Comparison metrics

Alex Malz
LINCC@CMU

DSFP Session 16

Adapted from a lesson at the Winter 2020 LSST-DESC collaboration meeting's "DE School"

LSSTC DATA SCIENCE FELLOWSHIP PROGRAM

Overview

The problem set aims to be specific to Bayesian model comparison, but this lecture is to give you some context for what metrics mean in the context of working with probabilities as the end goal.

How do we know if/when any estimated probabilities are good enough?

Conditional probability & forward models

Conditional probability & forward models

The universe has <u>true</u> physical parameters θ.

Conditional probability & forward models

The universe has true physical parameters θ.

There is a causal relationship $\boxed{p(x \mid \theta)}$ between data and the physical parameters; the parameters θ determine the data x.

Probability distribution function (PDF)

Conditional probability & forward models

The universe has true physical parameters θ.

↓

There is a causal relationship $p(x \mid \theta)$ between data and the physical parameters; the parameters θ determine the data x.

↓

We observe instances of data x <u>generated</u> by that model.

A tangible example of astrophysical parameter estimation under Bayesian statistics

Context: photo-*z*s

LSSTC DATA SCIENCE FELLOWSHIP PROGRAM

Spectroscopic redshift determination

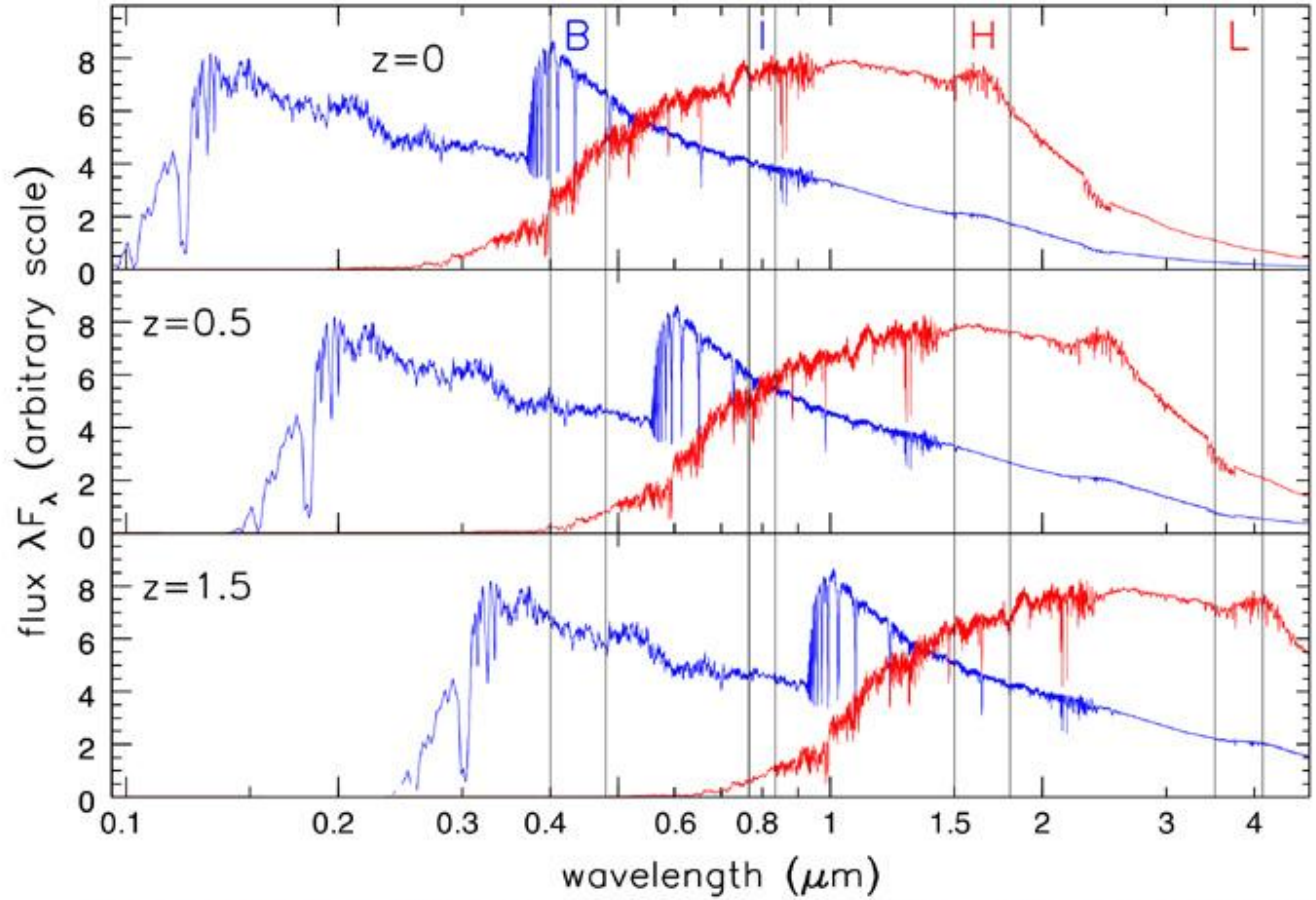
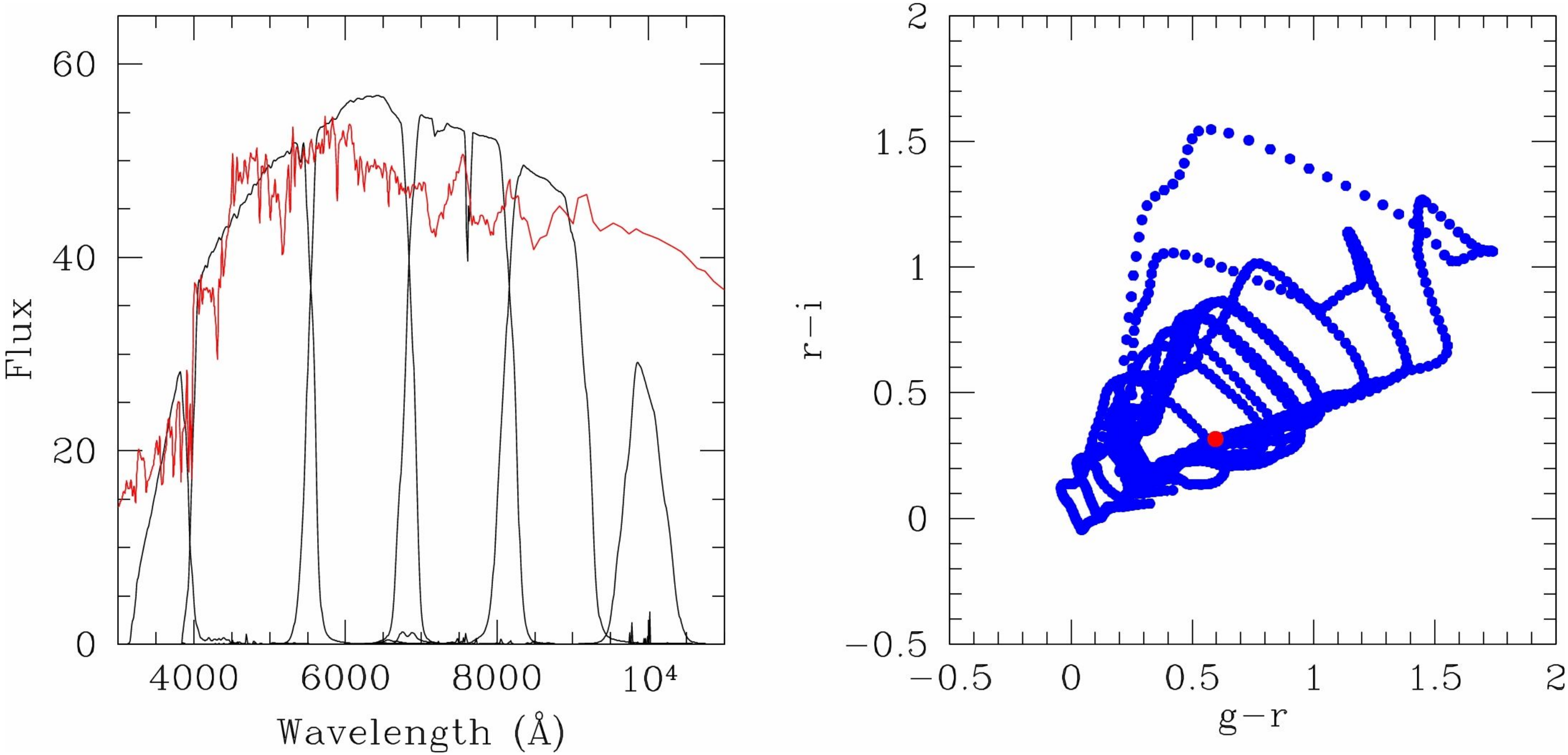


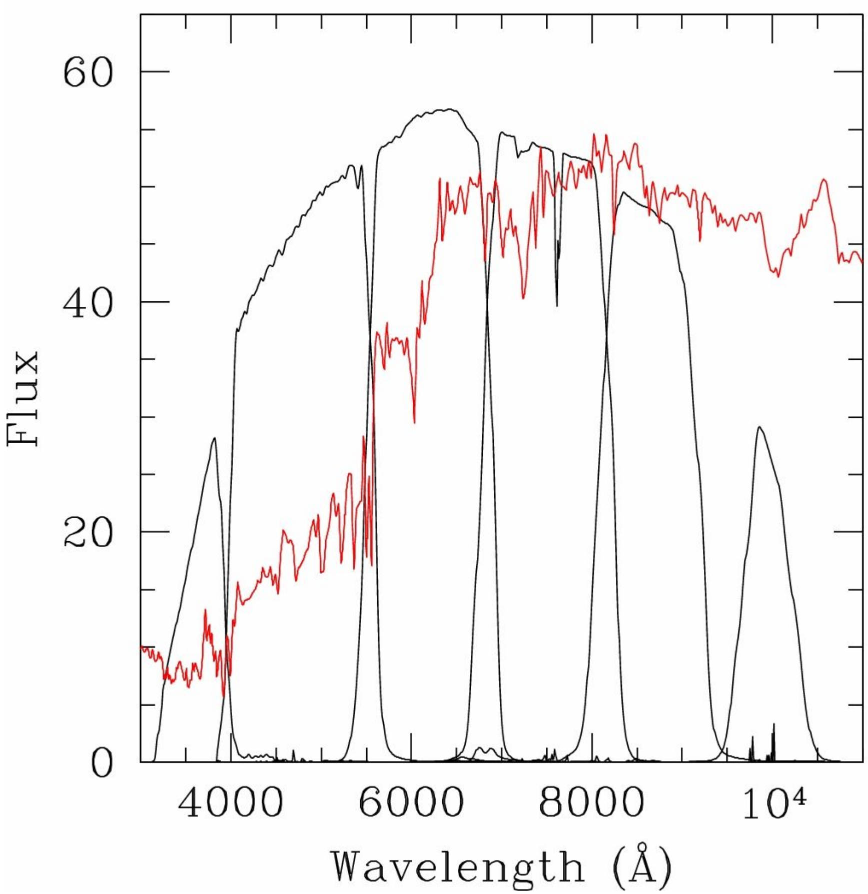
Fig 8.12 (S. Charlot) 'Galaxies in the Universe' Sparke/Gallagher CUP 2007

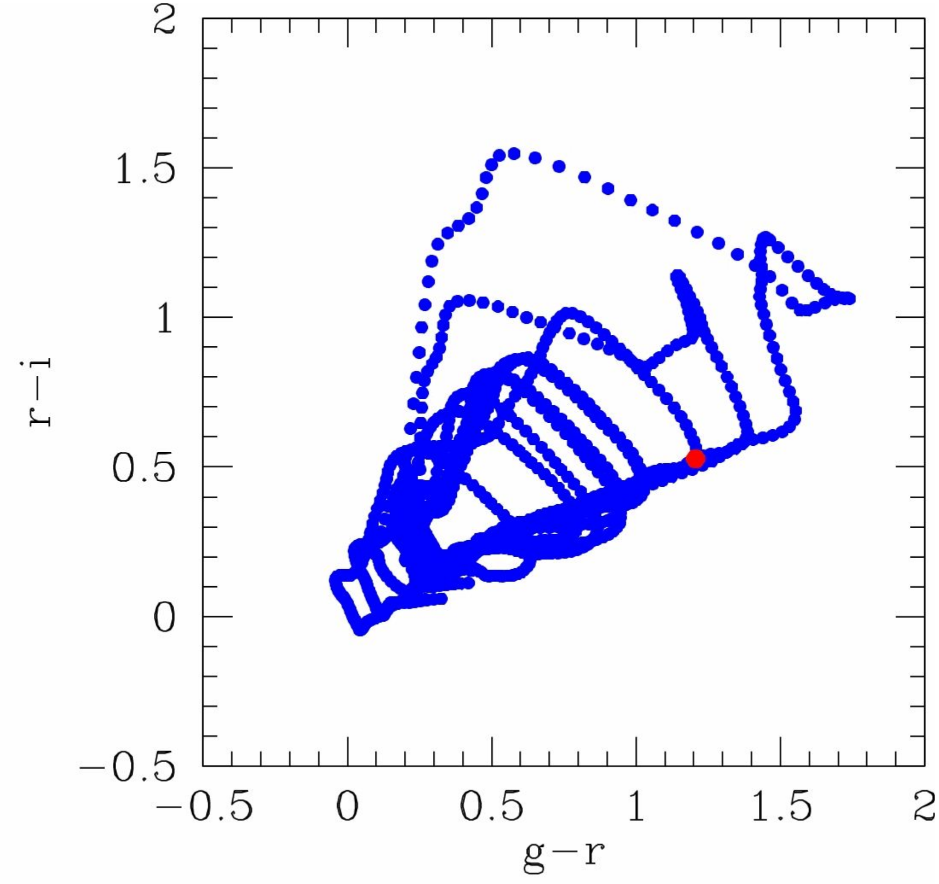
Photometric redshift (photo-z) estimation



credit:LSST

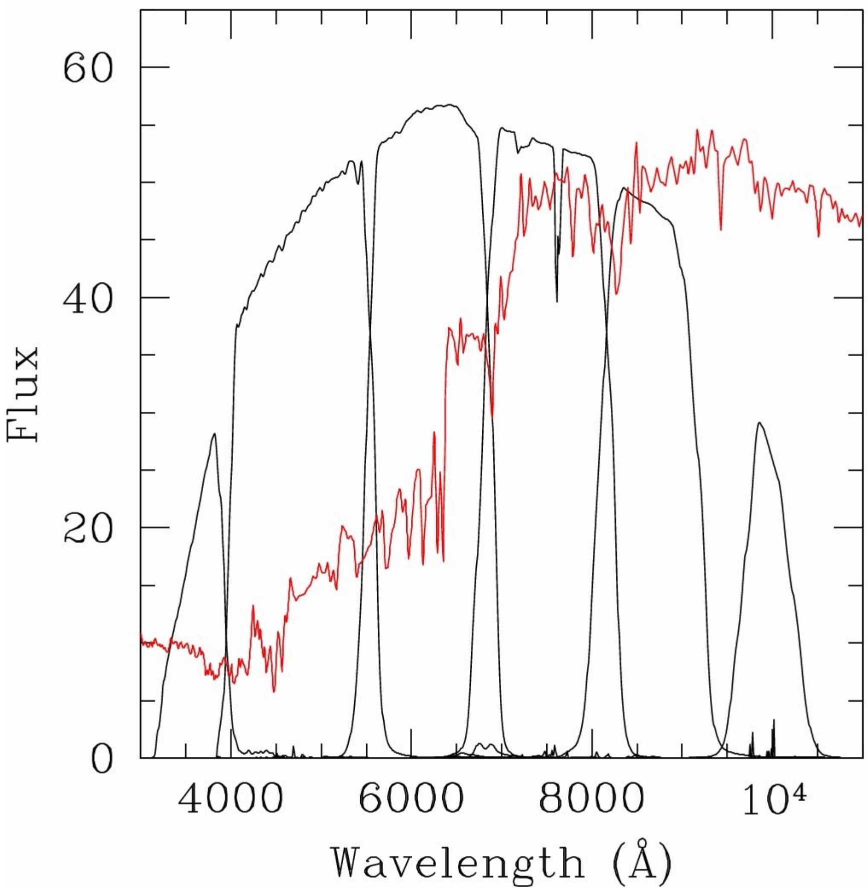
Photometric redshift (photo-z) estimation

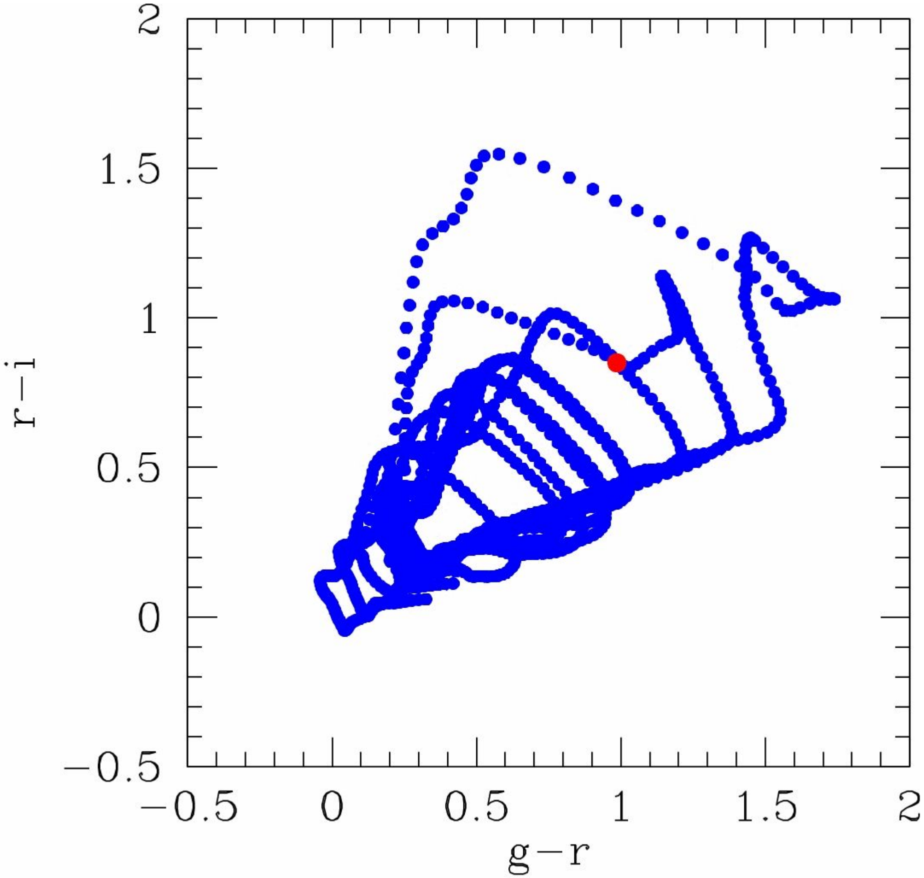




credit:LSST

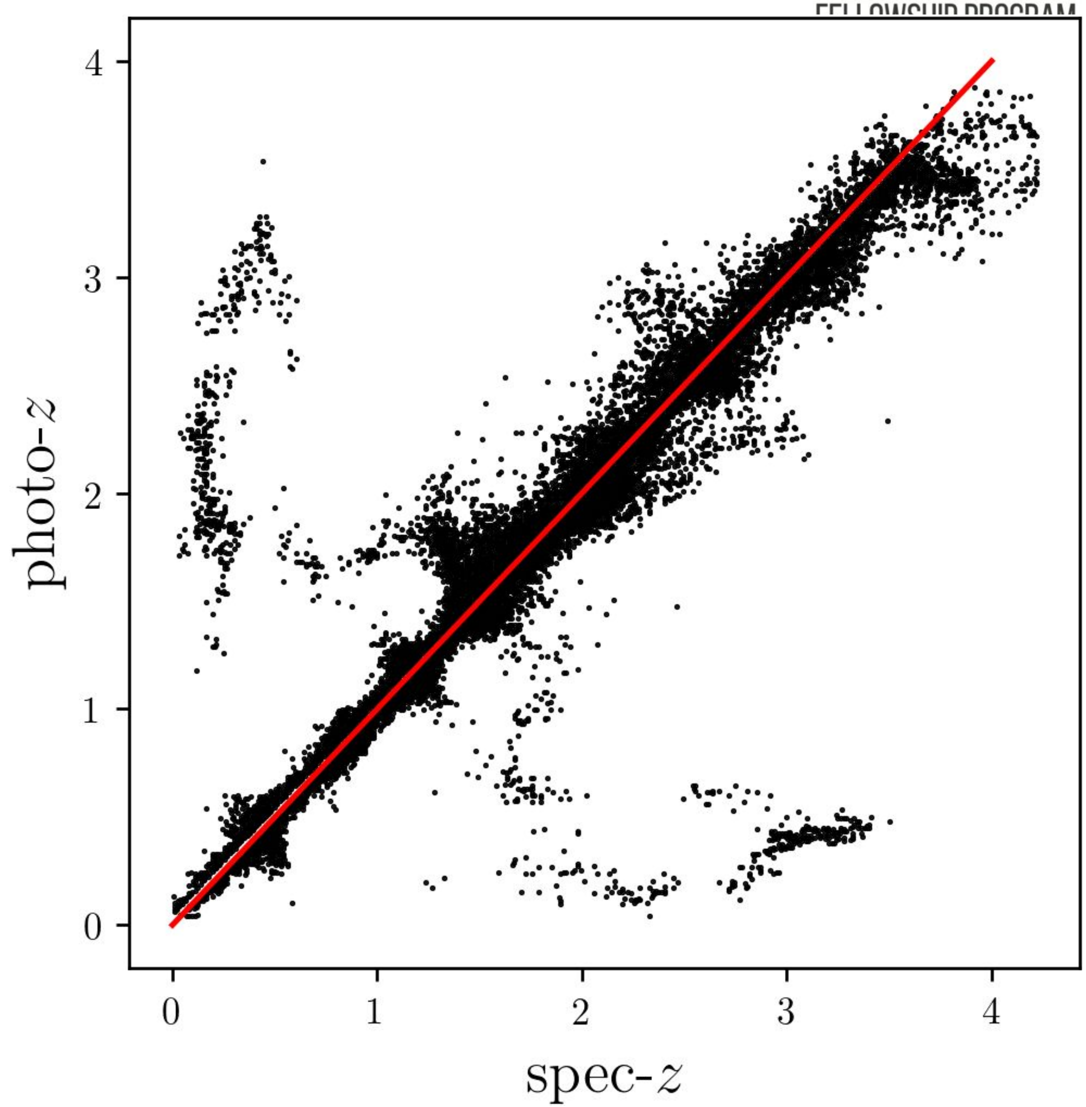
Photometric redshift (photo-z) estimation





credit:LSST

Photometric redshift (photo-z) estimation



arXiv:1501.07897

Traditional photo-*z* metrics

Intrinsic scatter

$$\sigma_z \;<\; 0.02(1+z)$$

Bias

$$\langle|z-\hat{z}|\rangle \;<\; 0.003(1+z)$$

Catastrophic outlier rate

$$\mathrm{N}_{|z-\hat{z}|>3\sigma_z} \;<\; 0.1\mathrm{N}_{\mathrm{LSST}}$$

https://docushare.lsstcorp.org/docushare/dsweb/Get/LPM-17

Traditional photo-*z* metrics

Intrinsic scatter

$$\sigma_z < 0.02(1+z)$$

Bias

$$\langle|z-\hat{z}|\rangle < 0.003(1+z)$$

Catastrophic outlier rate

$$\mathrm{N}_{|z-\hat{z}|>3\sigma_z} < 0.1\mathrm{N}_{\mathrm{LSST}}$$

https://docushare.lsstcorp.org/docushare/dsweb/Get/LPM-17

What do these metrics miss?

Intrinsic scatter

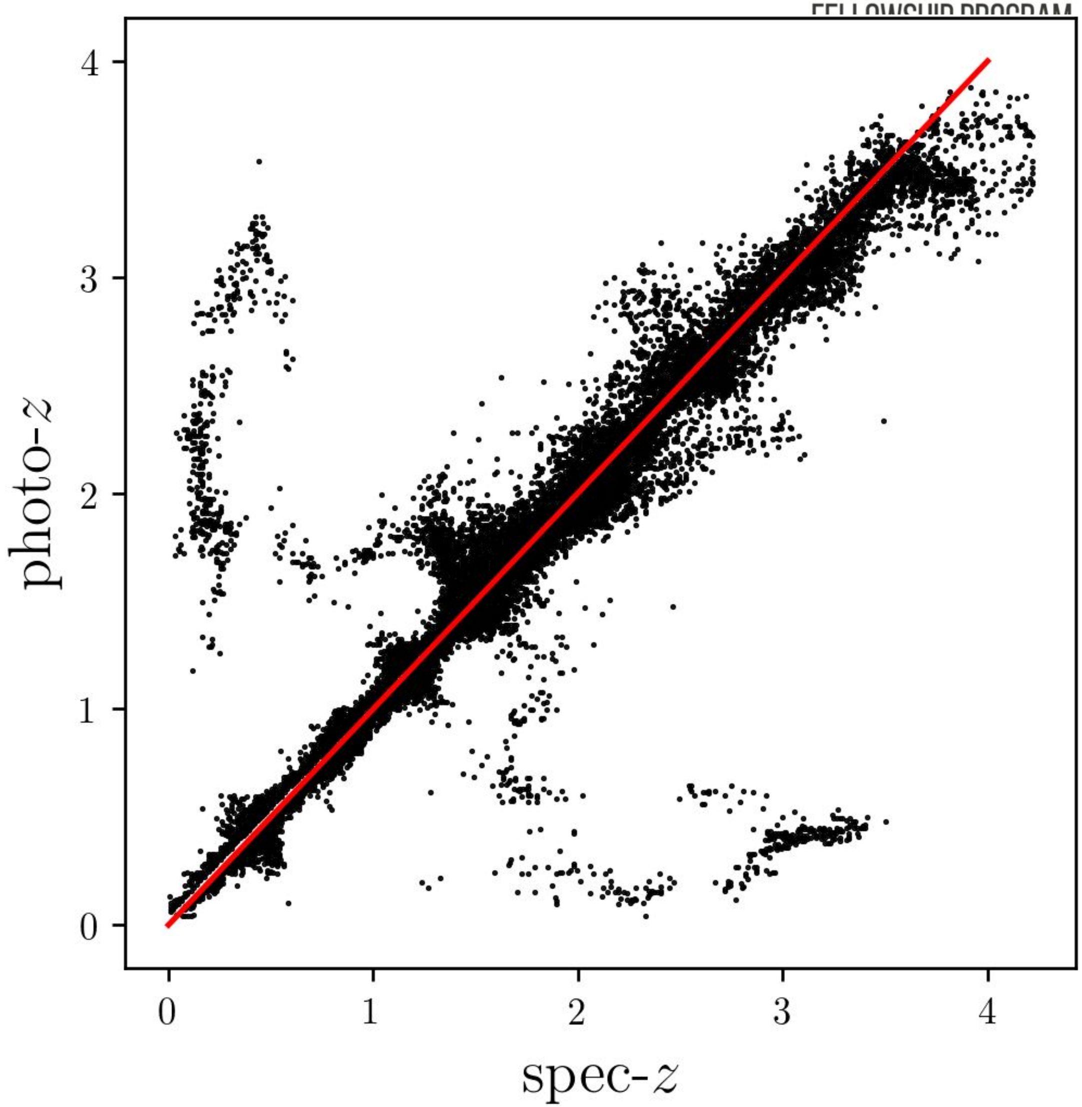
$$\sigma_z \ < \ 0.02(1+z)$$

Bias

$$\langle |z - \hat{z}| \rangle \ < \ 0.003(1+z)$$

Catastrophic outlier rate

$$\mathrm{N}_{|z-\hat{z}|>3\sigma_z} \ < \ 0.1\mathrm{N}_{\mathrm{LSST}}$$



arXiv:1501.07897

Conditional probability & forward models

The universe has true physical parameters θ.

There is a causal relationship p(x | θ) between data and the physical parameters; the parameters θ determine the data x.

We observe instances of data x generated by that model.

Conditional probability & forward models

The universe has true physical parameters θ.

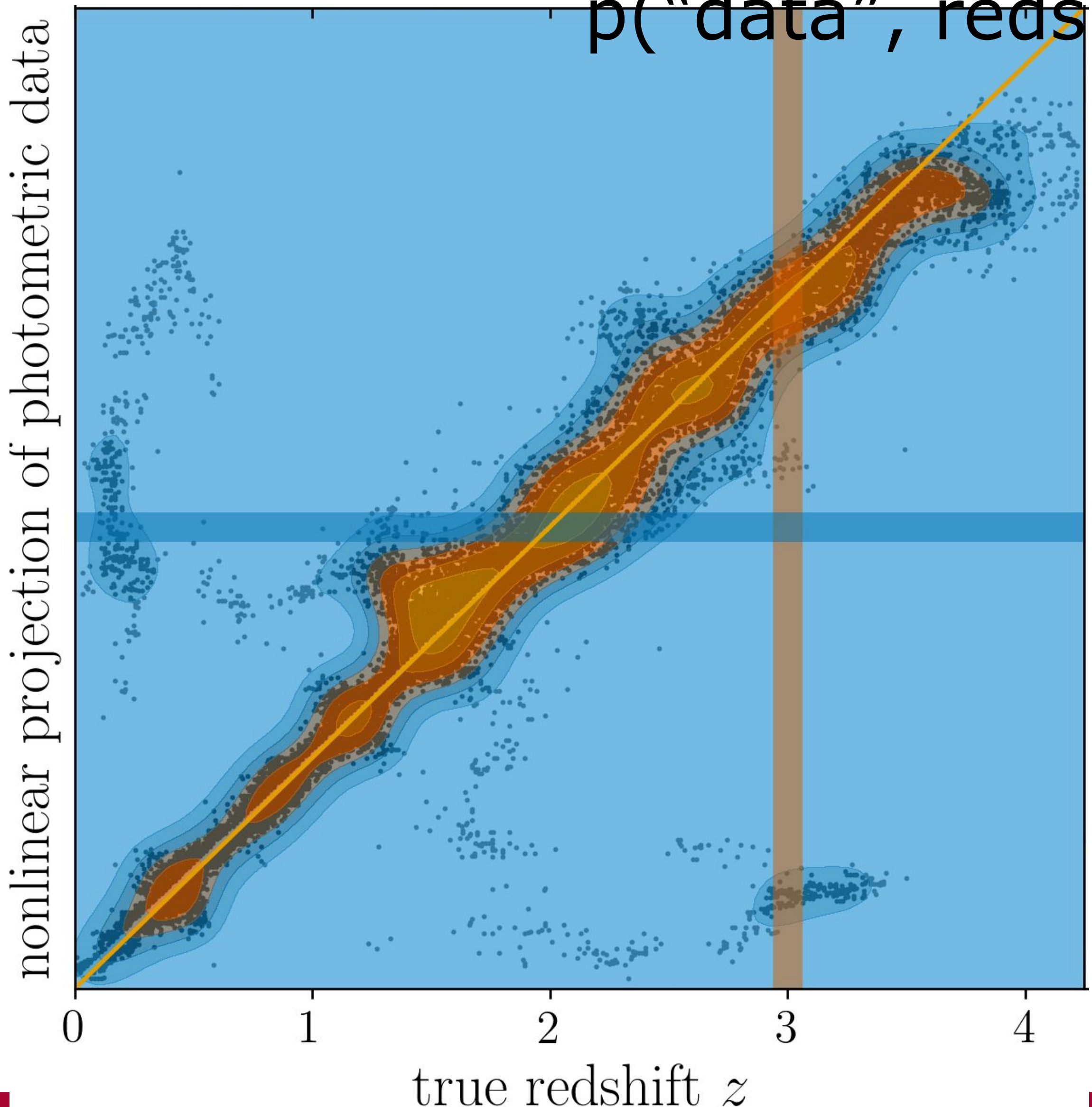
There is a causal relationship $p(x \mid \theta)$ between data and the physical parameters; the parameters θ determine the data x.

We observe instances of data x generated by that model.

We want to constrain the physical parameters θ that determined the observed data x, i.e. $p(\theta \mid x)$.

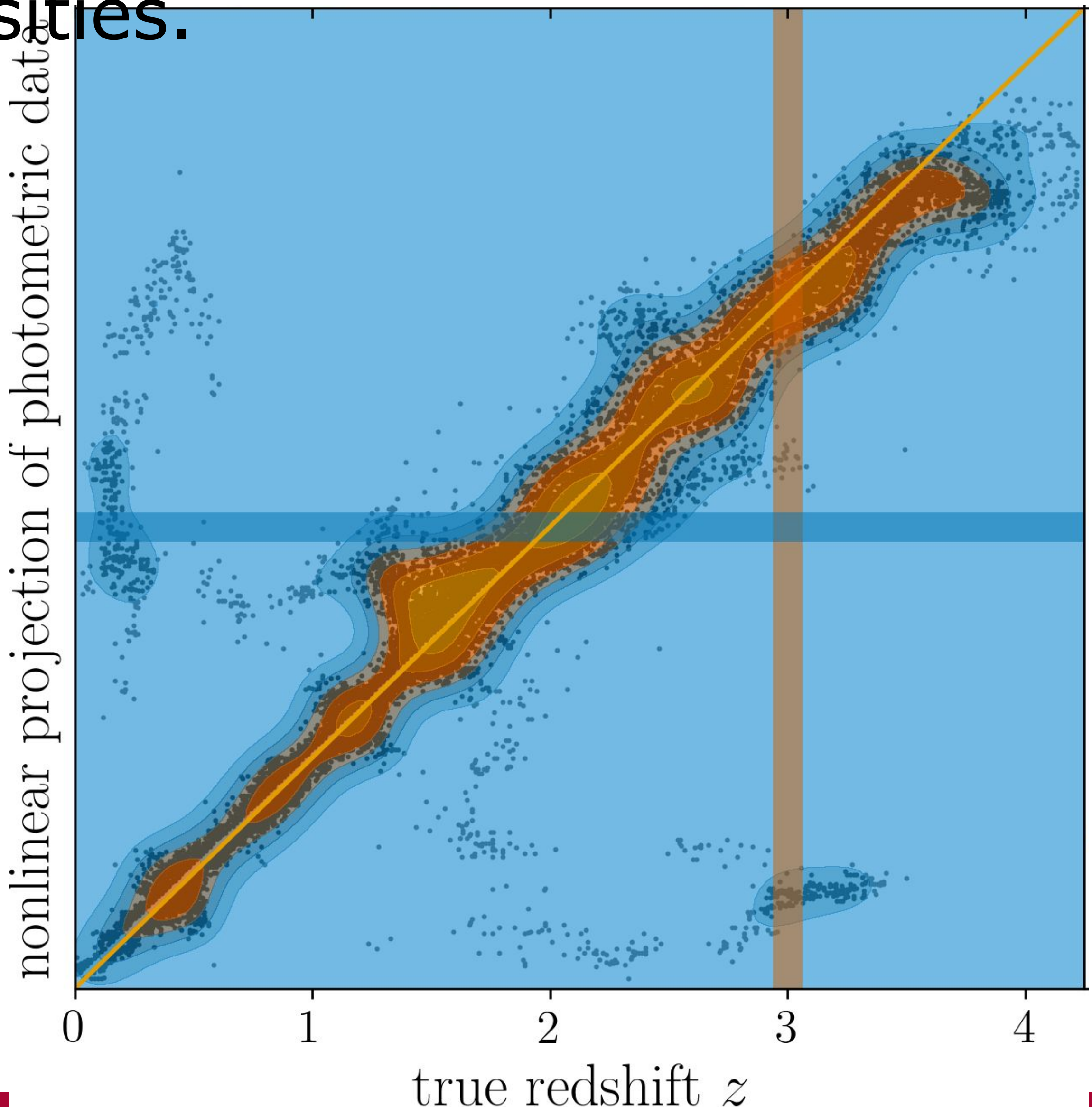
LSSTC DATA SCIENCE FELLOWSHIP PROGRAM

Photo-zs sample the joint PDF p("data", redshift)



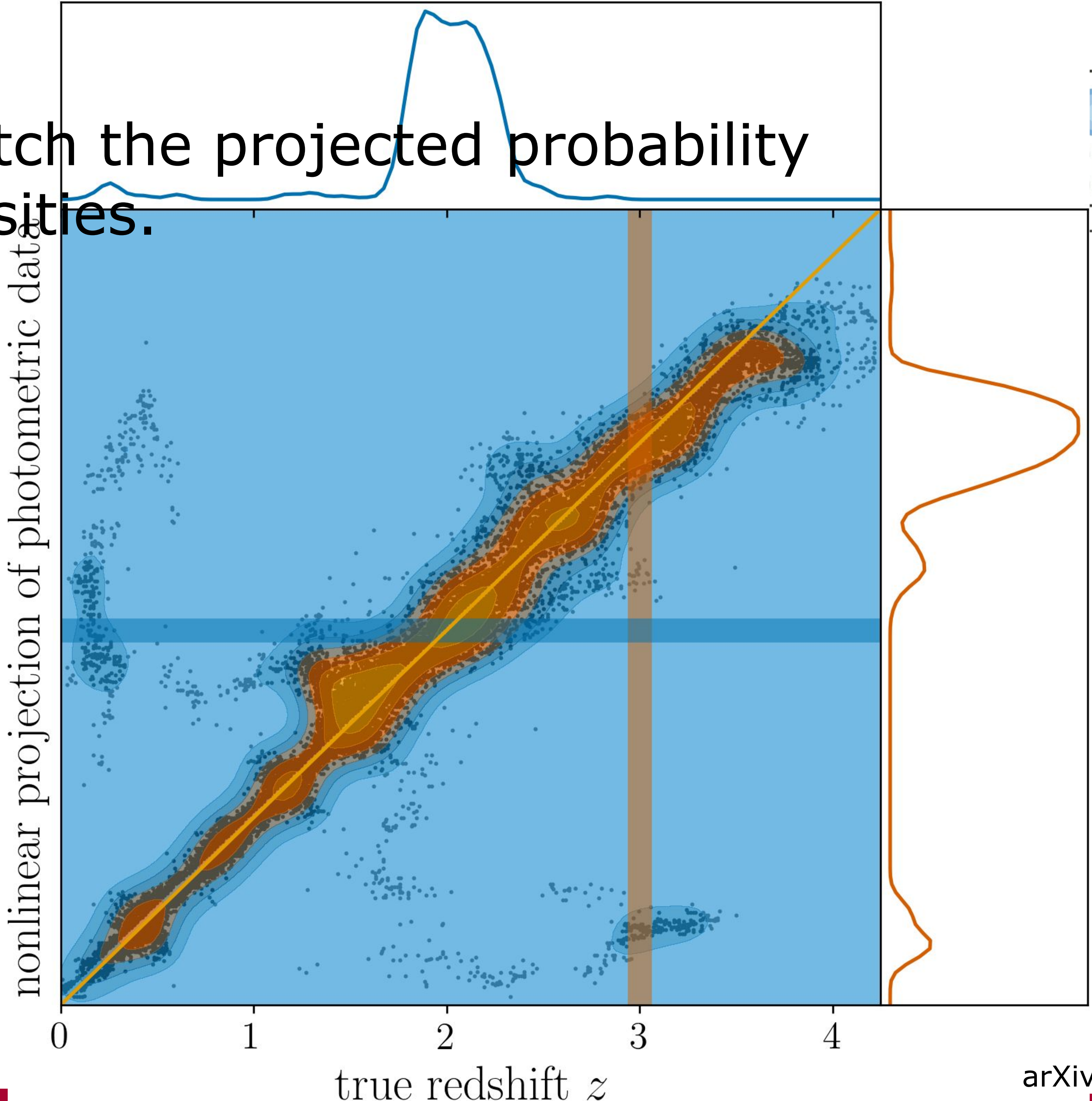
arXiv:2007.12178

Sketch the projected probability densities.



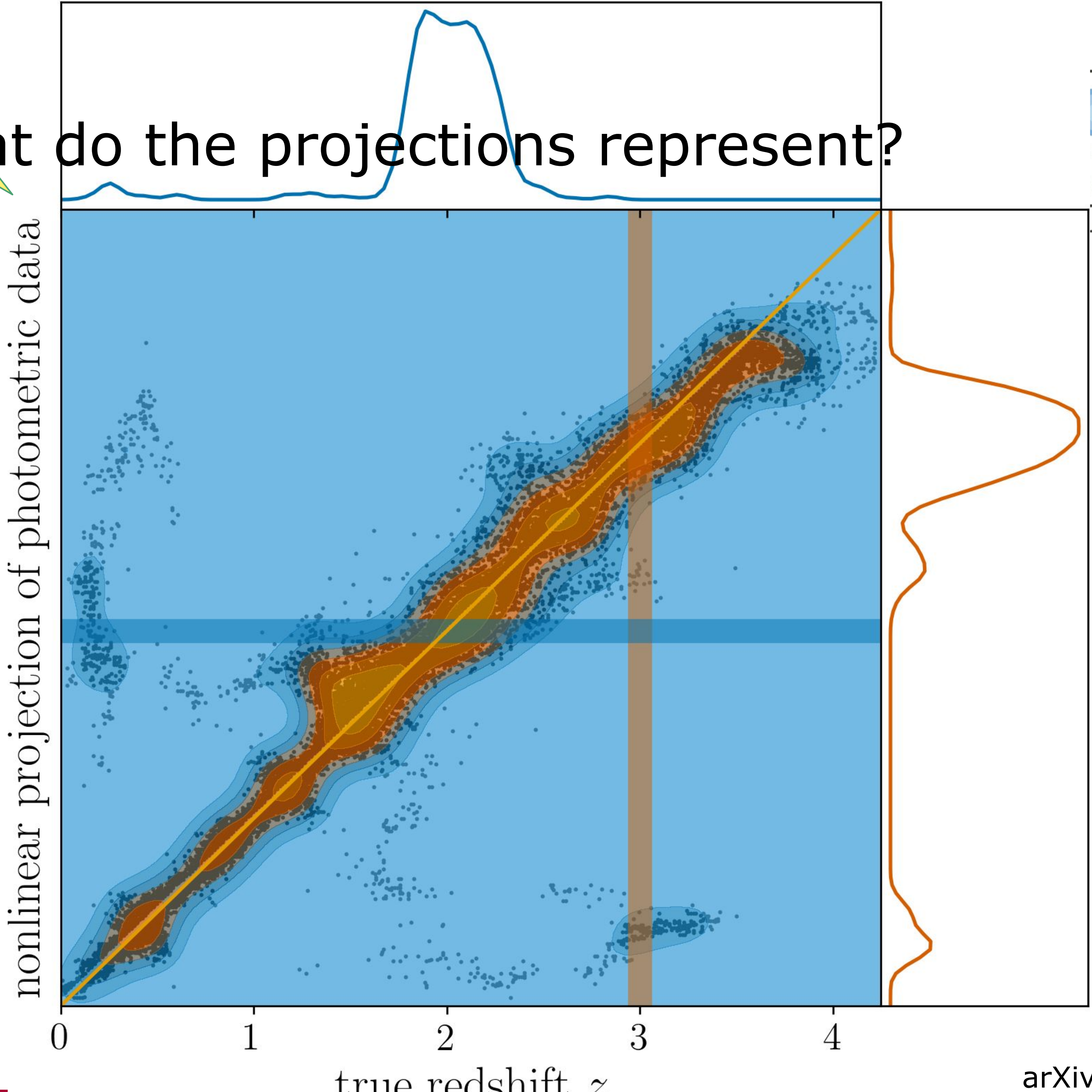
arXiv:2007.12178

Sketch the projected probability densities.



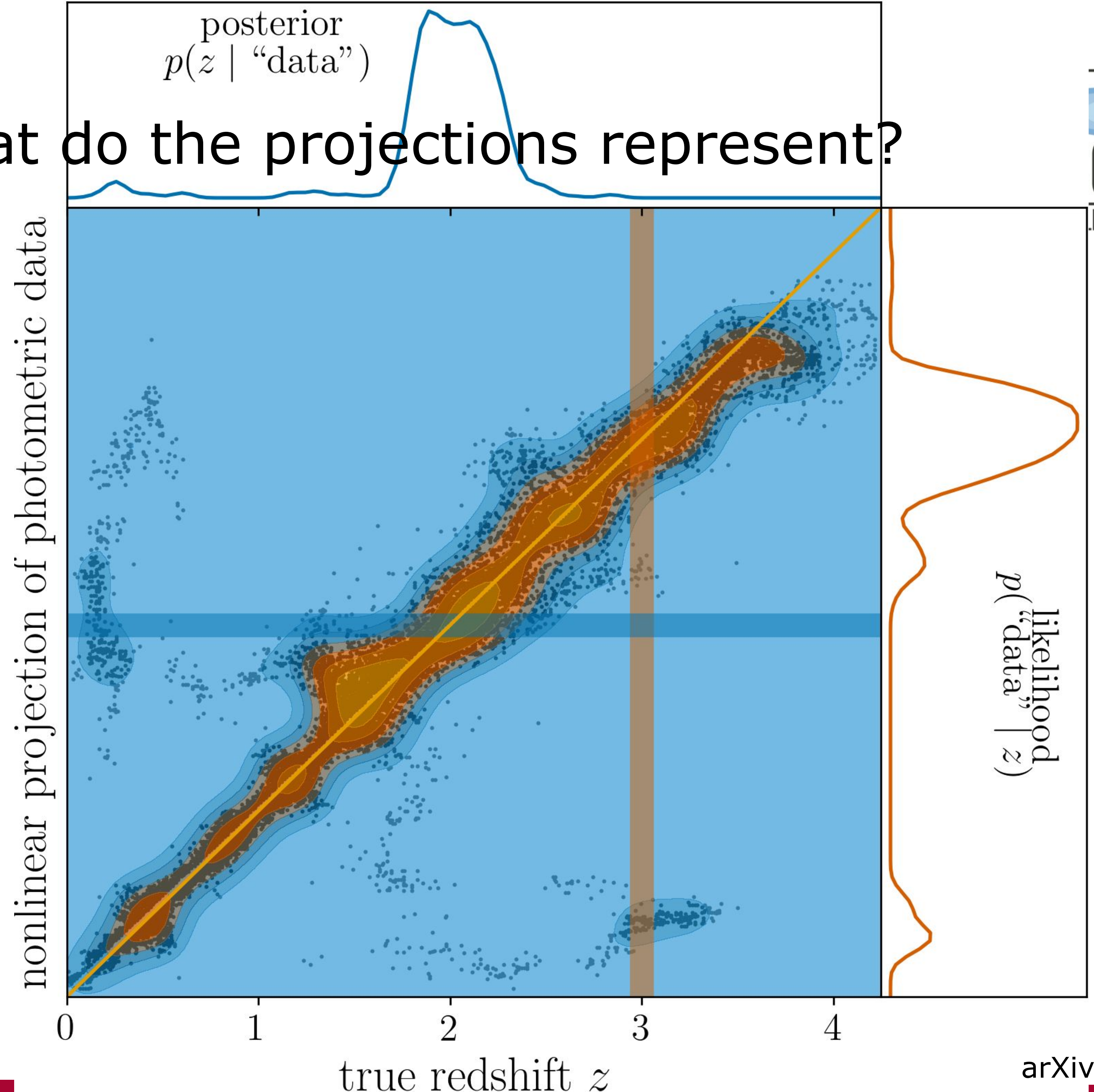
arXiv:2007.12178

What do the projections represent?



arXiv:2007.12178

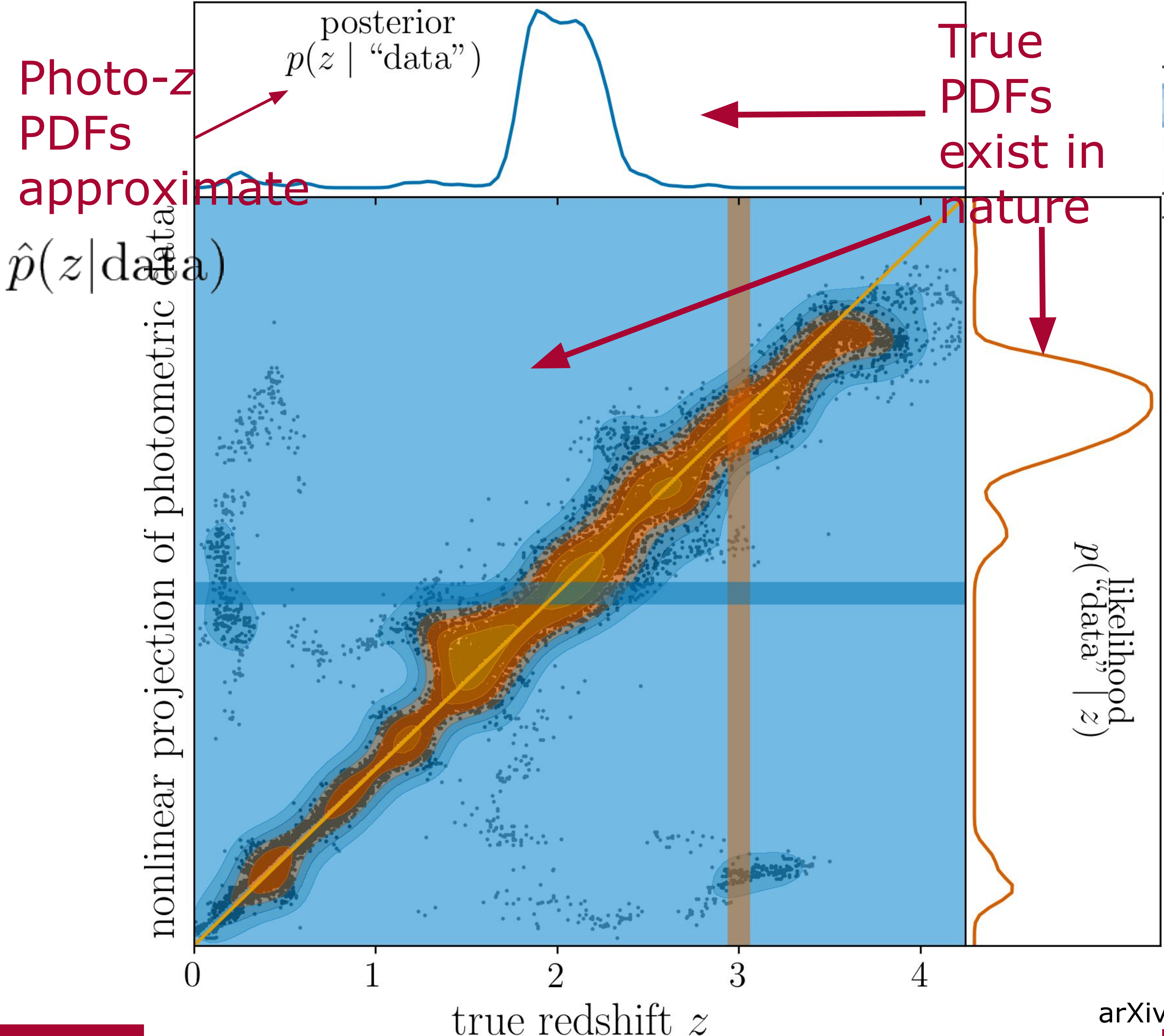
What do the projections represent?



LSSTC DATA SCIENCE FELLOWSHIP PROGRAM

arXiv:2007.12178

Photo-z PDFs approximate



LSSTC DATA SCIENCE FELLOWSHIP PROGRAM

arXiv:2007.12178

How do we compare PDFs?

Quantitative metrics of 1D PDFs

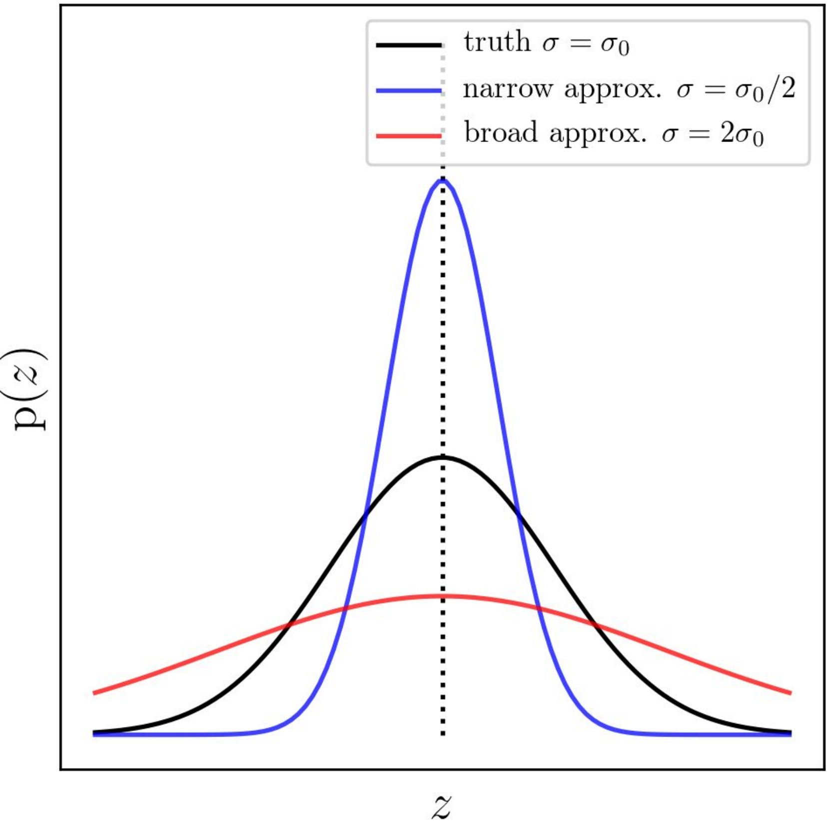
Root-mean-square Error (RMSE)

$$\mathrm{RMSE} = \sqrt{\int (p_{\mathrm{true}}(z) - \hat{p}_{\mathrm{est}}(z))^2 dz}$$

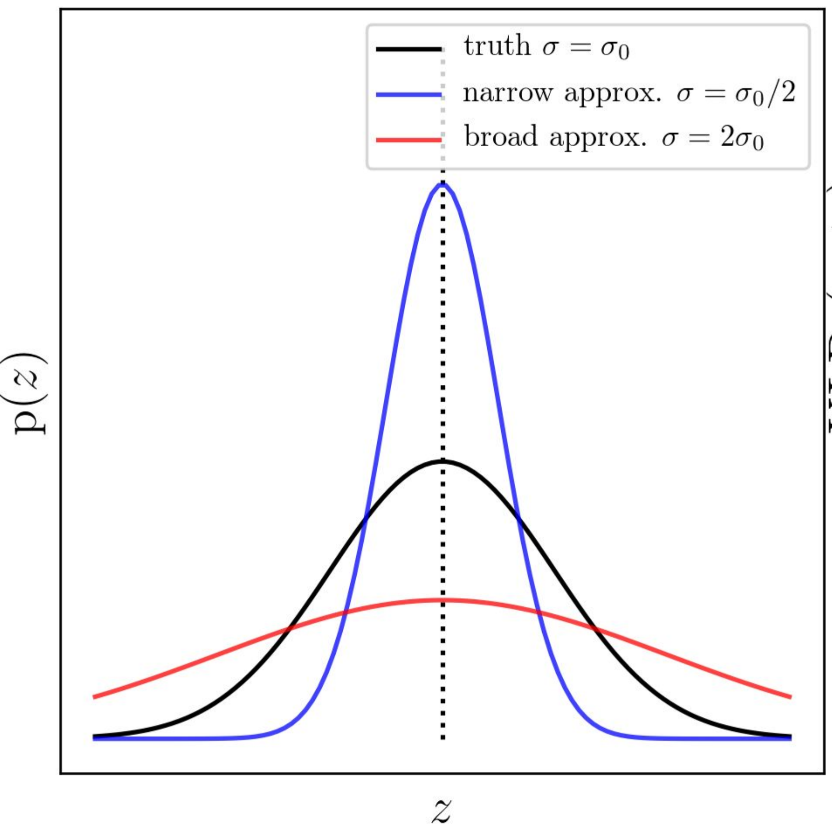
Kullback-Leibler Divergence (KLD)

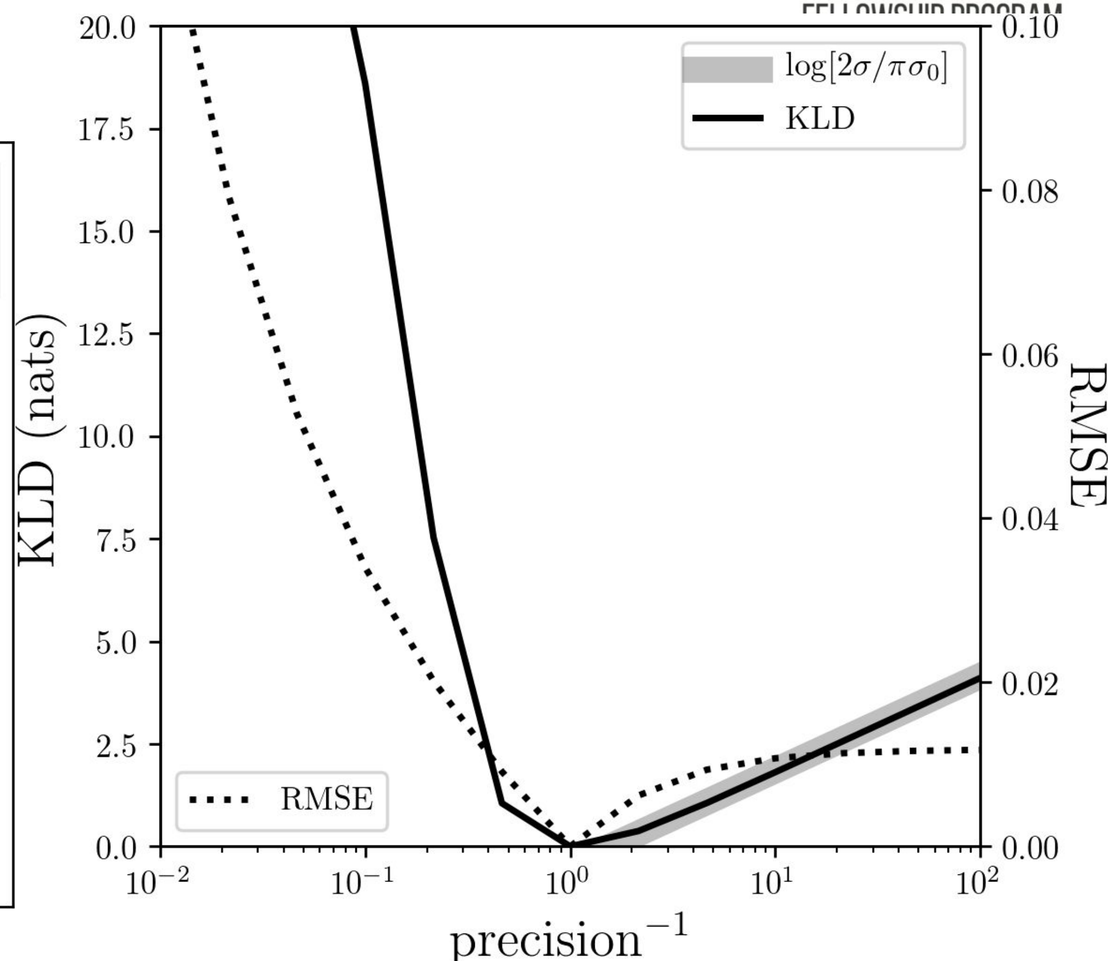
$$\mathrm{KLD}[\hat{p}_{\mathrm{est}}(z); p_{\mathrm{true}}(z)] = \int_{-\infty}^{\infty} p_{\mathrm{true}}(z) \ \log\left[\frac{p_{\mathrm{true}}(z)}{\hat{p}_{\mathrm{est}}(z)}\right] \ dz$$

Gaussian example: $\mathrm{precision}^{-1} = \sigma/\sigma_0$



Interpret the asymptotic behavior of the metrics.



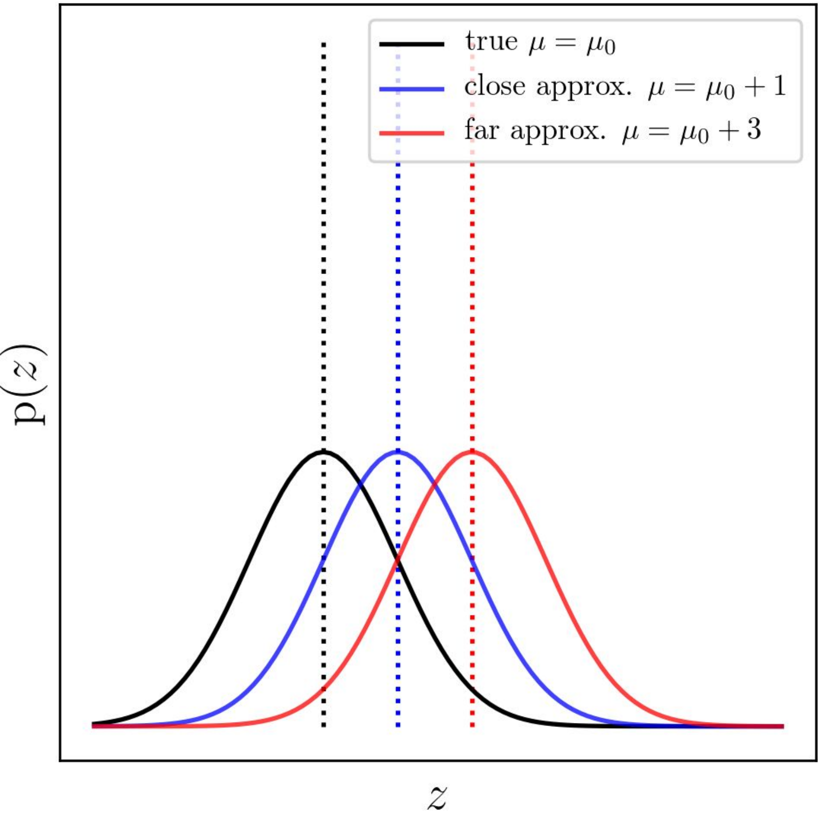


$$\text{precision}^{-1} = \sigma/\sigma_0$$

arXiv:1806.00014

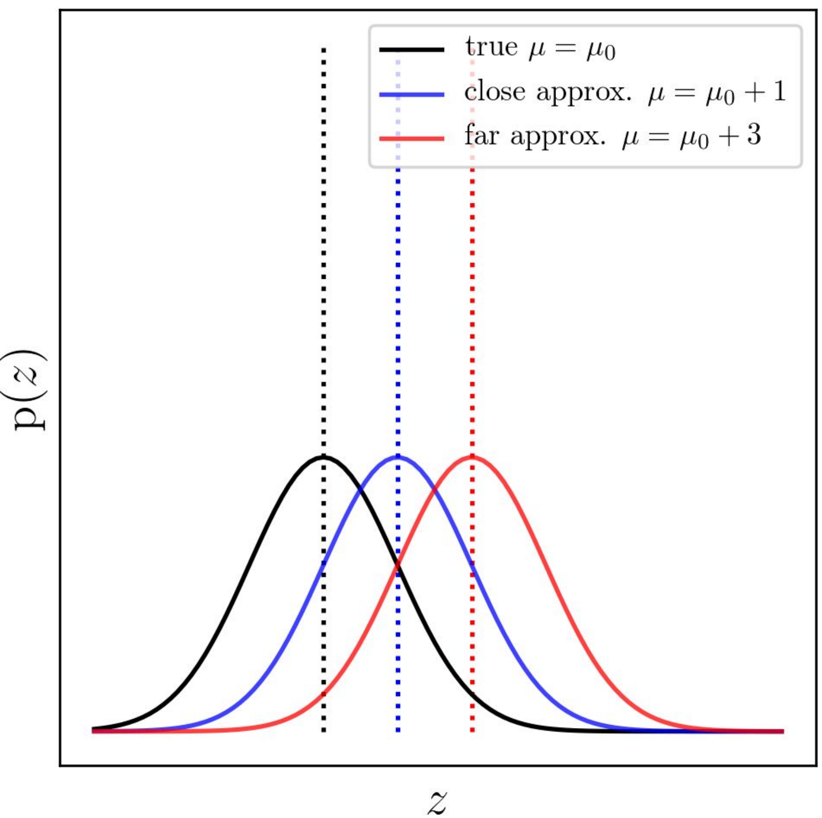
Gaussian example: $\text{tension} = \frac{|\mu - \mu_0|}{\sqrt{(\sigma_0^2 + \sigma^2)}}$

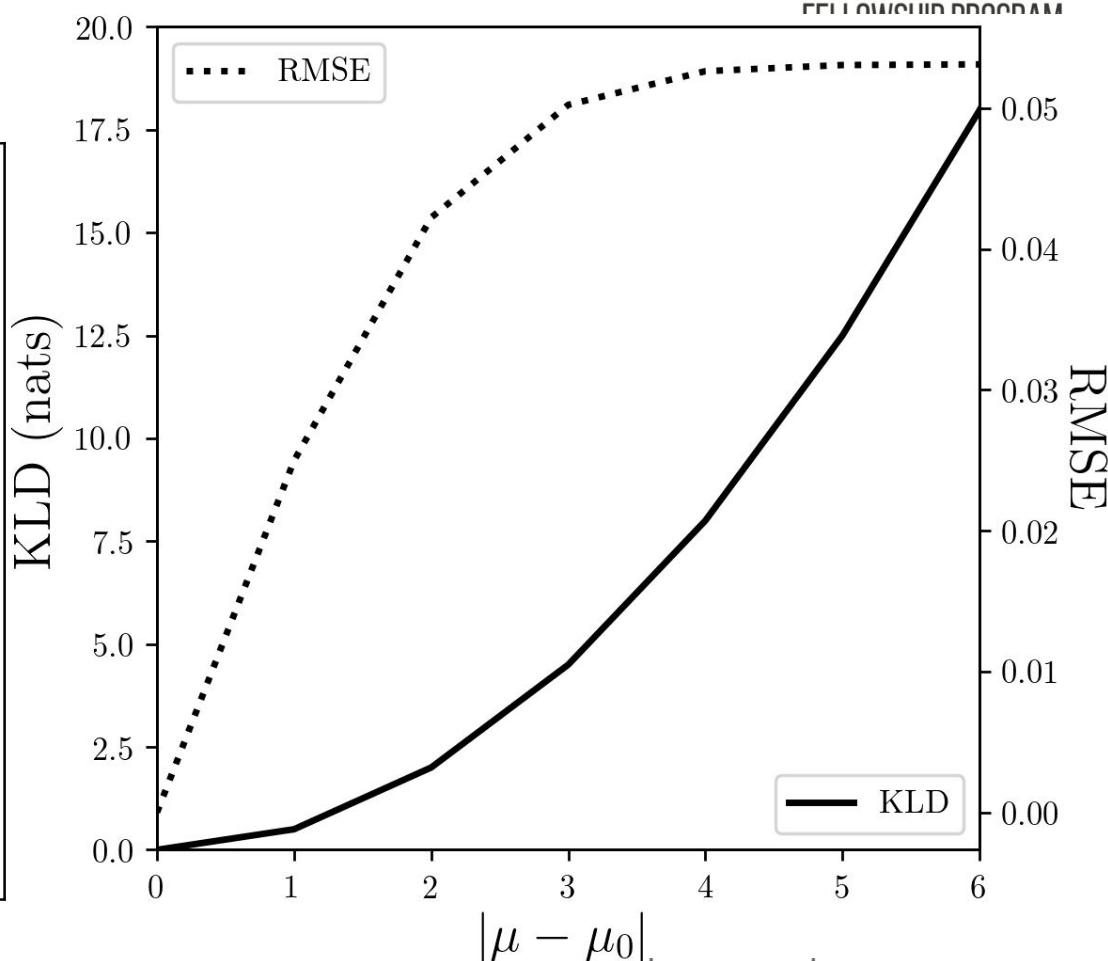
LSSTC DATA SCIENCE FELLOWSHIP PROGRAM



credit:A.I.Malz

Interpret the asymptotic behavior of the metrics.





$$\text{tension} = \frac{|\mu - \mu_0|}{\sqrt{\left(\sigma_0^2 + \sigma^2\right)}}$$

arXiv:1806.00014

What does this have to do with Bayesian model comparison?

Example: PZ DC1

The PZ DC1 experiment

Motivation: identify the best photo-z posterior code for LSST-DESC

Data: cosmological redshifts & photometry catalog painted on N-body simulation

Control: idealized, shared prior information

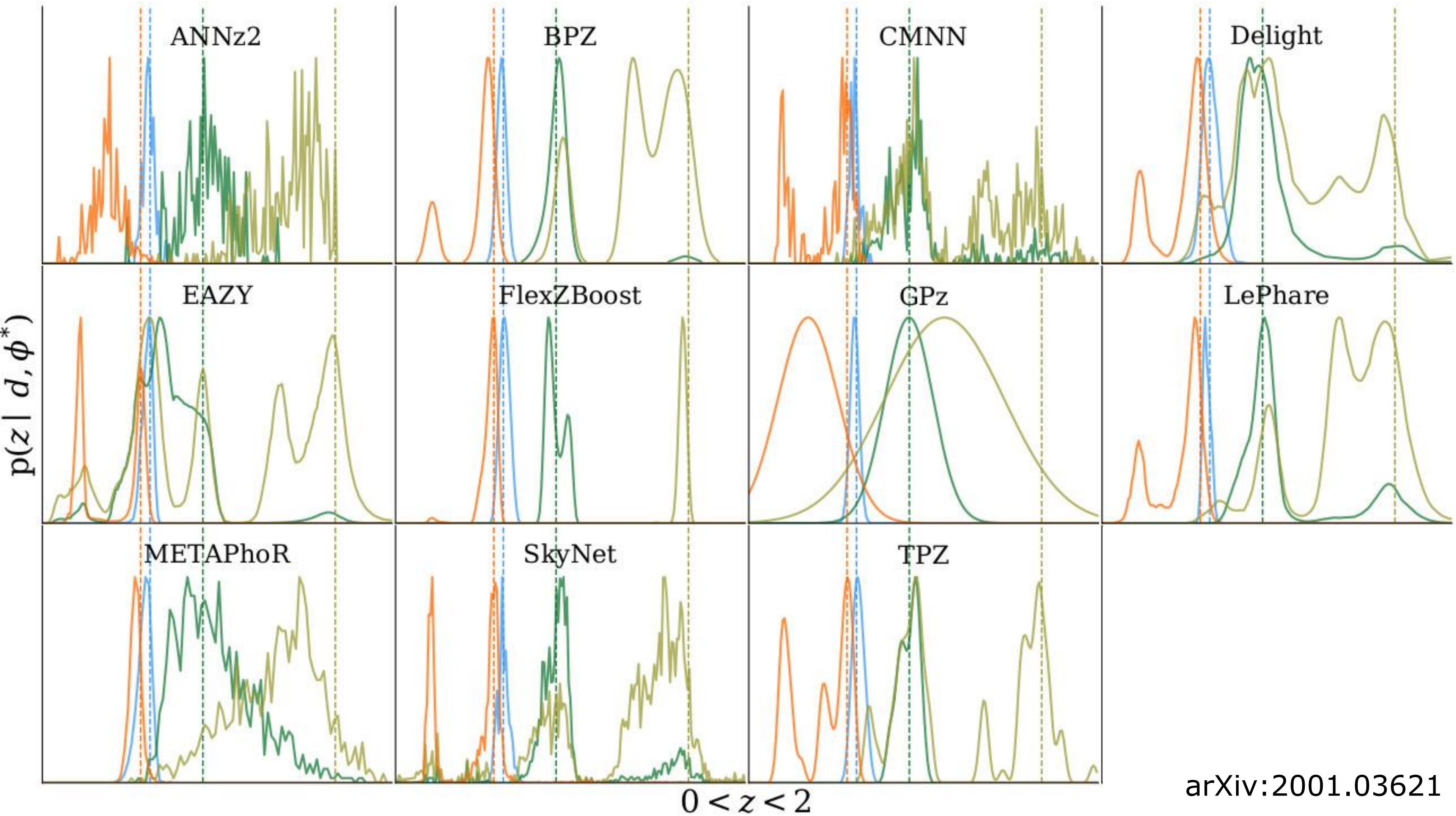
The PZ DC1 experiment

Motivation: identify the best photo-z posterior code for LSST-DESC

Data: cosmological redshifts & photometry catalog painted on N-body simulation

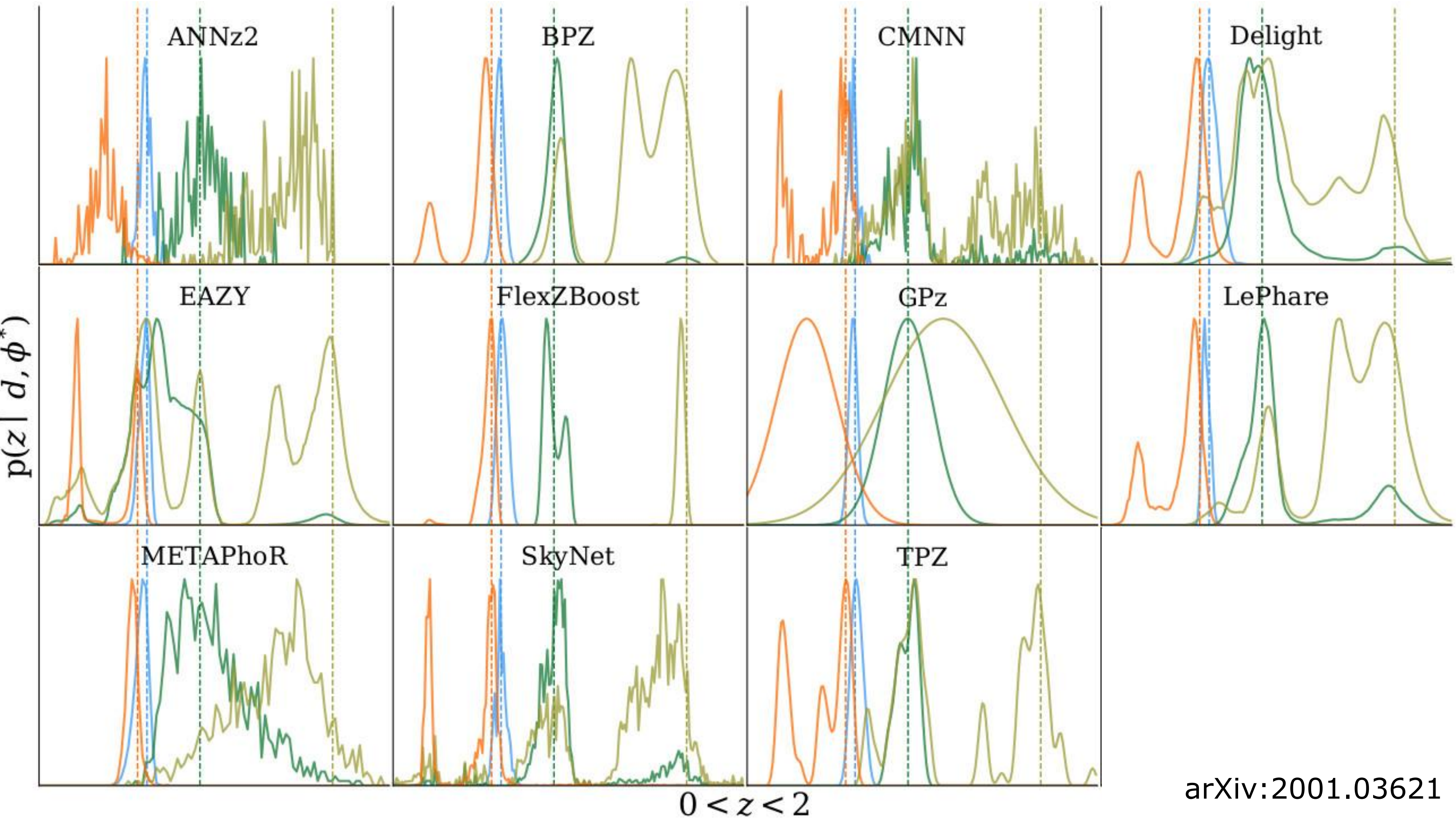
Control: idealized, shared prior information

Suggest possible causes for differences between photo-*z* posterior estimates.



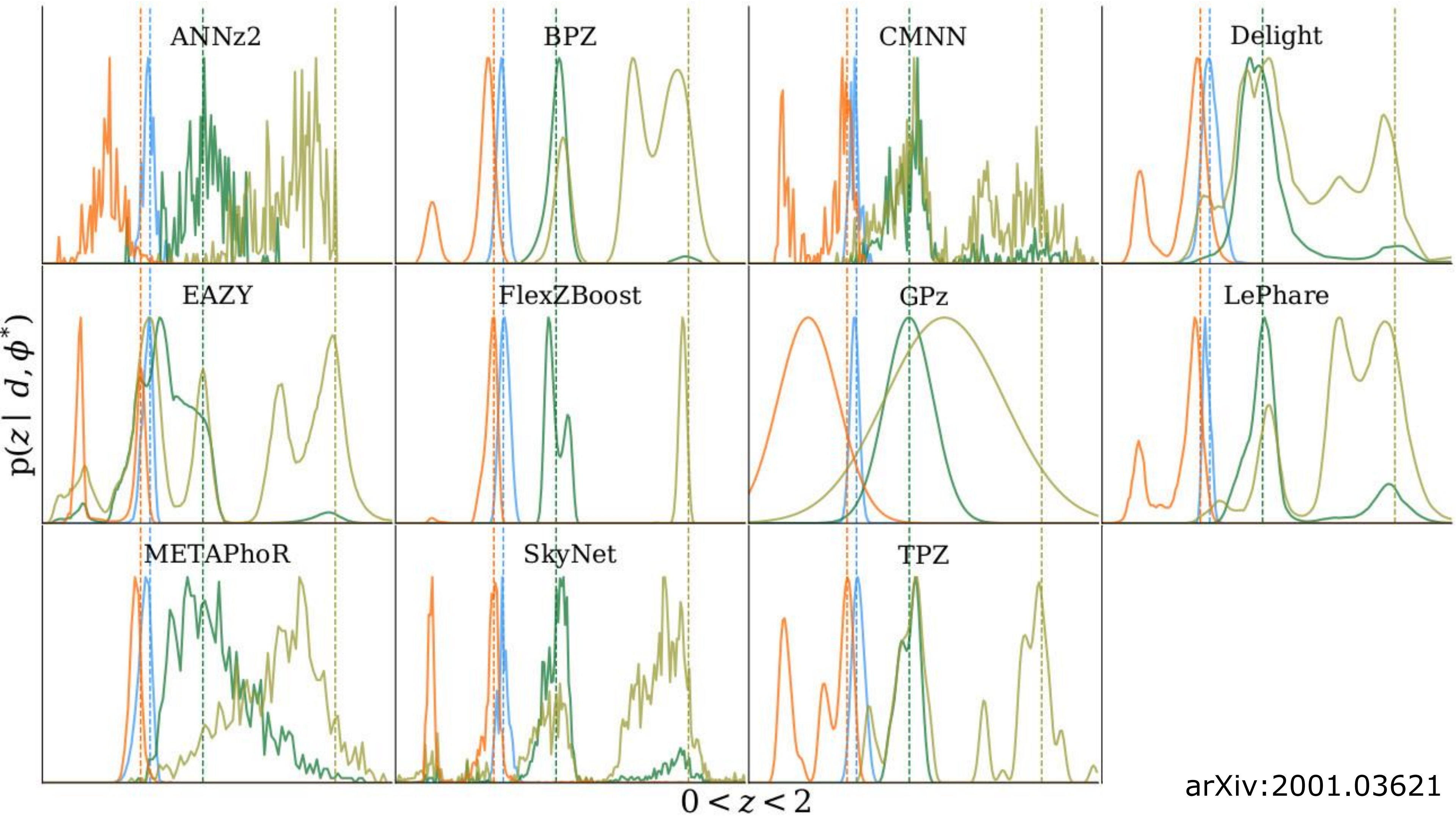
arXiv:2001.03621

Suggest possible causes for differences between photo-*z* posterior estimates.



arXiv:2001.03621

The photo-z posterior estimator is a model, distinct from the assumed prior.



arXiv:2001.03621

Quantitative metrics of 1D PDFs

Root-mean-square Error (RMSE)

$$\mathrm{RMSE} = \sqrt{\int (p_{\mathrm{true}}(z) - \hat{p}_{\mathrm{est}}(z))^2 dz}$$

Kullback-Leibler Divergence (KLD)

$$\mathrm{KLD}[\hat{p}_{\mathrm{est}}(z); p_{\mathrm{true}}(z)] = \int_{-\infty}^{\infty} p_{\mathrm{true}}(z) \ \log\left[\frac{p_{\mathrm{true}}(z)}{\hat{p}_{\mathrm{est}}(z)}\right] \ dz$$

The PZ DC1 experiment

LSSTC DATA SCIENCE FELLOWSHIP PROGRAM

Motivation: identify the best photo-z posterior code for LSST-DESC

Data: cosmological redshifts & photometry catalog painted on N-body simulation

Control: idealized, shared prior information

The PZ DC1 experiment

LSSTC DATA SCIENCE FELLOWSHIP PROGRAM

The universe has true physical parameters θ.

↓

There is a causal relationship p(x | θ) between data and the physical parameters; the parameters θ determine the data x.

↓

We observe instances of data x generated by that model.

Motivation: identify the best photo-z posterior code for LSST-DESC

Data: cosmological redshifts & photometry catalog painted on N-body simulation

Control: idealized, shared prior information

Quantitative metrics of 1D PDFs

Root-mean-square Error (RMSE)

$$\mathrm{RMSE} = \sqrt{\int (p_{\mathrm{true}}(z) - \hat{p}_{\mathrm{est}}(z))^2 dz}$$

Kullback-Leibler Divergence (KLD)

$$\mathrm{KLD}[\hat{p}_{\mathrm{est}}(z); p_{\mathrm{true}}(z)] = \int_{-\infty}^{\infty} p_{\mathrm{true}}(z)\ \log\left[\frac{p_{\mathrm{true}}(z)}{\hat{p}_{\mathrm{est}}(z)}\right]\ dz$$

No true posteriors available!

Quantitative metrics of 1D PDF ensembles

Root-mean-square Error (RMSE)

$$\mathrm{RMSE} = \sqrt{\int (p_{\mathrm{true}}(z) - \hat{p}_{\mathrm{est}}(z))^2 dz}$$

Kullback-Leibler Divergence (KLD)

$$\mathrm{KLD}[\hat{p}_{\mathrm{est}}(z); p_{\mathrm{true}}(z)] = \int_{-\infty}^{\infty} p_{\mathrm{true}}(z) \ \log\left[\frac{p_{\mathrm{true}}(z)}{\hat{p}_{\mathrm{est}}(z)}\right] \ dz$$

Cumulative Distribution Function (CDF)

$$\mathrm{CDF}[\hat{p}, z'] \equiv \int_{-\infty}^{z'} \hat{p}(z) dz$$

No true posteriors available!

Quantitative metrics of 1D PDF ensembles

Root-mean-square Error (RMSE)

$$\mathrm{RMSE} = \sqrt{\int (p_{\mathrm{true}}(z) - \hat{p}_{\mathrm{est}}(z))^2 dz}$$

Kullback-Leibler Divergence (KLD)

$$\mathrm{KLD}[\hat{p}_{\mathrm{est}}(z); p_{\mathrm{true}}(z)] = \int_{-\infty}^{\infty} p_{\mathrm{true}}(z)\ \log\left[\frac{p_{\mathrm{true}}(z)}{\hat{p}_{\mathrm{est}}(z)}\right]\ dz$$

No true posteriors available!

Cumulative Distribution Function (CDF)

$$\mathrm{CDF}[\hat{p}, z'] \equiv \int_{-\infty}^{z'} \hat{p}(z) dz$$

Probability Integral Transform (PIT)

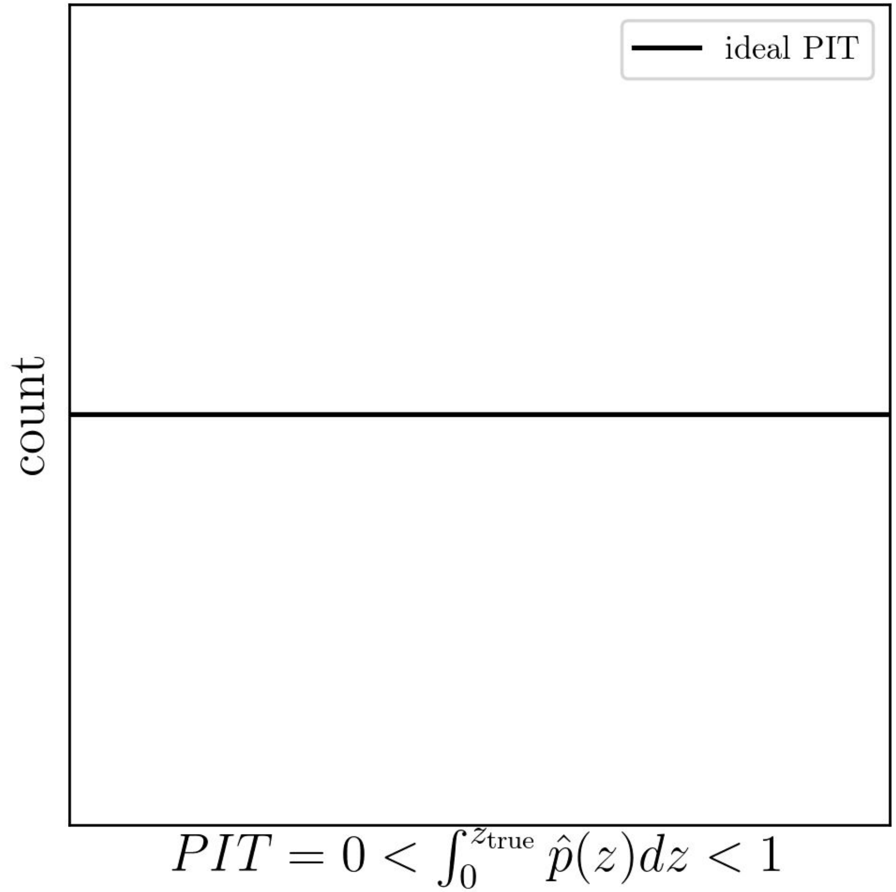
$$P(\mathrm{PIT} \equiv \mathrm{CDF}[\hat{p}, z_{true}])$$

Sketch the PIT histogram of an ideal photo-*z* posterior estimator.

$$\mathrm{CDF}[\hat{p}, z'] \equiv \int_{-\infty}^{z'} \hat{p}(z) dz$$

$$P(\mathrm{PIT} \equiv \mathrm{CDF}[\hat{p}, z_{\mathrm{true}}])$$

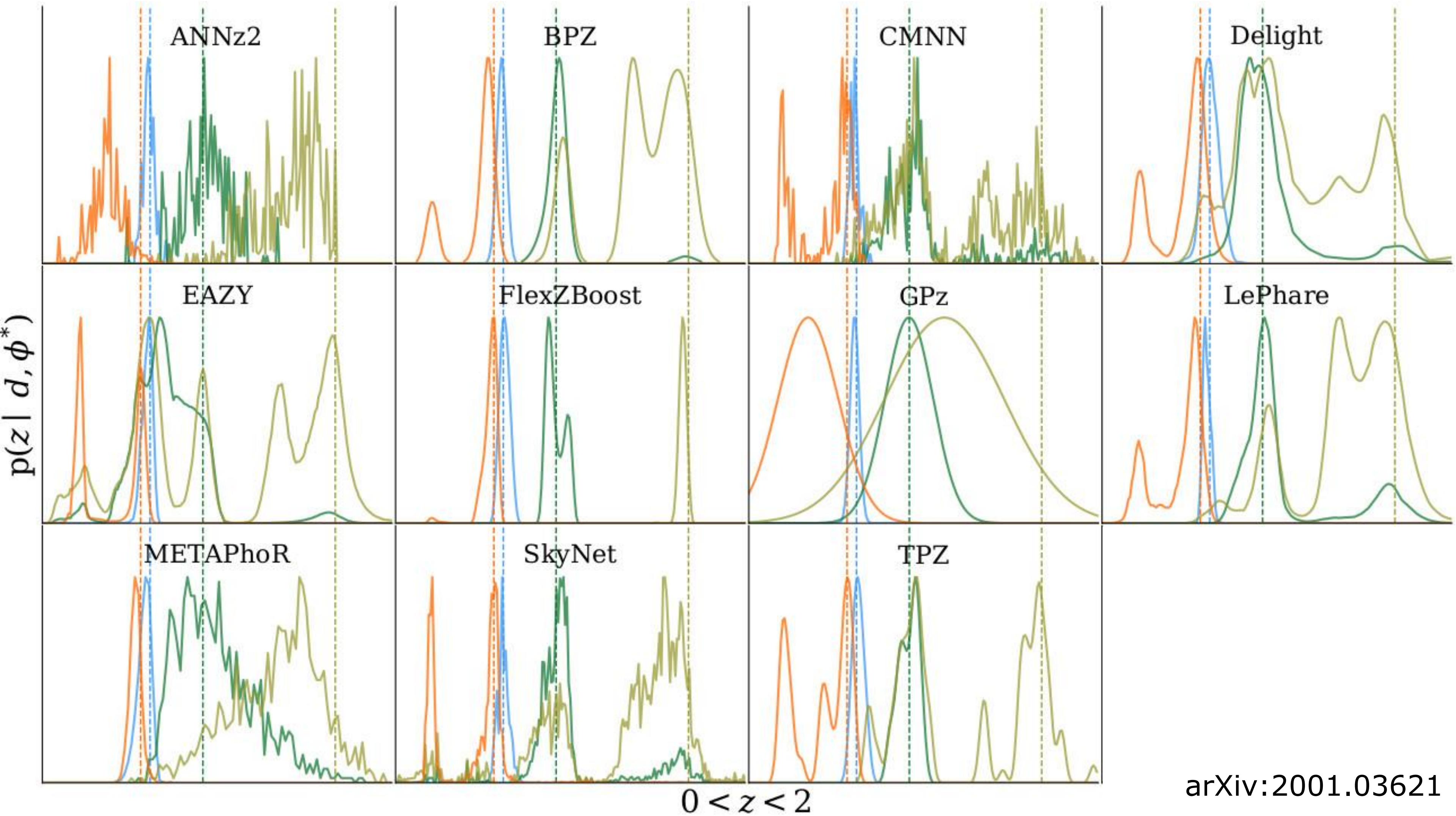
Sketch the PIT histogram of an ideal photo-*z* posterior estimator.



$$\mathrm{CDF}[\hat{p}, z'] \equiv \int_{-\infty}^{z'} \hat{p}(z)dz$$

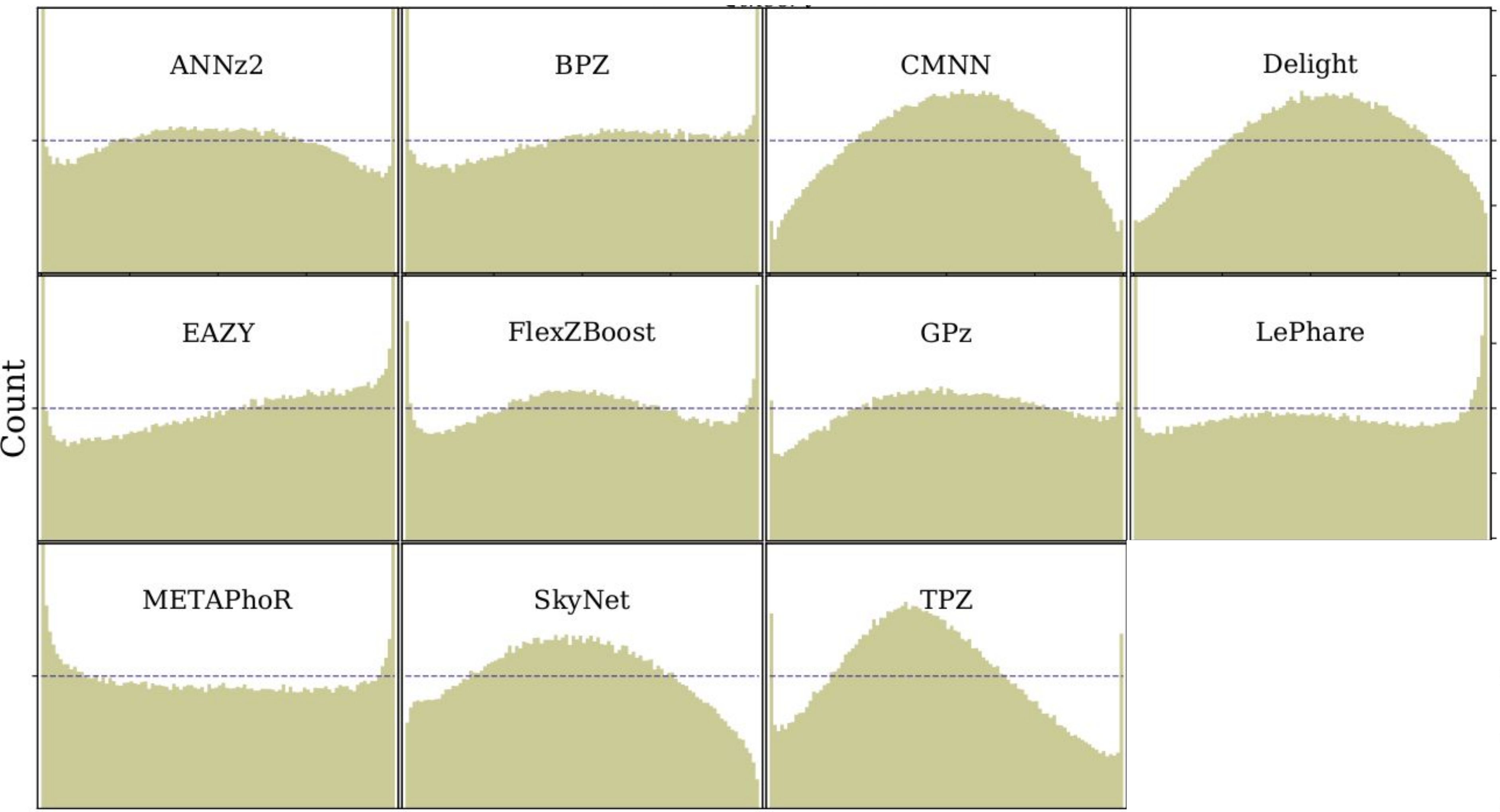
$$P(\mathrm{PIT} \equiv \mathrm{CDF}[\hat{p}, z_{true}])$$

What might the PIT look like for these photo-*z* posterior estimators?



arXiv:2001.03621

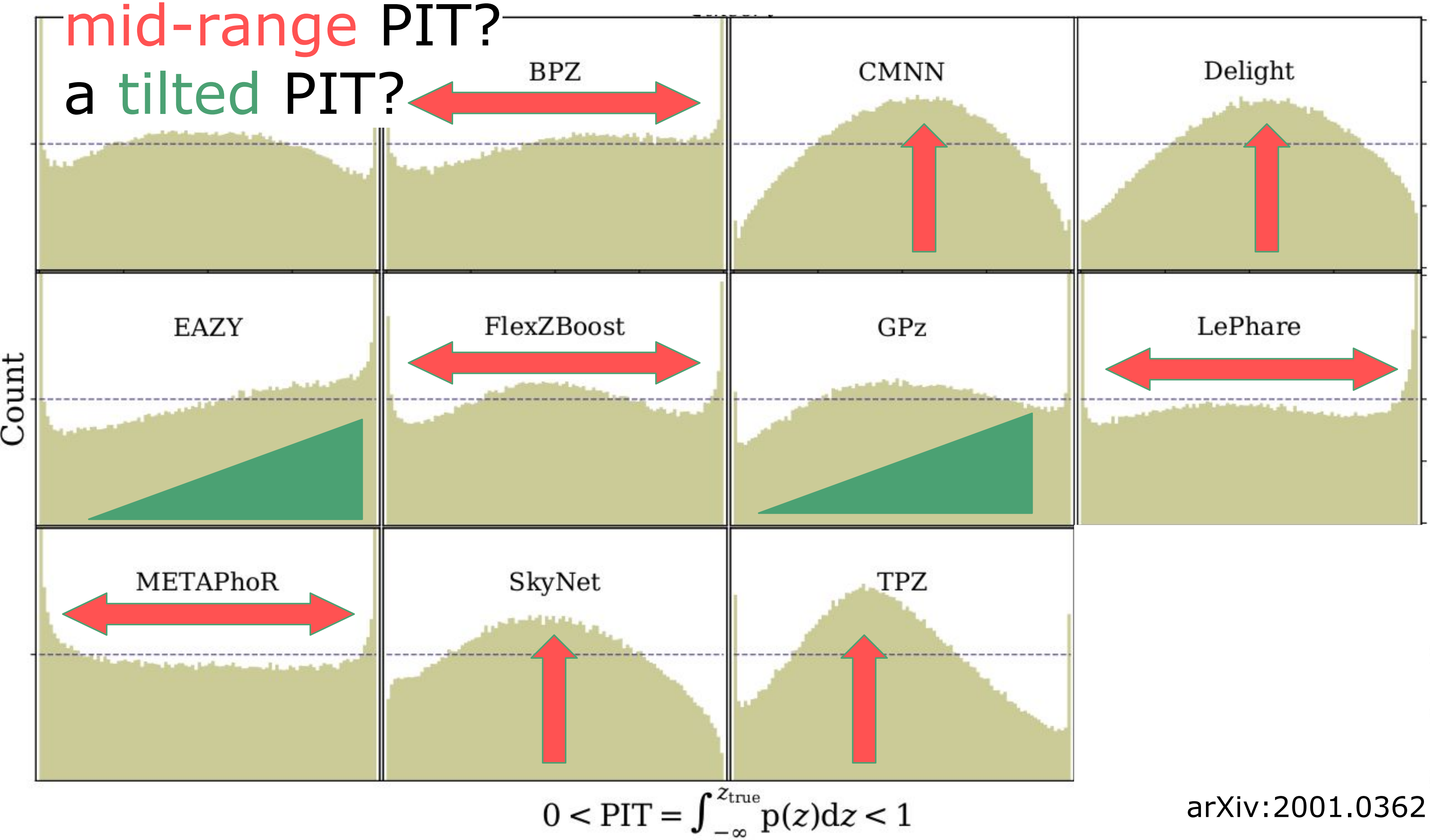
Photo-z posterior ensemble metrics



$$0 < \mathrm{PIT} = \int_{-\infty}^{z_{\mathrm{true}}} \mathrm{p}(z)\mathrm{d}z < 1$$

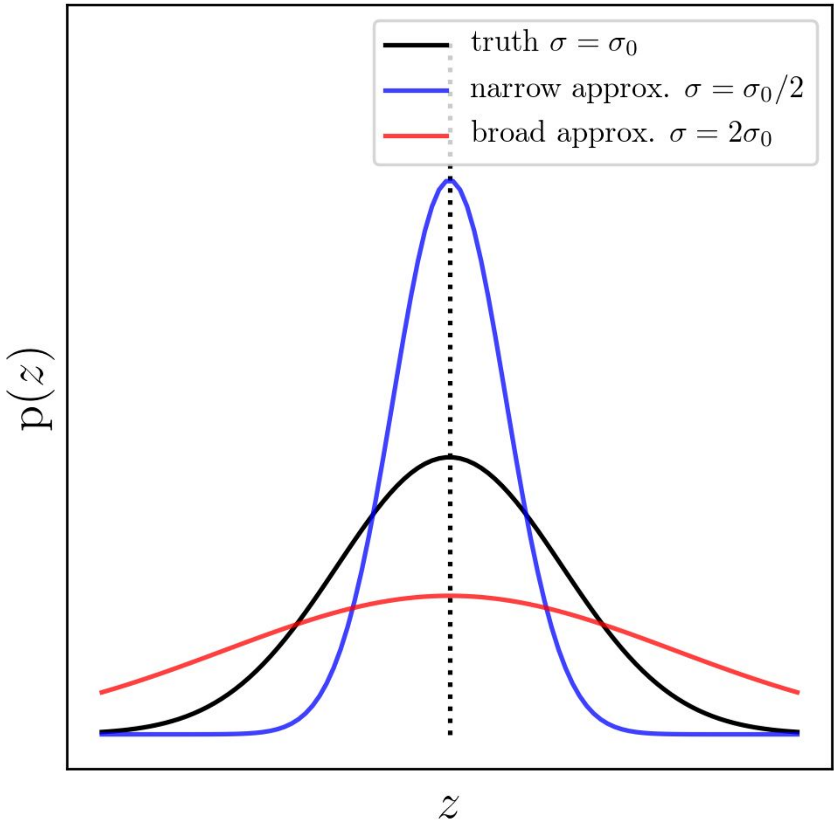
arXiv:2001.03621

What causes over-representation at extreme PIT? over-representation at mid-range PIT?
a tilted PIT?

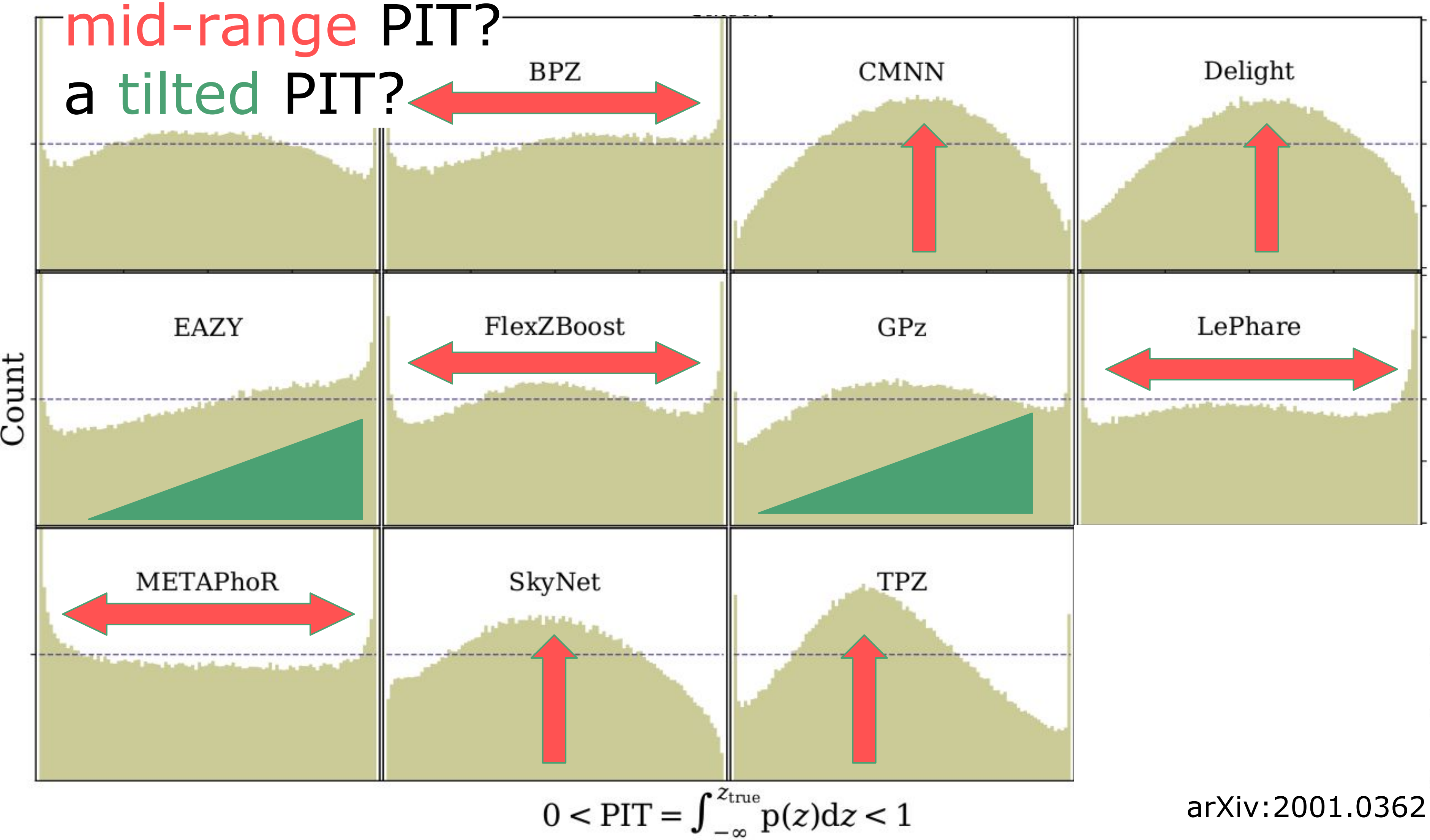


arXiv:2001.03621

(Hint: Imagine the PIT histograms in the Gaussian example.)



What causes over-representation at extreme PIT? over-representation at mid-range PIT?
a tilted PIT?



LSSTC DATA SCIENCE FELLOWSHIP PROGRAM

arXiv:2001.03621

Quantitative metrics of 1D PDF ensembles

Root-mean-square Error (RMSE)

$$\mathrm{RMSE} = \sqrt{\int (p_{\mathrm{true}}(z) - \hat{p}_{\mathrm{est}}(z))^2 dz}$$

No true posteriors available!

Kullback-Leibler Divergence (KLD)

$$\mathrm{KLD}[\hat{p}_{\mathrm{est}}(z); p_{\mathrm{true}}(z)] = \int_{-\infty}^{\infty} p_{\mathrm{true}}(z) \log\left[\frac{p_{\mathrm{true}}(z)}{\hat{p}_{\mathrm{est}}(z)}\right] dz$$

Cumulative Distribution Function (CDF)

$$\mathrm{CDF}[\hat{p}, z'] \equiv \int_{-\infty}^{z'} \hat{p}(z) dz$$

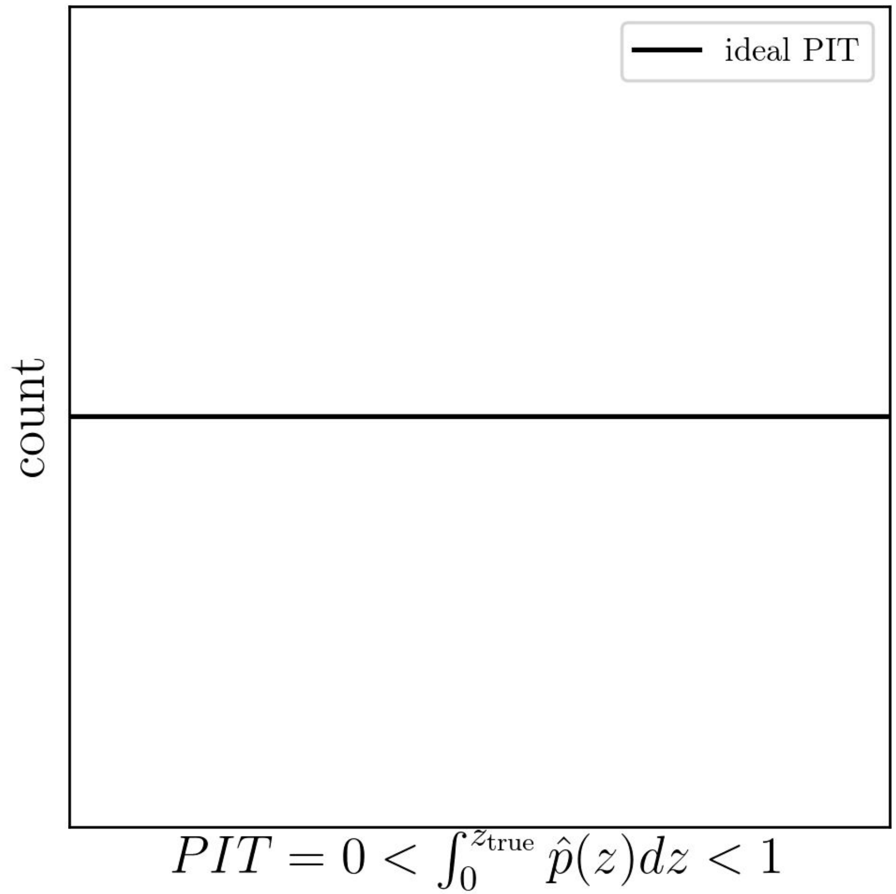
Probability Integral Transform (PIT)

$$P(\mathrm{PIT} \equiv \mathrm{CDF}[\hat{p}, z_{true}])$$

Quantile-quantile (QQ) Plot $\sim \int p(PIT) dPIT$

LSSTC DATA SCIENCE FELLOWSHIP PROGRAM

Sketch the Q-Q Plot of a hypothetical perfect photo-*z* estimator.

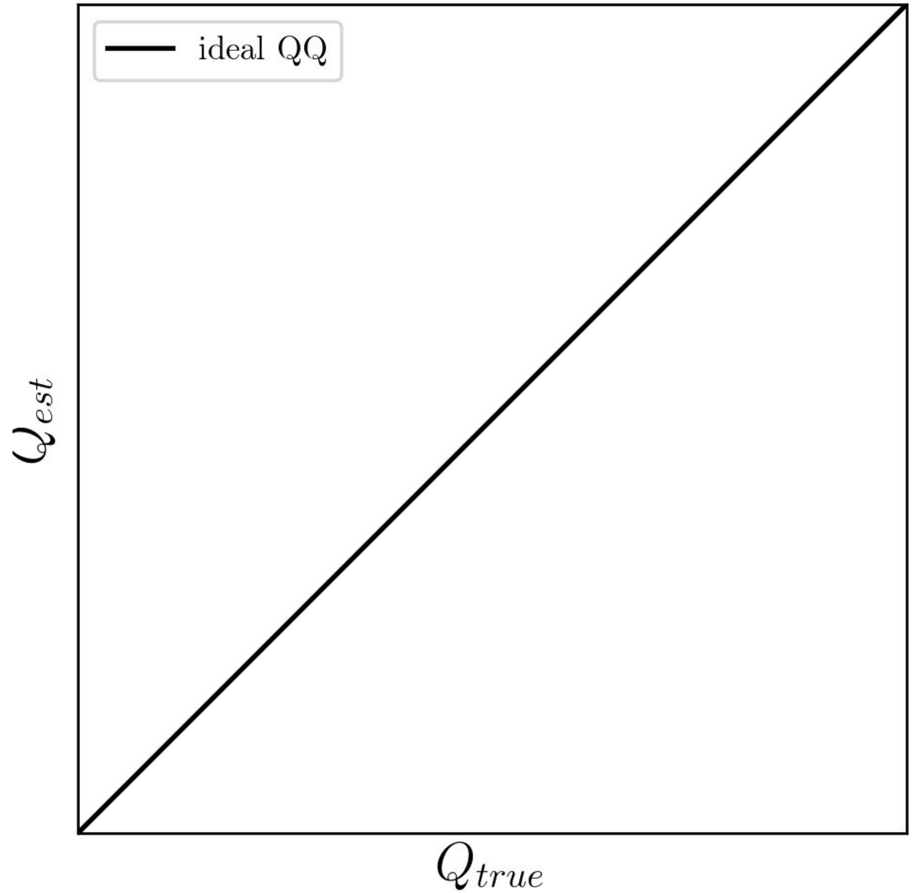


$$\mathrm{CDF}[\hat{p}, z'] \equiv \int_{-\infty}^{z'} \hat{p}(z)dz$$

$$P(\mathrm{PIT} \equiv \mathrm{CDF}[\hat{p}, z_{true}])$$

$$\sim \boxed{\int p(PIT)dPIT}$$

Sketch the Q-Q Plot of a hypothetical perfect photo-*z* estimator.

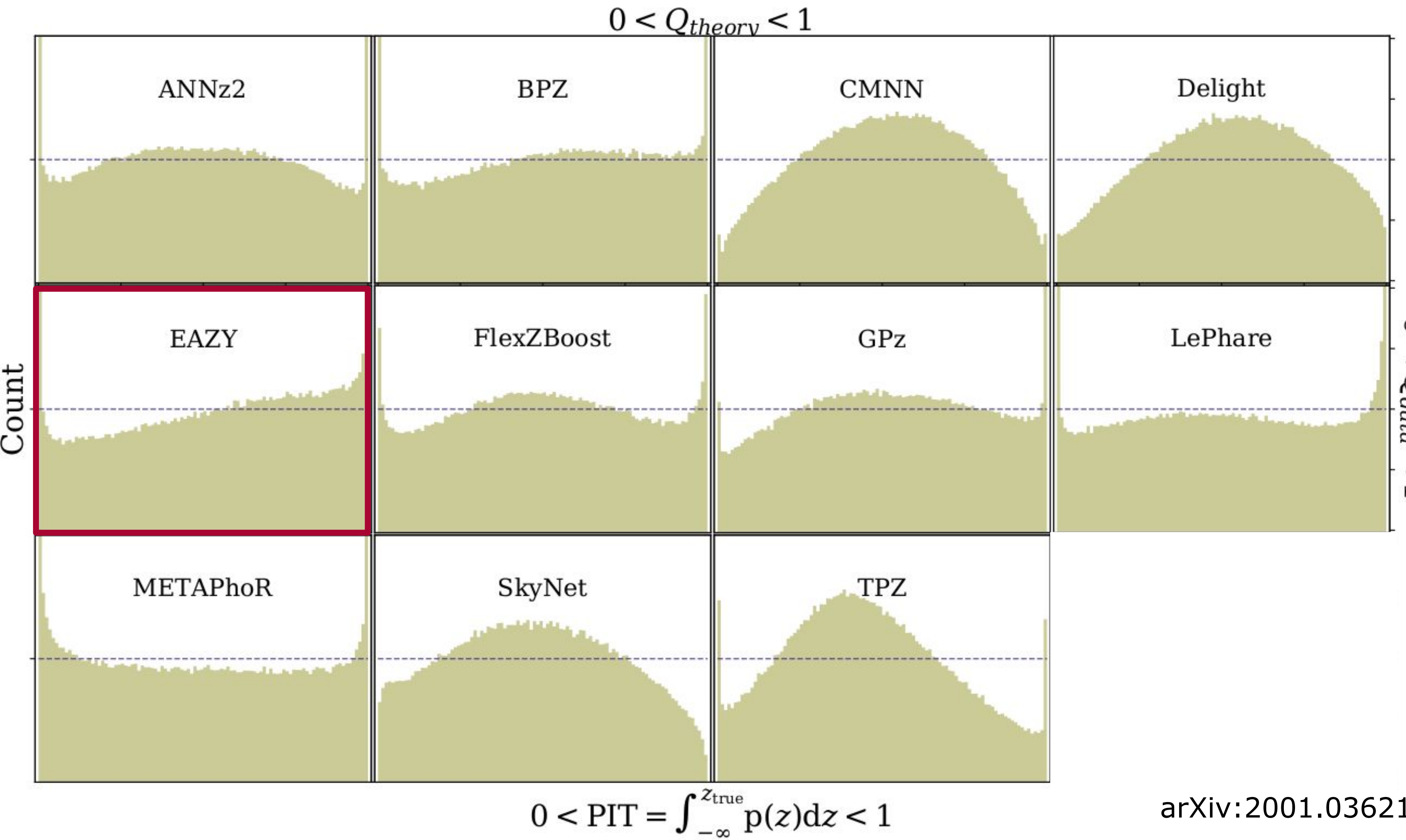


$$\mathrm{CDF}[\hat{p}, z'] \equiv \int_{-\infty}^{z'} \hat{p}(z) dz$$

$$P(\mathrm{PIT} \equiv \mathrm{CDF}[\hat{p}, z_{\mathrm{true}}])$$

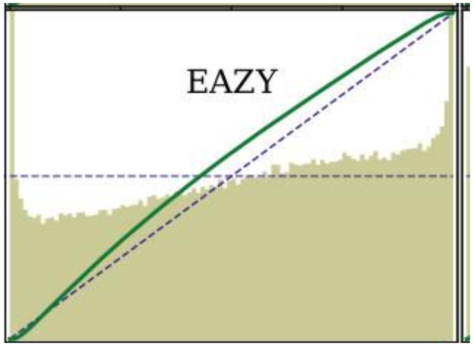
$$\sim \boxed{\int p(PIT) dPIT}$$

Sketch the QQ plot: $\sim\int p(PIT)dPIT$



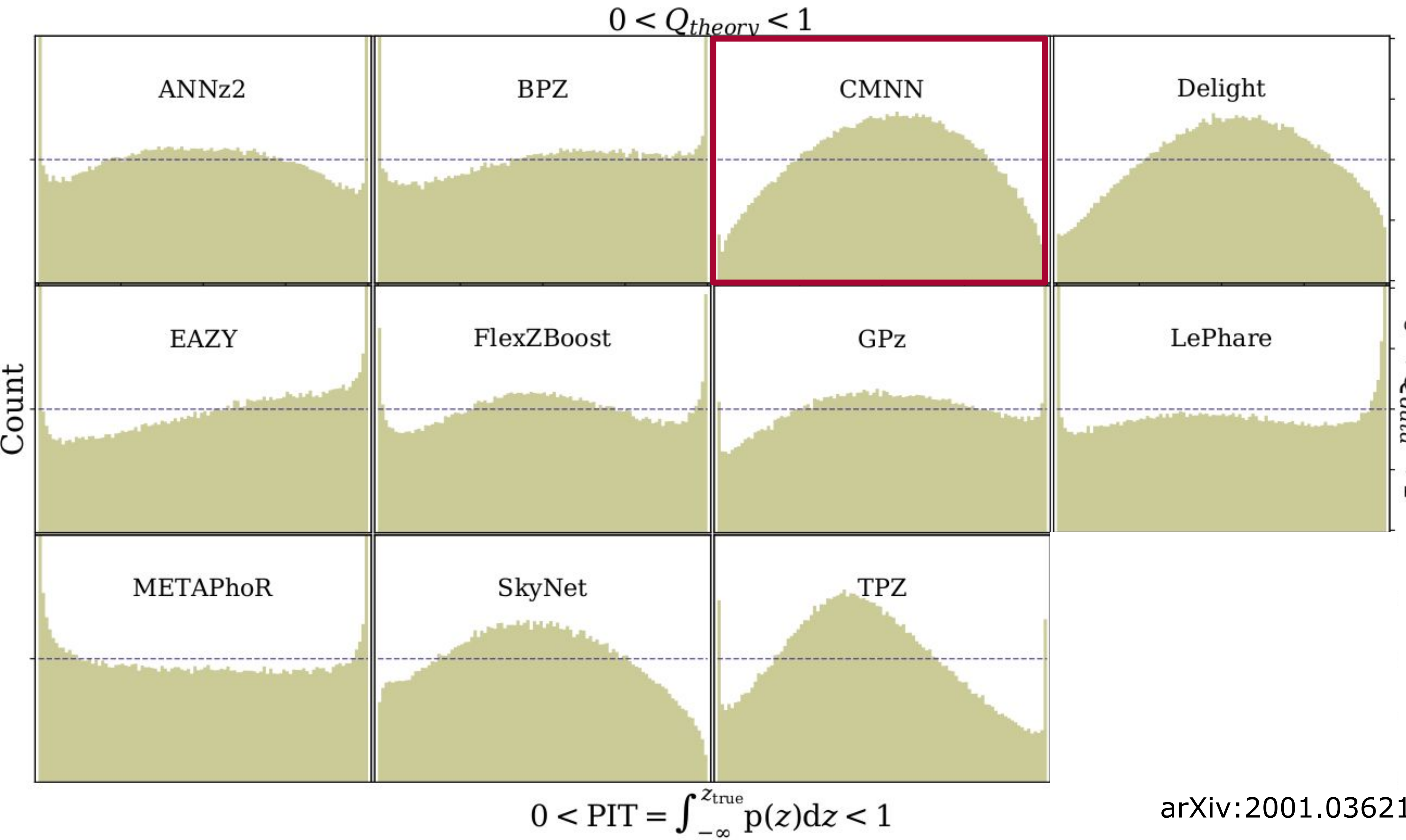
arXiv:2001.03621

Sketch the QQ plot: ~$\int p(PIT)dPIT$



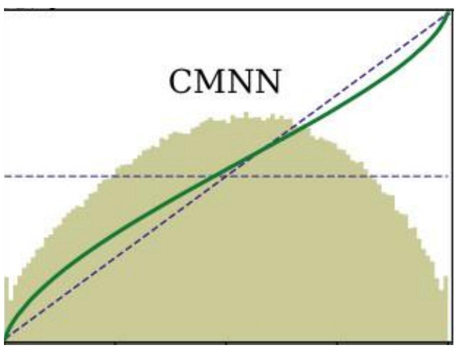
arXiv:2001.03621

Sketch the QQ plot:~$\boxed{\int p(PIT)dPIT}$



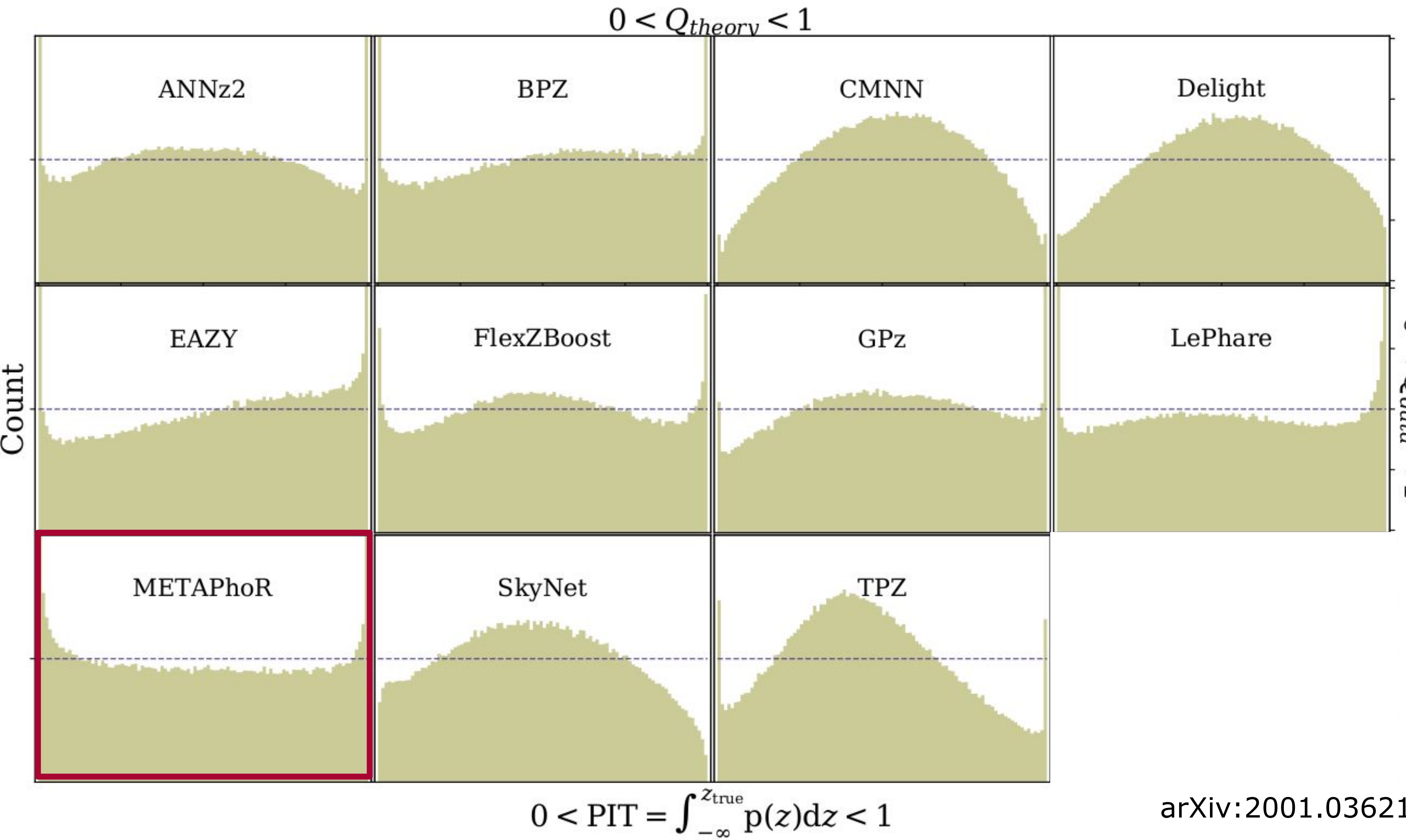
arXiv:2001.03621

Sketch the QQ plot: ~$\boxed{\int p(PIT)dPIT}$



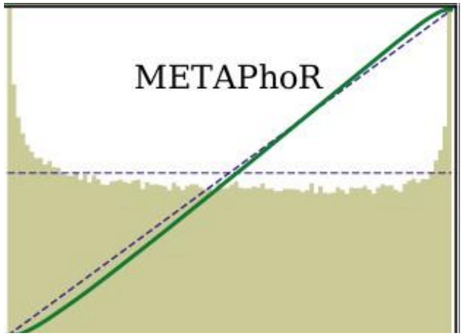
arXiv:2001.03621

Sketch the QQ plot:~$\boxed{\int p(PIT)dPIT}$



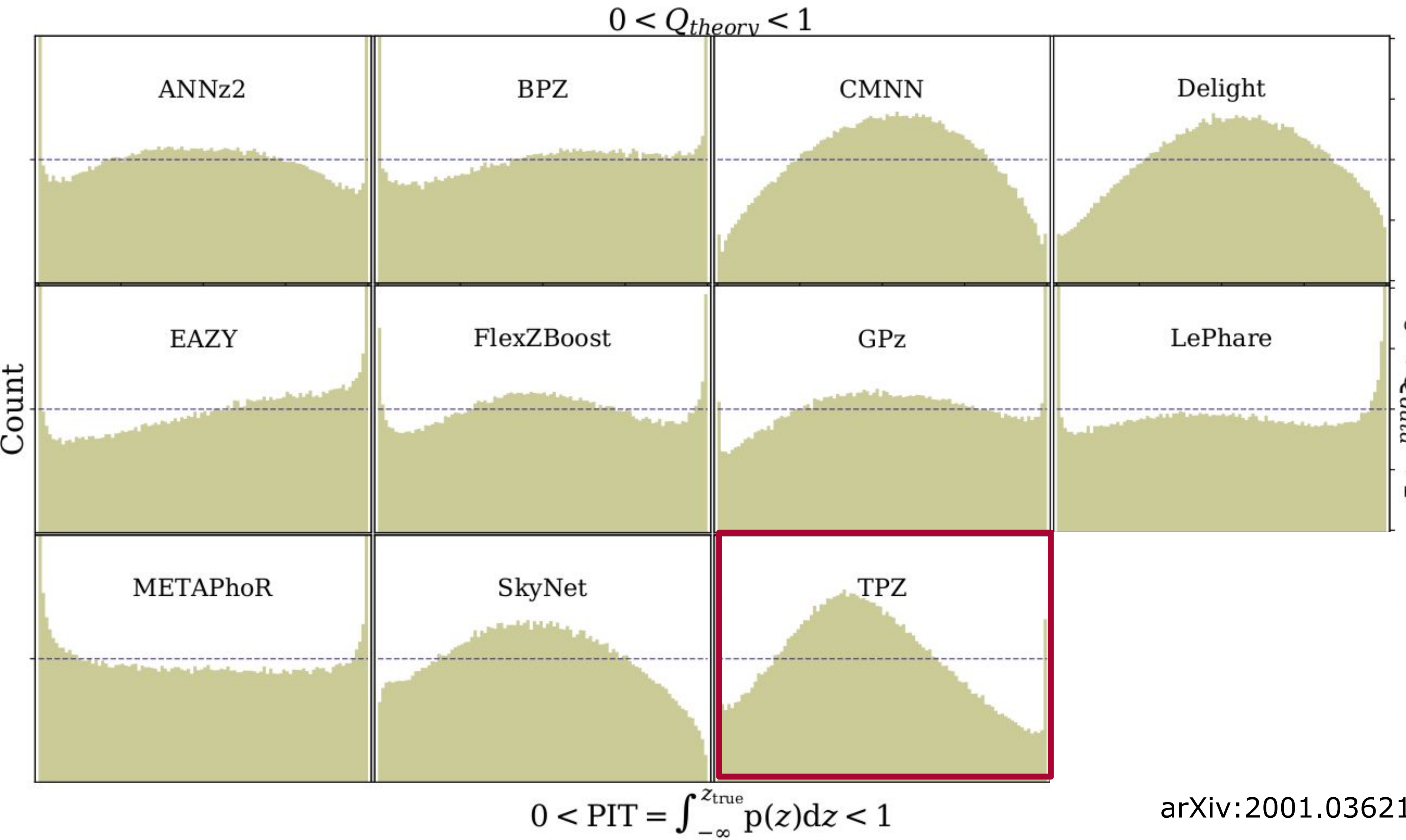
arXiv:2001.03621

Sketch the QQ plot.~$\int p(PIT)dPIT$



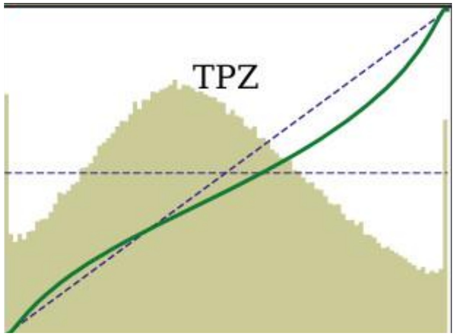
arXiv:2001.03621

Sketch the QQ plot: $\sim \boxed{\int p(PIT)dPIT}$



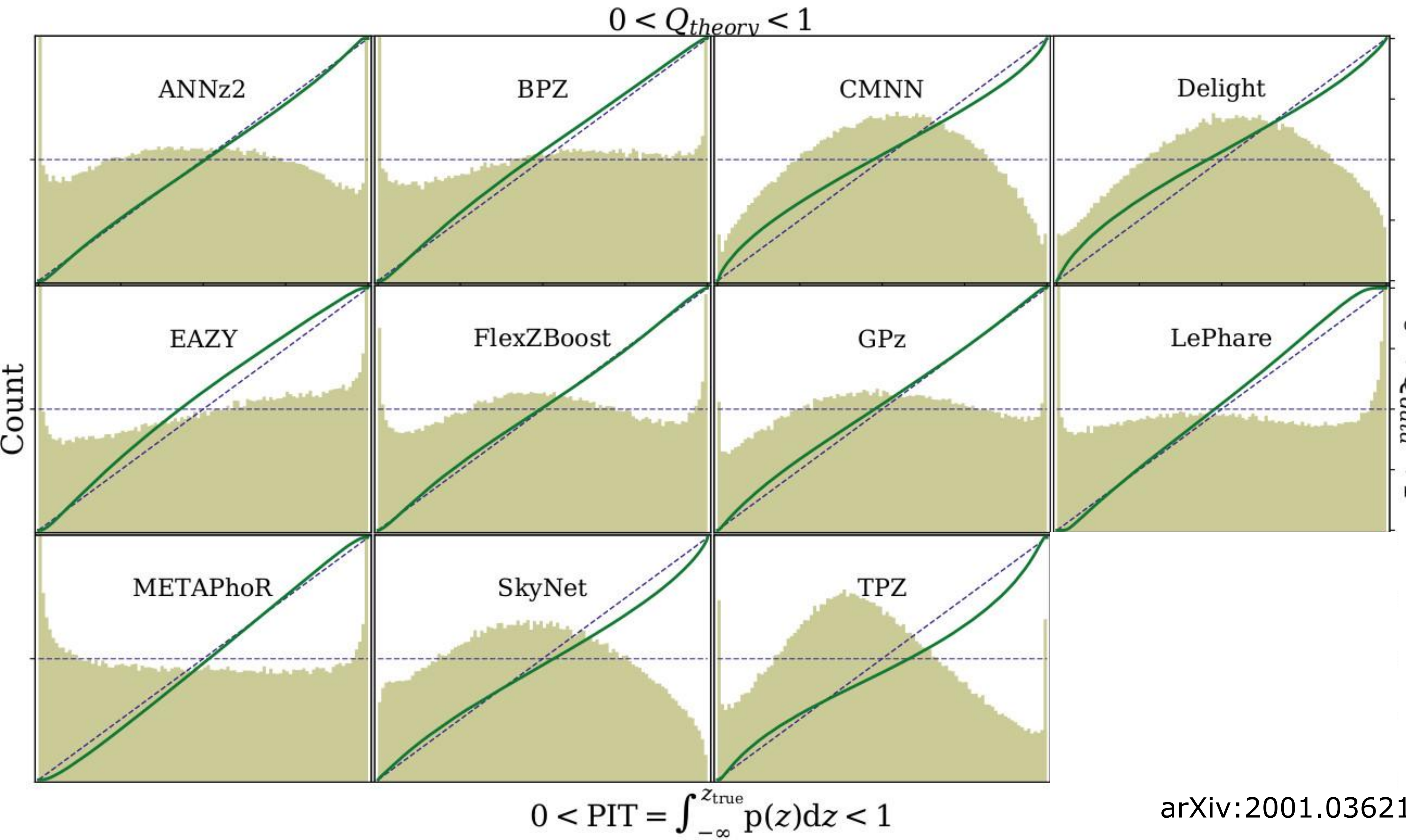
arXiv:2001.03621

Sketch the QQ plot: $\sim \int p(PIT)dPIT$



arXiv:2001.03621

Sketch the QQ plot.~$\boxed{\int p(PIT)dPIT}$



arXiv:2001.03621

Where is this going?

These metrics tell us about ensembles of 1D PDFs that correspond to different models.

In the problem session, we'll go over samples from 2D PDFs that correspond to different models (and data).