AWGN case

$$H_0 : \quad \begin{array}{ll} \text{Case } a_0 & f'_i = \overset{f_i + n_i}{\overset{\uparrow}{x_i}} \\ \text{Case } b_0 & f'_i = x_i + \gamma v_i \qquad (v \neq w) \end{array}$$

$$H_1 : \quad f'_i = x_i + \gamma w_i$$

<u>Note</u> cases a_0 and b_0 can be Treated Together if v is allowed To coincide with the null sequence.

The likelihood ratio is

$$\ell(f') = \frac{p(f'|w)}{\int_{R^n} p(f'|v)\,p(v)\,dv} \qquad \circledast$$

$\mathbb{R}^n$ perchè w è un insieme di misura zero in R^n, quindi

$$\int_{R^n} = \int_{R^n - w}$$

numerator

X is stationary ($f_i + n_i$ identically distributed random variables), white, normally distribute sequence <u>varianza σ_x^2</u> <u>media μ_x</u>

$$p(f'|w) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_x^2}} \exp\left(-\frac{(f'_i - \mu_x - \gamma w_i)^2}{2\sigma_x^2}\right)$$

$$p(f'|H_0) = \prod_{i=1}^{n} \int_{\mathbb{K}} \frac{1}{\sqrt{2\pi\sigma_x^2}} \exp\left(-\frac{(f_i' - \mu_x - \gamma v_i)^2}{2\sigma_x^2}\right) p(v_i)\,dv$$

usually $p(f'|H_0) = p(f'|0)$

$\sigma_x \gg \gamma v_i$ derives fro
the imperceptibility
constraint

$$p(f'|H_0) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_x^2}} \exp\left(-\frac{(f_i' - \mu_x)^2}{2\sigma_x^2}\right) = p(f'|$$

$$\Longrightarrow \quad l(f') = \frac{p(f'|w)}{p(f'|0)} = \frac{\prod_{i=1}^{n} \exp\left(\frac{-(f_i' - \mu_x - \gamma w_i)^2}{2\sigma_x^2}\right)}{\prod_{i=1}^{n} \exp\left(-\frac{(f_i' - \mu_x)^2}{2\sigma_x^2}\right)}$$

logarithmic formulation

$$\mathcal{L}(f') = \sum_{i=1}^{n} \frac{1}{2\sigma_x^2}\left[(f_i' - \mu_x)^2 - (f_i' - \mu_x - \gamma w_i)^2\right]$$

$$= \frac{1}{2\sigma_x^2}\left[\underline{\underline{\sum_{i=1}^{n} 2\gamma f_i' w_i}} - \sum_{i=1}^{2} 2\mu_x \gamma w_i + \right.$$

$$\left. - \sum_{i=1}^{n} \gamma^2 w_i^2 \right]$$

the only term that depends on f_i'

linear correlation between f' an

$$\Longrightarrow \quad \rho = \frac{1}{n}\sum_{i=1}^{n} f'_i w_i = \frac{f' \cdot w}{\boxed{n}} \longrightarrow$$ per comodità nei calcoli

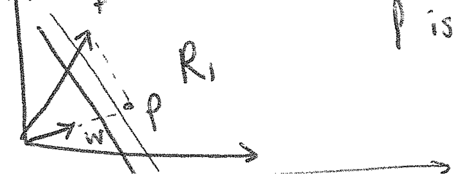
is sufficient statistic for watermarking detection

In order to decide whether a given watermark is present in A' or not, the detector needs only to look at the correlation between the to-be-searched watermark and the host feature vector extracted from A', and compare it against a **detection threshold T_ρ**

We apply Neyman-Pearson criterion

$$\int_{T_\rho}^{\infty} p(\rho \mid H_0)\, d\rho = \bar{P}_f$$

→ target value



ρ is a scaled projection of f' on w

When computing the false detection probability for setting T_ρ at the detector the watermark signal is known

→ we have to average over all possible host assets

When evaluating the performances of the whole watermarking system, we have to average over all possible watermarks as well

watermark samples zero mean i.i.d. random variab

$\left.\begin{array}{l}\text{attack noise}\\ \text{host feature}\end{array}\right\}$ i.i.d. normal variables

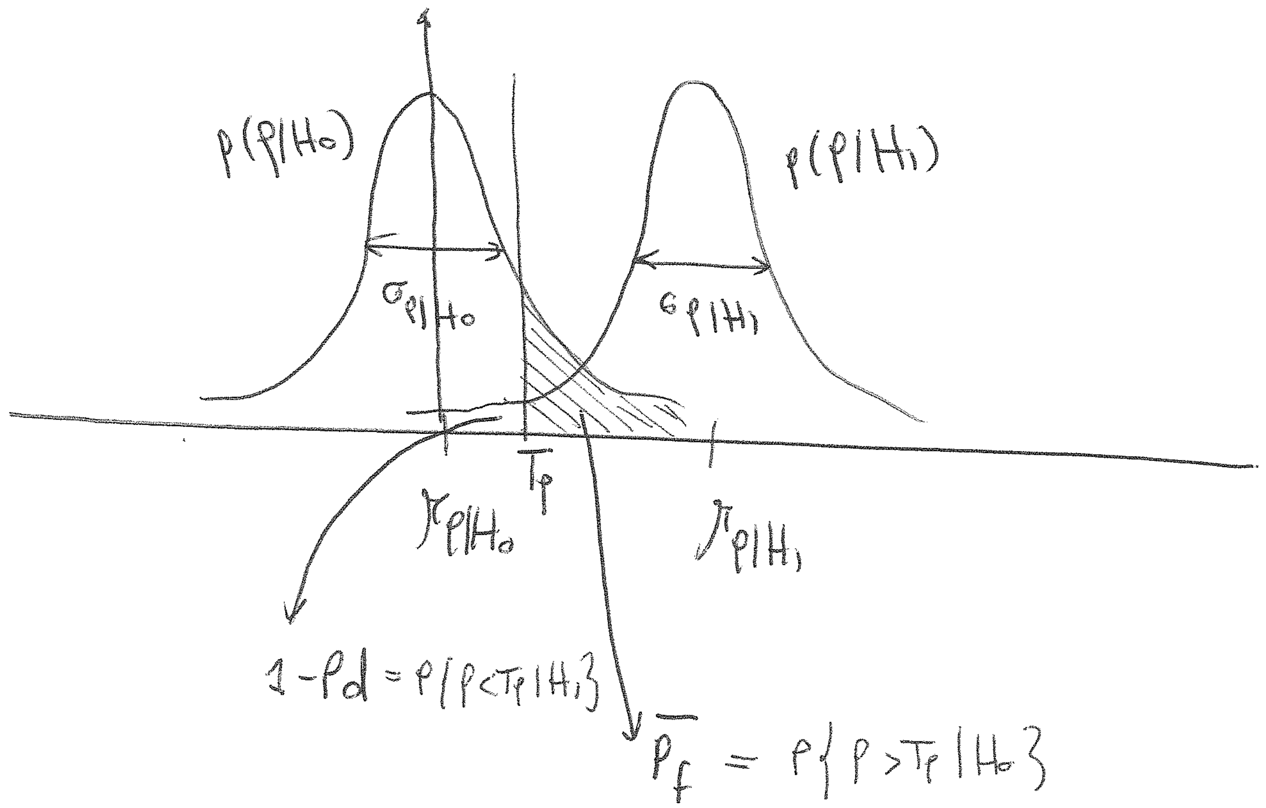
→ ρ normal distributes

→ it is sufficient to estimate its mean and variance

$p(\rho|H_0)$: $\mu_{\rho|H_0}$ and $\sigma^2_{\rho|H_0}$

$p(\rho|H_1)$: $\mu_{\rho|H_1}$ and $\sigma^2_{\rho|H_1}$

Devo stabilire la soglia T_ρ



Ho

$$\mu_{\rho|H_0} = E[\rho|H_0] \overset{f_i'=x_i}{=} \frac{1}{n}E\left[\sum x_i w_i\right] \overset{\text{marchio fisso}}{=} \frac{1}{n}\sum E[x_i]\, w_i$$

$$= \mu_x \sum \frac{w_i}{n} \underset{\text{zero-mean}}{=} 0$$

$$\sigma^2_{\rho|H_0} = var\left(\frac{1}{n}\sum x_i w_i\right) = \frac{1}{n^2}\sigma_x^2 \sum w_i^2 \cong \frac{\sigma_x^2\,\sigma_w^2}{n}$$

$$\sigma_x^2 = \sigma_{f_i}^2 + \sigma_n^2$$

for large values of n

$$\frac{1}{n}\sum w_i^2 = \frac{\|w\|^2}{n} \cong E[w^2$$

$$\| $$

$$\sigma_w^2$$

H_1

$$\mu_{\rho|H_1} = \frac{1}{n} E\left[\sum (x_i + \gamma w_i) w_i\right] = \mu_x 0 + \gamma \sigma_w^2 = \gamma \sigma_w^2$$

$$\sigma^2_{\rho|H_1} = var\left(\frac{1}{n} \sum (x_i + \gamma w_i) w_i\right] = \frac{1}{n} \sigma_x^2 \sigma_w^2$$

NB devo Togliere $\mu^2_{\rho|H_1}$ nel calcolo $\Rightarrow$ .. $\gamma^2 \sigma_w^4 - \gamma^2 \sigma$

$$\overline{P_f} = \int_{T_\rho}^{\infty} p(\rho|H_0)\, d\rho = \frac{1}{2} erfc\left(\sqrt{\frac{(T_\rho - \mu_{\rho|H_0})^2}{2\sigma^2_{\rho|H_0}}}\right)$$

⇝ invertendo trovo T_ρ

Fissata T_ρ per diminuire $1 - P_d$ posso spostare le gaussiane e aumentare $\mu_{\rho|H_1}$ ma così il watermark diviene visibile.

Posso anche diminuire $\sigma^2_{\rho|H_1}$ e $\sigma^2_{\rho|H_0}$ aumentando n e migliorare le performance quanto voglio.

$$T_\rho = \sqrt{2}\, \sigma_{\rho|H_0}\, \text{erfc}^{-1}(2\bar{P}_f) + \mu_{\rho|H_0}$$

In order To evaluate the probability of missing the watermark presence

$$1-P_d = \frac{1}{2}\,\text{erfc}\left(\sqrt{\frac{(\mu_{\rho|H_1}-T_\rho)^2}{2\sigma^2_{\rho|H_1}}}\right)$$

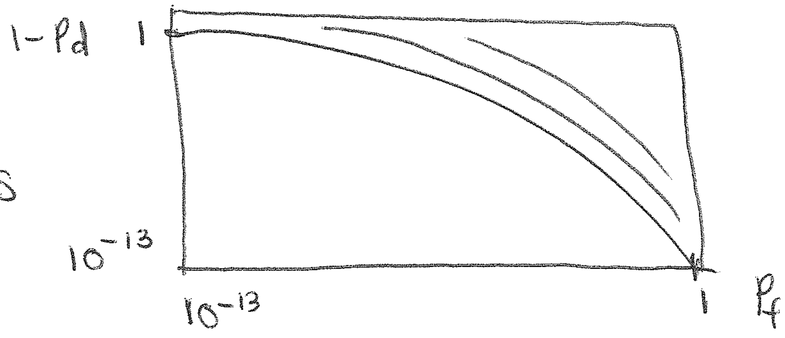
che posso esprimere in funzione di $\bar{P}_f$ → ROC

$$\Longrightarrow \quad 1-P_d = \frac{1}{2}\,\text{erfc}\left(\sqrt{\frac{n\,SNR}{2} - \text{erfc}^{-1}(2\bar{P}_f)}\right)$$

$$SNR = \frac{\gamma^2 \sigma_w^2}{\sigma_x^2}$$

per vari n e SNR

$\Longrightarrow$ ROC curves



We must average the probabilities derived so far over all possible watermarks:

≈ nothing changes in this case

NB: Se avessimo usato il criterio di minimizzazione della prob d'errore totale → T_ρ sarebbe stata fissata a metà fra $\mu_{\rho|H_0}$ e $\mu_{\rho|H_1}$

NB: The most important consequence of adopting a detector which does not depend on y is that the embedder can adjust the watermark strength to the asset at hand (trade-off between imperceptibility and robustness) without informing the detector

NB: Practical applicability: many assumptions are not matched in practical scenarios

- attacks not normally distributed
- host feature and noise not uncorrelated
- noise may depend on the host signal

→ correlation-based detection is far from being the optimum choice