BIRS systematics 2023
26/04/23

ROYAL
HOLLOWAY
UNIVERSITY
OF LONDON

Errors-on-errors

Enzo Canonero

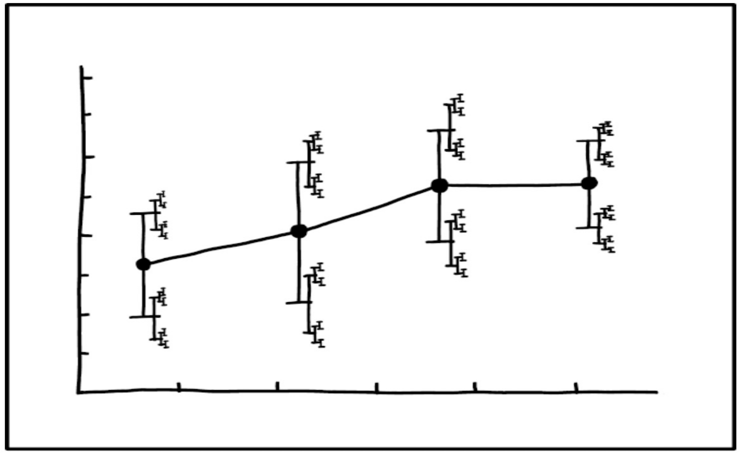
Glen Cowan
Alessandra Brazzale

Motivation

"*Uncertainties on Systematics*"

1) *Some* ***systematic uncertainties*** can be well estimated:
   - **Related to stat. error of control measurements**
   - **Related to size of MC event sample**

2) But they can also be ***quite uncertain***:
   - **Theory systematics *(∼50% relative error)***
   - **Two points systematics *(∼$\frac{1}{\sqrt{2}}$ ≅70% relative error)***



https://xkcd.com/2110/

Formulation of the problem

- Suppose measurements $\boldsymbol{y}$ have a probability density $P(\boldsymbol{y}|\boldsymbol{\mu},\boldsymbol{\theta})$
  - $\boldsymbol{\mu}$ = Parameters of interest
  - $\boldsymbol{\theta}$ = Nuisance parameters

- Auxiliary Measurements $\boldsymbol{u}$ are used to provide info on nuisance parameters and are (often) assumed to be independently Gaussian distributed

- The resulting Likelihood is:

Can be a real measurement or just our best guess based on theoretical reasons

$$L(\boldsymbol{\mu},\boldsymbol{\theta}) = P(\boldsymbol{y},\boldsymbol{u}|\boldsymbol{\mu},\boldsymbol{\theta}) = P(\boldsymbol{y}|\boldsymbol{\mu},\boldsymbol{\theta})\times\prod_i \frac{1}{\sqrt{2\pi}\sigma_{u_i}}\, e^{-(u_i-\theta_i)^2/2\sigma_{u_i}^2}$$

Formulation of the problem

- Suppose measurements $\boldsymbol{y}$ have a probability density $P(\boldsymbol{y}|\boldsymbol{\mu},\boldsymbol{\theta})$
  - $\boldsymbol{\mu}$ = Parameters of interest
  - $\boldsymbol{\theta}$ = Nuisance parameters

- Auxiliary Measurements $\boldsymbol{u}$ are used to provide info on nuisance parameters and are (often) assumed to be independently Gaussian distributed

- The resulting Likelihood is:

Can be a real measurement or just our best guess based on theoretical reasons

$$L(\boldsymbol{\mu},\boldsymbol{\theta}) = P(\boldsymbol{y},\boldsymbol{u}|\boldsymbol{\mu},\boldsymbol{\theta}) = P(\boldsymbol{y}|\boldsymbol{\mu},\boldsymbol{\theta})\times\prod_i \frac{1}{\sqrt{2\pi}\sigma_{u_i}}\, e^{-(u_i-\theta_i)^2/2\sigma_{u_i}^2}$$

- And the log Likelihood:

$$\log L(\boldsymbol{\mu},\boldsymbol{\theta}) = \log P(\boldsymbol{y}|\boldsymbol{\mu},\boldsymbol{\theta})\ -\sum\frac{(\boldsymbol{u_i}-\boldsymbol{\theta}_i)^2}{\boldsymbol{2\sigma_{u_i}^2}}$$

Formulation of the problem

- Suppose measurements $\boldsymbol{y}$ have a probability density $P(\boldsymbol{y}|\boldsymbol{\mu},\boldsymbol{\theta})$
  - $\boldsymbol{\mu}$ = Parameters of interest
  - $\boldsymbol{\theta}$ = Nuisance parameters

- Auxiliary Measurements $\boldsymbol{u}$ are used to provide info on nuisance parameters and are (often) assumed to be independently Gaussian distributed

- The resulting Likelihood is:

Can be a real measurement or just our best guess based on theoretical reasons

$$L(\boldsymbol{\mu},\boldsymbol{\theta}) = P(\boldsymbol{y},\boldsymbol{u}|\boldsymbol{\mu},\boldsymbol{\theta}) = P(\boldsymbol{y}|\boldsymbol{\mu},\boldsymbol{\theta})\times\prod_i \frac{1}{\sqrt{2\pi}\sigma_{u_i}}\, e^{-(u_i-\theta_i)^2/2\sigma_{u_i}^2}$$

- And the log Likelihood:

$$\log L(\boldsymbol{\mu},\boldsymbol{\theta}) = \log P(\boldsymbol{y}|\boldsymbol{\mu},\boldsymbol{\theta}) - \sum \frac{(\boldsymbol{u_i}-\boldsymbol{\theta}_i)^2}{\boldsymbol{2\sigma_{u_i}^2}}$$

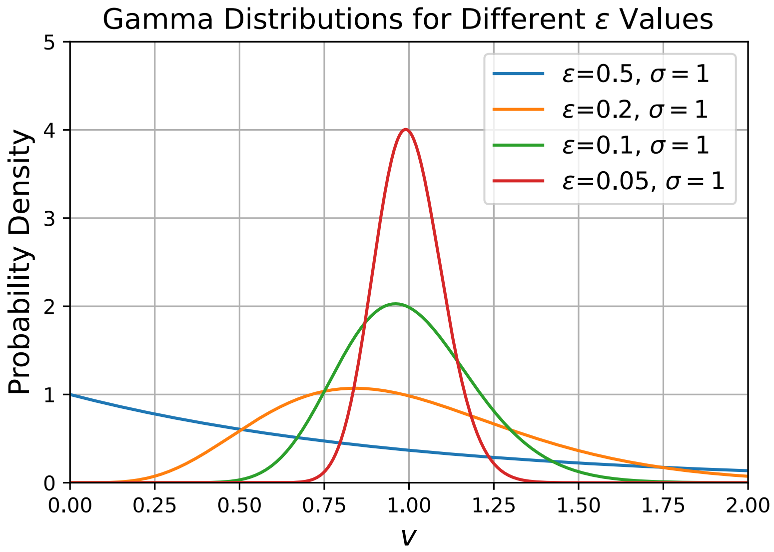
Let systematic errors be potentially uncertain!

Gamma distributions

To implement errors-on-errors suppose the systematic variances $\sigma_{u_i}^2$ are *adjustable parameters*, and their best estimates v_i are gamma distributed:

$$v \sim \frac{\beta^{\alpha}}{\Gamma(\alpha)} v^{\alpha-1} e^{-\beta v}$$

$$\alpha = \frac{1}{4\varepsilon_i^2} \qquad \beta = \frac{1}{4\varepsilon_i^2 \sigma_{u_i}^2}$$



- $\sigma_{u_i}^2$ Expectation value of v_i
- $\varepsilon_i = \frac{1}{2}\frac{\sigma_{v_i}}{\sigma_{u_i}^2} \cong \frac{\sqrt{v_i}}{\sigma_{u_i}}$: relative error on σ_{u_i}: "***Error on error***"*

*ε used to be r in previous references

Gamma Variance Model (GVM)

ROYAL
HOLLOWAY
UNIVERSITY
OF LONDON

- The likelihood is modified as follows:

$$L(\boldsymbol{\mu},\boldsymbol{\theta},\boldsymbol{\sigma}_{u_i}^2) = P(\boldsymbol{y}|\boldsymbol{\mu},\boldsymbol{\theta}) \times \prod_i \frac{1}{\sqrt{2\pi}\sigma_{u_i}} e^{-(u_i-\theta_i)^2/2\sigma_{u_i}^2} \times \frac{\boldsymbol{\beta}_i^{\boldsymbol{\alpha}_i}}{\boldsymbol{\Gamma}(\boldsymbol{\alpha}_i)} \boldsymbol{v}_i^{\boldsymbol{\alpha}_i-\mathbf{1}} \boldsymbol{e}^{-\boldsymbol{\beta}_i\boldsymbol{v}_i}$$

- One can profile over $\boldsymbol{\sigma}_{\boldsymbol{u}_i}^{\mathbf{2}}$ in closed form:

$$\log L_P(\boldsymbol{\mu},\boldsymbol{\theta}) = \log P(\boldsymbol{y}|\boldsymbol{\mu},\boldsymbol{\theta}) - \frac{1}{2}\sum_i \left(\mathbf{1} + \frac{\mathbf{1}}{\mathbf{2}\boldsymbol{\varepsilon}_i^{\mathbf{2}}}\right) \mathbf{log}\left(\mathbf{1} + \mathbf{2}\boldsymbol{\varepsilon}_i^{\mathbf{2}} \frac{(\boldsymbol{u}_i - \boldsymbol{\theta}_i)^{\mathbf{2}}}{\boldsymbol{v}_i}\right)$$

(see: G. Cowan, Eur. Phys. J. C (2019) 79:133; arXiv:1809.05778)

Gamma Variance Model (GVM)

- The original **quadratic terms** in the log likelihood replaced by a **logarithmic terms:**

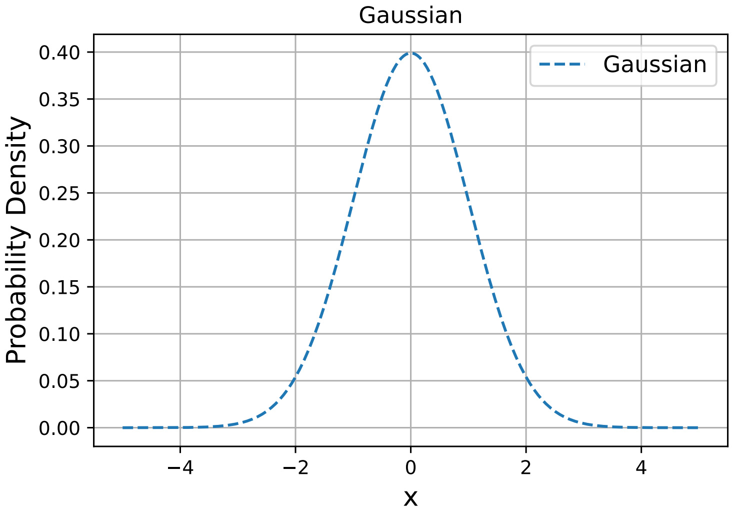
$$\sum_i \frac{(u_i - \theta_i)^2}{2\sigma_{u_i}^2} \longrightarrow \sum_i \left(1 + \frac{1}{2\varepsilon_i^2}\right) \log\left(1 + 2\varepsilon_i^2 \frac{(u_i - \theta_i)^2}{v_i}\right)$$

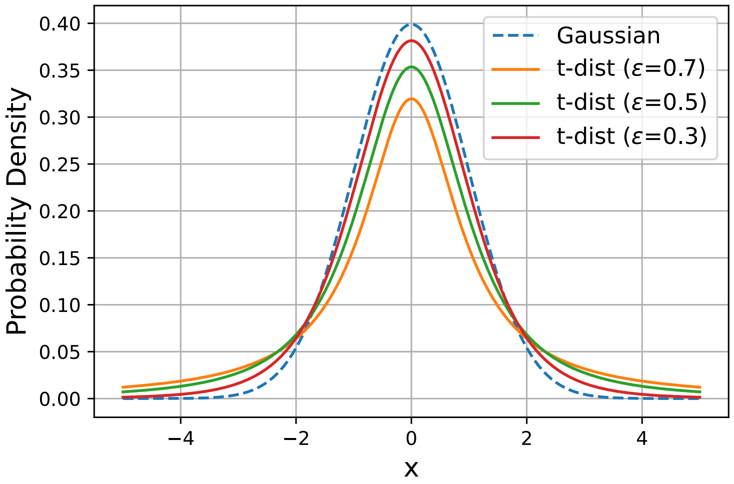
Gamma Variance Model (GVM)

- The original **quadratic terms** in the log likelihood replaced by a **logarithmic terms:**

$$\sum_i \frac{(u_i-\theta_i)^2}{2\sigma_{u_i}^2} \longrightarrow \sum_i \left(1+\frac{1}{2\varepsilon_i^2}\right)\log\left(1+2\varepsilon_i^2\frac{(u_i-\theta_i)^2}{v_i}\right)$$

- Equivalent to switch from **Gaussian constraints** to **Student's t constraints** for systematics:



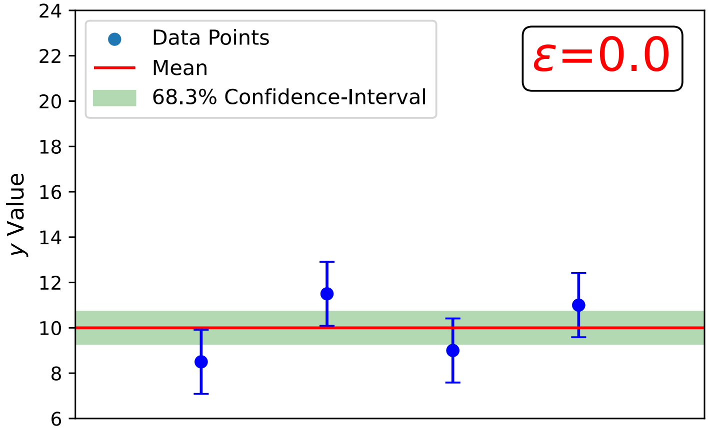


Sensitivity to outliers

- Suppose we want to average 4 measurements all with **statistical** and **syst errors** equal to **1**. Also assume they all have equal ***error on error*** $\boldsymbol{\varepsilon}$:

$$\log L_P(\boldsymbol{\mu}, \boldsymbol{\theta}) = -\frac{1}{2}\sum_i \frac{(y_i - \mu - \theta_i)^2}{\sigma_{y_i}^2} - \frac{1}{2}\sum_i \left(1 + \frac{1}{2\boldsymbol{\varepsilon}_i^2}\right) \log\left(1 + 2\boldsymbol{\varepsilon}_i^2 \frac{(u_i - \theta_i)^2}{v_i}\right)$$

- Suppose the measurements are internally compatible (no outliers), errors on errors have a small impact:

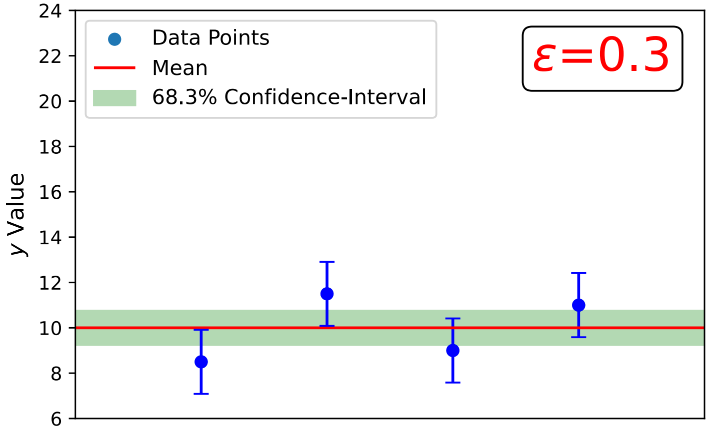


Sensitivity to outliers

- Suppose we want to average 4 measurements all with **statistical** and **syst errors** equal to **1**. Also assume they all have equal ***error on error*** $\boldsymbol{\varepsilon}$:

$$\log L_P(\boldsymbol{\mu},\boldsymbol{\theta}) = -\frac{1}{2}\sum_i \frac{(y_i-\mu-\theta_i)^2}{\sigma_{y_i}^2} - \frac{1}{2}\sum_i \left(1+\frac{1}{2\boldsymbol{\varepsilon}_i^2}\right)\log\left(1+2\boldsymbol{\varepsilon}_i^2\frac{(u_i-\theta_i)^2}{v_i}\right)$$

- Suppose the measurements are internally compatible (no outliers), errors on errors have a small impact:

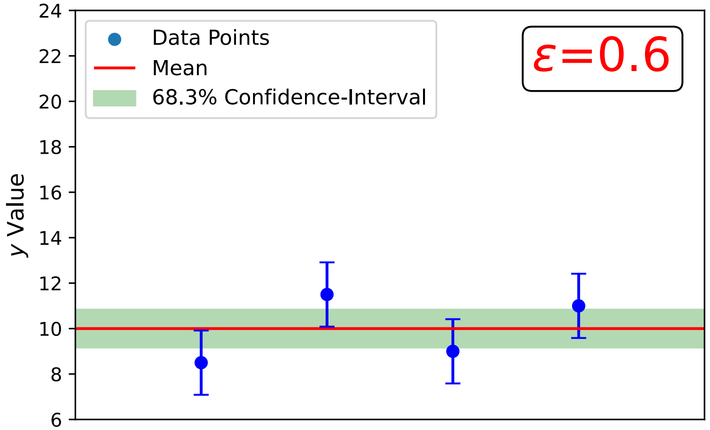


Sensitivity to outliers

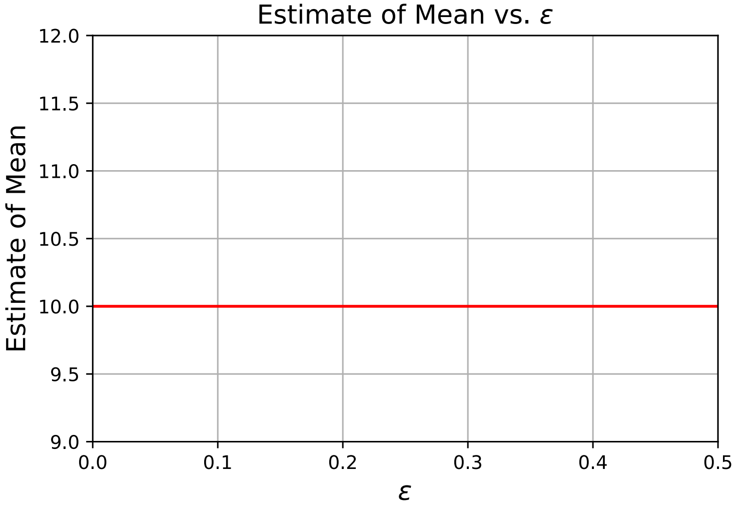
- Suppose we want to average 4 measurements all with **statistical** and **syst errors** equal to **1**. Also assume they all have equal ***error on error $\boldsymbol{\varepsilon}$***:

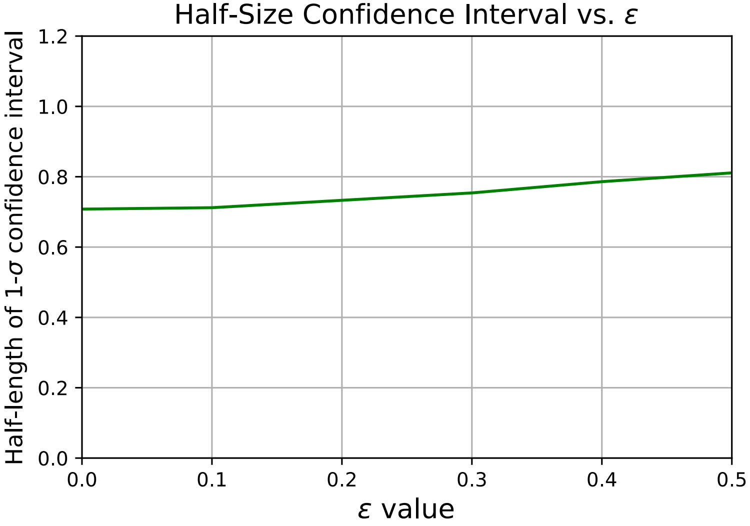
$$\log L_P(\boldsymbol{\mu}, \boldsymbol{\theta}) = -\frac{1}{2}\sum_i \frac{(y_i - \mu - \theta_i)^2}{\sigma_{y_i}^2} - \frac{1}{2}\sum_i \left(1 + \frac{1}{2\boldsymbol{\varepsilon}_i^2}\right) \log\left(1 + 2\boldsymbol{\varepsilon}_i^2 \frac{(u_i - \theta_i)^2}{v_i}\right)$$

- Suppose the measurements are internally compatible (no outliers), errors on errors have a small impact:



Sensitivity to outliers

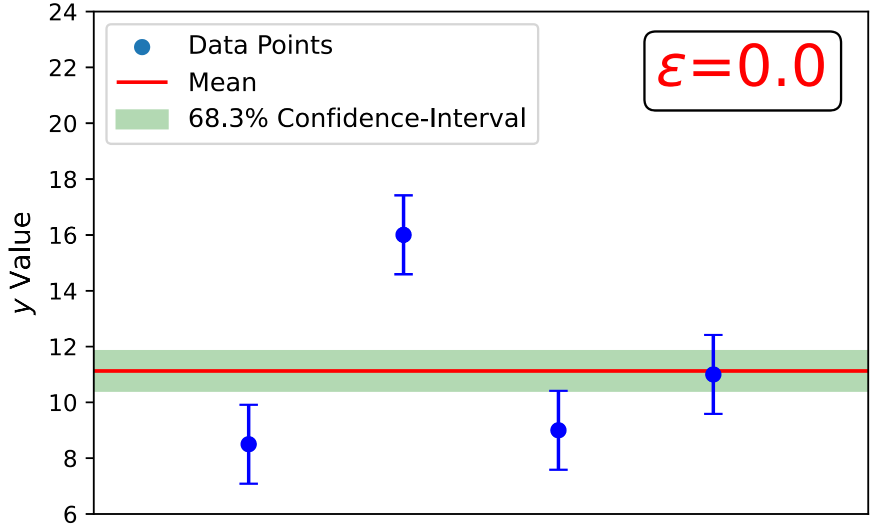

Estimate of Mean vs. ε


Half-Size Confidence Interval vs. ε

1. The estimate of the mean does not change when we increase ε
2. The size of the confidence interval for the mean only slightly increase, reflecting the extra degree of uncertainty introduced by the errors-on-errors
3. If data are internally compatible results are only slightly modified

Sensitivity to outliers

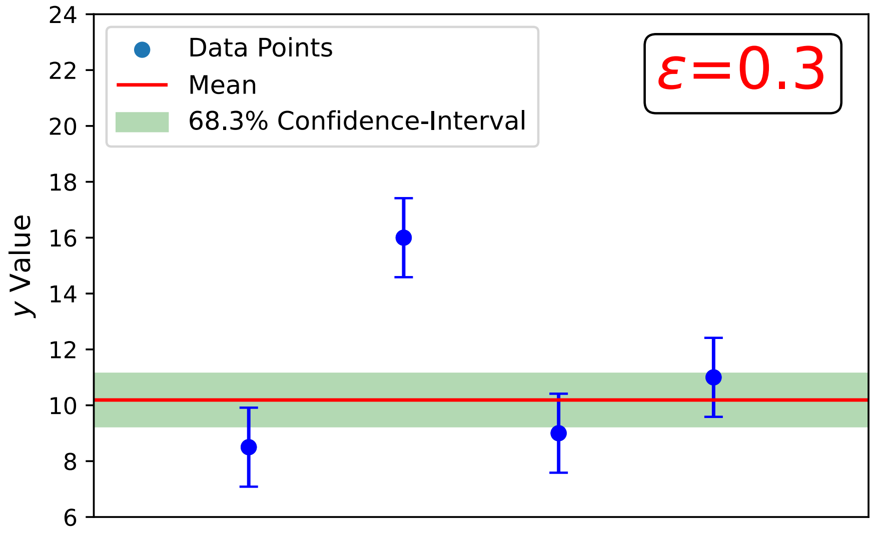
- Suppose one of the measurements is an outlier:



- If data are internally incompatible important changes can be observed

Sensitivity to outliers

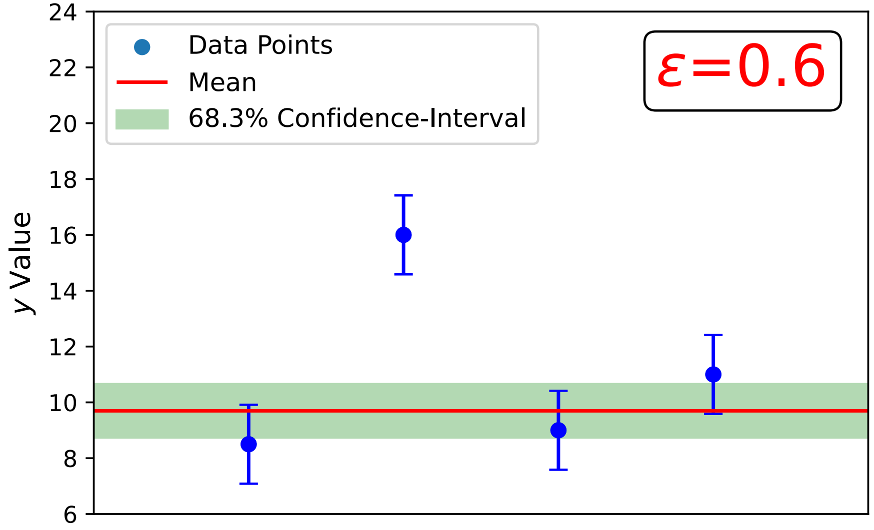
- Suppose one of the measurements is an outlier:



- If data are internally incompatible important changes can be observed

Sensitivity to outliers

- Suppose one of the measurements is an outlier:

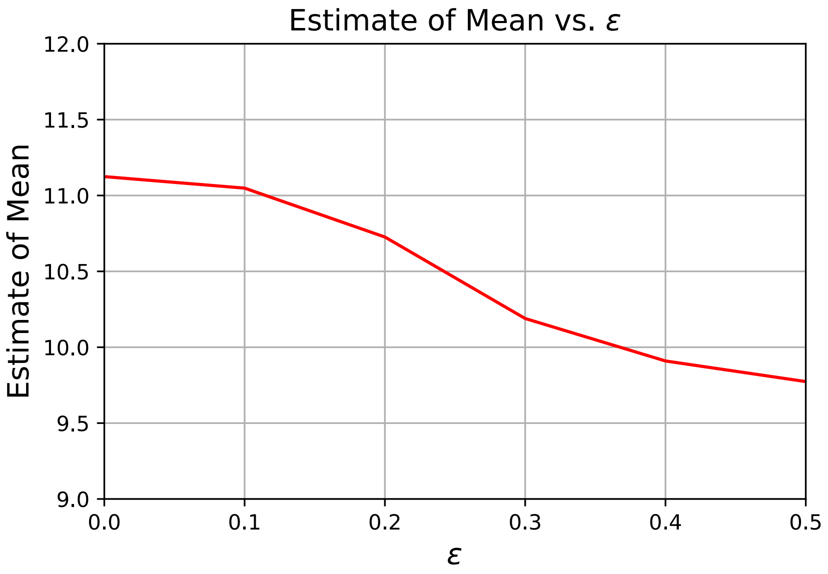


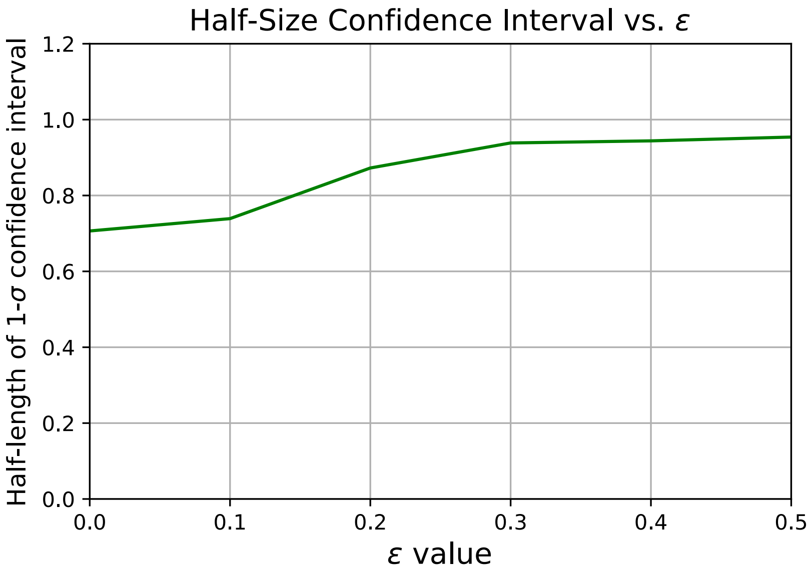
- If data are internally incompatible important changes can be observed

Sensitivity to outliers

1. With increasing ε, the estimate of mean is pulled less strongly by the outlier

2. The error bar grows more significantly: the GVM treats internal incompatibility as an additional source of uncertainty

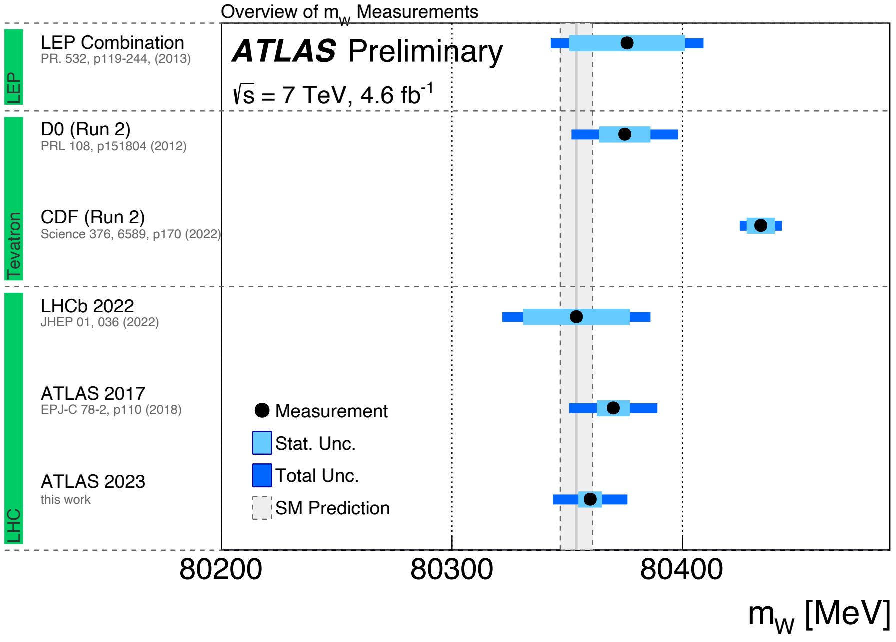

Estimate of Mean vs. ε


Half-Size Confidence Interval vs. ε

3. The model is sensitive to internal compatibility of the data

Application: W-mass

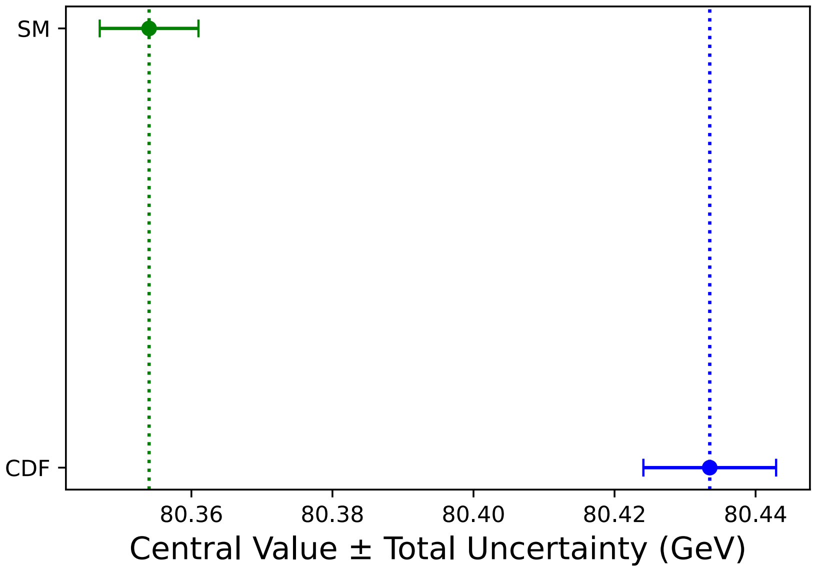
- The W-mass is one of the fundamental parameters of the Standard Model (SM)



See: ATLAS-CONF-2023-004

- The latest CDF W-mass measurement displayed a significant tension with the other measurements and the SM prediction

CDF vs SM



Potentially uncertain!

$M_{W-CDF} = 80433.5 \pm 6.4_{stat} \pm \mathbf{6.9}_{syst}$ MeV

Science 376 (2022) 170

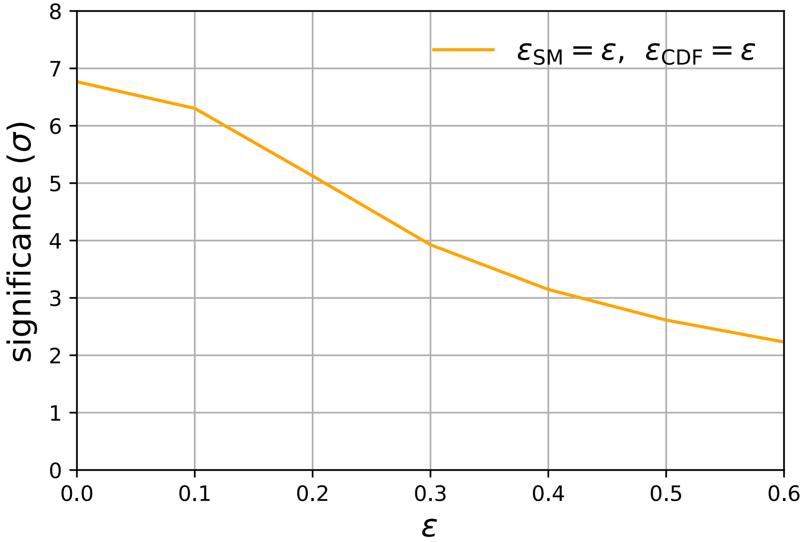
$M_{W-SM} = 80354 \pm \mathbf{7}$ MeV

PoS ICHEP2022 897, arXiv: 2211.07665

- CDF has a 7σ discrepancy with the SM
- We will associate an error-on-error to the uncertainties in ***orange.***

Significance of discrepancy

1. We assume $\boldsymbol{\varepsilon}_{\mathbf{CDF}}$ and $\boldsymbol{\varepsilon}_{\boldsymbol{SM}}$ to be equal to $\boldsymbol{\varepsilon}$ and we plot the significance as a function of $\boldsymbol{\varepsilon}$.



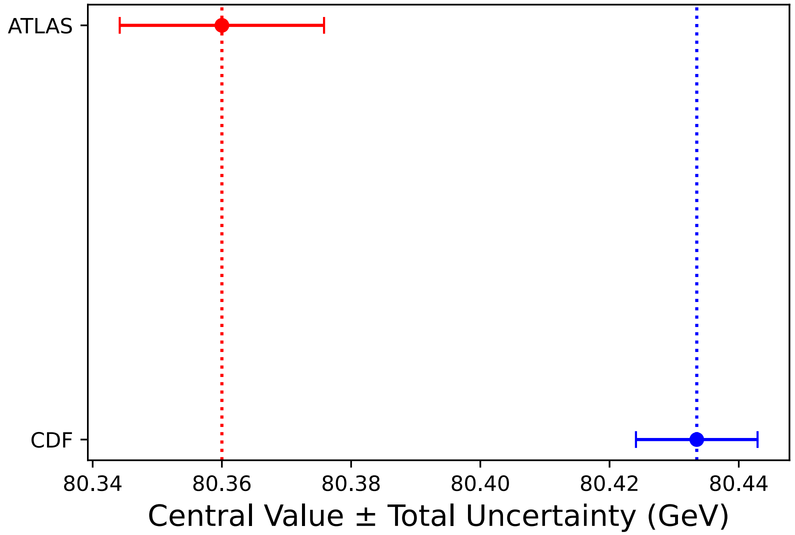
$$M_{W-CDF} = 80433.5 \ \pm 6.4_{stat} \pm \mathbf{6.9}_{syst} \text{ MeV}$$

$$M_{W-SM} = 80354 \ \pm \mathbf{7} \text{ MeV}$$

Assign ε to both

- Errors-on-errors substantially reduce the significance of the discrepancy
- Uncertainties in the assignment of systematics can account for some of the tension between the inputs

CDF vs Atlas



Potentially uncertain!

$$M_{W-CDF} = 80433.5 \pm 6.4_{stat} \pm \mathbf{6.9}_{syst} \text{ MeV}$$

Science 376 (2022) 170

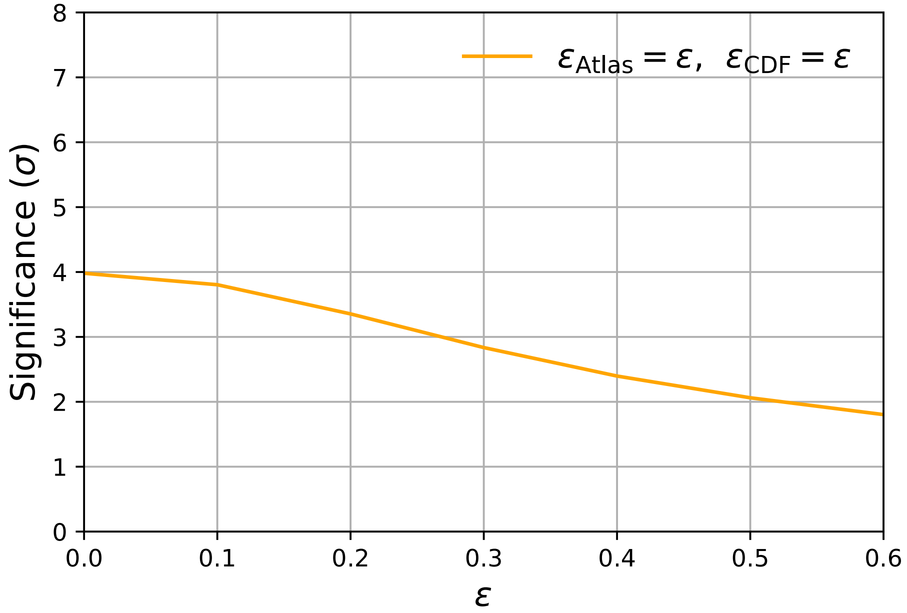
$$M_{W-Atlas} = 80360 \pm 5_{stat} \pm \mathbf{15}_{sys} \text{ MeV}$$

ATLAS-CONF-2023-004

- CDF and the latest Atlas measurements display a 4σ tension, even thought they measure the same SM parameter
- We will associate an error-on-error to the uncertainties in ***orange.***

Significance of discrepancy

1. $\boldsymbol{\varepsilon_{CDF} = \varepsilon_{Atlas} = \varepsilon}$

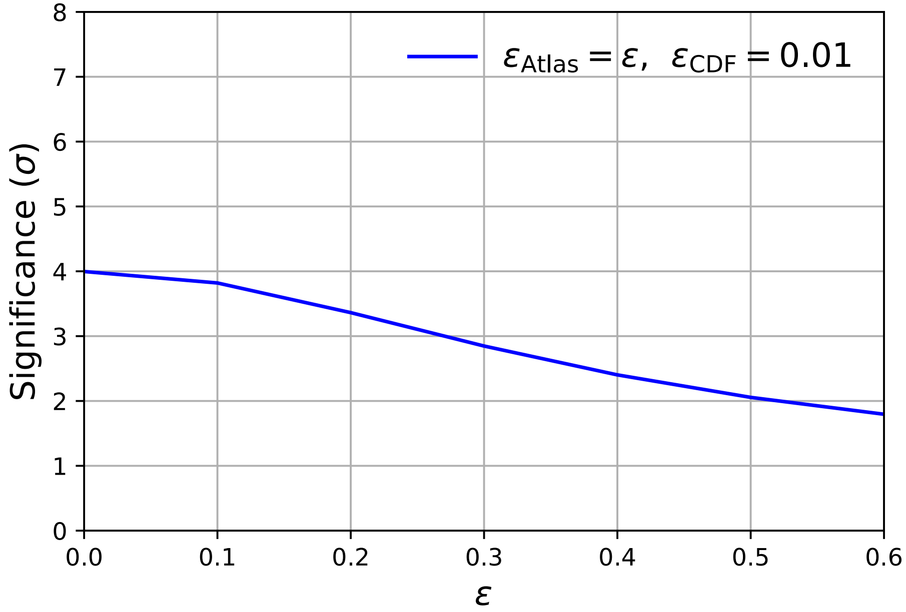


$$M_{W-CDF} = 80433.5 \pm 6.4_{stat} \pm \mathbf{6.9}_{syst}\ \text{MeV}$$

$$M_{W-Atlas} = 80360 \pm 5_{stat} \pm \mathbf{15}_{sys}\ \text{MeV}$$

Significance of discrepancy

2. $\boldsymbol{\varepsilon_{CDF}} = 0.01$ and $\boldsymbol{\varepsilon_{Atlas} = \varepsilon}$



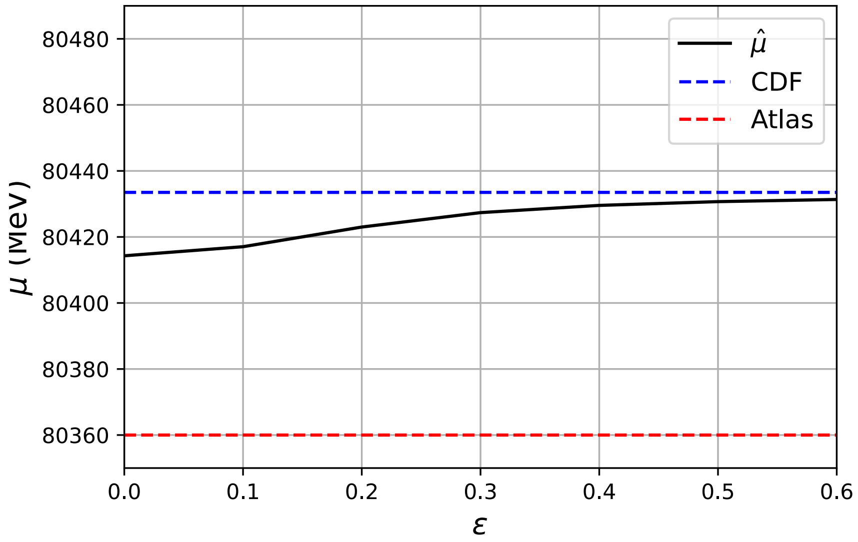
$$M_{W-CDF} = 80433.5\ \pm 6.4_{stat} \pm 6.9_{syst}\ \text{MeV}$$

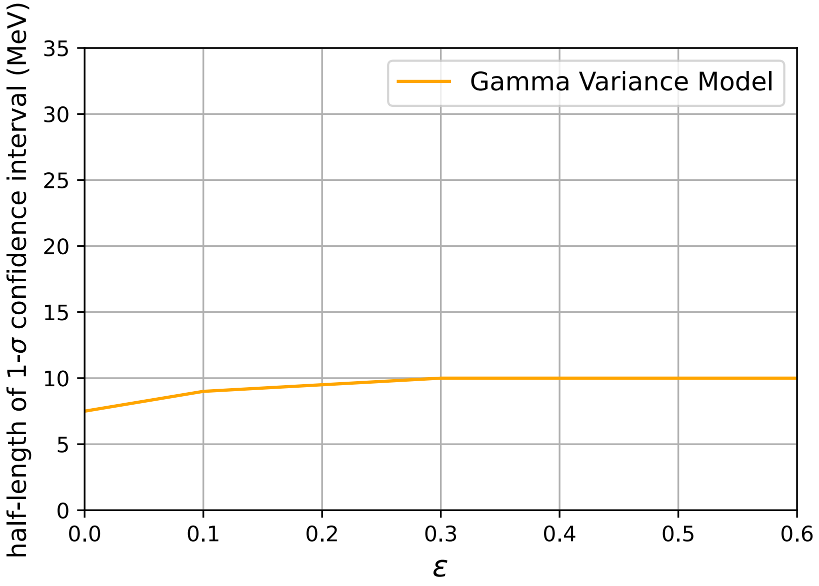
$$M_{W-Atlas} = 80360\ \pm 5_{stat} \pm \mathbf{15}_{sys}\ \text{MeV}$$

- Error on the largest uncertainty dominates
- The error on the Atlas systematic uncertainty dominates because is the one with largest systematic error

Average

1. $\varepsilon_{CDF} = \varepsilon_{Atlas} = \varepsilon$

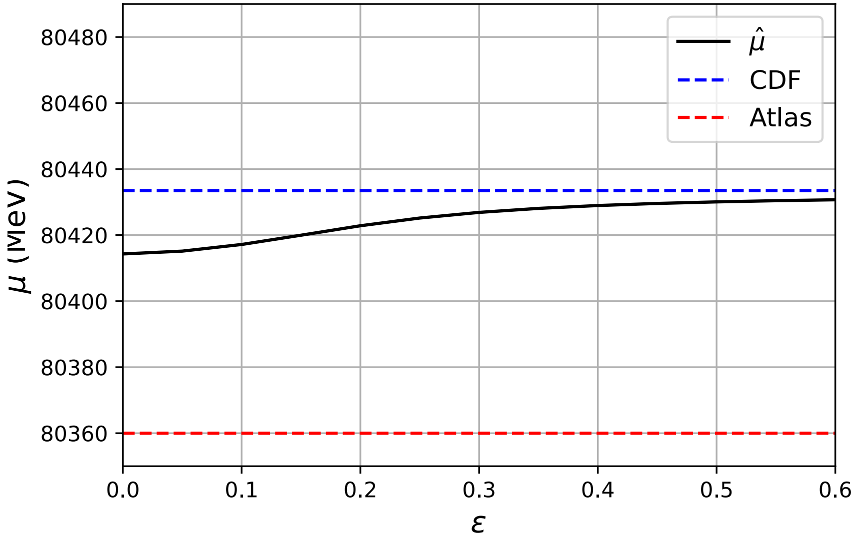


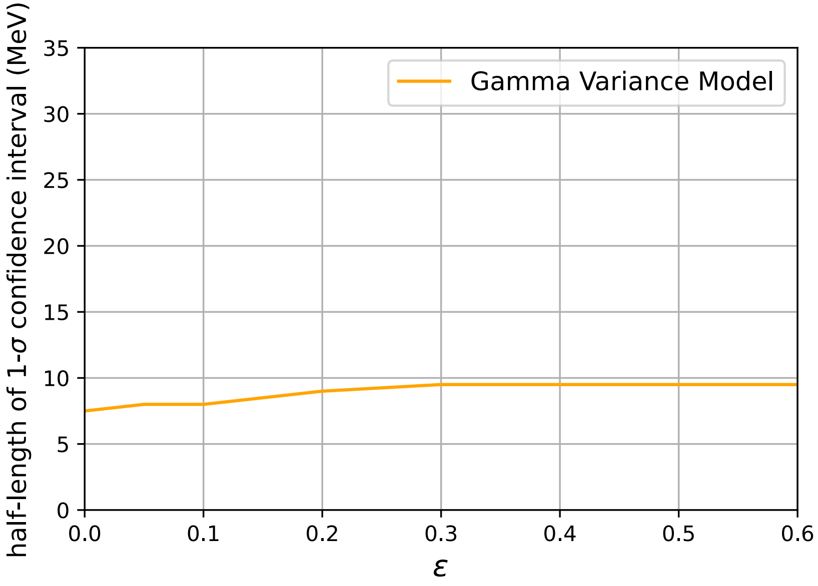


- Since the Atlas measurement has the bigger systematic error, it is treated by the GVM as the *outlier* of the dataset.

Average

2. $\varepsilon_{CDF} = 0.01$ and $\varepsilon_{Atlas} = \varepsilon$

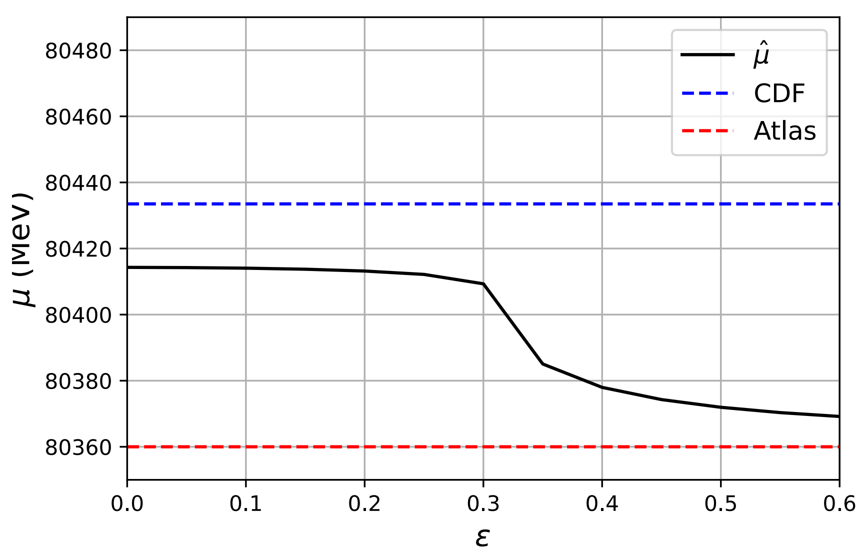


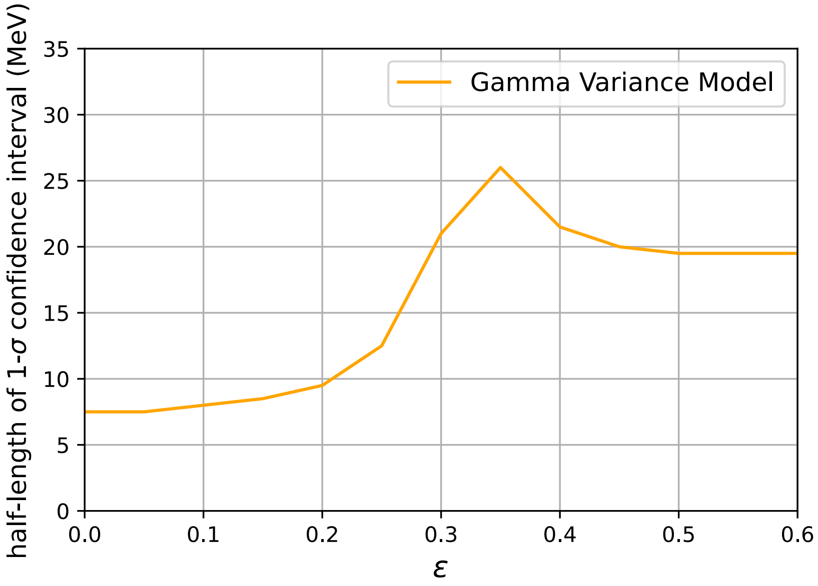


- As before the Atlas measurement is treated by the GVM as the *outlier* of the dataset.

Average

3. $\boldsymbol{\varepsilon_{Atlas}} = 0.01$ and $\boldsymbol{\varepsilon_{CDF} = \varepsilon}$





- This time, for sufficiently large error-on-error ($\boldsymbol{\varepsilon > 0.3}$) the GVM treats the CDF measurement as the *outlier* of the dataset.
- To accurately assign errors-on-errors experts knowledge is needed

Average

- If we want to reduce the significance of discrepancy from 4σ to 2σ large errors-on-errors are needed ($\varepsilon>0.5$)

- This implies a significant inflation of the confidence interval on the estimate of the mean
  - Examples 1&2: **8MeV** → **10MeV**
  - Example 3: **8MeV** → **20MeV**

- **The knowledge we claim to have on the W-mass can be significantly impacted by uncertainties on systematics.**

- How can confidence intervals be computed with the Gamma Variance Model?

- The GVM was found to deviate from the asymptotic limit when the errors-on-errors parameters ε_i are different from 0.

- Higher-order asymptotics provide an elegant solution.

Calculation of the confidence intervals

- The likelihood function can be used to construct the profile likelihood ratio test statistic:

$$w_{\boldsymbol{\mu}} = -2ln\frac{L\left(\boldsymbol{\mu},\hat{\hat{\boldsymbol{\theta}}}\right)}{L(\hat{\boldsymbol{\mu}},\ \hat{\boldsymbol{\theta}})}$$

- Use the p-value:

$$p_{\boldsymbol{\mu}} = \int_{w_{\boldsymbol{\mu},obs}}^{\infty} f\left(w_{\boldsymbol{\mu}}|\boldsymbol{\mu}\right) dw_{\boldsymbol{\mu}}$$

- Include $\boldsymbol{\mu}$ such that:

$$p_{\boldsymbol{\mu}} < \alpha$$

Confidence intervals

- The likelihood function can be used to construct the profile likelihood ratio test statistic:

$$w_{\boldsymbol{\mu}} = -2ln\frac{L\left(\boldsymbol{\mu}, \hat{\hat{\boldsymbol{\theta}}}\right)}{L\left(\hat{\boldsymbol{\mu}},\ \hat{\boldsymbol{\theta}}\right)}$$

PROBLEM: Compute $f\left(w_{\boldsymbol{\mu}}|\boldsymbol{\mu}\right)$

- Use the p-value:

$$p_{\boldsymbol{\mu}} = \int_{w_{\boldsymbol{\mu},obs}}^{\infty} f\left(w_{\boldsymbol{\mu}}|\boldsymbol{\mu}\right) dw_{\boldsymbol{\mu}}$$

- Include $\boldsymbol{\mu}$ such that:

$$p_{\boldsymbol{\mu}} < \alpha$$

Confidence intervals

- The likelihood function can be used to construct the profile likelihood ratio test statistic:

$$w_{\boldsymbol{\mu}} = -2ln\frac{L\left(\boldsymbol{\mu},\hat{\hat{\boldsymbol{\theta}}}\right)}{L\left(\hat{\boldsymbol{\mu}},\ \hat{\boldsymbol{\theta}}\right)}$$

PROBLEM: Compute $f\left(w_{\boldsymbol{\mu}}|\boldsymbol{\mu}\right)$

- Use the p-value:

$$p_{\boldsymbol{\mu}} = \int_{w_{\boldsymbol{\mu},obs}}^{\infty} f\left(w_{\boldsymbol{\mu}}|\boldsymbol{\mu}\right) dw_{\boldsymbol{\mu}}$$

SOLUTIONS:

- Use MC: Very long
- Use Asymptotic limit

- Include $\boldsymbol{\mu}$ such that:

$$p_{\boldsymbol{\mu}} < \alpha$$

Asymptotic limit

- The asymptotic limit is the limit where the MLEs are Gaussian distributed:

$$\hat{\mu} \sim \mathcal{N}(\mu, j^{-1/2}) + \mathcal{O}(\boldsymbol{n}^{-1/2})$$

$$j = -\frac{\partial^2 logL}{\partial \mu^2}$$

- $\boldsymbol{n}$ is usually, but not always, the sample size

- In the Asymptotic limit the likelihood ratio w_μ is χ^2 distributed (Wilks theorem)

$$w_\mu \sim \chi^2 + \mathcal{O}(\boldsymbol{n}^{-1})$$

- For the errors-on-errors model $\boldsymbol{n = 1 + \frac{1}{2\varepsilon^2}}$

Asymptotic limit

- The asymptotic limit is the limit where the MLEs are Gaussian distributed:

$$\hat{\mu} \sim \mathcal{N}(\mu, j^{-1/2}) + \mathcal{O}(\boldsymbol{n}^{-1/2})$$

$$j = -\frac{\partial^2 logL}{\partial\mu^2}$$

PROBLEM: What if $\boldsymbol{n}$ is small?

- $\boldsymbol{n}$ is usually, but not always, the sample size

- In the Asymptotic limit the likelihood ratio w_μ is χ^2 distributed (Wilks theorem)

$$w_\mu \sim \chi^2 + \mathcal{O}(\boldsymbol{n}^{-1})$$

- For the errors-on-errors model $\boldsymbol{n = 1 + \frac{1}{2\varepsilon^2}}$

Asymptotic limit

- The asymptotic limit is the limit where the MLEs are Gaussian distributed:

$$\hat{\mu} \sim \mathcal{N}(\mu, j^{-1/2}) + \mathcal{O}(\boldsymbol{n}^{-1/2})$$

$$j = -\frac{\partial^2 logL}{\partial \mu^2}$$

PROBLEM: What if $\boldsymbol{n}$ is small?

- $\boldsymbol{n}$ is usually, but not always, the sample size

SOLUTION:

- Use MC
- Higher order Asymptotics

- In the Asymptotic limit the likelihood ratio w_μ is χ^2 distributed (Wilks theorem)

$$w_\mu \sim \chi^2 + \mathcal{O}(\boldsymbol{n}^{-1})$$

- For the errors-on-errors model $\boldsymbol{n} = \boldsymbol{1} + \frac{\boldsymbol{1}}{\boldsymbol{2\varepsilon^2}}$

Higher order Asymptotics

Higher order-asymptotics are corrections to the likelihood ratio w_μ to make it “more” χ^2 distributed *(See: Applied Asymptotics Case Studies in Small-Sample Statistics by A. R. Brazzale, A. C. Davison and N. Reid)*

Higher order Asymptotics

Higher order-asymptotics are corrections to the likelihood ratio w_μ to make it “more” χ^2 distributed *(See: Applied Asymptotics Case Studies in Small-Sample Statistics by A. R. Brazzale, A. C. Davison and N. Reid)*

- **Bartlett Correction:** *(see: Bartlett, M. S. (1937) Proceedings of the Royal Society A, 160, 268–282)*
- $\boldsymbol{r^*}$: *(see: Barndorff-Nielsen, O. E. (1983) Biometrika, 70, 343–365)*

Higher order Asymptotics

Higher order-asymptotics are corrections to the likelihood ratio w_μ to make it “more” χ^2 distributed *(See: Applied Asymptotics Case Studies in Small-Sample Statistics by A. R. Brazzale, A. C. Davison and N. Reid)*

- **Bartlett Correction:** *(see: Bartlett, M. S. (1937) Proceedings of the Royal Society A, 160, 268–282)*
- $\boldsymbol{r^*}$: *(see: Barndorff-Nielsen, O. E. (1983) Biometrika, 70, 343–365)*

We will apply these corrections to the errors-on-errors model. But they can be a very useful tool in any analysis where the asymptotic distributions are a poor approximation, e.g., small data sample.

Bartlett Correction

- Modify the likelihood ratio w directly so that its distribution is closer to the asymptotic form:

$$w_{\mu} \longrightarrow w_{\mu}^{*} = w_{\mu}\frac{M}{E[w]}$$

$$w \sim \chi_M^2 + \mathcal{O}(\boldsymbol{n}^{-1})$$

$$w^{*} \sim \chi_M^2 + \mathcal{O}(\boldsymbol{n}^{-2})$$

Bartlett Correction

- Modify the likelihood ratio w directly so that its distribution is closer to the asymptotic form:

$$w_{\mu} \xrightarrow{\hspace{3cm}} w_{\mu}^{*} = w_{\mu} \frac{\textcircled{M}}{E[w]}$$

Expectation value in the asymptotic limit (degrees of freedom of χ^2)

$$\boxed{\begin{aligned} &w \sim \chi_M^2 + \mathcal{O}(\boldsymbol{n}^{-1}) \\ &w^* \sim \chi_M^2 + \mathcal{O}(\boldsymbol{n}^{-2}) \end{aligned}}$$

Bartlett Correction

- Modify the likelihood ratio w directly so that its distribution is closer to the asymptotic form:

$$w_{\boldsymbol{\mu}} \longrightarrow w_{\boldsymbol{\mu}}^{*} = w_{\boldsymbol{\mu}} \frac{M}{E[w]}$$

Expectation value in the asymptotic limit (degrees of freedom of χ^2)

*Exact expectation value**

$$w \sim \chi^2_M + \mathcal{O}(\boldsymbol{n}^{-1})$$

$$w^{*} \sim \chi^2_M + \mathcal{O}(\boldsymbol{n}^{-2})$$

**The expectation value can be computed at order $\mathcal{O}(\boldsymbol{n}^{-1})$ using a result by Lawley (Biometrika, Vol. 43, Issue 3-4, (1956) 295-303)*

r^*

- Define the likelihood root statistic:

$$r_\mu = sign(\mu - \hat{\mu})\sqrt{w_\mu} \;\; \sim \mathcal{N}(0,1) + \mathcal{O}(\boldsymbol{n}^{-1/2})$$

r^*

- Define the likelihood root statistic:

$$r_\mu = sign(\mu - \hat{\mu})\sqrt{w_\mu} \quad \sim\mathcal{N}(0,1) + \mathcal{O}(\boldsymbol{n}^{-1/2})$$

- Correct the likelihood root:

$$r_\mu^* = r_\mu + \frac{1}{r_\mu}\log\frac{q_\mu}{r_\mu}$$

r^*

- Define the likelihood root statistic:

$$r_\mu = sign(\mu - \hat{\mu})\sqrt{w_\mu} \quad \sim \mathcal{N}(0,1) + \mathcal{O}(\boldsymbol{n}^{-1/2})$$

- Correct the likelihood root:

$$r_\mu^* = r_\mu + \frac{1}{r_\mu}\log\frac{q_\mu}{r_\mu} \quad = \frac{r_\mu - \mathrm{E}[r_\mu]}{\mathrm{V}[r_\mu]^{1/2}} + \mathcal{O}(\boldsymbol{n}^{-3/2})$$

r^*

- Define the likelihood root statistic:

$$r_\mu = sign(\mu - \hat{\mu})\sqrt{w_\mu} \quad \sim\mathcal{N}(0,1) + \mathcal{O}(\boldsymbol{n}^{-1/2})$$

- Correct the likelihood root:

$$r_\mu^* = r_\mu + \frac{1}{r_\mu}\log\frac{q_\mu}{r_\mu} \quad = \frac{r_\mu - \mathrm{E}[r_\mu]}{\mathrm{V}[r_\mu]^{1/2}} + \mathcal{O}(\boldsymbol{n}^{-3/2})$$

$$r_\mu^* \sim \mathcal{N}(0,1) + \mathcal{O}(\boldsymbol{n}^{-3/2})$$

r^*

- Define the likelihood root statistic:

$$r_\mu = sign(\mu - \hat{\mu})\sqrt{w_\mu} \quad \sim \mathcal{N}(0,1) + \mathcal{O}(\boldsymbol{n}^{-1/2})$$

- Correct the likelihood root:

$$r_\mu^* = r_\mu + \frac{1}{r_\mu}\log\frac{q_\mu}{r_\mu} \quad = \frac{r_\mu - \mathrm{E}[r_\mu]}{\mathrm{V}[r_\mu]^{1/2}} + \mathcal{O}(\boldsymbol{n}^{-3/2})$$

$$r_\mu^* \sim \mathcal{N}(0,1) + \mathcal{O}(\boldsymbol{n}^{-3/2})$$

- r_μ^{*2} is χ^2 distributed

Simple error-on-error Model

- Suppose a measurement y is Gaussian distributed with mean μ and variance σ^2:

$$y \sim \mathcal{N}(\mu, \boldsymbol{\sigma})$$

- Suppose σ^2 is uncertain, with a relative error parameter $\boldsymbol{\varepsilon}$. The resulting likelihood is:

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

Simple error-on-error Model

- Suppose a measurement y is Gaussian distributed with mean μ and variance σ^2:

$$y \sim \mathcal{N}(\mu, \boldsymbol{\sigma})$$

- Suppose σ^2 is uncertain, with a relative error parameter $\boldsymbol{\varepsilon}$. The resulting likelihood is:

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

- The likelihood ratio is:

$$w_\mu = \left(1 + \frac{1}{2\varepsilon^2}\right) \log\left[1 + 2\varepsilon^2 \frac{(y-\mu)^2}{v}\right] \cong \frac{\boldsymbol{(y-\mu)^2}}{\boldsymbol{v}} + \mathcal{O}(\boldsymbol{\varepsilon^2})$$

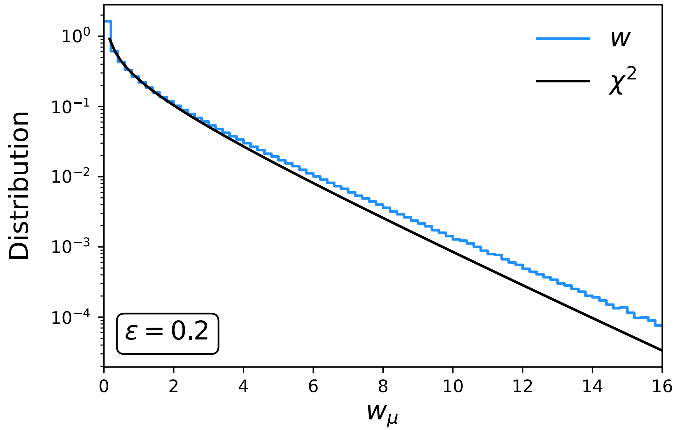
Asymptotic behaviour

- But terms like $\frac{(y-\mu)^2}{v}$ are χ^2 distributed, therefore if $\varepsilon \to 0$:

$$w_\mu \sim \chi^2 + \mathcal{O}(\varepsilon^2)$$

$$n \sim \frac{1}{\varepsilon^2}$$

- For values of $\varepsilon \neq 0$ we expect deviations of $\mathcal{O}(\varepsilon^2)$:



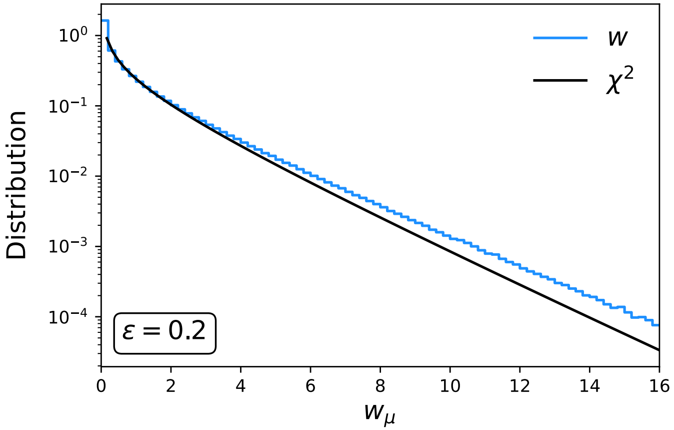
Asymptotic behaviour

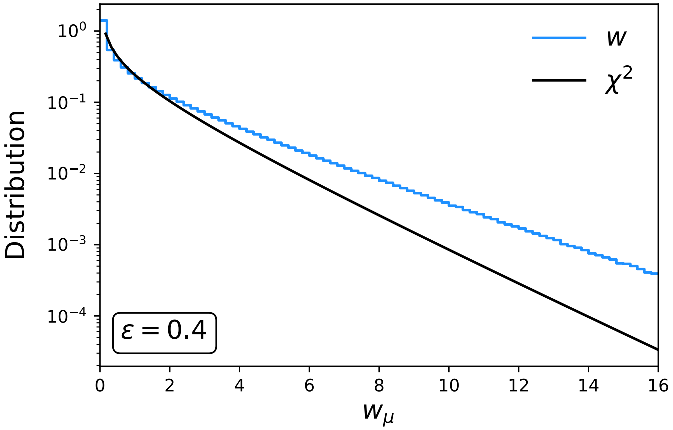
- But terms like $\frac{(y-\mu)^2}{v}$ are χ^2 distributed, therefore if $\varepsilon \to 0$:

$$w_\mu \sim \chi^2 + \mathcal{O}(\varepsilon^2)$$

$$n \sim \frac{1}{\varepsilon^2}$$

- For values of $\varepsilon \neq 0$ we expect deviations of $\mathcal{O}(\varepsilon^2)$:





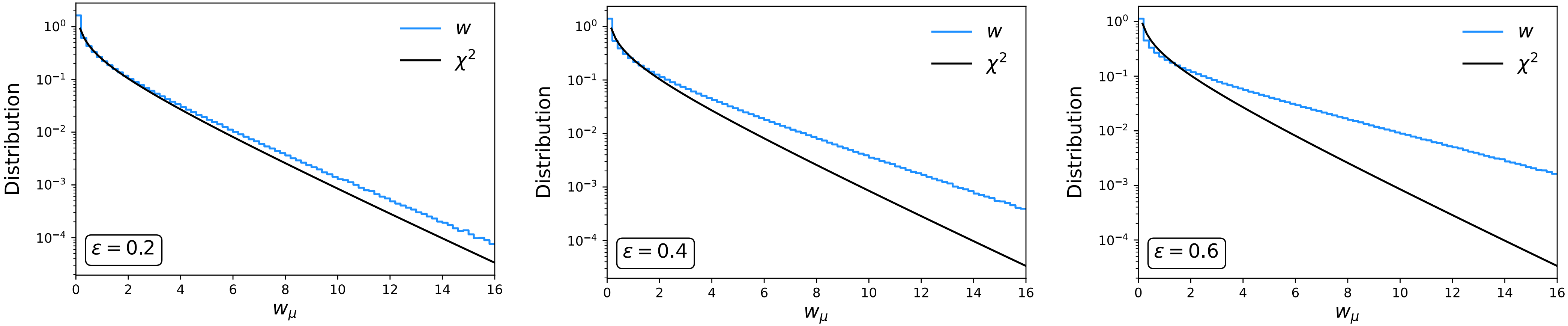
Asymptotic behaviour

- But terms like $\frac{(y-\mu)^2}{v}$ are χ^2 distributed, therefore if $\varepsilon \to 0$:

$$w_\mu \sim \chi^2 + \mathcal{O}(\varepsilon^2)$$

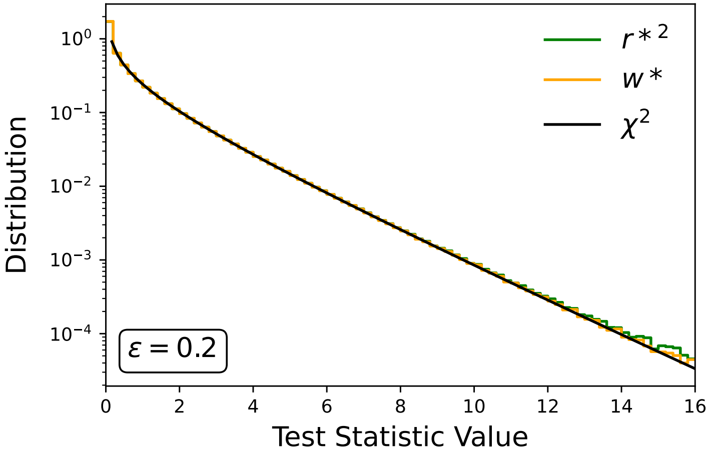
$$n \sim \frac{1}{\varepsilon^2}$$

- For values of $\varepsilon \neq 0$ we expect deviations of $\mathcal{O}(\varepsilon^2)$:



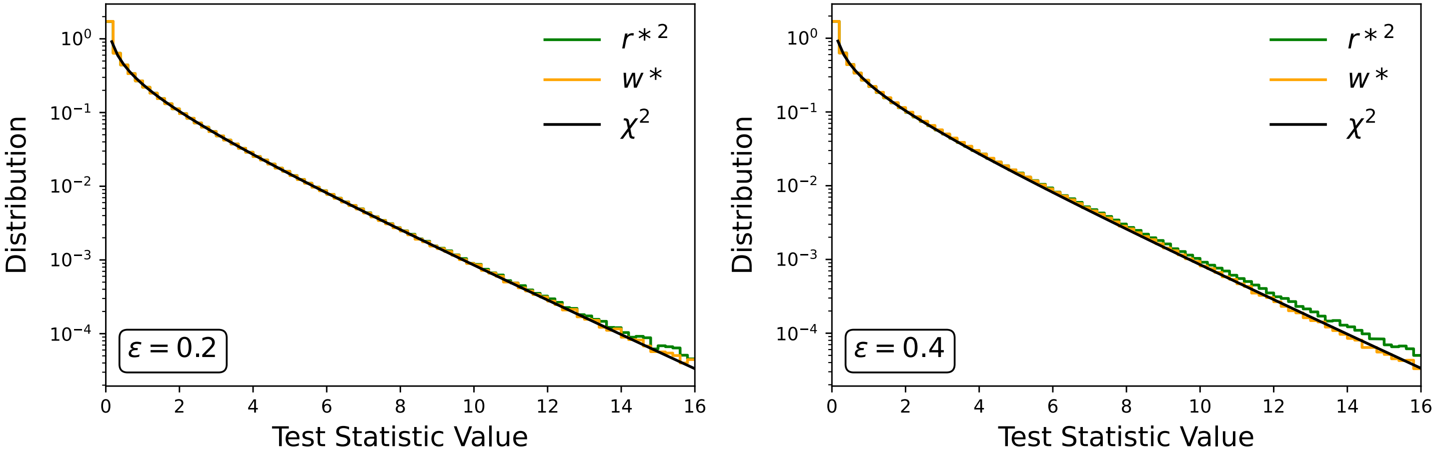
Asymptotic behaviour

- Higher order Asymptotics remarkably improve the χ^2 approximation:



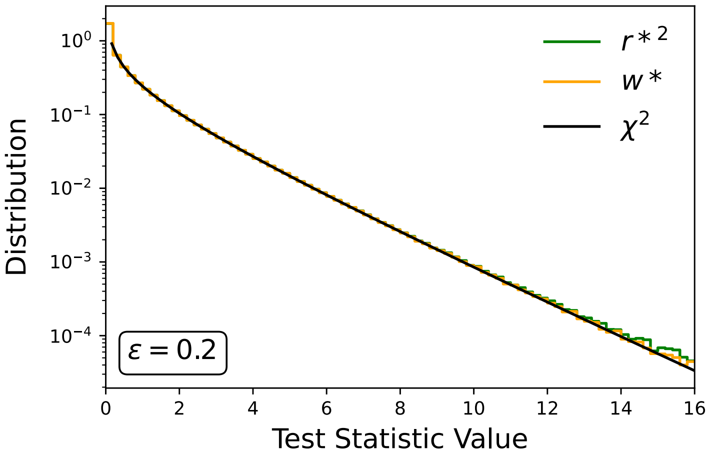
Asymptotic behaviour

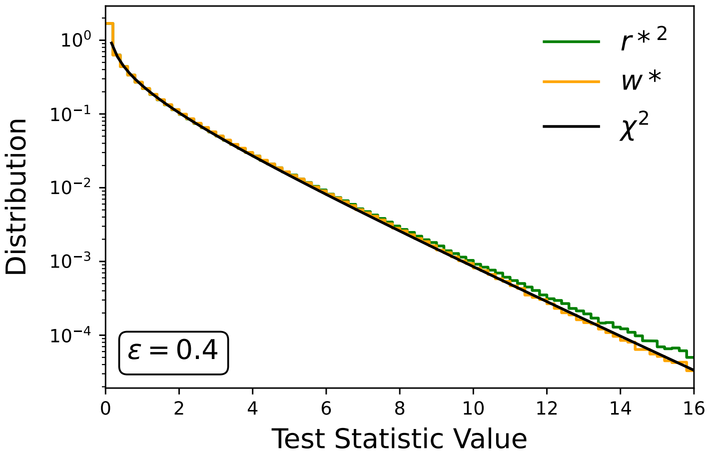
- Higher order Asymptotics remarkably improve the χ^2 approximation:

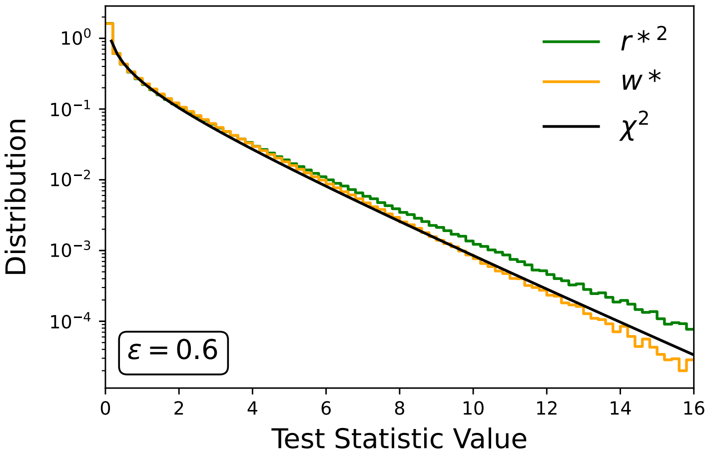


Asymptotic behaviour

- Higher order Asymptotics remarkably improve the χ^2 approximation:

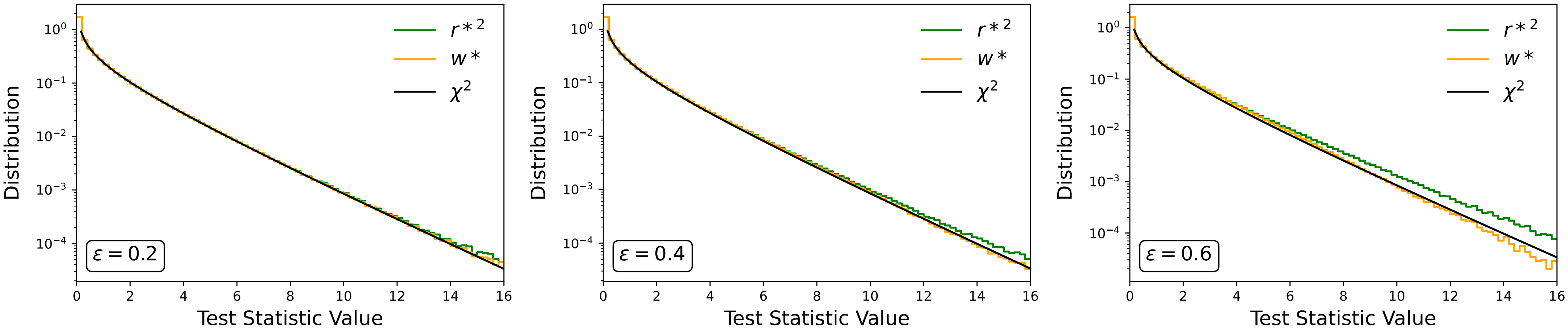






Asymptotic behaviour

- Higher order Asymptotics remarkably improve the χ^2 approximation:

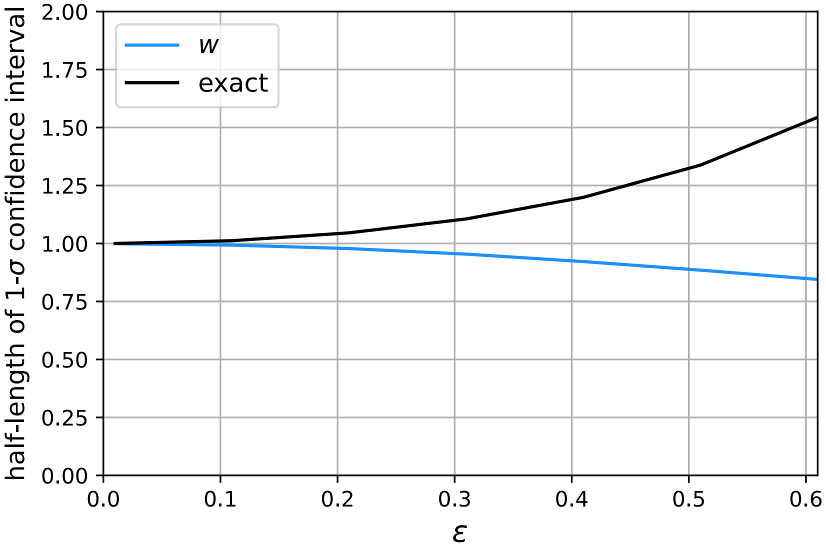


- MC not needed

Confidence Intervals

- We can use higher order asymptotics for a precise inference on the parameters of interest. For example:

$$p_\mu = \int_{w^*_{\boldsymbol{\mu},obs}}^{\infty} f\left(w^*_{\boldsymbol{\mu}}|\mu\right)\, dw^*_{\boldsymbol{\mu}}$$



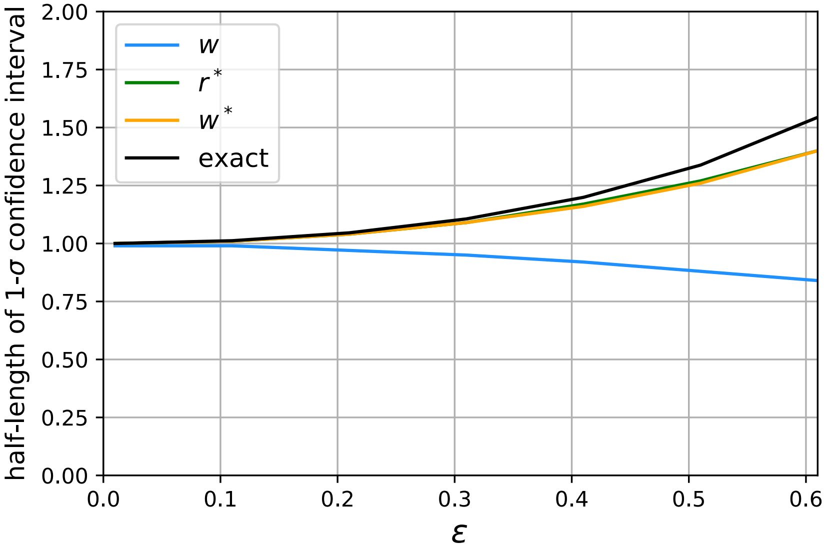
$\boldsymbol{y = 1, v = 1}$

(Half-length confidence interval as a function of ε)

Confidence Intervals

- We can use higher order asymptotics for a precise inference on the parameters of interest. For example:

$$p_\mu = \int_{w^*_{\boldsymbol{\mu},obs}}^{\infty} f\left(w^*_{\boldsymbol{\mu}}|\mu\right) dw^*_{\boldsymbol{\mu}}$$



$\boldsymbol{y = 1, v = 1}$

(Half-length confidence interval as a function of ε)

- We just published a paper on the arXiv (*arXiv:2304.10574*) where higher order asymptotics are applied to more types of errors-on-errors models
  - Simple averages without systematics.

$$L(\boldsymbol{\mu}, \boldsymbol{\sigma}^{\mathbf{2}}) = \prod_i \frac{1}{\sqrt{2\pi}\sigma_i} e^{-(y_i-\mu)^2/2\boldsymbol{\sigma}_{\boldsymbol{i}}^{\mathbf{2}}} \times \frac{\boldsymbol{\beta}_{\boldsymbol{i}}^{\boldsymbol{\alpha}_{\boldsymbol{i}}}}{\boldsymbol{\Gamma}(\boldsymbol{\alpha}_{\boldsymbol{i}})} \boldsymbol{v}_{\boldsymbol{i}}^{\boldsymbol{\alpha}_{\boldsymbol{i}}-\mathbf{1}} \boldsymbol{e}^{-\boldsymbol{\beta}_{\boldsymbol{i}}\boldsymbol{v}_{\boldsymbol{i}}}$$

  - Averages using the full GVM.

  - Correct goodness-of-fit statistics

Conclusions

- Including errors-on-errors has non-trivial consequences:
  - The model is sensitive to internal compatibility of the data
  - If data are internally compatible results are only slightly modified
  - If data are incompatible errors-on-errors modify both central values and confidence intervals in a non-linear way.

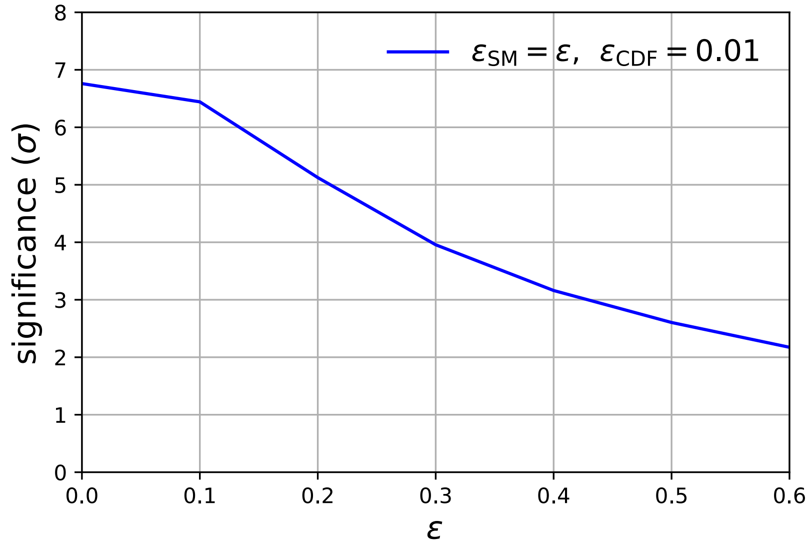
- The GVM deviates from asymptotic limit: higher-order asymptotics provide an elegant method to avoid (or reduce) the use of MC.

Thank you for your attention

Back-up slides

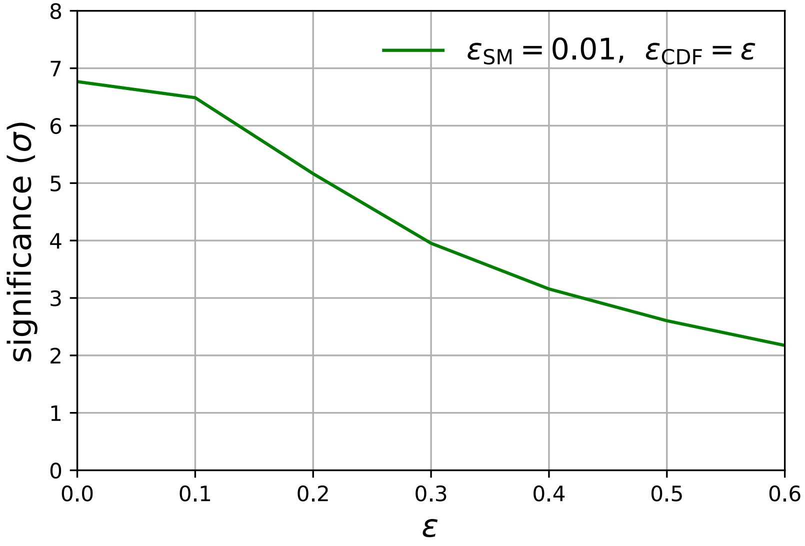
Significance of discrepancy

2. We assume $\boldsymbol{\varepsilon_{CDF}} = 0.01$ and $\boldsymbol{\varepsilon_{SM}}$ to be equal to $\boldsymbol{\varepsilon}$ and we plot the significance as a function of $\boldsymbol{\varepsilon}$.



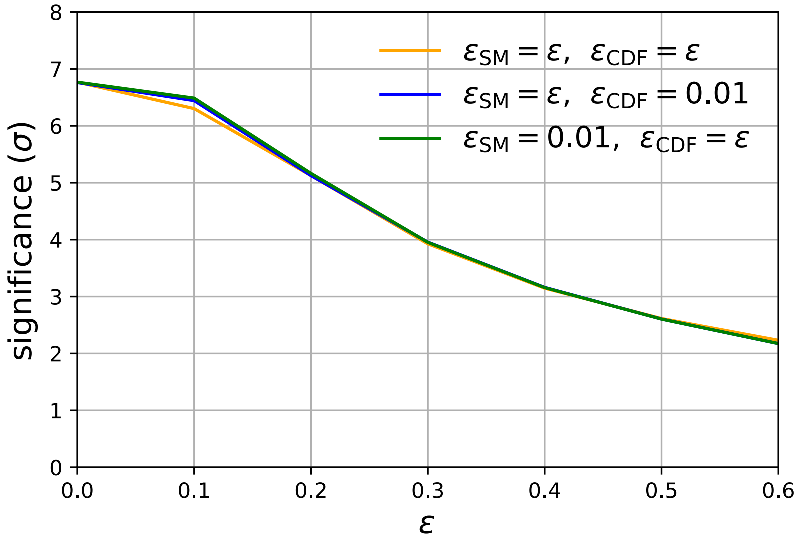
Significance of discrepancy

3. We assume $\boldsymbol{\varepsilon_{SM}} = 0.01$ and $\boldsymbol{\varepsilon_{CDF}}$ to be equal to $\boldsymbol{\varepsilon}$ and and we plot the significance as a function of $\boldsymbol{\varepsilon}$.



Significance of discrepancy

- The three examples gives very similar results:

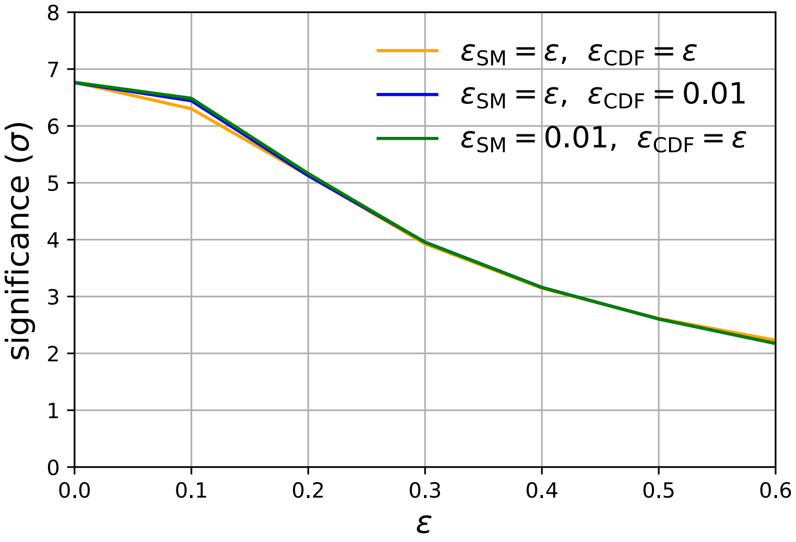


- Due to the high tension in the dataset, MLEs adjust to eliminate the quadratic term, directing tension in the dataset to the log-term with the larger uncertainty estimate v_i: *(below the goodness-of-fit statistic)*

$$q\ = \frac{(y_{CDF} - \hat{\mu} - \hat{\theta}_{CDF})^2}{\sigma^2_{y_{CDF}}} + \left(1 + \frac{1}{2\boldsymbol{\varepsilon}^2_{CDF}}\right)\log\left(1 + 2\boldsymbol{\varepsilon}^2_{CDF}\frac{\hat{\theta}_{CDF}{}^2}{v_{CDF}}\right) + \left(1 + \frac{1}{2\boldsymbol{\varepsilon}^2_{SM}}\right)\log\left(1 + 2\boldsymbol{\varepsilon}^2_{SM}\frac{(y_{SM} - \hat{\mu})^2}{v_{SM}}\right)$$

Significance of discrepancy

- The three examples gives very similar results:

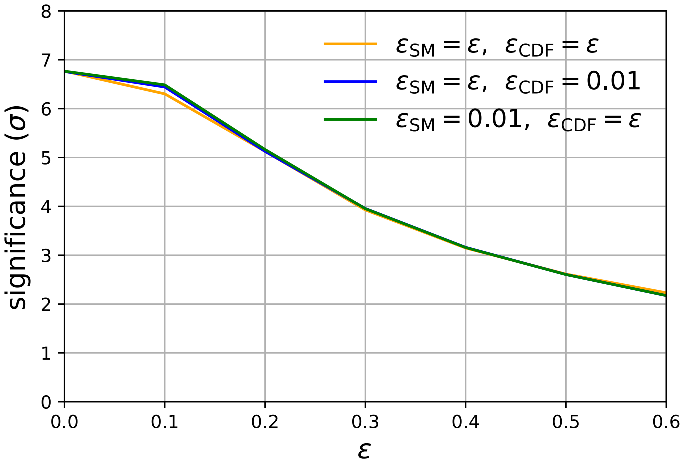


- Due to the high tension in the dataset, MLEs adjust to eliminate the quadratic term, directing tension in the dataset to the log-term with the larger uncertainty estimate v_i: *(below the goodness-of-fit statistic)*

$$q = \frac{(y_{CDF} - \hat{\mu} - \hat{\theta}_{CDF})^2}{\sigma^2_{y_{CDF}}} + \left(1 + \frac{1}{2\boldsymbol{\varepsilon}^2_{CDF}}\right)\log\left(1 + 2\boldsymbol{\varepsilon}^2_{CDF}\frac{\hat{\theta}_{CDF}{}^2}{\underset{= 6.9\ Mev}{v_{CDF}}}\right) + \left(1 + \frac{1}{2\boldsymbol{\varepsilon}^2_{SM}}\right)\log\left(1 + 2\boldsymbol{\varepsilon}^2_{SM}\frac{(y_{SM} - \hat{\mu})^2}{\underset{= 7\ Mev}{v_{SM}}}\right)$$

Significance of discrepancy

- The three examples gives very similar results.



- Due to the high tension in the dataset, the MLEs arrange themselves in order to set the quadratic term to zero and relieve the tension in the dataset in the log term with the greater estimate of the uncertainty:

1. First example: $\boldsymbol{\varepsilon_{\mathrm{CDF}} = \varepsilon}$, $\boldsymbol{\varepsilon_{\mathrm{SM}} = \varepsilon}$

$$q = \underbrace{\cancel{\frac{(y_{CDF}-\hat{\mu}-\hat{\theta}_{CDF})^2}{\sigma^2_{y_{CDF}}}}}_{=\mathbf{0}} + \left(1+\frac{1}{2\boldsymbol{\varepsilon}^2_{CDF}}\right)\log\underbrace{\cancel{\left(1+2\boldsymbol{\varepsilon}^2_{CDF}\right.}}_{=\mathbf{0}}\left.\underbrace{\frac{\hat{\theta}_{CDF}^{\;2}}{v_{CDF}}}_{=\ 6.9\ Mev}\right) + \left(1+\frac{1}{2\boldsymbol{\varepsilon}^2_{SM}}\right)\log\left(1+2\boldsymbol{\varepsilon}^2_{SM}\underbrace{\frac{(y_{SM}-\hat{\mu})^2}{v_{SM}}}_{=\ 7\ Mev}\right)$$

Significance of discrepancy

2. Second example: $\boldsymbol{\varepsilon}_{\mathbf{CDF}} = \mathbf{0.01}$, $\boldsymbol{\varepsilon}_{\mathbf{SM}} = \boldsymbol{\varepsilon}$

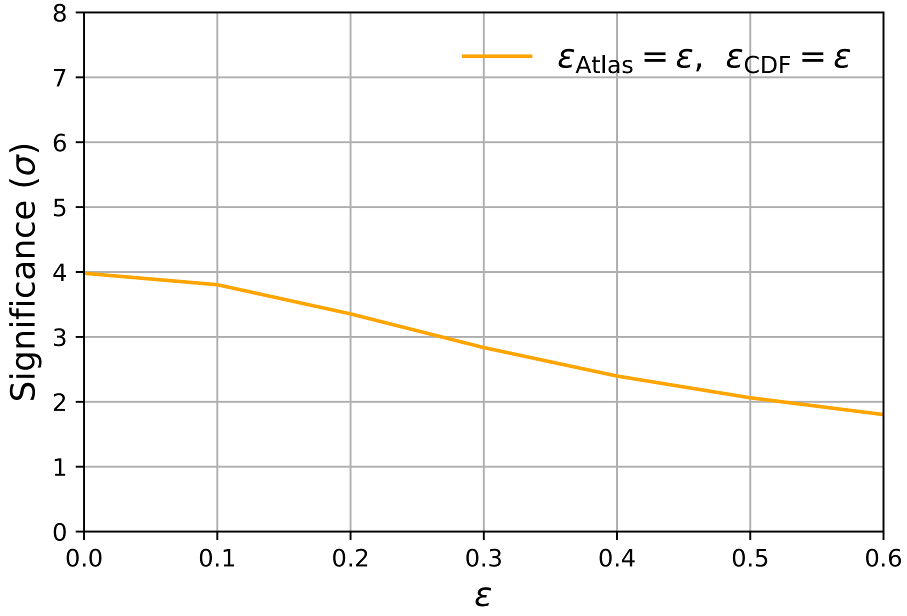
$$q = \underbrace{\cancel{\frac{(y_{CDF}-\hat{\mu}-\hat{\theta}_{CDF})^2}{\sigma^2_{y_{CDF}}}}}_{=\mathbf{0}} + \left(1+\frac{1}{2\boldsymbol{\varepsilon}^2_{CDF}}\right)\log\left(1+2\boldsymbol{\varepsilon}^2_{CDF}\frac{\hat{\theta}_{CDF}^{2}}{\underbrace{v_{CDF}}_{=6.9\ Mev}}\right) + \left(1+\frac{1}{2\boldsymbol{\varepsilon}^2_{SM}}\right)\log\left(1+\underbrace{\cancel{2\boldsymbol{\varepsilon}^2_{SM}\frac{(y_{SM}-\hat{\mu})^2}{\underbrace{v_{SM}}_{=7\ Mev}}}}_{=\mathbf{0}}\right)$$

3. Third example: $\boldsymbol{\varepsilon}_{\mathbf{CDF}} = \boldsymbol{\varepsilon}$, $\boldsymbol{\varepsilon}_{\mathbf{SM}} = \mathbf{0.01}$

$$q = \underbrace{\cancel{\frac{(y_{CDF}-\hat{\mu}-\hat{\theta}_{CDF})^2}{\sigma^2_{y_{CDF}}}}}_{=\mathbf{0}} + \left(1+\frac{1}{2\boldsymbol{\varepsilon}^2_{CDF}}\right)\log\left(1+\underbrace{\cancel{2\boldsymbol{\varepsilon}^2_{CDF}\frac{\hat{\theta}_{CDF}^{2}}{\underbrace{v_{CDF}}_{=6.9\ Mev}}}}_{=\mathbf{0}}\right) + \left(1+\frac{1}{2\boldsymbol{\varepsilon}^2_{SM}}\right)\log\left(1+2\boldsymbol{\varepsilon}^2_{SM}\frac{(y_{SM}-\hat{\mu})^2}{\underbrace{v_{SM}}_{=7\ Mev}}\right)$$

Significance of discrepancy

1. $\varepsilon_{CDF} = \varepsilon_{Atlas} = \varepsilon$

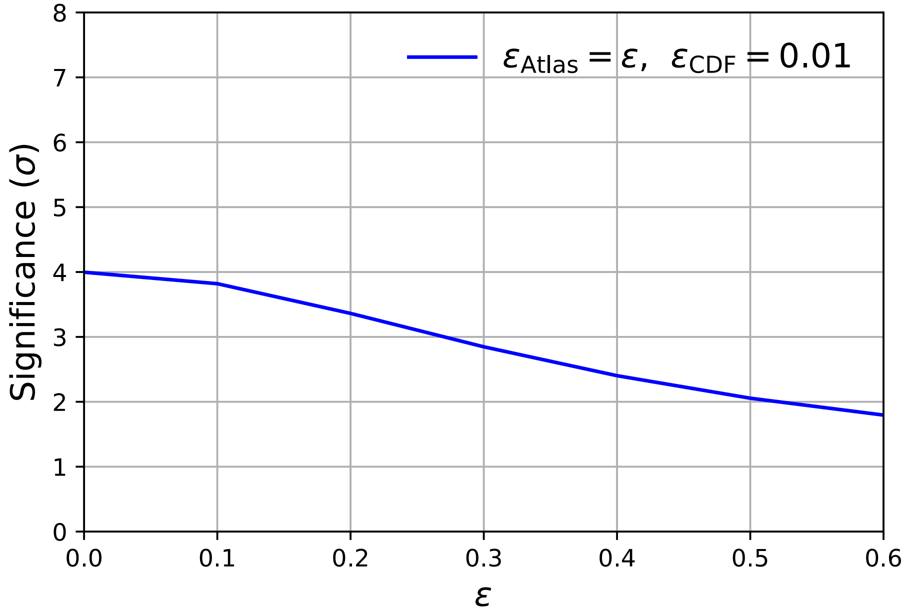


$$M_{W-CDF} = 80433.5\ \pm 6.4_{stat} \pm \mathbf{6.9}_{syst}\ \text{MeV}$$

$$M_{W-Atlas} = 80360\ \pm 5_{stat} \pm \mathbf{15}_{sys}\ \text{MeV}$$

Significance of discrepancy

2. $\boldsymbol{\varepsilon_{CDF}} = 0.01$ and $\boldsymbol{\varepsilon_{Atlas} = \varepsilon}$



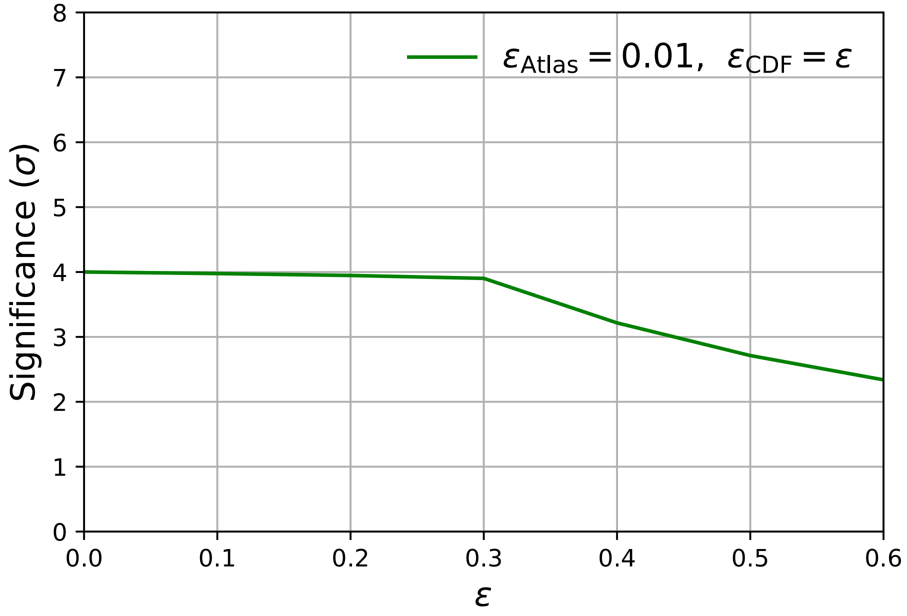
$$M_{W-CDF} = 80433.5 \ \pm 6.4_{stat} \pm 6.9_{syst} \text{ MeV}$$

$$M_{W-Atlas} = 80360 \ \pm 5_{stat} \pm \mathbf{15}_{sys} \text{ MeV}$$

- Error-on-error on the largest uncertainty dominates
- The error-on-error on the Atlas systematic uncertainty dominates because is the one with largest systematic error

Significance of discrepancy

3. $\boldsymbol{\varepsilon_{Atlas}} = 0.01$ and $\boldsymbol{\varepsilon_{CDF} = \varepsilon}$



$M_{W-CDF} = 80433.5\ \pm 6.4_{stat} \pm \mathbf{6.9}_{syst}$ MeV

$M_{W-Atlas} = 80360\ \pm 5_{stat} \pm 15_{sys}$ MeV

- If we set $\boldsymbol{\varepsilon_{Atlas}} = 0.01$ we can now see the effects of the uncertainty on CDF uncertainty

p^* Approximation

- If we integrate the p^* approximation for density function of $\hat{\mu}$:

$$f(\hat{\mu}) = p^*(\hat{\mu}) = \frac{1}{\sqrt{2\pi}}\, j\, e^{-w/2} + \mathcal{O}\left(n^{-3/2}\right)$$

- We find:

$$F(r) = \Phi(r) + (\frac{1}{r} - \frac{1}{q})\phi(r)$$

- With $q = \left(\frac{\partial logL}{\partial \hat{\mu}}(\hat{\mu}) - \frac{\partial logL}{\partial \hat{\mu}}(\mu)\right)\sqrt{j(\hat{\mu})}$

p^* Approximation

- Or equivalently we can define

$$r^* = r + \frac{1}{r} \log\left(\frac{q}{r}\right)$$

- Distributed as:

$$f(r^*) = \mathcal{N}(0,1) + \mathcal{O}(n^{-3/2})$$

- Therefore r^* is χ^2

Motivation for the GVM

- Gamma distributions allow to parametrize distributions of positive defined variables (like estimates of variances)

- Using Gamma distributions it is possible to profile in close form over σ_i^2

Motivation for the GVM

- Gamma distributions include the case where the variance is estimate from a real dataset of control measurements:

$$v_i = \frac{1}{n_i - 1} \sum \left(u_{i,j} - \overline{u_i}\right)^2$$

- $(n-1)v_i/\sigma_{u_i}^2$ follows a χ_{n-1}^2 distribution and v_{i} a Gamma distribution with:

$$\alpha_i = \frac{n_i - 1}{2}$$

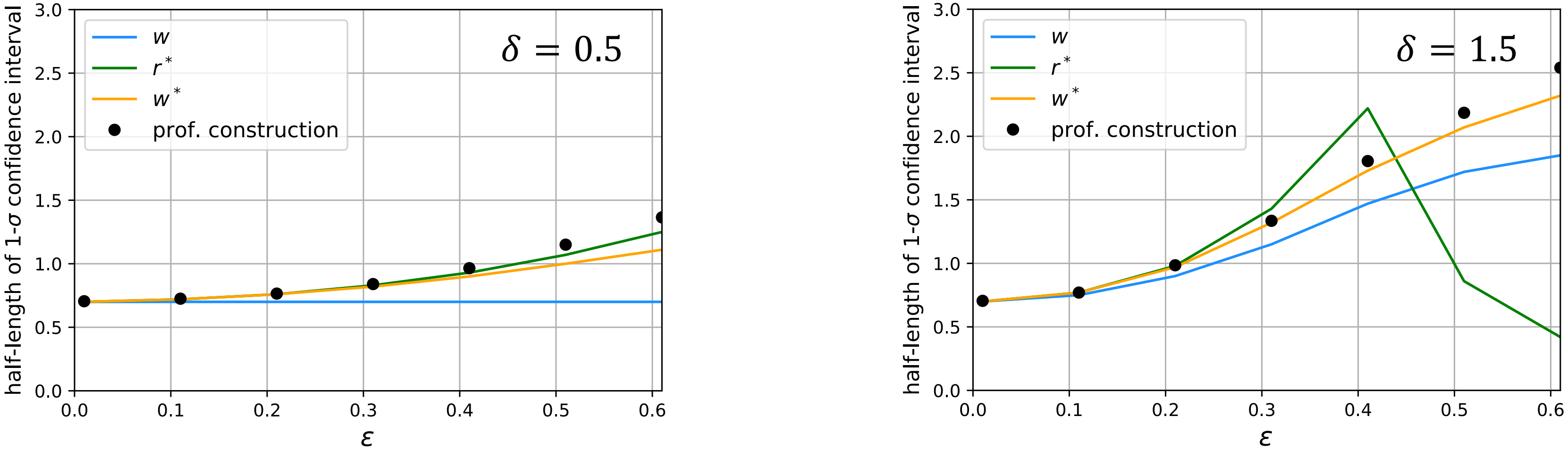
$$\beta_i = \frac{n_i - 1}{2\sigma_{u_i}^2}$$

- Therefore $n_i = 1 + 1/2r_i^2$

Simple-average-model

- $L(\boldsymbol{\mu}, \boldsymbol{\sigma^2}) = \prod_i \frac{1}{\sqrt{2\pi}\sigma_i} e^{-(y_i-\mu)^2/2\boldsymbol{\sigma_i^2}} \times \frac{\boldsymbol{\beta_i^{\alpha_i}}}{\boldsymbol{\Gamma(\alpha_i)}} \boldsymbol{v_i^{\alpha_i-1}} \boldsymbol{e^{-\beta_i v_i}}$

- $y_1 = +\delta$, $y_2 = -\delta$ and $v_1 = v_2 = 1$. Both the measurements are assumed to have $\varepsilon_1 = \varepsilon_2 = \varepsilon$



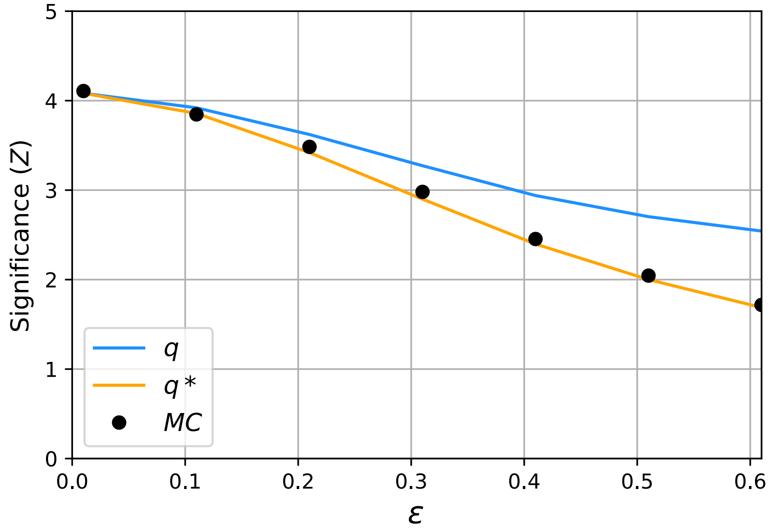
- $w^* = \frac{w}{b}$ and b can be computed in close form $b = \frac{4}{\sum\frac{1}{v_i}} \sum\frac{r_i^2}{v_i} - \frac{1}{\left(\sum\frac{1}{v_i}\right)^2} \sum\frac{r_i^2}{v_i^2}$

Simple-average-model G.O.F.

- Goodness-of-fit statistics:

$$q = \sum \left(1 + \frac{1}{2\boldsymbol{\varepsilon}_i^2}\right) \log\left(1 + 2\boldsymbol{\varepsilon}_i^2 \frac{(y_i - \hat{\mu})^2}{v_i}\right)$$

- Can be used to compute the significance of the p-value of q: $Z = \Phi^{-1}(1-p)$



(Significance computed for the same measurements of the slide before)

Simple-average-model G.O.F.

- q^* is the Bartlett-corrected version of q:

$$q^* = q * \frac{n-1}{E[q]}$$

- $E[q]$ was computed analytically at order ε^2 using the Lawley formula
- The Lawley formula was used defining q as the likelihood ratio of the saturated model

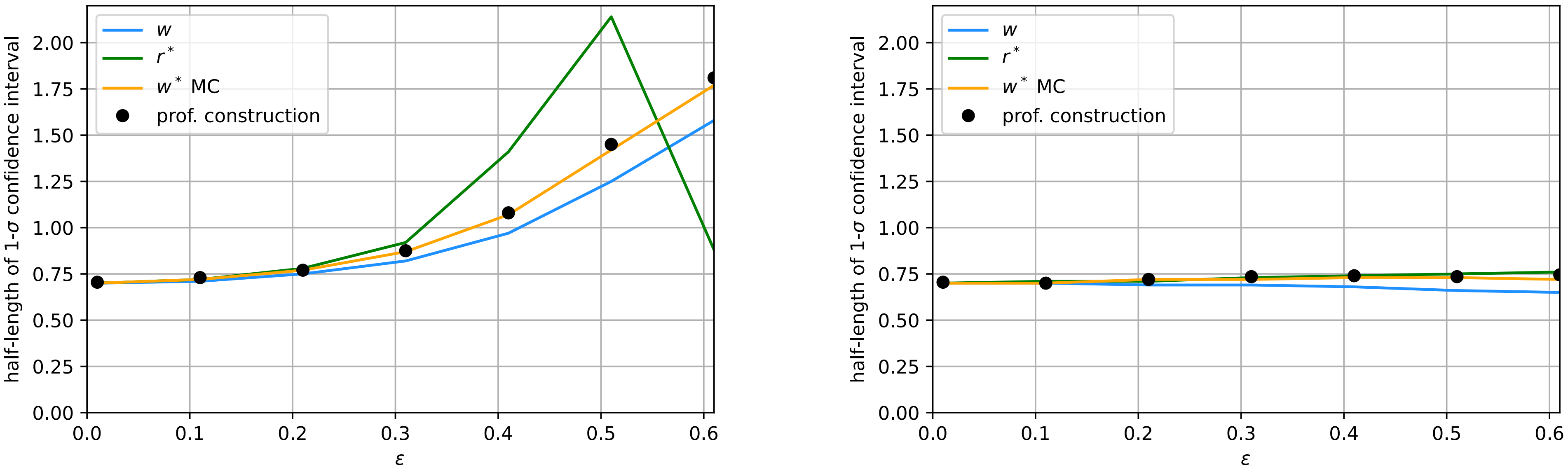
Combinations with the full GVM

- $L\left(\boldsymbol{\mu}, \boldsymbol{\theta}, \boldsymbol{\sigma}_{\boldsymbol{u}_i}^{\mathbf{2}}\right) = \prod_i \frac{1}{\sqrt{2\pi}\sigma_i} e^{-(y_i-\mu)^2/2\sigma_{y_i}^2} \prod_i \frac{1}{\sqrt{2\pi}\sigma_{u_i}} e^{-(u_i-\theta_i)^2/2\boldsymbol{\sigma}_{\boldsymbol{u}_i}^{\mathbf{2}}} \times \frac{\boldsymbol{\beta}_{\boldsymbol{i}}^{\boldsymbol{\alpha}_{\boldsymbol{i}}}}{\boldsymbol{\Gamma}(\boldsymbol{\alpha}_{\boldsymbol{i}})} \boldsymbol{v}_{\boldsymbol{i}}^{\boldsymbol{\alpha}_{\boldsymbol{i}}-\mathbf{1}} \boldsymbol{e}^{-\boldsymbol{\beta}_{\boldsymbol{i}}\boldsymbol{v}_{\boldsymbol{i}}}$

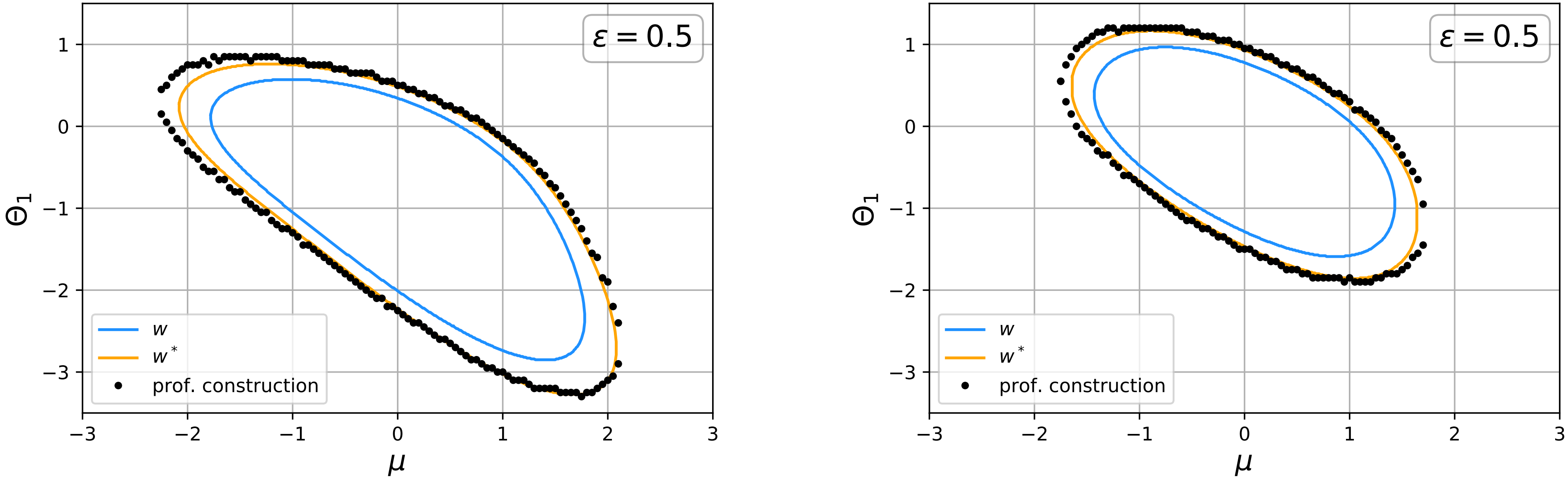
- $y_1 = +\delta, y_2 = -\delta$
- $u_1 = u_2 = 0$
- $\sigma_1^2 = \sigma_2^2 = 1/2$
- $v_1 = v_2 = 1/2$
- Both the measurements are assumed to have $\varepsilon_1 = \varepsilon_2 = \varepsilon$

Combinations with the full GVM

- Confidence interval for the parameter μ



- 2D Confidence interval for the parameter μ and θ_1

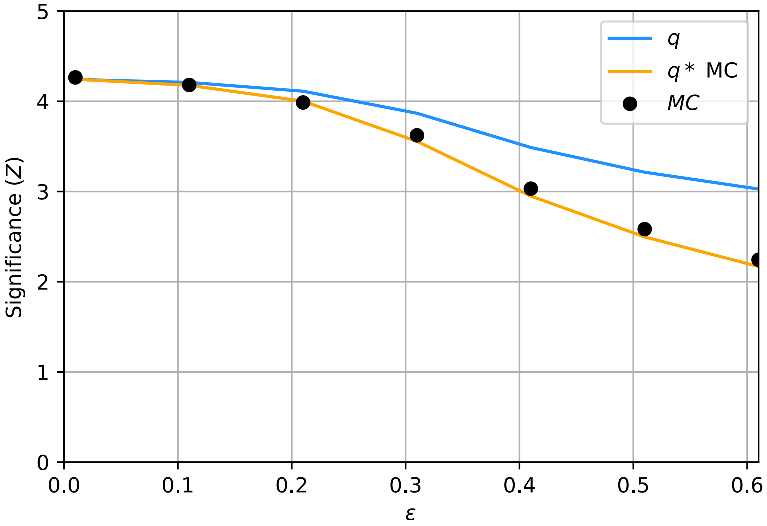


Simple-average-model G.O.F.

- Goodness-of-fit statistics:

$$q = \sum \frac{(y_i - \hat{\mu} - \widehat{\theta}_i)^2}{\sigma_{y_i}^2} + \sum \left(1 + \frac{1}{2\boldsymbol{\varepsilon}_i^2}\right) \log\left(1 + 2\boldsymbol{\varepsilon}_i^2 \frac{(u_i - \widehat{\theta}_i)^2}{v_i}\right)$$

- Can be used to compute the significance of the p-value of q: $Z = \Phi^{-1}(1 - p)$



(Significance computed for the same measurements of the slide before)