7-11-2014

Exact Tests for Singular Network Data

Ian H. Dinwoodie
Portland State University, ihd@pdx.edu

Kruti Pandya
Portland State University

Exact Tests for Singular Network Data

Ian H Dinwoodie
Kruti Pandya
Fariborz Maseeh Department of Mathematics and Statistics
Portland State University
PO Box 751, Portland, OR 97207-0751

January 2013
revised February 2014

Abstract

We propose methodology for exact statistical tests of hypotheses for models of network dynamics. The methodology formulates Markovian exponential families, then uses sequential importance sampling to compute expectations within basins of attraction and within level sets of a sufficient statistic for an overdispersion model. Comparisons of hypotheses can be done conditional on basins of attraction. Examples are presented.

Keywords: Basin of attraction, biological network, conditional test, polynomial dynamics, sequential importance sampling, sufficient statistic

1 Introduction

In this paper we develop methodology for exact statistical tests of hypotheses for models of network dynamics. We introduce statistical models that include a dispersion parameter to deal with real data, formulate conditional tests that respect a given test size, and develop practical methods for computing expectations within

level sets or fibers of a sufficient statistic. The methods are applied to examples of biological networks, including one on abscisic acid (ABA) signalling and another on cancer cell signalling.

Biological networks are often modeled as discrete dynamical systems in order to understand interactions and regulatory processes. Boolean models, a two-state conceptual simplification, continue to be developed and used (Albert et al., 2003; Klamt et al., 2006; Morris et al., 2010; Saez-Rodriguez et al., 2011; Stigler, 2006; Thomas, 1973, 1998). Extensions to discrete states with more than two levels have been of interest, so that on-off states may be refined to low-medium-high for example (Mendoza, 2006). In this paper we focus on the binary case but any discretization can be done similarly (Dinwoodie, 2012).

Regulatory network data has features that cause difficulties for rigorous statistical inference: high dimensionality, over-dispersion, and lack of ergodicity due to absorbing states and limit cycles. In addition, conceptually useful models are simplified to the point that data from experiments on real networks have probability 0 under the model, a situation that may be called singular data since the model probabilities and the empirical data are technically incompatible. Thus we introduce a straightforward dispersed version of idealized dynamics, and we can view the data as singular with respect to deterministic dynamics but lying within the support of the distributions in the dispersed dynamical model. We develop conditional inference to do tests of controlled size α even with unknown nuisance dispersion parameter ϕ in the model. Conditional tests generally require computation in level sets or fibers of a sufficient statistic. This is a classical subject now for contingency tables where the sufficient statistics are linear, and recent developments include connections with integer programming and commutative algebra (Aoki et al., 2012; Drton et al., 2008; Riccomagno et al., 2000), and sequential importance sampling (Chen et al., 2006).

Biological network models are very different than log-linear models in several key technical ways: the states are generally binary or ternary vectors instead of integer; and the constraint equations for sampling are not linear. On the plus

side, the dynamical equations are generally lightly coupled, and by this we mean imprecisely that most equations depend on only a few indeterminates so equations are not highly linked together or highly dependent. The methodology will use a combination of elementary computational commutative algebra, and sequential importance sampling for computing exact conditional p-values. Some of our methods require a lexicographic Gröbner basis (Kreuzer et al., 2000) for a set of polynomials, and while this is theoretically a hard and complex thing to compute it works well on many real examples from the systems biology literature including those in Section 6. Computations were done with Singular (Decker et al., 2011) but other software such as Macaulay 2 (Grayson et al., 2012) is also suitable. Section 2 is a self-contained discussion of attracting sets, which are simple for dynamical systems on a finite state space and fundamental for biological understanding. Section 3 presents a statistical model for dynamics that includes a dispersion parameter ϕ that makes idealized dynamics compatible with noisy data. Section 4 is a technical section on sequential importance sampling on a basin of attraction where some computational commutative algebra is used, but for practical purposes it is only necessary to understand the implementation in Example 2. This section is not new except for the extension of Theorem 1 to limit cycles from earlier fixed point assumptions.

Section 5 has the new results and this section formulates the probability model, sets up the problem of exact statistical inference, conditions on a sufficient statistic, and computes conditional p-values for exact tests. Finally Section 6 applies the method to network examples with published data.

2 Attractors and Basins of Attraction

Attracting sets in discrete Boolean dynamics can be steady states (fixed points) or limit cycles. In (Li et al., 2006), limit cycles appear in a model of stomatal closure, but in some cases only steady states are possible. In all of our real examples of Section 6 both steady states and cycles appear. In this section we set up the notation

for attracting sets and establish basic properties, as one goal of this paper is to extend a previous method (Dinwoodie, 2012) for sampling points in the basin of attraction of a steady state to any attracting set.

Consider a state space $\Omega := \{\mathbf{x} = (x_1, \ldots, x_d), x_j = 0, 1\} = \{0,1\}^d$, a d-fold product of binary on-off states.

Let $F = (F_1, \ldots, F_d)$ be a transition map or transition function or update function on Ω, where $F_j : \Omega \to \{0,1\}$ and $F : \Omega \to \Omega$. This map is deterministic, and the real time step which it represents can depend on many factors (Saez-Rodriguez et al., 2007). Randomized versions called asynchronous updates are of interest (Saadatpour et al., 2010), but we do not treat that extension in this paper.

For a state $\mathbf{x}$, define the limiting set

$$A_{\mathbf{x}} = \cap_{k=1}^{\infty} \cup_{n=k}^{\infty} F^n(\mathbf{x}) \tag{1}$$

where F^n is the n-fold composition of F.

The resulting sets, as $\mathbf{x}$ varies in the state space $\{0,1\}^d$ are disjoint and are the limiting sets or attractor sets of the system.

Proposition 1. *$A_{\mathbf{x}} \cap A_{\mathbf{y}} = \emptyset$ or $A_{\mathbf{x}} = A_{\mathbf{y}}$.*

Proof. Suppose $A_{\mathbf{x}} \cap A_{\mathbf{y}} \neq \emptyset$ and let $\mathbf{z} \in A_{\mathbf{x}} \cap A_{\mathbf{y}}$. This implies that $\mathbf{z} = F^{n_k}(\mathbf{x}) = F^{m_k}(\mathbf{y})$ for increasing sequences $n_k, m_k, k = 1, 2, 3, \ldots$. Then if $\mathbf{x}' \in A_{\mathbf{x}}$, it follows that $\mathbf{x}' = F^{i_k}(\mathbf{x}) = F^{i_k - n_1}(\mathbf{z}) = F^{i_k - n_1}(F^{m_1}(\mathbf{y}))$ and thus $\mathbf{x}' \in A_{\mathbf{y}}$. By symmetry, $A_{\mathbf{y}} \subset A_{\mathbf{x}}$ as well. □

A steady state $\mathbf{p} = (p_1, \ldots, p_d) \in \Omega$ has the defining property that $F(\mathbf{p}) = \mathbf{p}$, a cycle of length 1. Define the set of points that eventually lead to a steady state $\mathbf{p}$:

$$B_{\mathbf{p}} := \cup_{k=1}^{\infty} \{\mathbf{x} : F^k(\mathbf{x}) = \mathbf{p}\}$$

where F^k is the k-fold composition of the map F. More generally, define the basin of attraction B_A of any attractor (1) as

$$B_A := \cup_{k=1}^{\infty} \{\mathbf{x} : F^k(\mathbf{x}) \in A\}. \tag{2}$$

Clearly, if $\mathbf{p}$ is a steady state, then $A_{\mathbf{x}} = \{\mathbf{p}\}$ for all $\mathbf{x} \in B_{\mathbf{p}}$, the basin of attraction of $\mathbf{p}$.

The invariance of the attractor follows immediately from the definition (2.1).

Proposition 2. *If* $\mathbf{y} \in A_{\mathbf{x}}$, *then* $F(\mathbf{y}) \in A_{\mathbf{x}}$.

Proof. If $\mathbf{y} \in A_{\mathbf{x}}$ then there is an increasing sequence $n_1 < n_2 < n_3 < \cdots$ with $\mathbf{y} = F^{n_1}(\mathbf{x}) = F^{n_2}(\mathbf{x}) = \cdots$, and this then implies that $F(\mathbf{y}) = F^{n_k+1}(\mathbf{x}), k \geq 1$ and so $F(\mathbf{y}) \in A_{\mathbf{x}}$. □

Proposition 3. *All attractor sets* $A_{\mathbf{x}}$ *are fixed points or cycles.*

Proof. It is enough to show that the map F does not leave invariant any strict subset of $A_{\mathbf{x}}$. Let $B \subset A_{\mathbf{x}}$ with $F(B) \subset B$. If $\mathbf{y} \in A_{\mathbf{x}}$, then $\mathbf{y} = F^{n_k}(\mathbf{x}), n_1 < n_2 < n_3 < \ldots$ and similarly if $\mathbf{b} \in B$ then $\mathbf{b} = F^{m_k}(\mathbf{x})$. This means that $\mathbf{y} = F^{n_k - m_1}(\mathbf{b}), n_k > m_1$ and thus $\mathbf{y} \in A_{\mathbf{b}}$. By Proposition 1, it follows that $A_{\mathbf{x}} = A_{\mathbf{b}} \subset B$, where the last containment follows by the invariance of B. Thus any invariant subset B of $A_{\mathbf{x}}$ must be all of $A_{\mathbf{x}}$. □

Proposition 4. *For any basin of attraction* B_A, $\mathbf{x} \in B_A$ *if and only if* $A_{\mathbf{x}} = A$.

Proof. Suppose $\mathbf{x} \in B_A$. Then there is a $k \geq 1$ with $F^k(\mathbf{x}) \in A$. Since A is invariant, it follows that $F^n(\mathbf{x}) \in A$ for all $n \geq k$, and hence $A_{\mathbf{x}} = \cap_{k \geq 1} \cup_{n \geq k} F^n(\mathbf{x}) \subset A$ – then by Proposition 1 $A_{\mathbf{x}} = A$.

Conversely, suppose $A_{\mathbf{x}} = A$. To show that $\mathbf{x} \in B_A (= B_{A_{\mathbf{x}}})$, it is sufficient to show that $F_{n_k}(\mathbf{x}) = \mathbf{x}_0$ for a sequence $n_1 < n_2 < n_3 < \cdots$ and any point $\mathbf{x}_0$, because the point $\mathbf{x}_0$ must then be in $A_{\mathbf{x}}$. But this property is immediate since the infinite sequence $F^n(\mathbf{x})$ in the finite set $\{0,1\}^d$ must visit some point $\mathbf{x}_0$ an infinite number of times. □

Proposition 4 clarifies that a point $\mathbf{x}$ will hit its attracting set at some finite time (unlike the situation in continuous dynamics), and this is used in the algorithm of Theorem 1.

Example 1. An example of dynamics with limiting cycles is given in Table 1 of (Saadatpour et al., 2010) for a 13-node subnetwork of a guard cell ABA signalling network. With 13 nodes each getting an indeterminate $s_1, \ldots, s_{13}$, the dynamics are

$$F_1 = s_{11}, F_2 = s_1, F_3 = s_2, F_4 = s_2, F_5 = s_4, F_6 = s_3, F_7 = s_{11}, F_8 = s_7,$$
$$F_9 = (s_5 \cdot s_6) + s_8 - s_5 \cdot s_6 \cdot s_8, F_{10} = s_{11}, F_{11} = s_9 \cdot (1 - s_{10}), F_{12} = 1 - s_{11}, F_{13} = s_{11}.$$

There is one fixed point 0000000000010 with basin of attraction counting 108 points, and two attractors in the form of limit cycles of size 4, given by

1000001011001

0100000100010

0011000010010

0000110000110

1100001111001

0111000110010

0011110010110

1000111011101

The two basins of attraction have sizes 1704 and 6380.

3 One-parameter dispersion model

Idealized, simplified interaction and regulatory rules F are useful conceptual tools. However these dynamics usually do not fit data for several reasons: 1) the actual multivariate time series do not exactly follow the dynamics because the rules are only approximate, 2) there is noise in the original continuous measurements, which

leads to corrupted binary values in discretized data, 3) an intervention or experiment is deliberately stimulating or inhibiting the network to provide data for modeling. We may call data that is incompatible or inconsistent with a deterministic model *singular* data, as its probability or likelihood is 0. Comparing two idealized theories in this setting is our goal.

In this section we define a probability model that interpolates between pure iid noise and the exact deterministic dynamics. This will make the likelihood of the data positive, help account for uncertainties in measurement modeling, and then permit likelihood based methods of inference. We introduce a dispersion parameter in a way that is standard in generalized linear model theory and has some similarities with the categorical data version in (Diaconis et al., 1985).

For dispersion parameter ϕ, define a transition probability kernel on Ω by

$$K(\mathbf{x},\mathbf{y}) = \frac{e^{-\frac{1}{\phi}\|F(\mathbf{x})-\mathbf{y}\|^2}}{(1+e^{-1/\phi})^d},\ \phi \in (0,\infty). \tag{3}$$

When $\phi \to \infty$, the distribution approaches coin flipping for $\mathbf{y}$, and when $\phi \to 0$ it approaches the deterministic dynamics $\mathbf{y} = F(\mathbf{x})$. One may parametrize with $\theta = 1/\phi$ if desired, but using ϕ is more consistent with notation for dispersion parameters in exponential families where larger ϕ corresponds to more variance in the response.

Let μ denote a known initial probability distribution on Ω, giving probability distribution $P_{\mu,F,\phi}$ on Ω^{n+1}:

$$P_{\mu,F,\phi}(\mathbf{x}_{0:n} = (\mathbf{x}_0,\mathbf{x}_1,\ldots,\mathbf{x}_n)) = \mu(\mathbf{x}_0)\prod_{i=1}^{n} K(\mathbf{x}_{i-1},\mathbf{x}_i)$$

which simplifies to

$$P_{\mu,F,\phi}(\mathbf{x}_{0:n}) = \mu(\mathbf{x}_0)\frac{e^{-\frac{1}{\phi}\sum_{i=1}^{n}\|F(\mathbf{x}_{i-1})-\mathbf{x}_i\|^2}}{(1+e^{-1/\phi})^{dn}}. \tag{4}$$

We will consider ϕ to be a nuisance parameter, and the dynamics F to be the "parameter" of interest for testing.

To estimate ϕ (which is useful to determine how well the pure dynamics fit the data because a better map F is related to a smaller dispersion ϕ), note that we can solve explicitly for its maximum likelihood estimator $\hat{\phi}$:

$$\hat{p} := \frac{\sum_{i=1}^{n} ||F(\mathbf{x}_{i-1}) - \mathbf{x}_i||^2}{nd}$$
$$\frac{1}{\hat{\phi}} = \log(\frac{1-\hat{p}}{\hat{p}}),\ 0 < \hat{p} < 1/2.$$

The model (3) means that perturbations or errors occur with odds $e^{-1/\phi}$ homogeneously in time (index $i = 1,\ldots,n$) and space (coordinate indices $j = 1,\ldots,d$), and when they occur they are built into the process affecting future transitions (a state space model would be more appropriate for noisy observations where the true state is not randomly perturbed). This would roughly correspond to a situation where homogeneous interventions are made on a network to keep generating data for observation or reverse engineering. In experiments such as the hcc1954 data described in (Bender et al., 2011), there are many different interventions that affect the network in different ways, so homogeneity in the perturbations may be too idealized.

4 Sequential Importance Sampling

In this section we describe a sequential importance sampling algorithm for computing expectations on the initial state $\mathbf{x}_0 \in B_A \subset \Omega$, for basin of attraction B_A. The value of computations within basins of attraction is evident in work such as (Albert et al., 2003), (Saadatpour et al., 2010). The mathematical method is based on constructing the set of polynomials that vanish on the basin of attraction (its ideal), then sampling roots sequentially with a nonlinear version of back substitution. The algebraic tools are outlined in (Kreuzer et al., 2000) and (Riccomagno et al., 2000).

Before explaining the details, let us say how the approach in this section differs from existing methods for studying attractors, such as found in BoolNet (Müssel et

al., 2010). Rather than complete enumeration and listing of states in an attracting set, a process whose work grows exponentially in the number of dimensions or nodes d, the algebra constructs the polynomials that vanish on the attracting set. In many real examples, the polynomials are few and simple to understand. For example, in the signalling network of Example 7, each fixed point has an attracting basin of size 8192. Complete enumeration does not reveal that each is simply a cylinder set obtained by restricting three coordinates 1, 8, and 11, but the polynomial characterization shows sixteen polynomials only three of which say more than the states are binary. For the limit cycle in that example (the one of sixteen that was analyzed), just nineteen polynomials are needed, only three of which are nontrivial. While the algebra can in theory be hard or practically impossible, in real examples the standard polynomial basis typically has size on the order of d, rather than the 2^d states, and its computation is fast. Its use in sampling requires importance reweighting (Theorem 2), but that is a small inconvenience in return for the insight and memory efficiency.

Let A be an attractor of interest, possibly a limit cycle, and let μ have support on its basin of attraction B_A. We will use twice as many indeterminates as the number of coordinates d. Define the ring of polynomials $R := \mathbb{C}[s_1,\ldots,s_d,t_1,\ldots,t_d] = \mathbb{C}[\mathbf{s},\mathbf{t}]$, and define ideals

$$\begin{aligned}
I_{01} &= \langle s_1^2 - s_1, \ldots, s_d^2 - s_d, t_1^2 - t_1, \ldots, t_d^2 - t_d \rangle \\
F_{st} &= \langle F_1(\mathbf{s}) - t_1, F_2(\mathbf{s}) - t_2, \ldots, F_d(\mathbf{s}) - t_d \rangle \\
F_{ts} &= \langle F_1(\mathbf{t}) - s_1, F_2(\mathbf{t}) - s_2, \ldots, F_d(\mathbf{t}) - s_d \rangle \\
I_A &= \cap_{\mathbf{p}\in A} \langle t_1 - p_1, \ldots, t_d - p_d \rangle.
\end{aligned}$$

Define the ideal I_1 by

$$I_1 = (F_{st} + I_A + I_{01}) \cap \mathbb{C}[\mathbf{s}].$$

Define recursively a sequence of ideals $I_2, I_3, I_4, \ldots$ by

$$J = (F_{ts} + I_i + I_{01}) \cap \mathbb{C}[\mathbf{t}] \tag{5}$$

$$I_{i+1} = (F_{st} + J + I_{01}) \cap \mathbb{C}[\mathbf{s}], \quad i = 1,2,3,\ldots \tag{6}$$

Stop the iteration when $\dim R/(I_i + I_{01})$ repeats in order to get the polynomials that vanish on the basin of attraction B_A (see (Dinwoodie, 2012) for proofs in the case of a steady state and examples).

Theorem 1. *There exists* $i^\star < \infty$ *such that* $\dim R/(I_{i^\star} + I_{01}) = \dim R/(I_{i^\star+1} + I_{01})$, *and for such an integer*

$$I(B_A) = I_{i^\star}$$

as an ideal within $\mathbb{C}[\mathbf{s}]$.

Proof. Here we only sketch the main steps. Observe first that I_A is the ideal of the attracting set A containing a finite number of points $\mathbf{p} = (p_1, \ldots, p_d)$. The elimination ideal I_1 is the ideal for the points $\mathbf{x}$ that reach A in one time step, using indeterminates $\mathbf{s}$. Then the following ideal J is for the points $\mathbf{x}$ that reach A in two time steps, using indeterminates $\mathbf{t}$. The elimination operation does not add unwanted partial solutions (solutions that do not match up with points that reach A from the previous time step), because the Extension Theorem applies when the univariate polynomials in the ideal I_{01} are added. All the ideals $I_i + I_{01}$ are radical and 0-dimensional so the dimension of the vector space $R/(I_i + I_{01})$ counts solutions. When the number of solutions stops increasing, then the procedure has found all points that will reach A in forward iterations of F. □

Now map the polynomials in $I_{i^\star}$ to $\mathbb{C}[\mathbf{s}]$ in the obvious way ($s_j \to s_j, t_j \to 0$) so I_{B_A} is the ideal of polynomials in $s_1, \ldots, s_d$ that vanish on the basin of attraction B_A, the ideal of the variety. Note also that the univariate polynomials $s_j^2 - s_j$ all belong to the ideal $I(B_A)$.

For sequential sampling from B_A we adapt the “backward" method from (Dinwoodie, 2011). Let

$$\{f_1, \ldots, f_g\} \tag{7}$$

be a lexicographic Gröbner basis for $I(B_A)$ with indeterminate ordering $s_1 > s_2 > \cdots > s_d$.

The proposal distribution, from which we generate an iid sample of size N in $B_A \subset \Omega$, will be close to uniform. The proposal distribution q will be expressed as a product of successive conditional distributions

$$q(\mathbf{x}) = q_d(x_d) \cdot q_{d-1}(x_{d-1}|x_d) \cdot q_{d-2}(x_{d-2}|x_d, x_{d-1}) \cdots q_1(x_1|x_d, \ldots, x_2)$$

just as a random point $\mathbf{X}_k := (X_{k,1}, X_{k,2}, \ldots, X_{k,d}) \in B_A$ will be generated sequentially: $X_{k,d}, X_{k,d-1}, \ldots, X_{k,1}$, $k = 1, \ldots, n$.

The unnormalized weights w_k are defined by $w_k = \mu^\star(\mathbf{X}_k)/q(\mathbf{X}_k)$, where $\mu^\star$ is a convenient possibly unnormalized version of the probability distribution μ. The SIS Monte Carlo estimate for $E_{B_A}(f(\mathbf{X}))$ is given by

$$\hat{E}_{B_A}(f(\mathbf{X})) := \frac{1}{N} \sum_{k=1}^{N} f(\mathbf{X}_k) \frac{w_k}{\bar{w}}. \tag{8}$$

The law of large numbers says that

$$\bar{w} = \frac{1}{N} \sum_{k=1}^{N} \frac{\mu^\star(\mathbf{X}_k)}{q(\mathbf{X}_k)} \to \sum_{\mathbf{x} \in B_A} \frac{\mu^\star(\mathbf{x})}{q(\mathbf{x})} q(\mathbf{x}) = \mu^\star(B_A) \tag{9}$$

which implies the consistency of the estimator $\hat{E}_{B_A}(f(\mathbf{X}))$:

$$\hat{E}_{B_A}(f(\mathbf{X})) \to \frac{1}{\mu^\star(B_A)} \sum_{\mathbf{x} \in B_A} f(\mathbf{x}) \frac{\mu^\star(\mathbf{x})}{q(\mathbf{x})} q(\mathbf{x}) = \sum_{\mathbf{x} \in B_A} f(\mathbf{x}) \mu(\mathbf{x}) = E_{B_A}(f(\mathbf{X})). \tag{10}$$

When $\mu^\star$ is the unnormalized constant 1, then SIS can be used for approximate counting as is well-known: $\bar{w} \to |B_A|$.

The SIS procedure for sampling from a nonempty B_A using an initial Groebner basis computation is described next.

(SIS) Sequential Importance Sampling on B_A:

1. Compute a reduced lexicographic Groebner basis for $I(B_A)$ with variable order $s_1 > s_2 > \cdots > s_d$ in $\mathbb{C}[\mathbf{s}]$.

2. For sample size N, let the index k run from 1 to N:

(a) Using the polynomials from the lex basis that only involve s_d, determine which of $\{0,1\}$ solve the system and let $n_d \in \{1,2\}$ be the number of values in $\{0,1\}$ that solve the equations. Then uniformly sample X_d from the set of roots, and let $q_d(X_d) = 1/n_d$.

(b) Continue for indices $j = 1,\ldots,d-1$ to count (by substitution of 0 and 1) the number of solutions n_{d-j} to the equations in the lex basis that involve variables $s_{d-j},\ldots,s_d$, with $s_{d-j+1} = X_{d-j+1},\ldots,s_d = X_d$. Choose X_{d-j} uniformly from the n_{d-j} solutions, and set $q_{d-j}(X_{d-j}|X_{d-j+1},\ldots,X_d) = 1/n_{d-j}$.

(c) Complete $\mathbf{X} = (X_1,\ldots,X_d) \in B_A$ when X_1 is chosen and $q_1(X_1|X_d,\ldots,X_2)$ is computed.

(d) Set $\mathbf{X}_k = (X_1,\ldots,X_d)$ and $l_k = -\log(q_d(X_d)) - \cdots - \log(q_1(X_1|X_d,\ldots,X_2))$.

The following result is from (Dinwoodie, 2011), and is essentially an application of the Extension Theorem (Cox et al., 1998), using the elements of I_{01} to satisfy certain technical conditions, and accounting for the proposal probabilities. While the lexicographic Gröbner basis is considered computationally hard, the nature of the equations in biological networks usually gives tractable systems.

Theorem 2. *Sequential importance sampling in (SIS) above always produces an element $\mathbf{X}_k \in B_A$ if $B_A \neq \emptyset$, and when μ is constant on B_A the importance sampling weights w_k are*

$$w_k = e^{l_k}.$$

Example 2. To make the method above concrete, consider a simple example on $d = 2$ nodes, where the dynamics are $F_1(x_1,x_2) = F_2(x_1,x_2) = x_1x_2$. There are two fixed points, and attractor $A = \{00\}$ has basin of attraction equal to $B_{00} = \{00,01,10\}$. Its ideal is generated by lex Gröbner basis $\{s_1^2 - s_1, s_1 \cdot s_2, s_2^2 - s_2\}$ which is the key to sampling. There is one equation that involves only the last indeterminate s_2, and it is solved by both 0,1, so $n_d = n_2 = 2$. Suppose we choose 0 for its value, giving partial solution $*0$. Then replacing s_2 by 0 in the other

equations gives equations $s_1^2 - s_1, 0$, so again two choices are possible and $n_1 = 2$. Thus the weights w on 00 and 10 are both 4, the reciprocal of $\frac{1}{2} \cdot \frac{1}{2}$. On the other hand if the first choice was $x_2 = 1$, giving partial solution $*1$, the updated equations become $s_1^2 - s_1, s_1$. These are only solved by $s_1 = 0$ for complete solution 01 with weight 2, the reciprocal of $\frac{1}{2}\frac{1}{1}$. The sequential sampling will generate solutions 00, 01, 10, with frequencies proportional to 1/4, 1/2, 1/4, and the weights are the reciprocals.

Thus an expectation of a function f with the respect to the uniform distribution $\mu^\star = 1$ on B_{00} is computed as

$$E_{B_{00}}(f(\mathbf{X})) := \sum_{\mathbf{x} \in B_{00}} \frac{f(\mathbf{x})}{3} = f(00)\frac{w_{00}}{3} \cdot \frac{1}{4} + f(01)\frac{w_{01}}{3} \cdot \frac{1}{2} + f(10)\frac{w_{10}}{3} \cdot \frac{1}{4}$$

where the weights are given by $w_{00} = 4, w_{01} = 2, w_{10} = 4$, and the normalizing 3 corresponds to the average $\bar{w}$ from the expectation of the weights

$$\bar{w} \approx 3 = w_{00} \cdot \frac{1}{4} + w_{01} \cdot \frac{1}{2} + w_{10} \cdot \frac{1}{4}.$$

Then it is clear that the reweighting of the integrand f compensates for the unequal frequencies from the sampling procedure.

5 Exact Conditional Hypothesis Tests

We use the term “exact test" in the sense that the size α of the test is guaranteed to be as advertised – it does not come from asymptotic results with unknown convergence rates possibly not uniform over the parameter ϕ. The technical proof of the exactness is stated in Proposition 5 below. This result is generally not stated but exists as a folk theorem (Guo et al., 1992, p. 363). We state it completely to clarify the p-value formula and to show that the conditioning is not so much a Bayesian approach as one which makes a rejection region by considering each level set of a sufficient statistic.

The probability model of Section 3 gives likelihood function $L_{F,\phi}$ in the two unknown parameters F, ϕ of the form:

$$L_{F,\phi} := P_{\mu,F,\phi}(\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_n) = \mu(\mathbf{x}_0) \frac{e^{-\frac{1}{\phi}\sum_{i=1}^{n} \|F(\mathbf{x}_{i-1}) - \mathbf{x}_i\|^2}}{(1+e^{-1/\phi})^{dn}}.$$

Let $T_F(\mathbf{x}_{0:n}) := \sum_{i=1}^{n} \|F(\mathbf{x}_{i-1}) - \mathbf{x}_i\|^2$ measure the distance between ideal dynamics F and data $\mathbf{x}_{0:n}$. The data $\mathbf{x}^0_{0:n}$ is singular with respect to the ideal dynamics whenever $T_F(\mathbf{x}^0_{0:n}) > 0$. T_F is a sufficient statistic for ϕ. Then the conditional distribution on Ω^{n+1} given $T_F = t$ is proportional to $\mu(\mathbf{x}_0)$:

$$P_{\mu,F,\phi}\{\mathbf{x}_{0:n} \mid T_F(\mathbf{x}_{0:n}) = t\} = \frac{\mu(\mathbf{x}_0)}{\sum_{T_F(\mathbf{y}_{0:n})=t} \mu(\mathbf{y}_0)} \propto \mu(\mathbf{x}_0). \tag{11}$$

We will be interested in initial distributions μ that are supported on certain attractors.

Suppose the dynamics F could be one of two choices, G_0 or G_1 giving hypotheses

$$H_0 : F = G_0 \tag{12}$$

$$H_1 : F = G_1 \tag{13}$$

with unknown nuisance parameter $\phi \in (0, \infty)$.

A likelihood ratio test might be best if ϕ were known, but there are two practical difficulties: how to calibrate the test statistic for size α, and dealing with ϕ the unknown dispersion parameter. Note that the assumptions of Wilks' theorem that give a χ^2 asymptotic distribution for the likelihood ratio statistic are not satisfied (Bickel et al., 2007, p. 395). Another point here is that in maximizing a likelihood ratio $\max_\phi L_{G_1,\phi} / \max_\phi L_{G_0,\phi}$, the two maximizers $\hat{\phi}$ could be different, and the larger $\hat{\phi}$ should be considered as evidence against the corresponding dynamics G_1 or G_0. This information would not be considered in a standard likelihood ratio procedure but posterior densities of ϕ would be useful. Conditional inference handles both problems of exactness and unknown ϕ, but it is necessary to be able to compute in fibers $\{\mathbf{x}_0 \in B_A\} \cap \{T_{G_0} = t\}$. For this we use sequential importance sampling as described below.

The conditional p-value V can be used as a test statistic to give a size α test in the traditional frequentist sense. This is because the V test statistic cuts out a rejection subset $\{V \leq \alpha\}$ of size at most α (not depending on the nuisance parameter) from each level set of the sufficient statistic, and the parametric distribution just weights the various level sets differently depending on ϕ. A case study for practical issues of conditional p-values for categorical data is (Guo et al., 1992), and the proposition below is essentially in (Casella et al., 2002, p. 399).

Let $T = T_{G_0} - T_{G_1}$ and define the p-value test statistic V on observed data $\mathbf{x}^0_{0:n}$ by

$$V = V(\mathbf{x}^0_{0:n}) := P_{\mu,G_0,\phi}(T \geq T(\mathbf{x}^0_{0:n}) \mid T_{G_0} = T_{G_0}(\mathbf{x}^0_{0:n})) \tag{14}$$

which is a conditional p-value using the conditional likelihood ratio.

Proposition 5. *With p-value V defined above, the test that rejects H_0 when $V \leq \alpha$ has size at most α for any $0 < \alpha < 1$ regardless of ϕ.*

Proof. Suppose the true dynamics are given by G_0, so the probability distribution on Ω^{n+1} is $P_{\mu,G_0,\phi}$. Then the conditional distribution given $T_{G_0} = t$ is proportional to $\mu(\mathbf{x}_0)$. The test statistic V defines a rejection region R given by

$$R = \cup_{t \geq 0} R_t$$
$$R_t := \{\mathbf{x}^0_{0:n} \in T^{-1}_{G_0}(t) : V(\mathbf{x}^0_{0:n}) \leq \alpha\} = \{\mathbf{x}^0_{0:n} \in T^{-1}_{G_0}(t) : P_{\mu,G_0,\phi}(T \geq T(\mathbf{x}^0_{0:n}) \mid T_{G_0} = t) \leq \alpha\}.$$

Now using the mutual exclusiveness of the R_t, we get

$$P_{\mu,G_0,\phi}(R) = \sum_{t \geq 0} P_{\mu,G_0,\phi}(R_t \mid T_{G_0} = t) \cdot P_{\mu,G_0,\phi}(T_{G_0} = t)$$

and it is sufficient to show $P_{\mu,G_0,\phi}(R_t \mid T_{G_0} = t) \leq \alpha$ for each t. This is in fact a standard result for discrete random variables put into their own cdf, a slight variation on the continuous version where the resulting distribution is exactly uniform.

To simplify notation, fix t and let π denote the conditional mass function of $\mathbf{x}_{0:n} \in R_t$ and let g_t denote the mass function of T_{G_0} using the conditional distribu-

tion on R_t, and let x denote a trajectory $\mathbf{x}^0_{0:n}$. Then

$$\begin{aligned}
P_{\mu,G_0,\phi}(R_t \mid T_{G_0} = t) &= P_{\mu,G_0,\phi}\{\mathbf{x}^0_{0:n} \in T_{G_0}^{-1}(t) : P_{\mu,G_0,\phi}(T \geq T(\mathbf{x}^0_{0:n}) \mid T_{G_0} = t) \leq \alpha\} \\
&= \sum_{x:\sum_{t\geq t(x)} g_t \leq \alpha} \pi(x) \\
&= \sum_{s=0}^{\infty} I_{\{s:\sum_{t=s}^{\infty} g_t \leq \alpha\}} g_s \\
&= \sum_{s\in[s_\alpha,\infty)} g_s, \qquad [s_\alpha,\infty) := \{s : \sum_{t=s}^{\infty} g_t \leq \alpha\} \\
&\leq \alpha.
\end{aligned}$$

□

We now compute V with a Monte Carlo method that uses the SIS method of Section 4 for sampling B_A combined with a sampling method on $\{T_{G_0} = t\}$. For simplicity, the initial distribution μ will be uniform on basin B_A with unnormalized $\mu^\star = 1$. Recall that $T := T_{G_0} - T_{G_1}$ is defined before (14) and $\mathbf{x}^0_{0:n}$ is the actual data.

(MC) Monte Carlo Exact Test in $\{T_{G_0} = t\} \cap \{\mathbf{x}_0 \in B_A\}$:

1. Do (SIS) in B_A with dynamics $F = G_0$ to get an iid sample $\mathbf{X}_k \in B_A$ with weights w_k, $k = 1,\ldots,N$.

2. For each $k = 1,\ldots,N$:

   (a) Sample uniformly a subset S of size t from index set $\{(i,j), i = 1,\ldots,n, j = 1,\ldots,d\}$, set $S_i = \{j : (i,j) \in S\}$.

   (b) For $i = 1,\ldots,n$, set $\mathbf{x}_i = G_0(\mathbf{x}_{i-1}) \oplus \mathbf{1}_{S_i}$, with $\mathbf{x}_0 = \mathbf{X}_k$ and addition modulo 2 to switch the value $0 \leftrightarrow 1$.

   (c) Set $\mathbf{x}^k_{0:n} = (\mathbf{X}_k, \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$ to be the concatenation with $\mathbf{x}_0 = \mathbf{X}_k$.

3. Compute the estimator for p-value V defined at (14) by

$$\hat{V} = \frac{1}{N}\sum_{k=1}^{N} I_{\{T\geq T(\mathbf{x}^0_{0:n})\}}(\mathbf{x}^k_{0:n})\frac{w_k}{\bar{w}}. \tag{15}$$

Note that for each $\mathbf{x}^k_{0:n}$ constructed in (MC) above,

$$T_{G_0}(\mathbf{x}^k_{0:n}) = \sum_{i=1}^{n} \|G_0(\mathbf{x}^k_{i-1}) - \mathbf{x}^k_i\|^2 = \sum_{i=1}^{n} |S_i| = t.$$

Theorem 3. *The estimator $\hat{V}$ from Monte Carlo sampling (MC) converges to V as $N \to \infty$ when μ is uniform on B_A.*

Proof. As the Monte Carlo sample size $N \to \infty$,

$$\begin{aligned}
\hat{V} &= \frac{1}{N}\sum_{k=1}^{N} I_{\{T \geq T(\mathbf{x}^0_{0:n})\}}(\mathbf{x}^k_{0:n})\frac{w_k}{\bar{w}} \\
&\to \frac{1}{\mu^\star(B_A)} E[I_{\{T \geq T(\mathbf{x}^0_{0:n})\}}(\mathbf{x}_{0:n}) w(\mathbf{x}_0)] \\
&= \frac{1}{\mu^\star(B_A)} E\left[w(\mathbf{x}_0) E[I_{\{T \geq T(\mathbf{x}^0_{0:n})\}}(\mathbf{x}_{0:n}) \mid \mathbf{x}_0]\right] \\
&= \frac{1}{\mu^\star(B_A)} E\left[w(\mathbf{x}_0) \frac{|\{\mathbf{x}_{1:n} : T(\mathbf{x}_{0:n}) \geq T(\mathbf{x}^0_{0:n}), T_{G_0}(\mathbf{x}_{0:n}) = t\}|}{|\{\mathbf{x}_{1:n} : T_{G_0}(\mathbf{x}_{0:n}) = t\}|}\right] \\
&= \frac{1}{\mu^\star(B_A)} \sum_{\mathbf{x}_0 \in B_A} \frac{\mu^\star(\mathbf{x}_0)}{q(\mathbf{x}_0)} q(\mathbf{x}_0) \frac{|\{\mathbf{x}_{1:n} : T(\mathbf{x}_{0:n}) \geq T(\mathbf{x}^0_{0:n}), T_{G_0}(\mathbf{x}_{0:n}) = t\}|}{|\{\mathbf{x}_{1:n} : T_{G_0}(\mathbf{x}_{0:n}) = t\}|} \\
&= \frac{1}{|B_A|} \sum_{\mathbf{x}_0 \in B_A} \frac{1}{q(\mathbf{x}_0)} q(\mathbf{x}_0) \frac{|\{\mathbf{x}_{1:n} : T(\mathbf{x}_{0:n}) \geq T(\mathbf{x}^0_{0:n}), T_{G_0}(\mathbf{x}_{0:n}) = t\}|}{|\{\mathbf{x}_{1:n} : T_{G_0}(\mathbf{x}_{0:n}) = t\}|} \\
&= P_{\mu,G_0,\phi}(T \geq T(\mathbf{x}^0_{0:n}) \mid T_{G_0} = T_{G_0}(\mathbf{x}^0_{0:n})) \\
&= V.
\end{aligned}$$

□

6 Examples

The network examples are ABA signalling and a cancer cell network.

Example 3. Consider the network for stomatal closure from Example 1, and consider the first run from the Abscisic Acid Signaling Network Data Set at the UCI Machine Learning Repository (Frank et al., 2010) in Table 1. This data was simulated with the dynamics of Example 1 in an asynchronous fashion, meaning

that a coordinate j was chosen randomly, then that coordinate map F_j is applied to update that one coordinate while the others remain unchanged. The initial state (marked as time 1) was simulated uniformly over all states. This transition scheme has the same steady states as the pure dynamics, but introduces randomness differently than the perturbations of model (4) and slows the process by a factor of $1/d$ approximately. Therefore the data is quite different than what would be typical for model (4).

Table 1: ABA Signalling Data

	NOS	NO	GC	ADPRc	cADPR	cGMP	PLC	IP3	CIS	CaATPase	Ca	KAP	KEV
step+1	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}
1	0	1	1	0	0	1	0	0	0	1	1	0	1
2	1	1	1	1	1	1	0	0	0	0	0	0	1
3	0	0	0	0	1	0	0	0	1	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0	0	0	0	0

While the initial state above is in the basin of attraction of the smaller cycle limit described in Example 1, we will take μ to be uniform as was done in the original simulation. Then importance sampling is not needed as exact simulation of μ is straightforward . Let G_1 be a competing theory with map 9 given by $s_5 \cdot s_6$ with no appearance of s_8, while the null model G_0 is exactly map F from Example 1. Five Monte Carlo p-value computations with $N = 10000$ gave an average of 0.021, with standard error 0.001. Therefore this data would probably lead to rejection of the null model.

Example 4. Starting from the same initial state as Example 3 but running pure dynamics with no noise ($\phi = 0$), the data becomes

0 1 1 0 0 1 0 0 0 1 1 0 1
1 0 1 1 0 1 1 0 0 1 0 0 1
0 1 0 0 1 1 0 1 0 0 0 1 0
0 0 1 1 0 0 0 0 1 0 0 1 0
0 0 0 0 1 1 0 0 0 0 1 1 0
1 0 0 0 0 0 1 0 1 1 0 0 1
0 1 0 0 0 0 0 1 0 0 0 1 0
0 0 1 1 0 0 0 0 1 0 0 1 0
0 0 0 0 1 1 0 0 0 0 1 1 0
1 0 0 0 0 0 1 0 1 1 0 0 1
0 1 0 0 0 0 0 1 0 0 0 1 0
0 0 1 1 0 0 0 0 1 0 0 1 0
0 0 0 0 1 1 0 0 0 0 1 1 0
1 0 0 0 0 0 1 0 1 1 0 0 1
0 1 0 0 0 0 0 1 0 0 0 1 0
0 0 1 1 0 0 0 0 1 0 0 1 0
0 0 0 0 1 1 0 0 0 0 1 1 0
1 0 0 0 0 0 1 0 1 1 0 0 1
0 1 0 0 0 0 0 1 0 0 0 1 0
0 0 1 1 0 0 0 0 1 0 0 1 0
0 0 0 0 1 1 0 0 0 0 1 1 0

With the same maps G_0 and G_1 as Example 1, five runs of $N = 10000$ gave a mean for the p-value estimate of 0.104, with standard error 0.001, values normally consistent with keeping the null dynamics.

Our third example demonstrates the feasibility of the algebraic computations required to condition on a particular basin of attraction.

Example 5. Consider again the network for stomatal closure from Example 1, and hypothetically suppose $t = 0$, meaning the model $G_0 = F$ fits the data perfectly, and suppose $n = 1$ for one transition. Let G_1 again be the competing theory with map 9 given by $s_5 \cdot s_6$ with no appearance of s_8. Suppose the initial distribution μ is uniform on the basin of attraction B_A of size 6380 corresponding to the second

limit cycle.

Then the p-value V is simply the fraction of initial states $\mathbf{x}_0 \in B_A$ where $G_1(\mathbf{x}_0) \neq G_0(\mathbf{x}_0)$, a case treated algebraically in (Dinwoodie, 2012), and the answer is exactly 1 - 3740/6380 = .41. Employing the SIS method of Section 4, there is a lex Gröbner basis for sampling the set B_A of 19 polynomials which is found easily in Singular. Sampling with $N = 10000$ showed an estimated size $\hat{B}_A = 6413.3$ on one run for example, from the average of the importance sampling weights $\bar{w}$ (and a cv^2 value of approximately .12 indicates reasonable efficiency of importance sampling relative to perfect sampling, see (Liu, 2001)). A Monte Carlo estimate of V on five runs with $N = 10000$ is 0.414, which compares with the exact value of .41. The standard error on the five runs was 0.003, giving confidence interval $0.414 \pm 2 \cdot .003$ containing the true value.

Example 6. Consider again the network for stomatal closure from Example 1 with G_1 as above in Example 5. We generated data starting from an initial point in the larger basin of attraction of the second cycle, using the dynamics G_1 with zero random perturbations (zero perturbations are likely with $\phi < .1$ when $n = 20$ and $d = 13$) .

1 1 0 0 0 0 1 1 1 1 0 0 1
0 1 1 1 0 0 0 1 0 0 0 1 0
0 0 1 1 1 1 0 0 0 0 0 1 0
0 0 0 0 1 1 0 0 1 0 0 1 0
0 0 0 0 0 0 0 0 1 0 1 1 0
1 0 0 0 0 0 1 0 0 1 1 0 1
1 1 0 0 0 0 1 1 0 1 0 0 1
0 1 1 1 0 0 0 1 0 0 0 1 0
0 0 1 1 1 1 0 0 0 0 0 1 0
0 0 0 0 1 1 0 0 1 0 0 1 0
0 0 0 0 0 0 0 0 1 0 1 1 0
1 0 0 0 0 0 1 0 0 1 1 0 1
1 1 0 0 0 0 1 1 0 1 0 0 1
0 1 1 1 0 0 0 1 0 0 0 1 0
0 0 1 1 1 1 0 0 0 0 0 1 0

0 0 0 0 1 1 0 0 1 0 0 1 0
0 0 0 0 0 0 0 0 1 0 1 1 0
1 0 0 0 0 0 1 0 0 1 1 0 1
1 1 0 0 0 0 1 1 0 1 0 0 1
0 1 1 1 0 0 0 1 0 0 0 1 0
0 0 1 1 1 1 0 0 0 0 0 1 0

One can see that the map G_1 takes the starting state out of the limit cycle for G_0. Five runs of algorithm (MC) gave an estimated p-value of 0.020, with standard error 0.0005. Such values would normally lead to rejection of the incorrect dynamics G_0.

Example 7. Here we consider two 16-node signalling models for the cancer cell network of (Bender et al., 2011). We show that the exact test does not reject one in favor of the other using the hcc1954 signalling data in the the R package `ddepn` (Bender et al., 2011).

The hcc1954 data is described in (Bender et al., 2010). We use the EGF experiment, which has three real time measurements at 0, 4, 8, 12, 16, 20, 30, 40 50, 60 minutes. We first averaged the three replication values, then discretized to two states with the information-based method of (Scutari, 2010), giving time series

0 1 1 0 1 0 0 0 0 1 1 0 0 0 0 0
1 1 1 0 1 0 0 1 0 1 1 0 1 1 0 0
1 1 0 1 0 1 1 0 1 0 0 1 1 1 1 1
1 1 0 1 0 1 1 0 1 1 0 1 1 1 1 0
1 1 0 1 0 1 1 0 1 1 0 1 1 1 1 0
0 1 0 1 0 0 1 0 1 0 0 1 0 1 1 0
0 0 1 1 1 0 0 0 0 1 1 0 0 0 0 0
1 0 1 1 1 0 0 1 1 1 1 0 1 1 0 0
1 0 1 1 0 0 1 0 0 1 1 0 0 1 1 0
1 0 1 1 1 0 1 1 1 1 1 1 1 1 1 0

where the 16 column names are in the order of the proteins listed below, and each row corresponds to one time step.

The dynamics G_0 for the protein-signalling model are defined in Table 2, and are derived by logical disjunction of incoming nodes in the network, and an alternative model G_1 in Table 3 was learned with a Laplace prior (see Figure 6 of (Bender et al., 2011)).

Table 2: Cancer Cell Network Model

node	G_0 logical update	G_0 polynomial
1 EGF	EGF	x_1
2 ERBB2	EGF	x_1
3 ERK1/2	EGF	x_1
4 AKT	EGF	x_1
5 PDK1	ERBB3	x_{15}
6 MEK1/2	EGF	x_1
7 PLCg	EGF	x_1
8 PKC	PKC	x_8
9 P38	EGF or (not ERK1/2)	$x_1 + (1 - x_3) - x_1 \cdot (1 - x_3)$
10 SRC	ERBB3	x_{15}
11 mTOR	mTOR	x_{11}
12 P70	EGF or (not P38)	$x_1 + (1 - x_9) - x_1 \cdot (1 - x_9)$
13 GSK	not AKT	$1 - x_4$
14 PRAS	not ERBB4	$1 - x_{16}$
15 ERBB3	(not EGF) or PRAS	$(1 - x_1) + x_{14} - (1 - x_1) \cdot x_{14}$
16 ERBB4	PDK1	x_5

There are four steady states and sixteen limit cycles of size eight in the null

Table 3: Cancer Cell Network Alternative Model

node	G_1 logical update	G_1 logical formula
1 EGF	EGF	x_1
2 ERBB2	ERBB2	x_2
3 ERK1/2	MEK1/2	x_6
4 AKT	ERBB3 or (not PKC) or PDK1 or mTOR	$x_{15} \vee (!x_8) \vee x_5 \vee x_{11}$
5 PDK1	ERBB2 or ERBB3 or ERBB4	$x_2 \vee x_{15} \vee x_{16}$
6 MEK1/2	ERBB2 or ERBB3 or ERBB4	$x_2 \vee x_{15} \vee x_{16}$
7 PLCg	ERBB2 or ERBB3 or ERBB4	$x_2 \vee x_{15} \vee x_{16}$
8 PKC	PLCg	x_7
9 P38	ERK1/2	x_3
10 SRC	ERBB2 or ERBB3 or ERBB4	$x_2 \vee x_{15} \vee x_{16}$
11 mTOR	AKT or (not PRAS)	$x_4 \vee (!x_{14})$
12 P70	ERK1/2 or AKT or mTOR	$x_3 \vee x_4 \vee x_{11}$
13 GSK	not AKT	$!x_4$
14 PRAS	not AKT	$!x_4$
15 ERBB3	ERBB3	x_{15}
16 ERBB4	ERBB4	x_{16}

model G_0. The steady states are

0000100011001011

0000100111001011

0000100011101011

0000100111101011

with basins of attraction in the form of cylinders determined by coordinates 1, 8, and 11 and hence have size 8192. The data starts in one of the steady state basins because the first coordinate is 0, a condition which is immediate from

the polynomials. Importance sampling is not necessary for such sets, but the limit cycles are more interesting. The first as listed by (Müssel et al., 2010) includes the point 1111011010010000 and seven others that follow. The basin of attraction has 19 polynomials (reduced lexicographic basis), and counts 4096 points, a number which can be found by computing the vector space dimension with vdim in (Decker et al., 2011), or with BoolNet (Müssel et al., 2010), or by approximation with the average weights $\bar{w}$ from importance sampling of Section 4. For completeness and verification, these are the polynomials that define this basin of attraction: a quadratic polynomial $x^2 - x$ in x_j for each coordinate $j = 16, 15, 14, 13, 12, 10, 9, 7, 6, 5, 4, 3, 2$, three linear polynomials $x_{11}, x_8, x_1 - 1$ and three other quadratics: $x_{14} \cdot x_{15} - x_{14} \cdot x_{16} + x_{15} \cdot x_{16} - x_{15}, x_5 \cdot x_{15} + x_5 \cdot x_{16} - x_5 - x_{15} \cdot x_{16}, x_5 \cdot x_{14} + x_5 \cdot x_{16} - x_5 - x_{14} \cdot x_{16}$.

For comparing G_0 and G_1 on the four steady-state basins and the one cycle attracting basin we used five runs of size $N = 10000$. The results in Table 4 show not enough evidence to reject G_0 with this data.

Table 4: Analysis of hcc1954 Data on Five Attractor Basins

Initial Attractor Basin	Estimate of p-value	Standard Error
Steady State 1	.137	.002
Steady State 1	.141	.001
Steady State 3	.145	.001
Steady State 4	.142	.002
Cycle Limit 1	.184	.001

7 Conclusions and Further Problems

The method of conditional inference in Sections 4 and 5 makes rigorous inference possible for comparing non-ergodic dynamics F on discrete states. No asymptotics are used for calibrating the hypothesis test, rather a conditional p-value computation is done with a Monte Carlo sampling method on sets constrained by sufficient

statistics. While p-values are only one tool for inference, and are often criticized for many valid reasons, we believe it is worthwhile to have a method of inference that adheres to traditional notions of controlling Type I error probabilities, in addition to the wealth of learning algorithms available for discovery.

The probabilistic model (4) may not be rich enough to include realistic features of spatial and temporal inhomogeneity that arise when fusing data sets from experiments that perturb different parts of a network. A further model for investigation is an $n+d$ parameter model:

$$P_{\mu,F,\rho,\theta}(\mathbf{x}_{0:n}) = \mu(\mathbf{x}_0)\frac{e^{-\sum_{i=1}^{n}\|F(\mathbf{x}_{i-1})-\mathbf{x}_i\|^2_{\Phi_i}}}{\prod_{i=1}^{n}\prod_{j=1}^{d}(1+e^{-1/\phi_{ij}})}$$

with $1/\phi_{ij} = \rho_i + \theta_j, \Phi_i = (\phi_{i1},\ldots,\phi_{id})$ and norm $\|\mathbf{v}\|^2_{\Phi_i} := \sum_{j=1}^{d} v_{ij}^2/\phi_{ij}$, which gives richer spatial and temporal variability. The sufficient statistics now are the "row and column" sums of the error matrix $|F_j(\mathbf{x}_{i-1}) - \mathbf{x}_{i,j}|^2$. Uniform sampling can be done with sequential importance sampling (Chen et al., 2005).

Finally, rather than work on discretized data, which is necessary for simple Boolean models but raises further uncertainties in the discretization process, one may try a continuous Gaussian version of model (4), say

$$K(\mathbf{x},\mathbf{y}) = \frac{e^{-\frac{1}{2\phi}\|F(\mathbf{x})-\mathbf{y}\|^2}}{(2\pi\,\phi)^{d/2}},\ \ \mathbf{x},\mathbf{y}\in\mathbb{R}^d$$

or multiparameter variations. Further examples and applications to network models for Alzheimer's disease (Ramanan et al., 2012) would also be of interest.

References

Albert, R., and Othmer, H. G. (2003). The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in Drosophila melanogaster. *Journal of Theoretical Biology* **223**, 1-18.

Aoki, S., Hara, H., and Takemura, A. (2012). *Markov Bases in Algebraic Statistics.* Springer, New York.

Bender, C., Henjes, F., Fröhlich, H., Weimann, S., Korf, U., and Beissbarth, T. (2010). Dynamic deterministic effects propagation networks: learning signalling pathways from longitudinal protein array data. *Bioinformatics* **26**, 596-602.

Bender, C., Heyde, S., Henjes, F., Wiemann, S., Korf, U., and Beissbarth, T. (2011). Inferring signalling networks from longitudinal data using sampling based approaches in the R-package 'ddepn.' *BMC Bioinformatics* **12:**291.

Bickel, P. J., and Doksum, K. A. (2007). *Mathematical Statistics Volume I, Second Edition.* Pearson, Upper Saddle River NJ.

Casella, G., and Berger, R. L. (2002). *Statistical Inference, Second Edition.* Duxbury, Pacific Grove CA.

Chen, Y., Diaconis, P., Holmes, S. P., and Liu, J. S. (2005). Sequential Monte Carlo methods for statistical analysis of tables. *Journal of the American Statistical Association* **100**, 109-120.

Chen, Y., Dinwoodie, I. H., and Sullivant, S. (2006). Sequential importance sampling for multiway tables. *Annals of Statistics* **34**, 523-545.

Cox, D., Little, J., O'Shea, D. (1998). *Using Algebraic Geometry.* Springer, New York.

Decker, W., Greuel, G.-M., Pfister, G., and Schönemann, H. (2011). SINGULAR 3-1-3 — A computer algebra system for polynomial computations. http://www.singular.uni-kl.de.

Diaconis, P., and Efron, B. (1985). Testing for Independence in a Two-Way Table: New Interpretations of the Chi-Square Statistic. *Annals of Statistics* **13**, 845-874.

Dinwoodie, I. H. (2011). Sequential importance sampling of binary sequences. *Statistics and Computing* **22**, 53-63.

Dinwoodie, I. H. (2012). Conditional tests on basins of attraction with finite fields. *Methodology and Computing in Applied Probability*, to appear.

Drton, M., Sturmfels, B., and Sullivant, S. (2008). *Lectures on Algebraic Statistics.* Birkhäuser, Boston.

Frank, A. and Asuncion, A. (2010). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

Grayson, D. R. and Stillman, M. E. (2012). *Macaulay2, a software system for research in algebraic geometry.* Available at http://www.math.uiuc.edu/Macaulay2/.

Guo, S. W., and Thompson, E. A. (1992). Performing the Exact Test of Hardy-Weinberg Proportion for Multiple Alleles. *Biometrics* **48**, 361-372.

Hinkelmann, F., Murrugarra, D., Jarrah, A. S., and Laubenbacher, R. (2001). A Mathematical Framework for Agent Based Models of Complex Biological Networks. *Bulletin of Mathematical Biology* **73**, 1583 - 1602.

Klamt, S., Saez-Rodriquez, J., Lindquist, J. A., Simeoni, L., and Gilles, E. D. (2006). A methodology for the structural and functional analysis of signalling and regulatory networks. *BMC Bioinformatics* **7**, 1471-2105.

Kreuzer, M., and Robbiano, L. (2000). *Computational Commutative Algebra I.* Springer, New York.

Li, S., Assmann, S. M., Albert, R. (2006). Predicting Essential Components of Signal Transduction Networks: A Dynamic Model of Guard Cell Abscisic Acid Signaling. *PLoS Biology* **4**, 1733-1748.

Liu, J. S. (2001). *Monte Carlo Strategies in Scientific Computing.* Springer, New York.

Mendoza, L. (2006). A network model for the control of the differentiation process in Th cells. *BioSystems* **84**, 101-114.

Morris, M. K., Saez-Rodriguez, J., Sorger, P. K., and Lauffenburger, D. A. (2010). Logic-Based Models for the Analysis of Cell Signaling Networks. *Biochemistry* **49**, 3216-3224.

Müssel, C., Hopfensitz, M., and Kestler, H. A. (2010). BoolNet – an R package for generation, reconstruction and analysis of Boolean networks. *Bioinformatics* **26**, 1378-1380.

Ramanan, V. K., Kim, S., Holohan, K., Shen, L., Nho, K., Risacher, S. L., Foroud, T. M., Mukherjee, S., Crane, P. K., Aisen, P. S., Petersen, R. C., Weiner, M. W., and Saykin, A. J., (2012). Genome-wide pathway analysis of memory impairment in the Alzheimer's Disease Neuroimaging Initiative (ADNI) cohort implicates gene candidates, canonical pathways, and networks. *Brain Imaging and Behaviour* **6**, 634-648.

Riccomagno, E., Pistone, G., and Wynn, H. P. (2000). *Algebraic Statistics.* Chapman and Hall, London.

Saadatpour, A., Albert, I., and Albert, R. (2010). Attractor analysis of asynchronous Boolean models of signal transduction networks. *Journal of Theoretical Biology* **266**, 641-656.

Saez-Rodriguez, J., Simeoni, L., Lindquist, J. A., Hemenway, R., Bommhardt, U., Arndt, B., Haus, U.-U., Weismantel, R., Gilles, E. D., Klamt, S., and Schraven, B. (2007). A logical model provides insights into T cell receptor signalling. *PLOS Computational Biology* **3**, 1580-1590.

Saez-Rodriquez, J., Alexopoulos, L. G., Zhang, M., Morris, M., Lauffenburger, D. A., and Sorger, P. K. (2011). Comparing Signaling Networks between Normal and Transformed Hepatocytes Using Discrete Logical Models. *Cancer Research* **71**, 5400-5411.

Scutari, M. (2010). Learning Bayesian Networks with the bnlearn R Package, *Journal of Statistical Software* **35**, 1-22. http://www.jstatsoft.org/v35/i03/.

Stigler, B. (2006). Polynomial Dynamical Systems in Systems Biology. *AMS 2006 Proceedings of Symposia in Applied Mathematics* **64**, 59-84.

Thomas, R. (1973). Boolean formalization of genetic control circuits. *Journal of Theoretical Biology* **42**, 563 - 585.

Thomas, R. (1998). Laws for the dynamics of regulatory networks. *International Journal of Developmental Biology* **42**, 479-485.