Error curve & Neyman-Pearson rule

Recap: Binary hypothesis testing

Bayesian

$X=0$ & $X=1$ with priors $\underbrace{P_r(X=0)}_{=p_0}$ & $\underbrace{P_r(X=1)}_{=p_1}$

Observe X through $P_{Y|X}$ (or $f_{Y|X}$).

MAP rule: $\hat{x}_{MAP}(y) = \underset{x\in\{0,1\}}{\arg\max} \underbrace{P_{X|Y}(x|y)}_{\text{posterior prob.}}$

$\Downarrow$

min the prob. of error

Binary case: $$\Lambda(y) = \frac{P_{Y|X}(y|1)}{P_{Y|X}(y|0)} \underset{\hat{x}(y)=0}{\overset{\hat{x}(y)=1}{\gtrless}} \frac{p_0}{p_1} \quad \text{(LRT)}.$$

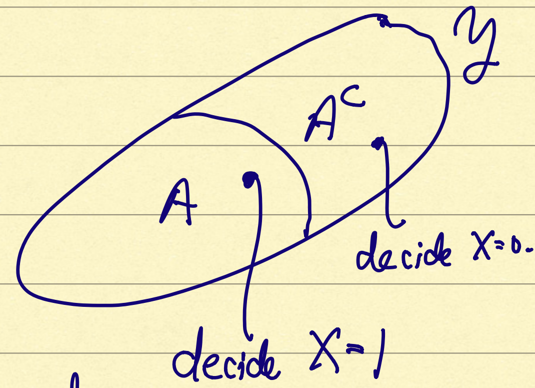
Sufficient statistic: y may be very high-dim. $y=(y_1,\dots,y_n)$

defined as any f^n $v(y)$ for which $\Lambda(y)$ can be computed.

Neyman-Pearson (non-Bayesian) HT. $(X=0,1)$

A test $A\subset\mathcal{Y}$ and $X=0$, an error is made whenever $y\in A$ (A: subset of $\mathcal{Y}$ for which we make a decision in favor of $X=1$)

$A \sqcup A^c = \mathcal{Y}$



If $X=1$, an error is made if $y \in A^c$.

$q_0(A) = \Pr(Y \in A \mid X=0)$

↑ prob. of false alarm.

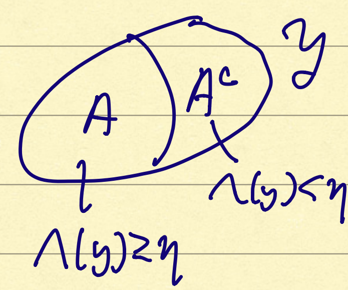
$q_1(A) = \Pr(Y \in A^c \mid X=1)$

↑ prob. of mis-detection

Special class of tests include threshold tests

$$A = \left\{ y : \frac{f_{Y|X}(y|1)}{f_{Y|X}(y|0)} \geq \eta \right\}.$$

$\frac{f_{Y|X}(y|1)}{f_{Y|X}(y|0)} = \Lambda(y) = LR.$ ↑ η: threshold

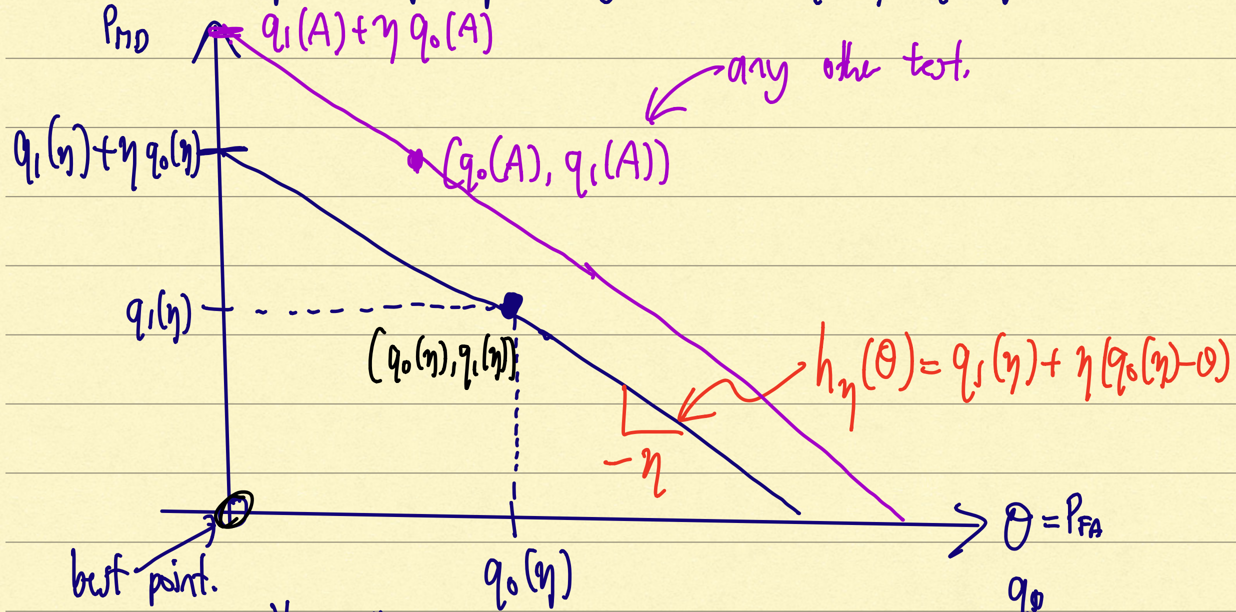


$$q_0(\eta) = \Pr(e_\eta \mid X=0) = \Pr\left(\frac{f_{Y|X}(Y|1)}{f_{Y|X}(Y|0)} \geq \eta \,\middle|\, X=0 \right)$$

$$q_1(\eta) = \Pr(e_\eta \mid X=1) = \Pr\left(\frac{f_{Y|X}(Y|1)}{f_{Y|X}(Y|0)} < \eta \,\middle|\, X=1 \right)$$

<u>Thm:</u> Consider a 2-dim plot in which $(q_0(A), q_1(A))$ is plotted for each test $A \subset \mathcal{Y}$.
Then for every threshold test given by $\eta \in [0,\infty)$ and any arb. test A, the point $(q_0(A), q_1(A))$ lies in

the closed halfspace to the top-right of a straight line of slope $-\eta$ passing thru. $(q_0(\eta), q_1(\eta))$.



We want "both" $(q_0(A), q_1(A))$ to be simultaneously small

Rmk: This theorem says that threshold tests achieve best tradeoff in terms of q_0, q_1.

$\eta = \frac{p_0}{1-p_0}$ $\qquad\qquad\qquad\qquad$ $p_0 + p_1 = 1$

Pf: For any $\eta \in [0, \infty)$ consider priors p_0 & p_1 s.t. $\eta = p_0/p_1$. For any test $A \subset \mathcal{Y}$, Bayesian err. prob.

$$\Pr(e(A)) = \underbrace{p_0}_{\overset{\Pr(X=0)}{\shortparallel}} \underset{\underset{\Pr(e(A)\mid X=0)}{\shortparallel}}{q_0(A)} + \underbrace{p_1}_{\overset{\Pr(X=1)}{\shortparallel}} \underset{\underset{\Pr(e(A)\mid X=1)}{\shortparallel}}{q_1(A)} = p_1\left[q_1(A) + \eta q_0(A)\right].$$

For the threshold test η, Bayesian error prob.

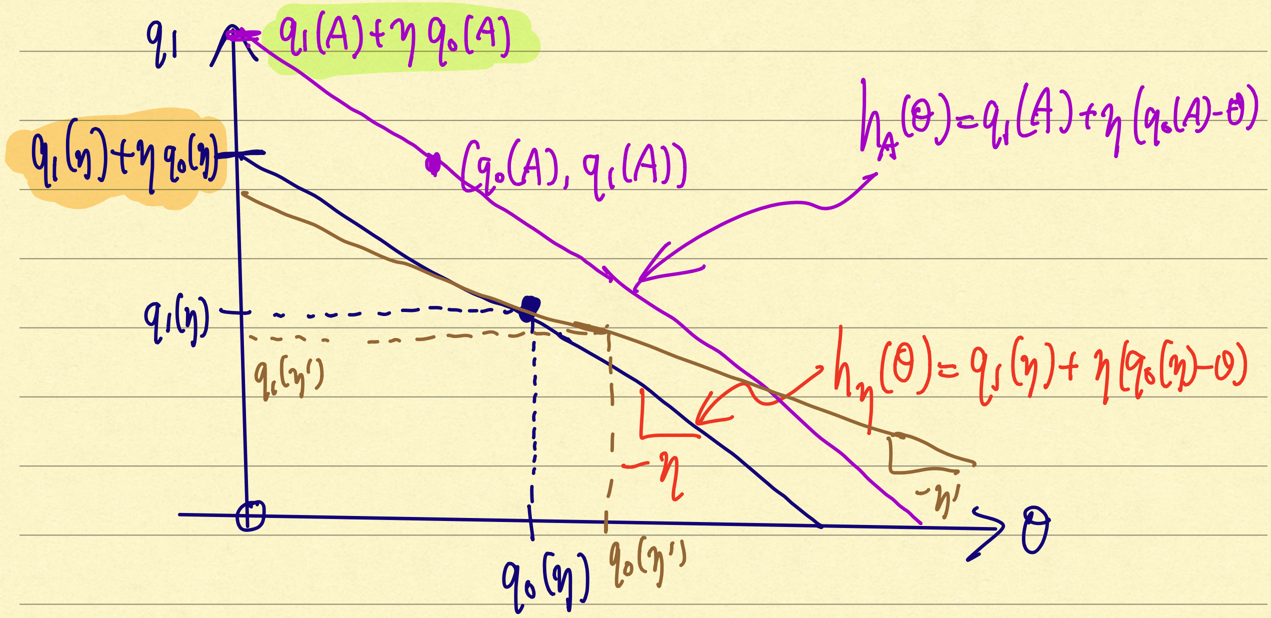
$$\Pr(e_\eta) = \underbrace{p_0\, q_0(\eta)}_{} + p_1\, q_1(\eta) = p_1[q_1(\eta) + \eta\, q_0(\eta)].$$

$$q_0(\eta) = \Pr(e_\eta \mid X=0) \qquad q_1(\eta) = \Pr(e_\eta \mid X=1)$$

By the optimality of threshold tests for Bayesian BHT,

$$\Pr(e_\eta) \le \Pr(e(A)) \qquad \text{(MAP rule)}$$

$$\Rightarrow \quad q_1(\eta) + \eta\, q_0(\eta) \le q_1(A) + \eta\, q_0(A).$$



The lines $h_\eta(\theta)$ & $h_A(\theta)$ have slope $-\eta$ and pass through $(q_0(\eta), q_1(\eta))$ & $(q_0(A), q_1(A))$ resp.

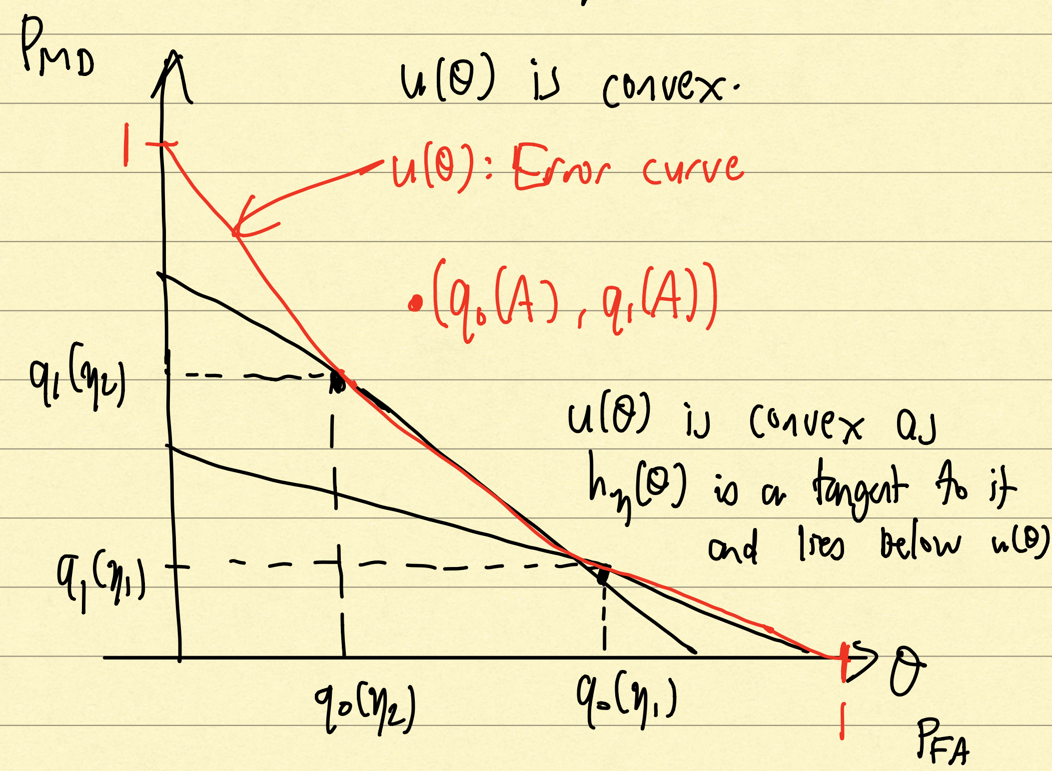
But <u>y-intercept of A</u> $\ge$ <u>y-intercept of η</u>.

All points on $h_A(\theta)$ lie in the closed halfspace to the top-right of all points on the threshold line $h_\eta(\theta)$. ∎

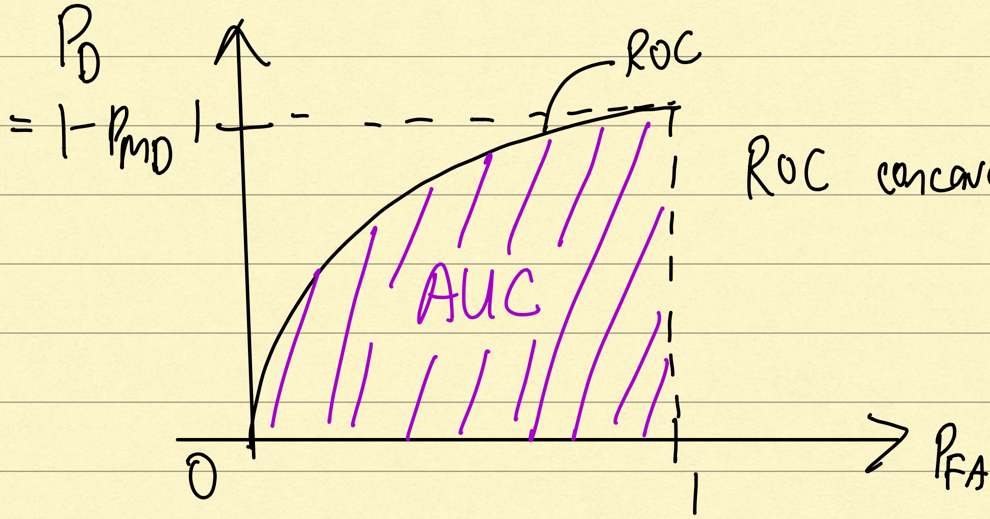
$(q_0(A), q_1(A))$ for an arb. test A. lies to the top right of straight line $h_\eta(\theta)$ for all $0 \le \eta < \infty$.

Define upper envelope of this family of straight lines

$$u(\theta) = \sup_{0 \le \eta < \infty} h_\eta(\theta) = \sup_{0 \le \eta < \infty} \{ q_1(\eta) + \eta (q_0(\eta) - \theta) \}$$



<u>Receiver operating characteristic.</u>

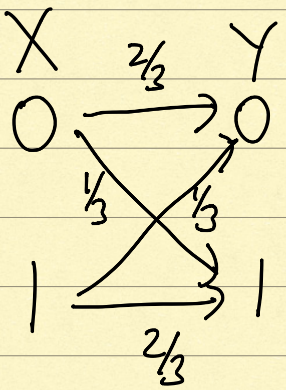


ROC concave.

Conclusion: Threshold tests are optimal points for the tradeoff. They lie on the error curve.

Are all optimal tests (deterministic) threshold tests?
<u>No!</u> : We need <u>mixtures</u> of <u>threhold</u> tests.

Eg:
BSC



$$P_{Y|X}(y|x) = \begin{cases} 2/3 & x = y \\ 1/3 & x \neq y \end{cases}$$

Threhold test: $$\Lambda(y) = \frac{P_{Y|X}(y|1)}{P_{Y|X}(y|0)} \underset{\text{'0'}}{\overset{\text{'1'}}{\gtrless}} \eta.$$

$y \in \{0,1\}$: When $y=1$, $\Lambda(1) = \frac{2/3}{1/3} = 2$
When $y=0$, $\Lambda(0) = \frac{1/3}{2/3} = 1/2$.

'x=1'

Under $\eta = 1$ (ML), $\Lambda(y) \underset{`X=0'}{\gtrless} 1$.

If $y=0$, $\hat{x}=0$
$\underbrace{\text{If } y=1,\ \hat{x}=1}$
$\hat{x}(y) = y$.

If $\eta \le \frac{1}{2}$, $\Lambda(y) \overset{`X=1'}{\underset{`X=0'}{\gtrless}} \eta \left(\le \frac{1}{2}\right) \Rightarrow \hat{x}(y) = 1 \ \forall y$.

If $\eta > 2$, $\Lambda(y) \gtrless \eta \ (>2)$ $\hat{x}(y) = 0 \ \forall y$

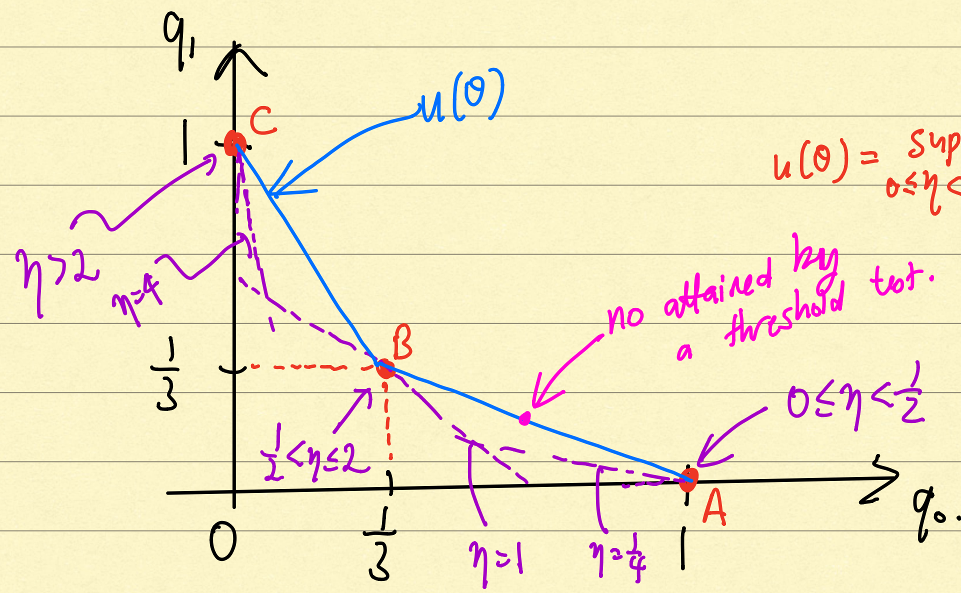
$$\hat{x}(y) = \begin{cases} 1 & 0 \le \eta \le \frac{1}{2} \\ y & \frac{1}{2} < \eta \le 2 \\ 0 & \eta > 2 \end{cases}$$

For $\eta \le \frac{1}{2}$, $\hat{x}(y) = 1$ so for both $y=0$ & $y=1$, error is made for $X=0$, but not $X=1$.
$q_0(\eta) = 1$, $q_1(\eta) = 0$ for $\eta \le \frac{1}{2}$.
$\quad\ \ \|\qquad\qquad\quad \|$
Pr(false alarm) $\quad$ Pr(mis-detection) = Pr(err | X=1)

For $\frac{1}{2} < \eta \le 2$, $\hat{x}(y) = y$, $q_0(\eta) = q_1(\eta) = \frac{1}{3}$.

For $\eta > 2$, $\hat{x}(y) = 0 \ \forall y = 0, 1$.
$q_1(\eta) = 1$, $q_0(\eta) = 0$.
$\quad \|$
$\Pr(`X=0' \mid X=1) = 1$ $\because$ we always declare $`X=0'$ when $\eta > 2$.

$$q_0(\eta) = \begin{cases} 1 & , \eta \le \frac{1}{2} \\ 1/3 & , \frac{1}{2} < \eta \le 2 \\ 0 & , \eta > 2, \end{cases} \qquad q_1(\eta) = \begin{cases} 0 & , \eta \le \frac{1}{2} \\ 1/3 & , \frac{1}{2} < \eta \le 2 \\ 1 & , \eta > 2, \end{cases}$$



Threshold tests attain the points A, B, C.

The error curve contains "many" points that are not achieved by threshold tests e.g. an "interior point bet. $(1/3, 1/3)$, $(1, 0)$

Requirement: $P_{FA} = q_0 \le 1/3$. Goal: min $P_{MD} = q_1$

Ans. $1/3$.

Requirement on $\overset{\text{"}q_0}{P_{FA}} \notin \{0, \frac{1}{3}, 1\}$.

Achieve the best (smallest) $q_1(A)$ for a given $q_0(A) \le 1/2$.

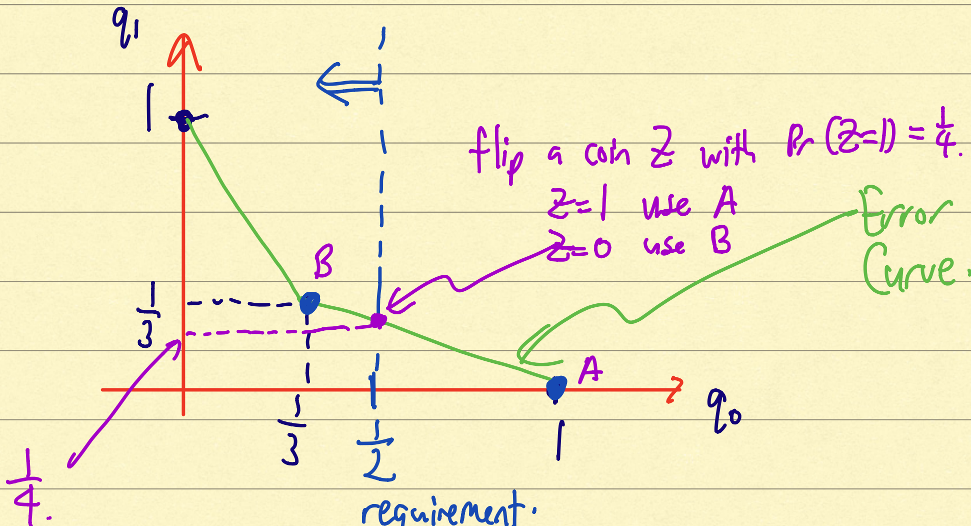
$\boxed{\text{Randomize}}$ between two tests $\hat{x}(y)=y$ (ML) &

$\hat{x}(y)=1.\ \forall y$

$(q_0, q_1) = (1, 0)$

$\Downarrow$

$(q_0, q_1) = (\frac{1}{3}, \frac{1}{3})$



$$q_0(A) = \frac{1}{2} = \alpha \cdot 1 + (1-\alpha)\frac{1}{3} \Rightarrow \alpha = 1/4.$$

α represents fraction of time we always declare 1.

Our final decision

$$\hat{x}(y) = \begin{cases} 1 & y=1 \\ 1 & y=0,\ Z=1 \\ 0 & y=0,\ Z=0 \end{cases} \qquad \begin{array}{l} Z \sim \text{Bern}(\frac{1}{4}=\alpha) \\ Pr(Z=1) = 1/4 \\ Pr(Z=0) = \frac{3}{4}. \end{array}$$

$1-\alpha$ fraction of time we use the ML rule.

Best MD q_1 prob $= \frac{1}{4}\cdot 0 + \frac{3}{4}\cdot\frac{1}{3} = \frac{1}{4} < \frac{1}{3}.$

Rmk: i) We need to use randomization to attain all points on error curve (ROC) if likelihood model is a pmf (i.e., $P_{Y|X}(\cdot|x)$ is a pmf $\forall x$).

ii) If $f_{Y|X}(\cdot|x)$ is a prob. density function, no randomization is needed & every point on error curve is attained by a threshold test.

Eg: (lec10.pdf)

These q_i's are not the same as the above prob.

$X=0 \quad q=q_0=1/2$

$X=1 \quad q=q_1=1/4.$

Likelihoods: $\begin{cases} P_{Y|X}(y|0) = q_0(1-q_0)^y, & y \geq 0 \\ P_{Y|X}(y|1) = q_1(1-q_1)^y, & y \geq 0. \end{cases}$

1. If $p_0 = p_1$ (ML) find the min prob. of error rule.

$$\underline{\Lambda(y)} = \frac{P_{Y|X}(y|1)}{P_{Y|X}(y|0)} = \frac{q_1(1-q_1)^y}{q_0(1-q_0)^y} \underset{`X=0'}{\overset{`X=1'}{\gtrless}} \eta = 1$$

$$\frac{\left(\frac{1}{4}\right)\left(\frac{3}{4}\right)^y}{\left(\frac{1}{2}\right)^{y+1}} \underset{}{\overset{`X=1'}{\gtrless}} 1 \iff y \overset{`X=1'}{\gtrless} \frac{\log\left(\frac{q_0}{q_1}\right)}{\log\left(\frac{1-q_1}{1-q_0}\right)}$$

$$y \underset{<}{\overset{\text{`}X=1\text{'}}{\geq}} 1.71$$

Decide in favor of $X=1$ if $y \geq 2$

Decide in favor of $X=0$ if $y < 2$.

2. Plot the $(P_{FA}, \underset{\underset{1-P_{MD}}{\parallel}}{P_D})$ curve (ROC).

$$\boxed{P_D = 1 - \underbrace{q_1(A)}_{\text{notation in book.}}}$$

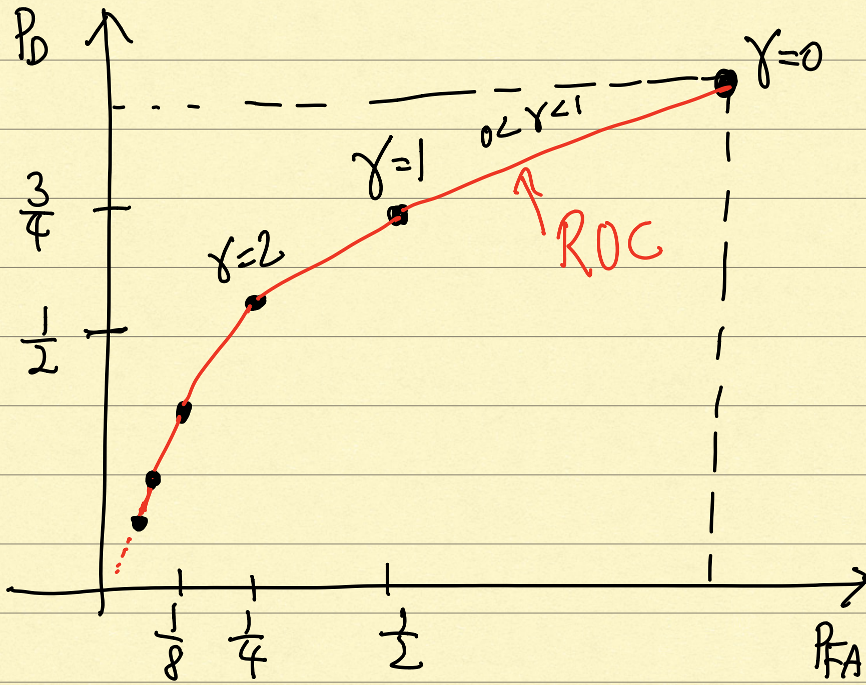
$$P_{FA} = P_r(Y \geq \gamma \mid X=0)$$

$$= \sum_{y \geq \gamma} q_0(1-q_0)^y = \sum_{y \geq \gamma} \left(\frac{1}{2}\right)\left(\frac{1}{2}\right)^y = \frac{1}{2}\,\frac{\left(\frac{1}{2}\right)^\gamma}{1-\frac{1}{2}} = \left(\frac{1}{2}\right)^\gamma.$$

$$P_D = P_r(Y \geq \gamma \mid X=1) = \sum_{y \geq \gamma} q_1(1-q_1)^y = \left(\frac{3}{4}\right)^\gamma.$$

The ROC (set of all (P_{FA}, P_D) points) is the countably infinite set of pairs of points

$$\begin{array}{llllllll} (P_{FA}, P_D): & (1,1), & \left(\frac{1}{2}, \frac{3}{4}\right), & \left(\frac{1}{4}, \frac{9}{16}\right), & \left(\frac{1}{8}, \frac{27}{64}\right), & \left(\frac{1}{16}, \frac{81}{256}\right), & \ldots \\ \gamma: & 0, & 1, & 2, & 3, & 4, & \ldots \end{array}$$



3. max P_D s.t. $P_F \le \frac{1}{128} = \frac{1}{2^7}$.

If $P_F \le \frac{1}{128}$, we choose $\gamma = 7$; corresponding $P_D = (3/4)^7$.

Rule: If $Y \ge 7$ declare $X=1$
$Y < 7$ declare $X = 0$

You are allowed to use a coin of any bias.

4. max P_D s.t. $P_F \le \frac{1}{100}$; Find the rule.

Ans: $0.01 \in \left(\frac{1}{128}, \frac{1}{64}\right) = \left(\frac{1}{2^7}, \frac{1}{2^6}\right)$.
Randomize bet. operating pts $\left(\frac{1}{2^7}, \left(\frac{3}{4}\right)^7\right)$ & $\left(\frac{1}{2^6}, \left(\frac{3}{4}\right)^6\right)$.

$P_F = 0.01 = \alpha \frac{1}{128} + (1-\alpha)\frac{1}{64} \Rightarrow \alpha = 0.72$.

$$\max P_D = \alpha\left(\tfrac{3}{4}\right)^7 + (1-\alpha)\left(\tfrac{3}{4}\right)^6 = \underline{0.1459}.$$

$$\Pr(Z=1) = \alpha = 0.72.$$

$$\hat{X} = \begin{cases} 1 & Y \geq 7 \quad Z=1 \\ 0 & Y < 7 \quad Z=1 \\ 1 & Y \geq 6 \quad Z=0 \\ 0 & Y < 6 \quad Z=0 \end{cases}$$

$$\Pr(Z=0) = 0.28.$$

$$\hat{X} = \begin{cases} 1 & Y \geq 7 \\ 0 & Y \leq 5 \\ 1 & Y = 6 \quad Z=0 \quad \text{w.p } 1-\alpha \\ 0 & Y = 6 \quad Z=1 \quad \text{w.p } \alpha \end{cases}$$

1) Optimality of threshold test (NP theorem)
2) Error curve
3) All threshold test $\in$ Error Curve
4) Error curve contain points that don't corr. threshold test.
5) Randomization.

<u>lec12.pdf.</u> <u>Cramér - Rao bound.</u>

fluctuation $\equiv$ Variance.