Making decisions with biometric systems: the usefulness of a Bayesian perspective

A. Nautsch[⋆], D. Ramos Castro[†], J. González Rodríguez[†], Christian Rathgeb[⋆], Christoph Busch[⋆]

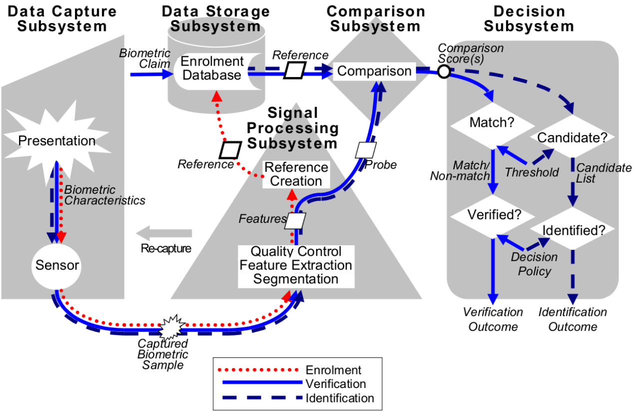
[⋆]Hochschule Darmstadt, CRISP, CASED, da/sec Security Research Group
[†]Universidad Autónoma de Madrid, ATVS Biometric Recognition Group

NIST IBPC'16, Gaithersburg, 03.05.2016

Outline

1. Decision Frameworks in Biometrics and Forensics
2. Bayesian Method: making good decisions
3. Metrics, operating points and examples
4. Conclusion

Biometric Systems in ISO/IEC JTC1 SC37 SD11



⇒ Note: separate decision subsystem

Making Decisions with Biometric Systems

Decisions are involved in most applications of biometric systems

- Access control
  Accepted-rejected decision

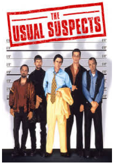
- Forensic Investigation
  Decide the k list to investigate
  e.g., AFIS
- Intelligence
  Decide where to establish
  relevant links in a database
- Forensic Evaluation
  Commnunicate for the court
  to decide a veredict

Making Decisions with Biometric Systems

Decisions are involved in most applications of biometric systems

- Access control
  Accepted-rejected decision

- Forensic Investigation
  Decide the k list to investigate
  e.g., AFIS

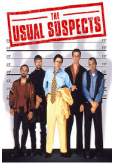


- Intelligence
  Decide where to establish
  relevant links in a database
- Forensic Evaluation
  Commnunicate for the court
  to decide a veredict

Making Decisions with Biometric Systems

Decisions are involved in most applications of biometric systems

- Access control
  Accepted-rejected decision



- Forensic Investigation
  Decide the k list to investigate
  e.g., AFIS

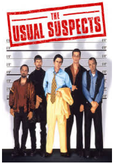
- Intelligence
  Decide where to establish
  relevant links in a database

- Forensic Evaluation
  Commnunicate for the court
  to decide a veredict

Making Decisions with Biometric Systems

Decisions are involved in most applications of biometric systems

- Access control
  Accepted-rejected decision



- Forensic Investigation
  Decide the k list to investigate
  e.g., AFIS

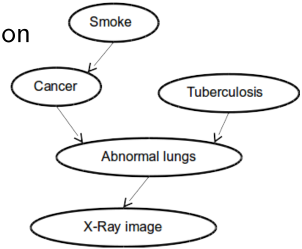
- Intelligence
  Decide where to establish
  relevant links in a database

- Forensic Evaluation
  Commnunicate for the court
  to decide a veredict

Making Decisions with Biometric Systems

- Decision maker faces multiple sources of information
  Biometric system is one of them, but also ...
  - Prior knowledge about users/impostors/suspects
  - Other evidence from other biometric systems
  - ...

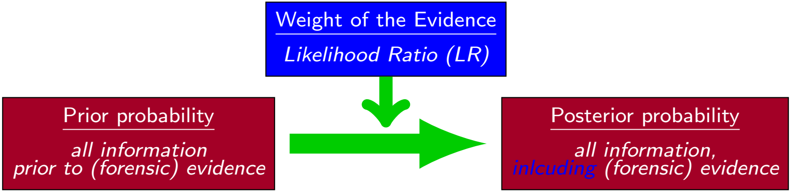
- Decisions must consider all that information
  - Formalizing decision framework helps
  - Especially in complex problems
  - Example: medical diagnosis support



Bayesian Decisions with Biometric Systems

- A proposal: Bayesian decision theory
  - Decisions are made based on posterior probabilities
  - Considering all the relevant information available
  - Updating strategy: likelihood ratios (LR)

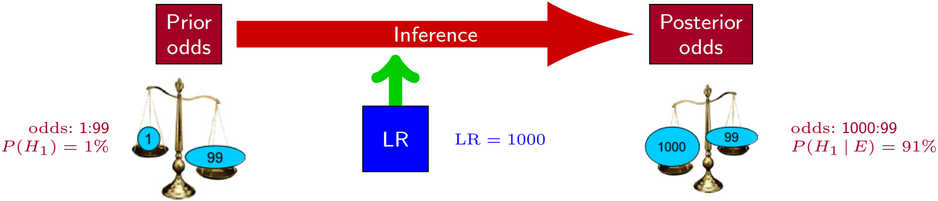
Example biometrics systems in forensic evaluation of the evidence



[1] I. Evett: *Towards a uniform framework for Reporting opinions in forensic science Casework*, Science and Justice, 1998.

Value of Evidence: Likelihood Ratio (LR)

- Two-class (H_1, H_2) decision framework
- Likelihood Ratio: probabilistic value of the evidence, also: the ratio of posterior to prior odds

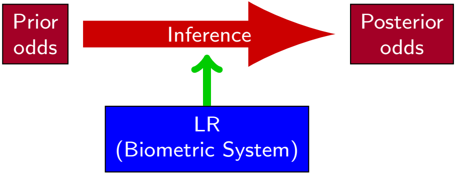


Prior odds — LR — Posterior odds

$$\boxed{\frac{P(H_1)}{P(H_2)}} \times \boxed{\frac{P(E \mid H_1)}{P(E \mid H_2)}} = \boxed{\frac{P(H_1 \mid E)}{P(H_2 \mid E)}}$$

Decisions Using Biometric Systems

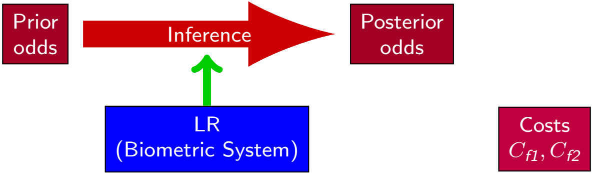
- Binary classes (hypotheses): H_1 and H_2
- Inference
  - Prior probability, before knowing the biometric system outcome
  - Posterior probability, after the biometric system outcome
  - LR is the value of the biometric evidence
  - $\Rightarrow$ Changes prior odds into posterior odds



$$\frac{P(H_1 \mid E)}{P(H_2 \mid E)}$$

Decisions Using Biometric Systems

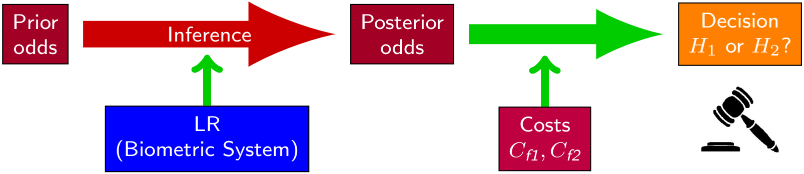
- Costs: Penalty of making a wrong decision towards H_1 (C_{f1}) or H_2 (C_{f2}).
- Can be different — example in access control:
  - is it better to accept an impostor (cost C_{f1})
  - or to reject a genuine user (cost C_{f2})?



Decisions Using Biometric Systems

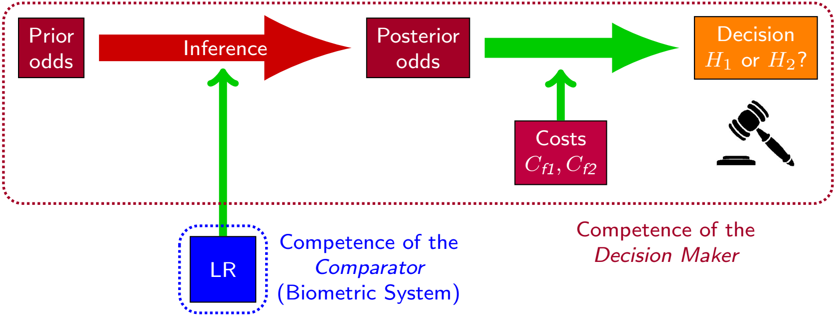
- Decision: Minimum-risk decision
  i.e.: minimum mean cost
- Decision rule considers
  - Posterior odds
  - Costs

$$\boxed{P(H_1 \mid E)\, C_{f1} \gtrless P(H_2 \mid E)\, C_{f2}}$$



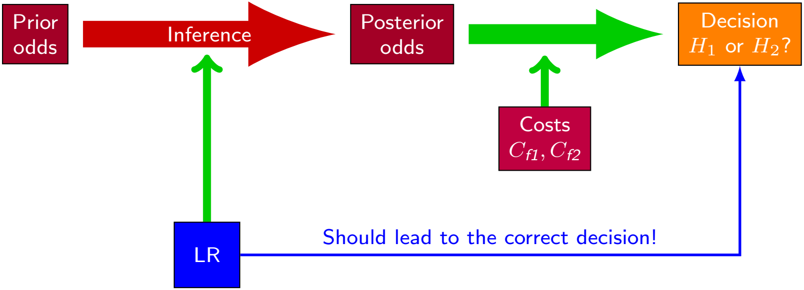
Decision Process: Competences

- Total separation between
  - The comparator (biometric system outputting a LR)
  - The decision maker (depends on the application)



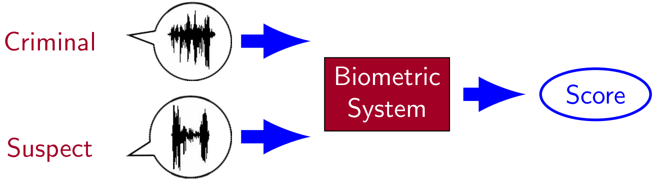
Decision Process: Consequences

- Duty of the biometric systems:
  yielding LR values that lead to the correct decisions
  - The LR should support H_1 when H_1 is actually true
  - The LR should support H_2 when H_2 is actually true
- LR values must be calibrated, which leads to better decisions



Biometric Systems

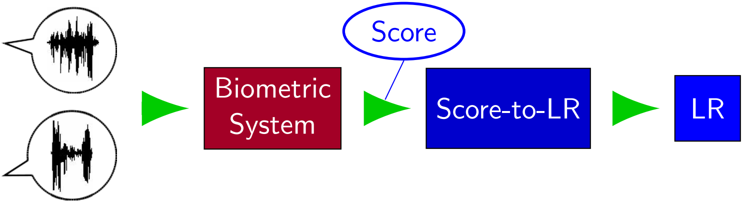
- Score-based architecture
  - Widely extended
  - Especially in black-box implementations (COTS)



- Score: in general the only output of the system
  - It may not be directly interpretable as a likelihood ratio
  - Depends on its calibration performance

LR-Based Computation with Biometric Systems

- A further stage is necessary: score-to-LR transformation



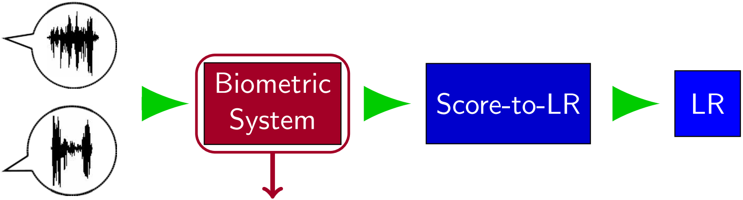
- Objective:
  output discriminating scores
  - Score-based architecture
  - Improve ROC/DET curves
- Objective:
  transforming the score into a meaningful LR
  $\Rightarrow$ Calibration of LRs [2,3]

[2] N. Brümmer and J. du Preez: *Application Independent Evaluation of Speaker Detection*, Computer Speech and Language, 2006.

[3] D. Ramos and J. González Rodríguez: *Reliable support: Measuring calibration of likelihood ratios*, Forensic Science International, 2013.

LR-Based Computation with Biometric Systems

- A further stage is necessary: score-to-LR transformation



- Objective:
  output discriminating scores
  - Score-based architecture
  - Improve ROC/DET curves

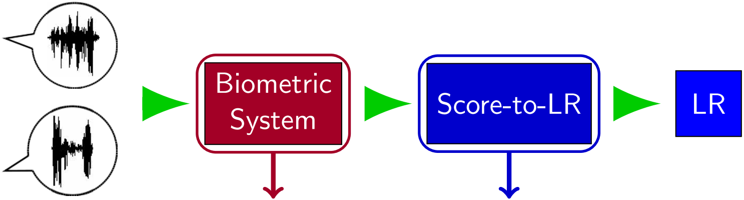
- Objective:
  transforming the score into a meaningful LR
  ⇒ Calibration of LRs [2,3]

[2] N. Brümmer and J. du Preez: *Application Independent Evaluation of Speaker Detection*, Computer Speech and Language, 2006.

[3] D. Ramos and J. González Rodríguez: *Reliable support: Measuring calibration of likelihood ratios*, Forensic Science International, 2013.

LR-Based Computation with Biometric Systems

- A further stage is necessary: score-to-LR transformation



- Objective:
  output discriminating scores
  - Score-based architecture
  - Improve ROC/DET curves
- Objective:
  transforming the score into a meaningful LR
  $\Rightarrow$ Calibration of LRs [2,3]

[2] N. Brümmer and J. du Preez: *Application Independent Evaluation of Speaker Detection*, Computer Speech and Language, 2006.

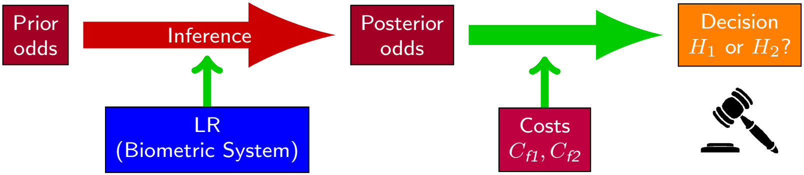
[3] D. Ramos and J. González Rodríguez: *Reliable support: Measuring calibration of likelihood ratios*, Forensic Science International, 2013.

Bayesian Decisions: Advantages

- Competences of the biometric system are delimited:
  - Biometric system: comparator
  - Decision maker: final decision considering all the information
  - Separation of roles: important in some fields (e.g. forensics)!
- Information is integrated formally

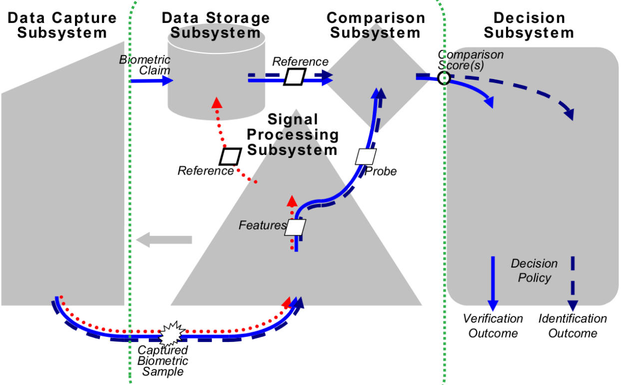
  $\Rightarrow$ LR into a probabilistic framework
- LR computation: great experience in other fields

  $\Rightarrow$ Example: forensic biometrics



Revisiting ISO/IEC JTC1 SC37 SD11

FNMR, FMR $\mapsto$ DET

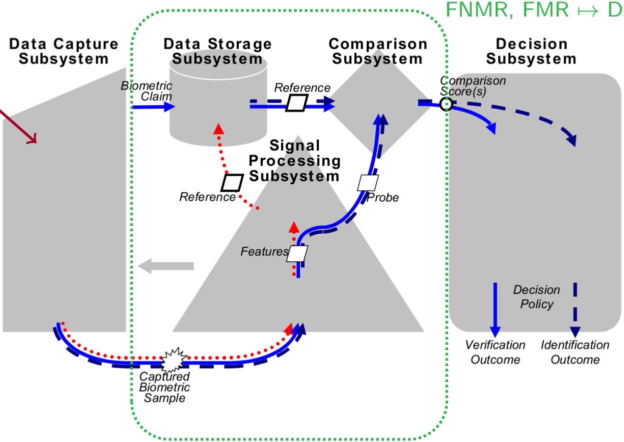


Revisiting ISO/IEC JTC1 SC37 SD11

$$\frac{P(H_1)}{P(H_2)} = \frac{\pi}{1-\pi}$$

$\Rightarrow \pi$

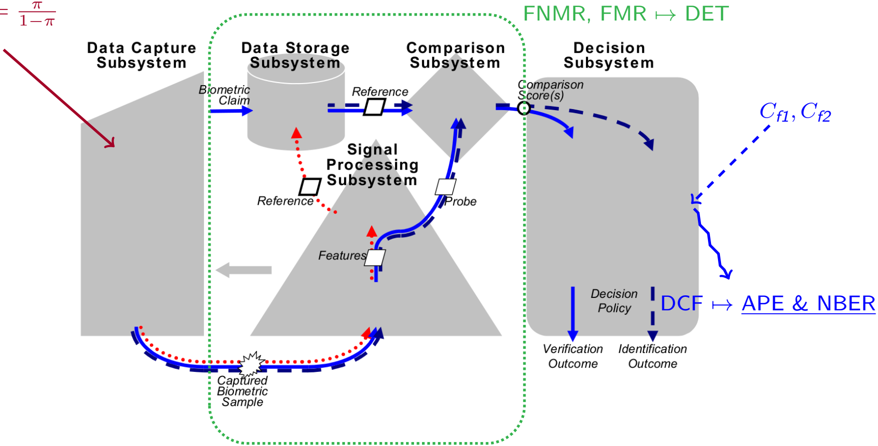
FNMR, FMR $\mapsto$ DET



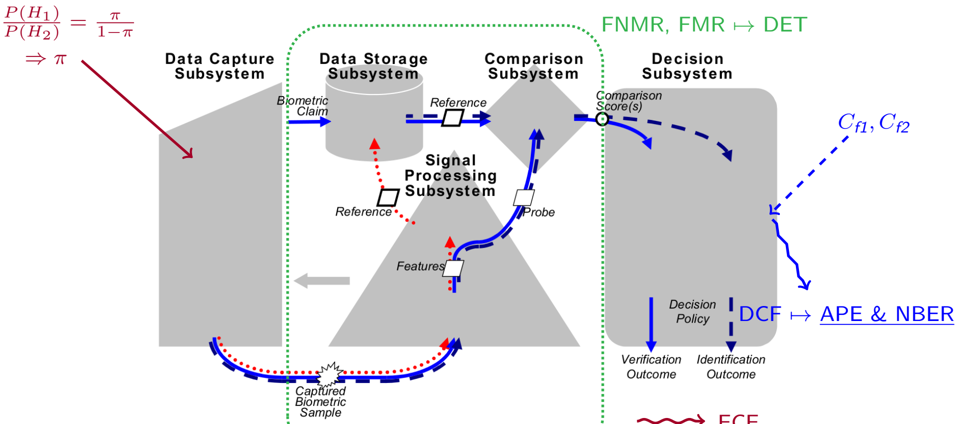
Revisiting ISO/IEC JTC1 SC37 SD11

$\frac{P(H_1)}{P(H_2)} = \frac{\pi}{1-\pi}$

$\Rightarrow \pi$



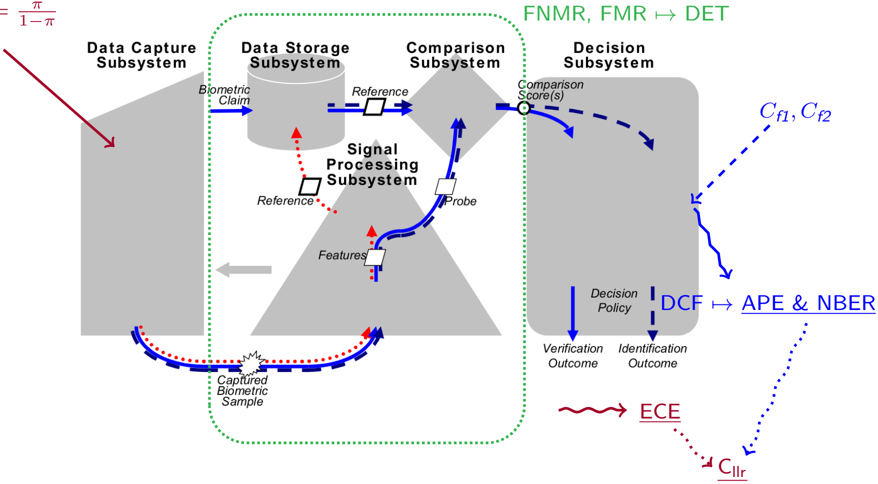
Revisiting ISO/IEC JTC1 SC37 SD11



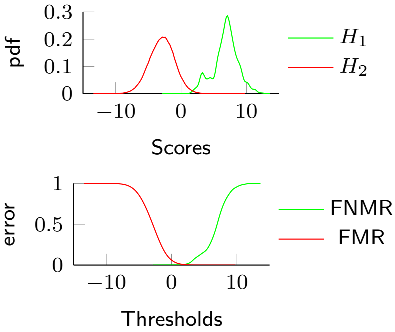
Revisiting ISO/IEC JTC1 SC37 SD11

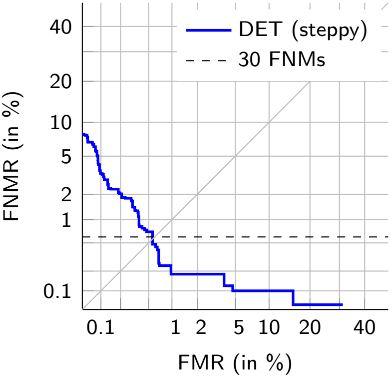
$\frac{P(H_1)}{P(H_2)} = \frac{\pi}{1-\pi}$

$\Rightarrow \pi$



Detection Error Trade-off (DET) diagrams





[4] N. Brümmer and E. de Villers: *The BOSARIS Toolkit User Guide: Theory, Algorithms and Code for Binary Classifier Score Processing*, Tech.Rep. AGNITIO Research, 2011.

From Bayesian Decisions to Cost Functions

- Bayes theorem

Prior odds, LR, Posterior odds

$$\boxed{\frac{P(H_1)}{P(H_2)}} \times \boxed{\frac{P(E\,|\,H_1)}{P(E\,|\,H_2)}} = \boxed{\frac{P(H_1\,|\,E)}{P(H_2\,|\,E)}}$$

- Decision rule

$$\boxed{P(H_1\,|\,E)\,C_{f1} \gtrless P(H_2\,|\,E)\,C_{f2}}$$

$$\Leftrightarrow \frac{P(H_1\,|\,E)}{P(H_2\,|\,E)} \gtrless \frac{C_{f2}}{C_{f1}}$$

- Bayesian threshold η for Log-LRs (LLRs) by posterior odds

$$\eta = \log \frac{C_{f2}}{C_{f1}} - \log \frac{P(H_1)}{P(H_2)} \gtrless \text{LLR}$$

From Bayesian Decisions to Cost Functions

- Bayes theorem

Prior odds LR Posterior odds

$$\frac{P(H_1)}{P(H_2)} \times \frac{P(E \mid H_1)}{P(E \mid H_2)} = \frac{P(H_1 \mid E)}{P(H_2 \mid E)}$$

- Decision rule

$$\boxed{P(H_1 \mid E)\, C_{f1} \gtrless P(H_2 \mid E)\, C_{f2}}$$

$$\Leftrightarrow \frac{P(H_1 \mid E)}{P(H_2 \mid E)} \gtrless \frac{C_{f2}}{C_{f1}}$$

- Bayesian threshold η for Log-LRs (LLRs) by posterior odds

$$\eta = \log \frac{C_{f2}}{C_{f1}} - \log \frac{P(H_1)}{P(H_2)} \gtrless \mathrm{LLR}$$

From Bayesian Decisions to Cost Functions

- Bayes theorem

$$\underbrace{\boxed{\frac{P(H_1)}{P(H_2)}}}_{\text{Prior odds}} \times \underbrace{\boxed{\frac{P(E \mid H_1)}{P(E \mid H_2)}}}_{\text{LR}} = \underbrace{\boxed{\frac{P(H_1 \mid E)}{P(H_2 \mid E)}}}_{\text{Posterior odds}}$$

- Decision rule

$$\boxed{P(H_1 \mid E)\, C_{f1} \gtrless P(H_2 \mid E)\, C_{f2}}$$

$$\Leftrightarrow \frac{P(H_1 \mid E)}{P(H_2 \mid E)} \gtrless \frac{C_{f2}}{C_{f1}}$$

- Bayesian threshold η for Log-LRs (LLRs) by posterior odds

$$\eta = \log \frac{C_{f2}}{C_{f1}} - \log \frac{P(H_1)}{P(H_2)} \gtrless \mathrm{LLR}$$

From Bayesian Decisions to Cost Functions

- Bayesian error rate: Decision Cost Function (DCF)

$$\mathrm{DCF}(P(H_1),\, P(H_2),\, C_{f1},\, C_{f2}) = P(H_1)\,\mathrm{FNMR}(\eta)\,C_{f1} + P(H_2)\,\mathrm{FMR}(\eta)\,C_{f2}$$

$$\eta = \log \frac{C_{f2}}{C_{f1}} - \log \frac{P(H_1)}{P(H_2)}$$

- Simplifying the operating point $(P(H_1),\, P(H_2),\, C_{f1},\, C_{f2}) \mapsto \tilde{\pi}$
  1. Mutually exclusive priors: $\log \frac{P(H_1)}{P(H_2)} = \log \frac{\pi}{1-\pi} = \operatorname{logit} \pi$

     $$\mathrm{DCF}(\pi, C_{f1}, C_{f2}) = \pi\,\mathrm{FNMR}(\eta)\,C_{f1} + (1-\pi)\,\mathrm{FMR}(\eta)\,C_{f2}$$

  2. Introducing an *effective prior*: $\tilde{\pi} = \frac{\pi\,C_{f1}}{\pi\,C_{f1} + (1-\pi)\,C_{f2}}$

     $$\mathrm{DCF}(\tilde{\pi}) = \tilde{\pi}\,\mathrm{FNMR}(\eta) + (1-\tilde{\pi})\,\mathrm{FMR}(\eta) = \mathrm{DCF}(\pi, 1, 1)$$

     $$\eta = -\operatorname{logit} \tilde{\pi}$$

$\Rightarrow$ meaningful LLR operating points: $\tilde{\pi}$ or η

[4] N. Brümmer and E. de Villiers: *The BOSARIS Toolkit User Guide: Theory, Algorithms and Code for Binary Classifier Score*, Tech.Rep., AGNITIO Research, December 2011.

From Bayesian Decisions to Cost Functions

- Bayesian error rate: Decision Cost Function (DCF)

$$\mathrm{DCF}(P(H_1), P(H_2), C_{f1}, C_{f2}) = P(H_1)\,\mathrm{FNMR}(\eta)\,C_{f1} + P(H_2)\,\mathrm{FMR}(\eta)\,C_{f2}$$

$$\eta = \log \frac{C_{f2}}{C_{f1}} - \log \frac{P(H_1)}{P(H_2)}$$

- Simplifying the operating point $(P(H_1), P(H_2), C_{f1}, C_{f2}) \mapsto \tilde{\pi}$
  1. Mutually exclusive priors: $\log \frac{P(H_1)}{P(H_2)} = \log \frac{\pi}{1-\pi} = \operatorname{logit} \pi$

     $$\mathrm{DCF}(\pi, C_{f1}, C_{f2}) = \pi\,\mathrm{FNMR}(\eta)\,C_{f1} + (1-\pi)\,\mathrm{FMR}(\eta)\,C_{f2}$$

  2. Introducing an *effective prior*: $\tilde{\pi} = \frac{\pi\, C_{f1}}{\pi\, C_{f1} + (1-\pi)\, C_{f2}}$

     $$\mathrm{DCF}(\tilde{\pi}) = \tilde{\pi}\,\mathrm{FNMR}(\eta) + (1-\tilde{\pi})\,\mathrm{FMR}(\eta) = \mathrm{DCF}(\pi, 1, 1)$$

     $$\eta = -\operatorname{logit} \tilde{\pi}$$

$\Rightarrow$ meaningful LLR operating points: $\tilde{\pi}$ or η

[4] N. Brümmer and E. de Villiers: *The BOSARIS Toolkit User Guide: Theory, Algorithms and Code for Binary Classifier Score*, Tech.Rep., AGNITIO Research, December 2011.

From Bayesian Decisions to Cost Functions

- Bayesian error rate: Decision Cost Function (DCF)

$$\mathrm{DCF}(P(H_1), P(H_2), C_{f1}, C_{f2}) = P(H_1)\,\mathrm{FNMR}(\eta)\,C_{f1} + P(H_2)\,\mathrm{FMR}(\eta)\,C_{f2}$$

$$\eta = \log \frac{C_{f2}}{C_{f1}} - \log \frac{P(H_1)}{P(H_2)}$$

- Simplifying the operating point $(P(H_1), P(H_2), C_{f1}, C_{f2}) \mapsto \tilde{\pi}$
  1. Mutually exclusive priors: $\log \frac{P(H_1)}{P(H_2)} = \log \frac{\pi}{1-\pi} = \operatorname{logit} \pi$

     $$\mathrm{DCF}(\pi, C_{f1}, C_{f2}) = \pi\,\mathrm{FNMR}(\eta)\,C_{f1} + (1-\pi)\,\mathrm{FMR}(\eta)\,C_{f2}$$

  2. Introducing an *effective prior*: $\tilde{\pi} = \frac{\pi\,C_{f1}}{\pi\,C_{f1} + (1-\pi)\,C_{f2}}$

     $$\mathrm{DCF}(\tilde{\pi}) = \tilde{\pi}\,\mathrm{FNMR}(\eta) + (1-\tilde{\pi})\,\mathrm{FMR}(\eta) = \mathrm{DCF}(\pi, 1, 1)$$

     $$\eta = -\operatorname{logit} \tilde{\pi}$$

$\Rightarrow$ meaningful LLR operating points: $\tilde{\pi}$ or η

[4] N. Brümmer and E. de Villiers: *The BOSARIS Toolkit User Guide: Theory, Algorithms and Code for Binary Classifier Score*, Tech.Rep., AGNITIO Research, December 2011.

From Bayesian Decisions to Cost Functions

- Bayesian error rate: Decision Cost Function (DCF)

$$\mathrm{DCF}(P(H_1),\, P(H_2),\, C_{f1},\, C_{f2}) = P(H_1)\,\mathrm{FNMR}(\eta)\,C_{f1} + P(H_2)\,\mathrm{FMR}(\eta)\,C_{f2}$$

$$\eta = \log\frac{C_{f2}}{C_{f1}} - \log\frac{P(H_1)}{P(H_2)}$$

- Simplifying the operating point $(P(H_1),\, P(H_2),\, C_{f1},\, C_{f2}) \mapsto \tilde{\pi}$
  1. Mutually exclusive priors: $\log\frac{P(H_1)}{P(H_2)} = \log\frac{\pi}{1-\pi} = \operatorname{logit}\pi$

$$\mathrm{DCF}(\pi, C_{f1}, C_{f2}) = \pi\,\mathrm{FNMR}(\eta)\,C_{f1} + (1-\pi)\,\mathrm{FMR}(\eta)\,C_{f2}$$

  2. Introducing an *effective prior*: $\tilde{\pi} = \frac{\pi\,C_{f1}}{\pi\,C_{f1} + (1-\pi)\,C_{f2}}$

$$\mathrm{DCF}(\tilde{\pi}) = \tilde{\pi}\,\mathrm{FNMR}(\eta) + (1-\tilde{\pi})\,\mathrm{FMR}(\eta) = \mathrm{DCF}(\pi, 1, 1)$$

$$\eta = -\operatorname{logit}\tilde{\pi}$$

$\Rightarrow$ meaningful LLR operating points: $\tilde{\pi}$ or η

[4] N. Brümmer and E. de Villiers: *The BOSARIS Toolkit User Guide: Theory, Algorithms and Code for Binary Classifier Score*, Tech.Rep., AGNITIO Research, December 2011.

From Bayesian Decisions to Cost Functions

- Bayesian error rate: Decision Cost Function (DCF)

$$\mathrm{DCF}(P(H_1),\, P(H_2),\, C_{f1},\, C_{f2}) = P(H_1)\,\mathrm{FNMR}(\eta)\, C_{f1} + P(H_2)\,\mathrm{FMR}(\eta)\, C_{f2}$$

$$\eta = \log \frac{C_{f2}}{C_{f1}} - \log \frac{P(H_1)}{P(H_2)}$$

- Simplifying the operating point $(P(H_1),\, P(H_2),\, C_{f1},\, C_{f2}) \mapsto \tilde{\pi}$
  1. Mutually exclusive priors: $\log \frac{P(H_1)}{P(H_2)} = \log \frac{\pi}{1-\pi} = \operatorname{logit} \pi$

     $$\mathrm{DCF}(\pi, C_{f1}, C_{f2}) = \pi\,\mathrm{FNMR}(\eta)\, C_{f1} + (1-\pi)\,\mathrm{FMR}(\eta)\, C_{f2}$$

  2. Introducing an *effective prior*: $\tilde{\pi} = \frac{\pi\, C_{f1}}{\pi\, C_{f1} + (1-\pi)\, C_{f2}}$

     $$\mathrm{DCF}(\tilde{\pi}) = \tilde{\pi}\,\mathrm{FNMR}(\eta) + (1-\tilde{\pi})\,\mathrm{FMR}(\eta) = \mathrm{DCF}(\pi, 1, 1)$$

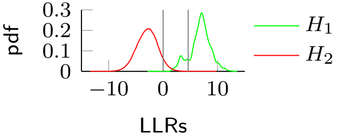
     $$\eta = -\operatorname{logit} \tilde{\pi}$$

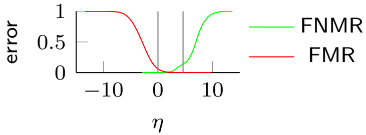
$\Rightarrow$ meaningful LLR operating points: $\tilde{\pi}$ or η

[4] N. Brümmer and E. de Villiers: *The BOSARIS Toolkit User Guide: Theory, Algorithms and Code for Binary Classifier Score*, Tech.Rep., AGNITIO Research, December 2011.

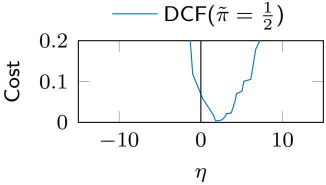
Example on Decision Cost Functions (DCFs)

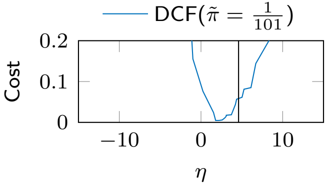
- Speaker recognition ivec/PLDA scores (I4U list/NIST SRE'12)





- Example: DCF(1:1, $\eta = 0$) vs. DCF(1:100, $\eta \approx 4.6$)

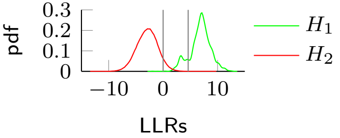


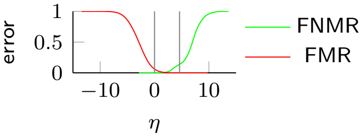


⇒ actual vs. minimum DCF: calibration loss
⇒ LLR meaning: aligning scores for Bayesian support

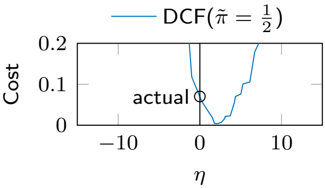
Example on Decision Cost Functions (DCFs)

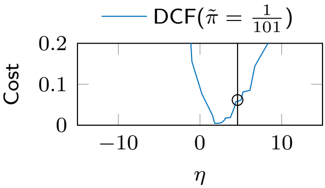
- Speaker recognition ivec/PLDA scores (I4U list/NIST SRE'12)





- Example: DCF(1:1, $\eta = 0$) vs. DCF(1:100, $\eta \approx 4.6$)



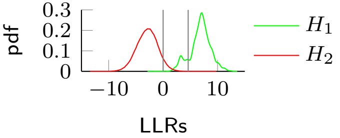


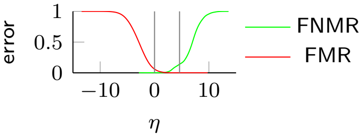
⇒ actual vs. minimum DCF: calibration loss

⇒ LLR meaning: aligning scores for Bayesian support

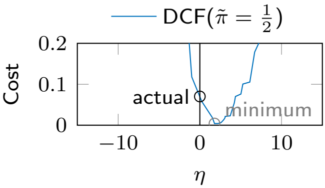
Example on Decision Cost Functions (DCFs)

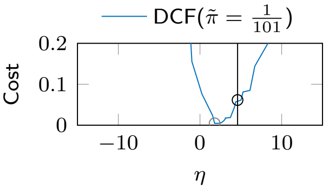
- Speaker recognition ivec/PLDA scores (I4U list/NIST SRE'12)





- Example: DCF(1:1, $\eta = 0$) vs. DCF(1:100, $\eta \approx 4.6$)



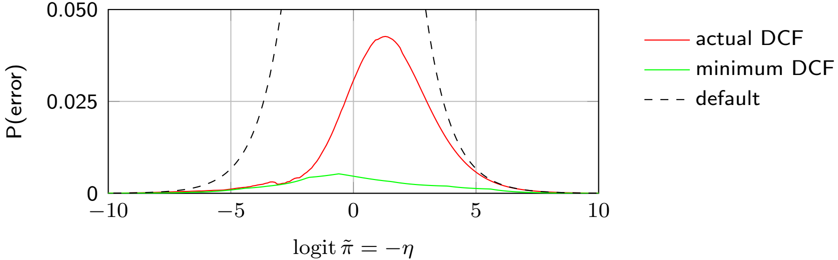


$\Rightarrow$ actual vs. minimum DCF: calibration loss
$\Rightarrow$ LLR meaning: aligning scores for Bayesian support

Visualizing DCFs

- Applied Probability of Error (APE) curve
  - Simulating DCFs on multiple operating points
  - default: all LLRs = 0, i.e.: $\mathrm{DCF} = \tilde{\pi} + (1 - \tilde{\pi})$
  - Area-under-APE: cost of LLR scores

    $\Rightarrow$ Goodness of LLRs: C_{llr}



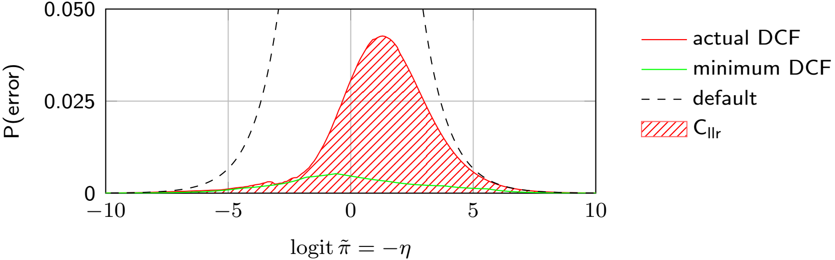
[5] N. Brümmer: *FoCal: Tools for Fusion and Calibration of automatic speaker detection systems*, Tech.Rep., 2005.

[6] D.A. van Leeuwen and N. Brümmer: *An Introduction to Application-Independent Evaluation of Speaker Recognition Systems*, Speaker Classification I: Fundamentals, Features, and Methods, Springer LNCS, 2007.

Visualizing DCFs

- Applied Probability of Error (APE) curve
  - Simulating DCFs on multiple operating points
  - default: all LLRs = 0, i.e.: $\mathrm{DCF} = \tilde{\pi} + (1 - \tilde{\pi})$
  - Area-under-APE: cost of LLR scores

    $\Rightarrow$ Goodness of LLRs: $\mathrm{C_{llr}}$



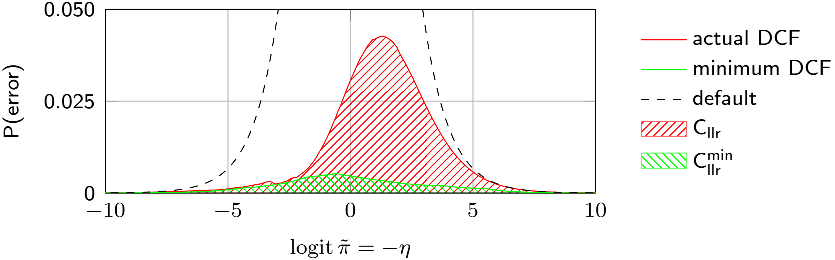
[5] N. Brümmer: *FoCal: Tools for Fusion and Calibration of automatic speaker detection systems*, Tech.Rep., 2005.

[6] D.A. van Leeuwen and N. Brümmer: *An Introduction to Application-Independent Evaluation of Speaker Recognition Systems*, Speaker Classification I: Fundamentals, Features, and Methods, Springer LNCS, 2007.

Visualizing DCFs

- Applied Probability of Error (APE) curve
  - Simulating DCFs on multiple operating points
  - default: all LLRs = 0, i.e.: $\mathrm{DCF} = \tilde{\pi} + (1 - \tilde{\pi})$
  - Area-under-APE: cost of LLR scores

  $\Rightarrow$ Goodness of LLRs: C_{llr}



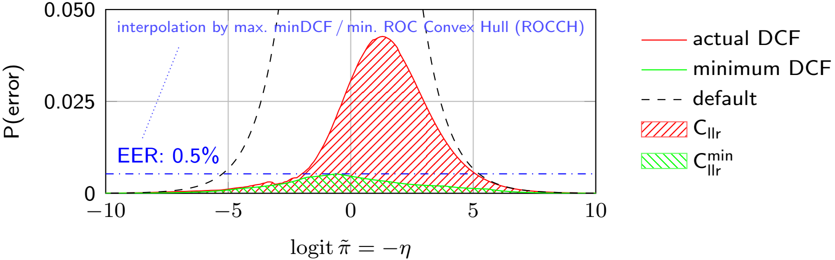
[5] N. Brümmer: *FoCal: Tools for Fusion and Calibration of automatic speaker detection systems*, Tech.Rep., 2005.

[6] D.A. van Leeuwen and N. Brümmer: *An Introduction to Application-Independent Evaluation of Speaker Recognition Systems*, Speaker Classification I: Fundamentals, Features, and Methods, Springer LNCS, 2007.

Visualizing DCFs

- Applied Probability of Error (APE) curve
  - Simulating DCFs on multiple operating points
  - default: all LLRs $= 0$, i.e.: $\mathrm{DCF} = \tilde{\pi} + (1 - \tilde{\pi})$
  - Area-under-APE: cost of LLR scores

    $\Rightarrow$ Goodness of LLRs: $\mathrm{C}_{\mathrm{llr}}$

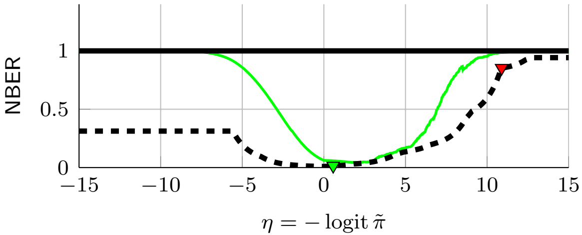


[5] N. Brümmer: *FoCal: Tools for Fusion and Calibration of automatic speaker detection systems*, Tech.Rep., 2005.

[6] D.A. van Leeuwen and N. Brümmer: *An Introduction to Application-Independent Evaluation of Speaker Recognition Systems*, Speaker Classification I: Fundamentals, Features, and Methods, Springer LNCS, 2007.

Normalized Bayesian Error Rate (NBER)

- APE-plot visually misleading on error impact
  - EER operating point: lots of scores to mismatch
  - FMR1000 operating point: few scores to mismatch
- Normalizing by default performance
  - $\Rightarrow$ wider range of operating points can be compared



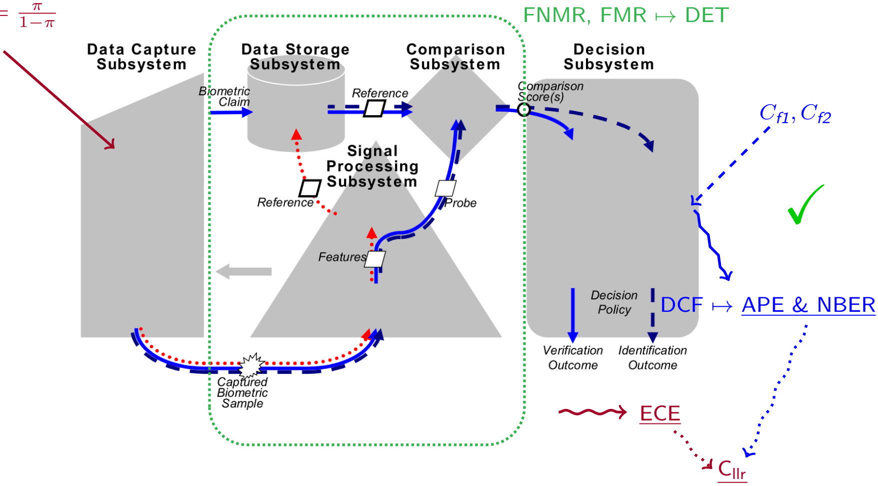
[4] N. Brümmer and E. de Villiers: *The BOSARIS Toolkit User Guide: Theory, Algorithms and Code for Binary Classifier Score*, Tech.Rep., AGNITIO Research, December 2011.

Note: in the BOSARIS toolkit, the x-axis is swapped, i.e.: depicting purely the effective prior.

Revisiting ISO/IEC JTC1 SC37 SD11

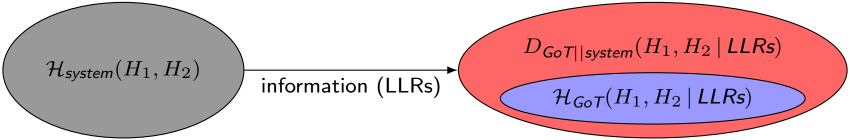
$\frac{P(H_1)}{P(H_2)} = \frac{\pi}{1-\pi}$

$\Rightarrow \pi$



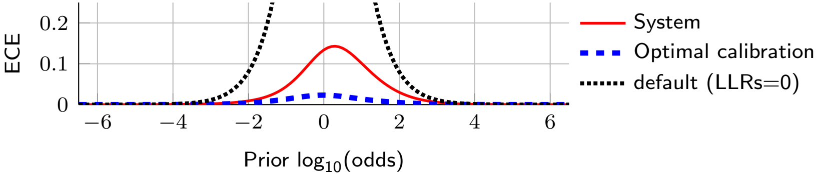
Empirical Cross-Entropy (ECE)

- Objective measure of performance
- Motivation by Information Theory
  - Prior entropy $\xrightarrow[\text{Information gain}]{\text{Evidence}}$ Posterior entropy
  - Divergence of system to Grund-of-Truth (GoT)
  - ECE: approximating Kullback-Leibler divergence $D_{GoT||system}$



Empirical Cross-Entropy (ECE)

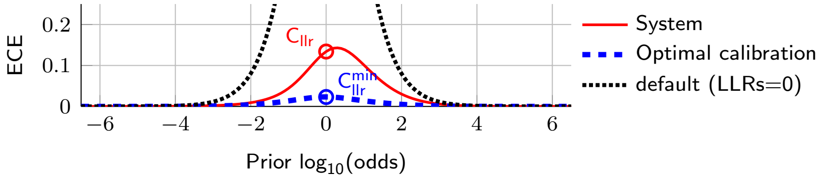
- We expect the reference, but obtain the system's LLRs
- Measuring performance of LR in terms of uncertainty
  - The lower the better
    Calibration loss: overall performance ⇔ discriminating power
  - C_{llr} at log(odds) = 0 ⇒ no information on H_1/H_2 prior



[7] D. Ramos Castro and J. González Rodríguez: *Cross-entropy Analysis of the Information in Forensic Speaker Recognition*, Odyssey, 2008.

Empirical Cross-Entropy (ECE)

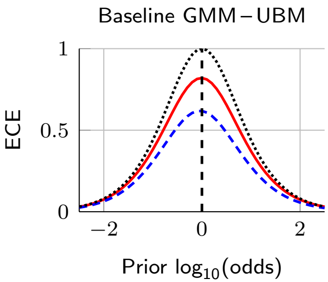
- We expect the reference, but obtain the system's LLRs
- Measuring performance of LR in terms of uncertainty
  - The lower the better
    Calibration loss: overall performance ⇔ discriminating power
  - C_{llr} at log(odds) $= 0$ $\Rightarrow$ no information on H_1/H_2 prior

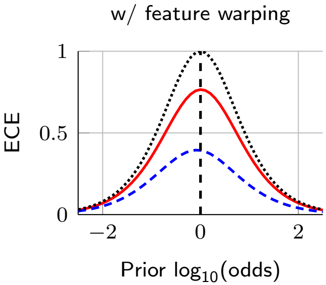


[7] D. Ramos Castro and J. González Rodríguez: *Cross-entropy Analysis of the Information in Forensic Speaker Recognition*, Odyssey, 2008.

Examples

- Signature recognition [8]
  - Performance of feature space normalization
  - Simulation of application-independent decision performances

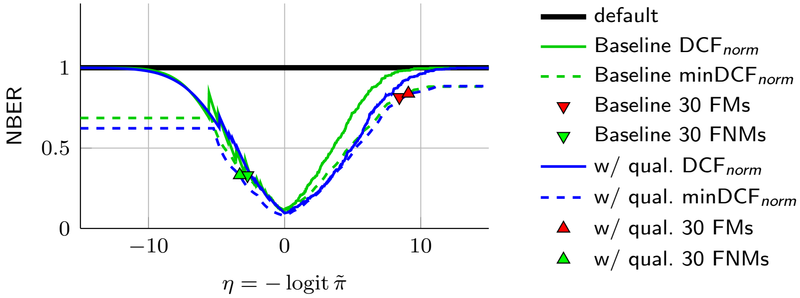




[8] A. Nautsch, C. Rathgeb, C. Busch: *Bridging Gaps: An Application of Feature Warping to Online Signature Verification*, ICCST, 2014.

Examples

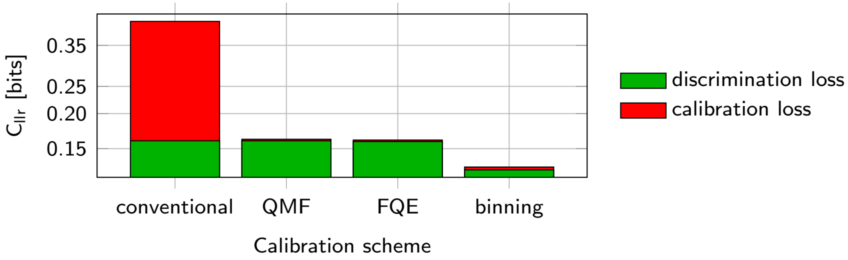
- Speaker recognition [9]
  - Overview of application-dependent decision costs in 10 dB/10 s
  - Conventional score normalization vs. quality-based



[9] A. Nautsch, R. Saeidi, C. Rathgeb, C. Busch: *Analysis of mutual duration and noise effects in speaker recognition: benefits of condition-matched cohort selection in score normalization*, Interspeech, 2015.

Examples

- Speaker recognition [10]
  - Examining calibration schemes in 55 quality conditions
  - Discrimination vs. calibration loss on 55-pooled
  - Goal: approx. binning performance, avoiding binning



[10] A. Nautsch, R. Saeidi, C. Rathgeb, C. Busch: *Robustness of Quality-based Score Calibration of Speaker Recognition Systems with respect to low-SNR and short-duration conditions*, Odyssey, 2016. *(to appear)*

Examples

- Recurring challenges in biometrics
  - NIST Speaker Recognition Evaluation (SRE)
    $\Rightarrow$ DCFs (since 1996) & C_{llr} (since 2006)
  - ICDAR Competition on Signature Verification and Writer Identification (SigWIcomp)
    $\Rightarrow$ C_{llr} & $C_{\mathrm{llr}}^{\mathrm{min}}$ (both since 2011)

- Non-biometric forensics [11]
  - Glass objects
  - Car paints
  - Inks

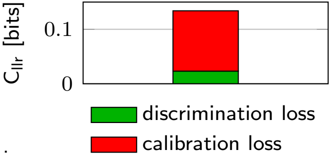
[11] G. Zadora, A. Martyna, D. Ramos, C. Aitken: *Statistical Analysis in Forensic Science: Evidential Values of Multivariate Physicochemical Data*, John Wiley and Sons, 2014.

Summary

- Bayesian decision framework
  - Bayes theorem & decision rule enploying costs
  - Biometric systems: generator of Bayesian support (LLRs)
  - Decisions by posterior knowledge of priors and LLR score
- Score-to-LLR calibration: meaningful LLRs
  - Necessary step, requiring a calibration data set
  - Essential for validation/accredetation
- Performance reporting
  - Decoupled decision policy
  - APE curves
  - NBER diagrams
  - ECE plots
  - Scalars: actDCF, minDCF, C_{llr} & C_{llr}^{min}

Summary

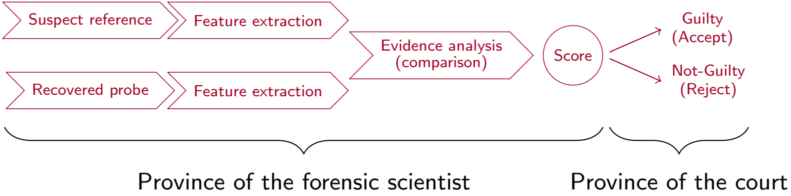
- Bayesian decision framework
  - Bayes theorem & decision rule enploying costs
  - Biometric systems: generator of Bayesian support (LLRs)
  - Decisions by posterior knowledge of priors and LLR score
- Score-to-LLR calibration: meaningful LLRs
  - Necessary step, requiring a calibration data set
  - Essential for validation/accredetation
- Performance reporting
  - Decoupled decision policy
  - APE curves
  - NBER diagrams
  - ECE plots
  - Scalars: actDCF, minDCF, C_{llr} & C_{llr}^{min}



Conclusion

Perspectives

- From forensics to biometrics in general
- Forensics: distinct separation of role provinces



⇒ Non-forensic biometric companion/equivalent

⋮ vendor system

⋮ customer decision policy

Note: neither forensic scientists nor courts shall be automated, its an analogue.

Application fields

- Operating point independent performance reporting
  - Discrimination loss $\mapsto$ Goodness of scores w/o calibration
  - System calibration (meaningful)
  - Forensic state-of-the-art

$\Rightarrow$ European Network of Forensic Science Institutre (ENFSI): adopted Bayesian methodology (strong recommendation)

- Fix-operational testing: no need

$\Rightarrow$ But: <u>fundamental</u> in technology testing

This work has been funded by the Center for Advanced Security Research Darmstadt (CASED), and the Hesse government (project no. 467/15-09, BioMobile).

LOEWE
Exzellente Forschung für Hessens Zukunft

Evaluation of evidence strength

- Metrics in the Bayesian Framework
  - Application-independent generalization [2]:

    Goodness of (Log-Likelihood Ratio) scores C_{llr}

    $$C_{llr} = \frac{0.5}{|H_1|} \sum_{S \in H_1} \mathrm{ld}\left(1 + e^{-S}\right) + \frac{0.5}{|H_2|} \sum_{S \in H_2} \mathrm{ld}\left(1 + e^{S}\right)$$

  - Information-theoretic generalization [7]:

    Empirical Cross-Entropy (ECE)

    $$\mathrm{ECE} = \frac{\pi}{|H_1|} \sum_{S \in H_1} \mathrm{ld}\left(1 + e^{-(S\ \frac{\pi}{1-\pi})}\right) + \frac{1-\pi}{|H_2|} \sum_{S \in H_2} \mathrm{ld}\left(1 + e^{S\ \frac{\pi}{1-\pi}}\right)$$

- Metrics represent (cross-) entropy in bits
- Performance reporting with decoupled decision layer

[2] N. Brümmer and J. du Preez: *Application Independent Evaluation of Speaker Detection*, Computer Speech and Language, 2006.

[7] D. Ramos Castro and J. González Rodríguez: *Cross-entropy Analysis of the Information in Forensic Speaker Recognition*, Odyssey, 2008.

Brief introduction to calibration

- Linear: logistic regression (robust model)
  - Transform: $S_{\text{cal.}} = w_0 + w_1\, S$
- Non-linear: Pool-Adjacent-Violator (PAV) algorithm (optimal)
  - Transform: monotonic, non-parametric mapping function