Unfolding Techniques: A Statistician's Perspective

Mikael Kuusela

Department of Statistics,
University of Chicago

Neutrino Seminar

Fermilab

April 13, 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

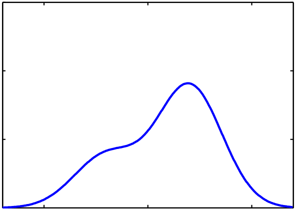
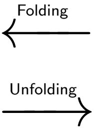


Figure: Smeared spectrum



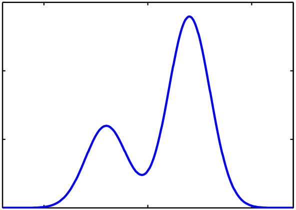
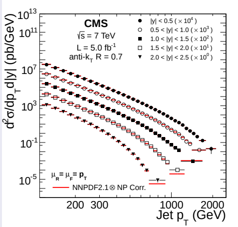


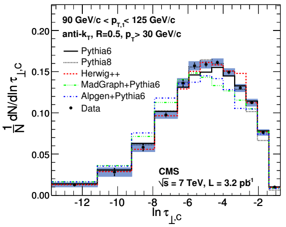
Figure: True spectrum

Examples of unfolding in LHC data analysis

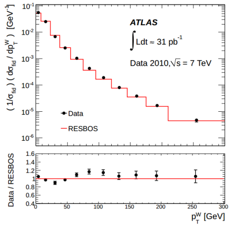
Inclusive jet cross section



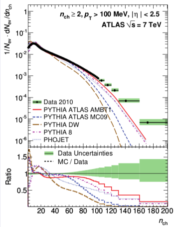
Hadronic event shape



W boson cross section



Charged particle multiplicity



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 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

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

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
- Smeared histogram $\boldsymbol{y} = [y_1, \ldots, y_n]^{\mathrm{T}}$ with mean

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

- Quantity of interest:

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

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

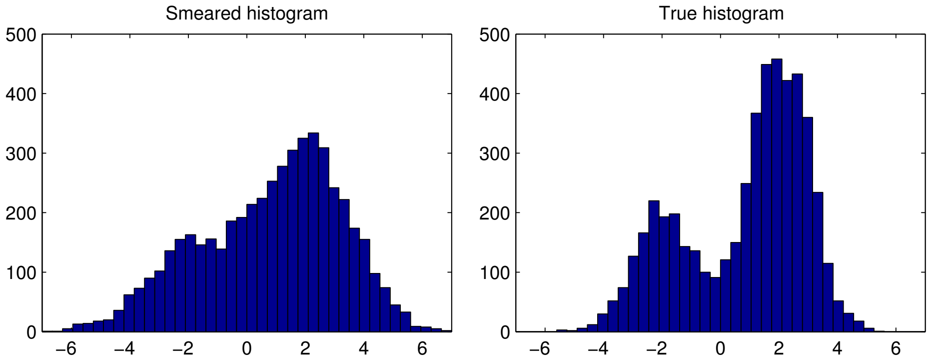
$$K_{i,j} = \frac{\int_{F_i} \int_{E_j} k(t,s) f(s)\,\mathrm{d}s\,\mathrm{d}t}{\int_{E_j} f(s)\,\mathrm{d}s} = P(\text{smeared event in bin } i \,|\, \text{true event in bin } j)$$

- The discretized statistical model becomes

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

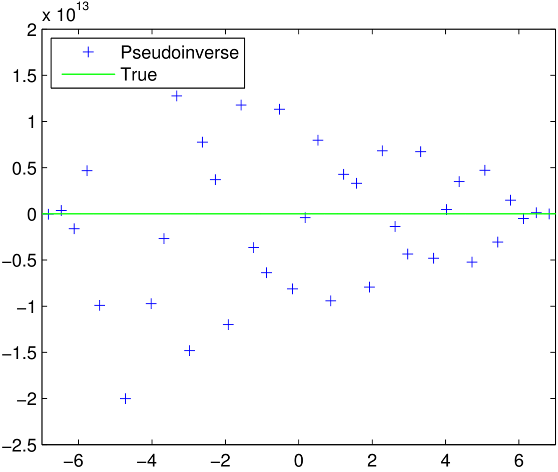
where $\boldsymbol{K}$ is an ill-conditioned matrix

Demonstration of ill-posedness

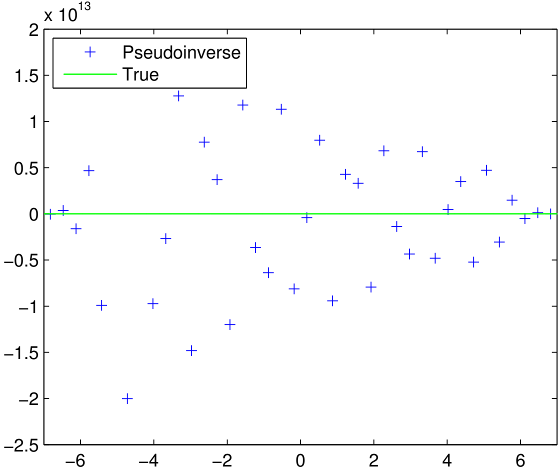


$$\boldsymbol{\mu} = \boldsymbol{K}\boldsymbol{\lambda}, \quad \boldsymbol{y} \sim \mathrm{Poisson}(\boldsymbol{\mu}) \quad \overset{??}{\Longrightarrow} \quad \hat{\boldsymbol{\lambda}} = \boldsymbol{K}^{-1}\boldsymbol{y}$$

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 $\boldsymbol{y} \sim \mathrm{Poisson}(\boldsymbol{K}\boldsymbol{\lambda})$
- 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, $k = 0$

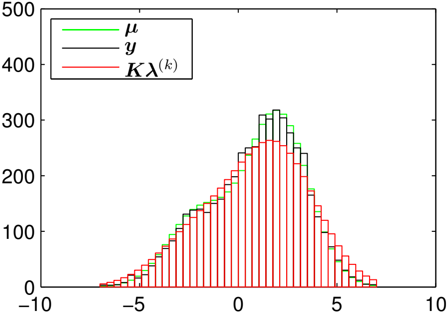


Figure: Smeared histogram

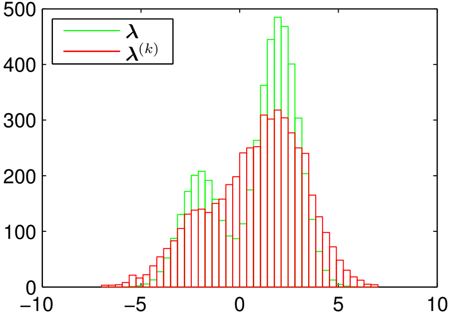


Figure: True histogram

D'Agostini demo, $k = 100$

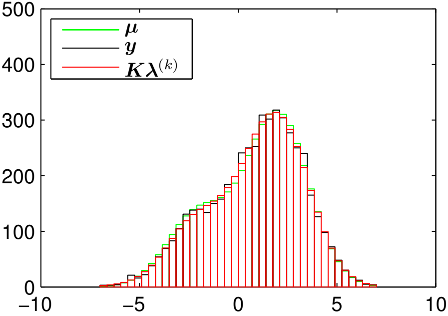


Figure: Smeared histogram

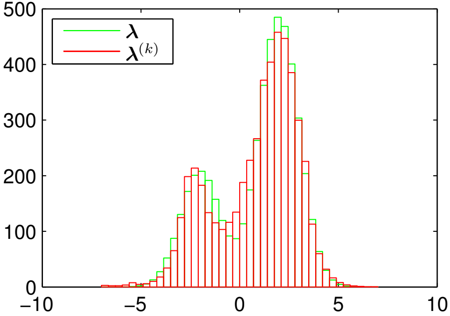


Figure: True histogram

D'Agostini demo, $k = 10000$

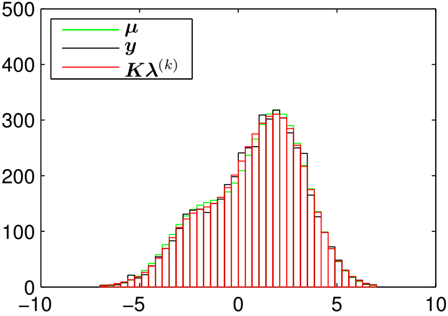


Figure: Smeared histogram

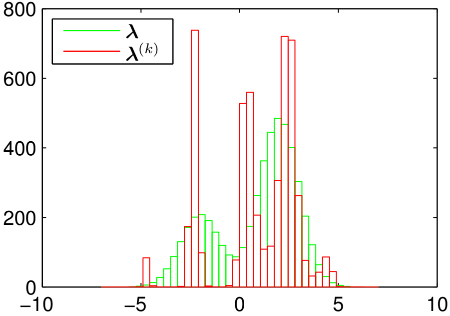


Figure: True histogram

D'Agostini demo, $k = 100000$

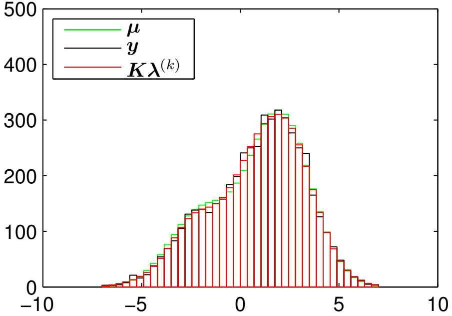


Figure: Smeared histogram

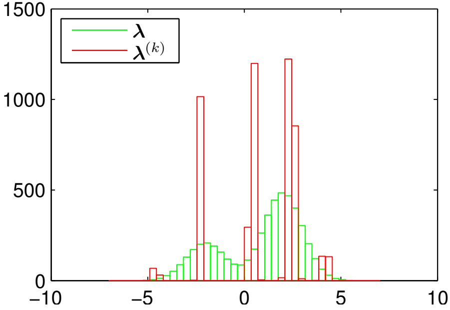


Figure: True histogram

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 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, 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 $\boldsymbol{K}$
- 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

[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

- 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...

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* $1 - \alpha$:

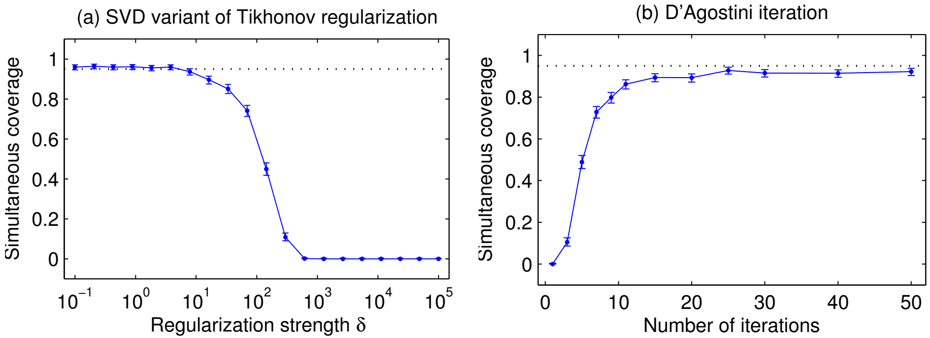
$$\mathrm{P}_f\left(\lambda_i \in \left[\underline{\lambda}_i(\boldsymbol{y}), \overline{\lambda}_i(\boldsymbol{y})\right]\right) \geq 1 - \alpha, \quad \forall f \in V$$

- 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

Undercoverage of existing methods (Kuusela, 2016)

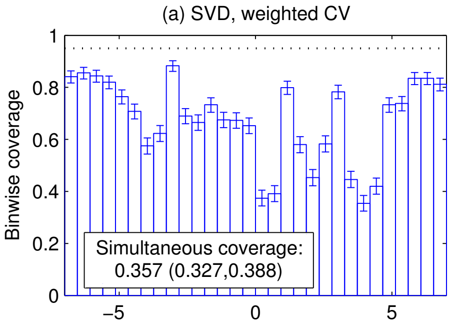


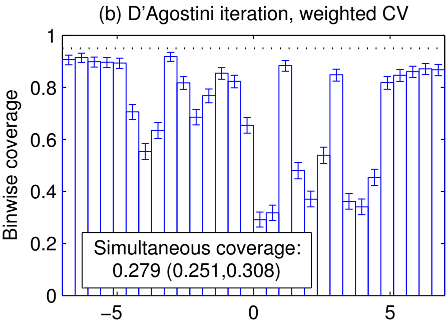
$$f(s) = \lambda_{\mathrm{tot}} \left\{ \pi_1 \mathcal{N}(s|-2,1) + \pi_2 \mathcal{N}(s|2,1) + \pi_3 \frac{1}{|E|} \right\}$$

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

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

Undercoverage of existing methods (Kuusela, 2016)





Towards improved uncertainty quantification

We have seen that standard UQ techniques can suffer from *severe undercoverage* in realistic unfolding scenarios.

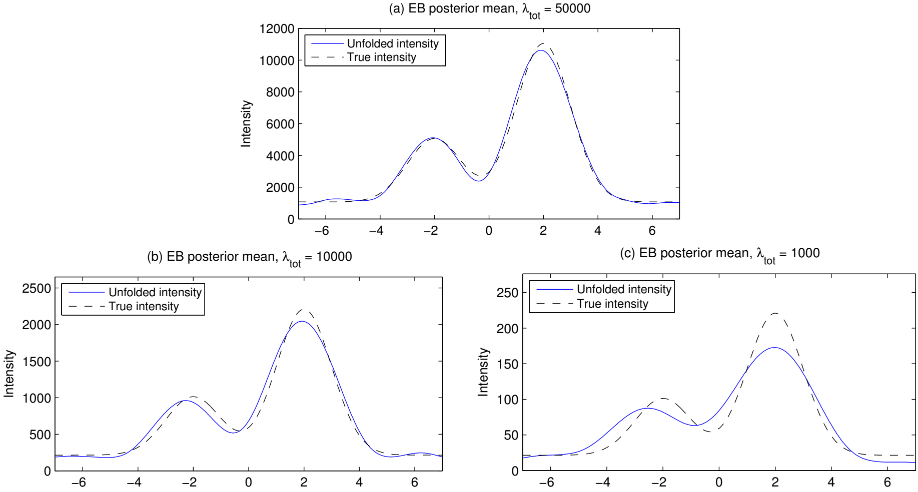
In the reminder of this talk, I will describe two complementary ways of obtaining improved unfolded uncertainty quantification:

1. **Debiased confidence intervals** for smooth spectra (Kuusela and Panaretos, 2015; Kuusela, 2016)
2. **Shape-constrained confidence intervals** for steeply falling spectra (Kuusela and Stark, 2017; Kuusela, 2016)

Debiased uncertainty quantification for smooth spectra

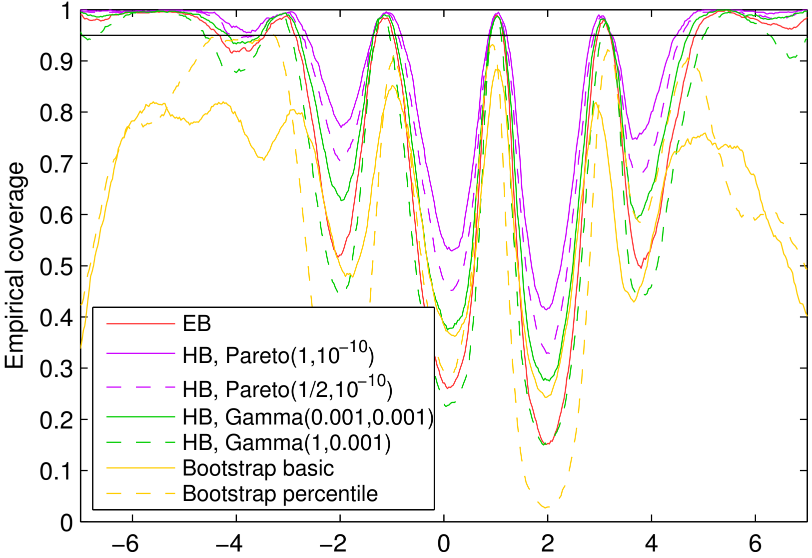
Joint work with Victor M. Panaretos (EPFL)

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



Debiasing by undersmoothing

- Let $\hat{\boldsymbol{\beta}}$ be the unfolded point estimator depending on the regularization strength δ (large $\delta \leftrightarrow$ strong regularization)
- A trivial way of debiasing $\hat{\boldsymbol{\beta}}$ 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
- This leads to major improvements in coverage at the expense of somewhat longer confidence intervals
- However: One can obtain even more powerful inferences by employing *iterative bias-corrections*

Iterative bias-correction

- By definition, the bias of $\hat{\boldsymbol{\beta}}$ is given by $\mathrm{bias}(\hat{\boldsymbol{\beta}}) = \mathsf{E}_{\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}) - \boldsymbol{\beta}$
- Denote the observed value of $\hat{\boldsymbol{\beta}}$ by $\hat{\boldsymbol{\beta}}^{(0)}$
- Plug-in estimate of the bias: $\widehat{\mathrm{bias}}^{(0)}(\hat{\boldsymbol{\beta}}) = \mathsf{E}_{\hat{\boldsymbol{\beta}}^{(0)}}(\hat{\boldsymbol{\beta}}) - \hat{\boldsymbol{\beta}}^{(0)}$
- Bias-corrected point estimate: $\hat{\boldsymbol{\beta}}^{(1)} = \hat{\boldsymbol{\beta}}^{(0)} - \widehat{\mathrm{bias}}^{(0)}(\hat{\boldsymbol{\beta}})$
- Repeat this iteratively:

Iterative bias-correction (Kuusela and Panaretos, 2015)

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

Confidence intervals from the bias-corrected estimator

- Let $\hat{\beta}_{\mathrm{BC}}$ be the bias-corrected point estimator obtained after N_{BC} bias-correction iterations
- The bias-corrected unfolded spectrum is $\hat{f}_{\mathrm{BC}}(s) = \sum_{j=1}^{p} \hat{\beta}_{\mathrm{BC},j} B_j(s)$
- We quantify the uncertainty of $f(s)$ using the variability of $\hat{f}_{\mathrm{BC}}(s)$
    - Percentile intervals (double bootstrap):

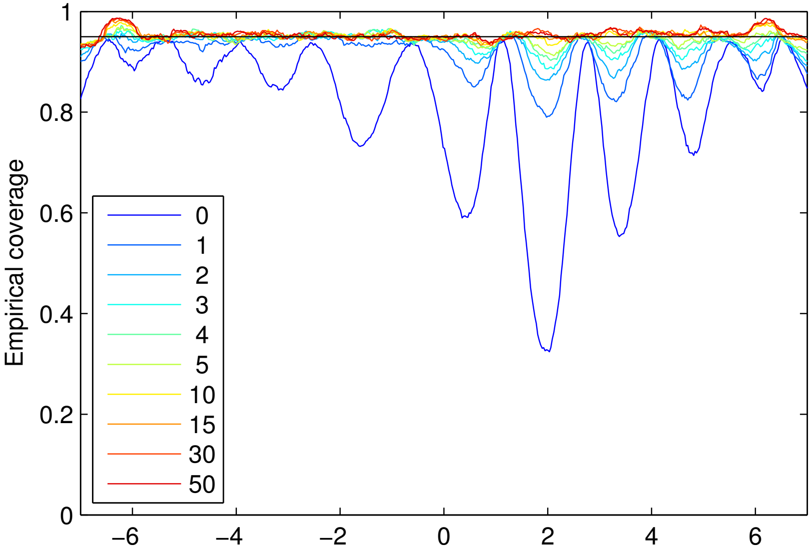
$$\left[\hat{f}^*_{\mathrm{BC},\alpha/2}(s), \hat{f}^*_{\mathrm{BC},1-\alpha/2}(s)\right]$$

    - Gaussian intervals:

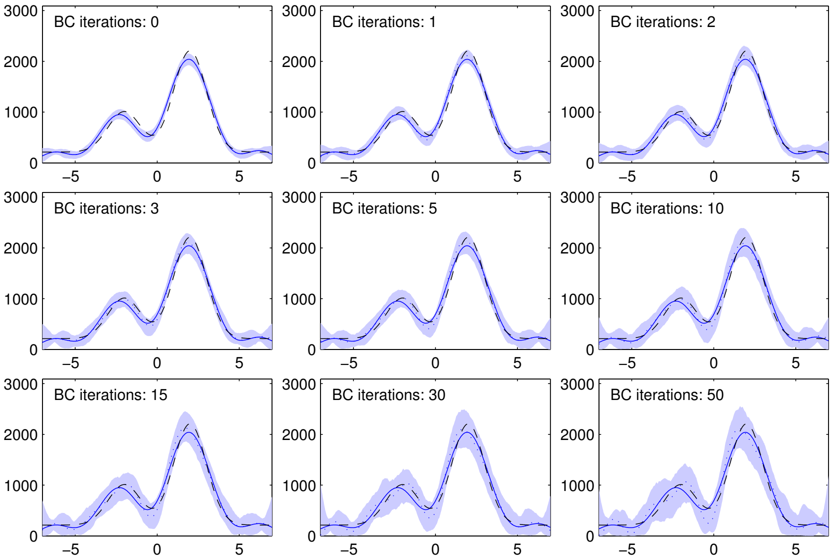
$$\left[\hat{f}_{\mathrm{BC}}(s) - z_{1-\alpha/2}\sqrt{\widehat{\mathrm{var}}\big(\hat{f}_{\mathrm{BC}}(s)\big)},\ \hat{f}_{\mathrm{BC}}(s) + z_{1-\alpha/2}\sqrt{\widehat{\mathrm{var}}\big(\hat{f}_{\mathrm{BC}}(s)\big)}\right]$$

- Percentile intervals take into account the (potential) asymmetry of the sampling distribution of $\hat{f}_{\mathrm{BC}}(s)$

Effect of bias-correction on coverage



Effect of bias-correction on interval length



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

[See Kuusela (2016) for details.]

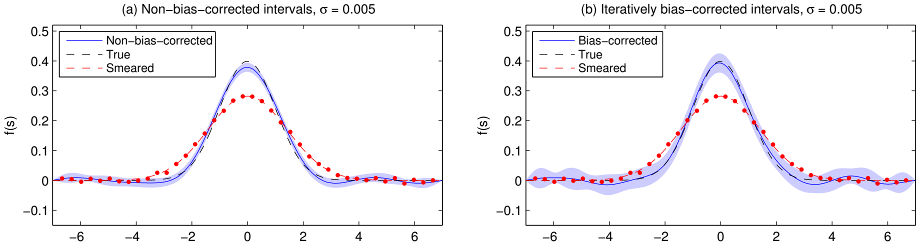


Figure: Gaussian intervals, 95 % nominal coverage, 94 % target coverage

Coverage study

Noise level	Method	Coverage at $s = 0$	Mean length
$\sigma = 0.005$	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

Similar conclusions hold for:

- Different noise levels / sample sizes
- Different choices of the true spectrum f
- Gaussian and Laplacian smearing kernels

Shape-constrained uncertainty quantification for steeply falling spectra

Joint work with Philip B. Stark (UC Berkeley)

Shape-constrained unfolding

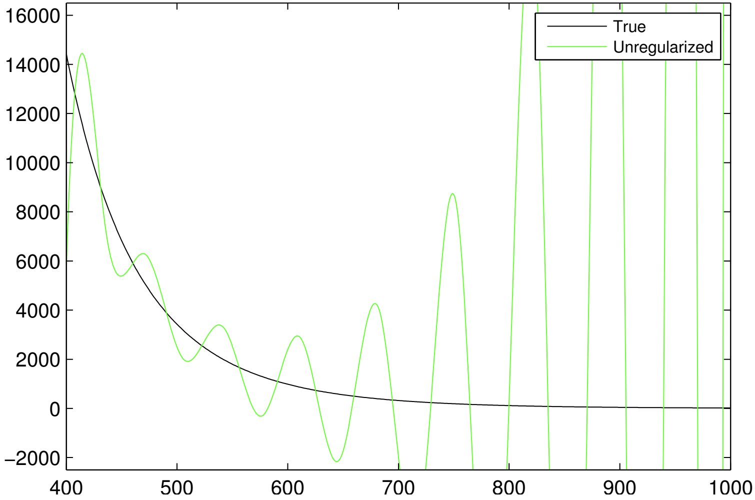
We present a technique for forming confidence intervals for $\boldsymbol{\lambda}$ that have *guaranteed simultaneous frequentist finite-sample coverage*, provided that f satisfies simple, physically justified shape constraints.

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

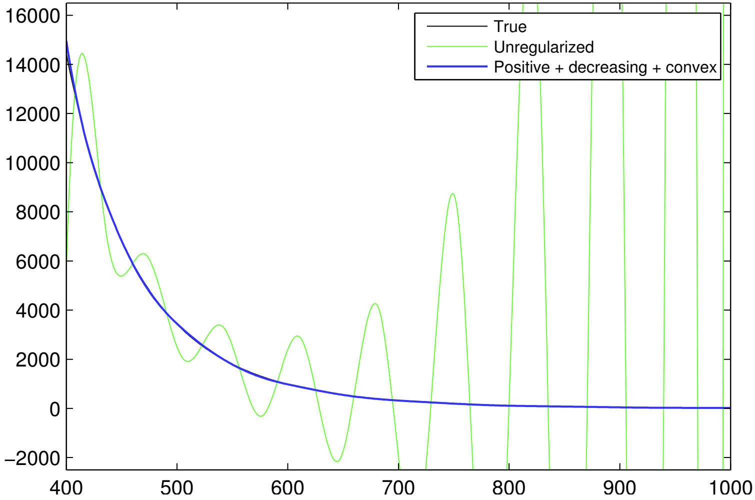
Examples from the LHC include the differential cross sections of:

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

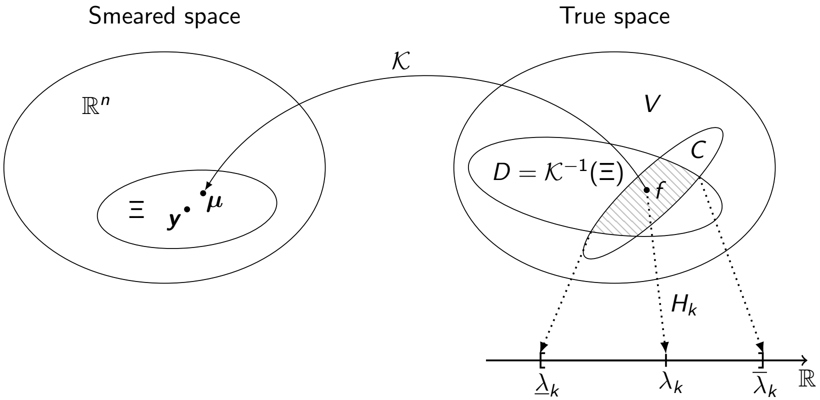
Regularization using shape constraints



Regularization using shape constraints



Strict bounds confidence intervals (Stark, 1992)



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

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

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

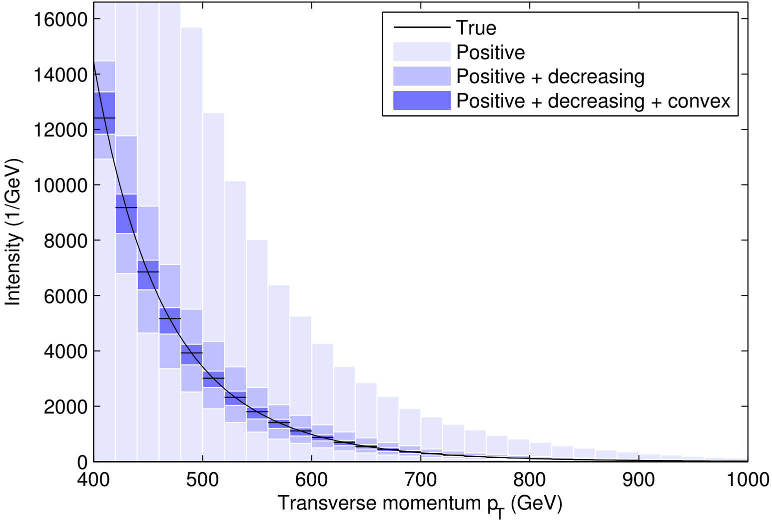


Figure: Shape-constrained unfolded confidence intervals for the inclusive jet p_{T} spectrum with *guaranteed* conservative 95 % simultaneous coverage.

Demonstration: Inclusive jet p_{T} spectrum

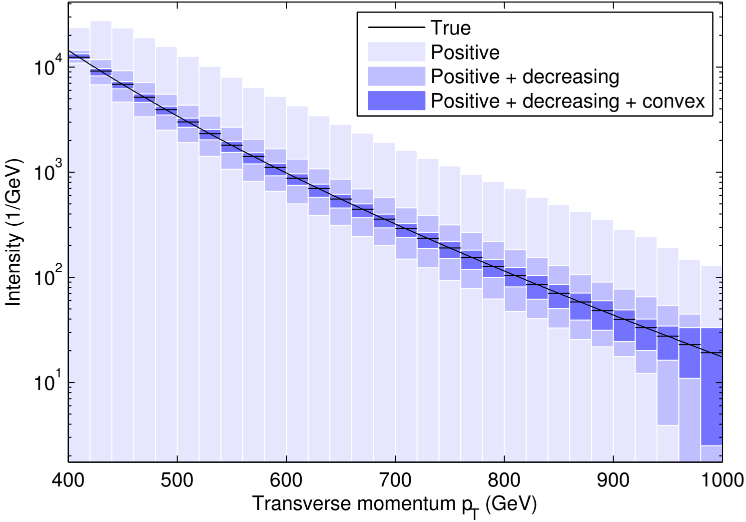


Figure: Shape-constrained unfolded confidence intervals for the inclusive jet p_{T} spectrum with *guaranteed* conservative 95 % simultaneous coverage.

Summary and conclusions

- Ill-posedness makes unfolding a very complex statistical problem
- 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
  - Uncertainties can be improved by debiasing (undersmoothing or iterative bias-correction) or by imposing qualitative shape constraints
- Many open issues remain:
  - How to handle systematic uncertainties?
  - How to properly compare, combine and propagate unfolded results?
- For more details, simulations and discussion see:

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, arXiv:1512.00905v3 [stat.AP], under minor revision for *The Annals of Applied Statistics*, 2017.

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

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.

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, 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, 2013a.

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, 2013b.

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.

References II

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.

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. `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.

References III

M. Kuusela and P. B. Stark, Shape-constrained uncertainty quantification in unfolding steeply falling elementary particle spectra, arXiv:1512.00905v3 [stat.AP], under minor revision, 2017.

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

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

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

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.

References IV

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.

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.