Powerweek Data Analysis

Helmholtz Research School for Quark Matter Studies in Heavy Ion Collisions

Klaus Reygers Physikalisches Institut Universität Heidelberg H-QM Powerweek Data Analysis I November 2010

Contents

Part I: Invariant mass analyses (example: π⁰ analysis)

- Kinematics
- Acceptance and efficiency
- Bin-shift correction
- Effects of energy scale uncertainties
- Part II: Statistics
 - Error propagation
 - Maximum likelihood method
 - Least squares method

Slides and initial versions of the macros for the hands-on tutorials: <u>http://www.physi.uni-heidelberg.de/~reygers/lectures/2010/Powerweek/pw.tgz</u>

Part I: Invariant mass analyses (example: π⁰ analysis)

Lorentz Invariant Phase Space Element

Lorentz transformation for a momentum space element $d^{3}\vec{p} = dp_{x} \cdot dp_{y} \cdot dp_{z}$

$$dp_{x}dp_{y}dp_{z} = \frac{\partial(p_{x}, p_{y}, p_{z})}{\partial(p_{x}', p_{y}', p_{z}')} \cdot dp_{x}'dp_{y}'dp_{z}' = \frac{E}{E'} \cdot dp_{x}'dp_{y}'dp_{z}'$$

Lorentz invariant momentum space element:

Invariant cross section:

$$\frac{d\sigma}{d^{3}\vec{p}/E} = E\frac{d\sigma}{d^{3}\vec{p}}$$

 $\frac{d^{3}\vec{p}}{E}$

4 Powerweek Data Analysis

Invariant Cross Section

$$\frac{d^{3}\sigma}{d\vec{p}^{3} / E} = E \frac{d^{3}\sigma}{d\vec{p}^{3}} = E \frac{1}{p_{T}} \frac{d^{3}\sigma}{dp_{T} dp_{L} d\varphi}$$

$$\frac{\frac{dp_{L}}{dy} = m_{T} \cosh y = E}{=} \frac{1}{p_{T}} \frac{d^{3}\sigma}{dp_{T} dy d\varphi}$$
symmetry in φ

$$= \frac{1}{2\pi p_{T}} \frac{d^{2}\sigma}{dp_{T} dy}$$

Integral of the invariant cross section:



Average number of particles per event 5 Powerweek Data Analysis

for the considered events

$$p + p \rightarrow \pi^{0} + X \text{ at } \sqrt{s} = 200 \text{ GeV}$$

Klaus Reygers

Invariant Mass

Consider the decay of a particle into two daughter particles:
In terms of CPU time this formula is better than the one with
$$\cos \vartheta$$

Invariant mass M : $M^2 = \left[\begin{pmatrix} E_1 \\ \vec{p}_1 \end{pmatrix} + \begin{pmatrix} E_2 \\ \vec{p}_2 \end{pmatrix} \right]^2 = (E_1 + E_2)^2 - (\vec{p}_1 + \vec{p}_2)^2$
 $= \frac{E_1^2 - \vec{p}_1^2}{m_1^2} + \frac{E_2^2 - \vec{p}_2^2}{m_2^2} + 2E_1E_2 - 2\vec{p}_1\vec{p}_2$
 $= m_1^2 + m_2^2 + 2E_1E_2 - 2p_1p_2\cos\vartheta$
Example: $\pi^0 \det \chi$ $\pi^0 \to \gamma + \gamma$, $m_1 = m_2 = 0$, $E_i = p_i$
rest frame:
 $M = \sqrt{2E_1E_2(1 - \cos\vartheta)}$

Invariant Mass



Two Ways to Measure Photons (I): With Calorimeters ...

WA98 - LEDA event display



Pb + Pb 160 A GeV central Nov. 3, 1995 - Pun 0001 - Evt Nr. 00001

- Two types of calorimeters
 - homogeneous calorimeters (e.g. lead glass)
 - sampling calorimeters (alternating layers of absorber material and scintillators)
- Energy resolution improves with increasing energy

$$\frac{\sigma_E}{E} \approx \frac{\sqrt{N_{tot}}}{N_{tot}} = \frac{1}{\sqrt{N_{tot}}} \propto \frac{1}{\sqrt{E}}$$

• Good homogeneous calorimeters reach $\frac{\sigma_E}{E} \approx \frac{6\%}{\sqrt{E/GeV}}$

Two Ways to Measure Photons (II): ... and via Photon Conversions



- Very good momentum resolution at low p_T
- However, photon conversion probability typically small (~ 8% in Alice)
- 9 Powerweek Data Analysis

Klaus Reygers

π^0 Decay Kinematics with Mathematica (1/7)



z component of the momentum of photon 1 as a function of the decay angle in the CMS:



Same for photon 2:

ln[4] = p2zcms = -p1zcms $out[4] = -\frac{1}{2}mCos[\Theta]$

10 Powerweek Data Analysis

π^0 Decay Kinematics with Mathematica (2/7)

Lorentz transformation of the z components of the momentum vector of the two decay photons

out[5]= $\frac{plzlab = \gamma (plzcms + \beta Egamcms)}{\frac{m\beta}{2} + \frac{1}{2} m \cos [\theta]}$ $\sqrt{1 - \beta^2}$

 $[n[6]:= p2zlab = \gamma (p2zcms + \beta Egamcms)$ $\frac{m\beta}{2} - \frac{1}{2} m Cos[\theta]$

Out[6]=

 $\frac{\frac{\mathfrak{m}\beta}{2} - \frac{1}{2} \mathfrak{m} \operatorname{Cos} \left[\Theta\right]}{\sqrt{1 - \beta^2}}$

3-vectors of the decay photons in the lab system

In[7]:= **pllab** = {Egamcms Sin[Θ], 0, plzlab} Out[7]= $\left\{\frac{1}{2} \text{ m Sin}[\Theta], 0, \frac{\frac{m\beta}{2} + \frac{1}{2} \text{ m Cos}[\Theta]}{\sqrt{1 - \beta^2}}\right\}$

 $|n[8]:= p2lab = \{-Egamcms Sin[\Theta], 0, p2zlab\}$

Out[8]=
$$\left\{ -\frac{1}{2} \operatorname{m} \operatorname{Sin}[\Theta], 0, \frac{\frac{\operatorname{m}\beta}{2} - \frac{1}{2} \operatorname{m} \operatorname{Cos}[\Theta]}{\sqrt{1 - \beta^2}} \right\}$$

11 Powerweek Data Analysis

π^0 Decay Kinematics with Mathematica (3/7)

Asymmetry of the photon energies in the lab system:

```
\alpha = \operatorname{Simplify}[\operatorname{Abs}[\operatorname{Norm}[\operatorname{pllab}] - \operatorname{Norm}[\operatorname{p2lab}]] / (\operatorname{Norm}[\operatorname{p1lab}] + \operatorname{Norm}[\operatorname{p2lab}]), \{m > 0, \beta > 0, \beta < 1, \gamma > 1, \theta \ge 0, \theta \le \operatorname{Pi}\}]
\operatorname{Out}[9]= \beta \operatorname{Abs}[\operatorname{Cos}[\theta]]
```

Cosine of the opening of the two decay photons in the lab system (costhetalab)

```
costhetalab[\Theta] =
 In[10]:=
             TrigFactor[Simplify[pllab.p2lab / (Norm[pllab] Norm[p2lab]),
                 \{\mathbf{m} > \mathbf{0}, \beta > \mathbf{0}, \beta < \mathbf{1}, \gamma > \mathbf{1}, \theta \ge \mathbf{0}, \theta \le \mathbf{Pi}\}]
           2 - 3 \beta^2 + \beta^2 Cos [ 2 \Theta ]
Out[10]=
             -2 + \beta^2 + \beta^2 \cos [2\theta]
           Simplify [costhetalab[\theta] /. {Cos[2x] -> 2Cos[x]^2 - 1}]
 In[11]:=
         \frac{1-2\beta^2+\beta^2 \operatorname{Cos}[\theta]^2}{-1+\beta^2 \operatorname{Cos}[\theta]^2}
Out[11]=
                                                                                                                                         \cos\vartheta_{lab} = \frac{\beta^2 \cos^2 \vartheta^{*2} + 1 - 2\beta^2}{\beta^2 \cos^2 \vartheta^{*2} - 1}
                                                                                                                    opening angle
                                                                                                                    in the lab system
 12 Powerweek Data Analysis
                                                                                                                                                                                    Klaus Reygers
```

 $\alpha \coloneqq \frac{\left|E_1 - E_2\right|}{E_1 + E_2}$

 $\alpha = \beta \left| \cos \vartheta^* \right|$

π^0 Decay Kinematics with Mathematica (4/7)

Write β in terms of the mass m and the momentum p of the particle

$$\ln[12] := \operatorname{Simplify}\left[\operatorname{costhetalab}\left[\Theta\right] / . \beta \to 1 / \sqrt{1 + \frac{m^2}{p^2}}\right]$$

Out[12]= $\frac{-2 m^2 + p^2 - p^2 \cos [2 \theta]}{2 m^2 + p^2 - p^2 \cos [2 \theta]}$

```
In[13]:= t2 = Simplify[t1 /. {Cos[2 x_] → 2 Cos[x]^2 - 1}]
Out[13]= -\frac{m^2 - p^2 + p^2 Cos[\theta]^2}{m^2 + p^2 - p^2 Cos[\theta]^2}
```

```
In[14]:= ctlab[\theta_{-}] = t2 /. {a_^2 - a_^2 Cos[x_]^2 → a^2 Sin[x]^2}

Out[14]= -\frac{m^2 - p^2 + p^2 Cos[\theta]^2}{m^2 + p^2 Sin[\theta]^2}
```

π^0 Decay Kinematics with Mathematica (5/7)



 9^* = 90 degrees corresponds to the minimum opening angle in the lab system

π^0 Decay Kinematics with Mathematica (6/7)



Most decays have opening angles close to the minimum

π^0 Decay Kinematics with Mathematica (7/7)

Also plot the minimum opening angle as a function of the momentum of the mother particle



Minimum opening angle as a function of the momentum of the neutral pion

16 Powerweek Data Analysis

Klaus Reygers

Asymmetry Cut



In the π^0 rest frame $\cos \theta^*$ is uniformly distributed

With E_1 and E_2 denoting the decay photon energies in the lab frame the energy asymmetry is defined as:

$$\alpha := \frac{\left|E_1 - E_2\right|}{E_1 + E_2}$$

We have shown that:

 $\alpha = \beta \left| \cos \theta^* \right|$

velocity of the pion in the lab frame (in units of c) $\int \frac{1}{\cos \theta^*}$

 β is typically close to unity. So for photons pairs from a π^0 decay the asymmetry α is approximately uniformly distributed between 0 and 1.

An Example from Phenix



- The asymmetry cut can help to improve the signal/background ratio of the π^0 peaks
 - Steeply falling spectra lead to large asymmetries in the combinatorial background
- Comparing π⁰ yields for different asymmetry cuts turned out to be a very useful systematic check in Phenix (e.g., no asymmetry cut vs. α < 0.7)
- **18** Powerweek Data Analysis

Combinatorial Background: Event Mixing

- Calculate inv. mass for combinations where photon 1 comes from the current event and photon 2 comes from an old event
- Make sure that the old event has the same global properties as the current one.
 Typically events are categorized according to
 - Event multiplicity
 - Position of the vertex
 - Angle w.r.t. reaction plane
- In Phenix, even within a vertex class the photon momentum vectors were recalculated with respect to a new vertex z_{mix} = (z_{current} + z_{old})/2

Peak Extraction: π^0 peak in Au+Au at $\sqrt{s} = 200$ GeV



Peak Extraction: η peak in Au+Au at \sqrt{s} = 200 GeV



η Peaks in Au+Au at \sqrt{s} = 200 GeV for various p_T Bins



Klaus Reygers

Peak Extraction: Statistical Error

Simple example: Measure count rate of a radioactive source in the presence background Measurement with Signal + Background: $O = \tilde{S} + \tilde{B}$ Background measurement: BExtracted signal: $S = O - B = \tilde{S} + \tilde{B} - B$ Statistical Error: $\sigma^2 = S + 2B$ In case of the π^0 yield extraction the background is estimated as $B = f \cdot \tilde{M}$ Background scaling factor

Extracted signal:
$$S = O - f \cdot M = \tilde{S} + \tilde{B} - f \cdot M$$

The statistical error of the extracted yield is then given by:

$$\sigma^2(S) = S + B + \sigma^2(f)M^2 + f^2M$$

Hands-on exercise 1: peak extraction

23 Powerweek Data Analysis

Geometrical Acceptance

Geometric acceptance:

$a = \frac{\#\pi^0 \text{ with both photons on the detector surface}}{\#\pi^0 \text{ generated in a certain } \eta \text{ window}}$

Low $p_T \pi^0$ on average have larger opening angles and therefore more like escape detection. Thus, the π^0 acceptance typically increases with p_T .



Hands-on exercise 2: acceptance calculation

Efficiency (Correction of Detector Effects)

Efficiency: (or better: correction function, it can be greater than unity)

$\varepsilon(p_T) = \frac{\text{reconstructed } \pi^0 \text{ spectrum}}{\text{true } \pi^0 \text{ spectrum in the acceptance}}$

Since the true spectrum is not know one can start with a reasonable guess an then iterate until the efficiency converges

Efficiency accounts for

- Any kind of signal loss
 - due to analysis cuts (and possibly dead detector areas)
 - intrinsic limitations (e.g., conversion probability in the conversion method)
- Distortions of the signal due to limited detector resolution
 - In particular important in case of steeply falling spectra
 - In heavy-ion collisions the presence of other particles can lead to additional distortions (calorimeter: showers from different particles start to overlap and merge)

Hands-on exercise 3: efficiency calculation

25 Powerweek Data Analysis

Limited Resolution and Steeply Falling Spectra (I)



- In case of steeply falling spectra finite energy resolution leads to an overall shift of the yield toward higher transverse momenta
- The "efficiency" can thus be larger than unity

Limited Resolution and Steeply Falling Spectra (II)



A consequence of limited energy resolution:

The measured peak position of a correctly calibrated detector lies above the nominal meson mass

A Tool to Study Effects of Large Detector Occupancy: Embedding



- In calorimeters showers are found by merging calorimeter cells ("towers") into so-called clusters
- Simulated showers are
 - analyzed on the empty detector
 - merged with a real event and the analyzed
- From this one can determine the multiplicity dependent energy smearing effect

Lorentz invariant yield:



Bin Shift Correction (1/6): Where to Plot Your Data Points within Wide Bins?

Lafferty, Wyatt, Nucl. Instr. and Meth. A 355, 541, 1995

What's the problem?

When measuring hadron yields as a function of p_T we have to deal with steeply falling spectra. Lack of statistics forces one to use wide bins at high p_T . Let f(x) denote the true spectrum. The measured quantity then is

$$\langle g_{meas} \rangle = \frac{1}{\Delta x} \int_{x_1}^{x_2} g(x) \, dx$$
 where $\Delta x = x_2 - x_1$

The question is where to plot the data point in this case. One frequently observes one of the following two methods

- the data point is plotted at the bin center $x_c = x_1 + \Delta x / 2$ the data point is plotted at the center-of-gravity within the bin: $\overline{x} = \frac{\int_{x_1}^{x_2} x g(x) dx}{\int_{y_2}^{y_2} g(x) dx}$

Both methods are wrong!

30 Powerweek Data Analysis

Bin Shift Correction (2/6): An Example



Data points at the bin center and the centerof-gravity (,barycentre') both don't lie on the curve!

The correct position can be calculated by solving (either analytically or numerically):

$$g(x_{lw}) = \left\langle g_{meas} \right\rangle \equiv \frac{1}{\Delta x} \int_{x_1}^{x_2} g(x) \, \mathrm{d}x$$

$$\uparrow_{lw = \text{ large width}}$$

g(x) is a priori not known so one has to work with a good approximation of g(x)

Bin Shift Correction (3/6): Analytical Solution for an Exponential

Typically an exponential is a good approximation within a bin and so the following result is useful:

$$g(x) = ae^{-bx} \quad \Rightarrow \quad x_{lw} = x_1 + \frac{1}{b} \left\{ \ln(b\Delta x) - \ln(1 - e^{-b\Delta x}) \right\}$$

Note that in the special case that g(x) varies linearly with x there is no ambiguity as to where to plot the data point:

$$g(x) = ax + b \implies x_{lw} = \overline{x} = x_c$$

So in case of small bins where the spectrum is well approximated by a linear function the problem disappears

Bin Shift Correction (4/6): Data Presentation: Three Examples with Drawbacks



A well defined solution would be to publish a histogram. For comparison with theory the theory curve would have to be binned in the same way.

Drawbacks:

- hard to read off the shape of the underlying distribution
- difficult to compare to other data with different binning

Better: Add points plotted at x_{lw} (easier to see the shape)



Vertical lines can be removed. However, horizontal lines indicating the bins can easily be confused with vertical error bars (and vice versa)

Bin Shift Correction (5/6): Suggestion by Lafferty and Wyatt

Use short vertical lines to indicate bins:



Bin Shift Correction (6/6): The Phenix Solution

The data point can also be moved along the ordinate (*y* axis). Lafferty and Wyatt do not recommend that

- Philosophical argument: The primary experimental result (g_{meas}) is modified
- Sum of all bins no longer gives the total number of entries

For Phenix the key argument to nevertheless do it this way was that it is the only methods that allows a straightforward calculation of ratios between spectra (e.g. R_{AA})

Here is the procedure:

1. Fit the raw spectrum (i.e., the spectrum not corrected for the bin shift effect) with g(x)2. For each bin calculate the correction as

$$g_{meas}^{corr} = g_{meas} / r$$
 where $r = \frac{\frac{1}{\Delta x} \int_{x_1}^{x_2} g(x) dx}{g(x_c)}$

3. Repeat steps 1. and 2. until $r \approx 1$ (typically less than ~ 5 iterations needed)

Numerical example (RHIC): $g(x) \propto 1/x^7$, $x_1 = 14$, $x_2 = 16 \implies r = 1.042$

Hands-on exercise 4: bin shift correction

35 Powerweek Data Analysis

Klaus Reygers

Sources of Systematic Errors

TABLE I. Summary of the dominant sources of systematic errors on the PbSc and PbGl π^0 yields and total errors on the combined measurements. The error ranges are quoted for the lowest to highest p_T values.

Source		Syst. error PbSc	Syst. error PbGl	
Yield extraction		10%	6%-7%	
Yield correction		8%	8%	
Energy scale		3%-11%	7%-13%	
Total error (%)	Stat.	Syst.	Normalization Central Peripheral	
Comb. π^0 spectra	2-40	10-17	5	5
R_{AA}	2-45	11–22	14	30

Phenix, pi0 in Au+Au at 200 GeV: <u>Phys.Rev.Lett.91:072301,2003</u>

Hands-on exercise 5: Energy scale uncertainties

In neutral pion measurements with calorimeters the uncertainty of the energy scale is typically among the dominant sources of systematic errors

Hands-On Exercise 1:

Extract Peak Content and Statistical Error



- Goto to "hands-on/inv_mass/01_peak_extr_CuCu"
- Run macro "real_mix_fit_v0.C"
- Modify it to extract the peak content of the π^0 and η peak with statistical errors
- **37** Powerweek Data Analysis

Klaus Reygers



- Starting point "hands-on/inv_mass/02_pi0_acc/pi0_toy_mc_v0.C"
- The macro simulates π⁰ decays
- Calculate the acceptance for π^0 s with $|\eta| < 0.5$ in the range $0 < p_T < 10$ GeV/c for a virtual calorimeter "VCal" which covers the full azimuth and $|\eta| < 0.5$
- Make the macro more efficient at high p_T by using a flat p_T distribution and p_T weights
- How important is the use of p_T weights in the acceptance calculation?
- Modify the macro to calculate the π^0 acceptance for the Alice detector "PHOS" $(|\eta| < 0.12, \Delta \varphi = 100^\circ)$

Hands-On Exercise 3: Calculate the Efficiency of VCal

- WW/
- The energy resolution of calorimeters can be parameterized as

$$\frac{\sigma_E}{E} = \sqrt{\left(\frac{a}{\sqrt{E/\text{GeV}}}\right)^2 + b^2} \equiv \frac{a}{\sqrt{E/\text{GeV}}} \oplus b$$

- The first term is related to the Poisson statistics of the electromagnetic shower (e.g., fluctuations in the number of produced scintillation photons)
- The seconds term corresponds to detector noise or tower-by-tower variations of the gain

VCal efficiency:

- Calculate the efficiency of VCal
 - a = 8% and b = 5%, lower energy threshold of VCAL: 100 MeV
 - Asymmetry cut $\alpha < 0.7$
- Try different functional forms for the input spectrum. How sensitive is the calculated efficiency to the shape of the input spectrum?

39 Powerweek Data Analysis

Klaus Reygers

Hands-On Exercise 4: Calculate Bin Shift Corrections



- The shown spectrum is not corrected for bin shift. It can be parameterized with a so-called Hagedorn function.
- Calculate (numerically)
 - The vertical bin shift correction for the bin from 6-7 GeV
 - The horizontal bin shift correction (i.e., x_{lw}) for the same bin
- 40 Powerweek Data Analysis

Hands-On Exercise 5: Systematic Errors Related to Energy Scale Uncertainties

Calculate the uncertainty of the yield resulting from a 1% uncertainty of the p_T scale for the following two functional forms. Plot the ratio of the p_T spectrum for the modified and correct energy scale.

a)
$$\frac{dN}{dp_T} = A \exp(-bp_T), \ b = 6 \ (\text{GeV/c})^{-1}$$
 b) $\frac{dN}{dp_T} = \frac{A}{p_T^n}, \ n = 7$

Hint:
$$\tilde{p}_T = (1 + \varepsilon) p_T$$
, $dN / d\tilde{p}_T = \dots$

- Monte Carlo exercise:
 - Employ the macro for the π^0 efficiency to estimate the uncertainty of the pi0 yields resulting from a 1% uncertainty of the photon energy scale. Use the two functions above as π^0 input spectra.

Hands-On Exercise 6: The final project (I): Putting it all together

Objective:

- Analyze the pi0_vcal_data.root data set (in directory "06_pi0_analysis")
- Determine a fully corrected invariant π^0 yield as a function of p_T
- Answer the following questions
 - What is the average π^0 multiplicity per event ?
 - What are the parameters of a Hagedorn fit in the range $1 < p_T < 8$ GeV/c

Hands-On Exercise 6: The final project (II): Putting it all together

Intermediate steps

- extract the π^0 yields in different p_T intervals
- apply acceptance and efficiency corrections (use an alpha cut of 0.7)
- apply a bin shift corrections

Detailed Instructions

- Form analysis teams of ~ 4 people to share the work
- Starting point: macro ,06_pi0_analysis/ana_v0.C'
- Task 1a) Add a subroutine that calculates the invariant mass
- Task 1b) ana_v0.C only contains a loop for mixed events, add the corresponding loop for real photon pairs
- Task 1c) Add a 2D histogram ,pT vs. m_{γγ}', write this histogram into an output file
- Task 2: Write a peak extraction macro that reads the output of ana.C (starting point: macro ,06_pi0_analysis/peak_extract_v0.C)
- Task 3: Write a macro that applies the corrections to the yields
- Compare your final spectrum with that of other analysis teams