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

Jeffrey Backus
Princeton University

10/03/25

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

- Using dijet samples located at **/ospool/uc-shared/project/futurecolliders/data/fmeloni/DataMuC_MAIA_v0/v5/reco/dijet_mjj_10000**

- Having some problems running my feature extractor for long periods of time on the cluster

- **Only was able to analyze 1k jets for today**

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

Training Features

Impact parameter features

10. **Max IP** (d0_max and z0_max) in mm

11. **Mean IP** (d0_mean and z0_mean) in mm

12. **Median IP** (d0_med and z0_med) in mm

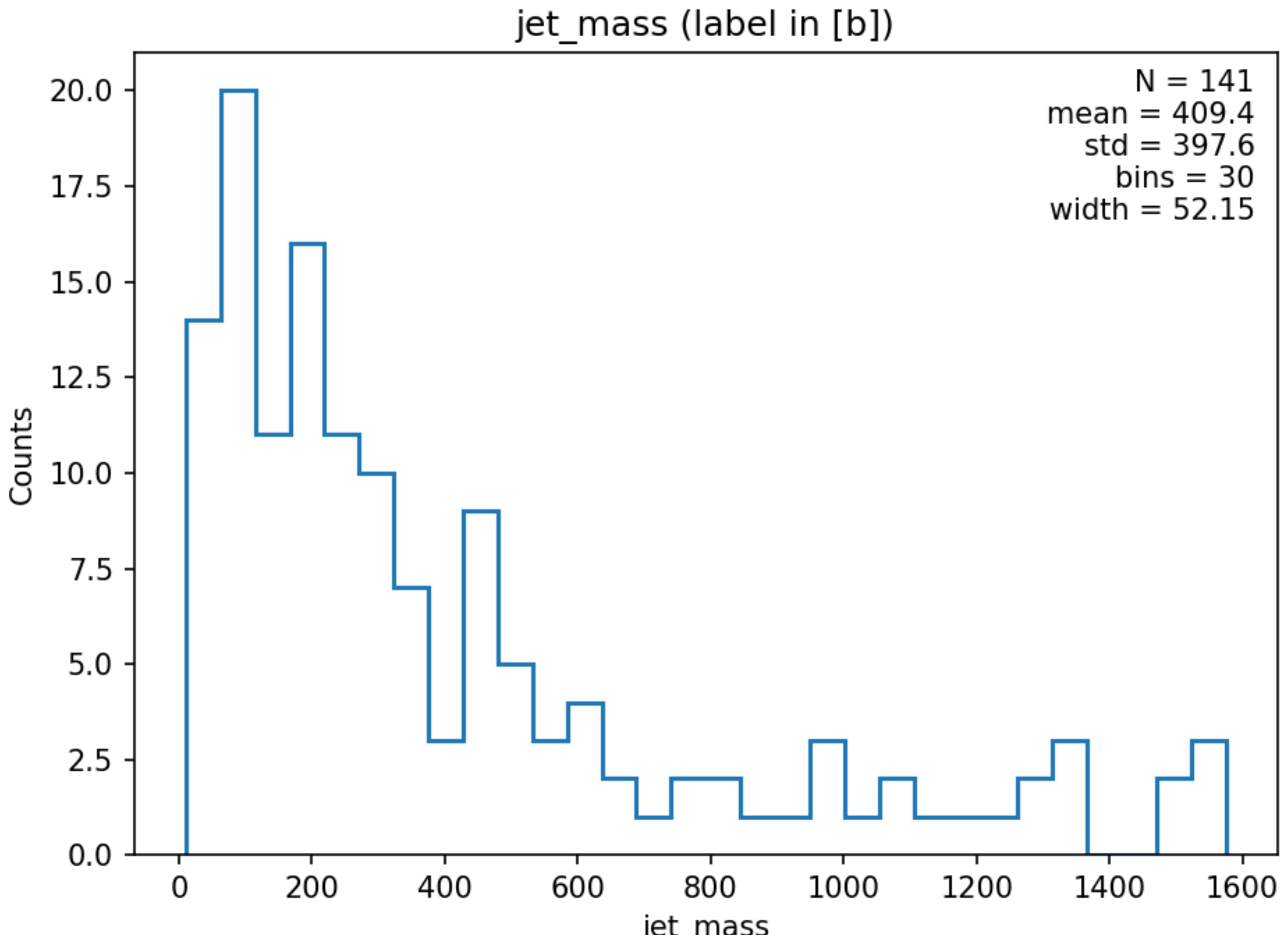
13. **IP significance max** (d0sig_max and z0sig_max)

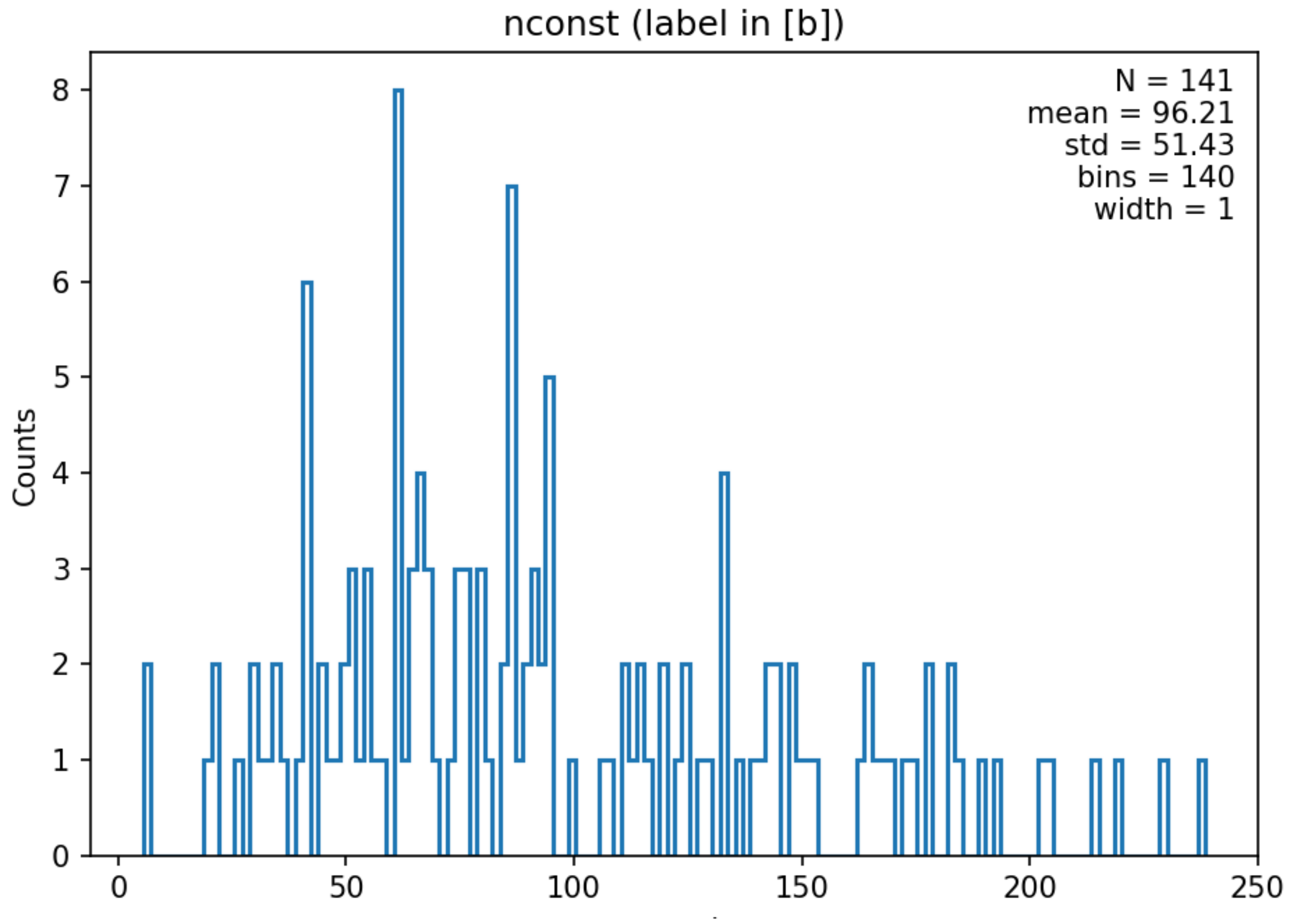
14. **IP significance mean** (d0sig_mean and z0sig_mean)

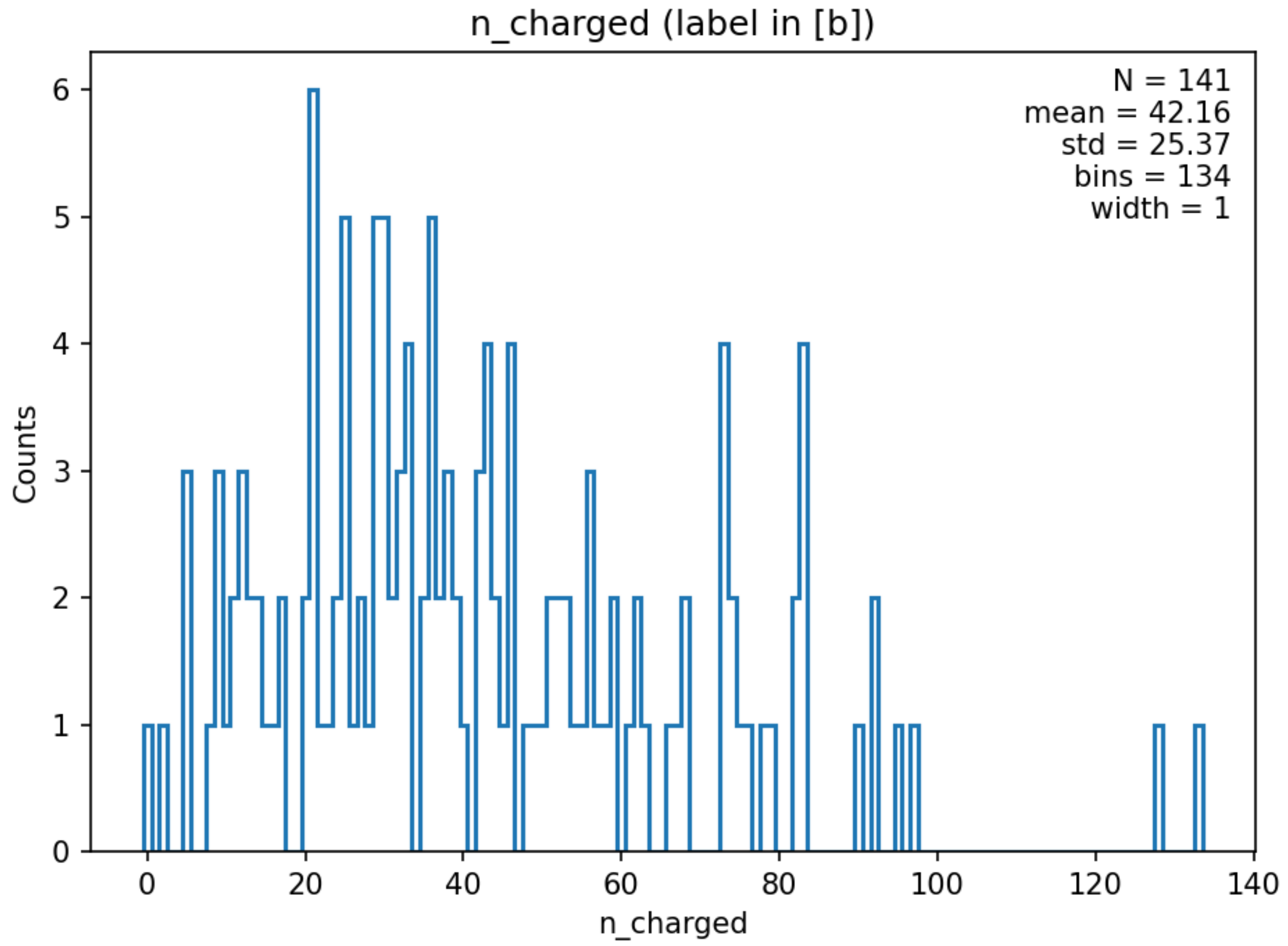
15. **IP significance median** (d0sig_med and z0sig_med)

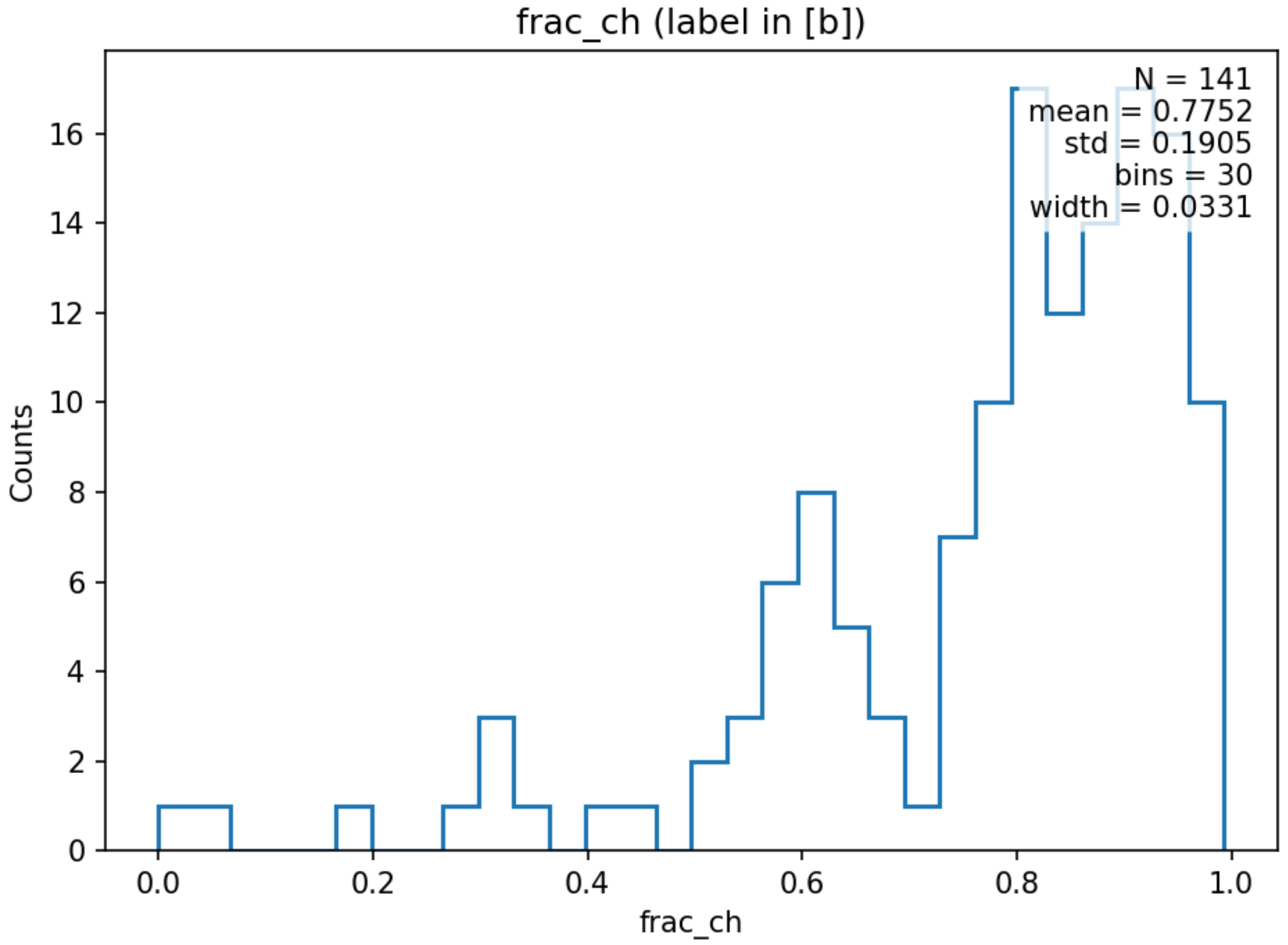
$$\mathrm{sig} = \frac{\mathrm{IP}}{\sigma(\mathrm{IP})}$$

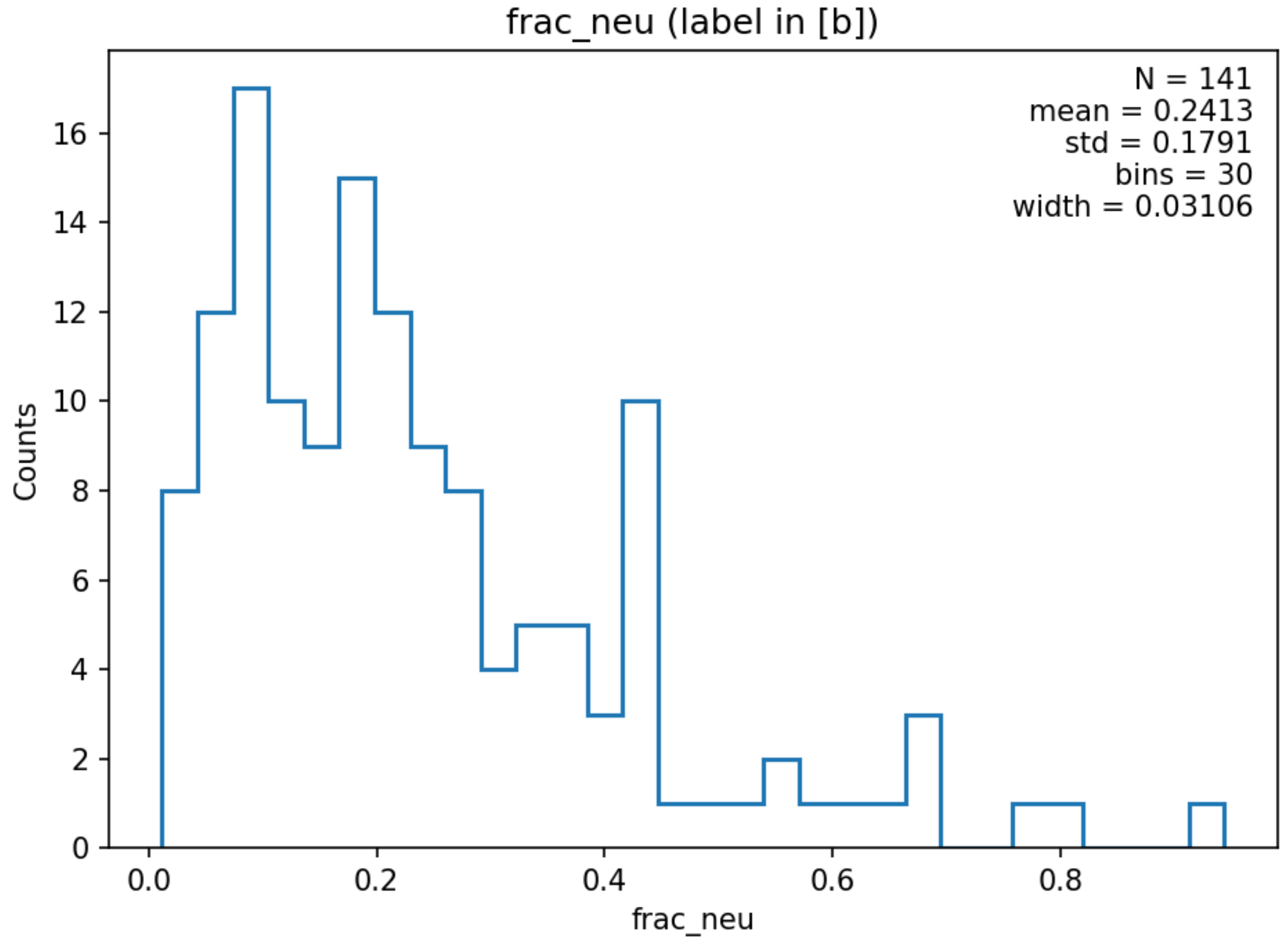
b-jet Feature Plots

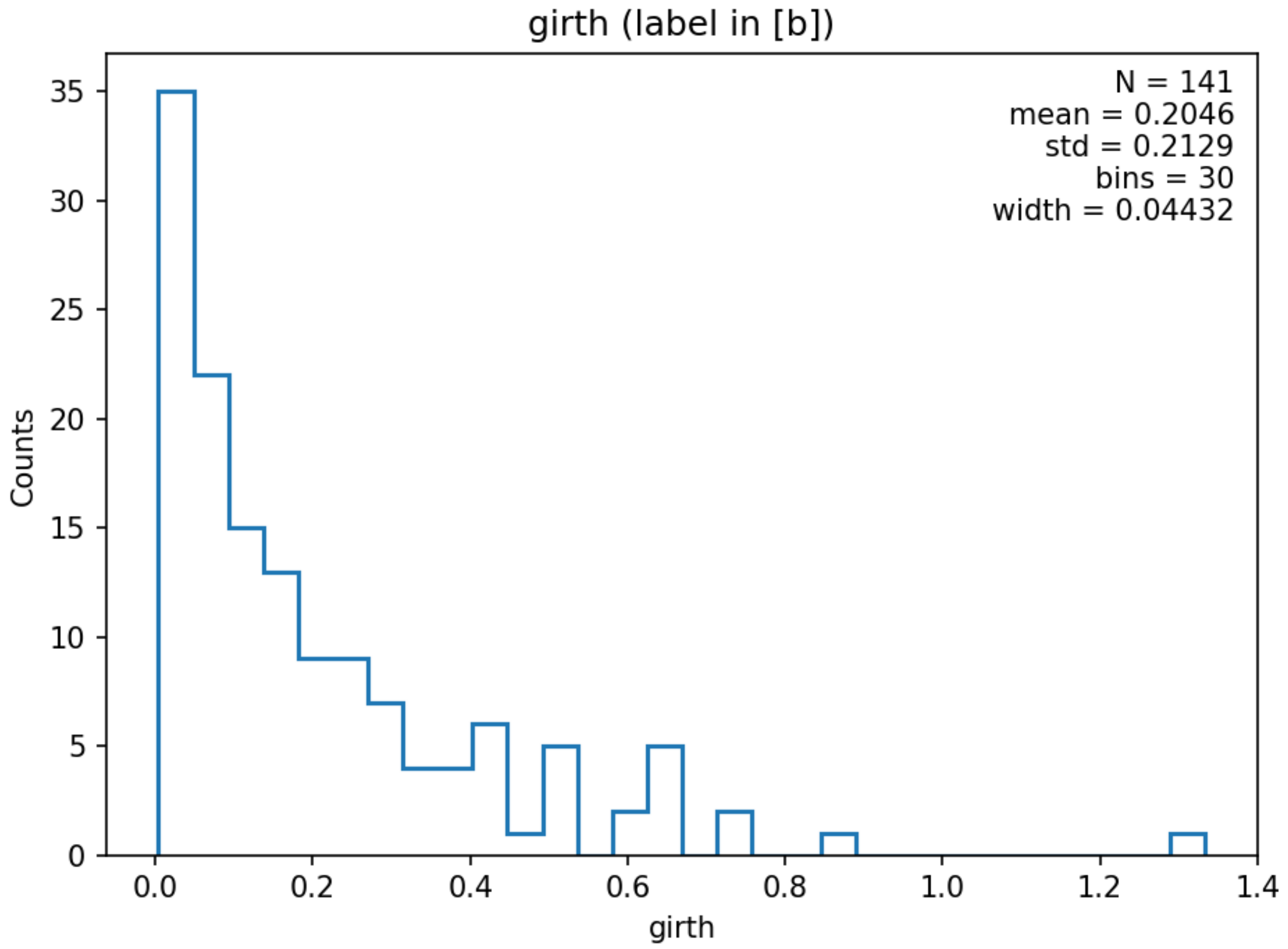




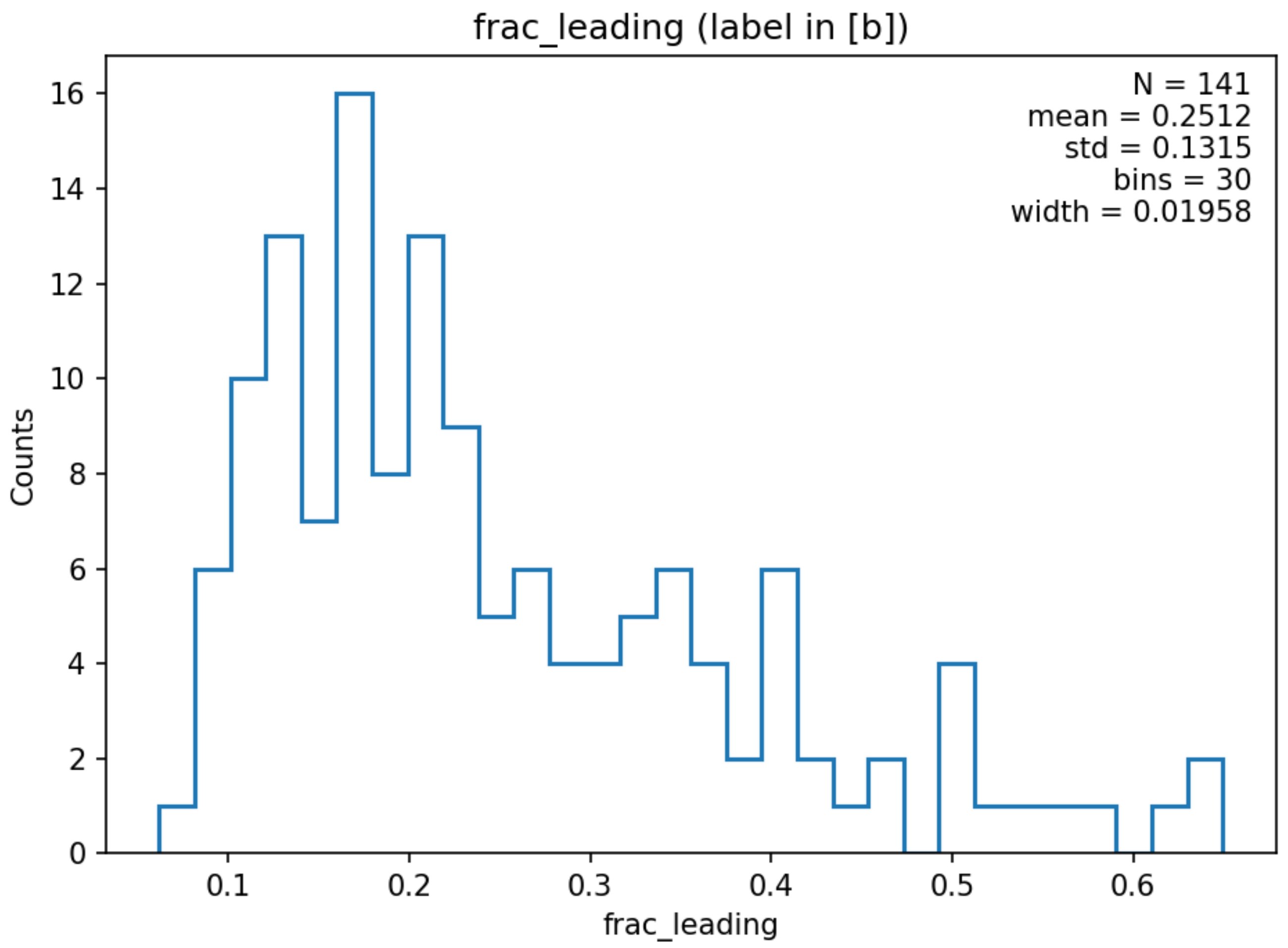


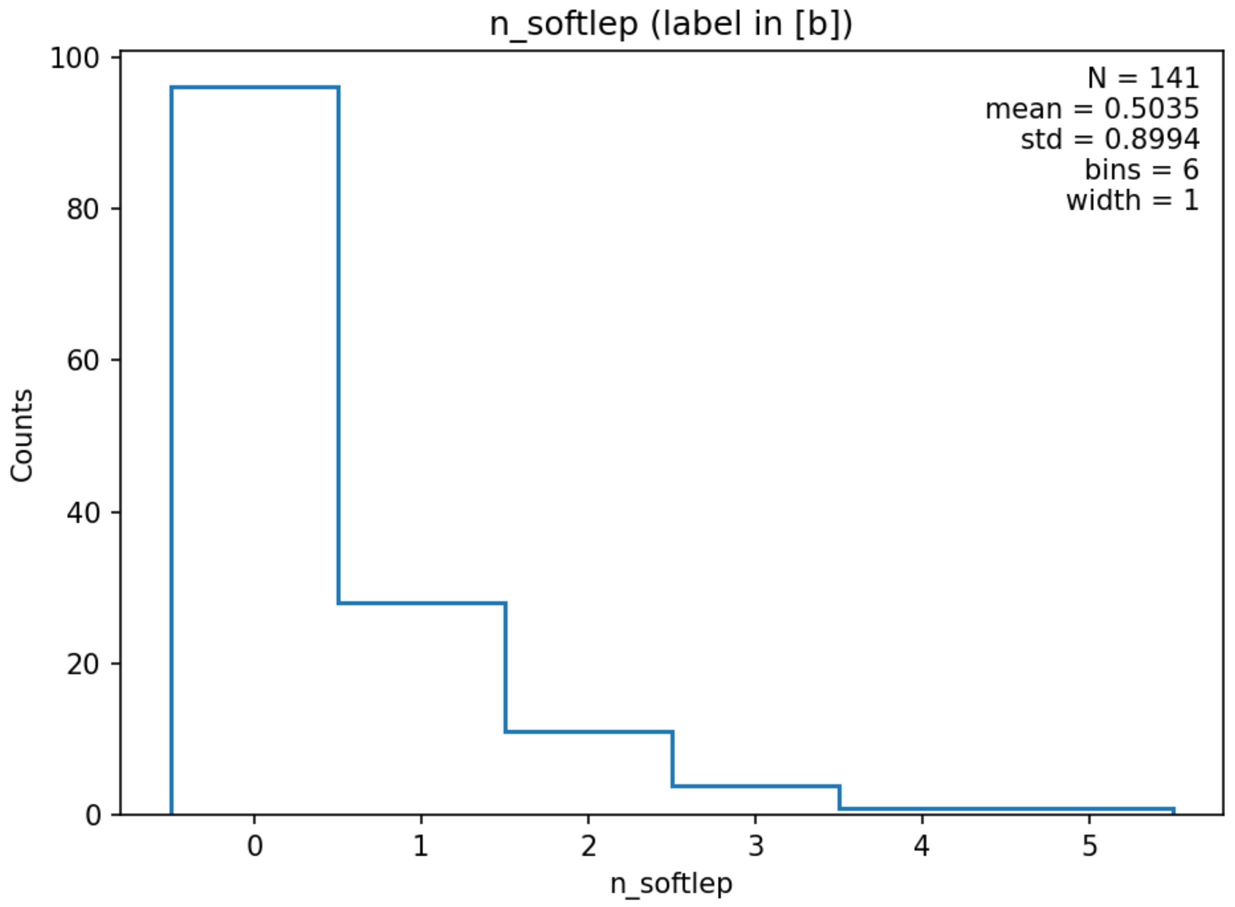


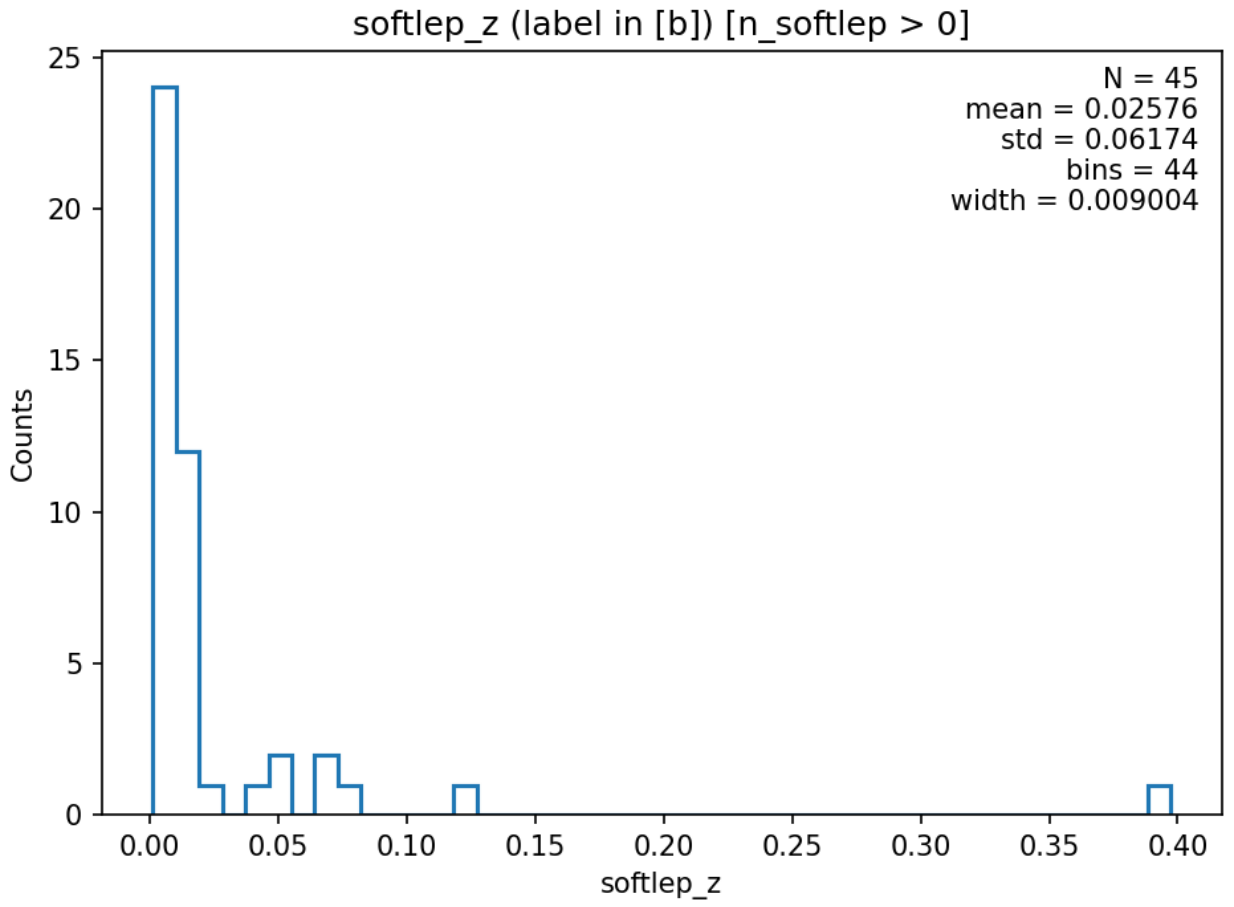


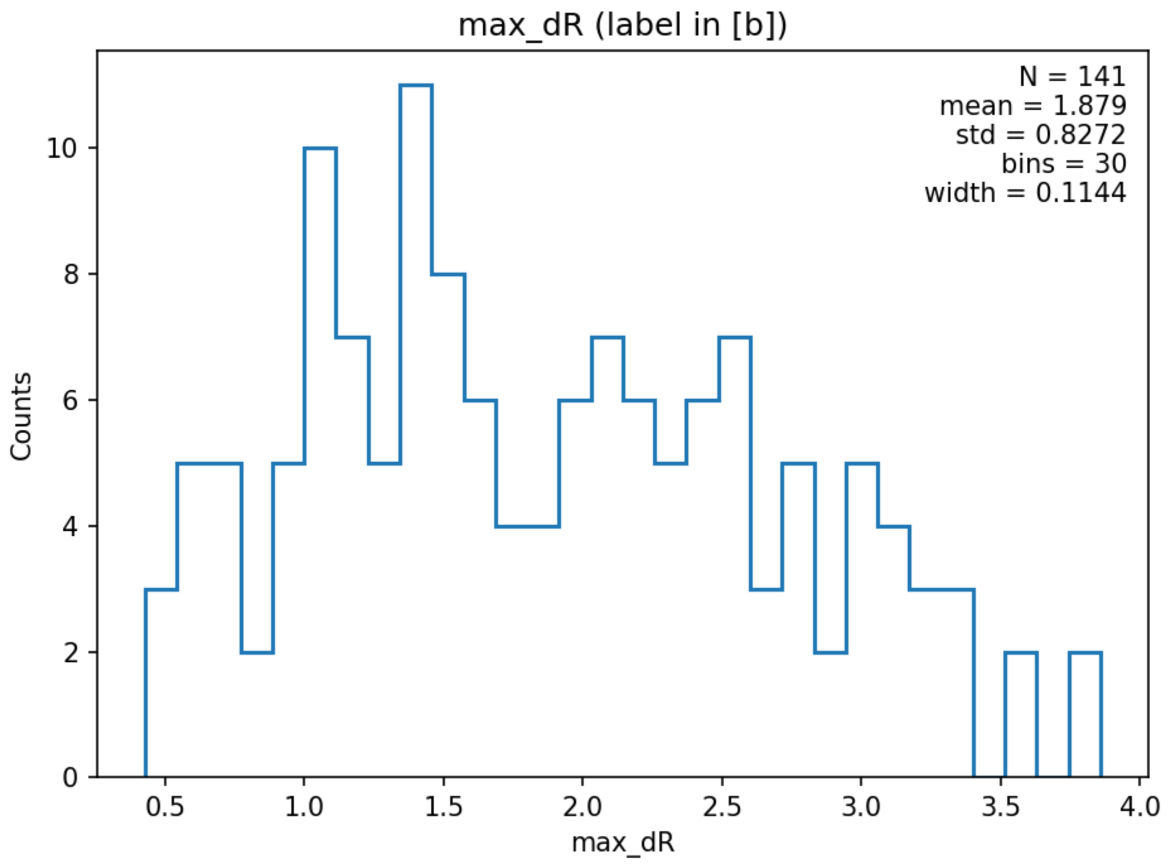


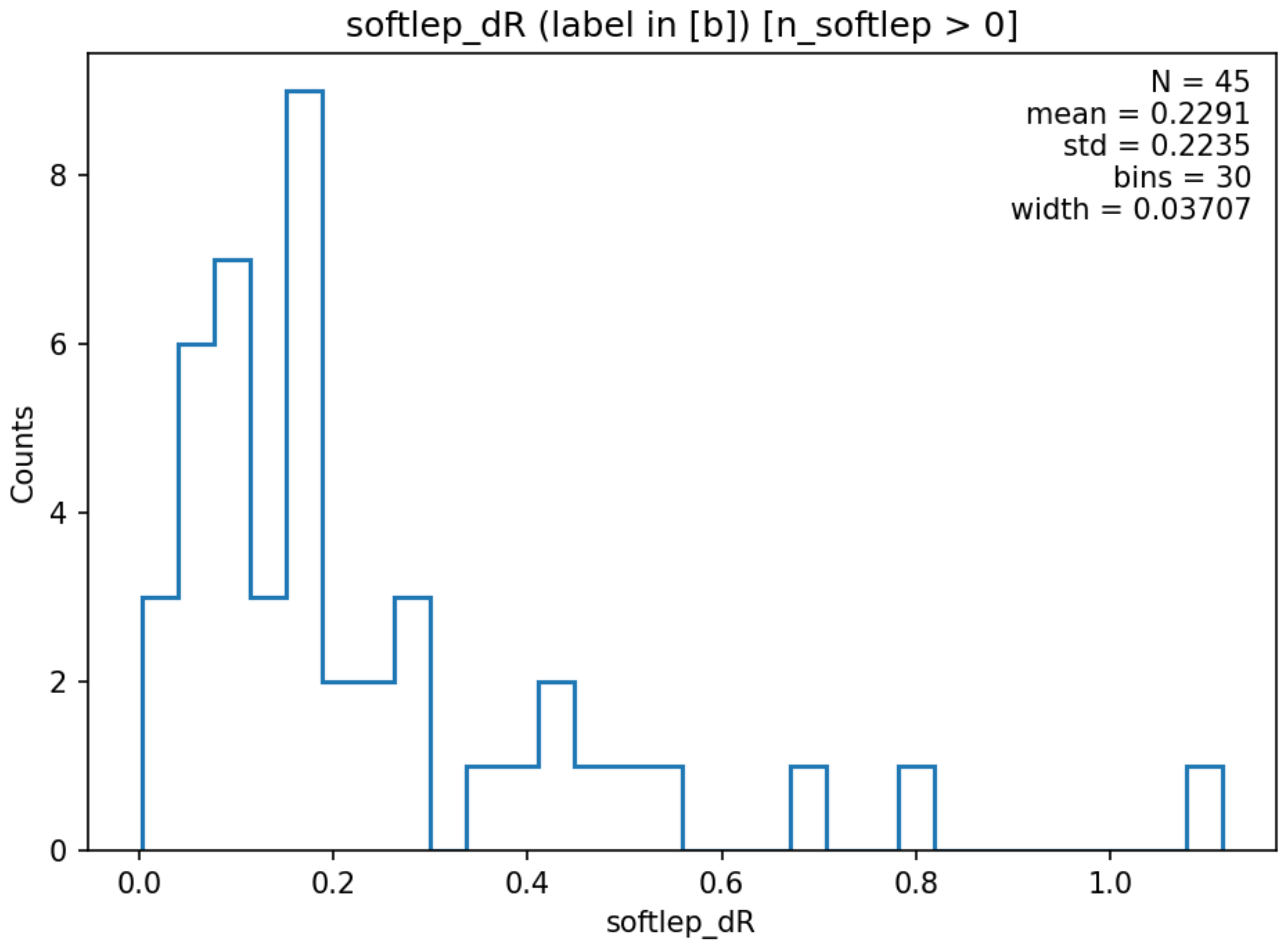
b-jet Feature Plots

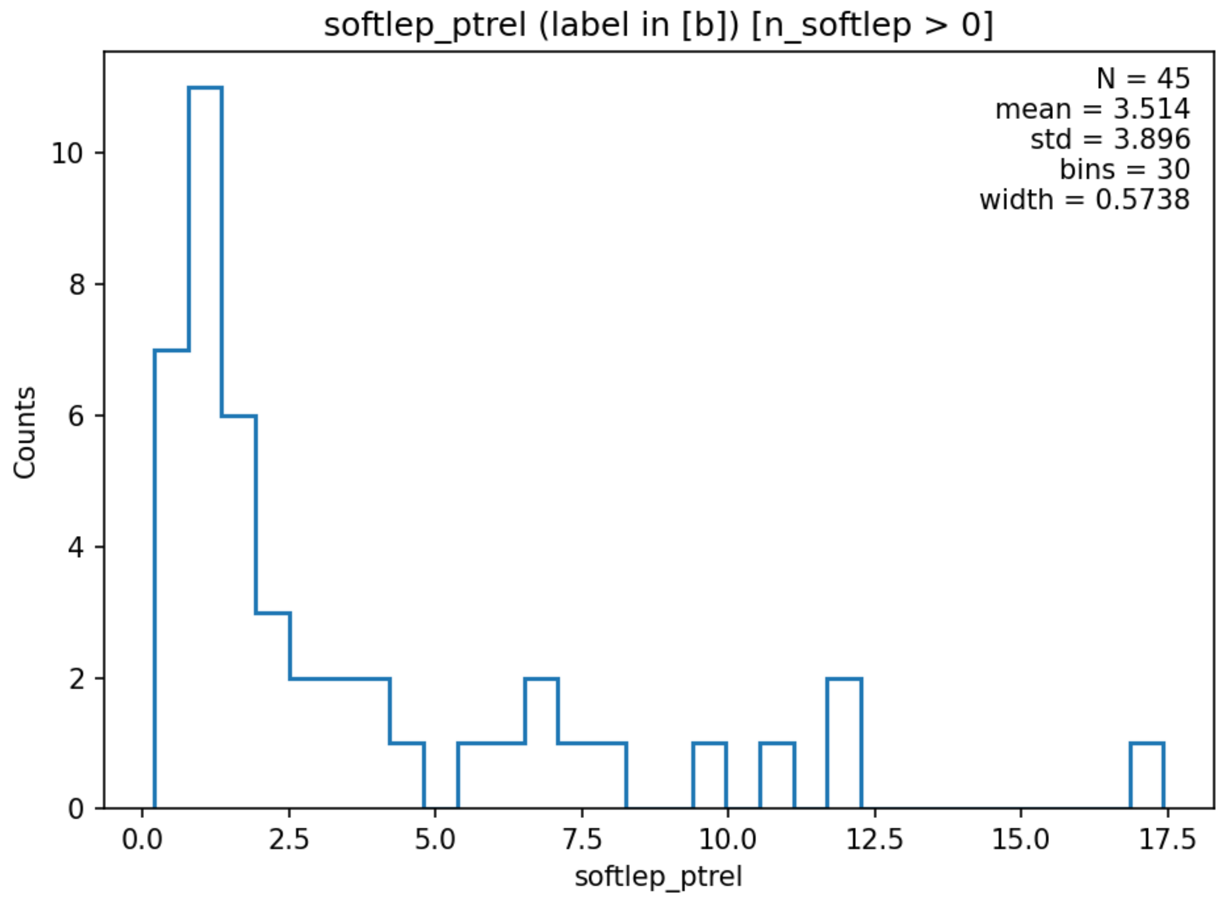

frac_leading (label in [b])

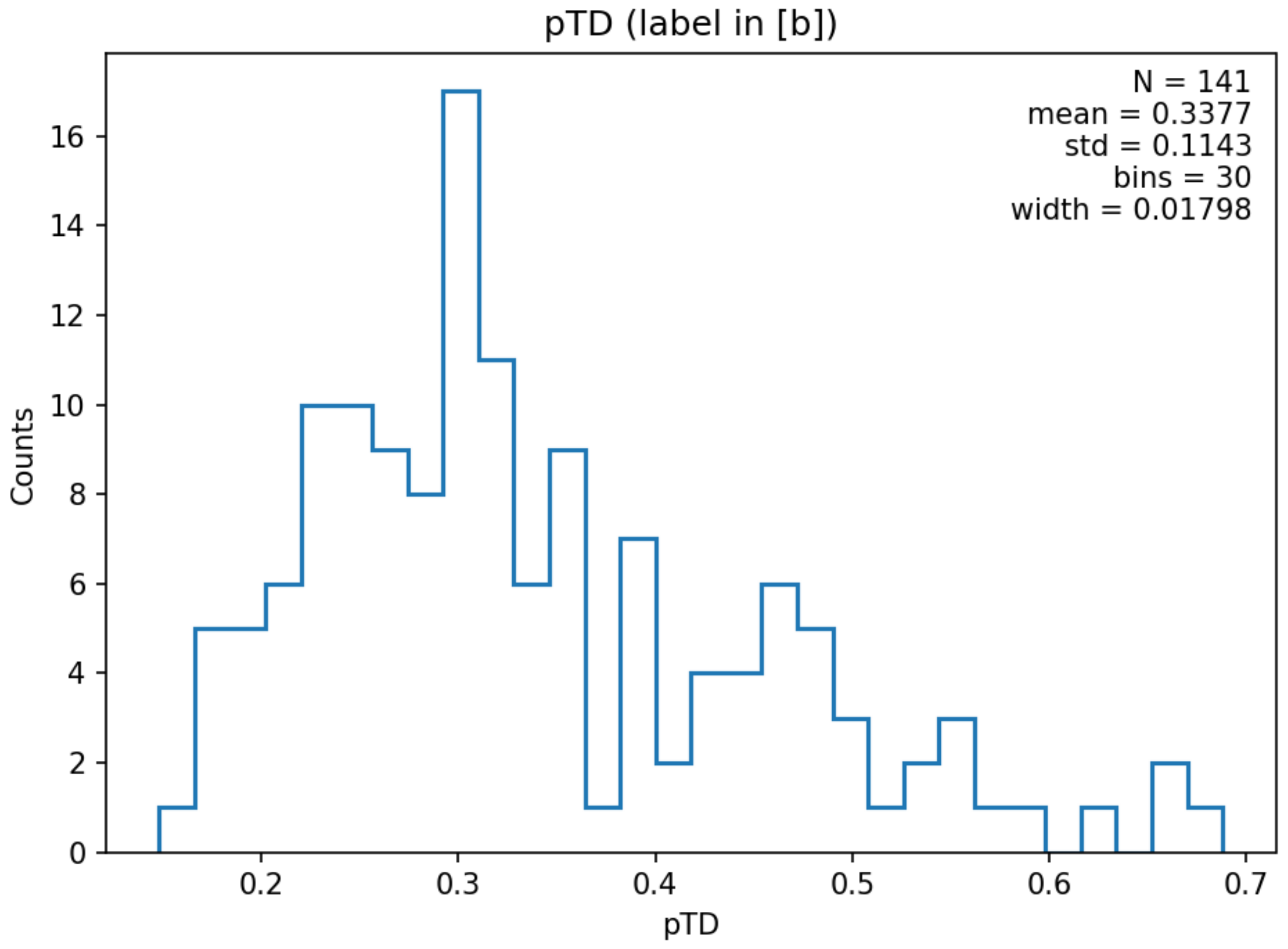

n_softlep (label in [b])


softlep_z (label in [b]) [n_softlep > 0]

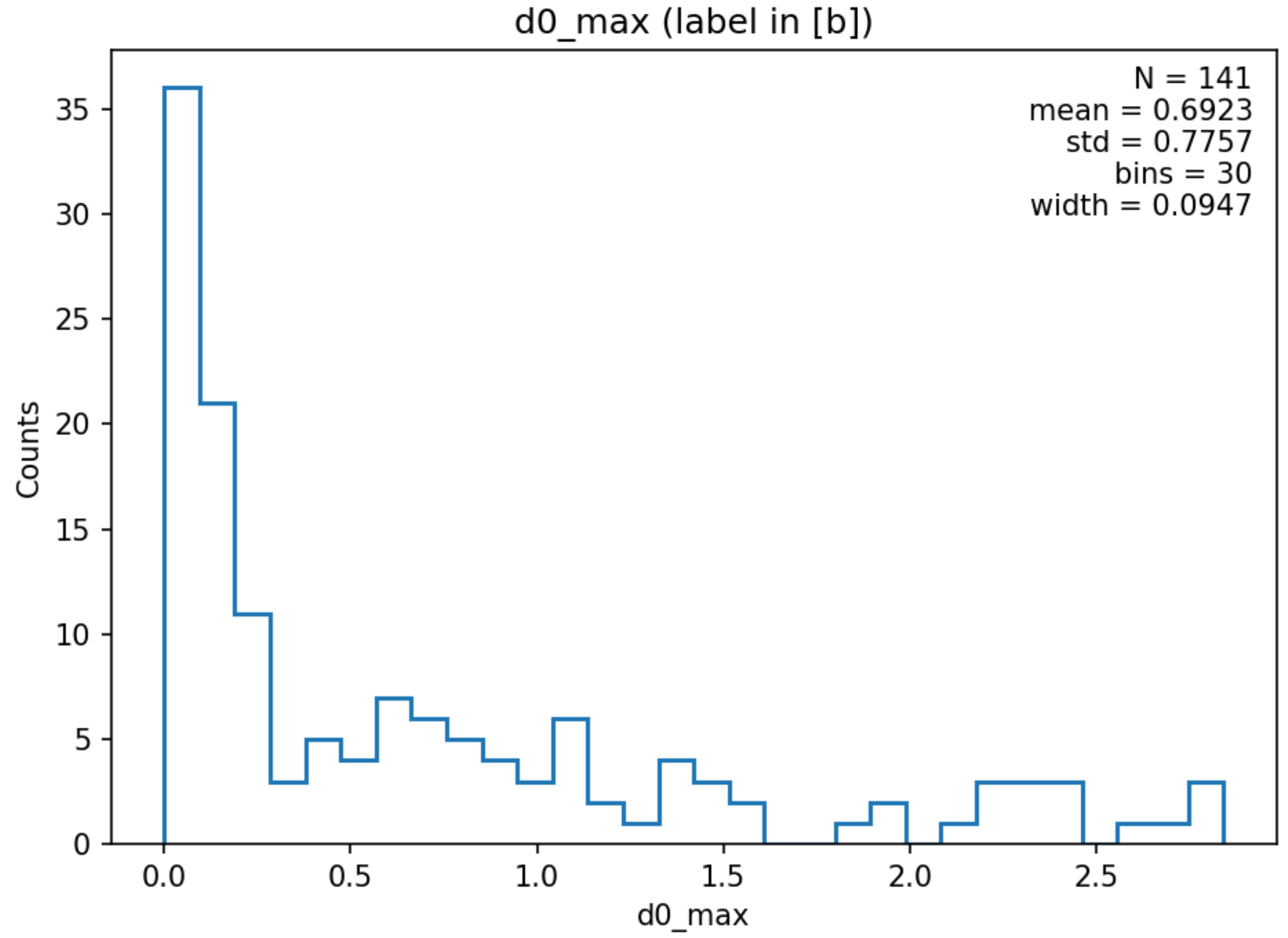

max_dR (label in [b])

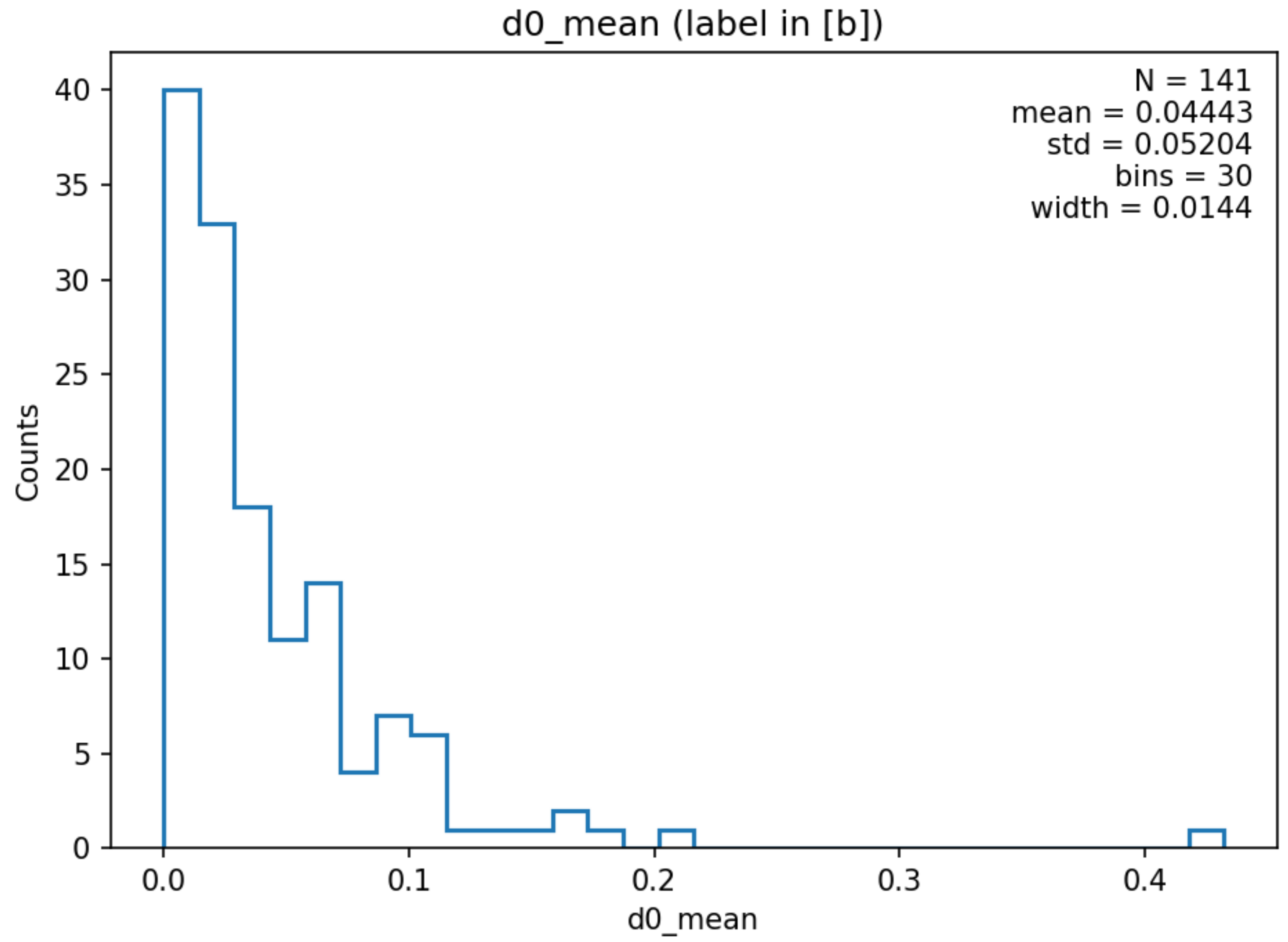

softlep_dR (label in [b]) [n_softlep > 0]

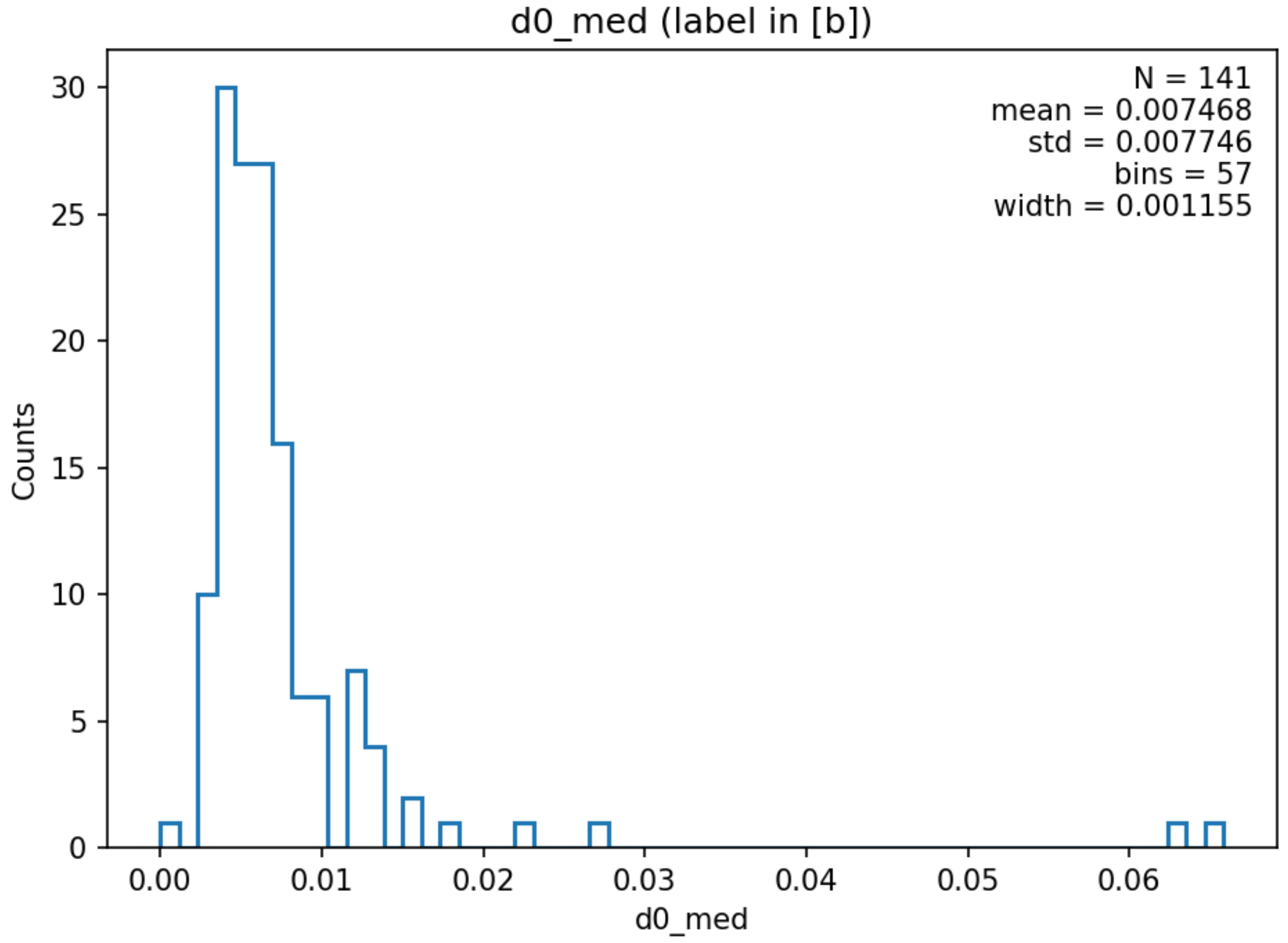

softlep_ptrel (label in [b]) [n_softlep > 0]

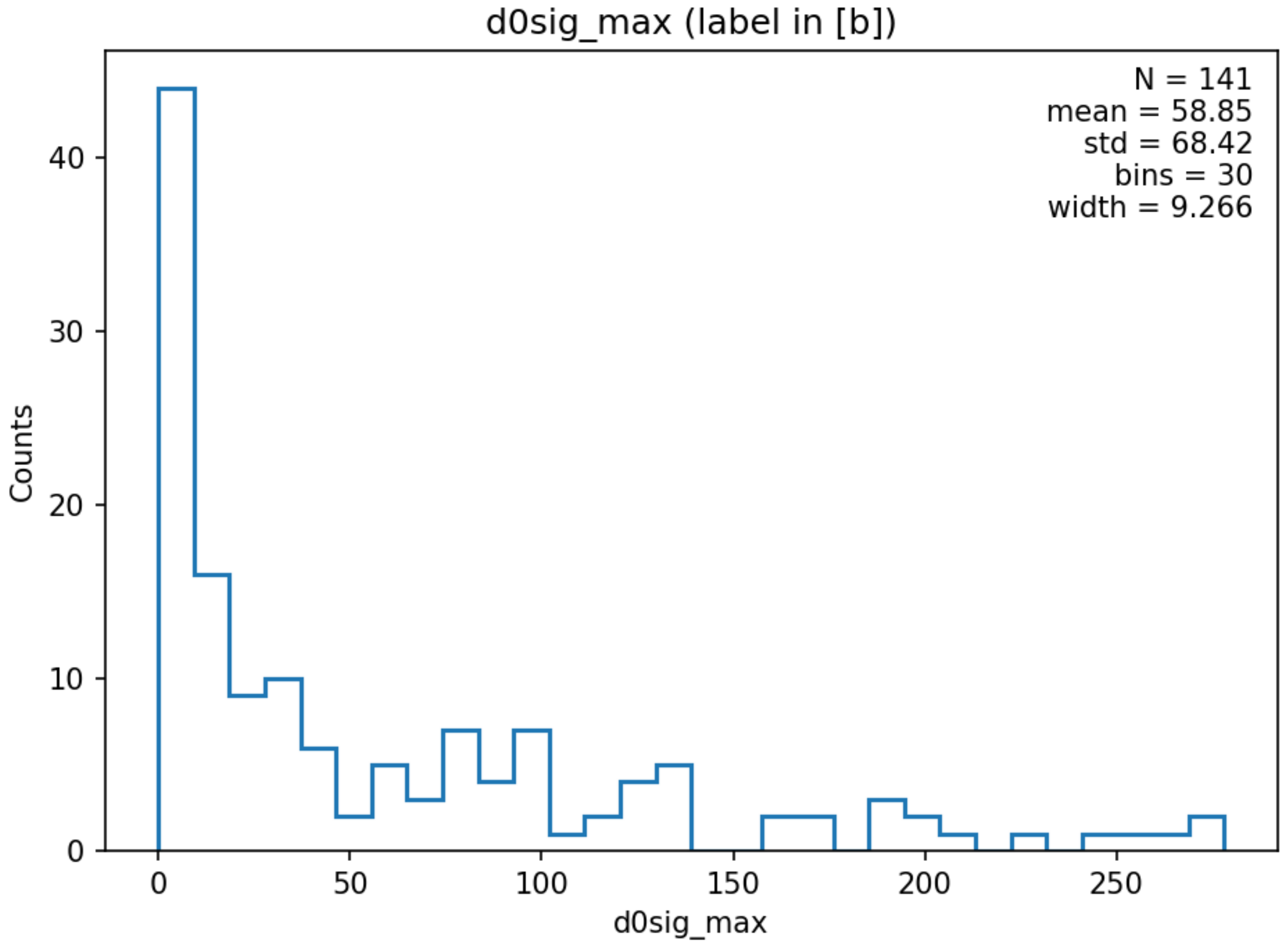

pTD (label in [b])

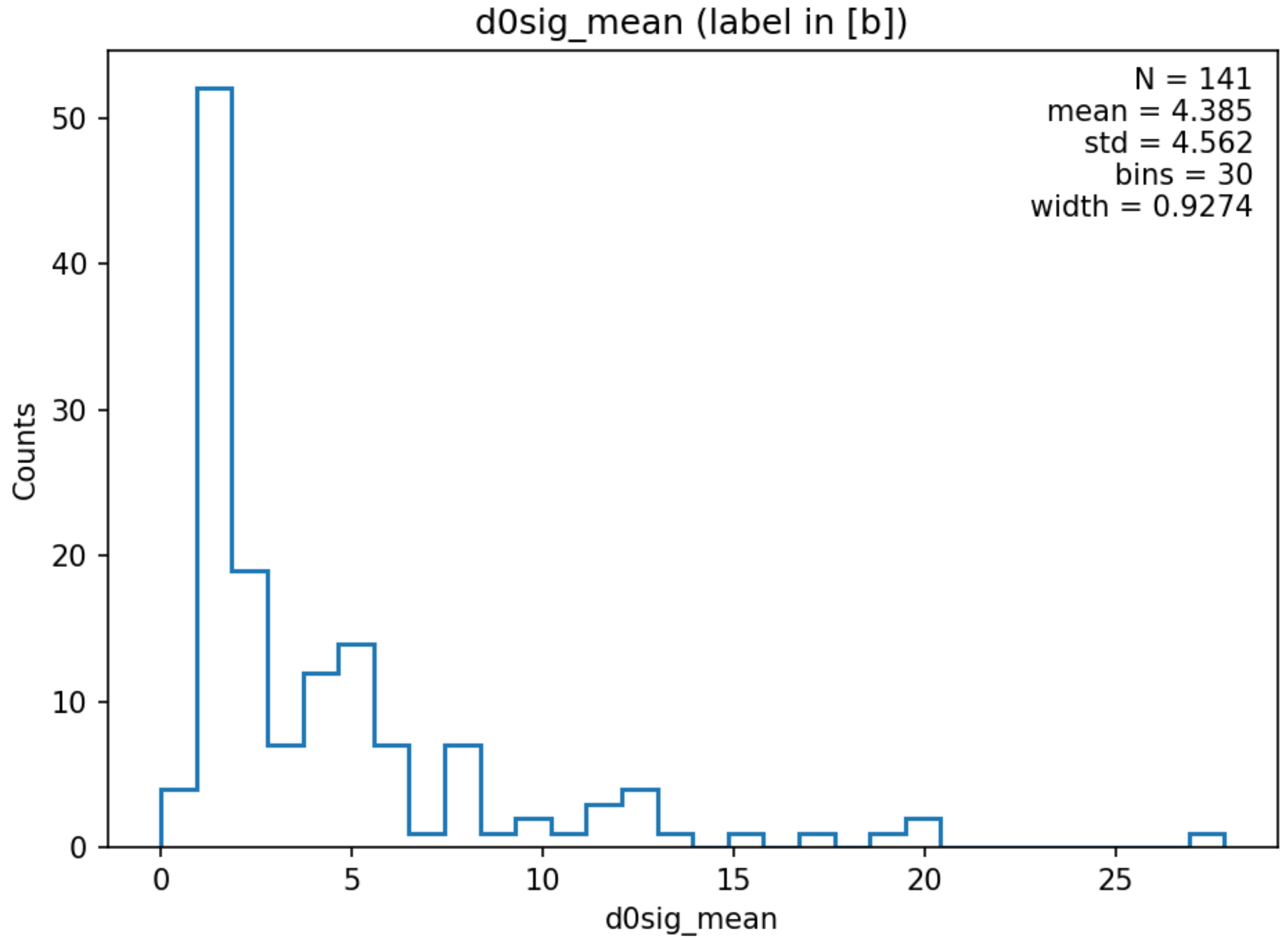
b-jet Feature Plots

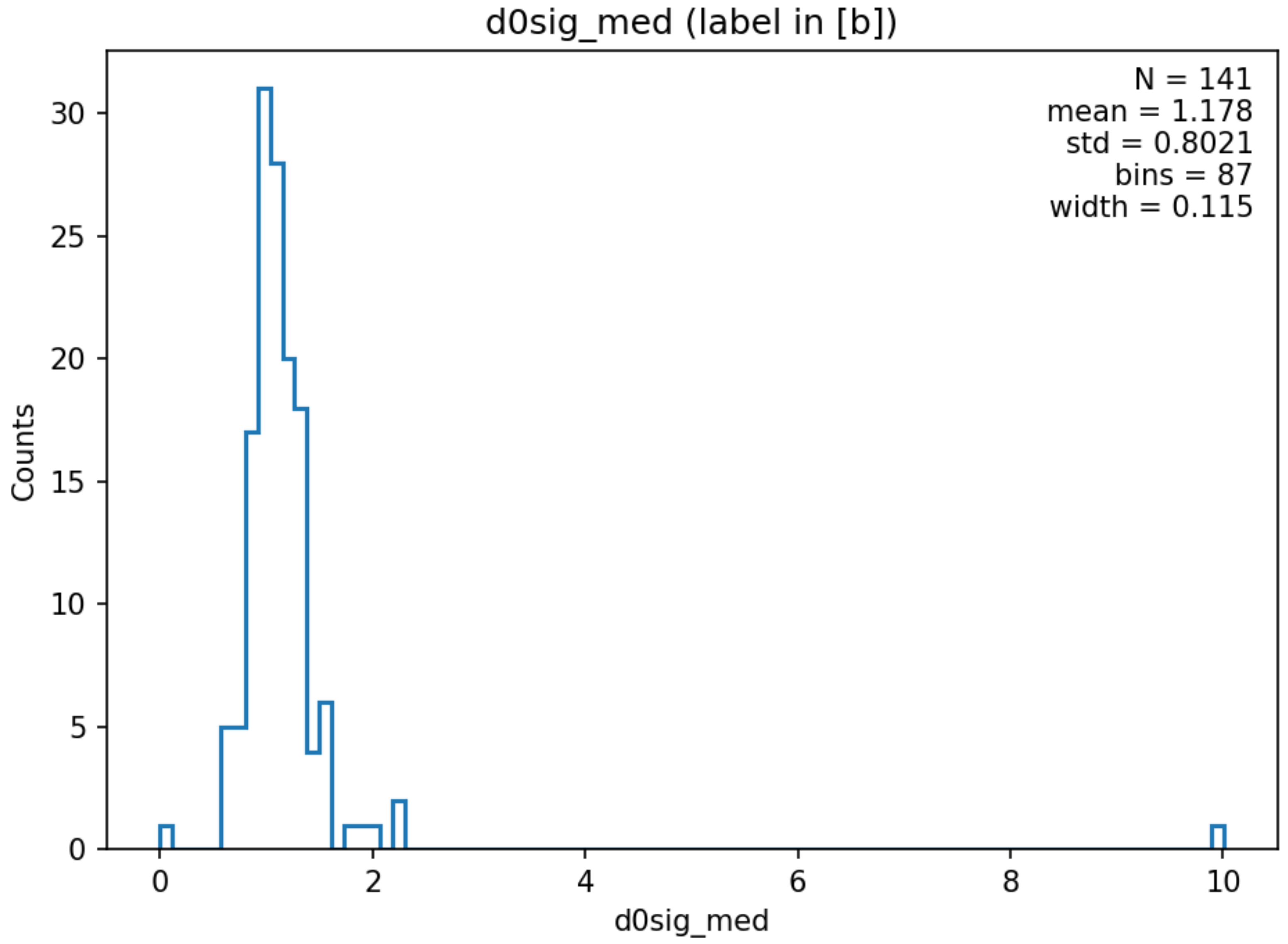

d0_max (label in [b])


d0_mean (label in [b])

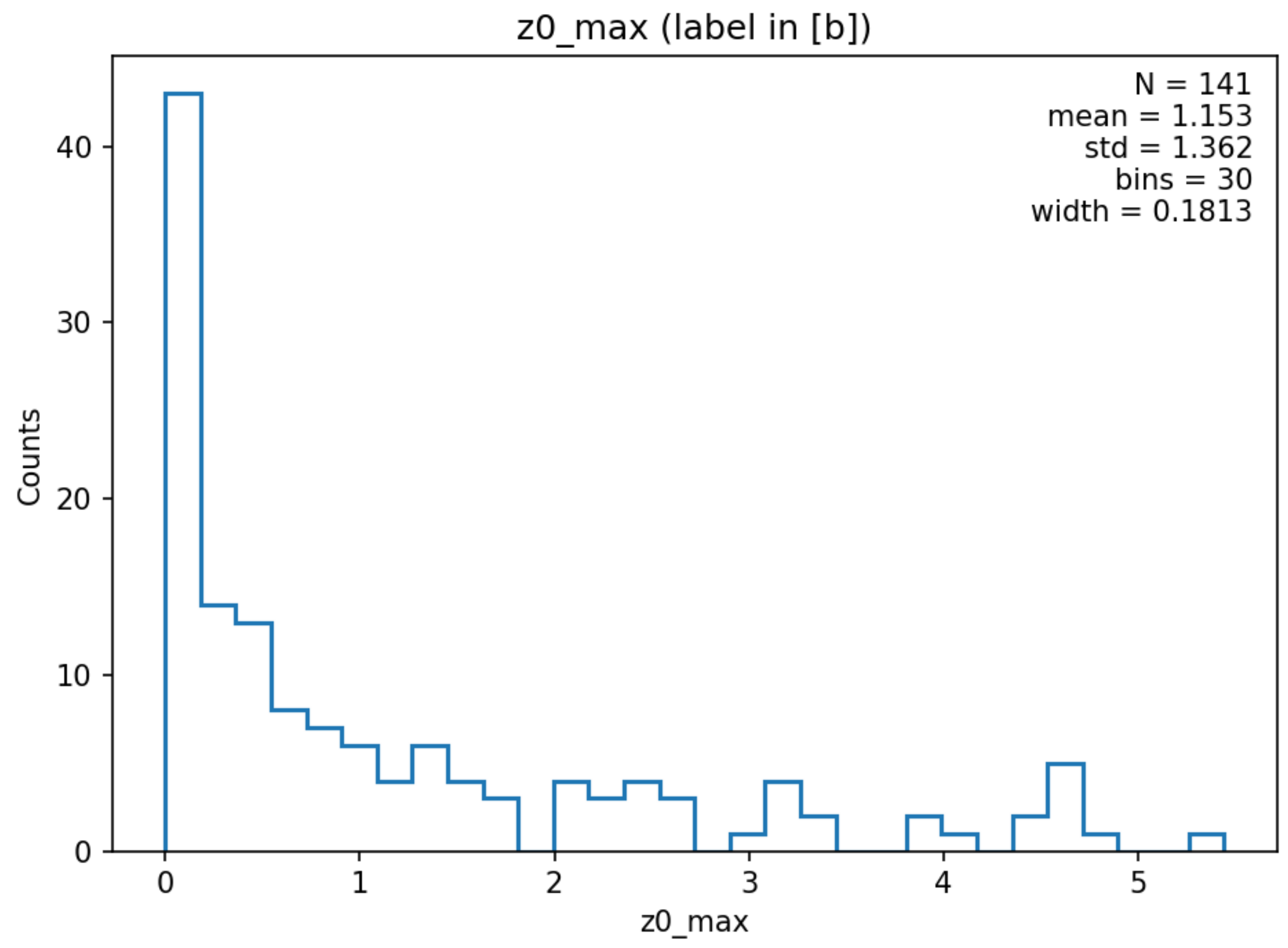

d0_med (label in [b])

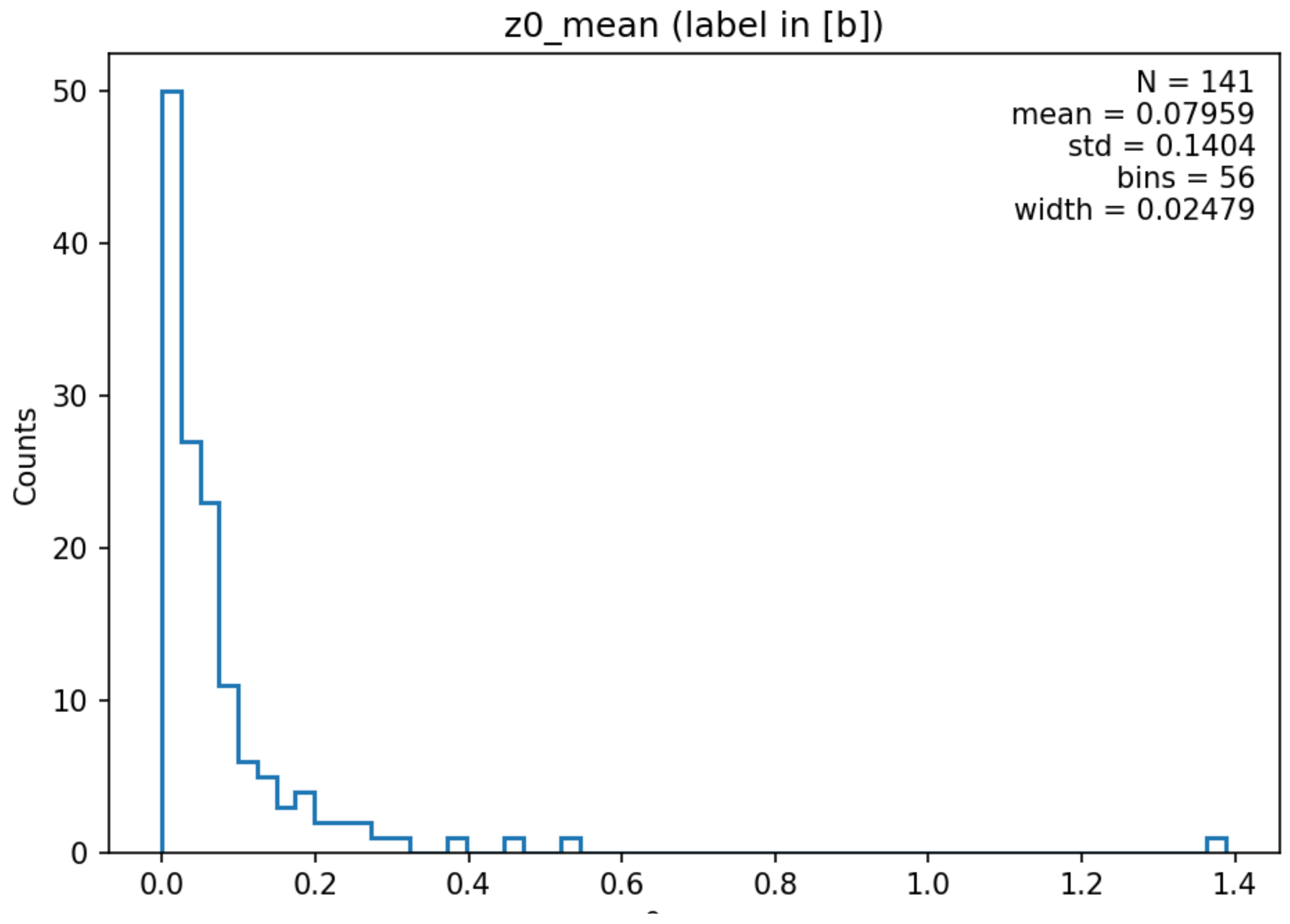

d0sig_max (label in [b])

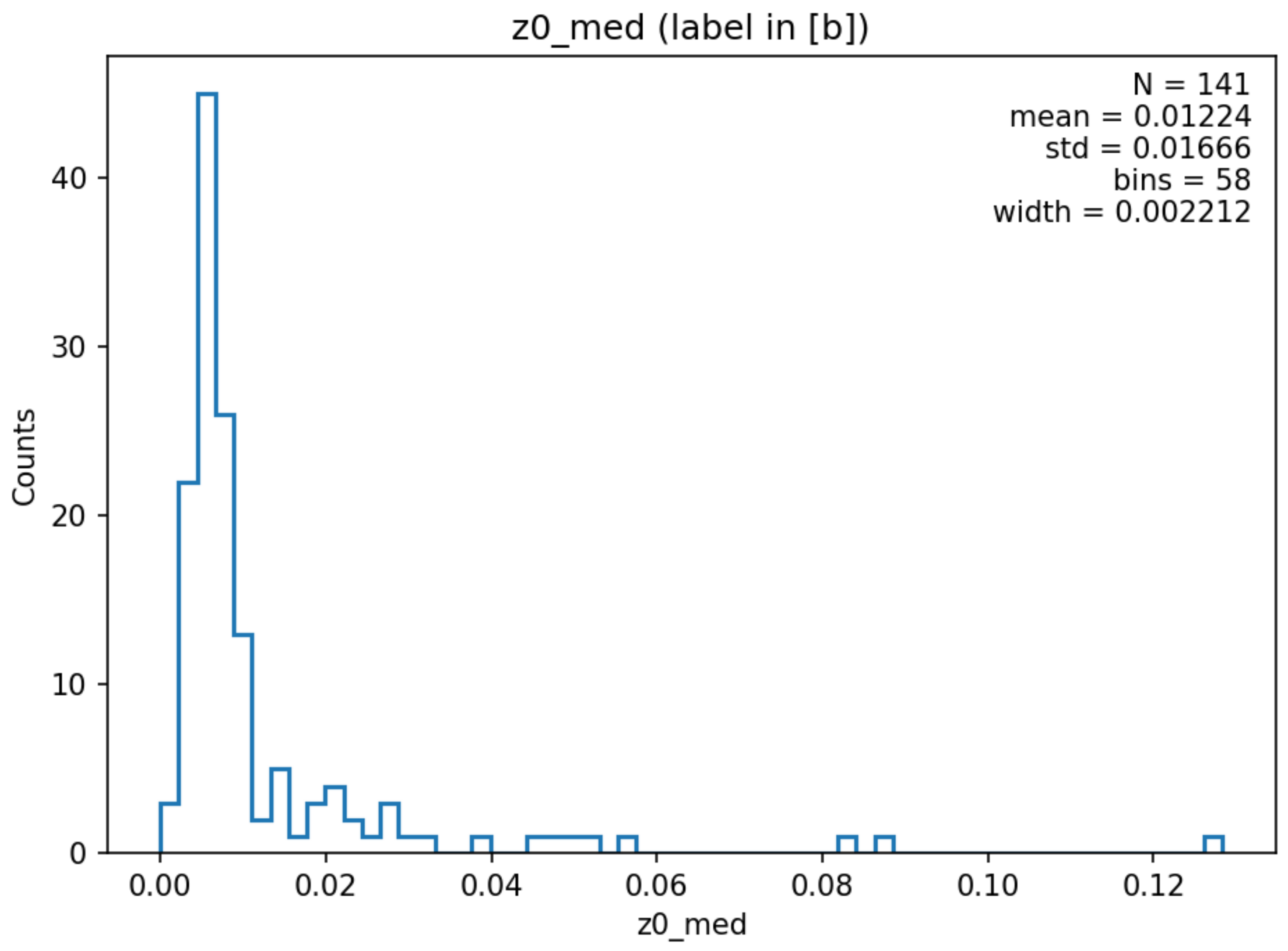

d0sig_mean (label in [b])

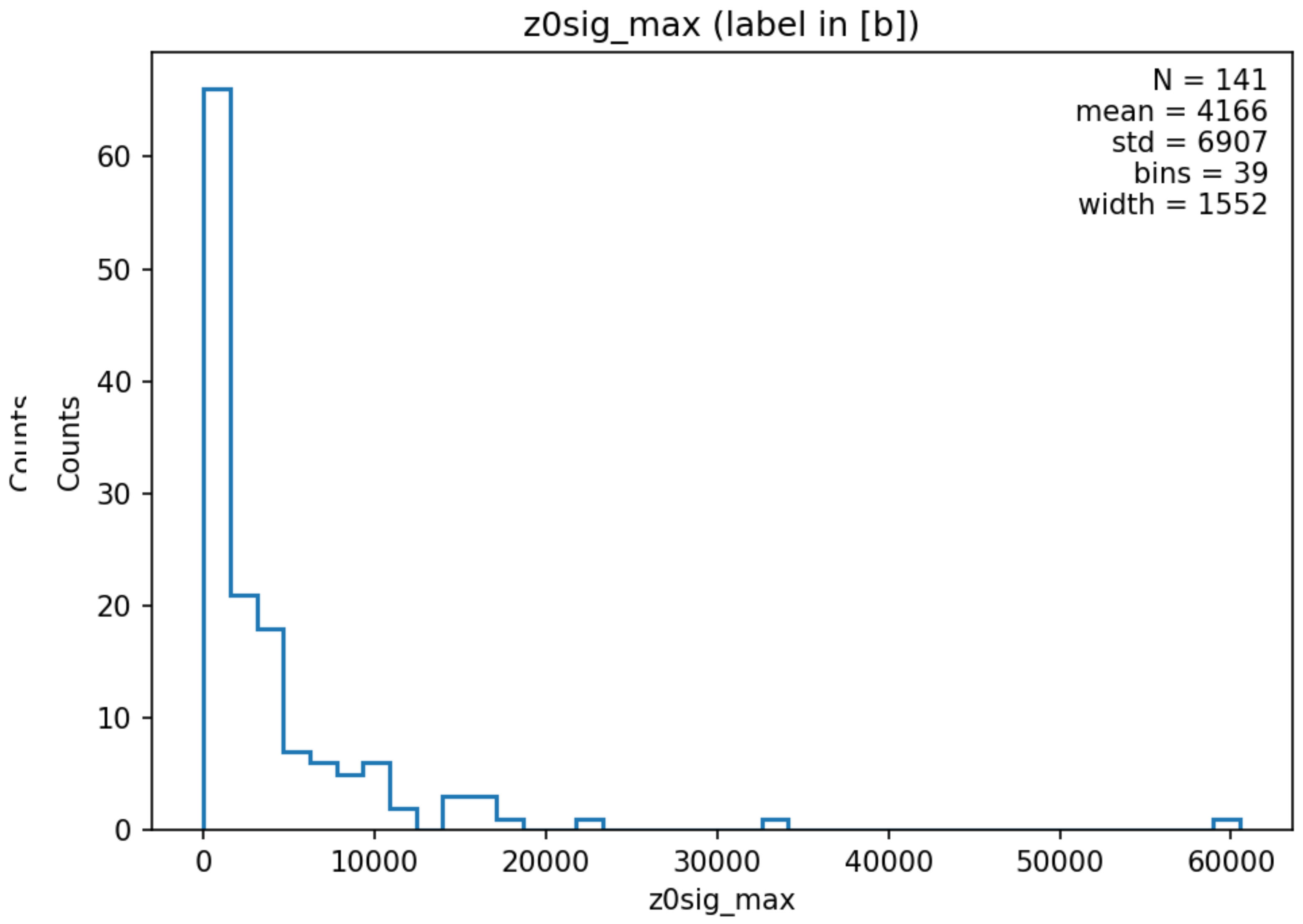

d0sig_med (label in [b])

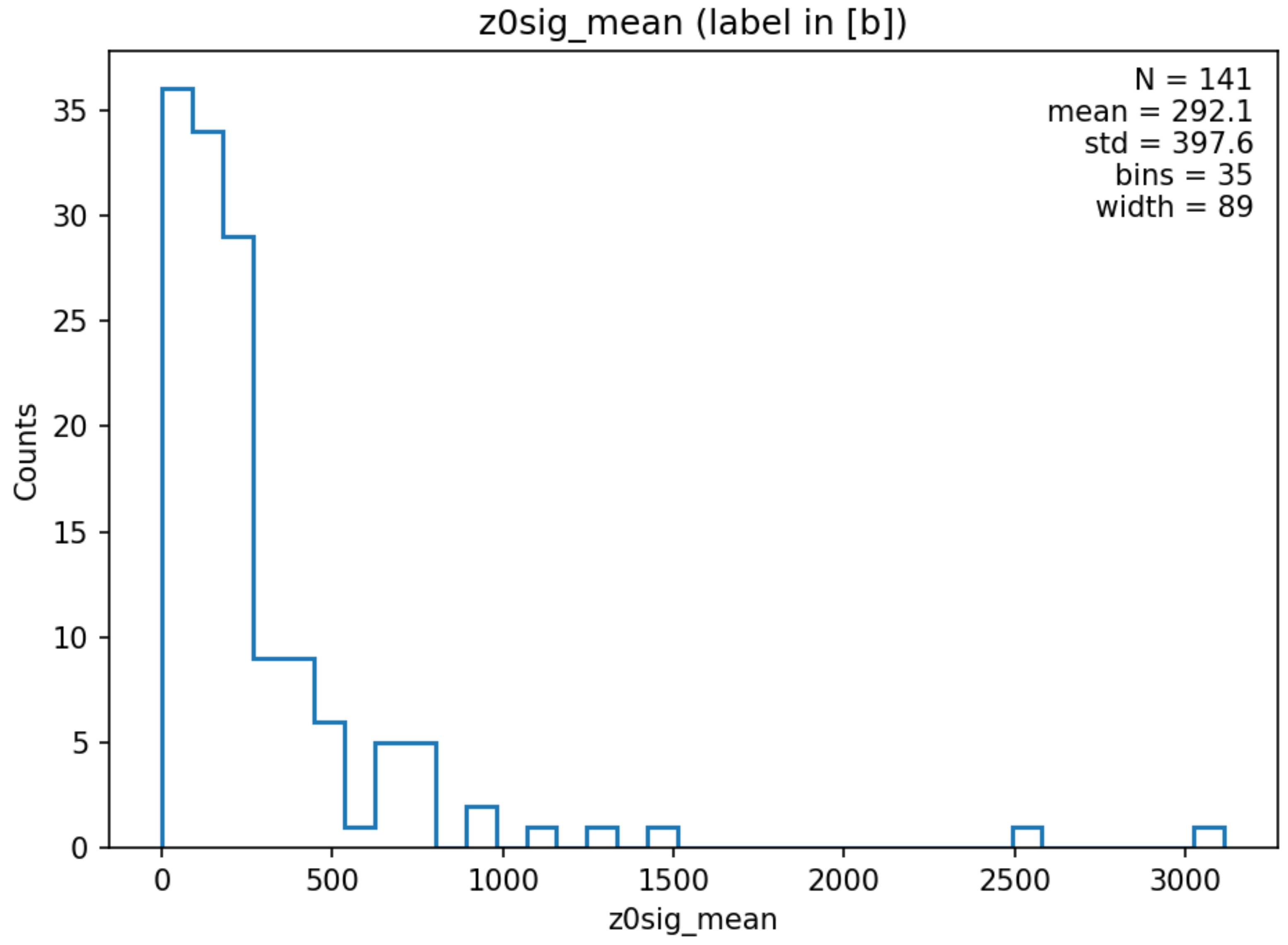
b-jet Feature Plots

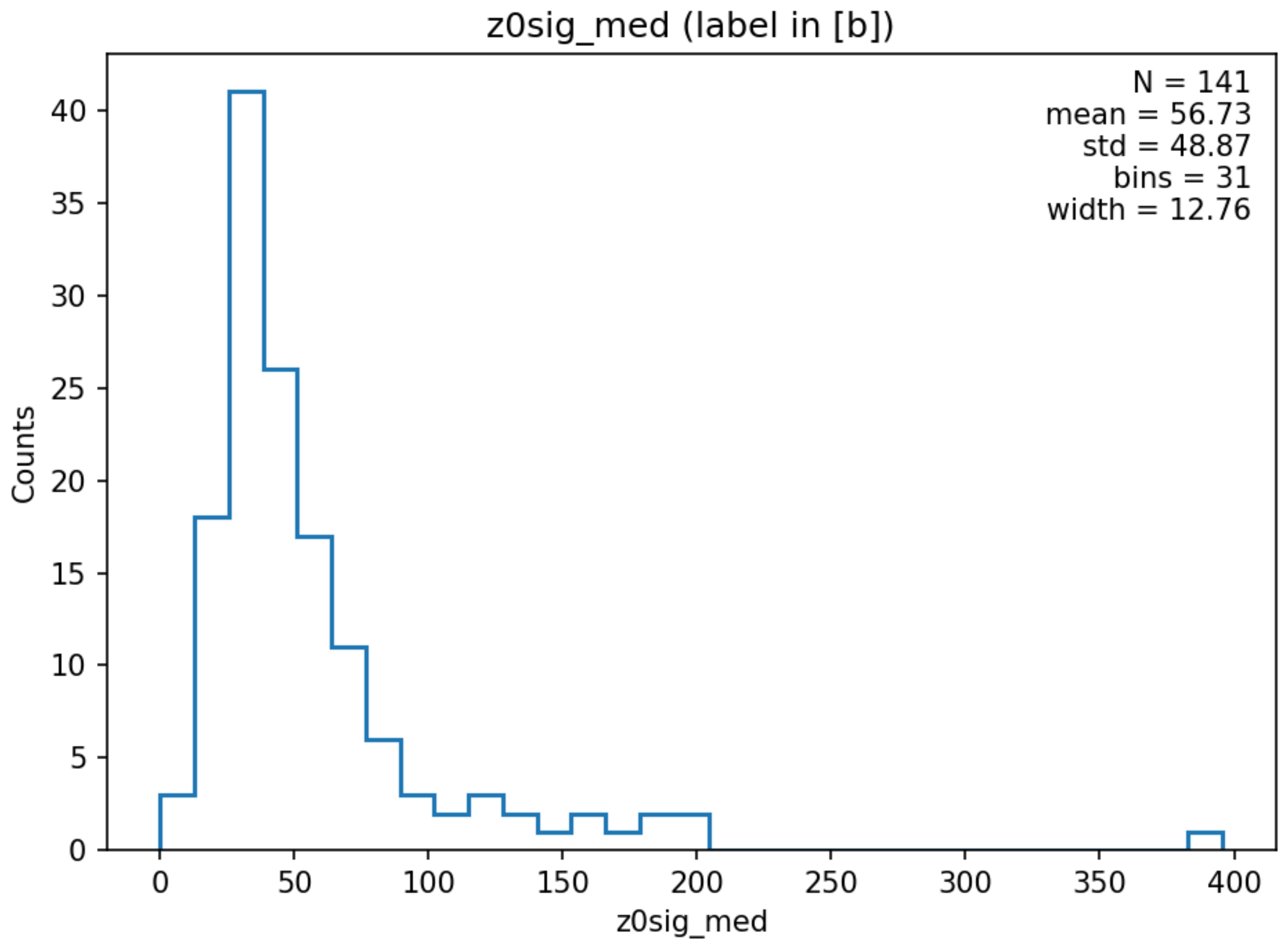




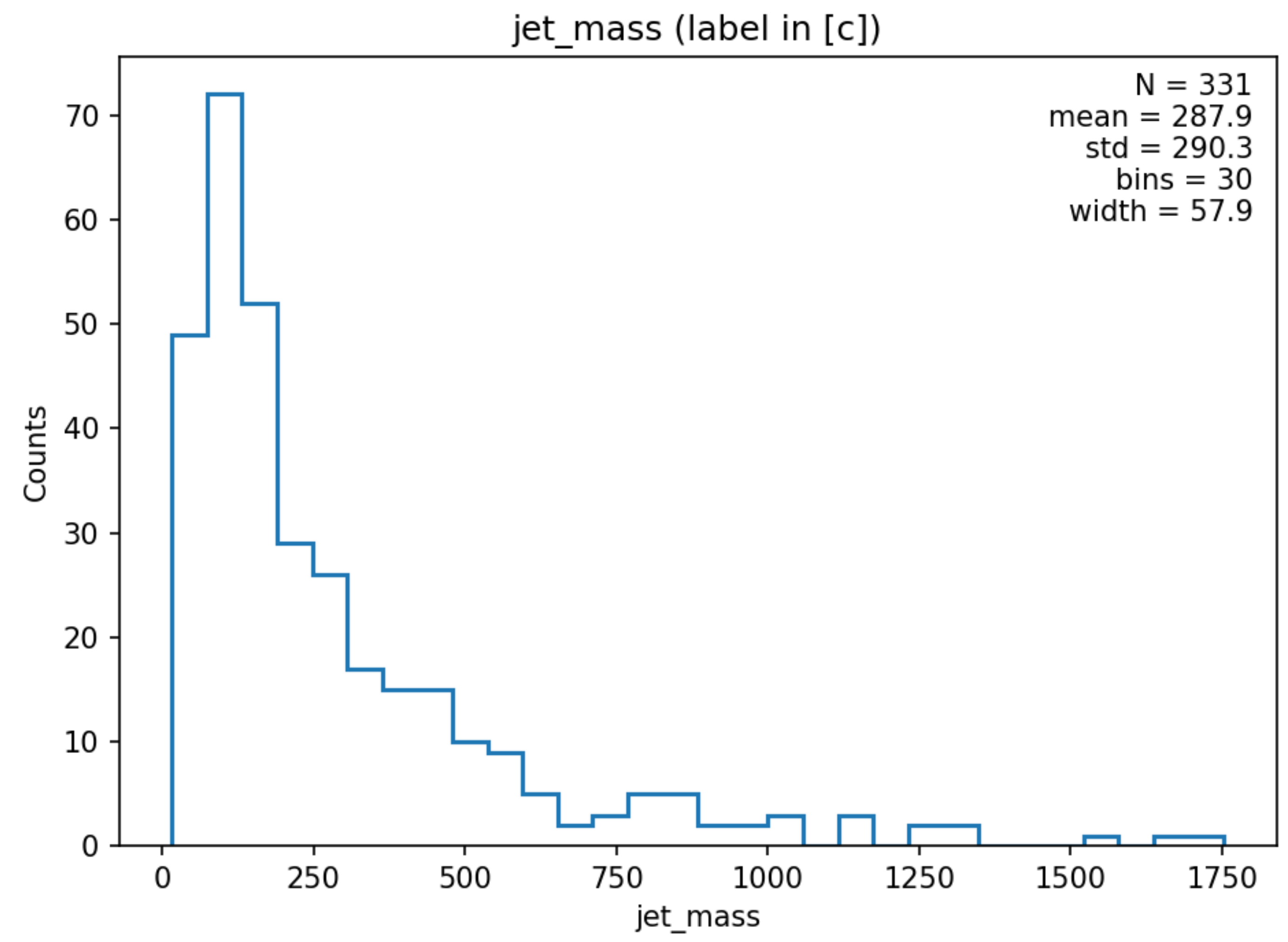


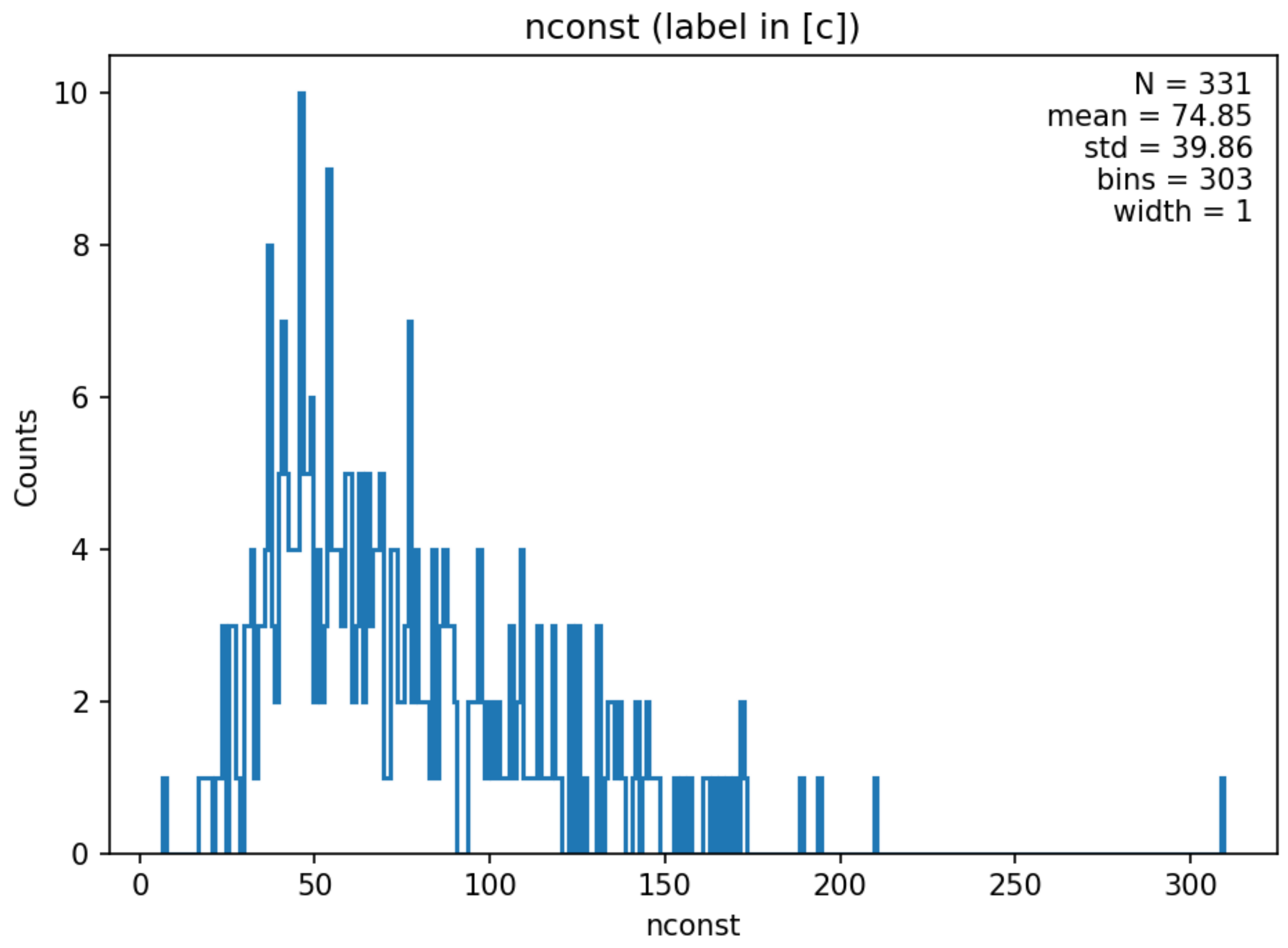


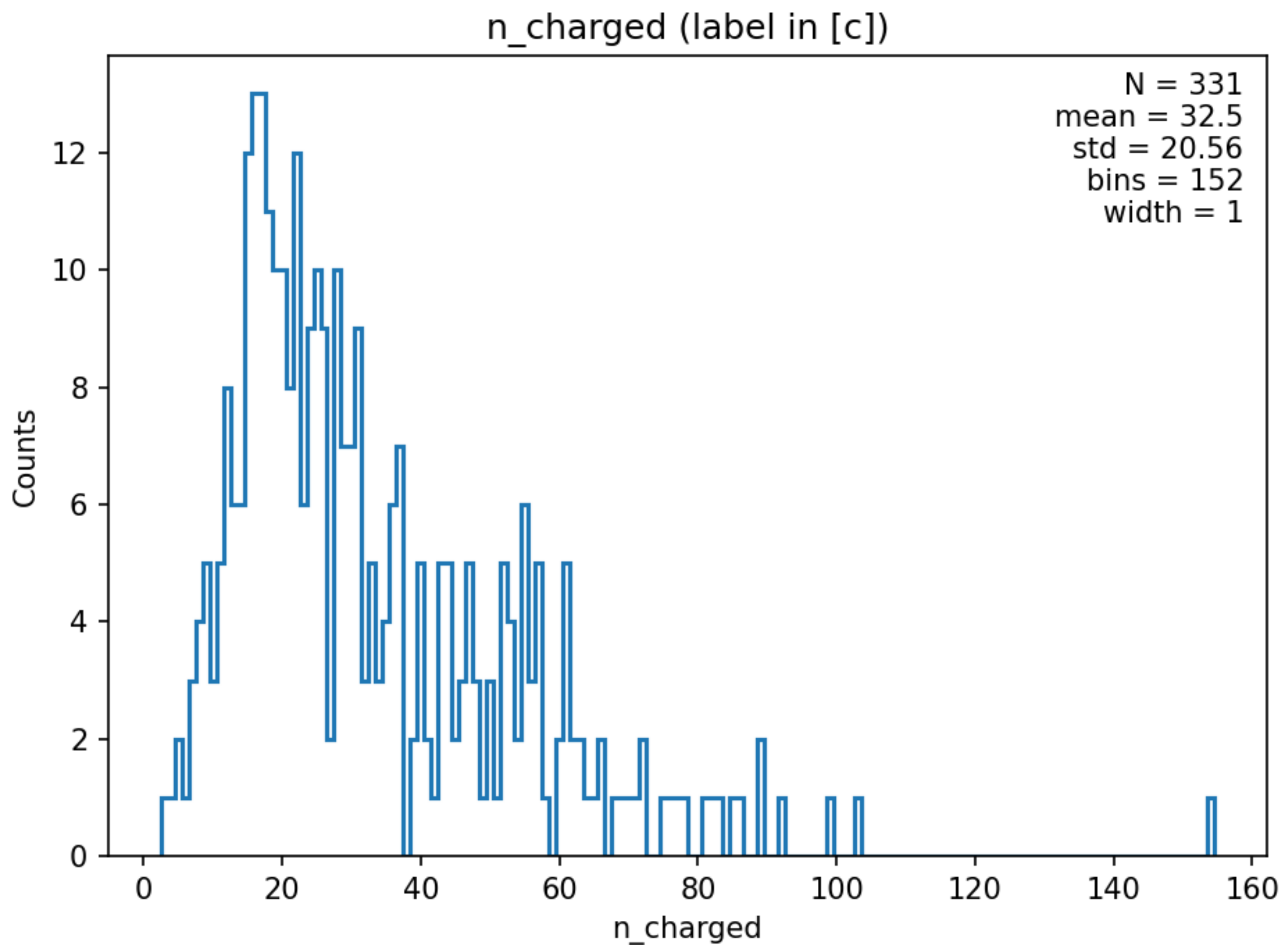


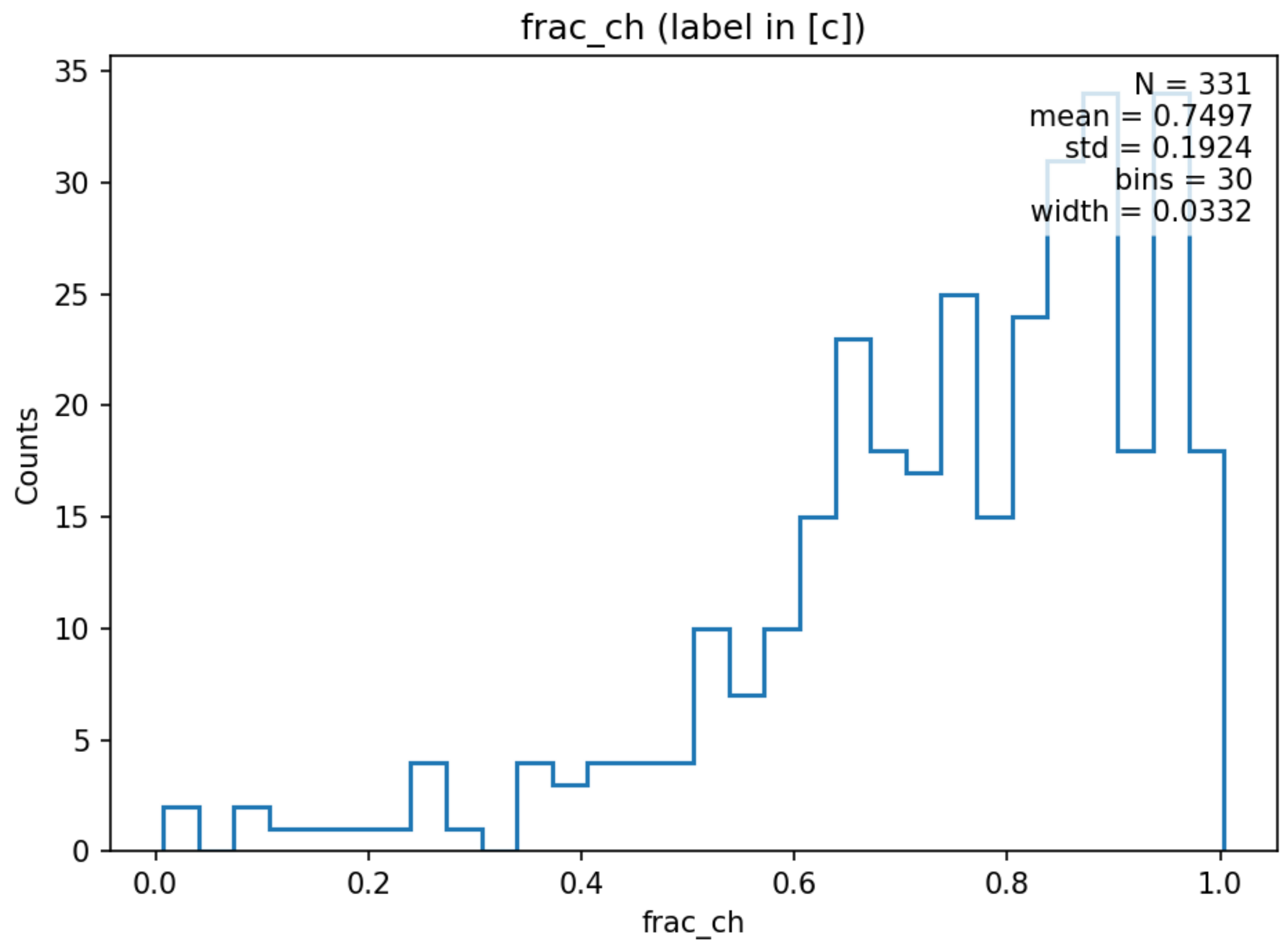


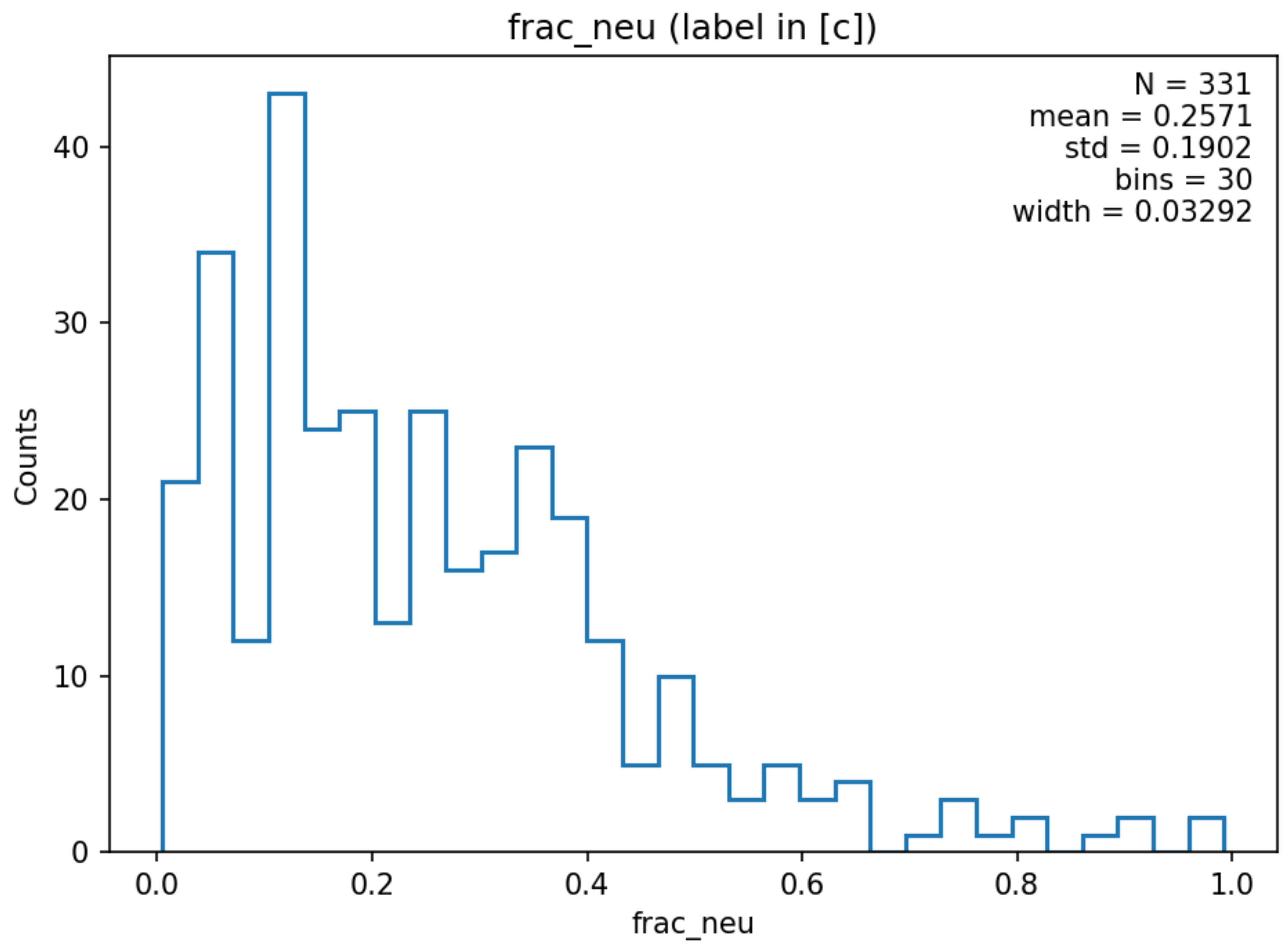
c-jet Feature Plots

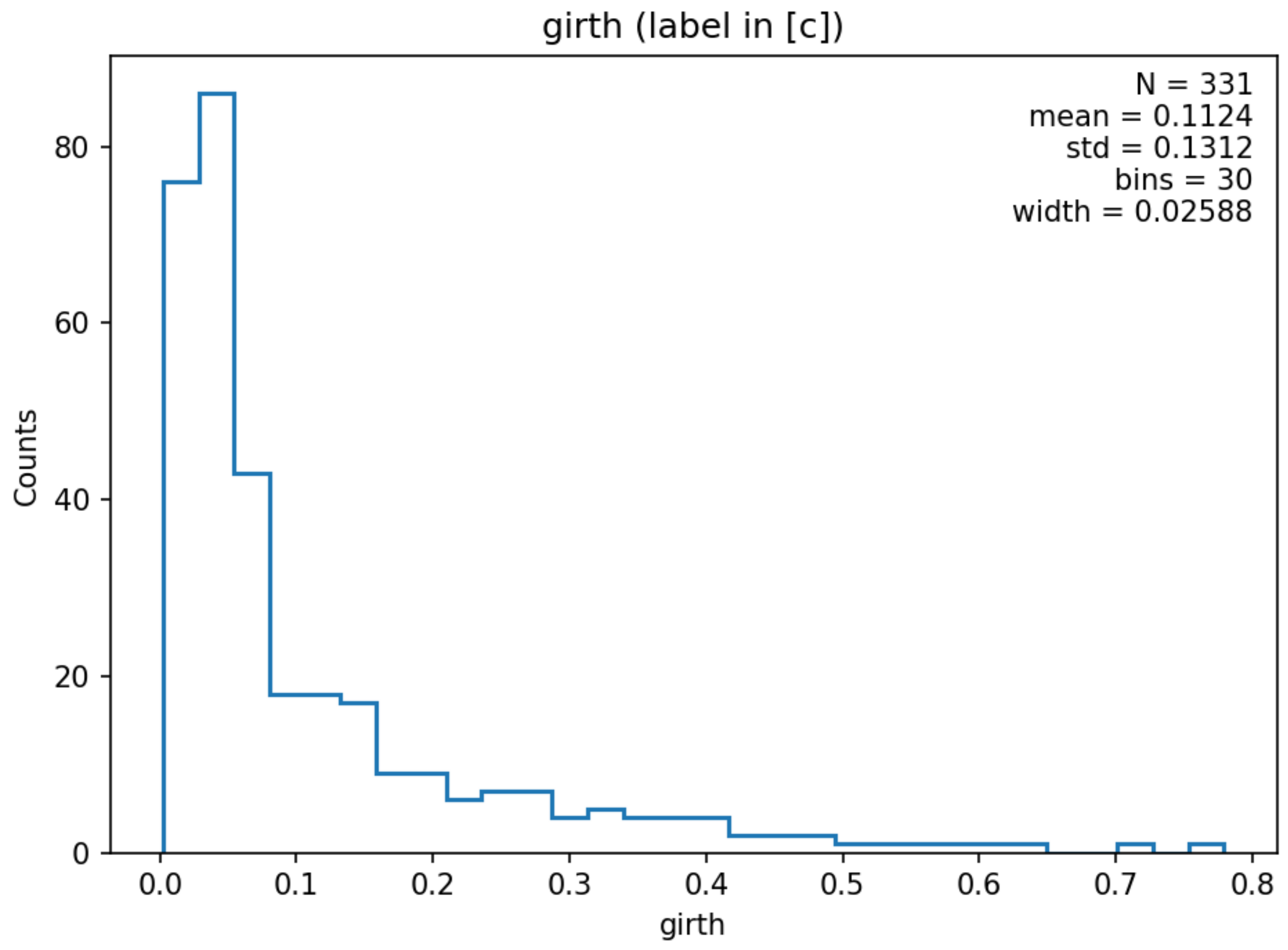




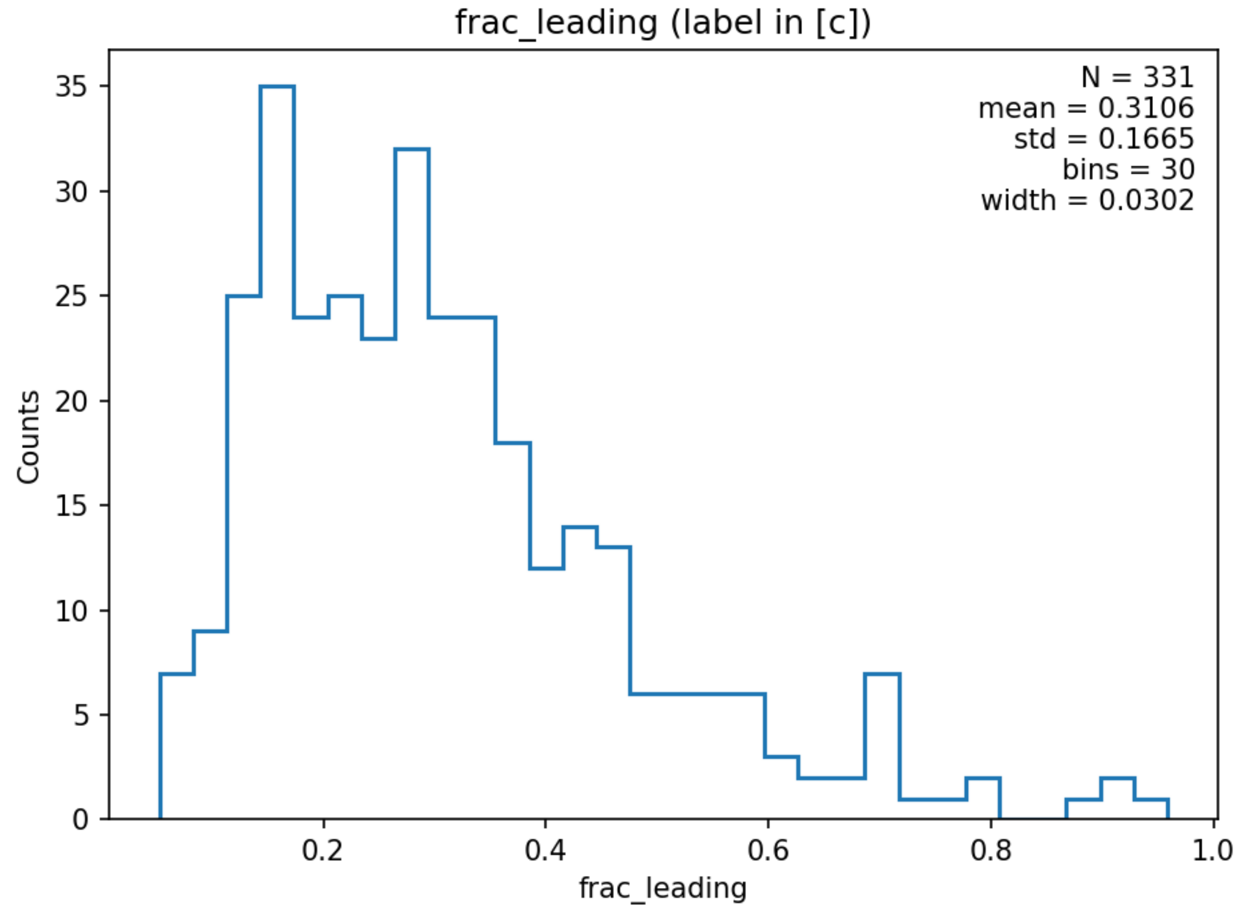


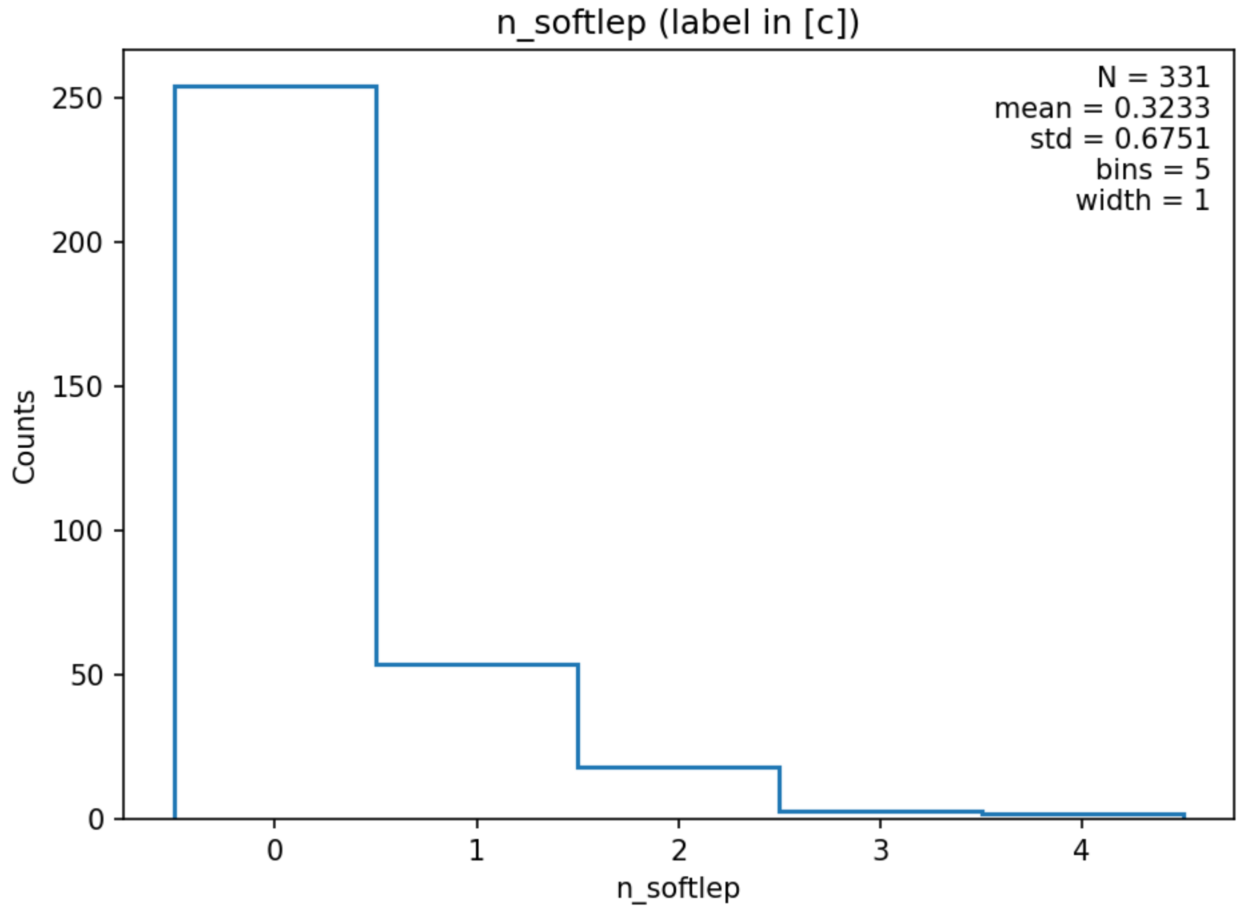


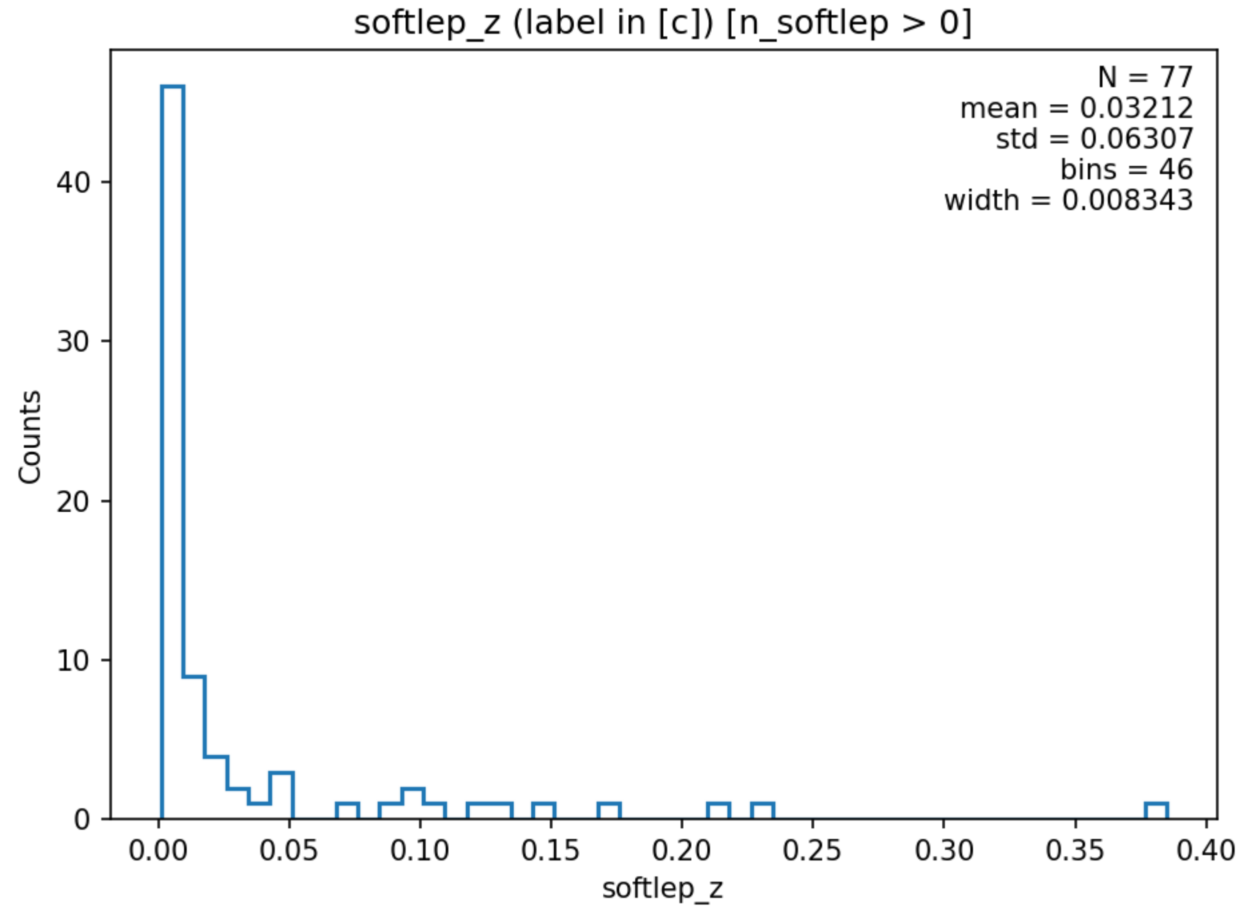


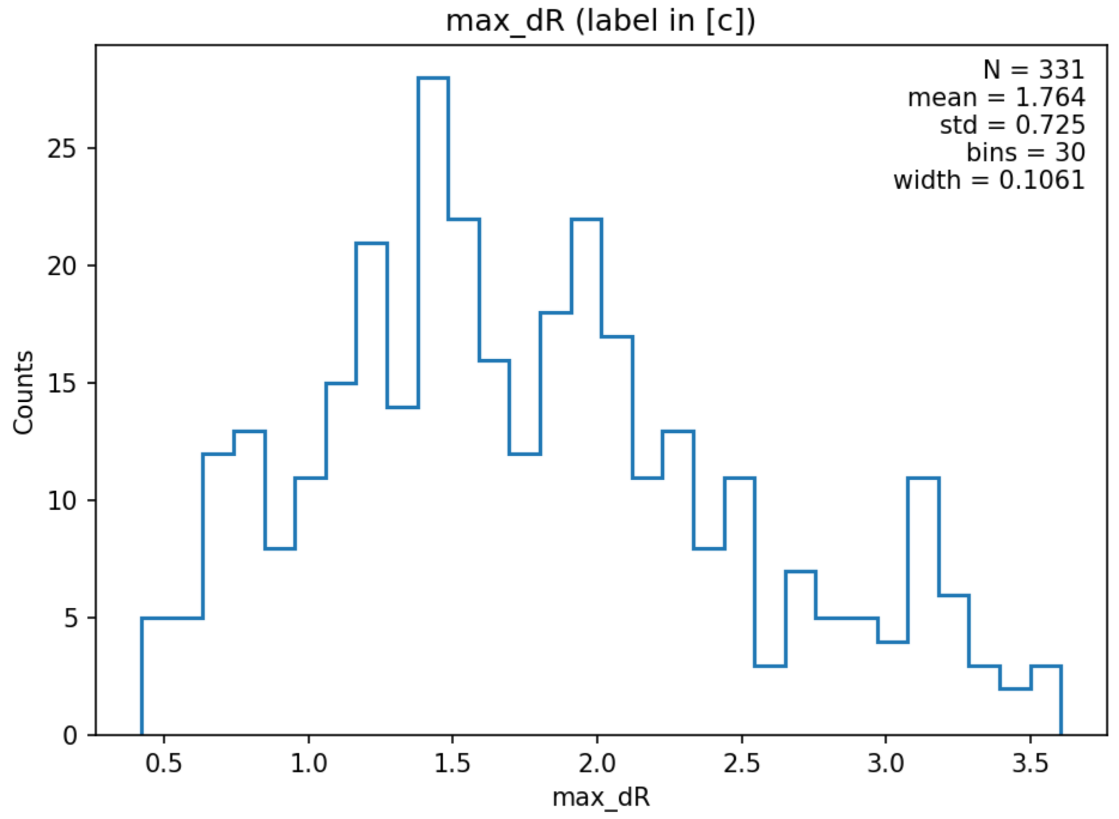


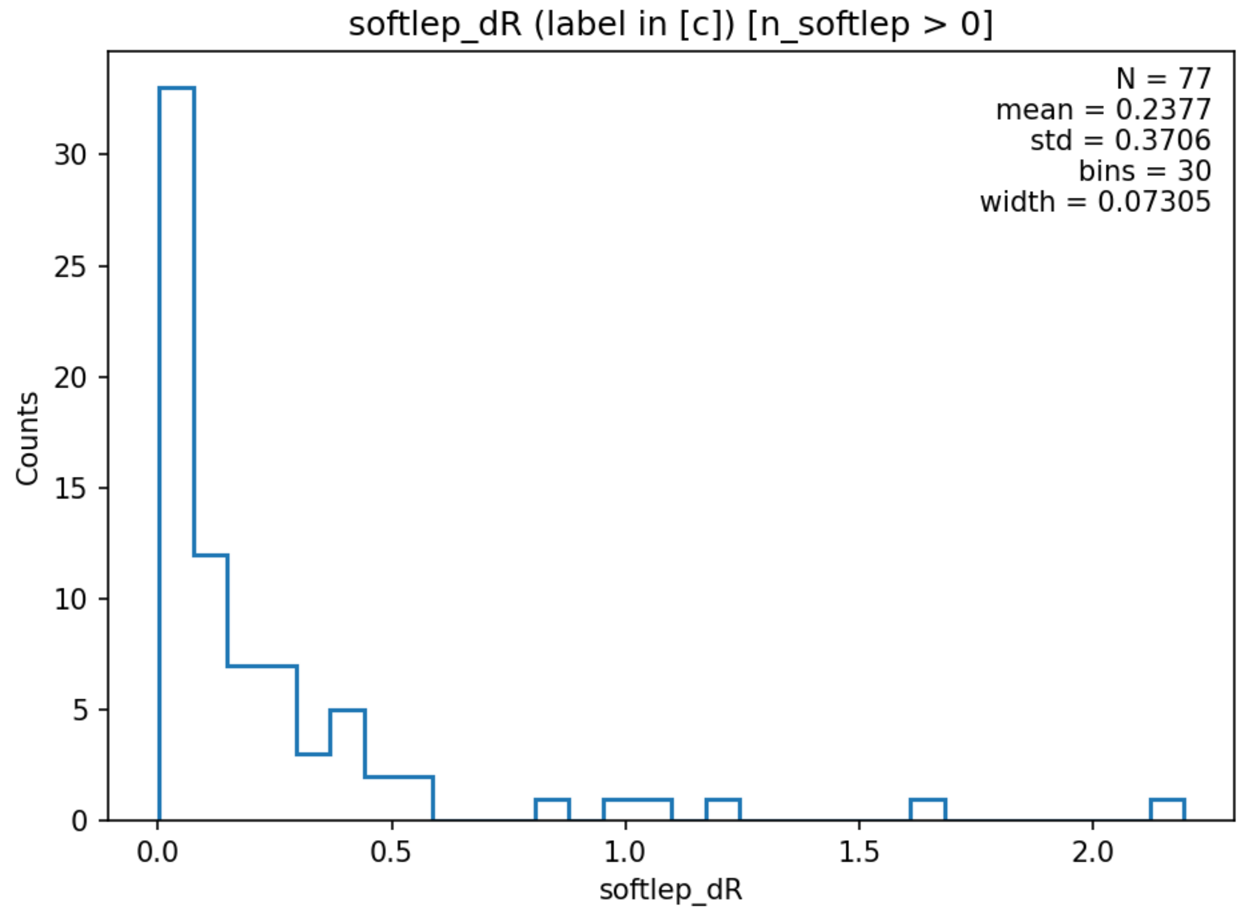
c-jet Feature Plots

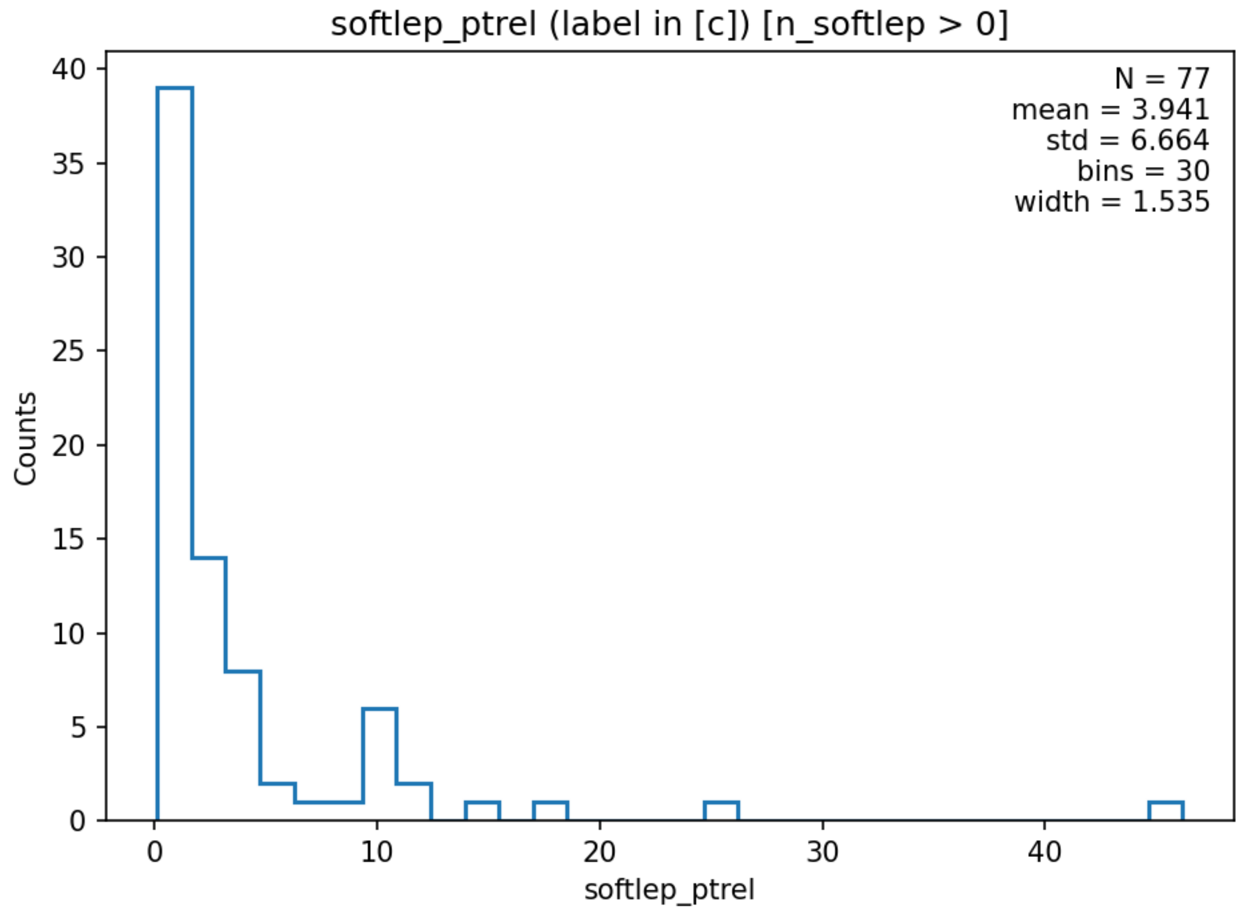


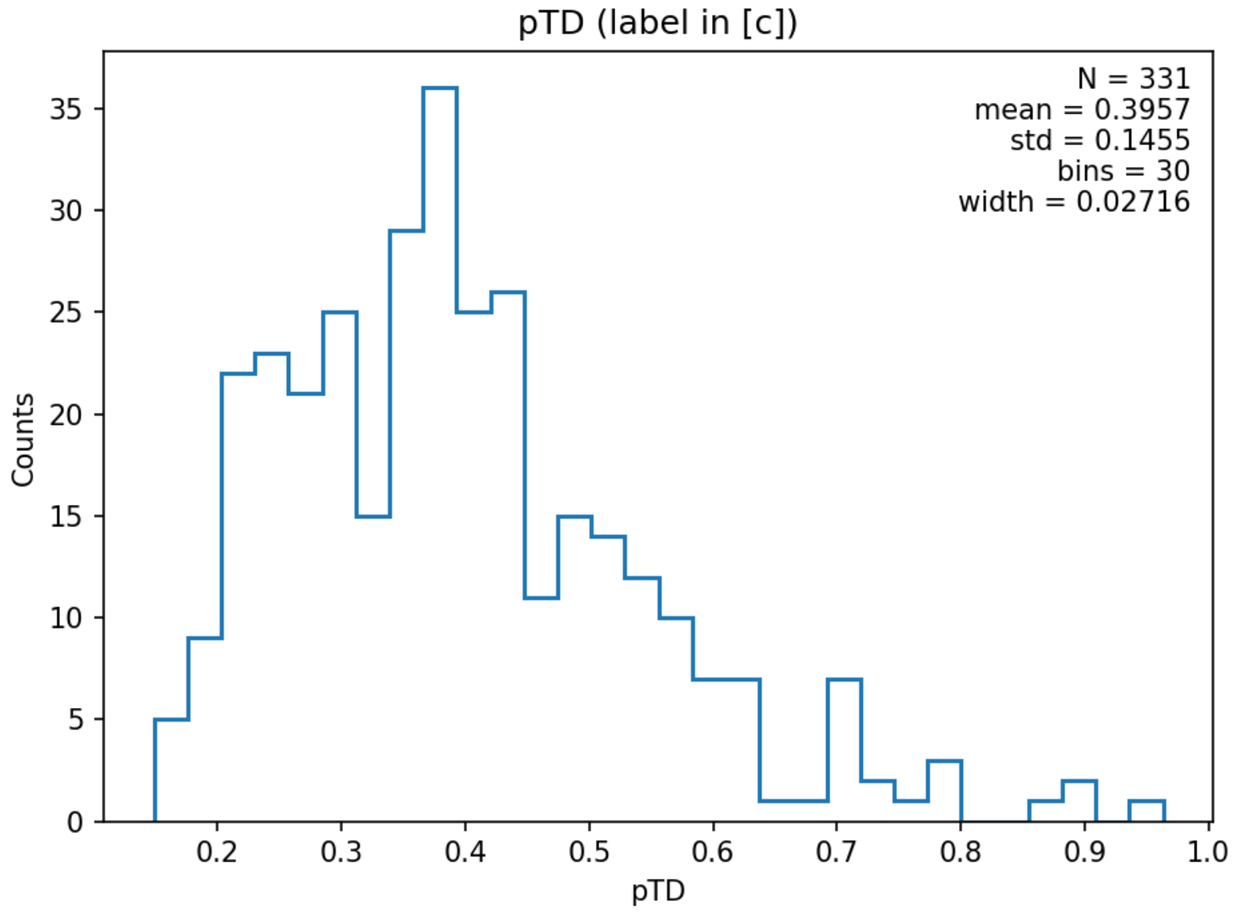




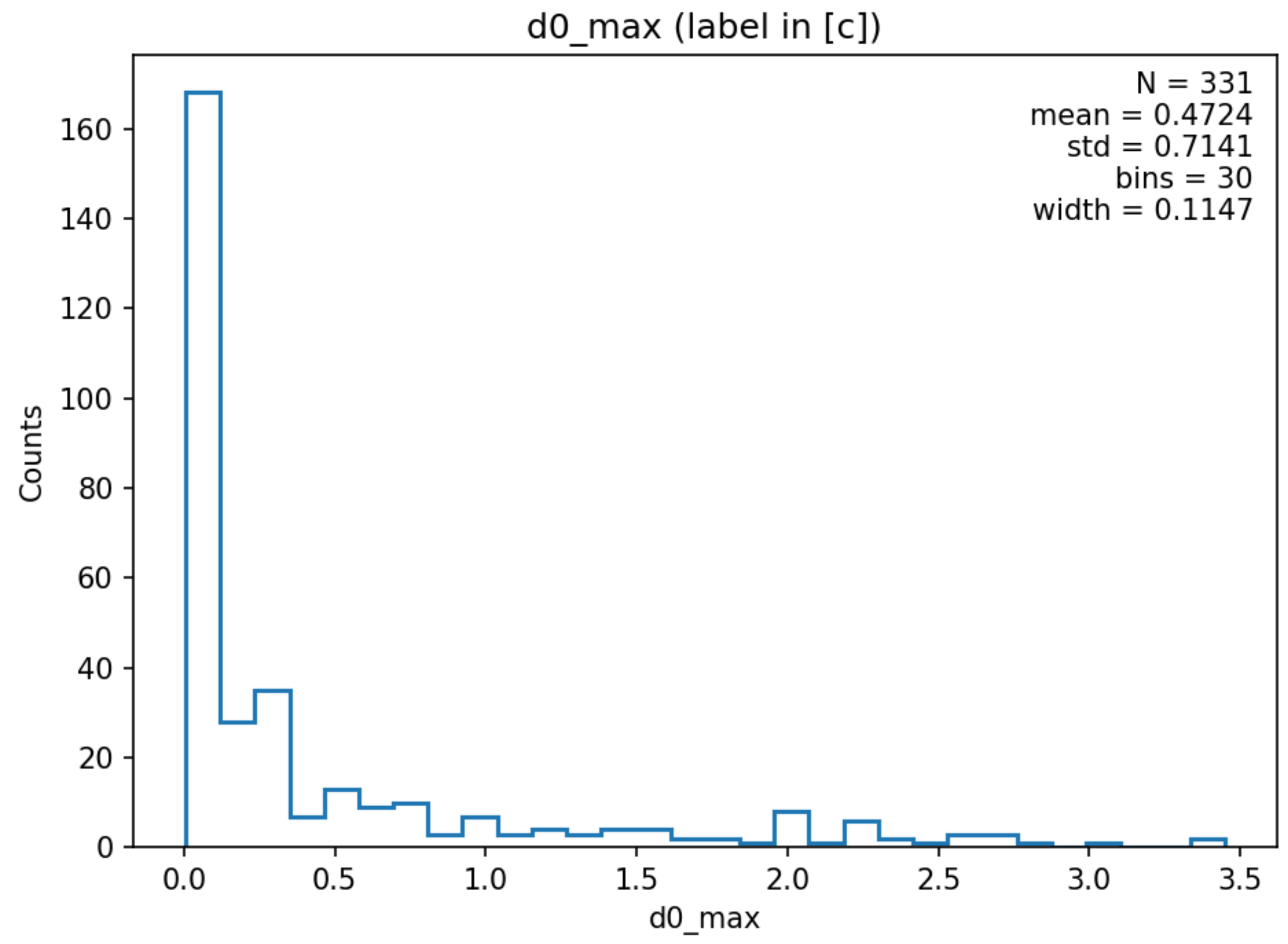


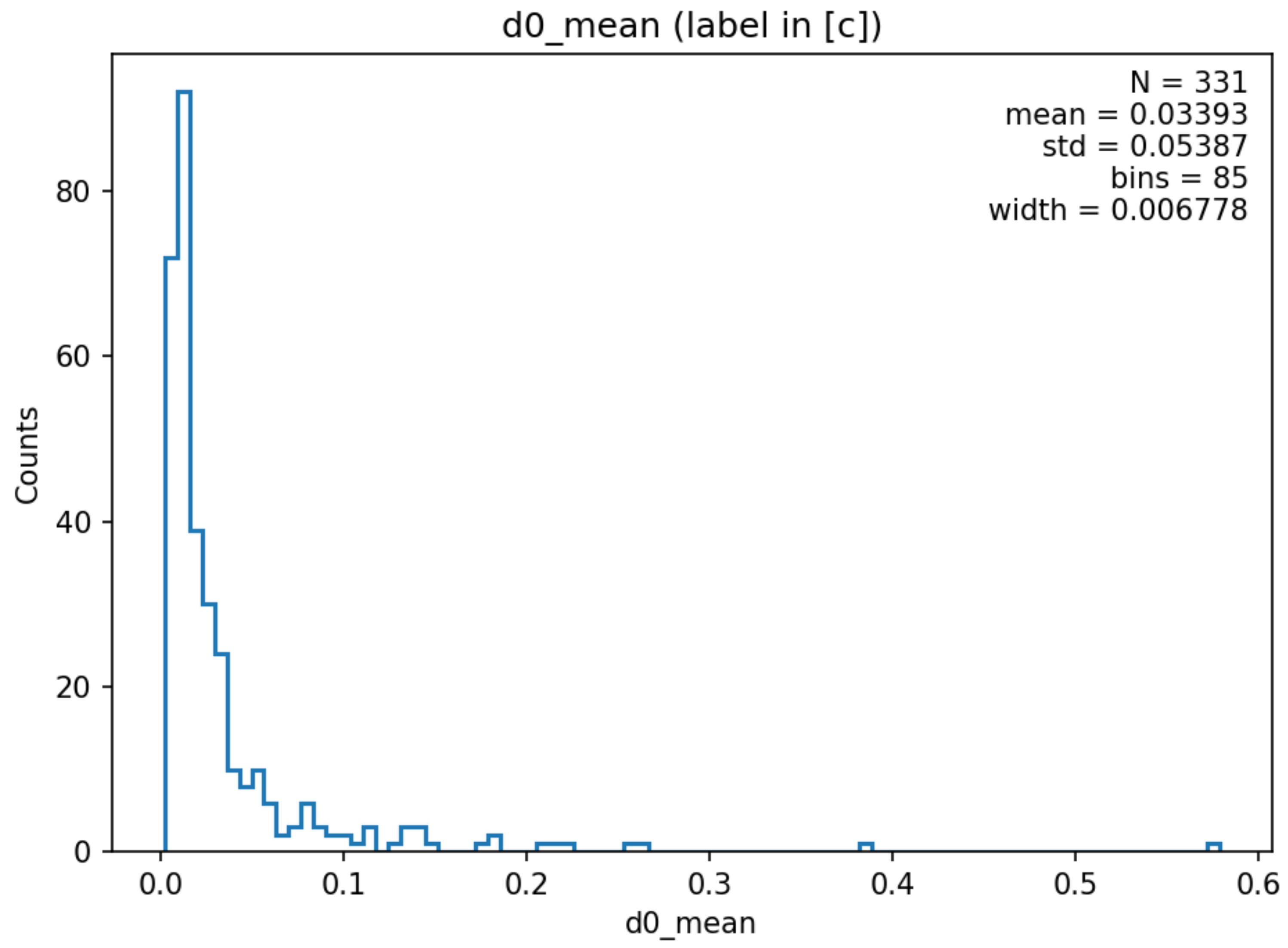


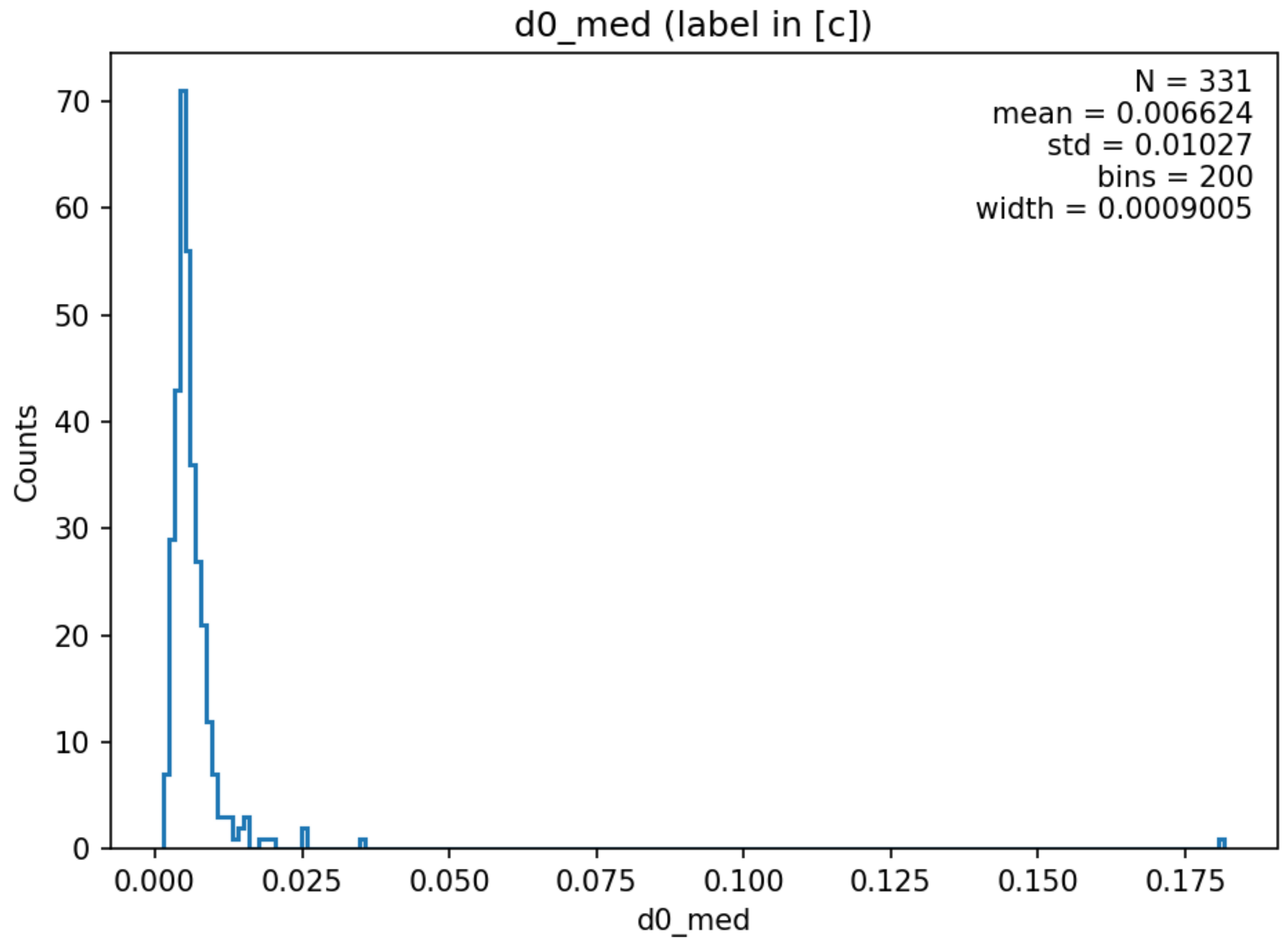


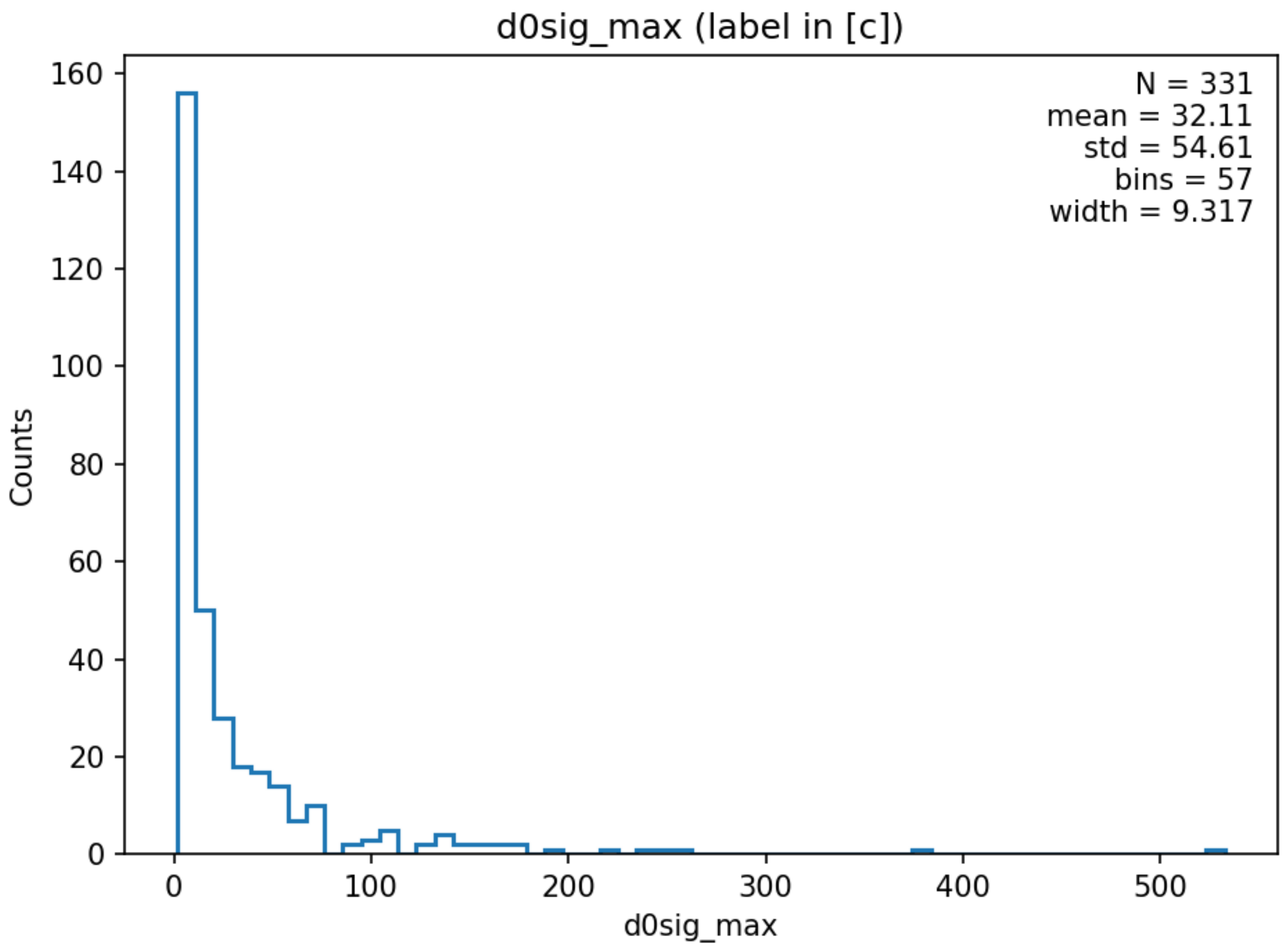


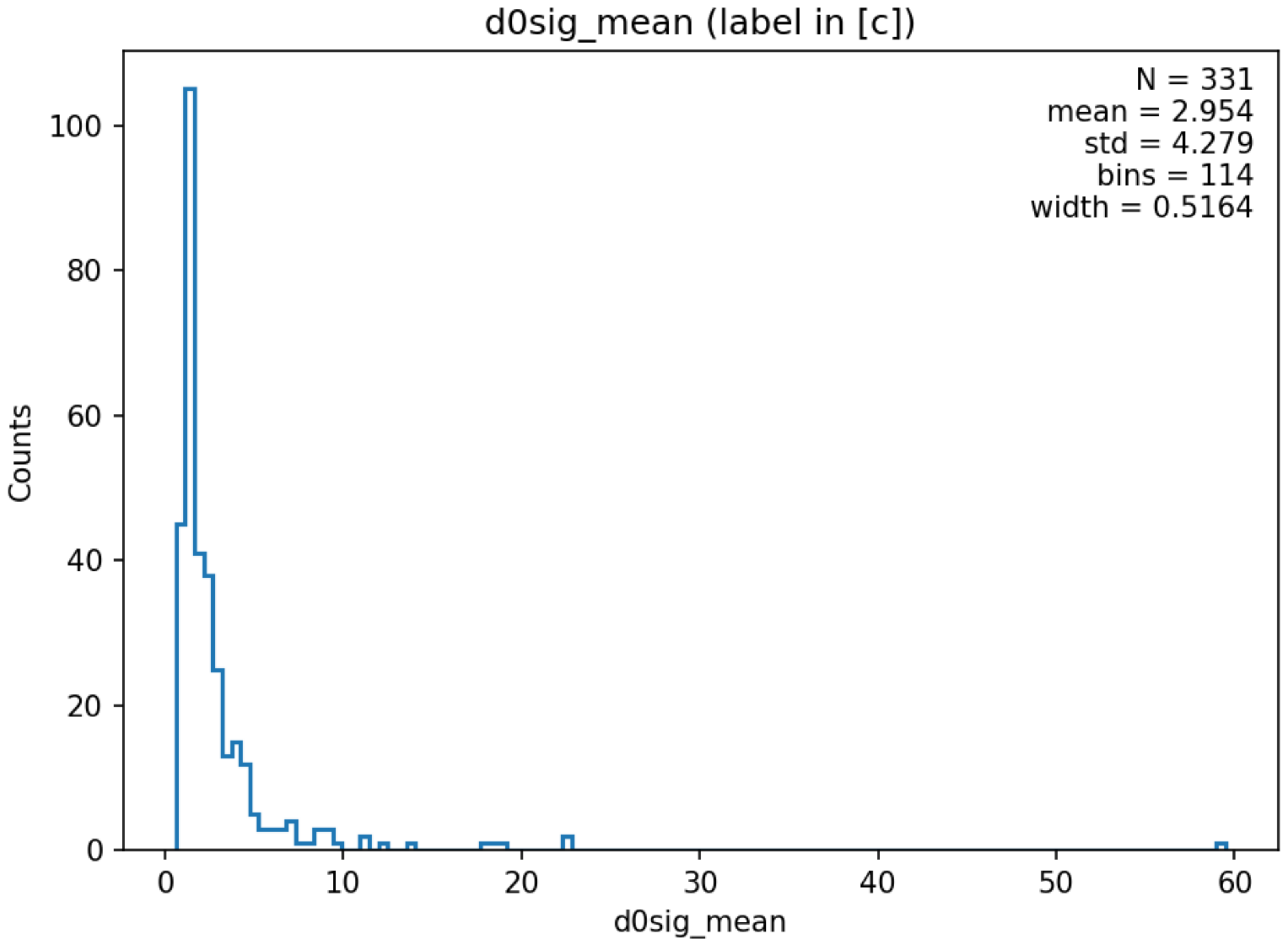
c-jet Feature Plots

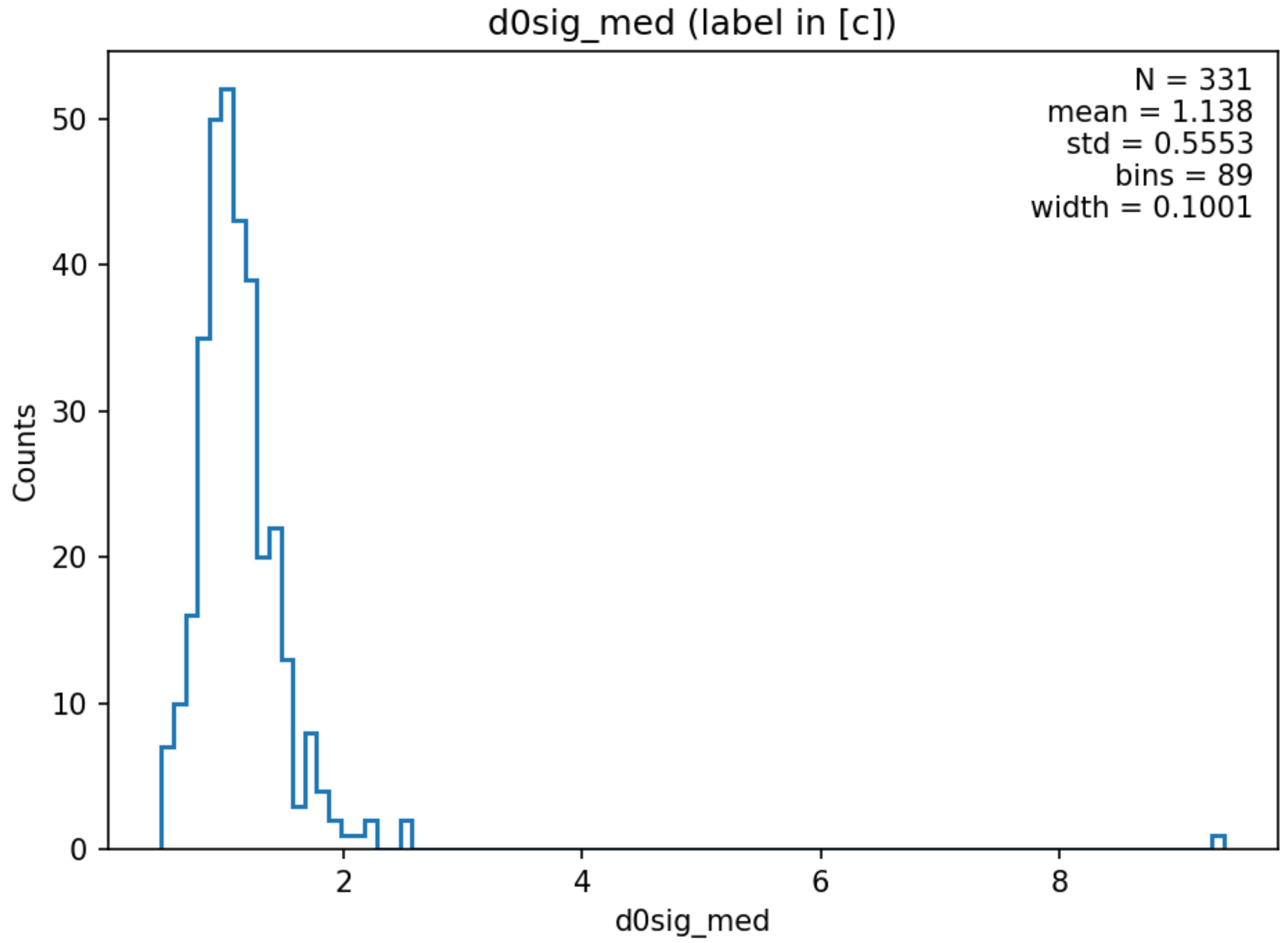




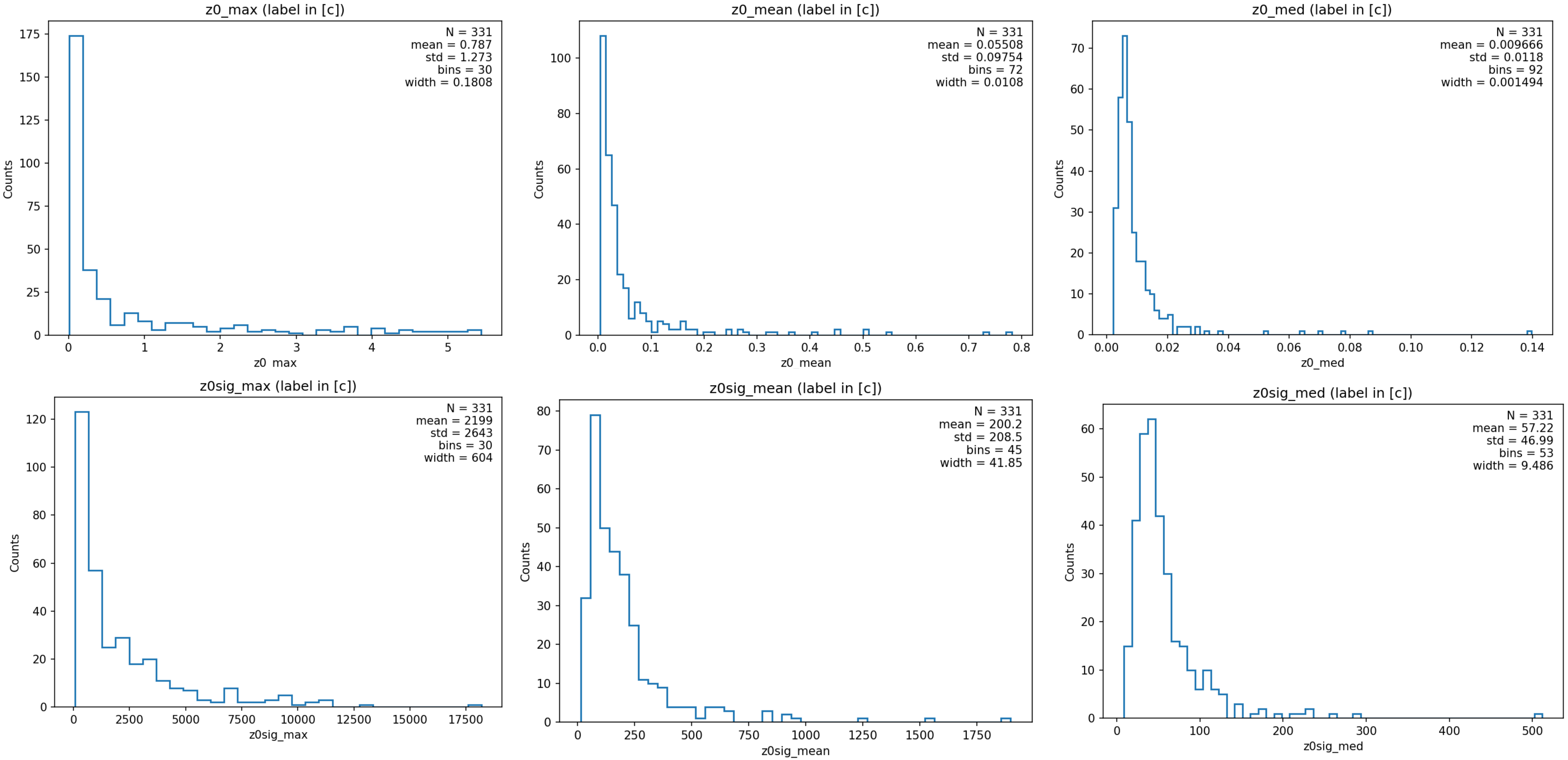




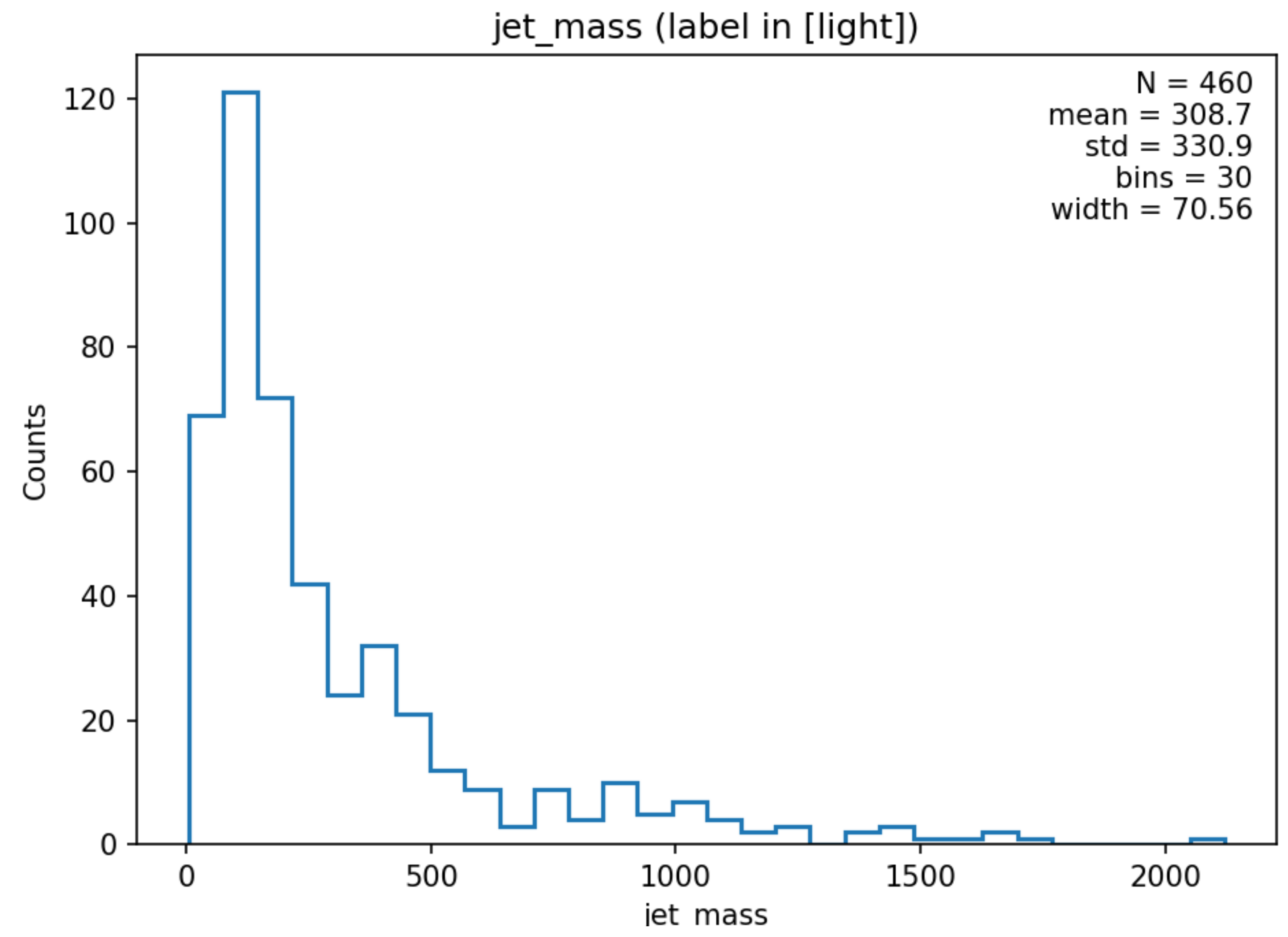


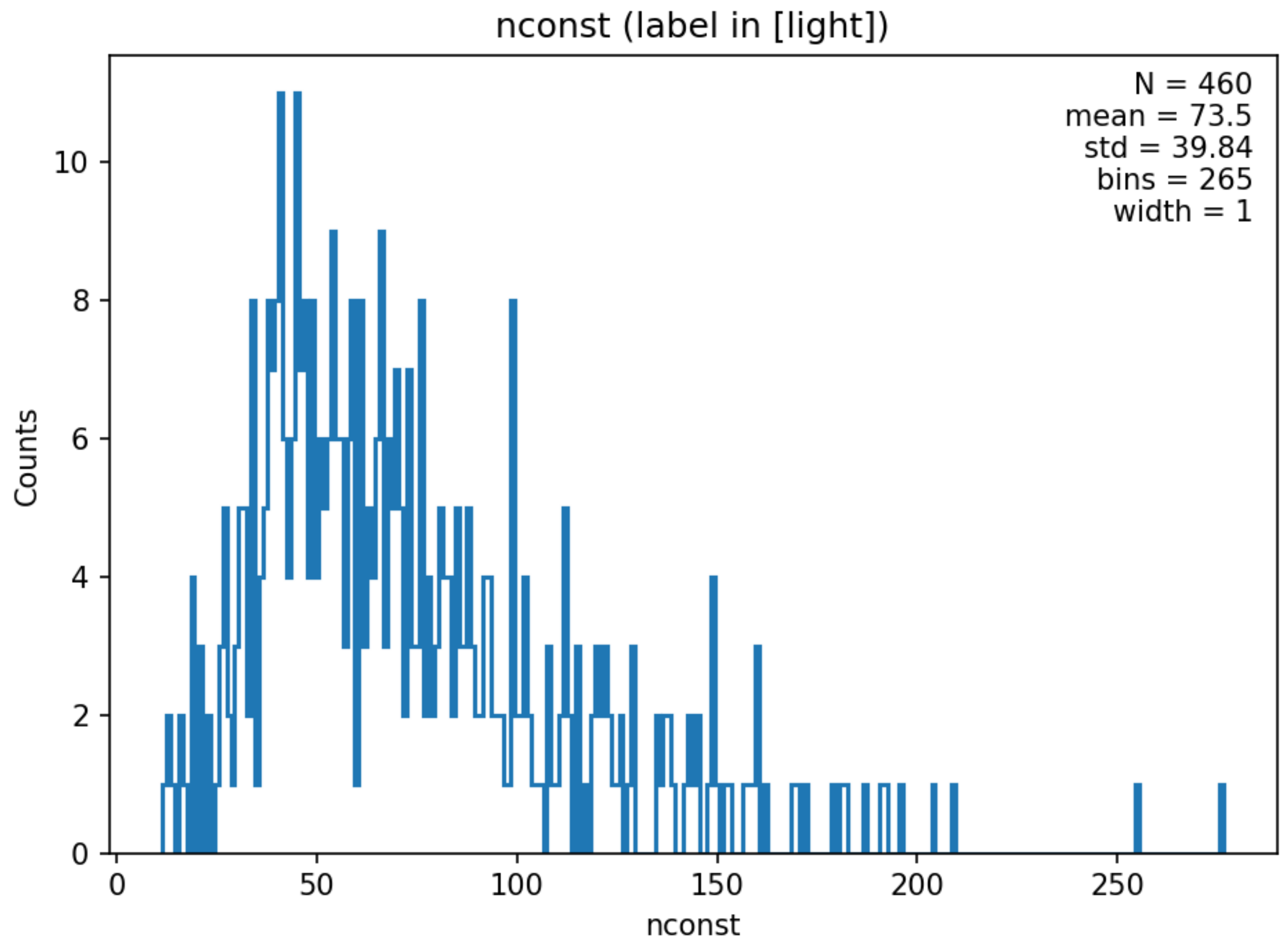


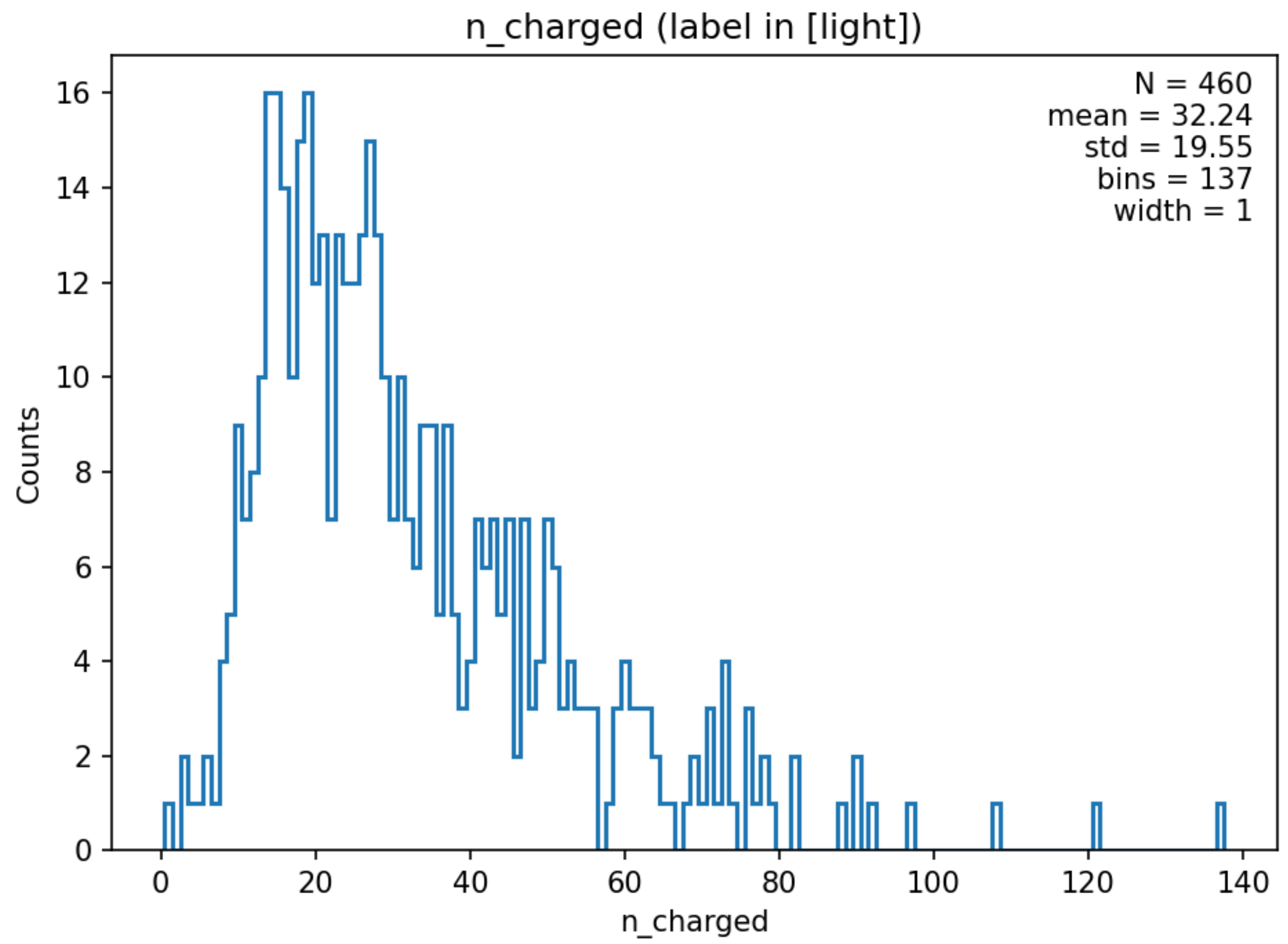
c-jet Feature Plots

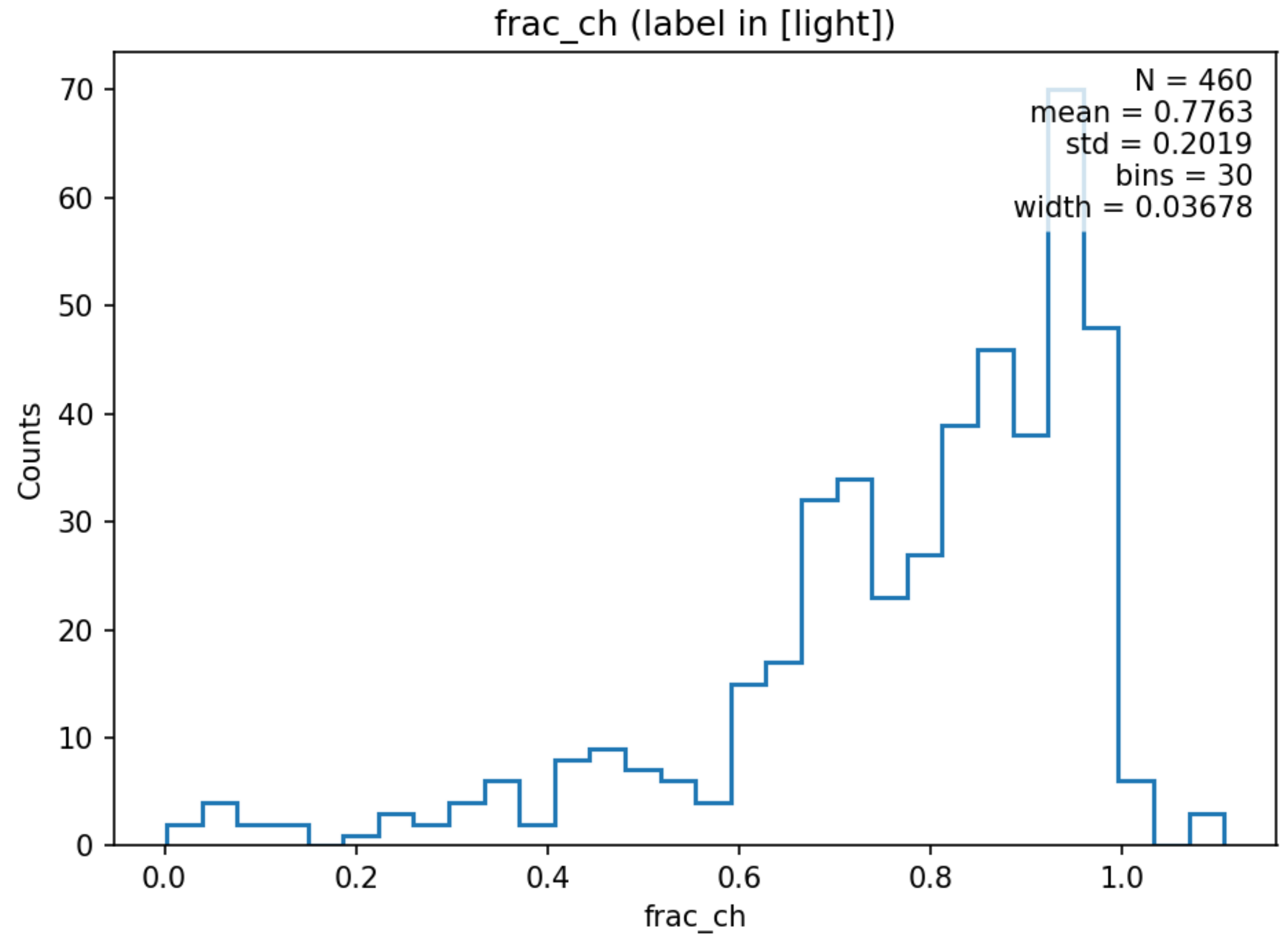


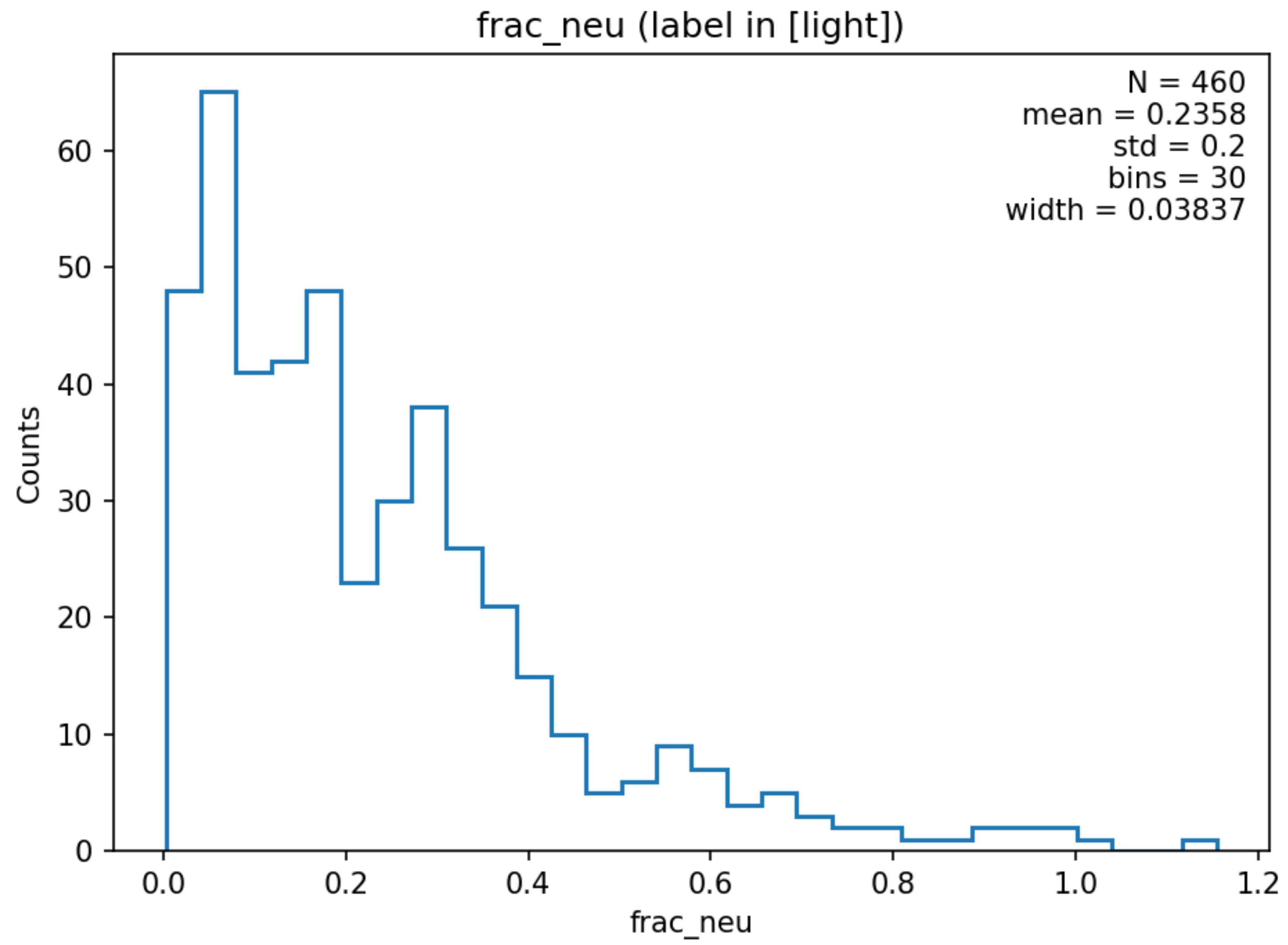
Light jets Feature Plots

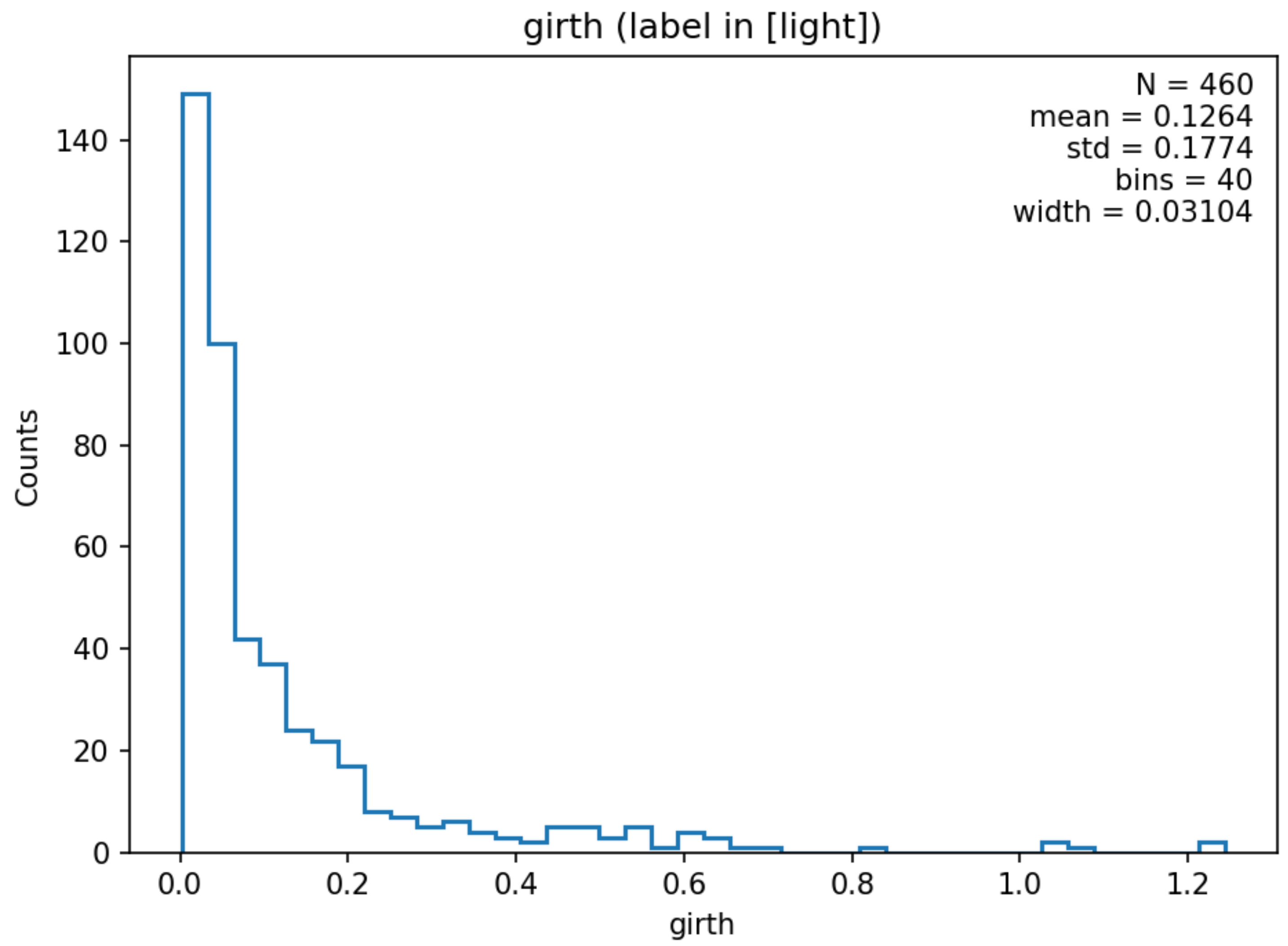






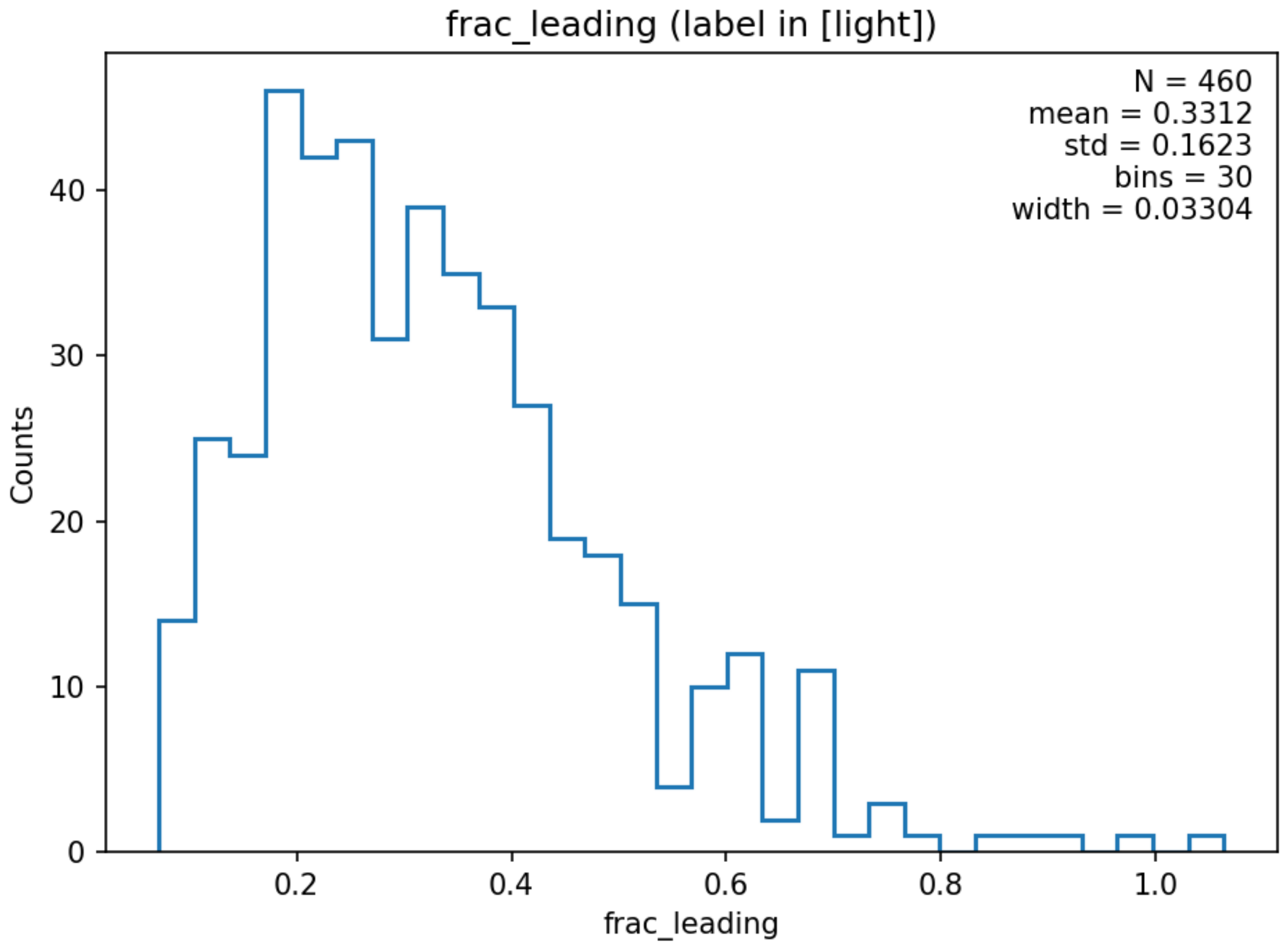


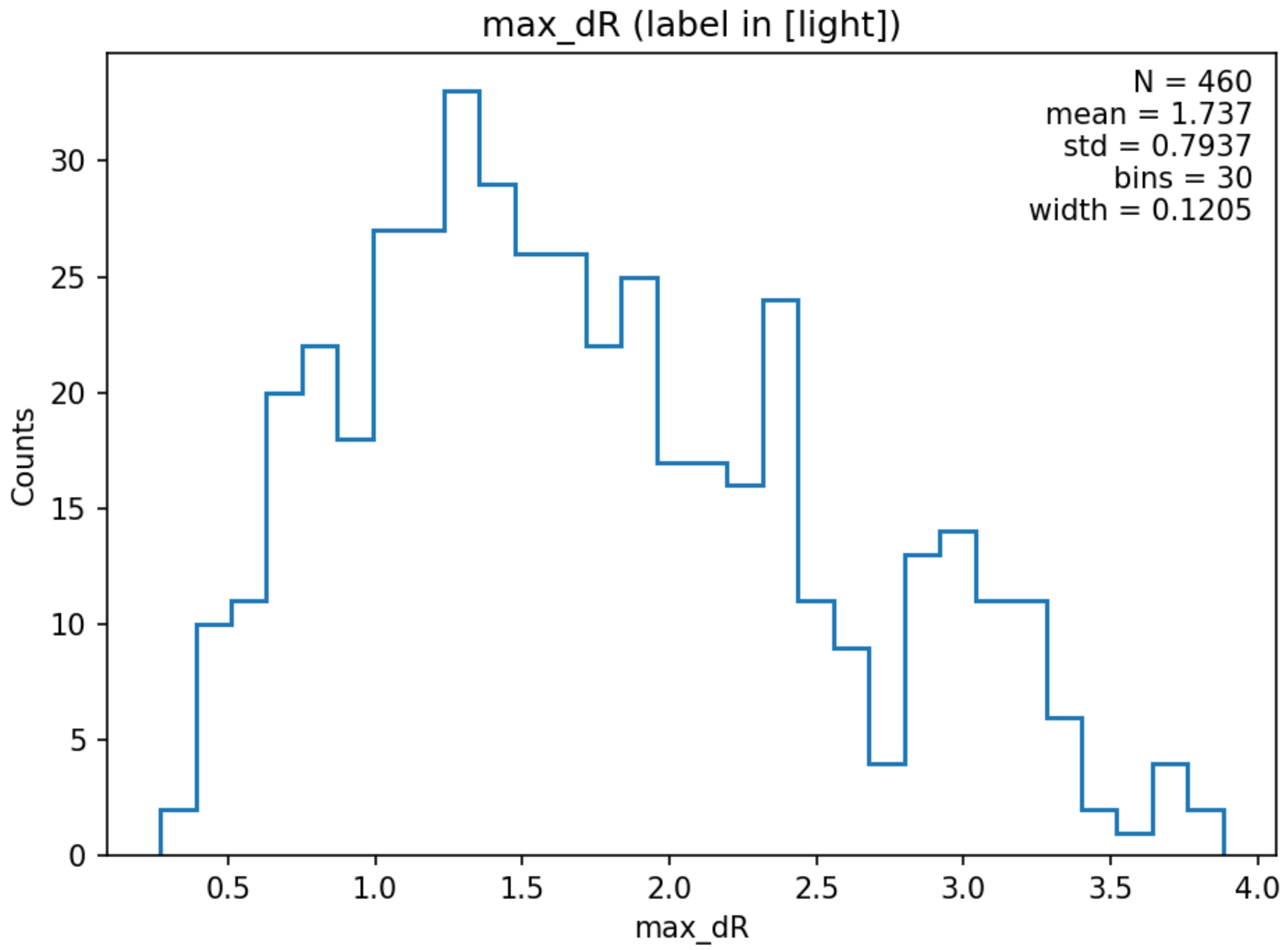


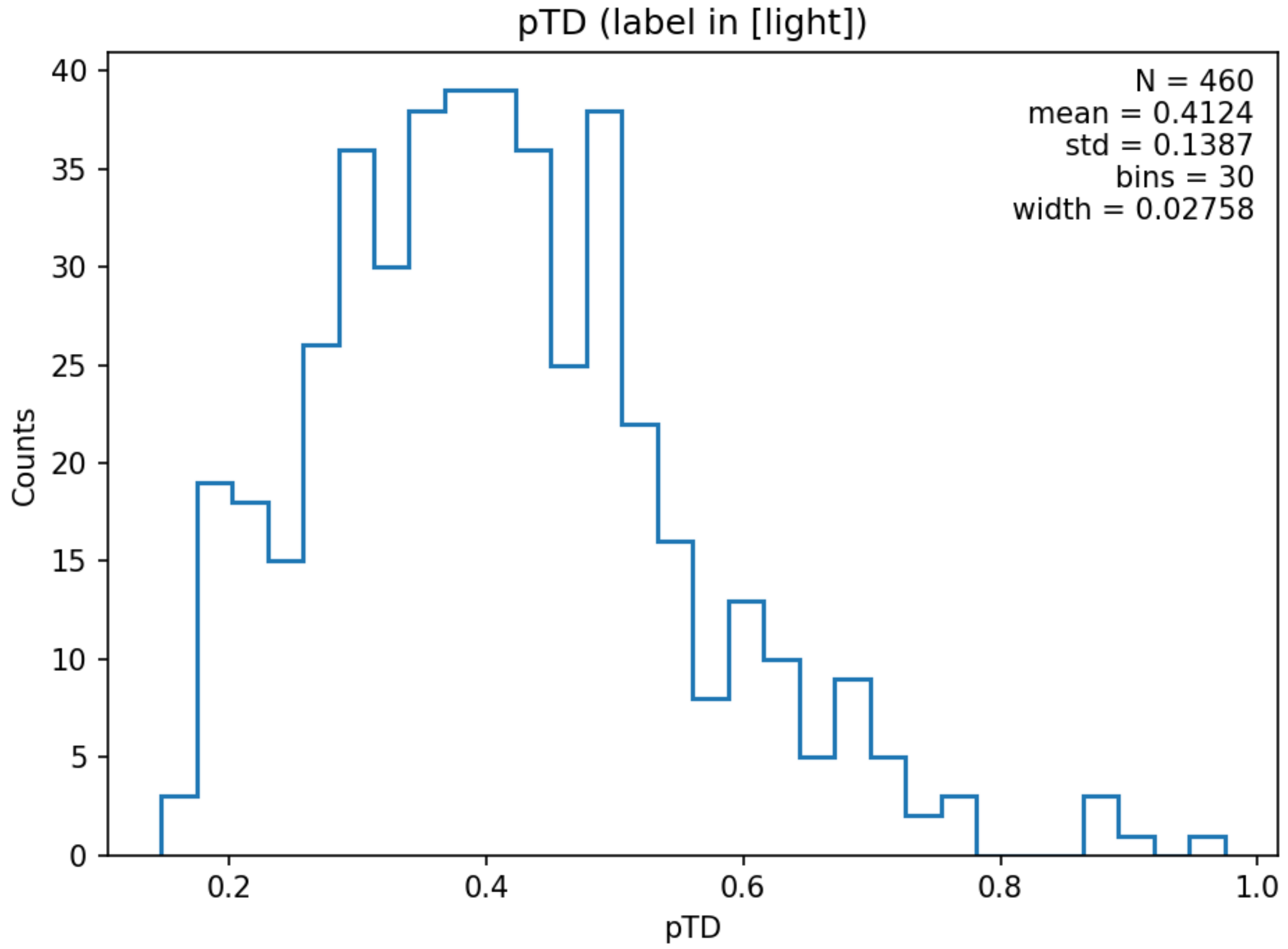


Light jet Feature Plots

For **light jets**



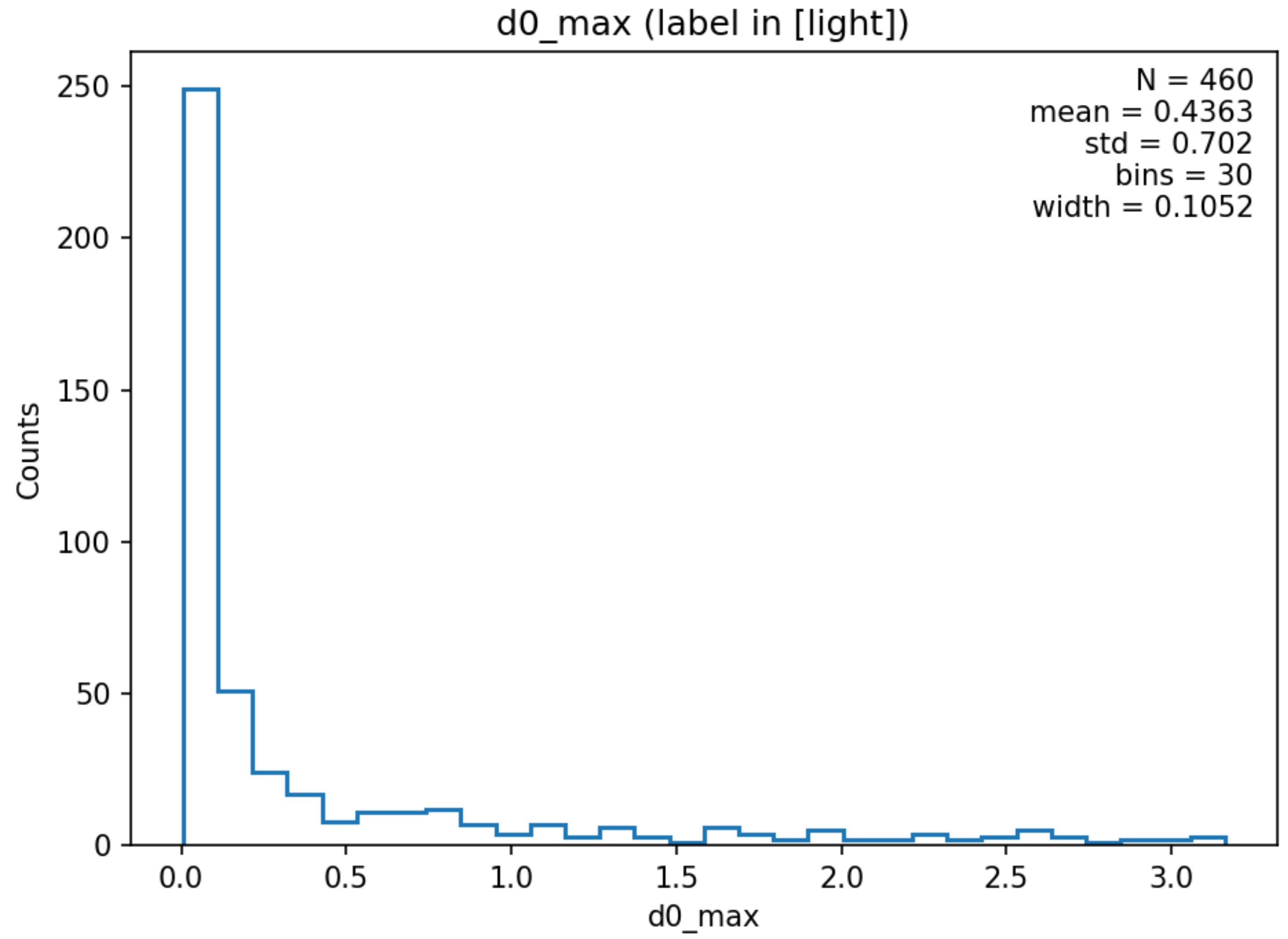


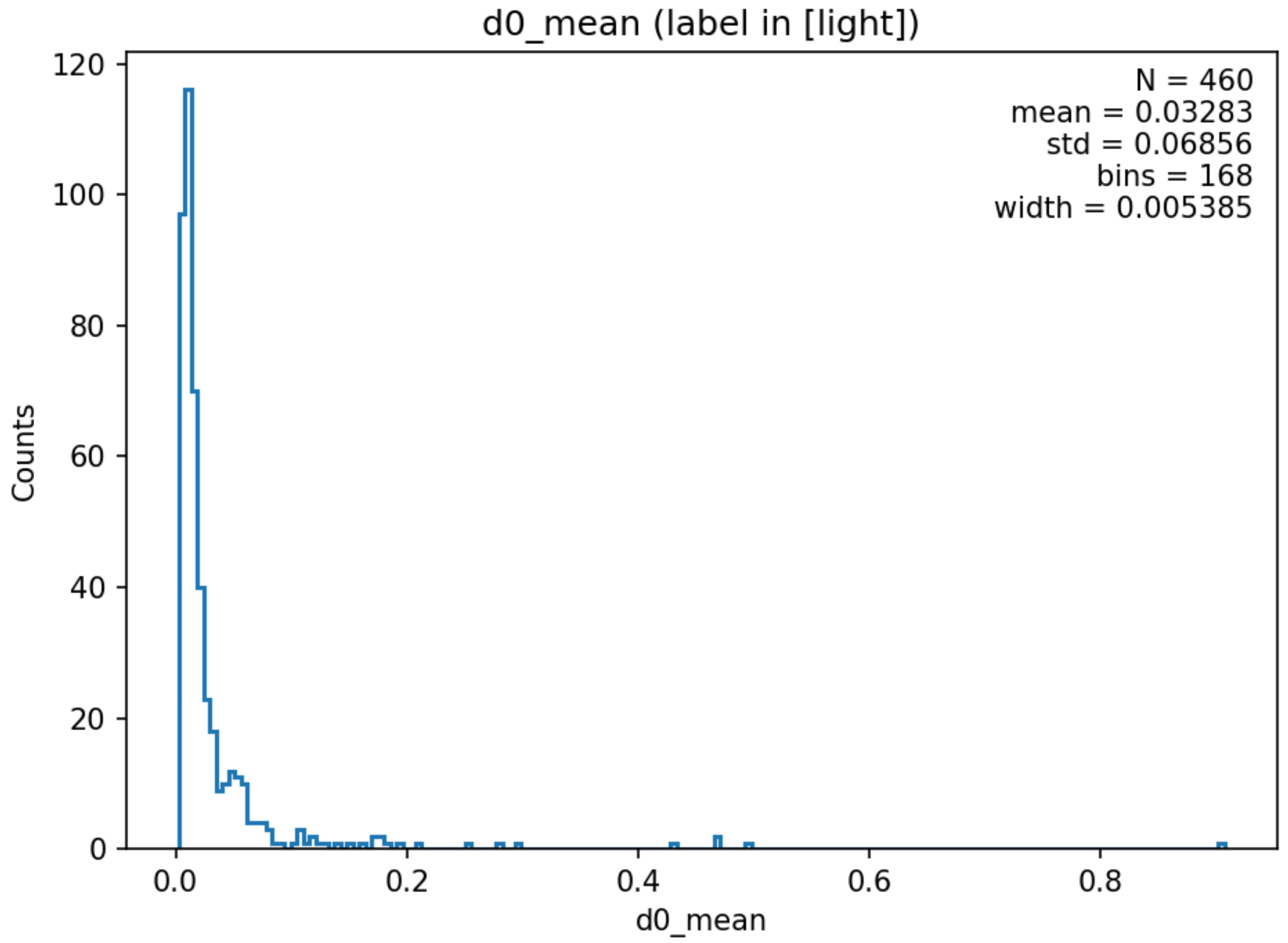


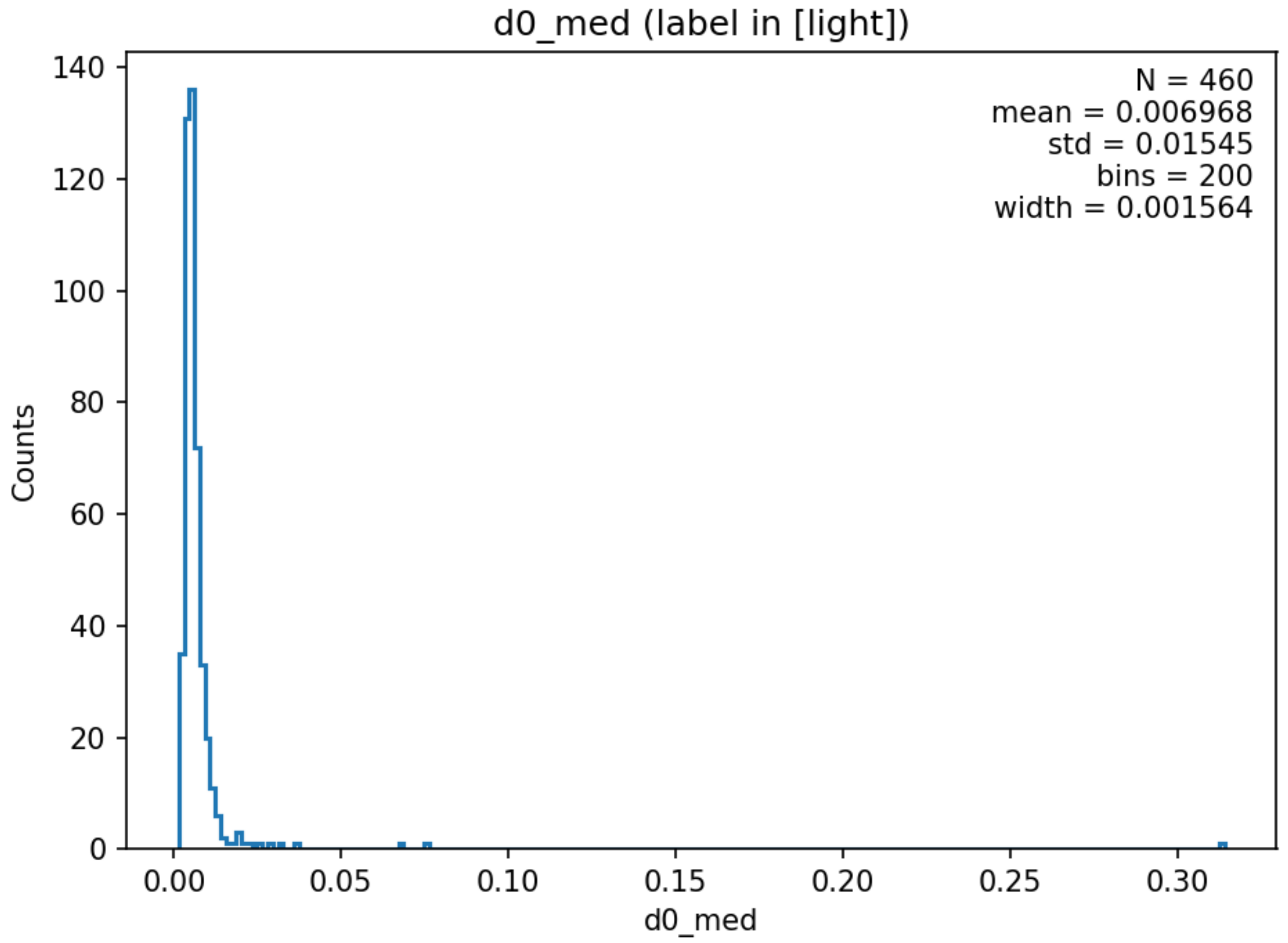
No soft leptons in light jets => no soft lepton plots

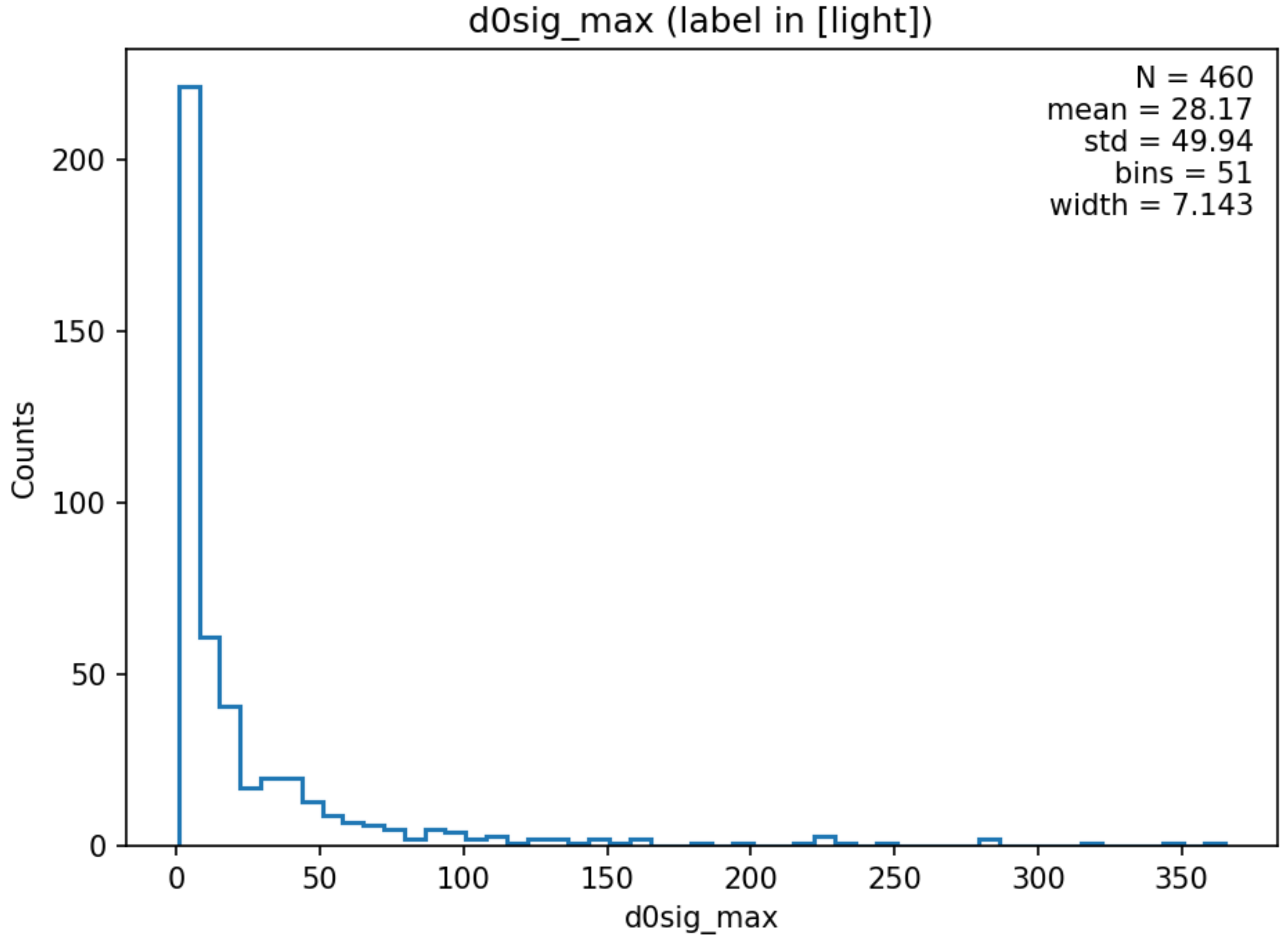
Light jet Feature Plots

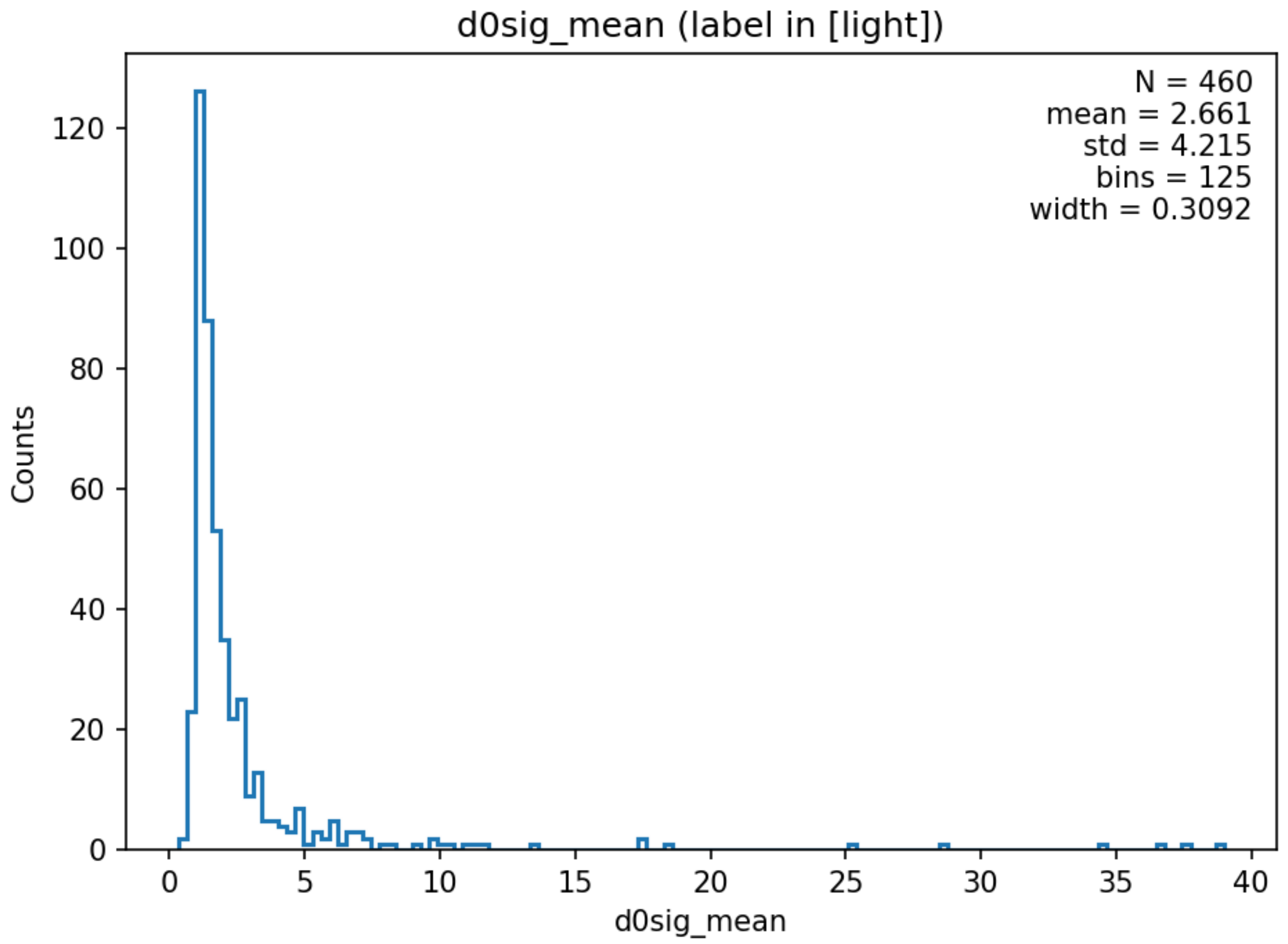
For **light jets**

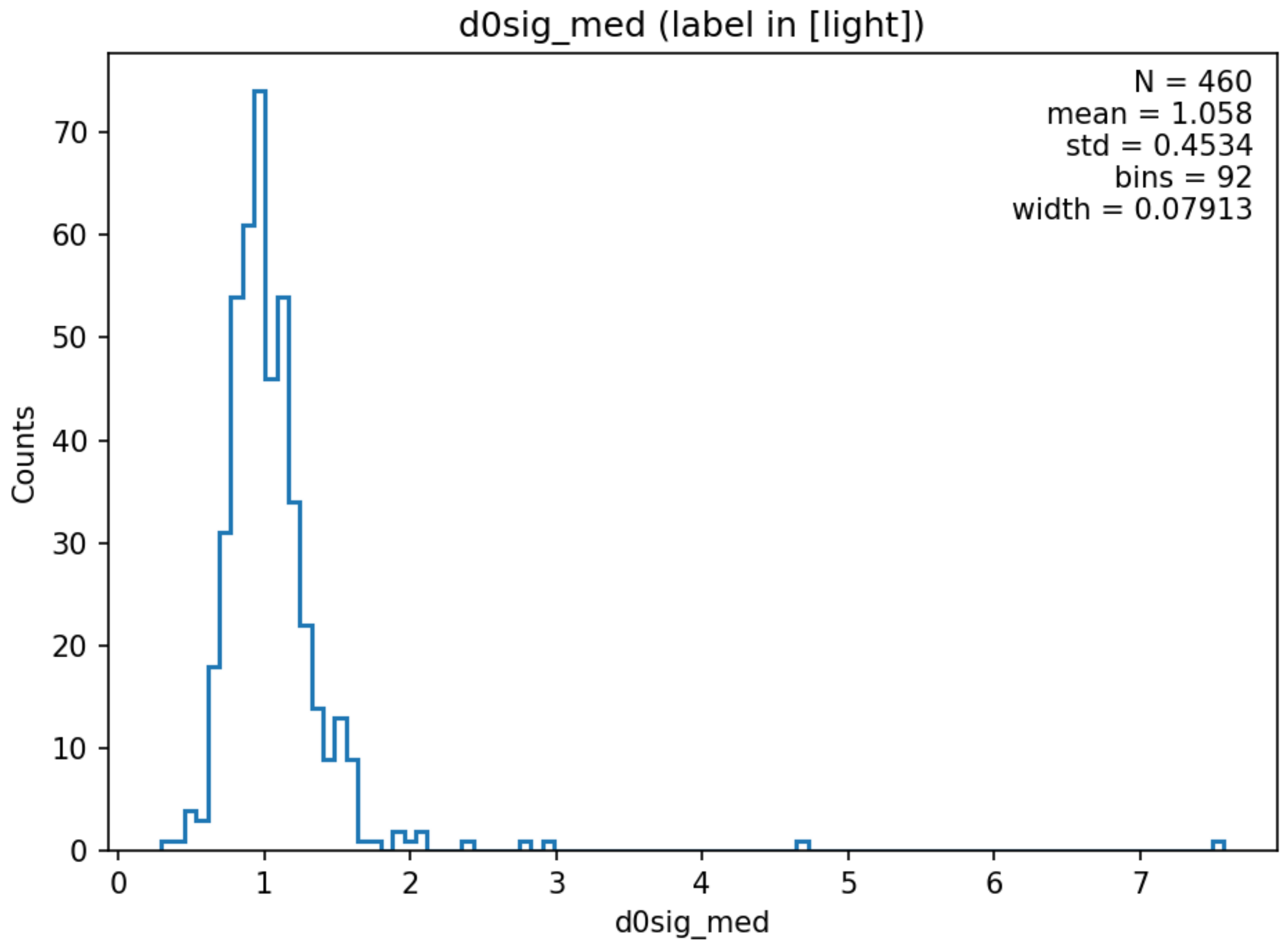

d0_max (label in [light])


d0_mean (label in [light])


d0_med (label in [light])

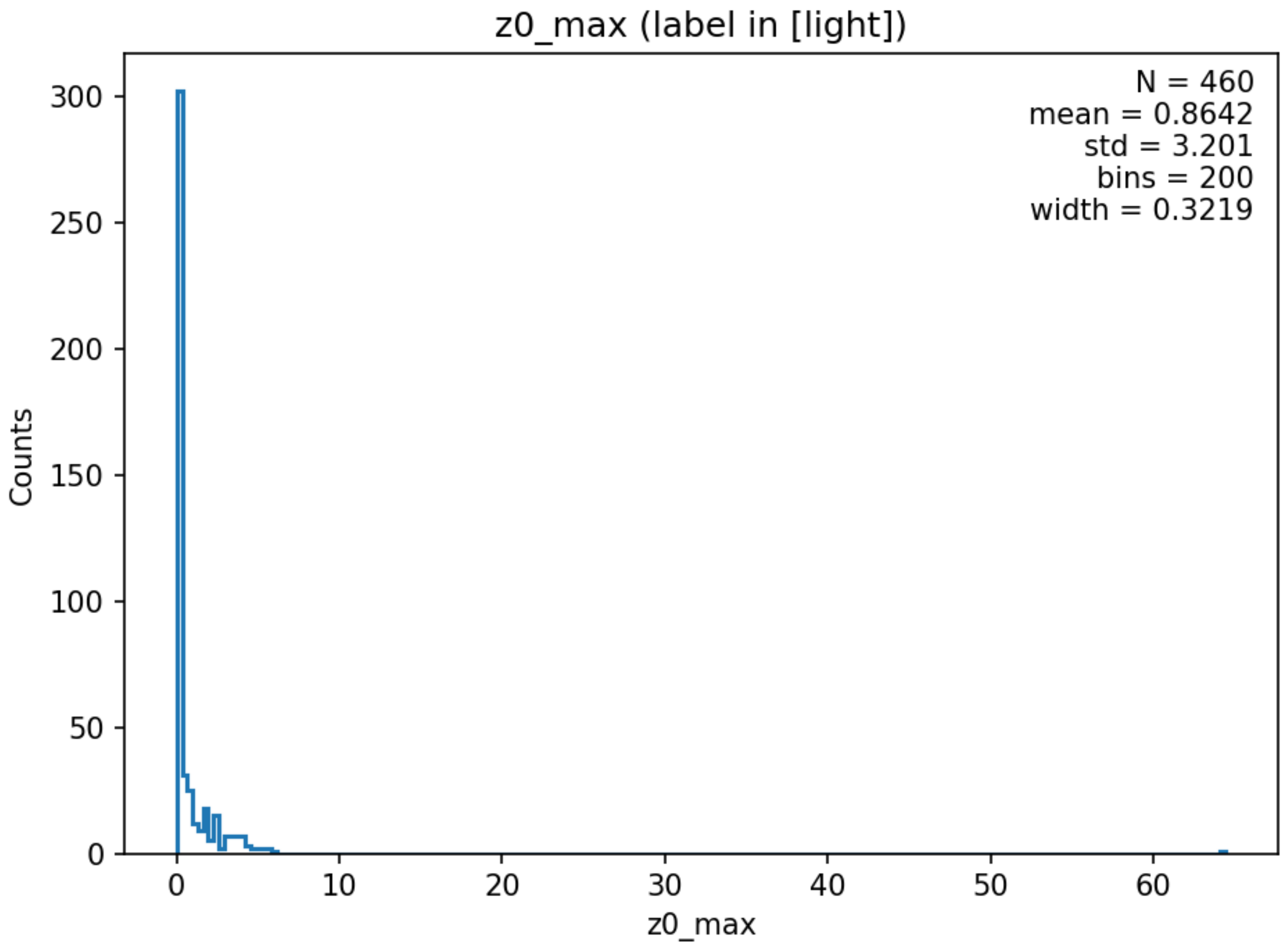

d0sig_max (label in [light])

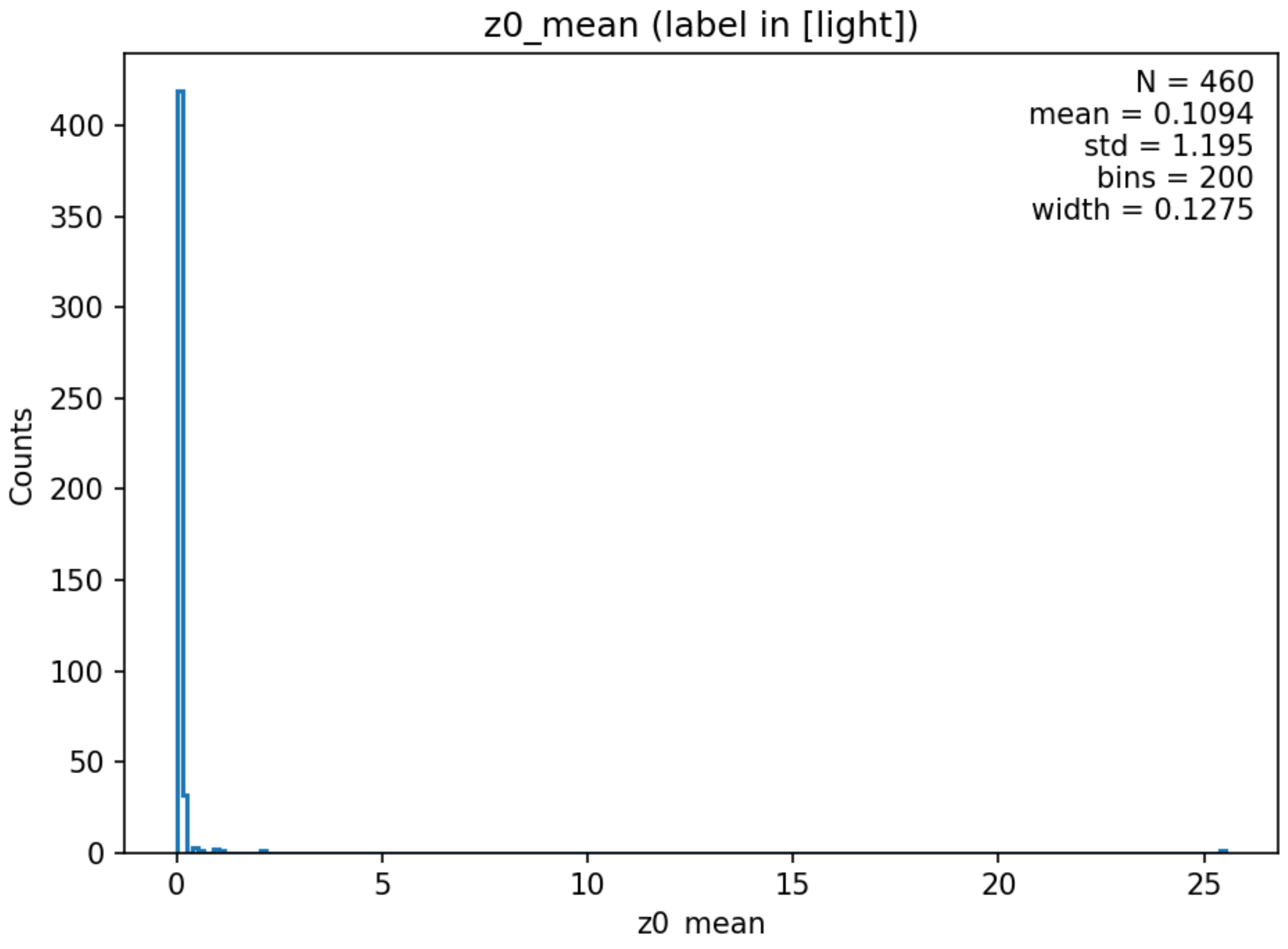

d0sig_mean (label in [light])

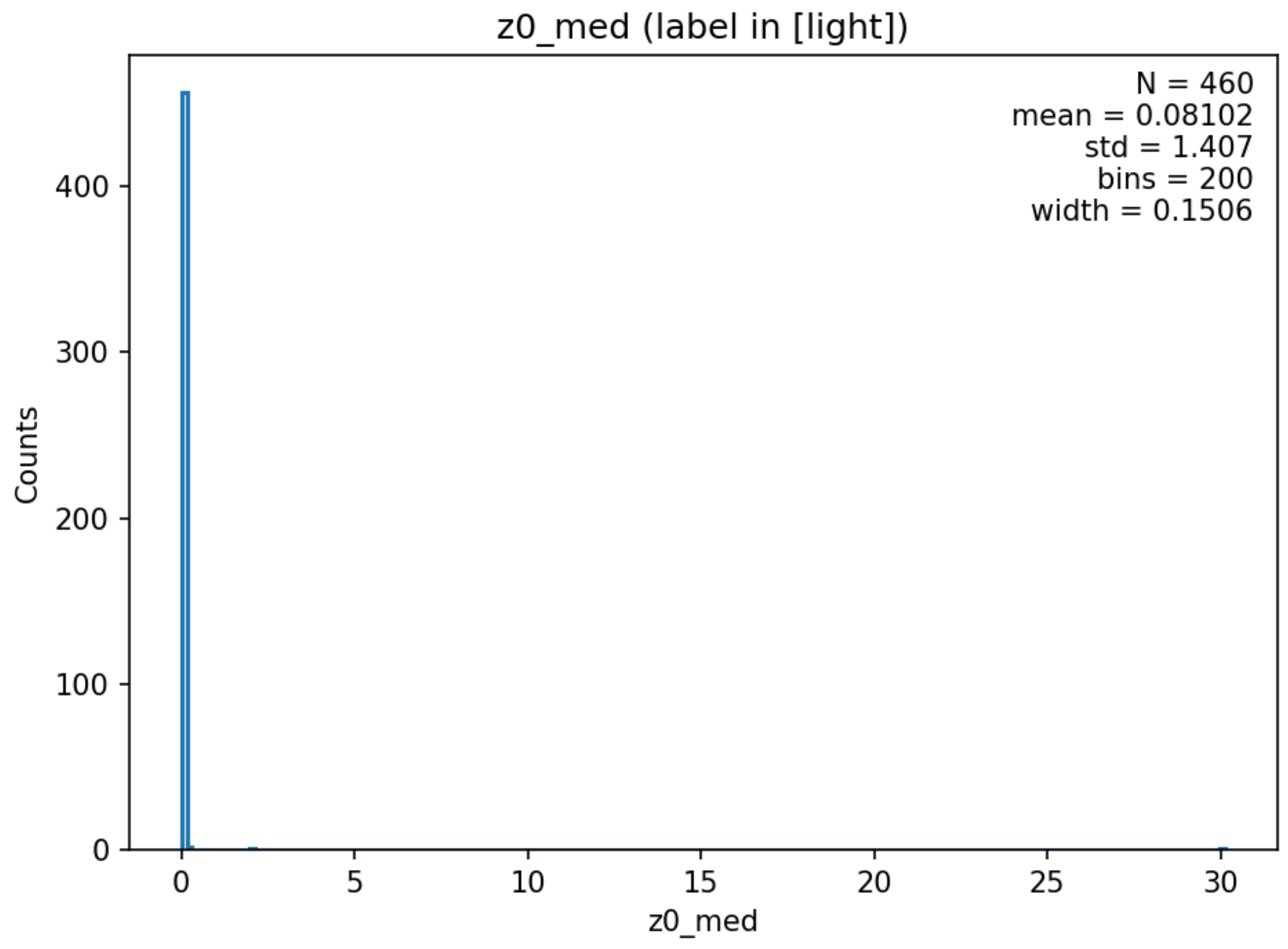

d0sig_med (label in [light])

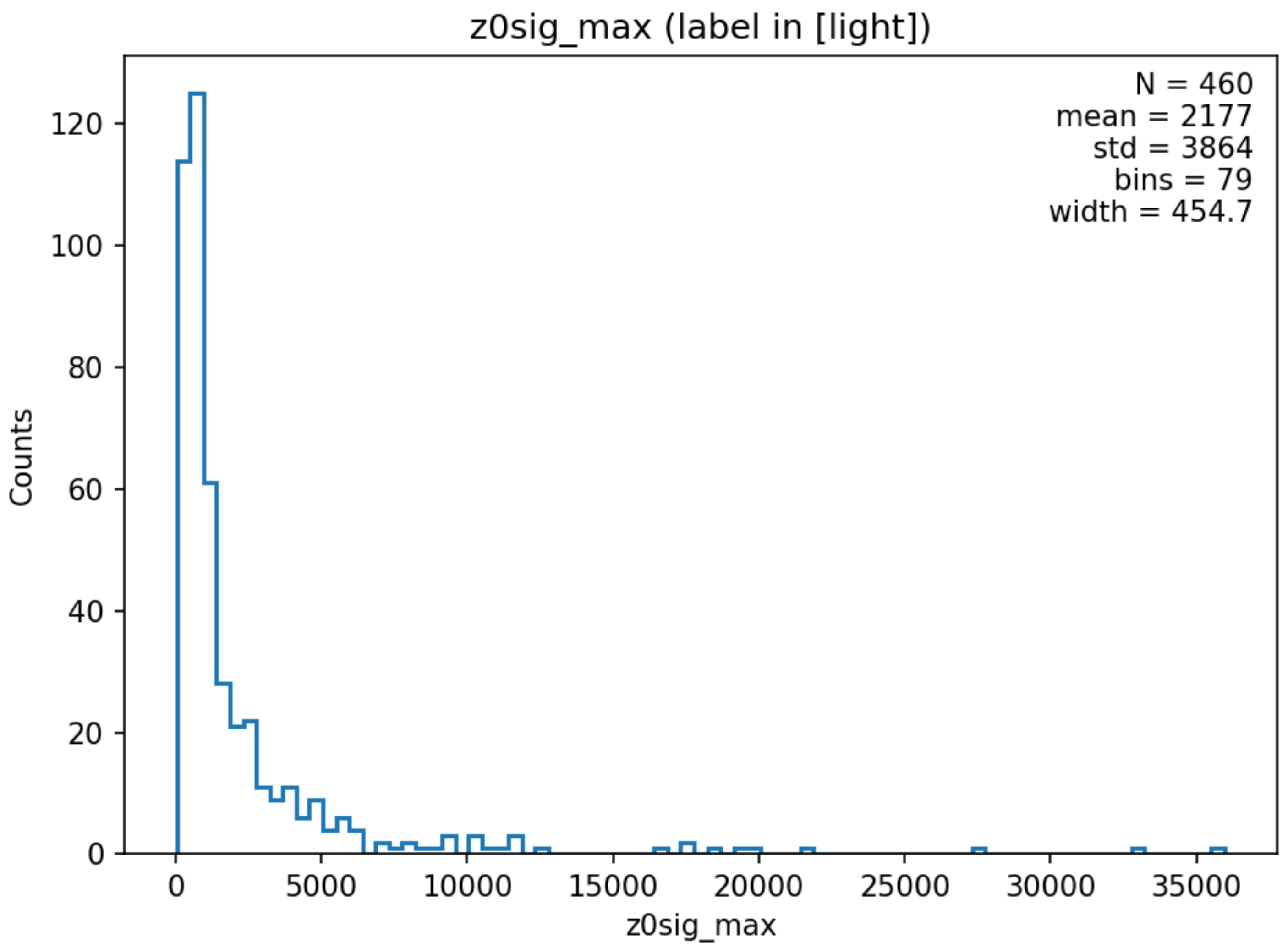
Light jet Feature Plots

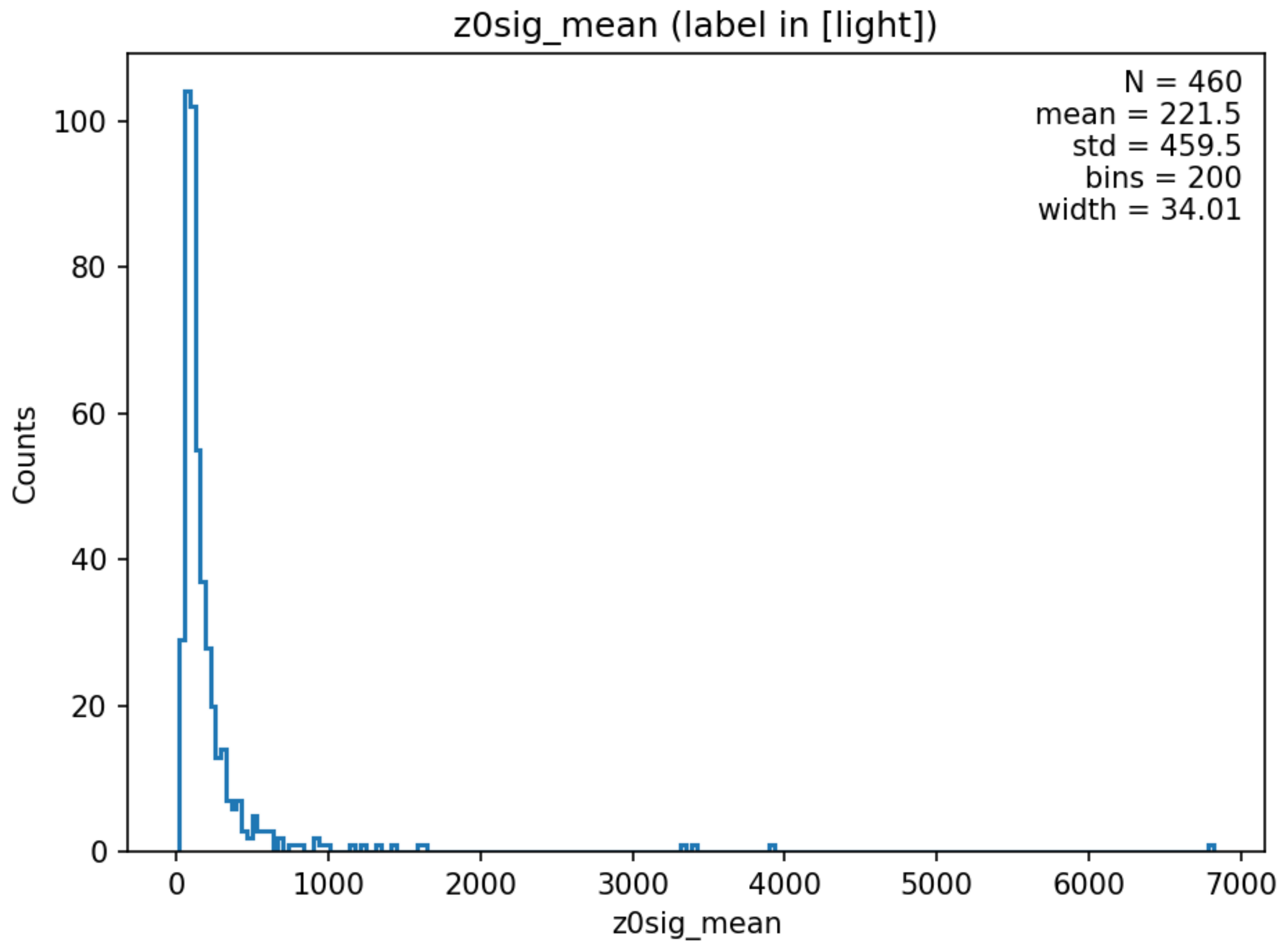
For **light jets**

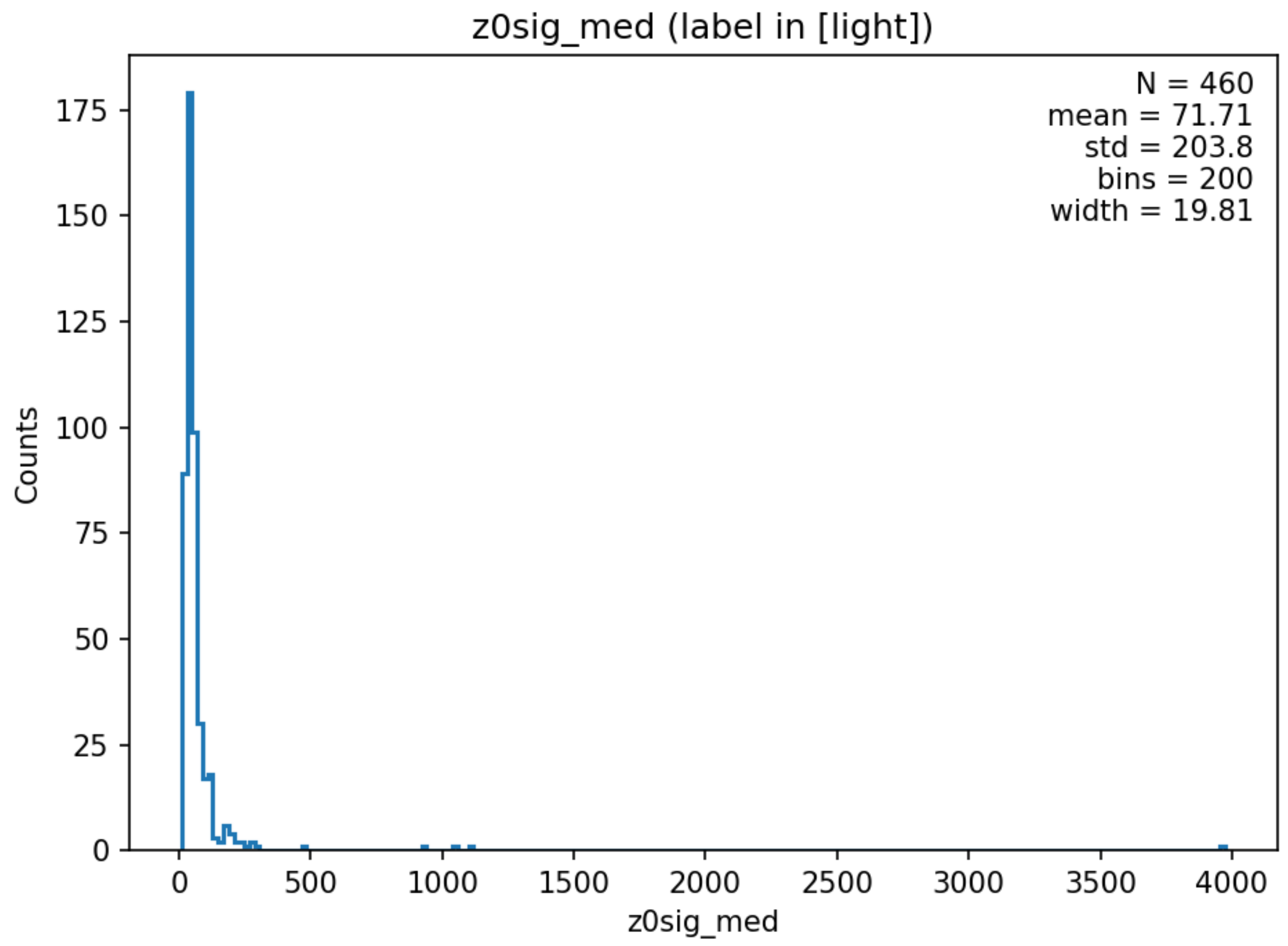










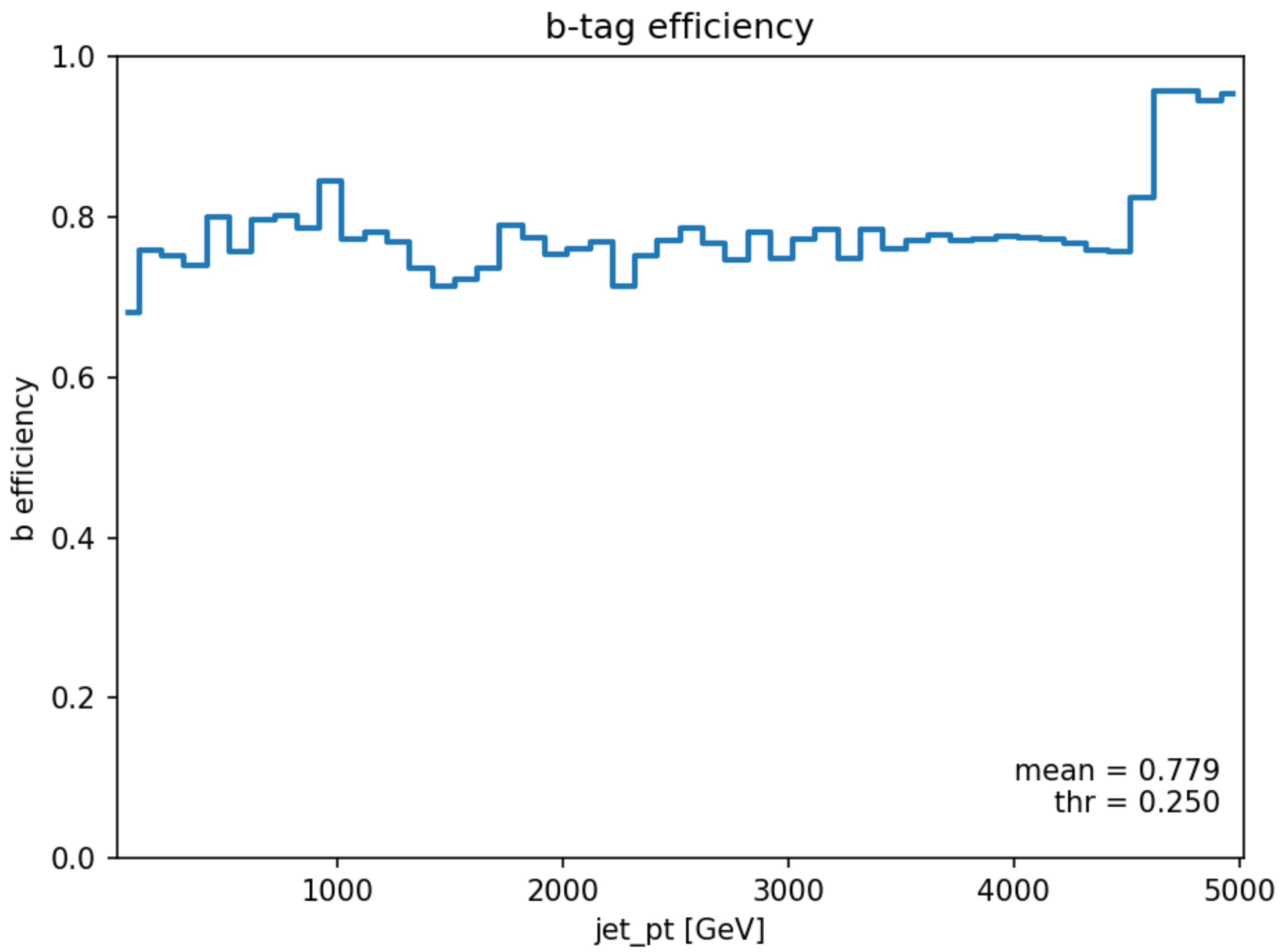


Neural Network Design

- **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 on $\mu\mu \to bb$, $\mu\mu \to cc$, $\mu\mu \to ll$ (each 100k) Pythia samples

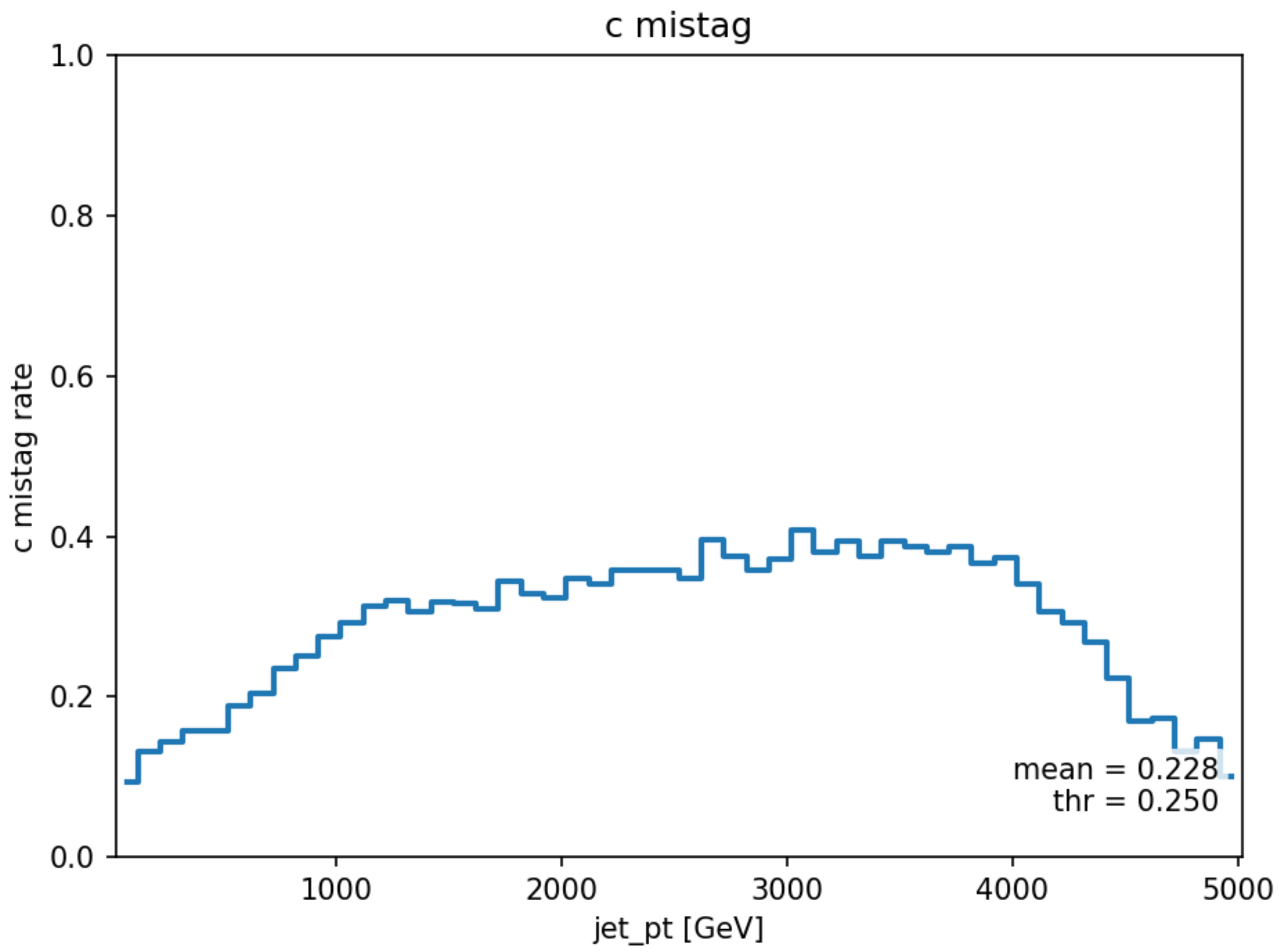
Results

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



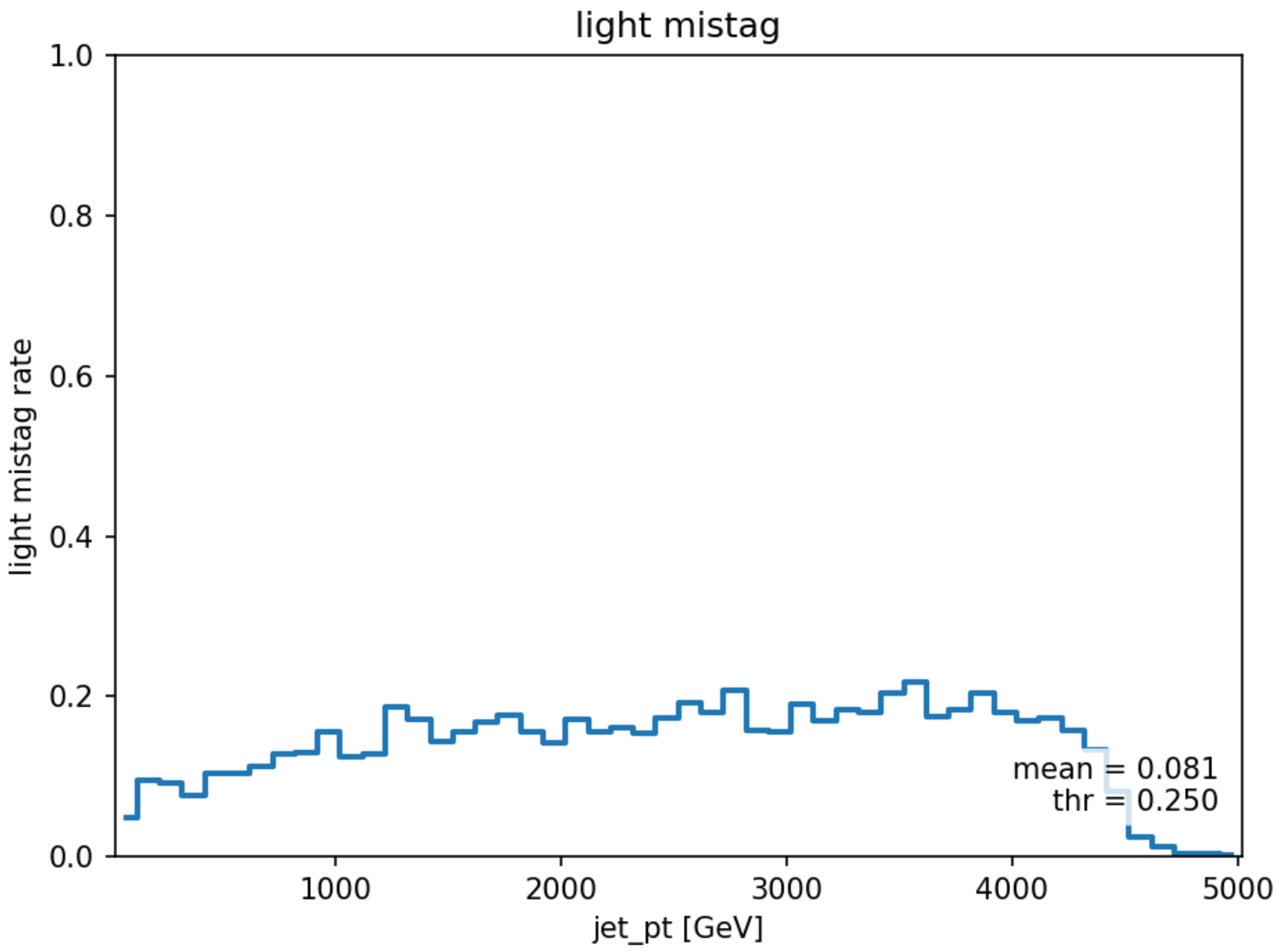
b-tag efficiency

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



c mistag

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

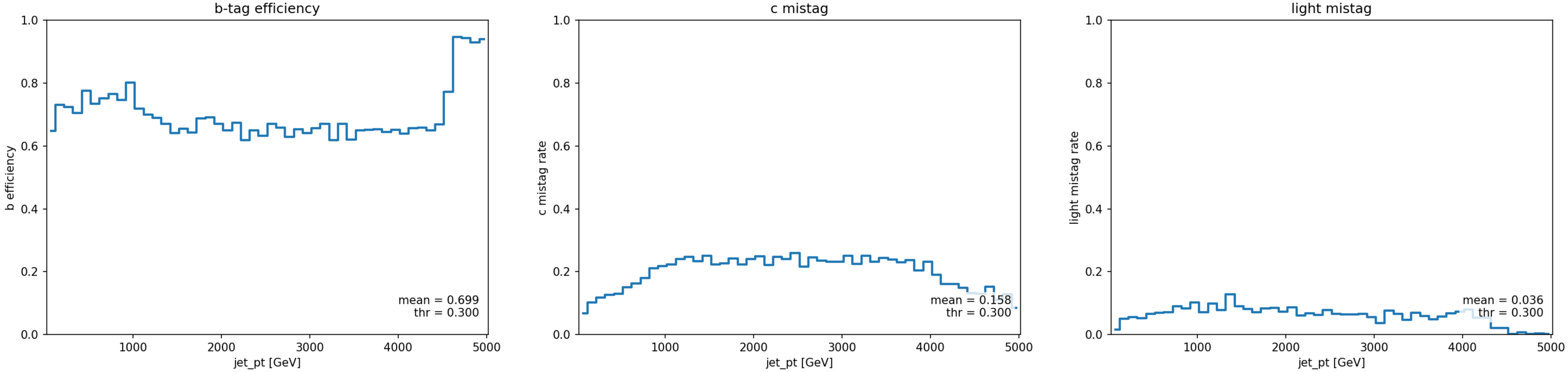


light mistag

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

Results

- Now using 0.30:



Next Steps

- **Analyze more samples**
  - Should be straightforward once I figure out why cluster keeps kicking me
- **Train on neural network**