SCHOOL OF OPERATIONS RESEARCH
AND INDUSTRIAL ENGINEERING
COLLEGE OF ENGINEERING
CORNELL UNIVERSITY
ITHACA, NEW YORK 14853

TECHNICAL REPORT NO. 685

February 1986

NONPARAMETRIC INFERIENCE FROM
POISSON-TYPE COUNTING PROCESSES

By

Michael J. Phelan
# Table of Contents

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.1</td>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1.2</td>
<td>Survey of Problems Involving a Counting Process</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>1.3</td>
<td>Counting Processes Involving an Auxiliary Process</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>1.4</td>
<td>A Probability Model</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>A Brief Survey of Literature</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>1.6</td>
<td>Summary of Thesis</td>
<td>21</td>
</tr>
<tr>
<td>2</td>
<td>2.1</td>
<td>Counting Processes and Their Compensators</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>2.2</td>
<td>Nonparametric Estimation in Poisson-Type Counting Processes</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>2.3</td>
<td>Consistency of the Estimator $\hat{b}$</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td>2.4</td>
<td>Asymptotic Normality</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>2.5</td>
<td>Empirical Estimation of the Covariance Function</td>
<td>51</td>
</tr>
<tr>
<td>3</td>
<td>3.1</td>
<td>Preliminaries</td>
<td>59</td>
</tr>
<tr>
<td></td>
<td>3.2</td>
<td>Functional Convergence Theorems: Processes of the Class LS</td>
<td>65</td>
</tr>
<tr>
<td>4</td>
<td>4.1</td>
<td>Further Inference Problems for Poisson-Type Counting Processes</td>
<td>81</td>
</tr>
<tr>
<td></td>
<td></td>
<td>One Sample Case: Tests of the Class $K$</td>
<td>81</td>
</tr>
</tbody>
</table>
4.2 Two Sample Case: Tests of the Class K

4.3 Applications to Survival Analysis

4.4 Likelihoods for Poisson-Type Counting Processes:
   Self-Excited Case

5 Directions for Future Research

Appendix: Review of Semimartingales

References
CHAPTER 1

A Family of Multivariate Counting Processes

1.1 Introduction

To a casual observer the problem of characterizing the eating pattern of a predator animal and of describing certain dynamical aspects of labor markets may appear to be two very different problems, indeed, as different from each other as is characterizing consumer repayment behavior of home mortgages different from describing internal patterns of migration in a developing country. Certainly, if we were to explicate these problems to their particular details we could not conclude otherwise. Yet, if we abandon the particular view of things and instead seek to abstract from these problems general features of structure and process their interrelatedness is revealed. This discovery of binding relationships among cross-disciplinary problems makes possible the formation of classes which unify diverse problems according to their underlying dynamical organization.

Our investigation of applied problems has revealed the usefulness of random counting processes in describing certain dynamical aspects of a number of scientific research problems. Consider as an example, a medical study of the survival times of recent heart attack sufferers. Denote by $N_t$ the number of deaths which have occurred by time $t$ among the group of heart patients, and let $N = (N_t), t \geq 0$
denote a counting process. Each time a death occurs the counting process \( N \) increases by one and we may try to understand how different factors influence the propensity of an individual to die. Thus we may demonstrate the relative effectiveness of competing treatment regimes or illustrate the effects of moderate exercise (e.g. walking) on the survival times of each patient. Our ultimate goal might be to specify a model describing the mechanism behind this interaction.

In this thesis we investigate a family of multivariate counting processes wide enough to provide a unifying framework for a diverse body of scientific research problems that occur, for example, in areas such as demographics, consumer economics, animal ecology and labor economics. The models in this family are both simple and effective, thus giving accessible descriptions of the natural phenomena we explore and providing navigable routes toward scientific discovery. The family is wide enough to include both discrete and continuous time counting processes and also counting process models based on Cox's proportional hazards model (see [14]). The theory of nonparametric inference we give for this family extends in a number of directions the work on multivariate counting processes done by Aalen [2,3].

The formulation of multivariate counting processes that we consider is based on a class of associated processes called the compensator. We view the compensator as a random measure on the positive real line and assume that the kernel for this measure has a relatively simple multiplicative form. The form will typically
be composed of the product of an observable stochastic process, \( Y \) say, and a deterministic function of time. The class includes the multiplicative intensity model considered by Aalen [3], who assumes that the intensity exits. This assumption means that the deterministic function in our model is absolutely continuous relative to the Lebesgue measure. We make no such assumption in the class of models we consider and consequently include both discrete and continuous time counting processes. Our treatment is thus more general.

Our approach to this class of models is based on Jacod [21] and relies on the martingale dynamics over point processes and the theory of stochastic integration and random measures. This theory is used to develop a theory of nonparametric inference for the family of counting processes under consideration. The statistical framework is to view the deterministic function as an unknown parameter in our probability model. Our main results concern the estimation of the unknown function in the model and the "asymptotic" distribution theory for those estimates. The asymptotic theory requires that we be given a sequence of related counting processes and is of particular usefulness when inferences are to be made from the observation of large numbers of, say, independent realizations of the counting process. To obtain the distributional or weak convergence results we cannot directly apply the martingale functional central limit theorems of Rebolledo [31] and McCleish [28] as was possible in Aalen [2]. Rather, to deal with the generality of our model we
need to adapt the semimartingale functional convergence theorems of Lipster and Shiryaev [27] and Jacod, Klopotowski and Mémin [22] to counting process applications. In the end, the theory allows us to develop nonparametric estimates and asymptotic confidence bands for unknown functions in the model, the comparison of subpopulations on these counting processes and tests of hypotheses with known asymptotic properties.

In the remainder of this chapter we survey some scientific research problems and give informal definition to the family of counting processes considered in this thesis. We show how the special features of these problems are naturally accommodated by these general counting processes and survey some literature relevant to our problem. To close this chapter we give a chapter-by-chapter summary of the thesis and highlight the main results.

1.2 Survey of Problems Involving a Counting Process

There are many problems in science involving the occurrence of a sequence of related events over time such as when observing the evolution of behavioral decisions or choices made by an organism. These observations may be represented by a sequence of epochs or chronological times at which events occur together with a mark to indicate the kind of event to have occurred. Such sequences are called point processes and we have found it convenient to equivalently express these processes using a multivariate counting process.
Although a counting process provides a convenient and concise way to represent sequences of interrelated events, in many problems this may not be enough. For instance, a predator's decision to hunt a given prey species depends on many factors including prey availability and cost of the hunt in light of the expected caloric intake. Thus, we typically find numerous auxiliary processes which interplay in complex ways evolving in concert with a point process. We consider four examples of this type of interplay here.

Example 1: Demographics. Newly developing countries, such as those of tropical Africa, have in recent decades experienced large increases in the size of their urban population. This growth is in part due to increases in their total populations, but it has also been recognized that a most important contributing factor has been a massive influx of migrants from rural areas into urban centers; see Todaro [36].

The decision to migrate into urban areas is based upon many considerations, all of which contribute to the prevailing perception among rural dwellers that a measurably better life is to be had in these growing urban centers. Specifically, rural migrants go to urban areas in search of employment and better wages. However, the economies of urban areas do not grow at rates as fast as many economic planners would like and this leads to slow growth in urban employment opportunities. The sluggish growth in urban employment possibilities coupled with rapid influx of migrants
induces extensive unemployment in urban centers. The burden these unhappy circumstances place on the already scarce resources can have destabilizing effects on a developing society. Thus African governments trying to alleviate the problems associated with growing unemployment must develop strategies to slow the rate of rural-to-urban migration.

Let \( N_t(1) \) and \( N_t(2) \) record the number of migrations from rural to urban areas and urban to rural areas, respectively, over the time interval \([0, t]\) and denote by \( N(1) = (N_t(1)), \ t \geq 0 \) and \( N(2) = (N_t(2)), \ t \geq 0 \) the corresponding counting processes which arise by recording these migrations for all times \( t \). The problem outlined above is for government planners to develop ways in which to slow the growth of the process \( N(1) \) and possibly increase the growth of the process \( N(2) \) until an optimum populational distribution is achieved. Clearly, an understanding of the factors governing the rate of flow between rural and urban areas is required to attain this goal.

Todaro [36] cites two principle economic determinants of an individual's decision to migrate. The most important of these seems to be the urban-rural real wage differential whereby urban wages usually exceed rural wages for jobs requiring similar skills. Secondly, individuals face different "probabilities" of securing a job in the two environments. In many developing countries, such as in sub-Saharan Africa, the wage differential favors the urban areas as does the chance of securing employment despite teeming
unemployment in urban centers. Thus a resourceful model of internal
migration would characterize the interplay between these factors
and the rate of growth of the counting processes introduced above. □

Example 2: Animal Ecology. In the ecology of natural systems one
of the more common forms of interspecies interactions is the
predator-prey relationship. From the perspective of the predator's
survivability, one may argue that when a predator harvests food
it must do so efficiently in order to help ensure reproductive
success (see for example Krebs [23]). The foraging process of an
animal involves many decisions and choices coordinating such
aspects as where to hunt, what to eat with the profitability of the
prey, the need for variation in its diet, and so on. Typically an
animal will have a diet set over which it varies its choice of
prey and it is this selection process we wish to focus on here.

Consider the example of the predator lynx on Newfoundland;
an island mass off the Atlantic coast of Canada near Quebec and Nova
Scotia. Here the lynx usually varies its diet over the set of
prey animals consisting of snowshoe hare, arctic hare and caribou
calves. The snowshoe hare is the obligatory prey species for the
lynx but occasionally the lynx will switch prey and hunt arctic hare
or caribou calves. In this century the numbers of arctic hare
and caribou began to rapidly decline due to predation by lynx, a
situation which had a dramatic effect upon the fragile ecosystem
on Newfoundland. For example, the caribou population declined from
about 40,000 in 1900 to about 1,000 or fewer by 1925; see Bergerud [8].
The ecosystem on Newfoundland is fragile and the effect of this prey switch was dramatic because the usual pyramidal configuration of species is inverted on the island leaving many predator species at the top to hunt fewer prey species at the bottom. Thus, it is relatively easy for an entire species to become extinct and permanently alter the island's ecology. Hence, in this situation there is a particularly acute need to understand the propensity for this kind of prey switch to occur and perhaps manage its determinants.

One way to summarize the eating pattern of the lynx is with a counting process \( N = \{ N_t(i,j), t \geq 0; 1 \leq i,j \leq 3 \} \), where 1, 2 and 3 represents snowshoe hare, arctic hare and caribou, respectively. Thus, for example, we use \( N_t(1,2) \) to count the number of prey switches made by the lynx from snowshoe to arctic hare over \([0,t]\). We may now hope to describe the mechanisms by which environmental pressures and the foraging habits of the lynx conspire to generate the dietary pattern we summarize with \( N \).

In his field investigation of lynx predation on Newfoundland, Bergerud [8] was able to uncover the primary factor governing a prey switching cycle for the lynx by looking into (among other things) the population dynamics of the snowshoe hare, the caribou, the lynx and the arctic hare. The main feature was that as the snowshoe hare entered a cyclic population decline there followed a subsequent decline in the lynx population size. This decline in lynx numbers is shortly thereafter followed by a recovery which coincides with decreases in the population levels of caribou and arctic hare. The
lynx, a resourceful predator determined to survive, would recover from its diminishing numbers by switching prey from the periodically scarce snowshoe hare to the readily available and easy prey animals, arctic hare and caribou. These observations certainly suggest that this dynamical system of interspecies interaction between predator and prey is, in large part, driven by the intrinsic population cycle of the snowshoe hare.

Example 3: Consumer Economics. The decision-making processes used by consumers in planning their consumption and leisure time activities over the course of their economic lives is perhaps not unlike the foraging process we encountered in example 2. In general, these problems of intertemporal choice can be quite complex, so we narrow our focus here to a discussion of a discrete choice problem arising in a study of consumer mortgage repayment behavior. The details of this study and the discussion that follows may be found in Zorn and Lea [38].

In the financial markets of Canada and the United States consumers may obtain mortgages on their homes having adjustable mortgage rates. This means that the interest rate paid by the consumers may vary with changes in the macroeconomic environment over the amortization period of the loan. The consumer (mortgage holder) will typically renegotiate with the bank the interest rate owed on a loan a number of times during this period. At the same time, the borrower and the bank fix a term ranging from 1 to 5 years which determines the period over which the latest negotiated interest rate is to apply. The usual contract with the bank asks the consumer to make monthly
payments on the principal of the loan until it is fully repaid with interest. The monthly repayment behavior of the consumer is what concerns us here.

Each month, when a payment is due, the consumer faces a decision with discrete choices as to what repayment action is to be taken. Basically, the consumer may take one of five alternative actions.

1 $\equiv$ normal payment of contractually required amount;
2 $\equiv$ prepayment, where a normal payment plus an additional amount less than the balance owed is paid;
3 $\equiv$ delinquency, where a payment is not made;
4 $\equiv$ default, where an uninterrupted sequence of delinquencies leads the bank to take possession of the house;
5 $\equiv$ repayment, where the current balance owed is fully repaid.

Normal payments are desirable from the point of view of the bank but this action may not always meet the utility requirements of the consumer.

We can describe the monthly decisions of the consumer by a sequence $D_1, D_2, D_3, \ldots$ where $D_t = i$ when action $i$ is taken in month $t$ ($i = 1, \ldots, 5$, $t = 1, 2, \ldots$). We thus obtain a multivariate counting process $N = \{(N_t(1), \ldots, N_t(5))\}$, where for example $N_t(3)$ counts the number delinquent actions taken by the consumer over $[0, t]$. This discrete time counting process $N$ is the process we would use to characterize consumer repayment behavior and to empirically investigate the factors in the economy and characteristics of the individual which govern this behavior.
In the United States lending institutions did not always offer adjustable rate mortgages but instead offered fixed rate mortgages where the interest rate on the loan is fixed. The fixed rate mortgage leaves the lending institution in the position of absorbing all the effects of interest rate fluctuations in the economy. During periods of high inflation that occurred in the late 1970's and early 1980's, this so-called interest rate risk together with other factors brought about a serious net worth problem for many lenders. To help alleviate this problem lending institutions have forced some of the interest rate risk upon consumers by switching to adjustable rate mortgages. In order to assess the effectiveness of their risk reduction strategy and to profitably plan on a portfolio level, lenders need to understand the mortgage borrowers' repayment behavior under adjustable rate mortgages.

As our discussion thus far suggests, one of the likely leading economic factors influencing repayment behavior (or N) is interest rates. Obviously, the pressures of changing interest rates will force consumers to adjust their financial behaviors, including the repayment action they take each month. Certainly, characteristics of the individual's financial history will greatly affect the ability of different consumers to adjust to these pressures. Therefore, both macroeconomic factors and individual characteristics of the borrower will interplay strongly in the dynamics of a consumers' repayment behavior. We believe, further,
that understanding the impact of these contributing factors will alert government policy makers to the need to help certain target groups adjust favorably (e.g. working poor people) and to help set conditions for lenders to plan fairly and soundly.

Example 4. Labor Economics. There exists a huge literature dealing with the problems and issues of the unemployment question in terms of government policy analysis, theoretical econometric models and so forth. We shall confine our attention to some econometric considerations arising in studies of unemployment durations and empirical estimation of the reemployment probability.

An individual member of the labor force may be described as being in one of three states - employed, unemployed and participating or unemployed and not participating (see for example [13]). The latter two states are formed by dividing the unemployed into those actively seeking work (i.e. participating) and those who are not. Thus, the employment history of an individual may be summarized as a sequence of occupied labor force participation states together with the corresponding sojourn time in those states. The three-state model of labor force dynamics may be characterized using a counting process in the manner of the prey switching problem of Example 2.

The duration of unemployment is the time it takes an unemployed individual to become reemployed and may be thought of as a positive random variable. To estimate the reemployment probability it is necessary to estimate the distribution function of unemployment duration. One issue in this context is whether or not an
individual's reemployment probability shows duration dependence in that the chance of future employment depends conditionally on how long an individual has been unemployed. In the case where duration dependence is absent then the distribution of unemployment duration is exponential:

\[ P(X > t) = e^{-\lambda t} \quad \text{for } \lambda > 0, \ X \equiv \text{duration and } \forall t \geq 0. \]

Lancaster [24] has argued that the exponential model gives reasonable correspondence to observed data when heterogeneity among unemployed persons is taken into account. Failure to subdivide the population of unemployed into relatively homogeneous subpopulations may lead to spurious conclusions about duration dependence as is believed to be the case in Clark and Sommers [15].

In addition to subdividing the unemployed according to race, sex and education, for example, Lancaster has further argued for the need to incorporate a reservation wage process. The reservation wage measures the minimum wage offer an unemployed person will accept in evaluating a job offer and will help to explain why some job offers are rejected and the individual continues to search.

Burdett, Kiefer, Mortensen and Neumann [13] used an extension and unification of classical participation and wage search theories to derive a Markov chain model of labor force dynamics over the three participation states mentioned earlier. The model is a time-homogeneous Markov chain and the transition rates they derive are functions of the "expected wage" and individual characteristics of workers such as age, education, race, sex and so on.
1.3 Counting Processes Involving an Auxiliary Process

Each of the examples described above gave rise to a counting process characterizing one of the observable components of the phenomenon discussed. In a natural way the counting process served as summary statistic for a sequence of discrete events occurring over time such as the sequence of prey eaten by the lynx in Example 2 (animal ecology). The common existence of a counting process component in each example serves as a basis for drawing them into a class.

There is a second feature that is common to the problems in these examples. It appears that we can identify in each case an auxiliary process influencing the course of events observed.

Example 1: Migration. Here we found that an individual's decision to migrate and thus the number of individuals migrating is influenced by the urban-rural real wage differential.

Example 2: Prey Switching. In this situation we found that the propensity for the lynx to switch to hunting caribou calves is closely tied to the availability of its obligatory prey species snowshoe hare. In this way the number of prey switches observed for the lynx over time depends on the population level of its primary prey species, the snowshoe hare.

Example 3: Mortgage Repayment. Here we found that the consumer repayment decision is influenced by such macro-economic factors as interest rates and thus the number of defaults, repays, etc. we observe depends on these rates.
Example 4: Unemployment. Here we found that an individual's unemployment duration is influenced by a host of demographic characteristics as well as job offer rates and a reservation wage process. Hence, the number of unemployed individuals finding a job will depend on all these factors.

Thus, we find that all the above examples give rise to a stochastic process \((N,Y)\) when \(N\) is a multivariate counting process recording the occurrence frequencies of discrete events whose propensity to variously manifest is governed by an auxiliary driving stochastic process \(Y\). This general framework unifies problems such as those described in our earlier examples into a class.

1.4 A Probability Model for \(N\)

In this thesis we develop a probability model for a family of multivariate counting processes which is described here in informal terms. It suffices to do this for \(m = 1\), namely, when all the events are of the same type. Thus we consider the process \(N = (N_t, t \geq 0)\) where now the counting process \(N\) is one dimensional.

We introduce a filtration \(F = (F_t, t \geq 0)\) (i.e. \(s \leq t \implies F_s \subset F_t\)) and assume that the process \(N\) is \(F\)-adapted, meaning \(N_t\) is an \(F_t\)-measurable random variable for each \(t \geq 0\). Denote by \(N = (N_t, F_t, t \geq 0)\) an \(F\)-adapted counting process \(N\) and recall that from the theory of counting processes there exists a nonnegative nondecreasing process \(A = (A_t, F_t, t \geq 0)\) called the compensator for
N relative to F. Notice that A is F-adapted and when A is $F_t$-predictable it is unique. The compensator A earns its name in that A compensates N up to a martingale, i.e. $N-A = (N_t-A_t, F_t)$, $t \geq 0$ is a martingale. This suggests the following decomposition:

$$ (1.1) \quad N = A + M $$

where A is the compensator to N and M denotes a zero-mean martingale. We may think of (1.1) as a decomposition of a counting process into the sum of a trend component A and a martingale noise component M. Alternatively, on taking expectations of (1.1) we may think of A as a running stochastic mean for N in the sense that $EN = EA$.

Our approach is to view the process A as a random measure on $\mathbb{R}^+$ and specify a probability model for N in terms of this random measure. Thus, the notation $A(\cdot)$ will be used to denote the random measure assigned to the set $\{\cdot\}$ by the process A. Consider the following definition.

**Poisson-Type Counting Process.** Let $b$ denote a nonnegative nondecreasing function on $\mathbb{R}^+$ with $b(0) = 0$ which we think of equivalently in terms of the induced measure $b$ generates on the Borel sets in $\mathbb{R}^+$. A counting process $N = (N_t,F_t)$, $t \geq 0$ is called a Poisson-type counting process whenever its compensator A satisfies:
(1.2) \( A\{dt\} = Y_t b\{dt\}, \quad \forall t \geq 0, \)

where the auxiliary process \( Y = (Y_t, F_t), \quad t \geq 0 \) is \( F \)-predictable and nonnegative. □

The appellation Poisson-type is also used in Lipster and Shiryayev [26] and is suggested by the fact that when \( Y_t = \lambda > 0 \) (\( \lambda \) is a constant) and \( b(t) = t, \quad \forall t \geq 0, \) then \( N \) is the ordinary Poisson process. More generally, if \( Y \) is an \( F_0 \)-measurable random function and \( b \equiv t \) then \( N \) is the so-called doubly-stochastic Poisson process. We note that the kernel \( A\{dt\} \) is also called the dual predictable projection by Jacod [21].

The general model for the compensator of a counting process may be thought of intuitively if we consider the following two special cases.

(i) Suppose that the function \( b \) is absolutely continuous with respect to the Lebesgue measure and let \( \alpha \) denote the Radon-Nikodym derivative of \( b \) relative to Lebesgue measure. In this case the measure \( A \) has the form \( A\{dt\} = Y_t \alpha(t) dt, \quad \forall t \geq 0. \) We further assume that the following relation holds:

\[
(1.3) \quad P(N_{t+h} - N_t = 1|F_t) = Y_t \alpha(t) h + o(h) \text{ as } h \downarrow 0.
\]

Hence, for small \( h \)'s \( Y \circ h \) approximates the conditional probability an event occurs in the time span \((t,t+h]\) given the history of the
process to time $t$; i.e. $F_t$. The product process $Y_\alpha$ is called the intensity process and gives the instantaneous rate at which events occur in $N$. $\square$

(ii) Suppose that $b$ has support on the discrete set 
$\{t_k\}, k \geq 1$ where $0 < t_1 < t_2 < \ldots < t_k \ldots$. For each $k$ let $\alpha_k$ denote the mass assigned to the point $t_k$ by the measure $b$ for $k = 1, 2, \ldots$. Now fix $k \geq 1$, put $F_{t_k} = \bigcap_{\varepsilon > 0} F_{t_{k+1} \varepsilon}$ and consider the following:

\begin{equation}
(1.4) \quad P(N_{t_{k+1}} - N_{t_k} = 1 \mid F_{t_k^-}) = E(N_{t_{k+1}} - N_{t_k} \mid F_{t_k^-})
\end{equation}

\begin{align}
&= E(Y_{t_{k+1}} \circ \alpha_{k+1}^+ \mid F_{t_k^-}) \\
&= Y_{t_{k+1}} \circ \alpha_{k+1}, \quad k = 1, 2, \ldots,
\end{align}

where the second equality uses the martingale property and the last uses the predictability of $Y$. Thus, $Y_\alpha$ gives the conditional probability of an event to occur in $N$ given the history of the process up to the instant before that event may happen which is roughly a rate of occurrence in discrete time. $\square$

These examples illustrate the way in which the auxiliary process $Y$ governs the propensity of events to occur in $N$. However, the auxiliary process serves not only as a driving process but as a mechanism for characterizing the interdependence of future events.
with past events. To see this recall that when \( Y \equiv 1 \) the compensator \( A \) is deterministic, implying that the counting process \( N \) has independent increments. Therefore, the presence of a general process \( Y \) induces a probabilistic dependence between future events and past events of \( N \).

1.5 A Brief Survey of Literature

In the remainder of this thesis we develop a theory of nonparametric inference for the family of Poisson-type counting processes. The statistical framework for this theory is to assume the process \((N, Y)\) is an observable stochastic process and to view the function \( b \) as an unknown parameter. The model is termed nonparametric since \( b \) is not specified beyond the minimal requirements of (1.2).

A theory of nonparametric inference for a family of multivariate counting processes is carried out in Aalen [3], who assumes that the intensity exits. This corresponds to assuming that the function \( b \) in our model is absolutely continuous relative to the Lebesgue measure as was done in equation (1.3). Motivated by the relation in (1.3) Aalen [3] defines an estimator \( \hat{b} \) for the function \( b \) as follows:

\[
(1.5) \quad \hat{b}_t = \sum_{s < t} \frac{Y_s - 1}{Y_s} \cdot 1(Y_s > 0) \Delta N_s, \quad t \in [0,1]
\]

(Aalen restricts his process to the interval \([0,1]\)). Then assuming we are given a sequence \((N^n, Y^n), n \geq 1\) of observed processes satisfying (1.3) for each \( n \) with the function \( \alpha \) being independent of \( n \),
Aalen develops an asymptotic theory for the sequence of estimators \( \hat{b}^n \) \( n \geq 1 \) defined by (1.5). In particular, Aalen [3] shows that \( \hat{b}^n \) converges in probability to \( b \) uniformly over \([0,1]\) and when suitably normalized converges in law to a continuous Gaussian process as \( n \) tends toward infinity. To develop this theory Aalen relied on the martingale dynamics over point processes and the functional central limit theorems of McCleish [28] (see [2]).

This general theory of counting processes is shown to provide a framework for the study of censored survival data by Aalen [3] and inhomogeneous Markov chains by Aalen and Johansen [4]. In addition, the ideas underlying this theory have proved useful in many other contexts. For example, Andersen and Gill [5] used Aalen's approach in developing a nonparametric estimation procedure for Cox's proportional hazards model extended to counting processes (see also Prentice and Self [30] for a related extension along these lines). Nonparametric \( k \)-sample tests are developed using counting process theory by Andersen, et al. [6] and the methods have also been used in deriving asymptotic normality of maximum likelihood estimators in parametric counting process models by Borgan [11]. Further, the tools of counting processes, stochastic integration and random measures, have been used by Gill [18] to study the product limit estimator of an arbitrary continuous distribution function \( F \) on \( \mathbb{R}^+ \) and by Slud [35] in a general treatment of two sample clinical trials involving arbitrary censorship of survival times.

In our treatment of multivariate counting processes we do not assume that the intensity exits nor do we restrict our processes to the interval \([0,1]\). In order to deal with this generality we have relied
the theory of multivariate counting processes in Jacod [21] and the general functional convergence theorems for semimartingales found in Jacod, Klopotowski and Mémin [22] and Lipster and Shiryaev [27]. Our general theory may be used to extend much of the work outlined here to a wider class of counting process models.

1.6 Summary of Thesis

In section 1.2 we describe four problems arising in the areas of demographics, economics and animal ecology. For example, in many developing countries there has been a massive influx of people from rural areas into overcrowded urban areas, causing extreme levels of unemployment, growing poverty and other societal hardships. The patterns of internal migration in these countries may be characterized as a flow of individuals between rural and urban areas and the problem for their governments is to manage these flows to best meet the demands of development. A principal economic influence on the rates of internal migration is the real wage differential whereby urban wages usually exceed rural wages for jobs of similar skills; clearly the prospect of better wages acts as an incentive to move to urban areas for individuals wanting a higher standard of living - a fact which government planners should take into account.

In each of the problems we describe the quantities of interest can be represented as \( \{ N_t, Y_t \} \), where \( N_t \) records the frequency of occurrences of events of different types and \( Y_t \) acts as a driving force on the rate of occurrences of these events. Thus \( N \) is a
multivariate counting process and Y an auxiliary process. In the migration problem above N has two components: one records the frequency of rural to urban migrations and the other the frequency of urban to rural migrations; Y measures the real wage differential. A second example is provided by the doubly-stochastic Poisson process of events (often called the Cox process) where the rate \( \lambda \) is a stochastic process and may be viewed as a driving force (that is, Y is identified as \( \lambda \)). In the general case, associated with the counting process \( \{ N_t \} \) there exists a stochastic process \( \{ A_t, t \geq 0 \} \) (called the compensator function) which represents the cumulative rate of occurrences of the events over a period of time, so that \( EA_t = EN_t \).

In the model we consider in this thesis the compensator A involves the process Y, the precise manner of which can be explained as follows in the case where we consider events of a single type. Here

\[
(1.6) \quad A_t = \int_0^t Y_s \, db(s)
\]

where Y is a nonnegative "predictable" stochastic process and b is a Borel measure on the real line. For the double-stochastic Poisson process b is the Lebesgue measure and conditional on Y, N is a non-homogeneous Poisson process. For this reason in the general situation a counting process N whose compensator A is given by (1.6) is called a Poisson-type counting process.

In Chapter 2 we consider the situation where N and Y are observable processes whereas b is unknown, beyond the fact that it is a Borel measure. The problem here is to estimate b from
observations of \((N,Y)\). Actually the observed data consists of \(n\) independent Poisson-type counting processes \(\{(N^i,Y^i), i = 1, \ldots, n\}\) where \(N^i\) has compensator \(A^i = \int Y^i db\), \(i = 1, \ldots, n\). We denote

\[
(1.7) \quad N^{(n)} = \sum_{i=1}^{n} N^i \quad \text{and} \quad Y^{(n)} = \sum_{i=1}^{n} Y^i,
\]

so that \(N^{(n)}\) is a random counting measure with compensator function \(A^{(n)} = \int Y^{(n)} db\). As the estimator of \(b\) we take \(\hat{b}^n\) where

\[
(1.8) \quad \hat{b}^n_t = \sum_{s < t} (Y^{(n)}_s)^{-1} 1(Y^{(n)}_s > 0) \Delta N^{(n)}_s, \quad \forall t > 0
\]

where \(\Delta N^{(n)}_s = N^{(n)}_s - N^{(n)}_{s-}\). Thus we estimate \(b\) on the set \(\{t: Y^{(n)}_t > 0\}\). This estimator \(\hat{b}^n\) was actually proposed by Aalen [3] to estimate the intensity function in the case where the measure \(b\) is assumed to be absolutely continuous relative to the Lebesgue measure (the so-called multiplicative intensity model). Accordingly we call (1.8) the Aalen estimator.

This problem is clearly one of nonparametric estimation from stochastic processes since \(b\) is left unspecified beyond the fact that it is a Borel measure and our only assumption on \(N\) is that it is a Poisson-type counting process with parameter \(b\) as shown in (1.6). Thus we rely mainly on asymptotic (that is, as \(n \uparrow \infty\)) arguments to develop properties of the estimator \(\hat{b}^n_t\). The most important observation in this regard is the following. Denote
(1.9) \[ \tilde{b}_t^n = \int_0^t 1(Y_s^{(n)} > 0) \, db(s) \quad \forall t \geq 0 \]

(the "true value" corresponding to the estimator \( \hat{b}_t^n \)). The error of estimation process, namely

\[ \hat{b}_t^n - \tilde{b}_t^n = \int_0^t (Y_s^{(n)})^{-1} 1(Y_s^{(n)} > 0)(dN_s^{(n)} - Y_s^{(n)}) \, db(s), \]
\[ \forall t \geq 0 \]

is a martingale (more generally a local martingale) so that, in particular it has zero-mean and orthogonal increments. Thus the estimator \( \hat{b}_t^n \) is unbiased and the estimation errors over disjoint intervals are uncorrelated. We establish the local uniform consistency and asymptotic normality of the estimator. Thus we have

(i) local uniform consistency: For each \( t \) such that
\[ Y_s^{(n)} \xrightarrow{p} \infty \text{ uniformly on } [0,t] \text{ as } n \uparrow \infty, \]

(2.1) \[ \sup_{s \leq t} |\hat{b}_s^n - b(s)| \overset{p}{\to} 0 \text{ as } n \uparrow \infty \]

where we have written \( b(s) = b([0,s]) \).

(ii) asymptotic normality: normalized error process
\[ n^{1/2}(\hat{b}_t^n - \tilde{b}_t^n) \text{ converges weakly to a Gaussian process} \]

of the independent increments as \( n \uparrow \infty \).
The result (ii) was proved earlier by Aalen [3] for his multiplicative intensity model. To deal with the more general Poisson-type counting processes treated in this thesis we have used the theory of functional convergence as developed by Jacod, Klopotowski and Mémin [22].

We note the general features of the Stieltjes integral shown in equation (2.0) as that of a stochastic integral of a suitable integrand $X^\mathbb{n}$ relative to a random measure $N^{(n)} - A^{(n)}$ ($A^{(n)} = \int Y^{(n)} db$) which is a martingale measure as well. It turns out that a large class of local martingales is formed by considering Stieltjes stochastic integrals of the form

\begin{equation}
M^\mathbb{n}_t = \int_0^t X^\mathbb{n}_s \left( dN^{(n)}_s - dA^{(n)}_s \right), \quad t \geq 0
\end{equation}

where $N^{(n)} - A^{(n)}$ is a counting process local martingale and $X^\mathbb{n} = \{(X^\mathbb{n}_s), s \geq 0\}$ satisfies certain measurability requirements of a technical nature. We use the structure of this class to prove a general theorem for the weak convergence of sequences $\{(M^\mathbb{n}), n \geq 1\}$ of the form (2.2) to $M^\infty$ (an arbitrary process of independent increments) by adapting the theory of Jacod, et al. [22] to this application. This is done in Chapter 3. The weak convergence result for $n^{1/2}(\tilde{b}^n - \tilde{b})$ derived in Chapter 2 then appears as a special case of this theorem.

In Chapter 4 we consider further inference problems for Poisson-type counting processes. First we consider one and two sample hypothesis tests for these processes and apply our results to
inference problems arising in survival analysis where our general theorem of weak convergence again finds application. In the hypothesis testing situations we outline a general procedure for constructing a broad class of asymptotic size \( \alpha (\alpha \in (0,1)) \) is the chosen level of significance) tests and showing them to be consistent. In the survival analysis setting this class would include common tests due to Wilcoxon, Savage and those considered by Gill [18], namely Efron's, Gehan's and Cox's. As a further application of our theory we give an elegant proof, avoiding a rather elaborate construction used by Gill [18], of the asymptotic normality of the Kaplan-Meier estimator of an unknown distribution function in a random censorship model.

A Poisson-type counting process \( N \) is also a self-excited counting process when the auxiliary process \( Y \) depends only on the history of \( N \) (i.e., for each \( t \), \( Y_t \) depends only on \( \{N_s : s \in [0,t]\} \)). For this case we give the necessary and sufficient conditions for explicit representation of the likelihood function for the process in a convenient exponential form. For certain subclasses of these counting processes this exponential representation can be used to exhibit sufficient and possibly complete statistics in the general case and in the parametric case it can be used in likelihood based estimation procedures such as maximum likelihood.

In Chapter 5 we indicate directions for further research in the areas of parametric, semi-parametric counting process models and applications to estimation from Markov renewal processes.
CHAPTER 2

Nonparametric Estimation for Poisson-Type Counting Processes

2.1 Counting Processes and their Compensators

Let \((\Omega, F, P)\) be a probability space and \(F = (F_t, t \geq 0)\) a given family of sub-\(\sigma\)-algebras of \(F\) satisfying the usual conditions (i.e., \(F\) is nondecreasing, right continuous and complete relative to \(P\)). A stochastic process \(X = (X_t, t \geq 0)\) is called \(F\)-adapted if for each \(t, X_t\) is \(F_t\)-measurable (written \(X_t \in F_t\)) which we denote by \(X = (X_t, F_t), t \geq 0\). An \(F\)-adapted process is also \(F\)-predictable if it is measurable with respect to the smallest \(\sigma\)-algebra over \(\Omega \times [0, \infty)\) generated by the left continuous, \(F\)-adapted processes.

The families \(M(F, P), M^2(F, P), M^{1}_{loc}(F, P)\) and \(M^{2}_{loc}(F, P)\) denote the classes of \(F\)-adapted processes which are uniformly integrable martingales, square integrable martingales and their local counterparts, respectively. Denote by \(b\) a deterministic Borel measure on \(\mathbb{R}^+ = [0, \infty)\) where we also use \(b(s) = b([0, s]), s \in \mathbb{R}^+\) to denote the corresponding right continuous nondecreasing function on \(\mathbb{R}^+\) and assume that \(b(0) = 0\).

Let \(E\) be a discrete space of \(m (m < \infty)\) distinct "events" and denote a generic event in \(E\) by \(i\). We define a multivariate counting process \(n\) over \([0, \infty) \times E\) as follows. Let \(T = (\tau_n), n \geq 1\) be a sequence of stopping times relative to \(F\) which satisfy the conditions:
(1.1) \( \tau_1 > 0 \quad \text{a.s.} \)
\( \tau_n < \tau_{n+1} \quad \text{a.s. on } \{ \tau_n < \infty \} \)
\( \tau_n = \tau_{n+1} \quad \text{a.s. on } \{ \tau_n = \infty \} \)

where a.s. means almost surely with respect to P. Also denote

\[ \tau_\infty = \lim_{n \to \infty} \tau_n \]

for the limit of the sequence \( T \).

Let \( X = (X_n) \), \( n \geq 1 \) be a sequence of \( E \)-valued random variables
and let \( (T,X) = (\tau_n, X_n) \), \( n \geq 1 \) be a multivariate point process
in the sense of Jacod [21]. For each \( \omega \in \Omega \) and Borel set \( A \) in
\( \sigma(\mathbb{R}^+ \times E) \) we define

(1.2) \[ \eta(\omega; A) = \sum_{n \geq 1} \int \varepsilon(\tau_n(\omega), X_n(\omega)) (dt, dx) 1(\tau_n(\omega) < \infty) \]

where \( \varepsilon(a) \) is the Dirac measure at the point \( a \) \[ \left( \int \varepsilon(a)(dt, dx) = \int_A 1(a \in A) \right) \]. Equation (1.2) defines a random measure over \( \mathbb{R}^+ \times E \) and
is called a multivariate counting process. By Jacod [21],
theorem 2.1, there exists a unique (up to stochastic equivalence)
predictable random measure \( \nu \) such that

(1.3) \( \nu([t] \times E) \leq 1 \)
\( \nu([\tau, \infty) \times E) = 0 \)
\( \nu \) for each subset \( B \) of \( E \)
\( a) \nu([0,t] \times B), t \geq 0 \) is \( F \)-predictable;
b) \( \eta([0,t] \times B) - \nu([0,t] \times B) \), \( t \geq 0 \) is a 
\( \tau_\infty \)-local martingale; i.e. for each \( n \)

\( \eta([0,t \tau_n] \times B) - \nu([0,t \tau_n] \times B) , t \geq 0 \)
is of the class \( M(F,P) \).

By virtue of the simple structure of the set \( E \) we can completely 
characterize the pair \( (\eta,\nu) \) by defining

\[
(1.4a) \quad N_t(i) := \eta([0,t] \times \{i\}), \forall t \geq 0 \text{ and } i \in E;
\]

and

\[
(1.4b) \quad A_t(i) := \nu([0,t] \times \{i\}), \forall t \geq 0 \text{ and } i \in E.
\]

For each \( t \) and \( i \), \( N_t(i) \) records the number of events of type \( i \) 
occurring over the time interval \([0,t]\) and \( A_t(i) \) may be viewed as 
the cumulative rate of occurrences of events of type \( i \). Thus for 
each \( i \) the counting process \( N(i) = (N_t(i), F_t), t \geq 0 \) is a uni-
variate counting process with \( F \)-predictable compensator function

\( A(i) = (A_t(i), F_t), t \geq 0 \).

We consider the case where the compensator function \( A \) has a 
particularly simple form defined as follows.

**Definition 1.1.** A multivariate counting process \( N \) is called 
a Poisson-type counting process if for each \( i \in E \) its compensator 
can be expressed:

\[
(1.5) \quad A_t(i) = \int_0^t Y_s(i)db_i(s), \forall t \geq 0
\]
where \( Y = (Y_t, F_t), t \geq 0 \) is a nonnegative F-predictable process and 

\( b_t \) is a deterministic Borel measure on \((\mathbb{R}^+, \sigma(\mathbb{R}^+))\). \( \square \)

When \( Y_s \equiv \lambda \) (a constant) and \( b(t) = t \) we obtain from (1.5) 

\( A_t = \lambda t \), so that \( N \) reduces to an ordinary Poisson process and thus 
in the general case \( N \) is called a Poisson-type point process by 

Lipster and Shiryayev [26]. Also when we assume \( b_t \) is absolutely 
continuous relative to Lebesgue measure (i.e. \( b_t = \int \alpha_t \, dt \)) 
(1.5) reduces to the multiplicative intensity model considered by 

Aalen [3].

Consider the following examples.

**Example 1.** Suppose \( \Omega \rightarrow \{1, 2, 3, \ldots\} \) is a random variable with 
the geometric distribution \( p_n = (1-p)^{n-1}p \) (\( p \epsilon (0, 1) \)) for \( n \geq 1 \). 
Define \( N = (N_t), t \geq 0 \) and \( F = (F_t), t \geq 0 \) by 

\[
(1.6a) \quad N_t = 1(X \leq t), \forall t \geq 0;
\]

and 

\[
(1.6b) \quad F_t = \sigma(N_s, s \leq t), \forall t \geq 0.
\]

The process \( N \) is called a simple counting process and it turns 
out that \( N \) has compensator \( A \) relative to \( F \) given by 

\[
(1.6c) \quad A_t = \int_0^t 1(X \geq s)p \, d\mu(s), \forall t \geq 0
\]

where \( \mu \) denotes the ordinary counting measure with support \( \{1, 2, 3, \ldots\} \).
and the process \( Y = (1(X \geq t)), \ t \geq 0 \) being left continuous is
\( F \)-predictable. \( \square \)

**Example 2.** Next consider a sequence \((X_n)\), \( n \geq 1 \) of independent
random variables each with the same distribution as \( X \) in example 1
and for each \( n = 1, 2, \ldots \) denote \( S_n = \sum_{k=1}^{n} X_k \). Then \( \{S_n\} \) is a
renewal process and we define \( N \) and \( F \) by

\[
(1.6d) \quad N_t = \sum_{n=1}^{\infty} 1(S_n \leq t), \ \forall t \geq 0;
\]

and

\[
(1.6e) \quad F_t = \sigma(N_s, s \leq t), \ \forall t \geq 0.
\]

The distribution of a Geometric random variable \( X \) is "memoryless"
meaning that for each \( s > 0 \)

\[
P(X > s+t | X \geq s) = P(X \geq t), \ \forall t \geq 0.
\]

Therefore by Jacod [21], proposition 3.1 it is easily shown that \( N \)
has compensator \( A \) given by

\[
(1.6f) \quad A_t = \int_{0}^{t} p \ d\mu(s) = \mu([0,t])p, \ \forall t \geq 0
\]

where \( \mu \) and \( p \) are defined in example 1. This result has intuitive
appeal in that for fixed \( t \), \( N_t \) is a Binomial random variable with
parameters \( \mu([0,t]) \) and \( p \), and the martingale property (see (1.3))
implies \( E N_t = E A_t = \mu([0,t])p. \) \( \square \)
Remark. We qualify example 2 by mentioning that in Chapter 5 we show that a general renewal process does not induce a Poisson-type counting process relative to the filtration $F$ defined as in $(1.6e)$. In fact example 2 works specifically by virtue of the memoryless property of the Geometric Law. In the general case we indicate an alternative approach which overcomes this limitation.

Let $N$ denote an $m$-variate Poisson-type counting process with discrete event space $E$ (here $m = \text{card}(E)$). In what follows denote
\[ \Delta_A_s = A_s - A_{s-} \quad (A_{s-} = \lim_{t \uparrow s} A_t), \]
and for $M \in \mathbb{M}_2^2$, $<M>$ denotes the quadratic characteristic of $M$ (i.e. $M^2 - <M> \in \mathbb{M}$). We have the following proposition.

**Proposition 1.1.** For each $i \in E$ the process $m(i) = N(i) - A(i)$ is a $\tau_\infty$-locally square integrable local martingale with quadratic characteristic
\[
(1.7) \quad <m(i)>_t = \int_0^t (1 - \Delta_A_s(i)) dA_s(i), \quad \forall t \geq 0;
\]
and for $i \neq i', i' \in E$
\[
(1.8) \quad <m(i), m(i')>_t = -\int_0^t \Delta_A_s(i) dA_s(i'), \quad \forall t \geq 0
\]
so that $m(i)m(i') - <m(i), m(i')>$ is a $\tau_\infty$-local martingale.

**Proof.** By Lipster and Shiryaev [26], corollary 18.12 $m(i) \in \mathbb{M}_\text{loc}^2(F,P)$ and (1.7) follows by [26], lemma 18.12 for $i \in E$. 


To prove (1.8) we fix $i$ and $i'$ in $E$ with $i = i'$ and for notational ease put $x_t = m_t(i)$ and $y_t = m_t(i')$, $\forall t \geq 0$. By Lipster and Shiryaev [26], Theorem 5.2 we have

$$
(1.9) \quad \langle x, y \rangle_t = \frac{1}{4} [\langle x+y \rangle_t - \langle x-y \rangle_t], \quad t \geq 0.
$$

It is a consequence of (1.2(3b)) and Lipster and Shiryaev [26], lemma 18.12, that the first term in the right-hand side of (1.9) is given by

$$
(1.10) \quad \langle x+y \rangle_t = \eta(\{i,i'\}) - \nu(\{i,i'\})_t
$$

$$
= \int_{(0,t]} (1 - \Delta A_s(i) - \Delta A_s(i'))(dA_s(i) + dA_s(i')),(0,t],
$$

$\forall t \geq 0$.

To determine the second term in the right-hand side of (1.9) we observe that for each $\omega \in \Omega$ $x-y$ is a right continuous function of $t$ of locally bounded variation. Therefore by virtue of equation (18.40), lemma 18.7 of Lipster and Shiryaev [26], we have

$$
(1.11) \quad (x_t - y_t)^2 = 2 \int_0^t (x_s - y_s) d(x_s - y_s) + \sum_{s < t} (\Delta(x_s - y_s))^2, \quad t \geq 0
$$

where (1.11) is to be interpreted pathwise (i.e. for each $\omega \in \Omega$) as a Stieltjes integral.
For each \( s \in \mathbb{R}^+ \) we have

\[
(1.12) \quad (\Delta(x_s - y_s))^2 = (\Delta N_s(i) - \Delta A_s(i) - (\Delta N_s(i') - \Delta A_s(i')))^2
\]

\[
= (\Delta N_s(i) - \Delta A_s(i))^2 + (\Delta N_s(i') - \Delta A_s(i'))^2
\]

\[
- 2(\Delta N_s(i) - \Delta A_s(i))(\Delta N_s(i') - \Delta A_s(i'))
\]

\[
= \Delta N_s(i) - 2\Delta N_s(i)\Delta A_s(i) + (\Delta A_s(i))^2
\]

\[
+ \Delta N_s(i') - 2\Delta N_s(i')\Delta A_s(i') + (\Delta A_s(i'))^2
\]

\[
- 2(\Delta N_s(i) - \Delta A_s(i))(\Delta N_s(i') - \Delta A_s(i')).
\]

By summing (1.12) over all \( s \leq t \) we obtain

\[
(1.13) \quad \sum_{s \leq t} (\Delta(x_s - y_s))^2 = \int_0^t (1 - 2\Delta A_s(i))dm_s(i) +
\]

\[
+ \int_0^t (1 - \Delta A_s(i))dA_s(i)
\]

\[
+ \int_0^t (1 - 2\Delta A_s(i'))dm_s(i')
\]

\[
+ \int_0^t (1 - \Delta A_s(i'))dA_s(i') - 2 \int_0^t \Delta A_s(i)dm_s(i')
\]

\[
+ 2 \int_0^t \Delta A_s(i')dm_s(i) + 2 \int_0^t \Delta A_s(i')dA_s(i)
\]

using equation (18.73) of [26] and the fact that \( \Delta N_s(i)\Delta N_s(i') = 0 \).
(P-a.s.) for \(i \neq i'\). Hence, on combining (1.13) with (1.11) we have

\[
\begin{align*}
(1.14) \quad (x_t - y_t)^2 &= 2 \int_0^t (x_s - y_s) d(x_s - y_s) \\
&+ \int_0^t (1 - 2A_s(i') - 2A_s(i)) dy_s \\
&+ \int_0^t (1 - 2A_s(i) + 2A_s(i')) dx_s \\
&+ \int_0^t (1 - A_s(i)) dA_s(i) \\
&+ \int_0^t (1 - A_s(i')) dA_s(i') + 2 \int_0^t A_s(i') dA_s(i) \\
&= \alpha_t + \beta_t + \gamma_t + \delta_t + \epsilon_t + \psi_t.
\end{align*}
\]

It is a straightforward application of Theorem 18.8 of [26] to show that \((\alpha_t + \beta_t + \gamma_t); t \geq 0\) is a \(\tau_\infty\)-local martingale. Also we have that \((\delta_t + \epsilon_t + \psi_t), t \geq 0\) is a nonnegative nondecreasing predictable process and therefore it is a natural increasing process (Dellacherie [16], theorem T27, Chapter 5). Now since \(((x_t - y_t)^2, F_t), t \geq 0\) is a \(\tau_\infty\)-local submartingale we have by the uniqueness of the Doob-Meyer decomposition that

\[
(1.15) \quad \langle m(i) - m(i') \rangle_t = \int_0^t (1 - A_s(i)) dA_s(i) \\
+ \int_0^t (1 - A_s(i')) dA_s(i') + 2 \int_0^t A_s(i') dA_s(i).
\]
Thus, upon subtracting (1.15) from (1.10) in the manner of (1.9) we obtain

\begin{equation}
(1.16) \, <m(i),m(i')>_t = \frac{1}{4} \left[<m(i)+m(i')>_t - <m(i)-m(i')>_t \right]
\end{equation}

\[= \frac{1}{4} \left[ - \int_0^t \Delta A_s(i')dA_s(i) - \int_0^t \Delta A_s(i')dA_s(i') ight. \]

\[+ 2 \int_0^t \Delta A_s(i')dA_s(i)] \]

\[= - \int_0^t \Delta A_s(i')dA_s(i), \forall t \geq 0 \text{ and } i,i' \in E, \quad i \neq i'. \]

This completes the proof. □

2.2 Nonparametric estimation for Poisson-type counting processes

We now address the problem of nonparametric estimation for the Poisson-type counting process defined in section 2.1. We assume that the process \((N,Y)\) is observable and that the Borel measure \(b\) in (1.5) is an unknown parameter. The statistical problem is to estimate \(b\) based on observation of the process \((N,Y)\) over a period of time.

For each \(n = 1, 2, \ldots\) suppose we observe \((N^i, Y^i)\), \(i = 1, \ldots, n\) where \(N^i(j)\) is a Poisson-type counting process with compensator \(A^i(j) = \int Y^i(j)db_j\), \(j = 1, \ldots, m\) and \(i = 1, \ldots, n\). Let \(F^n = (F^{(n)}_t), t \geq 0\)
where \( F_t^{(n)} = \sum_{i=1}^n F_i^t, \forall t \geq 0 \) and define \( N^{(n)} \) and \( Y^{(n)} \) by

\[
(1.1) \quad N^{(n)} = \sum_{i=1}^n N_i^{(n)},
\]
and

\[
(1.2) \quad Y^{(n)} = \sum_{i=1}^n Y_i^{(n)}.
\]

Denote \( A^{(n)} = \int Y^{(n)} db \) (componentwise), then it is easily shown that for each \( n = 1,2,... \) \( N^{(n)} - A^{(n)} \) is an \( m \)-dimensional \( F^n \)-local martingale (i.e. \( N^{(n)} - A^{(n)} \in M_{loc}(F^n, F^n), n = 1,2,... \)). Note that \( \Delta N^{(n)} > 1 \) occurs with positive probability at the mass points of \( b \) so that \( N^{(n)} \) is not a counting process and is more accurately called an integral-value random measure over \( (\mathbb{R}^+, \sigma \mathbb{R}^+) \).

Remark. The sequence outlined above may arise in the way described above. However, in Chapter 3 we work with a class of sequences of the form \( (N^{(n)}, Y^{(n)}, A^{(n)}) \), \( n \geq 1 \) of which the sequences of this section are a special case. \( \Box \)

To continue for each \( n = 1,2,... \) we define the \( F^n \)-predictable process \( X^n \) by

\[
(1.3) \quad X^n_t = (Y_t^{(n)})_{-1}^{1} (Y_t^{(n)} > 0), \forall t \geq 0
\]

where we use the convention \( \frac{0}{0} := 0 \). Now for each \( j = 1,..,m \) define the following Stieltjes integral (denoted by \( * \))
\begin{equation}
\hat{b}_t^n(j) = (X_t^n(j) * N_t^{(n)}(j))_t
= \int_0^t X_s^n(j) dN_s^{(n)}(j), \forall t \geq 0.
\end{equation}

Denote \( \hat{b}_t^n = (\hat{b}_t^n(1), \ldots, \hat{b}_t^n(m)) \) and consider the process \( \hat{b}_t^n = (\hat{b}_t^n, F_t^{(n)}) \), \( t \geq 0 \) (called the \( n \)th Aalen estimator) as an estimator of \( b \). Strictly speaking we can only estimate the function \( \tilde{b}_t^n \) defined by

\begin{equation}
\tilde{b}_t^n = \int \mathbb{1}(Y_t^{(n)} > 0) db
\end{equation}

which corresponds to the function \( b \) on \( \{ t : Y_t^{(n)} > 0 \} \).

In lemma 1.1 (below) we show that for each \( n = 1, 2, \ldots \) the error process \( \hat{b}_t^n - \tilde{b}_t^n \) is a local martingale where \( \hat{b}_t^n \) is defined by (1.4) and \( \tilde{b}_t^n \) by (1.5). We recall that for each \( n = 1, 2, \ldots \) \( N_t^{(n)} - A_t^{(n)} \) is a local martingale and therefore let \( (\tau_k^n)_k \), \( k \geq 1 \) denote a localizing sequence of \( F^n \)-stopping times for \( N_t^{(n)} - A_t^{(n)} \) (i.e. for each \( k \) the stopped process \( N_t^{(n)} - A_t^{(n)} \) is a martingale). Also let

\( \tau_\infty^n = \lim_{k \to \infty} \tau_k^n \) and consider the following lemma.

**Lemma 1.1.** Let \( \hat{b}_t^n \) and \( \tilde{b}_t^n \) be defined by (1.4) and (1.5), respectively. Then

i) If for each \( n = 1, 2, \ldots \)

\[ \int_0^t \mathbb{1}(Y_s^{(n)}(j) > 0) db_j(s) < \infty, \forall t < \tau_\infty^n, \forall j \leq m, p^n-a.s., \]

then \( \hat{b}_t^n - \tilde{b}_t^n \in M_{loc}(F^n, p^n) \).
ii) If for each \( n = 1, 2, \ldots \)

\[
\int_0^t (Y_s^{(n)}(j))^{-1}(Y_s^{(n)}(j) > 0) db_j(s) < \infty, \ \forall t < t_n, \ \forall j \leq m, \ P^n-a.s.,
\]

then \( \hat{b}^n - b^n \in H^n_{loc}(P^n, P^n) \) and for each \( t \in \mathbb{R}^+ \)

\[
<\hat{b}^n - b^n>_{ij} = \begin{cases} 
\int_0^t (Y_s^{(n)}(i))^{-1}(Y_s^{(n)}(i) > 0) db_i(s) \\
+ \sum_{s \leq t} (X_s^{n}(i))^2 \sum_{k=1}^n (Y_s^{k}(i))^2 (\Delta b_i(s))^2, \ i = j \\
- \sum_{k=1}^n \int_0^t X_s^{n}(i) X_s^{n}(j) Y_s^{k}(i) Y_s^{k}(j) \Delta b_i(s) db_j(s)
\end{cases}
\]

and

when \( i \neq j \).

\[\text{Proof.} \] Observe that since \( X^n \) is a measurable function of a predictable process \( Y^{(n)} \) it is predictable and that from equation (1.3)

\[ (1.7) \ P(X^n_t < \infty) = 1, \ \forall t > 0. \]

Next observe that
(1.8) \( \hat{b}^n - \tilde{b}^n = \int X^n(dN(n) - dA(n)) \)

\[= X^n \ast (N(n) - A(n)) \]

\[= \sum_{i=1}^{n} X^n \ast (N^i - A^i) \]

(here \( f \ast g \) denotes Stieltjes integration of \( f \) relative to \( g \)) and that for each \( j, 1 \leq j \leq m \), we have

(1.9) \( \int_{0}^{t} 1(Y(n) > 0)db_j(s) = \int_{0}^{t} X^j_s(j) \sum_{i=1}^{n} y^i_s(j)db_j(s) \)

\[= \sum_{i=1}^{n} \int_{0}^{t} X^j_s(j) y^i_s(j)db_j(s), \; \forall t \geq 0. \]

Recall that for each \( i, 1 \leq i \leq n \), \( N^i - A^i \in M_{loc}(F^n, P^n) \). Thus by (1.7), (i), (1.9) and Lipster and Shiryaev [26], theorem 18.7

\( X^n \ast (N^i - A^i) \) is a \( \tau^n_{\infty} \)-local martingale for \( i = 1, \ldots, n \) and therefore by (1.8) \( \hat{b}^n - \tilde{b}^n \in M_{loc}(F^n, P^n) \) for each \( n \geq 1. \)

Next recall that when \( A \) is a compensator for a counting process then

\( 0 \leq \Delta A_s \leq 1, \; \forall s \geq 0 \) (see section 2.1, equation (1.3)) so that the following inequality holds:

(2.0) \( \int_{0}^{t} (Y^j_s(n)(j))^{-1}1(Y^j_s(n)(j) > 0)db_j(s) \)

\[= \int_{0}^{t} (X^j_s(j))^2 dA_s(n)(j) \]
\[
\sum_{i=1}^{n} \int_{0}^{t} (X_s^n(j))^2 Y_s^i(j) dB_j(s)
\]

\[
> \sum_{i=1}^{n} \int_{0}^{t} (X_s^n(j))^2 (1 - Y_s^i(j) \Delta b_j(s)) Y_s^i(j) dB_j(s),
\]

\(\forall t > 0, \forall j \leq m.\)

By (ii) each term in the sum on the right hand side of (2.0) is almost surely finite for all \(t < \tau_n^\infty.\) Thus, by Lipster and Shiryayev [26], theorem 18.8 and arguing as earlier we obtain \(X^n \ast (N^i - A^i) \in M^2_{\text{loc}}(F^n, P^n)\) for \(i = 1, \ldots, n\) and hence by (1.8) \(b^n - b^n \in M^2_{\text{loc}}(F^n, P^n).\)

Finally, by the assumption of independence, the locally square integrable local martingales \([N^i - A^i], 1 \leq i \leq n\) are orthogonal and so using proposition 1.1, section 2.1 and Lipster and Shiryayev [26], theorem 18.8 we obtain (1.6). \(\square\)

### 2.3 Consistency of the estimator \(\hat{b}\)

We show that under general conditions the estimator \(\hat{b}^n\) is a consistent estimator in the sense that the error of estimation process \(\hat{b}^n - b^n\) converges to a zero process in some sense as \(n \uparrow \infty.\)

This result allows us to conclude that for \(n\) "large" \(\hat{b}^n\) is a close approximation to \(b\) on \(\{t: Y_t^n > 0\}.\)

For use in theorem 1.1 (below) we define two modes of uniform consistency of the estimator \(\hat{b}\).
Definition 1.1. For each $n = 1, 2, \ldots$ assume $\hat{b}^n - \tilde{b}^n \in M_{\text{loc}}(F^n, P^n)$ and let $(\tau^n_k)_k$, $k \geq 1$ be a localizing sequence of $F^n$-stopping times. For each $k = 1, 2, \ldots$ denote $\hat{b}^{nk} - \tilde{b}^{nk} = (\hat{b}^n_{t \wedge \tau^n_k} - \tilde{b}^n_{t \wedge \tau^n_k})$, $t \geq 0$ and if

$$\sup_{s \leq t} |\hat{b}^{nk}_s - \tilde{b}^{nk}_s| \xrightarrow{p} 0 \text{ as } n \uparrow \infty, \forall t \geq 0$$

then $(\hat{b}^n)$, $n \geq 1$ is called a local uniformly consistent estimator for $(\tilde{b}^n)$, $n \geq 1$. □

Definition 1.2. For each $n = 1, 2, \ldots$ assume $\hat{b}^n - \tilde{b}^n \in M_{\text{loc}}(F^n, P^n)$ and let $(\tau^n_k)_k$, $k \geq 1$ be a localizing sequence of $F^n$-stopping times with $\tau^n = \lim_{k \uparrow \infty} \tau^n_k$. We say $(\hat{b}^n)$, $n \geq 1$ is a $\tau^n$-uniformly consistent estimator for $(\tilde{b}^n)$, $n \geq 1$ when

$$\sup_{s \leq n} |\hat{b}^n_s - \tilde{b}^n_s| \xrightarrow{p} 0 \text{ as } n \uparrow \infty. \quad □$$

The relationship between these two modes of consistency is given in the following lemma.

Lemma 1.2. Suppose $(\hat{b}^n)$, $n \geq 1$ is a $\tau^n$-uniformly consistent estimator for $(\tilde{b}^n)$, $n \geq 1$, then $(\hat{b}^n)$, $n \geq 1$ is a local uniformly consistent estimator for $(\tilde{b}^n)$, $n \geq 1$.

Proof. To prove the lemma we trivially observe that for each $k \geq 1$

$$\sup_{t \geq 0} |\hat{b}^{nk}_t - \tilde{b}^{nk}_t| \leq \sup_{t \leq \tau^n_k} |\hat{b}^n_t - \tilde{b}^n_t|. \quad □$$
The idea behind these modes of consistency is that according to definitions 1.1 and 1.2 the estimator $\hat{b}_n$ need only be consistent locally over bounded intervals of the form $[0,t]$. We apply these definitions in the following theorem where all operations such as $|\cdot|$ on vector valued processes are to be interpreted componentwise.

**Theorem 1.1.** Suppose $(\hat{b}_n^n)$, $n \geq 1$ is a sequence of Aalen estimators for $(\tilde{b}_n^n)$, $n \geq 1$. Consider the following conditions.

- **[a1]** For each $t \in \mathbb{R}^+$ and $k = 1, 2, \ldots$

  \[
  E|\hat{b}_{nk}^n_t - \bar{b}_{nk}^n_t| \to 0 \text{ as } n \uparrow \infty;
  \]

- **[b1]** For each $n = 1, 2, \ldots$ $\tau_n^\infty < \infty$ almost surely and

  \[
  E|\hat{b}_n^n - b_n^n| \to 0 \text{ as } n \uparrow \infty;
  \]

- **[a2]** For each $k = 1, 2, \ldots$ and $t \in \mathbb{R}^+$

  \[
  \langle \hat{b}_{nk}^n_t - \bar{b}_{nk}^n_t \rangle \to 0 \text{ as } n \uparrow \infty \text{ for } i = 1, \ldots, m;
  \]

- **[b2]** For each $n = 1, 2, \ldots$ $\tau_n^\infty < \infty$ almost surely and

  \[
  \langle \hat{b}_n^n - b_n^n \rangle \to 0 \text{ as } n \uparrow \infty \text{ for } i = 1, \ldots, m.
  \]
a) When \( \hat{b}^n - \tilde{b}^n \in M_{\text{loc}}^1(F^n, P^n) \) for each \( n = 1, 2, \ldots \), then [\( \alpha_1 \)] and [\( \beta_1 \)] imply \( (\hat{b}^n)_n \), \( n \geq 1 \) is local uniformly consistent and \( \tau_{\infty} \)-uniformly consistent, respectively.

b) When \( \hat{b}^n - \tilde{b}^n \in M_{\text{loc}}^2(F^n, P^n) \) for each \( n = 1, 2, \ldots \), then [\( \alpha_2 \)] and [\( \beta_2 \)] imply \( (\hat{b}^n)_n \), \( n \geq 1 \) is local uniformly consistent and \( \tau_{\infty} \)-uniformly consistent, respectively.

Proof. Assume \( \hat{b}^n - \tilde{b}^n \in M_{\text{loc}}^1(F^n, P^n) \) for each \( n \) and that [\( \alpha_1 \)] holds. Observe that for each \( k = 1, 2, \ldots \) \( |\hat{b}^{nk} - \tilde{b}^{nk}| \) is a submartingale so that by Doob [17], theorem 3.2

\[
(2.4) \quad P(\sup_{s \leq t} |\hat{b}^{nk}_s - \tilde{b}^{nk}_s| > \lambda) \leq E|\hat{b}^{nk}_t - \tilde{b}^{nk}_t|
\]

for each real \( \lambda \) and \( t \geq 0 \). Since \( \lambda \) is arbitrary [\( \alpha_1 \)] implies \( \hat{b} \) is local uniformly consistent.

When for each \( n = 1, 2, \ldots \) \( \hat{b}^{nk} - \tilde{b}^{nk} \in M_{\text{loc}}^2(F^n, P^n) \) we recall that for each \( k = 1, 2, \ldots \) \( (\hat{b}^{nk} - \tilde{b}^{nk})^2 - <\hat{b}^{nk} - \tilde{b}^{nk}> \in M(F^n, P^n) \) and therefore for any finite \( F^n \)-stopping time \( \tau \)

\[
(2.5) \quad E(\hat{b}^{nk}_\tau(i) - \tilde{b}^{nk}_\tau(i))^2 = E<\hat{b}^{nk} - \tilde{b}^{nk}>_{\tau}, \quad i = 1, \ldots, m.
\]

Further, since \( (\hat{b}^{nk} - \tilde{b}^{nk})^2 \) and \( <\hat{b}^{nk} - \tilde{b}^{nk}> \) are nonnegative and \( <\hat{b}^{nk} - \tilde{b}^{nk}> \) is also nondecreasing and \( F^n \)-predictable it follows by the Lenglart inequality (see Lenglart [25]) that for all real \( a > 0 \) and \( b > 0 \)
\[(2.6) \quad \mathbb{P}(\sup_{s < t} |\hat{b}_{nk}(i) - \tilde{b}_{nk}(i)| > a^{1/2}) \leq \frac{1}{\alpha} E(<b_{nk} - \tilde{b}_{nk}, i_i \wedge t > b) \]

\[+ \mathbb{P}(<b_{nk} - \tilde{b}_{nk}, i_i > t > b)\]

for all $t \in \mathbb{R}^+$. 

Assume \([\alpha 2]\) holds so that $b$ may be taken arbitrarily small in (2.6) and therefore it follows that for any $a$ the l.h.s. of (2.6) may be made arbitrarily small as $n$ increases and so $\hat{b}$ is local uniformly consistent.

Next assume \([\beta 2]\) holds so that for each $n \tau_{\infty}^n < \infty$ almost surely and for $\hat{b}^n - \tilde{b}^n \in \mathcal{M}_{loc}^2 (\mathcal{F}^n, \mathcal{F}^n)$ we may replace $t$ by $\tau_{\infty}^n$ in (2.6) (see for example [27]) to obtain

\[(2.7) \quad \mathbb{P}(\sup_{s < \tau_{\infty}^n} |\hat{b}_{nk}(i) - \tilde{b}_{nk}(i)| > a^{1/2}) \leq \frac{1}{\alpha} E(<b_{nk} - \tilde{b}_{nk}, i_i \wedge \tau_{\infty}^n > b) \]

\[+ \mathbb{P}(<b_{nk} - \tilde{b}_{nk}, i_i > \tau_{\infty}^n > b)\]

for $k = 1, 2, \ldots$ and $i = 1, \ldots, m$. Now since $<\hat{b}^n - \tilde{b}^n>$ is non-decreasing and $\tau_{k}^n \uparrow \tau_{\infty}^n$ as $k \uparrow \infty$ we have

\[<\hat{b}_{nk} - \tilde{b}_{nk} > \uparrow <\hat{b}^n - \tilde{b}^n>_{\tau_{\infty}^n}\]

and

\[\sup_{s < \tau_{\infty}^n} |\hat{b}_{nk}(i) - \tilde{b}_{nk}(i)| = \sup_{s < \tau_{\infty}^n} |\hat{b}_{nk} - \tilde{b}_{nk}| \uparrow \sup_{s < \tau_{\infty}^n} |\hat{b}_n - \tilde{b}_n|\]

as $k \uparrow \infty$. Thus by an application of Fatou's lemma and the monotone
convergence theorem (2.7) implies
\begin{align*}
(2.8) \quad & \mathbb{P}\left( \operatorname{sup}_{s \leq \tau_n^{-}} |\widehat{b}_s^n(i) - \tilde{b}_s^n(i)| > a^{1/2} \right) \\
& < \lim_{\tau_n \to \infty} \inf_{k} \mathbb{P}\left( \operatorname{sup}_{s \leq \tau_n^{-}} |\widehat{b}_s^{nk}(i) - \tilde{b}_s^{nk}(i)| > a^{1/2} \right) \\
& < \lim_{\tau_n \to \infty} \inf_{k} \left\{ \frac{1}{a} \mathbb{E}(\langle \hat{b}^{nk} - \tilde{b}^{nk}, i \rangle_{\tau_n^{-}}) \\
& + \mathbb{P}(\langle \hat{b}^{nk} - \tilde{b}^{nk}, i \rangle_{\tau_n^{-}} > b) \right\} \\
& = \frac{1}{a} \mathbb{E}(\langle \hat{b}^{n} - \tilde{b}^{n}, i \rangle_{\tau_n^{-}}) \\
& + \mathbb{P}(\langle \hat{b}^{n} - \tilde{b}^{n}, i \rangle_{\tau_n^{-}} > b)
\end{align*}
for \( i = 1, \ldots, m \). Therefore [82] implies \( \hat{b} \) is \( \tau_{\infty} \)-uniformly consistent.

The case corresponding to [81] is proved in a manner similar to that of [82] by using \( \tau_n^{-} < \infty \) almost surely (according to [81]) and upon replacing \( t \) in (2.4) by a sequence \( (t_k) \), \( k \geq 1 \) such that \( t_k \uparrow \infty \) as \( k \uparrow \infty \). \( \Box \)

2.4 Asymptotic normality

We prove that when suitably normalized the sequence of estimators \( \hat{b}^n \), \( n \geq 1 \) defined in section 2.2 converges in law to a Gaussian process of independent increments. This result is an
application of a more general results proved in Chapter 3 where we have used the functional convergence theorems for semimartingales developed by Jacod, Klopotowski and Mémin [22] and Lipster and Shiryaev [27].

One of the distinguishing features of Poisson-type counting processes is that the measure \( b \) defined in section 2.1 is an arbitrary Borel measure. Thus the right-continuous function \( b(\cdot) := b([0,\cdot]) \) may have discontinuities corresponding to the mass points of the measure \( b \). Denote by \( J \) the set \( \{ t: t \in \mathbb{R}^+, b_i(t) > 0 \text{ for some } i = 1,2,\ldots,m \} \) and for convenience identify \( J \) with the countable set \( \{ t_j, j \geq 1 \} \) where \( 0 < t_1 < \ldots < t_j < \ldots \). Make the following assumption.

A1. Let \( \{t_j\} \) denote the set of discontinuities of the function \( b \) and assume the sequence \( t_1, t_2, \ldots \) has no finite points of accumulation. □

In theorem 1.2 (below) the limit process \( X \) is a zero-mean Gaussian process of independent increments which has specific features which are described as follows. Let \( C = (C^1, \ldots, C^m) \) be a continuous function defined on \( \mathbb{R}^+ \) and suppose that \( \xi = (\xi_t), t \geq 0 \) is a continuous \( m \)-variate Gaussian process such that

\[
(2.9) \quad \langle \xi \rangle_{t}^{i,j} = \begin{cases} 
C_{t}^{i} & i = j; \\
0 & i \neq j, \forall t \geq 0, 1 \leq i, j \leq m.
\end{cases}
\]

Note that equation (2.9) implies that the components of \( \xi \) are mutually
independent. For each \( j = 1,2,\ldots \) let \( U_j \) by an \( m \)-variate Gaussian random variable with mean zero and covariance matrix \((\sigma_{jk}^i)\), \( 1 \leq i, k \leq m \) and denote by \( \phi_j(\cdot) \) the distribution measure of \( U_j \) on \((\mathbb{R}^m, \sigma(\mathbb{R}^m))\) (i.e. for \( A \) a Borel set in \( \mathbb{R}^m \), \( P(U_j \in A) = \phi_j(A) \)). A measure \( \nu \) on \((\mathbb{R}_+ \times \mathbb{R}^m, \sigma(\mathbb{R}_+ \times \mathbb{R}^m))\) is defined as follows

\[
(3.0) \quad \nu([0,t] \times B) = \sum_{j: \tau_j < t} \phi_j(B),
\]

for \( t \in \mathbb{R}_+ \), \( B \in \sigma(\mathbb{R}^m) \) and \( \{\tau_j\} \) defined in A1. We define a Gaussian process \( X \) as follows.

**A2.** Assume \( \xi \) and \( \{U_j, j \geq 1\} \) are independent and define a Gaussian process \( X \) by

\[
(3.1) \quad X_t = \xi_t + \sum_{j: \tau_j < t} U_j, \quad \forall t \geq 0.
\]

Then \( X = (X_t), t \geq 0 \) has independent increments and triplet of local characteristics (see Appendix I) \((B,C,\nu)\) where \( B \equiv 0 \), \( C \) is given by \((2.9)\) and \( \nu \) by \((3.0)\). □

In the following theorem \( C_0 \) denotes the space of continuous functions \( g \) such that \( g: \mathbb{R}^m \rightarrow \mathbb{R} \), \( g \) has limits at infinity and is equal to zero in a neighborhood of zero.

**Theorem 1.2.** Let \( X \) satisfy A2 and let \( J \) (equivalently \( \{\tau_j\}, j \geq 1 \) satisfy A1 where \( b \) is a Borel measure defined in section 2.1. Let \( (\widehat{b_n}^m), n \geq 1 \) be a sequence of Aalen estimators for \( b \)
defined by (1.4) and let \( \tilde{b}^n \), \( n \geq 1 \) be defined by (1.5). Also let \( b^c = b - \Sigma \Delta b \) denote the continuous part of \( b \) and consider the following conditions.

\[ [\alpha] \quad \forall t \geq 0, \forall j \leq m, \forall \varepsilon \in (0,1] \]
\[ n^{1/2} \int_0^t 1((Y_s^{(n)}(j))^{-1}1(Y_s^{(n)}(j) > 0) > \varepsilon)1(Y_s^{(n)}(j) > 0)db_j^c(s) \]
\[ p \to 0; \]

\[ [\beta] \quad \forall t \geq 0, \forall j \leq m \]
\[ n \cdot \sum_{s \leq t, s \in \mathbb{R}_+^+/\{t_j\}} (Y_s^{(n)}(j))^{-2}1(Y_s^{(n)}(j) > 0)\Delta N_s(n) \to c_t^j, \]

\[ [\alpha]^2 \quad \forall t \geq 0, \forall j \leq m, \forall \varepsilon \in (0,1] \]
\[ n \cdot \int_0^t (Y_s^{(n)}(j))^{-1}1(Y_s^{(n)}(j) > 0)1((Y_s^{(n)}(j))^{-1}1(Y_s^{(n)}(j) > 0) > \varepsilon)db_j^c(s) \]
\[ p \to 0; \]

\[ [\beta]^2 \quad \forall t \geq 0, \forall j \leq m \]
\[ n \cdot \int_0^t (Y_s^{(n)}(j))^{-1}1(Y_s^{(n)}(j) > 0)db_j^c(s) \to c_t^j. \]

Also define \( W_j^n = n^{1/2}(Y_{t_j}^{(n)})^{-1}1(Y_{t_j}^{(n)} > 0)(\Delta N_{t_j}^{(n)} - Y_{t_j}^{(n)} \Delta b(t_j)) \)

and \( F_{j-1}^n = F_{t_j}^n \) for all \( j, n \geq 1 \). For \( x \in \mathbb{R}^m \) define the function \( f_x \).
by $f^\lambda_k(x) = x^\lambda_1(x^\lambda_2 \leq 1)$ where $x^\lambda_1$ is the $\lambda^{th}$ coordinate of $x$ and consider the following conditions:

$$[\gamma] \quad \forall t \geq 0, \forall i, j \leq m$$

$$\sum_{\lambda : t^\lambda_2 \leq t} E(f^\lambda_i(W^n_{\lambda_2}) f^\lambda_j(W^n_{\lambda_2}) | F^n_{\lambda_2-1}) - \sum_{\lambda : t^\lambda_2 \leq t} [E(f^\lambda_i(W^n_{\lambda_2}) | F^n_{\lambda_2-1})]$$

$$\cdot [E(f^\lambda_j(W^n_{\lambda_2}) | F^n_{\lambda_2-1})]$$

$$\mathbb{P} \rightarrow \sum_{\lambda : t^\lambda_2 \leq t} \int f^\lambda_i(s) f^\lambda_j(x) \phi_{\lambda_2}(dx) - \sum_{\lambda : t^\lambda_2 \leq t} [\int f^\lambda_i(x) \phi_{\lambda_2}(dx)]$$

$$\cdot [\int f^\lambda_j(x) \phi_{\lambda_2}(dx)];$$

$$[\delta] \quad \forall t \geq 0, \forall g \in C_0, \mathbb{P} = 1, 2,$$

$$\sum_{\lambda : t^\lambda_2 \leq t} [E(g(W^n_{\lambda_2}) | F^n_{\lambda_2-1})]^\mathbb{P} \rightarrow \sum_{\lambda : t^\lambda_2 \leq t} [\int \phi_{\lambda_2}(dx) g(x)]^\mathbb{P}$$

$$\sup \lambda \quad \forall t \geq 0, \forall j \leq m$$

$$\sup_{s \leq t} \left| \sum_{\lambda : t^\lambda_2 \leq s} E(f^\lambda_j(W^n_{\lambda_2}) | F^n_{\lambda_2-1}) \right|^\mathbb{P} \rightarrow 0.$$

a) If $\hat{b}^n - \tilde{b}^n \in M_{\text{loc}}(F^n, P^n)$ for each $n$, then $[\alpha]$, $[\beta]$, $[\gamma]$, $[\delta]$ and $[\sup \lambda]$ imply that $n^{-1/2}(\hat{b}^n - \tilde{b}^n)$ converges in law to $\hat{X}$ in the space $\mathbb{D}(\mathbb{R}^+, \mathbb{R}^m)$ endowed with the Skorokhod topology as $n \uparrow \infty$. 

b) If $\hat{b}^n - b^n \in M_{1,0}^2(F^n, P^n)$ for each $n$, then $[\alpha]^2$, $[\beta]^2$, $[\gamma]$, $[\delta]$ and $[\sup \lambda]$ imply that $n^{1/2} (\hat{b}^n - b^n)$ converges in law to $X$ in $D(\mathbb{R}^+, \mathbb{R}^m)$ endowed with the Skorokhod topology as $n \to \infty$.

Proof. To prove the theorem we need only observe that with $X^n = n^{1/2} Y^n, (Y^n > 0)$ and $N^{(n)} - A^{(n)} = \sum_{i=1}^{n} (N^i - \int Y^i db)$ then $n^{1/2} (\hat{b}^n - b^n) = \int X^n (dN^n - dA^n) \in M_{1,0}^2(F^n, P^n)$ is of class LS (see definition 1.2, section 3.3.1). Thus by Theorem 1.2, Chapter 3 $n^{1/2} (\hat{b}^n - b^n)$ converges in law to $X$ under (a) and (b) as $n$ tends toward infinity. □

2.5 Empirical estimation of the covariance function

In theorem 1.2 we proved that the normalized error of estimation process $n^{1/2} (\hat{b}^n - b^n)$ converges in law to a Gaussian process $X$ of independent increments where $X$ satisfies $A_2$. Under $A_2$ it is easy to show that the covariance function $\sigma = (\sigma_{ij}^t)$, $t \geq 0$ of $X$ is given by

$$
(3.3) \quad \sigma_{ij}^t = E X_i^t X_j^t = \begin{cases} 
C_i^t + \sum_{\ell: t_{\ell} \leq t} \sigma_{ii}^\ell, & i = j; \\
\sum_{\ell: t_{\ell} \leq t} \sigma_{ij}^\ell, & i \neq j,
\end{cases}
$$

for all $t \geq 0$, $1 \leq i, j \leq m$.

We consider an empirical estimation procedure for the covariance function $\sigma$ of equation (3.3) and the following assumption is made.
A3. Assume \( b^n - \tilde{b}^n \in M^2_{\text{loc}}(R^n, R^n) \) for each \( n \), that \([\beta]^2\) of Theorem 1.2 holds and that for \( \langle \hat{b}^n - \tilde{b}^n \rangle \) defined by (1.6), section 2.2

\[
\sum_{j=1}^{p} \sigma_{rj}^q \rightarrow \sigma_{rj}^q \text{ for } j \geq 1, \ l \leq r, \ q \leq m
\]

where \( \sigma_j = (\sigma_{rj}^q), \ l \leq r, \ q \leq m, \ j \geq 1 \) is defined in section 2.4. \( \square \)

Our approach to this estimation problem is to define two processes \( \hat{\sigma}^{nc} \) and \( \hat{\sigma}^{nd} \) as follows. For each \( t \in R^+ \) and \( 1 \leq i, j \leq m \) consider

\[
(3.4a) \quad \hat{\sigma}_{t}^{nc}(i, j) = \begin{cases} 
\sum_{s \leq t} (X^n_s(i))^2 \Delta N_s^{(n)}(i) \mathbb{1}(\Delta N_s^{(n)}(i) \leq 1), \ i = j; \\
0, \ i \neq j 
\end{cases}
\]

and

\[
(3.4b) \quad \hat{\sigma}_{t}^{nd}(i, j) = \begin{cases} 
\sum_{s \leq t} (X^n_s(i))^2 \Delta N_s^{(n)}(i) \mathbb{1}(\Delta N_s^{(n)} > 1) \\
- \sum_{\ell=1}^{n} \sum_{s \leq t} (Y^n_s(i))^2 (Y^n_s(i)) \Delta N_s^{(n)}(i) \\
\quad \times \mathbb{1}(\Delta N_s^{(n)} > 1), \ i = j; \\
-n[ \sum_{s \leq t} (X^n_s(i))^2 Y_s^\ell(i) (X^n_s(j))^2 \\
\quad \times Y_s^\ell(j) \Delta N_s^{(n)}(i) \Delta N_s^{(n)}(j) \mathbb{1}(\Delta N_s^{(n)}(i) > 1), \\
\quad \Delta N_s^{(n)}(i) > 1)], \ i \neq j. 
\end{cases}
\]
For each $n = 1, 2, \ldots$ the process formed of the sum $\hat{\sigma}^n = \hat{\sigma}_{nc} + \hat{\sigma}_{nd}$ is suggested as an empirical estimator of the covariance function $\sigma$ and below we give conditions under which $\hat{\sigma}^n$ converges in a suitable sense to $\sigma$ as $n \uparrow \infty$.

We define two processes $\hat{\sigma}_{nc}$ and $\hat{\sigma}_{nd}$ as follows

\begin{equation}
\hat{\sigma}_{nc}(i, j) = \begin{cases}
\frac{n}{t_0} \int_0^t (X^n_s(i)) db^c_1(s), & i = j; \\
0, & i \neq j
\end{cases}
\end{equation}

and

\begin{equation}
\hat{\sigma}_{nd}(i, j) = \begin{cases}
\frac{n}{t} \sum_{k: t_k < t_k < t} X^n_t(i) \Delta b_i(t_k) \\
\quad \quad - \sum_{k: t_k < t} (X^n_t(k)) \sum_{l=1}^n (Y^n_t(i)) (Y^n_t(j)) (\Delta b_i(t_k))^2, & i = j; \\
\quad \quad -n \sum_{k: t_k < t} X^n_t(i) X^n_t(j) (\sum_{l=1}^n Y^n_t(i)) (Y^n_t(j)) \\
\quad \quad \quad \times \Delta b_i(t_k) \Delta b_j(t_k), & i \neq j;
\end{cases}
\end{equation}

for all $t \geq 0$, $1 \leq i, j \leq m$ where $\{t_k\}$ is defined in A1.

The following definition is used in Theorem 1.3 (below).
Definition 1.3. For each $n = 1, 2, \ldots$ let $T^n_k = (\tau^n_k)$, $k \geq 1$ be a monotone sequence of $F^n$-stopping times with $\tau^n_\infty = \lim_{k \to \infty} \tau^n_k$ and let $\hat{\sigma}^n$ and $\tilde{\sigma}^n$ be $F^n$-adapted processes. If for each $k = 1, 2, \ldots$ and $t \in \mathbb{R}^+$

\begin{equation}
(3.6a) \frac{\hat{\sigma}^n - \tilde{\sigma}^n}{t^n_k - t^n_{k-1}} \to 0 \text{ as } n \to \infty,
\end{equation}

then $(\hat{\sigma}^n)$, $n \geq 1$ is called \textit{locally consistent} for $(\tilde{\sigma}^n)$, $n \geq 1$; and if for each $t \in \mathbb{R}^+$

\begin{equation}
(3.6b) \frac{\hat{\sigma}^n - \tilde{\sigma}^n}{t^n_\infty - t^n_{\infty-1}} \to 0 \text{ as } n \to \infty,
\end{equation}

then $(\hat{\sigma}^n)$, $n \geq 1$ is called $\tau_\infty$-\textit{locally consistent} for $(\tilde{\sigma}^n)$, $n \geq 1$. □

Theorem 1.3. Assume A1, section 2.4 and A3 hold and let $T^n_k = (\tau^n_k)$, $k \geq 1$ be a sequence of localizing stopping times for $\hat{b}^n - \tilde{b}^n$. Consider the following conditions.

[a] $(\hat{b}^n)$, $n \geq 1$ is locally consistent for $(\tilde{b}^n)$, $n \geq 1$;
[b] $(\tilde{\sigma}^n_{nc})$, $n \geq 1$ is locally consistent for $(\tilde{\sigma}^n_{nc})$, $n \geq 1$;
[b]' $(\hat{\sigma}^n_{nc})$, $n \geq 1$ is $\tau_\infty$-locally consistent for $(\hat{\sigma}^n_{nc})$, $n \geq 1$.

a) If [a] and [b] hold, then $(\hat{\sigma}^n)$, $n \geq 1$ is locally consistent for $\sigma$.

b) If [a]', [b]' and also

$[\delta] \quad (\hat{b}^n - \tilde{b}^n)_{t^n_\infty - t^n_{\infty-1}}$, $t \geq 0$ is a square integrable martingale,

then $(\hat{\sigma}^n)$, $n \geq 1$ is $\tau_\infty$-locally consistent for $\sigma$. 
Proof. First observe that under A3 $[\beta]^2$ of Theorem 1.2 implies that $(\tilde{\sigma}_{\text{nc}}^n)$, $n \geq 1$ converges in probability for all $t \geq 0$ to a function $C = (C_{ij}^t)$, $1 \leq i, j \leq m$ where $C_{ii}^t = C_i^t$ of (3.3) and $C_{ii'}^t = 0$, $i \neq i'$. Since for each $k = 1, 2, \ldots, t \geq 0$ and arbitrary $\varepsilon > 0$

\[(3.7) \quad P(|\tilde{\sigma}_{\text{nc}}^{t,n} - C_{t^{n}}^{t,k} - C_{t^{n}}^t| > \varepsilon)\]

\[= P(|\tilde{\sigma}_{\text{nc}}^t - C_t| > \varepsilon, \tau_k^n > t) + P(|\tilde{\sigma}_{\text{nc}}^{t,n} - C_{t^{n}}^{t,k} - C_{t^{n}}^t| > \varepsilon, \tau_k^n \leq t)\]

\[\leq P(|\tilde{\sigma}_{\text{nc}}^t - C_t| > \varepsilon) + P(\sup_{s \leq t} |\tilde{\sigma}_{\text{nc}}^n - C_s| > \varepsilon)\]

it follows (by A3 and lemma 1 of [28b] together with continuity of C) that (3.7) tends to zero as $n \to \infty$ which implies that $(\tilde{\sigma}_{\text{nc}}^n)$, $n \geq 1$ is locally consistent for C and a similar argument shows $\tau_\infty$-local consistency as well. Hence, it follows from the above that if $[\beta]$ and $[\beta]'$ hold then $(\tilde{\sigma}_{\text{nc}}^n)$, $n \geq 1$ is locally consistent, respectively, $\tau_\infty$-locally consistent for C.

It remains to show that $(\tilde{\sigma}_{\text{nd}}^n)$, $n \geq 1$ is consistent for $\sigma$-C. We note that on the set $\mathbb{R}^+/\{t_k\}$ where $b$ is continuous $\Delta N_{s}^{(n)} < 1$ with probability one and so we may replace $\sum_{s \leq t}$ on the right side of (3.4b) with $\sum_{k: t_k \leq t} \Delta N_{t_k}^{(n)}$ almost surely. Also under A1 $\{k: t_k \leq t\}$ is a finite set for each $t$ and so we need only show that $\Delta \tilde{\sigma}_{\text{nd}}^{t,n} \overset{p}{\to} \sigma_k$ for each $k$ (where convergence occurs in the appropriate modal sense).
Let \( \{F^n_0, w^n_k, F^n_k, k \geq 1\} \) be defined as in Theorem 1.2 and note that for each \( n = 1, 2, \ldots \)

\[
M^n_{\mathcal{F}} = \sum_{k: t_k \leq t} w^n_k, \ t \geq 0
\]

is a locally square integrable local martingale (see the proof of Theorem 1.2, Chapter 3). Therefore, by lemma 1.1 (equation (1.6)) and (3.5b) we have that \( \mathcal{G}^{nd} \), \( n \geq 1 \) is the unique predictable process such that

\[
M^n(i)M^n(j) - \mathcal{G}^{nd}(i, j) \in M_{loc}(F^n, p^n), \ n = 1, 2, \ldots; 1 \leq i, j \leq m.
\]

Observe that for each \( n = 1, 2, \ldots, t \in \mathbb{R}^+ \) and \( 1 \leq i, j \leq m 

\[
M^n_{\mathcal{F}}(i)M^n_{\mathcal{F}}(j) = \sum_{k: t_k \leq t} w^n_k(i)w^n_k(j) + \sum_{\ell \neq k} \sum_{t_{\ell}, t_k \leq t} w^n_{\ell}(i)w^n_k(j)
\]

and recall that \( \{w^n_k\} \) forms a local martingale difference array where for \( \ell \neq k \) \( w^n_{\ell} \) and \( w^n_k \) are increments over distinct time points. Thus using the fact that a martingale has orthogonal increments over disjoint time intervals it is easily shown that the second term on the r.h.s. of (3.9) is a local martingale whereby it follows that

\[
\sum_{k: t_k \leq t} w^n_k(i)w^n_k(j) - \mathcal{G}^{nd}(i, j), \ t \geq 0
\]

is a local martingale for each \( n = 1, 2, \ldots \) and \( 1 \leq i, j \leq m \).

Next observe that under A3

\[
E(w^n_k(i)w^n_k(j) | F^n_{k-1}), \ t \geq 0, 1 \leq i, j \leq m
\]
is both locally consistent and $\tau_\infty$-locally consistent for $\sigma$-c. To see this recall that under A3 [\gamma] of Theorem 1.2 implies that (4.1) converges in probability to $\sigma$-c, \forall t \geq 0. Also, under Al \{k : t_k \leq t\} is a fixed finite set for each $t$ and therefore it is easily argued that (4.1) converges in probability uniformly over bounded intervals of the form $[0,t]$. Thus, an argument analogous to that used with (3.7) verifies the statement above.

For each $r \geq 1$ (4.0) implies that

\[(4.2) \sum_{t : t_k < t} \nabla E(W_k(i)w_k^n(j) | F_k^{n-1}) = \tilde{\sigma}_{\text{nd}}^{n}(i,j), \]

\[t \geq 0, 1 \leq i, j \leq m\]

and if assumption [6] holds we may replace $\tau^n_r$ in (1.2) with $\tau^n_\infty$. Thus the conclusion of the previous paragraph implies that $(\tilde{\sigma}_{\text{nd}}, n \geq 1$ is locally (respectively $\tau_\infty$-locally) consistent for $\sigma$-c.

It remains to show that $(\tilde{\sigma}_{\text{nd}}, n \geq 1$ is correspondingly consistent for $\sigma$-c. Let us assume without loss of generality that the r.h.s. of (3.4b) has been modified according to the remarks of paragraph two of this proof. By [a] ([a']) $(\tilde{b}^n), n \geq 1$ is locally consistent ($\tau_\infty$-locally consistent) for $(\tilde{b}^n), n \geq 1$ and therefore for each $k = 1, 2, \ldots$

\[(4.3) \frac{\Delta \tilde{b}^n_k}{\xi_k} - \frac{\Delta \tilde{b}^n_k}{\xi_k} = \frac{x^n_k}{\xi_k} \Delta N_k^{(n)}(n) - 1(Y_k^{(n)} > 0) \Delta b(t_k)\]

converges in probability to zero as $n \uparrow \infty$ where convergence holds in accordance with definition 1.3. Hence, by reordering terms of the
products in the summations on the r.h.s. of (3.4b), comparing (3.4b) (with \( \sum_{s \leq t} \) replaced by \( \sum_{k: t_k \leq t} \)) to (3.5b), using A1 (in particular the fact that \( \{k: t_k \leq t\} \) is finite \( \forall t \geq 0 \)), statement (4.3) and the consistency of \( \hat{\sigma}^{nd} \), \( n \geq 1 \) it follows easily that \( \hat{\sigma}^{nd} \), \( n \geq 1 \) is locally consistent (respectively \( \tau_\infty \)-locally consistent) for \( \sigma \)-c.

Upon combining this result with the previous result for \( \hat{\sigma}^{nc} \), \( n \geq 1 \) it follows that \( \hat{\sigma}^n = \hat{\sigma}^{nc} + \hat{\sigma}^{nd} \) is correspondingly consistent for \( \sigma \). \( \square \)

Remark. In many applications it is possible to verify [\( \beta \)] and [\( \beta' \)] along with corresponding statements for \( \hat{\sigma}^{nd} \) by showing that \( \hat{\sigma}^n - \hat{\sigma}^n \) is a local martingale. This may typically lead to a stronger result of uniform local consistency. \( \square \)
CHAPTER 3

Functional Convergence Theorems

3.1 Preliminaries

We consider the convergence in law (or weak convergence) of a local martingale $M^n$ to a limit process $M^\infty$ of independent increments. For each $n = 1, 2, \ldots$ we assume $M^n$ belongs to a class of local martingales which may arise in applications of Poisson-type counting processes where the normalized error of estimation process $n^{1/2}(b^n - b^n)$ defined in section 2.2.2 provides an example. To define this class we extend the notion of Poisson-type counting process to a class of integral-valued random counting measures.

Denote $N = (N_t), t \geq 0$ an integral-valued random measure on $(\mathbb{R}^+, \mathcal{B}(\mathbb{R}^+))$ with $\Delta N_s : \Omega \mapsto \{0, 1, 2, \ldots\}$ for each $s \in \mathbb{R}^+$ and let $A$ denote a predictable random measure on $(\mathbb{R}^+, \mathcal{B}(\mathbb{R}^+))$ such that the process

\begin{equation}
(1.1) \quad m_t = N_t - A_t, \quad t \geq 0
\end{equation}

is a local martingale (here $A_t := A([0,t])$). We assume that the process $A = (A_t), t \geq 0$ is almost surely continuous as a function $t$ except perhaps at a countable but fixed set of discontinuities $\{t_j : t_j \in \mathbb{R}^+, j \geq 1\}$ such that $A(t_j) = \Delta A_{t_j} > 0$ with positive
probability. Clearly, examples of the pair \((N, A)\) are given by the Poisson-type counting process defined in section 2.2.1 as well as sums of \(n\) such counting processes as encountered in section 2.2.2. Consider the following definition.

**Definition 1.1.** Let \(N\) be an integral-valued random counting measure and \(A\) its predictable compensator such that the function \(A = A([0, \cdot])\) is almost surely continuous except at a countable but fixed set of discontinuities \(\{t_j, j \geq 1\}\). The local martingale \(m\) formed of the difference \(N - A\) is called a **class CP** local martingale. □

We next define the class of local martingales considered in the functional limit theorems of section 3.2.

**Definition 1.2.** Let \(X = (X_t), t \geq 0\) be a predictable process and \(m\) be a class CP local martingale and for each \(\omega \in \Omega\) let \(X \circ m\) denote the ordinary Stieltjes integral of \(X\) relative to \(m\). If \(M = X \circ m\) and \(M\) is a local martingale then \(M\) is called a **class LS** local martingale. □

When \(m \in \text{CP}\) it is the difference between two monotone real-valued processes \(N\) and \(A\) so that in general \(m\) is a process of local bounded variation (see Royden [33, p. 100]). Thus the Stieltjes integral \(X \circ m\) is in general well defined and we see in the next
lemma that the class LS inherits bounded variation from the class CP.

**Lemma 1.1.** If $M$ is of class LS, then $M$ is a process of local bounded variation.

**Proof.** Recall that $M$ is of the form $X \circ m$ where $m \in \text{CP}$. If necessary we may consider the positive part $X^+$ and the negative part $X^-$ of $X$. For each $\omega \in \Omega$ it is easy to verify that $M^+_t(\omega) = X^+ \circ m$ and $M^-_t(\omega) = X^- \circ m$ are each the difference between two monotone real-valued functions so that $M^+_t$ and $M^-_t$ are processes of local bounded variation. Now observe that

$$M_t(\omega) = M^+_t(\omega) - M^-_t(\omega)$$

$$= \int_0^t X^+_s(dN_s - dA_s)(\omega) - \int_0^t X^-_s(dN_s - dA_s)(\omega)$$

$$= [\int_0^t X^+_s dN_s + \int_0^t X^-_s dA_s](\omega)$$

$$- [\int_0^t X^-_s dN_s + \int_0^t X^+_s dA_s](\omega), \quad t < \tau_\omega(\omega)$$

is again the difference between two monotone real-valued functions and so $M$ is a process of local bounded variation as well. □
Recall that an arbitrary local martingale $M$ may be decomposed into the sum of a continuous local margingale $M^c$ plus a purely discontinuous local martingale $M^d$; see Appendix I. Let $\mu$ denote a random counting measure on $\mathbb{R}^+ \times \mathbb{R}/\{0\}$ generated by the jumps of $M$ (i.e. $\mu([0,t] \times A) = \sum_{s<t} 1(\Delta M_s \in A), A \in \sigma(\mathbb{R}/\{0\})$) and let $\nu$ denote its predictable compensator so that $\mu-\nu$ is a random martingale measure on $\mathbb{R}^+ \times \mathbb{R}/\{0\}$. The local martingale $M^d$ has the representation:

\[
M^d_t = \int_0^t \int_{\mathbb{R}} x(\mu-\nu)(ds, dx), \quad t \geq 0.
\]

The next lemma shows that if $M \in \text{LS}$, then $M^c \equiv 0$ in the fundamental decomposition of $M$.

**Lemma 1.2.** Let $M$ be a local martingale of class LS, then $M$ has representation $M^d$ where $M^d$ is defined as in (1.2).

**Proof.** Recall that any local martingale admits a representation $M = M^c + M^d$ and that the quadratic variation process $[M]$ for $M$ is given by:

\[
[M]_t = <M^c>_t + \sum_{s<t} (\Delta M_s)^2, \quad t \geq 0,
\]

where $<M^c>$ is the quadratic characteristic process for $M^c$. To prove the lemma it suffices to show that $<M^c> \equiv 0$ since this implies
that \( M^n \equiv 0 \). Recall that the following holds:

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} \left( M_{t\left(\frac{k+1}{n}\right)} - M_{t\frac{k}{n}} \right)^2 \overset{p}{\to} [M]_t, \quad \forall t > 0
\]

(see for example [27] or [20]). We fix \( t \) and for each \( n = 1,2,\ldots \) define \( Y_n \) by

\[
(1.5) \quad Y_n = \frac{1}{n} \sum_{k=0}^{n-1} \left( M_{t\left(\frac{k+1}{n}\right)} - M_{t\frac{k}{n}} \right)^2.
\]

Since \( Y_n \overset{p}{\to} [M]_t \), there exists a subsequence \( \{n_k\} \), \( k \geq 1 \) such that \( Y_{n_k} \overset{p}{\to} [M]_t \) P-a.s. as \( k \to \infty \) and therefore to prove the lemma it suffices to show that \( Y_{n_k} \to \sum_{s \leq t} (\Delta M_s)^2 \) P-a.s.

Let \( M_{j+1,k} = M_{t\left(\frac{j+1}{n_k}\right)} \) and \( M_{j,k} = M_{t\frac{j}{n_k}} \) for each \( t, k \) and \( j = 0,1,\ldots,n_k-1 \) and observe that

\[
(1.6) \quad (M_{j+1,k} - M_{j,k})^2 = M^2_{j+1,k} + M^2_{j,k} - 2M_{j+1,k}M_{j,k}
\]

\[
= M^2_{j+1,k} + M^2_{j,k} - 2[M_{j,k}(M_{j+1,k} - M_{j,k})]
\]

\[
+ M^2_{j,k}
\]

\[
= M^2_{j+1,k} - M^2_{j,k} - 2M_{j,k}\Delta M_{j,k}
\]

where \( \Delta M_{j,k} = M_{j+1,k} - M_{j,k} \). Therefore, by summing (1.6) over \( j \)
we obtain

\[ Y_n^{k} = \sum_{j=0}^{n_k-1} \left( M_{j+1}^{k} - M_j^{k} \right)^2 \]

By lemma 1.1 for each \( \omega \in \Omega \) \( M \) is a function of local bounded variation and since \( M \) is right continuous we have (by Lipster and Shiryaev [26], lemma 18.7 and the formula for integration by parts for Stieltjes integrals) that

\[ M_t^2 = 2 \int_0^t M_s^- dM_s + \sum_{s < t} (\Delta M_s)^2 \]

and so (1.7) becomes

\[ Y_n^{k} = 2 \int_0^t M_s^- dM_s + \sum_{s < t} (\Delta M_s)^2 - 2 \sum_{j=0}^{n_k-1} M_{j}^{k} \Delta M_j^{k} \]

Next observe that for each \( \omega \) outside a P-null set

\[ \lim_{k \to \infty} \sum_{j=0}^{n_k-1} M_{j}^{k} \Delta M_j^{k} = \lim_{k \to \infty} \sum_{j=0}^{n_k-1} M_{j}^{k} \left( \frac{M_{j+1}^{k} - M_j^{k}}{t_{n_k}^{j}} \right) \frac{j}{n_k} \]

satisfies the definition of a Stieltjes integral of the left continuous integrand \( (M_s^-)_s \), \( s \geq 0 \) with respect to \( M \). Hence the right hand side of (1.9) converges to \( \sum_{s < t} (\Delta M_s)^2 \) P-a.s. as \( k \to \infty \) which implies \( \langle M^c \rangle \equiv 0 \) P-a.s. \( \square \)
In the theorems below the limit process is a semimartingale of independent increments. These processes are reviewed in Appendix I so we proceed directly to the following definition.

**Definition 1.3.** Let \( \xi = (\xi_t), t \geq 0 \) be a continuous Gaussian process such that \( \langle \xi \rangle = C \) where \( C \) is a continuous function defined on \( \mathbb{R}^+ \), let \( \mu \) be a random counting measure on \( \mathbb{R}^+ \times \mathbb{R} \) with predictable compensator \( \nu \) and let \( B \) be a function of locally bounded variation on \( \mathbb{R}^+ \). Define the process \( X = (X_t), t \geq 0 \) by

\[
(2.1) \quad X_t = X_0 + \int_0^t \int_{|x|>1} x\mu(ds, dx) \\
+ \int_0^t \int_{|x|\leq1} x(\mu-\nu)(ds, dx), \ t \geq 0.
\]

The process \( X \) is a semimartingale and if the triplet \((B, C, \nu)\) is deterministic then \( X \) is a process of independent increments; see Jacod et al., [22]. □

### 3.2 Functional convergence theorems: processes of the class LS

As promised we formulate functional convergence theorems for sequences of class LS local martingales defined in section 3.1. Several applications of these theorems occur in Chapters 2 and 4 of this thesis and we take as our starting point a theorem due to Jacod, Klopotowski and Mémin [22] involving
functional convergence of sequence of semimartingales to an arbitrary process of independent increments.

The theorems below cover the k-dimensional case so that in general a process $X$ is a process in $\mathbb{R}^k$ for $k$ an integer. The space $\mathcal{D}(\mathbb{R}^+,\mathbb{R}^k)$ denotes the space of right continuous functions with left hand limits from $\mathbb{R}^+$ into $\mathbb{R}^k$ and $C^0$ denotes the space of continuous functions $g$ such that $g: \mathbb{R}^k \to \mathbb{R}$, $g$ has limits at infinity and $g$ is equal to zero in a neighborhood of zero.

**Theorem 1.1.** [22] Let $D$ denote a dense subset of $\mathbb{R}^+$ and suppose $X$ is a semimartingale of independent increments satisfying definition 1.3. For each $n = 1, 2, \ldots$ $(\Omega^n, F^n, P^n)$ is a probability space, $F^n = (F^n_t, t > 0$ is a filtration satisfying the usual conditions and $X^n$ is an $F^n$-semimartingale with triplet of local characteristics $(B^n, C^n, \nu^n)$ relative to $F^n$. Consider the following conditions.

\[ [\beta] \forall t \in D, \ E^n_B^{-1}_t \to B_t; \]

\[ [\sup \beta] \forall t > 0, \sup_{s \leq t} |E^n_{B_s} - B_s| \to 0; \]

\[ [\gamma] \forall t \in D, \forall i,j \leq k \]

\[
\begin{align*}
C^{n,ij}_t + \int_{1\leq |x^i|} \nu^n([0,t] \times dx)x^i x^j - \sum_{1\leq |x^i|} \int_{1\leq |x^i|} \nu^n([s] \times dx)x^j \\
\int_{1\leq |x^j|} \nu^n([s] \times dx)x^j 
\end{align*}
\]
\[ \begin{align*}
\mathbb{P} \left( C_t \right) + \int_{|x| \leq 1} \nu([0,t] \times dx)x_i x_j - \sum_{s \leq t} \left[ \int_{|x| \leq 1} \nu([s] \times dx)x_i \right] \\
\left[ \int_{|x| \leq 1} \nu([s] \times dx)x_j \right];
\end{align*} \]

\[ \mathbb{P} \left( \nu^{\pi}(0, t) \times dx \nu^{\pi}(x) \right) \]

\[ \mathbb{P} \left( \nu^{\pi}(0, t) \times dx \nu^{\pi}(x) \right) \]

\[ \sum_{s \leq t} \left[ \int_{\mathbb{R}^k} \nu([s] \times dx)g(x) \right]^2 \]

\[ \sum_{s \leq t} \left[ \int_{\mathbb{R}^k} \nu([s] \times dx)g(x) \right]^2. \]

\[ L_f (D) \]

a) Conditions [\( \beta \)], [\( \gamma \)], [\( \delta \)], [\( \delta \)] imply \( X^n \rightarrow X \) (convergence of finite dimensional distributions).

b) Conditions [\( \sup \beta \)], [\( \gamma \)], [\( \delta \)], [\( \delta \)] imply \( X^n \) converges in law to \( X \) in the space \( D(\mathbb{R}^+; \mathbb{R}^k) \) endowed with the Skorokhod topology (see Billingsley [10]).

Remark. Theorem 1.1 is proved in [22] as theorem 3.1 and is a quite general theorem of functional convergence to a process with independent increments since it includes as a special case many classical results of this type. We use Theorem 1.1 to develop a theory of weak convergence for sequences of local martingales of class LS. The conditions [\( \beta \)], [\( \gamma \)], [\( \delta \)], [\( \delta \)] define in precise terms the sense in which the local characteristics \( (\mathcal{B}^n, \mathcal{C}^n, \mathcal{V}^n) \) of \( X^n \) converge in probability to those of \( X; (B, C, \nu) \). Condition [\( \delta \)] and [\( \delta \)] is recognized as being equivalent to convergence in probability.
of the measure $\nu^n$ to $\nu$ with $[\delta I]$ guaranteeing this at the $t$-discontinuities of $\nu$. These conditions ensure that the sequence of characteristic functions $\phi^n$ of $X^n$ converges to that of $X$, implying convergence of the finite dimensional distributions of $X^n$ to those of $X$. In b) the added condition $[\sup \beta]$ in combination with the others ensures that the sequence of probability measures $\{p^n\}$ induced on $\mathcal{D}(\mathbb{R}^+ \times \mathbb{R}^k)$ by $\{X^n\}$ are relatively compact from which it follows that $X^n$ converges in law to $X$. \(\square\)

A process $X$ is of class CP or LS in $k$-dimensions if and only if $X$ is $k$-dimensional and each component of $X$ is of class CP or LS. In definition 1.1 we introduced a pair $(N,A)$ where $N$ is an integral-valued counting measure and $A$ its associated compensator. In the theorem below we assume that $N = (N(1), \ldots, N(k))$ and $A = (A(1), \ldots, A(k))$ are each $k$-dimensional processes where each component satisfies definition 1.1. Recall that associated with $A$ is a set $\{t_j, j \geq 1\}$ of fixed points of discontinuity where now we assume that $t_j$ is a point of discontinuity if $A(i)(t_j) > 0$ with positive probability for some $i = 1, \ldots, k$. Where convenient we write $A^c$ to denote the continuous part of the compensator function $A$ (i.e. $A^c = A - \sum_j \Delta A_{t_j}$) and in theorem 1.2 the limit process is a $k$-dimensional Gaussian process which satisfies the following hypothesis.
H1. Let $X$ be a $k$-dimensional process which satisfies the conditions of definition 1.3 with triplet of local characteristics $(B, C, \nu)$. Assume that the covariance function $(C_{ij}^t)$, $1 \leq i, j \leq k$ satisfies $C_{ij}^t = 0$, $i \neq j$ and $C_{ii}^t$ is a continuous function of $t$ and that for each $t$ $B_t$ and $\nu([0,t] \times \{\cdot\})$ have the following form:

$$B_t = \sum_{j: t_j < t} \delta_j \quad \text{(here } \delta_j \in \mathbb{R}^k \text{ for each } j)$$

and

$$\nu([0,t] \times \{\cdot\}) = \sum_{j: t_j < t} \nu([t_j] \times \{\cdot\}).$$

Thus $B_t$ and $\nu([0,\cdot], \cdot)$ are discrete functions of $t$ with support in $\{t_j, j \geq 1\}$ and for each $j = 1, 2, \ldots$, $\nu([t_j] \times \cdot)$ is a measure on $\mathbb{R}^k$. □

Theorem 1.2. Let $\{t_j, j \geq 1\}$ be a countable set in $\mathbb{R}^+ / \{0\}$ such that $0 < t_1 < \ldots < t_j < \ldots$ and suppose $X$ satisfies H1. For each $n = 1, 2, \ldots$ let $m^n$ be of class CP and $M^n$ be of class LS relative to $m^n$ (i.e. $M^n = X^n \circ m^n$ for some predictable $X^n$) defined on a filtered probability space $(\mathcal{F}^n, F_t^n, t \geq 0, P^n)$. If $m^n = N^n - A^n$ then assume $A^n$ has fixed points of discontinuity in $\{t_j\}$ independent of $n$. Consider the following conditions.

[\alpha] \forall t \in D, \forall j \leq k, \forall \varepsilon \in [0,1]

$$\int_0^t |X^n(j)|1(|X^n(j)| > \varepsilon) dA^n_{s}(j) \overset{\text{p}}{\rightarrow} 0;$$
\[ \forall t \in D, \forall j \leq k \]
\[ \sum_{s \leq t} (X^n_s(j))^2 \Delta N^n_s(j) \Rightarrow c_j(t); \]
\[ s \in \mathbb{R}^+\{/J\} \]

\[ \forall t \in D, \forall j \leq k, \forall \varepsilon \in (0,1) \]
\[ \int_0^t (x^n_s(j))^2 1(|X^n_s(j)| > \varepsilon)dA^n_{s,j}(j) \Rightarrow 0; \]

\[ \forall t \in D, \forall j \leq k \]
\[ \int_0^t (x^n_s(j))^2 dA^n_{s,j}(j) \Rightarrow c_j(t). \]

Let \( f = (f_1, \ldots, f_k) \) be a function such that \( f_N : \mathbb{R}^k \rightarrow \mathbb{R} \) and
\[ f_N(x) = x_1 1(|x_2| < 1), \quad \ell = 1, \ldots, k. \] For each \( n = 1, 2, \ldots \) and
\[ j = 1, 2, \ldots \] define \( W^n_j = X^n_{t_j} (\Delta N^n_{t_j} - \Delta A^n_{t_j}) \) and \( F^n_{j-1} = F^n_{t_{j-1}}. \) Consider the following conditions:

\[ \forall t \in D, \forall i,j \leq k \]
\[ \sum_{\ell \leq t \leq \ell^*} E(f_i f_j(W^n_\ell)|F^n_{\ell^*}) - \sum_{\ell \leq t \leq \ell} \left[ E(f_i(W^n_\ell)|F^n_{\ell^*}) \right] \left[ E(f_j(W^n_\ell)|F^n_{\ell^*}) \right] \]
\[ \Rightarrow \sum_{\ell \leq t \leq \ell^*} \int_{\mathbb{R}} \nu(t_\ell \times dx) f_i f_j(x) - \int_{\mathbb{R}} \nu(t_\ell \times dx) f_i(x) \]
\[ \times \int_{\mathbb{R}} \nu(t_\ell \times dx) f_j(x) \]
[δ] \( \forall t \in D, \forall g \in C_0, p = 1,2; \)

\[
\sum_{\lambda: t_\lambda \leq t} [E(g(W^n_\lambda | F^n_{\lambda-1})]^p \rightarrow \sum_{\lambda: t_\lambda \leq t} [\int_k \nu((\{t_\lambda\} \times dt)g(x)]^p
\]

[\sup \lambda] \( \forall t \in D, \)

\[
\sup_{s \leq t} \max_{1 \leq j < k} \left| \sum_{\lambda: t_\lambda \leq s} E(f_j(W^n_\lambda | F^n_{\lambda-1}) - B_j(s) \right|^p \rightarrow 0.
\]

a) If \( M^n \in M^1_{loc}(F^n, P^n) \) for each \( n \), then \([\alpha], [\beta], [\gamma], [\delta] \) and

[\sup \lambda] imply \( M^n \) converges in law to \( X \) in the space \( D(\mathbb{R}^+, \mathbb{R}^k) \)

endowed with the Skorokhod topology as \( n \to \infty. \)

b) If \( M^n \in M^2_{loc}(F^n, P^n) \) for each \( n \), then \([\alpha]^2, [\beta]^2, [\gamma], [\delta] \) and

[\sup \lambda] imply \( M^n \) converges in law to \( X \) in \( D(\mathbb{R}^+, \mathbb{R}^k) \)

endowed with the Skorokhod topology as \( n \to \infty. \)

**Proof.** Identify \( J \) with the set \( \{t_j, j \geq 1\} \) and let \( \chi_j \) denote the characteristic function of the set \( J \). Define the function \( f \) on \( \mathbb{R}^+ \) by

\[
f(s) = 1 - \chi_j(s), \quad s \in \mathbb{R}^+
\]

and observe that \( f \) is \( F^n \)-predictable (since it is deterministic) and bounded. Recall that by lemma 1.1 \( M^n \) is a process of local bounded
variation and therefore for each \( n = 1, 2, \ldots \) the Stieltjes integral
\( f \circ M^n \) agrees with the stochastic integral \( f \cdot M^n \) (see Shiryayev
[34, p. 209]). Thus \( f \circ M^n \) is a local martingale so that

\[
(1.1a) \quad M^{nc}_t = (f \circ M^n)_t, \quad t \geq 0
\]

and

\[
(1.1b) \quad M^{nd}_t = ((1-f) \circ M^n)_t = M^n_t - M^{nc}_t, \quad t \geq 0
\]

each define \( F^n \)-adapted local martingales.

Consider the process \( M^{nc} \) and for each \( n = 1, 2, \ldots \) and \( \varepsilon > 0 \)
define the process \( \alpha^{\varepsilon}[M^{nc}]^i \) and \( \sigma^{\varepsilon}[M^{nc}]^i \) for \( i = 1, \ldots, k \) by

\[
(1.2a) \quad \alpha^{\varepsilon}[M^{nc}]^i_t = \sum_{s \leq t} |\Delta M^{nc}_s(i)| 1(|\Delta M^{nc}_s(i)| > \varepsilon), \quad t \geq 0
\]

and

\[
(1.2b) \quad \sigma^{\varepsilon}[M^{nc}]^i_t = \sum_{s \leq t} (\Delta M^{nc}_s(i))^2 1(|\Delta M^{nc}_s(i)| > \varepsilon), \quad t \geq 0.
\]

When \( M^n \in M^1_{loc}(F^n, P^n) \) (\( M^n \in M^2_{loc}(F^n, P^n) \)) the process \( \alpha^{\varepsilon} (\sigma^{\varepsilon}) \) is
locally integrable and we can define its predictable compensator
\( \tilde{\alpha}^{\varepsilon} (\tilde{\sigma}^{\varepsilon}) \) (see Rebolledo [32]). Therefore observe that (1.2) may be
equivalently expressed as

\[
(1.3a) \quad \alpha^{\varepsilon}[M^{nc}]^i_t = \sum_{s \leq t} |X^n_s(i)| 1(|X^n_s(i)| > \varepsilon) \Delta N^{nc}_s
\]

\[
= \int_0^t |X^n_s(i)| 1(|X^n_s(i)| > \varepsilon) dN^{nc}_s, \quad t \geq 0
\]

and

\[
(1.3b) \quad \sigma^{\varepsilon}[M^{nc}]^i_t = \int_0^t (X^n_s(i))^2 1(|X^n_s(i)| > \varepsilon) dN^{nc}_s, \quad t \geq 0
\]
where \( N^{nc} = f \circ N^n \) for each \( n = 1, 2, \ldots \). Since the integrands in (1.3a and b) are predictable processes and \( N^{nc} \) has predictable compensator \( A^{nc} \) it follows immediately that for each \( n = 1, 2, \ldots \) and \( i = 1, \ldots, k \)

\[
(1.4a) \quad \tilde{\alpha}^{c}[M^{nc}]^i_t = \int_0^t |X^n_s(i)| \mathbb{1}(|X^n_s(i)| > \varepsilon) dA^{nc}_s(i), \quad t \geq 0
\]

and

\[
(1.4b) \quad \tilde{\sigma}^{c}[M^{nc}]^i_t = \int_0^t (X^n_s(i))^2 \mathbb{1}(|X^n_s(i)| > \varepsilon) dA^{nc}_s(i), \quad t \geq 0
\]

where (1.4b) requires \( M^n \) to be locally square integrable.

Recall that for each \( n \) \( N^{nc} \) is a nondecreasing random counting measure with continuous compensator \( A^{nc} \) and therefore it may be shown that when \( i \neq j \) \( \Delta N^{nc}(i)\Delta N^{nc}(j) = 0 \) P-a.s., \( 1 \leq i, j \leq k \). Thus for \( i \neq j \) and \( n = 1, 2, \ldots \) lemma 1.2 implies (see Helland [20], equation 4.12)

\[
(1.5) \quad [M^{nc}(i), M^{nc}(j)]_t = \sum_{s \leq t} \Delta M^{nc}_s(i) \Delta M^{nc}_s(j)
\]

\[
= \sum_{s \leq t} X^n_s(i) X^n_s(j) \Delta N^{nc}_s(i) \Delta N^{nc}_s(j)
\]

\[
= 0 \quad \text{almost surely,} \quad \forall t \geq 0.
\]

It turns out that by lemma 1.1 and Lipster and Shiryayev [26], lemma 18.7 and the formula for integration by parts that when \( M^n \in M^2_{loc}(P^n, P^n) \) for each \( n = 1, 2, \ldots \) and \( i \neq j \) then
(1.6) \( <M^{nc}(i), M^{nc}(j)>_t = 0 \) almost surely, \( t \geq 0, 1 \leq i, j \leq m \).

When \( M^n \in M^1_{loc}(F^n, P^n) \) then Rebolledo [32], theorem 2, Helland [20], theorem 5.4, (1.4a, (1.5), [\(\alpha\)] and [\(\beta\)] imply that \( M^{nc} \) converges in law to \( \xi \) in \( \mathcal{D}(\mathbb{R}^+, \mathbb{R}^k) \) as \( n \uparrow \infty \). Alternatively, when \( M^n \in M^2_{loc}(F^n, P^n) \) then Lipster and Shiryaev [27], corollary 2, Helland [20], theorem 5.4, (1.4b), (1.6), \( [\alpha]^2 \) and \( [\beta]^2 \) imply that \( M^{nc} \) converges in law to \( \xi \) in \( \mathcal{D}(\mathbb{R}^+, \mathbb{R}^k) \) as \( n \uparrow \infty \). In both cases \( \xi \) is a continuous Gaussian process of independent increments with covariance function defined in H1.

The remainder of the proof consists of showing that the preceding result plus conditions \([\gamma], [\delta]\) and \([\sup \lambda]\) imply the conditions of theorem 1.1(b). For each \( n = 1, 2, \ldots \) let \( \mu^{nc} \) and \( \mu^{nd} \) denote the random counting measures on \( (\mathbb{R}^+ \times \mathbb{R}^k, \sigma(\mathbb{R}^+ \times \mathbb{R}^k)) \) generated by the jumps of \( M^{nc} \) and \( M^{nd} \), respectively, and let \( \nu^{nc} \) and \( \nu^{nd} \) denote the \( F^n \)-predictable compensators of these measures. For \( x \in \mathbb{R}^k \) let \( x^i \) denote the \( i \)th coordinate of \( x \) and recall by lemma 1.2 that for \( i = 1, \ldots, k \)

\[
(1.9a) \quad M^{nc}_t(i) = \int_0^t \int_{\mathbb{R}^k} x^i (\mu^{nc} - \nu^{nc}) \{ds, dx\}, \quad t \geq 0
\]

and

\[
(1.9b) \quad M^{nd}_t(i) = \int_0^t \int_{\mathbb{R}^k} x^i (\mu^{nd} - \nu^{nd}) \{ds, dx\}, \quad t \geq 0.
\]

Recall that \( \xi \) is a continuous process so that the pair \((\mu^\xi, \nu^\xi)\) associated with the jumps of \( \xi \) are identically zero. Since \( M^{nc} \)
converges weakly to \( \xi \) it is easily shown that \( \nu^{nc} \overset{p}{\to} 0 \) and that if
\( \nu^n \) in Theorem 1.1 is replaced with \( \nu^{nc} \) then the r.h.s of condition [\( \delta \)] and [\( \delta_1 \)] of Theorem 1.1 is zero. Further the continuity of \( A^{nc} \) implies the quasi-continuity of \( M^{nc} \) so that

\[
(2.0) \quad \nu^{nc}(\{s\} \times \mathbb{R}^k/{(0)}) = 0 \quad \text{a.s. for all } s \in \mathbb{R}^+.
\]

By construction for each \( n = 1, 2, \ldots \) \( \nu^n = \nu^{nc} + \nu^{nd} \) and so the remarks above together with [\( \delta \)] for \( p = 1 \) and \( p = 2 \) imply [\( \delta \)] and [\( \delta_1 \)] of Theorem 1.1, respectively.

To verify condition [\( \gamma \)] of Theorem 1.1 we observe that [\( \alpha \)] and (1.4a), and [\( \alpha \)]\(^2\) and (1.4b), respectively imply (by the uniqueness of the predictable compensator and (1.9)) that for all \( t \in \mathbb{R}^+ \), \( \varepsilon \in (0, 1] \) and \( i = 1, \ldots, k \)

\[
(2.1a) \quad \int_0^{t} \int_{|x^i|>\varepsilon} |x^i| \nu^{nc}\{ds, dx\} \overset{p}{\rightarrow} 0 \quad \text{as } n \uparrow \infty;
\]

and

\[
(2.1b) \quad \int_0^{t} \int_{|x^i|>\varepsilon} (x^i)^2 \nu^{nc}\{ds, dx\} \overset{p}{\rightarrow} 0 \quad \text{as } n \uparrow \infty.
\]

Next let \( \{\hat{W}_j^n, j \geq 1\} \) be defined as in the theorem and observe that by construction

\[
(2.2) \quad M^{nd}_t = \sum_{j: t_j < t} \hat{W}_j^n.
\]

For each \( j = 1, 2, \ldots \) it is easy to show that the conditional law of \( \hat{W}_j^n \) given \( \tau^n_{t_j}^- \) is given by \( \nu^{nd}(\{t_j\} \times \cdot) \) and that for \( h: \mathbb{R}^k \rightarrow \mathbb{R} \);
a bounded measurable function, equation (1.9b) implies that

$$
(2.3) \ E(h(x) | F^n_{t_j^-} ) = E(\int_{\mathbb{R}^k} h(x)u^{nd}(\{t_j\},dx) | F^n_{t_j^-} )
= \int_{\mathbb{R}^k} h(x)v^{nd}(\{t_j\},dx) \quad (\text{a.s.})
$$

where the last equality follows by the predictability of $v^{nd}$. Thus

$[\alpha], [\beta], (2.3), (2.0), (2.1a), [\gamma], (1.5)$ and lemma 1.2 imply $[\gamma]

of Theorem 1.1. Alternatively, when $M^n$ is locally square integrable
then $[\alpha]^2, [\beta]^2, (2.3), (2.0), (2.1b), [\gamma], (1.6)$ and lemma 1.2 imply

$[\gamma]$ of Theorem 1.1.

It remains to verify condition $[\sup \beta]$ of Theorem 1.1. For each

$n = 1, 2, \ldots$. $M^n \in M_{loc}(F^n, P^n)$ so it follows that for each $i = 1, \ldots, k

$$
(2.4) \quad E^n_t(i) = \int_0^t \int_{\mathbb{R}} x^i \nu^n(ds,dx) = \int_0^t \int_{|x^i|>1} x^i (\nu^{nc} + \nu^{nd})(ds,dx)
= E^{nc}_t(i) + E^{nd}_t(i), \quad t \geq 0
$$

in obvious notation. By either condition $[\alpha]$ or $[\alpha]^2$ we have for

all $t \in D$

$$
(2.5) \sup_{s \leq t} \left| E^{nc}_s(i) \right| \leq \int_0^t \int_{|x^i|>1} x^i \nu^{nc}(ds,dx) \overset{p}{\rightarrow} 0, \quad \text{for } i = 1, \ldots, k.
$$

Further, since $\int_{\mathbb{R}} x^i \nu^n(\{t\},dx) = 0 \quad \forall t \geq 0$ we have by (2.3)
\[ B^n_{t_j}(i) = \sum_{j: t_j \leq t} \int_{\{x^i| > 1\}} x^i \nu^n(\{t_j\}, dx) \]
\[ = \sum_{j: t_j \leq t} \int_{\{x^i| < 1\}} x^i \nu^n(\{t_j\}, dx) \]
\[ = \sum_{j: t_j \leq t} E(f_i(w_j^n | F^n_{t_j}), \forall t \geq 0) \]

where \( f_i \) is defined in the statement of the theorem for \( i = 1, \ldots, k \).
Hence, by condition \([\text{sup } \lambda]\) and (2.5) condition \([\text{sup } \beta]\) of Theorem 1.1 is verified. This completes our proof. □

We now prove a second functional limit theorem, which is applicable to the case when for each \( n \) \( M^n \) is a locally square integrable local martingale. The conditions of the theorem appear different from those of Theorem 1.2 and are in many situations easier to verify.

**Theorem 1.3.** Let \( J \) be a fixed countable set of discrete points in \( \mathbb{R}^+ / \{0\} \) and \( D \) a dense subset in \( \mathbb{R}^+ \). Suppose \( X \) satisfies \( H_1 \) with \( \{t_j\} \) in \( H_1 \) contained in \( J \). For each \( n = 1, 2, \ldots, m^n = (m^n_{t_j}, F^n_{t_j}) \), \( t \geq 0 \) is of class \( CP \) with \( M^n = (M^n_{t_j}, F^n_{t_j}) \), \( t \geq 0 \) of class \( LS \) relative to \( m^n \).
Assume \( A^n \) (associated to \( m^n \)) has fixed points of discontinuity in \( J \) independent of \( n \). Consider the following conditions:

\[ [\alpha]^2 \quad \forall t \in D, \forall j \leq k, \forall \varepsilon \in (0, 1] \]
\[ \int_0^t (X^n_{s}(j))^2 1(|X^n_{s}(j)| > \varepsilon) dA^{nc}_{s}(j) \xrightarrow{p} 0; \]
\[ [\beta]^2 \quad \forall t \in D, \forall j \leq k \]

\[
\int_0^t (x_s^n(j))^2 dA_{s^n}(j) \overset{P}{\rightarrow} C_j(t);
\]

\[
 [\gamma] \quad \text{For each } t \in J \text{ and } u \in \mathbb{R}^k
\]

\[
E(e^{iu \cdot \tilde{W}^n_j}) \overset{P}{\rightarrow} E e^{iu \cdot U_j} \text{ as } n \uparrow \infty
\]

for each \( j = 1, 2, \ldots \) where \( \tilde{W}^n_j = x^n_{t_j} (\Delta N_t(n) - \Delta A^a_t) \).

\[
[\delta] \quad \forall t \in D, \forall r, q \leq k
\]

1) \[
\sum_{j:t_j \leq t} \Delta \langle M^n(r), M^n(q) \rangle_{t_j} \overset{P}{\rightarrow} \sum_{j:t_j \leq t} \sigma_{jq}^{r} \text{ as } n \uparrow \infty;
\]

2) \[
\Delta \langle M^n(r), M^n(q) \rangle_{t_j} \overset{P}{\rightarrow} \sigma_{jq}^{r} \text{ as } n \uparrow \infty, j = 1, 2, \ldots
\]

where \( \sigma_{jq}^{r} = (\sigma_{jq}^{r}) \), \( 1 \leq r, q \leq k \) is a nonnegative definite symmetric matrix with \( \sigma_{rr}^{r} \geq 0 \).

If \( M^n \in M_{\text{loc}}^{2}(F^n, F^n) \) for \( n = 1, 2, \ldots \) the conditions above imply that \( M^n \) converges in law to \( X \) in \( \mathcal{D} (\mathbb{R}^+, \mathbb{R}^k) \) endowed with the Skorokhod topology as \( n \uparrow \infty \).
Proof. The strategy of our proof is to show that the finite dimensional distributions of $M^n$ converge weakly to those of $X$ and then to show that the family of probability measures $\{p^n, n \geq 1\}$ is relatively compact from which it follows that $M^n$ converges in law to $X$.

Let $M_{\text{nc}}$ and $M_{\text{nd}}$ denote the locally square integrable local martingales defined by (1.2a) and (1.2b), respectively. For each $u \in \mathbb{R}^k$ define $\hat{A}(M_{\text{nc}}, u)$ by

\begin{equation}
\hat{A}(M_{\text{nc}}, u)_t = \int_{\mathbb{R}^k} (e^{iu \cdot x} - 1 - iu \cdot x)\nu_{\text{nc}}([0, t] \times dx), \quad t \geq 0
\end{equation}

and

\begin{equation}
\hat{A}(\xi, u)_t = -\frac{1}{2} \int_{\mathbb{R}^k} u_r^2 \xi^2 C_r(t), \quad t \geq 0
\end{equation}

where $\nu_{\text{nc}}$ is defined as in (1.9a) and $u \cdot x = \sum u_r x_r$. We show that $\hat{A}^n$ converges in probability to $\hat{A}$.

Recall that $\{M_{\text{nc}}^1, \ldots, M_{\text{nc}}^k\}$ are orthogonal local martingales so that for each $u \in \mathbb{R}^k$

\[(u \cdot M_{\text{nc}})^2 = \sum_{r=1}^{k} u_r^2 \int (X^n(r))^2 dA_{\text{nc}}(r) \in M_{\text{loc}}(F^n, p^n)\]

meaning that the compensator of $(u \cdot M_{\text{nc}})^2$ is the sum over $r$ of the compensator of $(u_r M_{\text{nc}}^r(r))^2$. Thus in terms of the random measure $\nu_{\text{nc}}$ we have

\begin{equation}
\sum_{r=1}^{k} u_r^2 \int_{\mathbb{R}^k} (X^n(r))^2 dA_{\text{nc}}(r) = \int_{\mathbb{R}^k} u^2 \nu_{\text{nc}}([0, \cdot] \times dx).
\end{equation}
Further, as in the proof of theorem 1, Lipster and Shiryayev [27] (see equation 65) \([u]^2\) implies via lemma 3 of [27] that

\[
(4.3) \quad \hat{A}(M^{nc}, u)_t + \frac{1}{2} \int_{\mathbb{R}^k} (u \cdot x)^2 \nu^{nc}([0, t] \times dx) \stackrel{P}{\to} 0, \quad P \ge 0
\]

as \(n \uparrow \infty\) and therefore (4.2) and \([\beta]^2\) imply that \(\hat{A}(M^{nc}, u)\) converge in probability to \(\hat{A}(\xi, u)\) for each \(u \in \mathbb{R}^k\).

On combining the above with \([\gamma]\) and the fact that \(\{t_j\}, j \ge 1\) is a fixed, hence predictable, sequence of stopping times we have by Jacod et al. [22], theorem 3.4, that the finite dimensional distributions of \(M^n\) converge to those of \(X\).

To prove that \(\{P^n, n \ge 1\}\) is relatively compact we define

\[
(4.4) \quad F^n_t = \sum_{j \le k} M^{nj}_t, \quad t \ge 0
\]

and observe that by \([\beta]^2\) and \([\delta]\) there exists an increasing deterministic function \(F^\infty\) such that

\[
(4.5) \quad F^n_t \stackrel{P}{\to} F^\infty, \quad P \ge 0 \text{ as } n \uparrow \infty
\]

and

\[
(4.6) \quad \Delta F^n_{t_j} \stackrel{P}{\to} \Delta F^\infty_{t_j} \quad \text{as } n \uparrow \infty
\]

with \(\Delta F^n_s = \Delta F^\infty_s = 0\) for \(s \notin J\). Since \(F^n\) is increasing and predictable an application of lemma 3.20, Jacod et al. [22] and Theorem 2.17, Jacod and Mémin [22a] shows that \(\{P^n, n \ge 1\}\) is relative compact. Thus \(M^n\) converges in law to \(X\) as \(n \uparrow \infty\). \(\square\)
Further Inference Problems for Poisson-Type Counting Processes

4.1 One sample case: tests of the class $K$

We apply the results of the two preceding chapters to the problem of testing a "simple" null hypothesis. We define a general class of test statistics and show how tests based on these statistics may be shown to be asymptotically consistent against specified alternatives.

Let $(N^i,Y^i)_i$, $i = 1,\ldots,n$ be $n$ independent $m$-variate Poisson-type counting processes each with parameter $b = (b_1,\ldots,b_m)$ so that $N^i(j)$ has compensator $\int Y^i db_j$ for $i = 1,\ldots,n$ and $j = 1,\ldots,m$. We are given a function $b^0 = (b^0_1,\ldots,b^0_m)$ such that $\forall j \leq m$ $b^0_j$ is right continuous and nondecreasing with $b^0_j(0) = 0$ and are to test the hypothesis $H_0 : b = b^0$ on $I \subset \mathbb{R}^+$; a possibly random subinterval.

Let $K$ denote the class of bounded predictable processes $K = (K_t)_t \geq 0$ such that $K$ is a known function of the observations $((N^i,Y^i), i = 1,\ldots,n)$ making $K$ an observable process. In general we allow for $K = (K(1),\ldots,K(m))$ to be a vector valued process.

Now let $b^n$ be the $n^{th}$ Aalen estimator of $b$ defined by (2.1.4) and let $R = \chi_I$ where $\chi_{\cdot}$ is the characteristic function of $\cdot$ and we require the process $R = (R_t)_t \geq 0$ to be adapted and predictable. For each $K \in K$ and $j = 1,2,\ldots,m$ define the process $W^n(j)$ by:
(1.1) \( \tilde{W}_t^n(j) = \int_0^t K_s(j) R_s \left( \tilde{b}_s^n(j) - \tilde{b}_s^0(j) \right), \quad t \geq 0 \)

where \( \tilde{b}_s^0 = \int I(Y^{(n)}(s) > 0)\tilde{b}^0 \) and the integral in (1.1) is an ordinary Stieltjes integral.

The process \( \tilde{W}^n = (\tilde{W}^n(1), \ldots, \tilde{W}^n(m)) \) may be used as a weighted measure of the difference between the idealized model or "true" \( \tilde{b} \) and the function \( b^0 \) specified in \( H_0 \) over the interval \( I \). For various choices of the weight function \( K \) we generate a class of tests of the hypothesis \( H_0 \). Often these tests will be based on the random variable \( \tilde{W}^n = (\tilde{W}^n_\infty(1), \ldots, \tilde{W}^n_\infty(m)) \) but in general we may consider tests based on \( \tilde{W}^n_t \) for \( t \) an arbitrary stopping time.

For compactness we summarize the main ideas above in the following assumption.

A1. \((N^i, Y^i), i = 1, \ldots, n\) are independent \( m \)-variate Poisson-type counting processes with common parameter \( b \). Given a right continuous nondecreasing function \( b^0 \) with \( b^0(0) = 0 \) and a possibly random set \( I \subset \mathbb{R}^+ \), we consider the null hypothesis

(1.2) \( H_0: b = b^0 \) on \( I \).

The family of one sample tests generated by the class \( K \) is broad so that in general we will not know the exact distribution of the test statistic under the hypothesis \( H_0 \). However, for some weight functions \( K \) the process \( \tilde{W}^n \) may have a limiting distribution under \( H_0 \) as \( n \uparrow \infty \).

In these cases our approach is to perform approximate tests of \( H_0 \) by
assuming that the law of $W^n$ is equal to its limiting law when $n$ is "large".

The use of a limiting law to perform tests based on $W^n$ suggests the following definition of an asymptotic size $\alpha$ test.

**Definition 1.1.** Let $K \in K$ and $W^n$ be defined by (1.1) and suppose $W^n \xrightarrow{P} W^\infty$ as $n \to \infty$, where $W^\infty$ is a Gaussian process of independent increments with covariance function $V$. Suppose $S$ is an $F^n$-stopping time and $S' \in \mathbb{R}^+$ such that $S \overset{P}{\to} S'$ and suppose further that for some sequence of processes $(V^n)_{n \geq 1}$ we have $V^n_S \overset{P}{\to} V_S', > 0$ as $n \to \infty$. Now choose $\alpha \in (0,1)$ and $\beta = (\beta_1, \ldots, \beta_m) \in \mathbb{R}^m$ so that

$$(1.3) \quad P(|\frac{W^n_S}{V^n_S'}(j)/\frac{V^n_{S'}(jj)}{V^n_{S'}} > \beta_j, j = 1, \ldots, m) = \alpha.$$ 

Under $A1$ a decision rule $\delta(S)$:

$$(1.4) \quad \text{reject } H_0 \text{ if } |\frac{W^n_S(j)}{V^n_S(jj)}| > \beta_j \text{ for some } j, 1 \leq j \leq m$$

is called an **asymptotic size $\alpha$ test** of $H_0$. $\square$

Recall that a sequence of tests $(T^n)$, $n \geq 1$ is called consistent if the power of the test $T^n$ under any fixed alternative tends to one as $n$ tends to infinity. This concept is extended to tests of the class $K$ involving Poisson-type counting processes. Our definition of consistency however requires that we show in the next lemma that $W^n$ is a local martingale.
Lemma 1.1. For each $n$ let $\hat{b}^n$ be the $n^{th}$ Aalen estimator of $b^n$ where $\hat{b}^n$ and $\hat{b}^n$ are defined by (2.2.1.4) and (2.2.1.5), respectively. Suppose $\hat{b}^n - \hat{b}^n \in M^2_{loc}(F^n, P^n)$ where $F^n$ is defined in section 2.2.2, then for each $K \in K$ the process $W^n = \int KR(db^n - dB^n) \in M^2_{loc}(F^n, P^n)$.

Proof. By assumption the process $K \cdot R$ is a bounded predictable process and observe that $\hat{b}^n - \hat{b}^n$ is a process of locally bounded local variation. Thus according to Shirayev [34, p. 209] $W^n \in M_{loc}(F^n, P^n)$ from which the boundedness of $KR$ implies $W^n \in M^2_{loc}(F^n, P^n)$ as well. \(\square\)

We use the conclusion of lemma 1.1 to define a consistent test based on the process $W^n$.

Definition 1.2. For fixed $K \in K$ let $W^n$ be the process defined in (1.1) and let $\tau$ be an arbitrary $F^n$-stopping time. Suppose $W^n \in M_{loc}(F^n, P^n)$ and let $\tau^n = (\tau^n_k), k \geq 1$ be a sequence of localizing stopping times for $W^n$. Given $A$ suppose $b^a$ is a fixed alternative to $H_0$ so that $b^a \neq b^0$ on $I$. We say $K$ induces at time $\tau$ a locally consistent test if for some $k$ the power of the test based on the statistic $W^n_{\tau_k^n}$ tends to one as $n \uparrow \infty$ under any alternative $b^a$.

Alternatively $K$ induces a $\tau_\infty$-consistent test if the power of the test based on $W^n_{\tau_\infty^n}$ tends to one as $n \uparrow \infty$ under any alternative $b^a$. \(\square\)

If $\hat{b}^n - \hat{b}^n \in M^2_{loc}(F^n, P^n)$ then lemma 1.1 implies that with the aid of the theory developed in Chapters 2 and 3 we can outline a general procedure to determine that a test of class $K$ is consistent. Roughly, suppose we show that under $H_0$, $W^n \xrightarrow{\mathbb{L}} W^\infty$ as $n \uparrow \infty$ where $W^\infty$ is a process
of independent increments and suppose further that $|W^n_T| \overset{P}{\to} \infty$ (in an appropriate sense) as $n \to \infty$ under any fixed alternative to $H_0$. If $|W^n_T| < \infty$ almost surely then a test based on $W^n_T$ is consistent.

Suppose we are interested in alternatives of the form

\[ H_a : b > b^0 \text{ on } I. \]

This motivates the restriction of $K$ to the subclass $K^+ \subset K$ where $K \in K^+$ is a nonnegative process. Thus a test of the class $K^+$ will be sensitive to departures from the hypothesis $H_0$ which are alternatives of the form (1.5). For any test of the class $K^+$ we can form a decision rule $\delta^+$ by a modification of definition 1.1 which omits the absolute value "|·|" from expressions (1.3) and (1.4).

For any weight function $K$ we can construct an asymptotic size $\alpha$ test if and only if the sequence of processes $W^n$ has a limiting distribution. In theorem 1.1 below we give a criterion for this when $K \in K^+$. We make the following assumption.

A2. For each $n \hat{b}^n - \hat{b}^n \in M^2_{loc}(F^n, p^n)$ and theorem 1.2, section 2.2.4 is valid for $n^{1/2}(\hat{b}^n - \hat{b}^n)$. Also, assumption A3, section 2.2.5 holds and $(\hat{c}^n)$, $n \geq 1$ (defined in section 2.2.5; see (3.4a) and (3.4b)) satisfies theorem 1.3, section 2.2.5. Thus $n^{1/2}(\hat{b}^n - \hat{b}^n) \overset{P}{\to} X$ where $X$ is a Gaussian process of independent increments and $n<\hat{b}^n - \hat{b}^n>$ (defined in (1.6), section 2.2.2) converges in probability to a deterministic function $\sigma$ for which $\hat{c}^n$ is a consistent estimator. □
Consider the following theorem.

Theorem 1.1. Assume A1 and A2 hold where in A1 I = \([t_0, t_1]\) for fixed \(t_0, t_1 \in \mathbb{R}^+\) with \(t_0 < t_1\). Let \(K \in \mathcal{K}^+\) and consider the following condition:

\[\alpha\] there exists a bounded function \(g = (g_1, \ldots, g_m)\) defined on I with

\[i)\] \(g_\ell(s) > 0, \forall s \in I, \forall \ell \leq m;\)

\[ii)\] \(\frac{K(\ell)}{n^{1/2}} \rightarrow g_\ell^{1/2}\) uniformly on I as \(n \rightarrow \infty, \forall \ell \leq m.\)

Form the process \(U^n\) defined in (1.1). Then under \(\alpha\) \(U^n\) converges in law to \(U^\infty\) where \(U^\infty\) is a Gaussian process of independent increments having covariance function \(V = (v_{ij}^\ell)\) given by

\[(1.6)\] \(v_{ij}^\ell_t = \int_0^t g_i(s)g_j(s)R(s) \, ds,\quad t \geq 0, 1 \leq i, j \leq m\)

where \(\sigma = (\sigma_{ij}^\ell)\) is defined by (3.3), section 2.2.5.

**Proof.** Fix \(K \in \mathcal{K}^+\) and let \(W^n = (W^n_t), t \geq 0\) be defined by (1.1) which by virtue of A2 and lemma 1.1 is of class \(W_2^{\text{loc}}(\mathbb{R}^n, \mathbb{P}^n)\). We drop the superscript "0" from the notation and assume \((N^i, Y^i), i = 1, \ldots, n\) are mutually independent Poisson-type counting processes with parameter \(b\) and form the process \(W^n\) relative to this parameter. Thus for each \(n\) \(W^n\) is written
\begin{equation}
\text{(1.7)} \quad W_t^n(i) = \int_0^t K_s(j) R(s) (dB_s^n(j) - dB_s^n(j)), \quad t \geq 0, \forall j \leq m.
\end{equation}

We show that the conditions of theorem 1.3 of Chapter 3 hold on I with \( X^n = K(Y^{(n)})^{-1}1(Y^{(n)} > 0) \).

By hypothesis A2 [\([\alpha]^2\) and [\([\beta]^2\) of theorem 1.2, Chapter 2 hold for the process \( n^{1/2}(\tilde{b}^n - \hat{b}^n) \) and so let \( h^n \) denote the l.h.s. of [\([\alpha]^2\) and [\([\beta]^2\) as needed and let \( h \) denote the respective in probability limit. Now to verify [\([\alpha]^2\) and [\([\beta]^2\) of theorem 1.3, Chapter 3 we observe that the respective left hand sides for \( W^n \) have the form:

\begin{equation}
\text{(1.8)} \quad Z^n = \int (K^2/n) R \, dh^n = \int (K^2/n-g) R \, dh^n + gR \, dh^n, \quad n = 1, 2, \ldots.
\end{equation}

By condition [\([\alpha]\) \( K^2/n-g \) converges uniformly on I to zero and g is bounded, thus simple calculations show that \( Z^n - \int gR \, dh \) converges in probability to zero. Hence [\([\alpha]^2\) and [\([\beta]^2\) of theorem 1.3, Chapter 3 are verified for \( W^n \).

Recall the definition of \( \{t_j\}, \, j \geq 1 \) as the ordered set of discontinuities of \( b \) as given in A1, section 2.2.4. To verify [\([\delta]\) we observe that

\[ \Delta M^n(r) M^n(q) \rangle_{t_j} = K^{(r)}(q) K^{(q)}(r) \langle b^n - \hat{b}^n \rangle_{t_j} \langle r \rangle_{t_j}, \quad t_j \in I, \, 1 \leq r, q \leq k, \]

where \( \langle b^n - \hat{b}^n \rangle \) is given by (1.6), section 2.2.2. Thus by hypothesis A3 and condition [\([\alpha]\) we have that [\([\delta]\) of theorem 1.3, Chapter 3 holds.
with covariance matrix \( Z_j = (g_i(t_j)g_j^k(t_j)g_k(t_j)), 1 \leq i, k \leq m \) and \( j = 1, 2, \ldots \).

Next let \( Z_j^n \) denote the random vector \( n^{1/2}(\Delta b_{t_j}^n - \Delta \tilde{b}_{t_j}^n) \) conditional on \( F_{t_j}^n \). Since the process \( K \) is predictable the conditional law of \( \Delta W_{t_j}^n \) given \( F_{t_j}^n \) is given by the law of \( (K_{t_j}/n^{1/2}) \cdot Z_j^n \). But under A2 condition [\( \delta \)] of theorem 1.2, Chapter 2 holds which implies that the law of \( Z_j^n \) converges in probability to the law of \( U_j \) for \( j \geq 1 \). Thus by [\( \alpha \)] and passing to subsequences it easily follows that the law of \( \Delta W_{t_j}^n \) converges in probability to the law of \( g_{t_j} \cdot U_j \) for \( t_j \in I \) and so [\( \gamma \)] of theorem 1.3, Chapter 3 holds as well (in the above \( a \cdot b = (a_1b_1, \ldots, a_mb_m) \) for vectors \( a \) and \( b \)).

Upon gathering together the arguments above we conclude via theorem 1.3, Chapter 3 that \( W^n \rightarrow W^\infty \) on \( I \) where \( W^\infty \) is a Gaussian process of independent increments.

We close this section by outlining a way in which theorem 1.1 may be used in the construction of an asymptotic size \( \alpha \) test and in verifying that the test is consistent. Again we assume A1 holds with \( I = [t_0, t_1] \) and suppose that our alternative hypothesis \( H_a \) has form (1.5) so that it is reasonable to consider a test of class \( K^+ \).

Also, assume the conditions of theorem 1.1 are in effect so that \( W^n \rightarrow W^\infty \).

To test \( H_0 \) we use \( W^\infty \) and \( W^n \) to form a decision rule \( \delta^+ \) according to definition 1.1 modified according to the remark made just before theorem 1.1. In practice the covariance function \( V \) of \( W^\infty \) may be
unknown making it necessary to estimate $V$ from the data. This will typically affect our ability to construct a test of exact size $\alpha$ asymptotically. However, if $(V^n, n \geq 1)$ is a consistent estimator for $V$ in some appropriate sense then for "large" $n$ we can substitute $V^n$ for $V$ in (1.3) to obtain an asymptotic test of approximate size $\alpha$.

Let $K \in K^+$ satisfy [a] of theorem 1.1 and define $V^n = (V^n(i,j))$ by

\[ V^n_t(i,j) = \int_0^t (K_s(i)K_s(j)/n)R(s)d\sigma^n_s(i,j), \quad \forall t \geq 0, \]

\[ 1 \leq i,j \leq m \]

where $(\sigma^n), n \geq 1$ is defined in section 2.2.5. By assumption A2 $(\sigma^n), n \geq 1$ is either locally or $t_\infty$-locally consistent for $\sigma$ and therefore by an argument analogous to that used with (1.8) $V^n$ is shown to be consistent for $V$ in the corresponding mode. Thus for large $n$ an asymptotic test of approximate size $\alpha$ may be constructed by substituting $V^n$ for $V$ in equation (1.3).

Earlier we mentioned that an asymptotic test based on $W^n$ is consistent if the test statistic tends to infinity as $n$ tends to infinity. For example let $b$ be an arbitrary alternative with $b > b^0$ on $I$ and assume that \{(N_i^i, Y_i^i)\} are Poisson-type counting processes with parameter $b$. Let $b^n = \int_l(Y_i^{(n)} > 0)db$ and consider $n^{-1/2}W^n$:
\[(2.0) \quad n^{-1/2} \tilde{w}^n = \int_n^{-1/2} \text{KR}(db^n - db^0) = \int_n^{-1/2} \text{KR}(db^n - db^0) + \int_n^{-1/2} \text{KR}(db^n - db^0).\]

Suppose that

\[(2.1) \quad \gamma(n) \overset{P}{\rightarrow} \infty \text{ over } I \text{ as } n \uparrow \infty \]

and recall that under A2 \((\hat{b}^n)\), \(n > 1\) is locally or \(\tau^*\)-locally consistent for \(b\). Therefore by the boundedness of KR it follows that \(n^{1/2}\) times the first term of the second equality in (2.0) tends to zero in probability in the appropriate mode. Now using (2.1) and \([\alpha]\) of theorem 1.1 it is easily shown that

\[(2.2) \quad \int_n^{-1/2} \text{KR}(db^n - db^0) = \int_n^{-1/2} \text{KR} \{\gamma(n) > 0\}(db - db^0) \overset{P}{\rightarrow} \int gR(db - db^0) > 0.\]

Hence \(\tilde{w}^n = n^{1/2} \int_n^{1/2} \text{KR}(db^n - db^0)\) tends to infinity over \(I\) and our informal argument suggests that \(\tilde{w}^n\) may be used to construct a consistent test.

For \(\tau\) an arbitrary \(F^n\)-stopping time consider the decision rule \(\delta^+(\tau \wedge \tau_{1}^n_k)\) for some \(k \in \{1, 2, \ldots, \infty\}\) where we use the substitution \(V^n\) for \(V\) as needed. Suppose that \(\tau \wedge \tau_{1}^n_k \overset{P}{\rightarrow} T\) for \(T \in \mathbb{R}^+\) and that \(V^n \overset{P}{\rightarrow} V_T > 0\) as \(n \uparrow \infty\). If \(\tilde{w}^n \overset{P}{\rightarrow} \infty\) as \(n \uparrow \infty\) then the rule \(\delta^+(\tau \wedge \tau_{1}^n_k)\) is consistent by theorem 1.1.
An elementary example of a weight function $K$ satisfying $[\alpha]$ is to choose $K$ so that

\[(2.3) \quad K_{s}(j) = n^{1/2}, \quad \psi_s \geq 0, \quad 1 \leq j \leq m.\]

In some situations a test based on $W_{\infty}^n$ may be shown to be consistent using this weight function.

4.2 Two sample case: tests of the class $K$

We extend the results of section 4.1 to the problem of comparing two subpopulations from two independent samples of Poisson-type counting processes. Let $(N^{ij}, Y^{ij})$, $i = 1, \ldots, n_j$ be $n_j$ independent $m$-variate Poisson-type counting processes each with parameter $b^j = (b_{1j}, \ldots, b_{mj})$ for $j = 1, 2$. Thus $N^{ij}(k)$ has compensator $\int Y^{ij}(k) \, db^j_k$, $1 \leq i \leq n_j$, $1 \leq k \leq m$, $j = 1, 2$ where we assume $\{(N^{i1}, Y^{i1}), i = 1, \ldots, n_1\}$ is independent of $\{(N^{i2}, Y^{i2}), i = 1, \ldots, n_2\}$. Suppose that we want to test

\[(1.1) \quad H_0: \quad b_1 = b_2 \text{ on } I \subset \mathbb{R}^+\]

where $I$ is a possibly random subinterval.

Let $K$ denote the class of bounded predictable processes $K = (K_t)$, $t \geq 0$ such that $K$ is a known function of the data $\{(N^{ij}, Y^{ij}), i = 1, \ldots, n_j; j = 1, 2\}$ making $K$ an observable process. Again, in general we allow $K$ to be an $m$-dimensional vector process.
For each sample \( j = 1, 2 \) let \( \hat{b}^{n_j}(j) \) denote the \( n_j \)th Aalen estimator of \( b_j \) defined by (2.2.1.4). Also put \( R = \chi_I \) where we require \( R = (R_t), t \geq 0 \) to be a predictable process. For each \( K \in K \) and \( k = 1, \ldots, m \), define the process \( W^{n_1n_2}_t(k) \) by:

\[
(1.2) \quad W^{n_1n_2}_t(k) = \int_0^t K_s(k) R_s(d\hat{b}^{n_1}_s(k, 1) - d\hat{b}^{n_2}_s(k, 2)), \quad t \geq 0
\]

where the integral in (1.2) is an ordinary Stieltjes integral.

The process \( W^{n_1n_2} = (W^{n_1n_2}_1, \ldots, W^{n_1n_2}_m) \) may be used as a weighted measure of the estimated difference between \( b_1 \) and \( b_2 \) over \( I \) in the idealized model for sample 1 and sample 2, respectively.

Again for various choices of the weight function \( K \) we generate a class of tests of the hypothesis \( H_o \) given in (1.1). Different choices of weight functions will show differing sensitivity to departures from \( H_o \) and we will want to choose \( K \) to reflect our prior interest in detecting certain departures.

The process \( W^{n_1n_2} \) will be used to construct tests of \( H_o \) in the two sample case analogous to what was done in the one sample case in section 4.1. As before the generality of the class \( K \) requires that we rely on the asymptotic properties of \( W^{n_1n_2} \) to develop a test. In this connection we given conditions under which the process \( W^{n_1n_2} \) defines a locally square integrable local martingale. In general we may write \( W^{n_1n_2} \) as

\[
(1.3) \quad W^{n_1n_2} = \int KR(dM^{n_1}_1 - dM^{n_2}_2) + \int KR(d\hat{b}^{n_2}_2 - d\hat{b}^{n_1}_1)
\]
where \( M_j^{(n)}(j) = \hat{b}_j^{(n)}(j) - \tilde{b}_j^{(n)}(j) \) and \( b_j^{(n)}(j) = \int 1(Y_j^{(n)} > 0) db_j \), where

\[
Y_j^{(n)} = \sum_{\ell=1}^{n_j} Y_{\ell j}^{(n)} \quad \text{for } j = 1, 2.
\]

Now let \( F_{t_1, t_2}^{n_1, n_2} = (F_{t_1}^{n_1} \vee F_{t_2}^{n_2}), t > 0 \) where \( F_{t}^{n_j} = (F_{t}^{n_j}), t > 0 \) is defined as in section 2.2.2 for samples \( j = 1 \) and \( j = 2 \), respectively. It follows directly from lemma 1.1, section 4.1 that if \( M_j^{(n)}(j) \in M^2_{\text{loc}}(F_{t_1}^{n_1}, n_1^{(n_1)}, n_2^{(n_2)}, F_{t_2}^{n_2}) \) for \( j = 1, 2 \) then the first term on the r.h.s. of (1.3) is of class \( M^2_{\text{loc}} \) as well.

Also, we may easily show that in general the second term on the r.h.s. of (1.3) is a process of local bounded variation and therefore \( W_1^{n_1, n_2} \) is a semimartingale (see Shirayev [34], page 201 or Meyer [29]). This suggests the following corollary to lemma 1.1, section 4.1.

**Corollary 1.1.** Let \( W_1^{n_1, n_2} \) be defined as in (1.3) with \( M_j^{(n)}(j) \in M^2_{\text{loc}} \) for \( j = 1, 2 \). Suppose that under the null hypothesis \( H_0 \) the function \( b_1 = b_2 \). If the process \( K \in K \) is modified so that

\[
(1.4) \quad \{ (\omega, t) : K(\omega, t) \neq 0 \} = \{ (\omega, t) : Y_1^{(n_1)}(\omega, t) > 0, Y_2^{(n_2)}(\omega, t) > 0 \}
\]

then \( W_1^{n_1, n_2} \) is a locally square integrable local martingale.

**Proof.** The proof follows from lemma 1.1 since the second term in (1.3) vanishes under the restriction (1.4). \( \square \)

The modification of the weight function \( K \) according to (1.4) is motivated by the fact that we can estimate \( b_1 \) and \( b_2 \) only when \( Y_{1}^{(n_1)} > 0 \) and \( Y_{2}^{(n_2)} > 0 \), respectively and thus we can compare \( b_1 \) with \( b_2 \) from the data only when \( Y_{1}^{(n_1)} \) and \( Y_{2}^{(n_2)} \) are simultaneously nonzero. Therefore
we assume that $K \in K$ has been accordingly modified throughout the remainder of this section.

Clearly definition 1.2 of a consistent test and definition 1.1 of an asymptotic size $\alpha$ test generalize to the two sample case in an obvious way. Further when the conclusion to corollary 1.1 holds then a straightforward extension to theorem 1.1, section 4.1 may be proved. In developing this extension one needs to have $\min(n_1, n_2) \to \infty$ and to have (ii) of $[\alpha]$ hold with $n$ replaced by $n_1$ and $n_2$ and to assume that $A_2$ holds in each sample. In this way we may obtain a limit Gaussian process $\bar{W}^\infty$ for $\bar{W}^{n_1n_2}$ under the hypothesis $H_0$ and use $\bar{W}^\infty$ to construct asymptotic tests of the class $K$ in the two sample case.

The asymptotic covariance function $V$ of $\bar{W}^\infty$ has the following form:

$$V^{ij}_t = \int_0^t g_s(i)g_s(j)(d\sigma^{ij}_s(1) + d\sigma^{ij}_s(2)), \ t \geq 0, \ 1 \leq i, j \leq m$$  

where $\sigma(j) = (\sigma^{ik}_s(j))$ is the limiting variance of sample $j$ defined in equation (3.3), section 2.2.5 for $j = 1,2$. Certainly if $\{(N^{11}, Y^{11})\}$ and $\{(N^{12}, Y^{12})\}$ are identically distributed then $\sigma(1) = \sigma(2)$ although this is not required in our theory. When $V$ is unknown there is no unique way to build a consistent estimate for (1.5). For example, if under the null hypothesis $H_0: b_1 = b_2$ and $\sigma(1) = \sigma(2)$ we may want to pool the data from the two samples together to form a single sample of size $n_1 + n_2$ and use a pooled estimate of $V$ having the form (1.9), section 4.1. Alternatively, we may define $V^{j}(j)$ as in (1.9),
section 4.1 for \( j = 1, 2 \) and use \( V_{12}^{n_1 n_2} \) to estimate \( V \) where

\[
(1.6) \quad V_{12}^{n_1 n_2} = V_{1}^{n_1} + V_{2}^{n_2}.
\]

In either case it is possible to establish consistency for these estimators as shown in section 4.1 for the one sample case.

The discussion above suggests that when \( V \) is unknown we can construct asymptotic tests of approximate size \( \alpha \) by using a consistent estimator \( V_{12}^{n_1 n_2} \) in place of \( V \) in the distribution of \( W_{n}^{\infty} \). For alternatives of the form \( H_a : b_1 > b_2 \) on \( I \), then we suggest that the weight function be chosen from the class \( K^+ \) and modified according to \( (1.4) \).

As earlier to establish that the test is consistent it suffices to shown that the test statistic tends to infinity in probability as \( \min(n_1, n_2) \to \infty \) for any fixed alternative. For example, choose a fixed alternative \( b_1 \) and \( b_2 \) with \( b_1 > b_2 \) and suppose the test is of class \( K^+ \). If \( (\hat{b}_1^{n_1}(1)), n_1 \geq 1 \) and \( (\hat{b}_2^{n_2}(2)), n_2 \geq 1 \) are consistent estimators for \( b_1 \) and \( b_2 \), respectively, then equation \( (1.3) \) and the argument surrounding \( (2.2) \), section 4.1 suggests that \( W_{12}^{n_1 n_2} \) will tend to plus infinity over \( I \) in an appropriate sense. In such a situation a decision rule \( \delta^+ \) based on \( W_{12}^{n_1 n_2} \) may be shown to be consistent as indicated in section 4.1.

An elementary example of a weight function \( K \) is given by

\[
(1.7) \quad K = \frac{1}{n_1^{1/4}} \frac{1}{n_2^{1/4}}
\]
where we assume \( n_1/n_2 \to 1 \) and \( \min(n_1,n_2) \to \infty \) and that \( I \) is a fixed subinterval. The weight function in (1.7) is a natural extension to the two sample case of the weight function (2.3), section 4.1 defined for the one sample case.

We note in closing that this outline of a general approach to two sample tests reduces to the approach outlined in section 7, Aalen [3] when the intensity exists and reduces to that outlined in Chapters 3 and 4, Gill [18] when the Poisson-type counting processes are applied to the life testing situations.

4.3 Application to Survival Analysis

We include under the heading survival analysis any application to life-testing, medical clinical trials, biological experimentation, etc. involving the observation of independent and possibly censored positive random variables. These random variables usually arise when measuring the time elapsed until some interesting event occurs to an observational unit and have been variously called life times, holding times, sojourns or durations. For example, we may measure the time elapsed from when a blue footed booby chick is first hatched from its egg until as a young fledgling it takes its first flight. If the young bird is eaten by a predator before its first flight the elapsed time cannot be observed and we may say it has been censored.

The situation described above is given in formal terms as follows:
Random Censorship Model: For each $i = 1, 2, \ldots, n$ and $U_i$, $i = 1, \ldots, n$ are $2n$ independent random variables with $X_i$ or $U_i$ almost surely finite for each $i$. $X_i$ has (sub-)distribution function $F$ and $U_i$ has (sub-)distribution function $L$. The observable random variables $\tilde{X}_i$ and $\delta_i$ are defined by $\tilde{X}_i = X_i \wedge U_i$ and $\delta_i = 1(X_i \leq U_i)$. When $\delta_i = 0$ we say the observation $X_i$ has been right censored at $U_i$.

To analyze problems arising in the random censorship model we define the following processes for each $i = 1, \ldots, n$:

\begin{align*}
(1.1) \quad & N^i_t = 1(\tilde{X}_i \leq t, \delta_i = 1), \ t \geq 0 \\
(1.2) \quad & Y^i_t = 1(\tilde{X}_i \geq t), \ t \geq 0 \\
(1.3) \quad & M^i_t = N^i_t - \int_0^t Y^i_s \, db(s), \ t \geq 0
\end{align*}

where $b$ is a deterministic function given by

\begin{equation}
(1.4) \quad b(t) = \int_0^t (1 - F(s))^\frac{1}{\delta} \, dF(s), \ t \geq 0.
\end{equation}

The process $N^i_t = (N^i_t), t \geq 0$ is called a simple counting process and is equal to zero until the $i^{th}$ observation time elapses and has not been censored while $Y^i_t = (Y^i_t), t \geq 0$ is called a risk process and is equal to one as long as the $i^{th}$ observation remains under observation (i.e. the observed time $\tilde{X}$ has not elapsed and it has not been censored). We show below that for each $i = 1, \ldots, n$ the process $N^i_t$ is a Poisson-
type counting process with parameter \( b \) and that \( M^i_t = (M^i_t) \), \( t \geq 0 \) is a martingale. The function \( b \) is sometimes called the cumulative hazard or risk function and it uniquely defines the (sub)-distribution function \( F \).

Let \( (\Omega^n, \mathcal{F}^n, P^n) \) be a probability space on which \( X_i, U_i \) are defined, \( i = 1, \ldots, n \). \( \mathcal{F}_t^n = (\mathcal{F}_t^n) \), \( t \geq 0 \) denotes a filtration given by

\[
(1.5) \quad \mathcal{F}_t^n = \mathcal{F}_0^n \vee \sigma(\{1(\tilde{X}_i \leq s), \delta_i 1(\tilde{X}_i \leq s), s \leq t, i = 1, \ldots, n\})
\]

where \( \mathcal{F}_0^n \) contains the \( P^n \)-null sets of \( \mathcal{F}^n \) and their subsets. Now by Gill [18], theorem 3.1.1 the process \( M^i = (M^i_t, F_t) \), \( t \geq 0 \) is a square integrable martingale so that \( N^i \) has compensator \( A^i = (A^i_t, F_t) \), \( t \geq 0 \) which satisfies

\[
(1.6) \quad A^i\{B\} = \int_B Y^i_s \, db(s) \quad \text{for any Borel set } B \text{ in } \mathbb{R}^+.
\]

Thus \( N^i \) is Poisson-type counting process with parameter \( b \) for each \( i = 1, \ldots, n \).

The counting process approach to the random censorship model in survival analysis is the modern approach to this model and was first introduced by Aalen [1] in his Ph.D. thesis at Berkeley. In his theory of counting processes Aalen assumes that the intensity exists and thus that the function \( F \) has a density relative to Lebesgue measure. Gill [18], on the other hand, uses Aalen's counting process approach to study a general random censorship model but he does not
assume that a density exists. Hence the theory of inference in the random censorship model developed by Gill [18] arises as a special case of our theory for Poisson-type counting processes.

Two estimation problems arise in the random censorship model, that of estimating the (sub)-distribution function $F$ and of estimating the cumulative hazard function $b$. To solve these problems we use the Aalen estimator $\hat{b}^n$ to estimate $b$ and the product limit estimator $\hat{F}^n$ to estimate $F$. For each $n = 1, 2, \ldots$ define the processes $M^n, Y^n, J^n$ by

$$
(1.5) \quad M^n = \sum_{i=1}^{n} M^i
$$

$$
(1.6) \quad Y^n = \sum_{i=1}^{n} Y^i
$$

$$
(1.7) \quad J^n = 1(Y^n > 0)
$$

where $M^i$ and $Y^i$ are defined in (1.3) and (1.2), respectively. Now the $n^{th}$ Aalen estimator $\hat{b}^n$ is the so-called empirical cumulative hazard function defined by

$$
(1.8) \quad \hat{b}^n_t = \int_0^t (Y^n_s)^{-1} J^n_s dN^n_s, \quad \forall t > 0
$$

where $N^n = \sum_{i=1}^{n} N^i$ (see (1.1)) whereas the product limit estimator $\hat{F}^n$ is defined by

$$
(1.9) \quad \hat{F}^n_t = 1 - \prod_{s \leq t} (1 - \Delta \hat{b}^n_s), \quad \forall t > 0.
$$
According to Gill [18], historically the estimators \( \hat{b}_n \) and \( \hat{F}_n \) were introduced in statistics by Nelson [29a] and Kaplan and Meier [22b], respectively, although versions had long been known in the fields of demography and actuarial science. In fact the estimator \( \hat{F}_n \), which reduces to the ordinary empirical distribution function when there is no censoring, is also known as the Kaplan-Meier estimator.

The product limit estimator and the empirical cumulative hazard function is studied by Gill [18] under a more general random censorship model than is given above. In Chapter 4 he develops asymptotic results for these estimators which under certain conditions include consistency and weak convergence to a Gaussian process of independent increments. Weak convergence for \( \hat{b}_n \) is proved explicitly in theorem 4.2.2 and weak convergence for \( \hat{F}_n \) is covered by a more general theorem 4.2.1. We are interested here to give a simpler and more elegant proof of weak convergence for these processes based on our theorem 1.3, Chapter 3.

First observe that the process \( Y^n \) defined by (1.6) is an integer valued process and for each \( \omega \in \Omega \) is decreasing in \( t \). For each \( n = 1, 2, \ldots \) we define the \( F^n \)-stopping times \( (\tau^n_k) \), \( k \leq n \) by \( \tau^n_1 = \inf\{t: t \geq 0, Y^n_t = Y^n_0 - 1\} \) and

\[
(2.0) \quad \tau^n_k = \inf\{t: t > \tau^n_{k-1}, Y^n_t = Y^n_{\tau^n_{k-1}} - 1\}, \quad k = 2, \ldots, n.
\]

Note that \( \tau^n_n = \inf\{t: t \geq 0, Y^n_t = 0\} = \sup\{t: t \geq 0, Y^n_t > 0\} \). We have the following lemma.
Lemma 1.1. Let $Y^n = (Y^n_t), t \geq 0$ be the process defined by (1.6) and $(\tau_k^n), k \leq n$ be the $F^n$-stopping times defined by (2.0). There exists a left continuous nonnegative function $y$ defined on $\mathbb{R}^+$ such that

$$
(2.1) \sup_{t>0} |n^{-1} Y^n_t - y(t)| \overset{P}{\to} 0 \text{ as } n \uparrow \infty
$$

and if $u = \sup \{t: y(t) > 0\}$ then

$$
(2.2) \tau_n \overset{P}{\to} u \text{ as } n \uparrow \infty.
$$

Proof. To prove (2.1) put $y(t) = E 1(\bar{X} \geq t), \forall t \geq 0$ and note that by the strong law of large numbers and a Glivenko-Cantelli type argument (2.1) can be shown to hold with probability one and therefore in probability.

To prove (2.2) we observe that $P(\tau_n \leq u) = 1$ for all $n$ and that for arbitrary $\varepsilon > 0$ such that $y(u-\varepsilon) > 0$

$$
(2.3) \{ \omega: \tau_n \leq u-\varepsilon \} \subset \{ \omega: n^{-1} Y^n_t = 0, \forall t \geq u-\varepsilon \}
$$

$$
\subset \{ \omega: \sup_{t>u-\varepsilon} |n^{-1} Y^n_t - y(t)| \geq y(u-\varepsilon) \},
$$

since $y$ is a nonincreasing function. Now since $\varepsilon$ may be chosen arbitrarily small and still maintain $y(u-\varepsilon) > 0$ (2.1) and (2.3) imply $P(\tau_n < u-\varepsilon) < \delta(n)$ where $\delta(n) \downarrow 0$ as $n \uparrow \infty$. Combining the two preceding statements proves the result. □
We define $\tilde{b}^n$ as follows:

$$ (2.4) \quad \tilde{b}^n_t = \int_0^t J^n_s \, db(s) $$

where the process $J^n$ is defined by (1.7) and the function $b$ by (1.4). To characterize the asymptotic distributional properties of $\hat{b}^n$ and $\hat{f}^n$ we consider the normalized processes $n^{1/2}(\hat{b}^n - \tilde{b}^n)$ and $n^{1/2}(\hat{f}^n - F)$ and apply the theory of Chapter 3. Therefore we need the following lemma.

**Lemma 1.2.** For each $n = 1, 2, \ldots$ let $\hat{b}^n$ and $\tilde{b}^n$ be defined by (1.8) and (2.4), respectively, and let $\hat{f}^n$ be defined by (1.9) for the random censorship model. Then we have

i) $\hat{b}^n - \tilde{b}^n \in M^2(F^n, P^n),$

ii) if $X^n = (\hat{f}^n - F(t^{\tau_n}))$, $t \geq 0$ then

$$ X^n \in \frac{1}{1-F} \, M^2(F^n, P^n) \text{ where } \tau_n \text{ is defined by (2.0)}. $$

**Proof.** Observe that for each $t \geq 0$ and $\omega \in \Omega$

$$ \hat{b}^n_t - \tilde{b}^n_t = \int_0^t Y^n_s J^n_s \, dM^n_s $$

where $M^n$ is a square integrable martingale defined by (1.5). Since $Y^n$ is an integer valued process it follows that
\[
\sup_t (Y^n_t)^{-1} 1(Y^n_t > 0) \leq 1
\]

where the convention \(0^0 = 0\) has been invoked. Further \(Y^n J^n\) is predictable so that \(\hat{b}^n - b^n \in M^2(F^n, p^n)\) follows as in lemma 1.1, section 4.1.

Next, according to Gill [18, p. 79]

\[
(2.5) \quad \hat{F}^n - F = (1-F) \int \frac{\chi_{(0,1)}(\Delta b)}{1 - \Delta b} \frac{1 - \hat{F}^n_\cdot}{1 - F_\cdot} (Y^n)^{-1}J^n \, dM^n
\]

on \(\{t: Y^n_t > 0\}\) where \(G_\cdot\) denotes a function such that \(G_\cdot(s) = \lim G(s) = G(s^-)\). Observe that \(\{t: Y^n_t > 0\} = [0, \tau^n]\) by definition of \(\tau^n\). Since \(\chi_{[0,1]}(\Delta b)/(1 - \Delta b) \cdot (1 - \hat{F}^-)(Y^n)^{-1}J^n\) is finite, and predictable (\(b\) is deterministic and \(1 - \hat{F}^-\) is left continuous and adapted) the result will follow if \(F^- < 1\) on \([0, \tau^n]\) almost surely.

To prove that \(F^- < 1\) on \([0, \tau^n]\) almost surely observe that for any \(t\)

\[
(2.6) \quad P(Y^n_t > 0) = \sum_{k=1}^{n} \binom{n}{k} y(t)^k(1 - y(t))^k
\]

where \(y(t) = E 1(\tilde{X} > t)\). Thus \(Y^n_t > 0\) with positive probability if and only if \(y(t) > 0\). By Gill [18, p. 79] \(y(t) > 0\) implies that \(F_\cdot(t) < 1\).
Therefore for each $\omega$ outside a $P^n$-null set $\tau_n^n = 1$ implies $y(\tau_n^n) > 0$ implies $F_-(\tau_n^n) < 1$ and since $F_-$ is nondecreasing $F_- < 1$ on $[0, \tau_n^n]$ $P^n$-almost surely. Hence it follows that the integrand in (2.5) is bounded and predictable $P^n$-almost surely from which (ii) is easily obtained. □

We proceed by stating a result due to Gill [18] establishing a kind of local uniform consistency of the estimator $\hat{n}$ and $\hat{b}^n$.

**Lemma 1.3.** Let $t \in (0, \infty]$ be such that

(2.6) \( Y_t^n \xrightarrow{P} \infty \) as $n \uparrow \infty$

and

(2.7) \( F(t-) < 1 \).

Then

(2.8) \( \sup_{s \in [0, t]} |\hat{n}_s^n - F(s)| \xrightarrow{P} 0 \) as $n \uparrow \infty$

and

(2.9) \( \sup_{s \in [0, t]} |\hat{b}_s^n - b(s)| \xrightarrow{P} 0 \) as $n \uparrow \infty$.

**Proof.** See theorem 4.1.1, page 56, Gill [18]. □

We apply theorem 1.3, Chapter 3 to prove weak convergence of the product limit estimator and the empirical cumulative hazard function to a certain Gaussian process of independent increments.
Theorem 1.1. Let $Y^n/n$ converge uniformly on $[0, \infty)$ to a function $y$ in probability as $n \uparrow \infty$. Then

1) $n^{1/2}(\hat{\beta}^n - \beta) \xrightarrow{L} [(1 - F)X[0,1](\Delta b)/(1 - \Delta b)] \cdot X$ as $n \uparrow \infty$

and

2) $n^{1/2}(\hat{b}^n - \tilde{b}^n) \xrightarrow{L} X$ as $n \uparrow \infty$

on $\mathbb{D}(I, \mathbb{R})$, where $I = \{t: y(t) > 0\}$ and $X$ is a zero-mean Gaussian process of independent increments and variance function $\mu$

$$
(3.0) \quad \langle X \rangle_t = \int_0^t y(s)^{-1}(1 - \Delta b_s)db_s, \quad \forall t \in I.
$$

Proof. If $u = \sup\{t: y(t) > 0\}$ we assume $u > 0$ otherwise the theorem is vacuously true. Thus $I = [0,u)$ when $y(u) = 0$ or $I = [0,u]$ when $y(u) > 0$. To prove (2) observe that by lemma 1.2 $n^{1/2}(\hat{b}^n - \tilde{b}^n)$ is a square integrable martingale on $I$ and so we will apply theorem 3.1.3 with $m = 1$. By lemma 2.2.1.1, equation (2.2.1.6) we have

$$
(3.1) \quad \langle n^{1/2}(\hat{b}^n - \tilde{b}^n) \rangle_t = \int_0^t n(Y^n_s)^{-1}1(Y^n_s > 0)(1 - \Delta b_s)db_s, \forall t \in I.
$$

Now by lemma 1.1, equation (2.1) it follows that (3.1) converges over $I$ to (3.0) in probability as $n \uparrow \infty$, thus proving $[\beta]^2$ and $[\delta]$ of theorem 3.1.3 with $C = \int y^{-1}db$ and for each $j$ such that $t_j \in I$ $O_j^2 = y(t_j)(1 - \Delta b(t_j))\Delta b(t_j)$ where $\{t_j\}$ is the set of discontinuities of $b$. 
To prove $[\alpha]^2$ we observe for each $t \in I$ and $\varepsilon \in (0, 1]$ the l.h.s. of $[\alpha]^2$ is given by

$$(3.2) \ n \int (Y^n)^{-1}((Y^n)^{-1}J^n > \varepsilon)db^c$$

where $b^c = b - \Sigma \Delta b$. Now since $y$ is strictly positive on $I$ and $n^{-1}Y^n \mathcal{P} y$ uniformly on $I$ it follows easily that (3.2) converges over $I$ to zero in probability as $n \uparrow \infty$, thus verifying $[\alpha]^2$.

To complete the proof of (2) it remains to be shown that the law of the random variable $Z_j = n^{1/2}(Y^n)^{-1}J^n_{t_j} (\Delta N^n_{t_j} - \Delta b(t_j))$ conditional on $F^n_{t_j}$ converges to the law of a Gaussian random variable $U_j$ with zero-mean and variance $\sigma_j^2 = y(t_j)(1 - \Delta b(t_j))\Delta b(t_j)$ in probability for all $t_j \in I$. But by Gill [18, theorem 3.1.1] $\Delta N^n_{t_j}$ conditional on $F^n_{t_j}$ is a binomial random variable with parameters $Y^n_{t_j}$ and $\Delta b(t_j)$ and therefore if $p_j = \Delta b(t_j)$ the Laplace transform of $Z_j$ conditional on $F^n_{t_j}$ is given by

$$(3.3) \ E(e^{rZ_j}|F^n_{t_j}) = \exp\{-n^{1/2}Y^n_{t_j} p_j r\} \times$$

$$(1-p_j + p_j \exp\{n^{1/2}Y^n_{t_j} - 1J^n_{t_j} r\})$$

$= [(1-p_j)\exp\{-n^{1/2}Y^n_{t_j} - 1J^n_{t_j} r\} + p_j \exp\{n^{1/2}Y^n_{t_j} - 1J^n_{t_j} r(1-p_j)\}]_{t_j}^r, \forall r \in \mathbb{R}.$

Now for each $t_j \in I$, $Y^n_{t_j}$ is of the order $ny(t_j)$ in probability so that the r.h.s of (3.3) is of the order
\[(3.4) \quad [\exp{-n^{-1/2}y(t_j)r} + p_j \exp{n^{-1/2}y(t_j)r(1-p_j)}]^\eta(y(t_j))\]

in probability as \(n \uparrow \infty\). But it is easily shown by calculation that

\[\frac{-\frac{1}{2}r^2(y(t_j)p_j(1-p_j))^{-1}}{2}\]

(3.4) converges to \(e^{-\frac{1}{2}r^2y(t_j)^2}\) as \(n \uparrow \infty\) and therefore

(3.3) converges to the Laplace transform of a Gaussian random variable

with zero-mean and variance \(\sigma_j^2 = y(t_j)(1 - \Delta b(t_j))\Delta b(t_j)\) in probability

as \(n \uparrow \infty\) for all \(t_j \in I\). On combining the results above (2) follows

from theorem 3.1.3.

To prove (1) observe that for each \(n = 1, 2, \ldots\)

\[(3.5) \quad n^{1/2}(\hat{\tau}_n^n - F(\tau_n^n))\]

\[= n^{1/2}(1 - F(\tau_n^n)) \int_0^{\tau_n^n} \frac{\chi_{[0,1]}(\Delta b(s))}{1 - \Delta b(s)} \frac{1 - \hat{F}(s-)}{1 - F(s-)} (\eta(s))^{-1} \frac{d\gamma_s^n}{s}\]

\[= n^{1/2}(1 - F(\tau_n^n)) \int_0^{\tau_n^n} H_s(\hat{b}_s^n - b^n_s), \quad \forall t \in I\]

where the second term is by lemma 1.2 a local square integrable martingale. With

\[H = [\chi_{[0,1]}(\Delta b)/(1 - \Delta b)][1 - F^n_\gamma)/(1 - F_\gamma)]\]

But \(n^{1/2}(\hat{b}_n - b^n)\)

converges weakly to \(X\) on \(D(I)\) as \(n \uparrow \infty\) and by lemma 1.3 \(H\) converges

uniformly on \(I\) to \(\chi_{[0,1]}(\Delta b)/(1 - \Delta b)\) in probability as \(n \uparrow \infty\).
Recalling that by lemma 1.1 \( \tau_n^P \xrightarrow{n} u \) as \( n \to \infty \), theorem 1.1, section 4.1 may be adapted straightforwardly to prove (1). □

Remark: The proof of (1) in theorem 1.1 is to be compared with that of theorem 4.2.2, Gill [18] which relies on Gill [18, theorem 4.2.1] which involves a rather elaborate construction designed to transform the problem into one in which the cumulative hazard function \( b \) is a continuous function. The building of this construction relies on the fact that the law of \( \Delta N^n_{t_j} \) conditional on \( F^n_{t_j} \) is Binomial with parameters \( Y^n_{t_j} \) and \( \Delta b(t_j) \), a fact which we relied on as well although in a much more elementary way. Our more elegant proof is adapted from Jacod, et al. [22] which as an extension of Lipster and Shiryaev [27] and was almost anticipated by Gill [18, page 78].

Under a general random censorship model the problem of comparing two subpopulations on the basis of two independent samples from those populations has been given careful attention by Gill [18, Chapters 3 and 4]. He introduces the family of tests of class \( K(K^+) \) considered in sections 4.1 and 4.2 of this thesis. A thorough examination is given to three tests of the class \( K^+ \), namely the tests due to Gehan, Efron and Cox, respectively. We remark that the weak convergence results he obtains for these tests may be proved straightforwardly by adaptation of the method we used to prove theorem 1.1. □
4.4 Likelihoods for Poisson-type counting processes: self-excited case

The likelihood principle says, roughly, that in making inferences about unknown parameters in a statistical model all relevant sample information about the parameter is contained in the likelihood (see for example Berger [7]). Therefore, in cases where the latter exists, inferences about unknown parameters should depend on the sample information through the data we observe by way of the likelihood. Such likelihood based inferences will often incorporate loss functions and prior information as contributing aspects of the inference process.

Believers in the likelihood principle will want to know when a likelihood can be explicitly written down for a Poisson-type counting process. We consider this problem now.

Let $\mathcal{X}$ denote the space of piecewise constant functions $x = (x_t, t \geq 0)$ such that $x_t = x_{t-} + (0$ or $1), x_0 = 0$ and $\sigma(\mathcal{X}) = \sigma(x: x_s, s > 0)$. In this section we identify $\Omega$ with $\mathbb{R}^m$ and $F$ with $\sigma(\mathbb{R}^m)$ where $m$ is an integer. For $P$ a probability measure on $(\Omega, F)$ let $N = (N_t, F_t, A_t, P), t \geq 0$ denote a point process $(N_t, F_t), t \geq 0$ with predictable compensator $(A_t, F_t), t \geq 0$ relative to the filtration $F = (F_t), t \geq 0$ and the measure $P$.

The internal history of a point process $N$ is given by

$$F^N_t = F_0 \vee \sigma-(N_s, s \in (0,t]), t \geq 0$$

where $F_0$ contains the $P$-null sets and their subsets in $F$. According to Brémaud [12], T25: Appendix A2 the family $F^N = (F^N_t), t \geq 0$
satisfies the usual conditions. A point process \( N = (N_t, F_t^N, A_t, P), t \geq 0 \) with \( F_t^N \) the internal history defined by (1.1) is called self-exciting.

The next theorem characterizes the likelihood or density function of a self-excited Poisson-type counting process. In the theorem assume \( F_0 = (\emptyset, \Omega) \) and let \( F_t^N \) be defined by (1.1).

**Theorem 1.1.** Let \( E \) be a discrete space of "events" denoted generically by \( i \) and suppose \( \text{card}(E) = m < \infty \). Assume \( N = (N_t, F_t^N, A_t, P) \) \( t \geq 0 \) is a Poisson-type counting process (see section 2.2.1) such that the following conditions hold \( P \)-a.s.

1. \( A_t(i) = \int_0^t Y_s(i)db_i(s), \forall t \geq 0 \) where \( b_i \) is a Borel measure on \( (\mathbb{R}^+, \sigma(\mathbb{R}^+)) \) and \( Y(i) \) is \( F_t^N \)-predictable for all \( i \in E \).

2. \( \sum_{i \in E} Y_t(i) \Delta b_i(t) \leq 1 \) and \( \sum_{i \in E} \Delta b_i(t) \leq 1 \) \( \forall t \);

3. \( \Delta b_i(t) = 1 \Rightarrow Y_t(i) = 1 \) for any such \( t \);

4. \( \sum_{i \in E} \int_0^\infty \left(1 - \sqrt{Y_s(i)}\right)^2 db_i(s) + \sum_{s \leq T} \left(1 - \sqrt{\frac{1-\lambda_s}{1-\lambda}}\right) (1-\alpha_s) < \infty, \quad 0 < \alpha_s < 1 \)
   
   where \( \alpha_s = \sum_i \Delta b_i(s), \lambda_s = \sum_i Y_s(i) \Delta b_i(s) \).

Then \( P \ll P' \), where \( P' \) is a probability measure on \( (\Omega, F) \) such that \( N' = (N_t, F_t^N, b(t), P') t \geq 0 \) is a Poisson-type counting process (here \( b(t) = (b_1(t), \ldots, b_m(t)) \) and \( b_i(t) = b_i([0,t]) \) for \( i = 1, \ldots, m \)).
In this case

\[ P\{d\omega\} = Z_{\tau_\infty}(\omega)P'\{d\omega\} \]

where

\[ Z_{\tau_\infty}(\omega) = \exp\left\{ \int_{\tau_\infty}^{\tau_\infty} \ln Y_s(i) dN_s(i) + \int_0^{\tau_\infty} (1 - Y_s(i)) d\beta_i^c(s) \right\} \]

\[ + \sum_{s \leq t} (1 - \Delta N_s) \ln \frac{1 - \lambda}{1 - \alpha_s}. \]

Proof. Since \( F_0 = \{\emptyset, \Omega\} \) and \( F_t^N = F_0 \vee \sigma(N_s, s \in (0, t)) \) we let \( P'_0 \) denote a probability measure on \((\Omega, F_0)\) such that \( P'_0(\emptyset) = 0 \) and \( P'_0(\Omega) = 1 \). By Jacod [21], theorem 3.6 there exists a unique probability measure \( P' \) on \((\Omega, F)\) whose restriction to \( F_0 \) is \( P'_0 \) and for which \( N' = (N_t, F_t^N, b, P') \), \( t \geq 0 \) is a Poisson-type counting process (here \( Y \equiv 1 \) \( P' \)-a.s.). By (1), (2), (3) and Jacod [21], theorem 4.5 the theorem follows for this \( P' \) provided \( P'Z_{\tau_\infty} = 1 \) (\( P'X \) denotes the expectation of \( X \) relative to \( P' \)). This problem is considered in the next lemma.

**Lemma 2.1.** Suppose the situation in theorem 2.1 is in effect, then in the presence of \{1), (2) and (3)\} (4) is a necessary and sufficient condition for \( P \ll P' \) or equivalently \( P'Z_{\tau_\infty} = 1 \).

**Remarks:** 1) The condition \( P'Z_{\tau_\infty} = 1 \) was given in theorem 4.5, Jacod [21] as a requirement for absolute continuity of \( P \) relative to \( P' \). However, conditions for this to hold were not given and so lemma 2.1 covering the discrete case (i.e. \( E \) is discrete) is of independent interest.
2) The method of proof is an extension of a similar proof used in the univariate case (\text{card}(E) = 1) by Lipster and Shiryayev [26, Chapter 19, II].

**Proof of lemma 2.1:** Put $\alpha_t = \sum_{i \in E} \Delta b_i(t)$ and $\lambda_t = \sum_{i \in E} Y_t(i) \Delta b_i(t)$, where. Define the process $Z = (Z_t, F^N_t, t \geq 0$ by

\[
(1.2) \quad Z_t = \exp \left\{ \sum_{i \in E} \int_0^t \ln Y_s(i) dN_s(i) + \int_0^t (1 - Y_s(i)) db^c_s(s) \right\} \\
+ \sum_{s \leq t} (1 - N_s) \ln \frac{1 - \lambda_s}{1 - \alpha_s},
\]

where $\ln$ denotes the natural logarithm, $b^c$ denotes the continuous part of $b$ and $N = N(1) + \ldots + N(m)$. By comparing equation (1.2) with equation (14) [21] we have, by the proof of proposition 4.3, Jacod [21], that $Z$ is a $\tau_\infty$-local martingale relative to $P'$. Therefore there exist an increasing sequence of stopping times $(\sigma_{n^*})$, $n \geq 1$ such that $\sigma_n \uparrow \tau_\infty$ and $Z^n = (Z_{\sigma_n^*}^{t}, F_{\sigma_n^*}^N, t \geq 0$ is a uniformly integrable martingale for $n = 1, 2, \ldots$.

Let $P_n$ ($P'_n$) be the restriction of the measure $P$ ($P'$) to the sub-$\sigma$-algebra $F_{\sigma_n}^N$ for $n = 1, 2, \ldots$. Since $P'_n Z_{\sigma_n} = P'_n Z = P'_1 = 1$, we have by theorem 4.5, [21] that $P_n \ll P'_n$ for each $n$. Therefore, by lemma 19.13, Lipster and Shiryayev [26] it remains to show that the following holds:
\begin{align}
(1.3) \quad \lim_{{n \to \infty}} \frac{dP}{dP_n} &= \lim_{{n \to \infty}} Z_n \quad \text{exists and is finite (P-a.s.).}
\end{align}

In the argument above we can replace $\sigma_n$, $n = 1, 2, \ldots$ with the sequence $\sigma_n \wedge t$, $n = 1, 2, \ldots$, $t \to \tau_\infty$ and so without loss of generality we show

\begin{align}
(1.4) \quad \lim_{{t \uparrow \tau_\infty}} Z_t \quad \text{exists and is finite (P-a.s.).}
\end{align}

An examination of equation (1.2) shows that the limit in (1.4) is defined and so the lemma is equivalent to showing

\begin{align}
(1.5) \quad P(\lim_{{t \to \tau_\infty}} Z_t < \infty) = 1.
\end{align}

To prove (1.5) we define, as in equation (19.106) [26, p. 312], the function $u$ by

\begin{align}
(1.6) \quad u(y) &= \begin{cases} 
  y & |y| \leq 1 \\
  \text{sign}(y) & |y| > 1,
\end{cases}
\end{align}

and similarly define $Z^u$ by:

\begin{align}
(1.7) \quad Z^u_t &= \exp\{ \sum_{{i \in E}} \int_0^t u(\ln Y_s(i))dN_s(i) + \int_0^t (1 - Y_s(i))db^c_1(s) \\
&\quad + \sum_{s < t} (1 - \Delta N_s)u(\ln \frac{1 - s}{1 - \alpha_s}), \quad t \geq 0.
\end{align}

Now comparing (1.7) with (1.2) shows that the exponent of $Z^u_t$ converges
to something finite as $t \to \tau_\infty$ if and only if the exponent of $Z_t$ converges to something finite as $t \to \tau_\infty$. Therefore, 

$$\{\omega: 0 < \lim_{t \to \tau_\infty} Z_t^u < \infty\} = \{\omega: 0 < \lim_{t \to \tau_\infty} Z_t < \infty\}$$

and therefore (1.5) is equivalent to 

$$P \left( 0 < \lim_{t \to \tau_\infty} Z_t^u < \infty \right) = 1.$$ 

To prove (1.8) we reexpress $\ln Z_t^u$ as follows:

$$\ln Z_t^u = \sum_{i \in E} \left[ \int_0^t 1(\Delta b_i(s) = 0)u(\ln Y_s(i))dN_s(i) + \sum_{s \leq t} \Delta N_s(i)u(\ln \frac{\Delta A_s(i)}{\Delta b_i(s)}) + \int_0^t (1 - Y_s(i))db_i^c(s) \right]$$

$$+ \sum_{s \leq t} (1 - \Delta N_s)u(\ln \frac{1-\lambda_s}{1-\alpha_s})$$

$$= \sum_{i \in E} \left[ \int_0^t 1(\Delta b_i(s) = 0)u(\ln Y_s(i)(dN_s(i) - dA_s(i))) + \int_0^t (u(\ln Y_s(i))Y_s(i) + 1 - Y_s(i))db_i^c(s) \right]$$

$$+ \sum_{i \in E} \sum_{s \leq t} \Delta N_s(i)u(\ln \frac{\Delta A_s(i)}{\Delta b_i(s)}) + \sum_{s \leq t} (1 - \Delta N_s)u(\ln \frac{1-\lambda_s}{1-\alpha_s})$$

$$= m_t + D_t.$$
where

\[
(2.0) \quad m_t = \sum_{i \in E} \int_0^t 1(\Delta b_i(s) = 0)u(\ln Y_s(i)(dN_s(i) - dA_s(i) + \frac{\Delta A_s(i)}{\Delta b_i(s)})u(\ln \frac{\Delta A_s(i)}{\Delta b_i(s)})] + \sum_{s \leq t} (1 - \Delta N_s - (1 - \lambda_s))u(\ln \frac{1 - \lambda_s}{1 - \alpha_s}),
\]

and

\[
(2.1) \quad D_t = \sum_{i \in E} \int_0^t (u(\ln Y_s(i))Y_s(i) + 1 - Y_s(i))db_i^c(s) + \sum_{s \leq t} \Delta A_s(i)u(\ln \frac{\Delta A_s(i)}{\Delta b_i(s)})] + \sum_{s \leq t} (1 - Y_s)u(\ln \frac{1 - \lambda_s}{1 - \alpha_s}),
\]

where equations (1.9) - (2.1) holds for all \( t \geq 0 \).

Since \( Y \) is a predictable process and \( u \) is bounded it is easily verified, using a straightforward extension of proposition 5.3, Jacod [21] via theorem 18.8, Lipster and Shiryaev [26] that \( m = (m_t, F_t, p) t \geq 0 \) is a \( \tau_{\infty} \)-local square integrable martingale. By proposition 2.2.1 and other facts about martingales we have for \( t \geq 0 \)

\[
(2.2) \quad <m>_t = \sum_{i \in E} \int_0^t u^2(\ln Y_s(i))Y_s(i)db_i^c(s)
+ \sum_{i \in E} \sum_{s \leq t} u^2(\ln \frac{\Delta A_s(i)}{\Delta b_i(s)})(\frac{\Delta A_s(i)}{\Delta b_i(s)})\Delta b_i(s)
\]

\[
- \sum_{s \leq t} (\sum_{i \in E} u(\ln \frac{\Delta A_s(i)}{\Delta b_i(s)})(\frac{\Delta A_s(i)}{\Delta b_i(s)})\Delta b_i(s))^2 + \]

\[
\]
\[ + \sum_{s \leq t} u^2(\ln \frac{1-\lambda}{1-\alpha_s})(1 - \lambda_s)\lambda_s \]
\[ + 2(\sum_{s \leq t} u(\ln \frac{1-\lambda}{1-\alpha_s})[\sum_{i \in E} u(\ln \frac{\Delta A_s(i)}{\Delta b_i(s)})](1 - \Delta A_s(i))\Delta A_s(i) \]
\[ - \sum_{i \in E} \Delta A_s(i)(\sum_{j \neq i} u(\ln \frac{\Delta A_s(j)}{\Delta b_j(s)})\Delta A_s(j)))], \forall t. \]

The reader should verify that (2.3) agrees with (19.113), Lipster and Shiryaev [26], when \( \text{card}(E) = 1. \)

Secondly, we note that the process \( D_t \) is a nondecreasing predictable function of \( t \), since \( yu(\ln y) + 1 - y \geq 0 \) for \( y \geq 0 \) and

\[ (2.3) \quad \Delta D_s = \sum_{i \in E} \Delta A_s(i)u(\ln \frac{\Delta A_s(i)}{\Delta b_i(s)}) + (1 - \lambda_s)u(\ln \frac{1-\lambda}{1-\alpha_s}) \]
\[ \geq \sum_{i \in E} \Delta b_i(s)(\frac{\Delta A_s(i)}{\Delta b_i(s)} - 1) + (1 - \lambda_s)(\frac{1-\lambda}{1-\alpha_s} - 1) = 0, \forall s \geq 0. \]

By lemma 18.14, Lipster and Shiryaev [26], (1.8) is satisfied if

\[ (2.4) \quad P(<m>_{t_\infty} + D_{t_\infty} < \infty) = 1 \quad (<m>_t = \lim_{t \to \infty} <m>_t, D_t = \lim_{t \to \infty} D_t). \]

By some algebraic manipulations we can write (2.2) equivalently as

\[ (2.5) \quad <m>_t = \sum_{i \in E} \int_0^t u^2(\ln Y_s(i))Y_s(i)db_i(s) \]
\[ + \sum_{i \in E} \sum_{s \leq t} u^2(\ln \frac{\Delta A_s(i)}{\Delta b_i(s)})\Delta A_s(i) + \sum_{s \leq t} \left( u^2(\ln \frac{1-\lambda_s}{1-\alpha_s})^2 \right)_{0 < \Delta b_i(s) < 1} \]
\[ - (\Delta D_s)^2, \]
where $\Delta D_s = \sum_{i \in \mathcal{E}} (\Delta A_s(i)u(\ln \frac{\Delta A_s(i)}{\Delta b_1(s)}) + (1-\lambda_s)u(\ln \frac{1-\lambda_s}{1-\alpha_s})$.

Now, from (2.1), (2.5) and using the fact that $\Delta D_s \geq 0$ we have that (2.4) is equivalent to

\[(2.6)\quad P(\sum_{i \in \mathcal{E}} \int_0^{\tau_{\infty}} g(Y_s(i))db_i^c(s) + \sum_{s \leq \tau_{\infty}} \frac{\Delta A_s(i)}{\Delta b_1(s)}\Delta b_1(s) \quad 0 < \Delta b_1(s) < 1) \]

\[+ \sum_{s \leq \tau_{\infty}} g\left(\frac{1-\lambda_s}{1-\alpha_s}\right)(1-\alpha_s) < \infty) = 1, \]

where $g(x) = xu(\ln x) + 1 - x$. To see the equivalence of (2.6) and (2.4) one needs only faithfully recall the definition of the function $u$ in equation (1.6). Now as in equation (19.117), Lipster and Shiryayev [26], there are constants $c$ and $C$ such that

\[(2.7)\quad c(1 - x^{1/2})^2 \leq g(x) \leq C(1 - x^{1/2})^2.\]

Hence (4) is equivalent to (2.6). Therefore, if (4) holds then (2.4) holds and $P \ll P'$; i.e. $P'(Z_{\tau_{\infty}} = 1).$ Conversely, if $P \ll P'$ then by the argument given on page 314, Lipster and Shiryayev [26] we have

\[(2.8)\quad P(\langle m \rangle_{\tau_{\infty}} + D_{\tau_{\infty}} < \infty) = 1\]

which implies (2.6) and the desired relation (4). Therefore, this concludes the proof of theorem 1.1. □
We consider some simple sufficient conditions for theorem 1.1 to hold in the following corollary:

**Corollary 1.1.** Suppose \( N = (N_t(i), F^N_t, A_t(i), i \in E, P), t \geq 0 \) is a Poisson-type counting process such that (1), (2) and (3) of theorem 1.1 hold (P-a.s.). In addition suppose either

(1) \( N \) is observed over the finite interval \([0,T]\) such that

(i) \[ \sum_{i \in E} b_i(T) < \infty \]

(ii) \[ \sum_{i \in E} N_T(i) < \infty \) (P-a.s.) so that \( \tau_\infty \wedge T = T \) (P-a.s.),

or

(2) \( \exists \) a constant \( C \) such that

(i) \[ \sum_{i \in E} \int_0^t |Y_i(s) + \frac{\lambda_i - 1}{1 - \alpha_i} db_i(s) | ds \leq C \) (P'-a.s.).

**Proof.** For (1) consider the stopped process by way of the sequence \( (\tau_n \wedge T), n \geq 1 \) and observe that by 1(i) and 1(ii) and an application of Lenglart's inequality (see for example lemma 1 [27]) it is easily shown that (4) of theorem 1.1 holds.

For (2) we observe that \( Z \), defined by (1.2) has the representation

(2.7) \[ Z_t = 1 + \sum_{i \in E} \int_0^t Z_{s-}(Y_i(s) + \frac{\lambda_i - 1}{1 - \alpha_i}) (dN_i(s) - db_i(s)), t \geq 0, \]

(see the proof of proposition 4.3, Jacod [21]). Clearly \( f_s(i) = \frac{\lambda_i - 1}{1 - \alpha_i} Z_{s-}(Y_i(s) + \frac{\lambda_i - 1}{1 - \alpha_i})(dN_i(s) - db_i(s)) \) is a predictable process and so by theorem 18.7,
Lipster and Shiryaev [26] it suffices to show

\[
(2.8) \quad P'[\int_0^{\tau_\infty} Z_{s-} Y_s(i) + \frac{\lambda s - 1}{1 - \alpha_s} |db_i(s)| < \infty \text{ for each } i \in E,
\]

where \( P'[\cdot] \) denotes expectation of [\( \cdot \)] relative to the measure \( P' \).

To show (2.8) holds for each \( i \in E \) we observe that

\[
H^i_t = \int_0^t |Y_s(i) + \frac{\lambda s - 1}{1 - \alpha_s} db_i(s), \quad t \geq 0 \text{ is a natural increasing process,}
\]

since it is predictable (see for example Dellacherie [16]). Also, recall that \( Z = (Z_t, F^N_t, P') \), \( t \geq 0 \) is a \( \tau_\infty \)-local martingale so there exists stopping times \( (\sigma_n) \), \( n \geq 1 \) such that \( \sigma_n \uparrow \tau_\infty (P'\text{-a.s.}) \) and

\[
Z^n = (Z_{t \wedge \tau_n}, F^n_t, P'), \quad t \geq 0 \text{ is a martingale with } P'Z_{\sigma_n} = 1 \text{ for } n = 1, 2, \ldots.
\]

Therefore, by lemma 3.2, Lipster and Shiryaev [26] and the monotone convergence theorem we have for each \( i \)

\[
(2.9) \quad P'\int_0^{\tau_\infty} Z_{s-} Y_s(i) + \frac{\lambda s - 1}{1 - \alpha_s} |db_i(s)
\]

\[
= \lim_{n \to \infty} P'\int_0^{\sigma_n} Z_{s-} Y_s(i) + \frac{\lambda s - 1}{1 - \alpha_s} |db_i(s)
\]

\[
= \lim_{n \to \infty} P'\int_0^{\sigma_n} |Y_s(i) + \frac{\lambda s - 1}{1 - \alpha_s} |db_i(s)
\]

\[
\leq C \lim_{n \to \infty} P'Z_{\sigma_n} < \infty \text{ for all } i \in E.
\]
Therefore, equation (2.7) is 1 plus the sum of a finite number of martingales with zero mean and so $Z = (Z_t, t \geq 0)$ is a martingale with mean one, which implies $P_Z = 1$ as required by theorem 1.1.

Remarks: 1) Theorem 1.1 may be used as a basis for the development of a theory of sufficiency and completeness for subclasses of Poisson-type counting processes. To do this we require for each subclass a suitable measure, $P'$ say, which dominates the entire subclass. Theorem 1.1 shows that each respective $P'$ is uniquely determined by its infinitesimal characteristics (i.e. the choice of compensator $b = (b_1, \ldots, b_m)$) and that the compensator associated to $P'$ must dominate the entire subclass of associated compensators as well. As an example, assume $b \ll dt$ where $dt$ denote Lebesgue measure. We recognize this to be the subclass of multiplicative intensity models considered by Aalen [3]. In this case we may take $P'$ to be a poissonian measure with constant rate $Y \equiv 1$. According to Aalen [3, section 5] we can develop a theory of sufficiency and completeness for this subclass.
CHAPTER 5

Directions for Future Research

Our interest in applications has revealed the relevance of a two component stochastic process \( \{N_t, Y_t\} \) arising in a number of areas of research where \( N_t \) records the frequency of occurrences of events of different types and \( Y_t \) acts as a driving force on the rates of occurrences of those events. Several examples of this were described in Chapter 1 of this thesis. There is a need to consider probability models for the class \((N,Y)\) where added attention is given to modeling the auxiliary process \( Y \) and thus to characterizing its impact on the point process generating \( N \). Short of this endeavor we may extend Poisson-type counting processes to a more informative class of auxiliary processes.

The Poisson-type counting process of an event of a single type assumes that the compensator \( A \) satisfies

\[
(1.1) \quad A(B) = \int_B Y_s \, db(s) \text{ for } B \text{ a Borel set in } \mathbb{R}^+,
\]

where \( Y \) is a nonnegative predictable process and \( b \) a Borel measure on \( \mathbb{R}^+ \). Let \( \Theta \) be a subset of \( \mathbb{R}^d \) (\( d \) an integer) and \( X = (X_t), t \geq 0 \) be an arbitrary \( d \)-dimensional predictable process and suppose \( g \) is a non-negative \( b \)-measurable function defined on \( \mathbb{R}^{2d} \). For each \( \theta \in \Theta \) the random measure \( A^\theta \) defined by

121
\[(1.2) \quad A^\theta \{B\} = \int_B g(\theta, X_s) \, db(s), \quad B \in \sigma(\mathbb{R}^+)\]

defines a compensator of the Poisson-type where \(Y\) is identified with \(g(\theta, X)\). Assume \(g\) is a known function parameterized by \(\theta \in \Theta\) whereby the impact on the point process of the components of \(X\) is given a functional form. For each \(\theta \in \Theta\) suppose \(N\) is a Poisson-type counting process with compensator \(A^\theta\) and \((N, X)\) are observable processes whereas \((\theta, b)\) are unknown parameters. We call (1.2) a semiparametric model to reflect the added parameterization of the auxiliary process and the statistical problem is to estimate \(\theta\) and \(b\) from observation of \((N, X)\) over a period of time. This problem indicates the first direction of our further research into Poisson-type counting processes and is an extension of the work of Andersen and Gill [5] and Prentice and Self [30] who assume \(\partial^2 g / \partial \theta^2\) exists and is continuous and that \(b \ll dt\).

In some problems the Borel measure \(b\) has a known form depending on some parameter \(\theta\) which is unknown and we are required to estimate \(\theta\) parametrically. Let \(\Theta\) be a subset in some Euclidean space and let \(P = \{b^\theta : \theta \in \Theta\}\) denote a parametric family of Borel measures. For each \(\theta \in \Theta\) suppose \(N\) is a Poisson-type counting process with compensator \(A^\theta\) given by

\[(1.3) \quad A^\theta \{B\} = \int_B Y_s^\theta \, db(s), \quad B \in \sigma(\mathbb{R}^+).\]

Assume that \(N\) is a self-excited counting process and that the density for
the process exists and is given by theorem 4.1.1. This gives the likelihood function for $\theta$ and in a second research direction we consider likelihood based inference procedures for $\theta$ in parametric families $P$ which extends the maximum likelihood theory of Brogan [11] who assumes that the intensity exists.

A counting process approach has been applied by Gill [19] to the problem of estimating transition distributions for a general Markov renewal process, although he does not employ counting processes of the Poisson-type. However, Voelkel and Crowley [37] who assume the intensity exists use these counting processes after further assuming that the corresponding semi-Markov process is "forward-going" (more suggestively thought of as an ascending ladder process). We find these assumptions unnecessary and propose what we call a local-time approach which allows us to express a general Markov renewal process as a Poisson-type counting process. This approach naturally incorporates the situation where a single process is observed over an expanding sequence of time intervals and so may be used to solve an estimation problem posed by Gill [19] for this case.
APPENDIX

Review of Semimartingales

In this thesis we work with families of stochastic processes called martingales and their local counterparts, processes of independent increments and processes of local bounded variation all of which give examples of a semimartingale. Since many of these processes may be unfamiliar to the reader we begin with the idea of a martingale and build our way toward a general definition of a semimartingale.

Let \((\Omega, F, \mathbb{P})\) denote a probability space and \(F = (F_t), t \geq 0\) a given family of sub-\(\sigma\)-algebras of \(F\) such that \(F\) satisfies the usual conditions (for each \(t \in \mathbb{R}^+\), \(F_t = \bigcap_{t+\varepsilon > 0} F_{t+\varepsilon}\), for all \(s \leq t\), \(F_s \subset F_t\) and \(F\) is complete relative to \(\mathbb{P}\)).

Martingales

We assume the reader is familiar with the notion of a martingale but may not be familiar with their local counterparts; the local martingale. Suppose \(M = (M_t), t \geq 0\) is an \(F\)-adapted (i.e. for each \(t, M_t\) is \(F_t\)-measurable) stochastic process and there exists a monotone sequence \((\tau_n), n \geq 1\) of \(F\)-stopping times such that the stopped process \(M^n\) defined by

\[
(1.1) \quad M^n_t = M^{t \wedge \tau_n}, \quad t \geq 0
\]

124
is a uniformly integrable martingale, then $M$ is called a local martingale and if $\tau_\infty = \lim_{n \to \infty} \tau_n$ we also call $M$ a $\tau_\infty$-local martingale. Clearly, if $M$ is a uniformly integrable martingale then the optional sampling theorem implies that $M$ is a local martingale as well. Another example occurs when $M = N - A$ where $N$ is a counting process and $A$ is its compensator; see section 2.2.1. Often, we need the concept of a square integrable martingale and with obvious definition a locally square integrable local martingale. The sequence $(\tau_n)$, $n \geq 1$ which gives rise to a martingale or square integrable martingale is called a localizing sequence. In the counting process example the sequence of event epochs (i.e. stopping times $T_n$ such that $N_{T_n} = n$ and $\Delta N_{T_n} = 1$) forms a localizing sequence from which it clearly follows that $M = N - A$ is a locally square integrable local martingale as well; see Lipster and Shiryayev [26, Chapter 19].

Let $X$ be a stochastic process so that for each $\omega \in \Omega$ $X(\omega)$ is a right continuous function of $t \in \mathbb{R}^+$ with left hand limits and denote $\Delta X_s = X_s - X_{s-}$ ($X_{s-} = \lim_{t \uparrow s} X_t$). For $A$ a Borel set in $\mathbb{R}/\{0\}$ define a random counting measure $\mu$ on $(\mathbb{R}^+ \times \mathbb{R}, \sigma(\mathbb{R}^+ \times \mathbb{R}))$ by

$$
(1.2) \quad \mu([0,t] \times A) = \sum_{s \leq t} 1(\Delta X_s \in A), \ t \geq 0.
$$

According to Jacod [21] there exists a unique predictable measure $\nu$ on $\mathbb{R}^+ \times \mathbb{R}$ called the compensator of $\mu$ such that $\mu - \nu$ is a random martingale measure on $(\mathbb{R}^+ \times \mathbb{R}/\{0\}, \sigma(\mathbb{R}^+ \times \mathbb{R}/\{0\}))$. Thus for $A$ a Borel set in $\mathbb{R}/\{0\}$
(1.3) \( \mu([0,t] \times A) - \nu([0,t] \times A), \ t \geq 0 \)

defines a local martingale.

It turns out that any local martingale \( M \) admits a decomposition
\( M = M^c + M^d \) where \( M^c \) is a continuous local martingale and \( M^d \) is called
a purely discontinuous local martingale; see Shiryayev [34]. The local
martingale \( M^d \) has the property that if \( Y \) is an arbitrary continuous
martingale then \( M^d \perp Y \) in the sense that \( EM^d_t Y_t = 0 \) for \( \tau \) an arbitrary
stopping time (here \( \perp \) denotes orthogonality). In addition \( M^d \) has the
following representation

\[
(1.4) \quad M^d_t = \int_0^t \int x(\mu - \nu)(ds, dx), \ t \geq 0
\]

where the pair \((\mu, \nu)\) are constructed relative to the jumps in \( M \).

In the decomposition of a local martingale \( M \), the continuous
part \( M^c \) is also locally square integrable and therefore we can define
a unique predictable process \(<M^c>\) (called the quadratic characteristic
of \( M^c \)) such that

\[
(1.5) \quad (M^c)^2 - <M^c>
\]

is a local martingale. For example, if \( W \) denotes an ordinary Weiner
process then \( W \) is a continuous local martingale and it is well known
that \( W^2 - t \) is a local martingale, thus \(<W>\) is identified with \( t \).

When \( M \) is an arbitrary locally square integrable local martingale
its quadratic characteristic \(<M>\) is shown to satisfy
(1.6) \( <M>_t = <M^c>_t + \int_0^t \int \nu(ds, dx), \; t \geq 0. \)

In general, however, \( M \) may not be square integrable in which case we define the quadratic variation \([M] \) for \( M \) as follows. For each \( t \in \mathbb{R} \) let

\[
(1.7) \; [M]_t = <M^c>_t + \sum_{s \leq t} (\Delta M_s)^2
\]

where \( M^c \) denotes the continuous part of \( M \). The process \([M] \) earns its name from the fact that for all \( t \in \mathbb{R}^+ \)

\[
(1.8) \lim_{k \to \infty} \sum_{k=0}^{k-1} (M_{t(k+1)/n} - M_{t(k)/n})^2 \overset{p}{\to} [M]_t.
\]

One virtue of the quadratic variation is that it is always defined whereas the quadratic characteristic may not be.

In the theory of stochastic integration with respect to local martingales the smoothness properties of the quadratic characteristic and the quadratic variation are used to determine the class of integrands for which such an integral is defined. However, in this thesis we mainly employ these processes to establish conditions guaranteeing the convergence in law of a sequence \( M^n \) of local martingales to a limit process \( M^\infty \) of independent increments.

**Processes of independent increments**

In statistical applications of probability theory the central limit theorem holds a special place. It is used for example in
estimation theory to show that certain classes of maximum likelihood estimators are asymptotically normally distributed. In most of these problems we are given a sequence \( \{ \xi_n \} \) of random variables and set out to show that as \( n \) tends toward infinity \( \xi_n \) tends in law or distribution to a Gaussian random variable or more generally a random variable with infinitely divisible law. In contrast, sometimes the demands of a problem require that we characterize the limiting law of sequences \( \{ (\xi_t^n) \}, \ t \geq 0 \) of stochastic processes as in the main problem of this thesis where \( (\xi_t^n), \ t \geq 0 \) is an error of estimation process which arises when estimating the compensator \( A \) of a counting process. In problems such as this functional central limit theorems have proved most useful in characterizing the limit law (see for example Aalen [3] and Chapter 3 of this sequel).

In general, functional central limit theorems are concerned with conditions under which sequences of stochastic processes converge weakly to another stochastic process called the limit process. A large body of this theory takes the limit process to be a process having independent increments; see [2], [20], [22], [27], and [32] for example.

Let \( X = (X_t), \ t \geq 0 \) be a stochastic process with independent increments so that by definition for all \( 0 < t_0 < t_1 < \ldots < t_k \) the random variables \( X_{t_0}, X_{t_1} - X_{t_0}, \ldots, X_{t_k} - X_{t_{k-1}} \) are mutually independent. Well known examples occur when \( X = W \) for \( W \) a Weiner process and when \( X \) is a compound Poisson process. These two processes
are fundamentally important in that every **stochastically continuous**
process of independent increments can be represented as a sum of a
Weiner process and an integral of Poisson processes where the role
played by the Poisson process is to characterize the jumps of the
process $X$.

Perhaps the greatest can be learned about processes of
independent increments by considering the general form of their
characteristic function. In general the characteristic function $\phi_t$
of $X_t$ at the point $\theta \in \mathbb{R}$ is given by:

$$
(1.8) \quad \phi_t(\theta) = E e^{i\theta X_t} \\
= \exp\left[i\theta B_t - \frac{1}{2} \theta^2 C_t + \int_{|x|>1} (e^{i\theta x} - 1 - i\theta x) \nu([0,t] \times dx) \right] \\
+ \int_{|x| \leq 1} (e^{i\theta x} - 1) \nu([0,t] \times dx),
$$

where $(B,C,\nu)$ are **deterministic** functions called the local characteris-
tics of $X$; $B$ is a function having bounded variation on finite
intervals, $C$ is a continuous function and $\nu$ is a Lévy measure satisfying

$$
\int_{\mathbb{R}} \min(1, x^2) \nu([0,t] \times dx) < \infty.
$$

Equation (1.8) gives the so-called Lévy-Khintchine representation of $\phi$.

**Example 1.** $B_t = \int_{|x| \leq 1} x \nu([0,t] \times dx)$, $C_t = 0 \ \forall t$ and
$\nu(dt, dx) = \lambda dt F(dx)$ for $\lambda > 0$ and $F$ a distribution measure gives the
characteristic function of compound Poisson process. $\blacksquare$
Example 2. Put \( B = 0, \nu = 0 \) and \( C_t = t \) giving the characteristic function of a Weiner process. \( \square \)

One of the more interesting aspects of the theory of processes of independent increments is concerned with representations for and properties of the sample functions of these processes. A most important result in this direction is a decomposition due to Lévy discussed by Doob [17, VIII.7 (6.5)] which we proceed to reconstruct.

Our aim is to build a process \( X \) of independent increments where for convenience we require \( X_0 = 0 \) P-a.s. and \( X \) is P-a.s. right continuous at all points \( t \in \mathbb{R}^+ \). In general \( X \) may have fixed points of discontinuity and we address this issue first. Let \( \{t_j\}, j \geq 1 \) be a countable set of time points and let \( \{U_j\}, j \geq 1 \) be a sequence of mutually independent random variables such that for any finite interval \( I \subset \mathbb{R}^+ \) \( \sum_{t_j \in I} U_j \) converges with probability 1 when centered. As in Doob [17, III, §2] we may choose to center the \( U_j \) by subtracting truncated expectations (e.g. \( \mathbb{E}[U_j 1(|U_j| \leq 1)] \)). Put \( \{U'_j\} \) equal to the centered \( \{U_j\} \) and consider the process \( X^d = (X^d_t), t \geq 0 \) defined by

\[
(1.9) \quad X^d_t = \sum_{0 < t_j < t} U'_j, \quad t \geq 0.
\]

Now according to Doob [17, Theorem 6.4], \( X^d \) is a right continuous and centered process of independent increments having almost all sample paths continuous except perhaps at the fixed points of discontinuity \( \{t_j\} \) with \( X^d_{t_j} - X^d_{t_j^-} = U'_j \). Next, consider a second process
\( X^c = (X^c_t), \ t \geq 0 \) such that \( X^c \) is a centered process of independent increments having no fixed points of discontinuity and assume \( X^c \) and \( X^d \) are mutually independent.

Suppose \( X \) is a process of independent increments and let \( f \) be a centering function for \( X \), then according to Doob [17, p. 417] \( X \) admits a decomposition (due to Lévy) given by:

\[
(2.0) \quad X_t = f(t) + X^c_t + X^d_t, \ t \geq 0
\]

where \( X^c \) and \( X^d \) are defined above. According to Doob [17, Theorem 7.2], the centered process \( X-f \) is bounded on every finite interval \( I \subset \mathbb{R}^+ \), has finite left and right hand limits and its discontinuities are jumps (except perhaps at the fixed points of discontinuity; although we have constructed \( X^d \) so that \( X \) "jumps" at the \( \{t_j\} \) as well).

Before closing this section we give one more representation of \( X \) and relate its structure to the characteristic function of \( X \) given in (1.8) and thus bring to light the significance of the triplet of local characteristics \((\mu, \nu, \gamma)\). First, we recall that \( X^c \) is a stochastically continuous process of independent increments and according to Gikhman and Skorokhod [172, VI, §3, Theorem 1], \( X^c \) admits the representation

\[
(2.0a) \quad X^c_t = \xi_t + \int_0^t \int_{|x| > 1} X^c(ds, dx) + \int_0^t \int_{|x| \leq 1} x(\mu^c - \nu^c)(ds, dx), \quad t \geq 0
\]

where the pair \((\mu^c, \nu^c)\) are constructed from the jumps of the process \( X^c \) and \( \xi = (\xi_t), \ t \geq 0 \) is a continuous Gaussian process of independent
increments where $\xi$ is independent of the counting measure $\mu^c$. This is extremely interesting if we similarly define $(\mu^d, \nu^d)$ with respect to the process $X^d$ so that for $j = 1, 2, \ldots$ $\mu^d(t_j \times \cdot) = 1(U_j \epsilon \cdot)$ and $\nu^d(t_j \times \cdot)$ is the distribution measure of $U_j$. Since we have used 

$$m_j = \int_{|x| \leq 1} x \nu^d(t_j \times dx)$$

as centering constants for the $\{U_j\}$ we may now write $X$ in the following form:

$$X_t = f(t) + \xi_t + \int_0^t \int_{|x| > 1} x \mu(ds, dx) + \int_0^t \int_{|x| \leq 1} x(\mu - \nu)(ds, dx), \quad t \geq 0$$

(2.1)

where $\mu = \mu^c + \mu^d$ and $\nu = \nu^c + \nu^d$. A most interesting aspect of the representation (2.1) is that $\nu$, which is the measure constructed from the jumps of $X$ is the Lévy measure appearing in the characteristic function for $X$ defined by (1.8). Further, since $\xi$ is a continuous Gaussian process of independent increments with zero-mean it is also a martingale and we may define its quadratic characteristic $<\xi>$ where $<\xi>_t = C_t$, $\forall t \geq 0$ for $C$ of (1.8). Thus, we have an elegant correspondence between the representation of $X$ in (2.1) and its characteristic function $\phi$ in (1.8) where the parameter $B$ may be identified with the function $f$. Moreover, it is seen by (2.1) and the discussion above that if $f(t) = -\int_0^t \int_{|x| > 1} x \nu(ds, dx)$ then $X$ is a local martingale as well.

**Semimartingales**

In the general theory of stochastic processes $X = (X_t)$, $t \geq 0$ is called a semimartingale if $X$ may be represented by:
(2.2) \[ X_t = \tilde{B}_t + X^c_t + X^d_t + \int_0^t \int_{|x|>1} x\tilde{\mu}(ds, dx), \forall t \geq 0 \]

where \( \tilde{B} \) is a process of local bounded variation and \( X^c \) is a continuous
local martingale (called the continuous martingale part of \( X \)) and
\( X^d \) is a purely discontinuous local martingale of the form

(2.3) \[ X^d_t = \int_0^t \int_{|x|<1} x(\tilde{\mu}-\tilde{\nu})(ds, dx), \ t \geq 0 \]

where the pair \( (\tilde{\mu}, \tilde{\nu}) \) are constructed from the jumps of \( X \). Recall
that we can define a continuous process \( \tilde{C}_t = \langle X^c \rangle_t \) called the quadratic
characteristic of \( X^c \). The triplet \( (\tilde{B}, \tilde{C}, \tilde{\nu}) \) is called the triplet of
local characteristics for \( X \) and are in general random functions or
random measures. It is interesting to note that \( X \) is a process
of independent increments if and only if the triplet \( (\tilde{B}, \tilde{C}, \tilde{\nu}) \) is
deterministic as might have been suspected by comparison of (2.2)
with equation (2.1). Also observe that when \( \tilde{B}_t = -\int_0^t \int_{|x|>1} x\tilde{\nu}(ds, dx) \)
then \( X \) is a local martingale and so the family of semimartingales
includes the family of local martingales.

Now suppose we are given a sequence \( (X^n) \), \( n \geq 0 \) of semimartingales
where for each \( n \) \( X^n \) admits the decomposition shown in (2.2)
with corresponding local characteristics \( (\tilde{B}^n, \tilde{C}^n, \tilde{\nu}^n) \). Then it is
natural to believe that \( X^n \) will converge in law to a limit process
\( X^\infty \) of independent increments provided that the triplet \( (\tilde{B}^\infty, \tilde{C}^\infty, \tilde{\nu}^\infty) \)
converges (in a sense that can be made precise) to a deterministic
triplet \( (\tilde{B}^\infty, \tilde{C}^\infty, \tilde{\nu}^\infty) \) of local characteristics for \( X^\infty \). This last remark,
though informal, can be made precise and captures the essence of the theory of weak convergence used in this sequel. Our hope is that the discussion of the last three sections provides enough of a basis for an intuitive interpretation of the conditions involved in our theorems.
References


