Unfolding Techniques: A Statistician's Perspective

Mikael Kuusela

Department of Statistics,
University of Chicago

State of the Nu-tion Workshop

Toronto, Canada

June 24, 2017

The unfolding problem

- Any differential cross section 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
- Ill-posed inverse problem with major methodological challenges

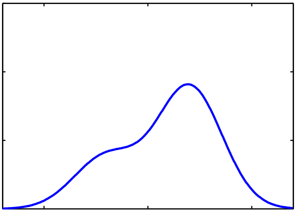
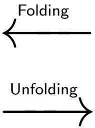


Figure: Smeared spectrum



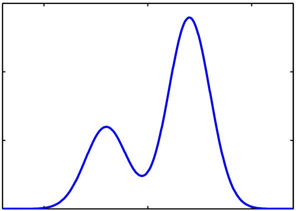


Figure: True spectrum

Problem formulation

- Let f be the true, particle-level spectrum and g the smeared, detector-level spectrum
  - Denote the true space by T and the smeared space by S (both taken to be 1D 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

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

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

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

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 $\{T_i\}_{i=1}^p$ and $\{S_i\}_{i=1}^n$ be binnings of the true space T and the smeared space S
- Smeared histogram $\boldsymbol{y} = [y_1, \ldots, y_n]^{\mathrm{T}}$ with mean

$$\boldsymbol{\mu} = \left[\int_{S_1} g(s)\,\mathrm{d}s, \ldots, \int_{S_n} g(s)\,\mathrm{d}s \right]^{\mathrm{T}}$$

- Quantity of interest:

$$\boldsymbol{\lambda} = \left[\int_{T_1} f(t)\,\mathrm{d}t, \ldots, \int_{T_p} f(t)\,\mathrm{d}t \right]^{\mathrm{T}}$$

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

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

- The discretized statistical model becomes

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

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

Discretization

- Problem primarily discretized using histograms (splines are also sometimes used)
- Let $\{T_i\}_{i=1}^p$ and $\{S_i\}_{i=1}^n$ be binnings of the true space T and the smeared space S
- Smeared histogram $\boldsymbol{y} = [y_1, \ldots, y_n]^{\mathrm{T}}$ with mean

$$\boldsymbol{\mu} = \left[\int_{S_1} g(s)\,\mathrm{d}s, \ldots, \int_{S_n} g(s)\,\mathrm{d}s\right]^{\mathrm{T}}$$

- Quantity of interest:

$$\boldsymbol{\lambda} = \left[\int_{T_1} f(t)\,\mathrm{d}t, \ldots, \int_{T_p} f(t)\,\mathrm{d}t\right]^{\mathrm{T}}$$

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

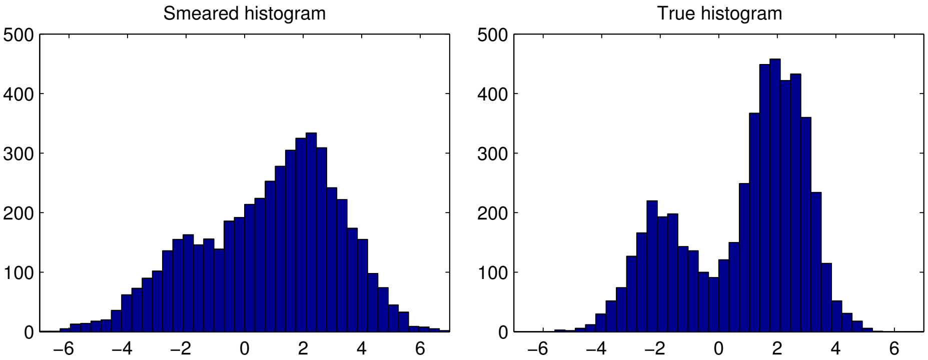
$$K_{i,j} = \frac{\int_{S_i}\int_{T_j} k(s,t) f^{\mathrm{MC}}(t)\,\mathrm{d}t\,\mathrm{d}s}{\int_{T_j} f^{\mathrm{MC}}(t)\,\mathrm{d}t} = P(\text{smeared event in bin } i \,|\, \text{true event in bin } j)$$

- The discretized statistical model becomes

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

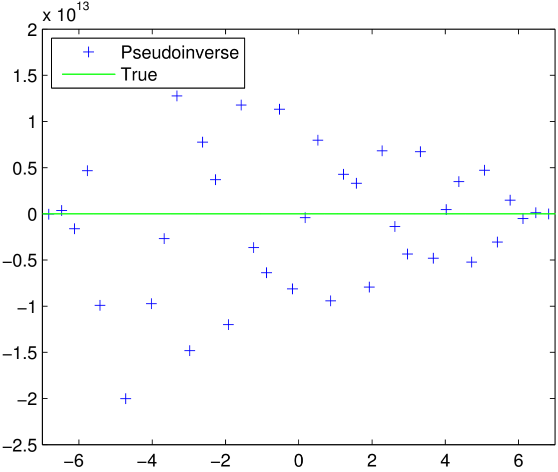
where $\boldsymbol{K}$ is an ill-conditioned matrix

Demonstration of ill-posedness

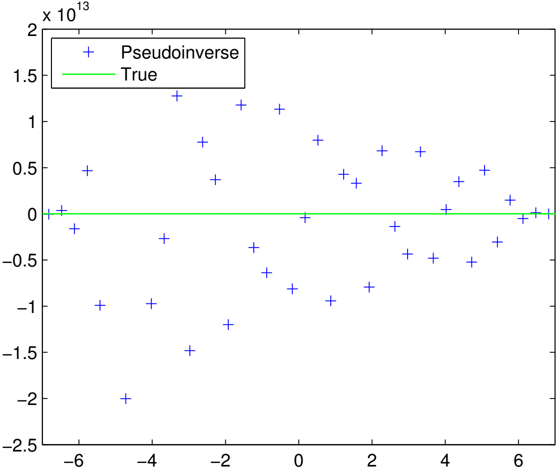


$$\boldsymbol{\mu} = \boldsymbol{K}\boldsymbol{\lambda}, \quad \boldsymbol{y} \sim \mathrm{Poisson}(\boldsymbol{\mu}) \quad \overset{??}{\Longrightarrow} \quad \hat{\boldsymbol{\lambda}} = \boldsymbol{K}^{-1}\boldsymbol{y}$$

Demonstration of ill-posedness



Demonstration of ill-posedness



$$\mathrm{MSE}(\hat{\theta}) = \mathsf{E}((\hat{\theta} - \theta)^2) = [\mathrm{bias}(\hat{\theta})]^2 + \mathsf{var}(\hat{\theta})$$

Regularization: bias $\uparrow$, variance $\downarrow$ $\Rightarrow$ $\mathrm{MSE} \downarrow$

Current methodology

Two main approaches to unfolding:

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:

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

- 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 $\boldsymbol{\lambda}^{(0)}$
  - Regularization strength controlled by the choice of K
- 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 $\boldsymbol{y} \sim \mathrm{Poisson}(\boldsymbol{K\lambda})$
- 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, $k = 0$

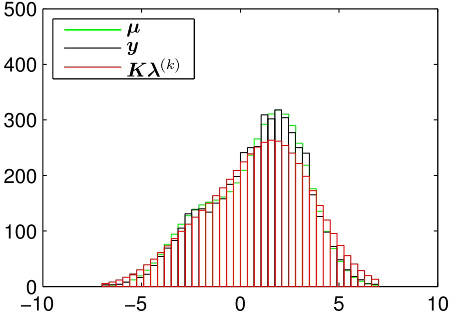


Figure: Smeared histogram

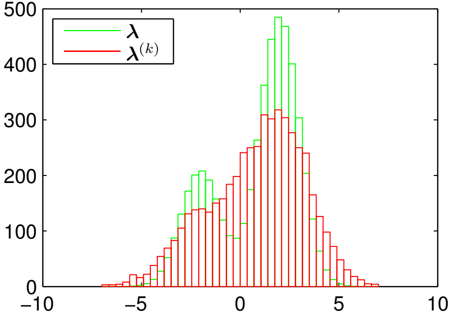


Figure: True histogram

D'Agostini demo, $k = 100$

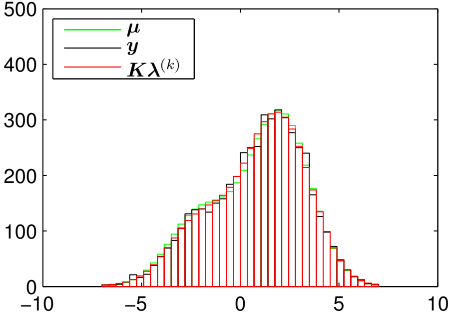


Figure: Smeared histogram

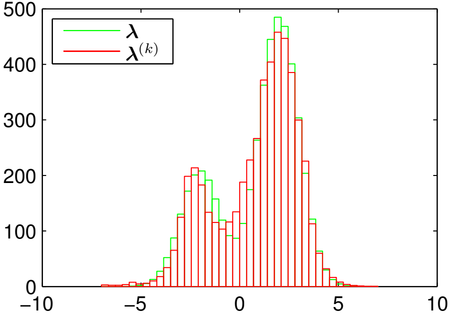


Figure: True histogram

D'Agostini demo, $k = 10000$

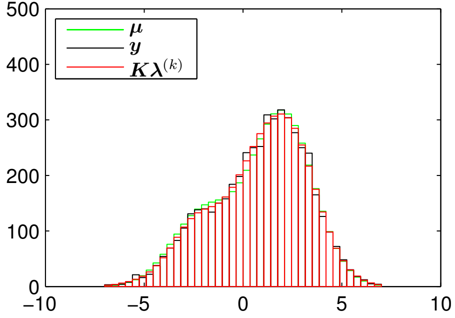


Figure: Smeared histogram

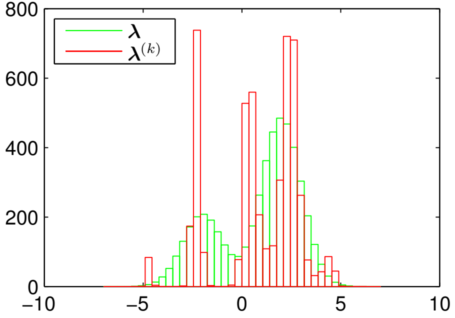


Figure: True histogram

D'Agostini demo, $k = 100000$

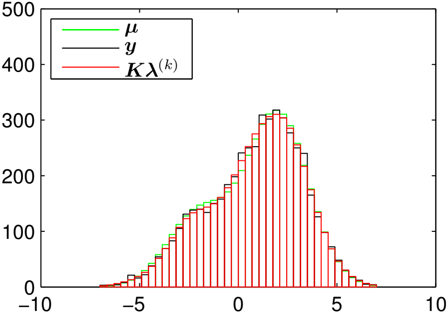


Figure: Smeared histogram

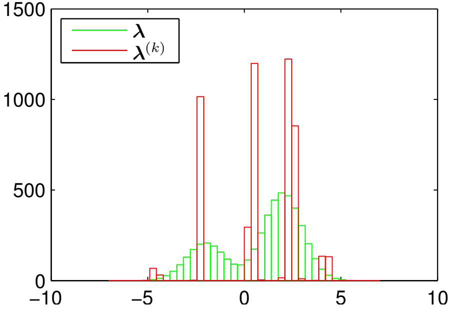


Figure: True histogram

Choice of the regularization strength

- A key issue in unfolding concerns the choice of the regularization strength (δ in Tikhonov, K in D'Agostini)
  - This choice has a major impact on the unfolded solution!
- To avoid MC dependence, the choice should ideally be done using a data-driven technique instead of MC studies
- 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)
- Note: All these are aiming to achieve optimal point estimation
  - Not necessarily optimal for uncertainty quantification!

Some remarks based on experience from the LHC

- One should think carefully if unfolding is *really* needed
  - E.g., if the goal of 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, cross section ratios, 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 $\boldsymbol{K}$

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

Some remarks based on experience from the LHC

- 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
  - In particular, the default 4 iterations for D'Agostini in `RooUnfold` is just an arbitrary choice and does not guarantee a good solution
- 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

Some remarks based on experience from the LHC

- `TUnfold` is more versatile and better documented than `RooUnfold`
  - In particular, `TUnfold` allows one to change the bias vector $\boldsymbol{\lambda}^{\mathrm{MC}}$, while in `RooUnfold` it is fixed to the same MC that is used to construct the response matrix
- One cannot simply do away with ill-posedness by using wider bins
  - The wider the bins, the more dependent the response matrix $\boldsymbol{K}$ becomes on the assumed shape of the spectrum inside the bins
- 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...

Uncertainty quantification in unfolding

- Aim: Find random intervals $\left[\underline{\lambda}_i(\boldsymbol{y}), \overline{\lambda}_i(\boldsymbol{y})\right]$, $i = 1, \ldots, p$, with *coverage probability* $1 - \alpha$:

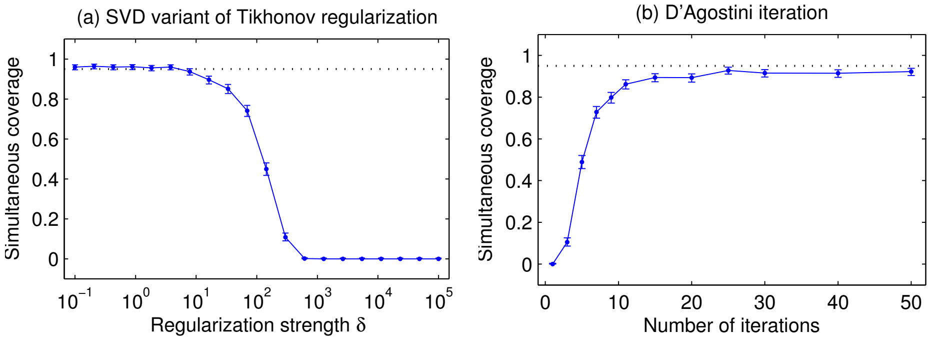
$$\mathrm{P}_f\left(\lambda_i \in \left[\underline{\lambda}_i(\boldsymbol{y}), \overline{\lambda}_i(\boldsymbol{y})\right]\right) \geq 1 - \alpha, \quad \forall f$$

- Current methods quantify the uncertainty using the binwise variance (estimated using either error propagation or resampling):

$$\left[\underline{\lambda}_i, \overline{\lambda}_i\right] = \left[\hat{\lambda}_i - z_{1-\alpha/2}\sqrt{\widehat{\mathrm{var}}\left(\hat{\lambda}_i\right)}, \, \hat{\lambda}_i + z_{1-\alpha/2}\sqrt{\widehat{\mathrm{var}}\left(\hat{\lambda}_i\right)}\right]$$

- But: These intervals may suffer from significant undercoverage since they ignore the regularization bias
- Two ways to obtain improved coverage performance:
  1. Debiased intervals (Kuusela and Panaretos, 2015; Kuusela, 2016)
  2. Shape-constrained intervals (Kuusela and Stark, 2017; Kuusela, 2016)

Undercoverage of existing methods (Kuusela, 2016)

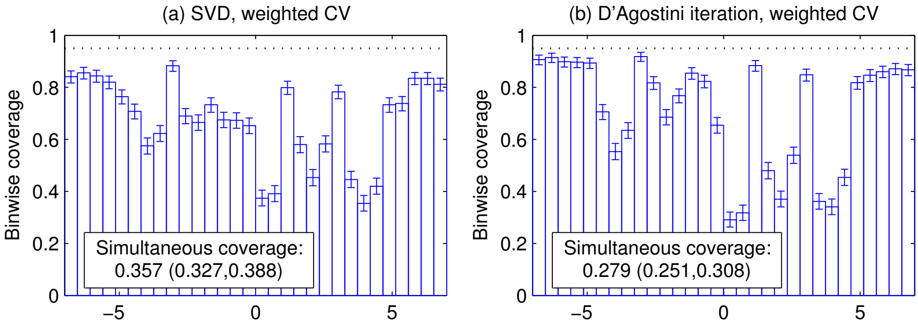


$$f(t) = \lambda_{\mathrm{tot}} \left\{ \pi_1 \mathcal{N}(t|-2,1) + \pi_2 \mathcal{N}(t|2,1) + \pi_3 \frac{1}{|E|} \right\}$$

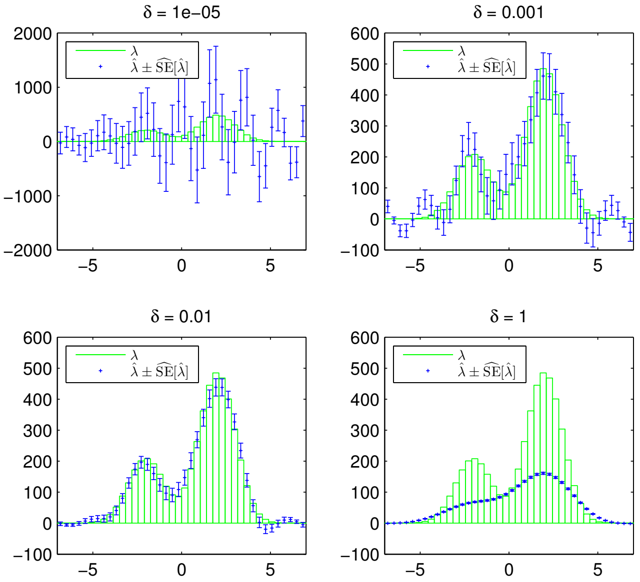
$$g(s) = \int_T \mathcal{N}(s-t|0,1) f(t)\,\mathrm{d}t$$

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

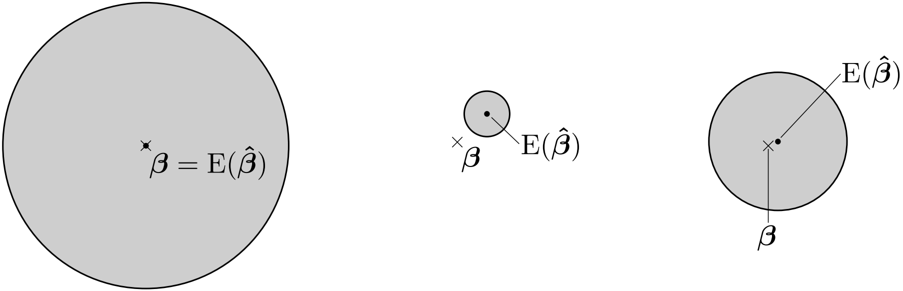
Undercoverage of existing methods (Kuusela, 2016)



UQ for Tikhonov regularization, $P(\boldsymbol{\lambda}) = \|\boldsymbol{\lambda}\|^2$



Improved uncertainty quantification by undersmoothing



Unbiased, coverage $= 1 - \alpha$

Optimal point estimation, coverage $\ll 1 - \alpha$

Optimal UQ, coverage $= 1 - \alpha - \varepsilon$

Undersmoothed UQ for unfolding (Kuusela, 2016)

1. Choose pilot estimate of δ using one of the standard data-driven methods (CV, MMLE, L-curve,...)
2. Reduce δ until intervals to have estimated coverage $1 - \alpha - \varepsilon$, for some small tolerance ε

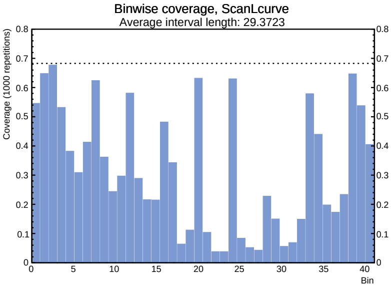
Coverage study

Method	Coverage at $t = 0$	Mean length
BC (data)	0.932 *(0.915, 0.947)*	0.079 *(0.077, 0.081)*
BC (oracle)	0.937 *(0.920, 0.951)*	0.064 *(0.064, 0.064)*
US (data)	0.933 *(0.916, 0.948)*	0.091 *(0.087, 0.095)*
US (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

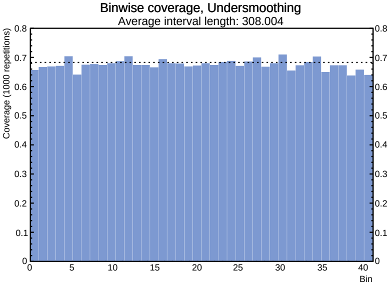
BC = iterative bias-correction
US = undersmoothing
MMLE = choose δ to maximize the marginal likelihood
MISE = choose δ to minimize the mean integrated squared error

This and further simulation studies in Kuusela (2016) show that data-driven debiasing performs robustly in many variants of the unfolding problem

TUnfold with data-driven undersmoothing

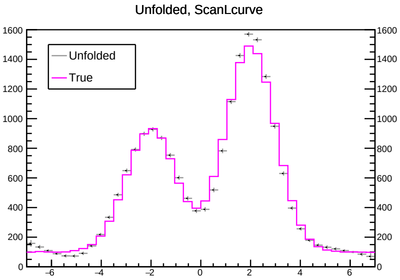


Coverage when δ chosen using L-curve.

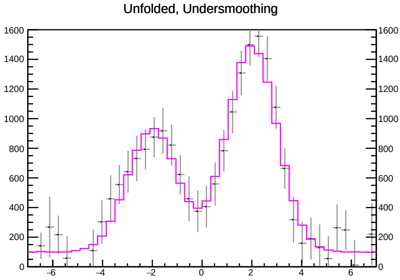


Coverage when δ undersmoothed to give 67 % coverage.

TUnfold with data-driven undersmoothing

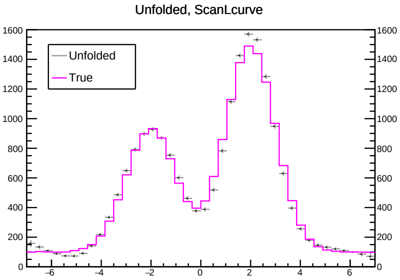


Intervals from L-curve; severe undercoverage.

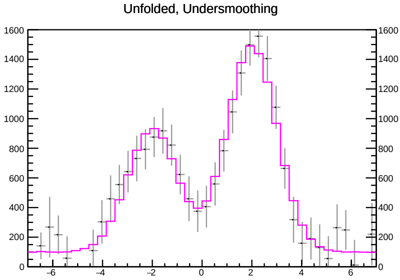


Undersmoothed intervals; nearly nominal coverage.

TUnfold with data-driven undersmoothing



Intervals from L-curve; severe undercoverage.



Undersmoothed intervals; nearly nominal coverage.

Undersmoothing code (modification of `TUnfold` by Lyle Kim and myself) available at: `https://github.com/lylejkim/UndersmoothedUnfolding`

Note: This is an early version, any feedback is most welcome!

Summary and conclusions

- Unfolding is a complex data analysis task with many potential pitfalls
  - It is crucial to understand all the ingredients that go into an unfolding procedure
- Tikhonov regularization and D'Agostini are the two most popular techniques
  - Personally I find it easier to interpret the 2nd derivative penalty in Tikhonov than the early stopping in D'Agostini
- Results from standard software (`RooUnfold`) depend strongly on the MC prediction (results biased towards the MC)
  - Safer to use MC-independent regularization (possible in `TUnfold`)
- Proper choice of the regularization strength is crucial
  - A choice that is optimal for point estimation might not be optimal for uncertainty quantification
- Statistical uncertainties from standard techniques can be unreliable
  - Improved uncertainty quantification can be achieved by debiasing
  - Data-driven undersmoothing now available for `ROOT`!
- Many open statistical issues remain:
  - How to properly present unfolded results? (Bins are correlated)
  - How to properly deal with systematics in the detector response?
  - How to properly compare, combine and propagate unfolded results?

References I

T. Adye. Unfolding algorithms and tests using RooUnfold. In H. B. Prosper and L. Lyons, 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, 17–20 January 2011.

ATLAS Collaboration, Measurement of the transverse momentum distribution of W bosons in *pp* collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector, *Physical Review D*, 85:012005, 2012.

G. Backus, Inference from inadequate and inaccurate data, I, *Proceedings of the National Academy of Sciences*, 65(1):1–7, 1970.

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

CMS Collaboration, Energy calibration and resolution of the CMS electromagnetic calorimeter in *pp* collisions at $\sqrt{s} = 7$ TeV, *Journal of Instrumentation*, 8(09): P09009, 2013a.

CMS Collaboration, Measurements of differential jet cross sections in proton-proton collisions at $\sqrt{s} = 7\ \mathrm{TeV}$ with the CMS detector, *Physical Review D*, 87:112002, 2013b.

References II

CMS Collaboration, Measurement of differential top-quark-pair production cross sections in *pp* collisions at $\sqrt{s} = 7\ \mathrm{TeV}$, *The European Physical Journal C*, 73:2339, 2013c.

CMS Collaboration, Measurement of differential cross sections for Higgs boson production in the diphoton decay channel in *pp* collisions at $\sqrt{s} = 8\ \mathrm{TeV}$, *The European Physical Journal C*, 76:13, 2016.

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

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

F. Garwood, Fiducial limits for the Poisson distribution, *Biometrika*, 28(3/4):437–442, 1936.

S. Geman and D. E. McClure. Bayesian image analysis: An application to single photon emission tomography. In *Proceedings of the American Statistical Association, Statistical Computing Section*, pages 12–18, 1985.

S. Geman and D. E. McClure, Statistical methods for tomographic image reconstruction, *Bulletin of the International Statistical Institute*, LII(4):5–21, 1987.

References III

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

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

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

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

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

M. Kuusela and P. B. Stark, Shape-constrained uncertainty quantification in unfolding steeply falling elementary particle spectra, *The Annals of Applied Statistics*, to appear, 2017. Preprint: `https://arxiv.org/abs/1512.00905`.

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

References IV

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

D. G. Luenberger. *Optimization by Vector Space Methods*. John Wiley & Sons, 1969.

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

S. S. Saquib, C. A. Bouman, and K. Sauer, ML parameter estimation for Markov random fields with applications to Bayesian tomography, *IEEE Transactions on Image Processing*, 7(7):1029–1044, 1998.

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

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

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

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

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

References V

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

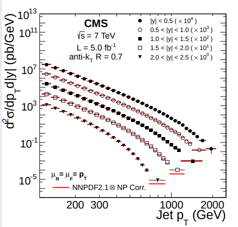
I. Volobouev. On the expectation-maximization unfolding with smoothing. arXiv:1408.6500v2 [physics.data-an], 2015.

S. N. Wood, 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, 2011.

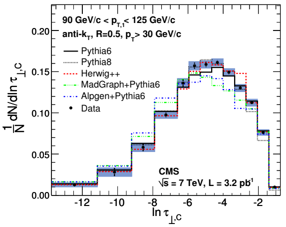
Backup

Examples of unfolding in LHC data analysis

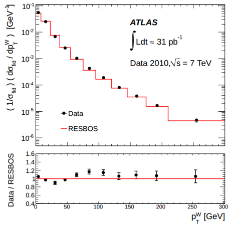
Inclusive jet cross section



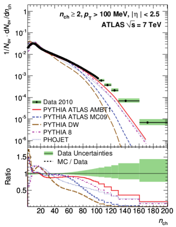
Hadronic event shape



W boson cross section



Charged particle multiplicity



Coverage for Gaussian observations

Proposition

Assume $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{K}\boldsymbol{\beta}, \boldsymbol{\Sigma})$*, where* $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$ *is a known covariance matrix and* $\boldsymbol{K} \in \mathbb{R}^{n \times p}$*. Let* $\hat{\boldsymbol{\beta}} = \boldsymbol{A}\boldsymbol{y}$ *with* $\boldsymbol{A} \in \mathbb{R}^{p \times n}$ *be a linear estimator of* $\boldsymbol{\beta}$ *and let* $\hat{\theta} = \boldsymbol{c}^{\mathrm{T}}\hat{\boldsymbol{\beta}}$ *be the corresponding estimator of the quantity of interest* $\theta = \boldsymbol{c}^{\mathrm{T}}\boldsymbol{\beta}$*. Then the confidence interval*

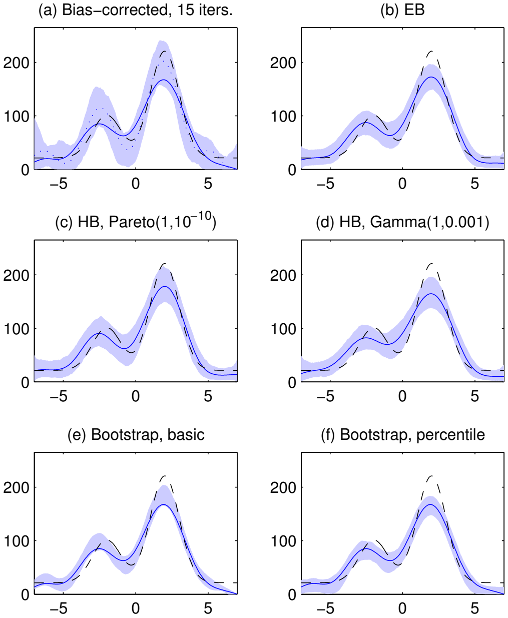
$$\left[\,\underline{\theta}, \overline{\theta}\,\right] = \left[\hat{\theta} - z_{1-\alpha/2}\sqrt{\mathrm{var}\,(\hat{\theta})},\, \hat{\theta} + z_{1-\alpha/2}\sqrt{\mathrm{var}\,(\hat{\theta})}\right]$$
$$= \left[\hat{\theta} - z_{1-\alpha/2}\sqrt{\boldsymbol{c}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{c}},\, \hat{\theta} + z_{1-\alpha/2}\sqrt{\boldsymbol{c}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{c}}\right]$$

has coverage probability

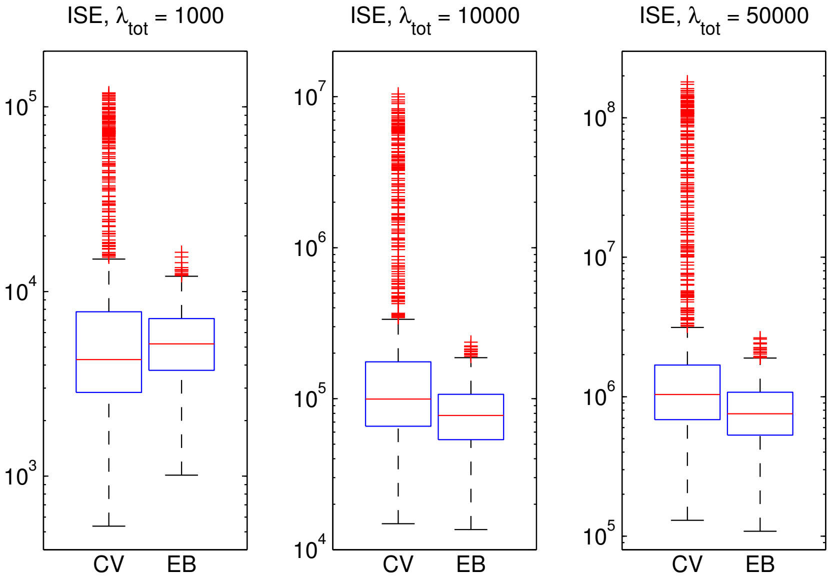
$$\mathrm{P}_{\boldsymbol{\beta}}\left(\underline{\theta} \le \theta \le \overline{\theta}\right) = \Phi\left(\frac{\mathrm{bias}(\hat{\theta})}{\mathrm{SE}(\hat{\theta})} + z_{1-\alpha/2}\right) - \Phi\left(\frac{\mathrm{bias}(\hat{\theta})}{\mathrm{SE}(\hat{\theta})} + z_{\alpha/2}\right),$$

where $\mathrm{bias}(\hat{\theta}) = \mathrm{E}_{\boldsymbol{\beta}}(\hat{\theta}) - \theta = \boldsymbol{c}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{K} - \boldsymbol{I})\boldsymbol{\beta}$ *is the bias of* $\hat{\theta}$*,* $\mathrm{SE}(\hat{\theta}) = \sqrt{\mathrm{var}\,(\hat{\theta})} = \sqrt{\boldsymbol{c}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{c}}$ *is the standard error of* $\hat{\theta}$ *and* Φ *is the standard normal cumulative distribution function.*

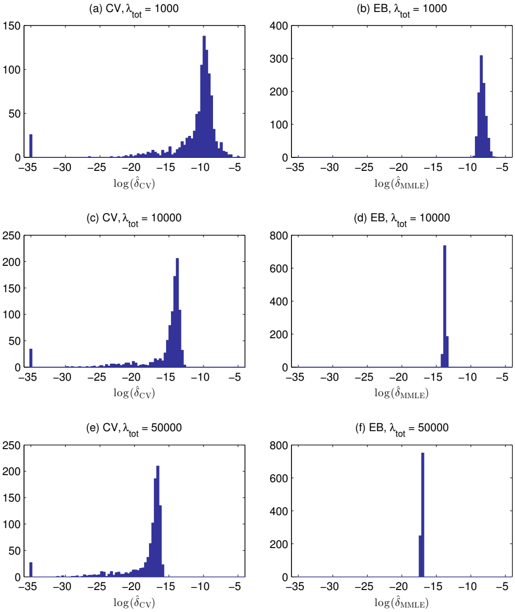
Comparison of intervals, $\lambda_{\text{tot}} = 1\,000$



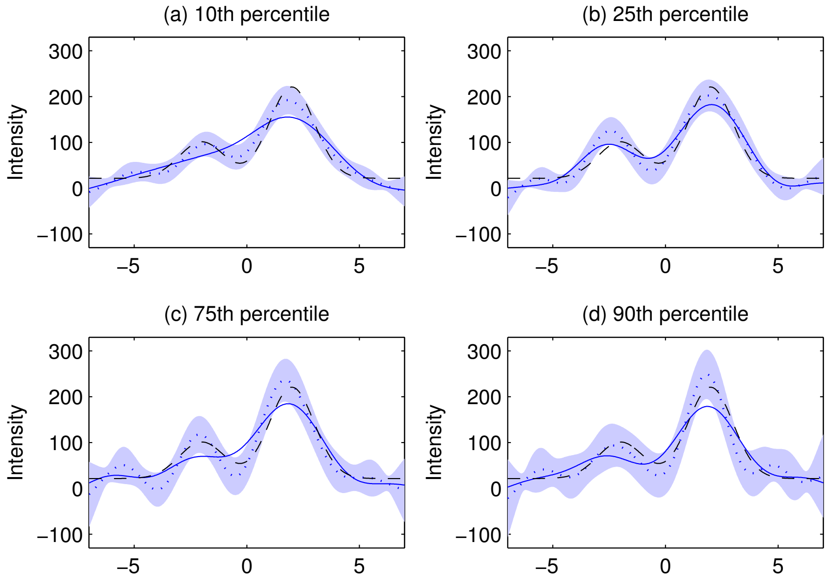
CV vs. EB


ISE, λ_{tot} = 1000
ISE, λ_{tot} = 10000
ISE, λ_{tot} = 50000
10^5
10^4
10^3
10^7
10^6
10^5
10^4
10^8
10^7
10^6
10^5
CV
EB
CV
EB
CV
EB

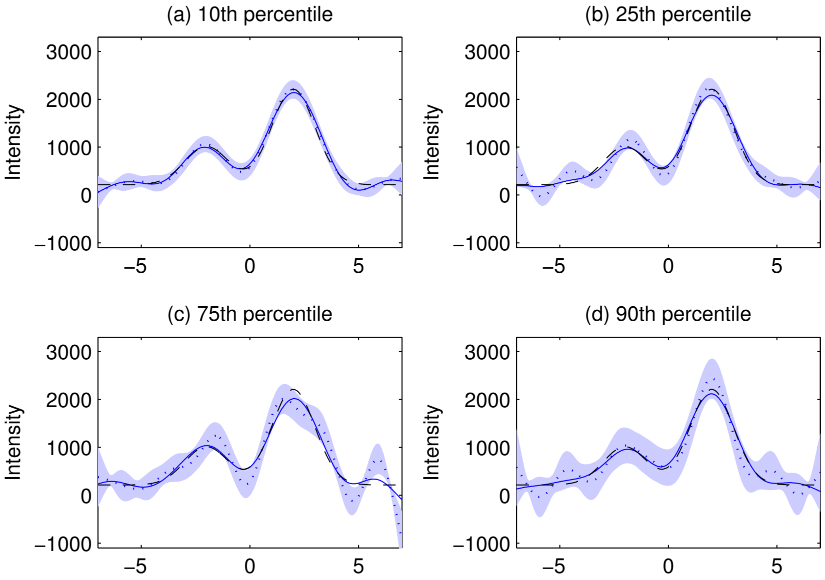
CV vs. EB


(a) CV, λtot = 1000
150
100
50
0
−35 −30 −25 −20 −15 −10 −5
log(δ̂CV)
(b) EB, λtot = 1000
400
300
200
100
0
−35 −30 −25 −20 −15 −10 −5
log(δ̂MMLE)
(c) CV, λtot = 10000
250
200
150
100
50
0
−35 −30 −25 −20 −15 −10 −5
log(δ̂CV)
(d) EB, λtot = 10000
800
600
400
200
0
−35 −30 −25 −20 −15 −10 −5
log(δ̂MMLE)
(e) CV, λtot = 50000
250
200
150
100
50
0
−35 −30 −25 −20 −15 −10 −5
log(δ̂CV)
(f) EB, λtot = 50000
800
600
400
200
0
−35 −30 −25 −20 −15 −10 −5
log(δ̂MMLE)

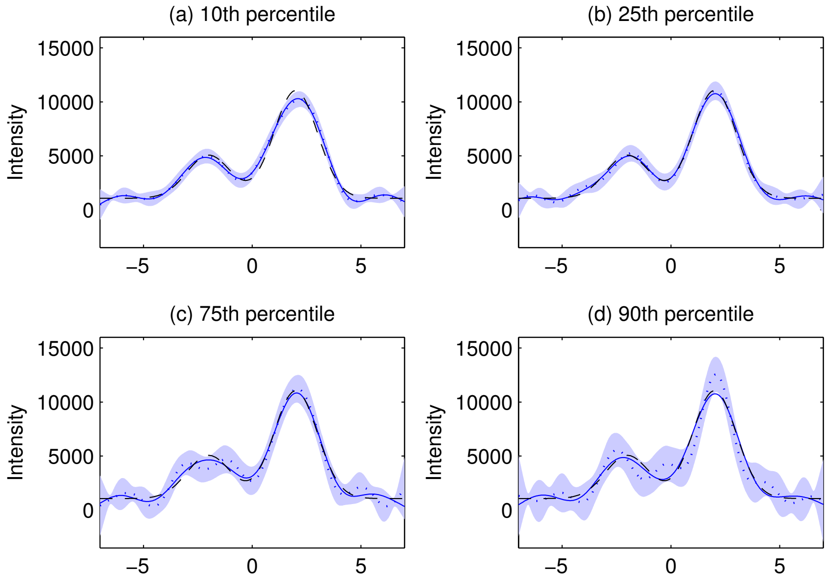
Variability of interval lengths, $\lambda_{\mathrm{tot}} = 1\,000$



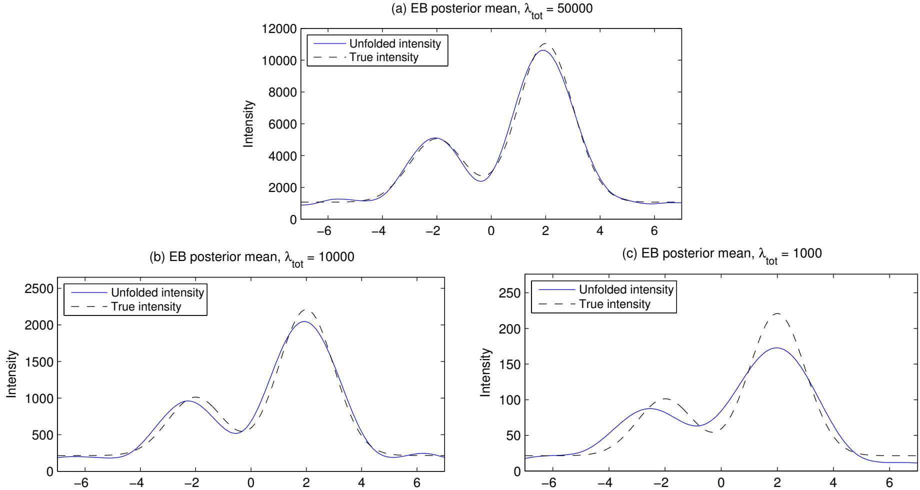
Variability of interval lengths, $\lambda_{\text{tot}} = 10\,000$



Variability of interval lengths, $\lambda_{\mathrm{tot}} = 50\,000$

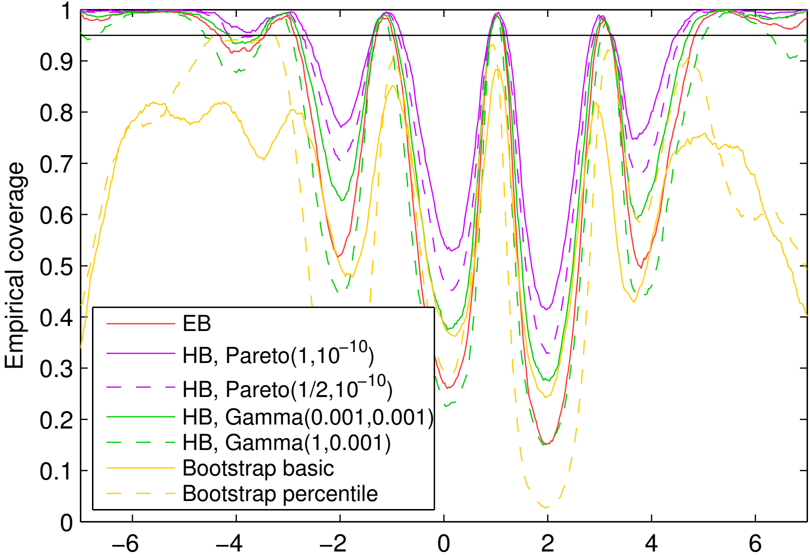


Point estimation demonstration

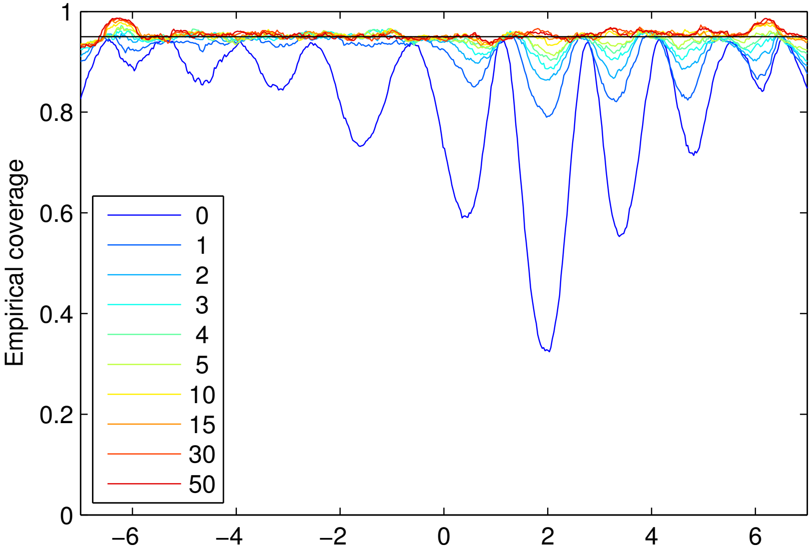


- B-spline basis expansion of f, i.e., $f(s) = \sum_{j=1}^{p} \beta_j B_j(s)$
- Regularization by penalizing $\|f''\|$
- Choice of the regularization strength by marginal maximum likelihood

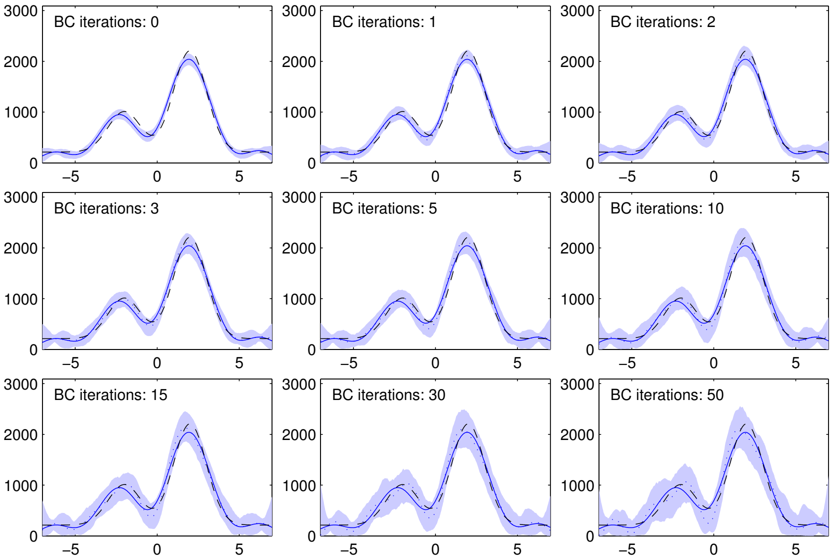
UQ hampered by the bias



Effect of bias-correction on coverage



Effect of bias-correction on interval length



Data-driven debiased confidence intervals

$\Rightarrow$ **Choose the amount of debiasing to calibrate** $1-\alpha$ **intervals to have coverage** $1-\alpha-\varepsilon$

[See Kuusela (2016) for details.]

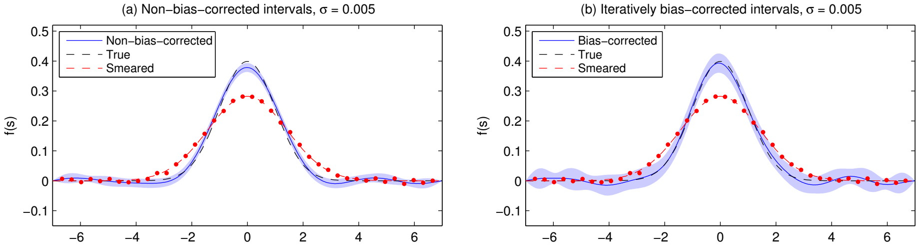
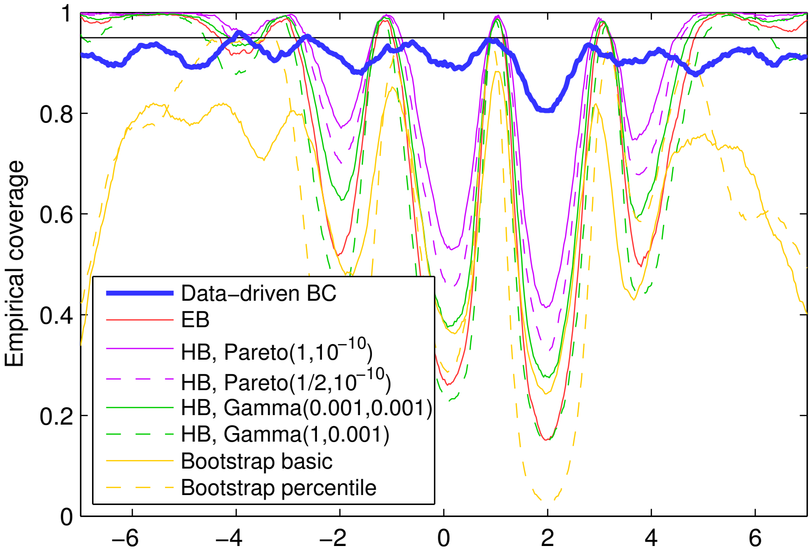
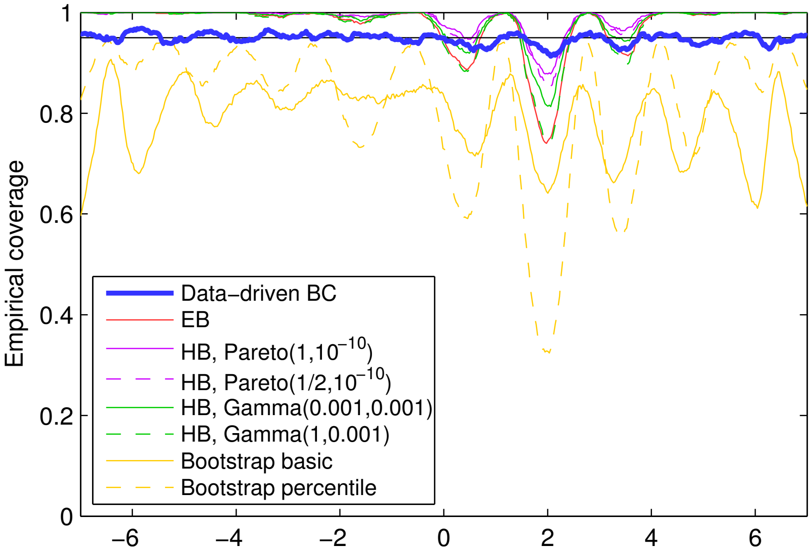


Figure: Gaussian intervals, 95 % nominal coverage, 94 % target coverage

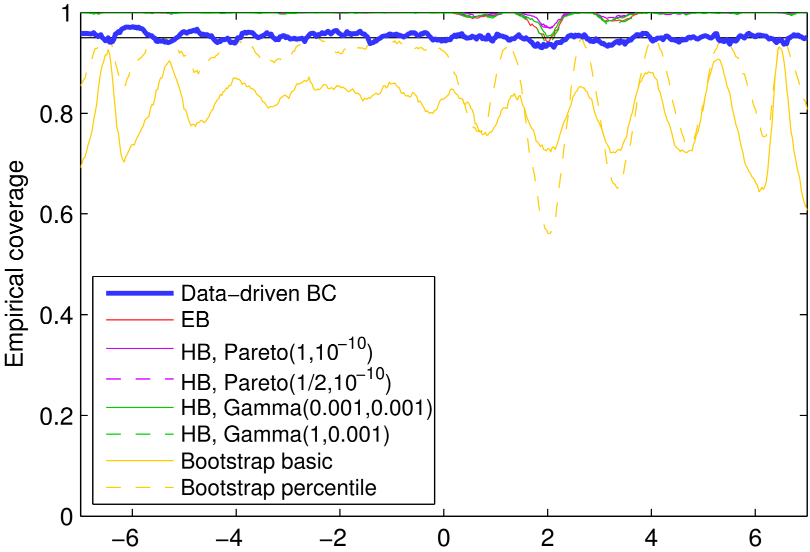
Comparison of coverage performance, $\lambda_{\mathrm{tot}} = 1\,000$



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



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



$Z \to e^+e^-$: Setup

- We demonstrate the proposed approach by unfolding the $Z \to e^+e^-$ invariant mass spectrum measured in the CMS experiment
- The data are published in CMS Collaboration (2013a) and correspond to integrated luminosity of $4.98\ \mathrm{fb}^{-1}$ collected in 2011 at $\sqrt{s} = 7\ \mathrm{TeV}$
- 67 778 "high quality" electron-positron pairs with invariant masses 65–115 GeV in 0.5 GeV bins
- Response: convolution with the Crystal Ball function

$$\mathrm{CB}(m|\Delta m, \sigma^2, \alpha, \gamma) = \begin{cases} Ce^{-\frac{(m-\Delta m)^2}{2\sigma^2}}, & \frac{m-\Delta m}{\sigma} > -\alpha, \\ C\left(\frac{\gamma}{\alpha}\right)^{\gamma} e^{-\frac{\alpha^2}{2}} \left(\frac{\gamma}{\alpha} - \alpha - \frac{m-\Delta m}{\sigma}\right)^{-\gamma}, & \frac{m-\Delta m}{\sigma} \leq -\alpha \end{cases}$$

- CB parameters estimated with maximum likelihood using 30 % of the data ("training data") assuming that the true spectrum is the non-relativistic Breit–Wigner with PDG values for the Z mass and width
  - The remaining 70 % used for unfolding ("test data")

$Z \to e^+e^-$: Unfolding results

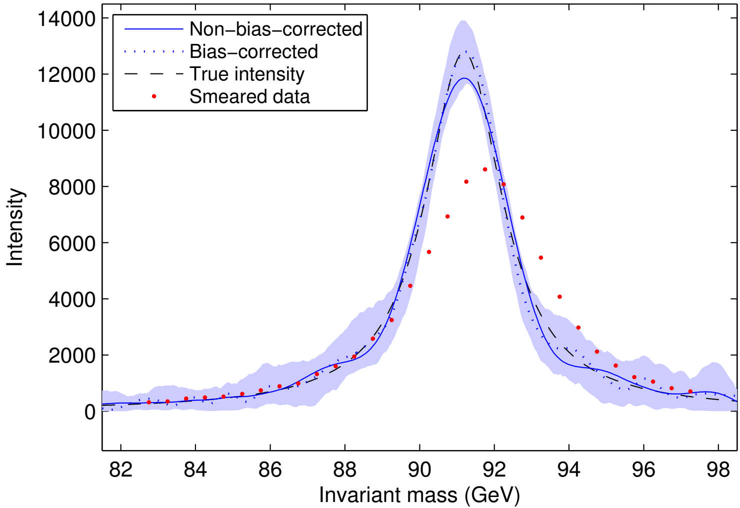


Figure: Z boson invariant mass spectrum, $\hat{N}_{\mathrm{BC}} = 14$, 95 % percentile intervals, 94 % target coverage

Shape-constrained unfolding

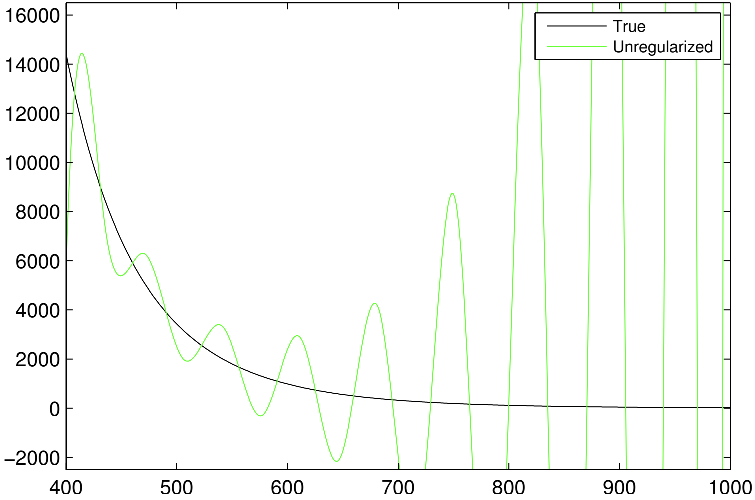
We present a technique for forming confidence intervals for $\boldsymbol{\lambda}$ that have *guaranteed simultaneous frequentist finite-sample coverage*, provided that f satisfies simple, physically justified shape constraints.

The shape constraints (positivity, monotonicity and convexity) are satisfied in the important and common class of unfolding problems with *steeply falling particle spectra*.

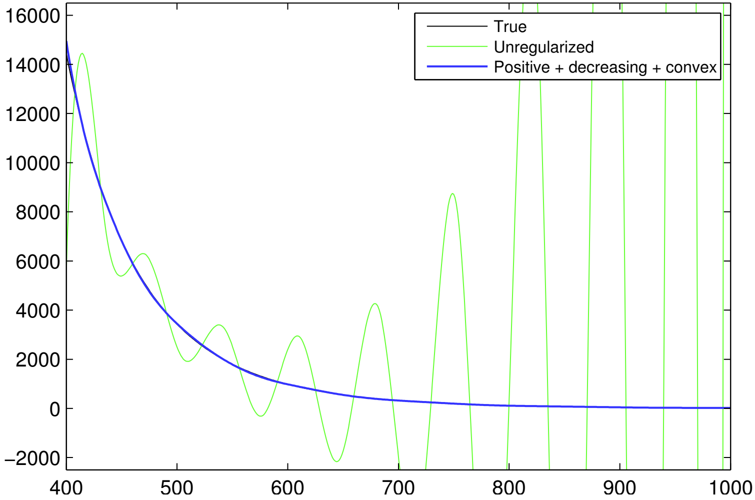
Examples from the LHC include the differential cross sections of:

- Jets (CMS Collaboration, 2013b)
- Top quark pairs (CMS Collaboration, 2013c)
- W boson (ATLAS Collaboration, 2012)
- Higgs boson (CMS Collaboration, 2016)
- ...

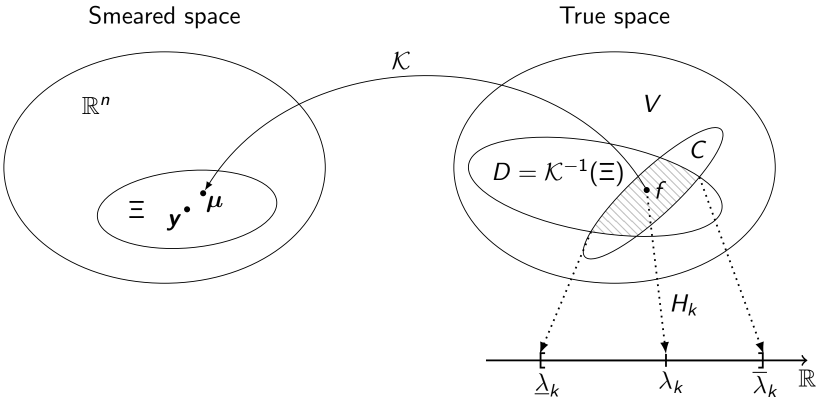
Regularization using shape constraints



Regularization using shape constraints



Strict bounds confidence intervals (Stark, 1992)



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

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

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:
  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 function and nonlinear constraints
- Strategy (Stark, 1992):
  1. Use *Fenchel duality* to turn the infinite-dimensional problem into a semi-infinite program with an n-dimensional unknown and an infinite set of constraints
  2. Discretize the constraints in such a way that the discretized problem is guaranteed to yield a conservative solution
- This enables us to efficiently 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

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)

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

  with $L = 5.1\,\mathrm{fb}^{-1}, \sqrt{s} = 7000\,\mathrm{GeV}, N_0 = 10^{17}\,\mathrm{fb/GeV}, \gamma = 10\,\mathrm{GeV},$ $\alpha = 5$ and $\beta = 10$
- We generate the smeared data by convolving this with the calorimeter resolution $\mathcal{N}(0, \sigma(p_{\mathrm{T}})^2)$, 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

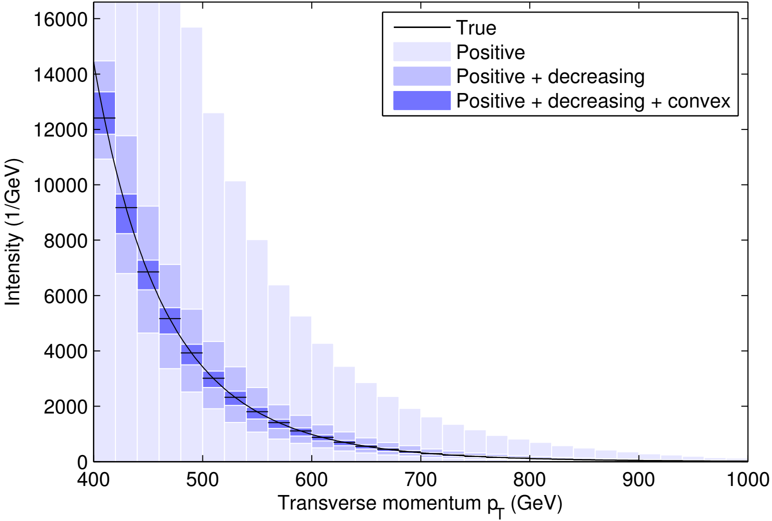


Figure: Shape-constrained unfolded confidence intervals for the inclusive jet p_T spectrum with *guaranteed* conservative 95 % simultaneous coverage.

Demonstration: Inclusive jet p_T spectrum

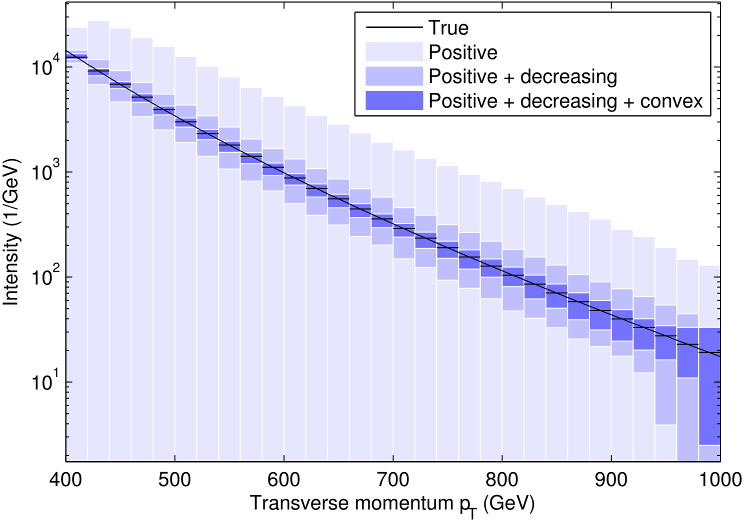
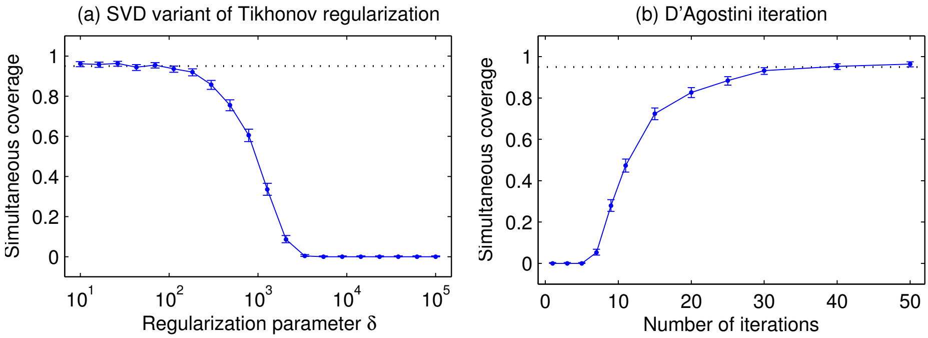
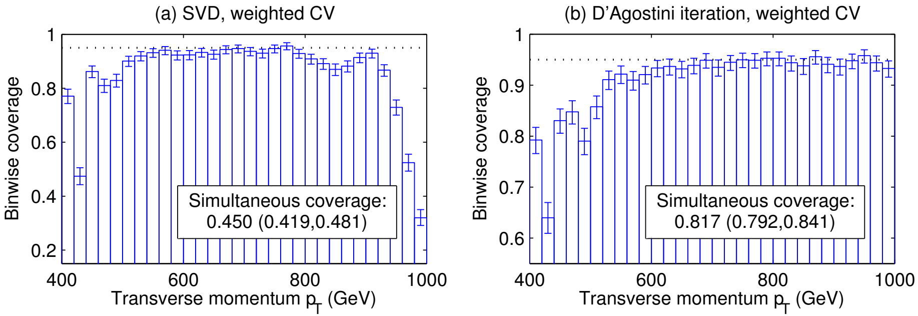


Figure: Shape-constrained unfolded confidence intervals for the inclusive jet p_T spectrum with *guaranteed* conservative 95 % simultaneous coverage.

Demonstration: Inclusive jet p_{T} spectrum

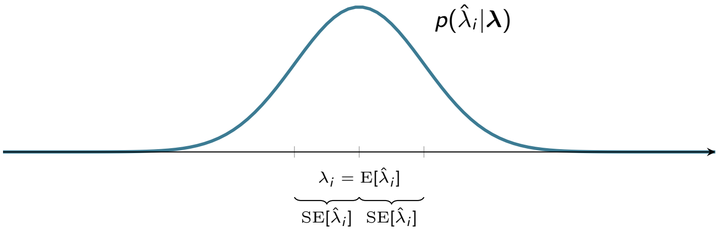


Demonstration: Inclusive jet p_{T} spectrum



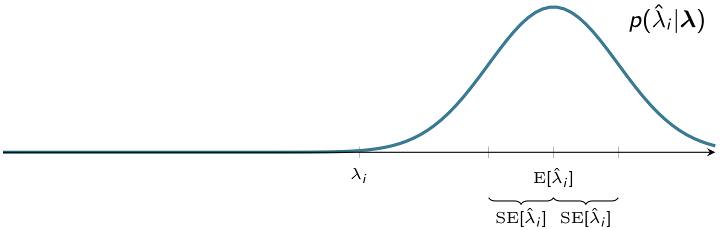
Uncertainty quantification

- Let $\mathrm{SE}[\hat{\lambda}_i]$ be the standard error of $\hat{\lambda}_i$ (i.e., the standard deviation of the sampling distribution of $\hat{\lambda}_i$)
- In many situations, $\hat{\lambda}_i \pm \widehat{\mathrm{SE}}[\hat{\lambda}_i]$ provides a reasonable 68% confidence interval
  - But this is only true when $\hat{\lambda}_i$ is unbiased and approximately Gaussian
- But in regularized unfolding the estimators are always biased!
  - Regularization reduces variance by increasing the bias (*bias-variance trade-off*)
  - Hence the SE confidence intervals may have lousy coverage



Uncertainty quantification

- Let $\mathrm{SE}[\hat{\lambda}_i]$ be the standard error of $\hat{\lambda}_i$ (i.e., the standard deviation of the sampling distribution of $\hat{\lambda}_i$)
- In many situations, $\hat{\lambda}_i \pm \widehat{\mathrm{SE}}[\hat{\lambda}_i]$ provides a reasonable 68% confidence interval
  - But this is only true when $\hat{\lambda}_i$ is unbiased and approximately Gaussian
- But in regularized unfolding the estimators are always biased!
  - Regularization reduces variance by increasing the bias (*bias-variance trade-off*)
  - Hence the SE confidence intervals may have lousy coverage



Problem formulation using Poisson point processes

The appropriate mathematical model for unfolding is that of an *indirectly observed Poisson point process*

Definition

A random point measure M is a *Poisson point process* with intensity function f and state space $E \subset \mathbb{R}$ iff

1. $M(B) \sim \mathrm{Poisson}(\lambda(B))$, where $\lambda(B) = \int_B f(s)\, \mathrm{d}s$, for every (Borel) set $B \subset E$;
2. $M(B_1), \ldots, M(B_n)$ are independent random variables for disjoint (Borel) sets $B_1, \ldots, B_n \subset E$.

Interpretation of frequentist confidence intervals

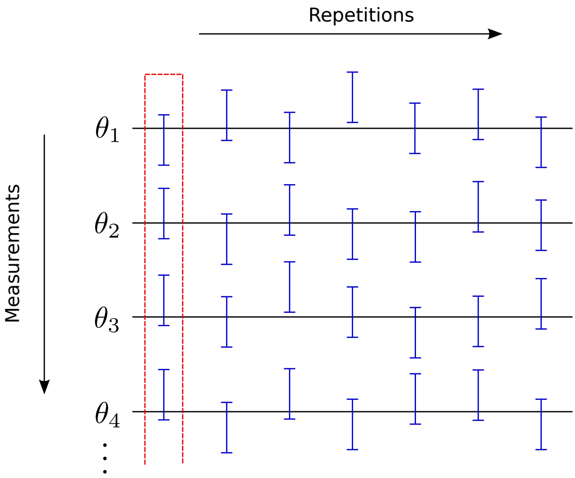
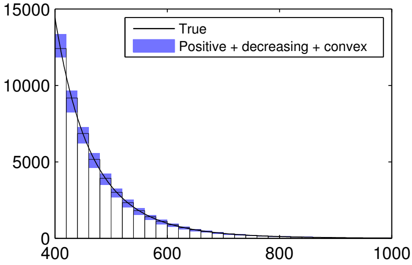


Figure: Frequentist confidence intervals imply coverage not only for independent repetitions of the same measurement, but also for independent measurements of unrelated quantities of interest $\{\theta_i\}$.

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 CMS 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