Unfolding: A Statistician's Perspective

Mikael Kuusela

Department of Statistics
University of Chicago

Phystat-ν

Fermilab
September 21, 2016

The unfolding problem

- Any measurement is affected by the finite resolution of the particle detectors
  - This causes the observed spectrum of events to be “smeared” or “blurred” with respect to the true one
- The *unfolding problem* is to estimate the true spectrum using the smeared observations
- Mathematically closely related to deblurring in optics and tomographic image reconstruction in medical imaging

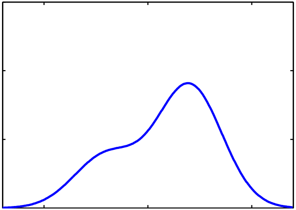
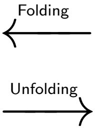


Figure : Smeared spectrum



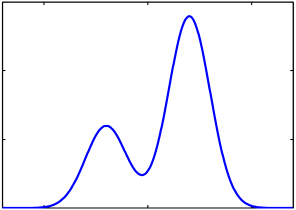


Figure : True spectrum

Problem formulation

- Let f be the true, particle-level spectrum and g the smeared, detector-level spectrum
  - Denote the true space by E and the smeared space by F (both taken to be intervals on the real line)
  - Mathematically f and g are the intensity functions of the underlying Poisson point process
- The two spectra are related by

$$g(t) = \int_E k(t,s) f(s)\, \mathrm{d}s,$$

where the smearing kernel k represents the response of the detector and is given by

$$k(t,s) = p(Y = t | X = s, X \text{ observed}) P(X \text{ observed} | X = s),$$

where X is a true event and Y the corresponding smeared event

Task: Infer the true spectrum f given smeared observations from g

Discretization

- Problem primarily discretized using histograms (splines are also sometimes used)
- Let $\{E_i\}_{i=1}^{p}$ and $\{F_i\}_{i=1}^{n}$ be binnings of the true space E and the smeared space F
- Let $\mathbf{y}$ be the corresponding histogram of smeared observations with mean vector

$$\boldsymbol{\mu} = \left[\int_{F_1} g(t)\,\mathrm{d}t, \ldots, \int_{F_n} g(t)\,\mathrm{d}t\right]^{\mathsf{T}}$$

- Let us set out to make inferences about the particle-level mean histogram

$$\boldsymbol{\lambda} = \left[\int_{E_1} f(s)\,\mathrm{d}s, \ldots, \int_{E_p} f(s)\,\mathrm{d}s\right]^{\mathsf{T}}$$

- The mean histograms are related by $\boldsymbol{\mu} = \mathbf{K}\boldsymbol{\lambda}$, where the elements of the *response matrix* $\mathbf{K}$ are given by

$$K_{i,j} = \frac{\int_{F_i}\int_{E_j} k(t,s) f(s)\,\mathrm{d}s\,\mathrm{d}t}{\int_{E_j} f(s)\,\mathrm{d}s} = P(\text{smeared event in bin } i \,|\, \text{true event in bin } j)$$

- The discretized statistical model becomes

$$\mathbf{y} \sim \mathrm{Poisson}(\mathbf{K}\boldsymbol{\lambda}),$$

where $\mathbf{K}$ is an ill-conditioned matrix

Discretization

- Problem primarily discretized using histograms (splines are also sometimes used)
- Let $\{E_i\}_{i=1}^p$ and $\{F_i\}_{i=1}^n$ be binnings of the true space E and the smeared space F
- Let $\mathbf{y}$ be the corresponding histogram of smeared observations with mean vector

$$\boldsymbol{\mu} = \left[\int_{F_1} g(t)\,\mathrm{d}t, \dots, \int_{F_n} g(t)\,\mathrm{d}t \right]^{\mathrm{T}}$$

- Let us set out to make inferences about the particle-level mean histogram

$$\boldsymbol{\lambda} = \left[\int_{E_1} f(s)\,\mathrm{d}s, \dots, \int_{E_p} f(s)\,\mathrm{d}s \right]^{\mathrm{T}}$$

- The mean histograms are related by $\boldsymbol{\mu} = \mathbf{K}\boldsymbol{\lambda}$, where the elements of the *response matrix* $\mathbf{K}$ are given by

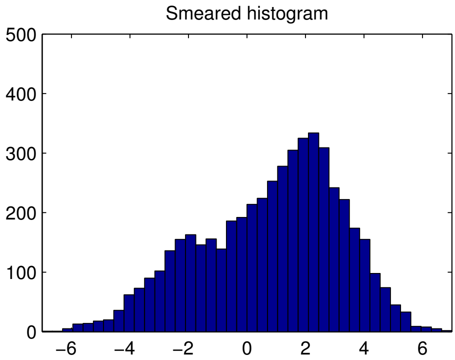
$$K_{i,j} = \frac{\int_{F_i}\int_{E_j} k(t,s) f^{\mathrm{MC}}(s)\,\mathrm{d}s\,\mathrm{d}t}{\int_{E_j} f^{\mathrm{MC}}(s)\,\mathrm{d}s} = P(\text{smeared event in bin } i \,|\, \text{true event in bin } j)$$

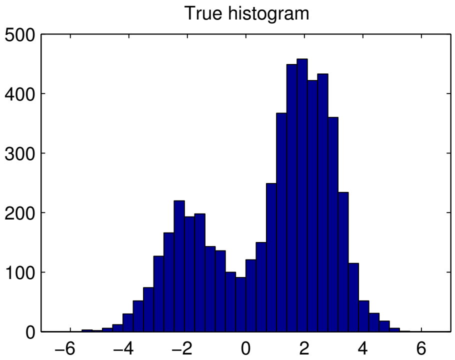
- The discretized statistical model becomes

$$\boxed{\mathbf{y} \sim \mathrm{Poisson}(\mathbf{K}\boldsymbol{\lambda}),}$$

where $\mathbf{K}$ is an ill-conditioned matrix

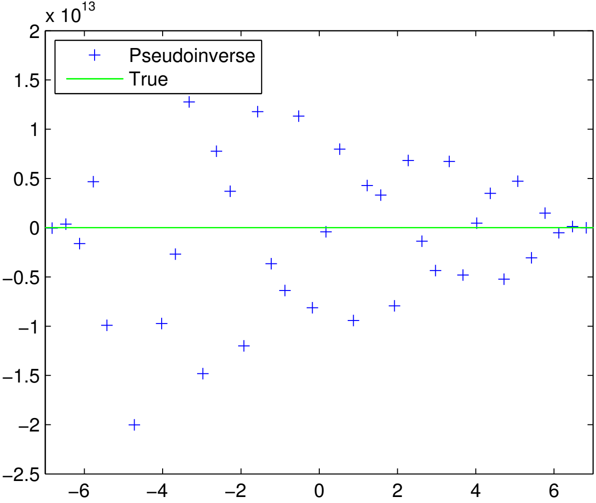
Demonstration of ill-posedness



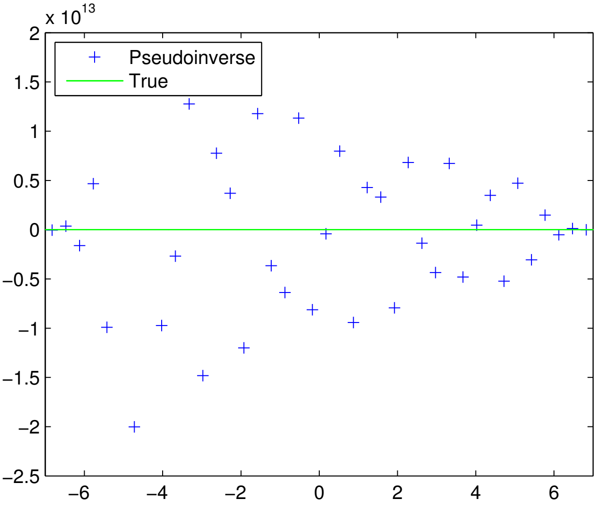


$$\boldsymbol{\mu} = \mathbf{K}\boldsymbol{\lambda}, \quad \mathbf{y} \sim \mathrm{Poisson}(\boldsymbol{\mu}) \quad \overset{??}{\Longrightarrow} \quad \hat{\boldsymbol{\lambda}} = \mathbf{K}^{-1}\mathbf{y}$$

Demonstration of ill-posedness



Demonstration of ill-posedness



Regularization: bias ↑, variance ↓

Current methodology

Two main approaches to unfolding (at least in LHC experiments):

1. Tikhonov regularization (Höcker and Kartvelishvili, 1996; Schmitt, 2012)
2. Expectation-maximization iteration with early stopping (D'Agostini, 1995; Richardson, 1972; Lucy, 1974; Shepp and Vardi, 1982; Lange and Carson, 1984; Vardi et al., 1985)

Tikhonov regularization

- Tikhonov regularization estimates $\boldsymbol{\lambda}$ by solving:

$$\min_{\boldsymbol{\lambda} \in \mathbb{R}^p} (\mathbf{y} - \mathbf{K}\boldsymbol{\lambda})^{\mathrm{T}} \hat{\mathbf{C}}^{-1} (\mathbf{y} - \mathbf{K}\boldsymbol{\lambda}) + \delta P(\boldsymbol{\lambda})$$

- The first term as a Gaussian approximation to the Poisson log-likelihood
- The second term penalizes physically implausible solutions
- Common penalty terms:
  - Norm: $P(\boldsymbol{\lambda}) = \|\boldsymbol{\lambda}\|^2$
  - Curvature: $P(\boldsymbol{\lambda}) = \|\mathbf{L}\boldsymbol{\lambda}\|^2$, where $\mathbf{L}$ is a discretized 2nd derivative operator
  - SVD unfolding (Höcker and Kartvelishvili, 1996):

$$P(\boldsymbol{\lambda}) = \left\| \mathbf{L} \begin{bmatrix} \lambda_1/\lambda_1^{\mathrm{MC}} \\ \lambda_2/\lambda_2^{\mathrm{MC}} \\ \vdots \\ \lambda_p/\lambda_p^{\mathrm{MC}} \end{bmatrix} \right\|^2,$$

    where $\boldsymbol{\lambda}^{\mathrm{MC}}$ is a MC prediction for $\boldsymbol{\lambda}$
  - `TUnfold`[1] (Schmitt, 2012): $P(\boldsymbol{\lambda}) = \|\mathbf{L}(\boldsymbol{\lambda} - \boldsymbol{\lambda}^{\mathrm{MC}})\|^2$

[1] `TUnfold` implements also more general penalty terms

D'Agostini iteration

- Starting from some initial guess $\boldsymbol{\lambda}^{(0)} > \mathbf{0}$, iterate

$$\lambda_j^{(k+1)} = \frac{\lambda_j^{(k)}}{\sum_{i=1}^{n} K_{i,j}} \sum_{i=1}^{n} \frac{K_{i,j} y_i}{\sum_{l=1}^{p} K_{i,l} \lambda_l^{(k)}}$$

- Regularization by stopping the iteration before convergence:
  - $\hat{\boldsymbol{\lambda}} = \boldsymbol{\lambda}^{(K)}$ for some small number of iterations K
  - I.e., bias the solution towards $\boldsymbol{\lambda}^{(0)}$
- In `RooUnfold` (Adye, 2011), $\boldsymbol{\lambda}^{(0)} = \boldsymbol{\lambda}^{\mathrm{MC}}$

D'Agostini iteration

$$\lambda_j^{(k+1)} = \frac{\lambda_j^{(k)}}{\sum_{i=1}^n K_{i,j}} \sum_{i=1}^n \frac{K_{i,j} y_i}{\sum_{l=1}^p K_{i,l} \lambda_l^{(k)}}$$

- This iteration has been discovered in various fields, including optics (Richardson, 1972), astronomy (Lucy, 1974) and tomography (Shepp and Vardi, 1982; Lange and Carson, 1984; Vardi et al., 1985)
- In particle physics, it was popularized by D'Agostini (1995) who called it "Bayesian" unfolding
- But: This is in fact an expectation-maximization (EM) iteration (Dempster et al., 1977) for finding the *maximum likelihood estimator* of $\boldsymbol{\lambda}$ in the Poisson regression problem $\mathbf{y} \sim \mathrm{Poisson}(\mathbf{K}\boldsymbol{\lambda})$
- As $k \to \infty$, $\boldsymbol{\lambda}^{(k)} \to \hat{\boldsymbol{\lambda}}_{\mathrm{MLE}}$ (Vardi et al., 1985)
- *This is a fully frequentist technique for finding the (regularized) MLE*
    - The name "Bayesian" is an unfortunate misnomer

D'Agostini demo, $k = 0$

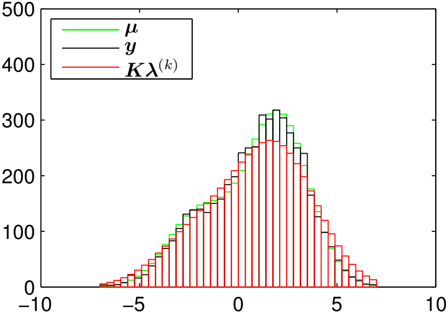


Figure : Smeared histogram

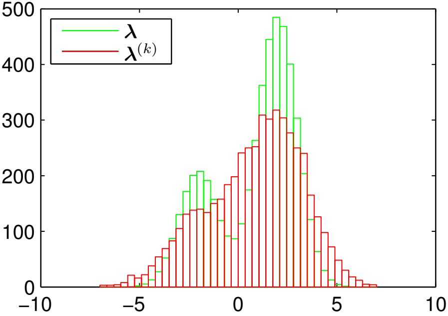


Figure : True histogram

D'Agostini demo, $k = 100$

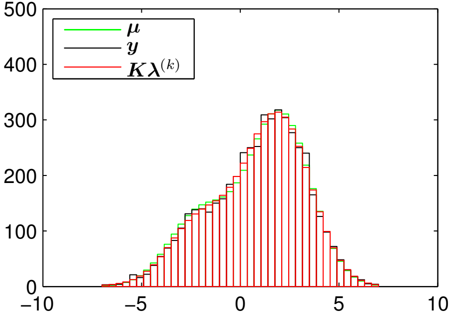


Figure : Smeared histogram

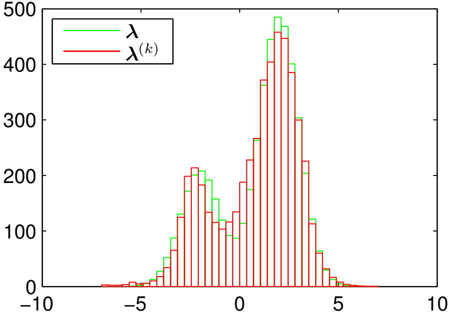


Figure : True histogram

D'Agostini demo, $k = 10000$

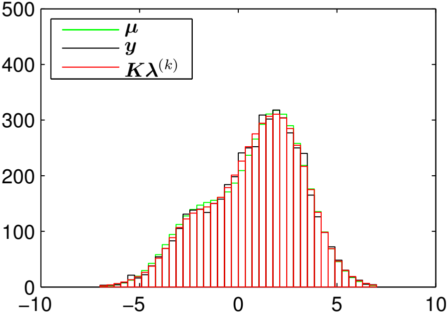


Figure : Smeared histogram

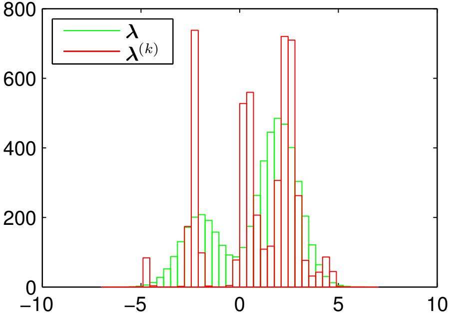


Figure : True histogram

D'Agostini demo, $k = 100000$

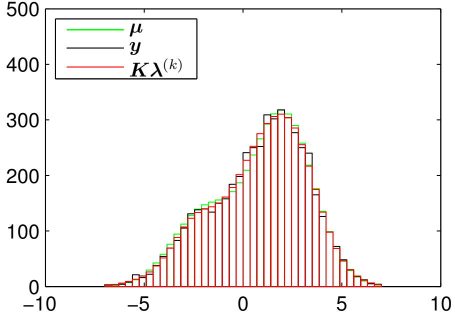


Figure : Smeared histogram

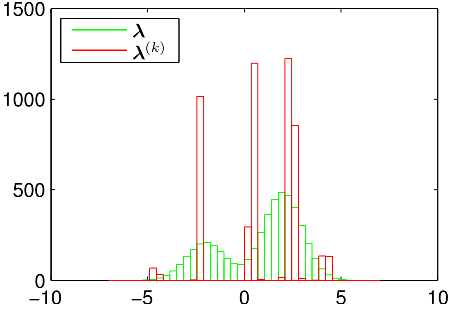


Figure : True histogram

Choice of the regularization strength

- A key issue in unfolding concerns the choice of the regularization strength (δ in Tikhonov, K in D'Agostini)
- Many data-driven methods have been proposed:
  - Cross-validation (Stone, 1974)
  - L-curve (Hansen, 1992)
  - Empirical Bayes estimation (Kuusela and Panaretos, 2015)
  - Goodness-of-fit test in the smeared space (Veklerov and Llacer, 1987)
  - Akaike information criterion (Volobouev, 2015)
  - Minimization of a global correlation coefficient (Schmitt, 2012)
  - ...
- Limited experience about the relative merits of these methods in typical unfolding problems
  - Some evidence that empirical Bayes tends to be more stable than cross-validation (Kuusela, 2016; Wood, 2011)
- Notice that all these are aiming to do optimal point estimation
  - Not necessarily optimal for uncertainty quantification!

Some comments based on experience from the LHC

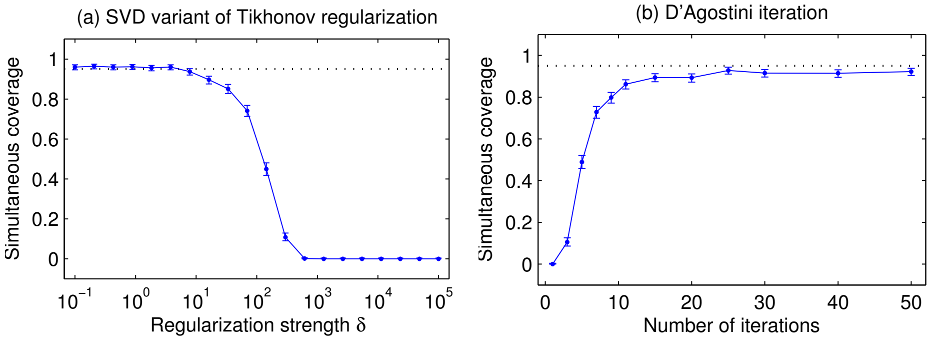
- One should think carefully if unfolding is *really* needed
  - E.g., if the goal if the experiment is to measure just a few 1-dimensional parameters, then one should perform the fit in the smeared space (as opposed to inferring the quantities from the regularized unfolded spectrum)
  - What about smearing the theory instead of unfolding the data? (Complicated by systematics in the response matrix)
  - Unfolding can be useful for comparison of experiments, propagation to further analyses, tuning of MC generators, exploratory data analysis,...
- One should analyze carefully if regularization is necessary
  - If there is little smearing (response matrix almost diagonal), then the MLE obtained by running D'Agostini until convergence will do the job[2]
  - Some insight can be obtained by studying the condition number of $\mathbf{K}$
- One must not rely on software defaults for the regularization strength
  - The unfolded solution is very sensitive to this choice and the optimal choice is very problem dependent

[2] The matrix inverse $\hat{\boldsymbol{\lambda}} = \mathbf{K}^{-1}\mathbf{y}$ also gives the MLE provided that $\mathbf{K}$ is invertible and $\hat{\boldsymbol{\lambda}} \geq \mathbf{0}$

Some comments based on experience from the LHC

- The standard methods (at least as implemented in `RooUnfold`) regularize by biasing the solution towards the MC prediction $\boldsymbol{\lambda}^{\mathrm{MC}}$
  - Danger of producing over-optimistic results, as too strong regularization will always make the unfolded histogram match the MC, whether the MC is correct or not
  - Safer to use MC-independent regularization (possible in `TUnfold`)
- Uncertainty quantification (i.e., providing confidence intervals) in the unfolded space is a very delicate matter
  - When regularization is used, the variance alone may not be a good measure of uncertainty because it ignores the bias
  - But the bias is needed to regularize the problem... (Cf. inference in lasso regression and spline smoothing)
  - Two ways to alleviate this situation:
    1. Debiased confidence intervals (Kuusela and Panaretos, 2015; Kuusela, 2016)
    2. Shape-constrained confidence intervals (Kuusela and Stark, 2016; Kuusela, 2016)

Undercoverage of existing methods



$$f(s) = \lambda_{\mathrm{tot}} \left\{ \pi_1 \mathcal{N}(s|-2,1) + \pi_2 \mathcal{N}(s|2,1) + \pi_3 \frac{1}{|E|} \right\}$$

$$g(t) = (Kf)(t) = \int_E \mathcal{N}(t-s|0,1) f(s)\,\mathrm{d}s$$

$$f^{\mathrm{MC}}(s) = \lambda_{\mathrm{tot}} \left\{ \pi_1 \mathcal{N}(s|-2,1.1^2) + \pi_2 \mathcal{N}(s|2,0.9^2) + \pi_3 \frac{1}{|E|} \right\}$$

Unfolding using debiased confidence intervals

Method	Coverage at $s = 0$	Mean length
Bias-correction (data-driven)	0.932 *(0.915, 0.947)*	0.079 *(0.077, 0.081)*
Bias-correction (oracle)	0.937 *(0.920, 0.951)*	0.064 *(0.064, 0.064)*
Undersmoothing (data-driven)	0.933 *(0.916, 0.948)*	0.091 *(0.087, 0.095)*
Undersmoothing (oracle)	0.949 *(0.933, 0.962)*	0.070 *(0.070, 0.070)*
MMLE	0.478 *(0.447, 0.509)*	0.030 *(0.030, 0.030)*
MISE	0.359 *(0.329, 0.390)*	0.028
Unregularized	0.952 *(0.937, 0.964)*	40316

MMLE = choose δ to maximize the marginal likelihood

MISE = choose δ to minimize the mean integrated squared error

Unfolding using shape-constrained confidence intervals

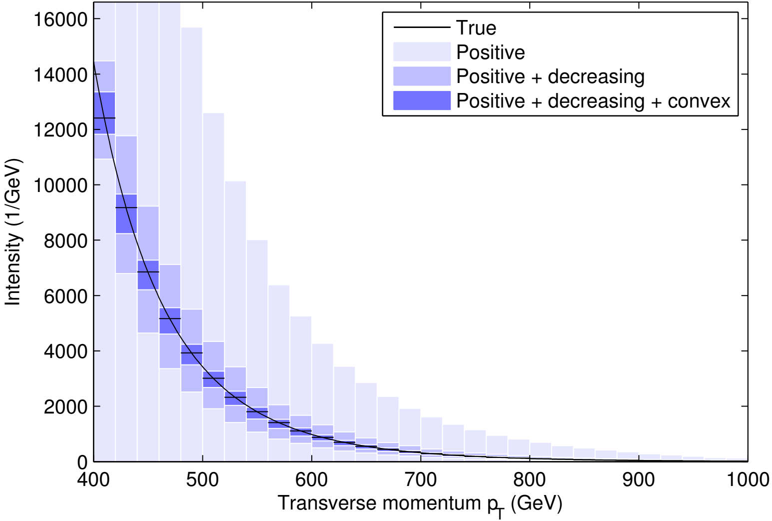


Figure : Shape-constrained unfolded confidence intervals for the inclusive jet p_{T} spectrum with *guaranteed* conservative 95 % simultaneous coverage.

Summary and conclusions

- Ill-posedness makes unfolding a very complex problem
- Tikhonov regularization and D'Agostini iteration are the most popular techniques
    - I personally find it easier to interpret the 2nd derivative penalty in Tikhonov than the early stopping in D'Agostini
- Proper choice of the regularization strength is crucial
    - A choice that is optimal for point estimation might not be optimal for uncertainty quantification
- Results from standard software (`RooUnfold`) depend strongly on the MC prediction (results biased towards the MC)
    - Better to use MC-independent regularization
- Uncertainties from standard techniques can be unreliable
    - Improved uncertainty quantification can be achieved by debiasing or by imposing qualitative shape constraints
- Many open statistical issues remain:
    - How to properly present unfolded results? (Bins are correlated)
    - How to properly deal with systematic uncertainties in the detector response?
    - How to properly compare, combine and propagate unfolded results?

References I

Adye, T. (2011). Unfolding algorithms and tests using RooUnfold. In Prosper, H. B. and Lyons, L., editors, *Proceedings of the PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding*, CERN-2011-006, pages 313–318, CERN, Geneva, Switzerland.

CMS Collaboration (2011). Measurement of the inclusive jet cross section in pp collisions at $\sqrt{s} = 7\ \mathrm{TeV}$. *Physical Review Letters*, 107:132001.

D'Agostini, G. (1995). A multidimensional unfolding method based on Bayes' theorem. *Nuclear Instruments and Methods A*, 362:487–498.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38.

Hall, P. (1992). Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density. *The Annals of Statistics*, 20(2):675–694.

References II

Hansen, P. C. (1992). Analysis of discrete ill-posed problems by means of the L-curve. *SIAM Review*, 34(4):561–580.

Höcker, A. and Kartvelishvili, V. (1996). SVD approach to data unfolding. *Nuclear Instruments and Methods in Physics Research A*, 372:469–481.

Kuusela, M. (2016). *Uncertainty quantification in unfolding elementary particle spectra at the Large Hadron Collider*. PhD thesis, EPFL. `https://infoscience.epfl.ch/record/220015`.

Kuusela, M. and Panaretos, V. M. (2015). Statistical unfolding of elementary particle spectra: Empirical Bayes estimation and bias-corrected uncertainty quantification. *The Annals of Applied Statistics*, 9(3):1671–1705.

Kuusela, M. and Stark, P. B. (2016). Shape-constrained uncertainty quantification in unfolding steeply falling elementary particle spectra. arXiv:1512.00905v3 [stat.AP], submitted.

References III

Lange, K. and Carson, R. (1984). EM reconstruction algorithms for emission and transmission tomography. *Journal of Computer Assisted Tomography*, 8(2):306–316.

Lucy, L. B. (1974). An iterative technique for the rectification of observed distributions. *Astronomical Journal*, 79(6):745–754.

Richardson, W. H. (1972). Bayesian-based iterative method of image restoration. *Journal of the Optical Society of America*, 62(1):55–59.

Schmitt, S. (2012). TUnfold, an algorithm for correcting migration effects in high energy physics. *Journal of Instrumentation*, 7:T10003.

Shepp, L. A. and Vardi, Y. (1982). Maximum likelihood reconstruction for emission tomography. *IEEE Transactions on Medical Imaging*, 1(2):113–122.

Stark, P. B. (1992). Inference in infinite-dimensional inverse problems: Discretization and duality. *Journal of Geophysical Research*, 97(B10):14055–14082.

References IV

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 36:111–147.

Vardi, Y., Shepp, L. A., and Kaufman, L. (1985). A statistical model for positron emission tomography. *Journal of the American Statistical Association*, 80(389):8–20.

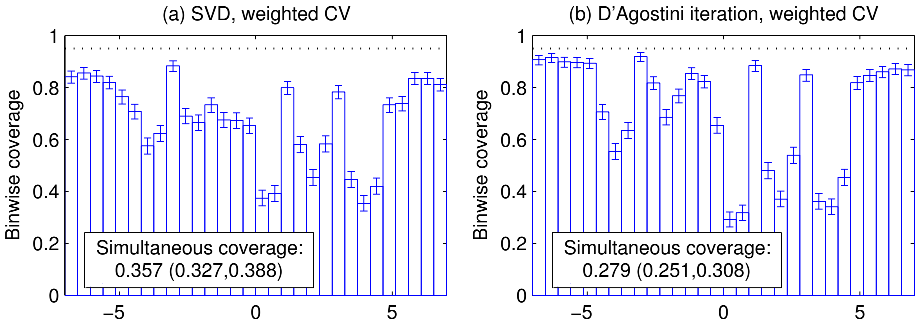
Veklerov, E. and Llacer, J. (1987). Stopping rule for the MLE algorithm based on statistical hypothesis testing. *IEEE Transactions on Medical Imaging*, 6(4):313–319.

Volobouev, I. (2015). On the expectation-maximization unfolding with smoothing. arXiv:1408.6500v2 [physics.data-an].

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 73(1):3–36.

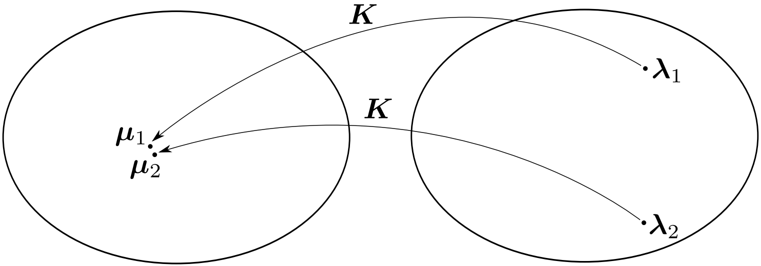
Backup

Undercoverage of existing methods

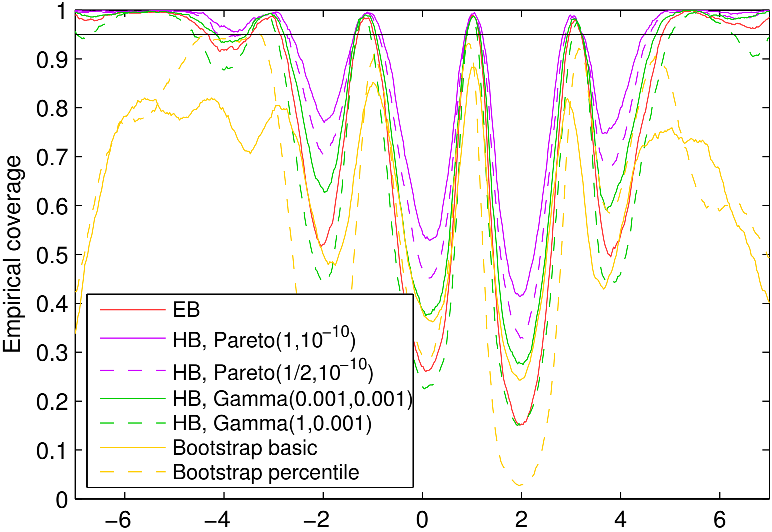


Unfolding is an ill-posed inverse problem

- The linear system $\boldsymbol{\mu} = \mathbf{K}\boldsymbol{\lambda}$ is typically ill-conditioned
  - That is, true histograms $\boldsymbol{\lambda}$ that are very different can map into smeared histograms $\boldsymbol{\mu}$ that are very similar
- As a result, the (pseudo)inverse of $\mathbf{K}$ is very sensitive to statistical fluctuations in the smeared data



UQ in inverse problems is challenging

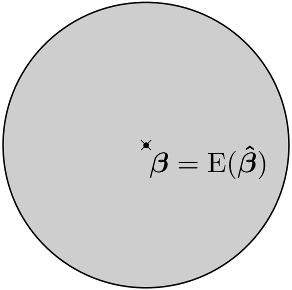


Details in Kuusela and Panaretos (2015) and Kuusela (2016)

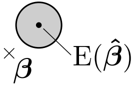
Towards improved uncertainty quantification

- Standard methods based on the variability of the unfolded point estimator do not adequately take the bias into account
  - As a result, they may suffer from *drastic undercoverage* in realistic unfolding scenarios (several examples given in Kuusela (2016))
- These exists very little previous work on how to solve this problem
- I have investigated two complementary approaches for improving this situation:
  1. Debiased confidence intervals based on an iterative bias-correction (Kuusela and Panaretos, 2015; Kuusela, 2016)
     - Yields significantly improved coverage performance with only a modest increase in the interval length
  2. Shape-constrained confidence intervals (Kuusela and Stark, 2016; Kuusela, 2016)
     - Yields guaranteed conservative coverage, provided that the spectrum satisfies simple shape constraints (positivity, monotonicity, convexity)

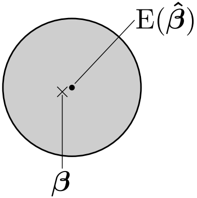
Optimal point estimation $\neq$ optimal interval estimation



(a) Unbiased



(b) Optimal point estimation



(c) Optimal UQ

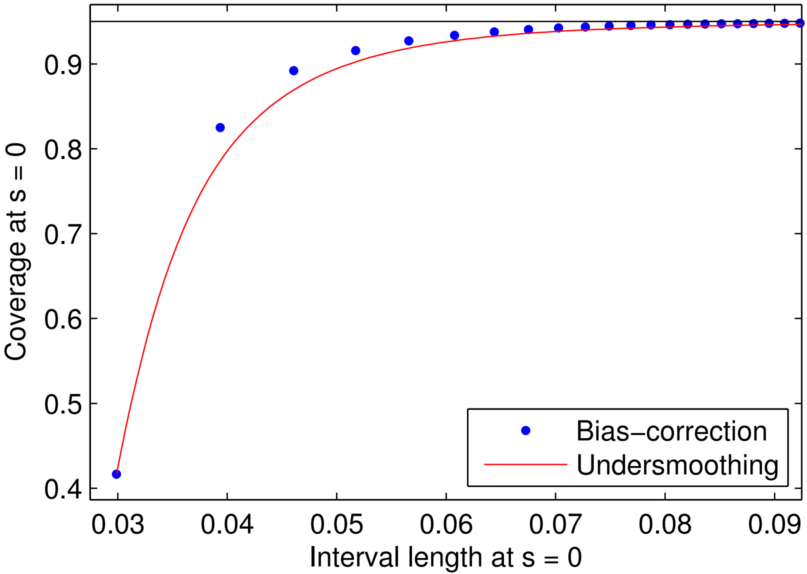
Debiasing

- Let $\hat{\beta}$ be the unfolded point estimator depending on the regularization strength δ (large $\delta \leftrightarrow$ strong regularization)
- A trivial way of debiasing $\hat{\beta}$ is to *undersmooth* (Hall, 1992) by choosing δ to be smaller than the value that would lead to optimal point estimation performance
  - The variability of the debiased point estimator is then used to construct confidence intervals
- Data-driven choice of δ: Calibrate $1-\alpha$ intervals to have coverage $1-\alpha-\varepsilon$
- But: One can obtain even more powerful inferences (i.e. shorter confidence intervals) by employing an iterative bias-correction:

Iterative bias-correction

1. Estimate the bias: $\widehat{\mathrm{bias}}^{(t)}(\hat{\beta}) = \mathrm{E}_{\hat{\beta}^{(t)}}(\hat{\beta}) - \hat{\beta}^{(t)}$
2. Compute the bias-corrected estimate: $\hat{\beta}^{(t+1)} = \hat{\beta}^{(0)} - \widehat{\mathrm{bias}}^{(t)}(\hat{\beta})$

Coverage-length trade-off



Demonstration: $Z \to e^+e^-$ invariant mass spectrum

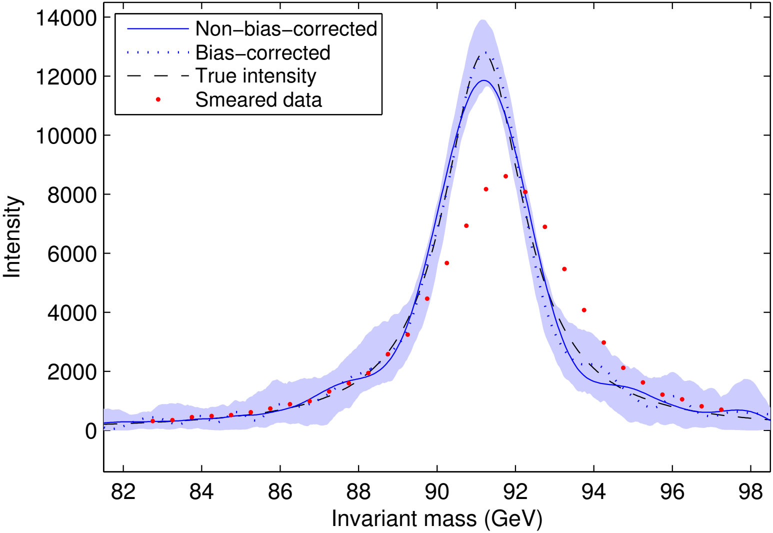
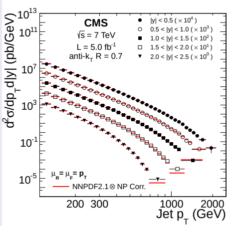


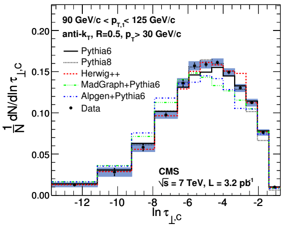
Figure : Unfolding of the Z boson invariant mass spectrum, 95 % iteratively bias-corrected percentile intervals, calibrated to have 94 % target coverage

Examples of unfolding in LHC data analysis

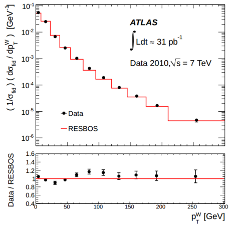
Inclusive jet cross section



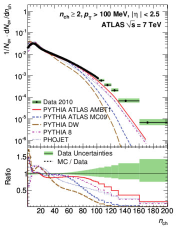
Hadronic event shape



W boson cross section



Charged particle multiplicity



Data-driven deconvolved confidence intervals

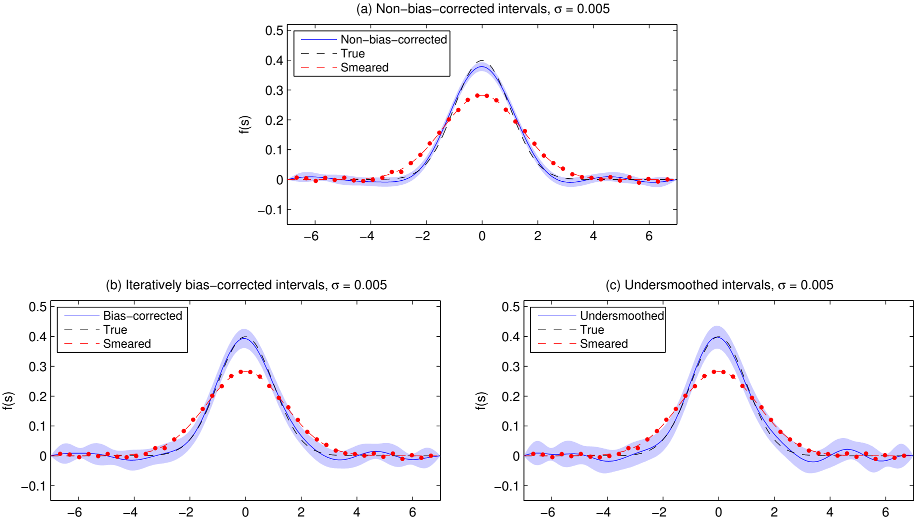
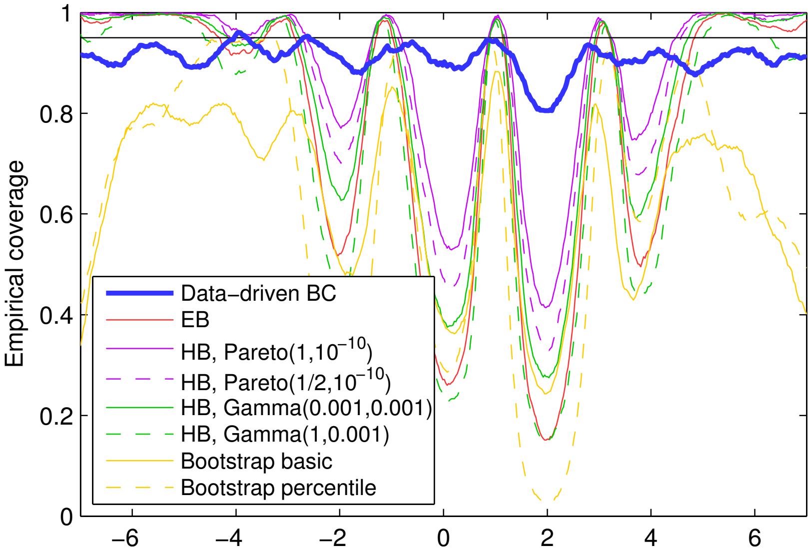
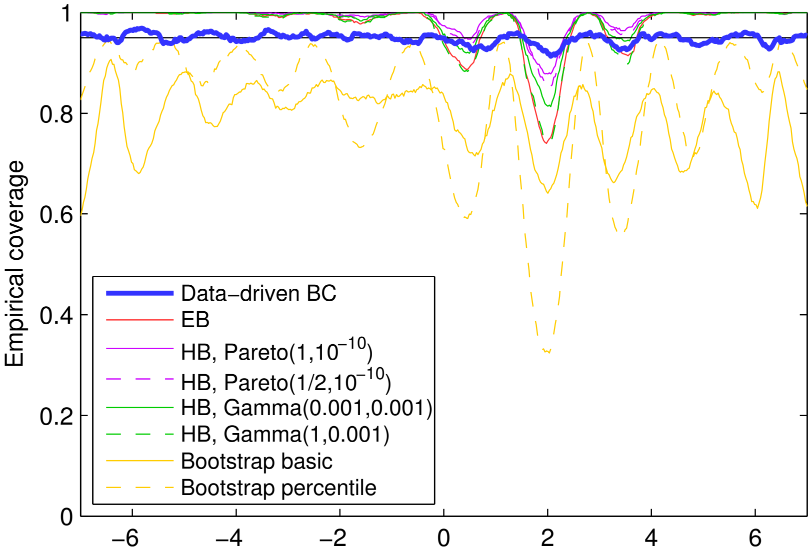


Figure : Gaussian intervals, 95 % nominal coverage, 94 % target coverage

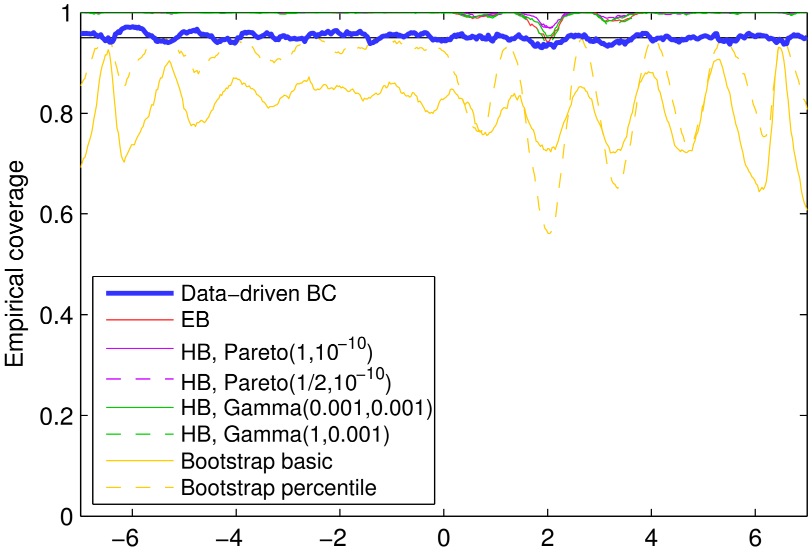
Comparison of coverage performance, $\lambda_{\text{tot}} = 1\,000$



Comparison of coverage performance, $\lambda_{\text{tot}} = 10\,000$



Comparison of coverage performance, $\lambda_{\text{tot}} = 50\,000$



Shape-constrained unfolding of steeply falling spectra

Joint work with Philip B. Stark (UC Berkeley)

Shape-constrained unfolding [arXiv:1512.00905]

We present a technique for forming unfolded confidence intervals with *guaranteed finite-sample simultaneous frequentist coverage*, provided that f satisfies simple, physically justified shape constraints.

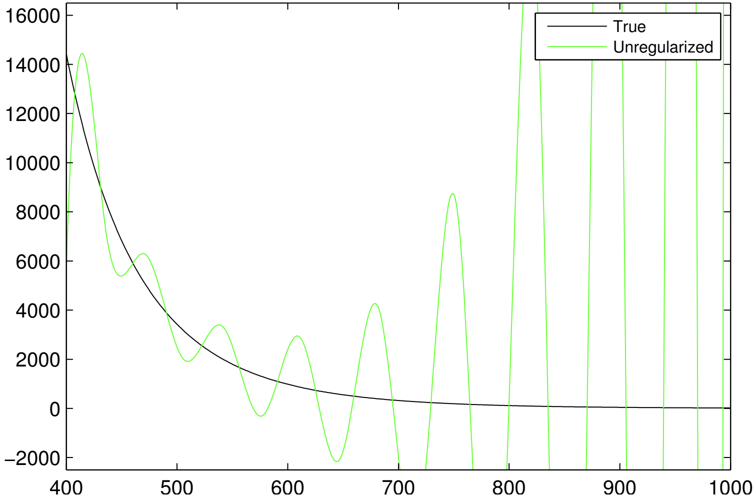
Current methods:

- How to choose the regularization parameter?
- The uncertainties can suffer from serious undercoverage
- There is a systematic uncertainty from the MC prediction which is extremely difficult to quantify rigorously
- Correlations make the unfolded graphical display difficult to interpret

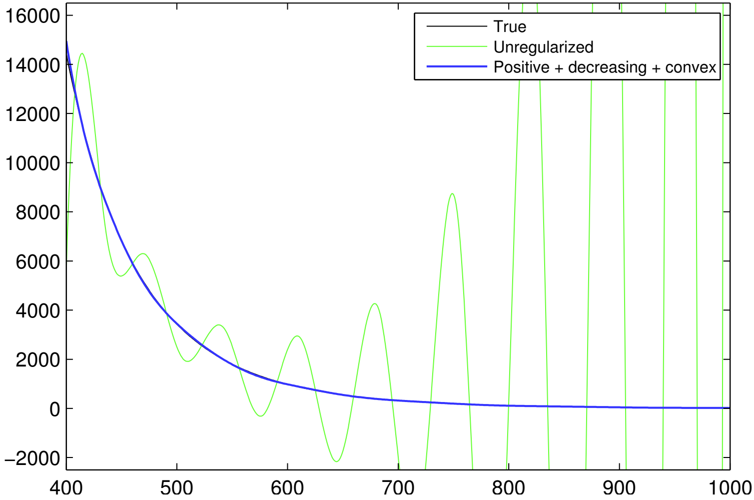
Shape-constrained unfolding:

- There is no regularization parameter to choose
- The uncertainties have guaranteed frequentist finite-sample coverage
- No need for a MC prediction and hence no associated systematic uncertainty
- The resulting simultaneous confidence envelope has a direct interpretation

Regularization using shape constraints



Regularization using shape constraints



Semi-discrete forward mapping

- The response matrix **K** depends on the MC prediction because of the discretization of the unfolded space
- We remove this dependence by considering a semi-discrete forward mapping, where only the smeared space is discretized
- Recall the continuous forward model $g(t) = \int_E k(t,s) f(s)\,\mathrm{d}s$
- As before, we assume histogrammed observations $\mathbf{y}$ with mean vector

$$\boldsymbol{\mu} = \left[\int_{F_1} g(t)\,\mathrm{d}t, \ldots, \int_{F_n} g(t)\,\mathrm{d}t \right]^{\mathrm{T}}$$

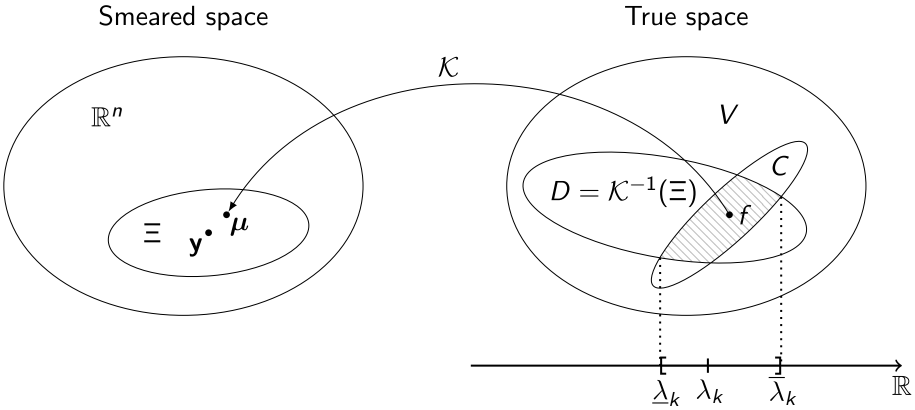
- The elements of $\boldsymbol{\mu}$ are given by

$$\mu_j = \int_{F_j} g(t)\,\mathrm{d}t = \int_{F_j} \int_E k(t,s) f(s)\,\mathrm{d}s\,\mathrm{d}t = \int_E k_j(s) f(s)\,\mathrm{d}s := K_j f$$

  with $k_j(s) = \int_{F_j} k(t,s)\,\mathrm{d}t$

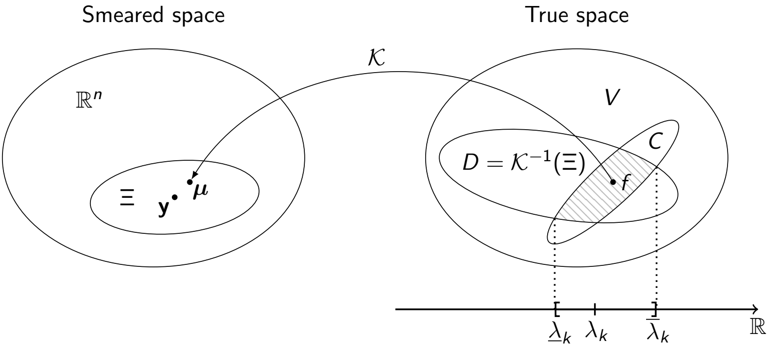
- The semi-discrete forward mapping $\mathcal{K}$ is then given by $\boldsymbol{\mu} = \mathcal{K}f = [K_1 f, \ldots, K_n f]^{\mathrm{T}}$
- The statistical model becomes $\mathbf{y} \sim \mathrm{Poisson}(\boldsymbol{\mu}),$ with $\boldsymbol{\mu} = \mathcal{K}f$
- Aim: Infer $\boldsymbol{\lambda} = \left[\int_{E_1} f(s)\,\mathrm{d}s, \ldots, \int_{E_p} f(s)\,\mathrm{d}s \right]^{\mathrm{T}} = [H_1 f, \ldots, H_p f]^{\mathrm{T}}$ with $H_k : f \mapsto \int_{E_k} f(s)\,\mathrm{d}s$

Strict bounds confidence intervals (Stark, 1992)



$$\lambda_k = H_k f = \int_{E_k} f(s)\,\mathrm{d}s, \quad \underline{\lambda}_k = \min_{f \in C \cap D} H_k f, \quad \overline{\lambda}_k = \max_{f \in C \cap D} H_k f$$

Strict bounds confidence intervals (Stark, 1992)



$$
\begin{aligned}
\mathrm{P}_f(\boldsymbol{\mu} \in \Xi) \geq 1 - \alpha \quad & \overset{(1)}{\Rightarrow} \quad \mathrm{P}_f(f \in D) \geq 1 - \alpha \\
& \Rightarrow \quad \mathrm{P}_f(f \in C \cap D) \geq 1 - \alpha \\
& \overset{(2)}{\Rightarrow} \quad \mathrm{P}_f(\boldsymbol{\lambda} \in [\underline{\lambda}_1, \overline{\lambda}_1] \times \cdots \times [\underline{\lambda}_p, \overline{\lambda}_p]) \geq 1 - \alpha
\end{aligned}
$$

(1) $\mathrm{P}_f(\boldsymbol{\mu} \in \Xi) = \mathrm{P}_f(f \in \mathcal{K}^{-1}(\Xi)) = \mathrm{P}_f(f \in D)$
(2) $\mathrm{P}_f(\boldsymbol{\lambda} \in [\underline{\lambda}_1, \overline{\lambda}_1] \times \cdots \times [\underline{\lambda}_p, \overline{\lambda}_p]) \geq \mathrm{P}_f(f \in C \cap D)$

Shape-constrained strict bounds

- Hence the problem reduces to solving the optimization problems

$$\min_{f \in C \cap D} H_k f \quad \text{and} \quad \max_{f \in C \cap D} H_k f$$

- We derive a conservative solution for the following shape constraints C:
  1. f positive $\Rightarrow$ finite-dimensional linear program
  2. f positive and decreasing $\Rightarrow$ finite-dimensional linear program
  3. f positive, decreasing and convex $\Rightarrow$ finite-dimensional program with a linear objective and nonlinear constraints
- This enables us to compute simultaneous confidence intervals for $\boldsymbol{\lambda}$
  - The coverage of the intervals is guaranteed for known smearing kernel k and for true f satisfying the shape constraints
  - The coverage probability will generally be larger than the nominal value, but the interval lengths are still orders of magnitude shorter than those of unregularized intervals
- Details of the construction described in arXiv:1512.00905

Demonstration: Inclusive jet p_{T} spectrum

- We demonstrate shape-constrained unfolding using the inclusive jet transverse momentum spectrum
- Let the true spectrum be (CMS Collaboration, 2011)

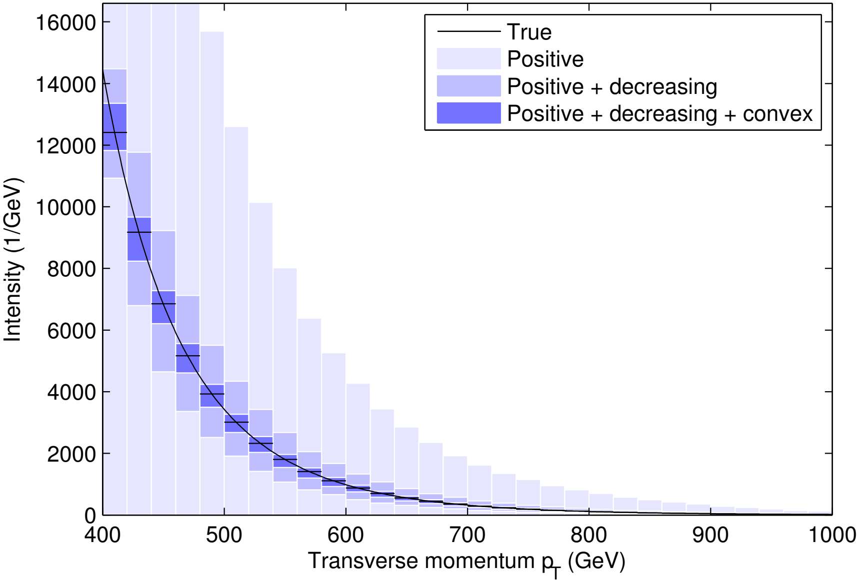
$$f(p_{\mathrm{T}}) = LN_0 \left(\frac{p_{\mathrm{T}}}{\mathrm{GeV}}\right)^{-\alpha} \left(1 - \frac{2}{\sqrt{s}} p_{\mathrm{T}} \cosh(y_{\mathrm{min}})\right)^{\beta} e^{-\gamma/p_{\mathrm{T}}}$$

with $L = 5.1\,\mathrm{fb}^{-1}, \sqrt{s} = 7000\,\mathrm{GeV}, N_0 = 10^{17}\,\mathrm{fb/GeV}, \gamma = 10\,\mathrm{GeV}$, $\alpha = 5, \beta = 10$ and $y_{\mathrm{min}} = 0$

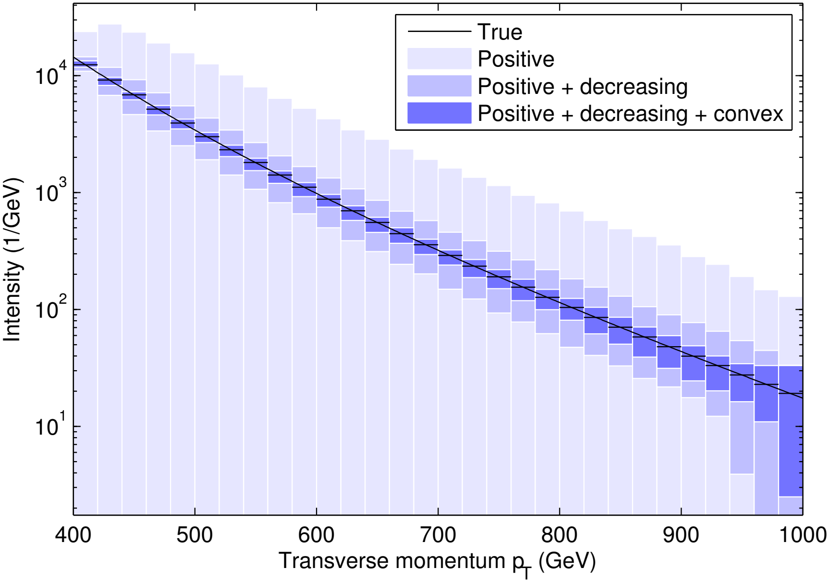
- We generate the smeared data by convolving this with the calorimeter resolution $N(0, \sigma(p_{\mathrm{T}})^2)$ where

$$\sigma(p_{\mathrm{T}}) = p_{\mathrm{T}} \sqrt{\frac{N^2}{p_{\mathrm{T}}^2} + \frac{S^2}{p_{\mathrm{T}}} + C^2}, \quad N = 1\,\mathrm{GeV}, S = 1\,\mathrm{GeV}^{1/2}, C = 0.05$$

Demonstration: Inclusive jet p_T spectrum

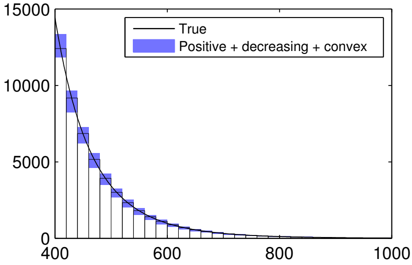


Demonstration: Inclusive jet p_T spectrum



Use of the unfolded simultaneous confidence intervals

- In my view, the best way of communicating the unfolded results is a simultaneous confidence envelope in the unfolded space
- This confidence envelope has a direct physical interpretation
    - The envelope contains the true $\boldsymbol{\lambda}$, whatever it may be, at least 95% of the time under repeated sampling



- The envelope can be used to perform a goodness-of-fit test of a new theory prediction by simply overlaying the prediction on the figure
    - If the prediction is contained within the envelope then it is consistent with the experimental data
    - If the prediction is outside the envelope at any bin, then it is rejected at 5% significance level
- The envelope can be used for propagating the unfolded measurements to further analyses (after an appropriate multiplicity correction)
    - Algorithms for doing this are not yet there, but can be developed

Conclusions and outlook

- We have presented a method that provides guaranteed coverage (for known k) in a large class of unfolding problems
- This is in contrast with the current state-of-the-art unfolding methods that can demonstrably fail to achieve nominal coverage
- Provides rigorous guarantees without having to worry about the choice of the regularization strength, MC dependence, unfolding biases,...
- Many extensions possible:
  - Uncertainty in $k(t, s)$
  - Unimodality constraint
  - Combination of several unfolded measurements
  - Propagation of unfolded measurements to e.g. PDFs
  - Optimality considerations
- Further details available in:

Kuusela, M. and Stark P. B. (2015). Shape-constrained uncertainty quantification in unfolding steeply falling elementary particle spectra. arXiv:1512.00905 [stat.AP].

- The code for the simulation study available at:

`https://github.com/mkuusela/ShapeConstrainedUnfolding`