Simplified Likelihoods (at least one approach)

Nicholas Wardle

Publication of statistical models: hands-on workshop

Simplified likelihoods

10/11/2021

Common choices for searches at the LHC

General form* for our experimental likelihood (for measurements, searches …) is

$$L(\boldsymbol{\alpha},\boldsymbol{\delta})\pi(\boldsymbol{\delta}) = \prod_{I=1}^{P} \mathrm{Pr}\Big(n_I^{\mathrm{obs}}\,\Big|\,n_I(\boldsymbol{\alpha},\boldsymbol{\delta})\Big)\pi(\boldsymbol{\delta})$$

Where $\boldsymbol{\alpha}$ are the "parameters of interest" (mass of a new hypothetical particle, cross-section for some new process …) and $\boldsymbol{\delta}$ are the "nuisance parameters".

[1] G. Cowan, K. Cranmer, E. Gross, O. Vitells Eur.Phys.J.C71:1554,2011

* For Bayesian approaches $\pi(\boldsymbol{\delta}) \rightarrow \pi(\boldsymbol{\alpha},\boldsymbol{\delta})$

Common choices for searches at the LHC

General form* for our experimental likelihood (for measurements, searches …) is

$$L(\boldsymbol{\alpha},\boldsymbol{\delta})\pi(\boldsymbol{\delta}) = \prod_{I=1}^{P} \mathrm{Pr}\Big(n_I^{\mathrm{obs}} \Big| n_I(\boldsymbol{\alpha},\boldsymbol{\delta})\Big)\pi(\boldsymbol{\delta})$$

Where **α** are the "parameters of interest" (mass of a new hypothetical particle, cross-section for some new process …) and **δ** are the "nuisance parameters".

$$\boldsymbol{\alpha} = \mu$$

At the LHC, the profiled likelihood ratio test statistic is the most common choice [1] → one parameter of interest μ – common multiplier for total signal yield

[1] G. Cowan, K. Cranmer, E. Gross, O. Vitells Eur.Phys.J.C71:1554,2011

* For Bayesian approaches $\pi(\boldsymbol{\delta}) \rightarrow \pi(\boldsymbol{\alpha},\boldsymbol{\delta})$

Common choices for searches at the LHC

General form* for our experimental likelihood (for measurements, searches …) is

$$L(\boldsymbol{\alpha},\boldsymbol{\delta})\pi(\boldsymbol{\delta}) = \prod_{I=1}^{P} \mathrm{Pr}\Big(n_I^{\mathrm{obs}} \Big| n_I(\boldsymbol{\alpha},\boldsymbol{\delta})\Big)\pi(\boldsymbol{\delta})$$

Where **α** are the "parameters of interest" (mass of a new hypothetical particle, cross-section for some new process …) and **δ** are the "nuisance parameters".

$$\boldsymbol{\alpha} = \mu$$

At the LHC, the profiled likelihood ratio test statistic is the most common choice [1] → one parameter of interest μ – common multiplier for total signal yield

$$n_I(\mu,\boldsymbol{\delta}) \rightarrow \mu \cdot \sum_{\mathrm{sigs}} n_{s_k,I} + \sum_{\mathrm{bkgs}} n_{b_k,I}(\boldsymbol{\delta}) \rightarrow \mu \cdot n_{s,I} + n_{b,I}(\boldsymbol{\delta})$$

Sum over the signals / background contributions

$$Pr(n|\lambda) = \frac{\lambda^n}{n!}e^{-\lambda}$$

Often use *binned* likelihood → Pr(.) are Poisson probabilities

[1] G. Cowan, K. Cranmer, E. Gross, O. Vitells Eur.Phys.J.C71:1554,2011

* For Bayesian approaches $\pi(\boldsymbol{\delta}) \rightarrow \pi(\boldsymbol{\alpha},\boldsymbol{\delta})$

Common choices for searches at the LHC

General form* for our experimental likelihood (for measurements, searches …) is

$$L(\boldsymbol{\alpha},\boldsymbol{\delta})\pi(\boldsymbol{\delta}) = \prod_{I=1}^{P} \mathrm{Pr}\Big(n_I^{\mathrm{obs}} \Big| n_I(\boldsymbol{\alpha},\boldsymbol{\delta})\Big)\pi(\boldsymbol{\delta})$$

Where **α** are the "parameters of interest" (mass of a new hypothetical particle, cross-section for some new process …) and **δ** are the "nuisance parameters".

$$\boldsymbol{\alpha} = \mu$$

At the LHC, the profiled likelihood ratio test statistic is the most common choice [1] → one parameter of interest μ – common multiplier for total signal yield

$$n_I(\mu,\boldsymbol{\delta}) \rightarrow \mu \cdot \sum_{\mathrm{sigs}} n_{s_k,I} + \sum_{\mathrm{bkgs}} n_{b_k,I}(\boldsymbol{\delta}) \rightarrow \mu \cdot n_{s,I} + n_{b,I}(\boldsymbol{\delta})$$

Sum over the signals / background contributions

$$Pr(n|\lambda) = \frac{\lambda^n}{n!}e^{-\lambda}$$

Often use *binned* likelihood → Pr(.) are Poisson probabilities

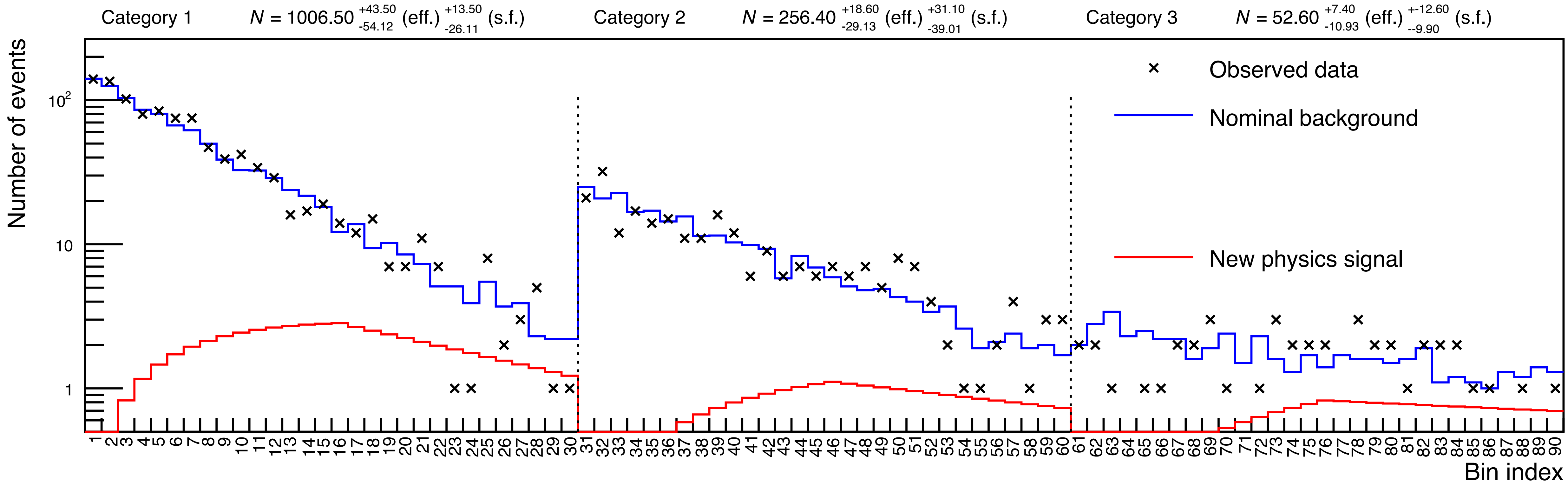
$$\pi(\boldsymbol{\delta})$$

Nuisance parameter priors and/or "in-situ" measurements of **δ**

[1] G. Cowan, K. Cranmer, E. Gross, O. Vitells Eur.Phys.J.C71:1554,2011

* For Bayesian approaches $\pi(\boldsymbol{\delta}) \rightarrow \pi(\boldsymbol{\alpha},\boldsymbol{\delta})$

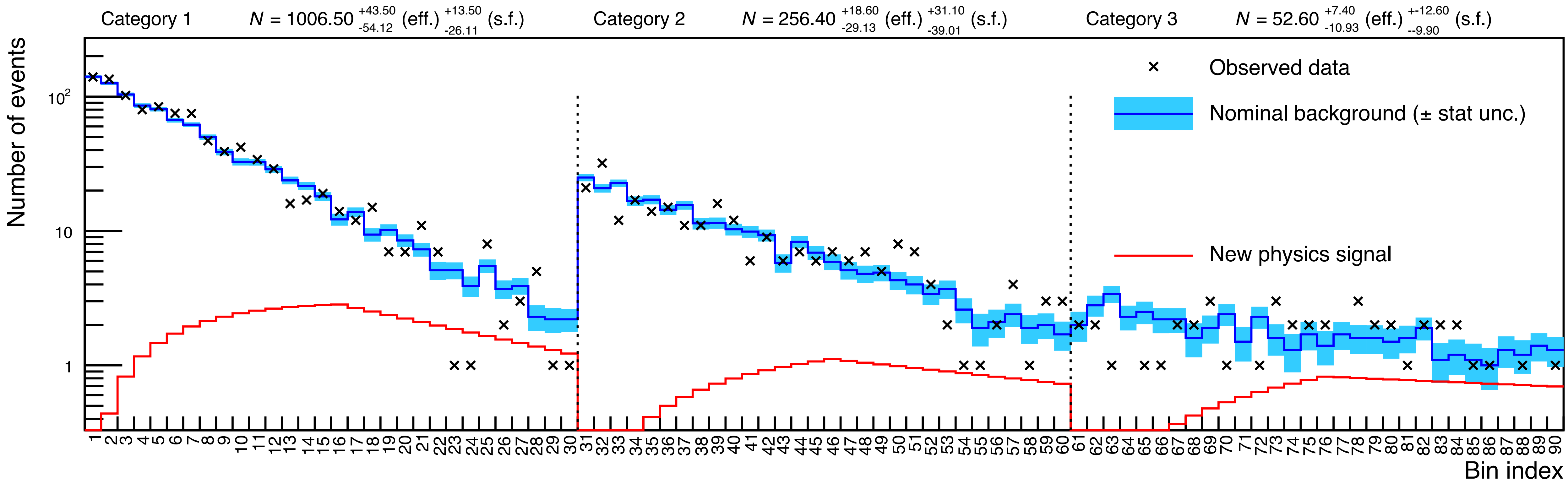
Toy search for new physics



Imagine a (rather simplified) model inspired by a typical search for some Supersymmetric particle or exotic signature.

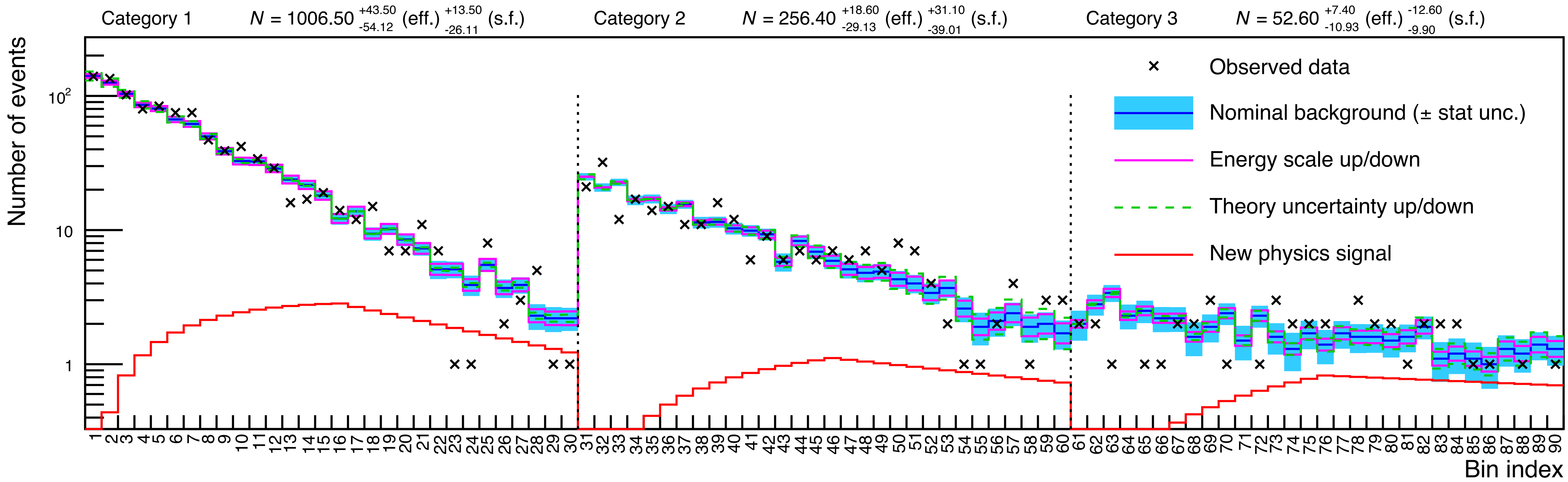
- There is a single source of background (can also think of this as the sum of all backgrounds)
- The data (observations) are divided into regions we have;
    - 3 categories for the data → each category has 30 bins
    - Increasing S/B with bin-number, within each category

Toy search for new physics



There are ***two*** uncertainties (labelled "efficiency" and "scale-factor") on the background yields (*N*), and ***each bin*** has an uncertainty which is uncorrelated between bins (e.g this could be from limited Monte Carlo statistics used to estimate n_l)

Toy search for new physics

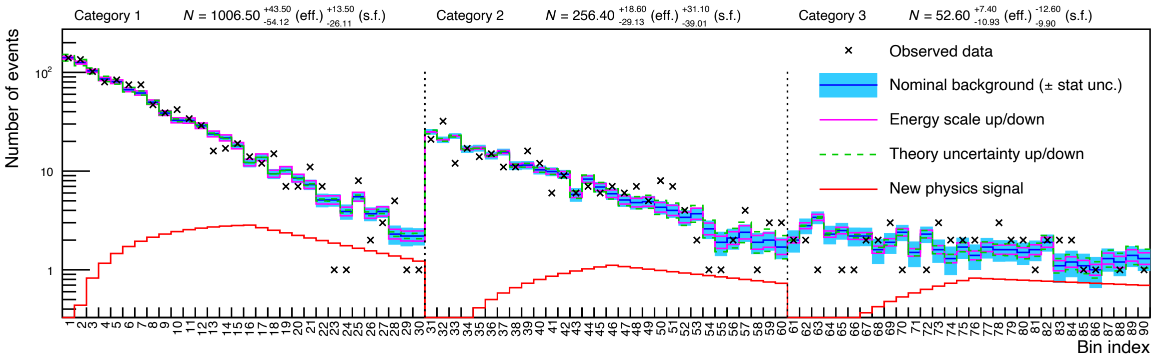


There are ***two*** uncertainties (labelled "efficiency" and "scale-factor") on the background yields (*N*), and ***each bin*** has an uncertainty which is uncorrelated between bins (e.g this could be from limited Monte Carlo statistics used to estimate n_l)

Another ***two*** uncertainties correlated between bins ("energy scale" and "theory" uncertainty)

In total this means 94 nuisance parameters

Toy search for new physics



Think of the expected number of background events in a given bin *I*, as the fraction of events in that bin (f_I) multiplied by the total number of events (*N*)

$$n_I(\boldsymbol{\delta}) \equiv f_I(\boldsymbol{\delta})N(\boldsymbol{\delta})$$

$\boldsymbol{\delta}$ are nuisance parameters representing ***independent*** sources of uncertainty (in our case 94 of them)

$$N(\boldsymbol{\delta}) = N^0 \cdot \prod_j (1+K_j)^{\delta_j}$$

Uncertainties in the normalisation (N) typically follow log-normals

$$\frac{n_I(\boldsymbol{\delta})}{n_I^0} = \prod_j (1+\epsilon_{Ij})^{\delta_j}$$

Similarly for un-correlated bin-by-bin uncertainties

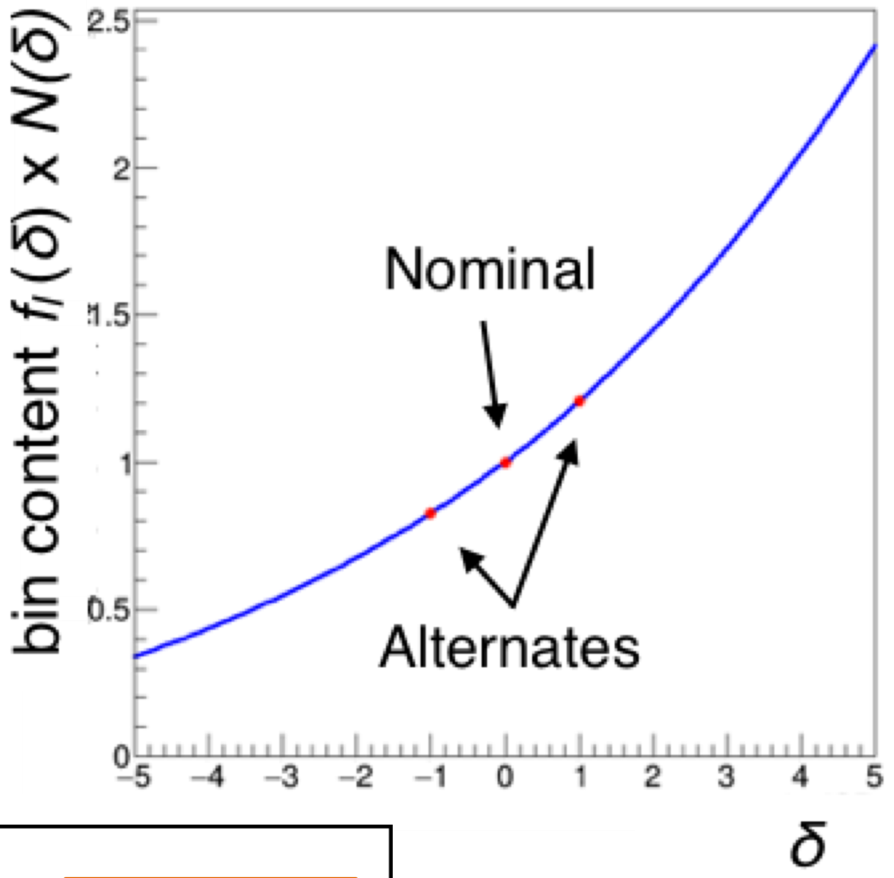
K_j *and* ε_{Ij} represent the relative size and direction of the uncertainty

Toy search for new physics

The effects of correlated systematic uncertainties on n_I are modelled using quadratic(linear) **interpo**(**extrapo**)lation function

$$\boxed{f_I(\boldsymbol{\delta}) = f_I^0 \cdot \frac{1}{F(\boldsymbol{\delta})} \prod_j p_{Ij}(\delta_j)}$$

$F(\boldsymbol{\delta}) = \sum_I f_I(\boldsymbol{\delta})$



$$p_{Ij}(\delta_j) = \begin{cases} \frac{1}{2}\delta_j(\delta_j - 1)\kappa^-_{Ij} - (\delta_j - 1)(\delta_j + 1) + \frac{1}{2}\delta_j(\delta_j + 1)\kappa^+_{Ij} & \text{for } |\delta_j| < 1 \\ \left[\frac{1}{2}(3\kappa^+_{Ij} + \kappa^-_{Ij}) - 2\right]\delta_j - \frac{1}{2}(\kappa^+_{Ij} + \kappa^-_{Ij}) + 2 & \text{for } \delta_j > 1 \\ \left[2 - \frac{1}{2}(3\kappa^-_{Ij} + \kappa^+_{Ij})\right]\delta_j - \frac{1}{2}(\kappa^+_{Ij} + \kappa^-_{Ij}) + 2 & \text{for } \delta_j < -1 \end{cases}$$

Experimental likelihood

Now we can write the likelihood for this search as follows;

$$L(\mu, \boldsymbol{\delta})\pi(\boldsymbol{\delta}) = \prod_{I=1}^{90} P(n_I^{\mathrm{obs}} | \mu \cdot n_{s,I} + n_{b,I}(\boldsymbol{\delta})) \cdot \prod_{j=1}^{94} e^{-\delta_j^2}$$

$$n_{b,I}(\boldsymbol{\delta}) = N_c^0 \cdot \prod_{k=1}^{2} (1 + K_k)^{\delta_k} \cdot f_I^0 \cdot \frac{1}{F(\boldsymbol{\delta})} \prod_{j=3}^{4} p_{I,j}(\delta_j) \cdot (1 + \epsilon_I)^{\delta_I}$$

Experimental likelihood

Now we can write the likelihood for this search as follows;

$$L(\mu, \boldsymbol{\delta})\pi(\boldsymbol{\delta}) = \prod_{I=1}^{90} P(n_I^{\mathrm{obs}} | \mu \cdot n_{s,I} + n_{b,I}(\boldsymbol{\delta})) \cdot \prod_{j=1}^{94} e^{-\delta_j^2}$$

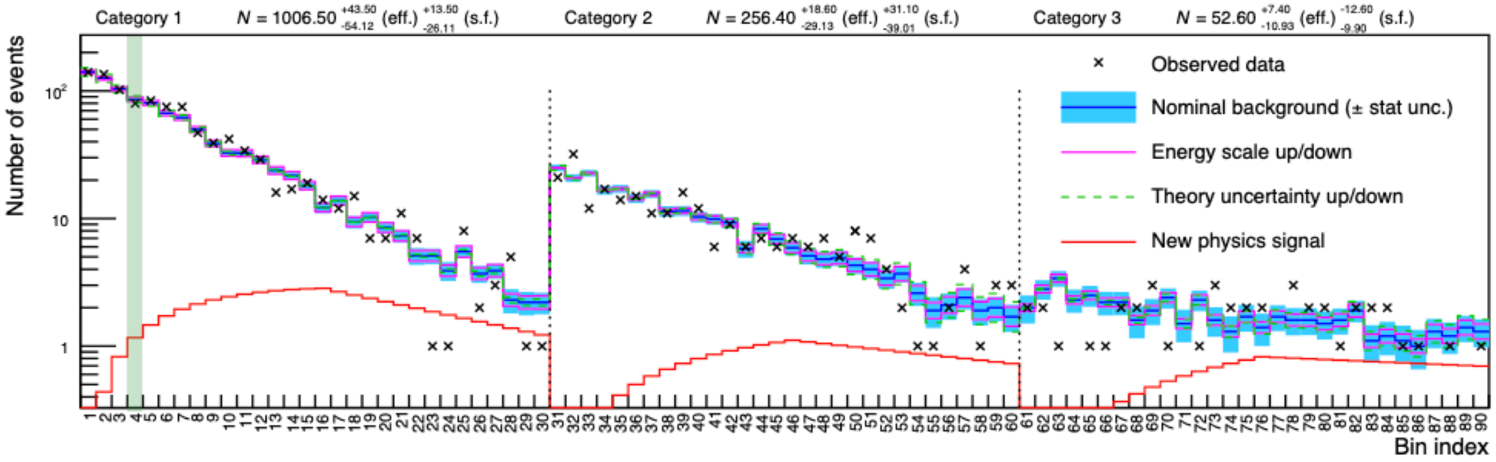
$$n_{b,I}(\boldsymbol{\delta}) = N_c^0 \cdot \prod_{k=1}^{2} (1 + K_k)^{\delta_k} \cdot f_I^0 \cdot \frac{1}{F(\boldsymbol{\delta})} \prod_{j=3}^{4} p_{I,j}(\delta_j) \cdot (1 + \epsilon_I)^{\delta_I}$$

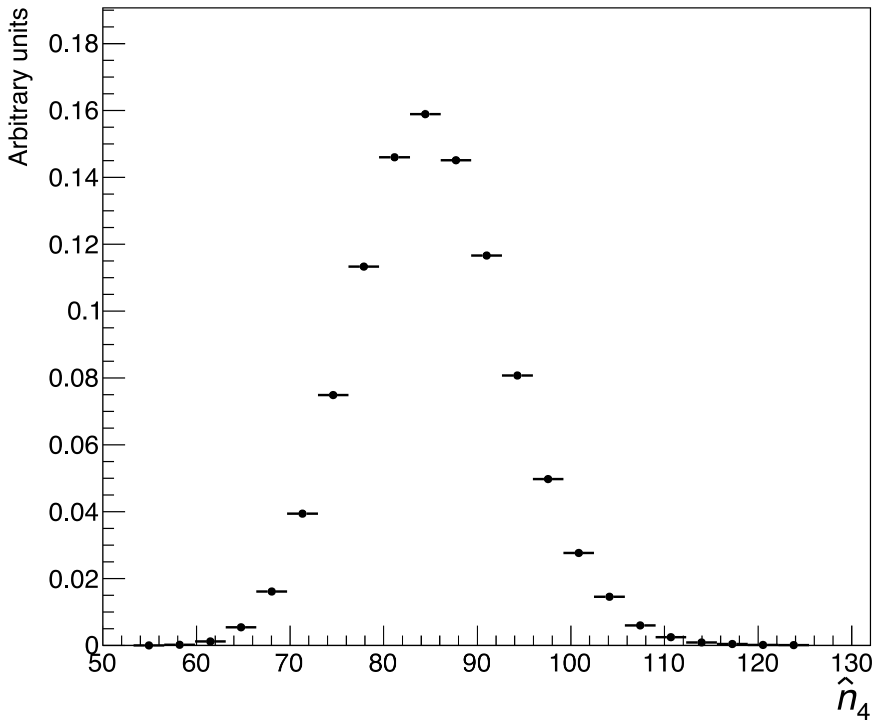
Specifying these terms with this generic form means the full likelihood can be communicated as plain text!

A lot of physicists' time working on an LHC search is spent on **these**!

Re-parameterize the backgrounds

We can generate pseudo-experiments for $n_{b,I}$ since we know $p(\boldsymbol{\delta}) := \pi(\delta) \sim e^{-\frac{1}{2}\boldsymbol{\delta}\cdot\boldsymbol{\delta}}$
Use randomly sampled $\boldsymbol{\delta}'$ and $\hat{n}_I = n_{b,I}(\boldsymbol{\delta}')$ to determine the distribution of the backgrounds…



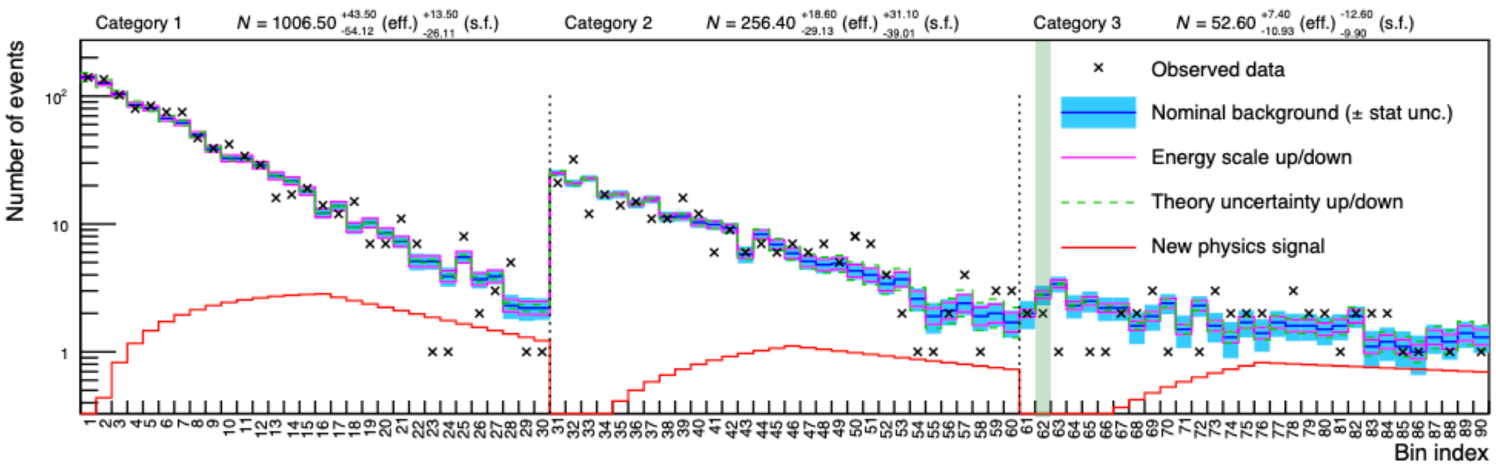


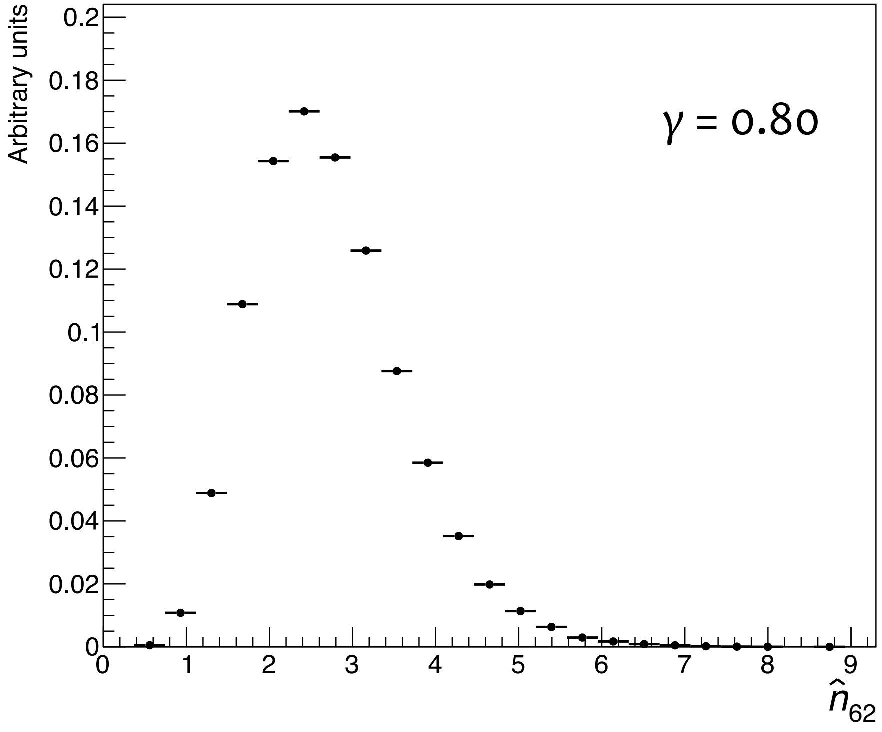
In some bins, distributions looks symmetric and Gaussian

→ can be described by 2 moments (mean and variance)

Re-parameterize the backgrounds

We can generate pseudo-experiments for $n_{b,I}$ since we know $p(\boldsymbol{\delta}) := \pi(\delta) \sim e^{-\frac{1}{2}\boldsymbol{\delta}\cdot\boldsymbol{\delta}}$
Use randomly sampled $\boldsymbol{\delta}'$ and $\hat{n}_I = n_{b,I}(\boldsymbol{\delta}')$ to determine the distribution of the backgrounds…





In other cases however, distributions are very asymmetric
→ Skewness (γ) provides a measure of asymmetry
→ 3rd moment relevant for describing backgrounds

$$\gamma = \frac{m_3}{(m_2)^{\frac{3}{2}}}$$

Simplifying the likelihood?

For statistical (re-) interpretation purposes we eliminate nuisance parameters (**δ**)

→ We are mainly interested in profiled / marginalized likelihoods $L(\mu, \boldsymbol{\delta}) \rightarrow L(\mu)$

Since the "backgrounds" are only dependent on the nuisance parameters, we can approximate in such a way that the profiled (or marginal) likelihood is preserved as follows [1];

1. Express $n_{b,I}$ as a simple expansion (quadratic) in terms of **combined nuisance parameters** ϑ_I

$$n_{b,I} \simeq a_I + b_I\theta_I + c_I\theta_I^2$$

$I = 1 \ldots 90$

[1] A. Buckley, M. Citron, S. Fichet, S. Kraml, W. Waltenberger, **NW** *J. High Energ. Phys.* **2019,** 64 (2019)

* We can restore $\mu \cdot n_{s,I} \rightarrow n_{s,I}(\boldsymbol{\alpha})$ if needed, but for this toy we keep μ

Simplifying the likelihood?

For statistical (re-) interpretation purposes we eliminate nuisance parameters (**δ**)

→ We are mainly interested in profiled / marginalized likelihoods $L(\mu, \boldsymbol{\delta}) \rightarrow L(\mu)$

Since the "backgrounds" are only dependent on the nuisance parameters, we can approximate in such a way that the profiled (or marginal) likelihood is preserved as follows [1];

1. Express $n_{b,I}$ as a simple expansion (quadratic) in terms of **combined nuisance parameters** ϑ_I

$$n_{b,I} \simeq a_I + b_I\theta_I + c_I\theta_I^2$$

$I = 1 \ldots 90$

2. Re-parameterize likelihood in terms of μ* and ϑ_I → Need to derive $\pi(\boldsymbol{\vartheta})$!

$$L(\mu, \boldsymbol{\delta})\pi(\boldsymbol{\delta}) \rightarrow L(\mu, \boldsymbol{\theta})\pi(\boldsymbol{\theta}) = \prod_{I=1}^{P=90} P(n_I^{\text{obs}} | \mu \cdot n_{s,I} + a_I + b_I\theta_I + c_I\theta_I^2) \cdot \frac{1}{\sqrt{(2\pi)^P}} e^{-\frac{1}{2}\boldsymbol{\theta}^T\boldsymbol{\rho}^{-1}\boldsymbol{\theta}}$$

P(*x*|*y*) = Poisson probability as before

These are the same as the *full* likelihood

$\rho_{I,J} = \rho_{J,I}$

[1] A. Buckley, M. Citron, S. Fichet, S. Kraml, W. Waltenberger, **NW** *J. High Energ. Phys.* **2019,** 64 (2019)

* We can restore $\mu \cdot n_{s,I} \rightarrow n_{s,I}(\boldsymbol{\alpha})$ if needed, but for this toy we keep μ

Nearly done with the formulae…

Coefficients obtained by matching moments and appealing to CLT at NLO.

Coefficients *a*, *b* and *c* are determined from the first 3 central moments of the joint distributions of $n_{b,I}$ - Mean, covariance ***and skew***

Solutions valid for $$\frac{8(m_{2,II})^3}{(m_{3,I})^2} \geq 1$$

$$c_I = -\mathrm{sign}(m_{3,I})\sqrt{2m_{2,II}}\cos\left(\frac{4\pi}{3} + \frac{1}{3}\arctan\left(\sqrt{8\frac{m_{2,II}^3}{m_{3,I}^2} - 1}\right)\right)$$

$$b_I = \sqrt{m_{2,II} - 2c_I^2}\,,$$

$$a_I = m_{1,I} - c_I\,,$$

$$\rho_{IJ} = \frac{1}{4c_I c_J}\left(\sqrt{(b_I b_J)^2 + 8c_I c_J\, m_{2,IJ}} - b_I b_J\right).$$

Nearly done with the formulae…

Coefficients obtained by matching moments and appealing to CLT at NLO.

$$c_I = -\mathrm{sign}(m_{3,I})\sqrt{2m_{2,II}}\cos\left(\frac{4\pi}{3} + \frac{1}{3}\arctan\left(\sqrt{8\frac{m_{2,II}^3}{m_{3,I}^2} - 1}\right)\right)$$

Coefficients *a*, *b* and *c* are determined from the first 3 central moments of the joint distributions of $n_{b,I}$ - Mean, covariance ***and skew***

$$b_I = \sqrt{m_{2,II} - 2c_I^2}\,,$$

$$a_I = m_{1,I} - c_I\,,$$

$$\rho_{IJ} = \frac{1}{4c_Ic_J}\left(\sqrt{(b_Ib_J)^2 + 8c_Ic_J\,m_{2,IJ}} - b_Ib_J\right).$$

Solutions valid for $\dfrac{8(m_{2,II})^3}{(m_{3,I})^2} \geq 1$

Moments can be calculated analytically or (my preference) using pseudo experiments

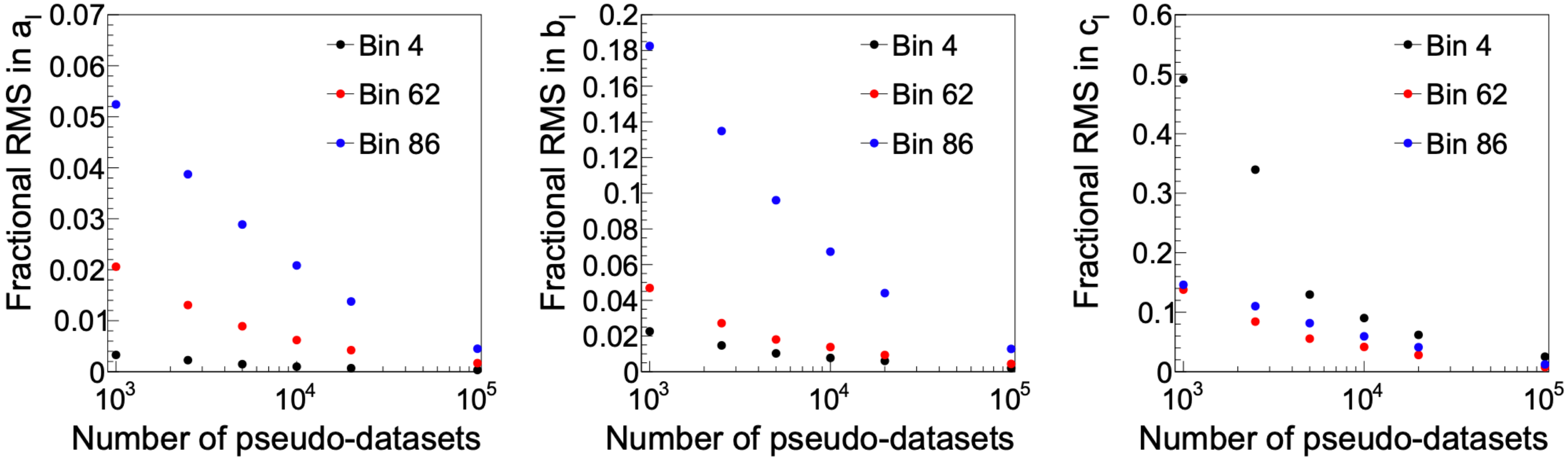
$$\boxed{m_{1,I}} = \mathbf{E}[\hat{n}_I]$$

$$\boxed{m_{2,IJ}} = \mathbf{E}[(\hat{n}_I - \mathbf{E}[\hat{n}_I])(\hat{n}_J - \mathbf{E}[\hat{n}_J])]$$

$$\boxed{m_{3,I}} = \mathbf{E}[(\hat{n}_I - \mathbf{E}[\hat{n}_I])^3]$$

These quantities are the inputs needed to determine the ***simplified likelihood***

Convergence of moment calculation with pseudo-data



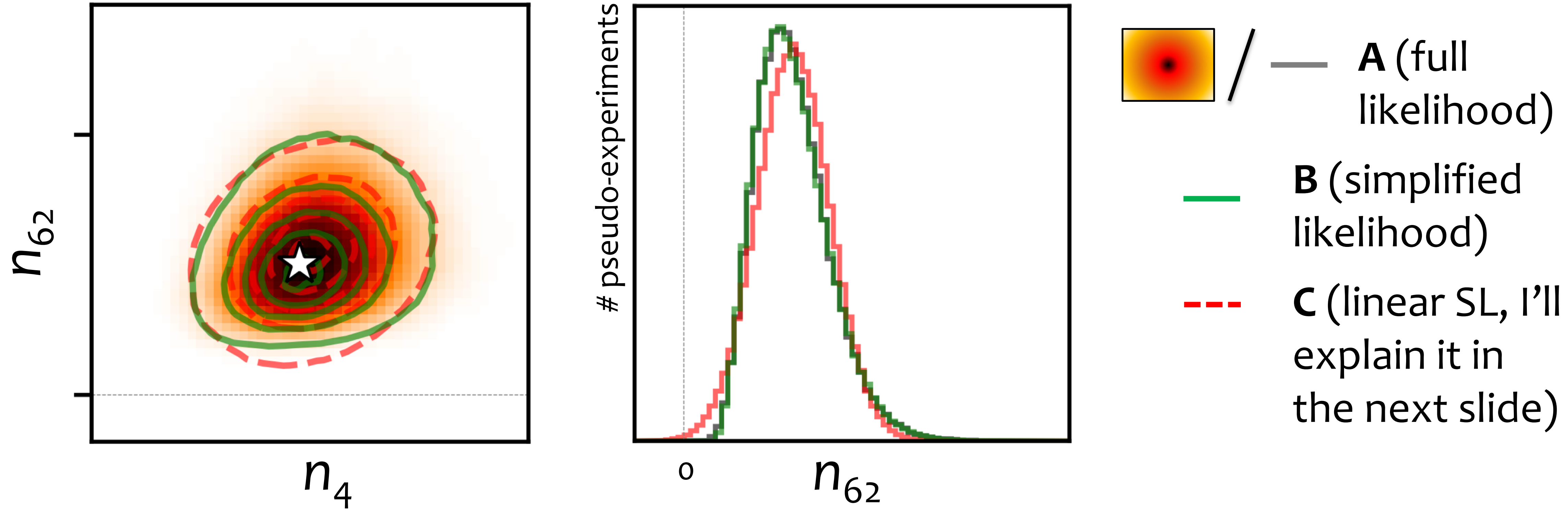
3rd Moment typically requires most toys to get accurate value, however this is mostly true when m_3 is small and therefore not so relevant!

How well does this approximate the distribution of n$_I$?

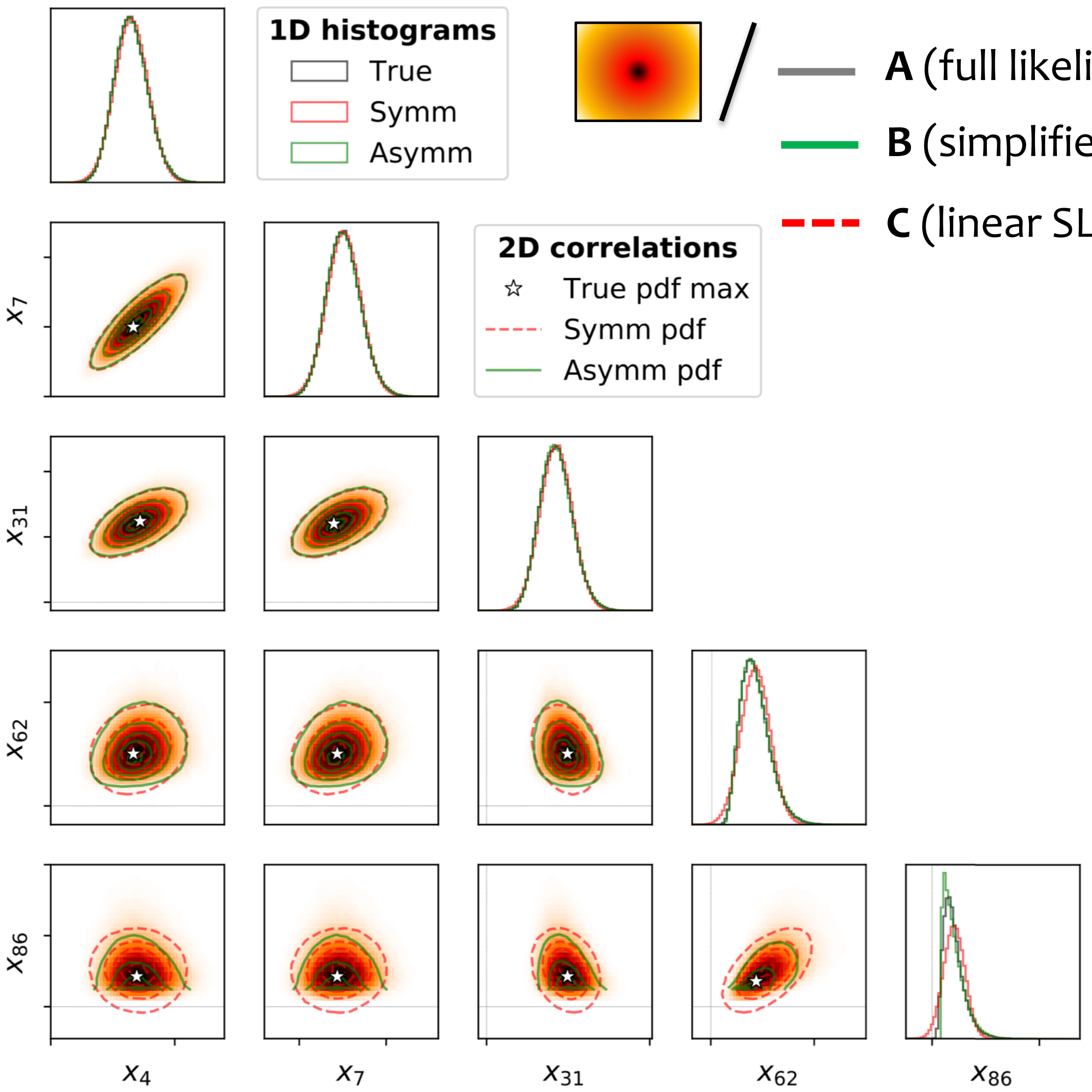
We can compare the distribution of $\hat{n}_I$ obtained in the pseudo-data from

A. $\hat{n}_I = n_{b,I}(\boldsymbol{\delta}')$ generating from $p(\boldsymbol{\delta}) := \pi(\delta) \sim e^{-\frac{1}{2}\boldsymbol{\delta}\cdot\boldsymbol{\delta}}$

B. $\hat{n}_I = n_{b,I}(\theta'_I)$ generating from $p(\boldsymbol{\theta}) \sim e^{-\frac{1}{2}\boldsymbol{\theta}^T\rho^{-1}\boldsymbol{\theta}}$



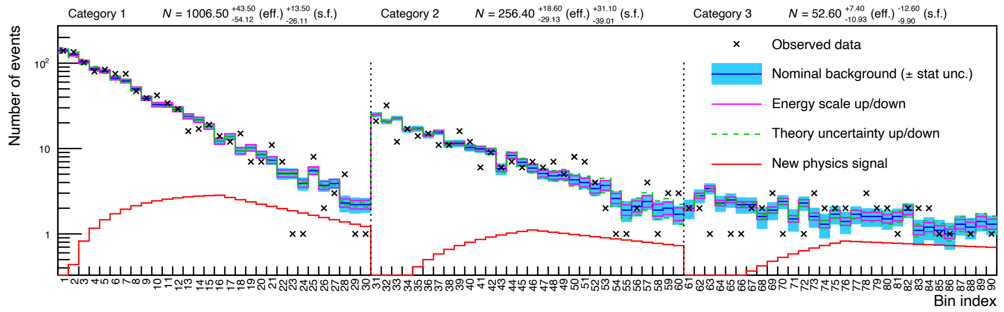
How well does this approximate the distribution of n_I?



A (full likelihood)

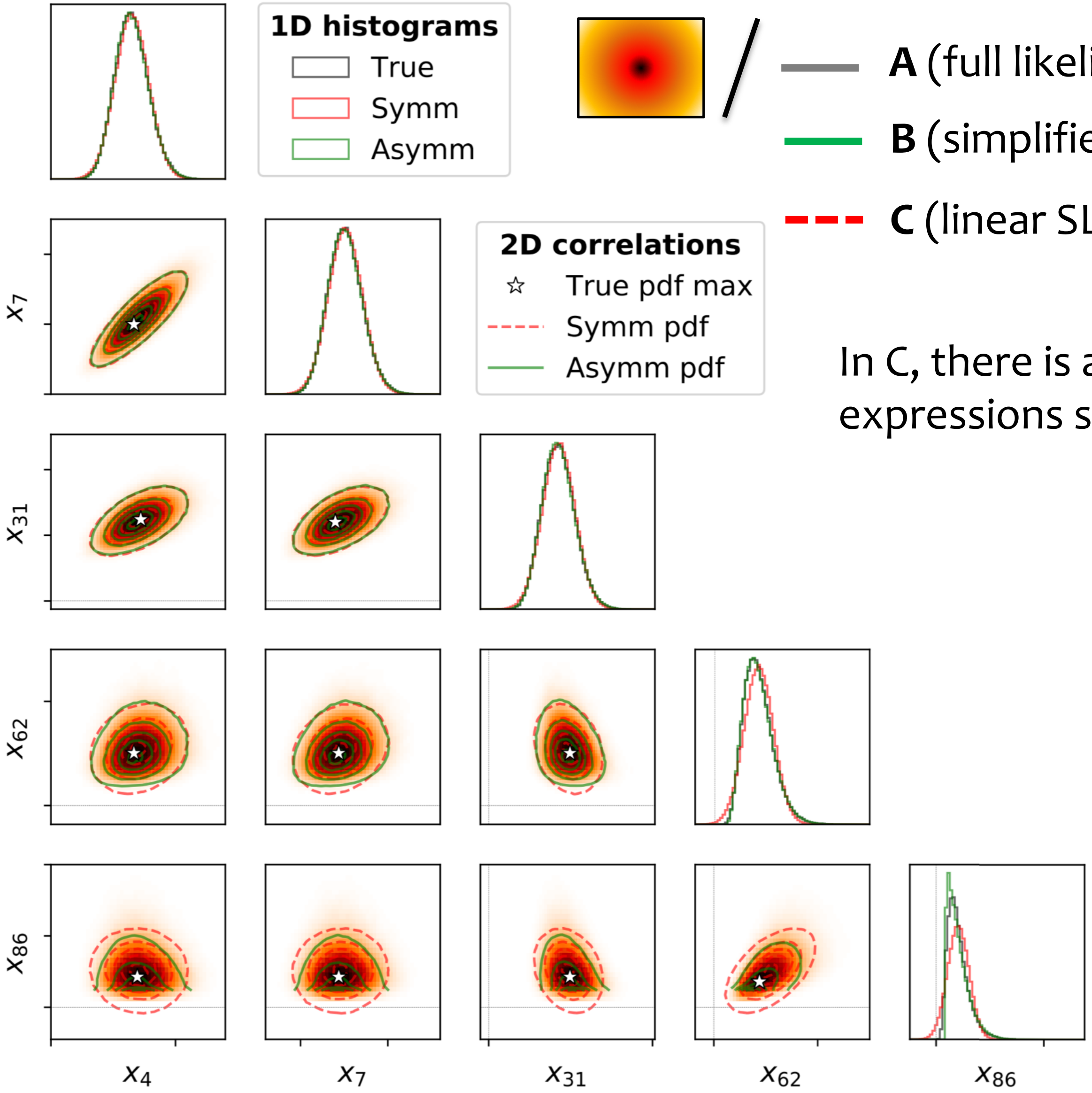
B (simplified likelihood)

C (linear SL)



*approach as in [CMS-NOTE-2017-001](), and K. Cranmer, S. Kreiss, D. López-Val, T. Plehn, [PhysRevD 91 054032]()

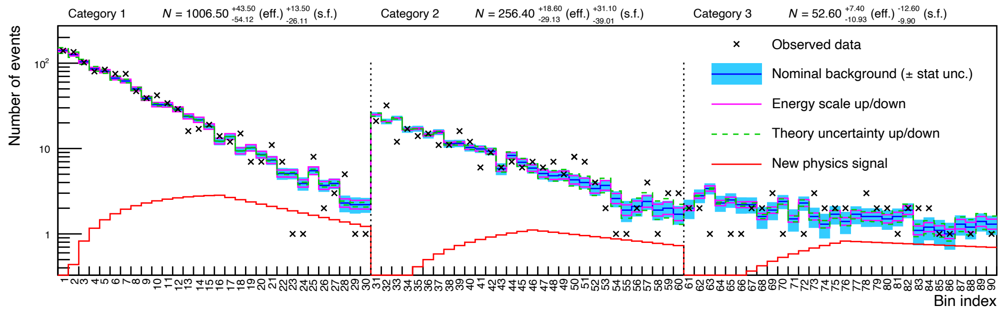
How well does this approximate the distribution of n_I?



A (full likelihood)

B (simplified likelihood)

C (linear SL)



In C, there is a further simplification that $m_{3,I}$ is 0. In this case, the expressions simplify to*

$$n_{b,I}(\theta_I) = A_I + B_I\theta_I$$

$$p(\boldsymbol{\theta}) \sim e^{-\frac{1}{2}\boldsymbol{\theta}^T\boldsymbol{v}^{-1}\boldsymbol{\theta}}$$

$$A_I = m_{1,I},\;\; B_I = m_{2,II},\;\; v_{IJ} = m_{2,IJ}$$

When $m_{3,I}/(m_{2,II})^{\frac{3}{2}}$ (the skew) is small, the linear approximation is fairly good, as expected.

*approach as in CMS-NOTE-2017-001, and K. Cranmer, S. Kreiss, D. López-Val, T. Plehn, PhysRevD 91 054032

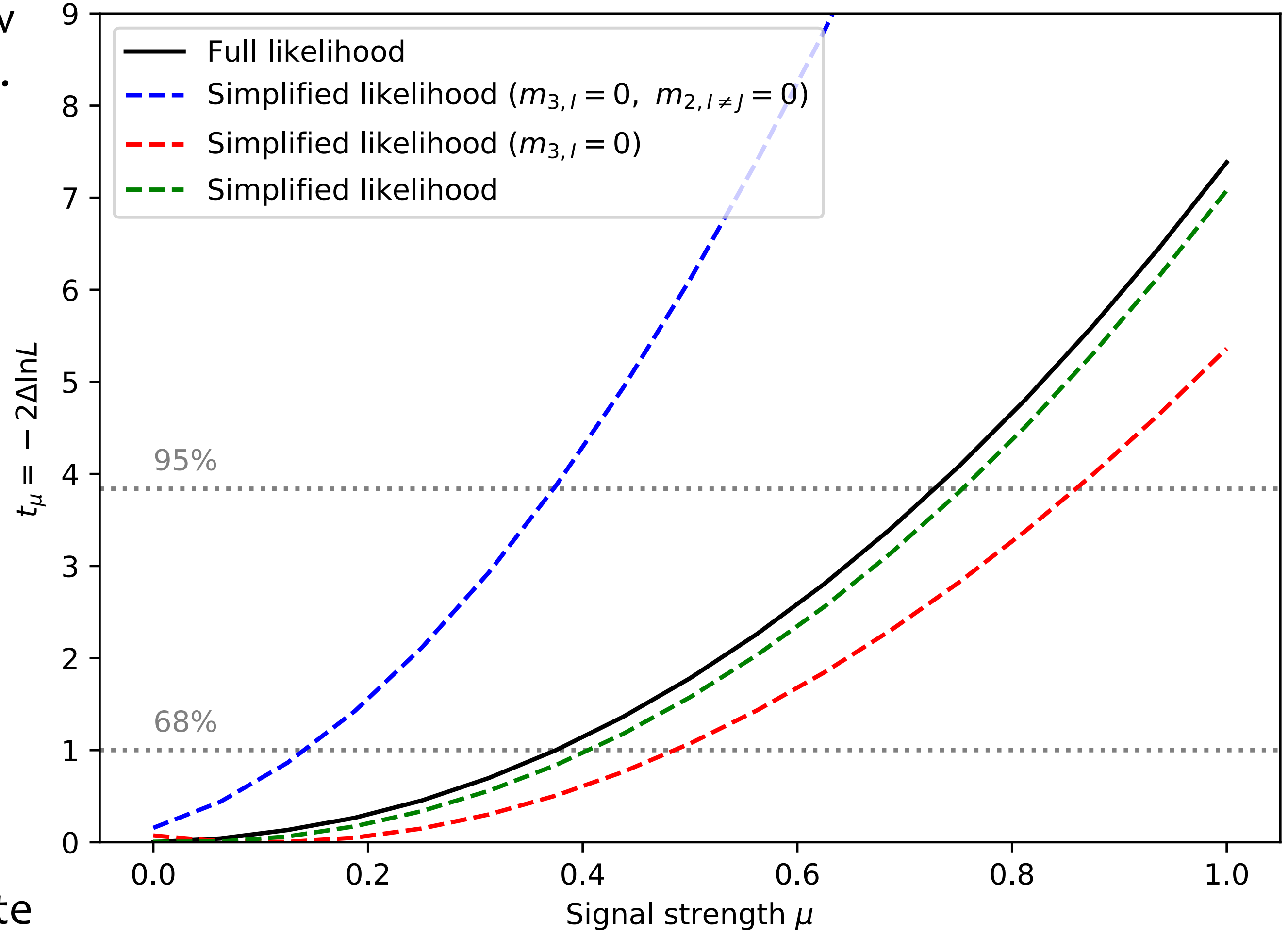
Get to the punchline already Nick …

Eliminating nuisance parameters (δ or θ) indicates how *accurately* we can reproduce statistical interpretations.

e.g. the profiled likelihood ratio test-statistic* is used to set limits on new physics processes at the LHC

$$t_\mu = -2\ln\frac{L_{\mathrm{S}}^{\max}(\mu)}{L_{\mathrm{S}}^{\max}}$$

$$L_{\mathrm{S}}^{\max}(\mu) = \max_{\theta_{\mathrm{I}}}\left\{\mathrm{L_S}(\mu,\boldsymbol{\theta})\right\}$$



Inputs for toy search uploaded to [HepData]

Public scipy-based code to calculate SL coefficients and run statistical tests on [GitLab]

*No reason why we couldn't have marginalised the likelihood to compare Bayesian posterior distributions instead of profiling.

Discussion

Can we implement this in phHF simplification routines ?

Some things to mull over

→ One only needs to calculate moments in different signal region bins : use MC (as we do in CMS) or propagate directly and use logL derivatives?

→ Signal region vs control regions : For simplification, assume only interested in signal region (control data summarized also in co-variance/skews)

→ If using CRs and not including in procedure, ideally use post-fit estimates for generating the toys (include CRs in fit but not SRs to avoid double counting!)

Backup slides

Simplified likelihood log-likelihood

$$\ln\big(L_{\mathrm{S}}(\mu,\boldsymbol{\theta})\pi(\boldsymbol{\theta})\big) = \sum_{I}^{P}\left[n_I^{\mathrm{obs}}\ln\left(\mu n_{s,I}+n_{b,I}(\boldsymbol{\theta})\right)-\left(\mu n_{s,I}+n_{b,I}(\boldsymbol{\theta})\right)-n_I^{\mathrm{obs}}!\right] - \frac{1}{2}\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{\rho}^{-1}\boldsymbol{\theta}-\frac{P}{2}\ln 2\pi \tag{B.1}$$

$$\frac{\partial\ln L_{\mathrm{S}}}{\partial\mu} = \sum_{I}^{P}\left(\frac{n_I^{\mathrm{obs}}}{\mu n_{s,I}+n_{b,I}(\boldsymbol{\theta})}-1\right)\cdot n_{s,I} \tag{B.2}$$

$$\frac{\partial\ln L_{\mathrm{S}}}{\partial\theta_A} = \left(\frac{n_A^{\mathrm{obs}}}{\mu n_{s,A}+n_{b,A}(\boldsymbol{\theta})}-1\right)\cdot\left(b_A+2c_A\theta_A\right)-\sum_{I}^{P}\rho_{AI}^{-1}\,\theta_I\ , \tag{B.3}$$

Analytic simplified likelihood coefficients

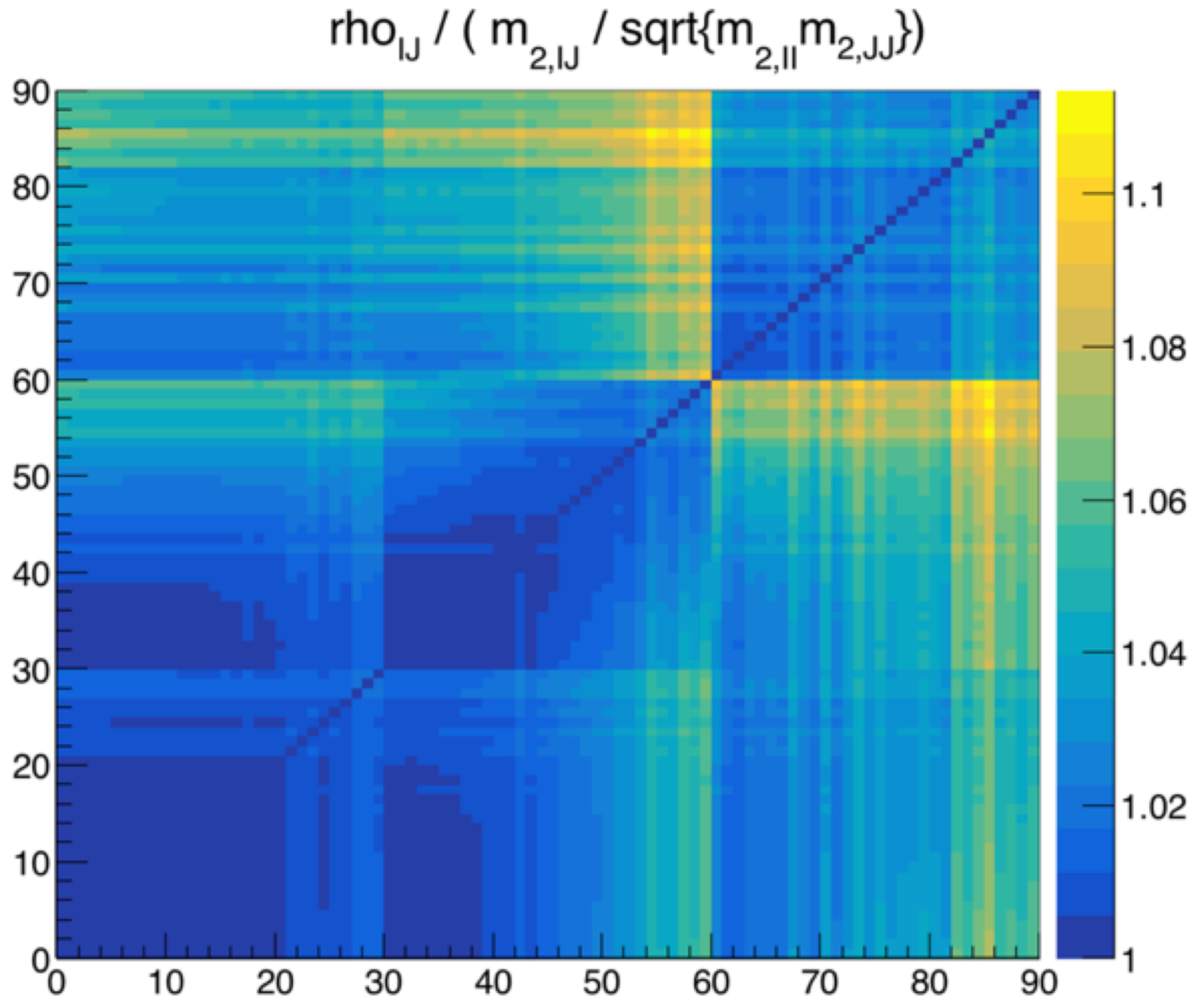
$$a_I = n_I^0 \left(1 + \mathrm{tr}\, \Delta_{2,I} - \frac{1}{6} \sum_{i=1}^{N} \gamma_i (\Delta_{1,I,i})^3 + O(\Delta^4) \right) \,,$$

$$b_I = a_I \left(\Delta_{1,I}^{\mathrm{T}} . \Delta_{1,I} + 2 \sum_{i=1}^{N} \gamma_i \Delta_{1,I,i} \Delta_{2,I,i} + O(\Delta^4) \right)^{1/2} \,,$$

$$\rho_{IJ} = \frac{a_I a_J}{b_I b_J} \left(\Delta_{1,I}^{\mathrm{T}} . \Delta_{1,J} + \sum_{i=1}^{N} \gamma_i (\Delta_{1,I,i} \Delta_{2,J,i} + \Delta_{1,J,i} \Delta_{2,I,i}) \right) + O(\Delta^4) \,,$$

$$c_I = \frac{a_I}{6} \sum_{i=1}^{N} \gamma_i (\Delta_{1,i})^3 + O(\Delta^4) \,,$$

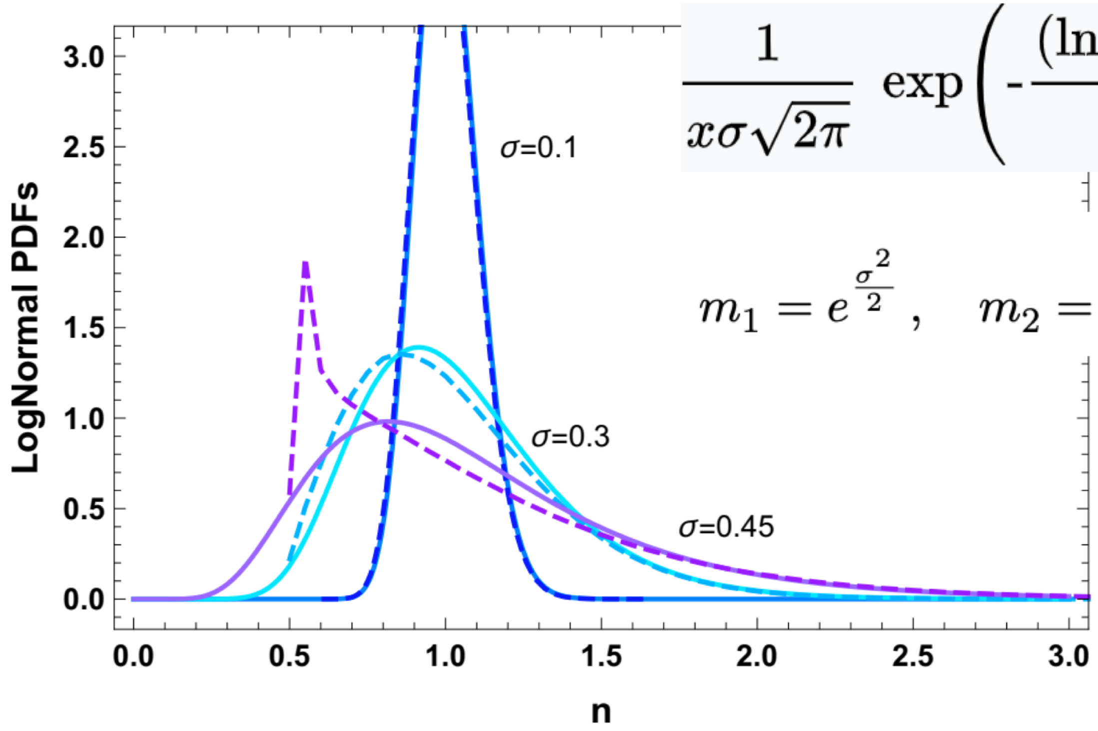
Corrections to correlations



NSL definition of correlation modified due to skew term

Ratio of ρIJ to linear correlation shows up to 15% correction in toy model

SL approximation for a log-normal



$$\frac{1}{x\sigma\sqrt{2\pi}}\exp\left(-\frac{(\ln x-\mu)^2}{2\sigma^2}\right)$$

$$m_1 = e^{\frac{\sigma^2}{2}}, \quad m_2 = e^{2\sigma^2} - e^{\sigma^2}, \quad m_3 = e^{\frac{9\sigma^2}{2}} - 3e^{\frac{5\sigma^2}{2}} + 2e^{\frac{3\sigma^2}{2}}$$

Figure 1. The log normal PDFs and corresponding normal approximations for $\sigma = 0.1$, 0.3 and 0.45 are shown in blue, cyan and purple respectively. Solid curves show the true distributions, dashed curves show the approximate distributions.