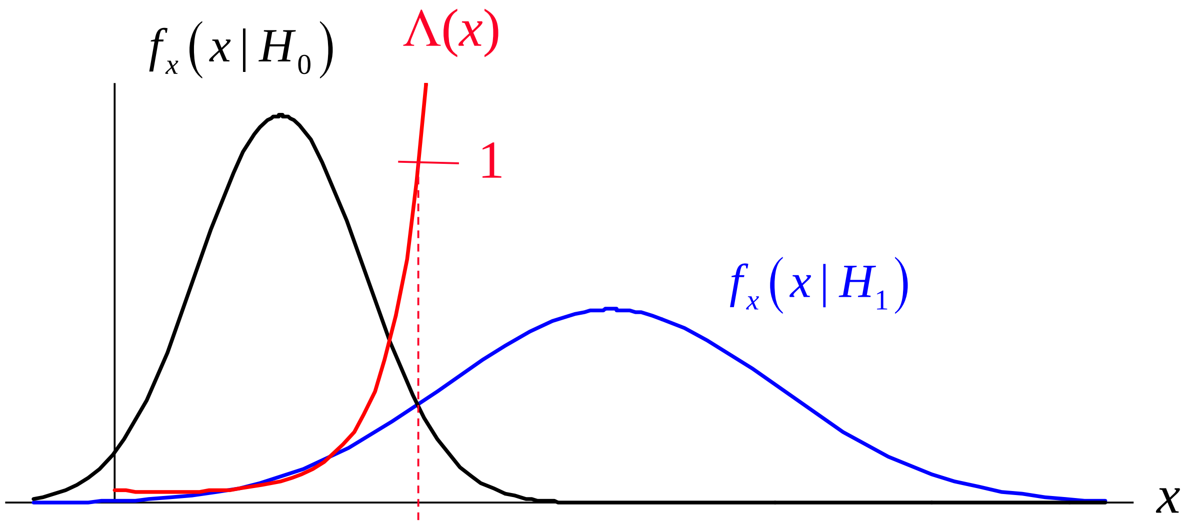
Estimation Theory

3 - Detection

Institut für Photogrammetrie und Fernerkundung, Fakultät für Bauingenieur-,
Geo- und Umweltwissenschaften



Detection

- Decision between 2 hypotheses H_0 and H_1, e.g:

  H_0 : no edge, bit "0", no airplane, no ship, …

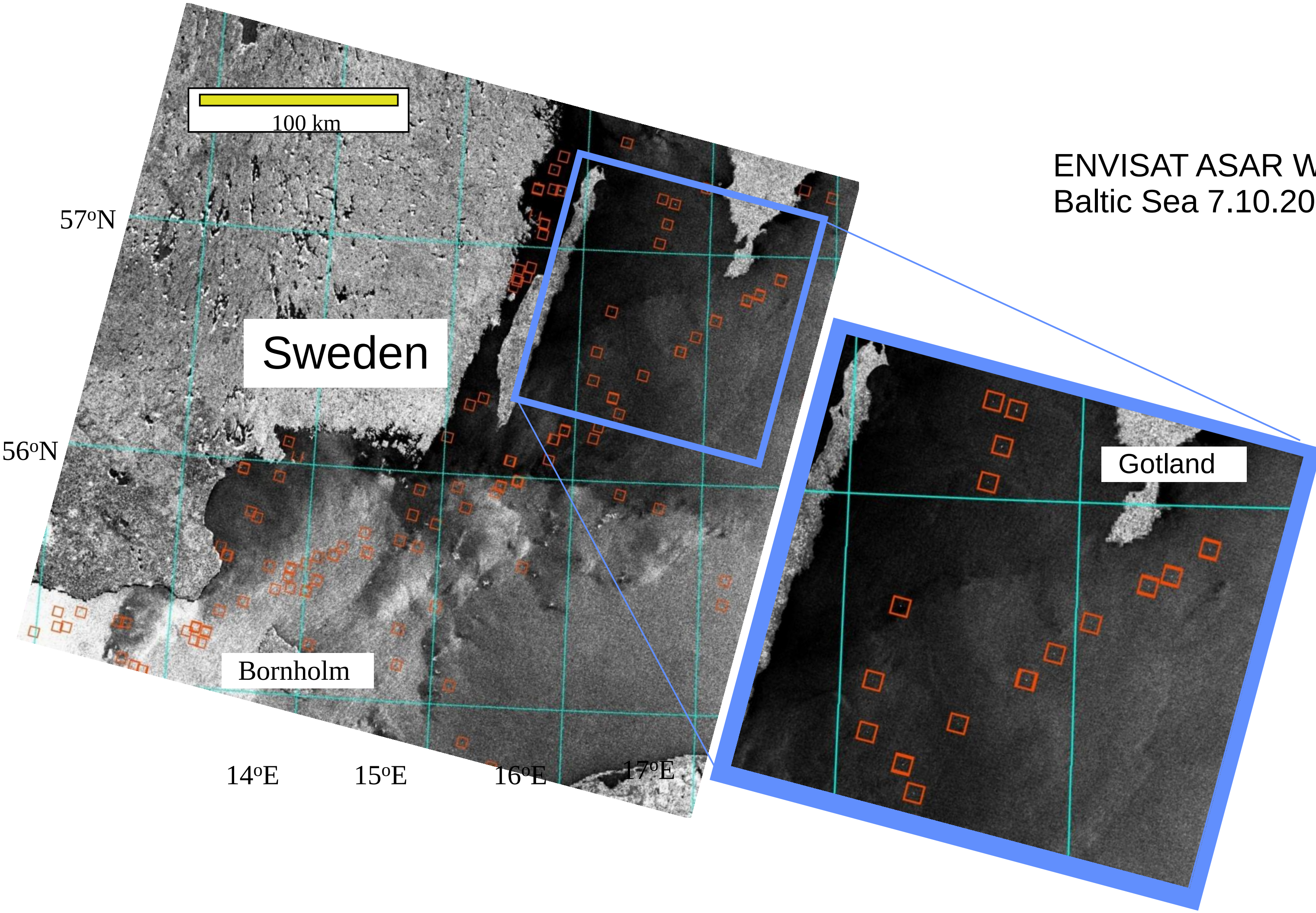
  H_1 : edge, bit "1", airplane, ship, …

- Equivalent to the problem of testing theory

- Optimum decision strategies require the formulation of a target function to be minimized or maximized

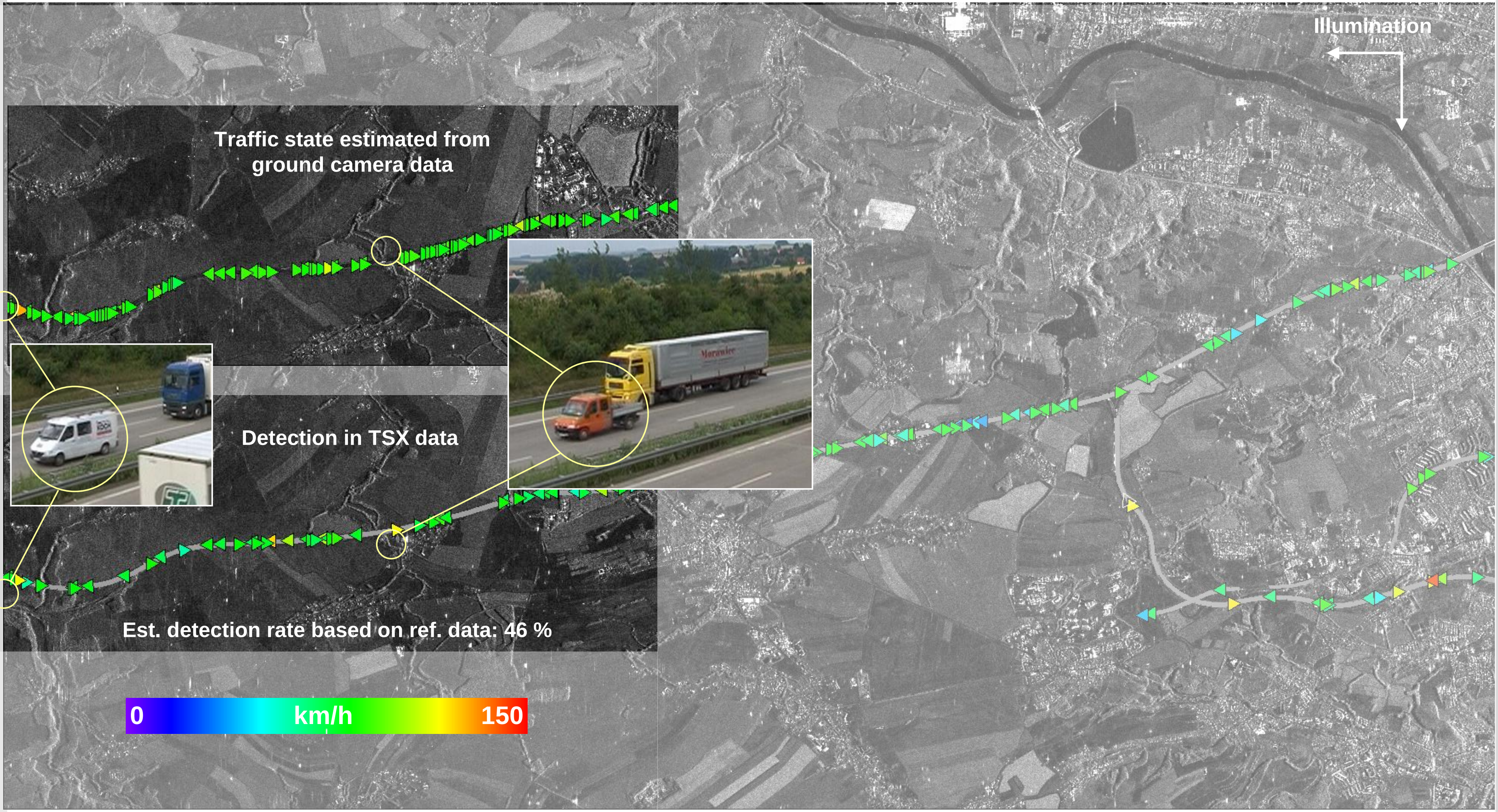
- Basis of decision: measurement vector $\mathbf{x}$, e.g. pixel values in an estimation window, signal record, …

Ship Detection (ENVISAT Data)



ENVISAT ASAR Wide Swath
Baltic Sea 7.10.2003 9:17 UTC

Vehicle Detection & Traffic Monitoring (TerraSAR-X Data)



Approx. 17 km

Number of detections for motorways: 156

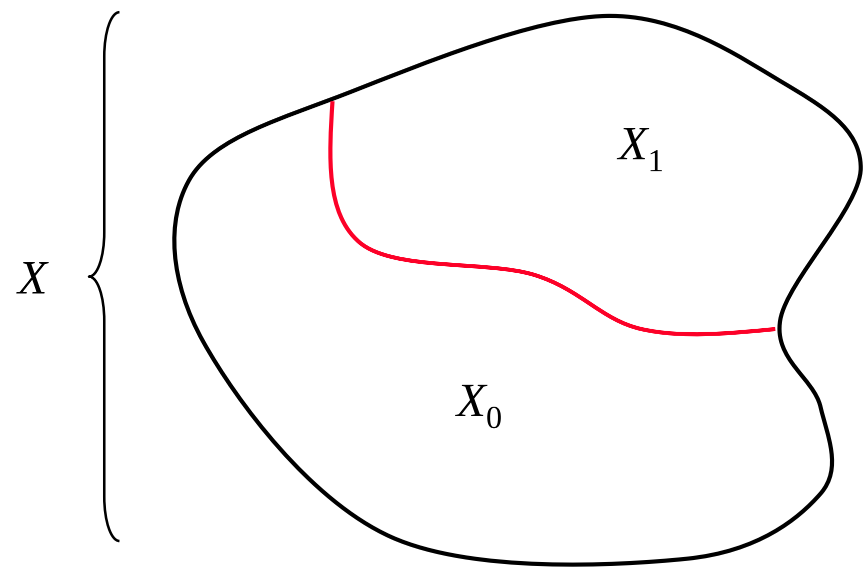
Measurement Space

- All possible $\mathbf{x}$ span the measurement (observation) space X :

- detection algorithm:
  division of X into X_0 and X_1, such that

  decision for H_0 if $\mathbf{x} \in X_0$

  decision for H_1 if $\mathbf{x} \in X_1$



- Decision Strategies:
  - Bayesian decision rule
  - Minimization of Expected Total Cost
  - Minimization of Error Probability
  - Minimization of Maximum Cost (Minimax Test)
  - Neyman-Pearson Test

  …

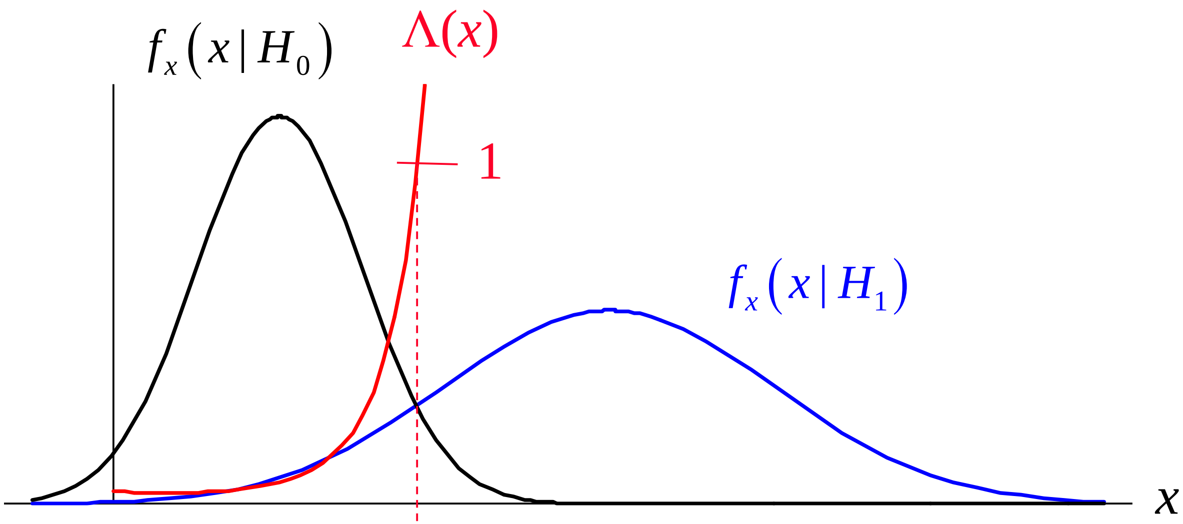
Likelihood Ratio

- Likelihoods $f_{\mathbf{x}}(\mathbf{x} \mid H_0)$ and $f_{\mathbf{x}}(\mathbf{x} \mid H_1)$ are assumed to be known

Likelihood ratio: $$\Lambda(\mathbf{x}) = \frac{f_{\mathbf{x}}(\mathbf{x} \mid H_1)}{f_{\mathbf{x}}(\mathbf{x} \mid H_0)}$$

- Often $\ln \Lambda(\mathbf{x})$ is used

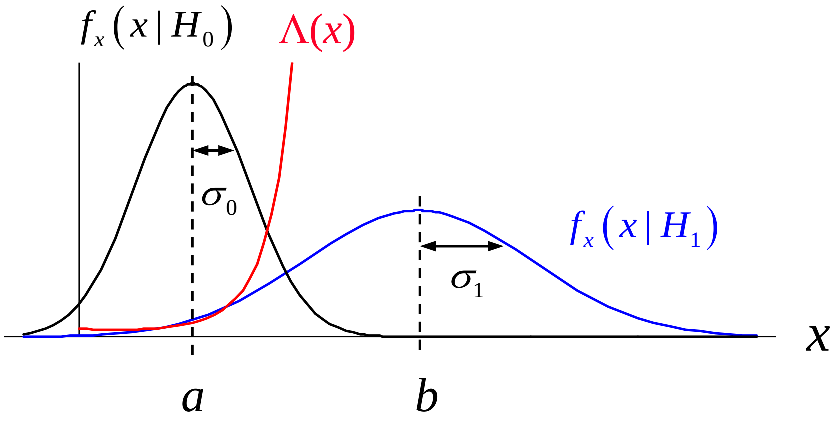
  „log-likelihood ratio“

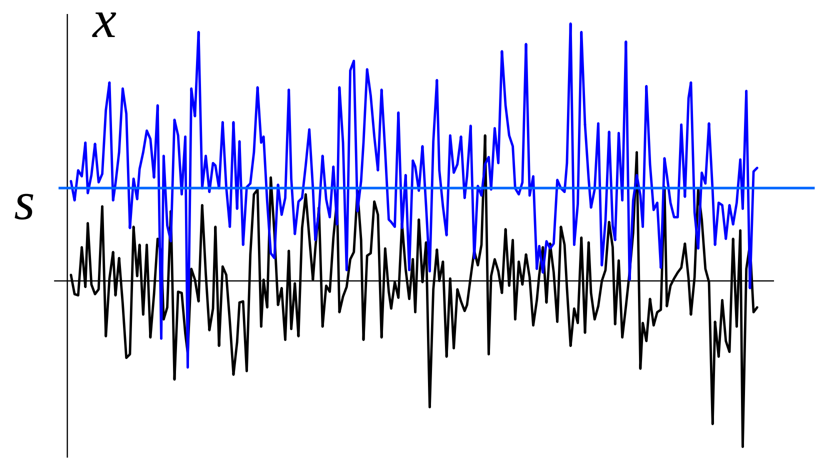


- $\Lambda(\mathbf{x})$ or $\ln \Lambda(\mathbf{x})$ are compared to some threshold:

$$\ln \Lambda(\mathbf{x}) \overset{H_1}{>} \lambda$$ **log-likelihood test**

Example: Detection of Signal in Gaussian noise (general case)





$$f_x(x|H_0) = \frac{1}{\sqrt{2\pi}\,\sigma_0}\exp\left(-\frac{(x-a)^2}{2\sigma_0^2}\right) \qquad f_x(x|H_1) = \frac{1}{\sqrt{2\pi}\,\sigma_1}\exp\left(-\frac{(x-b)^2}{2\sigma_1^2}\right)$$

$$\Lambda(x) = \frac{\sigma_1}{\sigma_0}\exp\left(-\frac{(x-b)^2}{2\sigma_1^2} + \frac{(x-a)^2}{2\sigma_0^2}\right) \qquad \ln\Lambda(x) = \ln\frac{\sigma_1}{\sigma_0} - \frac{(x-b)^2}{2\sigma_1^2} + \frac{(x-a)^2}{2\sigma_0^2}$$

Example: Detection of a Signal in Gaussian noise (specific)

$$\ln \Lambda(x) = -\frac{1}{2\sigma_n^2}\left((x-s)^2 - x^2\right)$$

$$= -\frac{1}{2\sigma_n^2}\left(-2xs + s^2\right) \overset{H_1}{>} \lambda$$

$$-2xs + s^2 < -2\sigma_n^2 \lambda$$

$$\Rightarrow \quad x \overset{H_1}{>} \frac{\sigma_n^2 \lambda}{s} + \frac{s}{2}$$

Assumptions:

- Noise is zero-mean (a = 0)
- Noise has same variance as the signal.
- Variance is not dependent on the signal's amplitude.

Example: Detection of an Edge based on the Values of two adjacent pixels (Additive Gaussian Noise)

- Given: two pixel values $x_1, x_2 \quad \Rightarrow \quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$

  - H_0: x_1 and x_2 belong to the <u>same</u> PDF (noise std = σ):

$$\frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2} \right) \quad \Rightarrow \text{ no edge}$$

Assumption:

PDFs have same variance

  - H_1: x_1 and x_2 belong to <u>different</u> PDFs:

$$\frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(x-\mu_1)^2}{2\sigma^2} \right) \quad \text{and} \quad \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(x-\mu_2)^2}{2\sigma^2} \right) \quad \Rightarrow \text{ edge}$$

Example: Detection of an Edge based on the values of two adjacent pixels (Additive Gaussian Noise) (cont.)

- x_1 and x_2 be independent. Then

$$f_{\mathbf{x}}\left(x_1, x_2 \mid H_0\right) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{1}{2\sigma^2}\left(\left(x_1-\mu\right)^2+\left(x_2-\mu\right)^2\right)\right)$$

$$f_{\mathbf{x}}\left(x_1, x_2 \mid H_1\right) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{1}{2\sigma^2}\left(\left(x_1-\mu_1\right)^2+\left(x_2-\mu_2\right)^2\right)\right)$$

- log-likelihood ratio:

$$\ln\Lambda\left(x_1, x_2\right) = -\frac{1}{2\sigma^2}\left(\left(x_1-\mu_1\right)^2+\left(x_2-\mu_2\right)^2-\left(x_1-\mu\right)^2-\left(x_2-\mu\right)^2\right)$$

$$= -\frac{1}{\sigma^2}\left(\underbrace{\frac{\mu_1^2+\mu_2^2}{2}-\mu^2}_{=:\,k}+\left(\mu\left(x_1+x_2\right)-\mu_1 x_1-\mu_2 x_2\right)\right)$$

Example: Detection of an Edge based on the values of two adjacent pixels (Additive Gaussian Noise) (cont.)

- let: $$\left.\begin{array}{l}\mu_1 = \mu - \Delta \\ \\ \mu_2 = \mu + \Delta\end{array}\right\} \Rightarrow \begin{array}{l}\mu = \frac{1}{2}(\mu_1 + \mu_2) \\ \\ \Delta = \frac{1}{2}(\mu_2 - \mu_1)\end{array} \quad \Rightarrow \quad k = \Delta^2$$

$$\Rightarrow \quad \ln\Lambda(x_1, x_2) = \frac{1}{\sigma^2}\left(\Delta(x_2 - x_1) - \Delta^2\right) \overset{H_1}{>} \lambda$$ for edge

- Leads to one of the classical edge detectors in image processing:

$$|x_2 - x_1| \overset{\text{edge}}{>} \frac{\lambda\sigma^2}{\Delta} + \Delta$$

=> i.e. edges are easier to detect for small σ or large Δ

=> Recap slide before, signal in Gaussian noise: $\Rightarrow \quad x \overset{H_1}{>} \frac{\sigma_n^2 \lambda}{s} + \frac{s}{2}$ for $s = 2\Delta$:

variance has *double* influence in edge detection, since *difference* of random variables x_1, x_2 is tested (see law of variance propagation)

Detection in the Case of Multivariate ND RVs

Karlsruher Institut für Technologie

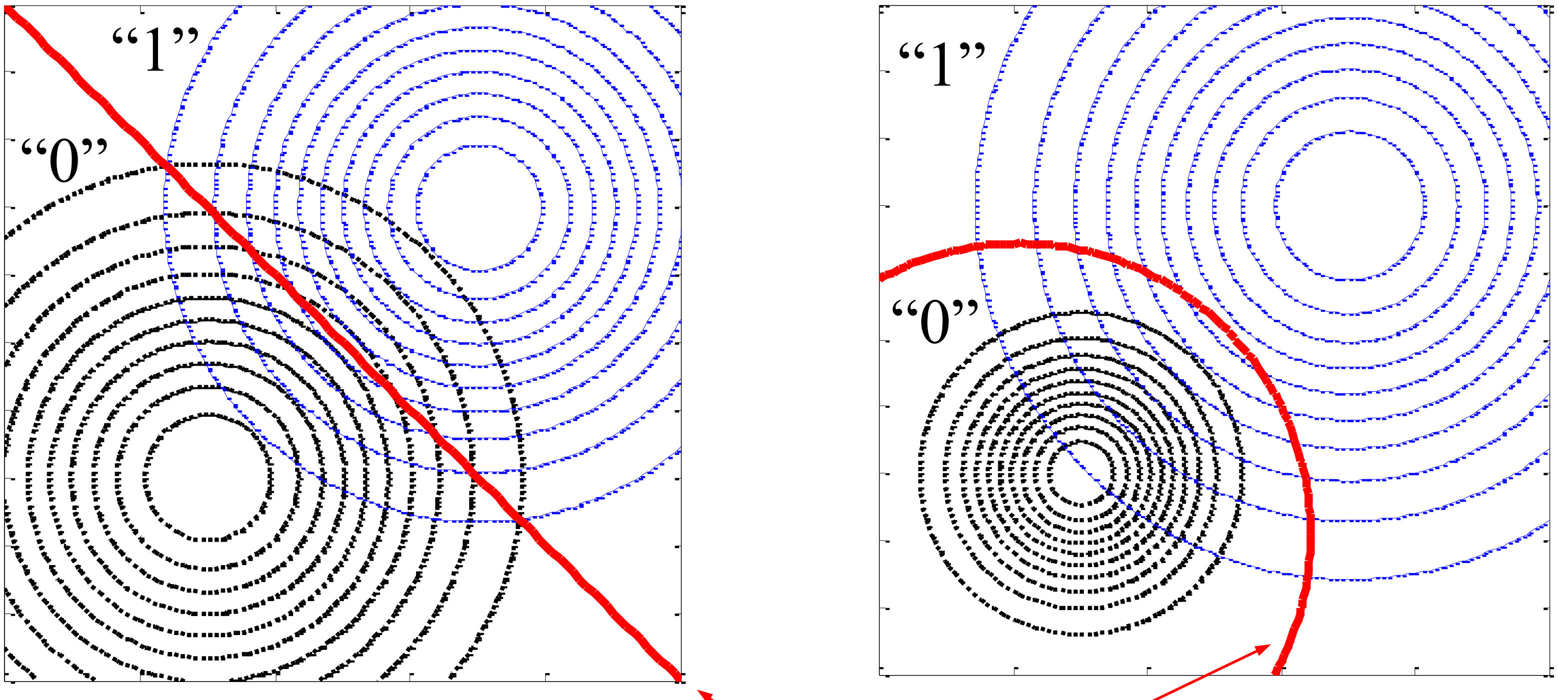
Notation for block matrix

$$f\left(\mathbf{x} \mid H_{0,1}\right) = \frac{1}{\sqrt{\left(2\pi\right)^n \det\left(\mathbf{C}_{\mathbf{xx},\mathbf{0},1}\right)}} \exp\left(-\frac{1}{2}\left(\mathbf{x}-\mathbf{m}_{\mathbf{0},1}\right)^T \mathbf{C}_{\mathbf{xx},\mathbf{0},1}^{-1}\left(\mathbf{x}-\mathbf{m}_{\mathbf{0},1}\right)\right)$$

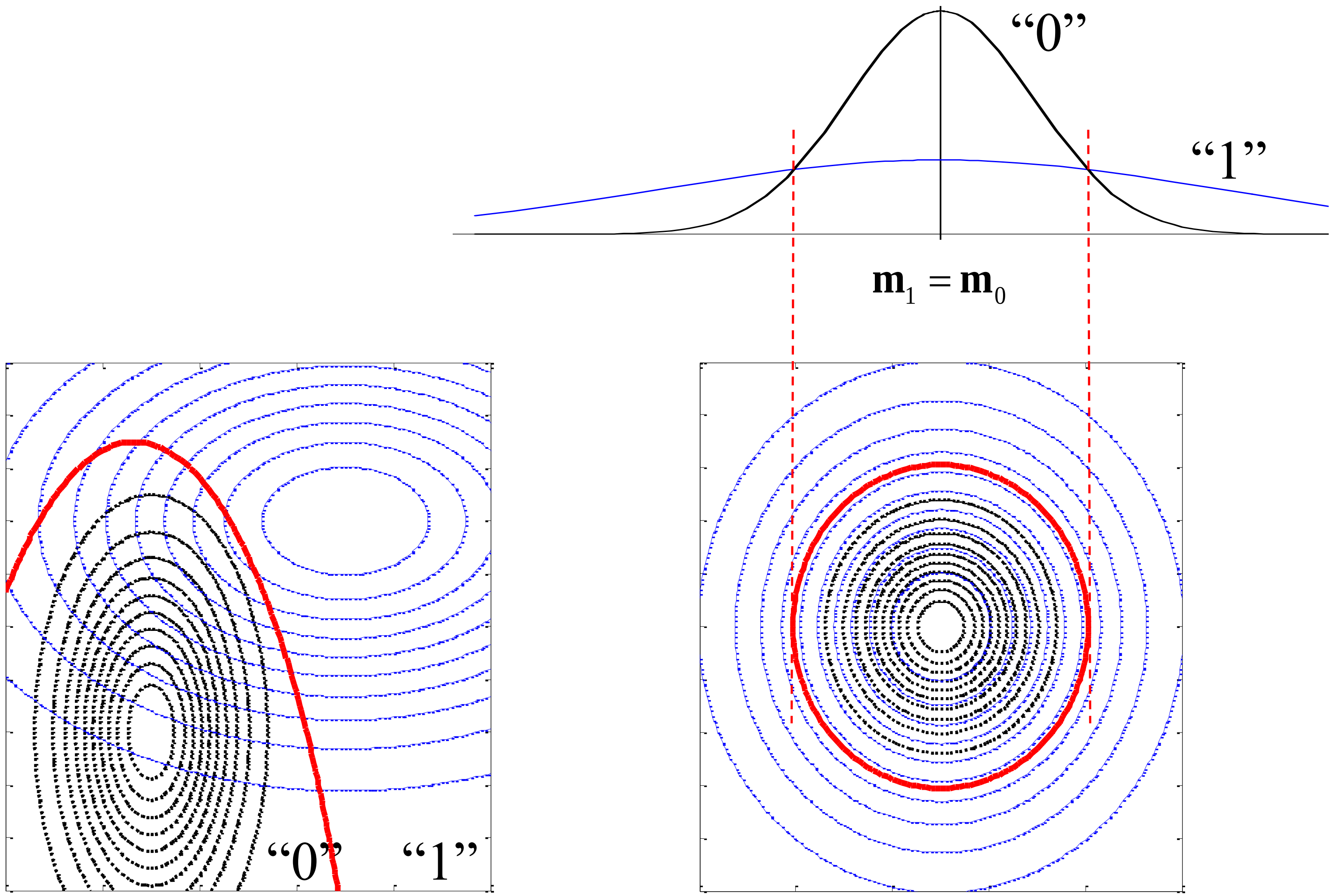
$$\ln\Lambda\left(\mathbf{x}\right) = -\frac{1}{2}\left(\ln\frac{\det\left(\mathbf{C}_{\mathbf{xx},1}\right)}{\det\left(\mathbf{C}_{\mathbf{xx},0}\right)} + \underbrace{\left(\mathbf{x}-\mathbf{m}_1\right)^T \mathbf{C}_{\mathbf{xx},1}^{-1}\left(\mathbf{x}-\mathbf{m}_1\right) - \left(\mathbf{x}-\mathbf{m}_0\right)^T \mathbf{C}_{\mathbf{xx},0}^{-1}\left(\mathbf{x}-\mathbf{m}_0\right)}\right)$$

quadratic form in $\mathbf{x}$ space

$\Rightarrow$ conic sections, e.g. planes, hyperboloids, ellipsoids …



$$\boxed{\left(\mathbf{x}-\mathbf{m}_1\right)^T \mathbf{C}_{\mathbf{xx},1}^{-1}\left(\mathbf{x}-\mathbf{m}_1\right)-\left(\mathbf{x}-\mathbf{m}_0\right)^T \mathbf{C}_{\mathbf{xx},0}^{-1}\left(\mathbf{x}-\mathbf{m}_0\right)=\text{const}}$$



“0”
“1”
$\mathbf{m}_1 = \mathbf{m}_0$
“0”
“1”

Detection of a (Known) Signal in Multivariate Noise

- Known signal

$$\mathbf{s} = \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{pmatrix} \qquad s_{\mathrm{i}} : \text{signal samples}$$

- Only one pdf (for “noise” or “background”)
- Noise: zero-mean ($m_0 = 0$), Gaussian with covariance $\mathbf{C}_{\mathbf{xx}}$
  => measurement and signal are shifted towards mean of noise beforehand

Detection of a (Known) Signal in Multivariate Noise (cont.)

Karlsruher Institut für Technologie

$$\ln \Lambda(\mathbf{x}) = -\frac{1}{2}\left((\mathbf{x}-\mathbf{s})^T \mathbf{C}_{\mathbf{xx}}^{-1}(\mathbf{x}-\mathbf{s}) - \mathbf{x}^T \mathbf{C}_{\mathbf{xx}}^{-1}\mathbf{x}\right) \quad \overset{H_1}{>} \lambda$$

$$\cancel{\mathbf{x}^T \mathbf{C}_{\mathbf{xx}}^{-1}\mathbf{x}} - \underbrace{\mathbf{s}^T \mathbf{C}_{\mathbf{xx}}^{-1}\mathbf{x}} - \underbrace{\mathbf{x}^T \mathbf{C}_{\mathbf{xx}}^{-1}\mathbf{s}} + \underbrace{\mathbf{s}^T \mathbf{C}_{\mathbf{xx}}^{-1}\mathbf{s}}_{\text{const.}} - \cancel{\mathbf{x}^T \mathbf{C}_{\mathbf{xx}}^{-1}\mathbf{x}} \quad \overset{H_1}{<} -2\lambda$$

=

$$\mathbf{s}^T \mathbf{C}_{\mathbf{xx}}^{-1}\mathbf{x} \quad \overset{H_1}{>} \quad \frac{1}{2}\mathbf{s}^T \mathbf{C}_{\mathbf{xx}}^{-1}\mathbf{s} + \lambda$$

- Detector:

$$\boxed{\mathbf{s}^T \mathbf{C}_{\mathbf{xx}}^{-1}\mathbf{x} \quad \overset{H_1}{>} \alpha}$$

Detection of a (Known) Signal in Multivariate Noise (cont.)

- Interpretation of the detector using

$$\mathbf{C}_{\mathbf{xx}}^{-1} = \mathbf{C}_{\mathbf{xx}}^{-\frac{1}{2}}\,\mathbf{C}_{\mathbf{xx}}^{-\frac{1}{2}} = \mathbf{C}_{\mathbf{xx}}^{-\frac{1}{2}T}\,\mathbf{C}_{\mathbf{xx}}^{-\frac{1}{2}}$$

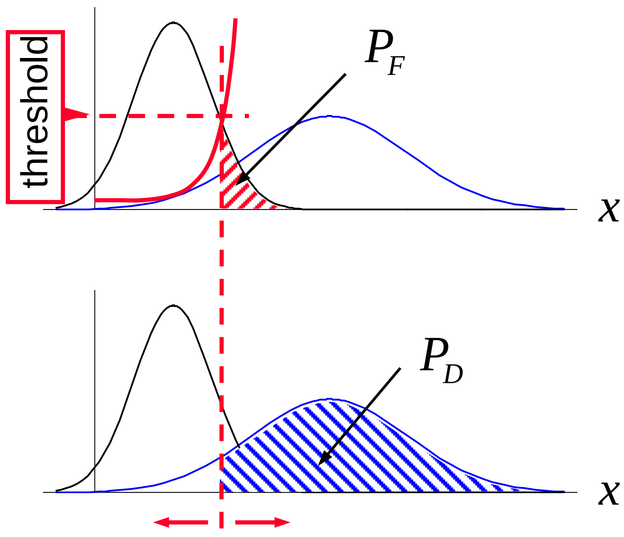
- Apply **whitening filter** , similar to **PCA**, see: $\mathbf{C}_{\mathbf{xx}}^{\frac{1}{2}}\,\mathbf{C}_{\mathbf{xx}}^{\frac{1}{2}} = \mathbf{P}\,\underbrace{\overbrace{\mathbf{\Lambda}_{\mathbf{xx}}^{\frac{1}{2}}}^{\text{Rotation and diagonal matrix}}\,\cancel{\mathbf{P}^{T}}\,\cancel{\mathbf{P}}\,\mathbf{\Lambda}_{\mathbf{xx}}^{\frac{1}{2}}}_{\mathbf{\Lambda}_{\mathbf{xx}}}\,\mathbf{P}^{T} = \mathbf{C}_{\mathbf{xx}}$

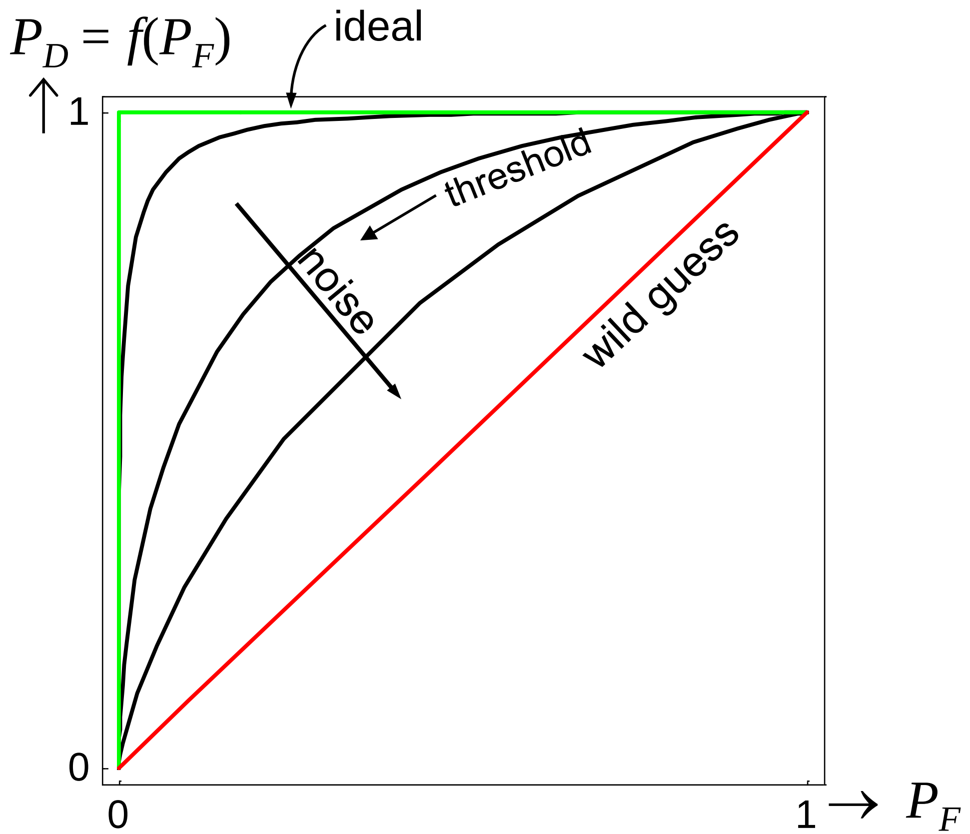
$$\left.\begin{aligned}\mathbf{x}' &= \mathbf{C}_{\mathbf{xx}}^{-\frac{1}{2}}\,\mathbf{x}\\ \mathbf{s}' &= \mathbf{C}_{\mathbf{xx}}^{-\frac{1}{2}}\,\mathbf{s}\end{aligned}\right\} \Rightarrow \quad \mathbf{s}^{T}\,\mathbf{C}_{\mathbf{xx}}^{-1}\,\mathbf{x} = \mathbf{s}^{T}\,\mathbf{C}_{\mathbf{xx}}^{-\frac{1}{2}T}\,\mathbf{C}_{\mathbf{xx}}^{-\frac{1}{2}}\,\mathbf{x} = \mathbf{s}'^{T}\mathbf{x}'$$

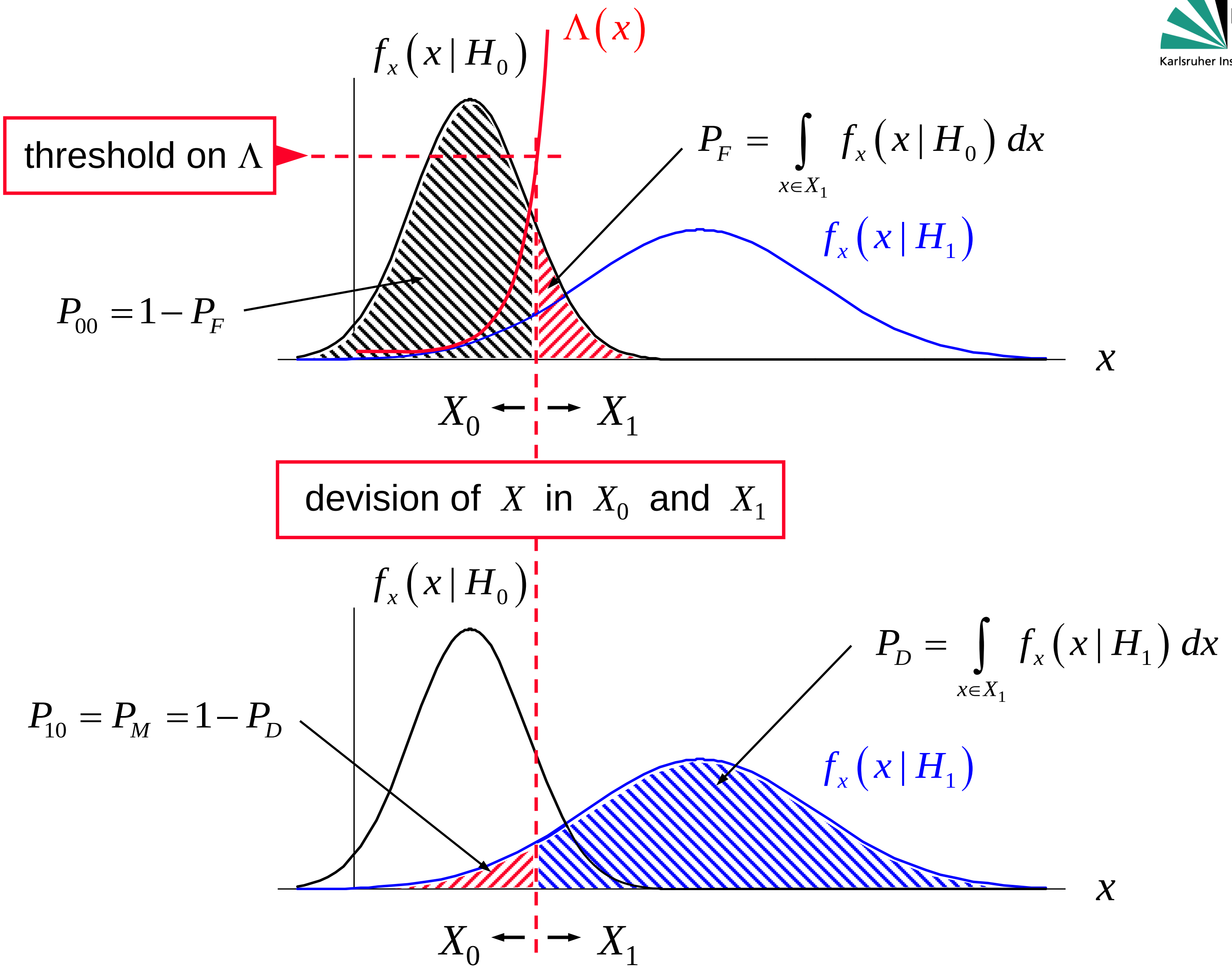
- Detector is scalar product of the whitened measurement and the whitened expected signal (= **matched filter**):

$$\mathbf{x}'\cdot\mathbf{s}' \quad \overset{H_1}{>} \alpha'$$

Receiver Operation Characteristics (ROC) Curve






$\Lambda(x)$
$f_x(x|H_0)$
threshold on Λ
$P_F = \int_{x \in X_1} f_x(x|H_0)\,dx$
$f_x(x|H_1)$
$P_{00} = 1 - P_F$
x
X_0
X_1
devision of X in X_0 and X_1
$f_x(x|H_0)$
$P_D = \int_{x \in X_1} f_x(x|H_1)\,dx$
$P_{10} = P_M = 1 - P_D$
$f_x(x|H_1)$
x
X_0
X_1

true is decision for

$$
\begin{tikzcd}[column sep=huge, row sep=huge]
H_0 \arrow[r, "P_{00}"] \arrow[dr, "P_\mathrm{F}"'] & H_0 \quad \text{i.e.} \quad \mathbf{x} \in X_0 \\
H_1 \arrow[r, "P_\mathrm{D}"'] \arrow[ur, "P_\mathrm{M}"' near end] & H_1 \quad \text{i.e.} \quad \mathbf{x} \in X_1
\end{tikzcd}
$$

$$
\Rightarrow \quad \begin{aligned} P_{00} &= 1 - P_F \\ P_M &= 1 - P_D \end{aligned}
$$

$$
P_F = \int_{\mathbf{x} \in X_1} f_{\mathbf{x}}\left(\mathbf{x} \mid H_0\right) d\mathbf{x}
$$

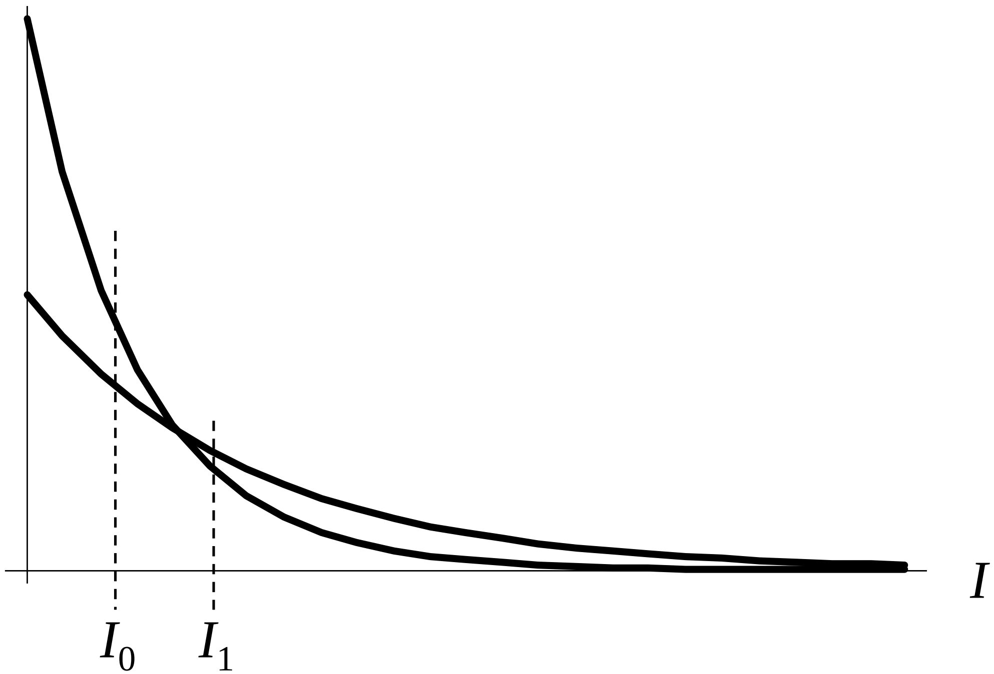
$$
P_D = \int_{\mathbf{x} \in X_1} f_{\mathbf{x}}\left(\mathbf{x} \mid H_1\right) d\mathbf{x}
$$

$$
P_{ij} = \int_{\mathbf{x} \in X_j} f_x\left(\mathbf{x} \mid H_i\right) d\mathbf{x} \qquad i, j \in \{0, 1\}
$$

Example: ROC for Decision Between Two Levels of „Brightness“ in a SAR Image

Brightness levels: I_0 and I_1

$$f_I\left(I \mid H_0\right)=\frac{1}{I_0}\exp\left(-\frac{I}{I_0}\right); \qquad f_I\left(I \mid H_1\right)=\frac{1}{I_1}\exp\left(-\frac{I}{I_1}\right) \qquad \forall \;\; I \geq 0$$



Example: ROC for Decision Between Two Levels of „Brightness“ in a SAR Image (cont.)

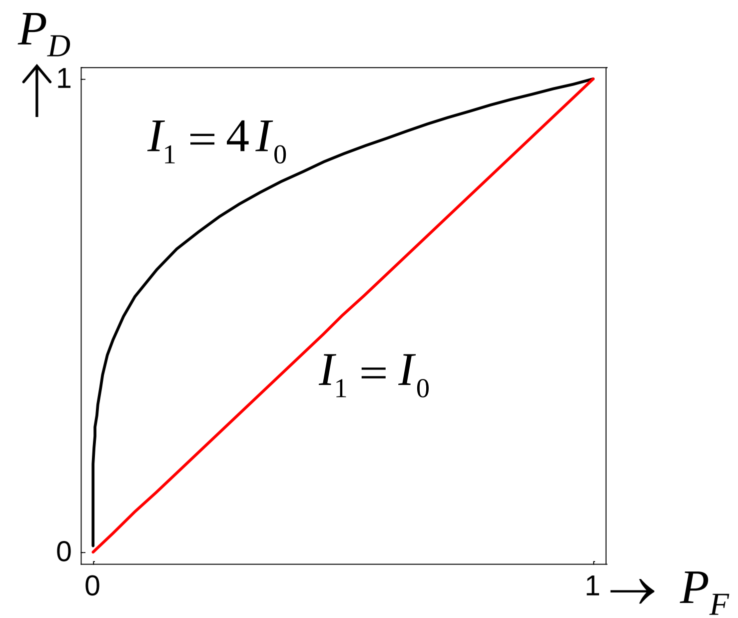
Threshold on I: β

$$P_D = \int_{\beta}^{\infty} \frac{1}{I_1} \exp\left(\frac{-I}{I_1}\right) dI = \frac{1}{I_1}\left(-I_1 \exp\left(\frac{-I}{I_1}\right)\right)\Bigg|_{\beta}^{\infty} = \exp\left(-\frac{\beta}{I_1}\right)$$

$$P_F = \int_{\beta}^{\infty} \frac{1}{I_0} \exp\left(\frac{-I}{I_0}\right) dI = \exp\left(-\frac{\beta}{I_0}\right)$$

eliminate β

$$P_D = \left(P_F\right)^{\frac{I_0}{I_1}} \qquad \frac{I_0}{I_1} < 1$$



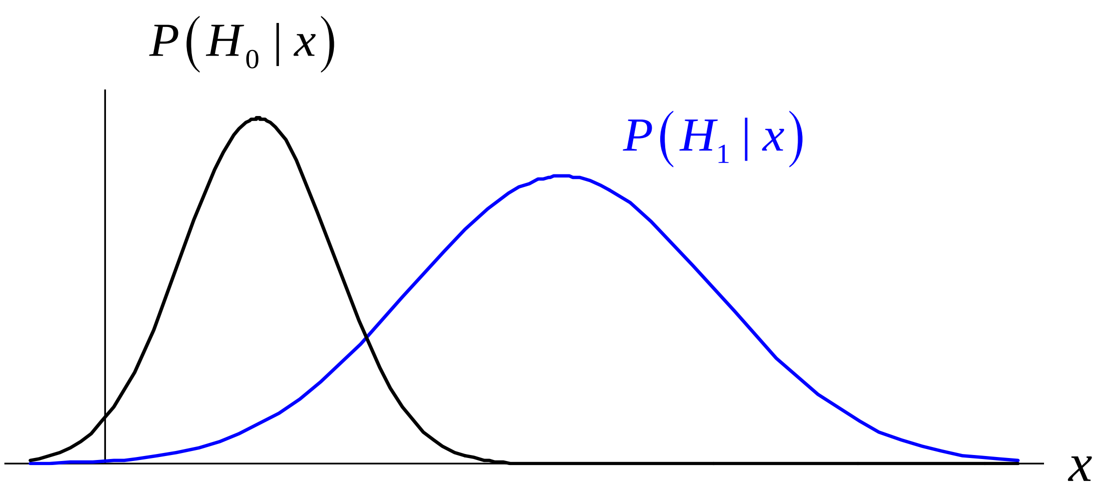
Likelihood Ratio: Incorporating Prior Knowledge

- Known prior probabilities: $P(H_0) = P_0$
  $P(H_1) = P_1$

- Posterior probabilities (Bayes‘ theorem):

$$P(H_0 \mid \mathbf{x}) = \frac{f_{\mathbf{x}}(\mathbf{x} \mid H_0) \cdot P_0}{f_{\mathbf{x}}(\mathbf{x})}$$

$$P(H_1 \mid \mathbf{x}) = \frac{f_{\mathbf{x}}(\mathbf{x} \mid H_1) \cdot P_1}{f_{\mathbf{x}}(\mathbf{x})}$$



Compare to likelihoods, here: $P_1 > P_0$

NB: $\int_{-\infty}^{\infty} P(H_{0,1} \mid x)\,dx \neq 1$

Bayesian Decision Rule incl. Prior Knowledge

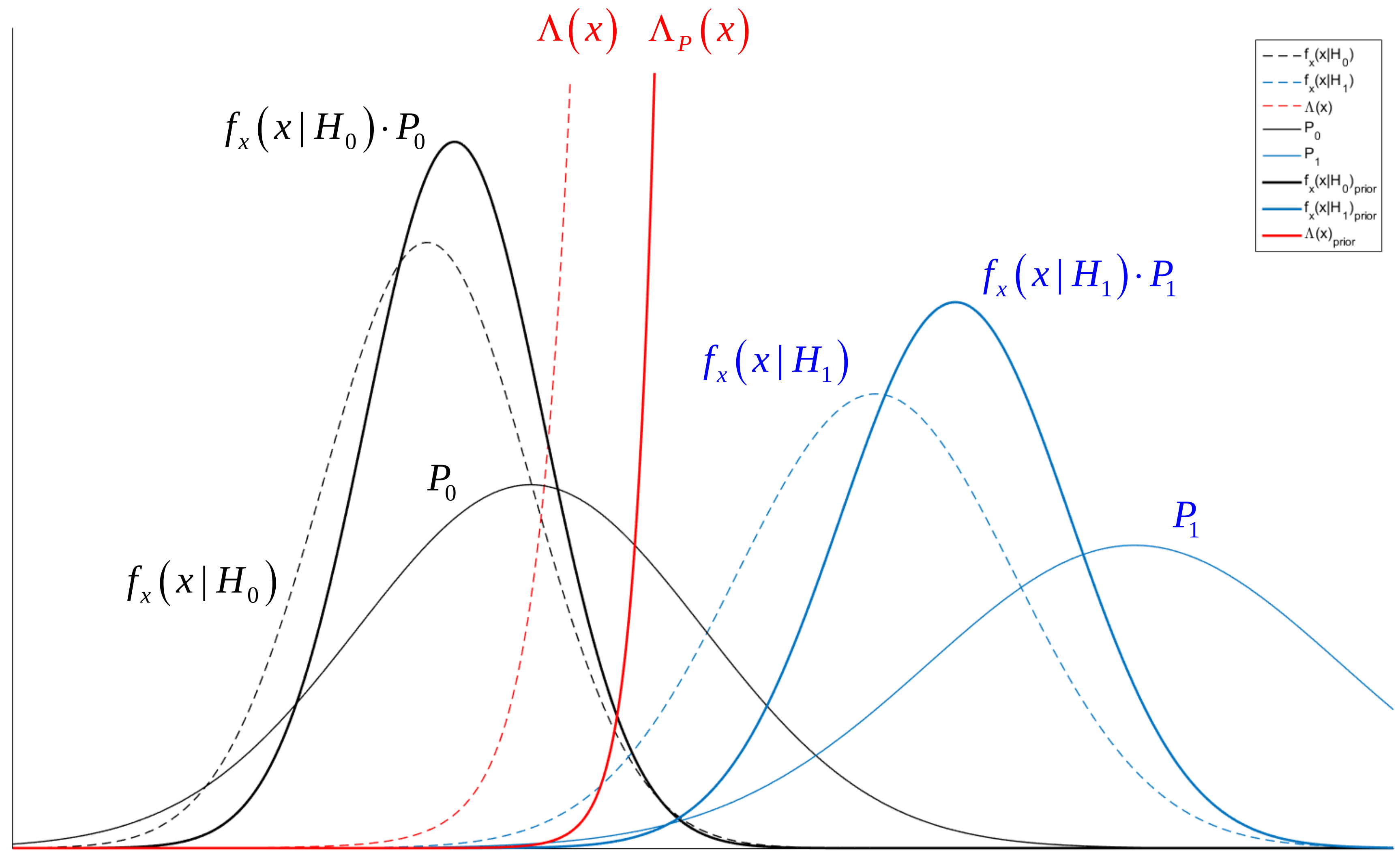
- Ratio of posterior probabilities:

$$\frac{P(H_1 \mid \mathbf{x})}{P(H_0 \mid \mathbf{x})} = \frac{f_{\mathbf{x}}(\mathbf{x} \mid H_1) \cdot P_1}{f_{\mathbf{x}}(\mathbf{x} \mid H_0) \cdot P_0} = \Lambda(\mathbf{x}) \cdot \frac{P_1}{P_0}$$

- Decide for H_1 if $P(H_1 \mid \mathbf{x}) > P(H_0 \mid \mathbf{x})$, i.e. $\frac{P(H_1 \mid \mathbf{x})}{P(H_0 \mid \mathbf{x})} \overset{H_1}{>} 1$

Bayesian decision rule: $\Lambda(\mathbf{x}) \overset{H_1}{>} \frac{P_0}{P_1}$

Influence of Prior Knowledge on Decision



Consideration of Costs in Bayesian Decision

	true hypothesis	a priori probability	decision for hypothesis	cost	conditional probability
correct	H_0	P_0	H_0	C_{00}	P_{00}
false alarm	$\boldsymbol{H_0}$	$\boldsymbol{P_0}$	$\boldsymbol{H_1}$	$\boldsymbol{C_{01}}$	$\boldsymbol{P_{01} = P_\mathrm{F}}$
missed hit	H_1	P_1	H_0	C_{10}	$P_{10} = P_\mathrm{M}$
correct detection	$\boldsymbol{H_1}$	$\boldsymbol{P_1}$	$\boldsymbol{H_1}$	$\boldsymbol{C_{11}}$	$\boldsymbol{P_{11} = P_\mathrm{D}}$

P_F : “false alarm probability” or “false alarm rate”
P_D : “detection probability”

Total Expected Cost of Decision

- Plausible assumptions: $C_{01} > C_{00}$

  $C_{10} > C_{11}$

- Total cost:

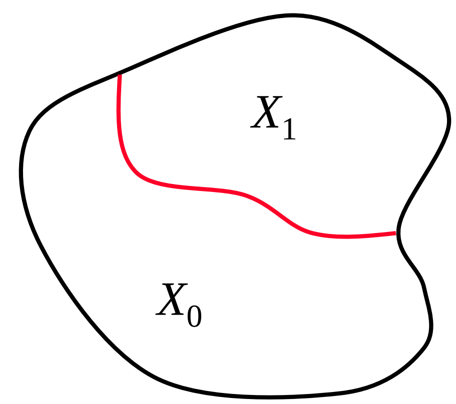
$$
\begin{aligned}
C &= C_{00}\,P_0\,P_{00} \;+\; C_{01}\,P_0\,P_F \;+\; C_{10}\,P_1\,P_M \;+\; C_{11}\,P_1\,P_D \\
&= C_{00}\,P_0\left(1-P_F\right) \;+\; C_{01}\,P_0\,P_F \;+\; C_{10}\,P_1\left(1-P_D\right) \;+\; C_{11}\cdot P_1\cdot P_D \\
&= \underbrace{C_{00}\,P_0 \;+\; C_{10}\,P_1}_{\text{independent of data, we cannot influence these costs}} \;+\; \underbrace{\left(C_{01}-C_{00}\right)}_{>0} P_0\;P_F \;-\; \underbrace{\left(C_{10}-C_{11}\right)}_{>0} P_1\;P_D \qquad \to \min
\end{aligned}
$$

Strategy 1: Minimize Total Expected Cost

- Choose the separation (e.g. threshold) between spaces X_0 and X_1 such that

$$\underbrace{\left(C_{01}-C_{00}\right)}_{>0} P_0\ P_F\ -\ \underbrace{\left(C_{10}-C_{11}\right)}_{>0} P_1\ P_D \quad \rightarrow \min$$

$$\Rightarrow \quad \left.\begin{array}{l} P_F \rightarrow \min \\ P_D \rightarrow \max \end{array}\right\} \text{ mutually exclusive}$$



- “Compromise” required between low P_F and high P_D

Minimize Total Expected Cost (cont.)

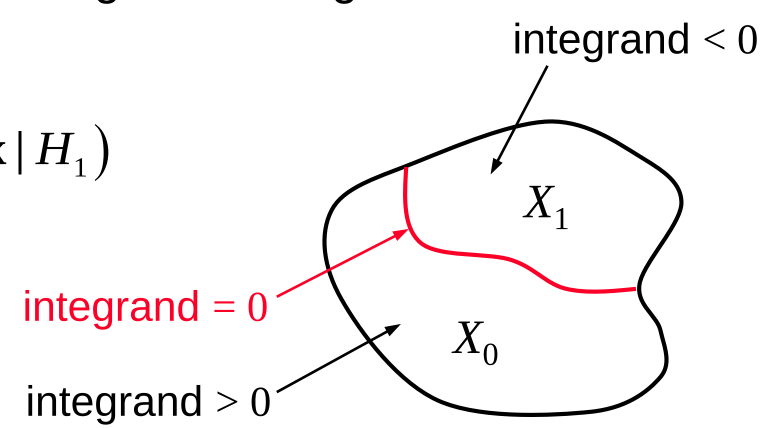
$$\left(C_{01}-C_{00}\right) P_0\, P_F \;-\; \left(C_{10}-C_{11}\right) P_1\, P_D$$

Replace probability by integral over pdf

$$= \int_{X_1} \Big(\underbrace{\left(C_{01}-C_{00}\right) P_0}_{>0}\; f_{\mathbf{x}}\left(\mathbf{x}\,|\,H_0\right) \;-\; \underbrace{\left(C_{10}-C_{11}\right) P_1}_{>0}\; f_{\mathbf{x}}\left(\mathbf{x}\,|\,H_1\right) \Big)\, d\mathbf{x} \quad \to \min$$

$\Rightarrow$ Choose X_1 as the set of those $\mathbf{x}$ where the integrand is negative,

i.e. decide for H_1, if
$\left(C_{01}-C_{00}\right) P_0\, f_{\mathbf{x}}\left(\mathbf{x}\,|\,H_0\right) < \left(C_{10}-C_{11}\right) P_1\, f_{\mathbf{x}}\left(\mathbf{x}\,|\,H_1\right)$



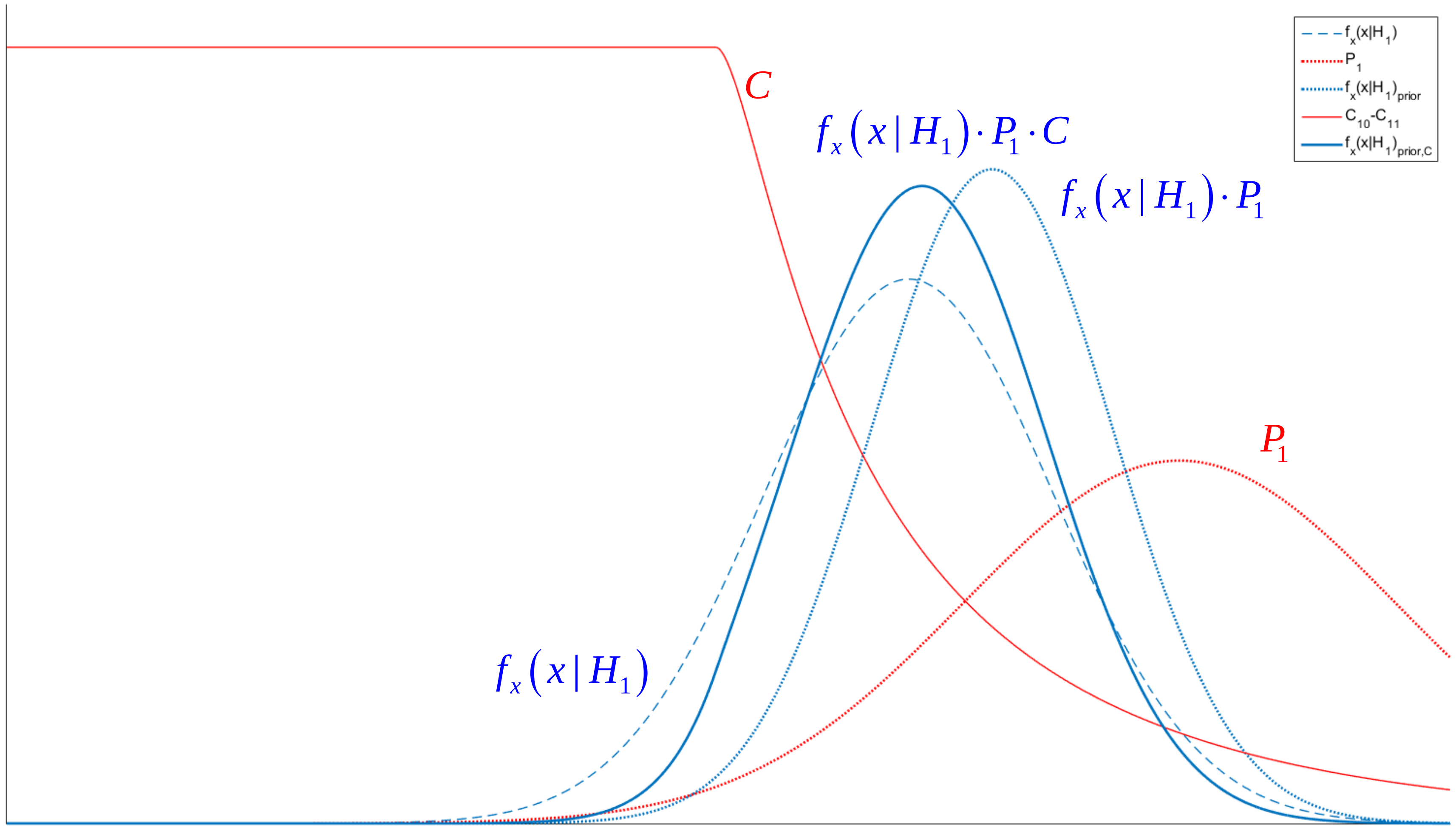
Decision Rule for Minimization of Total Expected Cost

$$\Rightarrow \quad \boxed{\Lambda(\mathbf{x}) = \frac{f_{\mathbf{x}}(\mathbf{x} \mid H_1)}{f_{\mathbf{x}}(\mathbf{x} \mid H_0)} \overset{H_1}{>} \frac{C_{01} - C_{00}}{C_{10} - C_{11}} \frac{P_0}{P_1}}$$

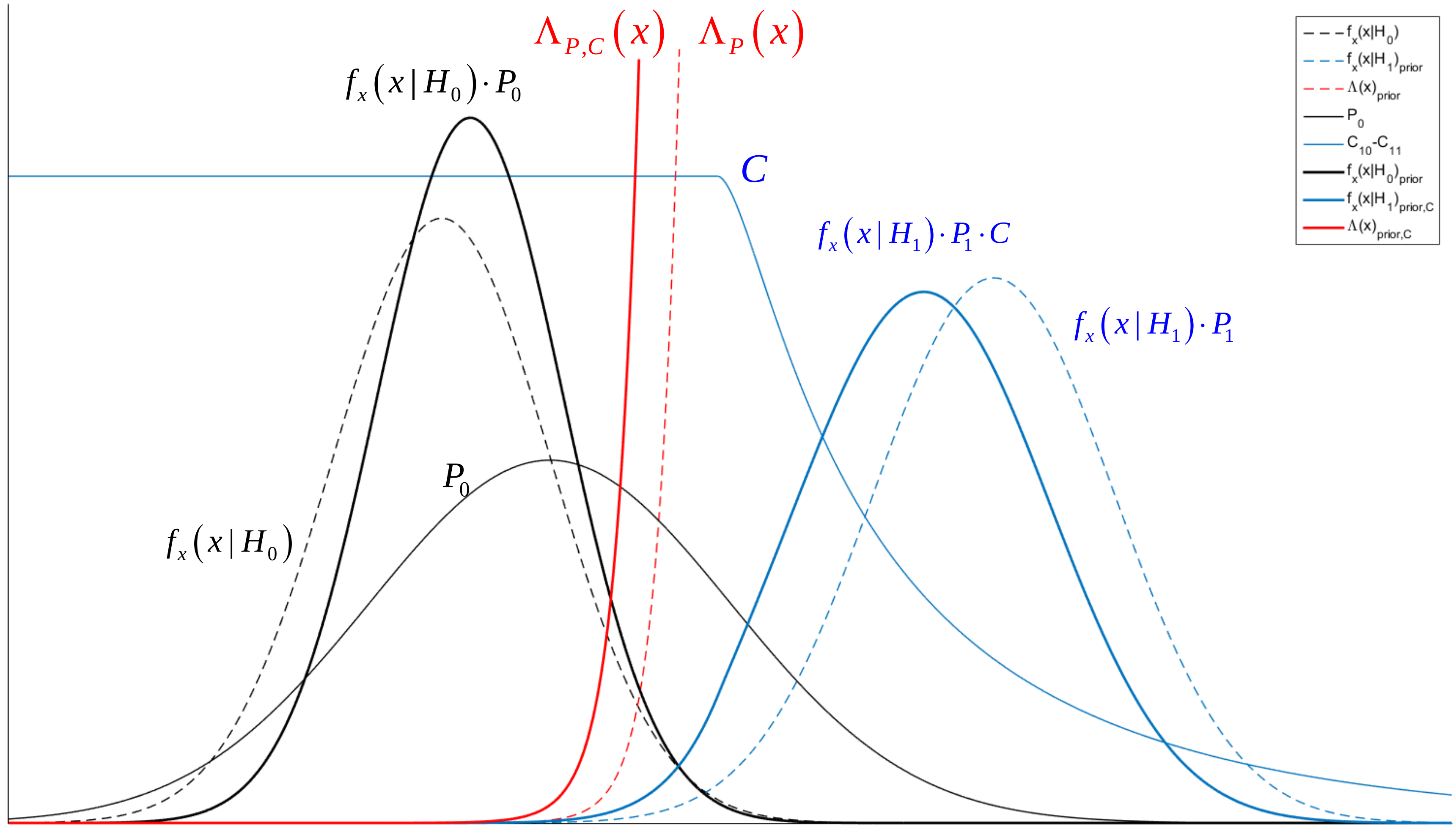
given by the measurement system (noise)

known a priori or arbitrarily chosen

Influence of Prior Knowledge and Costs on PDF



Influence of Prior Knowledge and Costs on Decision



Example: $C = C_{10}$, i.e. missed hits are penalized (e.g. detection of corsairs)

Relation between cost minimization and minimization of Error Probability

- Useful, if costs are not known (or symmetric)

- Minimization of Error Probability can be cast into the theory of cost minimization by setting

$$\left.\begin{array}{ll} C_{00} = C_{11} = 0 & \text{no cost for correct decision} \\ C_{01} = C_{10} = C_e & \text{identical costs for wrong decision} \end{array}\right\} \Rightarrow \frac{C_{01} - C_{00}}{C_{10} - C_{11}} = 1$$

$$\Rightarrow \boxed{\Lambda(\mathbf{x}) \overset{H_1}{>} \frac{P_0}{P_1} = \frac{1-P_1}{P_1}}$$

e.g.: $P_1 = P_0 = \frac{1}{2} \quad \Rightarrow \quad \Lambda(\mathbf{x}) \overset{H_1}{>} 1$

Strategy 2: Minimax Test

- Prior probabilities P_0 and P_1 are often unknown
- Strategy: pessimistic approach, i.e. <u>mini</u>mum cost <u>maxi</u>mization
- Assumption (for simplicity's sake): $C_{00} = C_{11} = 0$ (correct decision)

$\Rightarrow$ $$C = C_{01} P_0 P_F + C_{10} P_1 P_M \qquad \text{where} \qquad P_0 = 1 - P_1$$

$$= C_{01} P_F + P_1 \left(C_{10} P_M - C_{01} P_F \right)$$

Minimax Test (cont.)

- Given the likelihoods $f_{\mathbf{x}}\left(\mathbf{x} \mid H_0\right)$ and $f_{\mathbf{x}}\left(\mathbf{x} \mid H_1\right)$
- For any threshold λ (treated as free parameter) in the log-likelihood test we get

  $P_F\left(\lambda\right)$ and $P_M\left(\lambda\right)$
- These result in a total cost, that still depends (linearly) on the unknown P_1:

$$C\left(\lambda, P_1\right) = C_{01} P_F\left(\lambda\right) + P_1\left(C_{10} P_M\left(\lambda\right) - C_{01} P_F\left(\lambda\right)\right)$$

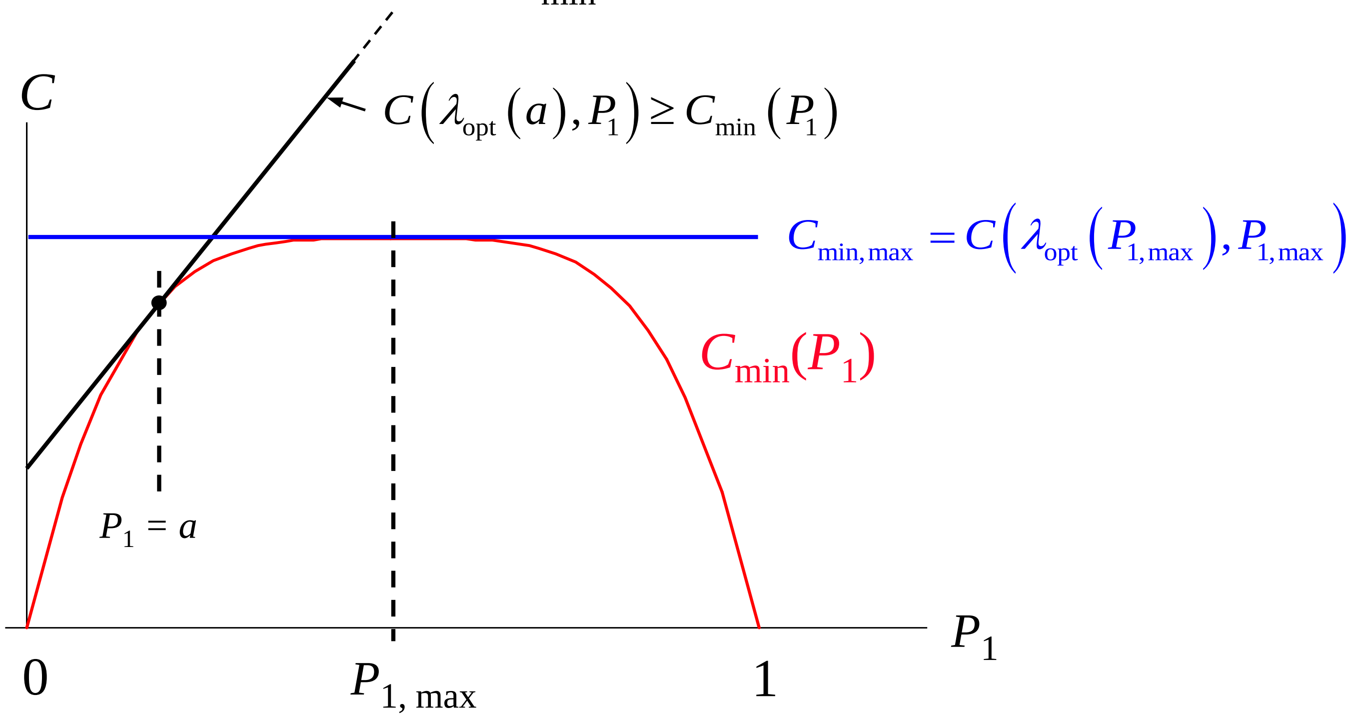
- P_1 is bounded by [0 ; 1]
- For every $0 \le P_1 \le 1$ determine numerically or analytically the optimum threshold $\lambda_{opt}(P_1)$ for minimum cost

$$C_{\min}\left(P_1\right) = C\left(\lambda_{opt}\left(P_1\right), P_1\right)$$

- Find the value of $P_{1,\max}$ with the largest minimum cost $C_{\min,\max}$ and choose this threshold for decision: $\lambda = \lambda_{opt}\left(P_{1,\max}\right)$

Minimax Test (cont.)

$$\left.\begin{array}{l} P_1 = 0 \Rightarrow \text{always} \quad H_0 \Rightarrow C_{\min} = 0 \\ P_1 = 1 \Rightarrow \text{always} \quad H_1 \Rightarrow C_{\min} = 0 \\ \qquad\qquad\qquad\qquad\qquad C_{\min} \geq 0 \end{array}\right\} \text{maximum of } C_{\min}(P_1) \text{ at } 0 < P_1 < 1$$



Minimax Test (cont.)

- We find $P_{1,\max}$ from

$$\frac{\partial}{\partial P_1} C_{\min}\left(P_1\right) = 0$$

- Leading to the **minimax decision rule**:

$$\boxed{C_{10} \cdot P_M = C_{01} \cdot P_F}$$

- Properties of this solution:
  - independent of P_1
  - often not analytically solvable
  - equal conditional costs for both wrong decisions

Strategy 3: Neyman-Pearson Criterion

- Prior probabilities P_0, P_1 unknown
- Costs C_{ij} unknown
- Strategy: Maximize the detection probability, given a maximum allowable false alarm rate, i.e.:

maximize $$P_D = \int_{X_1} f_x(\mathbf{x} \mid H_1)\, d\mathbf{x}$$

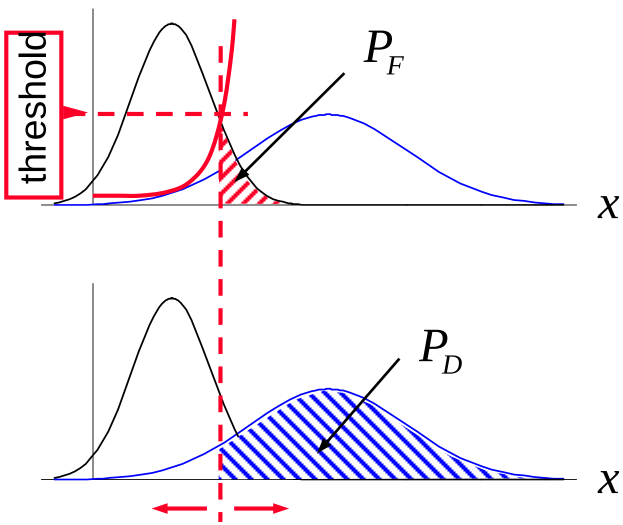
for $$P_F = \int_{X_1} f_x(\mathbf{x} \mid H_0)\, d\mathbf{x} \leq P_{F,\max}$$

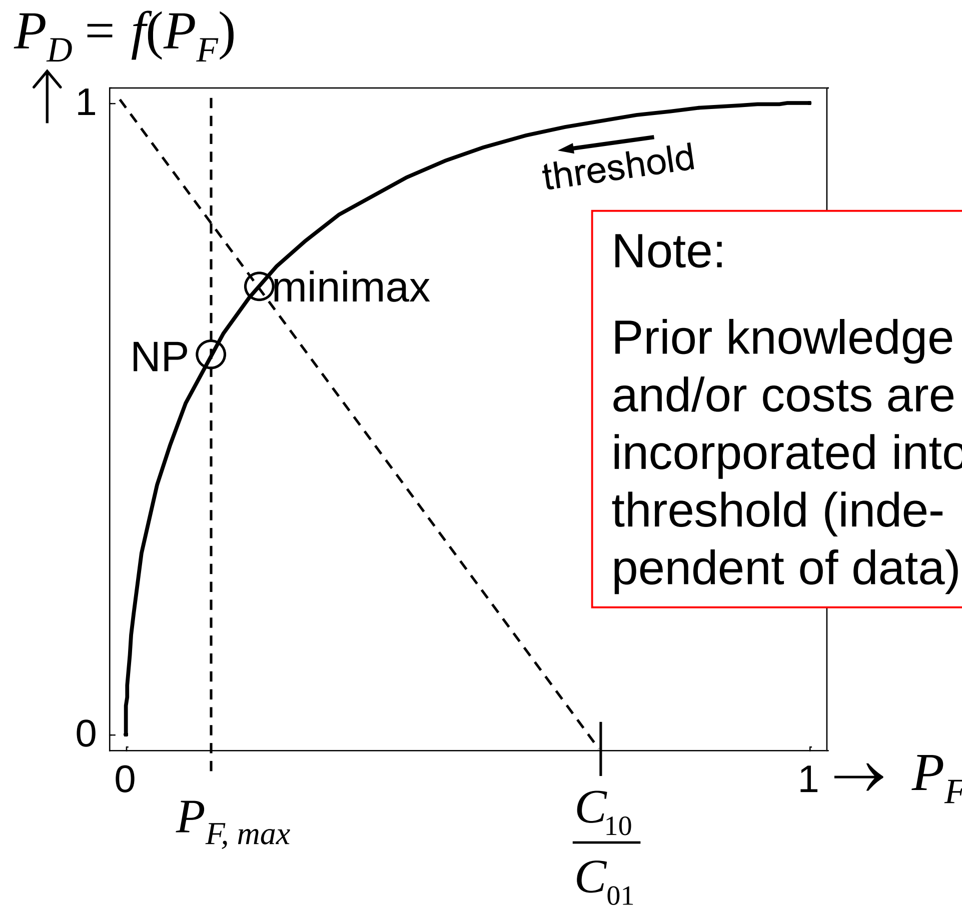
$\Rightarrow \quad \ln \Lambda(\mathbf{x}) \overset{H_1}{>} \lambda$ with λ chosen, such that $P_F = P_{F,\max}$

- Often: “Constant False Alarm Rate” (CFAR) detector

Recall: Receiver Operation Characteristics (ROC) Curve

- Parametric plot illustrating the trade-off between false alarm and detection rates





Note:

Prior knowledge and/or costs are incorporated into threshold (inde-pendent of data)

Minimax: $P_M = \frac{C_{01}}{C_{10}} \cdot P_F \quad \Rightarrow \quad P_D = 1 - P_M = 1 - \frac{C_{01}}{C_{10}} \cdot P_F$ for $P_D = 0 \Rightarrow P_F = \frac{C_{10}}{C_{01}}$

for $P_F = 0 \Rightarrow P_D = 1$

Concluding Remark

- So far, we have assumed that we know $f\left(\mathbf{x} \mid H_0\right)$ and $f\left(\mathbf{x} \mid H_1\right)$.
- In practice, this is not always the case.

$\Rightarrow$ $f\left(\mathbf{x} \mid H_0\right)$ and $f\left(\mathbf{x} \mid H_1\right)$ must be estimated from test data.

- See: Lecture module on classification.