The paper titled, *History of Application of Martingales in Survival Analysis*, provides a nice account of the various scientists, mathematicians, events and concepts that lead to the usage of martingales in Survival analysis. In this document, I will mention some of the main points of the paper.

Introduction

Time line of events

- **1662** : John Graunt came up with the first life table.
- **1958** : The paper by *Kaplan-Meier* was a major advance in the field of survival analysis.The KM curve is a nonparametric estimator of survival curve. The paper is one of the most cited papers in the history of statistics with more than 33,000 citations. The major difference between life table method and KM method is that the latter considered intervals of infinitesimal length. The method was also in response to the sparse clinical data that needed a statistical tool for inference.
- **1960 - 1970** : A majority of tests blossomed to compare survival curves. There was no coherent structure to all these tests that were being proposed.
- **1972** : Introduction of proportional hazards model by David Cox. The basic idea of the model was to adjust the survival curves for various covariates. This was a major advance in the field of survival analysis.

Given this background of KM estimator and Cox PH model, one might expect that it would find many applications. It did so. However theory lagged behind as there were many questions that were yet to answered, such as

- Why does the Cox model work ?
- How should one understand the plethora of tests ?
- What are the asymptotic properties of KM estimator ?
- What is meant by partial likelihood estimator ?
- How does one find the asymptotic properties of PL estimator ?

The authors of the paper credit themselves for the pioneering work in this area for the last 35 years. The main mathematical object that the authors applied in solving the above questions was **martingale**. Most statistics students are routinely taught the concepts of normal distributions, Markov processes, etc. But the concept of martingales is somehow avoided. The first time I came across **martingales** was in option pricing theory. Even though I had known and worked with time series, I had never taken effort to understand martingale objects that are pivotal to *time series analysis*. I think this paper's main intention is to demystify the term martingale by showing its varied applications to a specific domain, i.e. *Survival analysis*.

The hazard rate and a martingale estimator

Survival analysis essentially entails estimation of following functions:

1. Survival function

$$S(t) = P(T > t), \text{where} T \text{is the survival time}$$

2. Hazard function

$$\alpha(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} P(t \leq T \leq T + \Delta t | T \geq t)$$

3. Cumulative Hazard function

$$A(t) = \int_0^t \alpha(s)ds$$

Out of the three functions, it is simple to estimate survival curve and cumulative hazard rate, the estimation of hazard rate function is more difficult as it can be any arbitrary function of time.

Time line of events

- **1969, 1972** : A non-parametric estimator of $A(t)$ was first suggested by Nelson as a graphics tool
- **1972** : Altshuler suggested the same estimator as Nelson's
- **1972** : Aalen in his masters thesis suggested the same estimator for $A(t)$
- **1970's** : Rich collaboration between Norwegian statisticians and Berkley statisticians.
- **1975** : Aalen's PhD thesis introduces martingales for the first time. Aalen was influenced by his master thesis supervisor Jan M. Hoem who emphasized the importance of continuous-time Markov chains as a tool in the analysis when several events may occur to each individual. Aalen took the idea of CTMC and came up with a generic multiplicative model for rate function

  $$\lambda(t) = \alpha(t)Y(t)$$

  $\Rightarrow$

  $$dN(t) \approx \lambda(t)dt = \alpha(t)Y(t)dt$$

  $\Rightarrow$ (Nelson - Aalen estimator)

  $$\hat{A}(t) = \int_0^t \frac{dN(s)}{Y(s)}$$

  Even though CTMC was used, thanks to Aalen's advisor, the general formulation has nothing to do Markov property. The counting processes theory was not yet published by 1973. However, based on the suggestion from David Brillinger, Aalen referred to some papers and was immediately hooked. He knew that martingales were the right mathematical objects for understanding Survival analysis.

The basic idea of using martingales comes from the following equation:

$$E(dN(t) - \lambda(t)dt|past) = 0$$

By defining a martingale process as

$$M(t) = N(t) - \int_0^t \lambda(s)\, ds$$

one can see that the intensity process is equivalent to asserting that the counting process minus the integrated intensity process is a martingale.

Recollecting my Aha moment

For a long time I never had any intuition about martingales until one day when I realized that they are basically a richer version of noise and error term. If you look at any observation, a statistics 101 model is

$$X_{obs} = \mu_{true} + \epsilon, \quad \epsilon \sim N(0, \sigma)$$

where X_{obs} is the observed value, μ_{true} is the true value and ϵ is the error term associated with the observation. My *aha* moment was when I realized that one can write a stochastic process as a combination of a process dependent on the past and an innovative term, i.e. a zero-mean martingale.

The Nelson-Aalen estimator can be derived by using a multiplicative intensity model

$$dN(t) = \alpha(t)Y(t)dt + dM(t)$$

$\Rightarrow$

$$\hat{A}(t) = A(t) + \int_0^t \frac{dM(s)}{Y(s)}$$

One can transform the Nelson-Aalen estimator into an estimator of $S(t)$

$$S(t) = \prod_{(0,t]} \{1 - dA(s)\} = \exp\left(-\int_0^t \alpha(s)ds\right)$$

The following plug-in estimator is the Kaplan-meier estimator.

$$S(t) = \prod_{(0,t]} \{1 - d\hat{A}(s)\}$$

It is a finite product of the factors $1 - 1/Y(t_j), \forall t_k \leq t$. There is also a basic martingale representation is available for the Kaplan-Meier estimator as follows

$$\frac{\hat{S}(t)}{S(t)} - 1 = -\int_0^t \frac{\hat{S}(s-)}{S(s)Y(s)} dM(s)$$

where the right-hand side is a stochastic integral of a predictable process with respect to a zero-mean martingale, that is, itself a martingale.

Stochastic Integration and Statistical Estimation

The authors give a crash course on *Doob-Meyer decomposition*. It basically states that any submartingale can be decomposed into the sum of a martingale and a predictable process, which is often denoted as *compensator*. In the case of counting process, the following equation can be viewed from *Doob-Meyer decomposition* perspective

$$N(t) = M(t) + \int_0^t \lambda(s)ds$$

Many estimators can be written as a stochastic integral and hence to get to asymptotic properties of estimators, one needs to work with *variation processes* of martingales. In particular, there are two types of processes

1. predictable variation process $\langle M \rangle$

$$\langle M \rangle = \lim_{n\to\infty} \sum_{k=1}^{n} \text{Var}(\Delta M_k | \mathcal{F}_{(k-1)/n})$$

2. optional variation process $[M]$

$$[M] = \lim_{n\to\infty} \sum_{k=1}^{n} (\Delta M_k)^2$$

Two central results from the general theory of stochastic integration are important in the context of survival analysis. They are

1. A stochastic integral $\int_0^t H(s)dM(s)$ of a predictable process $H(t)$ with respect to a martingale $M(t)$ is itself a martingale
2. The variation processes satisfy

$$\langle HdM \rangle = \int H^2 dM$$
$$[HdM] = \int H^2 dM$$

The above formulae are used to derive variance formulas for estimators and tests in survival and event history analysis.

Stopping times, unbiasedness and independent censoring

The concept of *unbiasedness* pervades the field of classical statistics. One can intuitively connect the concept of unbiasedness with a fair game, i.e. martingale. The fundamental trait of survival data is censoring. If you assume certain censoring schemes, then it can be shown that martingale properties hold good for stopped random variables. The first time I learnt about the importance of stopping times was in Shreve's book, in which american options are priced using stopped martingales. Aalen's argument assumes that the censoring takes place in such a way that the intensity process with respect to the sigma algebra generated by the uncensored counting process is the same as intensity process with respect to sigma algebra generated by the censored counting process.

In many statistical models, there is an intrinsic assumption of independence between outcome variables. In the event history analysis, such an assumption may well be reasonable for the basic, uncensored observations, censoring may destroy the independence. In such cases, martingale properties are still preserved under the independent censoring assumption. This suggests that, for event history data, the counting process and martingale framework is, indeed, the natural framework and that the martingale property replaces the traditional independence assumption.

Martingale central limit theorems

The basic reason for the assumption of independence in statistics is to get some asymptotic distributional results of use in estimation and testing, and the martingale assumption can fulfill this need in the context of Stochastic processes. Central limit theorems for martingales can be traced back to the beginning of 1970's.

Time line of events

- **1974** : McLeish used the theory of counting processes in the discrete case
- **1977**: The first central limit theorem for continuous time martingales was published in Aalen

The central limit theorem for martingales is related to the fact that a martingale with continuous sample paths and a deterministic predictable variation process is a Gaussian martingale, i.e., with normal finite-dimensional distributions.

Where is the central limit theorem used? The difference between estimator and the true value can be represented as a stochastic integral. Since this difference process is a martingale, one needs to use predictable variation and optional variation to get a handle on the asymptotic properties of the estimator. If the number at risk process gets large, then there is a need to rescale the predictable variation process to get a non-degenerate result. All these results are extremely useful to get asymptotic results.

Two-sample tests for counting processes

The first connection to counting processes was made by Aalen in his Ph.D thesis.

TIME LINE OF EVENTS

- **1976** : Aalen represented the two-sample test as a martingale. The very simple idea was to write the test statistic as a weighted stochastic integral over the difference between two Nelson-Aalen estimators.

$$X(t) = \int_0^t L(s) d(\hat{A}_1(s) - \hat{A}_2(s)) = \int_0^t L(s) \left(\frac{dN_1(s)}{Y_1(s)} - \frac{dN_2(s)}{Y_2(s)} \right)$$

- **1978** : Aalen showed that all the previous proposals of test scores were special cases by judicious use of the weight function
- **1980** : Richard Gill in his Ph.D thesis did a thorough study of two-sample tests from a martingale perspective.

Copenhagen environment

The authors credit the statistics group at the university of Copenhagen for combining theory and applications, so as to further develop the theory.

From Kaplan-Meier to the empirical transition matrix

This section describes the development of competing risk models

TIME LINE OF EVENTS

- **1973** : Matrix version of Kaplan-Meier estimator for markov chains
- **1975-1976** : Soren Johansen used martingale theory along with product integral approach to non-homogeneous Markov chains. The theory of stochastic integrals could be then used in a simple and elegant way.
- **1978** : Fleming published the matrix version of KM estimator. He based his work on the martingale counting approach.

k-sample tests

TIME LINE OF EVENTS

- **1980**: Ornulf Bogan along with Aalen used the martingale approach to k-sample tests
- **1980** : Richard Gill suggested several changes to the earlier papers by Bogan and thus came up with a more robust theory behind the tests.

The Cox model

Time line of events

- **1950's, 1960's**: Development of clinical trials exploded and hence the need to analyze censored survival data increased.
- **1972** : Cox proportional hazards model
- **1974** : Cumulative baseline hazard rate estimate by Breslow
- **1978** : Projects on recurrent event analysis began at Copenhagen
- **1981** : Tsiatis used classical methods to provide a thorough treatment of large sample properties of estimators in the case of time invariant covariates
- **1979-1980** : Per Kragh Anderson obtained martingale stochastic representations for Cox score function and various asymptotic properties
- **1980** : Aalen additive hazards model served as a tool for analyzing survival data with changing effects of covariates
- **1982** : Richard Gill came up with general conditions for the asymptotic results
- **1985** : Anderson and Bogan extended the results to multivariate counting processes modeling the occurrence of several types of events in the same subjects
- **1988, 1990** : Martingale residuals were defined, that subsequently became the basis for a number of goodness-of-fit tests

Literature

The authors mention two monographs that lead to widespread knowledge dissemination of martingale methods in survival analysis

Time line of events

- **1991** : *Counting processes and Survival analysis* monograph published
- **1993** : *Statistical models based on counting process* monograph published

Takeaway

There are two major takeaways from this paper. One is of course the time line of all the developments in the field of survival analysis. The second takeaway from this paper is a good intuitive understanding of martingales and martingale stochastic integrals and their practical application in getting to asymptotic properties of many estimators. Survival analysis is one field where theoretical developments and practical applications have gone hand-in-hand. The paper, mirroring the developments in the field, gives a balanced view of the way martingales played a crucial role in all the developments.