

SCHOOL OF DATA ANALYSIS



NATIONAL RESEARCH
UNIVERSITY

Flavour Tagging Optimization

Denis Derkach^{1,3}, Tatiana Likhomanenko^{1,2,3}

¹YSDA, ²NIRC «Kurchatov Institute», ³HSE

Flavour Tagging Optimization

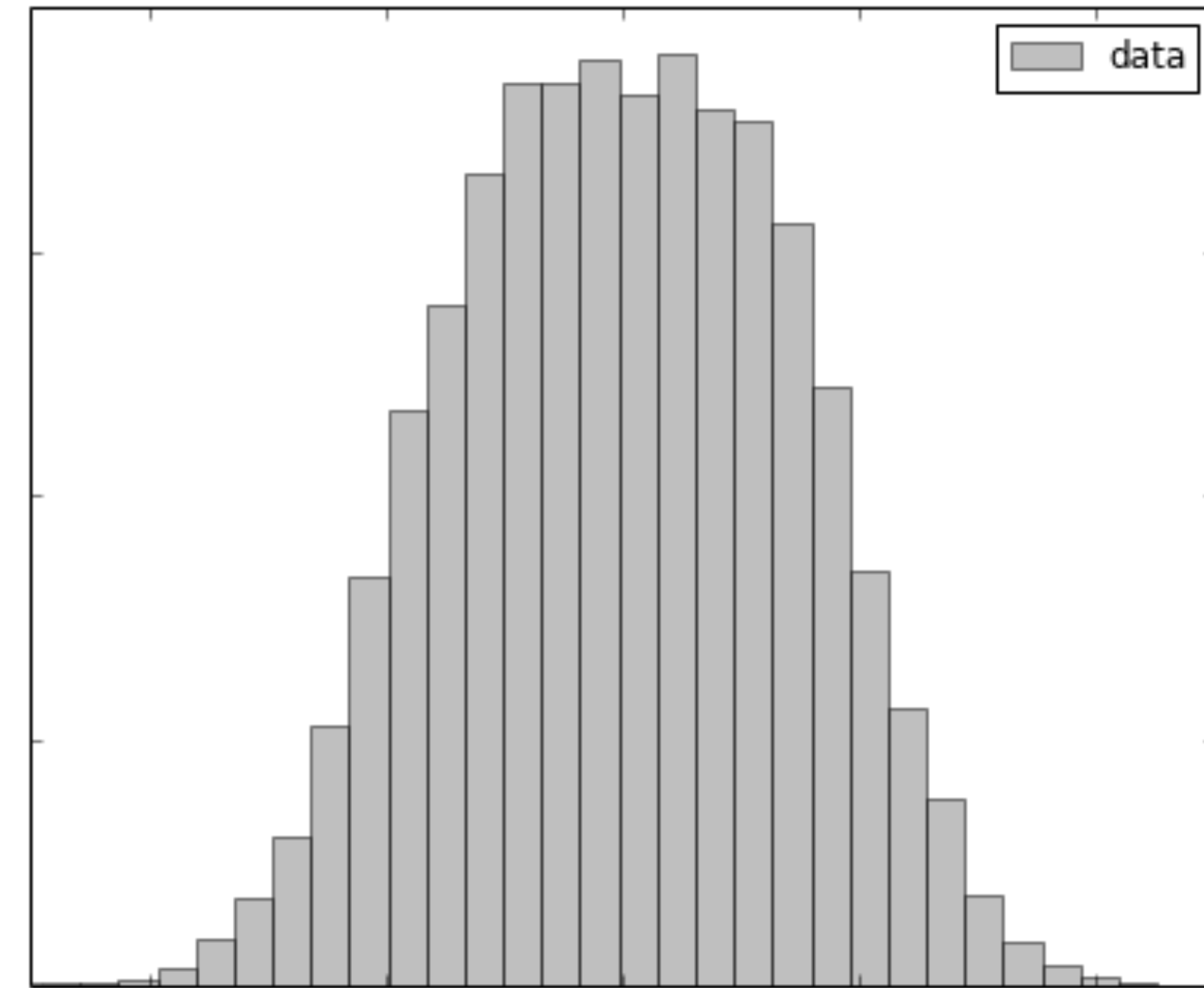
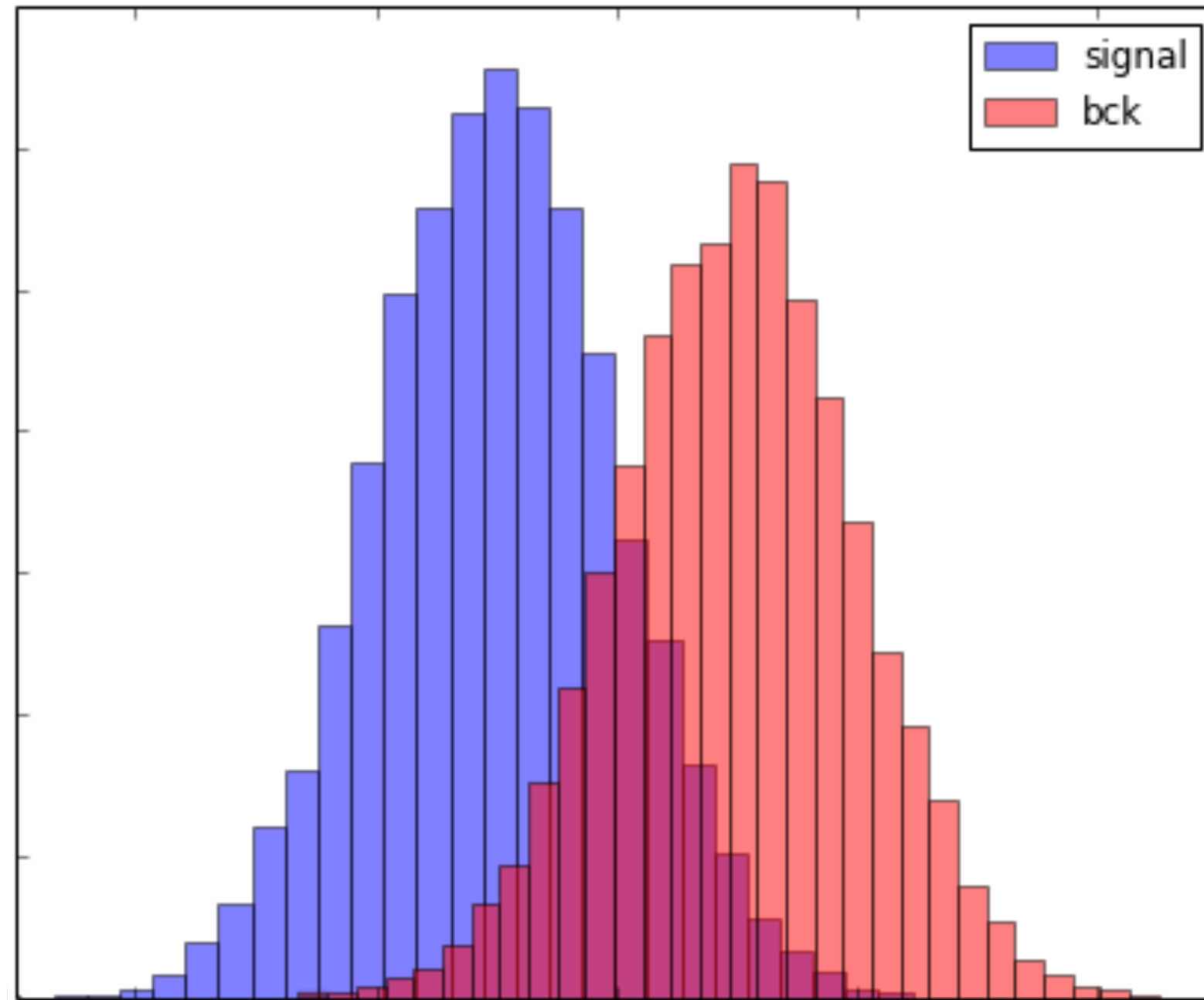
sPlot technique



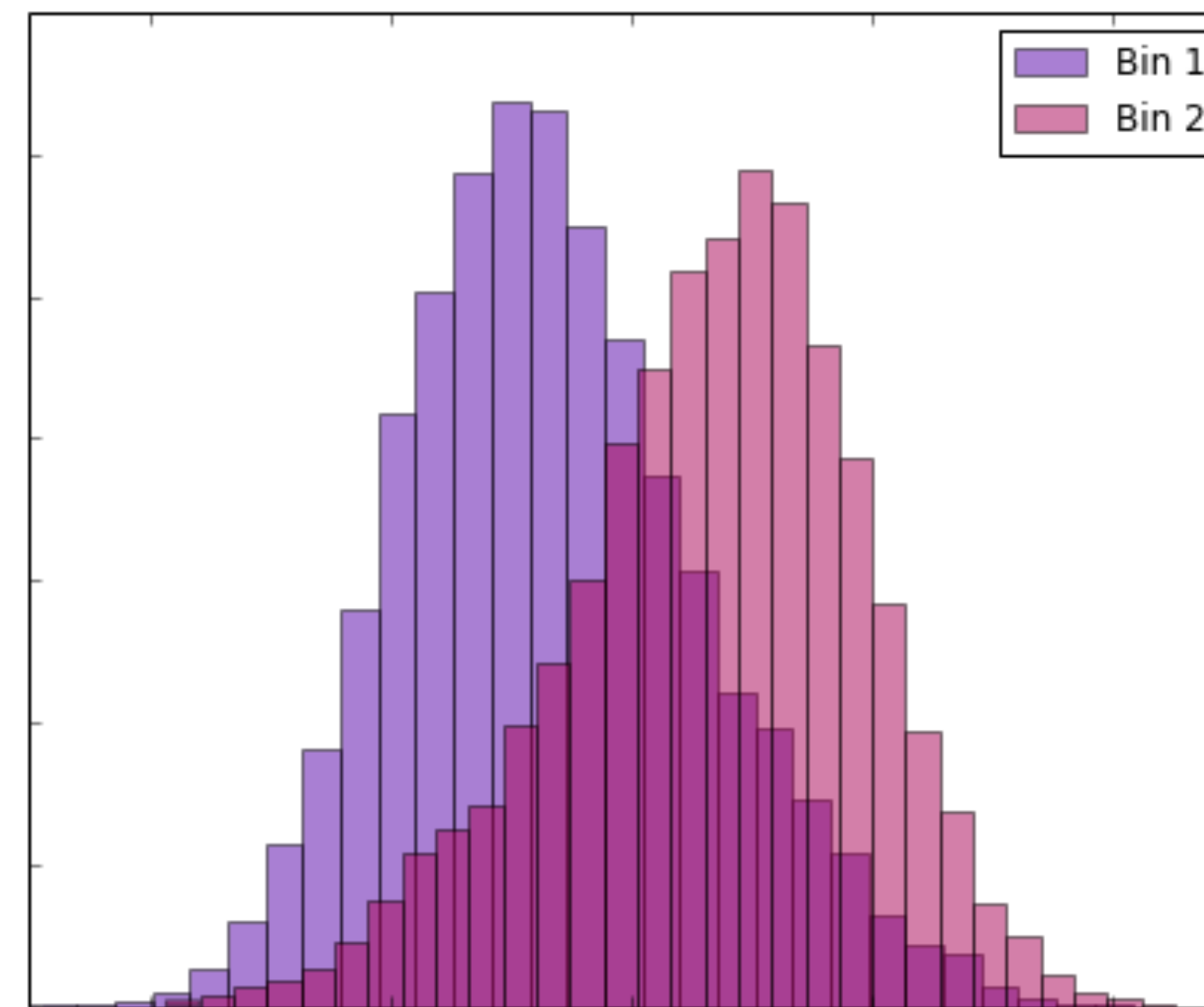
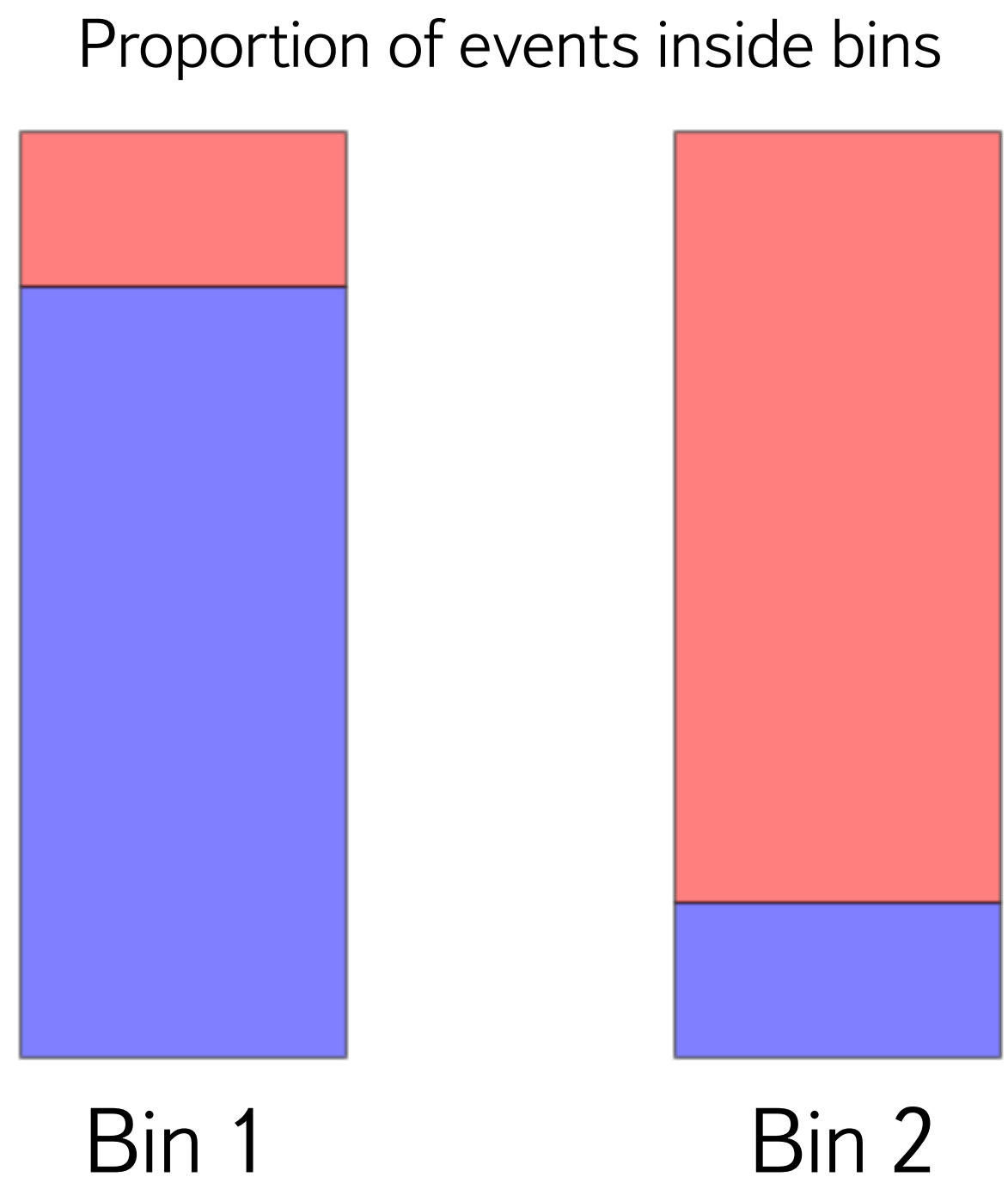
Solution for what?

- › Monte Carlo is not-well simulated
- › Need to work with real unlabeled data
- › Need somehow to label real data: want to restore for features their distributions for the signal and background data
- › Our main knowledge is the mass distribution for real data from which we can extract the mass pdfs for signal and background.
- › How to restore signal/bck pdfs for other features?

Feature initial distributions



Two mass bins



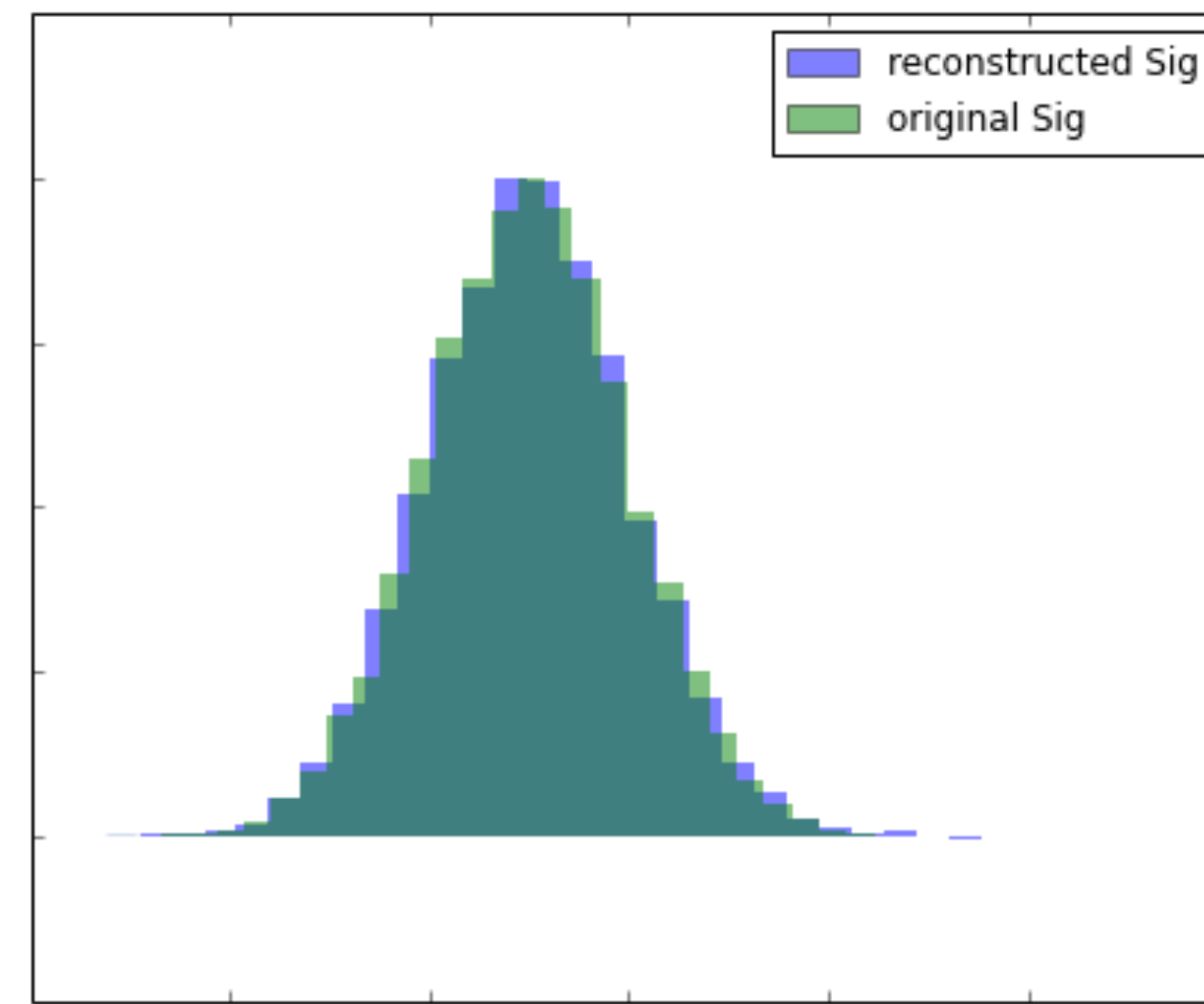
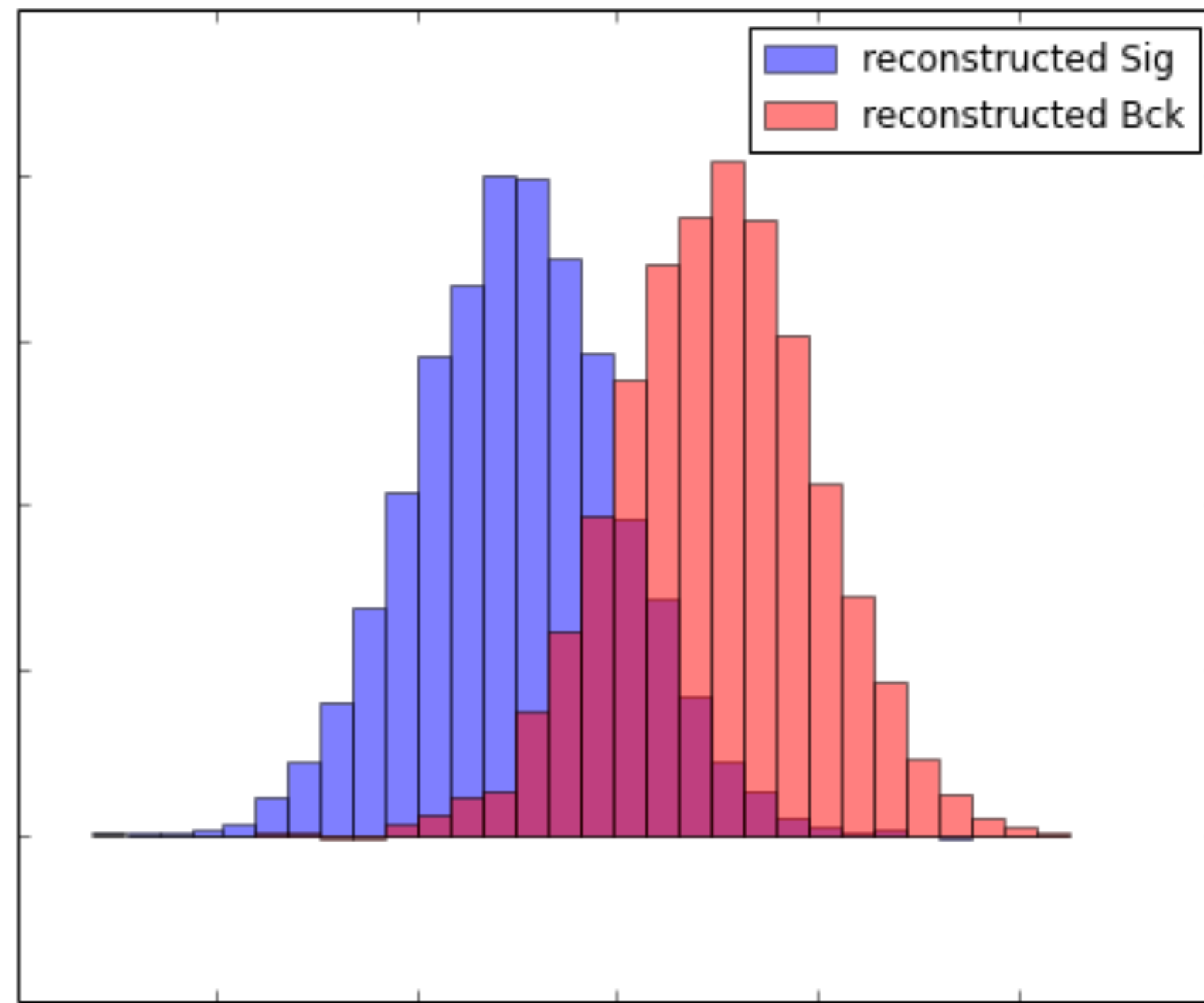
$$Bin1 : w_{b_1} f_b + w_{s_1} f_s$$

$$Bin2 : w_{b_2} f_b + w_{s_2} f_s$$

$$*w_{b_2} + \text{will obtain initial signal distribution}$$

$$*(-w_{b_1})$$

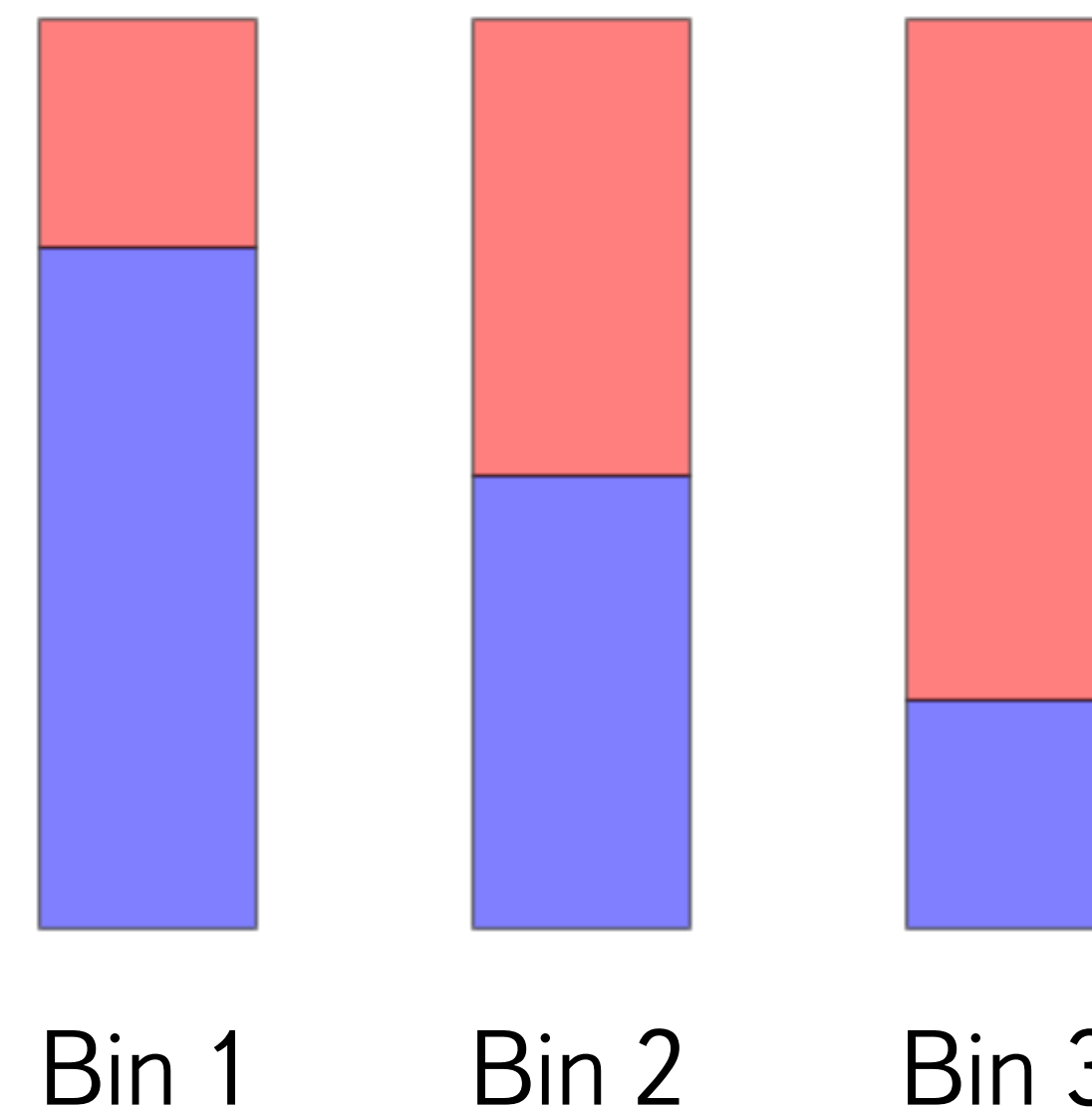
Reconstruction



More bins: sWeight

- › Equivalent to some optimization problem
- › Have explicit solution
- › Produce weights (sWeight) for each event
- › Feature pdf with sWeight will be signal pdf
- › Details for sPlot technique

Proportion of events inside bins



How to reweight?

Flavour Tagging Optimization

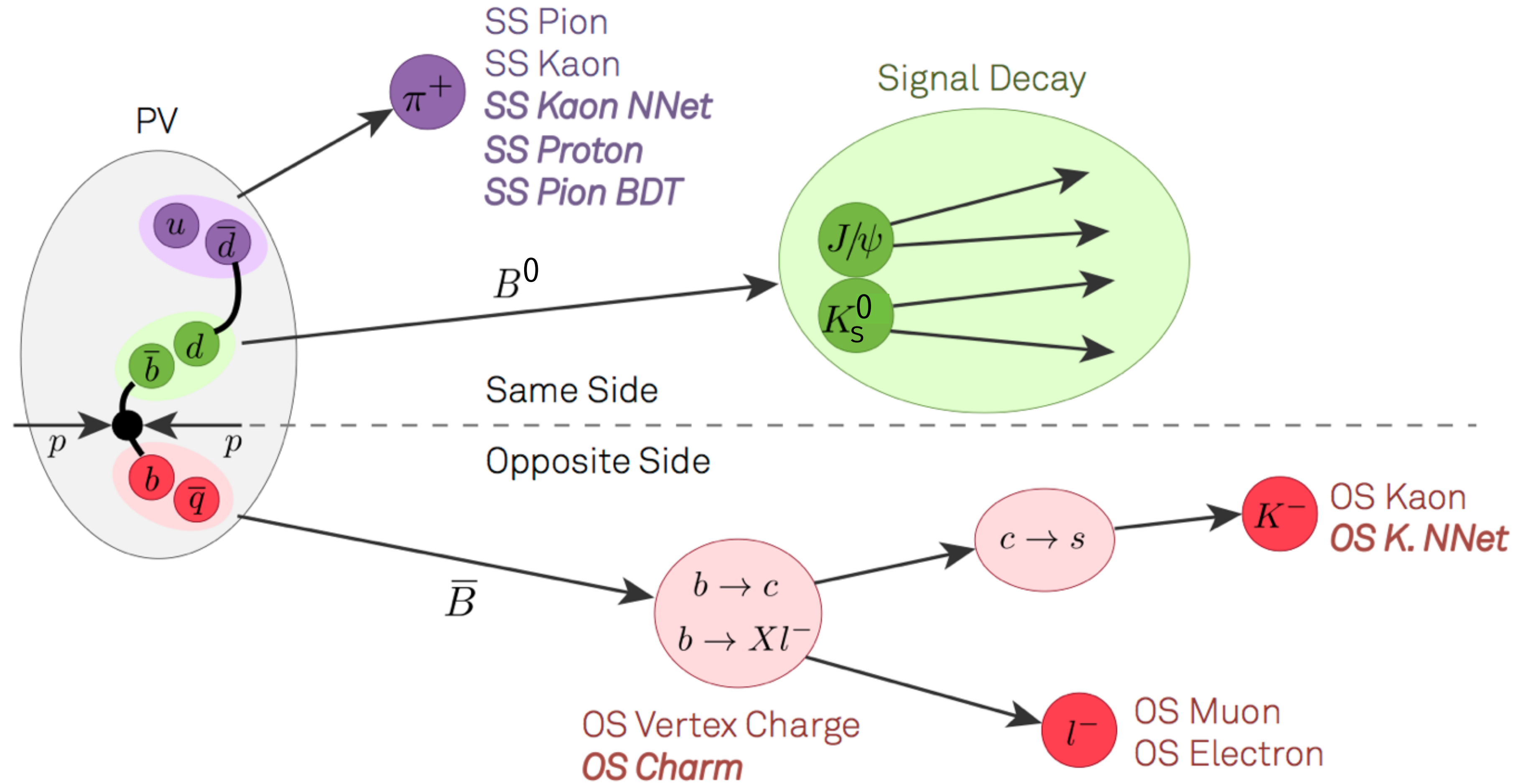
Tagging system



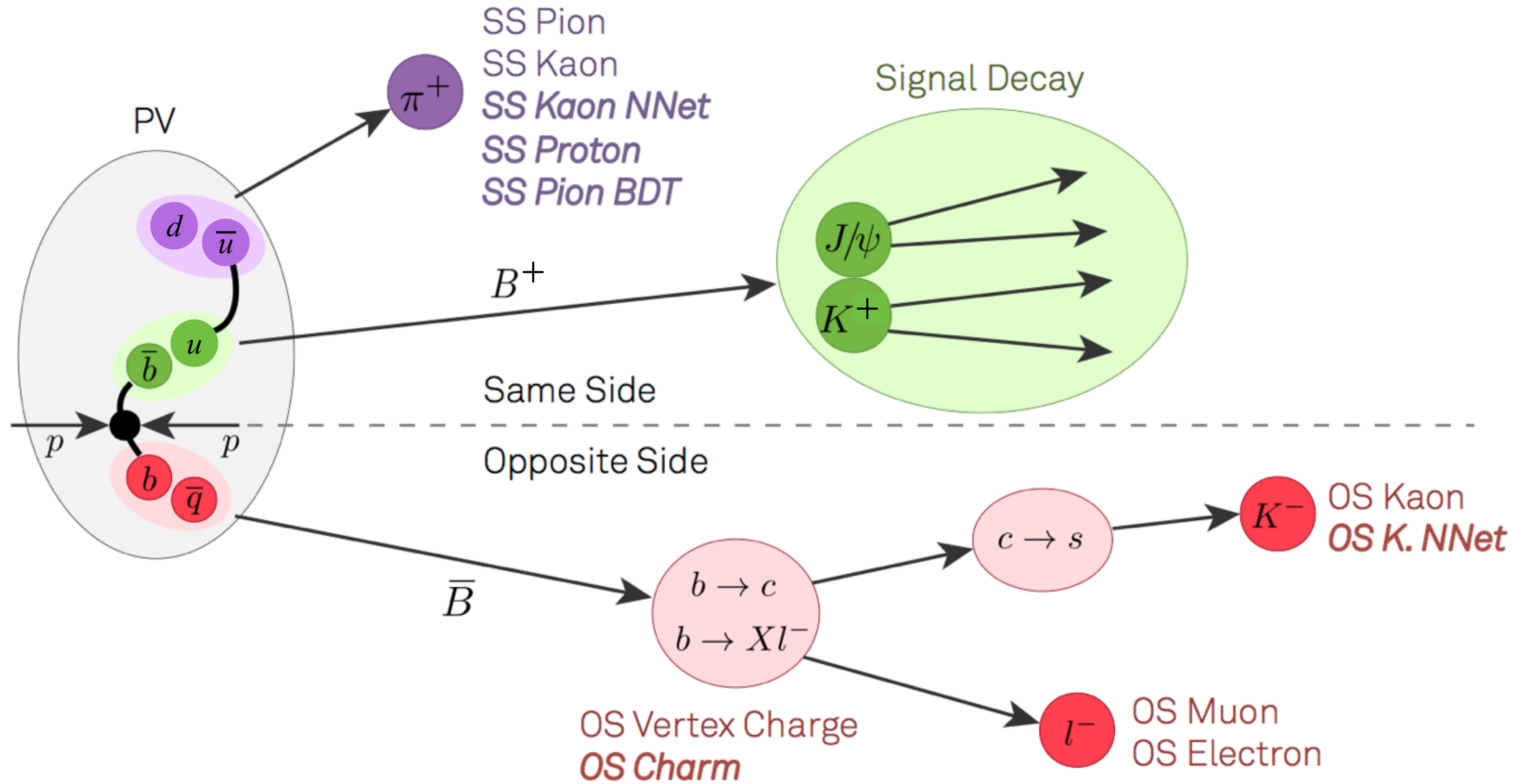
What is it?

- › Event has a signal decay part
- › The signal decay part can be produced from b quark or anti-b quark
- › The system should effectively predict the source of the signal decay (b quark or anti-b quark)
- › An intermediate B-meson in the signal decay part can oscillate
- › The tagging system prediction $P(\text{anti-b quark})$ will allow to measure the oscillation effects

Tagging particles (goal)



Tagging particles (training)



Confidence interval to asymmetry

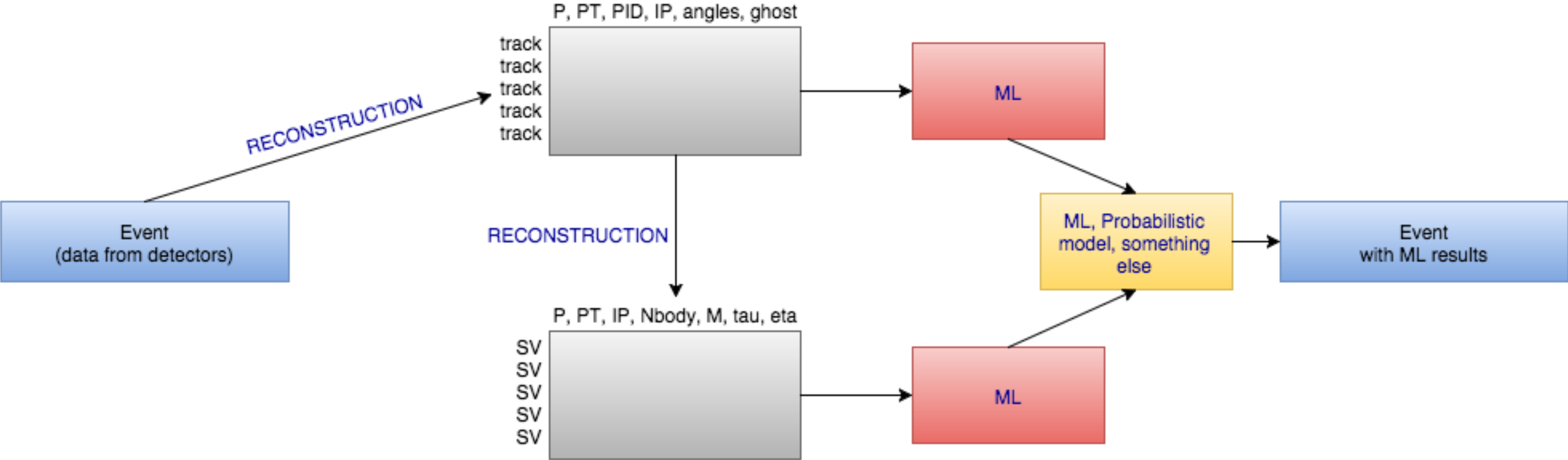
- › Construct probabilities for $b \rightarrow$ final state and anti- $b \rightarrow$ final state using $P(\text{anti-}b \text{ quark})$
- › Take relation of probabilities, called asymmetry
- › Confidence interval for parameter of interest (A_m - measured):

$$\sigma_A = \frac{\sigma_{A_m}}{1 - 2\omega} = \frac{\sqrt{1 - A_m^2}}{\sqrt{\varepsilon_{tag} N} (1 - 2\omega)}$$

- › Tagging system should maximize effective efficiency (ω - mistag probability):

$$\varepsilon_{eff} = \varepsilon_{tag} D^2 = \varepsilon_{tag} (1 - 2\omega)^2$$

Event processing



Current tagging system

- › Take $B^\pm \rightarrow J/\psi K^\pm$
- › Apply sPlot technique to real data to extract signal-like data (sWeights)
- › Choose:
 - one tagging track (PID selection and, if necessary, max PT track) for each event
 - one secondary vertex, which produces a tagging particle
- › Train exclusive taggers for SS (same side) tagging particles and OS (opposite side)
- › Target for classifier is ‘right tagged’ label
- › Combine all taggers to one (probabilistic model) to obtain $P(\text{anti-b quark})$

Flavour Tagging Optimization

Inclusive tagging system (not official)



Inclusive tagging system

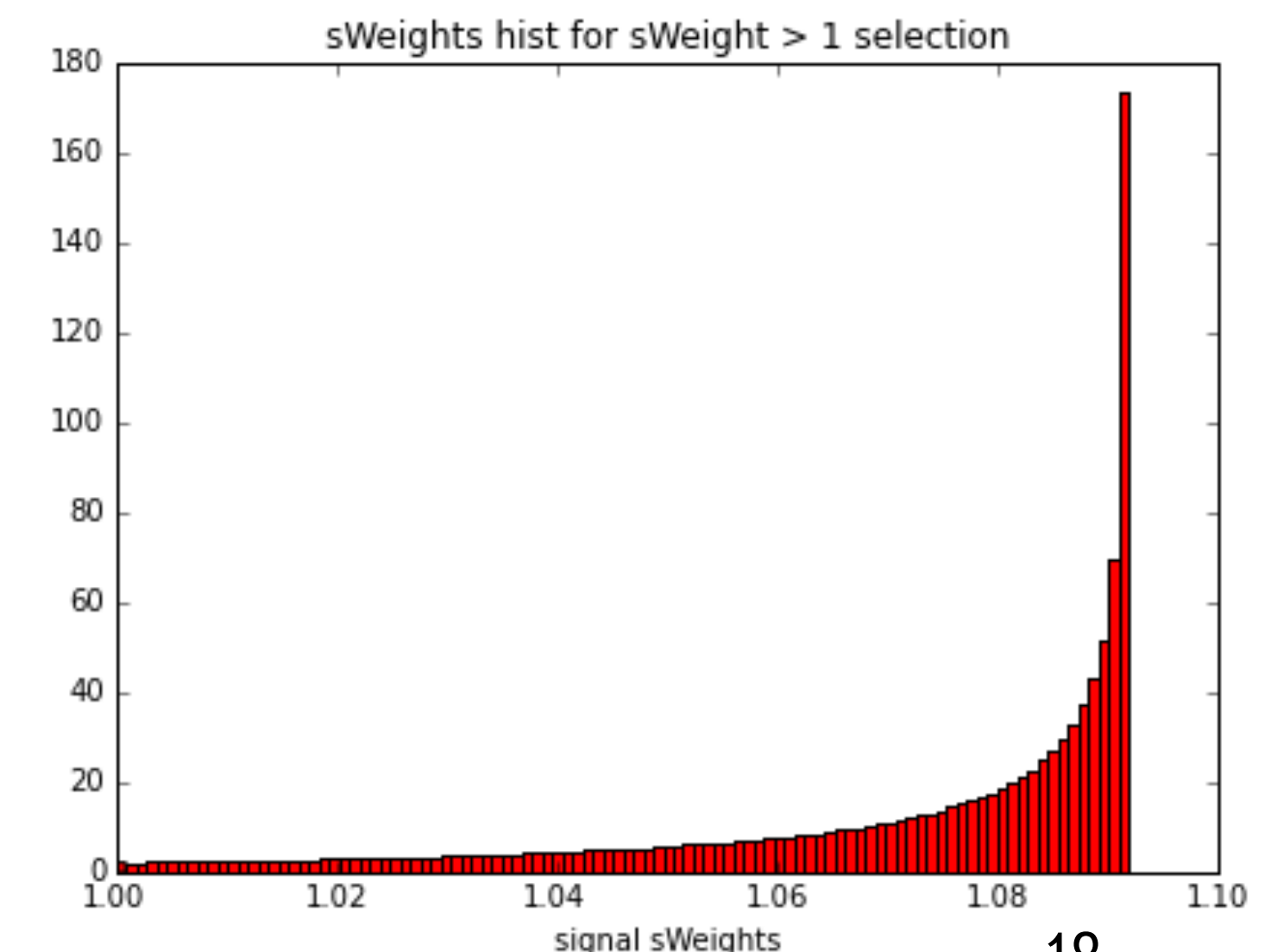
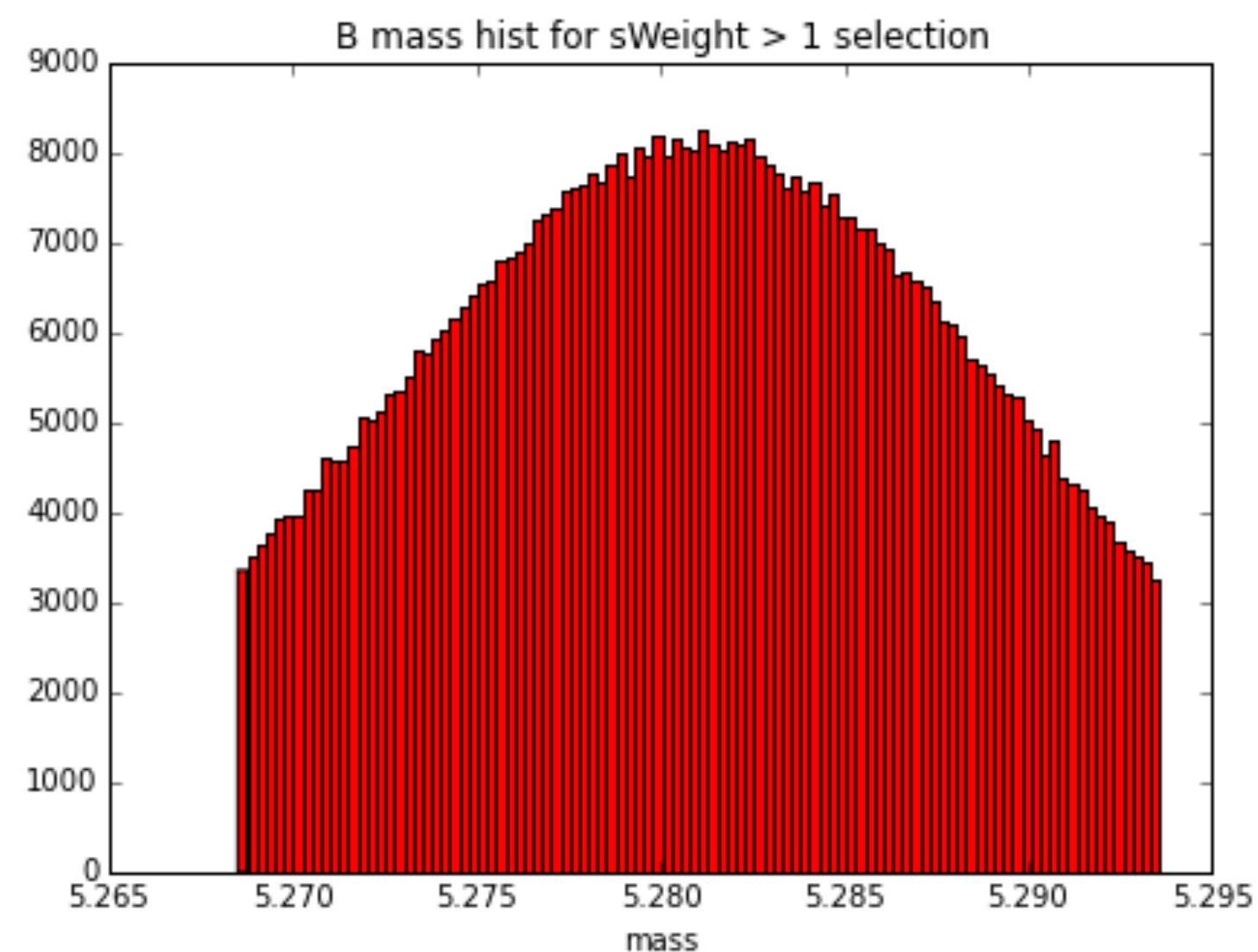
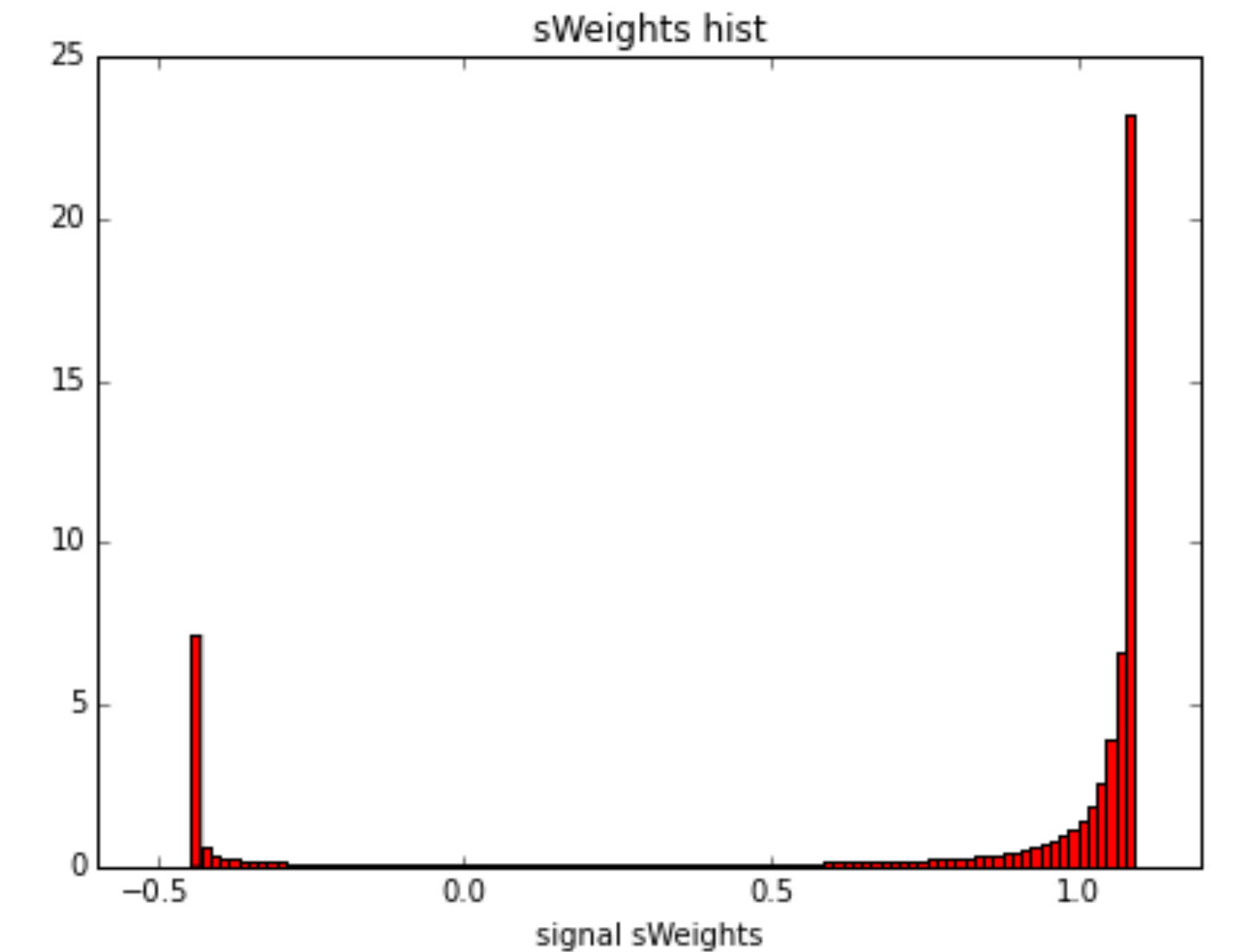
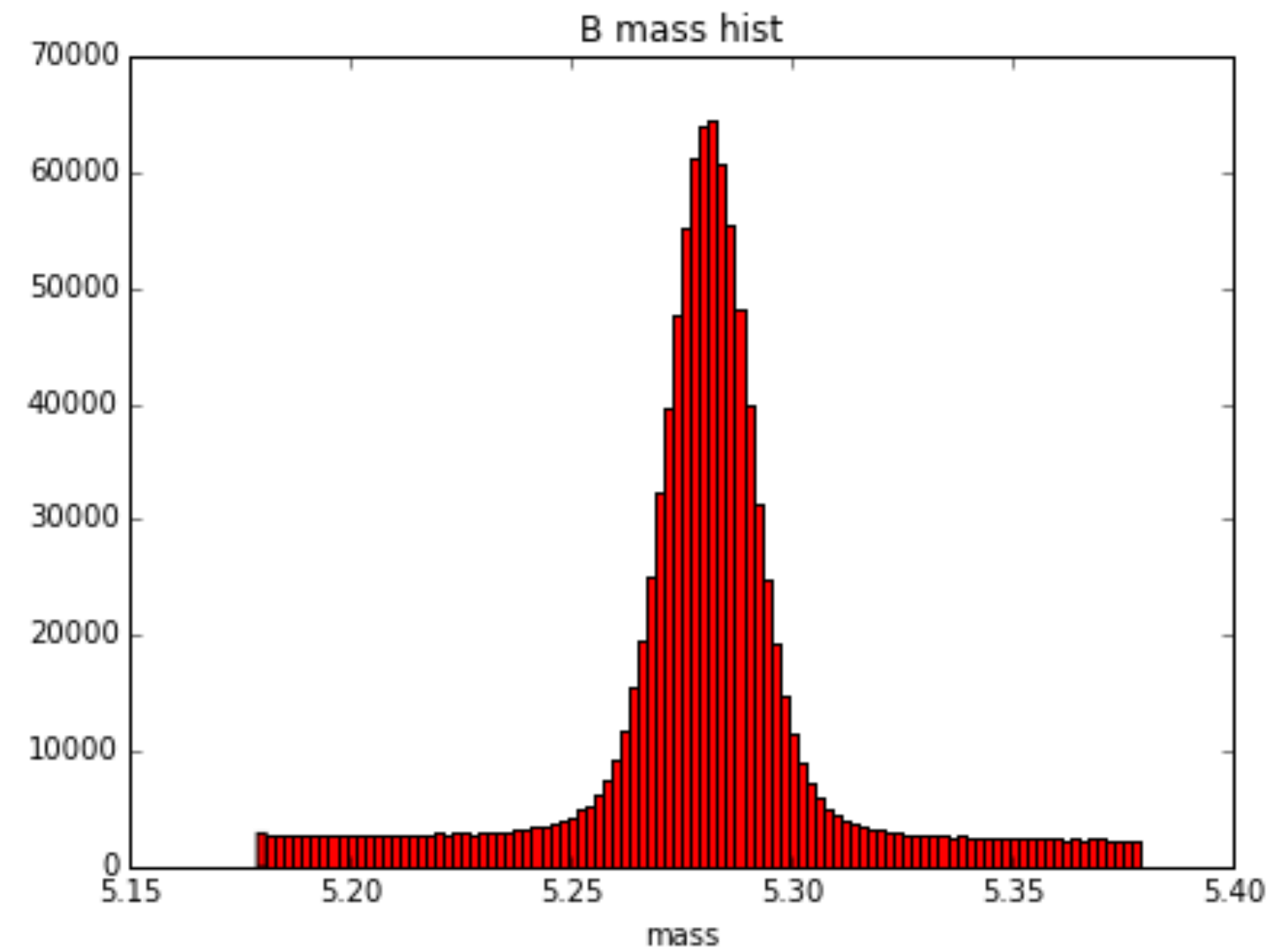
- › Don't apply physical selections, except loose cuts on PID and track ghost
- › For each event use full topological information: all tracks and possible vertices
- › Use probabilistic model for tracks and vertices to obtain $P(\text{anti-}b \text{ quark})$
- › Maximize ROC AUC score

Restrictions

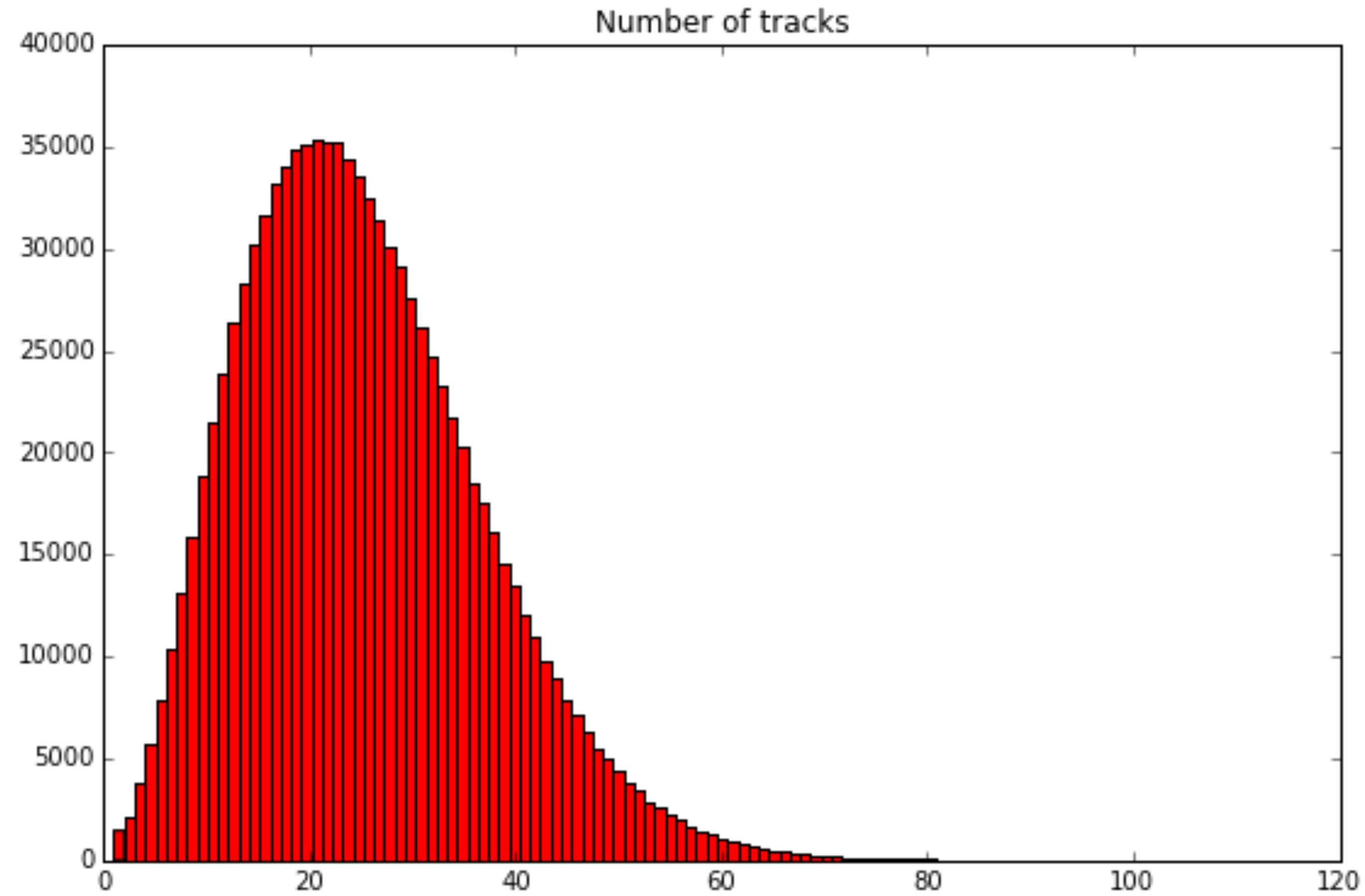
- › Should check that effective efficiency is maximized
- › Necessary to calibrate output to probability
- › Distribution should be symmetric for b-quark (B^-) and anti-b quark (B^+)
- › Flatness for B-mass, life time, life time error, momentum, transverse momentum (to simplify further analyses)

Training sample

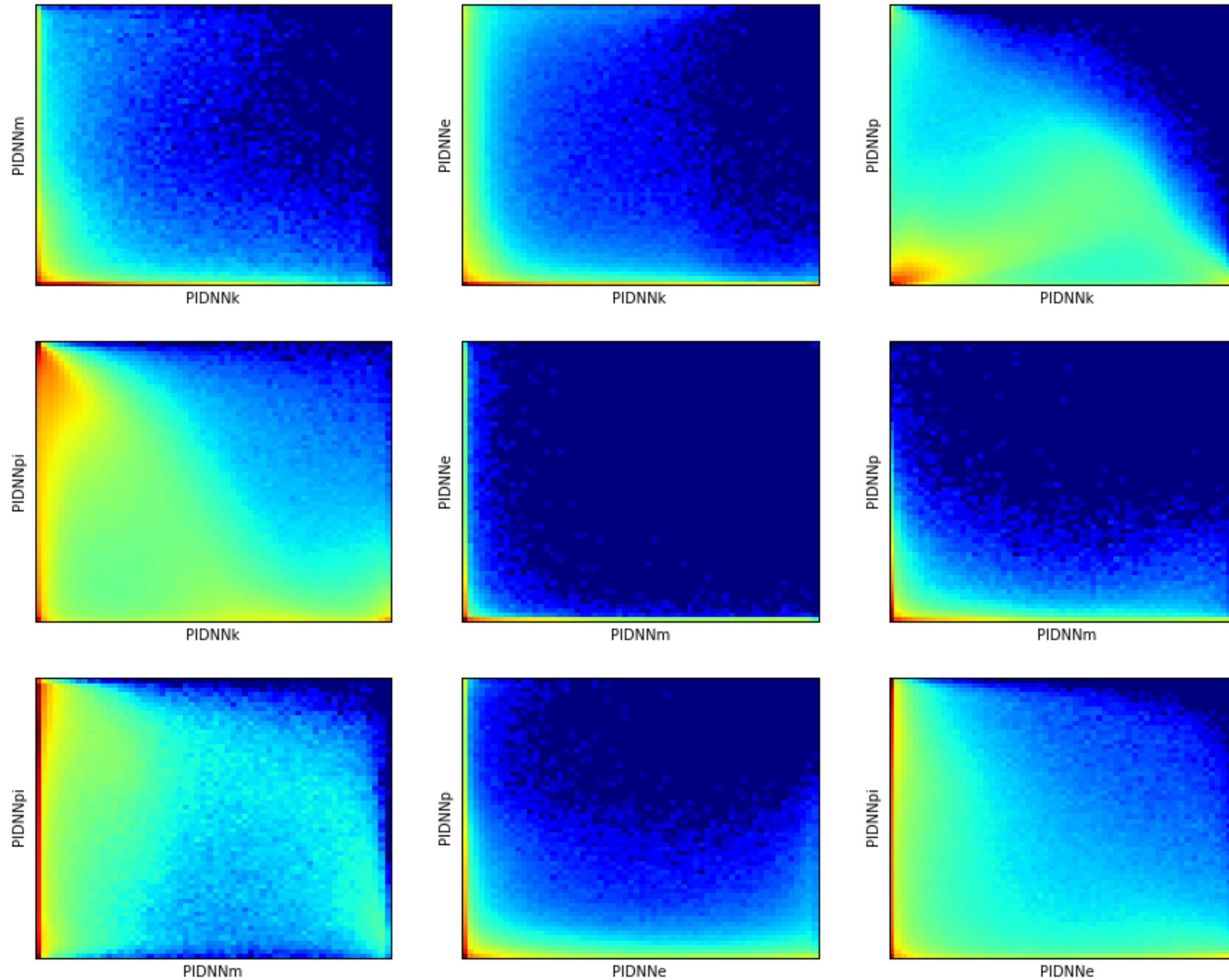
- › Take all events
- › Take events with $sWeight > 0$
- › Take events with $sWeight > 1$



Number of tracks



PID distributions



Flavour Tagging Optimization

Probabilistic model



Probabilistic model for events

$$\frac{P(B^+)}{P(B^-)} = \prod_{\text{track, vertex}} \frac{P(B^+ | \text{track/vertex})}{P(B^- | \text{track/vertex})}$$

How to compute conditional probabilities?

Assumptions

- › Assume that

$$\begin{aligned} &P(\text{track/vertex same sign as B| B sign}) = \\ &= P(\text{B same sign as track/vertex| track/vertex sign}) \end{aligned}$$

- › Classifier should reconstruct

$$P(\text{track/vertex same sign as B| B sign})$$

- › Target for classifier:

$$\text{sign B} * \text{sign track} > 0 \quad \text{or} \quad \text{sign B} * \text{sign vertex} > 0$$

Probabilistic model for events

$$\frac{P(B^+)}{P(B^-)} = \prod_{track, vertex} \frac{P(track/vertex|B^+)}{P(track/vertex|B^-)} = \alpha$$
$$\Rightarrow P(B^+) = \frac{\alpha}{1 + \alpha}, \quad [1]$$

where

$$\frac{P(B^+)}{P(B^-)} = \prod_{track, vertex} \begin{cases} \frac{P(track/vertex \text{ same sign as } B|B)}{P(track/vertex \text{ opposite sign as } B|B)}, & \text{if } track/vertex^+ \\ \frac{P(track/vertex \text{ opposite sign as } B|B)}{P(track/vertex \text{ same sign as } B|B)}, & \text{if } track/vertex^- \end{cases}$$

$$p_{mistag} = \min(p(B^+), p(B^-))$$

Flavour Tagging Optimization

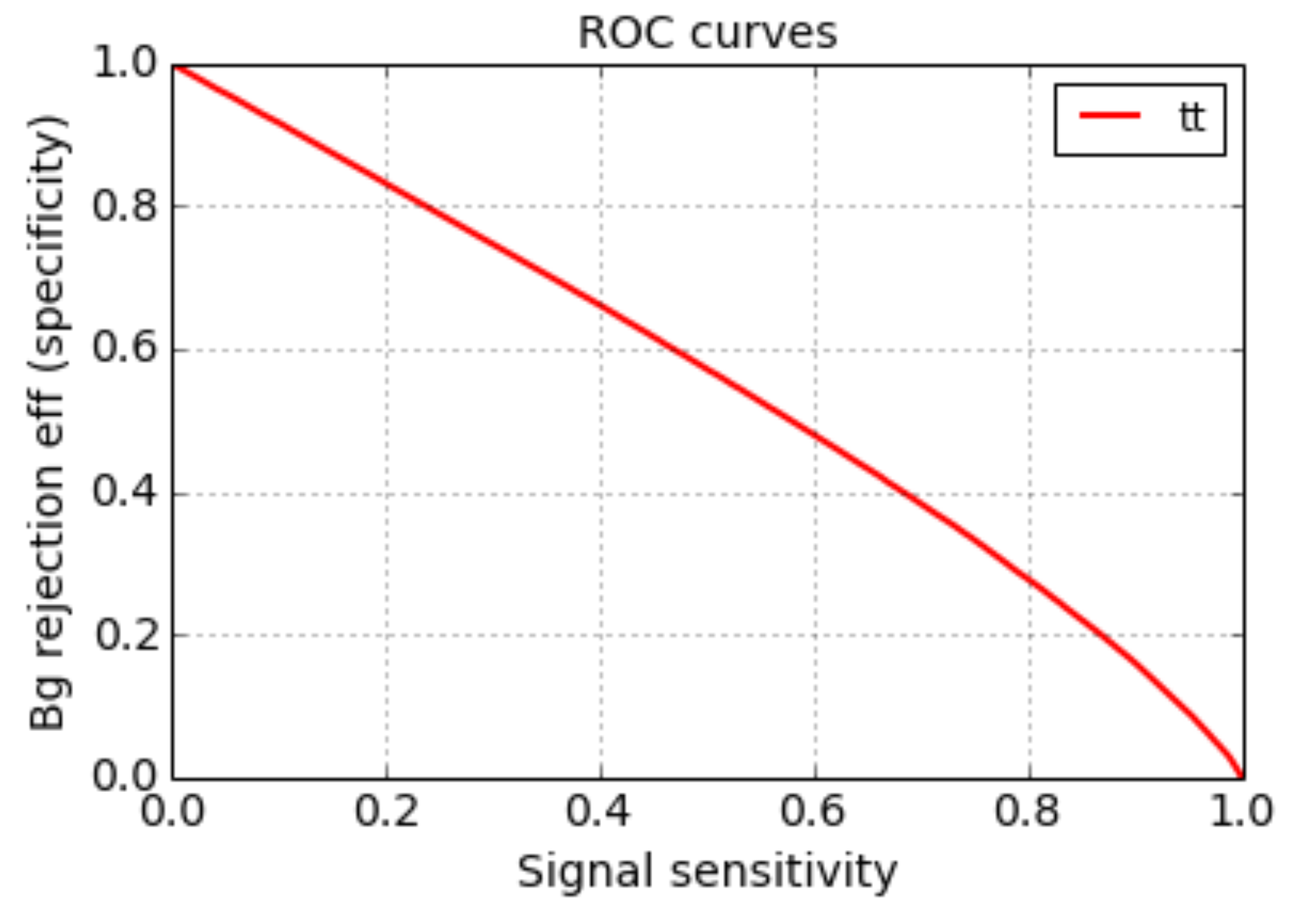
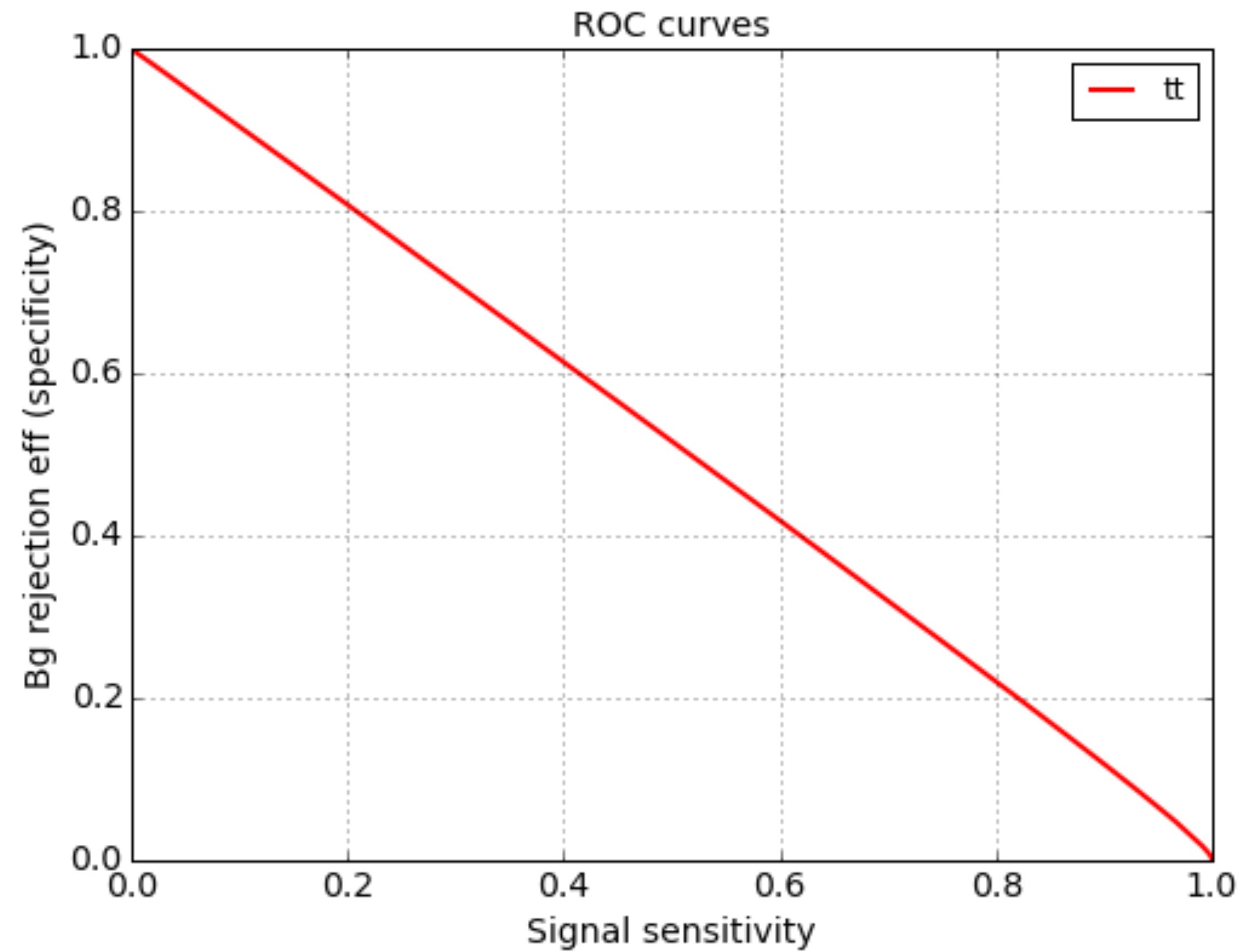
Training



Tracks and vertices training

- › Training on tracks: AUC 0.5134
- › Training on vertices: AUC 0.5544 (only one vertex for an event)
- › Classifier output is not strictly probability
- › Should calibrate output to probability

Tracks and vertices training: ROCs



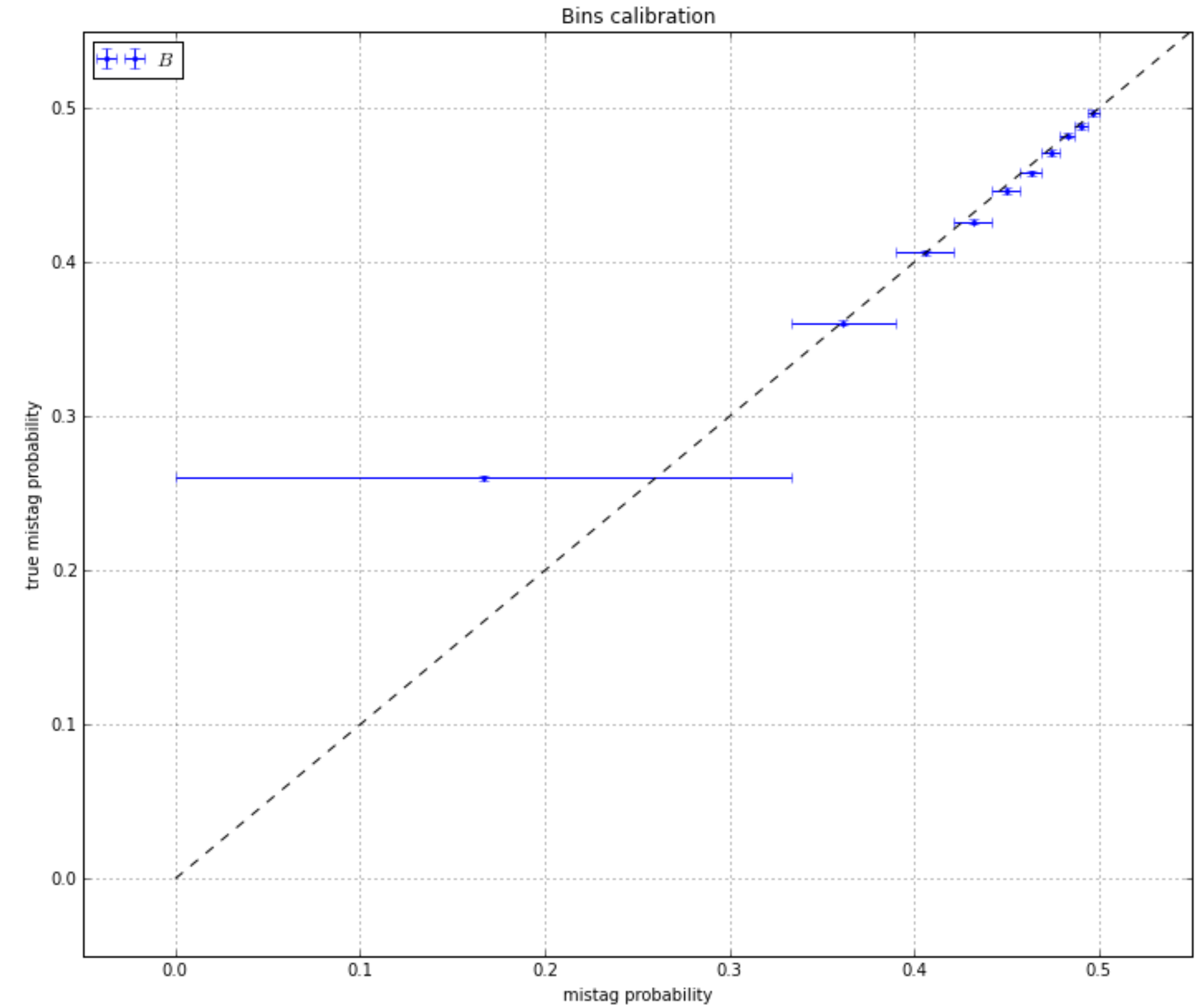
Flavour Tagging Optimization

Calibration

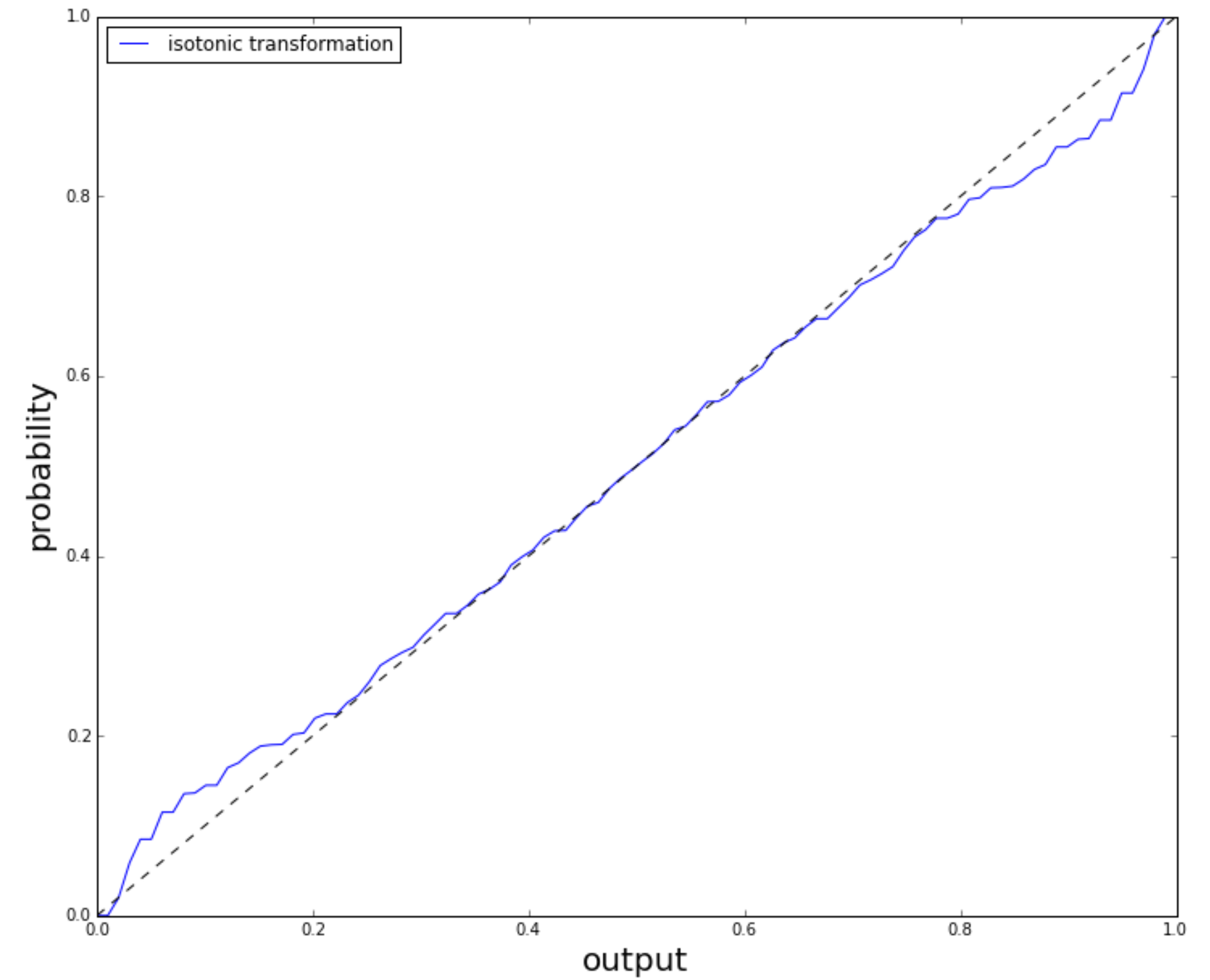
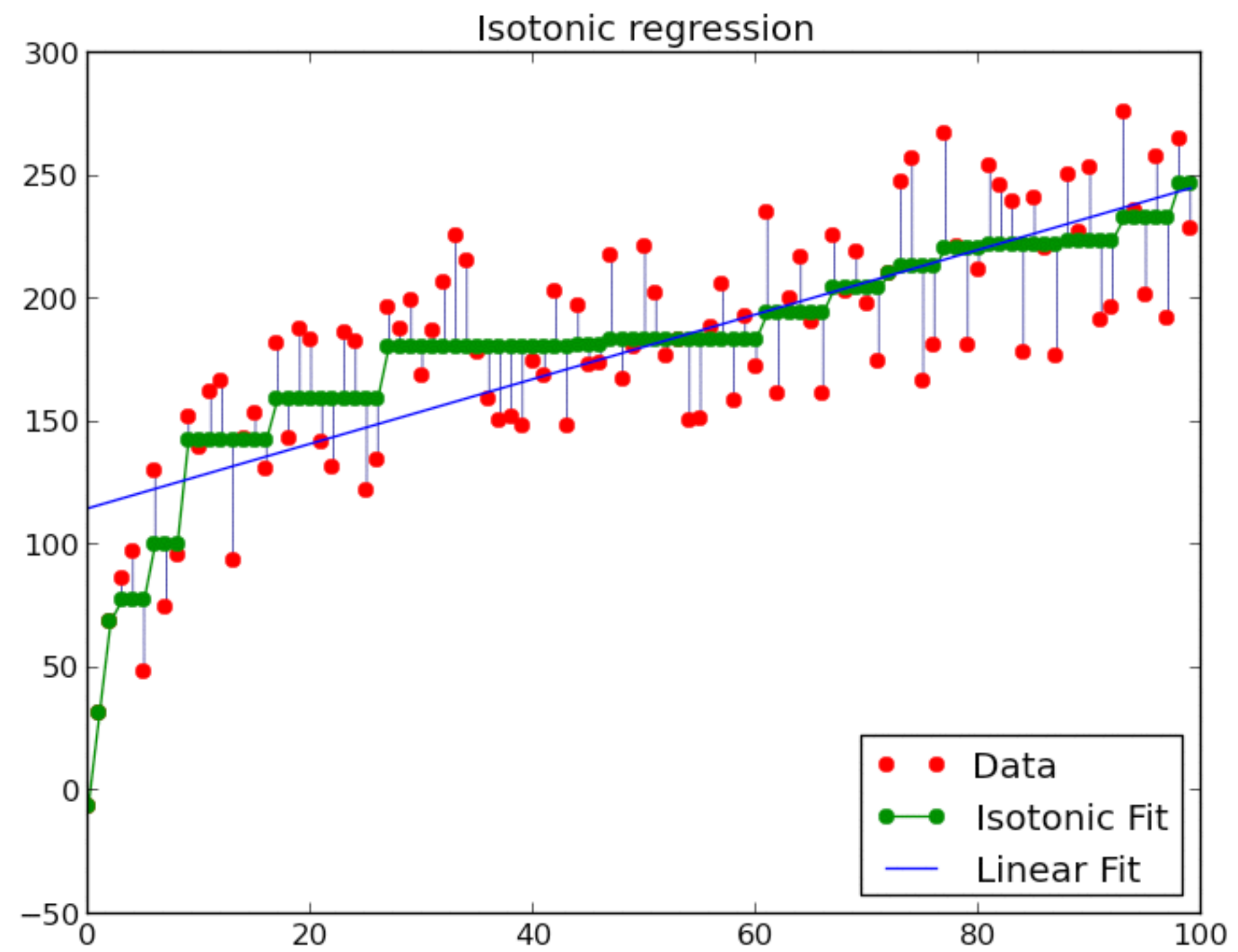


Calibration

- › Platt's calibration: logistic regression over the classifier output
- › Bin method:
 - $\#(\text{same sign in bin}) / \#(\text{in bin})$
 - fit by linear function
- › Isotonic regression: monotonic function (extends the bin method)



Isotonic calibration



Track and vertex calibration

- › Logistic calibration for tracks
- › Isotonic calibration for vertices
- › Compare with other possibilities
- › Calibration influences on ROC a little bit
- › Calibration is not necessary at this stage

Flavour Tagging Optimization

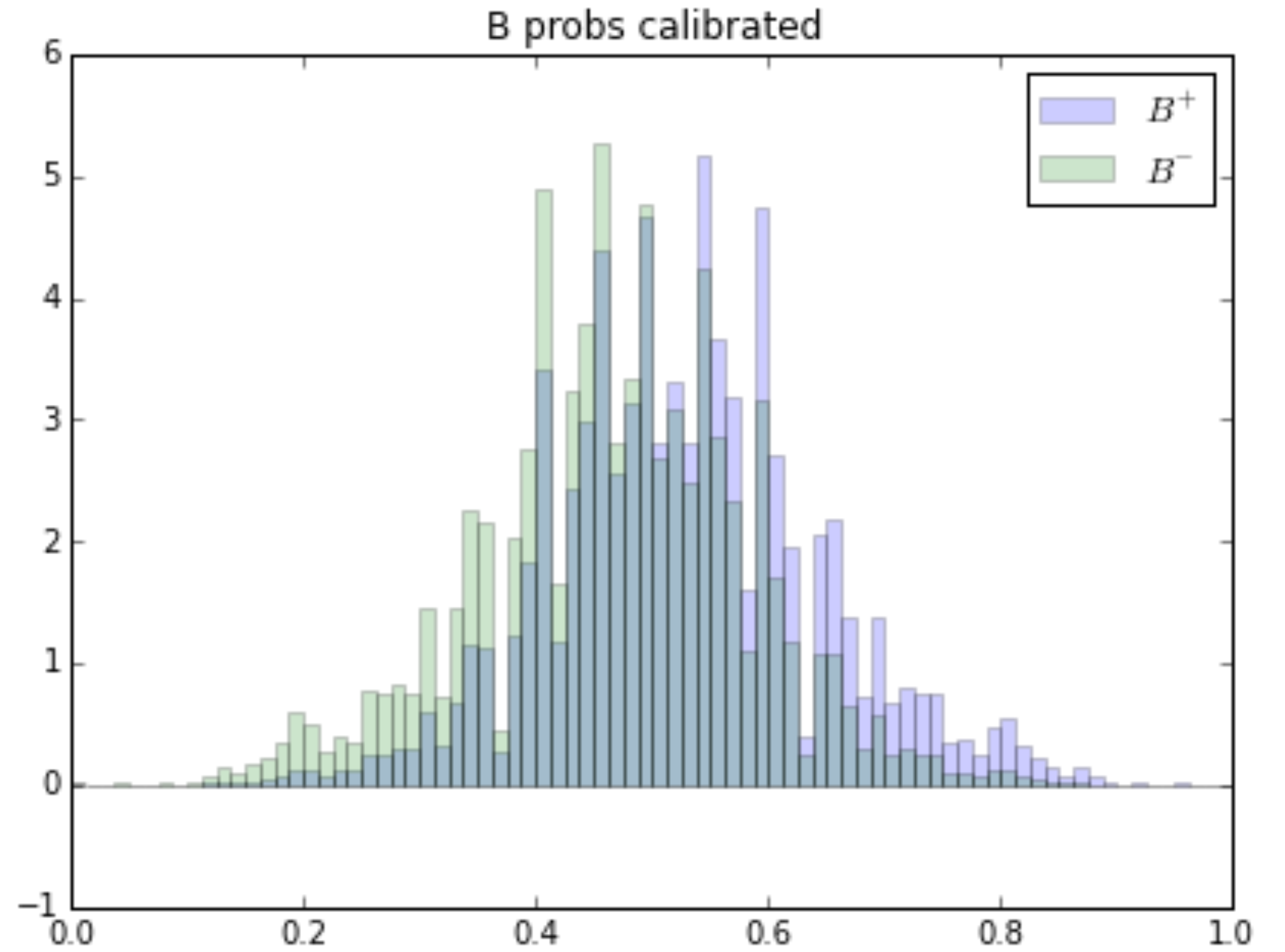
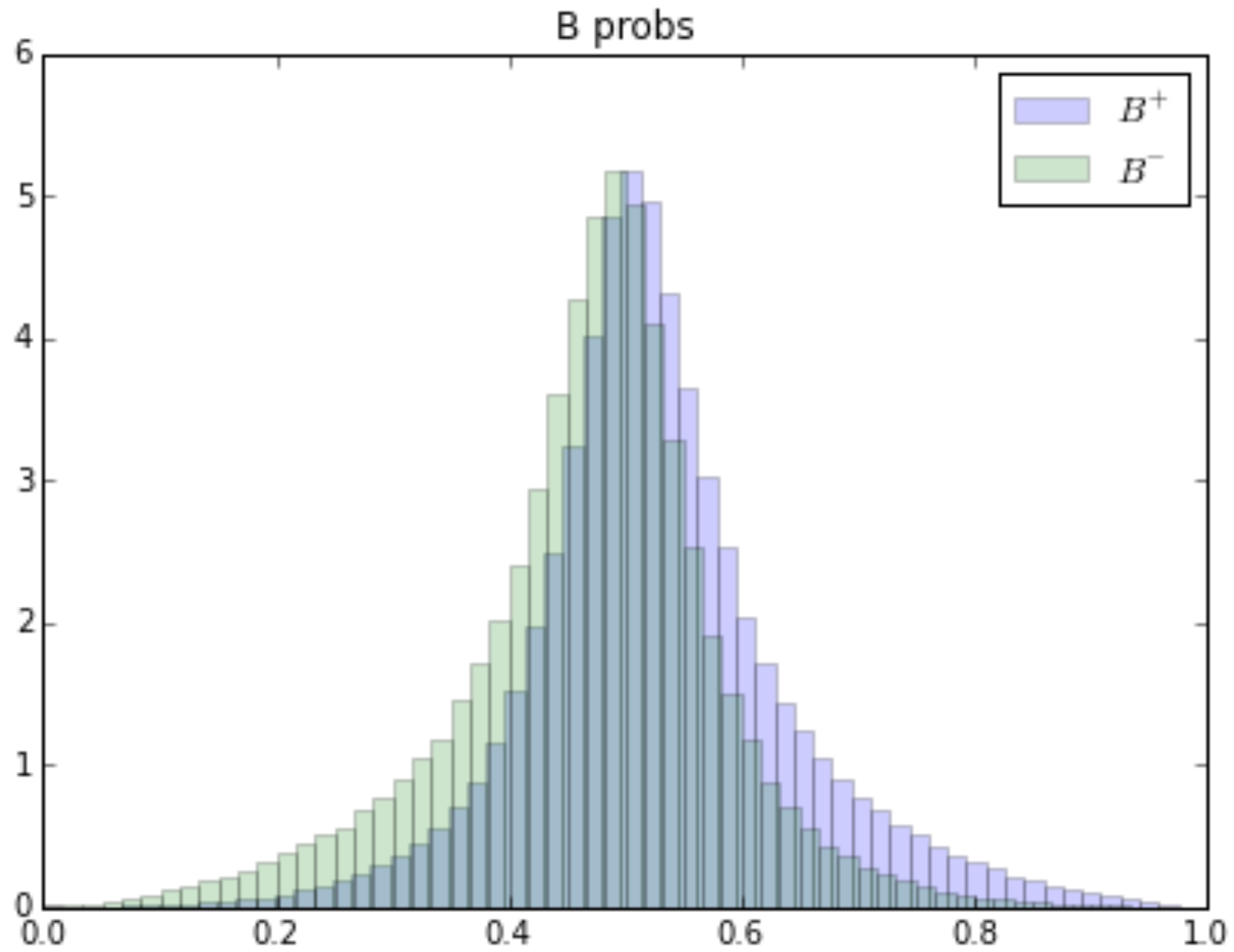
$P(\text{anti-b quark})$ asymmetry



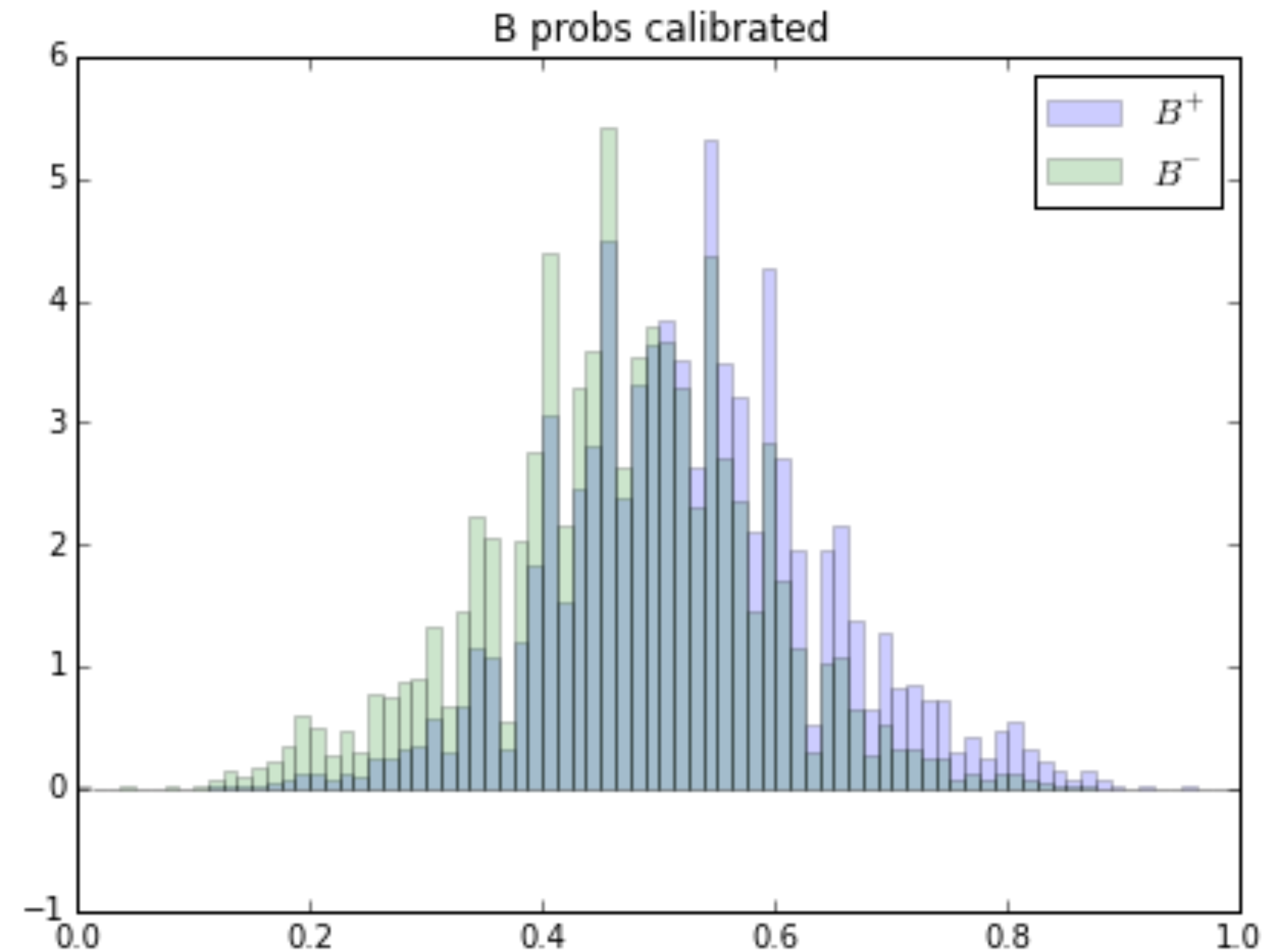
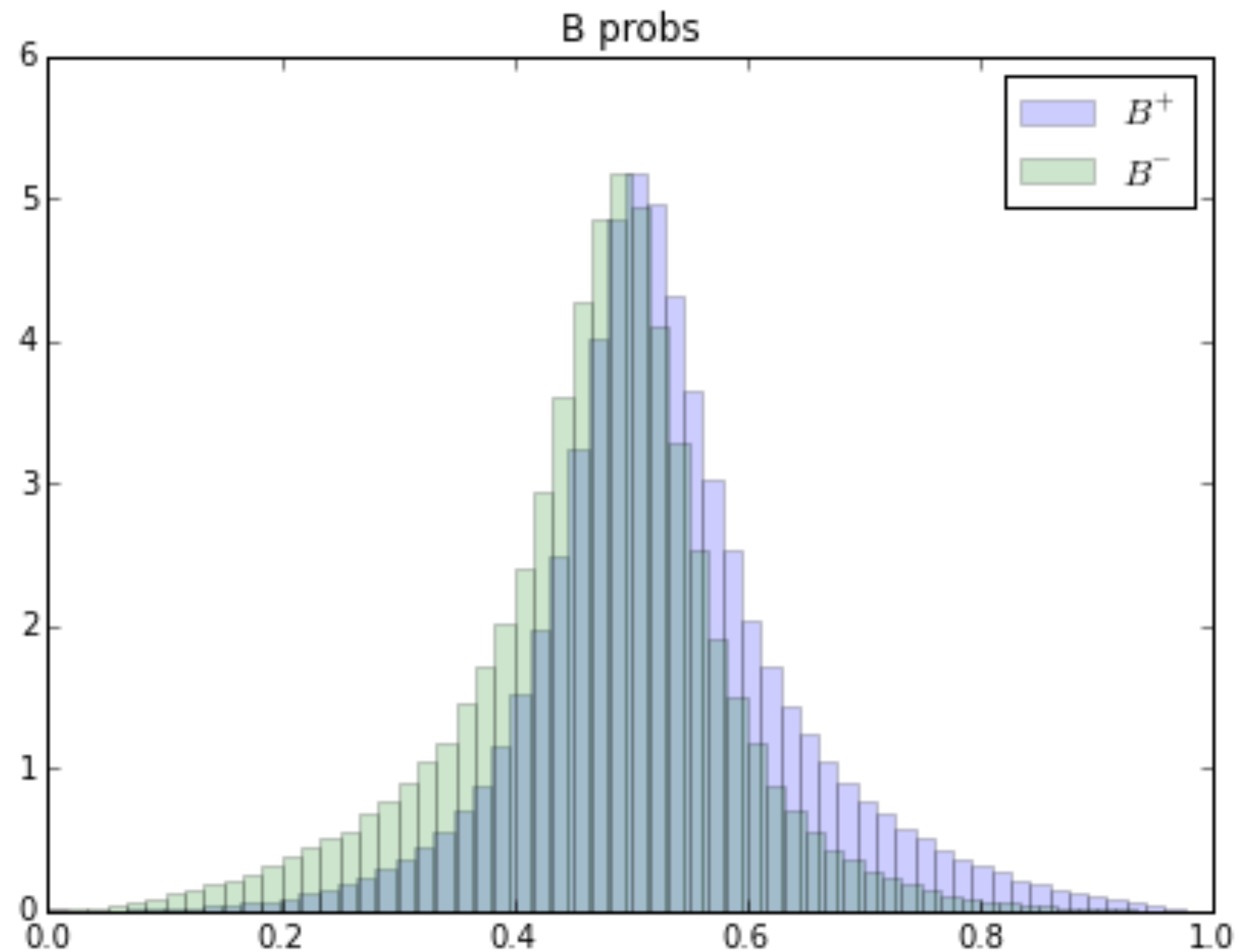
P(anti-b quark)

- › Compute $P(B^+)$ using calibrated output of classifier
- › This probability also should be calibrated (isotonic calibration is preferable)
- › Asymmetry between B^+ and B^- may appear after calibration
- › Use symmetric isotonic calibration (add inverse labels with inverse probability)

Anti-b quark probability



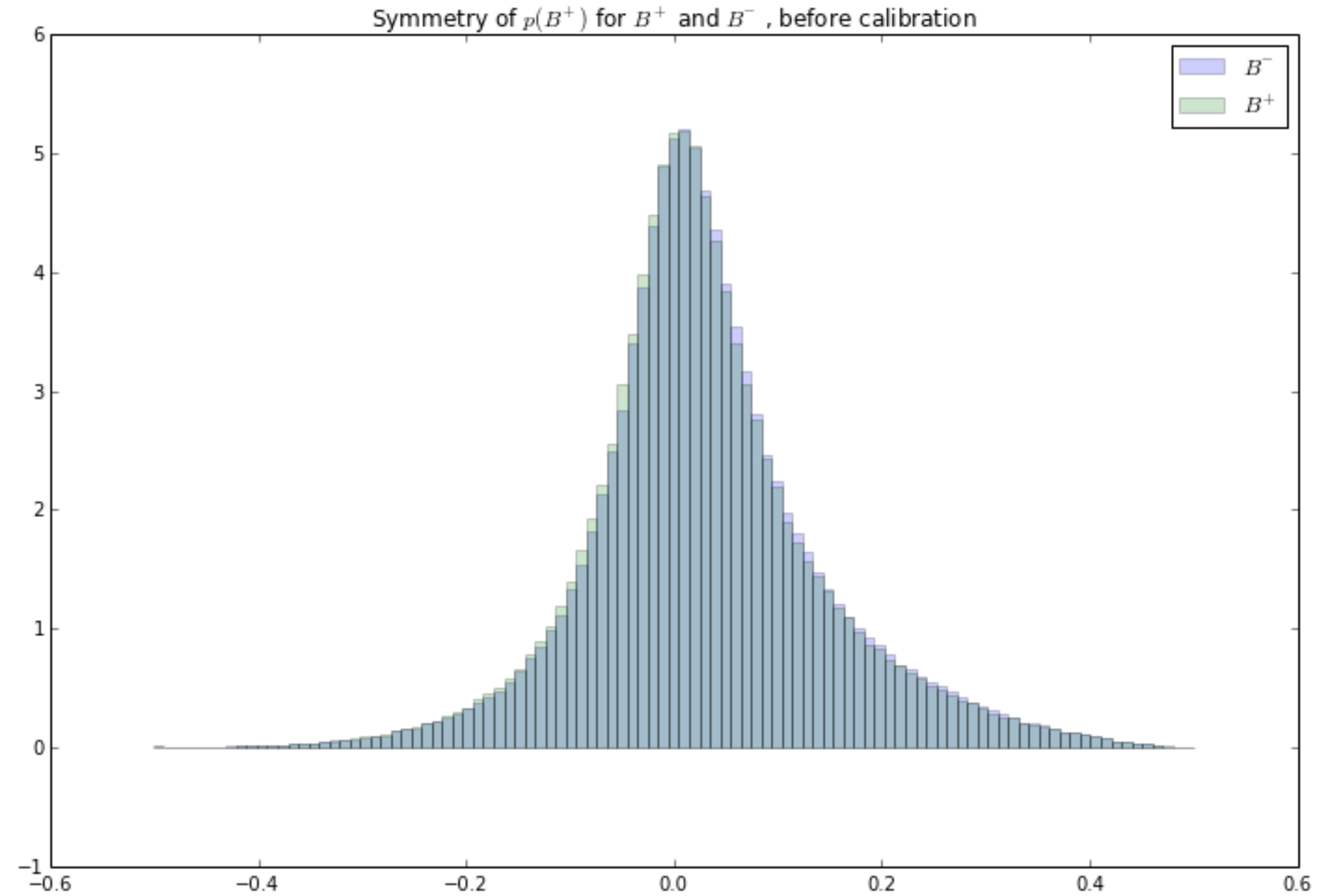
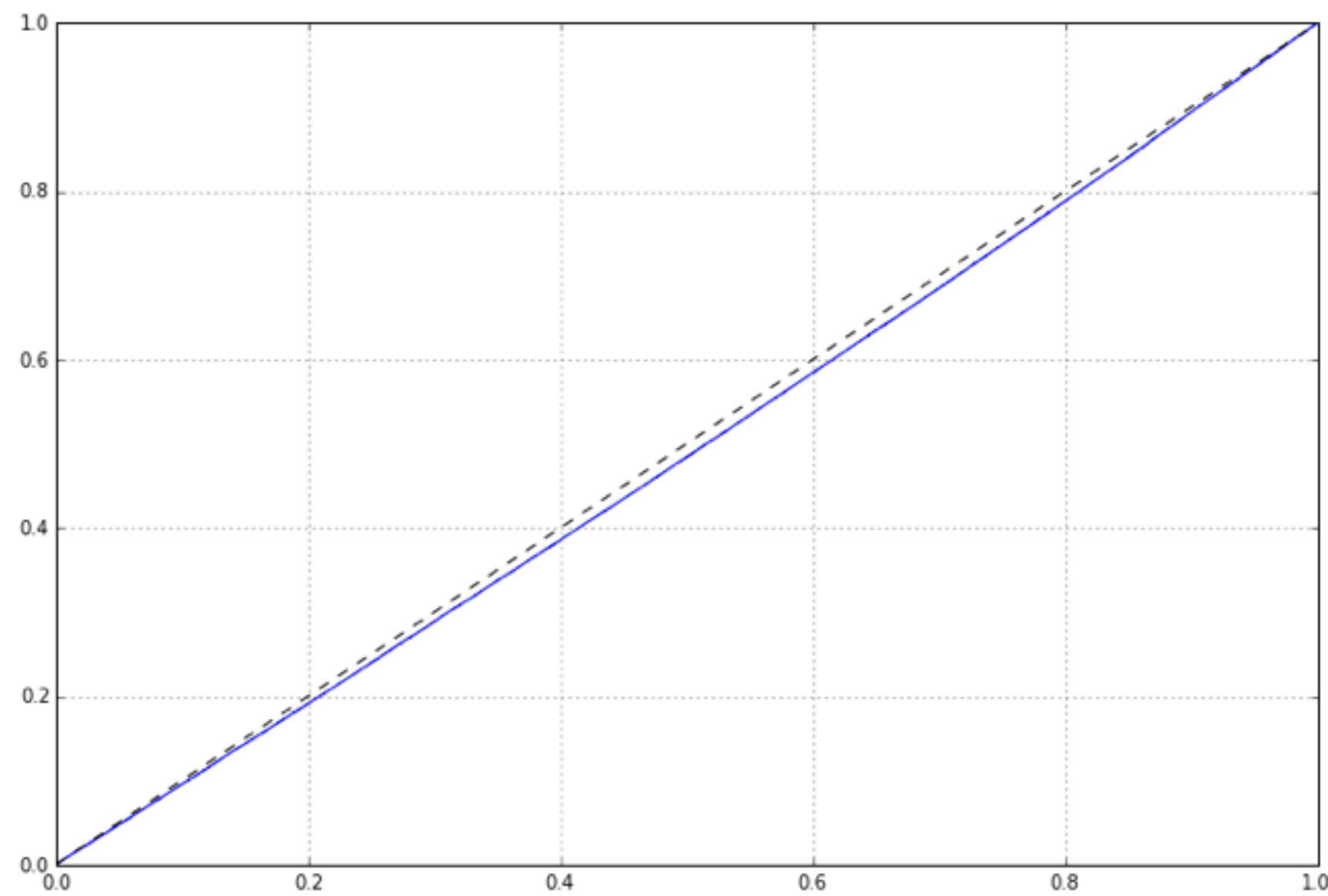
Anti-b quark probability



Add random noise after isotonic calibration for stability: $0.001 * \text{normal}(0,1)$

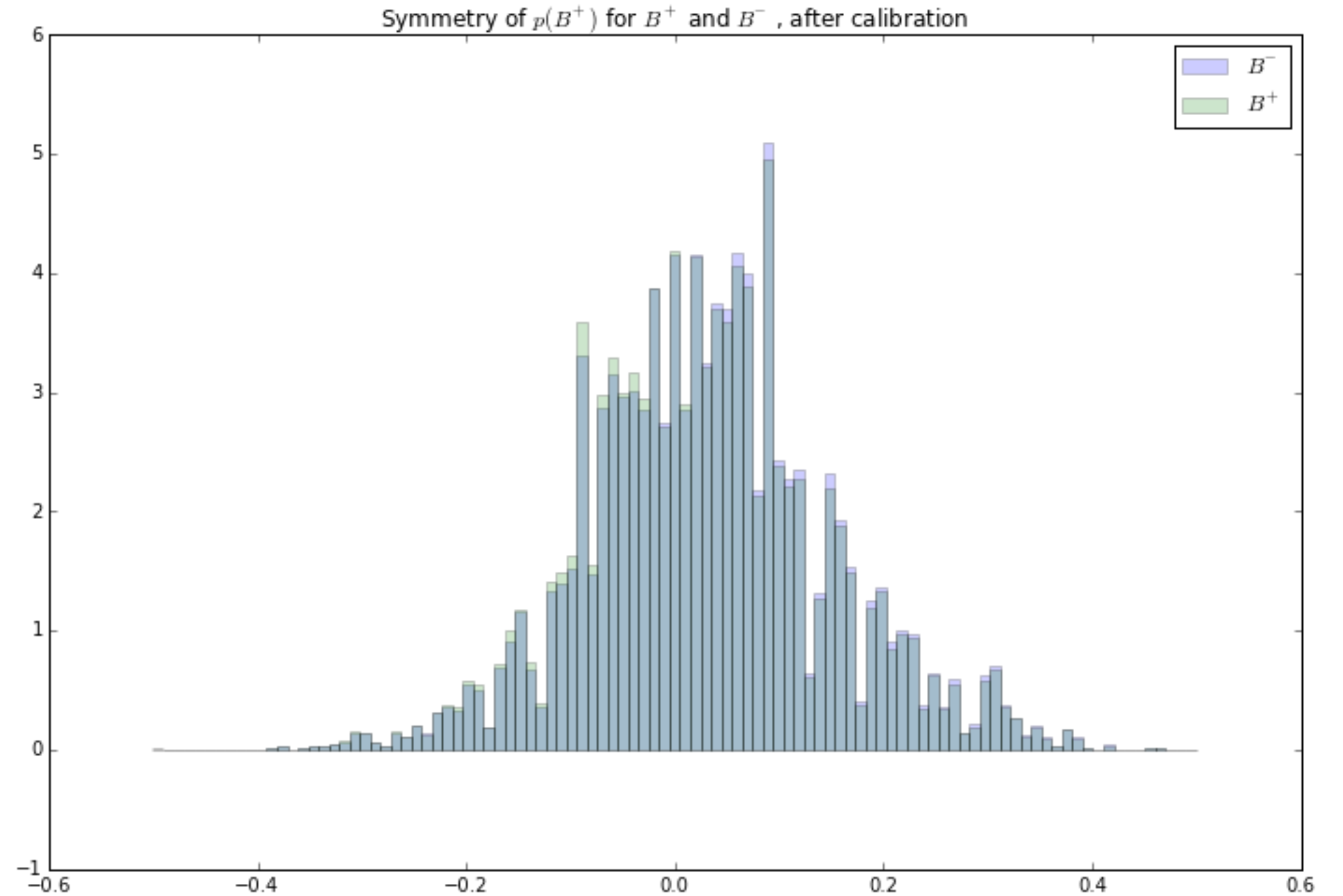
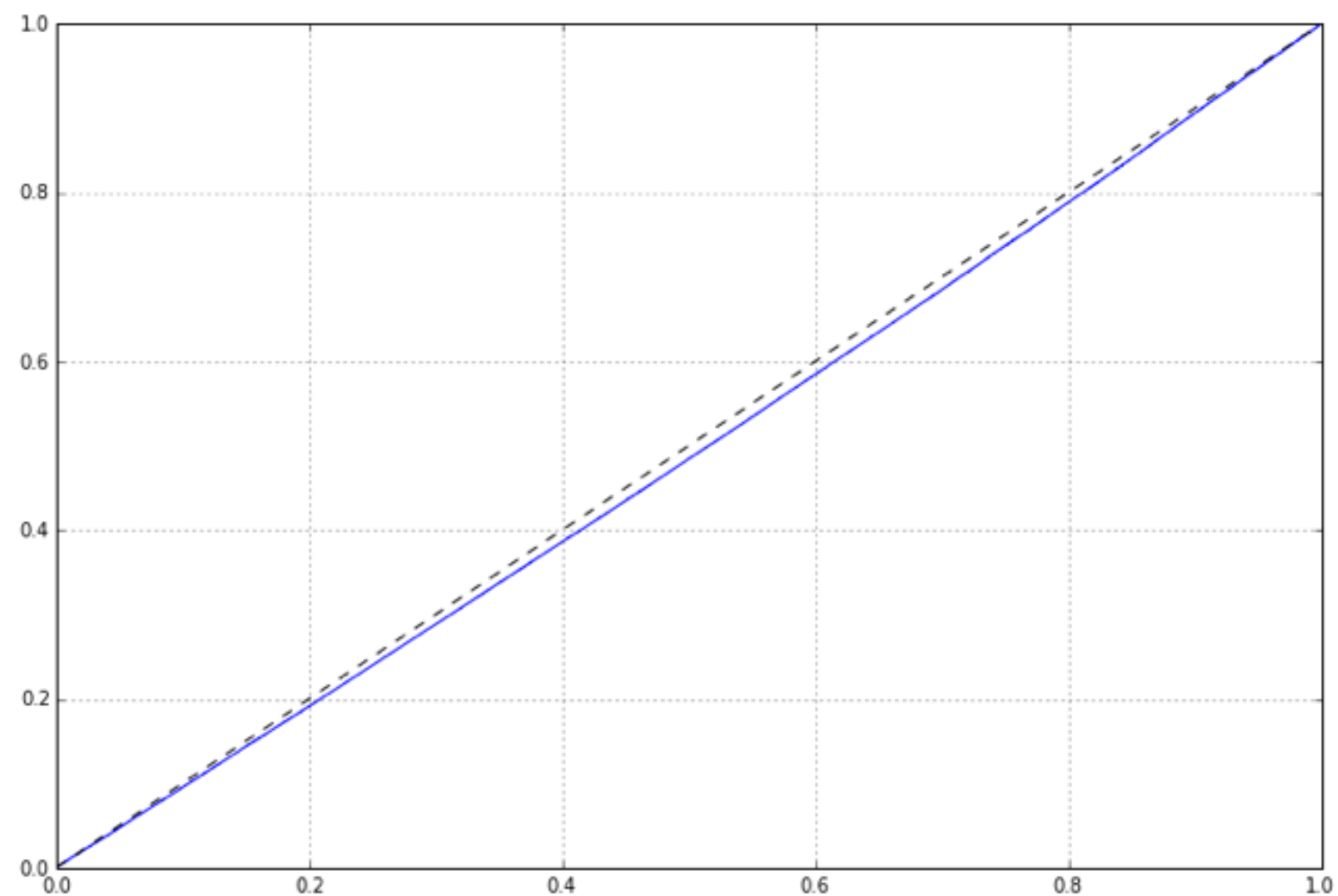
Check B symmetry (before calibration)

- > KS: 0.0163
- > compare PDFs using ROC curve
- > AUC: 0.489572

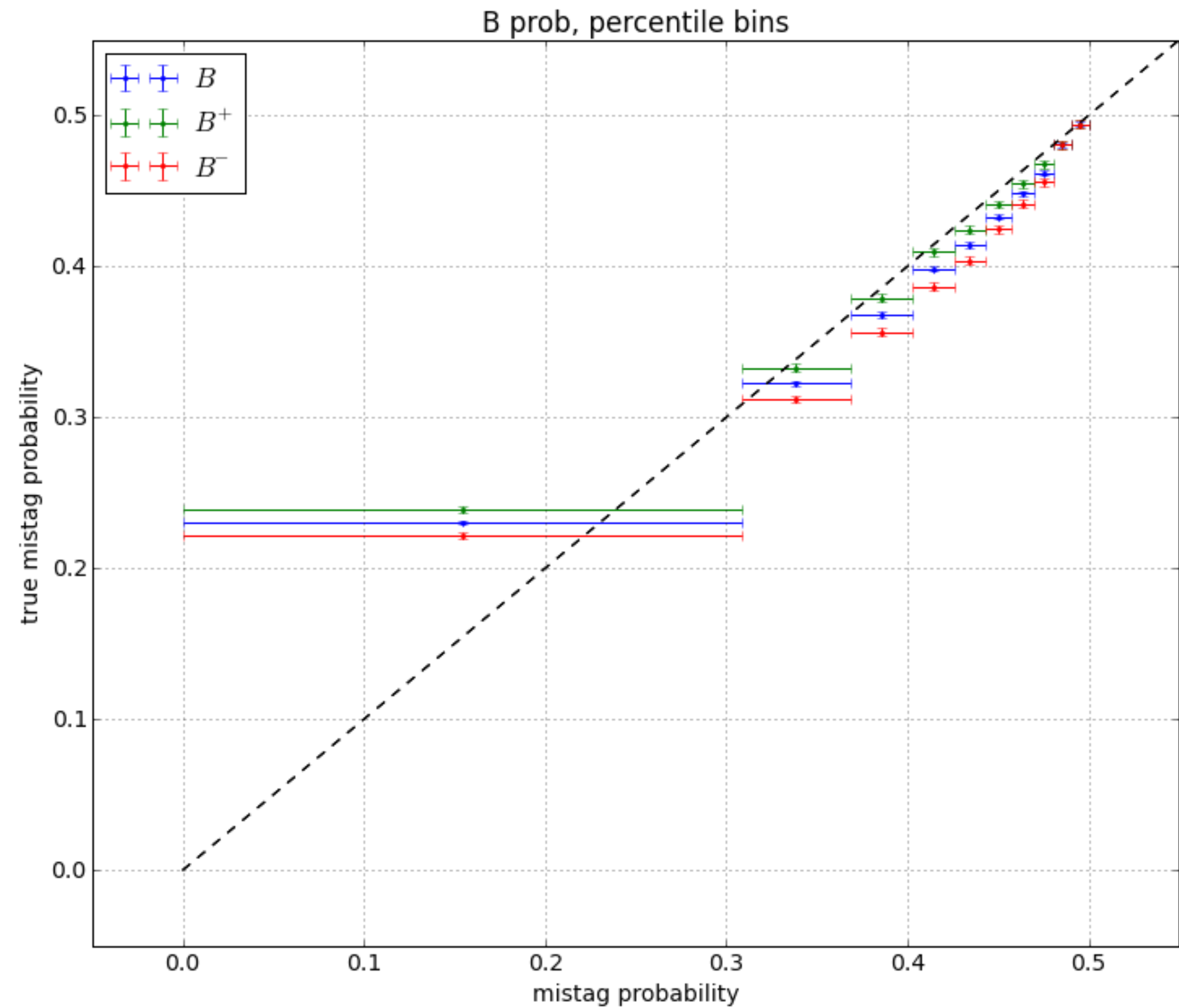
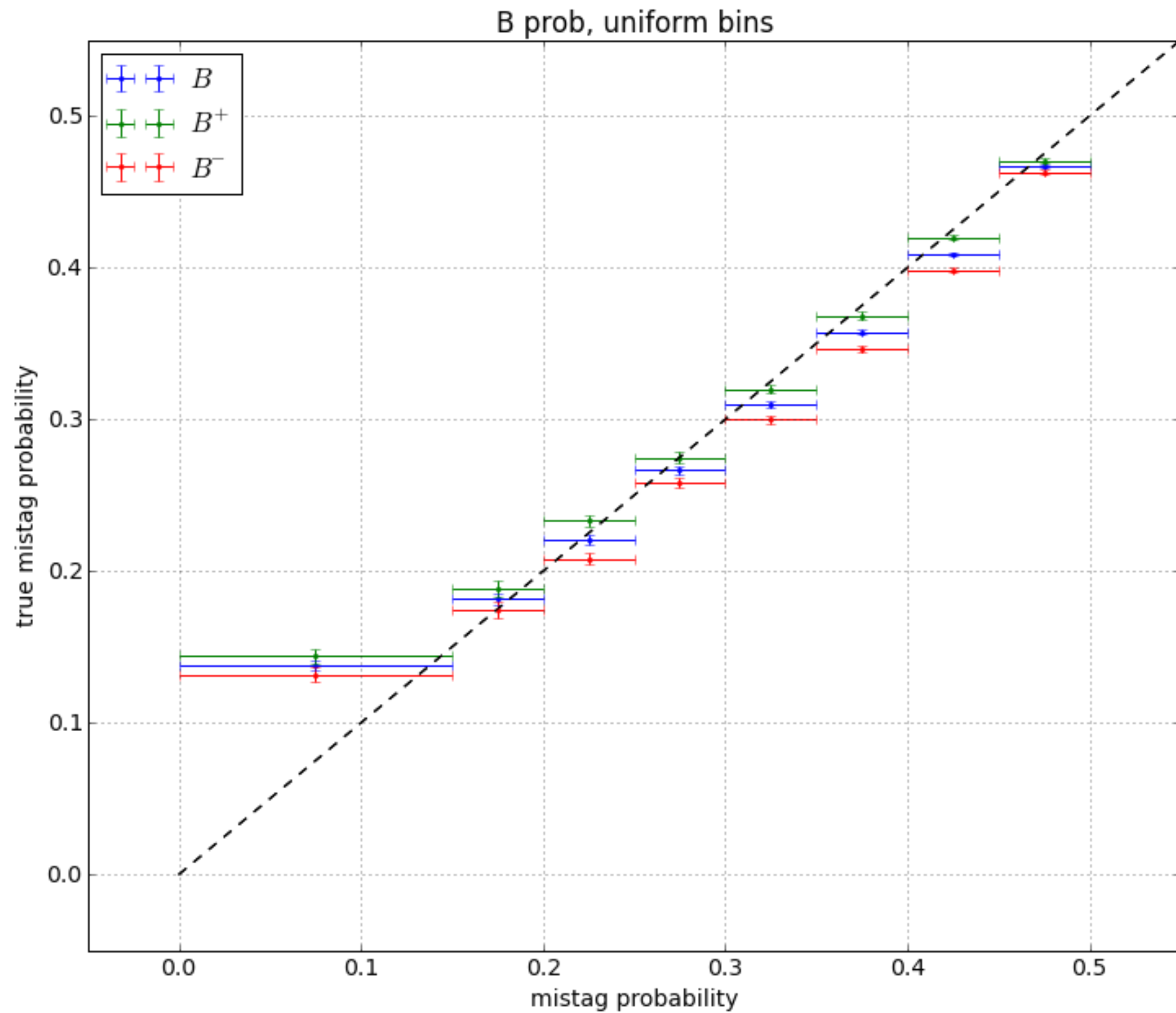


Check B symmetry (after calibration)

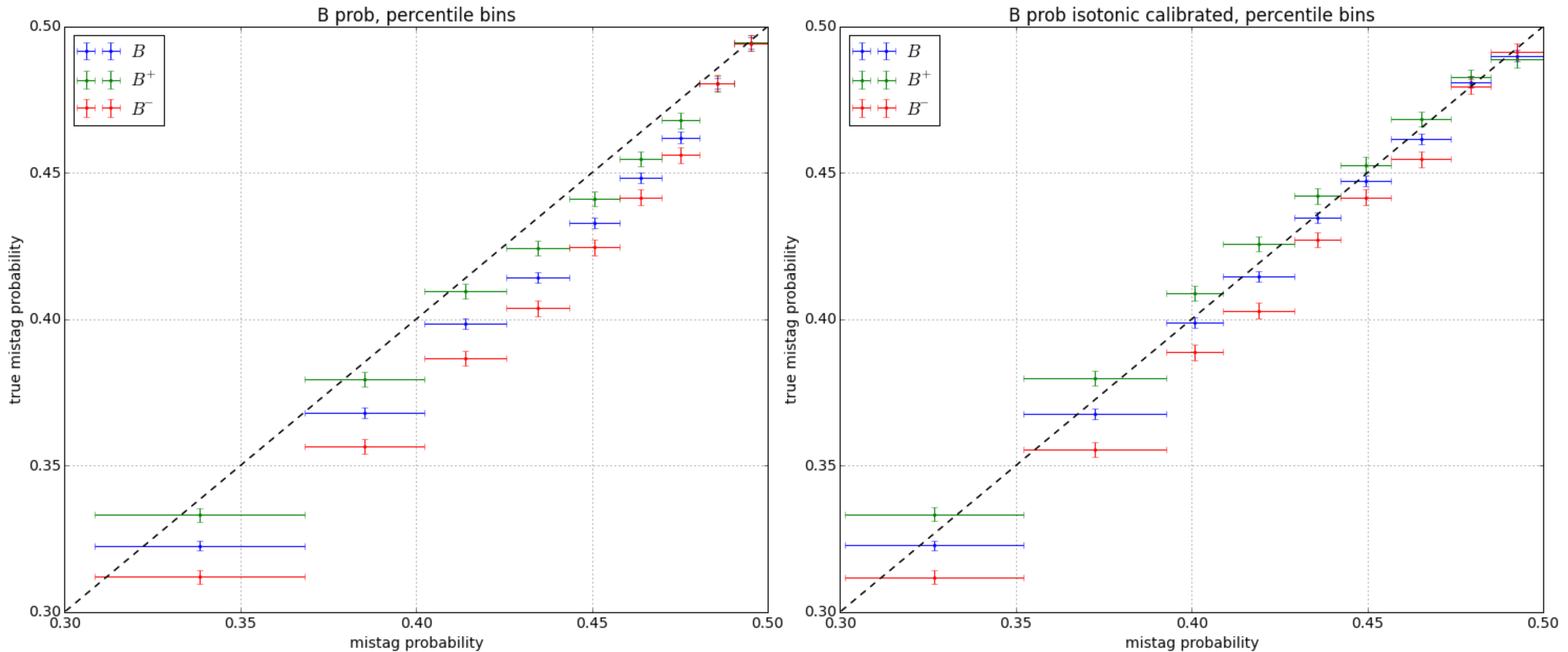
- › KS: 0.0165
- › compare PDFs using ROC curve
- › AUC: 489571



Check calibration (with B-symmetry)

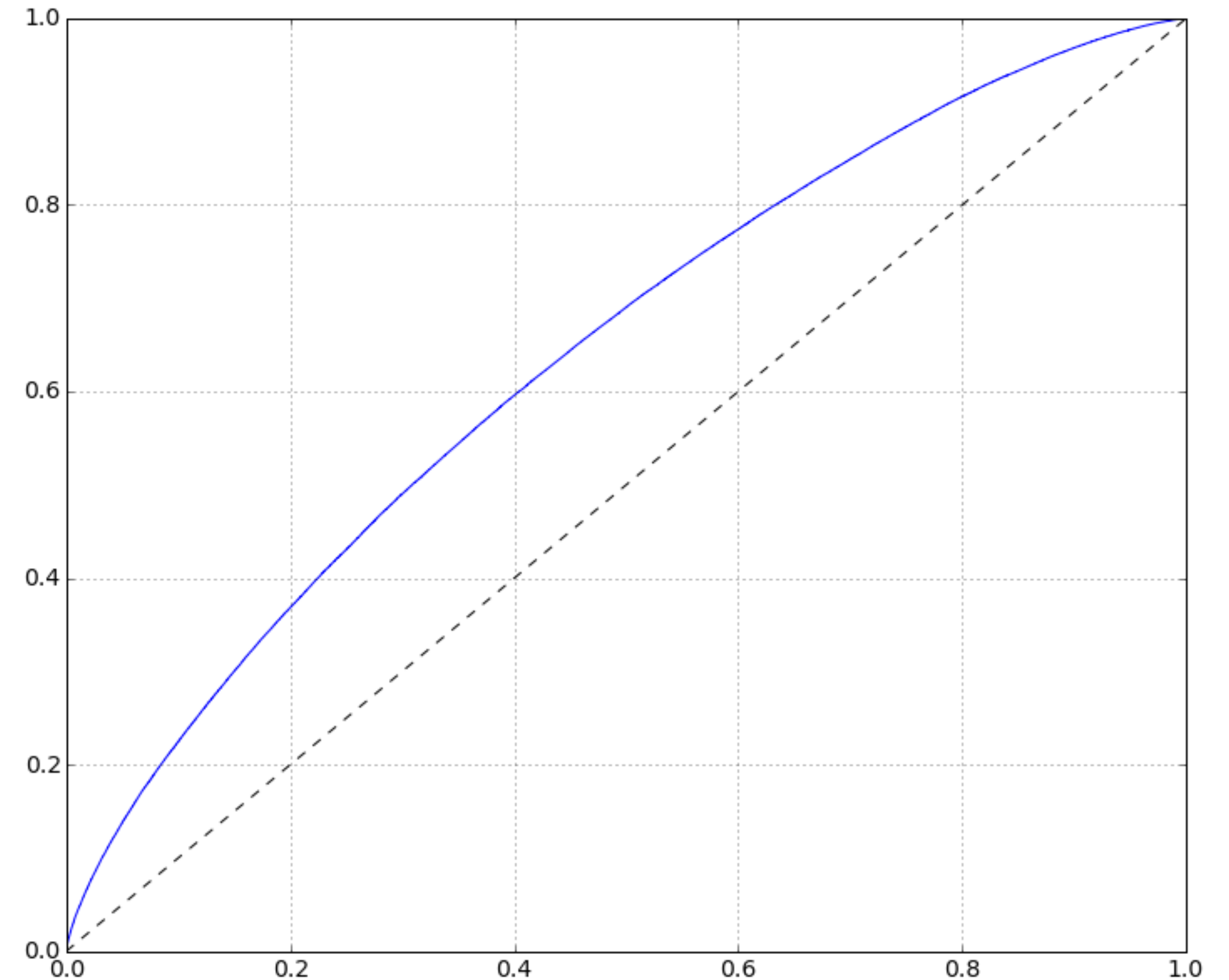


Check calibration (with B-symmetry)



Results, ROC for events

- › Compute ROC for all event including untagged (put for them 0.5 probability)
- › ROC AUC score 0.64
- › ROC AUC for current tagging system 0.566



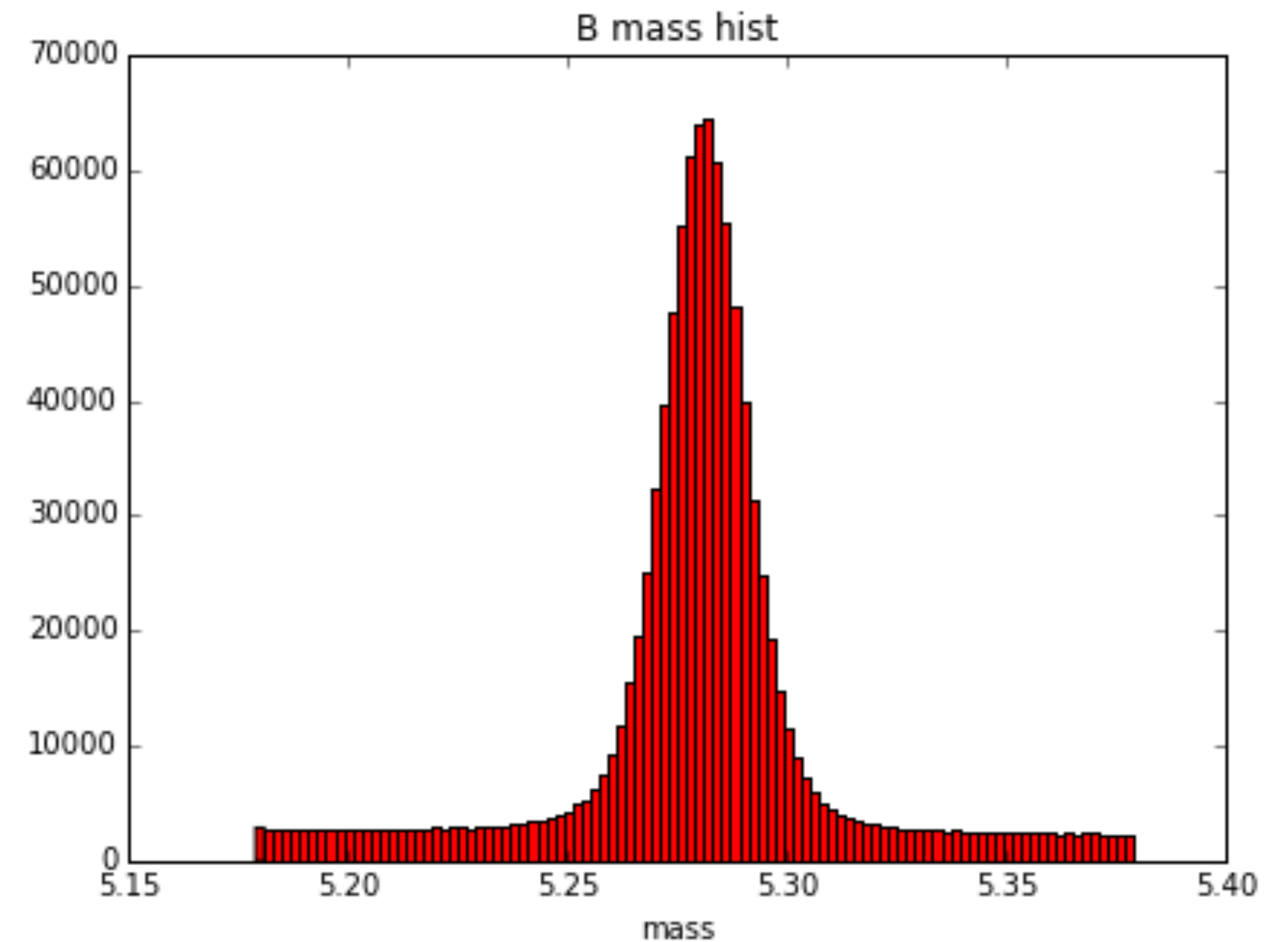
Flavour Tagging Optimization

Dependencies

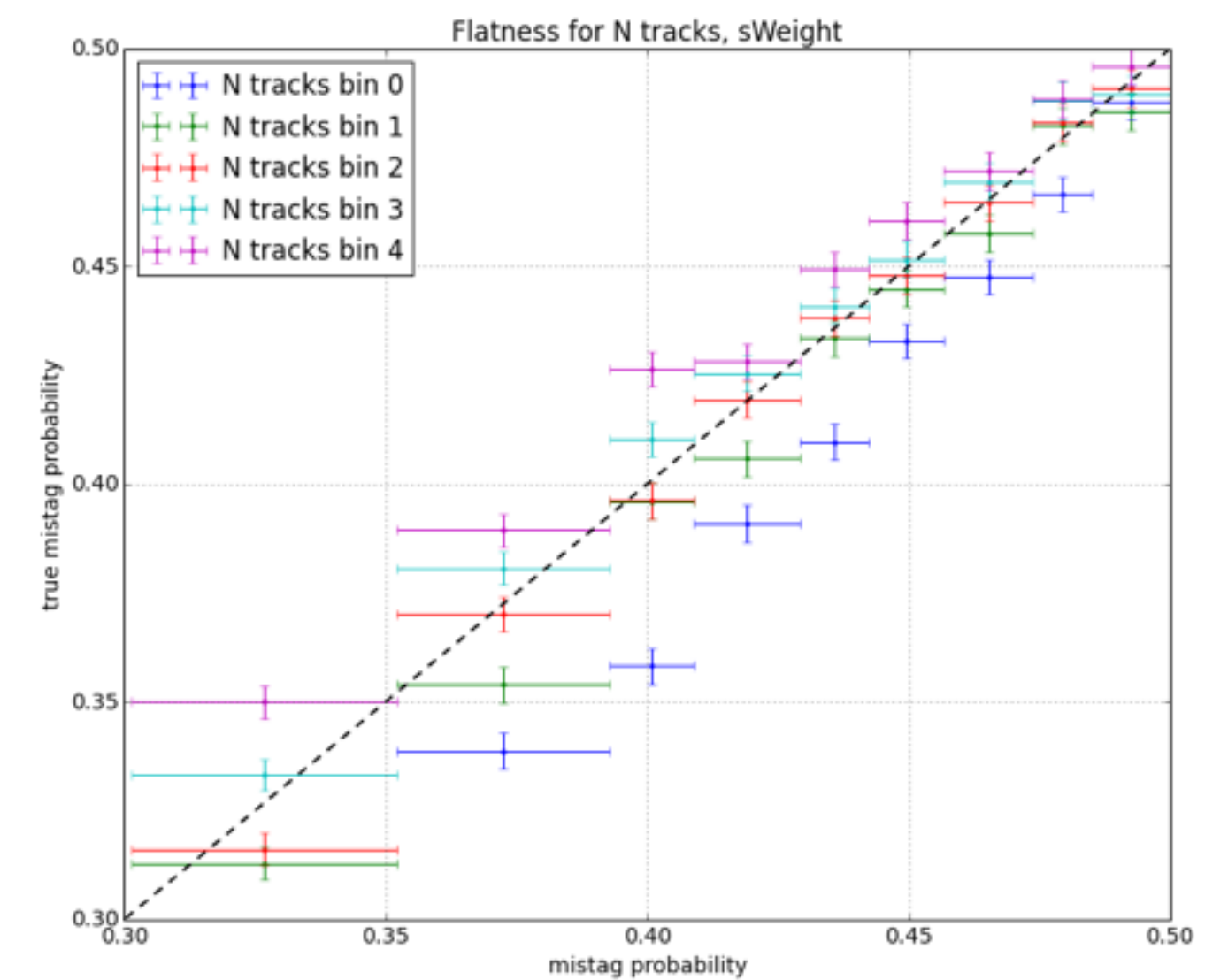
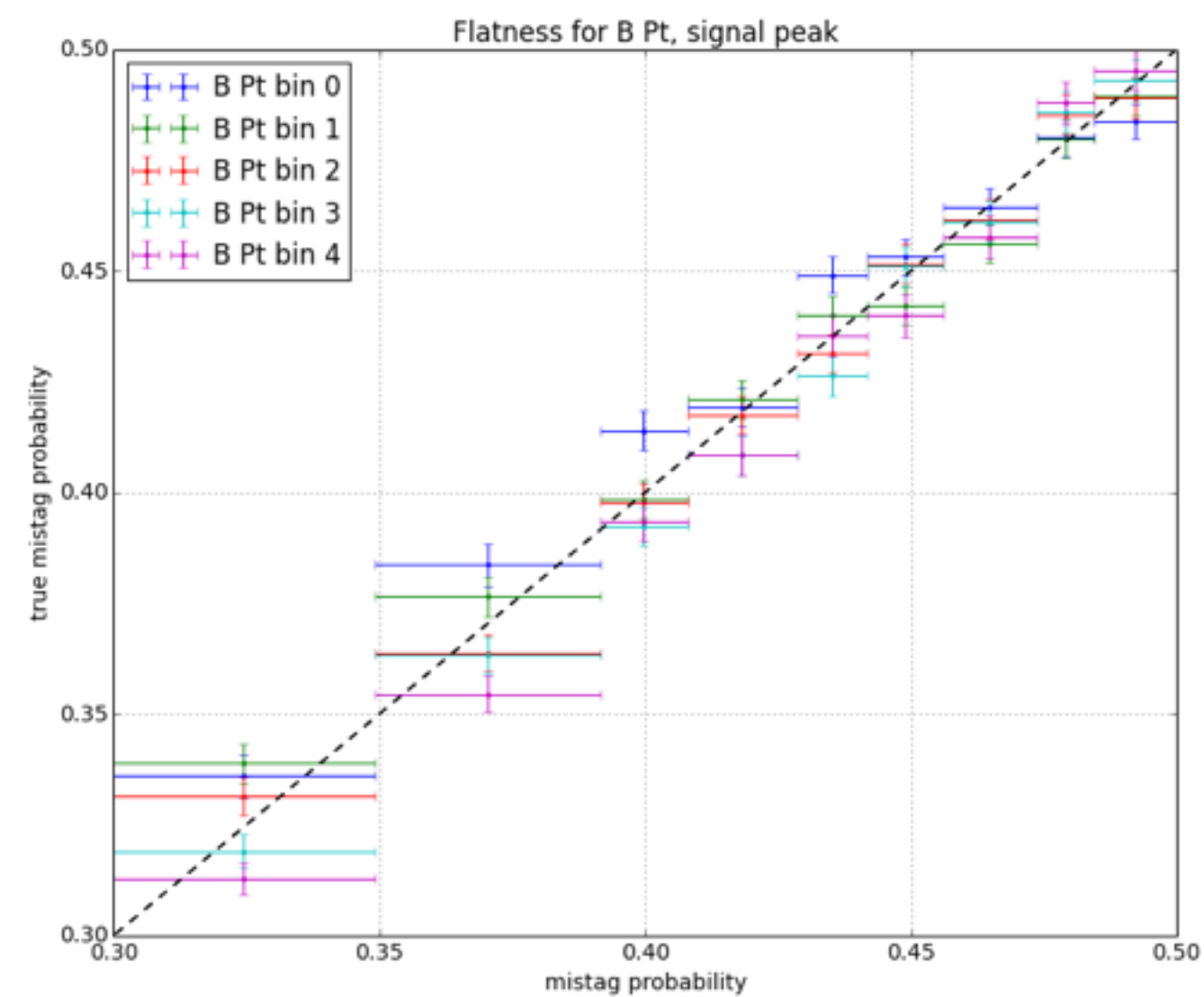
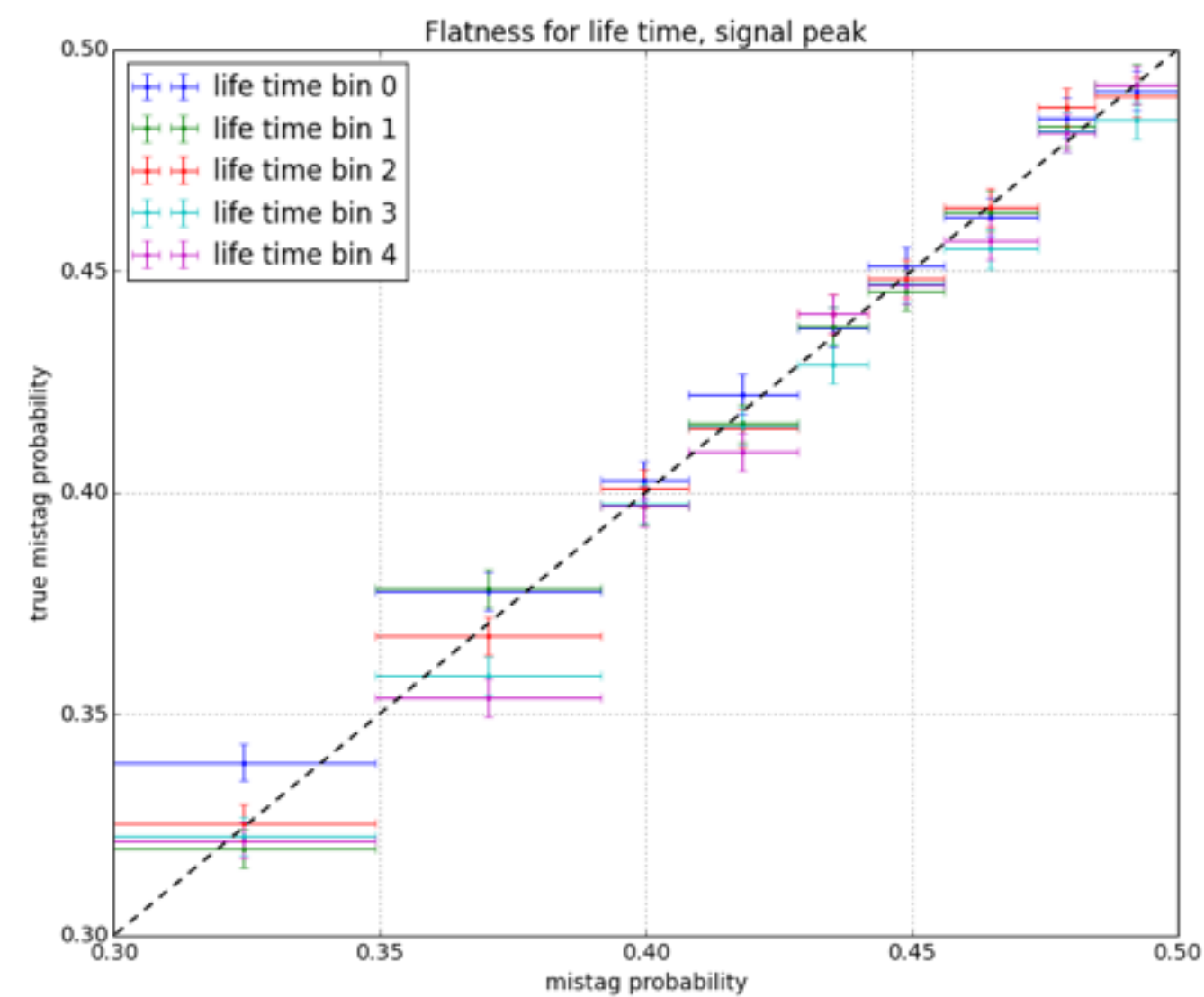
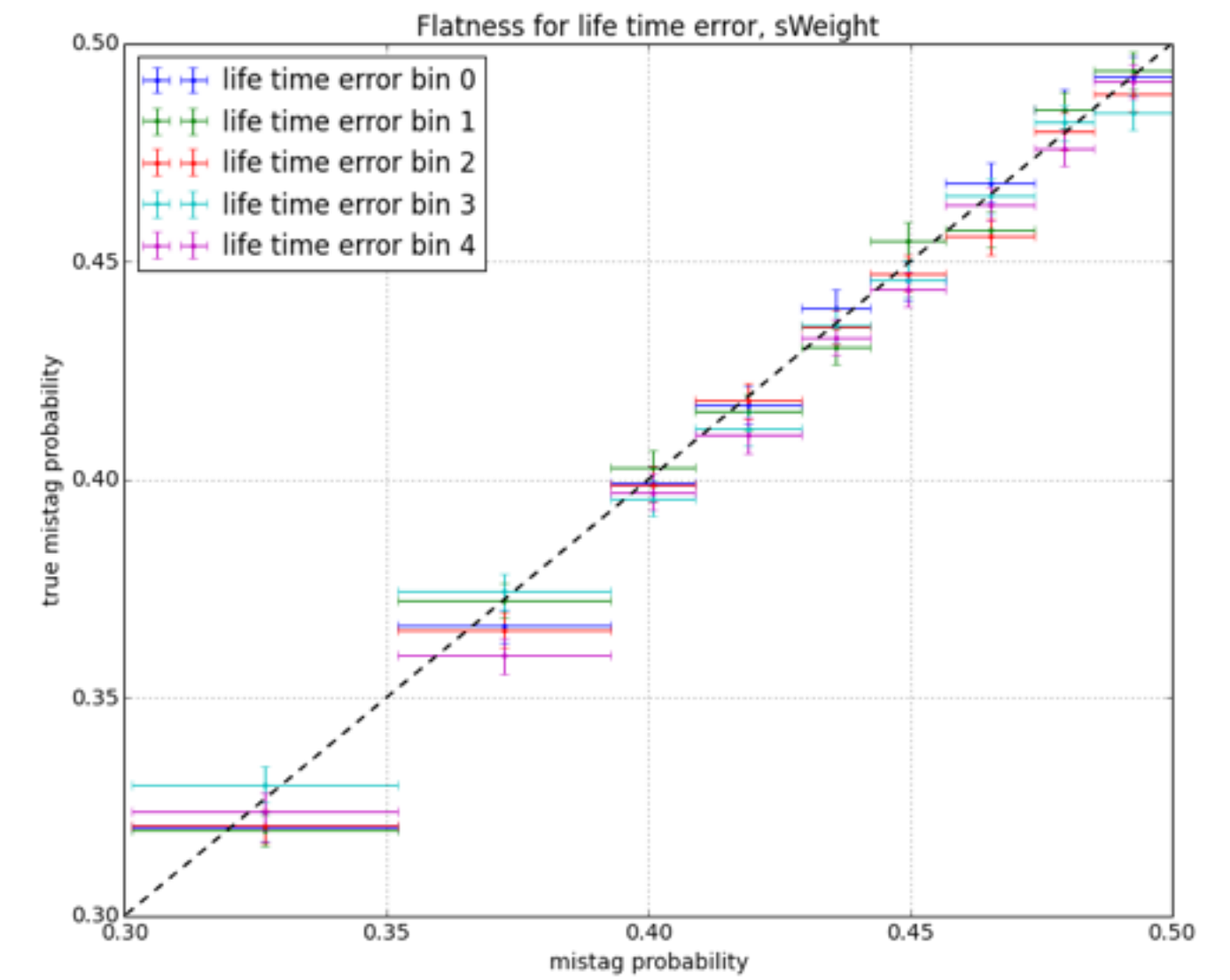
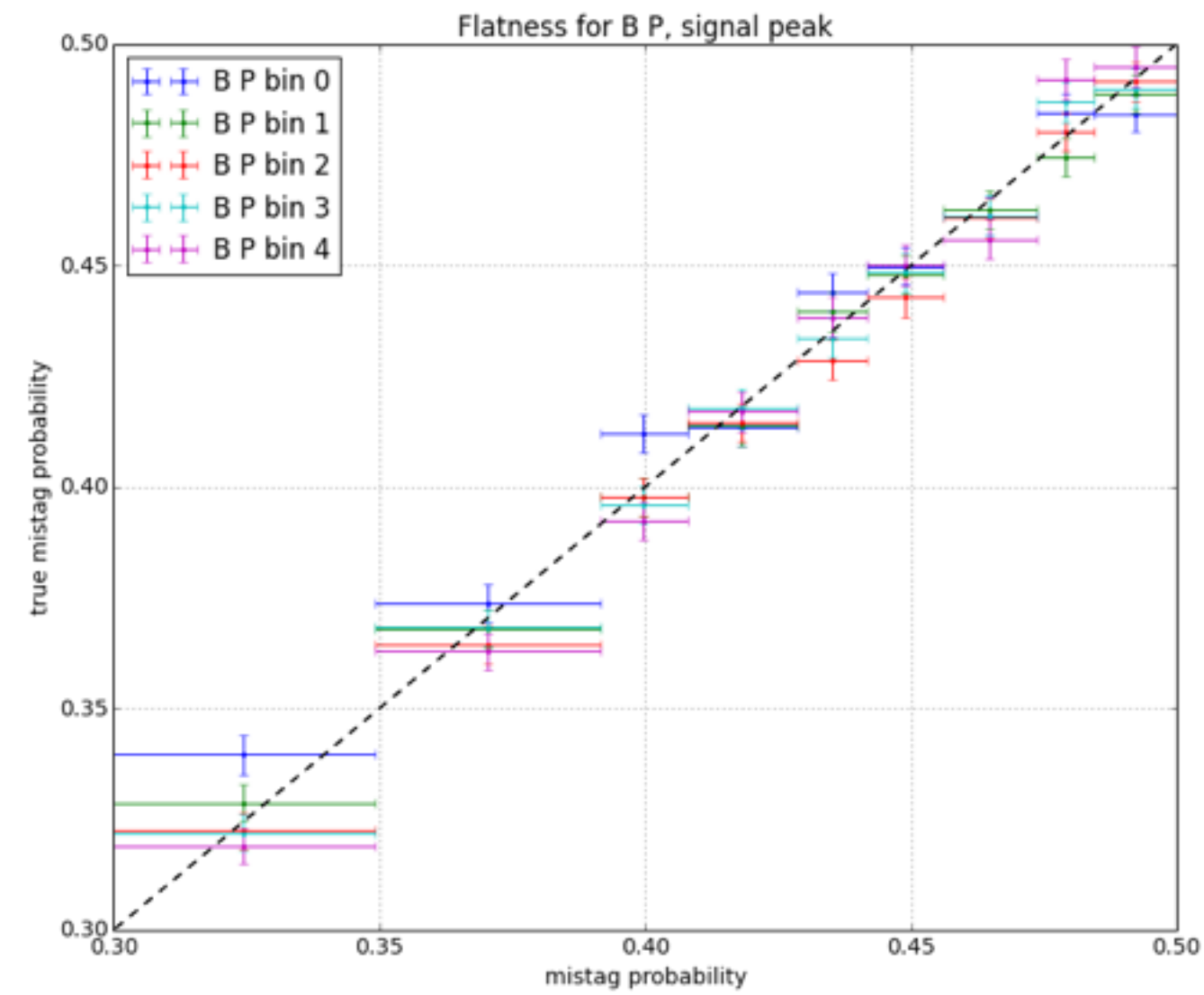
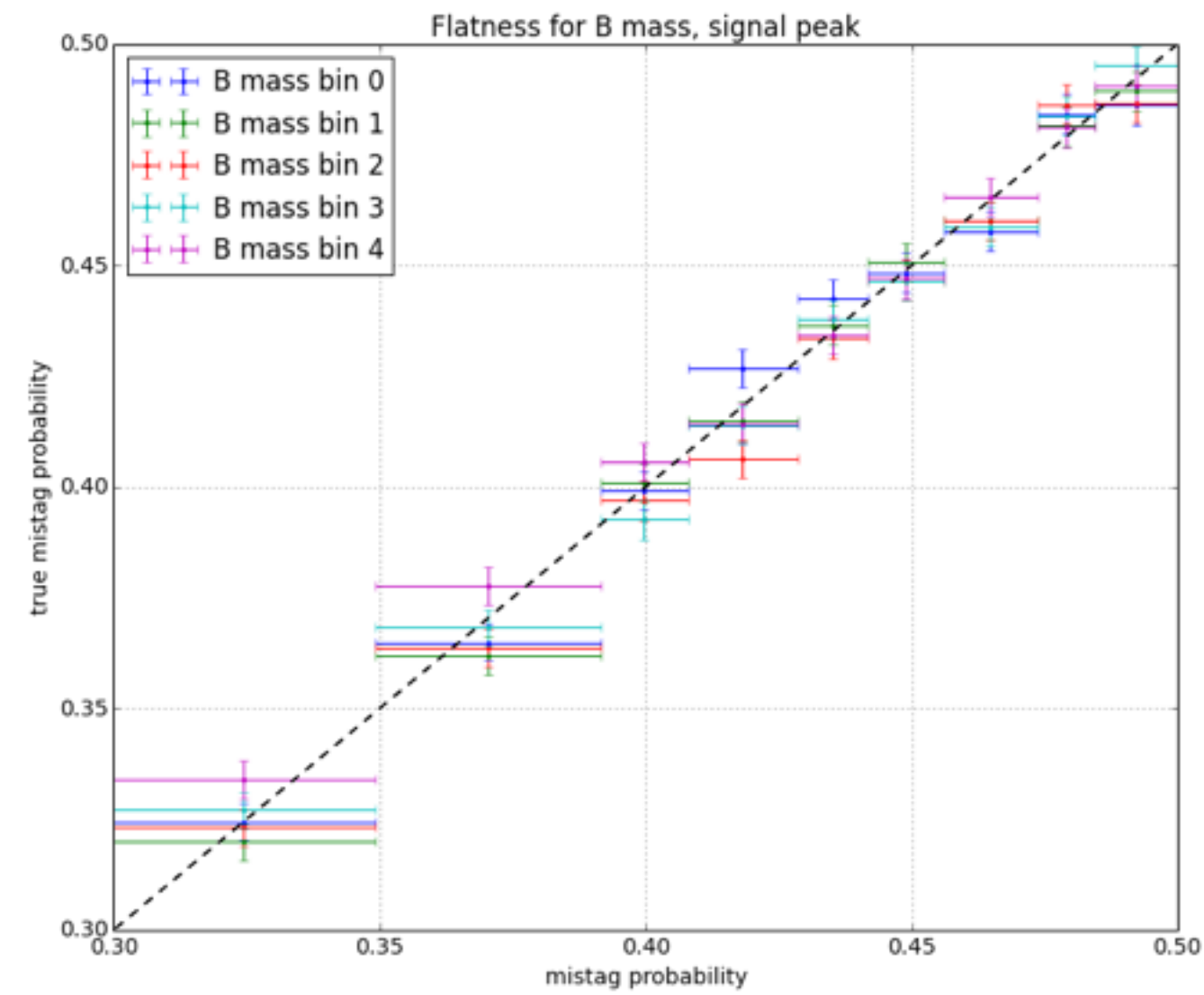


Flatness

- › For B mass, B momentum, B transverse momentum, B lifetime use sidebands as background and peak region as signal
 - they depend on the B mass
 - sWeight doesn't work in this case
- › For B lifetime error and number of tracks use sWeights
- › Procedure:
 - divide variable into 5 percentile bins
 - for each bin plot mistag vs true mistag



Flatness for B-events



Flavour Tagging Optimization

Main ideas



Summary & Tricks

- › sPlot technique (remember of variable independence on mass assumption!)
- › Probabilistic model to combine all tracks/vertices information
- › Calibration
- › Symmetric, stable isotonic calibration
- › Check similarity of 1d distributions using ROC curve
- › Flatness: check similarity of distributions (not only mean)
- › Model selection criteria: ROC curve (show the discriminative power)

References

- › https://github.com/tata-antares/tagging_LHCb
- › <https://inspirehep.net/record/1381330/files/CERN-THESIS-2015-040.pdf>
- › <http://fastml.com/classifier-calibration-with-platts-scaling-and-isotonic-regression/>,
- › <http://arxiv.org/pdf/1211.0025.pdf>
- › <http://arxiv.org/pdf/1511.00213>
- › <http://arxiv.org/pdf/physics/0402083>

Thanks for attention

Contacts

Likhomanenko Tatiana
researcher-developer



antares@yandex-team.ru, tatiana.likhomanenko@cern.ch

Flatness for background

