# Analytical distributions for stochastic gene expression: Supporting information

Vahid Shahrezaei and Peter S. Swain

## Derivation of the protein distribution for a two-stage model of gene expression

### From the master equation

The generating function for the master equation of the two-stage model satisfies (Eq. 1 in the main text)

$$\frac{1}{v}\frac{\partial F}{\partial \tau}+\frac{\partial F}{\partial v}-\gamma\left[b(1+u)-\frac{u}{v}\right]\frac{\partial F}{\partial u}=a\frac{u}{v}F \tag{23}$$

where $F(z',z)$ is defined as $\sum_{m,n}(z')^m z^n P_{m,n}$, and we have let $u=z'-1$ and $v=z-1$. If $r$ measure the distance along a characteristic which starts at $\tau=0$ with $u=u_0$ and $v=v_0$ for some constant $u_0$ and $v_0$, then Eq. 23 becomes

$$\begin{array}{lll}\frac{dv}{dr}=1 & ; & \frac{d\tau}{dr}=\frac{1}{v}\\ \frac{du}{dr}=-\gamma\left[b(1+u)-\frac{u}{v}\right] & ; & \frac{dF}{dr}=\frac{au}{v}F.\end{array} \tag{24}$$

Consequently, $v=r$ and

$$\frac{du}{dv}=-\gamma\left[b(1+u)-\frac{u}{v}\right] \tag{25}$$

which has solution

$$u(v)=\mathrm{e}^{-\gamma bv}v^{\gamma}\left[C-b\gamma\int^{v}dv'\frac{\mathrm{e}^{\gamma bv'}}{v'^{\gamma}}\right] \tag{26}$$

for a constant $C$ as can be verified by differentiation. By Taylor expanding $\mathrm{e}^{\gamma bv}$ so that $\mathrm{e}^{\gamma bv}=\sum_n\frac{(\gamma bv)^n}{n!}$, we can evaluate the integral in Eq. 26,

$$u(v)=\mathrm{e}^{-\gamma bv}\left[Cv^{\gamma}-\sum_{n=0}^{\infty}\frac{(\gamma bv)^{n+1}}{n!(n-\gamma+1)}\right]. \tag{27}$$

We can also carry out the sum in Eq. 27 in the limit of $\gamma\gg 1$ following Bender and Orzag [1]. By comparing the ratio of the $n-1$'th and the $n$'th term, we see that the elements of the sum have a maximum when $n\simeq\gamma bv$. For $\gamma\gg 1$, the sum will be dominated by terms with $n$ near $\gamma bv$. We therefore let $n=\gamma bv+s$ for some $s$, then $n!$ can be shown to be approximately [1]

$$n!\simeq(\gamma bv)^n\mathrm{e}^{-\gamma bv}\mathrm{e}^{\frac{s^2}{2\gamma bv}}\sqrt{2\pi\gamma bv} \tag{28}$$

using Stirling's approximation. Consequently, by approximating the sum as an integral and extending the range of the integral to $-\infty$,

$$\begin{aligned}
\sum_{n=0}^{\infty} \frac{(\gamma bv)^{n+1}}{n!(n-\gamma+1)} &\simeq \int_{-\infty}^{\infty} ds\, \frac{\mathrm{e}^{-\frac{s^2}{2\gamma bv}}}{\sqrt{2\pi\gamma bv}} \cdot \frac{\gamma bv \mathrm{e}^{\gamma bv}}{\gamma(bv-1)+s+1} \\
&= \int_{-\infty}^{\infty} ds\, \frac{\mathrm{e}^{-\frac{s^2}{2\gamma bv}}}{\sqrt{2\pi\gamma bv}} \cdot . \frac{bv\mathrm{e}^{\gamma bv}}{bv-1}\left[1+\gamma^{-1}\left(\frac{s+1}{bv-1}\right)\right]^{-1} \\
&= \frac{bv\mathrm{e}^{\gamma bv}}{bv-1}\int_{-\infty}^{\infty} ds\, \frac{\mathrm{e}^{-\frac{s^2}{2\gamma bv}}}{\sqrt{2\pi\gamma bv}} + O\left(\gamma^{-1}\right) \\
&\simeq \frac{bv\mathrm{e}^{\gamma bv}}{bv-1} \qquad (29)
\end{aligned}$$

to the lowest order in $\gamma$. From Eq. 27, $u$ satisfies

$$u(v) \simeq C\mathrm{e}^{-\gamma bv} v^{\gamma} + \frac{bv}{1-bv} \qquad (30)$$

when $\gamma \gg 1$. We evaluate $C$ using $u = u_0$ when $v = v_0$ giving

$$\begin{aligned}
u &\simeq \left(u_0 - \frac{bv_0}{1-bv_0}\right)\mathrm{e}^{-\gamma b(v-v_0)}\left(\frac{v}{v_0}\right)^{\gamma} + \frac{bv}{1-bv} \\
&\simeq \frac{bv}{1-bv}. \qquad (31)
\end{aligned}$$

when $\gamma \gg 1$ because $v = v_0\mathrm{e}^{\tau} > v_0$ from Eq. 24.

## Finding the generating function

Using Eq. 31, Eq. 24 becomes

$$\frac{dF}{dv} = \frac{ab}{1-bv}F \qquad (32)$$

or, on integrating,

$$\log\frac{F(v)}{F(v_0)} = -a\log\left(\frac{1-bv}{1-bv_0}\right) \qquad (33)$$

because $F(v_0) = F(\tau = 0)$. If initially we have $k$ proteins then

$$F(v_0) = \sum P_n(\tau=0)z^n = \sum \delta_{n,k} z^n = z^k = (1+v_0)^k. \qquad (34)$$

For our approximation, Eq. 31, to be valid, enough time must have passed for mRNA levels to have reached steady-state. Strictly, this initial condition is only valid for non-zero $\tau$ of the order of $d_1/d_0 = \gamma^{-1}$. Finally, inserting Eq. 34 into Eq. 33 gives

$$F(z,\tau) = \left[\frac{1-b(z-1)\mathrm{e}^{-\tau}}{1-bz+b}\right]^{a}\left[1+(z-1)\mathrm{e}^{-\tau}\right]^{k} \qquad (35)$$

because $v_0 = (z-1)\mathrm{e}^{-\tau}$. When $k = 0$, Eq. 35 becomes Eq. 7.

## Deriving the probability distribution for proteins

We can find $P_n(\tau)$, the probability of having $n$ proteins at time $\tau$ given initially zero proteins, by differentiating Eq. 35 when $k = 0$. By definition, $P_n$ satisfies $P_n = \frac{1}{n!}\frac{\partial^n}{\partial z^n}F(z,\tau)\Big|_{z=0}$. By writing

$$F(z,\tau) = \left(\frac{1+b\mathrm{e}^{-\tau}}{1+b}\right)^a \cdot \frac{\left[1-\frac{b}{1+b}z\right]^{-a}}{\left[1-\frac{b}{\mathrm{e}^{\tau}+b}z\right]^{-a}}, \tag{36}$$

we can make use of the identities

$$\frac{\partial^n}{\partial z^n}[1-qz]^{-a}\Big|_{z=0} = \frac{\Gamma(a+n)}{\Gamma(a)}q^n \tag{37}$$

and

$$\frac{\partial^n}{\partial z^n}\frac{x(z)}{y(z)} = n!\sum_{k=0}^{n}\frac{\partial^{n-k}}{\partial z^{n-k}}x(z)\cdot\sum_{j=0}^{k}\frac{(-1)^j(k+1)y(z)^{-j-1}}{(j+1)!(n-k)!(k-j)!}\frac{\partial^k}{\partial z^k}y(z)^j \tag{38}$$

which is given at Wolfram Research (`functions.wolfram.com/GeneralIdentities/9`).

Interpreting $x(z)$ as the numerator of the quotient in Eq. 36 and $y(z)$ as its denominator, we find

$$\begin{aligned} P_n(\tau) &= \left(\frac{1+b\mathrm{e}^{-\tau}}{1+b}\right)^a \sum_{k=0}^{n}\frac{\Gamma(a+n-k)}{\Gamma(a)}\left(\frac{b}{1+b}\right)^{n-k} \\ &\quad\times\sum_{j=0}^{k}\frac{(-1)^j(k+1)}{(j+1)!(n-k)!(k-j)!}\cdot\frac{\Gamma(aj+k)}{\Gamma(aj)}\cdot\left(\frac{b}{\mathrm{e}^{\tau}+b}\right)^k \end{aligned} \tag{39}$$

where we can use

$$\sum_{j=1}^{k}\frac{(-1)^j\Gamma(aj+k)}{\Gamma(aj)(j+1)!(k-j)!} = \frac{(-1)^k\Gamma(a+1)}{\Gamma(a-k+1)(k+1)!} \tag{40}$$

to simplify further. Eq. 40 can be verified by directly expanding the sum. Consequently,

$$P_n(\tau) = \left(\frac{b}{1+b}\right)^n\left(\frac{1+b\mathrm{e}^{-\tau}}{1+b}\right)^a\sum_{k=0}^{n}\frac{(-1)^k}{k!}\frac{\Gamma(a-k+n)}{\Gamma(n-k+1)\Gamma(a-k+1)}\left(\frac{1+b}{\mathrm{e}^{\tau}+b}\right)^k. \tag{41}$$

The hypergeometric function ${}_2F_1(a,b,c;z)$ obeys

$${}_2F_1(-n,b,c;z) = \sum_{k=0}^{n}(-1)^k\frac{\Gamma(n+1)}{\Gamma(n-k+1)}\frac{(b)_k}{(c)_k}\frac{z^k}{k!} \tag{42}$$

when $a$ is a negative integer and where $(b)_k$ and $(c)_k$ are Pochhammer symbols [2]. From their definition, $(a)_k = \Gamma(a+k)/\Gamma(a)$, the Pochhammer symbols satisfy

$$\Gamma(a+1) = (-1)^k(-a)_k\Gamma(a-k+1). \tag{43}$$

Writing $\Gamma(a-k+n) = \Gamma(a+n-1-k+1)$ and using Eq. 42 and Eq. 43, we find that

$$P_n(\tau) = \frac{1}{n!}\left(\frac{b}{1+b}\right)^n\left(\frac{1+b\mathrm{e}^{-\tau}}{1+b}\right)^a\frac{\Gamma(a+n)}{\Gamma(a)}\,{}_2F_1\left(-n,-a,1-a-n;\frac{1+b}{\mathrm{e}^{\tau}+b}\right) \tag{44}$$

which is valid for $\gamma \gg 1$, $\tau > \gamma^{-1}$, and $a$ and $b$ finite.

## Deriving the 'propagator' probability

By differentiating Eq. 35 for non-zero $k$, we can express the 'propagator' probability, $P_{n|k}(\tau)$, in terms of Eq. 44. From the definition of $P_n(\tau)$, Eq. 35 can be written as

$$F(z,\tau) = \left[\sum_{n=0}^{\infty} P_n(\tau) z^n\right] \left[1 - \mathrm{e}^{-\tau} + z\mathrm{e}^{-\tau}\right]^k \tag{45}$$

or

$$F(z,\tau) = \sum_{n=0}^{\infty} P_n(\tau) z^n \sum_{r=0}^{k} \binom{k}{r} \left(1 - \mathrm{e}^{-\tau}\right)^{k-r} \left(z\mathrm{e}^{-\tau}\right)^r \tag{46}$$

using the binomial theorem. From the coefficients of the powers of $z$, we find

$$P_{n|k}(\tau) = \sum_{r=0}^{k} \binom{k}{r} P_{n-r}(\tau) \left(1 - \mathrm{e}^{-\tau}\right)^{k-r} \mathrm{e}^{-r\tau} \tag{47}$$

because $F(z,\tau) = \sum_n P_{n|k}(\tau) z^n$ and remembering that $P_n(\tau) = 0$ if $n < 0$.

## Finding the probability distribution for the first passage time

With $P_n(\tau)$ and $P_{n|k}(\tau)$, we can find the distribution for the first time the number of proteins reaches a threshold $N$. We define this distribution to be $f_N(\tau)$. It obeys a renewal equation [3]

$$P_N(\tau) = \int_0^{\tau} d\tau' f_N(\tau') P_{N|N}(\tau - \tau'). \tag{48}$$

The probability of having $N$ proteins at time $\tau$ is equal to the sum of the probability of first reaching $N$ proteins at $\tau'$ and then returning to $N$ proteins at a time $\tau - \tau'$ later for all times $\tau'$ less than $\tau$. We have assumed that the initial number of proteins is zero, but this assumption is not necessary.

Eq. 48 is a Volterra integral equation of the first kind and can be straightforwardly solved numerically [4]. If $N > 0$ then $f_N(0) = 0$ and $P_{N|N}(0) = 1$ by definition. Consequently, by discretizing and letting $\tau_i = i\epsilon$ for integer $i$ and small $\epsilon$, we can write the integral in Eq. 48 as a trapezium rule:

$$\int_0^{\tau_i} d\tau' f_N(\tau') P_{N|N}(\tau_i - \tau') \simeq \epsilon \left[\frac{1}{2} f_N(\tau_i) + \sum_{j=1}^{i-1} P_{N|N}(\tau_i - \tau_j) f_N(\tau_j)\right]. \tag{49}$$

Inserting Eq. 49 into Eq. 48 gives a series of equations for $f_N(\tau_i)$ which we solve iteratively:

$$f_N(\tau_1) = \frac{2P_N(\tau_1)}{\epsilon} \tag{50}$$

$$f_N(\tau_i) = 2\left[\frac{P_N(\tau_i)}{\epsilon} - \sum_{j=1}^{i-1} P_{N|N}(\tau_{i-j}) f_N(\tau_j)\right]. \tag{51}$$

We implement Eqs. 50 and 51 in Matlab (The Mathworks, Natick, Massachusetts). Our code is available at `www.cnd.mcgill.ca/~swain`.

We use

$$\langle n(\tau_1) n(\tau_2)\rangle = \sum_{n,n'} n n' P_{n|n'}(\tau_2 - \tau_1) P_{n'}(\tau_1) \tag{52}$$

to find the auto-correlation function. We evaluate the sum in Eq. 52 numerically, cutting off the sums when $n$ is many times the mean steady-state value: $\langle n\rangle = ab$.

## High $\gamma$ implies bursts of protein synthesis

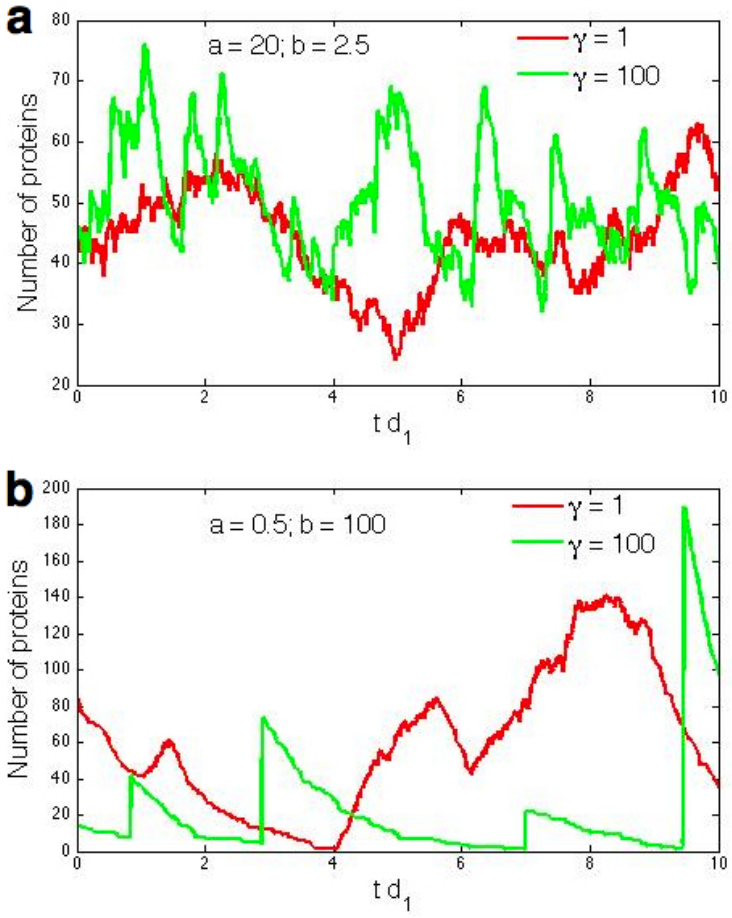


Figure 5: As $\gamma$ increases, protein synthesis occurs in bursts. Time courses of protein numbers from simulations of the two-stage model of of Fig. 1. When $\gamma$ is increased to 100 from 1, we see steep bursts of synthesis: short-lived mRNAs are only able to be occasionally translated before being degraded. The protein degradation rate is $d_1 = 0.0005\text{s}^{-1}$. **a** $a = 20$ and $b = 2.5$. **b** $a = 0.5$ and $b = 100$. Both examples have a mean protein number of 50.

## Solving the master equation for bursts of protein synthesis

When $\gamma \gg 1$, the distribution for protein numbers can also be derived by only considering $P_n(\tau)$, the probability of having $n$ proteins at time $\tau$, if this probability obeys a master equation where proteins are synthesized in bursts. We let the size $r$ of a burst obey a geometric distribution,

$$P(r) = \left(\frac{b}{1+b}\right)^r \left(1 - \frac{b}{1+b}\right). \tag{53}$$

The corresponding master equation is

$$\frac{\partial P_n}{\partial \tau} = a\left[\left(1 - \frac{b}{1+b}\right)\sum_{r=0}^{n}\left(\frac{b}{1+b}\right)^r P_{n-r} - P_n\right] + (n+1)P_{n+1} - nP_n \tag{54}$$

which can be converted into an equation for the generating function, $F(z) = \sum_n z^n P_n(\tau)$.

The generating function obeys

$$\frac{\partial F}{\partial \tau} = (1-z)\frac{\partial F}{\partial z} - aF + a\left(1 - \frac{b}{1+b}\right)\sum_{n=0}^{\infty}\sum_{r=0}^{n} z^n \left(\frac{b}{1+b}\right)^r P_{n-r} \tag{55}$$

where we need to evaluate the sums over $n$ and $r$. Relabelling and resuming

$$\begin{aligned}
\sum_{n=0}^{\infty}\sum_{r=0}^{n} z^n \left(\frac{b}{1+b}\right)^r P_{n-r} &= \sum_{n=0}^{\infty}\sum_{k=0}^{n} z^n \left(\frac{b}{1+b}\right)^{n-k} P_k \\
&= \sum_{k=0}^{\infty}\left(\frac{b}{1+b}\right)^{-k} P_k \sum_{n=k}^{\infty}\left(\frac{bz}{1+b}\right)^n \\
&= \sum_{k=0}^{\infty} \frac{P_k \left(\frac{bz}{1+b}\right)^k}{\left(1-\frac{bz}{1+b}\right)\left(\frac{b}{1+b}\right)^k} \\
&= \frac{F(z)}{1-\frac{bz}{1+b}}
\end{aligned} \tag{56}$$

where we use the definition of the generating function. Consequently, Eq. 55 becomes

$$\frac{\partial F}{\partial \tau} = (1-z)\frac{\partial F}{\partial z} + \left(\frac{1-\frac{b}{1+b}}{1-\frac{bz}{1+b}} - 1\right) aF \tag{57}$$

or

$$\frac{1}{v}\frac{\partial F}{\partial \tau} + \frac{\partial F}{\partial v} = \frac{ab}{1-bv}F \tag{58}$$

with $v = z - 1$. This partial differential equation is Eq. 23 when $\gamma \gg 1$ and Eq. 31 holds.

## Derivation of the gamma distribution for protein numbers

We can derive the gamma distribution for protein numbers found by Friedman *et al.* [5] when $n$ is large. If $P(n|a,b)$ is the negative binomial distribution and $\Gamma(n|a,b)$ is the gamma distribution, then

$$P(n|a,b) = \int_0^{\infty} d\lambda \, \frac{\mathrm{e}^{-\lambda}\lambda^n}{n!}\Gamma(\lambda|a,b) \tag{59}$$

which is a general relation between the negative binomial and gamma distributions. It can be verified by evaluating the integral using the definition of a gamma function [2]. If we approximate the Poisson distribution by a normal distribution and write $z = \lambda - n$, Eq. 59 becomes

$$\begin{aligned}
P(n|a,b) &\simeq \int_{-\infty}^{\infty} dz \, \frac{\mathrm{e}^{-\frac{z^2}{2(z+n)}}}{\sqrt{2\pi(z+n)}}\Gamma(z+n|a,b) \\
&= \int_{-\infty}^{\infty} dz \frac{\mathrm{e}^{-\frac{z^2}{2n}\left(1+\frac{z}{n}\right)^{-1}}}{\sqrt{2\pi n}} \cdot \left(1+\frac{z}{n}\right)^{-\frac{1}{2}} \Gamma\Big(n\Big[1+\frac{z}{n}\Big]\Big| a, b\Big).
\end{aligned} \tag{60}$$

We note that only values of $z$ close to zero contribute to the integral when $n \gg 1$ because $z = 0$ is the minimum of the exponent in the integrand. Then $n \gg 1$ implies $z/n \ll 1$, and so

$$\begin{aligned} P(n|a,b) &\simeq \int_{-\infty}^{\infty} dz \, \frac{\mathrm{e}^{-\frac{z^2}{2n}}}{\sqrt{2\pi n}} \Gamma(n|a,b) \\ &= \Gamma(n|a,b) \end{aligned} \tag{61}$$

for large $n$, as expected [5].

# Derivation of the protein distribution for a three-stage model of gene expression

We can use the same approximation of large $\gamma$ to find the protein distribution for the three-stage model. Let $P^{(0)}_{m,n}$ be the probability of having $m$ mRNAs and $n$ proteins when the DNA is inactive and $P^{(1)}_{m,n}$ be the probability of having $m$ mRNAs and $n$ proteins when the DNA is active. The master equation consists of two coupled equations:

$$\begin{aligned} \frac{\partial P^{(0)}_{n,m}}{\partial \tau} &= \kappa_1 P^{(1)}_{m,n} - \kappa_0 P^{(0)}_{m,n} + (n+1) P^{(0)}_{m,n+1} - n P^{(0)}_{m,n} \\ &\quad + \gamma \left[ (m+1) P^{(0)}_{m+1,n} - m P^{(0)}_{m,n} + bm \left( P^{(0)}_{m,n-1} - P^{(0)}_{m,n} \right) \right] \end{aligned} \tag{62}$$

$$\begin{aligned} \frac{\partial P^{(1)}_{n,m}}{\partial \tau} &= -\kappa_1 P^{(1)}_{m,n} + \kappa_0 P^{(0)}_{m,n} + (n+1) P^{(1)}_{m,n+1} - n P^{(1)}_{m,n} + a \left( P^{(1)}_{m-1,n} - P^{(1)}_{m,n} \right) \\ &\quad + \gamma \left[ (m+1) P^{(1)}_{m+1,n} - m P^{(1)}_{m,n} + bm \left( P^{(1)}_{m,n-1} - P^{(1)}_{m,n} \right) \right] \end{aligned} \tag{63}$$

where $\kappa_0 = k_0/d_1$ and $\kappa_1 = k_1/d_1$. By defining two generating functions

$$f^{(0)}(z',z) = \textstyle\sum_{m,n} (z')^m z^n P^{(0)}_{m,n} \quad ; \quad f^{(1)}(z',z) = \textstyle\sum_{m,n} (z')^m z^n P^{(1)}_{m,n}, \tag{64}$$

these equations become

$$\frac{1}{v} \frac{\partial f^{(0)}}{\partial \tau} = \frac{1}{v} \left[ \kappa_1 f^{(1)} - \kappa_0 f^{(0)} \right] - \frac{\partial f^{(0)}}{\partial v} + \gamma \left[ b(1+u) - \frac{u}{v} \right] \frac{\partial f^{(0)}}{\partial u} \tag{65}$$

$$\frac{1}{v} \frac{\partial f^{(1)}}{\partial \tau} = \frac{1}{v} \left[ -\kappa_1 f^{(1)} + \kappa_0 f^{(0)} \right] - \frac{\partial f^{(1)}}{\partial v} + a \frac{u}{v} f^{(1)} + \gamma \left[ b(1+u) - \frac{u}{v} \right] \frac{\partial f^{(1)}}{\partial u} \tag{66}$$

with $u = z' - 1$ and $v = z - 1$.

At steady-state $\frac{\partial f^{(0)}}{\partial \tau} = \frac{\partial f^{(1)}}{\partial \tau} = 0$, and we find using the method of characteristics that

$$\begin{array}{lll} \frac{dv}{dr} = 1 & ; & \frac{du}{dr} = -\gamma \left[ b(1+u) - \frac{u}{v} \right] \\ \frac{df^{(0)}}{dr} = \frac{1}{v} \left[ \kappa_1 f^{(1)} - \kappa_0 f^{(0)} \right] & ; & \frac{df^{(1)}}{dr} = \frac{1}{v} \left[ -\kappa_1 f^{(1)} + \kappa_0 f^{(0)} \right] + a \frac{u}{v} f^{(1)} \end{array} \tag{67}$$

where $r$ measures the distance along a characteristic. Both $u$ and $v$ obey Eq. 24 again. Consequently, $v = r$ and $u \simeq \frac{bv}{1-bv}$ from Eq. 31 when $\gamma \gg 1$. From Eq. 67, we therefore obtain the two coupled differential equations:

$$v \frac{df^{(0)}}{\partial v} = \kappa_1 f^{(1)} - \kappa_0 f^{(0)} \tag{68}$$

$$v \frac{df^{(1)}}{\partial v} = -\kappa_1 f^{(1)} + \kappa_0 f^{(0)} + \frac{abv}{1-bv} f^{(1)}. \tag{69}$$

Following Hornos *et al.* [6], Eqs. 68 and 69 can be reduced to one differential equation for $f^{(0)}(v)$ by solving Eq. 68 for $f^{(1)}$ in terms of $f^{(0)}$ and its derivative, and inserting the result into Eq. 69. This equation becomes a second-order differential equation:

$$v(bv-1)\frac{df^{(0)}}{dv^2}+\Big[(\kappa_0+\kappa_1)(bv-1)+bv(1+a)-1\Big]\frac{df^{(0)}}{dv}+ab\kappa_0 f^{(0)}=0. \tag{70}$$

Eq. 70 has solution

$$f^{(0)}(v)=C\,{}_2F_1(\alpha,\beta,1-\kappa_0-\kappa_1;bv) \tag{71}$$

where ${}_2F_1(a,b,c;z)$ is a hypergeometric function,

$$\alpha = \frac{1}{2}\left(a+\kappa_0+\kappa_1+\sqrt{(a+\kappa_0+\kappa_1)^2-4a\kappa_0}\right) \tag{72}$$

$$\beta = \frac{1}{2}\left(a+\kappa_0+\kappa_1-\sqrt{(a+\kappa_0+\kappa_1)^2-4a\kappa_0}\right), \tag{73}$$

and $C$ is a constant of integration.

We can find the generating function for protein numbers, $F(z)=f^{(0)}(z)+f^{(1)}(z)$, by using our solution for $f^{(0)}$ and Eq. 68 to find $f^{(1)}$. Determining the constant of integration $C$ from $F(1)=1$ and using the relation $c(c+1)\,{}_2F_1(a,b,c;z)=c(c+1)\,{}_2F_1(a,b,c+1;z)+abz\,{}_2F_1(a+1,b+1,c+2;z)$, we find that

$$F(z)=\,{}_2F_1\Big(\alpha,\beta,\kappa_0+\kappa_1;b(z-1)\Big), \tag{74}$$

replacing $v$ by $z-1$.

Expanding the generating function around $z=0$ determines the probabilities $P_n$. Using properties of the $n$-th derivatives with respect to $z$ of the hypergeometric function, ${}_2F_1^{(n)}(a,b,c;z)$, we can write

$$\begin{aligned} F(z) &= \sum_{n=0}^{\infty}{}_2F_1^{(n)}\Big(\alpha,\beta,\kappa_0+\kappa_1;-b\Big)\frac{b^n}{n!}z^n \\ &= \sum_{n=0}^{\infty}\frac{\Gamma(\alpha+n)\Gamma(\beta+n)\Gamma(\kappa_0+\kappa_1)b^n}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\kappa_0+\kappa_1+n)n!}\,{}_2F_1\Big(\alpha+n,\beta+n,\kappa_0+\kappa_1+n;-b\Big)z^n \end{aligned} \tag{75}$$

and $P_n$ can be found from the definition of $F(z)$: $F(z)=\sum_n P_n z^n$. With the linear transformation formulae for hypergeometric functions [2], we write $P_n$ as

$$\begin{aligned} P_n &= \frac{\Gamma(\alpha+n)\Gamma(\beta+n)\Gamma(\kappa_0+\kappa_1)}{\Gamma(n+1)\Gamma(\alpha)\Gamma(\beta)\Gamma(\kappa_0+\kappa_1+n)}\left(\frac{b}{1+b}\right)^n\left(1-\frac{b}{1+b}\right)^{\alpha} \\ &\quad\times{}_2F_1\left(\alpha+n,\kappa_0+\kappa_1-\beta,\kappa_0+\kappa_1+n;\frac{b}{1+b}\right). \end{aligned} \tag{76}$$

# The exact mRNA distributions

For completeness, we include the mRNA distributions for the two-stage and three-stage models. With initially zero mRNAs, the two-stage model has a Poisson distribution:

$$P_m(t)=\mathrm{e}^{-\langle m(t)\rangle}\frac{\langle m(t)\rangle^m}{m!} \tag{77}$$

where $\langle m(t)\rangle = m_s\left(1-\mathrm{e}^{-d_0 t}\right)$ and $m_s = v_0/d_0$ is the steady-state number of mRNAs. The propagator probability satisfies

$$P_{m|k}(t) = \sum_{r=0}^{k} \binom{k}{r} P_{m-r}(t)\left(1-\mathrm{e}^{-d_0 t}\right)^{k-r} \mathrm{e}^{-r d_0 t} \tag{78}$$

with $P_m(t) = 0$ if $m < 0$.

The steady-state distribution of mRNA for the three-stage model was first derived by Peccoud and Ycart, although they did not recognize it as such [7], and also by Raj *et al.* [8]. The exact probability of having $m$ RNAs at steady-state is

$$P_m = \frac{m_s^m \mathrm{e}^{-m_s}}{m!} \cdot \frac{\Gamma(\zeta_0+m)\Gamma(\zeta_0+\zeta_1)}{\Gamma(\zeta_0+\zeta_1+m)\Gamma(\zeta_0)}\, {}_1F_1(\zeta_1, \zeta_0+\zeta_1+m; m_s) \tag{79}$$

where $m_s = v_0/d_0$, $\zeta_0 = k_0/d_0$, and $\zeta_1 = k_1/d_0$, and ${}_1F_1(a,b;z)$ is the confluent hypergeometric function of the first kind [2]. Eq. 79 like Eq. 18 can be bimodal. For $\zeta_1 = k_1/d_0 \gg 1$, Eq. 79 tends to a negative binomial distribution [8], because then mRNA synthesis is more burst-like. The distribution becomes Poisson when $k_1$ is zero, and the three-stage model reduces to the two-stage model.

# References


[1] Bender, C. M. and Orzag, S. A. *Advanced Mathematical Methods for Scientists and Engineers.* Springer, New York, New York, (1999).

[2] Abramowitz, M. and Stegun, I. A. *Pocketbook of Mathematical Functions.* Harri Deutsch Publishing, Frankfurt am Main, (1984).

[3] van Kampen, N. G. *Stochastic Processes in Physics and Chemistry.* Elsevier, New York, New York, (1990).

[4] Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. *Numerical recipes in C++.* Cambridge University Press, New York, New York, (2002).

[5] Friedman, N., Cai, L., and Xie, X. S. Linking stochastic dynamics to population distribution: an analytical framework of gene expression. *Phys. Rev. Lett.* **97**, 16830 (2006).

[6] Hornos, J. E., Schultz, D., Innocentini, G. C., Wang, J., Walczak, A. M., Onuchic, J. N., and Wolynes, P. G. Self-regulating gene: an exact solution. *Phys. Rev. E.* **72**, 051907 (2005).

[7] Peccoud, J. and Ycart, B. Markovian modeling of gene-product synthesis. *Theor. Popul. Biol.* **48**, 222–234 (1995).

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