Stochastic Processes in Life Insurance: The Dynamic Approach

Jesper Lund Pedersen

Department of Mathematical Sciences
University of Copenhagen
Universitetsparken 5
DK-2100 Copenhagen
Denmark

Contents

CHAPTER I

Lebesgue-Stieltjes calculus

In context of life insurance, this chapter presents general mathematical definitions and results to describe (formalise) payment streams as well as valuation of payment streams. This chapter form the foundation for the next chapters. The definitions and results from measure and integration theory are outlined, where some are extensions of what is covered in a first course on the subject (see Appendix for a review of basic measure and integration theory).

1. CADLAG functions

Functions (sample paths) often considered in stochastic process theory are functions where discontinuities are jumps.

Definition 1.1. A right-continuous $\mathbb{R}$-valued function $x(t)$ with finite left-limits defined on $[0,\infty)$ is called CADLAG (abbreviates the French *continu à droite, limités à gauche*), that is, $x:[0,\infty)\to\mathbb{R}$ such that

(i) $x(t)=x(t+)=\lim_{s\downarrow t}x(s)$ for $t\geq 0$ (right-continuous).
(ii) $x(t-)=\lim_{s\uparrow t}x(s)$ exists finitely for $t>0$ (left-limit).

A left-continuous function on $(0,\infty)$ with right-limits on $[0,\infty)$ is called CAGLAD.

Example 1.2. A simple example of a CADLAG function is $x(t)=1_{[0,1)}(t)$. But $x(t)=1/(1-t)$ for $t<1$ and $x(t)=0$ for $t\geq 1$ is not a CADLAG function.

Let $x(t):[0,\infty)\to\mathbb{R}$ be a CADLAG function then $x(t-)=\lim_{s\uparrow t}x(s)$ is the left-continuous version of $x(t)$, that is, $x(t-)$ is CAGLAD and vice verse if $x(t)$ is CAGLAD then $x(t+)=\lim_{s\downarrow t}x(s)$ the right-continuous version of $x(t)$, that is, $x(t+)$ is CADLAG. The jump function of a CADLAG function $x(t)$ is defined by

$$\Delta x(t)=x(t)-x(t-)\text{ for }t>0$$

and is the size of the jump at point t. Thus a CADLAG function can only have jump discontinuities and the jump discontinuities is at most countable as stated in the next proposition. (The results below for CADLAG functions also hold for CAGLAD functions).

Proposition 1.3. *A CADLAG function $x:[0,\infty)\to\mathbb{R}$ has at most countable many discontinuity points, that is, the set of jump times $D_x=\{t>0:\Delta x(t)\neq 0\}$ is at most countable.*

Some basic properties of CADLAG functions are summarised in the next result.

Proposition 1.4. *Let $x,x_1,x_2:[0,\infty)\to\mathbb{R}$ be CADLAG functions, then*

(i) *$\alpha x_1(t)+\beta x_2(t)$ is a CADLAG function where α and β are constants.*
(ii) *x is a Borel (measurable) function.*
(iii) *x is continuous at the point $t>0$ if and only if $\Delta x(t)=0$.*
(iv) *Uniform limit of a sequence of CADLAG functions on bounded intervals is itself CADLAG.*
(v) *x can be approximated by a sequence of piecewise constant functions.* (*see Definition* 2.11).

2. Functions of finite variation

The mathematical concept to the notion of payment streams is variation. Moreover, variation is the foundation of integration theory. Variation measures the total up-and-down movement of a function, that is, the total vertical distance traveled by the function. Thus, a function of finite variation is a function that wiggles finitely. Variation is also applied in theory of stochastic processes, stochastic control theory, and other subjects.

Remark 2.1. Recall that the positive and negative parts of a number t are given by $t^+ = \max\{t, 0\}$ and $t^- = -\min\{t, 0\}$, so that $t = t^+ - t^-$ and $|t| = t^+ + t^- = 2t^- + t$.

Definition 2.2. The variation of the function $x : [0, \infty) \to \mathbb{R}$ on the interval $[0, t]$ is defined by

$$V^x(t) = \sup \sum_{i=1}^{n} |x(t_i) - x(t_{i-1})|$$

and the positive and negative variation of the function x is given by, respectively

$$V_+^x(t) = \sup \sum_{i=1}^{n} \big(x(t_i) - x(t_{i-1})\big)^+ \text{ and } V_-^x(t) = \sup \sum_{i=1}^{n} \big(x(t_i) - x(t_{i-1})\big)^-$$

where the supremums are over all finite partitions $0 = t_0 < t_1 < \cdots < t_{n-1} < t_n = t$ of $[0, t]$.

(i) The function x is of finite variation if $V^x(t) < \infty$ for all $t \geq 0$.

By the definition of variation we have the following basic properties.

Proposition 2.3. *Let $x, x_1, x_2 : [0, \infty) \to \mathbb{R}$ be functions.*

(i) *$V^x(t)$, $V_+^x(t)$, and $V_-^x(t)$ are increasing functions.*

(ii) *If x is increasing or decreasing then it is of finite variation.*

(iii) *If x_1 and x_2 are of finite variation then $\alpha x_1(t) + \beta x_2(t)$ is of finite variation where α and β are two constants.*

Remark 2.4. There is no connection between finite variation and continuity. Indeed a discontinuous increasing function is of finite variation. The function

$$x(t) = \begin{cases} t \sin(\pi/t) & \text{for } t > 0 \\ 0 & \text{for } t = 0 \end{cases}$$

is continuous without being of finite variation as it is infinite variation on e.g. $[0, 2]$.

If $a_1(t)$ and $a_2(t)$ are increasing functions then we have that the difference $a_1(t) - a_2(t)$ is a function of finite variation by Proposition 2.3. The Jordan decomposition below gives the opposite result that a function of finite variation can be expressed as the difference of two increasing functions. Using the two expressions in Remark 2.1, we have for any finite partition $0 = t_0 < t_1 < \cdots < t_{n-1} < t_n = t$ of $[0, t]$ that

$$\sum_{i=1}^{n} \big(x(t_i) - x(t_{i-1})\big)^+ = \sum_{i=1}^{n} \big(x(t_i) - x(t_{i-1})\big)^- + \big(x(t) - x(0)\big) \tag{2.1}$$

$$\sum_{i=1}^{n} |x(t_i) - x(t_{i-1})| = 2\sum_{i=1}^{n} \big(x(t_i) - x(t_{i-1})\big)^- + \big(x(t) - x(0)\big) \tag{2.2}$$

and taking supremum of each side of (2.1) and (2.2) one can derive the following result.

Lemma 2.5. *Let $x : [0, \infty) \to \mathbb{R}$ be a function of finite variation.*
(i) $V_+^x(t) - V_-^x(t) = x(t) - x(0)$.
(ii) $V^x(t) = V_+^x(t) + V_-^x(t)$.

By Lemma 2.5 we get the Jordan decomposition

$$x(t) = x(0) + V_+^x(t) - V_-^x(t)$$

which has the following minimal property.

Theorem 2.6. (Jordan decomposition). *A function $x : [0, \infty) \to \mathbb{R}$ is of finite variation if and only if $x(t)$ is the difference of two increasing functions $a_+(t)$ and $a_-(t)$, that is, $x(t) = a_+(t) - a_-(t)$. In that case $V_\pm^x(t) - V_\pm^x(s) \le a_\pm(t) - a_\pm(s)$ for $0 \le s < t$.*

Let $a(t)$ be an increasing function and hence the left- and right-limits exist for an increasing and has at most countable many discontinuity points. Since the variation function satisfies $V^a(t) = a(t) - a(0)$ we have that $V^a(t+) - V^a(t) = a(t+) - a(t)$ and $V^a(t) - V^a(t-) = a(t) - a(t-)$. Then Jordan decomposition implies the following result.

Proposition 2.7. *Let $x : [0, \infty) \to \mathbb{R}$ be a function of finite variation. Then the left-limit $x(t-)$ for $t \in (0, \infty)$ and right-limit $x(t+)$ for $t \in [0, \infty)$ exist and x has at most countable many discontinuity points. Moreover, the jumps of the variation function V^x coincide with the absolute value of the corresponding jumps of x, that is, $V^x(t+) - V^x(t) = |x(t+) - x(t)|$ for $t \ge 0$ and $V^x(t) - V^x(t-) = |x(t) - x(t-)|$ for $t > 0$.*

2.1. FV-functions. Right-continuous finite variation functions are an important class of functions for measure theory as well for sample paths of stochastic processes.

Definition 2.8. A function $x : [0, \infty) \to \mathbb{R}$ is an FV-function if
(i) $x(t)$ is of finite variation.
(ii) $x(t)$ is CADLAG.

Condition (ii) can be replaced by the weaker assumption that x is right-continuous and by Proposition 2.7 $x(t)$ is then CADLAG. By Proposition 1.4, Proposition 2.3, Lemma 2.5 and Proposition 2.7, FV-functions have the following properties, in particular the two increasing functions in Jordan decomposition can be chosen to be CADLAG.

Proposition 2.9. *Let x, x_1 and x_2 be FV-functions.*
(i) *V^x and $V_\pm^x$ are increasing CADLAG functions.*
(ii) *$\alpha x_1(t) + \beta x_2(t)$ is an FV-function where α and β are constants.*

For FV-functions there is another another decomposition. An FV-function can be decomposed in to the sum of a continuous FV-function and a discrete FV-function. Indeed, let $a(t)$ be an increasing CADLAG function. The series of jumps $\sum_{0<s\le t} \Delta a(s) \le a(t) - a(0) = V^a(t) < \infty$ is absolutely convergent where the sum extends over all jump times, that is, by Proposition 2.7 the set of jump points/times $D_a = \{t > 0 : \Delta a(t) \neq 0\}$ is at most countably and if $(0, t] \cap D_a = \emptyset$ then $\sum_{0<s\le t} \Delta a(s) = 0$ and if $(0, t] \cap D_a \neq \emptyset$ then $\sum_{0<s\le t} \Delta a(s) = \sum_{s\in(0,t]\cap D_a} \Delta a(s)$. The discrete part of a is defined by $a^d(0) = 0$ and $a^d(t) = \sum_{0<s\le t} \Delta a(s)$ for $t > 0$ and it is an increasing function. The number of jumps greater than $1/n$ is finite and define the increasing CADLAG function $a^{(n)}(0) = 0$ and $a^{(n)}(t) = \sum_{0<s\le t} 1_{\{\Delta a(s)>1/n\}} \Delta a(s)$. As $n \uparrow \infty$, $a^{(n)}$ converges uniformly to a^d on any finite interval $[0, t]$. Since $a^{(n)}$ is CADLAG then by Proposition 1.4, a^d is also CADLAG. If $\Delta a(t) > 1/k$ for some k then $\Delta a^{(n)}(t) = \Delta a(t)$ for all $n \geq k$ which implies that $\Delta a^d(t) = \Delta a(t)$. If $\Delta a(t) = 0$ then $\Delta a^{(n)}(t) = 0$ for all n and $\Delta a^d(t) = 0$. Thus a^d is an increasing CADLAG function with the same jumps as a. For this reason we can define the continuous part as $a^c(t) = a(t) - a(0) - a^d(t)$ which is a continuous increasing function, since $\Delta a^c(t) = \Delta a(t) - \Delta a^d(t) = 0$. Again by Jordan decomposition we have for an FV-function $x(t)$ that $\sum_{0<s\leq t} |\Delta x(s)| \leq V^x(t) < \infty$ and the series of jumps is absolutely convergent and we have the following result.

Theorem 2.10. *An FV-function $x : [0,\infty) \to \mathbb{R}$ can be written as*

$$x(t) = x(0) + x^d(t) + x^c(t)$$

where

(i) *The discrete part $x^d(0) = 0$ and $x^d(t) = \sum_{0<s\leq t} \Delta x(s)$ for $t > 0$ is an FV-function with same jumps as x, that is, $\Delta x(t) = \Delta x^d(t)$ for $t > 0$.*
(ii) *The continuous part $x^c(t) = x(t) - x(0) - x^d(t)$ is a continuous FV-function. Further, if x is increasing then x^c is also increasing.*

An FV-function x reduces to two special cases where either $x^c(t) = 0$ or $x^d(t) = 0$.

The first case that $x^c(t) = 0$, then x is a pure jump function given by $x(t) = x(0) + \sum_{0<s\leq t} \Delta x(s)$. We see that x only changes by jumps. A simple example of a pure jump function is the one-jump function $x(t) = 1_{[1,\infty)}(t)$, but there are more complicated, for example, the set of jump times may be dense (e.g. the sample paths of a Gamma process which is an increasing Lévy process with infinitely many jumps).

A pure jump function can be constructed with jumps at given times. Let $(t_n, n = 1, 2, \ldots)$ be a sequence of jump times with $t_n > 0$ but they might not be enumerable in increasing order. Let $(z_n, n = 1, 2, \ldots)$ be a sequence of jump sizes such that $\sum_{n=1}^\infty |z_n| < \infty$. Let $z : [0,\infty) \to \mathbb{R}$ be given by

$$z(t) = z(0) + \sum_{n=1}^{\infty} z_n 1_{[t_n,\infty)}(t). \tag{2.3}$$

The expression $z(t) = z(0) + \sum_{n=1}^\infty z_n^+ 1_{[t_n,\infty)}(t) - \sum_{n=1}^\infty z_n^- 1_{[t_n,\infty)}(t)$ shows that f is the difference of two increasing functions and hence of finite variation. Define $z^{(k)}(t) = z(0) + \sum_{n=1}^k z_n 1_{[t_n,\infty)}(t)$ As $k \uparrow \infty$, $z^{(k)}$ converges uniformly to z on any finite interval $[0,t]$. Since $z^{(k)}$ is CADLAG then by Proposition 1.4, z is also CADLAG and z is an FV-function. Since $z(t-) = \sum_{n=1}^\infty z_n 1_{(t_n,\infty)}(t)$ we have that the jump function is given by $\Delta z(t) = z(t) - z(t-) = \sum_{n=1}^\infty z_n 1_{\{t_n\}}(t)$ and the discrete part is then given by $z^d(t) = \sum_{0<s\leq t} \Delta z(s) = \sum_{n=1}^\infty z_n 1_{[t_n,\infty)}(t) = z(t) - z(0)$. Thus z is a pure jump function. A pure jump function with finite number of jumps in any finite interval is called a piecewise constant function (or a step function).

Definition 2.11. Let $(t_n, n = 1, 2 \ldots)$ be the sequence of jump times of a pure jump function z given in (2.3). The function z is a piecewise constant function if the jump times satisfies

(i) $0 < t_1 \leq t_2 \leq \cdots$.
(ii) $t_n < t_{n+1}$ if $t_n < \infty$.
(iii) $\lim_{n\uparrow\infty} t_n = \infty$.

The piecewise constant function z is constant on each interval $[t_n, t_{n+1})$ where $t_n < \infty$ and condition (iii) is equivalent to that only finitely many jumps can occur in any finite time interval [0,t].

The second case that $x^d(t) = 0$, then x is a continuous FV-function. A simple example of a continuous FV-function is $x(t) = t$. But in general a continuous FV-function may be rather complicated such as singular continuous (see Corollary 2.12 below), for example, the Cantor function or the sample paths of the maximum of a Brownian motion.

A function $x(t)$ is absolutely continuous if for some integrable (Borel) function φ such that

$$x(t) = x(0) + \int_0^t \varphi(s)\, ds. \tag{2.4}$$

If $x(t)$ is an absolutely continuous function then it is also a continuous FV-function. Indeed, the expression

$$x(t) - x(0) = \int_0^t \varphi(s)\, ds = \int_0^t \varphi^+(s)\, ds - \int_0^t \varphi^-(s)\, ds$$

shows that x is the difference of two increasing functions and x is a continuous FV-function.

A deep result in calculus is Lebesgue differential theorem: an FV-function x is almost everywhere differentiable and the derivative x' is integrable. Another result in calculus (Lebesgue's form of the Fundamental theorem of calculus) is that the derivative of the absolutely continuous in (2.4) is almost everywhere equal to φ and conversely an FV-function x is absolutely continuous if $x(t) = x(0) + \int_0^t x'(s)\, ds$. An FV-function is called singular if the derivative is zero almost everywhere. We want to identify the singular part and the absolutely continuous part of the decomposition of an FV-function in Theorem 2.10. Let x be an FV-function, then the derivative $x'(t)$ exists almost everywhere and is integrable. Define the absolutely continuous $x^{ac}(t) = \int_0^t x'(s)\, ds$, then $(x^{ac})'(t) = x'(t)$ almost everywhere. Note that the derivative of the discrete part x^d is zero almost everywhere and therefore the discrete part x^d is singular. The function $x^{sc}(t) = x^c(t) - x^{ac}(t) = x(t) - x(0) - x^d(t) - x^{ac}(t)$ is a continuous FV-function which derivative is zero almost everywhere. Thus x^{sc} is a singular continuous FV-function. Consequently, we get the following result.

Corollary 2.12. (Lebesgue decomposition of function of finite variation). *An FV-function $x : [0, \infty) \to \mathbb{R}$ can be written as*

$$x(t) = x(0) + x^d(t) + x^{ac}(t) + x^{sc}(t)$$

where

(i) *x^d is a (singular discrete) pure jump function.*
(ii) *x^{ac} is an absolutely continuous function.*
(iii) *x^{sc} is singular continuous FV-function.*

An attractive class of FV-functions are piecewise absolutely continuous functions.

Definition 2.13. A piecewise absolutely continuous function $x : [0, \infty) \to \mathbb{R}$ is an FV-function that can be written as

$$x(t) = x(0) + x^d(t) + x^{ac}(t)$$

where

(i) The discrete part x^d is a piecewise constant function.
(ii) The continuous part x^{ac} is an absolutely continuous function.

3. Lebesgue-Stieltjes integration

Lebesgue-Stieltjes integrals are at the core of this chapter. The starting point is to define integrals with respect to increasing functions using general theory of measures and integrations. Then using Jordan decomposition to deal with the case of integration with respect to functions of finite variation.

3.1. Lebesgue-Stieltjes integration with respect to increasing functions. Recall the definition of a measure (see Definition 25.18) and let $\mathcal{B}(0,\infty)$ be the Borel σ-algebra of $(0,\infty)$.

Definition 3.1. A measure μ on $((0,\infty),\mathcal{B}(0,\infty))$ is a Lebesgue-Stieltjes measure if

(i) $\mu(I) < \infty$ for every bounded interval $I \subseteq (0,\infty)$.

Given the Lebesgue measure λ (see Definition 25.22) which is determined by the length of an interval, that is, $\lambda((s,t]) = t-s$ for all $s < t$ and an increasing CADLAG function $a : [0,\infty) \to \mathbb{R}$ we can come from the Lebesgue measure to a Lebesgue-Stieltjes measure. Let $g(x) = \inf\{t > 0 : a(t) \geq x\}$ be the CAGLAD increasing (hence Borel, see Proposition 1.4) generalised inverse of a. Since g is Borel (measurable) we have by Lemma 25.20 that $\lambda^a(B) = \lambda(\{x : g(x) \in B\})$ is a unique measure on $((0,\infty),\mathcal{B}(0,\infty))$ with $\lambda^a((s,t]) = \lambda(\{x : g(x) \in (s,t]\}) = a(t) - a(s)$ for all $s < t$. Conversely, let μ be a Lebesgue-Stieltjes measure and define the function $a(t) = a(0) + \mu((0,t])$ for $t > 0$ where $a(0)$ is some constant that cannot be determined by μ. One can check that a is increasing and CADLAG with $\mu = \lambda^a$. The class of increasing functions corresponds to the class of Lebesgue-Stieltjes measures and a precise statement of this correspondence is the content in the following proposition.

Proposition 3.2. (Lebesgue-Stieltjes measure). *There is a one-to-one correspondence between CADLAG increasing functions (up to addition of a constant) and Lebesgue-Stieltjes measures, given by* $\mu((s,t]) = a(t) - a(s)$ *for all* $0 \leq s < t$.

Remark 3.3. In light of the above considerations, we have the following observations.

1. The Lebesgue measure λ is a Lebesgue-Stieltjes measure with corresponding increasing function $a(t) = t$.
2. The point mass of a Lebesgue-Stieltjes measure is given by $\lambda^a(\{t\}) = \lim_{s\uparrow t} \lambda^a((s,t]) = \lim_{s\uparrow t}(a(t) - a(s)) = a(t) - a(t-) = \Delta a(t)$ for $t > 0$.

For Lebesgue-Stieltjes measures, we can introduce integrals with respect to increasing functions (see Definition 26.3 for general definition of integrals with respect to measures).

Definition 3.4. Let $a : [0,\infty) \to \mathbb{R}$ be an increasing CADLAG function (increasing FV-function). A Borel function $f : (0,\infty) \to \mathbb{R}$ is a-integrable (f is integrable with respect to a) if $\int_{(0,\infty)} |f(t)|\, d\lambda^a(t) = \int_0^\infty |f(t)|\, da(t) < \infty$ and the Lebesgue-Stieltjes integral of f with respect to a is given by

$$\int_{(0,\infty)} f(t)\, d\lambda^a(t) = \int_0^\infty f(t)\, da(t).$$

For a Borel set $B \in \mathcal{B}(0,\infty)$, f is a-intgrable on B if $\int_B |f(t)| 1_B(t)\, da(t) < \infty$ and the Lebesgue-Stieltjes integral of f with respect to a on B is given by $\int_B f(t)\, da(t) = \int_0^\infty f(t) 1_B(t)\, da(t)$. In particular, for $B = (s,t]$ we write

$$\int_s^t f(u)\, da(u) = \int_{(s,t]} f(u)\, da(u).$$

The function f is the integrand and the function a is the integrator.

Example 3.5. (Expectation of random variables). The Lebesgue-Stieltjes integral is a generalisation of the following formula to compute mean values in probability theory. Let $\tau > 0$ be a strictly positive random variable and let $F(t) = \mathbf{P}(\tau \leq t)$ be the distribution function. Recall that F is an increasing CADLAG function. Then λ^F is a probability measure and we have that $\mathbf{P}(X \in B) = \lambda^F(B) = \int_B dF(t)$. Moreover, if $g(\tau)$ has finite expectation for some Borel function g, then $\mathbf{E}[g(\tau)] = \int_0^\infty g(t)\, d\lambda^F(t) = \int_0^\infty g(t)\, dF(t)$.

3.2. Lebesgue-Stieltjes integration with respect to functions of finite variation. Lebesgue-Stieltjes integrals can be extended to the case that integrators are of finite variation. Let $\mathcal{B}(0,t]$ be the Borel σ-algebra of the interval $(0,t]$.

Definition 3.6. A locally finite signed Lebesgue-Stieltjes measure ν is a finite signed measure on $((0,t], \mathcal{B}(0,t])$ for all $t > 0$, that is, ν is a set function on $((0,t], \mathcal{B}(0,t])$ satisfying

(i) $|\nu((0,t])| < \infty$.
(ii) ν is countable additive, that is, $\nu(\emptyset) = 0$ and $\nu(\cup_{n=1}^\infty B_n) = \sum_{n=1}^\infty \nu(B_n)$ whenever $B_1, B_2, \ldots$ is a sequence of disjoint sets in $\mathcal{B}(0,t]$ of pairwise disjoint sets.

Given an FV-function $x : [0,\infty) \to \mathbb{R}$ and then $x(t) = x(0) + V_+^x(t) - V_-^x(t)$ by Jordan decomposition where $V_\pm^x$ are the positive and negative variation of x. By Proposition 3.2, let $\lambda^{V_\pm^x}$ be the (unique) Lebesgue-Stieltjes measure such that $\lambda^{V_\pm^x}((s,t]) = V_\pm^x(t) - V_\pm^x(s)$. Then $\lambda^x = \lambda^{V_+^x} - \lambda^{V_-^x}$ is the signed Lebesgue-Stieltjes measure with

$$\lambda^x((s,t]) = \lambda^{V_+^x}((s,t]) - \lambda^{V_-^x}((s,t]) = (V_+^x(t) - V_+^x(s)) - (V_-^x(t) - V_-^x(s)) = x(t) - x(s) \tag{3.1}$$

for all $s < t$ and uniquely determines λ^x. Conversely, for the ν signed Lebesgue-Stieltjes measure, we define for $B \in \mathcal{B}(0,t]$

$$\nu_+(B) = \sup\{\nu(A) : A \in \mathcal{B}(0,t], A \subseteq B\} \text{ and } \nu_-(B) = -\inf\{\nu(A) : A \in \mathcal{B}(0,t], A \subseteq B\}$$

which are the positive and negative variations of ν, respectively. One can check that the positive variation ν_+ and the negative variation ν_- are Lebesgue-Stieltjes measures. For $A, B \in \mathcal{B}(0,t]$ and $A \subseteq B$ then $\nu(A) = \nu(B) - \nu(B \backslash A)$ and we have that $\nu(B) - \nu_+(B) \leq \nu(A) \leq \nu(B) + \nu_-(B)$. Take sup and inf in this expression we get that $\nu_+(B) \leq \nu(B) + \nu_-(B)$ and $\nu(B) - \nu_+(B) \leq -\nu_-(B)$ and it follows that $\nu(B) = \nu_+(B) - \nu_-(B)$. Hence the signed Lebesgue-Stieltjes measure ν is the difference of the two Lebesgue-Stieltjes measures $\nu_\pm$. The corresponding between signed Lebesgue-Stieltjes measures and FV-functions are given in the following proposition.

Proposition 3.7. (Signed Lebesgue-Stieltjes measure). *There is a one-to-one correspondence between FV-functions (up to addition of a constant) and locally finite signed Lebesgue-Stieltjes measures, given by $\nu((s,t]) = x(t) - x(s)$ for all $0 \leq s < t$. Furthermore, the decomposition $\nu = \nu_+ - \nu_-$ and the Jordan decomposition $x(t) = x(0) + V_+^x(t) - V_-^x(t)$ are related by $\nu_\pm((s,t]) = V_\pm^x(t) - V_\pm^x(s)$.*

Remark 3.8. Comments to above definition and proposition.

1. If x is an FV-function such that positive variation $V_+^x(\infty) < \infty$ or negative variation $V_-^x(\infty) < \infty$ then the associated signed Lebesgue-Stieltjes measure ν is the difference of two Lebesgue-Stieltjes measures $\lambda^{V_\pm^x}$ where at least one of is finite. In this case ν can be extended to be a signed measure on $((0,\infty), \mathcal{B}(0,\infty))$ with values in $[-\infty, \infty)$ or $(-\infty, \infty]$.
2. The bounded FV-function $x(t) = t - k$ for $k \leq t < k+1$, $k = 0,1,2,\ldots$ has positive variation $V_+^x(t) = t$ and the negative variation $V_-^x(t) = k$ for $k \leq t < k+1$, $k = 0,1,2,\ldots$. The associated signed measure ν is the difference of two infinite Lebesgue-Stieltjes measures $\lambda^{V_\pm^x}$ and it is not possible to extend ν to be a signed measure on $((0,\infty), \mathcal{B}(0,\infty))$.

Definition 3.9. Let x be an FV-function. A Borel function $f : (0,\infty) \to \mathbb{R}$ is x-integrable (f is integrable with respect to x) if $\int_0^\infty |f(t)|\, dV^x(t) = \int_0^\infty |f(t)|\, dV_+^x(t) + \int_0^\infty |f(t)|\, dV_-^x(t) < \infty$ and the Lebesgue-Stieltjes integral of f with respect to x is given by

$$\int_0^\infty f(t)\, dx(t) = \int_0^\infty f(t)\, dV_+^x(t) - \int_0^\infty f(t)\, dV_-^x(t).$$

The integrand f is x-integrable on $(s,t]$, for $0 \le s < t$, if $\int_s^t |f(u)|\, dV^x(u) = \int_s^t |f(u)|\, dV_+^x(u) + \int_s^t |f(u)|\, dV_-^x(u) < \infty$ and the Lebesgue-Stieltjes integral of f with respected to x on $(s,t]$ is given by

$$\int_s^t f(u)\, dx(u) = \int_s^t f(u)\, dV_+^x(u) - \int_s^t f(u)\, dV_-^x(u).$$

The value of the integral is independent of the decomposition of x (see Exercise 5.5 below). The Lebesgue-Stieltjes integral is linear both in the integrand and in the integrator.

Proposition 3.10. *Let x, x_1 and x_2 be FV-functions and let α and β be constants.*

(i) *If f_1 and f_2 are x-integrable on $(s,t]$, then $\alpha f_1 + \beta f_2$ is also x-integrable on $(s,t]$ and*

$$\int_s^t \big(\alpha f_1(u) + \beta f_2(u)\big)\, dx(u) = \alpha \int_s^t f_1(u)\, dx(u) + \beta \int_s^t f_2(u)\, dx(u).$$

(ii) *If f is x_1- and x_2-integrable on $(s,t]$ then f is $(\alpha x_1 + \beta x_2)$-integrable on $(s,t]$ and*

$$\int_s^t f(u)\, d\big(\alpha x_1(u) + \beta x_2(u)\big) = \alpha \int_s^t f(u)\, dx_1(u) + \beta \int_s^t f(u)\, dx_2(u).$$

Proof. Note that $\lambda^{\alpha x_1 + \beta x_2} = \alpha \lambda^{x_1} + \beta \lambda^{x_2}$, then proposition is be proved via Proposition 26.4 and Proposition 26.5. $\square$

Proposition 3.11. *Let x be an FV-function and assume that all integrals below are finite.*

(i) $\displaystyle\int_s^t dx(u) = \lambda^x((s,t]) = x(t) - x(s).$

(ii) $\displaystyle\int_{\{t\}} dx(u) = \Delta x(t)$ *and* $\displaystyle\int_{\{t\}} f(u)\, dx(u) = f(t)\Delta x(t).$

(iii) $\displaystyle\int_s^t f(u)\, dx(u) = 0$ *if $x(u) = c$ for $u \in (s,t]$.*

(iv) $\displaystyle\int_s^t f(u)\, dx^d(u) = \sum_{s<u\le t} f(u)\Delta x(u)$ *where x^d is the discrete part of x (see Theorem 2.10).*

(v) $\displaystyle\int_s^t f(u)\, dx(u) = \int_s^t f(u)\varphi(u)\, du + \sum_{s<u\le t} f(u)\Delta x(u)$ *when x is piecewise absolutely continuous (see Definition 2.13) and the absolutely continuous part is given by $x^c(t) = \int_0^t \varphi(s)\, ds$.*

(vi) $\displaystyle\int_s^t \Delta f(u)\, dx(u) = \sum_{s<u\le t} \Delta f(u)\Delta x(u)$ *if f is CADLAG.*

Proof. Let $0 \le s < t$ in the following calculations.

(i) Since $\int_s^t dx(u) = \lambda^x((s,t])$ then $\int_s^t dx(u) = x(t) - x(s)$ according to equation (3.1).

(ii) Apply $(s,t] \to \{t\}$ for $s \uparrow t$ to yield $\int_{\{t\}} dx(u) = \lim_{s\uparrow t} \int_s^t dx(u) = \lim_{s\uparrow t}(x(t) - x(s)) = \Delta x(t)$. Moreover,

$$\int_{\{t\}} f(u)\,dx(u) = \int_{\{t\}} f(u)1_{\{t\}}(u)\,dx(u) = \int_{\{t\}} f(t)1_{\{t\}}(u)\,dx(u) = f(t)\int_{\{t\}} dx(u) = f(t)\Delta x(t).$$

(iii) Since $\lambda^x((s,t]) = x(t) - x(s) = 0$, the interval $(s,t]$ has λ^x-measure zero and the integral is zero.

(iv) Recall that x can be written as (see Theorem 2.10) $x(t) = x(0) + x^d(t) + x^c(t)$ where the discrete part is given by $x^d(t) = \sum_{0<u\le t} \Delta x(u)$. Let $D_x = \{t_1, t_2, \ldots\} = \cup_{n=1}^{\infty}\{t_n\}$ be the set of jump times (discontinuity points). Since $\lambda^{x^d}(I) = \sum_{u\in I\cap D_x} \Delta x(u) = \sum_{n=1}^{\infty} \Delta x(t_n)1_I(t_n)$ where I is a bounded interval of $(0,\infty)$, the measure λ^{x^d} is discrete and is concentrated on D_x. Then we have that

$$\int_s^t f(u)\,dx^d(u) = \int_{(s,t]\cap D_x} f(u)\,dx^d(u) + \int_{(s,t]\setminus D_x} f(u)\,dx^d(u) = \sum_{n: t_n\in(s,t]} \int_{\{t_n\}} f(u)\,dx^d(u)$$

and hence $\int_s^t f(u)\,dx^d(u) = \sum_{s<u\le t} f(u)\Delta x(u)$.

(v) The result follows by Proposition 3.14 below and by (iv) above.

(vi) If D_f is the set of jump times of f, then

$$\int_s^t \Delta f(u)\,dx(u) = \int_s^t \Delta f(u)1_{D_f}(u)\,dx(u) = \sum_{u\in(s,t]\cap D_f} \Delta f(u)\int_{\{u\}} dx(u) = \sum_{s<u\le t} \Delta f(u)\Delta x(u).$$

□

Remark 3.12. The following facts will be used in Chapter III. If x is a continuous FV-function then by Proposition 3.11 (ii) any singleton $\{t\}$ has λ^x-measure zero. Hence $\int_s^t \tilde{f}(u)\,dx(u) = \int_s^t f(u)\,dx(u)$ if f and $\tilde{f}$ differ at most on a countable set of points, i.e. $f = \tilde{f}$ λ^x-everywhere. In particular, if f is CADLAG then recall that $f(t)$ and $f(t-)$ differ at most on a countable number of points (Proposition 1.3) and $\int_s^t f(u)\,dx(u) = \int_s^t f(u-)\,dx(u)$.

Lebesgue-Stieltjes integrals induce new FV-functions.

Definition 3.13. Let x be an FV-function. A Borel function $\varphi : [0,\infty) \to \mathbb{R}$ is locally x-integrable if $\int_0^t |\varphi(s)|\,dV^x(s) = \int_0^t |\varphi(s)|\,dV_+^x(s) + \int_0^t |\varphi(s)|\,dV_-^x(s) < \infty$ for all $t > 0$. Then the function

$$y(t) = \begin{cases} 0 & \text{for } t = 0 \\ \displaystyle\int_0^t \varphi(s)\,dx(s) & \text{for } t > 0 \end{cases} \tag{3.2}$$

is well-defined.

Note that if φ is locally bounded, that is, φ is bounded on every bounded interval of $[0,\infty)$ then φ is locally x-integrable for any FV-function x. Then the function y is well-defined for any FV-function x as integrator.

Proposition 3.14. *The function $y(t)$ defined in* (3.2) *has the following properties.*
(i) *$y(t)$ is an FV-function.*
(ii) *$\Delta y(t) = \varphi(t)\Delta x(t)$.*
(iii) *If $x(t)$ is continuous then $y(t)$ is also continuous.*
(iv) *If the integrals are well-defined and finite then $\displaystyle\int_0^t f(s)\,dy(s) = \int_0^t f(s)\varphi(s)\,dx(s)$.*

Proof. (i) The function y is of finite variation since it is the difference of two increasing functions by the decomposition

$$y(t) = \Big(\int_0^t \varphi^+(s)\,dV_+^x(s) + \int_0^t \varphi^-(s)\,dV_-^x(s)\Big) - \Big(\int_0^t \varphi^-(s)\,dV_+^x(s) + \int_0^t \varphi^+(s)\,dV_-^x(s)\Big).$$

If $t_n \uparrow t$ then $1_{(0,t_n]}(s) \uparrow 1_{(0,t)}(s)$ and if $t_n \downarrow t$ then $1_{(0,t_n]}(s) \downarrow 1_{(0,t]}(s)$. By dominated convergence we have that $y(t)$ is CADLAG where the left-limit is given by $y(t-) = \int_{(0,t)} \varphi(s)\,dx(s)$. Thus $y(t)$ is an FV-function.

(ii) The jump function at time $t > 0$ is

$$\Delta y(t) = \int_0^t \varphi(s)\,dx(s) - \int_{(0,t)} \varphi(s)\,dx(s) = \int_{\{t\}} \varphi(s)\,dx(s) = \varphi(t)\Delta f(t).$$

where Proposition 3.11 (ii) is used in the latter equality.

(iii) According to Proposition 1.4 (iii) $\Delta x(t) = 0$ and hence $\Delta y(t) = \varphi(t)\Delta x(t) = 0$. Again according to Proposition 1.4 (iii), y is continuous.

(iv) The result follows from Proposition 26.6. $\square$

3.3. Differential forms of FV-functions. Note that an FV-function $x(t)$ can be written on the integral form $x(t) = x(0) + \int_0^t dx(s)$. Then $dx(t)$ is the differential (dynamic) form of the integral form. The infinitesimal notation of the differential $dx(t)$ has the interpretation as the increment (change) $x((t+dt)-) - x(t-)$ over the infinitesimal interval $[t, t+dt)$. This informally (heuristically) interpretation is used to motivate some arguments later and is a short-hand notation for the formal integrals.

Formally, $f(t)\,dx(t) = g(t)\,dy(t)$ is the differential (dynamic) form of the integral form

$$\int_0^t f(s)\,dx(s) = \int_0^t g(s)\,dy(s)$$

for all $t > 0$.

Example 3.15. Below are examples of dynamics and some are used in the next chapters.
1. If $x(t) = c$ is a constant function then $dx(t) = 0$.
2. The dynamic form of Proposition 3.10 is

$$\begin{aligned}\big(\alpha f_1(t) + \beta f_2(t)\big)\,dx(t) &= \alpha f_1(t)\,dx(t) + \beta f_2(t)\,dx(t)\\ f(t)\,d\big(\alpha x_1(t) + \beta x_2(t)\big) &= \alpha f(t)\,dx_1(t) + \beta f(t)\,dx_2(t).\end{aligned}$$

3. The dynamics of the function $y(t)$ given in equation (3.2) is $dy(t) = \varphi(t)\,dx(t)$ and then Proposition 3.14 (iv) can be written as $f(t)\,dy(t) = f(t)\varphi(t)\,dx(t)$.
4. Let T be a fixed time, then the dynamics of the function $v(t) = \int_t^T \varphi(s)\,dx(s)$ is $dv(t) = -\varphi(t)\,dx(t)$ due to that $v(t) = \int_0^T \varphi(s)\,dx(s) - \int_0^t \varphi(s)\,dx(s)$.

Remark 3.16. Warning, differential form does not mean to differentiate, an FV-function might not be differentiable at all points. Differential form versus differentiability:

1. Let $y : [0,\infty) \to \mathbb{R}$ be a C^1-function, that is, the derivative y' exists and is continuous and y can be written as $y(t) = y(0) + \int_0^t y'(u)\,du$. Then y has dynamics (differential) given by $dy(t) = y'(t)\,dt$.
2. Conversely, let $y : [0,\infty) \to \mathbb{R}$ be an absolutely continuous function given by $dy(t) = \varphi(t)\,dt$. (Recall that the function is continuous). If $\varphi(t)$ is continuous at t_0 then $y(t)$ is differentiable at t_0 with $y'(t_0) = \varphi(t_0)$. Thus, in the case that φ is continuous then y is differentiable with $y'(t) = \varphi(t)$. An example of a continuous FV-function that is not differentiable at all points is the function $y(t) = |t-1|$ which is not differentiable at $t = 1$. But y has dynamics $dy(t) = \big(1_{(1,\infty)}(t) - 1_{(0,1]}(t)\big)\,dt$.

4. Change of variables formula (Ito formula for FV-functions)

In practice the definition of a Lebesgue-Stieltjes integral is not applied for a calculation of an integral. The Lebesgue-Stieltjes integral is useful by reason of its properties and rules which it can be manipulated. The previous section explores some of the properties of the Lebesgue-Stieltjes integral and this section explores the rules which integrals can be manipulated.

The next result is a cornerstone for FV-functions.

Theorem 4.1. (Integration by parts formula). *If x and y are two FV-functions then*

$$\begin{aligned} x(t)y(t) - x(0)y(0) &= \int_0^t y(s)\,dx(s) + \int_0^t x(s-)\,dy(s) \\ &= \int_0^t y(s-)\,dx(s) + \int_0^t x(s-)\,dy(s) + \sum_{0<s\le t} \Delta x(s)\Delta y(s). \end{aligned}$$

Or on the differential form

$$d\big(x(t)y(t)\big) = y(t)\,dx(t) + x(t-)\,dy(t) = y(t-)\,dx(t) + x(t-)\,dy(t) + \Delta y(t)\,dx(t).$$

Proof. By Fubini's theorem (Theorem 26.10) we have that

$$\begin{aligned} &x(t)y(t) - x(0)y(0) - x(0)\big(y(t)-y(0)\big) - y(0)\big(x(t)-x(0)\big) \\ &\quad= \big(x(t)-x(0)\big)\big(y(t)-y(0)\big) \\ &\quad= \int_0^t \Big(\int_0^t dx(u)\Big)\,dy(s) = \int_0^t \Big(\int_0^s dx(u)\Big)\,dy(s) + \int_0^t \Big(\int_s^t dx(u)\Big)\,dy(s) \\ &\quad= \int_0^t \big(x(s)-x(0)\big)\,dy(s) + \int_0^t \Big(\int_{(0,u)} dy(s)\Big)\,dx(u) \\ &\quad= \int_0^t x(s)\,dy(s) - x(0)\big(y(t)-y(0)\big) + \int_0^t \big(y(u-)-y(0)\big)\,dx(u) \\ &\quad= \int_0^t x(s)\,dy(s) - x(0)\big(y(t)-y(0)\big) + \int_0^t y(u-)\,dx(u) - y(0)\big(x(t)-x(0)\big) \end{aligned}$$

and the result follows. The latter equality in the theorem follows from Proposition 3.11 (vi) that $\int_0^t \Delta g(s)\,df(s) = \sum_{0<s\le t} \Delta g(s)\Delta f(s)$. $\square$

A function $f : \mathbb{R}^n \to \mathbb{R}$ is C^1 (continuously differentiable) if f has continuous first-order partial derivatives and

$$f_{x_i}(x) = \frac{\partial f}{\partial x_i}(x) \text{ for } i = 1, 2, \ldots, n$$

denotes the first-order partial derivative of f. The set of FV-functions is invariant under smooth transformations.

Theorem 4.2. (Change of variable formula—Ito formula for FV-functions). *Let $f : \mathbb{R} \to \mathbb{R}$ be a C^1-function and let $x(t)$ be an FV-function. Then $f(x(t))$ is an FV-function and*

$$f\big(x(t)\big) - f\big(x(0)\big) = \int_0^t f'\big(x(s)\big)\, dx^c(s) + \sum_{0<s\le t} \Big(f\big(x(s)\big) - f\big(x(s-)\big)\Big) \tag{4.1}$$

where x^c is the continuous part of x.

Proof. Let $\mathcal{C}$ denote the class of function in C^1 such that (4.1) holds. The class $\mathcal{C}$ is a vector space containing the function $f(x) = x$. The next step is to prove that $\mathcal{C}$ is a multiplicative class (closed under multiplication). Let $f_1, f_2 \in \mathcal{C}$ and set $f = f_1 f_2$. By the Integration by parts (Theorem 4.1) and (4.1)

$$\begin{aligned}
&f(x(t)) - f(x(0)) \\
&\quad= \int_0^t f_2(x(s))\, df_1(x(s)) + \int_0^t f_1(x(s-))\, df_2(x(s)) \\
&\quad= \int_0^t f_2(x(s)) f_1'(x(s))\, dx^c(s) + \int_0^t f_1 x(s)) f_2'(x(s))\, dx^c(s) \\
&\qquad+ \sum_{0<s\le t} f_2(x(s))\big(f_1(x(s)) - f_1(x(s-))\big) + \sum_{0<s\le t} f_1(x(s-))\big(f_2(x(s)) - f_2(x(s-))\big) \\
&\quad= \int_0^t f'(x(s))\, dx^c(s) + \sum_{0<s\le t} \big(f(x(s)) - f(x(s-))\big)
\end{aligned}$$

so that $f \in \mathcal{C}$. Hence $\mathcal{C}$ contains all polynomials.

Now let f be a C^1 function. By Weierstrass approximation theorem there is a sequence of polynomials $p_1, p_2, \ldots$ such that

$$\sup_{x \le N} \big((p_n(x) - f(x)) \vee (p_n'(x) - f'(x))\big) \to 0$$

for any $N > 0$. Thus $p_n \to f$ and $p_n' \to f'$ uniformly on $[-N, N]$. For given $t_0 > 0$, $x(t) \le N$ for all $t \le t_0$ for some N . Since (4.1) is true for the polynomials p_n and by dominated convergence, we get in the limit that f also satisfies (4.1) for $t \le t_0$. □

Below is the n-dimensional version of Change of variable formula. The proof is based on the same method to prove the one-dimensional version (Theorem 4.2).

Theorem 4.3. (Change of variable formula—Ito formula for FV-functions). *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a C^1-function and let $x(t) = \big(x_1(t), \ldots, x_n(t)\big)$ be a $\mathbb{R}^n$-valued function where each component is an FV-function. Then $f(x(t))$ is an FV-function and*

$$f\big(x(t)\big) - f\big(x(0)\big) = \sum_{i=1}^n \int_0^t f_{x_i}\big(x(s)\big)\, dx_i^c(s) + \sum_{0<s\le t} \Big(f\big(x(s)\big) - f\big(x(s-)\big)\Big)$$

where x_i^c is the continuous part of x_i.

Remark 4.4. It is possible to reformulate the Change of variables formula in different ways.

1. Note that

$$\begin{aligned}\int_0^t f_{x_i}\big(x(s-)\big)\,dx_i(s) &= \int_0^t f_{x_i}\big(x(s-)\big)\,d\big(x_i^c(s)+x_i^d(s)\big)\\ &= \int_0^t f_{x_i}\big(x(s)\big)\,dx_i^c(s) + \sum_{0<s\le t} f_{x_i}\big(x(s-)\big)\Delta x_i(s)\end{aligned}$$

   where we have used Remark 3.12 and Proposition 3.11 (iv) in the latter equality. Then the change of variable can be reformulated to

$$\begin{aligned}&f\big(x(t)\big)-f\big(x(0)\big)\\ &= \sum_{i=1}^n \int_0^t f_{x_i}\big(x(s-)\big)\,dx_i(s) + \sum_{0<s\le t}\Big(f\big(x(s)\big)-f\big(x(s-)\big)-\sum_{i=1}^n f_{x_i}\big(x(s-)\big)\Delta x_i(s)\Big)\end{aligned}$$

   which is the form of Ito formula for semimartingales (stochastic calculus).

2. For $s<t<T$ we have

$$f\big(x(t)\big)=f\big(x(s)\big)+\sum_{i=1}^n\int_s^t f_{x_i}\big(x(u)\big)\,dx_i^c(u)+\sum_{s<u\le t}\Big(f\big(x(u)\big)-f\big(x(u-)\big)\Big)$$

   and

$$f\big(x(t)\big)=f\big(x(T)\big)-\sum_{i=1}^n\int_t^T f_{x_i}\big(x(u)\big)\,dx_i^c(u)-\sum_{t<u\le T}\Big(f\big(x(u)\big)-f\big(x(u-)\big)\Big).$$

For the applications in the next chapters we use the following speciale version of Change of variable formula.

Corollary 4.5. *Let $f^j:\mathbb{R}^n\to\mathbb{R}$ be a C^1-function for $j=1,2,\ldots,m$ and let z be a piecewise constant function taking values in $\{1,2,\ldots,m\}$. Let $x=(x_1,x_2,\ldots,x_n)$ be an $\mathbb{R}^n$-valued function where each component is an FV-function such that $D_x\subseteq D_z$ where D_x and D_z are the sets of jump times for x and z respectively, that is, $D_x=\cup_{i=1}^n\{t>0:\Delta x_i(t)\neq 0\}$ and $D_z=\{t>0:\Delta z(t)\neq 0\}$. Then*

$$f^{z(t)}\big(x(t)\big)-f^{z(0)}\big(x(0)\big)=\sum_{i=1}^n\int_0^t f_{x_i}^{z(s)}\big(x(s)\big)\,dx_i^c(s)+\int_0^t\Big(f^{z(s)}\big(x(s)\big)-f^{z(s-)}\big(x(s-)\big)\Big)\,dn(s)$$

where $n(t)=\#\{0<s\le t|z(s-)\neq z(s)\}$. Or on the differential form

$$d\big(f^{z(t)}\big(x(t)\big)\big)=\sum_{i=1}^n f_{x_i}^{z(t)}\big(x(t)\big)\,dx_i^c(t)+\Big(f^{z(t)}\big(x(t)\big)-f^{z(t-)}\big(x(t-)\big)\Big)\,dn(t).$$

Proof. The jump times of z can be enumerable in increasing order by the assumption that z is a piecewise constant function and hence $n(t)$ is a well-defined increasing piecewise constant function. Note that $n(t)$ counts the number of jumps in the interval $[0,t]$ of z and that $dz^c(t)=0$. Then by Change of variable formula we get that

$$f^{z(t)}\big(x(t)\big)-f^{z(0)}\big(x(0)\big)=\sum_{i=1}^n\int_0^t f_{x_i}^{z(s)}\big(x(s)\big)\,dx_i^c(s)+\sum_{0<s\le t}\Big(f^{z(s)}\big(x(s)\big)-f^{z(s-)}\big(x(s-)\big)\Big).$$

Using that $D_x \subseteq D_z$ and $\Delta n(t) = 1$ for $t \in D_z$ and $\Delta n(t) = 0$ for $t \notin D_z$, the last term can be rewritten to

$$\begin{aligned}\sum_{0<s\le t}\Big(f^{z(s)}\big(x(s)\big) - f^{z(s-)}\big(x(s-)\big)\Big) &= \sum_{s\in D_z\cap(0,t]}\Big(f^{z(s)}\big(x(s)\big) - f^{z(s-)}\big(x(s-)\big)\Big)\cdot 1\\ &= \sum_{s\in D_z\cap(0,t]}\Big(f^{z(s)}\big(x(s)\big) - f^{z(s-)}\big(x(s-)\big)\Big)\Delta n(s)\\ &= \sum_{0<s\le t}\Big(f^{z(s)}\big(x(s)\big) - f^{z(s-)}\big(x(s-)\big)\Big)\Delta n(s)\\ &= \int_0^t \Big(f^{z(s)}\big(x(s)\big) - f^{z(s-)}\big(x(s-)\big)\Big)\, dn(s).\end{aligned}$$

In the latter equality we have used Proposition 3.11 (v). □

4.1. Exponential formula. As an application of the change of variables formula is Exponential formulas, also called the Doléans exponential. These solutions have many applications.

Theorem 4.6. (*Doléans exponential formula for FV-functions*). *Let x be an FV-function. The unique solution of*

$$dy(t) = y(t-)\, dx(t) \quad y(0) = y_0 \tag{4.2}$$

is given by

$$y(t) = y_0 \exp\big(x^c(t)\big) \prod_{0<s\le t}\big(1 + \Delta x(s)\big).$$

Proof. Note by Integration by parts (Theorem 4.1)

$$x^2(t) = 2\int_0^t x(s-)\, dx(s) + \sum_{0<s\le t}(\Delta x(s))^2. \tag{4.3}$$

To verify that $y(t)$ given in the theorem is a solution, set $y_1(t) = \exp\big(x^c(t)\big)$ and $y_2(t) = \prod_{0<s\le t}\big(1 + \Delta x(s)\big)$. First note that y_1 is a continuous FV-function which dynamics is given in Exercise 5.8. Next note that there are only finite number of $s \in (0, t]$ such that $|\Delta x(s)| > 1/2$ and $\bar{y}(t) = \prod_{0<s\le t}\big(1 + \Delta x(s1_{\{|\Delta x(s)|>1/2\}})\big)$ is a piecewise constant function and hence an FV-function. Then we have that $\log(\tilde{y}_2(t)) = \sum_{0<s\le t}\log\big(1 + \Delta x(s)1_{\{|\Delta x(s)|\le 1/2\}}\big)$, which is an absolutely convergent series, since

$$\sum_{0<s\le t}|\log\big(1 + \Delta x(s)1_{\{|\Delta x(s)|\le 1/2\}}\big)| \le c\sum_{0<s\le t}(\Delta x(s)1_{\{|\Delta x(s)|\le 1/2\}})^2 \le c\sum_{0<s\le t}(\Delta x(s))^2 < \infty$$

because $|\log(1 + x)| \le cx^2$ when $|x| \le 1/2$ for some constant $c > 0$ and due to (4.3). Thus $\log(\tilde{y}_2(t))$ is a pure jump function and so is $\tilde{y}(t)$. By integration by parts $y_2(t) = \bar{y}_2(t)\tilde{y}_2(t)$ is a pure jump function and hence an FV-function. One again by integration by parts we have that

$$\begin{aligned} y(t) = y_1(t)y_2(t) &= 1 + \int_0^t y_1(s)\, dy_2(s) + \int_0^t y_2(s-)\, dy_1(s) \\ &= 1 + \sum_{0<s\le t} y_1(s)y_2(s-)\,\Delta x(s) + \int_0^t y_2(s-)y_2(s)\, dx^c(s) \\ &= 1 + \int_0^t y(s-)\, dx(s). \end{aligned}$$

Hence $y(t)$ is a solution of (4.2).

To prove the uniqueness, let $y^1(t)$ and $y^2(t)$ be two solutions of (4.2). Set $u(t) = y^1(t) - y^2(t)$ we get that $u(t) = \int_0^t u(s-)\, dx(s)$. Then we have that $|u(t)| \le \int_0^t |u(s-)|\, dV^x(s) \le \sup_{0\le s\le t} |u(s)|\, V^x(t)$ where V^x is the variation function of x. This inequality implies that

$$|u(t)| \le \int_0^t |u(s-)|\, dV^x(s) \le \sup_{0\le s\le t} |u(s)| \int_0^t V^x(s-)\, dV^x(s) \le \sup_{0\le s\le t} |u(s)|(V^x(t))^2/2.$$

where we have made use of (4.3) in the latter inequality. By induction we get that $|u(t)| \le \sup_{0\le s\le t} |u(s)|(V^x(t))^n/n!$ and letting $n \uparrow \infty$ we see that $u(t) = 0$ and the uniqueness follows. □

Remark 4.7. The solution $y(t)$ is called the Doléans-Dade exponential of x.

1. In integral form, equation (4.2) becomes $y(t) = y + \int_0^t y(s-)\, dx(s)$.
2. If x is a C^1-function given by $x(t) = x(0) + \int_0^t \varphi(s)\, ds$ where $\varphi(t)$ is a continuous function. Then equation (4.2) is an ordinary first-order linear differential equation with solution $y(t) = \exp(\int_0^t \varphi(s)\, ds)$
3. If $\Delta x(t_0) = -1$ then $y(t)) = 0$ for $t \ge t_0$.

5. Exercises

Exercise 5.1. This exercise is a proof of Proposition 1.4(v). Let x be a CADLAG function and let y be a CAGLAD function. Define

$$x_r^{(n)}(t) = \sum_{i=1}^{\infty} x\big(i2^{-n}\big) 1_{[(i-1)2^{-n}, i2^{-n})}(t) \tag{5.1}$$

$$y_r^{(n)}(t) = \sum_{i=1}^{\infty} y\big((i-1)2^{-n}\big) 1_{[(i-1)2^{-n}, i2^{-n})}(t) \tag{5.2}$$

$$x_l^{(n)}(t) = x(0)1_{\{0\}}(t) + \sum_{i=1}^{\infty} x\big(i2^{-n}\big) 1_{((i-1)2^{-n}, i2^{-n}]}(t) \tag{5.3}$$

$$y_l^{(n)}(t) = y(0)1_{\{0\}}(t) + \sum_{i=1}^{\infty} y\big((i-1)2^{-n}\big) 1_{((i-1)2^{-n}, i2^{-n}]}(t). \tag{5.4}$$

(a) Show that $x_r^{(n)}(t)$ and $y_r^{(n)}(t)$ are CADLAG.
(b) Show that $x_l^{(n)}(t)$ and $y_l^{(n)}(t)$ are CAGLAD.
(c) Show that $x(t) = \lim_{n\uparrow\infty} x_r^{(n)}(t) = \lim_{n\uparrow\infty} x_l^{(n)}(t)$.
(d) Show that $y(t) = \lim_{n\uparrow\infty} y_r^{(n)}(t) = \lim_{n\uparrow\infty} y_l^{(n)}(t)$.

Exercise 5.2. This exercise is related to Remark 2.4. Show that the function

$$x(t) = \begin{cases} t\sin(\pi/t) & \text{for } t > 0 \\ 0 & \text{for } t = 0 \end{cases}$$

is of infinite variation on $[0,2]$. Hint: For the partition $0, 2/(2n-1), 2/(2n-3), \dots, 2/5, 2/3, 2$ show that

$$V^x([0,2]) \geq \frac{2}{2n-1} + \Big(\frac{2}{2n-3} + \frac{2}{2n-1}\Big) + \cdots + \Big(\frac{2}{5} + \frac{2}{3}\Big) + \Big(2 + \frac{2}{3}\Big) > \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}.$$

Exercise 5.3. Let $x : [0,\infty) \to \mathbb{R}$ be a function.

(a) Show that

$$\begin{aligned}\sum_{i=1}^{n} |x(t_i) - x(t_{i-1})| &= \sum_{i=1}^{n} \big(x(t_i) - x(t_{i-1})\big)^+ + \sum_{i=1}^{k} \big(x(t_i) - x(t_{i-1})\big)^- \\ &= 2\sum_{i=1}^{n} \big(x(t_i) - x(t_{i-1})\big)^+ - \big(x(t) - x(0)\big).\end{aligned}$$

(b) Show that if one of $V^x(t)$, $V^x_+(t)$ or $V^x_-(t)$ is finite then they are all finite, that is, $V^x_+(t) < \infty$ if and only if $V^x(t) < \infty$ if and only if $V^x_-(t) < \infty$. Hint: use (2.2) and question (a).

Exercise 5.4. This exercise gives a proof the minimal property of the Jordan decomposition, Theorem 2.6. Let $0 \leq s < t$ be given and let $s = t_0 < t_1 < \cdots < t_{n-1} < t_n = t$ is a finite partition of $[s,t]$.

(a) Show that

$$V^x(t) - V^x(s) = \sup \sum_{i=1}^{n} |x(t_i) - x(t_{i-1})| \text{ and } V^x_\pm(t) - V^x_\pm(s) = \sup \sum_{i=1}^{n} \big(x(t_i) - x(t_{i-1})\big)^\pm$$

where the supremum is taken over all finite partitions of $[s,t]$.

(b) Show that $\big(x(t_i) - x(t_{i-1})\big)^\pm \leq a_\pm(t_i) - a_\pm(t_{i-1})$.

(c) Show that $V^x_\pm(t) - V^x_\pm(s) \leq x_\pm(t) - x_\pm(s)$.

Exercise 5.5. Consider the Lebesgue-Stieltjes integral with respect to an increasing function (see Definition 3.4). Let $a : [0,\infty) \to \mathbb{R}$ be an increasing CADLAG function (increasing FV-function) and let $f : (0,\infty) \to \mathbb{R}$ be a-integrable. Recall that the integral of f is defined by (see Definition 26.3)

$$\int_0^\infty f(t)\, da(t) = \int_0^\infty f^+(t)\, da(t) - \int_0^\infty f^-(t)\, da(t)$$

where the representation of $f = f^+ - f^-$ is used. The function f can be written in many ways as the difference between two positive (Borel) functions f_1 and f_2, that is, $f = f_1 - f_2$.

(a) Show that $f^+ \leq f_1$ and $f^- \leq f_2$.

(b) Show: If f_1 and f_2 are a-integrable then

$$\int_0^\infty f^+(t)\, da(t) - \int_0^\infty f^-(t)\, da(t) = \int_0^\infty f_1(t)\, da(t) - \int_0^\infty f_2(t)\, da(t).$$

Argue that the integral $\int_0^\infty f(t)\, da(t)$ is uniquely defined, that is, the integral does not depend on representation of f as the difference of two positive functions.

Consider the Lebesgue-Stieltjes integral with respect to a function of finite variation (see Definition 3.9). Let x be an FV-function and let $f : (0, \infty) \to \mathbb{R}$ be x-integrable. Recall that the integral of f with respect to x is defined by

$$\int_0^\infty f(t)\, dx(s) = \int_0^\infty f(t)\, dV_+^x(t) - \int_0^\infty f(t)\, dV_-^x(t)$$

where Jordan decomposition of x is used (see Theorem 2.6) and Proposition 2.9. The FV-function x can be written in many ways as the difference of two increasing CADLAG functions a_+ and a_-, that is, $x(t) = a_+(t) - a_-(t)$. Recall from Jordan decomposition that $V_\pm^x(t) - V_\pm^x(s) \leq a_\pm(t) - a_\pm(s)$.

(c) Show: If f is also a_+- and a_--integrable then

$$\int_0^\infty f(t)\, dV_+^x(t) - \int_0^\infty f(t)\, dV_-^x(t) = \int_0^\infty f(t)\, da_+(t) - \int_0^\infty f(t)\, da_-(t).$$

Argue that the integral $\int_0^\infty f(t)\, dx(t)$ is uniquely defined, that is, the integral does not depend on decomposition of x as the difference of two increasing functions.

Exercise 5.6. Compute the integral $\int_0^4 t\, dx(t)$ in the following cases:

(a) $x(t) = k$ for $k - 1 \leq t < k$, $k = 1, 2, \ldots$
(b) $x(t) = e^t$.
(c) $x(t) = k + e^t$ for $k - 1 \leq t < k$, $k = 1, 2, \ldots$
(d) $x(t) = ke^t$ for $k - 1 \leq t < k$, $k = 1, 2, \ldots$

Exercise 5.7. Let x and y be two FV-functions such that $y(t) \neq 0$ for all t. Show that

$$d\Big(\frac{x(t)}{y(t)}\Big) = \frac{y(t)\, dx(t) - x(t)\, dy(t)}{y(t)y(t-)} = \frac{y(t-)\, dx(t) - x(t-)\, dy(t)}{y(t)y(t-)}.$$

Exercise 5.8. Let x be a continuous FV-function. Show the following differential forms:

(a) $dx^n(t) = nx^{n-1}(t)\, dx(t)$, $n \geq 1$.
(b) $de^{x(t)} = e^{x(t)}\, dx(t)$.

Assume that φ is locally integrable (with respect to Lebesgue measure).

(c) $d\exp\Big(\int_0^t \varphi(s)\, ds\Big) = \varphi(t)\exp\Big(\int_0^t \varphi(s)\, ds\Big)\, dt$.

CHAPTER II

Stochastic processes

In this chapter we will study the stochastic processes used for modelling in the insurance and statistical part of the note. All the life insurance and statistical models in this note are in continuous time. In what follows $(\Omega, \mathcal{F}, \mathbf{P})$ denotes a given probability space.

6. General theory of stochastic processes

This is a short introduction to a few concepts and terminology concerning the general theory of stochastic processes. Moreover basic notations are introduced which are used in the sequel.

6.1. Stochastic processes. A stochastic process is a mathematical model for the occurrence, at each moment after the initial time, of a random phenomenon. Thus, a stochastic process is a family of stochastic variables indexed by time. In this note the state space $\mathcal{S}$ is $\mathbb{R}$, $\mathbb{Z}$, $\mathbb{R}^d$, or a subset of those.

Definition 6.1. A continuous-time stochastic process with state space $\mathcal{S}$ is a family of stochastic variables $(X(t), t \geq 0)$ where $X(t)$ takes values in $\mathcal{S}$ for all $t \geq 0$.

There are three ways of looking at stochastic process.

- For fixed t, $\omega \mapsto X(t, \omega)$ as a stochastic variable (the definition of stochastic processes).
- For fixed ω, $t \mapsto X(t, \omega)$ as a function of t, also denoted as the sample path of the process. (A process takes functions as values whereas a random variable takes numbers as values).
- A map $(t, \omega) \mapsto X(t, \omega)$ from $[0, \infty) \times \Omega$ into the state space $\mathcal{S}$.

A stochastic process $X(t)$ is called

- Continuous, if all sample paths are continuous functions.
- CADLAG, if all sample paths are CADLAG functions (see Definition 1.1).
- Of finite variation if all the sample are functions of finte variation (see Definition 2.2).

If the process $X(t)$ is CADLAG, we can define two other processes by

- $X(t-) = \lim_{s \uparrow t} X(s)$ for $t > 0$ and set $X(0-) = X(0)$.
- $\Delta X(t) = X(t) - X(t-)$ for $t > 0$ and set $\Delta X(0) = 0$.

Then the process $X(t-)$ is CAGLAD (see Definition 1.1) and $\Delta X(t)$ is called the jump process.

6.2. Filtrations and stopping times. The definition of a stochastic process has the feature of a flow of time, in which, at any time $t \geq 0$, we can talk about a past, present, and future. Filtrations are heuristically the collection of events which may occur before or at time t.

Definition 6.2. A filtration $\mathcal{F}(t)$ is an increasing family of sub-σ-algebras of $\mathcal{F}$. That is, for each t there is a σ-algebra $\mathcal{F}(t)$ and $\mathcal{F}(s) \subseteq \mathcal{F}(t)$ for $0 \leq s < t < \infty$. Moreover

- $\mathcal{F}(t-) = \sigma(\cup_{s<t} \mathcal{F}(s))$ is the information strictly prior to t,
- $\mathcal{F}(t+) = \cap_{\varepsilon > 0} \mathcal{F}(t + \varepsilon)$ is the information immediately after t.

Note that $\mathcal{F}^{\pm}(t) = \mathcal{F}(t\pm)$ are also filtrations. The filtration $\mathcal{F}(t)$ is right-continuous if $\mathcal{F}(t) = \mathcal{F}^{+}(t)$ holds for all t and $\mathcal{F}^{+}(t)$ is the smallest right-continuous filtration containing $\mathcal{F}(t)$. The natural filtration of a stochastic process $X(t)$ is the simplest choice of a filtration that is generated by the process itself and is denoted by $\mathcal{F}^X(t) = \sigma(X(s)|0 \le s \le t)$.

Remark 6.3. A filtration $\mathcal{F}(t)$ satisfies the *usual conditions* (les conditions habituelles) if it is right-continuous and $\mathcal{F}(0)$ contains all $\mathbf{P}$-negligible events in $\mathcal{F}$, that is, if $A \subseteq B \in \mathcal{F}$ such that $\mathbf{P}(B) = 0$ then $A \in \mathcal{F}(0)$.

Definition 6.4. A stochastic process $X(t)$ is adapted to the filtration $\mathcal{F}(t)$ if, $X(t)$ is an $\mathcal{F}(t)$-measurable stochastic variable for each t.

Note that any process $X(t)$ is adapted to its natural filtration $\mathcal{F}^X(t)$. In fact, $\mathcal{F}^X(t)$ is the smallest filtration to which $X(t)$ is adapted.

Definition 6.5. A $[0, \infty]$-valued random variable τ is a stopping time with respect to the filtration $\mathcal{F}(t)$ if

$$\{\tau \le t\} \in \mathcal{F}(t)$$

for all $t \ge 0$. Moreover

$$\mathcal{F}(\tau) = \{A \in \mathcal{F} | A \cap \{\tau \le t\} \in \mathcal{F}(t), \text{ for all } t \ge 0\} \tag{6.1}$$

is a σ-algebra containing all information at the stopping time τ.

One has to show that $\mathcal{F}(\tau)$ is a σ-algebra. Indeed, $\emptyset \in \mathcal{F}(\tau)$ and $A^c \cap \{\tau \le t\} = \{\tau \le t\} \setminus (A \cap \{\tau \le t\})$ it follows that $A^c \in \mathcal{F}(\tau)$ as τ is a stopping time. For for sequence of set $A_n \in \mathcal{F}(\tau)$ then $(\cup_{n=1}^{\infty} A_n) \cap \{\tau \le t\} = \cup_{n=1}^{\infty}(A_n \cap \{\tau \le t\}) \in \mathcal{F}(t)$ for all $t \ge 0$ hence $\cup_{n=1}^{\infty} A_n \in \mathcal{F}(\tau)$.

Lemma 6.6. *Let τ be a $[0, \infty]$-valued random variable and let Z be a random variable. Then*

(i) *τ is a stopping time with respect to a filtration $\mathcal{F}(t)$ if and only if the process $t \mapsto \tau 1_{[\tau,\infty)}(t)$ is adapted to $\mathcal{F}(t)$ and $\{\tau = 0\} \in \mathcal{F}(0)$.*

(ii) *If τ is a stopping time with respect to a filtration $\mathcal{F}(t)$. Then Z is measurable with respect to $\mathcal{F}(\tau)$ if and only if the process $t \mapsto Z1_{[\tau,\infty)}(t)$ is adapted to $\mathcal{F}(t)$.*

(iii) *$\mathcal{F}(t) = \sigma(1_{\{\tau=0\}}, \tau 1_{[\tau,\infty)}(t), Z1_{[\tau,\infty)}(t))$ is the smallest filtration such that τ is a stopping time and Z is measurable with respect to $\mathcal{F}(\tau)$.*

Proof. Let $t \ge 0$ be given and fixed.

(i): Assume that τ is a stopping time. Then $\{\tau = 0\} = \{\tau \le 0\} \in \mathcal{F}(0)$ and $\{\tau 1_{[\tau,\infty)}(t) \le 0\} = \{\tau = 0\} \cup \{\tau > t\} \in \mathcal{F}(t)$. Moreover, for $0 < a < t$ we have that $\{\tau 1_{[\tau,\infty)}(t) \le a\} = \{\tau 1_{[\tau,\infty)}(t) = 0\} \cup \{0 < \tau 1_{[\tau,\infty)}(t) \le a\} = \{\tau 1_{[\tau,\infty)}(t) = 0\} \cup \{0 < \tau \le a\} \in \mathcal{F}(t)$ and for $a \ge t$ we have that $\{\tau 1_{[\tau,\infty)}(t) \le a\} = \Omega$. Hence $t \mapsto \tau 1_{[\tau,\infty)}(t)$ is an adapted process. Assume now that $t \mapsto \tau 1_{[\tau,\infty)}(t)$ is adapted and $\{\tau = 0\} \in \mathcal{F}(0)$. Then $\{\tau \le t\} = \{\tau = 0\} \cup \{0 < \tau \le t\} = \{\tau = 0\} \cup \{0 < \tau 1_{[\tau,\infty)}(t) \le t\}$ and τ is a stopping time.

(ii): By the definition $\mathcal{F}(\tau)$ in (6.1) we have that $A \in \mathcal{F}(\tau)$ if and only if the process $t \mapsto 1_A \cdot 1_{[\tau,\infty)}(t)$ is adapted. The result can be proved via the Monotone Class Theorem (Theorem 25.15).

(iii): First, we have to show that $\mathcal{F}(t) = \sigma(1_{\{\tau=0\}}, \tau 1_{[\tau,\infty)}(t), Z1_{[\tau,\infty)}(t))$ is a filtration, that is, we have to show that $\mathcal{F}(s) \subseteq \mathcal{F}(t)$ for $0 \le s < t$. If we know the value of $\tau 1_{[\tau,\infty)}(t)$ and can determine wether $\tau = 0$ or $\tau > 0$ the we also know the value of $\tau 1_{[\tau,\infty)}(s)$. Finally, $Z1_{[\tau,\infty)}(s) = (Z1_{[\tau,\infty)}(t))1_{[\tau,\infty)}(s)$ is $\mathcal{F}(t)$-measurable since the two terms are $\mathcal{F}(t)$-measurable.

Hence $\mathcal{F}(s) \subseteq \mathcal{F}(t)$ and $\mathcal{F}(t)$ is a filtration. By (i) and (ii) the filtration is the smallest filtration such that τ is a stopping time and Z is measurable with respect to $\mathcal{F}(\tau)$. $\square$

6.3. Progressive and predictable measurability. Implicit in the definition of a stochastic process is the assumption that each $X(t)$ is measurable. But a stochastic process is really a function of the pair of variables (t, ω) and it is convenient to have some measurability properties.

Definition 6.7. The stochastic process $X(t)$ is measurable if $(t, \omega) \mapsto X(t, \omega)$ is measurable with respect to $\mathcal{B}([0, \infty)) \otimes \mathcal{F}$.

As an immediate consequence of Lemma 26.9, all the sample paths of a measurable process are Borel functions. But the process does not need to be adapted. The following definition relates measurability in t and ω with the filtration.

Definition 6.8. Let $\mathcal{F}(t)$ be a filtration and define the progressive σ-algebra by

$$\Pr(\mathcal{F}) = \{A \in \mathcal{B}([0, \infty)) \otimes \mathcal{F} \mid A \cap ([0, s] \times \Omega) \in \mathcal{B}([0, s]) \otimes \mathcal{F}(s) \text{ for all } s \geq 0\}.$$

A stochastic process $X(t)$ is progressive if the mapping $(t, \omega) \mapsto X(t, \omega)$ is measurable with respect to $\Pr(\mathcal{F})$.

One has to show that $\Pr(\mathcal{F})$ is a σ-algebra.

Lemma 6.9. *A stochastic process $X(t)$ is progressive if and only if the map $(t, \omega) \mapsto X(t, \omega)$ restricted to $[0, s] \times \Omega$ is measurable with respect to $\mathcal{B}([0, s]) \otimes \mathcal{F}(s)$ for every $s \geq 0$.*

Proof. Let $B \in \mathcal{B}(S)$ and $s \geq 0$ be fixed and given. From the observation that

$$\{t \in [0, \infty), \omega \in \Omega \mid X(t, \omega) \in B\} \cap ([0, s] \times \Omega) = \{(t \in [0, s], \omega \in \Omega) \mid X(t, \omega) \in B\}$$

the result follows. $\square$

Thus, the process $X(t)$ is progressive if the restriction of the process X to $[0, s] \times \Omega$ is $\mathcal{B}([0, s]) \otimes \mathcal{F}(s)$-measurable for each s. By Lemma 6.9 and Lemma 26.9, a progressive process is automatically adapted, but the reverse implication does not hold. An important example of a progressive process is provided by the following lemma.

Lemma 6.10. *If an adapted process $X(t)$ is CADLAG or CAGLAD then the process $X(t)$ is progressive.*

Proof. Since the proof is similar, assume that $X(t)$ is CADLAG. For $n \in \mathbb{N}$, define the process

$$X^{(n)}(t) = \sum_{i=1}^{\infty} X(i2^{-n}) 1_{[(i-1)2^{-n}, i2^{-n})}(t).$$

By Exercise 5.1, we have that $\lim_{n \uparrow \infty} X^{(n)}(t) = X(t)$. For a fixed and given $s \geq 0$, the mapping $(t, \omega) \mapsto X^{(n)}(t, \omega)$ on $[0, s] \times \Omega$ is measurable with respect to $\mathcal{B}([0, s]) \otimes \mathcal{F}(s)$ since $X(t)$ is adapted. Hence $(t, \omega) \mapsto X(t, \omega)$ on $[0, s] \times \Omega$ is also measurable with respect to $\mathcal{B}([0, s]) \otimes \mathcal{F}(s)$. By Lemma 6.9 $X(t)$ is a progressive process. $\square$

Lemma 6.11. *Let $(X(t), t \geq 0)$ be a progressive process. If τ is a stopping time then the stopped process $(X(\tau \wedge t), t \geq 0)$ is progressive and the random variable $X(\tau) 1_{[0, \infty)}(\tau)$ is measurable with respect to $\mathcal{F}(\tau)$.*

Definition 6.12. Let $\mathcal{F}(t)$ be a filtration and let the elementary predictable sets be

$$\mathcal{E} = \{\{0\} \times F \mid F \in \mathcal{F}(0)\} \cup \{(s,t] \times F \mid F \in \mathcal{F}(s), 0 \leq s < t < \infty\}.$$

The predictable σ-algebra $\mathcal{P}(\mathcal{F})$ is the σ-algebra that is generated by the elementary predictable sets, that is, $\mathcal{P}(\mathcal{F}) = \sigma(\mathcal{E})$. A stochastic process $X(t)$ is predictable if the mapping $(t,\omega) \mapsto X(t,\omega)$ is measurable with respect to the predictable σ-algebra $\mathcal{P}(\mathcal{F})$.

The next proposition gives conditions for a process to be predictable.

Proposition 6.13. *The process $X(t)$ is predictable if one of the following conditions is satisfied:*
(i) *$X(t)$ is an adapted CAGLAD process.*
(ii) *$X(t)$ is a measurable (Borel) deterministic process.*
(iii) *$X(t)$ is a Borel-measurable function of a predictable process.*

Proof. (i): It can proved via the Monotone Class Theorem (Theorem 25.15) that the processes

$$\begin{aligned} X(t) &= X \cdot 1_{\{0\}}(t), \quad X \text{ is } \mathcal{F}(0)\text{-measurable} \\ X(t) &= X \cdot 1_{(u,v]}(t), \quad X \text{ is } \mathcal{F}(u)\text{-measurable} \end{aligned}$$

are predictable. The left-continuous process $X(t)$ is approxima by (see Exercise 5.1)

$$X(t) = \lim_{n\uparrow\infty} X^{(n)}(t) = \lim_{n\uparrow\infty} \Big(X(0)1_{\{0\}}(t) + \sum_{i=1}^{\infty} X\big((i-1)2^{-n}\big) 1_{((i-1)2^{-n}, i2^{-n}]}(t) \Big).$$

This is a limit of predictable processes.
(ii): Again, it can be proved via the Monotone Class Theorem (Theorem 25.15).
(iii): A measurable function of a measurable function is again measurable. □

Corollary 6.14. *If $X(t)$ is an adapted CADLAG process, then $X(t)$ is a progressive process and $X(t-)$ is a predictable process. Moreover, if $X(t)$ is predictable, then $\Delta X(t)$ is predictable.*

Next result gives the hierarchy of measurability. First observe that the elementary predictable sets are in the σ-algebra

$$\mathcal{E} \subseteq \{A \in \mathcal{B}([0,\infty)) \otimes \mathcal{F} \mid A \cap ([0,s] \times \Omega) \in \mathcal{B}([0,s]) \otimes \mathcal{F}(s-) \text{ for all } s \geq 0\}.$$

Hence

$$\begin{aligned} \mathcal{P}(\mathcal{F}) &\subseteq \{A \in \mathcal{B}([0,\infty)) \otimes \mathcal{F} \mid A \cap ([0,s] \times \Omega) \in \mathcal{B}([0,s]) \otimes \mathcal{F}(s-) \text{ for all } s \geq 0\} \\ &\subseteq \Pr(\mathcal{F}) \subseteq \mathcal{B}([0,\infty)) \otimes \mathcal{F} \end{aligned}$$

and we have the following proposition.

Proposition 6.15. *Let $\mathcal{F}(t)$ be a filtration.*
(i) *If $X(t)$ is a predictable process then $X(t)$ is adapted to the filtration $\mathcal{F}^-(t) = \mathcal{F}(t-)$. Particularly, $X(t)$ is adapted to the filtration $\mathcal{F}(t)$.*
(ii) *If $X(t)$ is a predictable process then $X(t)$ is also progressive.*
(iii) *If $X(t)$ is a progressive process then $X(t)$ is adapted to the filtration $\mathcal{F}(t)$.*
(iv) *If $X(t)$ is a progressive process then $X(t)$ is also measurable.*

Remark 6.16. Note that we have the following.
1. The first property in the proposition provides some motivation for the terminology "predictable": $X(t)$ is known strictly prior to t.
2. Since a predictable or progressive process is a measurable process, this ensures that the sample paths are Borel functions.

7. Markov processes

For modeling, Markov processes are mathematical tractable and allows for computations of for example reserves in the multi-state contract.

Definition 7.1. An adapted stochastic process $X(t)$ is a Markov process with respect to the filtration $\mathcal{F}(t)$ if

$$\mathbf{P}\big(X(t) \in A \big| \mathcal{F}(s)\big) = \mathbf{P}\big(X(t) \in A \big| X(s)\big) \qquad \text{(Markov property)}$$

for all $0 \leq s < t$ and all $A \in \mathcal{B}(\mathbb{R})$.

A process is Markov, intuitively speaking, if today to make a prediction on what is going to happen in the future, it is useless to know anything more about the whole past up to today than the present state.

In the insurance part we study present values of future payments, and therefore we need other formulations of the Markov property. Let $\mathcal{F}^t = \sigma(X(s)|s \geq t)$ be the information generated by the process in the future. The result has the interpretation that the future depends on the past only through the present.

Proposition 7.2. *Let $X(t)$ be a Markov process. If Y is a bounded $\mathcal{F}^t$ measurable random variable, then*

$$\mathbf{E}[Y|\mathcal{F}(t)] = \mathbf{E}[Y|X(t)].$$

Proof. Let $0 \leq t < t_1$ be given and fixed. The Markov property reads as

$$\mathbf{E}[1_F|\mathcal{F}(t)] = \mathbf{E}[1_F|X(t)] \text{ for } F = \{X(t_1) \in A\} \in \sigma(X(t_1)).$$

By Proposition 28.1, this implies that $\mathbf{E}[Z|\mathcal{F}(t)] = \mathbf{E}[Z|X(t)]$ for any bounded $\sigma(X(t_1))$-measurable random variable Z. Consider a random variable Y on the form $Y = Y_1Y_2$ where Y_i, $i = 1, 2$, is a bounded $\sigma(X(t_i))$-measurable random variable and $t \leq t_1 < t_2$. Then

$$\begin{aligned}
\mathbf{E}[Y|\mathcal{F}(t)] &= \mathbf{E}[Y_1Y_2|\mathcal{F}(t)] = \mathbf{E}\big[Y_1\mathbf{E}[Y_2|\mathcal{F}(t_1)]\big|\mathcal{F}(t)\big] \\
&= \mathbf{E}\big[Y_1\mathbf{E}[Y_2|X(t_1)]\big|\mathcal{F}(t)\big] = \mathbf{E}\big[Y_1\mathbf{E}[Y_2|X(t_1)]\big|X(t)\big] \\
&= \mathbf{E}\big[Y_1\mathbf{E}[Y_2|\mathcal{F}(t_1)]\big|X(t)\big] = \mathbf{E}[Y_1Y_2|X(t)] \\
&= \mathbf{E}[Y|X(t)].
\end{aligned}$$

By induction one has that $\mathbf{E}[Y|\mathcal{F}(t)] = \mathbf{E}[Y|X(t)]$ for Y on the form $Y = \prod_{i=1}^n Y_i$ where Y_i is a bounded $\mathcal{F}(t_i)$-measurable random variable and $t \leq t_1 < \cdots < t_n$. Thus, the set $\mathcal{C}$ given by

$$\Big\{\prod_{i=1}^n Y_i \,\Big|\, Y_i \text{ bounded } \sigma(X(t_i))\text{-measurable, } i = 1, \ldots, n,\ t \leq t_1 \leq \cdots \leq t_n,\ n = 1, 2, \ldots\Big\}$$

is a multiplicative class contained in the set

$$\mathcal{H} = \{Y|Y \text{ bounded random variable and } \mathbf{E}[Y|\mathcal{F}(t)] = \mathbf{E}[Y|X(t)]\}.$$

Since the set $\mathcal{H}$ is a Monotone Vector space and $\sigma(\mathcal{C}) = \mathcal{F}^t$ we have that the Monotone Class Theorem (Theorem 25.15) gives the result. ∎

8. Finite variation processes

As it is discussed in Section 14 below, the sample paths of a payment process (or function) is of finite variation. Thus processes of finite variation (FV-process), that is, the sample paths are of finite variation are building blocks for the theory of payment streams for an insurance policy. Below is the formal definition.

Definition 8.1. A process $X(t)$ is called an FV-process if

(i) $X(t)$ is adapted to a given filtration.
(ii) The sample paths $t \mapsto X(t)$ are FV-functions (see Definition 2.8).

By Proposition 2.3, if $X(t)$ and $Y(t)$ are FV-processes then $\alpha X(t) + \beta Y(t)$ is an FV-process where α and β are two constants.

Definition 8.2. An n-dimensional process $(X(t), t \geq 0) = ((X^1(t), \ldots, X^n(t)), t \geq 0)$ is called a n-dimensional FV-process if the component process $(X^i(t), t \geq 0)$ is an FV-process for $i = 1, \ldots, n$.

9. Finite variation martingales

The basic approach employed in the insurance part and in the statistical part is martingale methods. Let $\mathcal{F}(t)$ be a given filtration.

9.1. Basic martingale theory.

Definition 9.1. A process $M(t)$ is an $\mathcal{F}(t)$-martingale if

(i) $M(t)$ is adapted to the filtration.
(ii) $M(t)$ is integrable, that is, $\mathbf{E}[|M(t)|] < \infty$ for all t.
(iii) $\mathbf{E}[M(t)|\mathcal{F}(s)] = M(s)$ for all $0 \leq s \leq t$ (martingale property).

The process $M(t)$ is a submartingale if the latter condition is replaced by the inequality $\mathbf{E}[M(t)|\mathcal{F}(s)] \geq M(s)$ for all $0 \leq s \leq t$. With the inequality reversed, $M(t)$ is a supermartingale.

Remark 9.2. Taking expectation of the martingale property gives that

$$\mathbf{E}[M(t)] = \mathbf{E}[M(s)] = \mathbf{E}[M(0)].$$

This property will be used many times later in the note.

The next proposition gives a construction of martingales.

Proposition 9.3. *Let Y be an integrable random variable. Then $M(t) = \mathbf{E}[Y|\mathcal{F}(t)]$ is a martingale.*

Proof. See Exercise 13.1. □

9.2. FV-martingales. Martingales of finite variation are of partially interest in this note. For short notation we denote a martingale with finite variation as an FV-martingale and the formal definition is given below.

Definition 9.4. A process $M(t)$ is an FV-martingale if

(i) $M(t)$ is an integrable FV-process.
(ii) $\mathbf{E}[M(t)|\mathcal{F}(s)] = M(s)$ for all $0 \leq s \leq t$ (martingale property).

Let $M(t)$ be an integrable FV-process. Informally, we use $dM(t)$ for the increment $M((t+dt)-) - M(t-)$ over the infinitesimal interval $[t, t+dt)$. Then we have the following informal criterion: $M(t)$ is an FV-martingale if and only if $\mathbf{E}[dM(t)|\mathcal{F}(t-)] = 0$ for all t. Indeed, by adding up increments of $M(t)$ over small subintervals $[u, u+du)$ partitioning $[s+ds, t+dt) = (s,t]$ for $s < t$, we have informally that

$$\begin{aligned}\mathbf{E}[M(t)|\mathcal{F}(s)] - M(s) &= \mathbf{E}[M(t) - M(s)|\mathcal{F}(s)]\\ &= \mathbf{E}\Big[\int_s^t dM(u)\Big|\mathcal{F}(s)\Big]\\ &= \int_s^t \mathbf{E}[dM(u)|\mathcal{F}(s)]\\ &= \int_s^t \mathbf{E}\big[\mathbf{E}[dM(u)|\mathcal{F}(u-)]\big|\mathcal{F}(s)\big]\\ &= 0.\end{aligned}$$

The next result shows that only constant processes are continuous FV-martingales.

Proposition 9.5. *If $X(t)$ is a continuous FV-martingale, then $X(t)$ is a constant process, that is, $X(t) = X(0)$.*

Proof. By a localization argument (an argument based on stopping times) we can assume that $X(t)$ is a bounded process. The process $M(t) = X(t) - X(0)$ is again a continuous martingale with $M(0) = 0$. Integration by parts formula (Theorem 4.1) implies that

$$M^2(t) = 2\int_0^t M(s)\,dM(s).$$

Since $M(t)$ is continuous, it is predictable (Proposition 6.13) and the integral process is a martingale (see Proposition 10.5 below). Hence

$$\mathbf{E}[M^2(t)] = \mathbf{E}\Big[2\int_0^t M(s)\,dM(s)\Big] = 0$$

and so $M(t) = 0$ for any $t \geq 0$. □

Proposition 9.5 is a special case of the following result. Informally, the result is derived by the following argument. If $X(t)$ is a predictable FV-martingale we have that

$$0 = \mathbf{E}[dX(t)|\mathcal{F}(t-)] = dX(t)$$

and hence $X(t)$ is a constant (does not depend on time).

Proposition 9.6. *Let $X(t)$ be an FV-process. If $X(t)$ is predictable martingale then $X(t)$ is a constant process, that is, $X(t) = X(0)$.*

Proof. By Corollary 6.14 we have that the jump process $\Delta X(t)$ is predictable and by Proposition 10.5 below we have that the integral process

$$\int_0^t 1_{\{\Delta X(s)>0\}}\,dX(s) = \sum_{0<s\leq t} 1_{\{\Delta X(s)>0\}}\Delta X(s) \geq 0$$

is a martingale with mean zero. Therefore, the martingale cannot have positive jumps. The same argument shows that the martingale cannot have negative jumps and the martingale $X(t)$ must be continuous. By Proposition 9.5 we can conclude that $X(t)$ is a constant process. □

The two propositions will be applied in the insurance part with following formulation.

Corollary 9.7. *If $X(t)$ is a martingale with representation*

$$X(t) = X(0) + M(t) + B(t)$$

where $M(t)$ is an FV-martingale and $B(t)$ is a predictable FV-process, then $B(t)$ is a constant process. In the special case that $B(t) = \int_0^t H(s)\,ds$ where $H(t)$ is a progressive process. Then $H(t)$ is equal to zero.

Proof. We have that $B(t) = X(t) - X(0) - M(t)$ is a martingale. By Proposition 9.6, $B(t)$ is a constant process. Thus, in the case that $B(t)$ is the integral process $B(t) = \int_0^t H(s)\,ds$, the integral process is zero for all t and hence $H(t)$ also must be equal to zero for any t. □

9.3. Doob-Meyer decomposition: Predictable compensators. An FV-process can be decomposed into a martingale part and a predictable part. The decomposition is called the Doob-Meyer decomposition and is a deep result in martingale theory. The Doob-Meyer decomposition is in general formulated for submartingales (or the difference of two submartingales). Since an integrable increasing process is a submartingale and an integrable FV-process is the difference of two submartingales, the Doob-Meyer decomposition below is first formulated for increasing processes and then for FV-processes.

Theorem 9.8. (*Doob-Meyer decomposition for increasing processes*). *Assume that the filtration $\mathcal{F}(t)$ satisfies the usual conditions. Suppose $X(t)$ is an adapted increasing CADLAG process that is locally integrable (there is a sequence of stopping times $(\tau_n, n = 1, 2, \ldots)$, increasing to ∞, such that $\mathbf{E}[X(\tau_n) - X(0)] < \infty$ for every n). Then there is a unique predictable CADLAG increasing process $\Lambda^X(t)$ with $\Lambda^X(0) = 0$ such that $X(t) - \Lambda^X(t)$ is a local FV-martingale. Moreover, if $\mathbf{E}[X(t) - X(0)] < \infty$ or $\mathbf{E}[\Lambda^X(t)] < \infty$ then $X(t) - \Lambda^X(t)$ is a martingale.*

In other words, there exists a predictable CADLAG increasing process $\Lambda^X(t)$ with $\Lambda^X(0) = 0$ and a (local) FV-martingale $M(t)$ with $M(0) = 0$ such that increasing process $X(t)$ has the decomposition

$$X(t) = X(0) + M(t) + \Lambda^X(t).$$

In this note we avoid local martingales and we will always assume in what follows that we have enough integrability, $\mathbf{E}[X(t) - X(0)] < \infty$ for all t, such that the process $M(t)$ is a martingale. In this case note that Proposition 9.6 gives the uniqueness of the predictable compensator $\Lambda^X(t)$. Recall that an FV-process $X(t)$ of integrable variation ($\mathbf{E}[V^X(t)] < \infty$ for all t) can be decomposed into the difference of two (adapted CADLAG) integrable increasing processes. Then we can apply the Doob-Meyer decomposition for increasing processes on both increasing processes and get the following version of Doob-Meyer decomposition.

Theorem 9.9. (*Doob-Meyer decomposition for FV-processes*). *Assume that the filtration $\mathcal{F}(t)$ satisfies the usual conditions. Suppose $X(t)$ is an integrable FV-process ($\mathbf{E}[V^X(t)] < \infty$ for all t). Then there is a unique integrable predictable FV-process $\Lambda^X(t)$ with $\Lambda^X(0) = 0$ such that*

$$X(t) - \Lambda^X(t)$$

is an FV-martingale.

Again, $X(t)$ has the decomposition

$$X(t) = X(0) + M(t) + \Lambda^X(t)$$

where $M(t)$ is an FV-martingale with $M(0) = 0$.

Definition 9.10. The predictable FV-process $\Lambda^X(t)$ from Doob-Meyer decomposition is called the predictable compensator of $X(t)$.

Informally interpretation of the predictable compensator. The process $M(t)$ is an FV-martingale and we have informally that $0 = \mathbf{E}[dM(t)|\mathcal{F}(t-)] = \mathbf{E}[dX(t)|\mathcal{F}(t-)] - d\Lambda^X(t)$ where we use informally that $\Lambda^X(t)$ is predictable. Heuristically the infinitesimal form of the predictable compensator becomes the heuristic expression

$$d\Lambda^X(t) = \mathbf{E}[dX(t)|\mathcal{F}(t-)].$$

This informal form can also be motivated by a construction of the predictable compensator $\Lambda^X(t)$ as the following limit

$$\sum_{i=1}^{2^n} \mathbf{E}\big[X(i2^{-n}t) - X((i-1)2^{-n}t)\big|\mathcal{F}((i-1)2^{-n}t)\big] \xrightarrow[n\uparrow\infty]{\mathbf{P}} \Lambda^X(t) \tag{9.1}$$

where the limit is in probability.

As already used in the argument for extending Theorem 9.8 to Theorem 9.9, the predictable compensator is linear in the following sense.

Lemma 9.11. *Let $X_1(t)$ and $X_2(t)$ be two FV-processes of integrable variation and let $\Lambda^1(t)$ and $\Lambda^2(t)$ be the associated predictable compensators, respectively. Let α and β be two constants. Then the FV-process $X(t) = \alpha X_1(t) + \beta X_2(t)$ has predictable compensator given by $\Lambda(t) = \alpha\Lambda^1(t) + \beta\Lambda^2(t)$.*

Proof. See Exercise 13.3. □

Remark 9.12. By the martingale property of $M(t) = X(t) - \Lambda^X(t)$ we get that

$$\mathbf{E}[X(t) - X(s)|\mathcal{F}(s)] = \mathbf{E}[\Lambda^X(t) - \Lambda^X(s)|\mathcal{F}(s)] \text{ for } 0 \leq s < t. \tag{9.2}$$

Set $s = 0$ and take expectation then $\mathbf{E}[X(t) - X(0)] = \mathbf{E}[\Lambda^X(t)]$.

In this note we will be interested in the case that predictable intensities exists, that is, a so-called absolutely continuous case. By the short informal notation we have that the intensity process is given by $\lambda^X(t)\,dt = \mathbf{E}[dX(t)|\mathcal{F}(t-)]$. The formal definition is the following.

Definition 9.13. Let $X(t)$ be an FV-process with integrable variation. The FV-process $X(t)$ has intensity process $\lambda^X(t)$ if

(i) $\lambda^X(t)$ is a predictable process.
(ii) $\int_0^t |\lambda^X(s)|\,ds < \infty$ for all t.
(iii) $\Lambda^X(t) = \int_0^t \lambda^X(s)\,ds$ for all t.

In the special case that $X(t)$ is an increasing process, then $\lambda^X(t) \geq 0$ is a positive process.

Lemma 9.11 reads as follows for intensity processes.

Lemma 9.14. *Let $X_1(t)$ and $X_2(t)$ be two FV-processes with intensity processes $\lambda^1(t)$ and $\lambda^2(t)$, respectively. Let α and β be two constants. Then the FV-process $X(t) = \alpha X_1(t) + \beta X_2(t)$ has intensity process given by $\lambda(t) = \alpha\lambda^1(t) + \beta\lambda^2(t)$.*

Proof. See Exercise 13.3. □

9.4. Predictable variation and covariation processes. To be able to formulate the martingale central limit theorem for counting processes (see Section 11.5)—which we will use in the statistical part—we need the concept of variation processes, which results from the Doob-Meyer decomposition.

Let $M(t)$ be an square integrable FV-martingale that is $\mathbf{E}[M^2(t)] < \infty$ for all t. Then $M^2(t)$ also is an (integrable) FV-process. By the Doob-Meyer decomposition Theorem 9.9 there exists a unique predictable FV-process $\langle M\rangle(t)$ which is the predictable compensator of $M^2(t)$. The process $\langle M\rangle(t)$ is called the predictable variation process of $M(t)$ and $M^2(t) - \langle M\rangle(t)$ is a martingale. It follows that $\mathbf{E}\big[M^2(t) - M^2(0)\big] = \mathbf{E}\big[\langle M\rangle(t)\big]$. Informally arguments give that

$$\begin{aligned} d\big(M^2(t)\big) &= M^2((t+dt)-) - M^2(t-) \\ &= \big(M(t-) + dM(t)\big)^2 - M^2(t-) \\ &= 2M(t-)\,dM(t) + \big(dM(t)\big)^2 \end{aligned}$$

and then the predictable variation process (informally) is given by

$$\begin{aligned} d\langle M\rangle(t) &= \mathbf{E}\big[d\big(M^2(t)\big)\big|\mathcal{F}(t-)\big] \\ &= \mathbf{E}\big[\big(dM(t)\big)^2\big|\mathcal{F}(t-)\big] \\ &= \mathrm{var}(dM(t)|\mathcal{F}(t-)). \end{aligned}$$

Let $M_1(t)$ and $M_2(t)$ be two FV-martingales such that $\mathbf{E}[M_1(t)M_2(t)] < \infty$ for all t. The predictable covariation process $\langle M_1, M_2\rangle(t)$ of $M_1(t)$ and $M_2(t)$ is the unique predictable compensator of $M_1(t)M_2(t)$ such that $M_1(t)M_2(t) - \langle M_1, M_2\rangle(t)$ is a martingale. The predictable covariation process is bilinear and symmetric (as an ordinary covariance)

$$\begin{aligned} \langle aM_1 + bM_2, M_3\rangle(t) &= a\langle M_1, M_3\rangle(t) + b\langle M_2, M_3\rangle(t) \\ \langle M_1, M_2\rangle(t) &= \langle M_2, M_1\rangle(t). \end{aligned}$$

Note that $\langle M\rangle(t) = \langle M, M\rangle(t)$. By similar arguments as above, we have informally that

$$d\langle M_1, M_2\rangle(t) = \mathrm{cov}(dM_1(t), dM_2(t)|\mathcal{F}(t-)).$$

The predictable variation process can also be defined in terms of predictable variation processes. We have the following polarization identity

$$M_1(t)M_2(t) = \frac{1}{4}\Big(\big(M_1(t) + M_2(t)\big)^2 - \big(M_1(t) - M_2(t)\big)^2\Big)$$

and note that the two processes $M_1(t) + M_2(t)$ and $M_1(t) - M_2(t)$ are FV-martingales. By the polarization identity we get that

$$\langle M_1, M_2\rangle(t) = \frac{1}{4}\Big(\langle M_1 + M_2\rangle(t) - \langle M_1 - M_2\rangle(t)\Big).$$

In the statistical part we need to bound the probability of large values of $M(t)$ anywhere which we can use the predictable variation for.

Theorem 9.15. (*Lenglart's inequality*). *Let $M(t)$ be a square integrable FV-martingale and $\langle M\rangle(t)$ its predictable variation process. For any $\varepsilon > 0$ and any $\delta > 0$ then*

$$\mathbf{P}\Big(\sup_{0\le s\le t} |M(s)| > \varepsilon\Big) \le \frac{\delta}{\varepsilon^2} + \mathbf{P}\big(\langle M\rangle(t) > \delta\big).$$

10. Integral processes

For the present value of a payment process (FV-process) it is important to form an integral of one stochastic process with respect to another. In this note, it is the Lebesgue-Stieltjes integral defined ω-by-ω in the class of FV-processes. In this section, let $\mathcal{F}(t)$ be a given filtration.

10.1. Integral process and basic properties. Let $X(t)$ be an FV-process and let $H(t)$ be a progressive process. Let ω be fixed, then the sample path $t \mapsto X(t,\omega)$ is an FV-function and the sample path $t \mapsto X(t,\omega)$ is a Borel function (see Remark 6.16), then by Definition 3.9 we have the Lebesgue-Stieltjes integral with respect to the sample path of $H(t)$ with respect to the sample path of $X(t)$. We have then the following definition

Let $X(t)$ be an FV-process. We say that a progressive process $H(t)$ is locally X-integrable if $\int_0^t |H(t,\omega)|\, dV^X(t,\omega) = \int_0^t |H(t,\omega)|\, dV_+^X(t,\omega) + \int_0^t |H(t,\omega)|\, dV_-^X(t,\omega) < \infty$ for all $t > 0$ and all $\omega \in \Omega$. Then we define, ω-by-ω, the integral process of $H(t)$ with respect to $X(t)$ to be

$$Y(t,\omega) = \int_0^t H(s,\omega)\, dX(s,\omega)$$

the Lebesgue-Stieltjes integral.

By Proposition 3.14, the sample paths of the integral process have the following properties.

Proposition 10.1. *The integral process has the following properties.*

(i) *The sample paths* $t \mapsto \displaystyle\int_0^t H(s,\omega)\, dX(s,\omega)$ *are CADLAG.*

(ii) *The sample paths* $t \mapsto \displaystyle\int_0^t H(s,\omega)\, dX(s,\omega)$ *are continuous if* $t \mapsto X(t,\omega)$ *is continuous.*

(iii) *The sample paths* $t \mapsto \displaystyle\int_0^t H(s,\omega)\, dX(s,\omega)$ *have finite variation.*

By Proposition 3.10, the calculus of the integral process is the following.

Proposition 10.2. *Assume that all integrals below exists and are well defined, then we have the following properties where α and β are constants.*

(i) *If $H_1(t)$ and $H_2(t)$ are progressive processes and $X(t)$ is an FV-process then*

$$\int_0^t \big(\alpha H_1(s) + \beta H_2(s)\big)\, dX(s) = \alpha \int_0^t H_1(s)\, dX(s) + \beta \int_0^t H_1(s)\, dX(s).$$

(ii) *If $H(t)$ is a progressive process and $X_1(t)$ and $X_2(t)$ are two FV-processes then*

$$\int_0^t H(s)\, d\big(\alpha X_1(s) + \beta X_2(s)\big) = \alpha \int_0^t H(s)\, dX_1(s) + \beta \int_0^t H(s)\, dX_2(s).$$

The following proposition shows that the condition that $H(t)$ is progressive is made to insure that the integral process is adapted.

Proposition 10.3. *The integral process $Y(t) = \int_0^t H(s)\, dX(s)$ is adapted to the filtration $\mathcal{F}(t)$. Moreover, if $X(t)$ and $H(t)$ are predictable processes then the integral process $Y(t) = \int_0^t H(s)\, dX(s)$ is also predictable.*

Proof. Let $s \geq 0$ be fixed and given. Let $\mathcal{H}$ be the collection of bounded processes $H(t)$ which restricted to $[0,s]\times\Omega$ is measurable with respect to $\mathcal{B}([0,s])\otimes\mathcal{F}(s)$ satisfying $\int_0^s H(t)\,dX(t)$ is $\mathcal{F}(s)$-measurable. Then it can be shown that $\mathcal{H}$ is a Monotone Vector Space. Note that

$$\mathcal{C} = \{1_{(u,v]}(t)1_F \mid 0 \leq u < v \leq s, F \in \mathcal{F}(s)\}$$

is a multiplicative set and $\sigma(\mathcal{C}) = \mathcal{B}([0,s])\otimes\mathcal{F}(s)$. The processes in $\mathcal{C}$ are $\mathcal{B}([0,s])\otimes\mathcal{F}(s)$-measurable and

$$\int_0^s 1_{(u,v]}(t)1_F\,dX(t) = 1_F\big(X(v)-X(u)\big)$$

is $\mathcal{F}(s)$-measurable. Hence $\mathcal{C}\subseteq\mathcal{H}$ and Monotone Class Theorem (Theorem 25.15) gives that a bounded process $H(t)$ which restricted to $[0,s]\times\Omega$ is measurable with respect to $\mathcal{B}([0,s])\otimes\mathcal{F}(s)$ is contained in $\mathcal{H}$. If $H(t)$ is a bounded progressive process then by Lemma 6.9 $H(t)$ restricted to $[0,s]\times\Omega$ is measurable with respect to $\mathcal{B}([0,s])\otimes\mathcal{F}(s)$ for all $s\geq 0$. This implies that the integral process $\int_0^t H(s)\,dX(s)$ is adapted. If $H(t)\geq 0$ then by monotone convergence we have that

$$\int_0^t H(s)\,dX(s) = \lim_{n\uparrow\infty}\int_0^t \big(H(s)\wedge n\big)\,dX(s)$$

and measurability is preserved by pointwise convergence. The proposition is proved by means of $H(t) = H^+(t) - H^-(t)$.

Finally, assume that $X(t)$ and $H(t)$ are predictable. Set $Y(t) = \int_0^t H(s)\,dX(s)$ and by Proposition 10.1, the integral process is CADLAG. Then Corollary 6.14 implies that $Y(t-) = \int_{(0,t)} H(s)\,dX(s)$ is predictable and that the jump process $\Delta Y(t) = H(t)\Delta X(t)$ is also predictable. Hence, $Y(t) = Y(t-) + \Delta Y(t)$ is predictable. □

Finally, let us summing up the results in this section using Proposition 10.3, Proposition 10.1, and Proposition 3.14.

Proposition 10.4. *Let $X(t)$ be an FV-process and let $H(t)$ be a locally X-integrable progressive process. Then the integral process*

$$Y(t) = \int_0^t H(s)\,dX(s)$$

is an FV-process which is bilinear in $H(t)$ and $X(t)$ and has following properties.

(i) $\Delta Y(t) = H(t)\Delta X(t)$.

(ii) *Let $K(t)$ be a progressive process and provided that the integrals are well defined then*

$$\int_0^t K(s)\,dY(s) = \int_0^t K(s)H(s)\,dX(s).$$

10.2. FV-martingales as integrators. An important result is that the integral of a predictable process with respect to an FV-martingale is another FV-martingale. This shows how martingales, integral processes, and predictable processes are tightly linked. Informally, the following informal computations give some indication of how the following proposition can be derived. Since $H(t)$ is predictable then $H(t)$ is $\mathcal{F}(t-)$-measurable and we have in ial form that $\mathbf{E}[H(t)\,dM(t)|\mathcal{F}(t-)] = H(t)\,\mathbf{E}[dM(t)|\mathcal{F}(t-)] = 0$. By the informally criterion for an FV-martingale, $\int_0^t H(s)\,dM(s)$ is an FV-martingale. Note the crucial role of predictability in enabling one to bring $H(t)$ outside the conditional expectation. Here is the formal statement and more rigorous proof.

Proposition 10.5. *Let $M(t)$ be an FV-martingale and let $H(t)$ be a predictable process satisfying $\mathbf{E}\big[\int_0^t |H(s)|\,dV^M(s)\big] < \infty$ for every $t > 0$. Then the integral process $\int_0^t H(s)\,dM(s)$ is also an FV-martingale.*

Proof. Let $H(t) = 1_{(u,v]}(t)1_F$, $0 \le u < v$ and $F \in \mathcal{F}(u)$ be a simple predictable process. The associated integral process is given by

$$X(t) = \int_0^t H(s)\,dM(s) = 1_F\big(M(v \wedge t) - M(u \wedge t)\big).$$

The process is adapted and $\mathbf{E}[|X(t)|] < \infty$. For $0 \le s < t$ and $G \in \mathcal{F}(s)$ we have that

$$\mathbf{E}\big[\big(X(t) - X(s)\big)1_G\big] = \mathbf{E}\big[\big(M(v \wedge t) - M(v \wedge s)\big)1_{F\cap G}\big] = 0$$

by the definition of conditional expectation and shows that $X(t)$ is a martingale. An application of the Monotone Class Theorem (Theorem 25.15) completes the proof. □

10.3. Predictable compensator. The predictable compensator of an integral process over a predictable process $H(t)$ with respect to $X(t)$ is given in the following proposition. Let $\Lambda(t)$ be the predictable compensator for $X(t)$. Informally, computations of the differential form give that

$$\mathbf{E}[H(t)dX(t)|\mathcal{F}(t-)] = H(t)\mathbf{E}[dX(t)|\mathcal{F}(t-)] = H(t)d\Lambda(t)$$

which give us the predictable compensator of the integral process.

Proposition 10.6. *Let $X(t)$ be an FV-process of integrable variation with predictable compensator $\Lambda(t)$ and let $H(t)$ be a predictable process such that*

$$\mathbf{E}\Big[\int_0^t |H(s)|\,dV^X(s)\Big] < \infty \text{ or } \mathbf{E}\Big[\int_0^t |H(s)|\,dV^\Lambda(s)\Big] < \infty.$$

Then the integral process

$$\int_0^t H(s)\,d\big(X(s) - \Lambda(s)\big) = \int_0^t H(s)\,dX(s) - \int_0^t H(s)\,d\Lambda(s)$$

is an FV-martingale. Moreover, $\int_0^t H(s)\,d\Lambda(s)$ is the predictable compensator for the integral process $\int_0^t H(s)\,dX(s)$. In the special case that $X(t)$ has intensity process $\lambda(t)$, then

$$\int_0^t H(s)\,dX(s) - \int_0^t H(s)\lambda(s)\,ds$$

is an FV-martingale and $\int_0^t H(s)\,dX(s)$ has intensity process $H(t)\lambda(t)$.

Proof. The integrability condition ensures that the process

$$\int_0^t H(s)\,d\big(X(s) - \Lambda(s)\big) = \int_0^t H(s)\,dX(s) - \int_0^t H(s)\,d\Lambda(s)$$

is a martingale by Proposition 10.5. The latter process is a predictable FV-process and hence this is the Doob-Meyer decomposition. Thus, $\int_0^t H(s)\,d\Lambda(s)$ is the predictable compensator for $\int_0^t H(s)\,dX(s)$. □

For $0 \le s < t$ we have that

$$\int_0^t H(u)\, d\big(X(u) - \Lambda(u)\big) - \int_0^s H(u)\, d\big(X(u) - \Lambda(u)\big) = \int_s^t H(u)\, d\big(X(u) - \Lambda(u)\big)$$

and the martingale property implies that

$$\mathbf{E}\Big[\int_s^t H(u)\, dX(u)\Big|\mathcal{F}(s)\Big] = \mathbf{E}\Big[\int_s^t H(u)\, d\Lambda(u)\Big|\mathcal{F}(s)\Big]. \tag{10.1}$$

10.4. Predictable variation and covariation processes. Let $M(t)$ be an FV-martingale and use the informal notation

$$\begin{aligned} d\Big\langle \int_0^{\cdot} H(s)\, dM(s)\Big\rangle(t) &= \operatorname{var}\Big(d\Big(\int_0^t H(s)\, dM(s)\Big)\Big|\mathcal{F}(t-)\Big) \\ &= \operatorname{var}\Big(H(t)\, dM(t)\Big|\mathcal{F}(t-)\Big) \\ &= H^2(t)\ \operatorname{var}(dM(t)|\mathcal{F}(t-)) \\ &= H^2(t)\ d\langle M\rangle(t). \end{aligned}$$

Thus, we have the following result for integral processes.

Proposition 10.7. *Let $M(t)$ be a square integrable FV-martingale with predictable variation process $\langle M\rangle(t)$. Let $H(t)$ be a predictable process such that*

$$\mathbf{E}\Big[\int_0^t H^2(s)\, dV^{\langle M\rangle}(s)\Big] < \infty \ \text{ or } \mathbf{E}\Big[\Big(\int_0^t H(s)\, dV^{\langle M\rangle}(s)\Big)^2\Big] < \infty.$$

Then

$$\int_0^t H(s)\, dM(s)$$

is a square integrable FV-martingale and its predictable variation process is given by

$$\Big\langle \int_0^{\cdot} H(s)\, dM(s)\Big\rangle(t) = \int_0^t H^2(s)\, d\langle M\rangle(s).$$

For the predictable covariation process, we have a result corresponding to Proposition 10.7.

Proposition 10.8. *Let $M_1(t)$ and $M_2(t)$ be two FV-martingales such that $\mathbf{E}[M_1(t)M_2(t)] < \infty$. Let $H(t)$ and $K(t)$ be two predictable processes such that*

$$\mathbf{E}\Big[\int_0^t |H(s)K(s)|\, dV^{\langle M_1, M_2\rangle}(s)\Big] < \infty.$$

Then the predictable covariation process of the integral processes is given by

$$\Big\langle \int_0^{\cdot} H(s)\, dM_1(s), \int_0^{\cdot} K(s)\, dM_2(s)\Big\rangle(t) = \int_0^t H(s)K(s)\, d\langle M_1, M_2\rangle(s).$$

10.5. Different look-a-like Fubini results for integral processes. Let $X(t)$ be an FV-process and $H(t)$ a progressive process such that the integral process is well-defined. Due to that $X(t)$ is stochastic (depends on ω), it is not possible directly to apply Fubini Theorem (see Theorem 26.10). Indeed, the mean value $\mathbf{E}\big[\int_0^t H(s)\,dX(s)\big]$ is a number while $\int_0^t \mathbf{E}[H(s)]\,dX(s)$ is stochastic. However, if we assume that either the integrand or the integrator is deterministic, then there are results that gives condition for changing the order of integrations.

Before we state the condition for changing the order of integrations, we need a preliminary technical lemma.

Lemma 10.9. *Let $H(t)$ be a measurable process, that is, $H(t)$ is $\mathcal{B}([0,\infty))\otimes\mathcal{F}$ (see Definition 6.7) and let $\mathcal{G}\subseteq\mathcal{F}$ be a sub σ-algebra of $\mathcal{F}$. Consider $t\mapsto\mathbf{E}[H(t)|\mathcal{G}]$ as a stochastic process. Then $\mathbf{E}[H(t)|\mathcal{G}]$ admits a version which is $\mathcal{B}([0,\infty))\otimes\mathcal{G}$.If $H(t)$ furthermore is CADLAG then $\mathbf{E}[H(t)|\mathcal{G}]$ is also CADLAG.*

If the integrator is deterministic then we have the well known Fubini's theorem (Theorem 26.10) and moreover the conditional Fubini theorem stated in theorem below. We can change the (conditional) expectation into the integrand.

Theorem 10.10. *Let x be a deterministic FV-function (see Definition 2.8) and let $H(t)$ be a progressive process satisfying $\mathbf{E}\big[\int_0^t |H(s)|\,dV^x(s)\big]=\int_0^t\mathbf{E}[|H(s)|]\,dV^x(s)<\infty$ for all $t>0$. Then we have Fubini theorem*

$$\mathbf{E}\Big[\int_0^t H(s)\,dx(s)\Big]=\int_0^t\mathbf{E}[H(s)]\,dx(s). \tag{10.2}$$

Let $\mathcal{G}\subseteq\mathcal{F}$ be a sub σ-algebra of $\mathcal{F}$ then we have conditional Fubini

$$\mathbf{E}\Big[\int_0^t H(s)\,dx(s)\Big|\mathcal{G}\Big]=\int_0^t\mathbf{E}[H(s)|\mathcal{G}]\,dx(s). \tag{10.3}$$

Proof. The first result (10.2) is the usual Fubini's theorem, Theorem 26.10. In the latter result (10.3) using the usual Theorem 26.10 from (10.2), we find for each $A\in\mathcal{G}$ that

$$\begin{aligned}\mathbf{E}\Big[\int_0^t H(s)\,dx(s)1_A\Big]&=\int_0^t\mathbf{E}[H(s)1_A]\,dx(s)=\int_0^t\mathbf{E}\big[\mathbf{E}[H(s)|\mathcal{G}]1_A\big]\,dx(s)\\&=\mathbf{E}\Big[\int_0^t\mathbf{E}[H(s)|\mathcal{G}]\,dx(s)1_A\Big]\end{aligned}$$

which gives the result. □

If the integrand is deterministic then we can change the (conditional) expectation condition to the integrator.

Theorem 10.11. *Let $X(t)$ be an FV-process and let $f(t)$ be a measurable (Borel) function satisfying $\mathbf{E}\big[\int_0^t|f(s)|\,dV^X(s)\big]=\int_0^t\mathbf{E}[|f(s)|]\,dV^X(s)<\infty$ for all $t>0$. Then we have that*

$$\mathbf{E}\Big[\int_0^t f(s)\,dX(s)\Big]=\int_0^t f(s)\,d\mathbf{E}[X(s)]. \tag{10.4}$$

Let $\mathcal{G}\subseteq\mathcal{F}$ be a sub σ-algebra of $\mathcal{F}$ then we have σ-algebra of $\mathcal{F}$ then

$$\mathbf{E}\Big[\int_0^t f(s)\,dX(s)\Big|\mathcal{G}\Big]=\int_0^t f(s)\,d\mathbf{E}[X(s)|\mathcal{G}]. \tag{10.5}$$

Proof. For the first result (10.4), note that for $f(t) = 1_{(a,b]}(t)$ we have that

$$\mathbf{E}\Big[\int_0^t f(s)\,dX(s)\Big] = \mathbf{E}[X(b)] - \mathbf{E}[X(a)] = \int_0^t f(s)\,d\mathbf{E}[X(s)].$$

Then the standard argument using Monotone class theorem we get the result.

In the latter result (10.5) using the first result (10.4), we find for each $A \in \mathcal{G}$ that

$$\begin{aligned}\mathbf{E}\Big[\int_0^t f(s)\,dX(s)1_A\Big] &= \int_0^t f(s)\,d\mathbf{E}[X(s)1_A] \\ &= \int_0^t f(s)\,d\mathbf{E}\big[\mathbf{E}[X(s)|\mathcal{G}]1_A\big] \\ &= \mathbf{E}\Big[\Big(\int_0^t f(s)\,d\mathbf{E}[X(s)|\mathcal{G}]\Big)1_A\Big]\end{aligned}$$

which gives the result. $\square$

11. Counting processes and point processes

Suppose there is an (particular) event occurring repeatedly and random in time. A counting process is a stochastic process that counts the number of the given event that have occurred as time proceeds. For such a process, we will in this note use two equivalent descriptions of the sample paths.

- Increasing integer-valued piecewise constant functions: (multivariate) counting process.
- Sequence of points: (multivariate) point process.

A third way to view this process is as a random counting measure, but will not be used in this note.

11.1. Counting process and poin process. Let $\mathcal{F}(t)$ be a given filtration in this subsection. A (univariate) counting process is an increasing CADLAG process with sample paths which are piecewise constant having jumps of size one only and to make this precise the definition of a counting process is as follows.

Definition 11.1. An FV-process $(N(t), t \geq 0)$ is a (univariate) counting process if

(i) The initial value is $N(0) = 0$.
(ii) $N(t)$ takes values in $\mathbb{N}_0 \cup \{\infty\} = \{0, 1, 2, \ldots\} \cup \{\infty\}$ for all $t > 0$.
(iii) The sample paths of $(N(t), t \geq 0)$ increase by at most one at every point, that is, $\Delta N(t) \in \{0, 1\}$ on $\{N(t) < \infty\}$ for all $t > 0$.

To a counting process $(N(t), t \geq 0)$ corresponds jump times τ_n given by

$$\tau_n = \inf\{t \geq 0 \,:\, N(t) = n\} \quad (\inf \emptyset = \infty) \text{ for } n = 1, 2, \ldots. \tag{11.1}$$

The counting process $(N(t), t \geq 0)$ is recovered from the sequence of jump times $(\tau_n, n = 1, 2, \ldots)$ since

$$N(t) = \sum_{n=1}^{\infty} 1_{[\tau_n,\infty)}(t) = \sum_{n=1}^{\infty} 1_{\{\tau_n<\infty\}} 1_{[\tau_n,\infty)}(t). \tag{11.2}$$

Thus, there is an one-to-one corresponding between the counting process $(N(t), t \geq 0)$ and the sequence of jump times $(\tau_n, n = 1, 2, \ldots)$ given by (11.1) and (11.2). The sequence of jump times $(\tau_n, n = 1, 2, \ldots)$ is a (univariate) point process according to the following definition.

Definition 11.2. A sequence $(\tau_n, n = 1, 2, \ldots)$ of $(0, \infty]$-valued random variables is a point process if

(i) $0 < \tau_1 \leq \tau_2 \leq \cdots$.
(ii) $\tau_n < \tau_{n+1}$ on the set $\{\tau_n < \infty\}$ for $n = 1, 2, \ldots$
(iii) $\tau_n = \tau_{n+1} = \infty$ on the set $\{\tau_n = \infty\}$ for $n = 1, 2, \ldots$

As all ready noted, the counting process $(N(t), t \geq 0)$ may be identified with its point process $(\tau_n, n = 1, 2, \ldots)$. The explosion time of the counting process $(N(t), t \geq 0)$ is the random variable given by the limit $\tau_\infty = \lim_{n\uparrow\infty} \tau_n = \inf\{t \geq 0 : N(t) = \infty\}$. If $\tau_\infty = \infty$ then the counting process $N(t)$ is without explosion or equivalent that only finitely many events can occur in any finite time interval, that is, $N(t) < \infty$ for all $t \geq 0$. Throughout Chapter III and IV we will restricted to counting processes without explosion.

Lemma 11.3. *The counting process $N(t)$ is adapted to the filtration $\mathcal{F}(t)$ if and only if the point process τ_n is a sequence of stopping times with respect to $\mathcal{F}(t)$. In both cases, the explosion time τ_∞ is a stopping time with respected to $\mathcal{F}(t)$.*

Proof. By the definition of a stopping time (see Definition 6.5), the first part of the lemma follows by $\{\tau_n \leq t\} = \{N(t) \geq n\}$ and the latter part follows by $\{\tau_\infty \leq t\} = \cap_{n=1}^\infty \{\tau_n \leq t\} = \{N(t) = \infty\}$. $\square$

A multivariate counting process is a vector of (univariate) counting processes such that no two components jump at the same time and to make this precise the definition of a multivariate counting process is as follows.

Definition 11.4. A process $(N(t), t \geq 0) = ((N^1(t), \ldots, N^m(t)), t \geq 0)$ is a m-variate counting process if

(i) The component process $(N^k(t), t \geq 0)$ is a counting process for $k = 1, \ldots, m$.
(ii) $\Delta N^j(t)\, \Delta N^k(t) = 0$ on $\{N^j(t) < \infty, N^k(t) < \infty\}$ for all $j, k = 1, \ldots, m$, $j \neq k$ and all $t > 0$.

Due to condition (ii), the counting processes $N^1(t), \ldots, N^m(t)$ do not jump simultaneously. The total number of events is given by the sum of the components

$$N^\bullet(t) = \sum_{k=1}^m N^k(t)$$

and is a (univariate) counting process (see Exercise 13.4). To a m-variate counting process $N(t)$ corresponds jump times τ_n given by

$$\tau_n = \inf\{t \geq 0 : N^\bullet(t) = n\} \quad (\inf \emptyset = \infty) \text{ for } n = 1, 2, \ldots \tag{11.3}$$

and marks on events that occur given by

$$Z_n = \sum_{k=1}^m k 1_{\{\Delta N^k(\tau_n) = 1\}} \text{ on } \{\tau_n < \infty\} \text{ for } n = 1, 2, \ldots \tag{11.4}$$

For events that never occur we define the marks to be an arbitrary $\{1, \ldots, m\}$-valued random variables, that is, $Z_n = Z_\infty^{(n)}$ on $\{\tau_n = \infty\}$ where $Z_\infty^{(n)}$ is some $\{1, \ldots, m\}$-valued random variable. As for the univariate case, the counting process $(N^\bullet(t), t \geq 0)$ is recovered by the sequence of jump times $(\tau_n, n = 1, 2, \ldots)$ and the m-variate counting process $N(t) = (N^1(t), \ldots, N^m(t))$

is recovered by the sequence of jump times $(\tau_n, n = 1, 2, \ldots)$ and the sequence of marks $(Z_n, n = 1, 2, \ldots)$ on events that occur since

$$N^k(t) = \sum_{n=1}^{\infty} 1_{\{Z_n=k\}} 1_{[\tau_n,\infty)}(t) = \sum_{n=1}^{\infty} 1_{\{Z_n=k,\tau_n<\infty\}} 1_{[\tau_n,\infty)}(t) \text{ for } k = 1, \ldots, m. \tag{11.5}$$

The sequence of $((\tau_n, Z_n), n = 1, 2, \ldots)$ is a multivariate point process according to the following definition.

Definition 11.5. A double sequence $((\tau_n, Z_n), n = 1, 2, \ldots)$ of random variables is a m-variate point process if

(i) $(\tau_n, n = 1, 2, \ldots)$ is a (univariate) point process.
(ii) $(Z_n, n = 1, 2, \ldots)$ is a sequence of $\{1, \ldots, m\}$-valued random variables.

A multivariate point process is a marked point process with mark space the finite set $\{1, \ldots, m\}$. Again, the multivariate counting process $N(t)$ is adapted to the filtration if and only if the multivariate point process (τ_n, Z_n) is a sequence of stopping times τ_n and a sequence of $\mathcal{F}(\tau_n)$ measurable random variables Z_n.

11.2. Filtration. Let $((\tau_n, Z_n), n = 1, 2, \ldots)$ be an m-variate point process. We define the (multivariate) point process filtration by extending Lemma 6.6.

$$\mathcal{F}(t) = \sigma(\tau_n 1_{[\tau_n,\infty)}(t), Z_n 1_{[\tau_n,\infty)}(t) | n = 1, 2, \ldots). \tag{11.6}$$

Then $\mathcal{F}(t)$ is the smallest filtration such that τ_n is a stopping time and Z_n is $\mathcal{F}(\tau_n)$ measurable for all $n = 1, 2, \ldots$ Let $N(t) = (N^1(t), \ldots, N^m(t))$ be the associated m-variate counting process.

Proposition 11.6. *The point process filtration has the following properties:*

(i) $\mathcal{F}(t) = \mathcal{F}(t+)$ (*the filtration is right-continuous*).
(ii) $\mathcal{F}(t) = \mathcal{F}^N(t) = \sigma(N^1(s), \ldots, N^m(s) | s \leq t)$.
(iii) $\mathcal{F}(\tau_n) = \sigma((\tau_1, Z_1), \ldots, (\tau_n, Z_n))$.

Below we find a representation of the conditional expectation of a random variable given the point process filtration which will be a key lemma. To illustrate some of the ideas, we begin with a one-point process where $\tau > 0$ is a strictly positive random variable. In this case $\mathcal{F}(t)$ is generated by the random variable $\tau 1_{[\tau,\infty)}(t)$, that is,

$$\mathcal{F}(t) = \sigma(\tau 1_{[\tau,\infty)}(t)). \tag{11.7}$$

Lemma 11.7. *Let Y be an integrable random variable. Then*

$$\mathbf{E}[Y|\mathcal{F}(t)]1_{[0,\tau)}(t) = \frac{\mathbf{E}[Y 1_{[0,\tau)}(t)]}{\mathbf{P}(\tau > t)} 1_{[0,\tau)}(t) = \mathbf{E}[Y|\tau > t]1_{[0,\tau)}(t)$$

for $t \geq 0$.

Proof. By (11.7), the random variable $\mathbf{E}[Y|\mathcal{F}(t)] = \mathbf{E}[Y|\tau 1_{[\tau,\infty)}(t)]$ is measurable with respect to $\tau 1_{[\tau,\infty)}(t)$. Hence there is a (measurable) function φ such that $\mathbf{E}[Y|\mathcal{F}(t)] = \varphi(\tau 1_{[\tau,\infty)}(t)) = \varphi(\tau)1_{[\tau,\infty)}(t) + \varphi(0)1_{[0,\tau)}(t)$. Multiply this equation with $1_{[0,\tau)}(t)$, we get

$$\mathbf{E}\big[\mathbf{E}[Y|\mathcal{F}(t)]1_{[0,\tau)}(t)\big] = \mathbf{E}[\varphi(0)1_{[0,\tau)}(t)] = \varphi(0)\mathbf{P}(\tau > s).$$

Again due to (11.7) we have that $\{\tau > t\} \in \mathcal{F}(t)$ and hence

$$\varphi(0)\mathbf{P}(\tau > t) = \mathbf{E}\big[\mathbf{E}[Y|\mathcal{F}(t)]1_{[0,\tau)}(t)\big] = \mathbf{E}\big[\mathbf{E}[Y 1_{[0,\tau)}(t)|\mathcal{F}(t)]\big] = \mathbf{E}[Y 1_{[0,\tau)}(t)].$$

Thus $\varphi(0) = \mathbf{E}[Y 1_{[0,\tau)}(t)]/\mathbf{P}(\tau > t) = \mathbf{E}[Y|\tau > t]$. $\square$

The next result is for a general multivariate point process. We omit the proof, since it is more notationally cumbersome version of the proof of Lemma 11.7.

Lemma 11.8. *Let $((\tau_n, Z_n), n = 1, 2, \ldots)$ be an m-variate point process and let $\mathcal{F}(t)$ be the point process filtration given in (11.6) Let Y be an integrable random variable. Then*

$$\begin{aligned}\mathbf{E}[Y|\mathcal{F}(t)]1_{[\tau_n,\tau_{n+1})}(t) &= \frac{\mathbf{E}[Y1_{[0,\tau_{n+1})}(t)|(\tau_1, Z_1), \ldots, (\tau_n, Z_n)]}{\mathbf{P}(\tau_{n+1} > t|(\tau_1, Z_1), \ldots, (\tau_n, Z_n))}1_{[\tau_n,\tau_{n+1})}(t) \\ &= \mathbf{E}[Y|(\tau_1, Z_1), \ldots, (\tau_n, Z_n), 1_{[0,\tau_{n+1})}(t)]1_{[\tau_n,\tau_{n+1})}(t)\end{aligned}$$

for $n = 0, 1, 2, \ldots$ and $t \geq 0$ (with $\tau_0 = 0$).

The above lemma provides the following result.

Corollary 11.9. *Let Y be an integrable random variable. Then the martingale $M(t) = \mathbf{E}[Y|\mathcal{F}(t)]$ admits a version such that $M(t)$ is an FV-martingale.*

Proof. By Proposition 9.3 $M(t)$ is a martingale. By Lemma 11.8 we have that

$$\mathbf{E}[Y|\mathcal{F}(t)] = \sum_{n=0}^{\infty} \frac{\mathbf{E}[Y1_{[0,\tau_{n+1})}(t)|(\tau_1, Z_1), \ldots, (\tau_n, Z_n)]}{\mathbf{P}(\tau_{n+1} > t|(\tau_1, Z_1), \ldots, (\tau_n, Z_n))}1_{[\tau_n,\tau_{n+1})}(t) + \mathbf{E}[Y|\mathcal{F}(\infty)]1_{[\tau_\infty,\infty)}(t).$$

Note that $t \mapsto \mathbf{E}[Y1_{[0,\tau_{n+1})}(t)|(\tau_1, Z_1), \ldots, (\tau_n, Z_n)] = \mathbf{E}[Y^+1_{[0,\tau_{n+1})}(t)|(\tau_1, Z_1), \ldots, (\tau_n, Z_n)] - \mathbf{E}[Y^-1_{[0,\tau_{n+1})}(t)|(\tau_1, Z_1), \ldots, (\tau_n, Z_n)]$ is an FV-process as well as the remaining processes on the left-hand side. Then by Change of variable formula (Theorem 4.3) the process $M(t)$ must be an FV-process. □

From these remarkable result which implies that in the point process filtration given in (11.7) then any martingale on the form $M(t) = \mathbf{E}[Y|\mathcal{F}(t)]$ can be assumed to be an FV-martingale.

11.3. Predictable compensators and intensity processes. Since a counting process is CADLAG and increasing, it follows by the Doob-Meyer theorem 9.8 that there exists a predictable compensator (which is CADLAG and increasing). Below the compensator is expressed as a function of the point process

To illustrate some of the ideas for counting processes, we begin with one-jump process. Let $\tau > 0$ be a one-point process with distribution function $F(t) = \mathbf{P}(\tau \leq t)$ and survival functions $\bar{F}(t) = \mathbf{P}(\tau > t)$ for $t > 0$. Let $N(t) = 1_{[\tau,\infty)}(t)$ be the associated one-jump process.

Theorem 11.10. *Let $\tau > 0$ be a strictly positive random variable with distribution function F. Then the one-jump process $N(t) = 1_{[\tau,\infty)}(t)$ has predictable compensator*

$$\Lambda^N(t) = \int_0^{\tau\wedge t} \frac{1}{\bar{F}(u-)}\,dF(u) = \int_0^t 1_{(0,\tau]}(u)\frac{1}{\bar{F}(u-)}\,dF(u).$$

In particular, if F is absolute continuous with density f, that is, $dF(t) = f(t)\,dt$. Then the one-jump process $N(t)$ has intensity process

$$\lambda^N(t) = 1_{(0,\tau]}(t)\frac{f(t)}{\bar{F}(t)}.$$

Proof. The process $1_{(0,\tau]}(t)$ is predictable since it is adapted and left-continuous. Moreover, $dF(t)/\bar{F}(t-)$ is deterministic and therefore also is predictable, the process $\Lambda(t)$ is predictable. According to the definition we have to show that $N(t) - \Lambda(t)$ is a martingale. In other words, for $0 \leq s < t$, we want to show that $\mathbf{E}[N(t) - \Lambda(t)|\mathcal{F}(s)] = N(s) - \Lambda(s)$ or equivalent that $\mathbf{E}[N(t) - N(s)|\mathcal{F}(s)] = \mathbf{E}[\Lambda(t) - \Lambda(s)|\mathcal{F}(s)]$. By an application of Lemma 11.7, we have that

$$\begin{aligned}\mathbf{E}[N(t) - N(s)|\mathcal{F}(s)] = \mathbf{E}[1_{\{s<T\leq t\}}|\mathcal{F}(s)]1_{[0,\tau)}(s) &= \frac{\mathbf{E}[1_{\{s<T\leq t\}}]}{\bar{F}(s)}1_{[0,\tau)}(s) \\ &= \frac{F(t) - F(s)}{\bar{F}(s)}1_{[0,T)}(s).\end{aligned}$$

By an application of Fubini for conditional expectation (Theorem 10.10) and an application of Lemma 11.7, we have that

$$\begin{aligned}&\mathbf{E}[\Lambda(t) - \Lambda(s)|\mathcal{F}(s)] \\ &\quad= \mathbf{E}\Big[\int_s^t 1_{(0,\tau]}(u)\frac{1}{\bar{F}(u-)}\,dF(u)\Big|\mathcal{F}(s)\Big] = \int_s^t \mathbf{E}[1_{(0,\tau]}(u)|\mathcal{F}(s)]1_{[0,\tau)}(s)\frac{1}{\bar{F}(u-)}\,dF(u) \\ &\quad= \int_s^t \frac{\mathbf{E}[1_{(0,\tau]}(u)]}{\bar{F}(s)}1_{[0,\tau)}(s)\frac{1}{\bar{F}(u-)}\,dF(u) = \int_s^t \frac{\bar{F}(u-)}{\bar{F}(s)}1_{[0,\tau)}(s)\frac{1}{\bar{F}(u-)}\,dF(u) \\ &\quad= \frac{F(t) - F(s)}{\bar{F}(s)}1_{[0,\tau)}(s)\end{aligned}$$

which shows that $\Lambda(t)$ is the predictable compensator. Finally, when F is absolute continuous we have that

$$d\Lambda(t) = 1_{(0,\tau]}(t)\frac{1}{\bar{F}(t-)}\,dF(t) = 1_{(0,\tau]}(t)\frac{f(t)}{\bar{F}(t)}\,dt$$

which proves the result. □

The next result is the predictable compensator for a multivariate counting process. We omit the proof, since it is more notationally cumbersome version of the proof of Theorem 11.10. We will use the following notation $\mathbf{P}(Y \in dy)$ for the distribution of a random variable Y, that is,

$$\mathbf{P}(Y \in A) = \int_A \mathbf{P}(Y \in dy).$$

Theorem 11.11. *Let $((\tau_n, Z_n), n = 1, 2, \ldots)$ be a m-variate point process. Then the counting process*

$$N^k(t) = \sum_{n=1}^{\infty} 1_{\{Z_n=k\}}1_{[\tau_n,\infty)}(t)$$

has predictable compensator

$$\begin{aligned}\Lambda^k(t) = &\int_0^t 1_{(0,\tau_1]}(u)\frac{1}{\mathbf{P}(\tau_1 \geq u)}\,\mathbf{P}(\tau_1 \in du, Z_1 = k) \\ &+ \sum_{n=1}^{\infty}\int_0^t 1_{(\tau_n,\tau_{n+1}]}(u)\frac{\mathbf{P}(\tau_{n+1} \in du, Z_{n+1} = k|(\tau_1, Z_1), \ldots, (\tau_n, Z_n))}{\mathbf{P}(\tau_{n+1} \geq u|(\tau_1, Z_1), \ldots, (\tau_n, Z_n))}\end{aligned}$$

for $k = 1, \ldots, m$.

We shall always assume that $\mathbf{E}[N^k(t)] < \infty$ or $\mathbf{E}[\Lambda^k(t)] < \infty$ then $M^k(t) = N^k(t) - \Lambda^k(t)$ are martingales for $k = 1, \ldots, m$ and we can use the results in Section 9.3 to this special case. In infinitesimal form the predictable compensator becomes the heuristic expression

$$d\Lambda^k(t) = \mathbf{E}[dN^k(t)|\mathcal{F}(t-)] = \mathbf{P}(dN^k(t) = 1|\mathcal{F}(t-)).$$

Counting processes with predictable intensity process $\lambda^k(t) \geq 0$ are of particularly interest and by the short informal notation we have that the intensity process is given by

$$\lambda^k(t)\,dt = \mathbf{E}[dN^k(t)|\mathcal{F}(t-)] = \mathbf{P}(dN^k(t) = 1|\mathcal{F}(t-)).$$

From a modeling point of view, the traditional probabilistic approach to state the law of a counting process by the finite dimensional distributions is not appropriate (that is to state the distribution of $(N^k(t_1), \ldots, N^k(t_n))$ for $0 \leq t_1 < \cdots < t_n$). This approach can easily be incalculable. Another approach is to use intensity processes for modeling counting processes which also may be easier to interpret. This approach is applicable due to the following result: Given an intensity process $\lambda^k(t)$ then—under some regularity conditions—there is a unique counting process with $\lambda^k(t)$ as its intensity process.

11.4. Predictable variation and covariation processes. Let $N(t) = (N^1(t), \ldots, N^m(t))$ be a m-variate counting process with intensity process $\lambda(t) = (\lambda^1(t), \ldots, \lambda^m(t))$. Assume that the martingales

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

for $k = 1, \ldots, m$ are square integrable. By Section 9.4, $M^k(t)$ has a predictable variation process $\langle M^k \rangle(t)$ and $\langle M^j, M^k \rangle(t)$ is the predictable covariation process of $M^j(t)$ and $M^k(t)$. The next result shows the relation between the martingale and the predictable compensator of the counting process.

Proposition 11.12. *The predictable variation process and the predictable covariation process are given by.*

(i) $\langle M^k \rangle(t) = \int_0^t \lambda^k(s)\,ds$ *for* $k = 1, \ldots, m$.
(ii) $\langle M^j, M^k \rangle(t) = 0$ *for* $j, k = 1, \ldots, m$ *and* $j \neq k$.

Proof. Recall that $\Lambda^k(t) = \int_0^t \lambda^k(s)\,ds$ is the predictable compensator. Then $(M^k(t))^2 = \big(N^k(t) - \Lambda^k(t)\big)^2 = (N^k(t))^2 + (\Lambda^k(t))^2 - 2N^k(t)\Lambda^k(t)$. By integration by parts (Theorem 4.1), we have

$$\begin{aligned}
(M^k(t))^2 &= 2\int_0^t N^k(s-)\,dN^k(s) + N^k(t) + 2\int_0^t \Lambda^k(s)d\Lambda^k(s) \\
&\quad - 2\int_0^t N^k(s-)\,d\Lambda^k(s) - 2\int_0^t \Lambda^k(s)\,dN^k(s) \\
&= \int_0^t \Big(2M^k(s-) + 1\Big)\,dM^k(s) + \Lambda^k(t).
\end{aligned}$$

The integral process is a martingale and $\Lambda^k(t)$ is predictable. Thus the above equation is the Doob-Meyer decomposition of $(M^k(t))^2$. The result now follows from by the uniqueness of the Doob-Meyer decomposition. For the second part of the proposition see Exercise 13.5. □

11.5. Martingale central limit theorem for counting processes. The conditions in the martingale central limit theorem is formulated to the case of interest for counting processes models for life history. For simplicity we consider the one dimensional case. Let $N_{(n)}(t)$ be a counting process having intensity process $\lambda_{(n)}(t)$.

Theorem 11.13. (*Martingale central limit theorem for counting processes*). *Consider a sequence of counting processes $N_{(n)}(t)$ with intensity processes $\lambda_{(n)}(t)$, and a sequence $H_{(n)}(t)$ of predictable processes such that $M_{(n)}(t) = \int_0^t H_{(n)}(s)\big(dN_{(n)}(s) - \lambda_{(n)}(s)\,ds\big)$ are well defined martingales. Assume that*

(i) *For every $t \geq 0$, $\langle M_{(n)}\rangle(t) = \displaystyle\int_0^t \big(H_{(n)}(s)\big)^2 \lambda_{(n)}(s)\,ds \xrightarrow[n\uparrow\infty]{\mathbf{P}} V(t)$ where $V(t) \geq 0$ is a continuous deterministic function.*

(ii) *For every $t \geq 0$ and for all $\varepsilon > 0$,*
$$\int_0^t \big(H_{(n)}(s)\big)^2 1_{\{|H_{(n)}(s)|>\varepsilon\}} \lambda_{(n)}(s)\,ds \xrightarrow[n\uparrow\infty]{\mathbf{P}} 0.$$

Then the processes $M_{(n)}(t)$ converge in distribution to $M_{(\infty)}(t)$, a continuous Gaussian martingale with mean zero and variance $V(t)$ (see Definition 11.16*).*

Remark 11.14. Intuitively, a sequence of martingales converges to a Gaussian martingale if:

(i) The predictable variation process converges to the variance function of the limiting Gaussian martingale (the variation process converges to a deterministic function).

(ii) This condition is a kind of Lindeberg condition requiring that big jumps are asymptotically rare (sample paths becomes continuous in the limit).

Theorem 11.15. (*Martingale central limit theorem for multivariate counting processes*). *Let $N_{(n)}(t) = (N^1_{(n)}(t), \ldots, N^m_{(n)}(t))$ be a sequence of m-variate counting processes with intensity processes $\lambda_{(n)}(t) = (\lambda^1_{(n)}(t), \ldots, \lambda^m_{(n)}(t))$. Let $H^{jk}_{(n)}(t)$, $j,k = 1,\ldots,m$ be sequences of predictable processes such that*
$$M^j_{(n)}(t) = \sum_{k=1}^m \int_0^t H^{jk}_{(n)}(s)\big(dN^k_{(n)}(s) - \lambda^k_{(n)}(s)\,ds\big)$$
are well defined martingales. Assume that

(i) *For every $t \geq 0$ and $j,k = 1,\ldots,m$,*
$$\langle M^j_{(n)}, M^k_{(n)}\rangle(t) = \sum_{l=1}^m \int_0^t H^{jl}_{(n)}(s) H^{kl}_{(n)}(s) \lambda^l_{(n)}(s)\,ds \xrightarrow[n\uparrow\infty]{\mathbf{P}} V^{jk}(t)$$
where $V(t) = (V^{jk}(t),\ j,k = 1,\ldots,m)$ is a continuous deterministic semidefinite matrix.

(ii) *For every $t \geq 0$, for every $j = 1,\ldots,m$, and for all $\varepsilon > 0$*
$$\sum_{k=1}^m \int_0^t \big(H^{jk}_{(n)}(s)\big)^2 1_{\{|H^{jk}_{(n)}(s)|>\varepsilon\}} \lambda^k_{(n)}(s)\,ds \xrightarrow[n\uparrow\infty]{\mathbf{P}} 0.$$

Then the processes $M_{(n)}(t) = (M^1_{(n)}(t), \ldots, M^m_{(n)}(t))$ converges in distribution to a continuous Gaussian martingale (vector) $M_{(\infty)}(t) = (M^1_{(\infty)}(t), \ldots, M^m_{(\infty)}(t))$ with mean zero and covariance matrix $V(t)$ (see Definition 11.16*).*

Definition 11.16. A process $M(t) = (M^1(t), \ldots, M^m(t))$ is a continuous (m-dimensional) Gaussian martingale with mean zero and covariance matrix $V(t)$ if

(i) $M^1(t), \ldots, M^m(t)$ are continuous martingales with $M^k(0) = 0$ for $k = 1,\ldots,m$.

(ii) $M^j(t)M^k(t) - V^{jk}(t)$ are martingales for $j,k = 1,\ldots,m$ and $V(t) = (V^{jk}(t),\ j,k = 1,\ldots,m)$ is a continuous deterministic semidefinite matrix.

(iii) $M(t)$ has independent Gaussian increments.

12. Piecewise constant processes on finite state spaces

Piecewise constant processes are used for modeling the state of a multi-state life insurance contract. A piecewise constant process $Z(t)$ (or a step process) is CADLAG and it has finitely many jumps in each finite time interval, and is constant between jumps. For such a process, we will in this note use three equivalent description of the sample paths.

- Piecewise constant (step) functions: piecewise constant process
- Sequence of points: multivariate point process.
- Increasing integer-valued piecewise constant functions: multivariate counting process.

12.1. Piecewise constant process. A pure jump process with at most one explosion time has sample paths which are pure jump functions which are constant after the explosion time. The explosion time is finite if the process has infinitely many jumps with a finite accumulation point of jumps.

Definition 12.1. Let $\mathcal{F}(t)$ be a given filtration. An FV-process $(Z(t), t \geq 0)$ is a pure jump process with at most one explosion with finite state space $\mathcal{S} = \{1, 2, \ldots, m\}$ if

(i) $Z(t)$ takes values in $\mathcal{S} = \{1, 2, \ldots, m\}$ for all $t \geq 0$.
(ii) The sample paths of $(Z(t), t \geq 0)$ are pure jump functions with at most one explosion time τ_∞ such that $Z(t) = Z(\tau_\infty)$ for $t \geq \tau_\infty$.
(iii) $Z(t)$ is a piecewise constant process if there is no explosion, that is, $\tau_\infty = \infty$.

To the pure jump process $(Z(t), t \geq 0)$ corresponds jump times and jump values that can be inductively defined as follows. Let $\tau_0 = 0$ and $Z_0 = Z(0)$ and then inductively define $\tau_n = \inf\{t > \tau_{n-1} \,|Z(t) \neq Z(\tau_{n-1})\}$ ($\inf \emptyset = \infty$) and jump values on events that occur $Z_n = Z(\tau_n)$ on $\{\tau_n < \infty\}$ for $n = 1, 2, \ldots$ For events that never occur we define the jump values to be an arbitrary $\{1, 2, \ldots, m\}$-valued random variables, that is, $Z_n = Z_\infty^{(n)}$ on $\{\tau_n = \infty\}$ where $Z_\infty^{\{(n)\}}$ is some $\{1, \ldots, m\}$-valued random variable. The pure jump process $(Z(t), t \geq 0)$ is recovered from the sequence of jump times $(\tau_n, n = 0, 1, 2, \ldots)$ and the sequence of jump values $(Z_n, n = 0, 1, 2, \ldots)$ on events that occur since

$$Z(t) = \sum_{n=0}^{\infty} Z_n 1_{[\tau_n, \tau_{n+1})}(t) = Z_n \text{ if } t \in [\tau_n, \tau_{n+1}).$$

The sequence $((\tau_n, Z_n), n = 0, 1, 2, \ldots)$ is a multivariate point process with $Z_{n-1} \neq Z_n$ on $\tau_n < \infty$. The associated multivariate counting process to the multivariate point process (τ_n, Z_n) is given by

$$\begin{aligned} N^k(t) &= \sum_{n=1}^{\infty} 1_{\{Z_n = k\}} 1_{[\tau_n, \infty)}(t) = \sum_{n=1}^{\infty} 1_{\{Z_{n-1} \neq k, Z_n = k\}} 1_{[\tau_n, \infty)}(t) \\ &= \#\{0 < s \leq t | Z(s-) \neq k, Z(s) = k\} \end{aligned}$$

for $k = 1, 2, \ldots, m$. The counting process $N^k(t)$ counts the number of transitions into state k by $Z(t)$. For $j, k = 1, 2, \ldots, m$ and $j \neq k$, define $N^{jk}(t)$ by,

$$N^{jk}(t) = \sum_{n=1}^{\infty} 1_{\{Z_{n-1} = j, Z_n = k\}} 1_{[\tau_n, \infty)}(t) = \#\{0 < s \leq t | Z(s-) = j, Z(s) = k\})$$

that count the number of transitions from state j to state k by $Z(t)$ in the time interval $[0,t]$. Note that $N^k(t) = \sum_{j:j\neq k} N^{jk}(t)$ and

$$N^{jk}(t) = \int_0^t 1_{\{Z(s-)=j\}}\, dN^k(s) \text{ for } j \neq k.$$

Remark 12.2. Let $\mathcal{F}(t)$ be the point process filtration given in (11.6), let $\mathcal{F}^Z(t)$ be the natural filtration generated by the pure jump process $Z(t)$, and let $\mathcal{F}^N(t)$ be the natural filtration generated by the multivariate counting process $N(t) = (N^1(t), \ldots, N^m(t))$. Then $\mathcal{F}(t) = \mathcal{F}^Z(t) = \mathcal{F}^N(t)$.

Often a (general) piecewise constant process is a too complex model to be mathematically tractable. In order to be able to make computations of quantities (reserves) of interest we need smaller classes of processes. Below two type of piecewise constant processes are presented: Markov processes used as a model for standard multi-state contracts (see Section 16). Semi-Markov processes used in models where payments and transition intensities depend on state duration (see Section 17).

For the remainder of this section let $Z(t)$ be a piecewise constants process with state space $\mathcal{S} = \{1, 2, \ldots, m\}$ and let $((\tau_n, Z_n), n = 0, 1, 2, \ldots)$ be the associated multivariate point process.

12.2. Continuous time Markov chains.

Definition 12.3. A pure jump process $Z(t)$ is a continuous time Markov chain (or Markov jump process) if

$$\mathbf{P}(Z(t) = k | \mathcal{F}(s)) = \mathbf{P}(Z(t) = k | Z(s)) \tag{12.1}$$

for all $0 \leq s < t$ and all $k \in \mathcal{S}$.

The property (12.1) is the Markov property of $Z(t)$. Indeed, for any subset $A \subseteq \mathcal{S}$ and by (12.1) it follows that

$$\begin{aligned}\mathbf{P}\big(Z(t) \in A \big| \mathcal{F}(t)\big) &= \sum_{k\in A} \mathbf{P}\big(Z(t) = k \big| \mathcal{F}(t)\big) = \sum_{k\in A} \mathbf{P}\big(Z(t) = k \big| Z(t)\big) \\ &= \mathbf{P}\big(Z(t) \in A \big| Z(t)\big)\end{aligned}$$

which is the Markov condition Definition 7.1. The transition probabilities of the continuous time Markov chain are for $0 \leq s < t$ and $j, k \in \mathcal{S}$ denoted by

$$p^{jk}(s,t) = \begin{cases} \mathbf{P}(Z(t) = k | Z(s) = j) & \text{if } \mathbf{P}(Z(s) = j) > 0 \\ 0 & \text{else.} \end{cases}$$

Note that $p^{Z(s)k}(s,t) = \mathbf{P}(Z(t) = k | Z(s))$. Moreover, the finite dimensional distributions of the Markov chain are determined by the transition probabilities and the initial distribution, that is, the distribution of $Z(0)$. Indeed, by $\mathbf{P}(A|B)\mathbf{P}(B) = \mathbf{P}(A \cap B)$ and induction one gets that the marginal distributions are for $0 = t_0 < t_1 < \cdots < t_n$ and $j_0, j_1, \ldots, j_n \in \mathcal{S}$ given by

$$\begin{aligned}
&\mathbf{P}(Z(0) = j_0, Z(t_1) = j_1, \ldots, Z(t_n) = j_n) \\
&\quad = \mathbf{P}(Z(0) = j_0, Z(t_1) = j_1, \ldots, Z(t_{n-1}) = j_{n-1}) \\
&\qquad \times \mathbf{P}(Z(t_n) = j_n | Z(0) = j_0, Z(t_1) = j_1, \ldots, Z(t_{n-1}) = j_{n-1}) \\
&\quad = \mathbf{P}(Z(0) = j_0, Z(t_1) = j_1, \ldots, Z(t_{n-1}) = j_{n-1})\, p^{j_{n-1},j_n}(t_{n-1}, t_n) \\
&\quad = \mathbf{P}(Z(t_{j_1}) = j_1) | Z(0) = j_0) \prod_{i=2}^{n} p^{j_{i-1},j_i}(t_{i-1}, t_i) \\
&\quad = \prod_{i=1}^{n} p^{j_{i-1},j_i}(t_{i-1}, t_i).
\end{aligned}$$

For a modeling aspect it is more tractable to specify transition intensities rather than the transition probabilities. Let $\mu^{jk}(t)$ be the transition intensity (rate) of a jump from state j to state k at time t.

Definition 12.4. A family of Borel function $\mu^{jk} : [0, \infty) \to [0, \infty)$ for $j, k \in \mathcal{S}$ and $j \neq k$ are time-dependent transition intensities if μ^{jk} is locally integrable from the right, that is, for every $t \geq 0$ there is $\varepsilon > 0$ such that $\int_t^{t+\varepsilon} \mu^{jk}(u)\, du < \infty$. Moreover, the total intensity for a jump from state j at time t is given by $\mu^j(t) = \sum_{k:k\neq j} \mu^{jk}(t)$.

Below we will construct a continuous time Markov chain with given transition intensities $\mu^{jk}(t)$ out from a homogenous Markovian multivariate point process. A multivariate point process $((\tau_n, Z_n), n = 1, 2, \ldots)$ is a discrete time Markov process if

$$\mathbf{P}(\tau_{n+1} \in A, Z_{n+1} = k | (\tau_1, Z_1), \ldots, (\tau_n, Z_n)) = \mathbf{P}(\tau_{n+1} \in A, Z_{n+1} = k | (\tau_n, Z_n))$$

for all 1, all $A \in \mathcal{B}([0, \infty])$, and all $k \in \mathcal{S}$. If $\mathbf{P}(\tau_{n+1} \in A, Z_{n+1} = k | (\tau_n, Z_n))$ is independent of n then $((\tau_n, Z_n), n = 1, 2, \ldots)$ is a homogenous Markov process. The distribution of a homogenous Markovian multivariate point process is determined by the initial distribution and for $n = 1, 2, \ldots$ by the conditional distribution of τ_{n+1} given (τ_n, Z_n) and the conditional distribution of Z_{n+1} given (Z_n, τ_{n+1}).

Let $((\tau_n, Z_n), n = 0, 1, 2, \ldots)$ be a homogenous Markovian multivariate point process such that

1. Initial distribution: $\tau_0 = 0$ and $Z_0 = j_0$ for $j_0 \in \mathcal{S}$.
2. Next jump time: For $0 = t_0 < t_1 < \cdots < t_n < t_{n+1}$, and $j_0, j_1, \ldots, j_n \in \mathcal{S}$ then

$$\begin{aligned}
&\mathbf{P}(\tau_{n+1} > t_{n+1} | (\tau_0 = t_0, Z_0 = j_0), (\tau_1 = t_1, Z_1 = j_1), \ldots, (\tau_n = t_n, Z_n = j_n)) \\
&\quad = \mathbf{P}(\tau_{n+1} > t_{n+1} | \tau_n = t_n, Z_n = j_n) \\
&\quad = \exp\Big(- \int_{t_n}^{t_{n+1}} \mu^{j_n}(u)\, du \Big) = \bar{F}^{j_n}(t_n, t_{n+1}).
\end{aligned}$$

3. Next jump value: For $j_{n+1} \in \mathcal{S} \setminus \{j_n\}$ then

$$\begin{aligned}
&\mathbf{P}(Z_{n+1} = j_{n+1} | (\tau_0 = t_0, Z_0 = j_0), (\tau_1 = t_1, Z_1 = j_1), \ldots, (\tau_n = t_n, Z_n = j_n), \tau_{n+1} = t_{n+1}) \\
&\quad = \mathbf{P}(Z_{n+1} = j_{n+1} | \tau_{n+1} = t_{n+1}, Z_n = j_n) \\
&\quad = \mu^{j_n j_{n+1}}(t_{n+1}) / \mu^{j_n}(t_{n+1}) = \pi^{j_n, j_{n+1}}(t_{n+1})
\end{aligned}$$

The associated piecewise constant process is given by $Z(t) = Z_n$ for $\tau_n \le t < \tau_{n+1}$. Then $\mu^j(t)$ is the transition intensity for the waiting distribution in state j and $\pi^{jk}(t)$ is the jump probability from state j to state k at time t. So for $j \neq k$, $\mu^{jk}(t) = \mu^j(t)\pi^{jk}(t)$ is the transition intensity from state j to state k.

The transition probabilities of (τ_n, Z_n) are for $0 \le s < t$ and $j \neq k \in \mathcal{S}$ given by

$$\begin{aligned}
\bar{F}^{jk}(s,t) &= \mathbf{P}(\tau_{n+1} > t, Z_{n+1} = k | \tau_n = s, Z_n = j) \\
&= \int_t^\infty \mathbf{P}(Z_{n+1} = k | \tau_n = s, Z_n = j, \tau_{n+1} = u)\, \mathbf{P}(\tau_{n+1} \in du | \tau_n = s, Z_n = j) \\
&= \int_t^\infty \mathbf{P}(Z_{n+1} = k | Z_n = j, \tau_{n+1} = u)\, \mathbf{P}(\tau_{n+1} \in du | \tau_n = s, Z_n = j) \\
&= \int_t^\infty \pi^{jk}(u)\, d_u F^j(s,u).
\end{aligned}$$

For $0 \le s < t$ and $j \neq k \in \mathcal{S}$, the dynamics are given by $d_t F^j(s,t) = \mu^j(t) \exp\left(-\int_s^t \mu^j(u)\,du\right) = \mu^j(t)\bar{F}^j(s,t)\,dt$ and $d_t F^{jk}(s,t) = \pi^{jk}(t)\,d_t F^j(s,t) = \pi^{jk}(t)\mu^j(t)\bar{F}^j(s,t)\,dt = \mu^{jk}(t)\bar{F}^j(s,t)\,dt$. Using the general result in Theorem 11.11, we can derive the intensity processes of the associated counting processes of the given multivariate point process.

Proposition 12.5. *The counting process $N^{jk}(t) = \sum_{n=1}^\infty 1_{\{Z_{n-1}=j, Z_n=k\}} 1_{[\tau_n,\infty)}(t)$ has intensity process given by*

$$\lambda^{jk}(t) = \mu^{jk}(t) 1_{\{Z(t-)=j\}} \text{ for } j \neq k \in \mathcal{S}.$$

The counting process $N^k(t) = \sum_{n=1}^\infty 1_{\{,Z_{n-1}\neq k, Z_n=k\}} 1_{[\tau_n,\infty)}(t)$ has intensity process given by

$$\lambda^k(t) = \sum_{j:j\neq k} \mu^{jk}(t) 1_{\{Z(t-)=j\}} = \mu^{Z(t-)k}(t) 1_{\{Z(t-)\neq k\}} \text{ for } k \in \mathcal{S}.$$

Proof. By Theorem 11.11, the counting process $N^k(t)$ has predictable compensator

$$\begin{aligned}
\Lambda^k(t) &= \sum_{n=0}^\infty \int_0^t 1_{(\tau_n,\tau_{n+1}]}(u) \frac{\mathbf{P}(\tau_{n+1} \in du, Z_{n+1} = k | (\tau_0, Z_0), \ldots, (\tau_n, Z_n))}{\mathbf{P}(\tau_{n+1} \ge u | (\tau_0, Z_0), \ldots, (\tau_n, Z_n))} \\
&= \sum_{n=0}^\infty \int_0^t 1_{(\tau_n,\tau_{n+1}]}(u) 1_{\{Z_n \neq k\}} \frac{d_u F^{Z_n k}(\tau_n, u)}{\bar{F}^{Z_n}(\tau_n, u)} \\
&= \sum_{n=0}^\infty \int_0^t 1_{(\tau_n,\tau_{n+1}]}(u) 1_{\{Z_n \neq k\}} \frac{\mu^{Z_n k}(u)\bar{F}^{Z_n}(\tau_n, u)\,du}{\bar{F}^{Z_n}(\tau_n, u)} \\
&= \sum_{n=0}^\infty \int_0^t 1_{(\tau_n,\tau_{n+1}]}(u) 1_{\{Z_n \neq k\}} \mu^{Z_n k}(u)\,du \\
&= \int_0^t 1_{\{Z(u-)\neq k\}} \mu^{Z(u-)k}(u)\,du
\end{aligned}$$

and the intensity process is $\lambda^k(t) = \mu^{Z(t-)k}(t) 1_{\{Z(t-)\neq k\}}$. By Proposition 10.6, the counting process $N^{jk}(t) = \int_0^t 1_{\{Z(s-)=j\}}\,dN^k(s)$ has intensity process $\lambda^{jk}(t) = 1_{\{Z(t-)=j\}}\lambda^k(t) = \mu^{jk}(t) 1_{\{Z(t-)=j\}}$. □

Remark 12.6. We have that $\mathbf{E}[N^{jk}(t)] = \mathbf{E}[\Lambda^{jk}(t)] = \int_0^t \mu^{jk}(s)\mathbf{P}(Z(s-) = j)\,ds \leq \int_0^t \mu^{jk}(s)\,ds$ where in the first equality have used the martingale property. If we assume that $\int_0^t \mu^{jk}(s)\,ds < \infty$ then $N^{jk}(t) < \infty$ and hence there is finite number of transitions (events) from state j to state k in the time interval $[0, t]$. Moreover, $N^{jk}(t) - \int_0^t \lambda^{jk}(s)\,ds$ is a (true) martingale.

Based on this result, we can now verify that the associated piecewise constant process $Z(t) = Z_n$ for $\tau_n \leq t < \tau_{n+1}$ is a continuous time Markov chain with $\mu^{jk}(t)$ as transition intensities.

Theorem 12.7. *If $\mu^{jk}(t) \geq 0$ are the transition intensities as in Definition 12.4 then $Z(t)$ is Markov. Moreover the transition probabilities satisfy Kolmogorov backward equations*

$$p^{jk}(s,t) = \delta_{jk} \exp\Big(-\int_s^t \mu^j(v)\,dv\Big) + \int_s^t \exp\Big(-\int_s^u \mu^j(v)\,dv\Big) \sum_{i:i\neq j} \mu^{ji}(u) p^{ik}(u,t)\,du.$$

for $s < t$ where δ_{jk} is the usual Kronecker delta ($\delta_{jk} = 0$ if $j \neq k$ and $\delta_{jk} = 1$ if $j = k$).

Proof. Based on Lemma 11.8, one can show that $Z(t)$ is Markov, that is, $\mathbf{P}(Z(t) = k|\mathcal{F}(s)) = \mathbf{P}(Z(t) = k|Z(s))$ for $0 \leq s < t$ and $k \in \mathcal{S}$ (although the calculations are non-trivial).

To verify Kolmogorov equations, fix $t > 0$. By Corollary 11.9, $M(s) = \mathbf{E}[1_{\{Z(t)=k\}}|\mathcal{F}(s)]$ for $0 \leq s \leq t$ is an FV-martingale. Using the Markov property, we get that $M(s) = \mathbf{E}[1_{\{Z(t)=k\}}|\mathcal{F}(s)] = \mathbf{E}[1_{\{Z(t)=k\}}|Z(s)] = p^{Z(s)k}(s,t) = \sum_j p^{jk}(s,t)1_{\{Z(s)=j\}}$. Applying Integration by parts formula (Theorem 4.1) we get that

$$\begin{aligned}
dM(s) &= \sum_j 1_{\{Z(s-)=j\}}\,d_s p^{jk}(s,t) + \sum_j p^{jk}(s,t)\,d1_{\{Z(s)=j\}} \\
&= \sum_j 1_{\{Z(s-)=j\}}\,d_s p^{jk}(s,t) + \sum_j p^{jk}(s,t)\,d\big(N^j(s) - \sum_{l:l\neq j} N^{jl}(s)\big).
\end{aligned}$$

The first term on the right-hand side is predictable then by Proposition 10.6 and Proposition 12.5 we have that

$$\begin{aligned}
&\sum_j 1_{\{Z(s-)=j\}}\,d_s p^{jk}(s,t) \\
&\quad= \Big(\sum_j p^{jk}(s,t) \sum_{l:l\neq j} \lambda^{jl}(s) - \sum_i p^{ik}(s,t)\lambda^i(s)\Big)\,ds \\
&\quad= \Big(\sum_j p^{jk}(s,t) \sum_{l:l\neq j} \mu^{jl}(s)1_{\{Z(s-)=j\}} - \sum_i p^{ik}(s,t) \sum_{j:j\neq i} \mu^{ji}(s)1_{\{Z(s-)=j\}}\Big)\,ds \\
&\quad= \sum_j 1_{\{Z(s-)=j\}}\Big(p^{jk}(s,t)\mu^j(s) - \sum_{i:i\neq j} p^{ik}(s,t)\mu^{ji}(s)\Big)\,ds
\end{aligned}$$

where we have used that the predictable compensator of the martingale $M(t)$ is zero. From this we get Kolmogorov equations. □

Remark 12.8. If all the transitions functions $t \mapsto \mu^{jk}(t)$ are continuous at s_0 then then $p^{jk}(s,t)$ is differentiable in s at s_0 with

$$p_s^{jk}(s_0,t) = \mu^j(s_0)p^{jk}(s_0,t) - \sum_{i:i\neq k} \mu^{ji}(s_0)p^{ik}(s_0,t).$$

Thus, if all the transitions functions are continuous then $p^{jk}(s,t)$ is differentiable in s with

$$p_s^{jk}(s,t) = \mu^j(s)p^{jk}(s,t) - \sum_{i:i\neq k} \mu^{ji}(s)p^{ik}(s,t)$$

where $p^{jk}(t,t) = \delta_{jk}$.

Let s be given fixed time and denote the next jump time by

$$\tau(s) = \inf\{t > s : Z(t) \neq Z(s)\}.$$

from the (strong) Markov property of $Z(t)$ one gets the following result.

Proposition 12.9. *For $0 \leq s < t$ and $j \in \mathcal{S}$ with $\mathbf{P}(Z(s) = j) > 0$ then*

$$\mathbf{P}(\tau(s) > t | Z(s) = j) = \exp\Big(- \int_s^t \mu^j(u)\, du\Big)$$

or

$$\mathbf{P}(\tau(s) \in dt | Z(s) = j) = \mu^j(t) \exp\Big(- \int_s^t \mu^j(u)\, du\Big)\, dt.$$

The interpretation is that the greater the transition intensity, the higher the probability that the process performs this jump in a small time interval.

12.3. Semi-Markov processes. Let again $(Z(t), t \geq 0)$ by a (general) pure jump process on the state space $\mathcal{S} = \{1, 2, \ldots, m\}$ and let $((\tau_n, Z_n), n = 0, 1, 2, \ldots)$ be the associated multivariate point process (with $\tau_0 = 0$ and $Z_0 = j_0$). We define a new process $U(t)$ to be the process that measures the time spent in the present state (the time elapsed since entering the current state). The duration process $U(t)$ is then given by

$$\begin{aligned} U(t) &= \sum_{n=0}^{\infty} (t - \tau_n) 1_{[\tau_n, \tau_{n+1})}(t) \\ &= t - \tau_n \quad \text{for } \tau_n \leq t < \tau_{n+1}. \end{aligned}$$

Let $\mathcal{F}(t)$ be the point process filtration given in (11.6).

Definition 12.10. The pure jump process $Z(t)$ is a semi-Markov process if the bivariate process $(Z(t), U(t))$ is a Markov process, that is,

$$\mathbf{P}(Z(t) = k, U(t) \in A | \mathcal{F}(s)) = \mathbf{P}(Z(t) = k, U(t) \in A | Z(s), U(s))$$

for $0 \leq s < t$, $k \in \mathcal{S}$, and $A \in \mathcal{B}([0, \infty))$.

The state space of $(Z(t), U(t))$ is $\mathcal{S} \times [0, \infty)$ and takes into account information concerning the duration of presence and $Z(t)$ itself might not be a Markov process.

As for the Markov process, it is tractable for modeling to specify transition intensities for the semi-Markov process. Let $\mu^{jk}(t, u)$ be the transition intensity (rate) of a jump from state j to state k at time t also depending on duration u. Set

$$\mu^j(t, u) = \sum_{k:k\neq j} \mu^{jk}(t, u)$$

which is the intensity of transition from state j at time t and has spent time u in the state. For simplicity, we assume that the transition intensities $(t, u) \mapsto \mu^{jk}(t, u)$ are continuous functions.

Below we will construct a semi-Markov process with given transition intensities $\mu^{jk}(t,u)$ out from a homogenous Markovian multivariate point process.

Let $((\tau_n, Z_n), n = 1, 2, \ldots)$ be a homogenous Markovian multivariate point process given by:

1. Initial distribution: $\tau_0 = 0$ and $Z_0 = j_0$ for $j_0 \in \mathcal{S}$.
2. Next jump time: For $0 = t_0 < t_1 < \cdots < t_n < t_{n+1}$ and $j_0, j_1, \ldots, j_n \in \mathcal{S}$ then

$$
\begin{aligned}
&\mathbf{P}(\tau_{n+1} > t_{n+1} | (\tau_0 = t_0, Z_0 = j_0), (\tau_1 = t_1, Z_1 = j_1), \ldots, (\tau_n = t_n, Z_n = j_n)) \\
&\quad = \mathbf{P}(\tau_{n+1} > t_{n+1} | \tau_n = t_n, Z_n = j_n) \\
&\quad = \exp\Big(- \int_{t_n}^{t_{n+1}} \mu^{j_n}(v, v - t_n)\, dv \Big) = \bar{F}^{j_n}(t_n, t_{n+1}).
\end{aligned}
$$

3. Next jump value: For $j_{n+1} \in \mathcal{S} \setminus \{j_n\}$ then

$$
\begin{aligned}
&\mathbf{P}(Z_{n+1} = j_{n+1} | (\tau_0 = t_0, Z_0 = j_0), (\tau_1 = t_1, Z_1 = j_1), \ldots, (\tau_n = t_n, Z_n = j_n), \tau_{n+1} = t_{n+1}) \\
&\quad = \mathbf{P}(Z_{n+1} = j_{n+1} | \tau_{n+1} = t_{n+1}, Z_n = j_n) \\
&\quad = \frac{\mu^{j_n j_{n+1}}(t_{n+1}, t_{n+1} - t_n)}{\mu^{j_n}(t_{n+1}, t_{n+1} - t_n)} = \pi^{j_n, j_{n+1}}(t_{n+1}, t_{n+1} - t_n).
\end{aligned}
$$

The transition probabilities of (τ_n, Z_n) are for $0 \le s < t$ and $j \ne k \in \mathcal{S}$ given by (see Section 12.1 for all details)

$$
\bar{F}^{jk}(s,t) = \mathbf{P}(\tau_{n+1} > t, Z_{n+1} = k | \tau_n = s, Z_n = j) = \int_t^\infty \pi^{jk}(u, u - s)\, d_u F^j(s, u).
$$

For $0 \le s < t$ and $j \ne k \in \mathcal{S}$, the dynamics are given by $d_t F^j(s,t) = \mu^j(t, t-s) \bar{F}^j(s,t)\, dt$ and $d_t F^{jk}(s,t) = pi^{jk}(t, t-s)\, d_t \bar{F}(s,t) = \mu^{jk}(t, t-s) \bar{F}^j(s,t)\, dt$.

Let $Z(t)$ be the associated piecewise constant process given by $Z(t) = Z_n$ for $\tau_n \le t < \tau_{n+1}$. and we have the following result. Using the general result in Theorem 11.11, we can derive the intensity processes of the associated counting processes of the given multivariate point process.

Proposition 12.11. *The counting process $N^{jk}(t) = \sum_{n=1}^\infty 1_{\{Z_{n-1}=j, Z_n=k\}} 1_{[\tau_n, \infty)}(t)$ for $j \ne k \in \mathcal{S}$, has intensity process $\lambda^{jk}(t)$ given by*

$$
\lambda^{jk}(t) = \mu^{jk}(t, U(t-)) 1_{\{Z(t-)=j\}}.
$$

For $k \in \mathcal{S}$, the counting process $N^k(t) = \sum_{n=1}^\infty 1_{\{, Z_{n-1} \ne k, Z_n = k\}} 1_{[\tau_n, \infty)}(t)$ has intensity process $\lambda^k(t)$ given by

$$
\lambda^k(t) = \sum_{j : j \ne k} \mu^{jk}(t, U(t-)) 1_{\{Z(t-)=j\}} = \mu^{Z(t-)k}(t, U(t-)) 1_{\{Z(t-) \ne k\}}.
$$

Proof. By Theorem 11.11, the counting process $N^k(t)$ has predictable compensator

$$\begin{aligned}\Lambda^k(t) &= \sum_{n=0}^{\infty} \int_0^t 1_{(\tau_n,\tau_{n+1}]}(s) \frac{\mathbf{P}(\tau_{n+1} \in ds, Z_{n+1} = k | (\tau_0, Z_0), \ldots, (\tau_n, Z_n))}{\mathbf{P}(\tau_{n+1} \geq s | (\tau_0, Z_0), \ldots, (\tau_n, Z_n))} \\ &= \sum_{n=0}^{\infty} \int_0^t 1_{(\tau_n,\tau_{n+1}]}(s) 1_{\{Z_n \neq k\}} \frac{\mu^{Z_n k}(s, s-\tau_n) \bar{F}^{Z_n}(\tau_n, s)\, ds}{\bar{F}^{Z_n}(\tau_n, s)} \\ &= \sum_{n=0}^{\infty} \int_0^t 1_{(\tau_n,\tau_{n+1}]}(s) 1_{\{Z_n \neq k\}} \mu^{Z_n k}(s, s-\tau_n)\, ds \\ &= \int_0^t 1_{\{Z(s-) \neq k\}} \mu^{Z(s-)k}(s, U(s-))\, ds\end{aligned}$$

and the intensity process is $\lambda^k(t) = \mu^{Z(t-)k}(t, U(t-)) 1_{\{Z(t-) \neq k\}}$. By Proposition 10.6, the counting process $N^{jk}(t) = \int_0^t 1_{\{Z(s-)=j\}}\, dN^k(s)$ has intensity process $\lambda^{jk}(t) = 1_{\{Z(t-)=j\}} \lambda^k(t) = \mu^{jk}(t, U(t-)) 1_{\{Z(t-)=j\}}$. □

The transition probabilities of a semi-Markov process are for $0 \leq s < t$, $u, v \geq 0$ and $j, k \in \mathcal{S}$ denoted by

$$p^{jk}(s, t, u, v) = \mathbf{P}(Z(t) = k, U(t) \leq v | Z(s) = j, U(s) = u).$$

Theorem 12.12. *If the transition intensities $\mu^{jk}(t, u) \geq 0$ are continuous functions then $Z(t)$ is semi-Markov. Moreover the transition probabilities satisfy Kolmogorov backward equations*

$$p_s^{jk}(s, t, u, v) + p_u^{jk}(s, t, u, v) = \sum_{i: i \neq j} \mu^{ji}(s, u) \big(p^{jk}(s, t, u, v) - p^{ik}(s, t, 0, v)\big)$$

for $0 \leq s < t$, $u, v \geq 0$ and $j, k \in \mathcal{S}$.

Proof. We will only give a sketch of the proof. Based on Lemma 11.8, one can show that $(Z(t), U(t))$ is Markov (although the calculations are non-trivial).

To verify Kolmogorov equations, fix $t > 0$. By Corollary 11.9

$$M(s) = \mathbf{E}[1_{\{Z(t)=k\}} 1_{\{U(t) \leq v\}} | \mathcal{F}(s)]$$

is an FV-martingale. Using the Markov property, we get that

$$\begin{aligned}M(s) &= \mathbf{E}[1_{\{Z(t)=k\}} 1_{\{U(t) \leq v\}} | Z(s), U(s)] \\ &= p^{Z(s)k}(s, t, U(s), v).\end{aligned}$$

Applying Change of variables formula (Theorem 4.3) we get that

$$\begin{aligned}dM(s) &= \big(p_s^{Z(s)k}(s, t, U(s), v) + p_u^{Z(s)k}(s, t, U(s), v)\big)\, ds \\ &\quad + \sum_i \big(p^{ik}(s, t, 0, v) - p^{Z(s-)k}(s, t, U(s-), v)\big)\, dN^i(s)\end{aligned}$$

Then compute the intensity from this expression and one gets that

$$\begin{aligned}&p_s^{Z(s)k}(s, t, U(s), v) + p_u^{Z(s)k}(s, t, U(s), v) \\ &\quad = \sum_i \big(p^{Z(s)k}(s, t, U(s), v) - p^{ik}(s, t, 0, v)\big) \mu^{Z(s)i}(s, U(s)) 1_{\{Z(s) \neq i\}}.\end{aligned}$$

By setting $Z(s) = j$ and $U(s) = u$ one gets Kolmogorov equations. □

13. Exercises

Exercise 13.1. Prove Proposition 9.3. Hint: use tower property for conditional expectation.

Exercise 13.2. Let $M_1(t)$ and $M_2(t)$ be two martingales. For any two constants α and β show that $\alpha M_1(t) + \beta M_2(t)$ is a martingale.

Exercise 13.3. Prove Lemma 9.11 and Lemma 9.14.

Exercise 13.4. Let $N(t) = (N^1(t), \dots, N^m(t))$ be a m-variate counting process. Show that $N^\bullet(t) = \sum_{k=1}^m N^k(t)$ is a counting process.

Exercise 13.5. Prove the second part of Proposition 11.12.

Exercise 13.6. This exercise is to rewritte Proposition 10.6 to different cases of counting processes.

Let $N(t)$ be a counting process with intensity process $\lambda(t)$ and let $H(t)$ be a predictable process satisfying $\mathbf{E}\big[\int_0^t |H(s)|\lambda(s)\,ds\big] < \infty$.

(a) Show that

$$\int_0^t H(s)\,dN(s) - \int_0^t H(s)\lambda(s)\,ds$$

is a martingale and show for $0 \le s < t$ that

$$\mathbf{E}\Big[\int_s^t H(u)\,dN(u)\Big|\mathcal{F}(s)\Big] = \mathbf{E}\Big[\int_s^t H(u)\lambda(u)\,du\Big|\mathcal{F}(s)\Big] = \int_s^t \mathbf{E}[H(u)\lambda(u)|\mathcal{F}(s)]\,du.$$

Let $Z(t)$ be given as in Theorem 12.7 and let $N^k(t)$ be the associated counting process with intensity process $\lambda^k(t) = \mu^{Z(t-)k}(t)1_{\{Z(t-)\neq k\}}$. Let $H(t)$ be a predictable process satisfying $\mathbf{E}\big[\int_0^t |H(s)|\mu^{Z(s)k}(s)1_{\{Z(s)\neq k\}}\,ds\big] < \infty$ and let $f(t,j)$ be a (Borel) function that satisfies $\mathbf{E}\big[\int_0^t |f(s,Z(s))|\mu^{Z(s)k}(s)1_{\{Z(s)\neq k\}}\,ds\big] < \infty$.

(b) Show that

$$\int_0^t H(s)\,dN^k(s) - \int_0^t H(s)\mu^{Z(s)k}(s)1_{\{Z(s)\neq k\}}\,ds$$

is a martingale and show for $0 \le s < t$ that

$$\mathbf{E}\Big[\int_s^t H(u)\,dN^k(u)\Big|\mathcal{F}(s)\Big] = \mathbf{E}\Big[\int_s^t H(u)\mu^{Z(u)k}(u)1_{\{Z(u)\neq k\}}\,du\Big|\mathcal{F}(s)\Big].$$

(c) Show for $0 \le s < t$ that

$$\mathbf{E}\Big[\int_s^t f(u, Z(u-))\,dN^k(u)\Big|\mathcal{F}(s)\Big] = \sum_{j:j\neq k} \int_s^t p^{Z(s)j}(s,u)f(u,j)\mu^{jk}(u)\,du.$$

Exercise 13.7. Let $M(t)$ be a square integrable FV-martingale.

(a) Compute the dynamics of $M^2(t)$.
(b) Argue that $M^2(t)$ in general is not a martingale.
(c) Show that $M^2(t)$ is a martingale if $M(t)$ is continuous.

CHAPTER III

Life insurance models

This chapter gives applications of stochastic processes to life insurance models. The focus will be on different models of the multi-state policy. Throughout this chapter, the insurance contracts are issued at time 0 and terminating at time $T > 0$.

14. General definition of reserves

In this section we present some general concepts.

14.1. Interest. We need conversion factors to convert the value of money paid at one time to that paid at another. Let $v(s, t)$ be the value, at time s of 1 unit paid at time t (note 1 goes with the second coordinate). For $s < t$, $v(s, t)$ is the amount invested at time s in order to accumulate 1 unit at time t. For $u > t$, $v(u, t)$ is the amount that have accumulated at time u from an investment of 1 unit at time t. The function satisfies the relationship

$$v(s, u) = v(s, t)v(t, u) \quad \text{for all } s, t, u. \tag{14.1}$$

The basic properties for $v(\cdot, \cdot)$ are

- $v(t, t) = 1$
- $v(s, t) = v^{-1}(t, s)$.

For $s \leq t$, $v(s, t)$ is called the discount factor and $v^{-1}(s, t) = v(t, s)$ is called the accumulation factor. Define $v(t) = v(0, t)$ and by (14.1) we have that $v(s, t) = v(t)/v(s)$ and $v(0) = 1$. We assume that $v(t)$ is on the form

$$v(t) = \exp\Big(- \int_0^t r(v)\, dv\Big)$$

where $r(t)$ is the interest rate (or force of interest). The dynamics of the discount function and the accumulation function are $dv(t) = -v(t)r(t)\, dt$ and $dv^{-1}(t) = v^{-1}r(t)\, dt$. In general for $0 \leq s < t$, we have that

$$v(s, t) = \exp\Big(- \int_s^t r(v)\, dv\Big) \text{ and } v^{-1}(s, t) = \exp\Big(\int_s^t r(v)\, dv\Big).$$

Throughout this chapter, the investment portfolio of the insurance company bears interest with intensity $r(t)$ at time t, that is, $r(t)$ is the interest rate at time t.

Remark 14.1. We will always assume that the interest rate $r(t)$ is a deterministic function and it is integrable on $[0, T]$, that is,

$$\int_0^T |r(t)|\, dt < \infty.$$

14.2. Payment processes. A payment process (stream) describes payments floating between two parties, e.g. bank accounts or insurance contracts. In this note, payments are made continuously in time. Naturally, continuously paid payment processes do not appear in practice, but we can picture them as a limiting case of discrete payments (e.g. monthly payments). In context of insurance, the payments are the elements of the total payments in an insurance contract. The payment process is the difference of the total amount of benefits (outgoes) and premiums (incomes) which are positive and increasing. Therefore the mathematically definition of a payment process is of finite variation.

In this note there is two types of payments: payments that fall due continuously and lump sum payments. This introduces two types of FV-functions that will be used in this chapter.

- Lump sum payments. At time points $0 < t_1 < t_2 < \cdots < t_q \leq t$ we have the lump sum payment $\Delta x(t_i)$ then the piecewise constant function given by $x(t) = \sum_{0<s\leq t} \Delta x(t)$ is the accumulated payments in $(0, t]$ and is an FV-function.
- Continuously payments. If φ is the payment rate, then the integral $\int_0^t \varphi(s)\, ds$ is the accumulated payments in $(0, t]$ and is an FV-functions.

Therefore, in most cases the sample paths of payment processes are piecewise absolutely continuous (see Definition 2.13). We formalize this to mathematics.

The information currently available is given by a filtration $\mathcal{F}(t)$. The payment process $B(t)$ represents the payments commencing at time 0 and is the total amount paid in the time interval $[0, t]$. The payment process $B(t)$ is modelled as an adapted stochastic process where the sample paths are of finite variation and are CADLAG, that is, the payment process $B(t)$ is an FV-process. The payment process $B(t)$ represents the accumulated benefits and the accumulated premiums of the insurance contract. Let $B^{(b)}(t)$ be the accumulated benefits and let $B^{(p)}(t)$ be accumulated premiums. The payment process $B(t)$ represents the accumulated payments from the insurance company to the policy holder (benefits are positive and premiums are negative), that is, $B(t) = B^{(b)}(t) - B^{(p)}(t)$. Another decomposition of the payment process is payments that fall due continuously and lump sum payments. The accumulated lump sum payments are given by

$$B^d(t) = \sum_{0<s\leq t} \Delta B(s)$$

and the accumulated continuously payments are given by $B^c(t) = B(t) - B(0) - B^d(t)$. Note that for the continuously payments, we cannot specify an actual payment at any point of time but instead must speak of the rate (intensity) of payment, that is, $b(t)$ is the rate of payment if $dB^c(t) = b(t)\, dt$. Thus, we have the following decomposition

$$B(t) = B(0) + B^c(t) + B^d(t) = B(0) + \int_0^t b(s)\, ds + \sum_{0<s\leq t} \Delta B(s). \tag{14.2}$$

The decomposition in differential form (infinitesimal time interval), reads as

$$dB(t) = dB^c(t) + dB^d(t).$$

Remark 14.2. We will always assume that the payment process $B(t)$ is an integrable FV-process. Then by Doob-Meyer decomposition (Theorem 9.8) the payment process $B(t)$ has a predictable compensator $\Lambda^B(t)$ and $B(t) - \Lambda^B(t)$ is a martingale.

14.3. Present values and reserves. The future benefits are a liability for the insurance company. During the contract period the insurance company must provide a reserve to its future liabilities (benefits less premiums). The value at time t of the liabilities are the present value of future benefits less premiums over the time interval $(t,T]$ given by

$$Y(t,T] = \int_t^T \exp\Big(-\int_t^u r(v)\,dv\Big)\,dB(u). \tag{14.3}$$

This value is unknown at time t and cannot be used as a reserve. Technically, the present value $Y(t,T]$ depends only on the future and therefore it is a $\mathcal{F}^t$-measurable random variable (see Section 7 for a definition of $\mathcal{F}^t$). The assumptions in Remark 14.1 and Remark 14.2 insure that $Y(0,T]$ is integrable.

The prospective reserve is (defined to be) the expected present value of future benefits less premiums with regards to the information currently available, that is,

$$V(t) = \mathbf{E}\Big[\int_t^T \exp\Big(-\int_t^u r(v)\,dv\Big)\,dB(u)\Big|\mathcal{F}(t)\Big]. \tag{14.4}$$

This value is known at time t since the reserve $V(t)$ is adapted to the filtration $\mathcal{F}(t)$. Let $\Lambda^B(t)$ be the predictable compensator of the payment process $B(t)$ and using Proposition 10.6, we have another representation of the reserve

$$V(t) = \mathbf{E}\Big[\int_t^T \exp\Big(-\int_t^u r(v)\,dv\Big)\,d\Lambda^B(u)\Big|\mathcal{F}(t)\Big].$$

As the interest rate $r(t)$ is deterministic then by Theorem 10.11 we have that

$$V(t) = \mathbf{E}\Big[\int_t^T \exp\Big(-\int_t^u r(v)\,dv\Big)\,dB(u)\Big|\mathcal{F}(t)\Big] = \int_t^T \exp\Big(-\int_t^u r(v)\,dv\Big)\,d_u\mathbf{E}[B(u)|\mathcal{F}(t)].$$

By the martingale property we obtain that

$$V(t) = \int_t^T \exp\Big(-\int_t^u r(v)\,dv\Big)\,d_u\mathbf{E}[\Lambda^B(u)|\mathcal{F}(t)].$$

If the payment process has an intensity process $\lambda^B(t)$, that is, $d\Lambda^B(t) = \lambda^B(t)\,dt$ (see Definition 9.13) then the reserve in terms of the intensity process is given by

$$\begin{aligned} V(t) &= \mathbf{E}\Big[\int_t^T \exp\Big(-\int_t^u r(v)\,dv\Big)\lambda^B(u)\,du\Big|\mathcal{F}(t)\Big] \\ &= \int_t^T \mathbf{E}\Big[\exp\Big(-\int_t^u r(v)\,dv\Big)\lambda^B(u)\Big|\mathcal{F}(t)\Big]\,du. \end{aligned}$$

As the interest rate $r(t)$ is deterministic we have that

$$V(t) = \int_t^T \exp\Big(-\int_t^u r(v)\,dv\Big)\mathbf{E}\big[\lambda^B(u)\big|\mathcal{F}(t)\big]\,du.$$

In some of the models below it is possible to compute $\mathbf{E}[\lambda^B(u)|\mathcal{F}(t)]$.

14.4. Martingale method. We present a martingale method to compute the reserve. The method is to apply Corollary 9.7. Below are some general results that are applied in the next sections. Define the martingale (see Proposition 9.3) $M(t) = \mathbf{E}[Y(0,T)|\mathcal{F}(t)]$. In order to apply the martingale method we have to make the following assumption:

$$\text{Assumption 1: The martingale } M(t) \text{ is an FV-martingale.} \tag{14.5}$$

Then

$$\begin{aligned}
M(t) &= \mathbf{E}[Y(0,T)|\mathcal{F}(t)] \\
&= \mathbf{E}\Big[\int_0^T \exp\Big(-\int_0^s r(v)\,dv\Big)\,dB(s)\Big|\mathcal{F}(t)\Big] \\
&= \mathbf{E}\Big[\int_0^t \exp\Big(-\int_0^s r(v)\,dv\Big)\,dB(s) \\
&\qquad + \exp\Big(-\int_0^t r(v)\,dv\Big)\int_t^T \exp\Big(-\int_t^u r(v)\,dv\Big)\,dB(u)\Big|\mathcal{F}(t)\Big] \\
&= \int_0^t \exp\Big(-\int_0^s r(v)\,dv\Big)\,dB(s) \\
&\qquad + \exp\Big(-\int_0^t r(v)\,dv\Big)\mathbf{E}\Big[\int_t^T \exp\Big(-\int_t^u r(v)\,dv\Big)\,dB(u)\Big|\mathcal{F}(t)\Big] \\
&= \int_0^t \exp\Big(-\int_0^s r(v)\,dv\Big)\,dB(s) + \exp\Big(-\int_0^t r(v)\,dv\Big)V(t).
\end{aligned}$$

Since all processes in the expression are FV-processes except for the reserve $V(t)$ then the reserve must also be an FV-process. Then the martingale $M(t)$ has dynamics

$$\begin{aligned}
dM(t) = \exp\Big(-\int_0^t r(v)\,dv\Big)\,dB(t) + \exp\Big(-\int_0^t r(v)\,dv\Big)\,dV(t) \\
- r(t)\exp\Big(-\int_0^t r(v)\,dv\Big)V(t)\,dt.
\end{aligned}$$

Multiplying with accumulation function we get that

$$\exp\Big(\int_0^t r(v)\,dv\Big)\,dM(t) = dB(t) + dV(t) - r(t)V(t)\,dt.$$

Note that the right-hand side is an FV-martingale. Note also that the reserve is integrable and hence $V(t)$ has a predictable compensator $\Lambda^V(t)$ and let $\Lambda^B(t)$ be the predictable compensator of the payment process $B(t)$. Then the two processes $B(t) - \Lambda^B(t)$ and $V(t) - \Lambda^V(t)$ are martingales. We can use the two predictable compensators to decompose the martingale $\int_0^t \exp\big(\int_0^s r(v)\,dv\big)\,dM(s)$ into two martingales and a predictable term

$$\begin{aligned}
\exp\Big(\int_0^t r(v)\,dv\Big)\,dM(t) &= \big(dB(t) - d\Lambda^B(t)\big) + \big(dV(t) - d\Lambda^V(t)\big) \\
&\qquad + \big(d\Lambda^B(t) + d\Lambda^V(t) - r(t)V(t)\,dt\big).
\end{aligned}$$

From this expression we get that the predictable term $d\Lambda^B(t) + d\Lambda^V(t) - r(t)V(t)\,dt$ is an FV-martingale and by Proposition 9.6 we have that

$$d\Lambda^B(t) + d\Lambda^V(t) - r(t)V(t)\,dt = 0. \tag{14.6}$$

Assume that the payment process has an intensity process $\lambda^B(t)$ and the reserve $V(t)$ has an intensity process $\lambda^V(t)$. The equation (14.6) reads as $\big(\lambda^B(t) + \lambda^V(t) - r(t)V(t)\big)\,dt = 0$ which implies that

$$\lambda^B(t) + \lambda^V(t) - r(t)V(t) = 0. \tag{14.7}$$

If the two intensity processes have right-limits we can equivalent formulate the equation by

$$\lambda^B(t+) + \lambda^V(t+) - r(t)V(t) = 0. \tag{14.8}$$

Given a model, the problem is to find—if possible—the two intensity process and $\lambda^B(t)$ and $\lambda^V(t)$. Then from the equation

$$V(t) = \frac{\lambda^B(t) + \lambda^V(t)}{r(t)}$$

we might derive some (differentiable) equations for e.g. state-wise reserves.

14.5. Principle of equivalence. Let $B(t)$ be a payment process that represents premiums and benefits of an insurance contract, that is, $B(t) = B^{(b)}(t) - B^{(p)}(t)$ where $B^{(b)}(t)$ is the accumulated benefits and $B^{(p)}(t)$ is the accumulated premiums. The total value at time 0 of premiums less benefits is

$$-B(0) - \int_0^T \exp\Big(-\int_0^t r(s)\,ds\Big)\,dB(t)$$

where $-B(0)$ is a single premium paid at time 0. The principle of equivalence insures that the expected present value of premiums and benefits should be equal, that is, the payment process $B(t)$ satisfies the principle of equivalence if

$$\mathbf{E}\Big[\int_0^T \exp\Big(-\int_0^t r(s)\,ds\Big)\,dB(t)\Big] + B(0) = 0$$

or

$$B^+(0) + \mathbf{E}\Big[\int_0^T \exp\Big(-\int_0^t r(s)\,ds\Big)\,dB^{(b)}(t)\Big] = B^-(0) + \mathbf{E}\Big[\int_0^T \exp\Big(-\int_0^t r(s)\,ds\Big)\,dB^{(p)}(t)\Big].$$

The idea is that expected value is averaging the balance on profits and losses on the individual policies.

15. Multi-state policy, general model

The multi-state model will be a model for a life insurance policy where the payment of benefits and premiums depends on being in a given state or moving between states. The states represent different conditions for the life insurance policy such that the policy at any time is in one and only one state.

Consider an insurance policy issued at time 0 and expires at time T. There is a finite set $\mathcal{S} = \{1, 2, \ldots, m\}$ of states of the policy. Let $Z(t)$ denote the state of the policy at time t and assume that policy commencing in state 1, that is, $Z(0) = 1$.

15.1. Insurance model. The state of contract $Z(t)$ is modeled by a piecewise constant process with values in the finite state space $\mathcal{S}$, see Section 12. As in Section 12, let the jump time τ_n be the time point of the n-th occurrence of an (insurance) event and Z_n is the state of the policy of the n-th (insurance) event, that is, $((\tau_n, Z_n), n = 1, 2, \ldots)$ is the associated multivariate point process. Thus the state of the contract can be represented by $Z(t) = Z_n$ if $t \in [\tau_n, \tau_{n+1})$. Recall from Section 12, that there is an one-to-one connection between $Z(t)$ and the multivariate counting processes $N^{jk}(t) = \#\{0 < s \le t | Z(s-) = j, Z(s) = k\}$ for $j, k = 1, \ldots, m$, $j \neq k$ and $N^k(t) = \#\{0 < s \le t | Z(s-) \neq k, Z(s) = k\}$ for $k = 1, 2, \ldots, m$.

Remark 15.1. Since the filtration is a point process filtration then by Corollary 11.9 all martingales are FV-martingales. Then Assumption 1 (see (14.5)) in the martingale method is satisfied. Thus, the martingale method can be applied for the multi-state policy model.

16. Standard multi-state policy, Markov model

For the standard model, the state of the policy is model by a Markov chain and the payments might depend on the state of the policy and time.

16.1. Insurance Model. Consider the multi-state policy described in Section 15. The state of contract $Z(t)$ is modelled by a Markov chain with given transition intensities $j \neq k$ and set $\mu^j(t) = \sum_{k:k\neq j} \mu^{jk}(t)$. Assume that transition intensities $\mu^{jk}(t)$ are locally bounded functions and hence $\int_0^T \mu^{jk}(s)\, ds < \infty$. By Remark 12.6 we have that $N^{jk}(t)$ is an integrable counting process and there can only be finite number of (insurance) events in the model (and the state of the policy $Z(t)$ is a piecewise constant process).

16.2. The contractual payments. The contractual payment functions are specified by

- During sojourn in state j, the deterministic payment function is given by
$$dB^j(t) = d(B^j)^{(d)}(t) + b^j(t)\, dt$$
where the lump sum payments are $(B^j)^{(d)}(t) = \sum_{0<s\le t} \Delta B^j(s)$ at deterministic time points $0 < t_1^{(j)} < t_2^{(j)} < \cdots < t_{q_j}^{(j)} \le n$ and $b^j(t)$ is the payment rate at time t.
- For $j \neq k$, a sum assured of $b^{jk}(t)$ payable immediately on a transition from state j to state k at time t.

The dynamics of the payment process is described by
$$dB(t) = \sum_j 1_{\{Z(t)=j\}} dB^j(t) + \sum_k b^{Z(t-)k}(t)\, dN^k(t).$$
In details the payment process is given by
$$B(t) = B(0) + \int_0^t b^{Z(s)}(s)\, ds + \sum_{0<s\le t} \Delta B^{Z(s)}(s) + \sum_k \int_0^t b^{Z(s-)k}(s)\, dN^k(s)$$
where $B(0)$ is an initial lump sum payment, that is, $-B(0)$ is a premium paid when the contract is signed. The payment functions have the following interpretation: $\Delta B^j(t)$ represents a general endowment, $b^j(t)$ represents a general life annuity, and $b^{jk}(t)$ represents a general term insurance. The deterministic payment functions $b^j(t)$ and $b^{jk}(t)$ are assumed to be locally bounded functions. Then the payment process $B(t)$ is an integrable FV-process and the condition in Remark 14.2 Below we use the decomposition of the payment process $dB(t) = dB_1(t) + dB_2(t)$ where $dB_1(t) = \sum_j 1_{\{Z(t)=j\}} dB^j(t)$ and $dB_2(t) = \sum_k b^{Z(t-)k}(t)\, dN^k(t)$.

16.3. Reserves and Thiele equations. The present value of future benefits less premiums $Y(t,T]$ (see equation (14.3)) is a liability for the insurance company for which the insurance company is to provide a reserve (see equation (14.4)) given by

$$V(t) = \mathbf{E}[Y(t,T]|\mathcal{F}(t)].$$

Since the present value of future benefits less premiums $Y(t,T]$ only depends on the future, $Y(t,T]$ is $\mathcal{F}^t$-measurable and by the Markov property in Proposition 7.2(ii)) we have that $V(t) = \mathbf{E}[Y(t,T]|\mathcal{F}(t)] = \mathbf{E}[Y(t,T]|Z(t)]$. Given that $Z(t) = j$, the state-wise reserves given by

$$V^j(t) = \mathbf{E}[Y(t,T]|Z(t) = j]$$

are deterministic functions, $j = 1, 2, \ldots, m$. Moreover, we have that

$$V(t) = V^{Z(t)}(t).$$

Below, we derive Thiele equations for the state-wise reserves.

First we find the intensity process of the payment process $B_2(t)$ involving the counting processes. Recall, from Proposition 12.5 that $N^k(t)$ has intensity process $\lambda^k(t)$ given by $\lambda^k(t) = \mu^{Z(t-)k}(t)1_{\{Z(t-)\neq k\}}$. By Proposition 10.6 we have that the intensity process of $B_2(t)$ is given by

$$\lambda^{B_2}(t) = \sum_k b^{Z(t-)k}(t)\mu^{Z(t-)k}(t)1_{\{Z(t-)\neq k\}}.$$

Informally we have that $\mu^{Z(t-)k}(t)1_{\{Z(t-)\neq k\}}\,dt = \mathbf{E}[dN^k(t)|\mathcal{F}(t-)]$. The payment process $B_2(t)$ has an intensity process $\lambda^{B_2}(t)$ given by

$$\begin{aligned}\lambda^{B_2}(t)\,dt = \mathbf{E}[dB_2(t)|\mathcal{F}(t-)] &= \sum_k b^{Z(t-)k}(t)\,\mathbf{E}[dN^k(t)|\mathcal{F}(t-)]\\ &= \sum_k b^{Z(t-)k}(t)\mu^{Z(t-)k}(t)1_{\{Z(t-)\neq k\}}\,dt.\end{aligned}$$

For $t \leq u$ we have that

$$\begin{aligned}\mathbf{E}[\lambda^{B_2}(u)|\mathcal{F}(t)]\,du &= \mathbf{E}\Big[\sum_k b^{Z(u)k}(u)\mu^{Z(u)k}(u)1_{\{Z(u)\neq k\}}\Big|Z(t)\Big]\,du\\ &= \sum_j \Big(\sum_k b^{jk}(u)\mu^{jk}(u)1_{\{j\neq k\}}\Big)\mathbf{P}\big(Z(u) = j|Z(t)\big)\,du\\ &= \sum_j \Big(p^{Z(t)j}(t,u)\sum_{k:k\neq j} b^{jk}(u)\mu^{jk}(u)\Big)\,du.\end{aligned}$$

Next for $t \le u$ we compute

$$\begin{aligned}
\mathbf{E}[B_1(u)|\mathcal{F}(t)] &= \mathbf{E}\Big[\sum_j \int_0^u 1_{\{Z(v)=j\}}\, dB^j(v)\Big|\mathcal{F}(t)\Big] \\
&= \sum_j \int_0^u \mathbf{E}[1_{\{Z(v)=j\}}|\mathcal{F}(t)]\, dB^j(v) \\
&= \sum_j \int_0^u \mathbf{E}[1_{\{Z(v)=j\}}|Z(t)]\, dB^j(v) \\
&= \sum_j \int_0^u p^{Z(t)j}(t,v)\, dB^j(v)
\end{aligned}$$

and the dynamics is given by $d_u\mathbf{E}[B_1(u)|\mathcal{F}(t)] = \sum_j p^{Z(t)j}(t,u)\, dB^j(u)$. The reserve can now be computed

$$\begin{aligned}
V(t) &= \mathbf{E}\Big[\int_t^T \exp\Big(-\int_t^u r(v)\, dv\Big)\big(dB_1(u) + dB_2(u)\big)\Big|\mathcal{F}(t)\Big] \\
&= \int_t^T \exp\Big(-\int_t^u r(v)\, dv\Big)\big(d_u\mathbf{E}[B_1(u)|\mathcal{F}(t)] + \mathbf{E}[\lambda^{B_2}(u)|\mathcal{F}(t)]du\big) \\
&= \int_t^T \exp\Big(-\int_t^u r(v)\, dv\Big)\Big(\sum_j p^{Z(t)j}(t,u)\Big(dB^j(u) + \sum_{k:k\neq j} b^{jk}(u)\mu^{jk}(u)\, du\Big)\Big).
\end{aligned}$$

Thus, the state-wise $(Z(t) = i)$ reserves are given by

$$\begin{aligned}
V^i(t) &= \mathbf{E}[Y(t,T)|Z(t) = i] \\
&= \int_t^T \exp\Big(-\int_t^u r(v)\, dv\Big)\sum_j p^{ij}(t,u)\Big(dB^j(u) + \sum_{k:k\neq j} b^{jk}(u)\mu^{jk}(u)\, du\Big). \qquad (16.1)
\end{aligned}$$

By employing Kolmogorov backward integral equations (Theorem 12.7) we get Thiele's integral equations for the state-wise reserves:

Theorem 16.1. *The state-wise reserves $V^j(t)$, $j = 1, 2, \dots, m$, are solutions to the following system of integral equations*

$$\begin{aligned}
V^j(t) &= \int_t^T \exp\Big(-\int_t^u \big(\mu^j(v) + r(v)\big)\, dv\Big)\Big(dB^j(u) + \sum_{k:k\neq j} \mu^{jk}(u)\big(b^{jk}(u) + V^k(u)\big)\, du\Big) \\
&= \int_t^T \exp\Big(-\int_t^u \big(\mu^j(v) + r(v)\big)\, dv\Big)\Big(b^j(u) + \sum_{k:k\neq j} \mu^{jk}(u)\big(b^{jk}(u) + V^k(u)\big)\Big)\, du \\
&\quad + \sum_{t<u\le n} \exp\Big(-\int_t^u \big(\mu^j(v) + r(v)\big)\,\Delta B^j(u).
\end{aligned}$$

From the integral equations we see that the jumps of the state-wise reserves are $\Delta V^j(t) = -\Delta B^j(t)$. Moreover that the state-wise reserves are differentiable at t if $\Delta B^j(t) = 0$ and the functions $r(t)$, $\mu^{jk}(t)$, $b^j(t)$, and $b^{jk}(t)$ are continuous at t. In the theorem below, for simplicity we make assumptions such that state-wise reserves are differentiable in $[0, T]$ and by differentiating the integral equations we obtain Thiele differential equations.

Theorem 16.2. *Assume that interest rate $t \mapsto r(t)$, the transition intensities $t \mapsto \mu^{jk}(t)$, and the payment functions $t \mapsto b^j(t)$ and $t \mapsto b^{jk}(t)$ are continuous on $[0,T]$. Moreover, assume there are no lump sum payments in a state, that is, $\Delta B^j(t) = 0$. Then the state-wise reserves are solutions to Thiele differential equations*

$$V_t^j(t) = r(t)V^j(t) - b^j(t) - \sum_{k:k\neq j} \mu^{jk}(t)\big(b^{jk}(t) + V^k(t) - V^j(t)\big)$$

with boundary condition $V^j(T) = 0$ for $j = 1, 2, \ldots, m$.

16.4. Martingale method. We will use the martingale method from Section 14 to derive Thiele differential equations given in Theorem 16.2. We will explore the connection between the state-wise reserves and Thiele differential equations, that is, we will give a stochastic representation (state-wise reserves) of the differential equation.

As in Theorem 16.2, we assume that $t \mapsto r(t)$, $t \mapsto \mu^{jk}(t)$, $t \mapsto b^j(t)$, and $t \mapsto b^{jk}(t)$ are continuous on [0,n] and that $\Delta B^j(t) = 0$. In order to apply equation (14.7) (or equation (14.8)), we must derive the intensity process of the payment process $B(t)$ and the reserve $V(t)$.

Recall that $V(t) = V^{Z(t)}(t)$ where $V^j(t)$ are the state-wise reserves. We assume that the state-wise reserves are continuous differentiable, that is $V_t^j(t)$ exists and is continuous (note that this follows from Theorem 16.2 and cannot be proved by the martingale method). With this assumption we can apply the changes of variables formula (Corollary 4.5)

$$\begin{aligned} dV(t) &= d\big(V^{Z(t)}(t)\big) \\ &= V_t^{Z(t)}(t)\,dt + \sum_k \big(V^{Z(t)}(t) - V^{Z(t-)}(t-)\big)\,dN^k(t) \\ &= V_t^{Z(t)}(t)\,dt + \sum_k \big(V^k(t) - V^{Z(t-)}(t)\big)\,dN^k(s). \end{aligned} \tag{16.2}$$

It follows by Proposition 10.6 that the intensity process $\lambda^V(t)$ of the reserve $V(t)$ is given by

$$\lambda^V(t) = V_t^{Z(t-)}(t) + \sum_k \big(V^k(t) - V^{Z(t-)}(t)\big)\mu^{Z(t-)k}(t)1_{\{Z(t-)\neq k\}}.$$

Informally, the intensity process $\lambda^V(t)$ is given by

$$\begin{aligned} \lambda^V(t)\,dt &= \mathbf{E}[dV(t)|\mathcal{F}(t-)] \\ &= V_t^{Z(t-)}(t)\,dt + \sum_k \big(V^k(t) - V^{Z(t-)}(t)\big)\,\mathbf{E}[dN^k(t)|\mathcal{F}(t-)] \\ &= \Big(V_t^{Z(t-)}(t) + \sum_k \big(V^k(t) - V^{Z(t-)}(t)\big)\mu^{Z(t-)k}(t)1_{\{Z(t-)\neq k\}}\Big)\,dt. \end{aligned}$$

The dynamics of the payment process $B(t)$ is

$$dB(t) = b^{Z(t)}(t)\,dt + \sum_k b^{Z(t-)k}(t)\,dN^k(t), \quad B(0) = 0.$$

It follows by Proposition 10.6 that the intensity process $\lambda^B(t)$ of the payment process $B(t)$ is given by

$$\begin{aligned}\lambda^B(t)\,dt &= \mathbf{E}[dB(t)|\mathcal{F}(t-)]\\ &= b^{Z(t-)}(t)\,dt + \sum_k b^{Z(t-)k}(t)\mathbf{E}[dN^k(t)|\mathcal{F}(t-)]\\ &= \Big(b^{Z(t-)}(t) + \sum_k b^{Z(t-)k}(t)\mu^{Z(t-)k}(t)1_{\{Z(t-)\neq k\}}\Big)\,dt.\end{aligned}$$

Now we can apply equation (14.7) in the martingale method

$$\begin{aligned}0 &= \lambda^B(t+) + \lambda^V(t+) - r(t)V(t)\\ &= b^{Z(t)}(t) + V_t^{Z(t)}(t) - r(t)V(t) + \sum_k \big(b^{Z(t)k}(t) + V^k(t) - V^{Z(t)}(t)\big)\mu^{Z(t)k}(t)1_{\{Z(t)\neq k\}}.\end{aligned}$$

With initial value $Z(t) = j$, we have shown the following result (with some abuse of notation).

Theorem 16.3. *Assume that the state-wise reserves $V^j(t)$, $j = 1, 2, \ldots, m$, are continuous differentiable (C^1-functions). Then the state-wise reserves are solutions to Thiele differential equations*

$$V_t^j(t) = r(t)V^j(t) - b^j(t) - \sum_{k:k\neq j} \mu^{jk}(t)\big(b^{jk}(t) + V^k(t) - V^j(t)\big)$$

with boundary condition $V^j(T) = 0$ for $j = 1, 2, \ldots, m$.

16.5. Moments of present values. We want to compute higher order moments of the present value of future benefits less premiums. By Markov property we need the the state-wise conditional moments given by

$$(V^j)^{(n)}(t) = \mathbf{E}[(Y(t,T])^n|Z(t) = j]$$

and the functions are determined by the differential equations in the following propositions

Proposition 16.4. *Let the assumptions (setup) be as in Theorem 16.2. Then the state-wise conditional moments are solutions to the differential equations*

$$\begin{aligned}(V^j)_t^{(n)}(t) &= \big(nr(t) + \mu^j(t)\big)(V^j)^{(n)}(t) - nb^j(t)(V^j)^{(n-1)}(t)\\ &\quad - \sum_{k:k\neq j} \mu^{jk}(t) \sum_{p=0}^{n} \binom{n}{p} \big(b^{jk}(t)\big)^p (V^k)^{(n-p)}(t).\end{aligned}$$

17. Models with state duration, semi-Markov model

In this section we extend the analysis to models where payments and transition intensities depend also on the duration of time in the present state.

17.1. Insurance model. Consider the multi-state policy described in Section 15. The state of contract $Z(t)$ is modeled by a semi-Markov process (see Section 12). Recall that $(Z(t), U(t))$ is a Markov process where $U(t)$ is the duration process. The semi-Markov process is with given transition intensities $\mu^{jk}(t, u)$. The transition intensities $(t, u) \mapsto \mu^{jk}(t, u)$ are assumed to be continuous. Then $N^{jk}(t)$ is an integrable counting process and there can only be finite number of (insurance) events in the model (and the state of the policy $Z(t)$ is a piecewise constant process).

17.2. The contractual payments. The contractual payment functions are specified by

- $b^j(t,u)$ is the payment rate in state j at time t. Assume that the payment function $(t,u) \mapsto b^j(t,u)$ is continuous. There are no lump sum payments in a state: $\Delta B^j(t) = 0$.
- $b^{jk}(t,u)$ is the amount paid to the policy holder due upon transitions from state j to state k, $j \neq k$. Assume that the payment function $(t,u) \mapsto b^{jk}(t,u)$ is continuous.

Therefore the dynamics of the payment process $B(t)$ is

$$dB(t) = b^{Z(t)}(t, U(t))\, dt + \sum_k b^{Z(t-)k}(t, U(t-))\, dN^k(t), \quad B(0) = 0.$$

Then $B(t)$ is an integrable FV-process. Finally, assume that the interest rate $r(t)$ is also a continuous function.

17.3. Reserves and extended Thiele equations. The insurance company is to provide a reserve given by

$$\begin{aligned} V(t) &= \mathbf{E}\Big[\int_t^T \exp\Big(-\int_t^u r(v)\, dv\Big)\, dB(u)\Big|\mathcal{F}(t)\Big] \\ &= \mathbf{E}\Big[\int_t^T \exp\Big(-\int_t^u r(v)\, dv\Big)\, dB(u)\Big|Z(t), U(t)\Big] = V^{Z(t)}(t, U(t)) \end{aligned}$$

in the second equality, we have used the Markov property (see Proposition 7.2) since the present value $Y(t,T]$ depends only on the future. Condition on $Z(t) = k$ and $U(t) = u$, we have the state-wise reserves $V^k(t,u) = \mathbf{E}[Y(t,T]|Z(t) = k, U(t) = u]$ for $k = 1, 2, \ldots, m$ and $u \geq 0$. Below we find equations for the state-wise reserves using the martingale method.

It follows by Proposition 10.6 that the payment process $B(t)$ has an intensity process $\lambda^B(t)$ given by

$$\begin{aligned} \lambda^B(t)\, dt &= \mathbf{E}[dB(t)|\mathcal{F}(t-)] \\ &= b^{Z(t)}(t, U(t))\, dt + \sum_k b^{Z(t-)k}(t, U(t-))\, \mathbf{E}[dN^k(t)|\mathcal{F}(t-)] \\ &= \Big(b^{Z(t-)}(t, U(t-)) + \sum_k b^{Z(t-)k}(t, U(t-)) \mu^{Z(t-)k}(t, U(t-)) 1_{\{Z(t-)\neq k\}}\Big)\, dt. \end{aligned}$$

We assume that $V_t^j(t,u)$ and $V_u^j(t,u)$ exist and are continuous. With this assumption we can apply the changes of variables formula (Corollary 4.5)

$$\begin{aligned} &dV(t) \\ &= d\big(V^{Z(t)}(t, U(t))\big) \\ &= V_t^{Z(t)}(t, U(t))\, dt + V_u^{Z(t)}(t, U(t))\, dU^c(t) + \sum_k \big(V^{Z(t)}(t, U(t)) - V^{Z(t-)}(t-, U(t-))\big)\, dN^k(t) \\ &= \big(V_t^{Z(t)}(t, U(t)) + V_u^{Z(t)}(t, U(t))\big)\, dt + \sum_k \big(V^k(t, 0) - V^{Z(t-)}(t, U(t-))\big)\, dN^k(t). \end{aligned}$$

It follows by Proposition 10.6 that the reserve $V(t)$ has an intensity process $\lambda^V(t)$ given by

$$\begin{aligned}\lambda^V(t) &= V_t^{Z(t-)}(t,U(t-)) + V_u^{Z(t-)}(t,U(t-)) \\ &\quad + \sum_k \big(V^k(t,0) - V^{Z(t-)}(t,U(t-))\big)\mu^{Z(t-)k}(t,U(t-))1_{\{Z(t-)\neq k\}}.\end{aligned}$$

Informally, the intensity process $\lambda^V(t)$ is given by

$$\begin{aligned}\lambda^V(t)\,dt &= \mathbf{E}[dV(t)|\mathcal{F}(t-)] \\ &= \Big(V_t^{Z(t-)}(t,U(t-)) + V_u^{Z(t-)}(t,U(t-))\Big)\,dt \\ &\quad + \sum_k \big(V^k(t,0) - V^{Z(t-)}(t,U(t-))\big)\,\mathbf{E}[dN^k(t)|\mathcal{F}(t-)] \\ &= \Big(V_t^{Z(t-)}(t,U(t-)) + V_u^{Z(t-)}(t,U(t-)) \\ &\quad + \sum_k \big(V^k(t,0) - V^{Z(t-)}(t,U(t-))\big)\mu^{Z(t-)k}(t,U(t-))1_{\{Z(t-)\neq k\}}\Big)\,dt.\end{aligned}$$

Now we can apply equation (14.7) in the martingale method

$$\begin{aligned}0 &= \lambda^B(t+) + \lambda^V(t+) - r(t)V(t) \\ &= b^{Z(t)}(t,U(t)) + V_t^{Z(t)}(t,U(t)) + V_u^{Z(t)}(t,U(t)) - r(t)V(t) \\ &\quad + \sum_k \big(b^{Z(t)k}(t,U(t)) + V^k(t,0) - V^{Z(t)}(t,U(t))\big)\mu^{Z(t)k}(t,U(t))1_{\{Z(t)\neq k\}}\end{aligned}$$

Finally, condition on $Z(t) = j$ and $U(t) = u$ in the above expression we have derived the following theorem.

Theorem 17.1. *Assume that the state-wise reserves $V^j(t,u)$, $j = 1,2,\dots,m$, are continuous differentiable (in boths variables t and u). Then the state-wise reserves are solutions to the partial differential equations*

$$V_t^j(t,u) + V_u^j(t,u) = r(t)V^j(t,u) - b^j(t,u) - \sum_{k:k\neq j} \mu^{jk}(t,u)\big(b^{jk}(t,u) + V^k(t,0) - V^j(t,u)\big)$$

with boundary condition $V^j(T,u) = 0$ for $u \geq 0$ and $j = 1,2,\dots,m$.

18. Surplus and dividends

Life insurance policies are typically long term contracts, with possible unforseenable variations in intensities and interest in the contract period. Hence insurance company has a risk that the company only can protect against by build into contractual premium a safety loading. The safety loading will create a surplus which belongs to the policy holder. Thus dividends must be added to the contractual payments.

18.1. The insurance contract. Consider the standard multi-state present in Section 16 where the setup is as in Theorem 16.2. Recall that the state of the contract, $Z(t)$, is a Markov process with given continuous transition functions $\mu^{jk}(t)$ and the accumulated contractual payments are on differential form described by

$$dB(t) = b^{Z(t)}(t)\,dt + \sum_k b^{Z(t-)k}(t)\,dN^k(t)$$

where $b^j(t)$ and $b^{jk}(t)$ are continuous deterministic functions. Let $r(t)$ be a continuous deterministic interest rate.

We consider the basis (r, μ) which specifies an interest rate and a family of transition intensities. The basis is called the real basis or second order basis and represents the true mechanisms governing the insurance contract. On this basis we have the state-wise reserves given by

$$V^j(t) = \mathbf{E}\Big[\int_t^T \exp\Big(-\int_t^u r(v)\,dv\Big)\,dB(u)\Big|Z(t)=j\Big].$$

From Theorem 16.2 we have Thiele differential equations for the state-wise reserves given by

$$V_t^j(t) = r(t)V^j(t) - b^j(t) - \sum_{k:k\neq j} \mu^{jk}(t)\big(b^{jk}(t) + V^k(t) - V^j(t)\big).$$

18.2. Technical reserves. An important quantity in life insurance is the technical basis or first order basis (r^*, μ^*) (the expected value with intensity is denoted $\mathbf{E}^*$). The technical basis represents a prudent initial assessment for future development of the interest rate and transition intensities. The associated technical state-wise reserves are given by

$$V^{*j}(t) = \mathbf{E}^*\Big[\int_t^T \exp\Big(-\int_t^u r^*(v)\,dv\Big)\,dB(u)\Big|Z(t)=j\Big]$$

which are solutions to Thiele differential equations

$$V_t^{*j}(t) = r^*(t)V^{*j}(t) - b^j(t) - \sum_{k:k\neq j} \mu^{*jk}(t)\big(b^{jk}(t) + V^{*k}(t) - V^{*j}(t)\big).$$

The technical reserve is given by

$$V^*(t) = V^{*Z(t)}(t) = \mathbf{E}^*\Big[\int_t^T \exp\Big(-\int_t^u r^*(v)\,dv\Big)\,dB(u)\Big|Z(t)\Big].$$

The payment functions are based on principle of equivalence (see Section 14) under the first order basis

$$\mathbf{E}^*\Big[\int_0^T \exp\Big(-\int_0^s r^*(v)\,dv\Big)\,dB(s)\Big] = 0 \tag{18.1}$$

or (since $Z(0) = 1$) $V^{*1}(0) = 0$. The payment process which results from this equation is called the first order payment process and is the payments that are specified in the contract when it is issued. The principle of equivalence gives a balance between the first order expected present values of benefits and premiums.

18.3. Dividends. The reserves accumulate by the real basis, that is, the pair (r, μ) containing the real interest rate and the real transition intensities. With payment functions determined by principle of equivalence (18.1) based on prudent technical basis, then the first order payments under the real basis might not satisfy the principle of equivalence in the sense that $V^1(0) \neq 0$. Therefore, the insurance company adds dividends to the first order payments. The dividend process $D(t)$ are the accumulated dividend payments. The purpose of the dividends is to establish the principle of equivalence under the real basis

$$\mathbf{E}\Big[\int_0^T \exp\Big(-\int_0^s r(v)\,dv\Big)\,d(B+D)(s)\Big] = 0.$$

The dividend process $D(t)$ is modeled by

$$dD(t) = \delta^{Z(t)}(t)\,dt + \sum_k \delta^{Z(t-)k}(t)\,dN^k(t)$$

where $\delta^j(t)$ and $\delta^{jk}(t)$ are adapted processes.

18.4. Surplus. With premiums based on prudent first order basis there will be created a systematic surplus if everything goes well. We define the individual surplus at time t by

$$\begin{aligned} X(t) &= \int_0^t \exp\Big(\int_s^t r(v)\,dv\Big)\,d(-B-D)(s) - V^*(t) \\ &= -\exp\Big(\int_0^t r(v)\,dv\Big)\int_0^t \exp\Big(-\int_0^s r(v)\,dv\Big)\,d(B+D)(s) - V^*(t) \end{aligned}$$

which is past net income (premiums less benefits and dividends) compounded with the real interest minus the technical reserve. Using Integration by parts, Theorem 4.1, we get that the dynamics of the individual surplus is

$$\begin{aligned} dX(t) &= -d(B+D)(t) - dV^*(t) \\ &\qquad - r(t)\exp\Big(\int_0^t r(v)\,dv\Big)\int_0^t \exp\Big(-\int_0^s r(v)\,dv\Big)\,d(B+D)(s)\,dt \\ &= r(t)\big(X(t)+V^*(t)\big)\,dt - d(B+D)(t) - dV^*(t) \\ &= r(t)X(t)\,dt + d(C-D)(t) \end{aligned}$$

where the process $C(t)$ is given by

$$dC(t) = r(t)V^*(t)\,dt - dB(t) - dV^*(t).$$

Recall from equation (16.2) the dynamics of technical reserve and then use Thiele differential equation we get that

$$\begin{aligned} dV^*(t) &= V_t^{*Z(t)}(t)\,dt + \sum_k \big(V^{*k}(t) - V^{*Z(t-)}(t)\big)\,dN^k(t) \\ &= \big(r^*(t)V^*(t) - b^{Z(t)}(t) - \sum_k \mu^{Z(t)k}(t)(b^{Z(t)k}(t) + V^{*k}(t) - V^{*Z(t)}(t))\big)\,dt \\ &\qquad + \sum_k \big(V^{*k}(t) - V^{*Z(t-)}(t)\big)\,dN^k(t) \end{aligned}$$

and hence the dynamics of $C(t)$ has the form

$$dC(t) = c^{Z(t)}(t)\,dt + \sum_k c^{Z(t-)k}(t)\,dN^k(t)$$

with

$$\begin{aligned} c^j(t) &= \big(r(t) - r^*(t)\big)V^{*j}(t) + \sum_{k:k\neq j} \mu^{*jk}(t)\Big(b^{jk}(t) + V^{*k}(t) - V^{*j}(t)\Big) \\ c^{jk}(t) &= -\big(b^{jk}(t) + V^{*k}(t) - V^{*j}(t)\big). \end{aligned}$$

The process $C(t)$ is called the contribution process.

18.5. Surplus-linked dividends. We study a dividend process where the dividend payments are linked to surplus. Define the processes by

$$\begin{aligned}\delta^j(t) &= \bar{\delta}^j\big(t, X(t)\big)\\ \delta^{jk}(t) &= \bar{\delta}^{jk}\big(t, X(t-)\big).\end{aligned}$$

Thus, the dividend process has dynamics

$$dD(t) = \bar{\delta}^{Z(t)}\big(t, X(t)\big)\,dt + \sum_k \bar{\delta}^{Z(t-)k}\big(t, X(t-)\big)\,dN^k(t).$$

With this form of dividends, the purpose is to compute the reserve for the contractual payments $B(t)$ added with dividends $D(t)$ under the real basis. The reserve is given by

$$\begin{aligned}V(t) &= \mathbf{E}\Big[\int_t^T \exp\Big(-\int_t^u r(v)\,dv\Big)\,d(B+D)(u)\Big|\mathcal{F}(t)\Big]\\ &= \mathbf{E}\Big[\int_t^T \exp\Big(-\int_t^u r(v)\,dv\Big)\,d(B+D)(u)\Big|Z(t), X(t)\Big]\\ &= V^{Z(t)}\big(t, X(t)\big)\end{aligned}$$

where we have used that $(Z(t), X(t))$ is Markov (which we take as a fact) and the state-wise reserves are given by

$$V^j(t,x) = \mathbf{E}\Big[\int_t^T \exp\Big(-\int_t^u r(v)\,dv\Big)\,d(B+D)(u)\Big|Z(t)=j, X(t)=x\Big].$$

We will derive differential equation that characterizes the state-wise reserves $V^j(t,x)$ using the martingale methods in Section 14. In this case including the dividend process equation (14.8) reads as

$$\lambda^B(t+) + \lambda^D(t+) + \lambda^V(t+) - r(t)V(t) = 0. \tag{18.2}$$

As in previous sections we have that payment process $B(t)$ has intensity process given by

$$\lambda^B(t) = b^{Z(t-)}(t) + \sum_k b^{Z(t-)k}(t)\mu^{Z(t-)k}(t)1_{\{Z(t-)\neq k\}}$$

and by same method the dividend process $D(t)$ has intensity process given by

$$\lambda^D(t) = \bar{\delta}^{Z(t-)}\big(t, X(t-)\big) + \sum_k \bar{\delta}^{Z(t-)k}\big(t, X(t-)\big)\mu^{Z(t-)k}(t)1_{\{Z(t-)\neq k\}}.$$

Next step is to compute the intensity process of the reserve. From the dynamics

$$dX(t) = r(t)X(t)\,dt + d(C-D)(t)$$

we have the jump process of $X(t)$ is

$$\Delta X(t) = \Delta C(t) - \Delta D(t) = \sum_k \Big(c^{Z(t-)k}(t) - \bar{\delta}^{Z(t-)k}\big(t, X(t-)\big)\Big)\Delta N^k(t)$$

and hence

$$X(t) = X(t-) + \sum_k \Big(c^{Z(t-)k}(t) - \bar{\delta}^{Z(t-)k}\big(t, X(t-)\big)\Big)\Delta N^k(t).$$

We can apply the changes of variables formula (Corollary 4.5) if assume that $V_t^j(t,x)$ and $V_x^j(t,x)$ exist and are continuous and then we get that

$$\begin{aligned}
&dV(t)\\
&= d\big(V^{Z(t)}\big(t, X(t)\big)\big)\\
&= V_t^{Z(t)}\big(t, X(t)\big)dt + V_x^{Z(t)}\big(t, X(t)\big)\, dX^c(t) + \sum_k \Big(V^{Z(t)}\big(t, X(t)\big) - V^{Z(t-)}\big(t-, X(t-)\big)\Big)dN^k(t)\\
&= V_t^{Z(t)}\big(t, X(t)\big)\, dt + V_x^{Z(t)}\big(t, X(t)\big)\Big(r(t)X(t) + c^{Z(t)}(t) - \bar\delta^{Z(t)}\big(t, X(t)\big)\Big)\, dt\\
&\quad + \sum_k \Big(V^{Z(t)}\big(t, X(t)\big) - V^{Z(t-)}\big(t, X(t-)\big)\Big)\, dN^k(t)\\
&= V_t^{Z(t)}\big(t, X(t)\big)\, dt + V_x^{Z(t)}\big(t, X(t)\big)\Big(r(t)X(t) + c^{Z(t)}(t) - \bar\delta^{Z(t)}\big(t, X(t)\big)\Big)\, dt\\
&\quad + \sum_k \Big(V^k\big(t, X(t-) + c^{Z(t-)k}(t) - \bar\delta^{Z(t-)k}\big(t, X(t-)\big)\big) - V^{Z(t-)}\big(t, X(t-)\big)\Big)dN^k(t).
\end{aligned}$$

From this expression we see that the intensity process $\lambda^V(t)$ of the reserve $V(t)$ is

$$\begin{aligned}
\lambda^V(t) &= V_t^{Z(t-)}(t, X(t-)) + \Big(r(t)X(t-) + c^{Z(t-)}(t) - \bar\delta^{Z(t-)}\big(t, X(t-)\big)\Big)V_x^{Z(t-)}\big(t, X(t-)\big)\\
&\quad + \sum_k \Big(V^k\big(t, Y(t-)\big) - V^{Z(t-)}\big(t, X(t-)\big)\Big)\mu^{Z(t-)k}(t)1_{\{Z(t-)\neq k\}}
\end{aligned}$$

where we have set $Y(s) = X(s) + c^{Z(s)k}(s) - \bar\delta^{Z(s)k}(s, X(s))$. Substituting the intensity process into equation (18.2) we get that

$$\begin{aligned}
&V_t^{Z(t)}\big(t, X(t)\big) + \Big(r(t)X(t) + c^{Z(t)}(t) - \bar\delta^{Z(t)}\big(t, X(t)\big)\Big)V_x^{Z(t)}\big(t, X(t)\big) + b^{Z(t)}(t) + \bar\delta^{Z(t)}\big(t, X(t)\big)\\
&\quad = r(t)V(t) - \sum_k\Big(b^{Z(t)k}(t) + \bar\delta^{Z(t)k}\big(t, X(t)\big) + V^k\big(t, Y(t)\big) - V^{Z(t)}\big(t, X(t)\big)\Big)\mu^{Z(t)k}(t)1_{\{Z(t)\neq k\}}.
\end{aligned}$$

Finally, Condition on $Z(t) = j$ and $X(t) = x$ in the above expression we have derived the following theorem.

Theorem 18.1. *Assume that the state-wise reserves $V^j(t,x)$, $j = 1, 2, \dots, m$, are continuous differentiable (in boths variables t and x). Then the state-wise reserves are solutions to the partial differential equations*

$$\begin{aligned}
&V_t^j(t,x) + \Big(r(t)x + c^j(t) - \bar\delta^j(t,x)\Big)V_x^j(t,x)\\
&\quad = r(t)V^j(t,x) - b^j(t) - \bar\delta^j(t,x)\\
&\qquad - \sum_{k:k\neq j}\mu^{jk}(t)\Big(b^{jk}(t) + \bar\delta^{jk}(t,x) + V^k\big(t, x + c^{jk}(t) - \bar\delta^{jk}(t,x)\big) - V^j(t,x)\Big)
\end{aligned}$$

with boundary condition $V^j(T,x) = 0$ for $x \geq 0$ and $j = 1, \dots, m$.

18.6. Dividends linear in surplus. The dividend payments are linear functions of the surplus, that is,

$$\delta^j(t) = q^j(t)X(t)$$
$$\delta^{jk}(t) = q^{jk}(t)X(t-)$$

where $q^j(t)$ and $0 \le q^{jk}(t) < 1$ are continuous deterministic functions. Dividends linear in the surplus have applications to pension funding. In this case we find new stochastic representation for the solution the partial differential equations in Theorem 18.1.

With dividends linear in the surplus as above the partial differential equations in Theorem 18.1 read as

$$\begin{aligned}
&V^j_t(t,x) + \Big(r(t)x + c^j(t) - q^j(t)x\Big)V^j_x(t,x) \\
&\quad = r(t)V^j(t,x) - b^j(t) - q^j(t)x \\
&\qquad - \sum_{k:k\neq j} \mu^{jk}(t)\Big(b^{jk}(t) + q^{jk}(t)x + V^k\big(t, x + c^{jk}(t) - q^{jk}(t)x\big) - V^j(t,x)\Big).
\end{aligned}$$

We guess that the solution to the above partial differential equations are on the form $V^j(t,x) = f^j(t) + g^j(t)x$. Plug this expression into the partial differential equations we derive ordinary differential equations for $f^j(t)$ and $g^j(t)$.

The differential equation for $g^j(t)$ is

$$g^j_t(t) = \Big(q^j(t) + \sum_{k:k\neq j} \mu^{jk}(t)q^{jk}(t)\Big)g^j(t) - q^j(t) - \sum_{k:k\neq j} \big(1-q^{jk}(t)\big)\mu^{jk}(t)\Big(\frac{q^{jk}(t)}{1-q^{jk}(t)} + g^k(t) - g^j(t)\Big).$$

We want a stochastic representation for the solution. The intensity appears in the differential equation with a factor $1-q^{jk}(t)$. Thus, $\mathbf{E}^q$ denotes the expectation with respect to the probability measure under which $N^k(t)$ has intensity process $\big(1 - q^{Z(t-)k}(t)\big)\mu^{Z(t-)k}(t)1_{\{Z(t-)\neq k\}}$. Then $g^j(t)$ has the stochastic representation

$$g^j(t) = \mathbf{E}^q\Big[\int_t^T \exp\Big(-\int_t^u \big(q^j(v) + \sum_{k:k\neq j} \mu^{jk}(v)q^{jk}(v)\big)\,dv\Big)\,dG(u)\Big|Z(t) = j\Big]$$

where

$$dG(t) = q^{Z(t)}(t)\,dt + \sum_k \frac{q^{Z(t-)k}(t)}{1-q^{Z(t-)k}(t)}\,dN^k(t).$$

The differential equation for $f^j(t)$ is

$$f^j_t(t) = r(t)f^j(t) - b^j(t) - g^j(t)c^j(t) - \sum_{k:k\neq j} \mu^{jk}(t)\Big(b^{jk}(t) + c^{jk}(t)g^k(t) + f^k(t) - f^j(t)\Big).$$

Then $f^j(t)$ has the stochastic representation

$$f^j(t) = \mathbf{E}\Big[\int_t^T \exp\Big(-\int_t^u r(v)\,dv\Big)\,d(B+F)(u)\Big|Z(t) = j\Big]$$

where

$$dF(t) = g^{Z(t)}(t)c^{Z(t)}(t)\,dt + \sum_k c^{Z(t-)k}(t)g^k(t)\,dN^k(t).$$

19. Exercises

Exercise 19.1. consider the simplest example of the multi-state contract with two states and one possible transition. The two states of the policy are: "1"=Alive and "2"=Dead. Let $Z(t)$ denote the state of the policy holder at time t. The contract is issued at time 0 and expires at time T. We assume that policy commencing in state 1, that is $Z(0) = 1$. Assume that $Z(t)$ is a continuous time Markov chain with transition intensity $\mu^{12}(t) = \mu(t)$ where $\mu(t) \geq 0$ is a Borel function bounded on bounded intervals. Assume that the interest rate $r(t)$ is deterministic. The life length of the policy holder is the jump time τ_x and the survival probability is

$$\begin{aligned} \mathbf{P}(\tau_x > t | Z(s) = 1) &= \mathbf{P}(Z(t) = 1 | Z(s) = 1) \\ &= p^{11}(s, t) \\ &= \exp\Big(- \int_s^t \mu(v)\, dv \Big). \end{aligned}$$

Let $N(t)$ count the number of deaths of the policy holder, that is, $N(t) = 1_{\{Z(t)=1\}}$.

$$\boxed{\text{Alive : 1}} \quad \to \quad \boxed{\text{Dead : 2}}$$

The payment process is given by $B(t) = 1_{[T,\infty)}(t) 1_{\{Z(t)=1\}}$ (pure endowment).

(a) Give an interpretation of the payment process.
(b) Compute the present value $Y(t, T]$.
(c) Compute the reserve $V(t)$.
(d) Argue that $B(t)$ does not have an intensity process.

Exercise 19.2. Use Kolmogorov backward equations (Theorem 12.7) and equation (16.1) to derive Thiele integral equations in Theorem 16.1.

Exercise 19.3. Let the model be as Exercise 19.1 with two states and mortality intensity $\mu(t)$. The payment process is given by $\Delta B^1(t) = 1_{\{1,2,\dots,n\}}(t) 1_{\{Z(t)=1\}}$.

(a) Give an interpretation of the payment process.
(b) Compute the differential equation for $V^1(t)$.

Exercise 19.4. Consider the model in Section 14.4. Thus $Z(t)$ state of policy at time t and is Markov with continuous transition intensities $\mu^{jk}(t)$. Recall the counting process $N^k(t) = \#\{0 < s \leq t | Z(s-) \neq k, Z(s) = k\}$ and set

$$N^\bullet(t) = \sum_k N^k(t)$$

(a) What does $N^\bullet(t)$ counts?

Let the payment process be given by

$$dB(t) = b^{Z(t)}(t, N^\bullet(t))dt + \sum_k b^{Z(t-)k}(t, N^\bullet(t-))\, dN^k(t)$$

and then state-wise reserves are given by

$$V^j(t, n) = \mathbf{E}\Big[\int_t^T \exp\Big(\int_t^u r(v) dv \Big)\, dB(u) \Big| Z(t) = j, N^\bullet(t) = n \Big].$$

(b) Use the martingale method to compute the Thiele differential equations for the state-wise reserves.

CHAPTER IV

Introduction to survival and event history analysis

This chapter introduces some basic concepts and ideas of survival and event history analysis. The observations of moving among a finite number of states are assumed to happen in continuous time and often there are incomplete observations. The statistical models are based on counting processes.

20. Hazard rate (force of mortality)

The hazard rate is used to model survival times. The results below are a reminder to hazard rate and basic properties.

Let the survival time $\tau > 0$ be a positive random variable with distribution function $F(t) = \mathbf{P}(\tau \leq t)$ and survival function $\bar{F}(t) = \mathbf{P}(\tau > t)$ for $t \geq 0$. Suppose that τ is absolutely continuous with density $f(\cdot)$. This particularly implies that $F(t)$ is continuous. Assume also that $F(t)$ has the right-hand derivative $f(t)$, $t > 0$. Condition on that $\tau \geq t$, the probability that τ takes values in $[t, t + \Delta t)$ is given by

$$\begin{aligned}\mathbf{P}(t \leq \tau < t + \Delta t | \tau \geq t) &= 1_{\{\bar{F}(t-)>0\}} \frac{F(t + \Delta t) - F(t-)}{\bar{F}(t-)} \\ &= 1_{\{\bar{F}(t)>0\}} \frac{F(t + \Delta t) - F(t)}{\bar{F}(t)}\end{aligned}$$

(with the convention $0/0 = 0$). The hazard rate (force of mortality) of τ is

$$\mu(t) = \lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} \mathbf{P}(t \leq \tau < t + \Delta t | \tau \geq t) = 1_{\{\bar{F}(t)>0\}} f(t)/\bar{F}(t)$$

and for t satisfying $\bar{F}(t) > 0$ we write that as

$$\begin{aligned}\mu(t)\, dt &= \mathbf{P}(t \leq \tau < t + dt | \tau \geq t) \\ &= \frac{f(t)}{\bar{F}(t)}\, dt \\ &= -d\big(\log\big(\bar{F}(t)\big)\big).\end{aligned}$$

This is the probability of "dying" in the immediately future conditional on "survival" until time t. Integrating the above equation leads to

$$\bar{F}(t) = \exp\Big(- \int_0^t \mu(s)\, ds \Big)$$

and

$$f(t) = \alpha(t) \exp\Big(- \int_0^t \mu(s)\, ds \Big).$$

Hence, the hazard rate $\mu(t)$ determines the survival function $\bar{F}(t)$ and the density $f(t)$.

Example 20.1. Here are a few distributions used to model survival times.

1. Exponential distribution with parameter $\lambda > 0$: The density and hazard rate are
$$f(t;\lambda) = \lambda e^{-\lambda t} \text{ and } \mu(t;\lambda) = \lambda \text{ for } t > 0.$$
2. Gamma distribution with parameters $\beta > 0$ and $\lambda > 0$. The density and hazard rate are
$$f(t;\beta,\lambda) = \frac{\lambda^\beta}{\Gamma(\beta)} t^{\beta-1} e^{-\lambda t} \text{ and } \mu(t;\beta,\lambda) = \frac{\lambda^\beta}{\Gamma(\beta,\lambda t)} t^{\beta-1} e^{-\lambda t}$$
for $t > 0$ and $\Gamma(\beta, x) = \int_x^\infty y^{\beta-1} e^{-y}\, dy$ is the incomplete gamma function.
3. Gompertz-Makeham distribution with parameters $\alpha > 0$, $\beta > 0$ and $\gamma \in \mathbb{R}$. The density and hazard rate are
$$f(t;\alpha,\beta,\gamma) = (\alpha + \beta e^{\gamma t}) \exp\big(-\alpha t + \tfrac{\beta}{\gamma}(1 - e^{\gamma t})\big) \text{ and } \mu(t;\alpha,\beta,\gamma) = \alpha + \beta e^{\gamma t} \text{ for } t > 0.$$

21. Examples of basic counting processes models

In this section we introduce the counting processes associated to two concrete models—the survival model and the Markov model (multi-state model).

Incomplete observation can rarely be avoided in the study of life and event history data. The point is that one has to wait for the event to happen and when the study ends one will find the event has occurred for some individuals but not for others. The observations come as a mixture of complete and incomplete observations. In this note the form of incomplete observations is right-censoring at a given deterministic time, that is, the censoring cut off the time interval to the right-hand side at a given time. The concept is called right-censoring.

Example 21.1. (One uncensored survival time) The data in survival analysis measures the time to an event (e.g. death) and is denoted the survival time. The survival time τ is modeled as a positive random variable. Let τ be a survival time with hazard rate $\mu(t)$. We can associate a Markov chain to the survival time τ. Indeed, let the state 1 correspond to "alive" and let the (absorbing) state 2 correspond to "dead". Let $Z(t)$ be a Markov chain with state space $\mathcal{S} = \{1, 2\}$ and with transition intensity $\mu(t)$ from state 1 to state 2 (there is at most one transition) and let $Z(0) = 1$. The jump time is the survival time. From Proposition 12.9 we have that
$$\mathbf{P}(\tau > t) = \mathbf{P}(\tau > t | Z(0) = 1) = \exp\Big(-\int_0^t \mu(s)\, ds\Big).$$
We see that the transition intensity $\mu(t)$ is the hazard rate for the survival time. To the survival time, or equivalently, to the Markov chain we associate the counting process (one jump process)
$$N(t) = 1_{[\tau,\infty)}(t) = \#\{0 < s \le t | Z(s-) = 1, Z(s) = 2\}$$
that counts the number of deaths. The counting process $N(t)$ has intensity process given by (see Proposition 12.5)
$$\lambda(t) = \mu(t)\, 1_{[0,\tau]}(t) = \mu(t)\, 1_{\{Z(t-)=0\}}.$$
The survival data is represented by the counting process $N(t)$ registering the occurrence of the event of death (transition) and the predictable process $1_{[0,\tau]}(t)$ registering whether the individual under observation is alive just before time t.

Example 21.2. (One censured survival time) Let τ be the above (actual) survival time from Example 21.1 and let c be a given censoring time. The survival time is observed if it is smaller than the censoring time c and else we only know that τ exceeds c. We only observe the censored survival time $\tilde{\tau} = \tau \wedge c$ together with the censoring indicator $D = 1_{\{\tilde{\tau}=\tau\}}$ which is 1 if we observe the actual survival time and 0 if we observe the censoring time c. In other words, the Markov chain $Z(t)$ is observed continuously from time 0 to the time of censoring c. The censored counting process that counts the number of observed (actual) deaths is

$$\widetilde{N}(t) = 1_{[\tilde{\tau},\infty)}(t)1_{\{D=1\}} = \int_0^t 1_{[0,c]}(s)\, dN(s)$$

where $1_{[0,c]}(t)$ is the censoring process. The script $\tilde{}$ is to indicate that observations have been censored (the counting process is censored and is marked $\tilde{}$). By the intensity process of $N(t)$ and Proposition 10.6, the counting process $\widetilde{N}(t)$ has intensity process

$$\begin{aligned}\tilde{\lambda}(t) &= 1_{[0,c]}(t)\lambda(t)\\ &= \mu(t)\, 1_{\{Z(t-)=0\}}1_{[0,c]}(t)\\ &= \mu(t)1_{[0,\tau]}(t)1_{[0,c]}(t)\\ &= \mu(t)1_{[0,\tilde{\tau}]}(t).\end{aligned}$$

In this case the censored observation is represented by the counting processes $N(t)$ and the predictable process $1_{[0,c]}(t)1_{[0,\tau]}(t) = 1_{[0,\tilde{\tau}]}(t)$ (we do not observe the censoring process after death). Note that with the absorbing state 2 there is no more information to gain from observing the process after transition to that state.

Example 21.3. (Markov model) This is a generalization of the survival model. Let $Z(t)$ be a continuous time Markov chain with finite state space $\mathcal{S} = \{1, 2, \ldots, m\}$ and $Z(0) = 1$. We assume that the Markov chain has transition intensities denoted by $\mu^{jk}(t)$ for $j, k \in \mathcal{S}$, $j \neq k$. Let $N^{jk}(t)$ count the number of transitions from state j to state k, $j \neq k$, in $[0, t]$, that is,

$$N^{jk}(t) = \#\{0 < s \leq t | Z(s-) = j, Z(s) = k\}.$$

The intensity process of $N^{jk}(t)$ is given by (see Proposition 12.5) $\lambda^{jk}(t) = \mu^{jk}(t)\, 1_{\{Z(t-)=j\}}$. Then the multivariate counting process $N(t) = (N^{jk}(t), j, k \in \mathcal{S}, j \neq k)$ is equivalent to the Markov chain $Z(t)$ in the sense that observation of $Z(s)$, $0 \leq s \leq t$, gives the same data as observing the process $N(s)$ for $s \in [0, t]$.

Let c be a given censoring time. Then $Z(t)$ is observed until the censoring time c. Again, there is no information to gain from observing the process after transition to an absorbing state ("dead"). When an absorbing state is not reached, the end of the observation is the censoring time c. In this case the observed number of transitions is

$$\widetilde{N}^{jk}(t) = \int_0^t 1_{[0,c]}(s)\, dN^{jk}(s)$$

where $1_{[0,c]}(t)$ is the censoring process. Again, by Proposition 12.5, the intensity process of $\widetilde{N}^{jk}(t)$ is given by

$$\tilde{\lambda}^{jk}(t) = 1_{[0,c]}(t)\, \lambda^{jk}(t) = \mu^{jk}(t)\, 1_{[0,c]}(t)1_{\{Z(t-)=j\}}.$$

In this case the censored data is represented by the counting processes $\widetilde{N}^{jk}(t)$ and the predictable process $1_{[0,c]}(t)1_{\{Z(t-)=j\}}$ which contains information on whether the Markov process is at risk to make the transition from state j to state k.

22. Multiplicative intensity model

In the following sections we study statistical inference for general counting processes. The general statistical model for counting processes in this note is the multiplicative intensity model. Recall that given an intensity process then (under regularity conditions) there is a unique counting process with the given intensity process. In other words, the intensity process determines the counting process and we only need to model the intensity process.

Let $T > 0$ be the upper time limit for the observations/study. Let

$$(N(t), 0 \leq t \leq T) = ((N^1(t), \dots, N^m(t)), 0 \leq t \leq T)$$

be an m-variate counting process with intensity process given by

$$\lambda^k(t) = \alpha^k(t)\, Y^k(t), \quad k = 1, \dots, m, \quad t \in [0, T]. \tag{22.1}$$

The assumptions on the given $\alpha^k(t)$ and $Y^k(t)$ are for $k = 1, \dots, m$:

(i) $\alpha^k(t)$ is a positive deterministic function which is integrable on $[0, T]$, i.e. $\int_0^T \alpha^k(t)\, dt < \infty$.
(ii) $Y^k(t)$ is a positive predictable process and $1_{\{Y^k(t)>0\}}/Y^k(t)$ is locally bounded (with the convention $0/0 = 0$).

Recall that $1_{\{Y^k(t)>0\}}/Y^k(t)$ is locally bounded if $\sup_{0 \leq s \leq t} |1_{\{Y^k(s,\omega)>0\}}/Y^k(s, \omega)| < \infty$ for all t and ω. The process $Y^k(t)$ is observable in the sense that it does not change with parameterization of the model. The observations available at time t are described as $((N(s), Y(s)), 0 \leq s \leq t)$ where $Y(t) = (Y^1(t), \dots, Y^m(t))$. This class of statistical models is called the multiplicative intensity model for counting processes. The interpretation is that $\alpha^k(t)$ is the force of transition of type k and the observable process $Y^k(t)$ contains the information on the number at risks of making a transition of type k at time t, and if $Y^k(t) = 0$ there are no individuals to observe of making the transition. Below are examples of the multiplicative intensity model.

Example 22.1. (Uncensored survival times, complete data) Let $\tau_1, \dots, \tau_n$ be independent survival times of n individuals where τ_i has hazard rate $\mu_i(t)$. For each $i = 1, \dots, n$ define the counting process $N_i(t) = 1_{[\tau_i,\infty)}(t)$. From Example 21.1, $N_i(t)$ has intensity process $\lambda_i(t) = \mu_i(t)\, 1_{[0,\tau_i]}(t) = \mu_i(t)\, Y_i(t)$ where $Y_i(t) = 1_{[0,\tau_i]}(t) = 1 - N_i(t-)$ is CAGLAD and therefore is a predictable process. Thus, the n-variate counting process $N(t) = (N_1(t), \dots, N_n(t))$ satisfies the multiplicative intensity model. The process $Y_i(t) = 1$ if individual i is at risk for experiencing the event (alive) just before time t, otherwise $Y_i(t) = 0$ (we do not observe after death).

Consider also the case that the survival times also are independent and identical distributed (IID) with common hazard rate $\mu(t)$. Then the aggregated counting process

$$N_{(n)}(t) = \sum_{i=1}^n N_i(t) = \sum_{i=1}^n 1_{[\tau_i,\infty)}(t)$$

has intensity process

$$\lambda_{(n)}(t) = \sum_{i=1}^n \lambda_i(t) = \mu(t) \sum_{i=1}^n 1_{[0,\tau_i]}(t).$$

The process $Y_{(n)}(t) = \sum_{i=1}^n Y_i(t) = \sum_{i=1}^n 1_{[0,\tau_i]}(t) = n - N_{(n)}(t-)$ is predictable. The aggregated process $N_{(n)}(t)$ also satisfies the multiplicative intensity model. The process $Y_{(n)}(t)$ is the number of individuals at risk just before time t.

Example 22.2. (Censored survival times, incomplete data) Let $\tau_1, \tau_2, \ldots, \tau_n$ be IID survival times with given hazard rate $\mu(t)$. Let $c_1, c_2, \ldots, c_n$ be given deterministic censoring times. The observations are

- $\tilde{\tau}_i = \tau_i \wedge c_i$, the censored survival times.
- $D_i = 1_{\{\tilde{\tau}_i = \tau_i\}}$, the censoring indicators.

The censored counting processes that counts the actual number of deaths are

$$\widetilde{N}_i(t) = 1_{[\tilde{\tau}_i, \infty)}(t) 1_{\{D_i = 1\}} = \int_0^t 1_{[0,c_i]}(s)\, dN_i(s)$$

and its intensity process is given by $\tilde{\lambda}_i(t) = \mu(t) Y_i(t) 1_{[0,c_i]}(t) = \mu(t)\, \widetilde{Y}_i(t)$ where $\widetilde{Y}_i(t) = Y_i(t) 1_{[0,c_i]}(t) = 1_{[0,\tau_i]}(t) 1_{[0,c_i]}(t) = 1_{[0,\tilde{\tau}_i]}(t)$ is a CAGLAD process and hence predictable. Hence, the n-variate counting process $\widetilde{N}(t) = (\widetilde{N}_1(t), \ldots, \widetilde{N}_n(t))$ satisfies the multiplicative intensity model. The process $\widetilde{Y}_i(t) = 1$ if individual i is at risk for experiencing the event (we know that individual i is alive) just before time t, otherwise $\widetilde{Y}_i(t) = 0$ (we do not observe the censoring after death). Also in this case, the aggregated counting process $\widetilde{N}_{(n)}(t) = \sum_{i=1}^n \widetilde{N}_i(t)$ has a multiplicative intensity process given by

$$\tilde{\lambda}_{(n)}(t) = \mu(t) \sum_{i=1}^n 1_{[0,\tilde{\tau}_i]}(t) = \mu(t) \sum_{i=1}^n \widetilde{Y}_i(t) = \mu(t) \widetilde{Y}_{(n)}(t).$$

The process $\widetilde{Y}_\bullet(t)$ is the number of individuals that we know that are alive just before time t.

Example 22.3. (Markov model) Let the Markov chains $Z_1(t), \ldots, Z_n(t)$ be independent replicates of the Markov chain described in Example 21.3. The transition intensities are $\mu_{jk}(t)$. For individual i, the counting process that counts the number of transitions from state j to state k is denoted

$$N_i^{jk}(t) = \#\{0 < s \leq t | Z_i(s-) = j, Z_i(s) = k\}$$

and from Example 21.3 its intensity process is $\lambda_i^{jk}(t) = \mu^{jk}(t)\, 1_{\{Z_i(t-) = j\}} = \mu^{jk}(t)\, Y_i^j(t)$. The process $Y_i^j(t) = 1_{\{Z_i(t-) = j\}}$ is CAGLAD and therefore is predictable process. The multivariate counting process $N(t) = (N_i^{jk}(t), i = 1, \ldots, n,\, j, k \in \mathcal{S}, j \neq k)$ has a multiplicative intensity process. For individual i, the process $Y_i^{jk}(t)$ is the number at risk of making the transition from state j to state k just before time t.

Again, let $c_1, c_2, \ldots, c_n$ be given censoring times and the process $Z_i(t)$ is observed from time 0 to the time of censoring c_i. The censored counting processes are given by (see Example 21.3)

$$\widetilde{N}_i^{jk}(t) = \int_0^t 1_{[0,c_i]}(s)\, dN_i^{jk}(s)$$

with the intensity processes $\tilde{\lambda}_i^{jk}(t) = 1_{[0,c_i]}(t)\, \lambda_i^{jk}(t) = \mu^{jk}(t)\, 1_{[0,c_i]}(t) 1_{\{Z_i(t-) = j\}} = \mu^{jk}(t)\, \widetilde{Y}_i^j(t)$ where $\widetilde{Y}_i^j(t) = 1_{[0,c_i]}(t) Y_i^j(t) = 1_{[0,c_i]}(t) 1_{\{Z_i(t-) = j\}}$ is a predictable process. Again, the censored multivariate counting process $\widetilde{N}(t) = (\widetilde{N}_i^{jk}(t), i = 1, \ldots, n, j, k \in \mathcal{S}, j \neq k)$ has a multiplicative intensity process. As in the above examples the aggregated counting processes $\widetilde{N}_{(n)}^{jk}(t) = \sum_{i=1}^n \widetilde{N}_i^{jk}(t)$ has a multiplicative intensity process given by

$$\tilde{\lambda}_{(n)}^{jk}(t) = \sum_{i=1}^n \tilde{\lambda}_i^{jk}(t) = \mu^{jk}(t) \sum_{i=1}^n \widetilde{Y}_i^j(t) = \mu^{jk}(t) \widetilde{Y}_{(n)}^j(t).$$

23. Nonparametric models

In the present section, we will study nonparametric estimation of the multiplicative intensity model. One aim is to estimate the integrated "force of transition". An advantage of the nonparametric model is that it can handle any distribution, but the estimator is difficult to describe. Whereas a parametric model can be described by the values of the parameters. A disadvantage of the nonparametric method is that the "force of transition" cannot be directly estimated (it is not defined since the estimator is a piecewise constant function). The analogous to ordinary statistics with complete observations is the simplest nonparametric estimate of a distribution function: the empirical distribution (the distribution is equal to the observed distribution) where the true continuous distribution function is estimated by a discrete distribution function and the density cannot be estimated directly.

The set-up is the following. Let $T > 0$ be the upper time limit for the observations. Let the multivariate counting process $N(t) = (N^1(t), \dots, N^m(t))$ satisfy the multiplicative intensity model: the intensity process $\lambda(t) = (\lambda^1(t), \dots, \lambda^m(t))$ is given by

$$\lambda^k(t) = \alpha^k(t) Y^k(t), \qquad k = 1, \dots, m \text{ and } t \in [0, T]$$

where

(i) $\alpha^k(t)$ is a positive integrable deterministic function on $[0, T]$, that is $\int_0^T \alpha^k(t)\, dt < \infty$.
(ii) $Y^k(t)$ is a positive predictable process and $1_{\{Y^k(t)>0\}}/Y^k(t)$ is locally bounded.

Define the integrated function

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

and recall that $M^k(t) = N^k(t) - \int_0^t \alpha^k(s) Y^k(s)\, ds$ is a martingale.

23.1. Nelson-Aalen estimator. We will heuristically derive the Nelson-Aalen estimator for $A^k(t)$. The martingale $M^k(t)$ has dynamics

$$dM^k(t) = dN^k(t) - \alpha^k(t) Y^k(t)\, dt \tag{23.1}$$

that may be considered as "noise". Dividing by $Y^k(t)$ we have that

$$\alpha^k(t)\, dt = \frac{1}{Y^k(t)}\, dN^k(t) - \frac{1}{Y^k(t)}\, dM^k(t)$$

where the latter term (a martingale) again can be considered as "noise". Informally, we have that

$$\int_0^t \alpha^k(s)\, ds \approx \int_0^t \frac{1}{Y^k(s)}\, dN^k(s).$$

Since $Y^k(t)$ might be zero we introduce the indicator process $J^k(t) = 1_{\{Y^k(t)>0\}}$. With the convention $0/0 = 0$, the Nelson-Aalen estimator is given by

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

and it is an estimator of $A^k(t)$. The Nelson-Aalen estimator $\widehat{A}^k(t)$ is an increasing piecewise constant function with increments $1/Y^k(\tau_j^k)$ at a jump time τ_j^k of $N^k(t)$.

Next we will derive some basic (small sample) properties of the Nelson-Aalen estimator. Multiplying (23.1) by $J^k(t)/Y^k(t)$ yields

$$\frac{J^k(t)}{Y^k(t)}\,dM^k(t) = \frac{J^k(t)}{Y^k(t)}\,dN^k(t) - \alpha^k(t)J^k(t)\,dt$$

and on integral form the equation is

$$\int_0^t \frac{J^k(s)}{Y^k(s)}\,dM^k(s) = \int_0^t \frac{J^k(s)}{Y^k(s)}\,dN^k(s) - \int_0^t \alpha^k(s)J^k(s)\,ds = \widehat{A}^k(t) - A^{*k}(t)$$

where

$$A^{*k}(t) = \int_0^t \alpha^k(s)J^k(s)\,ds.$$

By Proposition 10.5

$$\widehat{A}^k(t) - A^{*k}(t) = \int_0^t \frac{J^k(s)}{Y^k(s)}\,dM^k(s) \tag{23.2}$$

is a martingale. The process $A^{*k}(t)$ is the predictable compensator of the Nelson-Aalen estimator $\widehat{A}^k(t)$. Note that $A^{*k}(t)$ is almost equal to $A^k(t)$ when there is a small probability for $Y^k(s) = 0$, $s \in [0,t]$. Since $\widehat{A}^k(t) - A^{*k}(t)$ is a martingale we have that

$$\begin{aligned}\mathbf{E}[\widehat{A}^k(t)] = \mathbf{E}[A^{*k}(t)] &= \int_0^t \alpha^k(s)\mathbf{E}[J^k(s)]\,ds = \int_0^t \alpha^k(s)\mathbf{P}(Y^k(s) > 0)\,ds\\ &= A^k(t) - \int_0^t \alpha^k(s)\mathbf{P}(Y^k(s) = 0)\,ds \le A^k(t).\end{aligned}$$

The Nelson-Aalen estimator is in general biased (downward). As written above, if there is a small probability for $Y^k(s) = 0$, $s \in [0,t]$, the bias is insignificant. It is not possible to find an unbiased nonparametric estimator of $A^k(t)$ since $\alpha^k(t)$ cannot be estimated when $Y^k(t) = 0$.

From Proposition 11.12, we have that the martingales $M^k(t)$, $k = 1, \ldots, m$ have predictable covariation processes given by

$$\langle M^j, M^k\rangle(t) = \begin{cases}\displaystyle\int_0^t \alpha^k(s)Y^k(s)\,ds & \text{for } j = k\\ 0 & \text{for } j \neq k.\end{cases}$$

By Proposition 10.7 and Proposition 10.8 we get that

$$\langle \widehat{A}^j - A^{*j}, \widehat{A}^k - A^{*k}\rangle(t) = \delta_{jk}\int_0^t \left(\frac{J^k(s)}{Y^k(s)}\right)^2 d\langle M^k\rangle(s) = \delta_{jk}\int_0^t \alpha^k(s)\frac{J^k(s)}{Y^k(s)}\,ds \tag{23.3}$$

where δ_{jk} is a Kronecker delta ($\delta_{jk} = 1$ for $j = k$ and $\delta_{jk} = 0$ for $j \neq k$). For $0 < s \le t$ and $j \neq k$ we have that

$$\begin{aligned}\mathbf{E}[(\widehat{A}^j(s) - A^{*j}(s))(\widehat{A}^k(t) - A^{*k}(t))] &= \mathbf{E}\big[\big(\widehat{A}^j(s) - A^{*j}(s)\big)\mathbf{E}[\widehat{A}^k(t) - A^{*k}(t)|\mathcal{F}(s)]\big]\\ &= \mathbf{E}\big[\big(\widehat{A}^j(s) - A^{*j}(s)\big)\big(\widehat{A}^k(s) - A^{*k}(s)\big)\big]\\ &= \mathbf{E}[\langle \widehat{A}^j - A^{*j}, \widehat{A}^k - A^{*k}\rangle(s)] = 0.\end{aligned}$$

It follows that $\widehat{A}^j(s) - A^{*j}(s)$ and $\widehat{A}^k(t) - A^{*k}(t)$ are uncorrelated for any $0 \leq$ and $j \neq k$ and plots of Nelson-Aalen estimators may be studied independently of each other.

The mean squared error function of $\widehat{A}^k(t)$ is defined as

$$\tilde{\sigma}_k^2(t) = \mathbf{E}[(\widehat{A}^k(t) - A^{*k}(t))^2].$$

Since $\langle \widehat{A}^k - A^{*k}\rangle(t)$ is the predictable compensator of $(\widehat{A}^k(t) - A^{*k}(t))^2$ we get that

$$\begin{aligned}\tilde{\sigma}_k^2(t) &= \mathbf{E}[\langle \widehat{A}^k - A^{*k}\rangle(t)] \\ &= \int_0^t \alpha^k(s)\mathbf{E}\Big[\frac{J^k(s)}{Y^k(s)}\Big]\, ds.\end{aligned}$$

To find an estimator for $\tilde{\sigma}_k^2(t)$ note that

$$\langle \widehat{A}^k - A^{*k}\rangle(t) = \int_0^t \frac{J^k(s)}{Y^k(s)}\, dA^k(s)$$

and as an estimator we use

$$\hat{\sigma}_k^2(t) = \int_0^t \frac{J^k(s)}{Y^k(s)}\, d\widehat{A}^k(s) = \int_0^t \frac{J^k(s)}{Y^k(s)^2}\, dN^k(s).$$

The process

$$\begin{aligned}\int_0^t \frac{J^k(s)}{Y^k(s)^2}\, dM^k(s) &= \int_0^t \frac{J^k(s)}{Y^k(s)^2}\, dN^k(s) - \int_0^t \frac{J^k(s)}{Y^k(s)^2}\alpha^k(s)Y^k(s)\, ds \\ &= \int_0^t \frac{J^k(s)}{Y^k(s)^2}\, dN^k(s) - \int_0^t \alpha^k(s)\frac{J^k(s)}{Y^k(s)}\, ds\end{aligned}$$

is a martingale and taking expectation it follows that $\tilde{\sigma}_k^2(t) = \mathbf{E}[\hat{\sigma}_k^2(t)]$ and $\hat{\sigma}_k^2(t)$ is an unbiased estimator. Heuristically, $A^{*k}(t)$ is "close" to the mean of $\widehat{A}^k(t)$ we can use $\hat{\sigma}_k^2(t)$ as an estimator for the variance of the Nelson-Aalen estimator $\widehat{A}^k(t)$.

23.2. Large sample properties. Let there be $i = 1, \ldots, n$ individuals and below we let $n \uparrow \infty$. As usual $T > 0$ is the upper time limit for the observations. Let the m-variate counting processes $N_i(t) = (N_i^1(t), \ldots, N_i^m(t))$, $t \in [0, T]$, have multiplicative intensity process $\lambda_i^k(t) = \alpha^k(t)Y_i^K(t)$. The aggregated counting processes $N_{(n)}^k(t) = \sum_{i=1}^n N_i^k(t)$ have multiplicative intensity process

$$\lambda_{(n)}^k(t) = \alpha^k(t)\sum_{i=1}^n Y_i^k(t) = \alpha^k(t)Y_{(n)}^k(t).$$

Define the process $J_{(n)}^k(t) = 1_{\{Y_{(n)}^k(t) > 0\}}$. Then the Nelson-Aalen estimator is given by

$$\widehat{A}_{(n)}^k(t) = \int_0^t \frac{J_{(n)}^k(s)}{Y_{(n)}^k(s)}\, dN_{(n)}^k(s)$$

and its predictable compensator is

$$A_{(n)}^{*k}(t) = \int_0^t \alpha^k(s)J_{(n)}^k(s)\, ds.$$

The next theorem states that the Nelson-Aalen estimator is uniformly consistent.

Theorem 23.1. *Let $t \in [0,T]$ and assume that*

$$\int_0^t \alpha^k(s) \frac{J^k_{(n)}(s)}{Y^k_{(n)}(s)}\, ds \xrightarrow[n\uparrow\infty]{\mathbf{P}} 0 \text{ and } \int_0^t \alpha^k(s)\big(1 - J^k_{(n)}(s)\big)\, ds \xrightarrow[n\uparrow\infty]{\mathbf{P}} 0.$$

Then

$$\sup_{0\le s\le t} |\widehat{A}^k_{(n)}(s) - A^k(s)| \xrightarrow[n\uparrow\infty]{\mathbf{P}} 0.$$

Proof. By Lenglart's inequality, Theorem 9.15

$$\begin{aligned}\mathbf{P}(\sup_{0\le s\le t} |\widehat{A}^k_{(n)}(s) - A^{*k}_{(n)}(s)| > \varepsilon) &\le \frac{\delta}{\varepsilon^2} + \mathbf{P}(\langle \widehat{A}^k_{(n)} - A^{*k}_{(n)}\rangle(t) > \delta) \\ &= \frac{\delta}{\varepsilon^2} + \mathbf{P}\Big(\int_0^t \alpha^k(s)\frac{J^k_{(n)}(s)}{Y^K_{(n)}(s)}\, ds > \delta\Big)\end{aligned}$$

and we get that

$$\sup_{0\le s\le t} |\widehat{A}^k_{(n)}(s) - A^{*k}_{(n)}(s)| \xrightarrow[n\uparrow\infty]{\mathbf{P}} 0.$$

Finally, we have that

$$\begin{aligned}\sup_{0\le s\le t} |\widehat{A}^k_{(n)}(s) - A^k(s)| &\le \sup_{0\le s\le t} |\widehat{A}^k_{(n)}(s) - A^{*k}_{(n)}(s)| + \sup_{0\le s\le t} |A^{*k}_{(n)}(s) - A^k(s)| \\ &\le \sup_{0\le s\le t} |\widehat{A}^k_{(n)}(s) - A^{*k}_{(n)}(s)| + \int_0^t \alpha^k(s)\big(1 - J^k_{(n)}(s)\big)\, ds \xrightarrow[n\uparrow\infty]{\mathbf{P}} 0.\end{aligned}$$

□

For the asymptotic distribution of the Nelson-Aalen estimator we have the following result.

Theorem 23.2. *Assume there is a sequence $(a_n, n = 1, 2, \ldots)$ such that $a_n \xrightarrow[n\uparrow\infty]{} \infty$ and assume that*

(i) *For any $t \in [0,T]$ and any $k = 1, \ldots, m$,*

$$a_n^2 \int_0^t \alpha^k(s) \frac{J^k_{(n)}(s)}{Y^k_{(n)}(s)}\, ds \xrightarrow[n\uparrow\infty]{\mathbf{P}} \sigma_k^2(t)$$

where $\sigma_k^2(t)$ is a continuous deterministic function.

(ii) *For $k = 1, \ldots, m$ and all $\varepsilon > 0$,*

$$a_n^2 \int_0^T \alpha^k(s) \frac{J^k_{(n)}(s)}{Y^k_{(n)}(s)} 1_{\{a_n J^k_{(n)}(s)/Y^k_{(n)}(s) > \varepsilon\}}\, ds \xrightarrow[n\uparrow\infty]{\mathbf{P}} 0.$$

(iii) *For $k = 1, \ldots, m$,*

$$a_n \int_0^T \alpha^k(t)\big(1 - J^k_{(n)}(t)\big)\, dt \xrightarrow[n\uparrow\infty]{\mathbf{P}} 0.$$

Then $((a_n(\widehat{A}^1_{(n)}(t) - A^1(t)), \ldots, a_n(\widehat{A}^m_{(n)}(t) - A^m(t)), 0 \le t \le T)$ converges in distribution to a process $((U^1(t), \ldots, U^m(t)), 0 \le t \le T)$ where $U^1(t), \ldots, U^m(t)$ are independent continuous Gaussian martingales with $U^k(0) = 0$ and $\operatorname{cov}(U^k(s), U^k(t)) = \sigma_k^2(s)$ for $s < t$. The variance $\sigma_k^2(t)$ may be estimated by $\hat{\sigma}_k^2(t)$.

Proof. We can apply the martingale central limit theorem to the martingales

$$\widetilde{M}^k_{(n)}(t) = a_n\big(\widehat{A}^j_{(n)}(t) - A^{*j}_{(n)}(t)\big) = \int_0^t a_n \frac{J^k_{(n)}(s)}{Y^k_{(n)}(s)}\, dM^k_{(n)}(s)$$

with the predictable (co)variation processes (see (23.3))

$$\langle \widetilde{M}^j_{(n)}, \widetilde{M}^k_{(n)}\rangle(t) = a_n^2 \langle \widehat{A}^j_{(n)} - A^{*j}_{(n)}, \widehat{A}^k_{(n)} - A^{*k}_{(n)}\rangle(t) = \delta_{jk} a_n^2 \int_0^t \alpha^k(s) \frac{J^k_{(n)}(s)}{Y^k_{(n)}(s)}\, ds.$$

The conditions (i) and (ii) ensure that the martingale central limit theorem (Theorem 11.15) can be applied and we get that the process $((a_n(\widehat{A}_{(n)}(t) - A^*_{(n)}(t))), 0 \le t \le T)$ converges in distribution to the continuous Gaussian martingale $(U(t), 0 \le t \le T)$ described in the theorem.

We have that $a_n |A^{*k}_{(n)}(t) - A^k(t)| = a_n \int_0^t \alpha^k(s)(1 - J^k_{(n)}(s))\, ds$ and then

$$\sup_{0\le t\le T} |a_n(A^{*k}_{(n)}(t) - A^k(t))| = a_n \int_0^T \alpha^k(t)(1 - J^k_{(n)}(t))\, dt \xrightarrow[n\uparrow\infty]{\mathbf{P}} 0.$$

Hence there is no asymptotically difference between $A(t)$ and $A^*_{(n)}(t)$ and it follows that $((a_n(\widehat{A}_{(n)}(t) - A(t))), 0 \le t \le T)$ converges in distribution to the continuous Gaussian martingale $(U(t), 0 \le t \le T$ described in the theorem. Finally, it can be shown that $\sup_{0\le t\le T} |a_n^2 \hat\sigma^2_k(t) - \sigma^2_k(t)| \xrightarrow{\mathbf{P}} 0$ for $n \uparrow \infty$ and $k = 1, \dots, m$. □

Example 23.3. (Markov model, revisited) Let the set-up be as in Example 22.3, that is, $Z(t)$ is a Markov process with finite state space $\mathcal{S} = \{1, 2, \dots, m\}$ and with transition intensities $\mu^{jk}(t)$, $j \ne k$. We assume that $Z(0) = 1$. Then $p^{1k}(0,t)$ is the probability that the Markov process is in state k at time t. We assume that the processes are observed in the time period $[0, T]$ which means that there is one common censoring time $c = T$. Moreover, we assume that $p^{1k}(0,t) > 0$ for all $t \in [0, T]$.

Let $Z_1(t), \dots, Z_n(t)$ be IID copies of $Z(t)$ and let

$$N^{jk}_{(n)}(t) = \sum_{i=1}^n \#\{0 < s \le t | Z_i(s-) = j, Z_i(s) = k\}, \quad j \ne k$$

be the number of transitions from state j to state k in $[0, t]$ for $Z_1(t), \dots, Z_n(t)$. Let $Y^j_{(n)}(t) = \sum_{i=1}^n 1_{\{Z_i(t-)=j\}}$ be the number of the $Z_i(t)$'s that are in state j at time $t-$. The multivariate counting process $(N^{jk}_{(n)}(t), j, k \in \mathcal{S}, j \ne k)$ has multiplicative intensity process $\lambda^{jk}_{(n)}(t) = \mu^{jk}(t)\, Y^j_{(n)}(t)$. The integrated transition intensities $A^{jk}(t) = \int_0^t \mu^{jk}(s)\, ds$ can be estimated by the Nelson-Aalen estimators

$$\widehat{A}^{jk}_{(n)}(t) = \int_0^t \frac{J^j_{(n)}(s)}{Y^j_{(n)}(s)}\, dN^{jk}_{(n)}(s)$$

with $J^j_{(n)}(t) = 1_{\{Y^j_{(n)}(t) > 0\}}$.

For the large sample properties, note that the Law of Large Numbers yields that

$$\frac{1}{n} Y^j_{(n)}(t) \xrightarrow[n\uparrow\infty]{\mathbf{P}} p^{1j}(0,t) > 0 \qquad \text{for } t \in [0, T]$$

and hence $Y^j_{(n)}(t) \xrightarrow[n\uparrow\infty]{\mathbf{P}} \infty$. Then we have that

$$n\frac{J^j_{(n)}(t)}{Y^j_{(n)}(t)}\mu^{jk}(t) = \frac{J^j_{(n)}(t)}{\frac{1}{n}Y^j_{(n)}(t)}\mu^{jk}(t) \xrightarrow[n\uparrow\infty]{\mathbf{P}} \frac{\mu^{jk}(t)}{p^{1j}(0,t)} \qquad \text{for } t \in [0,T].$$

It can be shown that

$$n\int_0^t \frac{J^j_{(n)}(s)}{Y^j_{(n)}(s)}\mu^{jk}(s)\,ds \xrightarrow[n\uparrow\infty]{\mathbf{P}} \int_0^t \frac{\mu^{jk}(s)}{p^{1j}(0,s)}\,ds.$$

This is condition (i) in Theorem 23.2 and using some similar arguments one can show that conditions (ii) and (iii) are satisfied. The process $((\sqrt{n}(\widehat{A}^{jk}_{(n)}(t) - A^{jk}(t)), j,k \in \mathcal{S}, j \neq k), 0 \leq t \leq T)$ converges in distribution to the process $((U^{jk}(t), j,k \in \mathcal{S}, j \neq k), 0 \leq t \leq T)$ by Theorem 23.2 where $U^{jk}(t)$'s are independent continuous Gaussian martingales with covariance given by $\operatorname{cov}(U^{jk}(s), U^{jk}(t)) = \int_0^s \mu^{jk}(u)/p^{1j}(0,u)\,du$ for $0 \leq s < t \leq T$ and $U^{jk}(0) = 0$.

23.3. Smoothing the Nelson-Aalen estimator. In the multiplicative intensity model, the Nelson-Aalen estimator is an estimator for $A^k(t) = \int_0^t \alpha^k(s)\,ds$. But it is $\alpha^k(t)$ that is of interest to estimate. When studying a plot of a Nelson-Aalen estimator, a rough estimate for $\alpha^k(t)$ is by looking at the slope. Here we provide an explicit estimator of $\alpha^k(t)$.

A kernel function $K(x)$ is a bounded function that vanishes outside $[-1,1]$ and has integral 1, that is,

- $0 \leq K(x) \leq$ constant for $x \in \mathbb{R}$.
- $K(x) = 0$ for $|x| > 1$.
- $\int_{-1}^1 K(x)\,dx = 1$.

From mathematical analysis we have that

$$\begin{aligned}\alpha^k(t) &= \lim_{b\downarrow 0}\int_{-1}^1 K(u)\alpha^k(t-bu)\,du = \lim_{b\downarrow 0}\int_{-(T-t)/b}^{t/b} K(u)\alpha^k(t-bu)\,du \\ &= \lim_{b\downarrow 0}\frac{1}{b}\int_0^T K\Big(\frac{t-s}{b}\Big)\alpha^k(s)\,ds = \lim_{b\downarrow 0}\frac{1}{b}\int_0^T K\Big(\frac{t-s}{b}\Big)\,dA^k(s).\end{aligned}$$

Examples of kernel functions are:

- $K(x) = \frac{1}{2}1_{[-1,1]}(x)$ (uniform kernel).
- $K(x) = \frac{3}{4}(1-x^2)1_{[-1,1]}(x)$ (Epanechnikov kernel).
- $K(x) = \frac{15}{16}(1-x^2)^2 1_{[-1,1]}(x)$ (biweight kernel).

The estimator for $\alpha^k(t)$ is obtained by smoothing the Nelson-Aalen estimator and is

$$\hat{\alpha}^k(t) = \frac{1}{b}\int_0^T K\Big(\frac{t-s}{b}\Big)\,d\widehat{A}^k(s)$$

where $K(\cdot)$ is a kernel function and $b > 0$ is the bandwidth. Both the kernel function and the bandwidth have to be chosen for concrete data. Let $\tau^k_1 < \tau^k_2 < \cdots$ be the jump times of $N^k(t)$, then

$$\hat{\alpha}^k(t) = \frac{1}{b}\sum_n K\Big(\frac{t-\tau^k_n}{b}\Big)\frac{1}{Y^k(\tau^k_n)}.$$

Only jump times τ_n^k satisfying $t-b \le \tau_n^k \le t+b$ contribute to the sum, since $K(x)=0$ for $|x|>1$. Therefore $\hat{\alpha}^k(t)$ is a weighted average of increments $1/Y^k(\tau_n^k)$ of the Nelson-Aalen estimator over $[t-b,t+b]$. Given b, a problem is to estimate $\alpha^k(t)$ in the tails (i.e. for $t\in[0,b]$ and $t\in[T-b,T]$), but we will only study the estimation of $\alpha^k(t)$ for $t\in[b,T-b]$.

Next we will study some statistical properties of $\hat{\alpha}^k(t)$. Define

$$\alpha^{*k}(t)=\frac{1}{b}\int_0^T K\Big(\frac{t-s}{b}\Big)\,dA^{*k}(s)=\frac{1}{b}\int_0^T K\Big(\frac{t-s}{b}\Big)\alpha^k(s)J^k(s)\,ds$$

and we have that

$$\hat{\alpha}^k(t)-\alpha^{*k}(t)=\frac{1}{b}\int_0^T K\Big(\frac{t-s}{b}\Big)\,d\big(\widehat{A}^k(s)-A^{*k}(s)\big)=\frac{1}{b}\int_0^T K\Big(\frac{t-s}{b}\Big)\frac{J^k(s)}{Y^k(s)}\,dM^k(s)$$

is a martingale (with T as the time parameter). Then the expected value of the estimator is

$$\mathbf{E}[\hat{\alpha}^k(t)]=\mathbf{E}[\alpha^{*k}(t)]=\frac{1}{b}\int_0^T K\Big(\frac{t-s}{b}\Big)\alpha^k(s)\mathbf{P}(Y^k(s)>0)\,ds.$$

We see that the estimator $\hat{\alpha}^k(t)$ is not even approximately unbiased. To find an estimator for the variance of $\hat{\alpha}^k(t)$, we will use the same type of arguments as for the Nelson-Aalen estimator. Recall that $T\mapsto\hat{\alpha}^k(t)-\alpha^{*k}(t)$ (with T as the time parameter) is a martingale and by Proposition 10.7 we get that the predictable variation process is

$$\langle\hat{\alpha}^k-\alpha^{*k}\rangle(t)=\int_0^T\Big(\frac{1}{b}K\Big(\frac{t-s}{b}\Big)\frac{J^k(s)}{Y^k(s)}\Big)^2\,d\langle M^k\rangle(s)=\frac{1}{b^2}\int_0^T K^2\Big(\frac{t-s}{b}\Big)\alpha^k(s)\frac{J^k(s)}{Y^k(s)}\,ds.$$

Define the function

$$\begin{aligned}\tilde{\psi}_k^2(t)&=\mathbf{E}[(\hat{\alpha}^k(t)-\alpha^{*k}(t))^2]=\mathbf{E}[\langle\hat{\alpha}^k-\alpha^{*k}\rangle(t)]\\&=\frac{1}{b^2}\int_0^T K^2\Big(\frac{t-s}{b}\Big)\alpha^k(s)\mathbf{E}\Big[\frac{J^k(s)}{Y^k(s)}\Big]\,ds=\frac{1}{b^2}\int_0^T K^2\Big(\frac{t-s}{b}\Big)\mathbf{E}\Big[\frac{J^k(s)}{Y^k(s)}\Big]\,dA^k(s)\end{aligned}$$

and note that

$$\begin{aligned}&\frac{1}{b^2}\int_0^T K^2\Big(\frac{t-s}{b}\Big)\frac{J^k(s)}{Y^k(s)^2}\,dM^k(s)\\&\qquad=\frac{1}{b^2}\int_0^T K^2\Big(\frac{t-s}{b}\Big)\frac{J^k(s)}{Y^k(s)^2}\,dN^k(s)-\frac{1}{b^2}\int_0^T K^2\Big(\frac{t-s}{b}\Big)\alpha^k(s)\frac{J^k(s)}{Y^k(s)}\,ds\end{aligned}$$

is a martingale. Define the estimator

$$\hat{\psi}_k^2(t)=\frac{1}{b^2}\int_0^T K^2\Big(\frac{t-s}{b}\Big)\frac{J^k(s)}{Y^k(s)^2}\,dN^k(s)$$

and using the above martingale we get that $\tilde{\psi}_k^2(t)=\mathbf{E}[\hat{\psi}_k^2(t)]$. If $\alpha^{*k}(t)$ is close to the expected value of $\hat{\alpha}^k(t)$, the variance of $\hat{\alpha}^k(t)$ is estimated by $\hat{\psi}_k^2(t)$.

The estimator $\hat{\alpha}^k(t)$ looks different for different choices of kernel functions and bandwidths. There are methods how to choose the kernel function and the bandwidth, but we will not study this point. There are also large sample properties of the estimator $\hat{\alpha}^k(t)$ stating that the estimator under some conditions is consistent and asymptotically normal distributed. This point will also not be studied.

24. Parametric models

In this section we will study parametrically specified multiplicative intensity models.

The general set-up is as follows. Let $T > 0$ be the upper time limit for the observations and let $N^k(t)$ count the number of type k events in $[0, t]$. The multiplicative intensity model is

$$\lambda^k(t;\theta) = \alpha^k(t;\theta)\, Y^k(t), \quad k = 1, \dots, m, \text{ and } t \in [0, T]$$

where $\alpha^k(t;\theta)$ is specified by a r-dimensional parameter $\theta = (\theta_1, \dots, \theta_r) \in \Theta$, where Θ is an open set of $\mathbb{R}^r$.

24.1. Likelihood. We will present the likelihood for the above parametric multiplicative intensity model. As a motivation for the general model we first consider some examples.

Example 24.1. The set-up and notation are as in Example 22.2. Let $\tau_1, \tau_2, \dots, \tau_n$ be IID survival times with a given parametric hazard rate (mortality rate) $\mu(t;\theta)$ parameterized by θ. Then the density and survival function are given by (see Section 20)

$$f(t;\theta) = \mu(t;\theta) \exp\Big(-\int_0^t \mu(s;\theta)\,ds\Big) \text{ and } \bar{F}(t;\theta) = \exp\Big(-\int_0^t \mu(s;\theta)\,ds\Big).$$

Let $c_1, c_2, \dots, c_n$ be given censoring times. The data are the censored survival times $\tilde{\tau}_i = \tau_i \wedge c_i$ and the survival indicators $D_i = 1_{\{\tilde{\tau}_i = \tau_i\}}$. The distribution of $(\tilde{\tau}_i, D_i)$ is

$$\mathbf{P}(\tilde{\tau}_i \le t, D_i = d_i) = \begin{cases} F(t \wedge c_i) & \text{if } d_i = 1 \\ \bar{F}(c_i) 1_{[c_i,\infty)}(t) & \text{if } d_i = 0. \end{cases}$$

If $D_i = 1$, the likelihood contribution is

$$f(\tilde{\tau}_i;\theta) = \mu(\tilde{\tau}_i;\theta) \exp\Big(-\int_0^{\tilde{\tau}_i} \mu(t;\theta)\,dt\Big) = \mu(\tilde{\tau}_i;\theta)^{D_i} \exp\Big(-\int_0^{\tilde{\tau}_i} \mu(t;\theta)\,dt\Big)$$

and if $D_i = 0$, the likelihood contribution is

$$\bar{F}(\tilde{\tau}_i;\theta) = \exp\Big(-\int_0^{\tilde{\tau}_i} \mu(t;\theta)\,dt\Big) = \mu(\tilde{\tau}_i;\theta)^{D_i} \exp\Big(-\int_0^{\tilde{\tau}_i} \mu(t;\theta)\,dt\Big).$$

The likelihood contribution of the i'th individual may be written as

$$L_i(\theta) = f(\tilde{\tau}_i;\theta)^{D_i} \bar{F}(\tilde{\tau}_i;\theta)^{1-D_i} = \mu(\tilde{\tau}_i;\theta)^{D_i} \exp\Big(-\int_0^{\tilde{\tau}_i} \mu(t;\theta)\,dt\Big).$$

The n individuals are independent and the full likelihood is

$$L(\theta) = \prod_{i=1}^n L_i(\theta) = \prod_{i=1}^n f(\tilde{\tau}_i;\theta)^{D_i} \bar{F}(\tilde{\tau}_i;\theta)^{1-D_i} = \prod_{i=1}^n \mu(\tilde{\tau}_i;\theta)^{D_i} \exp\Big(-\int_0^{\tilde{\tau}_i} \mu(t;\theta)\,dt\Big). \tag{24.1}$$

The likelihood can be rewritten in terms of the processes presented in Example 22.2: $\widetilde{N}_i(t) = 1_{[\tilde{\tau}_i,\infty)}(t) 1_{\{D_i=1\}}$ and $\widetilde{Y}_i(t) = 1_{[0,\tilde{\tau}_i]}(t)$. Then the likelihood $L_i(\theta)$ can be written as

$$L_i(\theta) = \Big(\prod_{0<t\le T} \big(\mu(t;\theta)\widetilde{Y}_i(t)\big)^{\Delta \widetilde{N}_i(t)}\Big) \exp\Big(-\int_0^T \mu(t;\theta)\widetilde{Y}_i(t)\,dt\Big)$$

and the full log-likelihood takes the form

$$
\begin{aligned}
l(\theta) = \log(L(\theta)) &= \sum_{i=1}^{n} \Big(\int_0^T \log\big(\mu(t;\theta)\widetilde{Y}_i(t)\big)\, d\widetilde{N}_i(t) - \int_0^T \mu(t;\theta)\widetilde{Y}_i(t)\, dt \Big) \\
&= \sum_{i=1}^{n} \int_0^T \log\big(\mu(t;\theta)\widetilde{Y}_i(t)\big)\, d\widetilde{N}_i(t) - \int_0^T \mu(t;\theta)\widetilde{Y}_{(n)}(t)\, dt.
\end{aligned}
$$

This shows the relationship between the process $\widetilde{N}(t) = (\widetilde{N}_1(t), \ldots, \widetilde{N}_n(t))$ and the likelihood.

Example 24.2. The set-up and the notation are as in Example 21.3 and Example 22.3. Let $Z(t)$ be the Markov chain given in Example 21.3 with parametric transition intensities $\mu^{jk}(t;\theta)$ parameterized by θ. From Section 7 we have that the Markov chain is determined by the jump times (transition times) $0 < \tau_1 < \tau_2 < \cdots$ and the state variables $Z(\tau_n) = Z_n$. Indeed, let

$$
N^{\bullet}(t) = \sum_j \sum_{k:k\neq j} N^{jk}(t)
$$

be the total number of jumps in $[0,t]$ where $N^{jk}(t)$ is the number of transitions from state j to state k (c.f. Example 21.3). Assume that the Markov chain is observed in the period $[0,c]$ where c is a censoring time. The distribution of the Markov chain $Z(t)$, or equivalently, the distribution of $N(t) = (N^{jk}(t),\, j,k \in \mathcal{S}, j \neq k)$ reduces to studying the joint distribution of the jump times and the state variables. For $n \in \{0,1,2,\ldots\}$, $0 = t_0 < t_1 < \cdots < t_n < t_{n+1} = c$, $j_0 = 1$ and $j_1, \ldots, j_n \in \mathcal{S}$ we have that

$$
\begin{aligned}
&\mathbf{P}(\tau_1 \in dt_1, \ldots, \tau_n \in dt_n, Z_1 = j_1, \ldots, Z_n = j_n, N^{\bullet}(c) = n) \\
&\quad = \mathbf{P}(\tau_1 \in dt_1, \ldots, \tau_n \in dt_n, Z_1 = j_1, \ldots, Z_n = j_n, \tau_{n+1} > c) \\
&\quad = \exp\Big(- \int_{t_n}^{c} \mu^{j_n j}(t;\theta)\, dt \Big) \Big(\prod_{i=1}^{n} \mu^{j_{i-1} j_i}(t_i;\theta) \exp\Big(- \int_{t_{i-1}}^{t_i} \mu^{j_{i-1}}(t;\theta)\, dt \Big)\, dt_i \Big) \\
&\quad = \exp\Big(\sum_{i=1}^{n} \log\big(\mu^{j_{i-1} j_i}(t_i;\theta)\big) - \sum_{i=1}^{n+1} \int_{t_{i-1}}^{t_i} \mu^{j_{i-1}}(t;\theta)\, dt \Big) dt_1 \cdots dt_n.
\end{aligned}
$$

Recall that $Y^j(t) = 1_{\{Z(t-)=j\}}$, then the likelihood contribution is

$$
\begin{aligned}
L(\theta) &= \exp\Big(\sum_j \sum_{k:k\neq j} \int_0^c \log\big(\mu^{jk}(t;\theta)\big)\, dN^{jk}(t) - \sum_j \int_0^c \mu^j(t;\theta) Y^j(t)\, dt \Big) \\
&= \exp\Big(\sum_j \sum_{k:k\neq j} \Big(\int_0^c \log\big(\mu^{jk}(t;\theta)\big)\, dN^{jk}(t) - \int_0^c \mu^{jk}(t;\theta) Y^j(t)\, dt \Big)\Big) \\
&= \exp\Big(\sum_j \sum_{k:k\neq j} \Big(\int_0^T \log\big(\mu^{jk}(t;\theta)\big)\, 1_{[0,c]}(t) dN^{jk}(t) - \int_0^T \mu^{jk}(t;\theta) Y^j(t) 1_{[0,c]}(t)\, dt \Big)\Big) \\
&= \exp\Big(\sum_j \sum_{k:k\neq j} \Big(\int_0^T \log\big(\tilde{\lambda}^{jk}(t;\theta)\big)\, d\widetilde{N}^{jk}(t) - \int_0^T \tilde{\lambda}^{jk}(t;\theta)\, dt \Big)\Big)
\end{aligned}
$$

where the censored (observed) counting processes are $\widetilde{N}^{jk}(t) = \int_0^t 1_{[0,c]}(s)\, dN^{jk}(s)$ with intensity processes $\tilde{\lambda}^{jk}(t;\theta) = \mu^{jk}(t;\theta)\widetilde{Y}^j(t)$ where $\widetilde{Y}^j(t) = 1_{[0,c]}(t) Y^j(t)$.

Let the Markov chains $Z_1(t), \ldots, Z_n(t)$ be independent replicates of the Markov chain described above. Again, let $c_1, c_2, \ldots, c_n$ be given censoring times. The process $Z_i(t)$ is observed from time 0 to the time of censoring c_i. Then the full log-likelihood may be written as

$$
\begin{aligned}
l(\theta) &= \log(L(\theta)) \\
&= \sum_{i=1}^{n} \sum_{j} \sum_{k:k\neq j} \Big(\int_0^T \log\big(\tilde{\lambda}_i^{jk}(t;\theta)\big)\, d\widetilde{N}_i^{jk}(t) - \int_0^T \tilde{\lambda}_i^{jk}(t;\theta)\, dt \Big)
\end{aligned} \tag{24.2}
$$

since the Markov processes are independent.

Based on the two examples 24.1 and 24.2, we now consider the multiplicative intensity model where the counting process $N(t) = (N^1(t), \ldots, N^m(t))$ is specified in the beginning of the section. Motivated by these examples, the (partial) log-likelihood for the general case is

$$
\begin{aligned}
l(\theta) &= \sum_{k=1}^{m} \Big(\int_0^T \log\big(\lambda^k(t;\theta)\big)\, dN^k(t) - \int_0^T \lambda^k(t;\theta)\, dt \Big) \\
&= \sum_{k=1}^{m} \Big(\int_0^T \log\big(\alpha^k(t;\theta) Y^k(t)\big)\, dN^k(t) - \int_0^T \alpha^k(t;\theta) Y^k(t)\, dt \Big).
\end{aligned} \tag{24.3}
$$

24.2. Maximum likelihood estimator. Consider the general multiplicative intensity model specified by the parameter $\theta = (\theta_1, \ldots, \theta_r) \in \Theta$. The maximum likelihood estimator (MLE) $\hat{\theta}$ of the unknown parameter θ is the value of the parameter which maximizes the likelihood of the given observations. In regular cases this maximum may be found by differentiation. If we assume that we may interchange the order of integration and differentiation then the score functions $U(\theta) = (U_1(\theta), \ldots, U_r(\theta))$ are given by

$$
U_q(\theta) = \frac{\partial}{\partial \theta_q} l(\theta) = \sum_{k=1}^{m} \Big(\int_0^T \frac{\partial}{\partial \theta_q} \log\big(\alpha^k(t;\theta)\big)\, dN^k(t) - \int_0^T \Big(\frac{\partial}{\partial \theta_q} \alpha^k(t;\theta)\Big) Y^k(t)\, dt \Big). \tag{24.4}
$$

The MLE $\hat{\theta}$ solves the likelihood equations $U_q(\theta) = 0$, $q = 1, \ldots, r$ and provides a global maximum (the likelihood equations may have multiple solutions). Below we will see that $\hat{\theta}$ has some of the properties for maximum likelihood estimators known from the cases of IID random variables. For example under some regularity conditions the likelihood ratio test statistics $2(l(\hat{\theta}) - l(\theta_0))$ is asymptotically χ^2 distributed with q degrees of freedom under the simple hypothesis of $H_0 : \theta = \theta_0$.

24.3. Large sample properties. The properties below are that under regularity conditions there exists a consistent solution $\hat{\theta}$ of the likelihood equations (with a probability tending to one) and $\hat{\theta}$ is asymptotically multi-normally distributed around its true parameter with a covariance matrix that may be estimated by $I(\hat{\theta})^{-1}$ where

$$
I_{pq}(\theta) = -\frac{\partial^2}{\partial \theta_p \partial \theta_q} l(\theta) = \sum_{k=1}^{m} \Big(\int_0^T \Big(\frac{\partial^2}{\partial \theta_p \partial \theta_q} \alpha^k(t;\theta)\Big) Y^k(t)\, dt - \int_0^T \frac{\partial^2}{\partial \theta_p \partial \theta_q} \log\big(\alpha^k(t;\theta)\big)\, dN^k(t) \Big)
$$

for $p, q = 1, \ldots, r$ is the observed information matrix. Below we present the main ideas of the derivations of these results.

The set-up is not the most general but it covers many cases of interesting applications. Moreover, not all the regularity conditions are stated in details. The true value of the parameter is denoted θ_0. Let there be $i = 1, \dots, n$ individuals and later we let $n \uparrow \infty$. The m-variate counting processes $N_i(t) = (N_i^1(t), \dots, N_i^m(t))$, $t \in [0, T]$, have the multiplicative intensity process $\lambda_i^k(t) = \alpha^k(t;\theta)Y_i^k(t)$. The aggregated counting processes $N_{(n)}^k(t) = \sum_{i=1}^n N_i^k(t)$ have multiplicative intensity process $\lambda_{(n)}^k(t) = \alpha^k(t;\theta)\sum_{i=1}^n Y_i^k(t) = \alpha^k(t;\theta)Y_{(n)}^k(t)$. In this case the log-likelihood can be written as

$$l^{(n)}(\theta) = \sum_{i=1}^n \sum_{k=1}^m \Big(\int_0^T \log\big(\alpha^k(t;\theta)Y_i^k(t)\big)\, dN_i^k(t) - \int_0^T \alpha^k(t;\theta)Y_i^k(t)\, dt \Big)$$

and the score function becomes

$$U_q^{(n)}(\theta) = \sum_{k=1}^m \Big(\int_0^T \frac{\partial}{\partial\theta_q} \log\big(\alpha^k(t;\theta)\big)\, dN_{(n)}^k(t) - \int_0^T \Big(\frac{\partial}{\partial\theta_q}\alpha^k(t;\theta)\Big) Y_{(n)}^k(t)\, dt \Big)$$

for $q = 1, \dots, r$.

Assumptions: (Not stated with all details). Let a_n be a sequence such that $a_n \uparrow \infty$ for $n \uparrow \infty$.

(a) There are functions $\sigma_{pq}(\theta)$ such that matrix $\Sigma(\theta_0) = \big(\sigma_{pq}(\theta_0),\, p, q = 1, \dots, r\big)$ is positive definite, and for all $p, q = 1, \dots, r$

$$\frac{1}{a_n^2} \sum_{k=1}^m \int_0^T \Big(\frac{\partial}{\partial\theta_p} \log\big(\alpha^k(t,\theta_0)\big)\Big)\Big(\frac{\partial}{\partial\theta_q} \log\big(\alpha^k(t,\theta_0)\big)\Big) \alpha^k(t;\theta_0)Y_{(n)}^k(t)\, dt \xrightarrow[n\uparrow\infty]{\mathbf{P}} \sigma_{jl}(\theta_0).$$

(b) For all $q = 1, \dots, r$ and all $\varepsilon > 0$, we have that

$$\frac{1}{a_n^2} \sum_{k=1}^m \int_0^T \Big(\frac{\partial}{\partial\theta_q} \log\big(\alpha^k(t;\theta_0)\big)\Big)^2 1_{\{|a_n^{-1}(\partial/\partial\theta_q)\log(\alpha^k(t;\theta_0))|>\varepsilon\}} \alpha^k(t;\theta_0)Y_{(n)}^k(t)\, dt \xrightarrow[n\uparrow\infty]{\mathbf{P}} 0.$$

(c) Regularity conditions on the partial derivatives of $\alpha^k(t;\theta)$ and $\log\big(\alpha^k(t;\theta)\big)$ of first, second and third order with respect to θ.

The notation $\xrightarrow[n\uparrow\infty]{\mathbf{P}}$ means converging in probability.

Theorem 24.3. *Under the assumptions above—with a probability tending to one—the equation $U^{(n)}(\theta) = 0$ has a solution $\hat\theta^{(n)}$ and $\hat\theta^{(n)} \xrightarrow[n\uparrow\infty]{\mathbf{P}} \theta_0$, that is, $\hat\theta^{(n)}$ is a consistence estimator.*

Theorem 24.4. *Under the assumptions above and $\hat\theta^{(n)}$ is the consistent MLE, $a_n(\hat\theta^{(n)} - \theta_0)$ converges in distribution to a multivariate normal distribution with mean zero and covariance matrix $\Sigma(\theta_0)^{-1}$ where $\Sigma(\theta_0)$ may be estimated by $a_n^{-2} I(\hat\theta^{(n)})$.*

Proof. This is a sketch of the proof of Theorem 24.3 and 24.4 where the main ideas are presented. Recall that $M_{(n)}^k(t) = N_{(n)}^k(t) - \int_0^t \lambda_{(n)}^k(s;\theta_0)\, ds$ is a martingale and then the score functions are given as

$$U_q^{(n)}(\theta_0) = \sum_{k=1}^m \int_0^T \frac{\partial}{\partial\theta_q} \log\big(\lambda_{(n)}^k(t;\theta_0)\big)\, dM_{(n)}^k(t) = \sum_{k=1}^m \int_0^T \frac{\partial}{\partial\theta_q} \log\big(\alpha^k(t;\theta_0)\big)\, dM_{(n)}^k(t).$$

By Proposition 10.5 we have that the score function is a martingale with predictable variation process (see Proposition 11.12 and Proposition 10.8)

$$\begin{aligned}\langle U_p^{(n)}(\theta_0), U_q^{(n)}(\theta_0)\rangle(T) &= \sum_{k=1}^{m}\int_0^T \Big(\frac{\partial}{\partial\theta_p}\log(\alpha^k(t;\theta_0))\Big)\Big(\frac{\partial}{\partial\theta_q}\log\big(\alpha^k(t;\theta_0)\big)\Big)\, d\langle M_{(n)}^k\rangle(t)\\ &= \sum_{k=1}^{m}\int_0^T \Big(\frac{\partial}{\partial\theta_p}\log\big(\alpha^k(t;\theta_0)\big)\Big)\Big(\frac{\partial}{\partial\theta_q}\log\big(\alpha^k(t;\theta_0)\big)\Big)\alpha^k(t;\theta_0)Y_{(n)}^k(t)\, dt.\end{aligned}$$

After some computations we get

$$I_{pq}^{(n)}(\theta_0) = \langle U_p^{(n)}(\theta_0), U_q^{(n)}(\theta_0)\rangle(T) - \sum_{k=1}^{m}\int_0^T \frac{\partial^2}{\partial\theta_p\partial\theta_q}\log\big(\alpha^k(t;\theta_0)\big)\, dM_{(n)}^k(t). \tag{24.5}$$

The latter term is a martingale with mean zero. One can show using assumption (a) (and assumption (c)) that

$$\frac{1}{a_n^2}U^{(n)}(\theta_0) \xrightarrow[n\uparrow\infty]{\mathbf{P}} 0, \quad \frac{1}{a_n^2}I^{(n)}(\theta_0) \xrightarrow[n\uparrow\infty]{\mathbf{P}} \Sigma(\theta_0) \quad \text{and} \quad \frac{1}{a_n^2}I^{(n)}(\hat\theta^{(n)}) \xrightarrow[n\uparrow\infty]{\mathbf{P}} \Sigma(\theta_0).$$

The MLE $\hat\theta^{(n)}$ satisfies the likelihood equation $U^{(n)}(\hat\theta^{(n)}) = 0$ and by a Taylor series expansion around θ_0 we get that

$$0 = U^{(n)}(\hat\theta^{(n)}) = U^{(n)}(\theta_0) - I^{(n)}(\theta_0)(\hat\theta^{(n)} - \theta_0) + \text{rest-term}. \tag{24.6}$$

Or $a_n^{-2}U^{(n)}(\theta_0) = a_n^{-2}I^{(n)}(\theta_0)(\hat\theta^{(n)} - \theta_0) + a_n^{-2} \times$ (rest-term) and we get that $\hat\theta^{(n)} \xrightarrow[n\uparrow\infty]{\mathbf{P}} \theta_0$. (Assumption (c) ensures that the rest term converges to zero in probability). This is the main ideas in the proof of Theorem 24.3.

For Theorem 24.4, assumptions (a) and (b) ensure that Martingale central limit theorem 11.15 can be applied. Hence, $a_n^{-1}U^{(n)}(\theta_0)$ converges in distribution to a multivariate normal distribution with mean zero and with covariance matrix $\Sigma(\theta_0)$. Using (24.6) we obtain that $a_n(\hat\theta^{(n)} - \theta_0) \approx (a_n^{-2}I^{(n)}(\theta_0))^{-1}\, a_n^{-1}U^{(n)}(\theta_0)$ and it follows that $\hat\theta$ is approximately multivariate normal distributed around θ_0 and the covariance matrix may be estimated by $I^{(n)}(\hat\theta)^{-1}$. □

Remark 24.5. Note by (24.5) we get that $\mathbf{E}[I_{pq}^{(n)}(\theta_0)] = \mathbf{E}[\langle U_p^{(n)}(\theta_0), U_q^{(n)}(\theta_0)\rangle(T)]$. That is the expected information matrix equals the covariance matrix of the score function.

Example 24.6. (Censored survival times, revisited) This is a continuation of Example 22.2 and Example 24.1. Let $\tau_1, \dots, \tau_n$ be IID survival times and $c_1, \dots, c_n \le T$ where T as usual is the upper time limit for the observations. We do not observe the survival times themselves, only the censored survival times $\tilde\tau_i = \tau_i \wedge c_i$ and the survival indicators $D_i = 1_{\{\tilde\tau_i = \tau_i\}}$. The given hazard rate $\mu(t;\theta)$ of the actual survival times τ_i is assumed to depend on a r-dimensional parameter $\theta = (\theta_1, \dots, \theta_r)$. The aggregated counting process $\widetilde{N}_{(n)}(t) = \sum_{i=1}^n 1_{[\tilde\tau_i,\infty)}(t)1_{\{D_i=1\}}$ has intensity process on the multiplicative form $\tilde\lambda_{(n)}(t;\theta) = \mu(t;\theta)\widetilde{Y}_{(n)}(t)$ where $\widetilde{Y}_{(n)}(t) = \sum_{i=1}^n 1_{[0,\tilde\tau_i]}(t)$. By (24.3), the log-likelihood is

$$l(\theta) = \int_0^T \log\big(\mu(t;\theta)\big)\, d\widetilde{N}_{(n)}(t) - \int_0^T \mu(t;\theta)\widetilde{Y}_{(n)}(t)\, dt$$

and the score function in (24.4) takes the form

$$U_q(\theta) = \int_0^T \frac{\partial}{\partial \theta_q} \log\big(\mu(t;\theta)\big)\, d\widetilde{N}_{(n)}(t) - \int_0^T \Big(\frac{\partial}{\partial \theta_q}\mu(t;\theta)\Big)\widetilde{Y}_{(n)}(t)\, dt.$$

The maximum likelihood estimator $\hat{\theta}$ is the solution to the likelihood equations $U_q(\theta) = 0$ for $q = 1, \ldots, r$. In this general case it is not possible to find analytic expression for the MLE. We have to numerically solve the likelihood equations that takes the form

$$\sum_{i=1}^n D_i \frac{\partial}{\partial \theta_q} \log\big(\mu(\tilde{\tau}_i;\theta)\big) = \sum_{i=1}^n \int_0^{\tilde{\tau}_i} \frac{\partial}{\partial \theta_q}\mu(t;\theta)\, dt$$

for $q = 1, \ldots, r$ to compute the MLE $\hat{\theta}$. Below we study two concrete hazard rates.

Exponential distribution: Assume that the survival times are exponentially distributed (see Example 20.1) with a hazard rate given by $\mu(t;\lambda) = \lambda$, $\lambda > 0$. (Note $q = 1$). In this case the score function is given by

$$\begin{aligned} U(\lambda) &= \int_0^T \frac{\partial}{\partial \lambda} \log\big(\mu(t;\lambda)\big)\, d\widetilde{N}_{(n)}(t) - \int_0^T \frac{\partial}{\partial \lambda}\mu(t;\lambda)\widetilde{Y}_{(n)}(t)\, dt \\ &= \frac{1}{\lambda}\widetilde{N}_{(n)}(T) - R(T) \end{aligned}$$

where $R(T) = \int_0^T \widetilde{Y}_{(n)}(t)\, dt = \sum_{i=1}^n \tilde{\tau}_i$ is the exposure (the total observation time). It follows that the MLE is the occurrence/exposure rate

$$\hat{\lambda} = \frac{\widetilde{N}_{(n)}(T)}{R(T)}.$$

In this case we can compute the asymptotic distribution of the MLE $\hat{\lambda}$. We need to verify the assumptions on page 84. Assume that λ_0 is the true value.

$$\begin{aligned} \sum_{i=1}^n \int_0^T \Big(\frac{\partial}{\partial \lambda} \log\big(\mu(t;\lambda_0)\big)\Big)^2 \mu(t;\lambda_0) Y_i(t)\, dt &= \sum_{i=1}^n \int_0^T \frac{1}{\lambda_0^2}\lambda_0 1_{[0,\tilde{\tau}_i]}(t)\, dt \\ &= \frac{1}{\lambda_0}\sum_{i=1}^n \tilde{\tau}_i. \end{aligned}$$

Set $a_n = \sqrt{\sum_{i=1}^n (1 - \exp(-\lambda_0 c_i))}$ and note that $a_n \uparrow \infty$ since the c_i's are bounded away from zero. We have that

$$\mathbf{E}[\tilde{\tau}_i] = \frac{1}{\lambda_0}(1 - e^{-\lambda_0 c_i}).$$

It can then be verified that $a_n^{-2}\sum_{i=1}^n \tilde{\tau}_i \xrightarrow[n\uparrow\infty]{\mathbf{P}} 1/\lambda_0$. In this case assumption (a) is satisfied with

$$a_n^{-2}\sum_{i=1}^n \int_0^T \Big(\frac{\partial}{\partial \lambda} \log\big(\mu(t;\lambda_0)\big)\Big)^2 \mu(t;\lambda_0) Y_i(t)\, dt \xrightarrow[n\uparrow\infty]{\mathbf{P}} \frac{1}{\lambda_0^2}.$$

For assumption (b): note that $1_{\{|a_n^{-1}(\partial/\partial\lambda)\log(\mu(t;\lambda_0))|>\varepsilon\}} = 1_{\{a_n^{-1}>\varepsilon\lambda_0\}} = 0$ for n large. Then we have that assumption (b) is satisfied. By Theorem 24.3 and Theorem 24.4 we have that $\hat{\lambda}$ is a consistent estimator and $a_n(\hat{\lambda} - \lambda_0)$ is asymptotically normal with mean zero and variance λ_0^2.

Gompertz-Makeham hazard rate: We will consider a model where the hazard rate is of Gompertz-Makeham (see Example 20.1)

$$\mu(t;\theta) = \alpha + \beta e^{\gamma t}$$

with $\theta = (\alpha, \beta, \gamma)$. In this case the score functions in (24.4) are given by

$$\begin{aligned}
U_\alpha(\theta) &= \sum_{i=1}^n \Big(\int_0^T \frac{1}{\alpha + \beta e^{\gamma t}}\, d\widetilde{N}_i(t) - \int_0^T \widetilde{Y}_i(t)\, dt \Big) = \sum_{i=1}^n \Big(\frac{D_i}{\alpha + \beta \exp(\gamma \tilde{\tau}_i)} - \tilde{\tau}_i \Big) \\
U_\beta(\theta) &= \sum_{i=1}^n \Big(\int_0^T \frac{e^{\gamma t}}{\alpha + \beta e^{\gamma t}} d\widetilde{N}_i(t) - \int_0^T e^{\gamma t}\widetilde{Y}_i(t)\, dt \Big) = \sum_{i=1}^n \Big(\frac{D_i \exp(\gamma \tilde{\tau}_i)}{\alpha + \beta \exp(\gamma \tilde{\tau}_i)} - \frac{\exp(\gamma \tilde{\tau}_i) - 1}{\gamma} \Big) \\
U_\gamma(\theta) &= \sum_{i=1}^n \Big(\int_0^T \frac{\beta t e^{\gamma t}}{\alpha + \beta e^{\gamma t}}\, d\widetilde{N}_i(t) - \int_0^T \beta t e^{\gamma t}\widetilde{Y}_i(t)\, dt \Big) \\
&= \sum_{i=1}^n \Big(\frac{D_i \beta \tilde{\tau}_i \exp(\gamma \tilde{\tau}_i)}{\alpha + \beta \exp(\gamma \tilde{\tau}_i)} - \beta \Big(\frac{\tilde{\tau}_i \exp(\gamma \tilde{\tau}_i)}{\gamma} - \frac{\exp(\gamma \tilde{\tau}_i) - 1}{\gamma^2} \Big) \Big).
\end{aligned}$$

Then the likelihood equations read as

$$\begin{aligned}
\sum_{i=1}^n \frac{D_i}{\alpha + \beta \exp(\gamma \tilde{\tau}_i)} &= \sum_{i=1}^n \tilde{\tau}_i \\
\sum_{i=1}^n \frac{D_i \exp(\gamma \tilde{\tau}_i)}{\alpha + \beta \exp(\gamma \tilde{\tau}_i)} &= \sum_{i=1}^n \frac{\exp(\gamma \tilde{\tau}_i) - 1}{\gamma} \\
\sum_{i=1}^n \frac{D_i \beta \tilde{\tau}_i \exp(\gamma \tilde{\tau}_i)}{\alpha + \beta \exp(\gamma \tilde{\tau}_i)} &= \beta \sum_{i=1}^n \Big(\frac{\tilde{\tau}_i \exp(\gamma \tilde{\tau}_i)}{\gamma} - \frac{\exp(\gamma \tilde{\tau}_i) - 1}{\gamma^2} \Big).
\end{aligned}$$

These equations have to be solved numerically. For asymptotic results, it can be shown that the assumptions are satisfied and the asymptotic covariance may be estimated using the observed information matrix $I(\hat{\theta})^{-1}$.

Example 24.7. (Markov model, revisited) Let $Z_1(t), \dots, Z_n(t)$ and other processes be defined as in Example 24.2. The log-likelihood is given in (24.2). The score functions are

$$U_q(\theta) = \sum_j \sum_{k:k\neq j} \Big(\int_0^T \frac{\partial}{\partial \theta_q} \log\big(\mu^{jk}(t;\theta)\big)\, d\widetilde{N}^{jk}_{(n)}(t) - \int_0^T \Big(\frac{\partial}{\partial \theta_q} \mu^{jk}(t;\theta)\Big) \widetilde{Y}^j_{(n)}(t)\, dt \Big). \tag{24.7}$$

Then the maximum likelihood estimator $\hat{\theta}$ is the solution to

$$\sum_j \sum_{k:k\neq j} \int_0^T \frac{\partial}{\partial \theta_q} \log\big(\mu^{jk}(t;\theta)\big)\, d\widetilde{N}^{jk}_{(n)}(t) = \sum_j \sum_{k:k\neq j} \int_0^T \Big(\frac{\partial}{\partial \theta_q} \mu^{jk}(t;\theta)\Big) \widetilde{Y}^j_{(n)}(t)\, dt.$$

If the assumptions (a)-(c) are satisfied we have the large sample properties.

Next is a concrete Markov model with piecewise constant intensities. Consider a partition $0 = t_0 < t_1 < \cdots < t_L = T$ of the observation time interval and assume that the transition intensities are on the form

$$\mu^{jk}(t;\theta) = \sum_{l=1}^L \theta_l^{jk} 1_{(t_{l-1}, t_l]}(t).$$

Hence it is a parametric model with parameter $\theta = (\theta_l^{jk}, l = 1, \ldots, L, j, k \in \mathcal{S}, j \neq k)$. For the score functions we use the notation $U_l^{jk}(\theta) = \frac{\partial}{\partial \theta_l^{jk}} l(\theta)$ and by (24.7) it is given by

$$U_l^{jk}(\theta) = \frac{1}{\theta_l^{jk}} \int_{t_{l-1}}^{t_l} d\widetilde{N}_{(n)}^{jk}(t) - \int_{t_{l-1}}^{t_l} \widetilde{Y}_{(n)}^{j}(t)\, dt = \frac{O_l^{jk}}{\theta_l^{jk}} - R_l^j$$

where

$$O_l^{jk} = \int_{t_{l-1}}^{t_l} d\widetilde{N}_{(n)}^{jk}(t) = \widetilde{N}_{(n)}^{jk}(t_l) - \widetilde{N}_{(n)}^{jk}(t_{l-1})$$

is the total number of transitions from state j to state k in the l'th subinterval and

$$R_l^j = \int_{t_{l-1}}^{t_l} \widetilde{Y}_{(n)}^{j}(t)\, dt$$

is the corresponding total observation time. The maximum likelihood estimators are the occurrence/exposure rates

$$\hat{\theta}_l^{jk} = \frac{O_l^{jk}}{R_l^j}.$$

For asymptotic results we will assume that there exist functions r_l^j such that

$$\frac{1}{n} R_l^j \xrightarrow[n\uparrow\infty]{\mathbf{P}} r_l^j.$$

Let the θ_{0l}^{jk}'s be the true values. Then we have that assumption (a) on page 84 is satisfied by

$$\frac{1}{n} \int_0^T \left(\frac{1}{\theta_{0l}^{jk}}\right)^2 \theta_{0l}^{jk} 1_{(t_{l-1}, t_l]}(t) \widetilde{Y}_{(n)}^{j}(t)\, dt \xrightarrow[n\uparrow\infty]{\mathbf{P}} \frac{r_l^j}{\theta_{0l}^{jk}}.$$

Assumption (b) is satisfied since the indicator function is zero for n large. Then the vector with elements $\sqrt{n}(\hat{\theta}_l^{jk} - \theta_{0l}^{jk})$ converges in distribution to a vector with components which are independent and normally distributed with means zero and variances θ_{0l}^{jk}/r_l^j. In other words the occurrence/exposure rates $\hat{\theta}_l^{jk}$, $j, k \in \mathcal{S}$, $j \neq k$, $l = 1, \ldots, L$, are approximately independent and normally distributed around their true values and the variance of $\hat{\theta}_l^{jk}$ can be estimated by $(\hat{\theta}_l^{jk})^2 / O_l^{jk} = O_l^{jk} / (R_l^{jk})^2$.

APPENDIX

Review of measure and integration theory

The aim of this appendix is to give a brief resumé of key notions on measure and integration theory that is used extensively throughout the notes.

25. Measurability and measure

25.1. σ-algebras. Let Ω be a given and fixed set.

Definition 25.1. A collection $\mathcal{F}$ of subsets of Ω is a σ-algebra if

(i) $\Omega \in \mathcal{F}$.
(ii) If $A \in \mathcal{F}$ then $A^c = \Omega \setminus A \in \mathcal{F}$.
(iii) If $A_n \in \mathcal{F}$, $n \in \mathbb{N}$ then $\cup_{n=1}^{\infty} A_n \in \mathcal{F}$.

The pair $(\Omega, \mathcal{F})$ is called a measurable space.

Thus, a σ-algebra on Ω is a family of subsets of Ω, stable under any countable collection of set operations.

Example 25.2. The collection of the empty set and Ω is denoted by $\mathcal{F}_* = \{\emptyset, \Omega\}$. The collection of all subsets of Ω –called the power set – is denoted by $2^{\Omega} = \{A | A \subseteq \Omega\}$. Note that, $\mathcal{F}_*$ is the trivial σ-algebra (smallest σ-algebra) and 2^{Ω} is the largest σ-algebra.

Given a collection $\mathcal{E}$ of subsets of Ω. Then $\sigma(\mathcal{E})$ is the smallest σ-algebra containing $\mathcal{E}$, that is, $\mathcal{E} \subseteq \sigma(\mathcal{E})$. The σ-algebra $\sigma(\mathcal{E})$ is generated by $\mathcal{E}$ and is the intersection of all σ-algebras containing $\mathcal{E}$ (the power set is a σ-algebra containing $\mathcal{E}$).

Example 25.3. Let Ω be a subset of $\mathbb{R}^d$. Let the Borel σ-algebra $\mathcal{B}(\Omega)$ on Ω, be the σ-algebra generated by the family of all open sets of Ω.

Definition 25.4. A collection $\mathcal{D}$ of subsets of Ω is a d-system (Dynkin-system) if

(i) $\Omega \in \mathcal{D}$.
(ii) If $A, B \in \mathcal{D}$ and $A \subseteq B$ then $B \setminus A \in \mathcal{D}$.
(iii) If $A_n \in \mathcal{D}$ and $A_n \subseteq A_{n+1}$ for $n \in \mathbb{N}$ then $\cup_{n=1}^{\infty} A_n \in \mathcal{D}$.

Note that the three conditions implies the following:

$$\text{If } A_n \in \mathcal{D} \text{ and } A_n \supseteq A_{n+1} \text{ for } n \in \mathbb{N} \text{ then } \cap_{n=1}^{\infty} A_n \in \mathcal{D}.$$

Thus $\mathcal{D}$ is a monotone-system.

Theorem 25.5. (Monotone class theorem for sets). *Let $\mathcal{C}$ and $\mathcal{D}$ be classes of subsets of Ω satisfying $\mathcal{C} \subseteq \mathcal{D}$. Suppose that $\mathcal{C}$ is closed under finite intersection and $\mathcal{D}$ is a d-system. Then $\sigma(\mathcal{C}) \subseteq \mathcal{D}$.*

25.2. Measurable functions. This section is largely a matter of acquainting the notation.

Let $(\Omega, \mathcal{F})$ and $(S, \mathcal{B})$ be measurable spaces, and let $f : \Omega \to S$ be a mapping between the two spaces Ω and S. The inverse image f^{-1} of a set is given by

$$f^{-1}(B) = \{\omega \in \Omega | f(\omega) \in B\}$$

for a subset B of S. Note that f^{-1} preserves all set operations and hence

$$\sigma(f) = f^{-1}(\mathcal{B}) = \{f^{-1}(B) | B \in \mathcal{B}\}$$

is a σ-algebra on Ω.

Definition 25.6. The function f is $(\mathcal{F}, \mathcal{B})$-measurable if $f^{-1}(B) = \{\omega \in \Omega | f(\omega) \in B\} \in \mathcal{F}$ for every $B \in \mathcal{B}$, that is, $\sigma(f) \subseteq \mathcal{F}$.

Note that $\sigma(f)$ is the smallest σ-algebra on Ω such that f measurable.

Lemma 25.7. (Composition Lemma). *Let $(\Omega, \mathcal{F})$, $(S_1, \mathcal{B}_1)$ and $(S_2, \mathcal{B}_2)$ be measurable spaces. Let $f : \Omega \to S_1$ be $(\mathcal{F}, \mathcal{B}_1)$-measurable and $g : S_1 \to S_2$ be $(\mathcal{B}_1, \mathcal{B}_2)$-measurable. Then the composition $h = g \circ f : \Omega \to S_2$ is $(\mathcal{F}, \mathcal{B}_2)$-measurable.*

Example 25.8. (Borel functions). If $\Omega \subseteq \mathbb{R}^d$ and $S \subseteq \mathbb{R}^m$ then a $(\mathcal{B}(\Omega), \mathcal{B}(S))$-measurable function $f : \Omega \to S$ is called a Borel function.

Lemma 25.9. (Functional representation). *Let $U \subseteq \mathbb{R}^d$ and let $h : \Omega \to U$ be a function. Then h is $(\sigma(f), \mathcal{B}(U))$-measurable if and only if there exists some $(\mathcal{B}, \mathcal{B}(U))$-measurable function $g : S \to U$ with $h = g \circ f$.*

Definition 25.10. Let $f_j : S \to T$, $j \in J$ be a collection of functions. Then

$$\sigma(f_j | j \in J) = \sigma(f_j^{-1}(B) | B \in \mathcal{B}, j \in J) = \sigma(\cup_{j \in J} \sigma(f_j))$$

is the smallest σ-algebra on S such that all the function f_j, $j \in J$ are measurable.

Let $f : \Omega \to \mathbb{R}$ be a real valued measurable function. f is a simple function if it only assumes finitely many different values. Let $f(\Omega) = \{a_1, \dots, a_n\}$ be the values of f and set $A_j = \{f = a_j\}$ for $j = 1, \dots, n$. Then

$$f = \sum_{j=1}^{n} a_j 1_{A_j}$$

and $A_j \in \mathcal{F}$ with $\Omega = \cup_{j=1}^{n} A_j$.

Lemma 25.11. *Let $f_m : \Omega \to \mathbb{R}$ be a sequence of measurable functions. Then*

(i) $\inf f_m$ *and* $\sup f_m$ *are measurable.*
(ii) $\liminf f_m$ *and* $\limsup f_m$ *are measurable.*
(iii) $\{\omega \in \Omega : \lim f_m(\omega)$ *exists in* $\mathbb{R}\} \in \mathcal{F}$.

Lemma 25.12. *Let $f : \Omega \to [0, \infty)$ be a positive measurable function. Then there exists simple functions $f_1 \le f_2 \le \cdots$ with $0 \le f_n \uparrow f$.*

Proof. For $n = 1, 2, \dots$, put

$$f_n(\omega) = \sum_{i=1}^{n2^n} (i-1)2^{-n} 1_{\{(i-1)2^{-n} \le f(\omega) < i2^{-n}\}} + n 1_{\{f(\omega) \ge n\}}.$$

Since f is measurable we have the sets $\{(i-1)2^{-n} \le f(\omega) < i2^{-n}\}$ and $\{f(\omega) \ge n\}$ belongs to $\mathcal{F}$. Hence by the construction f_n is a simple function. By construction $\le f_n \uparrow f$. $\square$

25.3. Monotone class theorem. A fundamental theorem of measure theory that is used from time to time is known as Monotone class theorem. There are several variants of this theorem, but the one given here is sufficient for our needs.

Definition 25.13. A set (collection) $\mathcal{H}$ of bounded real-valued functions f on a space Ω is called a monotone vector space (MVS) if the following three conditions are satisfied

(i) $\mathcal{H}$ is a vector (linear) space over $\mathbb{R}$.
(ii) $1_\Omega \in \mathcal{H}$ (that is,constant functions are in $\mathcal{H}$).
(iii) If $\{f_n\}_{n\geq 1} \subseteq \mathcal{H}$, and $f_1 \leq f_2 \leq \cdots \leq f_n \leq \cdots$, and $f = \lim_{n\to\infty} f_n$ is bounded, then $f \in \mathcal{H}$.

Definition 25.14. A set $\mathcal{M}$ of real functions on a space Ω is called a multiplicative set if $f, g \in \mathcal{M}$ implies that $fg \in \mathcal{M}$.

For the set $\mathcal{M}$, we let $\sigma(\mathcal{M})$ be the σ-algebra on Ω generated by $\{f^{-1}(A)|A \in \mathcal{B}(\mathbb{R}), f \in \mathcal{M}\}$, that is, $\sigma(\mathcal{M})$ is the smallest σ-algebra making all the functions in $\mathcal{M}$ measurable.

Theorem 25.15. (Monotone class theorem). *Let $\mathcal{H}$ be a monotone vector space on Ω and let $\mathcal{M}$ be a multiplicative class on Ω. If $\mathcal{M}$ is a subset of $\mathcal{H}$, $\mathcal{M} \subseteq \mathcal{H}$, then $\mathcal{H}$ contains all bounded $\sigma(\mathcal{M})$-measurable functions.*

For many applications of the Monotone class theorem we can use the following two corollaries of the theorem.

Corollary 25.16. *Let $\mathcal{A}$ be a σ-algebra on some space Ω. Let $\mathcal{H}$ be a monotone vector space of bounded real-valued functions f on Ω. If $1_A \in \mathcal{H}$ for all $A \in \mathcal{A}$ then $\mathcal{H}$ contains all bounded $\mathcal{A}$-measurable functions.*

Proof. Since $\mathcal{A}$ is stable under finite intersection and $\mathcal{A} = \sigma(\mathcal{A})$, the result follows from Corollary 25.17 below.

□

Corollary 25.17. *Let $\mathcal{A}$ be a σ-algebra on some space Ω such that $\mathcal{A} = \sigma(\mathcal{C})$ where $\mathcal{C}$ is a class of subsets of the space Ω which is stable under finite intersections, that is, $A, B \in \mathcal{C}$ implies $A \cap B \in \mathcal{C}$. Let $\mathcal{H}$ be a monotone vector space of bounded real-valued functions f on Ω. If $1_C \in \mathcal{H}$ for all $C \in \mathcal{C}$ then $\mathcal{H}$ contains all bounded $\mathcal{A}$-measurable functions.*

Proof. Let $\mathcal{M}$ be the set of indicator functions of every set in $\mathcal{C}$. Note that $1_A 1_B = 1_{A\cap B}$ and hence $\mathcal{M}$ is a multiplicative set. Moreover, we have that $\sigma(\mathcal{M}) = \sigma(\mathcal{C}) = \mathcal{A}$. Hence by Monotone class theorem, Theorem 25.15, the result follows. □

A rough outline how to apply Monotone class theorem is the following: Let $\mathcal{A}$ be a σ-algebra on Ω. Assume that $\mathcal{A} = \sigma(\mathcal{C})$ where $\mathcal{C}$ is given as in the corollary above. Note that $\mathcal{C}$ might be the σ-algebra $\mathcal{A}$ it self. We want to show that all bounded real-valued $\mathcal{A}$-measurable functions have a "property".

1. Show that the set $\mathcal{H} = \{f$ bounded real-valued $\mathcal{A}$-measurable function$|f$ has "property"$\}$ is a monotone vector space.
2. Show that $\{1_A | A \in \mathcal{C}\} \subset \mathcal{H}$.

Then any bounded real-valued $\mathcal{A}$-measurable function has "property".
See Section 28 for an application of Monotone class theorem with all details. Moreover, the following sections also demonstrated applications of Monotone class theorem.

25.4. Measures.

Definition 25.18. Let $(\Omega, \mathcal{F})$ be a measurable space. A measure on $(\Omega, \mathcal{F})$ is a mapping $\mu : \mathcal{F} \to [0, \infty]$ that satisfies

(i) $\mu(\emptyset) = 0$

(ii) $\mu(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n)$ for every sequence $A_n \in \mathcal{F}$, $n \in \mathbb{N}$, of mutually disjoint sets.

The triple $(\Omega, \mathcal{F}, \mu)$ is called a measure space.

- μ is a finite measure if $\mu(S) < \infty$.
- μ is a σ-finite measure if there is a sequence $A_n \in \mathcal{F}$, $n \in \mathbb{N}$, such that $S = \cup_{n=1}^{\infty} A_n$ and $\mu(A_n) < \infty$.

Some basic properties.

Lemma 25.19. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and let A_n, $n \in \mathbb{N}$ be a sequence of sets in $\mathcal{A}$. Then*

(i) *$A_n \uparrow A$ implies $\mu(A_n) \uparrow \mu(A)$.*

(ii) *$A_n \downarrow A$ and $\mu(A_1) < \infty$ implies $\mu(A_n) \downarrow \mu(A)$.*

Lemma 25.20. (i) *Let f be a measurable mapping between two measurable space $(\Omega, \mathcal{F})$ and $(S, \mathcal{B})$. Let μ be a measure on $(\Omega, \mathcal{F})$. Then*

$$\nu(B) = \mu(f^{-1}(B)) = \mu(\{\omega \in \Omega | f(\omega) \in B\}), \quad B \in \mathcal{B}$$

is a measure on $(S, \mathcal{B})$.

(ii) *Let μ_1 and μ_2 be two measures on $(\Omega, \mathcal{F})$ and let $\alpha \geq 0$ and $\beta \geq 0$ be positive constants. Then*

$$\mu(A) = \alpha\mu_1(A) + \beta\mu_2(A)$$

is a measure $(\Omega, \mathcal{F})$.

This note assumes that the reader has some knowledge of with theory of measures, including probability measures and the Lebesgue measure defined below.

Definition 25.21. (Probability measure). A measure $\mathbf{P}$ on $(\Omega, \mathcal{F})$ is a probability measure if

$$\mathbf{P}(\Omega) = 1$$

and $(\Omega, \mathcal{F}, \mathbf{P})$ is a probability space.

Definition 25.22. (Lebesgue measure). There is a unique σ-finite measure λ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that

$$\lambda((a, b]) = b - a$$

for all $-\infty < a < b < \infty$. The measure λ is called the Lebesgue measure on $\mathbb{R}$.

26. Integration

The aim is to define the integral

$$\int f \, d\mu = \int_{\Omega} f \, d\mu = \int_{\Omega} f(\omega) \, \mu(\omega)$$

of a real-valued measurable function f on a measurable space $(\Omega, \mathcal{F}, \mu)$. First, let f be positive simple function

$$f(\omega) = \sum_{j=1}^{n} a_j 1_{A_j}(\omega), \quad a_j \geq 0$$

and define the integral by

$$\int f\,d\mu = \sum_{j=1}^{n} a_j \mu(A_j).$$

This definition is consistent, in the sense that the integral is independent of the representation of f. For f and g positive simple functions we have the basic properties

$$\int (\alpha f + \beta g)\,d\mu = \alpha \int f\,d\mu + \beta \int g\,d\mu, \text{ for } \alpha, \beta \geq 0$$

$$\int f\,d\mu \leq \int g\,d\mu \text{ if } f \leq g.$$

Next, let f be a positive measurable function and by Lemma 25.12 there is a sequence of positive simple functions f_n with $f_n \uparrow f$. Since $\int f_n\,d\mu \leq \int f_{n+1}\,d\mu$ the limit exists and hence we define the integral $\int f\,d\mu$ by $\int f_n\,d\mu \uparrow \int f\,d\mu$. Again the definition is consistent in the sense that the limit is independent of choice of approximating sequence.

Lemma 26.1. *Let f be a positive measurable function. Let $f_1, f_2, \ldots$ and g be simple functions such that $f_n \uparrow f$ and $0 \leq g \leq f$. Then $\lim_n \int f_n\,d\mu \geq \int g\,d\mu$.*

Thus the definition of the integral is the following.

Definition 26.2. The integral of a positive measurable function f is given by

$$\int f\,d\mu = \lim_n \int f_n\,d\mu = \sup\{\int g\,d\mu | g \text{ positive simple function and } g \leq f\}$$

where f_n are positive simple functions with $f_n \uparrow f$.

Finally, consider the general case. A real-valued measurable function f is μ-integrable if $\int |f|\,d\mu = \int f^+\,d\mu + \int f^-\,d\mu < \infty$, where $f^\pm = (\pm f) \vee 0$. Thus f can be written as the difference of two positive integrable functions g and h ($f = f^+ - f^-$) and we define $\int f\,d\mu = \int g\,d\mu - \int h\,d\mu$. Again the extended integral is independent of choice of the representation.

Definition 26.3. A real valued measurable function f on $(\Omega, \mathcal{F}, \mu)$ is μ-integrable if

$$\int |f|\,d\mu = \int f^+\,d\mu + \int f^-\,d\mu < \infty$$

and the integral is given by

$$\int f\,d\mu = \int f^+\,d\mu - \int f^- d\mu.$$

Proposition 26.4. *Let f and g be μ-integrable functions. For $\alpha, \beta \in \mathbb{R}$, $\alpha f + \beta g$ is μ-integrable and*

$$\int (\alpha f + \beta g)\,d\mu = \alpha \int f\,d\mu + \beta \int g\,d\mu.$$

Proposition 26.5. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. Assume that μ is on the form $\mu = \alpha\mu_1 + \beta\mu_2$ where μ_1 and μ_2 are measures on $(\Omega, \mathcal{F})$ and $\alpha, \beta \geq 0$ are constants. If f is μ-integrable then*

$$\int f\,d(\alpha\mu_1 + \beta\mu_2) = \alpha \int f\,d\mu_1 + \beta \int f\,d\mu_2.$$

Let f be a measurable function and $A \in \mathcal{F}$ then we define the μ-integral of f over the set A by

$$\int_A f\,d\mu = \int 1_A f\,d\mu$$

assuming that $1_A f$ is μ-integrable.

Proposition 26.6. *Let $\varphi \geq 0$ be a positive measurable function on a measure space $(\Omega, \mathcal{F}, \mu)$ and set*

$$\nu(A) = \int_A \varphi\,d\mu, \quad A \in \mathcal{F}.$$

Then ν is a measure on $(\Omega, \mathcal{F})$ and if f is a real-valued measurable function then

$$\int f\,d\nu = \int (f\varphi)\,d\mu$$

whenever one of the integrals exists.

26.1. Product structures and Fubini's theorem. Let $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ be measurable spaces and let

$$\Omega_1 \times \Omega_2 = \{(\omega_1, \omega_2) | \omega_1 \in \Omega_1, \omega_2 \in \Omega_2\}$$

be the Cartesian product. For $j = 1, 2$, let $p_j : \Omega_1 \times \Omega_2 \to \Omega_j$ denote the jth coordinate map

$$p_j(\omega_1, \omega_2) = \omega_j.$$

Note that

$$p_1^{-1}(F_1) = F_1 \times \Omega_2, \quad F_1 \in \mathcal{F}_1$$
$$p_2^{-1}(F_2) = \Omega_1 \times F_2, \quad F_2 \in \mathcal{F}_2$$

and $p_1^{-1}(F_1) \cap p_2^{-1}(F_2) = (F_1 \times \Omega_2) \cap (\Omega_1 \times F_2) = F_1 \times F_2$.

Definition 26.7. The product σ-algebra on $\Omega_1 \times \Omega_2$ is defined by

$$\mathcal{F}_1 \otimes \mathcal{F}_2 = \sigma(p_1, p_2) = \sigma(F_1 \times F_2 | F_1 \in \mathcal{F}_1, F_2 \in \mathcal{F}_2).$$

Lemma 26.8. (Sections for sets). *Let $F \in \mathcal{F}_1 \otimes \mathcal{F}_2$ be a measurable set. Then*

- *for each (fixed and given) $\omega_1 \in \Omega_1$, the set $\{\omega_2 \in \Omega_2 | (\omega_1, \omega_2) \in F\} \in \mathcal{F}_2$ is measurable*
- *for each (fixed and given) $\omega_2 \in \Omega_2$, the set $\{\omega_1 \in \Omega_1 | (\omega_1, \omega_2) \in F\} \in \mathcal{F}_1$ is measurable.*

Before we state the basic condition for changing the order of integrations, we need a preliminary technical lemma

Lemma 26.9. (Sections for functions). *Let $f : \Omega_1 \times \Omega_2 \to \mathbb{R}$ be a $\mathcal{F}_1 \otimes \mathcal{F}_2$-measurable function. Then*

- *for each (fixed and given) $\omega_1 \in \Omega_1$, the map $\omega_2 \mapsto f(\omega_1, \omega_2)$ is $\mathcal{F}_2$-measurable on Ω_2*
- *for each (fixed and given) $\omega_2 \in \Omega_2$, the map $\omega_1 \mapsto f(\omega_1, \omega_2)$ is $\mathcal{F}_1$-measurable on Ω_1.*

Let μ_1 be a σ-finite measure on $(\Omega_1, \mathcal{F}_1)$ and μ_2 be a σ-finite measure on $(\Omega_2, \mathcal{F}_2)$. Then

- *the map $\omega_2 \mapsto \int_{\Omega_1} f(\omega_1, \omega_2)\,\mu_1(d\omega_1)$ is $\mathcal{F}_2$-measurable on Ω_2 if the integrals exists for all $\omega_2 \in \Omega_2$*
- *the map $\omega_1 \mapsto \int_{\Omega_2} f(\omega_1, \omega_2)\,\mu_2(d\omega_2)$ is $\mathcal{F}_1$-measurable on Ω_1 if the integrals exists for all $\omega_1 \in \Omega_1$.*

Theorem 26.10. (Fubini's theorem). *Let $(\Omega_1, \mathcal{F}_1, \mu_1)$ and $(\Omega_2, \mathcal{F}_2, \mu_2)$ be σ-finite measure spaces and let $f : \Omega_1 \times \Omega_2 \to [0, \infty)$ be a $\mathcal{F}_1 \otimes \mathcal{F}_2$-measurable function. Then*

$$\int_{\Omega_1} \Big[\int_{\Omega_2} f(\omega_1, \omega_2)\, \mu_2(d\omega_2) \Big]\, \mu_1(d\omega_1) = \int_{\Omega_2} \Big[\int_{\Omega_1} f(\omega_1, \omega_2)\, \mu_1(d\omega_1) \Big]\, \mu_2(d\omega_2).$$

If $f : \Omega_1 \times \Omega_2 \to \mathbb{R}$ is $\mathcal{F}_1 \otimes \mathcal{F}_2$-measurable, the relation remain valid if

$$\int_{\Omega_1} \Big[\int_{\Omega_2} |f(\omega_1, \omega_2)|\, \mu_2(d\omega_2) \Big]\, \mu_1(d\omega_1) = \int_{\Omega_2} \Big[\int_{\Omega_1} |f(\omega_1, \omega_2)|\, \mu_1(d\omega_1) \Big]\, \mu_2(d\omega_2) < \infty.$$

27. Conditional expectation

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space.

Definition 27.1. Let X be an integrable random variable and let $\mathcal{G} \subseteq \mathcal{F}$ be a sub σ-algebra of $\mathcal{F}$. The conditional expectation of X given $\mathcal{G}$ is defined to be any integrable random variable Y satisfying the properties

(i) Y is $\mathcal{G}$-measurable
(ii) for all $A \in \mathcal{G}$

$$\mathbf{E}[X; A] = \int_A X\, d\mathbf{P} = \int_A Y\, d\mathbf{P} = \mathbf{E}[Y; A].$$

The conditional expectation of X given $\mathcal{G}$ is denoted by $\mathbf{E}[X|\mathcal{G}]$.

Theorem 27.2. *Let X be an integrable random variable and let $\mathcal{G} \subseteq \mathcal{F}$ be a sub σ-algebra of $\mathcal{F}$. Then*

(i) *there exists a conditional expectation $\mathbf{E}[X|\mathcal{G}]$ of X given by $\mathcal{G}$.*
(ii) *any two conditional expectations of X given $\mathcal{G}$ are equal $\mathbf{P}$-a.s.*

Proposition 27.3. *Let X, Y be integrable random variables and let $\mathcal{G}, \mathcal{H} \subseteq \mathcal{F}$ be sub σ-algebras of $\mathcal{F}$. Then*

(i) $\mathbf{E}\big[\mathbf{E}[X|\mathcal{G}]\big] = \mathbf{E}[X]$.
(ii) *If $\alpha, \beta \in \mathbb{R}$ then $\mathbf{E}[\alpha X + \beta Y|\mathcal{G}] = \alpha\mathbf{E}[X|\mathcal{G}] + \beta\mathbf{E}[Y|\mathcal{G}]$.*
(iii) *If $X = a$ then $\mathbf{E}[X|\mathcal{G}] = a$.*
(iv) *If X is $\mathcal{G}$ then $\mathbf{E}[X|\mathcal{G}] = X$.*
(v) *If X is $\mathcal{G}$ and XY is integrable then $\mathbf{E}[XY|\mathcal{G}] = X\mathbf{E}[Y|\mathcal{G}]$.*
(vi) *If X is independent of $\mathcal{G}$ then $\mathbf{E}[X|\mathcal{G}] = \mathbf{E}[X]$.*
(vii) *If $\mathcal{G} = \mathcal{F}_* = \{\emptyset, \Omega\}$ (trivial σ-algebra) then $\mathbf{E}[X|\mathcal{G}] = \mathbf{E}[X]$.*
(viii) *If $(X, \mathcal{G})$ is independent of $\mathcal{H}$ then $\mathbf{E}[X|\sigma(\mathcal{G} \cup \mathcal{H})] = \mathbf{E}[X|\mathcal{G}]$.*
(ix) *If $\mathcal{H} \subseteq \mathcal{G}$ then $\mathbf{E}\big[\mathbf{E}[X|\mathcal{G}]\big|\mathcal{H}\big] = \mathbf{E}[X|\mathcal{H}]$.*

28. An application of the Monotone Class Theorem

In this section we have an application of Monotone class theorem with all details. We use the result in the Markov theory (see the proof of Proposition 7.2).

Proposition 28.1. *Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space and let $\mathcal{A}, \mathcal{G}_1, \mathcal{G}_2$ be sub-σ-algebras of $\mathcal{F}$. Assume that*

$$\mathbf{P}(A|\mathcal{G}_1) = \mathbf{P}(A|\mathcal{G}_2) \text{ for all } A \in \mathcal{A}.$$

Then for all $\mathcal{A}$-measurable random variables Z with $\mathbf{E}[|Z|] < \infty$ we have that

$$\mathbf{E}[Z|\mathcal{G}_1] = \mathbf{E}[Z|\mathcal{G}_2].$$

Proof. We apply the Monotone Class Theorem to verify the property for bounded Z. Let

$$\mathcal{H} = \{Z \text{ bounded } \mathcal{A}\text{-measurable random variable} \,|\, \mathbf{E}[Z|\mathcal{G}_1] = \mathbf{E}[Z|\mathcal{G}_2]\}.$$

(i): By linearity of conditional expectation, $\mathcal{H}$ is a vector space ($Z_1, Z_2 \in \mathcal{H}$ then $Z_1 + Z_2 \in \mathcal{H}$ and moreover $\alpha Z_1 \in \mathcal{H}$ for any $\alpha \in \mathbb{R}$).
(ii): $1_\Omega \in \mathcal{H}$ since $\mathbf{E}[1|\mathcal{G}] = 1$ for any σ-algebra $\mathcal{G}$.
(iii): For $Z_1 \leq Z_2 \leq \cdots$, $Z_n \to Z$ and Z bounded, then by monotone convergence theorem for conditional expectation we have that

$$\mathbf{E}[Z|\mathcal{G}_1] = \lim_{n\to\infty} \mathbf{E}[Z_n|\mathcal{G}_1] = \lim_{n\to\infty} \mathbf{E}[Z_n|\mathcal{G}_2] = \mathbf{E}[Z|\mathcal{G}_2].$$

Hence $Z \in \mathcal{H}$. Conclusion $\mathcal{H}$ is a MVS. The set $\mathcal{M} = \{1_A | A \in \mathcal{A}\}$ is multiplicative. Since

$$\mathbf{E}[1_A|\mathcal{G}_1] = \mathbf{P}(A|\mathcal{G}_1) = \mathbf{P}(A|\mathcal{G}_2) = \mathbf{E}[1_A|\mathcal{G}_2]$$

we have that $\mathcal{M} \subseteq \mathcal{H}$. Monotone Class Theorem implies that $\mathcal{H}$ contains all bounded $\mathcal{A}$-measurable random variables Z, that is, $\mathbf{E}[Z|\mathcal{G}_1] = \mathbf{E}[Z|\mathcal{G}_2]$.

Finally, let Z be a $\mathcal{A}$-measurable random variable with $\mathbf{E}[|Z|] < \infty$. Then

$$Z^+ \wedge n \uparrow Z^+ \text{ and } Z^- \wedge n \uparrow Z^- \text{ for } n \to \infty$$

where $Z^+ = Z \wedge 0$ is the positive part and $Z^- = (-Z) \wedge 0$ is the negative part. Recall that $Z = Z^+ - Z^-$. By monotone convergence theorem for conditional expectation and by the above result just verified we have that

$$\begin{aligned}
\mathbf{E}[Z|\mathcal{G}_1] &= \mathbf{E}[Z^+|\mathcal{G}_1] - \mathbf{E}[Z^-|\mathcal{G}_1] \\
&= \lim_{n\to\infty} \mathbf{E}[Z^+ \wedge n|\mathcal{G}_1] - \lim_{n\to\infty} \mathbf{E}[Z^- \wedge n|\mathcal{G}_1] \\
&= \lim_{n\to\infty} \mathbf{E}[Z^+ \wedge n|\mathcal{G}_2] - \lim_{n\to\infty} \mathbf{E}[Z^- \wedge n|\mathcal{G}_2] \\
&= \mathbf{E}[Z^+|\mathcal{G}_2] - \mathbf{E}[Z^-|\mathcal{G}_2] \\
&= \mathbf{E}[Z|\mathcal{G}_2].
\end{aligned}$$

$\square$

APPENDIX

Deriving Thiele integral equations

$$
\begin{aligned}
V^i(t) &= \sum_j \int_t^T e^{-\int_t^u r(z)\,dz} p^{ij}(t,u)\Big(dB^j(u) + \sum_{k:k\neq j} b^{jk}(u)\mu^{jk}(u)\,du\Big)\\
&= \int_t^T e^{-\int_t^u r(z)\,dz} e^{-\int_t^u \mu^i(z)\,dz}\Big(dB^i(u) + \sum_{k:k\neq i} b^{ik}(u)\mu^{ik}(u)\,du\Big)\\
&\quad + \sum_j \int_t^T e^{-\int_t^u r(z)\,dz}\Big[\int_t^u e^{-\int_t^v \mu^i(z)\,dz} \sum_{g:g\neq i} \mu^{ig}(v)p^{gj}(v,u)\,dv\Big]\Big(dB^j(u) + \sum_{k:k\neq j} b^{jk}(u)\mu^{jk}(u)\,du\Big)\\
&= \int_t^T e^{-\int_t^u ((r(z)+\mu^i(z))\,dz}\Big(dB^i(u) + \sum_{k:k\neq i} b^{ik}(u)\mu^{ik}(u)\,du\Big)\\
&\quad + \sum_j \int_t^T \Big[\int_t^u e^{-\int_t^u r(z)\,dz - \int_t^v \mu^i(z)\,dz} \sum_{g:g\neq i} \mu^{ig}(v)p^{gj}(v,u)\,dv\Big]\Big(dB^j(u) + \sum_{k:k\neq j} b^{jk}(u)\mu^{jk}(u)\,du\Big)\\
&= \int_t^T e^{-\int_t^u ((r(z)+\mu^i(z))\,dz}\Big(dB^i(u) + \sum_{k:k\neq i} b^{ik}(u)\mu^{ik}(u)\,du\Big)\\
&\quad + \sum_j \int_t^T \Big[\int_v^T e^{-\int_t^u r(z)\,dz - \int_t^v \mu^i(z)\,dz} \sum_{g:g\neq i} \mu^{ig}(v)p^{gj}(v,u)\Big(dB^j(u) + \sum_{k:k\neq j} b^{jk}(u)\mu^{jk}(u)\,du\Big)\Big]\,dv\\
&= \int_t^T e^{-\int_t^u ((r(z)+\mu^i(z))\,dz}\Big(dB^i(u) + \sum_{k:k\neq i} b^{ik}(u)\mu^{ik}(u)\,du\Big)\\
&\quad + \int_t^T e^{-\int_t^v ((r(z)+\mu^i(z))\,dz} \sum_{g:g\neq i} \mu^{ig}(v)\Big[\sum_j \int_v^T e^{-\int_v^u r(z)\,dz} p^{gj}(v,u)\Big(dB^j(u) + \sum_{k:k\neq j} b^{jk}(u)\mu^{jk}(u)\,du\Big)\Big]\,dv\\
&= \int_t^T e^{-\int_t^u ((r(z)+\mu^i(z))\,dz}\Big(dB^i(u) + \sum_{k:k\neq i} b^{ik}(u)\mu^{ik}(u)\,du\Big)\\
&\quad + \int_t^T e^{-\int_t^v ((r(z)+\mu^i(z))\,dz} \sum_{g:g\neq i} \mu^{ig}(v)V^g(v)\,dv\\
&= \int_t^T e^{-\int_t^u ((r(z)+\mu^i(z))\,dz}\Big(dB^i(u) + \sum_{k:k\neq i} \mu^{ik}(u)(b^{ik}(u) + V^k(u))\,du\Big).
\end{aligned}
$$

where we have used Kolmogorov backward equations given by

$$
p^{ij}(t,u) = \delta_{ij}\, e^{-\int_t^u \mu^i(z)\,dz} + \int_t^u e^{-\int_t^v \mu^i(z)\,dz} \sum_{g:g\neq i} \mu^{ig}(v)p^{gj}(v,u)\,dv.
$$

Bibliography

[1] Andersen, P.K., Borgan, Ø., Gill, R.D. & Keiding, N. (1993). *Statistical models based on counting processes*. Springer.

[2] Hoffmann-Jørgensen, J. (1994). *Probability with a view toward statistics. Vol. I.* Chapman & Hall.

[3] Jacobsen, M. (2006). *Point process theory and applications.* Birkhäuser.

[4] Jacod, J. & Shiryaev, A.N. (2003). *Limit theorems for stochastic processes.* (Second edition). Springer.

[5] Møller, T. & Steffensen, M. (2007). *Market-valuation methods in life and pension insurance.* Cambridge University Press.

[6] Norberg, R. (2000). *Basic life insurance mathematics.* Lecture notes, Laboratory of Actuarial Mathematics.

[7] Shiryaev, A.N. (1996). *Probability.* (Second edition). Springer.