Science & Technology Facilities Council

Imperial College London

Simplified Likelihoods for BSM re-interpretations

Andy Buckley, Matthew Citron, Sylvain Fichet, Sabine Kraml, Wolfgang Waltenberger, **Nicholas Wardle**

Reinterpretation2018: (Re)interpreting the results of new physics searches at the LHC

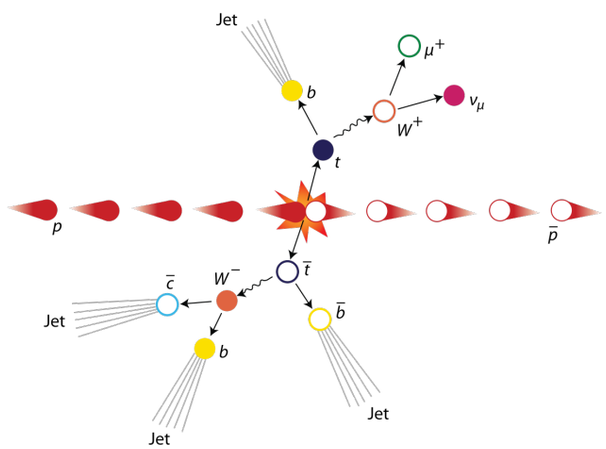
14-16 May 2018 - CERN

Searches for BSM

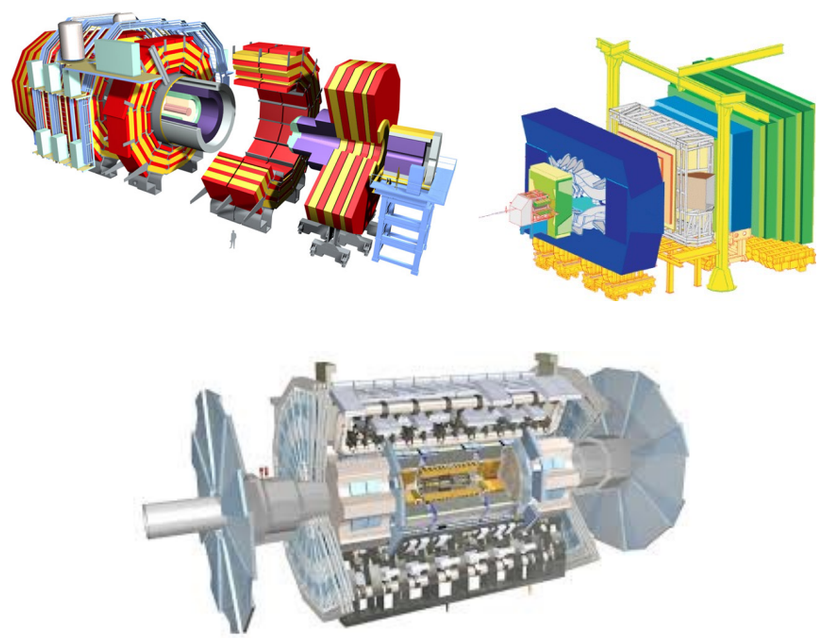
Truth/Gen Level

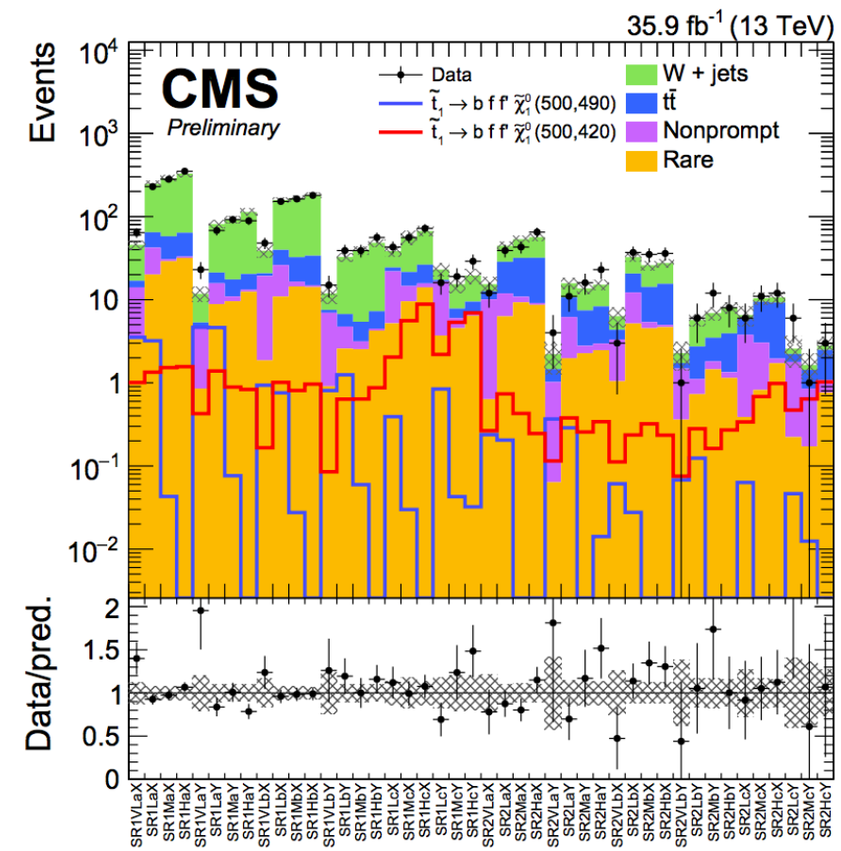
Smearing

Reconstructed



Physics Event





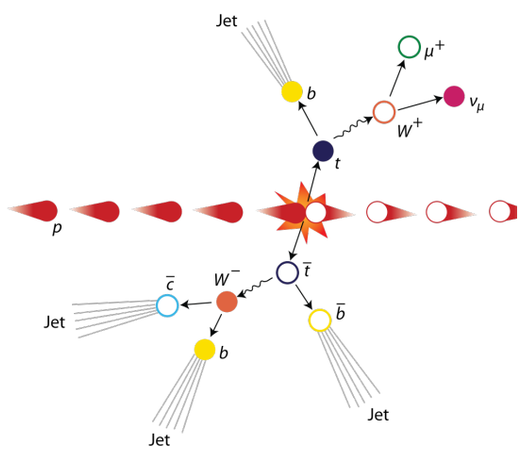
1-lepton compressed Stop CMS-SUS-17-005

Smeared MC + observed data

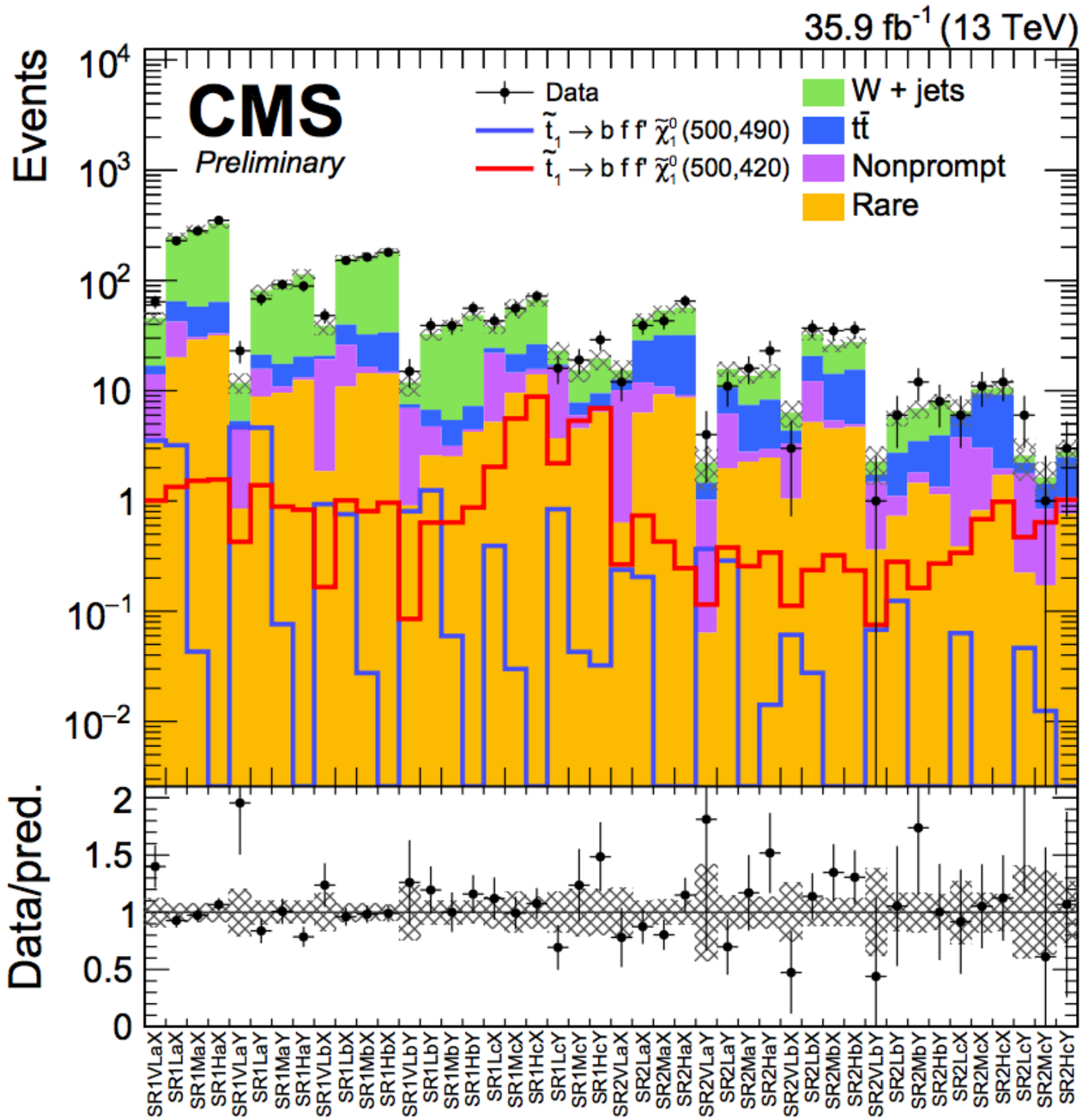
Searches for BSM

Truth/Gen

structed



Physics Eve



Stop CMS-SUS-17-005

d MC +
ed data

Experimental Likelihood

Experimental searches often employ the Likelihood for parameter estimation in some underlying physics model (also for discovery and placing limits)

$$L(\text{parameters}) = \Pr(\text{data}|\mathcal{M}, \text{parameters})$$

Searches often consider multiple bins/regions simultaneously

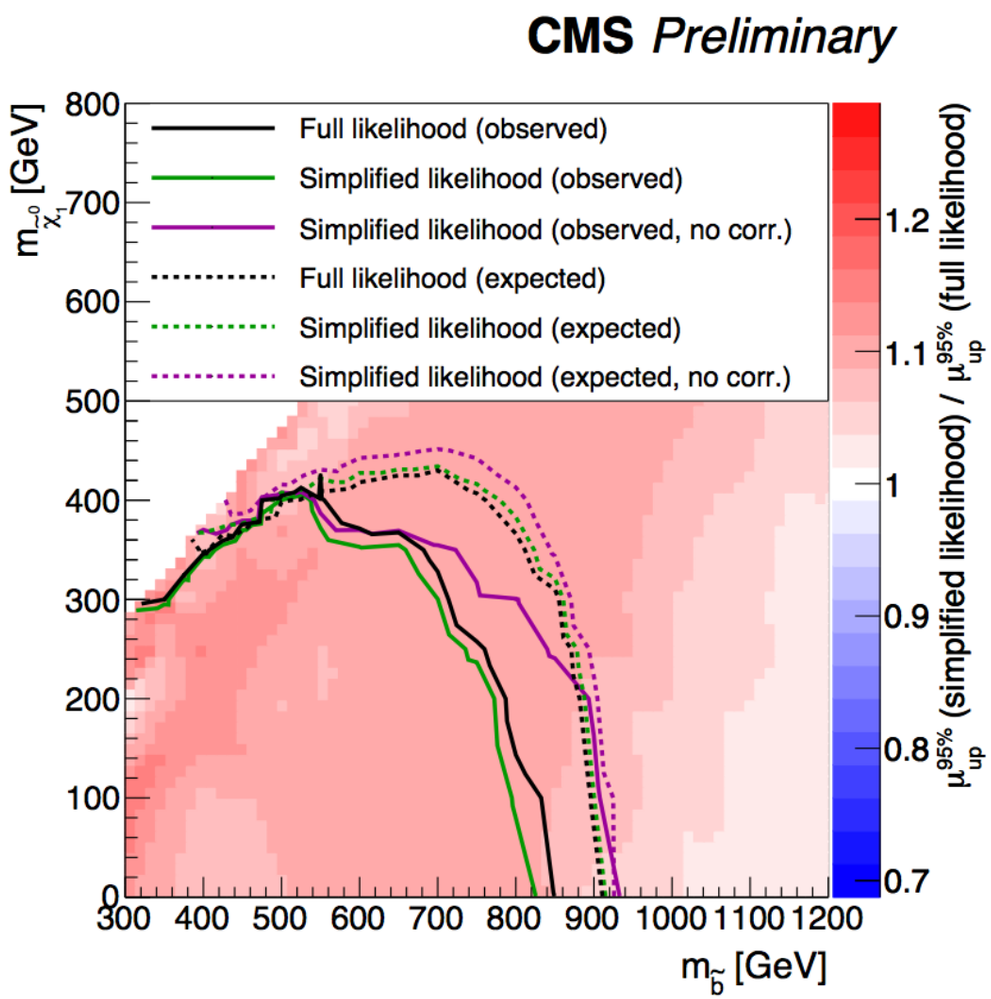
We are interested usually in the parameters of interest (POIs) = $\boldsymbol{\alpha}$

Additional parameters ($\boldsymbol{\delta}$) account for systematic uncertainties in the model (nuisance parameters) - these are often constrained in some external measurement/by theory/additional data - $\pi(\delta)$

$$L(\boldsymbol{\alpha}, \boldsymbol{\delta})\pi(\boldsymbol{\delta}) = \prod_{I=1}^{P} \Pr\left(\hat{n}_I \,\middle|\, n_I(\boldsymbol{\alpha}, \boldsymbol{\delta})\right)\pi(\boldsymbol{\delta})$$

Simplified Likelihood

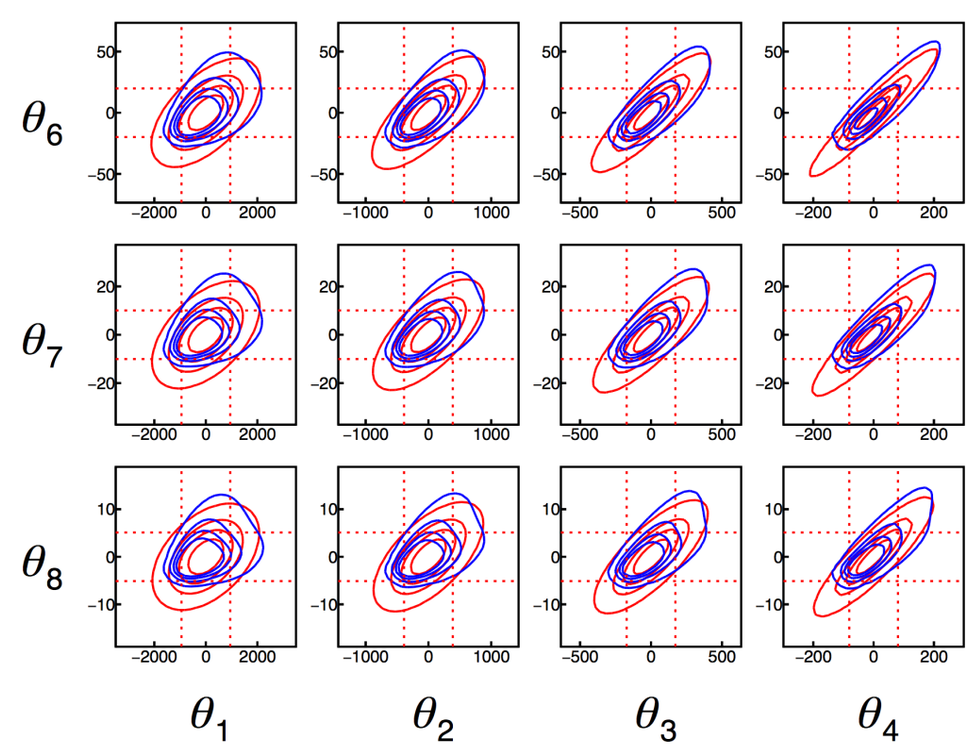
CMS proposal for providing covariance matrices* allows re-interpretation of searches with new signal models: CMS-NOTE-2017-001



Re-interpretation of SUSY search showed procedure performs well

Example studies highlighted some cases where covariance is insufficient

-> i.e where Gaussian approximation is poor



*Alternative approach using re-parameterisation templates in Phys.Rev.D91 (2015) 054032

*non-Gaussian approximations of LH scans: S. Gadatsch

Next-to Simplified Likelihood (NSL)

Today, I will present some recent work* intended to go beyond the Gaussian approximation (building on Nucl. Phys. B911 (2016))

Simplify experimental likelihood, but including quadratic term

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

$$\approx \prod_{I=1}^{P} \Pr\Big(\hat{n}_I \,\Big|\, a_I(\boldsymbol{\alpha}) + b_I(\boldsymbol{\alpha})\theta_I + c_I(\boldsymbol{\alpha})\theta_I^2\Big) \frac{e^{-\frac{1}{2}\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{\rho}^{-1}(\boldsymbol{\alpha})\boldsymbol{\theta}}}{\sqrt{(2\pi)^P}}$$

The NSL is ***fully*** defined given

- $\hat{n}$ - the measured (observed) results
- a, b, c - as function of the POIs
- $\boldsymbol{\rho} = \rho_{IJ}$ - correlations

* from the *LHC Chapter II: The Run for New Physics* workshop held at IIP Natal

Next-to Simplified Likelihood (NSL)

For the approximation we appeal to a modified CLT, keeping the “skew” term

Let Z be a random variable,

$$Z = N^{-1/2} \sum_{j=1}^{N} \delta_j$$

With δj iid variables (as in nuisance parameters) with variance σ, and skew γ

We find that Z can be approximated in the large N limit by

$$Z = \theta + \frac{\gamma}{3\sqrt{N}}\theta^2 , \quad \text{with} \quad \theta \sim \mathcal{N}(0, \sigma^2)$$

I.e as in the normal CLT but keeping the quadratic term

Note// Ff δj are *not* identically distributed, we can use the approximation by appealing to the Lyapunov CLT and defining $\sigma^2 = N^{-1} \sum_{j=1}^{N} \sigma_j^2$, $\gamma = N^{-1} \sum_{j=1}^{N} \gamma_j$

Next-to Simplified Likelihood (NSL)

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

$$\approx \prod_{I=1}^{P} \Pr\Big(\hat{n}_I \,\Big|\, a_I(\boldsymbol{\alpha}) + b_I(\boldsymbol{\alpha})\theta_I + c_I(\boldsymbol{\alpha})\theta_I^2\Big)\frac{e^{-\frac{1}{2}\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{\rho}^{-1}(\boldsymbol{\alpha})\boldsymbol{\theta}}}{\sqrt{(2\pi)^P}}$$

For ***searches*** we can typically have

- $\hat{n}_I$ are the observed number of events in each bin hence Pr == Poisson
- $n_I = n_{s,I} + n_{b,I}$
- Systematic uncertainties on signal are (at least one of)
    1. Sub-dominant
    2. ~Factorizable (uncorrelated with dominant bkg uncertainties)

  Hence we can assume $a_I(\alpha) \to a_I + n_{s,I}(\alpha),\ b_I(\alpha) \to b_I$ and $c_I(\alpha) \to c_I$

Note// signal systematics can be **added back in**, following similar treatment for backgrounds, if 1 and 2 are true (so its not a **hard** assumption)

Next-to Simplified Likelihood (NSL)

Finally our NSL for searches looks like…

Observation are just the number of events in each bin o_I $\Pr(o_I|n_I) \rightarrow \frac{(n_I)^{o_I} e^{-n_I}}{o_I!}$

Quadratic term allows to model asymmetric effects

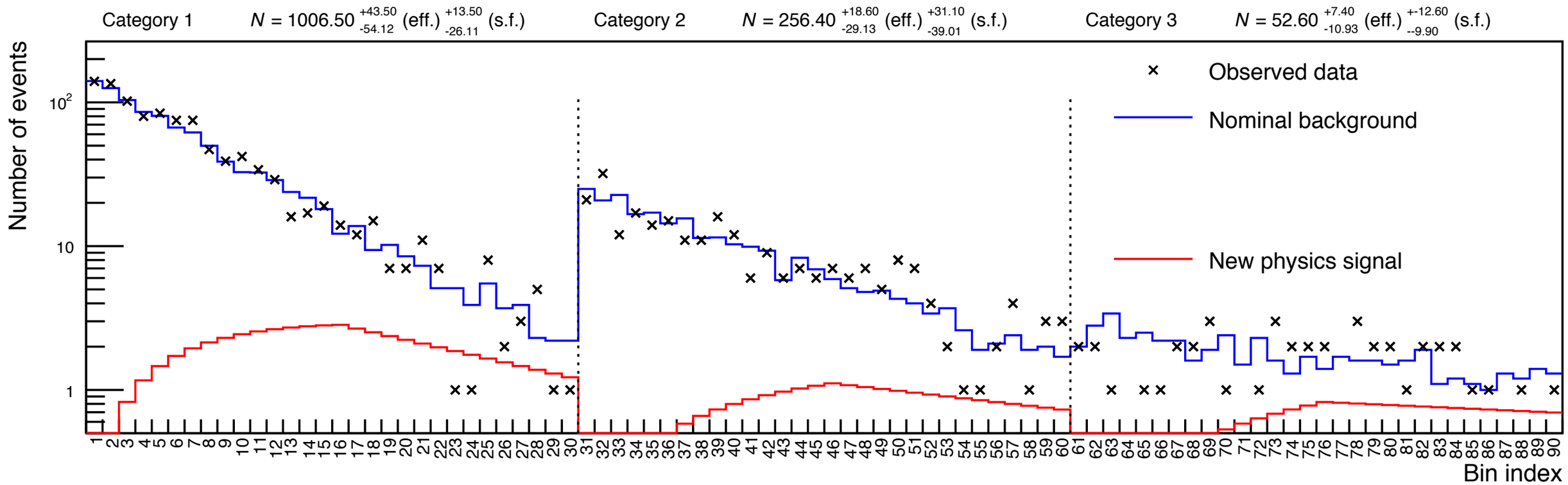
$$L(\boldsymbol{\alpha},\boldsymbol{\delta})\pi(\boldsymbol{\delta}) \approx L(\boldsymbol{\alpha},\boldsymbol{\theta})\pi(\boldsymbol{\theta}) = \prod_{I=1}^{P} \Pr\left(o_I \,\middle|\, n_{s,I}(\boldsymbol{\alpha}) + a_I + b_I\theta_I + c_I\theta_I^2\right) \frac{e^{-\frac{1}{2}\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{\rho}^{-1}\boldsymbol{\theta}}}{\sqrt{(2\pi)^P}}$$

Expected bin yield n_I split into

- signal (as function of POIs)
- background - parameterised as function of θ_I - one per bin

Multivariate normal constraint for θ_I

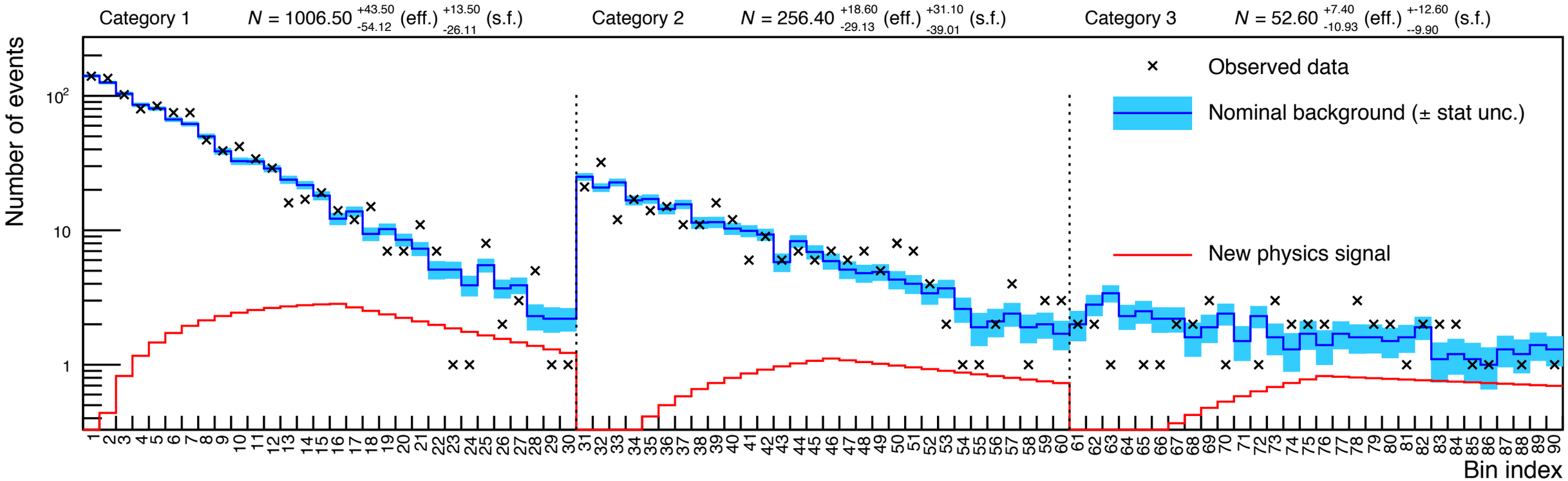
Toy Model



Setup a toy model inspired by a typical SUSY/Exotics search

- 3 categories
  - Each with 30 bins where observation (o_l) = number of events (Poisson counts)
  - Correlated "*normalisation*" uncertainties
- Increasing S/B with bin-number, within each category

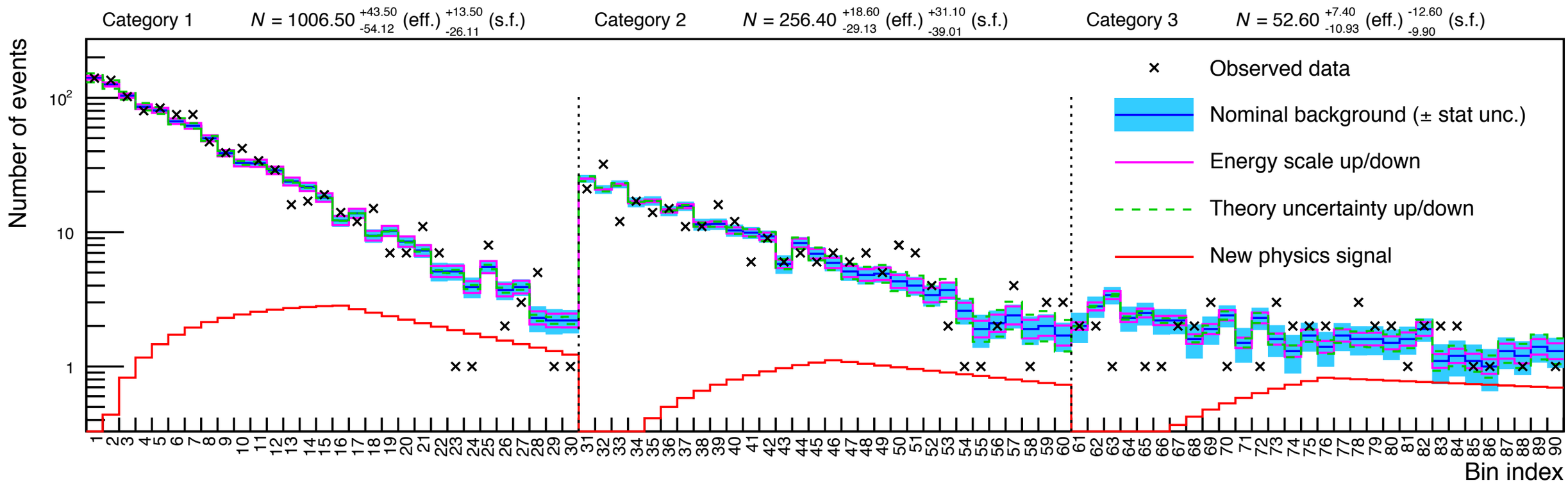
Toy Model



Setup a toy model inspired by a typical SUSY/Exotics search

- 3 categories
  - Each with 30 bins where observation (o_l) = number of events (Poisson counts)
  - Correlated "*normalisation*" uncertainties
- Increasing S/B with bin-number, within each category
- *Un-correlated* "bin-by-bin" uncertainties
  - Eg from limited MC statistics or data-driven control regions

Toy Model



Setup a toy model inspired by a typical SUSY/Exotics search

- 3 categories
  - Each with 30 bins where observation (o_l) = number of events (Poisson counts)
  - Correlated "*normalisation*" uncertainties
- Increasing S/B with bin-number, within each category
- *Un-correlated* "bin-by-bin" uncertainties
  - Eg from limited MC statistics or data-driven control regions
- *Correlated* "shape uncertainties"
  - Eg varied spectra from energy scales/ISR or PS models

Experimental Likelihood

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

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

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

Similarly for un-correlated bin-by-bin uncertainties

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

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

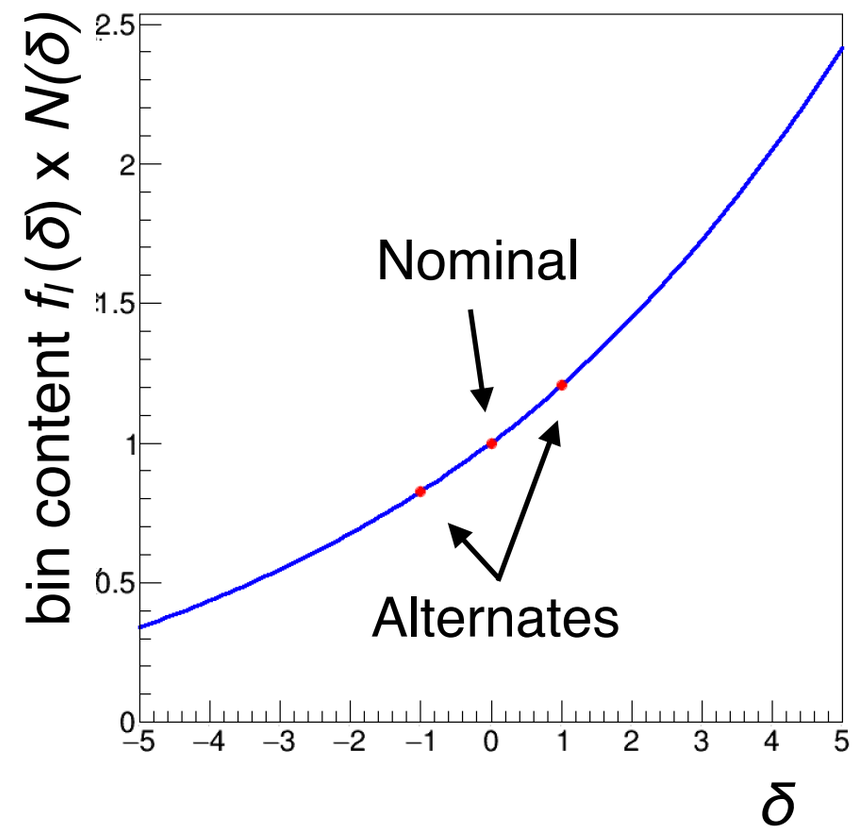
*easily generalises to per-process/per-category

Experimental Likelihood

"Shape" (correlated) systematics can be modelled as polynomial interpolations as a function of a single nuisance parameter (one independent parameter for each "source")*

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

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



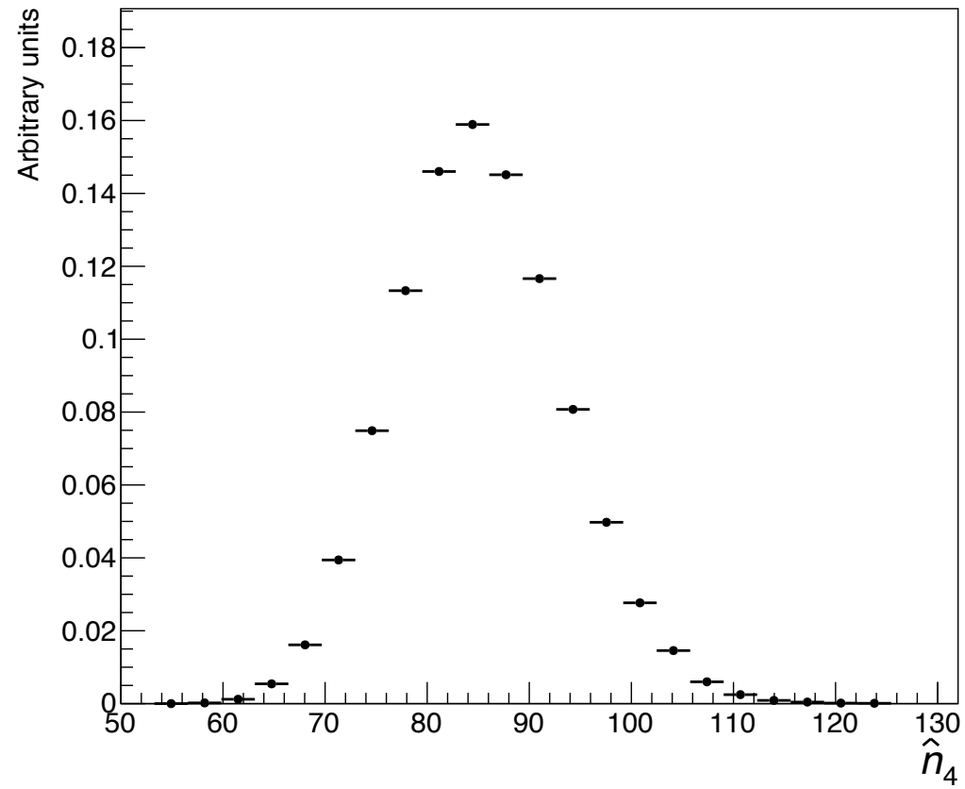
$$p_{Ij}(\delta_j) = \begin{cases} \frac{1}{2}\delta_j(\delta_j - 1)\kappa_{Ij}^- - (\delta_j - 1)(\delta_j + 1) + \frac{1}{2}\delta_j(\delta_j + 1)\kappa_{Ij}^+ & \text{for } |\delta_j| < 1 \\ \left[\frac{1}{2}(3\kappa_{Ij}^+ + \kappa_{Ij}^-) - 2\right]\delta_j - \frac{1}{2}(\kappa_{Ij}^+ + \kappa_{Ij}^-)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

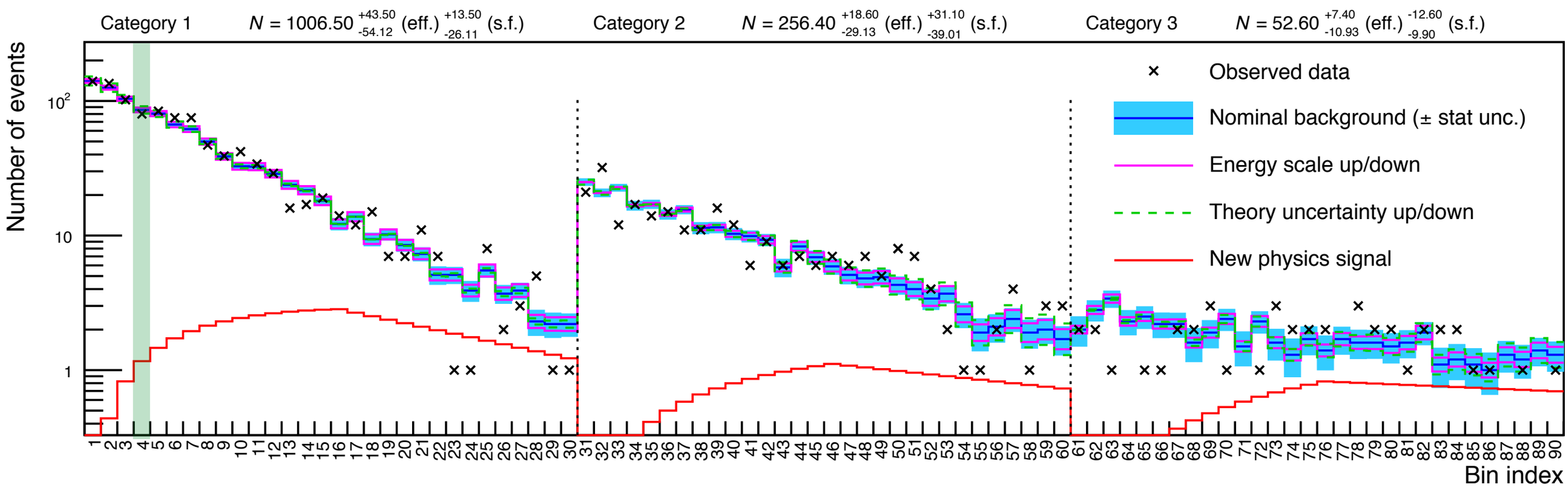
*see J. Conway arXiv.org/abs/1103.0354

Toy Model

Take the toy model and generate 100,000 pseudo-datasets for n_l

In some bins, distributions looks symmetric and Gaussian - can be described by 2 moments (*mean* and *variance*)



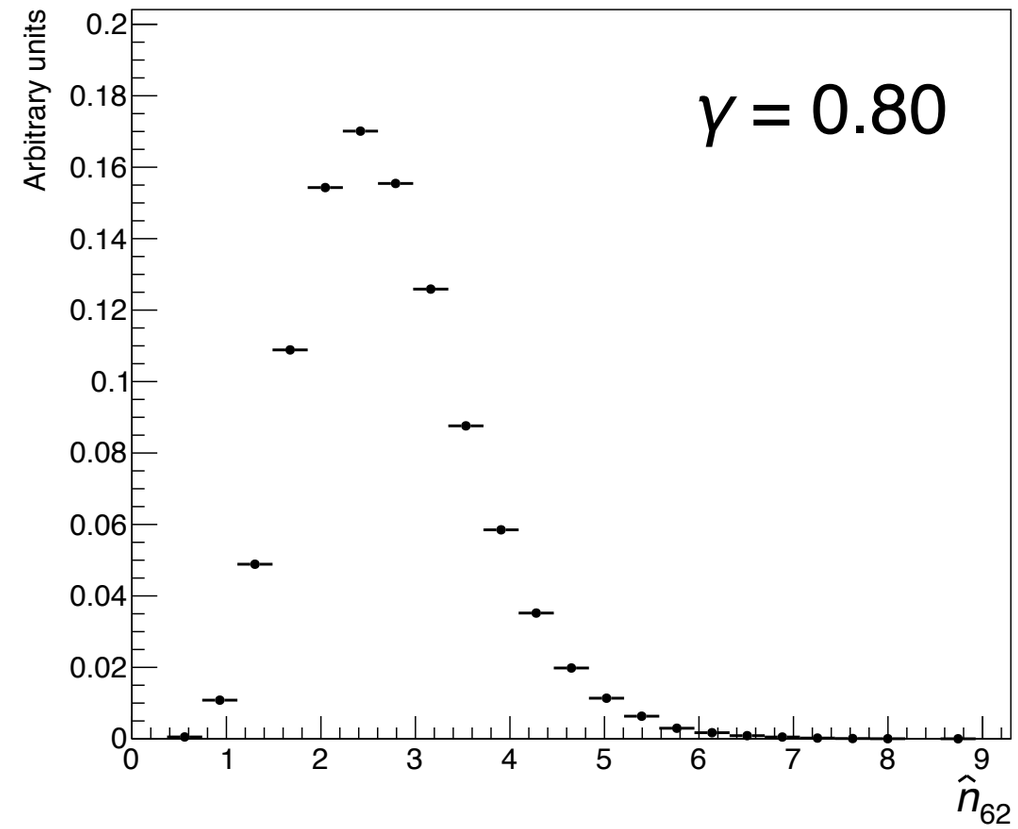


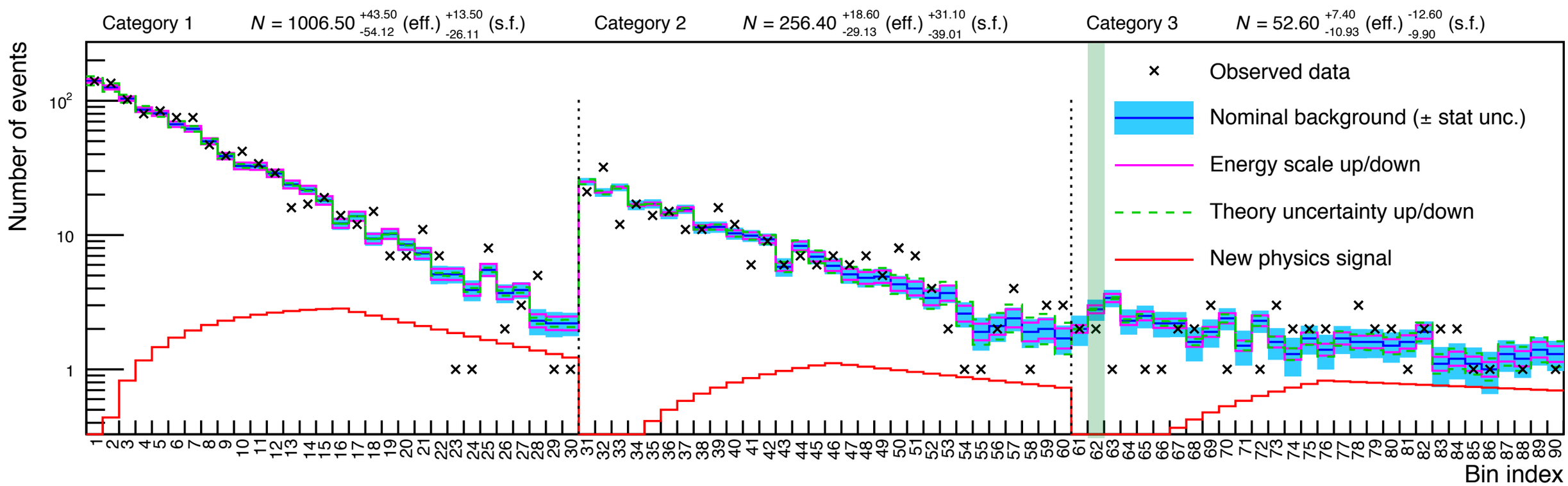
Toy Model

In other cases however, distributions are asymmetric

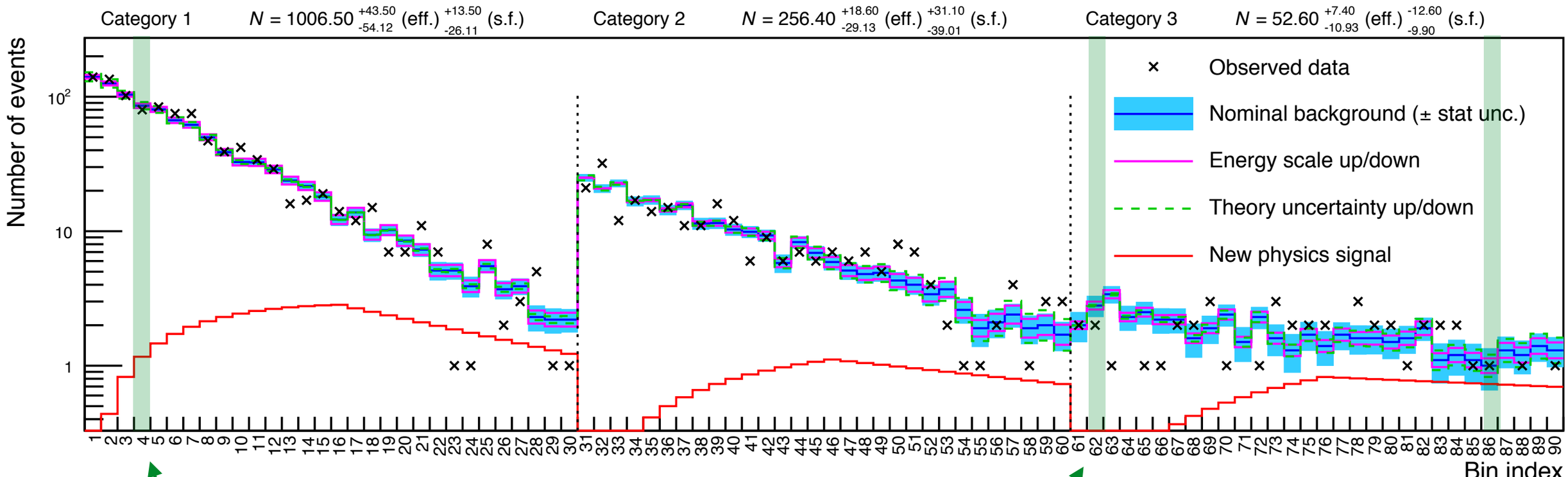
-> Skewness (γ) provides a measure of how important the 3rd moment is

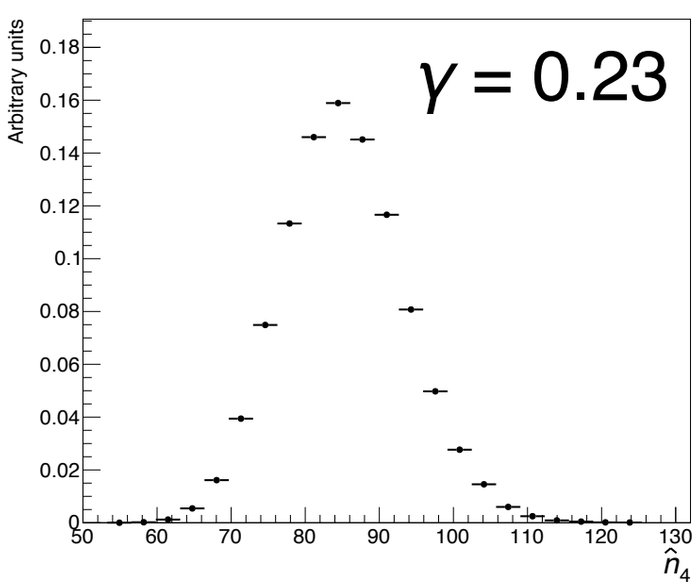
$$\gamma = \frac{m_3}{(m_2)^{\frac{3}{2}}}$$

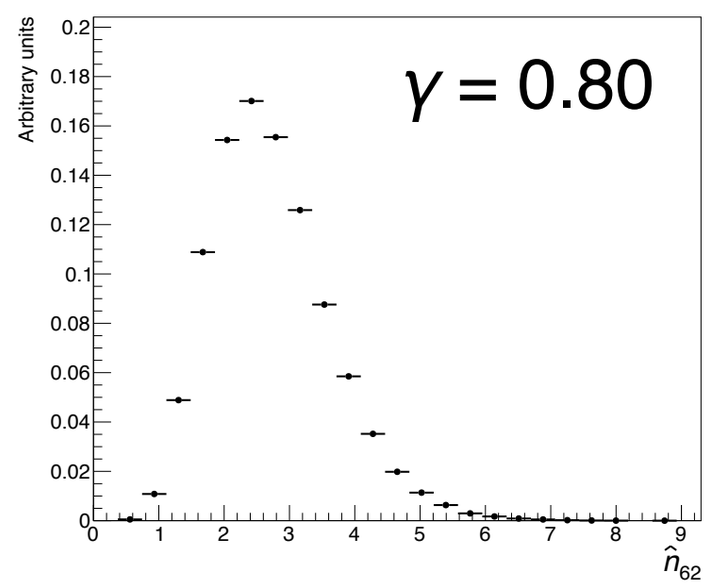


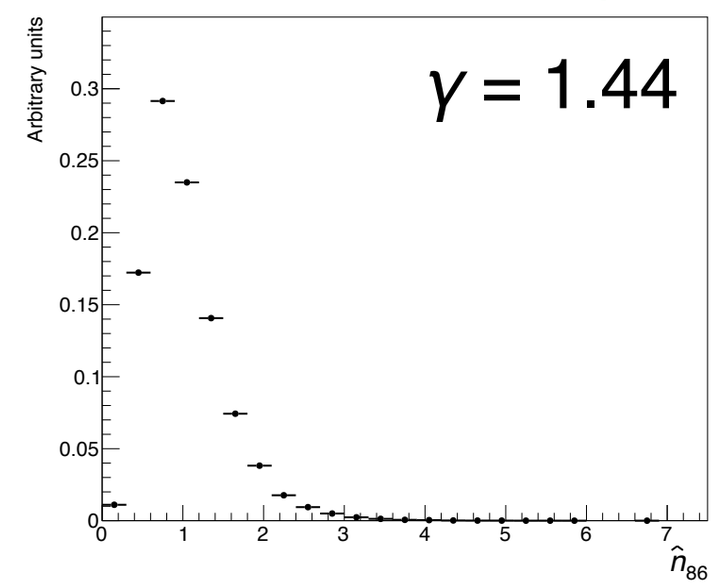


Skewness


Category 1
$N = 1006.50^{+43.50}_{-54.12}$ (eff.) $^{+13.50}_{-26.11}$ (s.f.)
Category 2
$N = 256.40^{+18.60}_{-29.13}$ (eff.) $^{+31.10}_{-39.01}$ (s.f.)
Category 3
$N = 52.60^{+7.40}_{-10.93}$ (eff.) $^{-12.60}_{-9.90}$ (s.f.)
Number of events
Observed data
Nominal background (± stat unc.)
Energy scale up/down
Theory uncertainty up/down
New physics signal
Bin index


$\gamma = 0.23$
Arbitrary units
$\hat{n}_4$


$\gamma = 0.80$
Arbitrary units
$\hat{n}_{62}$


$\gamma = 1.44$
Arbitrary units
$\hat{n}_{86}$

Matching moments

For the inputs to the NSL, we need to calculate the coefficients *a, b, c* and the matrix ***ρ***

If we let $(\theta_I, \theta_J) \sim \mathcal{N}(0, \rho_{IJ})$ and $n_I = a_I + b_I\theta_I + c_I\theta_I^2$

$$m_{1,I} = \mathbf{E}[\hat{n}_I] = a_I + c_I$$
$$m_{2,IJ} = \mathbf{E}[(\hat{n}_I - \mathbf{E}[\hat{n}_I])(\hat{n}_J - \mathbf{E}[\hat{n}_J])] = b_I b_J \rho_{IJ} + 2c_I c_J \rho_{IJ}^2$$
$$m_{3,I} = \mathbf{E}[(\hat{n}_I - \mathbf{E}[\hat{n}_I])^3] = 6b_I^2 c_I + 8c_I^3\,,$$

Meaning that the necessary components of the NSL can be calculated by matching to the 1st and 3rd moments and the covariance of the expected background distributions

Matching moments

We find …

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

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

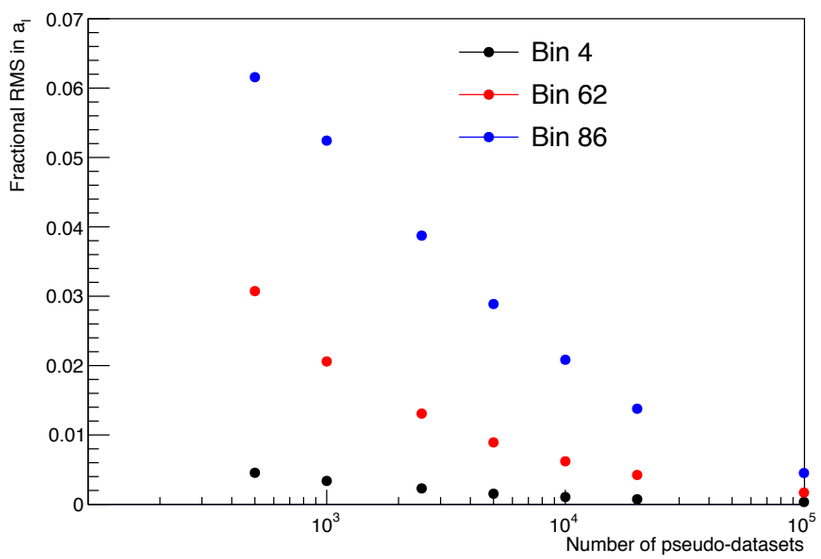
$$a_I = m_{1,I} - c_I$$

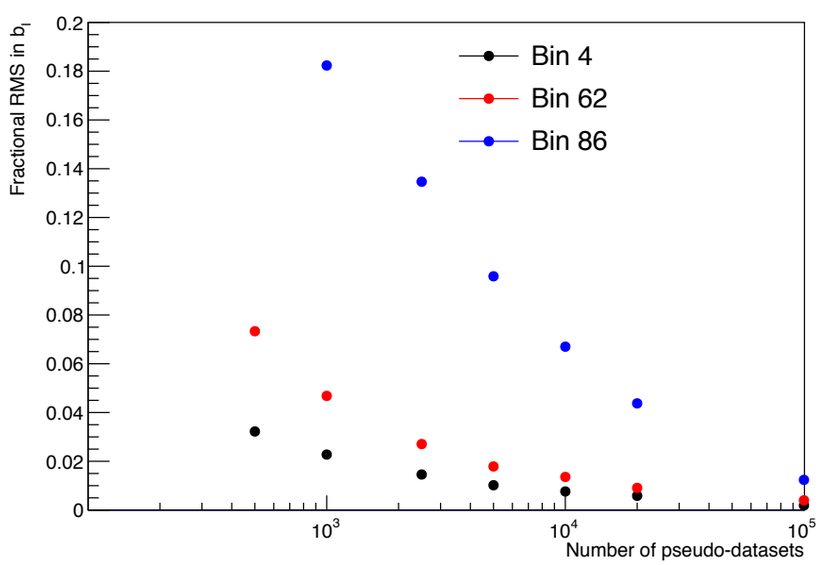
$$\rho_{IJ} = \frac{1}{4c_Ic_J}\left(\sqrt{(b_Ib_J)^2+8c_Ic_J\,m_{2,IJ}}-b_Ib_J\right).$$

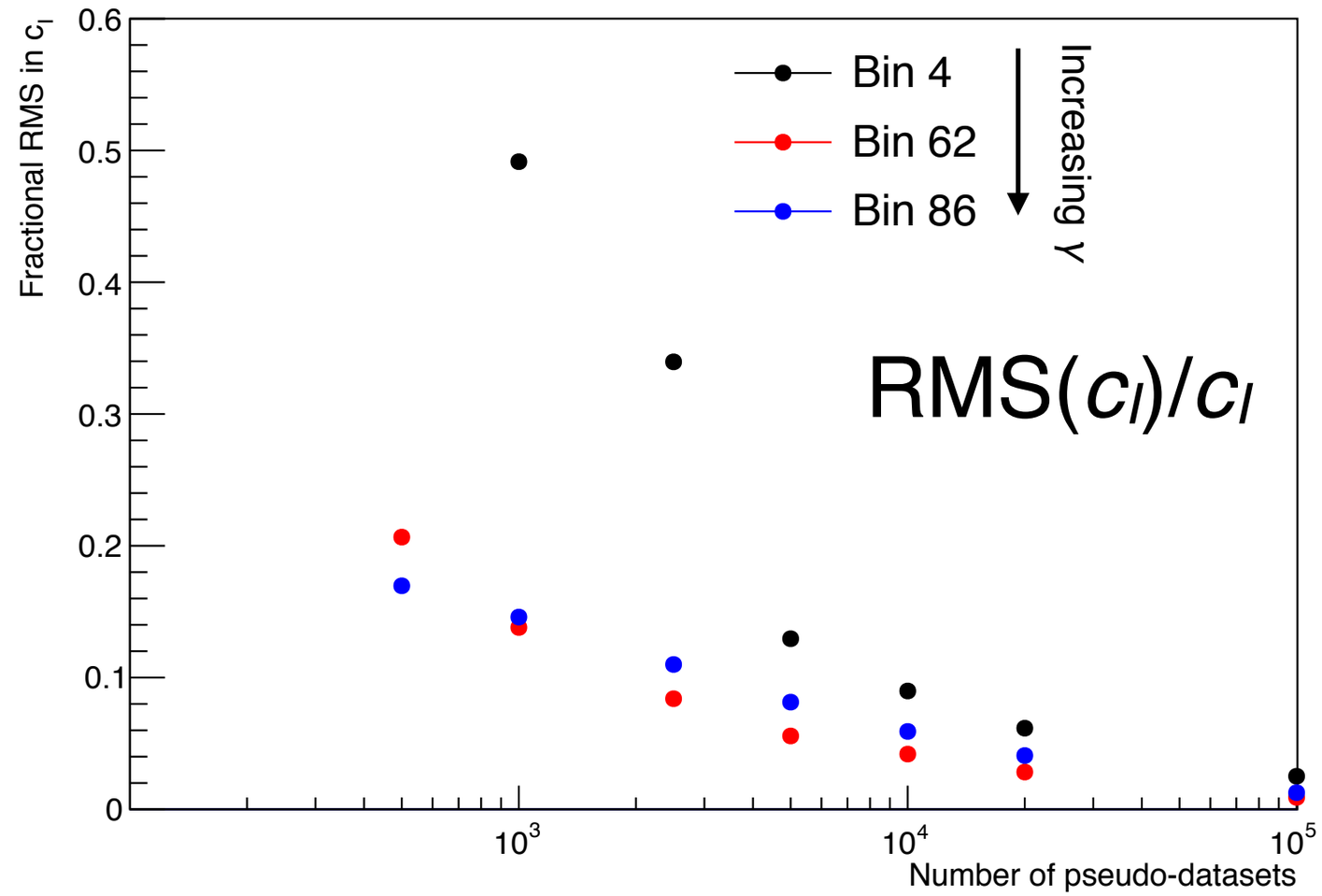
Provided $8m_{2,II}^3 \geq m_{3,I}^2$ or $\gamma \leq \sqrt{8} \approx 2.83$

Convergence

When calculating with toys, the rate of convergence can be slow for higher moments







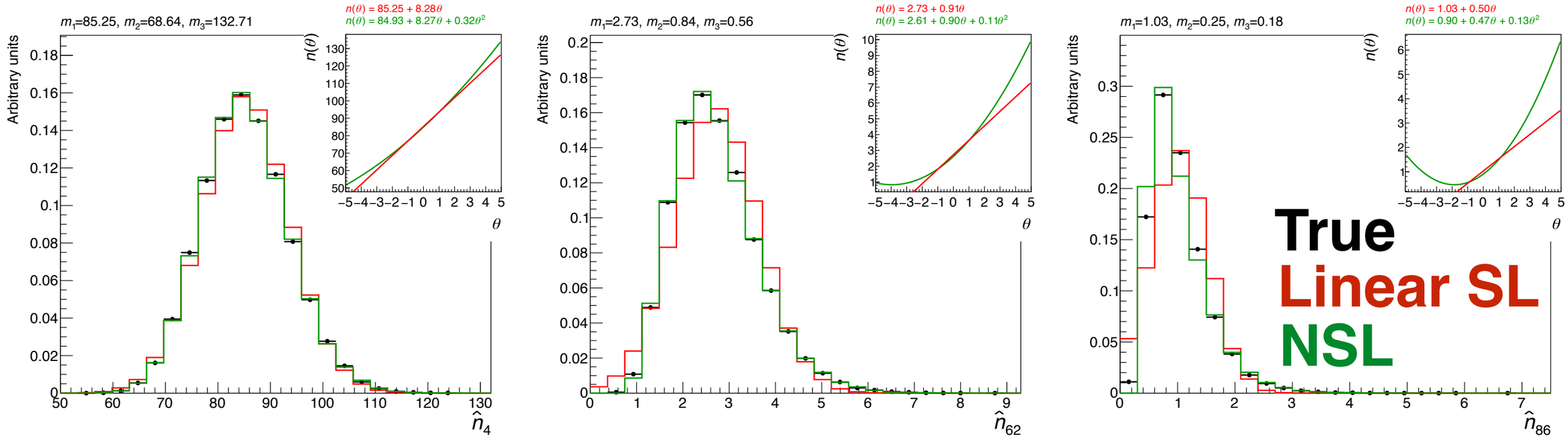
The coefficients can be calculated analytically instead if toy generation is slow (though other stability/speed issues can arise if experimental likelihood model is complex).

Validation studies

We applied the NSL to our toy model

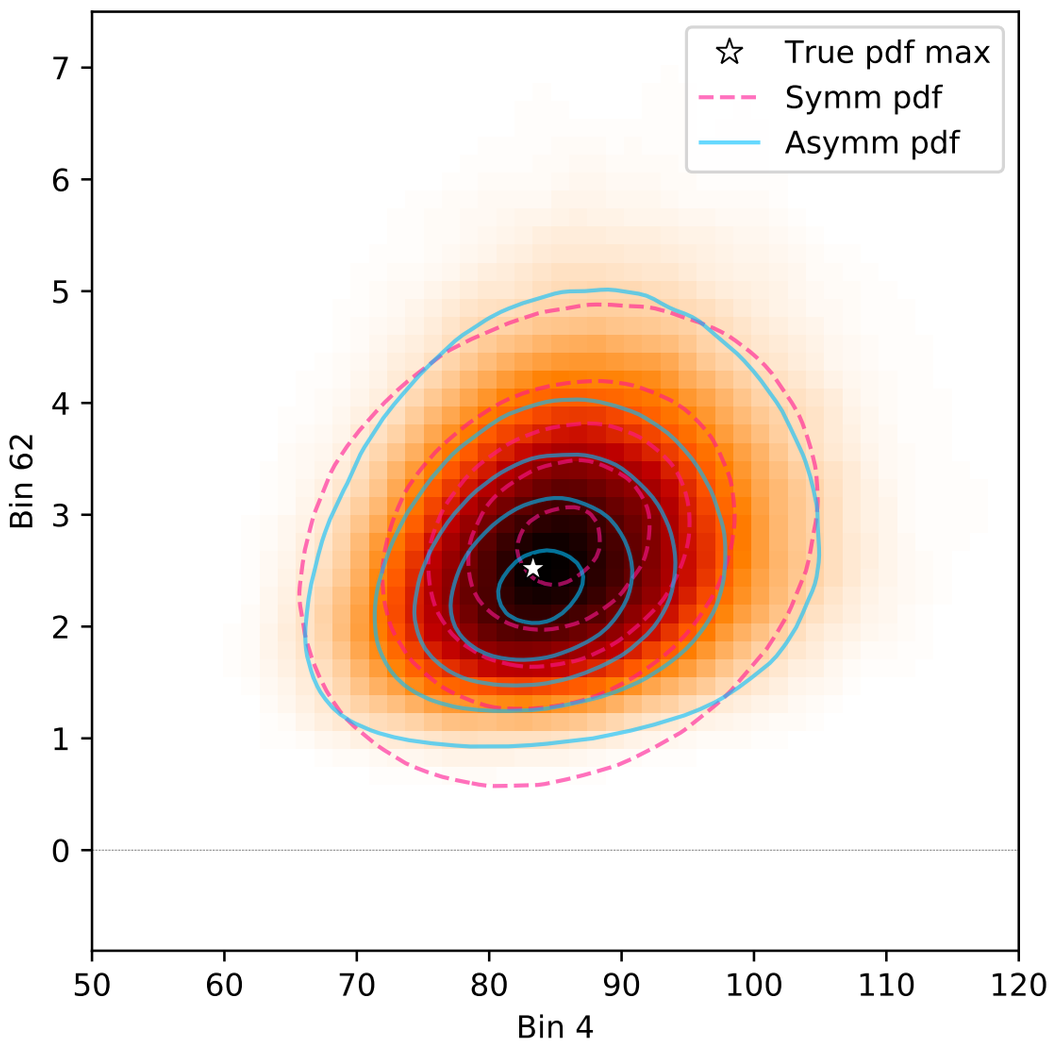
For bins with ~symmetric distributions, the NSL tends to the linear version (Gaussian approximation suffices)

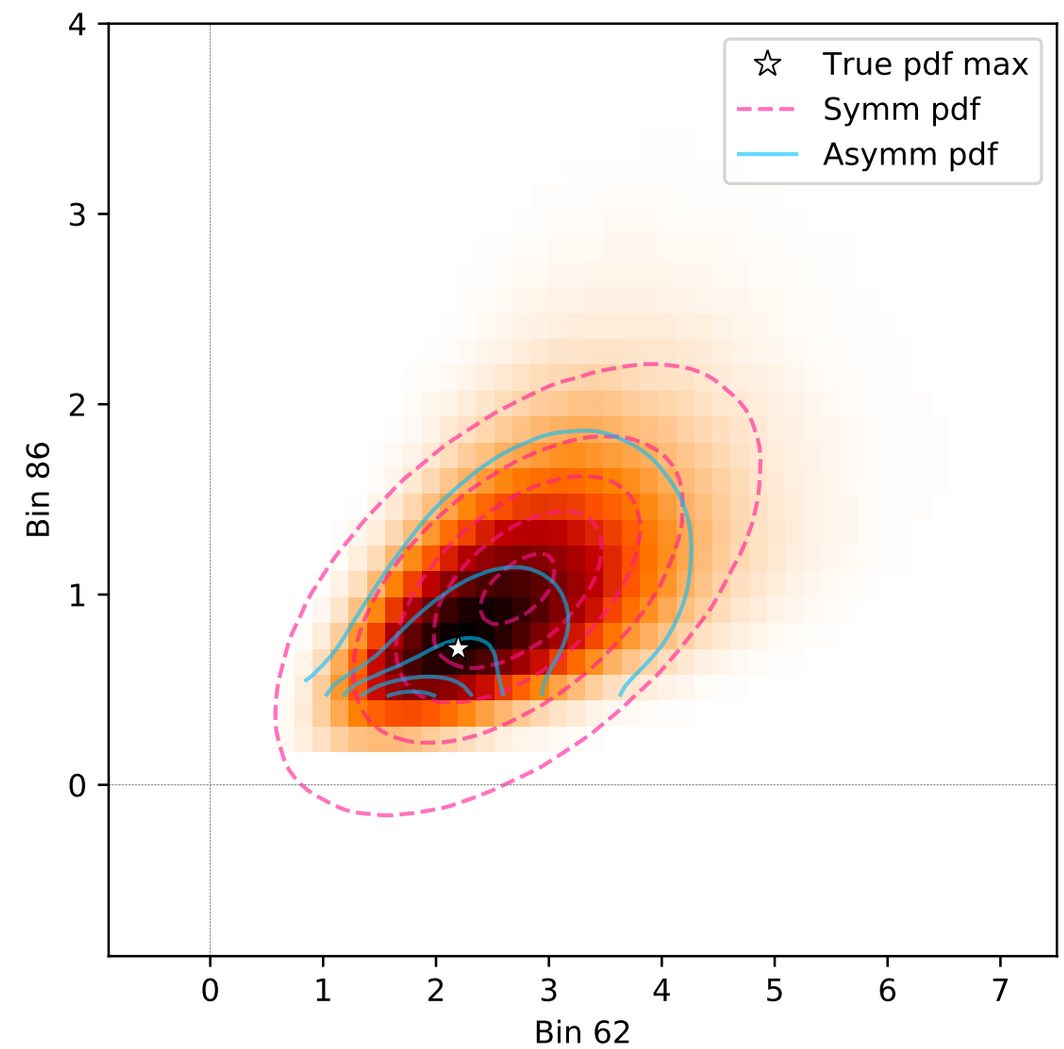
For bins with large skews, the NSL provides a better approximation to the true distribution



Validation studies

2D distributions show similar improvement when including quadratic term in the NSL expansion

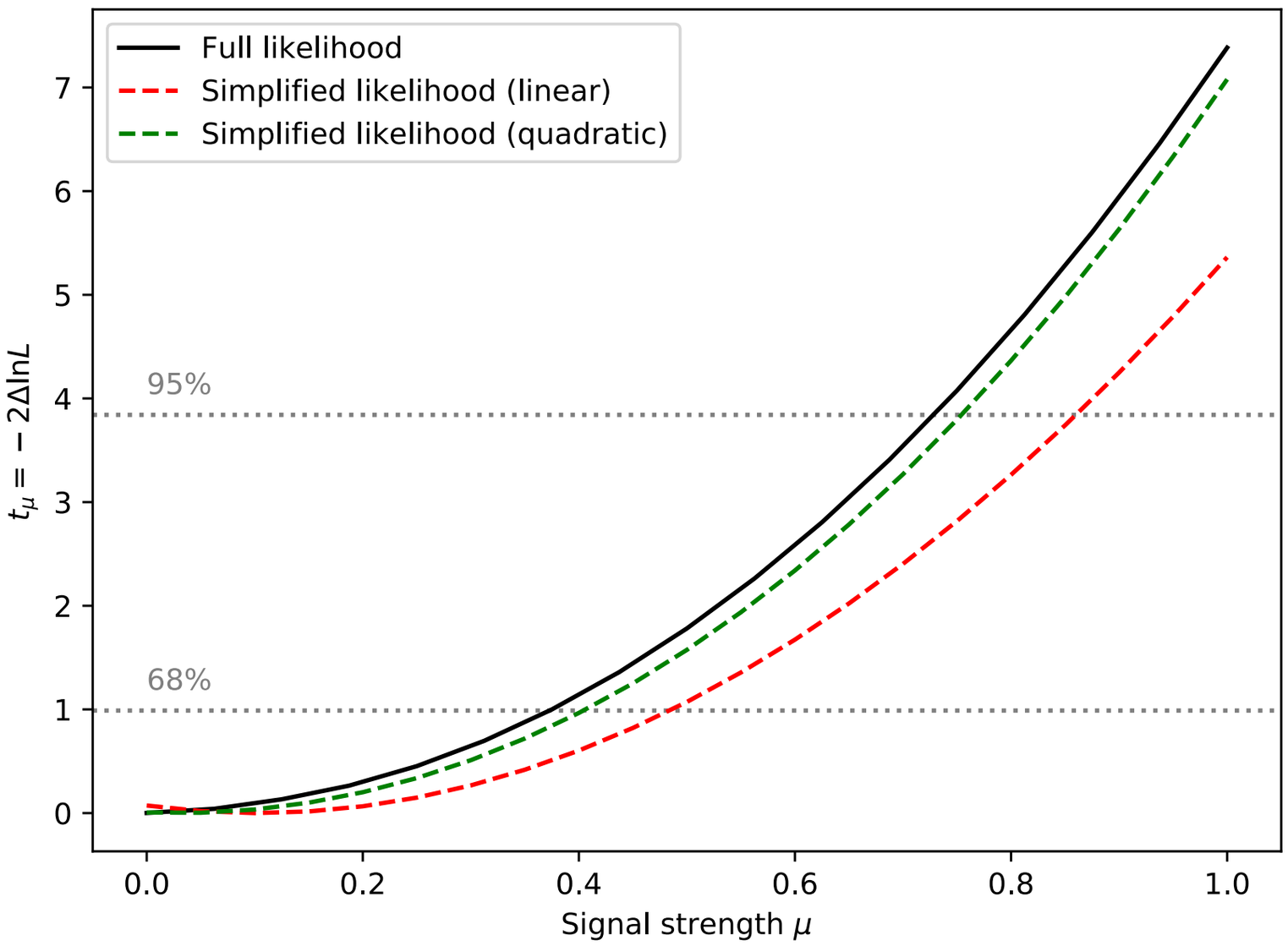




Validation studies

Profiled likelihood a common tool used in searches:

$$L(\mu, \theta) \rightarrow L(\mu, \hat{\theta}_\mu)$$



- Single POI = μ common signal scaling parameter.
- Likelihood maximised wrt nuisance parameters for each value of μ

$$t_\mu = -2\left[\ln L(\mu, \hat{\theta}_\mu) - \ln L(\hat{\mu}, \hat{\theta})\right]$$

- Log-likelihood differences used to set limits/find intervals on POIs

Distribution

https://www.hepdata.net/

HEPData

High Energy Physics Data Repository

Introduction (or extension) of simple HepData format to allow provision of

- Vectors (or histogram) of m_1, m_3
- 2D matrix for m_2
- observed data in each bin (as histogram/ vector)

Python based public code

- SciPy based optimisation routines widely used also outside of HEP
- Calculate coefficients of NSL given **moments**
- Include **data + signal model** as inputs to construct NSL
  - Performs marginalisation/ maximisation of nuisance parameters
  - Calculates Bayesian/Frequentist limits

python

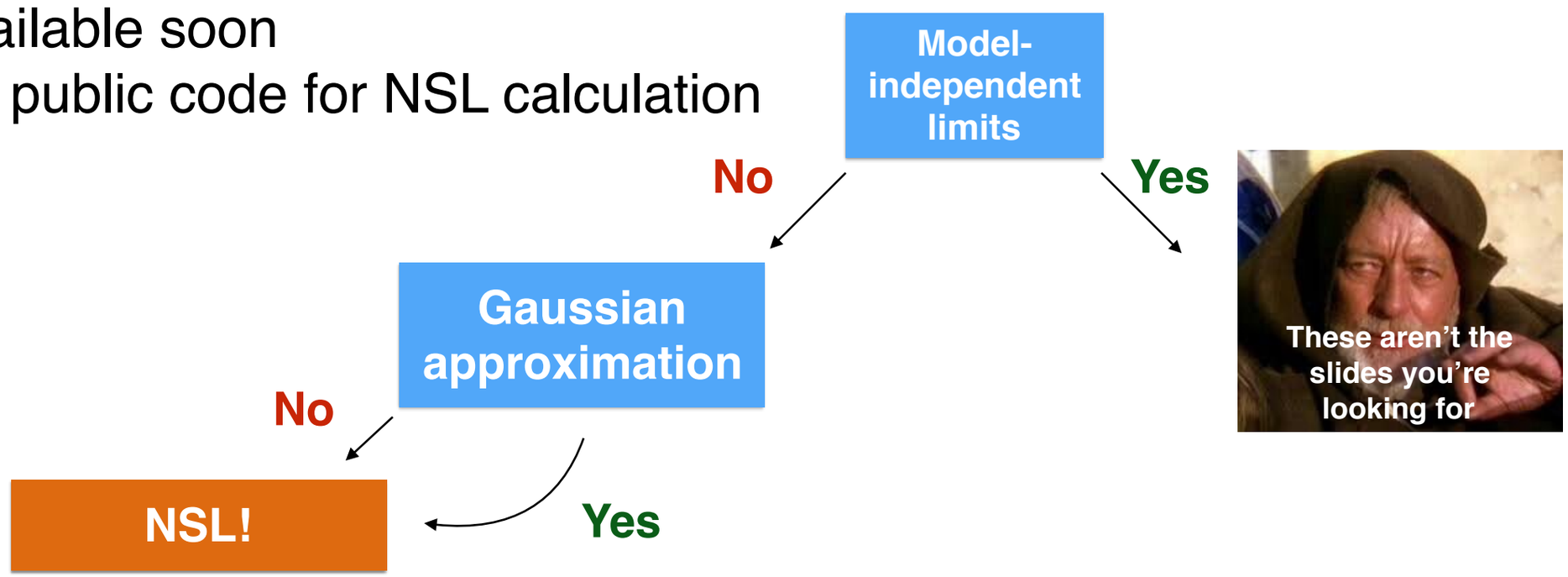
```
class SLParams(object):
    """Unified container for SL background moments, observed data, signal
    prediction, computed coefficients, and statistical functions"""

    def __init__(self, bg_m1, bg_m2, bg_m3=None, obs=None, sig=None, skew=True):
        import scipy.stats as st
        self.bg_m1 = np.array(bg_m1)
        self.bg_m2 = np.array(bg_m2).reshape([self.size,self.size])
        self.bg_m3 = np.array(bg_m3) if bg_m3 is not None else None
        self.obs = np.array(obs) if obs is not None else None
        self.sig = np.array(sig) if sig is not None else None
        #
        self.aparams, self.bparams, self.cparams = getCoeffsABC(self.bg_m1, self.bg_m2, self.bg_m3, skew=skew)
        self.rhoparams = getRho(self.bg_m1, self.bg_m2, self.bg_m3, skew=skew)
        self.thetadbn = st.multivariate_normal(np.zeros(self.size), self.rhoparams)
        #
        self.llmax = None #< for caching of unconditional max LL
        self.llnosig = None #< for caching of no-signal LL
```

Summary

Next-to Simplified Likelihoods allow re-interpretations of BSM searches

- Inputs are minimal (as in other approaches)
  - Allow for more robust re-interpretations (use experimental knowledge of systematic effects)
  - Go beyond Gaussian approximation with vector of 3rd moments to account for asymmetries
- Focused today on re-interpretations of BSM searches (with observed counts) but NSL can also be useful for other measurements (unfolded dists, signal yield fits …)
- Publication available soon
  - Will include public code for NSL calculation



Backup Slides

Experimental Likelihood

For each bin we have

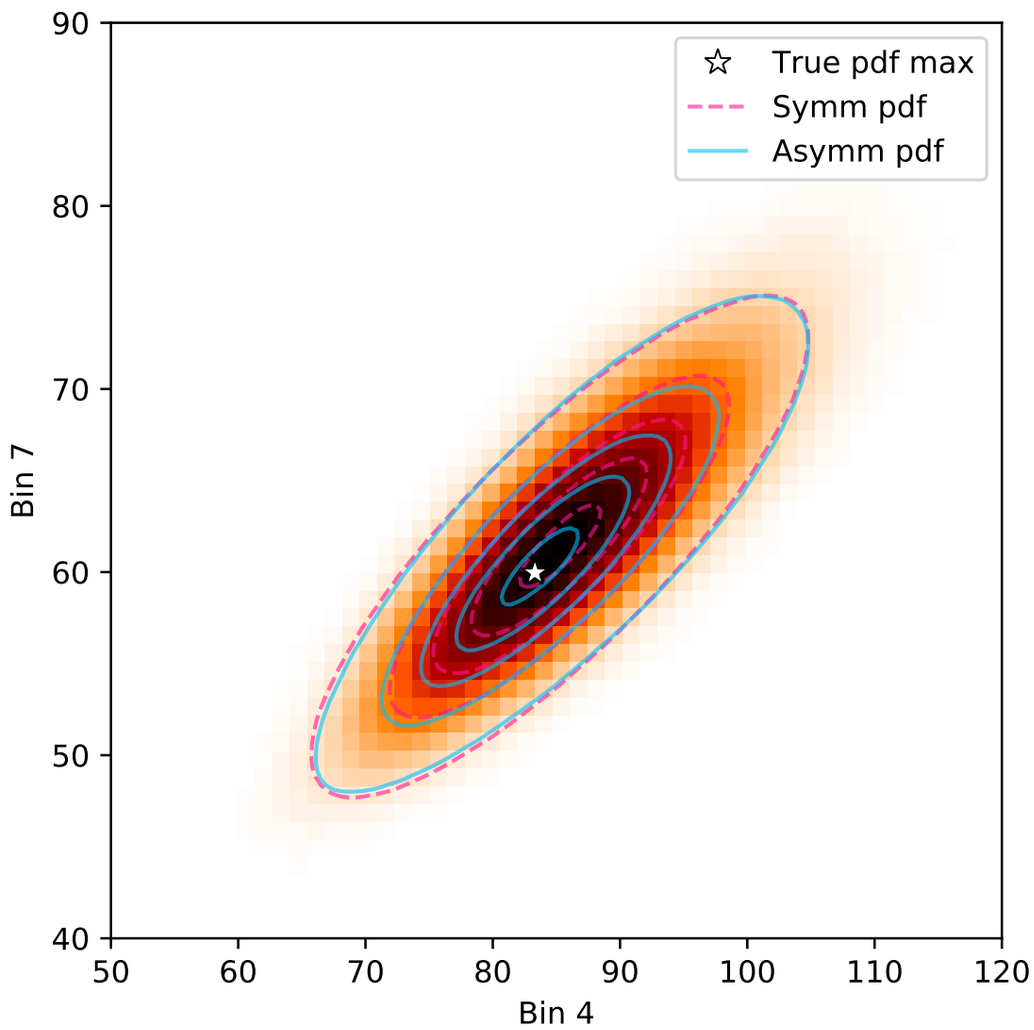
$$n_I(\boldsymbol{\delta}) = N^0 \cdot \prod_j (1+K_j)^{\delta_j} \cdot f_I^0 \cdot \frac{1}{F(\boldsymbol{\delta})} \prod_j p_{Ij}(\delta_j) \cdot \prod_j (1+\epsilon_{Ij}\delta_j)$$

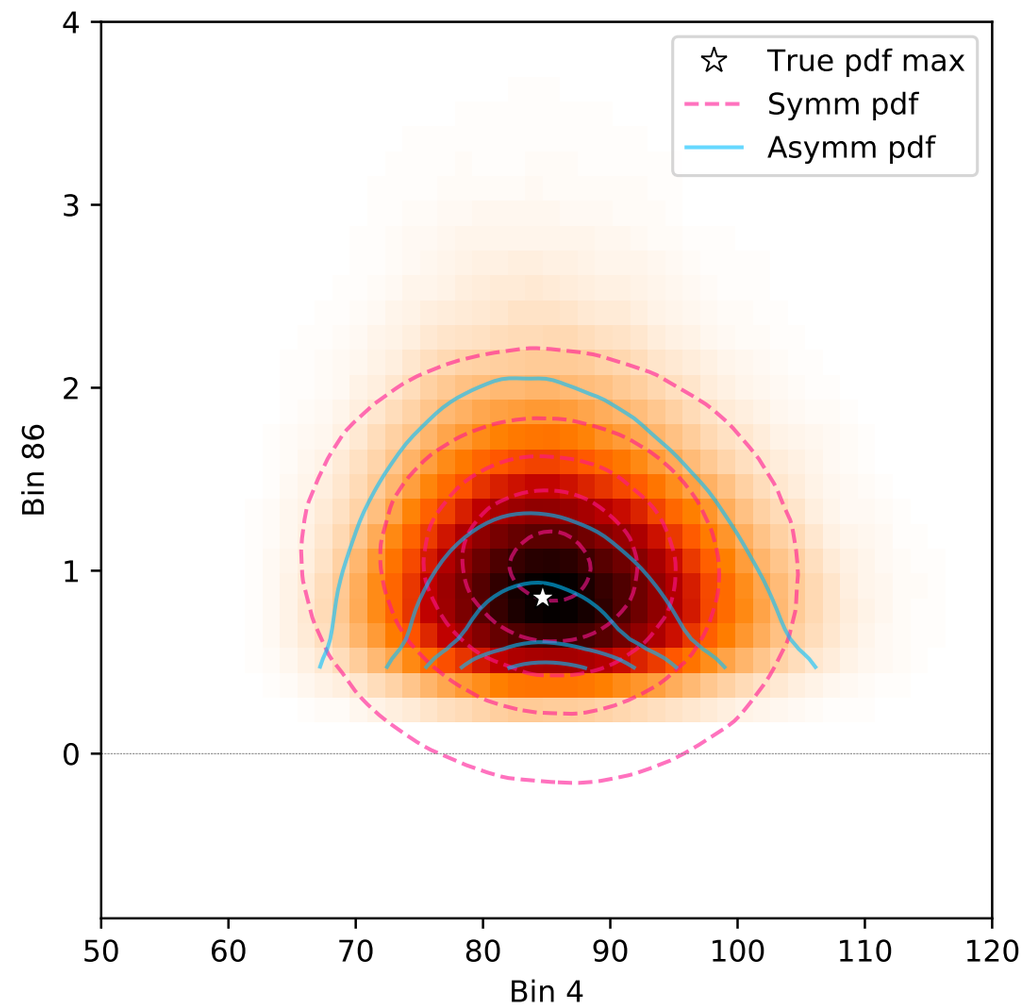
This means for a single process with P bins, we need

- P nominal values
- 1 value for each normalisation uncertainty
- P values for the uncorrelated uncertainties
- 2P values for each “shape” uncertainty

In this toy that means **452** values are needed.
In general, searches have >1 background process (x N backgrounds) and many more independent sources of uncertainty

More Correlations





Open issues

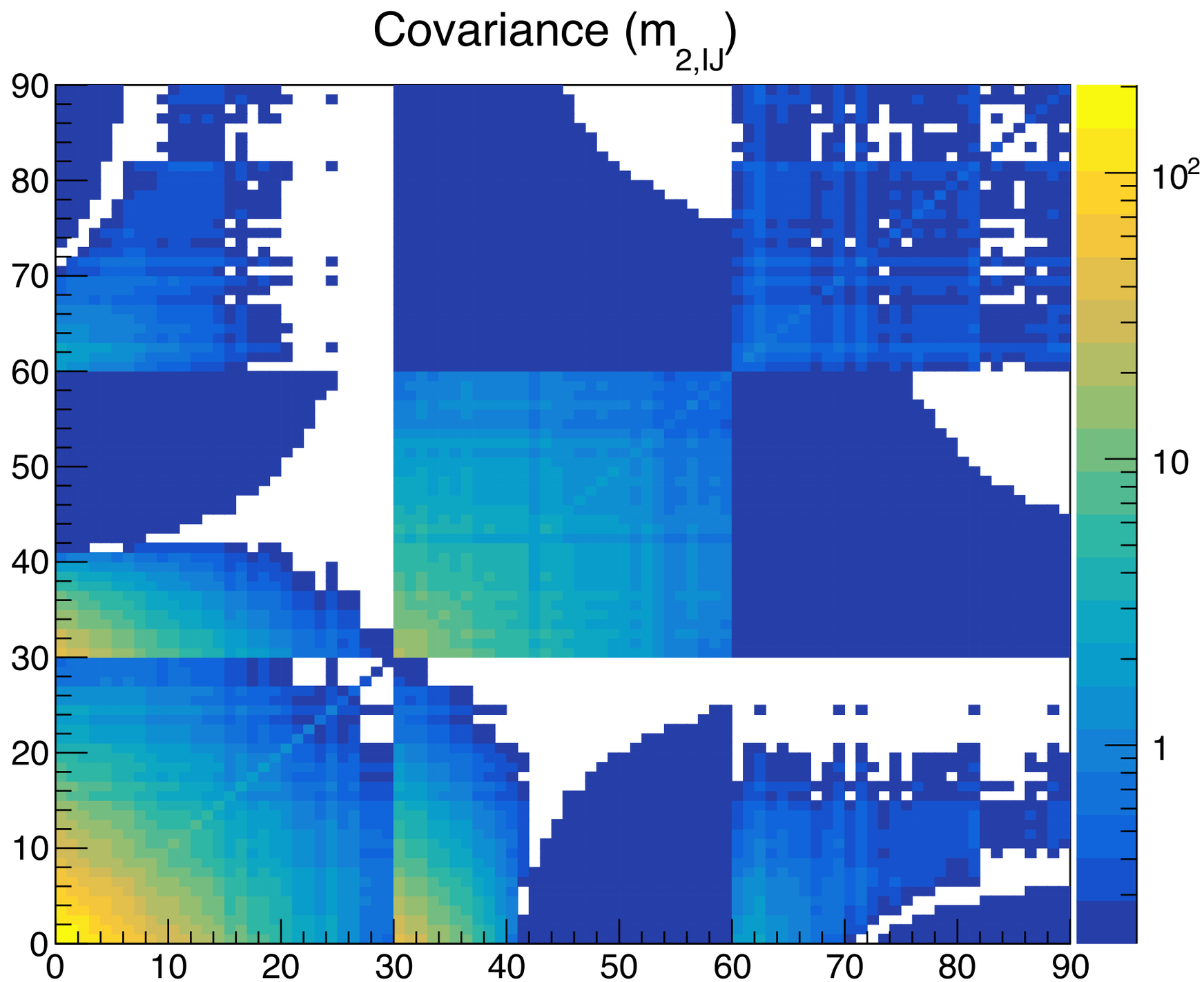
Beyond 3rd moment?

- How accurate is enough before distribution of shape is subdominant?
- Can releasing sampling toys lead to improved (non-analytic) constructions?

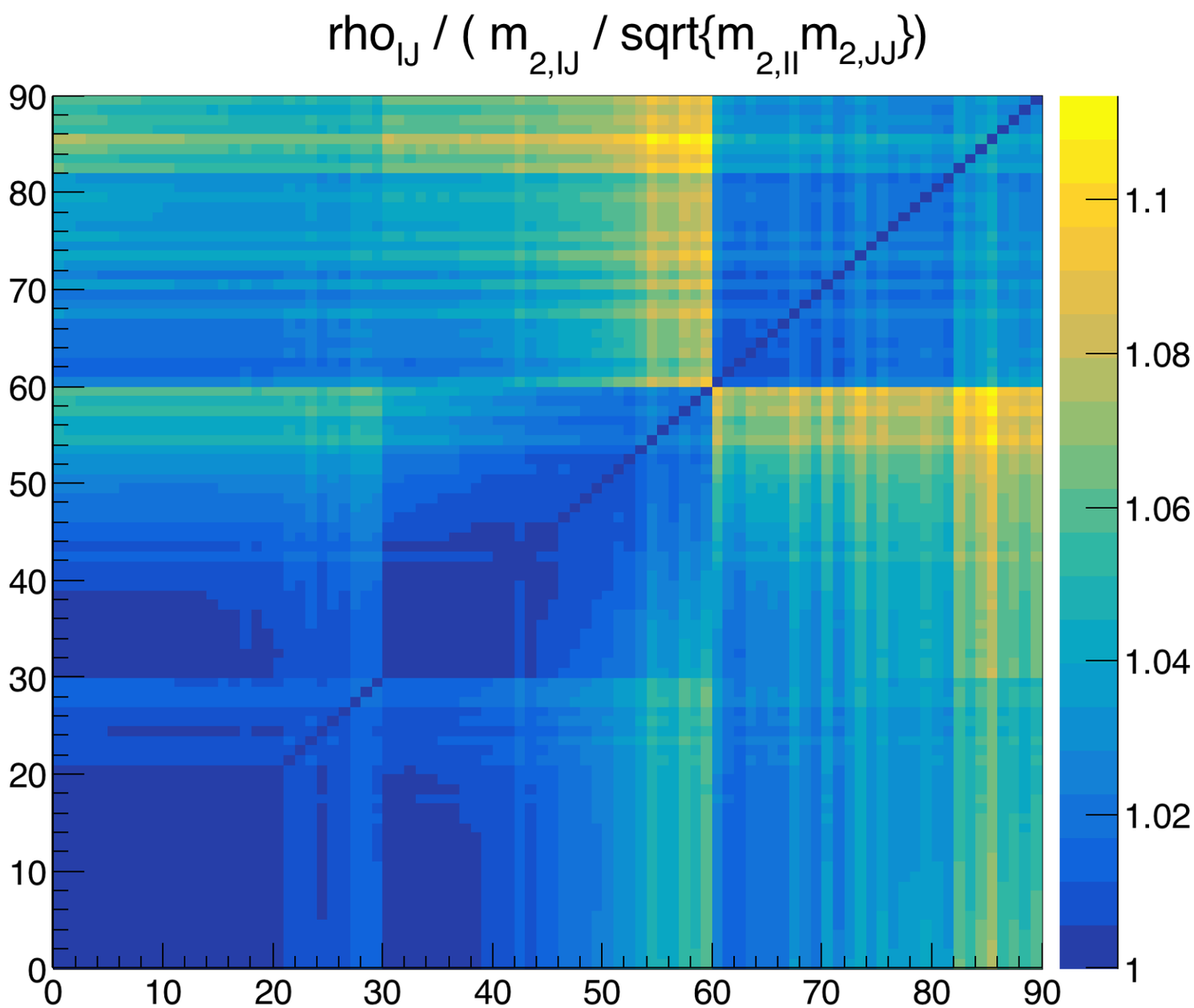
Systematic breakdowns

- SL/NSL allows for combination of covariances
  - Separate major components of uncertainty
  - Remove (and re-add) components correlated with signal systematics from the moment calculation

Toy Model Covariance



NSL correlation



NSL definition of correlation modified due to skew term

Ratio of ρ_{IJ} to linear correlation shows up to 15% correction in toy model