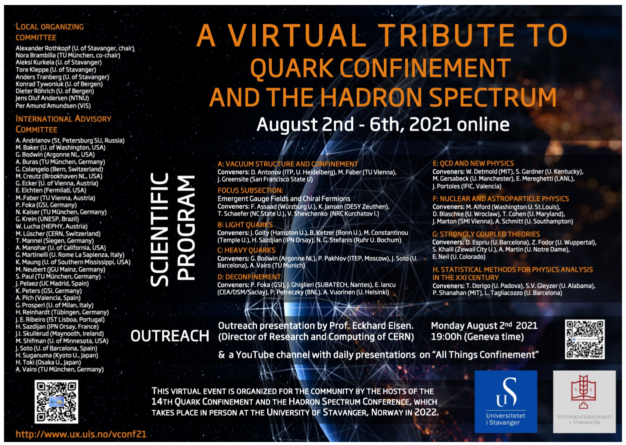
Effect of Systematic Uncertainty Estimation on the Muon $g-2$ Anomaly



A Virtual Tribute to Quark Confinement and the Hadron Spectrum

Stavanger / Online / 2-6 August 2021

https://indico.uis.no/event/12/

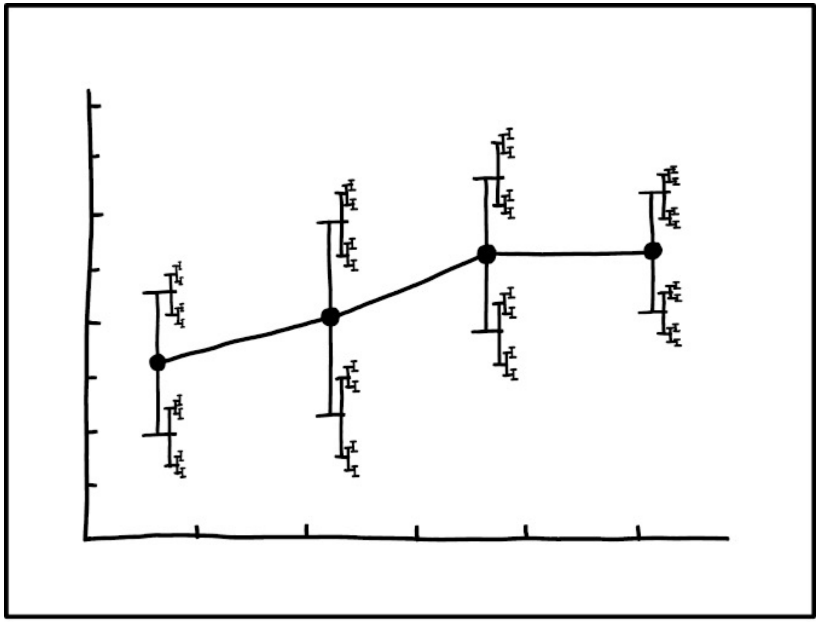
Glen Cowan
Physics Department
Royal Holloway, University of London
g.cowan@rhul.ac.uk
www.pp.rhul.ac.uk/~cowan

Motivation

The uncertainties on estimated systematic errors ("errors on errors") can in general play an important role in HEP analyses, see:

G. Cowan, *Statistical Models with Uncertain Error Parameters*, Eur. Phys. J. C (2019) 79:133, arXiv:1809.05778

https://xkcd.com/2110/



I DON'T KNOW HOW TO PROPAGATE ERROR CORRECTLY, SO I JUST PUT ERROR BARS ON ALL MY ERROR BARS.

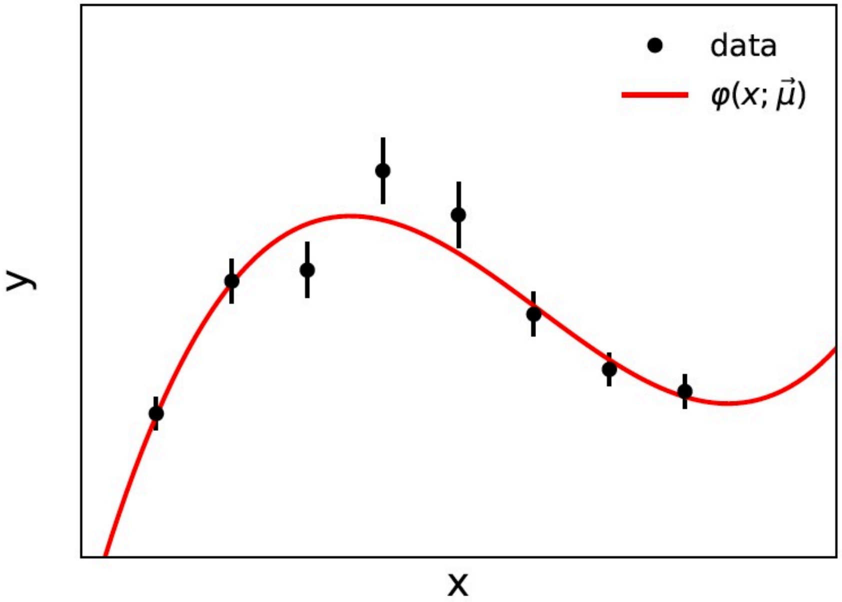
It turns out that models that use errors on errors have qualitatively new, interesting, desirable features:

Sensitivity to outliers reduced.

Confidence intervals sensitive to goodness of fit.

Effect on goodness of fit, p-values, significance.

Prototype example: curve fitting, averages



Suppose independent $y_i \sim$ Gauss, $i = 1,\ldots,N$, with

$$E[y_i] = \varphi(x_i; \boldsymbol{\mu})$$

$$V[y_i] = \sigma_i^2$$

$\boldsymbol{\mu} = (\mu_1,\ldots, \mu_M)$ are M parameters in the fit function $\varphi(x;\boldsymbol{\mu})$.

If we take the σ_i as known, we have the usual log-likelihood

$$\ln L(\boldsymbol{\mu}) = -\frac{1}{2}\sum_{i=1}^{N} \frac{(y_i - \varphi(x_i;\boldsymbol{\mu}))^2}{\sigma_i^2}$$

which leads to the Least Squares estimators for $\boldsymbol{\mu}$.

Goodness of fit for Least Squares

In the least-squares approach, the statistic

$$q = -2 \ln \frac{L(\hat{\boldsymbol{\mu}})}{L(\hat{\boldsymbol{\varphi}})} = \sum_{i=1}^{N} \frac{(y_i - \varphi(x_i; \hat{\boldsymbol{\mu}}))^2}{\sigma_i^2}$$

Likelihood of saturated model $L(\varphi_1,\ldots,\varphi_N)$

provides a measure of goodness of fit. The p-value of the composite hypothesis $\varphi(x_i;\boldsymbol{\mu})$ is

$$p = \int_{q_{\rm obs}}^{\infty} f(q)\, dq$$

If the $y_i \sim \text{Gauss}(\varphi(x_i;\boldsymbol{\mu}), \sigma_i)$ then $f(q)$ is chi-squared for $N-M$ degrees of freedom, independent of $\boldsymbol{\mu}$ (Wilks).

Special case: $\varphi(x_i;\boldsymbol{\mu}) = \mu$, i.e., test if the y_i have a common mean μ

$\rightarrow$ $q \sim$ chi-square($N-1$) $\rightarrow$ $p = 1 - F_{\chi^2_{N-1}}(q)$

What if the σ_i are not known?

The LS approach assumes that the standard deviations σ_i of the measurements are known.

σ_i = statistical error, usually well estimated from sample size.

σ_i = systematic error:

related to stat. error of control measurement – well estimated

related to size of MC event sample – well estimated

systematic uncertainty from modelling of experiment – could be poorly estimated

reflects uncertainty resulting from some mathematical approximation (theory error) – could be poorly estimated

In general, we should allow that the σ_i may not be exactly known.

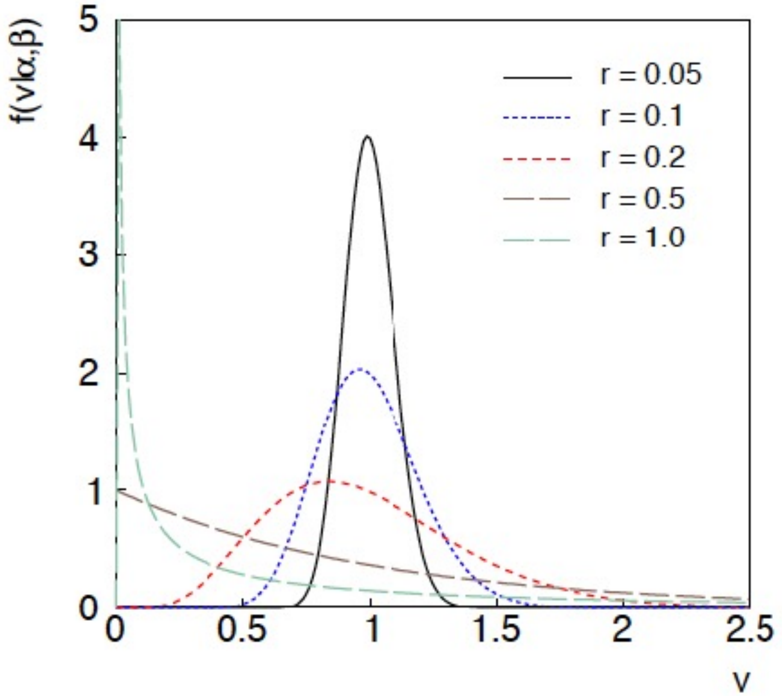
Gamma Variance Model

G. Cowan, EPJC (2019) 79:133

If the σ_i^2 are uncertain, we can take them as adjustable parameters.

The estimated variances $v_i = s_i^2$ are modeled as gamma distributed.

The likelihood becomes



$$L(\boldsymbol{\mu}, \boldsymbol{\sigma^2}) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-(y_i - \varphi(x_i;\boldsymbol{\mu}))^2/2\sigma_i^2} \frac{\beta_i^{\alpha_i}}{\Gamma(\alpha_i)} v_i^{\alpha_i - 1} e^{-\beta_i v_i}$$

Want $\quad E[v_i] = \sigma_i^2 \qquad r_i = \frac{\sigma_{v_i}}{2E[v_i]} \approx \frac{\sigma_{s_i}}{E[s_i]}$

$\rightarrow \quad \alpha_i = \frac{1}{4r_i^2} \qquad \beta_i = \frac{\alpha_i}{\sigma_i^2}$

Profile log-likelihood

One can profile over the σ_i^2 in close form.

The log-profile-likelihood is

$$\ln L'(\boldsymbol{\mu}) = \ln L(\boldsymbol{\mu}, \widehat{\widehat{\boldsymbol{\sigma}^2}}) = -\frac{1}{2} \sum_{i=1}^{N} \left(1 + \frac{1}{2r_i^2}\right) \ln \left[1 + 2r_i^2 \frac{(y_i - \varphi(x_i; \boldsymbol{\mu}))^2}{v_i}\right]$$

Quadratic terms replace by sum of logs.

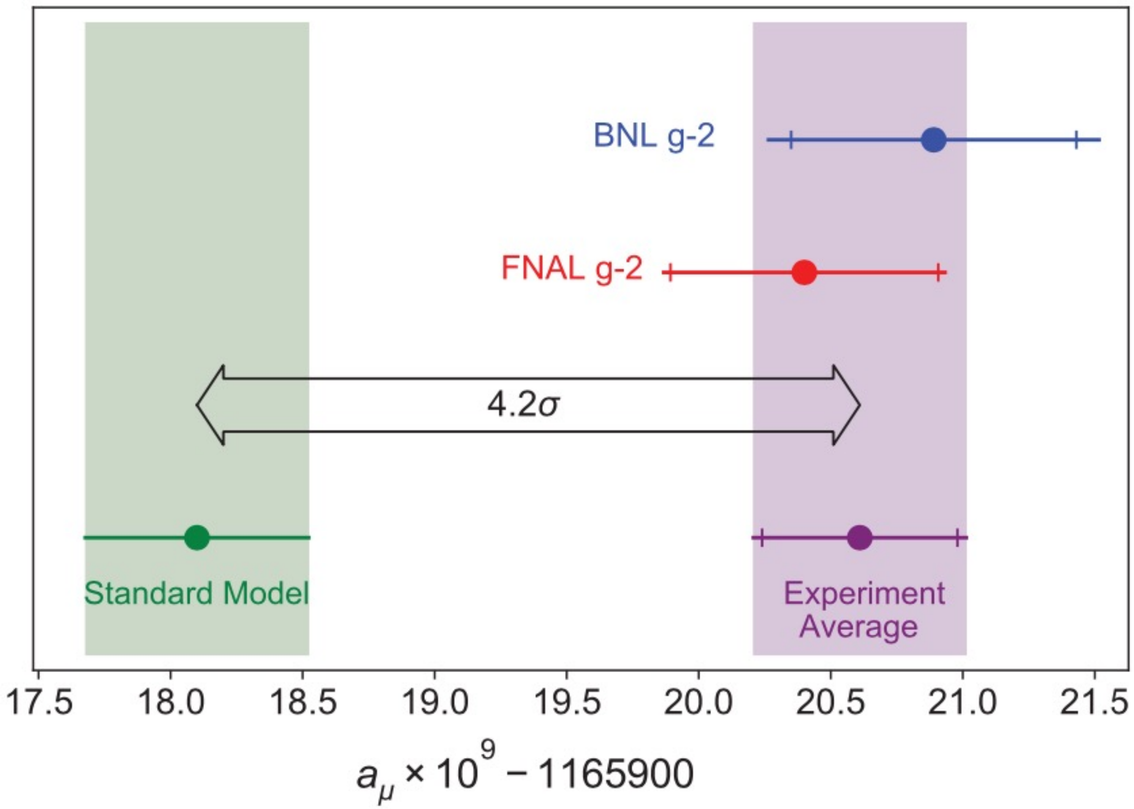
Equivalent to replacing Gauss pdf for y_i by Student's t, $\nu_{\mathrm{dof}} = 1/2r_i^2$

Simple program for Student's t average: stave.py
http://www.pp.rhul.ac.uk/~cowan/stat/stave/

Application to the muon $g-2$ anomaly

The recently measured muon $g-2$ (ave. of 2006, 2021) disagrees with the Standard Model prediction with a significance of 4.2σ.

Muon g-2 Collab., PRL 126, 141801 (2021)



Discrepancy significantly reduced by 2021 lattice-based prediction of Borsanyi et al. (BMW).

Current goal is to investigate sensitivity of significance to error assumptions, so for now focus on the 4.2σ problem.

Muon $g-2$ ingredients

Using $a_\mu = (g-2)/2$ $\quad y = a_\mu \times 10^9 - 1165900$

the ingredients of the 4.2σ effect are:

$$y_{\rm exp} = 20.61 \pm 0.41$$

(ave. of BNL 2006 and FNAL 2021)

0.37 (stat.) ± 0.17 (sys.)

B. Abi et al. (Muon $g-2$ Collaboration), *Measurement of the Positive Muon Anomalous Magnetic Moment to 0.46 ppm*, Phys. Rev. Lett. 126, 141801 (2021).

G. W. Bennett et al. (Muon $g-2$ Collaboration), *Final report of the E821 muon anomalous magnetic moment measurement at BNL*, Phys. Rev. D 73, 072003 (2006).

$$y_{\rm SM} = 18.10 \pm 0.43$$

(SM pred. by Muon $g-2$ theory initiative)

0.40 (Had. Vac. Pol.) ± 0.18 (Had. Light-by-Light)

T. Aoyama, N. Asmussen, M. Benayoun, J. Bijnens, and T. Blum et al., *The anomalous magnetic moment of the muon in the standard model*, Phys. Rep. 887, 1 (2020).

Suppose $\sigma_{\rm SM}$ uncertain

Suppose measurement errors well known, but that the SM theory error $\sigma_{\rm SM}$ (estimated 0.43) could be uncertain.

This is the largest systematic and probably hardest to estimate.

Treat estimate $v_{\rm SM} = (0.43)^2$ of variance $\sigma^2_{\rm SM}$ as gamma distributed, width from relative uncertainty parameter $r_{\rm SM}$.

Maximum-likelihood for mean from minimum of

$$Q(\mu) = -2 \ln \frac{L(\mu)}{L_{\rm sat}}$$

$$= \frac{(y_{\rm exp} - \mu)^2}{\sigma^2_{\rm exp}} + \left(1 + \frac{1}{2 r^2_{\rm SM}}\right) \ln \left[1 + 2 r^2_{\rm SM} \frac{(y_{\rm SM} - \mu)^2}{v_{\rm SM}}\right]$$

p-value/significance of common-mean hypothesis

Significance (goodness of fit) from $q = Q(\hat{\mu})$.

Because of non-quadratic term in $Q(\mu)$, distribution of q departs from chi-square(1) for increasing $r_{\rm SM}$.

Best to get distribution of q from Monte Carlo (and speed up with Bartlett correction – see EPJC (2019) 79:133).

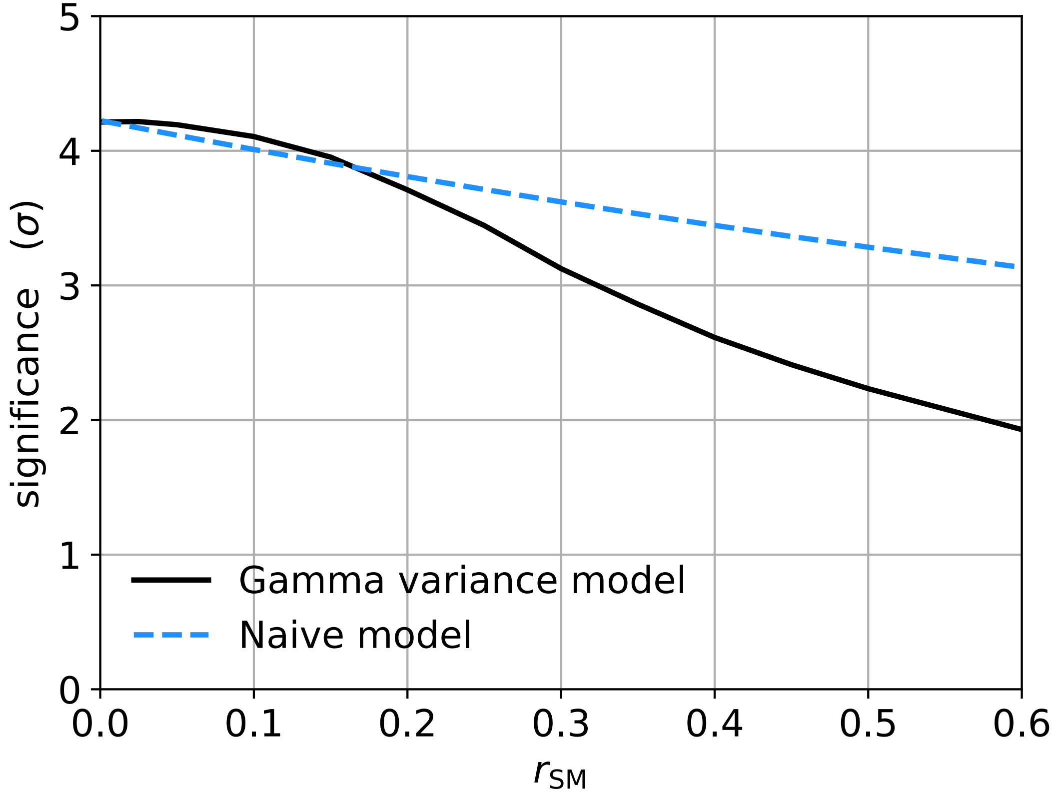
For $r_{\rm SM} > 0$ distribution of q depends on $\sigma^2_{\rm SM}$. For MC use Maximum-Likelihood estimate ("profile construction"):

$$\widehat{\sigma^2}_{\rm SM} = \frac{v_{\rm SM} + 2r_{\rm SM}^2 (y_{\rm SM} - \hat{\mu})^2}{1 + 2r_{\rm SM}^2}$$

$$\mathrm{MC} \;\rightarrow\; f(q) \;\rightarrow\; p = \int_{q,\mathrm{obs}}^{\infty} f(q)\,dq \;\rightarrow\; \text{significance } Z = \Phi^{-1}(1 - p/2)$$

# of sigmas

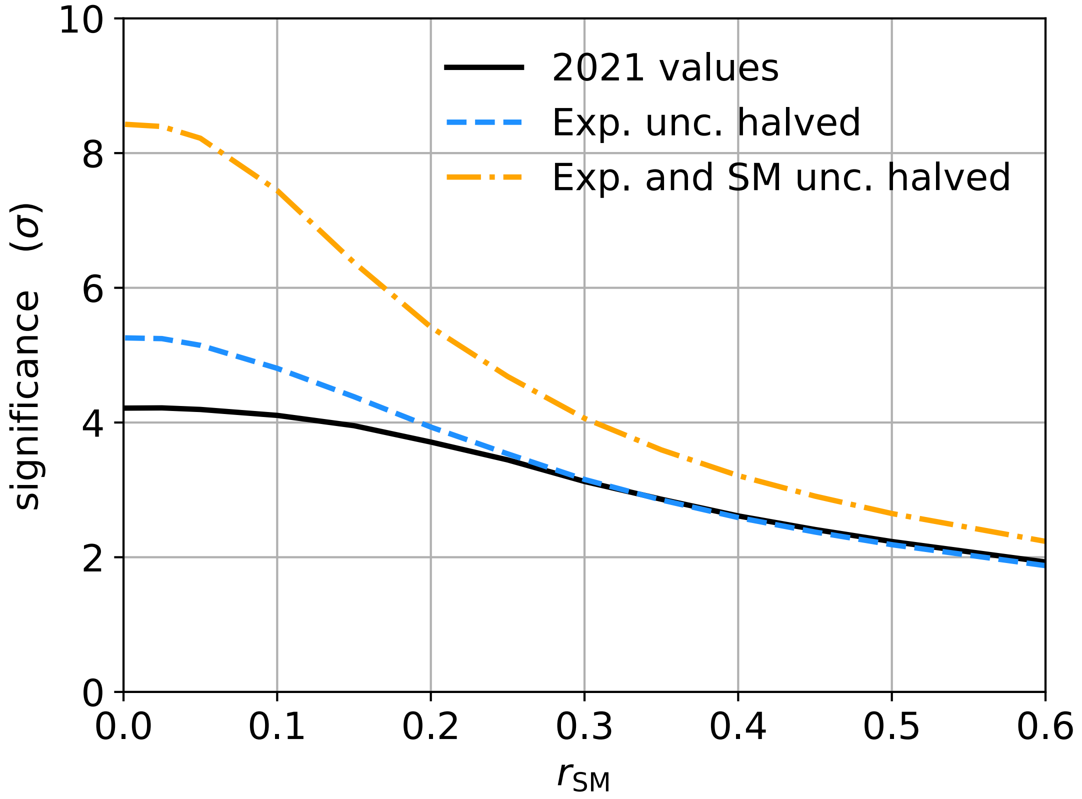
Significance of discrepancy versus r_{SM}



Naive model: use least squares but let $\sigma_{\text{SM}} \to (1 + r_{\text{SM}})\sigma_{\text{SM}}$

Gamma variance model gives greater decrease in significance for $r_{\text{SM}} \gtrsim 0.2$, e.g., 3.1σ for $r_{\text{SM}} = 0.3$, 2.0σ for $r_{\text{SM}} = 0.6$.

Significance of discrepancy versus r_{SM}



Establishing 4σ effect requires $r_{\mathrm{SM}} \lesssim 0.3$ even if nominal exp. and SM uncertainties become half of present values.

Discussion / conclusions

Including uncertainties on estimates of uncertainties can have large effect on hypothesis test, esp. for high significance.

To establish e.g. a 5σ effect it is crucial to have both:

small uncertainties

accurate estimates of those uncertainties ($\sim$ 20% level)

This is ultimately because the tails of the Gaussian fall off so quickly.

Gamma Variance Model $\sim$ Student's t likelihood with $\nu = 1/2r^2$ degrees of freedom $\rightarrow$ longer tails than Gaussian.

Work ongoing with Bogdan Malaescu of Muon g-2 Theory Initiative on the HVP uncertainty, see, e.g.,

B. Malaescu et al., https://indico.him.uni-mainz.de/event/11/contributions/80/attachments/50/51/amuWorkshop_Correlations_Malaescu.pdf

M. Davier et al., Eur. Phys. J. C 80 (2020) 241 , arXiv:1908.00921

Discussion / conclusions (2)

Other features of Gamma Variance Model (see EPJC (2019) 79:133 and the extra slides)

averages/fits become less sensitive to outliers;

confidence intervals linked to goodness of fit;

straightforward to include multiple correlated error sources.

But... is part of the reason for requiring 5σ for discovery not to account for uncertainties in assigned errors? Is there a trade-off between "errors on errors" and the requirement for discovery?

Best to have most realistic model. If the estimated errors are indeed uncertain, this should be reflected in the model.

Bottom line – it is very difficult to establish convincing evidence for a new physics if relevant uncertainties are estimated in an ad hoc way. We need robust procedures for their assignment.

Extra Slides

Full likelihood for gamma variance model

$$L(\boldsymbol{\mu}, \boldsymbol{\theta}, \boldsymbol{\sigma}_{\mathbf{u}}^2) = P(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\theta}) \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma_{u_i}^2}} e^{-(u_i-\theta_i)^2/2\sigma_{u_i}^2}$$

$$\times \quad \frac{\beta_i^{\alpha_i}}{\Gamma(\alpha_i)} v_i^{\alpha_i - 1} e^{-\beta_i v_i} \,,$$

$\alpha_i = 1/4r_i^2$
$\beta_i = \alpha_i/\sigma_{ui}^2$

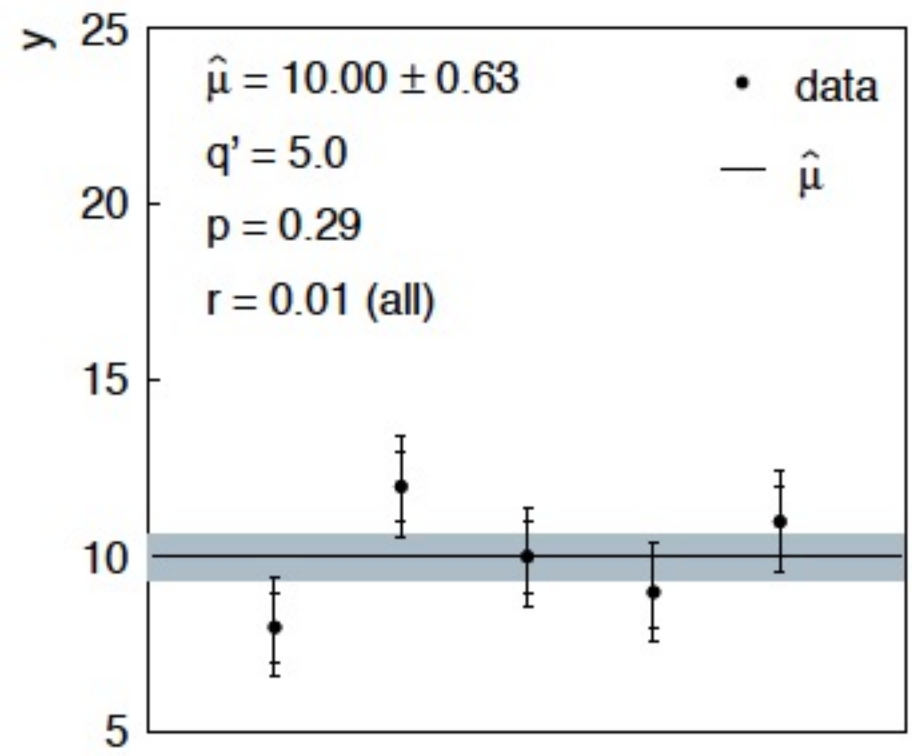
Treated like data: $y_1,\ldots,y_L$ (the primary measurements)
$u_1,\ldots,u_N$ (estimates of nuisance par.)
$v_1,\ldots,v_N$ (estimates of variances of estimates of NP)

Adjustable parameters: $\mu_1,\ldots,\mu_M$ (parameters of interest)
$\theta_1,\ldots,\theta_N$ (nuisance parameters)
$\sigma_{u,1},\ldots,\sigma_{u,N}$ (sys. errors = std. dev. of of NP estimates)

Fixed parameters: $r_1,\ldots,r_N$ (rel. err. in estimate of $\sigma_{u,i}$)

Sensitivity of average to outliers

Suppose we average 5 values, $y = 8, 9, 10, 11, 12$, all with stat. and sys. errors of 1.0, and suppose negligible error on error (here take $r = 0.01$ for all).

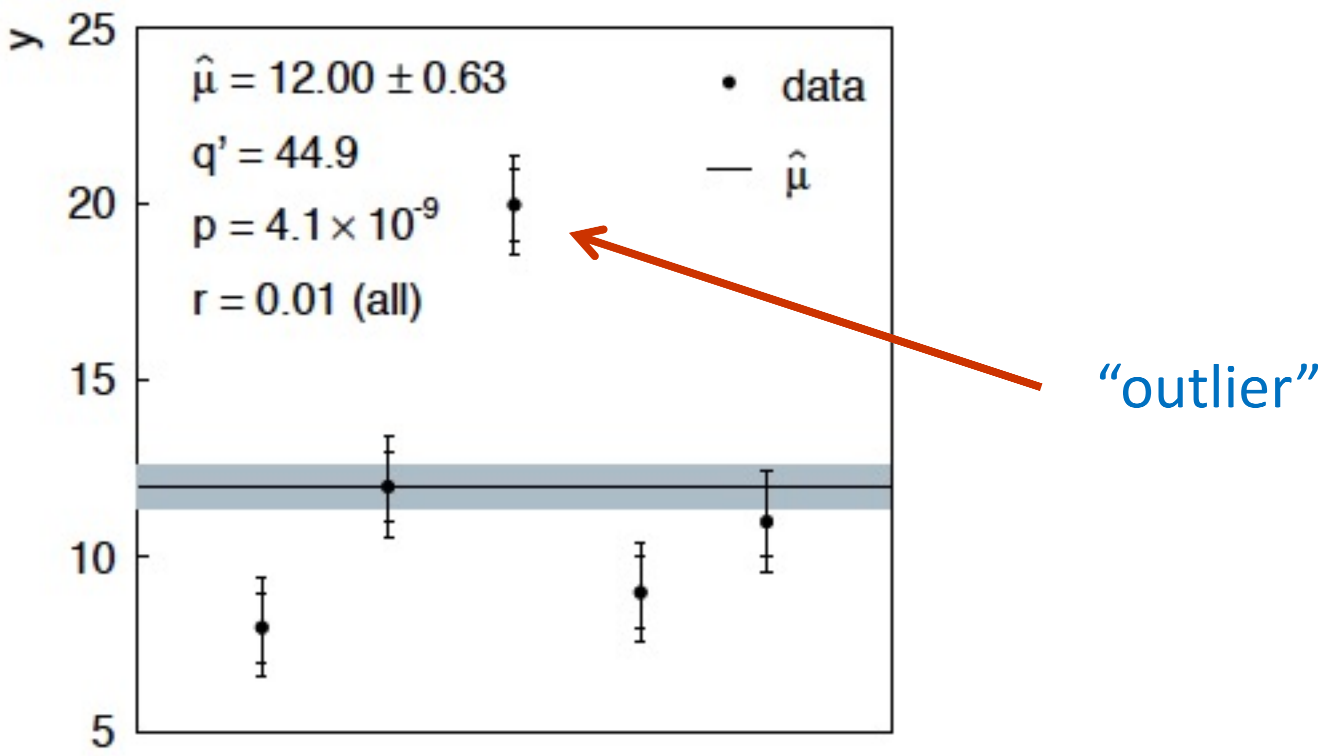


inner error bars $= \sigma_{y,i}$

outer error bars $= (\sigma_{y,i}^2 + \sigma_{u,i}^2)^{1/2}$

Sensitivity of average to outliers (2)

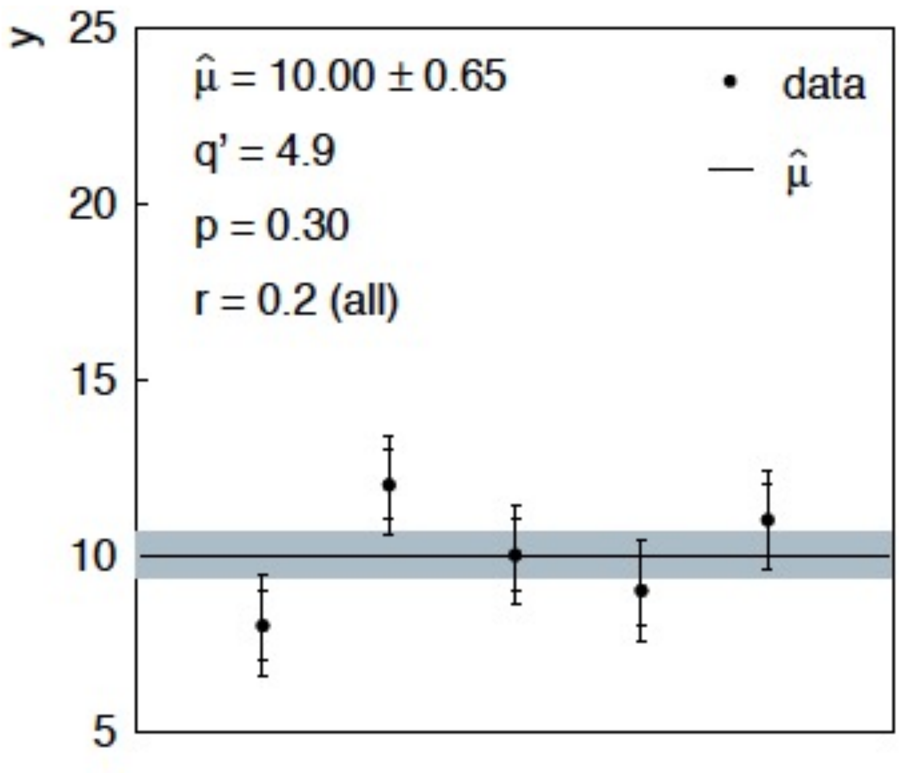
Now suppose the measurement at 10 had come out at 20:



Estimate pulled up to 12.0, size of confidence interval ~unchanged (would be exactly unchanged with $r \to 0$).

Average with all $r = 0.2$

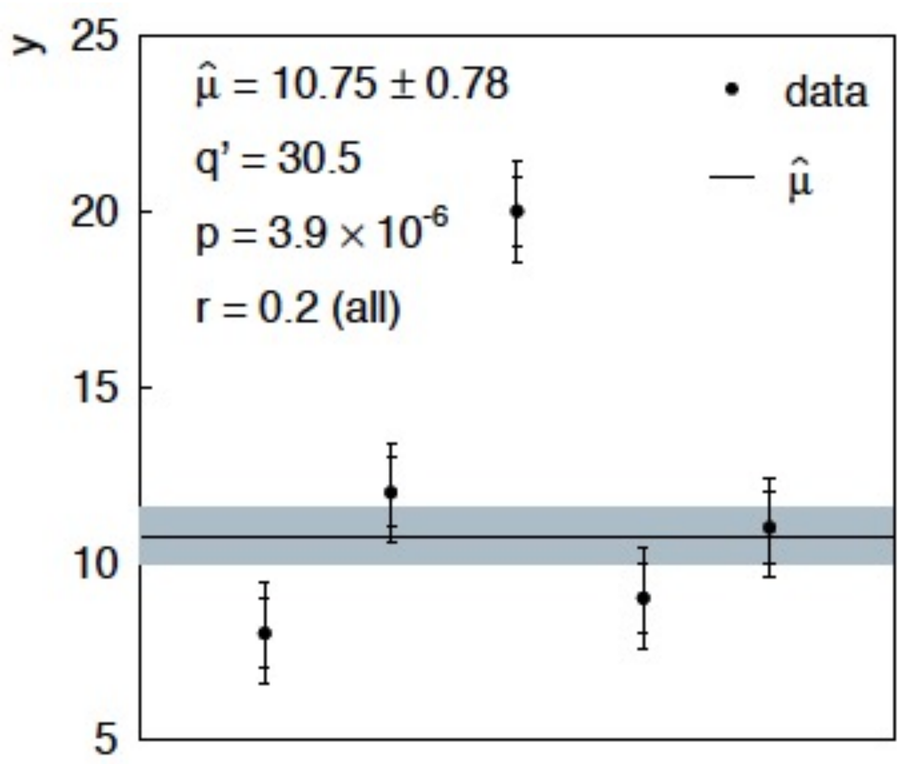
If we assign to each measurement $r = 0.2$,



Estimate still at 10.00, size of interval moves 0.63 → 0.65

Average with all $r = 0.2$ with outlier

Same now with the outlier (middle measurement $10 \to 20$)



Estimate $\rightarrow 10.75$ (outlier pulls much less).

Half-size of interval $\rightarrow 0.78$ (inflated because of bad g.o.f.).

Naive approach to errors on errors

Naively one might think that the error on the error in the previous example could be taken into account conservatively by inflating the systematic errors, i.e.,

$$\sigma_{u_i} \to \sigma_{u_i}(1 + r_i)$$

But this gives

$$\hat{\mu} = 10.00 \pm 0.70$$ without outlier (middle meas. 10)

$$\hat{\mu} = 12.00 \pm 0.70$$ with outlier (middle meas. 20)

So the sensitivity to the outlier is not reduced and the size of the confidence interval is still independent of goodness of fit.

Correlated uncertainties

The phrase "correlated uncertainties" usually means that a single nuisance parameter affects the distribution (e.g., the mean) of more than one measurement.

For example, consider measurements $\boldsymbol{y}$, parameters of interest $\boldsymbol{\mu}$, nuisance parameters $\boldsymbol{\theta}$ with

$$E[y_i] = \varphi_i(\boldsymbol{\mu}, \boldsymbol{\theta}) \approx \varphi_i(\boldsymbol{\mu}) + \sum_{j=1}^{N} R_{ij}\theta_j$$

That is, the θ_i are defined here as contributing to a bias and the (known) factors R_{ij} determine how much θ_j affects y_i.

As before suppose one has independent control measurements $u_i \sim$ Gauss(θ_i, σ_{ui}).

Correlated uncertainties (2)

The total bias of y_i can be defined as $$b_i = \sum_{j=1}^{N} R_{ij} \theta_j$$

which can be estimated with $$\hat{b}_i = \sum_{j=1}^{N} R_{ij} u_j$$

These estimators are correlated having covariance

$$U_{ij} = \mathrm{cov}[\hat{b}_i, \hat{b}_j] = \sum_{k=1}^{N} R_{ik} R_{jk} V[u_k]$$

In this sense the present method treats “correlated uncertainties”, i.e., the control measurements u_i are independent, but nuisance parameters affect multiple measurements, and thus bias estimates are correlated.