Strategies for Using the Back Button when Browsing the Web

by

Andrey Feuerverger[1] and Jeffrey S. Rosenthal[1]

(April 22, 2002.)

Abstract. Following the model of Fagin et al. (2001) for browsing the World Wide Web, we consider Markov chains with the option of moving "back" to the previous state. For finite state spaces we show that, with judicious use of the back button, essentially any limiting distribution can be achieved from essentially any Markov chain. We provided an algorithm, BACKMET, for accomplishing this.

1. Introduction.

Fagin et al. (2001) introduce the idea of a random walk with a "back" button, where the walk sometimes retreats to its previous state instead of moving to the next state. This modified walk (or "backoff process") is intended to model a user browsing the World Wide Web, where the web pages are the states of the walk, and where the user sometimes presses the "back button".

Fagin et al. (2001) show that if the probability of going back is less than 1/2 and is independent of the walk's history, then the limiting distribution of the backoff process is the same as that of the original walk. However, if the probability of going back is more than 1/2, or depends on the walk's history (e.g. through the current state), then the situation is more complicated. Fagin et al. (2001, Lemma 4.39) do indicate how to compute the limiting distribution in such cases, however the resulting expressions are very complicated (requiring solving cubic equations with coefficients like 3936 and 3180, even for a very simple example on a state space of size 2).

In this paper, we take a different approach, and ask the following question. Suppose we are given fixed random walk (or, more generally, Markov chain) transition probabilities, on a state space $\mathcal{X}$. Suppose we are also given a probability function $f(x)$ on $\mathcal{X}$, which represents our *desired* distribution of what fraction of our time should be spent where. At each step,

[1]Department of Statistics, University of Toronto, Toronto, Ontario, Canada M5S 3G3. Supported in part by NSERC of Canada. E-mail: `andrey@utstat.toronto.edu` and `jeff@math.toronto.edu`.

we may choose to (i) do nothing, (ii) go "back" one step, or (iii) go forward one step of the Markov chain. [Note that in (iii), the resulting jump is chosen randomly according to the underlying Markov chain, and we have no control over precisely where it jumps.] The question is, can we choose which of (i), (ii), and (iii) to perform at each step, in such a way that our long-term limiting distribution is precisely $f(x)$?

From the point of view of browsing the Web, this question may be interpreted as follows. Suppose $f(j)$ represents your subjective opinion (or some objective measure based on web site quality) of what fraction of your time you would like to spend at the Web site j. Then our question becomes, what back button scheme can you use when browsing the Web, so that the fraction of time you ultimately spend at web page j will indeed be equal to $f(j)$?

To address this, we shall introduce the BACKMET algorithm, which accomplishes our goal provided a certain non-negativity condition holds. Our algorithm is related to the *Metropolis-Hastings algorithm*, as we shall explain. The algorithm always chooses option (iii) at least once between every two choices of (ii), so that it never goes back more than one step before going forwards again; this makes the algorithm transient on history stacks (Proposition 1) in the sense of Fagin et al. (2001).

We shall prove (Theorem 7) that on a finite state space, essentially any limiting distribution can indeed be achieved with our BACKMET algorithm.

Acknowledgements. We are very grateful to Allan Borodin and Jon Kleinberg for introducing us to the work of Fagin et al., and for helpful advice.

2. Formal Definitions.

We assume there are fixed Markov chain transition probabilities $p(x, y)$ on a finite state space $\mathcal{X}$. We also assume that there is a fixed probability distribution $f(x)$ on $\mathcal{X}$, which represents the desired limiting distribution.

Following Fagin et al. (2001), we make formal definitions as follows. The backoff process keeps track of an entire *history stack* $H_n = \langle s_{n,0}, s_{n,1}, s_{n,2}, \ldots, s_{n,\ell(H_n)} \rangle$, of length $\ell(H_n) + 1$, at each time n. Here $X_n \equiv s_{n,\ell(H_n)}$ represents the current state of the chain, while $s_{n,0}, s_{n,1}, \ldots, s_{n,\ell(H_n)-1}$ represent previously-visited states that the process can go "back" to.

Initially, $H_0 = \langle s_{0,0} \rangle$ and $\ell(H_0) = 0$, where $s_{0,0}$ is an initial state. Then, at times $n = 0, 1, 2, \ldots$, given the current history stack H_n, the user may choose one of three options for selecting the next history stack H_{n+1}:

(i) Do nothing. That is, $H_{n+1} = H_n$ and $\ell(H_{n+1}) = \ell(H_n)$. [Alternatively, one could specify that if $H_n = \langle s_{n,1}, s_{n,2}, \ldots, s_{n,\ell(H_n)} \rangle$, then $H_{n+1} = \langle s_{n,1}, s_{n,2}, \ldots, s_{n,\ell(H_n)}, s_{n,\ell(H_n)} \rangle$, and $\ell(H_{n+1}) = \ell(H_n) + 1$, i.e. that the top of the stack gets repeated; this will not affect any of our results.]

(ii) Go "back" one step in the history stack. That is, if $H_n = \langle s_{n,1}, s_{n,2}, \ldots, s_{n,\ell(H_n)-1}, s_{n,\ell(H_n)} \rangle$, then $H_{n+1} = \langle s_{n,1}, s_{n,2}, \ldots, s_{n,\ell(H_n)-1} \rangle$, and $\ell(H_{n+1}) = \ell(H_n) - 1$. [If $\ell(H_n) = 0$, then this option is not permitted.]

(iii) Go forward one step of the Markov chain. That is, if $H_n = \langle s_{n,1}, s_{n,2}, \ldots, s_{n,\ell(H_n)-1}, s_{n,\ell(H_n)} \rangle$, then $H_{n+1} = \langle s_{n,1}, s_{n,2}, \ldots, s_{n,\ell(H_n)-1}, s_{n,\ell(H_n)}, Z \rangle$, and $\ell(H_{n+1}) = \ell(H_n) + 1$, where Z is chosen randomly, independent of all previous random variables, such that $\mathbf{P}(Z \in A) = P(s_{n,\ell(H_n)},\ A)$, i.e. with Z having the probability distribution $P(s_{n,\ell(H_n)}, \cdot)$ given by the Markov chain transition probabilities.

The question we ask is, by a judicious choice of which of (i), (ii), or (iii) to perform at each step, can the user ensure that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} P(X_i = a) \;=\; f(a)\,, \qquad a \in \mathcal{X}\,,$$

i.e. that the limiting distribution is given by $f(x)$. Recall that here $X_i = s_{i,\ell(H_i)}$ is the top of the history stack at time i, i.e. is the position of the random walk itself at time i.

3. The BACKMET Algorithm.

Let $p(x, y)$ be the transition probabilities for a Markov chain, and let $f(y)$ be the target probability distribution as above. Let $k \geq 2$ be an integer (to be a parameter in our algorithm), and let t_k be some probability function on $\mathcal{X}$ (so $t_k(x) \geq 0$ for $x \in \mathcal{X}$, with $\sum_{\mathcal{X}} t_k(x) = 1$) to be determined later (see Section 5). In terms of these quantities, we define

$$\alpha_k(x, y) \;=\; \min\left[1,\ \frac{t_k(y)\, p(y, x)}{t_k(x)\, p(x, y)}\right]\,, \tag{1}$$

with the convention that $\alpha_k(x, y) = 1$ if $t_k(x)\, p(x, y) = 0$.

We now define the BACKMET algorithm. Given the history H_n at time n, it proceeds as follows, for $n = 0, 1, 2, \ldots$.

(a) If $n \equiv 0 \pmod{k}$ (i.e. n is an integer multiple of k), then do a forward step of the Markov chain as in (iii) above.

(b) If $n \equiv 1 \pmod{k}$, then with probability $\alpha_k(X_n, X_{n-1})$ do nothing as in (i) above, otherwise with probability $1 - \alpha_k(X_n, X_{n-1})$ go back as in (ii) above.

(c) If $n \not\equiv 0 \pmod{k}$ and $n \not\equiv 1 \pmod{k}$, then do nothing as in (i) above.

In words, the BACKMET algorithm proceeds in blocks of k steps, as follows. For the first step it does one forward step of the Markov chain; for the second step it either does nothing or goes back one step; and for the following $k-2$ steps it does nothing. It then repeats this pattern during the next k steps. We shall justify this unusual pattern in the following sections.

Before leaving the construction of the BACKMET algorithm, we note the following.

Proposition 1. *For the BACKMET algorithm, suppose* $\min_{x,y\in\mathcal{X}} \alpha_k(x,y) > 0$. *Then* $\lim_{n\to\infty} \ell(H_n) = \infty$ *with probability 1, i.e. the process on history stacks is transient.*

Proof. Let $\delta = \min_{x,y\in\mathcal{X}} \alpha_k(x,y) > 0$. Then by construction, $\ell(H_{n+k}) \ge \ell(H_n)$, with $\mathbf{P}\Big(\ell(H_{n+k}) = \ell(H_n)+1\Big) \ge \delta$. It follows that the sequence $\{\ell(H_{(i+1)k}) - \ell(H_{ik})\}_{i=0}^{\infty}$ is i.i.d., with positive mean. The result then follows from the Law of Large Numbers. ∎

In the context of Fagin et al. (2001), Proposition 1 says that the BACKMET algorithm avoids the complication of returning to short history stacks infinitely often. In particular, assuming irreducibility, this shows (Fagin et al., 2001, Theorem 4.3) that the limiting distribution does not depend on the starting state $s_{0,0}$. Of course, this observation does not specify what the limiting distribution is equal to; and it is that question which shall occupy us in the sequel.

4. Relation to the Metropolis-Hastings Algorithm.

To understand the BACKMET algorithm, we recall the *Metropolis-Hastings algorithm* (Metropolis et al., 1953; Hastings, 1970). This algorithm requires a fixed "target" density f. It then proceeds for $n = 0, 1, 2, \ldots$ as follows. Given X_n, it chooses a proposal state Y_{n+1} so that $\mathbf{P}[Y_{n+1} = y] = p(X_n, y)$. It then *accepts* this proposal and sets $X_{n+1} = Y_{n+1}$ with probability $\min\left[1,\ \frac{f(X_n)\ p(X_n, X_{n-1})}{f(X_{n-1})\ p(X_{n-1}, X_n)}\right]$. Otherwise, it *rejects* this proposal and simply sets $X_{n+1} = X_n$.

It is well-known (and straightforward to see) that the Metropolis-Hastings algorithm is *reversible* with respect to the distribution $f(x)$, i.e. that

$$P(X_{n+1} = b \,|\, X_n = a)\, f(a) = P(X_{n+1} = a \,|\, X_n = b)\, f(b)\,, \qquad a, b \in \mathcal{X}\,.$$

It then follows from the general theory of Markov chains (e.g. Meyn and Tweedie, 1993; Tierney, 1994) that, assuming irreducibility, we will have

$$\lim_{n\to\infty} \frac{1}{n} \#\{1 \le i \le n;\ X_i = a\} \;=\; f(a)\,, \qquad a \in \mathcal{X}\,.$$

That is, the limiting distribution of the X_i will be given by $f(x)$. We remark that the Metropolis algorithm has been chosen as one of the top ten computer algorithms of all time (Dongarra and Sullivan, 2000).

The connection of the BACKMET algorithm to the Metropolis-Hastings algorithm is given by the following. (To avoid the problem of certain states being unobtainable, we shall assume the Markov chain $\{p(x, y)\}$ is *irreducible*, meaning that every state can eventually be reached from every other state with positive probability.)

Proposition 2. *Consider the BACKMET algorithm, with $\alpha_k(x, y)$ as in (1). Let $W_n = X_{kn}$ (so that the process $\{W_n\}$ samples the process $\{X_i\}$ once every k^{th} step). Then $\{W_n\}$ follows a Metropolis-Hastings algorithm with target density function t_k. Hence, assuming irreducibility, we have $\lim_{n\to\infty} \frac{1}{n}\#\{1 \le i \le n;\ W_i = a\} = t_k(a)$.*

Proof. By construction of the BACKMET algorithm, we will have $W_{n+1} = X_{kn+1}$ with probability $\alpha_k(X_{kn+1}, W_n)$, while $W_{n+1} = W_n$ with probability $1 - \alpha_k(X_{kn+1}, W_n)$. It then follows from (1) that the chain $\{W_n\}$ does indeed follow a Metropolis-Hastings algorithm with target density t_k. The convergence results then follow from general Markov chain theory (e.g. Meyn and Tweedie, 1993; Tierney, 1994). ∎

Now, under the BACKMET algorithm, we see that $X_{kn+i} = W_{n+1}$ as above for $i = 2, 3, \ldots, k$, i.e. for $k - 1$ out of every k steps. Furthermore, X_{kn+k+1} is constructed by going forward from W_n by one step of the Markov chain. Hence, from the above Proposition, we immediately obtain the following.

Proposition 3. *Consider the BACKMET algorithm. Assuming irreducibility, we have*

$$\lim_{n\to\infty} \frac{1}{n}\#\{1 \le i \le n;\ X_i = a\} \;=\; \frac{k-1}{k} t_k(a) + \frac{1}{k} \sum_{y \in \mathcal{X}} t_k(y) p(y, a)$$

$$\equiv\; \frac{k-1}{k} t_k(a) + \frac{1}{k}(t_k P)(a)\,, \qquad a \in \mathcal{X}\,,$$

where $(t_k P)(a) = \sum_{y \in \mathcal{X}} t_k(y) p(y, a)$. That is, the limiting fraction of time that the backoff process spends at a state a is given by a weighted average of $t_k(a)$ and $(t_k P)(a)$.

This Proposition tells us that, if we could choose the probabilities $t_k(x)$ so that $\frac{k-1}{k} t_k(a) + \frac{1}{k}(t_k P)(a) = f(a)$ for all $a \in \mathcal{X}$, then the probabilities in BACKMET would indeed converge to $f(x)$. The question then becomes, how should the probabilities $t_k(x)$ be chosen. We consider that next.

Remarks.

1. Note that $\frac{1}{n}\#\{1 \le i \le n;\ X_i = a\}$ is equal to the fraction of the time that the chain spends in the state a. Thus, Proposition 3 says that in the limit, the fraction of time spent at a is equal to $f(a)$, as desired.

2. It is possible that the Metropolis-Hastings algorithm $\{W_n\}$, and hence also the full BACKMET algorithm, will be periodic. This is true, for example, if P corresponds to simple symmetric random walk on the state space $\mathcal{X}$ consisting of M points arranged in a circle, and if $f(x) = 1/M$ for all $x \in \mathcal{X}$ (so that the Metropolis-Hastings algorithm will never reject). However, this is quite rare: As long as the Metropolis-Hastings algorithm has positive probability of rejecting at some point, there will be no periodicity problem. In this case, we will also have $\lim_{n\to\infty} P(X_n = a) = f(a)$.

5. How to Choose t_k.

Given a function $h : \mathcal{X} \to \mathbf{R}$, we define

$$(hP)(y) \;=\; \sum_{x\in\mathcal{X}} h(x)\, p(x,y)\,. \tag{2}$$

Thus $(hP^2)(z) \;=\; ((hP)P)(z) \;=\; \sum_{y\in\mathcal{X}}\sum_{x\in\mathcal{X}} h(x)\,p(x,y)\,p(y,z)$, etc. By inspection, if $h(x)$ is the probability function for the Markov chain at time n, then $(hP)(x)$ is the probability function for the Markov chain at time $n+1$.

Next, we define the function $t_k : \mathcal{X} \to [0,\infty)$ by

$$t_k \;=\; \frac{k}{k-1}\Big(f - \frac{1}{k-1} fP + \frac{1}{(k-1)^2} fP^2 - \frac{1}{(k-1)^3} fP^3 + \dots\Big)\,, \tag{3}$$

where P acts on functions as in (2). We shall require the following condition:

(C1) The series (3) converges for all $y \in \mathcal{X}$, with $t_k(y) \ge 0$ for all $y \in \mathcal{X}$.

Of course, if $k \ge 3$, then the convergence part of (C1) is automatic. Also, note that assuming convergence, we always have $\sum_{x\in\mathcal{X}} t_k(x) = 1$.

We then have

Lemma 4. *Assume (C1), and define t_k by (3). Then $\frac{k-1}{k} t_k + \frac{1}{k} t_k P = f$.*

Proof. We compute that

$$
\begin{aligned}
&\frac{k-1}{k}t_k + \frac{1}{k}t_k P \\
= \;& (\frac{k-1}{k})\frac{k}{k-1}\Big(f - \frac{1}{k-1}fP + \frac{1}{(k-1)^2}fP^2 - \frac{1}{(k-1)^3}fP^3 + \ldots\Big) \\
&+(\frac{1}{k}Q)\frac{k}{k-1}\Big(f - \frac{1}{k-1}fP + \frac{1}{(k-1)^2}fP^2 - \frac{1}{(k-1)^3}fP^3 + \ldots\Big) \\
= \;& \Big(f - \frac{1}{k-1}fP + \frac{1}{(k-1)^2}fP^2 - \frac{1}{(k-1)^3}fP^3 + \ldots\Big) \\
&+\Big(\frac{1}{k-1}fP - \frac{1}{(k-1)^2}fP^2 + \frac{1}{(k-1)^3}fP^3 - \ldots\Big) \\
= \;& f \, ,
\end{aligned}
$$

as claimed. ∎

This Lemma suggests that t_k as defined by the series (3) is indeed the right choice of the probabilities $t_k(x)$ for the BACKMET algorithm. However, the series (3) may fail to converge, or (more likely) may converge to a function which is sometimes *negative* (and hence not a valid density), either of which violates condition (C1). We shall consider this issue in the following section.

However, if (C1) does hold, then we easily conclude the following.

Theorem 5. *Consider the BACKMET algorithm. Assuming irreducibility and (C1), and with t_k defined by (3), we have*

$$\lim_{n\to\infty} \frac{1}{n}\#\{1 \le i \le n;\; X_i = a\} \;=\; f(a)\,, \qquad a \in \mathcal{X}\,.$$

Proof. We know from Proposition 3 that

$$\lim_{n\to\infty} \frac{1}{n}\#\{1 \le i \le n;\; X_i = a\} \;=\; \frac{k-1}{k}t_k(a) + \frac{1}{k}(t_k P)(a)\,.$$

But the previous lemma says that

$$f(a) \;=\; \frac{k-1}{k}t_k(a) + \frac{1}{k}(t_k P)(a)\,.$$

The result follows. ∎

We conclude from Theorem 5 that, assuming irreducibility and (C1), the BACKMET algorithm always achieves the correct limiting distribution $f(x)$. Hence, the one remaining question is, under what circumstances will (C1) hold? We consider that next.

6. The Existence of Solutions.

The above result leads to the following question. For which $\pi(\cdot)$ and P can we be sure that condition (C1) holds, so that Theorem 5 applies and the BACKMET algorithm guarantees that $\lim_{n\to\infty} \frac{1}{n}\#\{1 \leq i \leq n;\ X_i = a\} \ = \ f(a)$ for all $a \in \mathcal{X}$?

Obviously this answer depends on $f(x)$ and $p(x, y)$. In the case where $\mathcal{X}$ is finite, the answer is particularly simple. (Extensions to general state spaces are discussed in the full journal version Feuerverger and Rosenthal, 2002.)

Lemma 6. *Suppose the state space $\mathcal{X}$ is finite, and $f(x) > 0$ for all $x \in \mathcal{X}$. Then there exists k large enough that (C1) holds.*

Proof. Recall that for $k > 2$, $\sum_{j=1}^{\infty}(k-1)^{-j} = 1\,/\,(k-2)$. Furthermore, we note that that $|(fP^j)(x)| \leq 1$ for all j and x. Hence, if we choose k large enough that $\min_{x\in\mathcal{X}} f(x) > 1\,/\,(k-2)$, then we see by inspection of (3) and the triangle inequality, that the series for t_k will converge and will be non-negative. ∎

From Lemma 6, we obtain the following result, which solves our question (essentially completely) for finite state spaces.

Theorem 7. *Suppose the state space $\mathcal{X}$ is finite, and the chain is irreducible. Assume $f(x) > 0$ for all $x \in \mathcal{X}$. Then there exists k large enough, and t_k as given by (3), such that the BACKMET algorithm succeeds in ensuring that*

$$\lim_{n\to\infty} \frac{1}{n}\#\{1 \leq i \leq n;\ X_i = a\} \ = \ f(a)\,, \qquad a \in \mathcal{X}\,.$$

Theorem 7 shows that on a finite state space, provided one is willing to spend *some* positive limiting fraction of time in each state, one can always achieve any desired limiting distribution by use of a BACKMET algorithm for sufficiently large k. Recalling that the BACKMET algorithm does nothing for $k-2$ of every k steps, this says that by adding in enough "do nothing" steps, we can always achieve the limiting distribution $\pi(\cdot)$.

Remark. Another interpretation in the finite-state-space case is as follows. Using matrix notation, we wish to solve for $tM = \pi$, where t and π are row vectors, and M is the matrix having entries $M_{ij} = \frac{k-1}{k}\delta_{ij} + \frac{1}{k}P(i,j)$. Hence, assuming M is invertible (which it often will be), a solution is $t = \pi M^{-1}$. The question then becomes whether or not πQ^{-1} has all non-negative entries (which it often won't).

7. Summary and Extensions.

We summarise our results as follows.

Our goal was to find an algorithm regarding the use of the "back" button, to achieve a prescribed limiting distribution $\pi(\cdot)$, when given Markov chain transition probabilities $P(x, \cdot)$.

To accomplish this, we introduced the BACKMET algorithm, which accomplishes our goal provided the non-negativity condition (C1) holds.

We then proved (Theorem 7) that condition (C1) always holds for finite state spaces, provided only that the target distribution $\pi(\cdot)$ is everywhere-positive.

The case of infinite state space is considered in Feuerverger and Rosenthal (2002). Roughly, it is proved there that condition (C1) will continue to hold on infinite state spaces, provided that the target probabilities (or density) f are bounded away from 0 on compact sets, and also that a certain "tail condition" is satisfied (which says, roughly, that the tails of f must be lighter than those of the increment distribution of the Markov chain).

We feel that our results, taken together, give quite a complete picture of when the BACKMET algorithm will or will not work. However, a number of questions remain, including what happens if the target density is not bounded away from 0 (even on compact sets), or if the state space is infinite and the Markov chain is not a random walk. It is possible that other algorithms, besides BACKMET, could be used in such cases. Similarly, the BACKMET algorithm is somewhat inefficient in that it often does not move, and one can ask whether it can be made more efficient. We look forward to considering such questions in the future.

REFERENCES

Dongarra, J. and Sullivan, F., eds. (2000), The Top 10 Algorithms. Computing in Science and Engineering **2(1)**.

Fagin, R., Karlin, A.R., Kleinberg, J., Raghavan, P., Rajagopalan, S., Rubinfeld, R., Sudan, M., and Tomkins, A. (2001), Random Walks with "Back Buttons". Ann. Appl. Probab. **11**, 810–862. (Conference proceedings version in Proc. 32nd ACM Symposium on Theory of Computing, 2000.)

Feuerverger, A. and Rosenthal, J.S. (2002), Achieving Limiting Distributions for Markov Chains Using Back Buttons. In preparation.

Hastings, W.K. (1970), Monte Carlo sampling methods using Markov chains and their applications. Biometrika **57**, 97–109.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E. (1953), Equations of state calculations by fast computing machines. J. Chem. Phys. **21**, 1087–1091.

Meyn, S.P. and Tweedie, R.L. (1993), *Markov Chains and Stochastic Stability.* Springer-Verlag, London.

Tierney, L. (1994), Markov chains for exploring posterior distributions (with discussion). Ann. Stat. **22**, 1701–1762.