Is precise partition volume (monodispersity) required for good results in dPCR? Actually, no!

by

Joel Tellinghuisen
Department of Chemistry
Vanderbilt University
Nashville, TN 37235

4BIO Summit — San Francisco, September 13, 2018

Binomial statistics governs dPCR outcome

- Only two possible outcomes: success (probability p) and failure (probability q); $p + q = 1$. Consider N trials.
- Mean $\theta = Np$; variance $\sigma^2 = Npq$.
- Let p be probability for a null, and N_0 the number of nulls. Then $p \approx N_0/N$.

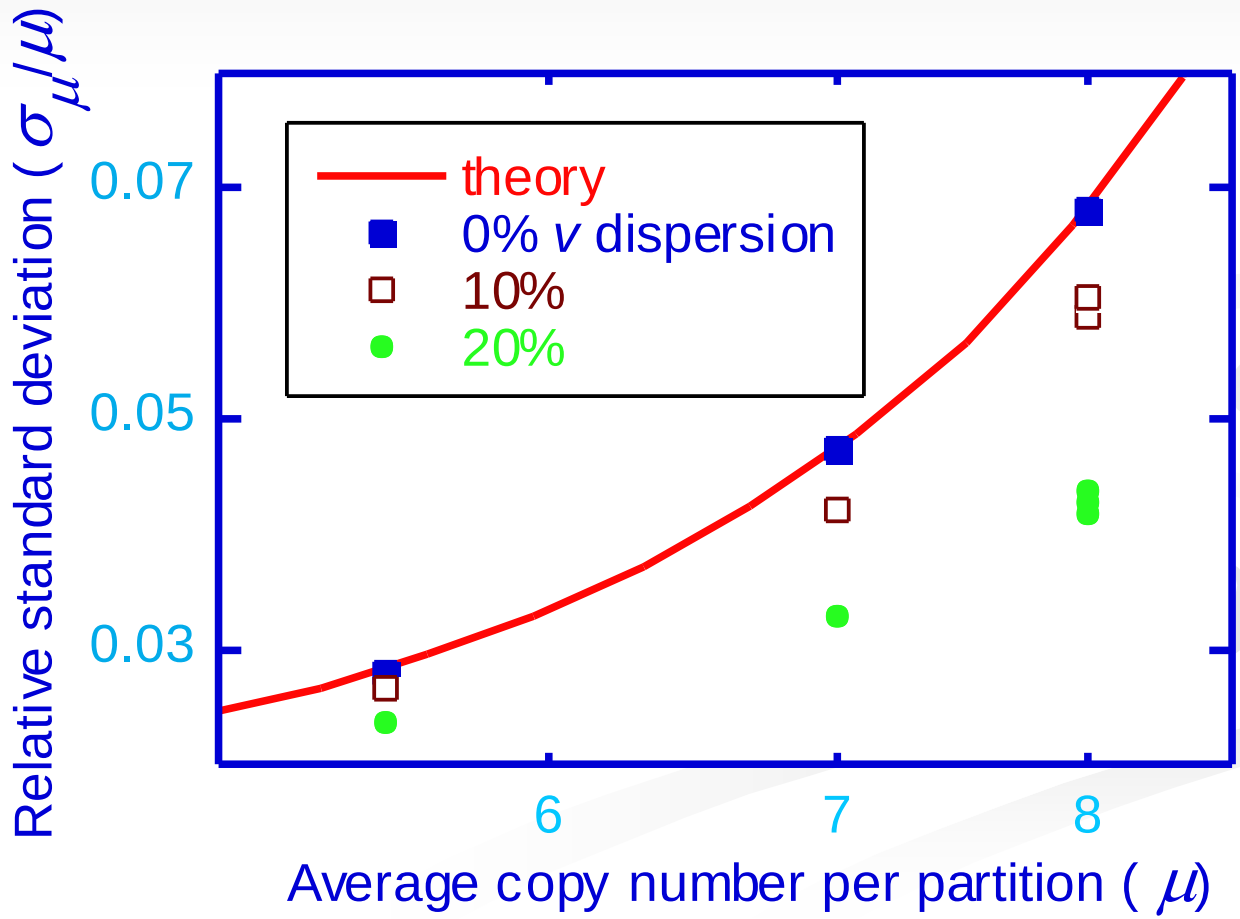
- Also, with fixed partition volume $p = e^{-\mu}$, where μ is the average copy number per partition (Poisson). Thus, $\mu = \ln(N/N_0)$. More complex with variable volume.

Monte Carlo simulations

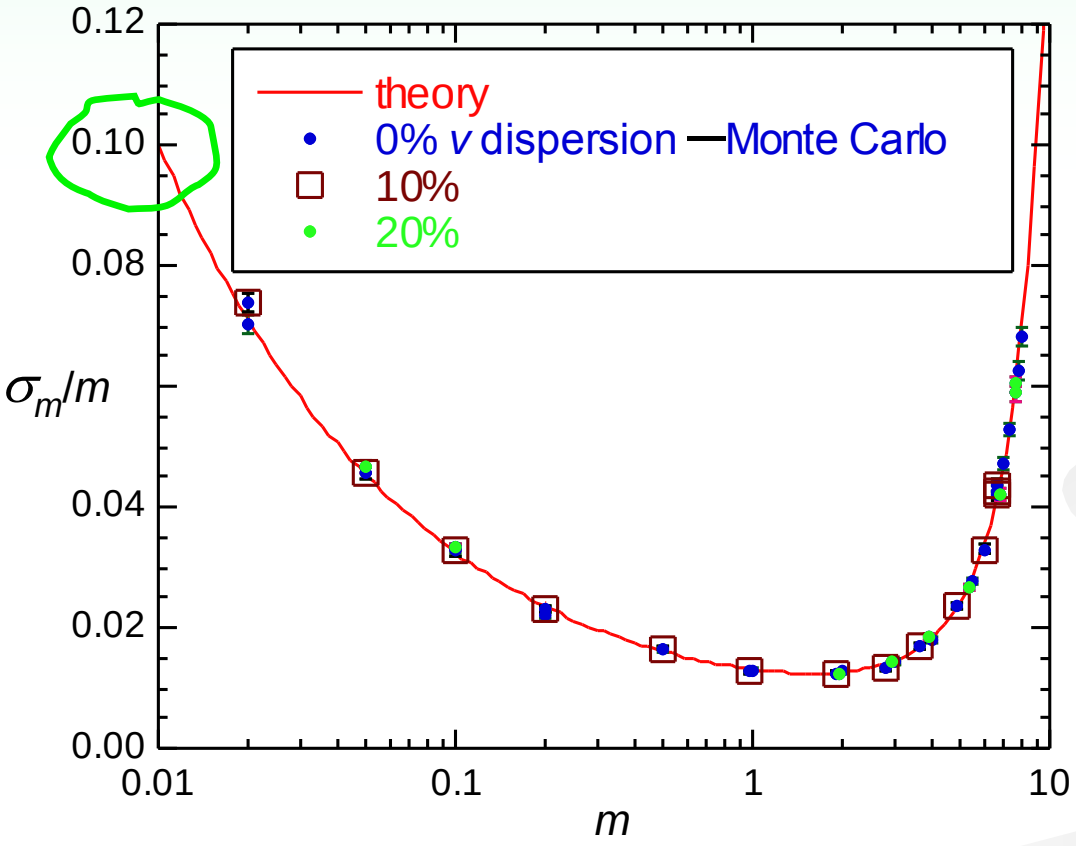
- Use simple random # (RAN) generator over interval 0-1. For constant v, RAN < $p \rightarrow$ null partition. Repeat for $N = 10^4$ partitions & accumulate statistics.
- For n such simulations, the MC standard error in N_0 is $\sigma_{N_0}/n^{1/2}$ and for estimated SDs, the relative SE is $\sim(2n)^{-1/2}$ (from properties of χ^2).
- For variable v w/ average $v = 1$, have $p(v) = e^{-\mu v}$. One RAN is used to pick v from the distribution $f(v)$ (*e.g.* Box-Muller algorithm for normal $f(v)$). Second RAN designates as null or hit, using $p(v)$.
- Now $m = \ln(N/N_0)$ is a **biased** estimator of μ.
- But precision of m is still **strictly binomial**.

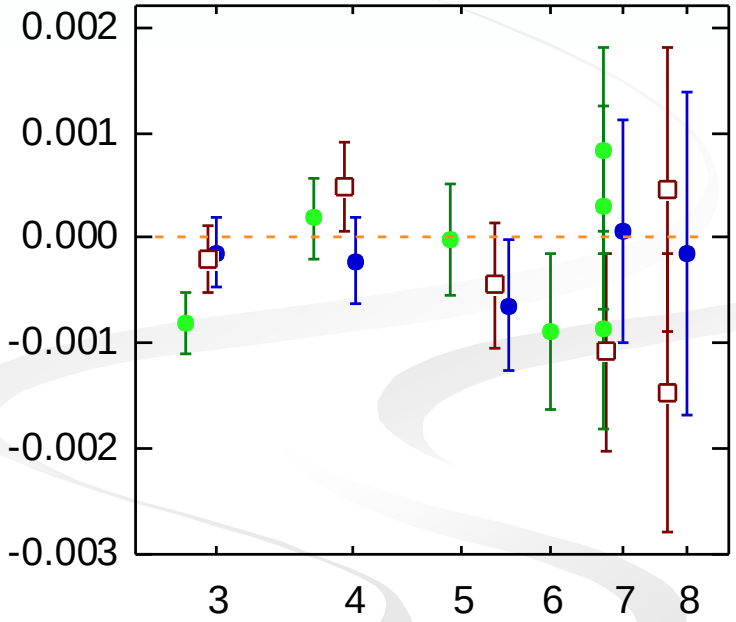
First glance: Polydispersity seems to IMPROVE precision!

But … the bias …



The bias in estimate *m* of true mean μ increases with partition volume dispersion. But *m* rigorously follows binomial statistics.*

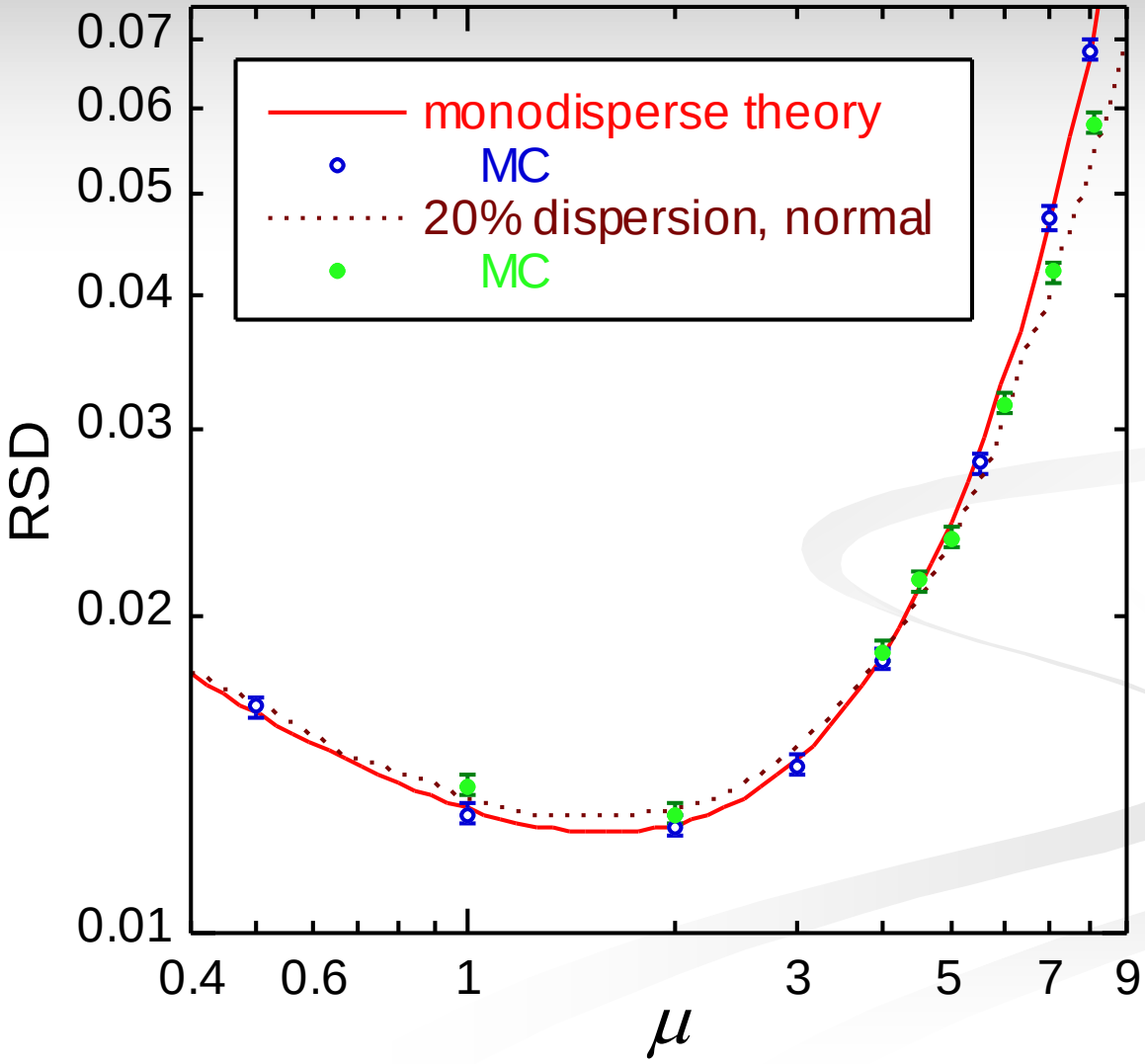




*Anal. Chem. **2016**, *88*, 12183-12187.

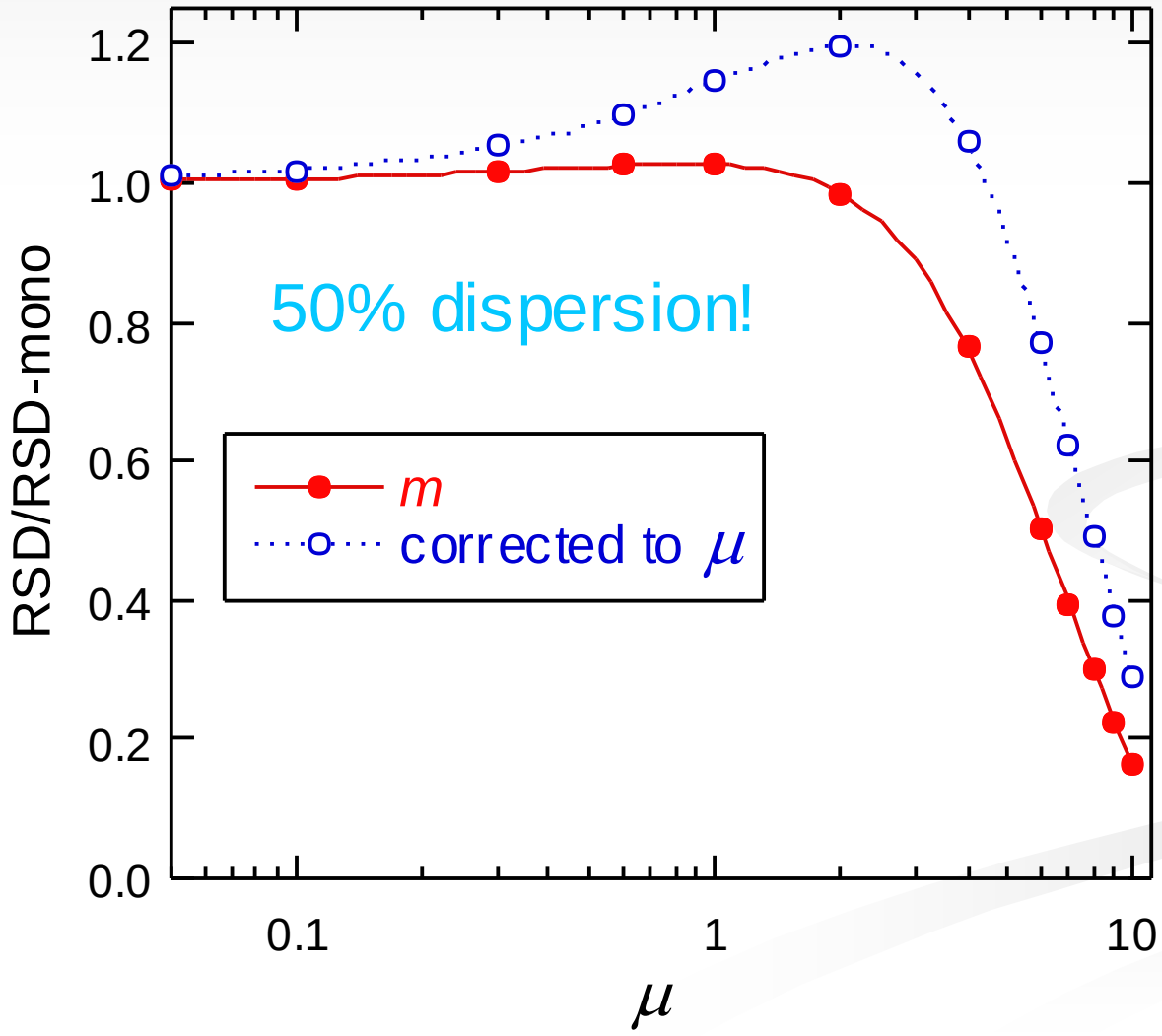
Correcting for bias decreases precision …

but for $\mu > 4$, precision still better for disperse v.



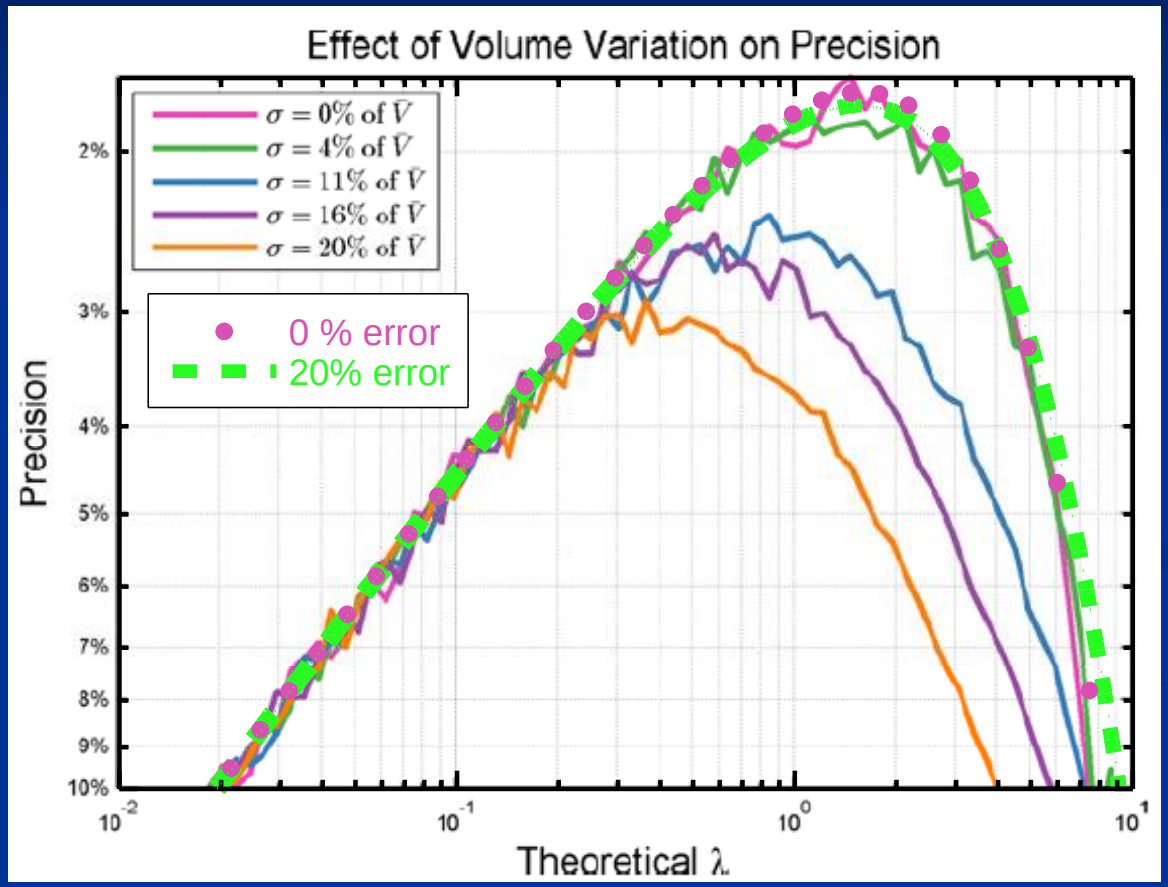
Displayed another way …

50% *v* dispersion ⟶ max. 20% precision loss!



Results obtained using **gamma distribution** for partition *v*; in full agreement with Supplement of Huggett, *et al.*, *Clin.Chem.* 2015, who displayed just range $\mu = 0.5\text{-}3$.

But ***not*** with Fig. 8 in Majumdar, *et al.* (2015),* reproduced as Fig. 1 by same team in 2017.†

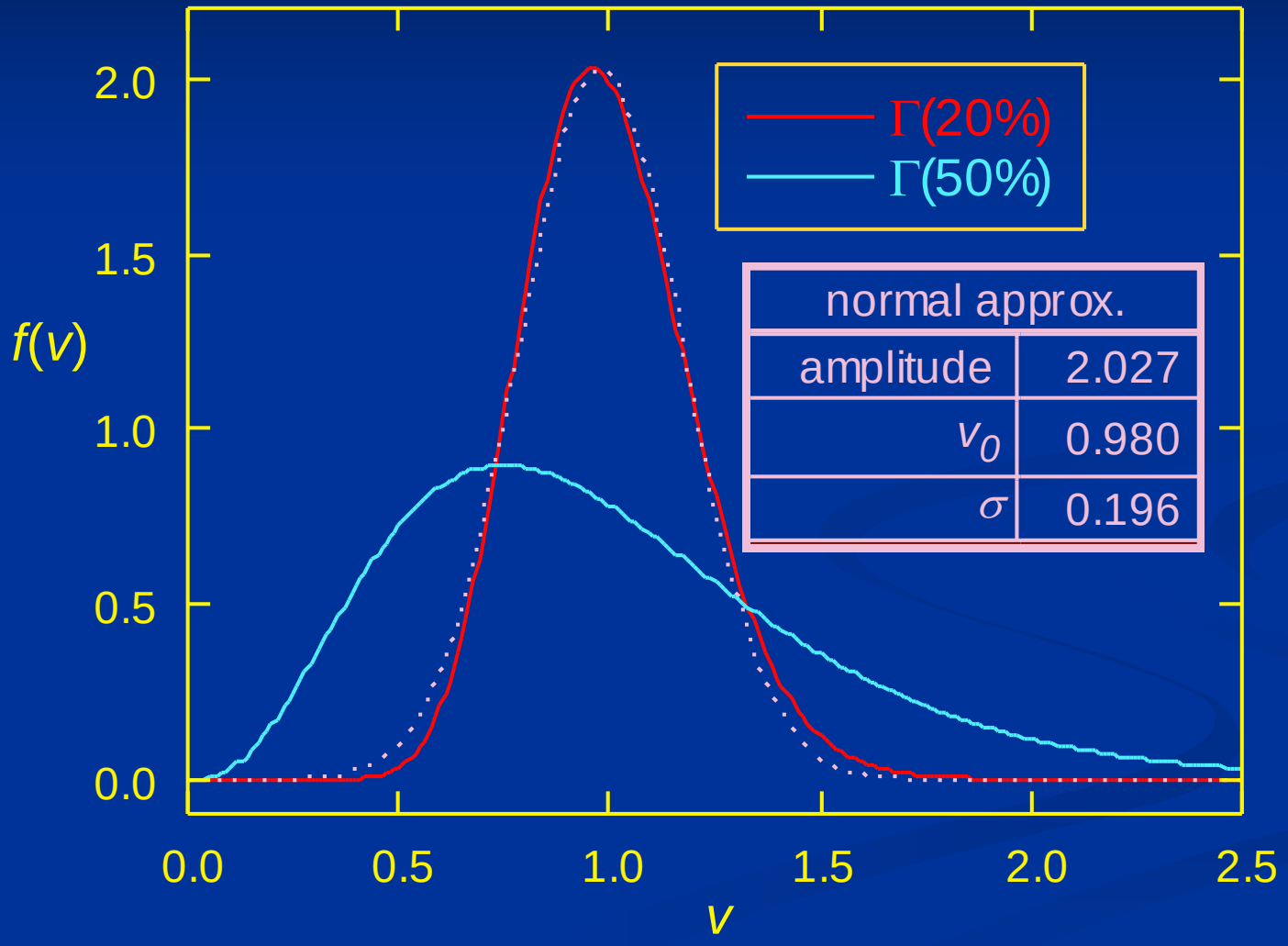


However, in 2017 paper, authors show apparent agreement with Huggett, *et al.* in other figures, up to high μ, where different $f(v)$ matters.

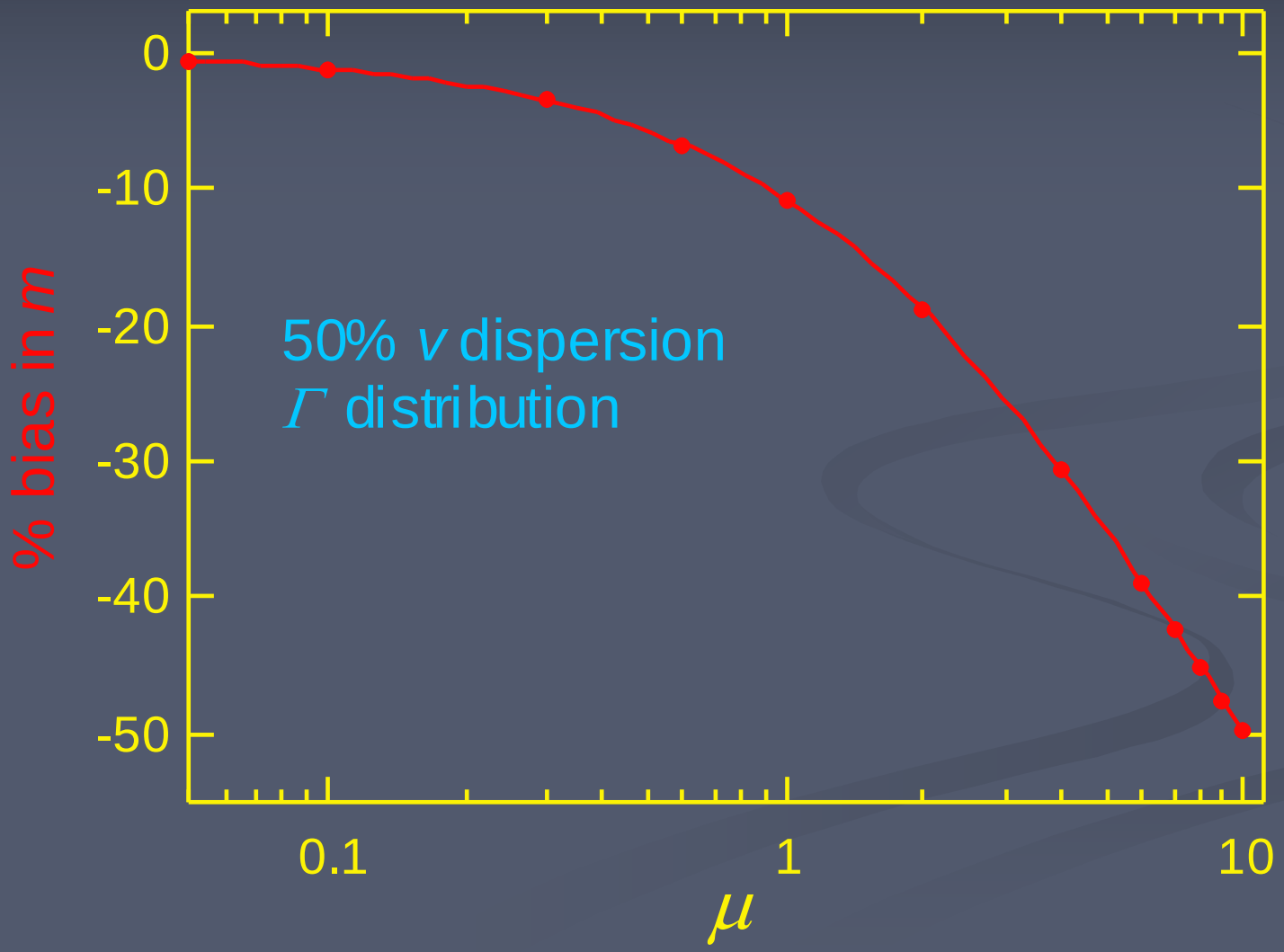
*Majumdar, Wessel, Marks, *PLoS One* **2015**, *10*, e0118833.

†Majumdar, *et al.*, *Sci. Repts.* **2017**, *7*:9617.

Gamma distribution



Bias much more important than imprecision



Probability Distributions

Uniform: $P(x)$ = constant ($a \le x \le b$); 0 otherwise

Normal:
$$P_G(\mu,\sigma;x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

Gamma:
$$\Gamma(x;\alpha,\beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$$

Γ has mean α/β and variance α/β^2, hence RSD $\alpha^{-1/2}$; *e.g.* $\alpha = \beta = 25$ gives mean 1 and $\sigma = 0.2$, while $\alpha = \beta = 4$ gives 50% dispersion.

Expand binomial treatment of dPCR – $f(v) = v_0$

Take "success" as no DNA, giving N_0 null partitions.

- Poisson stats apply for each partition; probability of a null partition is $p = P(0) = e^{-\mu}$ for constant-v partitions, where μ is the average number of copies per partition.
- Hence μ can be estimated from the fraction $N_0/N \approx p$ of null partitions, giving $\mu = \ln(N/N_0)$.
- Binomial theory SD in N_0 is $[N_0(1 - N_0/N)]^{1/2}$, giving

$$\sigma_\mu = \frac{\sigma_{N_0}}{N_0} = \left[\frac{1}{N_0} - \frac{1}{N}\right]^{1/2} = \left[\frac{e^{\mu} - 1}{N}\right]^{1/2}$$

- Relative SD (RSD) in μ is σ_μ/μ, which has minimum value (best % precision) near $\mu = 1.59$.

Next consider variable partition volume

Let μ apply for $v = v_0$. Then $M(v) = \mu v/v_0$ is average copy number for volume v. Take $v_0 = 1$ for simplicity.

- Now p = average over partition v, distribution $f(v)$

  $$\langle P(0) \rangle = \int_0^{v_{\max}} \mathrm{e}^{-\mu v} f(v) \mathrm{d}v$$

- *E.g.*, if $f(v)$ is uniform distribution, range $1 \pm s$ ($s<1$),

  $p = \mathrm{e}^{-\mu} \sinh(\mu s)$, giving $m_{\mathrm{uni}} = -\ln p \approx \mu - (\mu s)^2/6$.

- If $f(v)$ is normal distribution, centered at 1 w/ $\sigma = s$,

  $m_{\mathrm{norm}} \approx \mu - (\mu s)^2/2$, roughly valid for $s < 0.25$.

Gamma Distribution

- Γ goes to zero at $x = 0$ and is approximately normal for small s. Obtain

$$m_\Gamma = \frac{1}{s^2}\ln(1+\mu s^2) \approx \mu - \frac{\mu^2 s^2}{2}$$

(same as for normal dist.)

and

$$\sigma_\mu = (1+\mu s^2)\sigma_m$$

- These expressions are different versions of those given in the Supplement in Huggett, *et al*., *Clin.Chem.* 2015.
- Since m is negatively biased, its RSD is < that for μ at large μ, greater at small (from binomial stats).
- Solve for μ in terms of m and apply error propagation to $\mu(m) \rightarrow$ increased uncertainty in resulting estimate of μ.
- [Here error propagation is just $\sigma_\mu^2 = (\partial\mu/\partial m)^2\ \sigma_m^2$.]

Must characterize $f(v)$ to correct for bias.

Several groups have used optical microscopy:

- Pinheiro, *et al.*[1] measured >1100 droplets from 16 wells. Re-examination shows effects not random, but overall 3% SD means ~negligible estimation bias.
- Bhat, *et al.*[2] thought v dispersion a major contributor to overall 6% uncertainty.
- Corbisier, *et al.*[3] — results similar to Pinheiro.[1]
- Dangla *et al.*[4] — droplet radii ~3% RSD → 9% in v, hence possibly significant for bias.
- Košir, *et al.*[5] — interlaboratory comparisons.

1. *Anal. Chem.* **2012**, *84*, 1003.
2. *Anal. Bioanal. Chem.* **2009**, *394*, 457.
3. *Anal. Bioanal. Chem.* **2015**, *407*, 1831.
4. *Proc. Nat. Acad. Sci USA* **2013**, *110*, 853
5. *Anal. Bioanal. Chem.* **2017**, *409*, 6689.

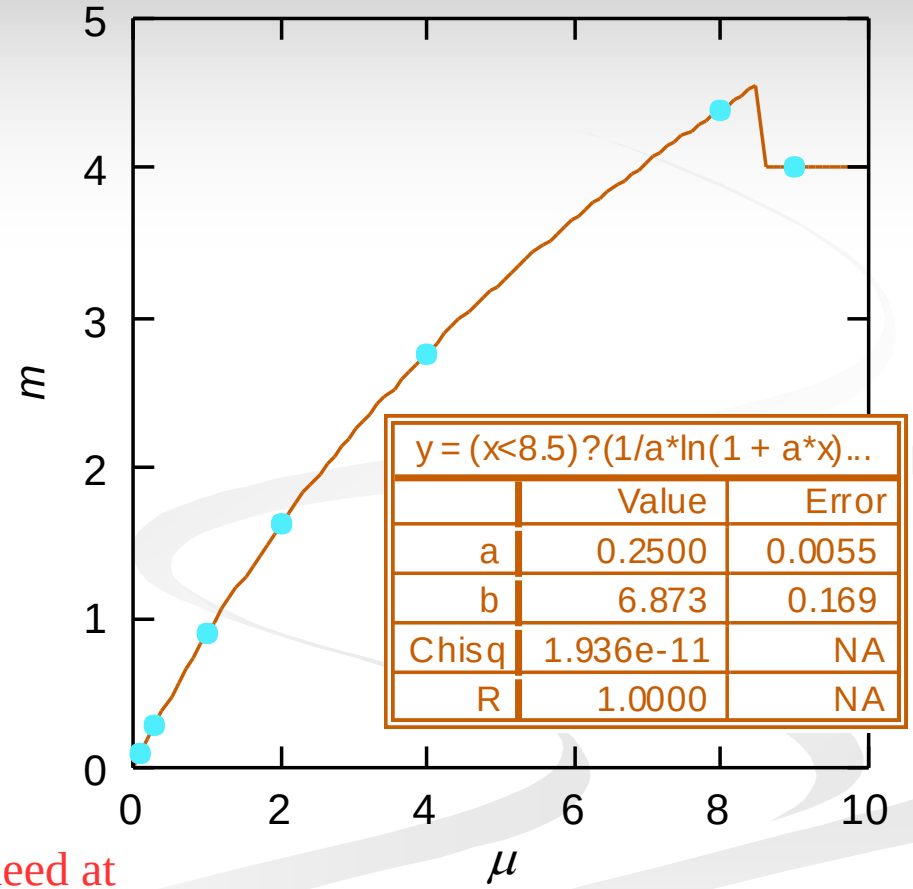
Best option … pending availability of suitable reference materials — **calibration**.

Suppose $f(v)$ is gamma-distributed. Then

$$m_\Gamma = \frac{1}{s^2}\ln(1+\mu s^2)$$

becomes a 1-parameter (s^2) calibration relation.

Determination of μ **and** its SE are straightforward, even if additional calibration parameters are needed.*



likely need at least one for μ.

* *Meth. Enzymol.* **2009**, *454*, 259-285.

Digression: Poisson stats rule in small-μ limit

- Say 10% precision $\rightarrow \mu \approx m = 0.01$; assume 10^4 partitions, each 1 nL. $N^+ \approx \mu N = 100$ = # copies in 10 μL.
- What if drop v to 0.1 nL? Now same # in 1 μL, *i.e.* need higher concentration!
- So increase N to $10^5 \rightarrow V = 10$ μL again. Now get 10% precision with $\mu = 0.001 \rightarrow 100$ copies again!
- **Poisson** stats at work here: $\sigma^2 = N^+$ (= # counts). From $\sigma_\mu = [(e^\mu - 1)/N]^{1/2} \approx (\mu/N)^{1/2} \approx (N^+/N^2)^{1/2}$. Hence $\sigma_\mu / \mu = (N/N^+)\,(N^+/N^2)^{1/2} = 1/\sqrt{N^+}$.
- Poisson stats also limits precision of sampling, *i.e.*, sample containing nominal 100 copies actually 100±10.

SUMMARY

Polydispersity a bias problem, not a precision problem.

Do need to characterize partition volume distribution; and it must not vary run to run.

Probably best done through calibration.

Relaxation of monodispersity demands can greatly reduce cost and increase throughput.

This is already being exploited experimentally, *e.g.*, Byrnes, Chang, Huynh, Astashkina, Weigl, and Nichols, *Anal. Chem.* **2018**, *90*, 9374 (see Poster 5).

Thanks!

to my collaborator, Andrej-Nikolai Spiess, for kindling my interest in PCR methods.

to the conference organizers, for this opportunity.

to you, for your interest and attendance.