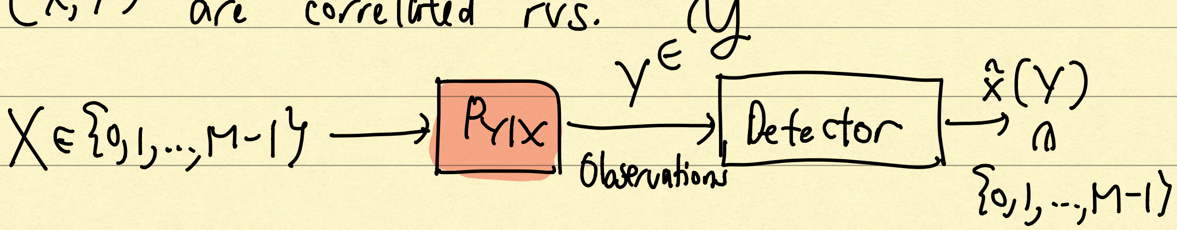
EE5137 Lecture 10 (Reading Secs 8.1 – 8.4).

Introduction to detection & hypothesis testing.

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

Observations Y can be discrete or cts.
(X,Y) are correlated rvs.



Design $\hat{x} : \mathcal{Y} \to \{0,1,\dots,M-1\}$

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

X has a <u>prior</u> distribution $\underbrace{p_X(x)}_{} = \Pr(X=x), x \in \{0,\dots,M-1\}$ $\;= p_x$

Likelihood (Observation) model

$$p_{Y|X}(y|x) = \Pr(Y=y \mid X=x) \text{ for discrete } Y$$
$$f_{Y|X}(y|x) \text{ for cts } Y.$$

<u>MAP criterion (Max a-posteriori)</u> Assume Y cts.

If $f_Y(y) > 0$, $\underbrace{p_{X|Y}(x|y)}_{\text{posterior}} = \dfrac{\overbrace{p_X(x)}^{\text{prior}}\,\overbrace{f_{Y|X}(y|x)}^{\text{likelihood}}}{f_Y(y)}$

$$f_Y(y) = \sum_{x'=0}^{M-1} p_X(x') f_{Y|X}(y|x')$$ model evidence.

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

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

The is the MAP rule!

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

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

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

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

Note that the <u>LHS</u>

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

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

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

$$= \sum_x p_X(x) \Pr(\hat{x}_{MAP}(Y) = X \mid X = x)$$

$$= \Pr(\hat{x}_{MAP}(Y) = X)$$

Similarly for the RHS, $= \Pr(\hat{x}_A(Y) = X)$

In summary, $\Pr(\hat{x}_{MAP}(Y) = X) \geq \Pr(\hat{x}_A(Y) = X)$ for all decoders/detectors A.

Binary MAP Detection $\quad M = 2, \; X \in \{0, 1\}$.

$p_X(0), p_X(1) > 0 \qquad p_X(0) = p_0, \; p_X(1) = p_1$.

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

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

$\ell \in \{0, 1\}$

$\arg\max_{\ell}$

What's the MAP rule?

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

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

$\|$

likelihood ratio for a binary detection / binary hypothesis testing problem.

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

Maximum Likelihood (ML) Decision Rule

$p_0 = p_1 = \frac{1}{2} \Rightarrow \eta = 1$

$$\Rightarrow \quad f_{Y|X}(y|1) \underset{\hat{x}_{MAP}(y)=0}{\overset{\hat{x}_{MAP}(y)=1}{\gtrless}} f_{Y|X}(y|0)$$

$\Rightarrow$ likelihood ratio test

$$\Rightarrow \quad \Lambda(y) \underset{`0'}{\overset{`1'}{\gtrless}} 1 \qquad `1' \equiv \text{Decide } 1.$$

Overall Prob. of Error (Bayesian X has a prior)

$$\underset{\text{thres need not be } 1.}{\Pr(e_\eta)} = \overset{p_0}{\Pr(X=0)}\, \underset{\text{null}}{\Pr(e_\eta | X=0)} + \overset{p_1}{\Pr(X=1)}\, \Pr(e_\eta | \underset{\text{alternative}}{X=1})$$

$\Pr(e_\eta | X=0) = \Pr($'patient has cancer' $|$ she does not have

$\parallel$ cancer)

false alarm. / type-I error

$Pr(e_\eta \mid X=1) = Pr($'patient does not have cancer' | she has cancer)

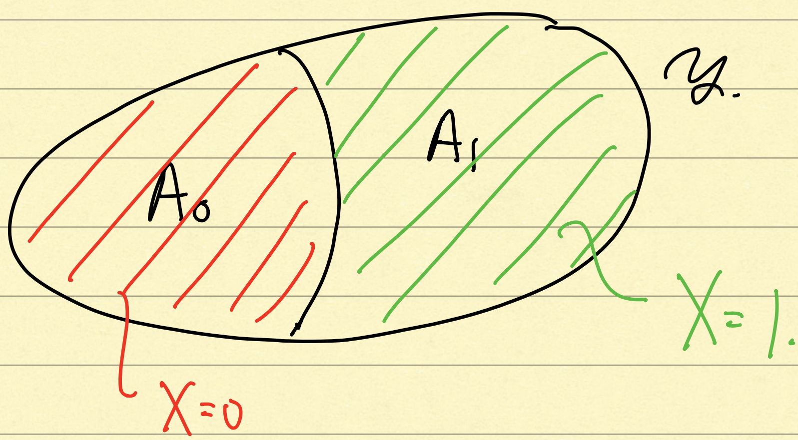
$\parallel$

mis-detection / type-II error.

Partition sample two disjoint parts, e.g. using MAP

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

$A_1 = \{y : \Lambda(y) \geq \eta\}$ $\Rightarrow$ —— " —— $\hat{x}_{MAP}(y) = 1$



$$Pr(e_\eta \mid X=0) = \int_{A_1} f_{Y|X}(y|0)\, dy = Pr(\Lambda(Y) \geq \eta \mid X=0)$$

$$Pr(e_\eta \mid X=1) = \int_{A_0} f_{Y|X}(y|1)\, dy = Pr(\Lambda(Y) < \eta \mid X=1)$$

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

Sufficient Statistics

Def: For BHT, a <u>sufficient statistic</u> is any function $v(y)$ of the obs y for which the likelihood ratio $\Lambda(y)$ can be computed.

In other words, $v(y)$ is a SS of y if $\exists$ function $u(v)$ s.t. $\Lambda(y) = u(v(y))$ $\forall y$.

Eg: $v(y) = \ln \Lambda(y)$ is a sufficient stat. for BHT. $\exists$ $u(v) = \exp(v)$ helps us to get from v to Λ.

$$u(v(y)) = \exp(\ln \Lambda(y)) = \Lambda(y).$$

Eg: Detection of antipodal signals in Gaussian noise.

$$Y = X + Z \qquad X \in \{-b, b\} \qquad \text{(PAM)}$$

$$Z \sim N(0, \sigma^2) \qquad Y \mid \{X = b\} \sim N(b, \sigma^2)$$

$$Y \mid \{X = -b\} \sim N(-b, \sigma^2)$$

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

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

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

Optimal test: $\Lambda(y) \underset{\text{`}\hat{x}(y)=-b\text{'}}{\overset{\text{`}\hat{x}(y)=b\text{'}}{\gtrless}} \eta = \frac{p_0}{p_1} = \frac{Pr(X=-b)}{Pr(X=+b)}$.

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

$$y \underset{}{\overset{\text{`}\hat{x}(y)=b\text{'}}{\gtrless}} \frac{\sigma^2 \ln \eta}{2b}.$$

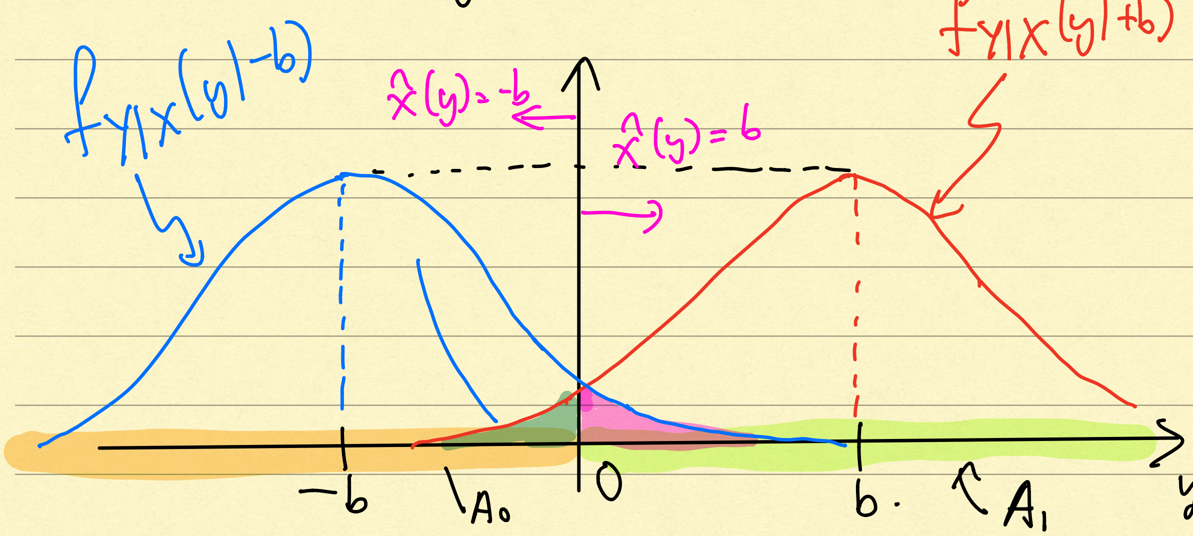
$\shortparallel$

is a sufficient statistic.

$$v(y) = y \overset{u}{\rightsquigarrow} \exp\left(\frac{2yb}{\sigma^2}\right)$$

In the special ML scenario, $p_0 = p_1 \Rightarrow \eta = 1, \ln \eta = 0.$

$$\Rightarrow \quad y \overset{\text{`}\hat{x}(y)=b\text{'}}{\gtrless} 0$$



$$\Pr(e_0 \mid X=-b) = \int_{A_1} f_{Y|X}(y|-b)\, dy$$

false alarm

$$= \int_0^\infty \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(y+b)^2}{2\sigma^2}\right) dy.$$

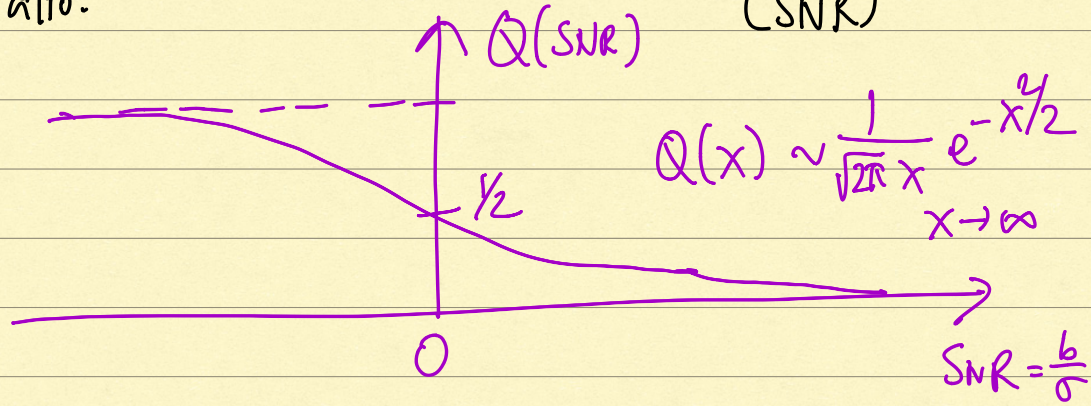
$$= Q\left(\frac{b}{\sigma}\right), \quad Q(u) = \int_u^\infty \frac{1}{\sqrt{2\pi}} e^{-s^2/2}\, ds.$$

Comp. cdf std Gaussian.

misdetection

$$\Pr(e_0 \mid X=b) = \int_{A_0} f_{Y|X}(y|+b)\, dy = Q\left(\frac{b}{\sigma}\right)$$

Rmk: $\frac{b}{\sigma}$ is known as the signal-to-noise ratio. (SNR)



$$Q(x) \sim \frac{1}{\sqrt{2\pi}\, x} e^{-x^2/2} \quad x \to \infty$$

$$\Pr(e_0) = \frac{1}{2}\Pr(e_0 \mid X=0) + \frac{1}{2}\Pr(e_0 \mid X=1)$$

$$= Q(\text{SNR}) \qquad \text{SNR} = b/\sigma.$$

Eg: Vector observation $P_X(0)=p_0$, $P_X(1)=p_1$

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

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

Observations $\underline{Y} = (Y_1, \dots, Y_n)$ are conditionally indep. of one another given the hypothesis.

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

$$\forall \underline{y} = (y_1, \dots, y_n) \in \mathbb{R}^n.$$

LRT: $$\Lambda(\underline{y}) = \prod_{j=1}^{n} \frac{f_{Y_j|X}(y_j|1)}{f_{Y_j|X}(y_j|0)} \overset{`1'}{\gtrless} \eta$$

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

$\uparrow$
log

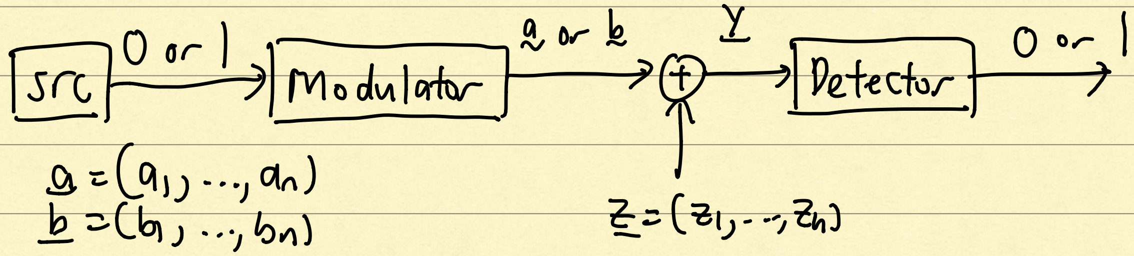
In addition if $LLR_j(Y_j)$ are conditionally identically distributed given X, $LLR_j(\cdot) = LLR(\cdot)$, the rule reduces to

$$\sum_{j=1}^{n} LLR(Y_j) \gtrless \ln \eta.$$

where $$LLR_j(y) = \ln \frac{f_{Y_j|X}(y|1)}{f_{Y_j|X}(y|0)}$$

and $$\text{LLR}(y) = \ln \frac{f_{Y|X}(y|1)}{f_{Y|X}(y|0)}$$

Binary detection of vector signals in Gaussian noise.



$\underline{z} = (z_1, \ldots, z_n)$ i.i.d. random vector $z_i \sim N(0, \sigma^2)$.

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

$$\text{LLR}(y_j) = \log\left(\underbrace{\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_j - b_j)^2}{2\sigma^2}\right)}_{P_{Y_j|X=1}(y_j|1)} \Big/ \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_j - a_j)^2}{2\sigma^2}\right)}_{P_{Y_j|X}(y_j|0)} \right)$$

j^{th} obs $j \in \{1, \ldots, n\}$

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

$$\boxed{\sum_{j=1}^{n} c_j y_j = c^T y}$$

Putting all the observations together

$$\text{LLR}(\underline{y}) = \sum_{j=1}^{n} \text{LLR}(y_j) = \sum_{j=1}^{n} \left(\frac{b_j - a_j}{\sigma}\right)\left[y_j - \left(\frac{b_j + a_j}{2}\right)\right].$$

$$= \left(\frac{\underline{b}-\underline{a}}{\sigma}\right)^T \left(\underline{y} - \left(\frac{\underline{b}+\underline{a}}{2}\right)\right) \underset{\text{'0'}}{\overset{\text{'1'}}{\gtrless}} \ln \eta$$

Rewrite this as $\qquad (b-a)^T(b+a) = b^Tb - a^Ta$

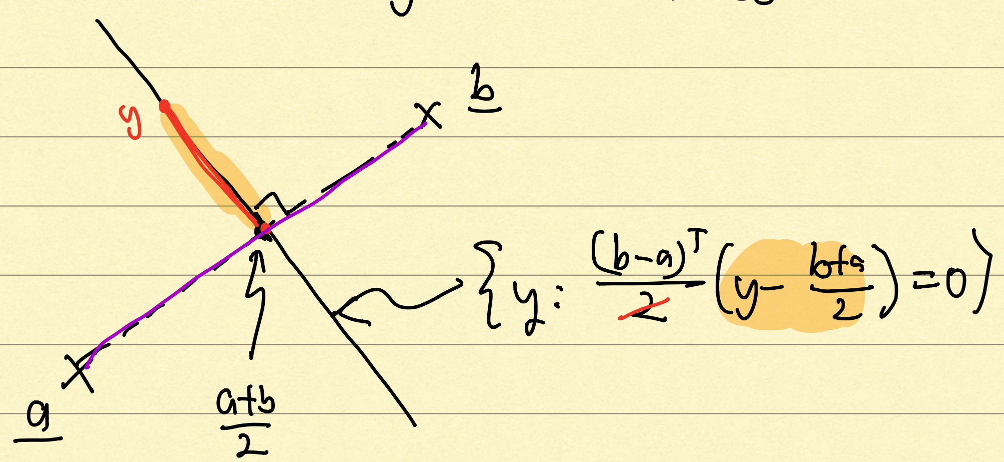
1-dim $\rightsquigarrow$

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

Originally our observation vector $\underline{y} \in \mathbb{R}^n$ is possibly very high-dim.
But, we formed a sufficient statistic $\underline{y}^T(\underline{b}-\underline{a})$ which is only 1-dim.

In the ML setting, $y^T(b-a) \overset{\text{'1'}}{\gtrless} \frac{1}{2}(\|b\|^2 - \|a\|^2)$

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

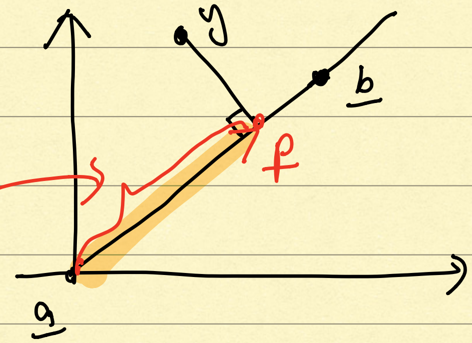


Fact: The 2 hypotheses can be distinguished by the component of y in the direction of "signal vector"

$\underline{b}-\underline{a}$.

i.e., $\frac{\underline{y}^T(\underline{b}-\underline{a})}{\|\underline{b}-\underline{a}\|}$ is a SS for BHT

$\underline{a}=\underline{0}$.



$f=\frac{(\underline{b}-\underline{a})^T\underline{y}}{\|\underline{b}-\underline{a}\|}(\underline{b}-\underline{a})$

Decision Eqn. as

$$LLR(\underline{y})=\frac{\|\underline{b}-\underline{a}\|}{\sigma^2}\Bigg(\underbrace{\frac{(\underline{b}-\underline{a})^T\underline{y}}{\|\underline{b}-\underline{a}\|}}-\frac{(\underline{b}-\underline{a})^T(\underline{b}+\underline{a})/2}{\|\underline{b}-\underline{a}\|}\Bigg)\gtrless \ln\eta$$

component of y in the signal direction

Problem reduces to the 1-dim case.

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

$$\frac{2\underline{b}^T\underline{y}}{\sigma^2}\overset{\text{'1'}}{\gtrless}\ln\eta.$$

$$Pr(e_\eta \mid X=-\underline{b})=Q\Big(\underbrace{\cancel{\frac{\ln\eta}{2\gamma}}}_{0}+\underbrace{\gamma}_{\text{scalar}}\Big)\qquad \gamma=\underbrace{\frac{\|\underline{b}\|}{\sigma}}_{\text{SNR}}$$

$Pr(e_\eta \mid X=\underline{b})=\ldots$

Consider $\Pr(e_\eta \mid X = -\underline{b})$ ← prob. of false alarm.

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

Under $X = -\underline{b}$, $\underline{Y} \overset{=\underline{X}+\underline{Z}}{\sim} N(-\underline{b}, \sigma^2 I_n)$

$$\underline{b}^T \underline{Y} \sim N(-\|b\|^2, \|b\|^2\sigma^2)$$

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

$w = (z+\|b\|^2)/\|b\|\sigma$

$$= Q\left(\frac{\ln\eta}{2\gamma} + \gamma\right), \quad \gamma = \frac{\|b\|}{\sigma}.$$

γ = Effective SNR

Eg: Poisson processes $\qquad X = 0 \Leftrightarrow PP(\lambda_0)$

$\lambda_0 \neq \lambda_1 \qquad X = 1 \Leftrightarrow PP(\lambda_1)$

Observations $(Y_1, \ldots, Y_n)$ $\qquad Y_i \equiv X_i$ interarrival times

$$f_{\underline{Y}|X}(y \mid \lambda_\ell) = \prod_{j=1}^{n} \underbrace{\lambda_\ell \exp(-\lambda_\ell y_j)}_{Exp(\lambda_\ell)}, \quad y_j \geq 0.$$

$\ell \in \{0, 1\}$

$$LLR(\underline{y}) = \log\left[\frac{f_{\underline{Y}|X}(\underline{y}|\lambda_1)}{f_{\underline{Y}|X}(\underline{y}|\lambda_0)}\right]$$

$$= \log\left[\prod_{j=1}^{n} \frac{\lambda_1 \exp(-\lambda_1 y_j)}{\lambda_0 \exp(-\lambda_0 y_j)}\right]$$

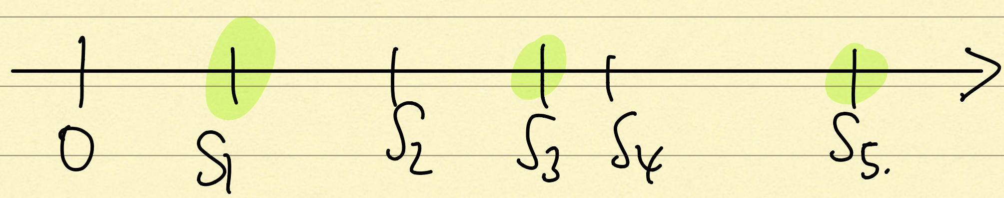
$$= n\log\left(\frac{\lambda_1}{\lambda_0}\right) + \sum_{j=1}^{n}(\lambda_0 - \lambda_1)\, y_j .$$

MAP test $\Rightarrow n\log\left(\frac{\lambda_1}{\lambda_0}\right) + (\lambda_0 - \lambda_1)\underbrace{\sum_{j=1}^{n} y_j}_{S_n} \gtrless \ln \eta .$

$S_n = \sum_{j=1}^{n} Y_j$ is a <u>suff. statistic</u> of this BHT.

$\therefore$ LR test depends only on S_n.

Intuitively true $\because$ first $n-1$ arrivals are unif. conditioned on n^{th} arrival time.



H_0: Interarrival times $\mathrm{Exp}(\lambda)$

H_1: —— " —————— $\mathrm{Erlang}(\lambda, 2)$

Sufficient statistics: Seek equivalent characterization of SS.

Recall: $V(Y)$ is a SS for BHT if $\exists\, u(v)$ s.t.

$$u(v(y)) = \Lambda(y).$$

Eg: $\underline{Y} = (Y_1, \dots, Y_n)$, $v(\underline{Y}) = \sum_{i=1}^{n} Y_i = S_n$

is easily seen to be a SS for the PP BHT.

Thm: Let $V = v(Y)$ be a $f^{\underline{n}}$ of Y for a BHT. The following are equivalent.

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

$$\boxed{I(X;Y) = I(X;V)}$$

2) For any $p_0, p_1 > 0$,

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

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

Rmk: V contain sufficient info about X compared to Y.

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

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

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

$(1 \Rightarrow 2)$

$$P_{X|Y}(0|y) = \frac{p_0 P_{Y|X}(y|0)}{p_0 P_{Y|X}(y|0) + p_1 P_{Y|X}(y|1)}$$

$p_0, p_1 > 0$

$$= \frac{p_0}{p_0 + p_1 \dfrac{P_{Y|X}(y|1)}{P_{Y|X}(y|0)}}$$

$$= \frac{p_0}{p_0 + p_1 \Lambda(y)} = \frac{p_0}{p_0 + p_1 u(v(y))}$$

making use of 1)

$P_{X|Y}(1|y)$

$\Rightarrow P_{X|Y}(0|y)$ is a function of y only through $v(y)$.

$$\boxed{\text{WTS}: P_{X|Y}(x|y) = P_{X|V}(x|v(y))}$$

$(2 \Rightarrow 3)$

$$\frac{P_{X|Y}(1|y)}{P_{X|Y}(0|y)} \overset{(2)}{=} \frac{P_{X|V}(1|v(y))}{P_{X|V}(0|v(y))}$$

Apply Bayes rule 4 times

$$\frac{P_{Y|X}(y|1)\cancel{P_X(1)}/\cancel{P_Y(y)}}{P_{Y|X}(y|0)\cancel{P_X(0)}/\cancel{P_Y(y)}} = \frac{P_{V|X}(v(y)|1)\cancel{P_X(1)}/\cancel{P_V(v(y))}}{P_{V|X}(v(y)|0)\cancel{P_X(0)}/\cancel{P_V(v(y))}}$$

$\parallel$

$$\overset{\shortparallel}{\underset{\shortparallel}{\Lambda(y)}} = \frac{P_{V|X}(v(y)|1)}{P_{V|X}(v(y)|0)} \qquad \text{—}(*)$$

The LR of Y is the LR of $v(Y)$.

$(3 \Rightarrow 1)$ The RHS of $(*)$ is a f^{n} of y only through $v(y)$.

$$\boxed{u(\cdot) = \frac{P_{V|X}(\cdot|1)}{P_{V|X}(\cdot|0)}} \Leftrightarrow \underset{v(y)}{\underset{\shortparallel}{u(v)}} = \frac{P_{V|X}(\overset{v(y)}{\overset{\shortparallel}{v}}|1)}{P_{V|X}(\underset{v(y)}{\underset{\shortparallel}{v}}|0)}$$

Neyman-Pearson Rule

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

$P_{FA} = \Pr(Y \in D | X=0)$ $= $ type-I error

$Y \in D \Rightarrow$ decision is made in favor of $X=1$.

$P_{MD} = \Pr(Y \in D^c | X=1)$ $=$ type-II error.

Always always possible to make $P_{FA} = 0$.
$(D = \phi \Leftrightarrow$ Never decide $X=1)$
$P_{MD} = 1 \Rightarrow$ very bad.

More meaningful to understand the "tradeoff" between the 2 error probs.

$$\boxed{\min \ P_{MD} \quad \text{s.t.} \quad P_{FA} \leq \varepsilon \qquad \varepsilon > 0.}$$

Abbreviate the probs as

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

A represents some decision region in favor of $X=1$.

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