Bi-Gaussianized Calibration of Likelihood Ratios

Geoffrey Stewart Morrison

Forensic Data Science Laboratory, Aston University

Forensic Evaluation Ltd

Aston University BIRMINGHAM UK

$$\frac{p(E|H_p)}{p(E|H_d)}$$

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.

Paper

- Morrison G.S. (2024). **Bi-Gaussianized calibration of likelihood ratios**. *Law, Probability & Risk*, 23, mgae004. https://doi.org/10.1093/lpr/mgae004

What is Calibration?

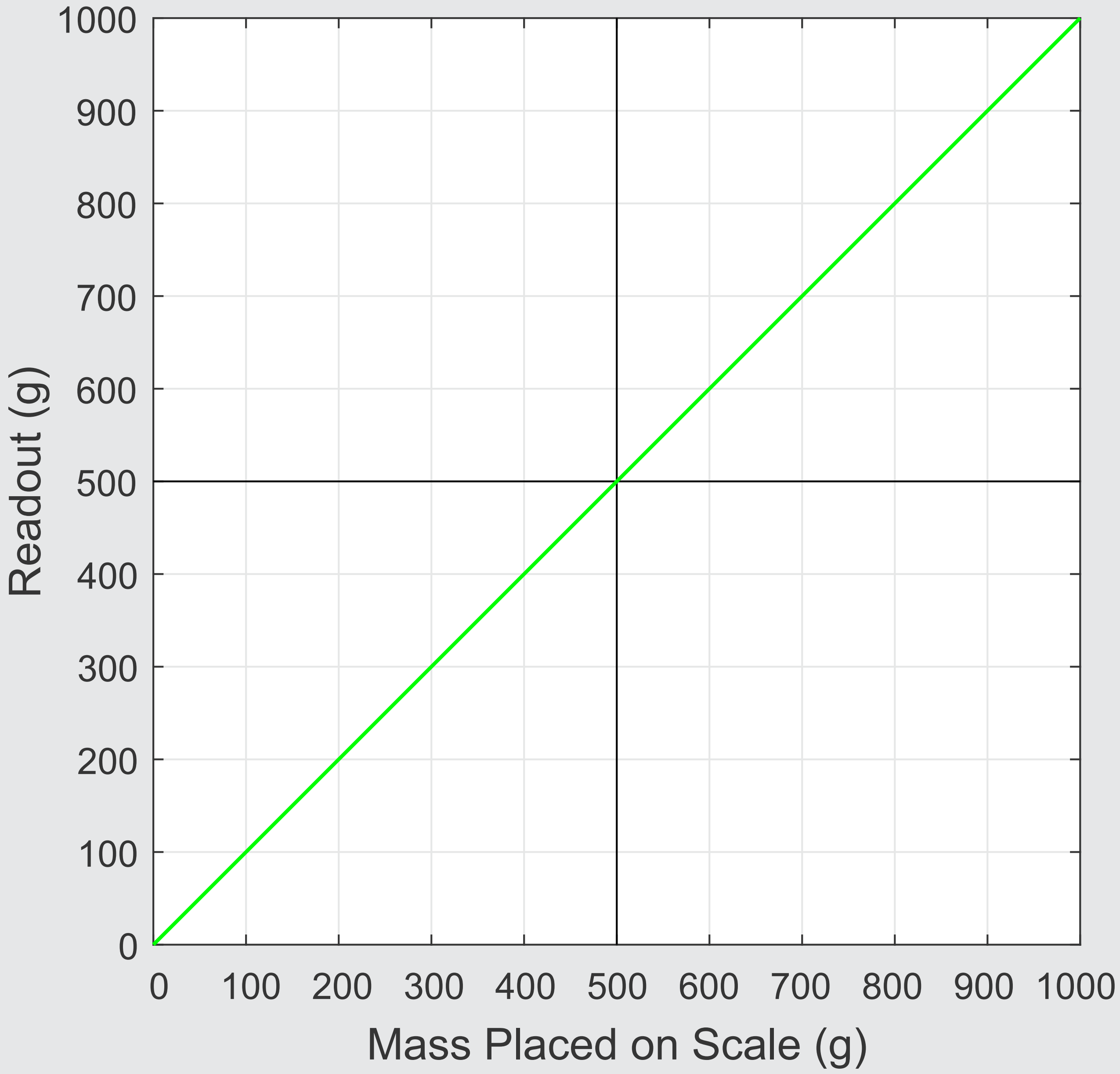
- What is a well-calibrated set of scales?

- 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?

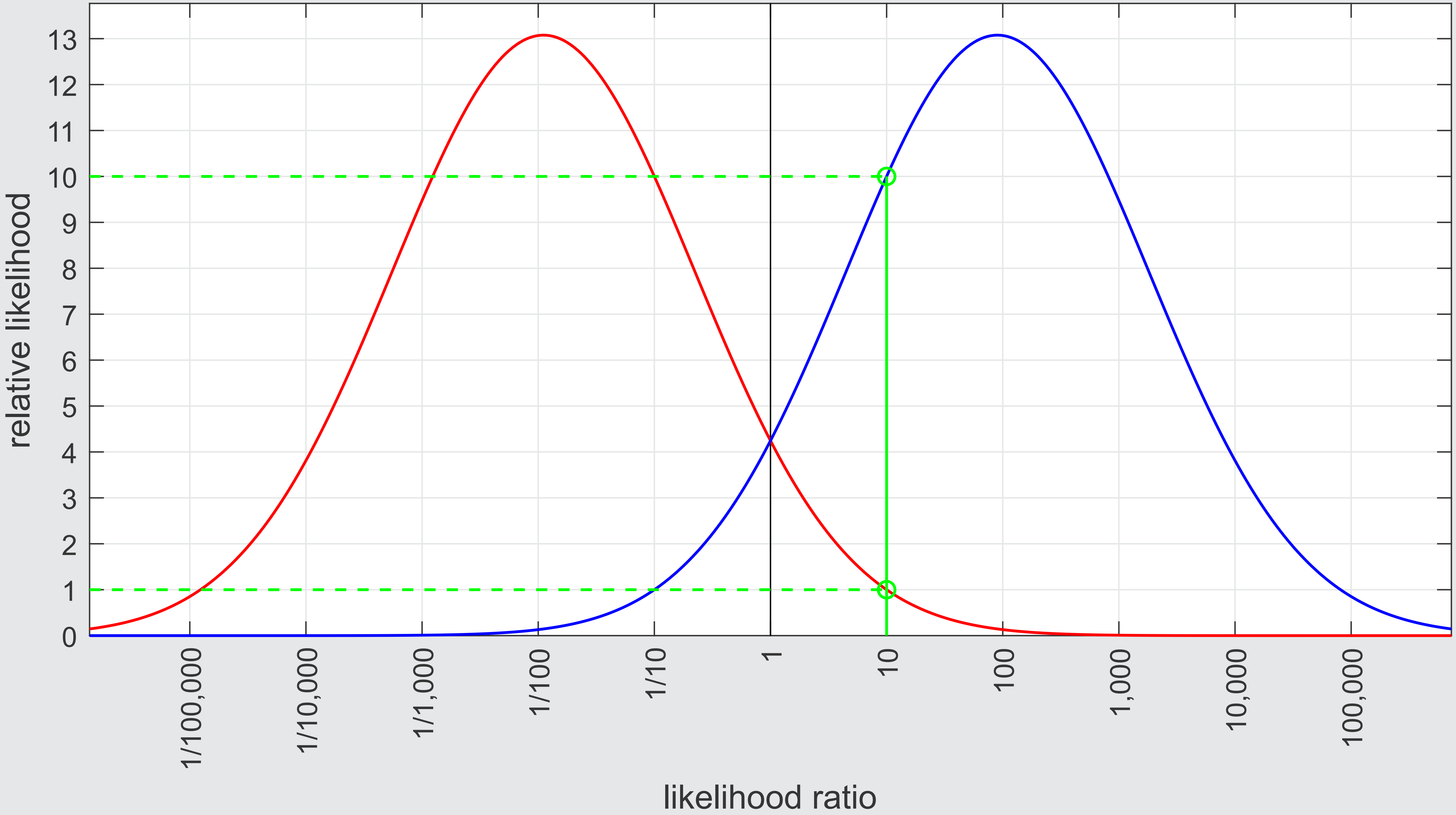
- What is a well-calibrated likelihood-ratio system?

- A system for which:

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

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

Perfectly calibrated likelihood ratios

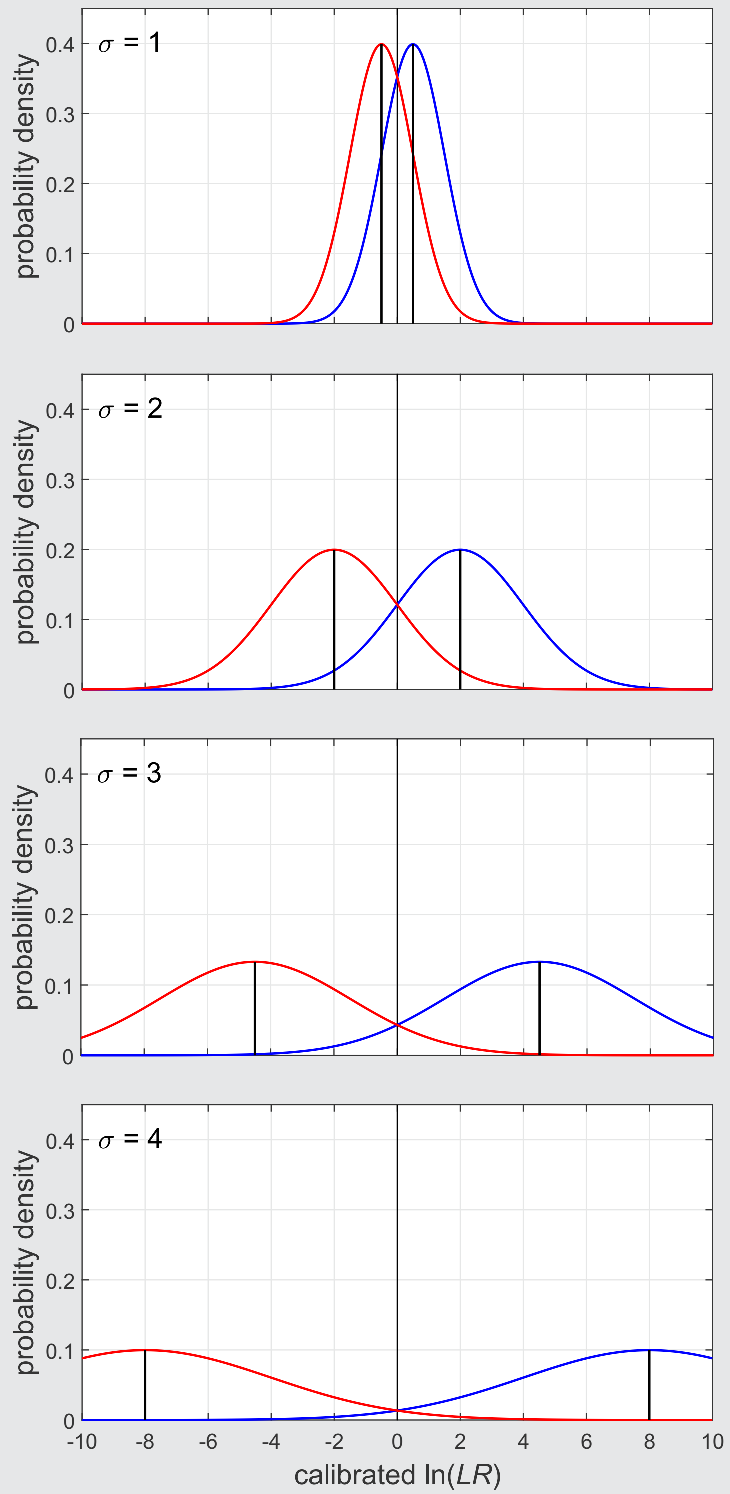


Perfectly calibrated likelihood ratios

- Perfectly calibrated ln(LR) distributions

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

$$\mu_d = -\frac{\sigma^2}{2} \qquad \mu_s = +\frac{\sigma^2}{2}$$



Perfectly calibrated likelihood ratios

- Perfectly calibrated ln(*LR*) distributions

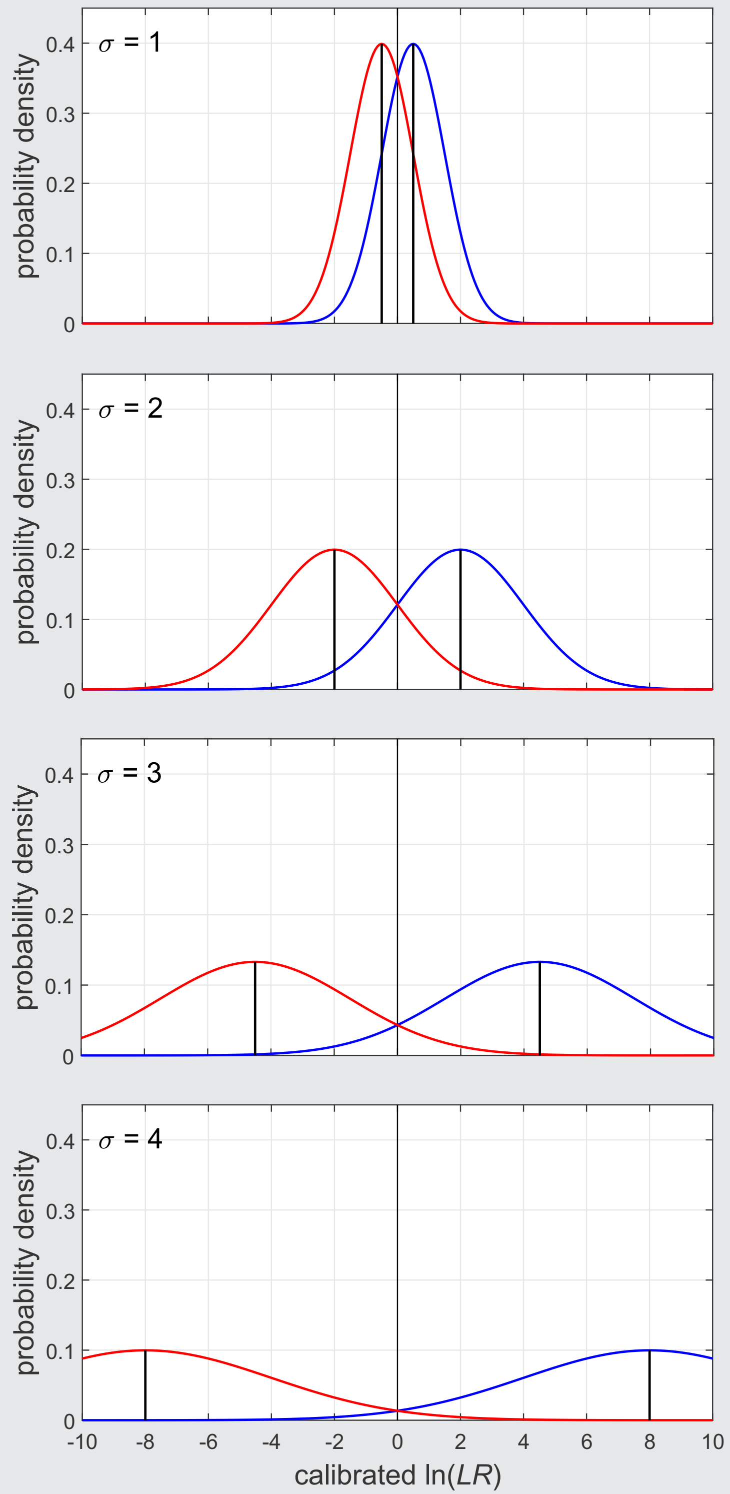
    - C_{llr} values

0.84

0.51

0.24

0.09



Perfectly calibrated likelihood ratios

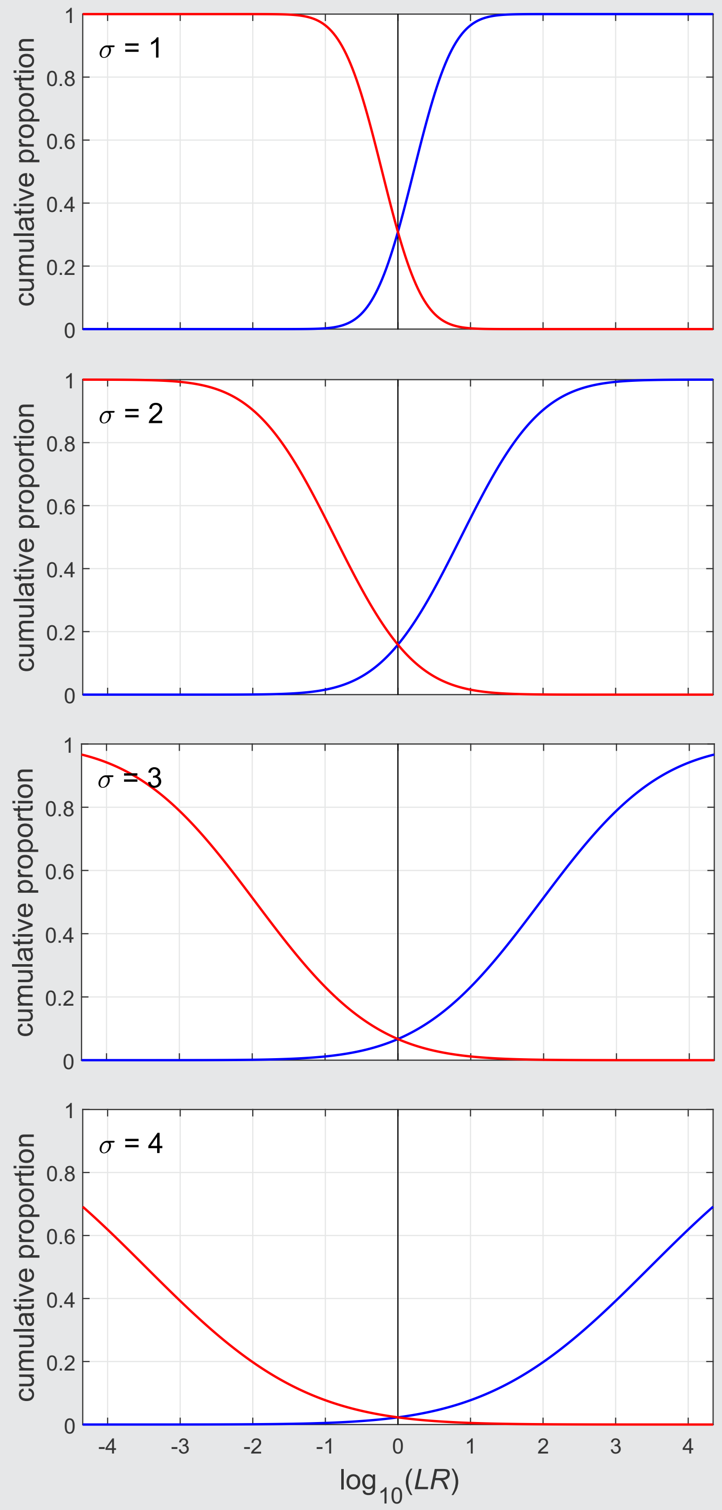
- Perfectly calibrated $\ln(LR)$ distributions
    - C_{llr} values
    - Tippett plots

0.84

0.51

0.24

0.09



Linear calibration

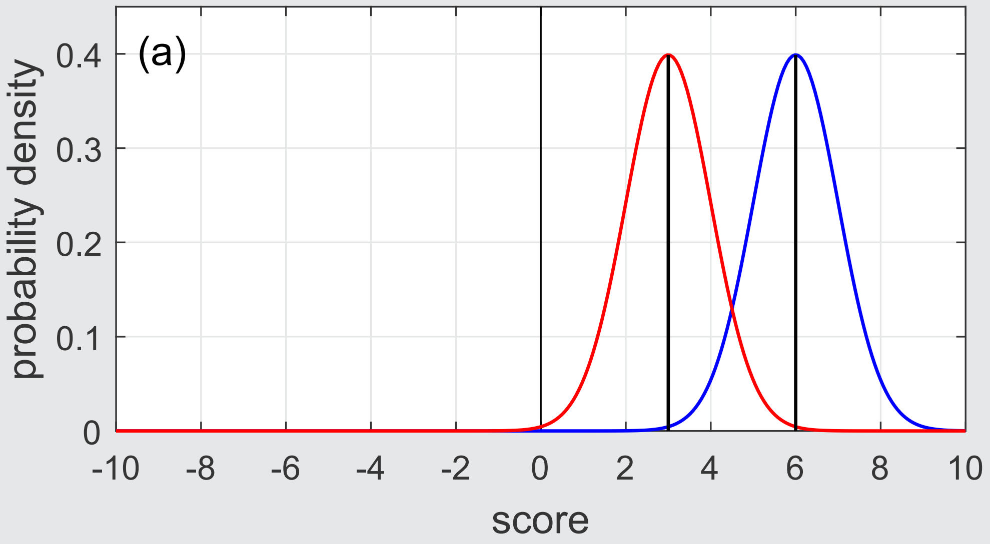
(a)

Uncalibrated scores

$\mu_d = 3$

$\mu_s = 6$

$\sigma = 1$



Linear calibration

(a)

Uncalibrated scores

$\mu_d = 3$

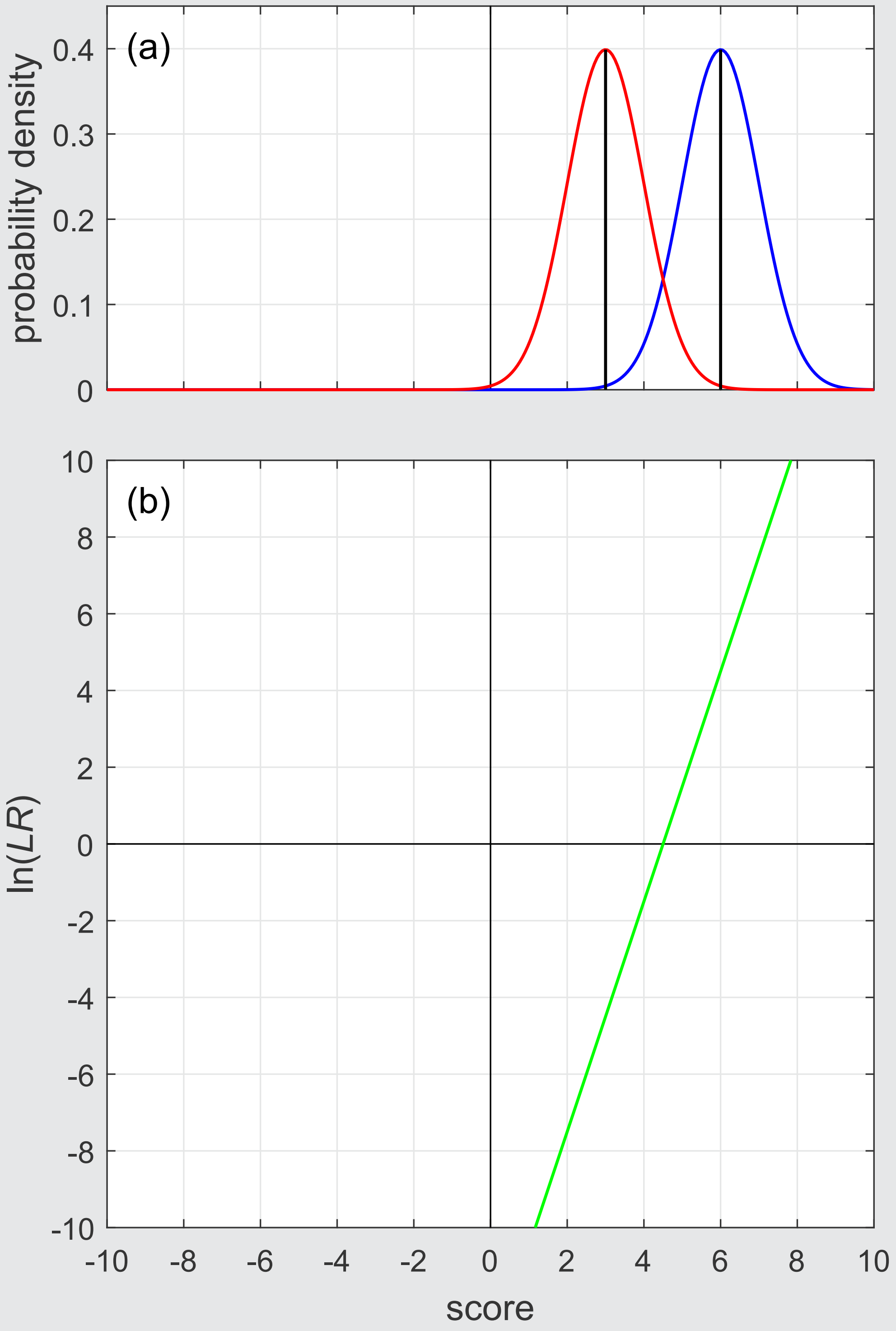
$\mu_s = 6$

$\sigma = 1$

(b)

Score to ln(*LR*)

mapping function



Linear calibration

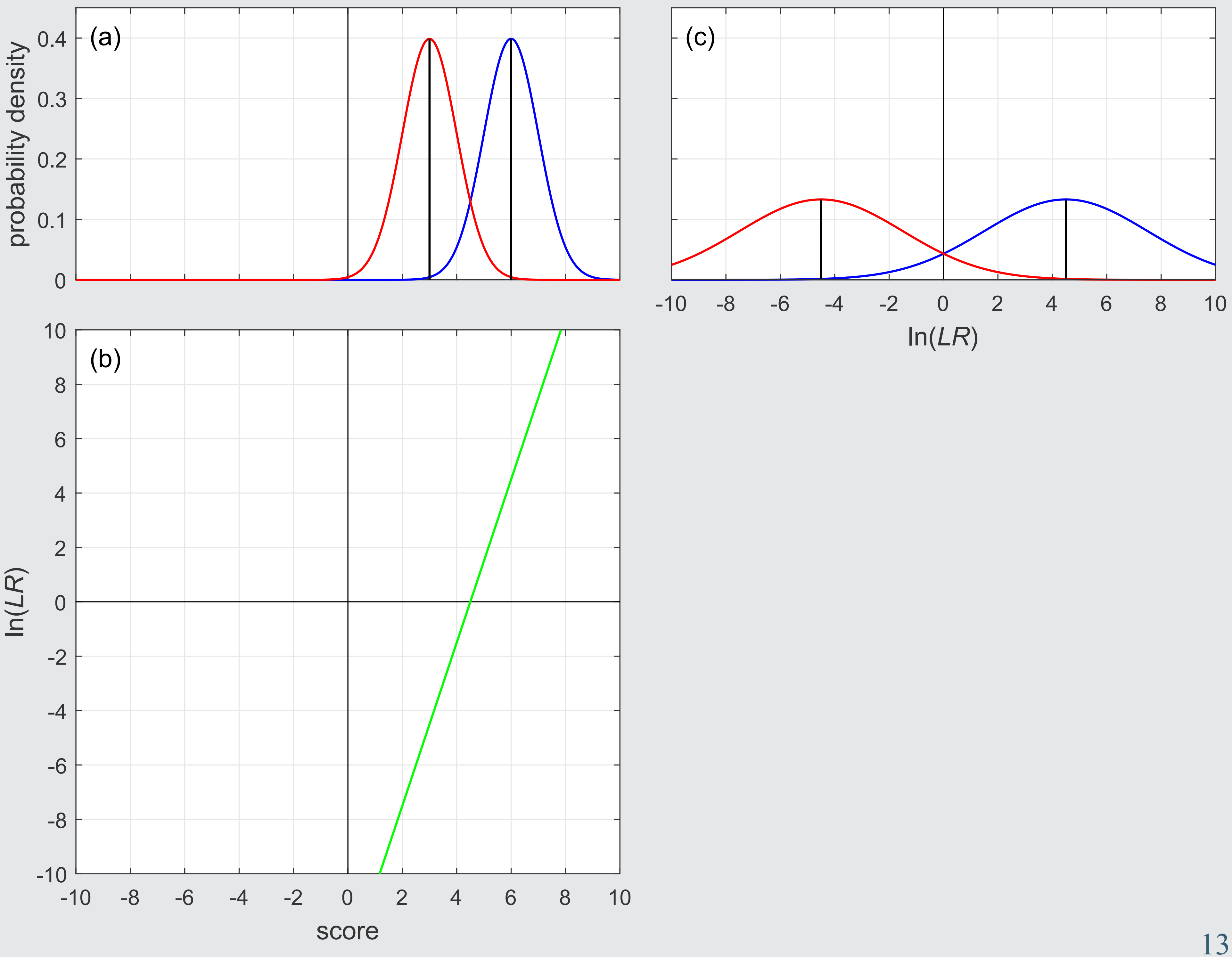
(c)

Calibrated ln(LR)

$\mu_d = -4.5$

$\mu_s = +4.5$

$\sigma = 3$



Linear calibration

(c)

Calibrated ln(*LR*)

$\mu_d = -4.5$

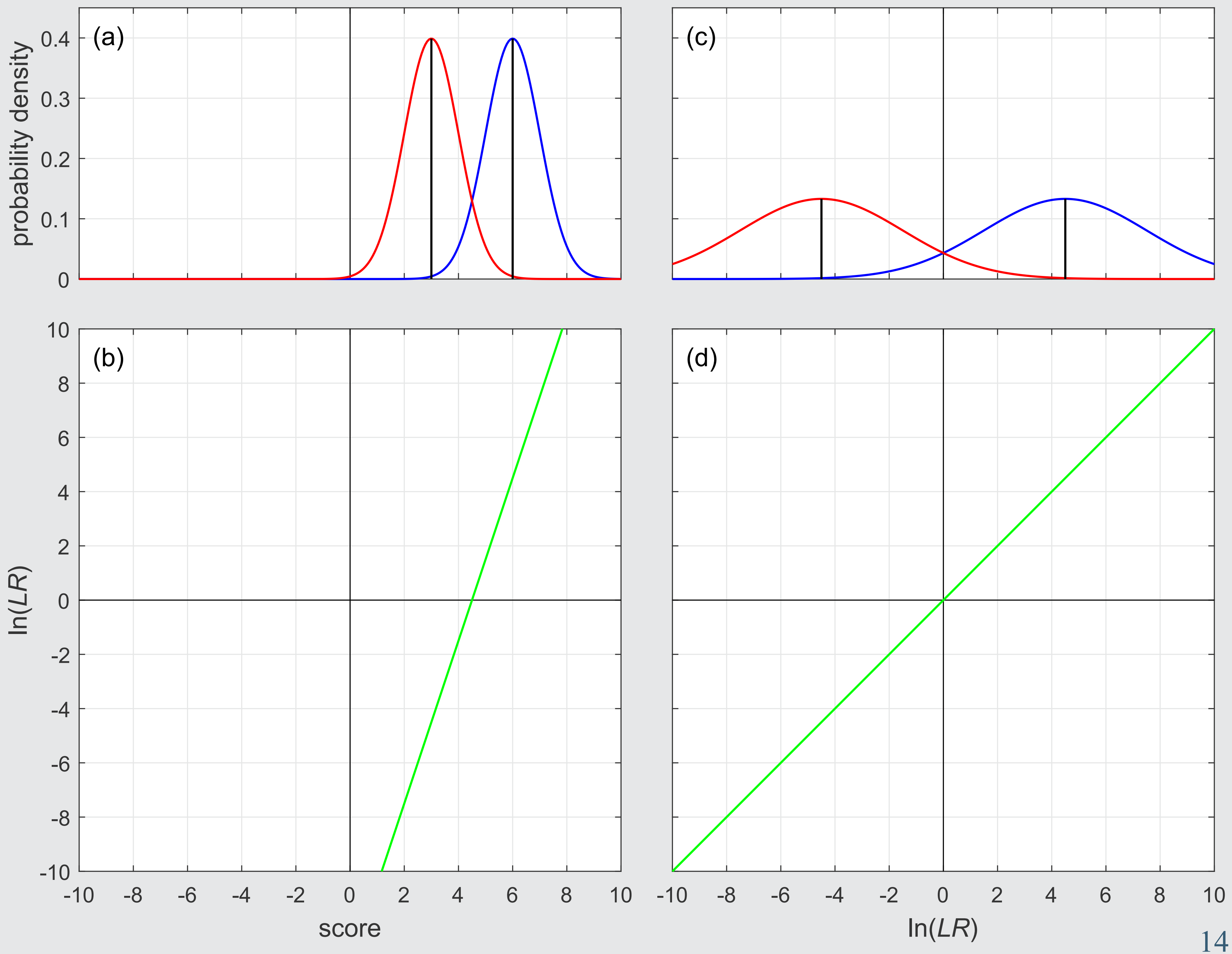
$\mu_s = +4.5$

$\sigma = 3$

(d)

ln(*LR*) to ln(*LR*)

mapping function



Linear calibration

- Score [x] to $\ln(LR)$ [y] mapping function:

$$y = a + bx$$

$$a = -b\frac{\mu_s + \mu_d}{2} \qquad b = \frac{\mu_s - \mu_d}{\sigma^2}$$

- Where μ_s, μ_d, σ are the statistics for the scores

Linear calibration

- Score [x] to $\ln(LR)$ [y] mapping function:

$$y = a + bx$$

- In practice, **logistic regression** is commonly used to calculate a and b

- It is more robust to violations of the assumptions of Gaussian distributions with the same variance

Non-linear calibration

- Logistic-regression calibration applies a linear transformation in the log-likelihood-ratio space.

- Unless the distributions of the different-source and same-source uncalibrated log likelihood ratios are both Gaussian and have the same variance, the calibrated log likelihood ratios could be far from a perfectly calibrated bi-Gaussian system.

- Bi-Gaussianized calibration applies a non-linear (but still monotonic) transformation designed to bring the distributions closer to those of a perfectly-calibrated bi-Gaussian system.

Non-linear calibration

- kernel-density estimation (KDE)
    - monotonicity not guaranteed

- pool adjacent violators (PAV), aka isotonic regression
    - overfits training data

Bi-Gaussianized calibration

1. Calculate uncalibrated log 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 log 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 log 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 log 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

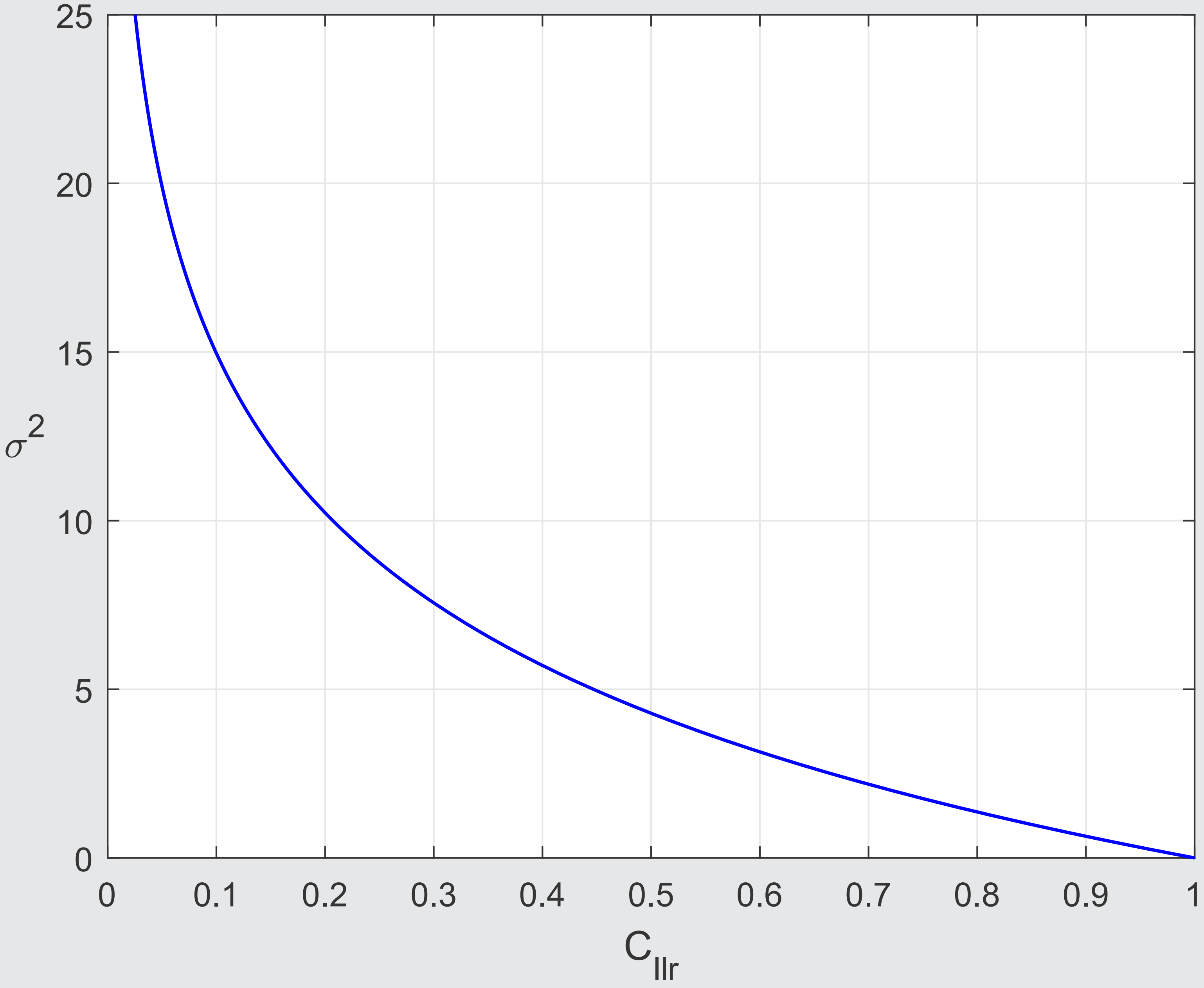
- for a perfectly-calibrated

bi-Gaussian system

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

$b = 17.7$

$c = 9.33\times10^{-3}$



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.

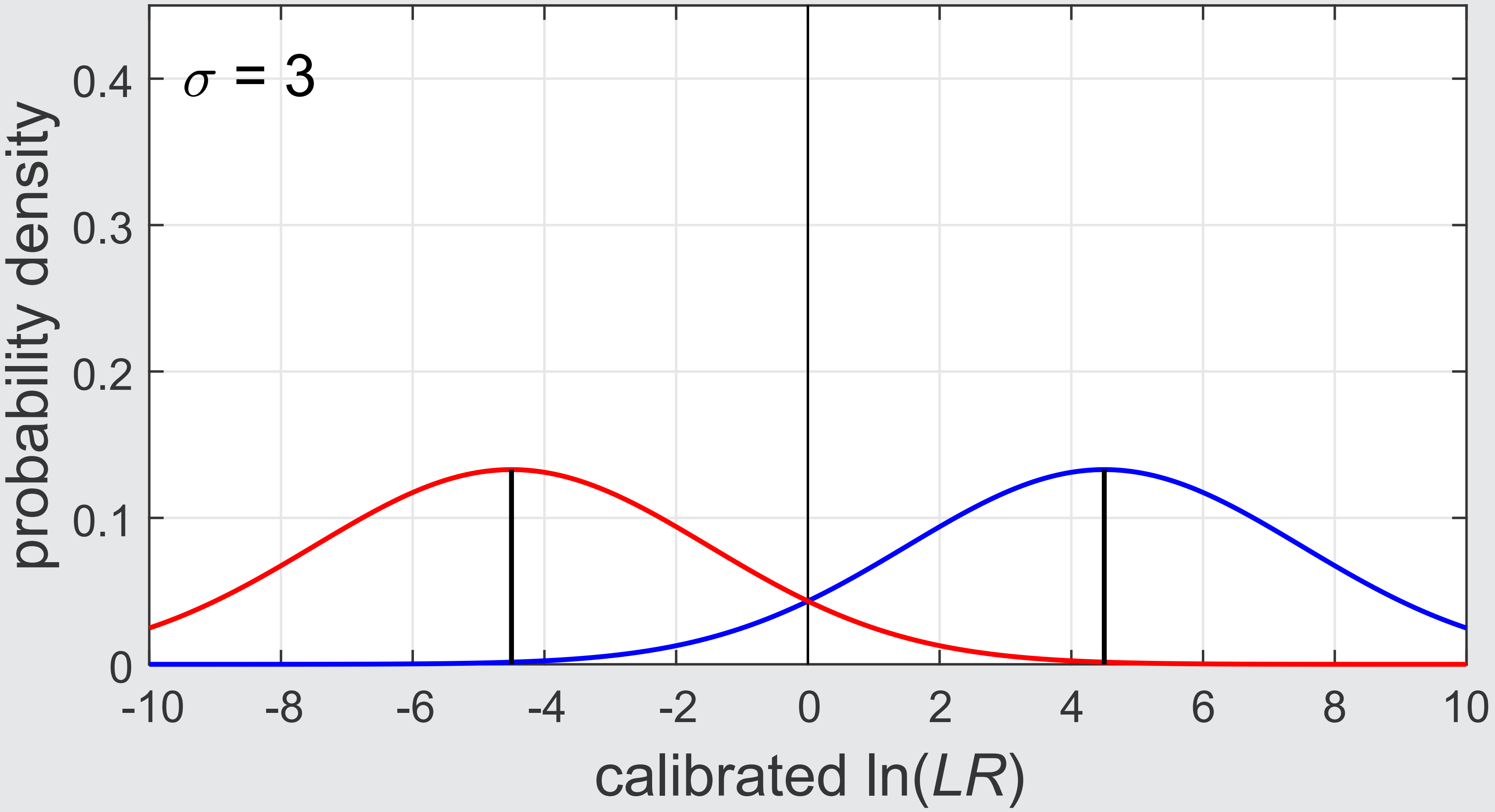
4. Determine the σ^2 of the perfectly-calibrated bi-Gaussian system with the C_{llr} calculated at Step 3.

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.

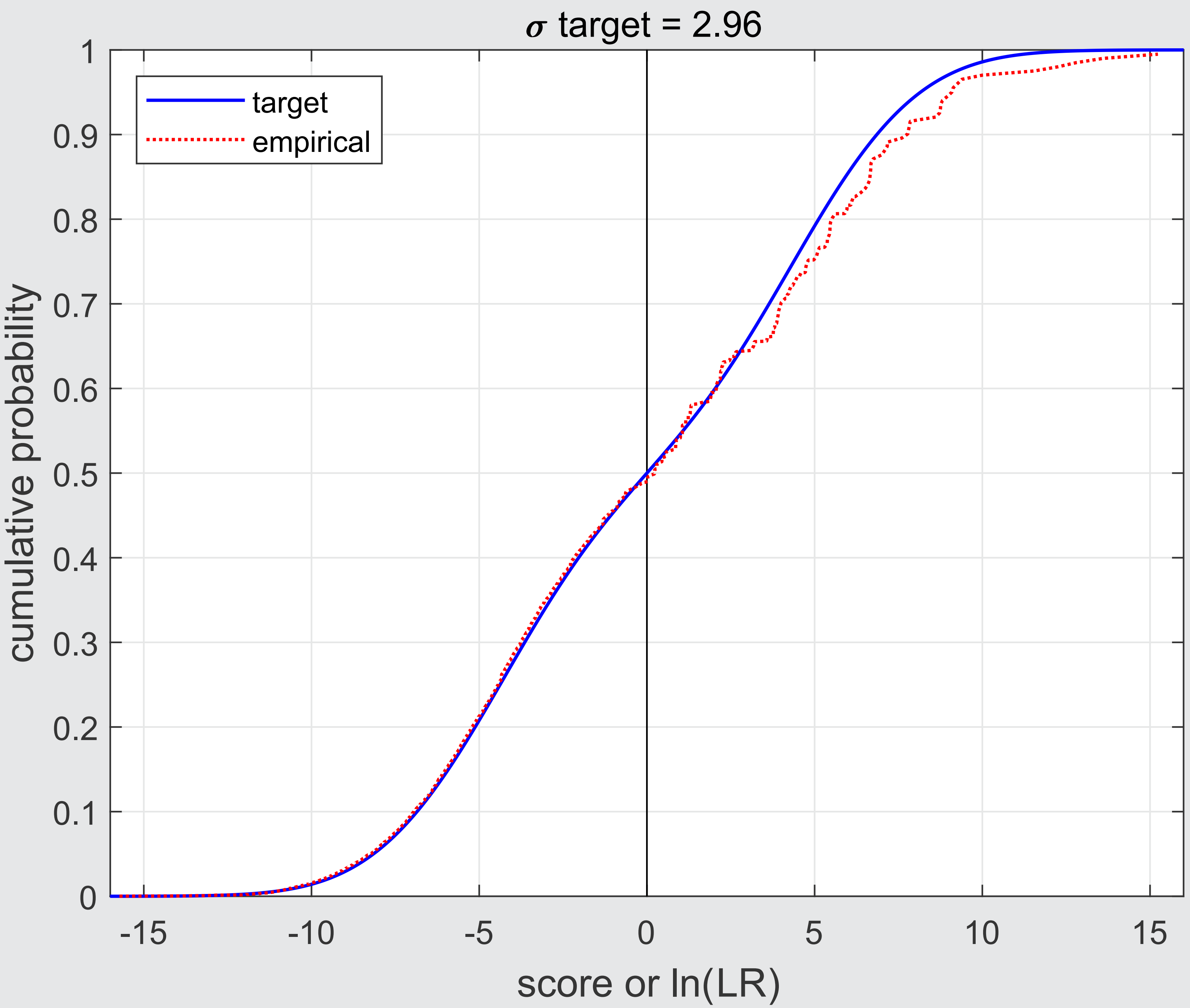
Simulated data: Gaussians with same variance

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



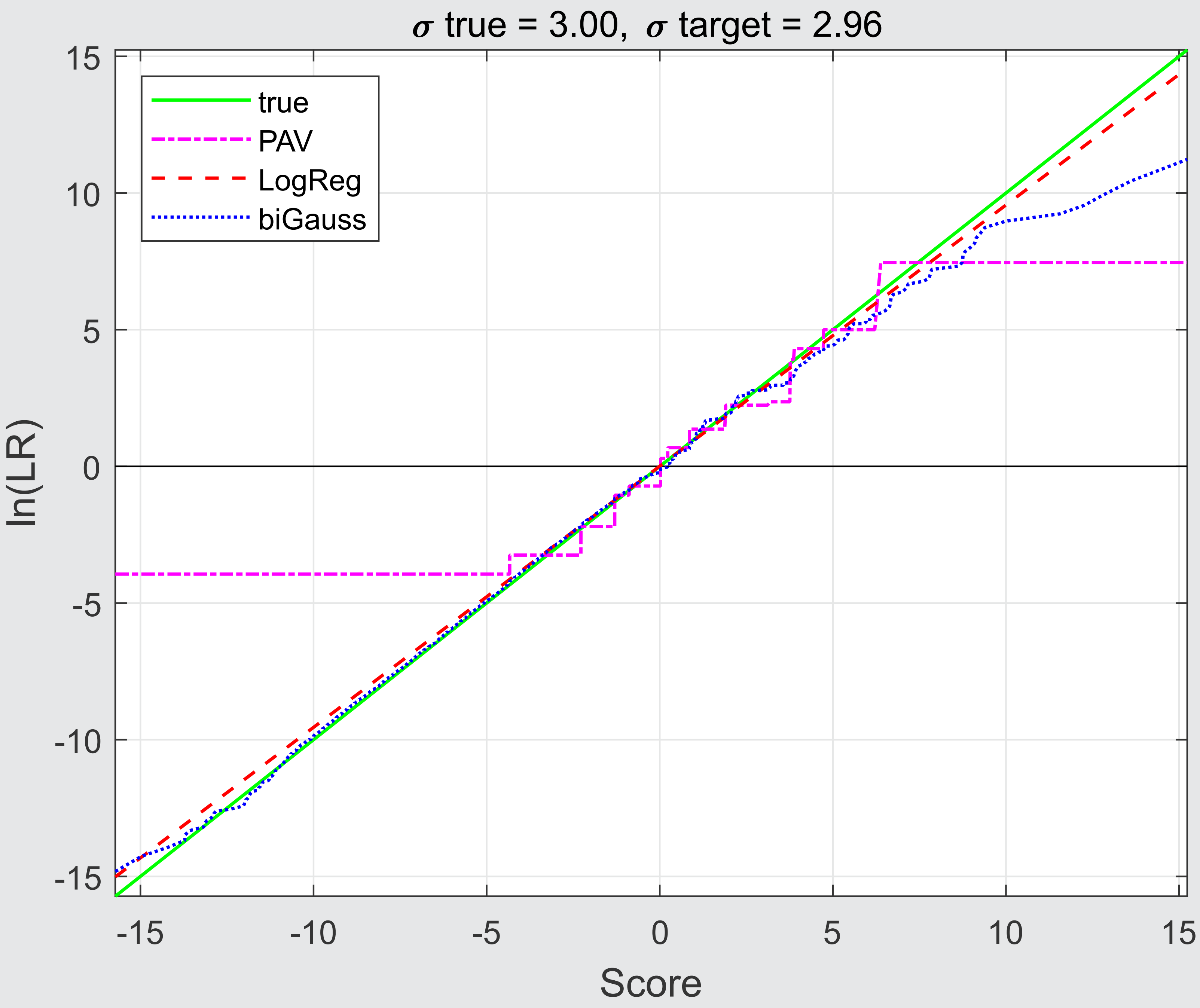
Simulated data: Gaussians with same variance

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



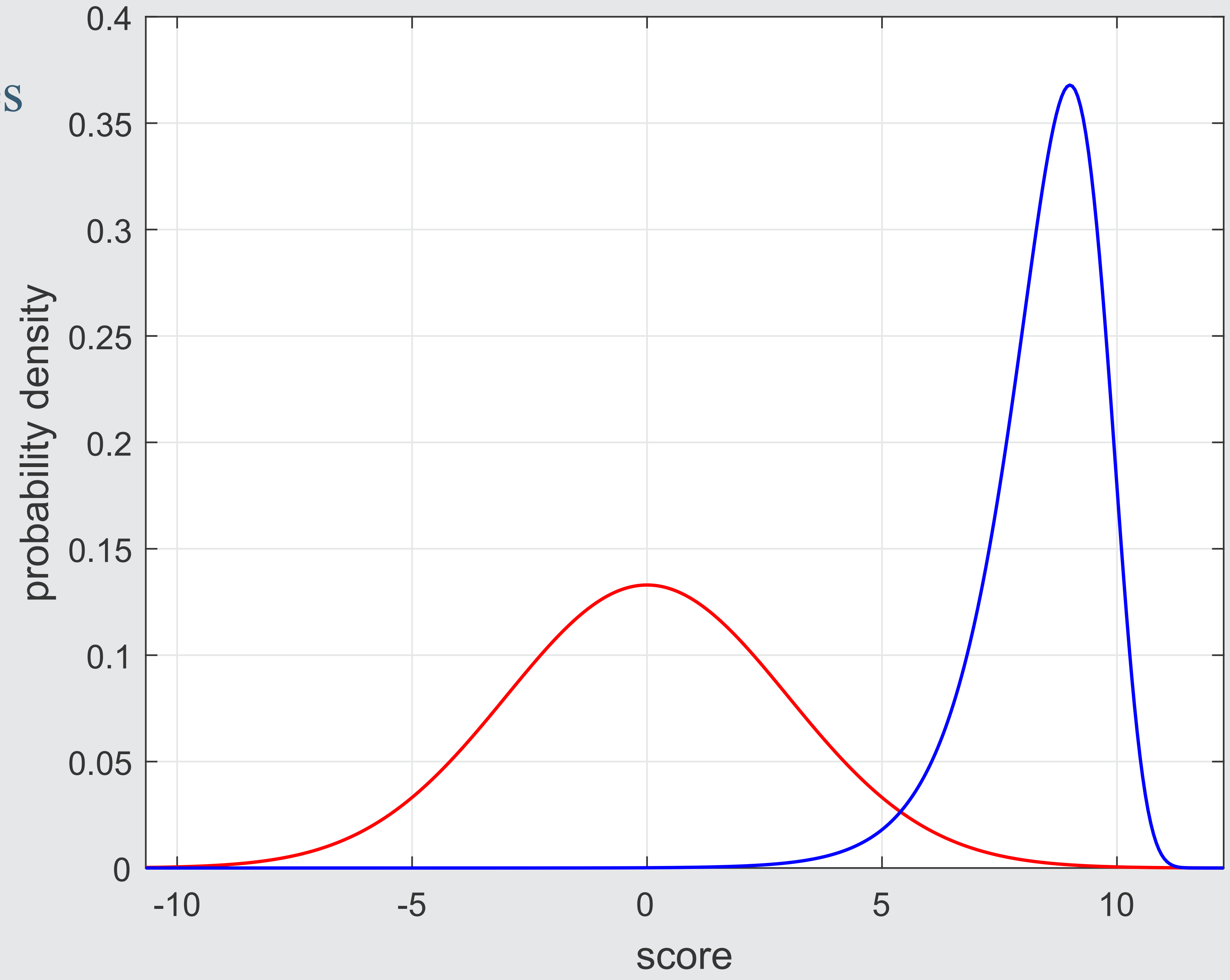
Simulated data: Gaussians with same variance

- Mapping functions



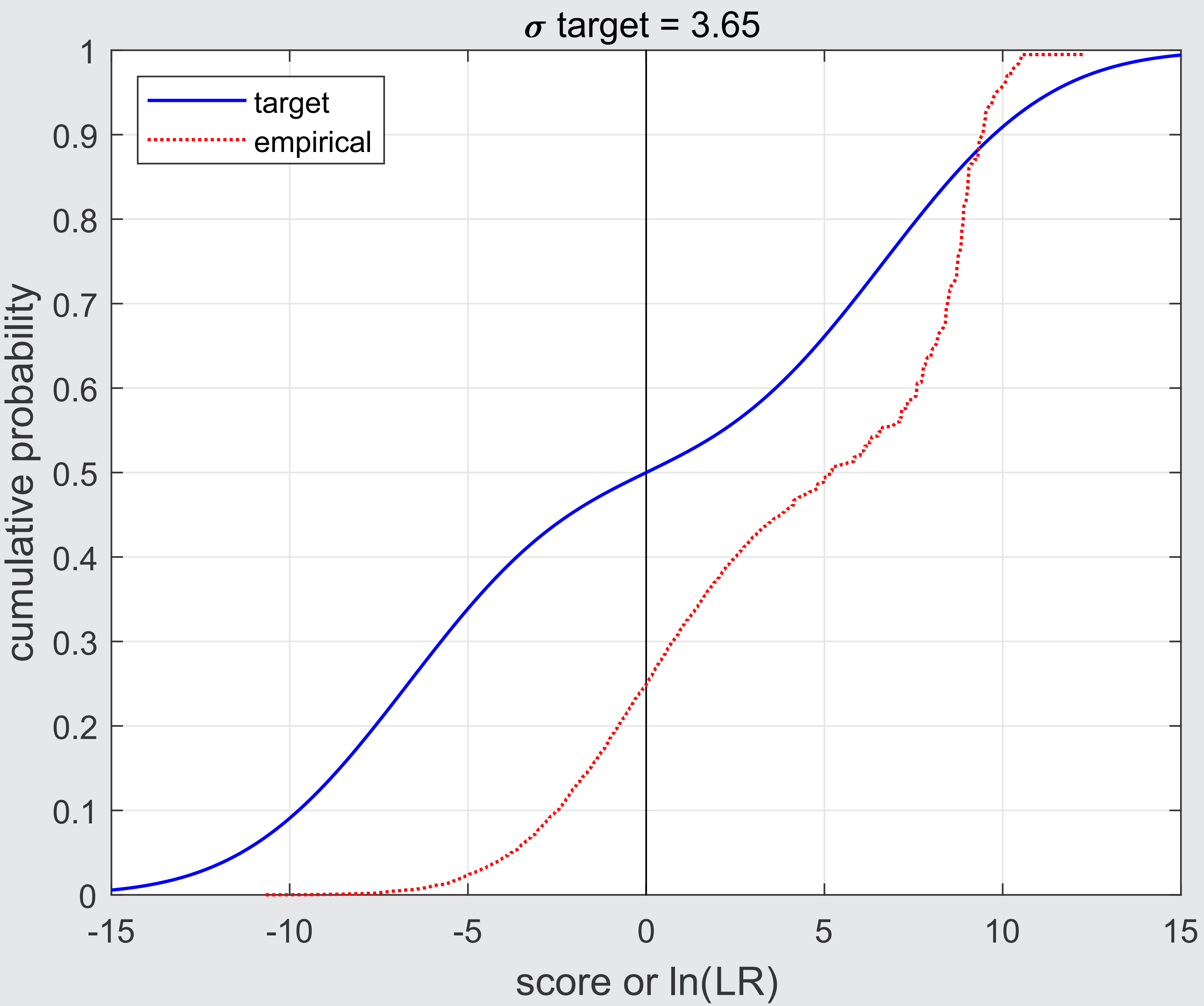
Simulated data: Gaussian & Gumbel with different variances

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



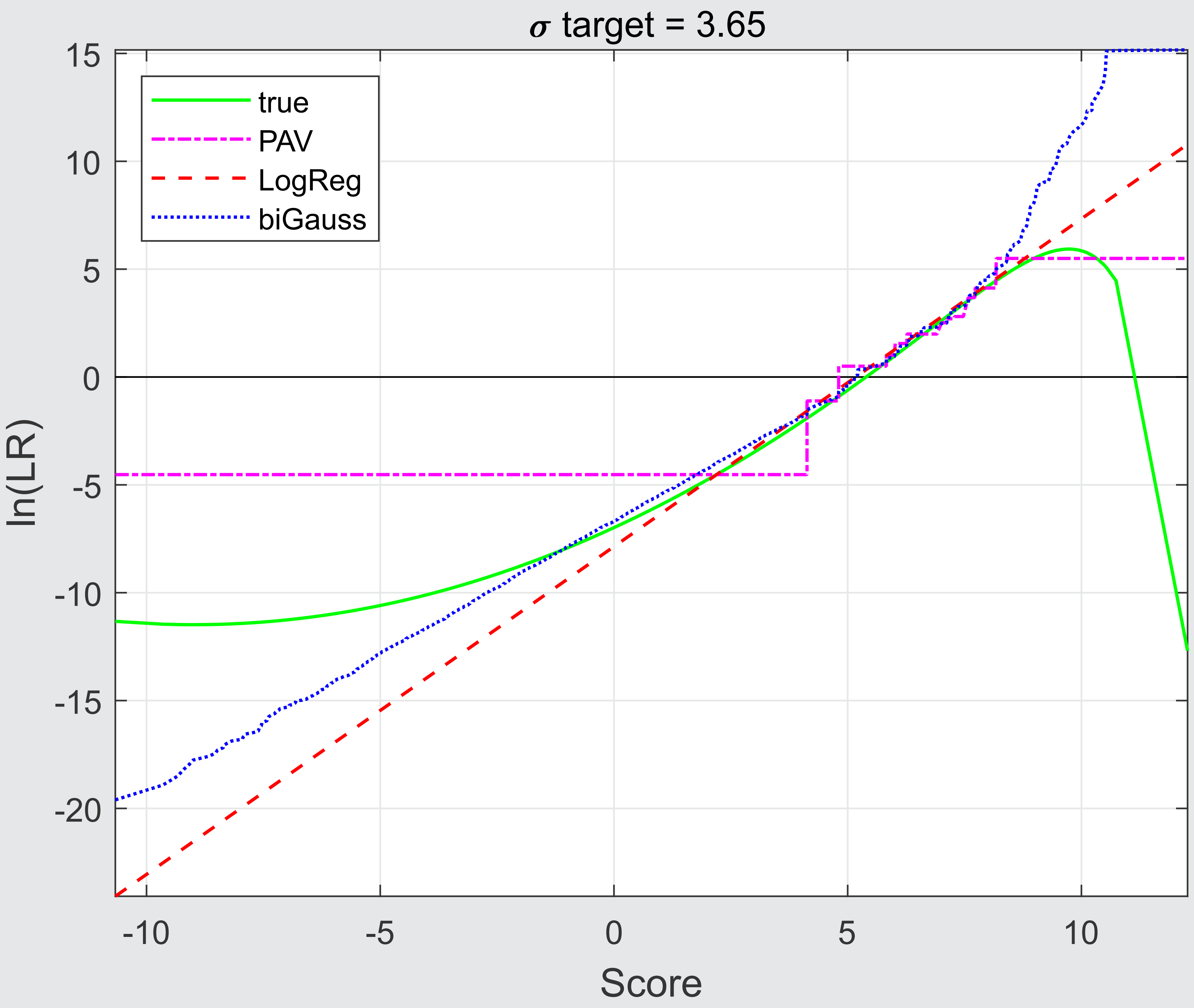
Simulated data: Gaussian & Gumbel with different variances

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



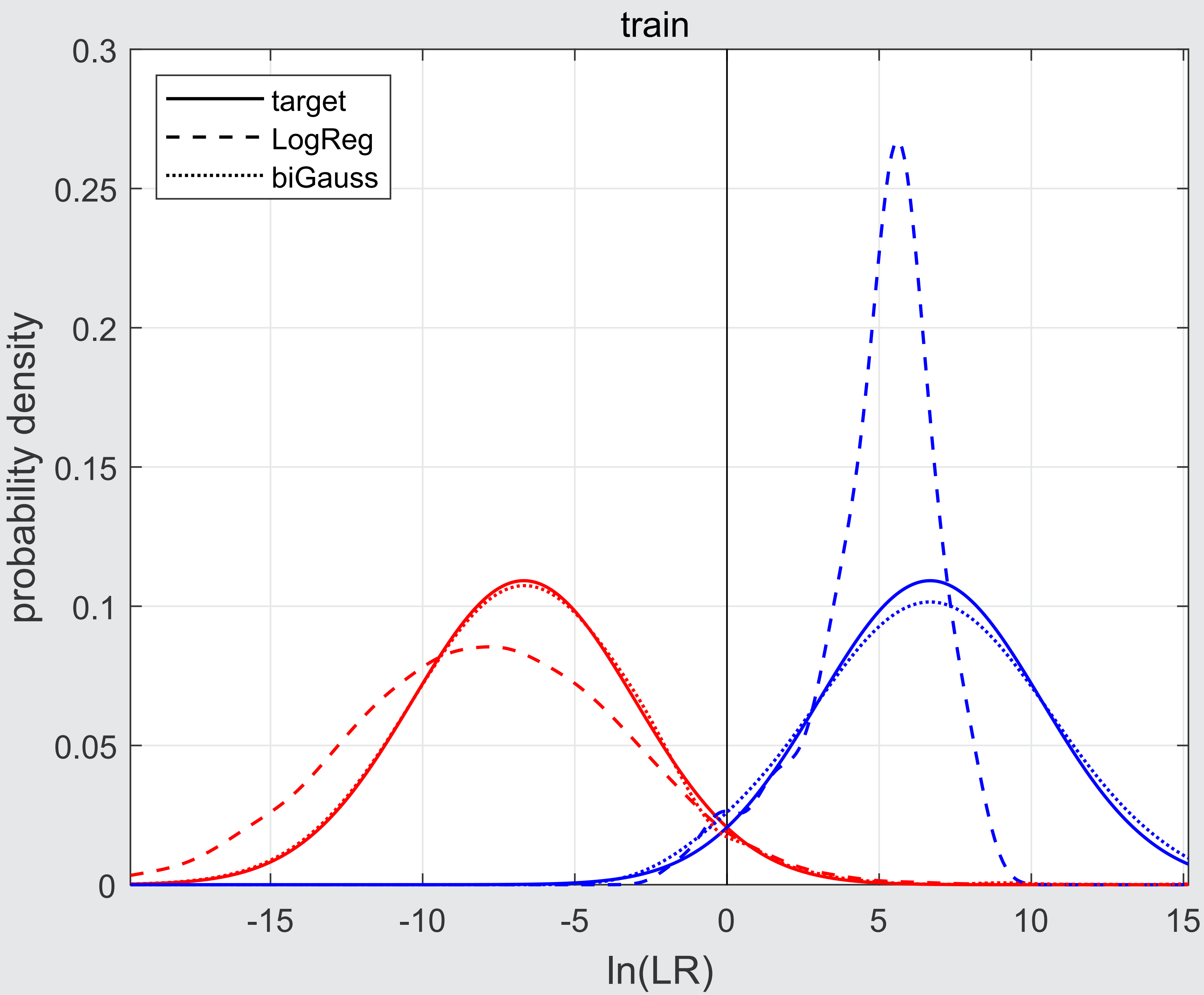
Simulated data: Gaussian & Gumbel with different variances

- Mapping functions



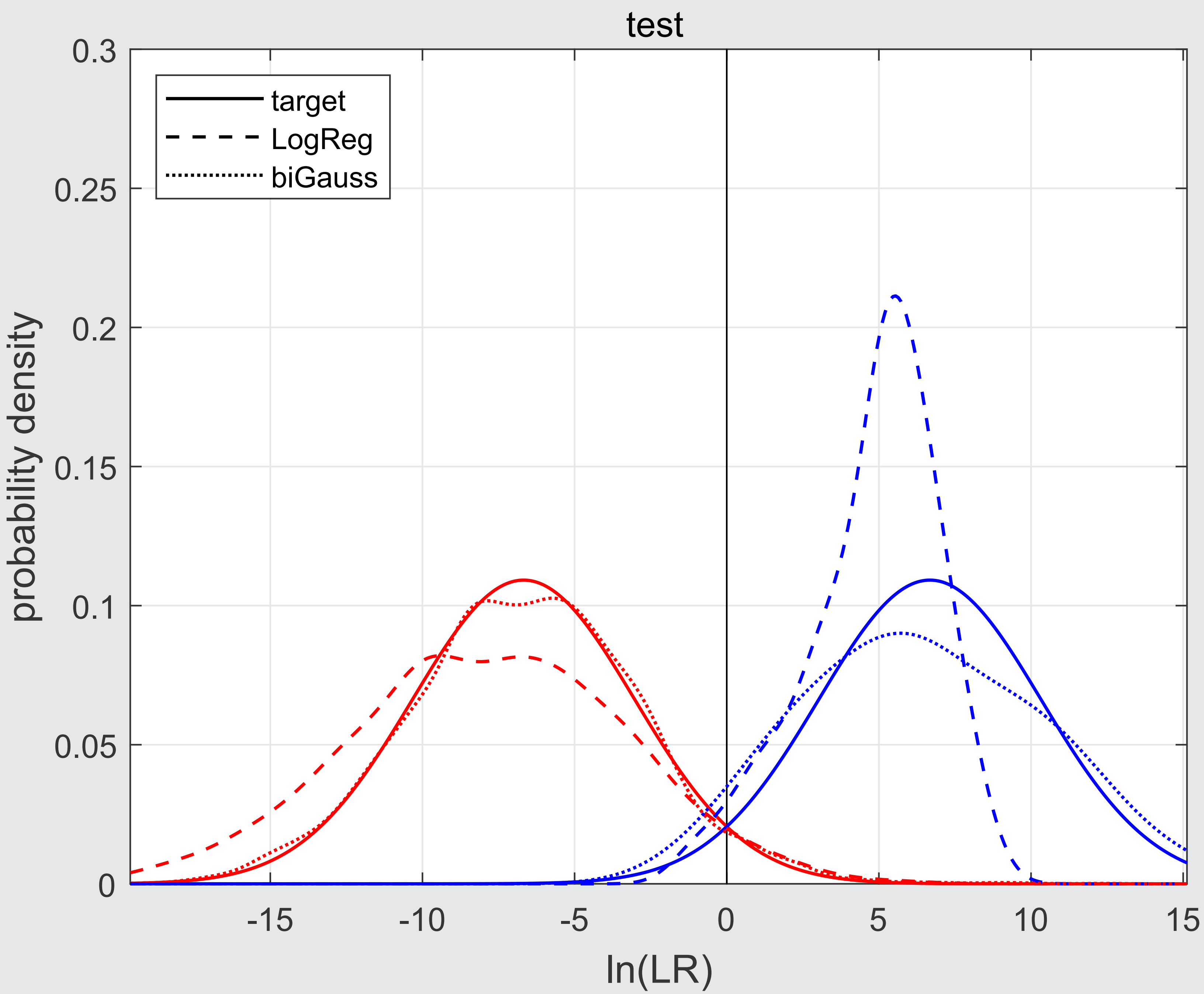
Simulated data: Gaussian & Gumbel with different variances

- Probability density functions
  - training data



Simulated data: Gaussian & Gumbel with different variances

- Probability density functions
  - test data

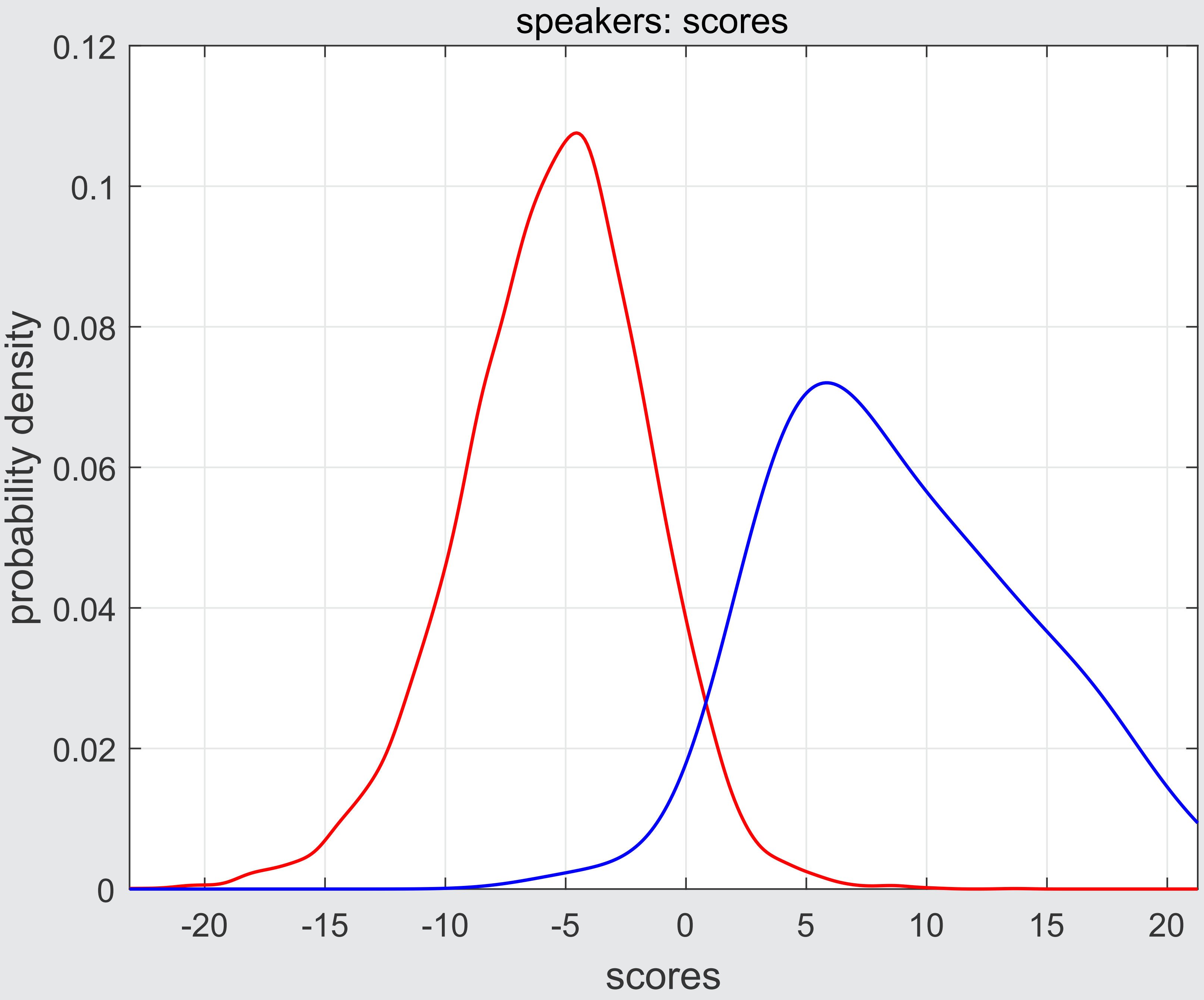


Real data: forensic voice comparison

- *forensic_eval_01*
  benchmark data

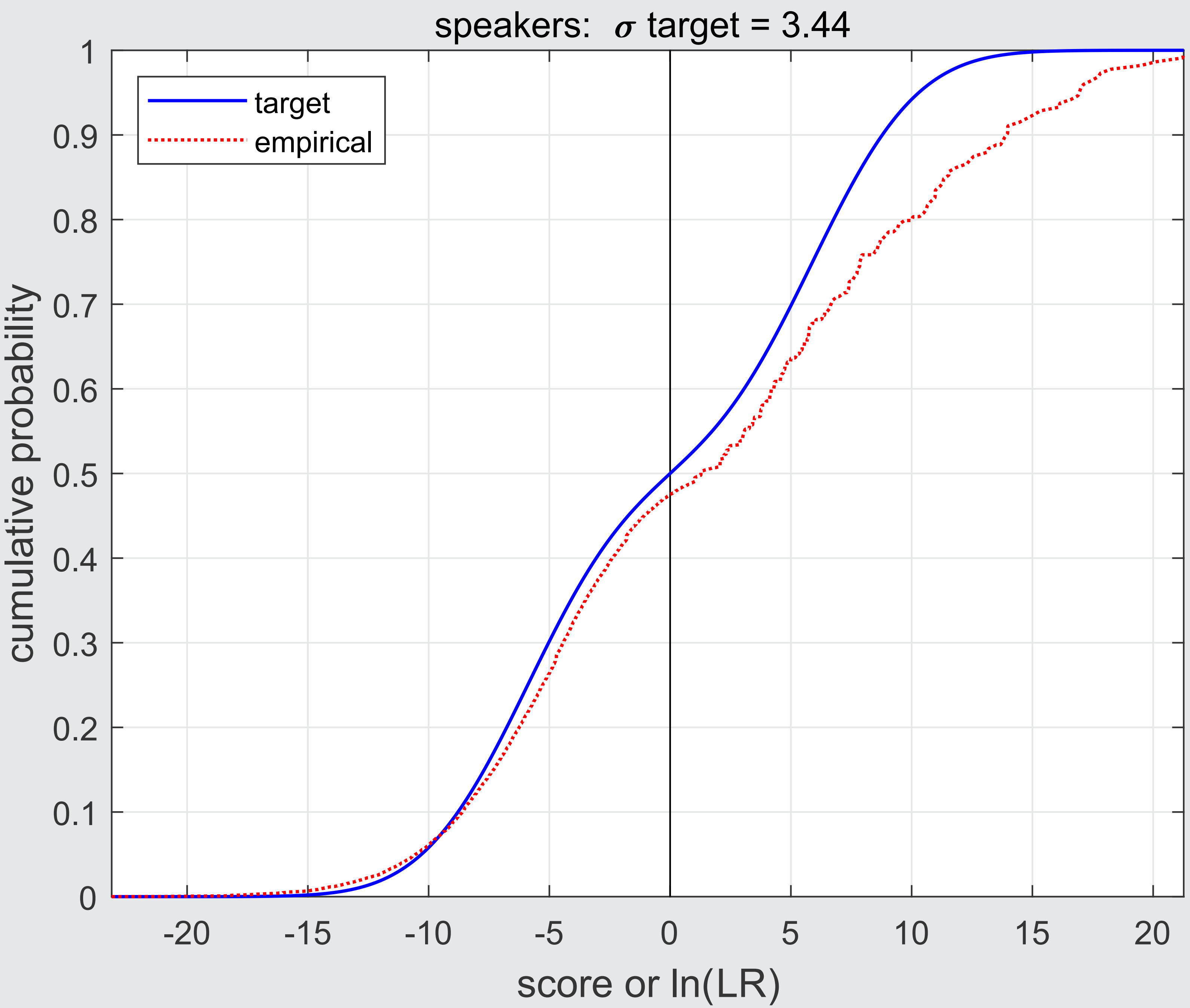
- E^3 Forensic Speech Science System (E^3FS^3)

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



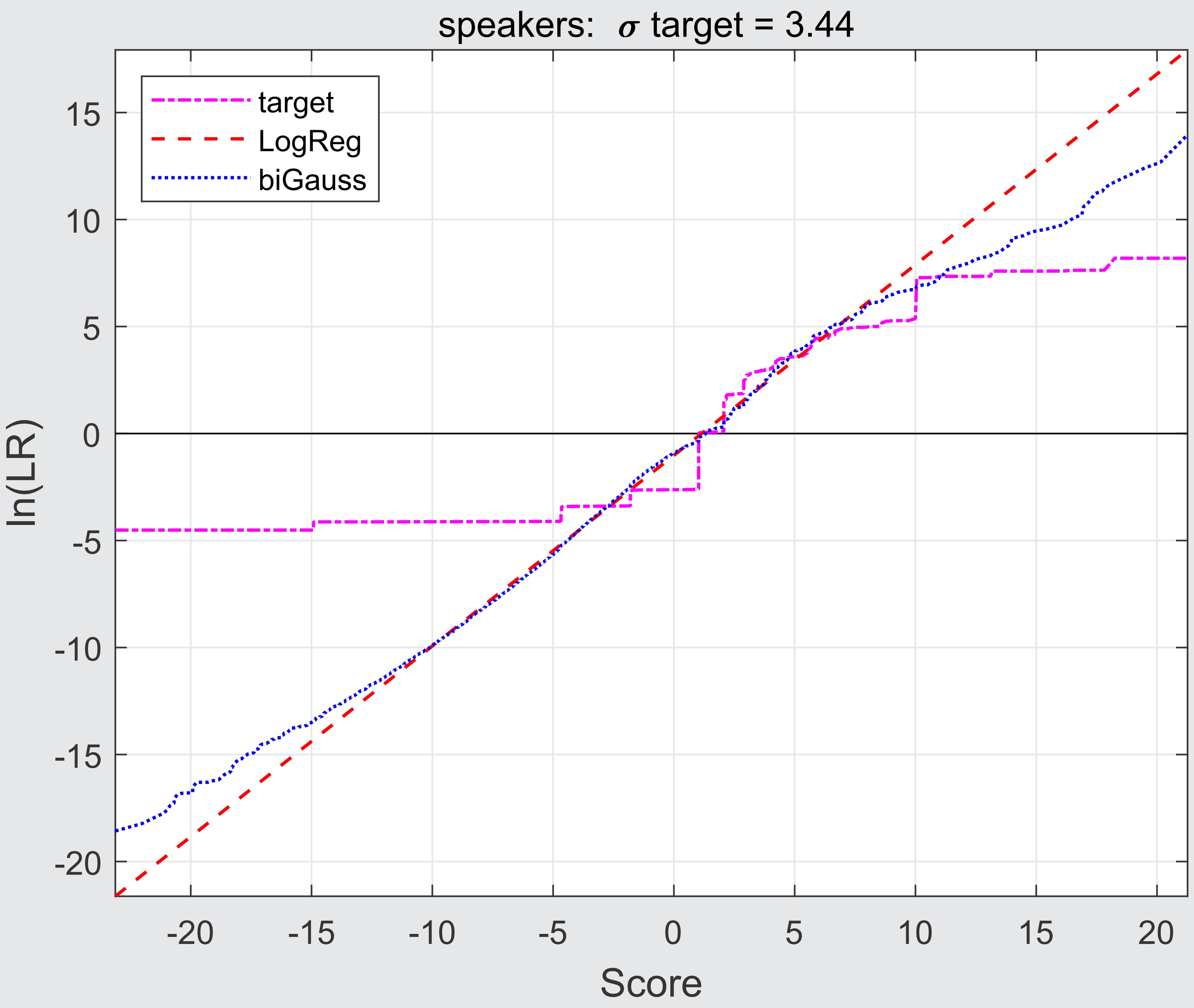
Real data: forensic voice comparison

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



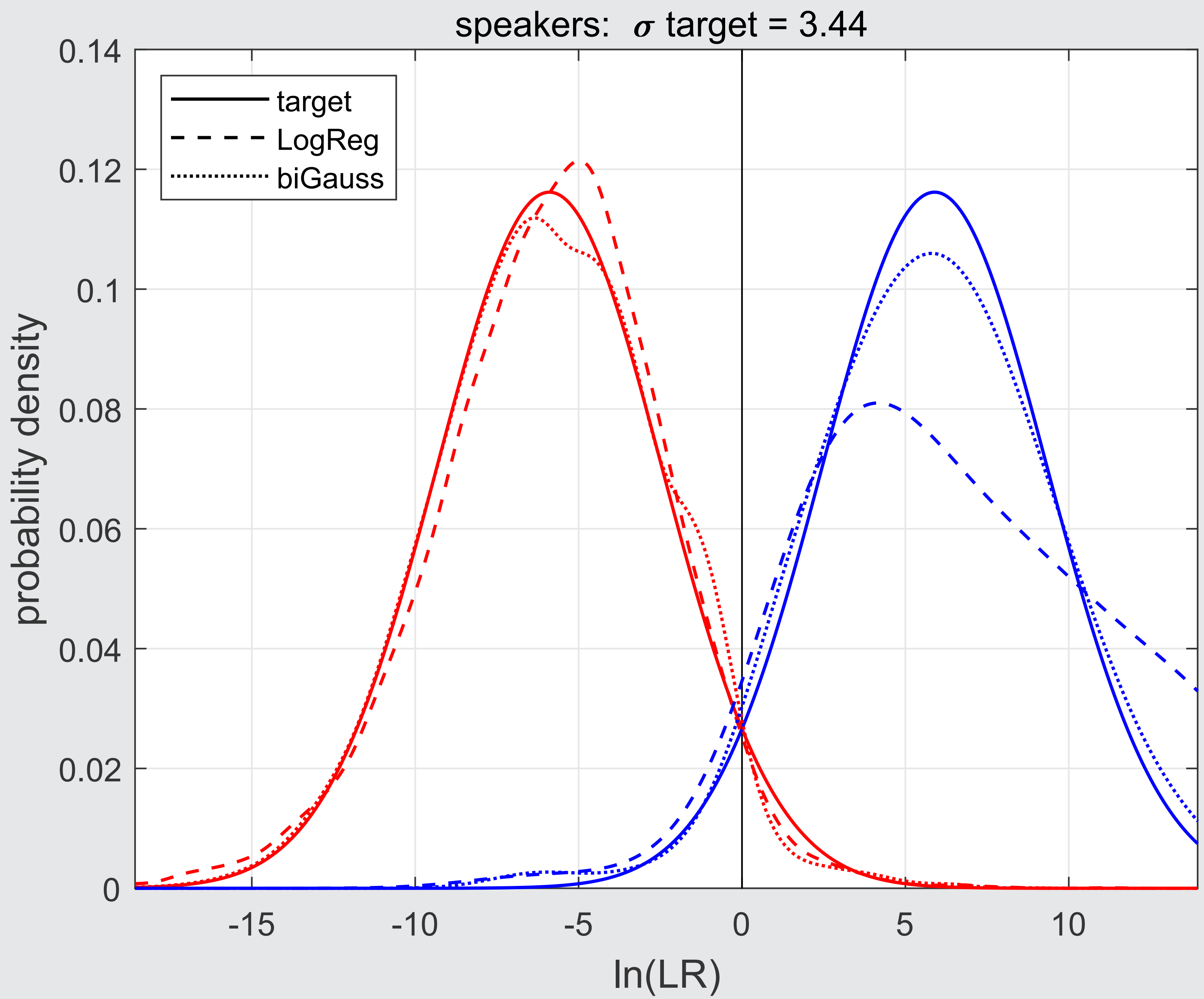
Real data: forensic voice comparison

- Mapping functions
    - cross-validated



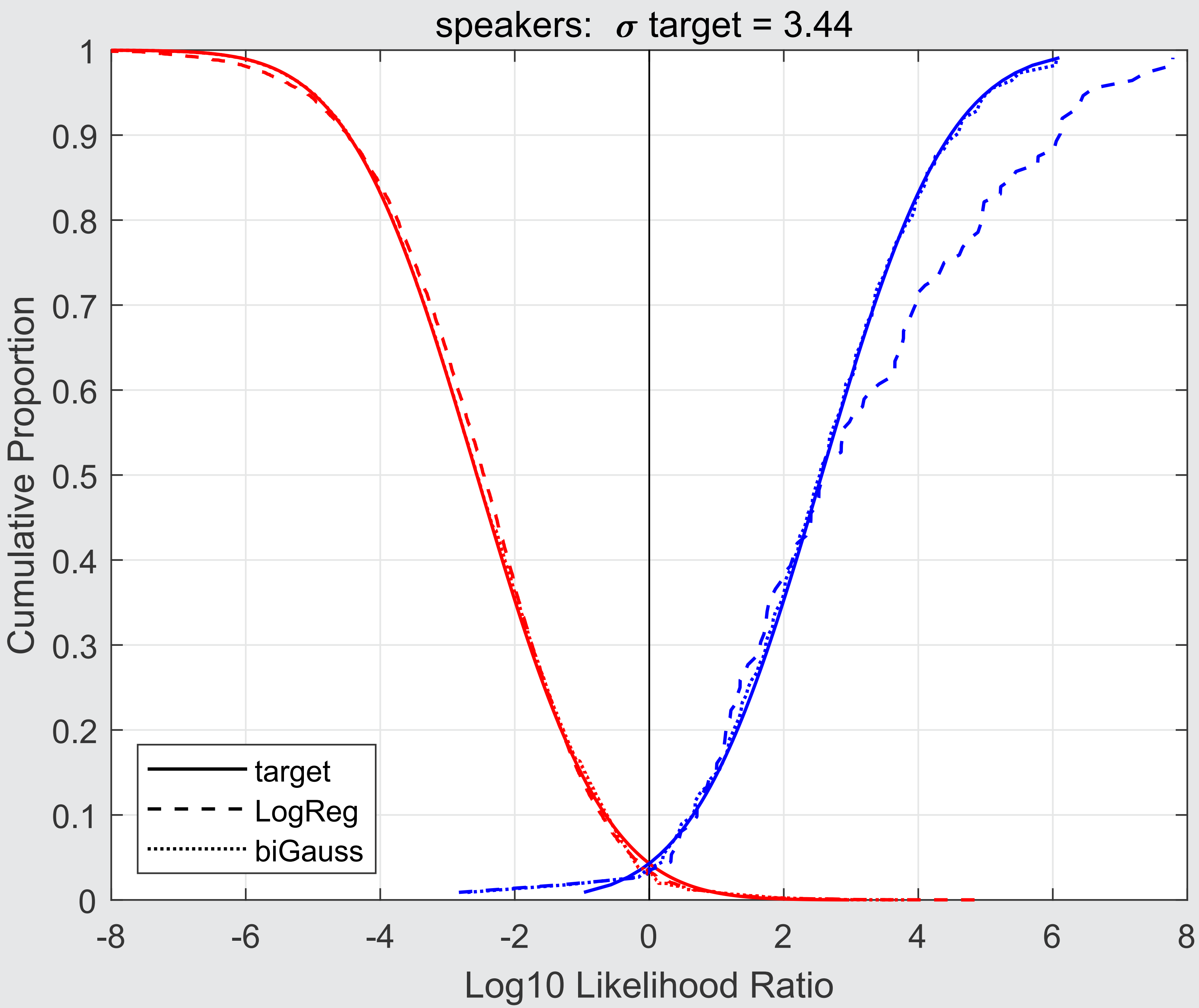
Real data: forensic voice comparison

- Probability density functions
  - cross-validated



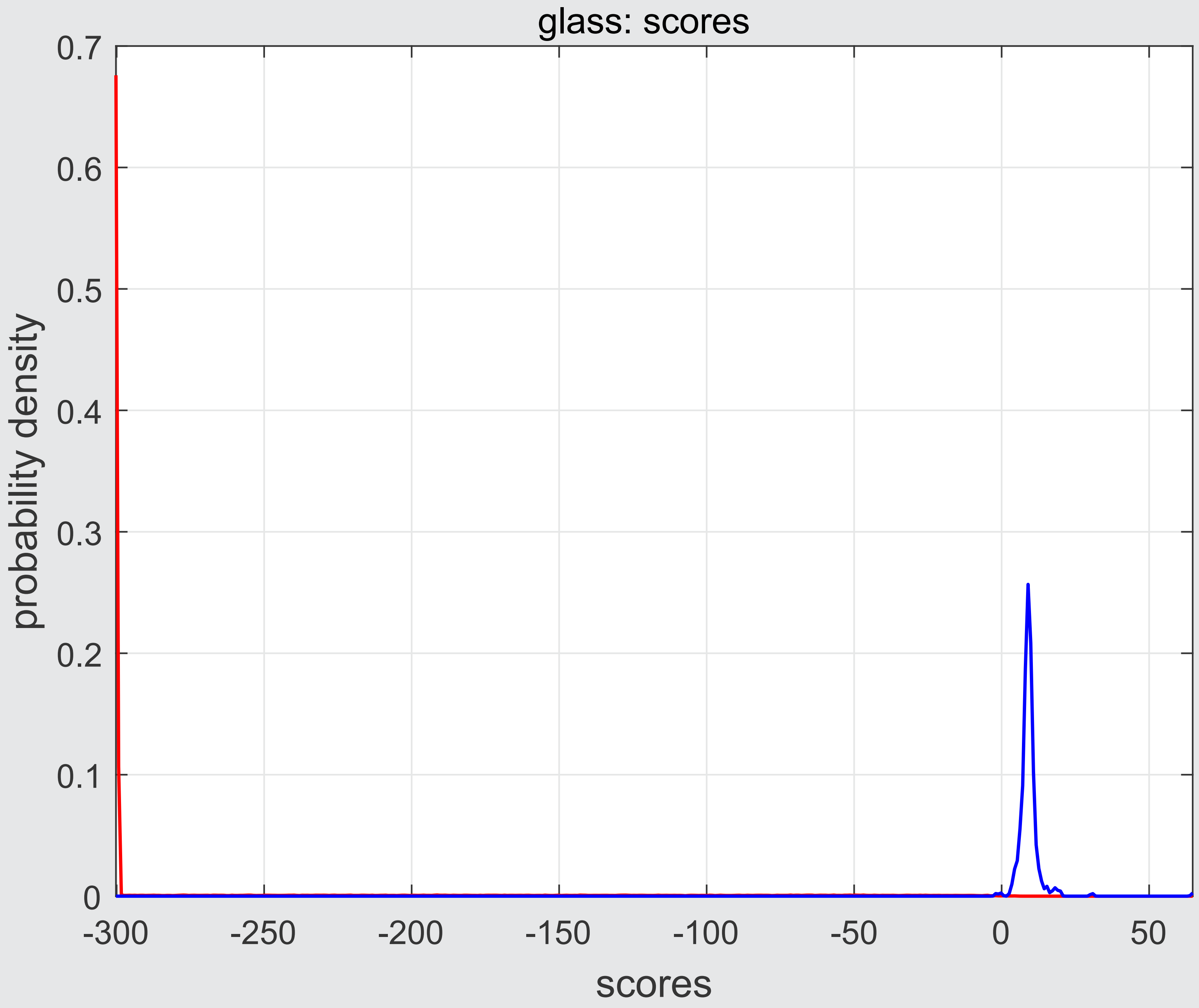
Real data: forensic voice comparison

- Tippett plots
    - cross-validated



Real data: glass

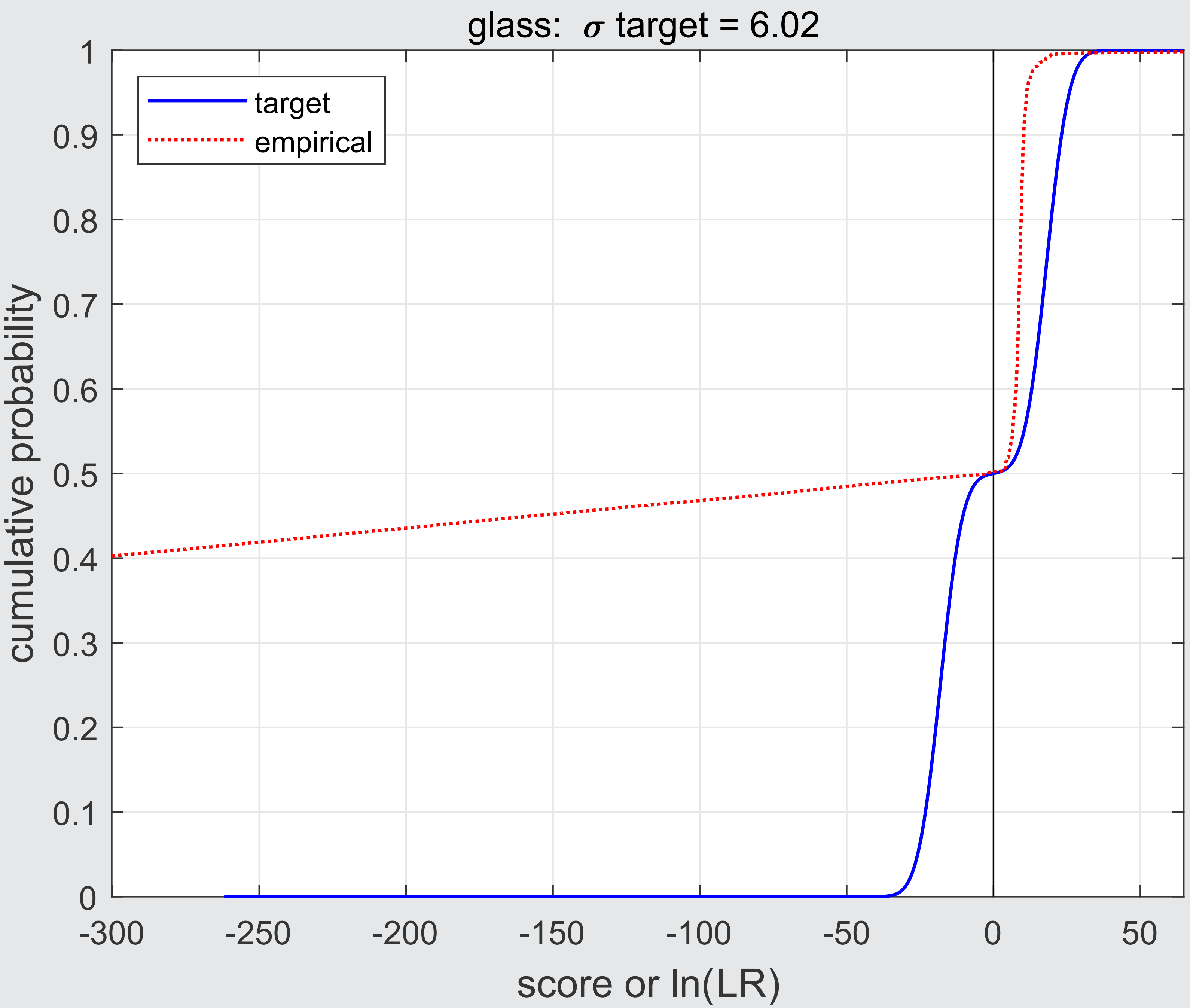
- glass scores from van Es et al. (2017)



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

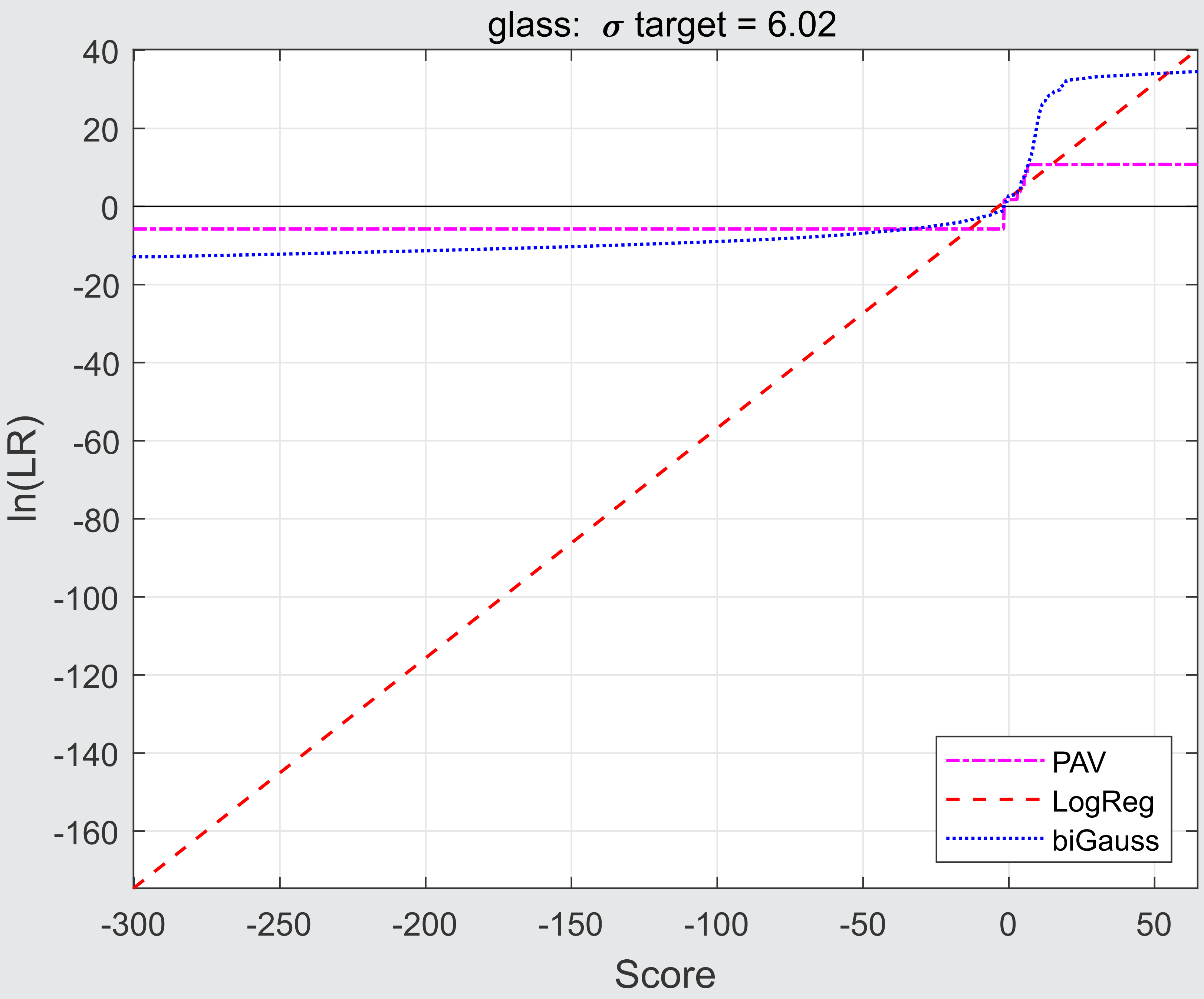
Real data: glass

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



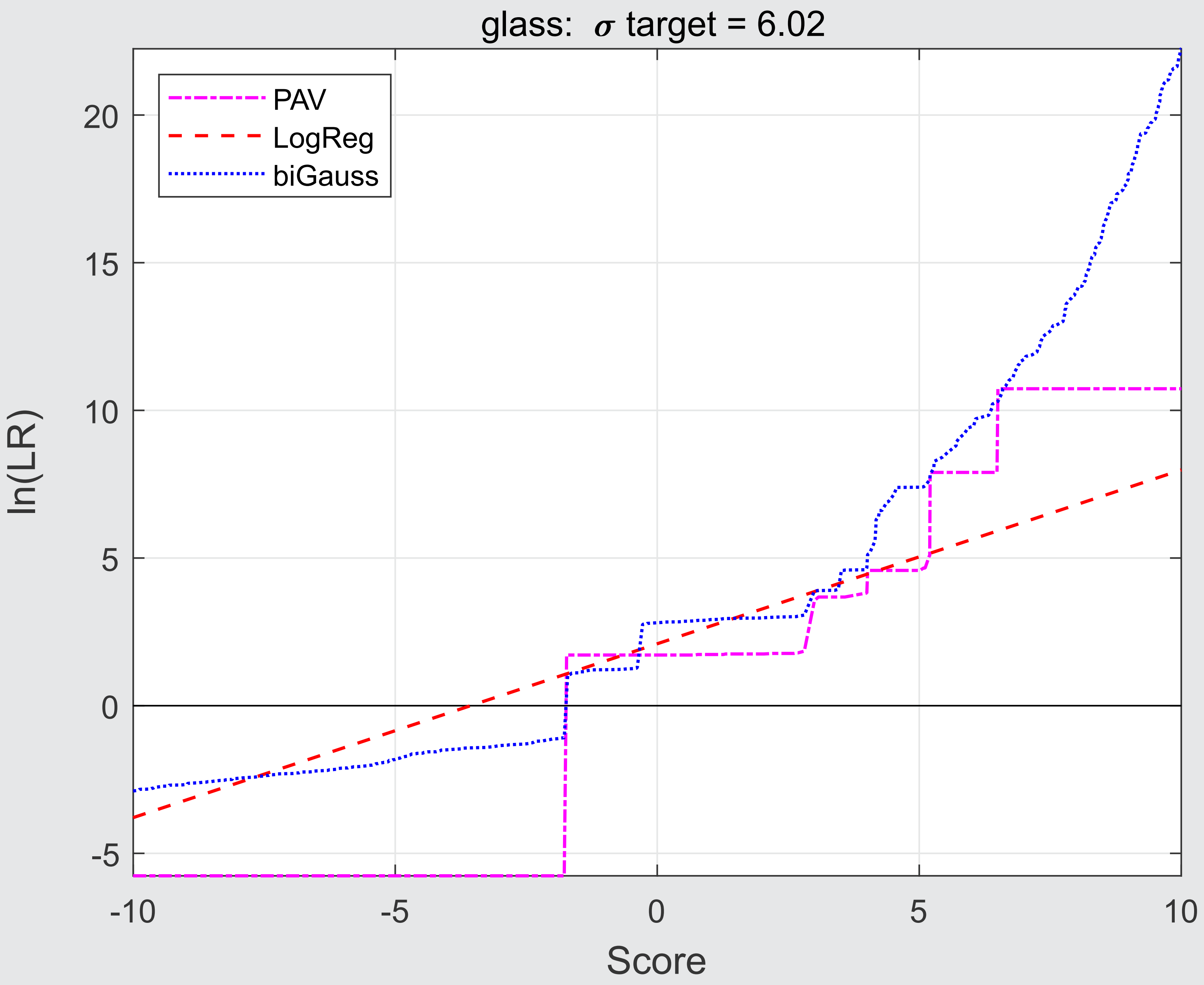
Real data: glass

- Mapping functions
    - cross-validated



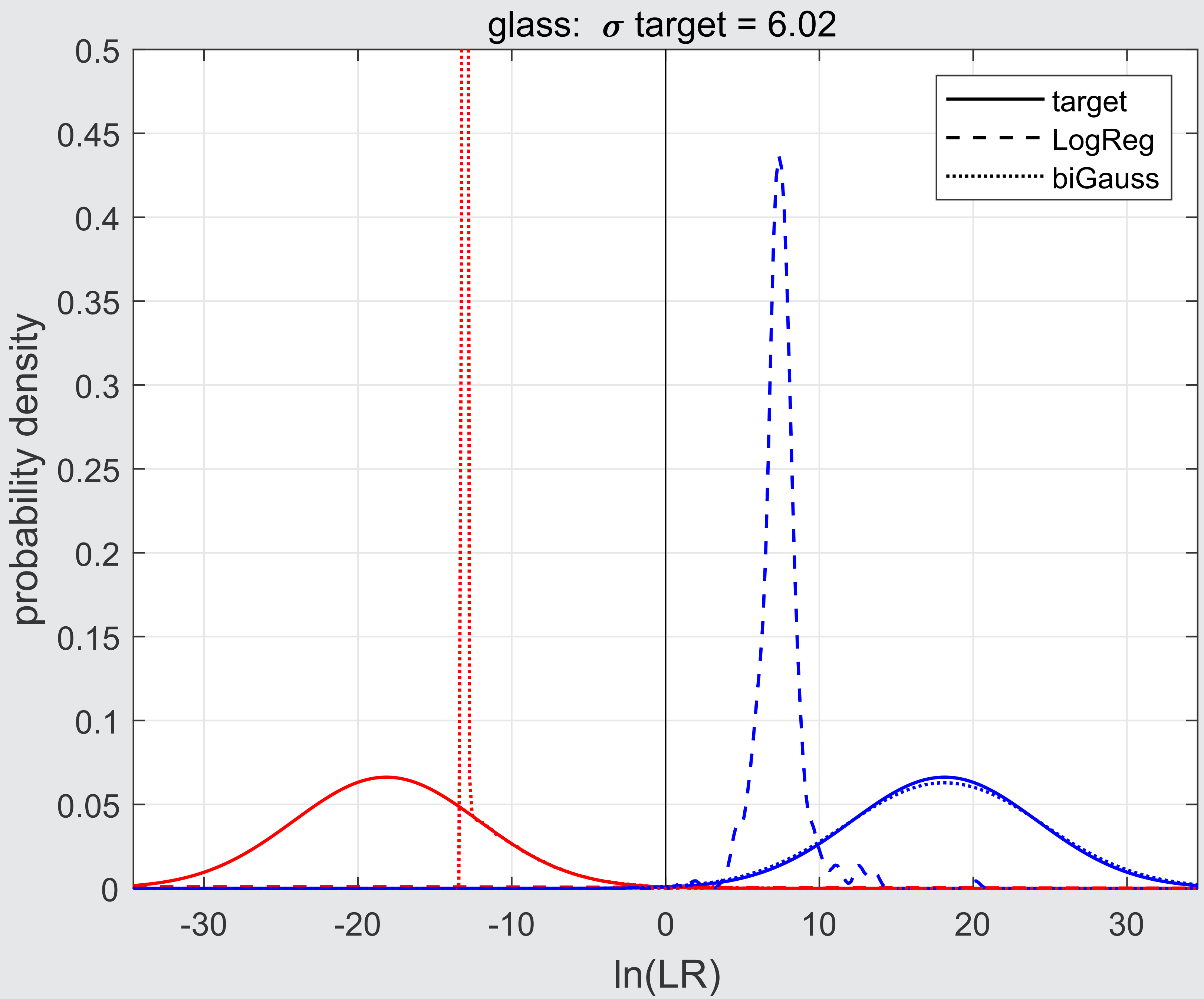
Real data: glass

- Mapping functions
    - cross-validated



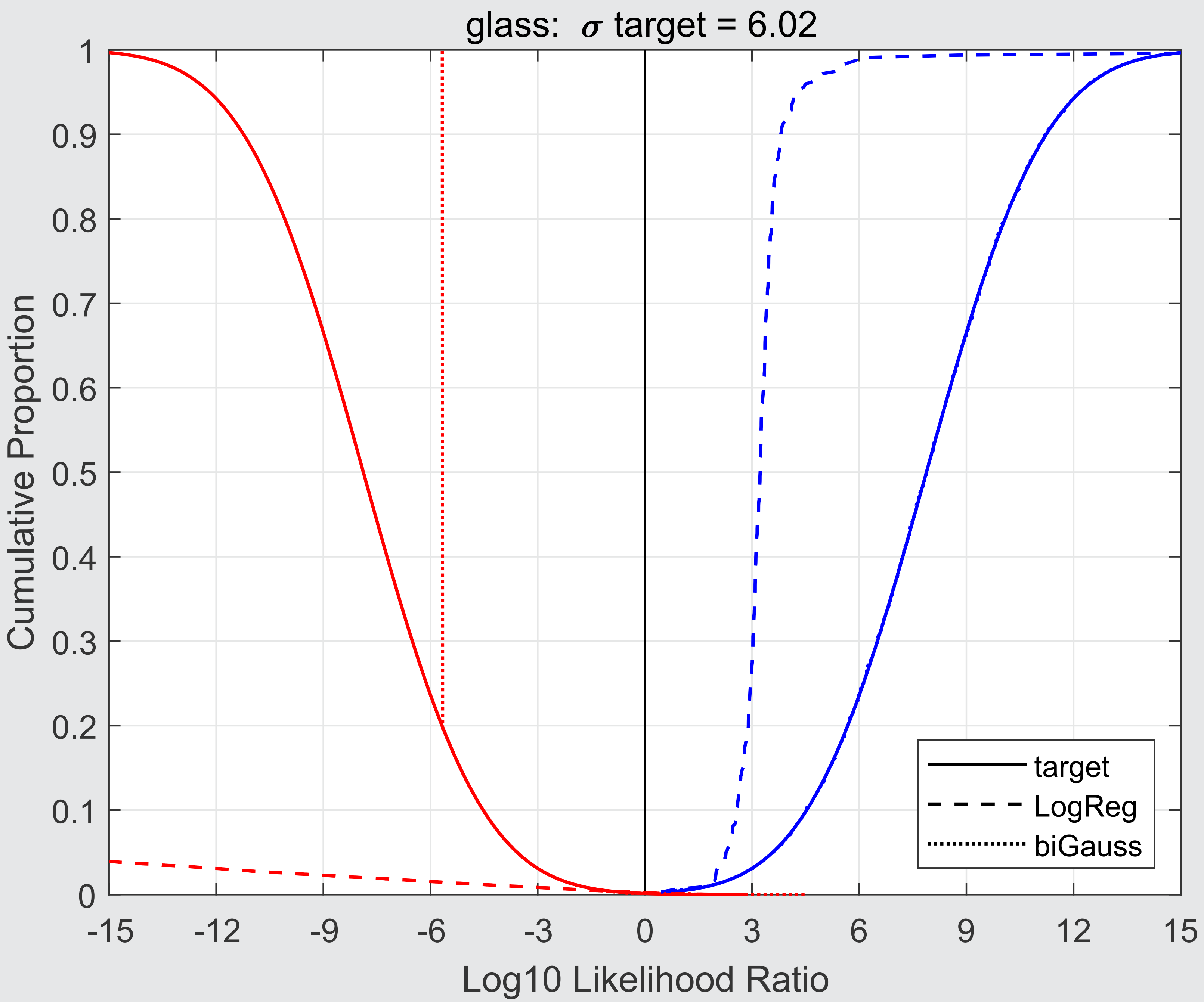
Real data: glass

- Probability density functions
  - cross-validated



Real data: glass

- Tippett plots
  - cross-validated



Conclusion

- If likelihood-ratio values are not calibrated:
    - their absolute values cannot be interpreted
    - they cannot be used in Bayes' theorem to update prior odds to posterior odds

- Logistic-regression calibration produces results that can deviate quite far from perfect calibration.

- Bi-Gaussininazed calibration produces results that are close to perfect calibration.

Thank You

The likelihood ratio

of

the likelihood ratio

is

the likelihood ratio