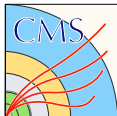


Statistical fluctuations and artificial constraints on systematic uncertainties

Andrey Popov
On behalf of CMS collaboration

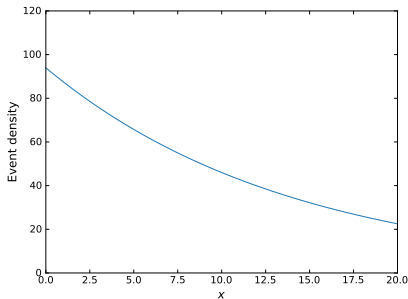
Université libre de Bruxelles

IRN Terascale
Brussels, 16–18 Oct 2019



Typical data analysis in HEP

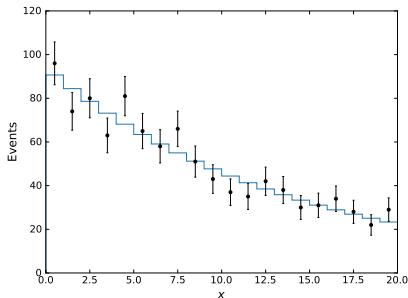
- Choose a **distribution** to study



Typical data analysis in HEP

- Choose a **distribution** to study
- Represented by a **histogram**

$$n_i \sim P(n_i; \lambda_i) = \frac{\lambda_i^{n_i}}{n_i!} e^{-\lambda_i}$$



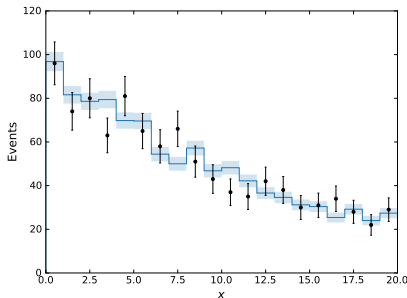
Typical data analysis in HEP

- Choose a **distribution** to study

- Represented by a **histogram**

$$n_i \sim P(n_i; \lambda_i) = \frac{\lambda_i^{n_i}}{n_i!} e^{-\lambda_i}$$

- Expected distribution ('template') constructed using **Monte-Carlo**
 - Associated per-bin uncertainties σ_i



Typical data analysis in HEP

- Choose a **distribution** to study

- Represented by a **histogram**

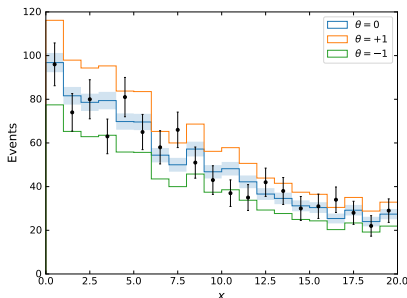
$$n_i \sim P(n_i; \lambda_i) = \frac{\lambda_i^{n_i}}{n_i!} e^{-\lambda_i}$$

- Expected distribution ('template') constructed using **Monte-Carlo**

- Associated per-bin uncertainties σ_i

- Systematic variations** given by alternative templates

- Nuisance parameters to control inter- and extrapolation from reference templates: $\lambda_i = \lambda_i(\theta; \lambda_i^0, \lambda_i^+, \lambda_i^-)$



Likelihood

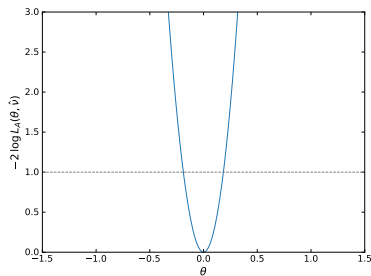
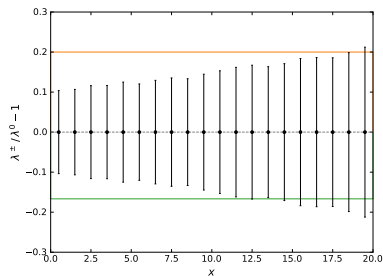
- Likelihood for toy model with one physical systematic uncertainty:

$$\log L(\theta, \boldsymbol{\nu}; \mathbf{n}) = \sum_{i=1}^m \log P(n_i; \lambda_i(\theta) + \nu_i \sigma_i) - \theta^2/2 - \boldsymbol{\nu}^2/2 + \text{const}$$

- Poissonian term: $\log P(n; \lambda) = n \log \lambda - \lambda - \text{const}$
 - Nuisances $\boldsymbol{\nu}$ control variations due to per-bin MC stat. uncertainties
- Can maximize $\log L$ with respect to $\boldsymbol{\nu}$ analytically
 - Barlow–Beeston light method
- Sensitivity is typically assessed with Asimov data set
 - Set \mathbf{n} to expectation, i.e. $n_i = \lambda_i^0$

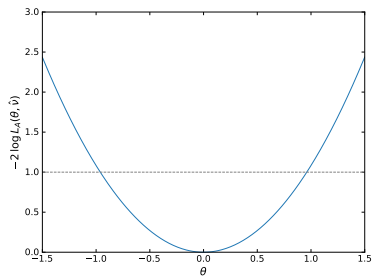
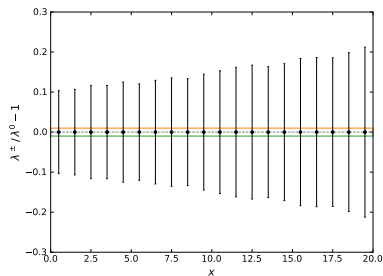
Constraints on systematic uncertainties

- Sensitivity to a systematic uncertainty is given by **profiled likelihood**
- If the variation is large compared to statistical uncertainties, it can be constrained



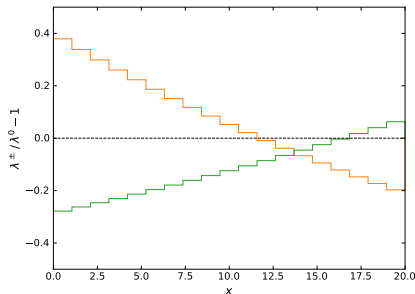
Constraints on systematic uncertainties

- Sensitivity to a systematic uncertainty is given by **profiled likelihood**
- No additional constraints if the variation is small



Fluctuations in systematic variations

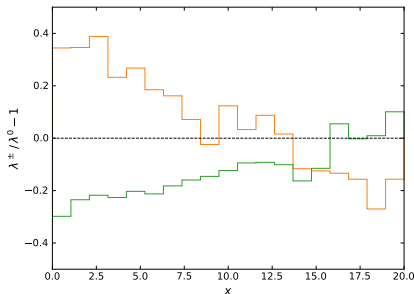
- There are different types of systematic variations:
 - Global or per-event **weights**
 - Do not change the set of MC events that enter a particular bin
 - Uncertainties in cross sections, lepton ID efficiencies, etc.



Fluctuations in systematic variations

- There are different types of systematic variations:

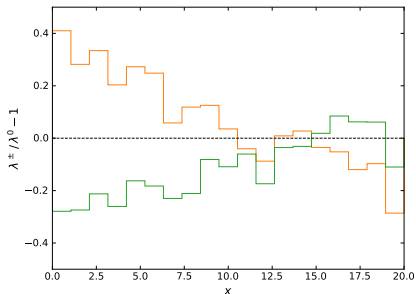
- Global or per-event **weights**
 - Do not change the set of MC events that enter a particular bin
 - Uncertainties in cross sections, lepton ID efficiencies, etc.
- **Independent variations**
 - Constructed from dedicated samples
 - Some theoretical uncertainties



Fluctuations in systematic variations

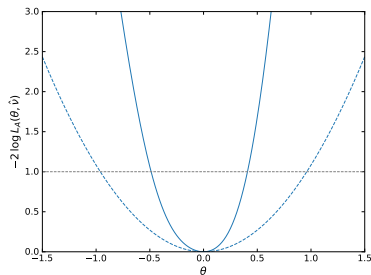
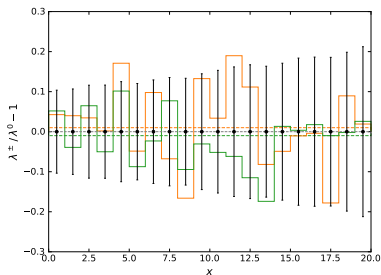
- There are different types of systematic variations:

- Global or per-event **weights**
 - Do not change the set of MC events that enter a particular bin
 - Uncertainties in cross sections, lepton ID efficiencies, etc.
- Independent variations**
 - Constructed from dedicated samples
 - Some theoretical uncertainties
- Inter-bin migrations**
 - Move events in and out of the signal region as well as between bins
 - Jet momentum calibration and like



Constraints in the presence of fluctuations

- Fluctuations in templates describing systematic uncertainties lead to tighter constraints on corresponding nuisances
 - These constraints do not represent sensitivity to underlying physical effect



No sensitivity case

- As a proxy for the constraint, can use profiled likelihood at $\theta = \pm 1$

$$\log R \equiv \max_{\nu} \log \frac{L_A(\pm 1, \nu)}{L_A^{\max}} = \sum_{i=1}^m \log \frac{P(\lambda_i^0; \lambda_i^{\pm} + \hat{\nu}_i \sigma_i)}{P(\lambda_i^0; \lambda_i^0)} - \hat{\nu}^2/2 - 1/2$$

- If $\log R < -1/2$, there is an additional constraint on θ
- Assuming parabolic dependence, $|\theta| < (-2 \log R)^{-1/2}$ at 68% CL
- With $|\Delta \lambda_i| \ll \lambda_i^0$ and setting $k = \mathcal{L}_{\text{eff}}^{\text{MC}} / \mathcal{L}^{\text{Data}}$,

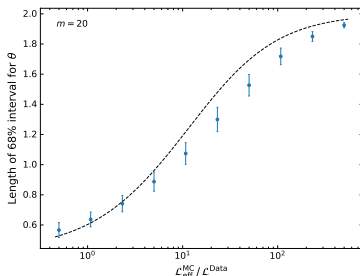
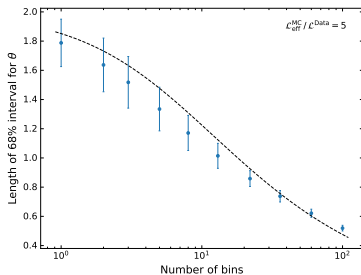
$$\log R \approx -\frac{1}{2(k+1)} \sum_{i=1}^m \frac{(\lambda_i^{\pm} - \lambda_i^0)^2}{\sigma_i^2} - 1/2$$

- If there is no real sensitivity, then $\lambda_i^{\pm} \sim \mathcal{N}(\lambda_i^0, \sigma_i^2)$, the sum follows χ_m^2 distribution, and

$$\langle \log R \rangle = -\frac{m}{2(k+1)} - 1/2$$

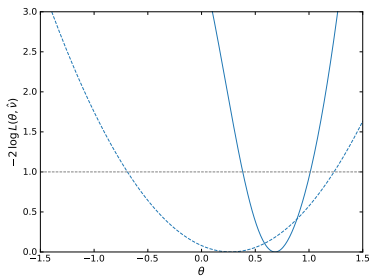
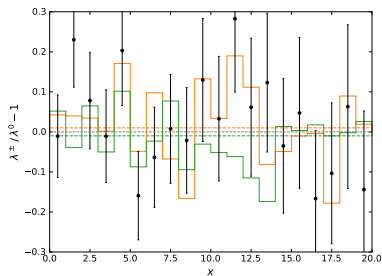
No sensitivity case

- **Numeric study** reproduces analytical results (shown with dashed lines)
 - The constraints become arbitrary tight as the number of bins grows
 - Even with $\mathcal{O}(10)$ bins, impractically large \mathcal{L}_{eff} might be needed to avoid the constraints



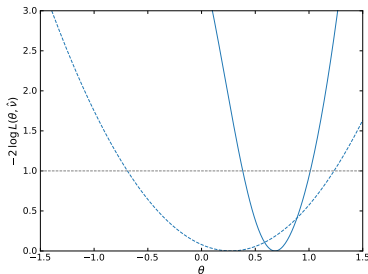
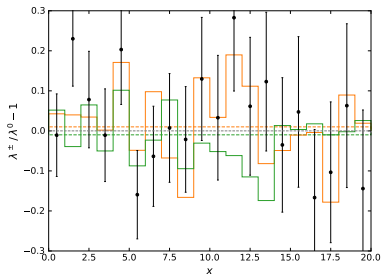
Constraints with data

- The constraints are not an artifact of Asimov data set and occur also with pseudodata:



Constraints with data

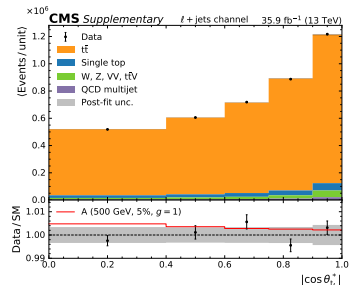
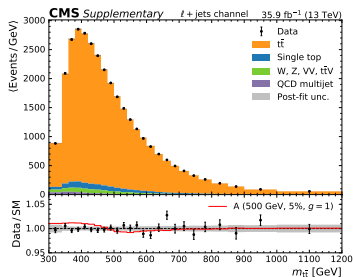
- The constraints are not an artifact of Asimov data set and occur also with **pseudodata**:



- Fit finds $\hat{\theta}$ such that $\lambda(\theta)$ resembles **noise in data** best. Any deviations from $\hat{\theta}$ get penalized by the $\nu^2/2$ term in the likelihood
 - Parameters ν have to be adjusted for a larger difference between the noise patters in data and in $\lambda(\theta)$

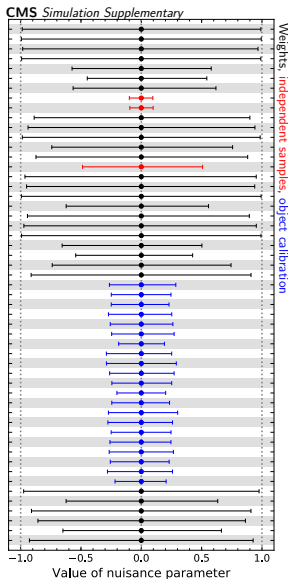
Real-life example: $H \rightarrow t\bar{t}$

- CMS search for heavy Higgs boson decaying to $t\bar{t}$
 - arXiv:1908.01115, submitted to JHEP
- Focus on $\ell + \text{jets}$ channel as an example
 - Reconstructed $m_{t\bar{t}}$ as main observable
 - Angular variable reflects spin
 - Use 2D distribution in statistical analysis
 - Separate $e + \text{jets}$ and $\mu + \text{jets}$ channels
 - 25×5 bins in each channel



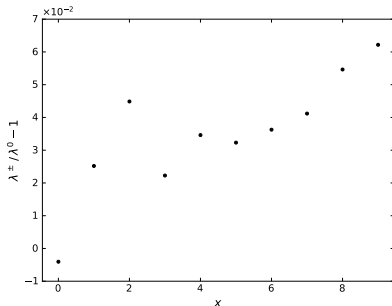
Real-life example: $H \rightarrow t\bar{t}$

- CMS search for heavy Higgs boson decaying to $t\bar{t}$
 - arXiv:1908.01115, submitted to JHEP
- Focus on $\ell + \text{jets}$ channel as an example
 - Reconstructed $m_{t\bar{t}}$ as main observable
 - Angular variable reflects spin
 - Use 2D distribution in statistical analysis
 - Separate $e + \text{jets}$ and $\mu + \text{jets}$ channels
 - 25×5 bins in each channel
- Initially observed unexpectedly tight constraints on some of nuisances
 - Asimov data set, MC stat. uncertainties not shown
 - All these templates affected by fluctuations



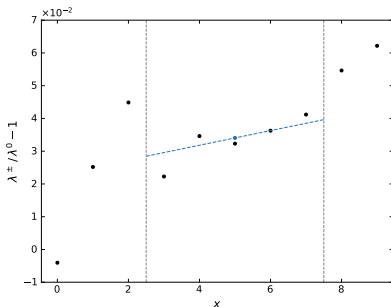
Smoothing

- Suppress fluctuations by smoothing relative deviations $\lambda_i^\pm / \lambda_i^0 - 1$
 - Denser binning than used in the analysis to avoid binning artefacts
 - Assume relative deviations are identical between $e + \text{jets}$ and $\mu + \text{jets}$
 - Assume up and down deviations are symmetric in shape



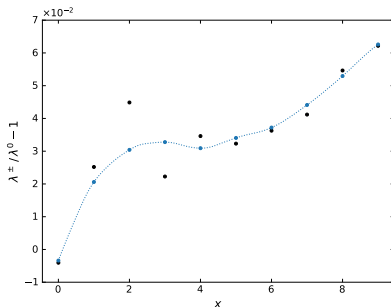
Smoothing

- **Suppress fluctuations** by smoothing relative deviations $\lambda_i^\pm / \lambda_i^0 - 1$
 - Denser binning than used in the analysis to avoid binning artefacts
 - Assume relative deviations are identical between $e + \text{jets}$ and $\mu + \text{jets}$
 - Assume up and down deviations are symmetric in shape
- **Local linear regression** as smoothing algorithm (LOWESS)
 - Weighted least squares fit to 2D rel. deviation with a linear function
 - Restrict the fit to rectangular window
 - Points further away from the centre of window receive smaller weights



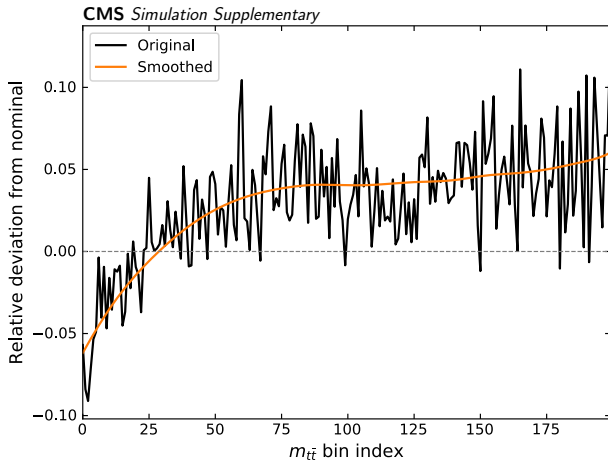
Smoothing

- **Suppress fluctuations** by smoothing relative deviations $\lambda_i^\pm / \lambda_i^0 - 1$
 - Denser binning than used in the analysis to avoid binning artefacts
 - Assume relative deviations are identical between $e + \text{jets}$ and $\mu + \text{jets}$
 - Assume up and down deviations are symmetric in shape
- **Local linear regression** as smoothing algorithm (LOWESS)
 - Weighted least squares fit to 2D rel. deviation with a linear function
 - Restrict the fit to rectangular window
 - Points further away from the centre of window receive smaller weights
 - Repeat while sliding the window
- In each bin, apply smoothed relative deviation to the nominal template to construct **new systematic variations**
 - Rescaled independently for up and down variations minimizing χ^2 error



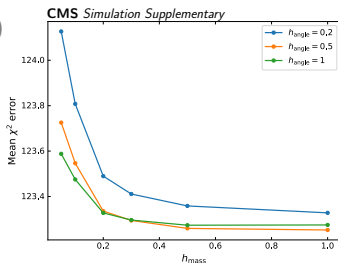
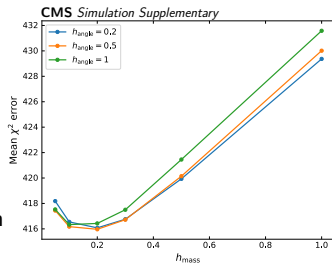
Smoothing

- Example of smoothing (a single angular bin shown)



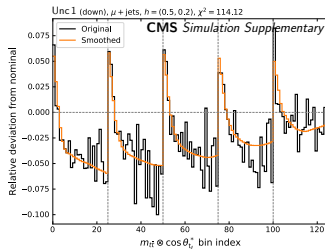
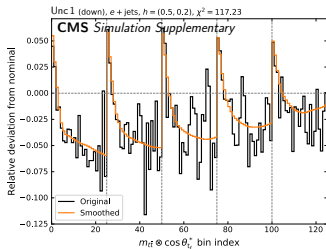
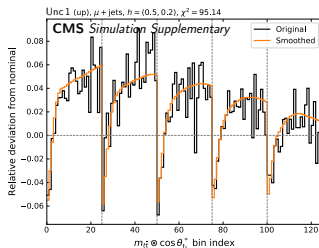
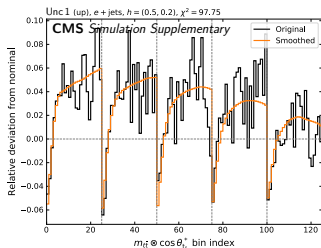
Parameters of smoothing algorithm

- Dimensions of the window ('bandwidths') are **free parameters** of the algorithm
- Chosen with repeated **cross-validation**
 - Split events into $k = 10$ partitions
 - Build smoothed rel. deviations from first $k - 1$ partitions and compute approximation χ^2 error on the k^{th} partition
 - Repeat for the other $k - 1$ possible choices of the test partition
 - Repeat everything with different (random) splittings of events into partitions
 - Choose bandwidths that give smallest average approximation error



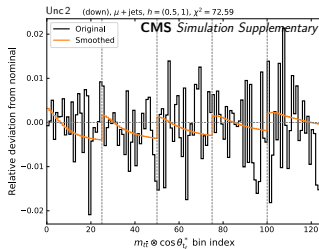
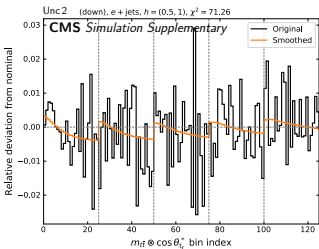
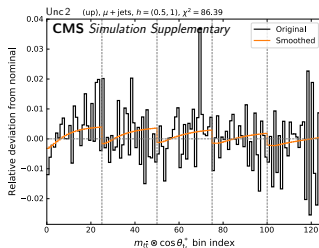
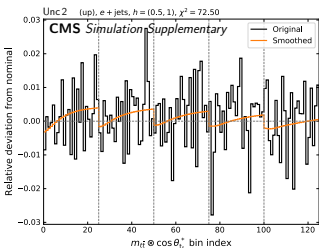
Example of smoothing

- An uncertainty with a real impact



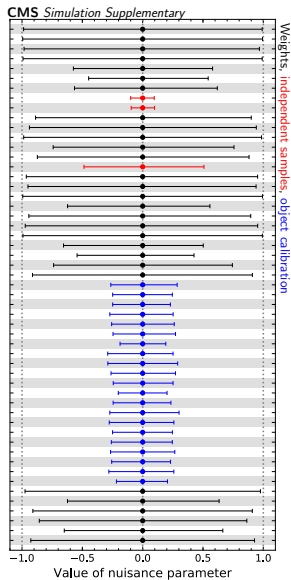
Example of smoothing

- An uncertainty with very little real impact

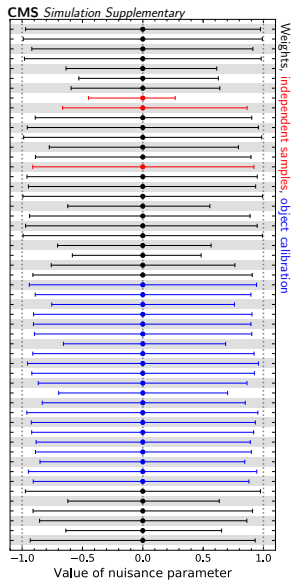


Constraints

Before smoothing



After smoothing



Summary

- Statistical fluctuations in templates representing systematic uncertainties can lead to severe **unphysical constraints** on corresponding nuisances
 - Get tighter as the number of bins grows or $\mathcal{L}_{\text{eff}}^{\text{MC}} / \mathcal{L}^{\text{Data}}$ decreases
 - Becomes especially important as large data sets are collected
 - Not addressed by MC stat. uncertainties in nominal templates
- Intuitive explanation: Fit takes into account not only the physical variation but also similarity between **noise patterns** in data and MC
- **Smoothing** of variations w. r. t. the nominal template can lift or reduce the constraints
 - Local regression was used in CMS search for $H \rightarrow t\bar{t}$, but other options are possible (e. g. smoothing splines or even (regularized) polynomial fit)