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
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 list to investigate
  e.g., AFIS

- **Intelligence**
  Decide where to establish
  relevant links in a database

- **Forensic Evaluation**
  Communicate for the court
to decide a verdict
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**
  Communicate for the court
to decide a verdict
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**
  Communicate for the court to decide a verdict
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**
  Communicate for the court
to decide a verdict
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

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

\[
P(H_1) = 1% \\
P(H_2) = 99% \\
\text{LR} = 1000 \\
P(H_1 | E) = 91% \\
\text{odds: 1000:99}
\]

\[
\frac{P(H_1)}{P(H_2)} \times \frac{P(E | H_1)}{P(E | H_2)} = \frac{P(H_1 | E)}{P(H_2 | E)}
\]
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
  - Changes prior odds into posterior odds

Prior odds \[\longrightarrow\text{Inference}\longrightarrow\text{Posterior odds}\]

\[\frac{P(H_1 | E)}{P(H_2 | E)}\]
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

\[
P(H_1 \mid E) C_{f1} \geq P(H_2 \mid E) C_{f2}
\]

Prior odds \(\rightarrow\) Inference \(\rightarrow\) Posterior odds

LR (Biometric System) \(\uparrow\) Costs \(C_{f1}, C_{f2}\) \(\uparrow\) Decision \(H_1\) or \(H_2\)?
Bayesian Method

Decision Process: Competences

- Total separation between
  - The comparator (biometric system outputing 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

\[
\begin{align*}
\text{Prior odds} & \quad \text{Inference} \quad \text{Posterior odds} \\
\text{LR} & \quad \text{Should lead to the correct decision!} \\
\text{Costs} & \quad \text{Decision} \\
C_{f1}, C_{f2} & \quad \text{H}_1 \text{ or H}_2?
\end{align*}
\]
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
Bayesian Method

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]


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 LR [2,3]


Bayesian Method

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]


Bayesian Decision: 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
  ⇒ LR into a probabilistic framework

- LR computation: great experience in other fields
  ⇒ Example: forensic biometrics
Metrics and Examples

Revisiting ISO/IEC JTC1 SC37 SD11

FNMR, FMR $\leftrightarrow$ DET

Data Capture Subsystem

Data Storage Subsystem

Comparison Subsystem

Decision Subsystem

Signal Processing Subsystem

Biometric Claim

Reference

Probe

Features

Captured Biometric Sample

Comparison Score(s)

Decision Policy

Verification Outcome

Identification Outcome

$\pi = F_{NMR}$, $F_{MR} \rightarrow \text{DET}$

Revisiting ISO/IEC JTC1 SC37 SD11

Metrics and Examples

Nautsch, Ramos, et al. Bayesian Biometrics / NIST IBPC'16, Gaithersburg, 03.05.2016
\[ \frac{P(H_1)}{P(H_2)} = \pi \frac{1}{1-\pi} \]

\[ \Rightarrow \pi \]

**Metrics and Examples**

Revisiting ISO/IEC JTC1 SC37 SD11

- FNMR, FMR $\leftrightarrow$ DET

Bayesian Biometrics / NIST IBPC'16, Gaithersburg, 03.05.2016
Metrics and Examples

Revisiting ISO/IEC JTC1 SC37 SD11

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

\[\Rightarrow \pi\]

Data Capture Subsystem

Data Storage Subsystem

Comparison Subsystem

Decision Subsystem

FNMR, FMR ↔ DET

DCF ↔ APE & NBER

Nautsch, Ramos, et al. Bayesian Biometrics / NIST IBPC'16, Gaithersburg, 03.05.2016

16/32
Revisiting ISO/IEC JTC1 SC37 SD11

\[ \frac{P(H_1)}{P(H_2)} = \frac{\pi}{1-\pi} \Rightarrow \pi \]

Metrics and Examples

FNMR, FMR ↔ DET

DCF ↔ APE & NBER

ECE

Nautsch, Ramos, et al. Bayesian Biometrics / NIST IBPC'16, Gaithersburg, 03.05.2016
Revisiting ISO/IEC JTC1 SC37 SD11

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

\[ \Rightarrow \pi \]

DCF $\leftrightarrow$ APE & NBER

C\textsubscript{F1}, C\textsubscript{F2}

ECE

Nautsch, Ramos, et al. Bayesian Biometrics / NIST IBPC’16, Gaithersburg, 03.05.2016
Metrics and Examples

Detection Error Trade-off (DET) diagrams


Nautsch, Ramos, et al. Bayesian Biometrics / NIST IBPC’16, Gaithersburg, 03.05.2016
**Metrics and Examples**

**From Bayesian Decisions to Cost Functions**

- **Bayes theorem**
  \[
  \frac{P(H_1)}{P(H_2)} \times \frac{P(E|H_1)}{P(E|H_2)} = \frac{P(H_1|E)}{P(H_2|E)}
  \]

- **Decision rule**
  \[
  P(H_1|E) C_{f1} \geq P(H_2|E) C_{f2}
  \]
  \[
  \iff \frac{P(H_1|E)}{P(H_2|E)} \geq \frac{C_{f2}}{C_{f1}}
  \]

- **Bayesian threshold \( \eta \)** for Log-LRs (LLRs) by posterior odds
  \[
  \eta = \log \frac{C_{f2}}{C_{f1}} - \log \frac{P(H_1)}{P(H_2)} \geq \text{LLR}
  \]
Metrics and Examples

From Bayesian Decisions to Cost Functions

- **Bayes theorem**

  \[
  \frac{P(H_1)}{P(H_2)} \times \frac{P(E | H_1)}{P(E | H_2)} = \frac{P(H_1 | E)}{P(H_2 | E)}
  \]

- **Decision rule**

  \[
  P(H_1 | E) \ C_{f1} \geq \ P(H_2 | E) \ C_{f2}
  \]

  \[
  \Leftrightarrow \frac{P(H_1 | E)}{P(H_2 | E)} \geq \frac{C_{f2}}{C_{f1}}
  \]

- **Bayesian threshold \( \eta \) for Log-LRs (LLRs) by posterior odds**

  \[
  \eta = \log \frac{C_{f2}}{C_{f1}} - \log \frac{P(H_1)}{P(H_2)} \geq \text{LLR}
  \]
From Bayesian Decisions to Cost Functions

- **Bayes theorem**

Prior odds \( \frac{P(H_1)}{P(H_2)} \) \times LR \( \frac{P(E|H_1)}{P(E|H_2)} \) = Posterior odds \( \frac{P(H_1|E)}{P(H_2|E)} \)

- **Decision rule**

\[ P(H_1|E)C_{f1} \geq P(H_2|E)C_{f2} \]

\[ \Leftrightarrow \frac{P(H_1|E)}{P(H_2|E)} \geq \frac{C_{f2}}{C_{f1}} \]

- **Bayesian threshold** \( \eta \) for Log-LRs (LLRs) by posterior odds

\[ \eta = \log \frac{C_{f2}}{C_{f1}} - \log \frac{P(H_1)}{P(H_2)} \geq \text{LLR} \]
Metrics and Examples

From Bayesian Decisions to Cost Functions

- Bayesian error rate: Decision Cost Function (DCF)

\[ DCF(P(H_1), P(H_2), C_{f1}, C_{f2}) = P(H_1) \cdot FNMR(\eta) \cdot C_{f1} + P(H_2) \cdot FMR(\eta) \cdot C_{f2} \]

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

- Simplifying the operating point \((P(H_1), P(H_2), C_{f1}, C_{f2}) \rightarrow \tilde{\pi}\)

1. Mutually exclusive priors: \(\log \frac{P(H_1)}{P(H_2)} = \log \frac{\pi}{1-\pi} = \text{logit} \pi\)

\[ DCF(\pi, C_{f1}, C_{f2}) = \pi \cdot FNMR(\eta) \cdot C_{f1} + (1-\pi) \cdot FMR(\eta) \cdot C_{f2} \]

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

\[ DCF(\tilde{\pi}) = \tilde{\pi} \cdot FNMR(\eta) + (1 - \tilde{\pi}) \cdot FMR(\eta) = DCF(\tilde{\pi}, 1, 1) \]

\[ \eta = - \text{logit} \tilde{\pi} \]

\(\Rightarrow\) meaningful LLR operating points: \(\tilde{\pi}\) or \(\eta\)

Metrics and Examples

From Bayesian Decisions to Cost Functions

- **Bayesian error rate: Decision Cost Function (DCF)**
  \[
  DCF(P(H_1), P(H_2), C_{f_1}, C_{f_2}) = P(H_1) \text{FNMR}(\eta) C_{f_1} + P(H_2) \text{FMR}(\eta) C_{f_2}
  \]
  \[
  \eta = \log \frac{C_{f_2}}{C_{f_1}} - \log \frac{P(H_1)}{P(H_2)}
  \]

- **Simplifying the operating point** \((P(H_1), P(H_2), C_{f_1}, C_{f_2}) \mapsto \tilde{\pi}\)
  1. Mutually exclusive priors:
     \[
     \log \frac{P(H_1)}{P(H_2)} = \log \frac{\pi}{1-\pi} = \logit \pi
     \]
     \[
     DCF(\pi, C_{f_1}, C_{f_2}) = \pi \text{FNMR}(\eta) C_{f_1} + (1 - \pi) \text{FMR}(\eta) C_{f_2}
     \]
  2. Introducing an *effective prior*:
     \[
     \tilde{\pi} = \frac{\pi C_{f_1}}{\pi C_{f_1} + (1 - \pi) C_{f_2}}
     \]
     \[
     DCF(\tilde{\pi}) = \tilde{\pi} \text{FNMR}(\eta) + (1 - \tilde{\pi}) \text{FMR}(\eta) = DCF(\pi, 1, 1)
     \]
     \[
     \eta = - \logit \tilde{\pi}
     \]

⇒ meaningful LLR operating points: \(\tilde{\pi}\) or \(\eta\)

From Bayesian Decisions to Cost Functions

- **Bayesian error rate: Decision Cost Function (DCF)**
  \[
  DCF(P(H_1), P(H_2), C_{f1}, C_{f2}) = P(H_1) \cdot \text{FNMR}(\eta) \cdot C_{f1} + P(H_2) \cdot \text{FMR}(\eta) \cdot C_{f2}
  \]
  \[
  \eta = \log \frac{C_{f2}}{C_{f1}} - \log \frac{P(H_1)}{P(H_2)}
  \]

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

  \[
  DCF(\pi, C_{f1}, C_{f2}) = \pi \cdot \text{FNMR}(\eta) \cdot C_{f1} + (1-\pi) \cdot \text{FMR}(\eta) \cdot C_{f2}
  \]

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

  \[
  DCF(\tilde{\pi}) = \tilde{\pi} \cdot \text{FNMR}(\eta) + (1-\tilde{\pi}) \cdot \text{FMR}(\eta) = DCF(\pi, 1, 1)
  \]

  \[
  \eta = - \text{logit} \tilde{\pi}
  \]

  \(\Rightarrow\) meaningful LLR operating points: \(\tilde{\pi}\) or \(\eta\)

---

From Bayesian Decisions to Cost Functions

- Bayesian error rate: Decision Cost Function (DCF)

\[
\text{DCF}(P(H_1), P(H_2), C_{f1}, C_{f2}) = P(H_1) \cdot \text{FNMR}(\eta) \cdot C_{f1} + P(H_2) \cdot \text{FMR}(\eta) \cdot C_{f2}
\]

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

- Simplifying the operating point \((P(H_1), P(H_2), C_{f1}, C_{f2}) \mapsto \tilde{\pi}\)

1. Mutually exclusive priors: \(\log \frac{P(H_1)}{P(H_2)} = \log \frac{\pi}{1-\pi} = \logit \pi\)

\[
\text{DCF}(\pi, C_{f1}, C_{f2}) = \pi \cdot \text{FNMR}(\eta) \cdot C_{f1} + (1 - \pi) \cdot \text{FMR}(\eta) \cdot C_{f2}
\]

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

\[
\text{DCF}(\tilde{\pi}) = \tilde{\pi} \cdot \text{FNMR}(\eta) + (1 - \tilde{\pi}) \cdot \text{FMR}(\eta) = \text{DCF}(\pi, 1, 1)
\]

\[
\eta = - \logit \tilde{\pi}
\]

⇒ meaningful LLR operating points: \(\tilde{\pi}\) or \(\eta\)

From Bayesian Decisions to Cost Functions

- Bayesian error rate: Decision Cost Function (DCF)

\[
DCF(P(H_1), P(H_2), C_{f1}, C_{f2}) = P(H_1) \text{FNMR}(\eta) C_{f1} + P(H_2) \text{FMR}(\eta) C_{f2}
\]

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

- Simplifying the operating point \((P(H_1), P(H_2), C_{f1}, C_{f2}) \mapsto \tilde{\pi}\)

1. Mutually exclusive priors: \(\log \frac{P(H_1)}{P(H_2)} = \log \frac{\pi}{1-\pi} = \logit \pi\)

\[
DCF(\pi, C_{f1}, C_{f2}) = \pi \text{FNMR}(\eta) C_{f1} + (1 - \pi) \text{FMR}(\eta) C_{f2}
\]

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

\[
DCF(\tilde{\pi}) = \tilde{\pi} \text{FNMR}(\eta) + (1 - \tilde{\pi}) \text{FMR}(\eta) = DCF(\pi, 1, 1)
\]

\[
\eta = - \logit \tilde{\pi}
\]

\(\Rightarrow\) meaningful LLR operating points: \(\tilde{\pi}\) or \(\eta\)

Example on Decision Cost Functions (DCF)

- Speaker recognition iVEC/PLDA scores (I4U list/NIST SRE’12)

  ![Graph of LLRs and pdf](image)

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

  ![Graph of Cost vs. $\eta$](image)

  $\Rightarrow$ actual vs. minimum DCF: calibration loss
  $\Rightarrow$ LLR meaning: aligning scores for Bayesian support
Example on Decision Cost Functions (DCFs)

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

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

⇒ 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: \( \text{DCF}(1:1, \eta = 0) \) vs. \( \text{DCF}(1:100, \eta \approx 4.6) \)

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

Nautsch, Ramos, et al. Bayesian Biometrics / NIST IBPC’16, Gaithersburg, 03.05.2016
Visualizing DCFs

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

\[ \logit \tilde{\pi} = -\eta \]


Metrics and Examples

Visualizing DCFs

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

![Graph showing visualizing DCFs]

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


Metrics and Examples

Visualizing DCFs

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

\[ \text{logit} \tilde{\pi} = -\eta \]


Visualizing DCFs

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

Logit $\tilde{\pi} = -\eta$

![Graph showing DCFs and ROCCH](image)

EER: 0.5%

Metrics and Examples

- Visualizing DCFs
  - Simulating DCFs on multiple operating points
  - default: all LLRs = 0, i.e.: $DCF = \pi + (1 - \pi)$
  - Area-under-APE: cost of LLR scores
    $\Rightarrow$ Goodness of LLRs: $C_{llr}$


Metrics and Examples

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
  ⇒ wider range of operating points can be compared

\[ \eta = - \text{logit} \tilde{\pi} \]


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 $\xrightarrow{\text{Evidence}}$ Posterior entropy
  - Divergence of system to Grund-of-Truth (GoT)
  - ECE: approximating Kullback-Leibler divergence $D_{\text{GoT}||\text{system}}$

$H_{\text{system}}(H_1, H_2)$

$D_{\text{GoT}||\text{system}}(H_1, H_2 | \text{LLRs})$

$H_{\text{GoT}}(H_1, H_2 | \text{LLRs})$
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 $\Leftrightarrow$ discriminating power
  - $C_{llr}$ at $\log(\text{odds}) = 0$ $\Rightarrow$ no information on $H_1/H_2$ prior

![Graph showing ECE vs Prior log10(odds) with System, Optimal calibration, and default (LLRs=0) curves]

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 $\Leftrightarrow$ discriminating power
  - $C_{llr}$ at log(odds) = 0 $\Rightarrow$ no information on $H_1/H_2$ prior

![Graph showing ECE vs Prior log10(odds)]

Examples

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

---

Examples

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

\[ \eta = - \log \tilde{\pi} \]

Examples

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

![Bar chart showing discrimination and calibration loss for conventional, QMF, FQE, and binning schemes.]

Examples

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

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

Conclusion

Summary

- Bayesian decision framework
  - Bayes theorem & decision rule employing 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/accreditation

- 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 employing 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/accreditation

- Performance reporting
  - Decoupled decision policy
  - APE curves
  - NBER diagrams
  - ECE plots
  - Scalars: actDCF, minDCF, $C_{llr}$ & $C_{llr}^{min}$

![Graph showing discrimination and calibration loss]
## Perspectives

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

\[
\text{Suspect reference} \rightarrow \text{Feature extraction} \rightarrow \text{Evidence analysis (comparison)} \rightarrow \text{Score} \rightarrow \text{Guilty (Accept)} \rightarrow \text{Not-Guilty (Reject)}
\]

Province of the forensic scientist \rightarrow \text{Province of the court}

\[ \Rightarrow \text{Non-forensic biometric companion/equivalent} \]

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

Nautsch, Ramos, et al.  
Bayesian Biometrics / NIST IBPC'16, Gaithersburg, 03.05.2016
Conclusion

Application fields

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

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

- Fix-operational testing: no need

$\Rightarrow$ But: fundamental 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).

Nautsch, Ramos, et al. Bayesian Biometrics / NIST IBPC'16, Gaithersburg, 03.05.2016
Evaluation of evidence strength

 Metrics in the Bayesian Framework
  ▶ Application-independent generalization [2]:
  
  \[
  \text{Goodness of (Log-Likelihood Ratio) scores } C_{llr} = \frac{0.5}{|H_1|} \sum_{S \in H_1} \log \left(1 + e^{-S}\right) + \frac{0.5}{|H_2|} \sum_{S \in H_2} \log \left(1 + e^S\right)
  \]

  ▶ Information-theoretic generalization [7]:
  
  \[
  \text{Empirical Cross-Entropy (ECE)}
  \]
  \[
  \text{ECE} = \frac{\pi}{|H_1|} \sum_{S \in H_1} \log \left(1 + e^{-(S \cdot \frac{\pi}{1-\pi})}\right) + \frac{1-\pi}{|H_2|} \sum_{S \in H_2} \log \left(1 + e^{S \cdot \frac{\pi}{1-\pi}}\right)
  \]

  ▶ Metrics represent (cross-) entropy in bits

  ▶ Performance reporting with decoupled decision layer


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