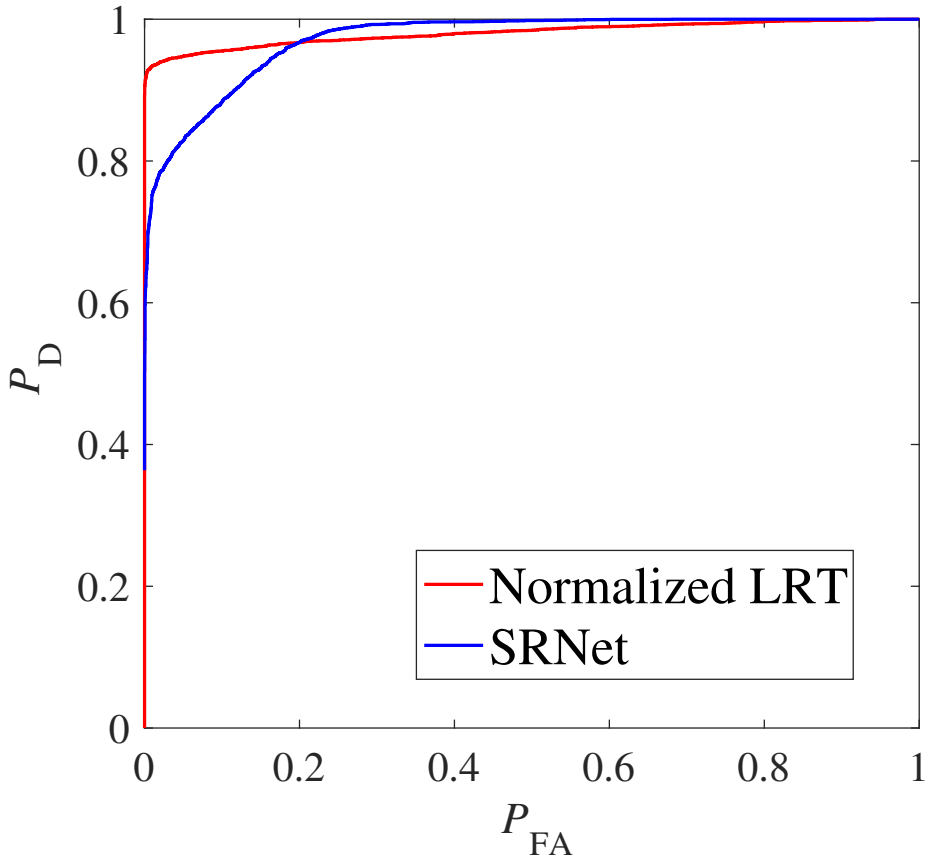
On Comparing Ad Hoc Detectors with Statistical Hypothesis Tests

Eli Dworetzky, Edgar Kaziakhmedov, Jessica Fridrich

IH&MMSEC 2023

BINGHAMTON
UNIVERSITY
STATE UNIVERSITY OF NEW YORK

LSBM at 0.03 bpp in artificial BOSSbase[1]



[1]M. Boroumand, J. Fridrich, R. Cogranne, "Are We There Yet?", IS&T Electronic Imaging, January 2019.

How we build detectors

Model-based detectors

- Statistically model a **single** image
- Different images have different models

ML-based detectors

- Trained on a multitude of images
- Implicitly learn a model for the **entire** cover and stego sources

We limit arguments to an acquisition noise model[2]. Such a model

- Captures the distribution of taking multiple acquisitions of a scene
- Parameterized by a "noise-free" image of the scene

[2] Alternative models: noise residuals, quantization errors, ...

Model-based detectors

Cover image is a sample from PDF $s_0(\mathbf{x};\mathbf{c})$
$\mathbf{c}$ is a noise-free scene

Embedding payload α changes PDF to $s_\alpha(\mathbf{x};\mathbf{c})$

Given image $\mathbf{y}$, Warden faces

$$\mathcal{H}_0 : \mathbf{y} \sim s_0(\mathbf{x};\mathbf{c})$$
$$\mathcal{H}_1 : \mathbf{y} \sim s_\alpha(\mathbf{x};\mathbf{c})$$

Assuming the distributions are known, the MP detector is

$$\ell_{\mathbf{c}}(\mathbf{y}) = \log \frac{s_\alpha(\mathbf{y};\mathbf{c})}{s_0(\mathbf{y};\mathbf{c})} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma$$

Factorizable models

MiPOD, Gaussian embedding, RJCA, JCA, sensor ISO noise model

$$s_0(\mathbf{x};\mathbf{c}) = \prod_{i=1}^{N} p_0^{(i)}(x_i;\mathbf{c})$$

$$s_\alpha(\mathbf{x};\mathbf{c}) = \prod_{i=1}^{N} p_{\beta_i}^{(i)}(x_i;\mathbf{c})$$

N number of pixels, α relative payload, $\beta_i(\mathbf{c})$ change rate of ith pixel

$$\ell_{\mathbf{c}}(\mathbf{y}) = \sum_{i=1}^{N} \log \frac{p_{\beta_i}^{(i)}(y_i;\mathbf{c})}{p_0^{(i)}(y_i;\mathbf{c})}$$

For small payloads

$$\mathbb{E}_0 \sum_{i=1}^{N} \log \frac{p_{\beta_i}^{(i)}(y_i;\mathbf{c})}{p_0^{(i)}(y_i;\mathbf{c})} = -\sum_{i=1}^{N} D_{\mathrm{KL}}(p_0^{(i)}||p_{\beta_i}^{(i)}) \approx -\frac{1}{2}\sum_{i=1}^{N} I_i\beta_i^2 = -\frac{1}{2}\delta_{\mathbf{c}}^2$$

$$\mathbb{E}_1 \sum_{i=1}^{N} \log \frac{p_{\beta_i}^{(i)}(y_i;\mathbf{c})}{p_0^{(i)}(y_i;\mathbf{c})} = \sum_{i=1}^{N} D_{\mathrm{KL}}(p_{\beta_i}^{(i)}||p_0^{(i)}) \approx \frac{1}{2}\sum_{i=1}^{N} I_i\beta_i^2 = \frac{1}{2}\delta_{\mathbf{c}}^2$$

$\delta_{\mathbf{c}}^2$ deflection coefficient, I_i Fisher information at pixel i

$$I_i = \int \frac{1}{p_0^{(i)}(x;\mathbf{c})} \left(\frac{\partial p_\beta^{(i)}(x;\mathbf{c})}{\partial \beta}\Big|_{\beta=0} \right)^2 \mathrm{d}x$$

Variances

$$\mathrm{Var}_0[\ell_{\mathbf{c}}(\mathbf{y})] = \mathrm{Var}_1[\ell_{\mathbf{c}}(\mathbf{y})] = \delta_{\mathbf{c}}^2$$

Asymptotic form of the MP detector

For large number of pixels N, by Lindeberg CLT

$$\ell_{\mathbf{c}}(\mathbf{y}) \dot{\sim} \begin{cases} \mathcal{N}\left(-\frac{1}{2}\delta_{\mathbf{c}}^2, \delta_{\mathbf{c}}^2\right) & \mathcal{H}_0 \\ \mathcal{N}\left(\frac{1}{2}\delta_{\mathbf{c}}^2, \delta_{\mathbf{c}}^2\right) & \mathcal{H}_1 \end{cases}$$

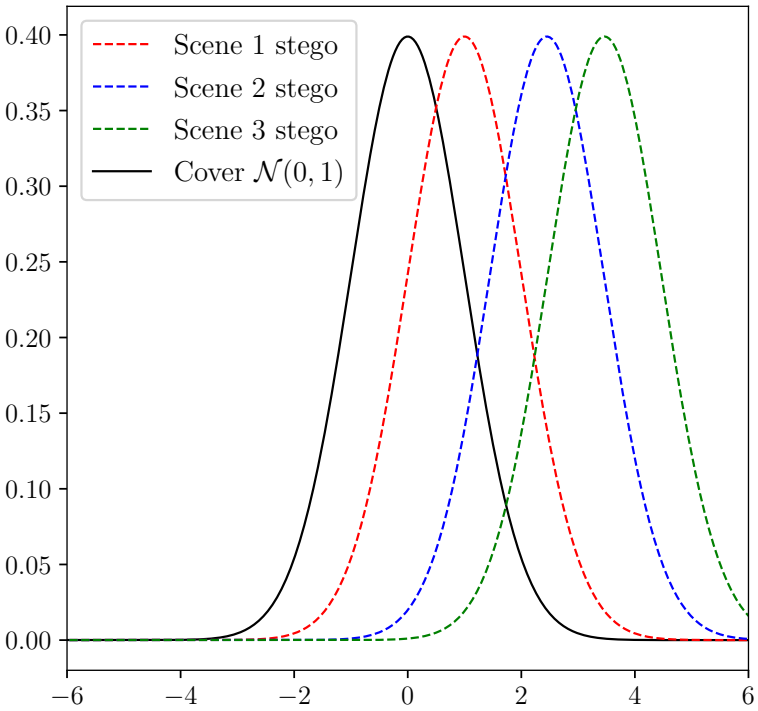
$$\bar{\ell}_{\mathbf{c}}(\mathbf{y}) = \frac{\ell_{\mathbf{c}}(\mathbf{y}) - \mathbb{E}_0\left[\ell_{\mathbf{c}}(\mathbf{y})\right]}{\sqrt{\mathrm{Var}_0\left[\ell_{\mathbf{c}}(\mathbf{y})\right]}} \dot{\sim} \begin{cases} \mathcal{N}(0,1) & \mathcal{H}_0 \\ \mathcal{N}(\delta_{\mathbf{c}},1) & \mathcal{H}_1 \end{cases}$$

ROC for scene $\mathbf{c}$

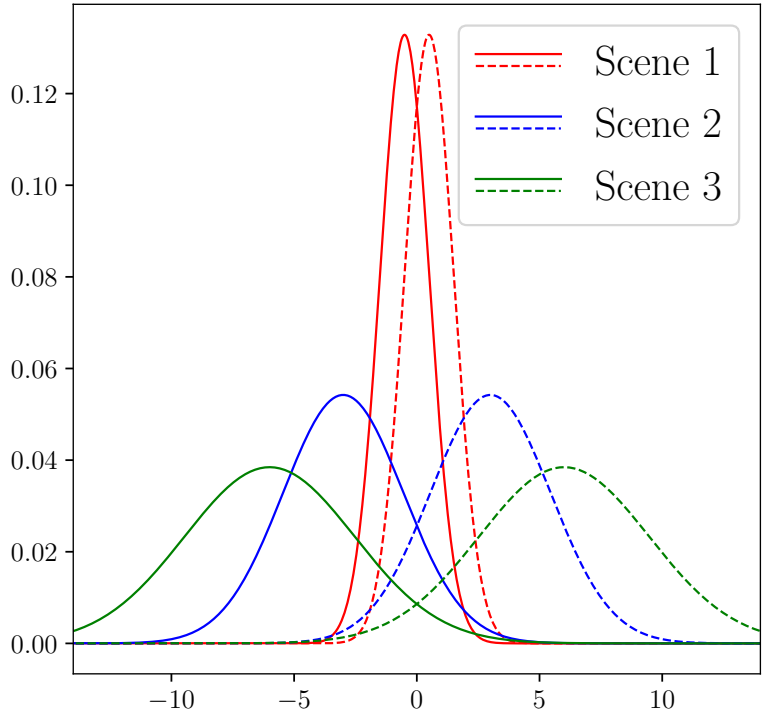
$$P_{\mathrm{D}}(P_{\mathrm{FA}}) = Q\left(Q^{-1}(P_{\mathrm{FA}}) - \delta_{\mathbf{c}}\right)$$

$Q(x) = \int_x^\infty (2\pi)^{-1} \exp(-z^2/2)\mathrm{d}z$ ... tail probability of $\mathcal{N}(0,1)$

Normalized vs. non-normalized test

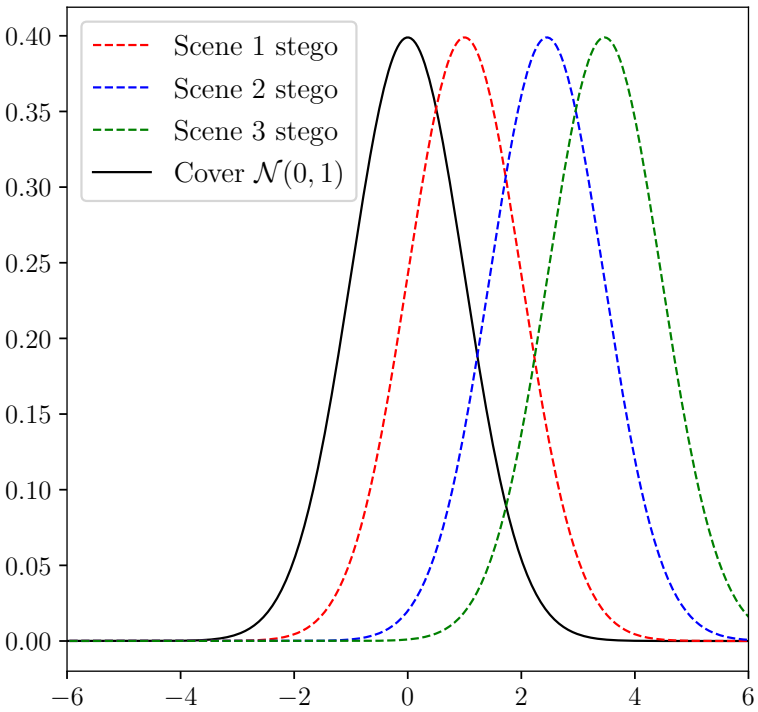


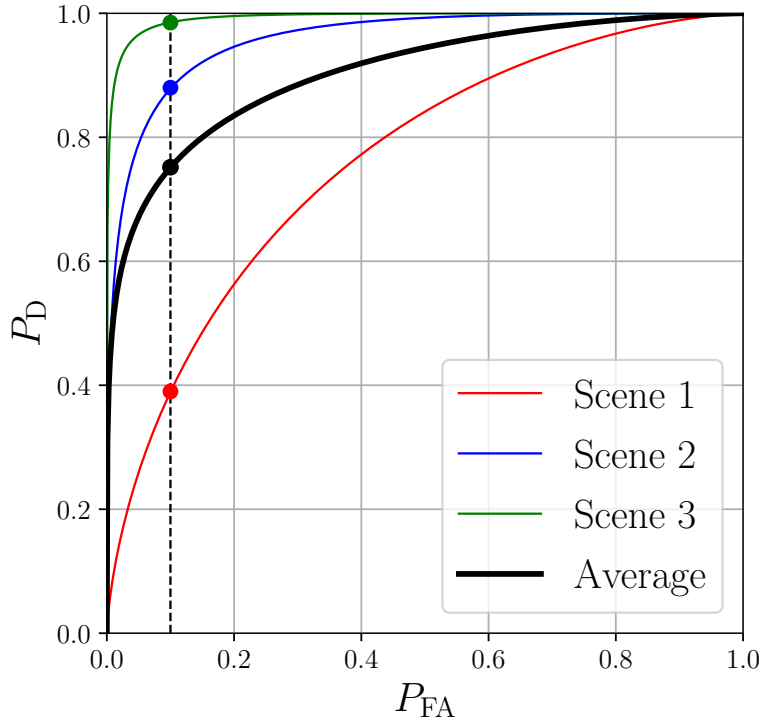
Normalized LRT



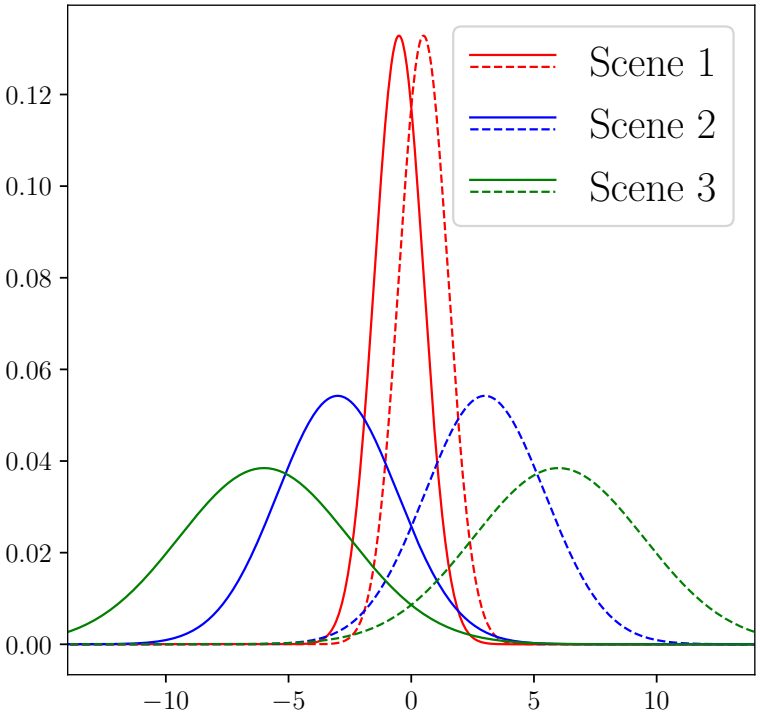
Non-normalized LRT

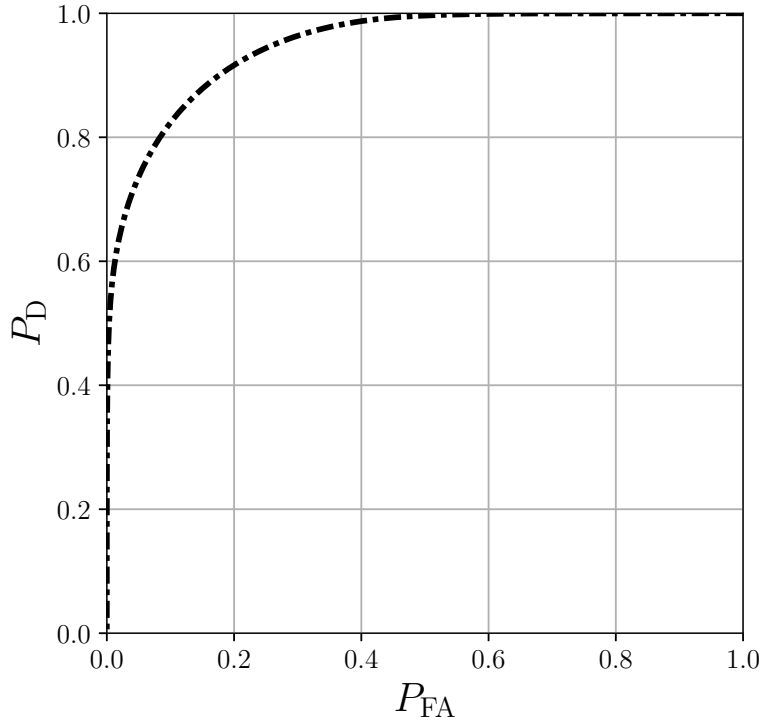
Normalized test



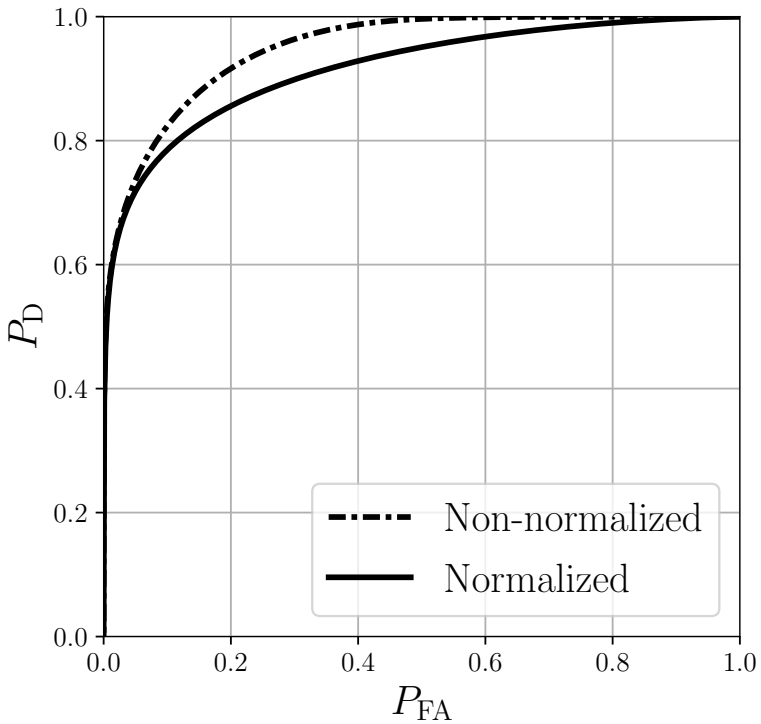


Non-normalized test





ROCs side by side



Cover source model

Sampling from the cover source involves sampling from the universe of all possible noise-free scenes $\mathbf{c}$

$$\mathbf{c} \sim \nu$$

and then acquiring actual digital image by sampling from $s_0(\mathbf{x}; \mathbf{c})$

Covers follow	$p_0(\mathbf{x}) = \int_{\mathcal{X}} s_0(\mathbf{x}; \mathbf{c}) \mathrm{d}\nu(\mathbf{c})$
Stegos follow	$p_\alpha(\mathbf{x}) = \int_{\mathcal{X}} s_\alpha(\mathbf{x}; \mathbf{c}) \mathrm{d}\nu(\mathbf{c})$

- ν infeasible to estimate
- ν induces a distribution on deflections $\delta_{\mathbf{c}} \sim \mu$
- μ feasible to estimate since $\delta_{\mathbf{c}}$ is a scalar

Case I: Test between mixtures

Given image $\mathbf{y} \in \mathcal{X}$, $\mathbf{c} \sim \nu$

$$\mathcal{H}_0 : \mathbf{y} \sim \int_{\mathcal{X}} s_0(\mathbf{x}; \mathbf{c}) \mathrm{d}\nu(\mathbf{c})$$
$$\mathcal{H}_1 : \mathbf{y} \sim \int_{\mathcal{X}} s_\alpha(\mathbf{x}; \mathbf{c}) \mathrm{d}\nu(\mathbf{c})$$

$$L(\mathbf{y}) = \log \frac{\int_{\mathcal{X}} s_\alpha(\mathbf{y}; \mathbf{c}) \mathrm{d}\nu(\mathbf{c})}{\int_{\mathcal{X}} s_0(\mathbf{y}; \mathbf{c}) \mathrm{d}\nu(\mathbf{c})} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma$$

ROC

$$P_{\mathrm{D}}(\gamma) = \mathbb{P}_1 \left(L(\mathbf{y}) > \gamma \right)$$
$$P_{\mathrm{FA}}(\gamma) = \mathbb{P}_0 \left(L(\mathbf{y}) > \gamma \right)$$

Case I: Test between mixtures

Given a dataset of n scenes $\mathbf{c}_1, \ldots, \mathbf{c}_n$

$$\log \sum_{i=1}^{n} s_\alpha(\mathbf{x}; \mathbf{c}_i) \approx \log s_\alpha(\mathbf{x}; \mathbf{c}_k) \text{ for some } k$$

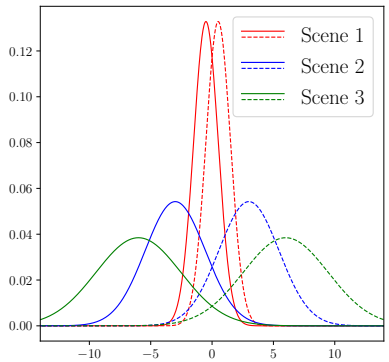
$L(\mathbf{y}) = \log \frac{\sum_{i=1}^{n} s_\alpha(\mathbf{x};\mathbf{c}_i)}{\sum_{i=1}^{n} s_0(\mathbf{x};\mathbf{c}_i)}$ approximated by mixture of

$\ell_{\mathbf{c}_i}(\mathbf{y}) = \log \frac{s_\alpha(\mathbf{y};\mathbf{c}_i)}{s_0(\mathbf{y};\mathbf{c}_i)}$

ROC

$$P_{\mathrm{D}}(\gamma) = 1/n \sum_{i=1}^{n} \mathbb{P}_1\left(\ell_{\mathbf{c}_i}(\mathbf{y}) > \gamma\right) \underset{n\to\infty}{\to} \mathbb{E}_{\delta_{\mathbf{c}}^2 \sim \mu} \mathbb{P}\left(\mathcal{N}\left(\delta_{\mathbf{c}}^2/2, \delta_{\mathbf{c}}^2\right) > \gamma\right)$$

$$P_{\mathrm{FA}}(\gamma) = 1/n \sum_{i=1}^{n} \mathbb{P}_0\left(\ell_{\mathbf{c}_i}(\mathbf{y}) > \gamma\right) \underset{n\to\infty}{\to} \mathbb{E}_{\delta_{\mathbf{c}}^2 \sim \mu} \mathbb{P}\left(\mathcal{N}\left(-\delta_{\mathbf{c}}^2/2, \delta_{\mathbf{c}}^2\right) > \gamma\right)$$



Case II: Random hypothesis test

Warden faces the **random** hypothesis test (or a **mixture of hypotheses**)

$$\mathcal{H}_0 : \mathbf{y} \sim s_0(\mathbf{x}; \mathbf{c})$$
$$\mathcal{H}_1 : \mathbf{y} \sim s_\alpha(\mathbf{x}; \mathbf{c})$$

and adjusts decision threshold for each scene to satisfy a given P_{FA}

$$\mathbb{P}_0 \left(\ell_{\mathbf{c}}(\mathbf{y}) > \gamma_{\mathbf{c}} \right) \leq P_{\text{FA}}$$

while maximizing

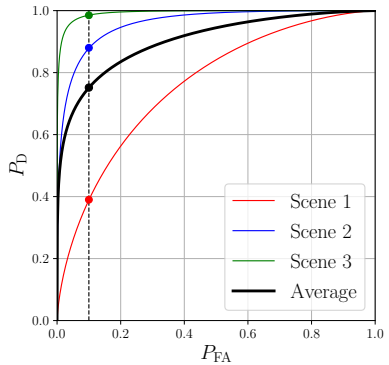
$$\mathbb{P}_1 \left(\ell_{\mathbf{c}}(\mathbf{y}) > \gamma_{\mathbf{c}} \right)$$

Case II: When test can be normalized

$$\overline{\ell}_{\mathbf{c}}(\mathbf{y}) \dot{\sim} \begin{cases} \mathcal{N}(0,1) \\ \mathcal{N}(\delta_{\mathbf{c}},1) \end{cases}$$

ROC

$$P_{\mathrm{D}}(\gamma) = 1/n \sum_{i=1}^{n} \mathbb{P}_1\left(\overline{\ell}_{\mathbf{c}_i}(\mathbf{y}) > \gamma\right) \to \mathbb{E}_{\mathbf{c}\sim\nu} Q(\gamma - \delta_{\mathbf{c}})$$
$$P_{\mathrm{FA}}(\gamma) = 1/n \sum_{i=1}^{n} \mathbb{P}_0\left(\overline{\ell}_{\mathbf{c}_i}(\mathbf{y}) > \gamma\right) = Q(\gamma)$$



$$\overline{P}_{\mathrm{D}}(P_{\mathrm{FA}}) = \mathbb{E}_{\delta_{\mathbf{c}}^2\sim\mu} Q(Q^{-1}(P_{\mathrm{FA}}) - \delta_{\mathbf{c}})$$

Case II: ROC is highly asymmetrical

$$\begin{aligned}
\overline{P}_{\mathrm{D}}(P_{\mathrm{FA}}) &= \mathbb{E}_{\delta_{\mathbf{c}}^2 \sim \mu} Q(Q^{-1}(P_{\mathrm{FA}}) - \delta_{\mathbf{c}}) \\
&= \sum_{k=0}^{\infty} \frac{(-1)^k c_k}{k!} Q^{(k)}(Q^{-1}(P_{\mathrm{FA}}) - \mathbb{E}[\delta_{\mathbf{c}}]) \\
&\approx \underbrace{Q(Q^{-1}(P_{\mathrm{FA}}) - \mathbb{E}[\delta_{\mathbf{c}}])}_{\text{symmetric Gaussian ROC}} + \underbrace{\frac{1}{2} c_2 Q''(Q^{-1}(P_{\mathrm{FA}}) - \mathbb{E}[\delta_{\mathbf{c}}])}_{\text{Term causing asymmetry}}
\end{aligned}$$

c_k ... kth central moment of $\delta_{\mathbf{c}}$

Second term positive iff $P_{\mathrm{FA}} < Q(\mathbb{E}[\delta_{\mathbf{c}}])$

Case I ROC $\geq$ Case II ROC

ROC for Case I (non-normalized test) $P_{\mathrm{D}}(P_{\mathrm{FA}})$ given by

$$P_{\mathrm{D}}(\gamma) = \mathbb{E}_{\delta_{\mathbf{c}}^2 \sim \mu} \mathbb{P}\left(\mathcal{N}\left(\delta_{\mathbf{c}}^2/2, \delta_{\mathbf{c}}^2\right) > \gamma\right)$$
$$P_{\mathrm{FA}}(\gamma) = \mathbb{E}_{\delta_{\mathbf{c}}^2 \sim \mu} \mathbb{P}\left(\mathcal{N}\left(-\delta_{\mathbf{c}}^2/2, \delta_{\mathbf{c}}^2\right) > \gamma\right)$$

bounds the ROC for Case II (normalized test)

$$\overline{P}_{\mathrm{D}}(P_{\mathrm{FA}}) = \mathbb{E}_{\delta_{\mathbf{c}}^2 \sim \mu} Q(Q^{-1}(P_{\mathrm{FA}}) - \delta_{\mathbf{c}})$$

$$\boxed{P_{\mathrm{D}}(P_{\mathrm{FA}}) \geq \overline{P}_{\mathrm{D}}(P_{\mathrm{FA}})}$$

Counter-intuitive since the normalized test decides for each scene rather than an entire source. However, the normalized test is ultimately crippled by the stringency of the FA constraint **for each scene**.

Outline of the proof (Case III)

Decision thresholds γ_i for each conditional test $\ell_{\mathbf{c}_i}(\mathbf{y}) = \log \frac{s_\alpha(\mathbf{y};\mathbf{c}_i)}{s_0(\mathbf{y};\mathbf{c}_i)}$ determined to maximize average P_{D}

$$\overline{P}_{\mathrm{D}}(\gamma_1, \ldots, \gamma_n) \triangleq \frac{1}{n} \sum_{i=1}^{n} \mathbb{P}_1 \left(\ell_{\mathbf{c}_i}(\mathbf{y}) > \gamma_i \right)$$

while satisfying the constraint for the average P_{FA}

$$\overline{P}_{\mathrm{FA}}(\gamma_1, \ldots, \gamma_n) \triangleq \frac{1}{n} \sum_{i=1}^{n} \mathbb{P}_0 \left(\ell_{\mathbf{c}_i}(\mathbf{y}) > \gamma_i \right) \leq P_{\mathsf{FA}}$$

- $\overline{P}_{\mathrm{D}}(\overline{P}_{\mathrm{FA}})$ bounds Case II since γ_i determined by $\mathbb{P}_0 \left(\ell_{\mathbf{c}_i}(\mathbf{y}) > \gamma_i \right) \leq P_{\mathsf{FA}}$ **for each** i
- Using Lagrange multipliers, $\gamma_i = \gamma$ across scenes, which corresponds to Case I (non-normalized test)

Lessons learned

Since the ROC of non-normalized test bounds the one for normalized test, ad hoc detectors are portrayed in better light—the normalized test is disadvantaged.

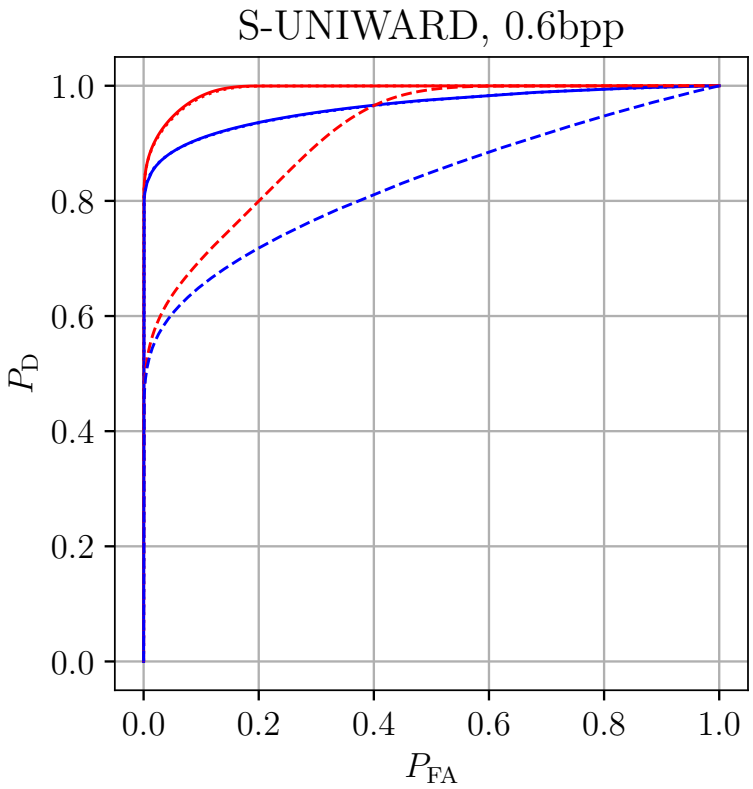
If data does not follow model exactly, compare ad hoc detectors with non-normalized test.

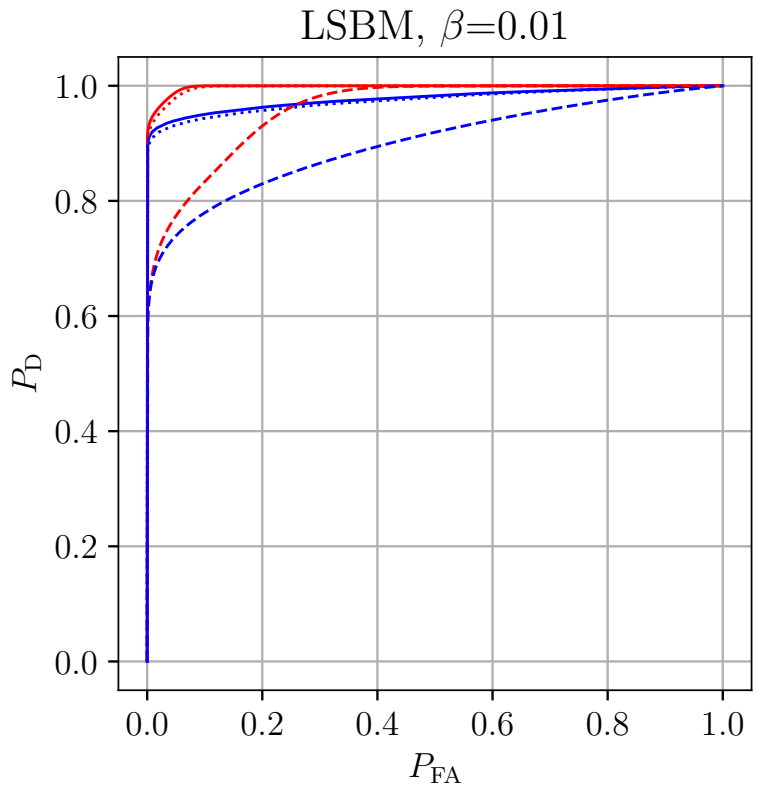
If data follows model exactly, one can choose to compare either form of the detector.

BOSSbase with a known scene model

1. BOSSbase, 10,000 256×256 grayscale images
2. Compute (MiPOD) pixel variance from BOSSbase images
3. Denoise all images aggressively
4. Reintroduce noise (i.d. Gaussian)
5. Optimal test is LRT

Comparing LRT with SRNet





Non-normalized test — LRT fine quant - - - SRNet

Normalized test ··· LRT sampled

Conclusions

ROC of ad hoc detector

- corresponds to a non-normalized test
- symmetrical

ROC of normalized test

- mixture of tests while guaranteeing constant FAR across images
- highly non-symmetrical

To compare fairly

- use non-normalized test OR
- normalize ad hoc detector (if cover model available)
- When data follows model exactly, ROC for mixture bounds ROC of normalized test