Blind DET Estimation

(or - How to cheat at the NIST evaluation)

Niko Brümmer

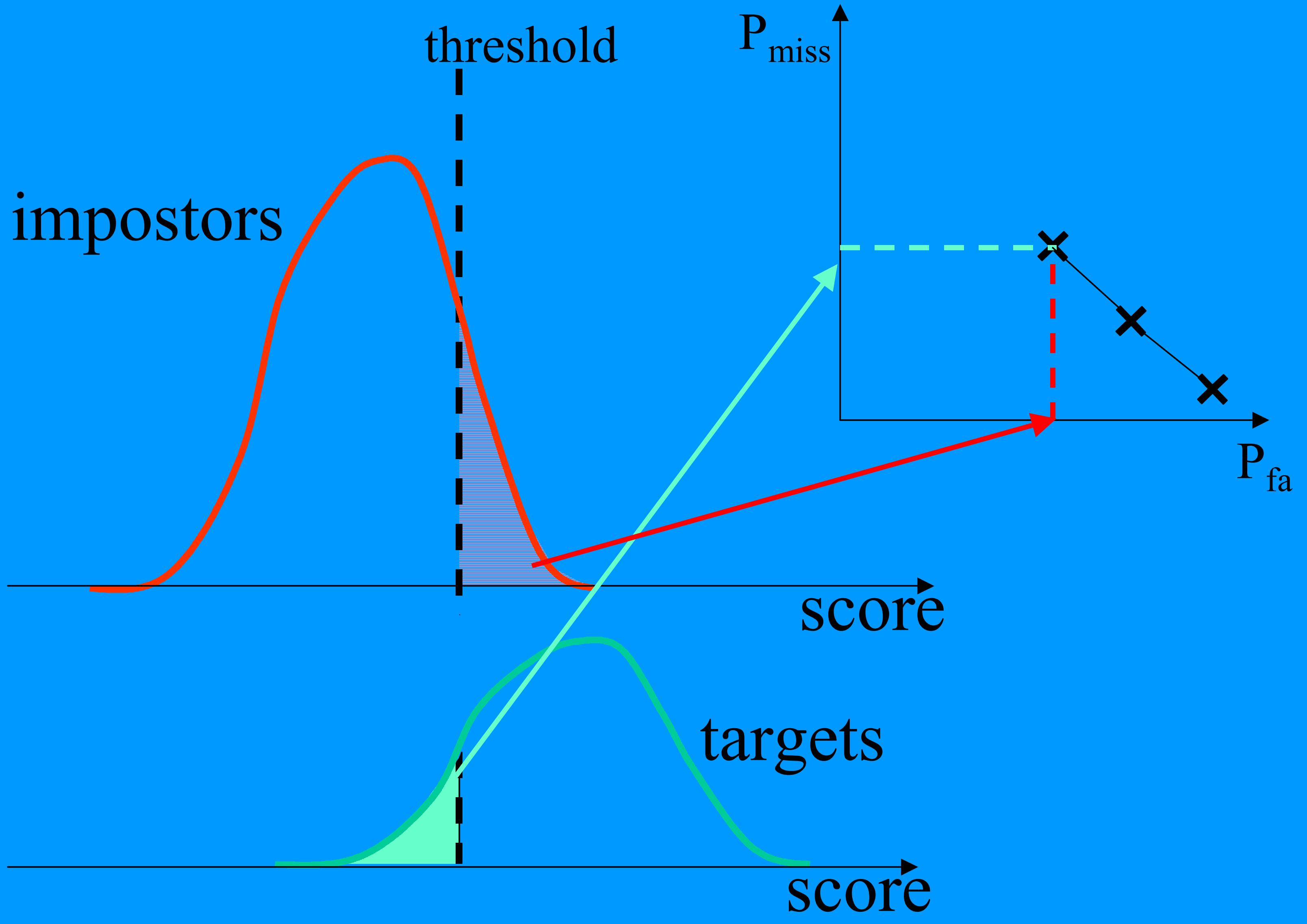
&

Jason Pelecanos

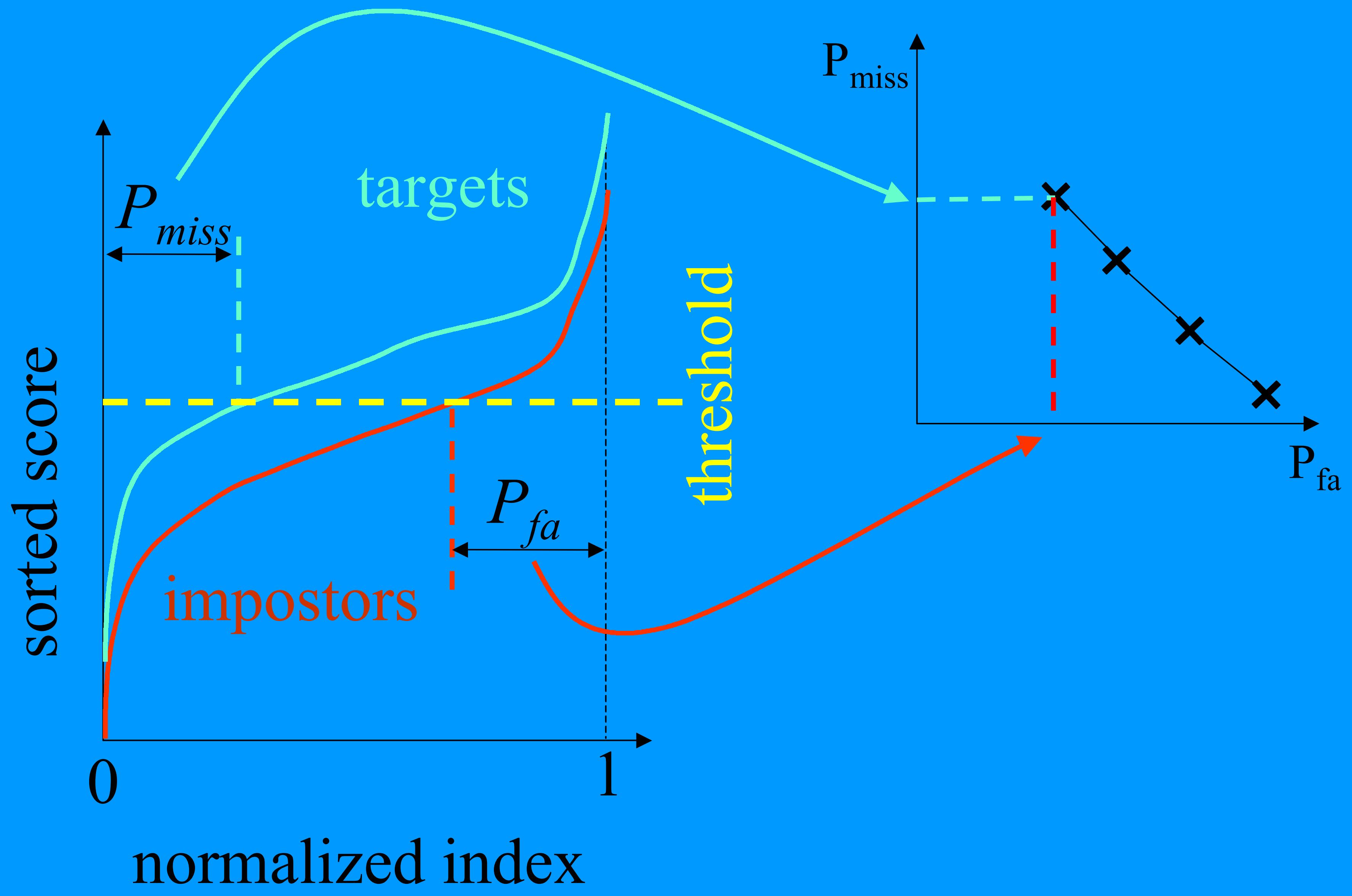
Introduction

- Problem: Databases for SR development & evaluation are expensive because:
  - They have to be *large*
  - They are *not portable* between environments
  - They have to be *supervised*
- The object of this study is to find out what can be done with an *unsupervised* database, where speaker ID's are not known.

Detection-Error-Tradeoff (DET)



DET: directly from data



Database Prerequisites

- contains multiple single-speaker speech segments from many speakers,
- speaker identities need not be known
- organized into *test pairs*

  = (training utterance, test utterance)

Database Prerequisites

- Must contain *significant proportion* of two kinds of test pairs:
  - *impostor*: training speaker ≠ test speaker
  - *target*: training speaker = test speaker

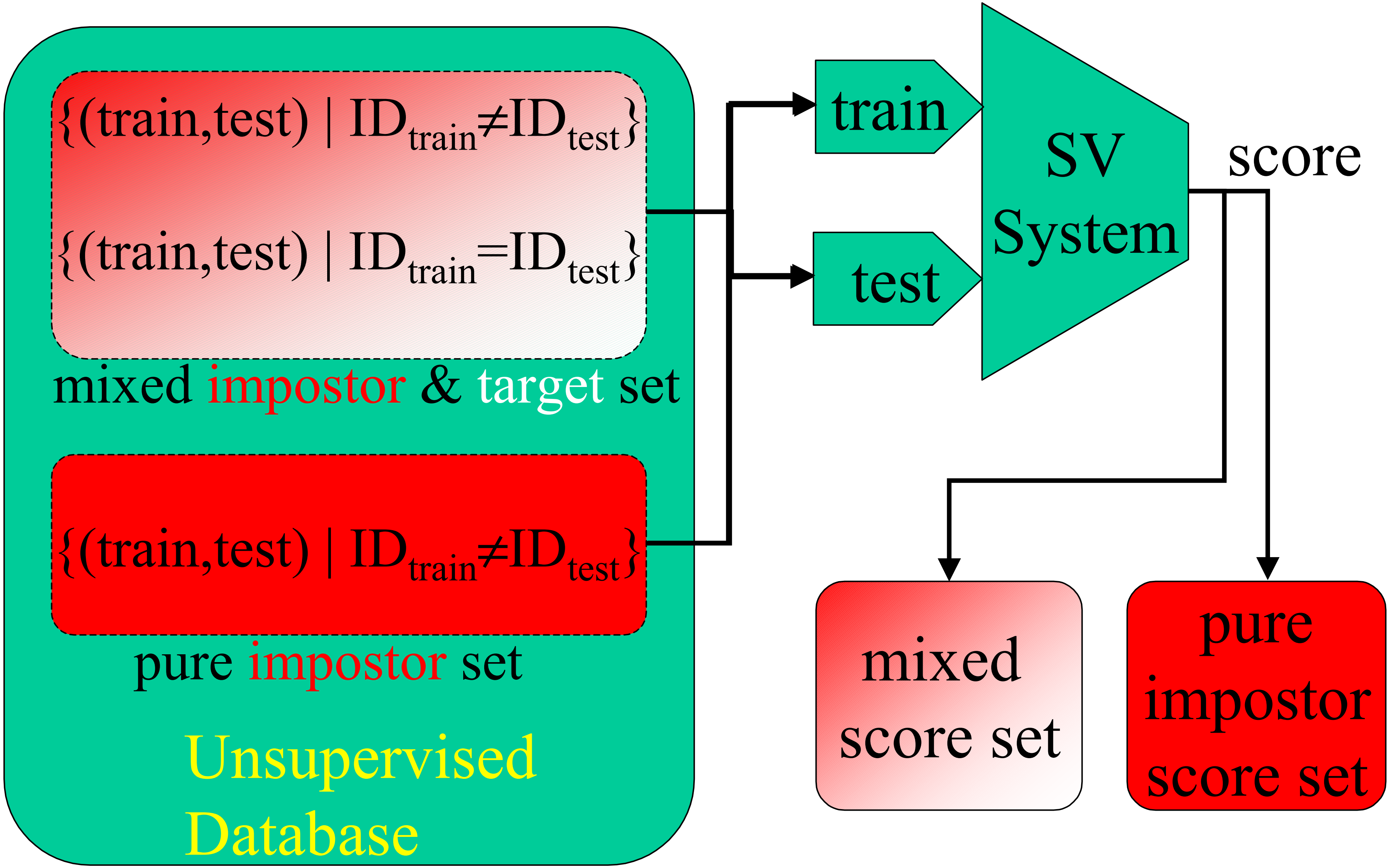
Note: It may be difficult in practice to ensure this requirement in an unsupervised way.

Database Prerequisites

- A separate set of pure impostor test pairs must be available

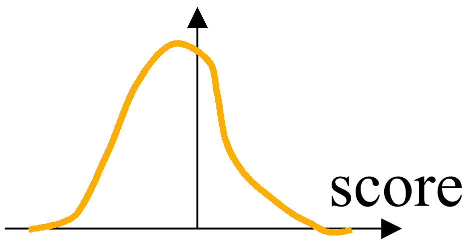
Note: Impostors are not so difficult to get hold of. This requirement is similar to that of SV systems that use impostor normalization schemes like H-norm and T-norm.

Prerequisites



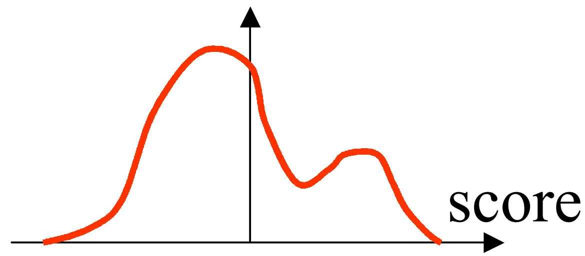
Problem summary:

Given:



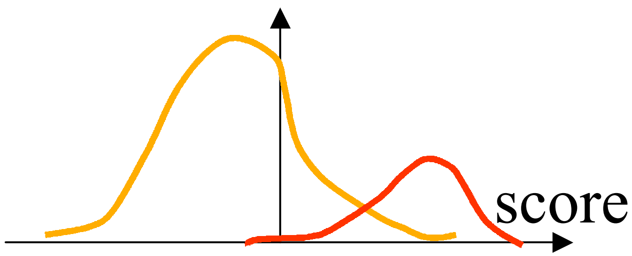
pure impostor scores

and

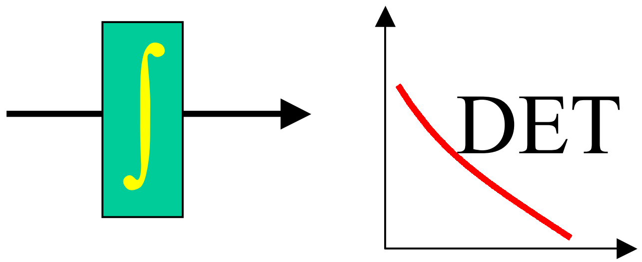


mixed scores

Estimate:



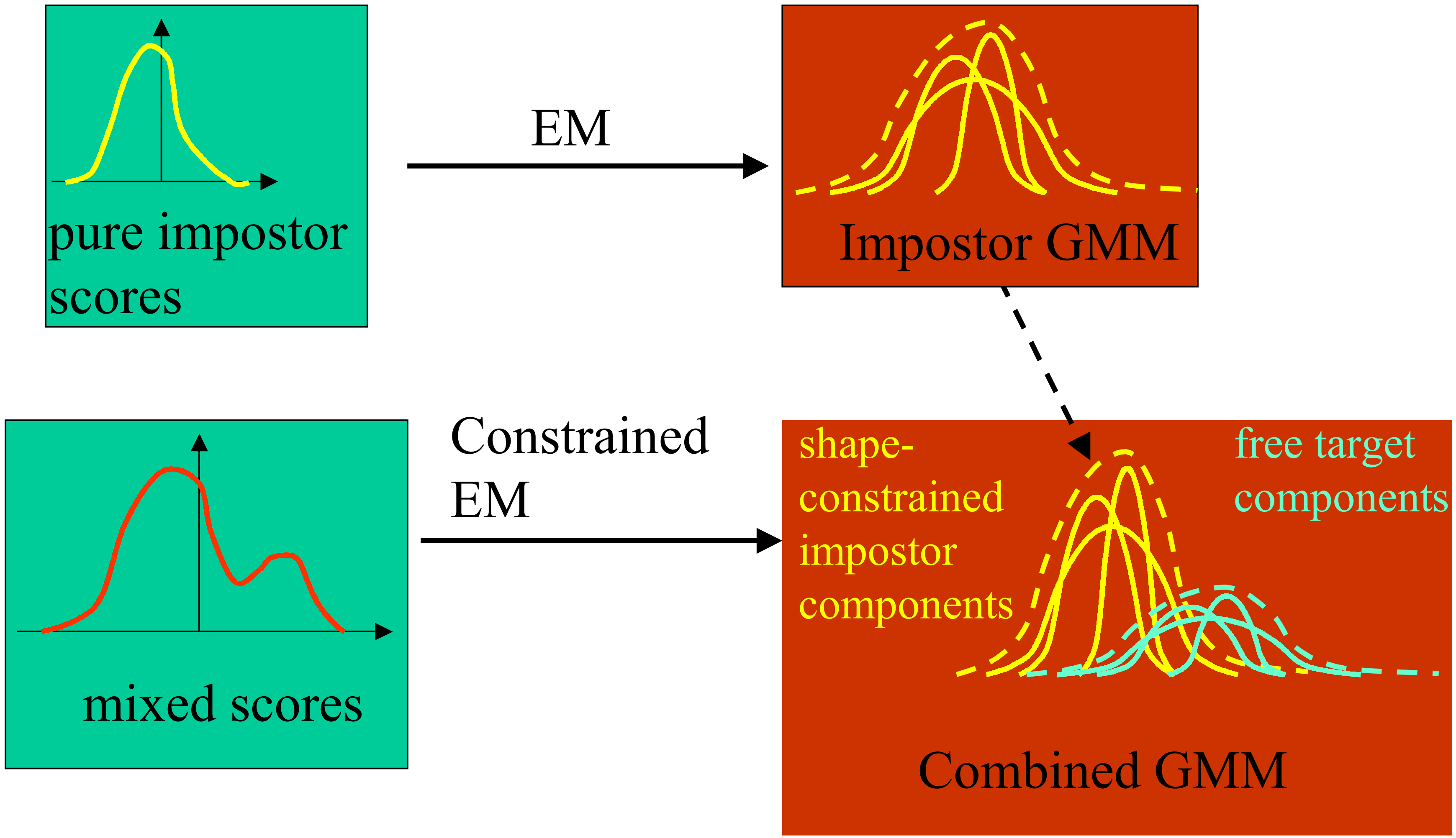
impostor & target
score distributions



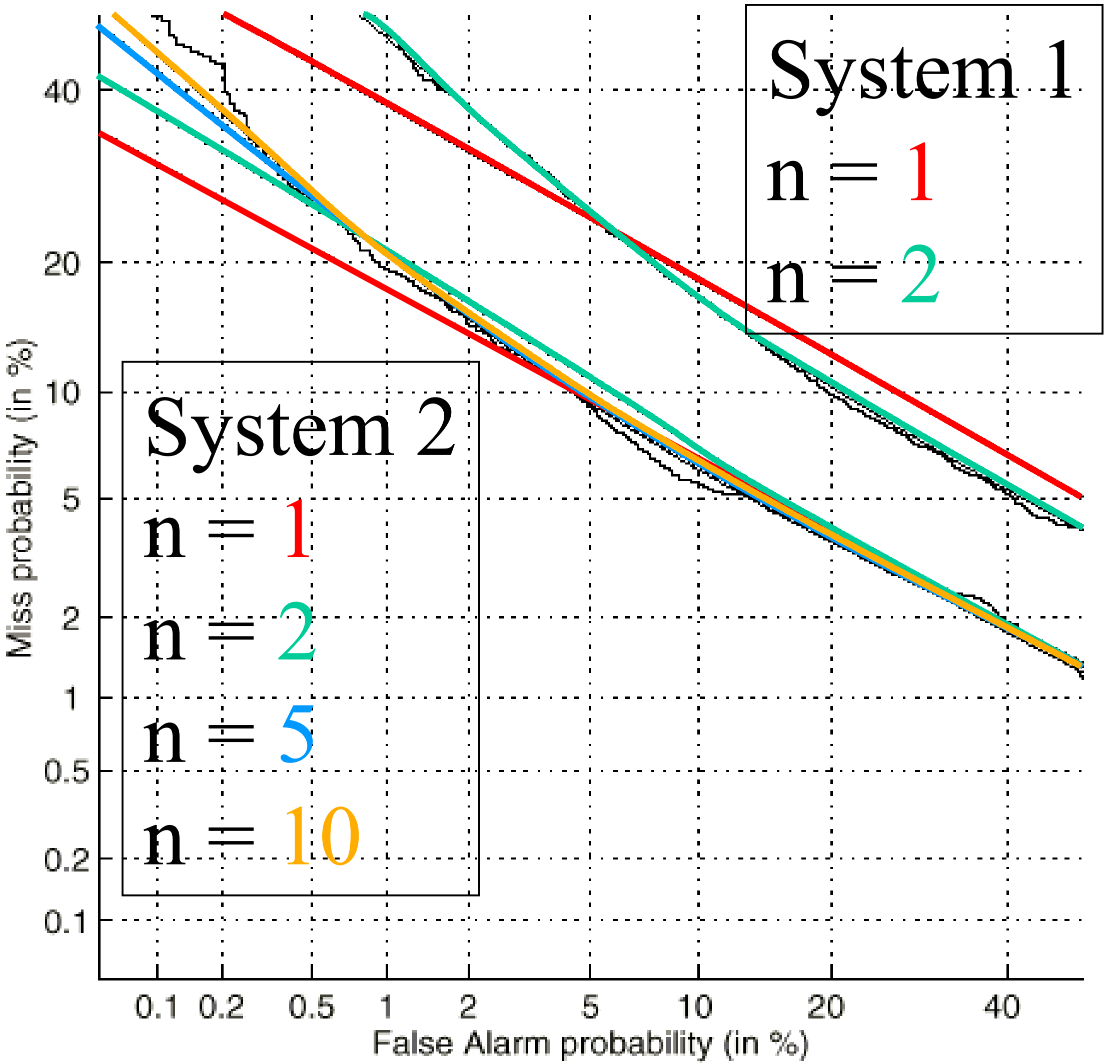
Complicating factors

- impostor distribution may change
- ratio of impostors to targets is unknown
- the problem is not a simple subtraction: we start with data sets, not distribution functions
- distributions are not Gaussian

Blind estimation approach



GMM DET Modelling



Comparison of DETs obtained directly from data vs. dual *n*-component GMM's trained separately on impostor & target scores

Impostor score model

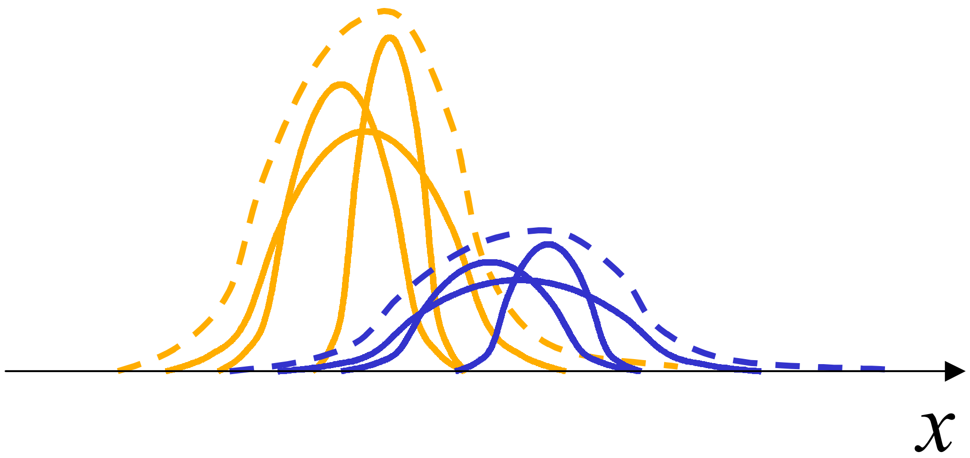
- One dimensional *n*-component GMM :

$$p(x) = \Sigma_{i \in I}\ q_i\ N(x\ ,\ \mu_i\ ,\ \sigma_i)$$

$N(\cdot)$ is Gaussian distribution

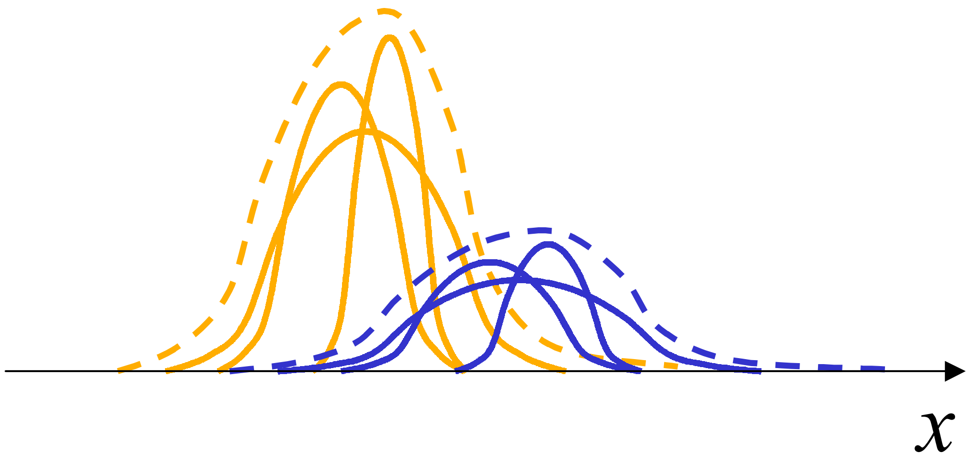
Mixed score model

$$p(x) = P_{imp} \; p_{imp}(x) + P_{tar} \; p_{tar}(x)$$



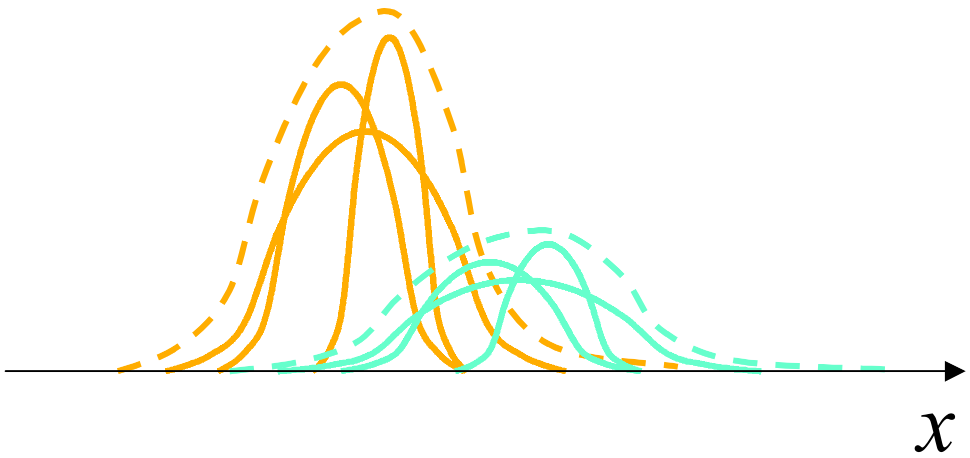
Mixed score model: ratio

$$P_{imp} + P_{tar} = 1$$



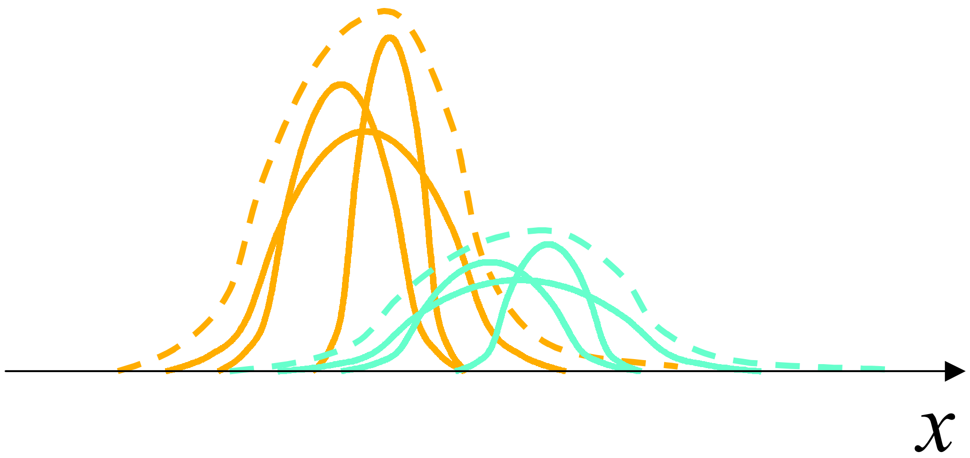
Mixed score model:
impostor offset

$$p_{imp}(x) = \Sigma_{i \in I}\ q_i\ N(x\ ,\ \alpha + \mu_i\ ,\ \sigma_i)$$



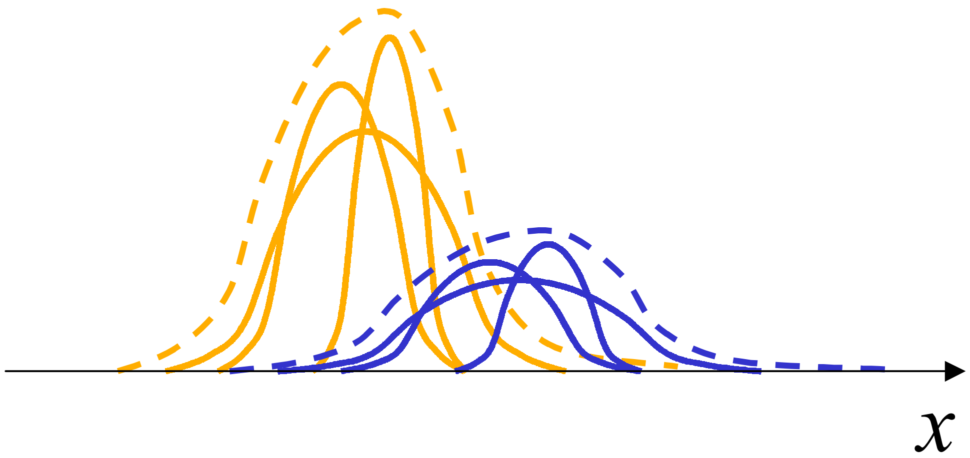
Mixed score model:
impostor variance

$$p_{imp}(x) = \Sigma_{i \in I}\ q_i\ N(x\ ,\ \alpha + \gamma \mu_i\ ,\ \gamma \sigma_i)$$

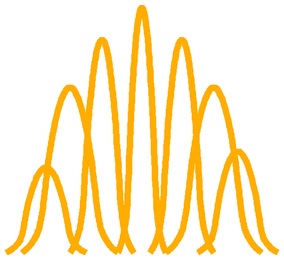


Mixed score model:
target distribution

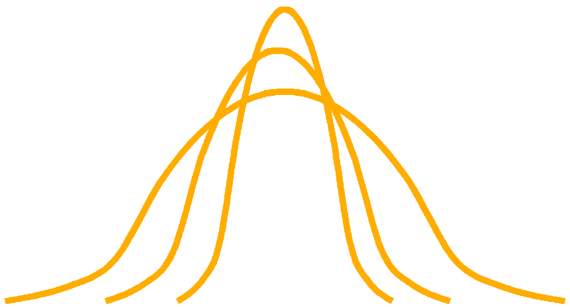
$$p_{tar}(x) = \Sigma_{i \in T}\, r_i\, N(x\, ,\, \beta_i\, ,\, \delta_i)$$



Impostor GMM Initialization

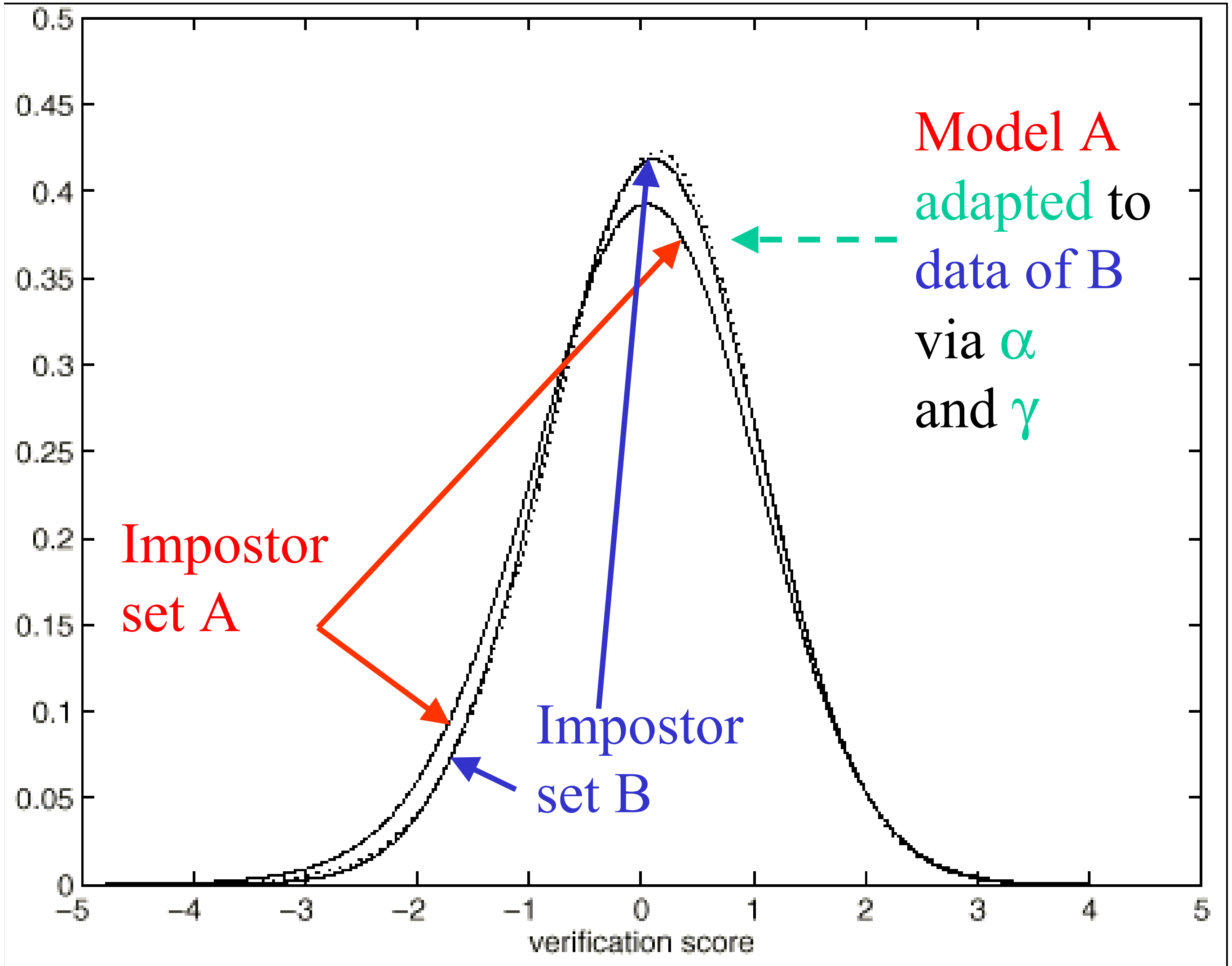


✘ Via k-means VQ



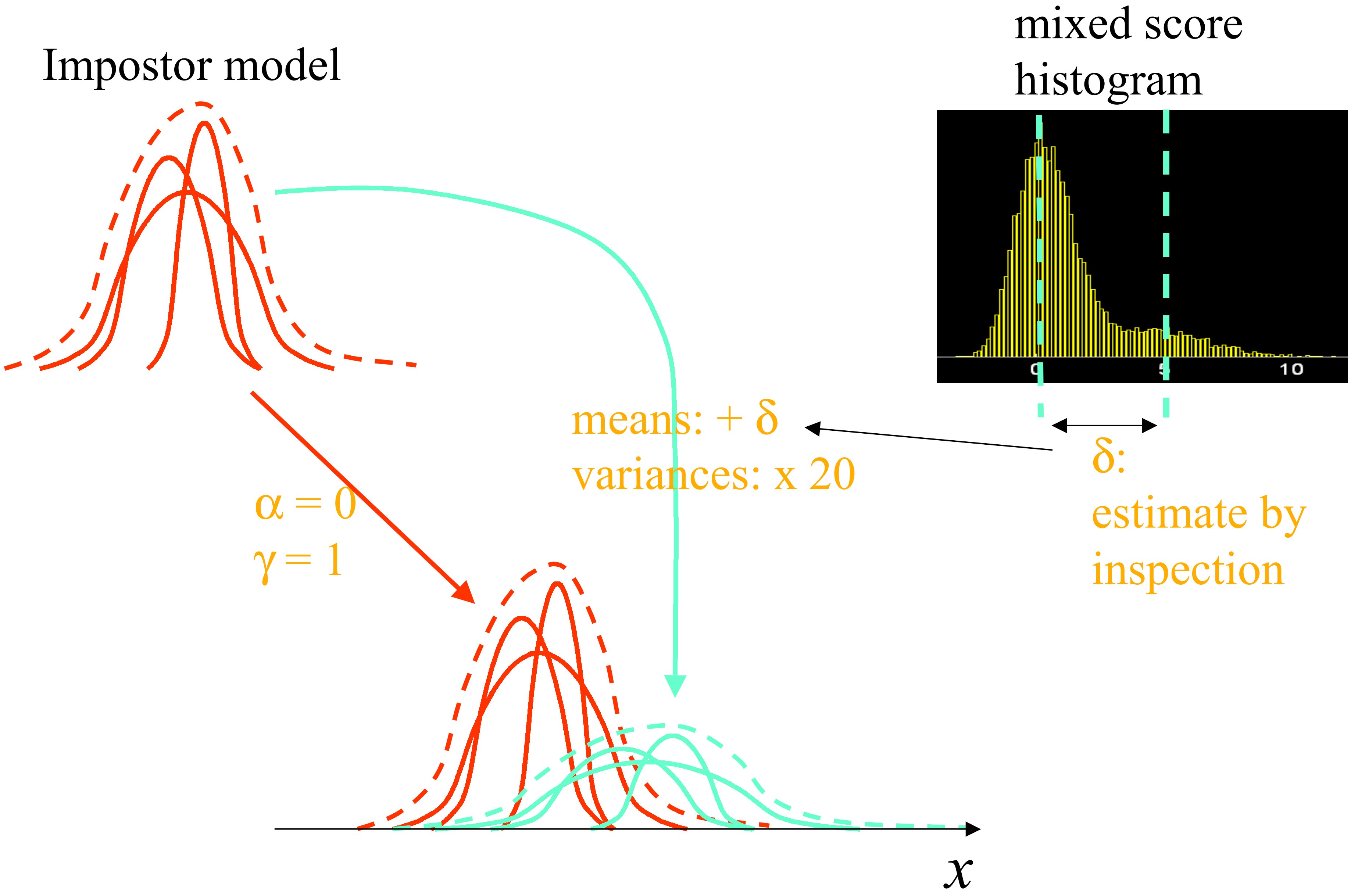
✔ Concentric means,
range of variances

Impostor model adaptation



real
data

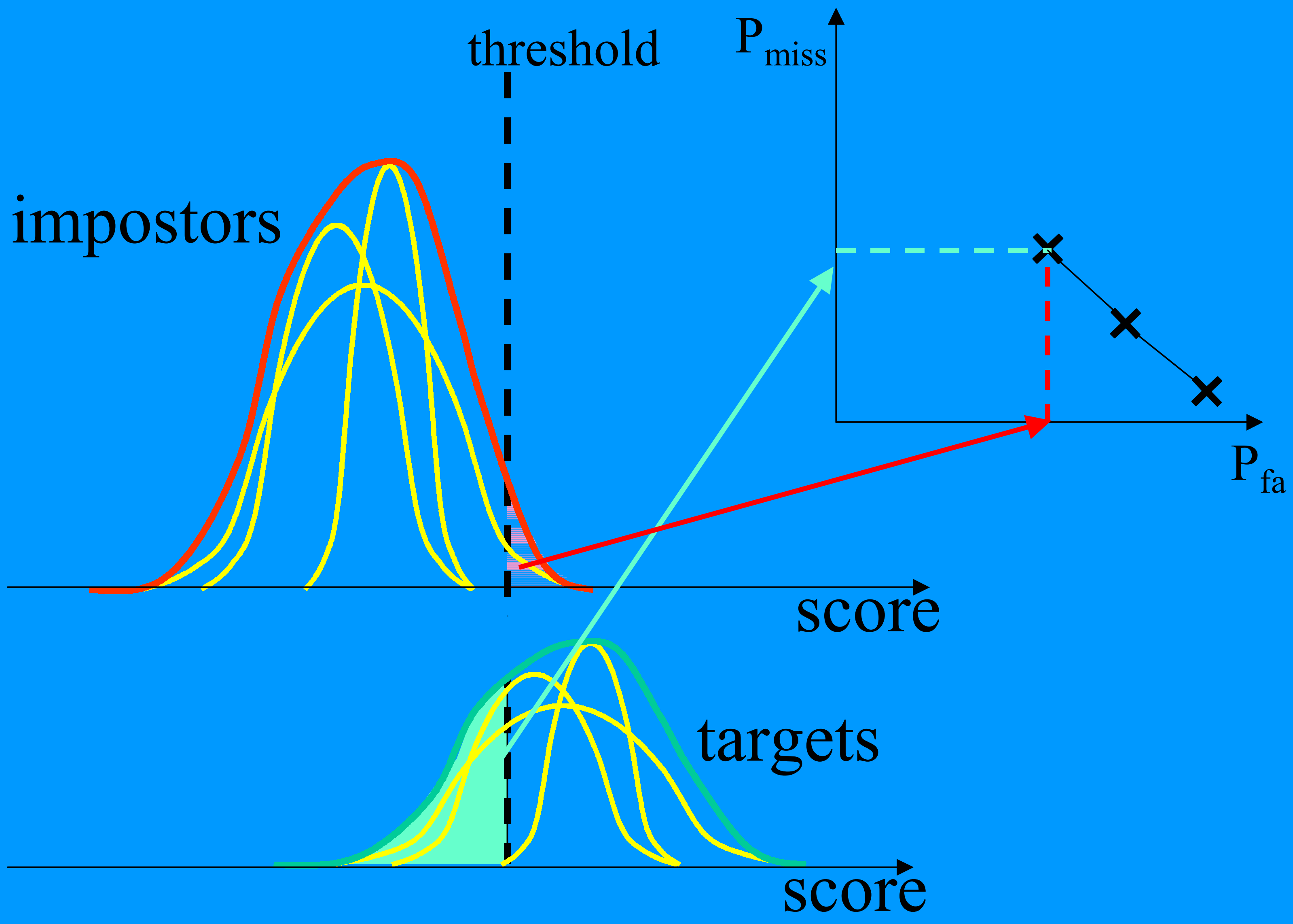
Combined model initialization



Combined model estimation

- Run several *constrained* EM iterations on mixed data:
  - impostor parameters stay fixed
  - α and γ are allowed to adapt
  - target parameters adapt freely
- See EM re-estimation formulae in main article. The formulae for α and γ are not trivial.

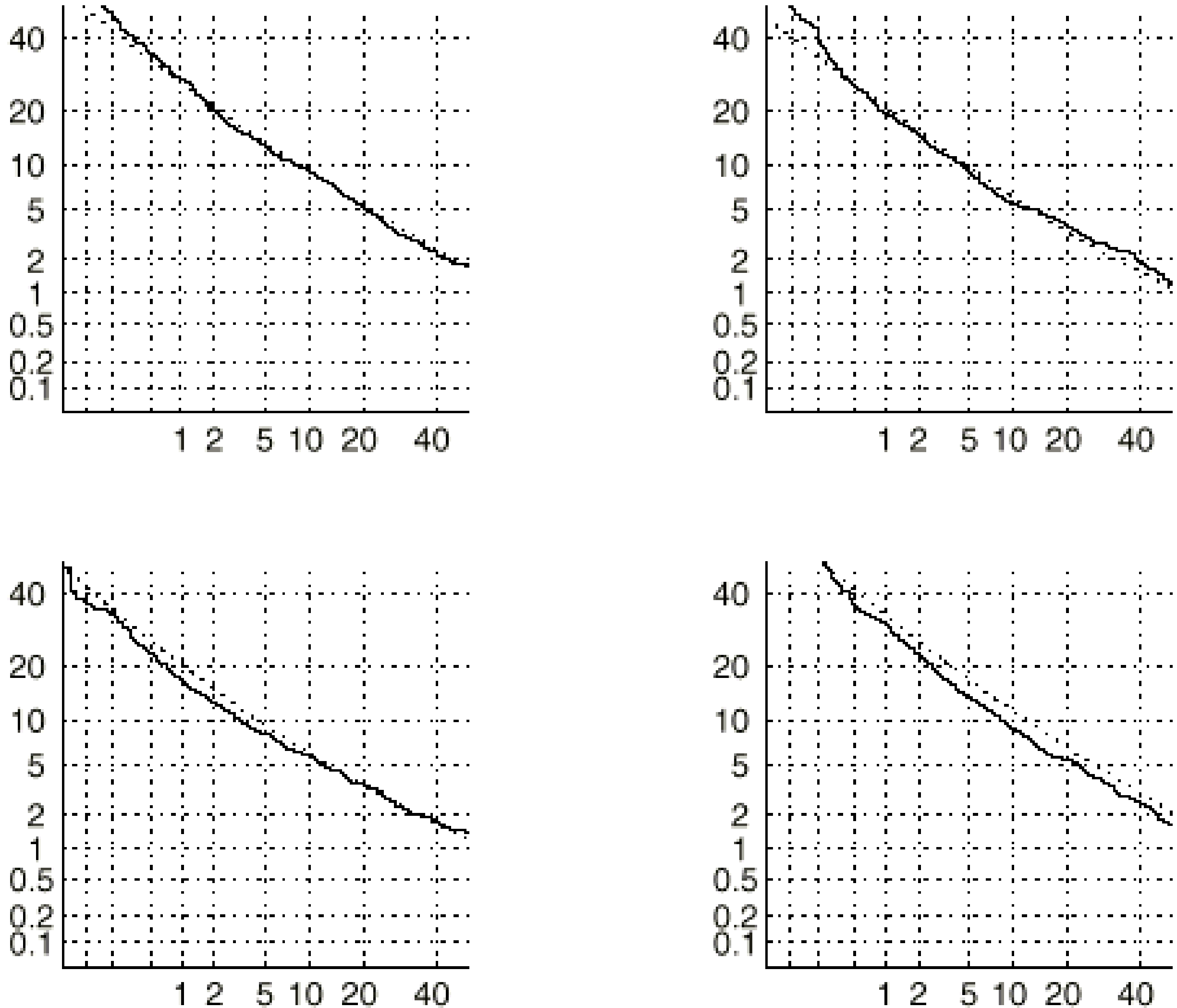
DET Calculation: Use erf(·)



Blind DET Estimation Experiments

- Tested on scores of the SV systems of 9 of the NIST 2000 participants, run on the *1-speaker detection* part of the NIST 2000 Evaluation
  - electret only training & testing
  - males only
  - impostors partitioned into two equal sets, to provide pure-impostor score set

Experimental results



true

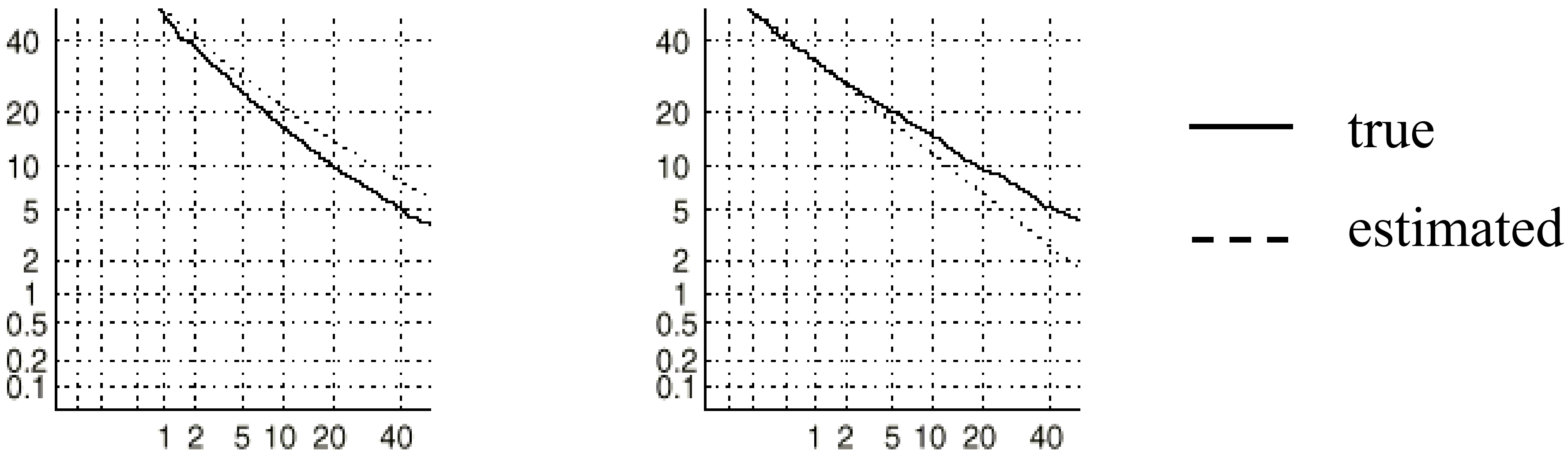
estimated

EER < 10%

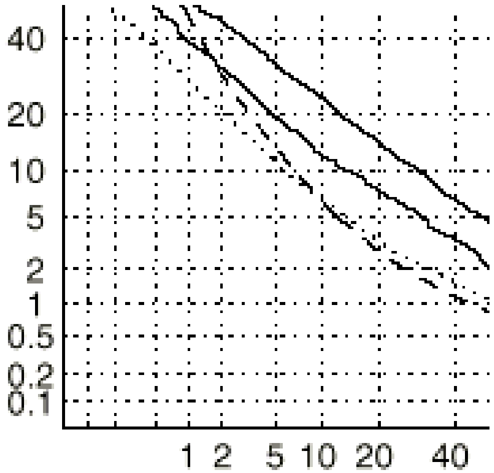
Experimental results

EER > 10%

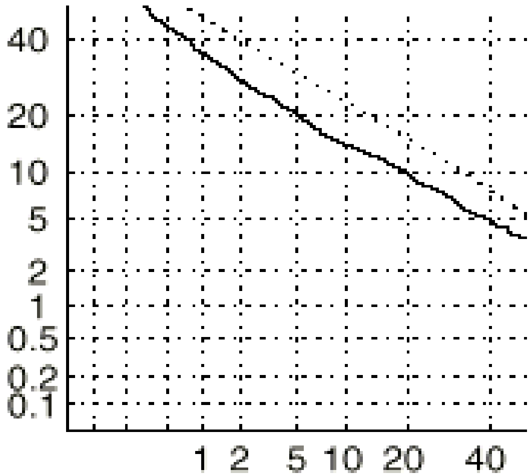
P_{tar} estimates close to correct



Experimental results: EER > 10%

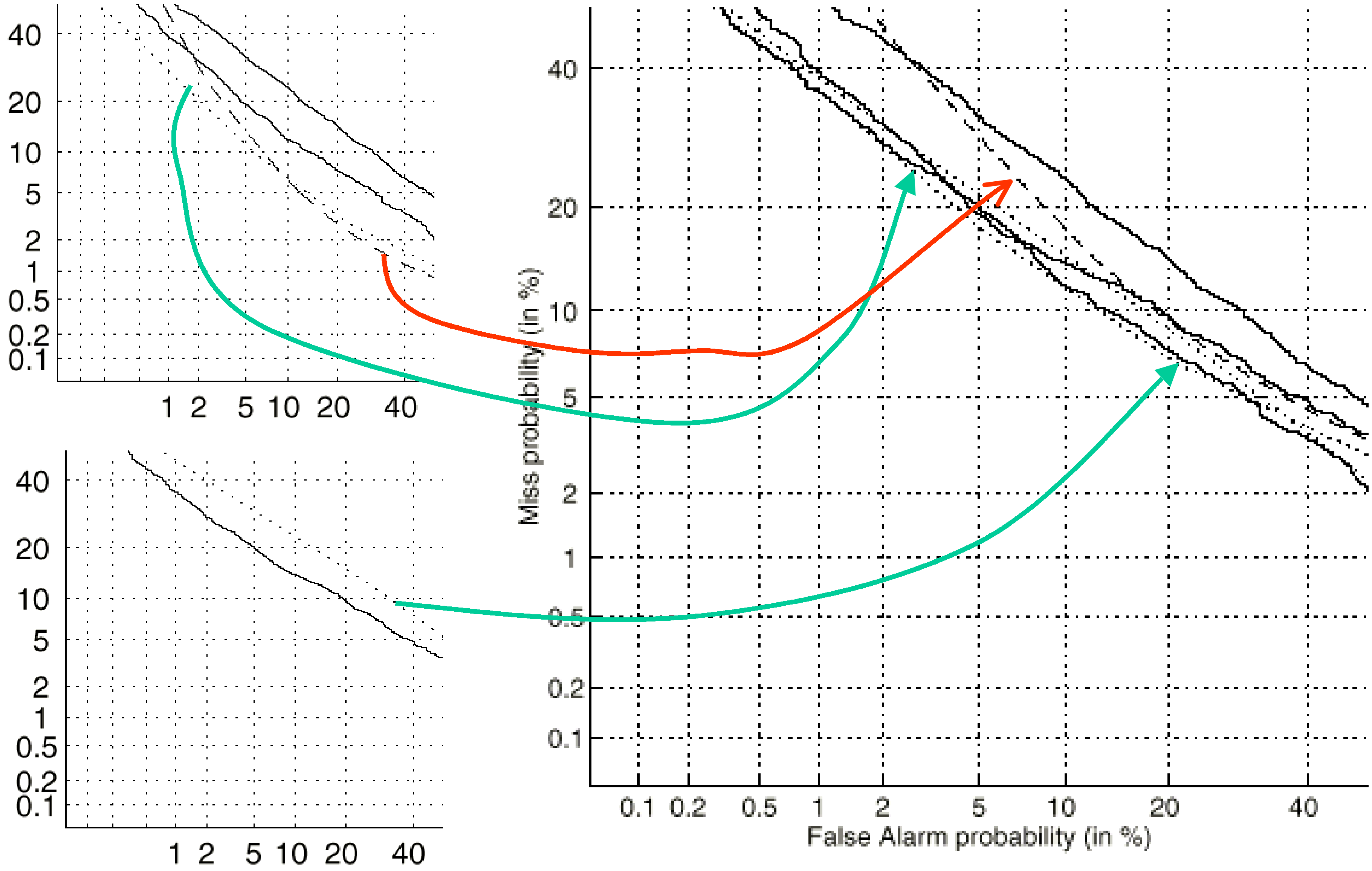


P_{tar} underestimated
- overoptimistic DET



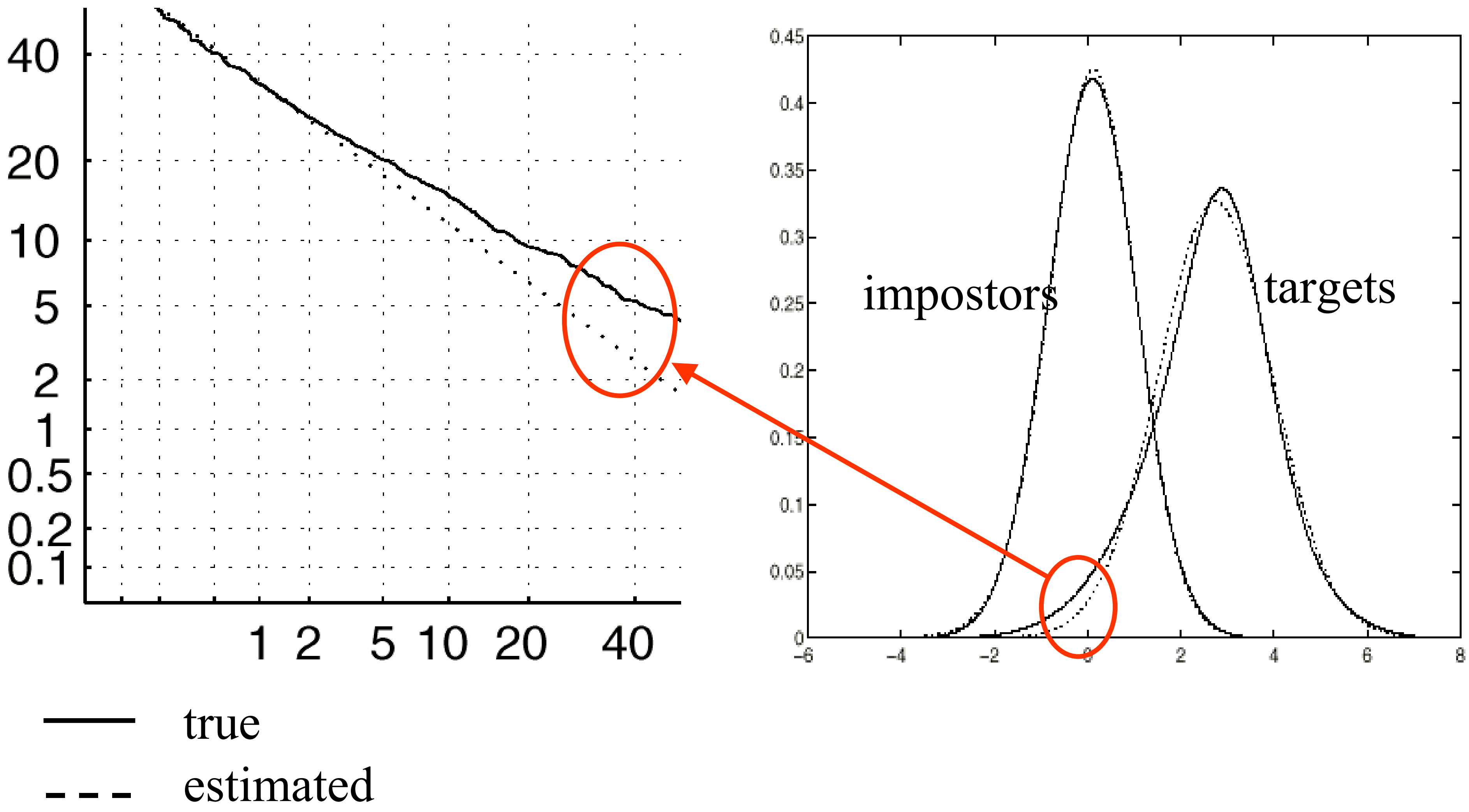
P_{tar} overestimated
- pessimistic DET

Fixing P_{tar}

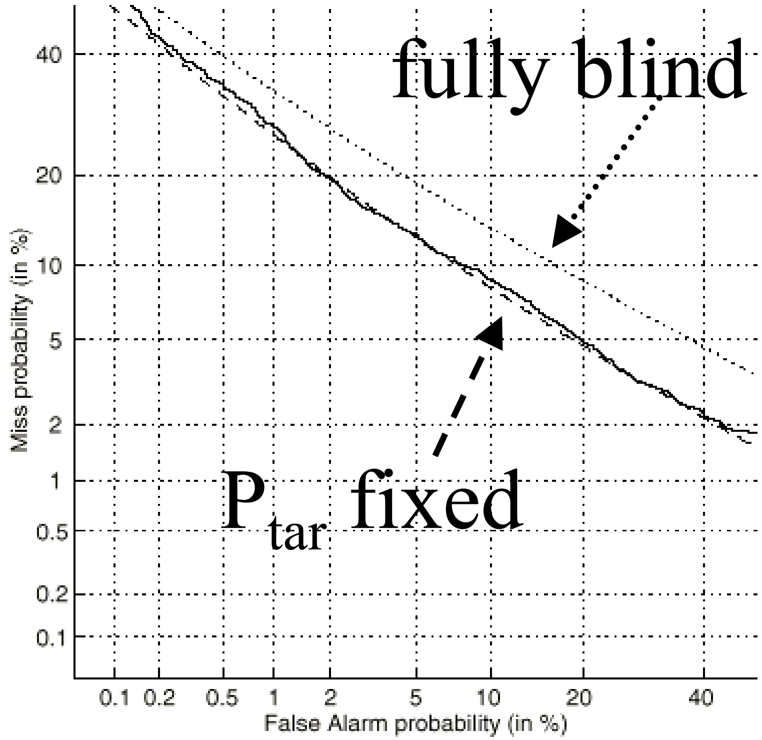


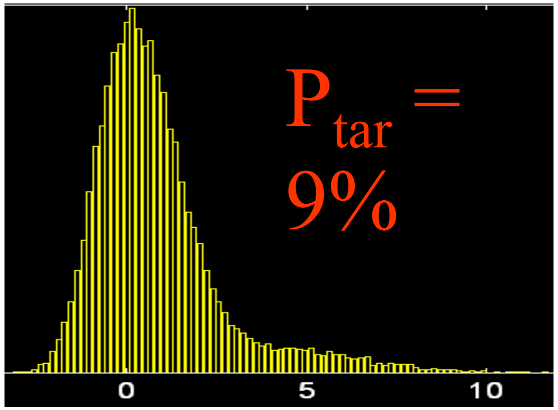
40
20
10
5
2
1
0.5
0.2
0.1
1 2 5 10 20 40
40
20
10
5
2
1
0.5
0.2
0.1
1 2 5 10 20 40
Miss probability (in %)
40
20
10
5
2
1
0.5
0.2
0.1
0.1 0.2 0.5 1 2 5 10 20 40
False Alarm probability (in %)

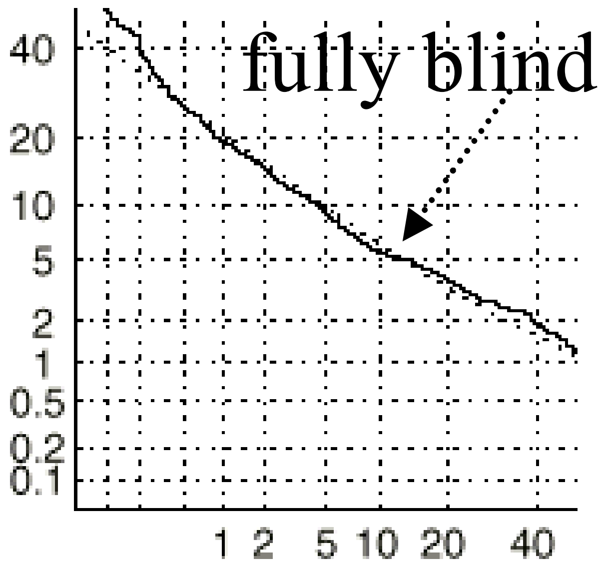
Estimated distributions

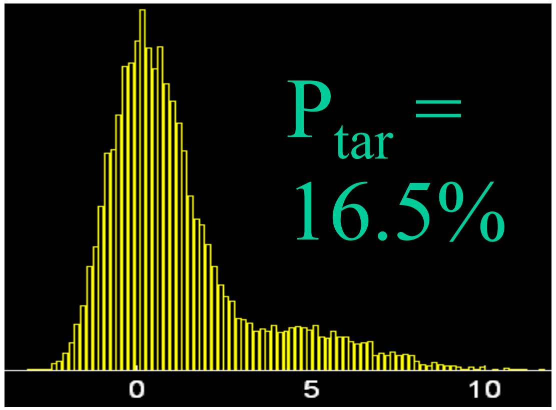


Smaller target subset









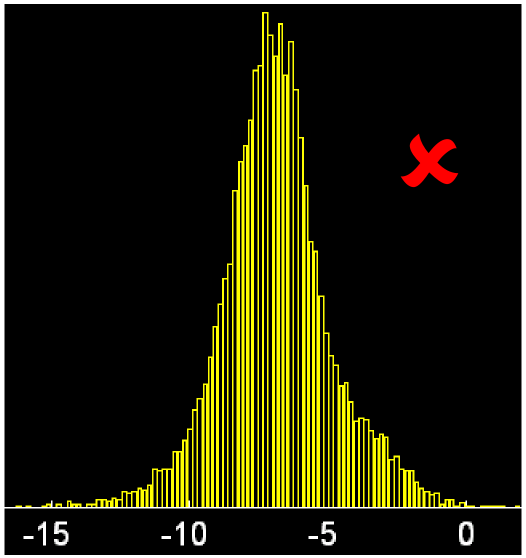
Conclusions

- Blind DET estimation is possible when:
  - There is small overlap of distributions when:
    - Error rates are low enough
    - When target:impostor ratio is not too extreme
  - Shape of the impostor distribution stays unchanged between pure impostor and mixed score sets. (Mean and variance changes can be compensated for.)
- Correct P_{tar} estimate is crucial. It helps if this value is known.

When can we use this method?

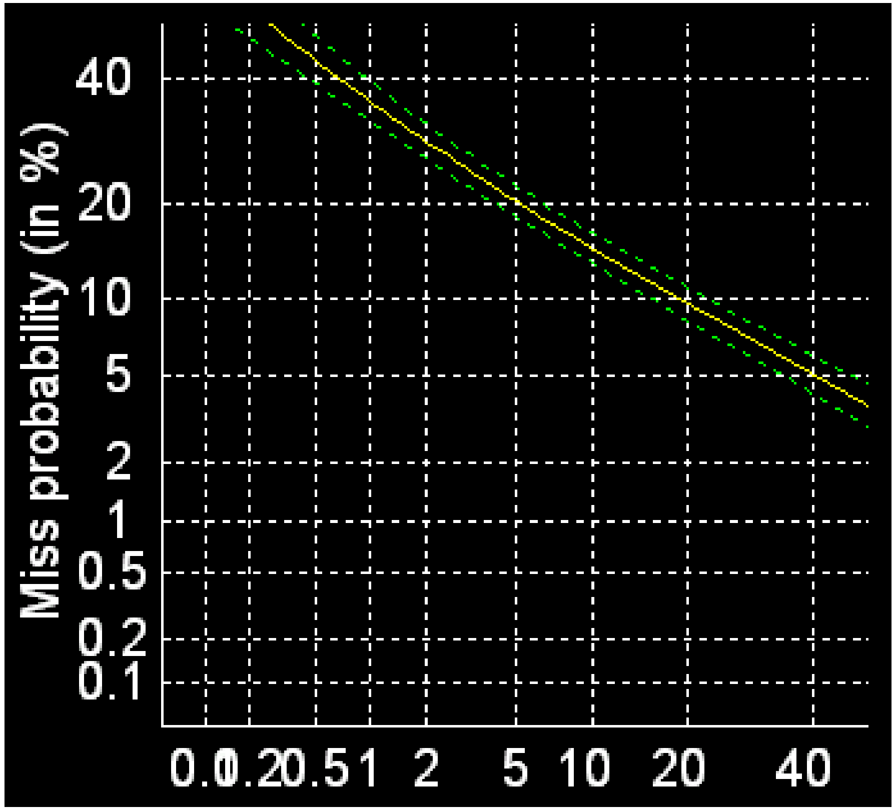
• Subject for future research to get a confidence estimate in the result.

• Look at the data. If you can clearly distinguish impostor & target distributions it will probably work.



Other uses

- Use GMMs to generate synthetic data sets, to estimate confidence intervals for DET curves. (Blind or supervised.)



90% Confidence interval

Other uses: speaker model adaptation

- Use multidimensional α-γ adaptation of speaker GMMs when recognizing speakers in mismatched conditions.
- For speaker recognition in conversations, when a GMM exists for one of the speakers. This is very similar to the blind DET estimation problem.