

# **Bin-width optimization in spectral analyses of the systems formed in high energy collisions**

**Aman Singh Katariya**

**Roll No: MS16088**

*A dissertation submitted for the partial fulfilment  
of BS-MS dual degree in Science*

Under the guidance of

**Dr. Satyajit Jena**



**April 2021**

**Indian Institute of Science Education and Research Mohali  
Sector - 81, SAS Nagar, Mohali 140306, Punjab, India**



## Certificate of Examination


This is to certify that the dissertation titled **“Bin-width optimization in spectral analyses of the systems formed in high energy collisions”** submitted by **Aman Singh Katariya** (Reg. No. MS16088) for the partial fulfillment of BS-MS dual degree programme of the Institute, has been examined by the thesis committee duly appointed by the Institute. The committee finds the work done by the candidate satisfactory and recommends that the report be accepted.



Dr. Ananth Venkatesan



Dr. Kinjalk Lochan



Dr. Satyajit Jena  
(Supervisor)

Dated: May 23, 2021



## Declaration

The work presented in this dissertation has been carried out by me under the guidance of Dr. Satyajit Jena at the Indian Institute of Science Education and Research Mohali.

This work has not been submitted in part or in full for a degree, a diploma, or a fellowship to any other university or institute. Whenever contributions of others are involved, every effort is made to indicate this clearly, with due acknowledgment of collaborative research and discussions. This thesis is a bonafide record of original work done by me and all sources listed within have been detailed in the bibliography.



Aman Singh Katariya  
(Candidate)

Dated: May 23, 2021

In my capacity as the supervisor of the candidate's project work, I certify that the above statements by the candidate are true to the best of my knowledge.



Dr. Satyajit Jena  
(Supervisor)



# Acknowledgement

First and foremost, I would like to thank my thesis supervisor, Dr. Satyajit Jena, without whose help and supervision, this thesis would have never been possible. The discussions that I had with him has enhanced my capabilities as a researcher.

I would also like to thank Mr. Rohit Gupta (Ph.D. student) profusely, who constantly helped me throughout my thesis. I also thank Mr. Rutik and Miss Nishat (Ph.D. students), who encouraged me and helped me. I thank my labmates who helped me with regular discussions while doing this work. Lastly, I thank my family and friends for their kind help and support throughout my thesis work.

Aman Singh Katariya

MS16088

IISER Mohali.





# List of Figures

1.1	The Standard Model of elementary particles[MissMJ ],[PDG ]. . . . .	2
1.2	$\alpha_s$ v/s E. $\alpha_s = g_s^2/4\pi$ [aus tirol ],[PDG ]. . . . .	4
1.3	Different stages of high energy nuclear collisions[Chaudhuri 14]. . . . .	5
1.4	A collision event in side view and beam view. Here, two nuclei A and B collide with nuclei $b$ distance apart. . . . .	7
1.5	A space time representation of evolution of a collision.[Chaudhuri 14] . . .	8
1.6	Variation of $\eta$ w.r.t. $\theta$ . As the polar angle increases, the pseudorapidity decreases. . . . .	11
2.1	A particle detector used in ATLAS experiment at LHC [det a][det c]. . . . .	14
3.1	Gaussian model . . . . .	22
3.2	Gaussian model log-scaled . . . . .	23
3.3	Exponential model . . . . .	23
3.4	Exponential model log-scaled . . . . .	24
3.5	A power law distribution (100000 samples) with optimal $\Delta = 0.06$ repre- sented on a log scale on the y-axis. . . . .	24
4.1	Cost vs Bin-width graph plotted for fixed bin-width $p_T$ spectra. . . . .	29
4.2	$p_T$ distribution for $\pi^+$ particles in range $ \eta  < 0.5$ for N (= 50K, 250K, 500K, 1M) events in the $p_T$ range of 0-3 GeV with the optimal bin-width $\Delta = 0.1111$ GeV/c. . . . .	30
4.3	$p_T$ distribution for $\pi^+$ particles in range $ \eta  < 0.5$ for N (= 50K, 250K, 500K, 1M) events in the $p_T$ range of 0-3 GeV with the optimal variable bin-width. . . . .	31
4.4	Cost v/s Bin-width graph for variable bin-width distribution. . . . .	32

4.5	A binning comparison for $p_T$ distribution for $\pi^+$ particles in range $ \eta  < 0.5$ for 250K events in the $p_T$ range of 0-3 GeV/c. . . . .	33
5.1	Cost v/s bin-width distribution for pseudorapidity( $\eta$ ) observable. . . . .	37
5.2	$\eta$ distribution for $\pi^+$ particles in range $ \eta  < 6.0$ for N (= 50K, 250K, 500K, 1M) events with $p_T$ cut from 0-3 GeV with the 0.8 optimal fixed bin-width. . . . .	38

# List of Tables

1.1	Fundamental forces and their mediators[Chaudhuri 14]. . . . .	2
4.1	Chi-squared/NDF values for the $p_T$ distributions. . . . .	29
4.2	Chi-squared/NDF values for the $p_T$ distributions with variable bin width. . .	32
4.3	Chi-squared/NDF values for the $p_T$ distributions binning comparison for 250K events. . . . .	34



# Contents

<b>Acknowledgement</b>	<b>i</b>
<b>List of Figures</b>	<b>iv</b>
<b>List of Tables</b>	<b>v</b>
<b>Abstract</b>	<b>ix</b>
<b>1 Introduction</b>	<b>1</b>
1.1 The Standard Model . . . . .	1
1.2 Quarks, Gluons and QGP . . . . .	3
1.3 High Energy collisions . . . . .	4
1.4 Kinematics of HIC . . . . .	6
1.4.1 Lorentz transformation . . . . .	8
1.4.2 Energy-Momentum four-vector . . . . .	8
1.4.3 The Rapidity variable . . . . .	9
1.4.4 Transverse momentum and Invariant yield . . . . .	10
<b>2 Measuring Observable from a collision</b>	<b>13</b>
2.1 Particle detection and measurements . . . . .	13
2.2 Experimental v/s Estimated Obseables . . . . .	15
2.2.1 Boltzmann distribution . . . . .	15
2.2.2 Tsallis distribution . . . . .	16
2.2.3 Pearson distribution . . . . .	16
2.3 Making distributions . . . . .	17

<b>3</b>	<b>Optimizing Bin Width</b>	<b>19</b>
3.1	Bin-width optimization . . . . .	19
3.1.1	Cost function . . . . .	20
3.1.2	Toy models and bin optimization . . . . .	22
3.2	Bin-width optimization for Variable bin-widths . . . . .	24
<b>4</b>	<b>Optimizing transverse momentum spectra</b>	<b>27</b>
4.1	Transverse momentum spectra . . . . .	27
4.1.1	Tsallis distribution for $p_T$ -spectra . . . . .	27
4.1.2	Pearson distribution for $p_T$ -spectra . . . . .	28
4.2	Optimizing bin-width . . . . .	28
<b>5</b>	<b>Optimizing Pseudorapidity spectra</b>	<b>35</b>
5.1	Pseudorapidity spectra . . . . .	35
5.1.1	Tsallis distribution for $\eta$ -spectra . . . . .	35
5.1.2	Pearson distribution for $\eta$ -spectra . . . . .	36
5.2	Optimizing bin-width . . . . .	37
<b>6</b>	<b>Conclusion</b>	<b>39</b>
<b>A</b>	<b>Python code for bin-width optimization</b>	<b>41</b>
A.1	Finding optimal fixed bin width: . . . . .	41
A.2	Finding optimal variable binwidth . . . . .	42
<b>B</b>	<b>PyROOT code for reading data and making histograms</b>	<b>45</b>
B.1	Code for $p_T$ data: . . . . .	45
B.2	Code for $\eta$ distribution: . . . . .	48
	<b>Bibliography</b>	<b>51</b>

# Abstract

To probe any heavy-ion collision(HIC), we need some observables to describe the final state particles. These observables are then fitted with different distribution functions to obtain some characteristic quantities using which comparison can be done with other collisions. The transverse momentum,  $p_T$ , is often used to describe final state particles as they come out of a collision. Various distribution functions can describe the transverse momentum spectra of final-state particles in a heavy-ion collision. The transverse momentum spectra obtained in HIC are often studied using the Tsallis, Boltzmann, Fermi-Dirac, and Bose-Einstein distribution functions. In a multi-source thermal model, it is essential to estimate the parameter as accurately as possible, for which bin-width correlation must be properly understood. This thesis work aims to optimize the bin width selection  $p_T$  distribution to enable a parameter correlation. We define a cost function using the mean integrated squared error function using the spike counts in each  $p_T$  bin separated by some bin-width  $\Delta$  and then minimize the cost function by changing the bin-width  $\Delta$ . In this thesis, we optimize the bin-width value for  $p_T$  spectra and fit it with *Tsallis distribution*. We further check for an optimized variable bin width histogram for  $p_T$  spectra. We also check for other observable such as  $\eta$  and optimize the bin-width for  $\eta$  distributions.





# Chapter 1

## Introduction

### 1.1 The Standard Model

In the universe, all the matter can be broken down into some elementary particles. In 1803, J. Dalton proposed the atomic theory according to which all the matter consisted of an atom which was the smallest individual and indestructible particle. However, in 1897, J. J. Thomson suggested that the atom can further be divided into subatomic particles when he discovered the electron in his cathode ray experiments. In the early 1900s, E. Rutherford, through his gold foil experiments, suggested that these electrons orbit a dense nucleus having a positive charge, which leads to the discovery of protons. In 1932, J. Chadwick discovered that the nucleus is actually made up of protons and neutrons, where these neutrons are neutral charge particles. In the 1960s, physicists discovered that these protons and neutrons were not themselves individual particles but instead, they were made up of quarks, the up and down quarks. The electron, the up quark, and the down quark make up the first generation of the particles present in the Standard model.

The Standard Model describes all the elementary particles and their interaction with the fundamentals forces, namely, electromagnetic force, weak nuclear force, strong nuclear force, and gravity. In the standard model, we have the first generation of particles consisting of the electrons, up and down quarks, and the neutrino particle. However, in the high-energy experiments, some heavy isotopes of these particles were found. These isotopes make the second and the third generation of the particles. Besides these matter elementary particles, there are also the mediator particles that govern the interaction of the matter particles with the fundamental forces. The mediator particles are the photons (governing the

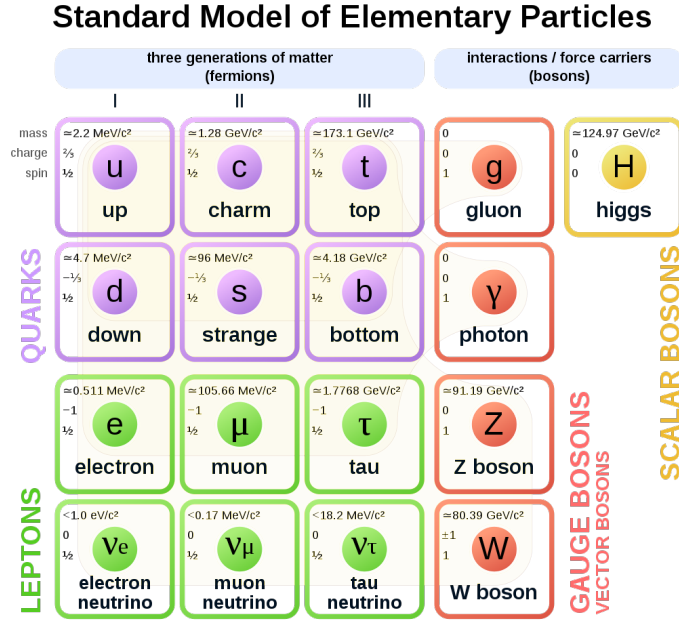


Figure 1.1: The Standard Model of elementary particles[MissMJ ],[PDG ].

electromagnetic interaction), the W and Z boson (governing the weak nuclear interaction), the gluons (governing the strong nuclear interaction), and the recently discovered Higgs boson (related to mass acquisition). Table 1.1 below shows the four fundamental forces along with their mediator particles, their relative strengths, and their interaction range. All these elementary matter particles also have their own anti-particle. The first generation particles are the lightest and do not decay; however, the heavier particles of the second generation and third generation decay into the first generation particles through mediator particles.

Interactions	Theoretical Description	Associated Particles	Relative Strength	Range of Interaction (m)
Strong	QCD	Gluons	$10^{38}$	$10^{-15}$
Electromagnetic	QED	Photons	$10^{36}$	infinity
Weak	Electro-weak	$W^\pm, Z$	$10^{25}$	$10^{-18}$
Gravitational	General Relativity	Graviton (hypothesised)	1	infinity

Table 1.1: Fundamental forces and their mediators[Chaudhuri 14].

In nuclear physics, the particles and their interactions are studied at zero temperature and densities of the order of nucleon densities. However, in high-energy experiments, the nu-

clear matter reaches a very high temperature and densities. At such extreme conditions of temperature and densities, the nuclear matter breaks into its constituent particles, namely, quarks and gluons. This hot soup of quarks and gluons is called "*Quark-Gluon Plasma*" or *QGP*.

## 1.2 Quarks, Gluons and QGP

The nuclear matter in extreme condition of temperature and pressure cannot contain its constituents particles together as protons and neutrons, rather they break down into quarks and gluons. This state of quarks and gluons interacting asymptotically freely is called Quark-Gluon Plasma. The quarks and gluons interact by strong nuclear force and have a very defining property of *color charge quantum number*. This makes their interaction different from other elementary particles and are described by using Quantum ChromoDynamics (QCD) lagrangian which is given as,

$$\mathcal{L} = -\frac{1}{4}F_a^{\mu\nu}F_{\mu\nu}^a + \sum_{\text{flavors}} \left[ i\bar{\psi}\gamma^\mu \left( \partial_\mu - ig\frac{\lambda_a}{2}A_\mu^a \right) \psi - m\bar{\psi}\psi \right] \quad (1.1)$$

with,

$$F_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + gf_{bc}^a A_\mu^b A_\nu^c \quad (1.2)$$

where,  $A_\mu^a$  represents the Gluon field of color  $a$  ( $= 1$  to  $8$ ), and  $m$  denotes the *bare quark mass*,  $f_{bc}^a$  is the structure constant of the group,  $\psi$  is the quark spinor, and  $g$  is the strong coupling constant [Chaudhuri 14].

***QGP is defined as the thermalized quarks and gluons state, where quarks and gluons are asymptotically free to move and interact over a nuclear volume instead of the nucleonic volume.***

QGP is the deconfined state of strongly interacting matter[Chaudhuri 14]. At low density or low temperature quarks are confined inside the hadronic volume and at high density or at high temperature, these quarks become deconfined. This confinement and deconfinement is explained using "*running coupling constant* which is a measure of the coupling strength of the quarks and gluons in hadrons. After solving the Lagrangian equation and renormalising the coupling constant in its condensed form can be written as,

$$\alpha_s(Q^2) = \frac{1}{\beta_0 \ln Q^2/\Lambda^2} \quad (1.3)$$

where  $\Lambda$  is the QCD scale parameter ( $\approx 200$  MeV) and  $\beta_0$  is given as,

$$\beta_0 = \frac{1}{4\pi} \left( 11 - \frac{2N_f}{3} \right) \quad (1.4)$$

where  $N_f$  is the total number of flavors present.

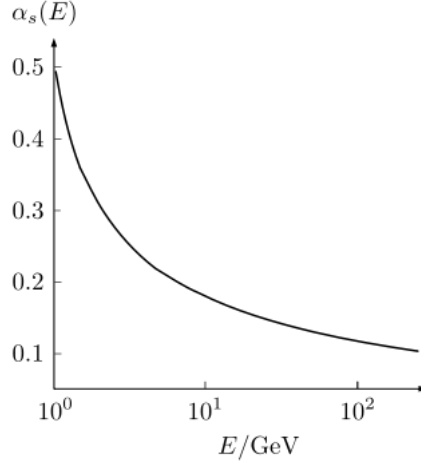


Figure 1.2:  $\alpha_s$  v/s  $E$ .  $\alpha_s = g_s^2/4\pi$  [aus tirol ],[PDG ].

Figure 1.2 shows the variation of the running coupling constant with the energy  $E$  (in GeV). As the energy increases, the coupling strength reduces. At low enough value, the quarks and gluons will become deconfined. And when the energy is lowered again, the running coupling constant goes high, and the quarks and gluons become confined again. A particular quark is not able to identify its partner quark, so the new quarks formed are different than those before.

### 1.3 High Energy collisions

During a high-energy collision, a large amount of energy is confined in a very small space for a very brief time period, approximately of microseconds, making the energy density in that region very high. This high energy density can reach the order higher than the nuclear energy density and often results in the formation of a deconfined fluid-like state, known as Quark-Gluon Plasma (QGP). This is one of the most important characteristics of high-energy collisions. In high-energy collisions, two ultra-relativistic nuclei undergo inelastic collisions resulting in the formation of QGP, which later results in the formation of new particles. For the two heavy nuclei, such as Pb or Au, the collisions are called heavy-ion

collisions(HIC) or AA-collisions, and for Hydrogen like nuclei, the collisions are stated as pp-collisions. Figure 1.3 shows a schematic diagram of the Hydrodynamic model with

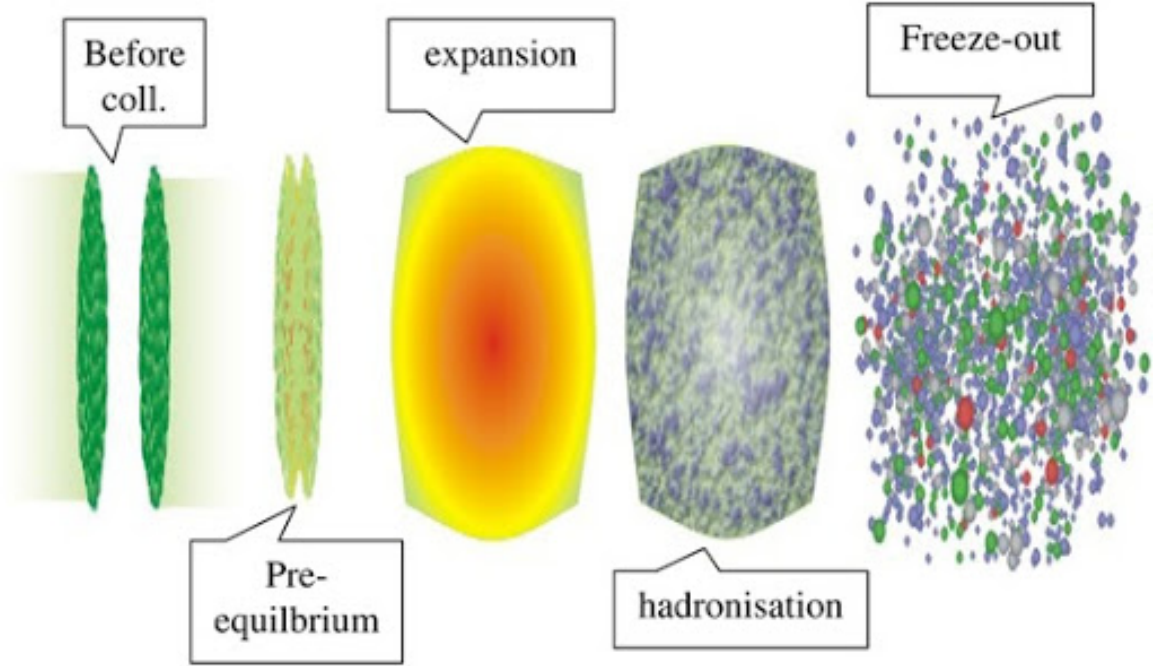


Figure 1.3: Different stages of high energy nuclear collisions[Chaudhuri 14].

different stages that occur in a collision, namely the Pre-equilibrium stage, the expansion stage, the hadronization stage, and the freeze-out[Chaudhuri 14].

- **Pre-equilibrium stage:** In this stage, initial partonic collisions produce a fireball in a non-equilibrium, highly excited state. Collisions frequently happen to reach a 'local' equilibrium' state. The time taken to reach this local equilibrium is known as thermalization time and is of the order of approximately 0.5-1.0 fm/c. This fast equilibration can be explained in reference to the Weibel instability of the electromagnetic plasma. A non-abelian version of the Weibel instability is often considered as the cause for the rapid thermalization.
- **Expansion stage:** In the thermalized state of equilibrium, the thermal pressure acts against the vacuum surroundings. As a result, the system undergoes expansion. Due to the expansion, the energy density of the system decreases with the temperature. The energy-momentum conservation laws govern the expansion and cooling, which involves an equation of state  $p = p(e, n_B)$  describing the system. In a QCD phase transition, at energy density lower than the critical energy density  $\rho_{cr} \sim 1\text{GeV}/\text{fm}^3$ ,

or temperature lower than the critical temperature  $T_{cr} \sim 200 \text{ MeV}$ , the partons (quarks and gluons) combine to form hadrons.

- **Hadronisation stage:** For a small temperature range, entropy density decreases very fast. By the third law of thermodynamics, the total entropy cannot decrease; as a result, the fireball expands rapidly with an approximately constant temperature. For a 1st order transition, the QGP and the hadrons co-exist in the system. In the 2nd transition, the collective flow grows slow, and all the partonic matter gets converted to hadronic matter.
- **Freeze-out stage:** The system remains in thermal equilibrium still. The hadrons continue to collide to maintain local equilibrium while the system expands and cools. When inelastic collisions, associated with the hadrons identity change, become too small to keep up with the expansion. This stage is called chemical freeze-out. The hadrons, after the chemical freeze-out, do not lose their identity in collisions (elastic). However, due to the elastic collisions, the local equilibrium remains, and the system expands with fixed hadrons. When the strong interaction range becomes comparatively smaller than the mean distance between two successive collisions, the collisions become infrequent, and the local thermal equilibrium is lost. The hadrons decouple or freeze-out, reaching the kinetic freeze-out. Hadrons from the freeze-out surface will be detected in the detector.

There are many models to describe high-energy collisions. Hydrodynamics gives a simple description of relativistic heavy-ion collisions, which assumes that a local (thermal) equilibrium is established in a heavy-ion collision. At each point in the space-time picture, a small region exists which maintains equilibrium, and a temperature  $T(x)$  can be defined for this small region at that point in space-time.

## 1.4 Kinematics of HIC

The figure 1.4 shows a sketch view of the collision inside a collider. The two Lorentz contracted disc collide with each other. The collisions can be "head-on" collisions, where the centers of two nuclei are aligned with each other or "peripheral" collisions, where the centers are not aligned with each other horizontally. To discuss the kinematics of a HIC, let

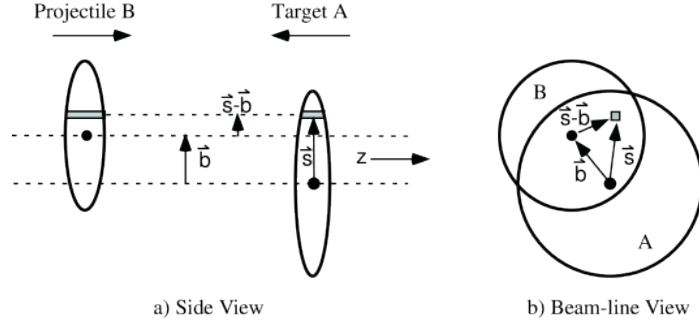


Figure 1.4: A collision event in side view and beam view. Here, two nuclei A and B collide with nuclei  $b$  distance apart.

us first define some standard notation and units. We will be using natural units through-out this thesis where,

$$\hbar = c = k_B = 1$$

$$Metric : g^{\mu\nu} = diag(1, -1, -1, -1)$$

In this work, the Einstein's summation convention is used (wherever needed), repeated indices are summed over, thus

$$\partial_\mu J^\mu = \sum_{i=x,y,z,t} \partial_i J^i$$

We define *proper time*  $\tau$ , as the time an observer observes in its rest frame.

$$d\tau = dt \sqrt{1 - \beta^2}, \quad \beta = \left| \frac{v}{c} \right|$$

In space time phase diagram the collision happens at the origin( $t = 0, z = 0$ ) and the QGP expands into the time-like region where  $\tau^2 = t^2 - z^2 > 0$ . The line  $t = z$  is the light-like region along which only light or massless particles can travel. The space-like region is defined as  $\tau^2 = t^2 - z^2 < 0$  and is not accessible to particle travelling at  $v < c$ . For time-like region, the space rapidity can be defined as,

$$\eta_s = \frac{1}{2} \ln \frac{t+z}{t-z} \quad (1.5)$$

which goes to infinity along the line  $t = \pm z$ . Figure 1.5 shows a space-time evolution picture for a collision event.

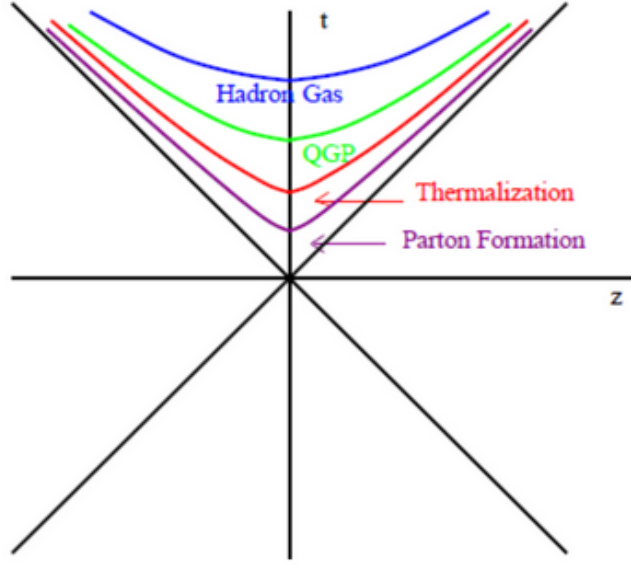


Figure 1.5: A space time representation of evolution of a collision.[Chaudhuri 14]

### 1.4.1 Lorentz transformation

A translational Lorentz transformation is known as Lorentz boost. Consider a Lorentz boost in  $z$ -direction in the lab frame. The particle at  $(t, x, y, z)$  in frame  $L$  is acted on with a Lorentz boost to  $L'$  frame where it is at  $(t', x', y', z')$  with velocity  $v$ . If the collision happens at  $t = t' = 0$ , then the two coordinates are related as,

$$\begin{pmatrix} t' \\ z' \end{pmatrix} = \begin{pmatrix} \gamma & -\beta\gamma \\ -\beta\gamma & \gamma \end{pmatrix} \begin{pmatrix} t \\ z \end{pmatrix} \quad (1.6)$$

and,

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} \quad (1.7)$$

where  $\beta = v/c$  and  $\gamma = 1/\sqrt{1 - \beta^2}$  is the Lorentz boost factor.[Chaudhuri 14][Sahoo 16]

### 1.4.2 Energy-Momentum four-vector

For a particle, its velocity is defined as distance travelled in a unit time,  $\vec{v} = \frac{d\vec{x}}{dt}$  where  $d\vec{x}$  is the distance travelled and  $dt$  is the time measured in the lab frame [Sahoo 16]. Using proper time, proper velocity is defined as,

$$\vec{\eta} = \frac{d\vec{x}}{d\tau} = \gamma \vec{v} \quad (1.8)$$



where  $d\vec{x}$  is the distance travelled in the laboratory frame and  $d\tau$  is the proper time. The momentum is defined as  $mass \times velocity$ . In relativity mechanics, momentum is defined using proper velocity [Sahoo 16].

$$p^\mu = m\eta^\mu \quad (1.9)$$

with,

$$\vec{p} = \gamma m \vec{v} \quad (1.10)$$

$$p^0 = \gamma mc \quad (1.11)$$

The relativistic energy,  $E$ , is given as,

$$E = \gamma mc^2 = p^0 c \quad (1.12)$$

Using the above equation, one can define the energy-momentum 4-vector as,

$$p^\mu = \left( \frac{E}{c}, p_x, p_y, p_z \right) \quad (1.13)$$

### 1.4.3 The Rapidity variable

Two successive Lorentz boost( $\beta$  and  $\beta'$ ) can be defined as a single Lorentz boost as,

$$\beta'' = \frac{\beta + \beta'}{1 + \beta\beta'}. \quad (1.14)$$

The velocity is not a linearly additive quantity. The rapidity  $y$  is a relative measure of this velocity, which is linearly additive for successive transformation [Sahoo 16][Chaudhuri 14].

It is defined in terms on  $\beta$  as,

$$y = \frac{1}{2} \ln \frac{1 + \beta}{1 - \beta} \quad (1.15)$$

and is additive in the sense,  $y'' = y + y'$ . In terms of rapidity  $y$ ,

$$\beta = \tanh(y)$$

$$\gamma = \cosh(y)$$

then the Lorentz transformation can be written as [Chaudhuri 14][Sahoo 16],

$$\begin{pmatrix} t' \\ z' \end{pmatrix} = \begin{pmatrix} \cosh(y) & -\sinh(y) \\ -\sinh(y) & \cosh(y) \end{pmatrix} \begin{pmatrix} t \\ z \end{pmatrix} \quad (1.16)$$

In terms of the rapidity variables, particle 4-momenta is defined as,

$$p^\mu = (E, p_x, p_y, p_z) = (m_T \cosh y, p_x, p_y, m_T \sinh y) \quad (1.17)$$

where,  $(m_T)$  is the transverse mass defined as,

$$m_T = \sqrt{m^2 + p_T^2} = \sqrt{m^2 + p_x^2 + p_y^2}$$

For  $p \gg m$ , i.e., very high energy,

$$y = \frac{1}{2} \ln \frac{E + p_z}{E - p_z} \quad (1.18)$$

$$= \frac{1}{2} \ln \frac{\sqrt{m^2 + p^2} + p \cos \theta}{\sqrt{m^2 + p^2} - p \cos \theta} \quad (1.19)$$

$$y = \frac{1}{2} \ln \frac{p + p \cos \theta}{p - p \cos \theta} \quad (1.20)$$

$$= -\ln \tan \theta/2 \equiv \eta \quad (1.21)$$

$\eta$  is called pseudorapidity. Figure 1.6 shows the variation of pseudorapidity with respect to the polar angle  $\theta$ . In emulsion experiments, where details of the particle, e.g., mass, momentum, etc., are unknown, except for the angle of emission, pseudorapidity analysis gives the significant results. In terms of momentum, pseudorapidity is defined as,

$$\eta = \frac{1}{2} \ln \left[ \frac{|\mathbf{p}| + p_z}{|\mathbf{p}| - p_z} \right] \quad (1.22)$$

#### 1.4.4 Transverse momentum and Invariant yield

The transverse momentum is defined as,

$$p_T = \sqrt{p_x^2 + p_y^2} \quad (1.23)$$

In a longitudinal Lorentz boost along the z-direction, the transverse momentum remains invariant. The Lorentz invariant yield is defined as

$$E \frac{d^3 N}{d^3 p} = E \frac{d^3 N}{d^2 p_T dp_z} = \frac{d^3 N}{d^2 p_T dy} \quad (1.24)$$

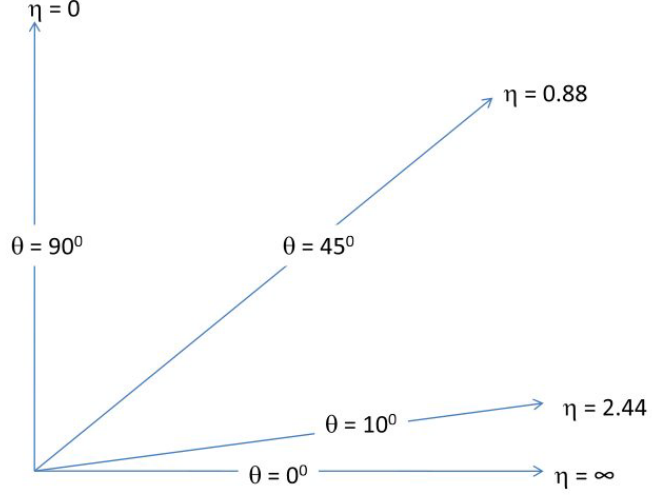


Figure 1.6: Variation of  $\eta$  w.r.t.  $\theta$ . As the polar angle increases, the pseudorapidity decreases.

where,  $dp_z = E dy$  comes from Lorentz invariance and  $p_T$  is the transverse momentum  $p_T = \sqrt{p_x^2 + p_y^2}$ . In terms of pseudorapidity, this can be transformed as,

$$\frac{dN}{d\eta dp_T} = \sqrt{1 - \frac{m^2}{m_t^2 \cosh^2 y}} \frac{dN}{dy dp_T} \quad (1.25)$$

In terms of measurable quantities, the invariant yield can be defined as,

$$E \frac{d^3 N}{dp^3} = \frac{1}{2\pi p_T} \frac{d^2 N}{dp_T dy} \quad (1.26)$$

This equation gives us the experimental invariant yield from the collision[Sahoo 16].



# Chapter 2

## Measuring Observable from a collision

### 2.1 Particle detection and measurements

In an accelerator, particles are boosted to high energies before their collision happens inside the detectors. The detectors then gather information about the particles, such as speed, mass, and charge – using which a particle is identified. The particle accelerators, which are circular, such as LHC, have radiofrequency cavities to accelerate charged particles and are fitted with magnets inside to maintain these particles' circular orbit[det b] as the particle moves inside the tube, its speed increases. The more energetic the particles are, the higher is the possibility of discoveries[det c].

Particles produced in collisions usually travel in straight lines, but in a magnetic field, their paths become curved. Since the detectors have electromagnets to generate the magnetic field, particles interact with this field and experience the magnetic Lorentz force.

$$\vec{F} = q(\vec{v} \times \vec{B}) \quad (2.1)$$

The curvature of its track tells about the particle's momentum. A particle with low momentum curves more than a particle with higher momentum.

A particle detector consists of different parts designed for different purposes. It consists of tracking devices that track the particles' course after the collision, calorimeters to stop or absorb a particle and measure its energy, and various particle-identification subdetectors using various techniques to identify a particle. Tracking devices form the pathway tracks of electrically charged particles as they pass through interacting with the device. However, most particle tracks are not visible to the naked eye but can be seen by reconstructing the

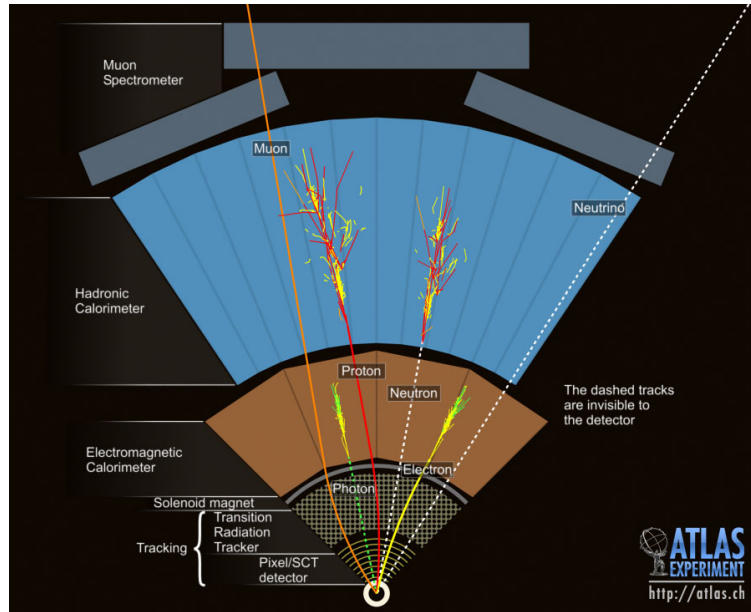


Figure 2.1: A particle detector used in ATLAS experiment at LHC [det a][det c].

recorded electrical signals triggered by particle interaction.

Calorimeters measure the energy of a particle as it passes through and loses it. Its main function is to stop entirely or "absorb" most of the particles coming from a collision, and collecting a deposit of their energy within the detector through layers of "passive" or "absorbing" high-density material, e.g., lead, combined with layers of an "active" material such as solid lead-glass or liquid argon. The electromagnetic calorimeters measure electromagnetic interacting particles' energy as they interact with the electrically charged particles in matter. The hadronic calorimeters measure the energy of hadrons (particles containing quarks, such as protons and neutrons) produced in the collision as they interact with atomic nuclei. Most known particles can be stopped in calorimeters except muons and neutrinos. After passing through the tracking devices and the calorimeters, particles are identified by detecting radiation emitted by the charged particles. A charged particle traveling at a speed greater than the phase velocity of light in the given medium emits Cherenkov radiation at an angle that depends on its velocity[det b]. This angle velocity dependence gives the particle's velocity, and then from momentum, its mass is calculated, which is a signature of its identity. Another method is based on the fact that a fast charged particle, crossing over two different electrical insulators (with different resistivity) to electric currents emits transition radiation. This radiation emitted is related to the energy of the particle and can distinguish between different particles.

The results from all these different parts of the detector are then collated to build up an image of how it might have been when the collision happened. The last step is to look for unusual particles from the collision or to look for results that deviate from what we know so far.

## 2.2 Experimental v/s Estimated Obseables

The quarks and gluons inside the QGP can be thought of as particles in a macroscopic system. Any macroscopic system can be represented in terms of its characteristic state variables, such as its number density ( $n$ ), pressure ( $P$ ), energy density ( $\epsilon$ ), and temperature ( $T$ ). The dynamics of this macro-system can be described in terms of its particle distribution function. Some of these observables we measure directly from experiments, e.g., momentum, particle's energy, charge, etc. These are called as *Experimental Observables* or *Exact Observables*. The other observables, such as temperature, pressure, etc., are obtained from the exact observables by fitting the data in some functional form. These are called as *Estimated Observables*. Some of the distribution functions that are used to give a statistical description are Maxwell-Boltzmann distribution, Tsallis distribution, Pearson distribution, and other functional forms.

### 2.2.1 Boltzmann distribution

A Boltzmann distribution is given by the follllowing equation,

$$f(p) = \frac{1}{(2\pi)^3} \exp\left(\frac{\mu - E}{T}\right) \quad (2.2)$$

This is the most fundamental distributions as other distributions, such as Fermi and Bose, can be written in terms of this distribution. For Boltzmann distribution, the energy density, number density and pressure is defined as [Chaudhuri 14],

$$\epsilon = \frac{1}{(2\pi)^3} \int d^3p E \exp\left(\frac{\mu - E}{T}\right) \quad (2.3)$$

$$n = \frac{1}{(2\pi)^3} \int d^3p \exp\left(\frac{\mu - E}{T}\right) \quad (2.4)$$

$$P = \frac{1}{(2\pi)^3} \int d^3p \frac{1}{3} \frac{|\vec{p}|^2}{E} \exp\left(\frac{\mu - E}{T}\right) \quad (2.5)$$

For dimensionless variables  $z = m/T$  and  $\tau = E/T = \sqrt{|\vec{p}|^2 + m^2}/T$ , the above three equations become,

$$\epsilon = \frac{T^4}{(2\pi)^2} e^{\mu/T} \int_z^\infty d\tau (\tau^2 - z^2)^{\frac{1}{2}} \tau^2 e^{-\tau} \quad (2.6)$$

$$n = 4\pi \frac{T^3}{(2\pi)^3} e^{\mu/T} \int_z^\infty d\tau (\tau^2 - z^2)^{\frac{1}{2}} \tau e^{-\tau} \quad (2.7)$$

$$P = \frac{T^4}{(2\pi)^2} e^{\mu/T} \int_z^\infty d\tau (\tau^2 - z^2)^{\frac{3}{2}} e^{-\tau} \quad (2.8)$$

### 2.2.2 Tsallis distribution

The Tsallis distribution is another functional form used in analysing the exact observables.

The Tsallis distribution for particles is given as [Chaudhuri 14],

$$n_T(E) \equiv \frac{1}{1 + \exp_q\left(\frac{E-\mu}{T}\right)} \quad (2.9)$$

where the function  $\exp_q(x)$  is defined as

$$\exp_q(x) \equiv \begin{cases} [1 + (q-1)x]^{1/(q-1)} & \text{if } x > 0 \\ [1 + (1-q)x]^{1/(1-q)} & \text{if } x \leq 0 \end{cases}$$

and, in the limit where  $q \rightarrow 1$  reduces to the standard exponential:

$$\lim_{q \rightarrow 1} \exp_q(x) \rightarrow \exp(x)$$

A Boltzmann form Tsallis distribution is given as,

$$n_T^B(E) = \left[ 1 + (q-1) \frac{E-\mu}{T} \right]^{-\frac{1}{q-1}} \quad (2.10)$$

This function reduces to Boltzmann distribution in the limit  $q \rightarrow 1$ . The Tsallis distribution is always greater than the Boltzmann distribution for  $q > 1$ . One of the advantages of using the Tsallis distribution is that it helps us deal with the non-extensivity of the system. For Boltzmann or Fermi, or Bose, we need some definite number of particles in the system to be able to get some definitive results. However, Tsallis lets us work in systems with lesser particles by introducing a non-extensivity parameter  $q$ .

### 2.2.3 Pearson distribution

The Pearson distribution, in general, is a solution to the following differential equation [Pearson ][Pollard ],

$$\frac{1}{p(x)} \frac{dp(x)}{dx} + \frac{a+x}{b_0 + b_1x + b_2x^2} = 0 \quad (2.11)$$



where  $a, b_0, b_1, b_2$  are defined using the first four moments ( $m_1, m_2, m_3, m_4$ ) of the variable  $x$  as,

$$\begin{aligned} a = b_1 &= \frac{m_3(m_4 + 3m_2^2)}{10m_2m_4 - 18m_2^3 - 12m_3^2} \\ b_0 &= \frac{m_2(4m_2m_4 - 3m_3^2)}{10m_2m_4 - 18m_2^3 - 12m_3^2} \\ b_2 &= \frac{2m_2m_4 - 6m_2^3 - 3m_3^2}{10m_2m_4 - 18m_2^3 - 12m_3^2} \end{aligned}$$

with  $m_1 = 0$  by construction. The roots to quadratic equation in the denominator leads to different solutions. The Pearson criterion is given as,

$$k = \frac{b_1^2}{4b_0b_2} \quad (2.12)$$

and determines the type of equation. A detailed discussion on Pearson distribution is presented in references [Gupta 20] and [Jena 20]. The solution to equation 2.11 is given as,

$$p(x) = C(e + x)^f(g + x)^h \quad (2.13)$$

## 2.3 Making distributions

As already discussed, to know about estimated observables or hidden observables, exact observables are analyzed as distributions with some functional forms. Most distributions used are particle number versus some observable type distributions. Such a type of distribution can be essentially be described in the form of a histogram, where a number of particles with the same property or same observable value are binned together. For a histogram, an observable is divided into bins covering the complete range of values for that observables. The ROOT library by CERN gives us various methods to make a histogram. These histograms can be 1D or 2D in nature, with one observable or two observables used for binning. In this thesis, 1D histograms are used with one observable required for binning. A histogram is defined using the ROOT TH1F or TH1D library as [ROO ],

```
1 h = TH1D("Name", "Title", n_bins, x_low, x_min,)
2 or
3 h = TH1D("Name", "Title", n_bins, bin_edges)
```

After that, the histogram is filled with particle data extracted from the event(.root) files. This histogram is then event normalized by dividing the bin counts by the total number of

events so as to picture the dynamics of a single event.

**Event Normalization:** When using the data for a different number of events for analysis, the data we get increases with the number of events, i.e., the larger the number of events included, the higher is the particle count. While analyzing, comparing two data sets from a different number of events does not give a meaningful result. At this point, event normalization is needed to bring the data to a common scale per event so that the analysis can be done on a single event basis. After event normalizing the datasets, two different datasets from different numbers of events can be compared as they represent the data scaled down for a single event.

After event normalization, further data analysis is performed.

# Chapter 3

## Optimizing Bin Width

The shape of a histogram depends greatly on the chosen bin width. With a bin width that is too large, one cannot represent the spike rate with good accuracy. And with a bin that is too small in width, the accuracy is increased but so do the fluctuations, and the spike rate is difficult to be observed. There is an appropriate bin size for a set of spike sequences, which is based on the goodness of the fit of the histogram to the hidden spike rate. So far in the  $p_T$  spectra analysis, the bin size has been subjectively selected by the authors. In this thesis, we use one such optimal bin size selection method to use in  $p_T$  spectra analysis.

### 3.1 Bin-width optimization

For a rate  $\hat{\lambda}_t$  estimated for the hidden spike rate  $\lambda_t$  over the total observation length  $T$ , its fit goodness is represented using the Mean Integrated Squared Error (MISE),

$$\text{MISE} \equiv \frac{1}{T} \int_0^T E(\hat{\lambda}_t - \lambda_t)^2 dt \quad (3.1)$$

where  $E$  is the expectation over the period  $T$ . Since the spike rate  $\lambda_t$  is unknown, we make a bar histogram with bin width  $\Delta$  over the whole observation period  $T$  with  $N=\lfloor T/\Delta \rfloor$  intervals. Given a bin size  $\Delta$ , the expected bar height for interval  $t \in [0, \Delta]$  is,

$$\theta = \frac{1}{\Delta} \int_0^\Delta \lambda_t dt \quad (3.2)$$

then the MISE function defined in eq.(3.1) can be written as [Shimazaki 07],

$$\text{MISE} = \frac{1}{\Delta} \int_0^\Delta \frac{1}{N} \sum_{i=1}^N \{E(\hat{\theta}_i - \lambda_{t+(i-1)\Delta})^2\} dt \quad (3.3)$$

where  $\hat{\theta}_i$  is the  $i^{th}$  bar height. The average over the segmented rate  $\lambda_{t+(i-1)\Delta}$  is same as doing the average over an ensemble of (segmented) rate functions  $\{\lambda_t\}$  and is defined, on an interval of  $t \in [0, \Delta]$ , as

$$\text{MISE} = \frac{1}{\Delta} \int_0^\Delta \langle E(\hat{\theta} - \lambda_t)^2 \rangle dt \quad (3.4)$$

### 3.1.1 Cost function

The MISE can be decomposed into two parts: the stochastic error and the error related to the function rate.

$$\text{MISE} = \langle E(\hat{\theta} - \theta)^2 \rangle + \frac{1}{\Delta} \int_0^\Delta \langle (\lambda_t - \theta)^2 \rangle dt \quad (3.5)$$

The first and second terms, respectively, denote the stochastic fluctuations of the fit  $\hat{\theta}$  around the expectation value  $\theta$  and the averaged fluctuations for  $\lambda_t$  around its mean value  $\theta$  over an interval of length  $\Delta$  [Shimazaki 07]. The second term can be decomposed further into two parts:

$$\frac{1}{\Delta} \int_0^\Delta \langle (\lambda_t - \langle \theta \rangle + \langle \theta \rangle - \theta)^2 \rangle dt = \frac{1}{\Delta} \int_0^\Delta \langle (\lambda_t - \langle \theta \rangle)^2 \rangle dt - \langle (\theta - \langle \theta \rangle)^2 \rangle \quad (3.6)$$

where  $\langle \theta \rangle$  is the mean rate and the first term is the mean squared fluctuation of  $\lambda_t$  from  $\langle \theta \rangle$ , and is independent of the bin size  $\Delta$ , since

$$\frac{1}{\Delta} \int_0^\Delta \langle (\lambda_t - \langle \theta \rangle)^2 \rangle dt = \frac{1}{T} \int_0^T (\lambda_t - \langle \theta \rangle)^2 dt. \quad (3.7)$$

#### Proof for Equation 3.6:

$$\begin{aligned} \frac{1}{\Delta} \int_0^\Delta \langle (\lambda_t - \langle \theta \rangle + \langle \theta \rangle - \theta)^2 \rangle dt &= \frac{1}{\Delta} \int_0^\Delta \langle ((\lambda_t - \langle \theta \rangle) - (\theta - \langle \theta \rangle))^2 \rangle dt \\ &= \frac{1}{\Delta} \int_0^\Delta \langle ((\lambda_t - \langle \theta \rangle)^2 + (\theta - \langle \theta \rangle)^2 - 2(\lambda_t - \langle \theta \rangle)(\theta - \langle \theta \rangle)) \rangle dt \\ &= \frac{1}{\Delta} \int_0^\Delta \langle (\lambda_t - \langle \theta \rangle)^2 \rangle dt + \frac{1}{\Delta} \int_0^\Delta \langle (\theta - \langle \theta \rangle)^2 \rangle dt - \frac{1}{\Delta} \int_0^\Delta \langle 2(\lambda_t - \langle \theta \rangle)(\theta - \langle \theta \rangle) \rangle dt \\ &= \frac{1}{\Delta} \int_0^\Delta \langle (\lambda_t - \langle \theta \rangle)^2 \rangle dt + \langle (\theta - \langle \theta \rangle)^2 \rangle - 2 \langle (\theta - \langle \theta \rangle)^2 \rangle \\ &= \frac{1}{\Delta} \int_0^\Delta \langle (\lambda_t - \langle \theta \rangle)^2 \rangle dt - \langle (\theta - \langle \theta \rangle)^2 \rangle \end{aligned}$$

Subtracting equation (3.7) from equation (3.5), the cost function is defined as,

$$\begin{aligned} C(\Delta) &\equiv \text{MISE} - \frac{1}{T} \int_0^T (\lambda_t - \langle \theta \rangle)^2 dt \\ &= \langle E(\theta - \theta)^2 \rangle - \langle (\theta - \langle \theta \rangle)^2 \rangle \end{aligned}$$

Using the decomposition rule for an unbiased estimator ( $E\hat{\theta} = \theta$ ),

$$\langle E(\theta - \langle E\hat{\theta} \rangle)^2 \rangle = \langle E(\theta - \theta)^2 \rangle + \langle (\theta - \langle \theta \rangle)^2 \rangle \quad (3.8)$$

then the cost function is transformed into

$$C(\Delta) = 2 \langle E(\hat{\theta} - \theta)^2 \rangle - \langle E(\hat{\theta} - \langle E\hat{\theta} \rangle)^2 \rangle \quad (3.9)$$

Since the hadronisation process and the collisions are random in nature, assuming the number of particles  $k$  counted for each bin obeys Poisson statistics for a random variable, which means the variance of  $k$  is equal to the mean [Shimazaki 07]. For the estimated rate,  $\hat{\theta} = k/(\Delta)$ , this mean-variance relation is analogous to

$$E(\hat{\theta} - \theta)^2 = \frac{1}{\Delta} E\hat{\theta} \quad (3.10)$$

By using equation 3.10 and equation 3.9 the cost function as a function of the estimator  $\hat{\theta}$  is defined as,

$$C(\Delta) = \frac{2}{\Delta} \langle E\hat{\theta} \rangle - \langle E(\hat{\theta} - \langle E\hat{\theta} \rangle)^2 \rangle \quad (3.11)$$

or in a much simpler form this can be written as,

$$C(\Delta) = \frac{2k - V}{\Delta} \quad (3.12)$$

where  $V$  is the variance and  $k$  is the mean of the sample.

Minimizing the cost function  $C(\Delta)$ , with respect to  $\Delta$ , gives the optimal bin width.

$$\Delta^* \equiv \arg \min_{\Delta} C(\Delta) \quad (3.13)$$

Using this cost, we define an algorithm given in appendix B and use it to get our optimal bin width or bin size.

### 3.1.2 Toy models and bin optimization

We try this method with some toy models for our better understanding and grasp of how this method benefits us. Different type of random data is produced using different distributions and then checked for the fit parameter accuracy using the aforementioned method of bin width optimization. Some distributions used here include Gaussian, Exponential, and Power-law distribution.

A Gaussian sample of 100000 entries with their mean  $\mu = 1.75$  and their standard deviation  $\sigma = 1.11$  was optimized using this method. Figure 3.2 shows the results for the gaussian model. In fig.3.1, spikes are visible for too small a bin width (the one with 500 bins), whereas, for too large bin width, the data doesn't represent the Gaussian model properly. In fig.3.2 the same model is represented on a log-scale in the y-axis. In this subfigure, for too small bin width, the edge values can't be inferred correctly. And for too large bin-width, the sample looks like an asymmetric step ladder. The sample is best presented with a bin-width of 0.06 with 134 bins.

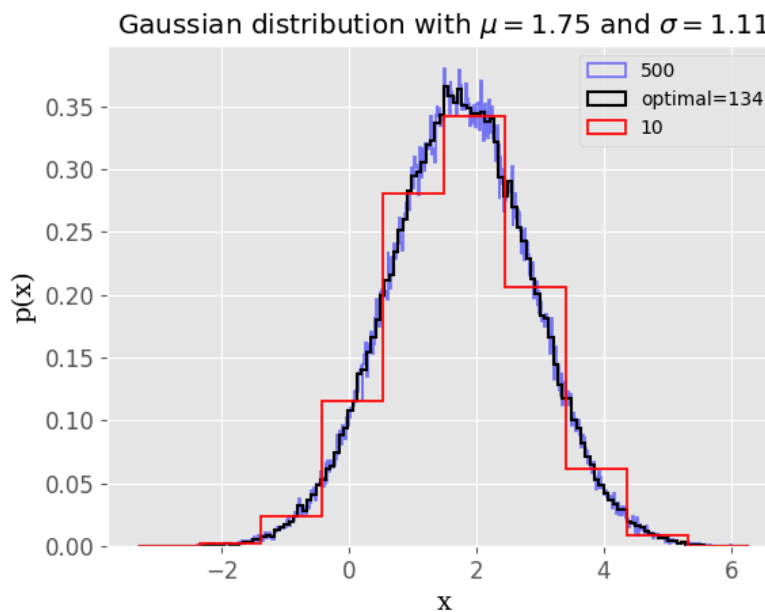


Figure 3.1: Gaussian model

For an exponential model of same size and approximately same range gap we find the optimal bin-width to be around 0.0989. The figure 3.4 shows the exponential samples presented with the optimal bin numbers (91 here). The subfigure 3.3 is the sample represented on a normal y-axis, and the subfigure 3.4 shows the sample on a log scaled y-axis. Again, for

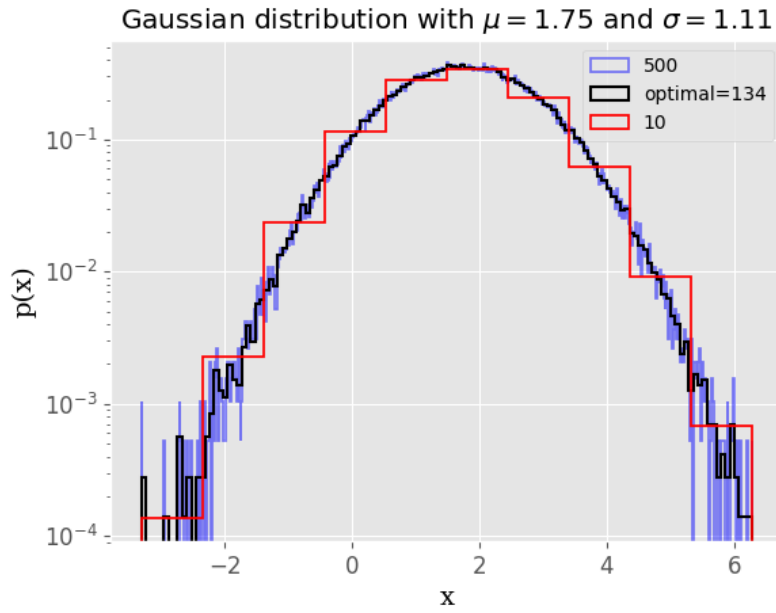


Figure 3.2: Gaussian model log-scaled

too small bin width too many spikes appear in sample representation and for too large bin-width not all characteristics of the sample are represented visually.

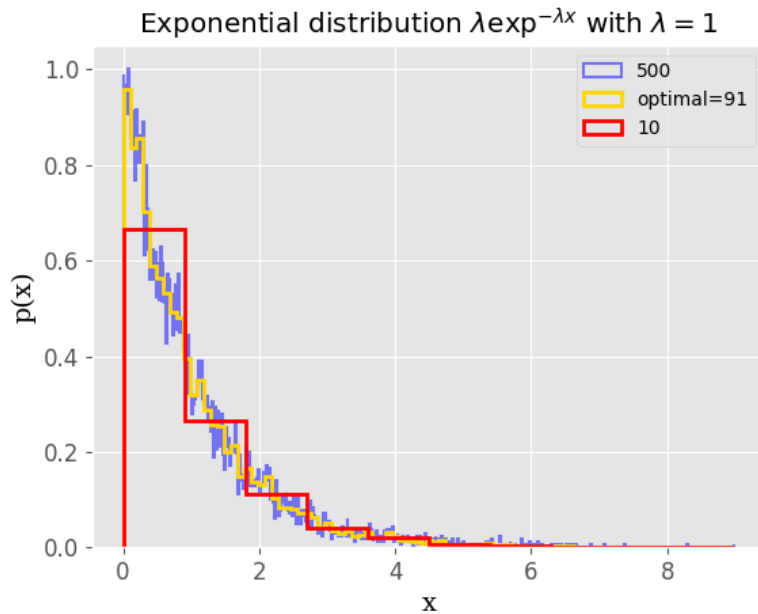


Figure 3.3: Exponential model

The figure 3.5 shows a power-law distribution of the form  $\lambda x^{\lambda-1}$  with 100000 samples. The optimal bin width for this sample was found to be 0.0127 for the x range of 0 - 1.

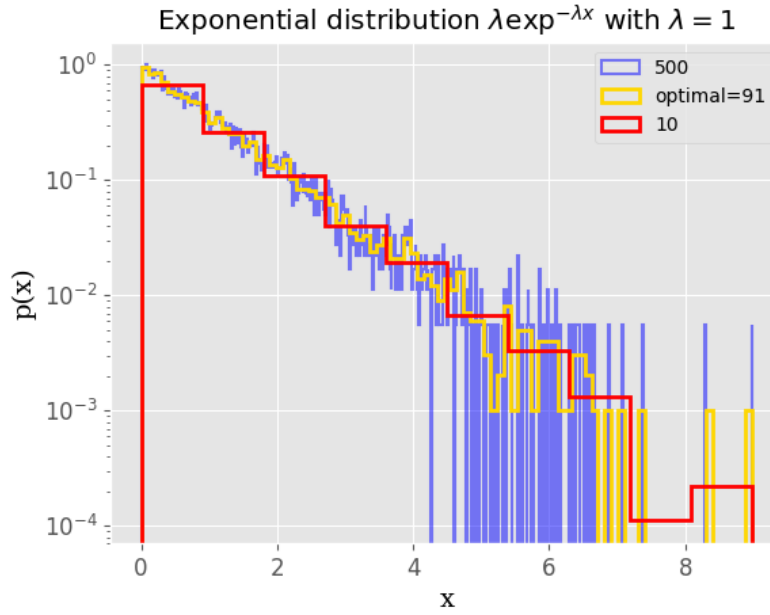


Figure 3.4: Exponential model log-scaled

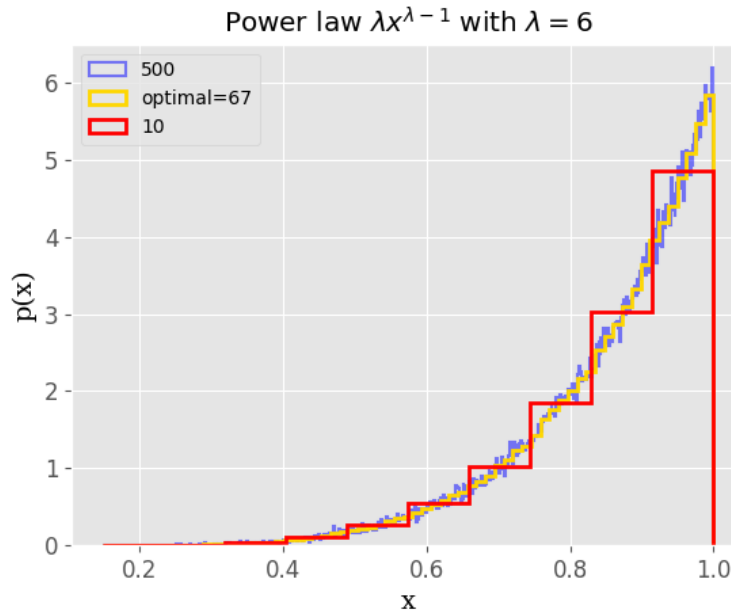


Figure 3.5: A power law distribution (100000 samples) with optimal  $\Delta = 0.06$  represented on a log scale on the y-axis.

## 3.2 Bin-width optimization for Variable bin-widths

Histograms are not always with the fixed bin-widths for all bins. Sometimes, to save time and reduce complexity, data is represented using the variable bin-width histograms. Statistically, a histogram is a graphical representation of tabulated frequencies shown in bins and



can be used as a density plot. The bin-width of a histogram is generally the same for all bins; however, sometimes, it may vary from bin to bin. Generally, a fine bin width gives a good approximation for the underlying distribution for a given dataset, but it also creates a large number of bins that may cost computation resources. While on the other hand, a larger bin width might not approximate the distribution well enough. Though, an advantage of large binning is that it creates lesser bins which can be used with lesser computation cost. This tradeoff between computational cost and accurate description of data distribution is balanced by using a hybrid histogram. It should have a relatively finer bin width setting on dense areas but not partition with too many bins on sparse areas. This motivates us to use a variable bin width, variable bin number histogram in our clustering algorithm [Gao 10]. For a variable bin histogram, the cost function is modified as,

$$C(N, \Delta) = \sum_{i=1}^N \left( \frac{(2k_i - (k_i - \Delta_i \bar{k}))^2}{\Delta_i} \right) \quad (3.14)$$

Using this equation, a variable bin width array can be found while minimizing the cost. In this thesis, we use the percentile binning method for getting the variable bin width array. In the percentile binning method, the bins are made in such a way that each bin contains the same percentage of the data. This makes the binning variable depending on the dataset. Only for uniform random distribution, the variable bin widths can be of approximately the same width.



# Chapter 4

## Optimizing transverse momentum spectra

In this thesis work, we optimized bin-width value for  $p_T$  histograms and checked whether it deviated from the analysis researchers have been doing using their own defined binning. First, let's discuss transverse momentum spectra or  $p_T$  spectra.

### 4.1 Transverse momentum spectra

After the kinetic freeze-out, elastic collisions become very low as the mean free path increases for the particles. There is no change in transverse momentum  $p_T$ , which convey the system kinetic freeze-out properties. The invariant yield plotted as a function of transverse momentum,  $p_T$  is called  $p_T$  spectrum, and the particle distribution represented as a function of  $p_T$  is called  $p_T$  distribution. Mathematically,

$$\frac{dN}{d\mathbf{p}_T} = \frac{dN}{2\pi |\mathbf{p}_T| d|\mathbf{p}_T|}$$

where  $dN$  is the particle numbers in a particular  $p_T$  -bin. Here  $p_T$  is used as a scalar quantity. To understand a  $p_T$  spectra, various mathematical distribution functions are used such as the Tsallis distribution and the Pearson distribution.

#### 4.1.1 Tsallis distribution for $p_T$ -spectra

The most simplest description for  $p_T$  distribution is given by the Boltzmann Gibbs statistics,

$$\frac{1}{2\pi p_T} \frac{d^2 N}{dp_T dy} = \frac{gV m_T}{(2\pi)^3} \exp\left(-\frac{m_T}{T}\right) \quad (4.1)$$

where  $m_T$  is the transverse mass of the particle given as  $\sqrt{p_T^2 + m^2}$ .

Boltzmann-Gibbs statistics is only good approximation to study an extensive system. However, it fails where the system is non-extensive or the entropy is non-additive.

The Tsallis distribution solves the problem of non-extensivity in the system by introducing a non-extensivity parameter  $q$ , and changing the exponential to a  $q$ -exponential which is given as,

$$e_q^x = [1 + (q - 1)x]^{\frac{1}{q-1}} \quad (4.2)$$

Therefore, in terms of Tsallis distribution the  $p_T$  distribution at mid rapidity( $y = 0$ ) is defined as [Cleymans 13],

$$\frac{1}{2\pi p_T} \frac{d^2 N}{dp_T dy} = \frac{gV m_T}{(2\pi)^3} \left[ 1 + (q - 1) \frac{m_T}{T} \right]^{-\frac{q}{q-1}} \quad (4.3)$$

#### 4.1.2 Pearson distribution for $p_T$ -spectra

The Pearson distribution is a further more generalization of the Tsallis distribution. As discussed in Chapter 2, a pearson distribution is a given as,

$$p(x) = B \left( 1 + \frac{x}{e} \right)^f \left( 1 + \frac{x}{g} \right)^h \quad (4.4)$$

where  $B = C e^f g^h$  is a normalization constant. Replacing  $e$ ,  $f$ ,  $g$ , and  $h$  with  $p_0$ ,  $-n$ ,  $\frac{T}{q-1}$ , and  $-\frac{q}{q-1}$  respectively gives us the pearson form for  $p_T$  distribution as,

$$\frac{1}{2\pi p_T} \frac{d^2 N}{dp_T dy} = B \left( 1 + \frac{p_T}{T} \right)^{-n} \left[ 1 + (q - 1) \frac{p_T}{p_0} \right]^{-\frac{q}{q-1}} \quad (4.5)$$

where,

$$B = C \frac{1}{p_0^n} \left( \frac{T}{q-1} \right)^{-\frac{q}{q-1}}$$

The Pearson distribution reduces to Tsallis distribution when  $n = -1$  and  $p_0 = 0$ , keeping all the thermodynamic properties intact [Jena 20] [Gupta 20][]. Pearson distribution is very helpful in explaining the higher- $p_T$  region which is mostly related to jets.

## 4.2 Optimizing bin-width

In this work, we have optimized  $p_T$  distribution binning over the low- $p_T$  range of 0-3 GeV/c. The data used is generated using the UrQMD KASCADE event simulator to simulate event by event Pb-Pb collisions at  $\sqrt{s_{NN}} = 2.76$  TeV. The data used in this work is

for  $\pi^+$  particles which are abundantly produced in collision events, with a pseudorapidity cut of -0.5 to 0.5. The optimized bin width, for the range 0-3 GeV/c, was found to be of 0.1111 GeV/c for fixed bin width histogram with a total of 27 bins. While optimizing the bin-width, changing the number of events doesn't change the optimal bin-width value. The figure 4.1 shows the cost v/s bin-width distribution for  $p_T$  distributions.

We used Tsallis and Pearson distributions for studying the distributions and for the chi-

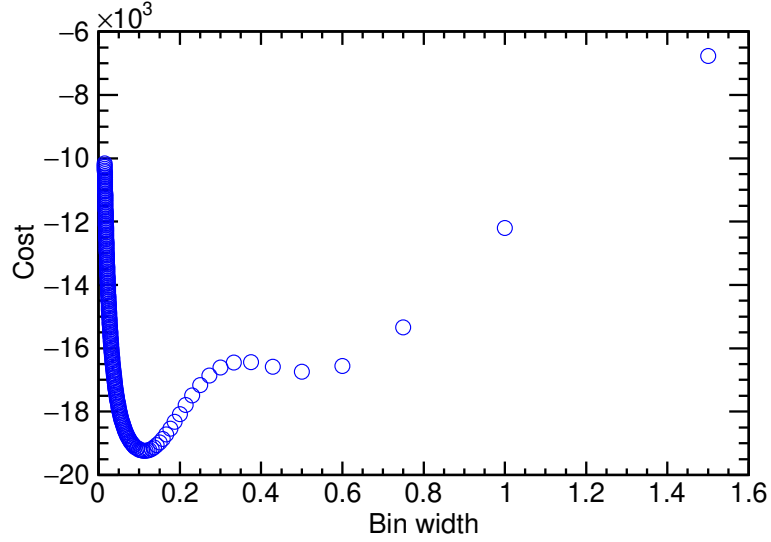


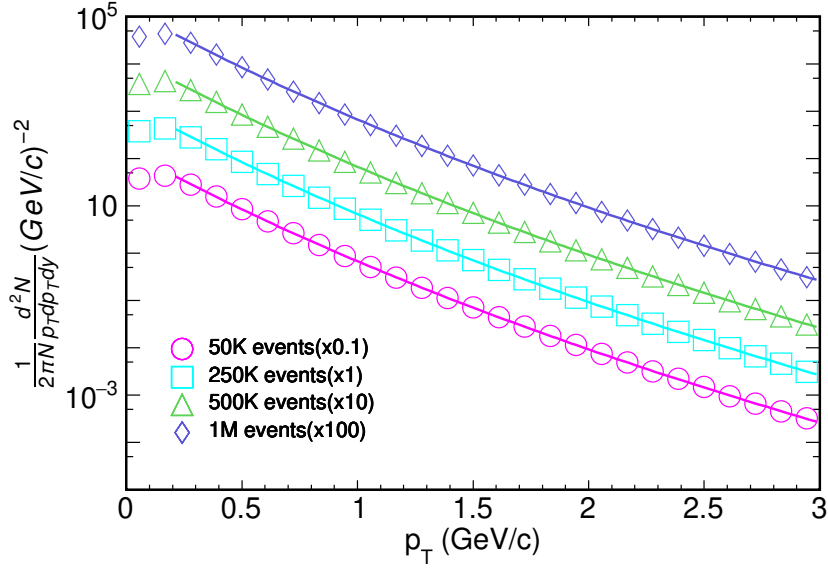
Figure 4.1: Cost vs Bin-width graph plotted for fixed bin-width  $p_T$  spectra.

squared check.

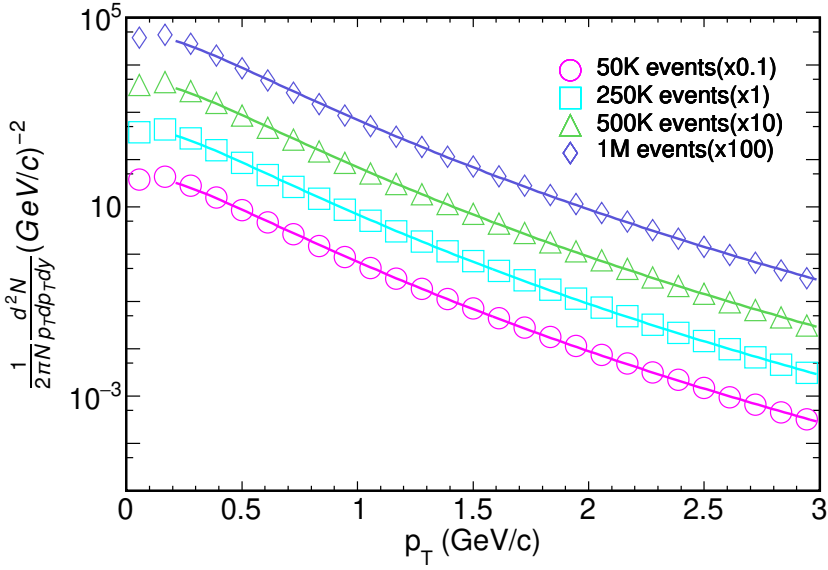
In the figure 4.2, are shown the  $p_T$  for different number of events with data points shifted by a multiplier for a clear understanding. On the y-axis is the event averaged invariant yield plotted against  $p_T$  on the x-axis, and fitted with Pearson (fig. 4.2a) and Tsallis (fig. 4.2b) distributions. The table 4.1 shows the chi-squared values for both the distributions.

No. of Events	Tsallis	Pearson
50,000	0.347685	0.049979
250,000	0.337071	0.026144
500,000	0.315487	0.023130
1,000,000	0.321809	0.022758

Table 4.1: Chi-squared/NDF values for the  $p_T$  distributions.



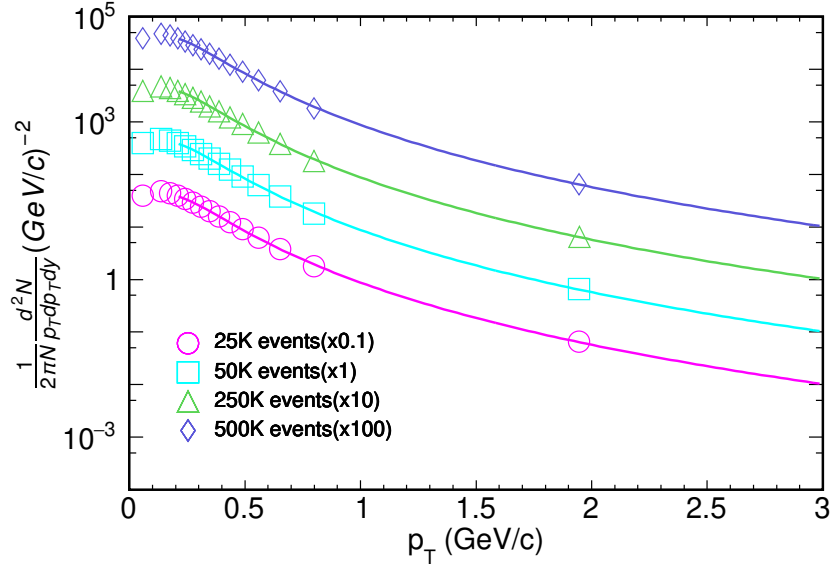
(a) Using Pearson distribution.



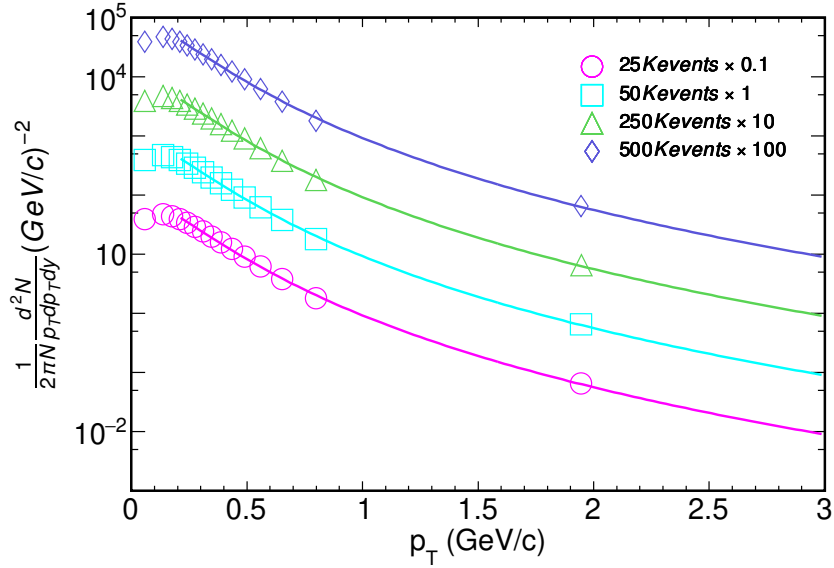
(b) Using Tsallis distribution.

Figure 4.2:  $p_T$  distribution for  $\pi^+$  particles in range  $|\eta| < 0.5$  for  $N$  ( $= 50\text{K}, 250\text{K}, 500\text{K}, 1\text{M}$ ) events in the  $p_T$  range of 0-3 GeV with the optimal bin-width  $\Delta = 0.1111$  GeV/c.

It is clear from the plots and the chi-squared values that optimal binning results in good fit values. Also, one can see that all data points for different number of events line coincide, which means there is no number of event dependence on binning.



(a) Using Pearson distribution.



(b) Using Tsallis distribution.

Figure 4.3:  $p_T$  distribution for  $\pi^+$  particles in range  $|\eta| < 0.5$  for  $N$  ( $= 50\text{K}, 250\text{K}, 500\text{K}, 1\text{M}$ ) events in the  $p_T$  range of 0-3 GeV with the optimal variable bin-width.

For variable binning, the percentile based binning method is used and bin edges were found to be 6.66 percentile apart with a total of 15 bins. The figure 4.3 shows the variable bin-width distribution of the  $p_T$  distribution with bin edges as  $[0.0, 0.117105, 0.158122,$

0.192790, 0.225549, 0.258199, 0.291905, 0.327654, 0.366533, 0.410062, 0.460290, 0.520606, 0.597546, 0.705596, 0.891956, 3.0]. The figure 4.4 shows the cost v/s bin-width distribution for  $p_T$  distributions.

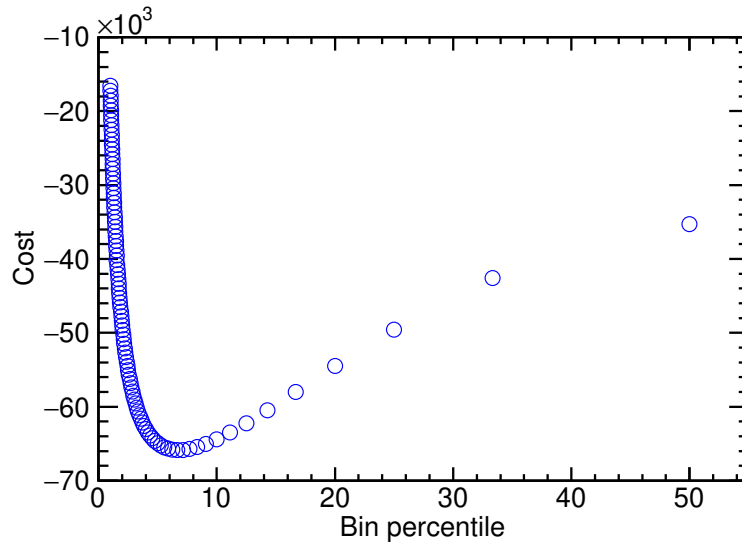


Figure 4.4: Cost v/s Bin-width graph for variable bin-width distribution.

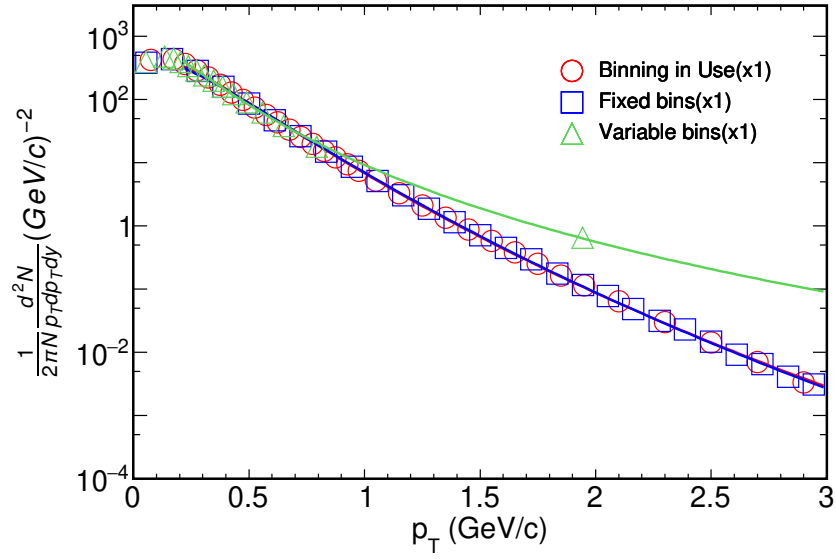
Following from the graphs and table, it is visible that variable binning lowers the accuracy of the fitting for both cases, Tsallis and Pearson. However, chi-squared/NDF values for Pearson distribution are still considerably low, which suggests that a rough estimate for the estimated observable such as  $T$  can still be made using Pearson distribution.//

No. of Events	Tsallis	Pearson
25,000	4.226795	0.644638
50,000	4.255385	0.649264
250,000	4.260996	0.629012
500,000	4.257882	0.635844

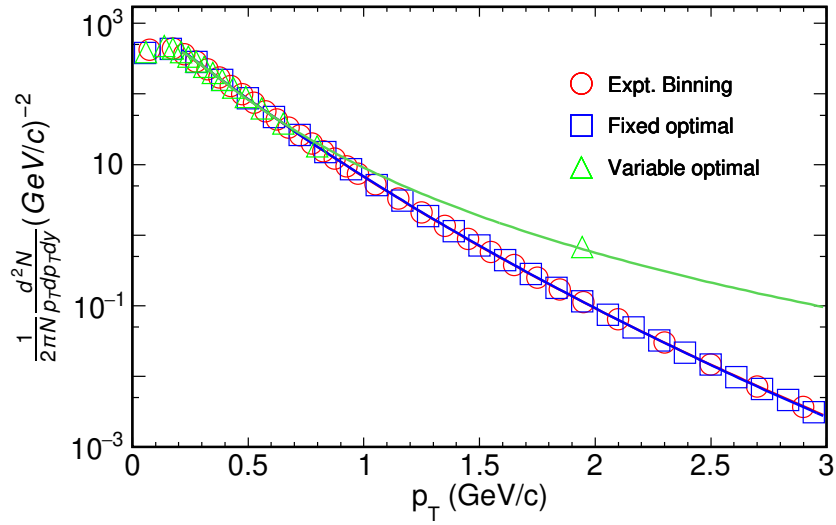
Table 4.2: Chi-squared/NDF values for the  $p_T$  distributions with variable bin width.

Further, a comparison of the binning used in transverse momentum spectrum analysis against the optimal fixed binning and the optimal variable binning for 250K events was made. The figure 4.5 shows the comparison and the table 4.3 shows the chi-squared value for the comparison.





(a) Using Tsallis distribution.



(b) Using Pearson distribution.

Figure 4.5: A binning comparison for  $p_T$  distribution for  $\pi^+$  particles in range  $|\eta| < 0.5$  for 250K events in the  $p_T$  range of 0-3 GeV/c.

This proves that using the optimal binning reduces the random error related to the data and the results obtained are more accurate. The optimal fixed binning presents better accuracy than both the binning, the binning used in experimental analysis by physicists and

	Tsallis	Pearson
Binning in Use	0.422864	0.050224
Optimal fixed binning	0.337084	0.026142
Optimal variable binning	0.593990	0.218882

Table 4.3: Chi-squared/NDF values for the  $p_T$  distributions binning comparison for 250K events.

the optimal variable binning. Using the variable binning reduces the accuracy, while the computation time does not reduce significantly for 250K events. However, it may reduce if number of events increases to millions or so.

# Chapter 5

## Optimizing Pseudorapidity spectra

Apart from optimizing bin-width for  $p_T$ , bin-width optimization for pseudorapidity,  $\eta$  was also done in this thesis work.

First let's discuss pseudorapidity spectra or  $\eta$  spectra.

### 5.1 Pseudorapidity spectra

We know, the invariant yield is given as,

$$E \frac{d^3 N}{dp^3} = \frac{1}{2\pi p_T} \frac{d^2 N}{dp_T dy} \quad (5.1)$$

Taking  $\frac{1}{2\pi p_T}$  to the other side and integrating with respect to  $p_T$  we get,

$$\frac{dN}{dy} = \int 2\pi p_T E \frac{d^3 N}{dp^3} dp_T \quad (5.2)$$

Using the rapidity to pseudorapidity transformation given in equation 1.25, we get

$$\frac{dN}{d\eta} = \int \sqrt{1 - \frac{m^2}{m_T^2 \cosh^2 y}} 2\pi p_T E \frac{d^3 N}{dp^3} dp_T \quad (5.3)$$

This equation defines the pseudorapidity spectra or the pseudorapidity distribution. Similar to transverse momentum, for pseudorapidity we apply same distribution models of tsallis and pearson.

#### 5.1.1 Tsallis distribution for $\eta$ -spectra

In terms of Tsallis, the  $p_T$  distribution is defined as,

$$\frac{1}{2\pi p_T} \frac{d^2 N}{dp_T dy} = \frac{gV m_T \cosh y}{(2\pi)^3} \left[ 1 + (q-1) \frac{m_T \cosh y - \mu}{T} \right]^{-\frac{q}{q-1}} \quad (5.4)$$

Around midrapidity region where  $y = 0$ , and setting  $\mu = 0$  (since very low chemical potentials are found at LHC energies), this equation reduces to equation 4.3 [Cleymans 13]. Integrating over the transverse momentum and using the rapidity and pseudorapidity transformation relation [Cleymans 13], we get

$$\begin{aligned} \frac{d^2 N}{p_T dp_T d\eta} &= \frac{N}{A} \sqrt{1 - \frac{m_0^2}{m_T^2 \cosh^2 y}} \\ &\times \int_{-\infty}^{\infty} \nu(y_f) \frac{m_T \cosh(y - y_f)}{(2\pi)^2} \left[ 1 + (q - 1) \frac{m_T \cosh(y - y_f)}{T} \right]^{-\frac{q}{q-1}} \end{aligned} \quad (5.5)$$

On doing numerical integration, we finally get

$$\begin{aligned} \frac{dN}{d\eta} &= \frac{N}{A} \int_{-\infty}^{\infty} dy_f \int_0^{\infty} dp_T p_T \sqrt{1 - \frac{m_0^2}{m_T^2 \cosh^2 y}} \\ &\times \nu(y_f) \frac{m_T \cosh(y - y_f)}{(2\pi)^2} \left[ 1 + (q - 1) \frac{m_T \cosh(y - y_f)}{T} \right]^{-\frac{q}{q-1}} \end{aligned} \quad (5.6)$$

where,  $y$  is the rapidity,  $y_f$  is the fireball rapidity in lab frame and  $\nu(y_f)$  is the distribution for fireball rapidity, defined as,

$$\nu(y_f) = \frac{1}{\sqrt{2\pi}\sigma} \left[ 1 + (q' - 1) \frac{(y_f - y_0)^2}{2\sigma^2} \right]^{-\frac{1}{q'-1}} + \frac{1}{\sqrt{2\pi}\sigma} \left[ 1 + (q' - 1) \frac{(y_f + y_0)^2}{2\sigma^2} \right]^{-\frac{1}{q'-1}} \quad (5.7)$$

In reference[], a complete derivation for tsallis distribution of pseudorapidity is discussed.

### 5.1.2 Pearson distribution for $\eta$ -spectra

The  $p_T$  distribution in pearson form is written as

$$\frac{1}{2\pi p_T} \frac{d^2 N}{dp_T dy} = B \left( 1 + \frac{m_T \cosh y}{p_0} \right)^{-n} \left[ 1 + (q - 1) \frac{m_T \cosh y}{T} \right]^{-\frac{q}{q-1}} \quad (5.8)$$

where,

$$B = C \frac{1}{p_0^n} \left( \frac{T}{q - 1} \right)^{-\frac{q}{q-1}}$$

For a fireball moving in lab frame with rapidity  $y_f$ , rapidity distribution is given as.

$$\frac{dN}{dy} = A \int_0^{\infty} dp_T p_T \left[ 1 + \frac{m_T \cosh(y - y_f)}{p_0} \right]^{-n} \left[ 1 + (q - 1) \frac{m_T \cosh(y - y_f)}{T} \right]^{-\frac{q}{q-1}} \quad (5.9)$$

Incorporating the rapidity distribution  $\nu(y_f)$  and integrating similar to what we did in tsallis distribution case, we get the pseudorapidity distribution as,

$$\begin{aligned} \frac{dN}{d\eta} = & A \int_{-\infty}^{\infty} dy_f \int_0^{\infty} dp_T p_T \sqrt{1 - \frac{m_0^2}{m_T^2 \cosh^2 y}} \\ & \times \nu(y_f) \left[ 1 + \frac{m_T \cosh(y - y_f)}{p_0} \right]^{-n} \left[ 1 + (q - 1) \frac{m_T \cosh(y - y_f)}{T} \right]^{-\frac{q}{q-1}} \end{aligned} \quad (5.10)$$

where  $y$  is the rapidity which is given as,

$$y = \frac{1}{2} \ln \left[ \frac{\sqrt{p_T^2 \cosh^2 \eta + m_0^2} + p_T \sinh \eta}{\sqrt{p_T^2 \cosh^2 \eta + m_0^2} - p_T \sinh \eta} \right] \quad (5.11)$$

A complete discussion is presented in reference [Gupta 21]

## 5.2 Optimizing bin-width

Again we use the data generated using the UrQMD KASCADE event simulator to simulate event by event Pb-Pb collisions at  $\sqrt{s_{NN}} = 2.76$  TeV. The data used here is for  $\pi^+$  particles which are abundantly produced in collision events, with a  $p_T$  cut from 0.0 to 3.0 GeV/c. The pseudorapidity bin-width is optimized for the range of -6.0 to 6.0, with optimal fixed bin-width of 0.8 resulting in 15 bins. The figure 5.1 shows the variation of cost with respect to bin-width, and it goes to a minimum value around 0.8 bin-width.

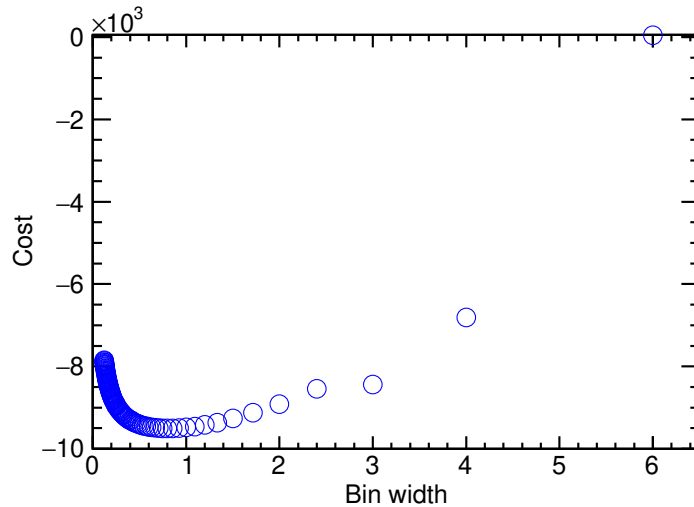


Figure 5.1: Cost v/s bin-width distribution for pseudorapidity( $\eta$ ) observable.

The pseudorapidity ( $\eta$ ) distribution, for fixed optimal bin-width of 0.8, with changing the number of events is shown in the figure 5.2. Here again, changing the number of events doesn't change the optimal value for bin-width for the distribution and the also the shape of the distribution is maintained. For variable binning, it is found that the minimum cost

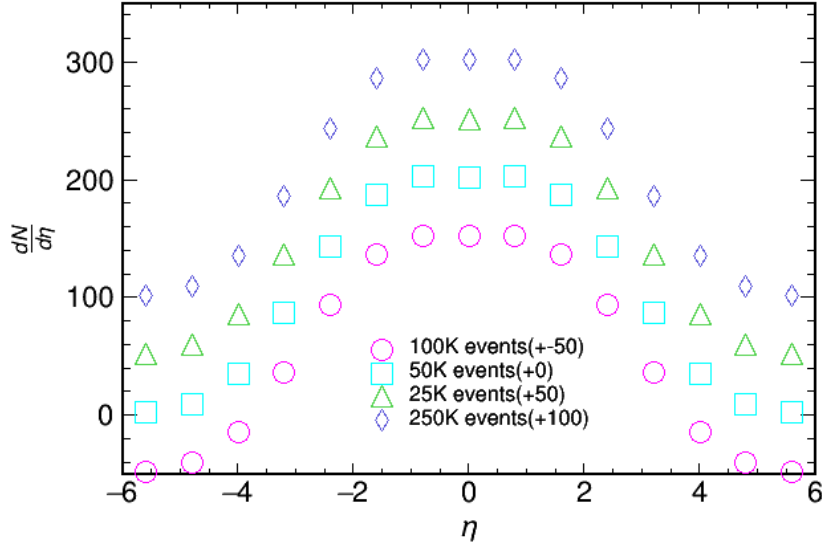


Figure 5.2:  $\eta$  distribution for  $\pi^+$  particles in range  $|\eta| < 6.0$  for  $N$  ( $= 50K, 250K, 500K, 1M$ ) events with  $p_T$  cut from 0-3 GeV with the 0.8 optimal fixed bin-width.

value occurs at three bins which cannot be used to do any meaningful analysis. A possible explanation for this can be stated as the symmetric double-Gaussian nature of the pseudorapidity distribution. Also, not much variation is seen in y-values for the  $\eta$  distribution in range -6.0 to 6.0 as compared to y-values for  $p_T$  distribution for 0-3 GeV/c range.

# Chapter 6

## Conclusion

This thesis work has looked at bin width optimization for data analysis for heavy-ion collision events. We began by talking about the QGP and the heavy-ion collision before moving on to collecting data from experimental setups and analysing it. The bin optimization approach for fixed and variable bin width histograms is then discussed. We begin by analysing the data using this technique for transverse momentum ( $p_T$ ) distributions and then moved on to pseudorapidity ( $\eta$ ) distributions.

For  $p_T$  distributions, a fixed bin width of 0.111 GeV/c was observed which produced better fit results as compared to the bin-width already being used. When using a variable bin-width optimization we see a 6.67 bin-percentile using our optimization method. The results were comparable though with lesser accuracy than both the fixed and the in-use bin-width distribution. However, variable binning gives an advantage of lesser memory resource being used. The time taken while fitting the results increases due to less number of points being used for fitting.

For  $\eta$  distributions, a fixed bin-width of 0.8 was obtained through the optimization method discussed in this work. We got only three bins for variable bin-width histogram, which cannot be used here for calculations and inference purposes. The reason for this could be the symmetric nature of the distribution or the less variation in the y-axis values for the x-axis range. However, a more concrete reason could only be stated after further analysis which is beyond the scope of this work.





# Appendix A

## Python code for bin-width optimization

### A.1 Finding optimal fixed bin width:

```
1 from __future__ import division
2 import numpy as np
3 from ROOT import TH1F
4
5 def optimizeroot(data, events):
6     mxp=3.0 # minimum of pt_data
7     mnp=0.0 # maximum of pt_data
8     N=np.array(range(2,100))
9     Delta = (mxp-mnp)/N #calculating bin widths for different bins, N
10
11     # Initialize cost array
12     Cost = np.zeros(shape=np.size(Delta))
13     for i in range(0,np.size(N)):
14         h=TH1F("h","",N[i],mnp,mxp)
15         for jj in data:
16             h.Fill(jj)
17         ki=np.zeros(shape=N[i])
18         for j in range(N[i]):
19             ki[j]=h.GetBinContent(j+1)
20         ki=ki/(events)
21
22         k=sum(ki)/N[i]
23
24         v=sum((ki-k)**2)/N[i]
```

```

25         # Calculate Cost for Delta[i] width
26         C[i]=(2*k-v)/(Delta[i]**2)
27         h=None
28
29     #finding minimum cost point
30     cmin=min(C)
31     idx=np.where(C==cmin)
32     idx=int(idx[0])
33     OD=Delta[idx]
34
35     return cmin,OD,N[idx],Cost,Delta

```

## A.2 Finding optimal variable binwidth

```

1 from __future__ import division
2 from ROOT import *
3
4 #defining cost function for variable bin width
5 def cost(bins,binedges,k,c=0):
6     for i in range(bins):
7         c = c+(2*k[i]-(k[i]-(binedges[i+1]-binedges[i])*np.mean(k))**2)
8         /(binedges[i+1]-binedges[i])
9     return c
10
11 #defining function to find variable optimal binning
12 def varbinopt(data, events):
13     #initializing bin edges
14     binedges=[]
15     for n in range(2,100):
16         p = 100/n
17         l=[];s=0
18
19         while s<100:
20             s=s+p
21             s=round(s,4)
22             if s>100 or 99<s<100:
23                 s=100

```

```

24         l.append(s)
25     # calculating bin edges based on the percentile
26     edges=list(np.percentile(data,l,axis=None))
27
28     # inserting the start and end points of pt_data
29     edges.insert(0,0.0)
30     edges.pop()
31     edges.insert(len(edges),3.0)
32     binedges.append(np.asarray(edges))
33
34 binedges=np.asarray(binedges)
35 print binedges
36
37 # initializing the cost array
38 cs=[]
39 for binedge in binedges:
40     bins =binedge.size -1
41     h = TH1D("h","",bins,binedge)
42     for val in data:
43         h.Fill(val)
44     h.Scale(1/events)
45     x=[];y=[];w=[]
46     for k in range(bins):
47         y.append(h.GetBinContent(k+1))
48         x.append(h.GetBinCenter(k+1))
49
50     # finding cost using cost function
51     cs.append(cost(bins,binedge,y))
52
53     h=None
54 # finding minimum cost
55 ind = cs.index(min(cs))
56 return ind,cs[ind], binedges[ind], binedges[ind].size-1, cs

```



# Appendix B

## PyROOT code for reading data and making histograms

### B.1 Code for $p_T$ data:

```
1 from __future__ import division
2 from optimization import optimizerooroot
3 import glob
4 import sys
5 import numpy as np
6 from ROOT import *
7
8 if len(sys.argv)!=2:
9     print "Use path"
10 path=sys.argv[1] # reading rootfiles path from terminal
11
12 # reading files data
13 starttime=time.time()
14
15 # getting a list of all the rootfiles
16 files=glob.glob(path+'/*/*.root')
17 pt=[]
18 events=0
19 for f in files:
20
21     if events >= 1000000:
```

```

22         break
23     if events%250000==0 and events>1:
24         print ("done reading",events)
25
26     # reading the rootfile
27     inFile = TFile.Open(f,"READ")
28     tree = inFile.Get("tree")
29     for entry in range(0,tree.GetEntries()):
30         if events >=1000000:
31             break
32         tree.GetEntry(entry)
33         Imp=getattr(tree,"fImp")
34
35         #checking for centrality
36         if Imp[0]<3.50:
37             fpx=np.asarray(tree.fPx)
38             fpy=np.asarray(tree.fPy)
39             fpz=np.asarray(tree.fPz)
40             pid=np.asarray(tree.fPid)
41             chg=np.asarray(tree.fChg)
42             # print fpx.size
43             for j in range(0,fpx.size):
44
45                 #cheeking for particle id and charge
46                 if pid[j]==101 and chg[j]==1:
47
48                     # calculating $p_T$ and storing in a list
49                     ptv=np.sqrt(fpx[j]**2+fpy[j]**2)
50                     theta=TMATH.ATan2(ptv,fpz[j])
51                     eta=-TMATH.Log(TMATH.Tan(0.5*theta))
52                     if (ptv <=3.0).all() and abs(eta)<0.5:
53                         pt.append(ptv)
54                 events+=1
55     print events
56
57     #converting list to a 1D array.
58     data_pt=np.squeeze(pt)
59
60     # find opt bin width using optimize function

```

```

61 pt_cmin , opt_pt_wid , opt_pt_bin , C, D = optimizeroot( data_pt , events )
62 print pt_cmin , opt_pt_wid , opt_pt_bin
63
64 # Root histogram plot
65 sf = (1 / events)
66
67 # initializing the histogram using TH1D function
68 h2 = TH1D("h2", "fit", opt_pt_bin, 0.0, 3.0)
69 for jj in range(0, len(data_pt)):
70     h2.Fill(data_pt[jj])
71
72 for i in range(0, opt_pt_bin):
73     fq = h2.GetBinContent(i+1)
74     w = h2.GetBinWidth(i+1)
75     x = h2.GetBinCenter(i+1)
76     newf = fq / (2 * np.pi * x * w)
77     h2.SetBinContent(i+1, newf)
78
79 # event normalizing
80 h2.Scale(sf)
81
82 x = np.asarray([h2.GetBinCenter(i+1) for i in range(0, opt_pt_bin)])
83 y = np.asarray([h2.GetBinContent(i+1) for i in range(0, opt_pt_bin)])
84
85 f = TFile.Open("Filename.root", "recreate")
86 can = TCanvas("c2")
87 can.cd()
88
89 # making the distribution graph
90 g = TGraph(x.size, x.astype(np.double), y.astype(np.double))
91 g.SetLineColor(kRed)
92 g.SetLineWidth(2)
93 g.SetMarkerColor(kBlue)
94 g.SetMarkerStyle(20)
95 g.SetTitle('pT distribution')
96 g.GetXaxis().SetTitle(r'pT')
97 g.GetYaxis().SetTitle(r'(1/Nevent)*(dN/dpT)*(1/2*pi*pT)')
98 can.SetGrid()
99

```

```

100 #drawing the distribution
101 can.Draw()
102 g.Draw()
103
104 #writing to the rootfile
105 g.Write()
106 f.Close()

```

## B.2 Code for $\eta$ distribution:

```

1 from __future__ import division
2 import glob
3 import sys
4 import numpy as np
5 from ROOT import *
6
7 if len(sys.argv)!=2:
8     print("Use path")
9 path=sys.argv[1]
10
11 # reading files list
12 files=glob.glob(path+'/*/*.root')
13
14 cent=[0.0, 3.50, 4.94, 6.98, 8.55, 9.88, 11.04]
15 etas=[]
16 events=0
17
18 for f in files:
19     if events>=1000000:
20         break
21     # reading data from file
22     inFile = TFile.Open(f,"READ")
23     tree = inFile.Get("tree")
24     for entry in range(0,tree.GetEntries()):
25
26         tree.GetEntry(entry)
27         Imp=getattr(tree,"fImp")
28         fpx=np.asarray(tree.fPx)

```



```

29     fpy=np.asarray( tree.fPy )
30     fpz=np.asarray( tree.fPz )
31     pid=np.asarray( tree.fPid )
32     chg=np.asarray( tree.fChg )
33
34     # checking centrality
35     if cent[0]<=Imp[0]<cent[1]:
36
37         # checking event count
38         if events >=1000000:
39             continue
40
41         else:
42             for j in range(0, fpx.size):
43                 if pid[j]==101 and chg[j]==1:
44                     ptv=np.sqrt( fpx[j]**2+fpy[j]**2)
45                     theta=TMATH.ATan2( ptv , fpz[j] )
46
47                     # calculate eta and append to the list
48                     eta=-TMATH.Log( TMATH.Tan(0.5*theta) )
49                     if ( ptv <=3.0 ). all() :
50                         if eta > -6.0 and eta < 6.0:
51                             etas.append( eta )
52
53                 events+=1
54
55 # converting list to 1D array
56
57 # find opt bin width using optimize function
58 cmin,opt_wid,opt_bin,C,D=optimizeroot( detas , events )
59 print cmin,opt_wid,opt_bin
60
61 # Root histogram plot
62 sf=(1/events)
63
64 hist = TH1D("h1"," ",opt_bin,-6.0,6.0)
65 for jj in range(0,len(detas)):
66     hist.Fill( detas[jj] )
67 hist.Scale(sf)

```

```

68
69 x=np.asarray([ hist.GetBinCenter(i+1) for i in range(0,opt_pt_bin)])
70 y=np.asarray([h.GetBinContent(i+1) for i in range(0,opt_pt_bin)])
71
72 f=TFile.Open("Filename.root","recreate")
73 can=TCanvas("c2")
74 can.cd()
75
76 # making the distribution graph
77 g=TGraph(x.size,x.astype(np.double),y.astype(np.double))
78 g.SetLineColor( kRed)
79 g.SetLineWidth( 2 )
80 g.SetMarkerColor( kBlue )
81 g.SetMarkerStyle( 20 )
82 g.SetTitle('pT distribution ')
83 g.GetXaxis().SetTitle(r'\eta ')
84 g.GetYaxis().SetTitle(r'(1/Nevent)*(dN/d \eta) ')
85 can.SetGrid()
86
87 #drawing the distribution
88 can.Draw()
89 g.Draw()
90
91 #writing to the rootfile
92 g.Write()
93 f.Close()

```

# Bibliography

- [aus tirol ] Franzl aus tirol. *Running Coupling constant*. [https://commons.wikimedia.org/wiki/File:Strong\\_coupling\\_as\\_function\\_of\\_energy.svg](https://commons.wikimedia.org/wiki/File:Strong_coupling_as_function_of_energy.svg).
- [Chaudhuri 14] A. K. Chaudhuri. *A short course on Relativistic Heavy Ion Collisions*. IOPP, 2014.
- [Cleymans 13] Jean Cleymans. *The Tsallis Distribution for  $p - p$  collisions at the LHC*. Journal of Physics Conference Series, vol. 455, pages 2049–, 08 2013.
- [det a] *ATLAS experiment*. <https://flickr.com/photos/atlasexperiment/>.
- [det b] *How detectors work*. <https://home.cern/science/experiments/how-detector-works>.
- [det c] *Particle Collision and Detection*. [https://stanford.edu/group/stanford\\_atlas/4Particle%20Collision%20and%20Detection](https://stanford.edu/group/stanford_atlas/4Particle%20Collision%20and%20Detection).
- [Gao 10] Song Gao, Chengcui Zhang & Wei-Bang Chen. *A Variable Bin Width Histogram Based Image Clustering Algorithm*. pages 166 – 171, 10 2010.
- [Gupta 20] R. Gupta, A. Menon & S. Jena. *A generalized thermodynamical approach to transverse momentum spectra in high energy collision*, 2020.
- [Gupta 21] Rohit Gupta, Aman Singh Katariya & Satyajit Jena. *A unified formalism to study the pseudorapidity spectra in heavy-ion collision*, 2021.

- [Jena 20] S. Jena & R. Gupta. *A unified formalism to study transverse momentum spectra in heavy-ion collision*. Physics Letters B, vol. 807, page 135551, 2020.
- [Marques 15] L. Marques, J. Cleymans & A. Deppman. *Description of high-energy pp collisions using Tsallis thermodynamics: Transverse momentum and rapidity distributions*. Phys. Rev. D, vol. 91, page 054025, Mar 2015.
- [MissMJ ] Cush MissMJ. *Standard Model of Particle Physics*. [https://en.wikipedia.org/wiki/File:Standard\\_Model\\_of\\_Elementary\\_Particles.svg](https://en.wikipedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg).
- [Oerter 06] R. Oerter. *The theory of almost everything: The standard model, the unsung triumph of modern physics*. 2006.
- [Olive 14] K. A. Olive *et al.* *Review of Particle Physics*. Chin. Phys., vol. C38, page 090001, 2014.
- [PDG ] P.A. Zyla *et al.* (Particle Data Group), *Prog. Theor. Exp. Phys.* 2020, 083C01 (2020). <http://pdg.lbl.gov/>.
- [Pearson ] K. Pearson. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 186, 343 (1895).
- [Pollard ] J. H. Pollard. *Numerical and Statistical Techniques* (Cambridge University Press, 1979).
- [ROO ] *ROOT Reference Documentation*. <https://root.cern.ch/root/html606/>.
- [Sahoo 16] Raghunath Sahoo. *Relativistic Kinematics*, 2016.
- [Shimazaki 07] Hideaki Shimazaki & Shigeru Shinomoto. *A Method for Selecting the Bin Size of a Time Histogram*. Neural computation, vol. 19, pages 1503–27, 07 2007.
- [Tao 20] J. Q. Tao, M. Wang, H. Zheng, W. C. Zhang, L. L. Zhu & A. Bonasera. *Pseudorapidity distributions of charged particles in pp( $\bar{p}$ ), p(d)A and AA collisions using Tsallis thermodynamics*, 2020.