Lecture 10

Continuous-time Markov chains

10.1 General Markov chains

Let $\mathbb{S}$ be a countable state space, and $M = (M_n)_{n\geq 0}$ or $X = (X_t)_{t\geq 0}$ a time-homogeneous Markov chain with unknown transition probabilities/rates. In this lecture, we will develop methods to construct maximum likelihood estimators for the transition probabilities/ rates. You may think of a population model where birth rates and death rates may depend on the population size, and also multiple births (or maybe immigration) and multiple deaths (accidents, disasters, emigration) are allowed.

10.1.1 Discrete time, estimation of Π-matrix

Suppose we start a Markov chain at $M_0 = i_0$ and then observe $(M_0, \ldots, M_n) = (i_0, \ldots, i_n)$. The general transition matrix $\Pi = (\pi_{ij})_{i,j\in\mathbb{S}}$ contains our parameters π_{ij}, $i, j \in \mathbb{S}$, and we can write down the likelihood (probability mass function) for our observations

$$\prod_{k=1}^{n} \pi_{i_{k-1},i_k} = \prod_{i\in\mathbb{S}}\prod_{j\in\mathbb{S}} \pi_{ij}^{n_{ij}}$$

where n_{ij} is the number of transitions from i to j. Now note that $p_{ii} = 1 - \sum_{j\neq i} p_{ij}$ so that, as before, the maximum likelihood estimators are

$$\hat{p}_{ij} = \frac{n_{ij}}{n_i}, \qquad \text{provided } n_i = \sum_{j\in\mathbb{S}} n_{ij} = \#\{0 \leq k \leq n-1 : i_k = i\} > 0.$$

If $n_i = 0$, then the likelihood is the same for all π_{ij}, $j \in \mathbb{S}$.

10.1.2 Estimation of the Q-matrix

Suppose we start a continuous-time Markov chain at $X_0 = i_0$ and observe $(X_s)_{0\leq s\leq T_n}$ where T_n is the nth transition time and record the data as successive holding times $T_j - T_{j-1} = h_{j-1}$ and sequence of states of the jump chain $(i_0, \ldots, i_n)$. The general Q-matrix $(q_{ij})_{i,j\in\mathbb{S}}$ contains the parameters q_{ij}, $i, j \in \mathbb{S}$, and the likelihood, as a product of densities of holding times and transition probabilities, is given by

$$\prod_{k=1}^{n} \lambda_{i_{k-1}} \exp\{-\lambda_{i_{k-1}} h_{j-1}\} \frac{q_{ik-1,i_k}}{\lambda_{i_{k-1}}} = \prod_{i\in\mathbb{S}}\prod_{j\neq i} q_{ij}^{n_{ij}} \exp\left\{-e_i q_{ij}\right\},$$

where $\lambda_i = -q_{ii} = \sum_{j \neq i} q_{ij}$, n_{ij} is the number of transitions from i to j and e_i is the total time spent in i. This is maximised factor by factor by

$$\hat{q}_{ij} = \hat{q}_{ij}(i_0, h_0, i_1, \dots, h_{n-1}, i_n) = \frac{n_{ij}}{e_i}, \qquad i \neq j.$$

In fact, the holding times and transitions may come from several chains (with the same unknown Q-matrix) without affecting the form of the likelihood, if we define

$$n_{ij} = n_{ij}^{(1)} + \ldots + n_{ij}^{(r)} \qquad \text{and} \qquad e_i = e_i^{(1)} + \ldots + e_i^{(r)}$$

for observed chains $(X_s^{(k)})_{0 \leq s \leq T_{n_k}^{(k)}}$, $1 \leq k \leq r$. This is useful to save time by simultaneous observation, and to reach areas of the state space not previously visited (e.g. for reducible or transient chains).

10.2 The induced Poisson process

In order to derive a rigorous estimation theory for a general finite-state Markov process, we consider the embedded Poisson processes that comes from considering the process only when it is in a given state. You might imagine a "state-x estimator" that is tasked with estimating the transition rates to all other states from state x; its clock runs while the process is in state x, and it counts the transitions, but it slumbers when the process is in any other state.

Suppose we run infinitely many copies of the Markov process, started in states $X_0^{(1)}, X_0^{(2)}, \ldots$ for some lengths of time $S^{(1)}, S^{(2)}, \ldots$. Suppose that the times $S^{(i)}$ are stopping times: That is, they do not depend upon knowing the future of the process. (For a precise technical definition, see any basic text on stochastic processes, such as [KT81].) The realisation i makes transitions at times $T_1^{(i)}, \ldots, T_{N_i}^{(i)}$ to states $X_1^{(i)}, \ldots, X_{N_i}^{(i)}$. (We do not wish to rule out the simple possibility that there is only one run of a positive recurrent Markov process. To admit that alternative, though, would complicate the notation. Instead, we may suppose that the single infinite run is broken into infinitely many pieces, for instance by stopping after each full unit of time, and restarting with $X_0^{(i+1)} = X_{N_i}^{(i)}$.) We assume that the realisations are independent, except that the starting state of a realisation may be dependent on

Consider some fixed state x. Suppose all K realisations visit state x a total of M_x times, and let $\tau_x(j)$ be the length of the j-th sojourn in x — so that $E_x := \tau_x(1) + \cdots \tau_x(M_x)$ is the total of all the time intervals when the process is in state x. Of the sojourns in x, some end with a transition to a new state, and some end because a stopping time $S^{(i)}$ intervened; define

$$\delta_x(j) = \begin{cases} 1 & \text{if sojourn } j \text{ ends with a transition;} \\ 0 & \text{if sojourn } j \text{ ends with a stopping time.} \end{cases}$$

Consider now the random interval $[0, E_x]$, and the set of points

$$\mathcal{S}_x := \big\{ \tau_x(1) + \cdots + \tau_x(j) : \text{ s.t. } \delta_x(j) = 1 \big\}.$$

The idea is that we take the events that occur only while the process is waiting at state x out, and stitch them together. Theorem 10.2.2 tells us that we obtain thereby a Poisson process. An illustration may be found in Figure 10.1. We start with a strong restatement of the "memoryless" property of the exponential distribution.

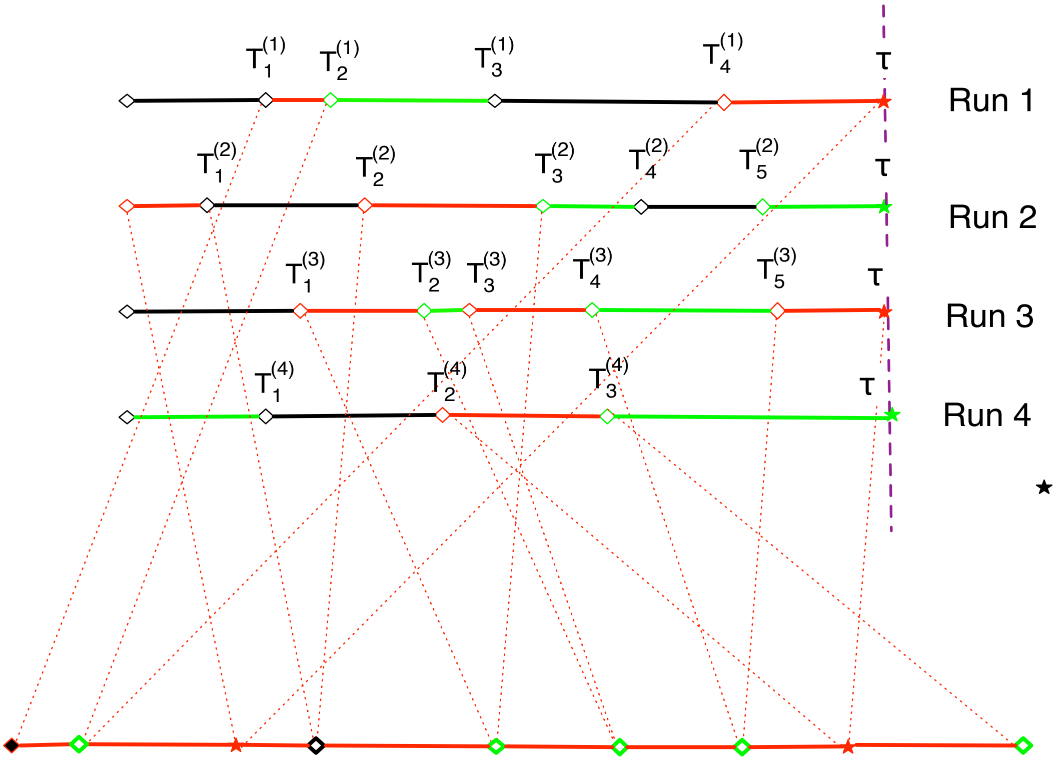

Figure 10.1: Illustration of the "stitching-together" construction, by which the process confined to a particular state generates a marked Poisson process. The Markov process has three states, represented by red, green, and black. We are estimating the transition rates from the red state. The diamond shapes represent the colour to which transitions are made. Stars represent censored observations; that is, times at which a realisation of the process was ended (at the time τ) in state R, without having transitioned out. The estimates based on these observations would be $\hat{q}_{RG} = 5/E$ and $\hat{q}_{RG} = 1/E$, where E is the total length of the red line at the bottom.

Lemma 10.2.1. *Suppose $T_1, T_2, \dots$ is an i.i.d. sequence of exponential random variables with parameter λ, and $S_1, S_2, \dots$ independent random variables such that each S_i is a stopping time with respect to T_i. That is, $T_i - t$ is independent of S_i on the event $\{S_i \le t\}$, for any fixed t. Let $K = \min\{k : T_k \le S_k\}$. Then*

$$T_* := T_K + \sum_{i=1}^{K-1} S_i$$

is exponential with parameter λ.

Proof. The stopping-time property tells us that $(T_i - S_i)$ is independent of S_i on the event $\{T_i > S_i\}$. Consequently, conditioned on $\{T_i > S_i = s\}$, $(T_i - S_i)$ has the distribution of $(T_i - s)$ conditioned on $\{T_i > s\}$, which is exponential with parameter λ; and conditioned on $\{T_i \le s\}$, T_i has exponential distribution with parameter λ. Conditioned on $\{K = 1\}$, then, it is immediately true that $T_* = T_1$ has the correct distribution. Suppose now that T_* has the correct distribution, conditioned on $\{K = k\}$. Then conditioned on $\{K = k+1\}$, $T_* := T_{k+1} + S_k + \sum_{i=1}^{k-1} S_i$. Note that $S_k + T_{k+1}$ conditioned on $\{K = k+1\}$ has the same distribution as T_k conditioned on $\{K = k\}$ (by the induction hypothesis). Since either of these is independent of $\sum_{i=1}^{k-1} S_i$, the distribution is the same T_* conditioned on $\{K = k+1\}$ is identical to the distribution conditioned on $\{K = k\}$, which completes the induction. ☐

Theorem 10.2.2. *The random set $\mathcal{S}_x$ is a Poisson process with rate $q_x := \sum_{y \ne x} q_{xy}$, and the total time E_x is a stopping time for the process. If we condition on $(E_x : x \in \mathcal{X})$, the processes corresponding to different states are independent. Finally, conditioned on $\mathcal{S}_x$, the transitions that take place at the times $\mathcal{S}_x$ are independent, with the probability of transitioning to y being q_{xy}/q_x.*

Proof. Consider the interarrival time between two points of $\mathcal{S}_x$. By Lemma 10.2.1 it is exponential with parameter q_x. By the Markov property, the interarrival times are all independent. Hence, these are independent Poisson processes. The independence of the transitions from the waiting times is standard. ☐

Define $N_x(s)$ to be the number of visits to state x up to total time s (where total time is measured by stitching together the processes $X^{(1)}$, followed by $X^{(2)}$, and so on) which end in a transition (as opposed to ending in a stopping time, and shift to a new realisation of the process). Let $N_{xy}(s)$ be the number of visits to state x up to total time s which end in a transition to state y. Let $E_x(s)$ be the total amount of time spent in state x up to total time s. (Thus, $\sum_{x \in \mathcal{X}} E_x(s) = s$ identically.)

Consequences of Theorem 10.2.2 are:

MLE The maximum likelihood estimator for the rate of a Poisson process is # events/total time. Thus, if we observe realisations of the process which add up to total time S (where S may be a random stopping time), the MLE for q_{xy} is

$$\hat{q}_{xy}(S) = \frac{N_{xy}(S)}{E_x(S)}. \tag{1}$$

Consistency $\lim_{s\to\infty} \frac{N_{xy}(S)}{E_x(S)} = q_{xy}$, on the event $\{\lim_{s\to\infty} E_x(s) = \infty\}$. (Question to consider: How would it be possible to arrange the realisations of the process so that the condition $\{\lim_{s\to\infty} E_x(s) = \infty\}$ does not have probability 1?)

Sampling dist. Suppose we run realisations of the process until a random time S that we will call $S_x(t) := \inf\{s : E_x(s) = t\}$; that is, we run the process (in its various successive realisations) until such time as the total time spent in x is exactly t. Then the estimator $\hat{q}_{xy}(S)$ is equal to $N_{xy}(S)/E_x(S) = N_{xy}(S)/t$. Since $N_{xy}(S)$ has Poisson distribution with parameter tq_{xy}, this tells us the distribution of $\hat{q}_{xy}(S)$. Its expectation is q_{xy}, and its variance is q_{xy}/t. As $t \to \infty$, $\hat{q}_{xy}(S_x(t))$ converges to a normal distribution.

The sampling distribution is a complicated matter. If we have observed up to time $S_x(t)$ then we know the exact distribution of $\hat{q}_{xy}$, and we can compute approximate 100α% confidence intervals as $\hat{q}_{xy} \pm z_{(1-\alpha/2}\sqrt{\hat{q}_{xy}/t}$, where z is the appropriate quantile of the normal distribution. Even then, we do not have the exact distribution even for the estimators of transition rates starting from any other state.

Can this be a serious problem? Suppose we observe instead up to a constant total time s. Is the distribution substantially different? In some respects it is, particularly in the tails. Suppose we decompose the estimate by the number of visits to x.

$$\hat{q}_{xy}(s) = \frac{N_{xy}(s)}{E_x(s)} = \sum_{n=0}^{\infty} \mathbf{1}_{\{N_x(s)=n\}} \frac{N_{xy}(s)}{E_x(s)}.$$

The summand corresponding to $n = 0$ is $0/0$, which is problematic. Moving on to $n = 1$, we have with probability q_{xy}/q_x the expression $1/E$, where E is exponential with parameter q_x. This has expectation ∞. Consequently, $\hat{q}_{xy}(s)$ also has infinite expectation. Other choices of a random time S at which to observe the process can similarly distort the distribution.

On the other hand, this is only a problem with the expectation and variance, not with the main bulk of the distribution. That is, as long as S is chosen so that there will be, with high probability, a large number of visits to x, the normal approximation should be fairly accurate.

The general rule for estimating transition rates is

$$\hat{q}_{xy} = \frac{\#\text{ transitions } x \to y}{\text{total time spent in state } x}$$
$$\mathrm{Var}(\hat{q}_{xy}) \approx \frac{\hat{q}_{xy}}{\text{total time spent in state } x}$$

We compute an approximate 100α% confidence interval as $\hat{q}_{xy} \pm z_{1-\alpha/2}\sqrt{\mathrm{Var}(\hat{q}_{xy})}$.
If the number of transitions is not very large, we may do better estimating $\hat{q}_{xy}$ from the Poisson parameter estimated by N_{xy}. Exact confidence intervals may be computed, using the identity

$$\mathbb{P}\big\{k \le N_s\big\} = \mathbb{P}\big\{T_k \le s\big\} = \mathbb{P}\big\{2k\mu T_k \le 2k\mu s\big\} = \alpha \qquad \text{for } \mu = \chi_{2k,\alpha}/(2ks).$$

This is carried out in the exercises.

10.3 Parametric and time-dependent models

So far, we have assumed that the Q-matrix was completely arbitrary, i.e. with entries $q_{ij} \ge 0$, $j \ne i$, and $q_{ii} = -\sum_{j:j\ne i} q_{ij} > -\infty$. We estimated all "parameters" q_{ij} by observing a

Markov chain with that unknown Q-matrix (or several independent Markov chains with the same unknown Q-matrix).

On the other hand, we studied the multiple decrement model, which we can view as a continuous-time Markov chain, where the Q-matrix contains lots of zeros, namely everywhere except in the first row. Here, the maximum likelihood estimates that we derived were of the same form

$$\frac{\text{numbers of transitions from } i \text{ to } j}{\text{total time spent in } i},$$

except that we did not specify the zero rows as estimates but as model assumptions.

Here, we will merge ideas and study more systematically Markov chain models where the Q-matrices are not completely unknown. Instead, we assume/know some structure, e.g. certain zero entries and/or that certain transition rates are the same or stand in a known relationship to each other.

Secondly, we will incorporate time-dependent transition rates (as we had in the multiple decrement model) into the general Markov model.

10.3.1 Example: Marital status model

A natural model for marital status consists of five states "bachelor" (B), "married" (M), "widowed" (W), "divorced" (D) and "dead" (Δ).

We can set up a model with 9 parameters corresponding to the 9 possible transitions ($B \to M$, $B \to \Delta$, $M \to W$, $M \to D$, $M \to \Delta$, $W \to M$, $W \to \Delta$, $D \to M$, $D \to \Delta$). Note that also state Δ is absorbing and there is no reason to continue to observe chains that have run into this state. This means that we agree that four entries vanish:

$$q_{\Delta B} = q_{\Delta M} = q_{\Delta W} = q_{\Delta D} = 0.$$

Furthermore, it is also impossible to have direct transitions between B, W and D or indeed to go from M to B, so we also know in advance that

$$q_{BD} = q_{DB} = q_{BW} = q_{WB} = q_{DW} = q_{WD} = q_{MB} = 0.$$

With states arranged in the above order, this gives a Q-matrix

$$Q = \begin{pmatrix} -\alpha - \mu_B & \alpha & 0 & 0 & \mu_B \\ 0 & -\nu - \delta - \mu_M & \nu & \delta & \mu_M \\ 0 & \sigma & -\sigma - \mu_W & 0 & \mu_W \\ 0 & \rho & 0 & -\rho - \mu_D & \mu_D \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Alternatively, we can *assume* that the death rate does not depend on the current state B, M, W or D, so that the Q-matrix only contains 6 parameters as

$$Q = \begin{pmatrix} -\alpha - \mu & \alpha & 0 & 0 & \mu \\ 0 & -\nu - \delta - \mu & \nu & \delta & \mu \\ 0 & \sigma & -\sigma - \mu & 0 & \mu \\ 0 & \rho & 0 & -\rho - \mu & \mu \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Finally, we can allow age-varying transition rates by having $Q(t)$, where now the parameters $\alpha(t)$, $\mu(t)$, $\nu(t)$, $\delta(t)$, $\sigma(t)$ and $\rho(t)$ depend on age t.

10.3.2 The general simple birth-and-death process

The general simple birth-and-death process is a continuous-time Markov chain with Q-matrix

$$Q = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & 0 & \cdots \\ \mu_1 & -\lambda_1 - \mu_1 & \lambda_1 & 0 & 0 & \cdots \\ 0 & \mu_2 & -\lambda_2 - \mu_2 & \lambda_2 & 0 & \cdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots \end{pmatrix},$$

where birth rates λ_j and death rates μ_j are in arbitrary (unknown) dependence on j. Note that this model has an infinite number of parameters, but just as with unspecified maximal ages in the single decrement model, we would only estimate $(\lambda_0, \mu_1, \ldots, \mu_{\max}, \lambda_{\max})$, where indeed $\lambda_{\max}$ may be left unspecified or equal to zero for the highest observed population size.

This model has the same maximum likelihood estimators as the general Q-matrix with all entries unknown, since that likelihood factorises completely, and given samples from simple birth-and-death processes, there are no multiple births and deaths, so the maximum likelihood estimates of the multiple birth and death rates (jumps of sizes two or higher) are zero. The (remaining) likelihood is

$$\prod_{k=1}^{n} q_{X_{T_{k-1}},X_{T_k}} \exp\{-\sum_{j:j\neq X_{T_{k-1}}} q_{X_{T_{k-1}},j}(T_k - T_{k-1})\} = \prod_{i\in\mathbb{N}} \prod_{|j-i|=1} q_{ij}^{N_{ij}} \exp\{-E_i q_{ij}\},$$

where now $q_{i,i+1} = \lambda_i$ and $q_{i,i-1} = \mu_i$, $i \geq 1$. N_{ij} is the number of transitions from i to j and E_i is the total time spent in i. Note that we have taken the liberty and written the likelihood in terms of the underlying random variables, not the realisations.

We note a general phenomenon: if transitions are impossible, and, in particular there are no observations of such transitions in the samples, their likelihood contribution is maximised by a zero rate. Note that vice versa, a given sample may not contain some other transitions although they are possible. In this case, the estimate of the corresponding transition rate is zero, but not usually the estimator, which is non-zero with positive probability for all transitions that are possible within the given number of steps from the given initial values.

10.3.3 Lower-dimensional parametric models of simple birth-and-death processes

Often population models come with some additional structure. The simplest structure is that of independent individuals each giving birth repeatedly at rate λ until their death at rate μ. Here $\lambda_j = j\lambda$ and $\mu_j = j\mu$, $j \in \mathbb{N}$, are all expressed in terms of two parameters λ and μ. The likelihood in this model is the same as in the general model, but has to be factorised as

$$\prod_{i\in\mathbb{N}} \prod_{|j-i|=1} q_{ij}^{N_{ij}} \exp\{-E_i q_{ij}\} = \left(\prod_{i\in\mathbb{N}} (i\lambda)^{N_{i,i+1}} \exp\{-E_i i\lambda\}\right) \left(\prod_{i\in\mathbb{N}} (i\mu)^{N_{i,i-1}} \exp\{-E_i i\mu\}\right)$$

to separate the two parameters. This can best be maximised via the log likelihood, which for the μ-factor is

$$\sum_{i=1}^{\infty} \left(N_{i,i-1}(\log(i) + \log(\mu)) - E_i i\mu\right).$$

Differentiation leads to the maximum likelihood estimator $\hat{\mu} = D/W$ where $D = \sum_i N_{i,i-1}$ is the total number of deaths and $W = \sum_i iE_i$ is the weighted sum of exposure times at population sizes $i \in \mathbb{N}$. This quantity has again the interpretation as the total time spent at risk since in state i there are i individuals at risk to die. Similarly, $\hat{\lambda} = B/W$, where B is the total number of births. (W is also the total time spent at risk to give birth.)

Note that the state 0 is absorbing, so an nth transition may never come. This can be helped by running several Markov chains. The likelihood function of the given sample is the one given, if we understand that N_{ij} are aggregate counts and E_i are aggregate waiting times, $i, j \in \mathbb{N}$.

10.4 Time-varying transition rates

10.4.1 Maximum likelihood estimation

We take up the setting of a general Markov model $(X_t)_{t\geq 0}$, but have the Q-matrix depend on t, $Q(t)$. For simplicity think of t as age. Denote the finite or countably infinite state space by $\mathbb{S}$. Given we have reached state $i \in \mathbb{S}$ aged exactly $y \in [0, \infty)$, the situation is exactly as for the multiple decrement model, there are competing hazards $q_{ij}(y + t)$, $t \geq 0$, $j \neq i$, and the total holding time in state i has hazard rate

$$-q_{ii}(y+t) = \sum_{j:j\neq i} q_{ij}(y+t), \qquad t \geq 0.$$

Given the holding time is $Z = t$, the transition is from i to j with probability

$$\mathbb{P}(X_{y+Z} = j | X_y = i, Z = t) = \frac{q_{ij}(y+t)}{\sum_{j:j\neq i} q_{ij}(y+t)}.$$

To be able to estimate time-varying transition rates, we require more than one realisation of X, say realisations $(X_t^{(m)})_{0\leq t\leq T^{(m)}_{n_m}}$, $m = 1, \ldots, r$, where T_{n_m} is the n_mth transition time of $X^{(m)}$. Then the likelihood is given by

$$\prod_{m=1}^{r} \left(\prod_{k=1}^{n_m} q_{X^{(m)}_{T^{(m)}_{k-1}}, X^{(m)}_{T^{(m)}_{k}}}(T^{(m)}_k) \exp\left\{ -\int_{T^{(m)}_{k-1}}^{T^{(m)}_k} \sum_{j:j\neq X^{(m)}_{T^{(m)}_{k-1}}} q_{X^{(m)}_{T^{(m)}_{k-1}}, j}(s)ds \right\} \right)$$

If we also postulate simplifying assumptions such as piecewise constant transition rates $q_{ij}(t) = q_{ij}(x + \frac{1}{2})$, $x \leq t < x+1$, $x \in \mathbb{N}$, we can reexpress this in a factorised form as

$$\prod_{x\in\mathbb{N}} \prod_{i\in\mathbb{S}} \prod_{j\neq i} q_{ij}(x+\tfrac{1}{2})^{N_{ij}(x)} \exp\left\{-E_i(x) q_{ij}(x+\tfrac{1}{2})\right\},$$

where $N_{ij}(x)$ is the number of transitions from i to j at an age x, i.e. aged t with $x \leq t < x+1$, and $E_i(x)$ is the total time spent in state i while aged x.

We read off the maximum likelihood estimators for all $x \in \mathbb{N}$ and $i \in \mathbb{N}$ with $E_i(x) > 0$:

$$\hat{q}_{ij}(x+\tfrac{1}{2}) = \frac{N_{ij}(x)}{E_i(x)}.$$

10.4.2 Example

Clearly, a reasonably complete set of reasonably reliable estimates can only be obtained if the state space $\mathbb{S}$ is small and the number of observations is very large, e.g., in the illness-death model with three states H=able, S=sick and D=dead, with age-dependent sickness rates σ_x from H to S, recovery rates ρ_x from S to H and death rates δ_x from H to D and γ_x from S to D.

Suppose, we observe r individuals over their whole life $[0, \tau_d^{(m)}]$, then we get estimates

$$\hat{\delta}_{x+\frac{1}{2}} = \frac{d_x}{v_x}, \quad \hat{\gamma}_{x+\frac{1}{2}} = \frac{c_x}{w_x}, \quad \hat{\sigma}_{x+\frac{1}{2}} = \frac{s_x}{v_x}, \quad \hat{\rho}_{x+\frac{1}{2}} = \frac{r_x}{w_x},$$

where v_x (w_x) is the total waiting time of lives aged x in the able (ill) state, and d_x, c_x, s_x, r_x are the aggregate counts of the respective transitions at age x.

10.4.3 Construction of the stochastic process $(X_t)_{t\geq 0}$

This section is *non-examinable* and deals with the Probability behind minimal $(\nu, (Q(t))_{t\geq 0})$ Markov chains, where ν is an initial distribution on $\mathbb{S}$ and $(Q(t))_{t\geq 0}$ is a time-dependent Q-matrix.

We first give a construction analogous to the maze construction for continuous-time Markov chains with constant transition rates, but note that the jump-chain holding description was rather vague, so we will not "prove" but only "indicate" why the process we construct does what we want. In fact, you may wish to take the maze construction as the definition of a minimal $(\nu, (Q(t))_{t\geq 0}$ Markov chain.

We construct counting processes $(N_t^{(ij)})_{t\geq 0}$ for all pairs $i, j \in \mathbb{S}$, $i \neq j$, independent. Fix i and j and consider a Poisson process $\tilde{N}^{(ij)}$ with unit rate. Then define

$$N_t^{(ij)} = \tilde{N}^{(ij)}_{\int_0^t q_{ij}(s)ds}$$

This is a time-inhomogeneous Poisson process. It is obvious that $N^{(ij)}$ is still a counting process since the jumps of N and $\tilde{N}$ are in $1-1$ correspondence

$$N_t^{(ij)} - N_{t-}^{(ij)} = \tilde{N}^{(ij)}_{\int_0^t q_{ij}(s)ds} - \tilde{N}^{(ij)}_{\int_0^t q_{ij}(s)ds-}.$$

N still has independent increments with Poisson distributions, since

$$\begin{aligned} N_{t_n}^{(ij)} - N_{t_{n-1}}^{(ij)} = \tilde{N}^{(ij)}_{\int_0^{t_n} q_{ij}(s)ds} - \tilde{N}^{(ij)}_{\int_0^{t_{n-1}} q_{ij}(s)ds} \quad &\sim \quad \mathrm{Poi}\left(\int_0^{t_n} q_{ij}(s)ds - \int_0^{t_{n-1}} q_{ij}(s)ds\right) \\ &= \quad \mathrm{Poi}\left(\int_{t_{n-1}}^{t_n} q_{ij}(s)ds\right), \end{aligned}$$

but note that these increments are no longer stationary for all increment lengths, unless $q_{ij} \equiv q_{ij}(s)$ does not depend on s, in which case $N^{(ij)}$ is simply a (homogeneous) Poisson process with rate q_{ij}.

Next we define aggregate processes

$$N_t^{(i)} = \sum_{j\neq i} N_t^{(ij)} \sim \mathrm{Poi}\left(\int_0^t \lambda_i(s)ds\right), \qquad t \geq 0, i \in \mathbb{S},$$

also inhomogeneous Poisson processes. Note that the first jump time $T_1^{(i)}$ of $N^{(i)}$ has survival function

$$\mathbb{P}(T_1^{(i)} > t) = \mathbb{P}(N_t^{(i)} = 0) = \exp\left\{-\int_0^t \lambda_i(s)ds\right\},$$

i.e. hazard rate $\lambda_i(t)$. Also $T_1^{(i)} = \inf\{T_1^{(ij)}, j \neq i\}$ is as for the multiple decrement model. Similarly, for $T_1^{(i)}(x) = \inf\{t \geq x : N_t^{(i)} \neq N_x^{(i)}\}$, we have

$$\begin{aligned}\mathbb{P}\left(T_1^{(i)}(x) > x+t\right) = \mathbb{P}\left(N_{x+t}^{(i)} - N_x^{(i)} = 0\right) &= \exp\left\{-\int_x^{x+t} \lambda_i(s)ds\right\} \\ &= \exp\left\{-\int_0^t \lambda_i(x+s)ds\right\},\end{aligned}$$

and this identifies a hazard rate of $\lambda_i(x+t)$. Furthermore, as calculated for $T = \min\{T_j : 1 \leq j \leq m\}$ in the multiple decrement model, we have for $T_1^{(i)}(x) = \inf\{T_1^{(ij)}(x), j \neq i\}$

$$\mathbb{P}\left(T_1^{(i)}(x) = T_1^{(ij)}(x) \middle| T_1^{(i)}(x) = t\right) = \frac{q_{ij}(x+t)}{\lambda_i(x+t)}.$$

The construction is as follows. Take $M_0 \sim \nu$, independent from all Poisson processes $(N_t^{(ij)})_{t\geq 0}$, $T_0 = 0$, and define inductively jump times

$$T_{n+1} = \inf\{t > T_n : N_t^{(M_n)} \neq N_{T_n}^{(M_n)}\},$$

and jump destinations

$$M_{n+1} = j \qquad \Longleftrightarrow \qquad N_{T_{n+1}}^{(M_n,j)} \neq N_{T_{n+1}-}^{(M_n,j)},$$

$n \in \mathbb{N}$. Then specify X as

$$X_t = M_n \qquad \Longleftrightarrow \qquad T_n \leq t < T_{n+1}.$$

In general, M is not a Markov chain, and holding times $T_{n+1} - T_n$ are not conditionally independent given M, but X has been constructed from independent Poisson processes only, as in the constant-Q case, and it can be shown that X has a Markov property, that we formulate below.

First note that we can, more generally, construct a $(\nu, x, (Q(t))_{t\geq 0}$ chain $(X_t)_{t\geq x}$ starting at time x (rather than 0) from an initial distribution $X_x \sim \nu$, simply by changing $T_0 := 0$ to $T_0 := x$, while keeping the remainder of the construction.

The Markov property of X now states that $(X_{x+t})_{t\geq 0}$ is conditionally independent of $(X_r)_{0\leq r\leq x}$ given $X_x = i$. Given $X_x = i$, the post-x process is a $(\delta_i, x, (Q(t))_{t\geq 0})$ Markov chain, where $\delta_i = (\delta_{ij})_{j\in\mathbb{S}}$ is the Dirac probability mass function putting all mass in i, $\delta_{ii} = 1$. This Markov property is again a consequence of the maze construction, since the post-x process only depends on the current state and the (inhomogeneous, but independent-increment) Poisson processes after time x.

We can then derive, under some further regularity conditions, as for the constant-rate case, an infinitesimal description,

$$\mathbb{P}(X_{x+t} = j | X_x = i) = \mathbb{P}(T_1^{(i)}(x) \leq x+t, T_1^{(ij)}(x) = T_1^{(i)}(x)) + o(t) = q_{ij}(x)t + o(t),$$

for $i \neq j$, as $t \downarrow 0$, and forward and backward equations

$$\frac{\partial}{\partial t}P(s,t) = P(s,t)Q(t) \qquad \text{and} \qquad \frac{\partial}{\partial s}P(s,t) = Q(s)P(s,t),$$

for the transition matrices $P(s,t) = (p_{ij}(s,t))_{i,j\in\mathbb{S}}$, where $p_{ij}(s,t) = \mathbb{P}(X_t = j | X_s = i)$.

10.5 Occupation times

The last topic we consider is the generalisation of the formulas that we had earlier for life expectancy. For a lifetime with constant mortality rate μ, the expected lifetime is μ^{-1}. Consider now the illness model described in section 10.4.2

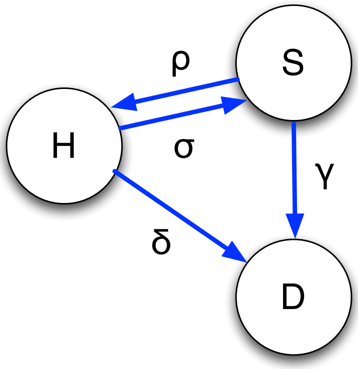

Figure 10.2: Diagram of the Illness model

We say that a matrix is **negative-definite** if all of its eigenvalues are negative. Define

$$T_x(t) := \text{total time spent in state } x \text{ up to time } t,$$

and ${}_yE_x(t) := \mathbb{E}_y[T_x(t)]$, where $\mathbb{E}_y$ represents the expectation given that the process starts in state y.

Theorem 10.5.1. *Let Q be the $(m+k)\times(m+k)$ transition matrix for a continuous-time discrete-space Markov process with m absorbing states and k non-absorbing states. Let Q_* be the $k\times k$ submatrix consisting of transition rates among the non-absorbing states. If Q_* is irreducible and some row has negative sum, then Q_* is negative definite, and the process is eventually absorbed with probability 1. The limit ${}_yE_x := \lim_{t\to\infty} {}_yE_x(t)$ is finite and given by the (y,x) entry of $-Q_*^{-1}$.*

Proof. The matrix of transition probabilities at time t is given by $P_t = e^{tQ}$. Then

$$\begin{aligned}
{}_yE_x(t) &= \mathbb{E}_y\left[\int_0^t \mathbf{1}_{\{X_s=x\}}ds\right] \\
&= \int_0^t P_s(y,x)ds \\
&= \left[\int_0^t E^{sQ}ds\right](y,x) \\
&= \left[\int_0^t E^{sQ_*}ds\right](y,x) \qquad \text{because the other states are absorbing} \\
&= Q_*^{-1}\left(e^{tQ_*}-I\right)(y,x).
\end{aligned}$$

By negative-definiteness of Q_* we have $\lim_{t\to\infty} e^{tQ_*}=0$, so this converges to $-Q_*^{-1}$. ☐

We are also interested in the state that the process ends up in. We state the following result in terms of the diagonalisation of Q. Of course, in the special case where Q is not diagonalisable, we can carry out the same construction with the Jordan Normal Form.

Theorem 10.5.2. *Let $v_1, \dots, v_m$ be the right-eigenvectors (columns) of Q with eigenvalue 0 (in other words, a basis for the kernel of Q), such that v_i has a 1 in coordinate $k+i$ and 0 for the remainder of the absorbing states. Then*

$$P_j\big\{\textit{absorbed in state } k+i\big\} = (v_i)_j.$$

Proof. Fix some i, and let $v_j = P_j\big\{\text{absorbed in state } k+i\big\}$. Let X_t be the position of the process at time t. Then $\mathbb{E}_{j'}[v_{X_t}] = P^t(j',\cdot)v$. Note that this function is constant in time (by the Chapman-Kolmogorov equation). By the Forward Equation, for all $t > 0$

$$0 = \frac{d}{dt}P^t(j',\cdot)v = P^t(j',\cdot)Qv,$$

which implies that $Qv = 0$, since $\lim_{t\downarrow 0} P^t$ is the identity matrix. Obviously v has the stated values on the absorbing states. □

10.5.1 The multiple decrements model

The simplest application of this formula is to the multiple decrements model. In that case, we have just a single non-absorbing state 0, and absorbing states $1, 2, \dots, m$. Then $Q_* = (-\lambda_+)$, so that the expected time spent in state 0 is $1/\lambda_+$, which we already knew.

The only non-trivial eigenvector is

$$(-1 \quad \frac{\lambda_1}{\lambda_+} \quad \cdots \quad \frac{\lambda_m}{\lambda_+}).$$

Thus

$$R^{-1} = \begin{pmatrix} -1 & \frac{\lambda_1}{\lambda_+} & \frac{\lambda_2}{\lambda_+} & \cdots & \frac{\lambda_m}{\lambda_+} \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}$$

Thus, the probability of ending up in state i is λ_i/λ_+.

10.5.2 The illness model

Consider now the illness model, with $\sigma = 0.1$, $\delta = 0.1$, $\gamma = 0.5$, and $\rho = 0$. The generator (taking the states in the order H, S, D, is

$$Q = \begin{pmatrix} -0.2 & 0.1 & 0.1 \\ 0 & -0.5 & 0.5 \\ 0 & 0 & 0 \end{pmatrix} \qquad Q_* = \begin{pmatrix} -0.2 & 0.1 \\ 0 & -0.5 \end{pmatrix}$$

We calculate

$$Q_*^{-1} = \begin{pmatrix} -5 & -1 \\ 0 & -2 \end{pmatrix}$$

Thus, a sick individual survives on average 2 years. A healthy individual survives on average 6 years, of which 1 year, on average, is spent sick.

There is only one absorbing state. If we want to study the state that individuals died from (sick or healthy), one approach is to make two absorbing states, one corresponding to death after being healthy, the other after being sick. The core Q_* matrix stays the same, but now Q becomes

$$\begin{pmatrix} -0.2 & 0.1 & 0.1 & 0 \\ 0 & -0.5 & 0 & 0.5 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

The eigenvectors with eigenvalue 0 are

$$\begin{pmatrix} 0.5 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \qquad \begin{pmatrix} 0.5 \\ 1 \\ 0 \\ 1 \end{pmatrix}$$

These tell us that someone who starts sick will with certainty die from the sick state (since there is no recovery in this model), while an initially healthy individual will have probability 1/2 of dying from the healthy or the sick state.

Suppose the recovery rate ρ now becomes 1. Then

$$Q = \begin{pmatrix} -0.2 & 0.1 & 0.1 & 0 \\ 1 & -1.5 & 0 & 0.5 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad Q_* = \begin{pmatrix} -0.2 & 0.1 \\ 1 & -1.5 \end{pmatrix} \qquad Q_*^{-1} = \begin{pmatrix} -7.5 & -.5 \\ -5 & -1 \end{pmatrix}$$

Thus, a healthy individual will now live, on average, 8 years, of which only 0.5 will be sick, and someone who is sick will have 6 years on average, with 1 of those sick.

The eigenvectors with eigenvalue 0 are now

$$\begin{pmatrix} 0.75 \\ 0.5 \\ 1 \\ 0 \end{pmatrix}, \qquad \begin{pmatrix} 0.25 \\ 0.5 \\ 0 \\ 1 \end{pmatrix}.$$

Thus we see that when starting from state H, the probability of transitioning to D from H has gone up to 3/4. Starting from S, the probability is now 1/2 of transitioning to D from S, and 1/2 of transitioning from H. This is consistent with the observation we make from the jump chain, that the healthy person transitions to sick or to dead with equal probabilities. Thus,

$$P_H(\text{last state H}) = \frac{1}{2} + \frac{1}{2} P_S(\text{last state H}) = \frac{1}{2} + \frac{1}{4}.$$