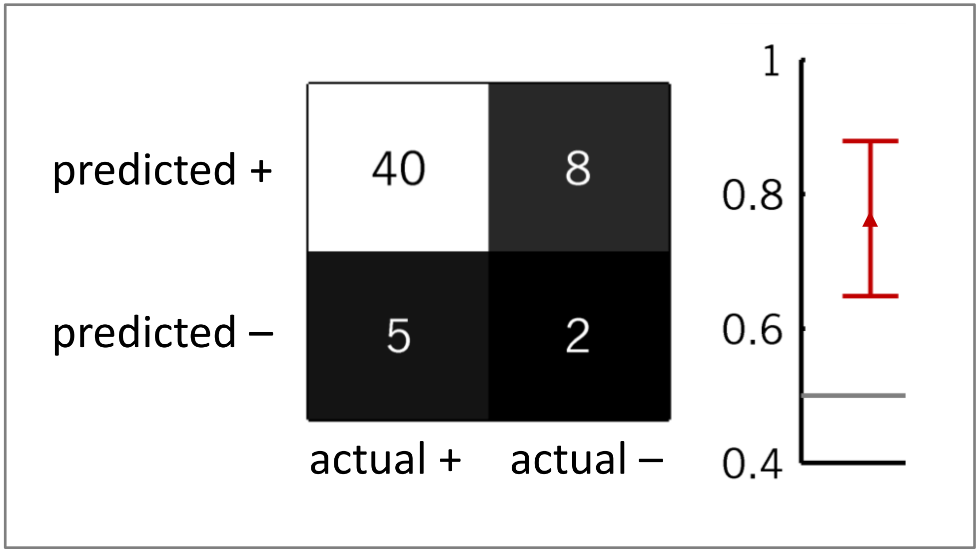
Evaluation of classification performance on small, imbalanced datasets

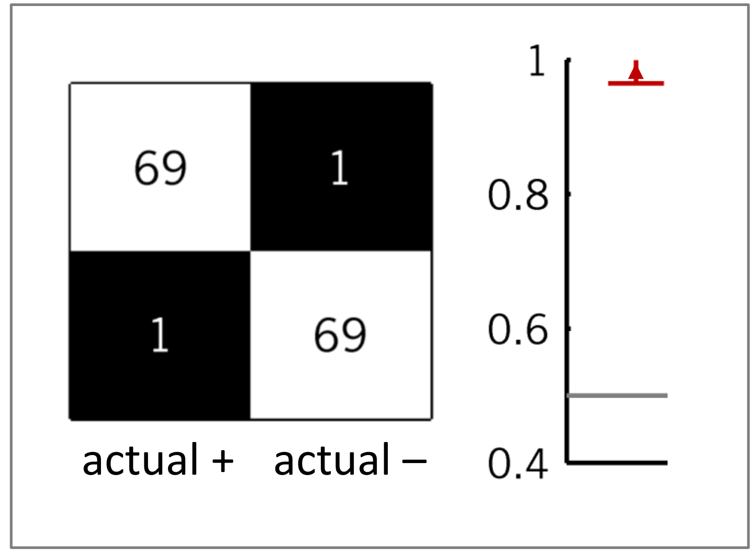
Kay H. Brodersen[1,2], Cheng Soon Ong[1], Klaas E. Stephan[1,2,3], Joachim M. Buhmann[1]

[1] Department of Computer Science, ETH Zurich, Switzerland
[2] Laboratory for Social and Neural Systems Research, University of Zurich, Switzerland
[3] Wellcome Trust Centre for Neuroimaging, University College London, United Kingdom

1 The balanced accuracy

Is the accuracy a faithful performance measure?





Assessing classification performance

Setting

- Observations $x \in \mathrm{X}$ with labels $y \in \{-1,+1\}$
- Classification-based confusion matrix:

$$C := TP + TN$$

$$I := FP + FN$$

	actual +	actual –
predicted +	TP	FP
predicted –	FN	TN
	P	N

Performance assessment

- Accuracy

$$A = \frac{TP + TN}{n}$$

- Balanced accuracy

$$B = \frac{1}{2}\left(\frac{TP}{TP + FN} + \frac{TN}{FP + TN}\right)$$

The posterior distribution of the accuracy

- Assuming a flat prior on the interval [0,1], the posterior of the accuracy follows a Beta distribution

$$A \sim Beta(a,b) \quad \text{with} \quad a = C+1, \quad b = I+1$$

$$p_A(x; C, I) = \frac{1}{B(C+1, I+1)} x^C (1-x)^I$$

- From this we can compute:
  - the mean: $\frac{C+1}{C+I+2}$
  - the mode: $\frac{C}{C+I}$
  - a posterior probability interval:

$$\left[F_B^{-1}\left(\tfrac{\alpha}{2}; C+1, I+1\right); \ F_B^{-1}\left(1-\tfrac{\alpha}{2}; C+1, I+1\right)\right]$$

The posterior distribution of the balanced accuracy

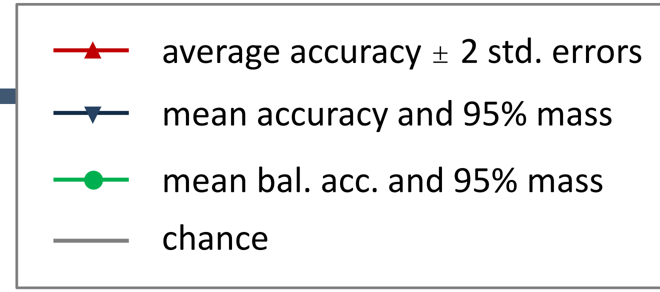
- Assuming a flat prior on the interval [0,1], the posterior of the balanced accuracy is given by the convolution of two Beta distributions

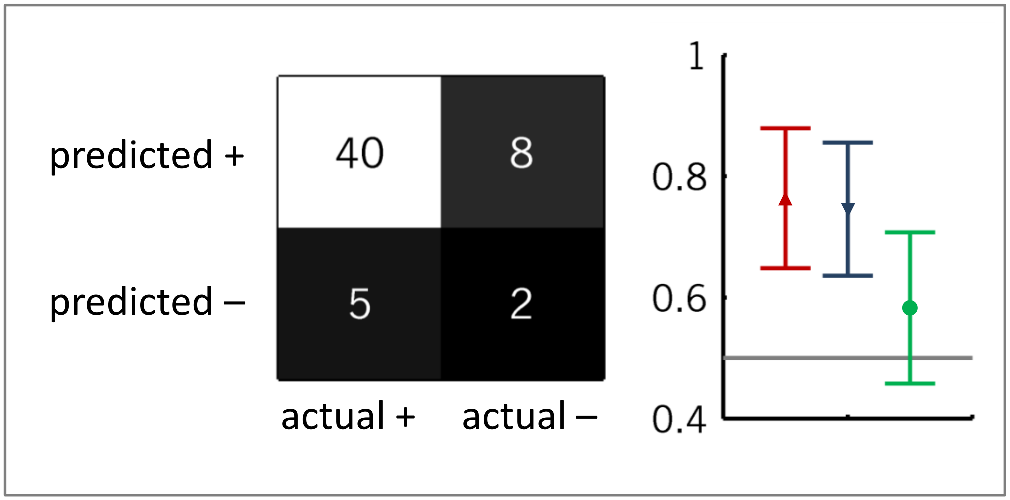
$$B = \tfrac{1}{2}(A_P + A_N) \sim Betaavg$$

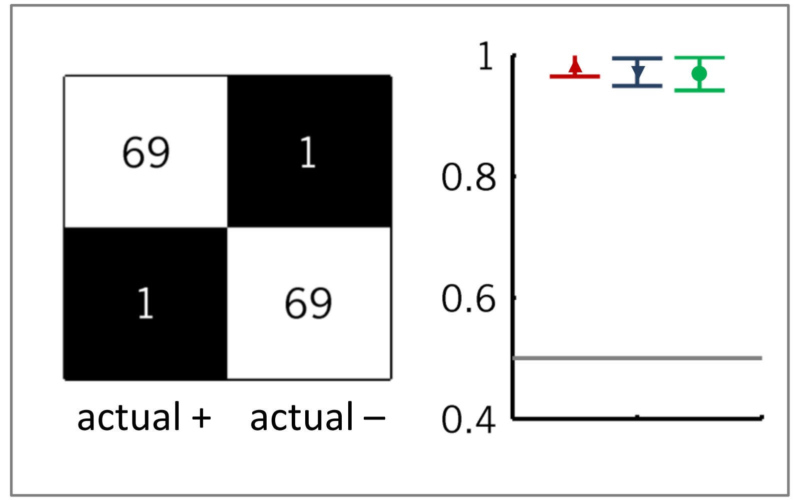
$$p_B(x) = \int_0^1 p_A\big(2(x-z); TP+1, FN+1\big) \times p_A\big(2z; TN+1, FP+1\big)\, dz$$

- Based on this density, we can compute:
  - the mean
  - the mode
  - a posterior probability interval

Two examples



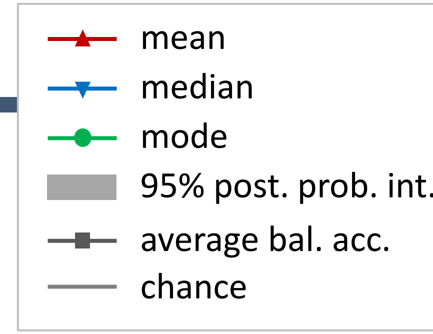




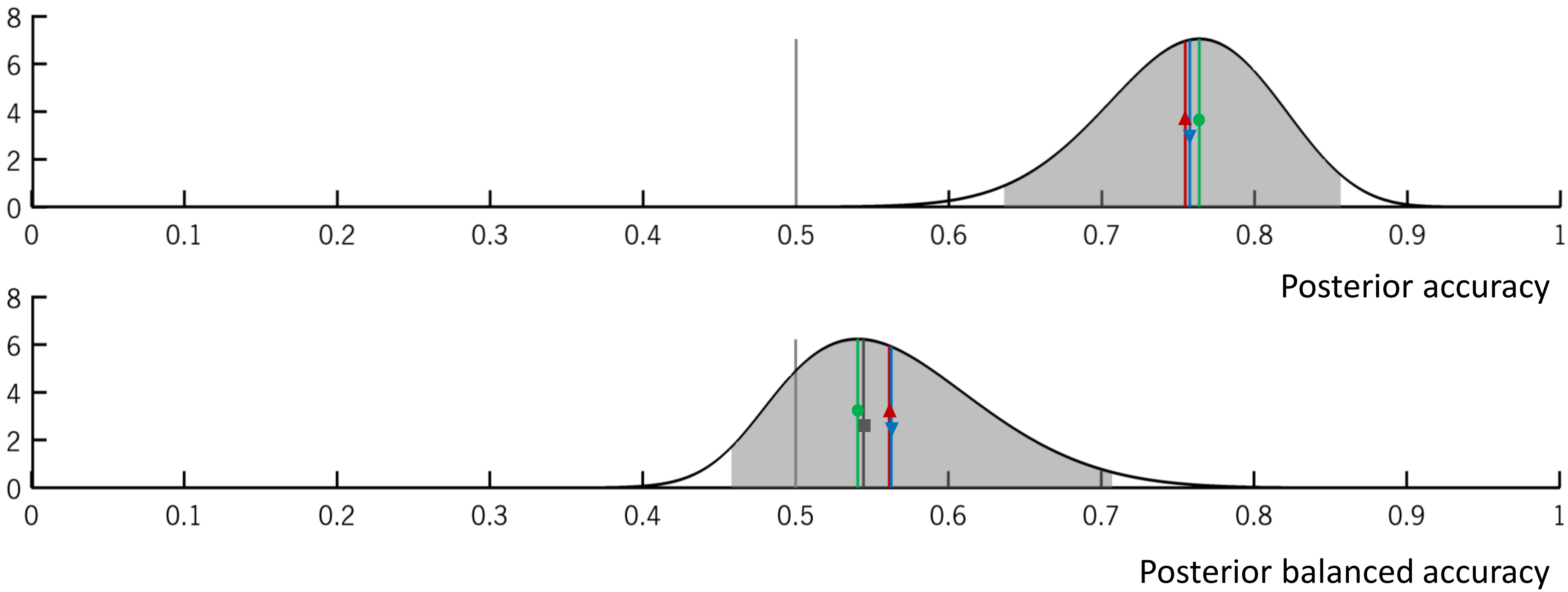
Example 1: fair overall accuracy, high class imbalance, strong prediction bias

Example 2: high accuracies on both classes, no imbalance, no bias

Posterior densities

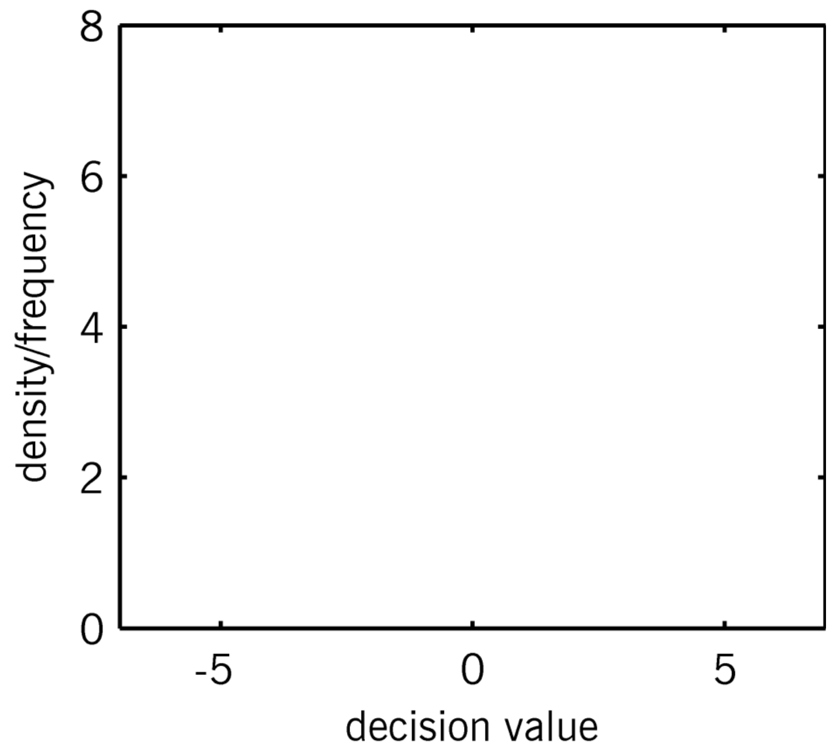


	actual +	actual –
predicted +	40	8
predicted –	5	2

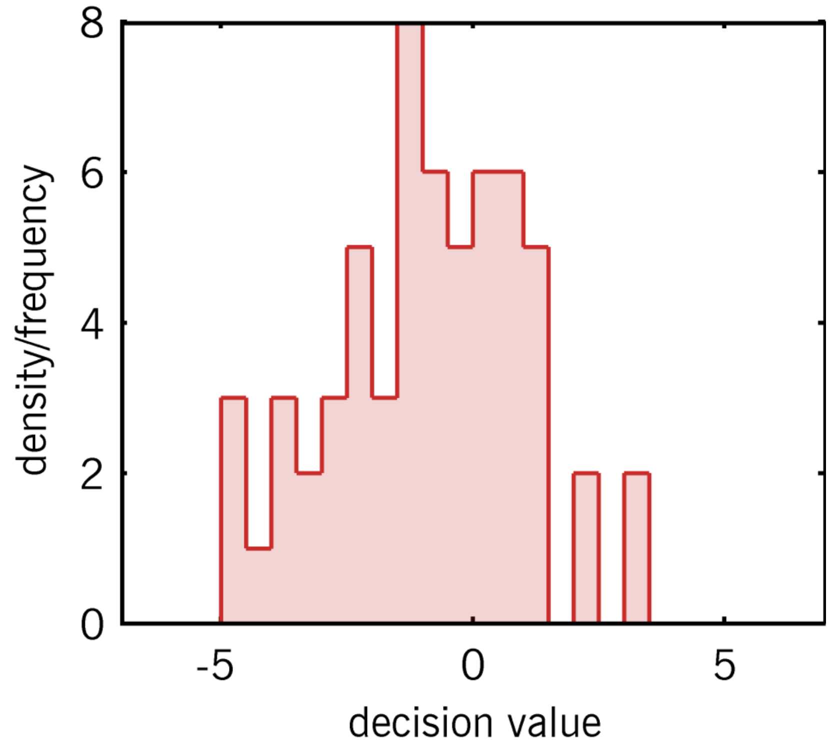


2 Smooth precision-recall curves

Decision values

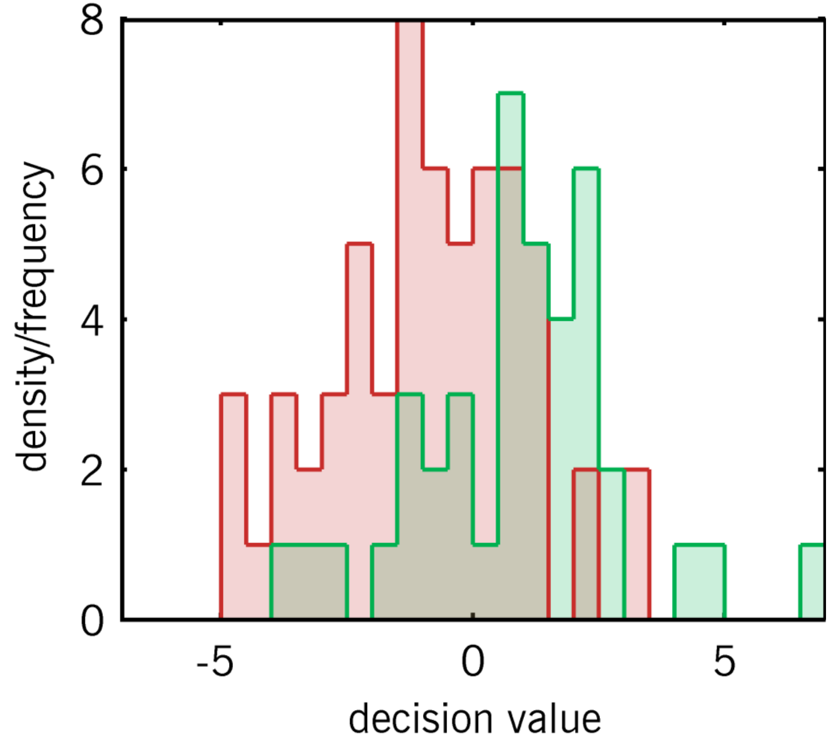


Decision values and the binormal assumption



decision values
of negative examples

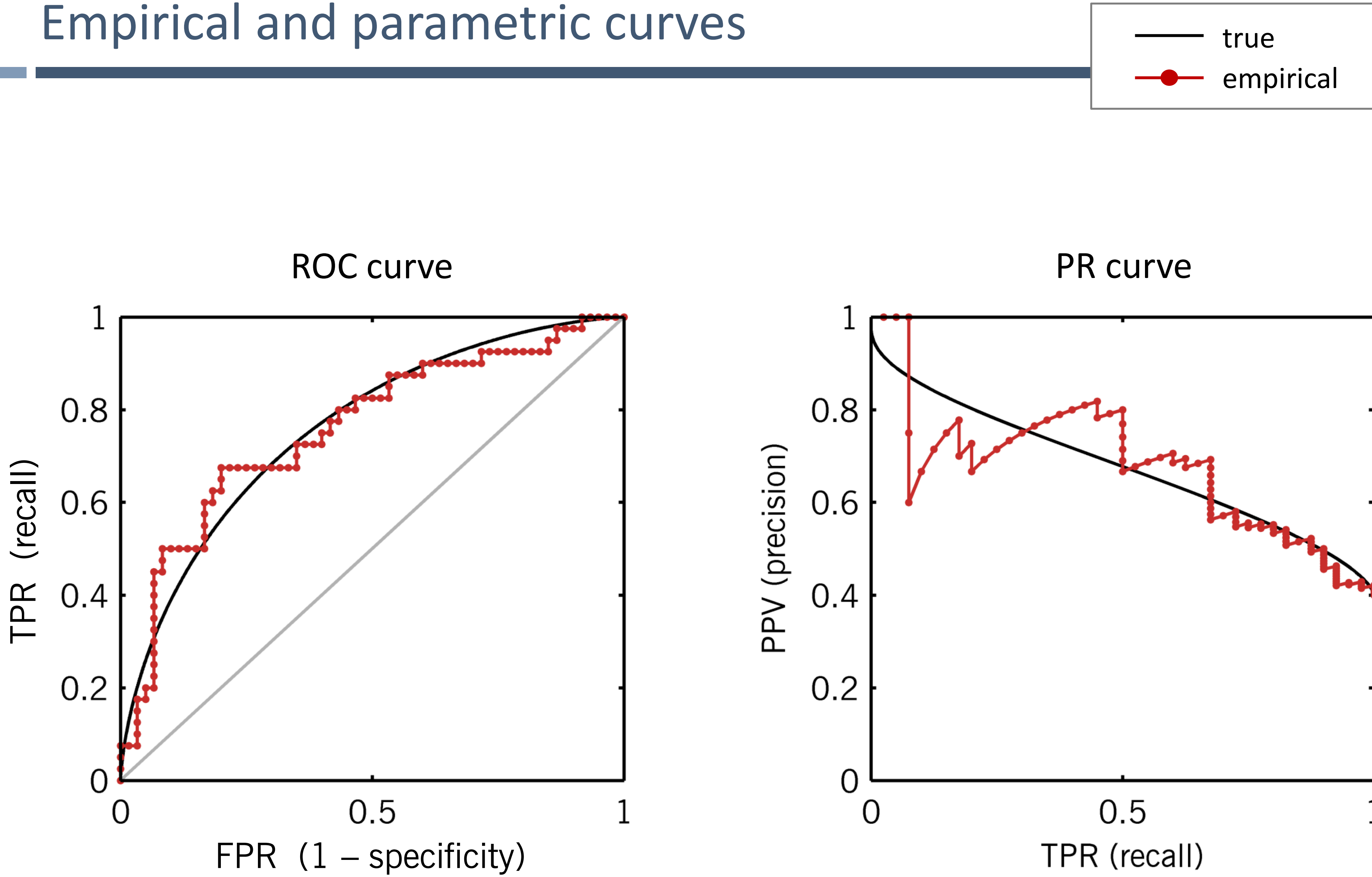
Decision values and the binormal assumption



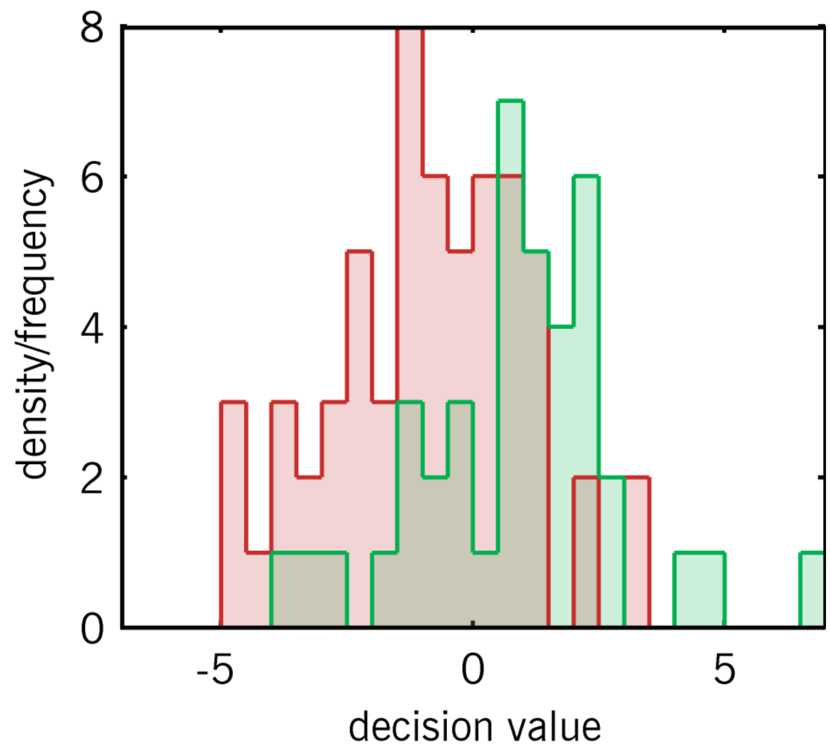
decision values
of negative examples

decision values
of positive examples

Empirical and parametric curves



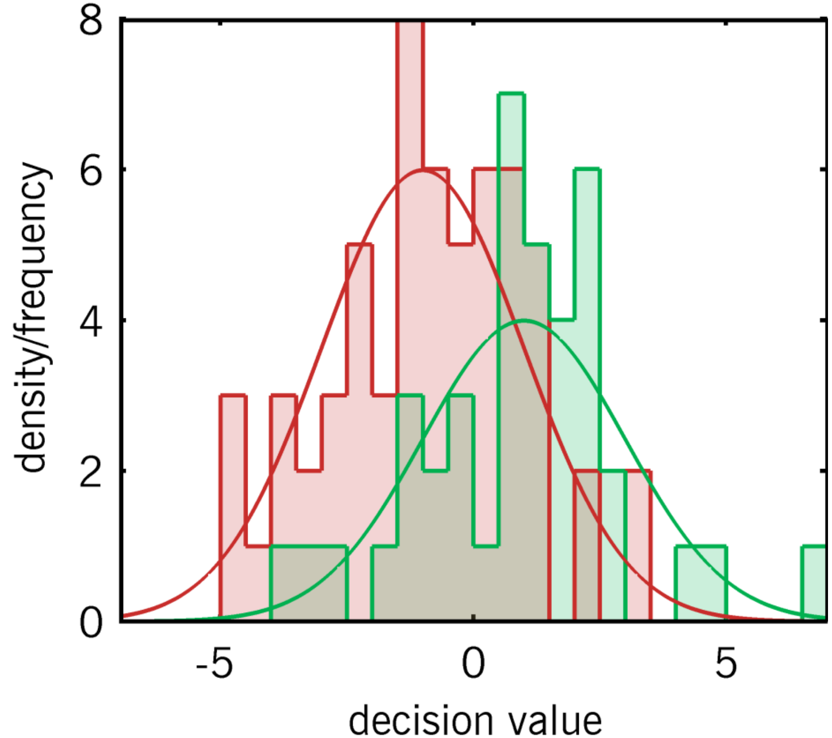
Decision values and the binormal assumption



decision values
of negative examples

decision values
of positive examples

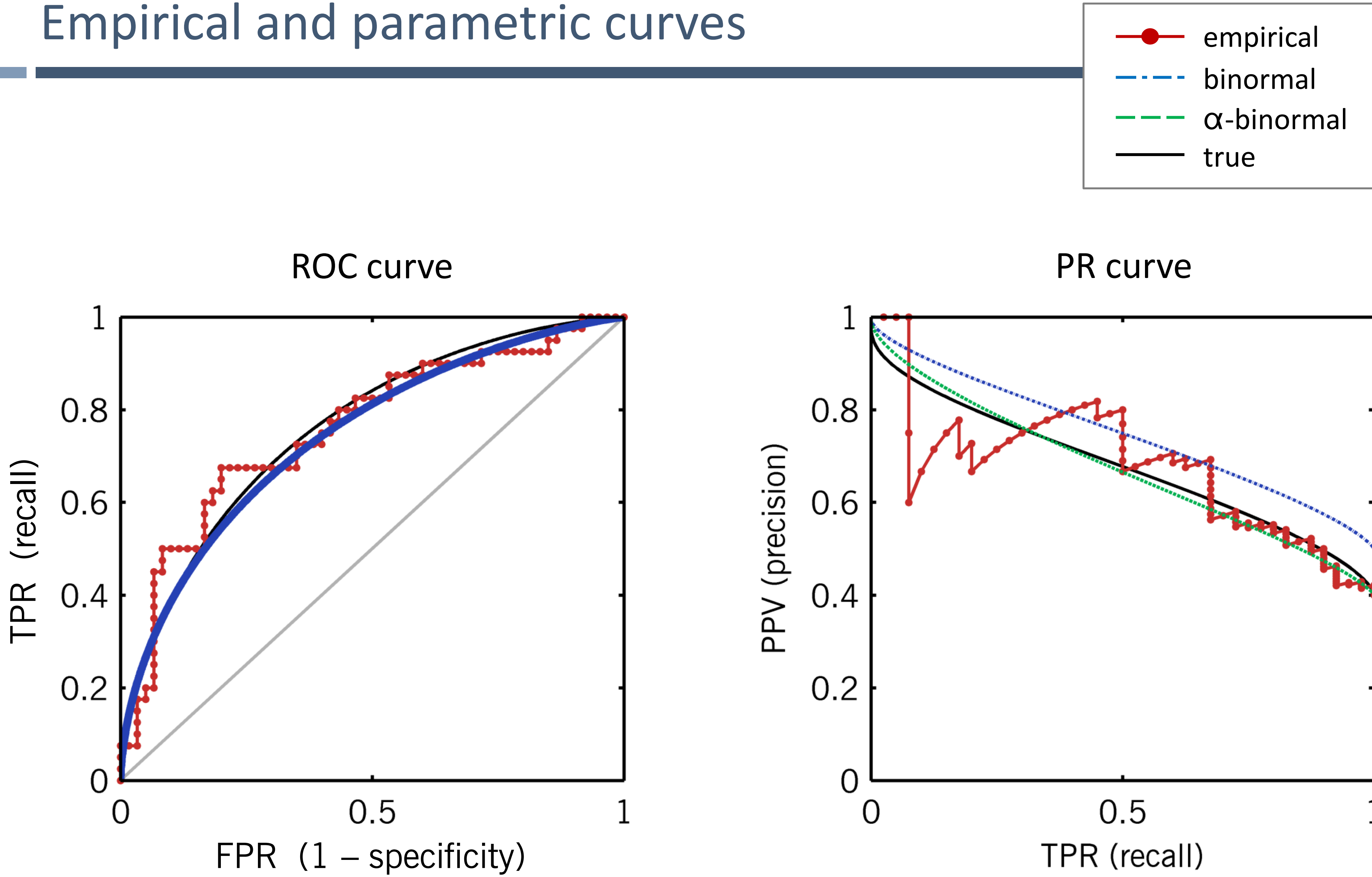
Decision values and the binormal assumption



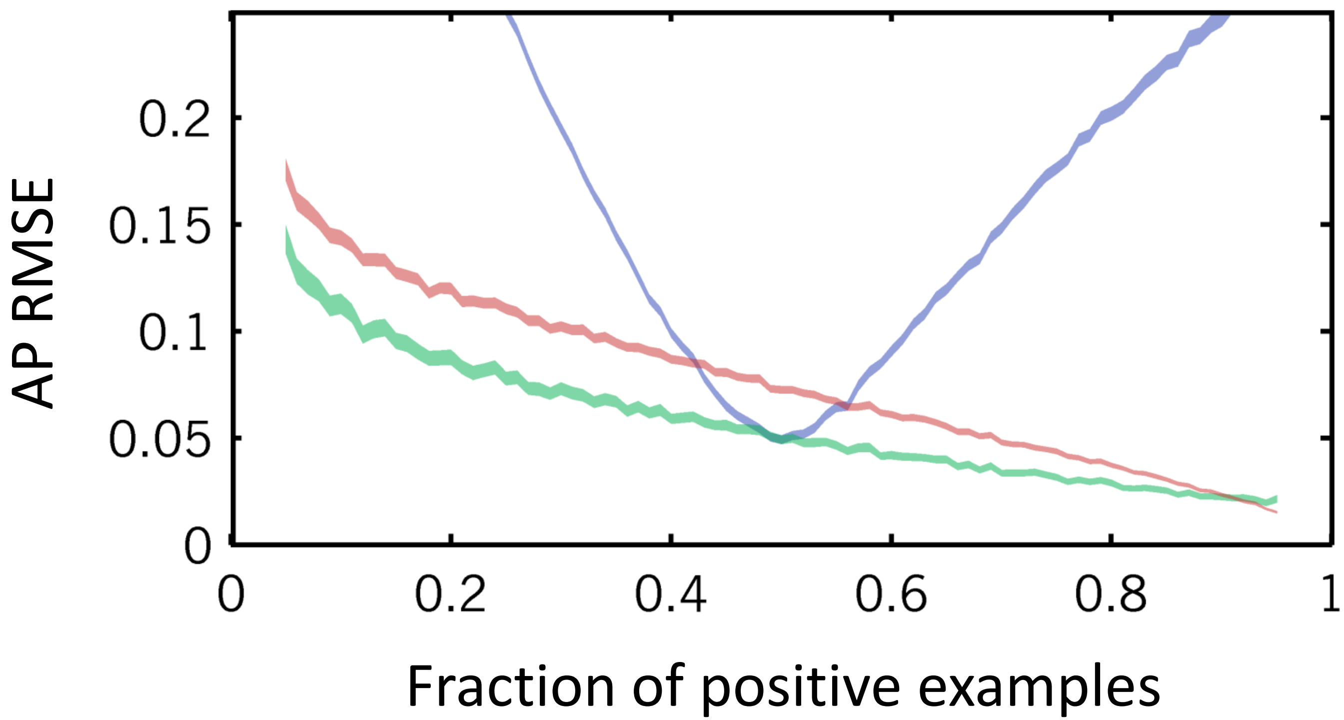
decision values
of negative examples

decision values
of positive examples

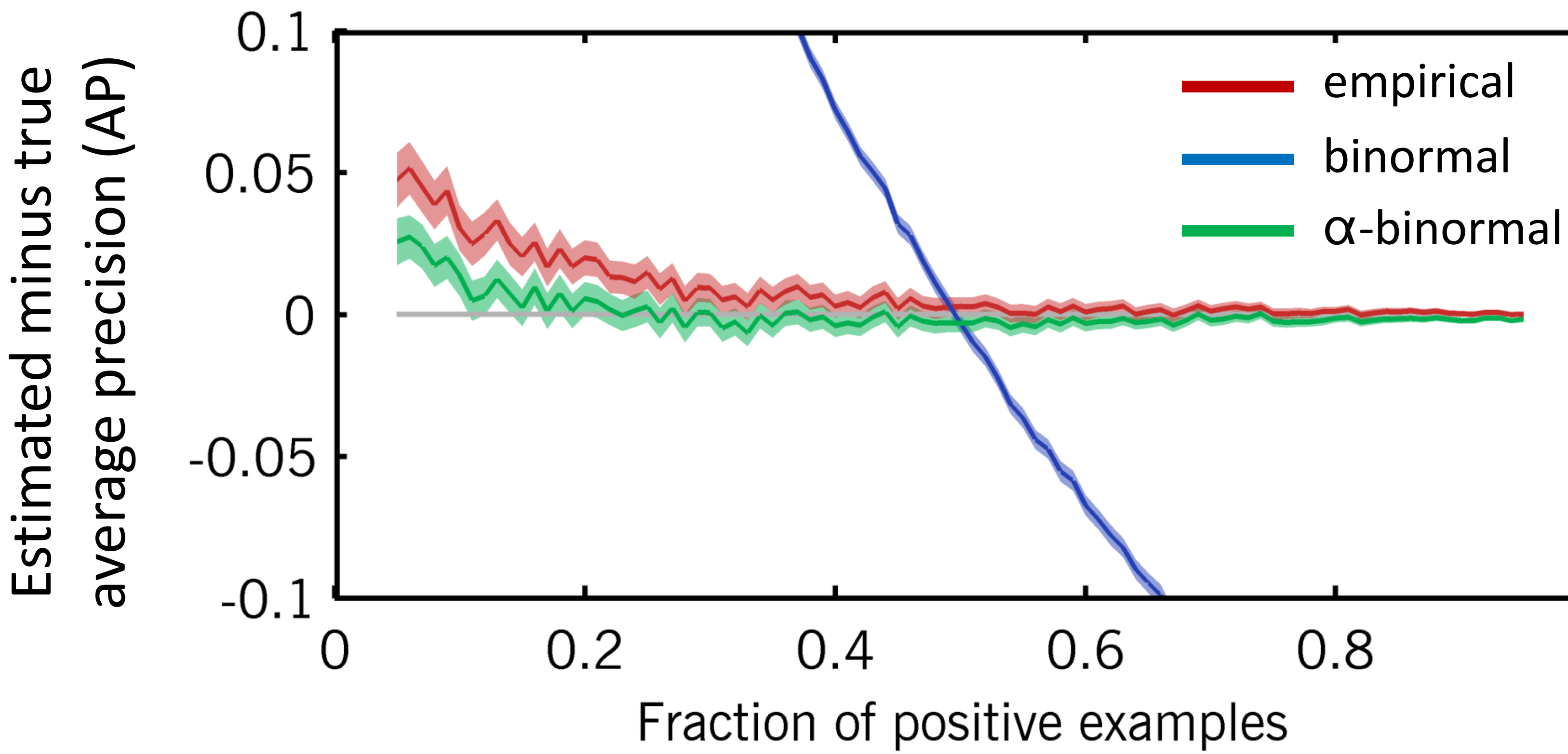
Empirical and parametric curves



The effect of class imbalance on the PR curve



The effect of class imbalance on the PR curve



Take-home messages

Dont's	Do's
report the average and the standard error of the accuracy across cross-validation folds	report a statistic of the posterior distribution of the balanced accuracy
look at empirical ROC or PR curves	compute a smooth ROC or PR curve under parametric assumptions

K.H. Brodersen, C.S. Ong, K.E. Stephan, J.M. Buhmann (2010)
The balanced accuracy and its posterior distribution.
Proceedings of the 20th International Conference on Pattern Recognition (in press).

K.H. Brodersen, C.S. Ong, K.E. Stephan, J.M. Buhmann (2010)
The binormal assumption on precision-recall curves.
Proceedings of the 20th International Conference on Pattern Recognition (in press).