Special Function Methods for Bursty Models of Transcription Supplementary Note

Gennady Gorin, Lior Pachter

Contents

Burst generating functions and their expansions

Let ϕ be the logarithm of the probability generating function (PDF) G. To determine ϕ, it is necessary to integrate $\frac{d\phi}{dt} = k_i(M(U(s)) - 1)$, where $M(U(s))$ is the factorial-moment generating function (FMGF) of the burst distribution. This function is represented as $M(U) = F(1 + U)$ where $F(x) = \sum_{\rho=0}^{\infty} P(B = \rho)x^\rho$ is the PGF of the burst distribution.
In what follows we use μ to denote the mean burst size for a single-parameter burst distribution.

Geometric distribution, $\mu = b$

As discussed by Singh and Bokes [1], the geometric distribution with mean burst size b has the probability mass function (PMF) $P(B = \rho) = p(1 - p)^\rho$, where $p = \frac{1}{1+b}$ and $\rho = 0, 1, 2, ...$ The resulting PGF is $F(x) = E[x^\rho] = \sum_{\rho=0}^{\infty} P(B = \rho)x^\rho = \sum_{i=0}^{\infty} p(1 - p)^\rho x^\rho = p\sum_{i=\rho}^{\infty}[(1 - p)x]^\rho = p\frac{1}{1-(1-p)x}$. This is exact for $|x| \leq 1$ and extends to $x \in \mathbb{C}$ by analytical continuation. Using the definition of p, $F(x) = \frac{1}{1+b}\frac{1}{1-\frac{b}{1+b}x} = \frac{1}{1+b-bx}$. The transformed PGF $M(u) = F(1 + u) = \frac{1}{1+b-b(1+u)} = \frac{1}{1-bu}$. Finally, $M(u) - 1 = \frac{1}{1-bu} - 1 = \frac{bu}{1-bu}$, which recapitulates previous work.
Defining $X = bu$, we have that $M(u) - 1 = \frac{X}{1-X}$. This expression has the well-known Laurent expansion $-\sum_{i=0}^{\infty}\frac{1}{X^i}$ for all $|X| > 1$. Since b is real-valued, $M(u) - 1 = \sum_{i=0}^{\infty}\Omega_i u^{-i}$, where $\Omega_i = -b^{-i}$. This expansion can be truncated at order N_L, yielding $M(u) - 1 \approx \sum_{i=0}^{N_L}\Omega_i u^{-i}$.
Within the region $|X+1| < 2$, the function $\frac{X}{1-X}$ has the Taylor expansion $\sum_{i=1}^{\infty}\frac{(1+X)^i}{2^{i+1}} - \frac{1}{2}$. Expanding the binomial yields $\sum_{i=1} 2^{-i-1}\sum_{j=0}^{i}\binom{i}{j}X^i - \frac{1}{2} = \sum_{i=1}\frac{1}{2^{i+1}}\sum_{j=1}^{i}\binom{i}{j}X^j = \sum_{i=1}\frac{1}{2^{i+1}}\sum_{j=1}^{i}\binom{i}{j}(bu)^j$.
This expansion can be truncated at order N_T, yielding $M(u) - 1 \approx \sum_{i=1}^{N_T}\frac{b^i}{2}(\sum_{j=i}^{N_T}\frac{1}{2^j}\binom{j}{i})u^i = \sum_{i=1}^{N_T}\Omega_i u^i$, where $\Omega_i = \frac{b^i}{2}\sum_{j=i}^{N_T}\frac{1}{2^j}\binom{j}{i}$.
Equivalently, $\Omega_i = b^i[1 - 2^{-N_T-2}\binom{N_T+1}{i}{}_2F_1(1, N_T + 2; -i + N_T + 2; \frac{1}{2})]$.

Shifted geometric distribution, $\mu = b$

A shifted geometric distribution with mean burst size b has the PMF $P(B = \rho) = p(1 - p)^{\rho-1}$, where $p = \frac{1}{b}$ and $\rho = 1, 2, 3, ...$ The resulting PGF over the relevant support is $F(x) = E[x^\rho] = \sum_{\rho=1}^{\infty} P(B = \rho)x^\rho = \sum_{\rho=1}^{\infty} p(1-p)^{\rho-1}x^\rho = p\sum_{\rho=0}^{\infty}(1-p)^\rho x^{\rho+1} = px\sum_{\rho=0}^{\infty}[(1-p)x]^\rho = px\frac{1}{1-(1-p)x}$.
Using the definition of p, $F(x) = \frac{x}{b}\frac{1}{1-\frac{b-1}{b}x} = \frac{x}{b-(b-1)x}$. The transformed PGF $M(u) = F(1 + u) = \frac{1+u}{b-(b-1)(1+u)} = \frac{1+u}{b-b+1-bu+u} = \frac{1+u}{1+(1-b)u}$. Finally, $M(u) - 1 = \frac{1+u-1-(1-b)u}{1+(1-b)u} = \frac{bu}{1+(1-b)u}$.
Within the region $|u| > \frac{1}{|b-1|}$, the function $M(u)-1$ has the Laurent expansion $-\frac{b}{b-1}\sum_{i=0}^{\infty}\frac{1}{(b-1)^i u^i} = \sum_{i=0}^{\infty}\Omega_i u^{-i}$, where $\Omega_i = -\frac{b}{(b-1)^{i+1}}$. This expansion can be truncated at order N_L, yielding $M(u) - 1 \approx \sum_{i=0}^{N_L}\Omega_i u^{-i}$.

Defining the intermediate variable $a = \frac{1}{b-1}$ and noting $a+1 = \frac{b}{b-1}$, $M(u) - 1 = \frac{b}{b-1}\frac{(b-1)u}{1+(1-b)u} = -\frac{b}{b-1}\frac{(1-b)u}{1+(1-b)u} = -\frac{b}{b-1}\frac{u}{u-\frac{1}{1-b}} = -\frac{b}{b-1}\frac{u}{u-a}$. The Taylor expansion of $\frac{u}{u-a}$ about $-a$, within the domain $|u+a| < 2|a|$, is $-\sum_{i=1}^{\infty}\frac{1}{2^{i+1}a^i}(a+u)^i + \frac{1}{2} = -\sum_{i=1}^{\infty}\frac{1}{2^{i+1}a^i}\sum_{j=0}^{i}\binom{i}{j}u^j a^{i-j} + \frac{1}{2}$
$= -\sum_{i=1}^{\infty}\frac{1}{2^{i+1}}\sum_{j=0}^{i}\binom{i}{j}u^j a^{-j} - \frac{1}{2} = -\sum_{i=1}^{\infty}\frac{1}{2^{i+1}}\sum_{j=1}^{i}\binom{i}{j}u^j a^{-j}$. Therefore, $M(u) - 1$
$= \frac{b}{2(b-1)}\sum_{i=1}^{\infty}\frac{1}{2^i}\sum_{j=1}^{i}\binom{i}{j}u^j(b-1)^j$. This expansion can be truncated at order N_T, yielding
$M(u) - 1 \approx \sum_{i=1}^{N_T}\Omega_i u^i$, where $\Omega_i = \frac{b}{2}(b-1)^{i-1}\sum_{j=i}^{N_T}\frac{1}{2^j}\binom{j}{i}$.
Equivalently, $\Omega_i = b(b-1)^{i-1}[1 - 2^{-N_T-2}\binom{N_T+1}{i}{}_2F_1(1, N_T+2; -i+N_T+2, \frac{1}{2})]$.

Degenerate (b-step)

The degenerate burst distribution that yields b pre-mRNA products with every burst has PMF $P(B = \rho) = \delta_{\rho b}$, where δ_{ij} is the Kronecker delta. The resulting PGF is $F(x) = E[x^\rho] = \sum_{\rho=0}^{\infty}P(B=\rho)x^\rho = x^b$. Therefore, $M(u) - 1 = (1+u)^b - 1$.
Expanding the binomial yields the expression $M(u) - 1 = \sum_{i=0}^{b}\binom{b}{i}u^i - 1 = \sum_{i=1}^{b}\binom{b}{i}u^i = \sum_{i=1}^{b}\Omega_i u^i$, where $\Omega_i = \binom{b}{i}$.

Uniform on $[a, b]$

The uniform distribution on $[a, b]$ has PMF $P(B = \rho) = \frac{1}{n}I(\rho \in [a, b])$, where $I(\cdot)$ is the indicator function and $n = b - a + 1$. The resulting PGF over the relevant support is $F(x) = E[x^\rho] = \frac{1}{n}\sum_{\rho=a}^{b}x^\rho$. Therefore, $M(u) - 1 = \frac{1}{n}\sum_{i=a}^{b}(1+u)^i - 1$.
Expanding the binomial yields $M(u) - 1 = \frac{1}{n}\sum_{i=a}^{b}\sum_{j=0}^{i}\binom{i}{j}u^j - 1 = \frac{1}{n}\sum_{i=a}^{b}\sum_{j=1}^{i}\binom{i}{j}u^j$. Assuming $a > 0$, reversing the order of summation yields $M(u) - 1 = \sum_{j=1}^{b}\frac{W_1(j) - W_2(j)}{n(j+1)}u^j$, where $W_1(j) = \frac{\Gamma(b+2)}{\Gamma(b-j+1)}$ for all j and $W_2(j) = 0$ for $j \in [a, b]$ and $\frac{\Gamma(a+1)}{\Gamma(a-j)}$ otherwise. Therefore, $M(u) - 1 = \sum_{i=1}^{b}\Omega_i u^i$, where $\Omega_i = \frac{W_1(i) - W_2(i)}{n(i+1)}$. Equivalently, $\Omega_i = \frac{1}{n}[\binom{b+1}{i+1} - \binom{a}{i+1}]$.

Special function solutions to $\int U^i ds$

As explored in the section "**Burst generating functions and their expansions**," the FCGFs of the burst models explored here can be represented in the common form $M(u) - 1 \approx \sum_i \Omega_i u^i$, where $i \in \mathbb{Z}$. Given $U(s)$, a characteristic solution representing the dynamics downstream of the gene locus, the integral $\frac{\phi}{k_i} = \int[M(U(s)) - 1]ds$ can be approximated as $\sum_i \Omega_i \int U(s)^i ds$ wherever the expansion holds. The determination of the appropriate domains is treated in the section "**Domain Decomposition**."

Degenerate case: $U(s; u, v) = e^{-\gamma s}(u + \gamma v s)$

Taylor expansion

From standard identities [2], the Taylor expansion of order $i \in \mathbb{Z}$ yields $T_i(s; u, v) = \int e^{-i\gamma s}(u + \gamma v s)^i ds = -\frac{e^{iu/v}v^i}{\gamma i^{i+1}}\Gamma(i+1, \frac{i}{v}(u + \gamma v s))$, where $\Gamma(a, z) = \int_z^\infty t^{a-1}e^{-t}dt$ is the upper incomplete Gamma function [2]. The evaluation of $\Gamma(a, z)$ for arbitrary complex arguments is computationally nontrivial. However, for $n \in \mathbb{N}$, $\Gamma(i+1, z) = i!e^{-z}e_n(z)$, where $e_n(z) = \sum_{k=0}^{n}\frac{z^k}{k!}$.

Therefore, $T_i = -\frac{e^{iu/v}v^i}{\gamma i^{i+1}} i! e^{-\frac{i}{v}(u+\gamma vs)} \sum_{k=0}^{i} \frac{1}{k!}(\frac{i}{v}(u+\gamma vs))^k = -\frac{e^{-i\gamma s}v^i}{\gamma i^{i+1}} i! \sum_{k=0}^{i} \frac{i^k}{k!}(\frac{u}{v}+\gamma s)^k$. The resulting definite integral from s_1 to s_2 is simply $-\frac{v^i i!}{\gamma i^{i+1}} \sum_{k=0}^{i} \frac{i^k}{k!}[(\frac{u}{v}+\gamma s_2)^k e^{-i\gamma s_2} - (\frac{u}{v}+\gamma s_1)^k e^{-i\gamma s_1}]$, which is directly computable without the use of special function routines.
This expression for T_i is inappropriate for the degenerate case $v = 0$, corresponding to the marginal with respect to the pre-mRNA. In that case, the definite integral reduces to $-\frac{u^i}{\gamma i}[e^{-i\gamma s_2} - e^{-i\gamma s_1}]$.

Laurent expansion

For $k = -i$, where the order of Laurent expansion $k \in \mathbb{Z}$, the identity $L_k(s; u, v)$
$= \int e^{i\gamma s}(u+\gamma vs)^{-i} ds = -\frac{e^{-iu/v}v^{-i}}{\gamma(-i)^{1-i}}\Gamma(1-i, -\frac{i}{v}(u+\gamma vs))$ holds.
$L_k = \frac{e^{-iu/v}(-1)^i}{\gamma v^i i^{1-i}}\Gamma(1-i, -\frac{i}{v}(u+\gamma vs))$. The direct computation of $\Gamma(1-i, z)$ is computationally non-trivial, and finite power series expansions are unavailable for this purpose. However, the following identity holds:
$\Gamma(1-i, z) = z^{1-i}E_i(z)$, where $E_i(z)$ is the generalized exponential integral of order i. Further, $E_i(z) = \frac{(-z)^{i-1}}{(i-1)!}E_1(z) + \frac{e^{-z}}{(i-1)!}\sum_{k=0}^{i-2}(i-k-2)!(-z)^k$, where $E_1(z) = \int_z^\infty \frac{e^{-t}}{t}dt$ is the first-order complex exponential integral.
Therefore, $L_k = \frac{e^{-iu/v}(-1)^i}{\gamma v^i i^{1-i}}[-\frac{i}{v}(u+\gamma vs)]^{1-i}E_i(-\frac{i}{v}(u+\gamma vs))$
$= -\frac{e^{-iu/v}}{\gamma v}[u+\gamma vs]^{1-i}E_i(-\frac{i}{v}(u+\gamma vs))$. Using the relation for $E_i(z)$,
$E_i(-\frac{i}{v}(u+\gamma vs)) = \frac{1}{\Gamma(i)}[\frac{i}{v}(u+\gamma vs)]^{i-1}E_1(z) + \frac{1}{\Gamma(i)}\exp(\frac{i}{v}(u+\gamma vs))\sum_{k=0}^{i-2}(i-k-2)!(\frac{i}{v}(u+\gamma vs))^k$, where $z = -\frac{i}{v}(u+\gamma vs)$.
This yields $L_k = -\frac{e^{-iu/v}}{\gamma v\Gamma(i)}(\frac{i}{v})^{i-1}E_1(z) - \frac{e^{i\gamma s}}{\gamma v\Gamma(i)}\sum_{k=0}^{i-2}\Gamma(i-k-1)(\frac{i}{v})^k(u+\gamma vs)^{k+1-i}$. Finally, the definite integral $L_k(s_2; u, v) - L_k(s_1; u, v)$ is
$-\frac{e^{-iu/v}}{\gamma v\Gamma(i)}(\frac{i}{v})^{i-1}[E_1(z_2) - E_1(z_1)]$
$-\frac{1}{\gamma v\Gamma(i)}\sum_{k=0}^{i-2}\Gamma(i-k-1)(\frac{i}{v})^k[(u+\gamma vs_2)^{k+1-i}e^{i\gamma s_2} - (u+\gamma vs_1)^{k+1-i}e^{i\gamma s_1}]$, where $z_l = -\frac{i}{v}(u+\gamma vs_l)$.
Equivalently, this is
$-\frac{e^{-iu/v}}{\gamma v\Gamma(i)}(\frac{i}{v})^{i-1}[E_1(z_2) - E_1(z_1)]$
$-\frac{e^{i\gamma s_1}}{\gamma v\Gamma(i)}\sum_{k=0}^{i-2}\Gamma(i-k-1)(\frac{i}{v})^k[(u+\gamma vs_2)^{k+1-i}e^{i\gamma(s_2-s_1)} - (u+\gamma vs_1)^{k+1-i}]$.
As in the Taylor expansion, this expression is inappropriate for the degenerate case $v = 0$. In that case, the definite integral reduces to $\frac{1}{iu^i\gamma}(e^{i\gamma s_2} - e^{i\gamma s_1})$.

Non-degenerate case: $U(s; u, v) = vfe^{-\gamma s} + (u - vf)e^{-\beta s}$

As throughout, f is defined as $\frac{\beta}{\beta-\gamma}$.

Taylor expansion

For a Taylor expansion of order $i \in \mathbb{Z}$, $T_i(s; u, v) = \int[vfe^{-\gamma s} + (u-vf)e^{-\beta s}]^i ds$
$= \int(Ae^{-\gamma s} + Be^{-\beta s})^i ds$. Expanding the binomial term yields T_i
$= \int\sum_{j=0}^{i}\binom{i}{j}(Ae^{-\gamma s})^j(Be^{-\beta s})^{i-j}ds = \int\sum_{j=0}^{i}\binom{i}{j}A^jB^{i-j}e^{-j\gamma s}e^{-(i-j)\beta s}ds$
$= \int\sum_{j=0}^{i}\binom{i}{j}A^jB^{i-j}e^{[-j\gamma-(i-j)\beta]s}ds = \sum_{j=0}^{i}\binom{i}{j}A^jB^{i-j}\int e^{[-j\gamma-(i-j)\beta]s}ds$
$= \sum_{j=0}^{i}\binom{i}{j}A^jB^{i-j}\frac{1}{-j\gamma-(i-j)\beta}e^{[-j\gamma-(i-j)\beta]s} = \sum_{j=0}^{i}\binom{i}{j}(vf)^j(u-vf)^{i-j}\frac{1}{-j\gamma-(i-j)\beta}e^{[-j\gamma-(i-j)\beta]s}$.
The resulting definite integral, which can be computed directly, is:

$\sum_{j=0}^{i}\binom{i}{j}(vf)^j(u-vf)^{i-j}\frac{1}{-j\gamma-(i-j)\beta}[e^{[-j\gamma-(i-j)\beta]s_2}-e^{[-j\gamma-(i-j)\beta]s_1}]$
$=(u-vf)^i\sum_{j=0}^{i}\binom{i}{j}(\frac{vf}{u-vf})^j\frac{1}{j(\beta-\gamma)-i\beta}[e^{[j(\beta-\gamma)-i\beta]s_2}-e^{[j(\beta-\gamma)-i\beta]s_1}].$
$=\frac{1}{\beta-\gamma}(u-vf)^i\sum_{j=0}^{i}\binom{i}{j}(\frac{vf}{u-vf})^j\frac{1}{j-if}[e^{[j(\beta-\gamma)-i\beta]s_2}-e^{[j(\beta-\gamma)-i\beta]s_1}]$
In the degenerate case $v=0$, the antiderivative is simply $\int([ue^{-\beta s}]^i ds=u^i\int e^{-i\beta s}ds=-\frac{u^i}{i\beta}e^{-i\beta s}$, with the definite integral $-\frac{u^i}{i\beta}[e^{-i\beta s_2}-e^{-i\beta s_1}]$.

Laurent expansion

For a Laurent expansion of order $i\in\mathbb{Z}$, $L_i(s;u,v)=\int[vfe^{-\gamma s}+(u-vf)e^{-\beta s}]^{-i}ds$
$=\int(Ae^{-\gamma s}+Be^{-\beta s})^{-i}ds$. Considering the integrand:
$l(s;u,v)=(Ae^{-\gamma s}+Be^{-\beta s})^{-i}=\frac{1}{(Ae^{-\gamma s}+Be^{-\beta s})^i}=\frac{1}{(Ae^{-\gamma s}+Be^{-\beta s})^i}=\frac{1}{(1+\frac{B}{A}e^{-(\beta-\gamma)s})^iA^ie^{-i\gamma s}}$. Defining $z=-\frac{B}{A}e^{-(\beta-\gamma)s}$, this yields $l=\frac{1}{(1-z)^iA^ie^{-i\gamma s}}$. Furthermore, $e^{-i\gamma s}=[e^{-(\beta-\gamma)s}]^\rho$ for $\rho=\frac{-(\beta-\gamma)}{-i\gamma}=\frac{\beta-\gamma}{i\gamma}$. Therefore, $z^\rho=(-\frac{B}{A}e^{-(\beta-\gamma)s})^\rho=(-\frac{B}{A})^\rho e^{-i\gamma s}$ and $A^ie^{-i\gamma s}=A^i(-\frac{B}{A})^{-\rho}z^\rho$. This, in turn, yields $l=(-\frac{B}{A})^\rho A^{-i}(1-z)^{-i}z^{-\rho}$.
Using variable substitution, $\int_{s_1}^{s_2}l(s;u,v)ds=\int_{z_1}^{z_2}l(z;u,v)dz$, where $z_i=-\frac{B}{A}e^{-(\beta-\gamma)s_i}$. $\frac{dz}{ds}=-(\beta-\gamma)z$, yielding $dz=-(\beta-\gamma)^{-1}z^{-1}ds$.
The solution to $\int_{z_1}^{z_2}l(s;u,v)dz$ is given by the incomplete beta function $B(z;\mu_1,\mu_2)$ [2]:
$r=(-\frac{B}{A})^\rho A^{-i}(\gamma-\beta)^{-1}[B(z_2;\mu_1,\mu_2)-B(z_1;\mu_1,\mu_2)]$, , $\mu_1=-\rho$, and $\mu_2=1-i$. This solution is, in principle, exact. However, methods for the computation of $B(z;\mu_1,\mu_2)$ for arbitrary complex arguments z are unavailable [2], presumably due to the ubiquity of the function in the field of statistical computation, which only requires evaluation of $B(x;\mu_1,\mu_2)$ for x on the real line. Therefore, it is more practical to use the Gaussian hypergeometric function representation $B(z;\mu_1,\mu_2)$
$=\frac{z^{\mu_1}}{\mu_1}{}_2F_1(1-\mu_2,\mu_1;\mu_1+1;z)$, yielding the desired result:
$L_i(s;u,v)=(-\frac{B}{A})^\rho A^{-i}(\gamma-\beta)^{-1}\rho^{-1}[z_2^{-\rho}{}_2F_1(i,-\rho;-\rho+1;z_2)-z_1^{-\rho}{}_2F_1(i,-\rho;-\rho+1;z_1)]$.
In the degenerate case $v=0$, the antiderivative is simply $\int([ue^{-\beta s}]^{-i}ds=u^{-i}\int e^{i\beta s}ds=\frac{1}{i\beta u^i}e^{i\beta s}$, with the definite integral $\frac{1}{i\beta u^i}[e^{i\beta s_2}-e^{i\beta s_1}]$.
If a robust evaluation routine for ${}_2F_1$ is available, it is also straightforward to compute the Taylor terms $T_i(s;u,v)=L_{-i}(s;u,v)$ without explicitly performing the summation described above.

Considerations for the numerical evaluation of $\phi(u,v)$

Numerical stability

Binomial coefficients

The expressions for Ω_i described in the section "**Burst generating functions and their expansions**" require the computation of binomial coefficients. For the geometric and shifted geometric distributions of burst sizes, which have binomial coefficients on the order of $\binom{N_T}{\cdot}$ (for small integer N_T), calculating the coefficients *via* the log-gamma function $\ln\Gamma$ is recommended, per $\binom{m}{n}=\exp(\ln\Gamma(m+1)-\ln\Gamma(n+1)-\ln\Gamma(m-n+1))$. For the degenerate and uniform distributions (with modest to large b), calculating the coefficients *via* $\ln\Gamma$ is required for practical implementation.

Overflow and underflow for large b

Although the approximation $\sum_i \Omega_i \int U(s)^i ds$ converges to the desired generating function $\frac{\phi}{k_i}$, the series as written is not ideal for numerical evaluation by the independent computation of Ω_i and $\int U(s)^i ds$. In particular, the Taylor coefficients for the geometric and shifted geometric distribution scale with b^i and $(b-1)^i$ respectively. Similarly, their Laurent coefficients scale with b^{-i} and $(b-1)^{-i}$. Since b may be fairly large (up to 300) for mammalian systems [3], the computation of its powers can produce overflow or underflow problems for Taylor and Laurent coefficients, respectively.
Conversely, the Taylor integrals for the degenerate system scale with $\frac{1}{i^i}$, and the Laurent ones scale with i^i.
In case of the degenerate system, issues can be ameliorated somewhat by computing $\frac{\Omega_i}{i^i}$ and $i^i \int U(s)^i ds$, which yields a $(\frac{b}{i})^i$ term and balances the growth behavior.
It is not clear whether an analogous approach is available for the non-degenerate case, although solutions based on Stirling's approximation ($i! \sim i^i \sqrt{2\pi i} e^{-i}$) may be viable for larger i.

Special function evaluation

Exponential integral

The solution for the Laurent expansion of the degenerate system, which only involves the evaluation of a parameter-free exponential integral $E_1(z)$, lends itself to optimization. We found that the following combination of approximations yields no more than 10^{-8} relative error with respect to the MATLAB function `expint` throughout the entire complex plane [4]. In this enumeration, $x \equiv Re(z)$ and $y \equiv Im(z)$. For benchmarking, $\tau = T_M/T_A$, where T_M is the runtime of the built-in function and T_A is the runtime of each approximation routine. The functions were run over a 1000×1000 point grid for $x, y \in [-35, 35]$.

1. Exterior region $[x/17+0.3824)^2+(y/13)^2 > 1]$: 6th order Padé approximation (Luke p.110-112 [5], table 4.3): `pade1`. $\tau = 45.7$.

2. Exterior region $[((x+10)/15)^2 + (y/9.5)^2 > 1]$: 10th order Padé approximation (Luke p.110-112 [6], table 4.3): `pade2`. $\tau = 40.4$.

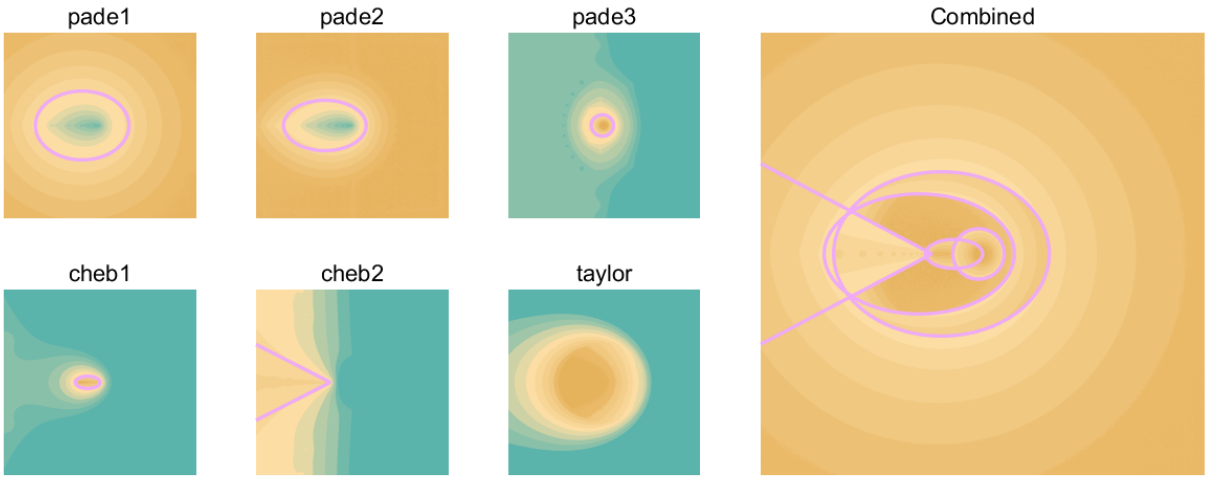
3. Elliptic region $[((x+0.65)/4.05)^2 + (y/4)^2 < 1]$: 10th order Padé approximation (Luke p.107-108 [6], table 4.2): `pade3`. $\tau = 33.8$.

4. Elliptic region $[((x+4.5)/4.5)^2 + (y/2.3)^2 < 1]$: 20th order Chebyshev approximation (Luke p.104 [6], table 4.1): `cheb1`. $\tau = 15.4$.

5. Radial region $[x < -8 \cap |y| < (-x-8)0.5294)]$: 20th order Chebyshev approximation (Luke p.105 [6], table 4.1): `cheb2`. $\tau = 12.6$.

6. Annular region (all other values of z): 55th order series approximation (Zhang and Jin [7]): `taylor`. $\tau = 7.3$.

The resulting performance is shown in Supplementary Figure 1. Teal denotes error above 10^{-8}; brown denotes error below 10^{-8}; the purple lines are the approximation regions. The combined approximation produced $\tau = 24.1$, with maximum error of $10^{-7.9}$.



Supplementary Figure 1: Exponential integral approximations.

Domain decomposition: geometrically-distributed bursts

The Taylor and Laurent approximations are only valid over some domains. For example, the Laurent approximation to $\frac{b\tilde{u}(s)}{1-b\tilde{u}(s)}$ is only valid for $|b\tilde{u}(s)| \equiv |U(s)| > 1$. We split the domain of integration into multiple sub-domains, and integrate each one separately. The domain boundaries are values of s $(s_1, s_2, s_3...)$ where $|U(s)| = \alpha$, such that α is within the domain of convergence for both Taylor and Laurent approximations.
The number of domains depends upon the number of times the function $|U(s)|$ crosses α. From integration from 0 to t, where the function crosses α at $s_1, s_2, s_3..., s_i$, the number of domains is $i + 1$. By the mean value theorem, for differentiable $|U(s)|$, $i \leq j + 1$, where j is the number of points such that $\frac{d|U(s)|}{ds} = 0$.
In the present section, we demonstrate that *at most* four domains are necessary; equivalently, $\frac{d|U(s)|}{ds} = 0$ for at most *two* $s \geq 0$.

Selection of threshold α

For the geometric distribution of burst sizes, the region of convergence for the Taylor expansion is a circle of radius 2 centered at -1, while the region of convergence for the Laurent expansion is a circle of radius 1 centered at 0. For the relevant case of $Re(U) \leq 0$, the minimum distance between the boundaries of the regions of convergence is achieved at $Re(U) = 0$. The corresponding region of convergence is $(i, \sqrt{3}i)$. As shown in **Figure 1c**, we select α equidistant from the two regions' boundaries, as a minimax strategy. This threshold for $|U|$ corresponds to $\alpha = \frac{1+\sqrt{3}}{2b}$.
For the shifted geometric distribution, the region of convergence for the Taylor expansion is a circle of radius $\frac{2}{b-1}$ centered at $-\frac{1}{b-1}$, while the region of convergence for the Laurent expansion is a circle of radius $\frac{1}{b-1}$ centered at 0. The minimum distance between the boundaries is again achieved at $Re(U) = 0$, defining the region of convergence $(\frac{1}{b-1}i, \frac{\sqrt{3}}{b-1}i)$. Therefore, the minimax optimal threshold corresponds to $\alpha = \frac{1+\sqrt{3}}{2(b-1)}$.

Degenerate case: $\beta = \gamma$

In the degenerate case where the export rate is equal to the degradation rate, $U(s) = e^{-\gamma s}(u+\gamma v s)$.
$|U|^2 = e^{-2\gamma s}(u+\gamma v s)(\bar{u}+\gamma\bar{v}s)$
$= e^{-2\gamma s}(u\bar{u}+\gamma^2 v\bar{v}s^2+\gamma[u\bar{v}+\bar{u}v]s)$
$= e^{-2\gamma s}(\gamma^2|v|^2s^2+\gamma[u\bar{v}+\bar{u}v]s+|u|^2)$.
Note that $|U(s)| = 0$ only if $U(s) = 0$. This may occur if $u = v = 0$ or $u = s = 0$ which are trivial edge cases. The alternative case $u = -\gamma v s$ is not relevant since $\gamma, s > 0$ and both u and v are only evaluated in the negative real half-plane. Thus, all nontrivial combinations of u, v, s yield a strictly positive $|U(s)|$ whose behavior is one-to-one with that of $|U(s)|^2$.
Differentiation yields $\frac{d|U|^2}{ds}$
$= e^{-2\gamma s}[2\gamma^2|v|^2s+\gamma[u\bar{v}+\bar{u}v]]-2\gamma e^{-2\gamma s}[\gamma^2|v|^2s^2+\gamma[u\bar{v}+\bar{u}v]s+|u|^2]$
$= e^{-2\gamma s}[(-2\gamma^3|v|^2)s^2+(2\gamma^2|v|^2-2\gamma^2[u\bar{v}+\bar{u}v])s+(\gamma[u\bar{v}+\bar{u}v]-2\gamma|u|^2)]$
$= 2\gamma e^{-2\gamma s}[(-\gamma^2|v|^2)s^2+(\gamma|v|^2-\gamma[u\bar{v}+\bar{u}v])s+(\frac{1}{2}[u\bar{v}+\bar{u}v]-|u|^2)]$
$= 2\gamma e^{-2\gamma s}[As^2+Bs+C]$,
where $A = -\gamma^2|v|^2 \leq 0$, $B = \gamma(|v|^2-[u\bar{v}+\bar{u}v])$, and $C = \frac{1}{2}[u\bar{v}+\bar{u}v]-|u|^2$
For $u = 0$ or $v = 0$, explicit solutions to $|U| = \alpha$ can be directly computed using the Lambert W function. For $u, v \neq 0$, a generalized Lambert W function is required for an analytical solution [8, 9]. Although approximation methods are available [10, 11], they are not generally necessary here. Instead, a numerical scheme can be used to compute coarse estimates for roots constrained by the zeroes of $\frac{d|U|^2}{ds}$. As long as the roots are within $\Delta\alpha$ of α, such that $(\alpha-\Delta\alpha, \alpha+\Delta\alpha)$ is within the radius of convergence for both expansions, the approximation is valid.

General case: $\gamma_e \neq \gamma_c$

An analogous demonstration may be provided for the general case where the export rate is not equal to the degradation rate.
As elsewhere, $f \equiv \frac{\beta}{\beta-\gamma}$. $U(s) = vfe^{-\gamma s}+(u-vf)e^{-\beta s}$.
$|U|^2 = [vfe^{-\gamma s}+(u-vf)e^{-\beta s}][\bar{v}fe^{-\gamma s}+(\bar{u}-\bar{v}f)e^{-\beta s}]$
$= |vf|^2e^{-2\gamma s}+|u-vf|^2e^{-2\beta s}+[vf(\bar{u}-\bar{v}f)+\bar{v}f(u-vf)]e^{-(\gamma+\beta)s}$
$= |vf|^2e^{-2\gamma s}+|u-vf|^2e^{-2\beta s}+f[v(\bar{u}-\bar{v}f)+\bar{v}(u-vf)]e^{-(\gamma+\beta)s}$
$= |vf|^2e^{-2\gamma s}+|u-vf|^2e^{-2\beta s}+f[v\bar{u}-|v|^2f+\bar{v}u-|v|^2f]e^{-(\gamma+\beta)s}$
$= f^2|v|^2e^{-2\gamma s}+|u-vf|^2e^{-2\beta s}+f([\bar{u}v+u\bar{v}]-2|v|^2f)e^{-(\gamma+\beta)s}$.
$|u-vf|^2 = (u-vf)(\bar{u}-\bar{v}f) = (|u|^2+f^2|v|^2-f[\bar{u}v+u\bar{v}])$.
Therefore, $|U|^2 = f^2|v|^2e^{-2\gamma s}+(|u|^2+f^2|v|^2-f[\bar{u}v+u\bar{v}])e^{-2\beta s}+f([\bar{u}v+u\bar{v}]-2|v|^2f)e^{-(\gamma+\beta)s}$.
Differentiation yields $\frac{d|U|^2}{ds}$
$= -2\gamma f^2|v|^2e^{-2\gamma s}-2\beta(|u|^2+f^2|v|^2-f[\bar{u}v+u\bar{v}])e^{-2\beta s}-(\gamma+\beta)f([\bar{u}v+u\bar{v}]-2|v|^2f)e^{-(\gamma+\beta)s}$
$= 0$ whenever $Ce^{-2\gamma s}+Ae^{-2\beta s}+Be^{-(\gamma+\beta)s} = 0$,
Where $C = -2\gamma f^2|v|^2 < 0$, $A = -2\beta(|u|^2+f^2|v|^2-f[\bar{u}v+u\bar{v}]) < 0$, and
$B = -(\gamma+\beta)f([\bar{u}v+u\bar{v}]-2|v|^2f) \in \mathbb{R}$. Multiplying both sides by $e^{2\gamma s}(>0)$,
$0 = C+Ae^{-2(\beta-\gamma)s}+Be^{-(\beta-\gamma)s}$
$= Ae^{2(\gamma-\beta)s}+Be^{(\gamma-\beta)s}+C$
Making the substitution $z = e^{(\gamma-\beta)s}$,
$0 = Az^2+Bz+C$

Note that two roots exist, and the solutions z^* may be converted back to the time domain per $s^* = \frac{1}{\gamma-\beta} \ln z^*$. Only real positive extrema s^* are pertinent to the domain partition; all others may be discarded.
It remains to note that the degenerate cases $u = 0$ or $v = 0$ yield at most one root, which is trivial to compute.

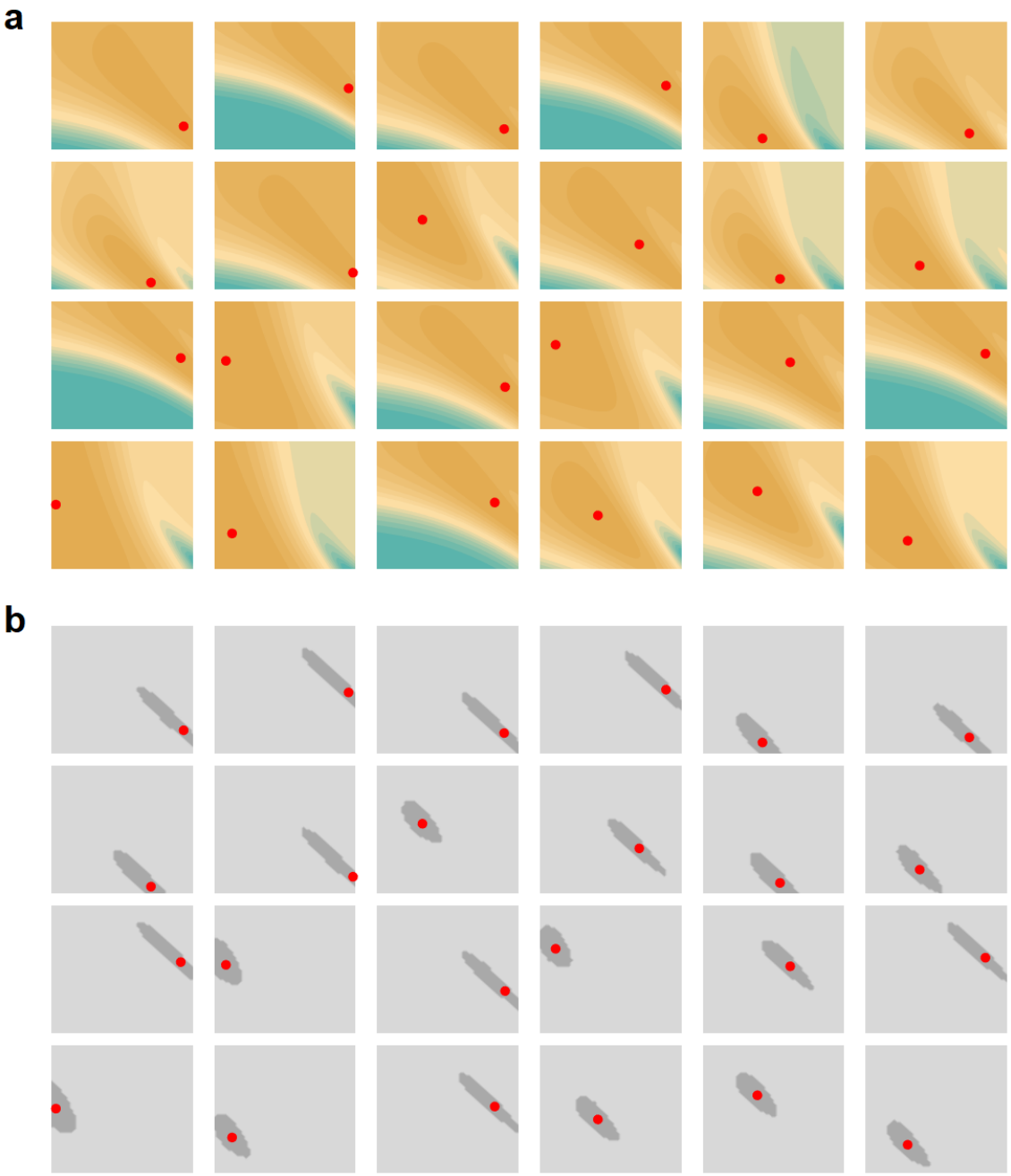
Root computation

Apart from the degenerate cases described above, where the roots can be computed *via* logarithms or the Lambert W function, the determination of domain boundaries requires the use of numerical routines. Specifically, we perform 20 iterations of the Newton-Raphson method to find locations where $|U|^2 = \alpha^2$. If the routine fails to converge, we switch to the bisection method and search in the domain bracketed by extrema of $|U|^2$; as discussed above, the locations of these extrema are known analytically. If the root is not bracketed (i.e., the rightmost extremum at the location s^* has a value $y^* > \alpha^2$), we use the heuristic bracket $(s^*, s^* + 10/\min(\beta, \gamma))$, which suffices due to the fast (exponential) decay of the function $|U|^2$.

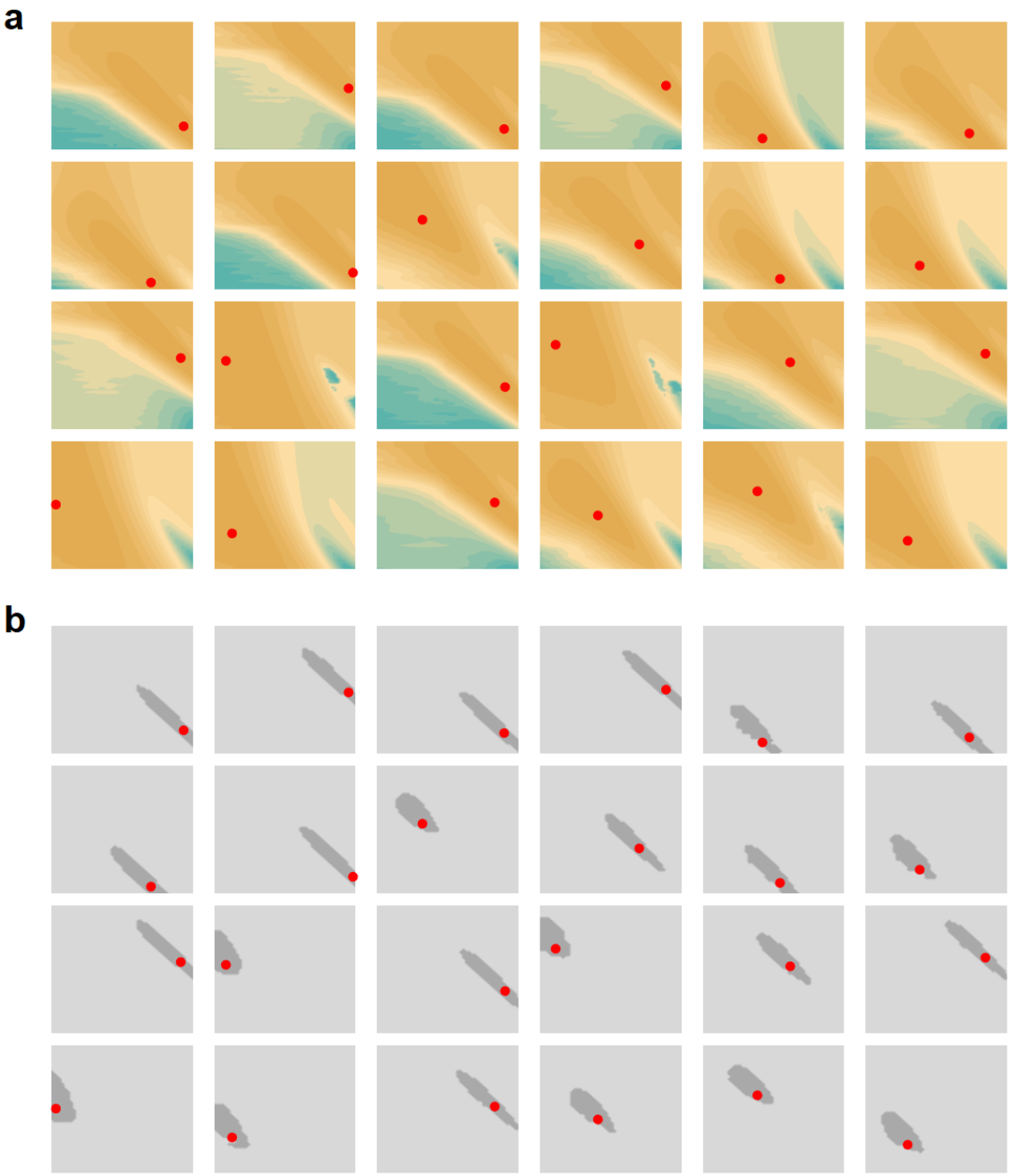
Validation of low-Laurent fidelity approximations

Concordance between quadrature and special function solutions

Using the stochastic simulation algorithm, we generated 24 synthetic datasets with 10^4 cells. The parameter space was restricted to $\gamma = \beta = 1$, $k_i \in \{-1, 1\}$, and $b \in \{0.1, 2\}$; parameter sets were selected by uniform sampling the space. The simulated model used a geometric burst size. As elsewhere, the simulation was run until the equilibration time horizon defined as $T = 5/\min(k_i, \gamma)$. To limit computational demands for the computation of proposed distributions, parameter sets that yielded more than 350 molecules of any particular mRNA species at any point in the simulation were rejected. This effectively depleted the region of parameter space with very high k_i and b.
The parameter space was scanned using a 50×50 grid. For each of the 2500 parameter sets, proposed distributions were generated using quadrature and an analytical approximation using a seventh-order Taylor series and first-order Laurent series (a *7/1 approximation schema*). The computation of each distribution was timed with and without the final call to `ifft2`, the MATLAB routine implementing the 2-dimensional inverse Fourier transform.
Each proposed distribution was compared to the synthetic data using Kullback-Leibler divergence, Kolmogorov-Smirnov distance, the Wasserstein metric, and an *ad hoc* characteristic function (chf) distance, described below.
As evident from a comparison of Supplementary Figures 2 and 3, the likelihood landscapes computed using the two landscapes are similar, although artifacts exist due to the low order of the Laurent approximation. Since the global behavior and the local landscape contours are preserved, we suggest that the approximation is sufficient to perform optimization for parameter estimation. For example, *ad hoc* confidence sets constructed by selecting the parameter sets with the lowest 5% K-L divergence values, are quite similar, and all contain the true parameter value.



Supplementary Figure 2: (a) Kullback-Leibler divergence landscapes for 24 synthetic datasets, computed using quadrature. (b) Set of parameters in the fifth percentile.



Supplementary Figure 3: (a) Kullback-Leibler divergence landscapes for 24 synthetic datasets, computed using special function approximations. (b) Set of parameters in the fifth percentile.

Concordance between likelihood and chf distance landscapes

The chf distance was defined by $M = \log_{10} \sum_{n,m} |G_{n,m} - G^*_{n,m}|^2$, where G is the computed characteristic function for a proposed set of parameters, G^* is the 2-dimensional Fourier transform of the data, and the sum is taken over all sample points u, v, indexed by n and m respectively. This implementation follows a simple interpretation of the chf divergence between discrete-valued distributions described in previous literature [12].

As evident from a comparison of Supplementary Figures 3 and 4, the landscapes produced using the chf distance are quite different from the K-L ones. Furthermore, confidence sets constructed using the chf distance are distorted compared to Bayesian confidence sets. However, examination of the landscapes also reveals that the global behavior of the chf distance yields a unimodal solution region, whereas the K-L landscapes have potential local minima. This mitigates the danger of finding local minima and facilitates the use of computationally efficient gradient-based optimization methods. Overall, we suggest that the approximation is sufficient to perform optimization for parameter estimation. For example, *ad hoc* confidence sets constructed by selecting the parameter sets with the lowest 5% chf divergence values all contain the true parameter value.

We hypothesize that the behavior of the chf landscapes - the diagonal degeneracies as well as the uniqueness properties - are related to the chf estimator's interpretation as a moment-based statistic [13, 14]. However, we anticipate the chf distance to be more broadly applicable in general inference problems due to the explicit use of the entire probability distribution (under the Fourier transform) as well as the challenges of constructing explicit analytical expressions for higher moments, especially in models with large numbers of parameters.

Timing and precision comparisons

Supplementary Figure 5 demonstrates the timing difference between the two methods, using the data from the 24 scans over the 2500-point grid described above. The runtimes were normalized by the state space size $\mathcal{N}$. The use of the 7/1 approximation schema yields a mean runtime decrease of 39%. As the state space sizes are small by design of the validation procedure, the benefit from forgoing the Fourier transform step is marginal, only on the order of 1%.

Supplementary Figure 6 demonstrates the precision of the special function method relative to quadrature using four different divergence measures. Finally, Supplementary Figure 7 demonstrates the dependence of these errors on the state space size. Integral measures of error, such as the chf and Wasserstein distances, tend to grow with increased state space, possibly due to accumulation of errors. The Kolmogorov-Smirnov distance, a supremum, does not show a distance-based effect, presumably because it is dominated by tail oscillatory phenomena observed at low Laurent approximation order. Kullback-Leibler divergence likewise shows low dependence on distance, which may be explained by concentration of error in the tail of the distribution, which has few datapoints.

Addenda

Solutions for arbitrary initial conditions

As described by Singh and Bokes [1], the generating function at time t for an arbitrary initial distribution is given by $\phi(u, v, t) = k_i \int_0^t (M(U(s)) - 1)ds + \phi(U(t), V(t), 0)$, where $\phi(u, v, 0)$ is the factorial cumulant generating function of the initial distribution. If $n(t = 0) = n_0$ and $m(t =$

$0) = m_0$, the corresponding generating function is $G(x, y, t = 0) = \sum_n \sum_m x^n y^m P(n, m, t = 0) = x^{n_m} y^{m_0}$. Its logarithm $\phi(u, v, 0)$ is simply $n_0 \ln(1 + u) + m_0 \ln(1 + v)$. To find the value of this contribution, all that remains is to compute $U(t)$ and $V(t)$, where $U(s)$ is given above and $V(s) = ve^{-\gamma s}$. This method can be trivially extended to an arbitrary initial distribution of molecules by summing over the appropriate n and m and using the observed initial copy-number distribution as the initial condition $P(n, m, t = 0)$.

Negative binomial burst generating function

A negative binomial (NB) burst distribution has $P(B = \rho) = \binom{\rho+a-1}{\rho}(1 - p)^a p^\rho$, where $p = \frac{b}{1+b}$ and the mean burst size is ba. The resulting generating function is $F(x) = (\frac{1-p}{1-px})^a$. Substituting $M(u) = F(1 + u)$ yields $M(u) - 1 = \frac{1}{(1-bu)^a} - 1$, which reduces to the geometric distribution for $a = 1$. Defining $X = (bu)^{-1}$, the first term can be rewritten as $\frac{X^a}{X^a(1-X^{-1})^a} = \frac{X^a}{(X-1)^a}$. This form affords the Taylor expansion in the form of $\sum_{i=0}^{\infty} X^{i+a}(-1)^{i-a}\binom{-a}{i}$, and an analogous Laurent expansion. Therefore, it is, in principle, possible to compute the CME solutions through a series of integrals of the form $\int U^{i+a} ds$, $i \in \mathbb{Z}$, which are given by $\Gamma(i + a, \cdot)$ for the degenerate case and ${}_2F_1(i + a, \cdot; \cdot; \cdot)$ otherwise.
Previous discussions of the NB model treat it as a natural generalization of the geometric model, but do not motivate its use from physical arguments. Therefore, we do not develop the solution to the model in further detail.

References

[1] Abhyudai Singh and Pavol Bokes. Consequences of mRNA Transport on Stochastic Variability in Protein Levels. *Biophysical Journal*, 103(5):1087–1096, September 2012.

[2] Milton Abramowitz and Irene Stegun, editors. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. United States National Bureau of Standards, 9 edition, 1970.

[3] R. D. Dar, B. S. Razooky, A. Singh, T. V. Trimeloni, J. M. McCollum, C. D. Cox, M. L. Simpson, and L. S. Weinberger. Transcriptional burst frequency and burst size are equally modulated across the human genome. *Proceedings of the National Academy of Sciences*, 109(43):17454–17459, October 2012.

[4] MATLAB R2019b, 2019.

[5] Yudell L Luke. *Mathematical Functions and their Approximations*. Academic Press, New York, San Francisco, London, 1975.

[6] Yudell L Luke. *The Special Functions and Their Approximations, Vol 2*. Academic Press, London ; New York, 1969.

[7] Shanjie Zhang and Jianming Jin. *Computation of Special Functions*. Wiley, New York, 1 edition, 1996.

[8] Tony C. Scott and Robert B. Mann. General Relativity and Quantum Mechanics: Towards a Generalization of the Lambert W Function. *arXiv:math-ph/0607011*, September 2006. arXiv: math-ph/0607011.

[9] Aude Maignan and Tony C. Scott. Fleshing out the generalized Lambert W function. *ACM Communications in Computer Algebra*, 50(2):45–60, August 2016.

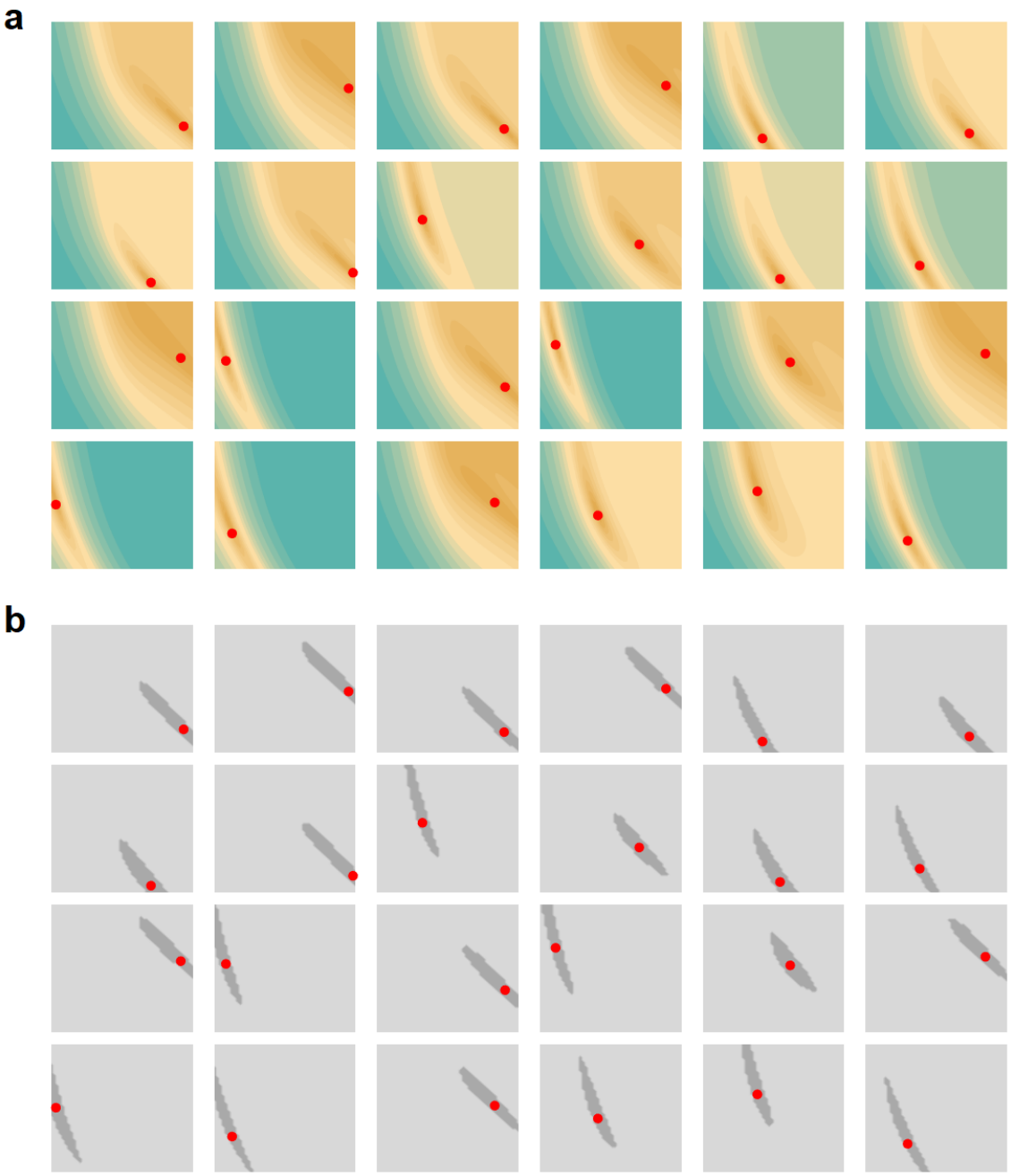
[10] Tony C. Scott, Greg Fee, and Johannes Grotendorst. Asymptotic series of generalized Lambert W function. *ACM Communications in Computer Algebra*, 47(3/4):75–83, January 2014.

[11] Paul Castle. Taylor series for generalized Lambert W functions. *arXiv:1801.09904 [math]*, January 2018. arXiv: 1801.09904.

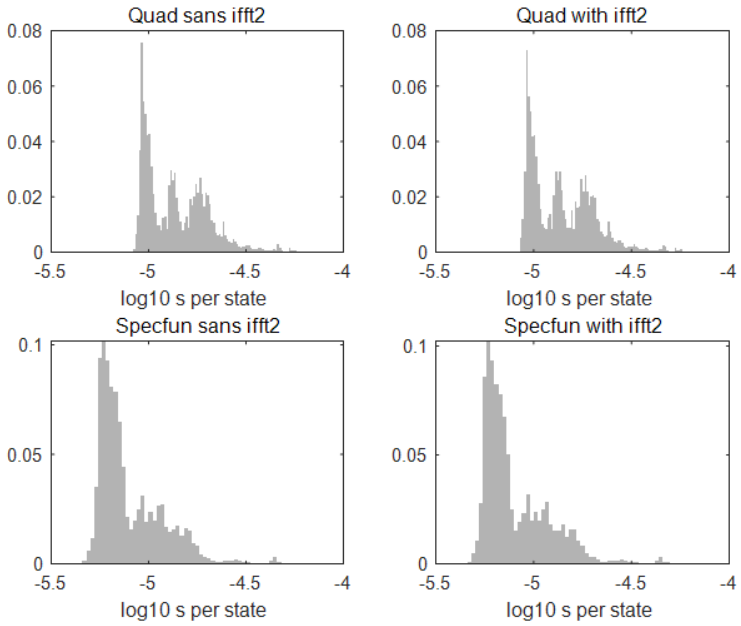
[12] M. Bee and L. Trapin. A characteristic function-based approach to approximate maximum likelihood estimation. *Communications in Statistics - Theory and Methods*, 47(13):3138–3160, July 2018.

[13] Jun Yu. Empirical Characteristic Function Estimation and Its Applications. *Econometric Reviews*, 23(2):93–123, December 2004.

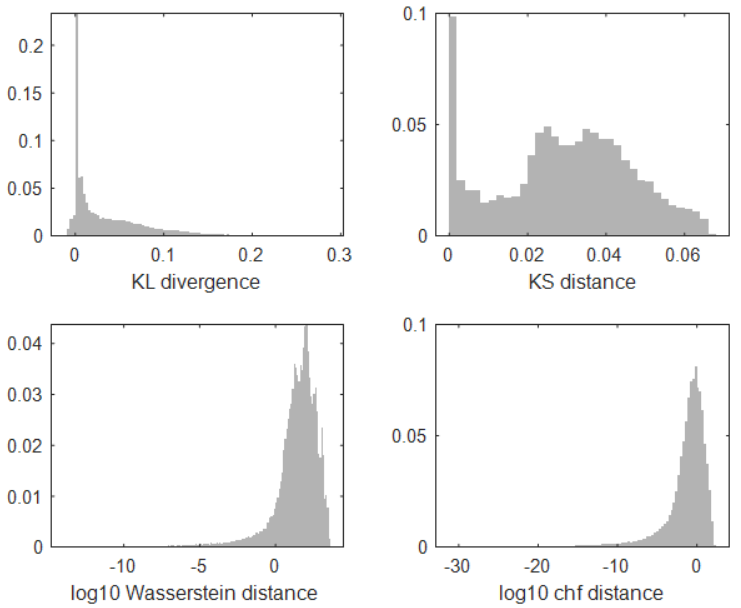
[14] Marine Carrasco and Rachidi Kotchoni. Efficient Estimation using the Characteristic Function. *Econometric Theory*, 33(2):479–526, April 2017.



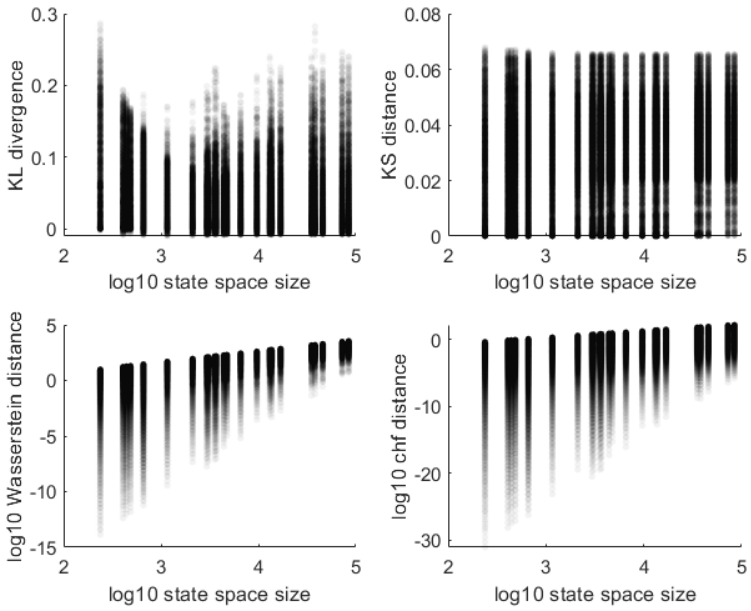
Supplementary Figure 4: (a) chf distance landscapes for 24 synthetic datasets, computed using special function approximations. (b) Set of parameters in the fifth percentile.



Supplementary Figure 5: Timing of quadrature and special function computations.



Supplementary Figure 6: Divergence of special function results with respect to quadrature results.



Supplementary Figure 7: Dependence of divergence measures on state space size.