How to advance cosmology with the data products of machine learning

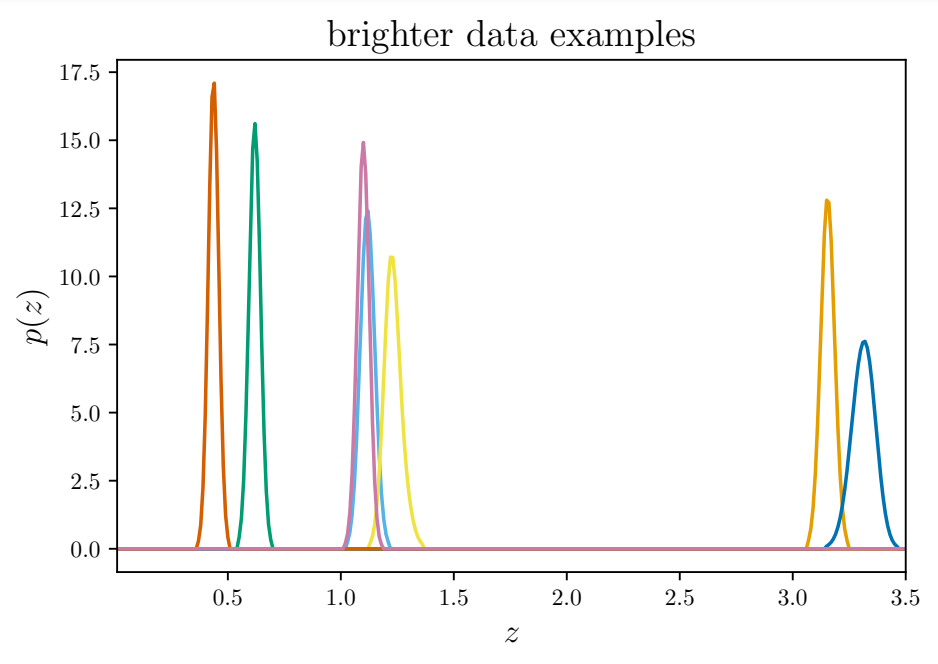
Alex Malz

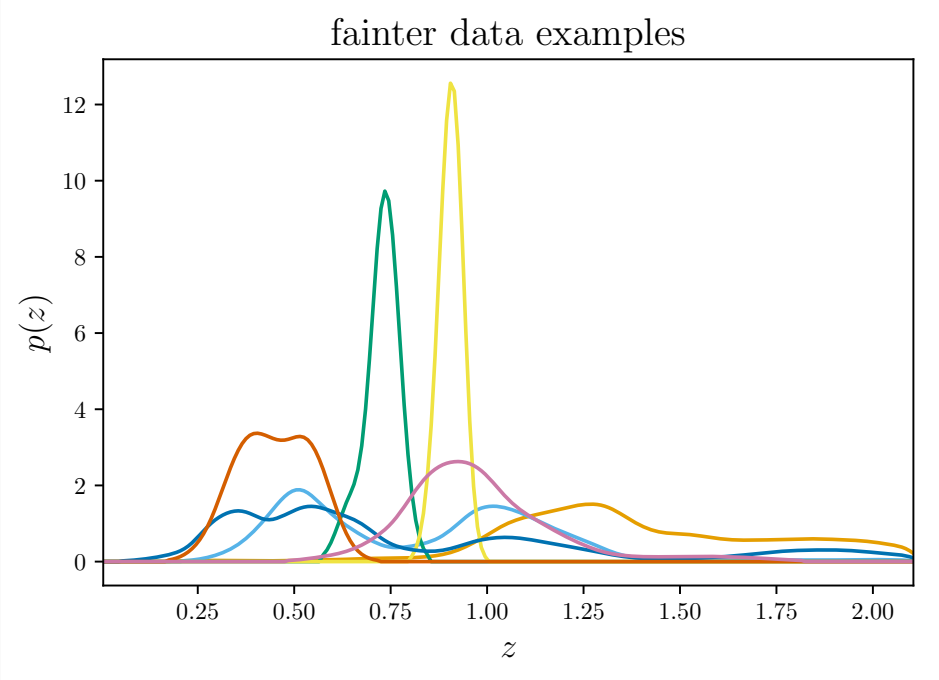
15 January 2019

Center for Cosmology and Particle Physics
New York University

What do cosmologists want from machine learning?

PDFs are more informative than point estimates.





Malz&Marshall+18

Cosmology pipelines are unprepared for photo-z PDFs.

Cosmology pipelines are unprepared for photo-z PDFs.

the stacked estimator of the redshift distribution

$$\hat{n}(z) \equiv \frac{1}{N} \sum_{i=1}^{N} p_i(z)$$

Cosmology pipelines are unprepared for photo-z PDFs.

the stacked estimator of the redshift distribution

$$\hat{n}(z) \equiv \frac{1}{N} \sum_{i=1}^{N} p_i(z)$$

and statistical independence

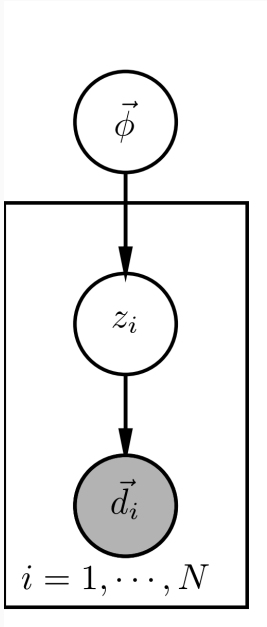
$$p(A \cap B) = p(A) \times p(B) \neq p(A) + p(B)$$

are incompatible!

CHIPPR: a model for using photo-z PDFs in inference

Cosmological
Hierarchical
Inference with
Probabilistic
Photometric
Redshifts

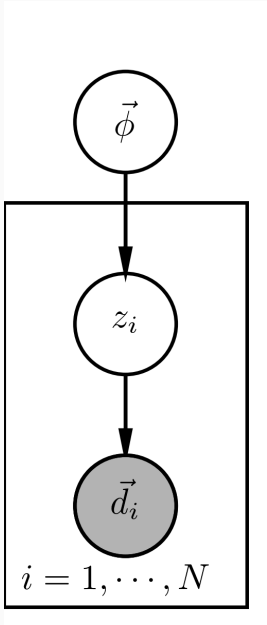
CHIPPR: a model for using photo-z PDFs in inference



Malz&Hogg(in prep)

Cosmological
Hierarchical
Inference with
Probabilistic
Photometric
Redshifts

CHIPPR: a model for using photo-z PDFs in inference



Cosmological
Hierarchical
Inference with
Probabilistic
Photometric
Redshifts

Malz&Hogg(in prep)

$$p(\vec{\phi} \mid \{\vec{d}_i\}) \propto p(\vec{\phi}) \prod_i^N \int p(z \mid \vec{d}_i, \vec{\phi}^*) \frac{p(z \mid \vec{\phi})}{p(z \mid \vec{\phi}^*)} dz$$

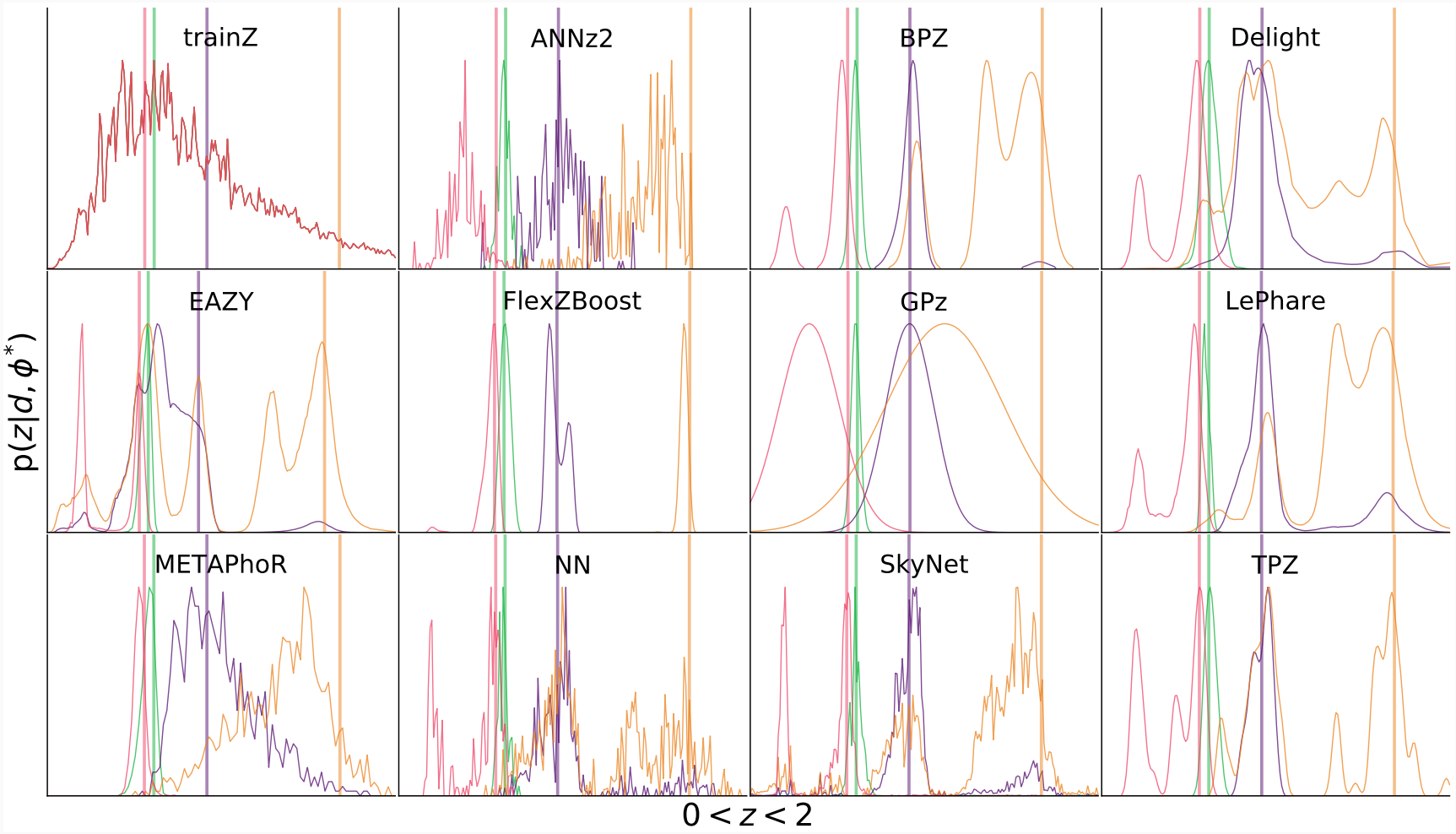
What do cosmologists want from ML?

Cosmologists want ML
to produce photo-z PDFs.

What does machine learning provide to cosmologists?

Photo-z PDFs vary despite shared, perfect training set.

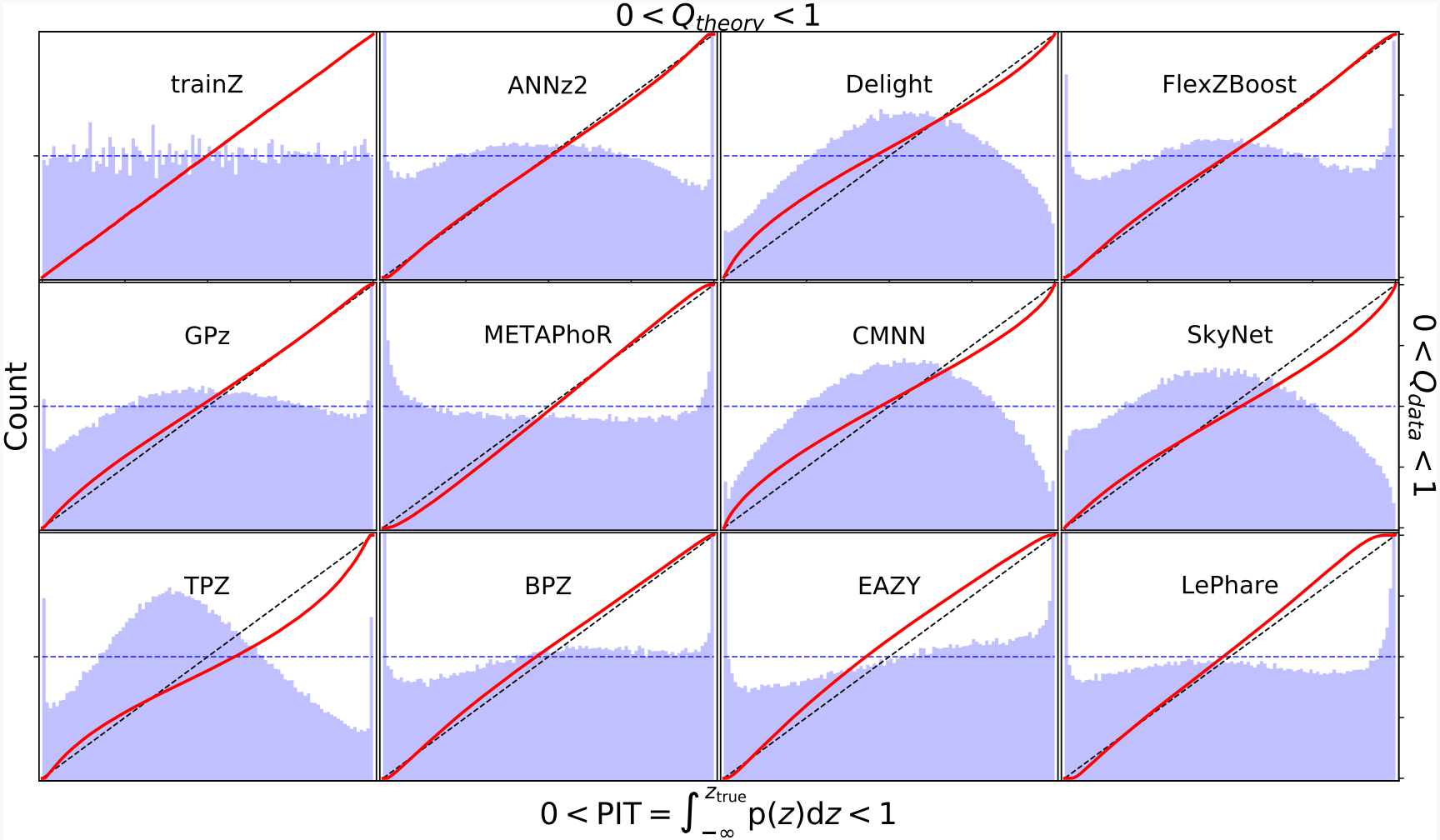
Photo-z PDFs vary despite shared, perfect training set.



Schmidt&Malz+(in review)

Established metrics of photo-z performance are inappropriate.

Established metrics of photo-z performance are inappropriate.



Schmidt&Malz+(in review)

What does ML provide to cosmologists?

ML yields photo-z posteriors weighted by an implicit prior.

What do cosmologists need from machine learning?

Cosmology pipelines are unprepared for the implicit prior.

the stacked estimator of the redshift distribution

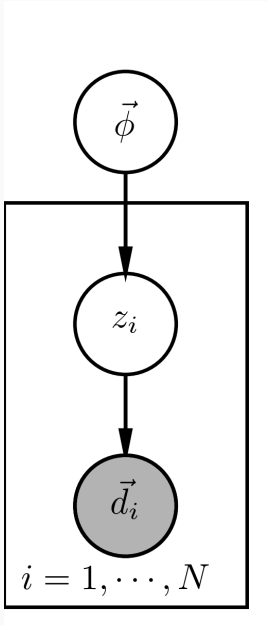
$$\hat{n}(z) \equiv \frac{1}{N} \sum_{i=1}^{N} p(z|\vec{d}_i, \vec{\phi}^*)$$

and statistical independence

$$p(A \cap B|C) = p(A|C) \times p(B|C) \neq p(A|C) + p(B|C)$$

are incompatible!

CHIPPR: a model for using photo-z PDFs in inference

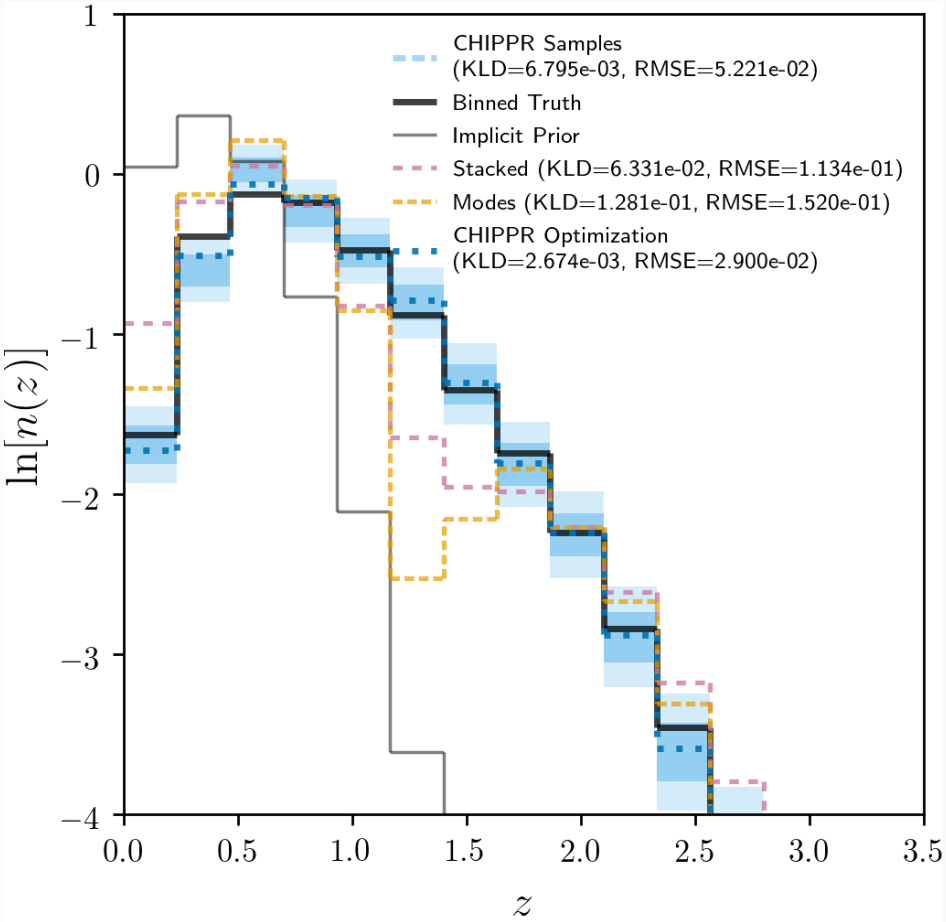


Cosmological
Hierarchical
Inference with
Probabilistic
Photometric
Redshifts

Malz&Hogg(in prep)

$$p(\vec{\phi} \mid \{\vec{d}_i\}) \propto p(\vec{\phi}) \prod_i^N \int p(z \mid \vec{d}_i, \vec{\phi}^*) \frac{p(z \mid \vec{\phi})}{p(z \mid \vec{\phi}^*)} dz$$

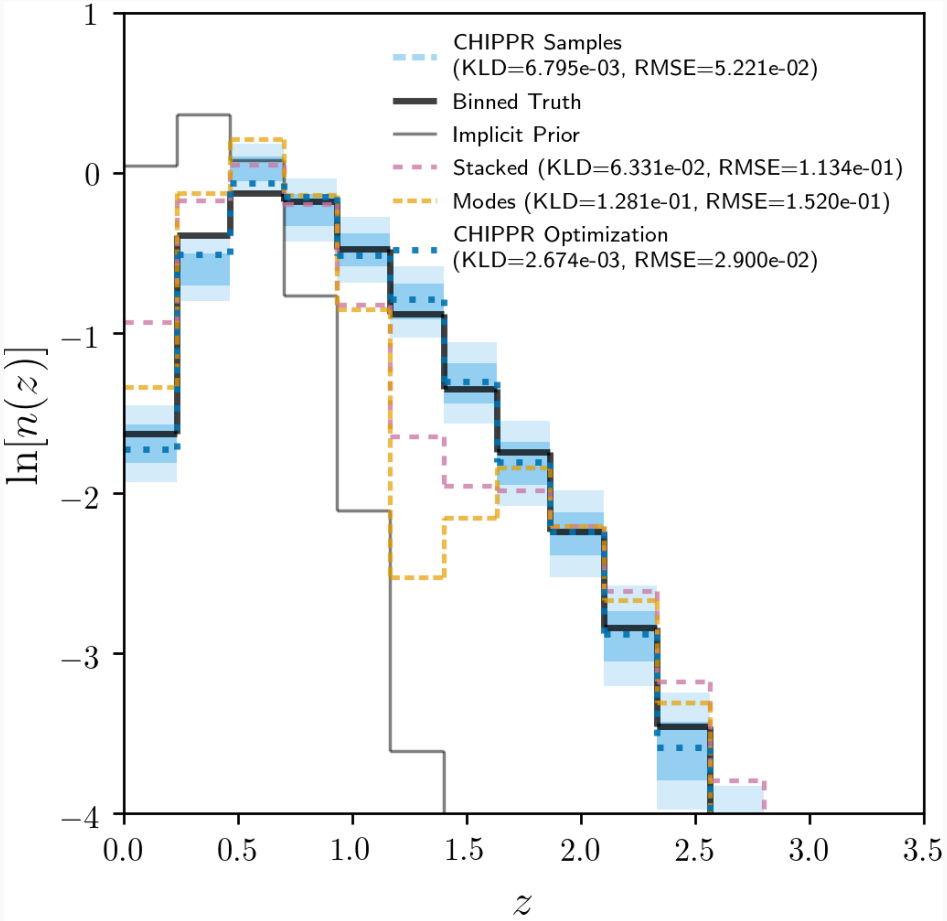
CHIPPR is robust to a nontrivial implicit prior.

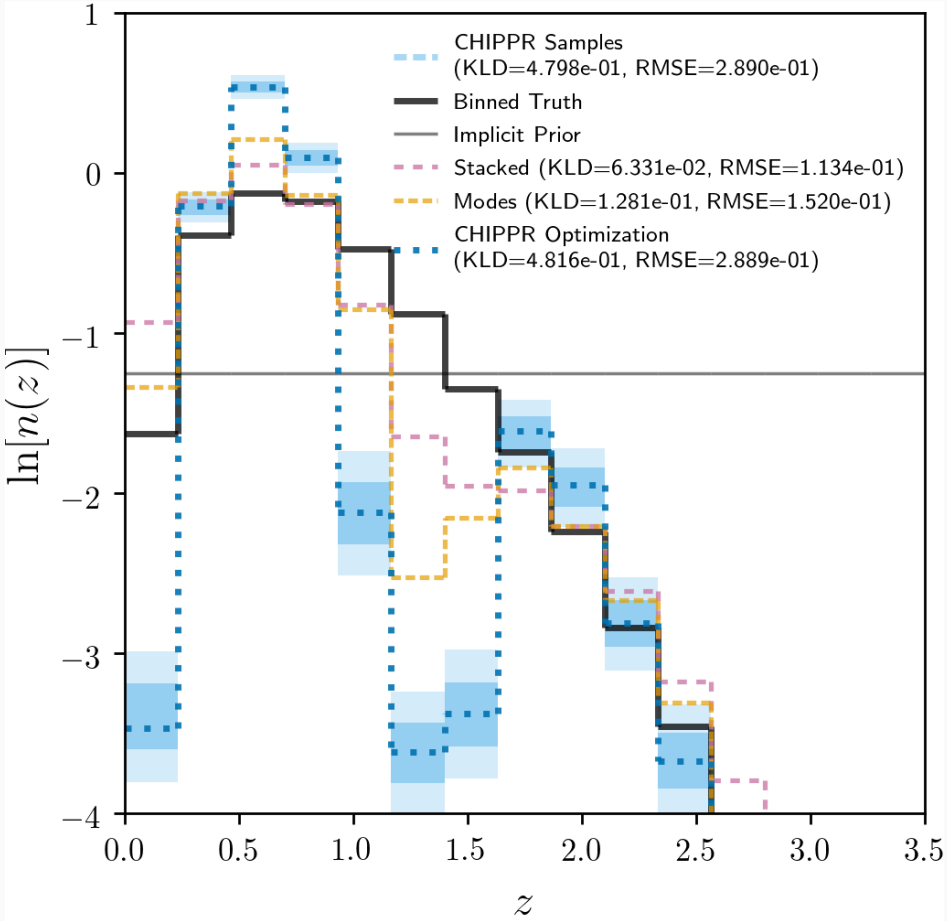


Malz&Hogg(in prep)

`github.com/aimalz/chippr`

Even CHIPPR fails under prior misspecification.





Malz&Hogg(in prep)

`github.com/aimalz/chippr`

What do cosmologists need from ML?

The implicit prior
must be known
in order to perform inference.

What do cosmologists need from ML?

A realistic implicit prior introduces further complications.

What do cosmologists need from ML?

A realistic implicit prior introduces further complications.

- The galaxy catalog may not share a single implicit prior.

What do cosmologists need from ML?

A realistic implicit prior introduces further complications.

- The galaxy catalog may not share a single implicit prior.
- The implicit prior may be derived nondeterministically.

What do cosmologists need from ML?

A realistic implicit prior introduces further complications.

- The galaxy catalog may not share a single implicit prior.
- The implicit prior may be derived nondeterministically.
- Retrieving the implicit prior after combining PDFs from multiple ML methods could be impossible.

ML is not a panacea for photo-zs.

What do cosmologists want from ML?

Cosmologists want ML to produce photo-z PDFs.

What does ML provide to cosmologists?

The photo-z PDFs of ML are at best implicit posteriors.

What do cosmologists need from ML?

The implicit prior must be known in order to perform inference.