Unfolding methods in ATLAS

Georgios Choudalakis[1]
on behalf of the **ATLAS** Collaboration

[1]University of Chicago

PHYSTAT2011
CERN, January 20, 2011

Overview

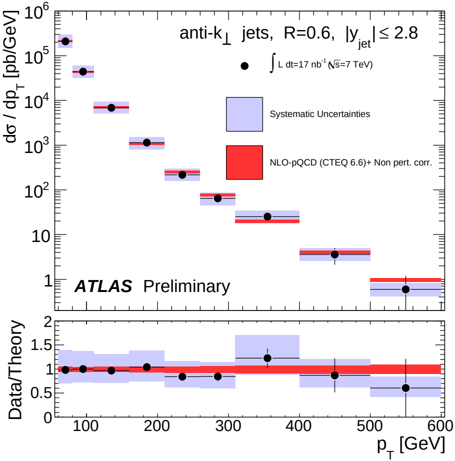
- So far, ATLAS uses *bin-by-bin*[1] and *iterative*[2] unfolding in SM measurements.
- Both (but particularly bin-by-bin) have received criticism.

  ATLAS is considering methods beyond bin-by-bin for the next round of analyses.

- Bin-by-bin unfolding:
  - ☞ Inclusive jet and dijet spectrum: (`arXiv:1009.5908v2 [hep-ex]`)
  - Inclusive γ spectrum: (`arXiv:1012.4389v2 [hep-ex]`)
  - Jet shape measurement (`arXiv:1101.0070 [hep-ex]`)
  - W+jets cross section measurements (`arXiv:1012.5382 [hep-ex]`)
- Iterative unfolding:
  - ☞ Charged-particle multiplicities measurement (`arXiv:1101.0598 [hep-ex]`)
- No unfolding in exotic searches, because it is *unnecessary* for making a discovery, or setting a limit, or estimating the parameters of a new particle. See when unfolding is necessary in Louis Lyons' earlier talk.

[1] A.k.a. "Correction Factors Method".
[2] A.k.a. "Bayesian" method, although it is not 100% Bayesian.

The Inclusive Jet p_T spectrum

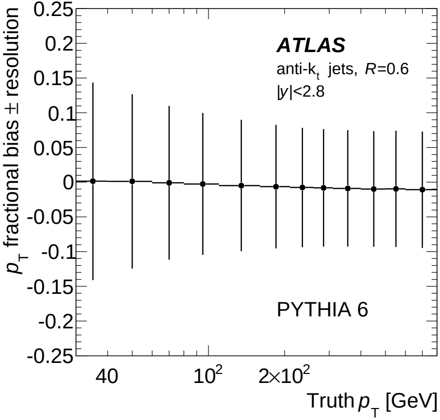


Jet p_T spectrum, *after* bin-by-bin unfolding.
Will return to this.

Binning

- At $p_T \lesssim 310$ GeV the bins are defined based on experimental criteria (trigger).

p_T resolution according to QCD MC + ATLAS simulation



Spread (error bars) and bias of reconstructed jet p_T, in bins of true p_T.
Indicative of migration matrix. Spread leads to off-diagonal elements.

Bin-by-bin correction factors

- T_i: Truth-level MC spectrum, without event selection.
- R_i: Reco-level MC spectrum, after event selection {Trigger, Jet reconstruction efficiency, primary vertex position, jet quality, etc.}.
- D_i: Data spectrum. (Integer values)
- C_i: Bin-by-bin correction factor

$$C_i = \frac{T_i}{R_i}.$$

- U_i: Unfolded spectrum

$$U_i = C_i \cdot D_i.$$

- Statistical standard deviation

$$\sigma_{U_i} \simeq C_i \sqrt{D_i}.$$

[E.g., if $C_i = 0.8$ and $D_i = 100$, we have $U_i = 0.8(100 \pm \sqrt{100}) = 80 \pm 8$.]

Side comment

- It ignores correlations: events don't just disappear / appear; they migrate.
- When $R_i > T_i$ (e.g. due to strong smearing), the relative statistical uncertainty becomes smaller than if we had an ideal detector.

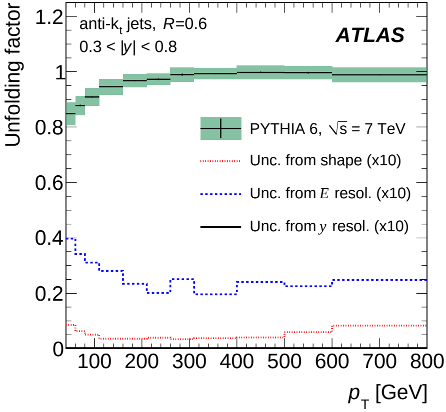
$$\frac{\sigma_{U_i}}{\langle U_i \rangle} = \frac{C_i \sqrt{R_i}}{C_i R_i} = \frac{1}{\sqrt{R_i}}.$$

If we had an ideal detector we wouldn't do any unfolding; D_i would itself be an estimator of T_i, which would follow a Poisson with mean T_i, so

$$\frac{\sigma_{D_i}}{\langle D_i \rangle} = \frac{\sqrt{T_i}}{T_i} = \frac{1}{\sqrt{T_i}}.$$

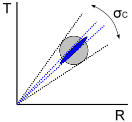
So, if $R_i > T_i$, we estimate T_i *more precisely* than if we had a perfect detector**!?**

Correction factors



Each corr. factor $C_i = \frac{T_i}{R_i}$ has uncertainty, from:

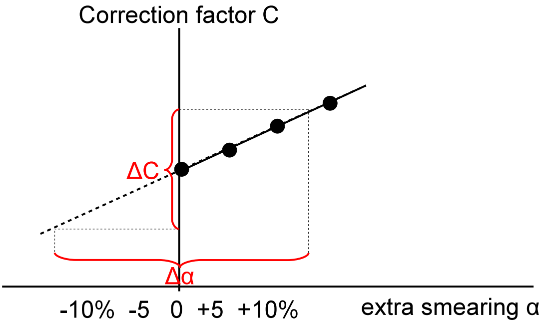
- Finite MC statistics to obtain T_i and R_i. [Tiny black error bars.] The cov(T_i, R_i) in each bin was taken into account.



- Uncertainty in R_i because the MC smearing may be unrealistic.
- Uncertainty in T_i and R_i, because physics in the MC may be unrealistic. [e.g. Pythia – Herwig]

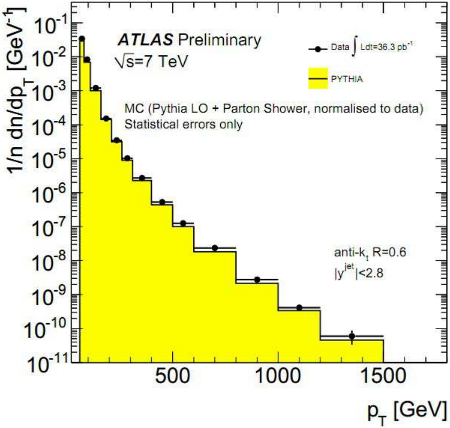
Syst. uncertainty on C_i due to MC p_T smearing uncertainty

- Took MC events.
- Smeared the reconstructed p_T of each jet by an extra α =15% (on top of the "nominal" smearing already present in the MC).
- Plotted the new R_i spectrum, and found the new $C_i = \frac{T_i}{R_i}$ factors.
- Repeated with extra smearing α =5%, 10%, 20%.
- Found that C_i increases linearly with α.
- Propagated smearing uncertainty into correction factor uncertainty as shown schematically here:



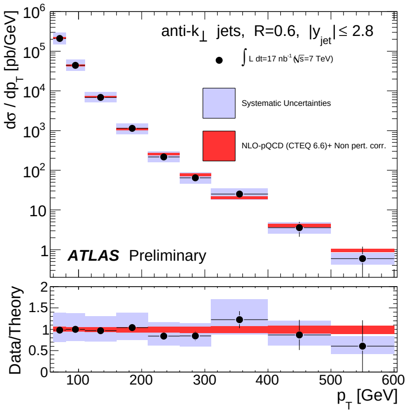
Syst. uncertainty on C_i due to MC spectrum uncertainty

- Re-weight the MC events according to their $\hat{p}_T$.
- The re-weighting functions are chosen to bracket the Pythia - Herwig - Alpgen spectral difference, and the difference between Pythia - data at detector-level.
- For each re-weighted MC new C_i were computed.
- The envelope of C_i variation was used as a systematic uncertainty.



Reconstructed p_T in data and in Pythia SM MC.

Back to the unfolded spectrum



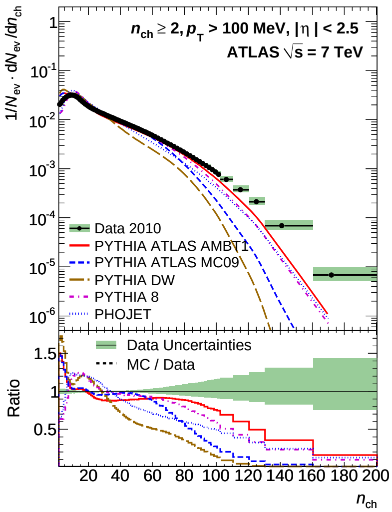
- Blue error bar: *Dominated* by jet energy scale uncertainty. Propagated by shifting jet p_T in MC, at detector-level (R), to find $\frac{\sigma_R}{R}$. The same relative uncertainty applies to the unfolded spectrum.

$$
\begin{aligned}
U &= \frac{T}{R} D \\
\sigma_U &= \frac{T}{R}\frac{\sigma_R}{R} D \Rightarrow \\
\frac{\sigma_U}{U} &= \frac{\sigma_R}{R}
\end{aligned}
$$

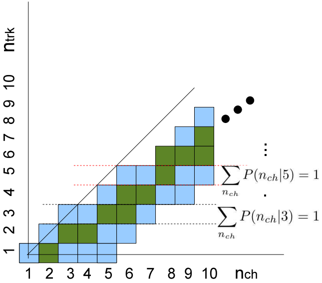
- The black error bar is $C_i\sqrt{D_i}$, divided by bin width and by luminosity.
- The other systematic uncertainties of C_i are $\lesssim 5\%$.
- [Luminosity uncertainty of 11% (not shown).]

2nd example: Iterative unfolding

Iterative unfolding of $n_{trk} \rightarrow n_{ch}$



- n_{ch}: *True* charged particles in an event.
- n_{trk}: Reconstructed charged particles (tracks).
- $n_{trk} \leq n_{ch}$ due to inefficiency. (Fake tracks are small in comparison.)
- Migration matrix, schematically:



ELSEVIER

Nuclear Instruments and Methods in Physics Research A 362 (1995) 487–498

NUCLEAR INSTRUMENTS & METHODS IN PHYSICS RESEARCH Section A

A multidimensional unfolding method based on Bayes' theorem

G. D'Agostini *

Università "La Sapienza" and INFN, Roma, Italy

Received 5 August 1994; revised form received 2 March 1995

$$\hat{n}(C_i) = \frac{1}{\epsilon_i} \sum_{j=1}^{n_E} n(E_j) P(C_i|E_j) \quad \epsilon_i \neq 0. \tag{4}$$

$$P(C_i|E_j) = \frac{P(E_j|C_i) P_0(C_i)}{\sum_{l=1}^{n_C} P(E_j|C_l) P_0(C_l)}. \tag{3}$$

Notation

"Cause" $C \to n_{ch}$

"Effect" $E \to n_{trk}$

"Efficiency" $\epsilon \to P(n_{trk} \geq 2 | n_{ch})$

The prior $P_0(n_{ch})$ and convergence

- **Prior used:** the n_{ch} spectrum in Pythia MC.
- It took **4** iterations to converge.
- Convergence means to reach

$$\frac{\chi^2}{N_{bins}} < 1,$$

  where

$$\chi^2 = \sum_{i=1}^{N_{bins}} \left(\frac{n_{ch}^{i,now} - n_{ch}^{i,before}}{\sqrt{n_{ch}^{i,before}}} \right)^2 .$$

- As a check, a flat prior was tried (which is a very unphysical assumption). *Still,* the result differed by less than 2% in all bins. This was included as a systematic uncertainty. However, convergence was slower (∼7 iterations). By the time the tail converged, the bulk started showing bin-to-bin fluctuations. [This may serve as a hint that a prior is too wrong.]

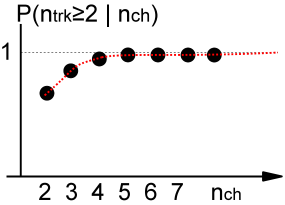
Correcting for acceptance

Example requiring $n_{trk} \geq 2$.

Find in MC the $P(n_{trk} \geq 2 | n_{ch})$. Overlay it with:

$$f(n_{ch}) = 1 - (1 - \epsilon_{eff})^{n_{ch}} - n_{ch}(1 - \epsilon_{eff})^{n_{ch}-1}\epsilon_{eff},$$

where ϵ_{eff} is adjusted to get $f(2) = P(n_{trk} \geq 2 | n_{ch} = 2)$.
This ϵ_{eff} is the *effective average* track-level efficiency. [$\sim$4% from actual $\langle \epsilon_{trk} \rangle$.]



(schematic, not using ATLAS MC)

This analytic expression was used to correct for acceptance:

$$N(n_{ch}) = \frac{1}{1 - (1 - \epsilon_{eff})^{n_{ch}} - n_{ch}(1 - \epsilon_{eff})^{n_{ch}-1}\epsilon_{eff}} \sum_{n_{trk} \geq 2} N(n_{trk})\, P(n_{ch} | n_{trk})$$

Uncertainty in the result of iterative unfolding

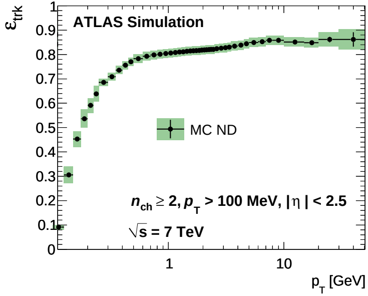
- The statistical uncertainty of observed data.

$$N(n_{ch}) = \frac{1}{\epsilon} \sum_{n_{trk} \geq 1} N(n_{trk}) P(n_{ch}|n_{trk})$$

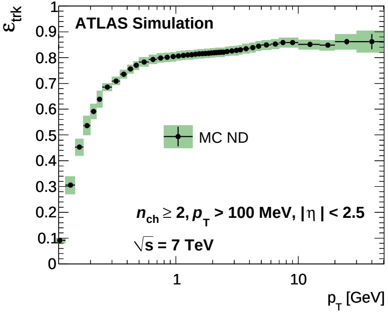
$$\sigma_{N(n_{ch})} = \frac{1}{\epsilon} \sqrt{\sum_{n_{trk} \geq 1} P^2_{(n_{ch}|n_{trk})} N(n_{trk})}$$

All $N(n_{trk})$ are $\geq$500, so we draw symmetric error-bars. (Invisibly small.)

- Dominant systematic uncertainties:
  - Track reco. eff/cy uncertainty (ϵ_{trk}).
  - The MC, at reco-level, doesn't reproduce the observed track p_T distribution. ϵ_{trk} depends on track p_T, so wrong p_T means wrong efficiency, and that's what we try to unfold.

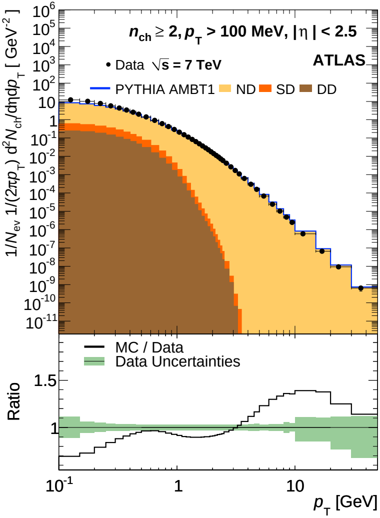


Uncertainty due to ϵ_{trk}

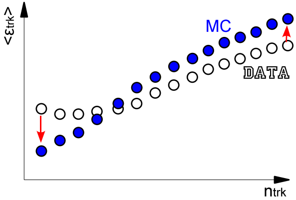


- Instead of modifying ϵ_{trk} in MC, changing $P(n_{ch}|n_{trk})$, we modify the data.
- A data event has n_{trk} tracks. Take the 1st. Its p_T is such that $\epsilon_{trk} = 0.80 \pm 0.05$. *Reducing* ϵ_{trk} by 1σ means we expect $\frac{1}{0.80} \times 0.75 = 0.9375$ tracks. Randomly keep the track, with $P = 0.9375$, else delete it. Repeat for all tracks. In the end, the event will have $n'_{trk} \leq n_{trk}$ tracks.
- Repeat for all events. Unfold the n'_{trk} spectrum, to obtain n'_{ch}. Compare n'_{ch} to n_{ch}.
- We only *reduced* the n_{trk} of each event, so symmetrize the difference: In each bin of the unfolded spectrum, if $n'_{ch} = n_{ch} \cdot (1 - x\%)$, assume n_{ch} has uncertainty $\pm x\%$.

Uncertainty due to different p_T^{trk} spectrum.



- The data have more low pt tracks than the MC.
- In bins of n_{trk}, we find the mean track efficiency for data and MC events.

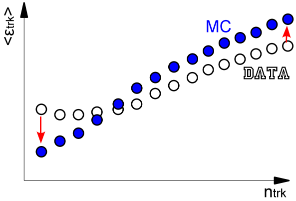


(Schematic; not actual data & MC)

Uncertainty due to different p_T^{trk} spectrum.

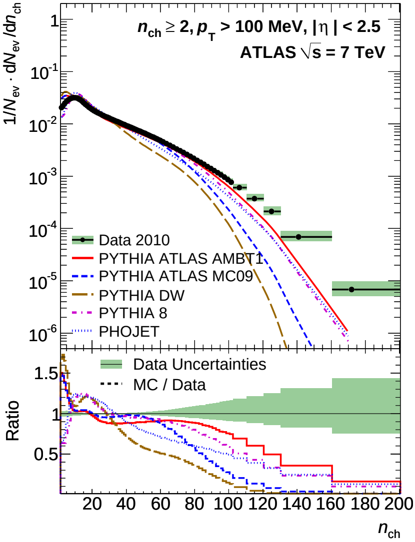
- Each data event has some n_{trk}, so we know if ϵ_{trk} should be shifted $\downarrow$ or $\uparrow$.
- When $\downarrow$, do as before: Delete some tracks in data $\rightarrow$ $n'_{trk} \xrightarrow{\text{unfold}} n'_{ch}$, compare to nominal n_{ch}. Don't symmetrize; asymmetric error.
- When $\uparrow$, *delete* some tracks as before, and flip the error sign.

- The data have more low pt tracks than the MC.
- In bins of n_{trk}, we find the mean track efficiency for data and MC events.



(Schematic; not actual data & MC)

Back to the unfolded spectrum



- The black error bars (invisible) are statistical errors.
- Full covariance matrix not given.
- 3% to 25% symmetric uncertainty from ϵ_{trk} uncertainty.
- -2% to +40% asymmetric uncertainty from p_T spectrum difference.

Summary

- We saw two detailed, real-life examples of unfolding in ATLAS.
- ATLAS has used extensively bin-by-bin, and in some cases iterative unfolding.
- We are considering methods beyond bin-by-bin for the next round of analyses.
- Unfolding has been used only in measurements.
- Unfolded spectra are usually compared to theory visually. [No goodness-of-fit tests / anything that requires the covariance matrix, worrying about low statistics, asymmetric stat. errors etc.] In some cases (jet shapes, MC tuning) a χ^2 is computed, but is not used to find a *p*-value; only to quantify which tuning is better.
- Exotic searches didn't unfold, to avoid unnecessary complications.
- The reason for unfolding is our desire to show truth-level spectra. For fitting/setting limits/hypothesis testing/future analysis, better fold than unfold, when det. simulation is unavailable. Experiments can be compared without unfolding, on the level of parameter estimates (not spectra).
- In ATLAS there are many views about unfolding, when/how to do it.
- We hope PHYSTAT2011 will clarify this young, often confusing subject.

Summary

- We saw two detailed, real-life examples of unfolding in ATLAS.
- ATLAS has used extensively bin-by-bin, and in some cases iterative unfolding.
- We are considering methods beyond bin-by-bin for the next round of analyses.
- Unfolding has been used only in measurements.
- Unfolded spectra are usually compared to theory visually. [No goodness-of-fit tests / anything that requires the covariance matrix, worrying about low statistics, asymmetric stat. errors etc.] In some cases (jet shapes, MC tuning) a χ^2 is computed, but is not used to find a p-value; only to quantify which tuning is better.
- Exotic searches didn't unfold, to avoid unnecessary complications.
- The reason for unfolding is our desire to show truth-level spectra. For fitting/setting limits/hypothesis testing/future analysis, better fold than unfold, when det. simulation is unavailable. Experiments can be compared without unfolding, on the level of parameter estimates (not spectra).
- In ATLAS there are many views about unfolding, when/how to do it.
- We hope PHYSTAT2011 will clarify this young, often confusing subject.

Summary

- We saw two detailed, real-life examples of unfolding in ATLAS.
- ATLAS has used extensively bin-by-bin, and in some cases iterative unfolding.
- We are considering methods beyond bin-by-bin for the next round of analyses.
- Unfolding has been used only in measurements.
- Unfolded spectra are usually compared to theory visually. [No goodness-of-fit tests / anything that requires the covariance matrix, worrying about low statistics, asymmetric stat. errors etc.] In some cases (jet shapes, MC tuning) a χ^2 is computed, but is not used to find a *p*-value; only to quantify which tuning is better.
- Exotic searches didn't unfold, to avoid unnecessary complications.
- The reason for unfolding is our desire to show truth-level spectra. For fitting/setting limits/hypothesis testing/future analysis, better fold than unfold, when det. simulation is unavailable. Experiments can be compared without unfolding, on the level of parameter estimates (not spectra).
- In ATLAS there are many views about unfolding, when/how to do it.
- We hope PHYSTAT2011 will clarify this young, often confusing subject.

Summary

- We saw two detailed, real-life examples of unfolding in ATLAS.
- ATLAS has used extensively bin-by-bin, and in some cases iterative unfolding.
- We are considering methods beyond bin-by-bin for the next round of analyses.
- Unfolding has been used only in measurements.
- Unfolded spectra are usually compared to theory visually. [No goodness-of-fit tests / anything that requires the covariance matrix, worrying about low statistics, asymmetric stat. errors etc.] In some cases (jet shapes, MC tuning) a χ^2 is computed, but is not used to find a *p*-value; only to quantify which tuning is better.
- Exotic searches didn't unfold, to avoid unnecessary complications.
- The reason for unfolding is our desire to show truth-level spectra. For fitting/setting limits/hypothesis testing/future analysis, better fold than unfold, when det. simulation is unavailable. Experiments can be compared without unfolding, on the level of parameter estimates (not spectra).
- In ATLAS there are many views about unfolding, when/how to do it.
- We hope PHYSTAT2011 will clarify this young, often confusing subject.

Summary

- We saw two detailed, real-life examples of unfolding in ATLAS.
- ATLAS has used extensively bin-by-bin, and in some cases iterative unfolding.
- We are considering methods beyond bin-by-bin for the next round of analyses.
- Unfolding has been used only in measurements.
- Unfolded spectra are usually compared to theory visually. [No goodness-of-fit tests / anything that requires the covariance matrix, worrying about low statistics, asymmetric stat. errors etc.] In some cases (jet shapes, MC tuning) a χ^2 is computed, but is not used to find a *p*-value; only to quantify which tuning is better.
- Exotic searches didn't unfold, to avoid unnecessary complications.
- The reason for unfolding is our desire to show truth-level spectra. For fitting/setting limits/hypothesis testing/future analysis, better fold than unfold, when det. simulation is unavailable. Experiments can be compared without unfolding, on the level of parameter estimates (not spectra).
- In ATLAS there are many views about unfolding, when/how to do it.
- We hope PHYSTAT2011 will clarify this young, often confusing subject.

Summary

- We saw two detailed, real-life examples of unfolding in ATLAS.
- ATLAS has used extensively bin-by-bin, and in some cases iterative unfolding.
- We are considering methods beyond bin-by-bin for the next round of analyses.
- Unfolding has been used only in measurements.
- Unfolded spectra are usually compared to theory visually. [No goodness-of-fit tests / anything that requires the covariance matrix, worrying about low statistics, asymmetric stat. errors etc.] In some cases (jet shapes, MC tuning) a χ^2 is computed, but is not used to find a *p*-value; only to quantify which tuning is better.
- Exotic searches didn't unfold, to avoid unnecessary complications.
- The reason for unfolding is our desire to show truth-level spectra. For fitting/setting limits/hypothesis testing/future analysis, better fold than unfold, when det. simulation is unavailable. Experiments can be compared without unfolding, on the level of parameter estimates (not spectra).
- In ATLAS there are many views about unfolding, when/how to do it.
- We hope PHYSTAT2011 will clarify this young, often confusing subject.

Summary

- We saw two detailed, real-life examples of unfolding in ATLAS.
- ATLAS has used extensively bin-by-bin, and in some cases iterative unfolding.
- We are considering methods beyond bin-by-bin for the next round of analyses.
- Unfolding has been used only in measurements.
- Unfolded spectra are usually compared to theory visually. [No goodness-of-fit tests / anything that requires the covariance matrix, worrying about low statistics, asymmetric stat. errors etc.] In some cases (jet shapes, MC tuning) a χ^2 is computed, but is not used to find a *p*-value; only to quantify which tuning is better.
- Exotic searches didn't unfold, to avoid unnecessary complications.
- The reason for unfolding is our desire to show truth-level spectra. For fitting/setting limits/hypothesis testing/future analysis, better fold than unfold, when det. simulation is unavailable. Experiments can be compared without unfolding, on the level of parameter estimates (not spectra).
- In ATLAS there are many views about unfolding, when/how to do it.
- We hope PHYSTAT2011 will clarify this young, often confusing subject.

Summary

- We saw two detailed, real-life examples of unfolding in ATLAS.
- ATLAS has used extensively bin-by-bin, and in some cases iterative unfolding.
- We are considering methods beyond bin-by-bin for the next round of analyses.
- Unfolding has been used only in measurements.
- Unfolded spectra are usually compared to theory visually. [No goodness-of-fit tests / anything that requires the covariance matrix, worrying about low statistics, asymmetric stat. errors etc.] In some cases (jet shapes, MC tuning) a χ^2 is computed, but is not used to find a *p*-value; only to quantify which tuning is better.
- Exotic searches didn't unfold, to avoid unnecessary complications.
- The reason for unfolding is our desire to show truth-level spectra. For fitting/setting limits/hypothesis testing/future analysis, better fold than unfold, when det. simulation is unavailable. Experiments can be compared without unfolding, on the level of parameter estimates (not spectra).
- In ATLAS there are many views about unfolding, when/how to do it.
- We hope PHYSTAT2011 will clarify this young, often confusing subject.

Summary

- We saw two detailed, real-life examples of unfolding in ATLAS.
- ATLAS has used extensively bin-by-bin, and in some cases iterative unfolding.
- We are considering methods beyond bin-by-bin for the next round of analyses.
- Unfolding has been used only in measurements.
- Unfolded spectra are usually compared to theory visually. [No goodness-of-fit tests / anything that requires the covariance matrix, worrying about low statistics, asymmetric stat. errors etc.] In some cases (jet shapes, MC tuning) a χ^2 is computed, but is not used to find a *p*-value; only to quantify which tuning is better.
- Exotic searches didn't unfold, to avoid unnecessary complications.
- The reason for unfolding is our desire to show truth-level spectra. For fitting/setting limits/hypothesis testing/future analysis, better fold than unfold, when det. simulation is unavailable. Experiments can be compared without unfolding, on the level of parameter estimates (not spectra).
- In ATLAS there are many views about unfolding, when/how to do it.
- We hope PHYSTAT2011 will clarify this young, often confusing subject.

Backup

Some past analyses that used bin-by-bin unfolding

- T. Aaltonen *et al.* [CDF Collaboration], "Search for new particles decaying into dijets in proton-antiproton collisions at $\sqrt{s} = 1.96$ TeV," Phys. Rev. D **79**, 112002 (2009) [arXiv:0812.4036 [hep-ex]].
- H. Abramowicz *et al.* [ZEUS Collaboration], "Inclusive dijet cross sections in neutral current deep inelastic scattering at HERA," Eur. Phys. J. C **70**, 965 (2010) [arXiv:1010.6167 [hep-ex]].
- T. Aaltonen *et al.* [CDF Collaboration], "Measurement of the Inclusive Jet Cross Section at the Fermilab Tevatron p-pbar Collider Using a Cone-Based Jet Algorithm," Phys. Rev. D **78**, 052006 (2008) [Erratum-ibid. D **79**, 119902 (2009)] [arXiv:0807.2204 [hep-ex]].

Small statistics treatment [in bin-by-bin, where $U_i = C_i \cdot (D_i \pm \sqrt{D_i})$.]

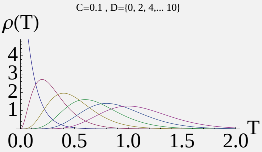
$\sqrt{D_i}$ is an approximation with obvious side-effects:

- When $D_i \to 0$: $U_i \to 0 \pm 0$ **!?**
- The uncertainty is not symmetric for low D_i. [Such approximations would matter in an exotic search, where D_i is small.]

Bayesian inference of T_i

For flat prior in T_i, the posterior is:

$$P(T_i|D_i, C_i) = \frac{1}{C_i} \cdot \frac{(T_i/C_i)^{D_i}}{D_i!} \cdot e^{-\frac{T_i}{C_i}},$$



$\langle T_i \rangle = C_i(D_i + 1)$,
$\sqrt{\langle T_i^2 \rangle - \langle T_i \rangle^2} = C_i\sqrt{D_i + 1}$.

Frequentist 68% CI of T_i

For large D_i, the CI of T_i is:

$$C_i(D_i \pm \sqrt{D_i}).$$

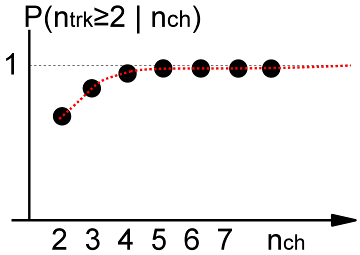
For small D_i, the CI is not symmetric around C_iD_i, and is not so simple. Well-discussed topic [e.g. Feldman-Cousins], we won't delve in it.

Bias of bin-by-bin

In cases with no background,

$$\begin{aligned}\mathrm{bias}_i &= \langle U_i \rangle - T_i^{\mathrm{nature}} \\ &= \left(\frac{T_i}{R_i} - \frac{T_i^{\mathrm{nature}}}{R_i^{\mathrm{nature}}} \right) R_i^{\mathrm{nature}}.\end{aligned}$$

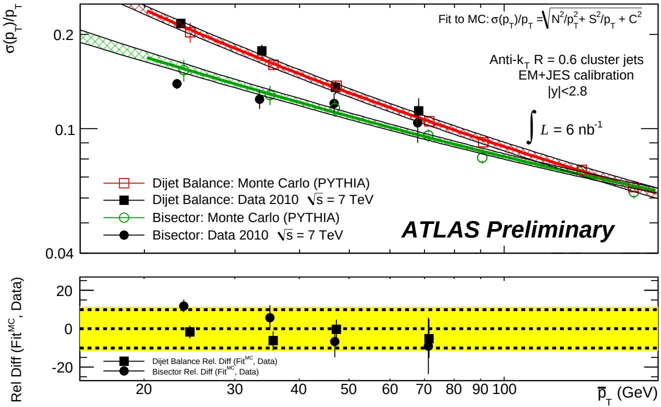
Correcting for acceptance



(schematic, not using ATLAS MC)

- Instead of the MC acceptance (black points), it was a chosen to use the analytic formula (red curve).
- The difference is negligible beyond $n_{ch} = 4$, where acceptance reaches 1.
- In the first 3-4 bins of n_{ch} the difference is $\lesssim 1\%$.

In situ JER validation



14% relative uncertainty in the amount of smearing, based on in-situ analysis.
[ATLAS-CONF-2010-054]

Systematic uncertainty on jet p_T