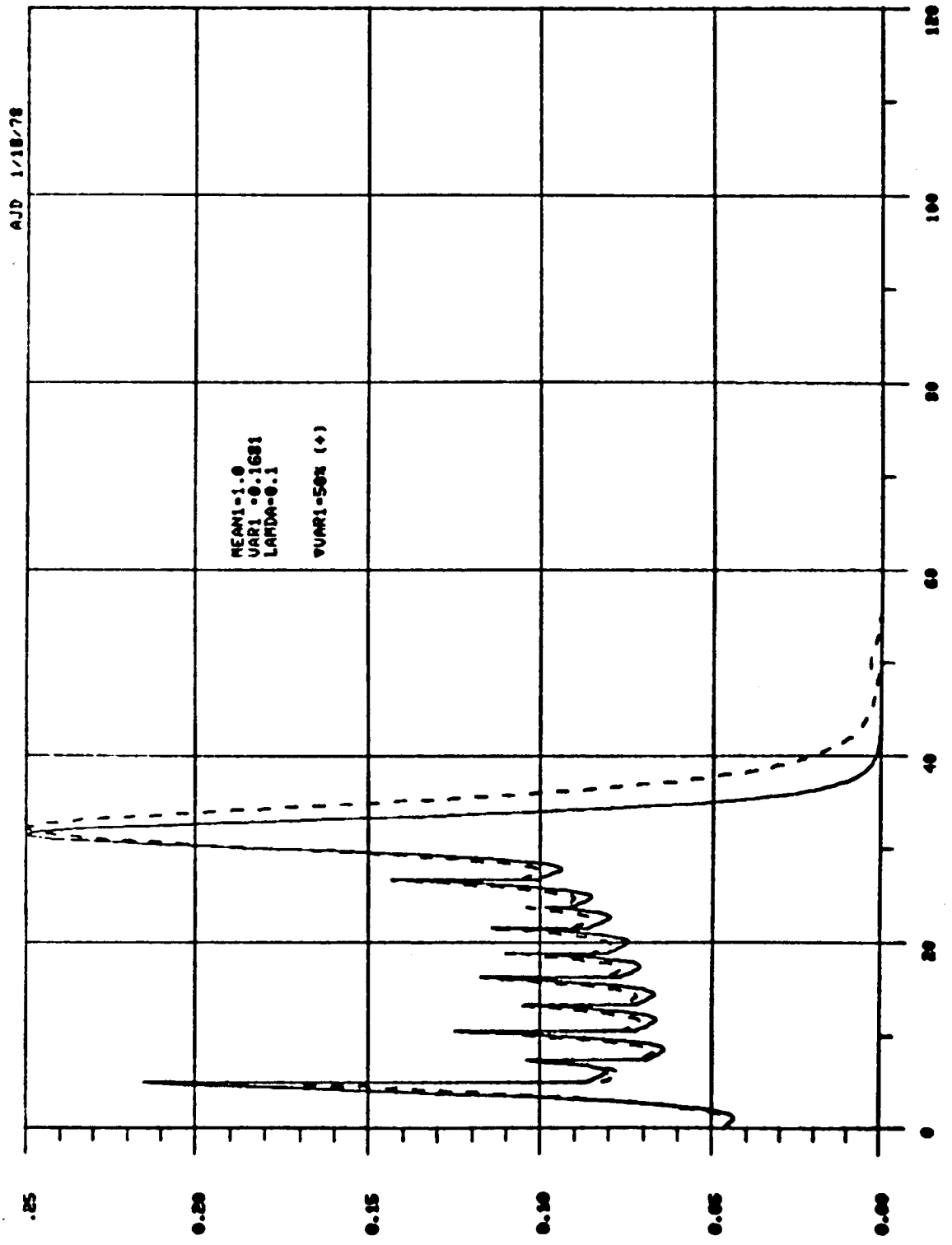


Fig. 3.21: Variation of Conditional Variance with respect to σ vs. time

The preliminary results in the parameter estimation show that at best the parameter estimates, needed in (3.3.2) are poor (see Table 3.1 and 3.2). Consequently alternative techniques need to be examined. Since the observed data is censored, the approach of Menhenhall et al.[54] seems promising. They obtain maximum likelihood parameter estimates, required in a mixed exponential density by using a modified likelihood function that incorporates the truncation in the observations. However, as Hill[55] noted, in general any estimate of the mixing parameter ψ is difficult to obtain when the two populations are not well separated. Although, these parameter estimates do not significantly effect the filter's performance (Fig. 3.18), the case is not the same for the a priori switch time density, p_n . We observed that for small queues (3 or less) or when a platoon of vehicles override a queue, the resulting estimate was unreliable. In other words, the uniform p_n, independent of the observations N_t, did not incorporate enough of the statistical information. Consequently to improve the filter's performance, the dependency of p_n on N_t needs to be examined with the more complex expressions (2.3.22),(2.3.24) replacing (2.3.14),' (2.3.16).

Besides these two areas, one may consider how the filter is effected by altering certain assumptions necessary in the above development. In our discussion, the headways were considered independent. Clearly though, following vehicles in urban traffic are influenced by the lead vehicle, i.e. ρ is nonzero.in (1.3.6) and the following rate λ_t^0 has the form (2.2.11). The hazard function will have the same shape (Fig. 2.11) but now depends on both the previous activation time and interarrival time and should improve the filter's performance. Assuming that the second rate λ_t^1 resulted from a homogeneous Poisson process, may have been premature. From the

filtering standpoint and because the observations are censored, the hazard function may have been more fittingly chosen to peak in a neighborhood to the left of λ_t^0 and decrease more rapidly for large t. Finally throughout this thesis, the observations have been considered counting processes and the estimate therefore is only applicable to single-lane traffic. The extension to multi-lane traffic and more general jump processes will be an involved matter and its fruitfulness is questionable.

In the onset of this research, we conjectured that the underlying headway probability was the same type for all different types of traffic (rural, freeway, urban, etc.) but the parameters may vary. In urban traffic, external inferences are induced into traffic and complicate the study of vehicle headways. Thus validation of the proposed model (3.3.2) against urban traffic has not been done previously and needs to be completed to insure any further extension. Also in our initial comments, we simplified are development by discarding the mark process associated with the activation times. The extension will result in summations over the mark space in the filter equations and the statistics required, derived from vehicle velocities, can be considered gaussian.

REFERENCES

[1] P. J. Tarnoff, "The Results of FHWA Urban Traffic Reasearch: An Interim Report," Traffic Engineering, vol. 45, pp. 27-35, April 1975.

[2] J. B. Kreer, "A Comparison of Predictor Algorithms for Computerized Traffic Control Systems," Traffic Engineering, vol. 45, pp. 51-56, April 1975.

[3] D. C. Gazis, Traffic Science, New York: John Wiley, 1974.

[4] D. J. Buckley, "Road Traffic Headway Distributions," Proc. Austr. Road Res. Board, vol. 1, pp. 153-187, 1962.

[5] L. C. Edie, "Flow Theories," pp. 1-108, In Traffic Science, New York: John Wiley, 1974.

[6] D. L. Gerlough and M. J. Huber, "Traffic Flow Theory," TRB Special Report 165, Transportation Research Board, National Research Counicil, Washington, D. C. 1975.

[7] W. F. Adams, "Road Traffic Considered as a Random Series," J. Inst. Civil Eng., vol. 4, pp. 121-130, 1936.

[8] A. D. May, "Gap Availability Studies," Highway Res. Record, no. 72, pp. 101-136, 1965.

[9] A. Daou, "On Flow Within Platoons," Austr. Road Res., vol. 2, pp. 4-13, July 1966.

[10] R. M. Oliver, "A Traffic Counting Distribution," Oper. Res., vol. 9, pp. 802-810, June 1961.

[11] G. F. Newell, "Statistical Analysis of the Flow of Highway Traffic through a Signalized Intersection," Q. Appl. Math., vol. 13, pp. 353-369, April 1956.

[12] A. J. H. Clayton, "Road Traffic Calculations," J. Inst. Civil Eng., vol. 16, pp. 247, 1941.

[13] C. J. Ancker, A. V. Gafarian, and R. K. Gray, "The Oversaturated Signalized Intersection-Some Statistics," Trans. Sci., vol. 2, pp. 340-361, Feb. 1968.

[14] A. J. Miler, "A Queuing Model for Road Traffic Flow," J. Roy. Statis. Soc. B, vol. 23, pp. 64-75, Jan. 1961.

[15] A. Schuhl, "The Probability Theory Applied to Distributions of Vehicles on Two-Lane Highways," pp. 59-75. In Poisson and Other Distributions in Traffic. Eno Foundation, 1955.

[16] A. Schuhl, "Probability Theory Applied to Vehicle Distributions on Two-Lane Highways," pp. 73-97. In Poisson and Other Distributions in Traffic. Eno Foundation, 1971.

[17] J.H. Kell, "Analyzing Vehicular Delay at Intersections Through Simulation," Highw. Res. Board Bull., no. 356, pp. 28-39, 1962.

[18] E.C. Sword, "Prediction of Parameters for Schuhl's Headway Distribution," Purdue Univ., 1967.

[19] R.E. Dawson and L.A. Chimini, "The Hyperlang Probability Distribution-A Generalized Traffic Headway Model," Highw. Res. Rec., no. 230, pp. 1-14, 1968.

[20] Highway Capacity Manual 1965. Highw. Res. Board Special Report 87, 1965.

[21] D.J. Buckley, "A Semi-Poisson Model of Traffic Flow," Trans. Sci., vol. 2, pp. 107-133, Feb. 1968.

[22] J. Pahl and T. Sands, "Vehicle Interaction Criteria from Time-Series Measurements," Trans. Sci., vol. 5, pp. 403-417, 1971.

[23] L.C. Edie, R.S. Foote, and R. Rothery, "Analysis of Single-Lane Traffic Flow," Traffic Eng., vol. 33, pp. 21-27, 1963.

[24] P. Athol, "Headway Grouping," Highw. Res. Record, no. 72, pp. 137-155, 1968.

[25] B.D. Greenshields and F.M. Weida, Statistics with Application to Highway Traffic Analysis. Eno Foundation, 1952.

[26] R.T. Underwood, "Traffic Flow and Bunching," J. Austr. Road Res., vol. 1, pp. 8-25, 1963.

[27] J.E. Tolle, "The Lognormal Headway Distribution Model," Traffic Engin. Control, vol. 13, pp. 22-24, 1971.

[28] N.L. Johnson and S. Kotz, Continuous Univariate Distributions-1. Boston: Houghton Mifflin, 1970.

[29] D. Branston, "Models of Single Lane Time Headway Distributions," Trans. Sci., vol. 10, pp. 125-148, Feb. 1976.

[30] D. Snyder, Random Point Processes, New York: Wiley Interscience, 1975.

[31] M.H.A. Davis, "The Representation of Martingales of Jump Processes," SIAM J. Control and Optim., vol. 14, pp. 623-638, 1976.

[32] P.K. Houpt, "Incident Detection by Generalized Likelihood Test: Interim Report," Federal Highw. Adm., Office of Research and Development, Washington, D.C., Sept. 1977.

[33] M.W. Szeto and D.C. Gazis, "Application of Kalman Filtering to Surveillance and Control of Traffic Systems," Trans. Sci., vol. 6, pp. 419-439, 1972.

[34] N.E. Naki and A.N. Trivedi, "Recursive Estimation of Traffic Section Density and Average Speed," Trans. Sci., vol. 7, pp. 269-286, 1973.

[35] L. Isaksen and H.J. Payne, "Freeway Traffic Surveillance and Control," Proc. IEEE, vol. 61, pp. 526-536, May 1973.

[36] P. Brémaud, "A Martingale Approach to Point Processes," Electronics Res. Lab. Memo M-345, Univ. of California, Berkeley, 1972.

[37] R. Boel, P. Varaiya, and E. Wong, "Martinagales on Jump Processes I: Representation Results," SIAM J. Control, vol. 13, pp. 999-1021, August 1975.

[38] R. Boel, P. Varaiya, and E. Wong, "Martingales on Jump Processes II: Applications," SIAM J. Control, vol. 13, pp. 1022-1061, August 1975.

[39] I. Rudin, "Regular Point Processes and their Detection," IEEE Trans. Inform. Theory, vol. IT-18, pp. 547-557, Sept. 1972.

[40] P.A. Meyer, Probabilités et Potential, Paris: Hermann, 1966, English translation, Probability and Potentials, Waltham, Mass.: Blaisdell, 1966.

[41] A. Segall, M.H.A. Davis, and T. Kailath, "Nonlinear Filtering with Counting Observations," IEEE Trans. Inform. Theory, vol. IT-21, pp. 125-134, March 1975.

[42] L.J. Galchuk and B.L. Rozovsky, "The Disorder Problem for a Poisson Process," Theory of Prob. and Appl., vol. 16, pp. 712-717, 1971.

[43] M.H.A. Davis, "A Note on the Poisson Disorder Problem," Pub. 74/8, Dept. of Computing and Control, Imperial College, London, Jan. 1974.

[44] C.B. Wan and M.H.A. Davis, "The General Point Process Disorder Problem," IEEE Trans. Inform. Theory, vol. 23, pp. 538-540, July 1977.

[45] D. E. Knuth, The Art of Computer Programming-Vol. 2, Reading, Mass: Addison-Wesley, 1973.

[46] C. Hastings, Numerical Approximations for Digital Computers, Princeton, 1955.

[47] Lieberman et. al., "Development, Validation and Testing of the UTCS-1 Network Simulation Model," Highw. Res. Record, no. 409, Jan. 1972.

[48] Lieberman et. al., "Logical Design and Demonstration of the UTCS-1 Simulation Model," Highw. Res. Record, no. 409, Jan. 1972.

[49] "Locating Detectors for Advanced Traffic Control Strategies," Federal Highw. Adm., Office of Research and Development, Washington, D. C., Final Report, August, 1975.

[50] A. P. Sage and J. L. Melsa, Estimation Theory with Applications Communications and Control, New York: McGraw-Hill, 1971.

[51] M. G. Kendall and A. Stuart, The Advanced Theory of Statistics-Vol. 2, New York: Hafner, 1971.

[52] P. L. Meyer, Introductory Probability and Statistical Applications, London: Addison-Wesley, 1970.

[53] D. B. Owen, Handbook of Statistical Tables, London: Addison-Wesley, 1962.

[54] W. Mendenhall and R. J. Hader, "Estimation of Parameters of Mixed Exponentially Distributed Failure Time Distributions from Censored Life Test Data," Biometrika, vol. 45, pp. 504-520, 1958.

[55] B. M. Hill, "Information for Estimating the Proportions in Mixture of Exponential and Normal Distributions," Amer. Statis. Assoc., vol. 58, pp. 918-932, 1963.