Quantifying Extrinsic Noise in Gene Expression Using the Maximum Entropy Framework

Purushottam D. Dixit*

Biosciences Department, Brookhaven National Laboratory, Upton, New York

Submitted April 27, 2013, and accepted for publication May 3, 2013.

*Correspondence: pdixit@bnl.gov

Supplementary materials

I. MAXIMUM ENTROPY FORMALISM

The maximum entropy formalism allows one to estimate the probability distribution $\{p_i\}$ of states $\{i\}$ of a system from limited information. Briefly, the ME framework for estimating probabilities $\{p_i\}$ involves maximizing the entropy function $\mathcal{S}(\{p_i\})$ subject to constraining the values of certain variables (1). For example, if $\langle X_1\rangle, \langle X_2\rangle, \ldots, \langle X_N\rangle$ are the mean values of variables $X_1, X_2, \ldots, X_N$ respectively, then the probabilities $\{p_i\}$ of states i are estimated by the maximizing the constrained objective function in Eq. 1,

$$\mathcal{S}(\{p_i\}) - \sum_k \lambda_k \left(\left(\sum_i p_i \cdot X_k(i) \right) - \langle X_k \rangle \right) + \alpha \left(\sum_i p_i - 1 \right) \tag{1}$$

Here, $\{\lambda_k\}$ and α are Lagrange multipliers that ensure that the constraints are satisfied and that the probabilities are normalized. The entropy is a non-negative convex function of the probabilities and is usually defined as (2, 3)

$$\mathcal{S}(\{p_i\}) = -\sum_i p_i \log p_i. \tag{2}$$

The maximization of Eq. 1 estimates probabilities

$$p_i = \frac{1}{\mathcal{Z}(\{\lambda_k\})} exp\left(-\sum_k \lambda_k X_k(i) \right). \tag{3}$$

Here,

$$\mathcal{Z}(\{\lambda_k\}) = \sum_i exp\left(-\sum_k \lambda_k X_k(i) \right) \tag{4}$$

is the partition function. The Lagrange multipliers λ_k are determined by solving

$$-\frac{\partial \log \mathcal{Z}}{\partial \lambda_k} = \langle X_k \rangle. \tag{5}$$

Notice that the probabilities depend exponentially on the constrained quantities (compare to Eq. 9, Eq. 13, and Eq. 15 in main text).

II. CALCULATION OF VARIOUS MOMENTS

If $P(k)$, $P(\mathrm{m},k)$, and $P(\mathrm{m})$ are given by Eq. 20, Eq. 21, and Eq. 23, in the main text, the various moments are,

$$\bar{\mathrm{m}}(k) = \sum_m mP(\mathrm{m}|k) = k, \tag{6}$$

$$\bar{\mathrm{m}}^2(k) = \sum_m m^2P(\mathrm{m}|k) = k^2 + k, \tag{7}$$

$$\langle \mathrm{m} \rangle = \int \bar{m}(k)P(k) = \langle k \rangle = \mu. \tag{8}$$

Similarly,

$$\langle \mathrm{m}^2 \rangle = \mu^2 + \mu + \frac{\mu}{\alpha} = \langle k^2 \rangle + \langle k \rangle, \tag{9}$$

$$\langle \mathrm{m}k \rangle = \mu^2 + \frac{\mu}{\alpha}. \tag{10}$$

The total noise is defined as

$$\eta_{\mathrm{T}} = \frac{\langle \mathrm{m}^2 \rangle - \langle \mathrm{m} \rangle^2}{\langle \mathrm{m} \rangle^2} = \frac{1}{\mu}\left(1 + \frac{1}{\alpha}\right) \tag{11}$$

The intrinsic noise is defined as

$$\eta_{\mathrm{I}} = \frac{1}{\langle \mathrm{m} \rangle^2}\int (\bar{\mathrm{m}}^2(k) - \bar{\mathrm{m}}(k)^2)P(k)dk = \frac{1}{\langle \mathrm{m} \rangle^2}\int kP(k)dk = \frac{1}{\mu} \tag{12}$$

III. HOW TO INCORPORATE PROMOTER FLUCTUATIONS?

When the promoter fluctuations are explicitly modeled, the distribution of mRNA copy numbers can be obtained in a closed form. Under simplifying conditions, the distribution of mRNA copy numbers becomes a negative binomial distribution (4, 5). Here, we sketch a rough outline of incorporating extrinsic noise beyond promoter fluctuations within the ME framework.

For simplicity, let us assume that the mRNA copy number distribution $P(\mathrm{m};\alpha,\beta)$ is given by the Gamma distribution (the continuous counterpart of the negative binomial distribution). The parameter α is related to the half life of the activated (transcribable) state of the DNA while β is the mean number of mRNA transcripts produced when the DNA is in the activated state. If there is no extrinsic noise beyond the promoter fluctuations, it is easy to show that

$$\langle \mathrm{m} \rangle = \alpha\beta, \tag{13}$$

$$\langle \mathrm{m}^2 \rangle = \alpha\beta^2 + \langle \mathrm{m} \rangle^2, \tag{14}$$

$$\langle \mathrm{m}^3 \rangle = \alpha(\alpha+1)(\alpha+2)\beta^3. \tag{15}$$

The Gamma distribution has only two free parameters and the skewness is not independent of the second moment and is given by

$$\gamma_1 = \frac{2}{\sqrt{\alpha}} = 2\sqrt{\eta_{\mathrm{T}}}. \tag{16}$$

Eq. 16 roughly holds when promoter fluctuations are the major contributor to extrinsic noise. A deviation from Eq. 16 should prompt an exploration of the cell-to-cell variation in the parameters of the Gamma distribution themselves.

In real cells, the parameters α and β of the Gamma distribution may be variable. The ME framework estimates the joint distribution $P(\alpha, \beta)$ from Eq. 15 of the main text. Even though the entropy of the Gamma distribution has a closed form, inserting the entropy in Eq. 15 of the main text and constraining the average value of α and β results in an expression for $P(\alpha, \beta)$ that does not have a closed form. Instead, if we assume that $S(\alpha, \beta) \sim \log \sigma_{\mathrm{m}}$ i.e. the entropy scales as the variance of the mRNA copy number m (which is a good approximation), we get

$$P(\alpha, \beta) = \frac{\zeta^{\lambda+1}\xi^{2\lambda+1}\left(\alpha\beta^2\right)^{\lambda} e^{-\alpha\zeta-\beta\xi}}{\Gamma(\lambda+1)\Gamma(2\lambda+1)}. \tag{17}$$

Eq. 17 is equivalent to Eq. 20 in the main text when promoter fluctuations are explicitly modeled. Notice that the distribution of the parameters α and β themselves is described by a product of two independent Gamma distribution. The variability in α and β can now be ascribed to other global extrinsic factors.

Notice that $P(\alpha, \beta)$ is parametrized by three parameters λ, ζ, and ξ. The resuling marginal distribution $P(\mathrm{m})$ for m will also be parametrized by three parameters. Unfortunately, this marginal distribution doesn't have a closed form either. Yet, we can indeed compute quantities such as the total, intrinsic, and extrinsic noise from Eq. 17. Notice that since $P(\mathrm{m})$ has three free parameters, the skewness estimated from Eq. 17 may not be equal to twice the square root of the total noise. Computing various moments from Eq. 17, we get

$$\langle \mathrm{m} \rangle = \frac{(\lambda+1)(2\lambda+1)}{\zeta\xi}, \tag{18}$$

$$\langle \mathrm{m}^2 \rangle = \frac{2(\lambda+1)^2(2\lambda+1)(\zeta+\lambda+2)}{\zeta^2\xi^2}, \tag{19}$$

$$\langle \mathrm{m}^3 \rangle = \frac{2(\lambda+1)^2(2\lambda+1)(2\lambda+3)\left(3\zeta\lambda+2\zeta(\zeta+3)+\lambda^2+5\lambda+6\right)}{\zeta^3\xi^3} \tag{20}$$

The intrinsic, extrinsic, and the total noise are given by,

$$\eta_{\mathrm{I}} = \frac{2\zeta}{2\lambda+1}, \tag{21}$$

$$\eta_{\mathrm{E}} = \frac{3}{2\lambda+1}, \tag{22}$$

$$\eta_{\mathrm{T}} = \frac{2\zeta+3}{2\lambda+1}. \tag{23}$$

And the skewness γ_1 is

$$\gamma_1 = \frac{2\sqrt{\lambda+1}\sqrt[4]{2\lambda+1}\left(6\zeta^2+(4\zeta(\zeta+3)+11)\lambda+15\zeta+13\right)}{(2\zeta+3)^{3/4}(\zeta\xi)^{3/2}}. \tag{24}$$

The developed framework can potentially parse intrinsic and extrinsic contributions if higher moments of the mRNA copy number are carefully estimated. Even though the theoretical framework allows it, unfortunately, currently published experimental data does not permit us to do the same.

IV. NUMERICAL SIMULATIONS

The synthesis and degradation of the mRNA of any given gene competes with the synthesis and degradation of all other co-expressed genes. Moreover, the cellular machinery that carries out these reactions itself comprises of proteins and mRNAs and is subject to cell to cell variation. We devise a simple scheme to mimic the coupled dynamics of synthesis and degradation of the cellular machinery with the dynamics of synthesis and degradation of the mRNA of a given gene.

The transcription apparatus is represented by a single protein RNAP and the mRNA degradation apparatus is represented by a single protein RNAase. The rate of synthesis γ of the given mRNA depends linearly on [RNAP] the concentration of the proxy for the RNA polymerase complex. Similarly, the rate of degradation δ depends linearly on the concentration [RNAase] of the

Parameters for the simulation

Parameter	Case 1	Case 2	Case 3
γ_1	2.0	2.0	2.0
γ_2	2.0	2.0	2.0
γ_0	0.9	1.6	1.0
κ_1	0.5	0.5	0.5
κ_2	0.5	0.5	0.5
δ_1	0.1	0.1	0.1
δ_2	0.1	0.1	0.1
δ_0	0.227	0.5	0.1
Δ_1	0.15	0.65	0.55
Δ_2	0.15	0.65	0.55
$[\text{DNA}]_{\text{RNAP}}$	5	5	5
$[\text{DNA}]_{\text{RNAase}}$	5	5	5
$[\text{DNA}]_{\text{Gene}}$	1	1	1

TABLE I. The details of the parameters for the numerical simulation of mRNA synthesis. All rates are in s^{-1} and all copy numbers are integers.

proxy for the RNAase enzyme ($\gamma = \gamma_0[\text{RNAP}]$ and $\delta = \delta_0[\text{RNAase}]$).

$$\begin{aligned}
\text{DNA} &\xrightarrow{\gamma_1} \text{rRNAP} \\
\text{DNA} &\xrightarrow{\gamma_2} \text{rRNAase} \\
\text{DNA} &\xrightarrow{\gamma} \text{rGene} \\
\text{rRNAP} &\xrightarrow{\kappa_1} \text{RNAP} \\
\text{rRNAase} &\xrightarrow{\kappa_2} \text{RNAase} \\
\text{rRNAP} &\xrightarrow{\delta_1} \phi \\
\text{rRNAase} &\xrightarrow{\delta_2} \phi \\
\text{rGene} &\xrightarrow{\delta} \phi \\
\text{RNAP} &\xrightarrow{\Delta_1} \phi \\
\text{RNAase} &\xrightarrow{\Delta_2} \phi
\end{aligned}$$

The dynamics of the synthesis and degradation of the mRNA of the given gene and RNAP and RNAase is propagated using the Gillespie's algorithm (6) for $2 \cdot 10^8$ steps. Data is stored every 5000th step after an initial equilibration of 50000 steps. The initial concentrations of all species

except the copy number of the each gene on the DNA at $t = 0$ was set to 0. Table I gives the details of the conditions that were employed to construct the histograms (red points in Fig. 2 of the main text).

---

[1] Jaynes, E. T., 1957. Information theory and statistical mechanics I. *Physical Review* 106:620–630.

[2] Shore, J., and R. Johnson, 1980. Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Trans. Info. Theory* 26:26–37.

[3] Shannon, C. E., and W. Weaver, 1949. The mathematical theory of communication. Univ. Illinois Press, Urbana Champagne.

[4] Raj, A., C. S. Peskin, D. Tranchina, D. Y. Vargas, and S. Tyagi, 2006. Stochastic mRNA synthesis in mammalian cells. *PLoS biology* 4:e309.

[5] So, L.-H., A. Ghosh, C. Zong, L. A. Sepúlveda, R. Segev, and I. Golding, 2011. General properties of transcriptional time series in Escherichia coli. *Nature genetics* 43:554–560.

[6] Gillespie, D. T., 1977. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem* 81:2340–2361.