Subject: Lecture 10: Introduction to detection & hypothesis testing

No:

Date:

Each model (state-of-the-world) is characterized by a discrete rv X (called a hypothesis).

Observations Y can be discrete or continuous.

$$X\in\{0,1,\dots,M-1\} \longrightarrow \boxed{P_{Y|X}} \xrightarrow{\;Y\;} \boxed{\text{Detector}} \longrightarrow \hat{x}(Y)\in\{0,1,\dots,M-1\}$$

Decision is correct if $\hat{x}(Y)=X$.

X has a prior probability $p_X(x)=\overset{=P_x}{P(X=x)}$, $x\in\{0,1,\dots,M-1\}$

Likelihood (Observation model) $p_{Y|X}(y|x)=P(Y=y|X=x)$ for discrete obs & $f_{Y|X}(y|x)$ for continuous obs.

<u>MAP Criterion (Maximum A-Posteriori)</u>

If $f_Y(y)>0$, $\quad p_{X|Y}(x|y)=\dfrac{p_X(x)\,f_{Y|X}(y|x)}{f_Y(y)}$, $f_Y(y)=\sum_{x=0}^{M-1} p_x f_{Y|X}(y|x)$

Fact: To maximize the prob. of choosing the true hypothesis, choose $\hat{x}(\cdot)$ to be

$$\hat{x}(y)=\underset{x\in\{0,1,\dots,M-1\}}{\operatorname{argmax}}\; p_{X|Y}(x|y) \quad \text{(MAP)}.$$

$$\hat{x}_{MAP}(y)=\underset{x}{\operatorname{argmax}}\; p_x f_{Y|X}(y|x)$$

Pf: Let $\hat{x}_A(y)$ be any other decision rule. Then

$$p_{X|Y}(\hat{x}_{MAP}(y)|y)\geq p_{X|Y}(\hat{x}_A(y)|y) \quad \forall y.$$

$$\Rightarrow \int f_Y(y)\,p_{X|Y}(\hat{x}_{MAP}(y)|y)\,dy \geq \int f_Y(y)\,p_{X|Y}(\hat{x}_A(y)|y)\,dy$$

Note that LHS $=\int \sum_x f_Y(y)\,p_{X|Y}(x|y)\,\mathbb{1}\{\hat{x}_{MAP}(y)=x\}\,dy$

$$=\sum_x p_X(x)\int f_{Y|X}(y|x)\,\mathbb{1}\{\hat{x}_{MAP}(y)=x\}\,dy$$

$$=\sum_x p_X(x)\,P(\hat{x}_{MAP}(Y)=x\,|\,X=x) = P(\hat{x}_{MAP}(Y)=X).$$

Subject: No:
                        Date:

Similarly, RHS $= P_r(\hat{x}_4(Y) = X)$

$\Rightarrow$ MAP rule maximizes prob. of detection.

<u>Binary MAP Detection.</u> $|\mathcal{X}| = 2$, $M = 2$, $\mathcal{X} = \{0, 1\}$

Say $p_X(0) > 0$ $p_X(1) > 0$ & $p_X(0) = p_0$, $p_X(1) = p_1$

$f_Y(y) = p_0 f_{Y|X}(y|0) + p_1 f_{Y|X}(y|1)$

Posterior prob $p_{X|Y}(\ell|y) = \dfrac{p_\ell f_{Y|X}(y|\ell)}{f_Y(y)}$

Writing out the MAP rule,

$$\frac{p_1 f_{Y|X}(y|1)}{f_Y(y)} \underset{\hat{x}(y)=0}{\overset{\hat{x}(y)=1}{\gtrless}} \frac{p_0 f_{Y|X}(y|0)}{f_Y(y)}.$$

$$\Leftrightarrow \quad \Lambda(y) = \frac{f_{Y|X}(y|1)}{f_{Y|X}(y|0)} \underset{\hat{x}(y)=0}{\overset{\hat{x}(y)=1}{\gtrless}} \eta = \frac{p_0}{p_1} \quad \text{(threshold)}$$

likelihood ratio for a binary decision/detection problem.

Optimal detection rule is a threshold rule (of the likelihood ratio) & the threshold is $\eta = p_0/p_1$.

<u>Maximum Likelihood (ML) Decision rule.</u>

$p_0 = p_1 = 1/2 \Rightarrow \eta = 1$ $\qquad f_{Y|X}(y|1) \underset{\hat{x}(y)=0}{\overset{\hat{x}(y)=1}{\gtrless}} f_{Y|X}(y|0)$

<u>Overall Prob. of Error</u>

$$P_r(e_\eta) = p_0 P_r(e_\eta | X=0) + p_1 P_r(e_\eta | X=1)$$

Partition of sample space given a rule such as the MAP rule.

$A_0 = \{y : \Lambda(y) < \eta\}$ corresponds to $\hat{x}(y) = 0$

$A_1 = \{y : \Lambda(y) > \eta\}$ —"— $\hat{x}(y) = 1$.

No:

Subject:

Date:

If $X=0$, error occurs if $y \in A_1 \Rightarrow P_r(e_\eta | X=0) = \int_{A_1} f_{Y|X}(y|0)\,dy$

If $X=1$ —— " —— $y \in A_0 \Rightarrow P_r(e_\eta | X=1) = \int_{A_0} f_{Y|X}(y|1)\,dy$

$$\Rightarrow P_r(e_\eta | X=0) = P_r(\Lambda(Y) \geq \eta \,|\, X=0)$$
$$P_r(e_\eta | X=1) = P_r(\Lambda(Y) < \eta \,|\, X=1)$$

Rmk: If the one-dim quantity $\Lambda(y)$ can be found, we can perform a threshold test on $\Lambda(y)$ without further reference to y (which may be a high-dim vector).

<u>Sufficient Statistics</u>

Def: For binary hypothesis testing, a sufficient statistic is any function $v(y)$ of the observation y for which the likelihood ratio ~~$\Lambda(y)$~~ $\Lambda(y)$ can be computed, i.e., $v(y)$ is a SS if $\exists$ function $u(v)$ s.t. $\Lambda(y) = u(v(y))$ for all y.

Eg: $v(y) = \ln \Lambda(y)$ is a sufficient statistic for BHT.

Eg. Detection of antipodal signals in Gaussian noise $Y = X + Z$
$Z \sim N(0, \sigma^2)$

$X \perp\!\!\!\perp Z$ and $X \in \{-b, b\}$, $b > 0$.

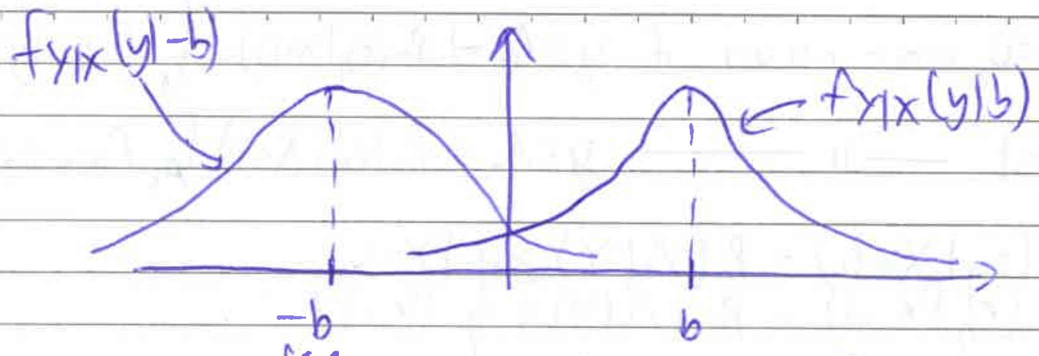
$$f_{Y|X}(y|b) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-b)^2}{2\sigma^2}\right), \quad f_{Y|X}(y|-b) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y+b)^2}{2\sigma^2}\right)$$

$$\Lambda(y) = \frac{f_{Y|X}(y|b)}{f_{Y|X}(y|-b)} = \exp\left(\frac{2yb}{\sigma^2}\right).$$

Optimal test $\quad \Lambda(y) = \exp\left(\frac{2yb}{\sigma^2}\right) \underset{\hat{x}(y)=-b}{\overset{\hat{x}(y)=b}{\gtrless}} \eta = \frac{p_0}{p_1} = \frac{p_X(-b)}{p_X(+b)}$

$$\Rightarrow LLR(y) = \frac{2yb}{\sigma^2} \gtrless \ln \eta$$

$$\Rightarrow y \gtrless \frac{\sigma^2 \ln \eta}{2b}. \text{ If ML, RHS} = 0.$$



$Pr(e_0|X=0) = \int_0^\infty f_{Y|X}(y|-b)\,dy = Q\left(\frac{b}{\sigma}\right)$, $\quad Q(u) = \int_u^\infty \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\,dt$

By symmetry, $Pr(e_0|X=1) = Q\left(\frac{b}{\sigma}\right)$.

b/σ: signal-to-noise ratio.

Eg. Vector Observation. $p_X(0)=p_0$, $p_X(1)=p_1$

Observations are vectors $\underline{Y} = (Y_1, \dots, Y_n) = Y^n$.

LRT: $$\Lambda(\underline{y}) = \frac{f_{\underline{Y}|X}(\underline{y}|1)}{f_{\underline{Y}|X}(\underline{y}|0)} \underset{\hat{x}(\underline{y})=0}{\overset{\hat{x}(\underline{y})=1}{\gtrless}} \eta = \frac{p_0}{p_1}$$

Suppose observation are conditionally independent of one another given the hypothesis.

$\forall x \in \{0,1\}$, $f_{\underline{Y}|X}(\underline{y}|x) = \prod_{j=1}^n f_{Y_j|X}(y_j|x) \quad \forall\, \underline{y} = (y_1, \dots, y_n) \in \mathbb{R}^n$

LRT becomes

$$\Lambda(\underline{y}) = \prod_{j=1}^n \frac{f_{Y_j|X}(y_j|1)}{f_{Y_j|X}(y_j|0)} \underset{\hat{x}(\underline{y})=0}{\overset{\hat{x}(\underline{y})=1}{\gtrless}} \eta$$

LLRT becomes

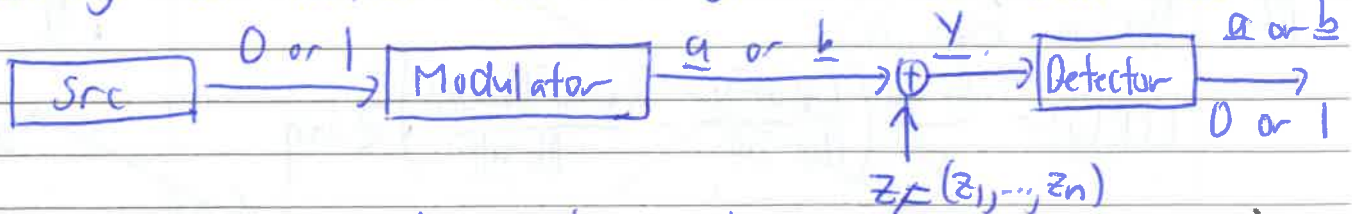
$$\sum_{j=1}^n \ln \frac{f_{Y_j|X}(y_j|1)}{f_{Y_j|X}(y_j|0)} = \sum_{j=1}^n LLR_j(y_j) \gtrless \ln \eta.$$

If $LLR_j(Y_j)$ are conditionally identically distributed given X,

Subject: No: Date:

$$\Rightarrow \frac{1}{n}\sum_{j=1}^{n} LLR_j(Y_j) \gtrless \frac{1}{n}\ln\eta$$

Binary detection of vectors signals in Gaussian noise



Z^n is an i.i.d random vector with components

$$f_{Z_i}(z) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{z^2}{2\sigma^2}\right), \quad z \in \mathbb{R}.$$

$$LLR(y_i) = \log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(y_j-b_j)^2}{2\sigma^2}\right) \Big/ \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(y_j-a_j)^2}{2\sigma^2}\right)\right)$$

$$= \left(\frac{b_j-a_j}{\sigma^2}\right)\left[y_j - \left(\frac{b_j+a_j}{2}\right)\right].$$

Putting all the observation together,

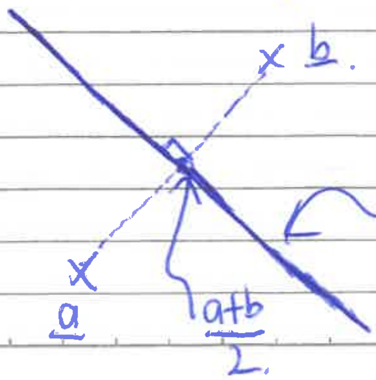
$$LLR(\underline{y}) = \sum_{j=1}^{n} LLR(y_j) = \frac{(\underline{b}-\underline{a})^T}{\sigma^2}\left[\underline{y} - \left(\frac{\underline{b}+\underline{a}}{2}\right)\right] \underset{\hat{x}(\underline{y})=\underline{a}}{\overset{\hat{x}(\underline{y})=\underline{b}}{\gtrless}} \ln\eta$$

Sufficient statistic: $\quad \underline{y}^T(b-a) \gtrless \eta' = \sigma^2\ln\eta + \frac{1}{2}(b^Tb - a^Ta).$

Interpretation for $\eta=1$, i.e, ML test becomes

$$y^T(b-a) \gtrless \frac{1}{2}(b^Tb - a^Ta)$$

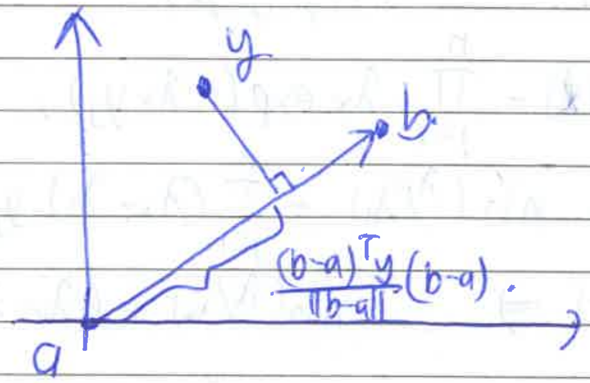
Rmk: Test involves observation only through its inner product with $b-a$. $\Rightarrow y^T(b-a)$ is a sufficient statistic for this BHT.



$$\left\{y : \frac{(b-a)^T}{2}\left(y - \frac{b+a}{2}\right) = 0\right\}.$$

Fact. The two hypotheses can be distinguished by the component of the obs vector y in the direction of the signal vector $b-a$, i.e., the projection length $y^T(b-a)/\|b-a\|$ is a sufficient statistic.

e.g. $a=0$



Decision Eqn is

$$LLR(y) = \frac{\|b-a\|}{\sigma^2}\Big(\underbrace{\frac{(b-a)^T y}{\|b-a\|}}_{\text{Comp. of } y \text{ in signal direction } b-a.} - \frac{(b-a)^T(b+a)/2}{\|b-a\|}\Big) \gtrless \ln \eta.$$

If $a=-b$, $LLR(y) = \frac{\|b\|}{\sigma^2}\left(\frac{2b^T y}{\|b\|} - 0\right) \gtrless \ln \eta$

$$\Rightarrow \frac{2b^T y}{\sigma^2} \gtrless \ln \eta.$$

Problem reduces to the 1-dim case:

$P_r(e_\eta \mid X=-b) = Q\left(\frac{\ln \eta}{2\gamma}\right)$, $P_r(e_\eta \mid X=b) = Q\left(-\frac{\ln \eta}{2\gamma}\right)$.

Why? $P_r(e_\eta \mid X=-b) = \int_{\frac{2b^T y}{\sigma^2} > \ln \eta} f_{Y|X}(y \mid -b)\, dy.$

Under $X=-b$, $Y \sim N(-\underline{b}, \sigma^2 I_n)$, $b^T Y \sim N(-\|b\|^2, \|b\|^2\sigma^2)$.

$$P_r(e_\eta \mid X=-b) = \int_{\frac{2z}{\sigma^2} > \ln \eta} \frac{1}{\sqrt{2\pi \|b\|^2 \sigma^2}} \exp\left(-\frac{(z+\|b\|^2)^2}{2\|b\|^2\sigma^2}\right) dz.$$

$$= \int_{w > \left(\frac{\sigma^2 \ln \eta}{2} + \|b\|^2\right)\frac{1}{\|b\|\sigma}} N(0,1)\, dw = Q\left(\frac{\ln \eta}{2\gamma} + \gamma\right), \quad \gamma = \frac{\|b\|}{\sigma}.$$

Subject: No: Date:

Eg. Poisson processes: Two diff. possibilities for rate λ_0, λ_1.
p_ℓ: $\ell \in \{0,1\}$ is the prior prob. for rate being λ_ℓ.

Observations $\underline{Y} = (Y_1, \dots, Y_n)$.

$$f_{\underline{Y}|X}(\underline{y}|x) = \prod_{j=1}^{n} \lambda_x \exp(-\lambda_x y_j), \quad \underline{y} \geq 0.$$

$$LLR(\underline{y}) = n \ln(\lambda_1/\lambda_0) + \sum_{j=1}^{n} (\lambda_0 - \lambda_1) y_j$$

MAP test $\Rightarrow$ $$n \ln(\lambda_1/\lambda_0) + (\lambda_0 - \lambda_1) \overbrace{\sum_{j=1}^{n} y_j}^{S_n} \gtrless \ln \eta.$$

Test depends only on $s_n = \sum_{j=1}^{n} y_j$, n^{th} arrival epoch.

S_n is a sufficient statistic; not surprising since under each hypothesis, first $n-1$ arrivals are unif. conditioned on n^{th} arrival time.

Sufficient Statistics: Seek other equivalent characterizations of SS.

Thm: Let $V = v(Y)$ be a f^n of Y for a BHT problem. The following are equiv. conditions for $v(Y)$ to be a SS.

1. A function $u(\cdot)$ exists s.t. $\Lambda(y) = u(v(y))$.

2. For any priors $p_0, p_1 > 0$, the a posteriori prob. satisfy

$$p_{X|Y}(x|y) = p_{X|V}(x|v(y))$$

$\boxed{X - V - Y \text{ forms a Markov chain}}$

3. Likelihood ratio of y is the same as that of $v(y)$

$$\Lambda(y) = \frac{p_{V|X}(v(y)|1)}{p_{V|X}(v(y)|0)}.$$

Pf $1 \Rightarrow 2 \Rightarrow 3 \Rightarrow 1$.

$1 \Rightarrow 2$. $$p_{X|Y}(0|y) = \frac{p_0 \, p_{Y|X}(y|0)}{p_0 \, p_{Y|X}(y|0) + p_1 \, p_{Y|X}(y|1)} = \frac{p_0}{p_0 + p_1 \Lambda(y)} = \frac{p_0}{p_0 + p_1 u(v(y))}.$$

$p_{X|Y}(0|y)$, the cond. prob. of 0 given y, is a f^n of y only through $v(y)$.

Subject: No: Date:

$2 \Rightarrow 3$ $$\frac{p_{X|Y}(1|y)}{p_{X|Y}(0|y)} = \frac{p_{X|V}(1|v(y))}{p_{X|V}(0|v(y))} = \frac{p_{X|V}(1|v(y))}{p_{X|V}(0|v(y))}$$

Applying Bayes rule 4 times,

$$\frac{p_{Y|X}(y|1)\, p_X(1)/p_Y(y)}{p_{Y|X}(y|0)\, p_X(0)/p_Y(y)} = \frac{p_{V|X}(v(y)|1)\, p_X(1)/p_V(v(y))}{p_{V|X}(v(y)|0)\, p_X(0)/p_V(v(y))}$$

$$\Rightarrow \Lambda(y) = \frac{p_{V|X}(v(y)|1)}{p_{V|X}(v(y)|0)}.$$

$3 \Rightarrow 1$ The RHS of $(*)$ is a function of y only through $v(y)$.

Choose $$u(\cdot) = \frac{p_{V|X}(\cdot|1)}{p_{V|X}(\cdot|0)}.$$ $\blacksquare$

Neyman-Pearson Rule.

No need to assign priors to $X=0$ & $X=1$.

Two error probabilities, $Y \in D$: decision made in favor of $X=1$,

$P_{FA} = \Pr(Y \in D \mid X=0)$ & $P_{MD} = \Pr(Y \in D^c \mid X=1)$.

Always possible to make $P_{FA}=0$ (i.e., set $D=\phi$) but then $P_{MD}=1$; more meaningful to understand the tradeoff between the two error probabilities.

Consider minimizing P_{MD} subject to $P_{FA} \le \varepsilon$ for some $\varepsilon > 0$.

Abbreviate the error probabilities as

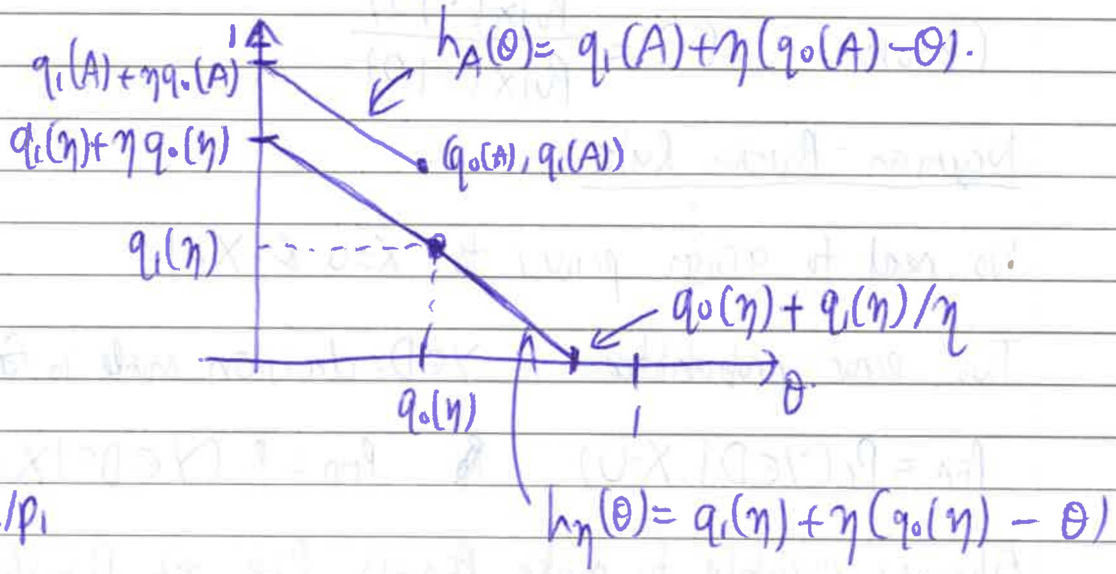
$q_0(A) = \Pr(Y \in A \mid X=0)$ $\quad q_1(A) = \Pr(Y \in A^c \mid X=1)$

Threshold test $$A = \left\{ y : \frac{f_{Y|X}(y|1)}{f_{Y|X}(y|0)} \ge \eta \right\}$$ in favor of $X=1$.

Since threshold tests are important, we write

$$q_0(\eta) = \Pr(e_\eta \mid X=0), \quad q_1(\eta) = \Pr(e_\eta \mid X=1) = \Pr\left(\frac{f_{Y|X}(Y|1)}{f_{Y|X}(Y|0)} < \eta \mid X=1\right)$$
$$= \Pr\left(\frac{f_{Y|X}(Y|1)}{f_{Y|X}(Y|0)} \geq \eta \mid X=0\right)$$

Lem: Consider a 2-dim plot in which the pair $(q_0(A), q_1(A))$ is plotted for each test A. For each threshold test with threshold $0 \leq \eta < \infty$ and each arbitrary test A, the point $(q_0(A), q_1(A))$ lies in the closed halfspace above and to the right of a straight line of slope $-\eta$ passing through $(q_0(\eta), q_1(\eta))$.



$\eta = p_0/p_1$

Punchline: Threshold tests provide optimal points for the tradeoff between the two types of errors.

This is because all other tests are inferior; they lie to the top right of threshold tests.

Will make this more precise next time.