v_2^{jet} - significance and p-values

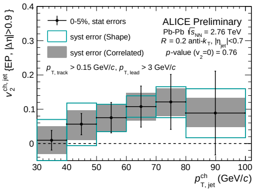
ALICE
A JOURNEY OF DISCOVERY

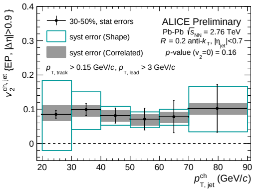
Redmer Alexander Bertens
Utrecht University

May 15, 2014

Problem at hand:

- How can we quantify the significance of the v_2^{jet} result?





Sources of uncertainty

- Statistical (uncorrelated)
- Shape uncertainty (related to steepness of jet spectra)
- Correlated uncertainty (global correlation)

Goal: devise a χ^2 statistic and give a p-value for $v_2^{jet} = 0$ in the final figures

- `‘the probability of finding data at least as incompatible with 0 as the actually observed data’`
- Use the statistic-, shape- and correlated uncertainty in a sensible way

Modified χ^2

Following `arXiv::0801.1665`, hypothesis $v_2^{jet} = 0$, assuming Gaussian distributed uncertainties

- Approach A

$$\tilde{\chi}^2(\epsilon_{corr}) = \left[\left(\sum_{i=1}^{n} \frac{(v_{2\,i} + \epsilon_{corr}\sigma_{corr,i})^2}{\sigma_i^2} \right) + \epsilon_{corr}^2 \right]$$

  where σ_i is the quadratic sum of the statistical and shape uncertainty, $\sigma_{corr,i}$ is the correlated uncertainty and $\epsilon_{corr,i}$ is a free parameter to be minimized

- Approach B

$$\tilde{\chi}^2(\epsilon_{corr}, \epsilon_{shape}) = \left[\left(\sum_{i=1}^{n} \frac{(v_{2\,i} + \epsilon_{corr}\sigma_{corr,i} + \epsilon_{shape,i}\sigma_{shape,i})^2}{\sigma_i^2} \right) + \epsilon_{corr}^2 + \epsilon_{shape}^2 \right]$$

  where σ_i is the statistical uncertainty, $\sigma_{shape,i}$ is the shape uncertainty, $\sigma_{corr,i}$ is the correlated uncertainty and $\epsilon_{corr,i}$, $\epsilon_{shape,i}$ are free parameters to be minimized

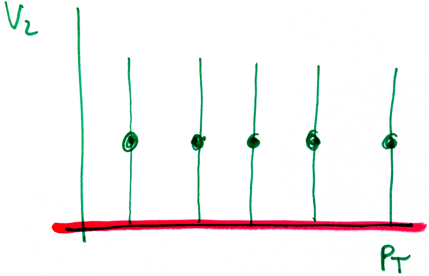
- Approach C somewhat preferred

$$\tilde{\chi}^2(\epsilon_{corr}, \epsilon_{shape}) = \left[\left(\sum_{i=1}^{n} \frac{(v_{2\,i} + \epsilon_{corr}\sigma_{corr,i} + \epsilon_{shape})^2}{\sigma_i^2} \right) + \epsilon_{corr}^2 + \frac{1}{n} \sum_{i=1}^{n} \frac{\epsilon_{shape}^2}{\sigma_{shape,i}^2} \right]$$

  where σ_i is the statistical uncertainty, $\sigma_{b\,i}$ is the shape uncertainty, $\sigma_{c\,i}$ is the correlated uncertainty and $\epsilon_{corr,i}$, $\epsilon_{shape,i}$ are free parameters to be minimized

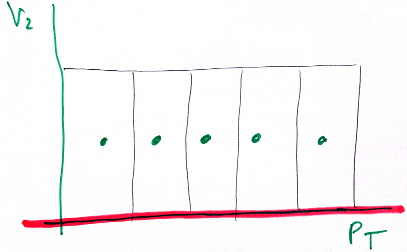
Contributions to χ^2

To get a feeling of differences between the methods, some cartoons



Statistical uncertainties

- In approaches A, B, C 'moving each point down with 1 σ' adds $\sqrt{n}$ to χ^2_{stat} (so in this cartoon $\sqrt{5}$)

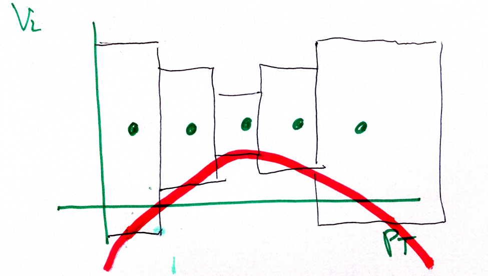


Correlated uncertainties

- In approaches A, B, C 'moving each point down with 1 σ' adds 1 to χ^2_{corr}, justified as points can only move up or down as an ensemble

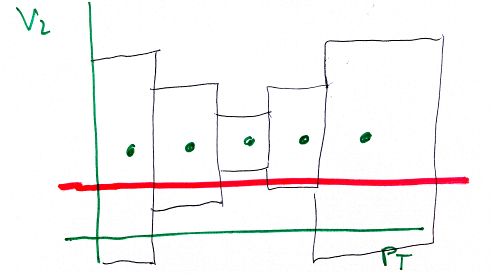
Contributions to χ^2 - shape uncertainty

The difference enters with the shape uncertainty



Shape uncertainties [B]

- In approache B moving each point 'down with 1 σ' adds 1 to χ^2_{shape}, however this is not a very realistic scenario



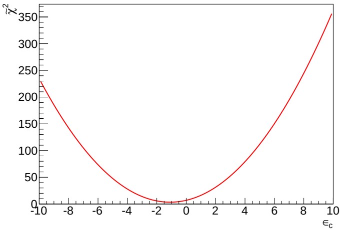
Shape uncertainties [C]

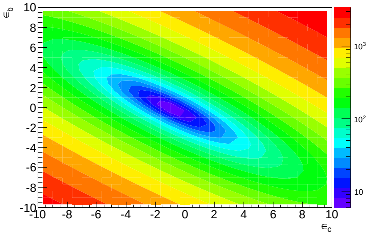
- In approaches C moving each point 'down with ϵ_{shape}' adds 1 to χ^2_{shape}, more physical result
- χ^2 contribution of ϵ_{shape} is the mean square deviation in terms of σ_{shape} over all points:

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\epsilon^2_{shape}}{\sigma^2_{shape,i}}$$

Hands on: minimizing ϵ_i

$\tilde{\chi}^2$ is found by minimizing equations 1, 2 or 3 w.r.t. ϵ_{corr} or $\epsilon_{corr}, \epsilon_{shape}$





Examples of $\tilde{\chi}^2$ functions
$\tilde{\chi}^2$ can be evaluated using either the full range of the results of using just a subset (i.e. omitting the first and last point). Corresponding *p*-values are different!

$p(v_2^{jet} = 0)$ different methods

p-values (`TMath::Prob(`$\tilde{\chi}^2$`, ndf)`) are given for method A, B, and C and different p_T ranges

0-5% collision centrality

method	$30 < p_T < 100$ GeV/c	$40 < p_T < 80$ GeV/c
p-value method A	0.761497	0.560675
p-value method B	0.521928	0.570259
p-value method C	0.586077	0.437825

30-50% collision centrality

method	$20 < p_T < 90$ GeV/c	$30 < p_T < 70$ GeV/c
p-value method A	0.161504	0.0771219
p-value method B	0.0722924	0.0051155
p-value method C	0.437825	0.132771

red values: preferred method

Conclusion

Three approaches to obtaining a p-value have been presented

- p-values are somewhat different but qualitatively similar
- Significance of results is not expected to increase

How to present

My preferred solution

- Present as-is, indicate p-value for $v_2^{jet} = 0$ on plots to have a clear message
- Use the p-value based on the 'narrow' p_T range, indicate this on the plot
- Also indicate systematic checks on poster: good to start discussion with e.g. ATLAS