Stochastics and Stochastics Reports, Vol. 29, pp. 221-226
Reprints available directly from the publisher
Photocopying permitted by license only

OPTIMAL ACCEPTANCE PROBABILITY FOR SIMULATED ANNEALING

GEORGE KESIDIS and EUGENE WONG

University of California at Berkeley, Berkeley, CA 94720, USA

(Received 25 May 1989; in final form 5 July 1989)

The objective of this note is two-fold: first, to prescribe a rather general form for the acceptance probability which will attain the Gibbs distribution for a stationary Markov chain; second, to find the particular one that will maximize the rate at which equilibrium is reached.

KEY WORDS: Gibbs field, Markov chain, Boltzmann machine, simulated annealing.

1. INTRODUCTION

Let S be a finite set and let μ be a probability measure on S such that $\mu(x)>0$ for every x. We define a *Gibbs field* as a Markov chain $\{X_t, t=0,1,2,\ldots\}$ with values in S and with its equilibrium distribution given by the Gibbs distribution

$$P_0(x)=\frac{1}{Z(T)}e^{-(1/T)U(x)}\mu(x) \tag{1.1}$$

where we interpret T as the temperature, $U(x)$ as an energy function, and

$$Z(T)=\sum_x e^{-(1/T)U(x)}\mu(x) \tag{1.2}$$

is known as the *partition function*.

To get (1.1), one can choose the transition probability

$$P(x'|x)=\text{Prob}(X_{t+1}=x'|X_t=x) \tag{1.3}$$

with considerable latitude. The objective of this note is two-fold: first, to prescribe a rather general family of $P(x'|x)$ that will result in (1.1); second, to find the one that optimizes (in a suitable sense) the rate at which the equilibrium distribution (1.1) is reached.

This problem is of considerable interest in simulated annealing (Holley and Stroock [3], Aarts and Jorst [1]) which is a stochastic algorithm for finding the minimum of the function $U(x)$. There, a Gibbs field in equilibrium is "cooled" by letting T be replaced by a function $T(t)$ that decreases to 0 as $t \to \infty$. The cooling must be sufficiently slow so that "quasi-equilibrium" is maintained. Intuitively, this means that the more quickly equilibrium is restored as temperature changes, the more quickly annealing can be achieved.

2. A GENERAL FORM FOR THE ACCEPTANCE PROBABILITY

The transition probability $P(x'|x)$ for a Gibbs field can be chosen as follows: Given $X_t = x$, we first choose a point $y \in S$ according to a probability $q(x, y)$, and set:

$$\begin{aligned} x_{t+1} &= y \qquad \text{with probability } p(\Delta U) \\ &= x \qquad \text{with probability } 1 - p(\Delta U) \end{aligned}$$

where $\Delta U = U(y) - U(x)$. We call $p(\Delta U)$ the *acceptance probability*. It follows that

$$P(y|x) = q(x, y)p(U(y) - U(x)) \qquad y \neq x$$

and

$$P(x|x) = 1 - \sum_{x' \neq x} q(x, x')p(U(x') - U(x)). \tag{2.1}$$

The conventional choice for the acceptance probability is

$$p(\Delta U) = \min(1, e^{-(1/T)\Delta U}) \tag{2.2}$$

and an alternative choice is

$$p(\Delta U) = (1 + e^{(1/T)\Delta U})^{-1}. \tag{2.3}$$

The first corresponds to the so-called heat-bath model and the second to the Metropolis algorithm (Metropolis *et al.* [4]). If q is μ symmetric and irreducible then either of these choices will yield the desired Gibbs distribution (1.1) for the stationary distribution. Our first result is to give a general form for the acceptance probability to achieve (1.1), a form that includes both (2.2) and (2.3) as special cases.

THEOREM 1 *Let q be μ-symmetric, i.e.,*

$$\mu(x)q(x, y) = \mu(y)q(y, x) \tag{2.4}$$

and irreducible, i.e.,

$$\sum_{n=0}^{\infty} q^{(n)}(x, y) = \infty \qquad \text{for all} \quad x, y \in S$$

where

$$q^{(n+1)}(x, y) = \sum_{x' \in X} q(x, x') q^{(n)}(x', y).$$

Let the transition probability be determined by (2.1). Then, the equilibrium distribution is given by (1.1) if the acceptance probability has the form

$$p(\Delta U) = e^{-(1/2T)\Delta U} g(|\Delta U|), \qquad \Delta U \in (-\infty, \infty). \tag{2.5}$$

Proof With $P_0(x)$ given by (1.1) and with the use of (2.5) in (2.1), we get

$$P(y|x)P_0(x) = \mu(x)q(x, y)g(|\Delta U|)1/Z\, e^{-(1/2T)[U(x)+U(y)]}.$$

It then follows from (2.4) that

$$P(y|x)P_0(x) = P(x|y)P_0(y)$$

and

$$\sum_x P(y|x)P_0(x) = P_0(y)$$

so that P_0 is a stationary distribution. Irreducibility then ensures that P_0 is also the equilibrium distribution, i.e. (Feller [2])

$$P_n(y|x) = \text{Prob}(X_{t+n} = y | X_t = x) \xrightarrow[n\to\infty]{} P_0(y). \qquad \text{Q.E.D.}$$

We note that the two choices of acceptance probability given by (2.2) and (2.3) can be put into the form (2.5) as follows:

$$p(v) = \min(1, e^{-(1/T)\Delta U}) = e^{-(1/2T)\Delta U} e^{-(1/2T)|\Delta U|} \tag{2.6}$$

$$p(v) = (1 + e^{(1/T)\Delta U})^{-1} = e^{-(1/2T)\Delta U} \left(2 \cosh \frac{|\Delta U|}{2T}\right)^{-1}. \tag{2.7}$$

3. CONVERGENCE TO EQUILIBRIUM

The rate of convergence to equilibrium can be estimated by considering the following eigenvalue problem:

$$\sum_x P(y|x)P_0(x)\theta(x) = \lambda P_0(y)\theta(y), \qquad y \in S. \tag{3.1}$$

With (2.4) and (2.5), we hve

$$P(y|x)P_0(x) = P(x|y)P_0(y)$$

and there exist orthonormal eigenfunction $\theta_v(x)$, $v = 0, 1, 2, \ldots, N$, such that

$$P_n(y|x) = P_0(y) \sum_{v=0}^{N} \lambda_v^n \theta_v(y)\theta_v(x).$$

The eigenvalues can be ordered so that

$$\lambda_0 = 1 > \lambda_1 \geqq \lambda_2 \geqq \cdots$$

and $\theta_0(x) \equiv 1$. Since

$$|P_n(y|x) - P_0(y)| = P_0(y)\left|\sum_{v=1}^{N} \lambda_v^n \theta_v(y)\theta_v(x)\right|$$

$$= \lambda_1^n \left|P_0(y) \sum_{v=1}^{N} \left(\frac{\lambda_v}{\lambda_1}\right)^n \theta_v(y)\theta_v(x)\right| \leqq K\lambda_1^n$$

the second largest eigenvalue λ_1 provides an estimate of the rate of convergence. Our objective is to find the particular choice of $g(\cdot)$ in (2.5) that would minimize λ_1.

THEOREM 2 *Under the condition of Theorem 1, the eigenvalue λ_1 is minimized by the choice*

$$g_0(v) = e^{-(1/2T)v}. \tag{3.2}$$

Proof We can write

$$\lambda_1 = \max_\theta \left[\sum_{x,y} P(y|x)P_0(x)\theta(x)\theta(y)\right] \tag{3.3}$$

subject to

$$\sum_x P_0(x)\theta(x) = 0 \tag{3.4a}$$

and

$$\sum_x P_0(x)\theta^2(x) = 1. \tag{3.4b}$$

With (2.1) and (2.4), we can rewrite (3.3) as

$$\lambda_1 = \max_\theta \left\{ \sum_y \sum_{x \neq y} P(y|x) P_0(x) \theta(x) \theta(y) + \sum_y P(y|y) P_0(y) \theta^2(y) \right\}$$

$$= 1 + \max_\theta \left\{ \sum_y \sum_x \alpha(x, y) g(|U(x) - U(y)|)[\theta(y)\theta(x) - \theta^2(y)] \right\} \quad (3.5)$$

where

$$\alpha(x, y) = \mu(x) q(x \cdot y)\, e^{-(1/2T)[U(x) + U(y)]}$$

and we have used (3.4b) and the symmetry of $\alpha(\cdot)$.

Since the sum in (3.5) is symmetric in x and y, we must have

$$\lambda_1 = 1 + \max_\theta \left\{ \frac{1}{2} \sum_{x,y} \alpha(x, y) g(|U(x) - U(y)|) \cdot [2\theta(y)\theta(x) - \theta^2(x) - \theta^2(y)] \right\}$$

$$= 1 - \frac{1}{2} \min_\theta \sum_{x,y} \alpha(x, y) g(|U(x) - U(y)|)[\theta(x) - \theta(y)]^2 \quad (3.6)$$

subject to the constraints in (3.4). Now, the function $g(\cdot)$ is related to the acceptance probability through (2.5). It follows that

$$p(u) = e^{-(1/2T)u} g(|u|) \leqq 1, \qquad u \in (-\infty, \infty).$$

Hence,

$$g(|u|) \leqq e^{(1/2T)u}, \qquad -\infty < u < \infty.$$

Taking u to be negative yields

$$g(|u|) \leqq e^{-(1/2T)|u|} = g_0(|u|). \quad (3.7)$$

Now, (3.6) can be written in the form

$$\lambda_1 = 1 - \min_\theta F(g, \theta)$$

where $g \geqq g'$ implies $F(g, \theta) \geqq F(g', \theta)$ for all θ. It follows that for every g

$$\min_\theta F(g_0, \theta) = F(g_0, \theta') \geqq F(g, \theta') \geqq \min_\theta F(g, \theta).$$

Hence,

$$\lambda_1 \geqq 1 - \min_{\theta} F(g_0, \theta)$$

and λ_1 is minimized by g_0. Q.E.D.

4. CONCLUSION

In this note we have established two results: (a) a general form of acceptance probability for constructing Gibbs fields, and (b) an optimality property for a particular choice of the acceptance probability, viz., that of the heat-bath model.

The results of this note can be generalized to continuous time Markov chains where (2.5) would apply to the acceptance *rate*, and under a boundedness condition on the rate (3.2) would yield an optimal acceptance rate.

References

[1] E. H. L. Aarts and J. H. M. Jorst, *Simulated Annealing and Boltzmann Machines...*, Wiley and Sons, New York, 1989.

[2] W. Feller, *An Introduction to Probability Theory and its Applications*, Wiley and Sons, New York, 1968, Vol. 1, 3rd ed.

[3] R. Holley and D. Stroock, Simulated annealing via Sobolev inequalities, *Communication in Math. Phys.* 115 (1988), 553–569.

[4] N. Metropolis, A. Rosenbluth, M. Rosenbluth and E. Teller, Equation of state calculation for fast computing machines, *J. Chem. Phy.* 6 (1953), 1087–1091.