# Bayesian cyclic networks, mutual information and reduced-order Bayesian inference

Robert K. Niven, Bernd R. Noack, Eurika Kaiser, Louis Cattafesta, Laurent Cordier, and Markus Abel

Citation: AIP Conference Proceedings **1757**, 020008 (2016);
View online: https://doi.org/10.1063/1.4959049
View Table of Contents: http://aip.scitation.org/toc/apc/1757/1
Published by the American Institute of Physics

# Bayesian Cyclic Networks, Mutual Information and Reduced-Order Bayesian Inference

Robert K. Niven*, Bernd R. Noack†, Eurika Kaiser†, Louis Cattafesta**, Laurent Cordier† and Markus Abel‡

*School of Engineering and Information Technology, The University of New South Wales, Canberra ACT 2600, Australia. Email r.niven@adfa.edu.au
†Institut Pprime, CNRS - Université de Poitiers - ENSMA, CEAT, 86036 Poitiers Cedex, France.
**Department of Mechanical Engineering, Florida State University, Tallahassee, FL 32301, USA.
‡Ambrosys GmbH / Institut für Physik und Astronomie, Universität Potsdam, Potsdam, Germany.

**Abstract.** We examine *Bayesian cyclic networks*, here defined as complete directed graphs in which the nodes, representing the domains of discrete or continuous variables, are connected by directed edges representing conditional probabilities between all pairs of variables. The prior probabilities associated with each domain are also included as probabilistic edges into each domain. Such networks provide a graphical representation of the *inferential connections* between variables, and substantially extend the standard definition of "Bayesian networks", usually defined as one-directional (acyclic) directed graphs. In a binary system, the proposed representation provides a graphical expression of Bayes' theorem. In higher-dimensional systems, further probabilistic relations can be recovered from the network cyclic properties and the joint probability of all variables. In particular, adopting a Markovian assumption leads to the theorem that the mutual information between any pair of variables on the network must be equivalent. Analysis of a hybrid Bayesian cyclic network of two continuous and two discrete variables – of the form of a commutative diagram – provides a framework for more computationally efficient *reduced-order Bayesian inference*, involving initial simplification by an order reduction process.

**Keywords:** Bayesian inference; Inferential chain; Reduced-order model; Clustering; Markov chain; Mutual information; Computational efficiency
**PACS:** 02.50.Cw, 02.50.Ga, 89.70.Cf, 02.50.Tt

## 1. INTRODUCTION

This study examines the concept of *Bayesian cyclic networks*, here defined as complete directed graphs composed of nodes representing domains, connected by edges representing conditional probabilities between elements of these domains. The prior probabilities associated with each domain are also represented as probabilistic "inflows" to each node. Such Bayesian cyclic networks are quite different to the standard definition of "Bayesian networks" as acyclic directed graphs [1, 2], which arguably omit one of the most important features of Bayesian inference: the calculation of inverse probabilities. The structures proposed here provide a graphical representation of the *inferential connections* between variables. The structures and resulting inference matrices exhibit some features of complete-graph forms of Markov chains [3, 4] and probabilistic automata [5].

In this analysis the major parameters are defined in §2.1, after which the binary, ternary, quaternary and $n$th-order forms of Bayesian cyclic networks are examined successively in §2.2-§2.5. The analysis provides some important results, including an intrinsic mutual information property of ternary and higher-order Bayesian cyclic networks, the result of a Markovian assumption. Detailed balance relations, and standard and augmented forms of the inference matrix, are also examined. Finally in §3, we propose a hybrid quaternary discrete-continuous Bayesian cyclic network as a framework for more computationally efficient Bayesian inference, referred to as *reduced-order Bayesian inference*, involving initial simplification by an order reduction (clustering or coarse-graining) process.

*Bayesian Inference and Maximum Entropy Methods in Science and Engineering*
AIP Conf. Proc. 1757, 020008-1–020008-9; doi: 10.1063/1.4959049
Published by AIP Publishing. 978-0-7354-1415-0/$30.00

# 2. ANALYSIS

## 2.1. Definitions

We first define the following variables and functions:

(i) The discrete variables $i \in \Omega_i, j \in \Omega_j, k \in \Omega_k, \ell \in \Omega_\ell$ defined on the countably finite discrete domains $\Omega_i, \Omega_j, \Omega_k, \Omega_\ell \subseteq \mathbb{N}$.

(ii) The continuous vector variables $\boldsymbol{x} \in \Omega_{\boldsymbol{x}}, \boldsymbol{y} \in \Omega_{\boldsymbol{y}}, \boldsymbol{z} \in \Omega_{\boldsymbol{z}}, \boldsymbol{w} \in \Omega_{\boldsymbol{w}}$, defined on the continuous domains $\Omega_{\boldsymbol{x}}$, $\Omega_{\boldsymbol{y}}$, $\Omega_{\boldsymbol{z}}$, $\Omega_{\boldsymbol{w}}$, each a subset of a multidimensional real domain $\mathbb{R}^N$ for (arbitrary) dimensionality $N$.

(iii) The joint, conditional and prior discrete probabilities (or probability mass functions, pmfs), defined by:

$$\begin{aligned} p(\alpha, \beta) &= \text{Prob}(\Upsilon_\alpha = \alpha, \Upsilon_\beta = \beta) \\ p(\beta|\alpha) &= \text{Prob}(\Upsilon_\beta = \beta | \Upsilon_\alpha = \alpha) \\ p(\alpha) &= \text{Prob}(\Upsilon_\alpha = \alpha) \end{aligned} \tag{1}$$

where scalars $\alpha, \beta \in \{i, j, k, \ell\}$ (as needed) and $\Upsilon_\mu$ denotes the random variable for $\mu \in \{\alpha, \beta\}$.

(iv) The joint, conditional and prior continuous probabilities, defined in terms of probability density functions (pdfs) by:

$$\begin{aligned} p(\boldsymbol{a}, \boldsymbol{b}) d\boldsymbol{a} d\boldsymbol{b} &= \text{Prob}(\boldsymbol{a} \leq \Upsilon_{\boldsymbol{a}} \leq \boldsymbol{a} + d\boldsymbol{a}, \boldsymbol{b} \leq \Upsilon_{\boldsymbol{b}} \leq \boldsymbol{b} + d\boldsymbol{b}) \\ p(\boldsymbol{b}|\boldsymbol{a}) d\boldsymbol{b} &= \text{Prob}(\boldsymbol{b} \leq \Upsilon_{\boldsymbol{b}} \leq \boldsymbol{b} + d\boldsymbol{b} | \boldsymbol{a} \leq \Upsilon_{\boldsymbol{a}} \leq \boldsymbol{a} + d\boldsymbol{a}) \\ p(\boldsymbol{a}) d\boldsymbol{a} &= \text{Prob}(\boldsymbol{a} \leq \Upsilon_{\boldsymbol{a}} \leq \boldsymbol{a} + d\boldsymbol{a}) \end{aligned} \tag{2}$$

where vectors $\boldsymbol{a}, \boldsymbol{b} \in \{\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{w}\}$ (as needed) and $\Upsilon_\mu$ denotes the random variable for $\mu \in \{\boldsymbol{a}, \boldsymbol{b}\}$.

(v) Where indicated, we adopt the discrete or continuous Markovian property that each conditional probability depends only on its immediate precursor, thus respectively for the chains $\alpha \mapsto \beta \mapsto \gamma$ or $\boldsymbol{a} \mapsto \boldsymbol{b} \mapsto \boldsymbol{c}$:

$$p(\gamma|\beta, \alpha) = p(\gamma|\beta), \qquad\qquad p(\boldsymbol{c}|\boldsymbol{b}, \boldsymbol{a}) d\boldsymbol{c} = p(\boldsymbol{c}|\boldsymbol{b}) d\boldsymbol{c} \tag{3}$$

## 2.2. Binary Networks

Consider the binary Bayesian cyclic networks illustrated in Figures 1(a)-(b), which respectively illustrate the (probabilistic or inferential) connections between the discrete domains $\Omega_i$ and $\Omega_j$ or continuous domains $\Omega_{\boldsymbol{x}}$ and $\Omega_{\boldsymbol{y}}$. In each graph, the two domains are represented as nodes, connected by directed edges in either direction, each representing the conditional probability of an element in the destination domain given an element in the source domain. More generally, each conditional probability can also be viewed as a map between the two entire domains. The illustrated networks should not be interpreted as Markov chains or cycles, composed of transition probabilities such as $p_{ij}(t)$ between mutually exclusive elements $i$ and $j$ at particular times $t$, but rather as inferential networks, composed of *inferential statements* such as $p(j|i)$ between dependent variables $i$ and $j$, valid irrespective of time (the inference process does not invoke a transition time, so all variables are examined contemporaneously). In the figures, it is also useful to include the probability by which an element of each domain is itself generated, or in other words the prior probabilities. These are included in Figures 1(a)-(b) as "inflow" edges to each domain. Finally, it is also possible to include (trivial) self-inferential edges such as $p(i|i)$, which connect each node to itself, but since these are always equal to 1, they are dropped from the representations drawn here.

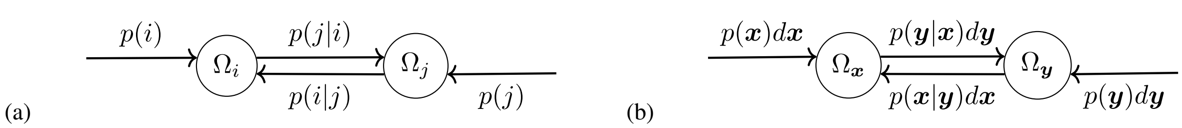


**FIGURE 1.** Schematic diagrams of (a) discrete and (b) continuous binary Bayesian cyclic networks.

Calculating the joint probability of elements in both domains in Figure 1(a) or (b), by multiplying the probabilities along the paths through the two domains in either direction, we obtain Bayes' rule [6, 7, 8]:

$$p(i,j) = p(i)\,p(j|i) = p(j)\,p(i|j) \tag{4}$$
$$p(\boldsymbol{x},\boldsymbol{y})d\boldsymbol{x}d\boldsymbol{y} = p(\boldsymbol{x})d\boldsymbol{x}\,p(\boldsymbol{y}|\boldsymbol{x})d\boldsymbol{y} = p(\boldsymbol{y})d\boldsymbol{y}\,p(\boldsymbol{x}|\boldsymbol{y})d\boldsymbol{x} \tag{5}$$

Figures 1(a)-(b) therefore provide *graphical representations of Bayes' rule*, respectively for discrete or continuous variables. From (4) or (5), Bayes' rule is more commonly written:

$$p(i|j) = \frac{p(i)\,p(j|i)}{p(j)}, \qquad p(\boldsymbol{x}|\boldsymbol{y}) = \frac{p(\boldsymbol{x})\,p(\boldsymbol{y}|\boldsymbol{x})}{p(\boldsymbol{y})} \tag{6}$$

but can also be expressed as:

$$\frac{p(i,j)}{p(i)p(j)} = \frac{p(j|i)}{p(j)} = \frac{p(i|j)}{p(i)}, \qquad \frac{p(\boldsymbol{x},\boldsymbol{y})}{p(\boldsymbol{x})\,p(\boldsymbol{y})} = \frac{p(\boldsymbol{y}|\boldsymbol{x})}{p(\boldsymbol{y})} = \frac{p(\boldsymbol{x}|\boldsymbol{y})}{p(\boldsymbol{x})} \tag{7}$$

These graphical and algebraic representations can be extended to higher-dimensional systems.

We also see that the probabilities in Figures 1(a)-(b), being inferential statements, do not involve "flows" of conserved quantities. In consequence, Bayesian cyclic networks do not satisfy Kirchhoff's laws, which here would imply that (I) the net probability flow into each node equals zero, and (II) there is zero probability flow around each cycle in the graph. For case (I) it makes little sense, for example, to calculate $p(i) - p(j|i) + p(i|j)$ for node $\Omega_i$, since these are non-independent probabilities with different meanings, which cannot be simply added or subtracted. For case (II), adding probabilities around a cycle would give, for example, $p(j|i) + p(i|j)$ which again makes little sense. Instead, as applied above, the probability arrows in Figures 1(a)-(b) and later graphs should primarily be considered to be *multiplicative*. Travelling around the cycles in Figure 1(a)-(b), for example, gives:

$$p(j|i)\,p(i|j) = \frac{p(i,j)^2}{p(i)p(j)}, \qquad p(\boldsymbol{y}|\boldsymbol{x})d\boldsymbol{y}\,p(\boldsymbol{x}|\boldsymbol{y})d\boldsymbol{x} = \frac{[p(\boldsymbol{x},\boldsymbol{y})d\boldsymbol{x}d\boldsymbol{y}]^2}{p(\boldsymbol{x})d\boldsymbol{x}\,p(\boldsymbol{y})d\boldsymbol{y}} \tag{8}$$

We can also express the graph structures of Figures 1(a)-(b) using the *probabilistic inference matrices*:

$$M_2^{disc} = \begin{bmatrix} 1 & p(j|i) \\ p(i|j) & 1 \end{bmatrix}, \qquad M_2^{cts} = \begin{bmatrix} 1 & p(\boldsymbol{y}|\boldsymbol{x})d\boldsymbol{y} \\ p(\boldsymbol{x}|\boldsymbol{y})d\boldsymbol{x} & 1 \end{bmatrix} \tag{9}$$

where each term $M_{ij} = p(j|i)$ indicates the inferential path from $i$ to $j$. These matrices include the (trivial) self-inferential statements such as $p(i|i) = 1$. Note that the quantities in the inference matrix are conditional probabilities with different (non-independent) conditions, so in contrast to a probability transition matrix, they need not add to 1 along any row or column. Indeed, the inference matrix is more akin to a (probabilistic) correlation matrix. The relations represented by (9) can be shown to be reflexive, symmetric and transitive [9], albeit with inverse probabilities rather than 1 in the non-diagonal elements.

The inference matrices can also be augmented with the priors, represented as flows from a fictitious "external" domain, to give the augmented inference matrices:

$$\widetilde{M}_2^{disc} = \begin{bmatrix} 0 & p(i) & p(j) \\ 0 & 1 & p(j|i) \\ 0 & p(i|j) & 1 \end{bmatrix}, \qquad \widetilde{M}_2^{cts} = \begin{bmatrix} 0 & p(\boldsymbol{x})d\boldsymbol{x} & p(\boldsymbol{y})d\boldsymbol{y} \\ 0 & 1 & p(\boldsymbol{y}|\boldsymbol{x})d\boldsymbol{y} \\ 0 & p(\boldsymbol{x}|\boldsymbol{y})d\boldsymbol{x} & 1 \end{bmatrix} \tag{10}$$

## 2.3. Ternary Cycles

We now consider the ternary Bayesian cyclic networks of discrete or continuous probabilities illustrated respectively in Figures 2(a)-(b). Once again these do not conform to the usual definition of (acyclic) Bayesian networks. For these systems, we further assume the Markovian property (3) that each probabilistic link does not depend on the history of previous links, and so is path-independent.

By multiplication of probabilities, depending on the entry point and cycle direction (clockwise or anticlockwise), these yield $3! = 6$ permutations of 3-node relations. Examining firstly the discrete case:

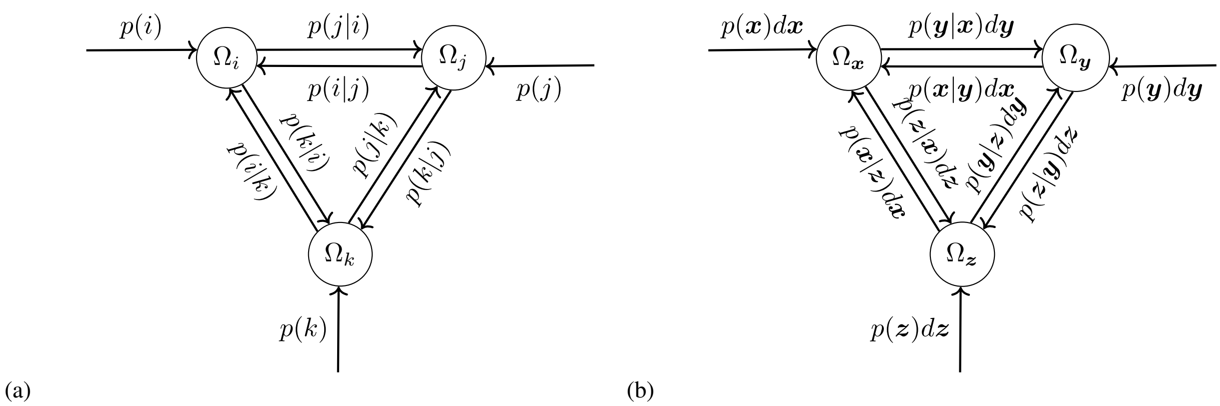


**FIGURE 2.** Schematic diagrams of (a) discrete and (b) continuous ternary Bayesian cyclic networks.

$$\begin{aligned} p(i,j,k) = p(i)\,p(j|i)\,p(k|j) &\quad = p(i)\,p(k|i)\,p(j|k) = p(j)\,p(k|j)\,p(i|k) \\ = p(j)\,p(i|j)\,p(k|i) &\quad = p(k)\,p(i|k)\,p(j|i) = p(k)\,p(j|k)\,p(i|j) \end{aligned} \tag{11}$$

By division:

$$\frac{p(i,j,k)}{p(i)\,p(j)\,p(k)} = \frac{p(j|i)}{p(j)}\,\frac{p(k|j)}{p(k)} = \frac{p(k|i)}{p(k)}\,\frac{p(j|k)}{p(j)} = \frac{p(k|j)}{p(k)}\,\frac{p(i|k)}{p(i)} = \frac{p(i|j)}{p(i)}\,\frac{p(k|i)}{p(k)} = \frac{p(i|k)}{p(i)}\,\frac{p(j|i)}{p(j)} = \frac{p(j|k)}{p(j)}\,\frac{p(i|j)}{p(i)} \tag{12}$$

Substitution of Bayes' rule (6) into (11) to eliminate the clockwise paths gives:

$$\begin{aligned} p(i,j,k) = p(i)\,p(j|i)\,p(k|j) &\quad = p(i)\,\frac{p(i|k)p(k)}{p(i)}\,\frac{p(k|j)p(j)}{p(k)} \\ = p(j)\,p(k|j)\,p(i|k) &\quad = p(j)\,\frac{p(j|i)p(i)}{p(j)}\,\frac{p(i|k)p(k)}{p(i)} \\ = p(k)\,p(i|k)\,p(j|i) &\quad = p(k)\,\frac{p(k|j)p(j)}{p(k)}\,\frac{p(j|i)p(i)}{p(j)} \end{aligned} \tag{13}$$

from which

$$p(i)\,p(j|i) = p(j)\,p(i|k), \qquad p(j)\,p(k|j) = p(k)\,p(j|i), \qquad p(k)\,p(i|k) = p(i)\,p(k|j) \tag{14}$$

whence by rearrangement and back-substitution of Bayes' rules (6):

$$\frac{p(i|j)}{p(i)} = \frac{p(i|k)}{p(i)} = \frac{p(j|i)}{p(j)} = \frac{p(j|k)}{p(j)} = \frac{p(k|i)}{p(k)} = \frac{p(k|j)}{p(k)} \tag{15}$$

from which, or directly from (12):

$$p(i|j) = p(i|k), \quad p(j|i) = p(j|k), \quad p(k|i) = p(k|j) \tag{16}$$

$$\frac{p(i,j)}{p(i)p(j)} = \frac{p(i,k)}{p(i)p(k)} = \frac{p(j,k)}{p(j)p(k)} \tag{17}$$

These have the continuous analogues:

$$p(\boldsymbol{x}|\boldsymbol{y}) = p(\boldsymbol{x}|\boldsymbol{z}), \quad p(\boldsymbol{y}|\boldsymbol{x}) = p(\boldsymbol{y}|\boldsymbol{z}), \quad p(\boldsymbol{z}|\boldsymbol{x}) = p(\boldsymbol{z}|\boldsymbol{y}) \tag{18}$$

$$\frac{p(\boldsymbol{x},\boldsymbol{y})}{p(\boldsymbol{x})p(\boldsymbol{y})} = \frac{p(\boldsymbol{x},\boldsymbol{z})}{p(\boldsymbol{x})p(\boldsymbol{z})} = \frac{p(\boldsymbol{y},\boldsymbol{z})}{p(\boldsymbol{y})p(\boldsymbol{z})} \tag{19}$$

We see that, firstly, each set of conditional probabilities are equivalent regardless of the conditioning variable, and secondly, the ratio of each joint probability to its constituent priors is equal throughout the network. These are both consequences of the Markovian assumption.

Now by equivalent functional and multiplicative operations, followed by summation or integration over all three variables, the latter relations can be rearranged. For the discrete case:

$$\begin{aligned}\sum_{i\in\Omega_i}\sum_{j\in\Omega_j}\sum_{k\in\Omega_k} p(i)p(j)p(k)\frac{p(i,j)}{p(i)p(j)}\log\frac{p(i,j)}{p(i)p(j)} \\ = \sum_{i\in\Omega_i}\sum_{j\in\Omega_j}\sum_{k\in\Omega_k} p(i)p(j)p(k)\frac{p(i,k)}{p(i)p(k)}\log\frac{p(i,k)}{p(i)p(k)} \\ = \sum_{i\in\Omega_i}\sum_{j\in\Omega_j}\sum_{k\in\Omega_k} p(i)p(j)p(k)\frac{p(j,k)}{p(j)p(k)}\log\frac{p(j,k)}{p(j)p(k)}\end{aligned} \tag{20}$$

where log is a logarithm of any consistent base. Simplifying and normalisation of the priors gives:

$$\sum_{i\in\Omega_i}\sum_{j\in\Omega_j} p(i,j)\log\frac{p(i,j)}{p(i)p(j)} = \sum_{i\in\Omega_i}\sum_{k\in\Omega_k} p(i,k)\log\frac{p(i,k)}{p(i)p(k)} = \sum_{j\in\Omega_j}\sum_{k\in\Omega_k} p(j,k)\log\frac{p(j,k)}{p(j)p(k)} \tag{21}$$

This has the continuous analogue:

$$\int\limits_{\Omega_{\boldsymbol{x}}}\int\limits_{\Omega_{\boldsymbol{y}}} p(\boldsymbol{x},\boldsymbol{y})\log\frac{p(\boldsymbol{x},\boldsymbol{y})}{p(\boldsymbol{x})p(\boldsymbol{y})}d\boldsymbol{x}d\boldsymbol{y} = \int\limits_{\Omega_{\boldsymbol{x}}}\int\limits_{\Omega_{\boldsymbol{z}}} p(\boldsymbol{x},\boldsymbol{z})\log\frac{p(\boldsymbol{x},\boldsymbol{z})}{p(\boldsymbol{x})p(\boldsymbol{z})}d\boldsymbol{x}d\boldsymbol{z} = \int\limits_{\Omega_{\boldsymbol{y}}}\int\limits_{\Omega_{\boldsymbol{z}}} p(\boldsymbol{y},\boldsymbol{z})\log\frac{p(\boldsymbol{y},\boldsymbol{z})}{p(\boldsymbol{y})p(\boldsymbol{z})}d\boldsymbol{y}d\boldsymbol{z} \tag{22}$$

These contain the *mutual information* between pairs of variables. In consequence, we obtain the interesting finding that the mutual information of each pair of variables on a Markovian Bayesian ternary cycle are equivalent. Eqs. (21)-(22) can be written in the compact forms:

$$I(\Upsilon_\alpha;\Upsilon_\beta) = \text{constant},\ \ \alpha,\beta\in\{i,j,k\},\alpha\neq\beta, \qquad I(\Upsilon_{\boldsymbol{a}};\Upsilon_{\boldsymbol{b}}) = \text{constant},\ \ \boldsymbol{a},\boldsymbol{b}\in\{\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}\},\boldsymbol{a}\neq\boldsymbol{b} \tag{23}$$

where $I$ is the applicable discrete or continuous mutual information, and we recall $\Upsilon_\mu$ is the random variable for parameter $\mu$.

We can also derive a different result, which exploits the cyclic property of the networks in Figures 2(a)-(b). From the Markov assumption (3), and equivalence of the clockwise and anticlockwise cyclic relations, we obtain the probabilistic *principle of detailed balance*:

$$p(j|i)\,p(k|j)\,p(i|k) = p(k|i)\,p(j|k)\,p(i|j) = p(i,j,k)\,\frac{p(\alpha,\beta)}{p(\alpha)p(\beta)}, \tag{24}$$

$$p(\boldsymbol{y}|\boldsymbol{x})\,p(\boldsymbol{z}|\boldsymbol{y})\,p(\boldsymbol{x}|\boldsymbol{z}) = p(\boldsymbol{z}|\boldsymbol{x})\,p(\boldsymbol{y}|\boldsymbol{z})\,p(\boldsymbol{x}|\boldsymbol{y}) = p(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})d\boldsymbol{x}d\boldsymbol{y}d\boldsymbol{z}\,\frac{p(\boldsymbol{a},\boldsymbol{b})d\boldsymbol{a}d\boldsymbol{b}}{p(\boldsymbol{a})d\boldsymbol{a}\,p(\boldsymbol{b})d\boldsymbol{b}} \tag{25}$$

again for $\alpha,\beta\in\{i,j,k\},\alpha\neq\beta$ and $\boldsymbol{a},\boldsymbol{b}\in\{\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}\},\boldsymbol{a}\neq\boldsymbol{b}$. These are usually written as the ratios:

$$\frac{p(i|j)\,p(j|k)\,p(k|i)}{p(j|i)\,p(k|j)\,p(i|k)} = 1, \qquad \frac{p(\boldsymbol{x}|\boldsymbol{y})\,p(\boldsymbol{y}|\boldsymbol{z})\,p(\boldsymbol{z}|\boldsymbol{x})}{p(\boldsymbol{y}|\boldsymbol{x})\,p(\boldsymbol{z}|\boldsymbol{y})\,p(\boldsymbol{x}|\boldsymbol{z})} = 1 \tag{26}$$

The detailed balance relations have no binary analogue for Figures 1(a)-(b), since they require distinct clockwise and anticlockwise cycles, but are related to (8).

We can also express the ternary graph structures of Figures 2(a)-(b) as the inference matrices:

$$M_3^{disc} = \begin{bmatrix} 1 & p(j|i) & p(k|i) \\ p(i|j) & 1 & p(k|j) \\ p(i|k) & p(j|k) & 1 \end{bmatrix}, \qquad M_3^{cts} = \begin{bmatrix} 1 & p(\boldsymbol{y}|\boldsymbol{x})d\boldsymbol{y} & p(\boldsymbol{z}|\boldsymbol{x})d\boldsymbol{z} \\ p(\boldsymbol{x}|\boldsymbol{y})d\boldsymbol{x} & 1 & p(\boldsymbol{z}|\boldsymbol{y})d\boldsymbol{z} \\ p(\boldsymbol{x}|\boldsymbol{z})d\boldsymbol{x} & p(\boldsymbol{y}|\boldsymbol{z})d\boldsymbol{y} & 1 \end{bmatrix} \tag{27}$$

or the augmented inference matrices:

$$\widetilde{M}_3^{disc} = \begin{bmatrix} 0 & p(i) & p(j) & p(k) \\ 0 & 1 & p(j|i) & p(k|i) \\ 0 & p(i|j) & 1 & p(k|j) \\ 0 & p(i|k) & p(j|k) & 1 \end{bmatrix}, \qquad \widetilde{M}_3^{cts} = \begin{bmatrix} 0 & p(\boldsymbol{x})d\boldsymbol{x} & p(\boldsymbol{y})d\boldsymbol{y} & p(\boldsymbol{z})d\boldsymbol{z} \\ 0 & 1 & p(\boldsymbol{y}|\boldsymbol{x})d\boldsymbol{y} & p(\boldsymbol{z}|\boldsymbol{x})d\boldsymbol{z} \\ 0 & p(\boldsymbol{x}|\boldsymbol{y})d\boldsymbol{x} & 1 & p(\boldsymbol{z}|\boldsymbol{y})d\boldsymbol{z} \\ 0 & p(\boldsymbol{x}|\boldsymbol{z})d\boldsymbol{x} & p(\boldsymbol{y}|\boldsymbol{z})d\boldsymbol{y} & 1 \end{bmatrix} \tag{28}$$

Relations (27) are again reflexive, symmetric and transitive [9], but again with inverse non-diagonal elements.

## 2.4. Quaternary Cycles

We now consider quaternary Bayesian cyclic networks for discrete or continuous probabilistic maps, illustrated in Figures 3(a)-(b). We again assume the Markovian property (3). Since every conditional probability can be formulated, at least in principle, we obtain a complete graph structure.

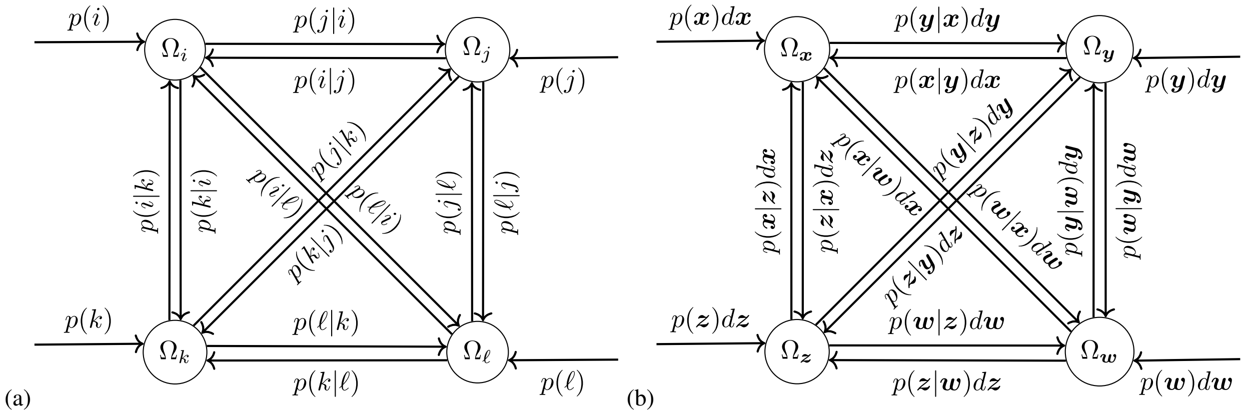


**FIGURE 3.** Schematic diagrams of (a) discrete and (b) continuous quaternary Bayesian cyclic networks.

By multiplication of probabilities along a chain through 4 nodes, we obtain in summary form:

$$\begin{aligned} &p(\alpha,\beta,\gamma,\delta) = p(\alpha)\, p(\beta|\alpha)\, p(\gamma|\beta)\, p(\delta|\gamma), \\ &\forall \alpha \neq \beta \neq \gamma \neq \delta, \quad \text{for } \alpha,\beta,\gamma,\delta \in \begin{cases} \{i,j,k,\ell\} & \text{if discrete} \\ \{\boldsymbol{x},\boldsymbol{y},\boldsymbol{z},\boldsymbol{w}\} & \text{if continuous} \end{cases} \end{aligned} \tag{29}$$

There are here $4! = 24$ permutations of 4-node paths. By division:

$$\frac{p(\alpha,\beta,\gamma,\delta)}{p(\alpha)\,p(\beta)\,p(\gamma)\,p(\delta)} = \frac{p(\beta|\alpha)}{p(\beta)}\,\frac{p(\gamma|\beta)}{p(\gamma)}\,\frac{p(\delta|\gamma)}{p(\delta)}, \quad \forall \alpha \neq \beta \neq \gamma \neq \delta \tag{30}$$

which gives quaternary analogues of (12) and their continuous form. Substitution of Bayes' rules and reduction in the manner of §2.3 yields the analogous results (with $\alpha,\beta$ defined as in (29)):

$$p(\alpha|\beta) = p(\alpha|\gamma), \quad \forall \alpha \neq \beta \neq \gamma \tag{31}$$

$$\frac{p(\alpha,\beta)}{p(\alpha)p(\beta)} = \text{constant}, \quad \forall \alpha \neq \beta \tag{32}$$

$$I(\Upsilon_\alpha;\Upsilon_\beta) = \text{constant}, \quad \forall \alpha \neq \beta \tag{33}$$

where the mutual information is defined for either the discrete or continuous case according to (21) or (22). There are ${}_4C_2 = \binom{4}{2} = 6$ combinations of relations in (32) and (33), corresponding to the six possible pairs of nodes.

We can also derive detailed balance relations for the ternary and quaternary cycles in Figures 3(a)-(b):

$$\frac{p(\alpha|\beta)\, p(\beta|\gamma)\, p(\gamma|\alpha)}{p(\beta|\alpha)\, p(\gamma|\beta)\, p(\alpha|\gamma)} = 1, \quad \forall \alpha \mapsto \beta \mapsto \gamma \tag{34}$$

$$\frac{p(\alpha|\beta)\, p(\beta|\gamma)\, p(\gamma|\delta)\, p(\delta|\alpha)}{p(\beta|\alpha)\, p(\gamma|\beta)\, p(\delta|\gamma)\, p(\alpha|\delta)} = 1, \quad \forall \alpha \mapsto \beta \mapsto \gamma \mapsto \delta \tag{35}$$

Only three of the five relations in (34)-(35) are independent, as can be shown by multiplication.

Once again we can express the networks in Figures 3(a)-(b) in the form of quaternary inference or augmented inference matrices, analogous to (9)-(10) or (27)-(28).

## 2.5. $n$th-Order Cycles

We now consider $n$th-order Bayesian cyclic networks for discrete or continuous probabilistic maps (with $n \in \mathbb{N}$), for which the above findings extend naturally. We again assume the Markovian property (3). The main results are:

$$p(\alpha_1, ..., \alpha_n) = p(\alpha_1) \prod_{m=2}^{n} p(\alpha_m | \alpha_{m-1}), \quad \forall \alpha_m \in \begin{cases} \{i_1, ..., i_n\} & \text{if discrete} \\ \{\boldsymbol{x}_1, ..., \boldsymbol{x}_n\} & \text{if continuous} \end{cases} \tag{36}$$

with $n!$ permutations of paths, whence:

$$\frac{p(\alpha_1, ..., \alpha_n)}{\prod\limits_{m=1}^{n} p(\alpha_m)} = \prod_{m=2}^{n} \frac{p(\alpha_m | \alpha_{m-1})}{p(\alpha_m)}, \tag{37}$$

Reduction gives:

$$p(\alpha_m | \alpha_j) = p(\alpha_m | \alpha_k), \quad \forall \alpha_m \neq \alpha_j \neq \alpha_k \tag{38}$$

$$\frac{p(\alpha_m, \alpha_j)}{p(\alpha_m) p(\alpha_j)} = \text{constant}, \quad \forall \alpha_m \neq \alpha_j \tag{39}$$

$$I(\Upsilon_{\alpha_m}; \Upsilon_{\alpha_j}) = \text{constant}, \quad \forall \alpha_m \neq \alpha_j \tag{40}$$

with ${}_nC_2 = \binom{n}{2}$ combinations of relations in (39) and (40). A variety of detailed balance relations are also obtained:

$$\prod_{m=1}^{c} \frac{p(\alpha_m | \alpha_{m+1})}{p(\alpha_m | \alpha_{m-1})} = 1, \tag{41}$$

where $c \in \{3, ..., n\}$, with a rubberbanded index so that $\alpha_0 = \alpha_c$ and $\alpha_{c+1} = \alpha_1$. Only some of these relations will be independent. Finally, the $n$th-order inference and augmented inference matrices are respectively:

$$M_n = \begin{bmatrix} 1 & ... & p(i_n|i_1) \\ \vdots & \ddots & \vdots \\ p(i_1|i_n) & ... & 1 \end{bmatrix}, \qquad \widetilde{M}_n = \begin{bmatrix} 0 & p(i_1) & ... & p(i_n) \\ \vdots & 1 & ... & p(i_n|i_1) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & p(i_1|i_n) & ... & 1 \end{bmatrix} \tag{42}$$

# 3. A HYBRID QUATERNARY CYCLE AND REDUCED-ORDER BAYESIAN INFERENCE

We now consider a hybrid quaternary cycle composed of two continuous parameters $\boldsymbol{x} \in \Omega_{\boldsymbol{x}}$ and $\boldsymbol{y} \in \Omega_{\boldsymbol{y}}$ and two discrete parameters $i \in \Omega_i$ and $j \in \Omega_j$, illustrated in the Bayesian cyclic network in Figure 4. (Although they exist, we omit the diagonal connections for convenience.) In particular, we are interested in Bayesian updating between a data space $\Omega_{\boldsymbol{x}}$ and model space $\Omega_{\boldsymbol{y}}$, shown in the solid arrow as the "direct" route. We propose an alternative "indirect" framework for this inference, referred to as *reduced-order Bayesian inference*, involving the following steps:

- A clustering or order reduction from the continuous data space $\Omega_{\boldsymbol{x}}$ to a discrete data space $\Omega_i$;
- Bayesian updating from this discrete data space to a clustered model space $\Omega_j$; and
- A "declustering" step or reverse Bayesian inference from the clustered model space to the original model space.

This framework is proposed to avoid the high computational cost of direct Bayesian updating, of particular importance for real-time applications such as turbulent flow control [10, 11]. The first step can involve any method for order reduction, including the k-means clustering method applied recently to the classification of dynamical and fluid mechanics systems [12]. We wish to examine the effectiveness of the reduced-order Bayesian inference compared to the direct method.

In contrast to the previous analyses, it is no longer reasonable to make the Markovian assumption (3) for transitions between continuous and discrete states. The very act of clustering (= "order reduction" = "coarse-graining") will cause

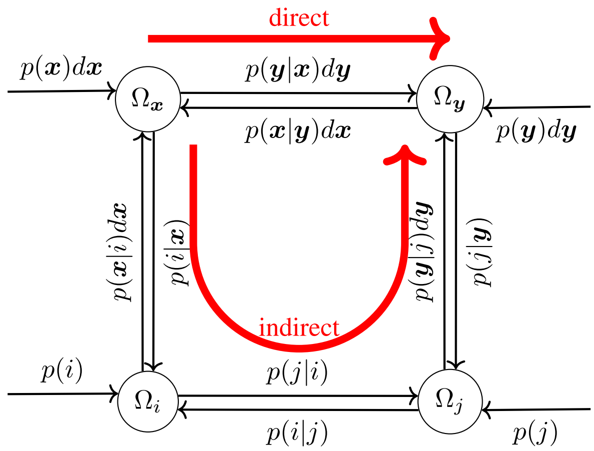


**FIGURE 4.** Incomplete schematic diagram of a hybrid discrete-continuous quaternary Bayesian cyclic network.

a *loss of information* during this process, which can never be recovered by any later operation along any path. This will be manifested in a net loss of information around the indirect path compared to the direct path:

$$\Delta I = I(\Upsilon_{\boldsymbol{x}};\Upsilon_{\boldsymbol{y}})_{direct} - I(\Upsilon_{\boldsymbol{x}};\Upsilon_{\boldsymbol{y}})_{indirect} \geq 0 \tag{43}$$

This loss of information can be used as a criterion for assessing the effectiveness of the reduced-order method for Bayesian inference.

## 4. CONCLUSIONS

This study defines the concept of *Bayesian cyclic networks*, consisting of complete directed graphs composed of nodes representing domains, connected by edges representing conditional probabilities or *inferential statements* between elements of these domains. The prior probabilities for each domain are also included as external inputs. Such Bayesian cyclic networks are quite different to the standard definition of “Bayesian networks” as acyclic directed graphs [1, 2], which omit the most important feature of Bayesian inference, that of inverse probabilities. In §2.2, by analysis of discrete and continuous binary systems, this representation is shown to provide a graphical expression of Bayes’ theorem. In higher-dimensional systems §2.3-§2.5, further probabilistic relations can be recovered. Most importantly, for Markovian networks, it is shown that the mutual information between any pair of variables on the network must be equivalent. Detailed balance relations, and standard and augmented forms of the inference matrix, are also examined. Finally in §3, we propose a hybrid discrete-continuous Bayesian cyclic network as a framework for more computationally efficient Bayesian inference – here termed *reduced-order Bayesian inference* – involving initial simplification by an order reduction (clustering or coarse-graining) process. An information-theoretic criterion for the effectiveness of this method is proposed.

## ACKNOWLEDGMENTS

This project was funded by the Australian Research Council Discovery Project DP140104402; the Go8 / DAAD Australia-Germany Joint Research Cooperation Scheme; Region Poitou-Charentes, France; and ANR Chair of Excellence TUCOROM, Institut PPrime, Poitiers, France. The authors also thank the two anonymous reviewers.

## REFERENCES

1. J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann, San Francisco, USA, 1988.
2. C.M. Bishop, Pattern Recognition and Machine Learning, Springer, NY, 2006.
3. J.R. Norris, Markov Chains, Cambridge UP, 1997.
4. G.R. Grimmett & D.R. Stirzaker, Probability and Random Processes, Oxford UP, 2001.
5. M.O. Rabin, Probabilistic Automata, Information and Control 6 (3): 230-245, 1963.
6. T. Bayes, R. Price, An essay towards solving a problem in the doctrine of chance, *Phil. Trans. Royal Soc. London*, 53: 370–418, 1763.

7. P. Laplace, Mémoire sur la probabilité des causes par les événements, *l'Academie Royale des Sciences*, 6: 621-656, 1774.
8. E.T. Jaynes, Probability Theory: The Logic of Science, Cambridge University Press, Cambridge, 2003.
9. S.S. Epp, Discrete Mathematics with Applications, 4th ed., Brooks Cole, 2010.
10. D. Cheng, X. Hu, T. Shen, Analysis and Design of Nonlinear Control Systems, Springer Press Beijing, 2010.
11. B.R. Noack, M. Morzynski, G. Tadmor (eds), *Reduced-Order Modelling for Flow Control*, Springer Wien, NY, 2011.
12. E. Kaiser, B.R. Noack, L. Cordier, A. Spohn, M. Segond, M. Abel., G. Daviller, J. Östh, S. Krajnović, R.K. Niven, Cluster-based reduced-order modelling of a mixing layer, *J. Fluid Mech.*, 754: 365–414, 2014.