Stochastic Processes and their Applications 35 (1990) 309-313
North-Holland

LAWS OF LARGE NUMBERS FOR THE ANNEALING ALGORITHM

Nina GANTERT

Institut für Angewandte Mathematik, Universität Bonn, 5300 Bonn 1, FR Germany

Received 29 May 1989
Revised 7 September 1989

In this paper we establish a weak and a strong law of large numbers for the algorithm of simulated annealing. To this end, we recall laws of large numbers for time-inhomogeneous Markov chains which are formulated in terms of Dobrushin's contraction coefficients, and we show how they lead to corresponding cooling constants in the annealing algorithm.

simulated annealing * time-inhomogeneous Markov chains * law of large numbers

1. Introduction

Let $(E, \mathscr{E})$ be a measurable state space. For two probability measures μ and ν on E, let

$$\|\mu - \nu\| := \sup_{A \subseteq E} |\mu(A) - \nu(A)|$$

denote the variational distance, and let

$$c(Q) := \sup_{x,y \in E} \|Q(x, \cdot) - Q(y, \cdot)\|$$

denote the contraction coefficient of a transition kernel $Q(x, \mathrm{d}y)$ on E. Now consider a time-inhomogeneous Markov chain given by a sequence of transition kernels

$$P_n(x, \mathrm{d}y) \quad (n = 1, 2, \ldots) \tag{1.1}$$

on $(E, \mathscr{E})$. Suppose that π_n is an invariant probability measure for P_n, and that

$$\sum_n \|\pi_{n+1} - \pi_n\| < \infty; \tag{1.2}$$

in particular, (1.2) implies that π_n converges to some probability measure π_∞. If, in addition,

$$\prod_{n \geq m} c(P_n) = 0 \tag{1.3}$$

for any m, then the sequence of distributions at time n,

$$\mu_n := \mu P_1 \cdots P_n \quad (n = 1, 2, \ldots) \tag{1.4}$$

converges to π_∞ for any initial distribution μ (see, e.g., Iosifescu and Theodorescu, 1969, p. 160). As a special case, one obtains a basic convergence result of Geman and Geman (1984) for the annealing algorithm. In this special context, condition (1.2) is satisfied automatically, and condition (1.3) corresponds to a condition of the form

$$\gamma \leqslant \gamma_0 \tag{1.5}$$

for the constant γ in the cooling scheme $\beta_n = \gamma \log n$ of the annealing algorithm; cf. Section 3 below.

We are interested in the law of large numbers, i.e., in the convergence

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f(X_i) = \int f \, \mathrm{d}\pi_\infty \tag{1.6}$$

for bounded measurable functions f on E, either in probability (*weak law*) or P_μ-almost surely (*strong law*), where P_μ denotes the distribution on $\Omega = E^{\{0,1,\ldots\}}$ of the Markov chain with initial distribution μ and transition kernels P_n $(n = 1, 2, \ldots)$, and X_i is the ith coordinate map on Ω. Gidas (1985) has addressed this question, but the proof of his theorem 1.3 is wrong; a counterexample is given in Section 2 below. Instead, we use known versions of the law of large numbers for time-inhomogeneous Markov chains where the assumptions are formulated in terms of the coefficients $c(P_n)$. For the annealing algorithm, this leads to the condition

$$\gamma < \gamma_0 \tag{1.7}$$

for the weak law, and to the condition

$$\gamma < \tfrac{1}{2}\gamma_0 \tag{1.8}$$

for the strong law of large numbers.

2. Some asymptotic properties of time-inhomogeneous Markov chains

Let us first show that conditions (1.2) and (1.3) are not sufficient to guarantee a law of large numbers; this shows in particular that Theorem 1.3 in Gidas (1985) is incorrect. The following example is due to H.-R. Künsch (oral communication).

Example. Let $E = \{0, 1\}$, and consider the time-inhomogeneous Markov chain given by

$$P_n(0, 1) = P_n(1, 0) = 1/n \quad (n = 1, 2, \ldots).$$

Then $\pi_n = \pi_\infty = (\frac{1}{2}, \frac{1}{2})$. Conditions (1.2) and (1.3) are satisfied; in particular, $\mu P_1 \cdots P_n$ converges to π_∞ for any initial distribution μ. But for $f(x) := x$,

$$\frac{1}{n} \sum_{i=1}^{n} f(X_i) \tag{2.1}$$

cannot converge in probability. In fact, for any $\alpha = 2, 3, \ldots$, and any μ we have

$$P_\mu[X_n = \cdots = X_{\alpha n}] = \prod_{j=1}^{\alpha n - 1} \left(1 - \frac{1}{j}\right) \geq \left(1 - \frac{1}{n}\right)^{n(\alpha - 1)} \to e^{-\alpha + 1} > 0. \tag{2.2}$$

Now assume that the sequence in (2.1) converges in probability to some random variable Z. It is easy to see that, due to (2.2), the values of Z can only be 0 and 1, P_μ-almost surely. On the other hand, condition (1.3) implies that P_μ satisfies a zero-one-law on the tail field $\mathscr{F} := \bigcap_m \sigma(X_m, X_{m+1}, \ldots)$ (cf. Ueno, 1957). In particular, Z is constant P_μ-almost surely, hence $P_\mu[Z = E_\mu[Z] = \frac{1}{2}] = 1$, and so we get a contradiction.

Let us now recall versions of the law of large numbers which are suitable for our purpose. For the given sequence (1.1) of transition kernels we set

$$c_n := \max_{1 \leq i \leq n} c(P_i).$$

Then the condition

$$\lim_{n \to \infty} n(1 - c_n) = \infty \tag{2.3}$$

is sufficient for $\mathscr{L}^2(P_\mu)$-convergence in (1.6); this follows from estimate (1.2.22) in Iosifescu and Theodorescu (1969, p. 53). In particular, (2.3) implies the *weak* law of large numbers. The *strong* law of large numbers holds if we have

$$\sum_{n=1}^{\infty} \frac{1}{2^n(1 - c_{2^n})} < \infty; \tag{2.4}$$

this is an immediate consequence of Theorem 1.2.23 in Iosifescu and Theodorescu (1969, p. 56).

3. Application to the annealing algorithm

Suppose that the state space is of the form $E = S^I$ where I and S are finite. What we are looking for is the set of global minima of some *energy function U* on the *configuration space E*. Simulated annealing corresponds to a time-inhomogeneous Markov chain on E which approaches this set of global minima.

For $\beta > 0$ let π_β denote the Gibbs measure on S^I given by

$$\pi_\beta(x) := e^{-\beta U(x)} / Z(\beta)$$

where $Z(\beta) := \sum_{y \in E} e^{-\beta U(y)}$. It is easy to verify that $\|\pi_\beta - \pi_\infty\| \to 0$ for $\beta \to \infty$, where π_∞ denotes the uniform distribution on the set of global minima. Moreover, condition (1.2) is satisfied for any sequence $\beta(n) \to \infty$. We fix $i \in I$ and $s \in S$ and denote by $x_{i,s} \in S^I$ the configuration which equals x off i and takes the value s in i. Define the conditional probability

$$\pi_i^\beta(s \mid x) := e^{-\beta U(x_{i,s})} / Z_i(\beta)$$

with normalizing factor $Z_i(\beta)$, and introduce the transition kernel

$$\Pi_i^\beta(x,\cdot)=\pi_i^\beta(s\,|\,x)\otimes\prod_{j\neq i}\delta_{x(j)}.$$

where $\delta_{x(j)}$ denotes Dirac measure on $x(j)$. For a fixed enumeration $I=\{1,\ldots,N\}$,

$$P_\beta:=\Pi_1^\beta\cdots\Pi_N^\beta$$

defines a transition kernel on S^I with unique invariant distribution π_β. Let us denote by $\Delta_i:=\max\{|U(x)-U(y)|;\ x,y\in S^I \text{ and } x=y \text{ off } i\}$ the oscillation of U in the ith coordinate, and set $\Delta:=\max_i\Delta_i$. Then it is easy to show that

$$c(P_\beta)\leq 1-\mathrm{e}^{-\beta N\Delta}. \tag{3.1}$$

Let us now specify the following version of the annealing algorithm. We fix a cooling scheme of the form

$$\beta(n)=\gamma\log n$$

and consider the time-inhomogeneous Markov chain with transition kernels $P_{\beta(n)}$ ($n=1,2,\ldots$). Using (3.1), it is easy to check that condition (1.3) is satisfied as soon as

$$\gamma\leq\gamma_0:=1/(N\delta). \tag{3.2}$$

Thus, we have the basic convergence result of Geman and Geman (1984) that (3.2) implies

$$\lim_{n\to\infty}\|\mu P_{\beta(1)}\cdots P_{\beta(n)}-\pi_\infty\|=0$$

for any initial distribution μ.

Let us now apply the laws of large numbers of Section 2. In our present case, $c(P_{\beta(n)})$ is increasing in n, hence $c_n=c(P_{\beta(n)})$, and (2.4) is equivalent to

$$\sum_{n=1}^{\infty}\frac{1}{n^2(1-c_n)^2}<\infty.$$

This leads to the following.

Theorem. *If*

$$\beta(n)=\gamma\log n\quad \textit{with } \gamma<\gamma_0, \tag{3.3}$$

then the weak law of large numbers holds. The strong law of large numbers holds if

$$\gamma<\tfrac{1}{2}\gamma_0. \tag{3.4}$$

Proof. We only have to check that (3.3) implies (2.3) and (3.4) implies (2.4). □

The same technique can be used for other annealing schemes, e.g. for the *Metropolis dynamics*, where the transition kernels on an arbitrary finite state space E are defined in the following way. Let Q be an irreducible symmetric matrix on E and define

$$P_\beta(x, y) = \begin{cases} Q(x, y)\, \mathrm{e}^{-\beta(U(y)-U(x))}, & \text{if } U(x) \leqslant U(y),\ x \neq y, \\ Q(x, y), & \text{if } U(x) > U(y), \\ 1 - \sum_{z \neq x} P_\beta(x, z), & \text{if } x = y. \end{cases}$$

Possibly $c(P_{\beta(n)}) = 1$ for all n; then we have to estimate $c(P_{\beta(n)} \cdots P_{\beta(n+M)})$ instead of $c(P_{\beta(n)})$, where $M := \min\{n \geqslant 1;\ Q^n > 0\}$. If we define

$$\Delta := \max\{|U(x) - U(y)|;\ x, y \in E \text{ and } Q(x, y) > 0\}$$

and $\gamma_0 := 1/(M\Delta)$, we can get the same theorem.

References

R.L. Dobrushin, Central limit theorems for nonstationary Markov chains, Theory Probab. Appl. 1(4) (1956) 329–383.

S. Geman and D. Geman, Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, IEEE Trans. Pattern Anal. Machine Intelligence 6 (1984) 721–741.

B. Gidas, Nonstationary Markov chains and convergence of the annealing algorithm, J. Statist. Phys. 39 (1985) 73–131.

M. Iosifescu and R. Theodorescu, Random Processes and Learning (Springer, Berlin, 1969).

D.L. Isaacson and R.W. Madsen, Markov Chains: Theory and Application (Wiley, New York, 1976).

T. Ueno, Some limit theorems for temporally discrete Markov processes, J. Fac. Sci. Univ. Tokyo Sect. I 7 (1957) 449–462.