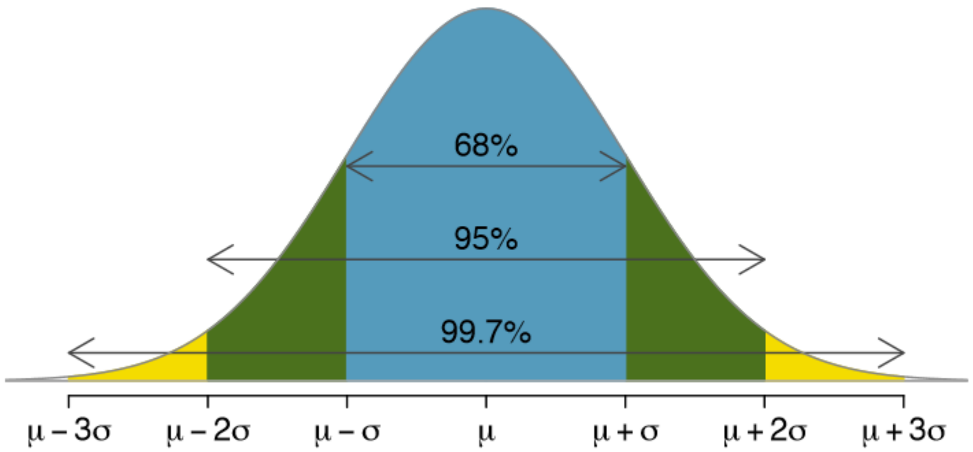
Error underestimation in high-statistics counting experiments with finite Monte Carlo samples

Cristina Alexe[1,3], Joshua Bendavid[4], Lorenzo Bianchini[2,3], Davide Bruschini[1,3]

1. Scuola Normale Superiore
2. Università di Pisa
3. INFN Sezione di Pisa
4. Massachusetts Institute of Technology



CERN School of Computing 2024

Student lightning talk - indico

DESY, Hamburg - 11th Sep 2024

Based on arXiv:2401.10542

Context

- **Task**: determine a parameter of interest (POI) μ from a binned distribution of event counts $\mathbf{y}$ given by a probability density function $\mathbf{f}$

- The model $\mathbf{f}$ depends on nuisance parameters (NPs) $\boldsymbol{\theta}$ describing systematic effects

$$-2\ln\mathcal{L}(\mu,\boldsymbol{\theta}) \approx (\mathbf{y}-\mathbf{f}(\mu,\boldsymbol{\theta}))^T\,\mathbf{V}^{-1}\,(\mathbf{y}-\mathbf{f}(\mu,\boldsymbol{\theta})) + \text{const.}$$

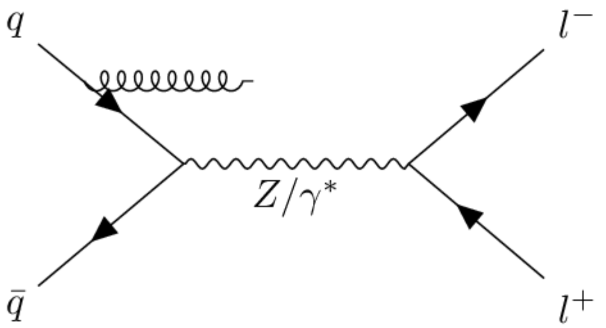
Neyman's χ2 test-statistic

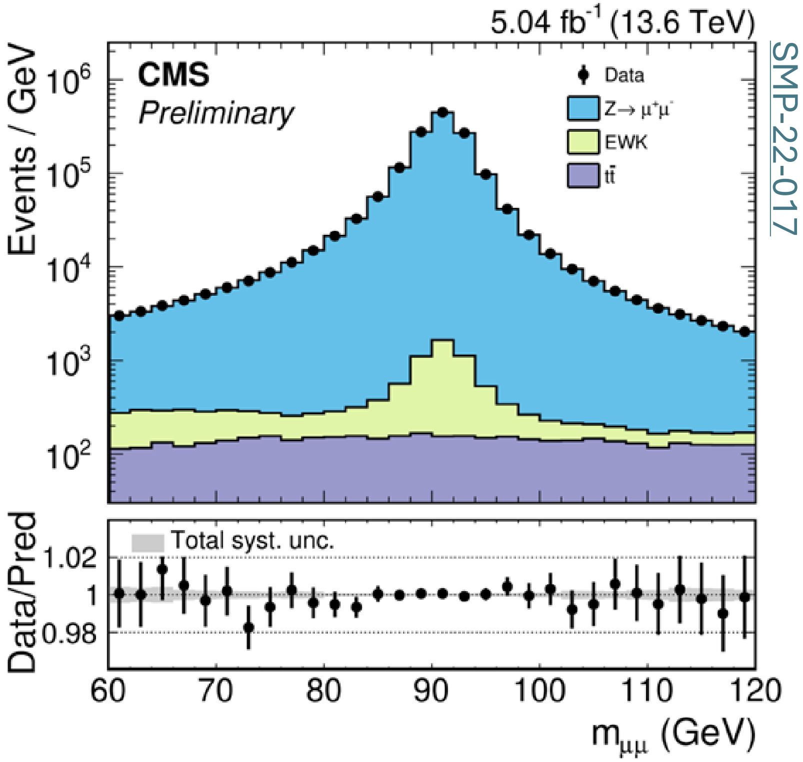
- **Example:** measurement of the cross section for a process with background

- **Figure of merit**: one-sigma confidence level interval on POI, $\hat{\sigma}$

$$\hat{\chi}^2(\hat{\mu}+\hat{\sigma}) - \hat{\chi}^2(\hat{\mu}) = 1$$

Maximum-likelihood estimators





SMP-22-017

Context

- **Task**: determine a POI μ from a binned distribution of event counts $\mathbf{y}$ given by a probability density function $\mathbf{f}$

$$-2\ln\mathcal{L}(\mu,\boldsymbol{\theta}) \approx (\mathbf{y}-\mathbf{f}(\mu,\boldsymbol{\theta}))^T\,\mathbf{V}^{-1}\,(\mathbf{y}-\mathbf{f}(\mu,\boldsymbol{\theta})) + \mathrm{const.}$$

- **Limitation:** most often, $\mathbf{f}$ is **not** a **perfectly known** function
  → instead use the prediction from a **MC simulation** with **finite statistics**
  → introduces **randomness in the extraction of μ**

theory model ⟶ theory model

- **Dealt with:** Barlow-Beeston approach
  → **introduce** as many **NPs** in the likelihood as data bins x MC processes
  → e.g. treat the **true, unknown values** of $\mathbf{f}(\mu,\boldsymbol{\theta})$ as **NPs**, constrained by a **pseudo-measurement** (the MC)

Ignoring this could lead to quoting wrong Physics results

Problem

- **Problem:** we show that the Barlow-Beeston approach **isn't sufficient**:
  → in fits with **large amounts of data** and **comparable MC statistics** & when the **model is complex**
  → leading to the underestimation of the error on the POI

- This high-statistics regime hasn't been studied much

$$\chi^2(\mu, \boldsymbol{\theta}) \approx (\mathbf{y} - \mathbf{f}(\mu, \boldsymbol{\theta}))^T \mathbf{V}^{-1} (\mathbf{y} - \mathbf{f}(\mu, \boldsymbol{\theta})) + \text{const.}$$

Suppose $\mathbf{f}(\mu, \boldsymbol{\theta}) \approx \mathbf{f}_0 + \mathbf{J}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)$

Initial value (→ $\mathbf{f}_0$) $\quad \mathbf{J} = \frac{\partial \mathbf{f}}{\partial \boldsymbol{\theta}}\Big|_{(\mu_0, \boldsymbol{\theta}_0)}$

$$\chi^2(\mu, \boldsymbol{\theta}) = \|\mathbf{b} - \mathbf{A}\mathbf{x}\|^2$$

$\mathbf{b} = \mathbf{V}^{-\frac{1}{2}}(\mathbf{y} - \mathbf{f}_0)$, $\quad \mathbf{A} = \mathbf{V}^{-\frac{1}{2}}\mathbf{J}$, $\quad \mathbf{x} = (\boldsymbol{\theta} - \boldsymbol{\theta}_0)$

Some linear algebra later…

- Write χ^2 at minimum $(\boldsymbol{\theta} - \boldsymbol{\theta}_0)$ in a simple form $\chi^2(\mu, \hat{\boldsymbol{\theta}}_\mu) = \mathbf{b}^T \mathbf{U} \mathbf{b} = \sum_{j=1}^{d} b_j^2$

Matrix **U** depends only on **A**

Components of vector **b** along eigenvectors of U

Problem

- Add statistical fluctuations to **b** and **A** $\begin{cases} \tilde{\mathbf{b}} = \mathbf{b} + \boldsymbol{\beta} \\ \tilde{\mathbf{A}} = \mathbf{A} + \boldsymbol{\alpha} \end{cases} \longrightarrow$

$$\chi^2(\mu, \boldsymbol{\theta}) = \|\mathbf{b} - \mathbf{A}\mathbf{x}\|^2$$

$$\mathbf{V}^{-\frac{1}{2}}(\mathbf{y} - \mathbf{f}_0) \quad \mathbf{V}^{-\frac{1}{2}}\mathbf{J} \quad (\boldsymbol{\theta} - \boldsymbol{\theta}_0)$$

Some more linear algebra later...

- χ2 with statistical **perturbations differs** from the **unperturbed** case by a positive offset (quadratic in μ)

$$\langle \hat{\chi}^2 \rangle = \sum_{j=1}^{d} \langle \tilde{b}_j^2 \rangle \approx \sum_{j=1}^{d} \left(b_j^2 + \boxed{\langle \nu_j \rangle^2 + 2b_j \left(\langle \nu_j \rangle + \langle \epsilon_j \rangle\right)}\right) \gtrsim \sum_{j=1}^{d} b_j^2$$

Perturbed χ2 (measured)

Positive difference (quadratic in μ)

Unperturbed χ2 (true)

→ the curvature of the perturbed χ2 is less than the unperturbed one

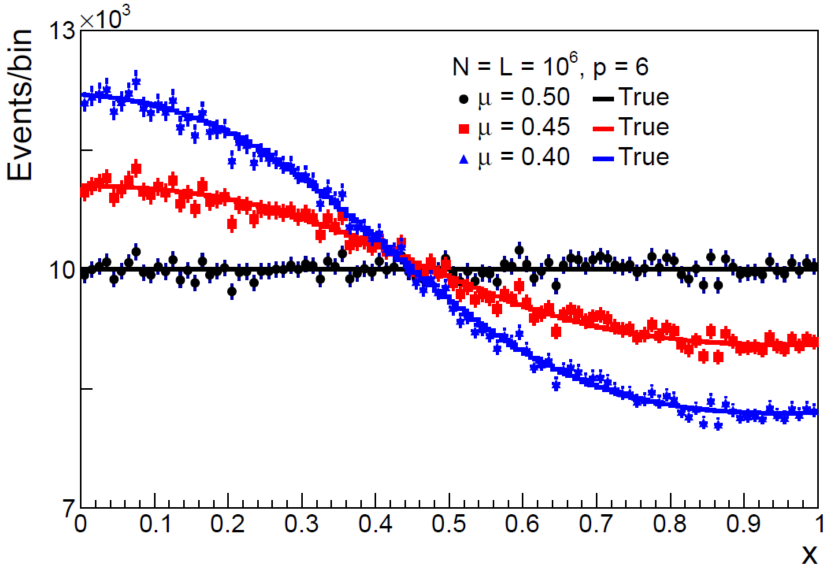
→ the **error on μ** will be **systematically underestimated**

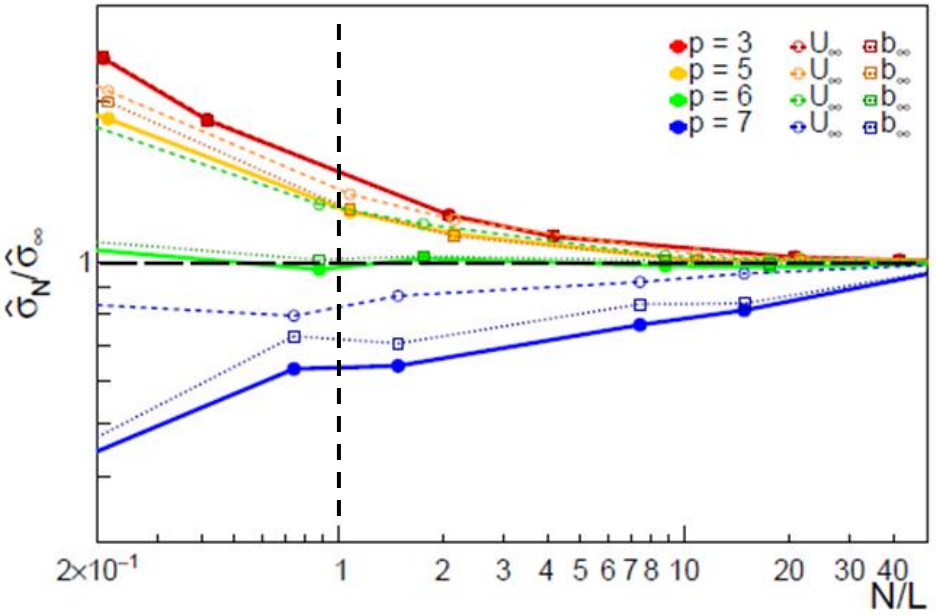
Example

- **Toy model** to compare **true** and **measured** errors
  → generate pseudo-data from model knowing the true value

$$f(x) \propto r_m(x, \boldsymbol{\theta}) \frac{z(x, \mu)}{z(x, \mu_0)}$$

Polynomial (r_m) — ~ Breit-Wigner mass distribution (z)





Solution: we can determine the scaling of $\hat{\sigma}$ as a function of MC size N at a fixed value of data luminosity L

$$\hat{\sigma}_N = \hat{\sigma}_\infty \left(1 + \frac{\delta}{N}\right)^{-\frac{1}{2}}$$

Measured error: $\hat{\sigma}_N$ — True error: $\hat{\sigma}_\infty$ — Constant: δ

2 unknowns

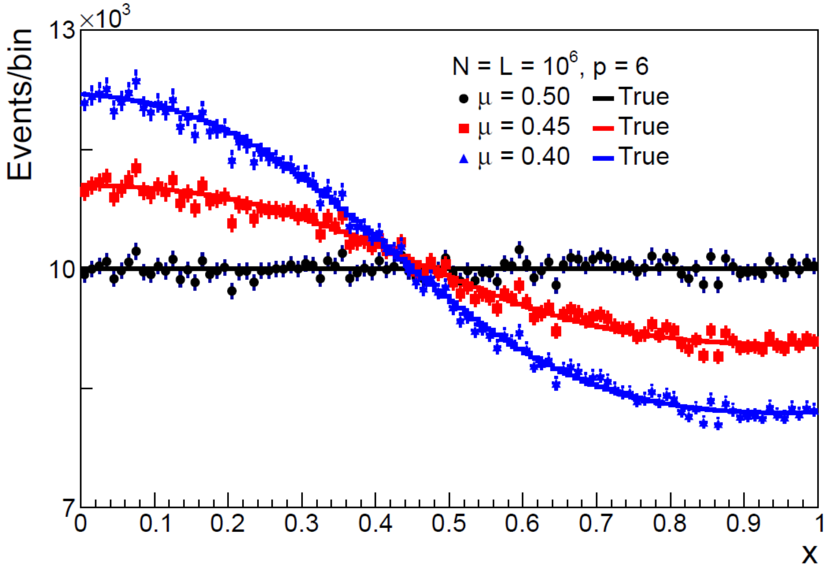
Estimate σ_N for 2 values of N and solve for $\hat{\sigma}_\infty$

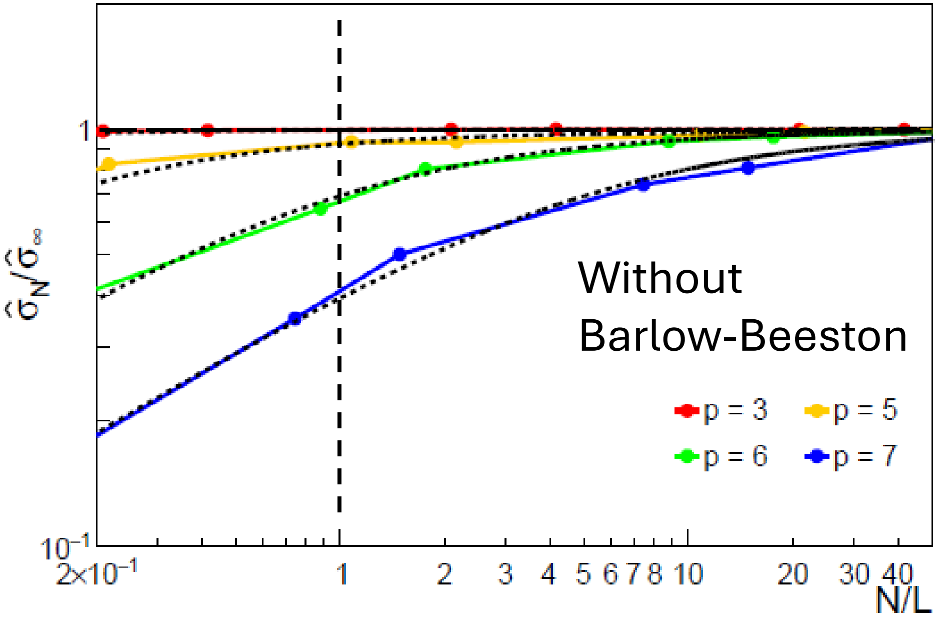
Example

- **Toy model** to compare **true** and **measured** errors
  → generate pseudo-data from model knowing the true value

$$f(x) \propto r_m(x, \boldsymbol{\theta}) \frac{z(x,\mu)}{z(x,\mu_0)}$$

Polynomial (r_m) ~ Breit-Wigner mass distribution (z)





Solution: we can determine the scaling of $\hat{\sigma}$ as a function of MC size N at a fixed value of data luminosity L

$$\hat{\sigma}_N = \hat{\sigma}_\infty \left(1 + \frac{\delta}{N}\right)^{-\frac{1}{2}}$$

Measured error ($\hat{\sigma}_N$), True error ($\hat{\sigma}_\infty$), Constant (δ)

2 unknowns

Estimate σ_N for 2 values of N and solve for $\hat{\sigma}_\infty$

Conclusions

Take care if:

- Reporting the **uncertainty on a parameter** using a profile likelihood test-statistic
- Model from **finite size MC samples**
- **Nuisance parameters** are profiled

Common task in Particle Physics

Even if:

- It's a **high-statistics** experiment
- **Size of the MC** and data sample are **comparable**
- **Barlow-Beeston** approach **is used**

You might quote an **artificially** smaller uncertainty

- Relevant for analyses with the **full data collected at the LHC** or B-factories
- **Solution:** evaluate the error underestimation for different MC samples sizes at given data luminosity

C. Alexe acknowledges financial support from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement N. 10100120)