A bi-Gaussian method for calibration of likelihood ratios

Geoffrey Stewart Morrison

Forensic Data Science Laboratory
Aston Institute for Forensic Linguistics

Aston University
BIRMINGHAM UK

Acknowledgment

- This research was supported by Research England's Expanding Excellence in England Fund as part of funding for the Aston Institute for Forensic Linguistics 2019–2024.

Disclaimer

- All opinions expressed are those of the presenter and, unless explicitly stated otherwise, should not be construed as representing the policies or positions of any organizations with which the presenters are associated.

Slides

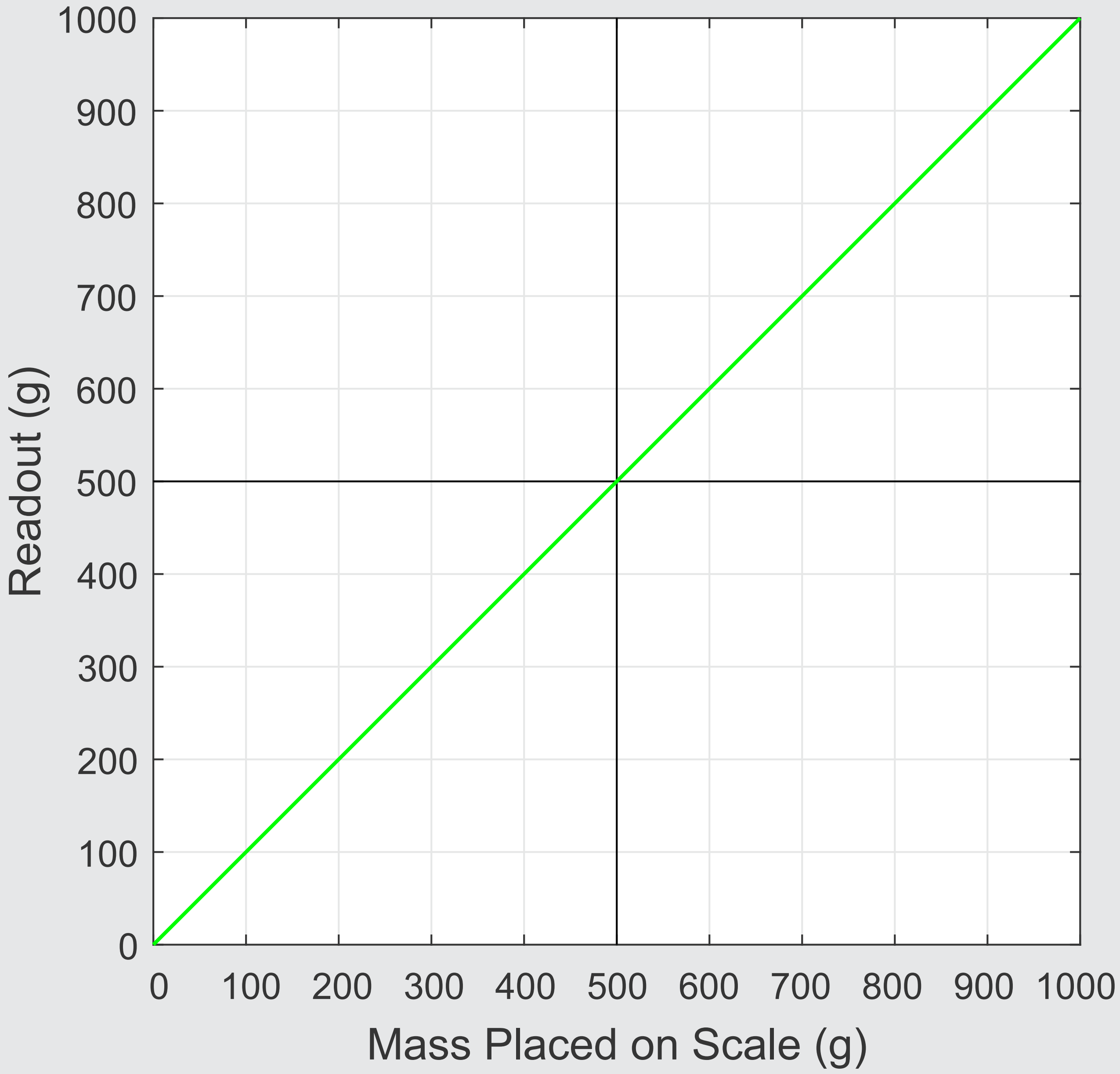
- https://geoff-morrison.net/#ICFIS2023

What is Calibration?

- What is a well-calibrated set of scales?
- A set of scales for which:
  - The mass stated in the readout is the same as the mass placed on the scale

What is Calibration?

- Calibration is the process of adjusting the set of scales so that its output is well calibrated.



What is Calibration?

- What is a well-calibrated likelihood-ratio system?

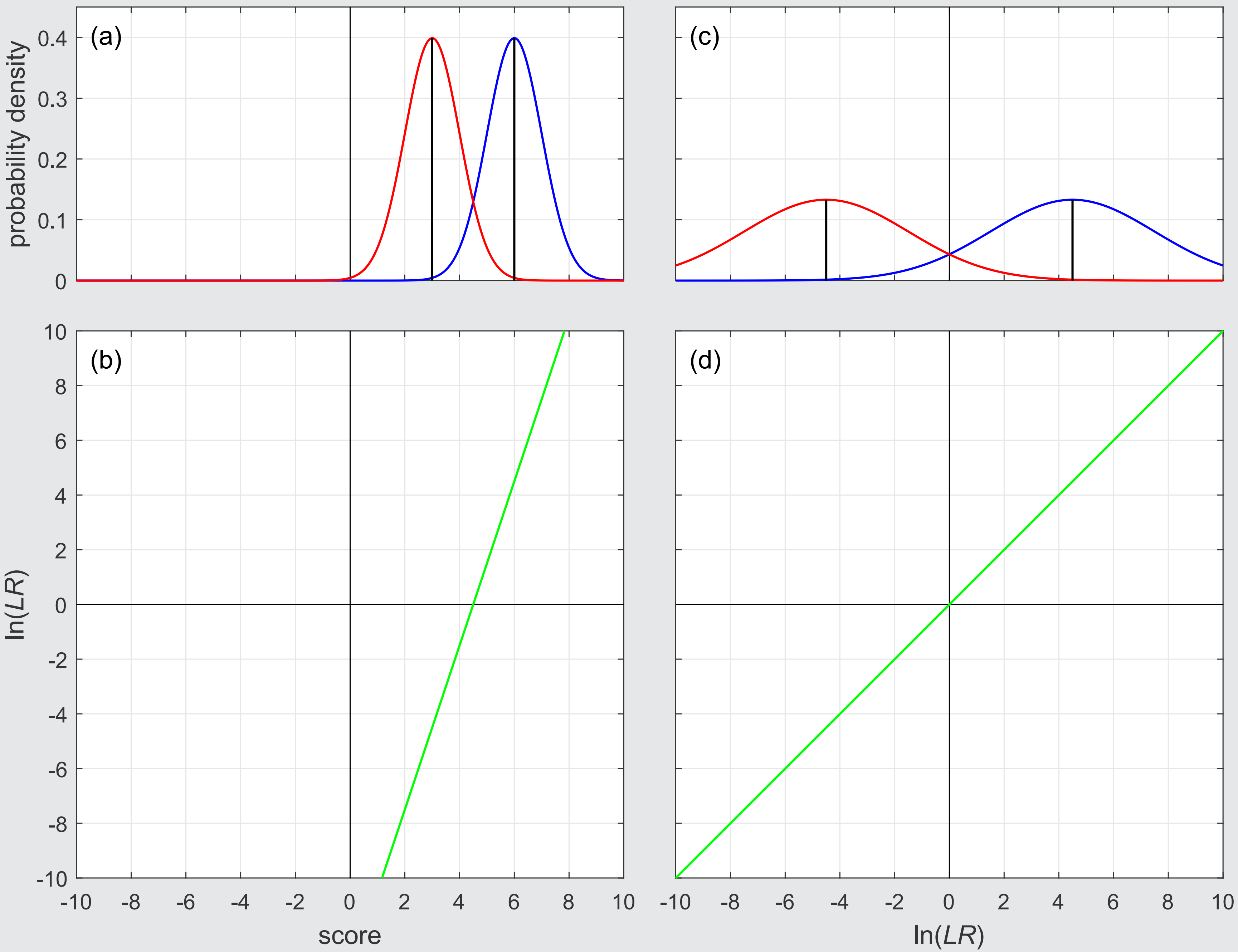
- A system for which:

    - The likelihood ratio of the likelihood ratio is the likelihood ratio

$$LR = \frac{f(LR \mid H_\mathrm{s})}{f(LR \mid H_\mathrm{d})}$$

What is Calibration?

- Calibration is the process of adjusting the system so that its output is well calibrated.

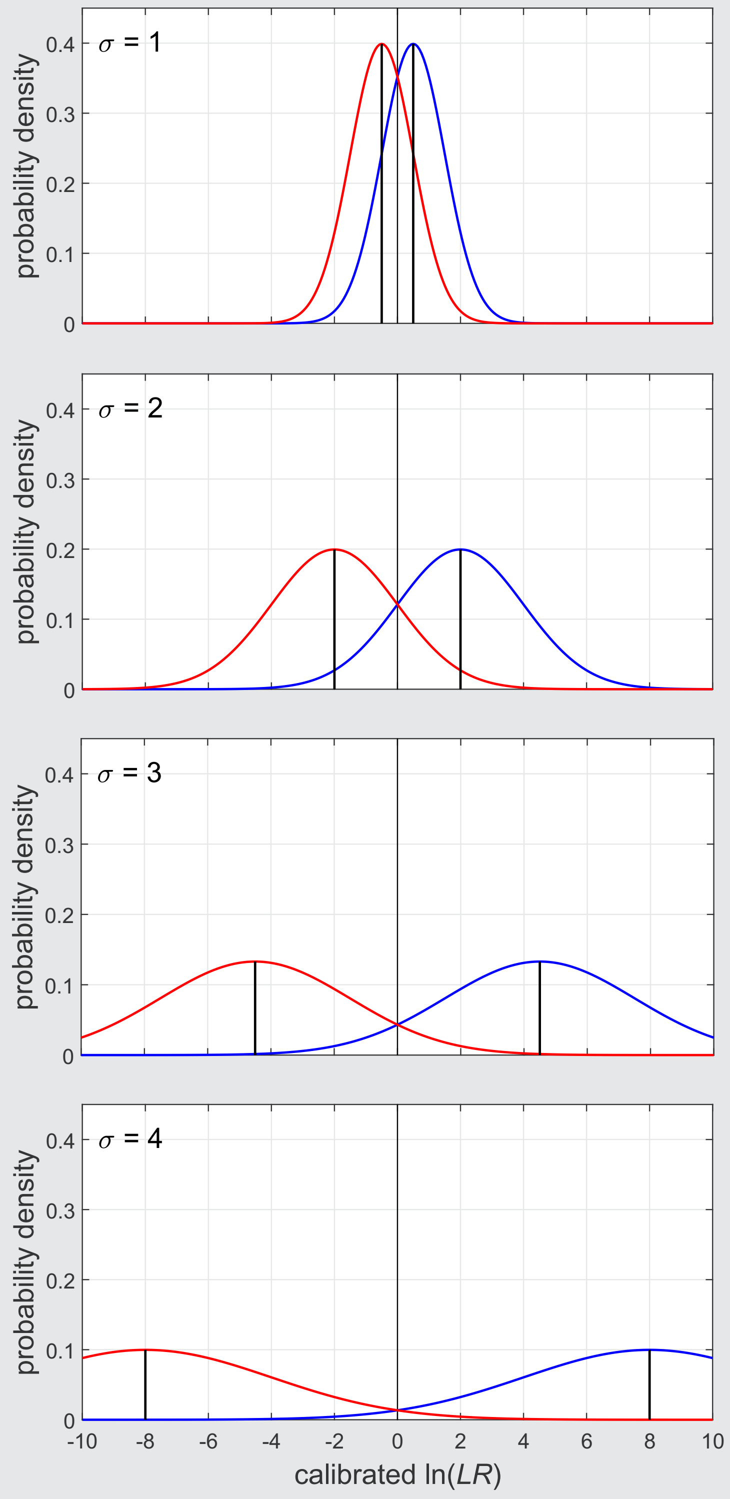


Well-calibrated likelihood ratios

- Perfectly calibrated ln(*LR*) distributions

- Both same-source and different-source distributions are Gaussian, and they have the same variance

$$\mu_\mathrm{d} = -\frac{\sigma^2}{2} \qquad \mu_\mathrm{s} = +\frac{\sigma^2}{2}$$



Well-calibrated likelihood ratios

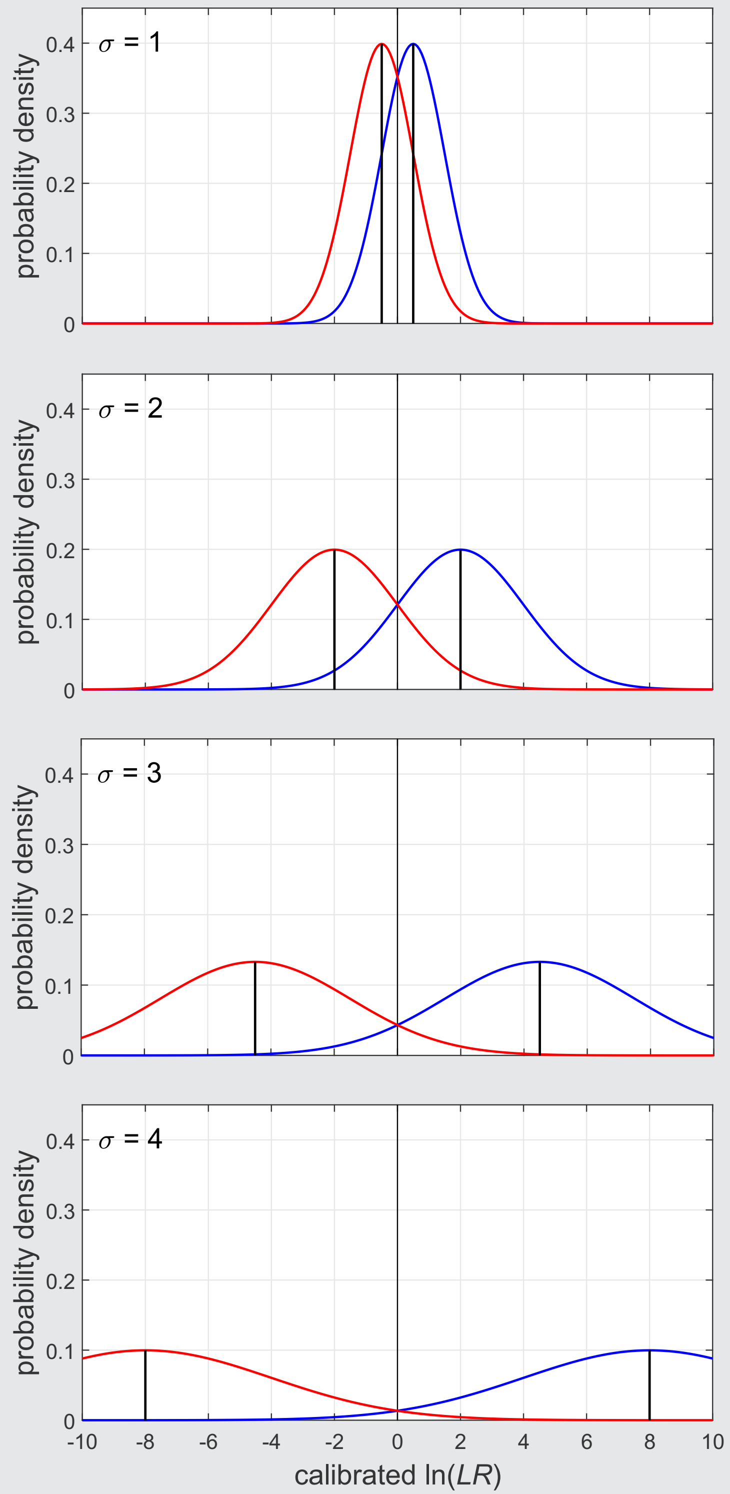
- Perfectly calibrated $\ln(LR)$ distributions
    - C_{llr} values

0.84

0.51

0.24

0.09



Well-calibrated likelihood ratios

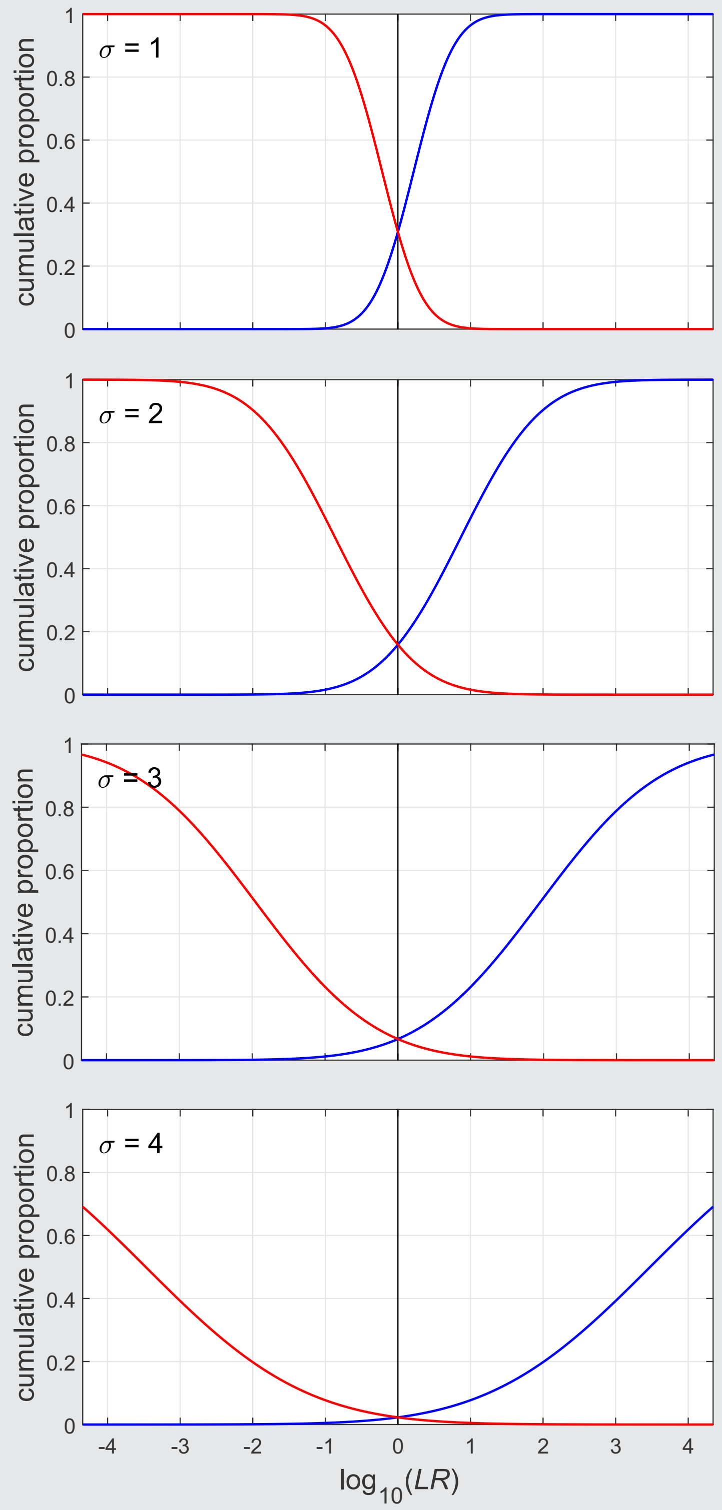
- Perfectly calibrated ln(*LR*) distributions
    - C_{llr} values
    - Tippett plots

0.84

0.51

0.24

0.09



Bi-Gaussianized calibration

1. Calculate uncalibrated likelihood ratios (scores) for training data and test data.

2. Calibrate the training-data output of Step 1 using logistic regression.

3. Calculate C_{llr} for the output of Step 2.

4. Determine the σ^2 of the perfectly-calibrated bi-Gaussian system with the C_{llr} calculated at Step 3.

5. Ignoring same-source and different-source labels, determine the mapping function from the empirical cumulative distribution of the training-data output of Step 1 to the cumulative distribution of the two-Gaussian mixture with the σ^2 determined at Step 4.

6. Apply the mapping function determined at Step 5 to the test-data output of Step 1.

Bi-Gaussianized calibration

1. Calculate uncalibrated likelihood ratios (scores) for training data and test data.
2. Calibrate the training-data output of Step 1 using logistic regression.
3. Calculate C_{llr} for the output of Step 2.
4. Determine the σ^2 of the perfectly-calibrated bi-Gaussian system with the C_{llr} calculated at Step 3.
5. Ignoring same-source and different-source labels, determine the mapping function from the empirical cumulative distribution of the training-data output of Step 1 to the cumulative distribution of the two-Gaussian mixture with the σ^2 determined at Step 4.
6. Apply the mapping function determined at Step 5 to the test-data output of Step 1.

Relationship between C_{llr} and σ^2

$$C_{\mathrm{llr}} = \frac{1}{2}\left(\frac{1}{N_s}\sum_{i}^{N_s}\log_2\left(1+\frac{1}{\Lambda_{s_i}}\right)+\frac{1}{N_d}\sum_{j}^{N_d}\log_2\left(1+\Lambda_{d_j}\right)\right)$$

$$C_{\mathrm{llr}} = \frac{1}{N_d}\sum_{j}^{N_d}\log_2\left(1+\Lambda_{d_j}\right)$$

$$C_{\mathrm{llr}} = \frac{1}{N_d}\sum_{j}^{N_d}\log_2\left(1+\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{\left(x_j-\mu_d\right)^2}{-2\sigma^2}}\right)$$

$$C_{\mathrm{llr}} = \frac{1}{N_d}\sum_{j}^{N_d}\log_2\left(1+\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{\left(x_j-\left(-\frac{\sigma^2}{2}\right)\right)^2}{-2\sigma^2}}\right)$$

$$C_{\mathrm{llr}} = \frac{1}{\ln(2)}\int \ln\left(1+\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x^2}{\sigma^2}+x+\frac{\sigma^2}{4}\right)}\right)dx$$

Relationship between C_{llr} and σ^2

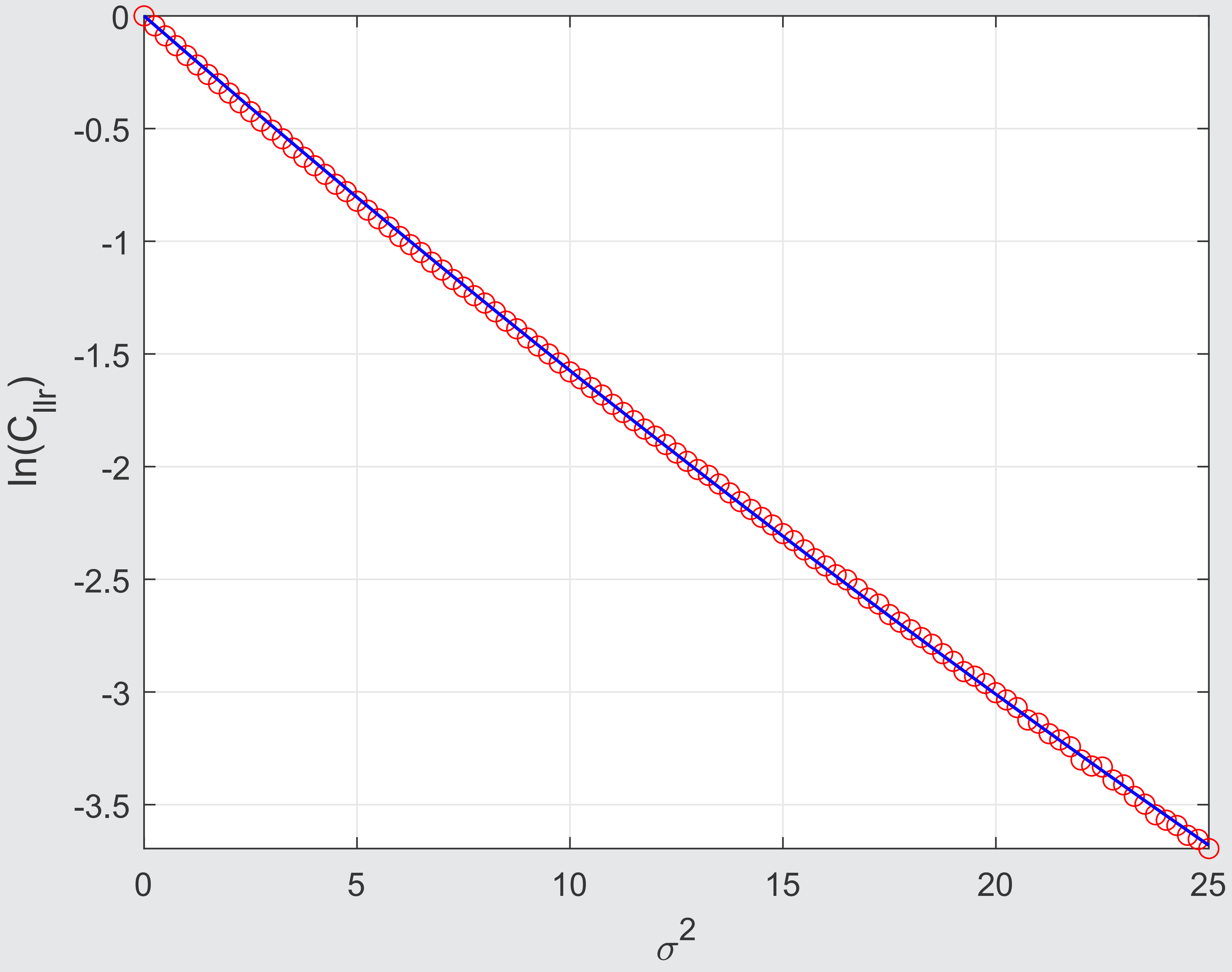
- Via Monte Carlo simulation

$$\ln(C_{\mathrm{llr}}) = a + be^{-c\sigma^2}$$

at $\sigma^2 = 0$, $\ln(C_{\mathrm{llr}}) = 0$, therefore:

$$\ln(C_{\mathrm{llr}}) = b\left(e^{-c\sigma^2} - 1\right)$$

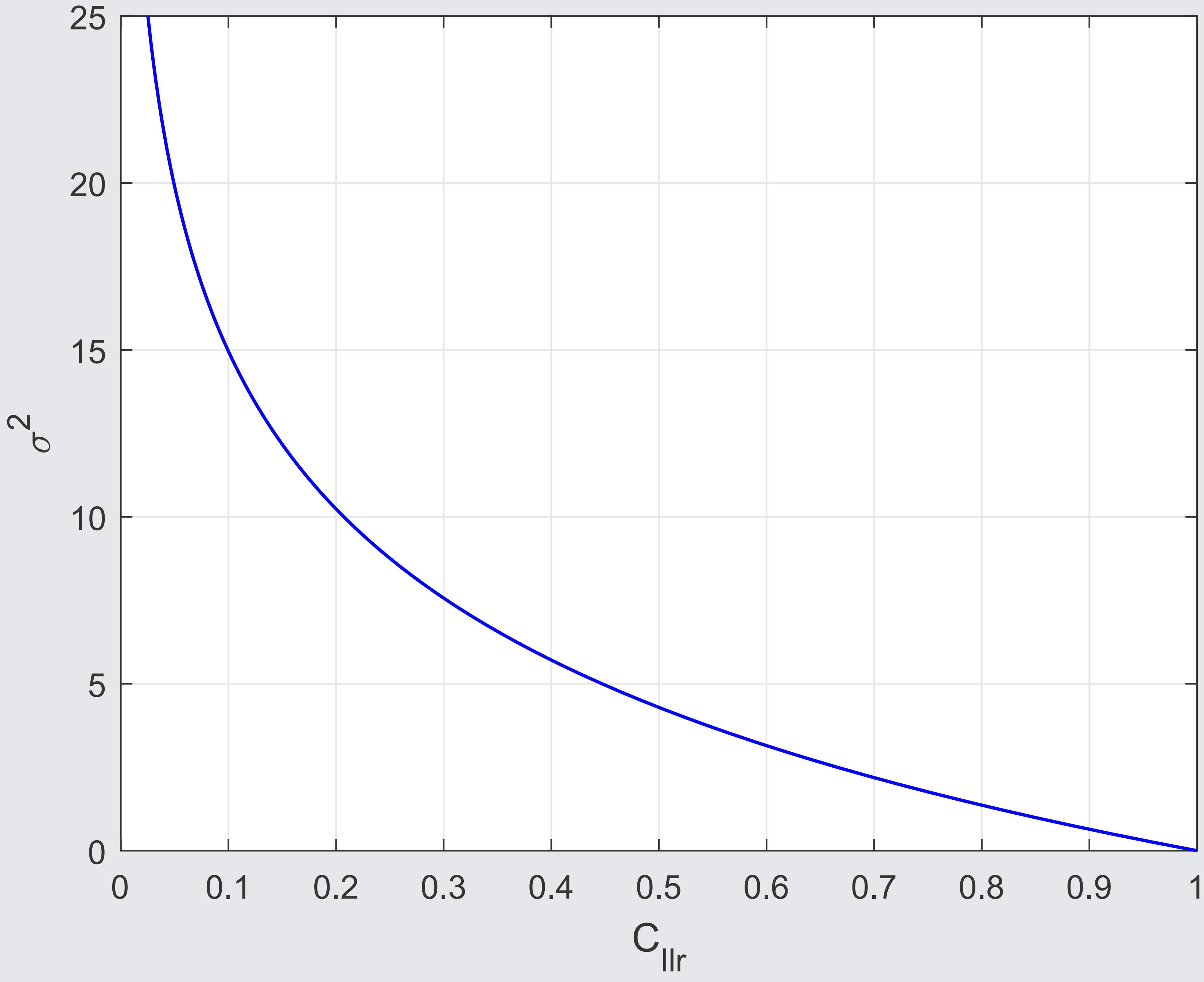
$b = 17.9$, $c = 9.22 \times 10^{-3}$



Relationship between C_{llr} and σ^2

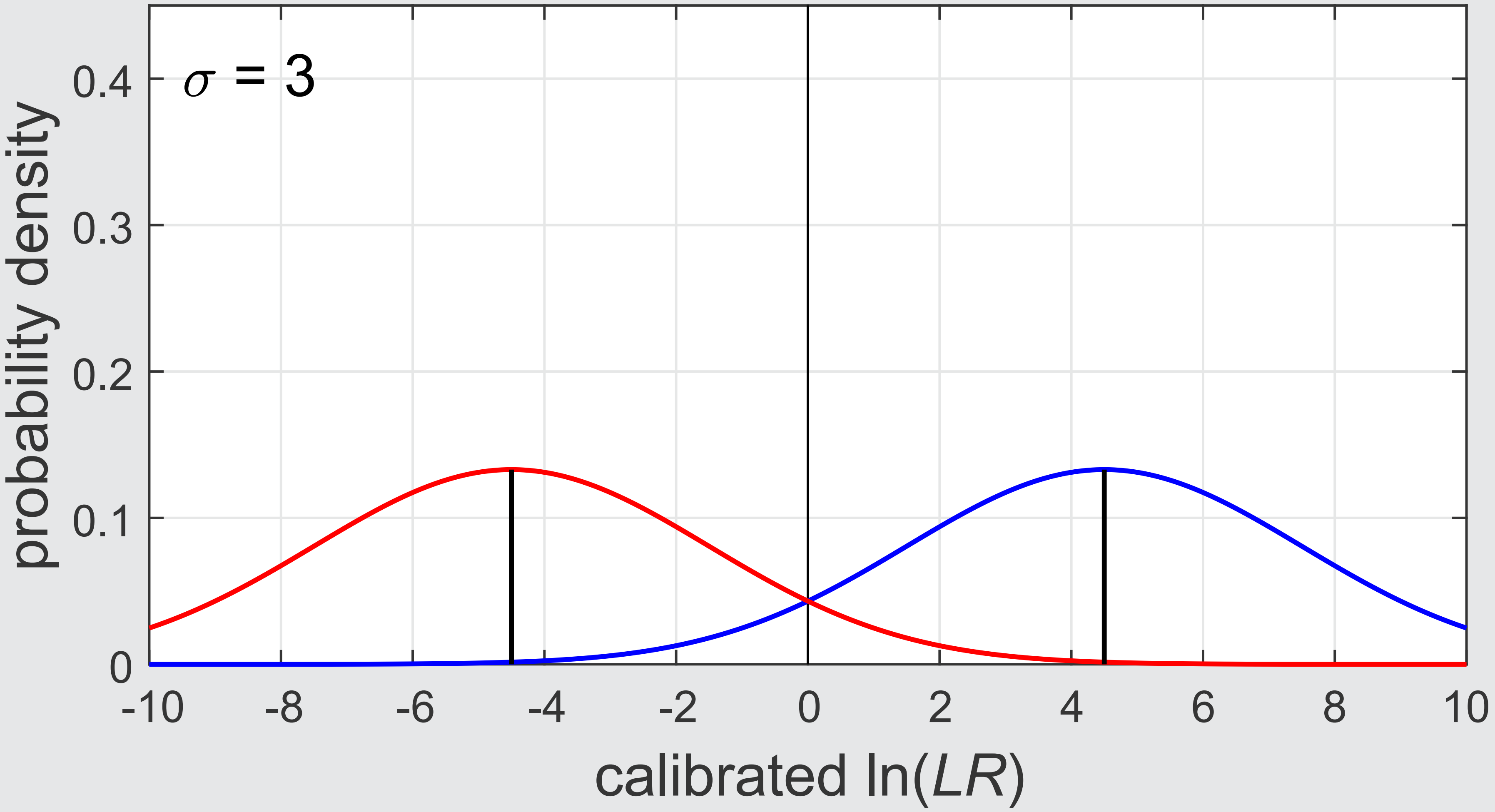
- Via Monte Carlo simulation

$$\sigma^2 = -\frac{\ln\left(\frac{\ln(C_{\text{llr}})}{b} + 1\right)}{c}$$



Simulated data: Gaussians with same variance

- 100 same-source samples
- 4950 different-source samples

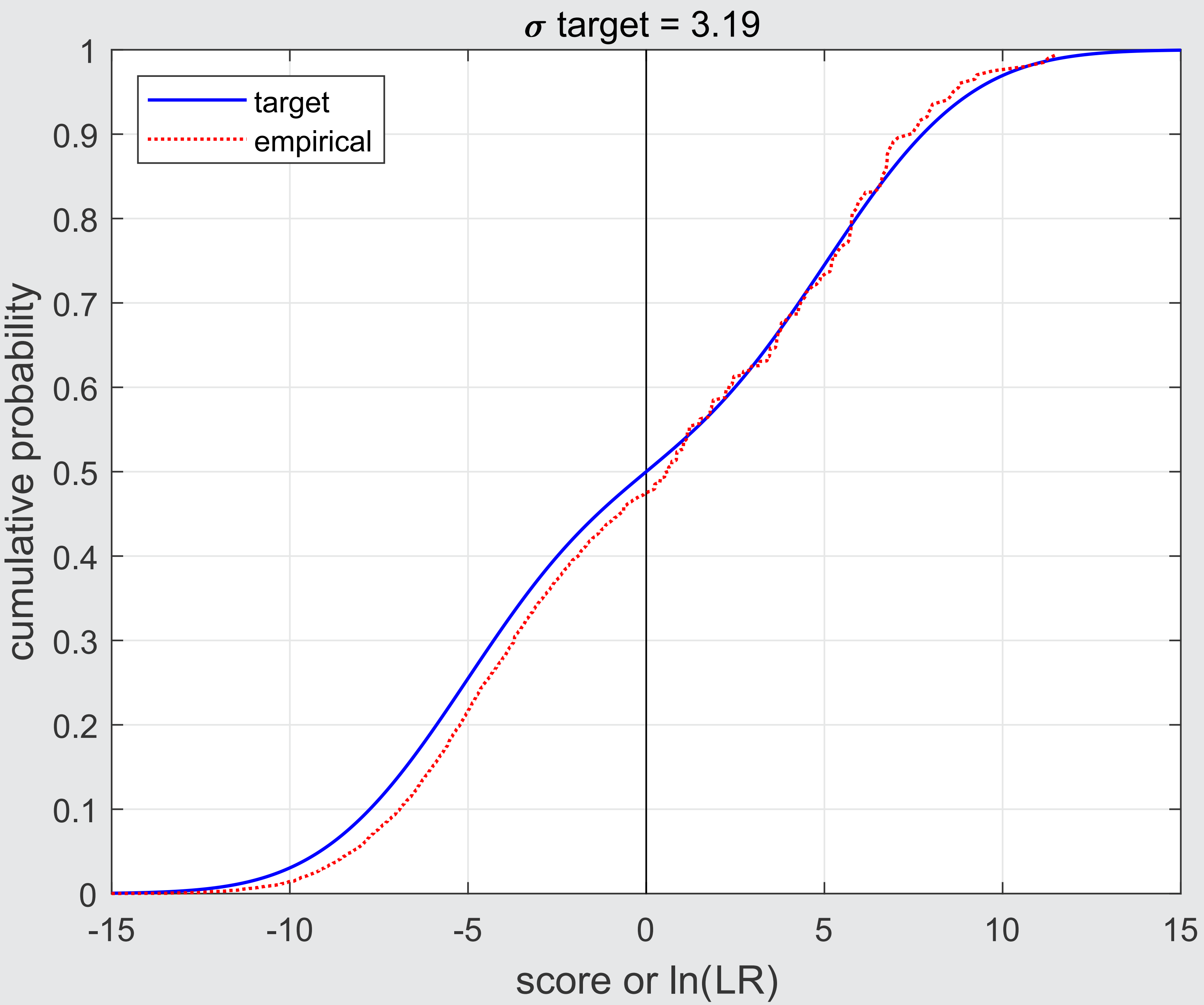


Bi-Gaussianized calibration

1. Calculate uncalibrated likelihood ratios (scores) for training data and test data.
2. Calibrate the training-data output of Step 1 using logistic regression.
3. Calculate C_{llr} for the output of Step 2.
4. Determine the σ^2 of the perfectly-calibrated bi-Gaussian system with the C_{llr} calculated at Step 3.
5. Ignoring same-source and different-source labels, determine the mapping function from the empirical cumulative distribution of the training-data output of Step 1 to the cumulative distribution of the two-Gaussian mixture with the σ^2 determined at Step 4.
6. Apply the mapping function determined at Step 5 to the test-data output of Step 1.

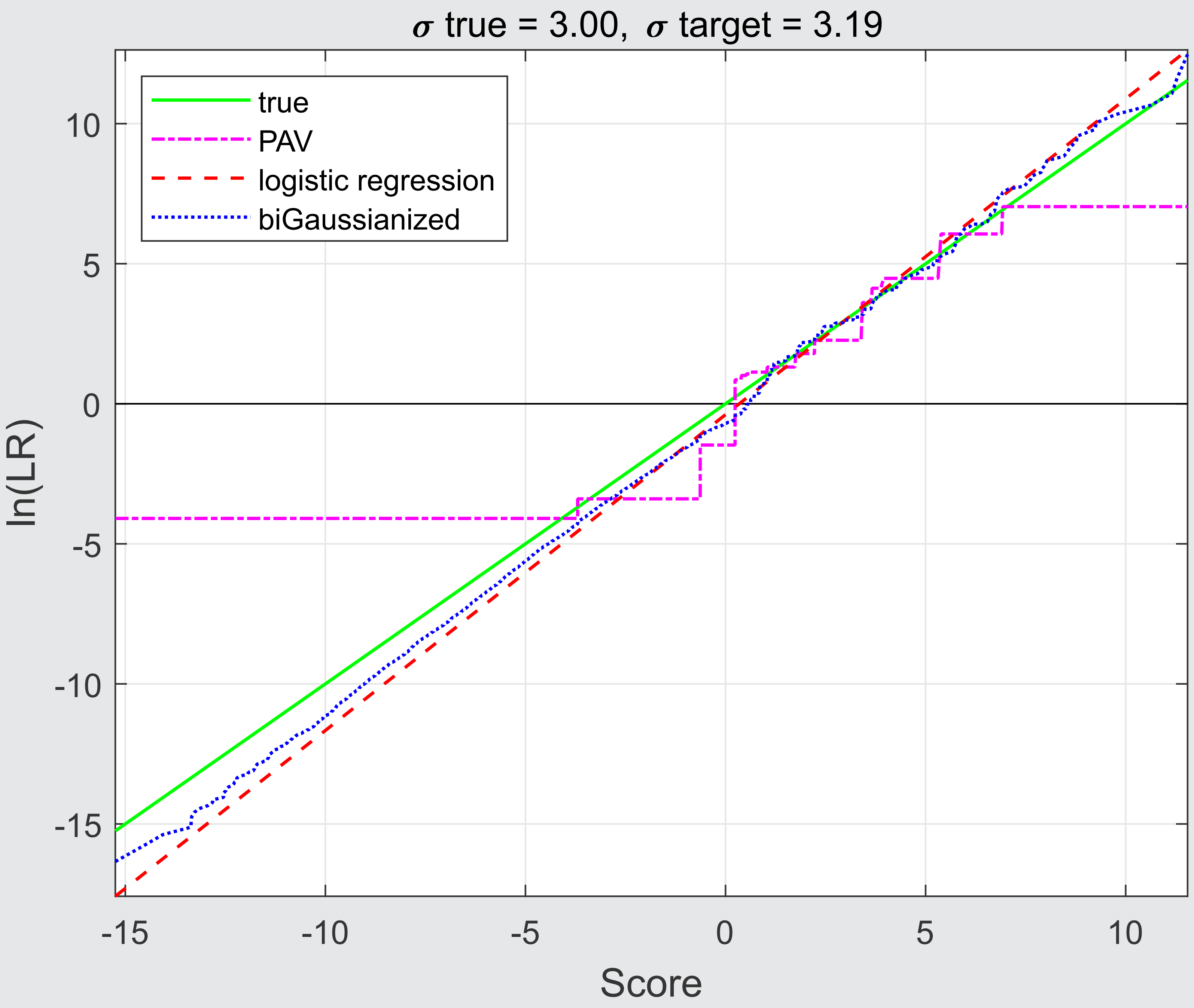
Simulated data: Gaussians with same variance

- Cumulative probability
  - with equal weight for same-source set and different-source set



Simulated data: Gaussians with same variance

- Mapping functions



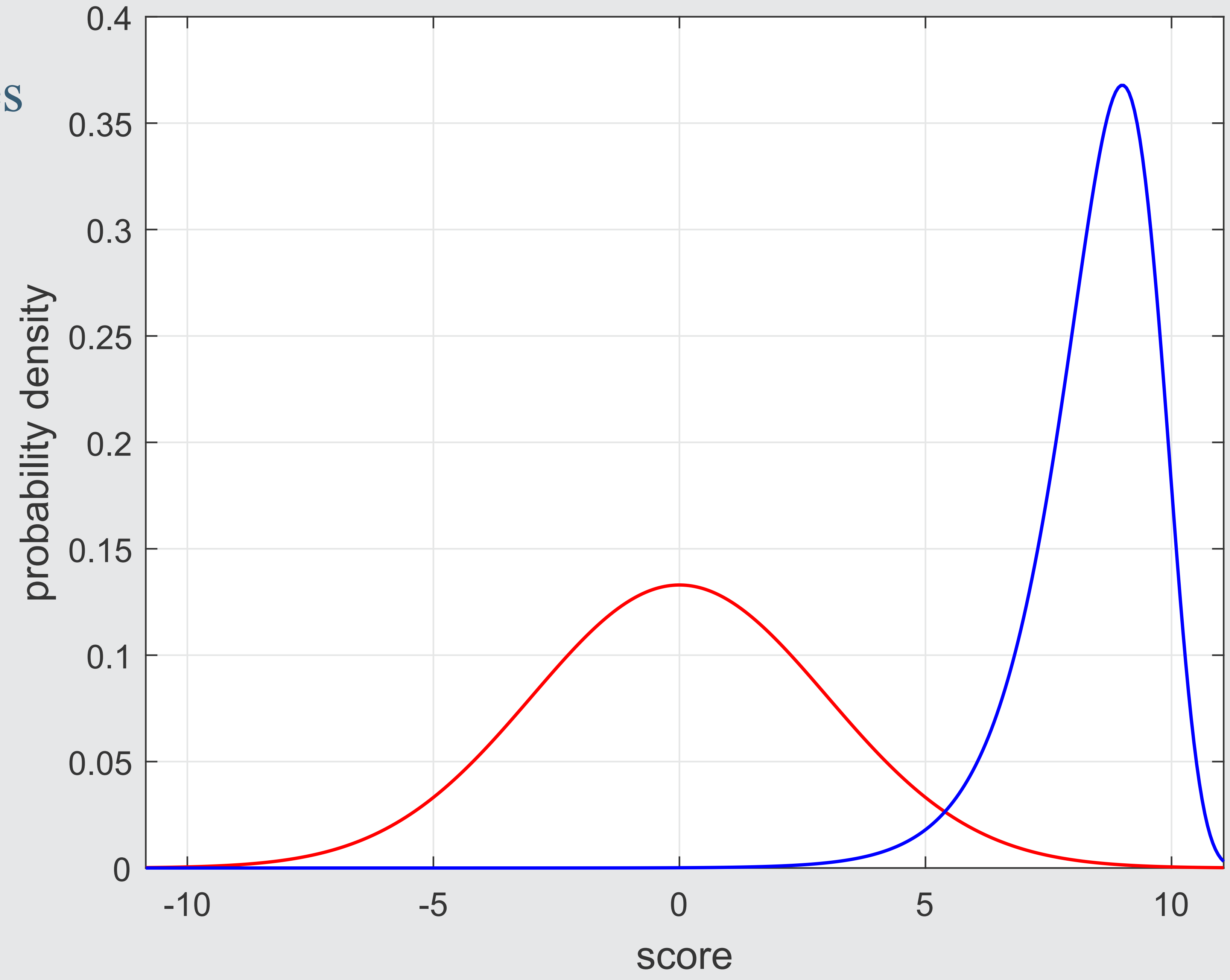
Simulated data: Gaussians with same variance

- C_{llr}

	LDF	LogReg	biGauss	PAV
train	0.21	0.20	0.20	0.18
test	0.25	0.26	0.26	0.28

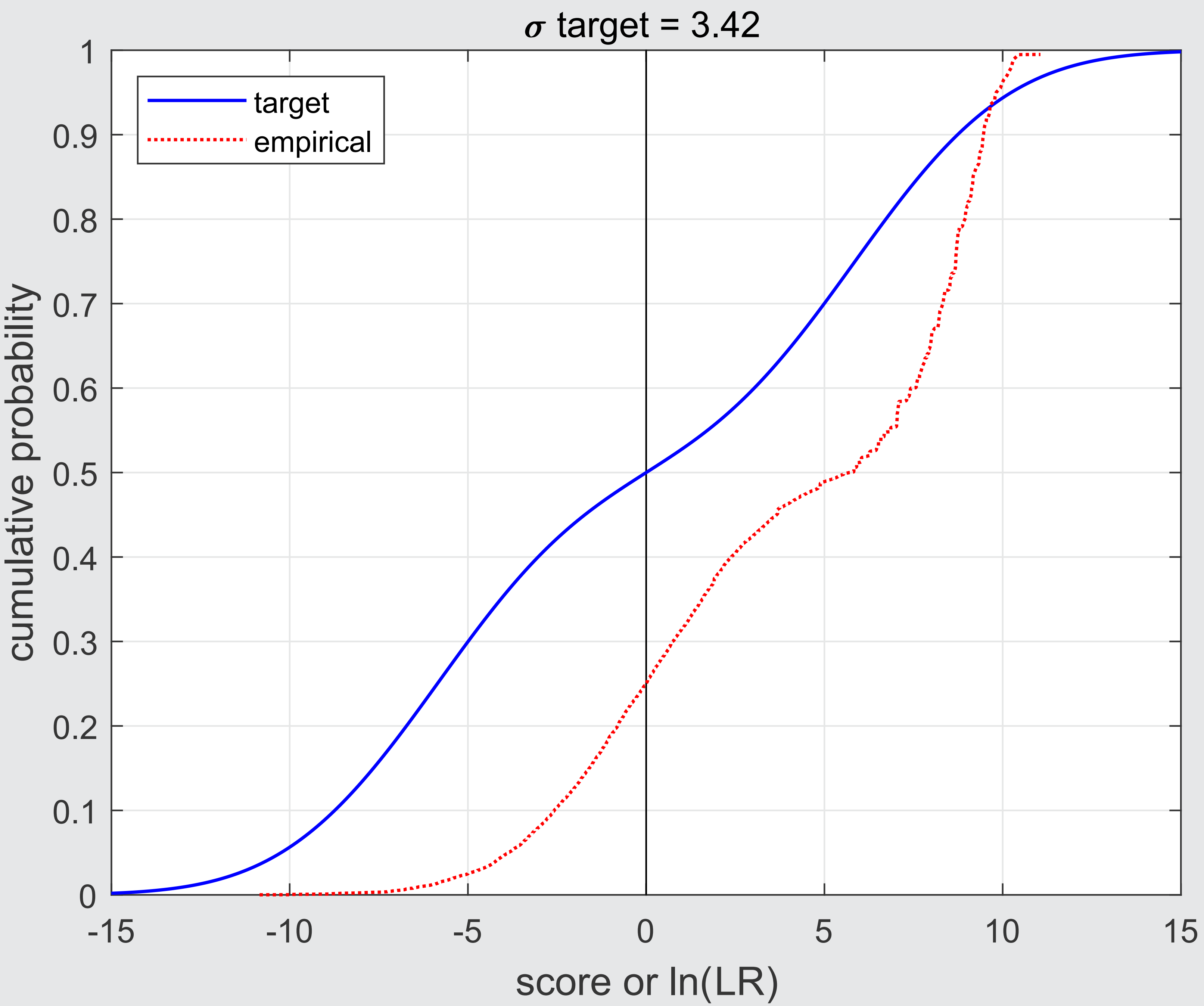
Simulated data: Gaussian & Gumbel with different variances

- 100 same-source samples
- 4950 different-source samples



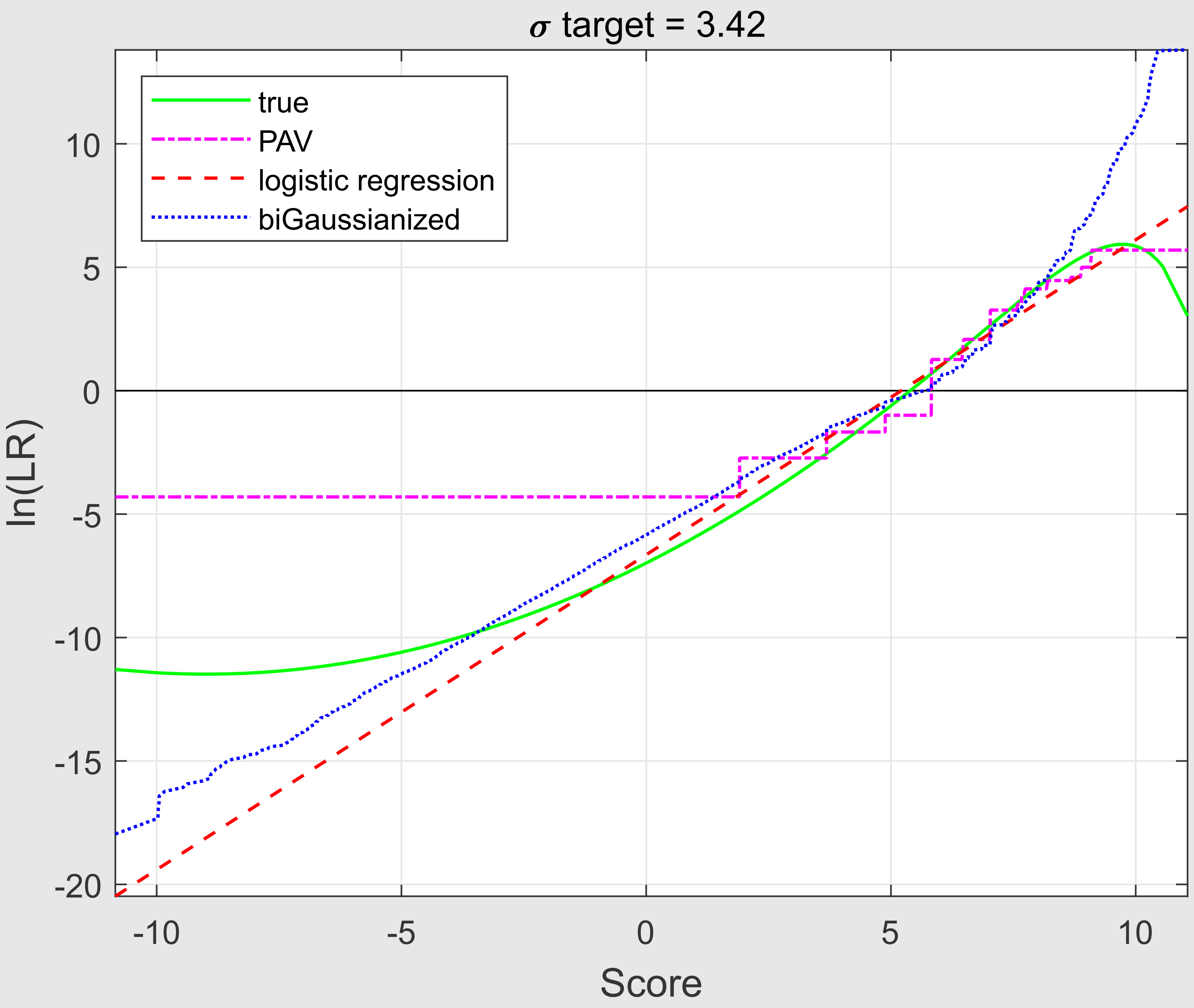
Simulated data: Gaussian & Gumbel with different variances

- Cumulative probability
  - with equal weight for same-source set and different-source set



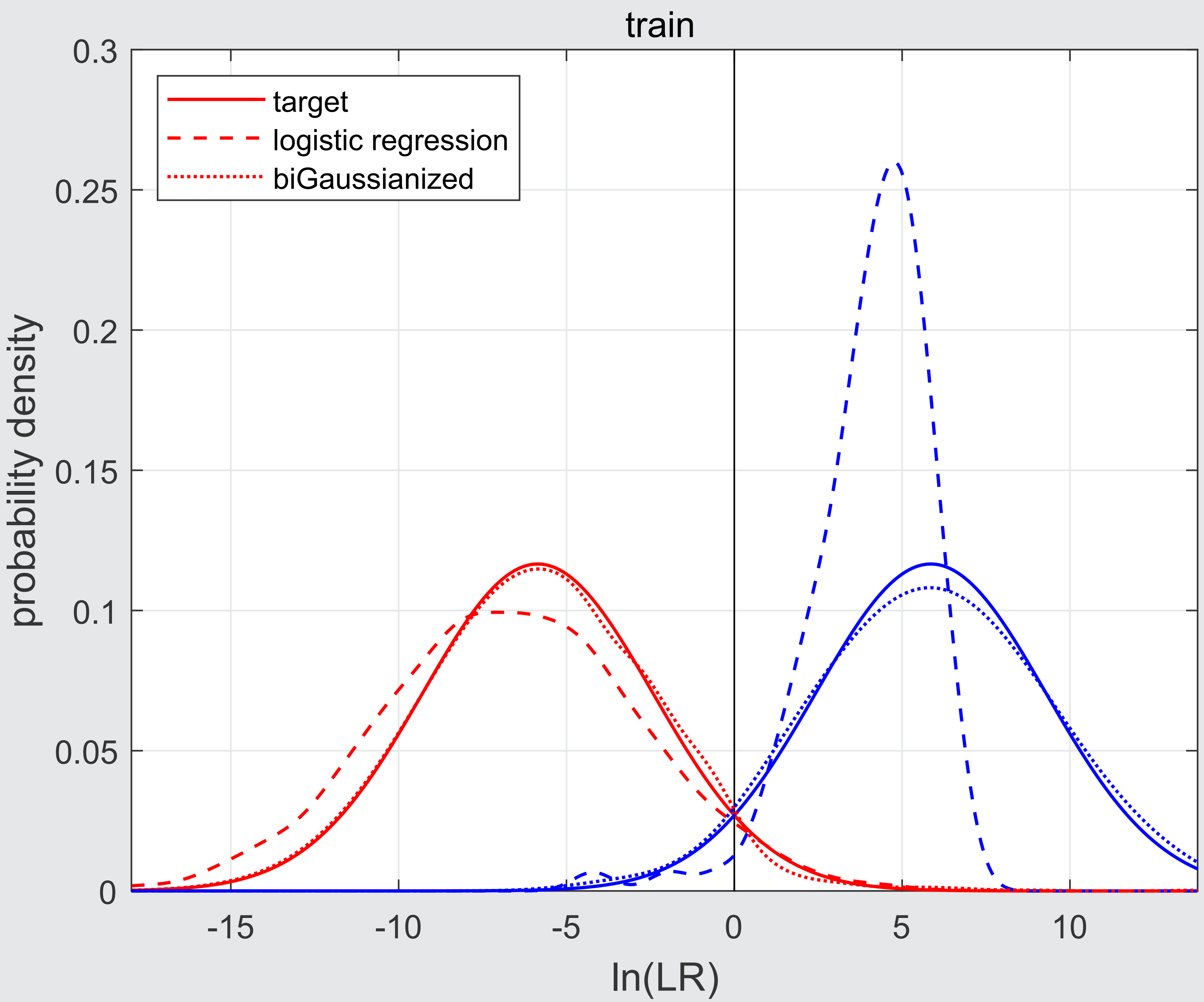
Simulated data: Gaussian & Gumbel with different variances

- Mapping functions



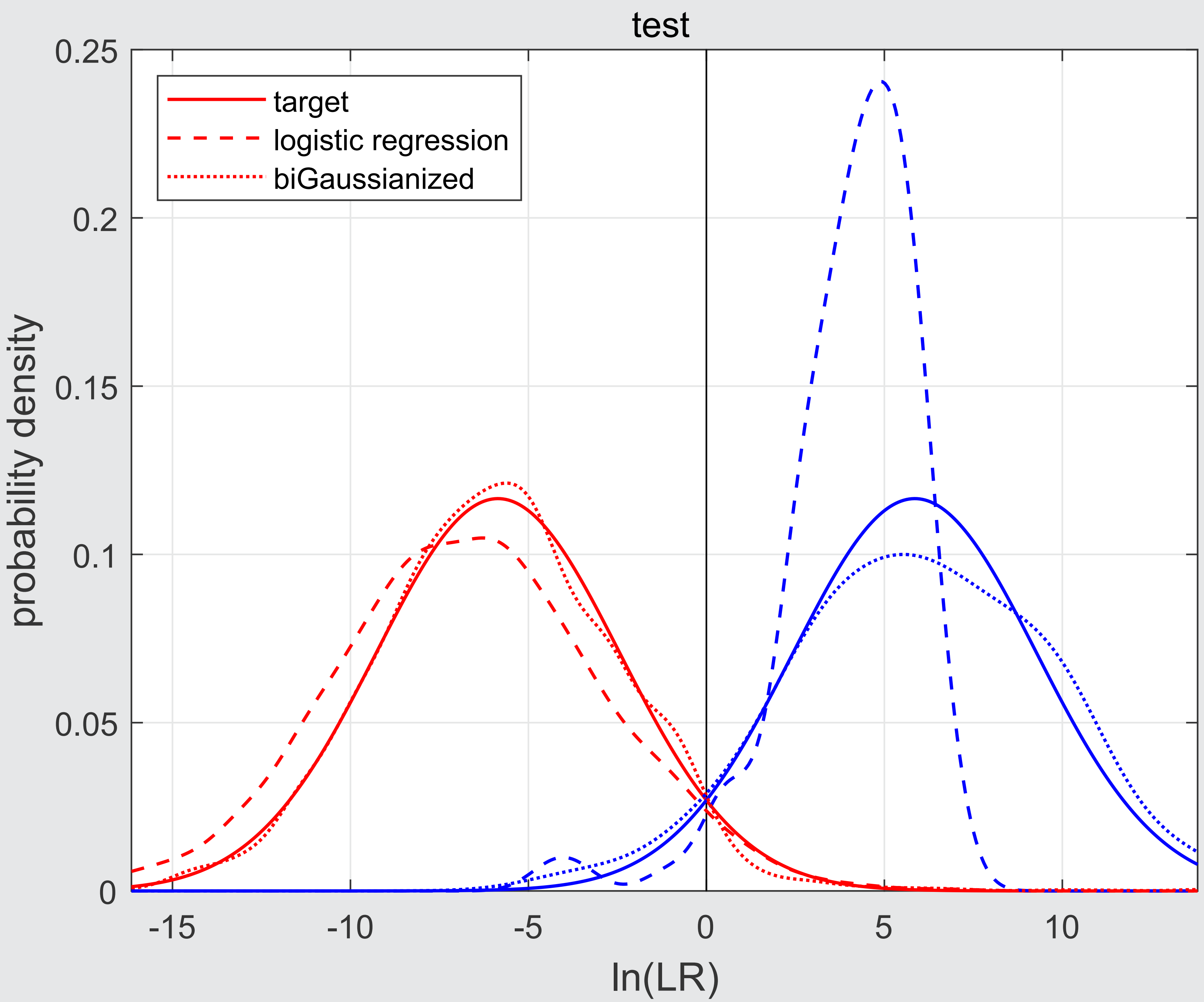
Simulated data: Gaussian & Gumbel with different variances

- Probability density functions
  - training data



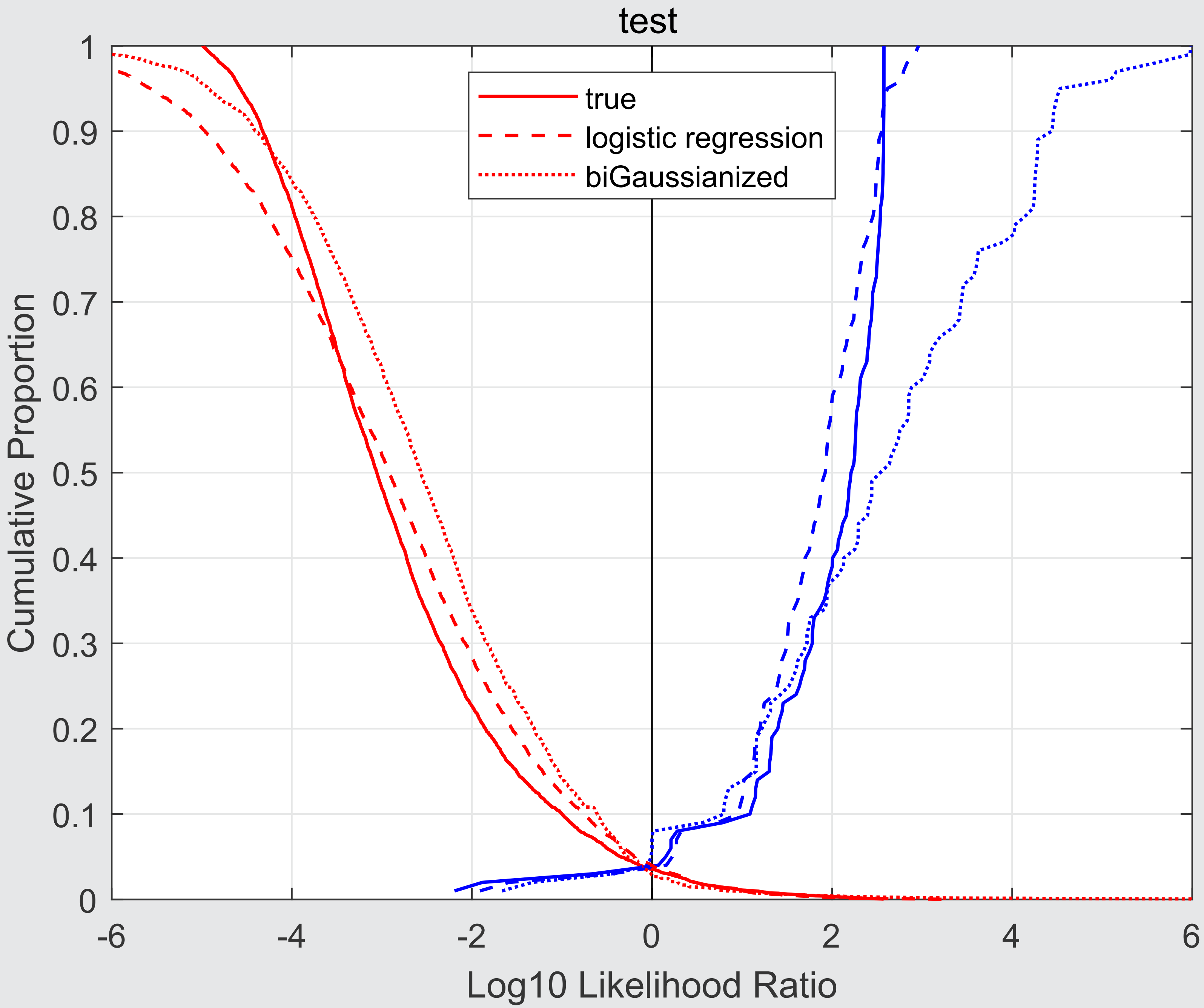
Simulated data: Gaussian & Gumbel with different variances

- Probability density functions
  - test data



Simulated data: Gaussian & Gumbel with different variances

- Tippett plots
  - test data



Simulated data: Gaussian & Gumbel with different variances

- C_{llr}

	true	LogReg	biGauss	PAV
train	0.16	0.16	0.16	0.14
test	0.18	0.18	0.18	0.20

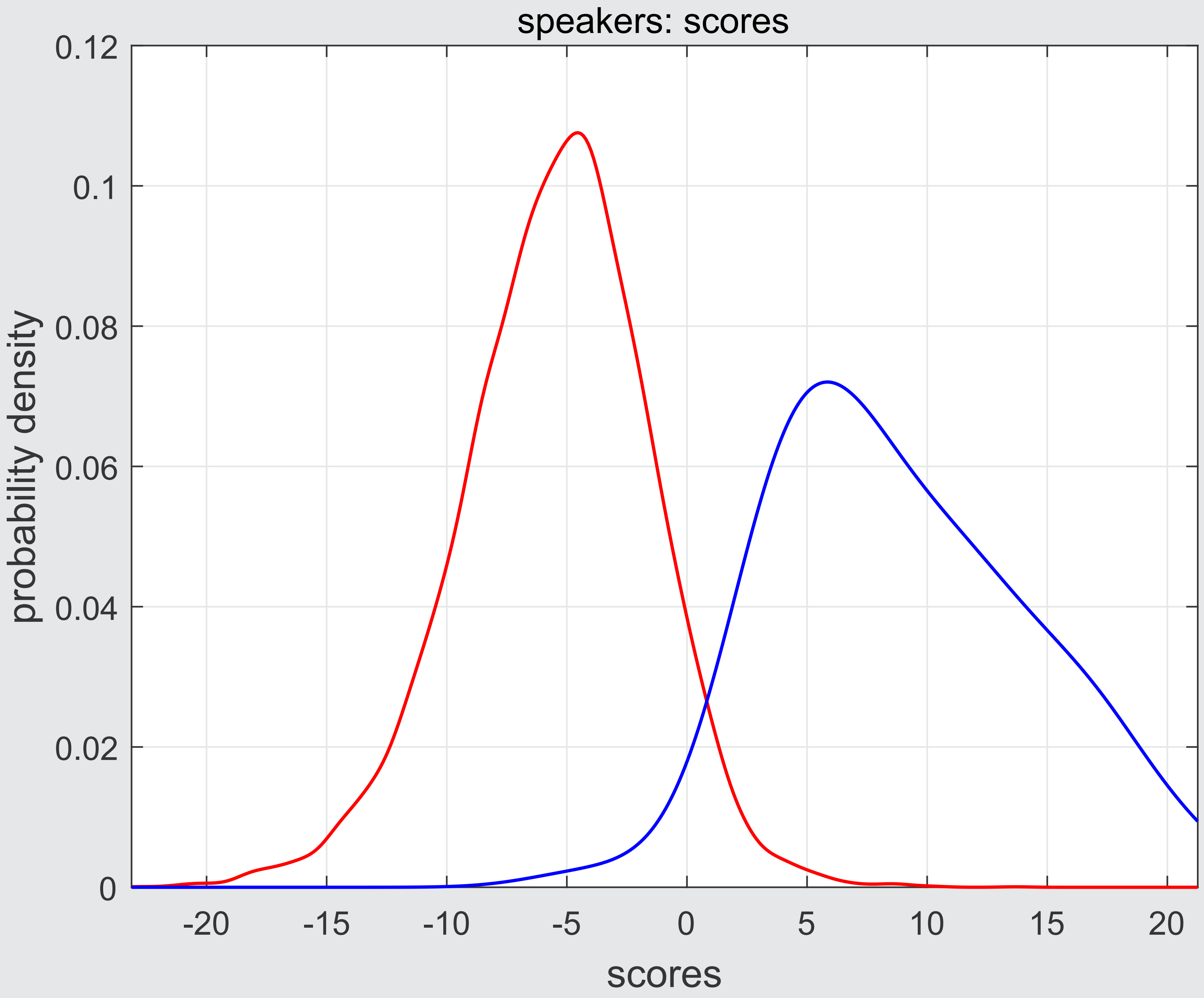
Real data: forensic voice comparison

- *forensic_eval_01*

  benchmark data

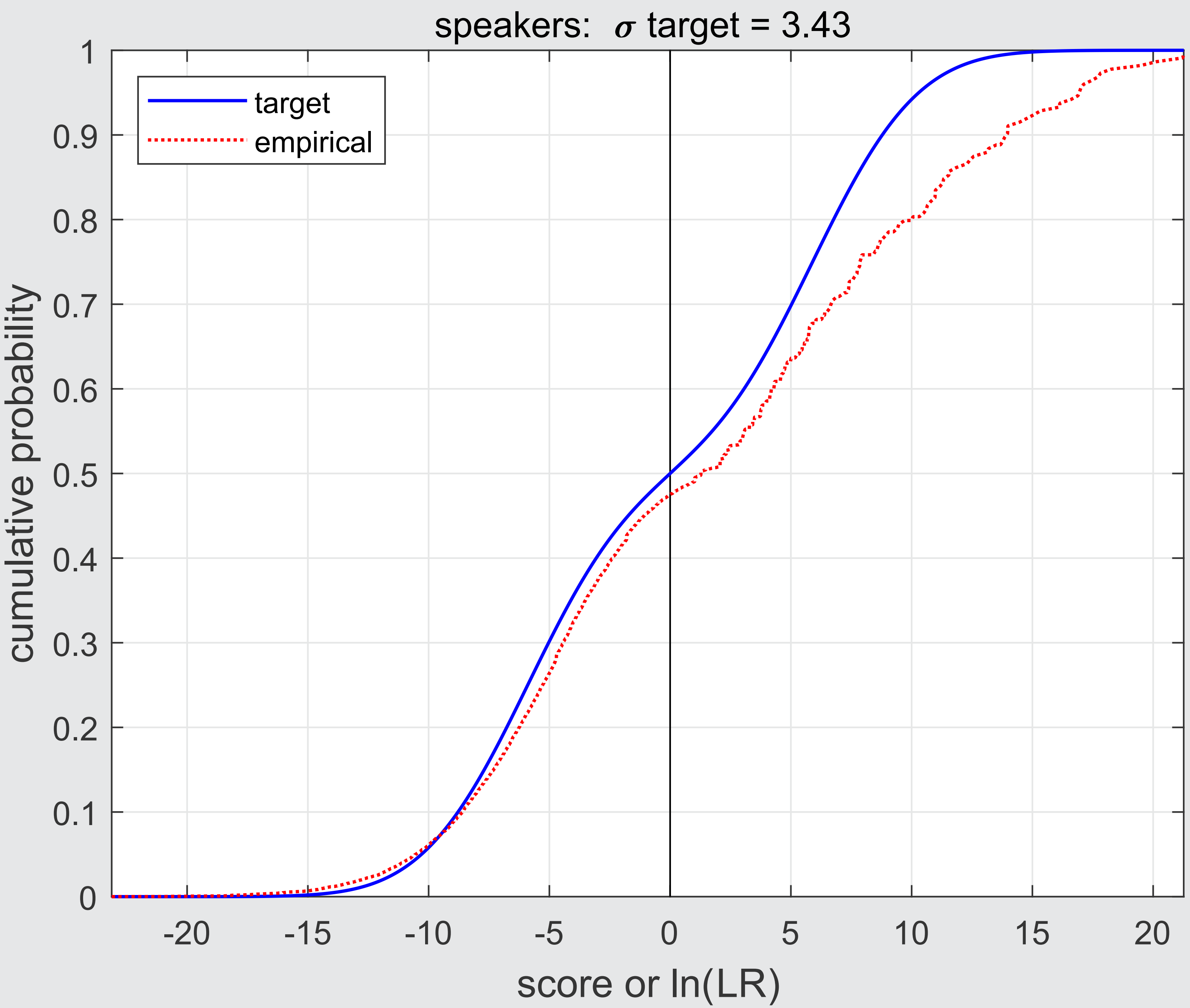
- E^3 Forensic Speech Science System (E^3FS^3)

Weber P., Enzinger E., Labrador B., Lozano-Díez A., Ramos D., González-Rodríguez J., Morrison G.S. (2022). Validation of the alpha version of the E^3 Forensic Speech Science System (E^3FS^3) core software tools. *Forensic Science International: Synergy*, 4, 100223. https://doi.org/10.1016/j.fsisyn.2022.100223



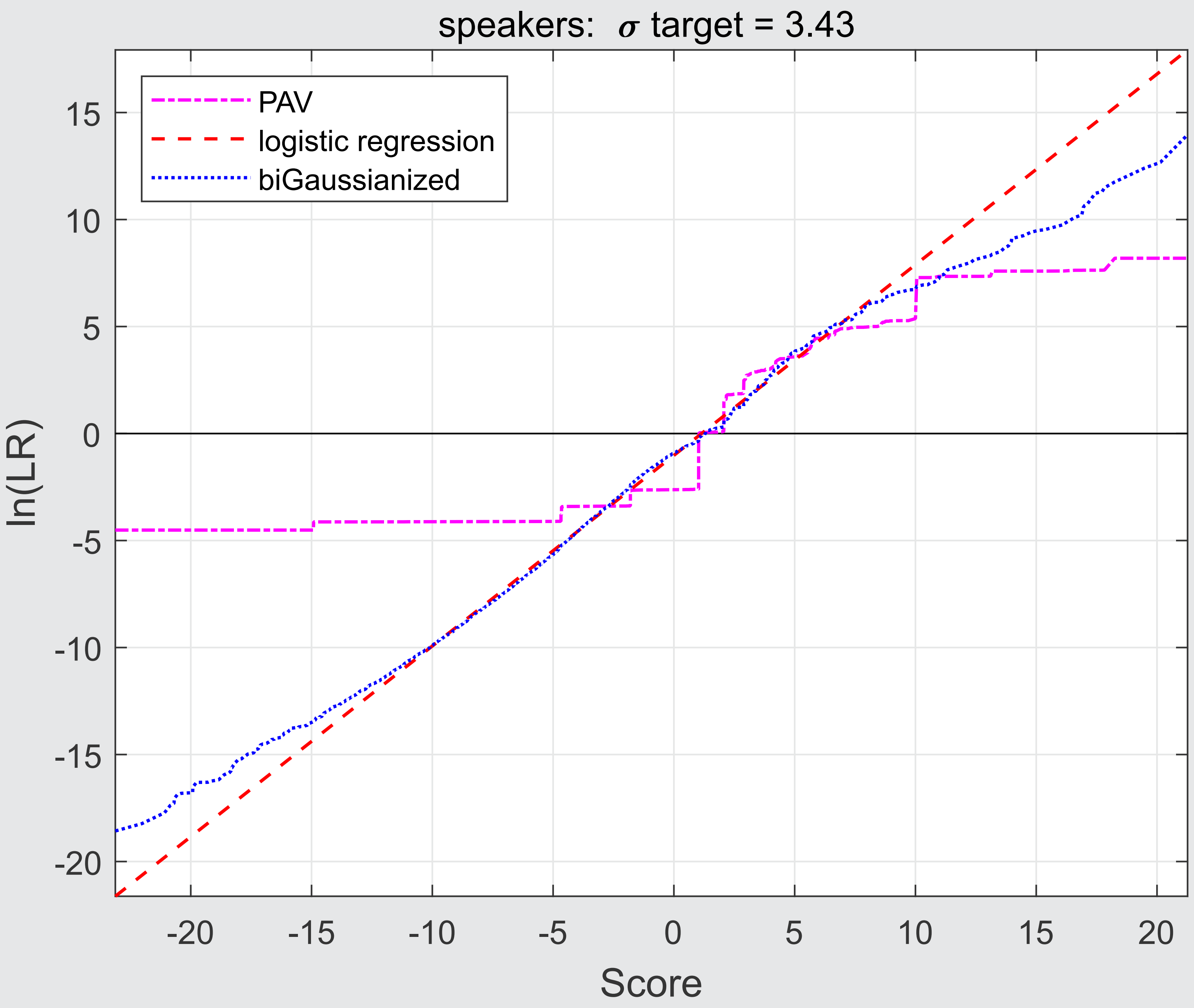
Real data: forensic voice comparison

- Cumulative probability
    - with equal weight for same-source set and different-source set



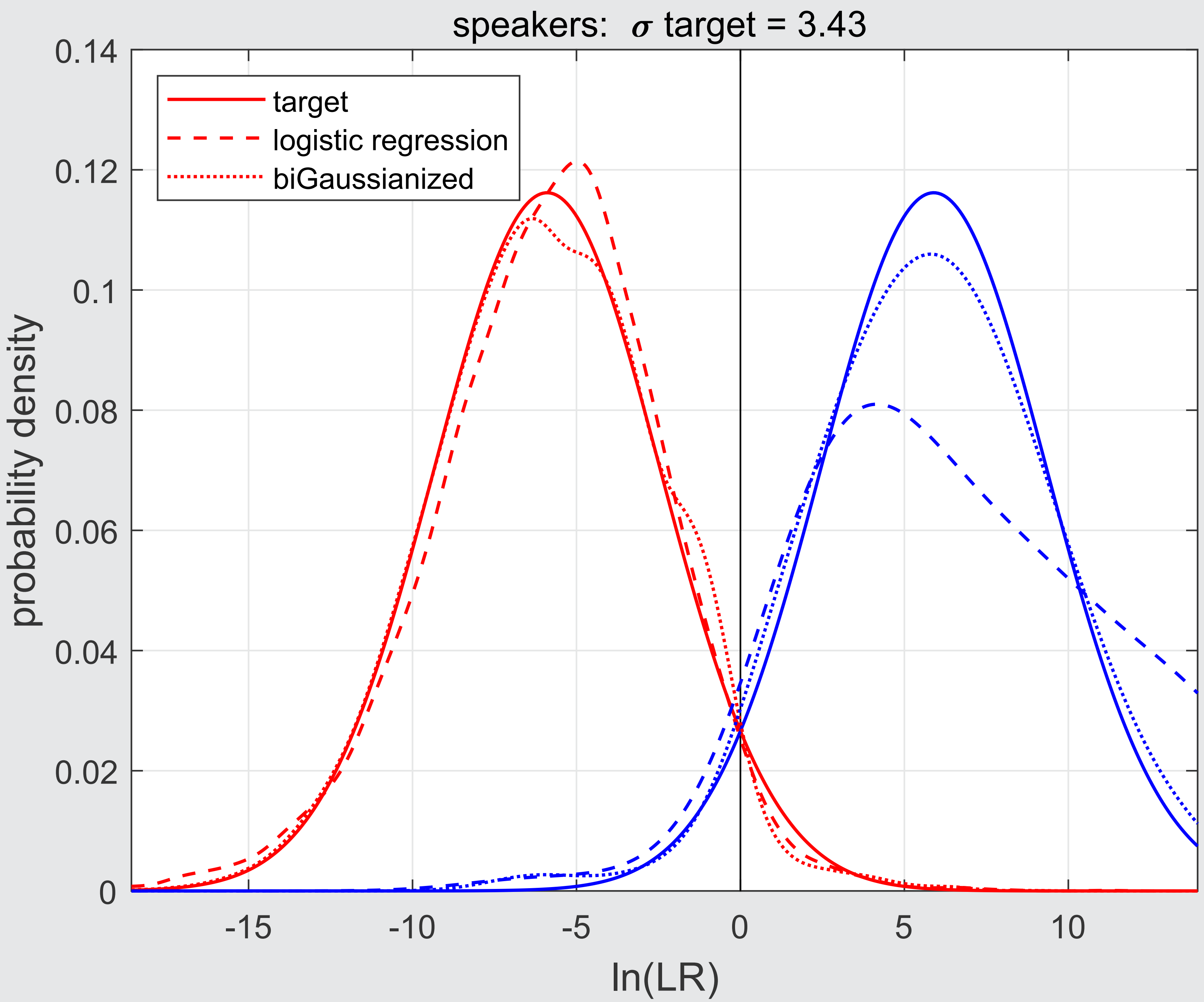
Real data: forensic voice comparison

- Mapping functions
    - cross-validated



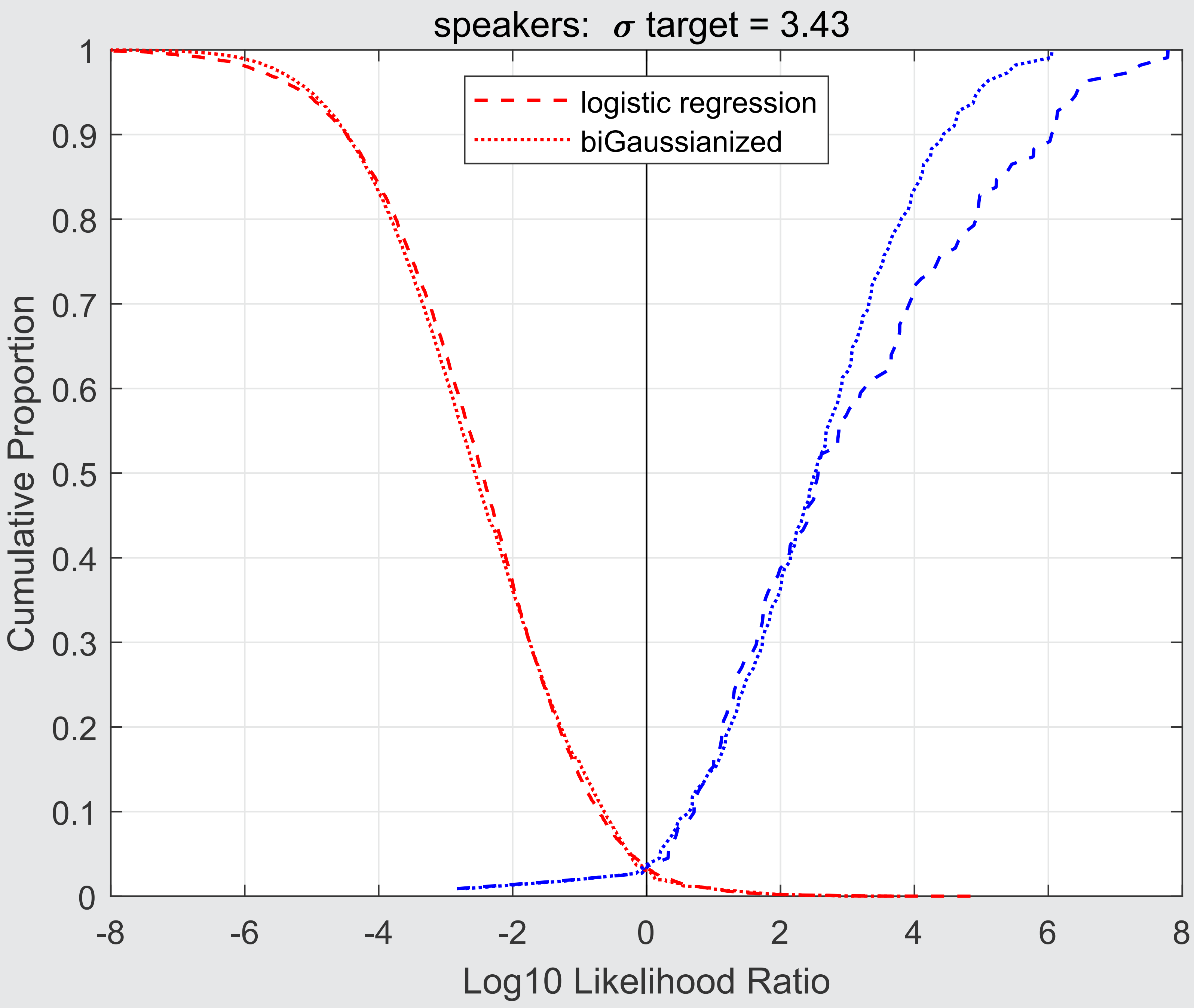
Real data: forensic voice comparison

- Probability density functions
  - cross-validated



Real data: forensic voice comparison

- Tippett plots
    - cross-validated



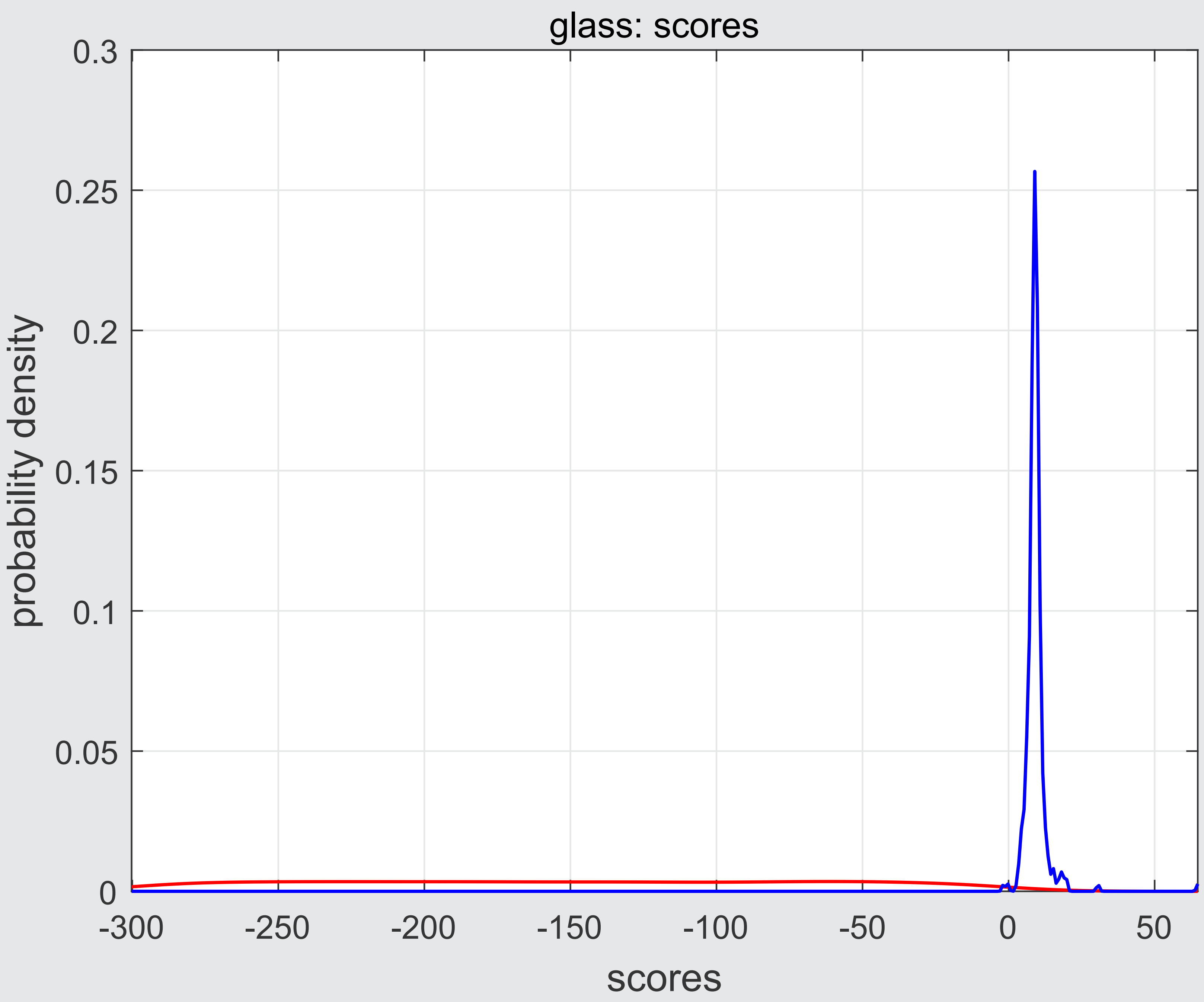
Real data: forensic voice comparison

- C_{llr}

	LogReg	biGauss	PAV
cross-val	0.17	0.17	0.17

Real data: glass

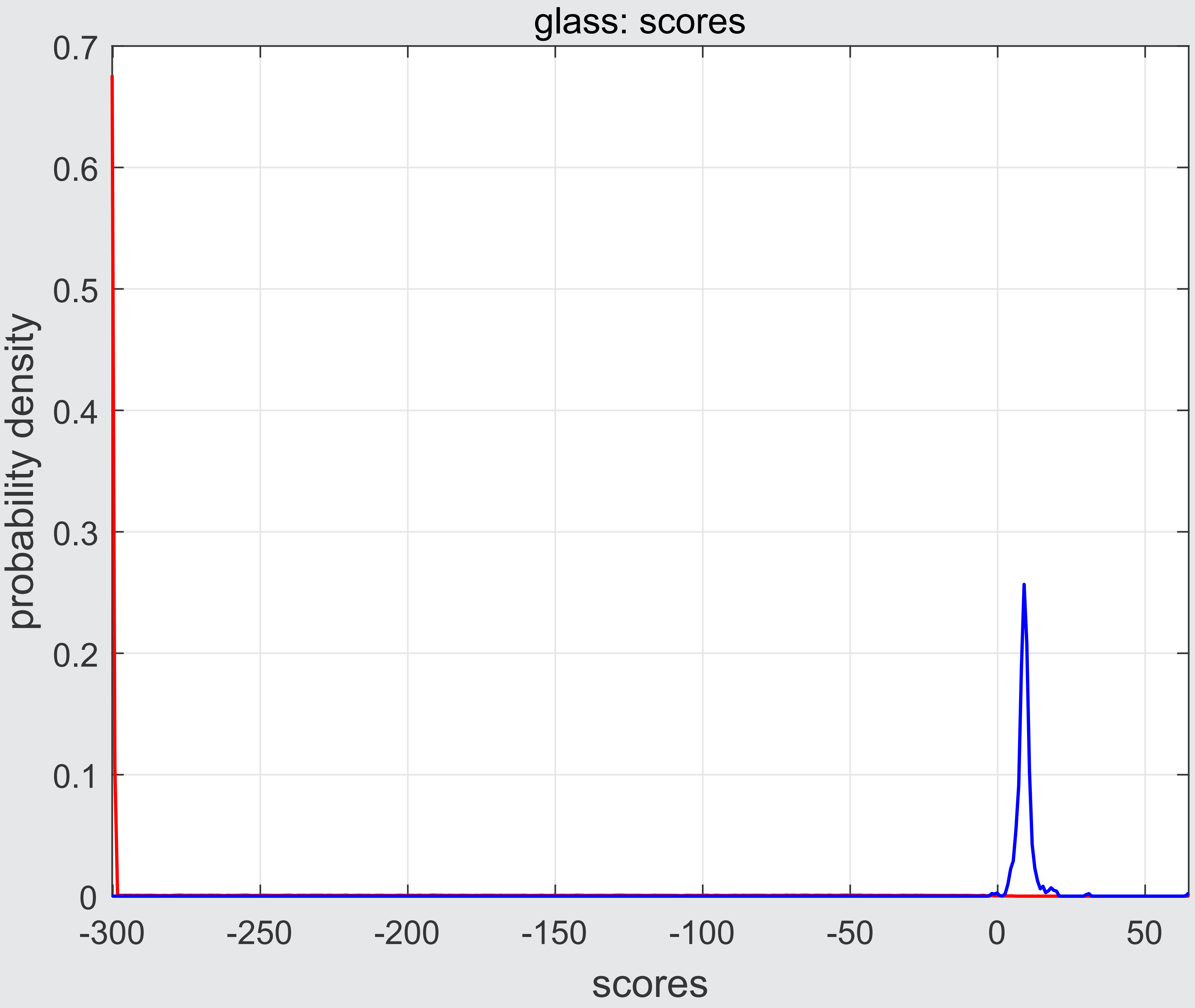
- glass scores from van Es et al. (2017)



van Es A., Wiarda W., Hordijk M., Alberink I., Vergeer P. (2017). Implementation and assessment of a likelihood ratio approach for the evaluation of LA-ICP-MS evidence in forensic glass analysis. *Science & Justice*, 57, 181–192. https://doi.org/10.1016/j.scijus.2017.03.002

Real data: glass

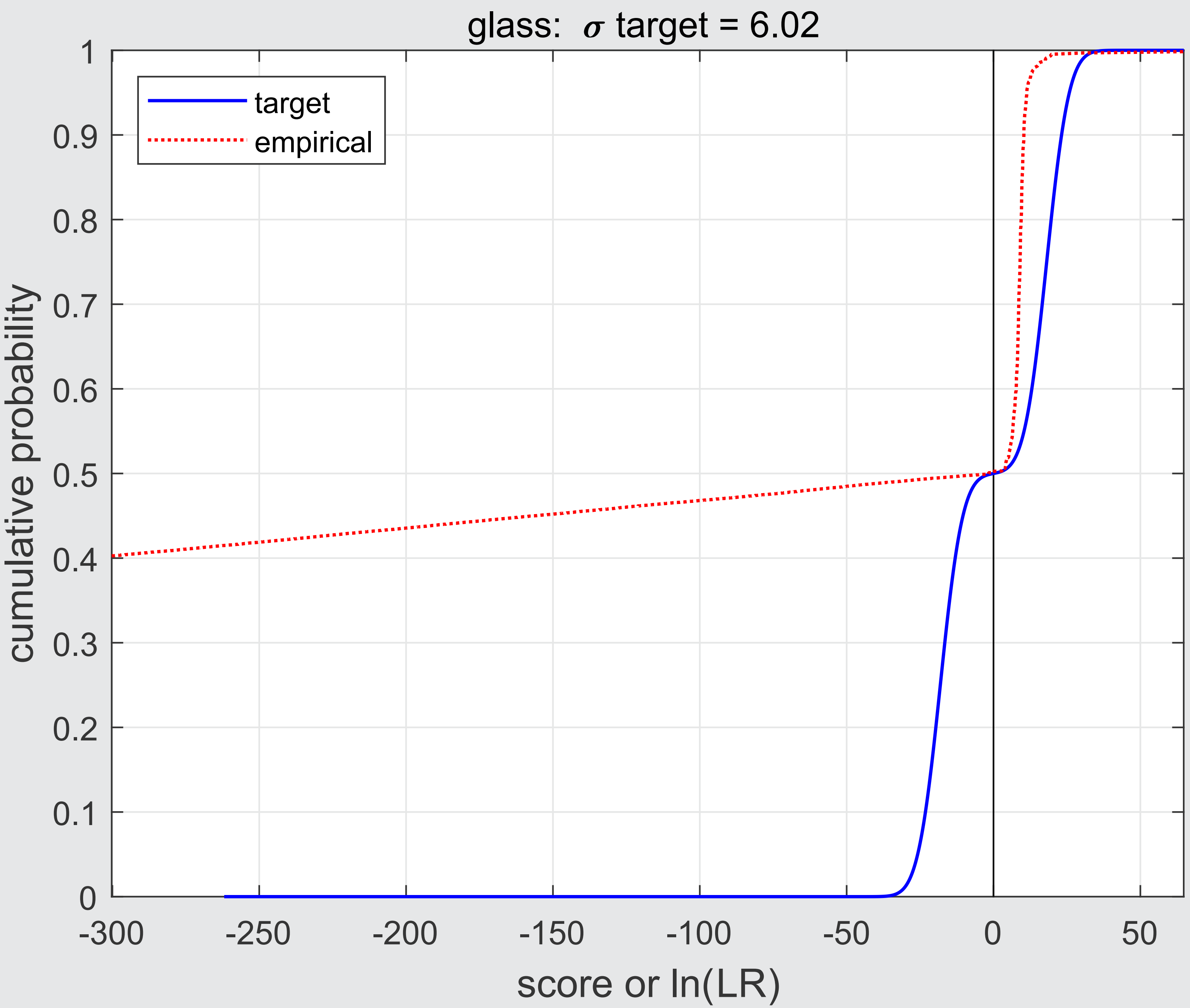
- glass scores from van Es et al. (2017)



van Es A., Wiarda W., Hordijk M., Alberink I., Vergeer P. (2017). Implementation and assessment of a likelihood ratio approach for the evaluation of LA-ICP-MS evidence in forensic glass analysis. *Science & Justice*, 57, 181–192. https://doi.org/10.1016/j.scijus.2017.03.002

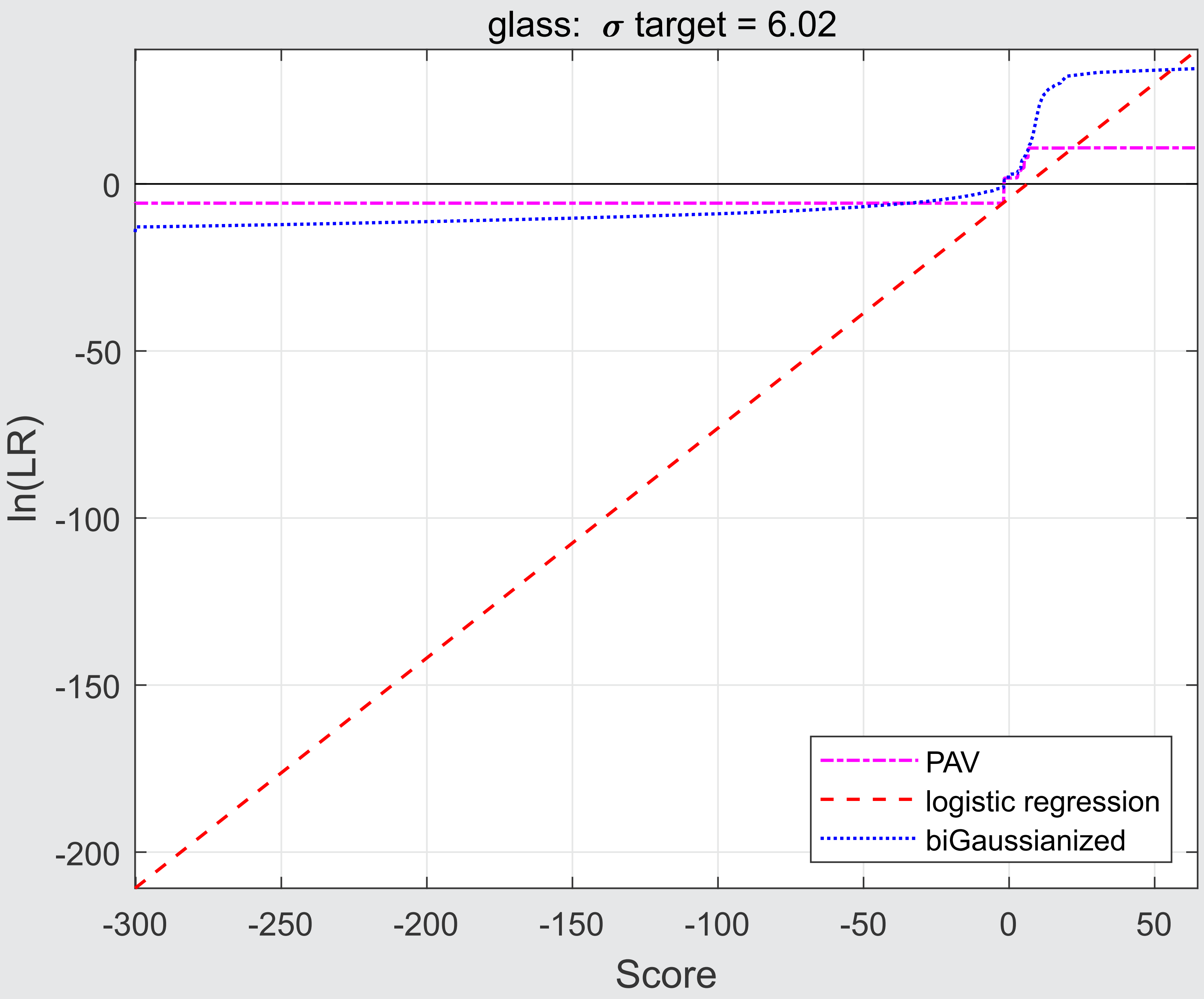
Real data: glass

- Cumulative probability
    - with equal weight for same-source set and different-source set



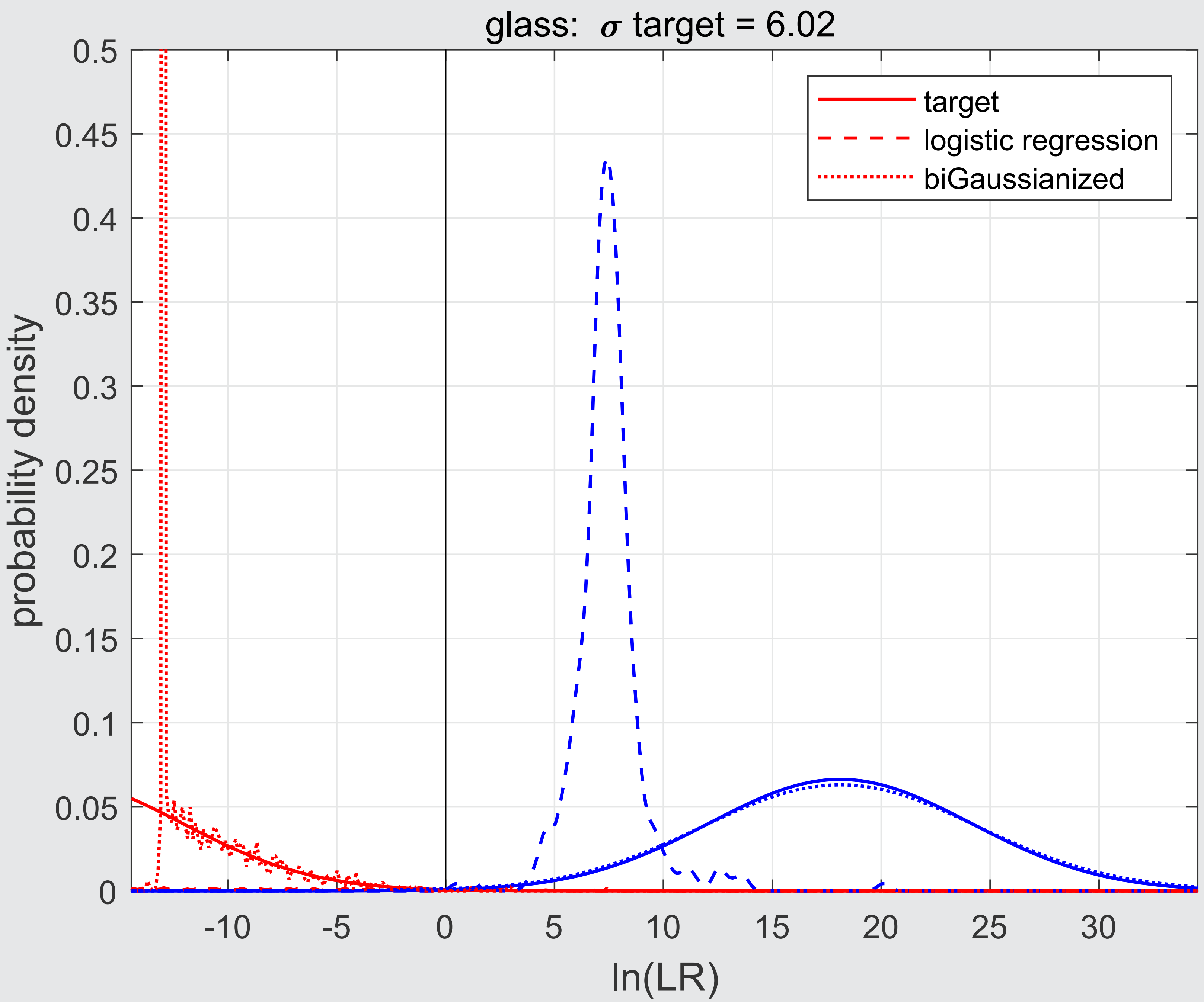
Real data: glass

- Mapping functions
    - cross-validated



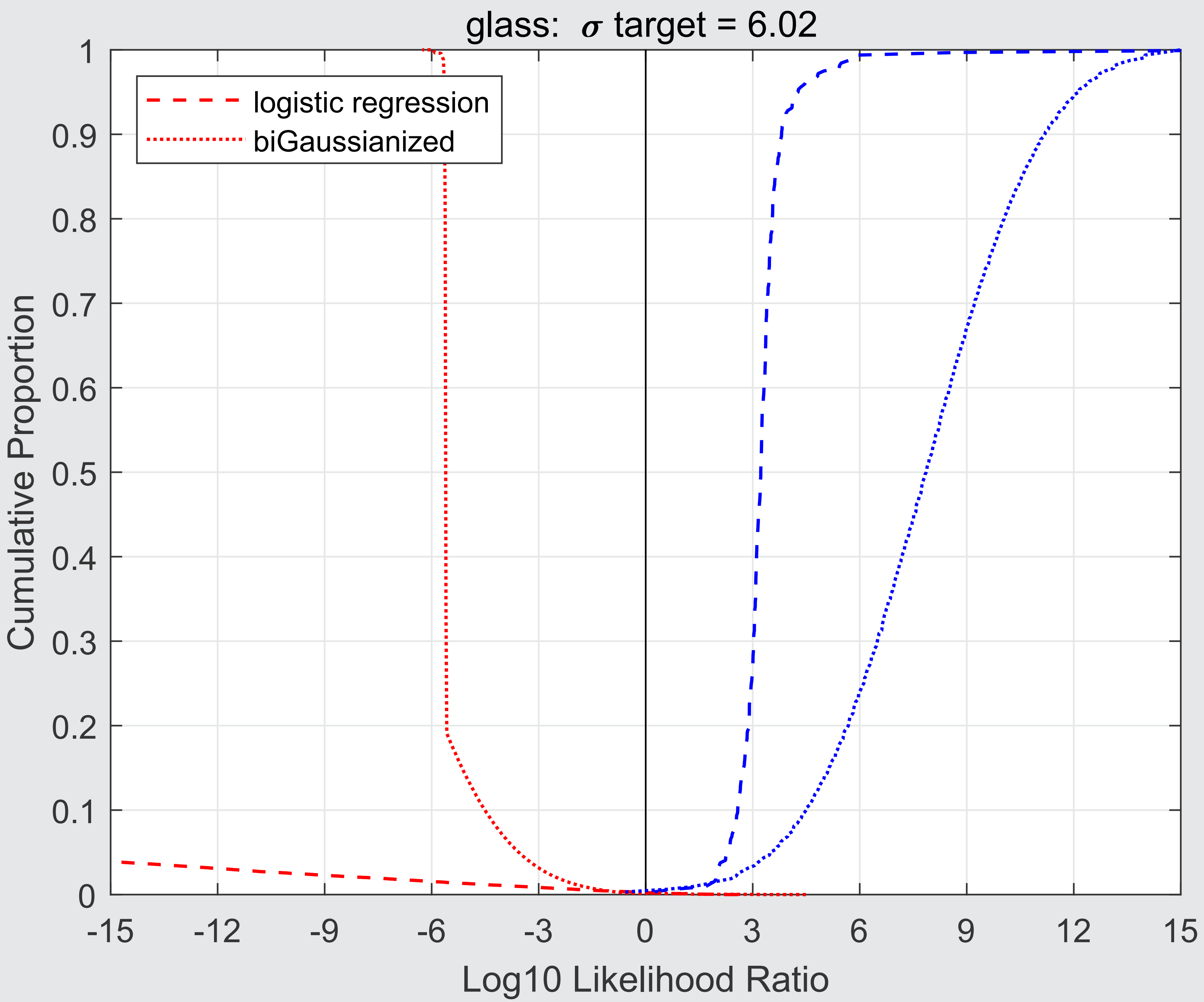
Real data: glass

- Probability density functions
  - cross-validated



Real data: glass

- Tippett plots
    - cross-validated

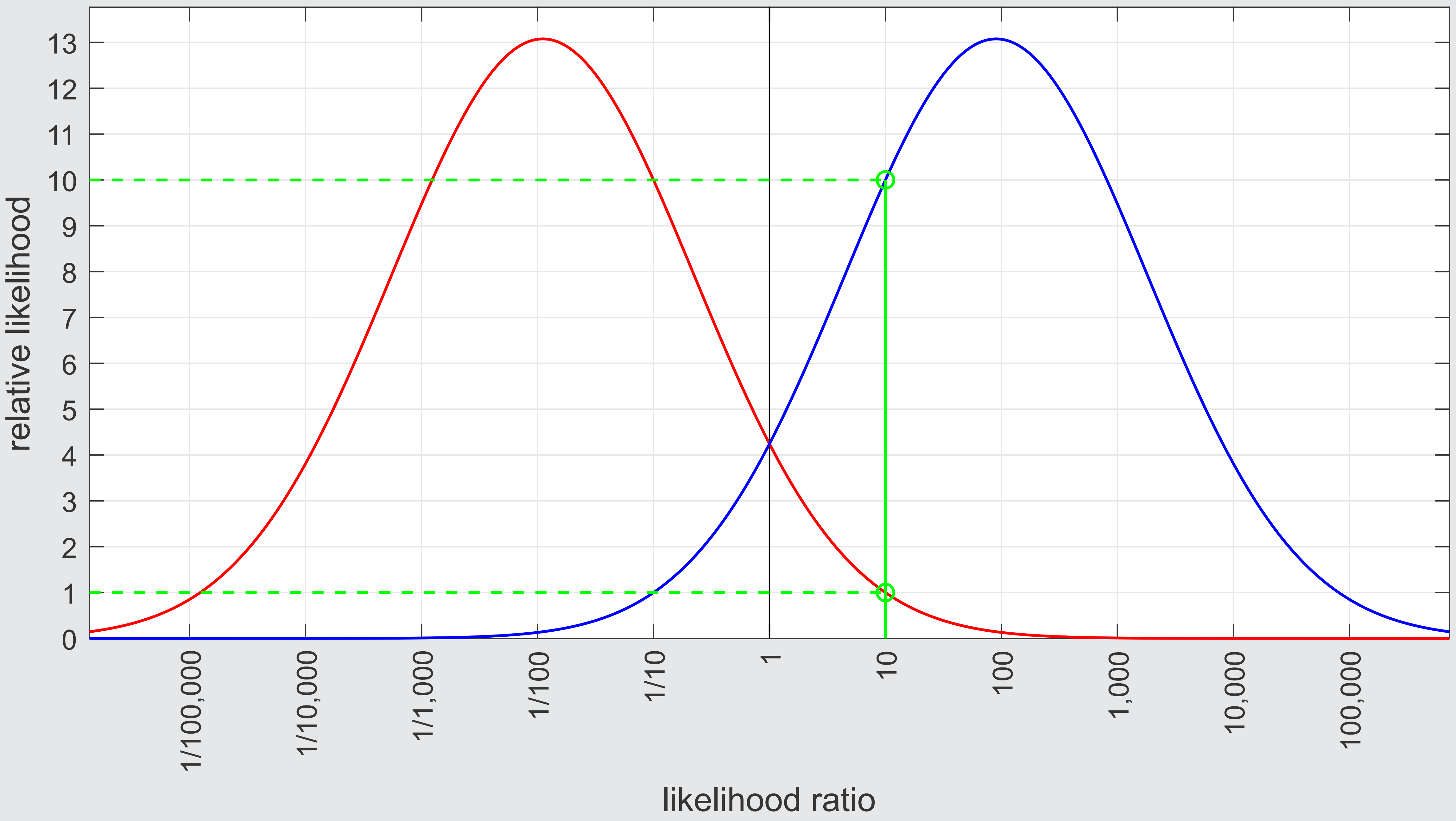


Real data: glass

- C_{llr}

	LogReg	biGauss	PAV
cross-val	0.007	0.009	0.020

Use in presentation of a likelihood-ratio value



Thank You