Fitting intrinsic and extrinsic noise contributions

Erik van Nimwegen
Computational and Systems Biology, Biozentrum, University of Basel,
Klingelbergstrasse 50-70, 4056 Basel, Switzerland

E-mail: erik.vannimwegen@unibas.ch

Contents

1 Relation between promoter mean and variance

For a promoter p the total variance (in the logarithm of the number of GFP per cell) is a combination of the intrinsic Poissonian noise (which would result if transcription, translation, mRNA decay, and dilution rates were all not fluctuating) plus an extrinsic part due to fluctuations in these rates. The intrinsic noise is proportional to the 'burst size' β (translation rate divided by the mRNA decay rate, i.e. the number of times an mRNA is translated on average before it decays) and inversely proportional to the mean number of GFP molecules per cell n. In general, the variance v_p of a particular promoter p satisfies

$$v_p = v(n_p, \sigma_p^2, \beta_p) = \sigma_p^2 \left(1 - \frac{n_{bg}}{n_p}\right)^2 + \frac{\beta_p}{n_p}\left(1 - \frac{n_{bg}}{n_p}\right), \tag{1}$$

where n_p is the mean of the logarithm of the number of GFP molecules per cell for promoter p, n_{bg} is the mean log-fluorescence of the empty plasmids (measured in units of GFP molecules per cell), σ_p^2 is the extrinsic noise, and β_p is the burst size.

Ideally we would like to infer how σ_p^2 and β_p vary across promoters, and across conditions. However, having measured the pair (n_p, v_p), we cannot uniquely solve for both σ_p^2 and β_p from equation (1). However, we prefer solutions in which σ^2 and β vary little across promoters and conditions compared to solutions in which these parameters vary a lot. Or to put it differently, we will use a prior probability distribution for both σ^2 and β that assumes these quantities vary as little as is consistent with the measurements and equation (1).

For example, if we are analyzing the means and variances of a large number of promoters in a single condition, we will assume that, for each promoter p, its extrinsic noise σ_p^2 is given by a 'typical' extrinsic noise σ^2 plus a deviation η_p

$$\sigma_p^2 = \sigma^2 + \eta_p. \tag{2}$$

Similarly, we assume that the burst size of promoter p is given by a typical burst size β plus a deviation δ_p:

$$\beta_p = \beta + \delta_p. \tag{3}$$

We will then assume that the deviations η_p and δ_p are drawn from some distribution (which has a maximum at zero) such that small deviations are favored.

Below we will need to integrate over the unknown deviations η_p and δ_p for each promoter. It is therefore helpful to write the function $v(n_p, \sigma_p^2, \beta_p)$ explicitly in terms of η_p and δ_p. If we define:

$$a_p = \left(1 - \frac{n_{bg}}{n_p}\right)^2, \tag{4}$$

and

$$b_p = \frac{1}{n_p}\left(1 - \frac{n_{bg}}{n_p}\right), \tag{5}$$

we have

$$v(n_p, \sigma_p^2, \beta_p) = a_p\sigma^2 + b_p\beta + a_p\eta_p + b_p\delta_p. \tag{6}$$

2 Probabilistic model

The first ingredient of our model is the measurement noise. That is, for each promoter p our data D_p consists of the measured values of the mean and variance, i.e. $D_p = (n_p, v_p)$ and we of course know that these measured values do not perfectly match the 'true' values, but have some measurement noise. Our experience has indicated that the noise in the measurements of the mean is much smaller than the noise in the measurements of the variance. To simplify our model, we will assume that, relative to the noise in variance and to the fluctuations in σ^2 and β, the noise in the measurement of the mean is negligible. Thus, we will assume the measured and true means n_p are equal. Given a mean n_p and given values of the extrinsic noise σ_p^2 and burst size β, the true variance is given by $v(n_p, \sigma_p^2, \beta_p)$. We will assume that the deviation between the true and measured variance is Gaussian distributed with variance ϵ^2:

$$P(v_p|n_p, \sigma^2, \beta, \eta_p, \delta_p) = \frac{1}{\epsilon\sqrt{2\pi}} \exp\left[-\frac{(v_p - a_p\sigma^2 - b_p\beta - a_p\eta_p - b_p\delta_p)^2}{2\epsilon^2}\right]. \tag{7}$$

Note that the probability of the data D_p is formally given by

$$P(D_p|\sigma^2, \beta, \eta_p, \delta_p) = P(v_p|n_p, \sigma^2, \beta, \eta_p, \delta_p)P(n_p), \tag{8}$$

where $P(n_p)$ is some prior probability that a randomly chosen promoter has mean expression level n_p. Since this distribution plays no role in our inference, we will assume a uniform distribution, i.e. $P(n_p) = \text{constant}$. We then have

$$P(D_p|\sigma^2, \beta, \eta_p, \delta_p) \propto P(v_p|n_p, \sigma^2, \beta, \eta_p, \delta_p). \tag{9}$$

In other words, equation (7) is, up to a multiplicative constant, also equal to the probability of the data D_p given the parameters $(\sigma^2, \beta, \eta_p, \delta_p)$. We will need that in our calculation below.

We now assume that each η_p is drawn from a Gaussian distribution with mean zero and variance λ^2, i.e.

$$P(\eta|\lambda) = \frac{1}{\sqrt{2\pi}\lambda} \exp\left[-\frac{\eta^2}{2\lambda^2}\right]. \tag{10}$$

Similarly, we assume that each δ_p is drawn from a distribution with mean zero and variance γ^2:

$$P(\delta|\gamma) = \frac{1}{\sqrt{2\pi}\gamma} \exp\left[-\frac{\delta^2}{2\gamma^2}\right]. \tag{11}$$

Using this, the joint probability for both the data and the deviations η_p and δ_p is given by

$$P(D_p, \eta_p, \delta_p|\sigma^2, \beta, \lambda, \gamma) \propto= \frac{1}{\epsilon\lambda\gamma} \exp\left[-\frac{(v_p - a_p\sigma^2 - b_p\beta - a_p\eta_p - b_p\delta_p)^2}{2\epsilon^2} - \frac{\delta_p^2}{2\gamma^2} - \frac{\eta_p^2}{2\lambda^2}\right]. \tag{12}$$

To obtain the probability of the data as a function of the parameters $(\sigma^2, \beta, \lambda, \gamma)$, i.e. $P(D_p|\sigma^2, \beta, \lambda, \gamma)$ we need to integrate out the unknown deviations η_p and δ_p, i.e.

$$P(D_p|\sigma^2, \beta, \lambda, \gamma) = \int_{-\infty}^{\infty} d\eta_p d\delta_p P(D_p, \eta_p, \delta_p|\sigma^2, \beta, \lambda, \gamma). \tag{13}$$

This integral can be performed analytically and we obtain

$$P(D_p|\sigma^2, \beta, \lambda, \gamma) \propto \frac{1}{\sqrt{\epsilon^2 + b_p^2\gamma^2 + a_p^2\lambda^2}} \exp\left[-\frac{(v_p - a_p\sigma^2 - b_p\beta)^2}{2(\epsilon^2 + b_p^2\gamma^2 + a_p^2\lambda^2)}\right]. \tag{14}$$

This equation is relatively straight-forward to understand. the difference between the measured value v_p and the predicted typical variance $a_p\sigma^2 + \beta b_p$ is Gaussian distributed with a variance that is the sum of the measurements variance ϵ^2 and the variance due to variation in σ^2, i.e. $a_p^2\lambda^2$, and the variance due to variation in β, i.e. $b_p^2\gamma^2$.

The probability of the entire data-set, given the parameters σ^2, β, λ, and γ is given by taking the product of the above expression over all promoters, i.e. the probability of the observed data for each promoter

$$P(D|\sigma^2, \beta, \lambda, \gamma) = \prod_{p=1}^{n} P(D_p|\sigma^2, \beta, \lambda, \gamma), \tag{15}$$

where n is the total number of promoters.

Assuming λ and γ as fixed and given for the moment, we will now determine the values of σ^2 and β that maximize the probability $P(D|\sigma^2, \beta, \lambda, \gamma)$, i.e. the maximum likelihood values of these two parameters. If we define the log-likelihood as $L(\sigma^2, \beta, \lambda, \gamma) = \log[P(D|\sigma^2, \beta, \lambda, \gamma)]$, we find that

$$L(\sigma^2, \beta, \lambda, \gamma) = \text{cons.} - \sum_p \left(\frac{(v_p - a_p\sigma^2 - b_p\beta)^2}{2(\epsilon^2 + b_p^2\gamma^2 + a_p^2\lambda^2)} + \frac{1}{2}\log\left[\epsilon^2 + b_p^2\gamma^2 + a_p^2\lambda^2\right] \right). \tag{16}$$

Note that the last term does not depend on σ^2 and β. Thus, to determine the optimum with respect to σ^2 and β we only need to consider the first term in the . For notational simplicity, we define

$$w_p = \frac{1}{\epsilon^2 + b_p^2\gamma^2 + a_p^2\lambda^2}, \tag{17}$$

keeping in mind that w_p depends on the variances ϵ^2, γ^2, and λ^2. We thus need to optimize the following expression with respect to σ^2 and β:

$$-\sum_p w_p(v_p - a_p\sigma^2 - b_p\beta)^2. \tag{18}$$

Taking the derivatives with respect to σ^2 and β, and demanding that these are zero, we find for the optimal values of σ^2 and β

$$\sigma_*^2 = \frac{\langle wbv\rangle\langle wab\rangle - \langle wb^2\rangle\langle wav\rangle}{\langle wab\rangle^2 - \langle wa^2\rangle\langle wb^2\rangle}, \tag{19}$$

and

$$\beta_* = \frac{\langle wav\rangle\langle wab\rangle - \langle wa^2\rangle\langle wbv\rangle}{\langle wab\rangle^2 - \langle wa^2\rangle\langle wb^2\rangle}. \tag{20}$$

In these equations we have used averages of products over the w_p, a_p, b_p, and v_p. For example

$$\langle wav\rangle = \frac{1}{n}\sum_{p=1}^{n} w_p a_p v_p, \tag{21}$$

$$\langle wa^2\rangle = \frac{1}{n}\sum_{p=1}^{n} w_p a_p^2, \tag{22}$$

and so on, where n is the number of promoters.

We can now plug the values σ_*^2 and β_* into the expression $\log[P(D|\sigma^2, \beta, \lambda, \gamma)]$ and obtain a log-likelihood as a function of λ and γ only, i.e.

$$L(\lambda, \gamma) = L(\sigma_*^2, \beta_*, \lambda, \gamma). \tag{23}$$

It is important to realize that, in this expression, the optimal values σ_*^2 and β_* are themselves functions of λ and γ, through the dependence of the w_p on λ and γ.

In summary, given our measurements (v_p, n_p) for each promoter p, we can determine the constants a_p and b_p for each promoter. For a given combination of the variances γ^2 and λ^2, we can then calculate all the w_p. Using the equations (19) and (20) above we can then determine the optimal σ_*^2 and β_*, and plugging these back into the expression for the log-likelihood (16), we obtain the log-likelihood of the data as a function of λ and γ (I am assuming we have a reasonable estimate of ϵ from replicate measurements of the noise, see below). My proposal is to implement this into a little computer program that calculates the likelihood as a function of λ and γ given an input data-set of values (v_p, n_p). The idea would be to then numerically determine the optimum in λ_* and γ_*.

2.1 Distribution of β and σ^2 given λ_* and γ_*

Once we have determined the optimum (λ_*, γ_*) we can calculate the distribution of σ^2 and β conditioned on (λ_*, γ_*). Note that, formally, if we wanted to exactly calculate the posterior distribution $P(\sigma^2, \beta|D)$ we would have to marginalize over λ and γ, i.e.

$$P(\sigma^2, \beta|D) = \int_0^\infty d\lambda d\gamma P(\sigma^2, \beta, \lambda, \gamma|D). \tag{24}$$

Unfortunately this integral cannot be done analytically. We will thus approximate the integral by the values at its maximum (λ_*, γ_*), i.e. by using the approximation $P(\sigma^2, \beta|D) \approx P(\sigma^2, \beta|\lambda_*, \gamma_*, D)$.

Inspection of equation (16) shows that the posterior distribution $P(\sigma^2, \beta|D)$ is a *multi-variate Gaussian* with mean (σ_*^2, β_*), as given by equations (19) and (20) and having the general form

$$P(\sigma^2, \beta|D) \propto \exp\left[-\frac{n\langle wa^2\rangle}{2}(\sigma^2 - \sigma_*^2)^2 - \frac{n\langle wb^2\rangle}{2}(\beta - \beta_*)^2 - n\langle wab\rangle(\sigma^2 - \sigma_*^2)(\beta - \beta_*)\right]. \tag{25}$$

From this, we find the components of the covariance matrix:

$$\text{var}(\sigma^2) = \frac{\langle wb^2\rangle}{n\left(\langle wa^2\rangle\langle wb^2\rangle - \langle wab\rangle^2\right)}, \tag{26}$$

$$\text{var}(\beta) = \frac{\langle wa^2\rangle}{n\left(\langle wa^2\rangle\langle wb^2\rangle - \langle wab\rangle^2\right)}, \tag{27}$$

and

$$\text{covar}(\sigma^2, \beta) = \frac{\langle wab\rangle}{n\left(\langle wa^2\rangle\langle wb^2\rangle - \langle wab\rangle^2\right)}. \tag{28}$$

Note that these scale as $1/n$, with the number of promoters n.

2.2 Estimating η_p and δ_p for individual promoters

Finally, we want to calculate best guesses and error bars for the deviations in extrinsice noise η_p and the deviation in burst size δ_p for each promoter p. To do this we go back to equation (12) for the probability $P(D_p, \eta_p, \delta_p|\sigma^2, \beta, \lambda, \gamma)$ of the data D_p as well as the promoter-specific deviations η_p and δ_p. First of all, we use the optimal values (λ_*, γ_*) for λ and γ. Second, we note that the posterior $P(\eta_p, \delta_p|D_p, \sigma^2, \beta, \lambda_*, \gamma_*)$ is simplify proportional to the joint probability $P(D_p, \eta_p, \delta_p|\sigma^2, \beta, \lambda_*, \gamma_*)$. Finally, we use the posterior distribution over (σ^2, β) using *all* data to calculate

$$P(\eta_p, \delta_p|D, \lambda_*, \gamma_*) = \int d\sigma^2 d\beta P(\eta_p, \delta_p|D_p, \sigma^2, \beta, \lambda_*, \gamma_*)P(\sigma^2, \beta|D, \lambda_*, \gamma_*). \tag{29}$$

Although the algebra is a bit tedious (and I hope I did not make errors), these integrals can also be performed analytically. The result shows that the posterior distribution of the deviations (η_p, δ_p) is a multi-variate Gaussian as well, with a maximum at

$$\eta_p^* = \langle \eta_p\rangle = v_p^* \frac{a_p\lambda_*^2}{\sigma_p^2 + a_p^2\lambda_*^2 + b_p^2\gamma_*^2}, \tag{30}$$

and

$$\delta_p^* = \langle \delta_p \rangle = v_p^* \frac{b_p \gamma_*^2}{\sigma_p^2 + a_p^2 \lambda_*^2 + b_p^2 \gamma_*^2}, \tag{31}$$

where we have defined

$$v_p^* = v_p - a_p \sigma_*^2 - b_p \beta_*, \tag{32}$$

and

$$\sigma_p^2 = \epsilon^2 + \frac{a_p^2 \langle wb^2 \rangle + b_p^2 \langle wa^2 \rangle - 2 a_p b_p \langle wab \rangle}{n \left(\langle wa^2 \rangle \langle wb^2 \rangle - \langle wab \rangle^2 \right)}. \tag{33}$$

Finally, the components of the covariance matrix of this multi-variate Gaussian (giving the error bars on η_p and δ_p) are given by

$$\mathrm{var}(\eta_p) = \lambda_*^2 \frac{\sigma_p^2 + b_p^2 \gamma_*^2}{\sigma_p^2 + a_p^2 \lambda_*^2 + b_p^2 \gamma_*^2}, \tag{34}$$

$$\mathrm{var}(\delta_p) = \gamma_*^2 \frac{\sigma_p^2 + a_p^2 \lambda_*^2}{\sigma_p^2 + a_p^2 \lambda_*^2 + b_p^2 \gamma_*^2}, \tag{35}$$

and

$$\mathrm{covar}(\eta_p, \delta_p) = - \frac{a_p b_p \lambda_*^2 \gamma_*^2}{\sigma_p^2 + a_p^2 \lambda_*^2 + b_p^2 \gamma_*^2}. \tag{36}$$

So what to do with these results? Well, given all the measurements (n_p, v_p) we first determine the optimal parameters (λ_*, γ_*). Then, given these optimal parameters, we determine the optimal 'typical' extrinsic noise and burst sizes (σ_*^2, β_*). Note that the extrinsic noise and burst size of an individual promoter are given by $\sigma_p^2 = \sigma_*^2 + \eta_p$ and $\beta_p = \beta_* + \delta_p$. So, with the distributions over (η_p, δ_p) derived above, we can get the distribution over $(\sigma_p^2, \beta_p$ for each promoter. One thing we can do with this is sum these distributions over all promoters, and get a joint 'histogram' $P(\sigma^2, \beta)$ over all promoters. This would show us how the extrinsic noise and burst size are globally distributed and help us answer questions like:

1. Can we assume one burst size for all promoters or does the data clearly reject this?
2. If the burst size varies across promoters, does it correlate with extrinsic noise? Does it correlate with the mean level of the promoter? And so on.
3. Do σ_p^2 values correlate well with our excess noise estimates?
4. If we do the analysis on data-sets from different conditions, do we find very similar β and σ^2? Or is there a systematic change?

3 Fitting changes in β and σ^2 for a single promoter across multiple conditions

In the sections above we are assuming the we have measured means n_p and variances v_p for many promoters in one or a few conditions. We there assumed that most promoters have similar values of σ^2 and β, and tried to explain the data with the minimal possible variation in these quantities. But we can also look at the complementary problem, where we have measured (v_p, n_p) for only a small set of promoters, but in a much larger number of (related) conditions, e.g. as the environment of the cells is changed. Some of the data Gwendoline have obtained are of this type. Here we will analyze this kind of data one promoter p at a time.

This case is in fact completely analogous to the case solved above, and we can use exactly the same formulas. Specifically, we will now fit the model one promoter at a time. So instead of putting data-points (v_p, n_p) of all promoters across a single condition into the equations, we put in data-points (v_c, n_c) of a single promoter across a set of (related) conditions. The inferred values (σ_*^2, β_*) correspond to the promoter's average values of (σ, β) across the conditions. The optimal values (λ_*, γ_*) correspond (approximately) to the amount of variation in the promoter's extrinsic noise σ^2 and burst size β across the conditions. My intuition is that we should find much *smaller* values (λ_*, γ_*). That is, I expect extrinsic noise and burst size to vary much *less* across related conditions for a single promoter, than across many promoters in a single condition. Possible exceptions to this is when a promoter is directly regulated across the conditions (like recA in the cipro induction experiments).

Once we have these results, we can check things like

1. Ideally we would find that, for each promoter, there is a typical extrinsic noise σ_p^2 and a typical burst size β. That is, that the noise v_p varies with n_p in a way that is consistent with constant σ_p^2 and β_p. We would then of course check consistency with the results from fitting (n_p, v_p) across a large set of promoters.

2. Do regulated promoters indeed show large variability in σ_p^2 across the conditions, i.e. in contrast to unregulated promoters?

3. How much do the (σ_p^2, β_p) vary across promoters? Do they vary in a correlated manner?

4 Some notes regarding the implementation

- It would be most appropriate to measure log-fluorescence everywhere in units of GFP molecules. To do this one has to add a term $\log(2.884)$ to all log-fluorescence measurements (I am already assuming that all logs are natural logs). Note that this will not effect the variance v_p, but it will affect the mean n_p, i.e. it will shift the mean by $\log(2.884)$.

- When fitting across promoters in one condition, do not include the empty plasmids or empty cells, because they likely behave differently than all the others (i.e. do not share the typical σ^2 and β).

- Make sure to use the appropriate n_{bg} in the formulas (i.e. the background in the condition being considered). This means that, when fitting across conditions, that each data-point has a (slightly) different n_{bg}. Maybe the easiest way to deal with this is to provide the input data as triples (v_p, n_p, n_{bg}) or (v_c, n_c, n_{bg}) when fitting across conditions.

- We need a numerical estimate of ϵ. The best way to do this is using data from replicates. Imagine that we have set of duplicate measuremetns (v_p^1, v_p^2) for a set of promoters. The probability of this data D given ϵ (and assuming Gaussian noise in the measurements) is given by

$$P(D|\epsilon) = \frac{1}{(2\pi)^n \epsilon^n} \exp\left[-\sum_{p=1}^{n} \frac{(v_p^1 - v_p^2)^2}{4\epsilon^2} \right], \tag{37}$$

where there is a 4 instead of 2 because the difference between the two measurements involves a noise ϵ on both measurements, so that the total variance is $2\epsilon^2$. From this equation, the maximum likelihood value of ϵ is

$$\epsilon_* = \sum_{p=1}^{n} \frac{(v_p^1 - v_p^2)^2}{2n}, \tag{38}$$

i.e. just half of the average squared-deviation between the measured variances.