Errors on Errors: Refining Particle Physics Analyses with the Gamma Variance Model

Carnegie Mellon University

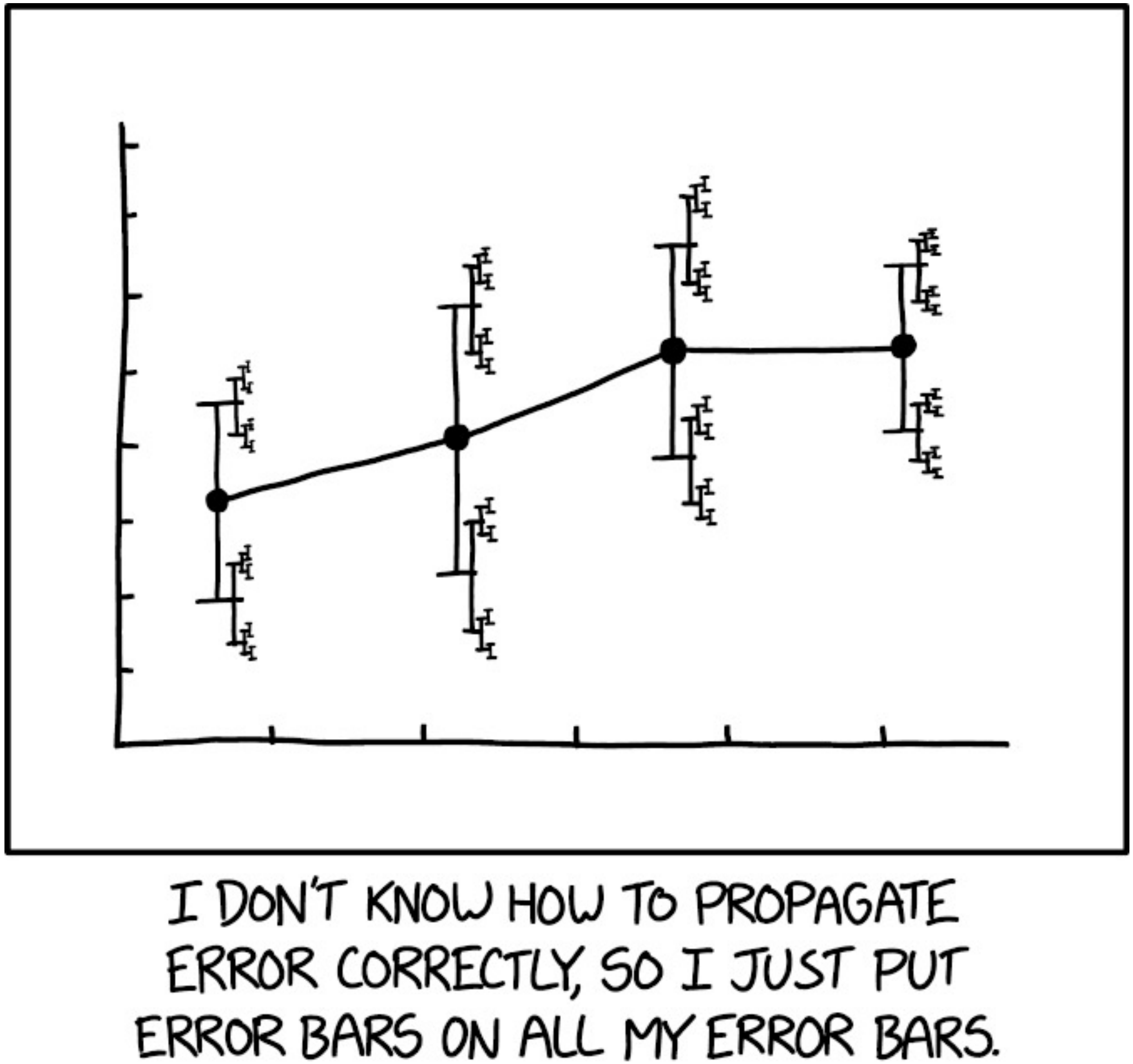
STAMPS Seminar
CMU via London
12 November 2021

ATLAS

ROYAL HOLLOWAY UNIVERSITY OF LONDON

INSIGHTS

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


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

Outline

Intro, motivation

Using measurements with "known" systematic errors:

Least Squares (BLUE)

Allowing for uncertainties in the systematic errors

Estimates of sys errors $\sim$ Gamma

Asymptotics, Bartlett correction

Curve fitting, averages

Confidence intervals, goodness-of-fit, outliers

Application to the muon $g\text{–}2$ anomaly (arXiv:2107.02652)

Discussion and conclusions

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

A typical particle physics measurement

A typical analysis involves primary measurements $\boldsymbol{y}$ modeled with probability $P(\boldsymbol{y}|\boldsymbol{\mu},\boldsymbol{\theta})$,

$\boldsymbol{\mu}$ = parameters of interest

$\boldsymbol{\theta}$ = nuisance parameters

To provide info on nuisance parameters, often treat their best estimates $\boldsymbol{u}$ as indep. Gaussian distributed with known std. dev. $\sigma_{\boldsymbol{u}}$:

$$L(\boldsymbol{\mu},\boldsymbol{\theta}) = P(\mathbf{y},\mathbf{u}|\boldsymbol{\mu},\boldsymbol{\theta}) = P(\mathbf{y}|\boldsymbol{\mu},\boldsymbol{\theta})P(\mathbf{u}|\boldsymbol{\theta})$$

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

or log-likelihood (up to additive const.)

$$\ln L(\boldsymbol{\mu},\boldsymbol{\theta}) = \ln P(\mathbf{y}|\boldsymbol{\mu},\boldsymbol{\theta}) - \frac{1}{2}\sum_{i=1}^{N} \frac{(u_i-\theta_i)^2}{\sigma_{u_i}^2}$$

assumes ”systematic errors” σ_{ui} are known

Motivation

Analyses are often sensitive to the values σ_u assigned to control measurements (the "systematic errors").

But these error estimates are also uncertain (→ errors on errors)

Could just try inflating the systematic error estimates, but this turns out not to be enough, especially if the analysis uses least squares (equivalent to assuming Gaussian pdfs in likelihood).

Need for "errors on errors" most visible when measurements are not internally consistent within their estimated uncertainties.

Candidate use cases in particle physics:

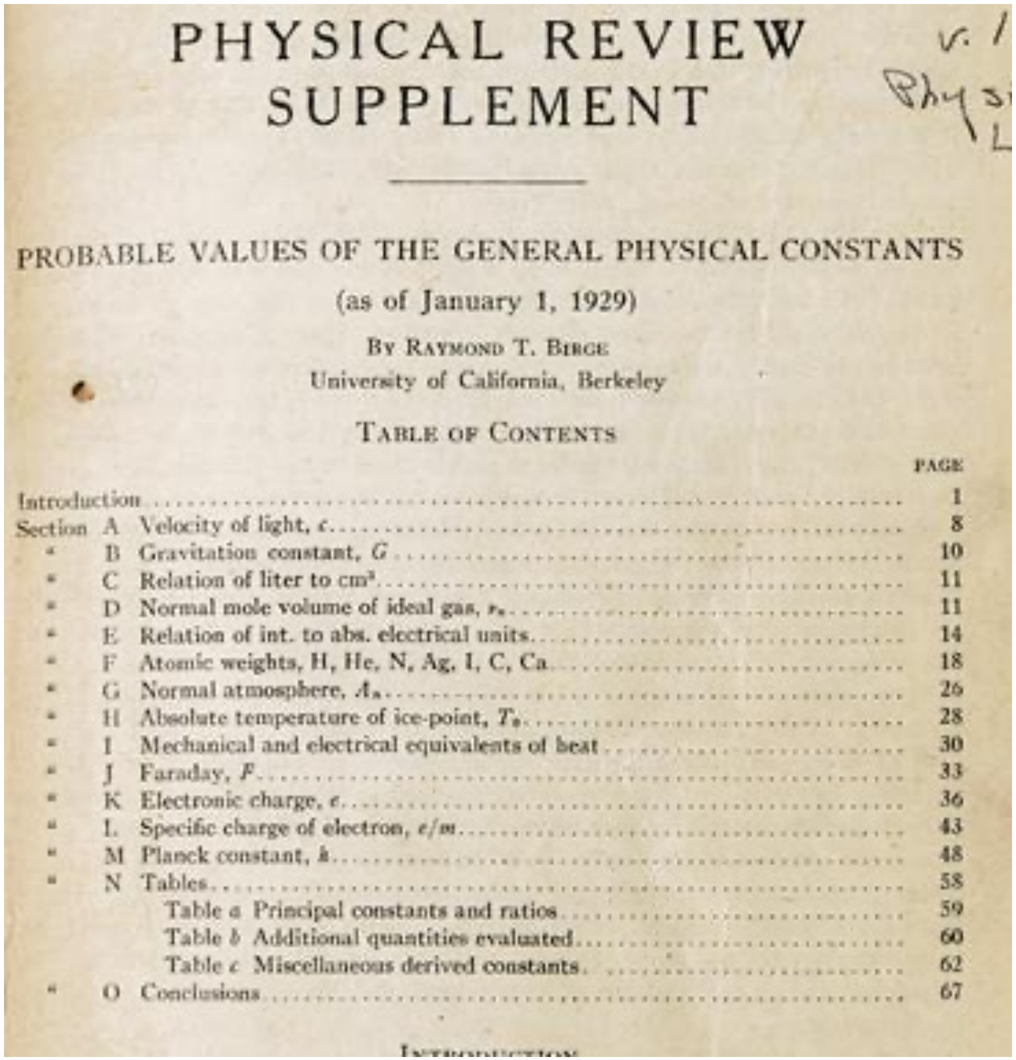
Combinations of inconsistent measurements

Analyses where systematic error assigned by ad hoc recipe

Any analysis where assigned systematic error is uncertain

Least Squares for Averaging

= fit of horizontal line



PHYSICAL REVIEW SUPPLEMENT

PROBABLE VALUES OF THE GENERAL PHYSICAL CONSTANTS

(as of January 1, 1929)

By Raymond T. Birge
University of California, Berkeley

TABLE OF CONTENTS

Raymond T. Birge, *Probable Values of the General Physical Constants (as of January 1, 1929)*, Physical Review Supplement, Vol 1, Number 1, July 1929

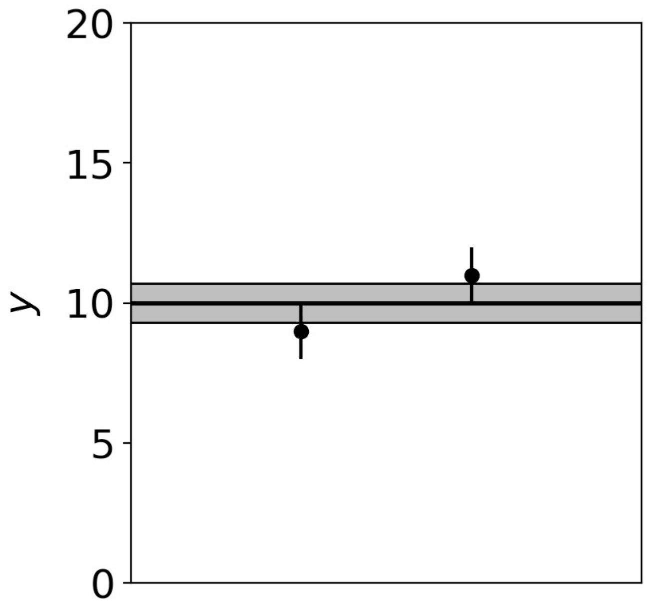
Forerunner of the Particle Data Group

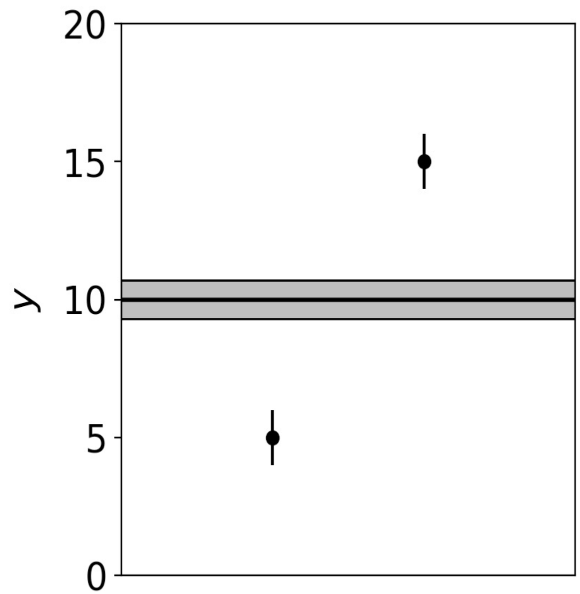
Motivation (2)

Assuming known standard deviations for least squares, uncertainty (e.g. confidence interval) does not reflect goodness of fit:

Least squares average of 9 ± 1 and 11 ± 1 is 10 ± 0.71

Least squares average of 5 ± 1 and 15 ± 1 is 10 ± 0.71





Width of confidence interval for the mean does not reflect the consistency of the values being averaged.

Systematic errors and their uncertainty

Sometimes $\sigma_{u,i}$ is well known, e.g., it is itself a statistical error known from sample size of a control measurement.

Other times the u_i are from an indirect measurement, Gaussian model approximate and/or the $\sigma_{u,i}$ are not exactly known.

Or sometimes $\sigma_{u,i}$ is at best a guess that represents an uncertainty in the underlying model (“theoretical error”).

In any case we can allow that the $\sigma_{u,i}$ are not known in general with perfect accuracy.

Gamma variance model

As the systematic errors $\sigma_{u,i}$ are uncertain, let them be adjustable nuisance parameters.

Treat their assigned values as estimates s_i for $\sigma_{u,i}$ or equivalently $v_i = s_i^2$, is an estimate of $\sigma_{u,i}^2$.

Model the v_i as independent and gamma distributed:

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

$$E[v] = \frac{\alpha}{\beta}$$

$$V[v] = \frac{\alpha}{\beta^2}$$

Set α and β so that they give desired relative uncertainty r in σ_u.

Leads to model equivalent to replacing Gaussian by Student's t with $\nu = 1/2r^2$ degrees of freedom.

Gamma model for estimates of variance

Suppose the estimated variance v was obtained as the sample variance from n observations of a Gaussian distributed bias estimate u.

In this case v is gamma distributed with

$$\alpha = \frac{n-1}{2} \qquad\qquad \beta = \frac{n-1}{2\sigma_u^2}$$

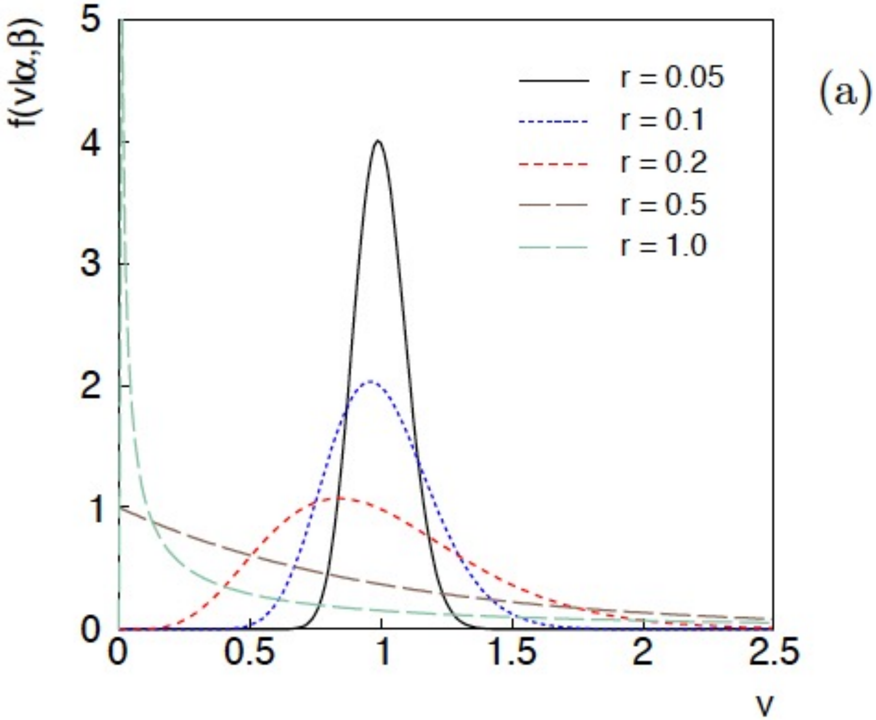
Furthermore choice of the gamma distribution for v allows one to profile over the nuisance parameters σ_u^2 in closed form and leads to a simple profile likelihood.

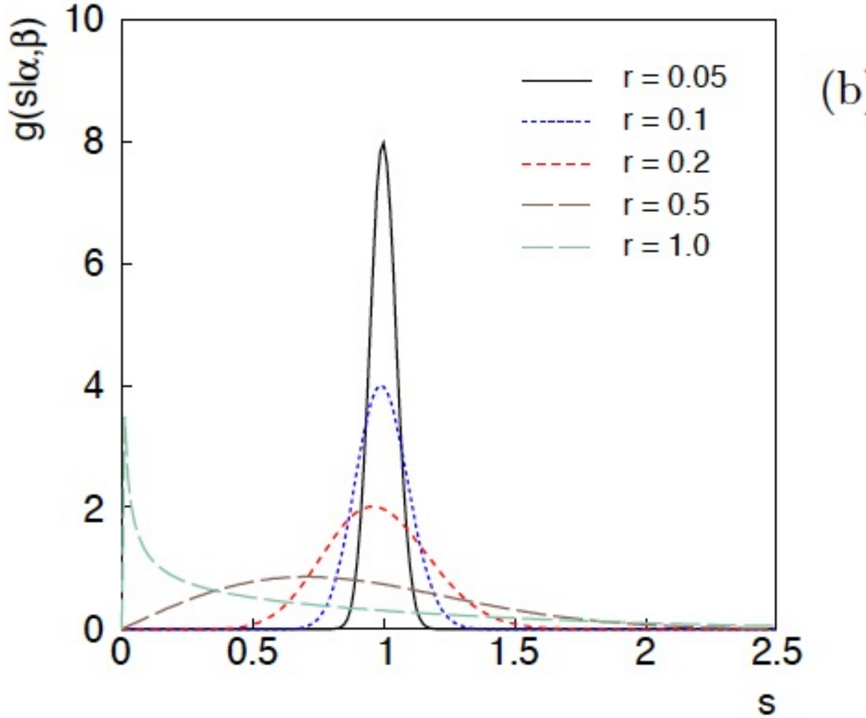
Distributions of v and $s = \sqrt{v}$

For α, β of gamma distribution, $\alpha_i = \frac{1}{4r_i^2}, \; \beta_i = \frac{1}{4r_i^2\sigma_{u_i}^2}$

$$r_i \equiv \frac{1}{2}\frac{\sigma_{v_i}}{E[v_i]} = \frac{1}{2}\frac{\sigma_{v_i}}{\sigma_{u_i}^2} \approx \frac{\sigma_{s_i}}{E[s_i]}$$

relative "error on error"





Full likelihood for gamma variance model

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

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

$$\alpha_i = \frac{1}{4r_i^2} \, ,$$

$$\beta_i = \frac{1}{4r_i^2 \sigma_{u_i}^2}$$

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: $\mu_1,\ldots,\mu_{\mathrm{M}}$ (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 $\sigma_{u,i}$)

Profiling over systematic errors

We can profile over the $\sigma_{u,i}$ in closed form

$$\widehat{\widehat{\sigma^2}}_{u_i} = \underset{\sigma^2_{u_i}}{\operatorname{argmax}} L(\boldsymbol{\mu}, \boldsymbol{\theta}, \boldsymbol{\sigma}^2_{\mathbf{u}}) = \frac{v_i + 2r_i^2(u_i - \theta_i)^2}{1 + 2r_i^2}$$

which gives the profile log-likelihood (up to additive const.)

$$\boxed{\begin{aligned} \ln L'(\mu, \boldsymbol{\theta}) &= \ln L(\mu, \boldsymbol{\theta}, \widehat{\widehat{\boldsymbol{\sigma}^2}}_{\mathbf{u}}) \\ &= \ln P(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\theta}) - \frac{1}{2}\sum_{i=1}^{N} \left(1 + \frac{1}{2r_i^2}\right) \ln\left[1 + 2r_i^2 \frac{(u_i - \theta_i)^2}{v_i}\right] \end{aligned}}$$

In limit of small r_i and $v_i \to \sigma_{u,i}{}^2$, the log terms revert back to the quadratic form seen with known $\sigma_{u,i}$.

Equivalent likelihood from Student's t

We can arrive at same likelihood by defining $$z_i \equiv \frac{u_i - \theta_i}{\sqrt{v_i}}$$

Since $u_i \sim$ Gauss and $v_i \sim$ Gamma, $z_i \sim$ Student's t

$$f(z_i|\nu_i) = \frac{\Gamma\left(\frac{\nu_i+1}{2}\right)}{\sqrt{\nu_i \pi}\,\Gamma(\nu_i/2)} \left(1 + \frac{z_i^2}{\nu_i}\right)^{-\frac{\nu_i+1}{2}}$$ with $$\nu_i = \frac{1}{2r_i^2}$$

Resulting likelihood same as profile $L'(\boldsymbol{\mu},\boldsymbol{\theta})$ from gamma model

$$L(\boldsymbol{\mu},\boldsymbol{\theta}) = P(\mathbf{y}|\boldsymbol{\mu},\boldsymbol{\theta}) \prod_{i=1}^{N} \frac{\Gamma\left(\frac{\nu_i+1}{2}\right)}{\sqrt{\nu_i \pi}\,\Gamma(\nu_i/2)} \left(1 + \frac{z_i^2}{\nu_i}\right)^{-\frac{\nu_i+1}{2}}$$

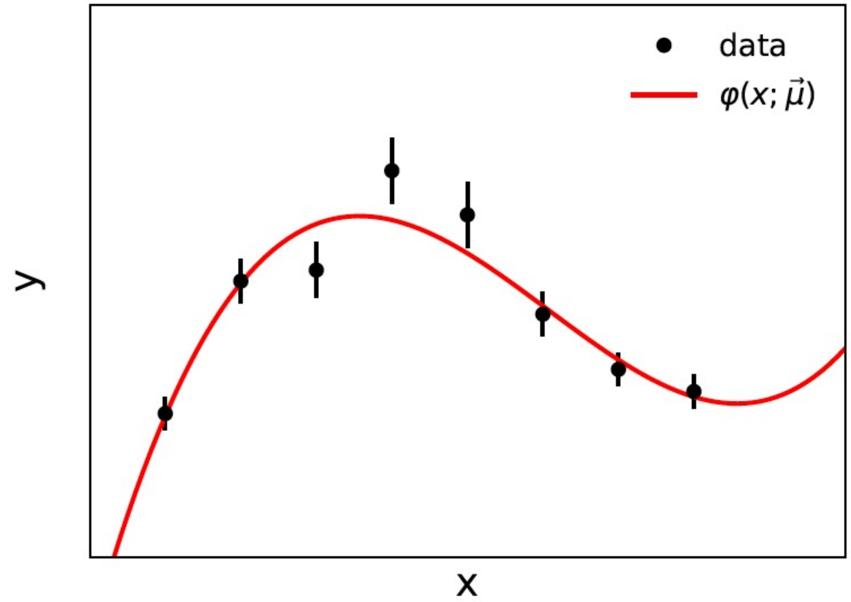
Curve fitting, averages

Suppose independent

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

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

$$V[y_i] = \sigma_{y_i}^2 \,.$$



$\boldsymbol{\mu}$ are the parameters of interest in the fit function $\varphi(x;\boldsymbol{\mu})$,

$\boldsymbol{\theta}$ are bias parameters constrained by control measurements $u_i \sim$ Gauss(θ_i, $\sigma_{u,i}$), so that if $\sigma_{u,i}$ are known we have

$$-2 \ln L(\boldsymbol{\mu}, \boldsymbol{\theta}) = \sum_{i=1}^{N} \left[\frac{(y_i - \varphi(x_i; \boldsymbol{\mu}) - \theta_i)^2}{\sigma_{y_i}^2} + \frac{(u_i - \theta_i)^2}{\sigma_{u_i}^2} \right]$$

Goodness of fit

Can quantify goodness of fit with statistic

$$q = -2 \ln \frac{L'(\hat{\boldsymbol{\mu}}, \hat{\hat{\boldsymbol{\theta}}})}{L'(\hat{\boldsymbol{\varphi}}, \hat{\boldsymbol{\theta}})}$$

$$= \min_{\boldsymbol{\mu}, \boldsymbol{\theta}} \sum_{i=1}^{N} \left[\frac{(y_i - \varphi(x_i; \boldsymbol{\mu}) - \theta_i)^2}{\sigma_{y_i}^2} + \left(1 + \frac{1}{2r_i^2}\right) \ln \left(1 + 2r_i^2 \frac{(u_i - \theta_i)^2}{v_i}\right) \right]$$

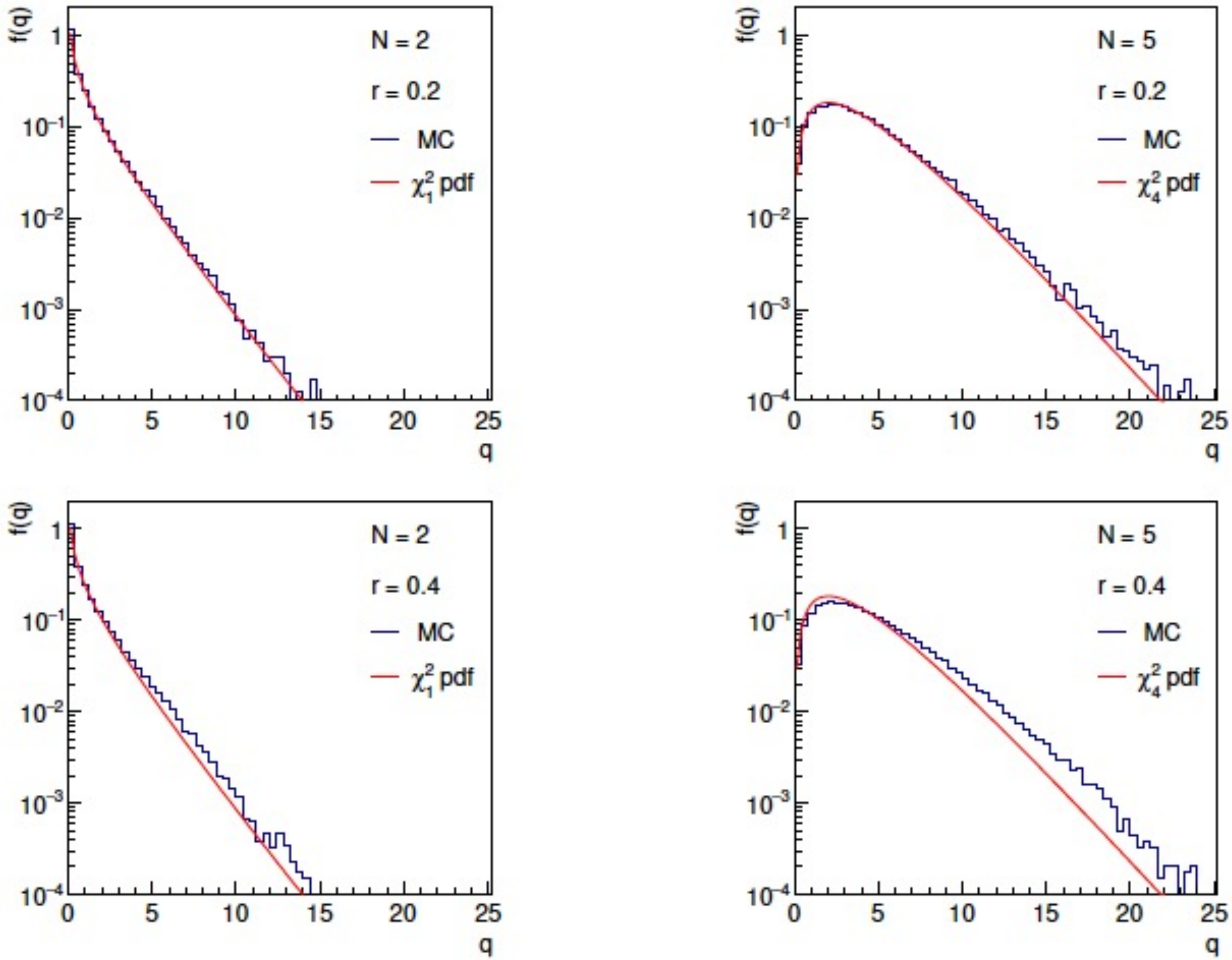
where $L'(\boldsymbol{\varphi}, \boldsymbol{\theta})$ has an adjustable φ_i for each y_i (the saturated model).

Asymptotically should have $q \sim$ chi-squared(N−M).

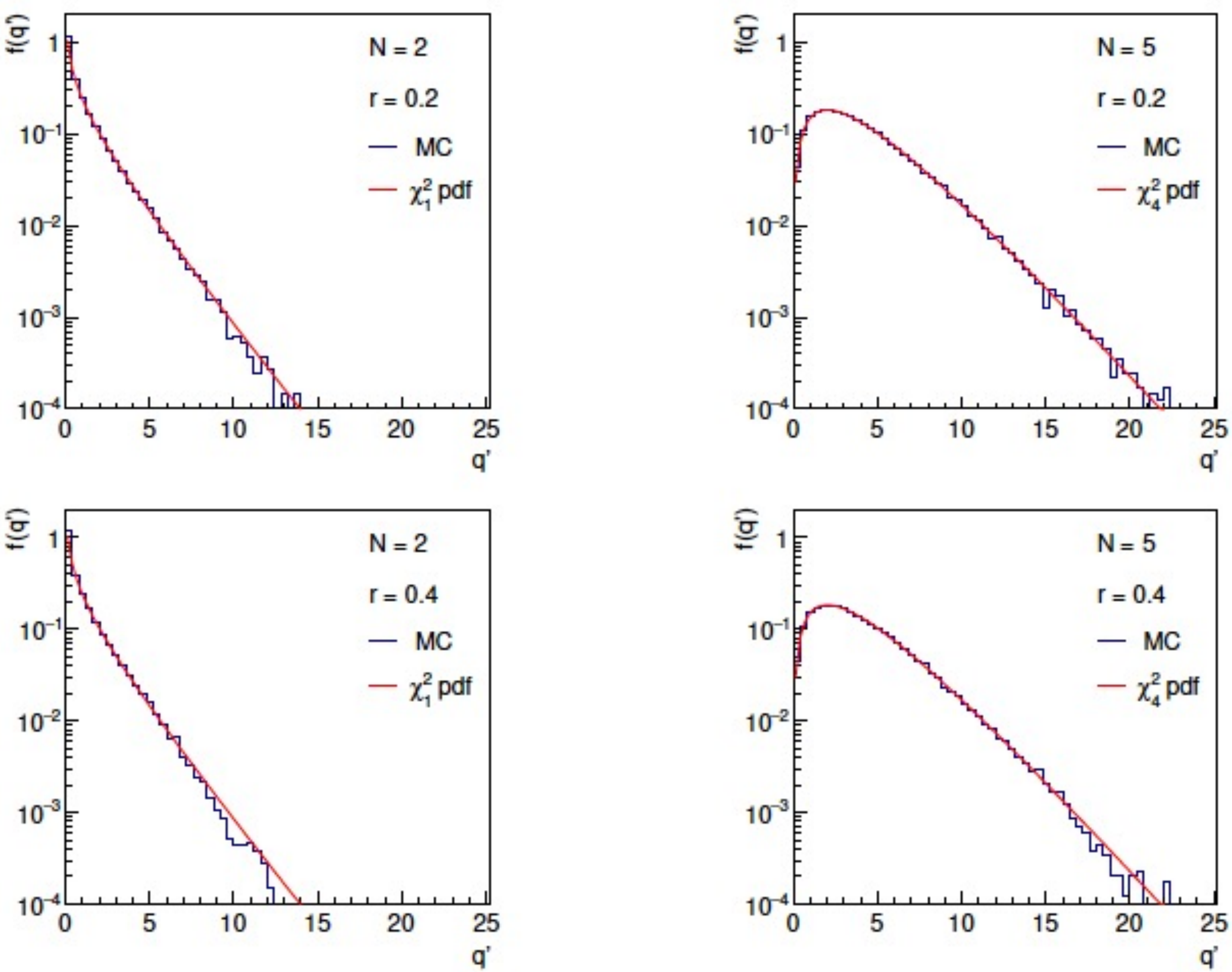
For increasing r_i, asymptotics breaks down: ongoing studies with Alessandra Brazzale (Uni. Padova) on higher order asymptotics.

Some improvement possible with Bartlett correction.

Distributions of q



Distributions of Bartlett-corrected $q' = \frac{n_{\rm d}}{E[q]} q$

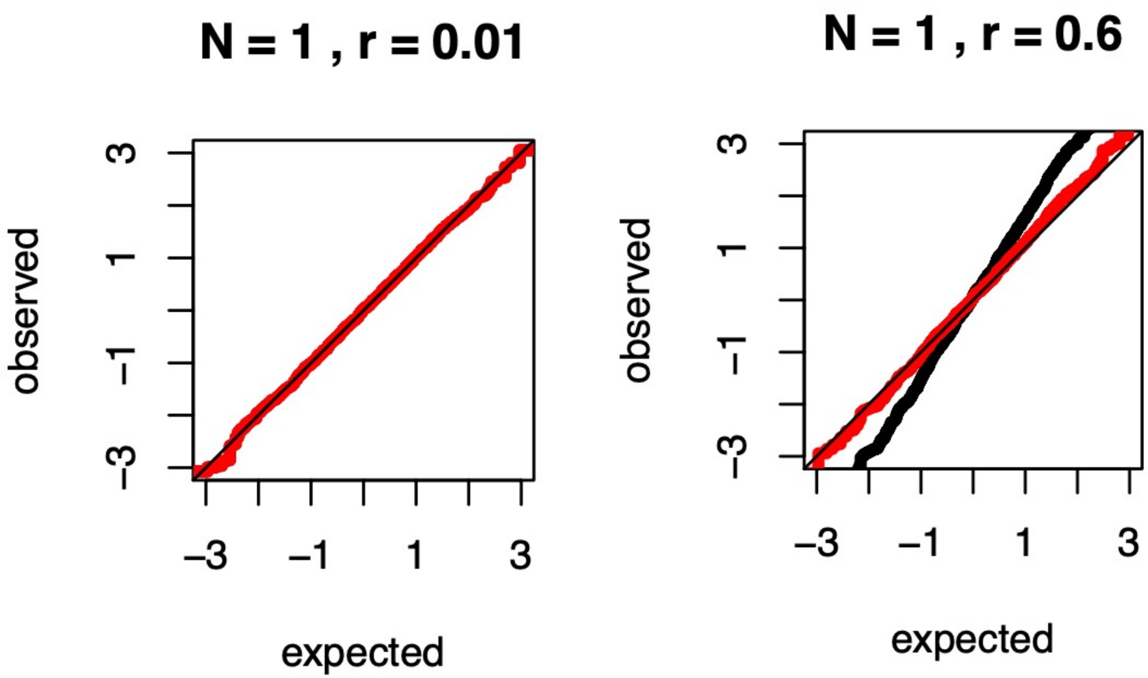


Higher-order asymptotics (A. Brazzale)

https://indico.cern.ch/event/1051224/contributions/4540321/attachments/2337195/3983787/PhyStatSyst_Brazzale%5B7%5D.pdf

Modified likelihood root

$$t^*(\mu) = t(\mu) + \frac{1}{t(\mu)} \log\left\{\frac{q(\mu)}{t(\mu)}\right\} \sim N(0,1) + O(n^{-3/2})$$

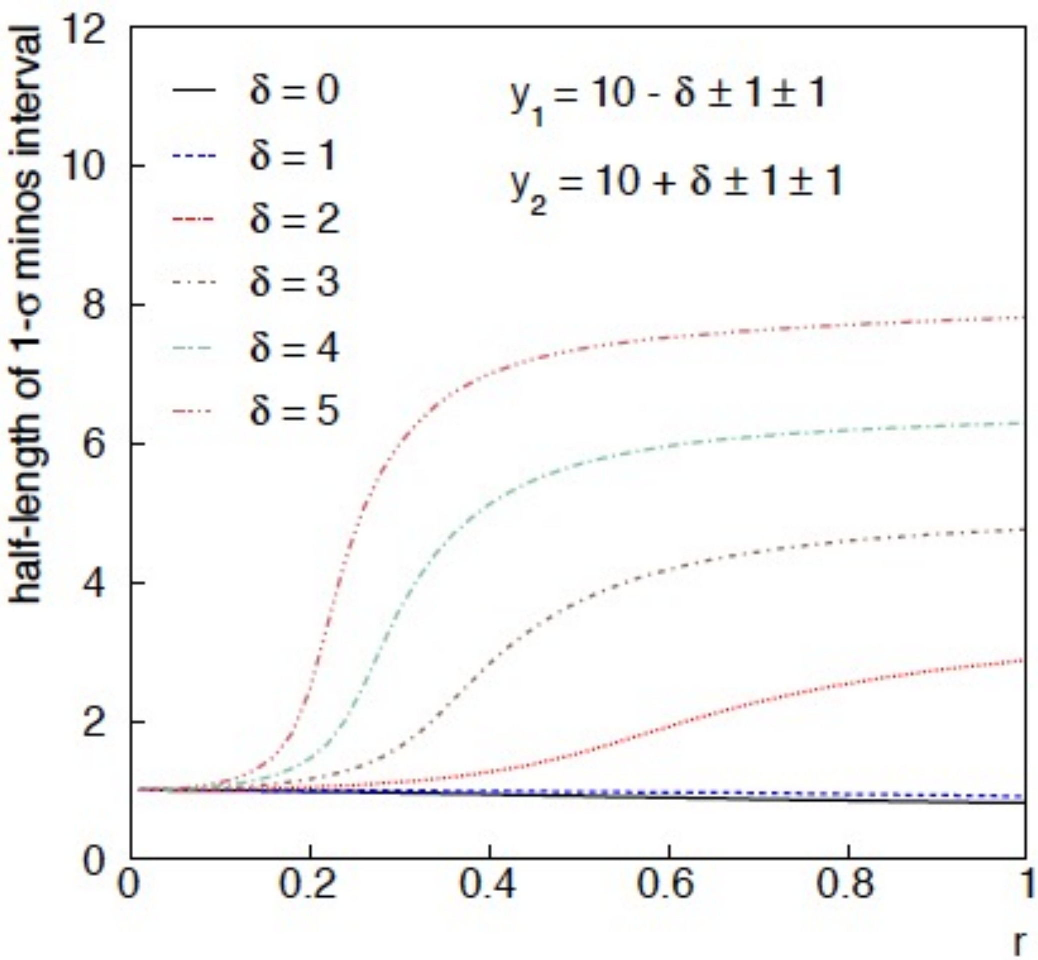


Modified statistic shows close to asymptotic distribution even for $r \sim 0.6$. No simulation required.

Example: average of two measurements

MINOS interval (= approx. confidence interval) based on

$$\ln L'(\mu) = \ln L'(\hat{\mu}) - Q_\alpha/2 \quad \text{with} \quad Q_\alpha = F^{-1}_{\chi^2}(1-\alpha; n)$$



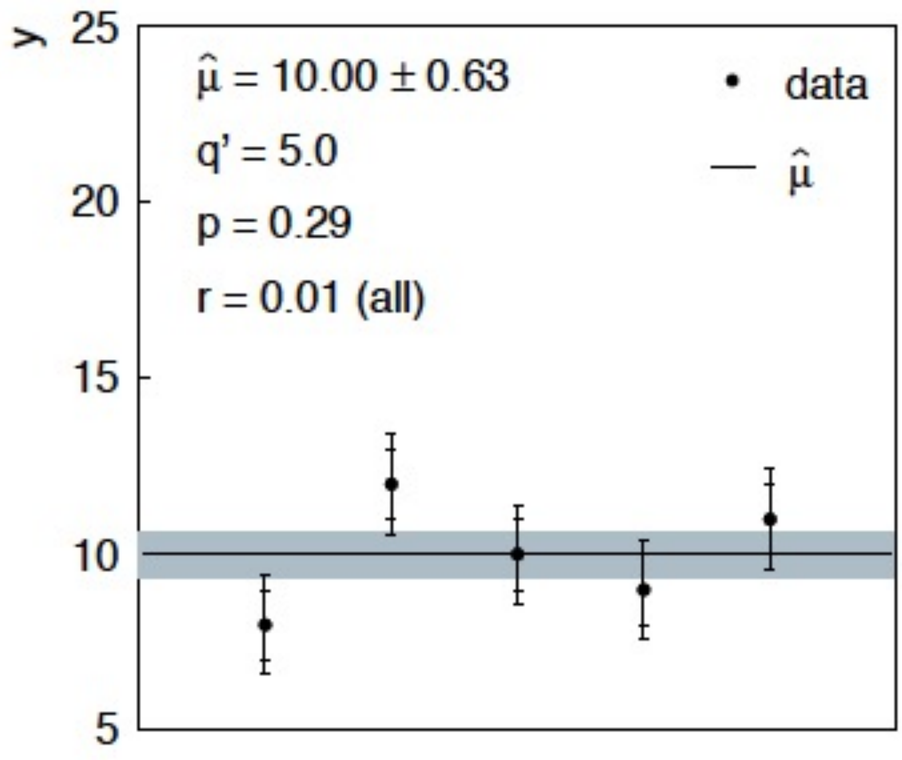
Increased discrepancy between values to be averaged gives larger interval.

Interval length saturates at ~level of absolute discrepancy between input values.

relative error on sys. error

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).



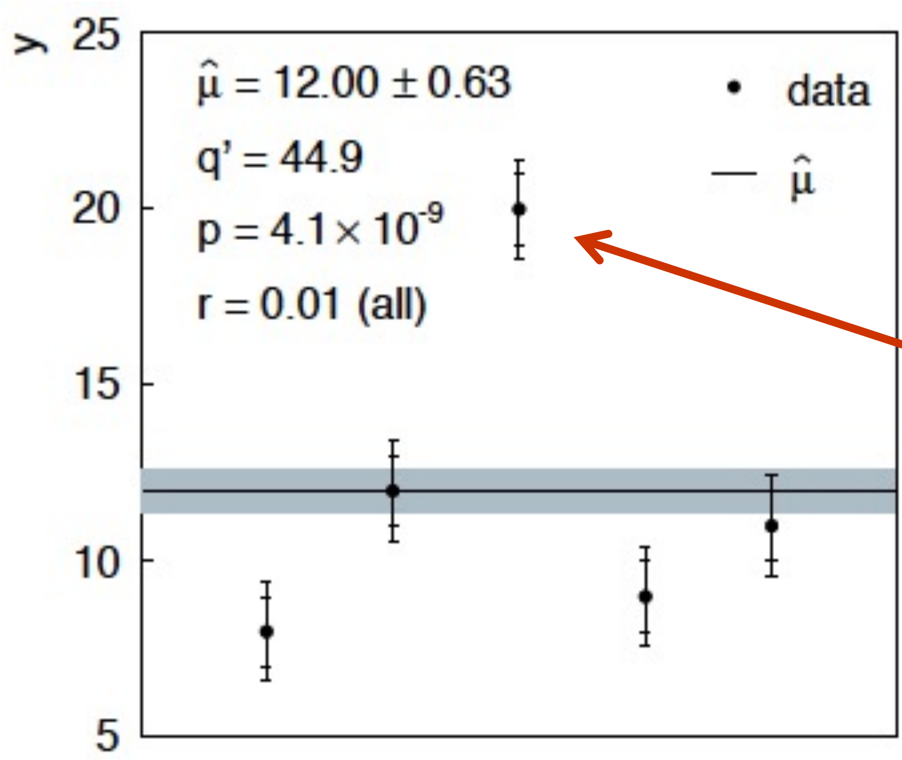
inner error bars = $\sigma_{y,i}$

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 was actually at 20:

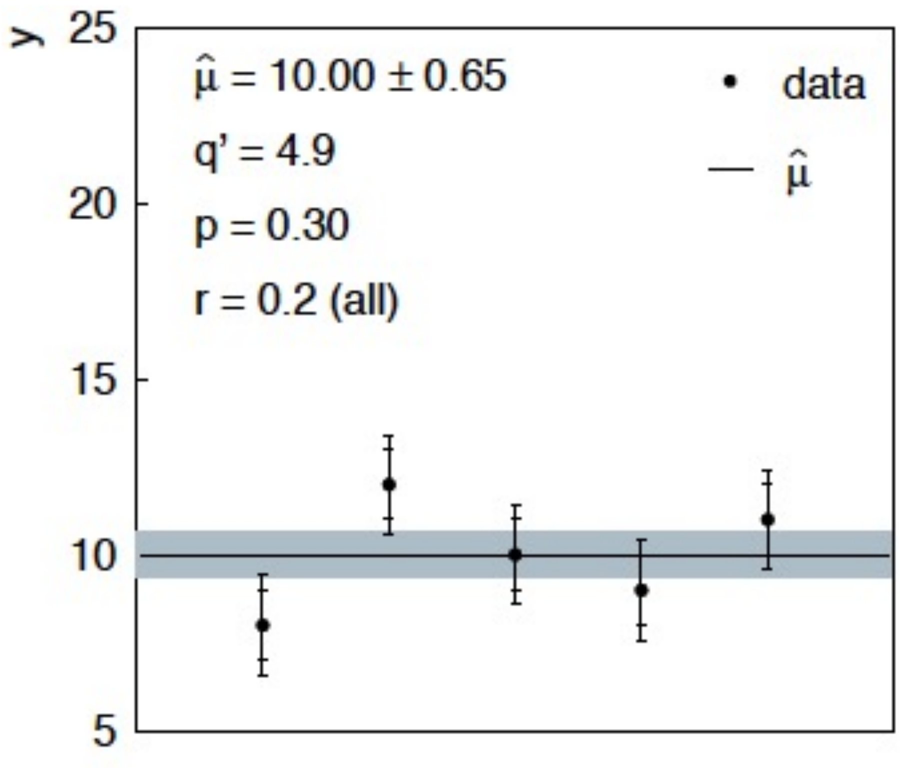


"outlier"

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

Average with all $r = 0.2$

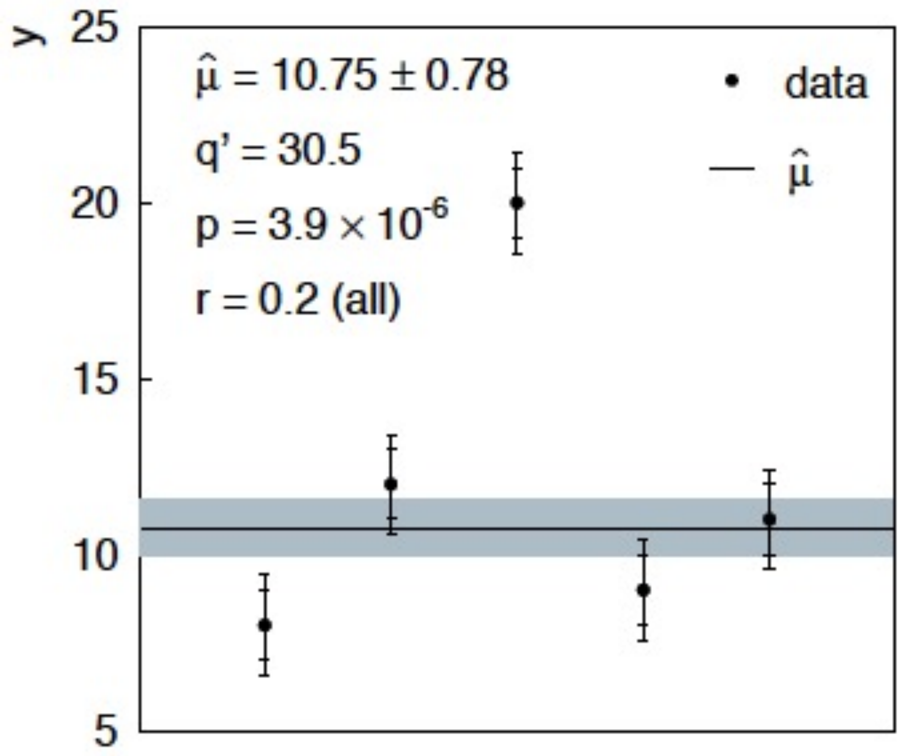
If we assign to each measurement $r = 0.2$,



Estimate still at 10.00, size of interval moves $0.63 \rightarrow 0.65$

Average with all $r = 0.2$ with outlier

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



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

Half-size of interval $\to 0.78$ (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)

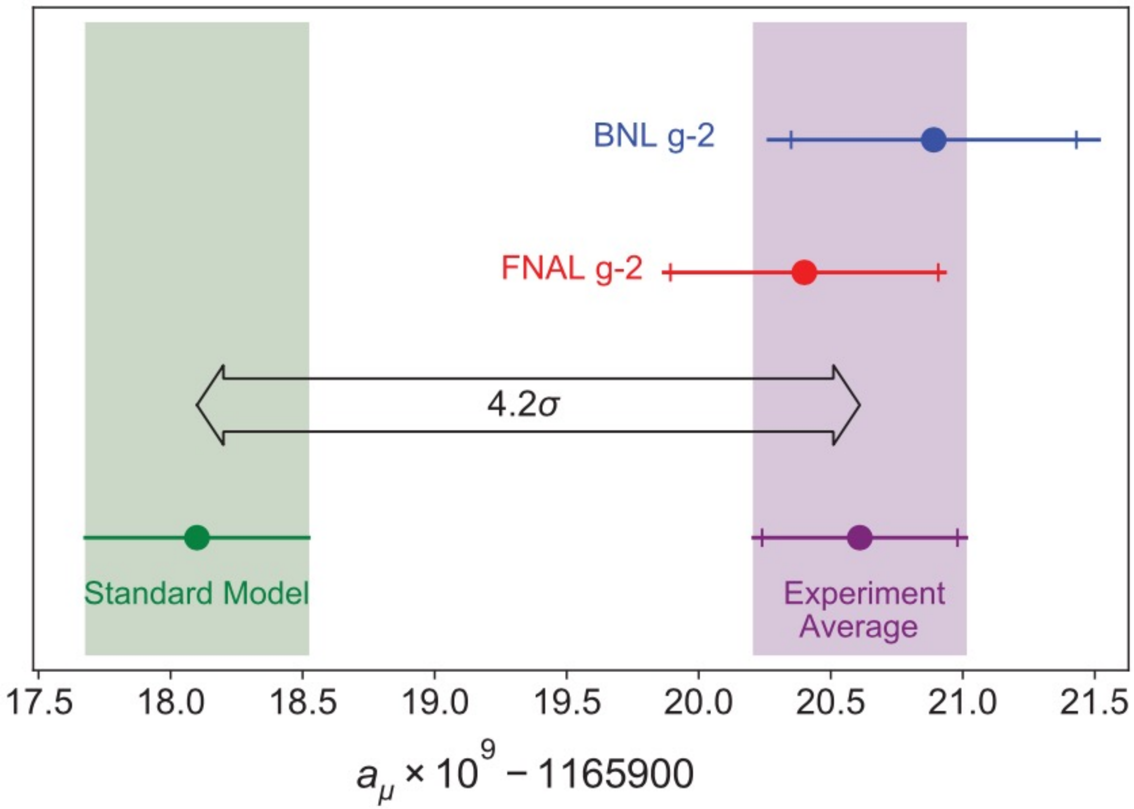
$\hat{\mu} = 12.00 \pm 0.70$ with outlier (middle meas. 20)

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

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

the ingredients of the 4.2σ effect are:

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

(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).

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

(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 σ_{SM} uncertain

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

This is the largest systematic and probably hardest to estimate.

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

Work ongoing with Bogdan Malaescu of Muon g-2 Theory Initiative on the HVP uncertainty, see, e.g., M. Davier et al., Eur. Phys. J. C 80 (2020) 241 , arXiv:1908.00921

Maximum-likelihood for mean from minimum of

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

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).

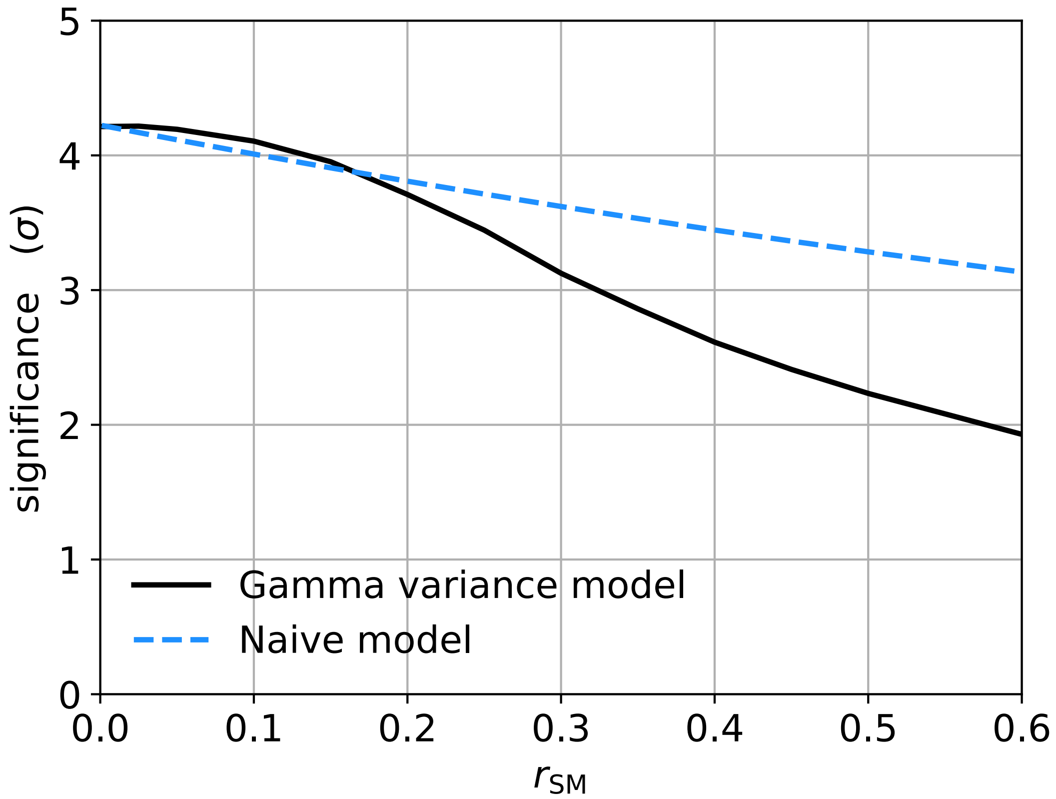
For $r_{\rm SM} > 0$ distribution of q depends on $\sigma^2_{\rm SM}$. For MC use Maximum-Likelihood estimate ("profile construction"):

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

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

# of sigmas

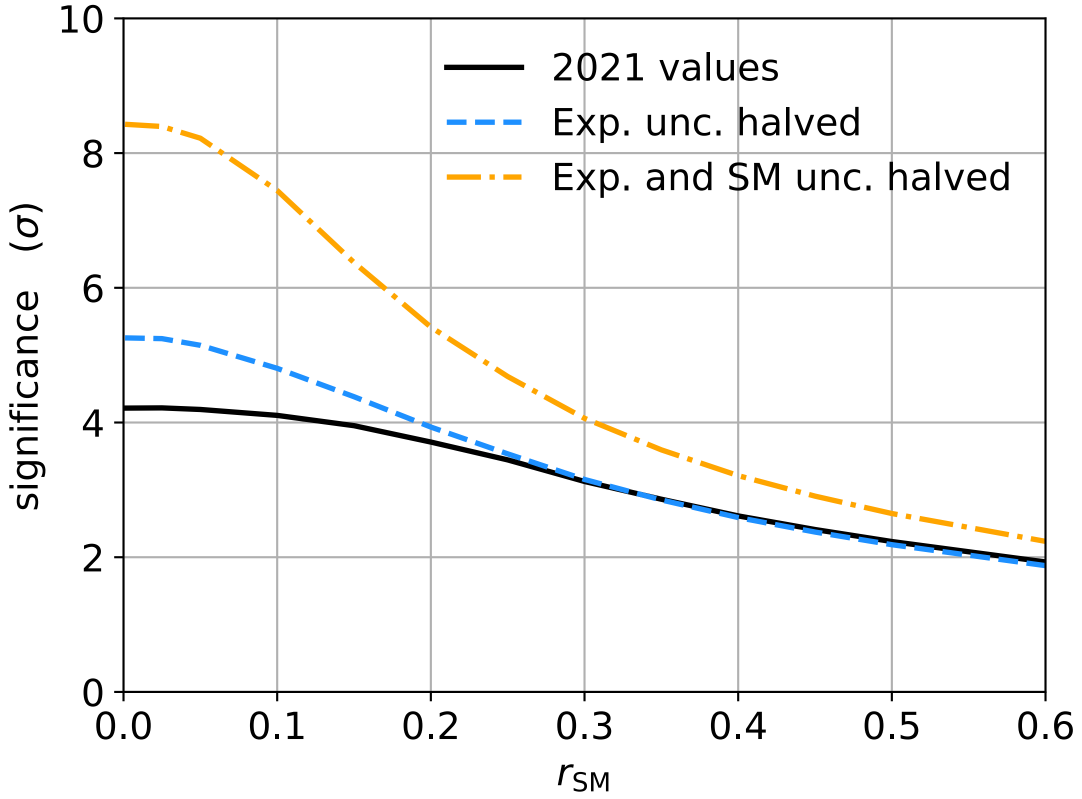
Significance of discrepancy versus $r_{\rm SM}$



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

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

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

Gamma model for variance estimates gives confidence intervals that increase in size when the data are internally inconsistent, and gives decreased sensitivity to outliers (known property of Student's t based regression).

Equivalence with Student's t model, $\nu = 1/2r^2$ degrees of freedom.

Simple profile likelihood – quadratic terms replaced by logarithmic:

$$\frac{(u_i - \theta_i)^2}{\sigma_{u_i}^2} \quad \rightarrow \quad \left(1 + \frac{1}{2r_i^2}\right) \ln\left[1 + 2r_i^2 \frac{(u_i - \theta_i)^2}{v_i}\right]$$

Discussion / Conclusions (2)

Method assumes that meaningful r_i values can be assigned and is valuable when systematic errors are not well known but enough "expert opinion" is available to do so.

Alternatively, one could try to fit a global r to all systematic errors, analogous to PDG scale factor method or meta-analysis à la DerSimonian and Laird. ($\rightarrow$ future work).

Could also use e.g. as "stress test" – crank up the r_i values until significance of result degrades and ask if you really trust the assigned systematic errors at that level.

Ongoing studies with Alessandra Brazzale (higher-order asymptotics), Enzo Canonero (ATLAS, top-quark mass), Bogdan Malaescu (muon $\mathrm{g}-2$).

Extra slides

Curve Fitting History: Least Squares

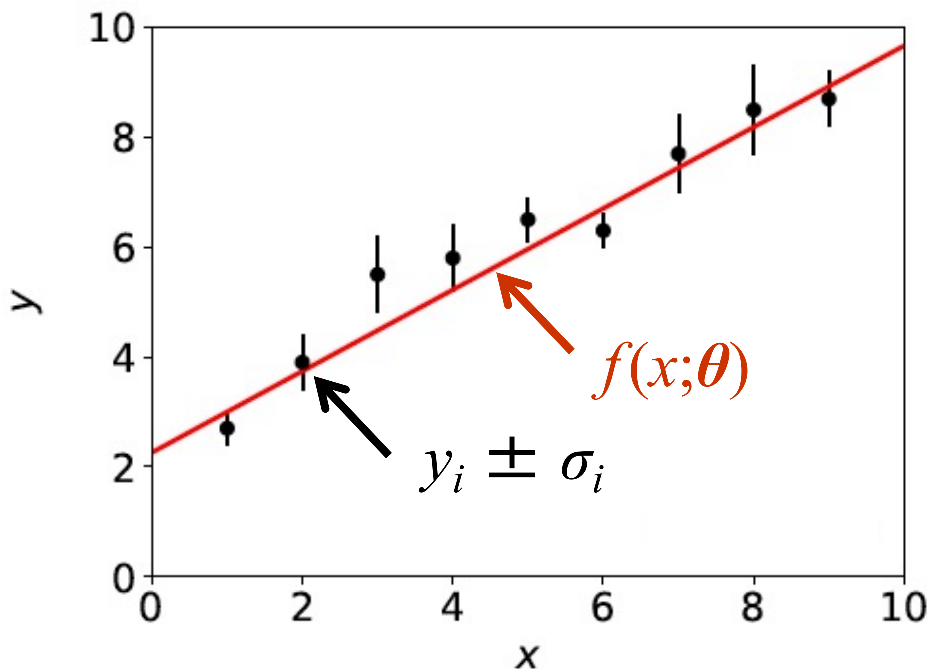
Method of Least Squares by Laplace, Gauss, Legendre, Galton...

C.F. Gauss, Theoria Combinationis Observationum Erroribus Minimis Obnoxiae, Commentationes Societatis Regiae Scientiarium Gottingensis Recectiores Vol. V (MDCCCXXIII).

To fit curve $f(x;\boldsymbol{\theta})$ to data $y_i \pm \sigma_i$, adjust parameters $\boldsymbol{\theta} = (\theta_1,\ldots,\theta_M)$ to minimize

$$\chi^2(\boldsymbol{\theta}) = \sum_{i=1}^{N} \frac{(y_i - f(x_i;\boldsymbol{\theta}))^2}{\sigma_i^2}$$

Assumes σ_i known.



Developments of LS for Averaging

Much work in HEP and elsewhere on application/extension of least squares to the problem of averaging or meta-analysis, e.g.,

A. C. Aitken, *On Least Squares and Linear Combinations of Observations*, Proc. Roy. Soc. Edinburgh **55** (1935) 42.

L. Lyons, D. Gibaut and P. Clifford, *How to Combine Correlated Estimates of a Single Physical Quantity*, Nucl. Instr. Meth. **A270** (1988) 110.

A. Valassi, *Combining Correlated Measurements of Several Different Physical Quantities*, Nucl. Instr. Meth. **A500** (2003) 391.

R. Nisius, *On the combination of correlated estimates of a physics observable*, Eur. Phys. J. C **74** (2014) 3004.

R. DerSimonian and N. Laird, *Meta-analysis in clinical trials*, Controlled Clinical Trials **7** (1986) 177-188.

"Errors on Errors"

APRIL 15, 1932 *PHYSICAL REVIEW* *VOLUME 40*

THE CALCULATION OF ERRORS BY THE METHOD OF LEAST SQUARES

By Raymond T. Birge
University of California, Berkeley

(Received February 18, 1932)

Abstract

Present status of least squares' calculations.—There are three possible stages in any least squares' calculation, involving respectively the evaluation of (1) the most probable values of certain quantities from a set of experimental data, (2) the reliability or probable error of each quantity so calculated, (3) the reliability or probable error of the probable errors so calculated. Stages (2) and (3) are not adequately treated in most texts, and are frequently omitted or misused, in actual work. The present article is concerned mainly with these two stages.

$\rightarrow$ PDG "scale factor method" $\approx$ scale sys. errors with common factor until $\chi^2_{\rm min}$ = appropriate no. of degrees of freedom.

Least Squares ← Maximum Likelihood

Method of Least Squares follows from method of Maximum Likelihood if independent measured $y_i \sim$ Gaussian($f(x_i;\boldsymbol{\theta})$, σ_i)

$$P(\mathbf{y}|\boldsymbol{\theta}) = L(\boldsymbol{\theta}) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma_i} e^{-(y_i - f(x_i;\boldsymbol{\theta}))^2/2\sigma_i^2}$$

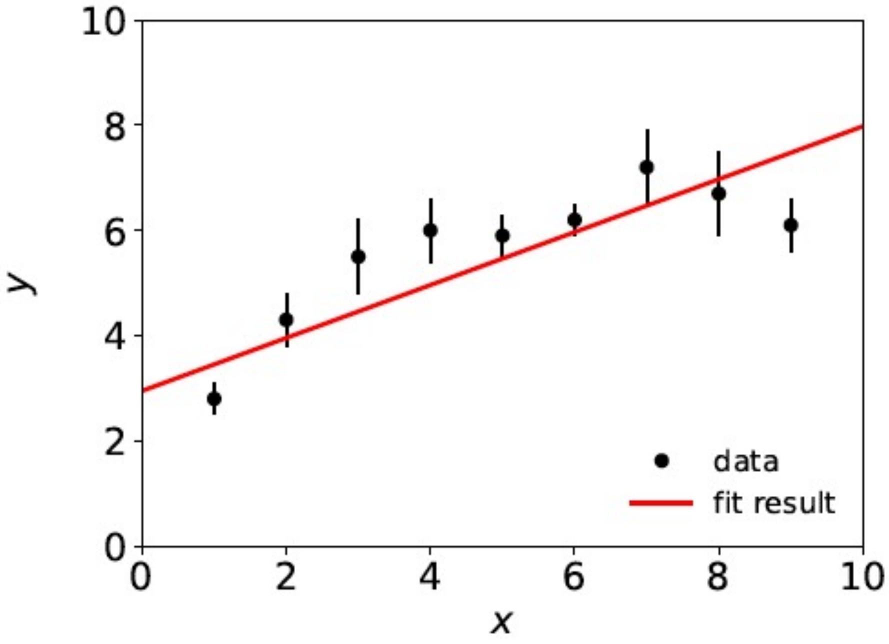
$$-2\ln L(\boldsymbol{\theta}) = \sum_{i=1}^{N} \frac{(y_i - f(x_i;\boldsymbol{\theta}))^2}{\sigma_i^2} + \text{const.}$$

Tails of Gaussian fall off very fast; points away from the curve ("outliers") have strong influence on parameter estimates.

Goodness of fit

If the hypothesized model $f(x;\boldsymbol{\theta})$ is correct, $\chi^2_{\min}$ should follow a chi-square distribution for N (# meas.) – M (# fitted par.) degrees of freedom; expectation value = $N - M$.

Suppose initial guess for model is: $f(x;\boldsymbol{\theta}) = \theta_0 + \theta_1 x$



$\chi^2_{\min} = 20.9$,
$N - M = 9 - 2 = 7$,
so goodness of fit is “poor”.

This is an indication that the model is inadequate, and thus the values it predicts will have a “systematic error”.

Systematic errors ↔ nuisance parameters

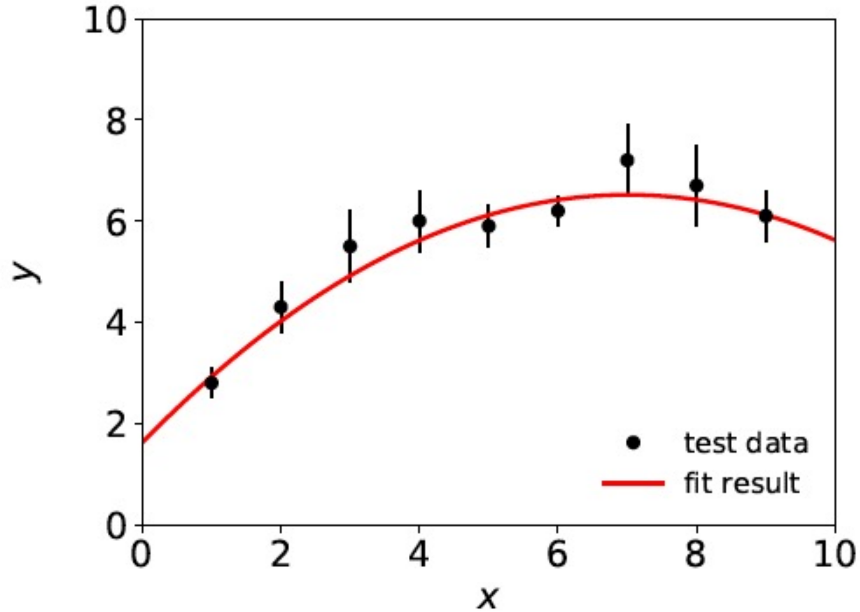
Solution: fix the model, generally by inserting additional adjustable parameters (“nuisance parameters”). Try, e.g.,

$$f(x;\boldsymbol{\theta}) = \theta_0 + \theta_1 x + \theta_2 x^2$$

$\chi^2_{\min} = 3.5,\ N - M = 6$

For some point in the enlarged parameter space we hope the model is now ~correct.

Sys. error gone?

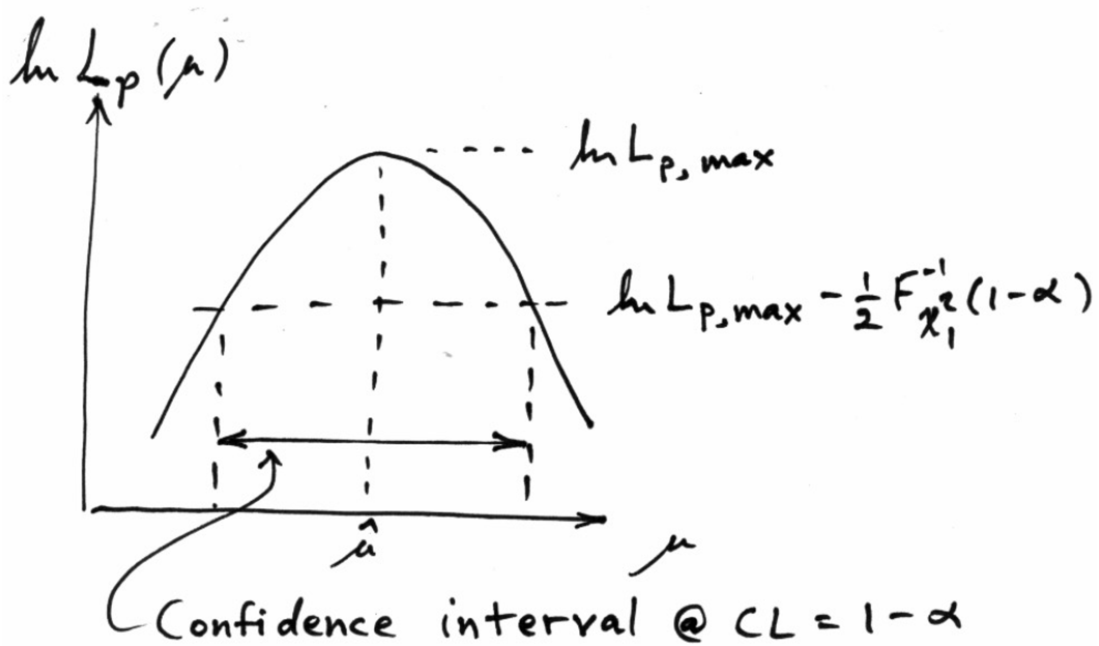


Estimators for all parameters correlated, and as a consequence the presence of the nuisance parameters inflates the statistical errors of the parameter(s) of interest.

Uncertainty of fitted parameters

Suppose parameter of interest μ, nuisance parameter θ, *confidence interval* for μ from "profile likelihood":

$$L_{\mathrm{p}}(\mu) = L(\mu, \hat{\hat{\theta}}) \qquad \hat{\hat{\theta}}(\mu) = \underset{\theta}{\operatorname{argmax}}\, L(\mu, \theta)$$



Width of interval in usual LS fit *independent* of goodness of fit.

Single-measurement model

As a simplest example consider

$y \sim \text{Gauss}(\mu, \sigma^2)$,

$v \sim \text{Gamma}(\alpha, \beta)$, $\quad \alpha = \frac{1}{4r^2}, \quad \beta = \frac{1}{4r^2\sigma^2}$

$$L(\mu, \sigma^2) = f(y, v|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-\mu)^2/2\sigma^2} \frac{\beta^\alpha}{\Gamma(\alpha)} v^{\alpha-1} e^{-\beta v}$$

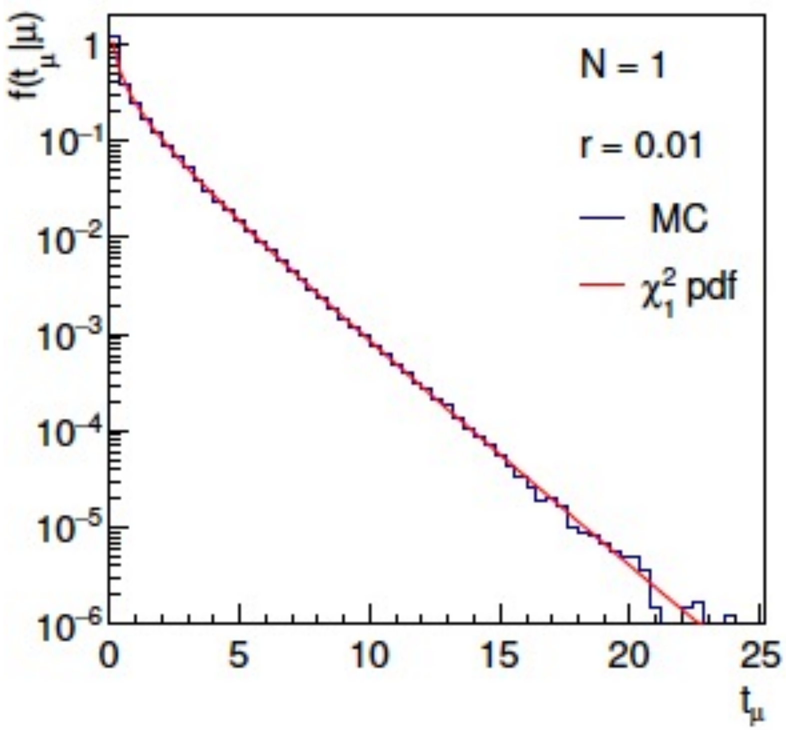
Test values of μ with $t_\mu = -2\ln\lambda(\mu)$ with $\quad \lambda(\mu) = \frac{L(\mu, \widehat{\widehat{\sigma^2}}(\mu))}{L(\hat{\mu}, \widehat{\sigma^2})}$

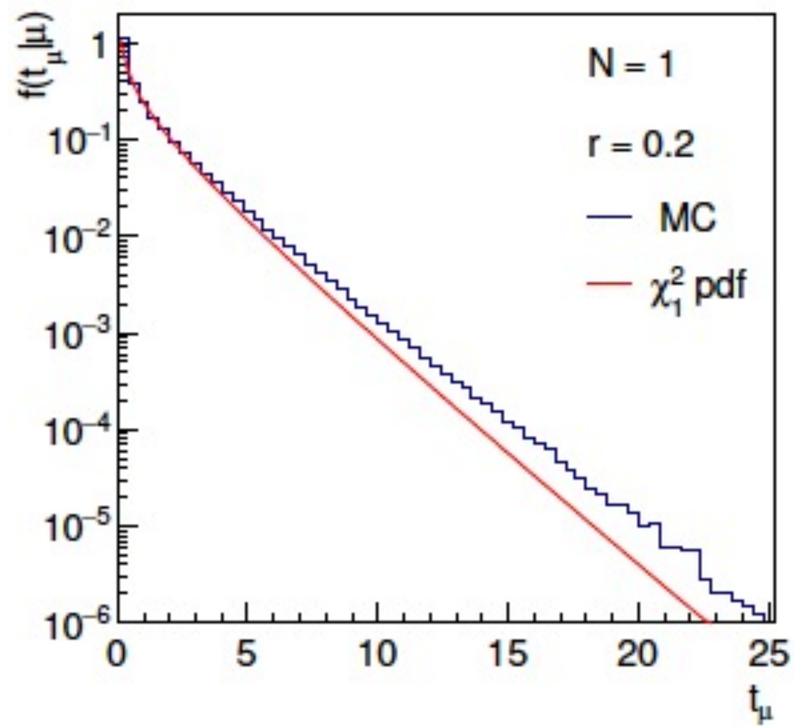
$$t_\mu = \left(1 + \frac{1}{2r^2}\right) \ln\left[1 + 2r^2 \frac{(y-\mu)^2}{v}\right]$$

Distribution of t_μ

From Wilks' theorem, in the asymptotic limit we should find $t_\mu \sim$ chi-squared(1).

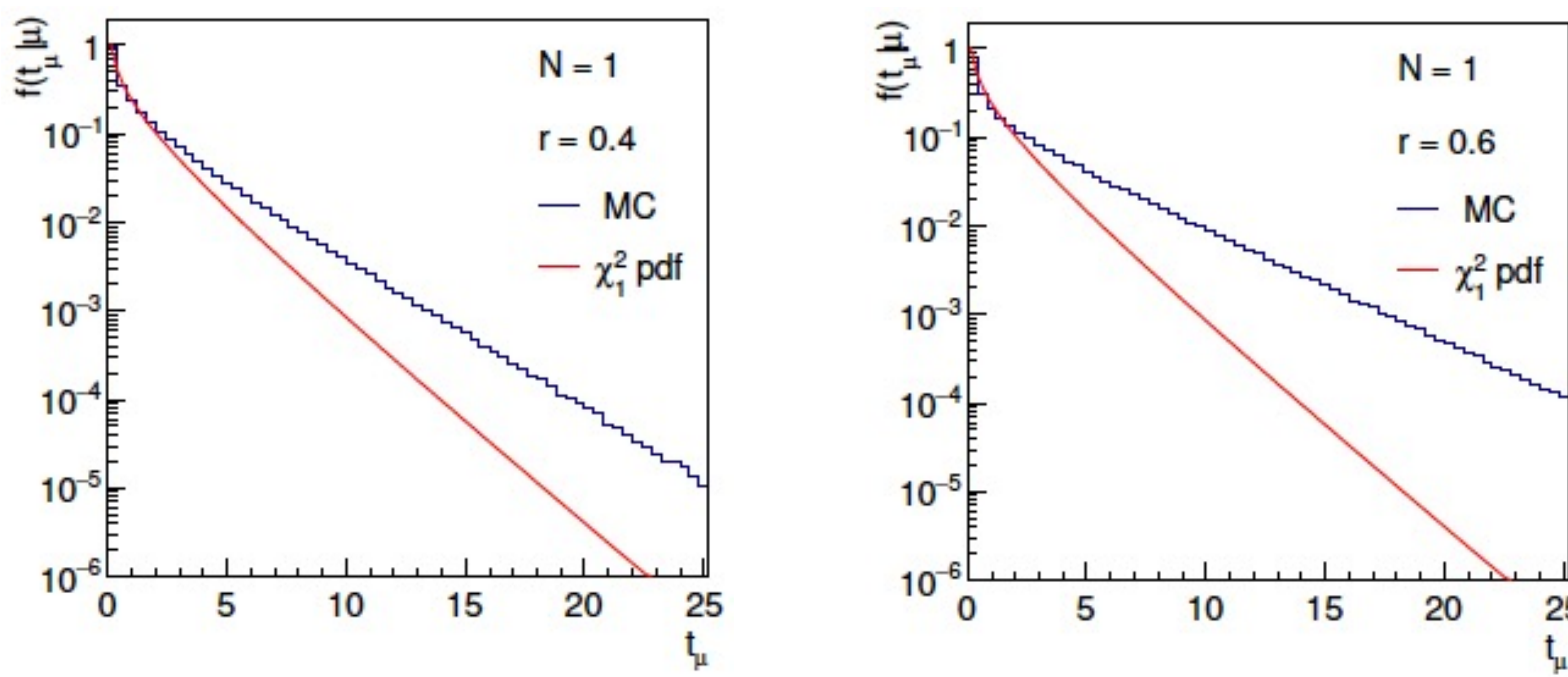
Here "asymptotic limit" means all estimators ~Gauss, which means $r \to 0$. For increasing r, clear deviations visible:





Distribution of t_μ (2)

For larger r, breakdown of asymptotics gets worse:



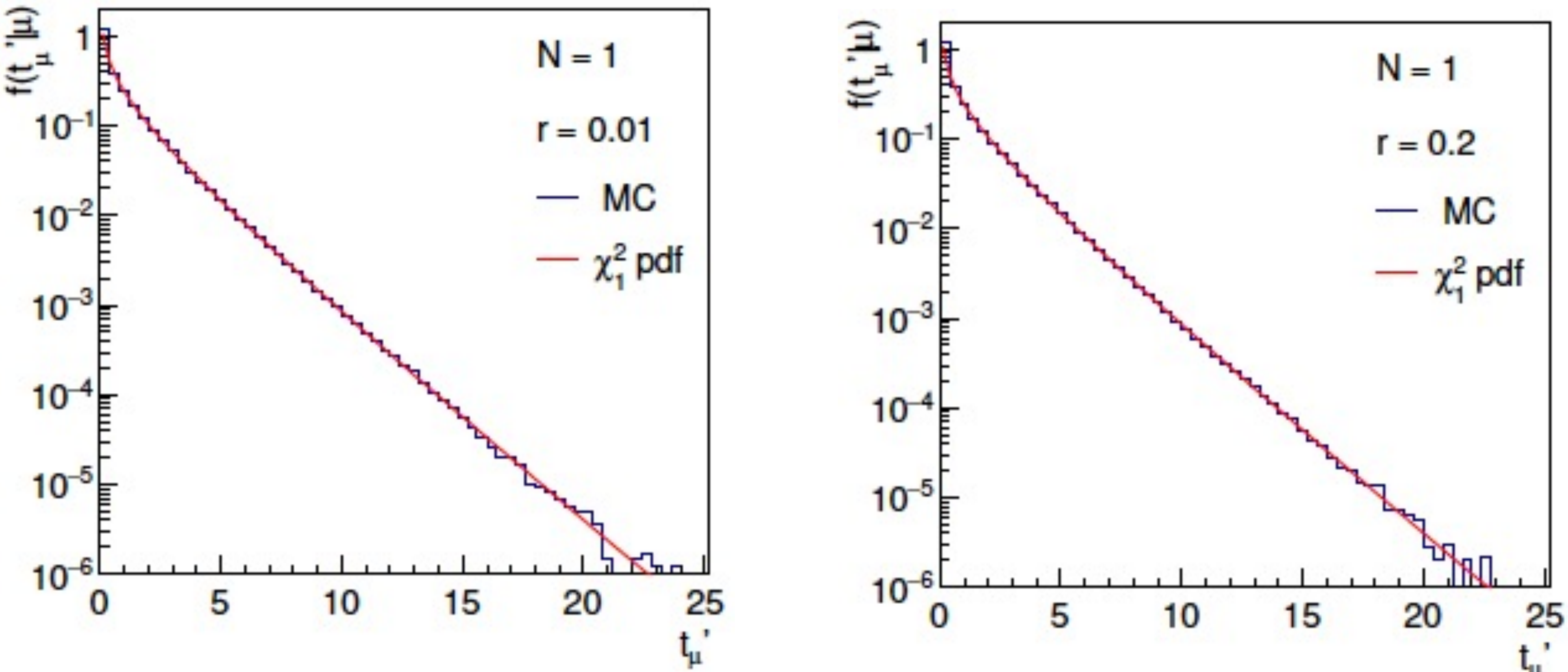
Values of $r \sim$ several tenths are relevant so we cannot in general rely on asymptotics to get confidence intervals, p-values, etc.

Bartlett corrections

One can modify t_μ defining $$t'_\mu = \frac{n_\mathrm{d}}{E[t_\mu]} t_\mu$$

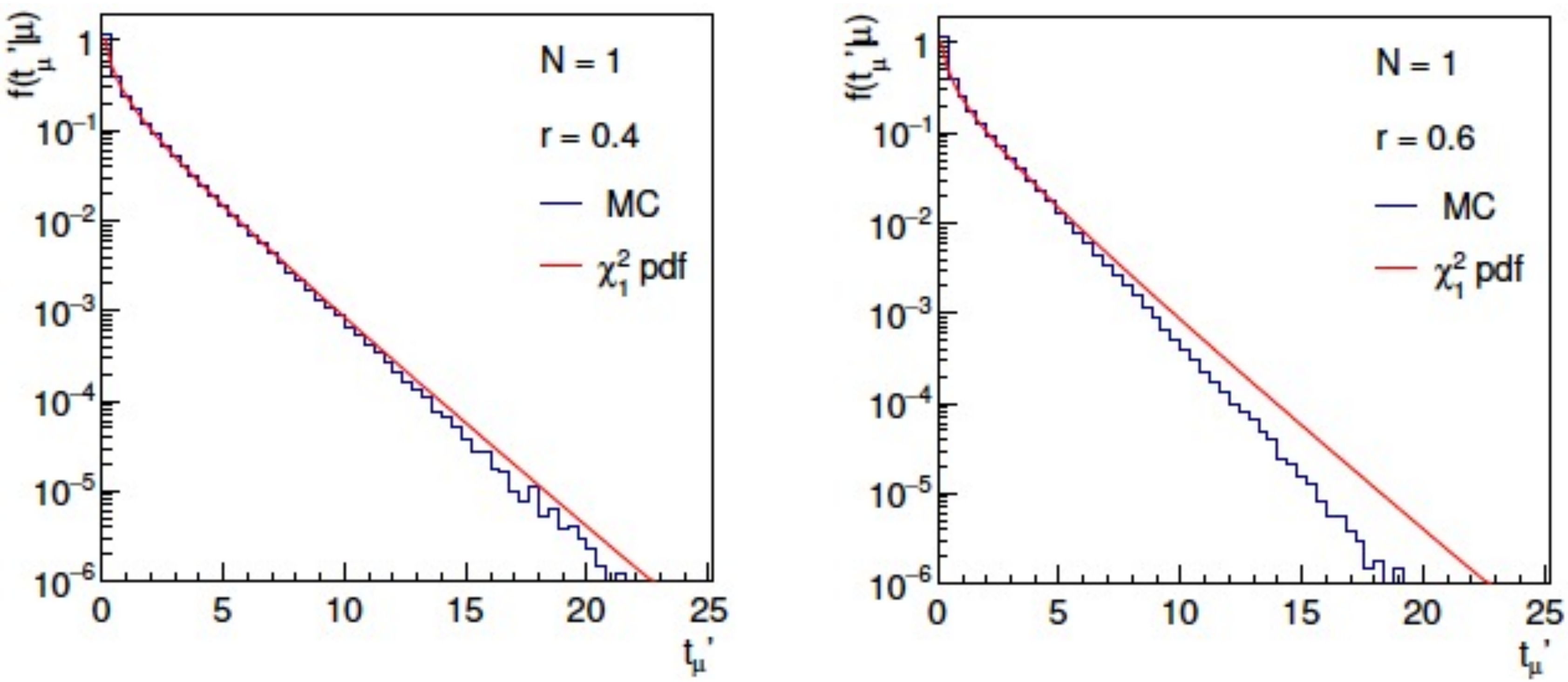
such that the new statistic's distribution is better approximated by chi-squared for n_d degrees of freedom (Bartlett, 1937).

For this example $E[t_\mu] \approx 1 + 3r^2 + 2r^4$ works well:

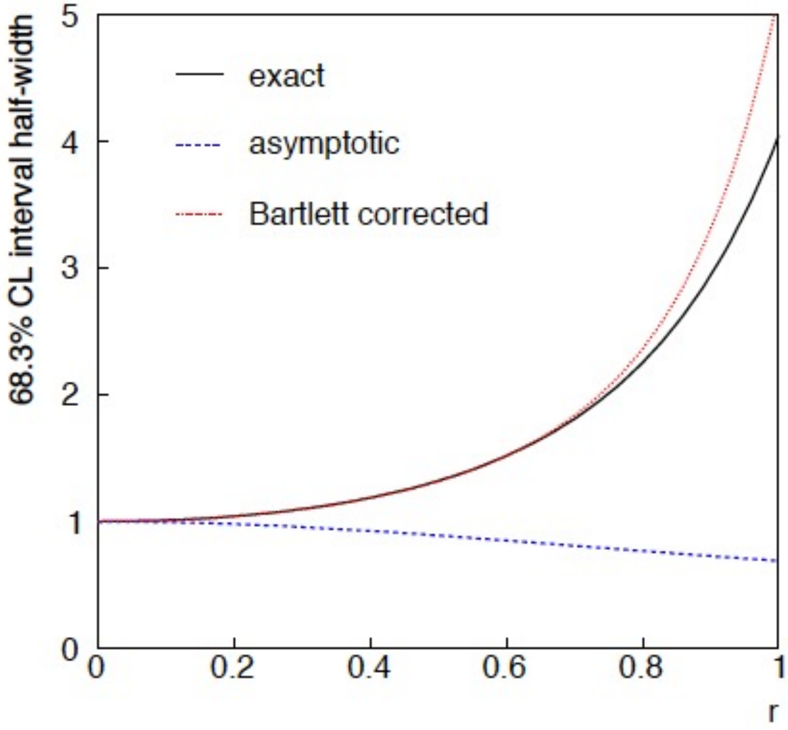


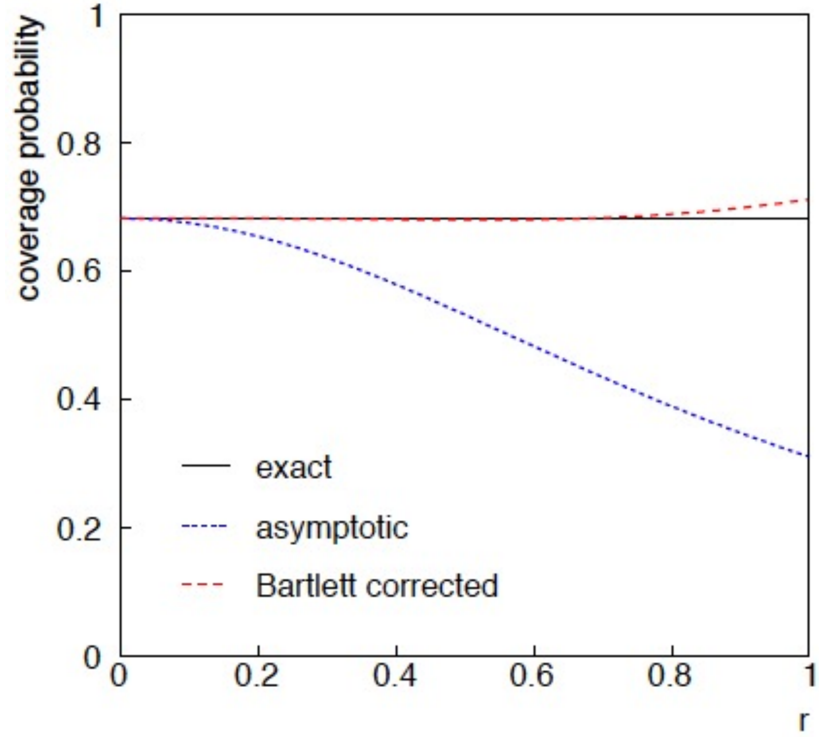
Bartlett corrections (2)

Good agreement for $r \sim$ several tenths out to $\sqrt{t_\mu}' \sim$ several, i.e., good for significances of several sigma:



68.3% CL confidence interval for μ





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.

Exact relation between *r* parameter and relative error on error

r parameter defined as: $$r \equiv \frac{1}{2}\frac{\sigma_v}{E[v]} \approx \frac{\sigma_s}{E[s]}$$

$v \sim$ Gamma(α, β) so $s = \sqrt{v}$ follows a Nakagami distribution

$$g(s|\alpha,\beta) = \left|\frac{dv}{ds}\right| f(v(s)|\alpha,\beta) = \frac{2\beta^\alpha}{\Gamma(\alpha)} s^{2\alpha-1} e^{-\beta s^2}$$

$$E[s] = \frac{\Gamma(\alpha+\frac{1}{2})}{\Gamma(\alpha)\sqrt{\beta}}$$

$$V[s] = \frac{\alpha}{\beta} - \frac{1}{\beta}\left(\frac{\Gamma(\alpha+\frac{1}{2})}{\Gamma(\alpha)}\right)^2$$

Exact relation between *r* parameter and relative error on error (2)

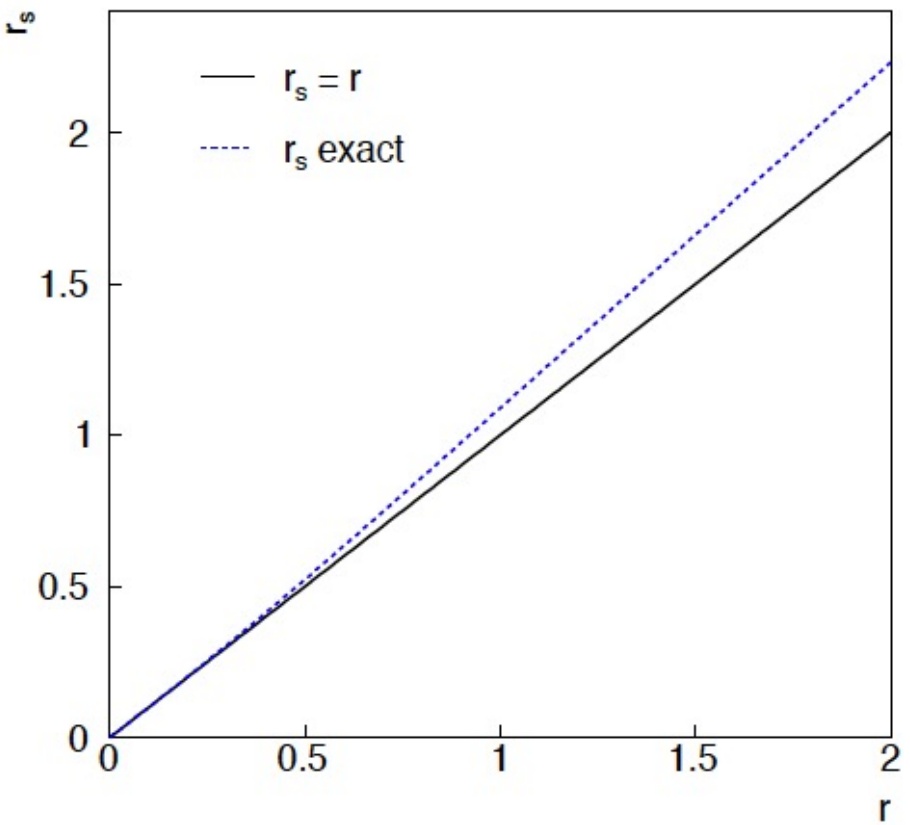
The exact relation between the error and the error r_s and the parameter r is therefore

$$r_s \equiv \frac{\sqrt{V[s]}}{E[s]}$$

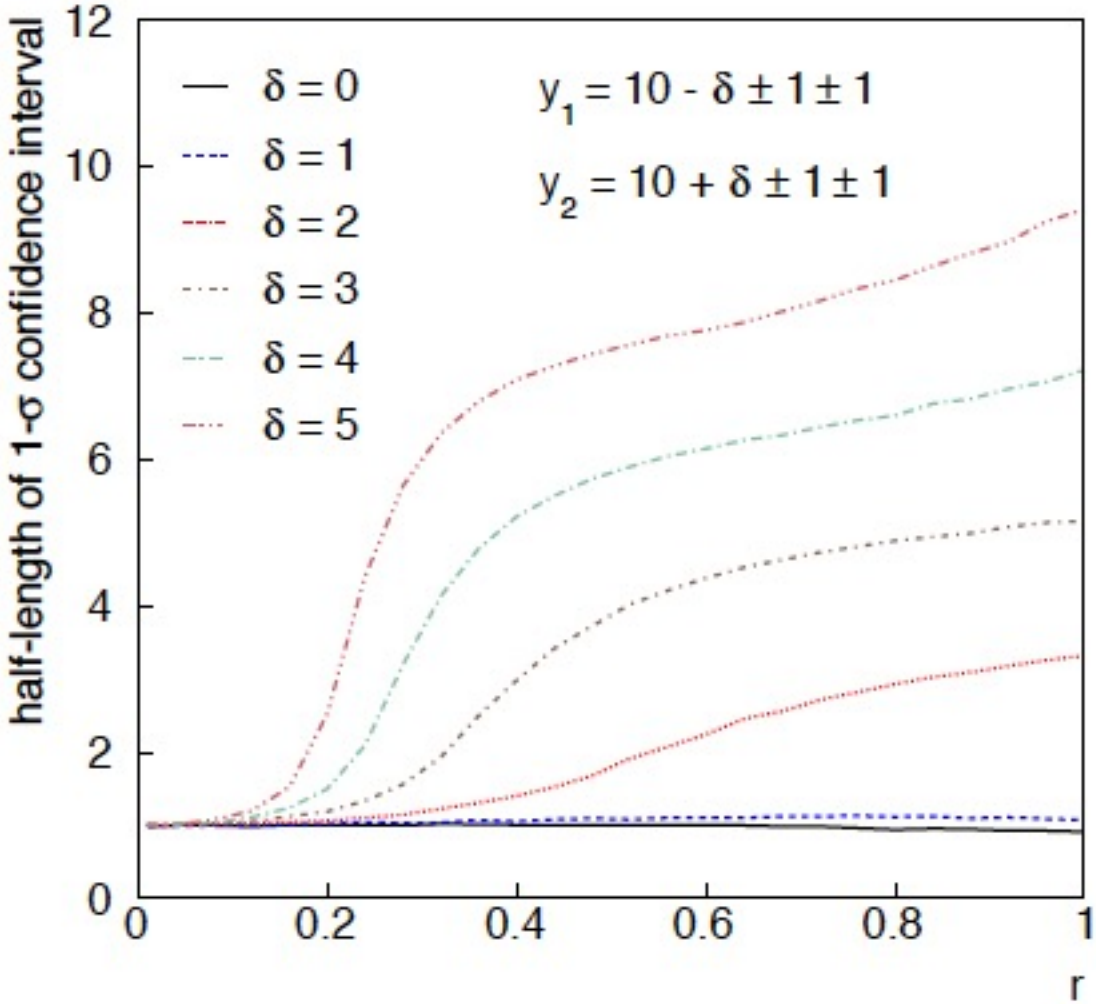
$$= \sqrt{\alpha \left(\frac{\Gamma(\alpha)}{\Gamma(\alpha + \frac{1}{2})} \right)^2 - 1}$$

$$\alpha = 1/4r^2$$

$\rightarrow$ $r_s \approx r$ good for $r \lesssim 1$.



Same with interval from $p_\mu = \alpha$ with nuisance parameters profiled at μ



Coverage of intervals

Consider previous average of two numbers but now generate for $i = 1, 2$ data values

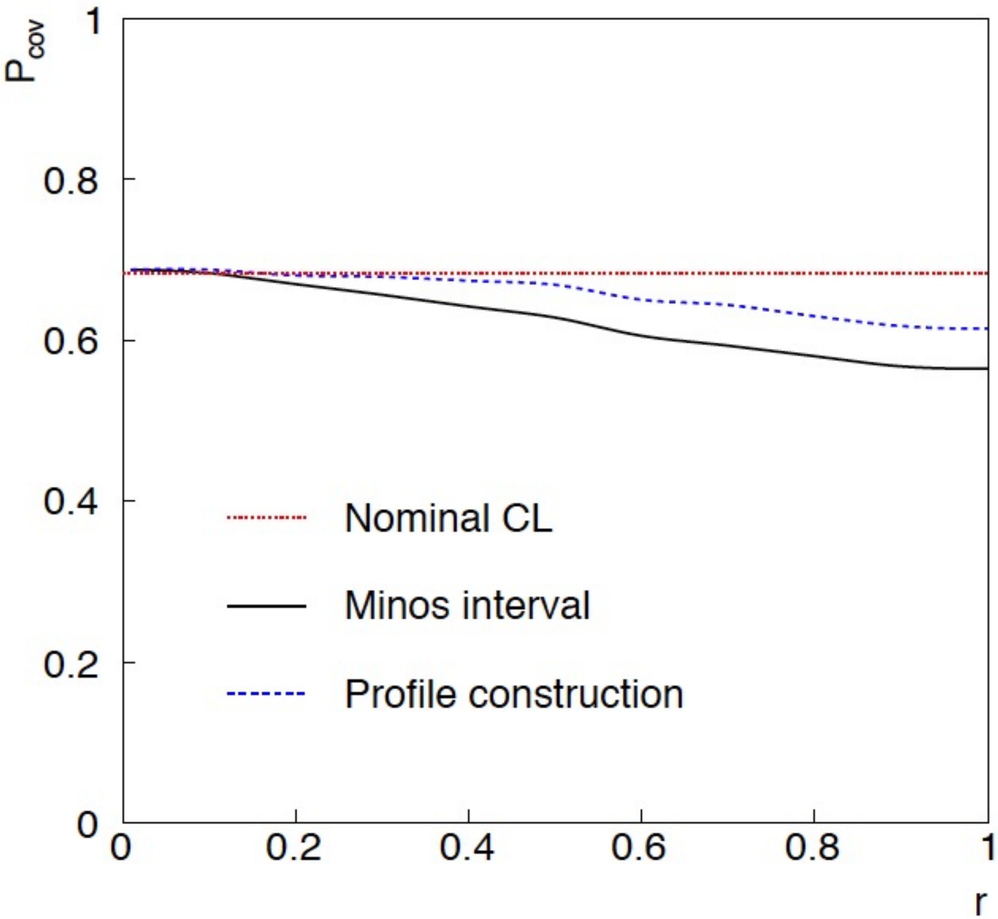
$y_i \sim$ Gauss(μ, $\sigma_{y,i}$)

$u_i \sim$ Gauss(0, $\sigma_{u,i}$)

$v_i \sim$ Gamma($\sigma_{u,i}$, r_i)

$\sigma_{y,i} = \sigma_{u,i} = 1$

and look at the probability that the interval covers the true value of μ.



Coverage stays reasonable to $r \sim 0.5$, even not bad for Profile Construction out to $r \sim 1$.

PDG scale factor

Suppose we do not want to take the quoted errors as known constants. Scale the variances by a factor ϕ,

$$\sigma_i^2 \to \phi \sigma_i^2$$

The likelihood function becomes

$$L(\mu, \phi) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\phi\sigma_i^2}} \exp\left[-\frac{1}{2} \frac{(y_i - \mu)^2}{\phi\sigma_i^2} \right]$$

The estimator for μ is the same as before; for ϕ the MLE is

$$\hat{\phi}_{\mathrm{ML}} = \frac{\chi^2(\hat{\mu})}{N}$$ which has a bias; $$\hat{\phi} = \frac{\chi^2(\hat{\mu})}{N-1}$$ is unbiased.

The variance of $\hat{\mu}$ is inflated by ϕ:

$$V[\hat{\mu}] = \frac{\phi}{\sum_{i=1}^{N} \frac{1}{\sigma_i^2}}$$

Bayesian approach

G. Cowan, *Bayesian Statistical Methods for Parton Analyses*, in *Proceedings of the 14th International Workshop on Deep Inelastic Scattering (DIS2006)*, M. Kuze, K. Nagano, and K. Tokushuku (eds.), Tsukuba, 2006.

Given measurements: $y_i \pm \sigma_i^{\text{stat}} \pm \sigma_i^{\text{sys}}, \quad i = 1, \dots, n$,

and (usually) covariances: $V_{ij}^{\text{stat}}, \; V_{ij}^{\text{sys}}$.

Predicted value: $\mu(x_i; \theta)$, expectation value $E[y_i] = \mu(x_i; \theta) + b_i$

control variable parameters bias

Frequentist approach: $V_{ij} = V_{ij}^{\text{stat}} + V_{ij}^{\text{sys}}$

Minimize $\chi^2(\theta) = (\vec{y} - \vec{\mu}(\theta))^T V^{-1} (\vec{y} - \vec{\mu}(\theta))$

Its Bayesian equivalent

Take $$L(\vec{y}|\vec{\theta},\vec{b}) \sim \exp\left[-\frac{1}{2}(\vec{y}-\vec{\mu}(\theta)-\vec{b})^T V_{\rm stat}^{-1}(\vec{y}-\vec{\mu}(\theta)-\vec{b})\right]$$

$$\pi_b(\vec{b}) \sim \exp\left[-\frac{1}{2}\vec{b}^T V_{\rm sys}^{-1}\vec{b}\right]$$

$$\pi_\theta(\theta) \sim \text{const.}$$

Joint probability for all parameters

and use Bayes' theorem: $$p(\theta,\vec{b}|\vec{y}) \propto L(\vec{y}|\theta,\vec{b})\pi_\theta(\theta)\pi_b(\vec{b})$$

To get desired probability for θ, integrate (marginalize) over b:

$$p(\theta|\vec{y}) = \int p(\theta,\vec{b}|\vec{y})\,d\vec{b}$$

→ Posterior is Gaussian with mode same as least squares estimator, σ_θ same as from $\chi^2 = \chi^2_{\rm min} + 1$. (Back where we started!)

Bayesian approach with non-Gaussian prior $\pi_b(b)$

Suppose now the experiment is characterized by

$$y_i, \quad \sigma_i^{\text{stat}}, \quad \sigma_i^{\text{sys}}, \quad s_i, \quad i = 1, \ldots, n \ ,$$

where s_i is an (unreported) factor by which the systematic error is over/under-estimated.

Assume correct error for a Gaussian $\pi_b(b)$ would be $s_i \sigma_i^{\text{sys}}$, so

$$\pi_b(b_i) = \int \frac{1}{\sqrt{2\pi} s_i \sigma_i^{\text{sys}}} \exp\left[-\frac{1}{2} \frac{b_i^2}{(s_i \sigma_i^{\text{sys}})^2} \right] \pi_s(s_i) \, ds_i$$

Width of $\sigma_s(s_i)$ reflects 'error on the error'.

Error-on-error function $\pi_s(s)$

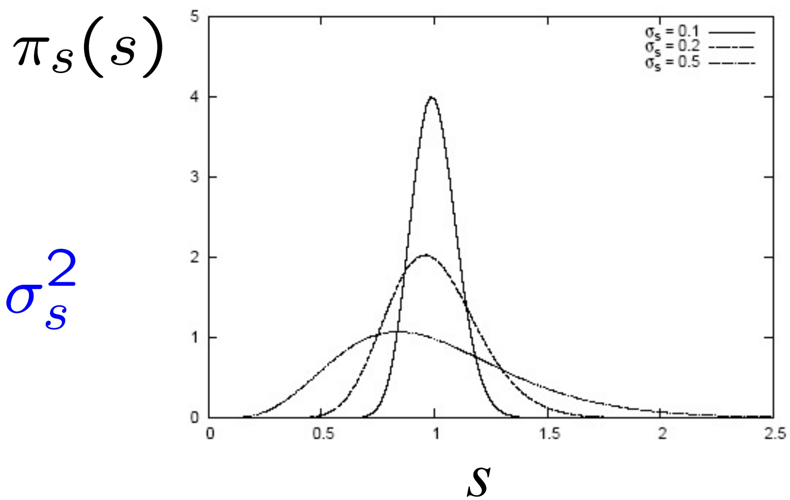
A simple unimodal probability density for $0 < s < 1$ with adjustable mean and variance is the Gamma distribution:

$$\pi_s(s) = \frac{a(as)^{b-1}e^{-as}}{\Gamma(b)}$$

mean $= b/a$

variance $= b/a^2$

Want e.g. expectation value of 1 and adjustable standard Deviation σ_s , i.e., $a = b = 1/\sigma_s^2$

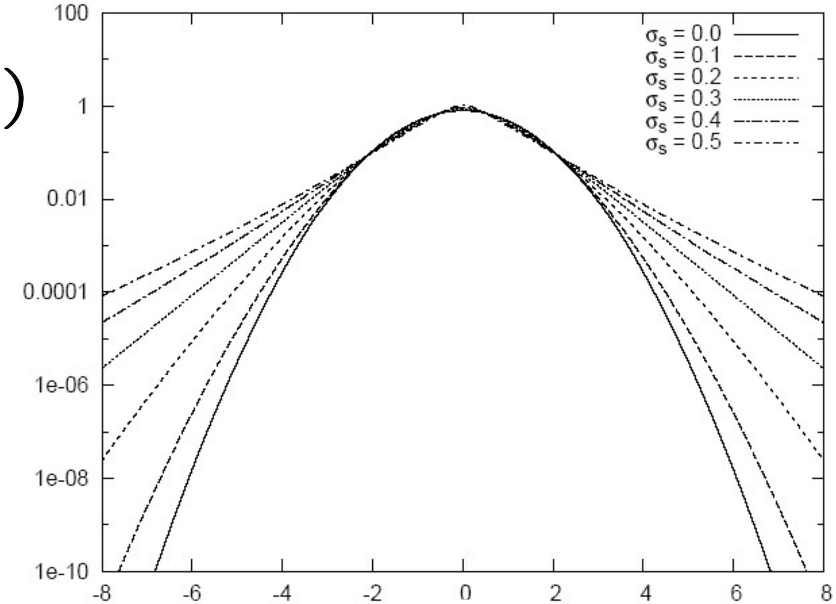


In fact if we took $\pi_s(s) \sim$ *inverse Gamma*, we could find $\pi_b(b)$ in closed form (cf. D'Agostini, Dose, von Linden). But Gamma seems more natural & numerical treatment not too painful.

Prior for bias $\pi_b(b)$ now has longer tails

$$\pi_b(b_i) = \int \frac{1}{\sqrt{2\pi} s_i \sigma_i^{\mathrm{sys}}} \exp\left[-\frac{1}{2}\frac{b_i^2}{(s_i \sigma_i^{\mathrm{sys}})^2}\right] \pi_s(s_i)\, ds_i$$

$\pi_b(b)$



b

Gaussian ($\sigma_s = 0$) $P(|b| > 4\sigma_{\mathrm{sys}}) = 6.3 \times 10^{-5}$

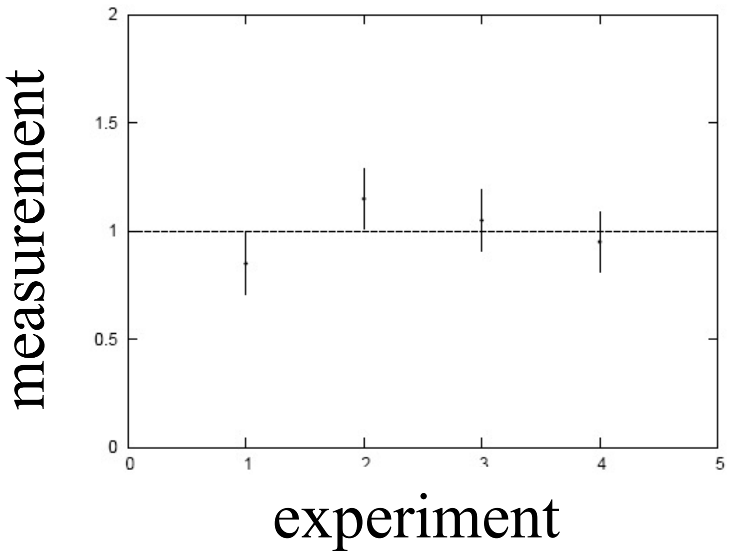
$\sigma_s = 0.5$ $P(|b| > 4\sigma_{\mathrm{sys}}) = 0.65\%$

A simple test

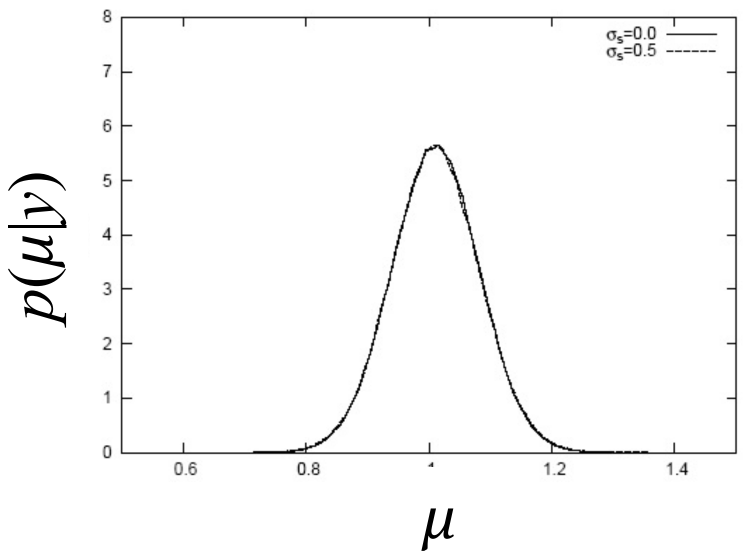
Suppose a fit effectively averages four measurements.

Take $\sigma_{\text{sys}} = \sigma_{\text{stat}} = 0.1$, uncorrelated.

Case #1: data appear compatible



Posterior $p(\mu|y)$:



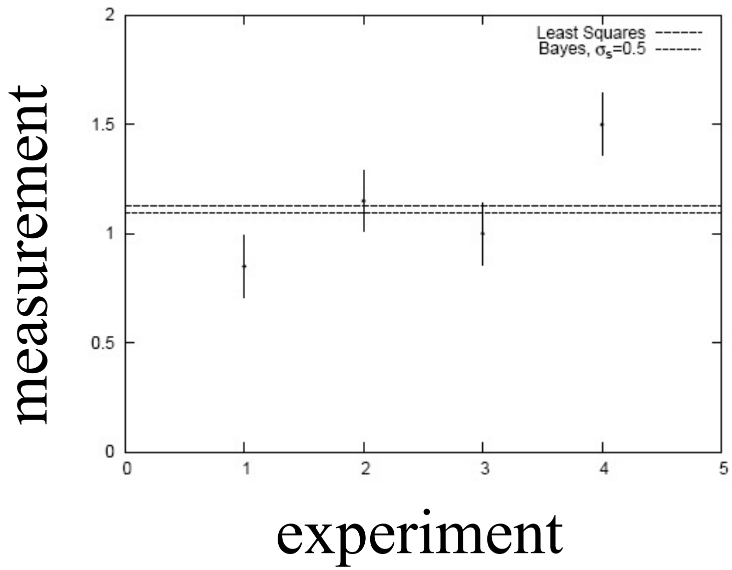
Usually summarize posterior $p(\mu|y)$ with mode and standard deviation:

$$\sigma_s = 0.0: \quad \hat{\mu} = 1.000 \pm 0.071$$

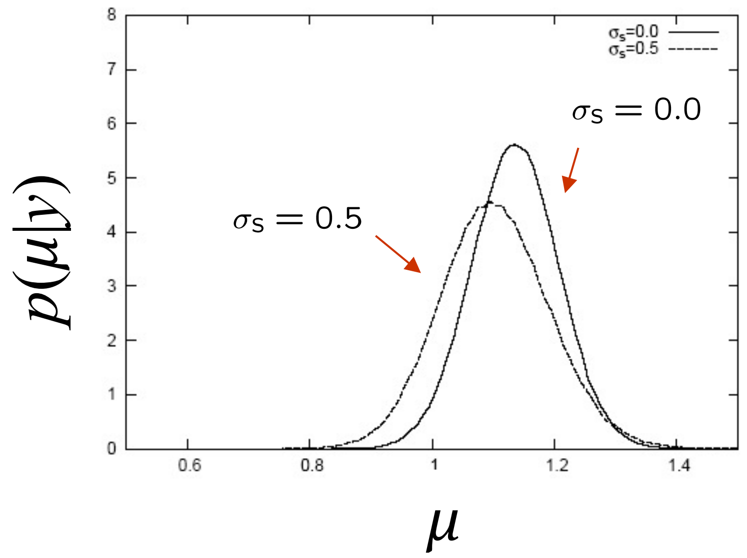
$$\sigma_s = 0.5: \quad \hat{\mu} = 1.000 \pm 0.072$$

Simple test with inconsistent data

Case #2: there is an outlier



Posterior $p(\mu|y)$:



$$\sigma_s = 0.0: \quad \hat{\mu} = 1.125 \pm 0.071$$

$$\sigma_s = 0.5: \quad \hat{\mu} = 1.093 \pm 0.089$$

→ Bayesian fit less sensitive to outlier. See also

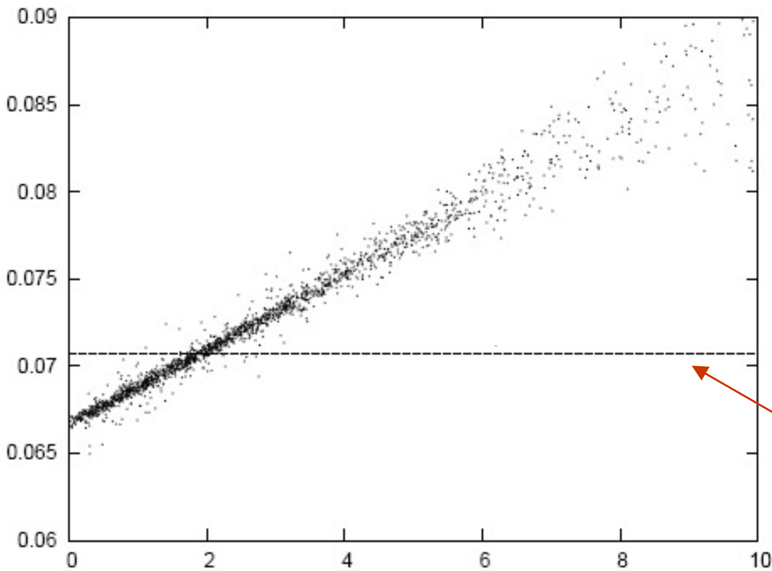
G. D'Agostini, *Sceptical combination of experimental results: General considerations and application to epsilon-prime/epsilon*, arXiv:hep-ex/9910036 (1999).

Goodness-of-fit vs. size of error

In LS fit, value of minimized χ^2 does not affect size of error on fitted parameter.

In Bayesian analysis with non-Gaussian prior for systematics, a high χ^2 corresponds to a larger error (and vice versa).

post-erior σ_μ



χ^2

2000 repetitions of experiment, $\sigma_s = 0.5$, here no actual bias.

σ_μ from least squares