Methodology and Computing in Applied Probability, 6, 109–127, 2004

Entropy Maximization for Markov and Semi-Markov Processes

VALERIE GIRARDIN girardin@math.unicaen.fr
LMNO, UMR6139, Campus II, Université de Caen, BP 5186, 14032 Caen, France

Received August 29, 2002; Revised September 8, 2003; Accepted September 23, 2003

Abstract. The literature about maximum of entropy for Markov processes deals mainly with discrete-time Markov chains. Very few papers dealing with continuous-time jump Markov processes exist and none dealing with semi-Markov processes. It is the aim of this paper to contribute to fill this lack.

We recall the basics concerning entropy for Markov and semi-Markov processes and we study several problems to give an overview of the possible directions of use of maximum entropy in connection with these processes. Numeric illustrations are presented, in particular in application to reliability.

Keywords: maximum of entropy, jump Markov processes, Markov chains, reliability, semi-Markov processes

AMS 2000 Subject Classification: 60K15, 60J25, 62B10, 94A17

1. Introduction

Markovian processes and entropy are linked since the introduction by Shannon (1948) of entropy in probability theory. He defined and studied the entropy (rate) of a discrete-time Markov chain with finite state space, giving the first statement of the ergodic theorem of information theory. The entropy of a finite continuous-time ergodic jump Markov process was later defined by Bad Dumitrescu (1988). Finally, the entropy of a finite ergodic semi-Markov process is defined in Girardin and Limnios (2003), thus extending the ergodic theorem of information to these processes. Semi-Markov processes generalize Markov and renewal processes and hence allow a wide range of applied systems to be modeled (see Howard, 1971; Ravichandran, 1990; Limnios and Oprişan, 2001).

Among a given family of processes, selecting the process with the maximum entropy is equivalent to adding the less of information possible to the considered problem or to choosing the process which can be realized in the most numerous number of ways (see, for example, Cover and Thomas, 1991). Many Markov chains problems have been studied via maximization of entropy but very few problems linked to jump Markov processes and none to semi-Markov processes. The purpose of this paper is to recall the existing results and to study several new problems to give an overview of the possible directions of use of maximum entropy methods in connection with these processes. We determine the process maximizing the entropy among jump Markov processes or semi-Markov processes satisfying the same moment-type constraints. This adds a justification to the use of semi-Markov models of the type

$$Q_{ij}(t) = P(i,j)H_i(t), \quad i \neq j \in E, \quad t \geq 0, \tag{1}$$

where $(P(i,j))$ is the transition matrix of the embedded chain and H_i the distribution of the sojourn time in state i. Most results are obtained by the method of the Lagrange multipliers and concern general Markov or semi-Markov processes. In applications, the form of the generator or of the semi-Markov kernel is generally known, constituting in itself a constraint on the process. We develop several examples to illustrate this, in particular in connection to reliability theory for Markov processes and seismic risk analysis for semi-Markov processes. We also present a maximum of entropy criterion of choice of the randomized process among the family linked to any jump Markov process.

The paper is organized as follows. In Section 2, we recall the necessary definitions of entropy for Markovian processes. In Section 3, we review the literature dealing with maximum of entropy in connection with Markov chains and we prove some new results. In Section 4, we study several maximum of entropy problems for jump Markov processes. We develop several applications to reliability systems, with numeric illustration. Finally, in Section 5, we determine the semi-Markov processes maximizing the entropy under different moment-type constraints. We study especially in Section 5.2 the alternating renewal process and kernels of the type (1) with an application to seismic risk analysis.

2. Entropy of Markovian Processes

The Shannon entropy of a real random variable with density f is defined as

$$\mathbb{S}(f) = -\int_{\mathbb{R}} f(t) \ln f(t) dt,$$

with the convention $0 \ln 0 = 0$.

For a discrete-time process $\mathbf{X} = (X_n)_{n \in \mathbb{N}}$, the entropy at time n is defined as the Shannon entropy of the n-dimensional marginal distribution of $\mathbf{X}$, namely

$$\begin{aligned}\mathbb{H}_n(\mathbf{X}) &= -\mathbb{E}[p_n^{\mathbf{X}}(X)] \\ &= -\int_{\mathbb{R}} p_n^{\mathbf{X}}(x_1, \ldots, x_n) \ln p_n^{\mathbf{X}}(x_1, \ldots, x_n) d\mu_n(x_1, \ldots, x_n),\end{aligned}$$

where $p_n^{\mathbf{X}}$ is the density function of the random vector $X = (X(1), \ldots, X(n))$ with respect to some reference measure μ with marginals μ_n. For discrete state spaces, this measure is the counting measure. For justifications of the choice of $x \ln x$ in the definition of entropy in connection with thermodynamics and reliability theory, see Moran (1961) and Rocchi (2002).

Under suitable conditions, $\mathbb{H}_n(\mathbf{X})/n$ converges to a finite limit $\mathbb{H}(\mathbf{X})$, called the entropy (or entropy rate) of the process and we have $\mathbb{H}(\mathbf{X}) = \inf \mathbb{H}_n(\mathbf{X})/n$.

The entropy rate was first defined by Shannon (1948) for an ergodic Markov chain $\mathbf{J}$

with a finite state space E as a function of the transition matrix $P = (P(i,j))$ and of the stationary distribution ν, namely

$$\mathbb{H}(\mathbf{J}) = -\sum_{i\in E} \nu_i \sum_{j\in E} P(i,j) \ln P(i,j). \tag{2}$$

Shannon proved the convergence of the sequence $(-\ln[p_n^{\mathbf{J}}(J)]/n)$ to $\mathbb{H}(\mathbf{J})$ in probability, and this constitutes the original version of the ergodic theorem of information theory. The convergence in mean for a stationary process was proven by McMillan (1953) and the almost sure convergence by Breiman (1957, 1960). The former is known as Shannon–McMillan theorem or asymptotic equirepartition property and the latter is called Shannon–McMillan–Breiman theorem or ergodic theorem of information theory.

For a continuous-time process $\mathbf{Z} = (Z(t))$ indexed by $\mathbb{R}_+$, the entropy at time T is defined as

$$\mathbb{H}_T(\mathbf{Z}) = -\mathbb{E}[f_T^{\mathbf{Z}}(Z)] = -\int_{\mathbb{R}} f_T^{\mathbf{Z}}(z) \ln f_T^{\mathbf{Z}}(z) d\mu_T(z),$$

where $f_T^{\mathbf{Z}}$ is the likelihood function of $Z = (Z(t))_{0\le t\le T}$ with respect to μ, with restriction μ_T on $[0, T]$. If $\mathbb{H}_T(\mathbf{Z})/T$ converges to a limit $\mathbb{H}(\mathbf{Z})$, this limit is called the entropy (rate) of the process. The extension of Shannon–McMillan theorem for stationary processes was obtained by Perez (1964) under integrability conditions (see also Kieffer, 1974 and the references therein).

The semi-Markov processes are not stationary, thus Perez's result does not apply to them. The entropy rate of an ergodic semi-Markov process $\mathbf{Z}$ with finite state space E and semi-Markov kernel $(Q_i j(t))$ absolutely continuous with respect to the Lebesgue measure, with Radon–Nikodym derivative $(q_{ij}(t))$ was proven in Girardin and Limnios (2003) to be

$$\mathbb{H}(\mathbf{Z}) = \frac{1}{\nu \cdot m} \sum_{i,j\in E} \nu_i \mathbb{S}(q_{ij}),$$

where m_i is the mean sojourn time in state $i \in E$, ν is the stationary distribution of $\mathbf{J}$ and $\nu \cdot m = \sum_{\ell\in E} \nu_\ell m_\ell$. The Shannon–McMillan–Breiman theorem is thus extended to this class of non-stationary continuous-time processes. If we set $q_{ij}(t) = P(i,j) f_{ij}(t)$, where $(P(i,j))$ is the transition matrix of the embedded Markov chain $\mathbf{J}$ of $\mathbf{Z}$, we obtain

$$\mathbb{H}(\mathbf{Z}) = \frac{1}{\nu \cdot m} \left[\mathbb{H}(\mathbf{J}) + \sum_{i,j\in E} \nu_i P(i,j) \mathbb{S}(f_{ij}) \right]. \tag{3}$$

The entropy rate of an ergodic jump Markov process first defined by Bad Dumitrescu (1988) appears as a special case of semi-Markov process entropy. It is a function of its infinitesimal generator $A = (a_{ij})$ and of its stationary distribution (π_i), namely

$$\mathbb{H}(\mathbf{Z}) = \sum_{i\in E} \pi_i \sum_{j\neq i} a_{ij}(1 - \ln a_{ij}). \tag{4}$$

3. Maximum of Entropy for Markov Chains

In the literature, the maximum of entropy has been determined for Markov chains under various constraints. Let us sketch some of the results.

First, the following results induce a particular interest for finite discrete-time Markovian processes in connection with entropy. Among processes of which the first and second moments are known, the maximum entropy is obtained for a Markov chain (see Berger, 1971). Among processes whose distribution of $(X_1, \ldots, X_p)$ is known, the maximum entropy is obtained for a Markov chain of order $p-1$ (see Politis, 1993). The maximum entropy among strictly stationary processes with the same energy (a functional of the second-order marginal distribution) is also obtained for a Markov chain (see Spitzer, 1972). Various problems involving Markov chains can be solved or their solution confirmed using entropy maximization. For example, one of the fundamental issues of Shannon's theory was the noiseless coding theorem. In a Markov chain source model, robust noiseless source encoding amounts to determining the maximum entropic transition matrix in a given convex set of transition matrices (see Kazakos, 1983). Many other examples can be found (see Cover and Thomas, 1991, for examples in the study of capacity of communication channels, etc.).

We state here some general results of maximum entropy selection of a Markov chain among others, for an ergodic Markov chain **J** with finite state space E, transition matrix P and stationary distribution ν. Its entropy is given by (2).

PROPOSITION 1 *Among finite ergodic Markov chains, the maximum entropy is obtained for uniform transition probabilities.*

Proof: Since ν is a stationary distribution for the transition matrix $P = (P(i,j))$, we have $\sum_{i \in E} \nu_i P(i,j) = \nu_j$ for $j \in E$ and $\sum_{i \in E} \nu_i = 1$. Equating to zero the derivative of the Lagrangian

$$-\sum_{i \in E} \nu_i \sum_{j \in E} P(i,j) \ln P(i,j) - \sum_{j \in E} x_j \left[\sum_{i \in E} \nu_i P(i,j) - \nu_j \right] - y \sum_{i \in E} \nu_i$$

with respect to $P(i,j)$ for $i,j \in E$ and ν_i for $i \in E$ yields

$$-\nu_i(1 + \ln P(i,j)) - x_j \nu_i = 0, \quad i,j \in E. \tag{5}$$

$$\sum_{j \in E} P(i,j) \ln P(i,j) - \sum_{j \in E} x_j P(i,j) + x_i - y = 0, \quad i \in E. \tag{6}$$

From (5), we obtain $P(i,j) = e^{-1-x_j}$, which in turn gives in (6) $x_i = x_j$ for all $i,j \in E$ and thus the transition probabilities are uniform and the stationary distribution ν is the uniform one. ■

PROPOSITION 2 (Gerchak, 1981) *Among finite ergodic Markov chains sharing the same stationary distribution ν, the maximum entropy is obtained for the chain with transition probabilities equal to the stationary ones, that is, $P(i,j) = \nu_j$ for all $i,j \in E$.*

We will extend both above results to jump Markov processes in Section 4.2.

Constraints on a finite number of moments or increments of a Markov chain can be seen as weighted constraints of the type

$$\sum_{i,j\in E} \nu_i P(i,j) w_{ij}(m) = W(m), \quad m \in M, \tag{7}$$

for given functions w_{ij} and W, for a finite $M \subset \mathbb{N}$.

PROPOSITION 3 (Hohöldt and Justesen, 1984) *Among finite ergodic Markov chains satisfying constraints (7), the maximum entropy is obtained for the chain with transition probabilities and stationary distribution*

$$P(i,j) = \frac{v_j}{v_i} e^{-\eta} \exp\left[\sum_{m\in M} s(m) w_{ij}(m)\right] \quad \text{and} \quad \nu_i = v_i v_i^*,$$

where $v = (v_i)$ and $v^* = (v_i^*)$ are the right and left eigenvectors for the eigenvalue $e^{-\eta}$ of the matrix

$$M = \left(\exp\left[\sum_{m\in M} s(m) w_{ij}(m)\right]\right)$$

and $(s(m))$ are the Lagrange multipliers determined by the constraints.

The associated entropy is $\mathbb{H}(\mathbf{J}) = -\eta - \sum_{m\in M} s(m) w_{ij}(m)$. Inequalities instead of equalities can be considered and convex programming methods can be used (see Balcău, 1993).

We will use this result for solving moments problems for jump Markov processes in Section 4.5.

4. Maximum of Entropy for Markov Processes

In this section, we consider maximization problems for an ergodic continuous-time jump Markov process $\mathbf{Z} = (Z(t))$, with finite state space E, infinitesimal generator $A = (a_{ij})$ and stationary distribution $\pi = (\pi_i)$.

Precisely, we determine the maximum entropy jump Markov process: without constraints in Section 4.1; with a known stationary distribution in Section 4.2, thus extending the result proven for Markov chains by Gerchak (1981); when the transition matrix of its embedded Markov chain is known in Section 4.3.

We establish a maximum of entropy criterion of selection of the parameter λ of the randomized processes $\{\mathbf{X}_\lambda\}$ linked to any jump Markov process $\mathbf{Z} = (Z(t))$ in Section 4.4. To our knowledge, no other criteria exist in the literature. This family is useful for studying $\mathbf{Z}$. Indeed, if $\mathbf{Z}_\lambda = (Z_\lambda(t))$ denotes the jump Markov process whose embedded Markov chain is $\mathbf{X}_\lambda$, then $Z(t)$ and $Z_\lambda(t)$ have the same probability distribution, but the probabilistic structure of $\mathbf{Z}_\lambda$ is much simpler than the probabilistic structure of $\mathbf{Z}$. The randomized processes are also basic in numerical applications of jump Markov processes, for example as in the method of randomization for solving numerically Kolmogorov's forward and backward equations (see Boucherie and van Doorn, 1998).

We determine the jump Markov process maximizing entropy under general moment constraints in Section 4.5. They include constraints on the moments of the process or of its linked randomized processes.

4.1. Maximum of Entropy Without Additional Constraints

The behavior of any finite ergodic jump Markov process $\mathbf{Z}$ is determined up to A, that is to say up to $(a_{ij})_{i\neq j}$ (and to the initial distribution). Its entropy is given by (4). The following proposition generalizes Proposition 1 to continuous-time jump Markov processes.

PROPOSITION 4 *Among finite ergodic jump Markov processes, the maximum entropy is obtained for the process with uniform generator.*

Proof: Since π is a stationary distribution for A, we have $\sum_{i\in E}\pi_i = 1$ and $\sum_{i\neq j}\pi_i a_{ij} = \pi_j \sum_{k\neq j} a_{jk}$ for $j\in E$. Equating to zero the derivative of the Lagrangian

$$\mathbb{H}(\mathbf{Z}) + y\sum_{i\in E}\pi_i + \sum_{j\in E}x_j\left[\sum_{i\neq j}\pi_i a_{ij} - \pi_j\sum_{k\neq j}a_{jk}\right],$$

with respect to a_{ij} for $i\neq j\in E$ and π_i for $i\in E$ yields

$$-\pi_i \ln a_{ij} + x_j\pi_i - x_i\pi_i = 0, \quad i\neq j. \tag{8}$$

$$\sum_{j\neq i} a_{ij}(1-\ln a_{ij}) + y + \sum_{j\neq i} x_j a_{ij} - x_i\sum_{j\neq i} a_{ij} = 0, \quad i\in E. \tag{9}$$

From (8), we obtain $a_{ij} = e^{x_j}e^{-x_i}$, which in turn gives in (9) $-ye^{x_i} + \sum_{j\neq i} e^{x_j} = 0$, for $i\in E$. This linear system with unknown e^{x_i} has a solution if and only if $y = |E| - 1$, and hence $e^{x_i} = e^{x_j}$ and finally $a_{ij} = 1$ for all $i\neq j$. The stationary distribution π is the uniform one. ∎

The above result concerns general Markov processes. In applications, generally the type of generator is known, for example in reliability modeling.

EXAMPLE 1 We consider several kinds of reliability systems with two identical components. The failure rate of each online component is λ and the repair rate of each component is μ, with perfect repair. Unless otherwise stated, a single repair facility is supposed to be available. The jump Markov process $\mathbf{Z} = (Z(t))$ modeling the system represents the number of failed machines at time $t \geq 0$, with values in $E = \{0, 1, 2\}$. See Howard (1971), Ravichandran (1990) and Girardin and Limnios (2001) for details on these systems.

a. For a cold standby system, the component in standby does not fail. The generator of the Markov process is

$$A = \begin{pmatrix} -\lambda & \lambda & 0 \\ \mu & -(\lambda+\mu) & \lambda \\ 0 & \mu & -\mu \end{pmatrix},$$

with stationary distribution $\pi = (\mu/d, \lambda\mu/d, \lambda^2/d)$, where $d = \lambda^2 + \lambda\mu + \mu^2$. We compute

$$\mathbb{H}(\mathbf{Z}) = \frac{-2\lambda\mu(-2 + \ln\lambda + \ln\mu)}{\mu+\lambda},$$

with maximum 4/3 obtained for $\lambda = \mu = 1$.

b. For a parallel system, the system fails only if both components fail. The generator of the Markov process is

$$A = \begin{pmatrix} -2\lambda & 2\lambda & 0 \\ \mu & -(\lambda+\mu) & \lambda \\ 0 & \mu & -\mu \end{pmatrix},$$

with stationary distribution $\pi = (\mu^2/d, 2\lambda\mu/d, 2\lambda^2/d)$, where $d = 2\lambda^2 + 2\lambda\mu + \mu^2$. We compute

$$\mathbb{H}(\mathbf{Z}) = \frac{2\lambda\mu(\mu(2 - \log 2) - (\lambda+\mu)(\ln\lambda\ \ln\mu) + 2\lambda}{(\mu+\lambda)^2 + \lambda^2}.$$

Numerically, the maximum can be proven to be 1.3728 and to be obtained for $\lambda \approx 0.6843$ and $\mu \approx 0.9728$.

c. For a two-machine-two-repairman system, the generator is

$$A = \begin{pmatrix} -2\lambda & 2\lambda & 0 \\ \mu & -(\lambda+\mu) & \lambda \\ 0 & 2\mu & -2\mu \end{pmatrix},$$

with stationary distribution $\pi = (\mu^2/d, 2\lambda\mu/d, \lambda^2/d)$, where $d = (\lambda+\mu)^2$. We compute

$$\mathbb{H}(\mathbf{Z}) = \frac{-2\lambda\mu(-2 + \ln\lambda + \ln 2 + \ln\mu)}{\mu+\lambda},$$

with maximum $\sqrt{2}$ obtained for $\lambda = \mu = 1/\sqrt{2}$.

4.2. Maximum of Entropy and Stationary Distribution

In many applications modeled by an ergodic Markov process, only the stationary behavior of the process is observed and thus it is interesting to suppose the stationary distribution to be known. The following proposition generalizes Proposition 2 to continuous-time Markov processes.

PROPOSITION 5 *Among finite ergodic jump Markov processes sharing the same stationary distribution π, the maximum entropy is obtained for the process with generator given by*

$$a_{ij}^2 = \pi_j/\pi_i, \quad i \neq j.$$

Proof: Let us suppose that the stationary distribution π is known. Equating to zero the derivative of the Lagrangian $\mathbb{H}(\mathbf{Z}) + \sum_{j\in E} x_j[\sum_{i\neq j} \pi_i a_{ij} - \pi_j \sum_{k\neq j} a_{jk}]$ with respect to a_{ij} for $i \neq j \in E$ yields (8). We compute

$$\frac{\partial^2 \mathbb{H}(\mathbf{Z})}{\partial a_{ij}^2} = \frac{-\pi_i}{a_{ij}} < 0, \quad i \neq j,$$

and hence $\mathbb{H}(\mathbf{Z})$ is a strictly concave function of $(a_{ij})_{i\neq j}$ with a unique maximum. Taking $a_{ij}^2 = \pi_j/\pi_i$ in (8) yields a solution $x_j = 0.5 \ln \pi_j$. Finally we have $a_{ii} = -\sum_{k\neq i} \sqrt{\pi_k}/\sqrt{\pi_i}$ for $i \in E$. ■

EXAMPLE 2 Let us consider again some reliability systems.

a. For the cold standby system, if π is known, then $\lambda = (w - \pi_0)\mu/2\pi_0$, where $v = \sqrt{\pi_0(4 - 3\pi_0)}$. The maximum entropy is obtained for $\mu = \sqrt{2\pi_0/(v - \pi_0)}$ and, necessarily, $\pi_1 = (v - \pi_0)/2$ and $\pi_2 = (v - \pi_0)^2/4\pi_0$.
b. For the parallel system, if π is known, then $\lambda = (v - \pi_0)\mu/2\pi_0$, where $v = \sqrt{\pi_0(2 - \pi_0)}$. The maximum entropy is obtained for

$$\mu = \exp\left[\frac{(v - \pi_0)\log 2 - (\pi_0 + v)\log[(v - \pi_0)/\pi_0]}{2(v + \pi_0)}\right],$$

and, necessarily, $\pi_1 = v - \pi_0$ and $\pi_2 = (v - \pi_0)^2/\pi_0$.
c. For the two-machine-two-repairman system, if π is known, then $\lambda = (1 - 1/\sqrt{\pi_0})\mu$. The maximum entropy is obtained for $\mu = (1 - 1/\sqrt{\pi_0})^{-1/2}/\sqrt{2}$ and, necessarily, $\pi_1 = 2(1 - \sqrt{\pi_0})\pi_0$ and $\pi_2 = (1 - \sqrt{\pi_0})^2$.
d. For a warm standby system, the component in standby can fail, with failure rate λ_s. The generator of the Markov process is

$$A = \begin{pmatrix} -\lambda - \lambda_s & \lambda + \lambda_s & 0 \\ \mu & -(\lambda + \mu) & \lambda \\ 0 & \mu & -\mu \end{pmatrix},$$

with stationary distribution $\pi = (\mu^2/d, (\lambda + \lambda_s)\mu/d, \lambda(\lambda + \lambda_s)/d)$, where $d = \lambda^2 + \lambda\lambda_s + \mu(\lambda + \lambda_s) + \mu^2$. We compute

$$\mathbb{H}(\mathbf{Z}) = -\frac{(\lambda + \lambda_s)\mu)[(-2 + \log\mu)(\lambda + \mu) + \lambda\log\lambda + \mu\log(\lambda + \lambda_s)]}{\mu^2 + \lambda^2 + \lambda\mu + \lambda_s(\lambda + \mu)}.$$

If π is known, then $\lambda = \pi_1\mu/\pi_0$ and $\lambda_s = (\pi_0\pi_2 - \pi_1^2)\mu/\pi_0\pi_1$. The maximum entropy is obtained for

$$\mu = \exp\left[\frac{(\pi_0^3 - \pi_0^2 + \pi_0^2\pi_1)\log\pi_2 + (\pi_0^2 - \pi_0^3 - \pi_0^2\pi_1 - \pi_1^3)\log\pi_1 - \pi_1^3\log\pi_0}{\pi_0^2 - \pi_0^3 - 2\pi_0^2\pi_1 + \pi_1^3 + \pi_0\pi_1}\right].$$

4.3. Maximum of Entropy and Embedded Markov Chain

An alternative to the knowledge of the limit behavior of a jump Markov process is the full knowledge of its embedded Markov chain or at least of the stationary distribution of its embedded Markov chain.

The embedded Markov chain **J** of a jump Markov process **Z** is defined by $J_n = Z(T_n)$, where T_n denote the time of the n-th jump of the process (see Girardin and Limnios, 2001). Its transition matrix $P = (P(i,j))$ and its stationary distribution $\nu = (\nu_i)$ are given by

$$P(i,i) = 0, \quad P(i,j) = \frac{a_{ij}}{a_i}, \quad i \neq j, \qquad P(i,i) = 0, \quad i \in E, \tag{10}$$

$$\nu_i = \pi_i \frac{a_i}{a \cdot \pi}, \quad i \in E. \tag{11}$$

PROPOSITION 6 *Among finite ergodic jump Markov processes with embedded Markov chains sharing the same stationary distribution ν such that $\nu_i < 1/2$ for all $i \in E$, the maximum entropy is obtained for the process with generator given by*

$$a_{ij} = \frac{(1 - 2\nu_j)^{-1}}{\sum_{k \neq i} (1 - 2\nu_k)^{-1}}, \quad i \neq j,$$

up to some multiplicative constant.

Proof: The constraints can be written as

$$\sum_{i \in E} \pi_i = 1, \qquad \sum_{j \neq i} a_{ij} \pi_i = \pi_j \sum_{k \neq j} a_{jk}, \quad j \in E,$$
$$\pi_i \sum_{l \neq i} a_{il} = \nu_i \sum_{j \in E} \pi_j \sum_{l \neq j} a_{jl}, \quad i \in E.$$

Equating to zero the derivative of the Lagrangian with respect to a_{ij} for $i \neq j \in E$ and π_i for $i \in E$ yields

$$-\pi_i \ln a_{ij} + x_j \pi_i - x_i \pi_i + y_i \pi_i - y_i \nu_i \pi_i = 0, \quad i \neq j. \tag{12}$$

$$\sum_{j \neq i} a_{ij}(1 - \ln a_{ij}) + z + \sum_{j \neq i} x_j a_{ij} - x_i \sum_{l \neq i} a_{il} + y_i \sum_{l \neq i} a_{il} - y_i \nu_i \sum_{l \neq i} a_{il} = 0, \quad i \in E. \tag{13}$$

From (12), we obtain $a_{ij} = e^{x_j} e^{-x_i + y_i(1 - \nu_i)}$, which in turn gives in (13)

$$z + e^{-x_i + y_i(\nu_i - 1)} \sum_{j \neq i} e^{x_j} = 0, \quad i \in E,$$

and hence $a_{ij} = -z e^{x_j} / \sum_{k \neq i} e^{x_k}$ for all $i \neq j$.

Using the constraints, we obtain $\pi = \nu$ and

$$\sum_{i \neq j} \nu_i \frac{e^{x_j}}{\sum_{k \neq i} e^{x_k}} = \nu_j, \quad j \in E,$$

which in turn gives the result. ∎

EXAMPLE 3 Let us consider again some reliability systems.

a. For the cold standby system, we have

$$\nu_0 = \frac{\mu}{2(\lambda+\mu)}, \quad \nu_1 = \frac{1}{2}, \quad \nu_2 = \frac{\lambda}{2(\lambda+\mu)}. \tag{14}$$

If ν is known, then $\lambda = \mu(1-2\nu_0)/2\nu_0$ and the maximum entropy is obtained for $\mu = \sqrt{2\nu_0/(1-2\nu_0)}$.

b. For the warm standby system, ν is also given by (14). If ν is known, then $\lambda = \mu(1-2\nu_0)/2\nu_0$ and the maximum entropy is obtained for

$$\lambda_s = \frac{(-4\nu_0^2 + 6\nu_0 - 1)\mu}{2\nu_0}, \quad \mu = \left(\frac{1}{2\nu_0} - 1\right)^{\nu_0 - 1/2} 2^{-\nu_0}(1-\nu_0)^{-\nu_0}.$$

PROPOSITION 7 *Among the finite ergodic jump Markov processes sharing the same embedded Markov chain, the maximum entropy is obtained for the process with generator given by*

$$a_{ij} = e^{\mathbb{H}(\mathbf{J})} P(i,j), \quad i \in E.$$

Proof: Once P is known, since $P(i,j) = a_{ij}/a_i$ for $i \neq j$, $\mathbf{Z}$ is determined up to (a_i) and π or equivalently up to (a_i) and $a.\pi$. The entropy of $\mathbf{Z}$ can be written as

$$\mathbb{H}(\mathbf{Z}) = a.\pi \sum_{i \in E} \nu_i \sum_{j \neq i} P(i,j)(1 - \ln P(i,j)a_i)$$

and the constraint is

$$a.\pi \sum_{i \in E} \frac{\nu_i}{a_i} = 1. \tag{15}$$

Equating to zero the partial derivatives of the Lagrangian $\mathbb{H}(\mathbf{Z}) - y(a.\pi \sum_{i \in E} \nu_i/a_i)$ with respect to $a.\pi$ and a_i for $i \in E$ yields

$$\sum_{i \in E} \nu_i \sum_{j \neq i} P(i,j)(1 - \ln P(i,j)a_i) - y \sum_{i \in E} \frac{\nu_i}{a_i} = 0, \tag{16}$$

$$a.\pi\, \nu_i \left(\frac{1}{a_i} - \frac{y}{a_i^2}\right) = 0, \quad i \in E. \tag{17}$$

(17) implies $a_i = y$ for $i \in E$, and then (16) yields $y = \exp \mathbb{H}(\mathbf{J})$. Finally, (15) gives $a_i = a.\pi = \exp \mathbb{H}(\mathbf{J})$ and (11) gives $\nu = \pi$. The associated entropy is $\mathbb{H}(\mathbf{Z}) = \exp[\mathbb{H}(\mathbf{J})]$. ∎

Note that the Markov chain $\mathbf{J}$ can be the maximum entropic chain determined under moment-type constraints such as studied in Section 3.

EXAMPLE 4 Let us consider a more complicated cold standby system with two identical components. If the online component fails, the second is switched online successfully with probability $p \in [0, 1]$. The failure rate of the online component is λ and the repair rate is μ, with two independent repair facilities for the two components. The generator of the Markov process is

$$A = \begin{pmatrix} -\lambda & p\lambda & (1-p)\lambda \\ \mu & -(\lambda+\mu) & \lambda \\ 0 & 2\mu & -2\mu \end{pmatrix}$$

with stationary distribution $\pi = (2\mu^2/d, 2\lambda\mu/d, \lambda(\lambda+\mu-p\mu)/d)$, where $d = 3\lambda\mu - p\mu\lambda + 2\mu^2 + \lambda^2$. Its entropy is

$$\begin{aligned}\mathbb{H}(\mathbf{Z}) = \frac{1}{d}[&-2\lambda\mu^2 p \log p + \lambda - 2\lambda\mu^2 p) \log \lambda + ((1-p) \\ &\log 2 + p - 3)\mu + \mu \times (1-p)\log(1-p) + \mu(1-p)\log \lambda + \\ &(2\mu + \lambda - p\mu)\log \mu + (\log 2 - 2)\lambda].\end{aligned}$$

The transition matrix of the embedded Markov chain is

$$P = \begin{pmatrix} 0 & p & (1-p) \\ \mu/(\lambda+\mu) & 0 & \lambda/(\lambda+\mu) \\ 0 & 1 & 0 \end{pmatrix}$$

with stationary distribution $\nu = (\mu/d, (\mu(1-p)+\lambda)/d, (\lambda+\mu)/d)$, where $d = (3-p)\mu + 3\lambda$.

If P is known, then p is known and $\lambda = \mu/[1 - P(1,2)]$. The maximum entropy is obtained for

$$\mu = \exp \left[\frac{p(P(1,2)-1)\log p - P(1,2)\log(1-P(1,2)) + (1-p)(P(1,2)-1)(\log(1-p)+\log 2)}{pP(1,2) - 3P(1,2) - p + 5}\right].$$

Note that if ν is known (and not P), then $\lambda = (\nu_1 - \nu_0)\mu/\nu_0$ and $p = (\nu_1 - \nu_2)/\nu_0$. If $\nu_0 \geq (3-\sqrt{5})/4$, the maximum entropy is obtained for

$$\mu = \frac{(\nu_1 - \nu_2)^{\nu_1+\nu_2}\nu_0^{\nu_0+\nu_1}(\nu_1 - \nu_0 - \nu_2)^{\nu_1 - \nu_0 - \nu_2}}{(\nu_1 - \nu_0)^{\nu_1 2\nu_2}}.$$

4.4. Maximum Entropic Randomized Process

The randomized Markov processes $(\mathbf{X}_\lambda)$ of an ergodic jump Markov process $\mathbf{Z}$ are ergodic Markov chains. The transition matrix $R = (r_{ij})$ of $\mathbf{X}_\lambda$ is $r_{ij} = \mathbf{1}_{(i=j)} + a_{ij}/\lambda$ for $i, j \in E$, and its stationary distribution is that of $\mathbf{Z}$. The parameter λ is supposed to belong to $[\max\{a_i : i \in E\}, +\infty[$, where $a_i = -a_{ii} = \sum_{j \neq i} a_{ij}$. We develop here an entropy criterion for choosing λ.

PROPOSITION 8 *Let $(\mathbf{X}_\lambda)$ be the family of randomized processes associated with a given finite ergodic jump Markov process. We have*

$$\mathbb{H}(\mathbf{X}_{\lambda_0}) = \max_{\lambda \geq \max\{a_i : i \in E\}} \mathbb{H}(\mathbf{X}_\lambda),$$

if and only if

$$\prod_{i\in E}(\lambda_0 - a_i)^{\pi_i a_i} = \exp\left[\sum_{i\in E}\pi_i\sum_{j\neq i}a_{ij}\ln a_{ij}\right]. \tag{18}$$

Proof: The entropy rate of the Markov chain $\mathbf{X}_\lambda$ (given by (2)) is

$$\mathbb{H}(\mathbf{X}_\lambda) = -\frac{1}{\lambda}\sum_{i\in E}\pi_i\sum_{j\neq i}a_{ij}\ \ln\ a_{ij} + \ln\ \lambda - \sum_{i\in E}\pi_i\left(1-\frac{a_i}{\lambda}\right)\ln(\lambda - a_i).$$

Therefore,

$$\frac{\partial}{\partial\lambda}\mathbb{H}(\mathbf{X}_\lambda) = \frac{\sum_{i\in E}\pi_i\sum_{j\neq i}a_{ij}\ \ln\ a_{ij}}{\lambda^2} - \frac{1}{\lambda^2}\sum_{i\in E}\pi_i a_i\ \ln(\lambda - a_i),$$

and (18) follows. Moreover,

$$\frac{\partial^2}{\partial\lambda^2}\mathbb{H}(\mathbf{X}_\lambda) = -\frac{2}{\lambda}\frac{\partial}{\partial\lambda}\mathbb{H}(\mathbf{X}_\lambda) - \frac{1}{\lambda^2}\sum_{i\in E}\frac{\pi_i a_i}{\lambda - a_i},$$

and hence $\partial^2/\partial\lambda^2\mathbb{H}(\mathbf{X}_{\lambda_0}) < 0$. Since λ_0 is the unique root of $\partial/\partial\lambda\mathbb{H}(\mathbf{X}_\lambda) = 0$, the result follows. ■

A detailed study of (18) with numerical illustration can be found in Merzéréaud (2002). Its root can always be computed numerically. In some cases, it takes a simple explicit analytic form.

EXAMPLE 5 Let us consider a reliability system with s machines and no repair until all the machines fail. The Markov process modeling this system is an s-state pure birth Markov process. Its generator is

$$A = \begin{pmatrix} -a_1 & a_1 & 0 & \cdots & \cdots & 0 \\ 0 & -a_2 & a_2 & \ddots & \cdots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & \ddots & -a_{s-1} & a_{s-1} \\ a_s & 0 & \cdots & \cdots & 0 & -a_s \end{pmatrix}$$

with stationary distribution $\pi = \left(\Pi_{k\neq i}a_i/\sum_{i\in E}\Pi_{k\neq i}a_i\right)_{i\in E}$. Hence (18) can be written as

$$\prod_{i\in E}(\lambda - a_i) = \prod_{i\in E}a_i.$$

If $s = 2n$, $a_i = \alpha$ for $i \leq n$ and $a_i = \beta$ for $i \geq n+1$, then $\lambda_0 = \alpha + \beta$. Note that this includes all binary Markov processes, with generators

$$A = \begin{pmatrix} -\alpha & \alpha \\ \beta & -\beta \end{pmatrix}.$$

If $s = 3$, for $a_1 = a_2 = 1$ and $a_3 = \alpha$, we compute using Maple

$$\lambda_0 = \frac{1}{6}u^{1/3} + \frac{2}{3}(\alpha - 1)^2 u^{-1/3} + \frac{1}{3}\alpha + \frac{2}{3},$$

where

$$u = -8 + 132\alpha - 24\alpha^2 + 8\alpha^3 + 12(-12\alpha + 117\alpha^2 - 36\alpha^3 + 12\alpha^4)^{1/2}.$$

For example, for $\alpha = 1/\sqrt{2}$, we compute $\lambda \approx 1.8032$, for $\alpha = \sqrt{2}$, we compute $\lambda_0 \approx 2.2789$, and, for $\alpha = 2$, we compute $\lambda_0 \approx 2.5956$.

4.5. *Maximum of Entropy and Moment Constraints*

Let $\mathbf{Z}$ be a finite ergodic jump Markov process with generator (a_{ij}) and let $(\mathbf{X}_\lambda)$ denote its family of randomized Markov processes, with transition probabilities (r_{ij}) and stationary distribution π. We consider here constraints of the form

$$\sum_{i,j \in E} \pi_i r_{ij} w_{ij}(m) = W(m), \quad m \in M, \tag{19}$$

for a finite $M \subset \mathbb{N}$. They include constraints on the moments of $\mathbf{X}_\lambda$ or $\mathbf{Z}$, such as $\sum_{i \in E} i\pi_i = m_1$, $\sum_{i \in E} i^2 \pi_i = m_2$, or on the variance of the increments of $\mathbf{X}_\lambda$, as

$$\sum_{i \in E} \pi_i \sum_{j \in E} r_{ij}(i - j)^2 = \sigma^2.$$

Inequalities instead of equalities can also be considered, for example for bounding the heights of the jumps by limiting the increments of $\mathbf{X}_\lambda$.

PROPOSITION 9 *Among the finite ergodic jump Markov processes satisfying Constraints (19), the maximum entropy is obtained for the process with generator given by*

$$a_{ij} = r_{ij}\lambda, \quad i \neq j \quad \textit{and} \quad a_i = (1 - r_{ii})\lambda, \tag{20}$$

where

$$\lambda = \exp - \left[\sum_{i \in E} \pi_i \frac{\sum_{j \neq i} r_{ij} \ln r_{ij}}{\sum_{i \in E} \pi_i (1 - r_{ii})} \right], \tag{21}$$

and

$$r_{ij} = \frac{v_j}{v_i} e^{-\eta} \exp\left[\sum_{k \in E} s(k) w_{ij}(k) \right], \tag{22}$$

with the notation of Proposition 3.

Proof: Under Constraints (19), Proposition 3 induces that the maximum entropy $\mathbb{H}(\mathbf{X}_\lambda)$ is obtained for the transition probabilities given by (22). The generator of the associated Markov process $\mathbf{Z}$ given by (20) is determined up to λ. We compute

$$\mathbb{H}(\mathbf{Z}) = -\sum_{i\in E}\pi_i\sum_{j\neq i} r_{ij}\lambda(1 - \ln r_{ij}\lambda),$$
$$\frac{\partial\mathbb{H}(\mathbf{Z})}{\partial\lambda} = -\sum_{i\in E}\pi_i\sum_{j\neq i} r_{ij} \ln r_{ij}\lambda,$$
$$\frac{\partial^2\mathbb{H}(\mathbf{Z})}{\partial\lambda^2} = -\frac{1}{\lambda}\sum_{i\in E}\pi_i(1 - r_{ii}).$$

Hence $\mathbb{H}(\mathbf{Z})$ is strictly concave and its derivative equals zero for λ given by (21). Together with (20) and (22), it yields the result. ■

5. Maximum of Entropy for Semi-Markov Processes

Let us consider now some maximization problems for an ergodic semi-Markov process $\mathbf{Z}$ with finite state space E, embedded Markov chain $\mathbf{J}$ with transition matrix P and stationary distribution ν. Firstly, we determine the maximum entropic semi-Markov process under different constraints on the sojourn times. Secondly, we develop some particular cases. We consider processes for which the distribution of the sojourn time in i if the next visited state is j is uniform in i or in j, with application to seismic risk analysis. Finally, we study the alternating renewal process case.

5.1. General Results

If the mean sojourn times are not known, the maximum of entropy is infinite. Indeed, the entropy of a semi-Markov kernel of type (1), with exponential distributions H_i with parameter $1/m_i$, is

$$\mathbb{H}(\mathbf{Z}) = \frac{1}{\nu\cdot m}\left[\mathbb{H}(\mathbf{J}) - \sum_{i,j\in E}\nu_i P(i,j)(\log m_i + 1)\right],$$

which tends to infinity when m_i tends to zero for some $i\in E$.

If the mean sojourn times are known and if the stationary distribution too is known, the maximum entropic process is the same as if the entropy of each distribution f_{ij} was maximized under mean constraint m_i and if the entropy of the embedded Markov chain was maximized separately under stationary distribution constraint, as proven by the following result.

PROPOSITION 10 *Among finite ergodic semi-Markov processes with known mean sojourn times and with embedded Markov chains sharing the same stationary distribution, the maximum entropy is obtained for a semi-Markov kernel with exponential distributions with parameter $1/m_i$ and transition matrix of the embedded Markov chain equal to the stationary one.*

Proof: The constraints are

$$\int_{\mathbb{R}+} f_{ij}(t)dt = 1, \quad i,j \in E \tag{23}$$

$$\sum_{j \in E} P(i,j) \int_{\mathbb{R}+} tf_{ij}(t)dt = m_i, \quad i \in E, \tag{24}$$

$$\sum_{i \in E} \nu_i P(i,j) = \nu_j, \quad j \in E, \qquad \sum_{j \in E} P(i,j) = 1, \quad i \in E. \tag{25}$$

If we write the entropy as in (2), the Lagrangian is

$$\frac{-1}{\nu \cdot m} \sum_{i \in E} \nu_i \left[\sum_{j \in E} P(i,j) \ln(P(i,j)) + \sum_{j \in E} P(i,j) \int_{\mathbb{R}+} f_{ij}(t) \ln f_{ij}(t) \right] + x_{ij} \int_{\mathbb{R}+} f_{ij}(t)dt$$
$$+ \sum_{i \in E} y_i \left[\sum_{j \in E} P(i,j) \int_{\mathbb{R}+} tf_{ij}(t)dt \right] + \sum_{j \in E} z_j \sum_{i \in E} \nu_i P(i,j) + \sum_{i \in E} v_i \sum_{j \in E} P(i,j).$$

Using Euler–Lagrange equations of calculus of variations, we obtain

$$\frac{-1}{\nu \cdot m} \nu_i P(i,j)[1 + \ln f_{ij}(t)] + x_{ij} + y_i P(i,j)t = 0, \quad i,j \in E, \tag{26}$$

$$\frac{-1}{\nu \cdot m} \left[\nu_i(\ln P(i,j) + 1) + \nu_i \int_{\mathbb{R}+} f_{ij}(t) \ln f_{ij}(t)dt \right]$$
$$+ y_i \left[\int_{\mathbb{R}+} tf_{ij}(t)dt \right] + z_j \nu_i + v_i = 0, \quad i,j \in E, \tag{27}$$

Setting $x'_{ij} = x_{ij}/P(i,j)$ in (26) yields $f_{ij}(t) = \exp[\nu \cdot m x'_{ij}/\nu_i + \nu \cdot m y_i t/\nu_i - 1]$. We deduce from (23) that f_{ij} depends only on i and thus from (24) that $x_{ij} = \nu_i(1 - \ln m_i)/\nu \cdot m$ and $y_i = -\nu_i/\nu \cdot m m_i$, which in turn yield $f_{ij}(t) = \exp[-t/m_i]/m_i$. Substituting in (27), we obtain $P(i,j) = \exp[-1 + \nu \cdot m(z_j + v_i/\nu_i)]/m_i$, and, finally, (25) implies $P(i,j) = \nu_j$ for $i,j \in E$. ■

The embedded Markov chain can be supposed to be known entirely, and the constraints on the sojourn times can be more general, for example as follows. The proof is similar to the first half of the proof of Proposition 10.

PROPOSITION 11 *Among finite ergodic semi-Markov processes sharing the same embedded Markov chain and under moment-type constraints including the mean sojourn times, the maximum entropy is obtained for a process with semi-Markov kernel of type (1).*

To be specific, if the constraints are

$$\sum_{j\in E} P(i,j) \int t f_{ij}(t)dt = m_i, \quad \textit{and} \quad \sum_{j\in E} P(i,j) \int \phi(t) f_{ij}(t)dt = \Phi_i, \quad i\in E,$$

for some given function ϕ, *then* f_{ij} *is given by* $f_{ij}(t) = \exp[x_i + y_i t + z_i \phi(t)]$.

For example, the maximum is obtained for the following type of distribution (the parameters of which are to be determined through the constraints):

- under geometric constraints, that is $\phi(t) = \ln t$, for a Gamma distribution (see Kapur, 1989);
- under geometric and harmonic constraints, that is $\phi_1(t) = \ln t$, and $\phi_2(t) = 1/t$, for an inverse Gaussian distribution (see Mohtashami Borzadaran, 2001);
- under variance constraints, that is $\phi(t) = t^2$, for a truncated normal distribution (see Kapur, 1989).

5.2. Examples and Applications

In applications, the type of semi-Markov kernel is often known *a priori*, as shown in the following examples.

EXAMPLE 6 In seismik risk analysis (see Grandori Guagenti and Molina, 1986), a time predictable model corresponds to a semi-Markov kernel of the type

$$q_{ij}(t) = P_j h_i(t) = \nu_j h_i(t), \quad i, j\in E$$

and a slip predictable model corresponds to a semi-Markov kernel of the type

$$q_{ij}(t) = P_j h_j(t) = \nu_j h_j(t), \quad i, j\in E.$$

For both above kernels, we have

$$\mathbb{H}(\mathbf{Z}) = \frac{1}{\nu \cdot m}\left[\mathbb{S}(\nu) + \sum_{i\in E} \nu_i \mathbb{S}(h_i)\right].$$

If the mean sojourn times m_i are known for $i\in E$, the maximum of entropy is thus obtained by maximizing the entropy of each h_i under the constraints considered, independently of the embedded Markov chain. The stationary distribution ν giving the maximum entropy can then be determined as a function of m, at least numerically.

If no constraints are added to the mean sojourn times, h_i is an exponential distribution with parameter $1/m_i$, which adds a justification to the model proposed by Grandori Guagenti and Molina (1986) in seismik analysis. The entropy of h_i is $\mathbb{S}(h_i) = -1 - \log m_i$ and the entropy of the process is

$$\mathbb{H}(\mathbf{Z}) = \frac{-1}{\nu \cdot m}\left[1 + \sum_{i \in E} \nu_i \log(\nu_i m i)\right].$$

For instance, for a three-state process, numerically, if $m_0 = 1$, $m_1 = 2$ and $m_2 = 3$, the maximum can be proven to be -0.2306 and to be obtained for $\nu_0 \approx 0.4633$, $\nu_2 \approx 0.2917$ and $\nu_1 \approx 0.2449$.

If a slip and time predictable model is considered, corresponding to a general semi-Markov process, Proposition 11 induces that the maximum under constraints on the sojourn times is obtained for distributions h_i which are again independent of the next visited state.

The maximum entropic process can also be determined for a given type of sojourn times distributions.

EXAMPLE 7 For a semi-Markov kernel (1), if the H_i are log-normal distributions with densities

$$h_i(t) = \frac{1}{\sigma_i t\sqrt{2\pi}} \exp \frac{-(\ln t - M_i)^2}{2\sigma_i^2}, \quad t \geq 0,$$

then

$$\begin{aligned}
\mathbb{H}(\mathbf{Z}) &= \frac{1}{\sum_{\ell \in E} \nu_\ell \exp(M_\ell + \sigma_\ell^2/2)} \\
&\quad \times \left[-\sum_{i,j \in E} (\nu_i P(i,j) \ln P(i,j)) + \sum_{i \in E} \nu_i (M_i + 1/2 + \ln \sigma_i \sqrt{2\pi})\right] \\
&= \frac{1}{A}[\mathbb{H}(\mathbf{J}) + B].
\end{aligned}$$

Let us suppose that the embedded Markov chain of the process is known. Equating to zero the partial derivatives of $\mathbb{H}(\mathbf{Z})$ with respect to M_i and σ_i, for $i \in E$, yields

$$A = [\mathbb{H}(\mathbf{J}) + B] \exp\left(\frac{M_i + \sigma_i^2}{2}\right) \quad \text{and} \quad A = \sigma_i^2 [\mathbb{H}(\mathbf{J}) + B] \exp\left(\frac{M_i + \sigma_i^2}{2}\right),$$

from which we derive $\sigma_i = 1$ and $M_i = -\mathbb{H}(\mathbf{J}) - (1 + \ln 2\pi)/2$ for $i \in E$.

Note that considering other distributions leads to similar results. The associated calculus can be more complex, depending on the form of the associated entropy. Explicit formulas for the entropy of several distributions are listed in Kapur (1989).

We will now determine the maximum entropic alternating renewal process.

EXAMPLE 8 Let $(X_n)_{n \geq 0}$ and $(Y_n)_{n \geq 0}$ be two i.i.d. independent sequences of random variables with respective distribution functions F and G with densities f and g with respect to the Lebesgue measure and finite mean values a and b. Set $S_0 = 0$ and $S_n = S_{n-1} + X_n + Y_n$ for $n \geq 1$. Let $\mathbf{J}$ be a Markov chain with state space $E = \{1, 2\}$, transition matrix

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

and stationary distribution $\nu = (1/2, 1/2)$.

Let $\mathbf{Z}$ be the semi-Markov process associated with the alternating renewal process $(J_n, S_n)_{n \geq 0}$, defined by $Z_t = J_{N(t)}$ for $t \geq 0$, where $N(t) = \sup\{n \geq 0 : S_n \leq t\}$. The state space of $\mathbf{Z}$ is E, its embedded Markov chain is $\mathbf{J}$ and its semi-Markov kernel is

$$Q(t) = \begin{pmatrix} 0 & F(t) \\ G(t) & 0 \end{pmatrix}.$$

We have $H_1(t) = F(t)$, $H_2(t) = G(t)$, $m_1 = a$ and $m_2 = b$. The entropy of $\mathbf{Z}$ is

$$\mathbb{H}(\mathbf{Z}) = \frac{1}{a+b}[\mathbb{S}(f) + \mathbb{S}(g)].$$

Maximization of entropy under constraints including the means of F and G thus amounts to separate maximizations of the Shannon entropies $\mathbb{S}(f)$ and $\mathbb{S}(g)$.

For example, the maximum entropic alternating renewal process is obtained under arithmetic constraints for a Poisson alternating process. The functions F and G are exponential distribution functions.

References

M. Bad Dumitrescu, ''Some informational properties of Markov pure jump processes,'' *Cas. Pestovani Mat.* vol. 113 pp. 429–434, 1988.

C. Balcău, ''L'optimisation de l'entropie des chaînes de Markov,'' *Bull. Math. Soc. Math. Roumanie* vol. 37 pp. 21–41, 1993.

T. Berger, *Rate Distortion Theory*, Prentice Hall: Englewood Cliffs, 1971.

R. J. Boucherie and E. A. van Doorn, ''Uniformization for l-Positive Markov Chains,'' *Commun. in Stat.: Stoch Mod.* vol. 14 pp. 171–186, 1998.

L. Breiman, ''The individual ergodic theorem of information theory,'' *Ann. Math. Stat.* vol. 28 pp. 809–811, 1957.

L. Breiman, ''Correction to: the individual ergodic theorem of information theory,'' *Ann. Math. Stat.* vol. 31 pp. 809–810, 1960.

L. Cover and J. Thomas, *Elements of Information Theory*, Wiley series in telecommunications: New-York, 1991.

Y. Gerchak, ''Maximal entropy of Markov chains with common steady-states probabilities,'' *J. Oper. Res. Soc.* vol. 32 pp. 233–234, 1981.

V. Girardin and N. Limnios, *Probabilités en vue des Applications*, Vuibert: Paris, 2001.

V. Girardin and N. Limnios, ''On the entropy for semi-Markov processes,'' *J. Applied Probab* vol. 40 pp. 1060–1068, 2003.

E. Grandori Guagenti and C. Molina, ''Semi-Markov processes in seismic risk analysis,'' In Janssen (ed.), *Semi-Markov Models: Theory and Applications*, Plenum: New York, pp. 487–506, 1986.

T. Hohöldt and J. Justesen, ''Maxentropic Markov chains,'' *IEEE Trans. Inf. Theory* vol. 30 pp. 665–667, 1984.

R. A. Howard, *Dynamic Probabilistic Systems*, vol. I: Markov Models. vol. II: Semi-Markov and Decision Processes, Series in Decision and Control, John Wiley & Sons: New York, 1971.

J. N. Kapur, *Maximum Entropy Models in Sciences and Engineering*, John Wiley: New York, 1989.

D. Kazakos, ''Robust noiseless source coding through a game theoretic approach,'' *IEEE Trans. Inf. Theory* vol. 29 pp. 576–583, 1983.

J. Kieffer, ''A simple proof of the Moy-Perez generalization of the Shannon-McMillan theorem,'' *Pacific J. Math.* vol. 51 pp. 203–206, 1974.

N. Limnios and G. Oprişan, *Semi-Markov Processes and Reliability*, Birkhauser: Boston, 2001.

M. McMillan, ''The basic theorems of information theory,'' *Ann. Math. Stat.* vol. 24 pp. 196–219, 1953.

M. Merzéréaud, *Entropie et processus de Markov : Un critère de choix du processus randomisé* T. E. R., Université de Caen, 2002.

G. R. Mohtashami Borzadaran ''Characterization of Pearsonian and bilateral power series distribution via maximum entropies,'' In A. Mohammad-Djafari (ed.), *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, AIPCP, pp. 145–150, 2001.

P. Moran, ''Entropy, Markov processes and Boltzmann's H-theorem,'' *Proc. Camb. Philos. Soc.* vol. 57 pp. 833–842, 1961.

A. Perez, ''Extensions of Shannon-McMillan's limit theorem to more general stochastic processes,'' *Inf. Theory, Stat. Decision Functions, Random Processes* pp. 545–574, 1964.

D. Politis, ''Nonparametric maximum entropy,'' *IEEE Trans. Inf. Theory* vol. 39 pp. 1409–1413, 1993.

N. Ravichandran, *Stochastic Methods in Reliability Theory*, John Wiley: New York, 1990.

P. Rocchi, ''Boltzmann-like entropy in reliability theory,'' *Entropy* vol. 4 pp. 142–150, 2002.

C. Shannon, ''A mathematical theory of communication,'' *Bell Syst., Techn. J.* vol. 27 pp. 379–423, 623–656, 1948.

F. Spitzer, ''A variational characterization of finite Markov chains,'' *Ann. Math. Stat.* vol. 43 pp. 303–307, 1972.