Metropolis-Hastings sampling of alternatives for route choice models

Michel Bierlaire
Gunnar Flötteröd

STRC 2010 **September 2010**

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

STRC 2010

Metropolis-Hastings sampling of alternatives for route choice models

Michel Bierlaire
Transport and Mobility Laboratory
Ecole Polytechnique Fédérale de Lausanne
1015 Lausanne
phone: +41-21-693.25.37
fax: +41-21-693.80.60
michel.bierlaire@epfl.ch

Gunnar Flötteröd
Transport and Mobility Laboratory
Ecole Polytechnique Fédérale de Lausanne
1015 Lausanne
phone: +41-21-693.24.29
fax: +41-21-693.80.60
gunnar.floetteroed@epfl.ch

September 2010

Abstract

We describe a new approach to the sampling of route choice sets using the Metropolis-Hastings algorithm.

Keywords

route choice, sampling of alternatives, Metropolis-Hastings

1 Introduction

The objective of a route choice model is to describe along which path an individual travels from an origin to a destination in a transportation network. A general review on the topic is provided in Frejinger (2008). The large number of unknown alternatives is one of the facets that renders route choice modeling a challenging problem.

Frejinger and Bierlaire (2010) identify two classes of approaches to the modeling of route choice sets: one based on behavioral considerations and one being econometrically motivated. In the behavioral approach, the analyst tries to identify choice sets that are behaviorally plausible. The main criticism of these methods is their disability to reproduce the actually chosen route. The econometric approach makes the (behaviorally questionable yet mathematically convenient) assumption that all elements of the universal set $\mathcal{C}$ are considered by the decision maker and that the elements of the modeled choice set $\mathcal{C}_n \subset \mathcal{C}$ are sampled from $\mathcal{C}$ according to some distribution $\{q(i)\}_{i\in\mathcal{C}}$ specified by the analyst.

Specifically, the econometric approach allows to estimate the parameters of logit models without bias if the choice probabilities are modified according to

$$P(i|\mathcal{C}_n) = \frac{e^{\mu V_{in} + \ln\left(\frac{k_{in}}{q(i)}\right)}}{\sum_{j\in C_n} e^{\mu V_{jn} + \ln\left(\frac{k_{jn}}{q(j)}\right)}} \tag{1}$$

where $\mathcal{C}_n$ consists of independent samples with replacement from $\{q(i)\}_{i\in\mathcal{C}}$, μ is a scale parameter, V_{in} is the deterministic utility of alternative i for decision maker n, and k_{in} is the number of times alternative i is sampled (McFadden, 1978). A generalization of this result to multivariate extreme value models has recently been presented by Bierlaire *et al.* (2008).

Frejinger *et al.* (2009) propose a random walk (RW) algorithm that, starting from the origin, incrementally and probabilistically adds links to a path until the destination is reached. The link addition distribution is such that the RW is biased towards the shortest path, resulting in an importance sampling strategy. The distribution $\{q(i)\}_{i\in\mathcal{C}}$ according to which entire RWs are generated is derived and used for correction. While being methodologically sound, this approach is computationally cumbersome for larger networks (Prato, 2009; Schüssler and Axhausen, 2009): often enough, the RW circles through the network and/or fails to reach the destination in a passable number of iterations.

We propose to replace RW in the approach of Frejinger *et al.* (2009) by a Metropolis-Hastings (MH) algorithm (Hastings, 1970). The MH method generates a Markov chain (MC) with a stationary distribution that coincides with an arbitrary, pre-specified distribution. In our case, the state of the chain comprises a route, the transition distribution of the MC copmrises a random variation of that route, and the stationary distribution is the sampling strategy $\{q(i)\}_{i\in\mathcal{C}}$.

Algorithm 1 Metropolis-Hastings algorithm

1. set iteration counter $k = 0$
2. select arbitrary initial state i^k
3. repeat beyond stationarity
   (a) draw candidate state j from $\{q(i^k, j)\}_j$
   (b) compute acceptance probability $\alpha(i^k, j) = \min\left(\frac{b(j)q(j,i^k)}{b(i^k)q(i^k,j)}, 1\right)$
   (c) with probability $\alpha(i^k, j)$, let $i^{k+1} = j$; else, let $i^{k+1} = i^k$
   (d) increase k by one

The MH algorithm allows to specify this distribution in un-normalized form $\{b(i)\}_{i\in C}$ where the $b(i) > 0$ are such that $q(i) = b(i)/B$ for all $i \in \mathcal{C}$ where $B = \sum_{j\in\mathcal{C}} b(j)$. That is, $b(i)$ is an un-normalized version of $q(i)$. Since B cancels out in (1), it is sufficient to know $\{b(i)\}_{i\in C}$ in order to correct for the sampling.

2 Framework

Our approach to choice set generation relies on the MH method. Algorithm 1 defines a generic MH algorithm with a proposal transition distribution $Q = (q(i, j))$, and stationary weights $\{b(i)\}_{i\in\mathcal{C}}$. Essentially, Q defines a MC that is run in a biased way that enforces the stationary weights, where the bias is implemented by accepting transitions from a state i to a state j with a probability $\alpha(i, j)$ that prefers transitions towards higher weights $b(j)$.

We specify the application of Algorithm 1 to route choice set generation in terms of the following notation:

Γ	a path
$\lvert\Gamma\rvert$	number of nodes in path Γ
$\Gamma(i)$	the ith node of path Γ
$\Gamma(a, b)$	sub-path of Γ between node positions $1 \leq a \leq b \leq \lvert\Gamma\rvert$
$\Gamma_1 + \Gamma_2$	concatenation of paths Γ_1 and Γ_2 (eliminates node duplications)
$\delta(v, w)$	distance between node v and w (may not be symmetric)

Let $\{q(i)\}_{i\in C}$ be the distribution from which the elements of the choice set are to be drawn, and let $\{b(i)\}_{i\in C}$ be such that $q(i) = b(i)/B$ for all $i \in \mathcal{C}$ where $B = \sum_{j\in\mathcal{C}} b(j)$. That is, $b(i)$ is a non-normalized version of $q(i)$. Since B cancels out in (1), it is sufficient to know $\{b(i)\}_{i\in C}$ in order to correct for the sampling.

In order to apply the MH algorithm, the state space, the proposal transition matrix Q, and the desired stationary distribution need to be specified. The stationary distribution is, apart from normalization, given by $\{b(i)\}_{i\in C}$. The state space and the proposal distribution are defined in Subsections 2.1 and 2.2, respectively.

2.1 Definition of MC state

For the purpose of choice set generation, the state space of the MC must contain the universal choice set $\mathcal{C}$. For technical reasons that are clarified further below, we define a state of the MC as a tuple (Γ, u, d, v) where $\Gamma \in \mathcal{C}$ is a path, u, d, are integer numbers with $1 \leq u < d \leq |\Gamma|$, and $v \in \mathcal{N}$ is a network node. The node v is arbitrary; in particular, it is allowed but not required to be an element of Γ.

2.2 Definition of MC proposal distribution

Apart from technical requirements verified further below, operational considerations affect the choice of the proposal distribution Q. If Q generates insufficient variability in that it creates long sequences of similar states, the MC needs many iterations to cover the relevant part of the state space (where relevance is defined through the weights $b(i)$). If Q generates too much variability, the MC frequently proposes jumps out of the relevant part of the state space, which results in low acceptance probabilities α and long persistence in the same state.

We define Q in terms of the two operations SPLICE and SHUFFLE described below.

SPLICE. Given a current state (Γ, u, d, v), a new state (Γ', u', d', v') is generated by (1) drawing a cycle-free path segment Γ_u that starts at $\Gamma(u)$ and ends at v from some distribution $P(\Gamma_u|\Gamma(u), v)$, (2) drawing a cycle-free path segment Γ_d that starts at v and ends at $\Gamma(d)$ from some distribution $P(\Gamma_d|v, \Gamma(d))$, (3) letting $\Gamma' = \Gamma(1, u) + \Gamma_u + \Gamma_d + \Gamma(d, |\Gamma|)$, (4) letting $u' = u$ and $v' = v$, and (5) updating $d' = d + |\Gamma'| - |\Gamma|$.

This operation randomly replaces the path segment $\Gamma(u, d)$ by a new one that goes through v. Given that u and d are not too far apart, v is not too far away from $\Gamma(u, d)$, and Γ_u and Γ_d are not too circuitous, the newly generated path segment constitutes a local detour from the original path. The probability of going from a feasible $i = (\Gamma, u, d, v)$ to a feasible $i' = (\Gamma', u', d', v')$ is

$$q_{SPLICE}(i, i') = \begin{cases} P(\Gamma'(u', z)|\Gamma(u), v)\cdot & \text{if } u' = u \text{ and } d' = d + |\Gamma'| - |\Gamma| \\ \ldots P(\Gamma'(z, d')|v, \Gamma(d)) & \ldots \text{ and } v' = v \text{ and } \Gamma'(z) = v \\ & \ldots \text{ and } \Gamma'(1, u') = \Gamma(1, u) \\ & \ldots \text{ and } \Gamma'(d', |\Gamma'|) = \Gamma(d, |\Gamma|) \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

SHUFFLE. Given a current state (Γ, u, d, v), a new state (Γ', u', d', v') is generated by (1) drawing the indices u' and d' from some distribution $P(u', d'|\Gamma)$ that ensures$1 \leq u' < d' \leq |\Gamma|$, (2) drawing the node v' from some distribution $P(v'|u', d', \Gamma)$, and (3) retaining the current path without modification in that $\Gamma' = \Gamma$.

This operation randomly shuffles the splicing positions as well as the splicing node, but it does not affect the currently generated path. The probability of going from a feasible $i = (\Gamma, u, d, v)$ to a feasible $i' = (\Gamma', u', d', v')$ is

$$q_{SHUFFLE}(i, i') = \begin{cases} P(u', d'|\Gamma)P(v|u', d', \Gamma) & \text{if } \Gamma' = \Gamma \\ 0 & \text{otherwise.} \end{cases} \quad (3)$$

The proposal distribution Q is such that one of these operations is randomly selected:

$$q(i, i') = \alpha q_{SPLICE}(i, i') + (1 - \alpha) q_{SHUFFLE}(i, i') \quad (4)$$

where $0 < \alpha < 1$ is the probability of selecting the SPLICE operation. We now propose concrete implementations of the distributions through which SPLICE and SHUFFLE are defined.

The SPLICE distribution (2) relies on $P(\Gamma|v, w)$, the probability of generating a path segment Γ between the nodes v and w. We resort here to the RW algorithm of Frejinger *et al.* (2009); the respective definition of $P(\Gamma|v, w)$ is given in Appendix A. Its main shortcoming, the high probability of generating circuitous routes that may even fail to reach the destination within a reasonable number of steps, occurs mainly if the RW is parametrized to generate paths with high variability. Since we are interested only in local modifications, we avoid this shortcoming by introducing a strong bias towards the shortest path. With appropriate parametrization, the RW can even be made to coincide with a shortest path calculation.

The SHUFFLE distribution (3) requires to define $P(u', d'|\Gamma)$, the probability of selecting an upstream/downstream splice position pair, and $P(v|u', d', \Gamma)$, the probability of selecting a splice node. We propose to generate u' and d' by drawing two uniform numbers without replacement from $\{1, \ldots, |\Gamma|\}$ and to assign the smaller one to u' and the larger one to d'. Hence,

$$P(u', d'|\Gamma) = \frac{2}{|\Gamma|^2 - |\Gamma|}. \tag{5}$$

The splice node v' is selected according to a distribution $P(v'|u', d', \Gamma)$ that prefers nodes that are near to the path segment $\Gamma(u', d')$. For this, the distance $\delta(v', \Gamma(u', d'))$ of v' to $\Gamma(u', d')$ is defined as the average length of a detour through v' that starts somewhere in $\Gamma(u', d')$ and ends somewhere later in $\Gamma(u', d')$:

$$\begin{aligned}\delta(v', \Gamma(u', d')) =& \frac{1}{D} \sum_{u''=u'}^{d'-1} \sum_{d''=u'+1}^{d'} [\delta(u'', v') + \delta(v', d'')] \\ =& \sum_{u''=u'}^{d'-1} \frac{d' - u''}{D} \delta(u'', v') + \sum_{d''=u'+1}^{d'} \frac{d'' - u'}{D} \delta(v', d'')\end{aligned} \tag{6}$$

where $D = |\Gamma(u', d')|(|\Gamma(u', d')| - 1)/2$. Based on this, the splice node is selected according to a logit model

$$P(v'|u', d', \Gamma) = \frac{e^{-\mu_{SPLICE}\delta(v', \Gamma(u', d'))}}{\sum_{w \in \mathcal{N}} e^{-\mu_{SPLICE}\delta(w, \Gamma(u', d'))}} \tag{7}$$

where the non-negative parameter μ_{SPLICE} defines the importance of "closeness": a value of zero results in a random node selection, a value approaching infinity results in the deterministic selection of a node that minimizes $\delta(v', \Gamma(u', d'))$.

3 Discussion

The proposed transition distribution (4) has a behavioral interpretation. It may be speculated that route choice sets are incrementally acquired by real travelers through an exploration step that takes an existing route as a starting point. This behavior is mimicked by a SHUFFLE/SPLICE sequence, where the traveler first decides which part of the known route to replace and then more or less randomly fills in the gap. A related interpretation is that new routes are learned by the need to visit intermediate destinations during a trip, which requires to detour from a habitual route through the intermediate destination node. Again, this process can be reflected by a SHUFFLE/SPLICE sequence.

The MH algorithm is guaranteed to converge to the desired stationary distribution if the transition distribution (4) defines an irreducible and aperiodic MC. Irreducibility means that every state can eventually be reached by every other state with a positive probability. Aperiodicity is guaranteed if there is at least one state i with $q(i, i) > 0$. Irreducibility of (4) can be shown by observing that the following three-step transition from any $(\Gamma^k, u^k, d^k, v^k)$ to any $(\Gamma^{k+3}, u^{k+3}, d^{k+3}, v^{k+3})$ is possible with nonzero probability:

1. SHUFFLE such that $u^{k+1} = 1$ and $d^{k+1} = |\Gamma^k|$ (the splice positions are the origin and the destination) and $v^{k+1} \in \Gamma^{k+3}$.
2. SPLICE such that Γ^{k+2} equals Γ^{k+3}.
3. SHUFFLE such u^{k+3}, d^{k+3}, and v^{k+3} are obtained.

Aperiodicity of (4) results immediately from the observation that every SHUFFLE and SPLICE operation has a nonzero probability of leaving the current state unmodified.

A computationally relevant instance of the proposed algorithm is when the RW collapses into a shortest path search. Aperiodicity is still guaranteed in this case because the SHUFFLE operation may still reproduce any given state. To show irreducibility, we introduce an expanded network that is created from the original network by introducing for every link $(v, w) \in \mathcal{N} \times \mathcal{N}$ in the network an additional node x_{vw} and by replacing the original link (v, w) by two link (v, x_{vw}) and (x_{vw}, w). The expanded network is a one-to-one representation of the original network, and hence every path found in the expanded network can be mapped back on an element of $\mathcal{C}$.

Given an initial state $(\Gamma^0, u^0, d^0, v^0)$, and an arbitrary target state (Γ', u', d', v') the following sequence of transitions is possible with positive probability (all variables refer to the expanded network):

1. choose initial state $(\Gamma^0, u^0, d^0, v^0)$ and target state (Γ', u', d', v')
2. for $c = 1 \ldots |\Gamma'| - 2$ do
    (a) $k = 2c - 1$
    (b) SHUFFLE such that $u^k = c$, $v^k = \Gamma'(c + 1)$
    (c) $k = 2c$
    (d) SPLICE such that $\Gamma^k(1, c + 1) = \Gamma'(1, c + 1)$
3. $k = 2|\Gamma'| - 3$
4. SHUFFLE such that $u^k = u'$, $d^k = d'$, $v^k = v'$

This builds the target path incrementally from the origin, and it establishes that it can be reached with positive probability in $2|\Gamma'| - 3$ transitions. The location of an intermediate node within each original link guarantees that every link can be reached by the shortest path algorithm (because the only path from and to an intermediate node goes across its respective link).

A Random walk probability of a given path

This presentation is taken from Frejinger *et al.* (2009).

Let (v, w) be the link from node v to node w and use ℓ as a link variable. The length of link ℓ is C_ℓ and the shortest path from node v to node s is $\text{SP}(v, s)$. The detour made from the shortest path towards s by selecting link (v, w) is measured by

$$x_{(v,w)} = \frac{\text{SP}(v, s)}{C_{(v,w)} + \text{SP}(w, s)}, \tag{8}$$

which is nonlinearly weighted by

$$\omega(\ell|b_1, b_2) = 1 - (1 - x_\ell^{b_1})^{b_2} \tag{9}$$

with b_1 and b_2 being real-valued parameters. The probability $P(\Gamma|r, s)$ of obtaining a path Γ from origin r to destination s through the random walk of Frejinger *et al.* (2009) can then be written as

$$P(\Gamma|r, s) = \begin{cases} \prod_{\ell\in\Gamma} \frac{\omega(\ell|b_1,b_2)}{\sum_{m\in\mathcal{E}_v} \omega(m|b_1,b_2)} & \text{if } \Gamma(1) = r \text{ and } \Gamma(|\Gamma|) = s \text{ and } s \notin \Gamma(2, |\Gamma| - 1) \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

References

Bierlaire, M., D. Bolduc and D. McFadden (2008) The estimation of generalized extreme value models from choice-based samples, *Transportation Research Part B*, **42** (4) 381–394.

Frejinger, E. (2008) Route choice analysis: Data, models, algorithms and applications, Ph.D. Thesis, École Polytechnique Fédérale de Lausanne.

Frejinger, E. and M. Bierlaire (2010) On path generation algorithms for route choice models, in *Choice modelling: the state-of-the-art and the state-of-practice*, 307–315, Emerald Group Publishing Limited.

Frejinger, E., M. Bierlaire and M. Ben-Akiva (2009) Sampling of alternatives for route choice modeling, *Transportation Research Part B*, **43** (10) 984–994.

Hastings, W. (1970) Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, **57** (1) 97–109.

McFadden, D. (1978) Modelling the choice of residential location, paper presented at *Spatial interaction theory and residential location*, 75–96, Amsterdam.

Prato, C. (2009) Assessing the influence of objective choice sets on route choice model estimates and predictions, paper presented at *Proceedings of 12th International Conference on Travel Behaviour Research*, Jaipur, India, December 2009.

Schüssler, N. and K. Axhausen (2009) Accounting for route overlap in urban and suburban route choice decisions derived from GPS observations, paper presented at *Proceedings of 12th International Conference on Travel Behaviour Research*, Jaipur, India, December 2009.