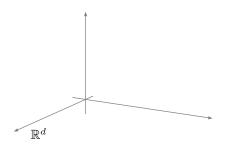
João Mota

UDRC Summer School, 2023

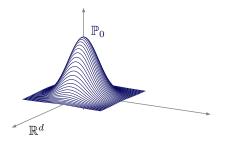
Heriot-Watt University

Problem



Problem

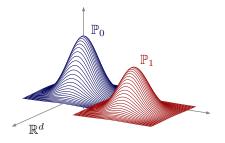
We observe $oldsymbol{X} \in \mathbb{R}^d$



Problem

We observe $oldsymbol{X} \in \mathbb{R}^d$

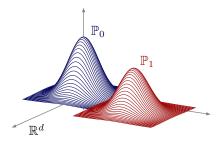
 $X\sim \mathbb{P}_0$



Problem

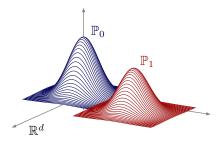
We observe $oldsymbol{X} \in \mathbb{R}^d$

 $X \sim \mathbb{P}_0$ or $X \sim \mathbb{P}_1$?



\mathbb{P}_1 ?

In classical *decision theory*, we *know* the distributions \mathbb{P}_0 and \mathbb{P}_1



Problem		
We observe	$oldsymbol{X}\in$	\mathbb{R}^d
$oldsymbol{X} \sim \mathbb{P}_0$	or	$oldsymbol{X}\sim \mathbb{P}_1?$

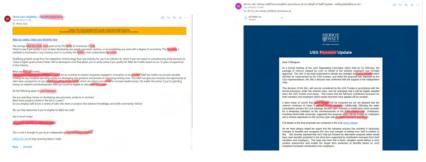
In classical *decision theory*, we *know* the distributions \mathbb{P}_0 and \mathbb{P}_1 In *machine learning*, we *have to estimate* \mathbb{P}_0 and \mathbb{P}_1 from data



۲

all rice she orkney shall bources@int serve.hwacuk on behall of Shall Lipitete + staff-polete@hwacuk.r

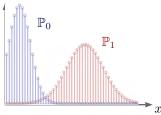
G	Check year eligibility educed an applied bits The second second The first second second	0 5 5 + -	B AFTERD And	v
	Eastern Fils and argument find a sensitivative film of the sensitive metals. Be an inference as your administrative is you should be administrative while another a metals.			HERIOT
	All tax codits. Check your alighting here			WATT
	The average field for Genes being paid and by the field is buildnesses in [16]. Check to see Furs: guiding Furs' endeed developing one proof, promotion, systems, or accompliciting any work with a day addicition to Comments in any Model and the Comments of the Section of Addicated Section (1997).	degree of uncertainty This last certain is		USS Pension Update
	Qualifying projects range from the adaptation of technology from one industry for use in an industry for which it uses not n create a higher qualit product factor. We've developed a load that allows pro to remly indire if you qualify for R&O Tac Cent in the industry.			Dear-Colleagues
	- ten many Weat are (bit). Weat are transition to mean a membra to mean a comparise simplicity in immediate. A UK sum of a factor of a second or any second symposity and the mean and the second or any second symposity and the mean is meaning with a second or any second symposity and the second or any second symposity and the se	her can give your business the apportunity to		As a lowest meeting of the Jans Stageslang Convention (JAG) and on 32 followary the paragage of motions instead by Jack on to head of the advense explayers and through approach. The JACA is the lodge authorized is include any changes to some taxable and these to improvements by the onlike lowests, and advents the proposal is many parent by the UCU separameterism. The JACA is blockness, and advect the proposal is many parent by the UCU separameterism. The JACA is blockness, and advect the proposal is many parent by the UCU separameterism. The JACA is blockness, and advect the proposal is many parent by the UCU separameterism.
	De the following apply to year landmitt			This decision of the JWC, will now be considered by the USS Trustee in accordance with the formal processes under the scheme rules, and we anticipate that it will be legally adopted when the USS Trustee next memis. This means that for full lask contribution schemes for both contributions and metallicity adoption adoption and and the scheme.
	And pile genoting-investig on exemptions for all processing products or services: The first frame project indicated in the last 2 stream of the projects that advance lossedwidge, and chefy unsertaining? Advice We can help determine if you've eligible for BAD's accord.	itan		It dies maan of course that appending will be repeated but on any pleased that the software community to which a please threader any software the software consistement and the please pleased appendix on tables at a software which appendix to a temporary transition to the commonweal of the 20m strategies with appendix to a temporary temporary to appendix to the commonweal of the 20m strategies and building temporary temporary temporary temporary temporary temporary building temporary t
	Get in touch today			and a hodest adjustment to the secondry plan with or additional cost is members. Full details of the final programming are considered in the ULK stress schedule.
	This + mail is brought to go by an independent of an independent of an independent of a second of the second of th			The means in the two puppeds are transmitted in the Out-Table States States At an In-Section States and an approximate the two shades processing the states and the states

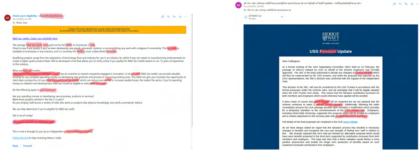


$X \in \mathbb{R}$: number of **spam words** in a message



$X \in \mathbb{R}$: number of spam words in a message

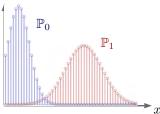


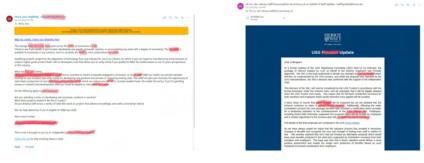


$X \in \mathbb{R}$: number of spam words in a message

Null Hypothesis

 H_0 : message isn't spam





 $X \in \mathbb{R}$: number of spam words in a message

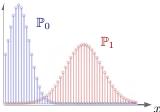
Null Hypothesis

 H_0 : message isn't spam

Alternative Hypothesis

 H_1 : message is spam

Decision Theory



Consider a test for detecting:

if given email is spam

Consider a test for detecting:

if given email is spam

presence of aircraft in radar

Consider a test for detecting:

if given email is spam

if defendant is guilty

presence of aircraft in radar

Consider a test for detecting:

if given email is spampresence of aircraft in radarif defendant is guiltypresence of tumor in an image

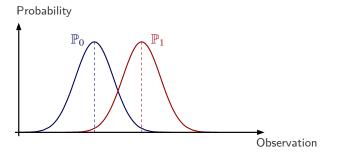
Probability

Consider a test for detecting:

if given email is spam

if defendant is guilty

presence of aircraft in radar

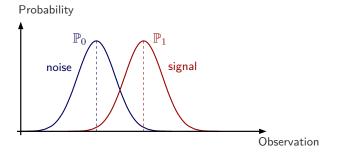


Consider a test for detecting:

if given email is spam

if defendant is guilty

presence of aircraft in radar

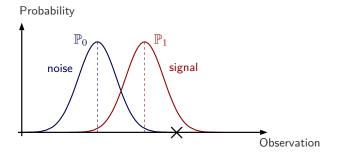


Consider a test for detecting:

if given email is spam

if defendant is guilty

presence of aircraft in radar

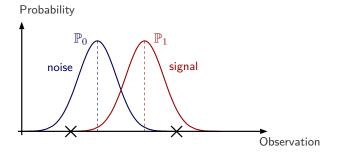


Consider a test for detecting:

if given email is spam

if defendant is guilty

presence of aircraft in radar

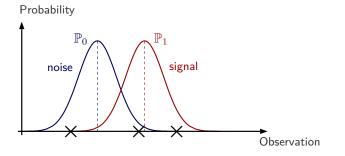


Consider a test for detecting:

if given email is spam

if defendant is guilty

presence of aircraft in radar

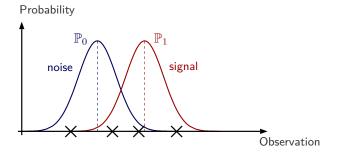


Consider a test for detecting:

if given email is spam

if defendant is guilty

presence of aircraft in radar

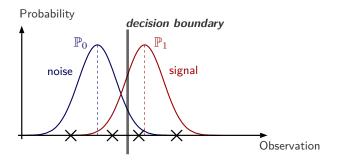


Consider a test for detecting:

if given email is spam prese

if defendant is guilty

presence of aircraft in radar presence of tumor in an image

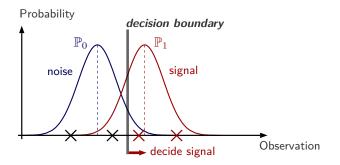


Consider a test for detecting:

if given email is spam presence

if defendant is guilty

presence of aircraft in radar presence of tumor in an image

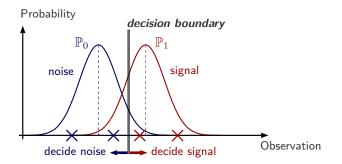


Consider a test for detecting:

if given email is spam presence of

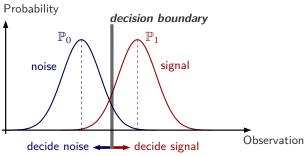
if defendant is guilty

presence of aircraft in radar presence of tumor in an image

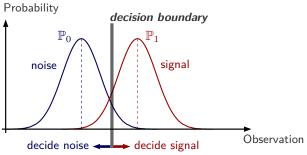


True hypothesis	Decide noise	Decide signal
noise		
signal		

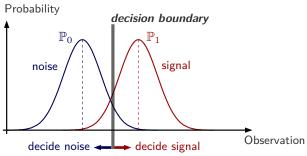
True hypothesis	Decide noise	Decide signal
noise		
signal		



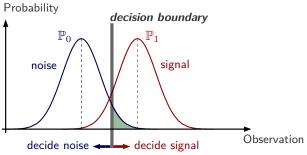
True hypothesis	Decide noise	Decide signal
noise	\checkmark	
signal		



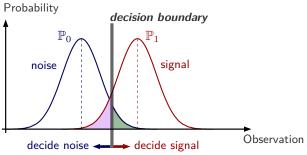
True hypothesis	Decide noise	Decide signal
noise	\checkmark	
signal		\checkmark



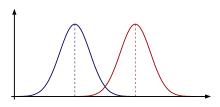
True hypothesis	Decide noise	Decide signal
noise	\checkmark	false alarm
signal		\checkmark



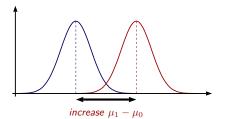
True hypothesis	Decide noise	Decide signal
noise	\checkmark	false alarm
signal	missed detection	\checkmark

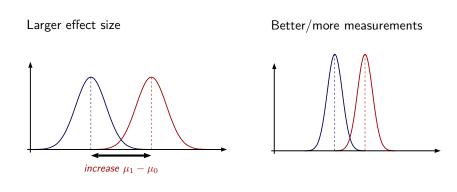


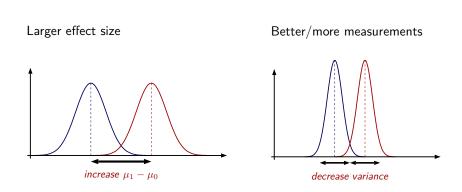
Larger effect size



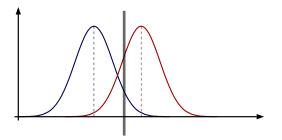
Larger effect size

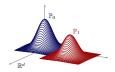


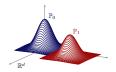




Where to place the decision boundary?

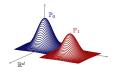






True label

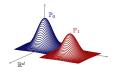
$$Y = \left\{ \begin{array}{rrr} 0 & , \mbox{ if } H_0 \mbox{ is true} \\ \\ 1 & , \mbox{ if } H_1 \mbox{ true} \end{array} \right.$$



True label

Decision function $f : \mathbb{R}^d \to \{0, 1\}$

$$Y = \begin{cases} 0 & , \text{ if } H_0 \text{ is true} \\ 1 & , \text{ if } H_1 \text{ true} \end{cases} \qquad f(\mathbf{X}) = \begin{cases} 0 & , \text{ if we } \underline{decide} \ H_0 \\ 1 & , \text{ if we } \underline{decide} \ H_1 \end{cases}$$

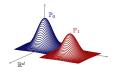


True label

Decision function $f : \mathbb{R}^d \to \{0, 1\}$

 $Y = \begin{cases} 0 & , \text{ if } H_0 \text{ is true} \\ 1 & , \text{ if } H_1 \text{ true} \end{cases} \qquad f(\mathbf{X}) = \begin{cases} 0 & , \text{ if we } \underline{decide} \ H_0 \\ 1 & , \text{ if we } \underline{decide} \ H_1 \end{cases}$

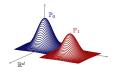
Loss function $\ell: \{0,1\} \times \{0,1\} \to \mathbb{R}$



True label

Decision function $f : \mathbb{R}^d \to \{0, 1\}$

 $Y = \begin{cases} 0 & , \text{ if } H_0 \text{ is true} \\ 1 & , \text{ if } H_1 \text{ true} \end{cases} \qquad f(\mathbf{X}) = \begin{cases} 0 & , \text{ if we } \underline{decide} \ H_0 \\ 1 & , \text{ if we } \underline{decide} \ H_1 \end{cases}$

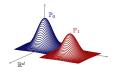


True label

Decision function $f : \mathbb{R}^d \to \{0, 1\}$

 $Y = \begin{cases} 0 & , \text{ if } H_0 \text{ is true} \\ 1 & , \text{ if } H_1 \text{ true} \end{cases} \qquad f(\mathbf{X}) = \begin{cases} 0 & , \text{ if we } \underline{decide} \ H_0 \\ 1 & , \text{ if we } \underline{decide} \ H_1 \end{cases}$

True hypothesis	$f(\boldsymbol{X}) = 0$	$f(\boldsymbol{X}) = 1$
H_0 is true		
H_1 is true		

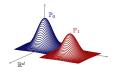


True label

Decision function $f : \mathbb{R}^d \to \{0, 1\}$

 $Y = \begin{cases} 0 & , \text{ if } H_0 \text{ is true} \\ 1 & , \text{ if } H_1 \text{ true} \end{cases} \qquad f(\mathbf{X}) = \begin{cases} 0 & , \text{ if we } \underline{decide} \ H_0 \\ 1 & , \text{ if we } \underline{decide} \ H_1 \end{cases}$

True hypothesis	$f(\boldsymbol{X})=0$	$f(\boldsymbol{X}) = 1$
H_0 is true	$\ell(0,0)$	
H_1 is true		

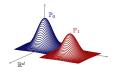


True label

Decision function $f : \mathbb{R}^d \to \{0, 1\}$

 $Y = \begin{cases} 0 & , \text{ if } H_0 \text{ is true} \\ 1 & , \text{ if } H_1 \text{ true} \end{cases} \qquad f(\mathbf{X}) = \begin{cases} 0 & , \text{ if we } \underline{decide} \ H_0 \\ 1 & , \text{ if we } \underline{decide} \ H_1 \end{cases}$

True hypothesis	$f(\boldsymbol{X})=0$	$f(\boldsymbol{X}) = 1$
H_0 is true	$\ell(0,0)$	$\ell(1,0)$
H_1 is true		

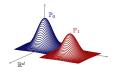


True label

Decision function $f : \mathbb{R}^d \to \{0, 1\}$

 $Y = \begin{cases} 0 & , \text{ if } H_0 \text{ is true} \\ 1 & , \text{ if } H_1 \text{ true} \end{cases} \qquad f(\mathbf{X}) = \begin{cases} 0 & , \text{ if we } \underline{decide} \ H_0 \\ 1 & , \text{ if we } \underline{decide} \ H_1 \end{cases}$

True hypothesis	$f(\boldsymbol{X}) = 0$	$f(\boldsymbol{X}) = 1$
H_0 is true	$\ell(0,0)$	$\ell(1,0)$
H_1 is true	$\ell(0,1)$	



True label

Decision function $f : \mathbb{R}^d \to \{0, 1\}$

 $Y = \begin{cases} 0 & , \text{ if } H_0 \text{ is true} \\ 1 & , \text{ if } H_1 \text{ true} \end{cases} \qquad f(\mathbf{X}) = \begin{cases} 0 & , \text{ if we } \underline{decide} \ H_0 \\ 1 & , \text{ if we } \underline{decide} \ H_1 \end{cases}$

True hypothesis	$f(\boldsymbol{X}) = 0$	$f(\boldsymbol{X}) = 1$
H_0 is true	$\ell(0,0)$	$\ell(1,0)$
H_1 is true	$\ell(0,1)$	$\ell(1,1)$

Given decision function $f : \mathbb{R}^d \to \{0,1\}$ and loss $\ell : \{0,1\}^2 \to \mathbb{R}$,

Given decision function $f : \mathbb{R}^d \to \{0,1\}$ and loss $\ell : \{0,1\}^2 \to \mathbb{R}$,

Risk:

Given decision function $f : \mathbb{R}^d \to \{0,1\}$ and loss $\ell : \{0,1\}^2 \to \mathbb{R}$,

Risk: $R[f] := \mathbb{E}_{XY} \Big[\ell \big(f(X), Y \big) \Big]$

Given decision function $f : \mathbb{R}^d \to \{0,1\}$ and loss $\ell : \{0,1\}^2 \to \mathbb{R}$,

Risk:
$$R[f] := \mathbb{E}_{XY} \Big[\ell \big(f(X), Y \big) \Big]$$

where $\mathbb{E}_{XY}[\cdot]$ is the expectation with respect to X and Y

Given decision function $f: \mathbb{R}^d \to \{0,1\}$ and loss $\ell: \{0,1\}^2 \to \mathbb{R}$,

Risk:
$$R[f] := \mathbb{E}_{XY} \Big[\ell \big(f(X), Y \big) \Big]$$

where $\mathbb{E}_{XY}[\cdot]$ is the expectation with respect to X and Y

Optimal decision problem:

Given decision function $f : \mathbb{R}^d \to \{0,1\}$ and loss $\ell : \{0,1\}^2 \to \mathbb{R}$,

Risk:
$$R[f] := \mathbb{E}_{XY} \left[\ell(f(X), Y) \right]$$

where $\mathbb{E}_{XY}[\cdot]$ is the expectation with respect to X and Y

Optimal decision problem: Given ℓ , find f that minimizes the risk:

Given decision function $f : \mathbb{R}^d \to \{0,1\}$ and loss $\ell : \{0,1\}^2 \to \mathbb{R}$,

Risk:
$$R[f] := \mathbb{E}_{XY} \left[\ell(f(X), Y) \right]$$

where $\mathbb{E}_{\boldsymbol{X}Y}[\cdot]$ is the expectation with respect to \boldsymbol{X} and Y

Optimal decision problem: Given ℓ , find f that minimizes the risk:

$$\underset{f: \mathbb{R}^d \to \{0,1\}}{\text{minimize}} \ \mathbb{E}_{\boldsymbol{X}Y} \Big[\ell \big(f(\boldsymbol{X}), Y \big) \Big]$$

Given decision function $f: \mathbb{R}^d \to \{0,1\}$ and loss $\ell: \{0,1\}^2 \to \mathbb{R}$,

Risk:
$$R[f] := \mathbb{E}_{XY} \left[\ell(f(X), Y) \right]$$

where $\mathbb{E}_{\boldsymbol{X}Y}[\cdot]$ is the expectation with respect to \boldsymbol{X} and Y

Optimal decision problem: Given ℓ , find f that minimizes the risk:

$$\underset{f:\mathbb{R}^d \rightarrow \{0,1\}}{\text{minimize}} \ \mathbb{E}_{\boldsymbol{X}Y} \Big[\ell \big(f(\boldsymbol{X}), Y \big) \Big]$$

... infinite-dimensional problem

 $\underset{f: \mathbb{R}^d \rightarrow \{0,1\}}{\text{minimize}} \ \mathbb{E}_{\boldsymbol{X}Y} \Big[\ell \big(f(\boldsymbol{X}), Y \big) \Big]$

```
Recall that f(\mathbf{X}) and Y are binary
```

 $\underset{f: \mathbb{R}^d \rightarrow \{0,1\}}{\text{minimize}} \ \mathbb{E}_{\boldsymbol{X}Y} \Big[\ell \big(f(\boldsymbol{X}), Y \big) \Big]$

 $\underset{f:\mathbb{R}^d \rightarrow \{0,1\}}{\text{minimize}} \ \mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}),Y\big)\Big]$

Recall that $f(\boldsymbol{X})$ and Y are binary

Conditioning on X,

Recall that $f(\mathbf{X})$ and Y are binary Conditioning on \mathbf{X} ,

$$\mathbb{E}_{\boldsymbol{X}Y}\left[\ell(f(\boldsymbol{X}),Y)\right] = \mathbb{E}_{\boldsymbol{X}}\left[\mathbb{E}_{Y}\left[\ell(f(\boldsymbol{X}),Y) \mid \boldsymbol{X}\right]\right]$$

Recall that $f(\boldsymbol{X})$ and Y are binary Conditioning on \boldsymbol{X} ,

$$\mathbb{E}_{\boldsymbol{X}Y}\left[\ell(f(\boldsymbol{X}), Y)\right] = \mathbb{E}_{\boldsymbol{X}}\left[\mathbb{E}_{Y}\left[\ell(f(\boldsymbol{X}), Y) \mid \boldsymbol{X}\right]\right]$$
$$= \int_{\mathbb{R}^{d}} \mathbb{E}_{Y}\left[\ell(f(\boldsymbol{X}), Y) \mid \boldsymbol{X} = \boldsymbol{x}\right] f_{\boldsymbol{X}}(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}$$

Recall that $f(\boldsymbol{X})$ and Y are binary Conditioning on \boldsymbol{X} ,

$$\mathbb{E}_{\boldsymbol{X}\boldsymbol{Y}}\Big[\ell\big(f(\boldsymbol{X}),\boldsymbol{Y}\big)\Big] = \mathbb{E}_{\boldsymbol{X}}\Big[\mathbb{E}_{\boldsymbol{Y}}\Big[\ell\big(f(\boldsymbol{X}),\boldsymbol{Y}\big)\,\Big|\,\boldsymbol{X}\Big]\Big]$$
$$= \int_{\mathbb{R}^d} \mathbb{E}_{\boldsymbol{Y}}\Big[\ell\big(f(\boldsymbol{X}),\boldsymbol{Y}\big)\,\Big|\,\boldsymbol{X} = \boldsymbol{x}\Big]f_{\boldsymbol{X}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$$

If $f(\boldsymbol{x}) = 0$,

Recall that $f(\boldsymbol{X})$ and Y are binary Conditioning on \boldsymbol{X} ,

$$\mathbb{E}_{\boldsymbol{X}\boldsymbol{Y}}\Big[\ell\big(f(\boldsymbol{X}),\boldsymbol{Y}\big)\Big] = \mathbb{E}_{\boldsymbol{X}}\Big[\mathbb{E}_{\boldsymbol{Y}}\Big[\ell\big(f(\boldsymbol{X}),\,\boldsymbol{Y}\big)\,\Big|\,\boldsymbol{X}\Big]\Big]$$
$$= \int_{\mathbb{R}^d} \mathbb{E}_{\boldsymbol{Y}}\Big[\ell\big(f(\boldsymbol{X}),\,\boldsymbol{Y}\big)\,\Big|\,\boldsymbol{X} = \boldsymbol{x}\Big]f_{\boldsymbol{X}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$$

If
$$f(\boldsymbol{x}) = 0$$
,

$$\mathbb{E}_{Y}\left[\ell(0, Y) \mid \boldsymbol{X} = \boldsymbol{x}\right] = \ell(0, 0) \mathbb{P}(Y = 0 \mid \boldsymbol{X} = \boldsymbol{x}) + \ell(0, 1) \mathbb{P}(Y = 1 \mid \boldsymbol{X} = \boldsymbol{x})$$

Recall that $f(\mathbf{X})$ and Y are binary Conditioning on \mathbf{X} ,

$$\mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}),Y\big)\Big] = \mathbb{E}_{\boldsymbol{X}}\Big[\mathbb{E}_{Y}\Big[\ell\big(f(\boldsymbol{X}),Y\big)\,\Big|\,\boldsymbol{X}\Big]\Big]$$
$$= \int_{\mathbb{R}^{d}} \mathbb{E}_{Y}\Big[\ell\big(f(\boldsymbol{X}),Y\big)\,\Big|\,\boldsymbol{X} = \boldsymbol{x}\Big]f_{\boldsymbol{X}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$$

If
$$f(\boldsymbol{x}) = 0$$
,

$$\mathbb{E}_{Y}\left[\ell(0, Y) \mid \boldsymbol{X} = \boldsymbol{x}\right] = \ell(0, 0) \mathbb{P}(Y = 0 \mid \boldsymbol{X} = \boldsymbol{x}) + \ell(0, 1) \mathbb{P}(Y = 1 \mid \boldsymbol{X} = \boldsymbol{x})$$
If $f(\boldsymbol{x}) = 1$,

Recall that $f(\mathbf{X})$ and Y are binary Conditioning on \mathbf{X} ,

$$\mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}),Y\big)\Big] = \mathbb{E}_{\boldsymbol{X}}\Big[\mathbb{E}_{Y}\Big[\ell\big(f(\boldsymbol{X}),Y\big) \mid \boldsymbol{X}\Big]\Big]$$
$$= \int_{\mathbb{R}^{d}} \mathbb{E}_{Y}\Big[\ell\big(f(\boldsymbol{X}),Y\big) \mid \boldsymbol{X} = \boldsymbol{x}\Big]f_{\boldsymbol{X}}(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x}$$

If $f(\boldsymbol{x}) = 0$, $\mathbb{E}_{Y}\left[\ell(0, Y) \mid \boldsymbol{X} = \boldsymbol{x}\right] = \ell(0, 0) \mathbb{P}(Y = 0 \mid \boldsymbol{X} = \boldsymbol{x}) + \ell(0, 1) \mathbb{P}(Y = 1 \mid \boldsymbol{X} = \boldsymbol{x})$ If $f(\boldsymbol{x}) = 1$, $\mathbb{E}_{Y}\left[\ell(1, Y) \mid \boldsymbol{X} = \boldsymbol{x}\right] = \ell(1, 0) \mathbb{P}(Y = 0 \mid \boldsymbol{X} = \boldsymbol{x}) + \ell(1, 1) \mathbb{P}(Y = 1 \mid \boldsymbol{X} = \boldsymbol{x})$

Decision Theory

Optimal decision

Optimal decision

$$f(\boldsymbol{x}) = 0$$
 if $\mathbb{E}_{Y}\left[\ell\left(0, Y
ight) \middle| \boldsymbol{X} = \boldsymbol{x}
ight] < \mathbb{E}_{Y}\left[\ell\left(1, Y
ight) \middle| \boldsymbol{X} = \boldsymbol{x}
ight]$

$$f(\boldsymbol{x}) = 0 \quad \text{if} \quad \mathbb{E}_{Y}\left[\ell(0, Y) \mid \boldsymbol{X} = \boldsymbol{x}\right] < \mathbb{E}_{Y}\left[\ell(1, Y) \mid \boldsymbol{X} = \boldsymbol{x}\right]$$
$$f(\boldsymbol{x}) = 1 \quad \text{if} \quad \mathbb{E}_{Y}\left[\ell(0, Y) \mid \boldsymbol{X} = \boldsymbol{x}\right] \ge \mathbb{E}_{Y}\left[\ell(1, Y) \mid \boldsymbol{X} = \boldsymbol{x}\right]$$

$$\begin{split} f(\boldsymbol{x}) &= 0 \quad \text{ if } \quad \mathbb{E}_{Y} \bigg[\ell \big(0, \, Y \big) \, \Big| \, \boldsymbol{X} = \boldsymbol{x} \bigg] \, < \, \mathbb{E}_{Y} \bigg[\ell \big(1, \, Y \big) \, \Big| \, \boldsymbol{X} = \boldsymbol{x} \bigg] \\ f(\boldsymbol{x}) &= 1 \quad \text{ if } \quad \mathbb{E}_{Y} \bigg[\ell \big(0, \, Y \big) \, \Big| \, \boldsymbol{X} = \boldsymbol{x} \bigg] \, \ge \, \mathbb{E}_{Y} \bigg[\ell \big(1, \, Y \big) \, \Big| \, \boldsymbol{X} = \boldsymbol{x} \bigg] \end{split}$$

$$\begin{aligned} f(\boldsymbol{x}) &= 0 \quad \text{if} \quad \mathbb{E}_{Y} \bigg[\ell \big(0, Y \big) \, \Big| \, \boldsymbol{X} = \boldsymbol{x} \bigg] \, < \, \mathbb{E}_{Y} \bigg[\ell \big(1, Y \big) \, \Big| \, \boldsymbol{X} = \boldsymbol{x} \bigg] \\ f(\boldsymbol{x}) &= 1 \quad \text{if} \quad \mathbb{E}_{Y} \bigg[\ell \big(0, Y \big) \, \Big| \, \boldsymbol{X} = \boldsymbol{x} \bigg] \, \ge \, \mathbb{E}_{Y} \bigg[\ell \big(1, Y \big) \, \Big| \, \boldsymbol{X} = \boldsymbol{x} \bigg] \end{aligned}$$

$$f(\boldsymbol{x}) = \begin{cases} 1 & \text{if } \mathbb{P}(Y = 1 \mid \boldsymbol{X} = \boldsymbol{x}) \geq \frac{\ell(0,0) - \ell(1,0)}{\ell(1,1) - \ell(0,1)} \mathbb{P}(Y = 0 \mid \boldsymbol{X} = \boldsymbol{x}) \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{split} f(\boldsymbol{x}) &= 0 \quad \text{ if } \quad \mathbb{E}_{Y} \bigg[\ell \big(0, Y \big) \, \Big| \, \boldsymbol{X} = \boldsymbol{x} \bigg] \, < \, \mathbb{E}_{Y} \bigg[\ell \big(1, Y \big) \, \Big| \, \boldsymbol{X} = \boldsymbol{x} \bigg] \\ f(\boldsymbol{x}) &= 1 \quad \text{ if } \quad \mathbb{E}_{Y} \bigg[\ell \big(0, Y \big) \, \Big| \, \boldsymbol{X} = \boldsymbol{x} \bigg] \, \ge \, \mathbb{E}_{Y} \bigg[\ell \big(1, Y \big) \, \Big| \, \boldsymbol{X} = \boldsymbol{x} \bigg] \end{split}$$

$$f(\boldsymbol{x}) = \begin{cases} 1 & \text{if } \mathbb{P}(\frac{\boldsymbol{Y} = 1}{\boldsymbol{Y} = 1} \mid \boldsymbol{X} = \boldsymbol{x}) \geq \frac{\ell(0,0) - \ell(1,0)}{\ell(1,1) - \ell(0,1)} \mathbb{P}(\boldsymbol{Y} = 0 \mid \boldsymbol{X} = \boldsymbol{x}) \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{split} f(\boldsymbol{x}) &= 0 \quad \text{if} \quad \mathbb{E}_{Y} \bigg[\ell \big(0, Y \big) \, \Big| \, \boldsymbol{X} = \boldsymbol{x} \bigg] \, < \, \mathbb{E}_{Y} \bigg[\ell \big(1, Y \big) \, \Big| \, \boldsymbol{X} = \boldsymbol{x} \bigg] \\ f(\boldsymbol{x}) &= 1 \quad \text{if} \quad \mathbb{E}_{Y} \bigg[\ell \big(0, Y \big) \, \Big| \, \boldsymbol{X} = \boldsymbol{x} \bigg] \, \ge \, \mathbb{E}_{Y} \bigg[\ell \big(1, Y \big) \, \Big| \, \boldsymbol{X} = \boldsymbol{x} \bigg] \end{split}$$

$$f(\boldsymbol{x}) = \begin{cases} 1 & \text{if } \mathbb{P}(\frac{\boldsymbol{Y} = 1}{\boldsymbol{Y} = 1} \mid \boldsymbol{X} = \boldsymbol{x}) \geq \frac{\ell(0,0) - \ell(1,0)}{\ell(1,1) - \ell(0,1)} \mathbb{P}(\frac{\boldsymbol{Y} = 0}{\boldsymbol{Y} = 0} \mid \boldsymbol{X} = \boldsymbol{x}) \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{split} f(\boldsymbol{x}) &= 0 \quad \text{if} \quad \mathbb{E}_{Y} \bigg[\ell \big(0, Y \big) \, \Big| \, \boldsymbol{X} = \boldsymbol{x} \bigg] \, < \, \mathbb{E}_{Y} \bigg[\ell \big(1, Y \big) \, \Big| \, \boldsymbol{X} = \boldsymbol{x} \bigg] \\ f(\boldsymbol{x}) &= 1 \quad \text{if} \quad \mathbb{E}_{Y} \bigg[\ell \big(0, Y \big) \, \Big| \, \boldsymbol{X} = \boldsymbol{x} \bigg] \, \ge \, \mathbb{E}_{Y} \bigg[\ell \big(1, Y \big) \, \Big| \, \boldsymbol{X} = \boldsymbol{x} \bigg] \end{split}$$

$$f(\boldsymbol{x}) = \begin{cases} \begin{array}{c} H_1 \\ 1 & \text{if } \mathbb{P}(\frac{\boldsymbol{Y} = 1}{\boldsymbol{Y} = 1} \mid \boldsymbol{X} = \boldsymbol{x}) \geq \frac{\ell(0,0) - \ell(1,0)}{\ell(1,1) - \ell(0,1)} \mathbb{P}(\frac{\boldsymbol{Y} = 0}{\boldsymbol{Y} = 0} \mid \boldsymbol{X} = \boldsymbol{x}) \\ 0 & \text{otherwise} \end{cases}$$
Bayes rule

$$\begin{aligned} f(\boldsymbol{x}) &= 0 \quad \text{if} \quad \mathbb{E}_{Y} \bigg[\ell \big(0, Y \big) \, \Big| \, \boldsymbol{X} = \boldsymbol{x} \bigg] \, < \, \mathbb{E}_{Y} \bigg[\ell \big(1, Y \big) \, \Big| \, \boldsymbol{X} = \boldsymbol{x} \bigg] \\ f(\boldsymbol{x}) &= 1 \quad \text{if} \quad \mathbb{E}_{Y} \bigg[\ell \big(0, Y \big) \, \Big| \, \boldsymbol{X} = \boldsymbol{x} \bigg] \, \ge \, \mathbb{E}_{Y} \bigg[\ell \big(1, Y \big) \, \Big| \, \boldsymbol{X} = \boldsymbol{x} \bigg] \end{aligned}$$

$$f(\boldsymbol{x}) = \begin{cases} 1 & \text{if } \mathbb{P}(\frac{\boldsymbol{Y} = 1}{\boldsymbol{Y} = 1} \mid \boldsymbol{X} = \boldsymbol{x}) \geq \frac{\ell(0,0) - \ell(1,0)}{\ell(1,1) - \ell(0,1)} \mathbb{P}(\frac{\boldsymbol{Y} = 0}{\boldsymbol{Y} = 0} \mid \boldsymbol{X} = \boldsymbol{x}) \\ 0 & \text{otherwise} \\ \frac{f_{\boldsymbol{X}|H_1}(\boldsymbol{x} \mid H_1) \mathbb{P}(H_1)}{f_{\boldsymbol{X}}(\boldsymbol{x})} \end{bmatrix} \text{Bayes rule}$$

$$\begin{aligned} f(\boldsymbol{x}) &= 0 \quad \text{if} \quad \mathbb{E}_{Y} \bigg[\ell \big(0, Y \big) \, \Big| \, \boldsymbol{X} = \boldsymbol{x} \bigg] \, < \, \mathbb{E}_{Y} \bigg[\ell \big(1, Y \big) \, \Big| \, \boldsymbol{X} = \boldsymbol{x} \bigg] \\ f(\boldsymbol{x}) &= 1 \quad \text{if} \quad \mathbb{E}_{Y} \bigg[\ell \big(0, Y \big) \, \Big| \, \boldsymbol{X} = \boldsymbol{x} \bigg] \, \ge \, \mathbb{E}_{Y} \bigg[\ell \big(1, Y \big) \, \Big| \, \boldsymbol{X} = \boldsymbol{x} \bigg] \end{aligned}$$

$$f(\boldsymbol{x}) = \begin{cases} 1 & \text{if } \mathbb{P}(\underline{Y=1} \mid \boldsymbol{X} = \boldsymbol{x}) \geq \frac{\ell(0,0) - \ell(1,0)}{\ell(1,1) - \ell(0,1)} \mathbb{P}(\underline{Y=0} \mid \boldsymbol{X} = \boldsymbol{x}) \\ 0 & \text{otherwise} \\ \frac{f_{\boldsymbol{X}\mid H_1}(\boldsymbol{x} \mid H_1) \mathbb{P}(H_1)}{f_{\boldsymbol{X}}(\boldsymbol{x})} \end{bmatrix} \text{Bayes rule} \frac{f_{\boldsymbol{X}\mid H_0}(\boldsymbol{x} \mid H_0) \mathbb{P}(H_0)}{f_{\boldsymbol{X}}(\boldsymbol{x})} \end{cases}$$

$$f(\boldsymbol{x}) = 0 \quad \text{if} \quad \mathbb{E}_{Y}\left[\ell(0, Y) \mid \boldsymbol{X} = \boldsymbol{x}\right] < \mathbb{E}_{Y}\left[\ell(1, Y) \mid \boldsymbol{X} = \boldsymbol{x}\right]$$
$$f(\boldsymbol{x}) = 1 \quad \text{if} \quad \mathbb{E}_{Y}\left[\ell(0, Y) \mid \boldsymbol{X} = \boldsymbol{x}\right] \ge \mathbb{E}_{Y}\left[\ell(1, Y) \mid \boldsymbol{X} = \boldsymbol{x}\right]$$

$$f(\boldsymbol{x}) = \begin{cases} 1 & \text{if } \mathbb{P}(\boldsymbol{Y}=\boldsymbol{1} \mid \boldsymbol{X}=\boldsymbol{x}) \geq \frac{\ell(0,0)-\ell(1,0)}{\ell(1,1)-\ell(0,1)} \mathbb{P}(\boldsymbol{Y}=\boldsymbol{0} \mid \boldsymbol{X}=\boldsymbol{x}) \\ 0 & \text{otherwise} \\ \frac{f_{\boldsymbol{X}\mid\boldsymbol{H}_{1}}(\boldsymbol{x}\mid\boldsymbol{H}_{1}) \mathbb{P}(\boldsymbol{H}_{1})}{f_{\boldsymbol{X}}(\boldsymbol{x})} \end{bmatrix} \begin{array}{l} \textbf{Bayes rule} \\ \frac{f_{\boldsymbol{X}\mid\boldsymbol{H}_{0}}(\boldsymbol{x}\mid\boldsymbol{H}_{0}) \mathbb{P}(\boldsymbol{H}_{0})}{f_{\boldsymbol{X}}(\boldsymbol{x})} \end{bmatrix}$$

$$f(\boldsymbol{x}) = 1 \quad \text{ if } \quad \frac{f_{\boldsymbol{X}|H_1}(\boldsymbol{x} \,|\, H_1)}{f_{\boldsymbol{X}|H_0}(\boldsymbol{x} \,|\, H_0)} \geq \frac{\ell(0,0) - \ell(1,0)}{\ell(1,1) - \ell(0,1)} \cdot \frac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)}$$

$$f(\boldsymbol{x}) = 0 \quad \text{if} \quad \mathbb{E}_{Y}\left[\ell(0, Y) \mid \boldsymbol{X} = \boldsymbol{x}\right] < \mathbb{E}_{Y}\left[\ell(1, Y) \mid \boldsymbol{X} = \boldsymbol{x}\right]$$
$$f(\boldsymbol{x}) = 1 \quad \text{if} \quad \mathbb{E}_{Y}\left[\ell(0, Y) \mid \boldsymbol{X} = \boldsymbol{x}\right] \ge \mathbb{E}_{Y}\left[\ell(1, Y) \mid \boldsymbol{X} = \boldsymbol{x}\right]$$

$$f(\boldsymbol{x}) = \begin{cases} 1 & \text{if } \mathbb{P}(\boldsymbol{Y}=\boldsymbol{1} \mid \boldsymbol{X}=\boldsymbol{x}) \geq \frac{\ell(0,0)-\ell(1,0)}{\ell(1,1)-\ell(0,1)} \mathbb{P}(\boldsymbol{Y}=\boldsymbol{0} \mid \boldsymbol{X}=\boldsymbol{x}) \\ 0 & \text{otherwise} \\ \frac{f_{\boldsymbol{X}\mid\boldsymbol{H}_1}(\boldsymbol{x}\mid\boldsymbol{H}_1) \mathbb{P}(\boldsymbol{H}_1)}{f_{\boldsymbol{X}}(\boldsymbol{x})} \end{bmatrix} \text{Bayes rule} \frac{f_{\boldsymbol{X}\mid\boldsymbol{H}_0}(\boldsymbol{x}\mid\boldsymbol{H}_0) \mathbb{P}(\boldsymbol{H}_0)}{f_{\boldsymbol{X}}(\boldsymbol{x})} \end{cases}$$

$$f(\boldsymbol{x}) = 1 \quad \text{if} \quad \frac{f_{\boldsymbol{X}|\boldsymbol{H}_1}(\boldsymbol{x} \mid \boldsymbol{H}_1)}{f_{\boldsymbol{X}|\boldsymbol{H}_0}(\boldsymbol{x} \mid \boldsymbol{H}_0)} \geq \frac{\ell(0,0) - \ell(1,0)}{\ell(1,1) - \ell(0,1)} \cdot \frac{\mathbb{P}(\boldsymbol{H}_0)}{\mathbb{P}(\boldsymbol{H}_1)}$$

$$f(\boldsymbol{x}) = 0 \quad \text{if} \quad \mathbb{E}_{Y}\left[\ell(0, Y) \mid \boldsymbol{X} = \boldsymbol{x}\right] < \mathbb{E}_{Y}\left[\ell(1, Y) \mid \boldsymbol{X} = \boldsymbol{x}\right]$$
$$f(\boldsymbol{x}) = 1 \quad \text{if} \quad \mathbb{E}_{Y}\left[\ell(0, Y) \mid \boldsymbol{X} = \boldsymbol{x}\right] \ge \mathbb{E}_{Y}\left[\ell(1, Y) \mid \boldsymbol{X} = \boldsymbol{x}\right]$$

Rearranging, the optimal decision is

$$f(\boldsymbol{x}) = \begin{cases} 1 & \text{if } \mathbb{P}(\boldsymbol{Y}=\boldsymbol{1} \mid \boldsymbol{X}=\boldsymbol{x}) \geq \frac{\ell(0,0)-\ell(1,0)}{\ell(1,1)-\ell(0,1)} \mathbb{P}(\boldsymbol{Y}=\boldsymbol{0} \mid \boldsymbol{X}=\boldsymbol{x}) \\ 0 & \text{otherwise} \\ \frac{f_{\boldsymbol{X}\mid\boldsymbol{H}_1}(\boldsymbol{x}\mid\boldsymbol{H}_1) \mathbb{P}(\boldsymbol{H}_1)}{f_{\boldsymbol{X}}(\boldsymbol{x})} \end{bmatrix} \begin{array}{l} \textbf{Bayes rule} \\ \frac{f_{\boldsymbol{X}\mid\boldsymbol{H}_0}(\boldsymbol{x}\mid\boldsymbol{H}_0) \mathbb{P}(\boldsymbol{H}_0)}{f_{\boldsymbol{X}}(\boldsymbol{x})} \end{bmatrix}$$

$$f(\boldsymbol{x}) = 1 \quad \text{if} \quad \begin{cases} f_{\boldsymbol{X}|\boldsymbol{H}_1}(\boldsymbol{x} | \boldsymbol{H}_1) \\ f_{\boldsymbol{X}|\boldsymbol{H}_0}(\boldsymbol{x} | \boldsymbol{H}_0) \\ \\ \\ \mathcal{L}(\boldsymbol{x}) \end{cases} \geq \frac{\ell(0,0) - \ell(1,0)}{\ell(1,1) - \ell(0,1)} \cdot \frac{\mathbb{P}(\boldsymbol{H}_0)}{\mathbb{P}(\boldsymbol{H}_1)} \end{cases}$$

Decision Theory

$$f(\boldsymbol{x}) = 0 \quad \text{if} \quad \mathbb{E}_{Y}\left[\ell(0, Y) \mid \boldsymbol{X} = \boldsymbol{x}\right] < \mathbb{E}_{Y}\left[\ell(1, Y) \mid \boldsymbol{X} = \boldsymbol{x}\right]$$
$$f(\boldsymbol{x}) = 1 \quad \text{if} \quad \mathbb{E}_{Y}\left[\ell(0, Y) \mid \boldsymbol{X} = \boldsymbol{x}\right] \ge \mathbb{E}_{Y}\left[\ell(1, Y) \mid \boldsymbol{X} = \boldsymbol{x}\right]$$

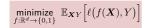
Rearranging, the optimal decision is

$$f(\boldsymbol{x}) = \begin{cases} 1 & \text{if } \mathbb{P}(\boldsymbol{Y}=\boldsymbol{1} \mid \boldsymbol{X}=\boldsymbol{x}) \geq \frac{\ell(0,0)-\ell(1,0)}{\ell(1,1)-\ell(0,1)} \mathbb{P}(\boldsymbol{Y}=\boldsymbol{0} \mid \boldsymbol{X}=\boldsymbol{x}) \\ 0 & \text{otherwise} \\ \frac{f_{\boldsymbol{X}\mid\boldsymbol{H}_{1}}(\boldsymbol{x}\mid\boldsymbol{H}_{1}) \mathbb{P}(\boldsymbol{H}_{1})}{f_{\boldsymbol{X}}(\boldsymbol{x})} \end{bmatrix} \begin{array}{l} \textbf{Bayes rule} \\ \frac{f_{\boldsymbol{X}\mid\boldsymbol{H}_{0}}(\boldsymbol{x}\mid\boldsymbol{H}_{0}) \mathbb{P}(\boldsymbol{H}_{0})}{f_{\boldsymbol{X}}(\boldsymbol{x})} \end{bmatrix}$$

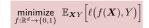
$$\begin{split} f(\boldsymbol{x}) &= 1 \quad \text{ if } \quad \left| \frac{f_{\boldsymbol{X}|\boldsymbol{H}_1}(\boldsymbol{x} \mid \boldsymbol{H}_1)}{f_{\boldsymbol{X}|\boldsymbol{H}_0}(\boldsymbol{x} \mid \boldsymbol{H}_0)} \geq \frac{\ell(0,0) - \ell(1,0)}{\ell(1,1) - \ell(0,1)} \cdot \frac{\mathbb{P}(\boldsymbol{H}_0)}{\mathbb{P}(\boldsymbol{H}_1)} \right| \\ \mathcal{L}(\boldsymbol{x}) : \text{ likelihood ratio} \end{split}$$

Decision Theory

 $\underset{f: \mathbb{R}^d \rightarrow \{0,1\}}{\text{minimize}} \ \mathbb{E}_{\boldsymbol{X}Y} \Big[\ell \big(f(\boldsymbol{X}), Y \big) \Big]$



The decision that minimizes the risk in a binary hypothesis test is



The decision that minimizes the risk in a binary hypothesis test is

$$f(\boldsymbol{x}) = \mathbbm{1}_{\{\mathcal{L}(\boldsymbol{x}) \geq \eta\}}(\boldsymbol{x})$$

The decision that minimizes the risk in a binary hypothesis test is

 $f(\boldsymbol{x}) = \mathbbm{1}_{\{\mathcal{L}(\boldsymbol{x}) \geq \eta\}}(\boldsymbol{x})$

• Indicator function of set
$$S$$
: $\mathbb{1}_{S}(s) = \begin{cases} 1 & , \text{ if } s \in S \\ 0 & , \text{ if } s \notin S \end{cases}$

 $\underset{f:\mathbb{R}^d \rightarrow \{0,1\}}{\text{minimize}} \ \mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}),Y\big)\Big]$

The decision that minimizes the risk in a binary hypothesis test is

 $f(\boldsymbol{x}) = \mathbbm{1}_{\{\mathcal{L}(\boldsymbol{x}) \geq \eta\}}(\boldsymbol{x})$

• Indicator function of set
$$S$$
: $\mathbb{1}_{\mathcal{S}}(s) = \begin{cases} 1 & , \text{ if } s \in S \\ 0 & , \text{ if } s \notin S \end{cases}$

• Likelihood ratio:
$$\mathcal{L}(\boldsymbol{x}) = rac{f_{\boldsymbol{X}|H_1}(\boldsymbol{x} \mid H_1)}{f_{\boldsymbol{X}|H_0}(\boldsymbol{x} \mid H_0)}$$

 $\underset{f: \mathbb{R}^d \rightarrow \{0, 1\}}{\text{minimize}} \ \mathbb{E}_{\boldsymbol{X}Y} \Big[\ell \big(f(\boldsymbol{X}), Y \big) \Big]$

The decision that minimizes the risk in a binary hypothesis test is

$$f(\boldsymbol{x}) = \mathbb{1}_{\{\mathcal{L}(\boldsymbol{x}) \geq \eta\}}(\boldsymbol{x})$$

• Indicator function of set
$$S$$
: $\mathbb{1}_{S}(s) = \begin{cases} 1 & , \text{ if } s \in S \\ 0 & , \text{ if } s \notin S \end{cases}$

• Likelihood ratio:
$$\mathcal{L}(m{x}) = rac{f_{m{X}|H_1}m{x} \mid H_1m{)}}{f_{m{X}|H_0}m{x} \mid H_0m{)}}$$

• Decision threshold:

$$\eta = \frac{\ell(1,0) - \ell(0,0)}{\ell(0,1) - \ell(1,1)} \cdot \frac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)}$$

Decision Theory

 $\min_{f: \mathbb{R}^d \to \{0,1\} } \mathbb{E}_{\boldsymbol{X}Y} \Big[\ell \big(f(\boldsymbol{X}), Y \big) \Big]$

 H_0 : X = W

 $H_1 : X = c + W$

 H_0 : X = W

 $H_1 : X = c + W$

no aircraft/tumor/spam innocent defendant

 H_0 : X = W

no aircraft/tumor/spam innocent defendant

 H_1 : X = c + Waircraft/tumor/spam guilty defendant

 H_0 : X = W

no aircraft/tumor/spam innocent defendant

 $\begin{array}{ll} H_1 \ : \ X = c + W \\ & {\rm aircraft/tumor/spam} \\ & {\rm guilty \ defendant} \end{array} \end{array}$

 $W \sim \mathcal{N}(0, 1)$

 H_0 : X = W

no aircraft/tumor/spam innocent defendant

 H_1 : X = c + Waircraft/tumor/spam guilty defendant

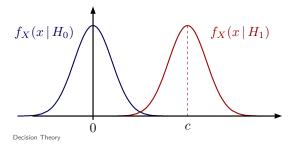
$$W \sim \mathcal{N}(0,1)$$
 $f_W(w) = \frac{1}{\sqrt{2\pi}} e^{-\frac{w^2}{2}}$

 H_0 : X = W

no aircraft/tumor/spam innocent defendant

 $\begin{array}{ll} H_1 \ : \ X = c + W \\ & {\rm aircraft/tumor/spam} \\ & {\rm guilty \ defendant} \end{array} \end{array}$

$$W \sim \mathcal{N}(0,1)$$
 $f_W(w) = \frac{1}{\sqrt{2\pi}} e^{-\frac{w^2}{2}}$



Assume

Assume

• c = 1

Assume

- c = 1
- Loss values

True hypothesis	$f(\boldsymbol{X})=0$	$f(\boldsymbol{X}) = 1$
H_0 is true	0	1
H_1 is true	25	0

Assume

- c = 1
- Loss values

True hypothesis	$f(\boldsymbol{X}) = 0$	$f(\boldsymbol{X}) = 1$
H_0 is true	0	1
H_1 is true	25	0

• Base rates: $\mathbb{P}(H_0) = 0.95$ $\mathbb{P}(H_1) = 0.05$

Assume

- c = 1
- Loss values

True hypothesis	$f(\boldsymbol{X})=0$	$f(\boldsymbol{X}) = 1$
H_0 is true	0	1
H_1 is true	25	0

• Base rates: $\mathbb{P}(H_0) = 0.95$ $\mathbb{P}(H_1) = 0.05$

Compute the decision threshold

Decision threshold occurs for

Decision threshold occurs for

 $\mathcal{L}(x) = \eta$

Decision threshold occurs for

$$\mathcal{L}(x) = \eta \qquad \iff \qquad \log \mathcal{L}(x) = \log \eta$$

Decision threshold occurs for

$$\mathcal{L}(x) = \eta \qquad \iff \qquad \log \mathcal{L}(x) = \log \eta$$

with

 $\mathcal{L}(x)$

Decision threshold occurs for

$$\mathcal{L}(x) = \eta \qquad \iff \qquad \log \mathcal{L}(x) = \log \eta$$

with

$$\mathcal{L}(x) = \frac{f_{X|H_1}(x \mid H_1)}{f_{X|H_0}(x \mid H_0)}$$

Decision threshold occurs for

$$\mathcal{L}(x) = \eta \qquad \iff \qquad \log \mathcal{L}(x) = \log \eta$$

with

$$\mathcal{L}(x) = \frac{f_{X|H_1}(x \mid H_1)}{f_{X|H_0}(x \mid H_0)} = \frac{\exp\left(-\frac{(x-1)^2}{2}\right)}{\exp\left(-\frac{x^2}{2}\right)}$$

Decision threshold occurs for

$$\mathcal{L}(x) = \eta \qquad \iff \qquad \log \mathcal{L}(x) = \log \eta$$

$$\mathcal{L}(x) = \frac{f_{X|H_1}(x \mid H_1)}{f_{X|H_0}(x \mid H_0)} = \frac{\exp\left(-\frac{(x-1)^2}{2}\right)}{\exp\left(-\frac{x^2}{2}\right)} = \exp\left(x - \frac{1}{2}\right)$$

Decision threshold occurs for

$$\mathcal{L}(x) = \eta \qquad \iff \qquad \log \mathcal{L}(x) = \log \eta$$

$$\mathcal{L}(x) = \frac{f_{X|H_1}(x \mid H_1)}{f_{X|H_0}(x \mid H_0)} = \frac{\exp\left(-\frac{(x-1)^2}{2}\right)}{\exp\left(-\frac{x^2}{2}\right)} = \exp\left(x - \frac{1}{2}\right)$$
$$\eta = \frac{\ell(0,0) - \ell(1,0)}{\ell(1,1) - \ell(0,1)} \cdot \frac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)}$$

Decision threshold occurs for

$$\mathcal{L}(x) = \eta \qquad \iff \qquad \log \mathcal{L}(x) = \log \eta$$

$$\mathcal{L}(x) = \frac{f_{X|H_1}(x \mid H_1)}{f_{X|H_0}(x \mid H_0)} = \frac{\exp\left(-\frac{(x-1)^2}{2}\right)}{\exp\left(-\frac{x^2}{2}\right)} = \exp\left(x - \frac{1}{2}\right)$$
$$\eta = \frac{\ell(0,0) - \ell(1,0)}{\ell(1,1) - \ell(0,1)} \cdot \frac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)} = \frac{0-1}{0-25} \cdot \frac{0.95}{0.05}$$

Decision threshold occurs for

$$\mathcal{L}(x) = \eta \qquad \iff \qquad \log \mathcal{L}(x) = \log \eta$$

$$\mathcal{L}(x) = \frac{f_{X|H_1}(x \mid H_1)}{f_{X|H_0}(x \mid H_0)} = \frac{\exp\left(-\frac{(x-1)^2}{2}\right)}{\exp\left(-\frac{x^2}{2}\right)} = \exp\left(x - \frac{1}{2}\right)$$
$$\eta = \frac{\ell(0,0) - \ell(1,0)}{\ell(1,1) - \ell(0,1)} \cdot \frac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)} = \frac{0-1}{0-25} \cdot \frac{0.95}{0.05} \simeq 0.76$$

Decision threshold occurs for

$$\mathcal{L}(x) = \eta \qquad \iff \qquad \log \mathcal{L}(x) = \log \eta$$

with

$$\mathcal{L}(x) = \frac{f_{X|H_1}(x \mid H_1)}{f_{X|H_0}(x \mid H_0)} = \frac{\exp\left(-\frac{(x-1)^2}{2}\right)}{\exp\left(-\frac{x^2}{2}\right)} = \exp\left(x - \frac{1}{2}\right)$$
$$\eta = \frac{\ell(0,0) - \ell(1,0)}{\ell(1,1) - \ell(0,1)} \cdot \frac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)} = \frac{0-1}{0-25} \cdot \frac{0.95}{0.05} \simeq 0.76$$

The decision threshold is then

Decision threshold occurs for

$$\mathcal{L}(x) = \eta \qquad \iff \qquad \log \mathcal{L}(x) = \log \eta$$

with

$$\mathcal{L}(x) = \frac{f_{X|H_1}(x \mid H_1)}{f_{X|H_0}(x \mid H_0)} = \frac{\exp\left(-\frac{(x-1)^2}{2}\right)}{\exp\left(-\frac{x^2}{2}\right)} = \exp\left(x - \frac{1}{2}\right)$$
$$\eta = \frac{\ell(0,0) - \ell(1,0)}{\ell(1,1) - \ell(0,1)} \cdot \frac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)} = \frac{0-1}{0-25} \cdot \frac{0.95}{0.05} \simeq 0.76$$

The decision threshold is then

$$x - \frac{1}{2} = \log 0.76$$

Decision Theory

Decision threshold occurs for

$$\mathcal{L}(x) = \eta \qquad \iff \qquad \log \mathcal{L}(x) = \log \eta$$

with

$$\mathcal{L}(x) = \frac{f_{X|H_1}(x \mid H_1)}{f_{X|H_0}(x \mid H_0)} = \frac{\exp\left(-\frac{(x-1)^2}{2}\right)}{\exp\left(-\frac{x^2}{2}\right)} = \exp\left(x - \frac{1}{2}\right)$$
$$\eta = \frac{\ell(0,0) - \ell(1,0)}{\ell(1,1) - \ell(0,1)} \cdot \frac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)} = \frac{0-1}{0-25} \cdot \frac{0.95}{0.05} \simeq 0.76$$

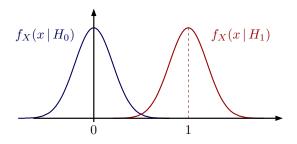
The decision threshold is then

$$x - \frac{1}{2} = \log 0.76 \qquad \Longleftrightarrow \qquad x \simeq 0.23$$

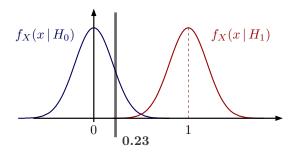
Decision Theory

c = 1		
True hypothesis	$f(\boldsymbol{X})=0$	$f(\boldsymbol{X}) = 1$
H ₀ is true	0	1
H_1 is true	25	0
$\mathbb{P}(H_0) = 0$ $\mathbb{P}(H_1) = 0$		

c = 1		
True hypothesis	$f(\boldsymbol{X}) = 0$	$f(\boldsymbol{X}) = 1$
H_0 is true	0	1
H_1 is true	25	0
$\mathbb{P}(H_0) = 0$ $\mathbb{P}(H_1) = 0$		



c = 1		
True hypothesis	$f(\boldsymbol{X}) = 0$	$f(\boldsymbol{X}) = 1$
H_0 is true	0	1
H_1 is true	25	0
$\mathbb{P}(H_0) = 0$ $\mathbb{P}(H_1) = 0$		



Recall the problem:

$$\underset{f:\mathbb{R}^{d}\to\{0,1\}}{\text{minimize}} \ \mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}),Y\big)\Big]$$

Recall the problem:

$$\min_{\boldsymbol{T}:\mathbb{R}^d \to \{0,1\}} \mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}),Y\big)\Big]$$

Expected value is w.r.t. joint distribution \mathbb{P}_{XY}

Recall the problem:

$$\min_{f:\mathbb{R}^d \to \{0,1\}} \mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}),Y\big)\Big]$$

Expected value is w.r.t. joint distribution \mathbb{P}_{XY}

When class $Y \in \{0,1\}$ is viewed as a parameter of $\mathbb{P}_{\boldsymbol{X}Y}$ to estimate,

Recall the problem:

$$\min_{f:\mathbb{R}^d \to \{0,1\}} \mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}),Y\big)\Big]$$

Expected value is w.r.t. joint distribution \mathbb{P}_{XY}

When class $Y \in \{0,1\}$ is viewed as a parameter of $\mathbb{P}_{\boldsymbol{X}Y}$ to estimate,

- Maximum a posteriori (MAP)
- Maximum likelihood (ML)

can be seen as likelihood ratio tests

Consider $\ell(0, 0) = \ell(1, 1) = 0$ and $\ell(1, 0) = \ell(0, 1) = 1$.

Consider $\ell(0, 0) = \ell(1, 1) = 0$ and $\ell(1, 0) = \ell(0, 1) = 1$. Then,

$$\mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}),Y\big)\Big]$$

Consider $\ell(0, 0) = \ell(1, 1) = 0$ and $\ell(1, 0) = \ell(0, 1) = 1$. Then,

$$\mathbb{E}_{\boldsymbol{X}Y}\left[\ell(f(\boldsymbol{X}),Y)\right] = \int_{\mathbb{R}^d} \mathbb{E}_Y\left[\ell(f(\boldsymbol{X}),Y) \mid \boldsymbol{X} = \boldsymbol{x}\right] f_{\boldsymbol{X}}(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}$$

Consider $\ell(0, 0) = \ell(1, 1) = 0$ and $\ell(1, 0) = \ell(0, 1) = 1$. Then,

$$\mathbb{E}_{\boldsymbol{X}Y}\left[\ell(f(\boldsymbol{X}),Y)\right] = \int_{\mathbb{R}^d} \mathbb{E}_Y\left[\ell(f(\boldsymbol{X}),Y) \mid \boldsymbol{X} = \boldsymbol{x}\right] f_{\boldsymbol{X}}(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}$$

If $f(\boldsymbol{x}) = 0$,

Consider $\ell(0, 0) = \ell(1, 1) = 0$ and $\ell(1, 0) = \ell(0, 1) = 1$. Then,

$$\mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}),Y\big)\Big] = \int_{\mathbb{R}^d} \mathbb{E}_Y\Big[\ell\big(f(\boldsymbol{X}),Y\big) \,\Big|\, \boldsymbol{X} = \boldsymbol{x}\Big] f_{\boldsymbol{X}}(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x}$$

If
$$f(\boldsymbol{x}) = 0$$
, $\mathbb{E}_{Y} \Big[\ell(0, Y) \, \big| \, \boldsymbol{X} = \boldsymbol{x} \Big] = \mathbb{P} \big(Y = 1 \, | \, \boldsymbol{X} = \boldsymbol{x} \big)$

Consider $\ell(0, 0) = \ell(1, 1) = 0$ and $\ell(1, 0) = \ell(0, 1) = 1$. Then,

$$\mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}),Y\big)\Big] = \int_{\mathbb{R}^d} \mathbb{E}_Y\Big[\ell\big(f(\boldsymbol{X}),Y\big) \,\Big|\, \boldsymbol{X} = \boldsymbol{x}\Big] f_{\boldsymbol{X}}(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x}$$

If
$$f(\boldsymbol{x}) = 0$$
, $\mathbb{E}_{Y}\left[\ell(0, Y) \mid \boldsymbol{X} = \boldsymbol{x}\right] = \mathbb{P}(Y = 1 \mid \boldsymbol{X} = \boldsymbol{x})$

If $f(\boldsymbol{x}) = 1$,

Consider $\ell(0, 0) = \ell(1, 1) = 0$ and $\ell(1, 0) = \ell(0, 1) = 1$. Then,

$$\mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}),Y\big)\Big] = \int_{\mathbb{R}^d} \mathbb{E}_Y\Big[\ell\big(f(\boldsymbol{X}),Y\big) \,\Big|\, \boldsymbol{X} = \boldsymbol{x}\Big] f_{\boldsymbol{X}}(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x}$$

If
$$f(\boldsymbol{x}) = 0$$
, $\mathbb{E}_{Y}\left[\ell(0, Y) \mid \boldsymbol{X} = \boldsymbol{x}\right] = \mathbb{P}(Y = 1 \mid \boldsymbol{X} = \boldsymbol{x})$
If $f(\boldsymbol{x}) = 1$, $\mathbb{E}_{Y}\left[\ell(1, Y) \mid \boldsymbol{X} = \boldsymbol{x}\right] = \mathbb{P}(Y = 0 \mid \boldsymbol{X} = \boldsymbol{x})$

Consider $\ell(0, 0) = \ell(1, 1) = 0$ and $\ell(1, 0) = \ell(0, 1) = 1$. Then,

$$\mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}),Y\big)\Big] = \int_{\mathbb{R}^d} \mathbb{E}_Y\Big[\ell\big(f(\boldsymbol{X}),Y\big) \,\Big|\, \boldsymbol{X} = \boldsymbol{x}\Big] f_{\boldsymbol{X}}(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x}$$

If
$$f(\boldsymbol{x}) = 0$$
, $\mathbb{E}_{Y}\left[\ell(0, Y) \mid \boldsymbol{X} = \boldsymbol{x}\right] = \mathbb{P}(Y = 1 \mid \boldsymbol{X} = \boldsymbol{x})$
If $f(\boldsymbol{x}) = 1$, $\mathbb{E}_{Y}\left[\ell(1, Y) \mid \boldsymbol{X} = \boldsymbol{x}\right] = \mathbb{P}(Y = 0 \mid \boldsymbol{X} = \boldsymbol{x})$

So, select $f(\boldsymbol{x}) = 1$ if

Consider $\ell(0, 0) = \ell(1, 1) = 0$ and $\ell(1, 0) = \ell(0, 1) = 1$. Then,

$$\mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}),Y\big)\Big] = \int_{\mathbb{R}^d} \mathbb{E}_Y\Big[\ell\big(f(\boldsymbol{X}),Y\big) \,\Big|\, \boldsymbol{X} = \boldsymbol{x}\Big] f_{\boldsymbol{X}}(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x}$$

If
$$f(\boldsymbol{x}) = 0$$
, $\mathbb{E}_{Y}\left[\ell(0, Y) \mid \boldsymbol{X} = \boldsymbol{x}\right] = \mathbb{P}(Y = 1 \mid \boldsymbol{X} = \boldsymbol{x})$
If $f(\boldsymbol{x}) = 1$, $\mathbb{E}_{Y}\left[\ell(1, Y) \mid \boldsymbol{X} = \boldsymbol{x}\right] = \mathbb{P}(Y = 0 \mid \boldsymbol{X} = \boldsymbol{x})$

So, select $f(\boldsymbol{x}) = 1$ if $\mathbb{P}(Y = 1 \mid \boldsymbol{X} = \boldsymbol{x}) \geq \mathbb{P}(Y = 0 \mid \boldsymbol{X} = \boldsymbol{x})$

So, select
$$f({m x})=1$$
 if $\mathbb{P}ig(Y=1\,|\,{m X}={m x}ig) \geq \mathbb{P}ig(Y=0\,|\,{m X}={m x}ig)$

So, select
$$f(\boldsymbol{x}) = 1$$
 if $\mathbb{P}ig(Y = 1 \,|\, \boldsymbol{X} = \boldsymbol{x}ig) \geq \mathbb{P}ig(Y = 0 \,|\, \boldsymbol{X} = \boldsymbol{x}ig)$

That is,

So, select
$$f(\boldsymbol{x}) = 1$$
 if $\mathbb{P}(Y = 1 \mid \boldsymbol{X} = \boldsymbol{x}) \geq \mathbb{P}(Y = 0 \mid \boldsymbol{X} = \boldsymbol{x})$

That is,
$$f(\boldsymbol{x}) = \operatorname*{arg\,max}_{i} \ \mathbb{P}\big(Y = i \, \big| \, \boldsymbol{X} = \boldsymbol{x} \big)$$

So, select
$$f(\boldsymbol{x}) = 1$$
 if $\mathbb{P}(Y = 1 \mid \boldsymbol{X} = \boldsymbol{x}) \geq \mathbb{P}(Y = 0 \mid \boldsymbol{X} = \boldsymbol{x})$

That is,
$$f(\boldsymbol{x}) = \operatorname*{arg\,max}_i \frac{\mathbb{P}ig(Y=i \mid \boldsymbol{X}=\boldsymbol{x}ig)}{posterior}$$

So, select
$$f(\boldsymbol{x}) = 1$$
 if $\mathbb{P}(Y = 1 \mid \boldsymbol{X} = \boldsymbol{x}) \geq \mathbb{P}(Y = 0 \mid \boldsymbol{X} = \boldsymbol{x})$

That is,
$$f(\boldsymbol{x}) = rg\max_i \frac{\mathbb{P} \big(Y = i \, \big| \, \boldsymbol{X} = \boldsymbol{x} \big)}{posterior}$$

So, select
$$f(\boldsymbol{x}) = 1$$
 if $\mathbb{P}(Y = 1 \mid \boldsymbol{X} = \boldsymbol{x}) \geq \mathbb{P}(Y = 0 \mid \boldsymbol{X} = \boldsymbol{x})$

That is,
$$f(\boldsymbol{x}) = rg\max_i \frac{\mathbb{P}ig(Y=i \mid \boldsymbol{X}=\boldsymbol{x}ig)}{posterior}$$

$$\mathbb{P}(Y=1 \,|\, \boldsymbol{X}=\boldsymbol{x}) \geq \mathbb{P}(Y=0 \,|\, \boldsymbol{X}=\boldsymbol{x})$$

So, select
$$f(\boldsymbol{x}) = 1$$
 if $\mathbb{P}(Y = 1 \mid \boldsymbol{X} = \boldsymbol{x}) \geq \mathbb{P}(Y = 0 \mid \boldsymbol{X} = \boldsymbol{x})$

That is,
$$f(\boldsymbol{x}) = rg\max_i \frac{\mathbb{P}ig(Y=i \mid \boldsymbol{X}=\boldsymbol{x}ig)}{posterior}$$

$$\mathbb{P}(Y = 1 \mid \boldsymbol{X} = \boldsymbol{x}) \geq \mathbb{P}(Y = 0 \mid \boldsymbol{X} = \boldsymbol{x})$$
$$\iff f_{\boldsymbol{X}\mid\boldsymbol{H}_{1}}(x \mid \boldsymbol{H}_{1}) \cdot \mathbb{P}(\boldsymbol{H}_{1}) \geq f_{\boldsymbol{X}\mid\boldsymbol{H}_{0}}(x \mid \boldsymbol{H}_{0}) \cdot \mathbb{P}(\boldsymbol{H}_{0})$$

So, select
$$f(\boldsymbol{x}) = 1$$
 if $\mathbb{P}(Y = 1 \mid \boldsymbol{X} = \boldsymbol{x}) \geq \mathbb{P}(Y = 0 \mid \boldsymbol{X} = \boldsymbol{x})$

That is,
$$f(\boldsymbol{x}) = rg\max_i \frac{\mathbb{P}ig(Y=i \mid \boldsymbol{X}=\boldsymbol{x}ig)}{posterior}$$

$$\mathbb{P}(Y = 1 | \boldsymbol{X} = \boldsymbol{x}) \geq \mathbb{P}(Y = 0 | \boldsymbol{X} = \boldsymbol{x})$$

$$\iff \quad f_{\boldsymbol{X}|H_1}(x | H_1) \cdot \mathbb{P}(H_1) \geq f_{\boldsymbol{X}|H_0}(x | H_0) \cdot \mathbb{P}(H_0)$$

$$\iff \quad \frac{f_{\boldsymbol{X}|H_1}(x | H_1)}{f_{\boldsymbol{X}|H_0}(x | H_0)} \geq \frac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)}$$

So, select
$$f(\boldsymbol{x}) = 1$$
 if $\mathbb{P}(Y = 1 \mid \boldsymbol{X} = \boldsymbol{x}) \geq \mathbb{P}(Y = 0 \mid \boldsymbol{X} = \boldsymbol{x})$

That is,
$$f(\boldsymbol{x}) = rg\max_i \frac{\mathbb{P}ig(Y=i \mid \boldsymbol{X}=\boldsymbol{x}ig)}{posterior}$$

$$\mathbb{P}(Y = 1 | \boldsymbol{X} = \boldsymbol{x}) \geq \mathbb{P}(Y = 0 | \boldsymbol{X} = \boldsymbol{x})$$

$$\iff f_{\boldsymbol{X}|H_1}(x | H_1) \cdot \mathbb{P}(H_1) \geq f_{\boldsymbol{X}|H_0}(x | H_0) \cdot \mathbb{P}(H_0)$$

$$\iff \mathcal{L}(\boldsymbol{x}) = \frac{f_{\boldsymbol{X}|H_1}(x | H_1)}{f_{\boldsymbol{X}|H_0}(x | H_0)} \geq \frac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)}$$

So, select
$$f(\boldsymbol{x}) = 1$$
 if $\mathbb{P}(Y = 1 \mid \boldsymbol{X} = \boldsymbol{x}) \geq \mathbb{P}(Y = 0 \mid \boldsymbol{X} = \boldsymbol{x})$

That is,
$$f(\boldsymbol{x}) = rg\max_i \frac{\mathbb{P}ig(Y=i \mid \boldsymbol{X}=\boldsymbol{x}ig)}{posterior}$$

$$\mathbb{P}(Y = 1 | \boldsymbol{X} = \boldsymbol{x}) \geq \mathbb{P}(Y = 0 | \boldsymbol{X} = \boldsymbol{x})$$

$$\iff f_{\boldsymbol{X}|H_1}(x | H_1) \cdot \mathbb{P}(H_1) \geq f_{\boldsymbol{X}|H_0}(x | H_0) \cdot \mathbb{P}(H_0)$$

$$\iff \mathcal{L}(\boldsymbol{x}) = \frac{f_{\boldsymbol{X}|H_1}(x | H_1)}{f_{\boldsymbol{X}|H_0}(x | H_0)} \geq \frac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)} = \eta$$

So, select
$$f(\boldsymbol{x}) = 1$$
 if $\mathbb{P} ig(Y = 1 \,|\, \boldsymbol{X} = \boldsymbol{x} ig) \, \geq \, \mathbb{P} ig(Y = 0 \,|\, \boldsymbol{X} = \boldsymbol{x} ig)$

That is,
$$f(\boldsymbol{x}) = rg\max_i \frac{\mathbb{P}ig(Y=i \mid \boldsymbol{X}=\boldsymbol{x}ig)}{posterior}$$

$$\mathbb{P}(Y = 1 | \boldsymbol{X} = \boldsymbol{x}) \geq \mathbb{P}(Y = 0 | \boldsymbol{X} = \boldsymbol{x})$$

$$\iff f_{\boldsymbol{X}|H_1}(x | H_1) \cdot \mathbb{P}(H_1) \geq f_{\boldsymbol{X}|H_0}(x | H_0) \cdot \mathbb{P}(H_0)$$

$$\iff \mathcal{L}(\boldsymbol{x}) = \frac{f_{\boldsymbol{X}|H_1}(x | H_1)}{f_{\boldsymbol{X}|H_0}(x | H_0)} \geq \frac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)} = \eta$$

Recall that MAP rule minimizes probability of incorrect decision:

So, select
$$f(\boldsymbol{x}) = 1$$
 if $\mathbb{P} ig(Y = 1 \,|\, \boldsymbol{X} = \boldsymbol{x} ig) \, \geq \, \mathbb{P} ig(Y = 0 \,|\, \boldsymbol{X} = \boldsymbol{x} ig)$

That is,
$$f(\boldsymbol{x}) = rg\max_i \frac{\mathbb{P}ig(Y=i \mid \boldsymbol{X}=\boldsymbol{x}ig)}{posterior}$$

This is a likelihood ratio test, because

$$\mathbb{P}(Y = 1 | \boldsymbol{X} = \boldsymbol{x}) \geq \mathbb{P}(Y = 0 | \boldsymbol{X} = \boldsymbol{x})$$

$$\iff f_{\boldsymbol{X}|H_1}(x | H_1) \cdot \mathbb{P}(H_1) \geq f_{\boldsymbol{X}|H_0}(x | H_0) \cdot \mathbb{P}(H_0)$$

$$\iff \mathcal{L}(\boldsymbol{x}) = \frac{f_{\boldsymbol{X}|H_1}(x | H_1)}{f_{\boldsymbol{X}|H_0}(x | H_0)} \geq \frac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)} = \eta$$

Recall that MAP rule minimizes probability of incorrect decision: $\mathbb{P}(\text{error}) = \mathbb{P}(f(\mathbf{X}) = 1, H_0) + \mathbb{P}(f(\mathbf{X}) = 0, H_1)$

Decision Theory

Consider $\ell(0, 0) = \ell(1, 1) = 0$ and $\ell(1, 0) = \ell(0, 1) = 1$

Consider $\ell(0, 0) = \ell(1, 1) = 0$ and $\ell(1, 0) = \ell(0, 1) = 1$

And $\mathbb{P}(H_0) = \mathbb{P}(H_1) = \frac{1}{2}$

Consider $\ell(0, 0) = \ell(1, 1) = 0$ and $\ell(1, 0) = \ell(0, 1) = 1$

And $\mathbb{P}(H_0) = \mathbb{P}(H_1) = \frac{1}{2}$

Consider $\ell(0, 0) = \ell(1, 1) = 0$ and $\ell(1, 0) = \ell(0, 1) = 1$ And $\mathbb{P}(H_0) = \mathbb{P}(H_1) = \frac{1}{2}$

$$f(\boldsymbol{x}) = \underset{i}{\operatorname{arg\,max}} \mathbb{P}(Y = i \mid \boldsymbol{X} = \boldsymbol{x})$$

Consider $\ell(0, 0) = \ell(1, 1) = 0$ and $\ell(1, 0) = \ell(0, 1) = 1$ And $\mathbb{P}(H_0) = \mathbb{P}(H_1) = \frac{1}{2}$

$$\begin{split} f(\boldsymbol{x}) &= \arg \max_{i} \quad \mathbb{P} \big(Y = i \, \big| \, \boldsymbol{X} = \boldsymbol{x} \big) \\ &= \arg \max_{i} \quad \frac{f_{\boldsymbol{X}|Y}(\boldsymbol{x} \,| \, Y = i) \cdot \mathbb{P}(Y = i)}{f_{\boldsymbol{X}}(\boldsymbol{x})} \quad (\text{Bayes rule}) \end{split}$$

Consider $\ell(0, 0) = \ell(1, 1) = 0$ and $\ell(1, 0) = \ell(0, 1) = 1$ And $\mathbb{P}(H_0) = \mathbb{P}(H_1) = \frac{1}{2}$

$$\begin{split} f(\boldsymbol{x}) &= \arg \max_{i} \quad \mathbb{P}\left(Y = i \mid \boldsymbol{X} = \boldsymbol{x}\right) \\ &= \arg \max_{i} \quad \frac{f_{\boldsymbol{X}|Y}(\boldsymbol{x} \mid Y = i) \cdot \mathbb{P}(Y = i)}{f_{\boldsymbol{X}}(\boldsymbol{x})} \qquad \qquad \left(\text{Bayes rule} \right) \\ &= \arg \max_{i} \quad f_{\boldsymbol{X}|Y}\left(\boldsymbol{x} \mid Y = i\right) \qquad \qquad \left(\mathbb{P}\left(Y = i\right) = \frac{1}{2} \right) \end{split}$$

Consider $\ell(0, 0) = \ell(1, 1) = 0$ and $\ell(1, 0) = \ell(0, 1) = 1$ And $\mathbb{P}(H_0) = \mathbb{P}(H_1) = \frac{1}{2}$

$$\begin{split} f(\boldsymbol{x}) &= \arg \max_{i} \quad \mathbb{P}\left(Y = i \mid \boldsymbol{X} = \boldsymbol{x}\right) \\ &= \arg \max_{i} \quad \frac{f_{\boldsymbol{X}|Y}(\boldsymbol{x} \mid Y = i) \cdot \mathbb{P}(Y = i)}{f_{\boldsymbol{X}}(\boldsymbol{x})} \qquad \qquad \left(\text{Bayes rule} \right) \\ &= \arg \max_{i} \quad f_{\boldsymbol{X}|Y}\left(\boldsymbol{x} \mid Y = i\right) \qquad \qquad \left(\mathbb{P}\left(Y = i\right) = \frac{1}{2} \right) \\ &= \max_{i} \min \text{ likelihood} \end{split}$$

Consider $\ell(0, 0) = \ell(1, 1) = 0$ and $\ell(1, 0) = \ell(0, 1) = 1$ And $\mathbb{P}(H_0) = \mathbb{P}(H_1) = \frac{1}{2}$

The optimal decision (MAP) is

$$\begin{split} f(\boldsymbol{x}) &= \arg \max_{i} \quad \mathbb{P}\left(Y = i \mid \boldsymbol{X} = \boldsymbol{x}\right) \\ &= \arg \max_{i} \quad \frac{f_{\boldsymbol{X}|Y}(\boldsymbol{x} \mid Y = i) \cdot \mathbb{P}(Y = i)}{f_{\boldsymbol{X}}(\boldsymbol{x})} \qquad \qquad \left(\text{Bayes rule} \right) \\ &= \arg \max_{i} \quad f_{\boldsymbol{X}|Y}\left(\boldsymbol{x} \mid Y = i\right) \qquad \qquad \left(\mathbb{P}\left(Y = i\right) = \frac{1}{2} \right) \\ &= \max \text{maximum likelihood} \end{split}$$

This corresponds to a likelihood ratio test with $\eta=1$

Decision Theory

	Table of probabilities	
True hypothesis	$f(\boldsymbol{X}) = 0$	$f(\boldsymbol{X}) = 1$
H_0 is true		
H_1 is true		

True Positive Rate (TPR)

	Table	of probabilities
True hypothesis	$f(\boldsymbol{X}) = 0$	$f(\boldsymbol{X}) = 1$
H_0 is true		
H_1 is true		

True Positive Rate (TPR)

power, sensitivity, recall

	Table	of probabilities
True hypothesis	$f(\boldsymbol{X}) = 0$	$f(\boldsymbol{X}) = 1$
H_0 is true		
H_1 is true		

True Positive Rate (TPR)

power, sensitivity, recall

	Table of probabilities	
True hypothesis	$f(\boldsymbol{X}) = 0$	$f(\boldsymbol{X}) = 1$
H_0 is true		
H_1 is true		TPR

True Positive Rate (TPR)

$$\mathsf{TPR} = \mathbb{P}\big(f(\boldsymbol{X}) = 1 \,\big|\, \boldsymbol{H}_1\big)$$

power, sensitivity, recall

	Table	of probabilities
True hypothesis	$f(\boldsymbol{X}) = 0$	$f(\boldsymbol{X}) = 1$
H_0 is true		
H_1 is true		TPR

True Positive Rate (TPR)

$$\mathsf{TPR} = \mathbb{P}\big(f(\boldsymbol{X}) = 1 \,\big|\, \boldsymbol{H}_1\big)$$

power, sensitivity, recall

False Positive Rate (FPR)

type I error, false alarm

	Table of probabilities	
True hypothesis	$f(\boldsymbol{X}) = 0$	$f(\boldsymbol{X}) = 1$
H_0 is true		
H_1 is true		TPR

True Positive Rate (TPR)

$$\mathsf{TPR} = \mathbb{P}\big(f(\boldsymbol{X}) = 1 \,\big|\, \boldsymbol{H}_1\big)$$

power, sensitivity, recall

False Positive Rate (FPR)

type I error, false alarm

	Table of probabilities	
True hypothesis	$f(\boldsymbol{X}) = 0$	$f(\boldsymbol{X}) = 1$
H_0 is true		FPR
H_1 is true		TPR

True Positive Rate (TPR)

power, sensitivity, recall

False Positive Rate (FPR)

type I error, false alarm

$$\mathsf{TPR} = \mathbb{P}(f(\mathbf{X}) = 1 \mid \mathbf{H}_1)$$

$$\mathsf{FPR} = \mathbb{P}(f(\boldsymbol{X}) = 1 \mid H_0)$$

	Table of probabilities	
True hypothesis	$f(\boldsymbol{X})=0$	$f(\boldsymbol{X}) = 1$
H_0 is true		FPR
H_1 is true		TPR

True Positive Rate (TPR)

power, sensitivity, recall

False Positive Rate (FPR)

type I error, false alarm

True Negative Rate (TNR) *specificity*

$$\mathsf{TPR} = \mathbb{P}\big(f(\boldsymbol{X}) = 1 \,\big|\, \boldsymbol{H}_1\big)$$

$$\mathsf{FPR} = \mathbb{P}(f(\boldsymbol{X}) = 1 \mid H_0)$$

	Table of probabilities	
True hypothesis	$f(\boldsymbol{X}) = 0$	$f(\boldsymbol{X}) = 1$
H_0 is true		FPR
H_1 is true		TPR

True Positive Rate (TPR)

power, sensitivity, recall

False Positive Rate (FPR)

type I error, false alarm

True Negative Rate (TNR) *specificity*

$$\mathsf{TPR} = \mathbb{P}\big(f(\boldsymbol{X}) = 1 \,\big|\, \boldsymbol{H}_1\big)$$

$$\mathsf{FPR} = \mathbb{P}\big(f(\boldsymbol{X}) = 1 \,\big|\, H_0\big)$$

	Table of probabilities	
True hypothesis	$f(\boldsymbol{X}) = 0$	$f(\boldsymbol{X}) = 1$
H_0 is true	TNR	FPR
H_1 is true		TPR

True Positive Rate (TPR)

power, sensitivity, recall

False Positive Rate (FPR) type I error, false alarm

True Negative Rate (TNR) *specificity*

$$\mathsf{TPR} = \mathbb{P}(f(\mathbf{X}) = 1 \mid \mathbf{H}_1)$$

$$\mathsf{FPR} = \mathbb{P}\big(f(\boldsymbol{X}) = 1 \,\big|\, H_0\big)$$

$$\mathsf{TNR} = \mathbb{P}\big(f(\boldsymbol{X}) = 0 \,\big|\, H_0\big)$$

	Table of probabilities	
True hypothesis	$f(\boldsymbol{X})=0$	$f(\boldsymbol{X}) = 1$
H_0 is true	TNR	FPR
H_1 is true		TPR

True Positive Rate (TPR)

power, sensitivity, recall

False Positive Rate (FPR) *type I error, false alarm*

True Negative Rate (TNR) *specificity*

False Negative Rate (FNR)

	Table of probabilities	
True hypothesis	$f(\boldsymbol{X})=0$	$f(\boldsymbol{X}) = 1$
H_0 is true	TNR	FPR
H_1 is true		TPR

$$\mathsf{TPR} = \mathbb{P}(f(\mathbf{X}) = 1 \mid \mathbf{H}_1)$$

$$\mathsf{FPR} = \mathbb{P}(f(\boldsymbol{X}) = 1 \mid H_0)$$

$$\mathsf{TNR} = \mathbb{P}\big(f(\boldsymbol{X}) = 0 \,\big|\, H_0\big)$$

True Positive Rate (TPR)

power, sensitivity, recall

False Positive Rate (FPR) *type I error, false alarm*

True Negative Rate (TNR) *specificity*

False Negative Rate (FNR)

	Table of probabilities	
True hypothesis	$f(\boldsymbol{X}) = 0$	$f(\boldsymbol{X}) = 1$
H_0 is true	TNR	FPR
H_1 is true	FNR	TPR

$$\mathsf{TPR} = \mathbb{P}(f(\mathbf{X}) = 1 \mid \mathbf{H}_1)$$

$$\mathsf{FPR} = \mathbb{P}(f(\boldsymbol{X}) = 1 \mid H_0)$$

$$\mathsf{TNR} = \mathbb{P}(f(\boldsymbol{X}) = 0 \,|\, H_0)$$

True Positive Rate (TPR) *power, sensitivity, recall*

False Positive Rate (FPR) *type I error, false alarm*

True Negative Rate (TNR) *specificity*

False Negative Rate (FNR)

$$\mathsf{TPR} = \mathbb{P}\big(f(\boldsymbol{X}) = 1 \,\big|\, \boldsymbol{H}_1\big)$$

$$\mathsf{FPR} = \mathbb{P}(f(\boldsymbol{X}) = 1 \mid H_0)$$

$$\mathsf{TNR} = \mathbb{P}\big(f(\boldsymbol{X}) = 0 \,\big|\, H_0\big)$$

$$\mathsf{FNR} = \mathbb{P}(f(\boldsymbol{X}) = 0 \,|\, \boldsymbol{H}_1)$$

	Table of probabilities	
True hypothesis	$f(\boldsymbol{X}) = 0$	$f(\boldsymbol{X}) = 1$
H_0 is true	TNR	FPR
H_1 is true	FNR	TPR

True Positive Rate (TPR) *power, sensitivity, recall*

False Positive Rate (FPR) *type I error, false alarm*

True Negative Rate (TNR) *specificity*

False Negative Rate (FNR)

$$\mathsf{TPR} = \mathbb{P}\big(f(\boldsymbol{X}) = 1 \,\big|\, \boldsymbol{H}_1\big)$$

$$\mathsf{FPR} = \mathbb{P}(f(\boldsymbol{X}) = 1 \mid H_0)$$

$$\mathsf{TNR} = \mathbb{P}\big(f(\boldsymbol{X}) = 0 \,\big|\, H_0\big)$$

$$\mathsf{FNR} = \mathbb{P}(f(\boldsymbol{X}) = 0 \,|\, \boldsymbol{H}_1)$$

	Table of probabilities	
True hypothesis	$f(\boldsymbol{X})=0$	$f(\boldsymbol{X}) = 1$
H_0 is true	TNR	$^{lpha}_{ m FPR}$
H_1 is true	FNR	TPR

True Positive Rate (TPR) *power, sensitivity, recall*

False Positive Rate (FPR) *type I error, false alarm*

True Negative Rate (TNR) *specificity*

False Negative Rate (FNR)

$$\mathsf{TPR} = \mathbb{P}(f(\boldsymbol{X}) = 1 \mid \boldsymbol{H}_1)$$

$$\mathsf{FPR} = \mathbb{P}(f(\boldsymbol{X}) = 1 \mid H_0)$$

$$\mathsf{TNR} = \mathbb{P}\big(f(\boldsymbol{X}) = 0 \,\big|\, H_0\big)$$

$$\mathsf{FNR} = \mathbb{P}(f(\boldsymbol{X}) = 0 \,\big|\, \boldsymbol{H}_1)$$

	Table of probabilities	
True hypothesis	$f(\pmb{X})=0$	$f(\boldsymbol{X}) = 1$
H_0 is true	TNR	$^{lpha}_{FPR}$
H_1 is true	ع FNR	TPR

Previous measures don't account for base rates: $\mathbb{P}(H_0)$, $\mathbb{P}(H_1)$

Previous measures don't account for base rates: $\mathbb{P}(H_0)$, $\mathbb{P}(H_1)$

That is, they are suitable when $\mathbb{P}(H_0) \simeq \mathbb{P}(H_1)$

Previous measures don't account for base rates: $\mathbb{P}(H_0)$, $\mathbb{P}(H_1)$

That is, they are suitable when $\mathbb{P}(H_0) \simeq \mathbb{P}(H_1)$

Alternatives

Previous measures don't account for base rates: $\mathbb{P}(H_0)$, $\mathbb{P}(H_1)$

That is, they are suitable when $\mathbb{P}(H_0) \simeq \mathbb{P}(H_1)$

<u>Alternatives</u>

Precision: $\mathbb{P}(H_1 \mid f(X) = 1)$

Previous measures don't account for base rates: $\mathbb{P}(H_0)$, $\mathbb{P}(H_1)$

That is, they are suitable when $\mathbb{P}(H_0) \simeq \mathbb{P}(H_1)$

<u>Alternatives</u>

Precision: $\mathbb{P}(H_1 \mid f(X) = 1) = \frac{\mathsf{TPR} \cdot \mathbb{P}(H_1)}{\mathsf{TPR} \cdot \mathbb{P}(H_1) + \mathsf{FPR} \cdot \mathbb{P}(H_0)}$

Previous measures don't account for base rates: $\mathbb{P}(H_0)$, $\mathbb{P}(H_1)$

That is, they are suitable when $\mathbb{P}(H_0) \simeq \mathbb{P}(H_1)$

Alternatives

Precision: $\mathbb{P}(H_1 \mid f(X) = 1) = \frac{\mathsf{TPR} \cdot \mathbb{P}(H_1)}{\mathsf{TPR} \cdot \mathbb{P}(H_1) + \mathsf{FPR} \cdot \mathbb{P}(H_0)}$

 F_1 -score: harmonic mean between precision and recall (TPR):

Previous measures don't account for base rates: $\mathbb{P}(H_0)$, $\mathbb{P}(H_1)$

That is, they are suitable when $\mathbb{P}(H_0) \simeq \mathbb{P}(H_1)$

Alternatives

Precision:
$$\mathbb{P}(H_1 \mid f(X) = 1) = \frac{\mathsf{TPR} \cdot \mathbb{P}(H_1)}{\mathsf{TPR} \cdot \mathbb{P}(H_1) + \mathsf{FPR} \cdot \mathbb{P}(H_0)}$$

 F_1 -score: harmonic mean between precision and recall (TPR):

$$F_1 = \frac{2}{\frac{1}{\mathbb{P}(H_1 \mid f(\mathbf{X}) = 1)} + \frac{1}{\mathsf{TPR}}}$$

	Table of probabilities	
True hypothesis	$f(\boldsymbol{X}) = 0$	$f(\boldsymbol{X}) = 1$
H_0 is true	TNR	lpha FPR
H_1 is true	β FNR	TPR

	Table of probabilities	
True hypothesis	$f(\pmb{X})=0$	$f(\boldsymbol{X}) = 1$
H_0 is true	TNR	lpha FPR
H_1 is true	β FNR	TPR

 α and β are in *conflict*:

	Table of probabilities	
True hypothesis	$f(\boldsymbol{X})=0$	$f(\boldsymbol{X}) = 1$
H_0 is true	TNR	lpha FPR
H_1 is true	β FNR	TPR

 α and β are in conflict:

•
$$\alpha \downarrow \implies \beta \uparrow$$
 : and vice-versa

	Table of probabilities	
True hypothesis	$f(\boldsymbol{X})=0$	$f(\boldsymbol{X}) = 1$
H_0 is true	TNR	lpha FPR
H_1 is true	β FNR	TPR

 α and β are in *conflict*:

- $\alpha \downarrow \implies \beta \uparrow$: and vice-versa
- Both can decrease only by observing more data

	Table of probabilities	
True hypothesis	$f(\boldsymbol{X})=0$	$f(\boldsymbol{X}) = 1$
H_0 is true	TNR	lpha FPR
H_1 is true	β FNR	TPR

 α and β are in *conflict*:

- $\alpha \downarrow \implies \beta \uparrow$: and vice-versa
- Both can decrease only by observing more data

It turns out that likelihood ratio tests are Pareto optimal

Theorem

Let f(x) be a decision rule (det. or prob.)

Theorem

Let f(x) be a decision rule (det. or prob.) with FPR and FNR

$$\alpha = \mathbb{P}\Big(f(\boldsymbol{x}) = 1 \,\big|\, H_0\Big) \qquad \qquad \beta = \mathbb{P}\Big(f(\boldsymbol{x}) = 0 \,\big|\, \boldsymbol{H}_1\Big)$$

Theorem

Let $f(\boldsymbol{x})$ be a decision rule (det. or prob.) with FPR and FNR

$$\alpha = \mathbb{P}\Big(f(\boldsymbol{x}) = 1 \mid H_0\Big) \qquad \qquad \beta = \mathbb{P}\Big(f(\boldsymbol{x}) = 0 \mid \boldsymbol{H}_1\Big)$$

Then, there exists an LRT decision rule $f_{\mathsf{LRT}}({m{x}})$ with

Theorem

Let f(x) be a decision rule (det. or prob.) with FPR and FNR

$$lpha = \mathbb{P}\Big(f(\boldsymbol{x}) = 1 \mid H_0\Big) \qquad \qquad \beta = \mathbb{P}\Big(f(\boldsymbol{x}) = 0 \mid H_1\Big)$$

Then, there exists an LRT decision rule $\, f_{\mathsf{LRT}}({m x}) \,$ with

$$\alpha_{\mathsf{LRT}} = \mathbb{P}\Big(f_{\mathsf{LRT}}(\boldsymbol{x}) = 1 \,\big|\, H_0\Big) \qquad \beta_{\mathsf{LRT}} = \mathbb{P}\Big(f_{\mathsf{LRT}}(\boldsymbol{x}) = 0 \,\big|\, \boldsymbol{H}_1\Big)$$

Theorem

Let f(x) be a decision rule (det. or prob.) with FPR and FNR

$$lpha = \mathbb{P}\Big(f(\boldsymbol{x}) = 1 \mid H_0\Big) \qquad \qquad \beta = \mathbb{P}\Big(f(\boldsymbol{x}) = 0 \mid H_1\Big)$$

Then, there exists an LRT decision rule $\, f_{\mathsf{LRT}}({m x}) \,$ with

$$\alpha_{\mathsf{LRT}} = \mathbb{P}\Big(f_{\mathsf{LRT}}(\boldsymbol{x}) = 1 \, \big| \, H_0\Big) \qquad \beta_{\mathsf{LRT}} = \mathbb{P}\Big(f_{\mathsf{LRT}}(\boldsymbol{x}) = 0 \, \big| \, \boldsymbol{H}_1\Big)$$

such that

Theorem

Let f(x) be a decision rule (det. or prob.) with FPR and FNR

$$lpha = \mathbb{P}\Big(f(\boldsymbol{x}) = 1 \mid H_0\Big) \qquad \qquad \beta = \mathbb{P}\Big(f(\boldsymbol{x}) = 0 \mid H_1\Big)$$

Then, there exists an LRT decision rule $\, f_{\mathsf{LRT}}({m x}) \,$ with

$$\alpha_{\mathsf{LRT}} = \mathbb{P}\Big(f_{\mathsf{LRT}}(\boldsymbol{x}) = 1 \, \big| \, H_0\Big) \qquad \beta_{\mathsf{LRT}} = \mathbb{P}\Big(f_{\mathsf{LRT}}(\boldsymbol{x}) = 0 \, \big| \, \boldsymbol{H}_1\Big)$$

such that

$$\alpha \leq \alpha_{\mathsf{LRT}} \qquad \Longrightarrow \qquad \beta \geq \beta_{\mathsf{LRT}}$$

Theorem

Let f(x) be a decision rule (det. or prob.) with FPR and FNR

$$lpha = \mathbb{P}\Big(f(\boldsymbol{x}) = 1 \mid H_0\Big) \qquad \qquad \beta = \mathbb{P}\Big(f(\boldsymbol{x}) = 0 \mid H_1\Big)$$

Then, there exists an LRT decision rule $\, f_{\mathsf{LRT}}({m x}) \,$ with

$$\alpha_{\mathsf{LRT}} = \mathbb{P}\Big(f_{\mathsf{LRT}}(\boldsymbol{x}) = 1 \, \big| \, H_0\Big) \qquad \beta_{\mathsf{LRT}} = \mathbb{P}\Big(f_{\mathsf{LRT}}(\boldsymbol{x}) = 0 \, \big| \, \boldsymbol{H}_1\Big)$$

such that

$\alpha \leq \alpha_{LRT}$	\implies	$\beta \geq \beta_{\rm LRT}$
$\beta\leq\beta_{\rm LRT}$	\implies	$\alpha \geq \alpha_{LRT}$

Theorem

Let $f(\boldsymbol{x})$ be a decision rule (det. or prob.) with FPR and FNR

$$lpha = \mathbb{P}\Big(f(\boldsymbol{x}) = 1 \mid H_0\Big) \qquad \qquad \beta = \mathbb{P}\Big(f(\boldsymbol{x}) = 0 \mid H_1\Big)$$

Then, there exists an LRT decision rule $\,f_{
m LRT}({m x})\,$ with

$$\alpha_{\mathsf{LRT}} = \mathbb{P}\Big(f_{\mathsf{LRT}}(\boldsymbol{x}) = 1 \, \big| \, H_0\Big) \qquad \beta_{\mathsf{LRT}} = \mathbb{P}\Big(f_{\mathsf{LRT}}(\boldsymbol{x}) = 0 \, \big| \, \boldsymbol{H}_1\Big)$$

such that

$\alpha \leq \alpha_{LRT}$	\implies	$\beta\geq\beta_{\rm LRT}$
$\beta\leq\beta_{\rm LRT}$	\implies	$\alpha \geq \alpha_{\rm LRT}$

And the same relations hold with strict inequalities (<,>)

Decision Theory

$$\mathbb{P}\left(\mathsf{error}_{\mathsf{MAP}}\right)$$
$$= \mathbb{P}\left(f_{\mathsf{MAP}}(\boldsymbol{X}) = 1, H_0\right) + \mathbb{P}\left(f_{\mathsf{MAP}}(\boldsymbol{X}) = 0, H_1\right)$$

$$\mathbb{P}\left(\operatorname{error}_{\mathsf{MAP}}\right)$$

$$= \mathbb{P}\left(f_{\mathsf{MAP}}(\boldsymbol{X}) = 1, H_0\right) + \mathbb{P}\left(f_{\mathsf{MAP}}(\boldsymbol{X}) = 0, H_1\right)$$

$$= \mathbb{P}\left(f_{\mathsf{MAP}}(\boldsymbol{X}) = 1 \mid H_0\right) \mathbb{P}(H_0) + \mathbb{P}\left(f_{\mathsf{MAP}}(\boldsymbol{X}) = 0 \mid H_1\right) \mathbb{P}(H_1)$$

$$\mathbb{P}\left(\operatorname{error}_{\mathsf{MAP}}\right)$$

$$= \mathbb{P}\left(f_{\mathsf{MAP}}(\boldsymbol{X}) = 1, H_0\right) + \mathbb{P}\left(f_{\mathsf{MAP}}(\boldsymbol{X}) = 0, H_1\right)$$

$$= \underbrace{\mathbb{P}\left(f_{\mathsf{MAP}}(\boldsymbol{X}) = 1 \mid H_0\right)}_{\alpha_{\mathsf{MAP}}} \mathbb{P}(H_0) + \mathbb{P}\left(f_{\mathsf{MAP}}(\boldsymbol{X}) = 0 \mid H_1\right) \mathbb{P}(H_1)$$

$$\mathbb{P}\left(\operatorname{error}_{\mathsf{MAP}}\right)$$

$$= \mathbb{P}\left(f_{\mathsf{MAP}}(\boldsymbol{X}) = 1, H_0\right) + \mathbb{P}\left(f_{\mathsf{MAP}}(\boldsymbol{X}) = 0, H_1\right)$$

$$= \underbrace{\mathbb{P}\left(f_{\mathsf{MAP}}(\boldsymbol{X}) = 1 \mid H_0\right)}_{\alpha_{\mathsf{MAP}}} \mathbb{P}(H_0) + \underbrace{\mathbb{P}\left(f_{\mathsf{MAP}}(\boldsymbol{X}) = 0 \mid H_1\right)}_{\beta_{\mathsf{MAP}}} \mathbb{P}(H_1)$$

• $f_{MAP}(x)$ minimizes probability of error over all rules (det. and prob.):

$$\mathbb{P}\left(\operatorname{error}_{\mathsf{MAP}}\right)$$

$$= \mathbb{P}\left(f_{\mathsf{MAP}}(\boldsymbol{X}) = 1, H_{0}\right) + \mathbb{P}\left(f_{\mathsf{MAP}}(\boldsymbol{X}) = 0, H_{1}\right)$$

$$= \underbrace{\mathbb{P}\left(f_{\mathsf{MAP}}(\boldsymbol{X}) = 1 \mid H_{0}\right)}_{\alpha_{\mathsf{MAP}}} \mathbb{P}(H_{0}) + \underbrace{\mathbb{P}\left(f_{\mathsf{MAP}}(\boldsymbol{X}) = 0 \mid H_{1}\right)}_{\beta_{\mathsf{MAP}}} \mathbb{P}(H_{1})$$

 $= \alpha_{\mathsf{MAP}} \, \mathbb{P}(H_0) + \beta_{\mathsf{MAP}} \, \mathbb{P}(H_1)$

$$\mathbb{P}\left(\operatorname{error}_{\mathsf{MAP}}\right)$$

$$= \mathbb{P}\left(f_{\mathsf{MAP}}(\boldsymbol{X}) = 1, H_{0}\right) + \mathbb{P}\left(f_{\mathsf{MAP}}(\boldsymbol{X}) = 0, H_{1}\right)$$

$$= \underbrace{\mathbb{P}\left(f_{\mathsf{MAP}}(\boldsymbol{X}) = 1 \mid H_{0}\right)}_{\alpha_{\mathsf{MAP}}} \mathbb{P}(H_{0}) + \underbrace{\mathbb{P}\left(f_{\mathsf{MAP}}(\boldsymbol{X}) = 0 \mid H_{1}\right)}_{\beta_{\mathsf{MAP}}} \mathbb{P}(H_{1})$$

$$= \alpha_{\mathsf{MAP}} \mathbb{P}(H_0) + \beta_{\mathsf{MAP}} \mathbb{P}(H_1)$$
$$\leq \mathbb{P}\Big(\mathsf{error}_f\Big)$$

$$\mathbb{P}\left(\operatorname{error}_{\mathsf{MAP}}\right)$$

$$= \mathbb{P}\left(f_{\mathsf{MAP}}(\boldsymbol{X}) = 1, H_0\right) + \mathbb{P}\left(f_{\mathsf{MAP}}(\boldsymbol{X}) = 0, H_1\right)$$

$$= \underbrace{\mathbb{P}\left(f_{\mathsf{MAP}}(\boldsymbol{X}) = 1 \mid H_0\right)}_{\alpha_{\mathsf{MAP}}} \mathbb{P}(H_0) + \underbrace{\mathbb{P}\left(f_{\mathsf{MAP}}(\boldsymbol{X}) = 0 \mid H_1\right)}_{\beta_{\mathsf{MAP}}} \mathbb{P}(H_1)$$

$$= \alpha_{\mathsf{MAP}} \mathbb{P}(H_0) + \beta_{\mathsf{MAP}} \mathbb{P}(H_1)$$

$$\leq \mathbb{P}\Big(\mathsf{error}_f\Big) = \mathbb{P}\Big(f(\mathbf{X}) = 1, H_0\Big) + \mathbb{P}\Big(f(\mathbf{X}) = 0, H_1\Big)$$

$$\mathbb{P}\left(\operatorname{error}_{\mathsf{MAP}}\right)$$

$$= \mathbb{P}\left(f_{\mathsf{MAP}}(\boldsymbol{X}) = 1, H_{0}\right) + \mathbb{P}\left(f_{\mathsf{MAP}}(\boldsymbol{X}) = 0, H_{1}\right)$$

$$= \underbrace{\mathbb{P}\left(f_{\mathsf{MAP}}(\boldsymbol{X}) = 1 \mid H_{0}\right)}_{\alpha_{\mathsf{MAP}}} \mathbb{P}(H_{0}) + \underbrace{\mathbb{P}\left(f_{\mathsf{MAP}}(\boldsymbol{X}) = 0 \mid H_{1}\right)}_{\beta_{\mathsf{MAP}}} \mathbb{P}(H_{1})$$

$$= \alpha_{\mathsf{MAP}} \mathbb{P}(H_0) + \beta_{\mathsf{MAP}} \mathbb{P}(H_1)$$

$$\leq \mathbb{P}\Big(\mathsf{error}_f\Big) = \mathbb{P}\Big(f(\mathbf{X}) = 1, H_0\Big) + \mathbb{P}\Big(f(\mathbf{X}) = 0, H_1\Big)$$

$$= \alpha \mathbb{P}(H_0) + \beta \mathbb{P}(H_1)$$

• $f_{MAP}(x)$ minimizes probability of error over all rules (det. and prob.):

$$\mathbb{P}\left(\text{error}_{\mathsf{MAP}}\right)$$

$$= \mathbb{P}\left(f_{\mathsf{MAP}}(\boldsymbol{X}) = 1, H_0\right) + \mathbb{P}\left(f_{\mathsf{MAP}}(\boldsymbol{X}) = 0, H_1\right)$$

$$= \underbrace{\mathbb{P}\left(f_{\mathsf{MAP}}(\boldsymbol{X}) = 1 \mid H_0\right)}_{\alpha_{\mathsf{MAP}}} \mathbb{P}(H_0) + \underbrace{\mathbb{P}\left(f_{\mathsf{MAP}}(\boldsymbol{X}) = 0 \mid H_1\right)}_{\beta_{\mathsf{MAP}}} \mathbb{P}(H_1)$$

$$= \alpha_{\mathsf{MAP}} \mathbb{P}(H_0) + \beta_{\mathsf{MAP}} \mathbb{P}(H_1)$$

$$\leq \mathbb{P}\Big(\mathsf{error}_f\Big) = \mathbb{P}\Big(f(\boldsymbol{X}) = 1, H_0\Big) + \mathbb{P}\Big(f(\boldsymbol{X}) = 0, H_1\Big)$$
$$= \alpha \mathbb{P}(H_0) + \beta \mathbb{P}(H_1)$$

That is,
$$(\alpha - \alpha_{\mathsf{MAP}}) \mathbb{P}(H_0) + (\beta - \beta_{\mathsf{MAP}}) \mathbb{P}(H_1) \ge 0$$

Decision Theory

•
$$(\alpha - \alpha_{\mathsf{MAP}}) \mathbb{P}(H_0) + (\beta - \beta_{\mathsf{MAP}}) \mathbb{P}(H_1) \ge 0$$

• $(\alpha - \alpha_{\mathsf{MAP}}) \mathbb{P}(H_0) + (\beta - \beta_{\mathsf{MAP}}) \mathbb{P}(H_1) \ge 0$

• $f_{MAP}(\boldsymbol{x})$ is characterized by $f_{MAP}(\boldsymbol{x}) = 1$ if $\mathcal{L}(\boldsymbol{x}) \ge \eta = \frac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)}$

- $(\alpha \alpha_{\mathsf{MAP}}) \mathbb{P}(H_0) + (\beta \beta_{\mathsf{MAP}}) \mathbb{P}(H_1) \ge 0$
- $f_{\mathsf{MAP}}(x)$ is characterized by $f_{\mathsf{MAP}}(x) = 1$ if $\mathcal{L}(x) \ge \eta = \frac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)}$
- Selecting $\eta > 0$ such that

$$\mathbb{P}(H_0) = \frac{\eta}{\eta + 1} \qquad \mathbb{P}(H_1) = \frac{1}{\eta + 1},$$

- $(\alpha \alpha_{\mathsf{MAP}}) \mathbb{P}(H_0) + (\beta \beta_{\mathsf{MAP}}) \mathbb{P}(H_1) \ge 0$
- $f_{MAP}(\boldsymbol{x})$ is characterized by $f_{MAP}(\boldsymbol{x}) = 1$ if $\mathcal{L}(\boldsymbol{x}) \geq \eta = \frac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)}$
- Selecting $\eta > 0$ such that

$$\mathbb{P}(H_0) = \frac{\eta}{\eta+1} \qquad \mathbb{P}(H_1) = \frac{1}{\eta+1},$$

we have $\,f_{\rm LRT}({\boldsymbol x}\,;\,\eta)=f_{\rm MAP}({\boldsymbol x})\,$

- $(\alpha \alpha_{\mathsf{MAP}}) \mathbb{P}(H_0) + (\beta \beta_{\mathsf{MAP}}) \mathbb{P}(H_1) \ge 0$
- $f_{MAP}(\boldsymbol{x})$ is characterized by $f_{MAP}(\boldsymbol{x}) = 1$ if $\mathcal{L}(\boldsymbol{x}) \geq \eta = \frac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)}$
- Selecting $\eta > 0$ such that

$$\mathbb{P}(H_0)=\frac{\eta}{\eta+1}\qquad \mathbb{P}(H_1)=\frac{1}{\eta+1}\,,$$
 we have $f_{\mathsf{LRT}}(\pmb{x}\,;\,\eta)=f_{\mathsf{MAP}}(\pmb{x})$

• Thus,

$$(\alpha - \alpha_{\mathsf{LRT}}) \frac{\eta}{\eta + 1} + (\beta - \beta_{\mathsf{LRT}}) \frac{1}{\eta + 1} \ge 0$$

- $(\alpha \alpha_{\mathsf{MAP}}) \mathbb{P}(H_0) + (\beta \beta_{\mathsf{MAP}}) \mathbb{P}(H_1) \ge 0$
- $f_{MAP}(\boldsymbol{x})$ is characterized by $f_{MAP}(\boldsymbol{x}) = 1$ if $\mathcal{L}(\boldsymbol{x}) \geq \eta = \frac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)}$
- Selecting $\eta > 0$ such that

$$\mathbb{P}(H_0)=\frac{\eta}{\eta+1}\qquad \mathbb{P}(H_1)=\frac{1}{\eta+1}\,,$$
 we have $f_{\rm LRT}({\pmb x}\,;\,\eta)=f_{\rm MAP}({\pmb x})$

• Thus,

$$(\alpha - \alpha_{\mathsf{LRT}}) \underbrace{\frac{\eta}{\eta + 1}}_{>0} + (\beta - \beta_{\mathsf{LRT}}) \frac{1}{\eta + 1} \ge 0$$

- $(\alpha \alpha_{\mathsf{MAP}}) \mathbb{P}(H_0) + (\beta \beta_{\mathsf{MAP}}) \mathbb{P}(H_1) \ge 0$
- $f_{MAP}(\boldsymbol{x})$ is characterized by $f_{MAP}(\boldsymbol{x}) = 1$ if $\mathcal{L}(\boldsymbol{x}) \geq \eta = \frac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)}$
- Selecting $\eta > 0$ such that

$$\mathbb{P}(H_0)=\frac{\eta}{\eta+1}\qquad \mathbb{P}(H_1)=\frac{1}{\eta+1}\,,$$
 we have $f_{\rm LRT}({\pmb x}\,;\,\eta)=f_{\rm MAP}({\pmb x})$

• Thus,

$$(\alpha - \alpha_{\mathsf{LRT}}) \underbrace{\frac{\eta}{\eta + 1}}_{>0} + (\beta - \beta_{\mathsf{LRT}}) \underbrace{\frac{1}{\eta + 1}}_{>0} \ge 0$$

- $(\alpha \alpha_{\mathsf{MAP}}) \mathbb{P}(H_0) + (\beta \beta_{\mathsf{MAP}}) \mathbb{P}(H_1) \ge 0$
- $f_{MAP}(\boldsymbol{x})$ is characterized by $f_{MAP}(\boldsymbol{x}) = 1$ if $\mathcal{L}(\boldsymbol{x}) \geq \eta = \frac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)}$
- Selecting $\eta > 0$ such that

$$\mathbb{P}(H_0)=\frac{\eta}{\eta+1}\qquad \mathbb{P}(H_1)=\frac{1}{\eta+1}\,,$$
 we have $f_{\rm LRT}({\pmb x}\,;\,\eta)=f_{\rm MAP}({\pmb x})$

• Thus,

$$(\alpha - \alpha_{\mathsf{LRT}}) \underbrace{\frac{\eta}{\eta + 1}}_{>0} + (\beta - \beta_{\mathsf{LRT}}) \underbrace{\frac{1}{\eta + 1}}_{>0} \ge 0$$

• Therefore,

- $(\alpha \alpha_{\mathsf{MAP}}) \mathbb{P}(H_0) + (\beta \beta_{\mathsf{MAP}}) \mathbb{P}(H_1) \ge 0$
- $f_{\mathsf{MAP}}(x)$ is characterized by $f_{\mathsf{MAP}}(x) = 1$ if $\mathcal{L}(x) \ge \eta = \frac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)}$
- Selecting $\eta > 0$ such that

$$\mathbb{P}(H_0)=\frac{\eta}{\eta+1}\qquad \mathbb{P}(H_1)=\frac{1}{\eta+1}\,,$$
 we have $f_{\rm LRT}({\pmb x}\,;\,\eta)=f_{\rm MAP}({\pmb x})$

• Thus,

$$(\alpha - \alpha_{\mathsf{LRT}}) \underbrace{\frac{\eta}{\eta + 1}}_{>0} + (\beta - \beta_{\mathsf{LRT}}) \underbrace{\frac{1}{\eta + 1}}_{>0} \ge 0$$

• Therefore, $\alpha \leq \alpha_{LRT} \implies \beta \geq \beta_{LRT}$

- $(\alpha \alpha_{\mathsf{MAP}}) \mathbb{P}(H_0) + (\beta \beta_{\mathsf{MAP}}) \mathbb{P}(H_1) \ge 0$
- $f_{\mathsf{MAP}}(x)$ is characterized by $f_{\mathsf{MAP}}(x) = 1$ if $\mathcal{L}(x) \ge \eta = \frac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)}$
- Selecting $\eta > 0$ such that

$$\mathbb{P}(H_0)=\frac{\eta}{\eta+1}\qquad \mathbb{P}(H_1)=\frac{1}{\eta+1}\,,$$
 we have $f_{\rm LRT}({\pmb x}\,;\,\eta)=f_{\rm MAP}({\pmb x})$

• Thus,

$$(\alpha - \alpha_{\mathsf{LRT}}) \underbrace{\frac{\eta}{\eta+1}}_{>0} + (\beta - \beta_{\mathsf{LRT}}) \underbrace{\frac{1}{\eta+1}}_{>0} \ge 0$$

• Therefore, $\alpha \leq \alpha_{LRT} \implies \beta \geq \beta_{LRT}$ $\beta \leq \beta_{LRT} \implies \alpha \geq \alpha_{LRT}$

- $(\alpha \alpha_{\mathsf{MAP}}) \mathbb{P}(H_0) + (\beta \beta_{\mathsf{MAP}}) \mathbb{P}(H_1) \ge 0$
- $f_{\mathsf{MAP}}(\boldsymbol{x})$ is characterized by $f_{\mathsf{MAP}}(\boldsymbol{x}) = 1$ if $\mathcal{L}(\boldsymbol{x}) \geq \eta = \frac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)}$
- Selecting $\eta > 0$ such that

$$\mathbb{P}(H_0) = rac{\eta}{\eta+1} \qquad \mathbb{P}(H_1) = rac{1}{\eta+1}\,,$$
 have $f_{\mathsf{LRT}}(oldsymbol{x}\,;\,\eta) = f_{\mathsf{MAP}}(oldsymbol{x})$

• Thus,

we

$$(\alpha - \alpha_{\mathsf{LRT}}) \underbrace{\frac{\eta}{\eta + 1}}_{>0} + (\beta - \beta_{\mathsf{LRT}}) \underbrace{\frac{1}{\eta + 1}}_{>0} \ge 0$$

• Therefore, $\alpha \leq \alpha_{LRT} \implies \beta \geq \beta_{LRT}$ $\beta \leq \beta_{LRT} \implies \alpha \geq \alpha_{LRT}$

Receiver Operating Characteristic (ROC)

Consider a likelihood ratio test with threshold η :

Consider a likelihood ratio test with threshold η : $f_{LRT}(\boldsymbol{x}; \eta)$

Consider a likelihood ratio test with threshold η : $f_{\mathsf{LRT}}(\boldsymbol{x};\eta)$

For each η , there is a pair

Consider a likelihood ratio test with threshold η : $f_{LRT}(\boldsymbol{x}; \eta)$

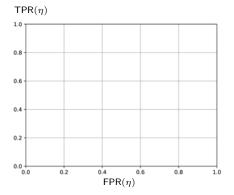
For each η , there is a pair $(\alpha(\eta), 1 - \beta(\eta))$

Consider a likelihood ratio test with threshold η : $f_{LRT}(\boldsymbol{x}; \eta)$

For each η , there is a pair $(\alpha(\eta), 1 - \beta(\eta)) = (\mathsf{FPR}(\eta), \mathsf{TPR}(\eta))$

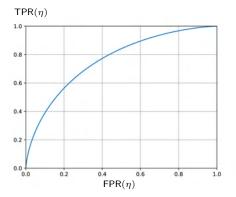
Consider a likelihood ratio test with threshold η : $f_{LRT}(\boldsymbol{x}; \eta)$

For each η , there is a pair $(\alpha(\eta), 1 - \beta(\eta)) = (FPR(\eta), TPR(\eta))$



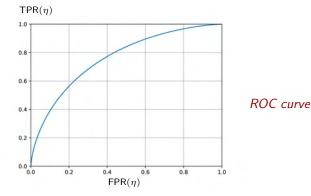
Consider a likelihood ratio test with threshold η : $f_{LRT}(\boldsymbol{x}; \eta)$

For each η , there is a pair $(\alpha(\eta), 1 - \beta(\eta)) = (\mathsf{FPR}(\eta), \mathsf{TPR}(\eta))$



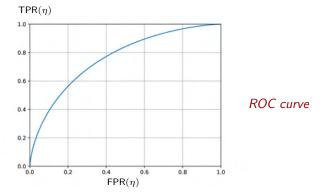
Consider a likelihood ratio test with threshold η : $f_{LRT}(x; \eta)$

For each $\eta,$ there is a pair $\left(\alpha(\eta), 1 - \beta(\eta)\right) = \left(\mathsf{FPR}(\eta), \mathsf{TPR}(\eta)\right)$



Consider a likelihood ratio test with threshold η : $f_{LRT}(x; \eta)$

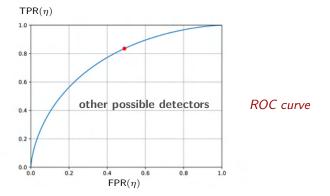
For each $\eta,$ there is a pair $\left(\alpha(\eta), 1 - \beta(\eta)\right) = \left(\mathsf{FPR}(\eta), \mathsf{TPR}(\eta)\right)$



Neyman-Pearson implies Pareto optimality

Consider a likelihood ratio test with threshold η : $f_{LRT}(\boldsymbol{x}; \eta)$

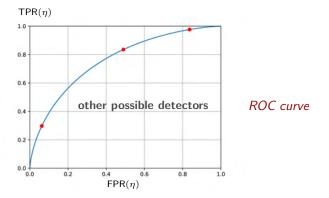
For each $\eta,$ there is a pair $\left(\alpha(\eta), 1 - \beta(\eta)\right) = \left(\mathsf{FPR}(\eta), \mathsf{TPR}(\eta)\right)$



Neyman-Pearson implies Pareto optimality

Consider a likelihood ratio test with threshold η : $f_{LRT}(\boldsymbol{x}; \eta)$

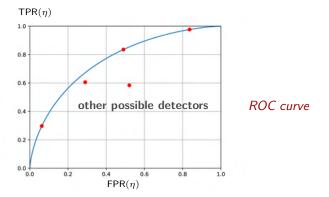
For each $\eta,$ there is a pair $\left(\alpha(\eta), 1 - \beta(\eta)\right) = \left(\mathsf{FPR}(\eta), \mathsf{TPR}(\eta)\right)$



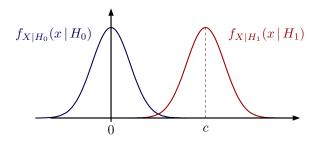
Neyman-Pearson implies Pareto optimality

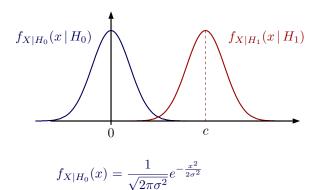
Consider a likelihood ratio test with threshold η : $f_{LRT}(\boldsymbol{x}; \eta)$

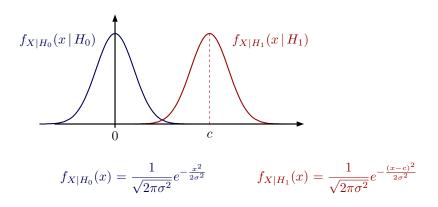
For each $\eta,$ there is a pair $\left(\alpha(\eta), 1 - \beta(\eta)\right) = \left(\mathsf{FPR}(\eta), \mathsf{TPR}(\eta)\right)$

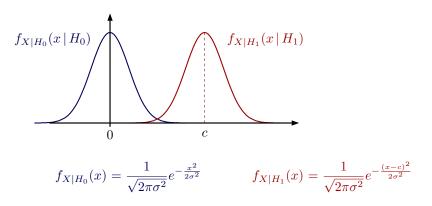


Neyman-Pearson implies Pareto optimality

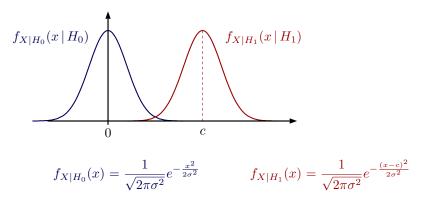






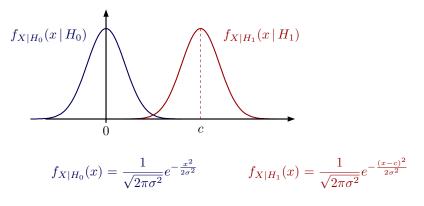


Likelihood ratio test:



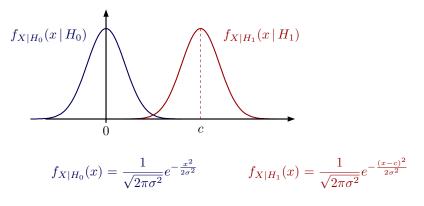
Likelihood ratio test:

$$\mathcal{L}(x) = \frac{f_{X|H_1}(x)}{f_{X|H_0}(x)} \ge \eta$$



Likelihood ratio test:

$$\mathcal{L}(x) = \frac{f_{X|H_1}(x)}{f_{X|H_0}(x)} \ge \eta \quad \Longleftrightarrow \quad x \ge \frac{c}{2} + \frac{\sigma^2}{c} \log \eta$$



Likelihood ratio test:

$$\mathcal{L}(x) = \frac{f_{X|H_1}(x)}{f_{X|H_0}(x)} \ge \eta \quad \Longleftrightarrow \quad x \ge \frac{c}{2} + \frac{\sigma^2}{c} \log \eta =: \gamma$$

 $x \geq \frac{c}{2} + \frac{\sigma^2}{c} \log \, \eta \, := \, \gamma$

 $x \geq \frac{c}{2} + \frac{\sigma^2}{c} \log \, \eta \, := \, \gamma$

 $\mathsf{TPR}(\eta)$

$$x \ge \frac{c}{2} + \frac{\sigma^2}{c} \log \eta := \gamma$$

$$\mathsf{TPR}(\eta) = \mathbb{P}\Big(X \ge \gamma \, \big| \, \underline{H}_1\Big)$$

$$x \ge \frac{c}{2} + \frac{\sigma^2}{c} \log \eta := \gamma$$

$$\mathsf{TPR}(\eta) = \mathbb{P}\Big(X \ge \gamma \, \big| \, H_1\Big) = \int_{\gamma}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-c)^2}{2\sigma^2}} \, \mathrm{d} \, x$$

$$x \geq \tfrac{c}{2} + \tfrac{\sigma^2}{c} \log \, \eta \, := \, \gamma$$

$$\mathsf{TPR}(\eta) = \mathbb{P}\Big(X \ge \gamma \, \big| \, H_1\Big) = \int_{\gamma}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-c)^2}{2\sigma^2}} \, \mathrm{d} x$$
$$= \int_{\frac{\gamma-c}{\sigma}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \, \mathrm{d} z$$

$$x \ge \frac{c}{2} + \frac{\sigma^2}{c} \log \eta := \gamma$$

$$\begin{aligned} \mathsf{TPR}(\eta) &= \mathbb{P}\Big(X \ge \gamma \,\big|\, H_1\Big) = \int_{\gamma}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-c)^2}{2\sigma^2}} \,\mathrm{d}\, x \\ &= \int_{\frac{\gamma-c}{\sigma}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \,\mathrm{d}z \\ &= Q\Big(\frac{\gamma-c}{\sigma}\Big) \end{aligned}$$

$$x \geq \frac{c}{2} + \frac{\sigma^2}{c} \log \, \eta \, := \, \gamma$$

$$\begin{aligned} \mathsf{TPR}(\eta) &= \mathbb{P}\Big(X \ge \gamma \,\Big|\, H_1\Big) = \int_{\gamma}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-c)^2}{2\sigma^2}} \,\mathrm{d}\,x \\ &= \int_{\frac{\gamma-c}{\sigma}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \,\mathrm{d}z \\ &= Q\Big(\frac{\gamma-c}{\sigma}\Big) \qquad \qquad \left[Q(x) := \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-\frac{z^2}{2}} \,\mathrm{d}z\right] \end{aligned}$$

$$x \geq \frac{c}{2} + \frac{\sigma^2}{c} \log \, \eta \, := \, \gamma$$

$$\begin{aligned} \mathsf{TPR}(\eta) &= \mathbb{P}\Big(X \ge \gamma \mid H_1\Big) = \int_{\gamma}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-c)^2}{2\sigma^2}} \,\mathrm{d}\,x \\ &= \int_{\frac{\gamma-c}{\sigma}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \,\mathrm{d}z \\ &= Q\Big(\frac{\gamma-c}{\sigma}\Big) \\ &= Q\bigg(\frac{\log \eta}{c/\sigma} - \frac{c/\sigma}{2}\bigg) \end{aligned} \qquad \begin{bmatrix} Q(x) := \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-\frac{z^2}{2}} \,\mathrm{d}z \\ &= Q\bigg(\frac{\log \eta}{c/\sigma} - \frac{c/\sigma}{2}\bigg) \end{aligned}$$

$$x \geq \tfrac{c}{2} + \tfrac{\sigma^2}{c} \log \, \eta \, := \, \gamma$$

$$\begin{aligned} \mathsf{TPR}(\eta) &= \mathbb{P}\Big(X \ge \gamma \,\big|\, H_1\Big) = \int_{\gamma}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-c)^2}{2\sigma^2}} \,\mathrm{d}\,x \\ &= \int_{\frac{\gamma-c}{\sigma}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \,\mathrm{d}z \\ &= Q\Big(\frac{\gamma-c}{\sigma}\Big) \\ &= Q\Big(\frac{\log\eta}{c/\sigma} - \frac{c/\sigma}{2}\Big) \\ &= Q\Big(\frac{\log\eta}{c/\sigma} - \frac{\mathsf{SNR}}{2}\Big) \end{aligned}$$

$$x \geq \tfrac{c}{2} + \tfrac{\sigma^2}{c} \log \, \eta \, := \, \gamma$$

$$\begin{aligned} \mathsf{TPR}(\eta) &= \mathbb{P}\Big(X \ge \gamma \mid H_1\Big) = \int_{\gamma}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-c)^2}{2\sigma^2}} \,\mathrm{d}\,x \\ &= \int_{\frac{\gamma-c}{\sigma}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \,\mathrm{d}z \\ &= Q\Big(\frac{\gamma-c}{\sigma}\Big) \qquad \qquad \left[Q(x) := \frac{1}{\sqrt{2\pi}} \int_{x}^{\infty} e^{-\frac{z^2}{2}} \,\mathrm{d}z\right] \\ &= Q\Big(\frac{\log\eta}{c/\sigma} - \frac{c/\sigma}{2}\Big) \\ &= Q\Big(\frac{\log\eta}{\mathsf{SNR}} - \frac{\mathsf{SNR}}{2}\Big) \qquad \qquad \left[\mathsf{SNR} := \frac{c}{\sigma}\right] \end{aligned}$$

$$x \geq \frac{c}{2} + \frac{\sigma^2}{c} \log \, \eta \, := \, \gamma$$

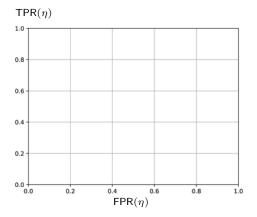
$$\begin{aligned} \mathsf{TPR}(\eta) &= \mathbb{P}\Big(X \ge \gamma \mid H_1\Big) = \int_{\gamma}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-c)^2}{2\sigma^2}} \,\mathrm{d}\,x \\ &= \int_{\frac{\gamma-c}{\sigma}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \,\mathrm{d}z \\ &= Q\Big(\frac{\gamma-c}{\sigma}\Big) \qquad \qquad \left[Q(x) := \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-\frac{z^2}{2}} \,\mathrm{d}z\right] \\ &= Q\Big(\frac{\log\eta}{c/\sigma} - \frac{c/\sigma}{2}\Big) \\ &= Q\Big(\frac{\log\eta}{\mathsf{SNR}} - \frac{\mathsf{SNR}}{2}\Big) \qquad \qquad \left[\mathsf{SNR} := \frac{c}{\sigma}\right] \end{aligned}$$

Similarly,

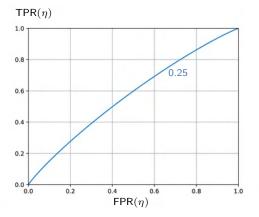
$$\mathsf{FPR}(\eta) = \mathbb{P}\left(X \ge \gamma \mid H_0\right) = \cdots = Q\left(\frac{\log \eta}{\mathsf{SNR}} + \frac{\mathsf{SNR}}{2}\right)$$

ROC curve for different values of SNR

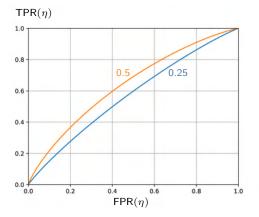
ROC curve for different values of SNR

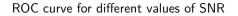


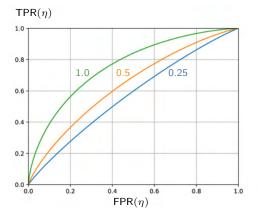






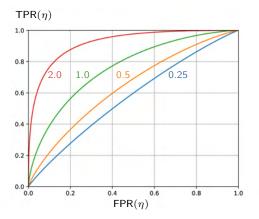






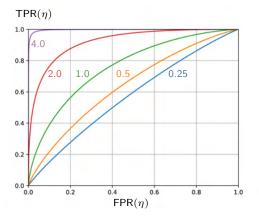
Example

ROC curve for different values of SNR



Example

ROC curve for different values of SNR



Decision Theory

Property 1: (0,0) and (1,1) are in the ROC curve

Property 1: (0,0) and (1,1) are in the ROC curve

Property 2: TPR \geq FPR

Property 1: (0,0) and (1,1) are in the ROC curve

Property 2: TPR \geq FPR

Property 3: The ROC curve is concave

Property 1: (0,0) and (1,1) are in the ROC curve

Proof:

Property 1: (0,0) and (1,1) are in the ROC curve

Proof:

• When $\eta \to +\infty$,

Property 1: (0,0) and (1,1) are in the ROC curve

Proof:

• When $\eta \to +\infty$,

$$f(x) = \mathbb{1}_{\mathcal{L}(\boldsymbol{x}) \ge \eta}(\boldsymbol{x})$$

Property 1: (0,0) and (1,1) are in the ROC curve

Proof:

• When $\eta \to +\infty$,

$$f(x) = \mathbb{1}_{\mathcal{L}(\boldsymbol{x}) \ge \eta}(\boldsymbol{x}) = 0$$

Property 1: (0,0) and (1,1) are in the ROC curve

Proof:

• When $\eta \to +\infty$,

$$f(x) = \mathbb{1}_{\mathcal{L}(\boldsymbol{x}) \ge \eta}(\boldsymbol{x}) = 0$$

for any \boldsymbol{x}

Property 1: (0,0) and (1,1) are in the ROC curve

Proof:

• When $\eta \to +\infty$,

$$f(x) = \mathbb{1}_{\mathcal{L}(\boldsymbol{x}) \ge \eta}(\boldsymbol{x}) = 0$$

for any x

Then,
$$(\mathsf{FPR}(+\infty), \mathsf{TPR}(+\infty)) = (0,0)$$

Property 1: (0,0) and (1,1) are in the ROC curve

Proof:

• When $\eta \to +\infty$,

$$f(x) = \mathbb{1}_{\mathcal{L}(\boldsymbol{x}) \ge \eta}(\boldsymbol{x}) = 0$$

for any x

Then,
$$(FPR(+\infty), TPR(+\infty)) = (0,0)$$

• Similarly, $\left(\mathsf{FPR}(-\infty), \, \mathsf{TPR}(-\infty)\right) = (1,1)$

Property 1: (0,0) and (1,1) are in the ROC curve

Proof:

• When $\eta \to +\infty$,

$$f(x) = \mathbb{1}_{\mathcal{L}(\boldsymbol{x}) \geq \eta}(\boldsymbol{x}) = 0$$

for any x

Then,
$$(FPR(+\infty), TPR(+\infty)) = (0,0)$$

• Similarly, $(FPR(-\infty), TPR(-\infty)) = (1,1)$

 \square

Consider two decision rules (omitting dependence on η):

$$\left(\mathsf{FPR}^{(1)},\,\mathsf{TPR}^{(1)}\right)$$
, $\left(\mathsf{FPR}^{(2)},\,\mathsf{TPR}^{(2)}\right)$

Consider two decision rules (omitting dependence on η):

$$\left(\mathsf{FPR}^{(1)}\,,\,\mathsf{TPR}^{(1)}
ight)\,\,,\,\,\left(\mathsf{FPR}^{(2)}\,,\,\mathsf{TPR}^{(2)}
ight)$$

Build a <code>randomized</code> decision rule by selecting 1 w.p. p, and 2 w.p. 1-p

Consider two decision rules (omitting dependence on η):

$$\left(\mathsf{FPR}^{(1)}\,,\,\mathsf{TPR}^{(1)}
ight)\,\,,\,\,\left(\mathsf{FPR}^{(2)}\,,\,\mathsf{TPR}^{(2)}
ight)$$

Build a <code>randomized</code> decision rule by selecting 1 w.p. p, and 2 w.p. 1-p

$$\left(\mathsf{FPR}\,,\,\mathsf{TPR}\right) = \left(p\,\mathsf{FPR}^{(1)} + (1-p)\mathsf{FPR}^{(2)}\,,\,\,p\,\mathsf{TPR}^{(1)} + (1-p)\mathsf{TPR}^{(2)}\right)$$

Consider two decision rules (omitting dependence on η):

$$\left(\mathsf{FPR}^{(1)}\,,\,\mathsf{TPR}^{(1)}
ight)\,\,,\,\,\left(\mathsf{FPR}^{(2)}\,,\,\mathsf{TPR}^{(2)}
ight)$$

Build a randomized decision rule by selecting 1 w.p. p, and 2 w.p. 1-p

$$(\mathsf{FPR}, \mathsf{TPR}) = (p \,\mathsf{FPR}^{(1)} + (1-p)\mathsf{FPR}^{(2)}, p \,\mathsf{TPR}^{(1)} + (1-p)\mathsf{TPR}^{(2)})$$

Property 2: TPR \geq FPR

Proof:

Consider two decision rules (omitting dependence on η):

$$\left(\mathsf{FPR}^{(1)}\,,\,\mathsf{TPR}^{(1)}
ight)\,\,,\,\,\left(\mathsf{FPR}^{(2)}\,,\,\mathsf{TPR}^{(2)}
ight)$$

Build a randomized decision rule by selecting 1 w.p. p, and 2 w.p. 1-p

$$(\mathsf{FPR}, \mathsf{TPR}) = \left(p \, \mathsf{FPR}^{(1)} + (1-p) \mathsf{FPR}^{(2)}, \ p \, \mathsf{TPR}^{(1)} + (1-p) \mathsf{TPR}^{(2)} \right)$$

Property 2: TPR \geq FPR

Proof:

• For any achievable $\alpha > 0$,

Consider two decision rules (omitting dependence on η):

$$\left(\mathsf{FPR}^{(1)}\,,\,\mathsf{TPR}^{(1)}
ight)\,\,,\,\,\left(\mathsf{FPR}^{(2)}\,,\,\mathsf{TPR}^{(2)}
ight)$$

Build a randomized decision rule by selecting 1 w.p. p, and 2 w.p. 1-p

$$(\mathsf{FPR}, \mathsf{TPR}) = \left(p \, \mathsf{FPR}^{(1)} + (1-p) \mathsf{FPR}^{(2)}, \ p \, \mathsf{TPR}^{(1)} + (1-p) \mathsf{TPR}^{(2)} \right)$$

Property 2: TPR \geq FPR

Proof:

• For any achievable $\alpha > 0$, we can always find a randomized rule s.t.

$$\mathsf{TPR} = \mathsf{FPR} = \alpha$$

Consider two decision rules (omitting dependence on η):

$$\left(\mathsf{FPR}^{(1)}\,,\,\mathsf{TPR}^{(1)}
ight)\,\,,\,\,\left(\mathsf{FPR}^{(2)}\,,\,\mathsf{TPR}^{(2)}
ight)$$

Build a randomized decision rule by selecting 1 w.p. p, and 2 w.p. 1-p

$$(\mathsf{FPR}, \mathsf{TPR}) = (p \, \mathsf{FPR}^{(1)} + (1-p) \, \mathsf{FPR}^{(2)}, p \, \mathsf{TPR}^{(1)} + (1-p) \, \mathsf{TPR}^{(2)})$$

Property 2: TPR \geq FPR

Proof:

• For any achievable $\alpha > 0$, we can always find a randomized rule s.t.

$$\mathsf{TPR} = \mathsf{FPR} = \alpha$$

• By the Neyman-Pearson lemma,

Consider two decision rules (omitting dependence on η):

$$\left(\mathsf{FPR}^{(1)}\,,\,\mathsf{TPR}^{(1)}
ight)\,\,,\,\,\left(\mathsf{FPR}^{(2)}\,,\,\mathsf{TPR}^{(2)}
ight)$$

Build a randomized decision rule by selecting 1 w.p. p, and 2 w.p. 1-p

$$\left(\mathsf{FPR}, \mathsf{TPR}\right) = \left(p \,\mathsf{FPR}^{(1)} + (1-p) \mathsf{FPR}^{(2)}, \ p \,\mathsf{TPR}^{(1)} + (1-p) \mathsf{TPR}^{(2)}\right)$$

Property 2: TPR \geq FPR

Proof:

• For any achievable $\alpha > 0$, we can always find a randomized rule s.t.

$$\mathsf{TPR} = \mathsf{FPR} = \alpha$$

• By the Neyman-Pearson lemma,

if an LRT has $FPR^* = \alpha$,

Consider two decision rules (omitting dependence on η):

$$\left(\mathsf{FPR}^{(1)}\,,\,\mathsf{TPR}^{(1)}
ight)\,\,,\,\,\left(\mathsf{FPR}^{(2)}\,,\,\mathsf{TPR}^{(2)}
ight)$$

Build a randomized decision rule by selecting 1 w.p. p, and 2 w.p. 1-p

$$\left(\mathsf{FPR}, \mathsf{TPR}\right) = \left(p \,\mathsf{FPR}^{(1)} + (1-p) \mathsf{FPR}^{(2)}, \ p \,\mathsf{TPR}^{(1)} + (1-p) \mathsf{TPR}^{(2)}\right)$$

Property 2: TPR \geq FPR

Proof:

• For any achievable $\alpha > 0$, we can always find a randomized rule s.t.

$$\mathsf{TPR} = \mathsf{FPR} = \alpha$$

• By the Neyman-Pearson lemma,

if an LRT has $\mathsf{FPR}^{\star} = \alpha$, then its $\mathsf{TPR}^{\star} \ge \alpha$

Decision Theory

Consider two decision rules (omitting dependence on η):

$$\left(\mathsf{FPR}^{(1)}\,,\,\mathsf{TPR}^{(1)}
ight)\,\,,\,\,\left(\mathsf{FPR}^{(2)}\,,\,\mathsf{TPR}^{(2)}
ight)$$

Build a randomized decision rule by selecting 1 w.p. p, and 2 w.p. 1-p

$$\left(\mathsf{FPR}, \mathsf{TPR}\right) = \left(p \,\mathsf{FPR}^{(1)} + (1-p) \mathsf{FPR}^{(2)}, \ p \,\mathsf{TPR}^{(1)} + (1-p) \mathsf{TPR}^{(2)}\right)$$

Property 2: TPR \geq FPR

Proof:

• For any achievable $\alpha > 0$, we can always find a randomized rule s.t.

$$\mathsf{TPR} = \mathsf{FPR} = \alpha$$

• By the Neyman-Pearson lemma,

if an LRT has $\mathsf{FPR}^* = \alpha$, then its $\mathsf{TPR}^* \ge \alpha$

Decision Theory

Property 3: The ROC curve is concave

Proof:

Property 3: The ROC curve is concave

Proof:

• Consider two achievable points in the ROC diagram:

 $(\mathsf{FPR}(\eta_1), \mathsf{TPR}(\eta_1))$ and $(\mathsf{FPR}(\eta_2), \mathsf{TPR}(\eta_2))$

Property 3: The ROC curve is concave

Proof:

• Consider two achievable points in the ROC diagram:

 $(\mathsf{FPR}(\eta_1), \mathsf{TPR}(\eta_1))$ and $(\mathsf{FPR}(\eta_2), \mathsf{TPR}(\eta_2))$

• For any $0 \le t \le 1$, we can form a randomized rule such that

$$(t \operatorname{FPR}(\eta_1) + (1-t)\operatorname{FPR}(\eta_2), t \operatorname{TPR}(\eta_1) + (1-t)\operatorname{TPR}(\eta_2))$$
 (1)

Property 3: The ROC curve is concave

Proof:

• Consider two achievable points in the ROC diagram:

 $(\mathsf{FPR}(\eta_1), \mathsf{TPR}(\eta_1))$ and $(\mathsf{FPR}(\eta_2), \mathsf{TPR}(\eta_2))$

- For any $0 \le t \le 1$, we can form a randomized rule such that $\left(t \operatorname{FPR}(\eta_1) + (1-t)\operatorname{FPR}(\eta_2), t \operatorname{TPR}(\eta_1) + (1-t)\operatorname{TPR}(\eta_2)\right)$ (1)
- By the Neyman-Pearson lemma, if an LRT has

$$\mathsf{FPR}^{\star} = t \, \mathsf{FPR}(\eta_1) + (1-t) \mathsf{FPR}(\eta_2) \,\,,$$

Property 3: The ROC curve is concave

Proof:

• Consider two achievable points in the ROC diagram:

 $(\mathsf{FPR}(\eta_1), \mathsf{TPR}(\eta_1))$ and $(\mathsf{FPR}(\eta_2), \mathsf{TPR}(\eta_2))$

- For any $0 \le t \le 1$, we can form a randomized rule such that $\left(t \operatorname{FPR}(\eta_1) + (1-t)\operatorname{FPR}(\eta_2), t \operatorname{TPR}(\eta_1) + (1-t)\operatorname{TPR}(\eta_2)\right)$ (1)
- By the Neyman-Pearson lemma, if an LRT has

$$\begin{split} \mathsf{FPR}^{\star} &= t \, \mathsf{FPR}(\eta_1) + (1-t) \mathsf{FPR}(\eta_2) \,, \qquad \text{then} \\ \mathsf{TPR}^{\star} &\geq t \, \mathsf{TPR}(\eta_1) + (1-t) \mathsf{TPR}(\eta_2) \end{split}$$

Property 3: The ROC curve is concave

Proof:

• Consider two achievable points in the ROC diagram:

 $(\mathsf{FPR}(\eta_1), \mathsf{TPR}(\eta_1))$ and $(\mathsf{FPR}(\eta_2), \mathsf{TPR}(\eta_2))$

- For any $0 \le t \le 1$, we can form a randomized rule such that $\left(t \operatorname{FPR}(\eta_1) + (1-t)\operatorname{FPR}(\eta_2), t \operatorname{TPR}(\eta_1) + (1-t)\operatorname{TPR}(\eta_2)\right)$ (1)
- By the Neyman-Pearson lemma, if an LRT has

$$\begin{split} \mathsf{FPR}^{\star} &= t \, \mathsf{FPR}(\eta_1) + (1-t) \mathsf{FPR}(\eta_2) \ , \qquad \text{then} \\ \mathsf{TPR}^{\star} &\geq t \, \mathsf{TPR}(\eta_1) + (1-t) \mathsf{TPR}(\eta_2) \end{split}$$

Thus, the ROC curve is above (1)

Decision Theory

Property 3: The ROC curve is concave

Proof:

• Consider two achievable points in the ROC diagram:

 $(\mathsf{FPR}(\eta_1), \mathsf{TPR}(\eta_1))$ and $(\mathsf{FPR}(\eta_2), \mathsf{TPR}(\eta_2))$

- For any $0 \le t \le 1$, we can form a randomized rule such that $\left(t \operatorname{FPR}(\eta_1) + (1-t)\operatorname{FPR}(\eta_2), t \operatorname{TPR}(\eta_1) + (1-t)\operatorname{TPR}(\eta_2)\right)$ (1)
- By the Neyman-Pearson lemma, if an LRT has

$$\begin{split} \mathsf{FPR}^{\star} &= t \, \mathsf{FPR}(\eta_1) + (1-t) \mathsf{FPR}(\eta_2) \,, \qquad \text{then} \\ \mathsf{TPR}^{\star} &\geq t \, \mathsf{TPR}(\eta_1) + (1-t) \mathsf{TPR}(\eta_2) \end{split}$$

Thus, the ROC curve is above (1)

Decision Theory

Example: SARS-CoV-2 Tests

Example: SARS-CoV-2 Tests

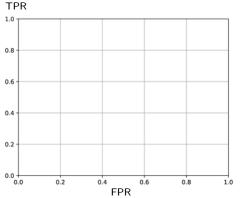
Lab-based tests (ELISA, CLIA) and rapid tests (lateral flow)

Example: SARS-CoV-2 Tests

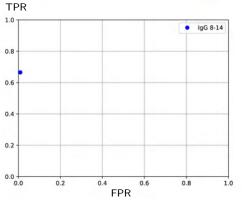
Lab-based tests (ELISA, CLIA) and rapid tests (lateral flow)

Detection of IgG, IgM, or IgG/IgM antibodies at days 8-14, 15-21, 22-35 (95% CI)

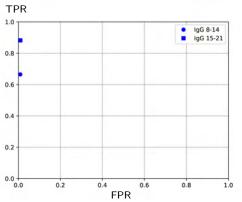
Lab-based tests (ELISA, CLIA) and rapid tests (lateral flow)



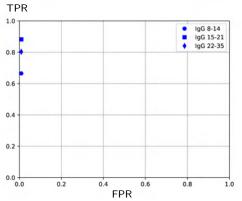
Lab-based tests (ELISA, CLIA) and rapid tests (lateral flow)



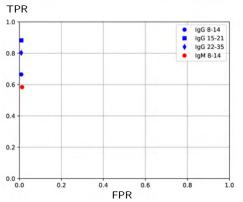
Lab-based tests (ELISA, CLIA) and rapid tests (lateral flow)



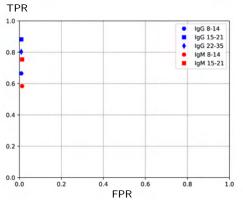
Lab-based tests (ELISA, CLIA) and rapid tests (lateral flow)



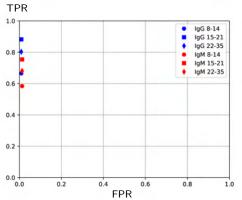
Lab-based tests (ELISA, CLIA) and rapid tests (lateral flow)



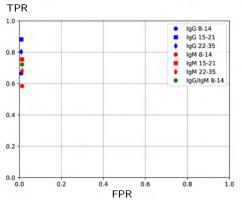
Lab-based tests (ELISA, CLIA) and rapid tests (lateral flow)



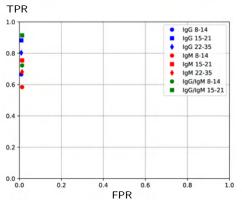
Lab-based tests (ELISA, CLIA) and rapid tests (lateral flow)



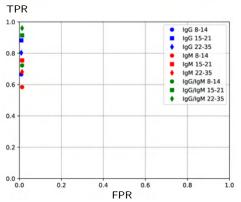
Lab-based tests (ELISA, CLIA) and rapid tests (lateral flow)



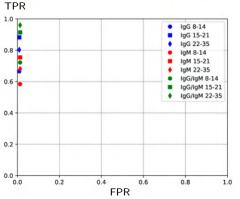
Lab-based tests (ELISA, CLIA) and rapid tests (lateral flow)



Lab-based tests (ELISA, CLIA) and rapid tests (lateral flow)



Lab-based tests (ELISA, CLIA) and rapid tests (lateral flow)





Decision Theory

We studied a (binary) decision problem:

$$\underset{f:\mathbb{R}^d \rightarrow \{0,1\}}{\text{minimize}} \ \mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}),Y\big)\Big]$$

We studied a (binary) decision problem:

$$\underset{f:\mathbb{R}^d\to\{0,1\}}{\text{minimize}} \mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}),Y\big)\Big]$$

Assumed known $f_{\boldsymbol{X}|H_i}(\boldsymbol{x} \,|\, H_i)$ and $\mathbb{P}(H_i)$

We studied a (binary) decision problem:

$$\underset{f:\mathbb{R}^d\to\{0,1\}}{\text{minimize}} \mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}),Y\big)\Big]$$

Assumed known $f_{\boldsymbol{X}|H_i}(\boldsymbol{x} \,|\, H_i)$ and $\mathbb{P}(H_i)$

```
What if they are unknown?
```

We studied a (binary) decision problem:

$$\underset{f:\mathbb{R}^d\to\{0,1\}}{\text{minimize}} \mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}),Y\big)\Big]$$

Assumed known $f_{\boldsymbol{X}|H_i}(\boldsymbol{x} \,|\, H_i)$ and $\mathbb{P}(H_i)$

What if they are unknown?

Empirical Risk Minimization (ERM):

We studied a (binary) decision problem:

$$\underset{f:\mathbb{R}^{d}\to\{0,1\}}{\text{minimize}} \mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}),Y\big)\Big]$$

Assumed known $f_{\boldsymbol{X}|H_i}(\boldsymbol{x} \mid H_i)$ and $\mathbb{P}(H_i)$

What if they are unknown?

Empirical Risk Minimization (ERM):

$$\underset{f:\mathbb{R}^d \rightarrow \{0,1\}}{\text{minimize}} \ \frac{1}{T} \sum_{t=1}^T \Big[\ell \big(f(\boldsymbol{x}_t), y_t \big) \Big]$$

We studied a (binary) decision problem:

$$\underset{f:\mathbb{R}^d\to\{0,1\}}{\text{minimize}} \mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}),Y\big)\Big]$$

Assumed known $f_{\boldsymbol{X}|H_i}(\boldsymbol{x} \,|\, H_i)$ and $\mathbb{P}(H_i)$

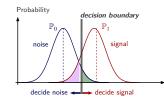
What if they are unknown?

Empirical Risk Minimization (ERM):

$$\underset{f:\mathbb{R}^d \rightarrow \{0,1\}}{\text{minimize}} \ \frac{1}{T} \sum_{t=1}^{T} \Big[\ell \big(f(\boldsymbol{x}_t), y_t \big) \Big]$$

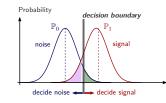
Assumption: we observe T samples $\{(\boldsymbol{x}_t,\,y_t)\}_{t=1}^T$

Optimal decision problems (binary case)



Optimal decision problems (binary case)

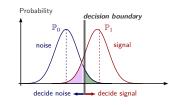
Likelihood ratio tests (LRT)



Optimal decision problems (binary case)

Likelihood ratio tests (LRT)

MAP and ML as particular cases

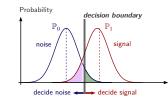


Optimal decision problems (binary case)

Likelihood ratio tests (LRT)

MAP and ML as particular cases

Optimality (Neyman-Pearson lemma)



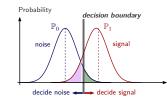
Optimal decision problems (binary case)

Likelihood ratio tests (LRT)

MAP and ML as particular cases

Optimality (Neyman-Pearson lemma)

Key assumption: known probability distributions



Optimal decision problems (binary case)

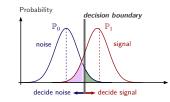
Likelihood ratio tests (LRT)

MAP and ML as particular cases

Optimality (Neyman-Pearson lemma)

Key assumption: known probability distributions

Types of errors



Optimal decision problems (binary case)

Likelihood ratio tests (LRT)

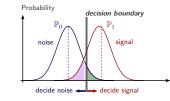
MAP and ML as particular cases

Optimality (Neyman-Pearson lemma)

Key assumption: known probability distributions

Types of errors

ROC curves and properties



Decision Theory

References

References



M. Hardt, B. Recht

Patterns, Predictions, and Actions

Princeton University Press, Oct, 2022

References



M. Hardt, B. Recht Patterns, Predictions, and Actions Princeton University Press, Oct, 2022



D. P. Bertsekas, J. N. Tsitsiklis Introduction to Probability

Athena Scientific, 2nd edition, 2008

T. Fauwcett, "An introduction to ROC analysis," Pattern Recognition Letters,

Vol. 27, pp. 861-874, 2006.

H. L. Van Trees, **Detection, Estimation, and Modulation: Part I**, John Wiley & Sons, 2001.