B-tagging with Machine Learning at a 10 TeV Muon Collider

Jeffrey Backus
Princeton University

Overview

- **Goal:** Train a neural network to discriminate between b-jets, c-jets and light (u, d, s, gluon) jets at a 10 TeV muon collider

- Started by generating the following samples using MadGraph and Pythia:

  - $\mu\mu \to bb$, $\mu\mu \to cc$, $\mu\mu \to ll$ (for l light), 100k events

  - $H \to bb$, with H produced via Higgstrahlung, 100k events

- Have truth-level data, still need to simulate detector dynamics but **unsure how to do this**

- **Trained only with** $\mu\mu \to bb$**,** $\mu\mu \to cc$**,** $\mu\mu \to ll$ **for now on truth data**

Extracting features

From the samples, I extracted relevant training features for b-tagging. Some details on how I did this from the Pythia output:

- **Jetting**: FastJet anti-kT with radius R = 0.4, pT >= 20 GeV

- Per-constituent pT floor: pT = 0.5 GeV

- Jets are **not** groomed.

- To label the jet: **ghost labeling** with **ancestry labeling**

- **Soft lepton features**: has pT > 3 GeV, only accepts leptons with b or c ancestry

Training Features

1. **Invariant mass of the jet** (jet_mass) in GeV
2. **Number of constituents** (nconst)
3. **Number of charged constituents** (n_charged)
4. **Fraction of jet pT carried by charged constituents** (frac_ch) in GeV
5. **Fraction of jet pT carried by neutral constituents** (frac_neu) in GeV
6. **Fraction of jet pT carried by leading particle** (leading_frac) in GeV
7. **Largest angular distance between constituents** (max_dR) in rads

Training Features

8. **Radial moment of constituents in jet** (girth) in GeV:

$$\frac{\sum_i p_{T,i} \Delta R_i}{p_T^{\mathrm{jet}}}$$

ΔR_i = angular distance of constituent from jet axis.

9. **Jet pT dispersion** (pTD) in GeV:

$$\frac{\sqrt{\sum_i p_{T,i}^2}}{\sum_i p_{T,i}}$$

Training Features

Soft lepton features

10. **Number of soft leptons in the jet** (n_softlep)

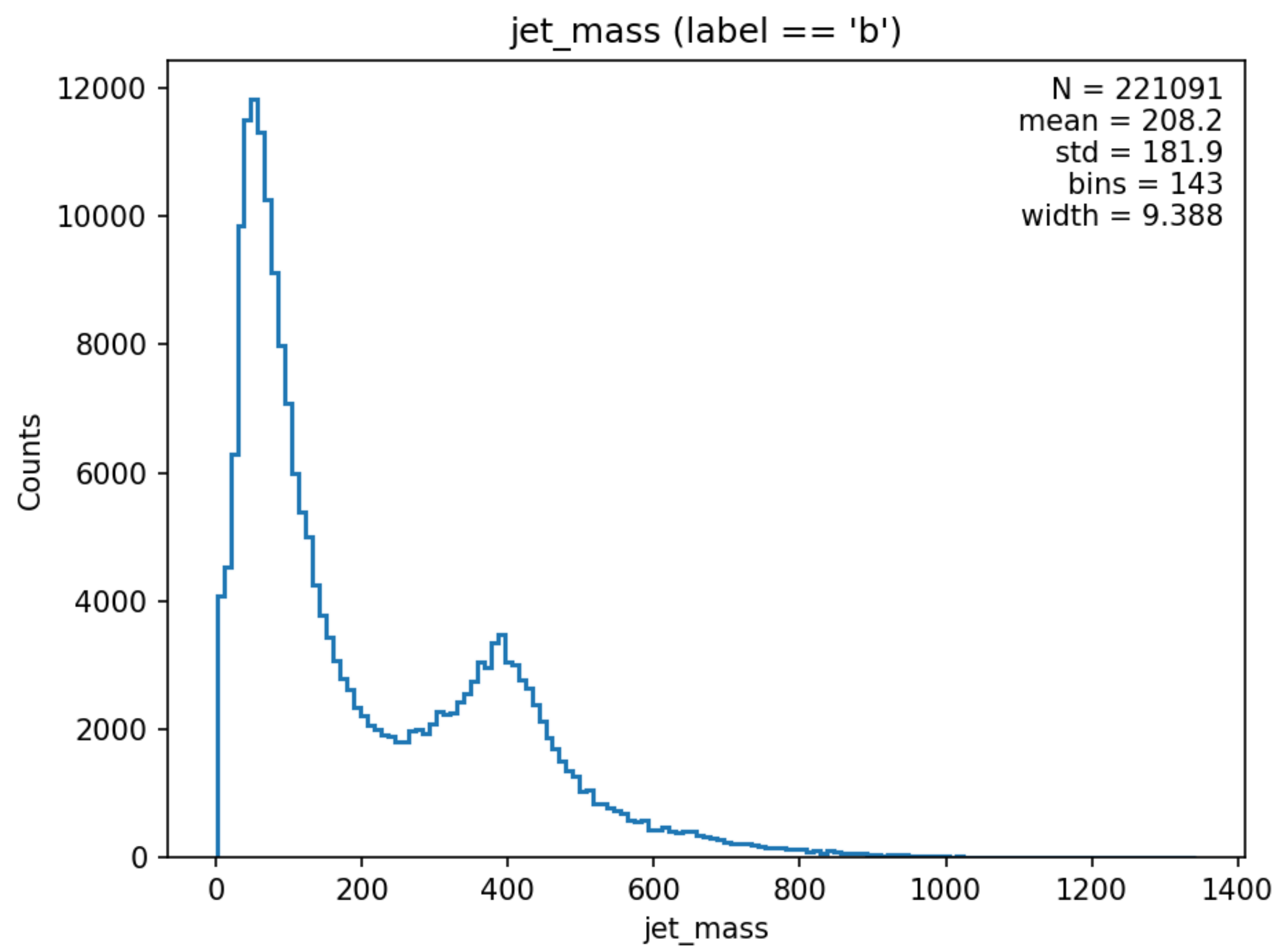
11. **pT of highest-pT lepton relative to jet axis** (softlep_ptrel) in GeV

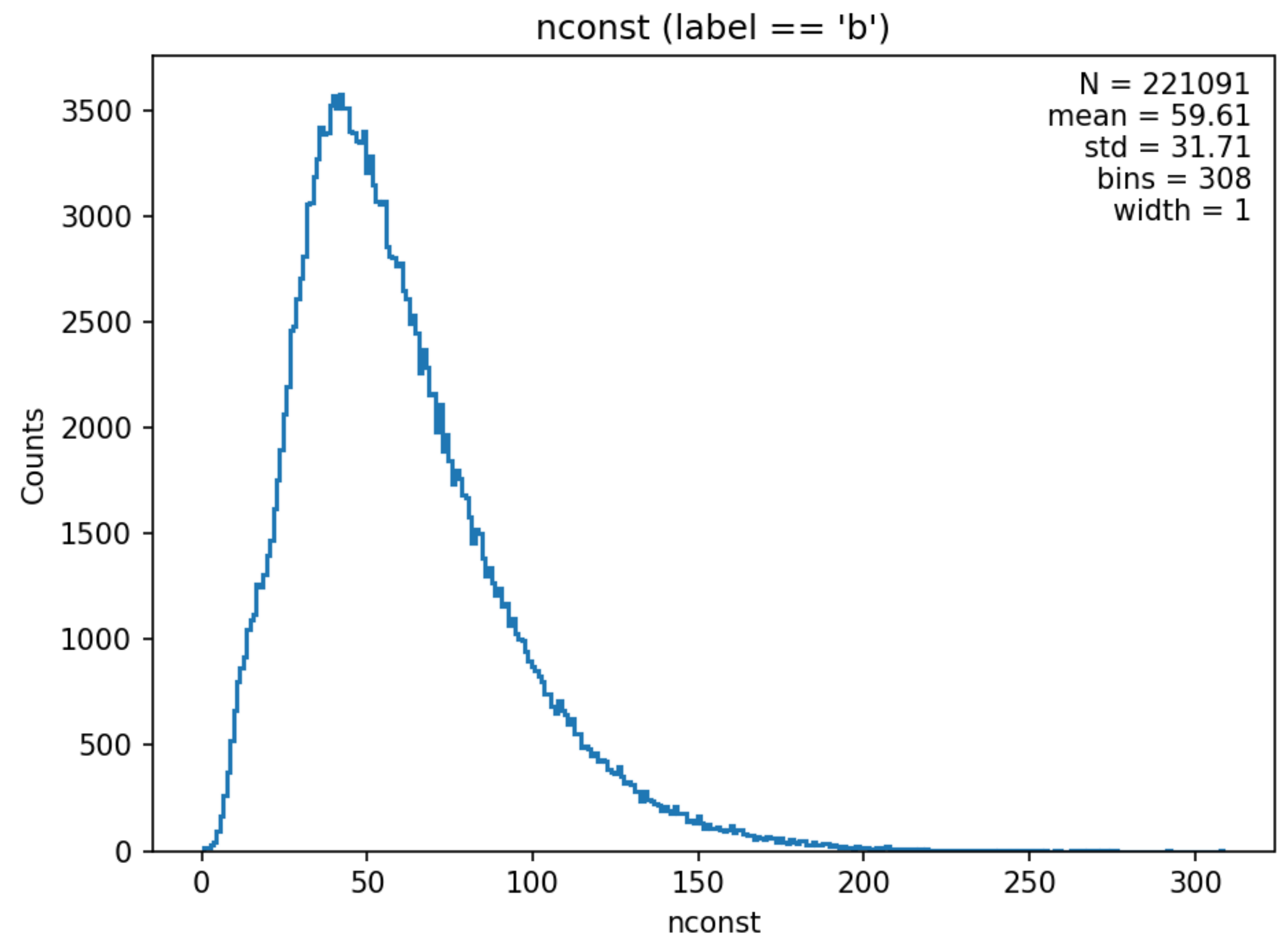
12. **Lepton pT fraction of the highest-pT lepton** (softlep_z)

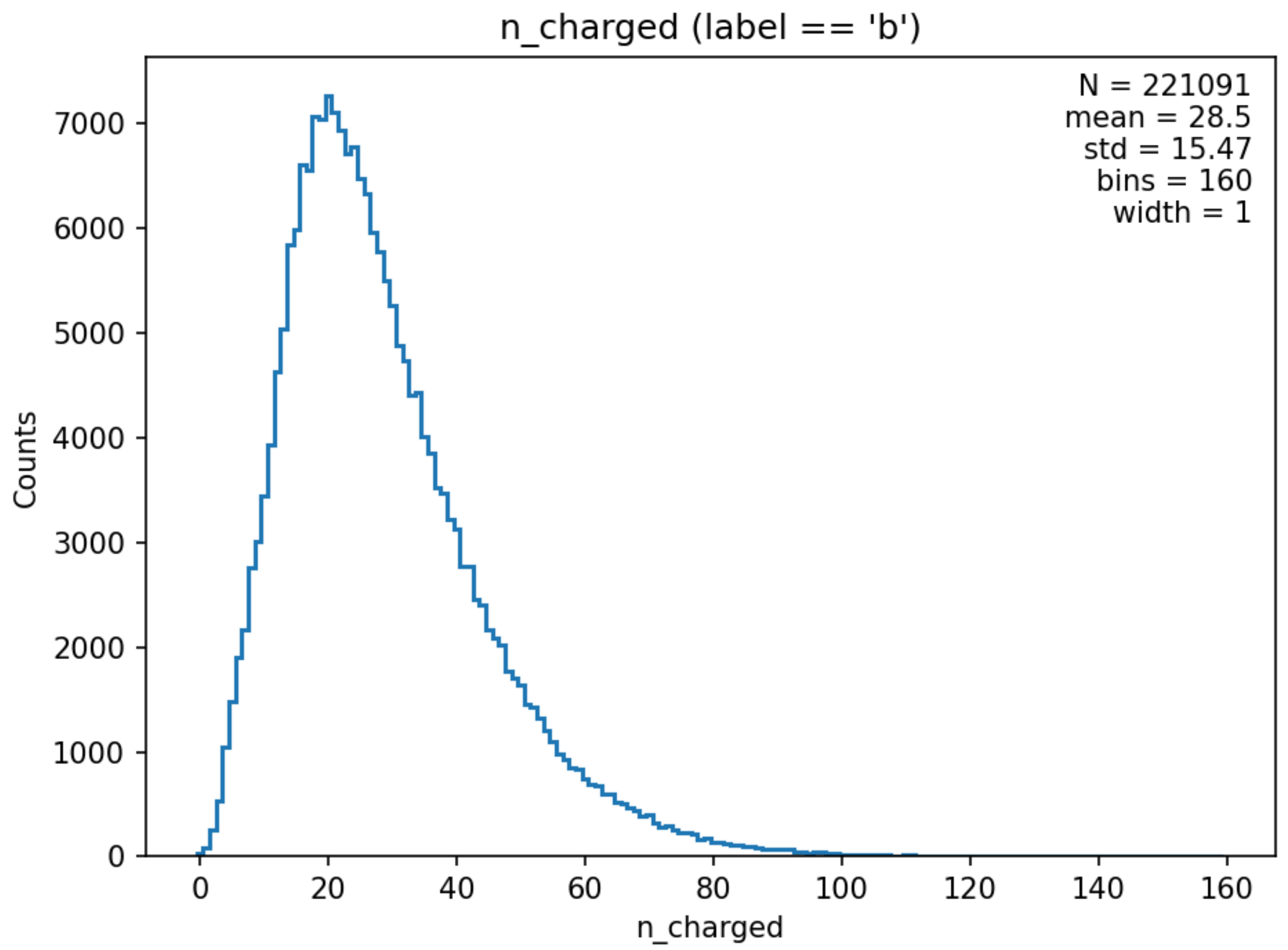
13. **Distance of highest-pT lepton to jet axis** (softlep_dR) in rads

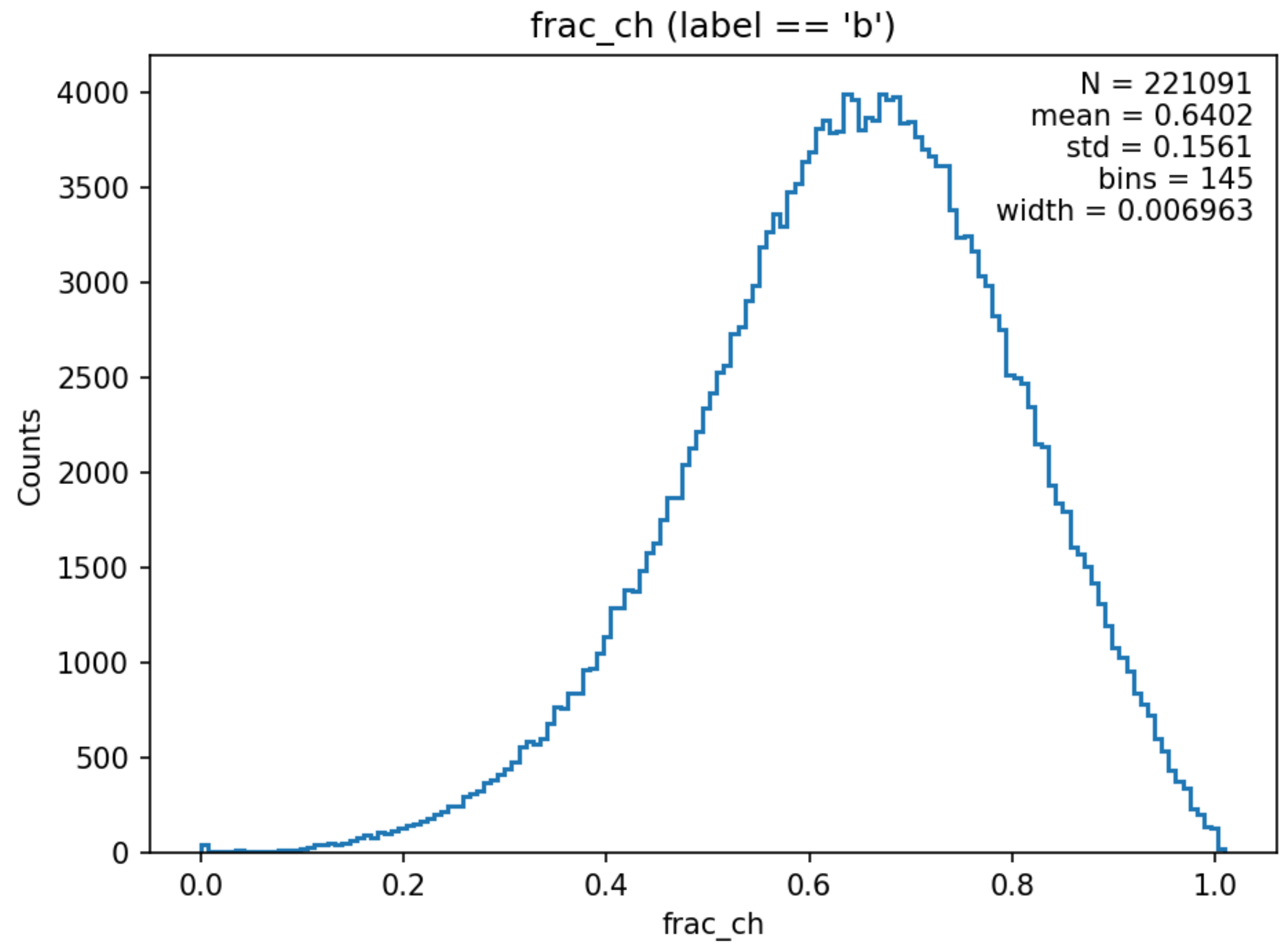
$\mu\mu \to bb$ Feature Plots

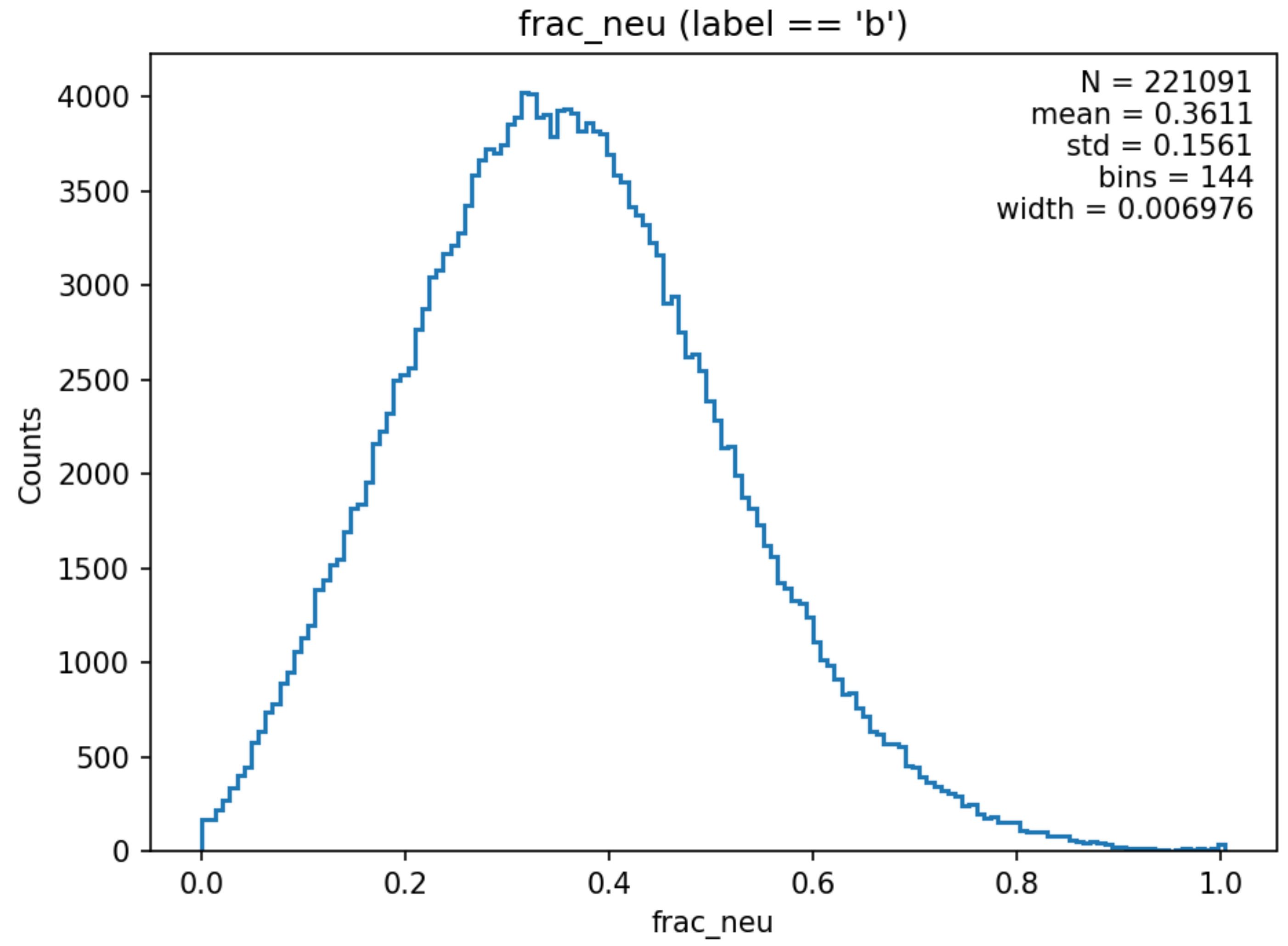
For **b-jets**

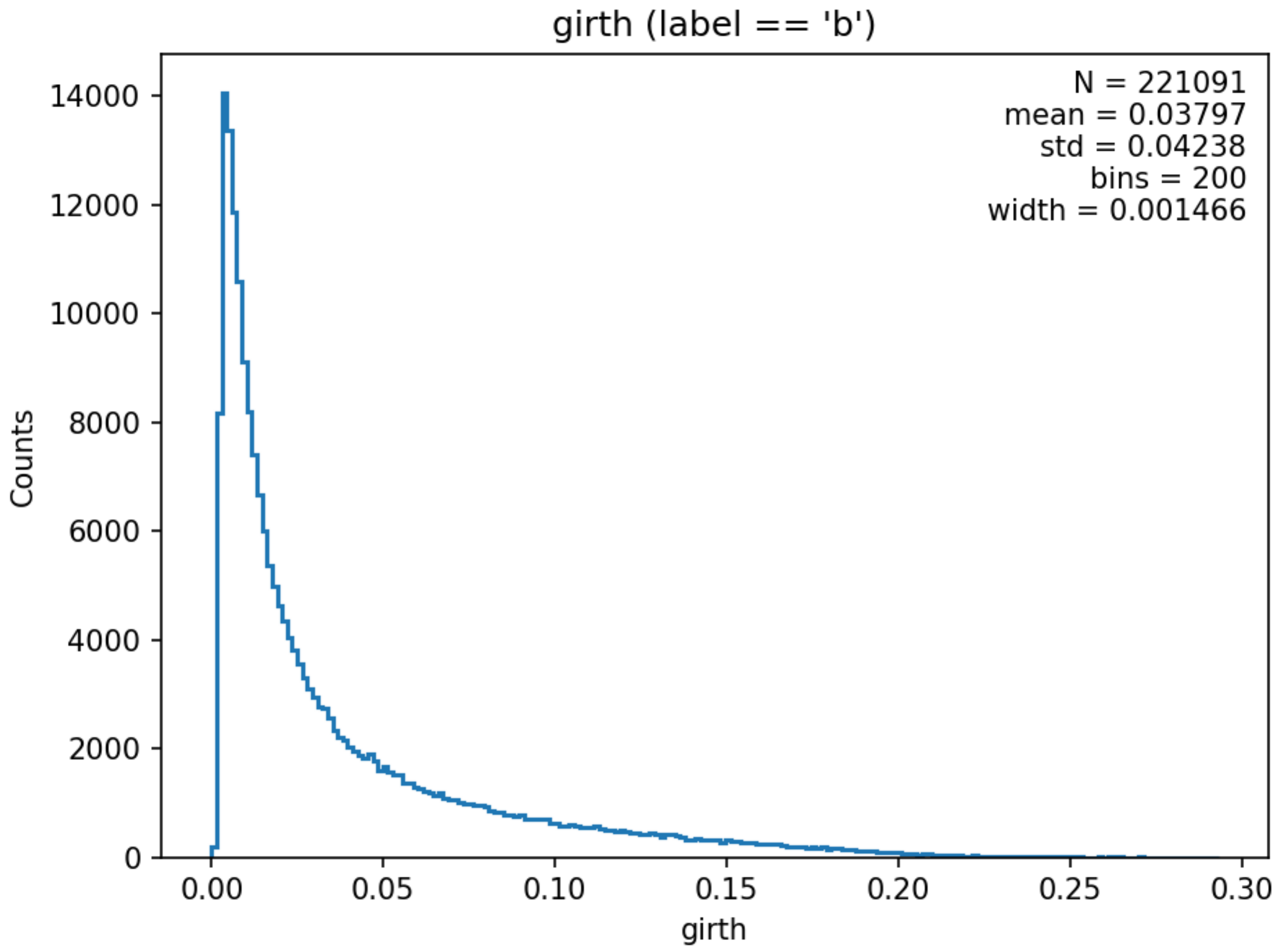






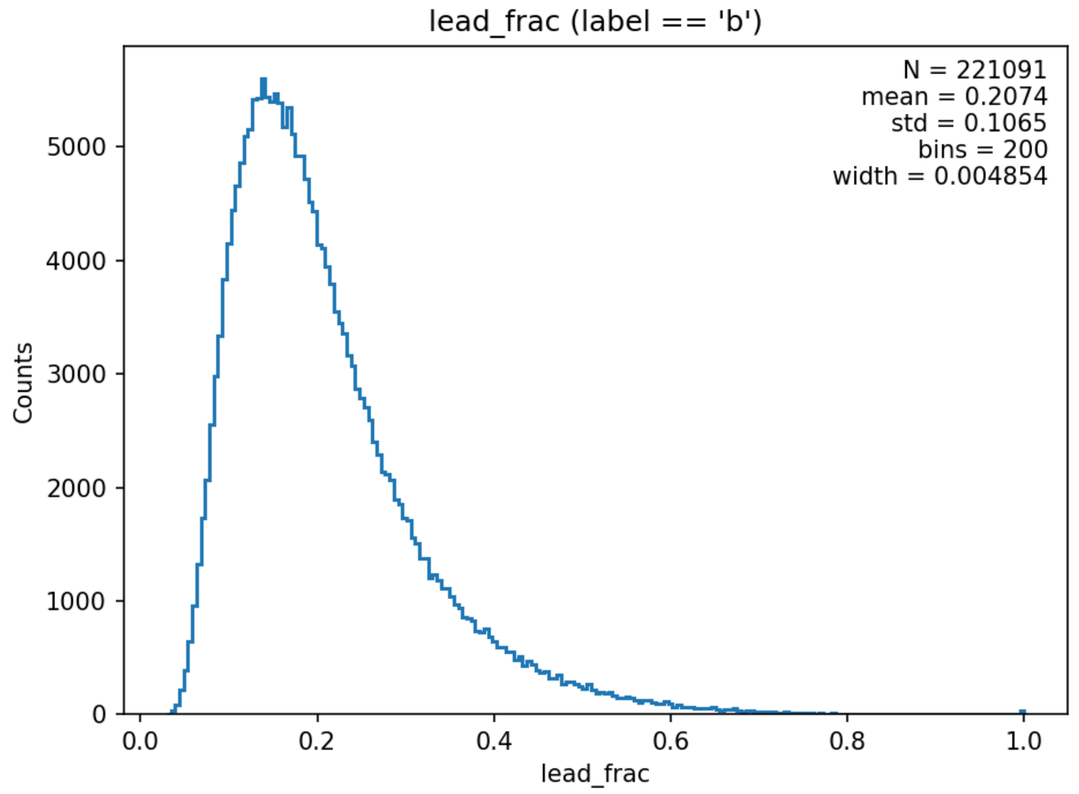


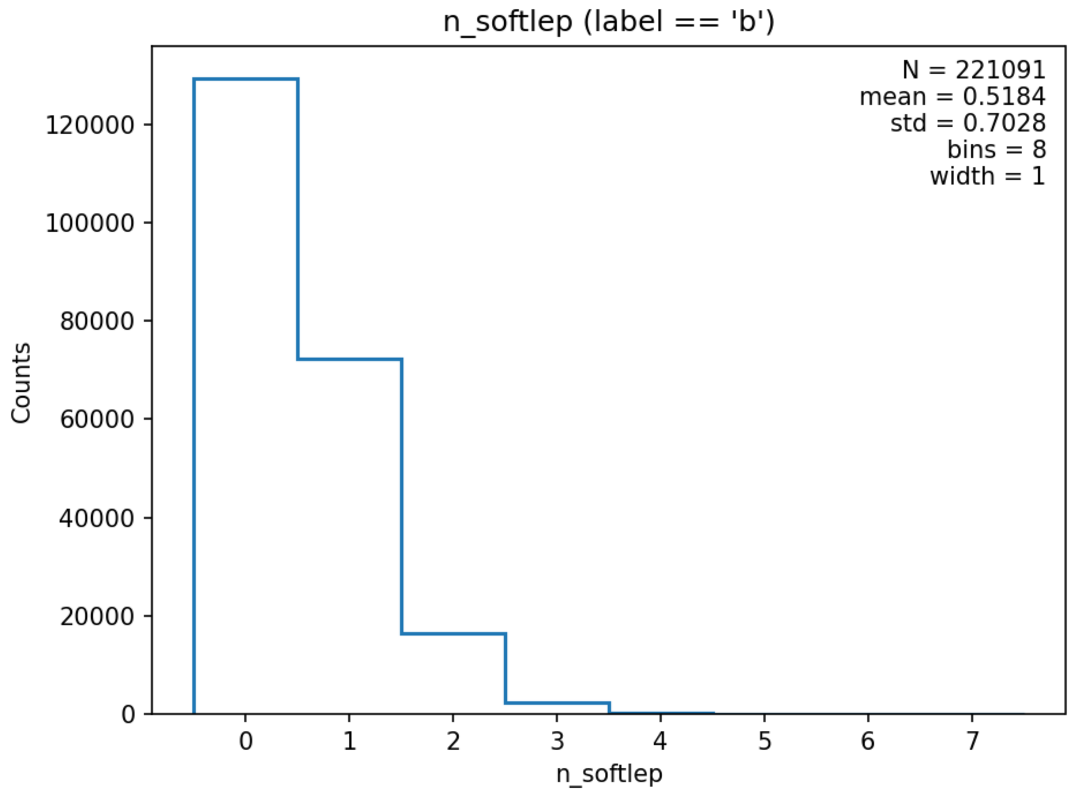


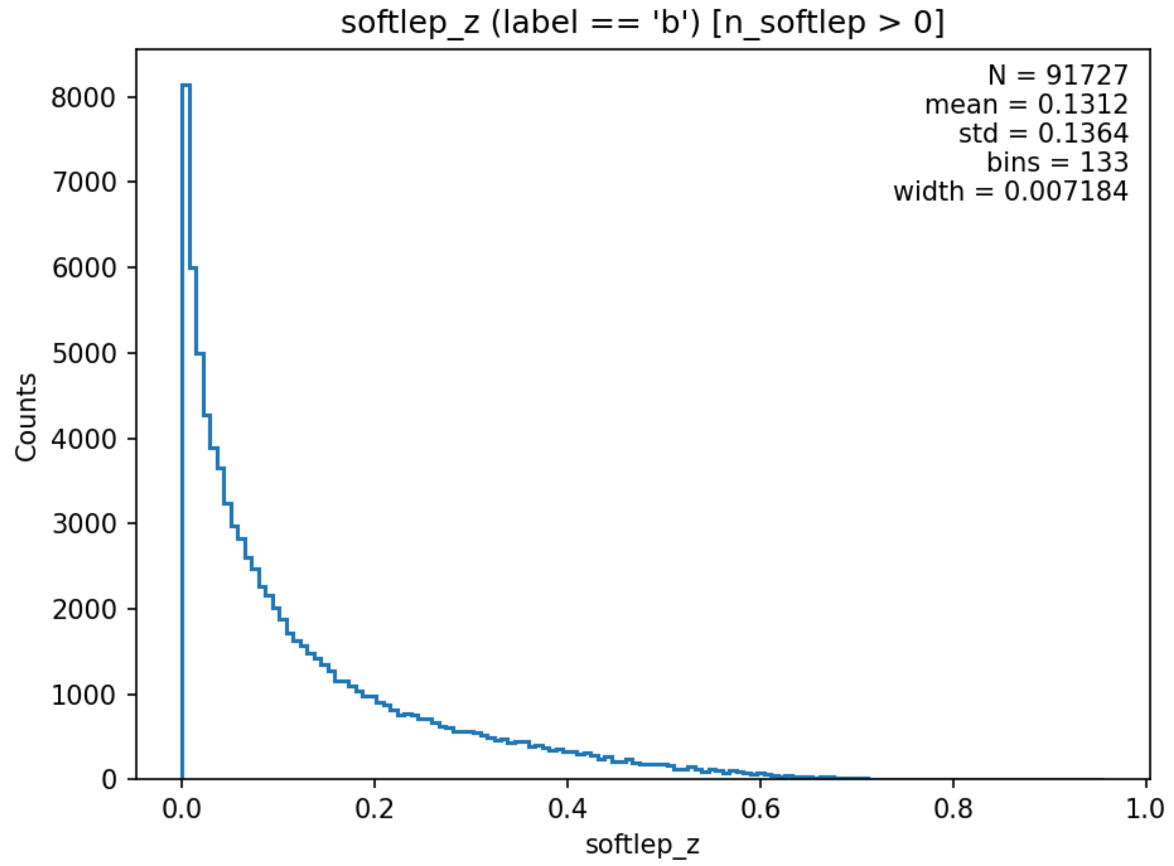


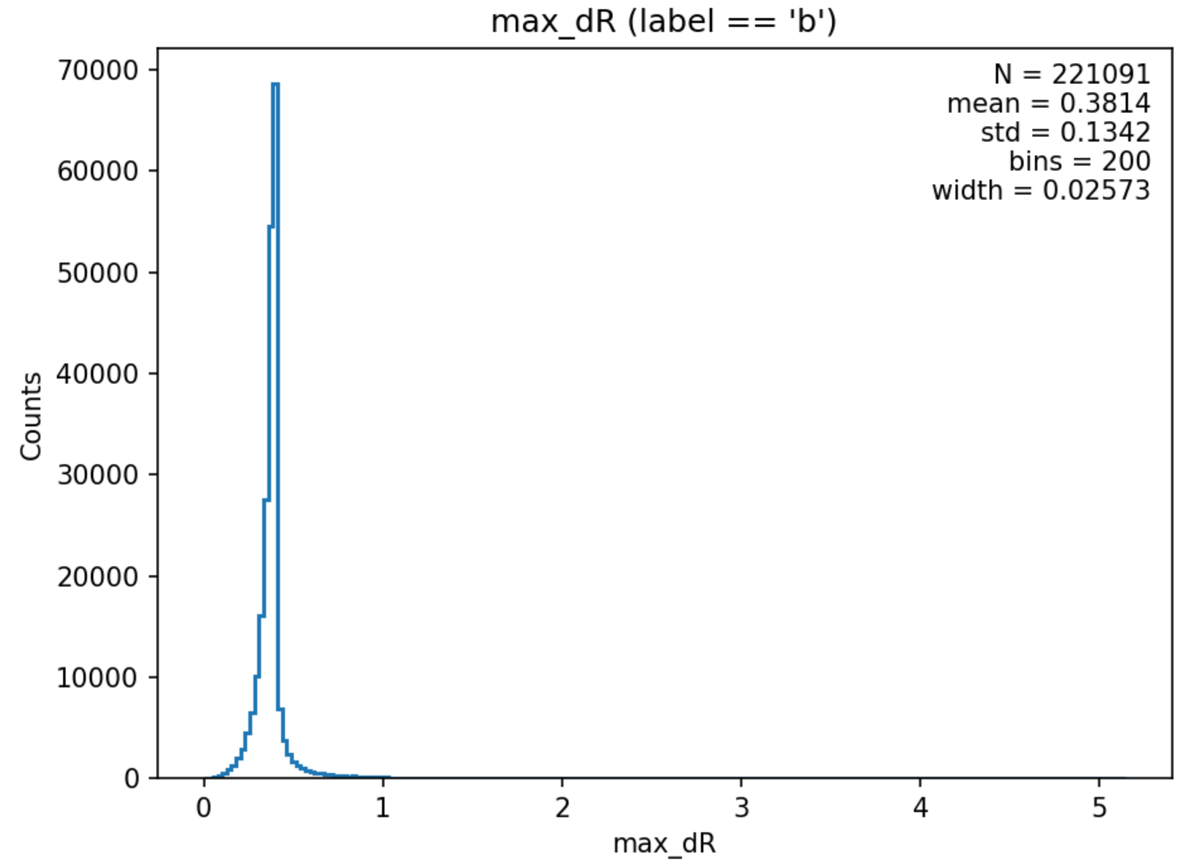
$\mu\mu \to bb$ Feature Plots

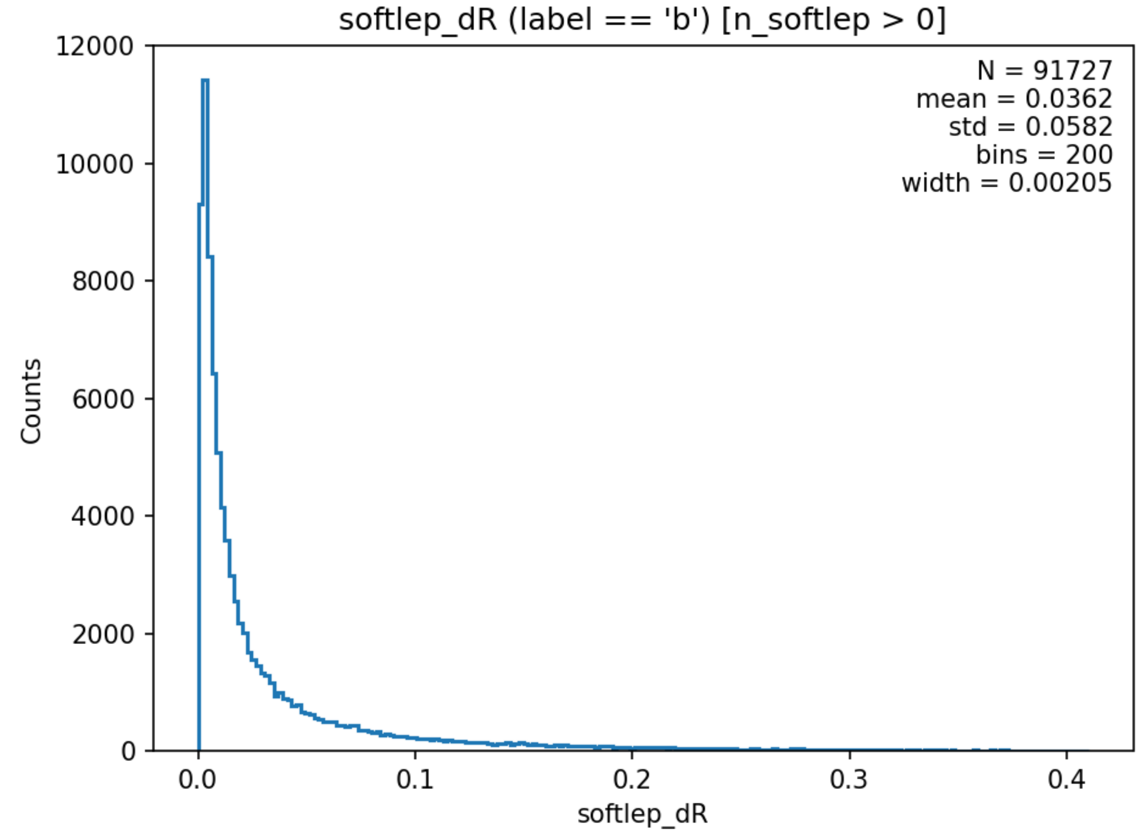
For **b-jets**

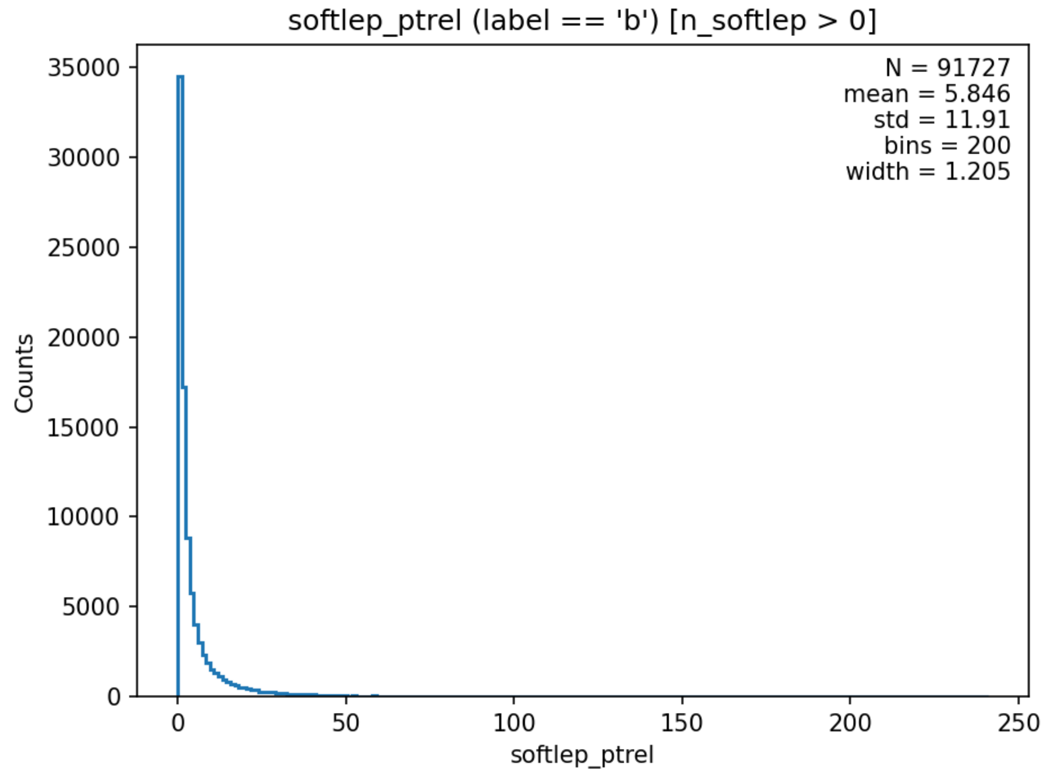


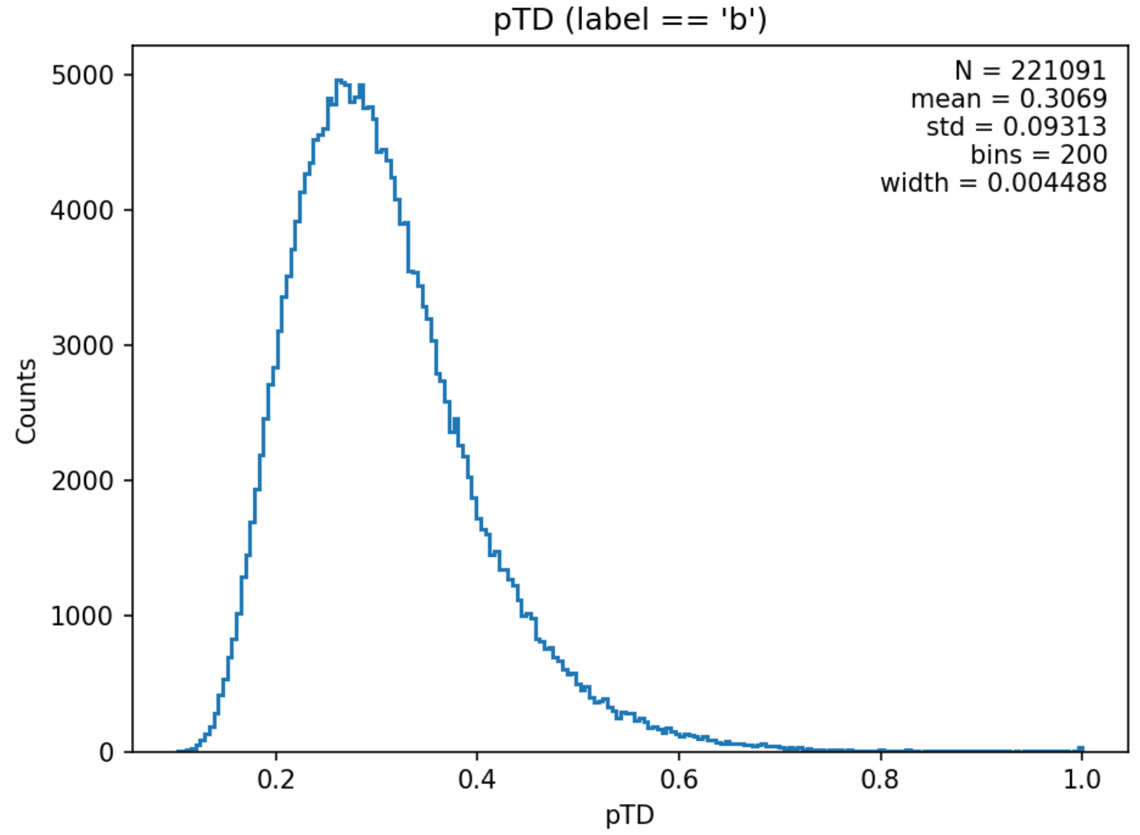






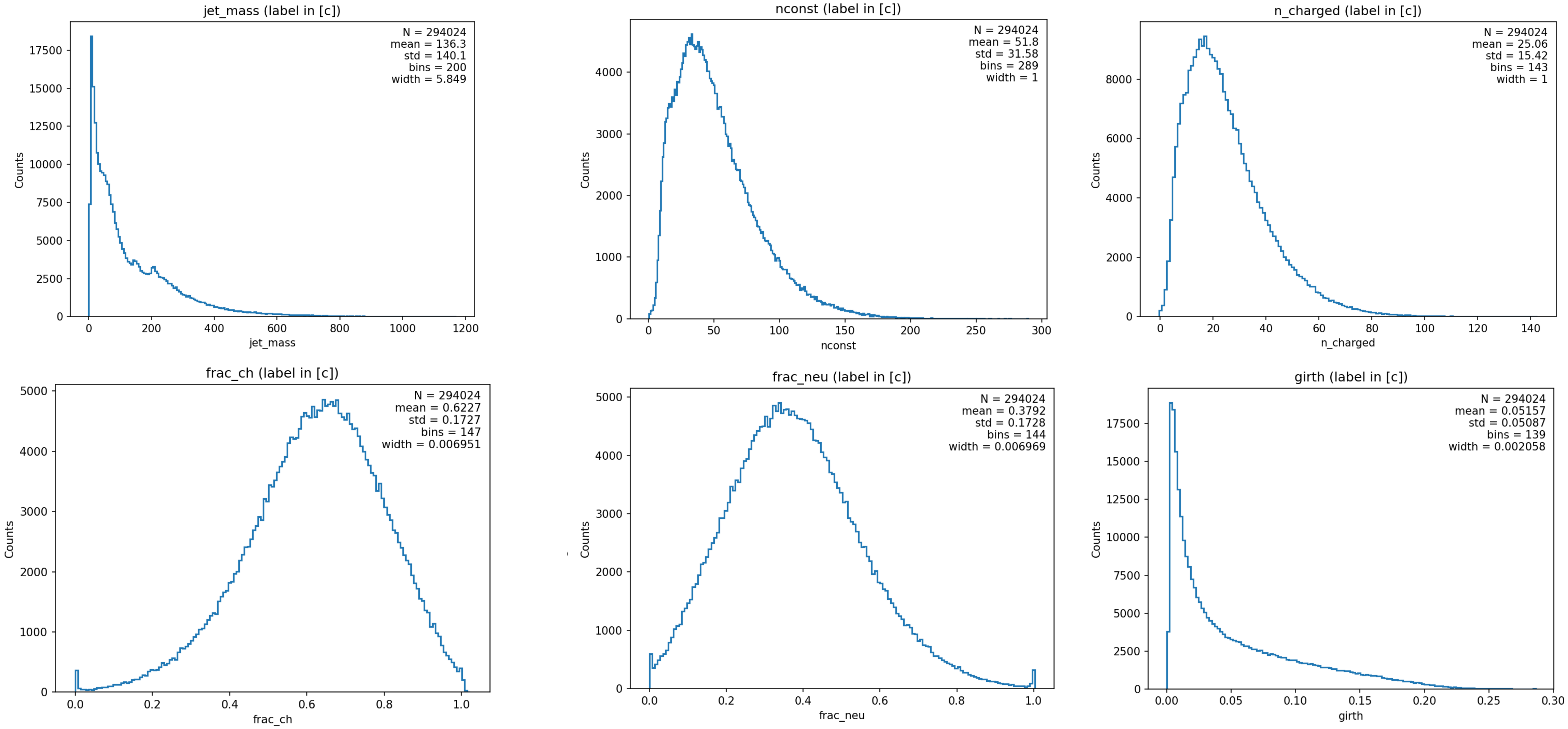






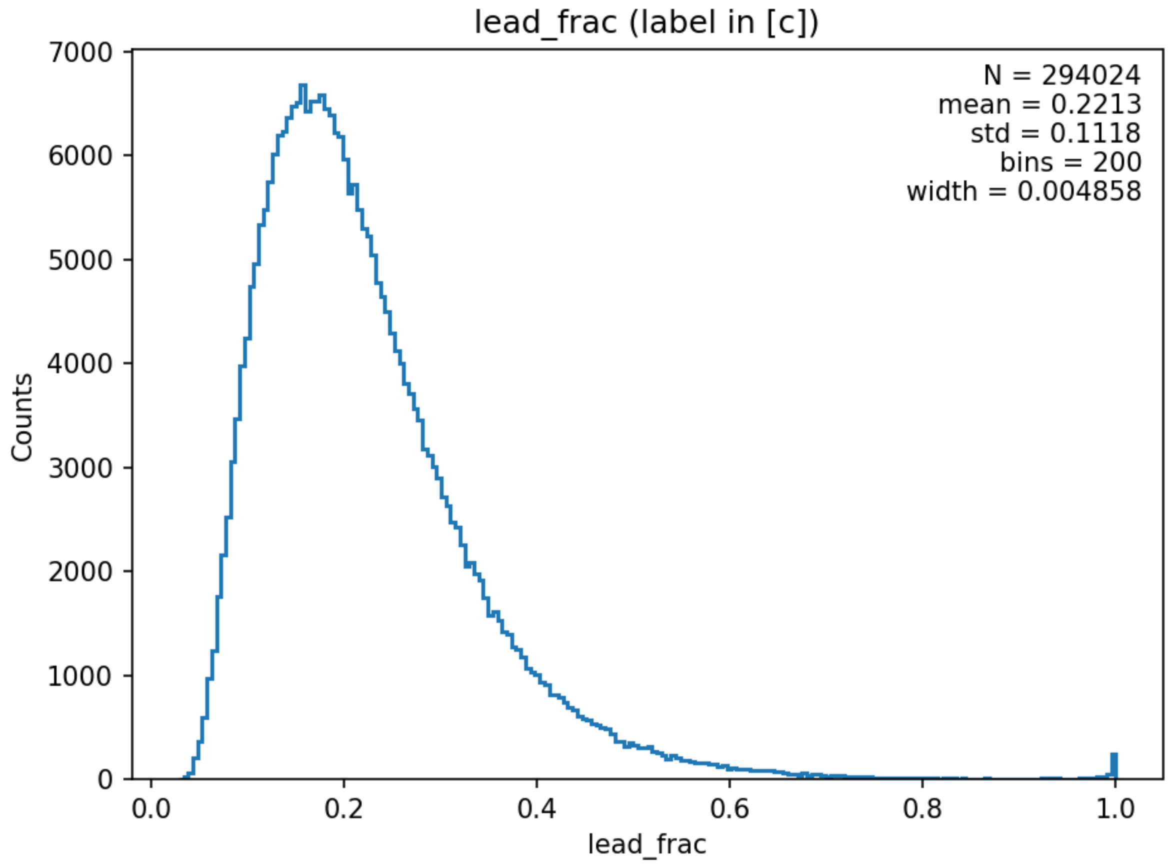
$\mu\mu \to cc$ Feature Plots

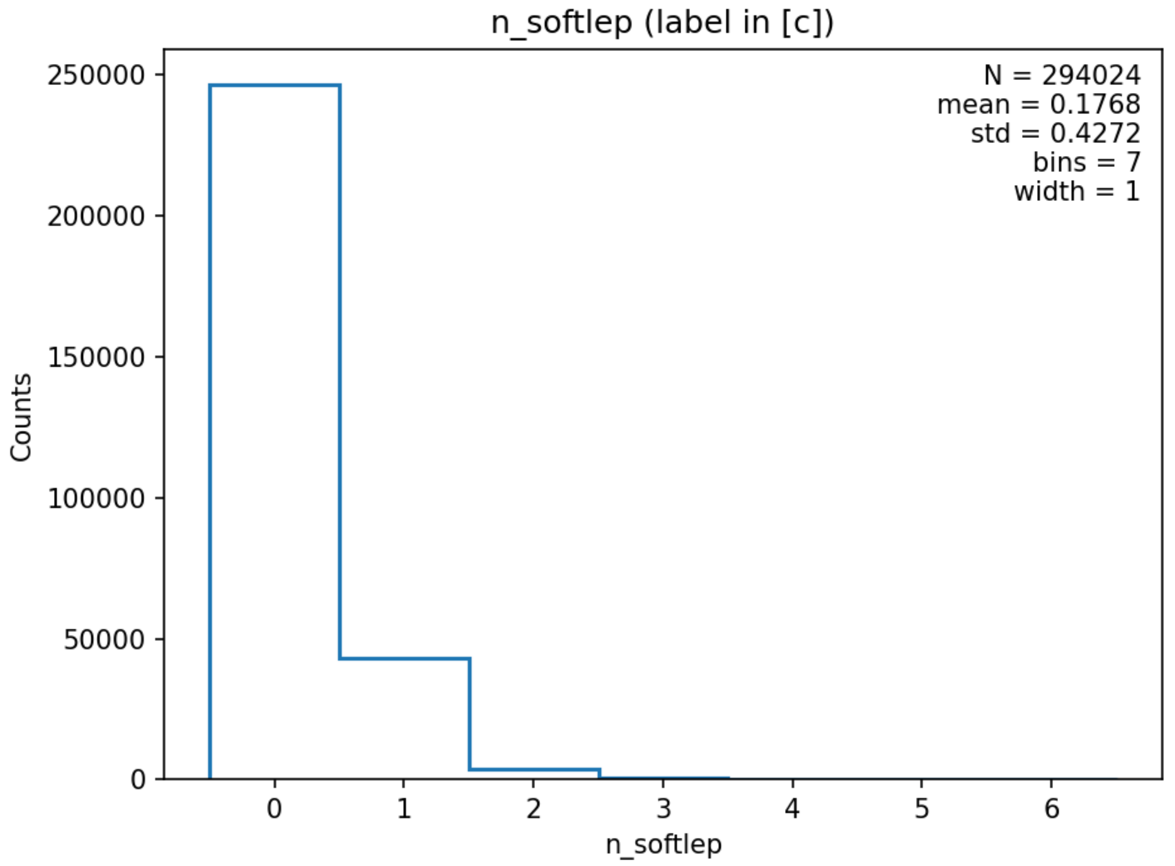
For **c-jets**

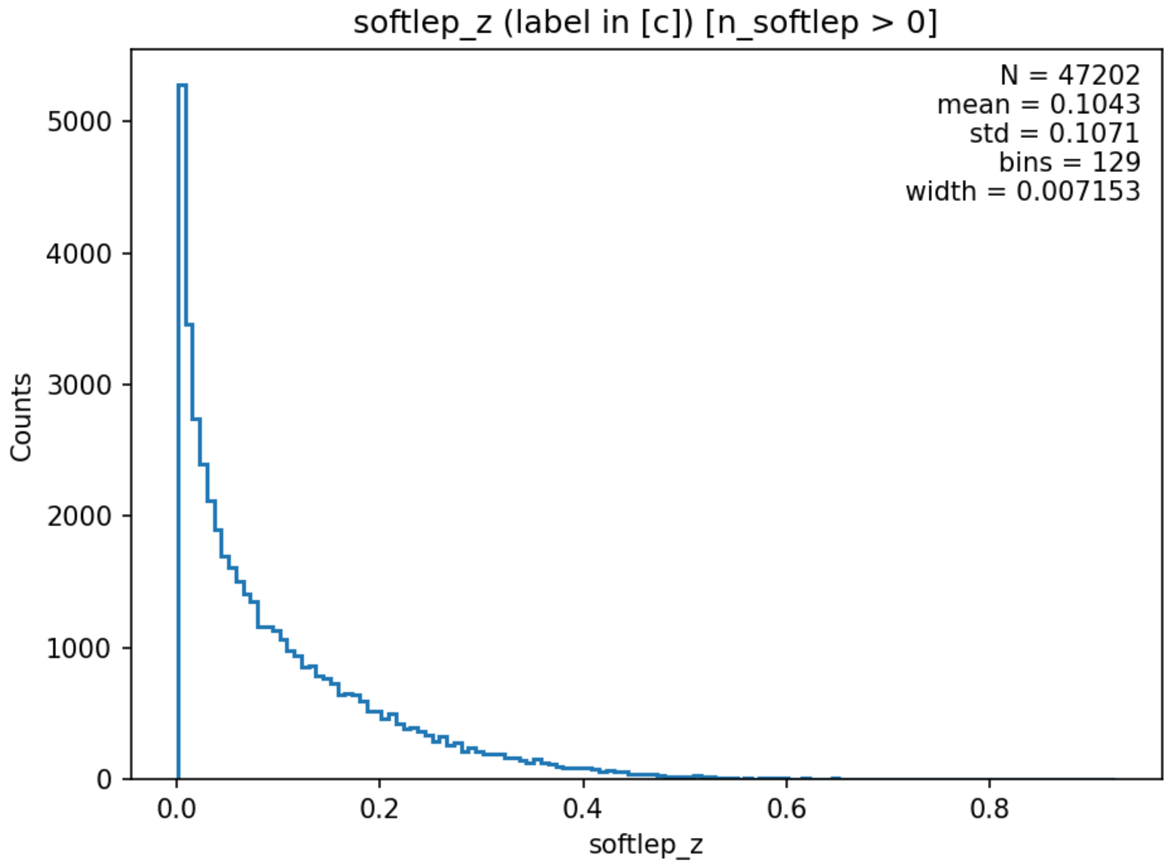


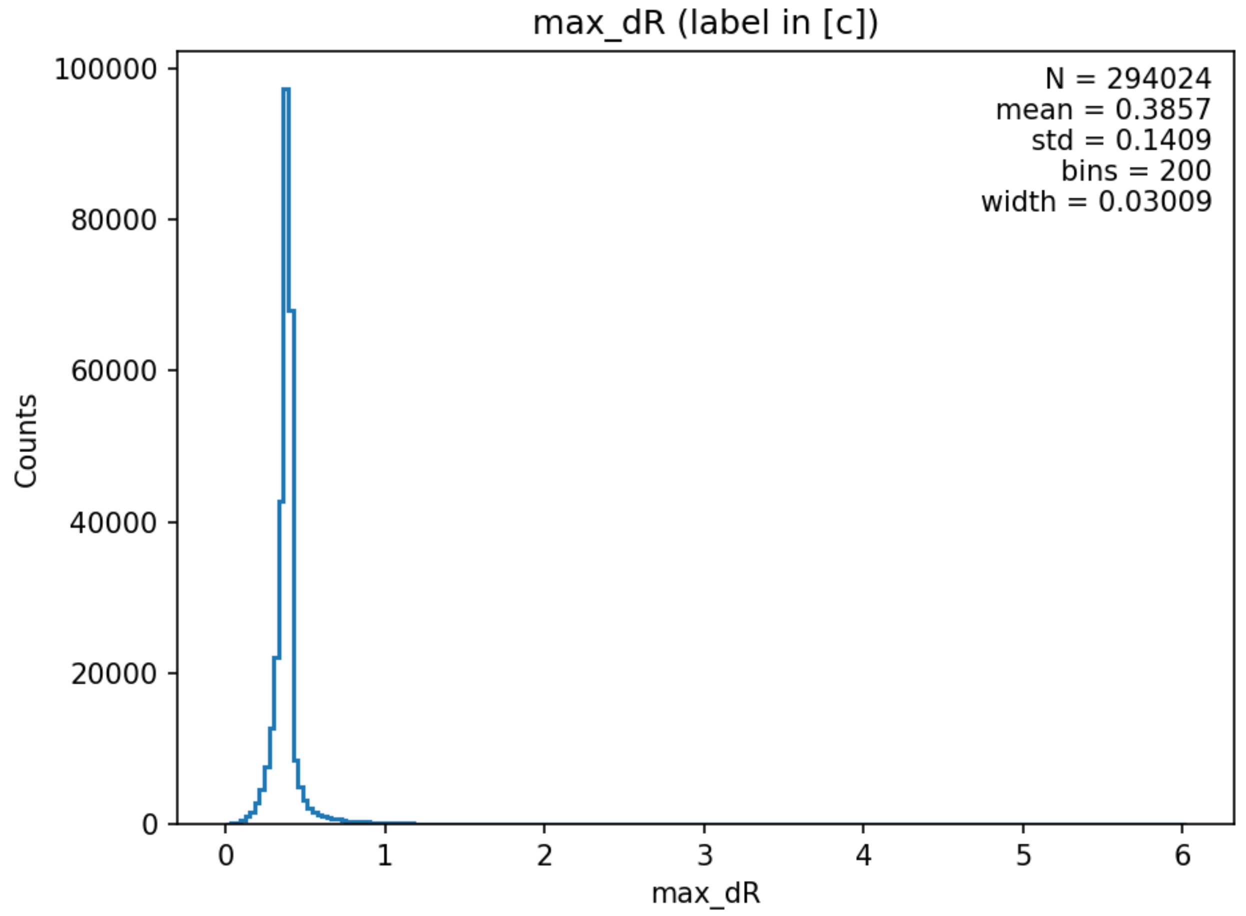
$\mu\mu \rightarrow cc$ Feature Plots

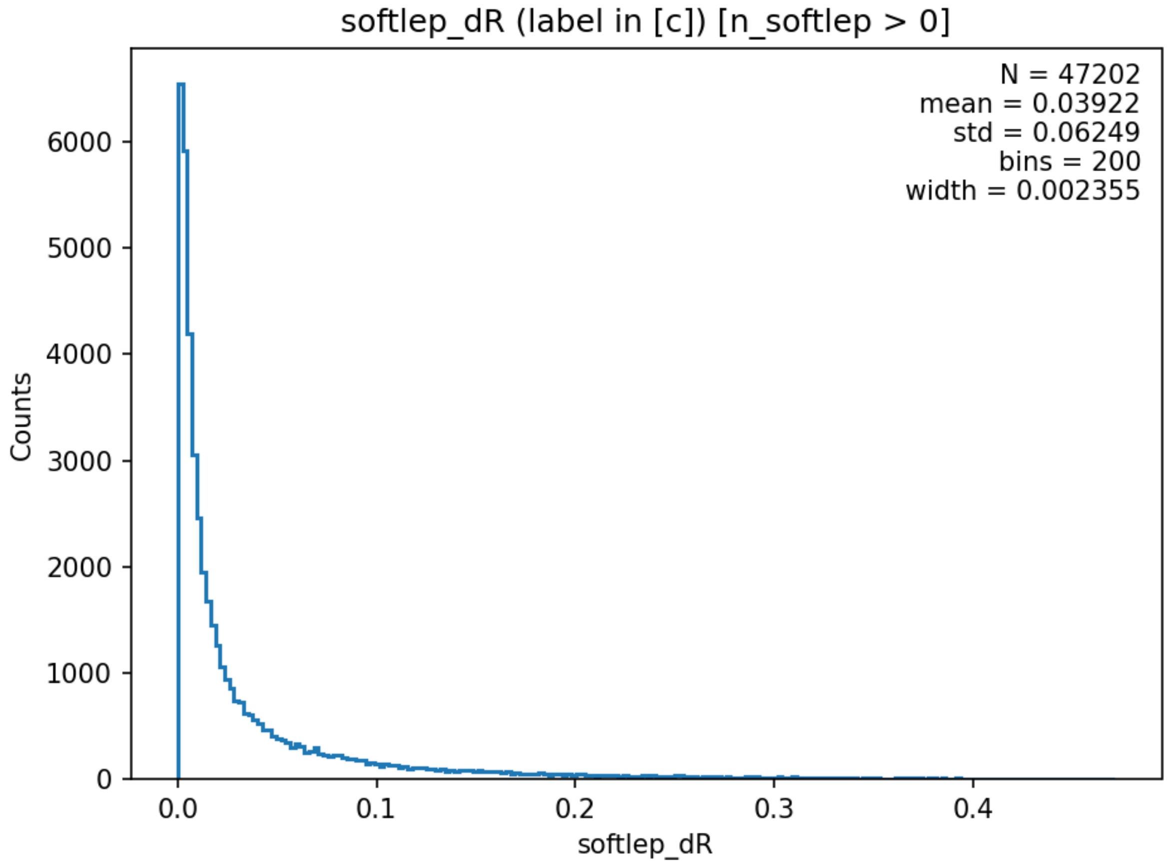
For **c-jets**

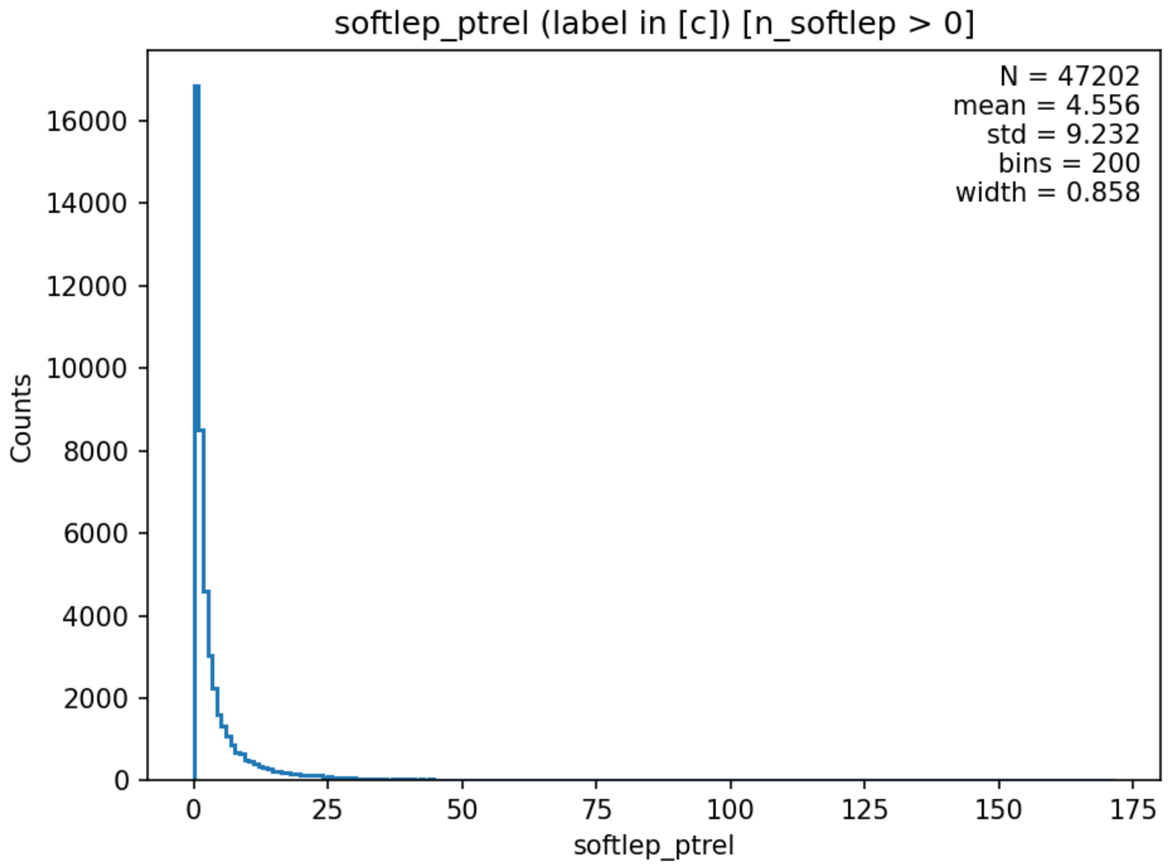


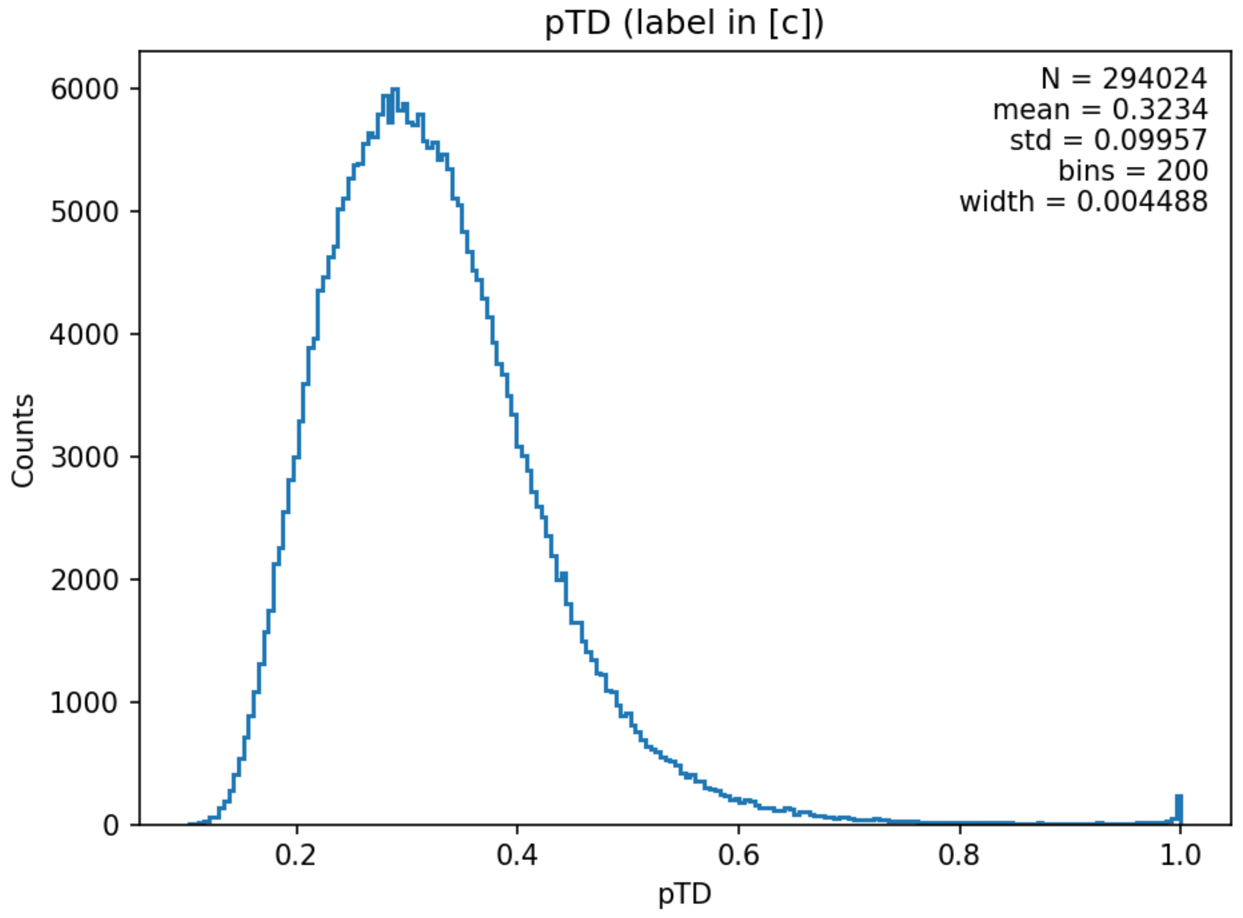






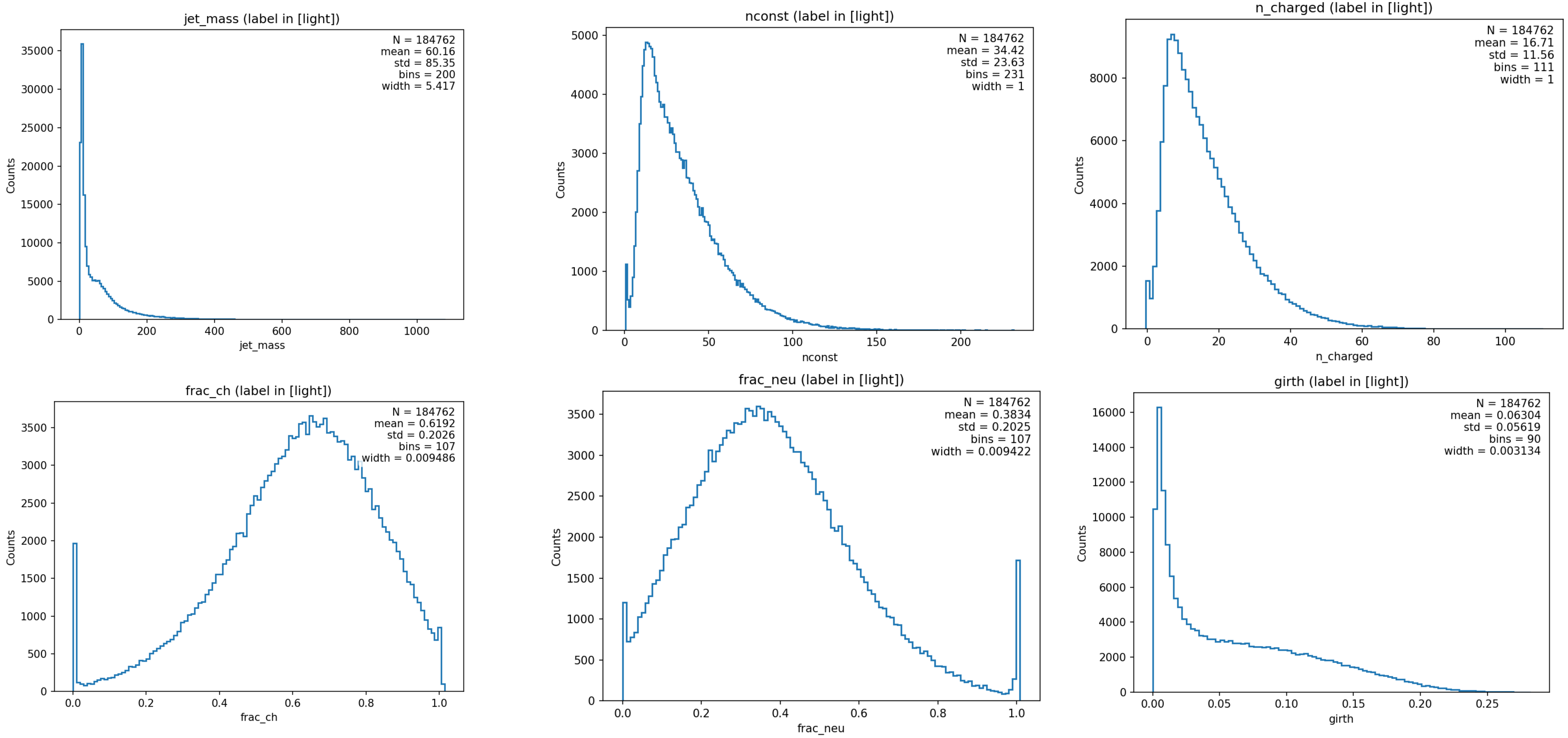






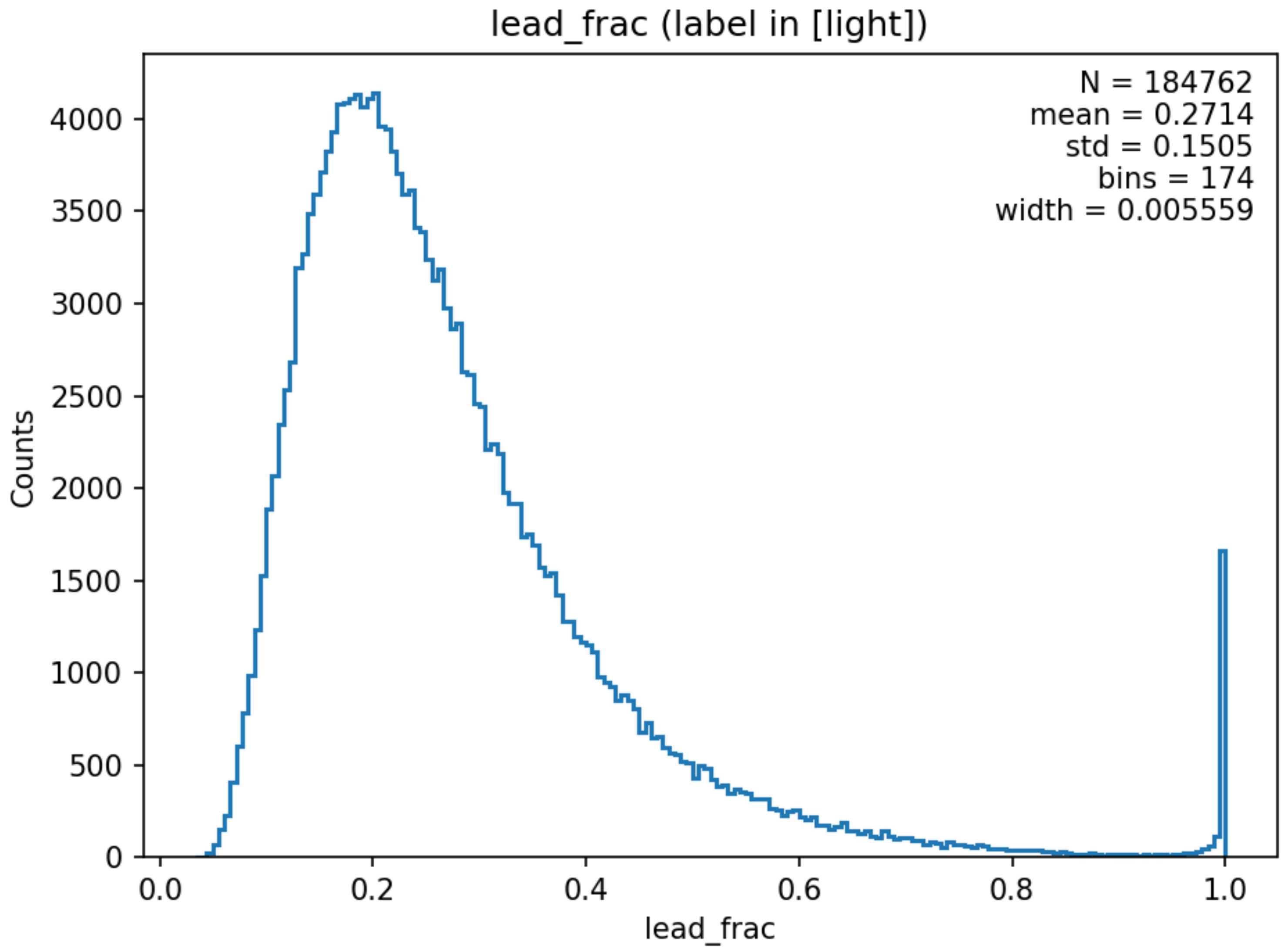
$\mu\mu \rightarrow ll$ Feature Plots

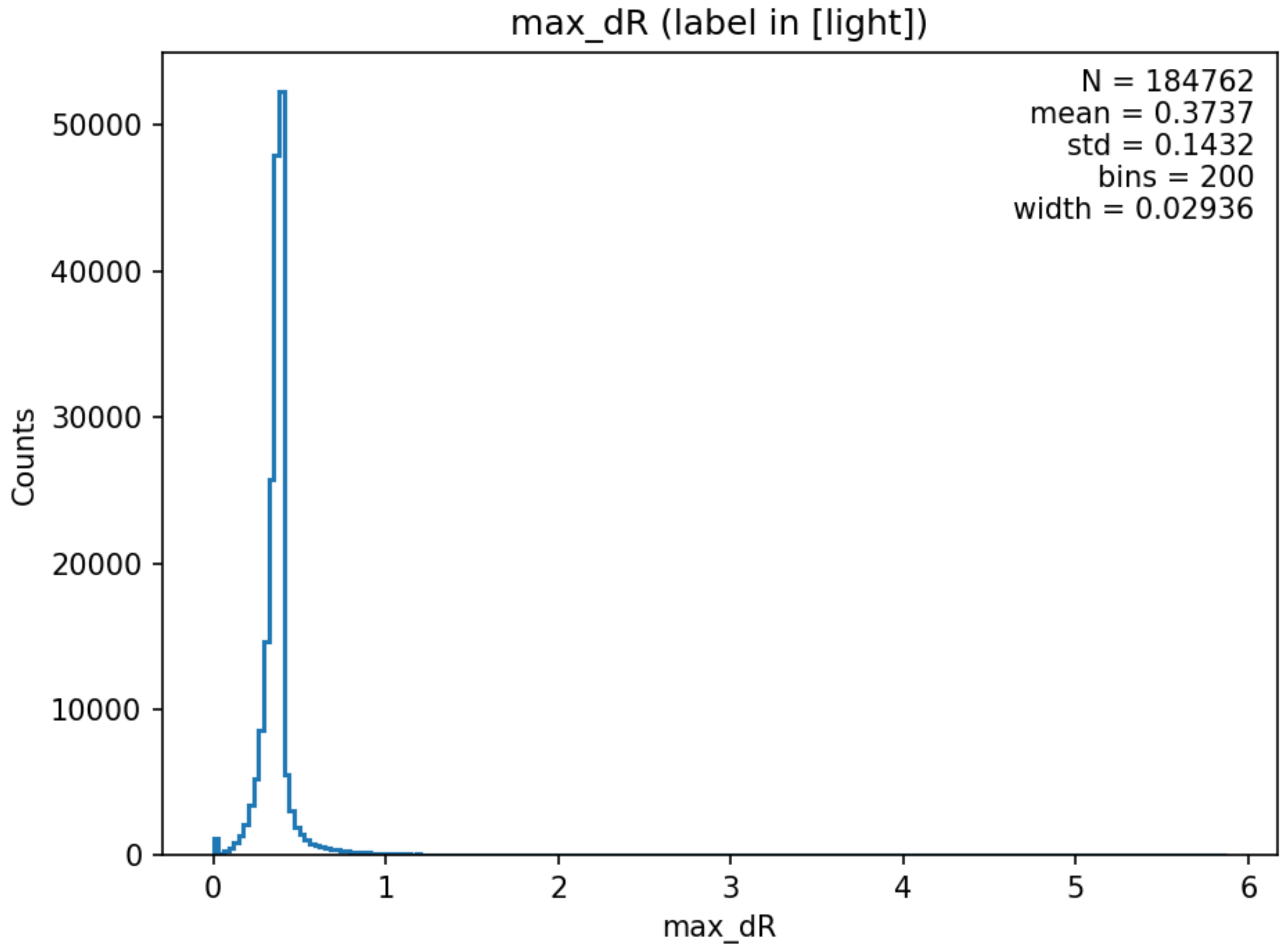
For **light jets**

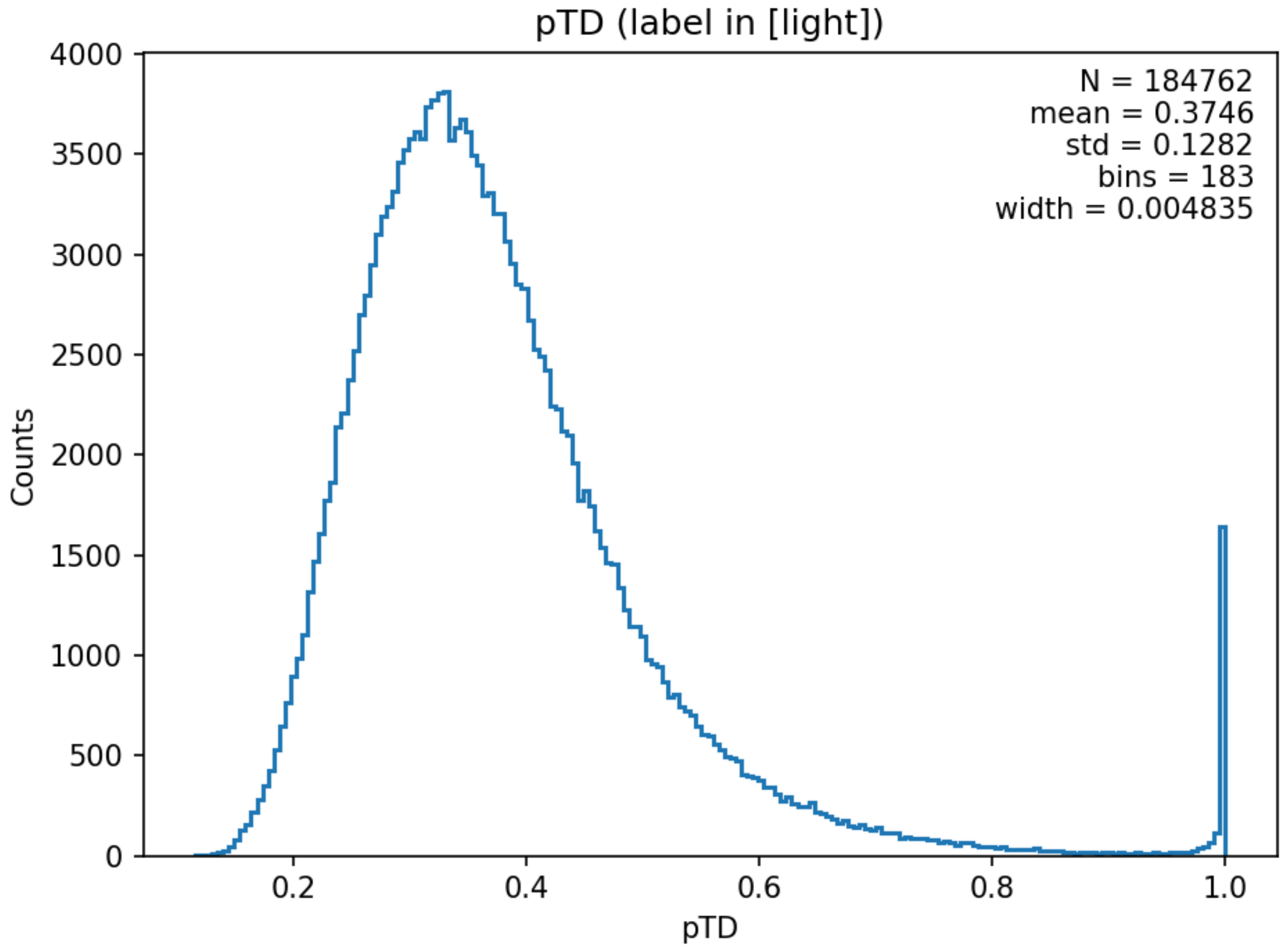


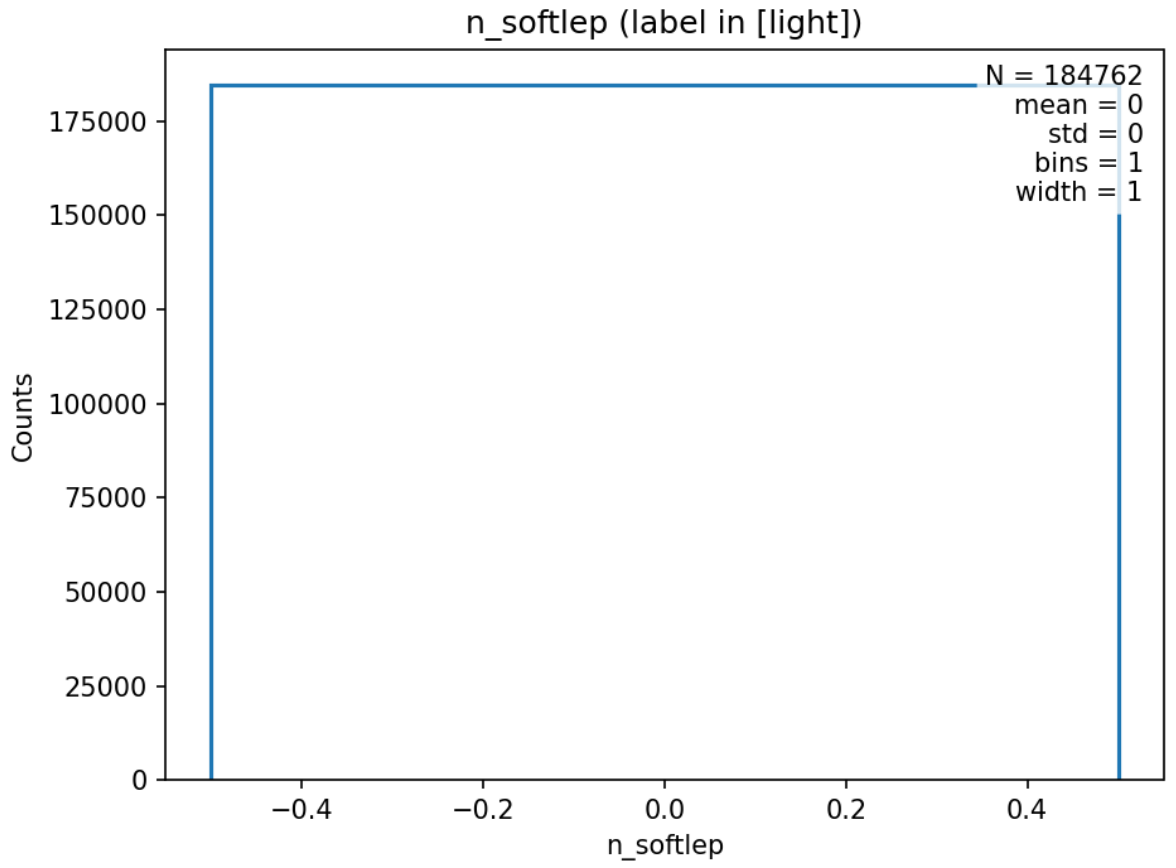
$\mu\mu \to ll$ Feature Plots

For **light jets**





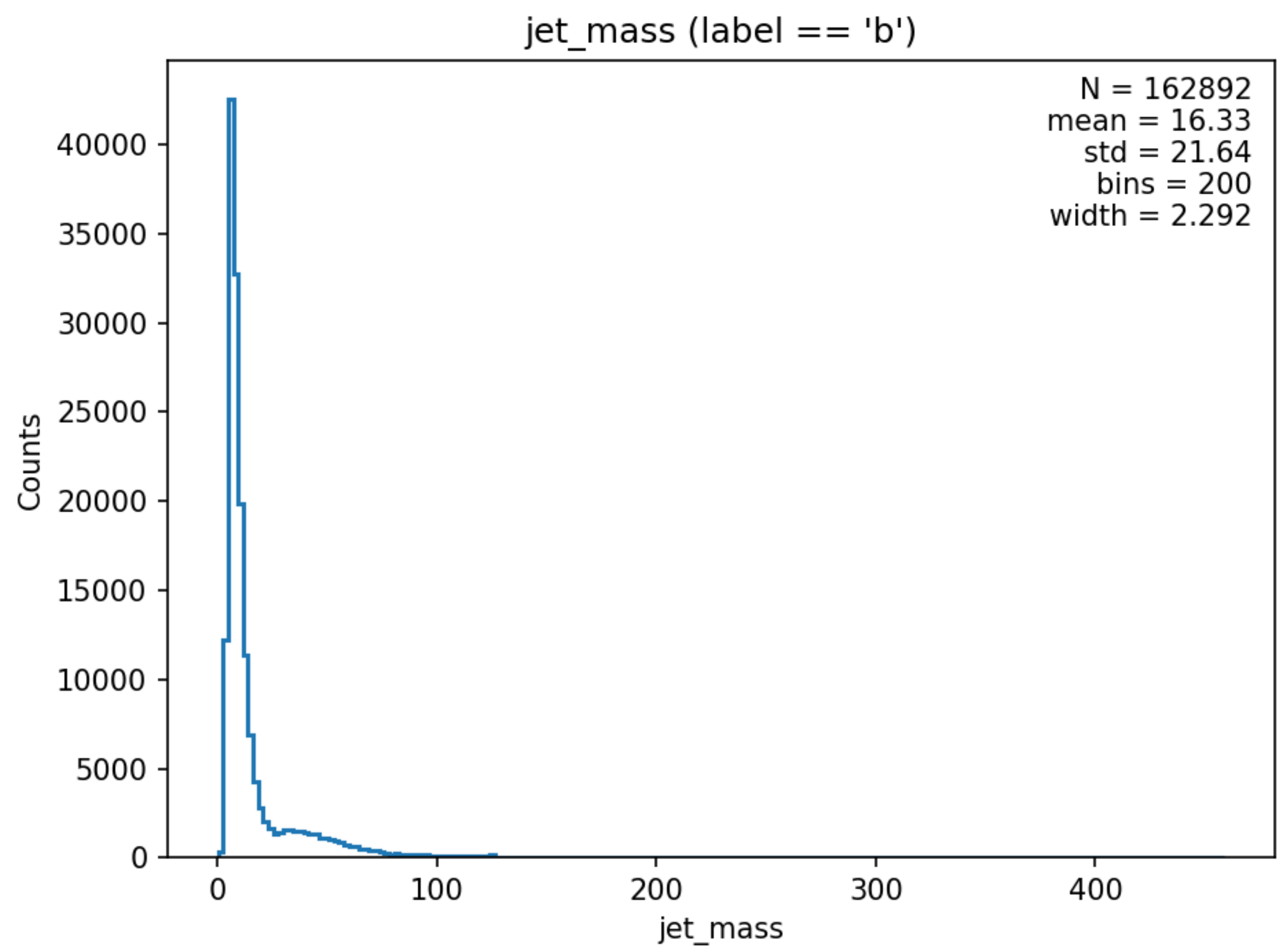


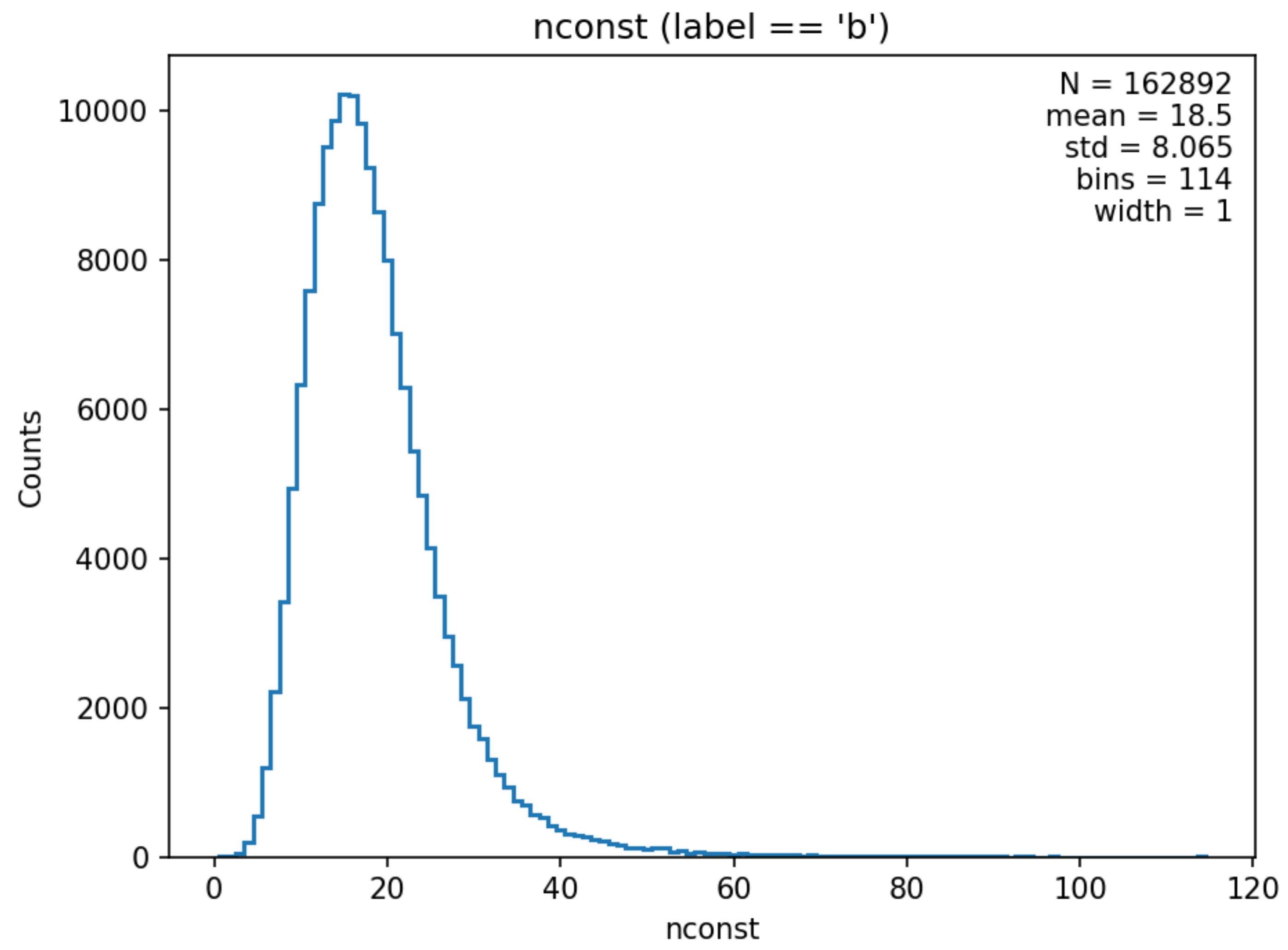


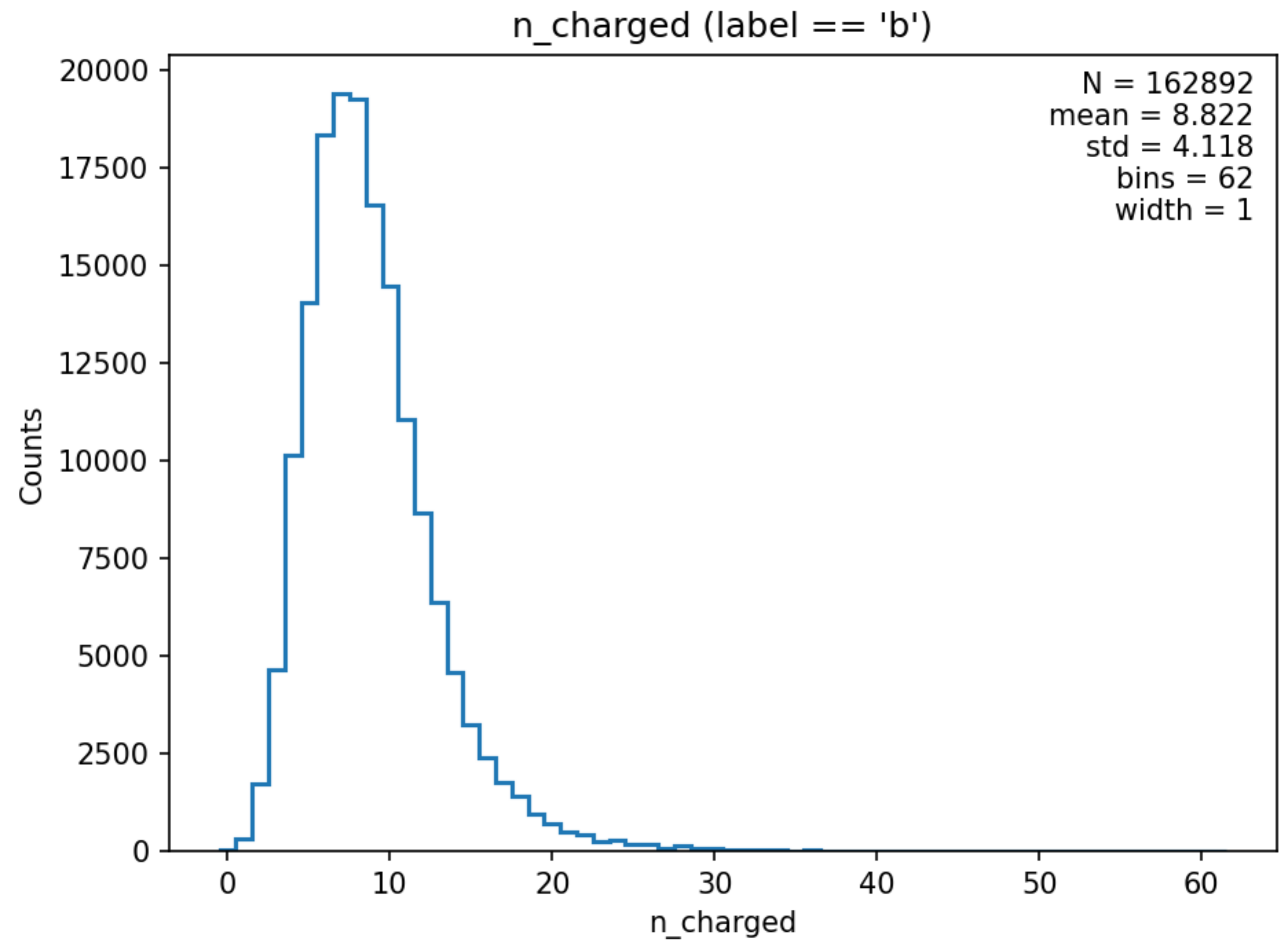
No soft leptons in light jets => no soft lepton plots

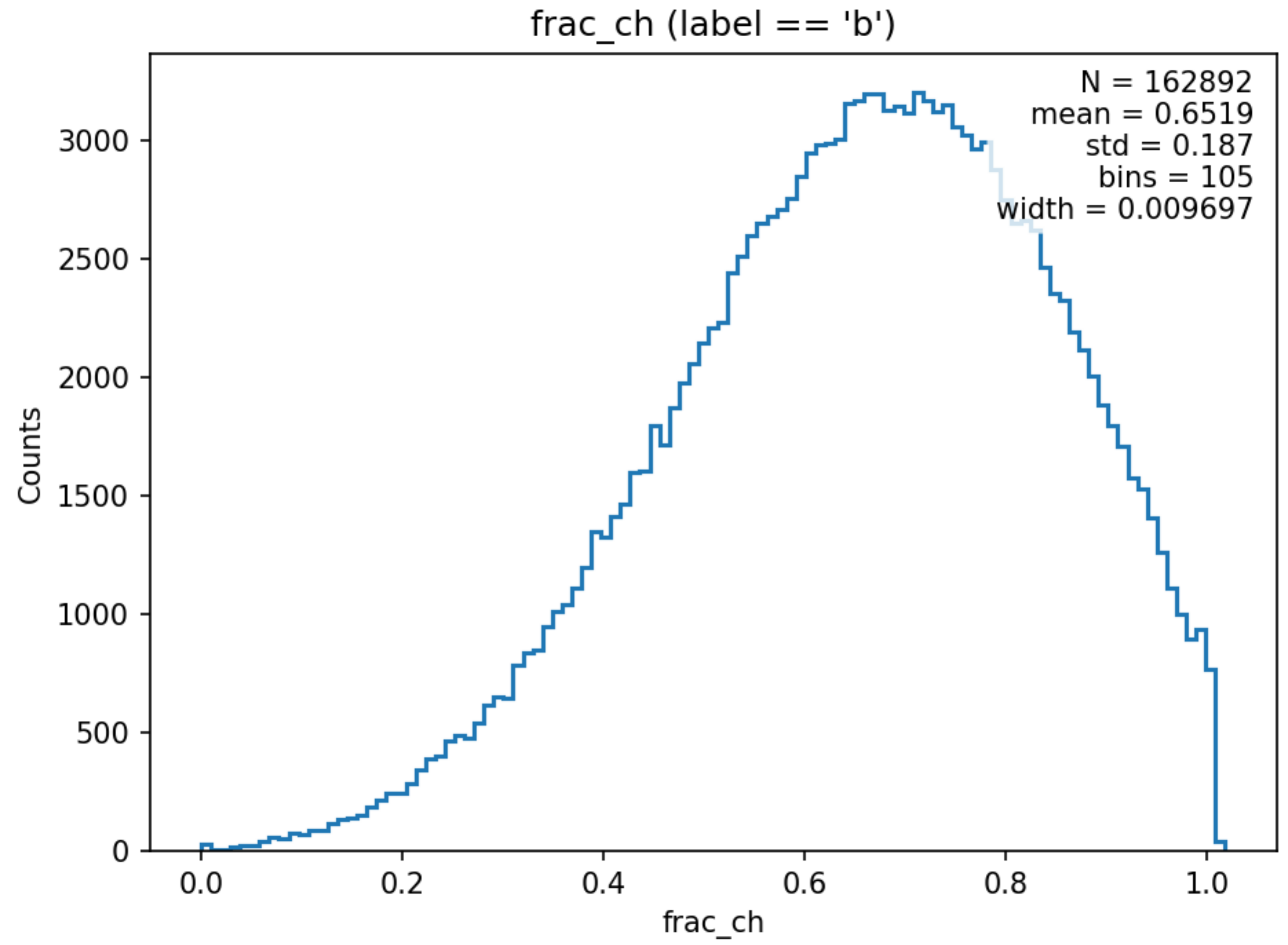
$H \to bb$ Feature Plots

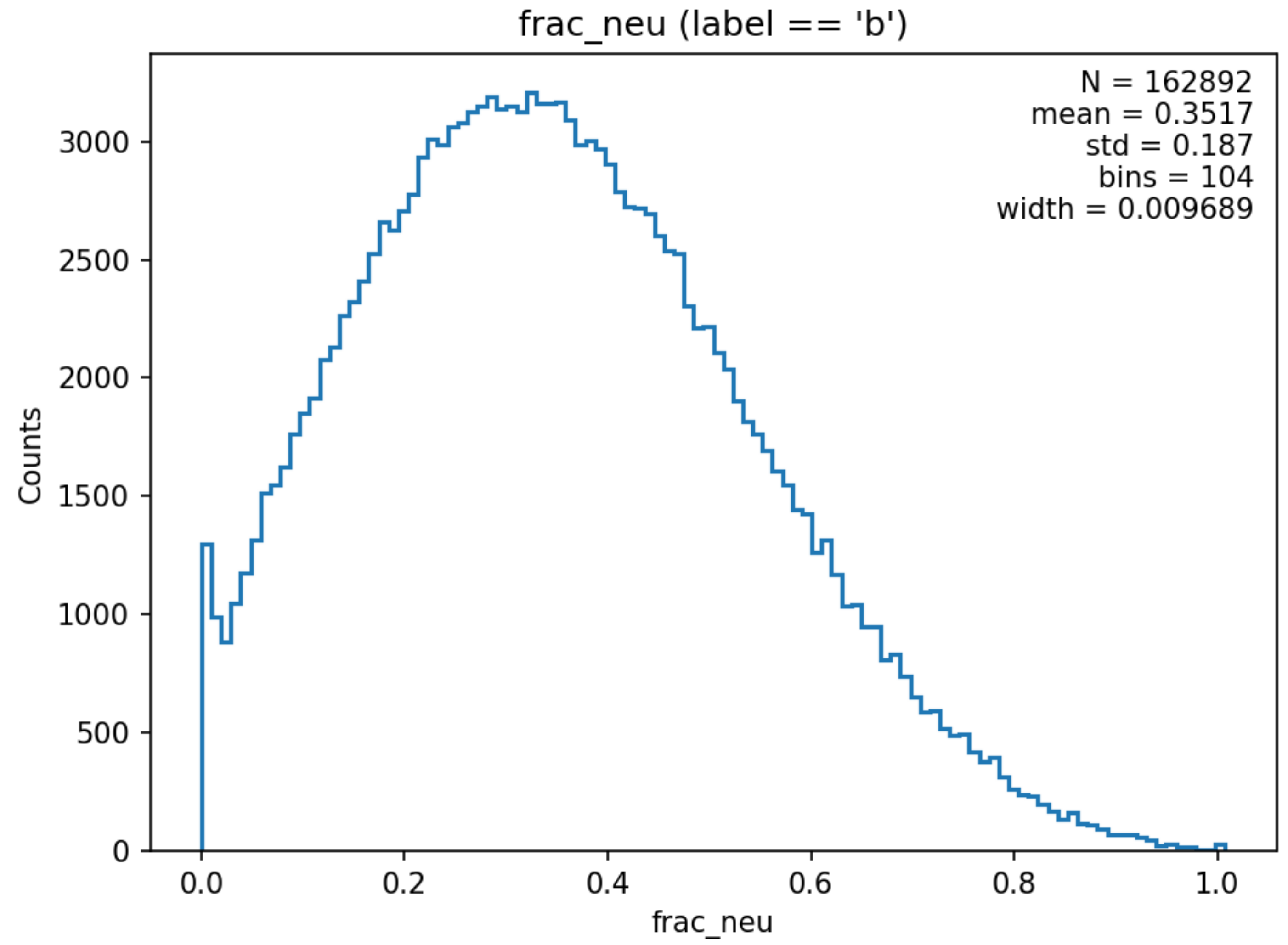
For **b-jets**

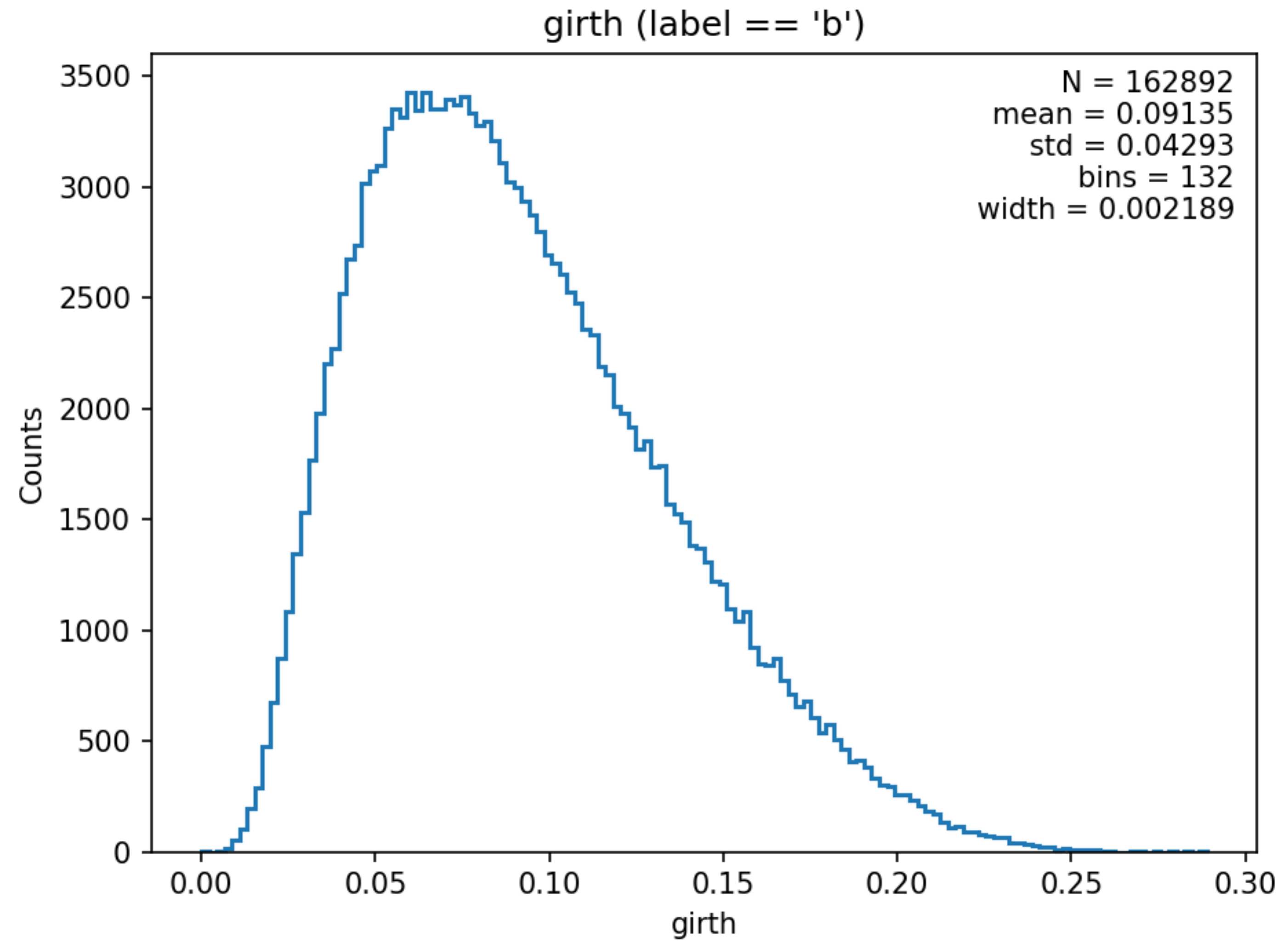






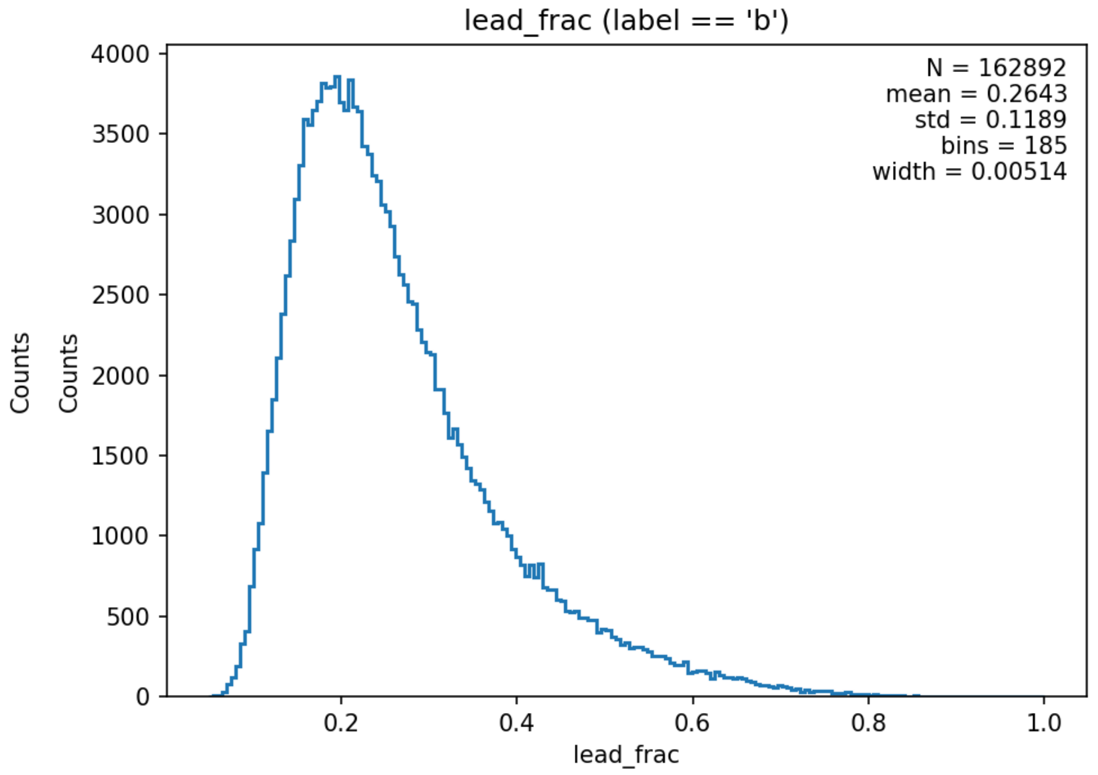


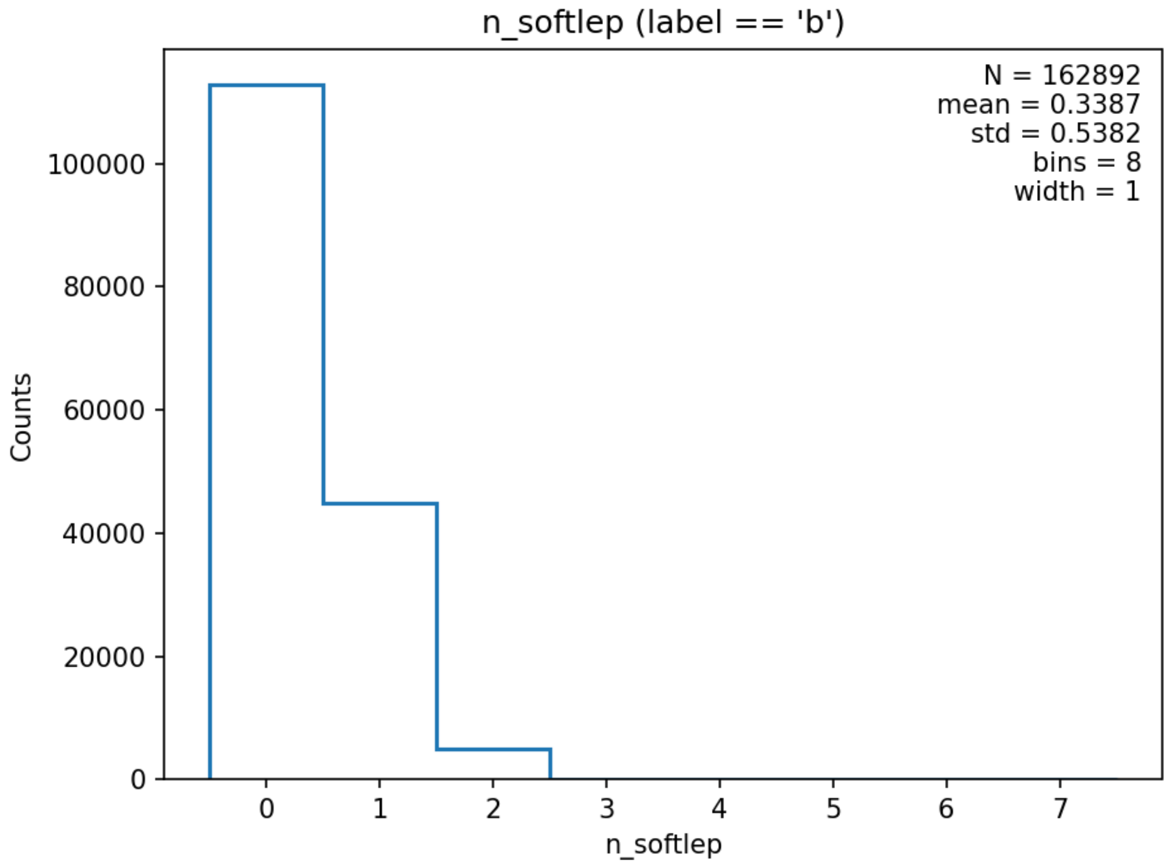


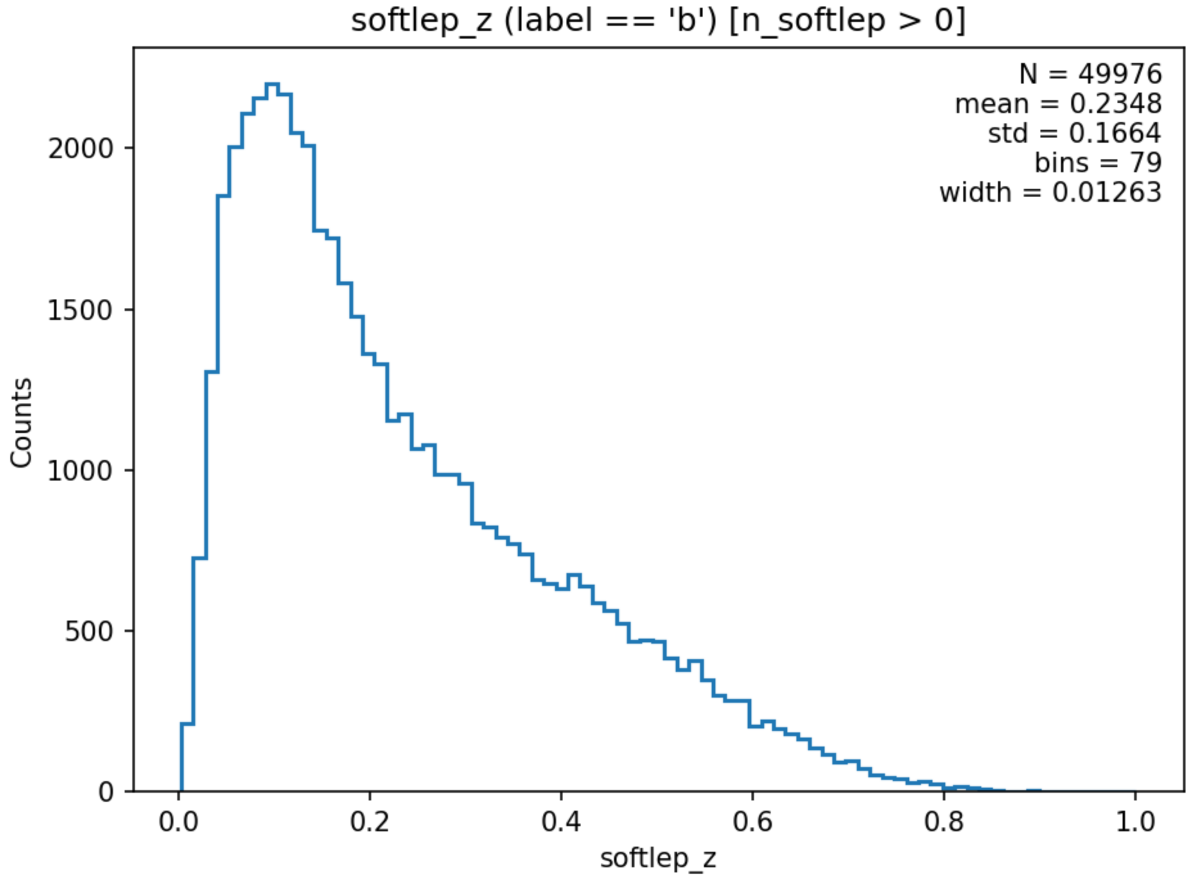


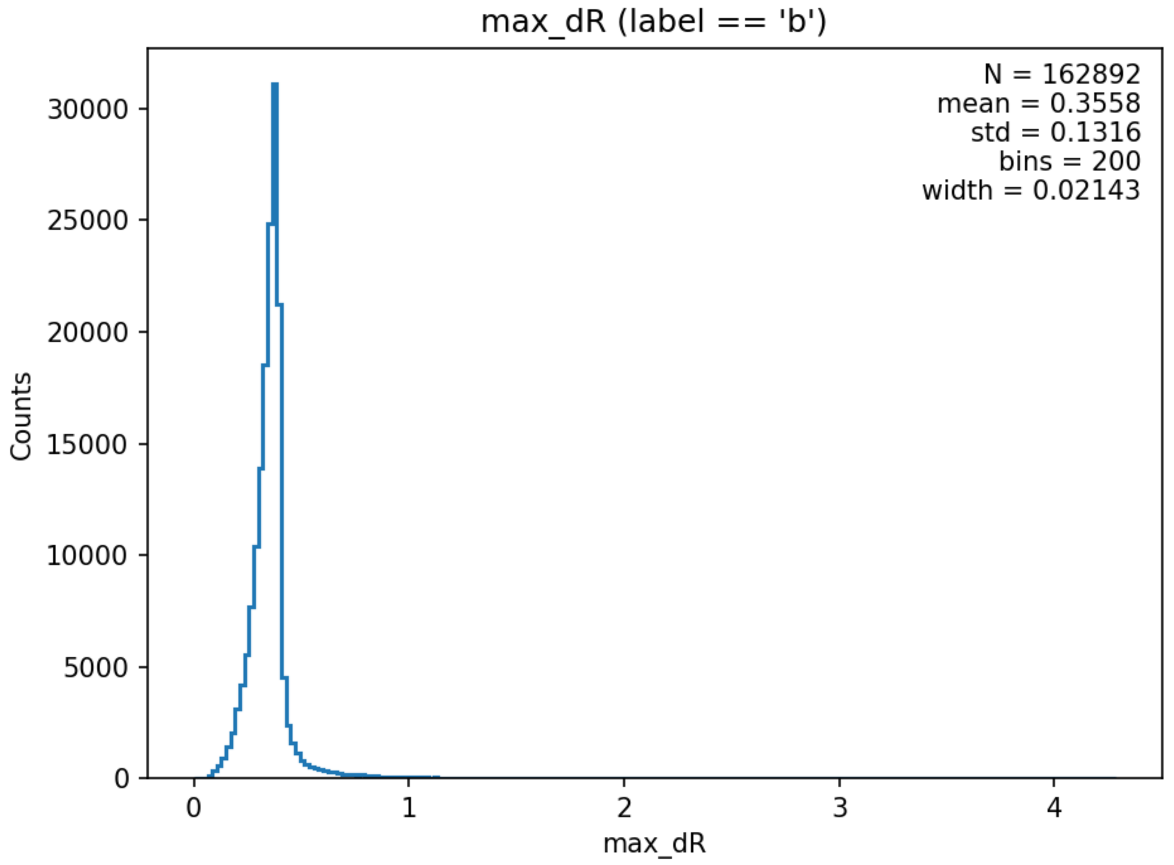
$H \to bb$ Feature Plots

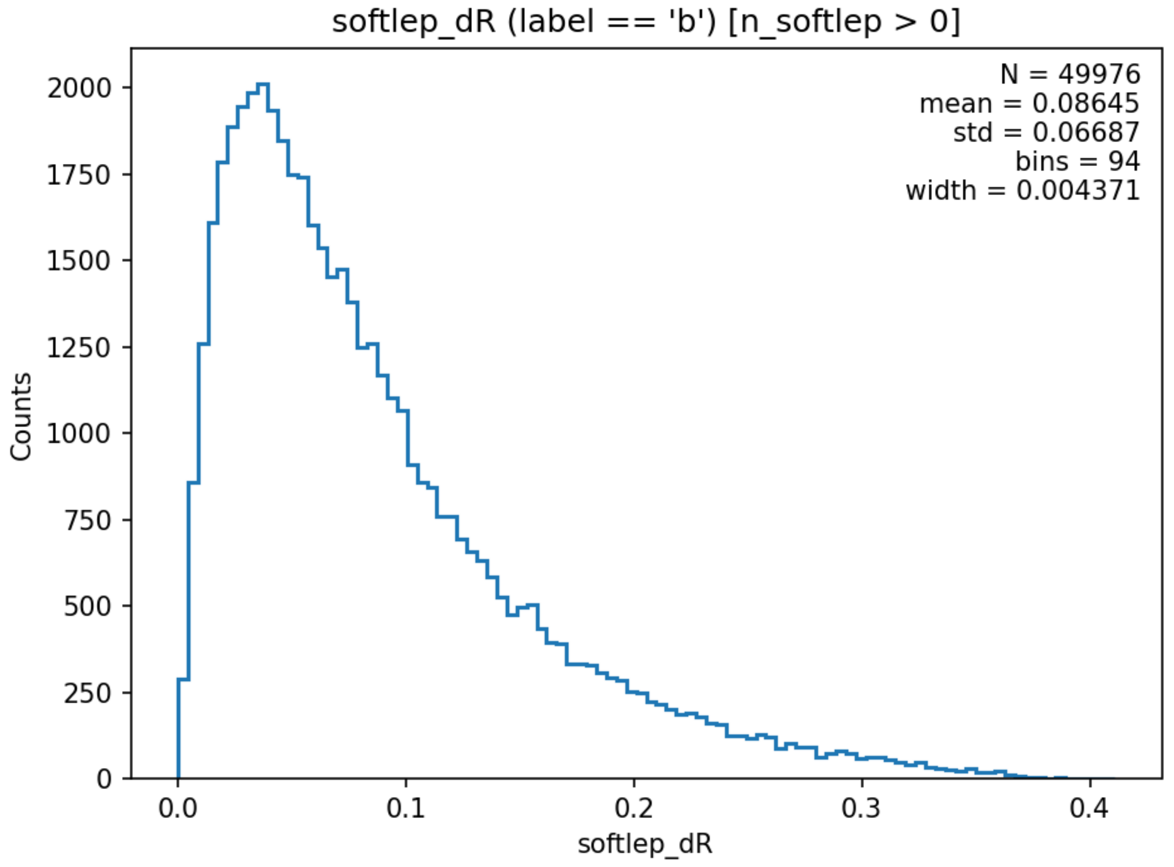
For **b-jets**

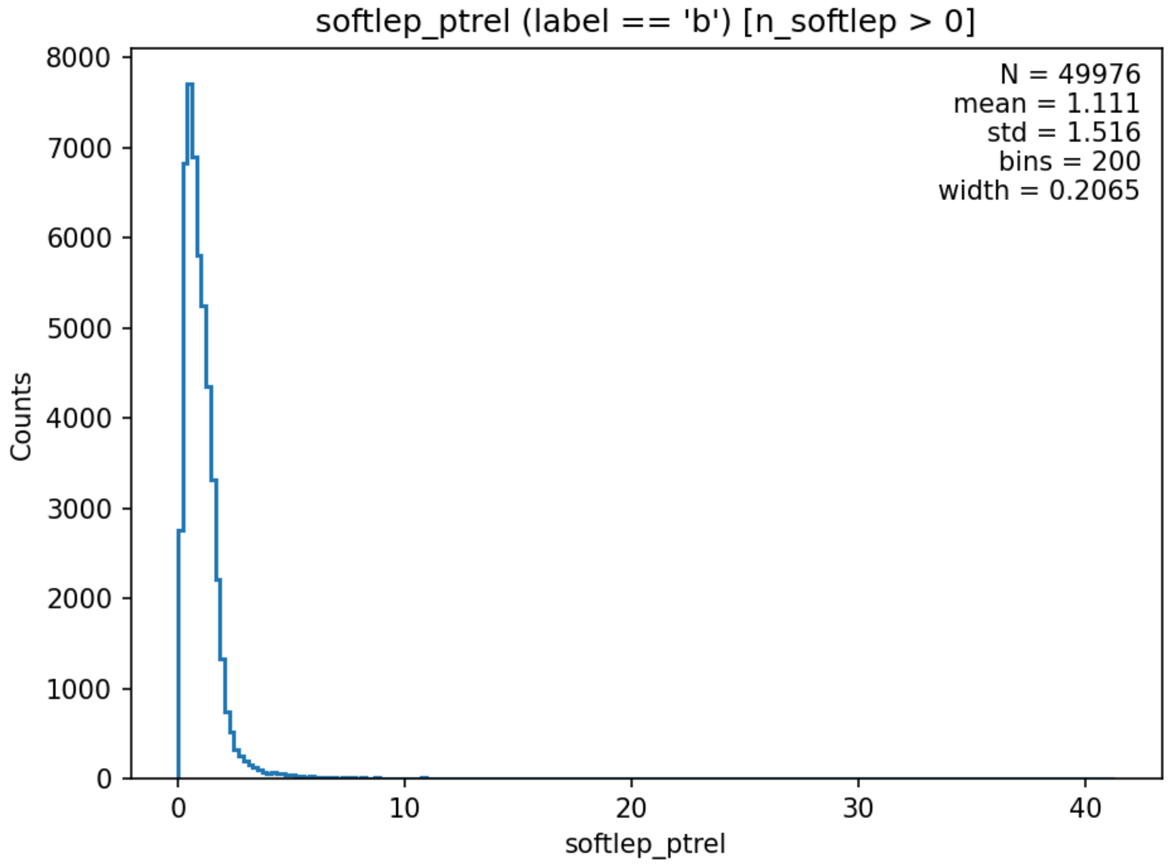


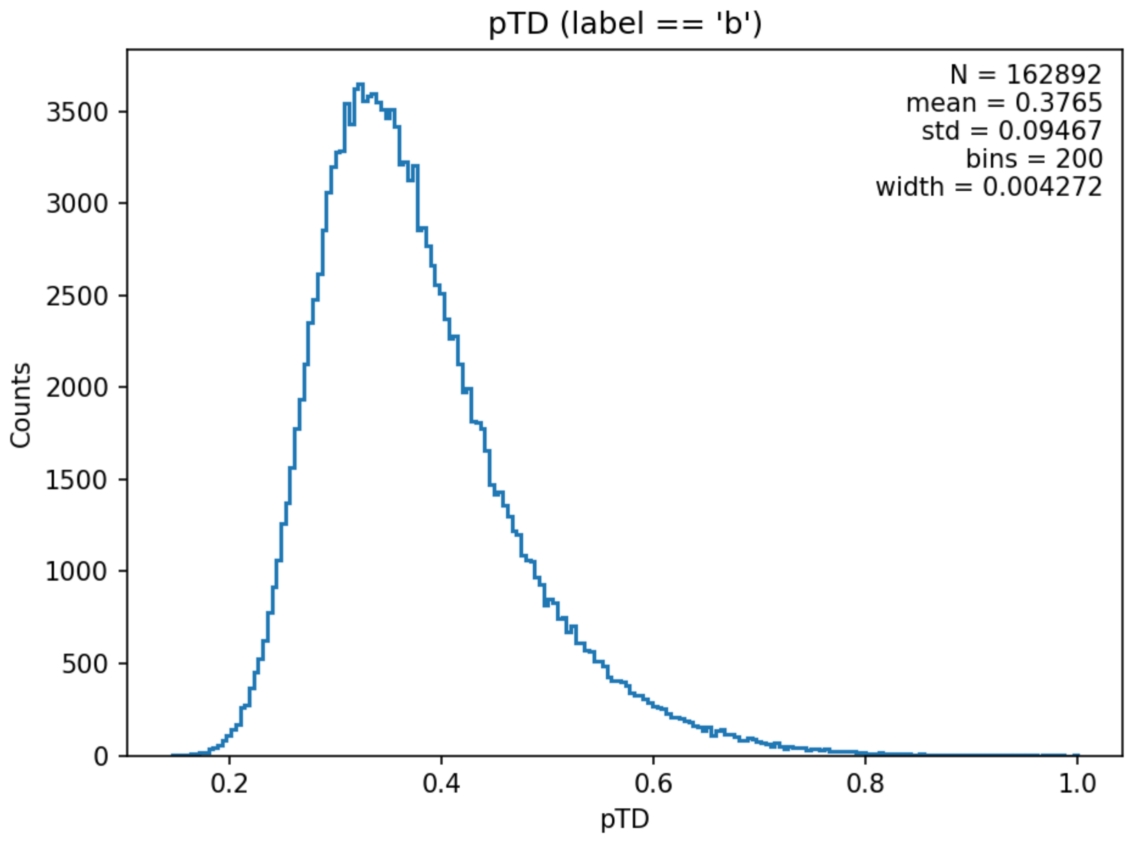










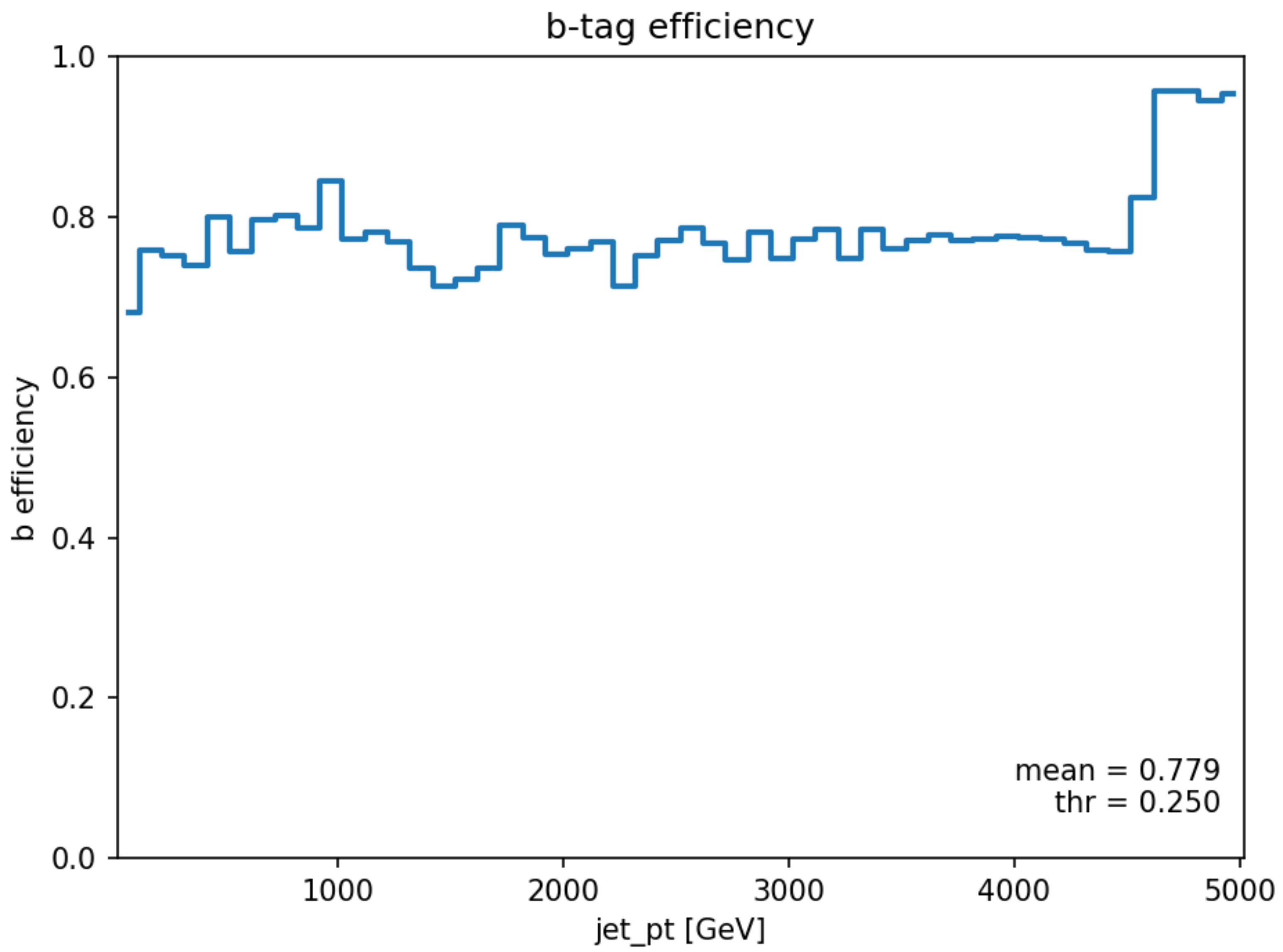


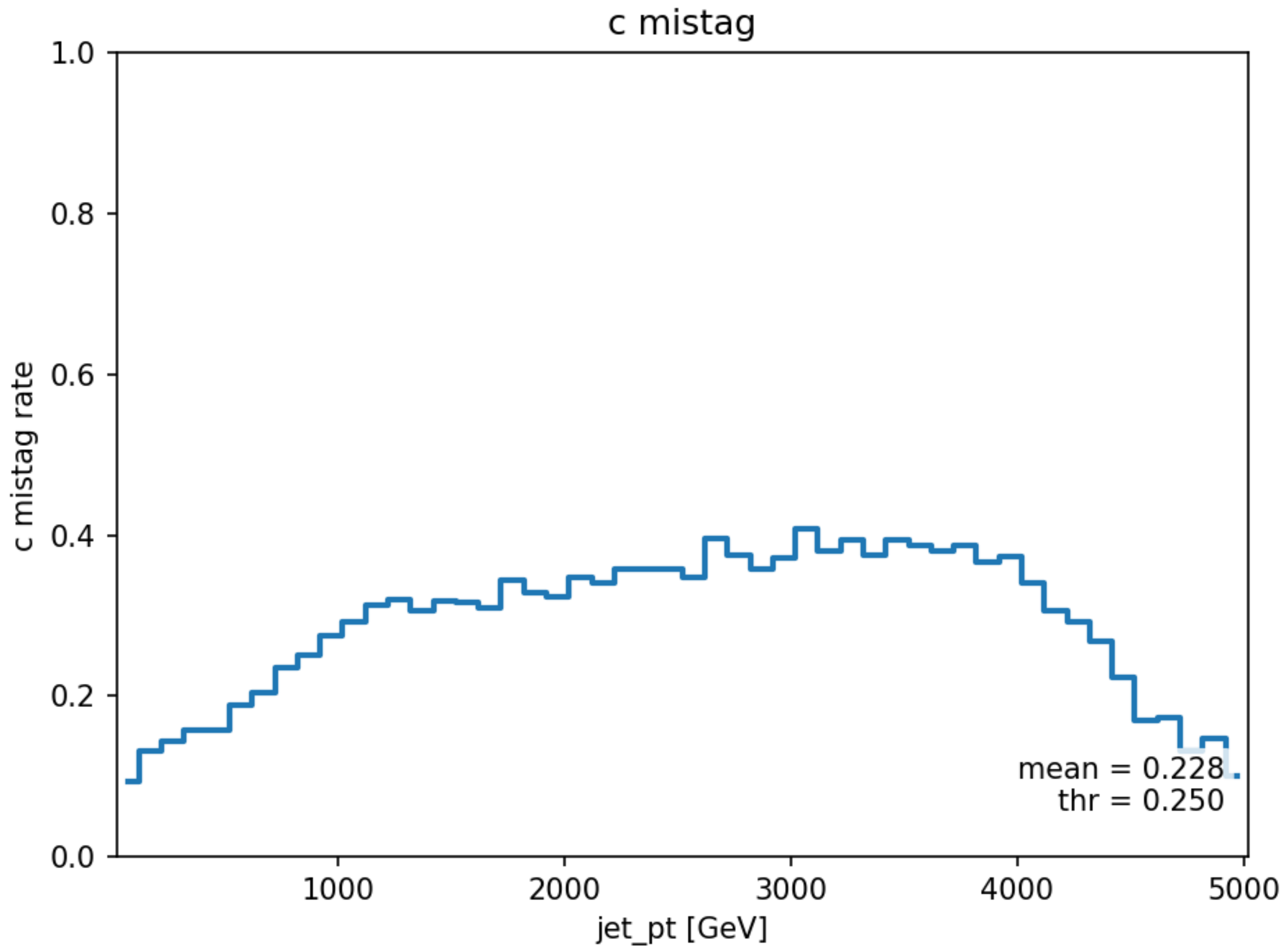
Neural Network Design

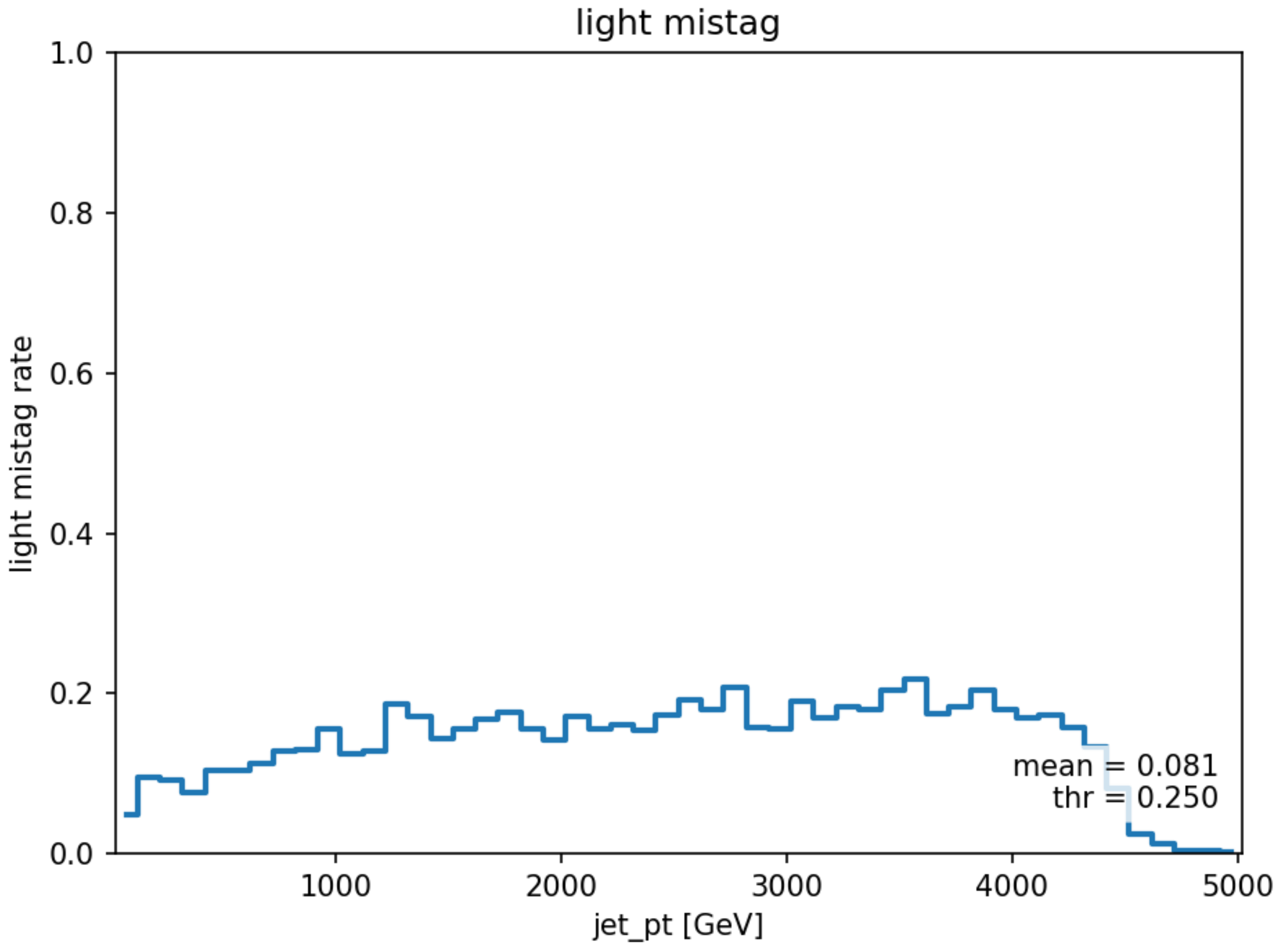
- **Simple setup:** Fully-connected MLP with three hidden layers
- **BatchNorm** + **ReLU** + **Dropout** after each hidden layer
- Features are **standardized** during pre-processing
- Outputs **three logits**: {b, c, light}
- **O(10k) parameters**
- Loss measured in **cross-entropy**
- **Re-weights by pT** so the model learns **flavor**, not kinematics
- Training **only** on $\mu\mu \to bb$, $\mu\mu \to cc$, $\mu\mu \to ll$ for now

Results

- Running on test set, using a b-score threshold of 0.25:







$$\frac{N(\text{true b, pred b})}{N(\text{true b})}$$

$$\frac{N(\text{true c, pred b})}{N(\text{true c})}$$

$$\frac{N(\text{true l, pred b})}{N(\text{true l})}$$

Results

- Now using 0.30: