Supplementary Information for

Quantifying biochemical reaction rates from static population variability within incompletely observed complex networks

Timon Wittenstein, Nava Leibovich, and Andreas Hilfinger

Andreas Hilfinger.
E-mail: andreas.hilfinger@utoronto.ca

This PDF file includes:

- Supplementary text
- Figs. S1 to S4
- SI References
Supporting Information Text

1. Simulation details

All simulations were performed using the standard Doob-Gillespie (1, 2) algorithm until each reaction occurred at least $10^8$ times. The resulting high-fidelity distributions from these simulations were considered as the underlying ‘true’ distribution for each respective system. To obtain the distributions used to test our inference algorithm, we randomly sampled $N$ individual data points ($N = 100,000$ if not otherwise specified) from the ‘true’ distributions corresponding to $N$ independent snapshots of a given system. For this sampling we used the `numpy.random.choice` function in Python.

Reaction dynamics for numerical proof-of-principle examples in Fig. 2 of main text. Below we specify the parameters used to simulate the numerical proof-of-principle examples introduced in Figure 2 of the main text. These parameters do not necessarily correspond to realistic cellular networks but were chosen to achieve a distinct set of system dynamics to make the mathematical point that we can successfully infer $f(x_2)$ for diverse system dynamics across comparable range for $X_2$-variability. All variables are subject to the following reactions

$$
\begin{align*}
    x_i & \xrightarrow{f_i(x)} x_i + 1 \\
    x_i & \xrightarrow{x_i/\tau_i} x_i - 1
\end{align*}
$$

with a linear degradation rate dependent on the lifetime $\tau_i$ of a specific molecule $X_i$ and a production rate $f_i(x)$ dependent on the state vector $x$.

As mentioned in the main text the production rate of $X_3$ was kept the same for all systems: $f_3(x_2) = \lambda_3 x_2^{n_3}/(K_3^{n_3} + x_2^{n_3})$ with a lifetime of $\tau_3 = 1$ and the Hill function parameters $\lambda_3 = 80, n_3 = 2, K_3 = 40$. The other rate functions were chosen as follows:

- **Bistable system**
  
  $$
  \begin{align*}
  f_1(x_2) &= \lambda_1 \frac{x_2^{n_1}}{K_1^{n_1} + x_2^{n_1}} + c_1 ; \\
  f_2(x_1) &= \lambda_2 x_1 ; \\
  \text{Lifetimes:} & \quad \tau_1 = \tau_2 = \tau_3 = 1
  \end{align*}
  $$
  
  $\lambda_1 = 50, n_1 = 6, K_1 = 37, c_1 = 15$
  $\lambda_2 = 1$

- **Oscillating system**
  
  $$
  \begin{align*}
  f_1(x_3) &= \lambda_1 \frac{K_1^{n_1}}{K_1^{n_1} + x_3^{n_1}} ; \\
  f_2(x_1) &= \lambda_2 \frac{x_1^{n_2}}{K_2^{n_2} + x_1^{n_2}} ; \\
  \text{Lifetimes:} & \quad \tau_1 = \tau_2 = \tau_3 = 1
  \end{align*}
  $$
  
  $\lambda_1 = 50000, n_1 = 10, K_1 = 0.1$
  $\lambda_2 = 80, n_2 = 1, K_2 = 100$

- **Noise controlling system**
  
  $$
  \begin{align*}
  f_1(x_3) &= \lambda_1 x_3 ; \\
  f_2(x_1) &= \lambda_2 \frac{K_2^{n_2}}{K_2^{n_2} + x_1^{n_2}} ; \\
  \text{Lifetimes:} & \quad \tau_1 = 50, \tau_2 = \tau_3 = 1
  \end{align*}
  $$
  
  $\lambda_1 = 50$
  $\lambda_2 = 3000, n_2 = 10, K_2 = 10$
• Noise enhancing system

\[ f_1 = \lambda_1 ; \]
\[ f_2(x_1, x_3) = \lambda_2 \frac{x_3^{n_2}}{K_2^{n_2} + x_3^{n_2}} + c_2 x_1 ; \]
\[ \lambda_1 = 5 \]
\[ \lambda_2 = 25, \quad n_2 = 4, \quad K_2 = 50, \quad c_2 = 8 \]

Lifetimes:
\[ \tau_1 = \tau_2 = \tau_3 = 1 \]

**Reaction dynamics for example systems in Fig. 3-5 of main text.** To illustrate how the inference method depends on time-scales and sampling (Fig. 3 of the main text) we simulated the “noise enhancing system” as defined above while logarithmically varying the life-time \( \tau_3 = 1.5^k \) with \( k \) ranging from -7 to 10. In order to keep system averages \( \langle x_3 \rangle \) comparable while varying the time-scales we simultaneously adjusted the value of \( \lambda_3 \) such that \( \lambda_3 \tau_3 = 80 \) remained constant.

To illustrate the dependence of the inference method on reaction rate function \( f(x_2) \) (Fig. 4 of the main text) we simulated the above “noise enhancing system”, with \( f_3(x_2) = \lambda_3 \frac{x_2^{n_3}}{K_3^{n_3} + x_2^{n_3}} \) for varying \( K_3, n_3 \) and fixed \( \lambda_3 = 80, \tau_3 = 1 \).

To illustrate that the nature of the degradation rate does not significantly affect the inference (Fig. 5 of the main text) we simulated the above “noise enhancing system” but replaced the degradation reaction of \( X_3 \) with
\[ x_3 \xrightarrow{\gamma x_3(x_3-1)} x_3-2 \] [S.2]

with a fixed value of \( \gamma = 2 \).

### 2. Modeling measurement errors – Fig. 6 of the main text

In the main text we analyzed how measurement noise affects our inference method by accounting for absolute and relative error terms as well as undercounting of molecules. The following error models \((x_2, x_3) \mapsto (x_2', x_3')\) were applied to each sampled data point:

- absolute error

\[(x_2', x_3') = (\mathcal{N}(x_2, \sigma_{abs}), \mathcal{N}(x_3, \sigma_{abs}))\],

- relative error

\[(x_2', x_3') = (x_2 \cdot \mathcal{N}(1, \sigma_{rel}), x_3 \cdot \mathcal{N}(1, \sigma_{rel}))\],

- binominal undercounting

\[(x_2', x_3') = (\mathcal{B}(x_2, p), \mathcal{B}(x_3, p))\],

where \( \mathcal{N}(\mu, \sigma) \) refers to a normal distribution with a mean of \( \mu \) and standard variation of \( \sigma \). \( \mathcal{B}(n, p) \) refers to a binomial distribution where \( p \) corresponds to the probability that a specific molecule gets detected. All resulting \((x_2', x_3')\) pairs were rounded to integers and pairs with negative molecule numbers were ignored.

When applying our inference method to simulated data with binomial undercounting the resulting rate function depends on the undercounted variable \( x_2' \). The correct rate function was obtained by rescaling the inferred rate function argument with the detection probability \( p \) such that \( f(x_2) = f(x_2'/p) \).
3. Derivation of the general probability balance relation

We consider a general stochastic system with state vector \( \mathbf{x} = (x_1, x_2, x_3, \ldots, x_\ell) \), where the \( k^{th} \) reaction changes levels of component \( X_j \) by \( d_{kj} \), with reaction propensities \( r_k(\mathbf{x}) \) that are arbitrarily non-linear functions of the state vector. The chemical master equation is then given by

\[
\frac{dP(\mathbf{X} = \mathbf{x}, t)}{dt} = \sum_k [r_k(\mathbf{x} - \mathbf{d}_k)P(\mathbf{X} = \mathbf{x} - \mathbf{d}_k, t) - r_k(\mathbf{x})P(\mathbf{X} = \mathbf{x}, t)].
\]  

[3.3]

Considering one component of interest \( X_i \) and summing over all other components \( X_j \neq X_i \) gives for each value of \( X_i = x_i \) where \( x_i = 0, 1, 2, \ldots \)

\[
\frac{d}{dt} \sum_{j \neq i} \sum_{x_j = 0} ^\infty P(\mathbf{X} = \mathbf{x}, t) = \sum_k \sum_{j \neq i} \sum_{x_j = 0} ^\infty [r_k(\mathbf{x} - \mathbf{d}_k)P(\mathbf{X} = \mathbf{x} - \mathbf{d}_k, t) - r_k(\mathbf{x})P(\mathbf{X} = \mathbf{x}, t)]
\]

\[
= \sum_k \sum_{j \neq i} \sum_{x_j = 0} ^\infty r_k(\mathbf{x} - \mathbf{d}_k)P(\mathbf{X} = \mathbf{x} - \mathbf{d}_k|x_i - d_{ki}, t)P(X_i = x_i - d_{ki}, t)
\]

\[
- \sum_k \sum_{j \neq i} \sum_{x_j = 0} ^\infty r_k(\mathbf{x})P(\mathbf{X} = \mathbf{x}|x_i, t)P(X_i = x_i, t),
\]

where Term 1 simplifies to conditional average rates through a shift of indices in the summation

\[
\sum_{j \neq i} \sum_{x_j = 0} ^\infty r_k(\mathbf{x} - \mathbf{d}_k)P(\mathbf{X} = \mathbf{x} - \mathbf{d}_k|x_i - d_{ki}, t)
\]

\[
= \sum_{x_1 = 0} ^\infty \cdots \sum_{x_{i-1} = 0} ^\infty \sum_{x_{i+1} = 0} ^\infty \cdots \sum_{x_\ell = 0} ^\infty r_k(x_1 - d_{k1}, \ldots, x_i - d_{ki}, \ldots, x_\ell - d_{k\ell})
\]

\[
\times P(X_1 = x_1 - d_{k1}, \ldots, X_i = x_i - d_{ki}, \ldots, X_\ell = x_\ell - d_{k\ell}|x_i - d_{ki}, t)
\]

\[
= \sum_{x_1 = d_{k1}} ^\infty \cdots \sum_{x_{i-1} = d_{k,i-1}} ^\infty \sum_{x_{i+1} = d_{k,i+1}} ^\infty \cdots \sum_{x_\ell = d_{k\ell}} ^\infty r_k(x_1, \ldots, x_i - d_{ki}, \ldots, x_\ell)
\]

\[
\times P(X_1 = x_1, \ldots, X_i = x_i - d_{ki}, \ldots, X_\ell = x_\ell|x_i - d_{ki}, t)
\]

\[
= \langle r_k|x_i - d_{ki}, t \rangle
\]

where the penultimate step follows because the missing terms are all equal to zero for systems with non-negative abundances: First, consider those reactions for which the step-size of \( X_j \) is positive, i.e., \( d_{kj} > 0 \) then the missing terms correspond to states \( X_j = x_j \) where \( x_j = -d_{kj}, -d_{kj} + 1, \ldots, 1 \). But for all those states the probability to find the system in \( X_j = x_j \) must be equal to zero because they all correspond to negative abundances of \( X_j \). Similarly, for reactions for which \( d_{jk} < 0 \) the extra terms correspond to states \( X_j = x_j \) where \( x_j = 0, 1, \ldots, -d_{jk} - 1 \). But in all those states the rate of this reaction must be equal to zero because the corresponding transition would take the system to negative abundances of \( X_j \).

Considering the above relations at stationarity we thus obtain

\[
\frac{dP_{ss}(X_i = x_i, t)}{dt} = 0 = \sum_k [\langle r_k|x_i - d_{ki}\rangle P_{ss}(X_i = x_i - d_{ki}) - \langle r_k|x_i\rangle P_{ss}(X_i = x_i)],
\]
where the conditional averages are now taken over the stationary state distribution \( P_{ss}(x) \), which corresponds to Eq. (2) presented in the main text, and previously derived in (3).

Overall, the above derivation is a trivial summation – the simplicity of the argument is obscured through the many indices and keeping track of the variables in general. In any specific system, the summation is readily performed similar to how the moment evolution equations can be readily derived from the master equation using a simple summation over states and shifting of the indices.

4. Derivation of probability balance relations for specific systems

Our starting point is the probability balance equation which must hold for any given molecular component \( X_i \) of a complex network that reaches a stationary state in which its probability distribution \( P_{ss}(x_i) \) no longer changes. This equations follows trivially from the master equation (see above) and was previously discussed elsewhere (3).

\[
0 = \sum_k \left[ \langle r_k(x) | x_i = m - d_{ki} \rangle P_{ss}(x_i = m - d_{ki}) - \langle r_k(x) | x_i = m \rangle P_{ss}(x_i = m) \right] \quad \forall m \in \mathbb{N}_0. \quad [S.4]
\]

Here, we denoted the states using the index variable \( m \) to avoid any confusion with the component itself.

**General systems with linear degradation of \( X_3 \).** First we derive the invariant balance equation for the following class of systems introduced in the main text.

\[
\begin{align*}
\begin{array}{c}
\xrightarrow{f(x_2)} & x_3 \\
\xrightarrow{x_3/\tau_3} & x_3 + 1
\end{array} + \left[ \begin{array}{c}
X_1, X_2 \\
\text{production & degradation}
\end{array} \right].
\end{align*}
\]

[S.5]

Applying Eq. S.4 to the specific dynamics of Eq. S.5 leads to

\[
\langle f(x_2) | x_3 = m-1 \rangle P(x_3 = m-1) + \frac{m+1}{\tau_3} P(x_3 = m+1) = \langle f(x_2) | x_3 = m \rangle P(x_3 = m) + \frac{m}{\tau_3} P(x_3 = m) \quad \forall m \in \mathbb{N}_0.
\]

[S.6]

where we have dropped the subscript to denote the stationary state probability distribution for notational convenience. For biochemical reaction systems, for the lowest state \( m = 0 \) the relationship simplifies to

\[
\langle f(x_2) | x_3 = 0 \rangle P(x_3 = 0) = \frac{1}{\tau_3} P(x_3 = 1),
\]

[S.7]

because molecular numbers cannot be negative (which implies that \( P(x_3 = -1) = 0 \)). Combining Eq. S.6 with Eq. S.7 then gives rise to the following recursion relation

\[
\langle f(x_2) | x_3 = m \rangle P(x_3 = m) = \frac{m+1}{\tau_3} P(x_3 = m + 1) \quad \forall m \in \mathbb{N}_0,
\]

[S.8]

which is Eq. (4) presented in the Main Text.

To get the matrix form of these relations which we use in our optimization algorithm, we re-wrote the conditional averages in Eq. S.8

\[
\sum_i f(x_2 = i) P(x_2 = i, x_3 = m) = \frac{m+1}{\tau_3} P(x_3 = m + 1) \quad \forall m \in \mathbb{N}_0,
\]

[S.9]

where \( P(x_2, x_3) \) is the joint probability of \( X_2 \) and \( X_3 \).

In experiments and computational simulations there will always be a maximum value \( m_{max} \) for which \( P(x_3 = m) = 0 \quad \forall m > m_{max} \) such that Eq. S.9 can be written in matrix form

\[
Gf = h
\]

[S.10]

where \( G_{ij} = P(x_2 = j, x_3 = i) \), \( f_i = f(x_2 = i) \) and \( h_i = (i+1)P(x_3 = i + 1)/\tau_3 \).
General systems with non-linear degradation of $X_3$. The derivation of the probability flux balance relations for systems in which $X_3$ is degraded as a dimer

$$
\begin{align*}
X_3 & \xrightarrow{\gamma x_3 (x_3 - 1)} x_3 + 1 \\
X_3 & \xrightarrow{x_3 - 2} x_3 - 2
\end{align*}
\tag{S.11}
$$

is equivalent to the one in the previous section. First, we apply Eq. S.4 the specified dynamics of Eq. S.11 to obtain

$$
\langle f(x_2) | x_3 = m - 1 \rangle P(x_3 = m - 1) + \gamma m (m - 1) P(x_3 = m) = \langle f(x_2) | x_3 = m \rangle P(x_3 = m) + \gamma (m + 2) (m + 1) P(x_3 = m + 2) \quad \forall m \in \mathbb{N}_0.
\tag{S.12}
$$

This is a second order recurrence equation that simplifies in the states $m = 0$ and $m = 1$ for which the degradation rate is zero, such that for $m = 0$

$$
\langle f(x_2) | x_3 = 0 \rangle P(x_3 = 0) = 2 \gamma P(x_3 = 2),
\tag{S.13}
$$

because $P(x_3 = -1) = 0$. For $m = 1$ we obtain

$$
\langle f(x_2) | x_3 = 0 \rangle P(x_3 = 0) - \langle f(x_2) | x_3 = 1 \rangle P(x_3 = 1) + 6 \gamma P(x_3 = 3)
= -\langle f(x_2) | x_3 = 1 \rangle P(x_3 = 1) + 2 \gamma P(x_3 = 2) + 6 \gamma P(x_3 = 3) = 0.
\tag{S.14}
$$

Combining these, then leads to the following general balance relation

$$
\langle f(x_2) | x_3 = m \rangle P(x_3 = m) = \gamma m (m + 1) P(x_3 = m + 1) + \gamma (m + 1) (m + 2) P(x_3 = m + 2) \quad \forall m \in \mathbb{N}_0,
\tag{S.15}
$$

as stated in the main text.

5. Heuristic approach to determine the regularization parameter

Following the assumption that biochemical reaction rates are reasonably smooth we introduced a regularization term in our optimization in order to penalize large jumps in the reaction rate $f(x_2)$

$$
\min_{\mathbf{f}} \left\{ \frac{||\mathbf{Gf} - \mathbf{h}||^2}{\text{unbiased objective function}} + \epsilon ||\mathbf{\Gamma f}||^2 \right\} \quad \text{s.t.} \quad \mathbf{f} \geq 0,
\tag{S.16}
$$

where the regularization term with $\Gamma_{ij} = \delta_{i,j} - 2 \delta_{i,j+1} + \delta_{i,j+2}$ essentially corresponds the discrete second derivative of $f(x_2)$. Throughout the paper we used $\epsilon = 1/\sqrt{N}$ which appropriately decreases the importance of the regularization term as the data become more precise. However, it is possible that for other systems another value for the regularization parameter may be advantageous. For completeness, we describe a general method to determine an “optimal” value for $\epsilon$ for arbitrary cases.

The goal is to pick a regularization parameter which is strong enough to smoothen our estimated reaction rate without introducing too strong a linear bias to the optimization. This can be achieved by determining the value of the unbiased objective function of Eq. S.16 after optimization as a function of the regularization parameter. In Fig. S1A we illustrate the result for the “bistable system” introduced earlier. We see that after a certain value for $\epsilon$ there is a sharp increase of the objective function which means that the optimization gets heavily biased by the regularization term beyond a certain value. As a heuristic we thus propose to pick a value for the regularization parameter just before the sharp increase. Large values of $\epsilon$ essentially force the resulting $f(x_2)$ to be a straight line leading to the plateau in the objective function.
**Fig. S1.** The optimal value for the regularization parameter can be heuristically determined through its effect on the objective function. A) Choosing a regularization parameter strong enough to smoothen the estimated reaction rate without introducing a significant bias leads to the best inference. In this case it was determined to be $\epsilon = 0.01$ (orange arrow). B) The regularization is essential to achieve optimal results. Shown here are the inferred rate with the chosen $\epsilon = 0.01$ (orange crosses) and without regularization, i.e., $\epsilon = 0$ (red crosses).
Fig. S2. For fast upstream variables the conditional average follows the reaction rate function. We consider the “noise enhancing” system defined above with changing lifetimes ratios. For such systems, inferring $f_3(x_2)$ is straightforward when $\tau_2 \gg \tau_3$ and the variability of $X_2$ is slow enough for $X_3$ to adjust to $X_2$-levels and the conditional average $\langle x_3 | x_2 \rangle$ directly identifies the production rate of $X_3$. In contrast, when upstream variability in $X_2$ is fast compared to the life-time of $X_3$ the conditional average no longer follows the production rate.
Fig. S3. Inference quality behaviour for individual states. For each of the four systems of Fig. 2 in the main text we analyzed the error behaviour for a given state $x_2 = m$. Initially, the relative error of the inferred reaction rate $f(x_2)$ approximately decreases $\propto 1/\sqrt{N}$ (dashed lines) as the number of sampling points $N$ increases. For large $N$, the relative error levels off, with higher probability states reaching a lower plateau than those only visited rarely. For very large sampling some of the errors can be seen to increase slightly as our choice of $\epsilon = 1/\sqrt{N}$ becomes suboptimal. The inset depicts the true function, the specific states considered here, and the respective probability distribution of $X_2$. 

Timon Wittenstein, Nava Leibovich, and Andreas Hilfinger 9 of 11
Fig. S4. Adding an additional monotonicity constraint improves inference in unfavourable regions. Analogous to Fig. 4 in the main text we change the shape of the production rate $f(x_2) = \lambda x_2^n/(x_2^n + K^n)$ by varying $K$ and $n$. Shown here is the error of the inferred $f(x_2)$ when applying our method with an additional monotonicity constraint. The quality of the inference is significantly improved in almost all regimes.
References

1. JL Doob, Markoff chains–denumerable case. *Transactions Am. Math. Soc.* **58**, 455–473 (1945).
2. DT Gillespie, A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. computational physics* **22**, 403–434 (1976).
3. A Hilfinger, TM Norman, G Vinnicombe, J Paulsson, Constraints on fluctuations in sparsely characterized biological systems. *Phys. Rev. Lett.* **116**, 058101 (2016).