TMA4275 Lifetime analysis
Obligatory project 1, Spring 2024

Out: Tuesday January 23rd
In: Wednesday February 7th at (latest) 21.00

Important information: Parts of this project are to be done using R. An introduction to R can be found in the course web page (see Statistical software). The project report should consist of one (and only one) pdf-file, and should be uploaded via Blackboard. The project report should include derivation of formulas that you are using in your implementations. The project report should also include the R code you have used to solve the project and the plots you have generated. Associated to the various plots there should be captions explaining the contents of the plots, and in addition all the plots should be explained and discussed in the main text of the report.

The project report should be formulated as a scientific report. In particular, it should be possible to understand what you have done without reading the questions in this problem text. Moreover, the text in the project report should consists of full sentences and proper punctuation should by used throughout, also in equations! All results you present should be discussed. What can you (and the world) learn from your results? The project text should be written so that it is easy to follow by your fellow students in TMA4275 Lifetime analysis.

The report can be written in English or Norwegian. The project can be done alone or in groups of two or three persons. If you do the project in a group, only one of the individuals in the group should hand in the solution. In your solution, specify your (full) names, NOT student or candidate numbers.

Your solution should be handed in in Blackboard. After having logged in to Blackboard click on "course information" to find where to hand in your solution.

Problem 1: Discrete time martingales

Let $\{X_n\}_{n=0}^{\infty}$ be a (first-order) homogeneous Markov chain, where for each $n = 0, 1, 2 \ldots$ we have $X_n \in \{-1, 0, 1\}$. We denote the initial distribution of the X_n chain by

$$P(X_0 = k) = \alpha_k \text{ for } k = -1, 0, 1,$$

and the transition probabilities we denote by

$$P(X_n = k|X_{n-1} = j) = \beta_{jk} \text{ for } j, k = -1, 0, 1.$$

Then let $\{Z_n\}_{n=0}^{\infty}$ be defined from the X_n chain by the relation

$$Z_n = \sum_{s=0}^{n} X_s.$$

Let $\mathcal{F}_n$ be the history containing information about the values of $\{X_s\}_{s=0}^{n}$ and $\{Z_s\}_{s=0}^{n}$.

Table 1: Values for $\{\beta_{jk}; j,k = -1,0,1\}$ which makes Z_n a zero-mean martingale.

$j\backslash k$	-1	0	1
-1	0.48	0.04	0.48
0	0.01	0.98	0.01
1	0.49	0.02	0.49

a) Identify what restrictions you need to put on $\{\alpha_k; k = -1,0,1\}$ and $\{\beta_{jk}; j,k = -1,0,1\}$ for $\{Z_n\}_{n=0}^{\infty}$ to be a zero-mean martingale with respect to $\mathcal{F}_n$. In particular use this to observe that Z_n is a zero-mean martingale when $\alpha_{-1} = \alpha_1 = 0, \alpha_0 = 1$ and the β_{jk}'s are as given Table 1.

In the following we restrict the values of $\{\alpha_k; k = -1,0,1\}$ and $\{\beta_{jk}; j,k = -1,0,1\}$ to be so that Z_n is a zero-mean martingale.

b) Find the predictable variation process $\langle Z \rangle_n$, and the optional variation process $[Z]_n$ of the zero-mean martingale Z_n.

c) Write an R function that simulates the Z_n process up to some specified time N, and outputs the simulated Z_n process and the corresponding predictable and optional variation processes. Input arguments to the function should be N and the values of $\{\alpha_k; k = -1,0,1\}$ and $\{\beta_{jk}; j,k = -1,0,1\}$.

For the values of $\{\alpha_k; k = -1,0,1\}$ and $\{\beta_{jk}; j,k = -1,0,1\}$ given in the end of Problem 1**a**), use the function to simulate one realisation of the Z_n process up to time $N = 50$. Make plots of the simulated Z_n and the two corresponding variation processes.

d) Use the R function you implemented in **c**) to simulate a large number of realisations of Z_n up to time $N = 100$. Make plots of some of the generated Z_n processes and corresponding variation processes. Include in the plots sufficient many realisations so that it is possible to get an impression of the variability of the processes, but not so many that everything becomes black.

For each $n = 0,1,\ldots,N$, estimate the variance of Z_n by forming the empirical mean of all the simulated predictable variation processes, and include the result in the plot containing the simulated predictable variation processes. Estimate also the variance of Z_n by forming the empirical mean of all the simulated optional variation processes and include the result in the plot containing the simulated optional variation processes. Finally, use the simulated Z_n processes to estimate the variance of Z_n for each $n = 0,1,\ldots,N$ and make a plot containing all the three estimated variances.

[*Remark: It is here also possible to compute the exact variance of Z_n. The simplest way to do this is perhaps to compute analytically the expected value of the optional variation process.*]

Using the X_n process defined above, define a process $\{U_n\}_{n=0}^{\infty}$ by

$$U_n = \sum_{s=0}^{n} \frac{X_s}{n-s+1}.$$

e) Find the Doob decomposition of $\{U_n\}_{n=0}^{\infty}$, i.e. find expressions for the predictable process $E[U_n|\mathcal{F}_{n-1}]$ and a zero-mean martingale $\{M_n\}_{n=0}^{\infty}$ so that

$$U_n = E[U_n|\mathcal{F}_{n-1}] + \Delta M_n$$

for $n = 0, 1, \ldots$, where $\Delta M_n = M_n - M_{n-1}$.

Problem 2: Continuous time martingales

Assume we have $n \geq 2$ components of a particular type, which we number from 1 to n. We are interested in failures for the components, and assume the different components fail independently of eachother. Component number 1 is different from the remaining components. We assume component number 1 has failures according to a non-homogenuous Poisson process with a (fixed) intensity function $\alpha(t)$ for $t \geq 0$. Letting $N_1(t)$ denote the number of failures of component number 1 from time zero til time t (including time t) we thereby have

$$P(dN_1(t) = 1|N(s); s \in [0,t)) = \alpha(t)dt,$$

where $dN_1(t) = N((t+dt)-) - N(t-)$. Whenever component number 1 has a failure we assume it is immediately repaired, so in particular one failure does not influence the intensity of new failures.

The remaining components, components number 2 to n, are all probabilitistically equal, so in the following we describe the assumed model for component number $i \in \{2, 3, \ldots, n\}$. Component number i is operational at time $t = 0$. Whenever the component is operational it fails with the same intensity as component number 1, namely $\alpha(t)$. When the component has failed it takes some time before it is repaired, and in this period it can not fail again. The time it takes to repair a failed component we assume to be exponentially distributed with mean value $1/\nu$. When the component is repaired it can again fail, and do so according to the same intensity function $\alpha(t)$. We let $Y_i(t)$ be an indicator function specifying whether or not component number i is operational just before time t, i.e.

$$Y_i(t) = \begin{cases} 1 & \text{if component } i \text{ is operational just before time } t, \\ 0 & \text{otherwise.} \end{cases}$$

Letting $N_i(t)$ denote the number of failures for component number i from time zero to time t (included time t) we thereby have

$$P(dN_i(t) = 1|N_i(s), Y_i(s); s \in [0,t)) = Y_i(t)\alpha(t)dt,$$

where $dN_i(t) = N_i((t+dt)-) - N_i(t-)$. The counting process of focus in this problem is then

$$N(t) = \sum_{i=1}^{n} N_i(t),$$

the total number of failures from time zero to time t (included time t). In your solution of the questions below you may also use the notation $Y(t)$ for the number of components at risk for failure just before time t, i.e.

$$Y(t) = 1 + \sum_{i=2}^{n} Y_i(t).$$

Moreover we define $A(t) = \int_0^t \alpha(s)ds$ and let $\mathcal{F}_t$ be the history that contains information about all $N_i(t), Y_i(t), i = 1, 2 \ldots, n$ from time zero to time t (including time t).

a) Starting from the general formulation of the Doob-Meyer decomposition, find the Doob-Meyer decomposition of $N(t)$. In particular, starting from

$$dN^\star(t) = E[dN(t)|\mathcal{F}_{t-}]$$

show that the compensator of $N(t)$ is

$$N^\star(t) = \int_0^t \alpha(t)Y(t)dt.$$

b) Write up the Doob-Meyer decomposition on incremental form. Divide by $Y(t)$ on both sides of this equation and thereafter integrate from time 0 to time t to get the expression

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

Explain why the second term on the right hand side of this expression is a zero-mean martingale, and explain why this in turn implies that the integral on the left hand side,

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

is an unbiased estimator of $A(t)$.

One should note that since $N(t)$ is a counting process, $dN(t)$ equals unity at times t where a failure occurs, and is otherwise equal to zero. The integral for $\widehat{A}(t)$ can therefore be written as a sum. Letting $T_1, T_2, \ldots$ denote the (ordered) failure times, we have

$$\widehat{A}(t) = \sum_{j:T_j<t} \frac{1}{Y(T_j)}.$$

c) Since $N(t)$ is a counting process we know that the optional variation process of the associated zero-mean martingale $M(t)$ is $[M](t) = N(t)$. Use this to find the optional variation process of

$$\mathbb{I}(t) = \int_0^t \frac{dM(s)}{Y(s)}.$$

Thereafter use the expression you found for the optional variation process of $\mathbb{I}(s)$, $[I](t)$, to show that an unbiased estimator of $\sigma^2(t) = \text{Var}[\widehat{A}(t)]$ is

$$\widehat{\sigma}^2(t) = \sum_{j:T_j<t} \frac{1}{Y(T_j)^2}.$$

{*Hint: Remember that Var*$[\mathbb{I}(t)] = E[[\mathbb{I}](t)]$}

In the last item of the project you should for specific choices of $\alpha(t)$, ν, n and τ simulate the failure processes defined above for $t \in [0, \tau]$. For this you may use

$$\alpha(t) = \begin{cases} \frac{t}{600} & \text{for } t \leq 60, \\ \frac{1}{10} & \text{otherwise,} \end{cases}$$

$\nu = 0.2$, $n = 25$ and $\tau = 100$, or something else of your choice. If you make other choices you should of course specify what values your are using.

d) Implement R code to simulate the failure processes defined above. Thereafter consider the simulated processes as observed and use them to compute the corresponding estimates $\widehat{A}(t)$ and $\widehat{\sigma}^2(t)$. Make a plot of $\widehat{A}(t)$ and, assuming $\widehat{A}(t)$ to be approximately Gaussian distributed, a 90%-confidence interval for $A(t)$ for each $t \in [0, \tau]$. Include in the plot also the true $A(t)$ you used when you simulated the data.