Uncertainty Estimation

5

Frank Noé and John D. Chodera

As only a finite quantity of data can be collected for the construction of Markov state models, the parameters characterizing the model and any properties computed from it will always be statistically uncertain. This chapter is concerned with the quantification of this statistical uncertainty, and its use in validation of model quality and prediction of properties using the model. In the following sections we proceed along Refs. [2, 7, 11] which should be used for reference purposes.

5.1 Uncertainties in Transition Matrix Elements

We first consider the uncertainty in the transition matrix $\mathbf{T}(\tau)$ itself estimated from a finite quantity of data. It may be the case that the uncertainty in individual elements $T_{ij}(\tau)$ may be of interest, in which case standard errors or confidence intervals of these estimates may be sufficient tools to quantify the uncertainty.

For a transition matrix estimated without the detailed balance constraint, the expectation and variance of individual elements follow from well-known properties of the distribution of stochastic matrices [1]. These uncertainties do, however, depend on the choice of prior used in modeling the full posterior for the transition matrix (Sect. 4.4). Under a uniform prior, the expectation and variance of an individual element T_{ij} is given by,

$$\mathbb{E}[T_{ij}] = \frac{c_{ij}+1}{c_i+n} \equiv \bar{T}_{ij}, \tag{5.1}$$

$$\begin{aligned}\mathrm{Var}[T_{ij}] &= \frac{(c_{ij}+1)((c_i+n)-(c_{ij}+1))}{(c_i+n)^2((c_i+n)+1)} \\ &= \frac{\bar{T}_{ij}(1-\bar{T}_{ij})}{c_i+n+1},\end{aligned} \tag{5.2}$$

where c_{ij} and c_i are the elements and row sums, respectively, of the observed count matrix $\mathbf{C}^{\mathrm{obs}}$ (Sect. 4.2).

To see the effect that the choice of prior has on the computed uncertainties, consider a trajectory of a given molecular system which is analyzed with two different state space discretizations. Assume one discretization uses $n = 10$ states, and the other $n = 1000$. Assume that a lag time τ has been chosen which is identical and long enough to provide Markov models with small discretization error for both n (as suggested in Sect. 4.7). With a uniform prior ($c_{ij} = c_{ij}^{\mathrm{obs}}$), the posterior expectation $\bar{T}_{ij}$ would be different for the two discretizations: While in the $n = 10$ case we can get a distinct transition matrix estimation, in the $n = 1000$ case, most c_{ij} are probably zero and $c_i \ll n$, such that the expectation value would be biased towards the uninformative $T_{ij} \approx 1/n \pm 1/n$ matrix, and many observed transitions would be needed to overcome this bias. This behavior is

F. Noé (✉)
Freie Universität Berlin, Arnimallee 6, 14195 Berlin, Germany
e-mail: frank.noe@fu-berlin.de

J.D. Chodera
Memorial Sloan-Kettering Cancer Center, New York, NY 10065, USA
e-mail: choderaj@mskcc.org

G.R. Bowman et al. (eds.), *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation*, Advances in Experimental Medicine and Biology 797,
DOI 10.1007/978-94-007-7606-7_5, © Springer Science+Business Media Dordrecht 2014

undesirable. Thus, for uncertainty estimation it is suggested to use a prior which allows the observation data to have more impact also in the low-data regime.

On the other hand, the "null prior" [10] defined by

$$c_{ij}^{\text{prior}} \to -1 \quad \forall i, j \in \{1, \ldots, n\}, \tag{5.3}$$

leans to the other extreme. Under the null prior, the expectation and the variance of the marginalized posterior for a single T_{ij} become,

$$\bar{T}_{ij} = \mathbb{E}[T_{ij}] = \frac{c_{ij}^{\text{obs}}}{c_i^{\text{obs}}} = \hat{T}_{ij}, \tag{5.4}$$

$$\begin{aligned} \text{Var}(T_{ij}) &= \frac{c_{ij}^{\text{obs}}(c_i^{\text{obs}} - c_{ij}^{\text{obs}})}{(c_i^{\text{obs}})^2(c_i^{\text{obs}} + 1)} \\ &= \frac{\hat{T}_{ij}(1 - \hat{T}_{ij})}{c_i^{\text{obs}} + 1}. \end{aligned} \tag{5.5}$$

Thus, with a null prior, the expectation value is located at the likelihood maximum. Both expectation value and variance are independent of the number of discretization bins used. The variance of any T_{ij} asymptotically decays with the number of transitions out of the state i, which is expected for sampling expectations from the central limit theorem.

5.2 Uncertainties in Computed Properties

In practice, one is often not primarily interested in the uncertainties of the transition matrix elements themselves, but rather in the uncertainties in properties computed *from* the transition matrix. Here, we review two different approaches for this purpose.

- **Linear error perturbation** [4, 12, 13]. Here, the transition matrix posterior distribution is approximated by a multivariate Gaussian, and the property of interest—taken to be a function of the transition matrix or its eigenvalues and eigenvectors—is approximated by a first-order Taylor expansion about the center of this Gaussian. This results in a Gaussian distribution of the property of interest, with a mean and a covariance matrix that can be computed in terms of the count matrix $\mathbf{C}$. This approach has the advantage that error estimates and their rates of reduction for different sampling strategies can be computed through a direct procedure. As a result, it is convenient for situations where uncertainty estimates are used as part of an adaptive sampling procedure [4, 8, 9, 13]. The disadvantage of this approach is that the Gaussian approximation of the transition matrix posterior in only asymptotically correct, and can easily break down when few counts have been observed. In the low-data regime, the resulting Gaussian distribution for the property of interest often gives substantial probability to unphysical or meaningless values, such as when transition matrix elements T_{ij} are allowed to assume values outside the range $[0, 1]$). Moreover, the property of interest is approximated linearly which can introduce a significant error when this property is nonlinear.
- **Markov chain Monte Carlo (MCMC) sampling of transition matrices** [2, 6, 7]. Here, transition matrices are sampled from the posterior distribution, and the property of interest is computed for each of these and stored as samples from the posterior distribution of the property. This approach requires that the sampling procedure be run sufficiently long that good estimates of standard deviations or confidence intervals of the posterior distribution of the property of interest can be computed, which may be time-consuming. The advantage of this approach is that no assumptions are made concerning the functional form of the distribution or the property being computed. Furthermore, this approach can be straightforwardly applied to any function or property of transition matrices, including complex properties such as transition path distributions [10] without deriving the expressions necessary for the linear error perturbation analysis—often a cumbersome task. However, for large state spaces, the transition matrix $\mathbf{T}$ may grow so large as to make this procedure impractical.

5.3 Linear Error Propagation

We start again with the posterior distribution of row-stochastic transition matrices without the detailed balance constraint, given by Eq. (4.10). Defining a new matrix $\mathbf{U}$,

$$\mathbf{U} = [u_{ij}] = [c_{ij} + 1], \tag{5.6}$$

and using that the posterior probability $p(\mathbf{T} \mid \mathbf{C}^{\mathrm{obs}})$ implicitly contains the prior probabilities Eq. (4.10) can be rewritten as:

$$p(\mathbf{T} \mid \mathbf{C}) = p\big(\mathbf{T} \mid \mathbf{C}^{\mathrm{obs}}\big) \propto \prod_i \prod_j T_{ij}^{u_{ij}-1} \tag{5.7}$$

such that

$$\mathbf{T}_{i*} \sim \prod_i \mathrm{Dir}(\mathbf{u}_{i*}) \tag{5.8}$$

where $\mathrm{Dir}(\boldsymbol{\alpha})$ denotes the Dirichlet distribution, and $\boldsymbol{\theta} \sim \mathrm{Dir}(\boldsymbol{\alpha})$ implies that $\boldsymbol{\theta}$ is drawn from the distribution

$$p(\boldsymbol{\theta}) \propto \prod_i \theta_i^{\alpha_i - 1}. \tag{5.9}$$

Based on well-established properties of this distribution, and using the abbreviation $u_i = \sum_j u_{ij}$, the moments of $p(\mathbf{T} \mid \mathbf{C})$ can be directly computed,

$$\big[\mathbb{E}(\mathbf{T})\big]_{ij} = \frac{u_{ij}}{u_i} = \frac{c_{ij}+1}{c_i + n} = \bar{T}_{ij},$$

$$\big(\arg\max p(\mathbf{T}|\mathbf{C})\big)_{ij} = \frac{u_{ij}-1}{u_{ij}-n} = \frac{c_{ij}}{c_i} = \hat{T}_{ij},$$

$$\begin{aligned}\mathrm{Var}(T_{ij}) &= \frac{u_{ij}(u_i - u_{ij})}{u_i^2(u_i+1)} \\ &= \frac{\bar{T}_{ij}(1-\bar{T}_{ij})}{(u_i+1)} \\ &= \frac{\bar{T}_{ij}(1-\bar{T}_{ij})}{c_i + n - 1},\end{aligned}$$

$$\mathrm{Cov}(T_{ij}, T_{ik}) = \frac{-u_{ij}u_{ik}}{u_i^2(u_i+1)} \quad \forall j \neq k.$$

Next, we determine how the uncertainties given by the variances and covariances of the transition matrix elements propagate onto uncertainties of functions derived from transition matrices, such as eigenvalues. If we do not have constraints between different rows, such as are imposed by detailed balance, the rows can be treated as independent random vectors, and thus,

$$\mathrm{Cov}(T_{ij}, T_{lk}) = 0, \quad i \neq l. \tag{5.10}$$

We can thus define a covariance matrix $\boldsymbol{\Sigma}^{(i)}$ separately for each row i as,

$$\begin{aligned}\Sigma^{(i)}_{jk} &:= \mathrm{Cov}(T_{ij}, T_{ik}) \\ &= \frac{1}{u_i^2(u_i+1)}[u_i \delta_{jk} u_{ij} - u_{ij}u_{ik}] \\ &= \frac{1}{c_i}\big[\delta_{jk}\bar{T}_{ij} - \bar{T}_{ij}\bar{T}_{ik}^T\big],\end{aligned}$$

where δ is the Kronecker delta. Alternatively, we can write the covariance matrix $\boldsymbol{\Sigma}^{(i)}$ in vector notation,

$$\begin{aligned}\boldsymbol{\Sigma}^{(i)} &= \frac{1}{u_i^2(u_i+1)}\big[u_i\,\mathrm{diag}(\mathbf{u}_{i*}) - \mathbf{u}_{i*}(\mathbf{u}_{i*})^T\big] \\ &= \frac{1}{c_i}\big[\mathrm{diag}(\bar{\mathbf{T}}_{i*}) - \bar{\mathbf{T}}_{i*}(\bar{\mathbf{T}}_{i*})^T\big].\end{aligned}$$

In the limit of many observed transition counts, the covariance for the Dirichlet processes scales approximately with the inverse of the total number of counts in a row, c_i.

With a sufficient number of counts c_i in each row i, the Dirichlet process resembles a multivariate Gaussian distribution, and we can approximate it as such using the mean and variance computed above,

$$\mathbf{T}_{i*} \sim \mathrm{Normal}\big(\hat{\mathbf{T}}_{i*}, \boldsymbol{\Sigma}^{(i)}\big). \tag{5.11}$$

This approximate distribution is used in a Gaussian error propagation for linear functions of the transition matrix. Let us assume that we are interested in computing the statistical error of a scalar functions $f(\mathbf{T}) : \mathbb{R}^{n\times n} \to \mathbb{R}$. The first order Taylor approximation is given by:

$$f(\mathbf{T}) = f(\hat{\mathbf{T}}) + \sum_{i,j} \frac{\partial f}{\partial T_{ij}}\bigg|_{\hat{\mathbf{T}}} (T_{ij} - \hat{T}_{ij}).$$

Since the uncertainty in the rows of $\mathbf{T}$ contribute independently to the uncertainty in f, we define a sensitivity vector $\mathbf{s}^{(i)}$ for each row separately

$$s_j^{(i)} = \frac{\partial f}{\partial T_{ij}}(\hat{\mathbf{T}})$$

that measures the sensitivity of the scalar function with respect to changes in the transition matrix elements. Then, with the function for the error propagation, we get

$$\hat{f} = f(\hat{\mathbf{T}})$$

obtaining an approximation for the variance in f,

$$\mathrm{Var}(f) = \mathrm{Cov}(f, f) = \sum_i \left(\mathbf{s}^{(i)}\right)^T \boldsymbol{\Sigma}^{(i)} \mathbf{s}^{(i)}.$$

or, more general, for the covariances between different scalar functions f, and g

$$\mathrm{Cov}(f, g) = \sum_i \left(\mathbf{s}[f]^{(i)}\right)^T \boldsymbol{\Sigma}^{(i)} \mathbf{s}[g]^{(i)}.$$

where $\mathbf{s}[f]^{(i)}$ and $\mathbf{s}[g]^{(i)}$ refer to the sensitivities of f and g respectively. The limitation of this approach is that it does not work well in situations where the Transition matrix distribution is far from Gaussian (especially in the situation of little data). Furthermore, the more nonlinear a given function of interest is in terms of T_{ij}, the more the estimated uncertainty on this function might be wrong.

5.3.1 Example: Eigenvalues

As an example, we consider the computation of statistical error in a particular eigenvalue λ_k of the transition matrix $\mathbf{T}$ using the linear error propagation scheme, closely following the approach described in Refs. [4, 13].

We start from the eigenvalue decomposition of the transition matrix $\mathbf{T}$, omitting the dependence on the lag time τ,

$$\boldsymbol{\Lambda} = \boldsymbol{\Phi}\mathbf{T}\boldsymbol{\Psi} \tag{5.12}$$

where $\boldsymbol{\Psi} = [\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_n]$ is the right eigenvector matrix, $\boldsymbol{\Phi} = [\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_n]^T = \boldsymbol{\Psi}^{-1}$ is the left eigenvector matrix, and $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_i)$ is the diagonal matrix of eigenvalues. For the kth eigenvalue-eigenvector pair, we have,

$$\lambda^{(k)} = \left(\boldsymbol{\phi}^{(k)}\right)^T \mathbf{T} \boldsymbol{\psi}^{(k)} = \sum_{i,j} \phi_i^{(k)} T_{ij} \psi_j^{(k)}.$$

We wish to compute the statistical error of the eigenvalues $\lambda^{(k)}$ via linear error perturbation. In general, both the eigenvalues and eigenvectors simultaneously depend on perturbations in the elements of $\mathbf{T}$ in a complex way. To first order, the partial derivatives of the eigenvalues with respect to the transition matrix elements is given by the inner product of left and right eigenvectors,

$$\frac{\partial \lambda^{(k)}}{\partial T_{ij}} = \phi_i^{(k)} \psi_j^{(k)}. \tag{5.13}$$

This expression for the eigenvalue sensitivity may be combined with Eq. (5.11) in order to yield the linear perturbation result,

$$\begin{aligned}
\mathrm{Var}\left(\lambda^{(k)}\right) &= \sum_{i=1}^{n} \sum_{a,b} \frac{\partial \lambda^{(k)}}{\partial T_{ia}} \mathrm{Cov}(T_{ab}) \frac{\partial \lambda^{(k)}}{\partial T_{ib}} \\
&= \sum_{i=1}^{n} \sum_{a,b} \phi_i^{(k)} \psi_a^{(k)} \left(\sum_a \frac{u_{ia}(u_i - u_{ia})}{u_i^2 (u_i + 1)} \right. \\
&\quad \left. + \sum_{a,b \neq a} \frac{-u_{ia} u_{ib}}{u_i^2 (u_i + 1)} \right) \phi_i^{(k)} \psi_b^{(k)}.
\end{aligned}$$

5.4 Sampling Transition Matrices Without Detailed Balance Constraint

In a full Bayesian approach, we sample the posterior distribution,

$$p(\mathbf{T} \mid \mathbf{C}) \propto p(\mathbf{T}) p(\mathbf{C} \mid \mathbf{T}) = \prod_{i,j} T_{ij}^{c_{ij}} \tag{5.14}$$

where we recall that the total count matrix $\mathbf{C} = \mathbf{C}^{\mathrm{obs}} + \mathbf{C}^{\mathrm{prior}}$, as discussed in Chap. 4, makes the use of different priors straightforward. If the only constraint of $\mathbf{T}$ is that it is a stochastic matrix, but we do not expect that $\mathbf{T}$ fulfills detailed balance, we can view Eq. (5.14) as a product of Dirichlet distributions, one for each row (see Eq. (5.7)).

We are then faced with the problem of sampling random variables from the distribution,

$$\mathbf{T}_{i*} \sim \mathrm{Dir}(\mathbf{u}_{i*}). \tag{5.15}$$

A fast way to generate Dirichlet-distributed random variables is to draw n independent samples $y_1, \ldots, y_n$ from univariate Gamma distributions, each with density,

$$y_j \sim \mathrm{Gamma}(c_{ij}+1, 1) = \frac{y_j^{c_{ij}} e^{-y_j}}{\Gamma(c_{ij}+1)}, \quad j = 1, \ldots, n, \tag{5.16}$$

and then obtain the T_{ij} by normalization of each row,

$$T_{ij} = \frac{y_j}{\sum_{m=1}^{n} y_k}. \tag{5.17}$$

Repeating this procedure independently for every row $i = 1, \ldots, n$ will generate a statistically independent sample of $\mathbf{T}$ from distribution (5.14).

5.5 Sampling the Reversible Transition Matrix Distribution

No similarly simple approach to direct generation of statistically independent samples of the distribution (5.14) exists when the transition matrix $\mathbf{T}$ is further constrained to satisfy that the transition matrices fulfill detailed balance. To include the detailed balance constraints, we consider sampling Eq. (5.14) using the Metropolis-Hastings algorithm, where we propose a change to the transition matrix, $\mathbf{T} \to \mathbf{T}'$. This proposal is accepted with probability given by the Metropolis-Hastings criterion,

$$\begin{aligned} p_{\mathrm{acc}} &= \frac{p(\mathbf{T}' \to \mathbf{T})}{p(\mathbf{T} \to \mathbf{T}')} \frac{p(\mathbf{T}'|\mathbf{C})}{p(\mathbf{T}|\mathbf{C})} \\ &= \frac{p(\mathbf{T}' \to \mathbf{T})}{p(\mathbf{T} \to \mathbf{T}')} \frac{p(\mathbf{C}|\mathbf{T}')}{p(\mathbf{C}|\mathbf{T})} \\ &= \frac{p(\mathbf{T}' \to \mathbf{T})}{p(\mathbf{T} \to \mathbf{T}')} \frac{\prod_{i,j} T'^{c_{ij}}_{ij}}{\prod_{i,j} T_{ij}^{c_{ij}}}. \end{aligned} \tag{5.18}$$

This scheme requires efficient schemes to generate proposals $\mathbf{T} \to \mathbf{T}'$ that maintain the detailed balance constraint and are likely to be accepted, as well as a method of efficiently computing the ratio of transition probabilities $p(\mathbf{T}' \to \mathbf{T})/p(\mathbf{T} \to \mathbf{T}')$ for each proposal. Such a scheme was worked out in detail in Ref. [7], and we summarize the resulting method as Algorithm 2.

Example 1 Every 2×2 transition matrix is reversible. To see this, we can compute the stationary distribution from the dominant eigenvector,

$$\boldsymbol{\pi} = \left(\frac{T_{21}}{T_{12}+T_{21}}, \frac{T_{12}}{T_{12}+T_{21}} \right), \tag{5.19}$$

from which we can see that detailed balance is always fulfilled,

$$\pi_1 T_{12} = \frac{T_{21}}{T_{12}+T_{21}} T_{12} = \frac{T_{12}}{T_{12}+T_{21}} T_{21} = \pi_2 T_{21}. \tag{5.20}$$

Indeed, for 2×2 matrices the nonreversible transition matrix sampling scheme (Sect. 5.4) generates the same distribution as the reversible transition matrix sampling scheme in Algorithm 2. See Fig. 5.1B for an illustration of this sampling scheme applied to a 2×2 matrix.

Example 2 Figure 5.2 illustrates how the distribution of a 3×3 transition matrix differs between the nonreversible (panels B, E, H) and reversible (panels C, F, I) cases. For the matrix studied here, the distribution of reversible matrices is slightly narrower.

5.5.1 Sampling with Fixed Stationary Distribution

In some cases, the stationary distribution, $\boldsymbol{\pi}$, may be known exactly or to very small statistical error. For example, an efficient equilibrium simulation scheme (such as parallel tempering or metadynamics) or a Monte Carlo method may have generated a very precise estimate of $\boldsymbol{\pi}$ by simulating a perturbed system or one with unphysical dynamics. It may be useful to incorporate this information about $\boldsymbol{\pi}$ when inferring the posterior distribution of transition matrices, since it may significantly reduce the uncertainty.

Algorithm 2 Metropolis Monte Carlo sampling of reversible stochastic matrices

Input: Transition count matrix $\mathbf{C} \in \mathbb{N}_0^{n \times n}$. Number of samples N.
Output: Ensemble of reversible transition matrices, $\mathbf{T}_1, \ldots, \mathbf{T}_N$.
1. Initialize $T_{ij}^{(0)} = (c_{ij} + c_{ji})/(\sum_{m=1}^{m} c_{ik} + c_{ki}) \forall i, j \in (1, \ldots, m)$.
2. Compute $\boldsymbol{\pi}$ as stationary distribution of $\mathbf{T}^{(0)}$ by solving $\boldsymbol{\pi}^{(0)} = \boldsymbol{\pi}^{(0)}\mathbf{T}^{(0)}$.
3. For $k = 1 \ldots N$
   3.1. Generate uniform random variables: $r_1, r_2 \sim$ Uniform[0, 1]
   3.2. $\mathbf{T}^{(k)} := \mathbf{T}^{(k-1)}$
   3.3. If $(r_1 < 0.5)$ *Reversible Element Shift:*
      3.3.1. Generate uniform random variables:
      $$i, j \in \{1, \ldots, n\},\ \Delta \in \left[\max\left\{ -T_{ii}^{(k)}, -\frac{\pi_j^{(k)}}{\pi_i^{(k)}} T_{jj}^{(k)} \right\}, T_{ij}^{(k)} \right].$$
      3.3.2. $p_{acc} := \left(\frac{(T_{ij}^{(k)} - \Delta)^2 + T_{ji}^{(k)} - \frac{\pi_i^{(k)}}{\pi_j^{(k)}}\Delta)^2}{(T_{ij}^{(k)})^2 + (T_{ji}^{(k)})^2} \right) \left(\frac{T_{ii}^{(k)} + \Delta}{T_{ii}^{(k)}} \right)^{c_{ii}} \left(\frac{T_{ij}^{(k)} - \Delta}{T_{ij}^{(k)}} \right)^{c_{ij}} \left(\frac{T_{jj}^{(k)} + \frac{\pi_i^{(k)}}{\pi_j^{(k)}}\Delta}{T_{jj}^{(k)}} \right)^{c_{jj}} \times \left(\frac{T_{ji}^{(k)} - \frac{\pi_i^{(k)}}{\pi_j^{(k)}}\Delta}{T_{ji}^{(k)}} \right)^{c_{ji}}$
      3.3.3. If $(r_2 \leq p_{acc})$:
      Set $T_{ii}^{(k)} := T_{ii}^{(k-1)} + \Delta$; $T_{ij}^{(k)} := T_{ij}^{(k-1)} - \Delta$
      and $T_{jj}^{(k)} := T_{jj}^{(k-1)} + \Delta\pi_i^{(k)}/\pi_j^{(k)}$; $T_{ji}^{(k)} := T_{ji}^{(k-1)} - \Delta\pi_i^{(k)}/\pi_j^{(k)}$
   else *Node Shift*:
      3.3.4. Generate uniform random variables: $i \in (1, \ldots, n)$, $\alpha \in \left[0, \frac{1}{1 - T_{ii}^{(k)}}\right]$.
      3.3.5. $p_{acc} := \alpha^{(n-2+c_i-c_{ii})} \left(\frac{1 - \alpha(1 - T_{ii}^{(k)})}{T_{ii}^{(k)}} \right)^{c_{ii}}$
      3.3.6. If $r_2 \leq p_{acc}$:
      For all $j \neq i$, set $T_{ij}^{(k)} := \alpha T_{ij}^{(k-1)}$.
      Set $T_{ii}^{(k)} = 1 - \sum_{j \neq i} T_{ij}^{(k)}$
      3.3.7. Update stationary distribution:
      For all $j \neq i$:, set $\pi_j^{(k)} := \frac{\alpha\pi_j^{(k-1)}}{\pi_i^{(k-1)} + \alpha(1 - \pi_i^{(k-1)})}$.
      Set $\pi_i^{(k)} := 1 - \sum_{j \neq i} \pi_j^{(k)}$.

To do this, we first note that the two types of Monte Carlo proposals utilized in Algorithm 2 above for sampling reversible transition matrices. One type of proposal (reversible element shifts) changes $\boldsymbol{\pi}$, while the other preserves $\boldsymbol{\pi}$ (node shift). We can suggest a straightforward modification of the $\mathbf{T}$-sampling algorithm that will ensure $\boldsymbol{\pi}$ is constrained to some specified value during the sampling procedure.

We first give an algorithm to construct an initial transition matrix $\mathbf{T}^{(0)}$ with a specified stationary distribution $\boldsymbol{\pi}$ from a given count matrix $\mathbf{C}$ (Algorithm 3), and then use this to initialize a Monte Carlo transition matrix sampling algorithm that preserves the stationary distribution (Algorithm 4).

5.6 Full Bayesian Approach with Uncertainty in the Observables

Suppose we are interested in some experimentally-measurable function of state $A(\mathbf{x})$. An experiment may be able to measure an expecta-

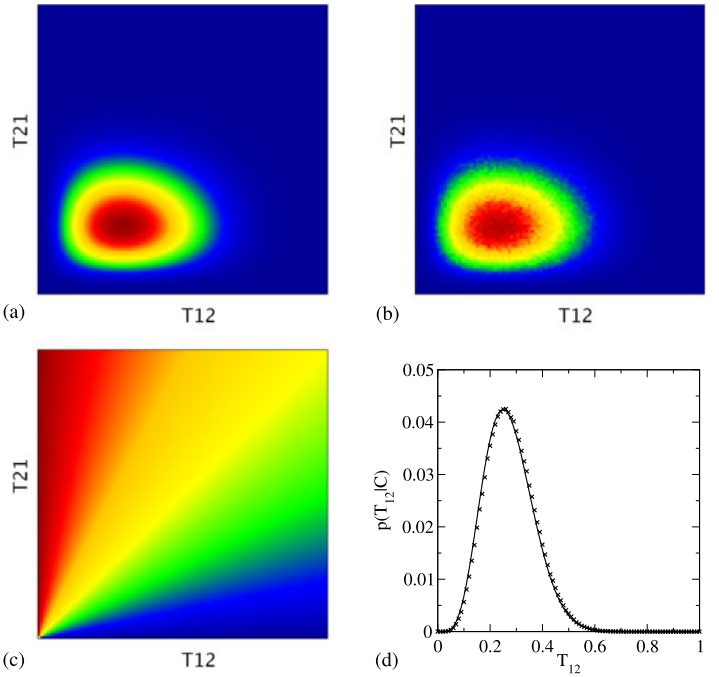

Fig. 5.1 Illustration of sampling of transition probability matrices for the observation $C = \left(\begin{smallmatrix} 5 & 2 \\ 3 & 10 \end{smallmatrix}\right)$ and a uniform prior. Panels (**a**), (**b**), and (**c**) show the probability distribution on the off-diagonal matrix elements. The color encodes the probability density, with *blue* $= 0$ and *red* $= 1$. Each density was scaled such that its maximum is equal to 1. (**a**) Analytic density of stochastic matrices. (**b**) Sampled density of stochastic matrices (these matrices automatically fulfill detailed balance). (**c**) Stationary probability of the first state π_1. When sampling with respect to a fixed stationary probability distribution π^*, the ensemble is fixed to the line $T_{21} = T_{12}\pi_1^*/(1-\pi_1^*)$. (**d**) Sampled and exact density of T_{12} of reversible matrices with fixed stationary distribution $\pi^* = (0.5, 0.5)$

tion $\langle A \rangle$ or correlation functions $\langle A(0)A(t) \rangle$, and we would like to compute the corresponding properties from the Markov model constructed from a molecular simulation and decide whether they agree with experiment to within statistical uncertainty, or if a prediction from the model is sufficiently precise to be useful. The previous framework for sampling transition matrices can be used in the following manner: (i) Assign the state-averaged value of the observable, $a_i = \int_{S_i} d\mathbf{x}\mu(\mathbf{x})A(\mathbf{x})$, to each discrete state. (ii) Generate an ensemble of $\mathbf{T}$-matrices according to the sampling scheme described above. (iii) Calculate the desired expectation or correlation function for each $\mathbf{T}$-matrix using the discrete vector $\mathbf{a} = [a_i]$. This approach involves several approximations that each deserve discussion. Here, we want to generalize the approach by eliminating one important approximation—that the values a_i are known exactly without statistical error themselves.

In a typical simulation scenario, the average a_i is itself calculated by a statistical sample. When a simulation trajectory $\mathbf{x}_t$ is available, then typically the time average

$$\hat{a}_i = \frac{\sum_t \chi_i(\mathbf{x}_t)A(\mathbf{x}_t)}{\sum_t \chi_i(\mathbf{x}_t)} \tag{5.21}$$

is employed, where χ_i is the indicator function of state i. The estimate $\hat{a}_i$ may in fact have significant statistical error because the number of uncorrelated samples of $\mathbf{x}_t$ inside any state i is finite, and possibly rather small. In order to estimate the distribution of expectation or correlation functions of A due to both, the statistical uncertainty of $\mathbf{T}$ and the statistical uncertainty of $\hat{a}_i$, we propose a full Bayesian approach using a Gibbs

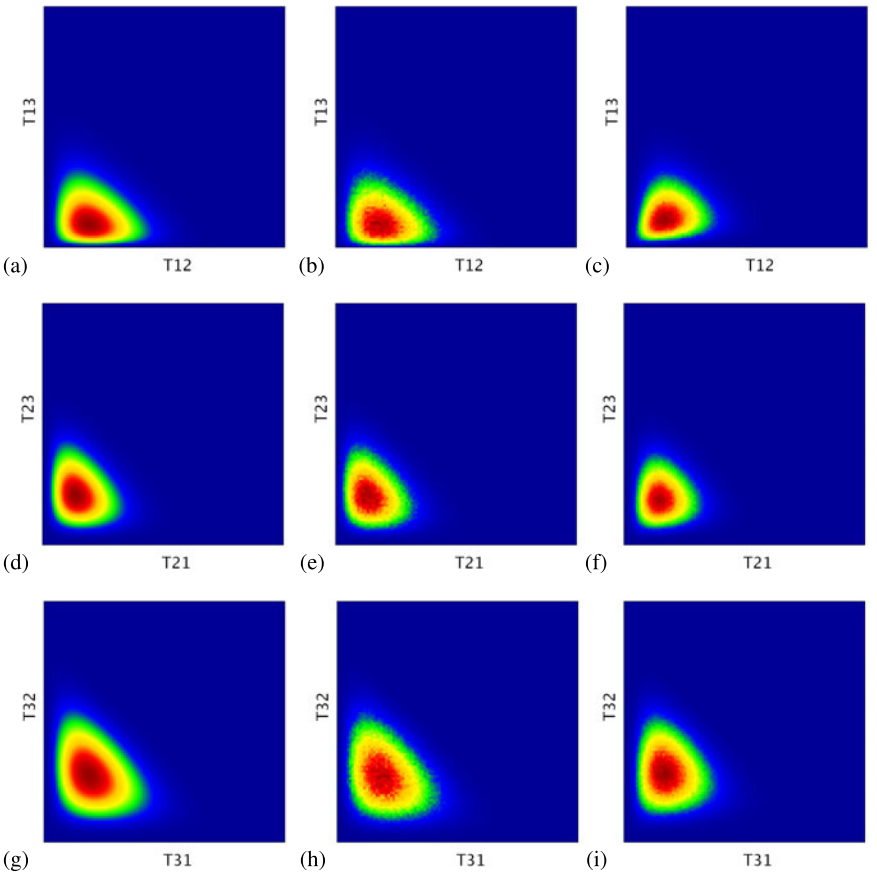

Fig. 5.2 Visualization of the probability density of transition matrices for the count matrix $\mathbf{C}^{\text{obs}} = \begin{pmatrix} 8 & 2 & 1 \\ 2 & 10 & 3 \\ 2 & 3 & 6 \end{pmatrix}$ and a uniform prior. Different two-dimensional joint marginal distributions are shown in the rows. The analytic and sampled distributions for stochastic matrices are shown in *columns 1* and *2*, respectively. *Column 3* shows the sampled distribution for stochastic matrices fulfilling detailed balance. Note how the peaks are more sharply peaked when the detailed balance constraint is imposed (*column 3*) compared to the corresponding transition matrices without detailed balance constraint (*column 2*)

sampling scheme, here illustrated for the expectation $\mathbb{E}[A]$ (Algorithm 5).

While the transition matrix $\mathbf{T}^{(k)}$ can be sampled using the framework described in the previous sections, an approach to sample $\mathbf{a}^{(k)}$ introduced in Ref. [2] is described subsequently.

5.6.1 Sampling State Expectations $\mathbf{a}^{(k)}$

Consider the expectation of some molecular observable $A(\mathbf{x})$ computed from Eq. (5.21). Temporally sequential samples $A_t \equiv A(\mathbf{x}_t)$ collected with a temporal resolution of the Markov time τ are subsequently presumed to be uncorrelated. We also assume that the set of samples $A(\mathbf{x}_t)$ for those configurations $\mathbf{x}_t$ appearing in state i are collected in the set $\{A_m\}_{m=1}^{N}$ in the remainder of this section, generally abbreviated as $\{A_m\}$.

Because only a finite number of samples N are collected for each state, there will be a degree of uncertainty in this estimate. Unlike the problem of inferring the transition matrix elements, however, we cannot write an exact expression for the probability of observing a single sample A_m in terms of a simple parametric form, since its probability distribution may be arbitrarily complex,

$$p_i(A_m) = \frac{1}{\pi_i} \int_{S_i} d\mathbf{x}\, \delta\big(A_m - A(\mathbf{x})\big)\mu(\mathbf{x}). \quad (5.22)$$

Algorithm 3 Generation of an initial transition matrix $\mathbf{T}^{(0)}$ given count matrix $\mathbf{C}$ and a specified stationary distribution $\boldsymbol{\pi}$

Input: Stationary distribution $\boldsymbol{\pi}$ and transition count matrix $\mathbf{C}$.
Output: Transition matrix $\mathbf{T}$ that has stationary distribution $\boldsymbol{\pi}$.
1. Define $\mathbf{Y} \in \mathbb{R}^{n \times n}$ as:

$$y_{ij} = \begin{cases} \frac{\pi_i c_{ij}}{2\sum_k c_{ik}} + \frac{\pi_j c_{ji}}{2\sum_k c_{jk}} & i \neq j, \\ 0 & i = j. \end{cases}$$

2. Define $\mathbf{X} \in \mathbb{R}^{n \times n}$ as:

$$o = \max_i \left\{ \sum_k x_{ik} \right\},$$

$$x_{ij} = \begin{cases} \frac{y_{ij}}{o} & i \neq j, \\ \pi_i - \sum_k \frac{y_{ik}}{o} & i = j. \end{cases}$$

3. Define $\mathbf{T}^{(0)} \in \mathbb{R}^{n \times n}$ as

$$T_{ij}^{(0)} = \frac{x_{ij}}{\sum_k x_{ik}}.$$

Algorithm 4 Metropolis-Hastings Monte Carlo sampling of reversible stochastic matrices with probability distribution of stationary distributions $p(\boldsymbol{\pi})$

Input: Transition count matrix $\mathbf{C} \in \mathbb{N}_0^{n \times n}$. Number of samples n_1, n_2. Stationary distribution $p(\boldsymbol{\pi})$
Output: Ensemble of reversible transition matrices, $\mathbf{T}_1, \ldots, \mathbf{T}_N$.
1. For $k = 1 \ldots n_1$
1.1. Draw $\boldsymbol{\pi}^{(k)}$ from $p(\boldsymbol{\pi})$
1.2. Initialize $\mathbf{T}^{(0)}$ using Algorithm 3.
1.3. For $l = 1 \ldots n_2$
1.3.1. Use *reversible element shift* from Algorithm 2 to update the transition matrix.

Algorithm 5 Gibbs sampler for the joint estimation of $p(\mathbb{E}[A])$

1. For $k = 1 \ldots N$
1.1. Sample observables
$\mathbf{a}^{(k)} \sim p(\mathbf{a} \mid \mathbf{x}_t)$.
1.2. Sample transition matrix
$\mathbf{T}^{(k)} \sim p(\mathbf{T} \mid \mathbf{x}_t) = p(\mathbf{T} \mid \mathbf{C}^{\mathrm{obs}})$.
1.3. Compute $\pi^{(k)}$ as the stationary distribution of $\mathbf{T}^{(k)}$ such that
$[\boldsymbol{\pi}^{(k)}]^{\mathbb{T}} = [\mathbf{T}^{(k)}][\boldsymbol{\pi}^{(k)}]^{\mathbb{T}}$.
1.3. Generate a sample of the expectation value:

$$A^{(k)} = \sum_{i=1}^{n} a_i^{(k)} \pi_i^{(k)}.$$

Despite this, the central limit theorem states that the behavior of $\hat{a}_i$ approaches a normal distribution (generally very rapidly) as the number of samples N increases. We will therefore make the assumption that $p_i(A_m)$ is *normal*—that is, we assume the distribution can be characterized by mean μ_i and variance σ_i^2,

$$A_m \sim \text{Normal}\left(\mu_i, \sigma_i^2\right) \tag{5.23}$$

where the normal distribution implies the probability density for A_m is approximated by

$$\tilde{p}_i\left(A_m; \mu_i, \sigma_i^2\right) = (2\pi)^{-1/2}\sigma_i^{-1}\exp\left[-\frac{1}{2\sigma_i^2}(A_m - \mu_i)^2\right]. \tag{5.24}$$

While this may seem like a drastic assumption, it turns out this approximation allows us to do a surprisingly good job of inferring the distribution of the error in $\delta\hat{a}_i \equiv \hat{a}_i - \langle A\rangle_i$ even for a small number of samples from each state, and generally gives an overestimate of the error (which is arguably less dangerous than an underestimate) for smaller sample sizes. While the validity of this approximation is illustrated in a subsequent example, we continue below to develop the ramifications of this approximation.

Consider the sample mean estimator for $\langle A\rangle_i$,

$$\hat{\mu} = \frac{1}{N}\sum_{m=1}^{N} A_m. \tag{5.25}$$

The asymptotic variance of $\hat{\mu}$, which provides a good estimate of the statistical uncertainty in $\hat{\mu}$ in the large-sample limit, is given as a simple consequence of the central limit theorem,

$$\delta^2\hat{\mu} \equiv \mathbb{E}\left[\left(\hat{\mu} - \mathbb{E}[\hat{\mu}]\right)^2\right] = \frac{\text{Var}\, A_m}{N} \approx \frac{\hat{\sigma}^2}{N} \tag{5.26}$$

where the unbiased estimator for the variance $\sigma^2 \equiv \text{Var}\, A_m$ is given by

$$\hat{\sigma}^2 \equiv \frac{1}{N-1}\sum_{m=1}^{N}(A_m - \hat{\mu})^2 \tag{5.27}$$

Suppose we now *assume* the distribution of A from state i is normal (Eq. (5.24)),

$$A|\mu, \sigma^2 \sim \text{Normal}\left(\mu, \sigma^2\right). \tag{5.28}$$

Were this to be a reasonable model, we could model the timeseries of the observable $A_t \equiv A(x_t)$ by the hierarchical process:

$$\begin{aligned} s_t|s_{t-1}, \mathbf{T} &\sim \text{Bernoulli}(T_{s_{t-1}1}, \ldots, T_{s_{t-1}N}), \\ A_t|\mu_{s_t}, \sigma_{s_t}^2 &\sim \text{Normal}\left(\mu_{s_t}, \sigma_{s_t}^2\right). \end{aligned} \tag{5.29}$$

Here, the notation $\text{Bernoulli}(\pi_1, \ldots, \pi_N)$ denotes a Bernoulli scheme where discrete outcome n has associated probability π_n of being selected. We will demonstrate below how this model does in fact recapitulate the expected behavior in the limit where there are sufficient samples from each state.

We choose the (improper) Jeffreys prior [5],

$$p\left(\mu, \sigma^2\right) \propto \sigma^{-2} \tag{5.30}$$

because it satisfies intuitively reasonable reparameterization [5] and information-theoretic [3] invariance principles. Note that this prior is uniform in $(\mu, \log\sigma)$.

The posterior is then given by

$$\begin{aligned} p\left(\mu, \sigma^2 \,\middle|\, \{A_m\}\right) &\propto \left[\prod_{n=1}^{N} p\left(A_m \,\middle|\, \mu, \sigma^2\right)\right] p\left(\mu, \sigma^2\right) \\ &\propto \sigma^{-(N+2)}\exp\left[-\frac{1}{2\sigma^2}\sum_{m=1}^{N}(A_m - \mu)^2\right]. \end{aligned} \tag{5.31}$$

Rewriting in terms of the sample statistics $\hat{\mu}$ and $\hat{\sigma}^2$, we obtain

$$p\left(\mu, \sigma^2 \,\middle|\, \{A_m\}\right) \propto \sigma^{-(N+2)}\exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{m=1}^{N}(A_m - \hat{\mu})^2 + N(\hat{\mu} - \mu)^2\right]\right\}$$

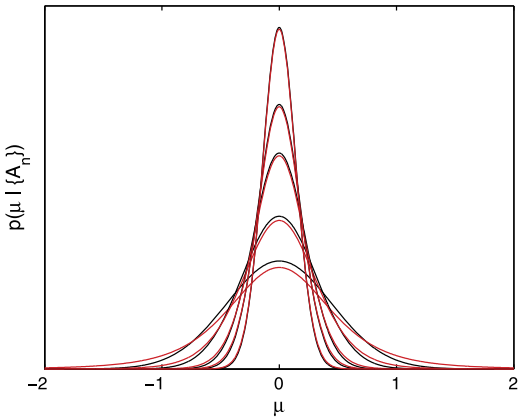

Fig. 5.3 Approach to normality for marginal distribution of the mean $p(\mu|\{A_m\})$. For fixed $\hat{\mu}$ and $\hat{\sigma}^2$, the marginal posterior distribution of μ (*red*), a scaled and shifted Student t-distribution, rapidly approaches the normal distribution (*black*) expected from asymptotic statistics. The PDF is shown for sample sizes of $N = 5$ (the broadest), 10, 20, and 30

$$\propto \sigma^{-(N+2)} \exp\left\{-\frac{1}{2\sigma^2}\left[(N-1)\hat{\sigma}^2 + N(\hat{\mu}-\mu)^2\right]\right\}. \tag{5.32}$$

The posterior has marginal distributions

$$\begin{aligned} \sigma^2|\{A_m\} &\sim \text{Inv}{-}\chi^2\left(N-1, \hat{\sigma}^2\right), \\ \mu|\{A_m\} &\sim \text{t}_{N-1}\left(\hat{\mu}, \hat{\sigma}^2/N\right) \end{aligned} \tag{5.33}$$

where σ^2 is distributed according to scaled inverse chi-square distribution with $N-1$ degrees of freedom, and μ according to Student's t-distribution with $N-1$ degrees of freedom that has been shifted to be centered about $\hat{\mu}$ and whose width has been scaled by $\hat{\sigma}^2/N$.

As can be seen in Fig. 5.3, as the number of degrees of freedom increases, the marginal posterior for μ approaches the normal distribution with the asymptotic behavior expected from standard frequentest analysis for the standard error of the mean, namely

$$\mu \to \text{N}\left(\hat{\mu}, \hat{\sigma}^2/N\right). \tag{5.34}$$

At low sample counts, the t-distribution is lower and wider than the normal distribution, meaning that confidence intervals computed from this distribution will be somewhat larger than those of the corresponding normal estimate for small samples. In some sense, this partly compensates for $\hat{\sigma}^2$ being a poor estimate of the true variance for small sample sizes, which would naturally lead to underestimates of the statistical uncertainty. In any case, this is also far from the asymptotic limit where the normal distribution with variance $\hat{\sigma}^2/N$ is expected to model the uncertainty well.

The posterior can also be decomposed as

$$\begin{aligned} &p\left(\mu, \sigma^2 \,\middle|\, \{A_m\}\right) \\ &\quad = p\left(\mu \,\middle|\, \sigma^2, \{A_m\}\right) p\left(\sigma^2 \,\middle|\, \{A_m\}\right). \end{aligned} \tag{5.35}$$

This readily suggests a two-step sampling scheme for generating uncorrelated samples of (μ, σ^2), in which we first sample σ^2 from its marginal distribution, and then μ from its distribution conditional on σ^2

$$\begin{aligned} \sigma^2|\{A_m\} &\sim \text{Inv}{-}\chi^2\left(N-1, \hat{\sigma}^2\right), \\ \mu|\sigma^2, \{A_m\} &\sim \text{N}\left(\hat{\mu}, \sigma^2/N\right). \end{aligned} \tag{5.36}$$

Alternatively, if the scaled inverse-chi-square distribution is not available, the χ^2-distribution (among others) can be used to sample σ^2:

$$(N-1)\left(\hat{\sigma}^2/\sigma^2\right) \,\middle|\, \{A_m\} \sim \chi^2(N-1) \tag{5.37}$$

where the first argument is the shape parameter and the second argument is the scale parameter.

5.6.2 Illustration of Fully Bayesian Sampling Scheme

Using the sampling procedures described previously, we are now equipped with a scheme to sample from the joint posterior describing our confidence in that a Markov model characterized by a transition matrix $\mathbf{T}$ and state expectations $\boldsymbol{\mu}_i$, $i = 1, \ldots, M$, produced the observed trajectory data. Using a set of models sampled from this posterior, we can characterize the statistical component of the uncertainty as it propagates into equilibrium averages, non-equilibrium relaxations, and (non-)equilibrium correlation measurements computed from the Markov model. To ensure the correctness of this procedure, however, we first test its ability to correctly characterize the

posterior distribution for a finite-size sample from a true Markovian model system.

How can we test a Bayesian posterior distribution? One of the more powerful features of a Bayesian model is its ability to provide confidence intervals that correctly reflect the level of certainty that the true value will lie within it. For example, if the experiment were to be repeated many times, the true value of the parameter being estimated should fall within the confidence interval for a 95 % confidence level 95 % of the time. As an illustrative example, consider a biased coin where the probability of turning heads is θ. From an observed sample of N coin flips, we can estimate θ using a Binomial model for the number of coin flips that turn up heads and a conjugate Beta Jeffreys prior [3, 5]. Each time we run an experiment and generate a new independent collection of N samples, we get a different posterior estimate for θ, and a different confidence interval (Fig. 5.4, top). If we run many trials and record what fraction of the time the true (unknown) value of θ falls within the confidence interval estimated from that trial, we can see if our model is correct. If correct, the observed confidence level should match the desired confidence level (Fig. 5.4, bottom right). Deviation from parity means that the posterior is either two broad or too narrow, and that the statistical uncertainty is being either over- or underestimated (Fig. 5.4, bottom left).

We performed a similar test on a three-state model system, using a model (reversible, row-stochastic) transition matrix for one Markov time is given by

$$\mathbf{T}(1) = \begin{bmatrix} 0.86207 & 0.12931 & 0.00862 \\ 0.15625 & 0.83333 & 0.01041 \\ 0.00199 & 0.00199 & 0.99602 \end{bmatrix}. \tag{5.38}$$

Each state is characterized by a mean value of the observable $A(x)$, fixed to 3, 2, and 1 for the first, second, and third states, respectively. The equilibrium populations are $\boldsymbol{\pi} \approx [0.16250.13450.7031]$. Simulation from this model involves a stochastic transition according to the transition element T_{ij} followed by observation of the value of $A(x)$ sampled i.i.d. from the current state's probability distribution $p_i(A)$. Multiple independent realizations of this process were carried out, and subjected to the Bayesian inference procedure for transition matrices and observables described above. The nonequilibrium relaxation $\langle A \rangle_{\rho_0}$ from the initial condition $\rho_0 = [100]$ in which all density is concentrated in state 1, as well as the autocorrelation function $\langle A(0)A(t) \rangle$, is shown in Fig. 5.5.

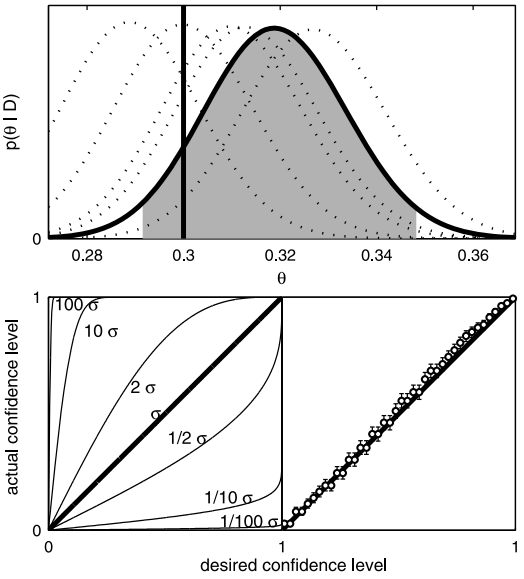

Fig. 5.4 Testing the posterior for inference of a biased coin flip experiment. *Top:* Posterior distribution for inferring the probability of heads, θ, for a biased coin from a sequence of $N = 1000$ coin flips (*dark line*) with 95 % symmetric confidence interval about the mean (*shaded area*). The true probability of heads is 0.3 (*vertical thick line*). Posteriors from five different experiments are shown as *dotted lines*. *Bottom left:* Desired and actual confidence levels for an idealized normal posterior distribution that either overestimates (*upper left curves*) or underestimates (*bottom right curves*) the true posterior variance by different degrees. *Bottom right:* Desired and actual confidence levels for the Binomial-Beta posterior for the coin flip problem depicted in *upper panel*. *Error bars* show 95 % confidence intervals estimates from 1000 independent experimental trials. For inference, we use a likelihood function such that the observed number of heads is $N_H|\theta \sim \text{Binomial}(N_H, N, \theta)$ and conjugate Jeffreys prior [3, 5] $\theta \sim \text{Beta}(1/2, 1/2)$ which produces posterior $\theta|N_H \sim \text{Beta}(N_H + 1/2, N_T + 1/2)$ along with constraint $N_H + N_T = N$

With the means of $p_i(A)$ within each state fixed as above, we considered models for $p_i(A)$ that were either *normal* or *exponential*, using the

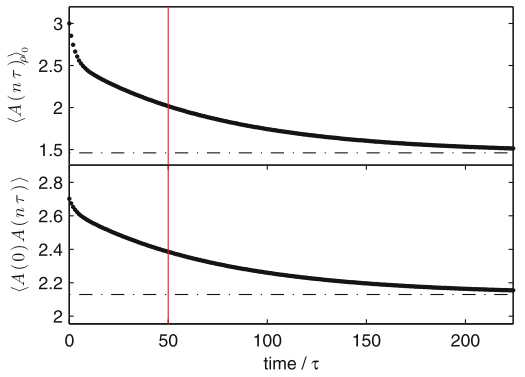

Fig. 5.5 Observables for three-state model system. *Top:* Relaxation of $\langle A(t)\rangle_{\rho_0}$ (*solid line*) from initial distribution $\rho_0 = [100]$ to equilibrium expectation $\langle A\rangle$ (*dash-dotted line*). *Bottom:* Equilibrium autocorrelation function $\langle A(0)A(t)\rangle$ (*solid line*) to $\langle A\rangle^2$ (*dash-dotted line*). The estimates of both $\langle A(t)\rangle_{\rho_0}$ and $\langle A(0)A(t)\rangle$ at 50 timesteps (*red vertical line*) were assessed in the validation tests described here

probability density functions:

$$p_i(A) = (2\pi)^{-1/2}\sigma_i^{-1}\exp\left[-\frac{1}{2\sigma_i^2}(A-\mu_i)^2\right],$$

normal

$$p_i(A) = \mu_i^{-1}\exp[-A/\mu_i],$$

$A \geq 0$. exponential

While the normal output distribution for $p_i(A)$ corresponds to the hierarchical Bayesian model that forms the basis for our approach, the exponential distribution is significantly different, and represents a challenging test case.

Figure 5.6 depicts the resulting uncertainty estimates for both normal (top) and exponential (bottom) densities for the observable A. In both cases, the confidence intervals are *underestimated* for short trajectory lengths (1 000 steps) where, in many realizations, few samples are ob-

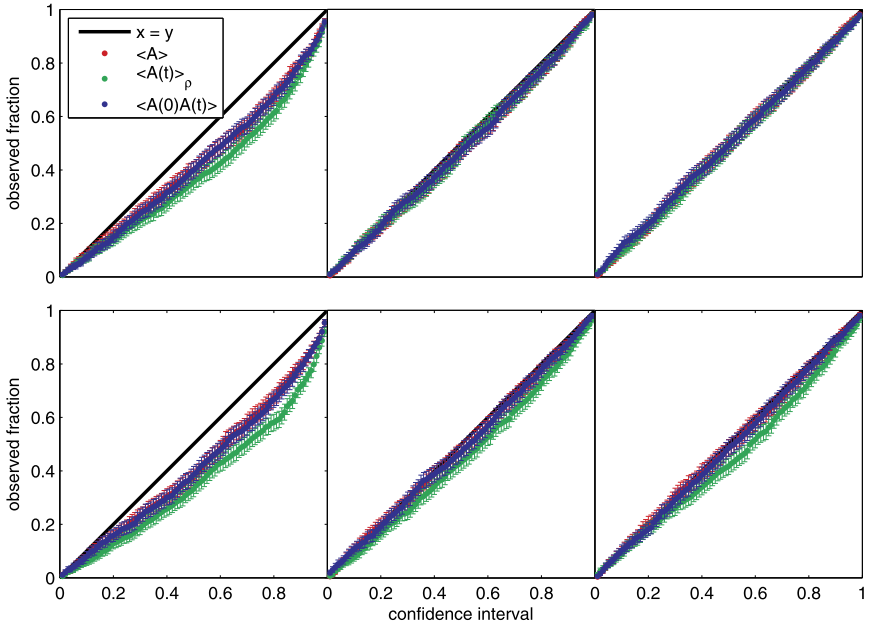

Fig. 5.6 Confidence interval tests for model system. *Top:* Expected and observed confidence intervals for three-state system with normal distribution for observable A with unit variance for simulations of length 1 000 (*left*), 10 000 (*middle*), and 100 000 (*right*) steps. Confidence intervals were estimated from generating 10 000 samples from the Bayesian posterior. Estimates of the fraction of observed times the true value was within the confidence interval estimated from the Bayesian posterior were computed from generating 1 000 independent experimental realizations. The resulting curves are shown for the equilibrium estimate $\langle A\rangle$ (*red*), nonequilibrium relaxation $\langle A\rangle_{\rho_0}$ (*green*), and the equilibrium correlation function $\langle A(0)A(t)\rangle$ (*blue*). *Bottom:* Same as top, except an exponential distribution with the same mean was used for the probability of observing a particular value of A within each state

served in one or more states, so that the variance is underestimated or the effective asymptotic limit has not yet been reached. As the simulation length is increased to 10 000 or 100 000 steps so that it is much more likely there are a sufficient number of samples in each state to reach the asymptotic limit, however, the confidence intervals predicted by the Bayesian posterior become quite good. For the exponential model for observing values of A (which might be the case in, say, fluorescence lifetimes), we observe similar behavior. Except for what appears to be a slight, consistent underestimation of $\langle A(t)\rangle_{\rho_0}$ (much less than half a standard deviation) there appears to be excellent agreement between the expected and observed confidence intervals, confirming that this method is expected to be a useful approach to modeling statistical uncertainties in equilibrium and kinetic observables.

References

1. Anderson TW, Goodman LA (1957) Statistical inference about Markov chains. Ann Math Stat 28:89–110
2. Chodera JD, Noé F (2010) Probability distributions of molecular observables computed from Markov models, II: uncertainties in observables and their time-evolution. J Chem Phys 133:105,102
3. Goyal P (2005) Prior probabilities: an information-theoretic approach. In: Knuth KH, Abbas AE, Morriss RD, Castle JP (eds) Bayesian inference and maximum entropy methods in science and engineering. American Institute of Physics, New York, pp 366–373
4. Hinrichs NS, Pande VS (2007) Calculation of the distribution of eigenvalues and eigenvectors in Markovian state models for molecular dynamics. J Chem Phys 126:244,101
5. Jeffreys H (1946) An invariant form for the prior probability in estimation problems. Proc R Soc A 186:453–461
6. Metzner P, Noé F, Schütte C (2009) Estimation of transition matrix distributions by Monte Carlo sampling. Phys Rev E 80:021,106
7. Noé F (2008) Probability distributions of molecular observables computed from Markov models. J Chem Phys 128:244,103
8. Noé F, Oswald M, Reinelt G (2007) Optimizing in graphs with expensive computation of edge weights. In: Kalcsics J, Nickel S (eds) Operations research proceedings. Springer, Berlin, pp 435–440
9. Noé F, Oswald M, Reinelt G, Fischer S, Smith JC (2006) Computing best transition pathways in high-dimensional dynamical systems: application to the alphaL–beta–alphaR transitions in octaalanine. Multiscale Model Simul 5:393–419
10. Noé F, Schütte C, Vanden-Eijnden E, Reich L, Weikl TR (2009) Constructing the full ensemble of folding pathways from short off-equilibrium simulations. Proc Natl Acad Sci USA 106:19,011–19,016
11. Prinz JH et al (2011) Markov models of molecular kinetics: generation and validation. J Chem Phys 134:174,105
12. Prinz JH, Held M, Smith JC, Noé F (2011) Efficient computation of committor probabilities and transition state ensembles. Multiscale Model Simul 9:545
13. Singhal N, Pande VS (2005) Error analysis and efficient sampling in Markovian state models for molecular dynamics. J Chem Phys 123:204,909