Application: Hypothesis testing

(see book Ch 8.5, example 8.23 page 444)

An example of inference, or learning

- can we distinguish among 2 or more possible situations?

when more than 2 situations, this is often called classification

---

signal detection (radar, communications, ...)

- detect whether signal present or absent.

H_1: hypothesis that signal is present

H_0: " " " " absent

Examples: detecting a whale in the ocean
detecting a person approaching your front door

Typical scenario: Known signal in additive noise

$Y = S + N$ if signal present (H_1)

$Y = N$ if signal absent (H_0)

Detector: observe Y and decide H_0 or H_1

$$g(Y) = \begin{cases} 1 & \text{if } Y \in A_1 \\ 0 & \text{if } Y \in A_1^c \end{cases}$$

A_1 is an event in the same experiment that generates Y

What makes a good decision? Low error.

2 kinds of mistakes

- deciding present when absent (false alarm)
- deciding absent when present (missed detection)

Typically measured through

P_F, probability of false alarm and
P_D, probability of detection

$$P_F = P(g(y) = 1 \mid H_0)$$ (a probability of one kind of mistake)

$$= P(A_1 \mid H_0)$$

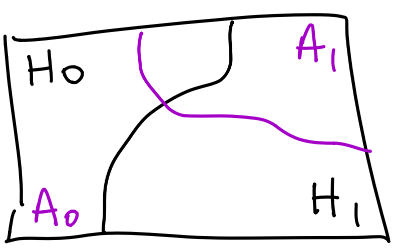
$$P_D = P(g(y) = 1 \mid H_1)$$ (a probability of one kind of correct decision)

$$= P(A_1 \mid H_1)$$

Another option P_m, probability of missed detection

$$P_m = P(g(y) = 0 \mid H_1)$$ (a probability of another kind of mistake)

$$= 1 - P_D = P(A_0 \mid H_1)$$



This terminology is just new names for concepts we covered back in Topic 1

What makes a good decision? Low cost

Each type of error may have a different cost or risk → depends on the applications

- too many false alarms may be worse than too many missed detections (or vice versa)

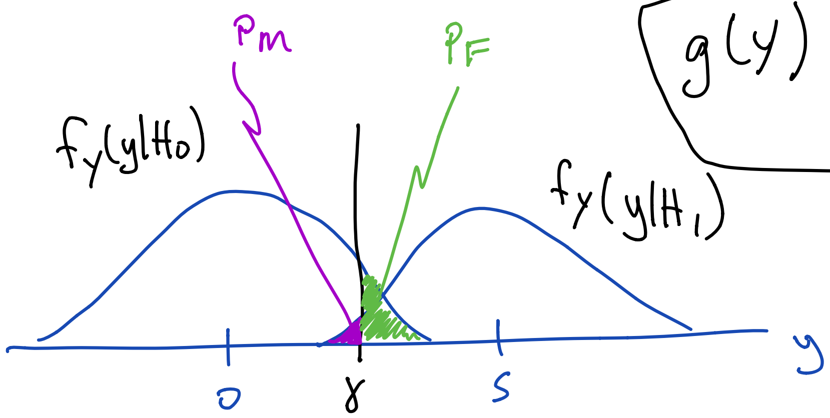
These different costs lead to different optimal decision functions $g(Y)$.

Suppose N is Gaussian noise with mean 0, variance σ^2,

and N is idependent of events H_0, H_1

Further, suppose our detector $g(Y)$ is just does thresholding.

$$g(y) = \begin{cases} 1 & y \geq \gamma \\ 0 & y < \gamma \end{cases}$$



$$P_F = P(g(Y) = 1 \mid H_0)$$

$$= P(Y \geq \gamma \mid H_0)$$

$$= P(N \geq \gamma) = 1 - \Phi\left(\frac{\gamma}{\sigma}\right)$$

$$P_D = P(g(Y) = 1 \mid H_1)$$

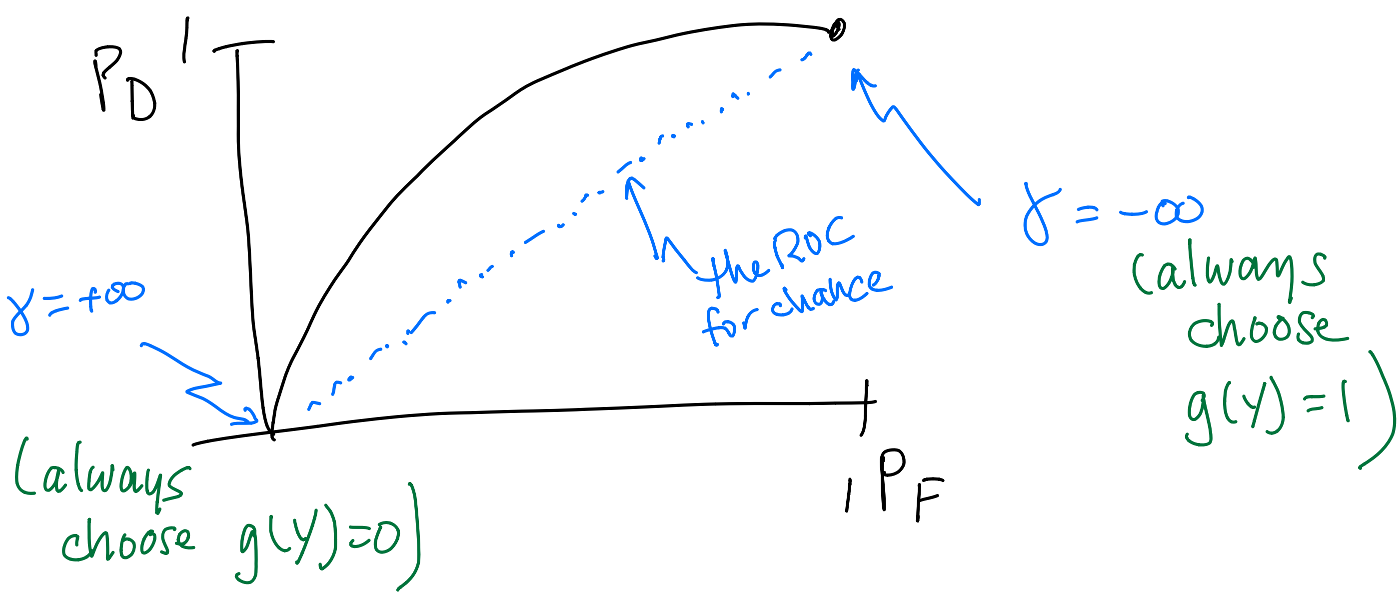
$$= P(Y \geq \gamma \mid H_1)$$

$$= P(S + N \geq \gamma)$$ since $Y = S + N$ when H_1 is true

$$= P(N \geq \gamma - S) = 1 - \Phi\left(\frac{\gamma - S}{\sigma}\right)$$

Can plot P_F vs. P_D for various γ

A Receiver Operating Characteristic (ROC)

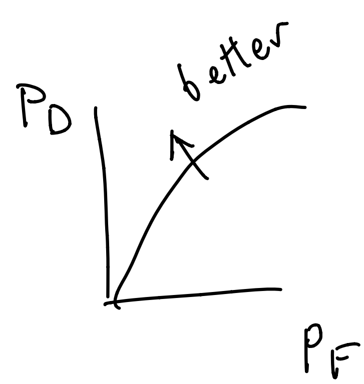


The threshold γ allows you to trade off the two types of errors, depending on the needs of your application.

$P_F = 0,\ P_D = 0 \longrightarrow$ never alarm

$P_F = 1,\ P_D = 1 \longrightarrow$ always alarm

A chance detector will have a diagonal ROC curve.
Always want your detector to be better than chance!



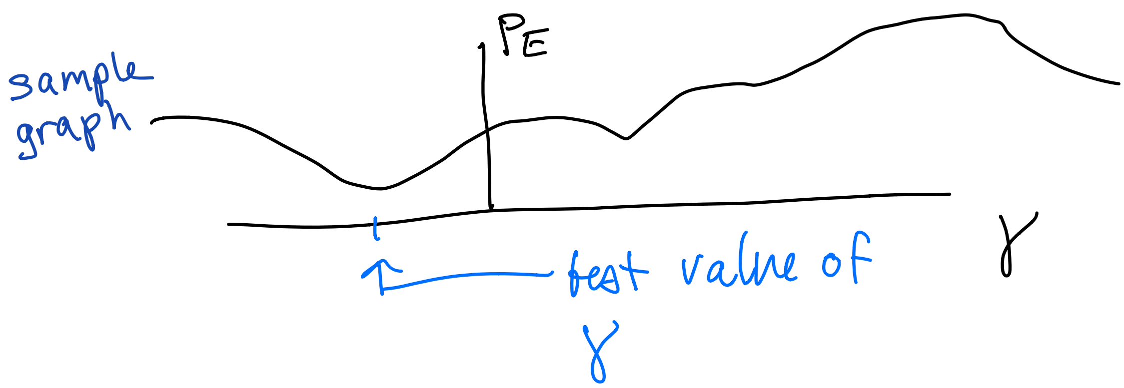
what threshold should you choose for your application?

Suppose you know $P(H_0)$ and $P(H_1)$, the underlying events.

Goal 1: minimize probability of error

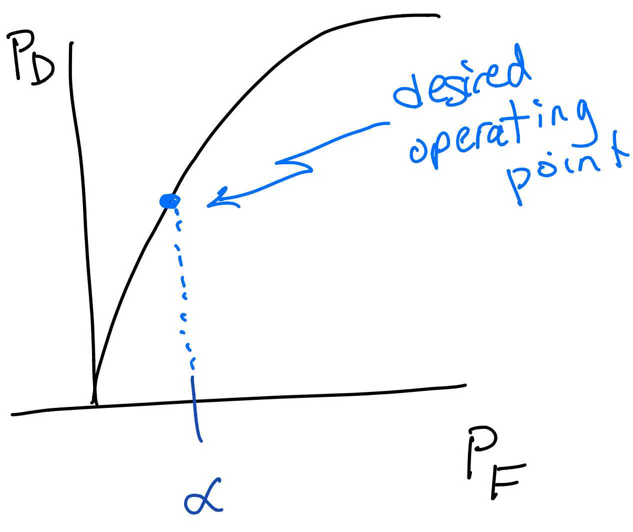
$$P_E = P(\text{error}) = P(\text{error}|H_0)\,P(H_0) + P(\text{error}|H_1)\,P(H_1)$$

$$= P_F\, P(H_0) + (1 - P_D)\, P(H_1)$$



Goal 2: minimize expected cost (omitted)

Goal 3: maximize P_D with constraint that $P_F \le \alpha$



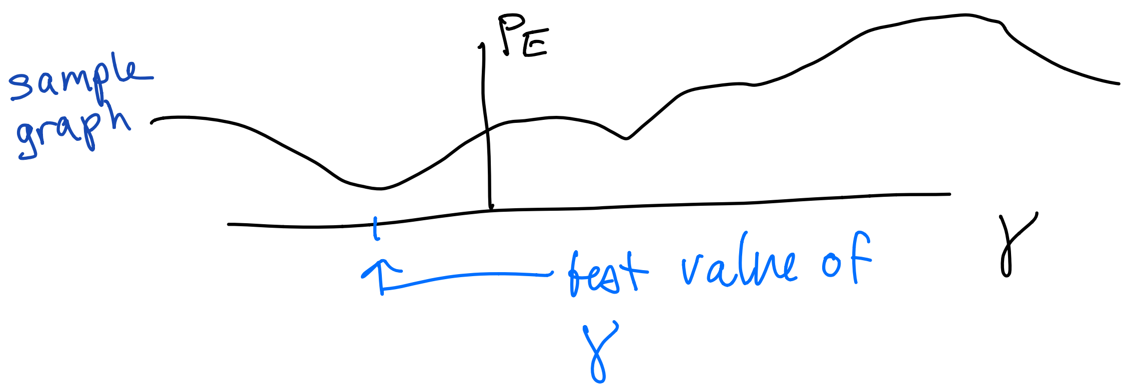
What threshold should you choose for your application? [A more full version w/mistakes]

Suppose you know $P(H_0)$ and $P(H_1)$, the underlying events.

Goal 1: minimize probability of error

$$P_E = P(\text{error}) = P(\text{error}|H_0)P(H_0) + P(\text{error}|H_1)P(H_1)$$

$$= P_F\, P(H_0) + (1 - P_D)\, P(H_1)$$



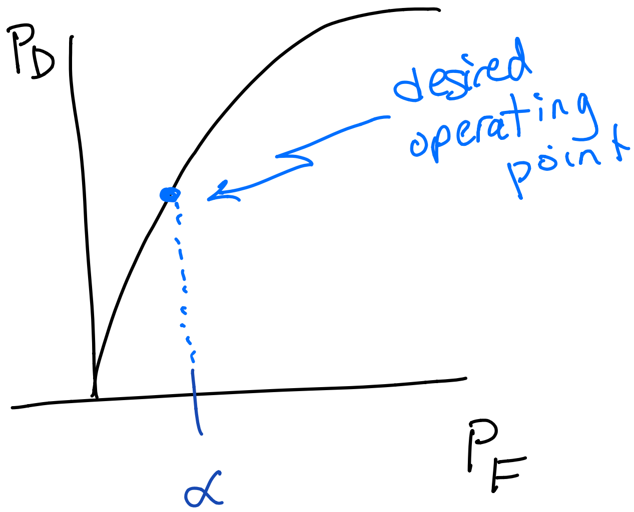
Goal 2: minimize expected cost

$$E(C) = E(C|H_0)P(H_0) + E(C|H_1)P(H_1)$$

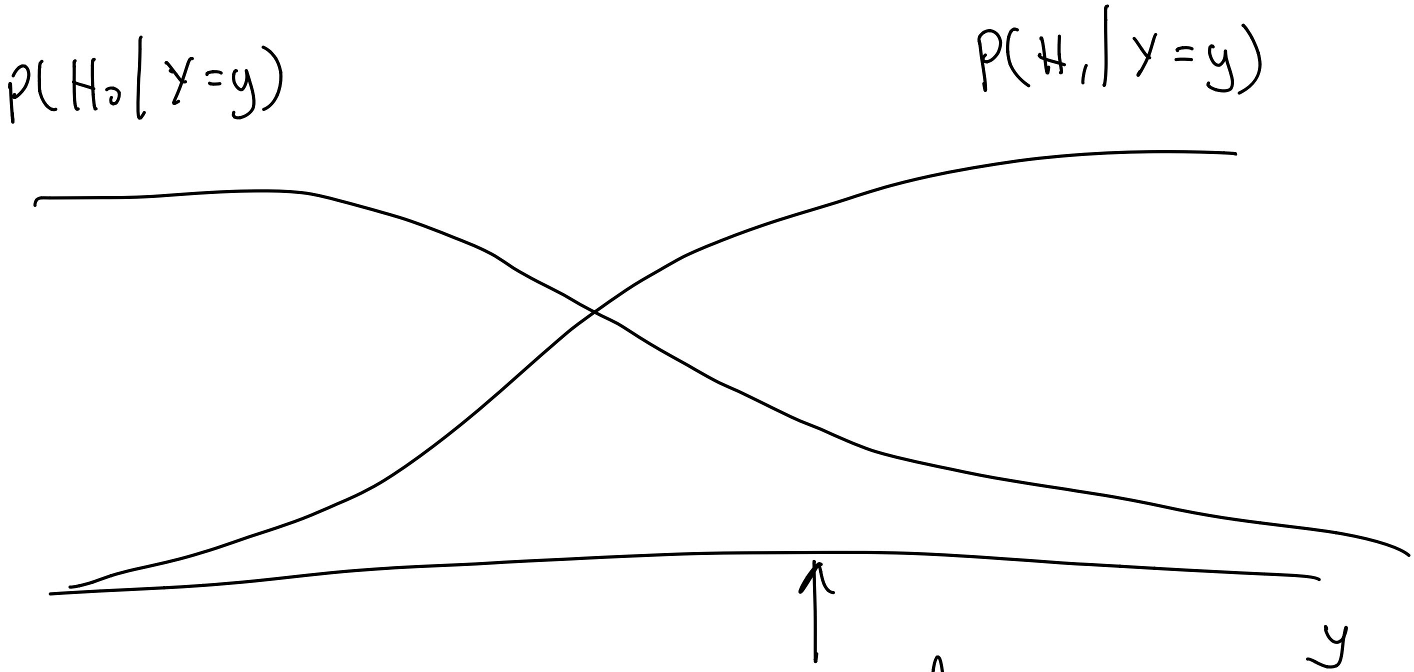
$$= C_{10}\, P(g(Y)=1|H_0)\, P(H_0) + C_{01}\, P(g(Y)=0|H_1)\, P(H_1)$$

$$= C_{10}\, P_F\, P(H_0) + C_{01}\, (1 - P_D)\, P(H_1)$$

Goal 3: maximize P_D with constraint that $P_F \le \alpha$



this example is not finished



observed value of y $\Rightarrow$ H_1 is more likely to be true

Classification

m hypotheses (not two)

ex: 3 colors of badges : R, G, B

wavelength in microns

- why not consistent?

fading from sun
inaccurate printer
light from bulb
measurement device

This example is not finished
- from T+T

X = color measured in microns

$X|B \sim N(436, 400)$ $\quad H_B \quad H_1$

$X|G \sim N(546, 400)$ $\quad H_G \quad H_2$

$X|R \sim N(700, 400)$ $\quad H_R \quad H_3$

Procedure: measure $X = x_0$

choose H_k to maximize $P(H_i \mid X = x_0)$

If all equally likely apriori, then

equiv to max $f_X(x_0 \mid H_i)$

ALG ch. 8.5.2

exa. 8.23

p 444

Gelfand 168-190 (corrected)

Y-G p 306

ACG p459

example 8.33

server allocation

mean square estimation

6.68

6.69

ch 6.5