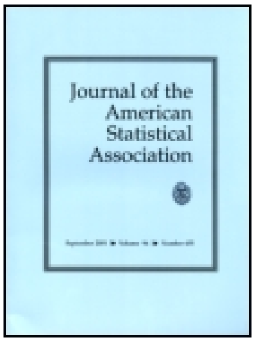

Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: https://amstat.tandfonline.com/loi/uasa20

Bayesian Inference on Network Traffic Using Link Count Data

Claudia Tebaldi & Mike West

To cite this article: Claudia Tebaldi & Mike West (1998) Bayesian Inference on Network Traffic Using Link Count Data, Journal of the American Statistical Association, 93:442, 557-573

To link to this article: https://doi.org/10.1080/01621459.1998.10473707

Bayesian Inference on Network Traffic Using Link Count Data

Claudia TEBALDI and Mike WEST

We study Bayesian models and methods for analysing network traffic counts in problems of inference about the traffic intensity between directed pairs of origins and destinations in networks. This is a class of problems very recently discussed by Vardi in a 1996 *JASA* article and is of interest in both communication and transportation network studies. The current article develops the theoretical framework of variants of the origin-destination flow problem and introduces Bayesian approaches to analysis and inference. In the first, the so-called fixed routing problem, traffic or messages pass between nodes in a network, with each message originating at a specific source node, and ultimately moving through the network to a predetermined destination node. All nodes are candidate origin and destination points. The framework assumes no travel time complications, considering only the number of messages passing between pairs of nodes in a specified time interval. The route count, or route flow, problem is to infer the set of actual number of messages passed between each directed origin-destination pair in the time interval, based on the observed counts flowing between all directed pairs of adjacent nodes. Based on some development of the theoretical structure of the problem and assumptions about prior distributional forms, we develop posterior distributions for inference on actual origin-destination counts and associated flow rates. This involves iterative simulation methods, or Markov chain Monte Carlo (MCMC), that combine Metropolis–Hastings steps within an overall Gibbs sampling framework. We discuss issues of convergence and related practical matters, and illustrate the approach in a network previously studied in Vardi's article. We explore both methodological and applied aspects much further in a concrete problem of a road network in North Carolina, studied in transportation flow assessment contexts by civil engineers. This investigation generates critical insight into limitations of statistical analysis, and particularly of non-Bayesian approaches, due to inherent structural features of the problem. A truly Bayesian approach, imposing partial stochastic constraints through informed prior distributions, offers a way of resolving these problems and is consistent with prevailing trends in updating traffic flow intensities in this field. Following this, we explore a second version of the problem that introduces elements of uncertainty about routes taken by individual messages in terms of Markov selection of outgoing links for messages at any given node. For specified route choice probabilities, we introduce the concept of a super-network—namely, a fixed routing problem in which the stochastic problem may be embedded. This leads to solution of the stochastic version of the problem using the methods developed for the original formulation of the fixed routing problem. This is also illustrated. Finally, we discuss various related issues and model extensions, including inference on stochastic route choice selection probabilities, questions of missing data and partially observed link counts, and relationships with current research on road traffic network problems in which travel times within links are nonnegligible and may be estimated from additional data.

KEY WORDS: Bayesian inference; Markov chain Monte Carlo; Network flow data; Origin-destination flow.

1. INTRODUCTION

In a recent article, Vardi (1996) addressed problems of inference in traffic intensity counts in networks based on observed traffic counts on directed network links between pairs of nodes. The archetype application context is a communication network. Here traffic counts on an individual directed link represent numbers of messages transmitted across that link. Interest lies, at least in part, in estimating the source, or origin, to destination counts; that is, for each pair of communicating nodes i and j in the network, estimate the number of packets transmitted from i to j based on the observed link counts. This is one of a collection of problems of interest in various areas, including, with extensions to incorporate nonnegligible and stochastic link transit times, studies of transportation networks. In this area there is long-standing interest in estimation of so-called "origin-destination (OD) matrices" that represent the actual or expected flows or counts of traffic between all possible pairs of origin ("O") and destination ("D") nodes or zones in urban road networks during a specific time interval. These matrices are used as inputs to network equilibrium models that are central tools in transportation modeling and policy studies (Berka and Boyce 1994; Sheffi 1995; West 1994). Our work directly connects with specific road transportation network problems, and we develop an applied study in the area in Section 4.

This article addresses the network count inference problem from a Bayesian perspective. We elaborate on the theoretical structure of the basic network count problem and develop analysis for the so-called fixed routing problem; that is, inferring flows of traffic between all directed OD pairs based on observed counts on all links in the network. This underlies the development of posterior analysis based on iterative simulation methods and is presented in Section 2 following introduction of the network specification, notation, terminology, and basic structure. Various difficulties of analysis encountered in non-Bayesian approaches are alleviated by approaching the problem in this way. Throughout Section 2 and in Section 3, we illustrate the ideas and present some computational results in an example network studied by Vardi (1996). Section 2 focuses on the algebraic and statistical structure of the basic network flow problem and discusses prior modeling and aspects of resulting pos-

Claudia Tebaldi is postdoctoral fellow, Climate Analysis Section/Geophysical Statistics Project, National Center for Atmospheric Research, Boulder, CO 80307. Mike West is Professor and Director, Institute of Statistics and Decision Sciences, Duke University, Durham, NC 27708. The research reported here was partially supported by National Science Foundation grant DMS-9313013 to the National Institute of Statistical Sciences (NISS). The authors acknowledge useful discussions with Yehuda Vardi of Rutgers University and Alan F. Karr of NISS. The Monroe datasets were provided by Nagui Rouphail and Steven Click of North Carolina State University.

Journal of the American Statistical Association
June 1998, Vol. 93, No. 442, Theory and Methods

terior distributions. Here we discuss and exemplify a general structural feature of the network problem that raises issues and challenges for any statistical analysis. This relates to unavoidable ambiguities in inferences based solely on likelihood or other non-Bayesian methods and whose solution requires imposition of additional constraints. We discuss and demonstrate how this is naturally achieved by imposition of partial stochastic constraints through informed prior distributions. Section 3 is concerned with issues of implementation and computation. Section 4 discusses an informative real application. Section 5 discusses extensions to incorporate stochastic route selection, also discussed by Vardi (1996), using Markov models for sequential selection of links to be traversed in moving between a specific OD pair. We develop a simple and natural embedding of such a problem in a larger, artificial network introduced to map the stochastic route choice problem onto the fixed route framework. This implies that a solution to the stochastic version of the problem lies in the fixed route problem already developed. This neat solution is attractive in its conceptual simplicity and rather interestingly reverses the roles of fixed and stochastic networks; whereas the fixed network problem is a special case of stochastic routing, the solution to the latter is achieved by reversing the roles and defining it as a special fixed routing problem. Section 6 concludes the article with discussions about extensions to include inference on stochastic route choice selection probabilities, questions of missing data and partially observed link counts, and relationships with current research on road traffic network problems.

2. THE BASIC NETWORK INFERENCE PROBLEM

2.1 Notation and Structure

We adopt the basic structure and notation of Vardi (1996). Consider a fixed network of n nodes, arbitrarily labelled $A, B \ldots$. The basic problem considers a fixed period of communication or passage of traffic during which a collection of messages moves through the network; each message originates at one node in the network (its origin or source node) and travels to another (its destination node) along a unique path as long as we deal with a "fixed routing problem." We are concerned with actual numbers, or counts, of messages travelling between pairs of nodes in the network. For any ordered pair of nodes $a = (i, j)$, write X_a for the number of messages originating at node i and terminating at its destination node j. X_a is the OD count between these two nodes. Given n nodes in the network, there are $c = n(n-1)$ OD pairs. Interest lies in estimating the full set of such counts based on observations of traffic counts on all individual directed links in the network; that is, of traffic counts between pairs of nodes that communicate directly, without intervening nodes. Let r be the total number of directed links in the network. Let $s = (i, j)$ represent the directed link from node i to node j, and write Y_s for the traffic count on this link. Then, based on observed link counts $\mathbf{Y} \stackrel{\text{def}}{=} (Y_1, \ldots, Y_r)'$, we are interested in inferring OD counts $\mathbf{X} \stackrel{\text{def}}{=} (X_1, \ldots, X_c)'$. Note that the number of observed link counts r is typically smaller than the number of OD pairs c, especially in very large networks. Throughout the article, we implicitly assume that $r < c$.

Following Vardi (1996), $\mathbf{Y}$ and $\mathbf{X}$ are related through the $r \times c$ routing matrix $\mathbf{A} = \{A_{s,a}\}$, where $A_{s,a} = 1$ if the directed link s belongs to the directed route through the network between OD pair a, and $A_{s,a} = 0$ otherwise. Note that $\mathbf{A}$ is typically singular, with a number of columns larger than the number of rows. With this definition, and with the links implicitly ordered $1, \ldots, r$, we have the defining identity:

$$\mathbf{Y} = \mathbf{AX}. \tag{1}$$

This simply expresses each link count Y_s as the sums of the OD counts for all routes that include directed link s. From an algebraic perspective, $\mathbf{Y}$ imposes a set of linear constraints on $\mathbf{X}$. The routing matrix summarizes the network traffic structure in useful ways. For example, $(\mathbf{AA}')_{a,a}$ counts the number of OD routes passing through link a, and $(\mathbf{AA}')_{a,b}$ counts the number of routes that pass through both links a and b. Similarly, $(\mathbf{A}'\mathbf{A})_{s,s}$ is the number of links in route s, and $(\mathbf{A}'\mathbf{A})_{s,t}$ is the number of links that routes s and t share in common.

A realistic problem will have a routing matrix that has no duplicate columns and in which each column has at least one nonzero entry. This relates to an identification issue in the estimation approach of Vardi (1996, sec. 2), though it is a practically relevant constraint independent of issues of estimation. A duplicate column means a redundancy in OD pair specification. A zero column is disallowed, as it would imply a noncommunicating OD pair. Similarly, a zero row would imply a link lying outside the network, and so is also disallowed.

Vardi (1996) gave several example networks. One of these is the $n = 4$ node network in Figure 1. This network, analysed further later, has $r = 7$ directed links and $c = 12$ OD pairs; the ordered sequence of nodes comprising these links and OD routes appears in Figure 2. The corresponding 7×12 routing matrix is

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix}.$$

2.2 Prior Models for X

Our problem is that of computing and summarizing the joint posterior distribution $p(\mathbf{X}|\mathbf{Y})$ for all route counts $\mathbf{X}$ given the observed link counts $\mathbf{Y}$. This requires a model for the prior distribution $p(\mathbf{X})$ to be tied together with the deterministic expression (1) that implies $\mathbf{Y}$ given $\mathbf{X}$. Consider first the kinds of models that we might assume for $p(\mathbf{X})$.

The statistical framework of Vardi (1996) involves an assumption that the origin-destination traffic $\mathbf{X}$ is generated

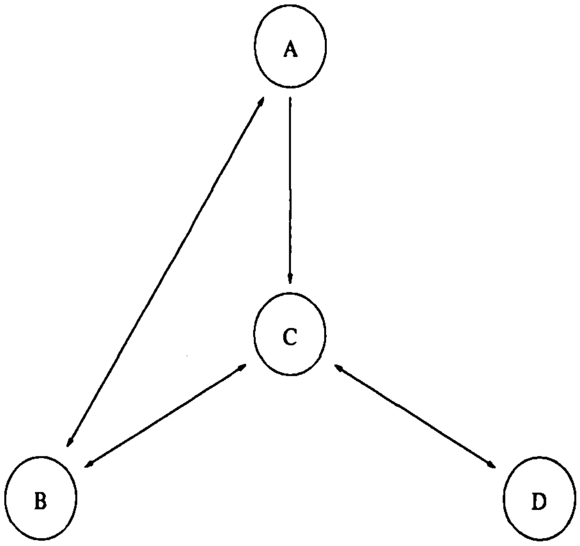

Figure 1. An Example Four-Node Network.

from a collection of independent Poisson distributions for the elements X_a. Assuming $X_a \sim Po(\lambda_a)$, independently over routes a, Vardi explores likelihood and method-of-moments estimation approaches to inference on the Poisson rates $\mathbf{\Lambda} \stackrel{\text{def}}{=} \{\lambda_1, \ldots, \lambda_c\}$. Both Poisson and independence assumptions have been common in transportation network flow applications, though both are questionable in many contexts. In developing prior estimates of OD matrices in connection with network equilibrium models, for example (Berka and Boyce 1994; Sheffi 1995), conditional Poisson models are typical, although dependencies between flows are often strong and may be modeled through hierarchical models for the rates $\mathbf{\Lambda}$ (see, e.g., Sen 1987; West 1994). Thus, for example, West (1994) developed Bayesian analyses under the Poisson assumption but with random-effects regression models relating the Poisson rates λ_a. His models, estimated based on survey data, incorporate origin and, separately, destination specific effects, together with interaction terms that modify the Poisson rates as a function of distance (in terms of estimated travel times) between origin and destination and with additional route specific random effects to model residual patterns of interactions and unexplained extra-Poisson variability. Thus more complex prior models than the independent Poisson forms are available by introducing structure among the Poisson rates through more or less standard hierarchical components. These kinds of models can be incorporated in our development later. For the current study, however, we restrict attention to the independent Poisson structure to simplify presentation of the key developments and to tie in directly with the previous work of Vardi.

Thus we assume, following Vardi (1996), that $X_a \sim Po(\lambda_a)$, independently over a. Prior specification is completed by a prior for $\mathbf{\Lambda}$, determining a joint model

$$p(\mathbf{X}, \mathbf{\Lambda}) = p(\mathbf{\Lambda}) \prod_{a=1}^{c} \lambda_a^{X_a} \exp(-\lambda_a)/X_a!, \tag{2}$$

as the starting point for analysis. Though we are primarily interested in estimating $\mathbf{X}$, estimation of $\mathbf{\Lambda}$ will be an enabling activity, and we will jointly infer $\mathbf{X}$ and $\mathbf{\Lambda}$ from (2) as a result. Begin with the assumption that the Poisson rates are independently drawn from specified marginal prior distributions $p(\lambda_a)$ for each a. As mentioned earlier, this is consistent with more realistic models, for some applied contexts, in which the priors $p(\lambda_a)$ will themselves be further structured hierarchically in terms of hyperparameters.

2.3 Conditional Posteriors for MCMC Posterior Simulation

Given the prior (2), we now observe the link counts $\mathbf{Y}$ and will condition to deliver the required posterior $p(\mathbf{X}, \mathbf{\Lambda}|\mathbf{Y})$. Naturally, posterior computations are difficult analytically in any other than trivial and quite unrealistic networks, so that our approach develops iterative MCMC simulation methods. Consider in particular Gibbs sampling, in which we iteratively resample from conditional posteriors for elements of the $\mathbf{X}$ and $\mathbf{\Lambda}$ variables.

First, consider $\mathbf{\Lambda}$. We note that

$$p(\mathbf{\Lambda}|\mathbf{X}, \mathbf{Y}) \equiv p(\mathbf{\Lambda}|\mathbf{X}) = \prod_{a=1}^{c} p(\lambda_a|X_a),$$

whose components have the form of the prior density $p(\lambda_a)$ multiplied by the gamma form arising in the Poisson-based likelihood function. Thus, conditional on $\mathbf{X}$, we can easily simulate new $\mathbf{\Lambda}$ values as a set of independent draws from the implied univariate posteriors. If $p(\lambda_a)$ is gamma, or a mixture of gammas, then these draws are trivially made from the corresponding gamma or mixture of gamma posteriors. Otherwise, we may use a rejection method or embed Metropolis–Hasting steps in the MCMC scheme in the by-now standard "Metropolis-within-Gibbs" framework (see, e.g., the many examples in Gilks, Richardson, and Spiegelhalter 1996).

Now turn to the conditional posterior $p(\mathbf{X}|\mathbf{\Lambda}, \mathbf{Y})$, viewing $\mathbf{\Lambda}$ as fixed. Our data $\mathbf{Y}$ are in the form of a set of linear constraints (1) on the route count vector $\mathbf{X}$, so that conditioning must be performed directly, algebraically, rather than via the usual application of Bayes's theorem. In the following subsection we develop a completely general and automatic approach to analysis of this conditional posterior.

$Y_1 : A \to B$	$X_1 : A \to B$
$Y_2 : B \to A$	$X_2 : A \to C$
$Y_3 : A \to C$	$X_3 : A \to C \to D$
$Y_4 : B \to C$	$X_4 : B \to A$
$Y_5 : C \to B$	$X_5 : B \to C$
$Y_6 : C \to D$	$X_6 : B \to A \to C \to D$
$Y_7 : D \to C$	$X_7 : C \to B \to A$
	$X_8 : C \to B$
	$X_9 : C \to D$
	$X_{10} : D \to C \to B \to A$
	$X_{11} : D \to C \to B$
	$X_{12} : D \to C$

Figure 2. The Link and Route Structure for the Network in Figure 1.

Here, to motivate that development and generate additional insights, we detail the structure of the posterior in a very simple example.

Take the network in Figure 3, just two directed links between nodes A, B, and C, and with routing matrix

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}.$$

This has three routes: $A \to B, B \to C$, and $A \to C$, with unobserved counts X_1, X_2, and X_3 and links counts Y_1 on $A \to B$ and Y_2 on $B \to C$. Trivially, we see that knowledge of X_3 in addition to $\mathbf{Y}$ implies that both X_1 and X_2 are known; that is, $X_1 = Y_1 - X_3$ and $X_2 = Y_2 - X_3$. Furthermore, under the independent Poisson priors for the X_a it easily follows that the marginal posterior density of X_3 is

$$p(X_3|\mathbf{\Lambda}, \mathbf{Y}) \propto p(X_3|\lambda_3) \prod_{a=1,2} p(X_a|\lambda_a) I_{Y_a - X_3}(X_a),$$

where $I_y(x)$ is the indicator of $x = y$ for fixed y. This reduces to

$$p(X_3|\mathbf{\Lambda}, \mathbf{Y}) \propto \frac{\lambda_3^{X_3}}{X_3!} \prod_{a=1,2} \frac{\lambda_a^{Y_a - X_3}}{(Y_a - X_3)!}$$

over the support $X_3 = 0, 1, \ldots, \min(Y_1, Y_2)$. Hence the full joint posterior $p(\mathbf{X}|\mathbf{\Lambda}, \mathbf{Y})$ can be simulated by drawing X_3 from the foregoing distribution, then directly evaluating X_1 and X_2.

This example, though trivial almost to the point of degeneracy, illuminates the general structure of the problem of conditioning on $\mathbf{Y}$ to deduce $p(\mathbf{X}|\mathbf{\Lambda}, \mathbf{Y})$. In general, the r vector $\mathbf{Y}$ and defining equation (1) implies a set of r linear constraints on $\mathbf{X}$. As a result, the posterior for $\mathbf{X}$ may be reduced to set of r linear equations that deliver precise values of r elements of $\mathbf{X}$ given specified values of the remaining $c - r$; the marginal posterior for the remaining $c - r$ elements may be directly evaluated and used as the centerpiece for posterior simulation in the MCMC framework. We move on with such developments later, but first describe and exemplify some further aspects of the model, and of resulting issues of ambiguity of inferences in the absence of additional constraints or truly informed priors.

2.4 Structural Ambiguity and the Need for Informative Priors

In raw algebraic form, the data $\mathbf{Y}$ simply provides a set of linear constraints on the unobserved flows $\mathbf{X}$. Under the assumed independent Poisson priors for the X_i, and conditional on the values of the Poisson means λ_i, the data simply constrain the conditional prior $p(\mathbf{X}|\mathbf{\Lambda})$ to the support defined by $\mathbf{Y} = \mathbf{AX}$, so defining "ridges" in the resulting likelihood function for Λ. This is best seen through simple

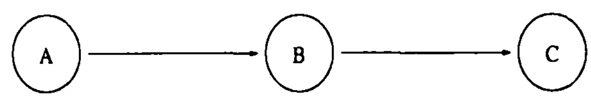

Figure 3. A Simple Three-Node Network With Just Two Directed Links.

examples, so we continue with the example of Figure 3. Here we have constraints $Y_1 = X_1 + X_3$ and $Y_2 = X_2 + X_3$. Assume that the underlying mean OD flows on links one and two are the same, $\lambda_1 = \lambda_2 = \lambda$, and are distinct from $\lambda_3 = \mu$. Clearly, there is an identification problem in that we can write $Y_1 = X_1 + X_3 = (X_1 + a) + (X_3 - a)$ and $Y_2 = X_2 + X_3 = (X_2 + a) + (X_3 - a)$ for any integer a less than or equal to $\min(Y_1, Y_2)$. This impacts the resulting likelihood function for (λ, μ) by inducing a diagonal ridge running north–west to south–east, reflecting the strong negative dependence induced by the constraints. Variation in the likelihood along this diagonal ridge depends on the actual data values $\mathbf{Y}$ observed and can lean toward either extreme or more highly support central values. Particularly critical are cases in which X_1 and X_2 are rather large and happen to have close values, but X_3 is small. The likelihood function based on large and similar values of (Y_1, Y_2) then tends to favor the central part of the ridge, so apparently overestimating μ and underestimating λ. This is induced partly by the relatively higher dispersion of the Poisson densities at higher levels, which leads to higher likelihood values for small means when "close" values of the Poisson counts are observed. As a result, analysis with uniform priors on μ and λ will lead to posteriors that tend to overestimate low rates and the corresponding OD flows, and to underestimate high rates and flows, though the extent of these biases is much more marked in the overestimation of low flows. Thus purely likelihood-based analyses are subject to such biases without additional constraints.

This phenomenon occurs quite generally and is responsible for gross overestimation of low OD flows and their means in some analyses, such as discussed in our application in Section 4. That application concerns a much more complex network that the simple example presented earlier, and several links with rather low OD flows are tied together in collections of OD routes that experience very high flow rates, leading to the gross overestimation of the low flow rates. This argues against the use of uniform priors and suggests that the standard practice in transportation studies of updating initial or prior estimates of OD flows based on previous data and experience (Maher 1983; McNeil and Hendrickson 1985) is partly helping to constrain the problem to overcome these kinds of identification difficulties. We note further that alternative estimation methods that do not recognize this problem, such as that of Vardi (1996), are immediately subject to the same kinds of biases in resulting inferences.

2.5 Algebraic Structure of General Network

We follow the foregoing example and discussion of Section 2.3 with further exploration of the theoretical structure of a general network. This leads to a theoretical result that underlies a generally applicable method of computing samples from conditional posteriors for route counts in an iterative Gibbs sampling framework. In addition to directly enabling simulation based analysis, the discussion is of independent interest in elaborating further on the structure of the network problem.

Consider $\mathbf{A}$ to be fixed and known, and focus attention on the conditional posterior $p(\mathbf{X}|\mathbf{A},\mathbf{Y})$. The following result, which is simply an algebraic deduction from the network structure and defining relation (1) among the traffic counts, is key and critical to the ensuing inferential development.

Theorem 1. In the network model (1), assume that $\mathbf{A}$ is of full rank r. Then we can reorder the columns of $\mathbf{A}$ so that the revised routing matrix has the form

$$\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2], \tag{3}$$

where $\mathbf{A}_1$ is a nonsingular $r \times r$ matrix. Also, similarly reordering the elements of $\mathbf{X}$ vector and conformably partitioning as $\mathbf{X}' = (\mathbf{X}_1', \mathbf{X}_2')$, it follows that

$$\mathbf{X}_1 = \mathbf{A}_1^{-1}(\mathbf{Y} - \mathbf{A}_2\mathbf{X}_2). \tag{4}$$

Proof. The partitioning of $\mathbf{A}$ in (3) is an essentially trivial consequence of the full rank assumption. Matrix $\mathbf{A}$ is $r \times c$ with $r < c$, and so has at least r linearly independent columns; simply reordering so that the first r columns are among the linearly independent subset produces (3). Then reorder the elements of $\mathbf{X}$ to correspond to the new ordering of the columns of $\mathbf{A}$. This gives $\mathbf{Y} = \mathbf{AX} = \mathbf{A}_1\mathbf{X}_1 + \mathbf{A}_2\mathbf{X}_2$, which, noting the invertibility of $\mathbf{A}_1$, easily leads to (4).

The full rank assumption is satisfied by all practical networks. Otherwise, there is a redundancy in specification and one or more rows of $\mathbf{A}$ that can be deleted to reduce to linear independence of the rows, hence full rank of the reduced matrix. Given the partitioned form (3), and with the implicit reordering of elements of $\mathbf{X}$, the result (4) implies that given $\mathbf{Y}$ and assumed values of the $(c-r)$ route counts in $\mathbf{X}_2$, we can directly compute the remaining r route flows, based simply on the algebraic structure of the defining routing matrix. In some cases, we may be able to directly identify $\mathbf{A}_1$ by inspection. More often, especially in larger networks, we need an algorithm to appropriately reorder routes and deduce the resulting routing matrix in the form (3) automatically. This may be done as follows.

Recall the QR decomposition of arbitrary, full-rank matrices (see, e.g., Golub and Van Loan 1983, sec. 6.2). Applied to the $r \times c$ routing matrix $\mathbf{A}$, this delivers the QR decomposition in terms of an $r \times r$ orthogonal matrix $\mathbf{Q}$ and an $r \times c$ upper triangular matrix $\mathbf{R}$ such that $\mathbf{A} = \mathbf{QR}$. The first r columns of $\mathbf{R}$ correspond to r linearly independent columns of $\mathbf{A}$; they are identified by a permutation of column indices that are typically delivered by iterative procedures for developing the QR decomposition. For example, the efficient Householder successive reflection procedure, as implemented in the `qr( )` routine in S-PLUS (1993), delivers a list of indices of columns of $\mathbf{A}$ in permutation such that the first r elements identify the linearly independent columns of $\mathbf{A}$. As a result, we can very simply, routinely, and automatically identify a reordering of the columns of $\mathbf{A}$ to achieve the form (3) with $\mathbf{A}_1$ of full rank: Simply apply the QR decomposition and extract the column permutation vector, then reorder the columns of $\mathbf{A}$ according to this vector. In passing, note that $\mathbf{A}_1$ is simply the first r columns of the matrix $\mathbf{R}$.

In our example network of Section 2.1, the `qr(A)` call in S-PLUS produces, among other things, the vector `pivot` of the permuted column indices—in this case 1, 2, 3, 4, 5, 7, 10, 11, 12, 9, 8, 6. Reordering columns 1–12 accordingly leads to the revised routing matrix:

$$\mathbf{A} = \begin{pmatrix} 1&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&1&0&1&1&0&0&0&0&1 \\ 0&1&1&0&0&0&0&0&0&0&0&1 \\ 0&0&0&0&1&0&0&0&0&0&0&0 \\ 0&0&0&0&0&1&1&1&0&0&1&0 \\ 0&0&1&0&0&0&0&0&0&1&0&1 \\ 0&0&0&0&0&0&1&1&1&0&0&0 \end{pmatrix}.$$

It is easily seen that the first 7×7 block is nonsingular with inverse

$$\mathbf{A}_1^{-1} = \begin{pmatrix} 1&0&0&0&0&0&0 \\ 0&0&1&0&0&-1&0 \\ 0&0&0&0&0&1&0 \\ 0&1&0&0&-1&0&0 \\ 0&0&0&1&0&0&0 \\ 0&0&0&0&1&0&-1 \\ 0&0&0&0&0&0&1 \end{pmatrix}.$$

With the general Theorem 1 and the resulting constructive method of applying it, we are now in a position to develop the conditional posteriors for route counts in a Gibbs sampling framework.

2.6 Conditional Posteriors for Route Counts

From the results of the previous section, we see that for any $\mathbf{\Lambda}$ and fixed link counts $\mathbf{Y}$, the conditional distribution $p(\mathbf{X}|\mathbf{\Lambda},\mathbf{Y})$ is concentrated in a subspace of dimension $c-r$ defined by the partition (3) of the routing matrix. Having reordered the columns of $\mathbf{A}$ to the form (3), this posterior has the form

$$p(\mathbf{X}_1|\mathbf{X}_2,\mathbf{\Lambda},\mathbf{Y})p(\mathbf{X}_2|\mathbf{\Lambda},\mathbf{Y}),$$

where $p(\mathbf{X}_1|\mathbf{X}_2,\mathbf{\Lambda},\mathbf{Y})$ is degenerate at $\mathbf{X}_1 = \mathbf{A}_1^{-1}(\mathbf{Y} - \mathbf{A}_2\mathbf{X}_2)$, and with $\mathbf{X}_2 = (X_{r+1},\dots,X_c)'$ defining $\mathbf{X}_1 = (X_1,\dots,X_r)'$ as earlier,

$$p(\mathbf{X}_2|\mathbf{\Lambda},\mathbf{Y}) \propto \prod_{a=1}^{c} \frac{\lambda_a^{X_a}}{X_a!} \tag{5}$$

over the support defined by $X_a \geq 0$ for all $a = 1,\dots,c$. This is simply the expression of the product of independent Poisson priors for the X_i constrained by the identity (1) rewritten in the form (4). The utility of this expression is in delivering the set of complete conditional posteriors for elements of the $\mathbf{X}_2$ vector to form part of the iterative simulation approach to posterior analysis. Consider each elements X_i of $\mathbf{X}_2$, $(i = r+1,\dots,c)$, and write $\mathbf{X}_{2,-i}$ for the remaining elements. Then, simply by inspection of (5), we see that the conditional distribution $p(X_i|X_{-i},\mathbf{\Lambda},\mathbf{Y})$ is given by

$$p(X_i|\mathbf{X}_{2,-i},\mathbf{\Lambda},\mathbf{Y}) \propto \frac{\lambda_i^{X_i}}{X_i!}\prod_{a=1}^{r}\frac{\lambda_a^{X_a}}{X_a!} \tag{6}$$

over the support defined by $X_i \geq 0$ and $X_a \geq 0$ for each $a = r+1,\ldots,c$; this holds for each $i = r+1,\ldots,c$.

Identifying the support of (6) requires study of the (at most r) linear constraints on X_i defined by $X_a \geq 0$ for all elements X_a of $\mathbf{X}_1 = \mathbf{A}_1^{-1}(\mathbf{Y} - \mathbf{A}_2\mathbf{X}_2)$. Given i in $r+1,\ldots,c$ (i.e., given a single component of the $\mathbf{X}_2$ vector), this implies a set of linear constraints as functions of the conditioning values of $\mathbf{X}_{2,-i}$ and $\mathbf{Y}$. The resulting constraints are the form $X_i \geq d_i$ or $X_i \leq e_i$, where the values d_i and e_i are functions of the conditioning values of $\mathbf{X}_{2,-i}$ and $\mathbf{Y}$. Hence, together with $X_i \geq 0$, we obtain a set of at most $r+1$ constraints on X_i. By directly evaluating these constraints and identifying their intersection, we may deduce the range of X_i over which (6) is nonzero, and hence we identify the unnormalized conditional posterior distribution. But in networks of even moderate size, this process can become computationally very burdensome, a point that is relevant in developing efficient computation in the following section. Finally, it can be shown that the support is a bounded and connected subset of nonnegative integers; details are given in the Appendix. This is important, as it enables rather direct simulation analysis as part of the overall posterior simulation analysis to which we now turn.

3. IMPLEMENTATION AND POSTERIOR SIMULATION

3.1 Gibbs and Metropolis–Hastings Algorithms

Iterative simulation of the full posterior $p(\mathbf{X},\mathbf{\Lambda}|\mathbf{Y})$ is now enabled. Fix starting values of the route counts $\mathbf{X}$ and proceed as follows:

a. Draw sampled values of the rates $\mathbf{\Lambda}$ from the c conditionally independent posteriors $p(\lambda_a|X_a)$ detailed in Section 2.3.
b. Conditioning on these values of $\mathbf{\Lambda}$, simulate a new $\mathbf{X}$ vector by sequencing through $i = r+1,\ldots,c$, and at each step sampling a new X_i from (6), with conditioning elements $\mathbf{X}_{2,-i}$ set at their most recent sampled values; at each step $\mathbf{X}_1$ is explicitly reevaluated via (4) as a function of the most recently sampled elements of $\mathbf{X}_2$.
c. Return to step a and iterate.

This is a standard Gibbs sampling setup in which the scalar elements of both $\mathbf{\Lambda}$ and $\mathbf{X}$ are resampled from the relevant distribution conditional on most recently simulated values of all other uncertain quantities. Sampling steps in a are easy. In the illustration in the following section, the λ_a are assigned independent priors uniform over a fixed, specified, and finite range, so that the posteriors are simply truncated gamma distributions that for λ_a have shape parameter X_i+1 and scale parameter 1; these are trivially sampled. Sampling steps in b appears to require evaluation of the support of (6), as previously discussed, and subsequent evaluation of the unnormalized posterior (6) at each step. Sampling may be performed directly, treating (6) as a simple multinomial distribution on this relevant range. This is easy to implement and has been explored in some example networks. But in larger, more realistic networks, the implied evaluation of (6) across what may be a very large support, at each iteration and for each element X_i, leads to a computational burden that may be excessive when compared to alternative approaches. Moreover, to do this requires identifying the support of (6) which, as mentioned earlier, can become computationally very burdensome in networks of even moderate size. Hence indirect but very much more efficient simulation methods become of interest.

In particular, more efficient algorithms are based on embedding Metropolis–Hastings steps within the Gibbs sampling framework. Here candidate values of the X_i are generated at each stage from suitable proposal distributions, and accepted or rejected according to the usual Metropolis–Hastings acceptance probabilities. Specifically, assume a specified and fixed proposal distribution with probability mass function $q_i(X_i)$ for each element X_i in step b. A candidate value X_i^* is drawn from $q_i(\cdot)$ and accepted with probability

$$\min\left[1, \frac{p_i(X_i^*)q_i(X_i)}{p_i(X_i)q_i(X_i^*)}\right],$$

where X_i is the current, most recently sampled value and $p_i(\cdot)$ is the unnormalized conditional posterior in equation (6). Note that the unnormalized density $p_i(\cdot)$ will be evaluated only at candidate draws, so that for a given proposal distribution $q_i(\cdot)$, it is not necessary either to identify the actual support of $p_i(\cdot)$ or to evaluate it completely across the support, in contrast to the approach via direct Gibbs sampling. If we use a proposal distribution $q_i(\cdot)$ whose support includes values of X_i lying outside the support of $p_i(\cdot)$, then candidate draws of such values will automatically be rejected, as they lead to zero acceptance probabilities.

From the structure of the network equations in (1), it is possible to identify bounds on each X_i so that a suitable range for the proposal distribution can be computed. For element X_a given $\mathbf{X}_{2,-a}, X_a$ of $\mathbf{X}_2$, $X_a = 0$ is a gross lower bound whatever the values in $\mathbf{X}_{2,-a}$. For an upper bound, $X_a \leq \min_i\{Y_i - \sum_{j\neq a} A_{i,j}X_j\}$, where the index i runs over the set of links whose counts include X_a; that is, those links i for which $A_{i,j} = 1$. Our algorithm is an iterative, trial-and-error search to identify upper and lower bounds for X_a given $\mathbf{X}_{2,-a}$ using these two values to start. Then, based on the specified bounds, the implied vector $\mathbf{X}_1$ is recomputed and checked for feasibility; that is, nonnegative values. If any element of $\mathbf{X}_1$ is negative, then the trial value of X_a is either incremented, in searching for the lower bound on its range, or decremented, in searching for the upper bound. This process terminates and delivers the resulting bounds once the $\mathbf{X}_1$ vector has r nonnegative entries.

This approach represents a standard application of "Metropolis-within-Gibbs" and, subject to verification of irreducibility of the resulting Markov chain on the space inhabited by $(\mathbf{X},\mathbf{\Lambda})$, leads to an overall simulation scheme that will ultimately generate samples from the required joint posterior (Tierney 1994). We have experimented with various such approaches, concentrating on efficient independence chain methods. One such method uses a proposal distribution for $q_i(\cdot)$ that is uniform over the identified support of (6); an alternative takes $q_i(\cdot)$ as the conditional Poisson prior $X_i \sim Po(\lambda_i)$. In each case, both generation of candidate values and evaluation of the acceptance probability are essentially trivial. In our examples to date, each of these approaches has been satisfactory in terms of reasonable acceptance rates and convergence, with the two being essentially undistinguished in most cases. Illustration in the next section is based on the Poisson proposal version.

Theoretical assurance that the MCMC algorithm so defined converges—that is, ultimately generates samples from the true joint posterior $p(\mathbf{X}, \boldsymbol{\Lambda}|\mathbf{Y})$—follows if we can determine that the Markov chain is irreducible. This reduces to the question of whether or not a current value $(\mathbf{X}, \boldsymbol{\Lambda})$ can "move" to any other point in the joint parameter space following a finite number of iterations of the scheme a and b. For the elements of $\boldsymbol{\Lambda}$, there is no problem, assuming continuous priors with fixed support. But for the $\mathbf{X}$, the support of the key conditional posteriors (6) depends on resampled values of elements of $\mathbf{X}_2$ and so changes with iterations; this complicates this issue somewhat. It can be shown, however, that in fact $\mathbf{X}_2$ is free to move arbitrarily around its parameter space in consecutive iterations, despite the support constraints and complications (details omitted here). Thus the resulting chain is irreducible, and convergence is assured.

3.2 Illustration

In simple networks, such as that of Figure 3, exact posterior distributions may be computed; this has been used to validate software implementing the MCMC analysis. We display here some summary inferences for the larger example network of Figures 1 and 2.

Taking $\boldsymbol{\Lambda} = \{1, \ldots, 12\}$, we simulated route counts $\mathbf{X}$ from the resulting independent Poisson distributions, producing $\mathbf{X} = (2, 2, 0, 8, 5, 7, 6, 4, 9, 7, 17, 11)'$; these lead to the data values of link counts $\mathbf{Y} = (2, 28, 9, 5, 34, 16, 35)'$. The MCMC scheme is initialized with values of $\mathbf{X}_2$ chosen by inspection of the observed link flows, an easy task in this small network but one that will take more work in larger and more complex cases. Generally, identifying suitable initial values will involve coming up with a nonnegative solution to an underdetermined system of equations, which can be easily set up to automatically generate ranges of possible solutions and hence possible starting values. For example, using Mathematica, we can automatically find a (possibly negative) solution to the underdetermined system $\mathbf{Y} = \mathbf{A}\mathbf{X}$ and the basis for the null space associated to the matrix $\mathbf{A}$. Playing with different linear combinations of the first with the last, we were able to easily determine starting values for our vector $\mathbf{X}$ (actually a number of them), so that we could start our simulation from a widely ranging set of points. It should be remarked that integer programming methods also could provide an approach to assigning initial values, although we have not as yet pursued this formally. Notice that we are assuming observations without error for the link counts, so we rely on the existence of at list one nonnegative solution. This ensures, in cases other than trivial ones, the existence of other possible nonnegative solutions.

In our analysis we reinitialized and reran the MCMC simulation from these several starting values for $\mathbf{X}_2$, to assess and validate the rapid convergence experienced, whatever the starting values. Our reported analysis uses finite uniform priors for the λ_a with common range $(0, L)$ for an appropriately large upper bound L whose value need not be specified. The analysis summarized is based on the Metropolis–Hastings variant with Poisson proposals for the X_a, based on the current values of the respective λ_a. The reported results, and questions of convergence and efficiency, have been explored in various ways. In particular, and most convincing, we have repeated the analysis using several variants of the MCMC scheme; one uses proposal distributions that are uniform on the earlier mentioned "large support," another has uniform proposals on the current support for each X_a conditional on the remaining values $\mathbf{X}_{2,-a}$, and a third reanalysis uses "exact" Gibbs sampling in which the full conditional multinomial distributions of each X_a are evaluated exactly and sampled at each stage. These analyses are in consistent agreement and validate the summary inferences reported using the Poisson proposal. Across all Metropolis–Hastings variants, the rates of acceptance were satisfactorily between 23% and 40% for the different components of vector $\mathbf{X}_2$, the differences primarily reflecting the varying degrees of latitude arising in identifying appropriate bounds on the support of proposal distributions.

The aforementioned assurances on convergence of the MCMC algorithms are complemented by positive results from various standard convergence investigations as implemented in CODA (Best, Cowles, and Vines 1995). Following this, we summarize analysis by running the MCMC algorithm for a total 50,000 iterations and computing posterior estimates based on only 5,000 values. The full 50,000 iterations are performed rapidly on this small network, running in a real time of 2–2.5 minutes on an otherwise idle, standard desktop workstation (DEC AlphaStation 200). How computing time, as well as convergence characteristics, scale up to larger networks is the subject of current investigations. Figures 4 and 5 display approximate posteriors for $\boldsymbol{\Lambda}$ and $\mathbf{X}$. For each λ_a, we compute the posterior density approximation via the Monte Carlo average of known conditional distributions. Because $p(\lambda_a|\mathbf{X}, \mathbf{Y}) = p(\lambda_a|X_a)$ is simply $Ga(X_a + 1, 1)$ truncated to $(0, L)$, we have

$$
\begin{aligned}
p(\lambda_a|\mathbf{Y}) &= \sum_{X_a} p(\lambda_a|X_a)p(X_a|\mathbf{Y}) \\
&\approx m^{-1}\exp(-\lambda_a)\sum_{i=1}^{m} \lambda^{X_a^{(i)}}/X_a^{(i)}!,
\end{aligned}
$$

where $m = 5,000$, the Monte Carlo sample size, and $\{X_a^{(i)}, i = 1, \ldots, m\}$ is the set of sampled values of X_a. These densities appear as full lines in Figure 4. As we are dealing with a simulated dataset and know the true underlying values of each X_a, we can compute the truncated gammas $p(\lambda_a|X_a)$ for comparison; these appear as the dashed lines in Figure 4. The differences between these and the full posteriors are indicative of the uncertainties about the λ_a due to lack of precise knowledge of the X_a. In the cases of $a = 1$ and $a = 5$, the structure of the network implies $X_1 = Y_1 = 2$ and $X_5 = Y_4 = 5$; hence there is no uncertainty about these two X_a values, so the posteriors coincide, $p(\lambda_a|X_a) = p(\lambda_a|\mathbf{Y})$. Figure 5 displays histograms of the 5,000 posterior samples of the X_a, representing the marginal posterior $p(X_a|\mathbf{Y})$. Note the degenerate posteriors for X_1 and X_5, as these two values are known. Otherwise, full marginal uncertainties about the route flows are reflected in these histogram approximations to posterior densities; the underlying X_a values from the simulated data are marked as triangles on the lower axis, for comparison.

4. THE MONROE NC TRANSPORTATION NETWORK

Estimation, and prediction, of OD flows in real road networks is a problem central to policy-oriented studies of a range of traffic engineering problems, ranging from real-time traffic management to implementation of congestion-pricing mandates (Sheffi 1985). The archetype context is one of updating estimates of OD flows; based on "today's" initial OD flow estimates and link traffic since observed, make forecasts of "tomorrow's" OD flows to feed into traffic management and decision systems, including network equilibrium models that are central tools in transportation modelling and policy studies (Berka and Boyce 1994; Sheffi 1995; Smith 1987; West 1994). This problem has an inherent Bayesian flavor, and although most approaches have been quite ad hoc, some have adopted basic Bayesian methods (Maher 1983). One interest here is to assess the efficacy of the full Bayesian solution in the context assumed earlier on a real network. From transportation engineers at North Carolina State University working with the National Institute of Statistical Sciences, we have a section of a road network from the township of Monroe in North Carolina. This network is structured so that all OD pairs have just one route between them and so falls under the fixed rout-

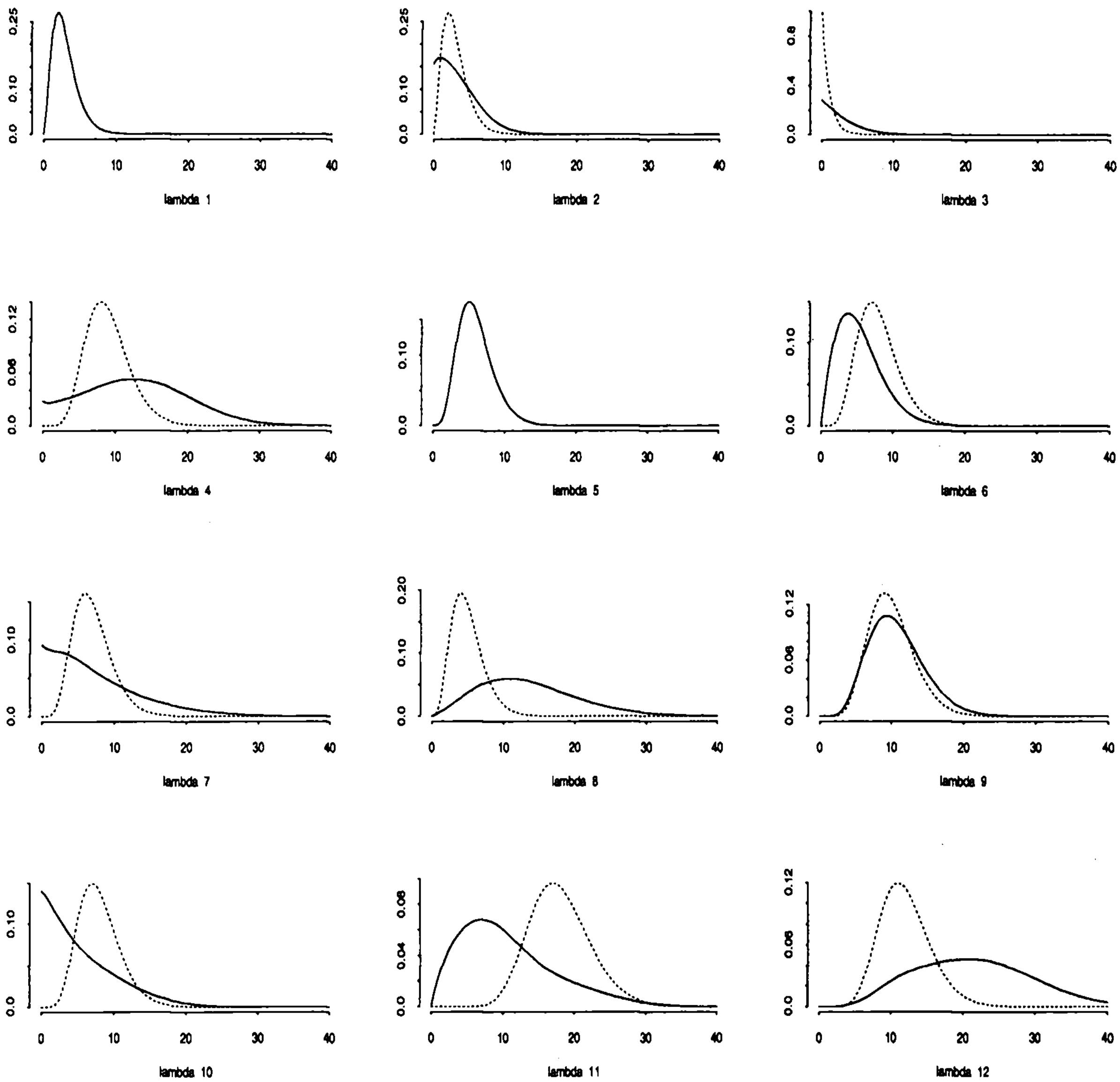

Figure 4. Posterior Distributions for the 12 Components of Λ, as Full Lines. – – –, Conditional posteriors were the actual route counts **X** *known.*

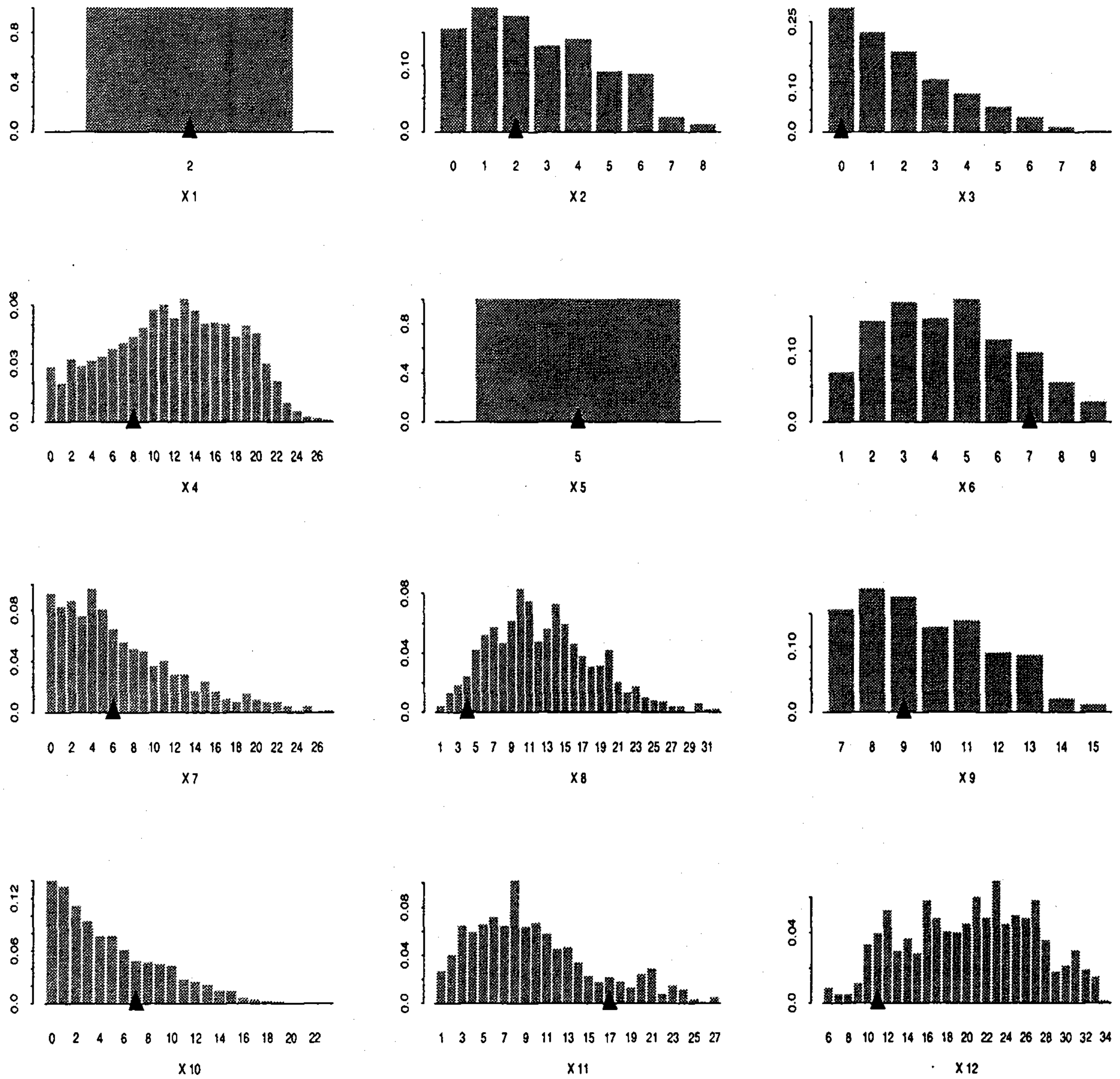

Figure 5. Posterior Distributions for the 12 Components of **X**. *The actual values are marked by the triangles on the axes.*

ing assumptions of our model. The network is schematically represented in Figure 6. This network has been adapted into the operating "real-time" simulation system *Integration* (Van Aerde et al. 1996); this system is extensively used in studying effects on traffic networks of potential network and road traffic policy changes. Based on previous studies of the Monroe road system, OD flows are generated to "load" the network which runs through a specified time period of the day and are subject to the usual delays due to traffic congestion, traffic lights, left and right turns, merging of lanes, and so on. All aspects of road traffic are continuously monitored and recorded. In particular, observed link counts are recorded on all links in the network. Here we report on some aspects of analyses of observed link counts over a 2-hour morning period; this analysis illuminates application of our models and generates additional insights into the structure and difficulties of the OD estimation problem in a real context but one in which we have the true OD flows to compare with model-based inferences.

In this network, most pairs of the nodes at the boundaries constitute OD pairs; the exceptions are pairs (M, G), (M, L), (G, M), (G, L), (N, I), (L, M), (L, G), and (L, N). Observed link counts in the 2-hour period appear in Figure 7. After deleting four of the original 24 link count observations to eliminate observed linear dependencies, the resulting matrix **A** is 20×64. The problem thus involves the 44-dimensional vector $\mathbf{X}_2$ and the 20-dimensional vector $\mathbf{X}_1$ in the earlier notation.

Our first analysis is exactly as detailed in Sections 2 and 3, using independent uniform priors for the OD rates λ_i. Note that in a real "on-line" application, observation and analysis from previous days would generate information on which to base more informative priors. This turns out to be absolutely critical in resolving the inherent tendencies toward biased inferences discussed in Section 2.4. As discussed there, this problem is ever-present but is exacerbated in cases of "disbalanced" flows; notice the wide variability in observed link counts here, indicative of comparable diversity among the underlying OD flows of interest. In such

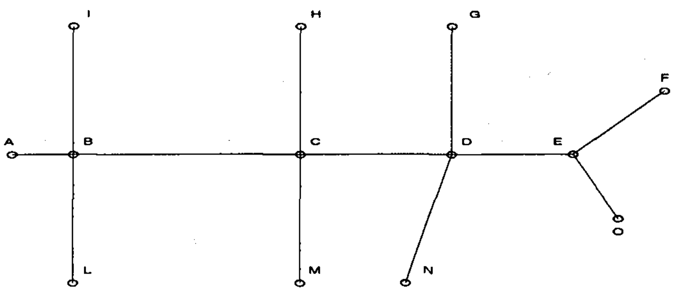

Figure 6. Physical Node-Link Structure of the Monroe Network.

cases analysis may be subject to significant biases in inferences unless it is appropriately constrained via informed prior distributions.

We summarize some basic results of this first analysis using uniform priors for the λ_i. We ran the MCMC analysis from starting values computed as described in Section 3, for a large number of iterations in view of the unavoidably high dependencies among OD flows and rates link. After a number of experiments, we ran a final chain for 1 million iterations, and summarize for posterior inferences here a "dependence-breaking" subsample of size 10,000. Across repeat experiments and with different subsamples, the results are consistent. Figure 8 summarizes marginal posteriors for 16 of the full 64 OD flows. Because in this case we actually know the realized OD values, we can assess the accuracy of posterior inferences by comparing the true values, denoted by X_i^* for OD pair i, against the margins; the X_i^* are indicated in the figures. The key point to note here is just how poorly several of the smaller X_i^* values are estimated; they lie way down in the lower tail of their marginal posterior distributions and so are grossly overestimated. The higher flows, in comparison, are consistently adequately estimated. This is an example of the general phenomenon discussed in Section 2.4 and is exhibited across repeat analyses of other data from the Monroe network, taken from differing time periods, and also of simulated data. The biased inferences are due to the inherent structural ambiguity in the likelihood function for the Poisson rates.

Now consider reanalysis using more informed priors for the λ_i. We now take independent gamma priors; note that the analysis of Sections 2 and 3 can be developed with minor modification under (conditionally conjugate) gamma priors, and so the easy details are omitted. To mirror an on-line, OD flow "updating" context, we base the priors for "today's" X_i on the known values X_i^*; the idea here is that the numbers X_i^* represent estimate based on previous days analyses and observations. Specifically, we choose the prior gamma distribution for λ_i to have shape parameter aX_i^* and scale parameter a for some $a > 0$. Small values of a discount the prior estimate X_i^* and lead to a relatively diffuse prior. One analysis summarized here is based on $a = .02$. For each of the 16 OD pairs chosen in the foregoing analysis, Figure 9 graphs the corresponding gamma prior densities (as dashed lines). Following MCMC-based analysis with these priors, the estimates of corresponding posteriors are computed and also graphed in Figure 9. Figure 10 displays the corresponding marginal posteriors for the OD flows X_i, again with true values X_i^* marked. Evidently, even very weak prior information, roughly "correctly" located based on previous OD estimates, is sufficient to overcome the distortions and biases inherent in "disbalanced" networks. Figure 11 summarizes all marginal posterior distributions for all 64 OD flows in terms of box plots. Boxplots are graphed for the two analyses: uniform priors and gamma priors on the λ_i. Superimposing the true X_i^* values, we note uniform consistency of the data with the priors in the latter analysis and the corresponding uniform correction of the overestimation bias for smaller flows. We also indicate (with the symbol "V") the point estimates delivered using the algorithm of Vardi (1996) on this network. As is clear from the figure, this algorithm, though not a direct likelihood-based algorithm, suffers precisely the same problem of overestimating low flows in the context of dominating high flow rates on subsets of the network and should be used with caution unless explicitly adjusted to overcome this problem.

5. RANDOM (MARKOVIAN) ROUTING

One important model extension relaxes the assumption that all messages or trips between a specified OD pair take the single route specified in the 0/1 routing matrix **A**. Vardi (1996) discussed this and developed the case of Markovian routing, in which a message travelling between a specified OD pair exits its "current" node in the network on one of possibly several links consistent with a route to the specified destination. In communications networks, such random routing may arise as a result of control procedures to avoid or redress queuing questions. In urban road transportation networks, there typically exist one or a small number of primary routes between specified OD pairs, such routes being used by much of the traffic, but with several additional,

Y_1 :	$A \to B$:	1980
Y_2 :	$B \to C$:	1966
Y_3 :	$C \to D$:	1788
Y_4 :	$D \to E$:	2600
Y_5 :	$E \to F$:	2100
Y_6 :	$F \to E$:	1816
Y_7 :	$E \to D$:	2880
Y_8 :	$D \to C$:	2052
Y_9 :	$C \to B$:	1954
Y_{10} :	$B \to A$:	1772
Y_{11} :	$I \to B$:	338
Y_{12} :	$L \to B$:	284
Y_{13} :	$H \to C$:	306
Y_{14} :	$M \to C$:	176
Y_{15} :	$G \to D$:	68
Y_{16} :	$N \to D$:	1000
Y_{17} :	$O \to E$:	1104
Y_{18} :	$B \to I$:	488
Y_{19} :	$C \to M$:	274
Y_{20} :	$D \to N$:	926

Figure 7. Link Flows for the Monroe Network of Figure 6.

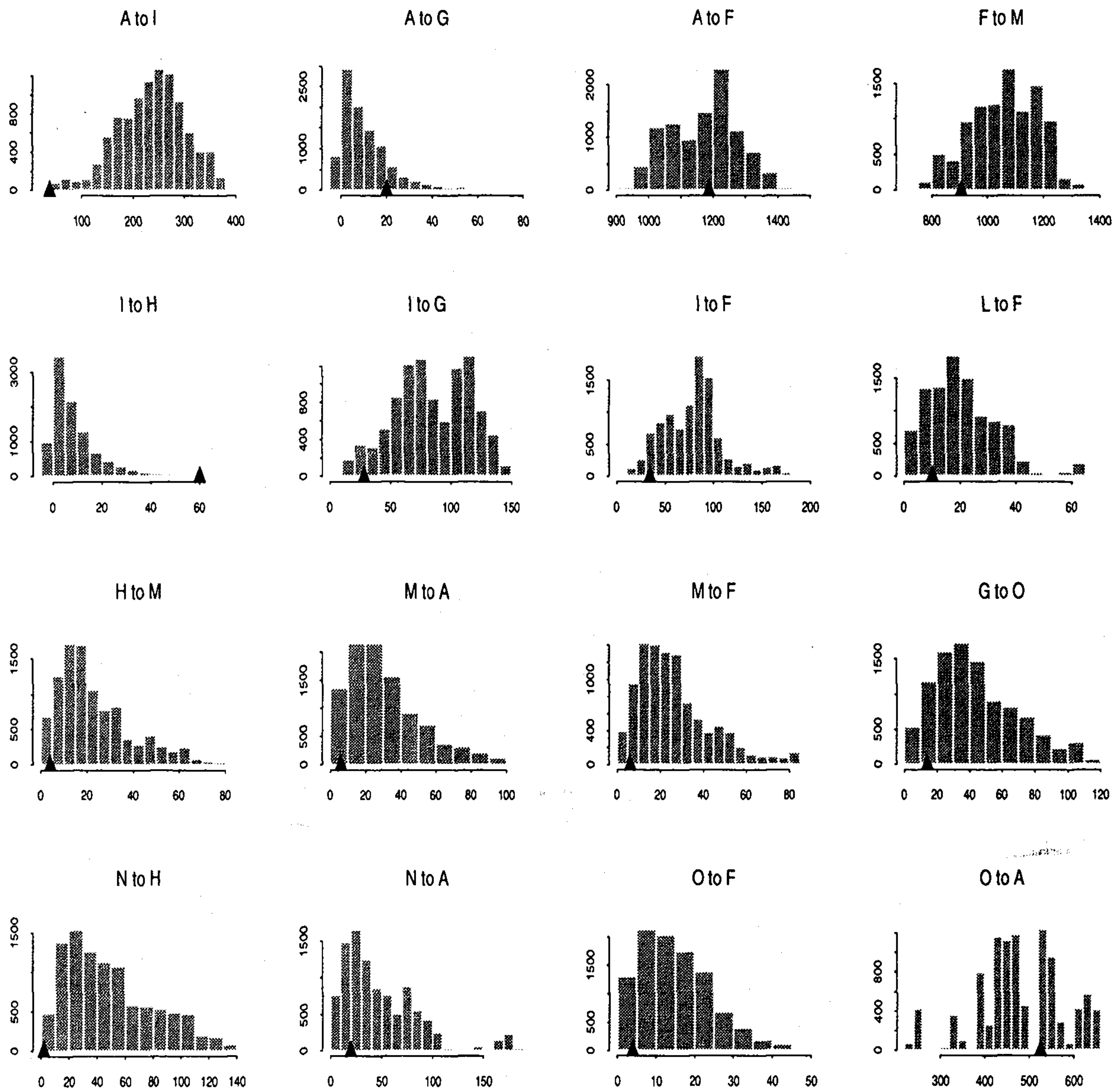

Figure 8. Posterior Distributions for 16 Components of **X** *Under Uniform Priors on the* λ_i. *The true values* X_i *appear on the x axis.*

secondary routes used less frequently either as alternatives in times of congestion on primary routes or as occasional alternatives for other reasons. Primary and secondary routes between a specified OD pair will typically share a common subset of links, with a few additional links being specific to one or a few routes (Figure 12).

A neat extension of the development in Sections 2 and 3 provides for analysis using the Bayesian approach developed for the fixed routing network. We discuss this later, following a precise definition and discussion of the random routing model.

Given a specific OD pair $a = (i, j)$, messages leaving node i may now take differing routes to node j. The Markovian routing model simply assumes that at any node on the way to the destination, each message exits on a link determined by a set of link choice probabilities, independently of the path taken to the current node and independently across messages. In some cases there will be just one exit link possible, in other cases there may be several. As in Vardi (1996), the setup can be summarized through a modified routing matrix that summaries the link choice probabilities for all possible OD pairs. As in the fixed routing case, the matrix has rows indexing links in the network and columns indexing OD pairs; now, however, the entries are probabilities determining the selection of links (rows) on trips between specified OD pairs (columns). We study an example matrix from Vardi (1996) further later; with rows and columns labelled by links and OD pairs, this has the probability matrix **A** given in Figure 6.

Here we have 4 nodes, $r = 9$ directed links and $c = 12$ OD pairs. Consider, for example, the OD pair AB. Each trip from A to B has an 80% chance of moving directly along link $A \to B$ and terminating there, with the complementary 20% chance it travels from $A \to C$. Assuming that it follows this latter path, it then travels along $C \to B$ directly and terminates with an 80% chance; otherwise, it moves along the two consecutive links $C \to D \to B$ and terminates. By elementary probability computations, we can

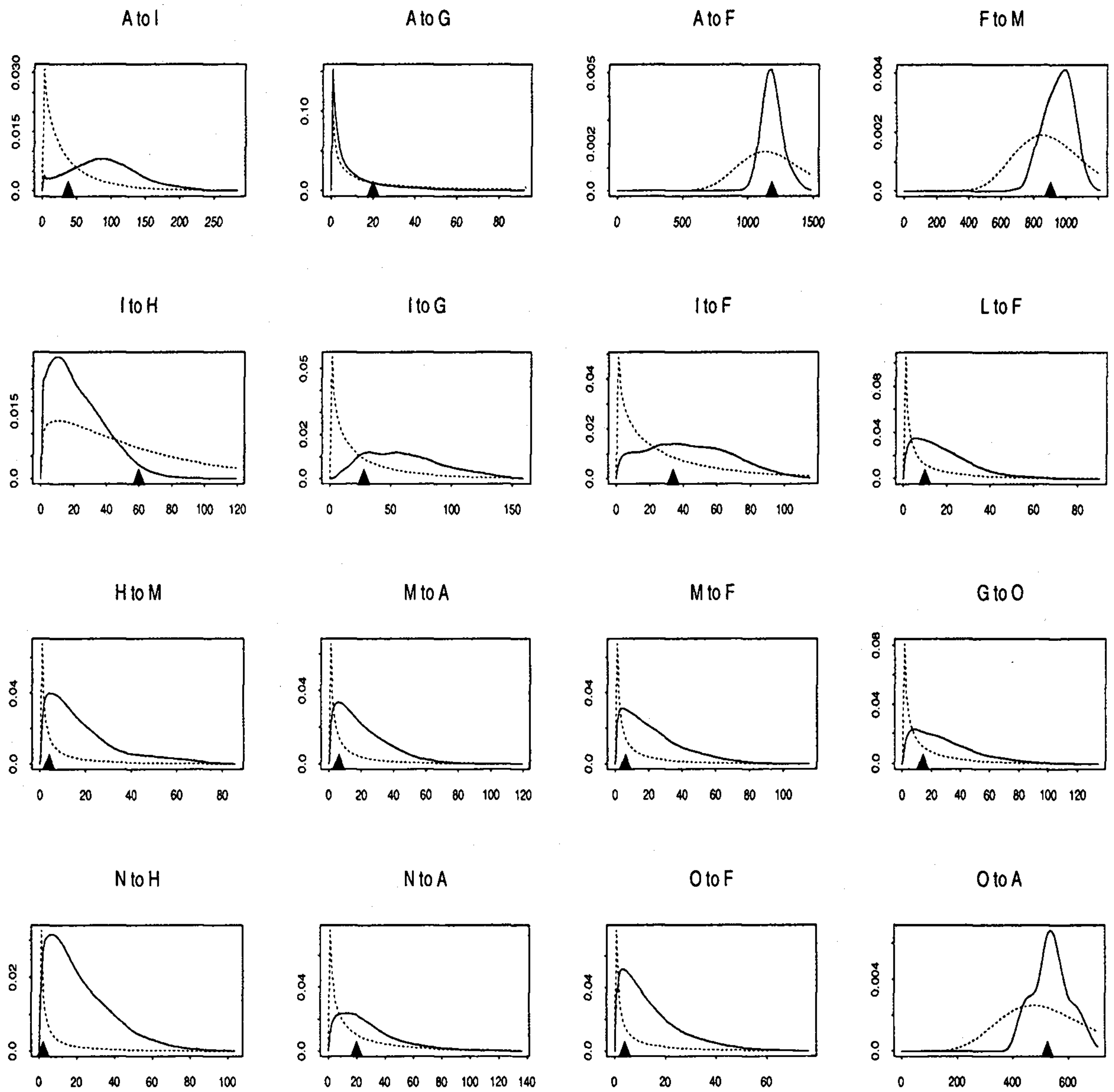

Figure 9. Marginal Prior (- - -) and Posterior (——) Densities for 16 Components of Λ. The true values X_i appear on the x axis.

deduce various interesting marginal and conditional probabilities, such as the probability that a message between a specified OD pair j passes through any given link i (or through a specified pair or subset of links).

Now consider the questions of inference about OD counts $\mathbf{X}$ based on observed link counts $\mathbf{Y}$, as earlier. Retain the same modeling assumptions, so that the X_a are conditionally independent $X_a \sim Po(\lambda_a)$ with specified independent priors on the rates λ_a. Our approach is based on the idea of embedding the random routing problem in a fixed routing problem on an artificial "super-network" and then applying the theory and methods of Sections 2 and 3 to the latter. To motivate this, consider the matrix of Figure 6 and focus on the single OD pair AB. As earlier, trips from A to B travel one of three routes: $A \to B, A \to C \to B$, or $A \to C \to D \to D$, with corresponding marginal probabilities .8, .16, and .04. So we can view the pair AB as comprising three artificial OD pairs corresponding to these three distinct routes. The X_1 trips originating at A and travelling to B can be viewed as initially assigned to one of these three routes according to the marginal route selection probabilities .8, .16, and .04. Once all assignments are made, the subsets of the X_1 trips travel their allocated routes, and we have a framework of fixed routing.

This leads to the general approach to constructing a super-network with fixed routing. For each column of the probabilistic routing matrix, we identify all possible routes between the corresponding OD pair, and simply list these as a subset of possible fixed routes; the corresponding probabilities assigned to these routes are easily computed (Figure 13). Then we create a new 0/1 fixed routing matrix that has a column for each of these new fixed routes (Figure 14). This can be easily automated. (A simple S-PLUS function is available from the first author.) In the earlier example, the 9×12 probabilistic routing matrix in Figure 2 generates the extended fixed routing matrix in Figure 7. There, in addition to labelling groups of columns with their

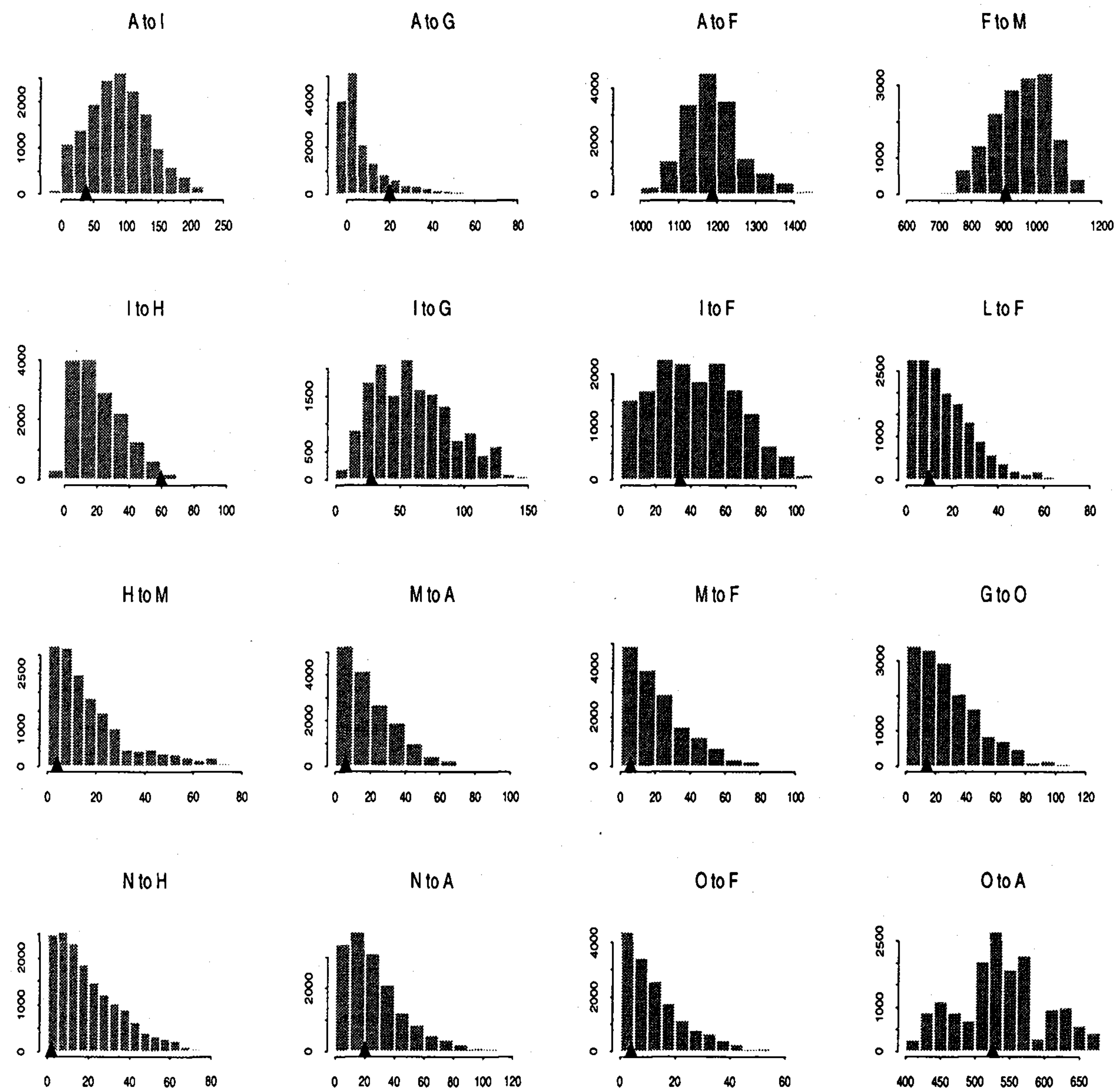

Figure 10. Posterior Distributions for 16 Components of **X** *Under Informed Gamma Priors on the* λ_i*. The true values* X_i *appear on the x axis.*

common OD pair, we add the probabilities of assignment to the columns (i.e., fixed routes) within such subsets. In this case, the original 12 OD pairs generates 27 in the super-network.

Now turn to the issue of modeling and inferring route counts. Consider any OD pair $a = (i, j)$ in the original network, with corresponding counts X_a. The process of creating a corresponding set of fixed routes between i and j generates some number, say k_a, of such routes. Given the original probability routing matrix, we can trivially compute the resulting probabilities $\mathbf{p}_a \overset{\text{def}}{=} (p_{a,1}, \ldots, p_{a,k_a})'$ of selecting each of these fixed routes; note that these k_a probabilities sum to unity. Write $X_{a,t}$ for the number of trips that take fixed route t from i to j. Then, given X_a, the disaggregated counts $X_{a,1}, \ldots, X_{a,k_a}$ are conditionally multinomially distributed among the k_a fixed routes out of the total X_a and with route selection probabilities $\mathbf{p}_a$. Under the earlier assumption that $X_a \sim Po(\lambda_a)$, this trivially implies that the disaggregated, fixed route counts are themselves marginally independent and Poisson distributed, with $X_{a,t} \sim Po(p_{a,t}\lambda_a)$ for $t = 1, \ldots, k_a$. Hence, independently across OD pairs a, we have

$$p(X_{a,1}, \ldots, X_{a,k_a} | \lambda_a, \mathbf{p}_a) = \prod_{t=1}^{k_a} \frac{(p_{a,t}\lambda_a)^{X_{a,t}} \exp(-p_{a,t}\lambda_a)}{X_{a,t}}, \quad (7)$$

which reduces to

$$p(X_{a,1}, \ldots, X_{a,k_a} | \lambda_a, \mathbf{p}_a) \propto \lambda_a^{X_a} \exp(-\lambda_a) \prod_{t=1}^{k_a} p_{a,t}^{X_{a,t}} \quad (8)$$

as a function of $(\lambda_a, \mathbf{p}_a)$, noting that $X_a = \sum_{t=1}^{k_a} X_{a,t}$.

The immediate consequence of (7) and (8) is that the original approach to inference about route counts in the fixed routing problem in Sections 2 and 3 applies directly to the super-network. Conditional on the parameters $\mathbf{\Lambda}$ and

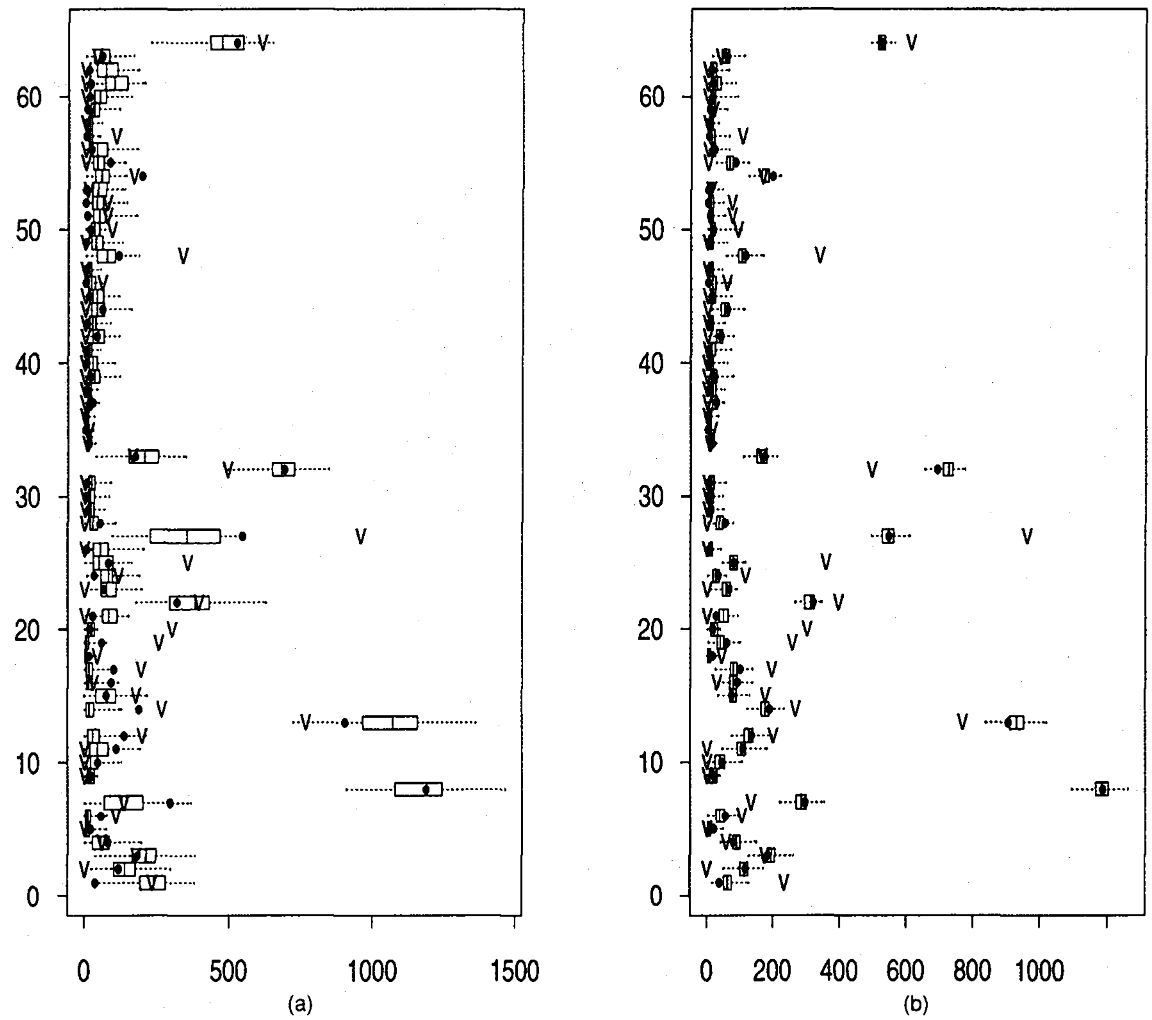

Figure 11. Boxplots of Marginal Posteriors for all 64 Elements of **X** *in the Monroe Network. Separate displays are given for analysis based on uniform priors (a) on the λ_i, and based on informed but very diffuse gamma priors (b). The X_i^* are marked as "•," and the point estimates of Vardi's algorithm as "V."*

each of the $\mathbf{p}_a$ vectors, we simply have an expanded network with implied route counts for each of the $\sum_{a=1}^{c} k_a$ fixed routes. The compatibility of the independent Poisson models for the priors on route counts implies that the construction of conditional posterior distributions for route counts is structurally unchanged. Hence the components of the MCMC analysis for simulating route counts apply and will produce samples of the full set of route counts,

$$\mathbf{X} = \{X_{a,t}, t = 1, \ldots, k_a; a = 1, \ldots, c\}. \quad (9)$$

From these we can trivially deduce the totals X_a for each OD pair a.

Now turn to inference on the Poisson rates. It is clear that, conditional on all route counts $\mathbf{X}$ in (9), the likelihood function for the underlying rates $\mathbf{\Lambda}$ is just the function (8). This is, again, of exactly the form arising in the original fixed routing problem, and so the construction of posterior samples for $\mathbf{\Lambda}$ follows that development.

This discussion is all conditional on known and fixed routing probabilities, and hence known and fixed vectors $\mathbf{p}_a$ for each OD pair a. We note that extensions to incorporate inference on these probabilities are essentially direct. Note that, from (8), we obtain a conditional likelihood function for each of the $\mathbf{p}_a$ vectors given imputed route counts in the super-network and the λ_a. On this basis, the likelihood function factorizes into a set of c components of the form in (8), and so the iterative simulation analysis is trivially generalized by linking in a component to sample each $\mathbf{p}_a$ from the corresponding posterior distributions.

	1 *AB*	2 *AC*	3 *AD*	4 *BA*	5 *BC*	6 *BD*	7 *CA*	8 *CB*	9 *CD*	10 *DA*	11 *DB*	12 *DC*
1 $A \to B$	.8	.2	.2	0	0	0	0	0	0	0	0	0
2 $A \to C$	.2	.8	.8	0	1	1	0	0	0	0	0	0
3 $B \to A$	0	0	0	1	.2	.1	1	0	0	1	0	0
4 $B \to C$	0	.8	0	0	.8	.1	0	0	0	0	0	1
5 $B \to D$	0	.2	1	0	0	.8	0	0	1	0	0	0
6 $C \to B$	.8	0	.2	0	0	0	.8	.8	.2	1	1	0
7 $C \to D$	.2	0	.8	0	0	1	.2	.2	.8	0	0	0
8 $D \to B$	1	0	0	0	0	0	1	1	0	.8	.8	.2
9 $D \to C$	0	1	0	0	0	0	0	0	0	.2	.2	.8

Figure 12. An Example of a Probability Routing Matrix. As in the fixed routing case, rows and columns correspond to directed links and OD pairs. The entries are now conditional probabilities of traversing the link (row) during a trip between the OD pair (column).

	1			2			3			4	5	
	AB			AC			AD			BA	BC	
	.8	.16	.04	.16	.04	.8	.2	.16	.64	1	.2	.8
1 $A \to B$	1	0	0	1	1	0	1	0	0	0	0	0
2 $A \to C$	0	1	1	0	0	1	0	1	1	0	1	0
3 $B \to A$	0	0	0	0	0	0	0	0	0	1	1	0
4 $B \to C$	0	0	0	1	0	0	0	0	0	0	0	1
5 $B \to D$	0	0	0	0	1	0	1	1	0	0	0	0
6 $C \to B$	0	1	0	0	0	0	0	1	0	0	0	0
7 $C \to D$	0	0	1	0	0	0	0	0	1	0	0	0
8 $D \to B$	0	0	1	0	0	0	0	0	0	0	0	0
9 $D \to C$	0	0	0	0	1	0	0	0	0	0	0	0

Figure 13. Super-Network Fixed Routing Matrix for the Stochastic Network Problem of Figure 6. Here we display just the first 12 columns corresponding to the first 5 OD pairs.

Assumptions about prior distributions will depend on context, but no essential difficulties arise. Further details will be explored and reported elsewhere.

6. SUMMARY COMMENTS

We have elaborated on the structure of both fixed and random (Markovian) routing problems introduced and analyzed by Vardi (1996). In the fixed routing context, our theoretical development has elaborated on aspects of the structure of the network count inference problem that underpin the approach to Bayesian analysis based on MCMC simulation. Our contributions include the isolation of deficiencies in likelihood and other non-Bayesian approaches to inference in these problems due to structural indeterminacies that may lead to systematic and significant biases in inferences in some applications. We demonstrate how this may be remedied through the use of informed prior distributions that provide appropriate "constraints" and note that this strategy is perfectly consistent with current practices in "updating" OD flow estimates in transportation networks (Maher 1983; McNeil and Hendrickson 1985). Other existing algorithms, including that of Vardi, run the unavoidable risk of overestimating low flows on OD routes that share links with OD routes experiencing high flows. The Bayesian approach here naturally, and satisfactorily, deals with this problem and forms a natural extension of some of the traditional, ad hoc methods used in traffic engineering solutions.

	6			7		8		9		10		11		12	
	BD			CA		CB		CD		DA		DB		DC	
	.1	.1	.8	.8	.2	.8	.2	.2	.8	.8	.2	.8	.2	.2	.8
1 $A \to B$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2 $A \to C$	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3 $B \to A$	1	0	0	1	1	0	0	0	0	1	1	0	0	0	0
4 $B \to C$	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0
5 $B \to D$	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0
6 $C \to B$	0	0	0	1	0	1	0	1	0	0	1	0	1	0	0
7 $C \to D$	1	1	0	0	1	0	1	0	1	0	0	0	0	0	0
8 $D \to B$	0	0	0	0	1	0	1	0	0	1	0	1	0	1	0
9 $D \to C$	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1

Figure 14. The Final 15 Columns, Corresponding to the Final 7 OD Pairs, of the Fixed Routing Matrix for the Super-Network Representation of the Stochastic Network Problem of Figure 6.

The embedding of the Markovian routing problem in an elaborated fixed routing context enables similar analyses to be performed there, with evident potential to extend to problems of inference on route choice probabilities. This latter context is very relevant in urban transportation networks, and some developments for application in that area are envisaged. It is clear from the development that the Bayesian view clarifies the structure of these specific network inference problems, and enable relatively direct and easily implementable solutions. Several variations of the specific model assumptions, including the questions of prior models for the underlying Poisson rates $\mathbf{\Lambda}$, have been mentioned and will represent no substantial additional technical nor computational complications.

The development assumes fully observed link counts in the fixed time period. This is trivially relaxed. Note that were we to observe counts on only a subset of links, the problem remains essentially the same but with the corresponding rows of the routing matrix $\mathbf{A}$ deleted. In transportation networks, automated traffic detectors, including in-road loop detectors and video cameras, typically monitor traffic on very sparse subsets of key links in large road networks, so this kind of issue is relevant. In large networks, it will be important to focus attention on a subset of likely key routes (i.e., a subset of OD pairs), in addition to a subset of links. Model modifications to partition large-scale networks hierarchically into subsets of smaller, manageable networks and develop inference on route counts and flows based on partial network structure, are then of interest. To our knowledge, no such developments currently exist.

Additional extensions include questions of monitoring link traffic in several time periods in sequence, and taking into consideration nonnegligible transit time over links. We are currently studying link count data from a Seattle highway network as part of a current study of network structure underway at the National Institute of Statistical Sciences (NISS), and will report results of that study elsewhere.

APPENDIX: CONVEXITY AND IRREDUCIBILITY

We prove convexity of the support of the unidimensional conditional posterior distributions, and connectedness of the support of the joint posterior, of the vector $\mathbf{X}_2$ in the fixed routing analysis of Sections 2 and 3.

First, consider the conditional distribution $p(X_{2,a}|\mathbf{X}_{2,-a})$ for any $a = 1, \ldots, n-k$. That the corresponding support is convex is proved as follows.

Suppose that two $(n-k)$-dimensional vectors $\mathbf{x}, \mathbf{y}$ differ in just one of their components; that is, $\mathbf{x} = (x_1, x_2, \ldots, x_i, \ldots, x_{n-k})'$ and $\mathbf{y} = (x_1, x_2, \ldots, y_i, \ldots, x_{n-k})'$. Suppose also that for any vector $\mathbf{Y}$, the inequalities $\mathbf{A}_1^{-1}(\mathbf{Y} - \mathbf{A}_2\mathbf{x}) \geq 0$ and $\mathbf{A}_1^{-1}(\mathbf{Y} - \mathbf{A}_2\mathbf{y}) \geq 0$ hold. For any number $\alpha \in [0,1]$, define the vector

$$\mathbf{z} = \alpha\mathbf{x} + (1-\alpha)\mathbf{y} = (x_1, x_2, \ldots, \alpha x_i + (1-\alpha)y_i, \ldots, x_{n-k})'.$$

Then $\mathbf{z}$ also satisfies $\mathbf{A}_1^{-1}(\mathbf{Y} - \mathbf{A}_2\mathbf{z}) \geq 0$. In fact,

$$\begin{aligned}
&\mathbf{A}_1^{-1}(\mathbf{Y} - \mathbf{A}_2\mathbf{z}) \\
&\quad = \mathbf{A}_1^{-1}(\mathbf{Y} - \mathbf{A}_2(\alpha\mathbf{x} + (1-\alpha)\mathbf{y}))
\end{aligned}$$

$$= \mathbf{A}_1^{-1}(\alpha\mathbf{Y} + (1-\alpha)(\mathbf{Y}) - \mathbf{A}_2(\alpha\mathbf{x} + (1-\alpha)\mathbf{y}))$$

$$= \mathbf{A}_1^{-1}\alpha(\mathbf{Y} - \mathbf{A}_2\mathbf{x}) + \mathbf{A}_1^{-1}(1-\alpha)(\mathbf{Y} - \mathbf{A}_2\mathbf{y}).$$

This is the mean of two nonnegative quantities by assumption, and so is itself a nonnegative quantity. This ends the proof.

We now show that the support of the $n-k$ dimensional vector $\mathbf{X}_2$ is fully connected, by ruling out the possibility that two or more of its subsets, even if marginally convex, can be separated. The result implies that the MCMC analysis can "smoothly" move around the support, and that the chain is irreducible. Begin with the simplified case of a two-dimensional vector $\mathbf{X}_2 = (X_1, X_2)'$. The proof is by contradiction, showing that a configuration such as displayed in Figure A.1 is not consistent with our fundamental assumptions: that is, if we assume that both $\mathbf{X}_2$ and $\mathbf{X}_2 + (1,1)'$ are feasible, then either $\mathbf{X}_2 + (1,0)$ is feasible as well, or $\mathbf{X}_2 + (0,1)$ is, if not both.

In fact, the two pairs of assumptions

$$\{\mathbf{A}_1^{-1}(\mathbf{Y} - \mathbf{A}_2\mathbf{X}_2) \geq 0, \mathbf{A}_1^{-1}(\mathbf{Y} - \mathbf{A}_2(\mathbf{X}_2 + (1,1)') \geq 0\}$$

and

$$\{\mathbf{A}_1^{-1}(\mathbf{Y} - \mathbf{A}_2(\mathbf{X}_2 + (1,0)') < 0, \mathbf{A}_1^{-1}(\mathbf{Y} - \mathbf{A}_2(\mathbf{X}_2 + (0,1)') < 0\}$$

are incompatible. To begin, note that $\mathbf{A}_1^{-1}(\mathbf{Y} - \mathbf{A}_2(\mathbf{X}_2 + (1,1)')$ ≥ 0 reduces to

$$\mathbf{A}_1^{-1}(\mathbf{Y} - \mathbf{A}_2\mathbf{X}_2 - (A_2^{11}, A_2^{21})' - (A_2^{12}, A_2^{22})') \geq 0,$$

where the matrix elements are denoted by double superscript. Now add $\mathbf{A}_1^{-1}(\mathbf{Y} - \mathbf{A}_2\mathbf{X}_2) \geq 0$ by assumption, to lead to

$$\mathbf{A}_1^{-1}(\mathbf{Y} - \mathbf{A}_2\mathbf{X}_2) + \mathbf{A}_1^{-1}(\mathbf{Y} - \mathbf{A}_2\mathbf{X}_2 - (A_2^{11}, A_2^{21})' - (A_2^{12}, A_2^{22})') \geq 0.$$

This reduces to

$$\mathbf{A}_1^{-1}(\mathbf{Y} - \mathbf{A}_2\mathbf{X}_2 + \mathbf{Y} - \mathbf{A}_2\mathbf{X}_2 - (A_2^{11}, A_2^{21})' - (A_2^{12}, A_2^{22})') \geq 0$$

or

$$\mathbf{A}_1^{-1}(\mathbf{Y} - \mathbf{A}_2\mathbf{X}_2 - (A_2^{11}, A_2^{21})' + \mathbf{Y} - \mathbf{A}_2\mathbf{X}_2 - (A_2^{12}, A_2^{22})') \geq 0.$$

This last term then cannot be the sum of two negative quantities, and this establishes the contradiction between the two pairs of inequalities.

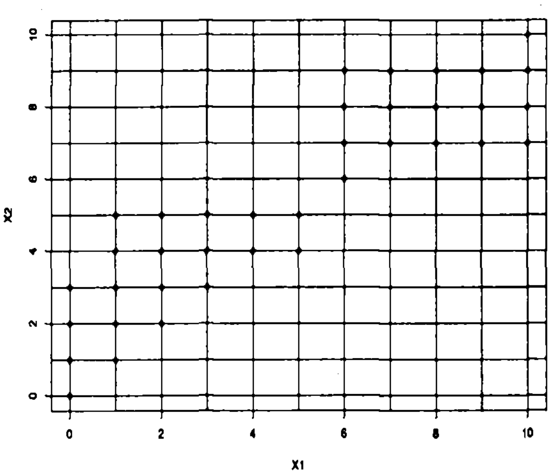

Figure A1. An Example of Bidimensional Support, Fully Connected in its Two Marginal Components, But Not Jointly So. In such a case, a Markov chain sampler starting in the bottom-left corner will not visit the region Top-right.

The same sequence of steps proves the result for cases when $\mathbf{X}_2$ and $\mathbf{X}_2 - (1,1)'$ are feasible, implying then that at least one of $\mathbf{X}_2 - (1,0)$ or $\mathbf{X}_2 - (0,1)$ is feasible—or, analogously, that $\mathbf{X}_2$ and $\mathbf{X}_2 + (-1,1)'$ or $\mathbf{X}_2$ and $\mathbf{X}_2 + (1,-1)'$ are feasible. As a result, the support is fully connected and irreducibility of the Markov chain sampler is ensured.

In the general multidimensional case, we proceed in a similar fashion. By assumption,

$$\mathbf{A}_1^{-1}(\mathbf{Y} - \mathbf{A}_2\mathbf{X}_2) \geq 0$$

and

$$\mathbf{A}_1^{-1}(\mathbf{Y} - \mathbf{A}_2(\mathbf{X}_2 + (1,1,\ldots,1)')) \geq 0$$

where now the vectors $\mathbf{X}_2$, and $(1,1,\ldots,1)'$ are n dimensional. Then

$$\mathbf{A}_1^{-1}(\mathbf{Y} - \mathbf{A}_2(\mathbf{X}_2 + (1,1\ldots,1)')) \geq 0$$

can be rewritten as

$$\mathbf{A}_1^{-1}(\mathbf{Y} - \mathbf{A}_2\mathbf{X}_2 - (A_2^{11}, A_2^{21}, \ldots, A_2^{n1})' - \cdots - (A_2^{1n}, A_2^{2n}, \ldots, A_2^{nn})') \geq 0,$$

and adding $n\mathbf{A}_1^{-1}(\mathbf{Y} - \mathbf{A}_2\mathbf{X}_2) \geq 0$ (by assumption) leads to

$$n(\mathbf{A}_1^{-1}(\mathbf{Y} - \mathbf{A}_2\mathbf{X}_2)) + \mathbf{A}_1^{-1}(\mathbf{Y} - \mathbf{A}_2\mathbf{X}_2 - (A_2^{11}, A_2^{21}, A_2^{n1})' - \cdots - (A_2^{1n}, A_2^{2n}, \ldots, A_2^{nn})') \geq 0.$$

As before, we redistribute the terms to obtain a sum of negative quantities and hence reach a contradiction: that is,

$$\begin{aligned} &\mathbf{A}_1^{-1}(\mathbf{Y} - \mathbf{A}_2\mathbf{X}_2 + \cdots + \mathbf{Y} - \mathbf{A}_2\mathbf{X}_2 \\ &\quad + \mathbf{Y} - \mathbf{A}_2\mathbf{X}_2 - (A_2^{11}, A_2^{21}, \ldots, A_2^{n1})' \\ &\quad - (A_2^{12}, A_2^{22}, \ldots, A_2^{2n})' - \cdots - (A_2^{n1}, A_2^{n2}, \ldots, A_2^{nn})') \\ &= \mathbf{A}_1^{-1}(\mathbf{Y} - \mathbf{A}_2\mathbf{X}_2 - (A_2^{11}, A_2^{21}, \ldots, A_2^{n1})') \\ &\quad + \mathbf{A}_1^{-1}(\mathbf{Y} - \mathbf{A}_2\mathbf{X}_2 - (A_2^{12}, A_2^{22}, \ldots, A_2^{2n})') \\ &\quad + \cdots + \mathbf{A}_1^{-1}(\mathbf{Y} - \mathbf{A}_2\mathbf{X}_2 - (A_2^{n1}, A_2^{n2}, \ldots, A_2^{nn})' \end{aligned}$$

must be nonnegative.

This proves irreducibility in the general n-dimensional case.

[Received June 1996. Revised October 1997.]

REFERENCES

Berka, S., and Boyce, D. E. (1994), "Implementation and Solution of a Large Asymmetric Network Equilibrium Model," working paper, Urban Transportation Center, University of Illinois at Chicago.

Best, N. G., Cowles, M. K., and Vines, S. K. (1995), *CODA: Convergence Diagnosis and Output Analysis Software for Gibbs Sampler Output* (Version 0.3), Cambridge, U.K.: Medical Research Council Biostatistics Unit.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (1996), *Markov Chain Monte Carlo in Practice*, London: Chapman and Hall.

Golub, G. H., and Van Loan, C. F. (1983), *Matrix Computations*, Baltimore: Johns Hopkins University Press.

Maher, M. J. (1983), "Inferences on Trip Matrices From Observations on Link Volumes: A Bayesian Statistical Approach," *Transportation Research*, 20B, 435–447.

MathSoft, Inc. (1993), *S-PLUS Guide to Statistical and Mathematical Analysis* (Version 3.2), Seattle: Author.

McNeil, S., and Hendrickson, C. (1985), "A Regression Formulation of the Matrix Estimation Problem," *Transportation Science*, 19, 278–292.

Sen, A. (1986), "Maximum Likelihood Estimation of Gravity Model Parameters," *Journal of Regional Science*, 26, 461–474.

Sheffi, Y. (1985), *Urban Transportation Networks: Equilibrium Analysis With Mathematical Programming Methods*, Englewood Cliffs, NJ: Prentice-Hall.
Smith, T. E. (1987), "Poisson Gravity Models of Spatial Flows," *Journal of Regional Science*, 27, 315–340.
Tierney, L. (1994), "Markov Chains for Exploring Posterior Distributions" (with discussion), *The Annals of Statistics*, 22, 1701–1762.
Van Aerde, M., Hellinga, B., Baker, M., and Rakha, H. (1996), "INTEGRATION: An Overview of Traffic Simulation Features," Technical Report, Dept. of Civil Engineering, Queen's University, Kingston, Canada.
Van Zuylen, H. J., and Willumsen, L. G. (1980), "The Most Likely Trip Matrix Estimated From Traffic Counts," *Transportation Research*, 14B, 281–293.
Vardi, Y. (1996), "Network Tomography: Estimating Source-Destination Traffic Intensities From Link Data," *Journal of the American Statistical Association*, 91, 365–377.
West, M. (1994), "Statistical Inference for Gravity Models in Transportation Flow Forecasting," Discussion Paper 94-40, ISDS, Duke University.

Comment

Y. VARDI

Networks are typically high-dimensional objects with complicated correlation and dependency structure and thus require high sampling rates for proper statistical analyses. Tebaldi and West (1998) (TW heretofore) give some insight and MCMC sampling tools for small networks with low sampling rates. As presented, the focus is on a *sample of size one*, and they need to better explain how, and if, their analyses and tools are applicable to even moderate-sized networks with moderate sampling rates. I do not see it. Their observation of "structural ambiguity" in the problem is a characteristic of the extremely low sampling rate that they consider and not of the structure of the problem itself. No statistical method except Bayesian with a *truly* informative prior could or should be attempted for such a small sample.

Consider first the fixed-routing problem on a typical (strongly connected, directed) network of n nodes. The number of OD pairs, c, is equal to $n(n-1)$, and the number of directed links, r, is of order $O(n)$. For instance, if each node has at most five outbound links, then $r \leq 5n$. It makes very little statistical sense to try to estimate the OD rates $\mathbf{\Lambda} = (\lambda_1, \ldots, \lambda_c)'$, a vector of dimension $c = n(n-1)$, from a *single* "snapshot" of link-traffic data $\mathbf{Y} = (Y_1, \ldots, Y_r)'$, which is a vector of dimension $r = O(n)$. The exception is, of course, if there is a priori information on the distribution of $\mathbf{\Lambda}$, but even then a single observation $\mathbf{Y}$ is likely to be overwhelmed by the prior. Thus it is crucial to take a large number of statistical replications in the form of K observations, $\mathbf{Y}^{(1)}, \ldots, \mathbf{Y}^{(K)}$, all iid from the same model:

$$\mathbf{Y}^{(k)} = \mathbf{A}\mathbf{X}^{(k)}, \qquad \mathbf{X}^{(k)} \sim \text{Poisson}(\mathbf{\Lambda}), \qquad k = 1, \ldots, K, \quad \text{independent}. \quad (1)$$

$\mathbf{A}$ is a known $r \times c$, 0–1 routing matrix (the Markovian routing case is discussed later) and $X_i^{(k)} \sim \text{Poisson}(\lambda_i), i = 1, \ldots, c, k = 1, \ldots, K$ independent. TW focus on estimating $\mathbf{\Lambda}$ when $K = 1$ and $\mathbf{\Lambda}$ has a known prior distribution. A key question is: What in their analysis carries over to the general case, when $K > 1$ (and typically large)? In trying to answer this, the first thing to note is that despite the "general friendliness" of Poisson modeling, the mean link-traffic,

$$\bar{\mathbf{Y}} = K^{-1} \sum_{k=1}^{K} \mathbf{Y}^{(k)},$$

is *not* a sufficient statistic and so we cannot base statistical inference on it or on any function of it. In fact, the likelihood function is

$$\begin{aligned} L(\mathbf{\Lambda} | \mathbf{Y}^{(1)}, \ldots, \mathbf{Y}^{(K)}) &= \prod_{k=1}^{K} \sum_{\{\mathbf{x} \in \mathbf{Y}^{(k)}\}} \prod_{j=1}^{c} e^{-\lambda_j} \lambda_j^{x_j} / x_j! \\ &\propto \exp\left\{ -K \sum_{j=1}^{c} \lambda_j \right\} \prod_{k=1}^{K} \sum_{\{\mathbf{x} \in \mathbf{Y}^{(k)}\}} \prod_{j=1}^{c} \lambda_j^{x_j}, \quad (2) \end{aligned}$$

where $\{\mathbf{x} \in \mathbf{Y}^{(k)}\} \equiv \{\mathbf{x} \in \mathbb{N}^c; \mathbf{Y}^{(k)} = \mathbf{A}\mathbf{x}\}, \mathbb{N} = \{0, 1, 2, \ldots\}$. I do not see how it can be factorized to yield any reduction of the sufficient statistic beyond the entire sample $\mathbf{Y}^{(1)}, \ldots, \mathbf{Y}^{(K)}$, modulo ordering.

Given that we cannot replace the data with a sufficient statistic of lower dimension, the next thing to consider is the form of the posterior distribution of Λ conditioned on the entire sample $\mathbf{Y}^{(1)}, \ldots, \mathbf{Y}^{(K)}$, and see to what extent the case $K = 1$ is generalizable:

$$\begin{aligned} P\{\mathbf{\Lambda} | \mathbf{Y}^{(1)}, \ldots, \mathbf{Y}^{(K)}\} &= P\{\mathbf{Y}^{(1)}, \ldots, \mathbf{Y}^{(K)} | \mathbf{\Lambda}\} P\{\mathbf{\Lambda}\} / P\{\mathbf{Y}^{(1)}, \ldots, \mathbf{Y}^{(K)}\} \\ &= P\{\mathbf{\Lambda}\} \prod_{k=1}^{K} (P\{\mathbf{Y}^{(k)} | \mathbf{\Lambda}\} / P\{\mathbf{Y}^{(k)}\}) \end{aligned}$$

Yehuda Vardi is Professor, Department of Statistics, Rutgers University, Piscataway, NJ 08854-8019.

Journal of the American Statistical Association
June 1998, Vol. 93, No. 442, Theory and Methods