Physics 403

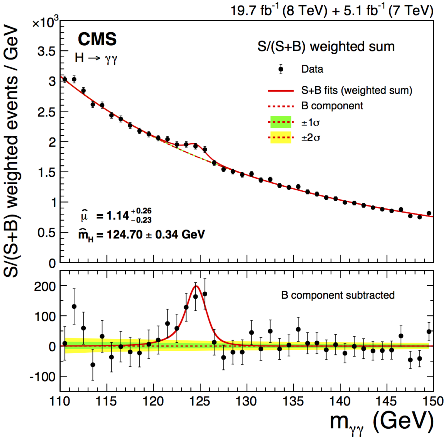
Unfolding

Segev BenZvi

Department of Physics and Astronomy
University of Rochester

Table of Contents

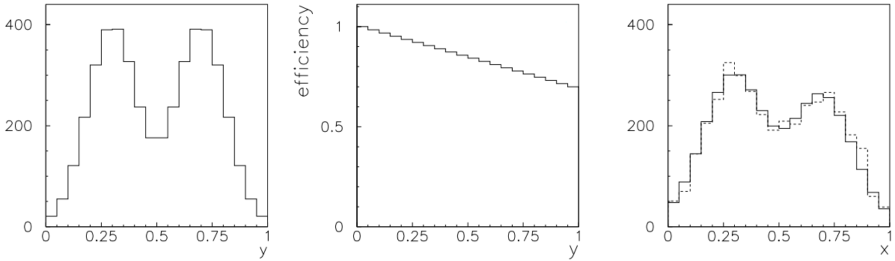
Motivation



- One of the most common tasks in physics is the publication of spectra, by which we mean a binned histogram showing the distribution of events in some observable quantity
- Example: diphoton mass spectrum $m_{\gamma\gamma}$ from CMS [1] showing the Higgs resonance
- **Problem**: spectra are often smeared and distorted by the finite resolution and thresholds of your apparatus
- Unfolding is used to fix these distortions

Example: A Bin-Dependent Instrumental Response

Suppose we have a "true" spectrum given by the histogram on the left [2]



The center plot shows efficiency as a function of bin. E.g., the instrumental efficiency decreases as a function of the parameter y

The right plot shows the measured spectrum (dashed) and expectation from Monte Carlo (solid) showing distortion in the counts

Unfolding in 2D

- Note that this problem also applies to 2D data such as binned images

- We can try to deconvolve smearing and blur from an image [3]
- What's required is some model of the smearing effect. By "smearing" we refer to an effect that results in an event being classified or reconstructed in the wrong bin

Forward Folding and Unfolding

Smeared data can be analyzed in a couple of ways:

1. **Forward Folding**: take a theoretical spectrum, smear it, and then compare the result to the data. The best fit gives you the true spectrum
2. **Unfolding**: take an observation which has smearing and other detector effects and try to deconvolve those effects

Forward folding is considerably easier than unfolding, so when possible it's best to do that. This is most appropriate when you just want to compare data from one experiment to a theoretical prediction

However, sometimes it is necessary to unfold your data. For example, if you want to compare spectra across several experiments, you need to remove the instrument response function to make unambiguous comparisons

Table of Contents

Unfolding Formalism

Define the following terms (following the notation of [4]):

- **Truth**: a "truth" spectrum $\boldsymbol{T} = (T_1, T_2, \ldots, T_{N_t})$ represents the binned counts that would be observed without smearing
- Truth spectra are usually estimated with Monte Carlo. Let the true counts be $\hat{\boldsymbol{T}}$ and the Monte Carlo truth be $\tilde{\boldsymbol{T}}$. Ideally, $\tilde{\boldsymbol{T}} = \hat{\boldsymbol{T}}$
- **Reco**: a "reco" spectrum $\boldsymbol{R} = (R_1, R_2, \ldots, R_{N_r})$ is the number of events expected to be reconstructed in a bin
- **Data**: a "data" spectrum $\boldsymbol{D} = (D_1, D_2, \ldots, D_{N_d})$ is the number of events observed in a bin after smearing. We expect $\boldsymbol{D}$ to follow a Poisson distribution with mean $\boldsymbol{R}$
- **Migration matrix**: a matrix $\mathcal{M}_{tr}$ defined by the joint PDF $p(t, r)$ of an event being produced in true bin t and reconstructed in bin r
- **Response matrix**: a matrix $\mathcal{P}_{tr}$ defined by the conditional probability $p(r|t)$

The Unfolding Problem

- Fundamentally, the unfolding problem requires us to calculate

$$p(\boldsymbol{T}|\boldsymbol{D},\mathcal{M}) = \frac{1}{Z}\mathcal{L}(\boldsymbol{D}|\boldsymbol{T},\mathcal{M})\,\pi(\boldsymbol{T},\mathcal{M})$$

- If the data follow a Poisson distribution, then

$$\mathcal{L}(\boldsymbol{D}|\boldsymbol{T},\mathcal{M}) = \prod_{r=1}^{N_r}\frac{R_r^{D_r}}{D_r!}e^{-R_r}$$

where the reconstructed counts are related to the true counts by

$$R_r = \sum_{t=1}^{N_t} T_t \cdot p(r|t)$$

- Recalling that $p(r|t)$ is the probability of reconstructing an event in bin r given that is should have been in bin t before smearing, we have

$$p(r|t) = \frac{p(t,r)}{p(t)} = \frac{\mathcal{M}_{tr}}{\epsilon_t^{-1}\sum_{k=1}^{N_r}\mathcal{M}_{tk}}, \qquad \epsilon_t = \frac{\sum_{r=1}^{N_r}\mathcal{M}_{tr}}{p(t)}$$

Accounting for the Presence of Background

- It is typical for the data to be contaminated by sources of background
- Example: in gamma-ray experiments, the cosmic-ray background reduction efficiency is typically 99.9%, but since the cosmic-ray flux is 1000$\times$ larger than the gamma-ray flux, the resulting signal/noise ratio is 1:1
- We can account for background by adding it to the expectation of the reconstructed counts:

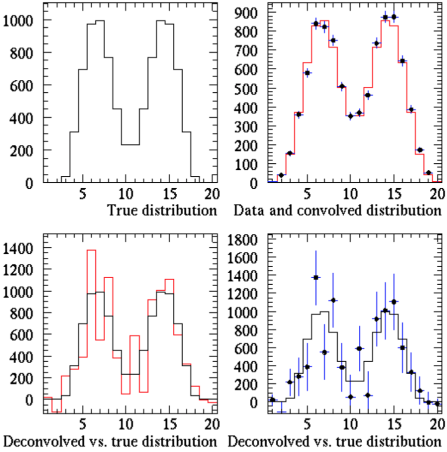
$$R_r = B_r + \sum_{t=1}^{N_t} T_t \cdot p(r|t)$$

where B_r is the number of background events in bin r. In matrix notation this is written

$$\boldsymbol{R} = \boldsymbol{B} + \mathcal{P}^\top \boldsymbol{T}$$

The Noise Amplification Problem

- Unfolding involves inverting the migration matrix to go from observed counts $\boldsymbol{D}$ to true counts $\boldsymbol{T}$
- Unfortunately this procedure tends to amplify noise in the data
- Right: 3-10% random scatter in the data is amplified to $> 20\%$ errors in unfolded counts (from Oser)
- The off-diagonal terms in $\mathcal{P}^\top \boldsymbol{D}$ may be large, causing oscillations (zig-zagging) in the unfolding



Regularization

- The maximum likelihood estimator (MLE) of $\boldsymbol{T}$ is $\mathcal{P}^\top \boldsymbol{D}$
- The MLE is the unbiased estimator with the smallest possible variance. Unfortunately the variance is still huge!
- Solution: introduce a bias into the estimator. The fit is now worse (w.r.t. the likelihood) but is smoothed by an amount that we specify
- Introduce a regularization parameter $S(\boldsymbol{T})$ that increases as $\boldsymbol{T}$ oscillates, and maximize

$$\tilde{\mathcal{L}}(\boldsymbol{T}) = \mathcal{L}\left(\boldsymbol{D}|\boldsymbol{T}\right) \cdot e^{-\alpha \cdot S(\boldsymbol{T})}$$

- Essentially we are defining the prior on $\boldsymbol{T}$ as

$$\pi(\boldsymbol{T}) = e^{-\alpha \cdot S(\boldsymbol{T})}$$

If the prior is constant then $p(\boldsymbol{T}|\boldsymbol{D})$ may be too wide, so different $\boldsymbol{T}$'s are equally likely. So we penalize certain solutions with $S(\boldsymbol{T})$

Tikhonov Regularization

- There is freedom to define α and the regularization function, which determines the smoothness of the unfolded counts
- Tikhonov regularization is a common approach for defining smoothing functions. We define

$$S(f) = -\int dx \left(\frac{d^k f}{dx^k}\right)^2$$

  where k represents the k^{th}-order derivative of f and f is the deconvolved distribution
- Common approach: set $k = 2$, which penalizes curvature in f (nonzero second derivatives)
- If we don't want to just favor linear functions of f we can use higher-order derivatives, or some linear combination of derivaties

Tikhonov Regularization

- In a binned spectrum, if we want to penalize the curvature between bins, we would write

$$S(\boldsymbol{T}) = \sum_{t=2}^{N_t-1} (\Delta_{t+1,t} - \Delta_{t,t-1})^2$$

where

$$\Delta_{t_1,t_2} = T_{t_1} - T_{t_2}$$

- One can also try to penalize variations in the first derivative and account for varying bin sizes:

$$S(\boldsymbol{T}) = \sum_{t=2}^{N_t-1} \frac{|\delta_{t+1,t} - \delta_{t,t-1}|}{|\delta_{t+1,t} + \delta_{t,t-1}|}$$

where δ_{t_1,t_2} is related to the bin width w_t and bin center c_t by

$$\delta_{t_1,t_2} = \frac{T_{t_1}/w_{t_1} - T_{t_2}/w_{t_2}}{c_{t_1} - c_{t_2}}, \quad w_t = m_t - m_{t-1}, \quad c_t = (m_t + m_{t-1})/2$$

Maximum Entropy Regularization

- Without any prior knowledge about the distribution of $\boldsymbol{T}$ in the bins, a reasonable choice for S is the maximum entropy

$$S(\boldsymbol{T}) = -\sum_{t=1}^{N_t} \frac{T_t}{\sum T_t} \ln \frac{T_t}{\sum T_t}$$

- Recall that the maximum entropy distribution is the one that favors the largest possible value for the bin multiplicity
- This means that it favors a relatively flat distribution, since entropy tends to be maximized when the bin counts are relatively equal
- If you don't want a flat distribution, you shouldn't use the maximum entropy. Instead, you can try cross-entropy

$$S(\boldsymbol{T}) = -\sum_{t=1}^{N_t} \frac{T_t}{\sum T_t} \ln \frac{T_t}{q_t \sum T_t}$$

  where q contains any prior knowledge you may have about the true distribution. If $q_i = 1/N_t$ this reduces back to the entropy

Regularization using Monte Carlo

- If you really trust your Monte Carlo $\tilde{\boldsymbol{T}}$ you could set your prior to [4]

$$\pi(\boldsymbol{T}) = \prod_{t=1}^{N_t} \exp\left[-\frac{(T_t - \tilde{T}_t)^2}{2(\tilde{T}_t/\alpha)^2}\right]$$

- Here the prior is proportional to a multivariate Gaussian with no correlations between bins
- The effect of the prior is to disfavor $\boldsymbol{T}$ far from $\tilde{\boldsymbol{T}}$
- The free parameter α adjusts the width of the Gaussian. Larger values of α imply a stricter constraint, forcing the unfolded counts to match the result of the Monte Carlo

Choosing a Regularization Parameter

There is no recipe for choosing a regularization parameter, but you can pick one of several criteria [2, 5]:

1. Minimize the mean squared error (MSE):

$$MSE = \frac{1}{N_t}\sum_{t=1}^{N_t} \text{var}\,(T_t) + \hat{b}_t^2, \qquad \hat{b}_t = \text{E}\,(T_t) - T_t$$

2. Tune α and S so that for each bin

$$2\Delta \ln \mathcal{L} = 2(\ln \mathcal{L}_{\text{max}} - \ln \mathcal{L}) = N_t$$

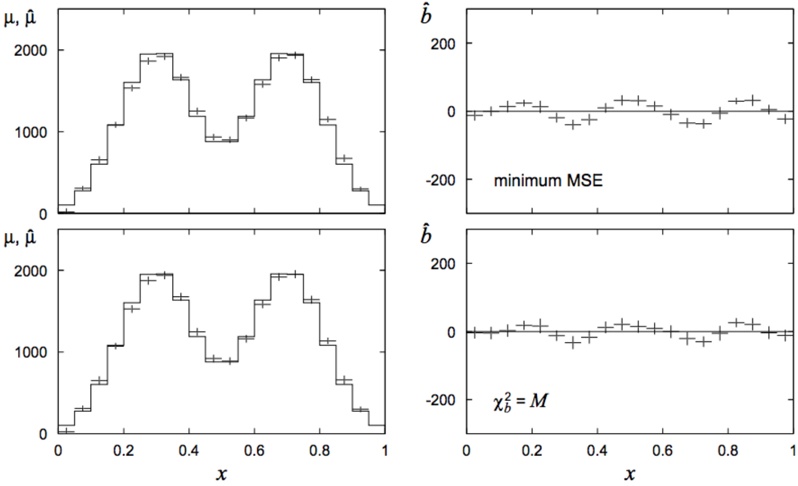
3. Tune α and S so that the biases and variances are balanced:

$$\chi_b^2 = \sum_{t=1}^{N_t} \frac{\hat{b}_t^2}{\text{var}\,(\hat{b}_t)} \approx N_t$$

There is a trade-off between bias and variance that you have to choose for your application

Example: Unfolding with Tikhonov Regularization

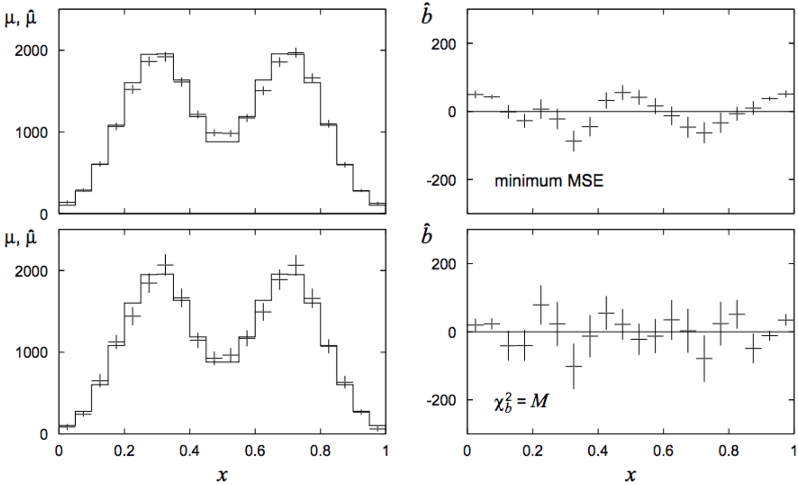
Unfolded distributions using Tikhonov regularization [2]



The parameter α was tuned using the MSE and $\chi_b^2 = N_t$

Example: Unfolding with Maximum Entropy Regularization

Unfolded distributions using MaxEnt regularization [2]

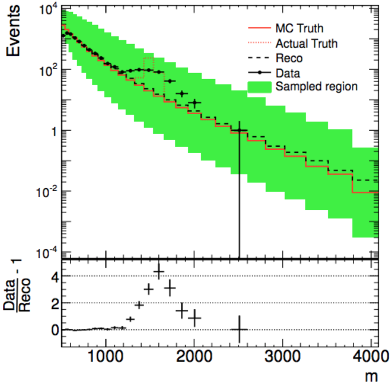


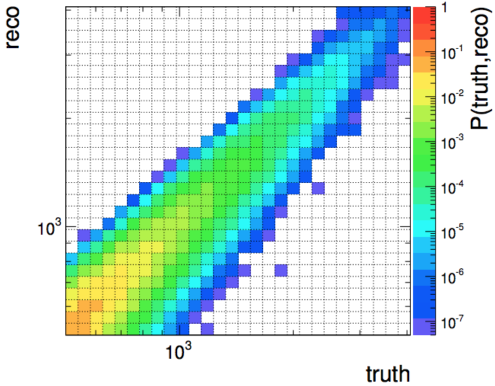
The parameter α was tuned using the MSE and $\chi^2_b = N_t$

Table of Contents

Example: Steeply Falling Spectrum with a Bump

A steeply falling spectrum with a bump, from [4]





The true spectrum has a bump. The Monte Carlo truth does not. The data are smeared by the bin-to-bin migration matrix shown at right

MCMC Sampling

The procedure for unfolding is as follows:

- Choose a prior for $\boldsymbol{T}$
- Sample the N_t-dimensional hypercube using MCMC
- For each bin, find the mode (maximum) of the posterior $p(\boldsymbol{T}|\boldsymbol{D})$

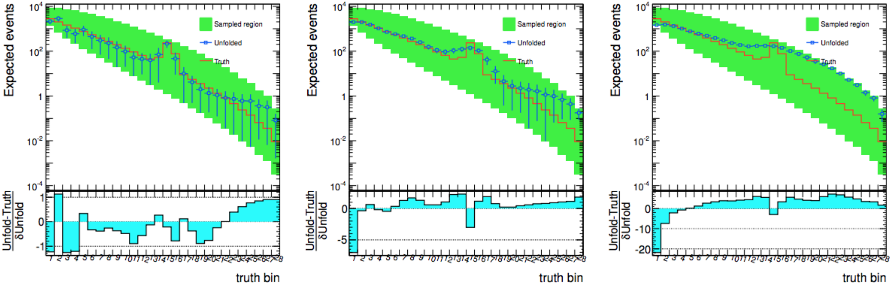
It's good practice to try different smoothing penalties $S(\boldsymbol{T})$ and smoothing factors α

Usual caveats about MCMC: only use data after burn-in, and plot the marginal distributions of $\boldsymbol{T}$ to see if they are unimodal

Unfolding with a MaxEnt Penalty

The unfolded spectrum is reconstructed using maximum entropy regularization

$$S(\boldsymbol{T}) = -\sum_{t=1}^{N_t} \frac{T_t}{\sum T_t} \ln \frac{T_t}{\sum T_t}$$

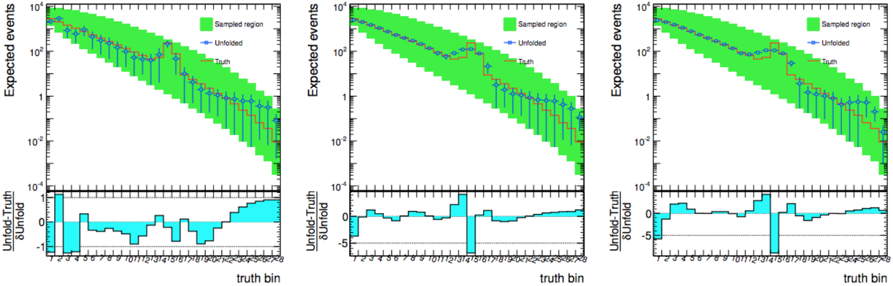


Left: no penalty ($\alpha = 0$). Middle: $\alpha = 10^3$. Right: $\alpha = 3 \times 10^3$.

Unfolding with a Curvature Penalty

The unfolded spectrum is reconstructed using the curvature penalty

$$S(\boldsymbol{T}) = \sum_{t=2}^{N_t-1} (\Delta_{t+1,t} - \Delta_{t,t-1})^2$$

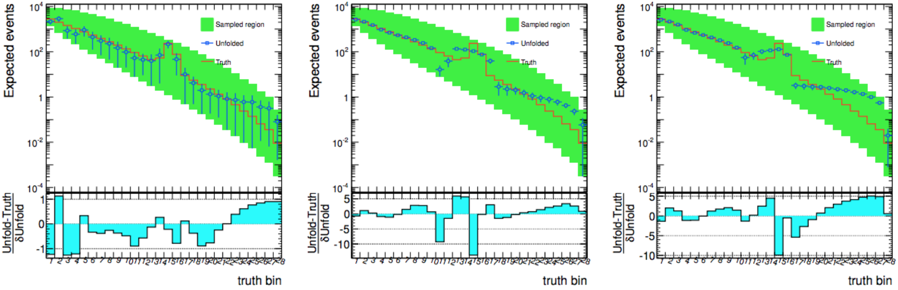


Left: no penalty ($\alpha = 0$). Middle: $\alpha = 3 \times 10^{-4}$. Right: $\alpha = 6 \times 10^{-4}$.

Unfolding Accounting for Nonuniform Bins

The unfolded spectrum is reconstructed using the penalty

$$S(\boldsymbol{T}) = \sum_{t=2}^{N_t-1} \frac{|\delta_{t+1,t} - \delta_{t,t-1}|}{|\delta_{t+1,t} + \delta_{t,t-1}|}$$

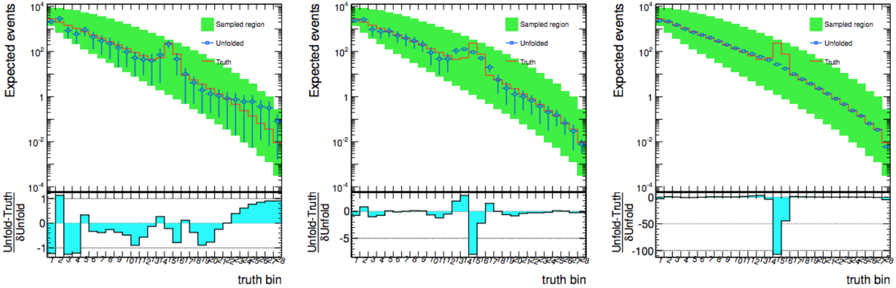


Left: no penalty ($\alpha = 0$). Middle: $\alpha = 10$. Right: $\alpha = 20$.

Unfolding with Gaussian Regularization

The unfolded spectrum is reconstructed using the prior

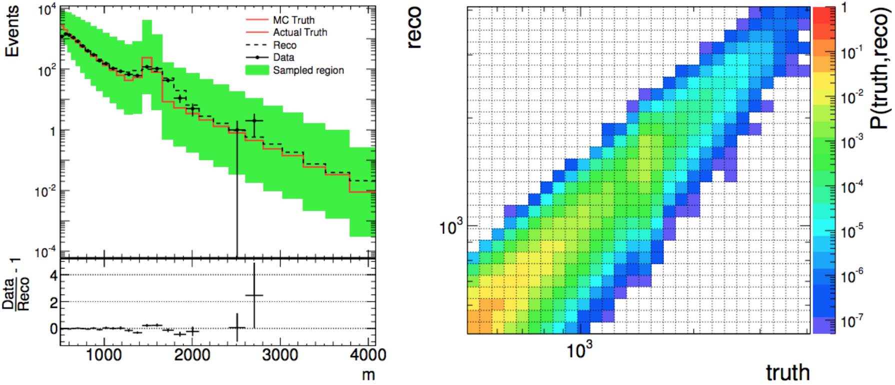
$$\pi(\boldsymbol{T}) = \prod_{t=1}^{N_t} \exp\left[-\frac{(T_t - \tilde{T}_t)^2}{2(\tilde{T}_t/\alpha)^2}\right]$$



Left: no penalty ($\alpha = 0$). Middle: $\alpha = 1$. Right: $\alpha = 10$.

Example: Steeply Falling Spectrum with an Expected Bump

A steeply falling spectrum with a bump, from [4]

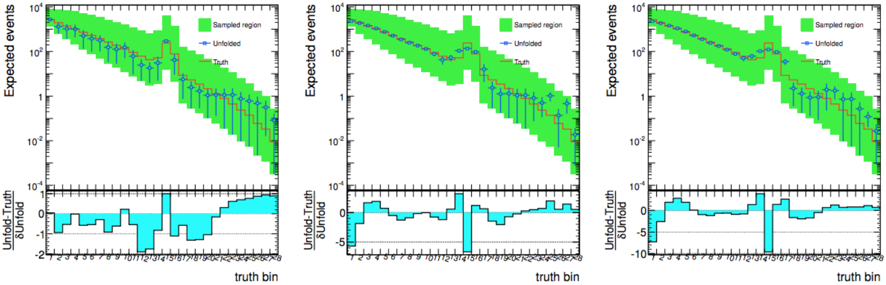


This time, the bump is expected and included in the MC truth. Is there any difference in the unfolding?

Unfolding with a Curvature Penalty

The unfolded spectrum is reconstructed using the curvature penalty

$$S(\boldsymbol{T}) = \sum_{t=2}^{N_t-1} (\Delta_{t+1,t} - \Delta_{t,t-1})^2$$



No real improvement in appearance of the bump w.r.t. case where $\tilde{\boldsymbol{T}}$ did not contain a bump

Summary

- Unfolding is a technique used to remove instrumental smearing and efficiency artifacts from a binned spectrum
- After unfolding, the unbiased maximum likelihood estimator tends to have big variances which show up as zig-zagging between neighboring bins
- The fix for oscillations is to apply a smoothing function that penalizes zig-zagging. There is a lot of freedom in how to do this
- There is a trade off between the bias in the estimator and the variance. You have to decide what is appropriate; there is no recipe
- Best approach: run a data challenge to see if the kind of effect you are looking for is washed out by how you unfold
- Cowan suggests several figures of merit for balancing variance and bias [2, 5] that are good starting points for this kind of analysis

References I

[1] Vardan Khachatryan et al. "Observation of the diphoton decay of the Higgs boson and measurement of its properties". In: *Eur.Phys.J.* C74.10 (2014), p. 3076. arXiv: 1407.0558 [hep-ex].

[2] G. Cowan. "A survey of unfolding methods for particle physics". In: *Conf. Adv. Stat. Tech. in Part. Phys.* Durham, England, 2002, pp. 248–257.

[3] *MaxEnt: Maximum Entropy Data Consultants Ltd.* 2007. URL: http://www.maxent.co.uk/.

[4] Georgios Choudalakis. "Fully Bayesian Unfolding". In: (2012). arXiv: 1201.4612 [physics.data-an].

[5] Glen Cowan. *Statistical Data Analysis*. New York: Oxford University Press, 1998.