LECTURE NOTES

MSF 200/ MVE 330 Stochastic Processes

3rd Quarter Spring 2010

By Patrik Albin

March 5, 2010

Lecture 1, Thursday 21 January

Chapter 6 Markov chains

6.1 Markov processes

Definition 6.1.1. *A family of integer valued random variables* $X = \{X_n\}_{n=0}^{\infty}$ *defined on a common probability space is a* Markov[1] chain *if*

$$\mathbf{P}\big\{X_{n+1} = i_{n+1} \big| X_n = i_n, \ldots, X_0 = i_0\big\} = \mathbf{P}\{X_{n+1} = i_{n+1} | X_n = i_n\}$$

for all $n \geq 0$ *and* $i_0, \ldots, i_{n+1} \in \mathbb{Z}$.

The Markov chain property means that the future X_{n+1} of the stochastic process X depends on the history of the process $X_n, \ldots, X_0$ only through the value (or *state*) of the process right now X_n (but not on how that value was obtained).

Markov chains may also take values in other countable sets (so called *state spaces*) than the integers, but when developing the theory it is enough to consider the integers.

Definition 6.1.4. *A Markov chain* X *is (time)* homogeneous *if*

$$\mathbf{P}\{X_{n+1} = j \,|\, X_n = i\} = \mathbf{P}\{X_1 = j \,|\, X_0 = i\}$$

does not depend on $n \geq 0$ *for any* $i, j \in \mathbb{Z}$.

We will (unless otherwise is explicitely stated) only deal with homogeneous Markov chains. Thus Markov chain will mean homogeneous Markov chain for us in the sequel.

Definition 6.1.4. *The* transition matrix $P = (p_{ij})$ *of a (homogeneous) Markov chain* X *is made up of the* transition probabilities

$$p_{ij} = \mathbf{P}\{X_{n+1} = j \,|\, X_n = i\} = \mathbf{P}\{X_1 = j \,|\, X_0 = i\}.$$

Theorem 6.1.5. *The transition matrix* P *is a stochastic matrix, which is to say that* $p_{ij} \geq 0$ *for all* i, j *and* $\sum_j p_{ij} = 1$ *for all* i.

Proof. Trivial. □

[1] Andrey Andreyevich Markov, Russian mathematician 1856-1922. Tried to develop the theory of stochastic processes.

Example 6.1.9. (Simple random walk) Let $X_n = \sum_{i=1}^n Y_i$ for $n \geq 0$ where $\{Y_i\}_{i=1}^\infty$ are iid. random variables with $\mathbf{P}\{Y_i = 1\} = 1 - \mathbf{P}\{Y_i = -1\} = p$. We have

$$\begin{aligned}\mathbf{P}\big\{X_{n+1} = j \big| X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0\big\} \\ &= \mathbf{P}\big\{Y_{n+1} = j - i \big| X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0\big\} \\ &= \mathbf{P}\{Y_{n+1} = j - i\} \\ &= \mathbf{P}\{Y_{n+1} = j - i \,|\, X_n = i\} \\ &= \mathbf{P}\{X_{n+1} = j \,|\, X_n = i\},\end{aligned}$$

so that X is a Markov chain with transition probabilities

$$p_{ij} = \mathbf{P}\{Y_{n+1} = j - i\} = \begin{cases} p & \text{for } j - i = 1, \\ 1 - p & \text{for } j - i = -1. \end{cases} \qquad \#$$

Definition 6.1.6. *The* n-th step transition matrix $P(n) = (p_{ij}(n))$ *is made up of the* n-th step transition probabilities

$$p_{ij}(n) = \mathbf{P}\{X_{m+n} = j \,|\, X_m = i\}.$$

Theorem 6.1.7. (Chapman-Kolmogorov[2]) $P(n) = P^n$.

Proof. We have

$$\begin{aligned}p_{ij}(n) &= \mathbf{P}\{X_{m+n} = j \,|\, X_m = i\} \\ &= \sum_k \mathbf{P}\{X_{m+n} = j, X_{m+1} = k \,|\, X_m = i\} \\ &= \sum_k \frac{\mathbf{P}\{X_{m+n} = j, X_{m+1} = k, X_m = i\}}{\mathbf{P}\{X_{m+1} = k, X_m = i\}} \, \frac{\mathbf{P}\{X_{m+1} = k, X_m = i\}}{\mathbf{P}\{X_m = i\}} \\ &= \sum_k p_{kj}(n-1)\, p_{ik} \\ &= (P\,P(n-1))_{ij},\end{aligned}$$

so that $P(n) = P\,P(n-1) = P\,P\,P(n-2) = \ldots = P^n$. □

Definition. *The* distribution *at time* n *is the row matrix* $\mu^{(n)} = (\mu_i^{(n)})$ *with entries*

$$\mu_i^{(n)} = \mathbf{P}\{X_n = i\}.$$

[2]Andrey Nikolaevich Kolmogorov, Russian mathematician 1903-1987. The most important probabilist ever by a wide margin.

Lemma 6.1.8. $\mu^{(m+n)} = \mu^{(m)} P^n$ *and in particular* $\mu^{(n)} = \mu^{(0)} P^n$.

Proof. We have

$$\begin{aligned}\mu_i^{(m+n)} &= \mathbf{P}\{X_{m+n} = i\} \\ &= \sum_k \mathbf{P}\{X_{m+n} = i \,|\, X_m = k\}\, \mathbf{P}\{X_m = k\} \\ &= \sum_k p_{ki}(n)\, \mu_k^{(m)} \\ &= (\mu^{(m)} P^n)_i.\end{aligned}$$

□

Example 6.1.12. (Bernoulli[3] process) Let $X_n = \sum_{k=1}^n Y_k \bmod 10$, where $\{Y_k\}_{k=1}^\infty$ are iid. Bernoulli distributed random variables with $\mathbf{P}\{Y_k = 1\} = 1 - \mathbf{P}\{Y_k = 0\} = p$. Then X can take values in $\{0, \dots, 9\}$ and P is the 10×10-matrix

$$P = \begin{pmatrix} 1-p & p & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1-p & p & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1-p & p & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1-p & p & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1-p & p & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1-p & p & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1-p & p & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1-p & p & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1-p & p \\ p & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1-p \end{pmatrix}. \quad \#$$

6.2 Classification of states

Definition 6.2.1. *The state (value)* i *of a Markov chain* X *is* persistent *(same as* reccurent*) if*

$$\mathbf{P}\{X_n = i \text{ for some } n \geq 1 \,|\, X_0 = i\} = 1.$$

States that are not persistent are transient.

Definition. *We write*

$$f_{ij}(n) = \mathbf{P}\big\{X_1 \neq j, \dots, X_{n-1} \neq j, X_n = j \big| X_0 = i\big\} \quad \text{for } n \geq 1$$

with the convention $f_{ij}(0) = 0$ *for all* i *and* j. *Further,* $f_{ij} = \sum_{n=1}^\infty f_{ij}(n)$.

[3] Jakob Bernoulli, Swiss mathematician 1654-1705. Wrote the first essential work on probability theory.

Corollary 6.2.4. **(a)** *A state j is persistent if (and only if) $\sum_{n=1}^{\infty} p_{jj}(n) = \infty$, and in that case $\sum_{n=1}^{\infty} p_{ij}(n) = \infty$ for all states i with $f_{ij} > 0$.*
(b) *A state j is transient if (and only if) $\sum_{n=1}^{\infty} p_{jj}(n) < \infty$, and in that case $\sum_{n=1}^{\infty} p_{ij}(n) < \infty$ for all states i.*

Proof. To prove (a) we employ the generating functions

$$P_{ij}(s) = \sum_{n=0}^{\infty} s^n p_{ij}(n) \quad \text{and} \quad F_{ij}(s) = \sum_{n=0}^{\infty} s^n f_{ij}(n) \quad \text{for } s \in [0,1).$$

Note that $\lim_{s\uparrow 1} P_{ij}(s) = \sum_{n=0}^{\infty} p_{ij}(n)$ and $\lim_{s\uparrow 1} F_{ij}(s) = f_{ij}$. Further, we have

$$p_{ij}(n) = \sum_{k=1}^{n} f_{ij}(k)\, p_{jj}(n-k) \quad \text{for } n \geq 1,$$

so that

$$\begin{aligned} P_{ij}(s) - \delta_{ij} &= \sum_{n=1}^{\infty} s^n p_{ij}(n) \\ &= \sum_{n=1}^{\infty} s^n \sum_{k=1}^{n} f_{ij}(k)\, p_{jj}(n-k) \\ &= \sum_{k=1}^{\infty} s^k f_{ij}(k) \sum_{n=k}^{\infty} s^{n-k} p_{jj}(n-k) \\ &= F_{ij}(s) P_{jj}(s). \end{aligned}$$

It follows that $\sum_{n=1}^{\infty} p_{jj}(n)$ is infinite if and only if $\lim_{s\uparrow 1} P_{jj}(s) = \lim_{s\uparrow 1} 1/(1 - F_{jj}(s)) = \infty$, which in turn holds if and only if $\lim_{s\uparrow 1} F_{jj}(s) = f_{jj}$ is 1. However, $f_{jj} = 1$ is equivalent with persistence. □

Exercise. Prove option of (b) of Corollary 6.2.4.

Corollary 6.2.5. *If j is transient then $p_{ij}(n) \to 0$ as $n \to \infty$ for all states i.*

Proof. Trivial consequence of Corollary 6.2.4. □

Example 6.2.12. For the simple random walk in Example 6.1.9 we have

$$p_{jj}(n) = \begin{cases} 0 & \text{if } n \text{ is odd,} \\ \binom{n}{n/2} p^{n/2} (1-p)^{n/2} & \text{if } n \text{ is even.} \end{cases}$$

We can check how $p_{jj}(n)$ behaves for large even n by means of Stirling's[4] formula. From this in turn we can check whether $\sum_{n=1}^{\infty} p_{jj}(n) = \infty$ or not (depending on the value of p) to find out whether the random walk is persistent or transient. #

[4]James Stirling, Scottish mathematician 1692-1770.

Exercise. Prove that a simple random walk is persistent if and only if it is symmetric ($p = 1/2$).

Definition 6.2.8. *The* mean reccurence time μ_i[5] *of a state i is defined*

$$\mu_i = \mathbf{E}\big\{ \inf\{n \geq 1 : X_n = i\} \,\big|\, X_0 = i \big\} = \begin{cases} \sum_{n=1}^{\infty} n\, f_{ii}(n) & \textit{if } i \textit{ is persistent,} \\ \infty & \textit{if } i \textit{ is transient.} \end{cases}$$

A persistent state i is non-null *if $u_i < \infty$ while i is* null *if $u_i = \infty$.*

Theorem 6.2.9. *A persistent state j is null if and only if $p_{jj}(n) \to 0$ as $n \to \infty$. In that case $p_{ij}(n) \to 0$ as $n \to \infty$ for all states i.*

Proof. Omitted! □

Example 6.2.12. (CONTINUED) A symmetric simple random walk is null. #

Exercise. Prove that a symmetric simple random walk is null.

Definition 6.2.10. *A state i has* period $d(i)$ *if*

$$d(i) \equiv gcd\{n : p_{ii}(n) > 0\} > 1.$$

Otherwise the state i is aperiodic.

Definition 6.2.11. *A state i is* ergodic *if it is persistent, non-null and aperiodic*[6]*.*

[5]Note the silly convention to use μ to denote two different things!

[6]I.e., 3 "good" properties out of 3.

Lecture 2, Friday 22 January

6.3 Classification of chains

Definition 6.3.1. *State i* communicates *with state j for a Markov chain X, denoted $i \to j$, if $p_{ij}(n) > 0$ for some $n \geq 0$. When i does not communicate with j we write $i \not\to j$. States i and j* intercommunicate*, denoted $i \leftrightarrow j$, when $i \to j$ and $j \to i$.*

Proposition. *Intercommunication $\leftrightarrow$ is an equivalence relation on the state space.*

Proof. Since it is obvious that $i \leftrightarrow i$ [as $p_{ii}(0) > 0$], it is enough to prove that $i \leftrightarrow j$ and $j \leftrightarrow k$ imply $i \leftrightarrow k$. However, when $i \leftrightarrow j$ and $j \leftrightarrow k$ we get $i \to k$ from the fact that $p_{ik}(m+n) \geq p_{ij}(m)\, p_{jk}(n) > 0$ whenever $p_{ij}(m), p_{jk}(n) > 0$, while $k \to i$ follows from the symmetric argument. □

Theorem 6.3.2. *If $i \leftrightarrow j$, then i and j have the same period (if any), i is transient if and only if j is, and i is null (non-null) persistent if and only if j is.*

Proof. We have

$$p_{ii}(k+m+n) \geq p_{ij}(k)\, p_{jj}(m)\, p_{ji}(n) = \alpha\, p_{jj}(m),$$

where $\alpha > 0$ for some k and n when $i \leftrightarrow j$. Hence

$$\sum_{m=1}^{\infty} p_{jj}(m) \leq \frac{1}{\alpha} \sum_{m=1}^{\infty} p_{ii}(k+m+n) \leq \frac{1}{\alpha} \sum_{m=1}^{\infty} p_{ii}(m) < \infty$$

when i is transient, so that also j is transient (recall Corollary 6.2.4). By the symmetric argument i is transient if j is. From the above we also see that $p_{jj}(n) \to 0$ as $n \to \infty$ if $p_{ii}(n) \to 0$ as $n \to \infty$, so that j is null persistent if i is. By the symmetric argument i is null persistent if j is. The statement about periods is proved in a similar way and is left as a difficult exercise. □

Exercise. (Difficult) Prove the statement about periods in Theorem 6.3.2.

Definition 6.3.3. *A set of states C is* closed *if $p_{ij} = 0$ whenever $i \in C$ and $j \notin C$. A set of states C is* irreducible *if $i \leftrightarrow j$ for all $i, j \in C$.*

Obviously, a chain which enters into a closed set of states will never leave that closed set (although in can start up outside it).

Theorem 6.3.4. (Decomposition theorem) *The set of possible values S of a Markov chain can be partioned as $S = T \cup C_1 \cup C_2 \cup \ldots$, where T are the transient states of the chain and $C_1, C_2, \ldots$ are disjoint irreducible closed sets of persistent states. The partition is unique except for the ordering of the closed sets.*

Proof. Let T be the transient states. Partition the set of all persistent states of the chain into irreducible sets $C_1, C_2, \ldots$ of persistent states that are equivalence classes of $\leftrightarrow$. This decomposition is possible by Theorem 6.3.2 together with basic properties of equivalence relations. It remains to show that the sets $C_1, C_2, \ldots$ are closed. If some C_k is not closed so that we can move out of it, then there exists an $i \in C_k$ and an $j \notin C_k$ such that $p_{ij} > 0$. We cannot have $p_{j\ell}(n) > 0$ for any $\ell \in C_k$, because this together with $i \leftrightarrow \ell$ would give $i \leftrightarrow j$, which contradicts the fact that $j \notin C_k$. Hence we can never come back to C_k from j, which in turn violates the fact that i is persistent, because then we can never come back to i either. Hence C_k is closed. □

Lemma 6.3.5. *If the set of possible values S of a Markov chain is finite, then at least one state is persistent and all persistent states are non-null.*

Proof. Pick an $i \in S$. As $\sum_{j\in S} p_{ij}(n) = 1$ for each $n \geq 0$, there must exist an $j \in S$ such that $p_{ij}(n) \not\to 0$ as $n \to \infty$. Therefore $\sum_{n=1}^{\infty} p_{ij}(n) = \infty$, so that Corollary 6.2.4 gives $\sum_{n=1}^{\infty} p_{jj}(n) = \infty$ (as finiteness of the latter sum implies finiteness of the former by that corollary). Hence j is persistent by Corollary 6.2.4.

Now consider a decomposition $S = T \cup C_1 \cup \cdots \cup C_m$ of the state space S of the chain where T are the transient states and $C_1, \ldots, C_m$ are disjoint irreducible closed sets of persistent states. Assume that some C_i has a null persistent state. As C_i is irreducible Theorem 6.3.2 then shows that all states in C_i are null persistent. By Theorem 6.2.9 we therefore have $p_{kj}(n) \to 0$ as $n \to \infty$ for all $j \in C_i$ for all $k \in S$. However, picking an $k \in C_i$, the fact that C_i is closed (and finite) then gives $1 = \sum_{j\in S} p_{kj}(n) = \sum_{j\in C_i} p_{kj}(n) \to 0$ as $n \to \infty$. As this is a contradiction it follows that all persistent states are non-null. □

Example 6.3.6. For the chain with state space $S = \{1, 2, 3, 4, 5, 6\}$ and transition matrix

$$P = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 & 0 \\ 1/4 & 3/4 & 0 & 0 & 0 & 0 \\ 1/4 & 1/4 & 1/4 & 1/4 & 0 & 0 \\ 1/4 & 0 & 1/4 & 1/4 & 0 & 1/4 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 \end{pmatrix}$$

we have $S = T \cup C_1 \cup C_2$, where $T = \{3, 4\}$ are transient aperiodic states and C_1 and C_2 are irreducible closed sets of ergodic states. #

6.4 Stationary distributions and the limit theorem

Definition 6.4.1. *A row matrix* π *is a* stationary distribution *for a Markov chain* X *if* π *is a probability distribution on the integers such that* $\pi P = \pi$.

Proposition. *If* $\mu^{(m)} = \pi$ *then* $\mu^{(m+n)} = \pi$ *for all* $n \geq 0$.

Proof. $\mu^{(m+n)} = \mu^{(m)} P^n = \mu^{(m)} P\, P^{n-1} = \pi\, P^n = \pi\, P^{n-1} = \ldots = \pi$. □

Theorem 6.4.3. *An irreducible chain has a stationary distribution* π *if and only if all states are non-null persistent. In that case* $\pi_i = 1/\mu_i$ *for each state* i.

Proof. Omitted! □

Example 6.1.12. (CONTINUED) The Bernoulli process is clearly irreducible and non-null persistent. We want to find the expectation $\mathbf{E}\{T_i\}$ of the time $T_i = \inf\{n \geq 1 : X_n = i\}$ it takes to the state $i \in \{1, \ldots, 9\}$, given that $X_0 = 0$. To that end we note that $\mathbf{E}\{T_i\} = \mu_i - 1$ where μ_i is the mean reccurance time for the modified chain with state space $\{0, \ldots, i\}$ and transistion matrix

$$P = \begin{pmatrix} 1-p & p & 0 & \cdots & 0 & 0 \\ 0 & 1-p & p & \cdots & 0 & 0 \\ 0 & 0 & 1-p & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1-p & p \\ 1 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix}.$$

As $\mu_i = 1/\pi_i$ (where π refers to the modified chain), we can determine μ_i by determining π. This in turn is done by solving the equation

$$\pi P = \pi \equiv \begin{cases} (1-p)\,\pi_0 + \pi_i &= \pi_0, \\ p\,\pi_{k-1} + (1-p)\,\pi_k &= \pi_k \quad \text{for } k = 1, \ldots, i-1, \\ p\,\pi_{i-1} &= \pi_i, \end{cases}$$

which boils down to

$$p\,\pi_{k-1} = p\,\pi_k \quad \text{for } k = 1, \ldots, i-1 \quad \text{and} \quad p\,\pi_0 = p\,\pi_{i-1} = \pi_i.$$

Hence we have

$$\pi = (\pi_0 \ \cdots \ \pi_0 \ p\,\pi_0) = \Big(\frac{1}{i+p} \ \cdots \ \frac{1}{i+p} \ \frac{p}{i+p}\Big),$$

giving

$$\mathbf{E}\{T_i\} = \mu_i = \frac{i+p}{p} - 1 = \frac{i}{p}.$$

It should be noted that this result can be established in an entirely different way making use of the fact that the waiting time in each state before transition to the next state for the Bernoulli process is geometrically distributed with expected value $1/p$. #

Theorem 6.4.17. *For an irreducible aperiodic chain we have* $p_{ij}(n) \to 1/\mu_j$ *as* $n \to \infty$ *for all states* i *and* j.

Proof. Omitted! □

Lecture 3, Thursday 28 January

6.8 Birth processes and the Poisson process

This section belongs to the course but is covered by the treatment of Section 6.9.

6.9 Continuous-time Markov chains

Definition 6.9.1. *A family of integer valued random variables* $X = \{X(t)\}_{t\geq 0}$ *defined on a common probability space is a* Markov chain *if*

$$\mathbf{P}\big\{X(t_{n+1}) = i_{n+1}\big|X(t_n) = i_n, \ldots, X(t_1) = i_1\big\} = \mathbf{P}\{X(t_{n+1}) = i_{n+1}|X(t_n) = i_n\}$$

for all $0 \leq t_1 \leq \ldots \leq t_n \leq t_{n+1}$ *and* $i_1, \ldots, i_{n+1} \in \mathbb{Z}$.

A continuous-time Markov chain may take values in any countable state space, but when developing the theory it is enough to let the state space be the integers.

Definition. *A Markov chain* $\{X(t)\}_{t\geq 0}$ *is (time)* homogeneous *if*

$$\mathbf{P}\{X(t+s) = j\,|\,X(s) = i\} = \mathbf{P}\{X(t) = j\,|\,X(0) = i\}$$

does not depend on $s \geq 0$ *but only on* $t \geq 0$ *for all* $i, j \in \mathbb{Z}$.

We will (unless otherwise is explicitely stated) only deal with homogeneous Markov chains. Thus Markov chain will mean homogeneous Markov chain for us in the sequel.

Definition 6.9.2. *The* transition matrices $\{P_t\}_{t\geq 0} = \{(p_{ij}(t))\}_{t\geq 0}$ *of a (homogeneous) Markov chain* $\{X(t)\}_{t\geq 0}$ *is made up of the* transition probabilities

$$p_{ij}(t) = \mathbf{P}\{X(t+s) = j\,|\,X(s) = i\} = \mathbf{P}\{X(t) = j\,|\,X(0) = i\}$$

for $s, t \geq 0$ *and* $i, j \in \mathbb{Z}$.

Theorem 6.9.3. *The family of transition matrices* $\{P_t\}_{t\geq 0}$ *is a* stochastic semigroup, *which is to say that*
(a) $P_0 = I$ *(the identity matrix);*
(b) P_t *is a stochastic matrix (cf. Theorem 6.1.5), which is to say that all* $p_{ij}(t) \geq 0$ *and* $\sum_j p_{ij}(t) = 1$ *for all* $i \in \mathbb{Z}$ *and* $t \geq 0$;
(c) *the* Chapman-Kolmogorov equations $P_{s+t} = P_s P_t$ *holds for* $s, t \geq 0$.

Proof. The properties (a) and (b) are immediate, while the property (c) follows as (cf. the proof of Theorem 6.1.7)

$$\begin{aligned}(P_{s+t})_{ij} &= \mathbf{P}\{X(s+t)=j\,|\,X(0)=i\}\\ &= \sum_k \mathbf{P}\{X(s+t)=j, X(s)=k\,|\,X(0)=i\}\\ &= \sum_k \frac{\mathbf{P}\{X(s+t)=j, X(s)=k, X(0)=i\}}{\mathbf{P}\{X(s)=k, X(0)=i\}}\,\frac{\mathbf{P}\{X(s)=k, X(0)=i\}}{\mathbf{P}\{X(0)=i\}}\\ &= \sum_k p_{kj}(t)\,p_{ik}(s)\\ &= (P_sP_t)_{ij}. \qquad \square\end{aligned}$$

We will henceforth assume that the transition semigroup $\{P_t\}_{t\geq 0}$ is element wise differentiable from the right at $t=0$. The corresponing matrix of derivatives

$$G = \lim_{t\downarrow 0}\frac{P_t - I}{t}$$

is called the *generator* of the continuous-time Markov chain and takes over the role of the transition matrix P for discrete-time chains.

Theorem 6.9.5. *We may express $P_t = (p_{ij}(t))$ in terms of the generator $G = (g_{ij})$ as*

$$p_{ij}(h) = \begin{cases} g_{ij}h + o(h) & \text{as } h\downarrow 0 \quad \text{for } i\neq j,\\ 1 + g_{jj}h + o(h) & \text{as } h\downarrow 0 \quad \text{for } i=j.\end{cases}$$

Proof. This is just another way to phrase the fact that $G = \lim_{t\downarrow 0}(P_t - I)/t$. $\square$

Note that we must have

$$\boxed{g_{ij}\geq 0 \quad \text{for } i\neq j \quad \text{while} \quad g_{jj}\leq 0}\,.$$

Furthermore, the following formal calculation

$$1 = \sum_j p_{ij}(h) = 1 + g_{jj}h + o(h) + \sum_{j\neq i}\big(g_{ij}h + o(h)\big) = 1 + \big(\sum_j g_{ij}\big)h + o(h)$$

gives

$$\boxed{\sum_j g_{ij} = 0}\,.$$

However, occasionally this formula may fail, which is explained by the fact that the formal calculation uses a change of order of limit operation that might not be correct when $o(h)$ is moved outside the sum. We will resolve this issue by restricting our attention to continuous-time Markov chains for which the above kind of formal arguments apply (which happens to be the case for virtually all chains of any interest).

Example 6.9.7. (Birth process) A birth process has state space $S = \mathbb{N}$ and generator

$$G = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & 0 & \cdots \\ 0 & -\lambda_1 & \lambda_1 & 0 & 0 & \cdots \\ 0 & 0 & -\lambda_2 & \lambda_2 & 0 & \cdots \\ 0 & 0 & 0 & -\lambda_3 & \lambda_3 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

A particular case of a birth process is a Poisson process for which $\lambda_i = \lambda$ for all $i \in \mathbb{N}$, for some $\lambda > 0$. In addition, $\mu^{(0)} = (1 \ 0 \ 0 \ \ldots)$ for the Poisson process, while no such restriction applies to a more general birth process. #

Theorem 6.9.9. (Kolmogorov forward equations) *$P_t' = P_t G$ for $t \geq 0$, where $'$ denotes right derivative.*

Proof. By the Chapman-Kolmogorov equations we have

$$\frac{P_{t+h} - P_t}{h} = P_t \frac{P_h - P_0}{h} = P_t \frac{P_h - I}{h} \to P_t G \quad \text{as } h \downarrow 0. \qquad \square$$

By completely analogous methods we prove

Theorem 6.9.10. (Kolmogorov backward equations) *$P_t' = G P_t$ for $t \geq 0$.*

Definition. *For an entire analytic function $\mathbb{C} \ni z \curvearrowright f(z) = \sum_{k=0}^{\infty} a_k z^k$ and a square matrix M we define the matrix function $f(M) = \sum_{k=0}^{\infty} a_k M^k$.*

Exercise. (Difficult) Prove that the above definition makes rigorous sense in terms of elementwise convergence.

Proposition. (a) *We have $e^{M_1+M_2} = e^{M_1} e^{M_2}$ for square matrices M_1 and M_2.*
(b) *We have $\lim_{h\downarrow 0}(e^{(t+h)M} - e^{tM})/h = M e^{tM} = e^{tM} M$ for a square matrix M.*

Proof. (a) This one is a classic:

$$e^{M_1+M_2} = \sum_{k=0}^{\infty} \frac{(M_1+M_2)^k}{k!} = \sum_{k=0}^{\infty} \sum_{\ell=0}^{k} \binom{k}{\ell} \frac{M_1^{\ell} M_2^{k-\ell}}{k!} = \sum_{\ell=0}^{\infty} \sum_{k=\ell}^{\infty} \frac{M_1^{\ell} M_2^{k-\ell}}{\ell!\,(\ell-k)!} = e^{M_1} e^{M_2}.$$

(b) By (a) we have

$$\frac{e^{(t+h)M}-e^{tM}}{h}=\frac{e^{(h+t)M}-e^{tM}}{h}=\frac{e^{tM}(e^{hM}-I)}{h}=\frac{(e^{hM}-I)\,e^{tM}}{h}\to e^{tM}M=M\,e^{tM}$$

as $h\downarrow 0$ since $(e^{hM}-I)/h=\sum_{k=1}^{\infty}h^{k-1}M^k/k!$. □

Theorem 6.9.12. $P_t=e^{tG}$ *for* $t\geq 0$.

Proof. By the Kolmogorov backward equations and the above proposition we have

$$0=P_t'-GP_t=e^{tG}\frac{d}{dt}e^{-tG}P_t \implies \frac{d}{dt}e^{-tG}P_t=0 \implies e^{-tG}P_t=I \implies P_t=e^{tG}. \quad \square$$

In view of Theorem 6.9.12 a continuous-time Markov chain is completely specified by (the state space,) the start distribution $\mu^{0)}$ together with the generator G, recall Example 6.9.7.

Theorem 6.9.13. *The time* $T_i=\inf\{t>0: X(t)\neq i\,|\,X(0)=i\}$ *spent in state* i *for a continuous-time Markov chain is exponetially distributed with parameter* $-g_{ii}$.

Proof. For the non-increasing function

$$f(t)=\mathbf{P}\{X(r)=i \text{ for } r\in(0,t]\,|\,X(0)=i\}$$

the Markov property and time homogenity give

$$\begin{aligned}\frac{f(t+s)}{f(s)}&=\frac{\mathbf{P}\{X(r)=i \text{ for } r\in(0,t+s]\,|\,X(0)=i\}}{\mathbf{P}\{X(r)=i \text{ for } r\in(0,s]\,|\,X(0)=i\}}\\&=\frac{\mathbf{P}\{X(r)=i \text{ for } r\in[0,t+s]\}}{\mathbf{P}\{X(r)=i \text{ for } r\in[0,s]\}}\\&=\mathbf{P}\big\{X(r)=i \text{ for } r\in(s,t+s]\}\big|X(r)=i \text{ for } r\in[0,s]\big\}\\&=\mathbf{P}\{X(r)=i \text{ for } r\in(s,t+s]\}\,|\,X(s)=i\}\\&=f(t)\quad \text{ for } s,t\geq 0.\end{aligned}$$

Hence the non-decreasing function $g(t)=-\log(f(t))$ satisfies the so called *Cauchy functional equation* $g(t+s)=g(t)+g(s)$ for $s,t\geq 0$, the only solutions of which take the form $g(t)=C\,t$ for some constant $C\in\mathbb{R}$.

Now note that for small $h>0$ and a grid $0=h_0<h_1<\ldots<h_n=h$ such that $\max_{1\leq i\leq n}h_i-h_{i-1}\to 0$ as $n\to\infty$ we have

$$\begin{aligned}\mathbf{P}\{X(r)=i \text{ for } r\in(0,h]\,|\,X(0)=i\}&\leftarrow\mathbf{P}\big\{\cap_{i=1}^n X(h_i)=i\big|X(0)=i\big\}\\&=\prod_{i=1}^n p_{ii}(h_i-h_{i-1})\end{aligned}$$

$$= \prod_{i=1}^{n} \big(1 + g_{ii}(h_i - h_{i-1}) + o(h_i - h_{i-1})\big)$$
$$= 1 + g_{ii} \sum_{i=1}^{n} (h_i - h_{i-1}) + o(h)$$
$$= 1 + g_{ii}h + o(h).$$

Sending $h \downarrow 0$ we conclude that

$$g(h) = -\log\big(\mathbf{P}\{X(r) = i \text{ for } r \in (0,h] \,|\, X(0) = i\}\big) = -g_{ii}h + o(h).$$

Hence we must have $C = -g_{ii}$, so that $f(t) = \mathrm{e}^{-g(t)} = \mathrm{e}^{g_{ii}t}$. □

The converse to Theorem 6.9.13 also holds, which is to say that a continuous-time integer valued stochastic process that spends independent exponentially distributed times in its states (where the parameter of the exponential distribution only depends on the state) is a Markov chain.

Theorem 6.9.14. *For a continuous-time Markov chain we have*

$$\mathbf{P}\{\textit{next value of } X(t) \textit{ is } j \,|\, X(0) = i\} = \frac{g_{ij}}{-g_{ii}} \quad \textit{for } i \neq j.$$

Proof. Consider a grid $0 = t_0 < t_1 < t_2 < \ldots < t_n \to \infty$ as $n \to \infty$ that becomes finer and finer so that in the limit $\max_{i \geq 1} t_i - t_{i-1} \to 0$. As the grid becomes finer and finer we then have by Theorem 6.9.13 together with a Riemann sum argument

$$\begin{aligned}
&\mathbf{P}\{\text{next value of } X(t) \text{ is } j \,|\, X(0) = i\} \\
&\quad \leftarrow \sum_{n=1}^{\infty} \mathbf{P}\big\{X(t_n) = j, \cap_{i=0}^{n-1}\{X(s) = i\}\big| X(0) = i\big\} \\
&\quad = \sum_{n=1}^{\infty} \mathbf{P}\{X(t_n) = j \,|\, X(t_{n-1}) = i\}\, \mathbf{P}\big\{ \cap_{i=0}^{n-1} \{X(s) = i\}\big| X(0) = i\big\} \\
&\quad \approx \sum_{n=1}^{\infty} p_{ij}(t_n - t_{n-1})\, \mathrm{e}^{g_{ii}t_{n-1}} \\
&\quad \approx \sum_{n=1}^{\infty} g_{ij}\,(t_n - t_{n-1})\, \mathrm{e}^{g_{ii}t_{n-1}} \\
&\quad \to \int_0^{\infty} g_{ij}\, \mathrm{e}^{g_{ii}t}\, dt \\
&\quad = \frac{g_{ij}}{-g_{ii}}.
\end{aligned}$$
□

Note that Theorem 6.9.13 together with Theorem 6.9.14 give a method to simulate a continuous-time Markov chain with a given generator G: Just start the chain according to the appropriate initial distribution $\mu^{(0)}$, and then proceed step wise using Theorem 6.9.13 to selcet the times at which to switch values and Theorem 6.9.14 to select what values to switch to at those times.

Example. Consider a reliability system consisting of two iid. components with exponentially distributed life times with parameter $\lambda > 0$. The system is started at time $t = 0$ with two functioning components $X(0) = 2$. Then there is a waiting time equals the minimum of two iid. exponentially distributed life times with parameter λ, which is to say an exponentially distributed random waiting time with parameter 2λ, until the system changes to one functioning component $X(t) = 1$. Then, by the lack of memory proprty for exponential distributions, there is a final exponentially distributed random waiting time with parameter λ until the system changes to its terminal state with no functioning components $X(t) = 0$. The process has state space $S = \{0, 1, 2\}$, starting distribution $\mu^{(0)} = (0\ \ 0\ \ 1)$ and generator

$$G = \begin{pmatrix} 0 & 0 & 0 \\ \lambda & -\lambda & 0 \\ 0 & 2\lambda & -2\lambda \end{pmatrix}. \qquad \#$$

Definition 6.9.17. *A continuous-time Markov chain is* irreducible *if $p_{ij}(t) > 0$ for some $t \geq 0$ for every pair of states i, j.*

Theorem 6.9.16. *For an irreducible chain we have $p_{ij}(t) > 0$ for all $t > 0$ for every pair of states i, j.*

Proof. By Theorem 6.9.13 we have

$$p_{ii}(t) \geq \mathbf{P}\{X(r) = i \text{ for } r \in (0, t]\,|X(0) = i\} = \mathrm{e}^{g_{ii}t} > 0$$

for some $t > 0$. Now, as the chain is irreducible we have $p_{ij}(t) > 0$ for some $t > 0$ for $i \neq j$ (since $p_{ij}(0) = 0$). Hence Theorem 6.9.13 shows that there must exist states $i \neq k_1 \neq \ldots \neq k_n \neq j$ such that $g_{ik_1}, g_{k_1k_2}, \ldots, g_{k_nj} > 0$. For all sufficiently small $h > 0$ we therefore have $p_{ij}((n+1)h) \approx g_{ik_1}h\, g_{k_1k_2}h \cdot \ldots \cdot g_{k_nj}h > 0$. From this in turn we get $p_{ij}(t) \geq p_{ij}((n+1)h)\, p_{jj}(t-(n+1)h)$ for all $t > 0$ (picking $h > 0$ samll enough). □

Definition 6.9.18. *A row matrix π is a* stationary distribution *for a continuous-time Markov chain if $\pi P_t = \pi$ for all $t \geq 0$.*

Theorem 6.9.19. *A row matrix π is a stationary distribution if and only if $\pi G = 0$.*

Proof. We have $\pi P_t = \sum_{k=0}^{\infty} \pi t^k G^k / k! = \pi$ if $\pi G = 0$. If on the other hand $\pi P_t = \pi$

for all $t \geq 0$, then we must have $\sum_{k=1}^{\infty} \pi t^k G^k / k! = 0$ for all $t \geq 0$, which can only happen (e.g, by differetiating at $t=0$) if $\pi G = 0$. □

We have the following important continuous-time version of the discrete-time limit Theorem 6.4.17:

Theorem 6.9.21. *For an irreducible chain we have* $\lim_{t\to\infty} p_{ij}(t) = \pi_j$ *for all states* i *and* j *if there exists a stationary distribution* π*. Otherwise we have* $\lim_{t\to\infty} p_{ij}(t) = 0$ *for all states* i *and* j*.*

Proof. Omitted! □

6.11 Birth-death processes (but no imbedding)

Example. (Birth-death process) A birth-death process is the general form of a continuous-time Markov chain with state space $S = \mathbb{N}$ that can only change its values one unit at a time (upwards or downwards). In other words the generator is given by

$$G = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & 0 & \cdots \\ \mu_1 & -(\lambda_1+\mu_1) & \lambda_1 & 0 & 0 & \cdots \\ 0 & \mu_2 & -(\lambda_2+\mu_2) & \lambda_2 & 0 & \cdots \\ 0 & 0 & \mu_3 & -(\lambda_3+\mu_3) & \lambda_3 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \qquad \#$$

Example 6.11.6. (Simple death with immigration) A birth-death process with death rate $\mu_i = i\,\mu$ proportional to the current state i and constant birth rate $\lambda_i = \lambda$. #

Example. To find the stationary distribution π for a birth-death process (if it exists) we use Theorem 6.9.19 and note that the equation $\pi\, G = 0$ takes the form

$$\begin{cases} -\pi_0\lambda_0 + \pi_1\mu_1 = 0 \\ \pi_{i-1}\lambda_{i-1} - \pi_i(\mu_i+\lambda_i) + \pi_{i+1}\lambda_{i+1} = 0 \quad \text{for } i \geq 1 \end{cases} \Leftrightarrow \pi_i = \frac{\lambda_0 \ldots \lambda_{i-1}}{\mu_1 \ldots \mu_i}\pi_0 \quad \text{for } i \geq 1.$$

We see that this solution is a probability distribution if and only if

$$\sum_{i=1}^{\infty} \frac{\lambda_0 \ldots \lambda_{i-1}}{\mu_1 \ldots \mu_i} < \infty \quad \text{and} \quad \pi_0 = 1 \Big/ \bigg(1 + \sum_{i=1}^{\infty} \frac{\lambda_0 \ldots \lambda_{i-1}}{\mu_1 \ldots \mu_i}\bigg). \qquad \#$$

Example 6.11.6. (CONTINUED) For the birth-death process with simple death and immigration we have the stationary distribution is Poisson distributed with parameter λ/μ. #

Lecture 4, Friday 29 January

Chapter 8 Random processes

8.1 Introduction

Definition. *Given a sample space* Ω *a* stochastic process $X = \{X(t)\}_{t\in T}$ *with* parameter set T *is a function* $X : \Omega\times T \to \mathbb{R}$ *such that* $X(\cdot,t) : \Omega \to \mathbb{R}$ *is a random variable for each* $t\in T$.

Note that all memebers $X(t)$, $t \in T$, of a stochastic process should be random variables defined on a common sample space.

The dependence of $\omega \in \Omega$ for a stochastic process X is usually supressed in the notation (as we did already in the definition), so that we write $X(t)$ or $\{X(t)\}_{t\in T}$ instead of $X(\omega,t)$ or $\{X(\omega,t)\}_{(\omega,t)\in\Omega\times T}$.

It is natural to belive that a stochastic process is more or less "determined" by its univariate marginal distributions $F_{X(t)}(x) = \mathbf{P}\{X(t) \le x\}$ for $x \in \mathbb{R}$, for each $t \in T$. But this is not true at all as the following exercise demonstrates:

Exercise. Consider the stochastic process $X(t)=\xi$ for $t\in\mathbb{R}$, where ξ is a single standard normal $\mathrm{N}(0,1)$ distributed random variable. Let $Y(t)$ be a stochastic process that is $\mathrm{N}(0,1)$ distributed at each $t\in\mathbb{R}$, but with all random values of the process at different times independent of each other. Find the univariate marginal distributions $F_{X(t)}$ and $F_{Y(t)}$. Pick a "typical" $\omega \in \Omega$ and, for that choice of ω, plot a likely appearance of the graphs, the so called *realisations*,

$$\mathbb{R}\ni t \curvearrowright X(t)=X(\omega,t)\in\mathbb{R} \quad \text{and} \quad \mathbb{R}\ni t \curvearrowright Y(t)=Y(\omega,t)\in\mathbb{R}.$$

The univariate marginal distributions do not give any information at all about the dependence structure of the process at hand. To get such information we consider

Definition. *The* finite dimensional distributions (fidi's) $\{F_{X(t_1),\ldots,X(t_n)} : t_1,\ldots,t_n \in T, n\in\mathbb{N}\}$ *of a stochastic process* $\{X(t)\}_{t\in T}$ *are given by*

$$F_{X(t_1),\ldots,X(t_n)}(x_1,\ldots,x_n) = \mathbf{P}\{X(t_1)\le x_1,\ldots,X(t_n)\le x_n\} \quad \textit{for } x_1,\ldots,x_n\in\mathbb{R}.$$

Although the fidi's of a process usually offer sufficient information about the process for probabilistic analysis, it might happen that not even all this information is sufficient as the following exercise demonstrates:

Exercise. Find two stochastic processes $\{X(t)\}_{t\in[0,1]}$ and $\{Y(t)\}_{t\in[0,1]}$ that have common fidi's but that satisfy

$$\mathbf{P}\{X(t)\neq Y(t)\ \text{ for some }\ t\in[0,1]\}=1.$$

8.2 Stationary processes

Definition 8.2.1. *A stochastic process* $\{X(t)\}_{t\in T}$ *with* $T\subseteq\mathbb{R}$ *is* strongly stationary *if*

$$\big(X(t_1+h),\ldots,X(t_n+h)\big)=_D\big(X(t_1),\ldots,X(t_n)\big)$$

for all $t_1+h,\ldots,t_n+h,t_1,\ldots,t_n\in T$ *(where* $=_D$ *denotes equality of distributions).*

Definition 8.2.2. *A stochastic process* $\{X(t)\}_{t\in T}$ *with* $T\subseteq\mathbb{R}$ *is* weakly stationary *if the function* $T\ni t\curvearrowright\mathbf{E}\{X(t)\}$ *is a (finite) constant and if the function* $T\times T\ni(s,t)\curvearrowright\mathbf{Cov}\{X(s),X(t)\}$ *(is finite and) only depends on the difference* $t-s$ *between* s *and* t*.*

Grimmett and Stirzaker adopt the practice to understand stationary (without any additional information) as weakly stationary. In doing so they go against common practice which instead is to understand stationary as strongly stationary. However, as we use their book we have to follow their practice ...

Example 8.2.8. An iid. sequence of random variables $\{X_n\}_{n\in\mathbb{Z}}$ is quite obviously strongly stationary. #

Exercise. Show that strongly stationary processes are stationary when their expectations and covariances are well-defined.

Exercise. Exemplify that strongly stationary processes need not necessarily be stationary.

Exercise. Exemplify that stationary processes need not necessarily be strongly stationary.

Example 8.2.4. A continuous-time Markov chain that has stationary distribution π is strongly stationary if $\mu^{(0)}=\pi$. This is so since we then have $\mu^{(h)}=\pi$ for all $h\geq 0$, so that

$$\mathbf{P}\big\{X(t_1+h)\leq x_1,\ldots,X(t_n+h)\leq x_n\big\}$$

$$= \sum_k \mathbf{P}\big\{X(t_1+h) \le x_1, \ldots, X(t_n+h) \le x_n \big| X(h) = k\big\}\, \mu_k^{(h)}$$
$$= \sum_k \sum_{k_1 \le x_1} \cdots \sum_{k_n \le x_n} p_{kk_1}(t_1)\, p_{k_1k_2}(t_2-t_1) \cdot \ldots \cdot p_{k_{n-1}k_n}(t_n-t_{n-1})\, \pi_k,$$

where the right-hand side does not depend on $h \ge 0$. #

Example 8.2.5. For A and B uncorrelated zero-mean unit variance random variables and $\lambda \in \mathbb{R}$ a constant the *cosine process*

$$X(t) = A\cos(\lambda t) + B\sin(\lambda t), \quad t \in \mathbb{R},$$

is stationary since $\mathbf{E}\{X(t)\} = 0$ is a constant and

$$\begin{aligned}\mathbf{Cov}\{X(s), X(t)\} = \mathbf{Var}\{A\}\cos(\lambda s)\cos(\lambda t) & \\ + \mathbf{Cov}\{A, B\}\cos(\lambda s)\sin(\lambda t) & \\ + \mathbf{Cov}\{B, A\}\sin(\lambda s)\cos(\lambda t) & \\ + \mathbf{Var}\{B\}\sin(\lambda s)\sin(\lambda t) & \\ = \cos(\lambda s)\cos(\lambda t) + \sin(\lambda s)\sin(\lambda t) & \\ = \cos(\lambda(t-s)) &\end{aligned}$$

only depends on the difference $t-s$ between s and t. #

8.3 Renewal processes

Definition 8.3.3. *A* renewal process $N = \{N(t)\}_{t \ge 0}$ *is given by*

$$N(t) = \max\{n \in \mathbb{N} : T_n \le t\} \quad \textit{for } t \ge 0,$$

where $T_n = \sum_{i=1}^n X_i$ *and* $\{X_i\}_{i=1}^\infty$ *is an iid. sequence of non-negative random variables.*

Example 8.3.1. A room is lit by a single light bulb with a random life time X_1. When the bulb fails it is replaced immediately with an iid. copy with life time X_2, and so on .... Then $T_n = \sum_{i=1}^n X_i$ is the time to failure of the n-th bulb, and the renewal process $N(t) = \max\{n \in \mathbb{N} : T_n \le t\}$ is the number of bulbs that have failed up to time t. #

Example. A renewal process is a Poisson process if and only if the X_i's are exponentially distributed. #

Example 8.3.2. Let $\{Y_n\}_{n=0}^\infty$ be a discrete-time Markov chain with $Y_0 = i$ for some non-random $i \in \mathbb{Z}$ and define $T_0 = 0$ and

$$T_n = \inf\{n > T_{n-1} : Y_n = i\} \quad \text{for } n \geq 1.$$

Then $X_n = T_n - T_{n-1}$, $n \geq 1$, are iid. non-negative random variables so that the number of revisits $N(t) = \max\{n \in \mathbb{N} : T_n \leq t\}$ to the state i up to time t is a renewal process. #

Theorem 8.3.5. *A renewal process is a Markov chain if and only if it is a Poisson process.*

Proof. This is Exercise 8.3.5. □

8.4 Queues

A queue process $\{Q(t)\}_{t\geq 0}$ counts the total number of customers in a queuing system at time $t \geq 0$. The probabilistic behaviour of the queue process will depend on factors such as

- in what manner do customers arrive to the queuing system?
- what is the queuing disciplin for customers that have arrived to the queue and are waiting to be served?
- how long are service times for customers?

The simplest kind of queues are birth-and-death continuous-time Markov process. More complicated queues, albeit not Markov, may be analyzed through an imbedded discrete-time Markov chain. The most complicated queues are more complicated still.

8.5 The Wiener process

Definition. *A* Lévy[7] process *is a stochastic process* $\{X(t)\}_{t\geq 0}$ *with* $X(0) = 0$ *such that*

$$X(t_1) - X(t_0), \ldots, X(t_n) - X(t_{n-1}) \text{ are independent random variables}$$

for $0 \leq t_0 < t_1 < \ldots < t_{n-1} < t_n$, *that is,* independent increments, *and*

$$\text{the distribution of } X(t+s) - X(s) \text{ does not depend on } s \geq 0 \text{ but only on } t \geq 0,$$

that is, stationary increments.

[7] Paul Pierre Lévy 1886-1971, very important French probabilist.

Definition. *A Lévy process* $\{W(t)\}_{t\geq 0}$ *the increments* $W(t+s)-W(s)$ *of which are zero-mean normal distributed is called a* Wiener[8] process or Brown[9]ian motion.

The Wiener process is the most important stochastic process in the world.

Theorem. *The vector of process values* $(W(t_1),\ldots,W(t_n))$ *of a Wiener process has a zero-mean multivariate normal distribution with covariance matrix*

$$\mathbf{Cov}\{W(t_i),W(t_j)\} = \mathbf{Var}\{W(1)\}\min\{t_i,t_j\} \quad \textit{for } i,j=1,\ldots,n.$$

Proof. Assume without loss of generality that $0\leq t_1\leq \ldots \leq t_n$. We have to check that the linear combination $\sum_{i=1}^n a_i W(t_i)$ is univariate normal distributed for any choice of constants $a_1,\ldots,a_n\in\mathbb{R}$. To that end we note that

$$\sum_{i=1}^n a_i W(t_i) = \sum_{i=1}^n \Big(\sum_{j=i}^n a_j\Big)(W(t_i)-W(t_{i-1})),$$

where $t_0=0$. As the increments of W are independent normal distributed it follows from elementary properties of the normal distribution that the linear combination $\sum_{i=1}^n a_i W(t_i)$ is univariate normal distributed.

As $W(0)=0$ the fact that increments are zero-mean implies that all process values $W(t_i)=(W(t_i+0)-W(0))+W(0)$ are zero-mean. As for covariances, independence and stationarity of increments give

$$\begin{aligned}\mathbf{Cov}\{W(s),W(s+t)\} &= \mathbf{Cov}\{W(s),W(s)\}+\mathbf{Cov}\{W(s),W(s+t)-W(s)\}\\ &= \mathbf{Var}\{W(s)\}+0 \quad \text{for } s,t\geq 0.\end{aligned}$$

Here independence and stationarity of increments give

$$\begin{aligned}&\mathbf{Var}\{W(s+t)\}\\ &\quad = \mathbf{Var}\{W(s+t)-W(t)\}+2\,\mathbf{Cov}\{W(s+t)-W(t),W(t)\}+\mathbf{Var}\{W(t)\}\\ &\quad = \mathbf{Var}\{W(s)\}+\mathbf{Var}\{W(t)\} \quad \text{for } s,t\geq 0.\end{aligned}$$

We see that $f(t)=\mathbf{Var}\{W(t)\}$, $t\geq 0$, is a non-decreasing function of t that solves the Cauchy functional equation $f(s+t)=f(s)+f(t)$ for $s,t\geq 0$. However, recalling the proof of Theorem 6.9.13, the only possible solutions to this equation are $f(t)=C\,t$ for some constant $C\in\mathbb{R}$. In our case that constant must be $C=\mathbf{Var}\{W(1)\}$. □

[8]Norbert Wiener, American scientist 1894-1964 who gave important contributions to mathematics as well as many other areas of science including early computer science.

[9]Robert Brown 1773-1858, Scottish botanist who made the first empirical observations of Brownian motion as the movement pattern of small particles in a fluid.

Lecture 5, Thursday 4 February

Chapter 9 Stationary processes

9.1 Introduction

Recall Definitions 8.2.1 and 8.2.2 of strong stationarity and (weak) stationarity, respectively.

In this chapter it will sometimes be natural to consider complex valued stochastic processes: A complex valued stochastic process is a family $\{X(t)\}_{t\in T}$ of complex valued random variables defined on a common sample space Ω. A complex valued random variable X, in turn, simply is a function $X:\Omega\to\mathbb{C}$. Eqivalently, a complex valued random variable X is given by $X=X_1+iX_2$, where the real and imaginary part of X, X_1 and X_2, respectively, are ordinary real valued random variables.

The Definition 8.2.1 of strong stationarity that

$$\big(X(t_1{+}h),\ldots,X(t_n{+}h)\big) =_D \big(X(t_1),\ldots,X(t_n)\big)$$

for all $t_1+h,\ldots,t_n+h,t_1,\ldots,t_n\in T$ formally remains valid for a complex valued process $\{X(t)\}_{t\in T}$. Now the above equality in distribution $=_D$ simply means that

$$\mathbf{P}\big\{X(t_1{+}h)\in A_1,\ldots,X(t_n{+}h)\in A_n\big\} = \mathbf{P}\big\{X(t_1)\in A_1,\ldots,X(t_n)\in A_n\big\}$$

for all subsets $A_1,\ldots,A_n$ of $\mathbb{C}$.

The expected value of a complex valued random variable $X=X_1+iX_2$ with real and imaginary part X_1 and X_2, respectively, is defined $\mathbf{E}\{X\}=\mathbf{E}\{X_1\}+i\,\mathbf{E}\{X_2\}$.

Definition 9.1.1. *The* covariance *between two complex valued random variables X and Y defined on a common probability space is given by*

$$\mathbf{Cov}\{X,Y\} = \mathbf{E}\Big\{\big(X-\mathbf{E}\{X\}\big)\,\overline{\big(Y-\mathbf{E}\{Y\}\big)}\Big\}.$$

Having resolved the of issue how to define expectations and covariaces for complex valued random variables, the Definition 8.2.2 of (weak) stationarity that the function $T\ni t\curvearrowright \mathbf{E}\{X(t)\}$ is a constant and that the function $T\times T\ni(s,t)\curvearrowright \mathbf{Cov}\{X(s), X(t)\}$ only depends on the difference $t-s$ between s and t remains valid for a complex valued process $\{X(t)\}_{t\in T}$.

Definition 9.1.2. *The* variance *of a complex valued random variable X is given by*

$$\mathbf{Var}\{X\} = \mathbf{Cov}\{X,X\}.$$

Definition 9.1.3. *Two complex valued random variables X and Y are* orthogonal *if*

$$\mathbf{Cov}\{X,Y\}=0.$$

Example 9.1.4. Pick an integer $m\geq 2$ and let $X(t)=\mathrm{e}^{2\pi i(Z+N(t))/m}$, where $\{N(t)\}_{t\geq 0}$ is a Poisson process with intensity $\lambda>0$ and Z is a random variable that is independent of N and uniformly distributed over the integers $\{1,\ldots,m\}$. By symmetry considerations it is clear that $\mathbf{E}\{X(t)\}=0$. Further, we have

$$\begin{aligned}\mathbf{Cov}\{X(t),X(t+h)\}&=\mathbf{E}\big\{X(t)\,\overline{X(t+h)}\big\}\\&=\mathbf{E}\big\{\mathrm{e}^{2\pi i(N(t)-N(t+h))/m}\big\}\\&=\mathbf{E}\big\{\mathrm{e}^{-2\pi i\,N(h)/m}\big\}\\&=\mathrm{e}^{-\lambda h(1+2\pi i/m)}\quad\text{for }\ h\geq 0\end{aligned}$$

(by basic properties of Poisson processes and elementary calculations), so that X is stationary with covariance $\mathbf{Cov}\{X(t),X(t+h)\}=\mathrm{e}^{-\lambda|h|-2\pi i\lambda h/m)}$ for $h\in\mathbb{R}$, see the following exercise. #

Exercise. Explain how the formula $\mathbf{Cov}\{X(t),X(t+h)\}=\mathrm{e}^{-\lambda|h|-2\pi i\lambda h/m)}$ for $h\in\mathbb{R}$ in Example 9.1.4 follows from the fact that the formula holds for $h\geq 0$.

9.2 Linear prediction

Theorem 7.9.17. *Given random variables $X_1,\ldots,X_n$ and a square-integrable random variable Y, the function $\phi:\mathbb{R}^n\to\mathbb{R}$ that minimizes the* mean-square prediction error

$$\mathbf{E}\big\{[Y-\phi(X_1,\ldots,X_n)]^2\big\}$$

is given by the minimum mean-square error predictor

$$\phi(x_1,\ldots,x_n)=\mathbf{E}\big\{Y\big|X_1\!=\!x_1,\ldots,X_n\!=\!x_n\big\}.$$

Proof. For the linear space H of square-integrable functions $h(X_1,\ldots,X_n)$ we have

$$\begin{aligned}\mathbf{E}\big\{[Y-\phi(X_1,\ldots,X_n)]Z\big\}&=\mathbf{E}\Big\{\mathbf{E}\big\{[Y-\phi(X_1,\ldots,X_n)]Z\big|X_1,\ldots,X_n\big\}\Big\}\\&=\mathbf{E}\Big\{\mathbf{E}\big\{Y-\phi(X_1,\ldots,X_n)\big|X_1,\ldots,X_n\big\}Z\Big\}\\&=\mathbf{E}\Big\{\big[\mathbf{E}\{Y\big|X_1,\ldots,X_n\}-\mathbf{E}\{Y\big|X_1,\ldots,X_n\}\big]Z\Big\}\\&=0\quad\text{for }\ Z\in H.\end{aligned}$$

From this in turn it follows that

$$
\begin{aligned}
&\mathbf{E}\big\{[Y-Z]^2\big\} \\
&\quad= \mathbf{E}\big\{[Y-\phi(X_1,\ldots,X_n)]^2\big\} + 2\,\mathbf{E}\big\{[Y-\phi(X_1,\ldots,X_n)]\,[\phi(X_1,\ldots,X_n)-Z]\big\} \\
&\qquad\qquad\qquad\qquad + \mathbf{E}\big\{[\phi(X_1,\ldots,X_n)-Z]^2\big\} \\
&\quad= \mathbf{E}\big\{[Y-\phi(X_1,\ldots,X_n)]^2\big\} + \mathbf{E}\big\{[\phi(X_1,\ldots,X_n)-Z]^2\big\} \\
&\quad\geq \mathbf{E}\big\{[Y-\phi(X_1,\ldots,X_n)]^2\big\} \quad \text{for any } Z\in H. \qquad \square
\end{aligned}
$$

Given the information about the values of the random variables $X_1,\ldots,X_n$, the minimum mean-square error predictor of the random variable Y is given by Theorem 7.9.17. However, the actual evaluation of that predictor as a rule requires knowledge about the joint distribution of $(Y,X_1,\ldots,X_n)$, which is usually not available in a typical statistical application. In practice, one therefore often instead employ the following alternative predictor which only requires knowledge about the covariance structure between the random variables $Y,X_1,\ldots,X_n$:

Theorem 9.2.1. *Given square-integrable random variables $X_1,\ldots,X_n$ and Y, the linear function $(x_1,\ldots,x_n)\ni\mathbb{R}^n \curvearrowright \sum_{i=1}^n a_i x_i \to \mathbb{R}$ that minimizes the mean-square prediction error*

$$\mathbf{E}\big\{[Y-\textstyle\sum_{i=1}^n a_i X_i]^2\big\}$$

is given by the minimum mean-square error linear predictor *for which $\{a_i\}_{i=1}^n$ satisfy*

$$\mathbf{E}\{X_iY\} = \textstyle\sum_{j=1}^n a_j\,\mathbf{E}\{X_iX_j\} \quad \textit{for } i=1,\ldots,n.$$

In particular, if the random variables $X_1,\ldots,X_n$ are zero-mean, then

$$\mathbf{Cov}\{X_i,Y\} = \textstyle\sum_{j=1}^n a_j\,\mathbf{Cov}\{X_i,X_j\} \quad \textit{for } i=1,\ldots,n.$$

Proof. The idea of the proof is the same as that in the proof of Theorem 7.9.17: For any linear predictor $\sum_{j=1}^n b_i X_i$ we have

$$\mathbf{E}\big\{[Y-\textstyle\sum_{j=1}^n a_jX_j]\sum_{i=1}^n b_iX_i\big\} = \displaystyle\sum_{i=1}^n b_i\left(\mathbf{E}\{X_iY\}-\textstyle\sum_{j=1}^n a_j\,\mathbf{E}\{X_iX_j\}\right) = 0,$$

so that

$$
\begin{aligned}
&\mathbf{E}\big\{[Y-\textstyle\sum_{i=1}^n b_iX_i]^2\big\} \\
&\quad= \mathbf{E}\big\{[Y-\textstyle\sum_{j=1}^n a_jX_j]^2\big\} + 2\,\mathbf{E}\big\{[Y-\textstyle\sum_{j=1}^n a_jX_j]\sum_{i=1}^n (a_i-b_i)X_i\big\} \\
&\qquad\qquad\qquad\qquad + \mathbf{E}\big\{[\textstyle\sum_{i=1}^n (a_i-b_i)X_i]^2\big\} \\
&\quad= \mathbf{E}\big\{[Y-\textstyle\sum_{j=1}^n a_jX_j]^2\big\} + \mathbf{E}\big\{[\textstyle\sum_{i=1}^n (a_i-b_i)X_i]^2\big\} \\
&\quad\geq \mathbf{E}\big\{[Y-\textstyle\sum_{j=1}^n a_jX_j]^2\big\}. \qquad \square
\end{aligned}
$$

Example 9.2.5. (AUTOREGRESSIVE SCHEME) Let $\{Z_n\}_{n\in\mathbb{Z}}$ be a sequence of independent zero-mean random variables with common variances σ^2. Let $\{X_n\}_{n\in\mathbb{Z}}$ be a stationary so called *AR-process* such that for some constant $\alpha\in(-1,1)$

$$\begin{cases} Z_n \text{ is independent of } X_{n-1}, X_{n-2}, \ldots & \text{for } n\in\mathbb{Z}, \\ X_n = \alpha\, X_{n-1} + Z_n & \text{for } n\in\mathbb{Z}. \end{cases}$$

Then stationarity gives

$$\mathbf{E}\{X_n\} = \alpha\,\mathbf{E}\{X_{n-1}\} + \mathbf{E}\{Z_n\} = \alpha\,\mathbf{E}\{X_n\} + 0 \implies \mathbf{E}\{X_n\} = 0,$$

while the autocovariance function (see the next definition below) is given by

$$c(k) \equiv \mathbf{Cov}\{X_{n-k}, X_n\} = \mathbf{Cov}\{X_{n-k}, \alpha\, X_{n-1} + Z_n\} = \alpha\, c(k-1) + 0 \quad \text{for } k\geq 1,$$

so that $c(k) = c(0)\,\alpha^{|k|}$ for $k\in\mathbb{Z}$. Here stationarity gives

$$c(0) = \mathbf{Var}\{X_n\} = \mathbf{Var}\{\alpha\, X_{n-1}+Z_n\} = \alpha^2\,\mathbf{Var}\{X_{n-1}\}+\mathbf{Var}\{Z_n\} = \alpha^2\, c(0)+\sigma^2,$$

so that

$$c(k) = \frac{\sigma^2}{1-\alpha^2}\,\alpha^{|k|} \quad \text{for } k\in\mathbb{Z}.$$

Hence Theorem 9.2.1 shows that the minimum mean-square error linear predictor $\sum_{i=1}^n a_i X_i$ of X_{n+k} is given by the equations

$$\mathbf{Cov}\{X_i, X_{n+k}\} = \textstyle\sum_{j=1}^n a_j\,\mathbf{Cov}\{X_i, X_j\} \quad \text{for } i=1,\ldots,n,$$

which is to say that

$$\alpha^{n+k-i} = \textstyle\sum_{j=1}^n a_j\,\alpha^{|j-i|} \quad \text{for } i=1,\ldots,n.$$

This system of equations is solved by $a_1 = \ldots = a_{n-1} = 0$ and $a_n = \alpha^k$. #

9.3 Autocovariances and spectra

Definition. *The* autocovariance function c *of a complex valued (weakly) stationary process* $X=\{X(t)\}_{t\in T}$ *is defined as*

$$c(t) = \mathbf{Cov}\{X(s), X(s+t)\} \quad \textit{for } s, s+t\in T.$$

When considering stationary processes we will henceforth assume (unless otherwise is explicitely stated) that the (constant) expected value of the process is zero, as that only amounts to a shift of a constant level as compared with the general case.

Theorem 9.3.2. *The autocovariance function c of a complex valued stationary process $X=\{X(t)\}_{t\in T}$ is* hermitian, *which is to say that*

$$c(-t)=\overline{c(t)} \quad \textit{for all } t,$$

and non-negative definite, *which is to say that*

$$\sum_{i,j=1}^{n} c(t_i-t_j)\, z_i\overline{z_j}\geq 0 \quad \textit{for all } n\in\mathbb{N},\ z_1,\ldots,z_n\in\mathbb{C} \textit{ and } t_1,\ldots,t_n\in T.$$

Proof. By Definition 9.1.1 we have

$$c(-t)=\mathbf{Cov}\{X(s),X(s-t)\}=\mathbf{Cov}\{X(s+t),X(s)\}=\overline{\mathbf{Cov}\{X(s),X(s+t)\}}=\overline{c(t)},$$

while Definition 9.1.2 gives

$$0\leq \mathbf{Var}\Big\{\sum_{i=1}^{n} z_i X(t_i)\Big\}=\mathbf{Cov}\Big\{\sum_{i=1}^{n} z_i X(t_i),\overline{\sum_{j=1}^{n} z_j X(t_j)}\Big\}=\sum_{i,j=1}^{n} c(t_i-t_j)\, z_i\overline{z_j}. \qquad \square$$

Definition 9.3.3. *The* autocorrelation function *ρ of a complex valued stationary process $X=\{X(t)\}_{t\in T}$ is defined as*

$$\rho(t)=\frac{\mathbf{Cov}\{X(s),X(s+t)\}}{\sqrt{\mathbf{Var}\{X(s)\}\,\mathbf{Var}\{X(s+t)\}}}=\frac{c(t)}{c(0)} \quad \textit{for } s,s+t\in T$$

whenever $c(0)>0$.

The following important theorem is given without proof because the proof belongs to a course in advanced calculus (in particular Fourier transforms) and do not have any probabilistic content:

Theorem 9.3.4. (Spectral theorem for autocorrelations) *If a stationary process has variance $c(0)>0$ and autocorrelation function $\rho(t)$ that is continuous at $t=0$, then there exists a probability distribution function F, the so called* spectral distribution function *of the process, such that ρ is the characteristic function of F, which is to say that*

$$\rho(t)=\int_{-\infty}^{\infty} \mathrm{e}^{it\lambda}\, dF(\lambda) \quad \textit{for all } t.$$

Proof. Omitted. $\square$

The interpretation of the spectral representation of the autocorrelation function is just the usual one of Fourier transforms as the build up of the autocorrelation function by means of its frequency content.

The integral in the spectral representation is a so called *Riemann-Stieltjes integral* which is obtained as the limit of approximating Riemann-Stieltjes sums

$$\int_{-\infty}^{\infty} e^{it\lambda}\, dF(\lambda) = \lim_{\max_n(\lambda_n-\lambda_{n-1})\downarrow 0} \sum_n e^{it\lambda_{n-1}}\big(F(\lambda_n)-F(\lambda_{n-1})\big)$$

as the mesh $-\infty \leftarrow \ldots < \lambda_{-2} < \lambda_{-1} < \lambda_0 < \lambda_1 < \lambda_2 \ldots \to \infty$ becomes finer and finer.

If the spectral distribution is a continuous distribution with probability density function $f(\lambda) = F'(\lambda)$, so that

$$F(\lambda_n) - F(\lambda_{n-1}) = (\lambda_n - \lambda_{n-1})\, f(\lambda_{n-1}) + o(\lambda_n - \lambda_{n-1}) \quad \text{as} \quad \max_n(\lambda_n - \lambda_{n-1}) \downarrow 0,$$

the we have

$$\rho(t) = \int_{-\infty}^{\infty} e^{it\lambda}\, f(\lambda)\, d\lambda \quad \text{for all } t,$$

as a Riemann sum argument shows that

$$\begin{aligned}\int_{-\infty}^{\infty} e^{it\lambda}\, dF(\lambda) &= \lim_{\max_n(\lambda_n-\lambda_{n-1})\downarrow 0} \sum_n e^{it\lambda_{n-1}}\big(F(\lambda_n)-F(\lambda_{n-1})\big)\\ &= \lim_{\max_n(\lambda_n-\lambda_{n-1})\downarrow 0} \sum_n e^{it\lambda_{n-1}}(\lambda_n-\lambda_{n-1})\, f(\lambda_{n-1})\\ &= \int_{-\infty}^{\infty} e^{it\lambda}\, f(\lambda)\, d\lambda.\end{aligned}$$

In this case we call f the *spectral density function* of the process.

If X is a real-valued process, then Theorem 9.3.2 shows that ρ is a symmetric function (as c is). From this we get that the spectral distribution must be a symmetric distribution, by uniqueness of Fourier-Stieltjes transforms together with the fact that

$$\int_{-\infty}^{\infty} e^{it\lambda}\, dF(\lambda) = \rho(t) = \rho(-t) = \int_{-\infty}^{\infty} e^{-it\lambda}\, dF(\lambda) = \int_{-\infty}^{\infty} e^{it\lambda}\, dF(-\lambda).$$

In this case we may also use the following simplified spectral representation

$$\begin{aligned}\rho(t) = \frac{\rho(t)+\rho(-t)}{2} &= \int_{-\infty}^{\infty} \frac{e^{it\lambda}+e^{-it\lambda}}{2}\, dF(\lambda)\\ &= \int_{-\infty}^{\infty} \cos(t\lambda)\, dF(\lambda)\\ &= 2\int_0^{\infty} \cos(t\lambda)\, dF(\lambda).\end{aligned}$$

Definition 9.3.10. *The* spectrum *of a stationary process with spectral distribution function F is the set*

$$\{\lambda \in \mathbb{R} : F(\lambda+\varepsilon) - F(\lambda-\varepsilon) > 0 \text{ for all } \varepsilon > 0\}.$$

Remembering the definition of the Riemann-Stieltjes integral as the limit of approximating Riemann-Stieltjes sums we see that the spectrum simply consists of those frequencies that contributes to the build up of the autocorrelations function (by means of its frequencies).

Theorem. *For a continuous time stationary process* $\{X(t)\}_{t\in\mathbb{R}}$ *possessing an integrable autocorrelation function*

$$\int_{-\infty}^{\infty} |\rho(t)|\, dt < \infty$$

that is continuous at zero, we have

$$\rho(t) = \int_{-\infty}^{\infty} \mathrm{e}^{it\lambda} f(\lambda)\, d\lambda \quad \text{for } t \in \mathbb{R},$$

where the spectral density function f *is given by*

$$f(\lambda) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{e}^{-it\lambda} \rho(t)\, dt \quad \text{for } \lambda \in \mathbb{R}.$$

Proof. This simply is Fourier transform theory. □

We will return to examples of application of this theorem later, e.g., when we study Gaussian processes. We will now instead consider discrete time processes.

Theorem. *For a discrete time stationary process* $\{X(t)\}_{t\in\mathbb{Z}}$ *possesing an autocorrelation function* ρ *that is continuous at zero, we have*

$$\rho(n) = \int_{(-\pi,\pi]} \mathrm{e}^{it\lambda}\, dF(\lambda) \quad \text{for } n \in \mathbb{Z},$$

for a spectral distribution function F *such that* $F(-\pi) = 0$ *and* $F(\pi) = 1$.

Proof. Writing G for the spectral distribution function for the process X that is provided by Theorem 9.3.4, we have

$$\begin{aligned}
\rho(n) &= \int_{-\infty}^{\infty} \mathrm{e}^{in\lambda}\, dG(\lambda) \\
&= \sum_{k=-\infty}^{\infty} \int_{((2k-1)\pi,(2k+1)\pi]} \mathrm{e}^{in\lambda}\, dG(\lambda) \\
&= \sum_{k=-\infty}^{\infty} \int_{(-\pi,\pi]} \mathrm{e}^{in(\lambda+2k\pi)}\, dG(\lambda+2k\pi) \\
&= \int_{(-\pi,\pi]} \mathrm{e}^{in\lambda} \left(\sum_{k=-\infty}^{\infty} dG(\lambda+2k\pi) \right).
\end{aligned}$$

□

Exercise. (DIFFICULT) There is a possible element of "cheating" in the application of Theorem 9.3.4 in the above proof – explain what can be the problem!

Theorem 9.3.15. *For a discrete time stationary process* $\{X(t)\}_{t\in\mathbb{Z}}$ *possesing a summable autocorrelation function*

$$\sum_{n=-\infty}^{\infty} |\rho(n)| < \infty$$

that is continuous at zero, we have

$$\rho(n) = \int_{-\pi}^{\pi} \mathrm{e}^{in\lambda} f(\lambda)\, d\lambda \quad \textit{for } n\in\mathbb{Z},$$

where the spectral density function f *is given by*

$$f(\lambda) = \frac{1}{2\pi}\sum_{n=-\infty}^{\infty} \mathrm{e}^{-in\lambda}\rho(n) \quad \textit{for } \lambda\in[-\pi,\pi].$$

Proof. This simply is Fourier series theory. □

Example 9.3.19. (DISRETE TIME WHITE NOISE) A sequence $\{X_n\}_{n\in\mathbb{Z}}$ of uncorrelated zero-mean random variables with unit variance is stationary with autocovariance function and autocorrelation function given by $c(n) = \rho(n) = \delta(n)$ (Kronecker's[10] delta). As ρ is summable the process has spectral density function

$$f(\lambda) = \frac{1}{2\pi}\sum_{n=-\infty}^{\infty} \mathrm{e}^{-in\lambda}\rho(n) = \frac{1}{2\pi} \quad \text{for } \lambda\in[-\pi,\pi]. \qquad \#$$

Example 9.3.20. (IDENTICAL SEQUENCE) A process $\{X_n\}_{n\in\mathbb{Z}}$ given by $X_n = Y$ for $n\in\mathbb{Z}$ for a single zero-mean and unit variance random variable Y is stationary with autocovariance function and autocorrelation function given by $c(n) = \rho(n) = 1$ for $n\in\mathbb{Z}$. This ρ is not summable, but we see that the spectral distribution function F corresponds to a unit mass at zero as that gives

$$\rho(n) = \int_{(-\pi,\pi]} \mathrm{e}^{it\lambda}\, dF(\lambda) = \mathrm{e}^{it\cdot 0} = 1 \quad \text{for } n\in\mathbb{Z}. \qquad \#$$

Example 9.2.5. (CONTINUED) The AR-process $\{X_n\}_{n\in\mathbb{Z}}$ in Example 9.2.5 has autocorrelation fucntion $\rho(n) = \alpha^{-|n|}$, so that the spectral density function is

$$f(\lambda) = \frac{1}{2\pi}\sum_{n=-\infty}^{\infty} \mathrm{e}^{-in\lambda}\alpha^{-|n|} = \cdots = \frac{1-\alpha^2}{2\pi\,(1+\alpha^2-2\alpha\cos(\lambda))} \quad \text{for } \lambda\in[-\pi,\pi]. \quad \#$$

[10]Leopold Kronecker, German mathematician 1823-1891.

Lecture 6, Friday 5 February

9.4 Stochastic integration and the spectral representation

This section is in essence a single very long proof of the so called *spectral theorem*. The details of that proof are argubly not terribly important on their own, and they will therefore be omitted. Also, no hand-in exercise is required for this section.

Theorem 9.4.2. (SPECTRAL THEOREM) *If a stationary process X has autocorrelation function $\rho(t)$ that is continuous at $t=0$ with* spectral distribution function F, *then there exists a complex valued zero-mean stochastic process $\{S(\lambda)\}_{\lambda\in\mathbb{R}}$ called the* spectral process of X, *that has orthogonal increments, which is to say*

$$\mathbf{Cov}\big\{S(\lambda)-S(\mu), S(\eta)-S(\nu)\big\}=0 \quad \textit{for} \quad -\infty<\mu\leq\lambda\leq\nu\leq\eta<\infty,$$

and that satisfies

$$\mathbf{Var}\{S(\lambda)-S(\mu)\}=F(\lambda)-F(\mu) \quad \textit{for} \quad -\infty<\mu\leq\lambda<\infty,$$

such that X has spectral representation

$$X(t)=\int_{-\infty}^{\infty}\mathrm{e}^{-it\lambda}\,dS(\lambda) \quad \textit{for all } t.$$

Proof. Omitted. □

The *stochastic integral* $\int_{-\infty}^{\infty}\mathrm{e}^{-it\lambda}\,dS(\lambda)$ in the spectral representation is a (stochastic) Riemann-Stieltjes integral obtained as the limit

$$X(t)=\int_{-\infty}^{\infty}\mathrm{e}^{-it\lambda}\,dS(\lambda)=\lim_{\max_n(\lambda_n-\lambda_{n-1})\downarrow 0}\sum_n \mathrm{e}^{-it\lambda_{n-1}}\big(S(\lambda_n)-S(\lambda_{n-1})\big)$$

of approximating Riemann-Stieltjes sums as the mesh $-\infty\leftarrow\ldots<\lambda_{-2}<\lambda_{-1}<\lambda_0<\lambda_1<\lambda_2\ldots\rightarrow\infty$ becomes finer and finer, see also the text following Theorem 9.3.4. As we are talking about convergence of random variables here, we must specify in which sense that convergence is. And that is in *mean-square*, which is to say that the convergence displayed in the previous formula really means (by definition) that

$$\lim_{\max_n(\lambda_n-\lambda_{n-1})\downarrow 0}\mathbf{E}\left\{\left(X(t)-\sum_n \mathrm{e}^{-it\lambda_{n-1}}\big(S(\lambda_n)-S(\lambda_{n-1})\big)\right)^2\right\}=0.$$

Example. The spectral theorem implies the spectral representation in Theorem 9.3.4 of the autocorrelation function, as

$$\begin{aligned}&\rho(t)\\ &=\mathbf{Cov}\{X(s),X(s+t)\}\end{aligned}$$

$$
\begin{aligned}
&= \lim \mathbf{Cov}\Big\{\sum_m \mathrm{e}^{-is\lambda_{m-1}}\big(S(\lambda_m)-S(\lambda_{m-1})\big), \sum_n \mathrm{e}^{-i(s+t)\lambda_{n-1}}\big(S(\lambda_n)-S(\lambda_{n-1})\big)\Big\} \\
&= \lim \sum_m \sum_n \mathrm{e}^{i(s+t)\lambda_{n-1}-is\lambda_{m-1}} \mathbf{Cov}\big\{S(\lambda_m)-S(\lambda_{m-1}), S(\lambda_n)-S(\lambda_{n-1})\big\} \\
&= \lim \sum_n \mathrm{e}^{it\lambda_{n-1}} \mathbf{Var}\big\{S(\lambda_n)-S(\lambda_{n-1})\big\} \\
&= \lim \sum_n \mathrm{e}^{it\lambda_{n-1}} \big(F(\lambda_n)-F(\lambda_{n-1})\big) \\
&= \int_{-\infty}^{\infty} \mathrm{e}^{it\lambda}\, dF(\lambda). \qquad \#
\end{aligned}
$$

9.5 The ergodic theorem

Also the treatment of this section will be shortened and rather summary. However, for this section a hand-in exercise is required. (We will instead use our time to give a rather extensive treatment of Gaussian processes in the next section.)

The *strong law of large numbers* states that for an iid. sequence of random variables $\{X_n\}_{n=1}^{\infty}$ that are integrable, $\mathbf{E}\{|X_1|\} < \infty$, we have

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to \mathbf{E}\{X_1\} \quad \text{as } n \to \infty.$$

Here the convergence is *strong*, which is to say that the sequence on the left-hand side converges to the limit on the right-hand side for all $\omega \in \Omega$ (or for all ω except those in an event with probability zero).

Ergodic theorems deal with generalizations of the strong law of large numbers to sequences of random variables that are strongly or weakly stationary, but not necessarily iid.

Theorem 9.5.2. (Ergodic theorem for strongly stationary processes) *If $\{X_n\}_{n=1}^{\infty}$ is a strongly stationary sequence of random variables that are integrable, $\mathbf{E}\{|X_1|\} < \infty$, then there exists an integrable random variable Y such that*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to Y \quad \textit{as } n \to \infty$$

in the sense of strong convergence (see above) as well as in the sense of mean convergence, which is to say that

$$\mathbf{E}\bigg\{\bigg|\frac{1}{n}\sum_{i=1}^{n} X_i - Y\bigg|\bigg\} \to 0 \quad \textit{as } n \to \infty.$$

Proof. Omitted. □

Theorem 9.5.3. (Ergodic theorem for stationary processes) *If $\{X_n\}_{n=1}^{\infty}$ is a stationary sequence of random variables, then there exists a square-integrable random variable Y with $\mathbf{E}\{Y\} = \mathbf{E}\{X_1\}$ such that*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to Y \quad \text{as } n \to \infty$$

in the sense of mean-square convergence, which is to say that

$$\mathbf{E}\left\{\left(\frac{1}{n}\sum_{i=1}^{n} X_i - Y\right)^2\right\} \to 0 \quad \text{as } n \to \infty.$$

Proof. Omitted. □

Exercise. Prove that mean-square convergence implies mean convergence.

The following example illustrates the usefulness of ergodicity results:

Example 9.5.28. Let $X = \{X_n\}_{n=0}^{\infty}$ be an irreducible ergodic (that is, all states are non-null persistent aperiodic) Markov chain with stationary distribution π (recall Theorem 6.4.3). Start X according to the stationary distribution, $\mu^{(0)} = \pi$, so that X is strongly stationary (recall Example 8.2.4). Define a new strongly stationary sequence $I = \{I_n\}_{n=0}^{\infty}$ by $I_n = 1$ if $X_n = k$ while $I_n = 0$ if $X_n \neq k$. (Why is I strongly stationary?) Then we readily see that I has autocovariance function

$$c(n) = \mathbf{Cov}\{X_m, X_{m+n}\} = \mathbf{E}\{X_m X_{m+n}\} - \mathbf{E}\{X_m\}\mathbf{E}\{X_{m+n}\} = \pi_k p_{kk}(n) - \pi_k^2.$$

(Why?) As $p_{kk}(n) \to \pi_k$ as $n \to \infty$ by Theorem 6.4.17, we have $c(n) \to 0$ as $n \to \infty$. From this it follows readily that

$$\mathbf{Var}\left\{\frac{1}{n}\sum_{i=0}^{n-1} I_i\right\} = \frac{1}{n^2}\sum_{i=0}^{n-1}\sum_{j=0}^{n-1} c(i-j) \to 0 \quad \text{as } n \to \infty.$$

(Why?) On the other hand Theorem 9.5.3 shows that

$$\frac{1}{n}\sum_{i=1}^{n} I_i \to Y \quad \text{as } n \to \infty$$

in the sense of mean-square convergence for some square-integrable random variable Y with $\mathbf{E}\{Y\} = \mathbf{E}\{I_1\} = \pi_k$. It follows that $\mathbf{Var}\{Y\} = 0$ so that

$$\frac{1}{n}\sum_{i=1}^{n} I_i \to \pi_k \quad \text{as } n \to \infty.$$

(Why?) And so, using ergodicity theory, we can estimate the expected value $\mathbf{E}\{I_1\}$ $= \pi_k$ by the sample average of the stationary sequence $I = \{I_i\}_{i=0}^{n-1}$. That is to say that we can move along a single sample path of I to estimate an theoretical expected value by means of a smaple average, instead of (as in elementary statistics) requiring several independent copies of I. #

Exercise. Answer the four whys in Example 9.5.28.

Lecture 7, Thursday 11 February

9.6 Gaussian processes

Definition 9.6.3. *A stochastic process* $\{X(t)\}_{t\in T}$ *is* Gaussian[11] *if each vector of process values* $(X(t_1),\ldots,X(t_n))$ *is multivariate normal distributed, which is to say,*

$$\sum_{i=1}^{n} a_i X(t_i) \quad \textit{is univariate normal distributed}$$

for each choice of $n\in\mathbb{N}$ *and* $a_1,\ldots,a_n\in\mathbb{R}$.

By Theorem 9.3.2 the autocovariance function of a real-valued stationary process $\{X(t)\}_{t\in T}$ is symmetric and non-negative definite. This result has the converse that every symmetric and non-negative definite function is the autocovariance function for some stationary process. More precisely, we have the following result:

Theorem 9.6.1. *If* $T\backslash T \ni t \curvearrowright c(t)\in\mathbb{R}$ *is a symmetric and non-negative definite function, then there exists a zero-mean stationary Gaussian process* $\{X(t)\}_{t\in T}$ *that has autocovariance function* c.

Proof. The proof of this result is very closely linked to the so called *Kolmogorov Consistency theorem*; a result that we have omitted as it goes into the very hart of Lebesgue measure theory. □

Example 9.6.13. The treatment of the Wiener process in Section 8.5 shows that the Wiener process is a Gaussian process. #

It is an immediate consequence of Definition 9.6.3 that each process value $X(t)$ of a Gaussian process $\{X(t)\}_{t\in T}$ is normal distributed (take $n=1$, $t_1=t$ and $a_1=1$). However, the converse is *not true*, which is to say that a process for which each process value $X(t)$ is normal distributed need not be Gaussian.

Example. Let ξ and η be independent standard normal distributed random variables and consider the process $X=\{X(t)\}_{t\in\{0,1\}}$ defined by $X(0)=\text{sign}(\xi)\eta$ and $X(1)=\text{sign}(\eta)\xi$. Then X is not Gaussian albeit $X(0)$ and $X(1)$ are both standard normal distributed, see Exercise below. #

Exercise. (DIFFICULT) The process in the above example is not Gaussian.

[11] Johann Carl Friedrich Gauss, German mathematcian 1777-1855. Argubly the greatest mathematician of all time (albeit, also argubly, an equally distinguished bore).

Theorem. *The finite dimensional distributions of a Gaussian process* $\{X(t)\}_{t\in T}$ *are determined by the mean function* $T \ni t \curvearrowright m(t) = \mathbf{E}\{X(t)\}$ *and autocovariance function* $T\times T \ni (s,t) \curvearrowright c(s,t) = \mathbf{Cov}\{X(s),X(t)\}$ *of the process.*

Proof. The distribution of the random variable $(X(t_1),\ldots,X(t_n))$ is determined by its characteristic function $\mathbb{R}^n \ni (\theta_1,\ldots,\theta_n) \curvearrowright \mathbf{E}\{\mathrm{e}^{i\sum_{j=1}^n \theta_j X(t_j)}\} \in \mathbb{C}$. That characteristic function in turn is determined by the univariate distribution of the linear combination $\sum_{j=1}^n \theta_j\, X(t_j)$ for each $(\theta_1,\ldots,\theta_n) \in \mathbb{R}^n$. As X is Gaussian that linear combination is univariate normal distributed with expected value $\mathbf{E}\{\sum_{j=1}^n \theta_j X(t_j)\} = \sum_{j=1}^n \theta_j\, m(t_j)$ and variance $\mathbf{Cov}\{\sum_{j=1}^n \theta_j\, X(t_j), \sum_{k=1}^n \theta_k\, X(t_k)\} = \sum_{j=1}^n \sum_{k=1}^n \theta_j\theta_k\, c(t_j,t_k)$. □

As Gaussian processes are determined (in the sense of uniqueness of their finite dimensional distributions) by their mean and covariance functions, it is convenient to specify a Gaussian processes simply by saying what these functions are. But note that only symmetric non-negative definite functions can be covariance functions!

Theorem. *Two process values* $X(s)$ *and* $X(t)$ *of a Gaussian process* X *are independent if and only if they are uncorrelated.*

Proof. The implication to the right is elementary. Now, assume that $X(s)$ and $X(t)$ are uncorrelated. By the previous theorem the distribution function $F_{X(s),X(t)}(x,y)$ of $(X(s),X(t))$ is determined by the expected values $\mathbf{E}\{X(s)\}$ and $\mathbf{E}\{X(t)\}$, together with the variances $\mathbf{Var}\{X(s)\}$ and $\mathbf{Var}\{X(t)\}$ and the covariance $\mathbf{Cov}\{X(s),X(t)\}$. As the latter covariance is zero the distribution function $F_{X(s),X(t)}(x,y)$ is the same as the distribution function $F_{\xi,\eta}(x,y)$ of two independent normal distributed random variables ξ and η with expected values $\mathbf{E}\{X(s)\}$ and $\mathbf{E}\{X(t)\}$ and variances $\mathbf{Var}\{X(s)\}$ and $\mathbf{Var}\{X(t)\}$, respectively (as such random variables will have zero covariance). □

Theorem 9.6.4. *A Gaussian process is stationary if and only if it is strongly stationary.*

Proof. As Gaussian processes have well-defined moments of all orders the implication to the left is immediate (recall Section 8.2). Now, if X is a stationary Gaussian process, then the distribution of $(X(t_1+h),\ldots,X(t_n+h))$ is determined by the expected value $\mathbf{E}\{\sum_{j=1}^n \theta_j X(t_j+h)\} = \sum_{j=1}^n \theta_j\, m(t_j+h) = \sum_{j=1}^n \theta_j\, m$ and variance $\mathbf{Cov}\{\sum_{j=1}^n \theta_j\, X(t_j+h), \sum_{k=1}^n \theta_k\, X(t_k+h)\} = \sum_{j=1}^n \sum_{k=1}^n \theta_j\theta_k\, c(t_j+h, t_k+h) =$

$\sum_{j=1}^n \sum_{k=1}^n \theta_j\theta_k\, c(t_k-t_j)$ for any $\theta_1,\ldots,\theta_n\in\mathbb{R}$, see the proof of the theorem on the top of the previous page. As these quantities do not depend on h the process is strongly stationary. □

The above established properties for Gaussian processes are very special and cannot at all be expected to hold for stochastic processes in general!

Example. If $X=\{X_n\}_{n\in\mathbb{N}}$ consists of iid. standardized normal distributed random variables, then by Example 9.3.19 X is stationary with mean function 0 and covariance function Kronecker's delta. This process is Gaussian since $\sum_{i=1}^n a_iX_i$ is zero-mean normal distributed with variance $\sum_{i=1}^n a_i^2$ for $a_1,\ldots,a_n\in\mathbb{R}$. Hence (unsurprisingly by Example 8.2.8), X is strongly stationary by Theorem 9.6.4. #

Example 9.6.10. (Ornstein-Uhlenbeck process) Let $\{W(t)\}_{t\geq 0}$ be a Wiener process with $\mathbf{Var}\{W(1)\}=\sigma^2$. For a constant $\alpha>0$, put $X(t)=\mathrm{e}^{-\alpha t}W(\mathrm{e}^{2\alpha t})$ for $t\in\mathbb{R}$. Then X is zero-mean Gaussian (as W is), as well as (strongly) stationary since it has covariance function

$$c(s,t)=\mathrm{e}^{-\alpha(s+t)}\sigma^2\min\{\mathrm{e}^{2\alpha s},\mathrm{e}^{2\alpha t}\}=\sigma^2\mathrm{e}^{-\alpha(\min\{s,t\}+\max\{s,t\})}\mathrm{e}^{2\alpha\min\{s,t\}}=\sigma^2\mathrm{e}^{-\alpha|t-s|}$$

(recall Section 8.5). A (stationary) zero-mean Gaussian process with this covariance function is called an *Ornstein*[12]*-Uhlenbeck*[13]*process.* #

Definition 9.6.5. *A stochastic process $X=\{X(t)\}_{t\in\mathbb{R}}$ is a* Markov process *if*

$$\mathbf{P}\big\{X(t)\in A\big|X(t_n)=x_n,\ldots,X(t_1)=x_1\big\}=\mathbf{P}\{X(t)\in A\,|\,X(t_n)=x_n\}$$

for all choices of $-\infty<t_1\leq\ldots\leq t_n\leq t<\infty$, $x_1,\ldots,x_n\in\mathbb{R}$ and $A\subseteq\mathbb{R}$.

Example 9.6.10. (continued) For the Ornstein-Uhlenbeck process we have

$$\mathbf{Cov}\big\{X(t)-\mathrm{e}^{-\alpha(t-t_n)}X(t_n),X(t_i)\big\}=\sigma^2\mathrm{e}^{-\alpha(t-t_i)}-\mathrm{e}^{-\alpha(t-t_n)}\sigma^2\mathrm{e}^{-\alpha(t_n-t_i)}=0$$

for $-\infty<t_1\leq\ldots\leq t_n\leq t<\infty$, so that $X(t)-\mathrm{e}^{-\alpha(t-t_n)}X(t_n)$ is independent of $X(t_1),\ldots,X(t_{n-1})$. From this we see that X is a Markov process as follows:

$$\begin{aligned}&\mathbf{P}\big\{X(t)\in A\big|X(t_n)=x_n,\ldots,X(t_1)=x_1\big\}\\&=\mathbf{P}\big\{\big(X(t)-\mathrm{e}^{-\alpha(t-t_n)}X(t_n)\big)+\mathrm{e}^{-\alpha(t-t_n)}X(t_n)\in A\big|X(t_n)=x_n,\ldots,X(t_1)=x_1\big\}\end{aligned}$$

[12]Leonard Salomon Ornstein, Dutch physicist 1880-1941. (Not the same person as Donald Samuel Ornstein, American mathematician born 1934 (advisor of Jeff Steiff).)

[13]George Eugene Uhlenbeck, Dutch-American theoretical physicist 1900-1988.

$$= \mathbf{P}\big\{\big(X(t)-\mathrm{e}^{-\alpha(t-t_n)}X(t_n)\big)+\mathrm{e}^{-\alpha(t-t_n)}X(t_n)\in A\big|X(t_n)=x_n\big\}$$
$$= \mathbf{P}\{X(t)\in A\,|\,X(t_n)=x_n\}. \qquad \#$$

Theorem. *A stationary zero-mean Gaussian process $\{X(t)\}_{t\in\mathbb{R}}$ is Markov if and only if it is an Ornstein-Uhlenbeck process.*

Proof. The implication to the left follows from Example 9.6.10. Conversely, if X is stationary zero-mean Gaussian with covariance function c, then we have

$$\mathbf{E}\{X(t+s)\,|\,X(s)=y\} = \mathbf{E}\bigg\{X(t+s)-\frac{c(t)}{c(0)}X(s)\,\bigg|\,X(s)=y\bigg\}+\frac{c(t)}{c(0)}y=\frac{c(t)}{c(0)}y$$

for $t\geq 0$, since $X(t+s)-(c(t)/c(0))\,X(s)$ is independent of $X(s)$ (as these random variables are uncorrelated). If in addition X is Markov, it follows that

$$\begin{aligned}
\frac{c(t+s)}{c(0)} &= \frac{\mathbf{E}\{X(t+s)X(0)\}}{c(0)}\\
&= \frac{1}{c(0)}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\mathbf{E}\{X(t+s)X(0)\,|\,X(0)=x,\,X(s)=y\}\,dF_{X(0),X(s)}(x,y)\\
&= \frac{1}{c(0)}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x\,\mathbf{E}\{X(t+s)\,|\,X(s)=y\}\,dF_{X(0),X(s)}(x,y)\\
&= \frac{c(t)}{c(0)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xy\,dF_{X(0),X(s)}(x,y)\\
&= \frac{c(t)}{c(0)^2}\,\mathbf{E}\{X(s)X(0)\}\\
&= \frac{c(t)}{c(0)}\,\frac{c(s)}{c(0)}\quad \text{for } s,t\geq 0.
\end{aligned}$$

Hence we must have $c(t)/c(0)=\mathrm{e}^{-\alpha t}$ for $t\geq 0$ for some constant $\alpha\geq 0$ (recall Theorem 6.9.13 and see Exercise below), so that $c(t)=c(0)\,\mathrm{e}^{-\alpha|t|}$ for $t\in\mathbb{R}$ by symmetry of c. Hence X is an Ornstein-Uhlenbeck process. □

Exercise. Show that if a stationary stochastic process $\{X(t)\}_{t\in\mathbb{R}}$ has covariance function $c(t)=c(0)\,\mathrm{e}^{-\alpha|t|}$ for $t\in\mathbb{R}$ for some constant $\alpha\in\mathbb{R}$, then it must hold that $\alpha\geq 0$.

Remeber that you must do two hand-in exercises for this lecture!

Lecture 8, Friday 12 February

Chapter 10 Renewals

10.1 The renewal equation

Definition 10.1.1. *Given a sequence* $\{X_i\}_{i=1}^{\infty}$ *of iid. non-negative random variables and writing* $T_n = \sum_{i=1}^{n} X_i$*, a* renewal process $N = \{N(t)\}_{t\geq 0}$ *is given by*

$$N(t) = \max\{n : T_n \leq t\} \quad \text{for } t \geq 0.$$

Theorem. *We have* $N(t) \geq n$ *if and only if* $T_n \leq t$*.*

Proof. The implication to the left is immediate from Definition 10.1.1. Conversely, if $T_n > t$, then Definition 10.1.1 shows that $N(t) < n$. □

Theorem 10.1.2. *If* $\mathbf{E}\{X_1\} > 0$ *then* $\mathbf{P}\{N(t) < \infty\} = 1$*.*

Proof. If $\mathbf{E}\{X_1\} > 0$ then $\mathbf{P}\{X_1 \geq \delta\} \geq \varepsilon$ for some selection of $\delta, \varepsilon > 0$, so that

$$\mathbf{P}\{N(t) < n\} = \mathbf{P}\{T_n > t\} = 1 - \mathbf{P}\{T_n \leq t\} \geq 1 - \mathbf{P}\left\{\mathrm{Bin}\,(n, \varepsilon) \leq \frac{t}{\delta}\right\} \to 1$$

as $n \to \infty$ (see Exercise below). □

Exercise. Explain the last two steps in the equation in the proof of Theorem 10.1.2.

Henceforth we always assume not only that $\mu = \mathbf{E}\{X_1\}$ is strictly positive (and finite), but that X_1 is strictly positive, which is to say that $\mathbf{P}\{X_1 > 0\} = 1$.

Example 10.1.15. A Poisson process is a renewal process with X_1 exponentially distributed. Recall from Theorem 8.3.5 that this is the only renewal process that is a Markov chain. #

Definition. *The distribution functions of* X_1 *and* T_n *are denoted* F *and* F_n*, respectively.*

Lemma 10.1.4. $F_{n+1}(x) = \int_0^x F_n(x-y)\, dF(y)$ *for* $x > 0$ *and* $n \geq 1$*.*

Proof. We have

$$\begin{aligned} F_{n+1}(x) &= \mathbf{P}\{T_{n+1} \leq x\} \\ &= \mathbf{P}\{T_n + X_{n+1} \leq x\} \\ &= \int_0^x \mathbf{P}\{T_n + y \leq x\}\, dF(y) \\ &= \int_0^x F_n(x-y)\, dF(y). \end{aligned}$$ □

Lemma 10.1.5. $\mathbf{P}\{N(t) = n\} = F_n(t) - F_{n+1}(t)$ *for* $t > 0$ *and* $n \geq 1$.

Proof. We have

$$\begin{aligned} \mathbf{P}\{N(t) = n\} &= \mathbf{P}\{N(t) \geq n\} - \mathbf{P}\{N(t) \geq n+1\} \\ &= \mathbf{P}\{T_n \leq t\} - \mathbf{P}\{T_{n+1} \leq t\} \\ &= F_n(t) - F_{n+1}(t). \end{aligned}$$ □

Definition 10.1.6. *The* renewal function m *is given by* $m(t) = \mathbf{E}\{N(t)\}$.

Lemma 10.1.7. $m(t) = \sum_{n=1}^{\infty} F_n(t)$ *for* $t > 0$.

Proof. We have

$$\begin{aligned} \mathbf{E}\{N(t)\} &= \sum_{k=1}^{\infty} k\, \mathbf{P}\{N(t) = k\} \\ &= \sum_{k=1}^{\infty} \sum_{n=1}^{k} \mathbf{P}\{N(t) = k\} \\ &= \sum_{n=1}^{\infty} \sum_{k=n}^{\infty} \mathbf{P}\{N(t) = k\} \\ &= \sum_{n=1}^{\infty} \mathbf{P}\{N(t) \geq n\} \\ &= \sum_{n=1}^{\infty} \mathbf{P}\{T_n \leq t\} \\ &= \sum_{n=1}^{\infty} F_n(t). \end{aligned}$$ □

Lemma 10.1.8. $m(t) = F(t) + \int_0^t m(t-x)\, dF(x)$ *for* $t > 0$.

Proof. We have

$$\begin{aligned}\mathbf{E}\{N(t)\} &= \int_0^\infty \mathbf{E}\{N(t)\,|\,X_1 = x\}\,dF(x)\\ &= 0 + \int_0^t \mathbf{E}\{N(t-x)+1\}\,dF(x)\\ &= \int_0^t m(t-x)\,dF(x) + F(t).\end{aligned}$$ □

Lemma 10.1.8 can be expressed in terms of convolution language as

$$\boxed{m = F + m\star F}\,.$$

Definition 10.1.10. *Given a bounded function H we call the equation*

$$\mu(t) = H(t) + \int_0^t \mu(t-x)\,dF(x) \quad \textit{for } t>0$$

a renewal-type equation *(where a solution μ is sought after).*

Note that the renewal-type equation in Definition 10.1.10 can be expressed as

$$\boxed{\mu = H + \mu\star F}\,.$$

Theorem 10.1.11. *The renewal-type equation in Definition 10.1.10 has solution $\mu = H + H\star m$.*

Proof. As $m = F + m\star F$ by Lemma 10.1.8 it follows that $\mu = H + H\star m$ satisfies

$$\mu = H + H\star m = H + H\star(F+m\star F) = H + (H+H\star m)\star F = H + \mu\star F.$$ □

Example 10.1.15. A Poisson process $\{N(t)\}_{t\geq 0}$ with intensity λ has expected value $\mathbf{E}\{N(t)\} = \lambda t$. We may recover this result using that N is a renewal process with X_1 exponentially distributed with parameter λ, so that T_n is $\Gamma(n,\lambda)$-distributed and Lemma 10.1.7 gives

$$m(t) = \sum_{n=1}^\infty \mathbf{P}\{T_n\leq t\} = \sum_{n=1}^\infty \int_0^t \frac{\lambda\,(\lambda s)^{n-1}}{(n-1)!}\,\mathrm{e}^{-\lambda s}\,ds = \int_0^t \lambda\,ds = \lambda t.$$ #

10.2 Limit theorems

Theorem 10.2.1. *We have $N(t)/t \to 1/\mu$ as $t\to\infty$.*

Proof. By the strong law of large numbers we have $\sum_{i=1}^n X_i/n \to \mu$ as $n\to\infty$. It follows that

$$\begin{aligned}\frac{N(t)}{t} &= \frac{\max\{n : \sum_{i=1}^n X_i \le t\}}{t}\\ &= \max\left\{\frac{n}{t} : \frac{1}{n}\sum_{i=1}^n X_i \le \frac{t}{n}\right\}\\ &\to \max\left\{\frac{n}{t} : \mu \le \frac{t}{n}\right\}\\ &= \frac{1}{\mu} \quad \text{as } t\to\infty. \qquad \square\end{aligned}$$

Theorem 10.2.2. *If* $\sigma^2 = \mathbf{Var}\{X_1\} < \infty$, *then*

$$\mathbf{P}\left\{\frac{N(t)-t/\mu}{\sqrt{t\sigma^2/\mu^3}} \le x\right\} \to \Phi(x) \quad \textit{as } t\to\infty \ \textit{ for } x\in\mathbb{R}.$$

Proof. Noting that

$$s = \frac{t}{\mu} + \frac{x\sigma\sqrt{t}}{\sqrt{\mu^3}} \implies t = \mu s - x\sigma\sqrt{s} + o(\sqrt{s}) \quad \text{as } s,t\to\infty$$

the central limit theorem gives

$$\begin{aligned}\mathbf{P}\left\{\frac{N(t)-t/\mu}{\sqrt{t\sigma^2/\mu^3}} \le x\right\} &= \mathbf{P}\left\{N(t) \le \frac{t}{\mu} + \frac{x\sigma\sqrt{t}}{\sqrt{\mu^3}}\right\}\\ &= \mathbf{P}\big\{N\big(\mu s - x\sigma\sqrt{s} + o(\sqrt{s})\big) \le s\big\}\\ &= \mathbf{P}\big\{T_s \ge \mu s - x\sigma\sqrt{s} + o(\sqrt{s})\big\}\\ &= \mathbf{P}\left\{\frac{T_s-\mu s}{\sigma\sqrt{s}} \ge -x + o(\sqrt{1})\right\}\\ &\to 1-\Phi(-x)\\ &= \Phi(x) \quad \text{as } s,t\to\infty. \qquad \square\end{aligned}$$

Definition 10.2.4. *A random variable X is* arithmetic with span $\lambda > 0$ *if X takes values in the set $\{n\lambda : n\in\mathbb{Z}\}$ and λ is the maximal real number with this property.*

The proof of the following important result is long and complicated and must therefore be omitted.

Theorem 10.2.5. (Renewal theorem) *If X_1 is not arithmetic, then*

$$m(t+h) - m(t) \to \frac{h}{\mu} \quad \textit{as } t\to\infty \ \textit{ for } h>0.$$

If X_1 is arithmetic with span λ, then the above limit holds for h a multiple of λ.

Proof. Omitted. □

Exercise. Construct a counter example to the fisrt part of the renewal theorem for X_1 arithmetic.

Theorem 10.2.3. (Elementary renewal theorem)

$$\frac{m(t)}{t} \to \frac{1}{\mu} \quad \textit{as } t \to \infty.$$

Proof. Let λ be the span of X_1 if X_1 is arithmetic and take $\lambda = 1$ otherwise. By the renewal theorem (see also Exercise below) we have

$$\frac{m(t)}{t} = \frac{1}{t}\sum_{k=1}^{t/\lambda}\big(m(k\lambda) - m((k-1)\lambda)\big) \to \frac{1}{\lambda}\frac{\lambda}{\mu} = \frac{1}{\mu} \quad \text{as } t \to \infty.$$ □

Exercise. Attend to the technical details of the proof of Theorem 10.2.3.

Theorem 10.2.7. (Key renewal theorem) *If X_1 is not arithmetic and $g : [0,\infty) \to [0,\infty)$ is a non-increasing integrable function, then*

$$\int_0^t g(t-x)\,dm(x) \to \frac{1}{\mu}\int_0^\infty g(x)\,dx \quad \textit{as } t \to \infty.$$

Proof. Note that $\int_0^t g(t-x)\,dm(x) = \int_0^t g(x)\,dm(t-x)$, where the renewal theorem states that $dm(t-x)$ works like dx/μ for large t. □

10.3 Excess life

We start this section by making the following elementary but important observation:

Theorem 10.2.8. $T_{N(t)} \le t \le T_{N(t)+1}$ *for* $t > 0$.

Proof. The number of jumps up to inclduing time t is exactly $N(t)$. This in turn is exactly the claim of the theorem but phrased in other words. □

Definition 10.3.1. **(a)** *The* excess lifetime $E(t)$ *at time* $t > 0$ *is* $E(t) = T_{N(t)+1} - t$.
(b) *The* current lifetime $C(t)$ *at time* $t > 0$ *is* $C(t) = t - T_{N(t)}$.
(c) *The* total lifetime $D(t)$ *at time* $t > 0$ *is* $D(t) = E(t) + C(t) = X_{N(t)+1}$.

Theorem 10.3.3. $\mathbf{P}\{E(t)\le y\} = F(t+y) - \int_0^t (1-F(t+y-x))\,dm(x)$ *for* $t,y>0$.

Proof. As

$$\mathbf{P}\{E(t)>y\,|\,X_1=x\} = \begin{cases} \mathbf{P}\{E(t-x)>y\} & \text{for } 0<x<t,\\ 0 & \text{for } 0<t<x\le t+y,\\ 1 & \text{for } 0<t<t+y<x,\end{cases}$$

we have

$$\begin{aligned}\mathbf{P}\{E(t)>y\} &= \int_0^\infty \mathbf{P}\{E(t)>y\,|\,X_1=x\}\,dF(x)\\ &= \int_0^t \mathbf{P}\{E(t-x)>y\}\,dF(x) + \int_{t+y}^\infty dF(x)\\ &= \int_0^t \mathbf{P}\{E(t-x)>y\}\,dF(x) + (1-F(t+y)).\end{aligned}$$

It follows that the function $\mathbf{P}\{E(t)>y\}$ satisfies a renewal-type equation with $H(t) = 1-F(t+y)$. Hence Theorem 10.1.11 shows that

$$\mathbf{P}\{E(t)>y\} = \int_0^t (1-F(t+y-x))\,dm(x) + (1-F(t+y)).$$ □

Corollary 10.3.4.

$$\mathbf{P}\{C(t)\ge y\} = \begin{cases} 1-F(t) + \int_0^{t-y}(1-F(t-x))\,dm(x) & \textit{for } 0<y<t,\\ 0 & \textit{for } 0<t\le y.\end{cases}$$

Proof. Use Theorem 10.3.3 together with the fact that

$$\mathbf{P}\{C(t)\ge y\} = \begin{cases} \mathbf{P}\{E(t-y)\ge y\} & \text{for } 0<y<t,\\ 0 & \text{for } 0<t\le y.\end{cases}$$ □

Theorem 10.3.5. *If* X_1 *is not arithmetic, then*

$$\mathbf{P}\{E(t)\le y\} \to \frac{1}{\mu}\int_0^y (1-F(x))\,dx \quad \textit{as } t\to\infty \ \textit{for } y>0.$$

Proof. By application of the key renewal theorem with $g(x) = 1-F(x+y)$ to the statement of Theorem 10.3.3 we obtain

$$\mathbf{P}\{E(t)\le y\} \to 1 - \frac{1}{\mu}\int_0^\infty (1-F(x+y))\,dx \quad \text{as } t\to\infty.$$

The claim now follows readily from the elmentary fact that $\mu = \int_0^\infty (1-F(x))\,dx$. □

Lecture 9, Friday 19 February

10.4 Applications

Example 10.4.1. (Dead periods) Let $N(t)$ be a renewal process with excess life-time process $E(t)$. Assume that the renewal process is locked a time L_i after the i'th jump of N, so that no new jumps can be made during that locking time. Here $\{L_i\}_{i=0}^{\infty}$ are iid. random variables that are independent of N and jumps during locking times are simply neglected. The process $\tilde{N} = \{\tilde{N}(t)\}_{t\geq 0}$ that results from this is started up with an initial locking time L_0. Writing $\{\tilde{X}_i\}_{i=1}^{\infty}$ for the interarrival times of $\tilde{N}$ Theorem 10.3.3 gives

$$\begin{aligned}
\mathbf{P}\{\tilde{X}_1 \leq x\} &= \int_0^x \mathbf{P}\big\{L_0 + E(L_0) \leq x \,\big|\, L_0 = \ell\big\}\, dF_L(\ell) \\
&= \int_0^x \mathbf{P}\{E(\ell) \leq x - \ell\}\, dF_L(\ell) \\
&= \int_0^x F(\ell + x - \ell)\, dF_L(\ell) - \int_0^x \bigg(\int_0^\ell (1 - F(\ell + x - \ell - z))\, dm(z)\bigg) dF_L(\ell) \\
&= F(x)F_L(x) - \int_0^x m(\ell) dF_L(\ell) + \int_0^x \bigg(\int_0^{\ell\wedge x} F(x - z)\, dm(z)\bigg) dF_L(\ell) \\
&= F(x)F_L(x) - \int_0^x m(\ell) dF_L(\ell) + \int_0^x \bigg(\int_z^x F(x - z)\, dF_L(\ell)\bigg) dm(z) \\
&= F(x)F_L(x) - \int_0^x m(\ell) dF_L(\ell) + \int_0^x F(x - z)\,(F_L(x) - F_L(z))\, dm(z).
\end{aligned}$$

Here Lemma 10.1.8 shows that $\int_0^x F(x-z)\, dm(z) = m(x) - F(x)$, so that we get

$$\begin{aligned}
&\mathbf{P}\{\tilde{X}_1 \leq x\} \\
&= F(x)F_L(x) - \int_0^x m(\ell) dF_L(\ell) + F_L(x)\,(m(x) - F(x)) - \int_0^x F(x - z) F_L(z)\, dm(z) \\
&= \int_0^x F_L(\ell)\, dm(\ell) - \int_0^x F(x - z) F_L(z)\, dm(z) \\
&= \int_0^x (1 - F(x - z))\, F_L(\ell)\, dm(\ell).
\end{aligned}$$

Now, this is just the distribution of $\tilde{X}_1$, where in general the interarrival times $\{\tilde{X}_i\}_{i=1}^{\infty}$ are neither independent nor identically distributed, so more work is required to find the distribution of the n'th arrival $\tilde{T}_n = \tilde{X}_1 + \ldots + \tilde{X}_n$. However, when N is a Poisson process the lack of memory property for the exponential distribution makes $\tilde{N}$ a renewal process with interarrival times $\{L_{i-1} + X_i\}_{i=1}^{\infty}$. #

Example 10.4.4. (Alternating renewal process) A machine breaks down repeatedly, where after the n'th breakdown it is repaired a time Y_n after which it runs a time Z_n before its $n+1$'th breakdown. Here $\{Y_n\}_{n=1}^{\infty}$ and $\{Z_n\}_{n=0}^{\infty}$ are independent iid. sequences of random variables, and the renewal process N with interarrival times $\{Y_n+Z_{n-1}\}_{n=1}^{\infty}$ counts the number of start ups after repair of the machine up to time t.

Lemma 10.4.5. *For an alternating renewal process the probability $p(t)$ that the machine works at time t is given by*

$$p(t) = 1 - F_Z(t) + \int_0^t p(t-x)\,dF(x) = 1 - F_Z(t) + \int_0^t (1 - F_Z(t-x))\,dm(x).$$

Proof. The second equation follows from the first using Theorem 10.1.11. Further,

$$\begin{aligned} p(t) &= \mathbf{P}\{\text{work at time } t,\ Z_0 > t\} + \mathbf{P}\{\text{work at time } t,\ Z_0 \le t\} \\ &= \mathbf{P}\{Z_0 > t\} + \int_0^\infty \mathbf{P}\{\text{work at time } t,\ Z_0 \le t \,|\, X_1 = x\}\,dF(x) \\ &= 1 - F_Z(t) + \int_0^t \mathbf{P}\{\text{work at time } t,\ Z_0 \le t \,|\, X_1 = x\}\,dF(x) \\ &= 1 - F_Z(t) + \int_0^t p(t-x)\,dF(x). \end{aligned}$$ □

Corollary 10.4.6. *For an alternating renewal process such that Y_1+Z_0 is not arithmetic we have*

$$p(t) = \frac{1}{1 + \mathbf{E}\{Y_1\}/\mathbf{E}\{Z_0\}} \quad \text{as } t \to \infty.$$

Proof. By the key renewal theorem we have

$$\int_0^t g(t-x)\,dm(x) \to \frac{1}{\mu}\int_0^\infty g(x)\,dx \quad \text{as } t \to \infty.$$

Using this together with Lemma 10.4.5 we get

$$p(t) \to 0 + \frac{1}{\mu}\int_0^\infty (1 - F_Z(x))\,dx = \frac{\mathbf{E}\{Z_0\}}{\mathbf{E}\{Y_1\} + \mathbf{E}\{Z_0\}} \quad \text{as } t \to \infty.$$ □

Example 10.4.7. (Superposition of renewal processes) Let $\tilde{N}(t) = N_1(t) + N_2(t)$ where N_1 and N_2 are iid. renewal processes.

Theorem 10.4.8. *$\tilde{N}$ is a renewal process if and only if N_1 and N_2 are Poisson processes.*

Proof. The implication to the left is elementary while that to the right is omitted. □

Definition 10.4.12. *A (version of a) renewal process* $N^d = \{N^d(t)\}_{t\geq 0}$ *where the first interarrival time* X_1 *has another distribution function* F^d *than the following ones* $\{X_i\}_{i=2}^{\infty}$ *is called a* delayed renewal process.

Theorem. *For a delayed renewal process we have*

$$m^d(t) \equiv \mathbf{E}\{N^d(t) = F^d(t) + \int_0^t m(t-x)\, dF^d(x).$$

Proof. By analogy with the proof of Lemma 10.1.8 we have

$$\begin{aligned} m^d(t) &= \int_0^\infty \mathbf{E}\{N^d(t)\,|\,X_1 = x\}\, dF^d(x) \\ &= 0 + \int_0^t \mathbf{E}\{N(t-x)+1\}\, dF^d(x) \\ &= \int_0^t m(t-x)\, dF^d(x) + F^d(t). \end{aligned}$$

□

Theorem 10.4.15. *For a delayed renewal process we have*

$$\frac{m^d(t)}{t} \to \frac{1}{\mu} \quad \text{as } t \to \infty.$$

If X_2 *is not arithmetic the we further have*

$$m^d(t+h) - m^d(t) \to \frac{h}{\mu} \quad \text{as } t \to \infty \text{ for } h \geq 0.$$

If X_2 *is arithmetic with span* λ *then the above limit holds for* h *a mutiple of* λ.

Proof. The time X_1 to the first jump of N^d does not affect limit properties of $N^d(t)$ as $t \to \infty$, which are thus the same as those of N. □

Theorem 10.4.17. *A delayed renewal process* N^d *has stationary increments if and only if*

$$F^d(y) = \frac{1}{\mu}\int_0^y (1 - F(x))\, dx \quad \text{for } y \geq 0.$$

Proof. The implication to the left is omitted. For that to the right we note that if N^d has stationary increments, then we must have $m^d(t+s) = m^d(t) + m^d(s)$, so that $m^d(t) = ct$ for some constant $c \geq 0$. As a previous theorem gives

$$m^d(t) = F^d(t) + \int_0^t F^d(t-x)\, dm(x),$$

we might use Theorem 10.1.11 to conclude that

$$F^d(t) = m^d(t) - \int_0^t m^d(t-x)\,dF(x) = m^d(t) - \int_0^t F(t-x)\,dm^d(x) = c\int_0^t (1-F(x))\,dx,$$

where we get $c = 1/\mu$ by sending $t \to \infty$. $\square$

10.5 Renewal-reward processes

Definition. *Given a sequence* $\{(X_i, R_i)\}_{i=1}^{\infty}$ *of pairs of iid. random variables such that* $\mathbf{P}\{X_1 > 0\} = 1$ *and writing* $T_n = \sum_{i=1}^{n} X_i$, *a* renewal-reward process $C = \{C(t)\}_{t\geq 0}$ *is given by*

$$C(t) = \sum_{i=1}^{N(t)} R_i \quad \textit{where} \quad N(t) = \max\{n : T_n \leq t\} \quad \textit{for } t \geq 0.$$

Definition. *The* reward function c *of a renewal-reward process* C *is given by* $c(t) = \mathbf{E}\{C(t)\}$.

Theorem 10.5.1. (Renewal-reward theorem) *If* $\mathbf{E}\{X_1\} < \infty$ *and* $\mathbf{E}\{|R_1|\} < \infty$, *then we have*

$$\frac{C(t)}{t} \to \frac{\mathbf{E}\{R_1\}}{\mathbf{E}\{X_1\}} \quad \textit{and} \quad \frac{c(t)}{t} \to \frac{\mathbf{E}\{R_1\}}{\mathbf{E}\{X_1\}} \quad \textit{as } t \to \infty.$$

Proof. By Theorem 10.2.1 together with the strong law of large numbers we have

$$\frac{C(t)}{t} = \frac{1}{t}\sum_{i=1}^{N(t)} R_i = \frac{N(t)}{t}\frac{1}{N(t)}\sum_{i=1}^{N(t)} R_i \to \frac{1}{\mathbf{E}\{X_1\}}\mathbf{E}\{R_1\} \quad \text{as } t \to \infty.$$

Since the event that $\{N(t) \geq i-1\} = \{X_1 + \ldots + X_{i-1} \leq t\}$ is independent of R_i the elementary renewal Theorem 10.2.3 further gives that

$$\begin{aligned}
\frac{c(t)}{t} &= \frac{1}{t}\mathbf{E}\Big\{\sum_{i=1}^{N(t)} R_i\Big\} \\
&= \frac{1}{t}\mathbf{E}\Big\{\sum_{i=1}^{N(t)+1} R_i\Big\} - \frac{\mathbf{E}\{R_{N(t)+1}\}}{t} \\
&= \frac{1}{t}\mathbf{E}\Big\{\sum_{i=1}^{\infty} \mathbf{1}_{\{N(t)+1\geq i\}} R_i\Big\} - \frac{\mathbf{E}\{R_{N(t)+1}\}}{t}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{t}\sum_{i=1}^{\infty}\mathbf{E}\{\mathbf{1}_{\{N(t)+1\geq i\}}R_i\} - \frac{\mathbf{E}\{R_{N(t)+1}\}}{t}\\
&= \frac{\mathbf{E}\{R_1\}}{t}\sum_{i=1}^{\infty}\mathbf{P}\{N(t)+1\geq i\} - \frac{\mathbf{E}\{R_{N(t)+1}\}}{t}\\
&= \frac{\mathbf{E}\{R_1\}\,\mathbf{E}\{N(t)+1\}}{t} - \frac{\mathbf{E}\{R_{N(t)+1}\}}{t}\\
&\to \frac{\mathbf{E}\{R_1\}}{\mathbf{E}\{X_1\}} + 0 \quad \text{as } t\to\infty,
\end{aligned}
$$

provided that $\mathbf{E}\{R_{N(t)+1}\}/t \to 0$. The proof that the latter limit relation holds is trivial for X_i and R_i independent, but is omitted for X_i and R_i dependent. □

In the particular case when the X_i's are exponentially distributed and independent of the R_i's, so that in particular N is a Poisson process, we call the corresponding renewal reward process C a *compound Poisson process*.

Unsurprisingly, one can do a lot of complicated as well as less complicated calculations for a renewal reward process C. We leave it as homework to the reader to study enough of Section 10.5 in order to be able to do one hand-in exercise for that section.

Lecture 10, Thursday 25 February

Chapter 11 Queues

11.1 Single server queues

A *single server queue* is a system specified by two independent sequences of non-negative iid. *inter arrival times* $\{X_n\}_{n=1}^{\infty}$ and *service times* $\{S_n\}_{n=1}^{\infty}$, respectively. An arriving customer either goes directly to the server, if the server is free, or joins the queue otherwise. When the server finishes serving a customer he (or she) picks the next customer from the queue if the queue is not empty, or otherwise becomes idle until the next customer arrives. We write $Q(t)$ for the total number of customers in the queuing system at time $t \geq 0$ (i.e., the difference between the number of customers that have arrived and the number of customers that have been served).

A single server queue is conveniently denoted A/B/s, where A is the distribution of inter arrival times, B the distribution of service times and s the number of servers, that is, $s = 1$. Important examples of common choices of A and B include

- $\mathrm{D}(d)$ = deterministic with value d;
- $\mathrm{M}(\lambda)$ = exponentially distributed with parameter λ;
- $\Gamma(n, \lambda)$ = gamma distributed;
- G = some fixed but unspecified general distribution.

Here M stands for Markov, the explanation of which will become evident later.

11.2 M/M/1

The $\mathrm{M}(\lambda)/\mathrm{M}(\mu)/1$ queue simply is a birth-death process with birth intensities $\lambda_n = \lambda$ and death intensities $\mu_n = \mu$, recall Section 6.11.

Theorem 11.2.3. *For a* $\mathrm{M}(\lambda)/\mathrm{M}(\mu)/1$ *queue the probability*

$$p_n(t) = \mathbf{P}\{Q(t) = n\} \quad \textit{for } t \geq 0 \textit{ and } n \in \mathbb{N}$$

has Laplace transform

$$\hat{p}_n(\theta) \equiv \int_0^{\infty} \mathrm{e}^{-\theta t} p_n(t)\, dt = \frac{\alpha(\theta)^n}{\lambda + \theta - \mu\, \alpha(\theta)} \quad \textit{for } \theta > 0 \textit{ and } n \in \mathbb{N},$$

where

$$\alpha(\theta) = \frac{\lambda + \mu + \theta - \sqrt{(\lambda + \mu + \theta)^2 - 4\lambda\mu}}{2\mu} \quad \textit{for } \theta > 0.$$

Proof. Let P_t and G denote the transition matrix and generator, respectively, for the above mentioned birth-death process. Note that $\mu^{(0)}$ is the one-point distribution at zero, so that $Q_t = \mu^{(0)} P_t = (p_0(t)\ \ p_1(t)\ \ \dots\ )$ satisfies $Q_t' = Q_t G$ by the Kolmogorov forward equations. Spelled out explicitely that system of equations in turn become

$$\begin{cases} p_n'(t) = \lambda\, p_{n-1}(t) - (\lambda+\mu)\, p_n(t) + \mu\, p_{n+1}(t) & \text{for } n \geq 1, \\ p_0'(t) = -\lambda\, p_0(t) + \mu\, p_1(t). \end{cases}$$

Noting that

$$\int_0^\infty e^{-\theta t} p_n'(t)\, dt = \left[e^{-\theta t} p_n(t)\right]_{t=0}^\infty + \theta \int_0^\infty e^{-\theta t} p_n(t)\, dt = -\delta(n) + \theta\, \hat{p}_n(\theta),$$

we may Laplace transform the above system of equations to obtain

$$\begin{cases} \lambda\, \hat{p}_{n-1}(\theta) - (\lambda+\mu+\theta)\, \hat{p}_n(\theta) + \mu\, \hat{p}_{n+1}(\theta) = 0 & \text{for } n \geq 1, \\ (\lambda+\theta)\, \hat{p}_0(\theta) - \mu\, \hat{p}_1(\theta) = 1. \end{cases}$$

From difference equation techniques it is known that the first of these equations has two solutions, namely

$$\hat{p}_n(\theta) = \left(\frac{\lambda+\mu+\theta - \sqrt{(\lambda+\mu+\theta)^2 - 4\lambda\mu}}{2\mu}\right)^n \hat{p}_0(\theta) = \alpha(\theta)^n\, \hat{p}_0(\theta) \quad \text{for } n \geq 1$$

and

$$\hat{p}_n(\theta) = \left(\frac{\lambda+\mu+\theta + \sqrt{(\lambda+\mu+\theta)^2 - 4\lambda\mu}}{2\mu}\right)^n \hat{p}_0(\theta) \quad \text{for } n \geq 1.$$

Here the second solution must be invalid as the value of $\hat{p}_n(\theta)$ because $\hat{p}_n(\theta) \leq 1/\theta$ for $n \geq 0$, while the second solution goes to ∞ as $n \to \infty$ for $\theta > 0$ sufficiently large. (As $\hat{p}_0$ is an analytic function it can not be zero for large θ.) Now we may insert $\hat{p}_1(\theta) = \alpha(\theta)\, \hat{p}_0(\theta)$ in the equation $(\lambda+\theta)\, \hat{p}_0(\theta) - \mu\, \hat{p}_1(\theta) = 1$ to obtain $\hat{p}_0(\theta) = 1/(\lambda+\theta-\mu\, \alpha(\theta))$. This proves the claim of the theorem. □

Definition. *The* traffic intensity ρ *of a queue is given by* $\rho = \mathbf{E}\{S\}/\mathbf{E}\{X\}$.

Unsurprisingly the qualitative behaviour of an $\mathrm{M}(\lambda)/\mathrm{M}(\mu)/1$ queue depends on whether $\rho < 1$ or $\rho \geq 1$:

Theorem 11.2.8. *Consider a* $\mathrm{M}(\lambda)/\mathrm{M}(\mu)/1$ *queue with traffic intensity* $\rho = \lambda/\mu$.
(a) *If* $\rho < 1$ *then* $\mathbf{P}\{Q(t) = n\} \to \pi_n$ *as* $t \to \infty$ *where* $\pi_n = (1-\rho)\, \rho^n$ *is the unique stationary distribution.*
(b) *If* $\rho \geq 1$ *then* $\mathbf{P}\{Q(t) = n\} \to 0$ *as* $t \to \infty$ *and there exists no stationary distribution.*

Proof. This follows from the limit Theorem 6.9.21 for continuous time Markov chains together with the second last example of Section 6.11. □

Let $\{U_n\}_{n=1}^{\infty}$ be the sequence of times at which the number of customers $Q(t)$ in a $\mathrm{M}(\lambda)/\mathrm{M}(\mu)/1$ queue changes. Setting $Q_n = Q(U_n^+)$ for $n \geq 1$ and $Q_0 = 0$ we see that $\{Q_n\}_{n=0}^{\infty}$ is a discrete time Markov chain such that

$$Q_{n+1} = \begin{cases} Q_n + 1 & \text{with probability } \dfrac{\lambda}{\lambda+\mu} = \dfrac{\rho}{1+\rho} \\ Q_n - 1 & \text{with probability } \dfrac{\mu}{\lambda+\mu} = \dfrac{1}{1+\rho} \end{cases} \quad \text{for } Q_n \geq 1,$$

while $Q_{n+1} = 1$ for $Q_n = 0$. In other words, $\{Q_n\}_{n=0}^{\infty}$ is a random walk on the non-negative integers with a reflecting barrier at zero.

Looking for a stationary distribution for the Markov chain $\{Q_n\}_{n=0}^{\infty}$ we study the system of equations $\pi P = \pi$, which spelled out becomes

$$\begin{cases} \dfrac{1}{1+\rho}\,\pi_1 = \pi_0, \\ \pi_0 + \dfrac{1}{1+\rho}\,\pi_2 = \pi_1, \\ \dfrac{\rho}{1+\rho}\,\pi_{n-1} + \dfrac{1}{1+\rho}\,\pi_{n+1} = \pi_n \quad \text{for } n \geq 2. \end{cases}$$

From difference equation techniques it is known that the last of these equations has two solutions, namely

$$\pi_n = \pi_1 \quad \text{for } n \geq 2 \quad \text{and} \quad \pi_n = \rho^{n-1}\pi_1 \quad \text{for } n \geq 2.$$

In order for π to be a stationary distribution we must disregard the first of these solution, while the second solution will give a stationary distribution if and only if $\rho < 1$. In the latter case we may insert $\pi_2 = \rho\,\pi_1$ in the first two equations of the system of equations $\pi P = \pi$ to see that they reduce to $\pi_1 = (1+\rho)\,\pi_0$. Using that

$$\sum_{n=0}^{\infty} \pi_n = \pi_0 + (1+\rho)\,\pi_0 + \sum_{n=2}^{\infty}(1+\rho)\,\rho^{n-1}\pi_0 = \ldots = \frac{2}{1-\rho}\,\pi_0,$$

we conclude that the chain is non-null persistent with stationary distribution

$$\pi_0 = \frac{1-\rho}{2} \quad \text{and} \quad \pi_n = \frac{1-\rho^2}{2}\,\rho^{n-1} \quad \text{for } n \geq 1$$

when $\rho < 1$. Unsurprisingly, it can be shown that the chain is null persistent for $\rho = 1$, while it is transistent for $\rho > 1$. We leave this as an exercise to the reader.

Exercise. (Difficult) Prove that the Markov chain $\{Q_n\}_{n=0}^{\infty}$ is null persistent when $\rho = 1$ and transistent for $\rho > 1$.

The chain $\{Q_n\}_{n=0}^{\infty}$ is called an *imbedded Markov chain*, and we shall see in the following two sections that such imbedding techniques are important for the study of M/G/1 and G/M/1 queues.

11.3 M/G/1

Let D_n be the time at which the n'th customer departs from a $\mathrm{M}(\lambda)/\mathrm{G}/1$ queue for $n \geq 1$. Set $Q_n = Q(D_n^+)$ for $n \geq 1$ and $Q_0 = 0$.

Theorem 11.3.4. *For a* $\mathrm{M}(\lambda)/\mathrm{G}/1$ *queue the sequence* $\{Q_n\}_{n=0}^{\infty}$ *is a discrete time Markov chain with transition matrix*

$$P = \begin{pmatrix} \delta_0 & \delta_1 & \delta_2 & \delta_3 & \delta_4 & \cdots \\ \delta_0 & \delta_1 & \delta_2 & \delta_3 & \delta_4 & \cdots \\ 0 & \delta_0 & \delta_1 & \delta_2 & \delta_3 & \cdots \\ 0 & 0 & \delta_0 & \delta_1 & \delta_2 & \cdots \\ 0 & 0 & 0 & \delta_0 & \delta_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

where δ *is the probability distribution on* $\mathbb{N}$ *with entries*

$$\delta_j = \mathbf{E}\bigg\{\frac{(\lambda S)^j}{j!}\,\mathrm{e}^{-\lambda S}\bigg\} \quad \textit{for } j \in \mathbb{N}.$$

Proof. Writing A_n for the number of customers that arrive during the service time of the n'th customer, we have

$$Q_{n+1} = Q(D_{n+1}^+) = \begin{cases} A_{n+1} + Q(D_n^+) - 1 & \text{if } Q_n > 0, \\ A_{n+1} & \text{if } Q_n = 0. \end{cases}$$

It follows that

$$\mathbf{P}\{Q_{n+1} = j \,|\, Q_n = i\} = \begin{cases} \mathbf{P}\{A_{n+1} = j - i + 1\} & \text{for } j \geq i > 0, \\ \mathbf{P}\{A_{n+1} = j\} & \text{for } j \geq i = 0. \end{cases}$$

As each A_n is $\mathrm{Po}(\lambda S)$ distributed the claim of the theorem follows. □

Theorem 11.3.5. *Consider a* $\mathrm{M}(\lambda)/\mathrm{G}/1$ *queue with traffic intensity* $\rho = \lambda\,\mathbf{E}\{S\}$*.*
(a) *If* $\rho < 1$ *then the chain* $\{Q_n\}_{n=0}^{\infty}$ *is ergodic with a unique stationary distribution* π *having generating function*

$$G_\pi(s) = \sum_{j=0}^{\infty} s^j\,\pi_j = (1-\rho)\,(s-1)\,\frac{M_S(\lambda\,(s-1))}{s - M_S(\lambda\,(s-1))} \quad \textit{for } s \in [0,1),$$

where $M_S(t) = \mathbf{E}\{\mathrm{e}^{tS}\}$ *is the moment generating function of* S*.*
(b) *If* $\rho = 1$ *then the chain* $\{Q_n\}_{n=0}^{\infty}$ *is null persistent.*
(c) *If* $\rho > 1$ *then the chain* $\{Q_n\}_{n=0}^{\infty}$ *is transient.*

Proof. We omit the proof of claims b and c, but note that they are very natural and expected. As for the proof of claim a, we note that Theorem 11.3.4 shows that the equation $\pi = \pi P$ takes the form

$$\pi_j = \pi_0 \delta_j + \sum_{i=1}^{j+1} \pi_i \, \delta_{j-i+1} \quad \text{for } j \in \mathbb{N}.$$

Writing $\Delta(s) = \sum_{j=0}^{\infty} s^j \, \delta_j$ for the generating function of δ it follows that

$$\begin{aligned} G_\pi(s) &= \sum_{j=0}^{\infty} s^j \, \pi_j \\ &= \pi_0 \, \Delta(s) + \sum_{j=0}^{\infty} s^j \left(\sum_{i=1}^{j+1} \pi_i \, \delta_{j-i+1} \right) \\ &= \pi_0 \, \Delta(s) + \sum_{i=1}^{\infty} \pi_i \, s^{i-1} \left(\sum_{j=i-1}^{\infty} s^{j-(i-1)} \, \delta_{j-(i-1)} \right) \\ &= \pi_0 \, \Delta(s) + \frac{G_\pi(s) - \pi_0}{s} \, \Delta(s), \end{aligned}$$

so that by rearrangement

$$G_\pi(s) = \frac{\Delta(s)\,(s-1)}{s - \Delta(s)} \, \pi_0.$$

Here we may use the Taylor expansion

$$\Delta(s) = \Delta(1) + (s-1)\,\Delta'(1) + o(s-1) = 1 + (s-1)\,\Delta'(1) + o(s-1) \quad \text{as } s \uparrow 1$$

to conclude that

$$1 \leftarrow G_\pi(s) = \frac{\Delta(s)\,(s-1)}{s - \Delta(s)} \, \pi_0 \to \frac{1}{1 - \Delta'(1)} \, \pi_0 \quad \text{as } s \uparrow 1,$$

so that

$$G_\pi(s) = \frac{\Delta(s)\,(s-1)}{s - \Delta(s)} \, (1 - \Delta'(1)).$$

Now it is sufficient to check that $\Delta(s) = M_S(\lambda\,(s-1)) = \mathbf{E}\{e^{\lambda S(s-1)}\}$, because then $\Delta'(1) = \mathbf{E}\{\lambda S\} = \rho$, and the claim of the theorem follows. However,

$$\Delta(s) = \sum_{j=0}^{\infty} s^j \, \mathbf{E}\left\{ \frac{(\lambda S)^j}{j!} \, e^{-\lambda S} \right\} = \mathbf{E}\{ e^{\lambda s S} e^{-\lambda S} \} = M_S(\lambda\,(s-1)). \qquad \square$$

Corollary 11.3.6. *A* busy period B *during which the server is continuously occupied for a* $\mathrm{M}(\lambda)/\mathrm{G}/1$ *queue satisfies*

(a) $\mathbf{E}\{B\} < \infty$ *for* $\rho < 1$;

(b) $\mathbf{E}\{B\} = \infty$ *and* $\mathbf{P}\{B < \infty\} = 1$ *for* $\rho = 1$;

(c) $\mathbf{P}\{B < \infty\} < 1$ *for* $\rho > 1$.

Proof. As B is the reccurence time for the state 0 for the chain $\{Q_n\}_{n=0}^{\infty}$, the corollary follows immediately from Theorem 11.3.5. □

Theorem 11.3.16. *The* waiting time W *a customer has to wait in the queue after arrival before coming to the server for a* $\mathrm{M}(\lambda)/\mathrm{G}/1$ *queue with traffic intensity* $\rho < 1$ *in equilibrium has moment generating function*

$$M_W(s) = \frac{(1-\rho)\, s}{\lambda + s - \lambda\, M_S(s)}.$$

Proof. Equilibrium means that the queue follows its stationary distribution, which is to say that it is started according to that distribution. Then with obvious notation a customer that waits a time W leaves a $\mathrm{Po}(\lambda(W{+}S))$ distributed number of customers behind him (or her) in the queue. Using this together with Theorem 11.3.5 we get

$$\begin{aligned}
(1-\rho)\,(s-1)\,\frac{M_S(\lambda\,(s-1))}{s - M_S(\lambda\,(s-1))} &= \mathbf{E}\big\{s^{Q_{n+1}}\big\} \\
&= \mathbf{E}\big\{s^{\mathrm{Po}(\lambda(W+S))}\big\} \\
&= \sum_{k=0}^{\infty} s^k\, \mathbf{E}\bigg\{\frac{(\lambda(W{+}S))^k}{k!}\, \mathrm{e}^{-\lambda(W+S)}\bigg\} \\
&= \mathbf{E}\big\{\mathrm{e}^{\lambda(s-1)(W+S)}\big\} \\
&= M_W(\lambda\,(s{-}1))\, M_S(\lambda\,(s{-}1)),
\end{aligned}$$

so that

$$M_W(\lambda\,(s{-}1)) = \frac{(1-\rho)\,\lambda\,(s-1)}{\lambda + \lambda\,(s{-}1) - \lambda\, M_S(\lambda\,(s{-}1))}. \qquad \square$$

Theorem 11.3.17. *The moment generating function* M_B *for the busy period* B *of a* $\mathrm{M}(\lambda)/\mathrm{G}/1$ *queue with traffic intensity* $\rho < 1$ *in equilibrium satisfies the equation*

$$M_B(s) = M_S\big(s - \lambda + \lambda\, M_B(s)\big).$$

Proof. Defining A as in the proof of Theorem 11.3.4 we have with obvious notation that $B = S + \sum_{j=1}^{A} B_j$. Recalling that A is $\mathrm{Po}(\lambda S)$ distributed we may condition on the value of S to obtain

$$\begin{aligned}
M_B(s) &= \mathbf{E}\big\{\mathrm{e}^{s(S+\sum_{j=1}^{A} B_j)}\big\} \\
&= \int_0^{\infty} \sum_{k=0}^{\infty} \mathbf{E}\big\{\mathrm{e}^{s(x+\sum_{j=1}^{k} B_j)}\big\}\, \frac{(\lambda x)^k}{k!}\, \mathrm{e}^{-\lambda x}\, dF_S(x) \\
&= \int_0^{\infty} \mathrm{e}^{sx+(M_B(s)-1)\lambda x}\, dF_S(x) \\
&= M_S\big(s - \lambda + \lambda\, M_B(s)\big). \qquad \square
\end{aligned}$$

Lecture 11, Friday 26 February

11.4 G/M/1

Let T_n be the time at which the n'th customer arrives to the queuing system for a $\mathrm{G}/\mathrm{M}(\mu)/1$ queue for $n \geq 1$. Set $Q_n = Q(T_{n+1}^-)$ for $n \geq 0$.

Theorem 11.4.3. *For a* $\mathrm{G}/\mathrm{M}(\mu)/1$ *queue the sequence* $\{Q_n\}_{n=0}^\infty$ *is a discrete time Markov chain with transition matrix*

$$P = \begin{pmatrix} 1-\alpha_0 & \alpha_0 & 0 & 0 & 0 & \cdots \\ 1-\alpha_0-\alpha_1 & \alpha_1 & \alpha_0 & 0 & 0 & \cdots \\ 1-\alpha_0-\alpha_1-\alpha_2 & \alpha_2 & \alpha_1 & \alpha_0 & 0 & \cdots \\ 1-\alpha_0-\alpha_1-\alpha_2-\alpha_3 & \alpha_3 & \alpha_2 & \alpha_1 & \alpha_0 & \cdots \\ 1-\alpha_0-\alpha_1-\alpha_2-\alpha_3-\alpha_4 & \alpha_4 & \alpha_3 & \alpha_2 & \alpha_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

where α *is the probability distribution on* $\mathbb{N}$ *with entries*

$$\alpha_j = \mathbf{E}\left\{\frac{(\mu X)^j}{j!}\,\mathrm{e}^{-\mu X}\right\} \quad \textit{for } j \in \mathbb{N}.$$

Proof. If V_n is the number of customers that depart from the queuing system during the time interval $[T_n, T_{n+1})$, then we have $Q_{n+1} = Q_n + 1 - V_n$ for $n \geq 0$. Here V_n has a truncated $\mathrm{Po}(\mu X_{n+1})$ distribution, which is to say

$$\mathbf{P}\{Q_{n+1} = j \,|\, Q_n = i\} = \mathbf{P}\{V_n = i+1-j\} = \begin{cases} \mathbf{E}\left\{\dfrac{(\mu X)^{i+1-j}}{(i+1-j)!}\,\mathrm{e}^{-\mu X}\right\} & \text{if } 1 \leq j \leq i+1, \\ \displaystyle\sum_{k>i} \mathbf{E}\left\{\dfrac{(\mu X)^k}{k!}\,\mathrm{e}^{-\mu X}\right\} & \text{if } j = 0. \end{cases} \qquad \square$$

Theorem 11.4.4. *Consider a* $\mathrm{G}/\mathrm{M}(\mu)/1$ *queue with traffic intensity* $\rho = 1/(\mu\mathbf{E}\{X\})$.
(a) *If* $\rho < 1$ *then the chain* $\{Q_n\}_{n=0}^\infty$ *is ergodic with a unique stationary distribution* π *given by* $\pi_j = (1-\eta)\,\eta^j$ *for* $j \in \mathbb{N}$*, where* η *is the smallest positive root to the equation* $\eta = M_X(\mu\,(\eta-1))$ *and* M_X *is the moment generating function for* X*.*
(b) *If* $\rho = 1$ *then the chain* $\{Q_n\}_{n=0}^\infty$ *is null persistent.*
(c) *If* $\rho > 1$ *then the chain* $\{Q_n\}_{n=0}^\infty$ *is transient.*

Proof. Taking off from Theorem 11.4.3, the proof of this theorem is very similar to that of Theorem 11.3.5 (taking off from Theorem 11.3.4). The proof is therefore omitted. $\square$

Theorem 11.4.12. *The* waiting time W *a customer has to wait in the queue after arrival before coming to the server for a* $\mathrm{G}/\mathrm{M}(\mu)/1$ *queue with traffic intensity* $\rho<1$ *in equilibrium has distribution function*

$$\mathbf{P}\{W\leq x\}=\begin{cases}0 & \text{if } x<0,\\ 1-\eta & \text{if } x=0,\\ 1-\eta\,\mathrm{e}^{-\mu(1-\eta)x} & \text{if } x>0,\end{cases}$$

where η *is as in Theorem 11.4.4.*

Proof. By the lack of memory property the waiting time W_n for the n'th customer that arrives to the queue is the sum of Q_{n-1} independent exponentially distributed random variables with parameter μ. Using this together with Theorem 11.4.4 we get

$$M_W(s)=\sum_{j=0}^{\infty}(\mathbf{E}\{\exp(\mu)\})^j\pi_j=\sum_{j=0}^{\infty}\left(\frac{\mu}{\mu-s}\right)^j(1-\eta)\,\eta^j=\cdots=(1-\eta)+\eta\,\frac{\mu(1-\eta)}{\mu(1-\eta)-s},$$

which is to say a mixture of a distribution that is zero with probability $1-\eta$ and an exponential distribution with parameter $\mu(1-\eta)$ with probability η. □

11.5 G/G/1

Theorem 11.5.1. (Lindley's[14] equation) *For a* $\mathrm{G}/\mathrm{G}/1$ *queue we have*

$$W_{n+1}=\max\{0,W_n+S_n-X_{n+1}\}.$$

Proof. Writing T_n for the arrival time of the n'th customer to the queuing system and noting that $T_{n+1}-T_n=X_{n+1}$ we readily see that

$$T_{n+1}+W_{n+1}=\max\{T_{n+1},T_n+W_n+S_n\}.$$

Subtracting T_{n+1} from both sides of this equation we get the claim of the theorem. □

Theorem 11.5.2. *For a* $\mathrm{G}/\mathrm{G}/1$ *queue we have*

$$\mathbf{P}\{W_{n+1}\leq x\}=\begin{cases}0 & \text{if } x<0,\\ \displaystyle\int_{-\infty}^{x}\mathbf{P}\{W_n\leq x-y\}\,dG(y) & \text{if } x\geq 0,\end{cases}$$

where G *is the distribution function of* $U_n=S_n-X_{n+1}$. *In particular it follows that the limit* $\lim_{n\to\infty}\mathbf{P}\{W_n\leq x\}$ *exists.*

[14] Dennis Victor Lindley, British statistician born 1923.

Proof. The first claim follows from Lindley's equation. As for the second claim, it is sufficient to prove that $\mathbf{P}\{W_{n+1} \leq x\} \leq \mathbf{P}\{W_n \leq x\}$ for all n and x. Now this is trivially true for $n = 1$ as $\mathbf{P}\{W_1 \leq 0\} = 1$. Further, if the inequality holds for all x for $n = k$, then it holds for all x for $n = k+1$ by the first claim of the theorem. □

Theorem 11.5.7. *For a* G/G/1 *queue the random variable* W_{n+1} *has the same distribution as*

$$W'_{n+1} \equiv \max\left\{0, U_1, U_1+U_2, \ldots, U_1+\ldots+U_n\right\}.$$

Proof. Use Lindley's equation to deduce that $W_1 = 0$, $W_2 = \max\{0, W_1 + U_1\} = \max\{0, U_1\}$, $W_3 = \max\{0, W_2+U_2\} = \max\{0, U_2, U_1+U_1\}$, ... , $W_{n+1} = \max\{0, U_n, U_n+U_{n-1}, \ldots, U_n+\ldots+U_1\}$. Then use that the vectors $(U_1, \ldots, U_n)$ and $(U_n, \ldots, U_1)$ are identically distributed. □

Theorem 11.5.4. *Consider a* G/G/1 *queue with traffic intensity* $\rho = \mathbf{E}\{S\}/\mathbf{E}\{X\}$.
(a) *If* $\rho < 1$ *then the limit* $F(x) = \lim_{n\to\infty} \mathbf{P}\{W_n \leq x\}$ *is a distribution function.*
(b) *If* $\rho = 1$ *and* $\mathbf{Var}\{U\} > 0$ *then* $F(x) = 0$ *for* $x \in \mathbb{R}$.
(c) *If* $\rho > 1$ *then* $F(x) = 0$ *for* $x \in \mathbb{R}$.

Proof. We omit the proof of claims b and c, but note that they are very natural and expected. As for claim a we note that $W'_n \leq W'_{n+1}$ for all n, so that the limit $\lim_{n\to\infty} W'_n = W'$ exists. Using Theorem 11.5.7 we now get

$$F(x) = \lim_{n\to\infty} \mathbf{P}\{W'_n \leq x\} = \mathbf{P}\{W' \leq x\} = \mathbf{P}\left\{\textstyle\sum_{j=1}^n U_j \leq x \text{ for all } n\right\}.$$

As we assume $\rho < 1$, which is to say $\mathbf{E}\{U\} < 0$, the strong law of large numbers gives

$$\mathbf{P}\left\{\textstyle\sum_{j=1}^n U_j > 0 \text{ for infinitely many } n\right\}$$
$$= \mathbf{P}\left\{\frac{1}{n}\sum_{j=1}^n U_j - \mathbf{E}\{U\} > -\mathbf{E}\{U\} \text{ for infinitely many } n\right\} = 0.$$

Hence W' is with probability 1 the maximum of only finitely many members of the sequence $\{\sum_{j=1}^n U_j\}_{n=1}^\infty$, and thus a well-defined finite random variable. □

11.6 Heavy traffic

Many real-world queues are designed to have a traffic intensity ρ strictly less than one (for reasons of stability), but just barely so (for reasons of economy). It is therefore interesting to investigate the behaviour of a queue as $\rho \uparrow 1$. The following theorem exemplifies such an investigation:

Theorem 11.6.1. *Let a* $\mathrm{M}(\lambda)/\mathrm{D}(d)/1$ *queue have traffic intensity* $\rho = d\lambda < 1$. *If* Q_ρ *is a random variable with the stationary distribution (recall Theorem 11.3.5), then* $(1-\rho)\,Q_\rho$ *converges in distribution to an exponential distribution with expected value* $1/2$ *as* $\rho \uparrow 1$.

Proof. As $M_S(t) = \mathrm{e}^{td}$ Theorem 11.3.5 shows that

$$\begin{aligned}
M_{(1-\rho)Q_\rho}(s) &= M_{Q_\rho}((1-\rho)\,s) \\
&= G_\pi\big(\mathrm{e}^{(1-\rho)s}\big) \\
&= (1-\rho)\,(\mathrm{e}^{(1-\rho)s}-1)\,\frac{\exp\big\{\rho\,(\mathrm{e}^{(1-\rho)s}-1)\big\}}{\mathrm{e}^{(1-\rho)s}-\exp\big\{\rho\,(\mathrm{e}^{(1-\rho)s}-1)\big\}} \\
&= \frac{(1-\rho)\,(\mathrm{e}^{(1-\rho)s}-1)}{\exp\big\{(1-\rho)\,s-\rho\,(\mathrm{e}^{(1-\rho)s}-1)\big\}-1} \\
&= \frac{(1-\rho)\,\big((1-\rho)\,s+o(1-\rho)\big)}{(1-\rho)^2 s-(1-\rho)^2 s^2/2+o((1-\rho)^2)} \\
&\to \frac{2}{2-s} \\
&= \int_0^\infty \mathrm{e}^{sx}\,2\,\mathrm{e}^{-2x}\,dx \quad \text{as } \rho\uparrow 1.
\end{aligned}$$

□

Lecture 12, Thursday 4 March

Chapter 12 Martingales

In this chapter we will be much concerned with conditional expectations and probabilities. Given random variables $X_0,\ldots,X_n,Y$, recall that

$$\mathbf{E}\{Y\,|\,X_0{=}x_0,\ldots,X_n{=}x_n\} = f(x_0,\ldots,x_n)$$

and

$$\mathbf{P}\{Y\le y\,|\,X_0{=}x_0,\ldots,X_n{=}x_n\} = g(y,x_0,\ldots,x_n)$$

are functions of $x_0,\ldots,x_n$ and $y,x_0,\ldots,x_n$, respectively. Hence we may define

$$\mathbf{E}\{Y\,|\,X_0,\ldots,X_n\} = f(X_0,\ldots,X_n)$$

and

$$\mathbf{P}\{Y\le y\,|\,X_0,\ldots,X_n\} = g(y,X_0,\ldots,X_n),$$

so that $\mathbf{E}\{Y\,|X_0,\ldots,X_n\}$ and $\mathbf{P}\{Y\le y\,|X_0,\ldots,X_n\}$ become random variables.

12.1 Introduction

Definition 12.1.1. *A sequence $\{Y_n\}_{n=0}^{\infty}$ of random variables is a* martingale[15] *with respect to a sequence $\{X_n\}_{n=0}^{\infty}$ of random variables if*

$$\mathbf{E}\{|Y_n|\}<\infty \quad \textit{and} \quad \mathbf{E}\{Y_{n+1}|X_0,\ldots,X_n\}=Y_n \quad \textit{for all } n.$$

The most common selection of the sequence $\{X_n\}_{n=0}^{\infty}$ is to take it to be equal to $\{Y_n\}_{n=0}^{\infty}$, so that the martingale definition boils down to

$$\mathbf{E}\{|Y_n|\}<\infty \quad \text{and} \quad \mathbf{E}\{Y_{n+1}|Y_0,\ldots,Y_n\}=Y_n \quad \text{for all } n.$$

Whatever the choice of the sequence $\{X_n\}_{n=0}^{\infty}$ we write $\mathcal{F}_n$ to denote the information corresponding to the knowledge of the random variables $\{X_n\}_{n=0}^{\infty}$, so that

$$\mathbf{E}\{Y\,|\,X_0,\ldots,X_n\} = \mathbf{E}\{Y\,|\,\mathcal{F}_n\}.$$

Two important formulae associated with this formalism is that

$$\mathbf{E}\big\{\mathbf{E}\{Y\,|\,\mathcal{F}_{m+n}\}\big|\,\mathcal{F}_n\big\} = \mathbf{E}\{Y\,|\,\mathcal{F}_n\} \quad \text{and} \quad \mathbf{E}\big\{\mathbf{E}\{Y\,|\,\mathcal{F}_n\}\big|\,\mathcal{F}_{m+n}\big\} = \mathbf{E}\{Y\,|\,\mathcal{F}_n\}$$

for $m{+}n\ge n\ge 0$.

[15]See `http://www.emis.de/journals/JEHPS/juin2009/Mansuy.pdf` for information on the origin of the word martingale. (It is not a person!)

Example 12.1.2. (Simple random walk) Let $Y_n = \sum_{i=1}^n X_i - n\,(p-q)$ where $\{X_n\}_{n=0}^\infty$ are iid. $\{-1,1\}$-valued random variables such that $\mathbf{P}\{X_n = 1\} = p$ and $\mathbf{P}\{X_n = -1\} = q = 1-p$. Then it is clear that $\{Y_n\}_{n=0}^\infty$ is a martingale with respect to $\{X_n\}_{n=0}^\infty$, see exercise below. #

Exercise. Prove the claim of Example 12.1.2 in full detail.

Example 12.1.4. (De Moivre's[16] martingale) Take $\{X_n\}_{n=0}^\infty$ as in Example 12.1.2 and define $Y_n = (q/p)^{X_1+\ldots+X_n}$. Then $\{Y_n\}_{n=0}^\infty$ is a martingale with respect to $\{X_n\}_{n=0}^\infty$ because

$$\mathbf{E}\{Y_{n+1}|X_0,\ldots,X_n\} = (q/p)^{X_1+\ldots+X_n}\,\mathbf{E}\big\{(q/p)^{X_{n+1}}\big\} = Y_n\left(\frac{q}{p}\,p + \frac{p}{q}\,q\right) = Y_n. \quad \#$$

Definition 12.1.9. (a) *A sequence $\{Y_n\}_{n=0}^\infty$ of random variables is a* submartingale *with respect to $\{\mathcal{F}_n\}_{n=0}^\infty$ if*

$$\mathbf{E}\{|Y_n|\} < \infty^{17}, \quad \mathbf{E}\{Y_n|\mathcal{F}_n\} = Y_n \quad \textit{and} \quad \mathbf{E}\{Y_{n+1}|\mathcal{F}_n\} \geq Y_n \quad \textit{for all } n.$$

(b) *A sequence $\{Y_n\}_{n=0}^\infty$ of random variables is a* supermartingale *with respect to $\{\mathcal{F}_n\}_{n=0}^\infty$ if*

$$\mathbf{E}\{|Y_n|\} < \infty, \quad \mathbf{E}\{Y_n|\mathcal{F}_n\} = Y_n \quad \textit{and} \quad \mathbf{E}\{Y_{n+1}|\mathcal{F}_n\} \leq Y_n \quad \textit{for all } n.$$

Clearly, a process that is both a submartingale and a supermartingale (with respect to a common sequence) is a martingale (with respect to that sequence). Also, if $\{Y_n\}_{n=0}^\infty$ is a submartingale then $\{-Y_n\}_{n=0}^\infty$ is a supermartingale, and vice versa.

Proposition. *A submartingale have non-decreasing mean.*

Proof. We have

$$\mathbf{E}\{Y_{n+1}\} = \mathbf{E}\{\mathbf{E}\{Y_{n+1}|\mathcal{F}_n\}\} \geq \mathbf{E}\{Y_n\}. \qquad \square$$

By the above proposition together with the text following Definition 12.1.9 martingales have constant means while supermartingales have non-increasing means.

The next theorem is quit basic and follows from a few simple arguments, but is of fundamental importance anyway:

[16] Abraham de Moivre, French mathematician 1667-1754.

[17] The weaker integrability condition employed by Grimmett and Stirzaker in their definition of submartingales and supermartingales is wrong!

Theorem. **(a)** *If* $\{Y_n\}_{n=0}^{\infty}$ *is a martingale and* $g:\mathbb{R}\to\mathbb{R}$ *a convex function such that* $\mathbf{E}\{|g(Y_n)|\}<\infty$ *for all* n*, then* $\{g(Y_n)\}_{n=0}^{\infty}$ *is a submartingale.*
(b) *If* $\{Y_n\}_{n=0}^{\infty}$ *is a submartingale and* $g:\mathbb{R}\to\mathbb{R}$ *a non-decreasing convex function such that* $\mathbf{E}\{|g(Y_n)|\}<\infty$ *for all* n*, then* $\{g(Y_n)\}_{n=0}^{\infty}$ *is a submartingale.*

Proof. This is Exercises 12.1.6 and 12.1.7, respectively, in the book. □

The next theorem turns out to be one of the few most important results for the build up of the more advanced martingale theory:

Theorem 12.1.10. (Doob[18] decomposition) *A submartingale* $\{Y_n\}_{n=0}^{\infty}$ *with respect to* $\{\mathcal{F}_n\}_{n=0}^{\infty}$ *may be expressed in a unique way as* $Y_n = M_n + S_n$ *where* $\{M_n\}_{n=0}^{\infty}$ *is a martingale with respect to* $\{\mathcal{F}_n\}_{n=0}^{\infty}$ *and* $\{S_n\}_{n=0}^{\infty}$ *is a non-decreasing process with* $S_0=0$ *such that* $\mathbf{E}\{S_{n+1}|\mathcal{F}_n\} = S_{n+1}$ *for all* n*.*

Proof. As for uniqueness we note that if we have two representations $Y_n = M_n+S_n$ and $Y_n = M_n'+S_n'$ for Y of the stipulated type, then

$$S_{n+1}'-S_n' = \mathbf{E}\{S_{n+1}'-S_n'|\mathcal{F}_n\} = \mathbf{E}\{Y_{n+1}-Y_n|\mathcal{F}_n\} = \mathbf{E}\{S_{n+1}-S_n|\mathcal{F}_n\} = S_{n+1}-S_n.$$

From this together with a telescope sum argument and the fact that $S_0=S_0'=0$ we get $S_n=S_n'$, which in turn gives $M_n=M_n'$.

As for existence, define M and S recursively by $M_0=Y_0$, $S_0=0$,

$$M_{n+1} = M_n+Y_{n+1}-\mathbf{E}\{Y_{n+1}|\mathcal{F}_n\} \quad \text{and} \quad S_{n+1} = S_n-Y_n+\mathbf{E}\{Y_{n+1}|\mathcal{F}_n\} \quad \text{for } n\geq 1.$$

By adding the last two equations we see that $M_{n+1}-M_n+S_{n+1}-S_n = Y_{n+1}-Y_n$ for $n\geq 1$, so that a telescope sum argument together with the fact that $M_0+S_0 = Y_0$ give $Y_n = M_n+S_n$. Further, $\mathbf{E}\{M_{n+1}|\mathcal{F}_n\} = M_n$ holds for $n=0$, since

$$\mathbf{E}\{M_1|\mathcal{F}_0\} = \mathbf{E}\{M_0|\mathcal{F}_0\} = \mathbf{E}\{Y_0|\mathcal{F}_0\} = Y_0 = M_0.$$

If $\mathbf{E}\{M_{n+1}|\mathcal{F}_n\} = M_n$ holds for $n=k$ then we further have

$$\begin{aligned}
\mathbf{E}\{M_{k+2}|\mathcal{F}_{k+1}\} &= \mathbf{E}\big\{M_{k+1}+Y_{k+2}-\mathbf{E}\{Y_{k+2}|\mathcal{F}_{k+1}\}\big|\mathcal{F}_{k+1}\big\}\\
&= \mathbf{E}\{M_{k+1}|\mathcal{F}_{k+1}\}\\
&= \mathbf{E}\big\{M_k+Y_{k+1}-\mathbf{E}\{Y_{k+1}|\mathcal{F}_k\}\big|\mathcal{F}_{k+1}\big\}\\
&= \mathbf{E}\{M_k|\mathcal{F}_{k+1}\}+Y_{k+1}-\mathbf{E}\{Y_{k+1}|\mathcal{F}_k\}\\
&= M_k+Y_{k+1}-\mathbf{E}\{Y_{k+1}|\mathcal{F}_k\}\\
&= M_{k+1},
\end{aligned}$$

[18]Joseph Leo Doob, American mathematician 1910-2004, who made founding contributions to the theory of stoachastic processes and in particular to martingale theory.

because

$$\mathbf{E}\{M_k|\mathcal{F}_{k+1}\} = \mathbf{E}\big\{\mathbf{E}\{M_{k+1}|\mathcal{F}_k\}\big|\mathcal{F}_{k+1}\big\} = \mathbf{E}\{M_{k+1}|\mathcal{F}_k\} = M_k.$$

As it is obvious that the S process is non-decreasing (when Y is a submartingale), it remains to prove that $\mathbf{E}\{S_{n+1}|\mathcal{F}_n\} = S_{n+1}$, which in turn by inspection of the definition of S holds if $\mathbf{E}\{S_n|\mathcal{F}_n\} = S_n$. As the latter equation holds if $\mathbf{E}\{S_n|\mathcal{F}_{n-1}\} = S_n$ (see exercise below), it is sufficient to prove that $\mathbf{E}\{S_1|\mathcal{F}_0\} = S_1$, which in turn holds by inspection of the definition of S. □

Exercise. Show that $\mathbf{E}\{S_n|\mathcal{F}_n\} = S_n$ if $\mathbf{E}\{S_n|\mathcal{F}_{n-1}\} = S_n$ in the proof of Theorem 12.1.10.

12.2 Martingale differences (but not Hoeffding's inequality)

Definition 12.2.1. *A sequence* $\{D_n\}_{n=1}^{\infty}$ *of random variables are* martingale differences *if*

$$\mathbf{E}\{|D_n|\} < \infty, \quad \mathbf{E}\{D_n|\mathcal{F}_n\} = D_n \quad \text{and} \quad \mathbf{E}\{D_{n+1}|\mathcal{F}_n\} = 0 \quad \text{for all } n.$$

Theorem. *$\{D_n\}_{n=1}^{\infty}$ are martingale differences if and only if $\{\sum_{i=1}^{n} D_i\}_{n=0}^{\infty}$ is a martingale.*

Exercise. Prove the above theorem.

12.3 Crossings and convergence

Historically, martingale theory emerged from a desire to generalize the classical limit theorems in probability from iid. sequences of random variables to more general settings. The following theorem is just one example of a limit theorem for martingales:

Theorem 12.3.1. *If $\{Y_n\}_{n=0}^{\infty}$ is a submartingale such that $\sup_{n\geq 0}\mathbf{E}\{|Y_n|\} < \infty$, then $Y_n \to Y_\infty$ as $n\to\infty$ for some random variable Y_∞.*

Proof. Omitted. □

Theorem 12.3.7. *If $\{Y_n\}_{n=0}^{\infty}$ is either a non-positive submartingale or a non-negative supermartingale, then $Y_n \to Y_\infty$ as $n\to\infty$ for some random variable Y_∞.*

Proof. The first statement follows from Theorem 12.3.7 together with the fact that submartingales have non-decreasing expectations, so that

$$\sup_{n\geq 0}\mathbf{E}\{|Y_n|\} = \sup_{n\geq 0}\mathbf{E}\{-Y_n\} = -\inf_{n\geq 0}\mathbf{E}\{Y_n\} = -\mathbf{E}\{Y_0\} < \infty.$$

The second statement follows from applying the first statement to $\{-Y_n\}_{n=0}^{\infty}$. □

Example 12.3.9. (Doob's martingale) Let Z be an integrable random variable $\mathbf{E}\{|Z|\} < \infty$ and set $Y_n = \mathbf{E}\{Z\,|\,\mathcal{F}_n\}$ for $n\in\mathbb{N}$. Then we have

$$\mathbf{E}\{|Y_n|\} = \mathbf{E}\big\{\big|\mathbf{E}\{Z\,|\,\mathcal{F}_n\}\big|\big\} \leq \mathbf{E}\big\{\mathbf{E}\{|Z|\,|\,\mathcal{F}_n\}\big|\big\} = \mathbf{E}\{|Z|\} < \infty \quad \text{for } n\in\mathbb{N},$$

as well as

$$\mathbf{E}\{Y_{n+1}|\,\mathcal{F}_n\} = \mathbf{E}\big\{\mathbf{E}\{Z\,|\,\mathcal{F}_{n+1}\}\big|\mathcal{F}_n\big\} = \mathbf{E}\{Z\,|\,\mathcal{F}_n\} = Y_n \quad \text{for } n\in\mathbb{N}.$$

Hence $\{Y_n\}_{n=0}^{\infty}$ is a martingale that statisfies the hypothesis of Theorem 12.3.7, so that $Y_n \to Y_\infty$ as $n\to\infty$ for some random variable Y_∞. #

12.4 Stopping times

Definition 12.4.1. *An $\mathbb{N}$-valued random variable T is called a* stopping time *(or* optional time*) if*

$$\mathbf{P}\{T=n\,|\,\mathcal{F}_n\} = \mathbf{1}_{\{T=n\}} \quad \textit{for } n\in\mathbb{N}.$$

Exercise. Show that if T is a stopping time, then we have

$$\mathbf{P}\{T\leq m\,|\,\mathcal{F}_n\} = \mathbf{1}_{\{T\leq m\}} \quad \text{for } m\leq n \quad \text{and} \quad \mathbf{P}\{T>n\,|\,\mathcal{F}_n\} = \mathbf{1}_{\{T>n\}}.$$

Example 12.4.3. (Coin tossing) Let $X_n = 1$ if the n'th throw of a fair coin gives a head while $X_n = 0$ if the n'th throw gives tails. Then the first time T for a head is a stopping time with respect to $\{X_n\}_{n=0}^{\infty}$, because

$$\mathbf{P}\{T=n\,|\,X_0,\ldots,X_n\} = \mathbf{1}_{\{X_0=\ldots=X_{n-1}=0,X_n=1\}} = \mathbf{1}_{\{T=n\}}. \qquad \#$$

Example 12.4.4. (First passage time) The argument employed in Example 12.4.3 extends to show that $T = \inf\{n\in\mathbb{N} : X_n\in B\}$ is a stopping time with respect to a given sequence of random variables $\{X_n\}_{n=0}^{\infty}$ for any set $B\subseteq\mathbb{R}$. #

Theorem 12.4.5. *If $\{Y_n\}_{n=0}^{\infty}$ is a submartingale and T a stopping time, then $\{Y_{n\wedge T}\}_{n=0}^{\infty}$ is a submartingale.*

Proof. We have

$$Y_{n\wedge T} = \sum_{k=0}^{n} Y_k\, \mathbf{1}_{\{T=k\}} + Y_n\, \mathbf{1}_{\{T>n\}}$$

From this it follows that

$$\mathbf{E}\{|Y_{n\wedge T}|\} \leq \sum_{k=0}^{n} \mathbf{E}\{|Y_k|\} + \mathbf{E}\{|Y_n|\} < \infty$$

as well as (by inspection) $\mathbf{E}\{Y_{n\wedge T}|\mathcal{F}_n\} = Y_{n\wedge T}$. Further, we have

$$\begin{aligned}
\mathbf{E}\{Y_{(n+1)\wedge T}|\mathcal{F}_n\} &= \mathbf{E}\left\{\sum_{k=0}^{n+1} Y_k\, \mathbf{1}_{\{T=k\}} + Y_{n+1}\, \mathbf{1}_{\{T>n+1\}} \,\middle|\, \mathcal{F}_n\right\} \\
&= \mathbf{E}\left\{ Y_{n\wedge T} + Y_{n+1}\, \mathbf{1}_{\{T=n+1\}} - Y_n\, \mathbf{1}_{\{T>n\}} + Y_{n+1}\, \mathbf{1}_{\{T>n+1\}} \,\middle|\, \mathcal{F}_n\right\} \\
&= Y_{n\wedge T} + \mathbf{E}\big\{(Y_{n+1} - Y_n)\, \mathbf{1}_{\{T>n\}} \big| \mathcal{F}_n\big\} \\
&= Y_{n\wedge T} + \mathbf{E}\{Y_{n+1} - Y_n|\mathcal{F}_n\}\, \mathbf{1}_{\{T>n\}} \\
&= Y_{n\wedge T} + \big(\mathbf{E}\{Y_{n+1}|\mathcal{F}_n\} - Y_n\big)\, \mathbf{1}_{\{T>n\}} \\
&\geq Y_{n\wedge T}
\end{aligned}$$

as Y is a submartingale. □

Theorem 12.4.7. (Optional switching) *If $\{X_n\}_{n=0}^{\infty}$ and $\{Y_n\}_{n=0}^{\infty}$ are martingales and T a stopping such that $X_T = Y_T$, then the following process $\{Z_n\}_{n=0}^{\infty}$ is a martingale:*

$$Z_n = \begin{cases} X_n & \text{for } T \leq n, \\ Y_n & \text{for } T > n, \end{cases} \quad \text{for } n \in \mathbb{N}.$$

Proof. By the martingale properties of X and Y and the fact that $X_T = Y_T$ we have

$$\begin{aligned}
Z_n &= \mathbf{E}\{X_{n+1}|\mathcal{F}_n\}\, \mathbf{1}_{\{T\leq n\}} + \mathbf{E}\{Y_{n+1}|\mathcal{F}_n\}\, \mathbf{1}_{\{T>n\}} \\
&= \mathbf{E}\big\{X_{n+1}\, \mathbf{1}_{\{T\leq n\}} + Y_{n+1}\, \mathbf{1}_{\{T>n\}} \big| \mathcal{F}_n\big\} \\
&= \mathbf{E}\big\{Z_{n+1} - X_{n+1}\, \mathbf{1}_{\{T=n+1\}} + Y_{n+1}\, \mathbf{1}_{\{T=n+1\}} \big| \mathcal{F}_n\big\} \\
&= \mathbf{E}\{Z_{n+1}|\mathcal{F}_n\} - \mathbf{E}\big\{(X_T - Y_T)\, \mathbf{1}_{\{T=n+1\}} \big| \mathcal{F}_n\big\} \\
&= \mathbf{E}\{Z_{n+1}|\mathcal{F}_n\}.
\end{aligned}$$

□

Theorem 12.4.11. (Optional sampling theorem) *For a submartingale $\{Y_n\}_{n=0}^{\infty}$ and a stopping time T that is bounded above by a deterministic finite constant we have*

$$\mathbf{E}\{|Y_T|\} < \infty \quad \textit{and} \quad \mathbf{E}\{Y_T|\mathcal{F}_0\} \geq Y_0.$$

Proof. By Theorem 12.4.5 $Z_n = Y_{n\wedge T}$, $n \in \mathbb{N}$, is a submartingale. Selecting an $N \in \mathbb{N}$ such that $T \leq N$ it follows that $\mathbf{E}\{|Y_T|\} < \infty$ in the same way as the first part of the proof of Theorem 12.4.5. Further, we have

$$\mathbf{E}\{Y_T|\mathcal{F}_0\} = \mathbf{E}\{Z_N|\mathcal{F}_0\} \geq Z_0 = Y_0. \qquad \square$$

Theorem 12.4.13. *For a martingale* $\{Y_n\}_{n=0}^{\infty}$ *we have*

$$\mathbf{P}\Big\{\max_{0\leq k\leq n} Y_k \geq x\Big\} \leq \frac{\mathbf{E}\{Y_n^+\}}{x} \quad \textit{and} \quad \mathbf{P}\Big\{\max_{0\leq k\leq n} |Y_k| \geq x\Big\} \leq \frac{\mathbf{E}\{|Y_n|\}}{x} \quad \textit{for } x > 0.$$

Proof. The second inequality follows from the first one as

$$\begin{aligned}\mathbf{P}\Big\{\max_{0\leq k\leq n} |Y_k| \geq x\Big\} &\leq \mathbf{P}\Big\{\max_{0\leq k\leq n} Y_k \geq x\Big\} + \mathbf{P}\Big\{\max_{0\leq k\leq n} -Y_k \geq x\Big\}\\ &\leq \frac{\mathbf{E}\{Y_n^+\}}{x} + \frac{\mathbf{E}\{Y_n^-\}}{x}\\ &= \frac{\mathbf{E}\{|Y_n|\}}{x}.\end{aligned}$$

As for the first inequality, write $T = \inf\{n \in \mathbb{N} : Y_n \geq x\}$. Then Example 12.4.4 shows that T is a stopping time with respect to $\{Y_n\}_{n=0}^{\infty}$, from which it follows that T is also a stopping time with respect to $\{\mathcal{F}_n\}_{n=0}^{\infty}$, see the exercise below. Hence $T \wedge n$ is a bounded stopping time (see the exercise below), so that

$$\mathbf{E}\{Y_{T\wedge n}\} = \mathbf{E}\big\{\mathbf{E}\{Y_{T\wedge n}|\mathcal{F}_0\}\big\} = \mathbf{E}\{Y_0\} = \mathbf{E}\{Y_n\}$$

by the optional sampling theorem together with the martingale property of Y. It follows that

$$\begin{aligned}\mathbf{E}\{Y_n^+\} &\geq \mathbf{E}\{Y_n^+ \mathbf{1}_{\{T\leq n\}}\}\\ &\geq \mathbf{E}\{Y_n \mathbf{1}_{\{T\leq n\}}\}\\ &= \mathbf{E}\{Y_n\} - \mathbf{E}\{Y_n \mathbf{1}_{\{T>n\}}\}\\ &= \mathbf{E}\{Y_{T\wedge n}\} - \mathbf{E}\{Y_{T\wedge n} \mathbf{1}_{\{T>n\}}\}\\ &= \mathbf{E}\{Y_{T\wedge n} \mathbf{1}_{\{T\leq n\}}\}\\ &= \mathbf{E}\{Y_T \mathbf{1}_{\{T\leq n\}}\}\\ &\geq x\,\mathbf{E}\{\mathbf{1}_{\{T\leq n\}}\}\\ &= x\,\mathbf{P}\{T \leq n\}.\end{aligned} \qquad \square$$

Exercise. In the proof of Theorem 12.4.13, show that T is a stopping time with respect to $\{\mathcal{F}_n\}_{n=0}^{\infty}$ when it is a stopping time with respect to $\{Y_n\}_{n=0}^{\infty}$, and that $T \wedge n$ is a stopping time when T is a stopping time.

Lecture 13, Friday 5 March

12.5 Optional stopping

Theorem 12.5.1. (Optional stopping theorem) *Let $\{Y_n\}_{n=0}^{\infty}$ be a martingale and T a stopping time such that*

$$\mathbf{P}\{T<\infty\}=1, \quad \mathbf{E}\{|Y_T|\}<\infty \quad \textit{and} \quad \lim_{n\to\infty}\mathbf{E}\big\{Y_n\,\mathbf{1}_{\{T>n\}}\big\}=0.$$

Then we have $\mathbf{E}\{Y_T\}=\mathbf{E}\{Y_0\}$.

Proof. As Theorem 12.4.5 shows that $\{Y_{n\wedge T}\}_{n=0}^{\infty}$ is a martingale we have

$$\mathbf{E}\{Y_T\}=\mathbf{E}\{Y_{n\wedge T}\}+\mathbf{E}\big\{(Y_T-Y_n)\,\mathbf{1}_{\{T>n\}}\big\}=\mathbf{E}\{Y_0\}+\mathbf{E}\{Y_T\,\mathbf{1}_{\{T>n\}}\}-\mathbf{E}\{Y_n\,\mathbf{1}_{\{T>n\}}\}.$$

Sending $n\to\infty$ we conclude that it is enough to prove that $\lim_{n\to\infty}\mathbf{E}\{Y_T\,\mathbf{1}_{\{T>n\}}\}=0$. However, this follows by elementary considerations from the facts that

$$\mathbf{E}\{Y_T\}=\sum_{i=0}^{\infty}\mathbf{E}\{Y_T\,\mathbf{1}_{\{T=i\}}\} \quad \text{and} \quad \mathbf{E}\{Y_T\,\mathbf{1}_{\{T>n\}}\}=\sum_{i=n+1}^{\infty}\mathbf{E}\{Y_T\,\mathbf{1}_{\{T=i\}}\}. \qquad \square$$

Example 12.5.6. (Symmetric simple random walk) Let $Y_n=\sum_{i=1}^{n}X_i$ for $n\geq 0$ where $\{X_n\}_{n=0}^{\infty}$ are iid. Rademacher[19] distributed random variables, i.e., $\mathbf{P}\{X_i=1\}=\mathbf{P}\{X_i=-1\}=\frac{1}{2}$. Let $T=\inf\{n\in\mathbb{N}:Y_n=a \text{ or } Y_n=b\}$ for integers $a<0<b$. Then it is obvious that the martingale Y (with respect to $\{X_n\}_{n=0}^{\infty}$) and the stopping time T satisfy the hypothesis of Theorem 12.5.1 (recall from Chapter 6 that Y is persistent). Hence we have

$$0=\mathbf{E}\{Y_0\}=\mathbf{E}\{Y_T\}=a\,\mathbf{P}\{Y \text{ hits } a \text{ before } b\}+b\,(1-\mathbf{P}\{Y \text{ hits } a \text{ before } b\}),$$

so that

$$\mathbf{P}\{Y \text{ hits } a \text{ before } b\}=\frac{b}{b-a} \quad \text{and} \quad \mathbf{P}\{Y \text{ hits } b \text{ before } a\}=\frac{-a}{b-a}.$$

Now note that $\{Z_n\}_{n=0}^{\infty}$ given by $Z_n=Y_n^2-n$ also is a martingale, because

$$\begin{aligned}\mathbf{E}\{Z_{n+1}|\mathcal{F}_n\}&=\mathbf{E}\big\{(Y_{n+1}-Y_n)^2+2\,(Y_{n+1}-Y_n)\,Y_n+Y_n^2\big|\mathcal{F}_n\big\}-(n+1)\\&=\mathbf{E}\{X_{n+1}^2\}+2\,\mathbf{E}\{X_{n+1}\}\,Y_n+Y_n^2-(n+1)\\&=1+2\cdot 0\cdot Y_n+Y_n^2-(n+1)\\&=Z_n \quad \text{for } n\in\mathbb{N}.\end{aligned}$$

Now it might be shown that also Z and T satisfy the hypothesis of Theorem

[19] Hans Adolph Rademacher, German mathematician 1892-1969.

12.5.1, see exercise below. Hence we have

$$0 = \mathbf{E}\{Z_0\} = \mathbf{E}\{Z_T\} = a^2 \frac{b}{b-a} + b^2 \frac{-a}{b-a} - \mathbf{E}\{T\},$$

so that we get $\mathbf{E}\{T\} = -ab$ by rearranging. #

Exercise. Explain in detail why the process Y and the stopping time T in Example 12.5.6 satisfy the hypothesis of Theorem 12.5.1.

Exercise. (DIFFICULT) Explain in detail why the process Z and the stopping time T in Example 12.5.6 satisfy the hypothesis of Theorem 12.5.1.

Example 12.5.12. (WALD'S[20] EQUATION) Let $\{X_n\}_{n=0}^{\infty}$ be iid. integrable random variables and T a stopping time with finite mean. Then we have

$$\mathbf{E}\left\{\sum_{n=1}^{T} X_n\right\} = \mathbf{E}\{X_1\}\,\mathbf{E}\{T\}.$$

This can be proved by application of Theorem 12.5.1. However, as it is not easy to verify the hypothesis of that theorem for the setting under consideration it turns out to be more convenient to prove the result by direct calculations as follows:

$$\begin{aligned}
\mathbf{E}\left\{\sum_{n=1}^{T} X_n\right\} - \mathbf{E}\{X_1\}\,\mathbf{E}\{T\} &= \mathbf{E}\left\{\sum_{n=1}^{T}(X_n - \mathbf{E}\{X_1\})\right\} \\
&= \mathbf{E}\left\{\sum_{i=1}^{\infty}\sum_{n=1}^{i}(X_n - \mathbf{E}\{X_1\})\,\mathbf{1}_{\{T=i\}}\right\} \\
&= \mathbf{E}\left\{\sum_{n=1}^{\infty}\sum_{i=n}^{\infty}(X_n - \mathbf{E}\{X_1\})\,\mathbf{1}_{\{T=i\}}\right\} \\
&= \sum_{n=1}^{\infty}\mathbf{E}\big\{(X_n - \mathbf{E}\{X_1\})\,\mathbf{1}_{\{T\geq n\}}\big\} \\
&= -\sum_{n=1}^{\infty}\mathbf{E}\big\{(X_n - \mathbf{E}\{X_1\})\,\mathbf{1}_{\{T\leq n-1\}}\big\} \\
&= -\sum_{n=1}^{\infty}\mathbf{E}\big\{\mathbf{E}\big\{(X_n - \mathbf{E}\{X_1\})\,\mathbf{1}_{\{T\leq n-1\}}\big|\mathcal{F}_{n-1}\big\}\big\} \\
&= -\sum_{n=1}^{\infty}\mathbf{E}\big\{\mathbf{E}\big\{(X_n - \mathbf{E}\{X_1\})\big|\mathcal{F}_{n-1}\big\}\,\mathbf{1}_{\{T\leq n-1\}}\big\} \\
&= 0.
\end{aligned}$$

#

[20] Abraham Wald, mathematician 1902-1950 born in Austria-Hungary.

12.6 The maximal inequality

We have the following important generalization of Theorem 12.4.13:

Theorem 12.6.1. (Doob maximal inequality) *For a submartingale* $\{Y_n\}_{n=0}^{\infty}$ *we have*

$$\mathbf{P}\Big\{\max_{0\le k\le n} Y_k \ge x\Big\} \le \frac{\mathbf{E}\{Y_n^+\}}{x} \quad \textit{for } x>0.$$

Proof. As $\{Y_n^+\}_{n=0}^{\infty}$ is a non-negative submartingale (since $(\,\cdot\,)^+$ is a non-decreasing convex function), we may define a stopping time $T = \inf\{n \in \mathbb{N} : Y(t) \ge x\}$ and deduce that

$$\begin{aligned}
\mathbf{E}\{Y_n^+\} &\ge \mathbf{E}\big\{Y_n^+ \mathbf{1}_{\{\max_{0\le k\le n} Y_k^+ \ge x\}}\big\} \\
&= \mathbf{E}\Big\{Y_n^+ \sum_{k=0}^{n} \mathbf{1}_{\{T=k\}}\Big\} \\
&= \mathbf{E}\Big\{\sum_{k=0}^{n} \mathbf{E}\big\{Y_n^+ \mathbf{1}_{\{T=k\}} \big| \mathcal{F}_k\big\}\Big\} \\
&= \mathbf{E}\Big\{\sum_{k=0}^{n} \mathbf{E}\{Y_n^+ | \mathcal{F}_k\}\, \mathbf{1}_{\{T=k\}}\Big\} \\
&\ge \mathbf{E}\Big\{\sum_{k=0}^{n} Y_k^+ \mathbf{1}_{\{T=k\}}\Big\} \\
&= \mathbf{E}\big\{Y_T^+ \mathbf{1}_{\{\max_{0\le k\le n} Y_k^+ \ge x\}}\big\} \\
&= \mathbf{E}\big\{Y_T\, \mathbf{1}_{\{\max_{0\le k\le n} Y_k \ge x\}}\big\} \\
&\ge x\, \mathbf{E}\big\{\mathbf{1}_{\{\max_{0\le k\le n} Y_k \ge x\}}\big\} \\
&= x\, \mathbf{P}\Big\{\max_{0\le k\le n} Y_k \ge x\Big\}.
\end{aligned}$$

□

Example 12.6.7. (Doob-Kolmogorov inequality) For a square-integrable martingale $\{Y_n\}_{n=0}^{\infty}$ we may apply Theorem 12.6.1 to the non-negative submartingale $\{Y_n^2\}_{n=0}^{\infty}$ to obtain

$$\mathbf{P}\Big\{\max_{0\le k\le n} |Y_k| \ge x\Big\} = \mathbf{P}\Big\{\max_{0\le k\le n} Y_k^2 \ge x^2\Big\} \le \frac{\mathbf{E}\{Y_n^2\}}{x} \quad \text{for } x>0.$$

#

12.7 Backward martingales and continuous-time martingales

Definition 12.7.1. *A sequence* $\{Y_n\}_{n=0}^{\infty}$ *of random variables is a* backward martingale *with respect to a sequence* $\{X_n\}_{n=0}^{\infty}$ *of random variables if*

$$\mathbf{E}\{|Y_n|\} < \infty \quad \textit{and} \quad \mathbf{E}\{Y_n | X_{n+1}, X_{n+2}, \ldots\} = Y_{n+1} \quad \textit{for all } n.$$

Example 12.7.2. (Strong law of large numbers) Let $\{X_n\}_{n=1}^{\infty}$ be iid. integrable random variables. Then $S_n = \sum_{i=1}^{n} X_i$ is integrable and satisfies

$$\mathbf{E}\{S_n | S_{n+1}, S_{n+2}, \ldots\} = \mathbf{E}\{S_n | S_{n+1}\} = \frac{n}{n+1}\,\mathbf{E}\{S_{n+1} | S_{n+1}\} = \frac{n}{n+1}\,S_{n+1},$$

so that $\{S_n/n\}_{n=1}^{\infty}$ is a backward martingale. By employing limit theorems for backward martingales the strong law of large numbers may thus be recovered. #

Definition. *A stochastic process* $\{Y(t)\}_{t\geq 0}$ *is a* martingale *with respect to a stochastic process* $\{X(t)\}_{t\geq 0}$ *if* $\mathbf{E}\{|Y(t)|\} < \infty$ *for* $t \geq 0$ *and*

$$\mathbf{E}\{Y(t) | X(s_1), \ldots, X(s_n), X(s)\} = Y(s) \quad \textit{for } 0 \leq s_1 \leq \ldots \leq s_n \leq s \leq t.$$

Unlike discrete time martingale theory, continuous time martingale theory is very heavily burdened with difficult technicalities that cannot at all be adressed rigorously at the undergraduate level. We will feel satisfied with giving the most important example of a continuous time martingale.

Example. If $\{X(t)\}_{t\geq 0}$ is a finite mean independent increment process, then by writing $X(t) = (X(t) - X(s)) + X(s)$ and using independence of the increments it is readily seen that $Y(t) = X(t) - \mathbf{E}\{X(t)\}$ is a martingale with respect to $\{X(t)\}_{t\geq 0}$. #

12.8 Some examples

From this section you should read just enough to make it possible for you to complete one hand-in exercise.