Perturbation Analysis of Inhomogeneous Markov Processes

Bernd Heidergott, Haralambie Leahu, Andreas Löpker, Georg Pflug

Working Paper, Version of February 23, 2012

1 Introduction

Markov processes are widely used in applied probability for studying the timedependent behavior of stochastic models. Typically, Markov processes are analyzed under the assumption that the process is homogeneous. In words, the transition dynamic is time dependent. This assumption is often violated in applications. For example, consider a simple call center with c agents and a single arrival stream of callers. Customers are helped in first-come-first-served order and we are interested in the system behavior over a fixed time period of length T, say, a day. Then, while the arrival stream can be assumed to be Poisson the arrival rate of the process is time dependent: low intensity in the morning hours, which increases over the morning and may fall down around lunch time. This can be modeled as a inhomogeneous Poisson process with rate $\lambda(t)$, for a formal definition see later in the text.

In this paper we will provide a perturbation analysis for inhomogeneous Markov processes. Let $\mathcal{X}_\theta = \{X_\theta(t), t \geq 0\}$ be a continuous-time Markov process on a (at most) denumerable state space S depending on some real-valued parameter θ. Throughout this paper we will denote the transition probability matrix of $X(0)$ to (t) by $P_\theta(0,t)$ and the associated generator by $Q_\theta(t)$, i.e.,

$$\lim_{\Delta \to 0} \frac{1}{\Delta}(P_\theta(t, t+\Delta) - I) = Q_\theta(t).$$

Throughout the paper we assume that $Q(t)$ is conservative, i.e., $\sum_{ij} q_{ij}(t) = 0$ for all $j \in S$. In our analysis we investigate the dependence of $P_\theta(0,t)$ on θ. Specifically, our starting point will be that $Q_\theta(t)$ is differentiable with respect to θ in a suitable sense (to be defined later in the text) and we will deduct formulas for $dP_\theta(0,t)/d\theta$ from this derivative. We will illustrate our approach with models from biology, finance and the afore mentioned call-center model.

The contribution of the paper is the following. We derive a closed form representation for $dP(0,t)/d\theta$. We show how this formula can be used to (i) derive a simple gradient estimators for transient performance characteristics, and (ii) to obtain bounds on the transient performance sensitivities. Numerical examples will illustrate the numerical behavior of the estimators. In addition, we show that for relevant case the derivative of $Q_\theta(t)$ commutes in a suitable way with $Q_\theta(t)$, which leads to a significant simplification of the estimators.

The paper is organized as follows. In Section 2 we provide a discussion of the literature. Perturbation analysis of Markov processes with time-dependent generators is discussed in Section 3. The particular class of Markov processes for which the derivative of the generator commutes with the generator itself is analyzed in Section 4. Higher oder derivatives are discussed in Section 5. In Section 6 we discuss several examples in detail. We conclude the paper with a numerical study of the application of our results to sensitivity analysis of a (simple) call-center in Section 7.

2 Discussion of the literature

Ad an quasi stationary distributions, also cite Massey and Whitt...

The main approaches for gradient estimation are infinitesimal perturbation analysis (IPA) and its variants/extensions, which is a as sample-path based approach [18, 14, 15, 16, 17, **?**], the score function method [19, 20, 21], with relates gradient estimation to differentiation the likelihood ratio of a sample realization, and measure-valued differentiation, which is an operator based approach to sensitivity analysis of Markov processes, [21, 22, **?**]. While perturbation analysis for homogeneous Markov processes is covered by the afore mentioned methods, best to our knowledge no results are known for gradient estimation of inhomogeneous Markov processes. The gap in the literature stems from the fact that for inhomogeneous processes a perturbation of a rate function cannot in a direct way be interpreted as a perturbation of the holding time of a state in the sample path (as IPA would require) nor as a change in the likelihood ration (as the score function would require).

It is worth noting that there exists a stream of active research on perturbation analysis of inhomogeneous Markov processes for so-called singularly perturbed Markov processes, see [13] and the references therein. Here, the Markov processes is assumed to have several ergodic classes and θ parameterizes the rate with which the process jumps from one ergodic class to another. Letting θ tend to zero the process will get stuck in one of the ergodic classes. Investigating the limiting behavior of the Markov processes as θ tends to zero is the topic of this research.

3 Perturbation Analysis of Inhomogenuous Markov Processes

Let $\mathcal{X} = \{X_t, t \geq 0\}$ be a continuous-time ergodic Markov process on a denumerable state space S describing the nominal system. Throughout this paper we will denote its transition matrix by $P(u,t)$, for $0 \le u < t$, more specifically,

$$[P(u,t)](ij) = \mathbb{E}\big[X_t = j \big| X_u = i\big], \quad i,j \in S,$$

and we set $P(0,0) = I$. We first summarize basic properties of homogeneous Markov processes. The infinitesimal generator of an homogeneous Markov transition matrix $P(0,t)$ is denoted by Q, i.e.,

$$\lim_{\Delta \downarrow 0} \frac{1}{\Delta}(P(t) - P(0)) = Q,$$

where $P(0)$ is the identity operator. We assume that $\mathcal{X}$ has a unique stationary distribution, denoted by π. It is well known that this implies $\pi Q = 0$. Moreover it holds that

$$P(t) = e^{Qt} = \sum_{n=0}^{\infty} \frac{t^n}{n!} Q^n, \quad t \ge 0, \tag{1}$$

which implies that

$$P(t) = I + Qt + o(t^2),$$

where I denotes the identity matrix.

Definition 1. *A generator matrix Q is called uniformizable with rate μ if $\lambda = \sup_j |q_{jj}| < \infty$.*

While any finite dimensional generator matrix is uniformizable a classical example of a Markov process on denumerable state space that fails to have this property is the M/M/∞ queue. Note that if Q is uniformizable with rate λ, then Q is uniformizable with rate η for any $\eta > \lambda$.

Let Q be uniformizable with rate μ and introduce the Markov chain P_λ as follows

$$[P_\lambda]_{ij} = \begin{cases} q_{ij}/\lambda & i \ne j \\ 1 + q_{ii}/\lambda & i = j, \end{cases} \tag{2}$$

for $i,j \in S$, or, in shorthand notation,

$$P_\lambda = I + \frac{1}{\lambda} Q,$$

then it holds that

$$P(0,t) = e^{-\lambda t} \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} (P_\lambda)^n, \quad t \ge 0. \tag{3}$$

The Markov chain $\mathcal{X}_\lambda = \{X_n^\lambda : n \ge 0\}$ with transition probability matrix P_λ is called the *sampled chain*. The relationship between $\mathcal{X}$ and $\mathcal{X}_\lambda$ can be expressed as follows. Let $N_\lambda(t)$ denote a Poisson process with rate λ, then $X^\lambda_{N_\lambda(t)}$ and X_t are equal in distribution for all $t \ge 0$.

The classical theory of homogeneous Markov chains extends the the class of inhomogeneous continuous time Markov chains as follows. Let $p(t)$ denote the distribution of

X_t, for $t \geq 0$, with $p(0)$ being the initial distribution, then by the Kolmogorov forward equations for $p(t)$ it holds that

$$\frac{d}{dt}p(t) = p_\theta(t)Q_\theta(t), \quad t > 0,$$

and also

$$P_\theta(s,t) = \exp\left(\int_s^r Q_\theta(u)du\right), \quad t > s \geq 0.$$

This defines a family of operators $\{Q_\theta(t) : t \geq 0\}$, where we assume that $Q_\theta(t)$ is measurable in t with respect to the Borel field on $[0,\infty)$, and bounded as a mapping in t. For non-homogeneous Markov chains the concept of uniformizability (see Definition 1) is extended as follows.

Definition 2. *The bounded family of operators $\{Q(t) : t \geq 0\}$ is called time-varying uniformizable if*

$$\lambda = \sup_{0\leq s\leq t} \sup_j |q_{jj}(s)| < \infty.$$

The family of operators $\{Q(t) : t \geq 0\}$ is said to have finite support if for any row of $Q(t)$ contains only finitely many non-zero elements for any t.

A time-inhomogeneous Markov chain that is time-varying uniformizable can be interpreted as time-inhomogeneous discrete time Markov chain, where the jump times follows a Poisson-λ-process. Provided there is jump at time t, then the transition triggered by this jump is given by

$$P_\lambda(t) = I + \frac{1}{\lambda}Q(t)\,. \tag{4}$$

Suppose that n points of the Poisson-λ-process fall into $[u, t+u]$, for $u \geq 0$, and denote these points by $t_1, \ldots, t_n$. Then the transition probability $P(u, t+u)$, from the initial state to the state at time t is given by $P_\lambda(t_1)\, P_\lambda(t_2), \ldots, P_\lambda(t_n)$. It can be shown that

$$P(u, u+t) = \underbrace{\sum_{n=0}^{\infty} \frac{e^{-\lambda t}(\lambda t)^n}{n!}}_{\text{number of jumps}} \underbrace{\int \cdots \int_{u\leq t_1\leq t_2\leq\ldots\leq t_n\leq t+u} \frac{n!}{t^n}}_{\text{location of jumps}} \quad \underbrace{\prod_{i=1}^{n} P_\lambda(t_i)}_{\text{embedded jump chain}} \quad dt_1 \ldots dt_n,$$

see [11]. Recall that we denote the inhomogeneous Markov processes with generator $Q(t)$ by X_t. In this paper we carefully distinguish the *over all transition probability* from X_0 to X_t given by

$$\left[\prod_{i=1}^{n} P_\lambda(t_i)\right](X_0, X_t),$$

provided that there are n Poisson-λt-process epochs at time instances $t_1 \leq t_2 \leq \cdots \leq t_n$, and the *sample path probability* of observing the state sequence $X_0, X_{t_1}, \ldots, X_{t_n}$ given by

$$\prod_{i=1}^{n} \left[P_\lambda(t_i)\right](X_{t_{i-1}}, X_{t_i}),$$

provided that there are n Poisson-λt-process epochs at time instances $t_1 \leq t_2 \leq \cdots \leq t_n$, From the above it follows that

$$\mathbb{E}\left[f(X_T)\right] = \sum_{n=0}^{\infty} \frac{e^{-\lambda t}(\lambda t)^n}{n!} \frac{n!}{t^n} \times \int \cdots \int_{0 \leq t_1 \leq t_2 \leq \ldots \leq t_n \leq T} \sum_{x \in S} f(x) \left[\prod_{i=1}^{n} P_\lambda(t_i)\right](x_0, x) dt_1 \ldots dt_n,$$

where x_0 is the initial value of X_t, i.e., $X_0 = x_0$, and

$$\mathbb{E}\left[\int_0^T f(X_t)dt\right] = \sum_{n=0}^{\infty} \frac{e^{-\lambda t}(\lambda t)^n}{n!} \frac{n!}{t^n} \times \int \cdots \int_{0 \leq t_1 \leq t_2 \leq \ldots \leq t_n \leq T} \sum_{x_1, \ldots, x_n \in S} \sum_{k=0}^{n} f(x_k)(t_{k+1} - t_k) \left[\prod_{i=1}^{n} (P_\lambda(t_i))(x_{i-1}, x_i)\right] dt_1 \ldots dt_n,$$

with $t_0 = 0$ and $t_{n+1} = T$. **[0]**

This leads to the following simulation algorithm for time integrals of $f(X_t)$ over $[0, T]$.

Algorithm 1 *Let $p(0)$ denote the initial distribution of $\{X_t : 0 \leq t \leq T\}$, and sample $X(0)$ according to $p(0)$.*

- *Simulate a Poisson-λ-process up to time T, which yields N_λ time instances $0 < t_1 < t_2 < \cdots < t_{N_\lambda} < T$, and let $t_0 = 0$ and $t_{N_\lambda + 1} = T$.*
- *For $k = 0$ to $k = n$, sample $\{X_k : 1 \leq k \leq n\}$ where the transition probability from X_{k-1} to X_k is given by $P_\lambda(t_k)$.*

Then it holds that

$$\mathbb{E}\left[\int_0^T f(X_t)dt\right] = \mathbb{E}\left[\sum_{k=0}^{N_\lambda} f(X_k)(t_{k+1} - t_k)\right].$$

Note that one can alternative simulate a Poisson-λ-process, which yields by construction a sequence $t_1, \ldots, t_{N_\lambda}$, and sample N_λ.

Let v be real-valued mapping from S to $\mathbb{R}$ with $\inf_x v(x) = 1$. The v-norm on S of a function f from S to $\mathbb{R}$ is defined as

$$||f||_v = \sup_{x \in S} \frac{|f(x)|}{v(x)}.$$

Note that the above definition implies

$$|f(x)| \leq ||f||_v \, v(x), \quad x \in S. \tag{5}$$

The v-norm is extend to measures μ on S as follows. For $\mu \in S^{\mathbb{R}}$, we set

$$||\mu||_v = \sum_{x \in S} v(x)|\mu(x)|,$$

and the v-norm of a matrix A in $S^{n\times n}$ is defined as

$$||A||_v = \sup_x \frac{1}{v(x)} \sum_y v(y)|A_{xy}|.$$

The following lemma establishes a result for v-norms that we will frequently use in the following.

Lemma 1. *For $\mu, f \in S^{\mathbb{R}}$ and A in $S^{n\times n}$ it holds that*

$$|\mu A f| \leq ||\mu||_v \, ||A||_v \, ||f||_v.$$

Proof: By computation,

$$\begin{aligned}
|\mu A f| &= \left| \sum_i \mu_i \left(\sum_j A_{ij} f_j \right) \right| \\
&\leq \sum_i |\mu_i| \left(\sum_j |A_{ij}| \, |f_j| \right) \\
&= \sum_i |\mu_i| \left(\sum_j |A_{ij}| \, v_j \frac{|f_j|}{v_j} \right) \\
&\leq \sum_i |\mu_i| \left(\sum_j |A_{ij}| \, v_j \right) ||f||_v \\
&\leq \sum_i |\mu_i| \left(v_i \frac{1}{v_i} \sum_j |A_{ij}| \, v_j \right) ||f||_v \\
&\leq \sum_i |\mu_i| v_i ||A||_v \, ||f||_v \\
&= ||\mu||_v \, ||A||_v \, ||f||_v,
\end{aligned}$$

which proves the claim. □

We denote the set of real-valued mappings from S to $\mathbb{R}$ with v-norm bounded by some fixed K by

$$\mathcal{D}_v = \{ f \in S^{\mathbb{R}} : ||f||_v \leq K \}$$

In the following we turn to sensitivity analysis. To begin with the analysis we first have to reflect on the concept of differentiability as we are dealing with Markov chains on a denumerable state space. Indeed, for a finite state space differentiability of Q (resp.

P_λ) can be defined as element-wise differentiability of Q. For a denumerable state-space we have to define differentiability in the weak sense. We call Q $\mathcal{D}_v$-differentiable if

$$\frac{d}{d\theta}[Qf](i) = \sum_j \frac{d}{d\theta} Q(i,j) f(j), \tag{6}$$

for all $j \in S$ and all $f \in \mathcal{D}_v$. We define v-differentiability of P_λ in the same way.

Remark 1. If S is finite, then element-wise differentiability of $Q(t)$ with respect to θ implies v-differentiability for any $v : S \to \mathbb{R}$. If S is denumerable, the v-differentiability of Q is implied by the following condition:

(C) *The entries of $Q(t)$ are continuously differentiable with respect to θ for all t on an open neighborhood Θ of θ and the element-wise derivatives in reach row of $Q(t)v$ are uniformly bounded on Θ for all t i.e., for each t and each row i there exists a sequence $\{a^i_j(t)\}$ such that*

$$\sup_{\theta \in \Theta} \frac{d}{d\theta} \Big| [Q(t)](ij) \Big| v(j) \leq a^i_j(t)$$

for all j and t such that

$$\sum_j a^i_j(t) < \infty$$

for all j and t.

Indeed, condition **(C)** implies (6) since interchanging summation and differentiation in (6) as the series of element-wise derivatives $\sum v(j) d[Q(t)](i,j)/d\theta$ converges uniformly on Θ and the partial sums of $v(j) d[Q(t)](i,j)/d\theta$ are continuous.

Lemma 2. *Consider the family $\{Q(t) : t \geq 0\}$. If $Q(t)$ is $\mathcal{D}_v$-differentiable then there exist transition probability matrices $P^+_Q(t)$ and $P^-_Q(t)$ such that for all $f \in \mathcal{D}_v$ it holds that*

$$\frac{d}{d\theta}\big[Q(t)f\big] = C_Q(t)\big(P^+_Q(t)f - P^-_Q(t)f\big),$$

with $C_Q(t)$ being a matrix with diagonal elements

$$[C_Q(t)](i,i) = \sum_j \max\left(\frac{d}{d\theta}[Q(t)](i,j), 0\right)$$

for $i \in S$ and otherwise zero

Proof: By definition the row sums of $Q(t)$ are zero. Hence, differentiation the elements of $Q(t)$, the row sums of the derivative matrix are also zero. Collect the positive elements of $Q'(t)$ in a matrix $Q^+(t)$ and the negative elements in a matrix $Q^-(t)$, i.e.,

$$[Q^+(t)](i,j) = \max\left(\frac{d}{d\theta}[Q(t)](i,j), 0\right) \quad \text{and} \quad [Q^-(t)](i,j) = \max\left(-\frac{d}{d\theta}[Q(t)](i,j), 0\right),$$

for $i, j \in S$. Since

$$\frac{d}{d\theta} \sum_j [Q(t)](i,j) = \sum_j \frac{d}{d\theta}[Q(t)](i,j) = 0, \tag{7}$$

it holds that

$$\sum_j [Q^+(t)](i,j) = \sum_j [Q^-(t)](i,j) = c_i(t),$$

for all i and t. We now introduce probability transition matrices $P_Q^{\pm}(t)$ such that

$$[P_Q^+(t)](i,j) = \frac{1}{c_i(t)}[Q^+(t)](i,j)$$

and

$$[P_Q^-(t)](i,j) = \frac{1}{c_i(t)}[Q^-(t)](i,j)$$

and we let $C_Q(t)$ be a diagonal matrix with entries $[C_Q(t)](i,i) = c_i(t)$. In case that $\sum_j [Q^+(t)](i,j) = 0$, we let $[P_Q^{\pm}(t)](i,j) = 0$ for $j \neq i$ and $[P_Q^{\pm}(t)](i,i) = 1$, and $[C_Q(t)](i,i) = 1$.

Note that in the above line of argument we have implicitly used that interchanging summation and differentiation in (7) is justified. While this is obvious for matrices with finite support on each row, it follows matrices with infinite support on at least one row by $\mathcal{D}_v$-differentiability. $\square$

Lemma 3. *If $Q(t)$ is $\mathcal{D}_v$-differentiable, then so is $P_\lambda(t)$ and it's $\mathcal{D}_v$-derivative is given by*

$$\frac{d}{d\theta} P_\lambda(t) = \frac{1}{\lambda} C_Q(t)(P_Q^+(t) - P_Q^-(t)).$$

In addition. $P_\lambda(t)$ is $||\cdot||_v$-Lipschitz continuous, i.e., there exists a finite constant M such that

$$||P_{\lambda,\theta+\Delta}(t) - P_{\lambda,\theta}(t)||_v \leq |\Delta|\, M.$$

Proof: Differentiating the expression in (4) and replacing $dQ(t)/d\theta$ by the difference between probability transition kernels, see Lemma 2, yields

$$\frac{d}{d\theta}[P_\lambda(t) f] = \frac{d}{d\theta}\left(\left(I + \frac{1}{\lambda} Q(t) \right) f \right) = \frac{1}{\lambda} C_Q(t)(P_Q^+(t) f - P_Q^-(t) f),$$

which proves the first part of the lemma.

For the proof of the second part of the lemma, note that $(\mathcal{D}_v, ||\cdot||_v)$ is a Banach space. The proof then follows from evoking the basic fact that weak differentiability on a Banach space implies norm Lipschitz continuity with respect to the norm of the Banach space; for details we refer to [7]. $\square$ **[1]**

Lemma 4. *Let $P_\lambda(t)$ be $\mathcal{D}_v$-differentiable for $t \geq 0$. For any $f \in \mathcal{D}_v$ it holds for $0 \leq t_1 \leq \cdots \leq t_n$ that*

$$\frac{d}{d\theta}\left(\prod_{i=1}^{n} P_\lambda(t_i)\right) f = \sum_{j=1}^{n} \prod_{i=1}^{j-1} P_\lambda(t_i)\left(\frac{d}{d\theta} P_\lambda(t_j)\right) \prod_{i=1}^{j-1} P_\lambda(t_i) f$$

Proof: This is the product rule of $\mathcal{D}_v$-differentiation. □ **[2]**

Recall that $p(0)$ denotes the initial distribution which is assumed to be independent of θ.

Theorem 1. *There exists an open neighborhood of θ, called Θ, such that $\{Q_\theta(t) : t \geq 0\}$ is uniformly time-varying uniformizable on Θ, i.e.,*

$$\lambda = \sup_{0 \leq s \leq t} \sup_{j} \left| Q_\theta(s)](j,j) \right| < \infty,$$

for all $\theta \in \Theta$. If $Q_\theta(t)$ is $\mathcal{D}_v$-differentiable of a neighborhood Θ of θ, and if

$$\sup_{u \in [0,t]} \sup_{\theta \in \Theta} ||P_{\lambda,\theta}(t)||_v < \infty,$$

then for all $f \in \mathcal{D}_v$ it holds that

$$\begin{aligned} \frac{d}{d\theta} p(0) P(0,t) f \quad &= \quad \frac{1}{\lambda} \sum_{n=1}^{\infty} \frac{e^{-\lambda t} (\lambda t)^n}{n!} \frac{n!}{t^n} \sum_{j=1}^{n} \int \cdots \int_{0 \leq t_1 \leq t_2 \leq \ldots \leq t_n} C_Q(t_j) \times \\ & p(0) \left(\left(\prod_{i=1}^{j-1} P_\lambda(t_i) P_Q^{+}(t_j) \prod_{i=j+1}^{n} P_\lambda(t_i) \right) f - \left(\prod_{i=1}^{j-1} P_\lambda(t_i) P_Q^{-}(t_j) \prod_{i=j+1}^{n} P_\lambda(t_i) \right) f \right) dt_1 \end{aligned}$$

Proof: By simple algebra it holds

$$\begin{aligned} & \left| p(0) \prod_{i=1}^{n} P_{\lambda,\theta+\Delta}(t_i) f - p(0) \prod_{i=1}^{n} P_{\lambda,\theta}(t_i) f \right| \\ &= \sum_{j=1}^{n} \left| p(0) \prod_{i=1}^{j-1} P_{\lambda,\theta+\Delta}(t_i) (P_{\lambda,\theta+\Delta}(t_j) - P_{\lambda,\theta}(t_j)) \prod_{i=j+1}^{n} P_{\lambda,\theta}(t_i) f \right|, \end{aligned}$$

with is bounded by Lemma 1 by

$$\leq ||p(0)||_v |\Delta| n D^{n-1} M ||f||_v,$$

with

$$D = \sup_{u \in [0,t]} \sup_{\theta \in \Theta} ||P_\lambda(t)||_v.$$

Hence,

$$\frac{n!}{t^n}\int_{0\le t_1\le t_2\le \ldots \le t_n}\!\!\cdots\int \frac{1}{|\Delta|}\sum_{j=1}^{n}\left|\int_{0\le t_1,\ldots\le t_n}\left(\prod_{i=1}^{n}P_{\lambda,\theta+\Delta}(t_i)f-\prod_{i=1}^{n}P_{\lambda,\theta}(t_i)\right)f\,dt_1\cdots dt_n\right|$$
$$\le ||p(0)||_v\,||f||_v nD^{n-1}M.$$

Let $N(t)$ be Poisson-λt-distributed. Since

$$\mathbb{E}\big[N(t)\,D^{N(t)-1}\big]=\lambda tDe^{\lambda t(D1)}<\infty,$$

it follows from the Dominated Convergence Theorem that differentiation with respect to θ and integrating with respect to the Poisson distribution can be interchanged. The particular expression for the derivative then follows from Lemma 4. □

Theorem 2. *Under the conditions put forward in Theorem 1, it holds for all $f\in\mathcal{D}_v$ that*

$$\begin{aligned}\frac{d}{d\theta}\mathbb{E}\left[\int_0^T f(X_t)dt\right] &= \frac{1}{\lambda}\sum_{n=0}^{\infty}\frac{e^{-\lambda t}(\lambda t)^n}{n!}\frac{n!}{t^n}\int_{0\le t_1\le t_2\le\ldots\le t_n\le T}\!\!\cdots\int\sum_{x_1,\ldots,x_n\in S}C_Q(t_j)\sum_{k=0}^{n}f(x_k)(t_{k+1}-t_k)\times\\ &\left(\prod_{i=1}^{j-1}[P_\lambda(t_i)](x_{i-1},x_i)[P_Q^+(t_j)](x_{j-1},x_j)\prod_{i=j+1}^{n}[P_\lambda(t_i)](x_{i-1},x_i)\right.\\ &\left.-\prod_{i=1}^{j-1}[P_\lambda(t_i)](x_{i-1},x_i)[P_Q^-(t_j)](x_{j-1},x_j)\prod_{i=j+1}^{n}[P_\lambda(t_i)](x_{i-1},x_i)\right)dt_1\ldots dt_n.\end{aligned}$$

Proof: By simple algebra it holds

$$\begin{aligned}&\left|\sum_{x_1,\ldots,x_n\in S}\sum_{k=0}^{n}f(x_k)(t_{k+1}-t_k)\left(p(0)\prod_{i=1}^{n}[P_{\lambda,\theta+\Delta}(t_i)](x_{i-1},x_i)-p(0)\prod_{i=1}^{n}[P_{\lambda,\theta}(t_i)](x_{i-1},x_i)\right)\right|\\
&\le T\sum_{x_1,\ldots,x_n\in S}\sum_{k=0}^{n}|f(x_k)|\left|p(0)\prod_{i=1}^{n}[P_{\lambda,\theta+\Delta}(t_i)](x_{i-1},x_i)-p(0)\prod_{i=1}^{n}[P_{\lambda,\theta}(t_i)](x_{i-1},x_i)\right|\\
&=T\sum_{k=1}^{n}\sum_{x_1,\ldots,x_k\in S}\left|p(0)\prod_{i=1}^{k}[P_{\lambda,\theta+\Delta}(t_i)](x_{i-1},x_i)-p(0)\prod_{i=1}^{k}[P_{\lambda,\theta}(t_i)](x_{i-1},x_i)\right||f(x_k)|\\
&=T\sum_{k=1}^{n}\sum_{x_1,\ldots,x_k\in S}\left|p(0)\sum_{j=1}^{k}\prod_{i=1}^{j-1}[P_{\lambda,\theta+\Delta}(t_i)](x_{i-1},x_i)\times\right.\\
&\qquad\left.[P_{\lambda,\theta+\Delta}(t_i)](x_{i-1},x_i)-[P_{\lambda,\theta}(t_i)](x_{i-1},x_i))\prod_{i=j+1}^{k}[P_{\lambda,\theta}(t_i)](x_{i-1},x_i)\right||f(x_k)|\end{aligned}$$

which is bounded by Lemma 1 by

$$\le T||p(0)||_v|\Delta|n^2D^{n-1}M||f||_v,$$

with

$$D = \sup_{u\in[0,t]} \sup_{\theta\in\Theta} ||P_\lambda(t)||_v.$$

Let $N(t)$ be Poisson-λt-distributed. Since

$$\mathbb{E}\big[(N(t))^2 \, D^{N(t)-1}\big] = \lambda t D e^{\lambda t(D-1)}(\lambda t D + 1) < \infty,$$

it follows from the Dominated Convergence Theorem that differentiation with respect to θ and integrating with respect to the Poisson distribution can be interchanged. The particular expression for the derivative then follows from Lemma 4. □

The result put forward in Theorem 2 leads to the following estimation algorithm.

Algorithm 2 *Let $p(0)$ denote the initial distribution.*

- *Simulate a Poisson-λ-process up to time T, which yields N_λ time instances $0 < t_1 < t_2 < \cdots < t_{N_\lambda} < T$, and let $t_0 = 0$ and $t_{N_\lambda+1} = T$. Construct $\{X_k : 0 \le k \le N_\lambda + 1\}$ according to Algorithm 1.*
- *For given state x, let $X_1^+(x)$ be distributed according to $P_\lambda^+(x,\cdot)$ and $X_1^-(x)$ be distributed according to $P_\lambda^-(x,\cdot)$.*
- *For $k \ge 1$, let the transition probability from $X_k^\pm(x)$ to $X_{k+1}^\pm(x)$ is given by $P_\lambda(t_{k+1})$.*

Then it holds that

$$\frac{d}{d\theta}\mathbb{E}\left[\int_0^T f(X_t)dt\right] = \mathbb{E}\left[\sum_{j=1}^{N_\lambda} c_j(X_{j-1}) \sum_{k=1}^{N_\lambda - j} \Big(f(X_k^+(X_{j-1})) - f(X_k^-(X_{j-1}))\Big)(t_{j+k} - t_{j+k-1})\right],$$

with $c_j(y) = (C_Q(t_j)/\lambda)_{yy}$. Note the the perturbed processes have the same jump epochs, namely, t_k, then the nominal process.

The above algorithm yields the following bound for the derivative.

Corollary 1. *If $|f|$ is bounded by c, then*

$$\left|\frac{d}{d\theta}\mathbb{E}\left[\int_0^T f(X_t)dt\right]\right| \le 4c(\lambda T)^2 \sup_{u\in[0,t]} \max_y (C_Q(u))_{yy}.$$

Theorem 3. *Under the conditions put forward in Theorem 1, it holds that*

$$\frac{d}{d\theta}P(0,t) = \int_0^t P(0,u)Q'(u)P(u,t)du,$$

or, equivalently,

$$\frac{d}{d\theta}P(0,t) = t\mathbb{E}\big[P(0,U)Q'(U)P(U,t)\big],$$

with U uniformly distributed on $[0,t]$ independent of everything else.

Proof: By Lemma 2, we replace $C_Q(t)(P_Q^+(t) - P_Q^-(t))$ in the statement of Theorem 1 by $(1/\lambda)Q'(t)$, which yields

$$\frac{d}{d\theta}P(0,t) \;=\; \sum_{n=1}^{\infty} e^{-\lambda t}\lambda^{n-1} \sum_{j=1}^{n} \int_{0\le t_1\le t_2\le \ldots \le t_n}\!\!\cdots\int \left(\prod_{i=1}^{j-1} P_\lambda(t_i) Q'(t_j) \prod_{i=j+1}^{n} P_\lambda(t_i) \right) dt_1 \ldots dt_n.$$

Note that

$$\begin{aligned}
&\sum_{j=1}^{n} \int_{0\le t_1\le t_2\le \ldots \le t_n}\!\!\cdots\int \left(\prod_{i=1}^{j-1} P_\lambda(t_i) Q'(t_j) \prod_{i=j+1}^{n} P_\lambda(t_i) \right) dt_1 \ldots dt_n \\
&= \sum_{j=1}^{n} \int_{0\le t_1\le\cdots\le t_{j-1}\le u\le t_{j+1}\le\cdots\le t_n}\!\!\cdots\int \left(\prod_{i=1}^{j-1} P_\lambda(t_i) Q'(u) \prod_{i=j+1}^{n} P_\lambda(t_i) \right) dt_1 \ldots dt_{j-1}\, du\, dt_{j+1} \ldots dt_n,
\end{aligned}$$

relabeling the variables t_j yields

$$= \sum_{j=1}^{n} \int_{0\le t_1\le\cdots\le t_{j-1}\le u\le t_j\le\cdots\le t_{n-1}}\!\!\cdots\int \left(\prod_{i=1}^{j-1} P_\lambda(t_i) Q'(u) \prod_{i=j}^{n-1} P_\lambda(t_i) \right) dt_1 \ldots dt_{j-1}\, du\, dt_j \ldots dt_{n-1}$$

and the overall expression for the derivative becomes

$$\begin{aligned}
&\frac{d}{d\theta}P(0,t) \\
&= \sum_{n=0}^{\infty} e^{-\lambda t}\lambda^{n} \sum_{j=1}^{n} \int_{0\le t_1\le\cdots\le t_{j-1}\le u\le t_j\le\cdots\le t_n}\!\!\cdots\int \left(\prod_{i=1}^{j-1} P_\lambda(t_i) Q'(u) \prod_{i=j}^{n} P_\lambda(t_i) \right) dt_1 \ldots dt_{j-1}\, du\, dt_j \ldots dt_n,
\end{aligned}$$

where we set $\sum_{j=1}^{0} a_j = 0$; rearranging sums yields

$$= \sum_{n=0}^{\infty}\sum_{k=0}^{\infty} e^{-\lambda t}\lambda^{n+k} \int_{0\le t_1\le\cdots\le t_n\le u\le t_{n+1}\le\cdots\le t_{n+k}}\!\!\cdots\int \left(\prod_{i=1}^{n} P_\lambda(t_i) Q'(u) \prod_{i=n+1}^{n+k} P_\lambda(t_i) \right) dt_1 \ldots dt_n\, du\, dt_{n+1} \ldots dt_{n+k}$$

which can be written as

$$\begin{aligned}
&= \int_0^t \sum_{n=0}^{\infty} \frac{e^{-\lambda u}(\lambda u)^n}{n!} \frac{n!}{u^n} \sum_{k=0}^{\infty} \frac{e^{-\lambda(t-u)}(\lambda(t-u))^k}{k!} \\
&\quad \times \frac{k!}{(t-u)^k} \int_{0\le t_1\le\cdots\le t_n\le u\le t_{n+1}\le\cdots\le t_{n+k}}\!\!\cdots\int \left(\prod_{i=1}^{n} P_\lambda(t_i) Q'(u) \prod_{i=n+1}^{n+k} P_\lambda(t_i) \right) dt_1 \ldots dt_{n+k} du.
\end{aligned}$$

Rewriting the above expression in operator form, yields

$$\frac{d}{d\theta}P(0,t) = \int_0^t P(0,u)Q'(u)P(u,t)dt.$$

Letting U be uniformly distributed on $[0,t]$ independent of everything else, we arrive at

$$\frac{d}{d\theta}P(0,t) = t\mathbb{E}\big[P(0,U)Q'(U)P(U,t)\big],$$

which concludes the proof. $\square$

Theorem 3 yields the following algorithm.

Algorithm 3 *Let $p(0)$ denote the initial distribution.*

- *Generate U uniformly distributed on $[0,t]$.*
- *Simulate a Poisson-λ-process up to time U, which yields N_λ time instances $0 < t_1 < t_2 < \cdots < t_{N_\lambda} < U$, and let $t_0 = 0$ and $t_{N_\lambda+1} = U$. Construct $\{X_k : 0 \le k \le N_\lambda\}$ according to Algorithm 1 and set $X(U) = X_{N_\lambda}$.*
- *For given state x, let $X_1^+(x)$ be distributed according to $P_\lambda^+(x,\cdot)$ and $X_1^-(x)$ be distributed according to $P_\lambda^-(x,\cdot)$.*
- *Simulate a Poisson-λ-process from time U up to time T, which yields N'_λ time instances $U < t_1 < t_2 < \cdots < t_{N'_\lambda} < T$, and let $t_{N'_\lambda+1} = T$.*
- *For $1 \le k \le N'$, let the transition probability from $X_k^\pm(x)$ to $X_{k+1}^\pm(x)$ be given by $P_\lambda(\tau_{k+1})$. Construct $\{X_k^\pm(x) : 0 \le k \le N'_\lambda\}$ according to Algorithm 1.*

Then it holds that

$$\frac{d}{d\theta}\mathbb{E}\left[\int_0^T f(X_t)dt\right] = T\,\mathbb{E}\left[c(X(U))\sum_{k=1}^{N'_\lambda+1}\Big(f(X_k^+(X(U))) - f(X_k^-(X(U))\Big)(\tau_k - \tau_{k-1})\right],$$

with $c(y) = (1/\lambda)[C_Q(t_N)](y,y)$.

4 Perturbation Analysis for Commutative Inhomogeneous Markov Processes

Definition 3. *Let $Q'_\theta(t) = Q'_\theta$ be independent of t. The family $\{Q_\theta(t) : t \ge 0\}$ is said to be commutative if*

$$Q_\theta(t)\,Q'_\theta = Q'_\theta Q_\theta(t),$$

for all $t \ge 0$.

Lemma 5. *If the family $\{Q_\theta(t) : t \ge 0\}$ is commutative, then it holds that*

$$P_\lambda(t)Q'_\theta = Q'_\theta P_\lambda(t),$$

for all $t \ge 0$.

Proof: By (4) it follows

$$P_\lambda(t)Q'_\theta = \left(I + \frac{1}{\lambda}Q_\theta(t)\right)Q'_\theta = Q'_\theta + \frac{1}{\lambda}Q'_\theta Q_\theta(t) = Q'_\theta\left(I + \frac{1}{\lambda}Q_\theta(t)\right) = Q'_\theta P_\lambda(t),$$

for any t, where the second equality follows from the fact that $\{Q_\theta(t) : t \geq 0\}$ is commutative. □

Theorem 4. *Let $\{Q_\theta(t) : t \geq 0\}$ be commutative such that $Q'_\theta(t)$ is independent of t. Under the conditions put forward in Theorem 1 it then holds that*

$$\frac{d}{d\theta}P(0,t) = t\,Q'_\theta P(0,t) = t\,P(0,t)Q'_\theta,$$

for any t.

Proof We have assumed that Q'_θ is independent of t. Hence, Theorem 1 yields

$$\frac{d}{d\theta}P(0,t) \;=\; \sum_{n=1}^{\infty} e^{-\lambda t}\lambda^{n-1}\sum_{j=1}^{n} \int\cdots\int_{0\leq t_1\leq t_2\leq\ldots\leq t_n} \left(\prod_{i=1}^{j-1} P_\lambda(t_i)Q'_\theta \prod_{i=j+1}^{n} P_\lambda(t_i)\right) dt_1\ldots dt_n.$$

By commutativity of $\{Q_\theta(t) : t \geq 0\}$, Lemma 5 yields

$$\frac{d}{d\theta}P(0,t) \;=\; Q'_\theta \sum_{n=1}^{\infty} e^{-\lambda t}\lambda^{n-1}\sum_{j=1}^{n} \int\cdots\int_{0\leq t_1\leq t_2\leq\ldots\leq t_n} \prod_{\substack{i=1,\\ i\neq j}}^{n} P_\lambda(t_i)dt_1\ldots dt_n,$$

relabeling the t_i's and noting that integrating the free variable over $[0,t]$ yields t, gives

$$\begin{aligned}
&= tQ'_\theta \sum_{n=1}^{\infty} e^{-\lambda t}\lambda^{n-1} \int\cdots\int_{0\leq t_1\leq t_2\leq\ldots\leq t_{n-1}} \prod_{i=1}^{n-1} P_\lambda(t_i)dt_1\ldots dt_{n-1} \\
&= tQ'_\theta \sum_{n=0}^{\infty} e^{-\lambda t}\lambda^{n} \int\cdots\int_{0\leq t_1\leq t_2\leq\ldots\leq t_n} \prod_{i=1}^{n} P_\lambda(t_i)dt_1\ldots dt_n \\
&= tQ'_\theta \sum_{n=0}^{\infty} \frac{e^{-\lambda t}(\lambda t)^n}{n!}\frac{n!}{t^n} \int\cdots\int_{0\leq t_1\leq t_2\leq\ldots\leq t_n} \prod_{i=1}^{n} P_\lambda(t_i)dt_1\ldots dt_n \\
&= tQ'_\theta P_\theta(0,t).
\end{aligned}$$

The proof of the second equality follows from the same line of argument and is therefore omitted.

□

We call

$$\frac{d}{d\theta}P(0,t) = t\,Q'_\theta P(0,t)$$

the *backward sensitivity expression*, and

$$\frac{d}{d\theta}P(0,t) = t\,P(0,t)Q'_\theta,$$

the *forward sensitivity expression.*

The result put forward in Theorem can be interpreted as follows. Let $X_t(x)$ denote the Markov processes started at time 0 with initial value $X(0) = x$. Writing $Q' = C_Q(P_Q^+ - P_Q^-)$ the derivative of $\mathbb{E}[\int_0^T f(X_t)dt]$ can be obtained from

$$\mathbb{E}\left[c(X(0))\left(\int_0^T f(X_t(X^+))dt - \int_0^T f(X_t(X^-))dt\right)\right], \tag{8}$$

whereX^+ is distributed according to $p(0)P_Q^+$ and whereX^- is distributed according to $p(0)P_Q^-$. The normalizing variable $c(X(0))$ is given by $(C_Q)_{yy}$ for $y = X(0)$, the initial state according to $p(0)$. The derivative expression in (8) is called the *backward sensitivity estimator* as it is built on the backward sensitivity expression for $dP(0,t)/d\theta$. In the same vein, a *forward sensitivity estimator* can be obtained by pushing the perturbation Q' to the end of time interval $[0,T]$.

Theorem 5. *Let $\{Q_\theta(t) : t \geq 0\}$ be commutative such that $Q'_\theta(t)$ is independent of t. Then for any t, the forward sensitivity estimator is given by*

$$\frac{d}{d\theta}\mathbb{E}^x\left[\int_0^T f(X_t)dt\right] = T\sum_y Q'(x,y)\mathbb{E}^y\left[\int_0^T f(X_t)dt\right]$$

and backward sensitivity estimator is given by

$$\frac{d}{d\theta}\mathbb{E}^x\left[\int_0^T f(X_t)dt\right] = T\,\mathbb{E}^x\left[\int_0^T (Q'f)(X_t)dt\right].$$

Proof: We first proof the backward sensitivity estimator. Using Theorem 4, we may compute as follows

$$\begin{aligned}
\frac{d}{d\theta}\mathbb{E}^x\left[\int_0^T f(X_t)dt\right] &= \frac{d}{d\theta}\int_0^T\left(\sum_y f(y)[P(0,t)](x,y)\right)dt \\
&= \int_0^T\left(\sum_y f(y)[Q'P(0,t)](x,y)\right)dt \\
&= \int_0^T\left(\sum_y f(y)\sum_z Q'(x,z)[P(0,t)](z,y)\right)dt \\
&= \sum_z Q'(x,z)\int_0^T\sum_y f(y)[P(0,t)](z,y)\,dt,
\end{aligned}$$

which concludes the first part of the proof. As for the forward sensitivity estimator,

$$\begin{aligned}
\frac{d}{d\theta}\mathbb{E}^x\left[\int_0^T f(X_t)dt\right] &= \frac{d}{d\theta}\int_0^T \left(\sum_y f(y)[P(0,t)](x,y)\right) dt \\
&= \int_0^T \left(\sum_y f(y)[P(0,t)Q'](x,y)\right) dt \\
&= \int_0^T \left(\sum_y f(y)\sum_z [P(0,t)](x,z)Q'(z,y)\right) dt \\
&= \int_0^T \sum_z \left(\sum_y Q'(z,y)f(y)\right) P(0,t)(x,z)\, dt,
\end{aligned}$$

which proves the claim $\square$

5 Higher Order Derivatives

Repeating the arguments put forward in the previous section, expressions for higher order derivatives can be obtained. The general statement is as follows

$$\frac{d^n}{d\theta^n}P(0,t) = \sum_{\substack{l_1,\ldots,l_k \\ l_1+l_2+\cdots+l_k=n}} \frac{n!}{l_1!l_2!\cdots l_k!}\int_{[0,T]^k} P(U_0,U_1)\prod_{i=1}^n Q^{(l_i)}(U_i)P(U_i,U_{i+1})\,dU_1\cdots dU_n,$$

with $U_{n+1} = T$, and $Q^{(k)}$ denoting the kth order derivative of Q. For many models, the generator is an affine linear mapping in the parameters, and in this case the above expression for the nth order derivative simplifies to

$$\frac{d^n}{d\theta^n}P(0,t) = \frac{n!}{t^n}\int_{0\le t_1\le\cdots\le t_n\le t} P(0,t_1)Q'(t_1)P(t_1,t_2)Q'(t_2)\cdots Q'(t_n)P(t_n,t)dt_1\cdots dt_n.$$

In the commutative case it holds that

$$\frac{d^n}{d\theta^n}P(0,t) = \lambda^n\, t^n\,(Q')^n P(0,t) = \lambda^n\, t^n\, P(0,t)(Q')^n,$$

for any t and any n. In the case that $(Q')^n = 0$ for sufficiently large n, this already yields a Taylor series expansion for $P(0,t)$.

6 Applications

In this paper we will discuss three models out of different areas of applied probability. The models have been chosen to illustrate the application of our approach to different types of models.

6.1 DNA Model form Biology

The following two examples are inspired by DNA substitution models in biology/genetics.

Example 1. *The K80 Model: Let $\{X_t : t \geq 0\}$ on $\mathbb{S} = \{1, 2, 3, 4\}$ be governed by the following generator*

$$Q_\theta = \begin{bmatrix} -(\theta+2) & \theta & 1 & 1 \\ \theta & -(\theta+2) & 1 & 1 \\ 1 & 1 & -(\theta+2) & \theta \\ 1 & 1 & \theta & -(\theta+2) \end{bmatrix}.$$

One can easily see that Q_θ is weakly differentiable w.r.t. θ, having weak derivative Q'_θ given by

$$Q'_\theta = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 1 & -1 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} - \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = Q^+_\theta - I,$$

i.e., Q^-_θ is the identity matrix. Hence the "negative" version of the process will be the original process. Moreover, Q^+_θ represents a deterministic jump from the current state to a neighboring one. In addition, one can check that Q_θ and Q'_θ commute in this case, using the backward sensitivity estimator yields

$$\frac{d}{d\theta}\mathbb{E}[f(X_t)] = \mathbb{E}[f(\phi(X_t)) - f(X_t)],$$

where $\phi(1) = 2, \phi(2) = 1, \phi(3) = 4,$ and $\phi(4) = 3$.

Example 2. *The JC69 Model: Let $\{X_t : t \geq 0\}$ on $\mathbb{S} = \{1, 2, 3, 4\}$ be governed by the following generator*

$$A_\theta = \begin{bmatrix} -3\theta & \theta & \theta & \theta \\ \theta & -3\theta & \theta & \theta \\ \theta & \theta & -3\theta & \theta \\ \theta & \theta & \theta & -3\theta \end{bmatrix}.$$

Again, A_θ is weakly differentiable w.r.t. θ, having weak derivative A'_θ given by

$$A'_\theta = \begin{bmatrix} -3 & 1 & 1 & 1 \\ 1 & -3 & 1 & 1 \\ 1 & 1 & -3 & 1 \\ 1 & 1 & 1 & -3 \end{bmatrix} = 3\begin{bmatrix} 0 & 1/3 & 1/3 & 1/3 \\ 1/3 & 0 & 1/3 & 1/3 \\ 1/3 & 1 & 0 & 1 \\ 1/3 & 1/3 & 1/3 & 0 \end{bmatrix} - \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Note first that A_θ and A'_θ commute in this case. Therefore, the differential formula reduces to

$$\partial_\theta P_\theta(t) = tA'_\theta P_\theta(t) = tP_\theta(t)A'_\theta.$$

For $1 \le i \le 4$, introduce $\phi(i)$ uniformly distributed on $\{1,2,3,4\}\backslash\{i\}$. Them, the backward sensitivity formula leads to

$$\frac{\partial}{\partial\theta}\mathbb{E}[f(X_t)] = 3\mathbb{E}[f(\phi(X_t)) - f(X_t)].$$

In this case, one can check directly the validity of the above estimator since the expression of $P_\theta(t)$ can be obtained in closed form. Indeed, in this case we have $P_\theta(t) = \exp(tA_\theta)$; direct calculation yields

$$P_\theta(t) = \frac{1}{4}\begin{bmatrix} 1+3e^{-4t\theta} & 1-e^{-4t\theta} & 1-e^{-4t\theta} & 1-e^{-4t\theta} \\ 1-e^{-4t\theta} & 1+3e^{-4t\theta} & 1-e^{-4t\theta} & 1-e^{-4t\theta} \\ 1-e^{-4t\theta} & 1-e^{-4t\theta} & 1+3e^{-4t\theta} & 1-e^{-4t\theta} \\ 1-e^{-4t\theta} & 1-e^{-4t\theta} & 1-e^{-4t\theta} & 1+3e^{-4t\theta} \end{bmatrix} = \frac{1}{4}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} - \frac{1}{4}e^{-4t\theta}A'_\theta.$$

We conclude that $\partial_\theta P_\theta(t) = tA'_\theta P_\theta(t) = te^{-4t\theta}A'_\theta$, hence

$$\partial_\theta \mathbb{E}^x_\theta[f(X_t)] = tA'_\theta P_\theta(t) f(x) = te^{-4t\theta}[A'_\theta f](x) = te^{-4t\theta}\sum_{y \ne x}[f(y) - f(x)]. \tag{9}$$

On the other hand, the distribution of $\phi(x)$ is given by[1] $[0, 1/3, 1/3, 1/3]^*$ *while that of x is $[1, 0.0.0]^*$. Since $P_\theta(t)$ is known, the distributions of $X_t^{\phi(x)}$ and X_t^x are given by the vectors*

$$P_\theta(t)^*[0, 1/3, 1/3, 1/3]^* = \left[\frac{1-e^{-4t\theta}}{4}, \frac{3+e^{-4t\theta}}{12}, \frac{3+e^{-4t\theta}}{12}, \frac{3+e^{-4t\theta}}{12}\right]^*,$$

and

$$P_\theta(t)^*[1,0,0,0]^* = \left[\frac{1+3e^{-4t\theta}}{4}, \frac{1-e^{-4t\theta}}{4}, \frac{1-e^{-4t\theta}}{4}, \frac{1-e^{-4t\theta}}{4}\right]^*,$$

respectively. To see now that $\partial_\theta \mathbb{E}^x_\theta[f(X_t)] = t\mathbb{E}^x_\theta[W_f(t,0)]$, we calculate the latter, obtaining

$$t\mathbb{E}^x_\theta[W_f(t,0)] = 3t\left[\frac{3+e^{-4t\theta}}{12} - \frac{1-e^{-4t\theta}}{4}\right]\sum_{y\ne x} f(y) + 3t\left[\frac{1-e^{-4t\theta}}{4} - \frac{1+3e^{-4t\theta}}{4}\right] f(x).$$

After performing straightforward calculations, the expression above is the same as the r.h.s. in (9).

6.2 The M(t)/M/c+M Queue

Let $\mathcal{X}^*$ be the ergodic queue-length process with states $(x_1, x_2)^t \in S = \mathbb{N}_0 \times \mathbb{N}_0$, where x_1 denotes the number of customers either in service or waiting in the queue and x_2 refers to the impatient customers intending to recall. We regard this model as an open Jackson

[1] We may re-label the elements of $\mathbb{S}$ so that x comes on the first position.

network with two nodes. External arrivals - modeling first callers - enter the system with rate $\lambda(t)$ at the first node where they are served by c servers each providing service at rate μ. We model the timedependence of the arrival rate through the sinusoidal arrival rate function

$$\lambda(t) = a + b\sin(t), \quad t \geq 0,$$

see, for example, [3] for an motivation of this rate function from call-center analysis. However, callers abandon if their waiting time exceeds their exponentially-α distributed patience. Customers who hung up are considered to enter a second node - the orbit - which they leave by recalling after an exponentially-β distributed time. Therefore the first node is an M(t)/M/c queue with abandonments while the latter one is an M/M/∞ queue. Transition rates for $x, y \in S$ are given as follows

$$Q_{x,y}(t) = \begin{cases} \lambda(t) & y = (x_1+1, x_2),\ x_1, x_2 \geq 0 \\ \min\{x_1, c\}\mu & y = (x_1-1, x_2),\ x_1 \geq 1, x_2 \geq 0 \\ \max\{x_1 - c, 0\}\alpha & y = (x_1-1, x_2+1),\ x_1 \geq 1, x_2 \geq 0 \\ x_2\beta & y = (x_1+1, x_2-1),\ x_1 \geq 0, x_2 \geq 1 \\ -(\lambda(t) + \min\{x_1, c\}\mu + \max\{x_1 - c, 0\}\alpha + x_2\beta) & y = (x_1, x_2),\ x_1, x_2 \geq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

An overview of the system is provided in Figure 1.

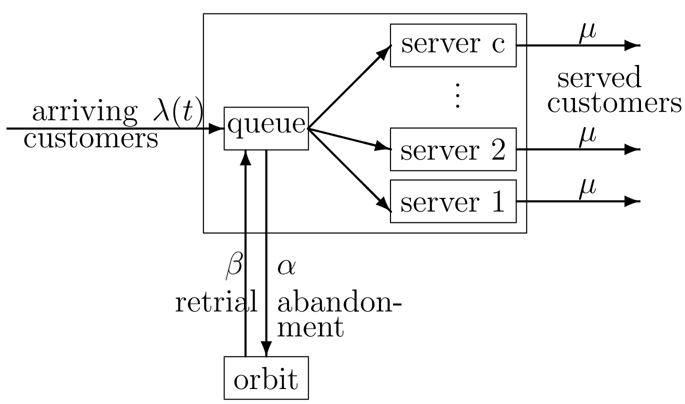

Figure 1: Structure of an M(t)/M/c+M queueing system with abandonment and retrial

Retrial queues have been intensively studied in the literature, see, for example, [5], and [8] and the references therein, as well as the two survey papers [12] and [4]. Retrial queues are especially useful for modelling call centers, see, for example, [6], but have also some further applications of which four are presented in [12]. For a recent overview on retrial queues we refer to [1]. Even for time-independent arrival rates, i.e., for $\lambda(t) = a, t \geq 0$, a closed form solution for the stationary distribution of the M/M/c retrial queue is only available for $c = 1, 2$, see [5], for larger values of c only approximations are known [1].

Differentiating Q with respect to $\theta = b$ yields

$$\frac{\partial}{\partial b} Q_{x,y}(t) = \begin{cases} \sin(t) & y = (x_1+1, x_2),\ x_1, x_2 \geq 0 \\ -\sin(t) & y = (x_1, x_2),\ x_1, x_2 \geq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

In words, perturbing b in the arrival rate has only effect if an arrival occurs or the state remains unchanged. Hence $\partial Q / \partial \mu$ can be written as the difference between two re-scaled transition kernels P^+ and P^-, where P^+ leaves forces an arrival, i.e., $P^+(x_1, x_2; x_1 + 1, x_2) = 1$ and all other entries are zero, and P^- is the identify transition kernel that leaves the state unchanged. More specifiaclly.

$$\frac{\partial}{\partial b} Q(t) = \sin(t)(P^+ - P^-).$$

and we obtain

$$\frac{\partial}{\partial b} P(0;T) = \int_0^T \sin(t) P(0;t) P^+ P(0;T)\, dt - T(1-\cos(t))\, P(0,T),$$

The above representation leads to the following simulation scheme. For $0 \leq U \leq t$, let $X^+(T,U)$ be defined as follows. The process starts at time zero in x. Until time U the process evolves according to $P(0,U)$. At time U an instantaneous transition takes place where one customer is added to the queue (i.e., one additional caller is generated). After this instantaneous perturbation, the process evolves during the remaining $T-U$ time units according to $P(T-U)$. Choosing U uniformly distributed on $[0,T]$ and independent of everything else, we arrive at

$$\frac{\partial}{\partial \mu} \mathbb{E}\big[f(X(t)) \big| X(0) = x\big] = T\, \mathbb{E}\big[f(X(T,U)) - (1-\cos(t)) f(X(T)) \big| X(0) = x\big].$$

7 A Numerical Study

7.1 Sensitivity Analysis

7.2 Fitting the Model

Compute the gradient of $\mathbb{E}[f(S_t)]$ with respect to a and b. Minimize the distance of $\mathbb{E}_{a,b}[f(S_t)]$ estimated by the model the true observed value of the performance, say, d. Using straightforward stochastic approximation we can now find (a,b) that minimizes $(\mathbb{E}_{a,b}[f(S_t)] - d)^2$.

Notes To Ourselves

Here is a list of issues we have to discuss. As a general remark, the notation has to be unified but up to now I still haven't found a good notation that satisfies all needs.

(0) Massey and Whitt only prove the formula for $P(0,t)$. The extension of this result to $E[\int f(X_t)dt]$ seems logical but, to be frank, I wouldn't know how to prove it.

(1) The fact that weak differentiability implies v-norm Lipschtiz continuity is only established for measures and not for Markov chains. I am not sure whether this actually holds for P_λ. If we cannot prove this, then we still can use the following result. If condition **(C)** holds and in addition for $a^i_j(t)$ in condition **(C)** it holds that

$$\left|\left|\sum_j a^i_j(t)\right|\right| = \left(\sum_j a^i_j(t)\right) v_i < \infty,$$

then P_λ is $||\cdot||_v$-norm Lipschitz continuous.

(2) Should we prove this? In an article in TOMACS we have proved this v-norms where we assumed that v is of polynomial form.

References

[1] J. Artalejo: *Retrial Queueing Systems*, Springer, New York, 2008.

[2] Cinlar, E.: *Introduction to Stochastic Processes*, Prentice-Hall, New-Jersey, 1975.

[3] Z. Feldman, W. Massey, W. Whitt: Staffing of time-varying queues to achieve time-stable performance, *Management Science*, **54**: 324-338, 2008.

[4] G. Falin: A survey of retrial queues, *Queueing Systems*, 7:127–168, 1990.

[5] G. Falin, J. Templeton: *Retrial Queues*, Chapman and Hall, London, 1997.

[6] N. Gans, G. Koole, A. Mandelbaum: Telephone call centers: tutorial, review, and research prospects, *Manufacturing & Service Operations Management*, 5:79–141, 2003.

[7] B. Heidergott, H. Leahu: Weak differentiability of product measures, *Mathematics of Operations Research*, **35**: 27-51, 2010.

[8] V. Kulkarni, H. Liang: Retrial queues revisited. In *Frontiers in Queueing: Models and Applications in Science and Engineering*, J. Dshalalow (ed.), CRC press, Boca Raton, 19–34, 1997.

[9] Lewis, P., G. Shedler: Simulation of nonhomogeneous Poisson processes by thinning, *Naval Research Logistics Quaterly*, **26**: 403-413, 1979.

[10] R. Mamon: A time-varying Markov chain model of term structure, *Statisitics and Probability Letters*, **60**, 309-312, 2002.

[11] W. Massey, W. Whitt: Uniform acceleration expansions for Markov chains with time-varying rates, *Journal of Applied Probability*, **8**, 1130-1155, 1998.

[12] T. Yang, J. Templeton: A Survey on Retrial Queues, *Queueing Systems*, 2:201–233, 1987.

[13] Yin, G.,Q. Zhang: *Continuous Time Markov Chains and Applications.* Springer, New York, 1998.

[14] M. Fu, J. Q. Hu. *Conditional Monte Carlo: Gradient Estimation and Optimization Applications*, Kluwer Academic Publishers, Boston, 1997.

[15] p. Glasserman. *Gradient Estimation via Perturbation Analysis*, Kluwer Academic Publishers, Boston, 1991.

[16] P. Glasserman, W. Gong. Smoothed perturbation analysis for a class of discrete-event systems. *IEEE Transactions on Automatic Control*, **35**, 1218 - 1230, 1990.

[17] W. Gong, Y. Ho. Smoothed (conditional) perturbation analysis of discrete event dynamical systems. *IEEE Transactions on Automatic Control*, **32**, 858 - 866, 1987.

[18] Y. Ho, X. Cao. *Perturbation Analysis of Discrete Event Dynamic Systems.* Kluwer Academic Publisher, Boston, 1991.

[19] R. Rubinstein. Sensitivity analysis of discrete event systems by the score function method. *Annals of Operations Research*, **39**, 229-250, 1992.

[20] R. Rubinstein. *Modern Simulation and Modeling.* Kluwer Academic Publishers, Boston. 1998.

[21] R. Rubinstein, D. Kroese. *Simulation and the Monte Carlo Method.* Second Edition, Wiley, 2007.

[22] M. Fu. Stochastic Gradient Estimation. Chapter 19, in *Handbook on Operations Research and Management Science: Simulation*, (S.G. Henderson and B.L. Nelson, editors), Elsevier, pages 75-616, 2006.