Extrinsic noise and heavy-tailed laws in gene expression

Lucy Ham [1,†], Rowan D. Brackston [2,†], Michael P.H. Stumpf [1,2]

1 *School of BioSciences and School of Mathematics and Statistics, University of Melbourne, Parkville VIC 3010, Australia*
2 *Centre for Integrative Systems Biology and Bioinformatics, Department of Life Sciences, Imperial College London, London SW7 2AZ, UK*
†*L.H. and R.D.B. contributed equally to this work.*
lucy.ham@unimelb.edu.au, r.brackston13@imperial.ac.uk, mstumpf@unimelb.edu.au

Abstract

Noise in gene expression is one of the hallmarks of life at the molecular scale. Here we derive analytical solutions to a set of models describing the molecular mechanisms underlying transcription of DNA into RNA. Our *Ansatz* allows us to incorporate the effects of extrinsic noise – encompassing factors external to the transcription of the individual gene – and discuss the ramifications for heterogeneity in gene product abundance that has been widely observed in single cell data. Crucially, we are able to show that heavy-tailed distributions of RNA copy numbers cannot result from the intrinsic stochasticity in gene expression alone, but must instead reflect extrinsic sources of variability.

Trancription is one of the canonical examples of a stochastic process in biology; and as the first step in the central dogma, is of fundamental phenotypic importance. Recent single-cell analysis methods now allow this heterogeneity to be observed in terms of distributions of transcript copy numbers across ensembles of cells. Conveniently, the process is also amenable to the application of elegant mathematical models reminiscent of those in statistical physics Benecke (2008); Schnoerr et al. (2017).

The *Telegraph process* is widely used to model stochastic RNA transcription initiation, originally detailed in Peccoud and Ycart (1995) and discussed in recent reviews Munsky et al. (2012); Jones and Elf (2018). In the slightly generalised form we consider here, the gene is either active or inactive: when active, mRNA is transcribed as a Poisson process with rate K_1, while when inactive, basal transcription may still occur at lower rate, K_0. The mRNA degradation is modelled by a first-order Poissonian degradation process with rate, δ, while switching between the two states occurs at rates ν_0 (turn off) and ν_1 (turn on), as shown schematically in Fig. 1(a). This leads to a Markov process for the copy number n of the mRNA molecules at time t and gene state i, with an associated master equation for the probability $p_i(n, t)$,

$$
\begin{aligned}
(\forall n \geq 1)\, \partial_t p_i(n, t) = & - (\nu_{i'} + K_i + \delta n)p_i(n, t) \\
& + \delta(n + 1)p_i(n + 1, t) \\
& + K_i p_i(n - 1, t) \\
& + \nu_i p_{i'}(n, t),
\end{aligned}
\tag{1}
$$

where $i \in \{0, 1\}$, and $i' = 1 - i$. For, $n = 0$ terms involving $n - 1$ are set to 0.

The master equation Eq. (1) for this "leaky gene" model coincides with (Kepler and Elston, 2001, Eq.'s (2),(3)), though a steady state solution is not given there. When $K_0 = 0$, the equation reduces to that given in (Peccoud and Ycart, 1995, Eq. (5)). Following the approach of Peccoud and Ycart (1995), generating functions yield an
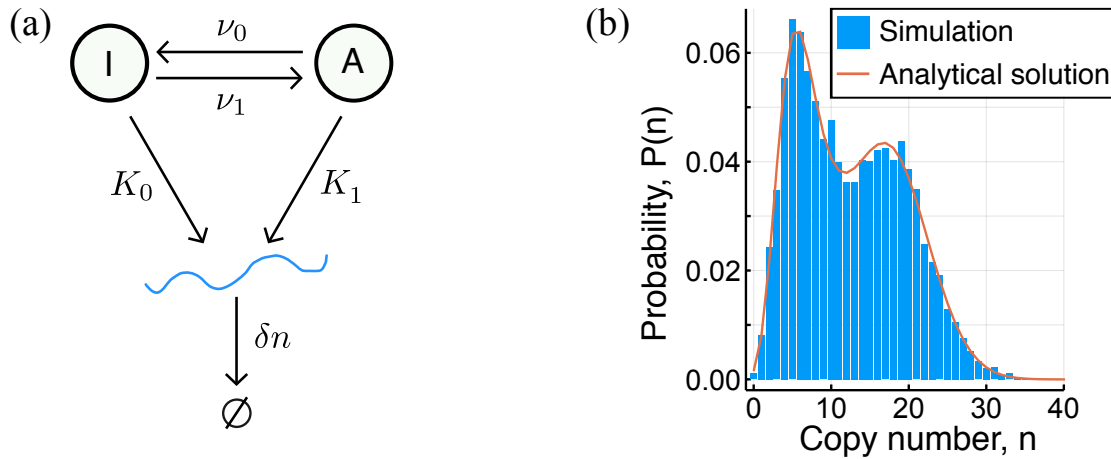
Figure 1: *(a) A schematic of the leaky Telegraph model. (b) A comparison of the analytical solution Eq. (3) and a probability mass function from simulated data.*

exact solution for the probability generating function (pgf) for the mRNA copy number [1]:

$$g(z) = e^{K_1(z-1)}{}_1F_1\left(\nu_0, \nu_0 + \nu_1, w(z-1)\right), \tag{2}$$

where $w := K_1 - K_0$ and rates are scaled so that $\delta = 1$. The probability mass function (pmf) is then recovered as $\tilde{p}(n) = \frac{1}{n!}g^{(n)}(0)$, which by the general Leibniz rule gives

$$\tilde{p}(n) = \frac{1}{n!}\sum_{r=0}^{n}\left[\binom{n}{r}K_1^{n-r}e^{-K_1}w^r\frac{\nu_0^{(r)}}{(\nu_0+\nu_1)^{(r)}}\right.$$

$$\left.{}_1F_1\left(\nu_0+r, \nu_0+\nu_1+r, -w\right)\right], \tag{3}$$

where, for real number x and positive integer n, the notation $x^{(n)}$ abbreviates the rising factorial of x.

A useful limiting case of this generalized model is obtained when the active transcription is sufficiently rare so that transcription can be considered to occur in instantaneous bursts ($\nu_0 \gg \nu_1$), and the degradation rate is sufficiently small ($\nu_0 \gg \delta$). This model simultaneously encompasses the two well-known extremes of bursty transcription and constitutive transcription, as we now show.

Under the assumptions $\nu_0 \gg \delta, \nu_1$, the burst size is geometrically distributed, as is already understood Paulsson and Ehrenberg (2000); Ingram et al. (2008):

$$G(M; r) = r(1-r)^M, \tag{4}$$

where $r = \nu_0/(\nu_0 + K_1)$. The master equation Eq. (1) may then be rewritten as,

$$\partial_t p(n, t) = -\left(\nu_1(1-r) + K_0 + \delta n\right)p(n, t)$$

$$+ \nu_1\sum_{j=1}^{n}G(j; r)p(n-j, t)$$

$$+ \delta(n+1)p(n+1, t)$$

$$+ K_0 p(n-1, t). \tag{5}$$

The steady-state solution is again obtained using generating functions, with the exact expression for the pgf given by

$$g(z) = \left(\frac{r}{1 - z(1-r)}\right)^{\nu_1/\delta}\exp\left(\frac{K_0}{\delta}(z-1)\right), \tag{6}$$
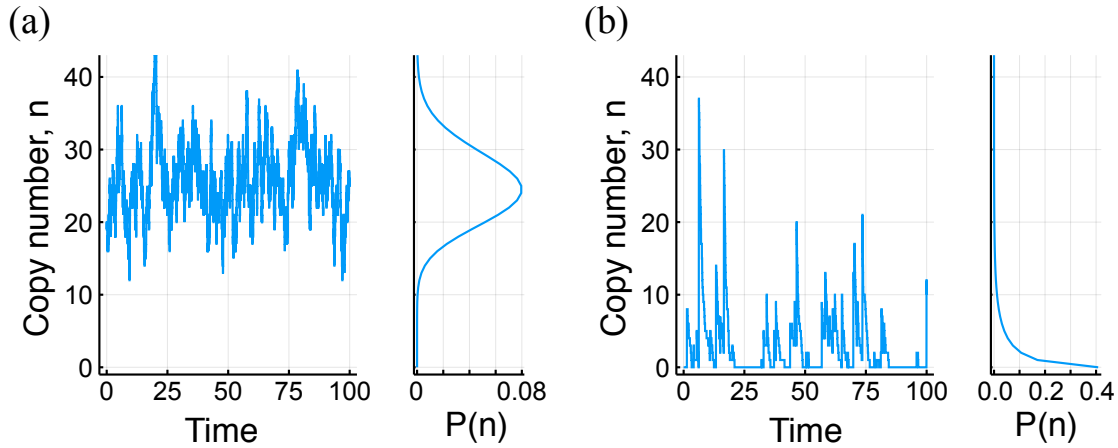
Figure 2: *Time series and corresponding probability distributions for (a) constitutive and (b) bursty transcription.*

which is the product of the pgf for the negative binomial distribution $\text{NegBin}(\nu_1/\delta, r)$ and the Poisson distribution $\text{Pois}(K_0/\delta)$. Thus ?,

$$\tilde{p}(n) = \frac{1}{n!} \sum_{i=0}^{n} \left[\binom{n}{s} \frac{\Gamma(\nu_1/\delta + (n-s))}{\Gamma(\nu_1/\delta)(n-s)!} (1-r)^{(n-s)} r^{\nu_1/\delta} \right.$$

$$\left. \frac{e^{-K_0/\delta} K_0^s}{\delta^s} \right]. \quad (7)$$

When $K_0 = 0$, corresponding to bursty gene expression, Eq. (S.23) is the pgf for $\text{NegBin}(\nu_1/\delta, r)$, agreeing with the solution of Paulsson and Ehrenberg (2000). Alternatively, use Eq. (7), using $0^0 := 1$. When $K_1 = 0$ (or indeed is kept constant and $\nu_0 \to \infty$) the burst height parameter $r = \nu_0/(\nu_0 + K_1)$ becomes 1, and the steady state solution Eq. (7) agrees with the solution for constitutive gene expression, namely the $\text{Pois}(K_0/\delta)$ distribution; see Fig. 2 for the pdfs of both of these.

Similarly, when $K_0 = 0$, the generating function Eq. (S.13) for the leaky gene model reduces to the solution for $g(z)$ in (Peccoud and Ycart, 1995, Eq. 20) by way of Kummer's transformation (Olver et al., 2010, §13.2.39). This then yields the following well-known analytical expression for the Telegraph model Peccoud and Ycart (1995); Raj et al. (2006)

$$\tilde{p}(n) = \frac{K_1^n \nu_1^{(n)}}{n!(\nu_0 + \nu_1)^{(n)}} {}_1F_1(\nu_1 + n, \nu_0 + \nu_1 + n, -K_1). \quad (8)$$

The standard Telegraph model for transcription describes the effect of intrinsic noise at the level of a single gene, yet the process will often also be influenced by other sources of variability. Such extrinsic noise has been widely observed experimentally Elowitz (2002); Swain et al. (2002); Raser and O'Shea (2004); Gasch et al. (2017), and considered theoretically Dattani and Barahona (2016); Fu and Pachter (2016); Bressloff and Levien (2019); Thomas (2019), but incorporating these effects into the Master equations has generally proven challenging. The approach we take in the context of the Telegraph model is to consider the model parameters themselves to vary between cells, and therefore to be drawn from probability distributions Lenive et al. (2016). The mRNA copy number then follows a compound distribution,

$$\tilde{q}(n; \eta) = \int \tilde{p}(n; \theta) f(\theta; \eta) \, d\theta, \quad (9)$$

where θ is the vector of parameters $[\nu_0, \nu_1, K_0, K_1]$ and the distribution f is a multivariate distribution for θ with hyperparameters η. This model is valid provided that parameter values are static for individual cells but vary across an ensemble of cells according to f, or if the parameter values are dynamic, but change at substantially slower timescales (adiabatically) relative to the transcriptional dynamics. An example of the latter is variation in upstream biological drivers, which is a special case of the extrinsic noise considered
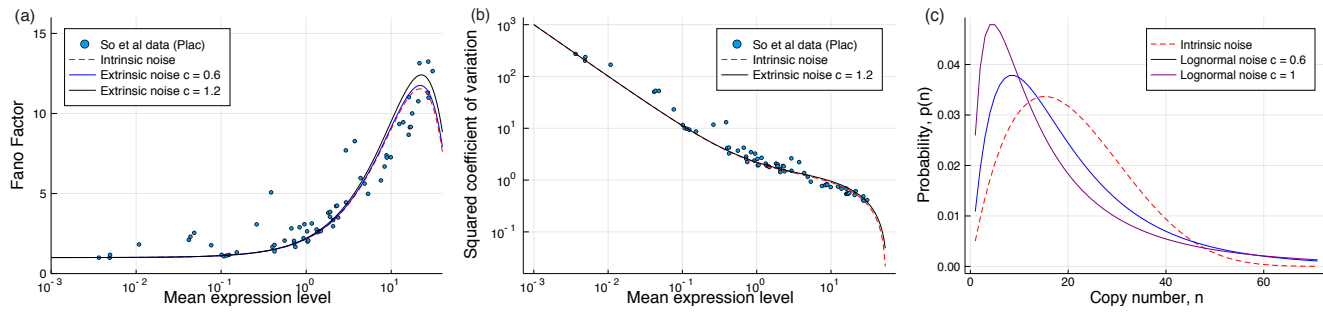
Figure 3: *In figures (a) and (b), the parameters are* $E(K) = K = 54$, $\nu_1 = 0.83$ *and* $\delta = 1$. *(a) The Fano factor of the compound distribution as a function of the mean mRNA copy number (given by Eq.* (15)*), varied by tuning the parameter* ν_0*. Different values of c are plotted against the universal noise scaling curve (intrinsic noise only) given in So et al.* (2011)*. Data is the* P_{lac} *promoter also from So et al.* (2011)*. (b) The squared coefficient of variation as a function of the mean copy number for intrinsic and extrinsic noise* $c = 1.2$*. (c) The effect of lognormal extrinsic noise on K on the copy number distribution, with* $K = 60$, $\nu_0 = 4$, $\nu_1 = 2$ *and* $\delta = 1$*.*

in Dattani and Barahona (2016). For the remainder of the paper, when it is clear from the context that only one rate of transcription is being considered, we will use K in place of K_0 or K_1.

An interesting example arises from Eq. (8) when K is Gamma(α, β) distributed. In this case, the compound distribution $\tilde{q}(n; \alpha, \beta)$ can be shown to be

$$\tilde{q}(n; \alpha, \beta) = \frac{\alpha^{(n)} \nu_1^{(n)}}{n! \beta^n (\nu_0 + \nu_1)^{(n)}} \left(\frac{1}{1+\beta}\right)^n \left(\frac{\beta}{1+\beta}\right)^{\nu_1}$$
$$ {}_2F_1\left(\nu_1 + n, \nu_0 + \nu_1 - \alpha, \nu_0 + \nu_1 + n, \frac{1}{1+\beta}\right). \quad (10)$$

Rather intriguingly, this distribution is also the steady-state protein number distribution found in Shahrezaei and Swain (2008). Gene expression there is modelled as a three-stage stochastic process, but the model can be considered as Telegraph noise, with distribution Eq. (8), on the rate parameter of a Poissonian process. It is surprising that the resultant steady-state distribution coincides with that of Gamma-distributed noise on a Telegraph process.

Another striking example arises by way of constitutive transcription ($\tilde{p}(n) = \text{Pois}(K/\delta)$) again with $K \sim$ Gamma(α, β). In this case, the compound distribution yields the negative binomial distribution,

$$\tilde{q}(n; \alpha, \beta) = \text{NegBin}\left(\alpha, \frac{\beta}{1+\beta}\right). \quad (11)$$

Thus the copy number distribution alone cannot distinguish constitutive expression with extrinsic noise, from a bursty expression without extrinsic noise, which has distribution NegBin $(\nu_1/\delta, r)$. A similar example is given in Dattani and Barahona (2016), where constitutive expression with Beta-distributed K is shown to agree with the solution of the Telegraph model, Eq. (8).

Whether or not the moments of compound distributions converge is a problem of considerable practical importance Willinger et al. (2004). We can provide simple formulæ for all moments of the copy number distribution Eq. (8) under extrinsic noise on the transcription rate K. Noise in K is of particular relevance, as will become apparent in due course. Let $X = X_K$ denote a random variable from $\tilde{p}(n; K)$ (from Eq. (8)) and $Y = Y_\eta$ a random variable from the compound distribution $\tilde{q}(n; K)$ (from Eq. (9)). It can be shown that the nth moment of Y is given by

$$E(Y^n) = \sum_{i=1}^{n} \frac{\nu_1^{(i)_\delta}}{\delta^i (\nu_0 + \nu_1)^{(i)_\delta}} S(n, i) E(K^i), \quad (12)$$

where $S(n, i)$ is a Stirling number of the second kind; the notation $x^{(n)_y}$ abbreviates $x(x+y) \ldots (x + y(n-1))$ for real numbers x, y and positive integer, n. It follows from Eq. (S.31) that if the first two moments of the compounding distribution $f(K; \eta)$ are known, then the mean, variance and Fano factor of the compound

distribution $\tilde{q}(n; \eta)$ can be easily calculated. From Eq. (S.31), and noting that $S(1, 1) = S(2, 1) = S(2, 2) = 1$, the Fano factor of $\tilde{q}(n; \eta)$ is given by

$$\mathrm{FF}(Y) = 1 - \frac{\nu_1}{\delta(\nu_0 + \nu_1)} E(K) + \frac{\nu_1 + \delta}{\delta(\nu_0 + \nu_1 + \delta)} \frac{E(K^2)}{E(K)}. \tag{13}$$

It is also possible to obtain formulæ for the nth moments and Fano factor in the case for constitutive expression with extrinsic noise on K, as well as in the case for bursty expression with extrinsic noise on K. In the former, we can alternatively obtain a formula for the Fano factor from Eq. (13) by taking $\frac{\nu_1}{\nu_0 + \nu_1} = 1$ (corresponding to $\nu_1 \to \infty$, or $\nu_0 \to 0$). This gives

$$\mathrm{FF}(Y) = 1 + \frac{1}{\delta} \frac{\mathrm{Var}(K)}{E(K)}. \tag{14}$$

It has previously been suggested that heterogeneity in mRNA copy numbers and a universal scaling between the mean and Fano factor is attributable primarily to intrinsic noise, as opposed to extrinsic noise So et al. (2011). The purported observational evidence for these claims is: (i) in the limit of low mean mRNA copy number, the Fano factor is approximately one; (ii) at the other extreme of high mean mRNA copy number, the Fano factor decreases sharply rather than approaching a plateau; (iii) the square of the coefficient of variation decreases monotonically with the mean.

These claims do not stand up against the analytical solutions for the Fano factor obtained above. From Eq. (13), the Fano factor of the compound distribution $\tilde{q}(n; K)$ depends only on $E(K)$ and $\mathrm{Var}(K)$ (or equivalently $E(K^2)$) and the values of the parameters ν_0, ν_1, δ. Throughout we let c denote the coefficient of variation, $\sigma(K)/E(K)$, for the noise distribution, $f(K; \eta)$. Noting that $E(K^2)/E(K) = (1 + c^2) E(K)$, Eq. (13) becomes

$$1 \quad - \quad \frac{\nu_1}{\delta(\nu_0 + \nu_1)} E(K) \quad + \quad \frac{\nu_1 + \delta}{\delta(\nu_0 + \nu_1 + \delta)} (1 \quad + \quad c^2) E(K). \tag{15}$$

The situation $E(Y) \to 0$ corresponds to $\nu_0 \to \infty$, which from Eq. (15) easily gives $\mathrm{FF}(Y) \to 1$. Thus, item (i) holds true for all values of c, and therefore for any extrinsic noise distribution on K, provided that the second moment exists.

Figure 3(a) shows qualitatively identical behaviour of the Fano factor as a function of the mean mRNA copy number for values of c close to 1, even for the same parameter values of $\nu_1, \delta, E(K) = K$. We remark that, as in So et al. (2011), the mean is varied by regulating ν_0 only. Values of c below 0.6 have been omitted, as they are visually indistinguishable from the intrinsic noise case at the present resolution.

Finally, we consider the squared coefficient of variation as a function of the mean copy number for different values of c; see Fig. 3(b), and note that values of c smaller than $c = 1.2$ are again visually (and practically) indistinguishable from intrinsic noise and so are omitted. We see that the behaviour is effectively identical to that of intrinsic noise only. Even for unrealistic values of c (for example $c = 200,000$), the squared noise curve continues to satisfy (iii).

We next examine the effect of extrinsic noise on the noise scaling curve in the case for constitutive transcription, again considering only noise on K. In this case, it can be easily shown that the mean copy number $E(Y)$ is equal to $E(K)$. Thus, from Eq. (14), scaling the Fano factor with mean copy number is dependent only on the noise distribution of K. If the coefficient of variation is fixed at c as the noise distribution on K is varied, then the Fano factor is given by $1 + c^2 E(K)$, which is linear in $E(Y) = E(K)$. On semilogarithmic axes and plotted as a function of the mean copy number, this yields the same qualitative behaviour as that found in (Jones et al., 2014, Fig. 2B). The observations after Eq. (11) are pertinent here: with $c = 1/\sqrt{\nu_1}$, identical noise scaling behaviour arises from the two extremes—constitutive expression with extrinsic noise and bursty expression without noise.

Thus far, extrinsic noise, as modelled by the compound distribution, exhibits behaviours that are similar to intrinsic noise alone. We now present a potential qualitative identifier for extrinsic noise: we show that contrary to previous claims Iyer-Biswas et al. (2009), intrinsic noise alone never leads to a heavy-tailed copy number distribution, but find many cases in which extrinsic noise does so. Formally, we take heavy tailed to mean that the moment generating function (mgf) is undefined for positive t, which implies that the tail of the distribution decays more slowly than that of the exponential distribution.

If K and δ are fixed, then the copy number is maximized when the gene remains permanently active, which has distribution $\mathrm{Poisson}(K/\delta)$ and is not heavy tailed. Intuitively then, no compounding of ν_0 and ν_1 alone results in a heavy tailed distribution. A more robust argument is obtained by establishing the following inequality for the Telegraph model **?**: for all positive t,

$$\mathrm{M}_{\mathrm{Pois}\left(\frac{\nu_1}{\nu_0+\nu_1}\frac{K}{\delta}\right)}(t) \leq \mathrm{M}_{\tilde{p}}(t) \leq \mathrm{M}_{\mathrm{Pois}\left(\frac{K}{\delta}\right)}(t), \tag{16}$$

where M_g denotes the mgf for distribution g. In particular, $\mathrm{M}_{\tilde{p}}(t)$ is bounded above by a Poissonian mgf that does not depend on ν_0 or ν_1. Thus $\tilde{p}(n)$ itself is not heavy tailed, and we require compounding of K or δ to make it so. On the other hand, any extrinsic noise on K or δ that renders the mgf for $\mathrm{Pois}\left(\frac{\nu_1}{\nu_0+\nu_1}\frac{K}{\delta}\right)$ undefined, will also result in $\mathrm{M}_{\tilde{p}}(t)$ being undefined and the resulting compound distribution will be heavy tailed. We now demonstrate this for $K \sim \mathrm{LogNormal}(\mu,\sigma)$. The result relies on the following well-known property of mixture distributions, here interpreted in the context of Eq. (9):

$$\mathrm{M}_{\tilde{q}}(t) = \mathrm{E}_\theta\big(\mathrm{M}_{\tilde{p}(\_;\theta)}(t)\big). \tag{17}$$

It follows from this and Eq. (16) that the compounding integrand is bounded below by

$$\exp\left(\frac{\nu_1}{\nu_1+\nu_0}\frac{K}{\delta}(e^t-1)\right)\frac{1}{\sqrt{\sigma 2\pi}K}\exp\left(-\frac{\ln^2(K-\upsilon)}{2\sigma^2}\right) \tag{18}$$

which diverges to infinity as $K \to \infty$ provided $t > 0$. Thus log-normal extrinsic noise on K renders the compound distribution $\tilde{q}(n;\mu,\sigma)$ heavy tailed; cf. Fig. 3 (c). These results extend to the leaky gene model, with log-normal noise on K_1 (conditional on $K_0 < K_1$), as it is trivial that $K_0 > 0$ only increases the probability of large copy number in comparison to the standard Telegraph model.

From Eq. (16) we see that when $\frac{\nu_1}{\nu_0+\nu_1} = 1$ (that is, $\nu_1 \to \infty$, or $\nu_0 \to 0$) we obtain the constitutive distribution, $\mathrm{Poisson}(K/\delta)$. When K is subject to log-normal noise, the argument using Eq. (18) carries through unchanged.

For bursty expression we require $\nu_0 \gg \nu_1, \delta$ so consider extrinsic noise on K only. The effect of extrinsic noise here is qualitatively different to the other cases: we observe that the mgf for the negative binomial distribution is given by,

$$\mathrm{M}_{\mathrm{nb}}(t) = \left(\frac{r}{1-(1-r)e^t}\right)^{\nu_1/\delta},$$

for $t < -\ln(1-r) = -\ln(K/(K+\nu_0))$, and is infinite otherwise, where $r = \nu_0/(\nu_0+K)$. Thus, the range of positive t for which $\mathrm{M}_{\mathrm{nb}}(t)$ is finite approaches 0 as $K \to \infty$, implying that any unbounded distribution $f(K;\eta)$ results in the moment generating function of the compound distribution

$$\mathrm{M}_{\tilde{q}}(t) = \int_0^\infty \mathrm{M}_{\mathrm{nb}}(t)f(K,\eta)\,dK$$

undefined for positive t. Now in contrast to the previous result, Eq. (11), Gamma distributed noise on K yields a heavy-tailed distribution.

In summary, our approach has provided a range of new analytical solutions to long-standing problems in gene expression modelling. Crucially, we have been able to include extrinsic noise and found that this alone provides complementary and alternative explanations for many empirical observations So et al. (2011). Further to this, we have demonstrated that extrinsic noise can explain observations of heavy-tailed distributions Bengtsson (2005), which intrinsic noise alone cannot. Given the notoriously noisy environment within cells and the intricate organisation of gene regulatory networks, noise extrinsic to a given gene (or gene model) is almost certainly ubiquitous. The framework and results provided here allow us to get a better, more detailed handle on the origins and implications of noise in molecular systems and beyond.

Acknowledgments

References

Arndt Benecke. *HFSP Journal*, 2(4):183–188, 2008. doi: 10.1140/epje/i2005-10068-8.

M. Bengtsson. *Genome Res.*, 15(10):1388–1392, 2005. doi: 10.1101/gr.3820805.

P. C. Bressloff and E. Levien. *Bull. Math. Biol.*, 81(3):800–829, 2019. doi: 10.1007/s11538-018-00538-0.

J. Dattani and M. Barahona. *J. R. Soc. Interface*, 14:20160833, 2016. doi: 10.1098/rsif.2016.0833.

M. B. Elowitz. *Science*, 297(5584):1183–1186, 2002. doi: 10.1126/science.1070919.

Audrey Qiuyan Fu and Lior Pachter. *Stat. Appl. Genet. Mol. Biol.*, 15(6):447–471, 2016. doi: 10.1515/sagmb-2016-0002.

Audrey P. Gasch, Feiqiao Brian Yu, James Hose, Leah E. Escalante, Mike Place, Rhonda Bacher, Jad Kanbar, Doina Ciobanu, Laura Sandor, Igor V. Grigoriev, Christina Kendziorski, Stephen R. Quake, and Megan N. McClean. *PLOS Biol.*, 15(12):e2004050, 2017. doi: 10.1371/journal.pbio.2004050.

R. Grima, D. R. Schmidt, and T. J. Newman. *J. Chem. Phys.*, 137(3):035104, 2012. doi: 10.1063/1.4736721.

P. J. Ingram, M. P. H. Stumpf, and J. Stark. *PLoS Comput. Biol.*, 4(10):e1000192, 2008. doi: 10.1371/journal.pcbi.1000192.

Srividya Iyer-Biswas, F. Hayot, and C. Jayaprakash. *Phys. Rev. E*, 79(3):031911, 2009. doi: 10.1103/PhysRevE.79.031911.

D. Jones and J. Elf. *Curr. Opin. in Microbiol.*, 45:124–130, 2018. doi: 10.1016/j.mib.2018.04.001.

Daniel L. Jones, Robert C. Brewster, and Rob Phillips. *Science*, 346(6216):1533–1536, 2014. doi: 10.1126/science.1255301.

Thomas B. Kepler and Timothy C. Elston. *Biophys. J.*, 81:3116–3136, 2001. doi: 10.1016/S0006-3495(01)75949-8.

O. Lenive, P. D. W. Kirk, and M. P. H. Stumpf. *BMC Syst. Biol.*, 10(1):81, 2016. doi: 10.1186/s12918-016-0324-x.

B. Munsky, G. Neuert, and A. van Oudenaarden. *Science*, 336(6078):183–187, 2012. doi: 10.1126/science.1216379.

Frank W. Olver, Daniel W. Lozier, Ronald F. Boisvert, and Charles W. Clark. *NIST Handbook of Mathematical Functions*. Cambridge University Press, 2010.

J. Paulsson and M. Ehrenberg. *Phys. Rev. Lett.*, 84(23):5447–5450, 2000. doi: 10.1103/PhysRevLett.84.5447.

Jean Peccoud and Bernard Ycart. *Theor. Popul. Biol.*, 48:222–234, 1995. doi: 10.1006/tpbi.1995.1027.

Arjun Raj, Charles S Peskin, Daniel Tranchina, Diana Y Vargas, and Sanjay Tyagi. *PLoS Biol.*, 4(10):e309, 2006. doi: 10.1371/journal.pbio.0040309.

J. M. Raser and E. K. O'Shea. *Science*, 304(5678):1811–1814, 2004. doi: 10.1126/science.1098641.

David Schnoerr, Guido Sanguinetti, and Ramon Grima. *J. Phys. Math. Theor.*, 50(9):093001, 2017. doi: 10.1088/1751-8121/aa54d9.

V. Shahrezaei and P. S. Swain. *Proc. Natl. Acad. Sci.*, 105(45):17256–17261, 2008. doi: 10.1073/pnas.0803850105.

L.-H. So, A. Ghosh, C. Zong, L. A. Sepúlveda, R. Segev, and I. Golding. *Nature Genetics*, 43(6):554–560, 2011. doi: 10.1038/ng.821.

P. S. Swain, M. B. Elowitz, and E. D. Siggia. *Proc. Natl. Acad. Sci.*, 99(20):12795–12800, 2002. doi: 10.1073/pnas.162041399.

Philipp Thomas. *Sci. Rep.*, 9(1):474, 2019. doi: 10.1038/s41598-018-35927-x.

W. Willinger, D. Alderson, J.C. Doyle, and Lun Li. In *Proceedings of the 2004 WSC*, volume 1. IEEE, 2004. doi: 10.1109/WSC.2004.1371310.

A The leaky gene model

A.1 Steady-state solution

In this section, we provide a full derivation of the steady-state solution to the leaky gene model described by the following master equation, given as Eq. 1 in the main text.

$$(\forall n \geq 1)\, \partial_t p_i(n,t) = -(\nu_{i'} + K_i + \delta n)p_i(n,t) + \delta(n+1)p_i(n+1,t) + K_i p_i(n-1,t) + \nu_i p_{i'}(n,t), \quad \text{(S.1)}$$

where $i \in \{0,1\}$, and $i' = 1-i$. For $n=0$, terms involving $n-1$ are set to 0. As mentioned, the development closely follows Peccoud and Ycart (1995). Letting $p_i(n)$ denote the steady-state probability of copy number n in state $i \in \{0,1\}$, we introduce the following generating functions:

$$g_0(z) = \sum_{n=0}^{\infty} z^n p_{0,n}; \qquad g_1(z) = \sum_{n=0}^{\infty} z^n p_{1,n}. \quad \text{(S.2)}$$

Multiplying Equation (S.1) through by z^n and then summing these over n from 0 to ∞ gives rise to the following pair of symmetric first-order differential equations:

$$\delta(z-1)g_0'(z) = -(\nu_1 + K_0(1-z))g_0(z) + \nu_0 g_1(z), \quad \text{(S.3)}$$

and

$$\delta(z-1)g_1'(z) = -(\nu_0 + K_1(1-z))g_1(z) + \nu_1 g_0(z). \quad \text{(S.4)}$$

By differentiation, then back substitution, we arrive at the following symmetric pair of second-order differential equations:

$$\delta^2(z-1)g_0''(z) + \delta(\delta+\nu_0+\nu_1-(K_0+K_1)(z-1))g_0'(z) - (K_0\delta+\nu_0 K_0+\nu_1 K_1-K_0 K_1(z-1))g_0(z) = 0 \quad \text{(S.5)}$$

and

$$\delta^2(z-1)g_1''(z) + \delta(\delta+\nu_0+\nu_1-(K_0+K_1)(z-1))g_1'(z) - (K_1\delta+\nu_0 K_0+\nu_1 K_1-K_0 K_1(z-1))g_1(z) = 0. \quad \text{(S.6)}$$

As in Peccoud and Ycart (1995), the substitution $x = 1-z$ and defining $h_i(x) = g_i(1-x)$, for $i \in \{0,1\}$, leads to conceptually simpler equations

$$\delta^2 x h_0''(x) + \delta(\delta+\nu_0+\nu_1+(K_0+K_1)x)h_0'(x) + (K_0\delta+\nu_0 K_0+\nu_1 K_1+K_0 K_1 x)h_0(x) = 0 \quad \text{(S.7)}$$

and

$$\delta^2 x h_1''(x) + \delta(\delta+\nu_0+\nu_1+(K_0+K_1)x)h_1'(x) + (K_1\delta+\nu_0 K_0+\nu_1 K_1+K_0 K_1 x)h_1(x) = 0. \quad \text{(S.8)}$$

In comparison to Peccoud and Ycart (1995), these equations have an extra term in the $h_i(z)$ coefficient which makes the solution more complicated. Nevertheless, both (S.7) and (S.8) can be represented as extended confluent hypergeometric equations and so have a known general solution. At this point it is notationally convenient to set $\delta = 1$, which is equivalent to scaling time by δ and there is no loss of generality in doing this. Let

$$C = \nu_0 + \nu_1 + 1; \quad D = K_0 + K_1; \quad E_i = \nu_0 K_0 + \nu_1 K_1 + K_i \text{ for } i \in \{0,1\}; \quad F = K_0 K_1.$$

Then from (S.7) and (S.8), we obtain the equations

$$x h_i''(x) + (C + Dx)h_i'(x) + (E + Fx)h_i(x) = 0. \quad \text{(S.9)}$$

Given the initial conditions $h_0(0) = \nu_0/(\nu_0+\nu_1)$ and $h_1(0) = \nu_1/(\nu_0+\nu_1)$, these can be solved as

$$h_i(x) = \frac{\nu_0}{\nu_0+\nu_1}\exp\left(-\frac{x}{2}\left(D+\sqrt{D^2-4F}\right)\right){}_1F_1\left(\frac{CD+C\sqrt{D^2-4F}-2E_i}{2\sqrt{D^2-4F}}, C, x\sqrt{D^2-4F}\right). \quad \text{(S.10)}$$

Note that (S.7) and (S.8) have a second solution in terms of Kummer's second hypergeometric function, however this is not defined at the initial conditions and fails other physical properties of the system; this is detailed also in Grima et al. (2012). We observe that this solution uses the assumption $K_0 \neq K_1$, however the special case of $K_0 = K_1$ is a degenerate form of the system (a standard birth-death model). Thus we assume now that $K_0 < K_1$, while the opposite case $K_1 < K_0$ follows by symmetry. This assumption leads to enormous simplification, because $\sqrt{D^2 - 4F} = |K_0 - K_1|$ and then $D + \sqrt{D^2 - 4F} = 2K_1$. After further cancelling in (S.10), and replacing x by $1 - z$ we obtain

$$g_0(z) = \frac{\nu_0 e^{K_1(z-1)}}{\nu_0 + \nu_1} {}_1F_1\left(\nu_0 + 1, \nu_0 + \nu_1 + 1, w(z-1)\right), \tag{S.11}$$

and

$$g_1(z) = \frac{\nu_1 e^{K_1(z-1)}}{\nu_0 + \nu_1} {}_1F_1\left(\nu_0, \nu_0 + \nu_1 + 1, w(z-1)\right), \tag{S.12}$$

where w denotes $K_0 - K_1$. For the copy number itself, we want $g(z) = g_0(z) + g_1(z)$. Using the functional identity (Olver et al., 2010, §13.3.3) for the confluent hypergeometric function

$$(a - b + 1){}_1F_1(a, b, z) - a{}_1F_1(a + 1, b, z) - (1 - b){}_1F_1(a, b - 1, z) = 0$$

we obtain

$$g(z) = e^{K_1(z-1)} {}_1F_1\left(\nu_0, \nu_0 + \nu_1, w(z-1)\right). \tag{S.13}$$

The probability mass function is then recovered by way of $\tilde{p}(n) = \frac{1}{n!}g^{(n)}(0)$. The general Leibniz rule now gives

$$\tilde{p}(n) = \frac{1}{n!}\sum_{r=0}^{n}\left[\binom{n}{r}K_1^{n-r}e^{-K_1}w^r\frac{\nu_0^{(r)}}{(\nu_0 + \nu_1)^{(r)}}{}_1F_1(\nu_0 + r, \nu_0 + \nu_1 + r, -w)\right] \tag{S.14}$$

A.2   A special case

We now provide a full derivation of the solution to the limiting case of the leaky gene model; refer to the master equation Eq. 5 in the main article, or alternatively see (S.20) below. Recall that this model is justified only when the active transcription is rare enough to be considered instantaneous ($\nu_0 \gg \nu_1$) and the degradation rate is sufficiently small ($\nu_0 \gg \delta$). Consider then an interval of length T in which the gene is active at the higher rate K_1 and in which no degradation occurs. The probability of transcribing M mRNA molecules is given by a Poisson distribution as,

$$p(M; T, K_1) = \frac{(K_1 T)^M e^{-K_1 T}}{M!}. \tag{S.15}$$

The time period T is itself exponentially distributed according to,

$$p(T; \nu_0) = \nu_0 e^{-\nu_0 T}. \tag{S.16}$$

Marginalising (S.15) over T leads to,

$$p(M; K_1, \nu_0) = \frac{\nu_0 K_1^M}{M!}\int_0^\infty T^M e^{-(\nu_0 + K_1)T}\,dT, \tag{S.17}$$

which with the substitution $u = (\nu_0 + K_1)T$ can be solved to yield,

$$p(M; K_1, \nu) = \frac{\nu_0 K_1^M}{(\nu_0 + K_1)^{M+1}}. \tag{S.18}$$

As is already understood, this is the geometric distribution Paulsson and Ehrenberg (2000); Ingram et al. (2008),

$$G(M; r) = r(1 - r)^M, \tag{S.19}$$

where $r = \nu_0/(\nu_0 + K_1)$. Given instantaneous geometric bursts of transcription, the master equation (S.1) may then be rewritten as,

$$\partial_t p(n,t) = -(\nu_1(1-r) + K_0 + \delta n)p(n,t) + \nu_1 \sum_{j=1}^{n} G(j;r)p(n-j,t) + \delta(n+1)p(n+1,t) + K_0 p(n-1,t).$$
(S.20)

The first term in this equation gives fluxes away from $p(n)$, occurring due to bursts of non-zero size or degradation of a single mRNA. The second term gives the total probability of a burst occurring at any lower molecule number and raising the number to n, while the third term gives the probability of degradation reducing the molecule number from $n + 1$. The fourth term is the probability of raising the number to n when inactive. To solve this system in the steady state we introduce the following generating function:

$$g(z) = \sum_{n=0}^{\infty} z^n p(n)$$
(S.21)

Multiplying (S.20) through by z^n and summing over n from 0 to ∞ gives the following equation:

$$\delta(z-1)g'(z) = \nu_1 \left(\sum_{n=0}^{\infty} \left(z^n \sum_{j=0}^{n} G(j;r)p(n-j) \right) - \sum_{n=0}^{\infty} z^n p(n) \right) + K_0(z-1)g(z).$$

which after regrouping becomes

$$\delta(z-1)g'(z) = \nu_1 \left(g(z) \left(\sum_{n=0}^{\infty} G(n;r)z^n \right) - g(z) \right) + K_0(z-1)g(z).$$

The term $\sum_{n=0}^{\infty} G(n;r)z^n$ is the probability generating function for the geometric distribution, which can be written as $\frac{r}{1-z(1-r)}$. Thus we obtain the first-order differential equation

$$g'(z) = \frac{\nu_1}{\delta(z-1)} \left(\frac{r}{1-z(1-r)} - 1 \right) g(z) + \frac{K_0}{\delta} g(z)$$

which simplifies to

$$g'(z) = \left[\frac{\nu_1}{\delta} \left(\frac{(1-r)}{1-z(1-r)} \right) + \frac{K_0}{\delta} \right] g(z).$$
(S.22)

As previously mentioned, this is routinely solved by the integrating factor method (using the initial condition $g(1) = 1$) to give

$$g(z) = \left(\frac{r}{1-z(1-r)} \right)^{\nu_1/\delta} \exp\left(\frac{K_0}{\delta}(z-1) \right)$$
(S.23)

which is the product of the probability generating function for the negative binomial distribution $\text{NegBin}(\nu_1/\delta, r)$ and the Poisson distribution $\text{Pois}(K_0/\delta)$. This gives the following analytical expression for the steady-state solution, presented as Eq. 7 in the main article:

$$\tilde{p}(n) = \frac{1}{n!} \sum_{i=0}^{n} \left[\binom{n}{s} \frac{\Gamma(\nu_1/\delta + (n-s))}{\Gamma(\nu_1/\delta)(n-s)!} (1-r)^{(n-s)} r^{\nu_1/\delta} \frac{e^{-K_0/\delta} K_0^s}{\delta^s} \right].$$
(S.24)

B   Incorporating extrinsic noise

As mentioned in the main article, an interesting compound distribution arises from Eq. 8 when K is distributed according to the distribution $\text{Gamma}(\alpha, \beta)$; see Eq. 10. We now provide the missing details of its derivation. From Eq. 9, the compound distribution $\tilde{q}(n; \alpha, \beta)$ is given by

$$\tilde{q}(n; \alpha, \beta) = \int_0^{\infty} \frac{K_1^n \nu_1^{(n)}}{n!(\nu_0 + \nu_1)^{(n)}} {}_1F_1(\nu_1 + n, \nu_0 + \nu_1 + n, -K_1) \frac{\beta^\alpha}{\Gamma(\alpha)} K^{\alpha-1} e^{-\beta K} \, dK.$$
(S.25)

Using the Laplace transformation (Olver et al., 2010, §13.10.3),

$$\int_0^\infty t^{b-1} e^{-zt} {}_1F_1\left(a,c,-t\right) \mathrm{d}t = z^{-b}\Gamma(b)\, {}_2F_1\left(a,b,c,\frac{-1}{z}\right),$$

the Eq. (S.25) becomes

$$\tilde{q}(n;\alpha,\beta) = \frac{\alpha^{(n)}\nu_1^{(n)}}{n!\beta^n(\nu_0+\nu_1)^{(n)}}\, {}_2F_1\left(\nu_1+n,\alpha+n,\nu_0+\nu_1+n,\frac{-1}{\beta}\right). \qquad (S.26)$$

Now applying the linear transformation (Olver et al., 2010, §15.8.1)

$$_2F_1\left(a,b,c,z\right) = (1-z)^{-a}{}_2F_1\left(a,c-b,c,\frac{z}{z-1}\right)$$

to (S.26), we obtain

$$\tilde{q}(n;\alpha,\beta) = \frac{\alpha^{(n)}\nu_1^{(n)}}{n!\beta^n(\nu_0+\nu_1)^{(n)}}\left(\frac{1}{1+\beta}\right)^n\left(\frac{\beta}{1+\beta}\right)^a\, {}_2F_1\left(\nu_1+n,\nu_0+\nu_1-\alpha,\nu_0+\nu_1+n,\frac{1}{1+\beta}\right), \qquad (S.27)$$

C Existence of moments under extrinsic noise

In the main article we provide simple formulæ for the nth moments of copy number distributions under extrinsic noise on the rate parameter K. The following three subsections are dedicated to obtaining these: the first covers the Telegraph model under extrinsic noise, and the second and third cover constitutive and bursty expression under extrinsic noise, respectively. Throughout, we let $f(K;\eta)$ denote the compounding distribution of K.

C.1 Telegraph model

Let $X = X_K$ denote a random variable with distribution $\tilde{p}(n;K)$ (from Eq. 8) and $Y = Y_\eta$ a random variable with the compound distribution $\tilde{q}(n;K)$ (from Eq. 9). In the following it will be useful to recall that, for real numbers x,y and positive integer n, the notation $x^{(n)_y}$ abbreviates $x(x+y)\ldots(x+y(n-1))$, the rising factorial of x with respect to δ. The notation $x_{(n)}$ abbreviates $x(x-1)\ldots(x-(n-1))$, the falling factorial of x. Our results rely on the following property of mixture distributions.

$$\mathrm{E}(Y^n) = \mathrm{E}_K(\mathrm{E}(X^n)) \qquad (S.28)$$

The proof of (S.28) is straightforward, relying only on switching the order of integration/sum. Following Peccoud and Ycart Peccoud and Ycart (1995), we let

$$e_n := \mathrm{E}(X(X-1)(X-2)\ldots(X-n+1)) \qquad (S.29)$$

and obtain directly from the generating function (Peccoud and Ycart, 1995, Equation 20),

$$e_n = \frac{K^n}{\delta^n}\frac{\nu_1^{(n)_\delta}}{(\nu_0+\nu_1)^{(n)_\delta}}.$$

Before we continue, we recall that Stirling numbers of the first and second kind relate expansions of the falling factorial $x_{(n)} := x(x-1)\ldots(x-n+1)$ to conventional powers of x. Stirling numbers of the first kind are the coefficients in the expansion of $x_{(n)}$, and Stirling numbers of the second kind arise as the coefficients when x^n is written as a linear combination of $x_{(1)},\ldots,x_{(n)}$. Thus, from (S.29) we have that $e_n = \sum_{k=0}^n s(n,k)\mathrm{E}(X^k)$, where $s(n,k)$ are Stirling numbers of the first kind. We obtain

$$\mathrm{E}(X^n) = e_n - \sum_{k=0}^{n-1} s(n,k)\,\mathrm{E}(X^k) = \frac{K^n}{\delta^n}\frac{\nu_1^{(n)_\delta}}{(\nu_0+\nu_1)^{(n)_\delta}} - \sum_{k=0}^{n-1} s(n,k)\,\mathrm{E}(X^k).$$

Thus $E(X^n)$, as a function of K, is a polynomial of degree n, which we can write as

$$
\begin{aligned}
E(X^n) &= \sum_{i=1}^n S(n,i)\, e_i \\
&= \sum_{i=1}^n S(n,i) \frac{\nu_1^{(i)\delta}}{\delta^i (\nu_0 + \nu_1)^{(i)\delta}} K^i,
\end{aligned}
\tag{S.30}
$$

where $S(n,i)$ is a Stirling number of the second kind. Returning to (S.28), we obtain

$$
E(Y^n) = \sum_{i=1}^n S(n,i) \frac{\nu_1^{(i)\delta}}{\delta^i (\nu_0 + \nu_1)^{(i)\delta}} E(K^i).
\tag{S.31}
$$

Noting that $S(1,1) = S(2,1) = S(2,2) = 1$, it follows from (S.31) that

$$
E(Y) = \frac{\nu_1}{\delta(\nu_0 + \nu_1)} E(K) \quad \text{and}
$$

$$
\mathrm{Var}(Y) = \frac{\nu_1(\nu_1 + \delta)}{\delta^2(\nu_0 + \nu_1)(\nu_0 + \nu_1 + \delta)} E(K^2) + \frac{\nu_1}{\delta(\nu_0 + \nu_1)} E(K) - \left(\frac{\nu_1}{\delta(\nu_0 + \nu_1)} E(K) \right)^2.
$$

This then yields the following formula for the Fano factor of $\tilde{q}(n; K)$:

$$
\mathrm{FF}(Y) = 1 - \frac{\nu_1}{\delta(\nu_0 + \nu_1)} E(K) + \frac{\nu_1 + \delta}{\delta(\nu_0 + \nu_1 + \delta)} \frac{E(K^2)}{E(K)},
\tag{S.32}
$$

which is presented as Eq. 13 in the main article.

C.2 Constitutive expression

In a similar way to the Telegraph model, we can obtain closed-form formulæ for the nth moments and Fano factor in the case for constitutive expression with extrinsic noise on K. Let $X = X_k$ be a random variable with distribution $\mathrm{Pois}(K/\delta)$ distribution, let $\tilde{q}_{\mathrm{Pois}}(n; \eta)$ denote the compound distribution obtained by compounding $\mathrm{Pois}(K/\delta)$ by $f(K; \eta)$, and let $Y = Y_\eta$ denote a random variable for $\tilde{q}_{\mathrm{Pois}}(n; \eta)$. The nth moments of a $\mathrm{Pois}(K/\delta)$ random variable are given by

$$
E(X^n) = \sum_{k=1}^n S(n,k) \left(\frac{K}{\delta} \right)^k,
\tag{S.33}
$$

where $S(n,k)$ is a Stirling number of the second kind. From (S.28), it follows immediately that

$$
E(Y^n) = \sum_{k=1}^n \frac{S(n,k)}{\delta^k} E(K^k).
\tag{S.34}
$$

Thus,

$$
E(Y) = \frac{E(K)}{\delta},
$$

$$
\mathrm{Var}(Y) = \frac{E(K)}{\delta} + \frac{E(K^2)}{\delta^2} - \left(\frac{E(K)}{\delta} \right)^2, \quad \text{and}
$$

$$
\mathrm{FF}(Y) = 1 + \frac{1}{\delta} \frac{\mathrm{Var}(K)}{E(K)}.
\tag{S.35}
$$

C.3 Bursty expression

Let $X = X_K$ denote a random variable with distribution the $\mathrm{NegBin}(\nu_1/\delta, r)$ distribution, where r is equal to $\nu_0/(\nu_0 + K)$. Let $\tilde{q}_{\mathrm{nb}}(n; \eta)$ denote the compound distribution obtained by compounding $\mathrm{NegBin}(\nu_1/\delta, r)$ by

$f(K; \eta)$, and let $Y = Y_\eta$ denote a random variable for $\tilde{q}_{nb}(n; \eta)$. Recall that the moment generating function for the $NegBin(\nu_1/\delta, r)$ distribution is given by

$$M_{nb}(t) = \left(\frac{r}{1 - (1 - r)e^t} \right)^{\nu_1/\delta}$$

for $t < -\ln(1 - r) = -\ln(K/(K + \nu_0))$, and is infinite otherwise. Using the recurrence relation,

$$S(n + 1, k) = kS(n, k) + S(n, k - 1) \tag{S.36}$$

a straightforward proof by induction on the number n of derivatives of $M_{nb}(t)$ with respect to t (with base case $n = 0$ given by $M_{nb}(t)$) shows that

$$\frac{d^n M_{nb}(t)}{dt^n} = r^{\frac{\nu_1}{\delta}} \sum_{k=1}^n S(n, k) \left(\frac{\nu_1}{\delta} \right)^{(k)} q^k (1 - qe^t)^{\frac{\nu_1}{\delta} - k} e^{kt}, \tag{S.37}$$

where $q = 1 - r$. By substituting $t = 0$ into (S.37), we immediately obtain

$$E(X^n) = \sum_{k=1}^n S(n, k) \left(\frac{\nu_1}{\delta} \right)^{(k)} \left(\frac{q}{r} \right)^k,$$

which simplifies to

$$E(X^n) = \sum_{k=1}^n S(n, k) \left(\frac{\nu_1}{\delta} \right)^{(k)} \left(\frac{K}{\nu_0} \right)^k. \tag{S.38}$$

It then follows from (S.28) that

$$E(Y^n) = \sum_{k=1}^n S(n, k) \left(\frac{\nu_1}{\delta} \right)^{(k)} \nu_0^{-k} E(K^k), \tag{S.39}$$

which gives the following formula for the Fano factor of $\tilde{q}_{nb}(n; \eta)$:

$$FF(Y) = 1 - \frac{\nu_1}{\delta \nu_0} E(K) + \frac{1}{\nu_0} \left(\frac{\nu_1}{\delta} + 1 \right) \frac{E(K^2)}{E(K)}. \tag{S.40}$$

D   Heavy-tailedness of the compound distribution

In the final section of the main article we showed that heavy-tailedness in the copy number distribution is a potential qualitative identifier of extrinsic noise. The arguments rely on the following inequality for the Telegraph model, given as Eq. 16 in the main article: for all positive t,

$$M_{Pois\left(\frac{\nu_1}{\nu_0 + \nu_1} \frac{K}{\delta} \right)}(t) \leq M_{\tilde{p}}(t) \leq M_{Pois\left(\frac{K}{\delta} \right)}(t), \tag{S.41}$$

where M_g denotes the mgf for distribution g. We now give a detailed proof of this result. In the following, let $X_1 \sim Poisson\left(\frac{\nu_1}{\nu_0 + \nu_1} \frac{K}{\delta} \right)$, let $X_2 \sim Poisson(\frac{K}{\delta})$, and let $X_3 \sim \tilde{p}(n)$ (from Eq. 8). For $n \in \mathbb{N}$, we have

$$
\begin{aligned}
E(X_1^n) &= \sum_{i=1}^n S(n, i) \left(\frac{\nu_1}{\nu_0 + \nu_1} \right)^i \left(\frac{K}{\delta} \right)^i \\
&\leq \sum_{i=1}^n S(n, i) \left(\frac{\nu_1^{(i)_\delta}}{(\nu_0 + \nu_1)^{(i)_\delta}} \right) \left(\frac{K}{\delta} \right)^i \text{ as } \frac{\nu_1 + i\delta}{\nu_1 + \nu_0 + i\delta} \leq \frac{\nu_1}{\nu_0 + \nu_1} \, (\forall i) \\
&= E(X_3^n) \text{ (from (S.30))} \\
&\leq \sum_{i=1}^n S(n, i) \left(\frac{K}{\delta} \right)^i \text{ as } \frac{\nu_1^{(i)_\delta}}{(\nu_0 + \nu_1)^{(i)_\delta}} \leq 1 \, (\forall i) \\
&= E(X_2^n)
\end{aligned}
$$

As each moment $E(X_3^n)$ is sandwiched between two Poissonian n^{th} moments, it immediately follows from the least upper bound property that the mgf for the copy number distribution $\tilde{p}(n)$ is bounded below by the mgf for the Pois $\left(\frac{\nu_1}{\nu_0+\nu_1}\frac{K}{\delta}\right)$ distribution and bounded above by the mgf for the Pois$\left(\frac{K}{\delta}\right)$ distribution. Thus, Equation (S.41) holds.