HYPOTHESIS TESTING

MANY REAL-WORLD SITUATIONS INVOLVE OBSERVING A SET OF IID RANDOM SAMPLES $\{X_1, \dots X_n\}$ AND TESTING A HYPOTHESIS LIKE:

- THIS COIN IS FAIR.
- THIS MANUFACTURING PROCESS PRODUCES A BETTER BATTERY THAN OUR OLD PROCESS.
- THIS SIGNAL HAS A DIFFERENT DISTRIBUTION THAN THIS OTHER SIGNAL.

SIGNIFICANCE TESTING

WE HAVE A HYPOTHESIS CALLED THE <u>NULL HYPOTHESIS</u> H_0 THAT WE WANT TO EITHER ACCEPT AS TRUE OR REJECT AS FALSE.

<u>EX</u>: H_0: A COIN IS FAIR.

OBSERVE A COLLECTION OF IID BERNOULLI R.V.S.

$$\{X_1, \ldots, X_n\}$$

$$S_n = \sum_{i=1}^{n} X_i = \text{\# OF HEADS.}$$

S CLOSE TO 50

$\Rightarrow$ PROBABLY FAIR

S = 23, 78 $\Rightarrow$ PROBABLY NOT FAIR.

WHAT WE WANT:

A REGION R SO THAT

$S_n \in R \Rightarrow$ REJECT H_0

$S_n \notin R \Rightarrow$ ACCEPT H_0.

THIS IS A DECISION RULE.

TYPE I ERROR:

WE REJECTED H_0 WHEN H_0 WAS TRUE.

TYPE II ERROR:

WE ACCEPTED H_0 WHEN H_0 WAS FALSE.

WE ASSUME THAT UNDER H_0, WE KNOW THE PDF OF THE OBSERVATIONS.

$$f_X(x \mid H_0)$$

IF $\alpha =$ TYPE I ERROR

$$\int_R f_X(x \mid H_0)\, dx = \alpha$$

WE CAN'T SAY MUCH ABOUT THE TYPE II ERROR (YET).

$\alpha =$ SIGNIFICANCE LEVEL OF THE TEST / DECISION RULE.

WE DETERMINE R BASED ON A DESIRED VALUE OF α.

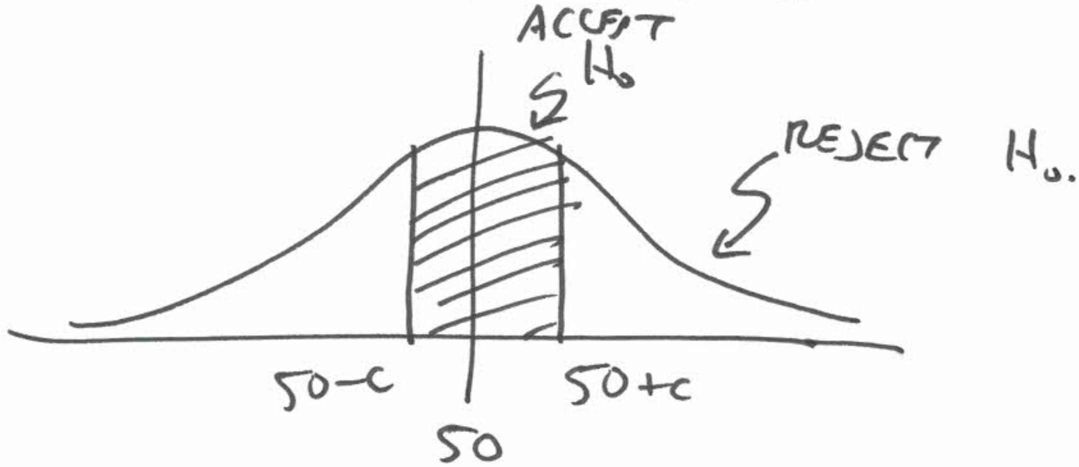
SMALLER $\alpha \Rightarrow$ MORE CONSERVATIVE RANGE FOR ACCEPTING H_0.

$\Rightarrow$ MORE LIKELY TO ACCEPT H_0 TO AVOID TYPE I MISTAKES.

EX H_0: A COIN IS FAIR $\left(p = \frac{1}{2}\right)$

S_{100} = # HEADS IN 100 FLIPS.

DETERMINE SIGNIFICANCE TEST AT A LEVEL OF 5%.



$R = \{$ INTEGERS OUTSIDE OF $[50-c,\ 50+c]$

$\alpha = 0.05$

$= P(\text{REJECT } H_0 \text{ WHEN } H_0 \text{ IS TRUE})$

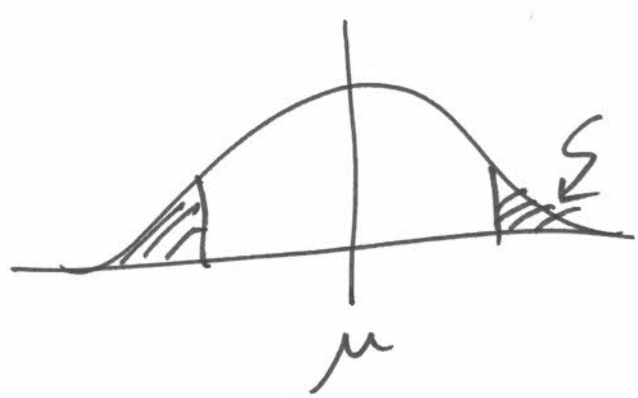
$= P(S_{100} \notin [50-c, 50+c] \mid H_0)$

$= P(|\underbrace{S_{100}}_{\text{SUM}} - 50| > c \mid H_0)$

$= P(|\underbrace{M_{100}}_{\text{MEAN}} - \frac{1}{2}| > \frac{c}{100} \mid H_0)$

CENTRAL LIMIT THEOREM

$$P\left(|M_n - \mu| > \frac{z\sigma}{\sqrt{n}}\right) \approx 2Q(z)$$



$$0.05 = 2Q(z)$$

$$Q(z) = 0.025$$

$$z = 1.96 \quad \text{(FROM TABLE)}$$

$$\frac{z\sigma}{\sqrt{n}} = \frac{c}{100}$$

FROM TABLE FOR BINOMIAL

$$\frac{(1.96)\left(\frac{1}{2}\right)}{\sqrt{100}} = \frac{c}{100}$$

$$c = 10.$$

$R =$ OUTSIDE OF $[40, 60]$.

SIMPLE HYPOTHESIS TESTING

IF WE CAN CHARACTERIZE THE ALTERNATIVE HYPOTHESIS H_1, THEN WE CAN PERFORM A HYPOTHESIS TEST,

e.g.

H_0: TARGET IS ABSENT

$$f_X(x \mid H_0)$$

H_1: TARGET IS PRESENT

$$f_X(x \mid H_1)$$

IN THIS CASE, TYPE I AND TYPE II ERRORS HAVE SPECIAL NAMES:

TYPE I: DECIDE H_1 WHEN H_0 IS TRUE. <u>FALSE ALARM</u>

TYPE II: DECIDE H_0 WHEN H_1 IS TRUE. <u>MISS</u>

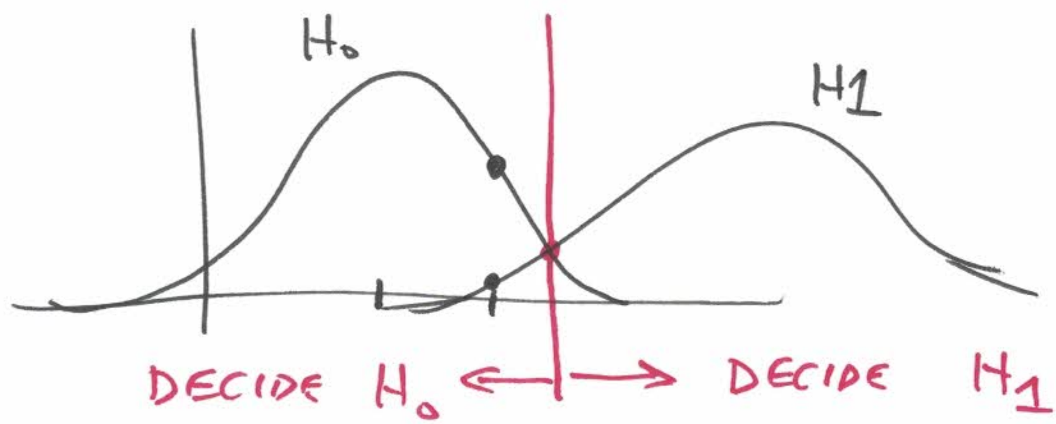
WE PREVIOUSLY TALKED ABOUT THE <u>MAXIMUM LIKELIHOOD</u> TEST FOR A 2-HYPOTHESIS CASE,

$$f_X(x \mid H_1) > f_X(x \mid H_0) \Rightarrow$$

CHOOSE H_1.

$$f_X(x \mid H_0) > f_X(x \mid H_1) \Rightarrow$$

CHOOSE H_0



MORE CONCISELY,

$$\frac{f_X(x \mid H_1)}{f_X(x \mid H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} 1$$

LIKELIHOOD RATIO

DEPENDS ON α.

ALSO SOMETHING ~~VERY~~ CALLED NEYMAN–PEARSON TEST:

MINIMIZE THE MISS (TYPE II) PROBABILITY GIVEN ~~THE~~ A SPECIFIED VALUE OF THE FALSE ALARM (TYPE I) PROBABILITY.

BAYESIAN (MAXIMUM A POSTERIORI) HYPOTHESIS TEST:

WE HAVE THE LIKELIHOODS

$$f_X(x \mid H_0), \quad f_X(x \mid H_1)$$

AND THE PRIOR PROBABILITIES

$$P(H_0), \quad P(H_1)$$

AND ASSIGN COSTS:

C_{00} = H_0 TRUE, DECIDE H_0

C_{01} = H_0 TRUE, DECIDE H_1 TYPE I FALSE ALARM

C_{10} = H_1 TRUE, DECIDE H_0 TYPE II MISS

C_{11} = H_1 TRUE, DECIDE H_1.

WE WANT A DECISION RULE THAT MINIMIZES THE EXPECTED VALUE OF THE COST.

$$C = C_{00}\, P(\text{CHOOSE } H_0 \mid H_0 \text{ TRUE})\, P(H_0)$$

$$+ C_{01}\, P(\text{CHOOSE } H_1 \mid H_0)\, P(H_0)$$

$$+ C_{10}\, P(\text{CHOOSE } H_0 \mid H_1)\, P(H_1)$$

$$+ C_{11}\, P(\text{CHOOSE } H_1 \mid H_1)\, P(H_1)$$

WE CAN SHOW THAT THE DECISION RULE THAT MINIMIZES THE EXPECTED COST IS:

$$\frac{f_X(x \mid H_1)}{f_X(x \mid H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{P(H_0)(C_{01} - C_{00})}{P(H_1)(C_{10} - C_{11})}$$

Ex: $C_{00} = C_{11} = 0$
$C_{01} = C_{10} = 1$

$$\frac{f_X(x \mid H_1)}{f_X(x \mid H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{P(H_0)}{P(H_1)}$$

SAME AS MAXIMUM A POSTERIORI RULE:

CHOOSE LARGER OF

$$P(H_0 \mid x) \quad \text{AND} \quad P(H_1 \mid x)$$

EX BINARY COMMUNICATION SYSTEM.

SYSTEM TRANSMITS EITHER $+1$ OR -1.

RECEIVED SIGNAL = TRANSMITTED BIT + $N(0,1)$ GAUSSIAN NOISE

MEAN 0

$\sigma^2 = 1$.

GAUSSIAN (NORMAL)

SAY WE TRANSMIT THE SAME BIT n TIMES AND RECEIVE $X_1, \ldots X_n$

DETERMINE THE MAP RECEIVER / DECISION RULE.

H_0: n $\quad -1$ BITS

H_1: n $\quad +1$ BITS.

$$f_X(x \mid H_0) = f_X(x_1, \dots x_n \mid H_0)$$

$$= f_X(x_1 \mid H_0) \cdot f_X(x_2 \mid H_0) \cdot \cdots$$

$$= \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2} \sum_{i=1}^{n} (x_i - {}^-1)^2}$$

$$f_X(x \mid H_1) = \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2} \sum_{i=1}^{n} (x_i - 1)^2}$$

THE LIKELIHOOD RATIO IS

$$\frac{f_X(x \mid H_1)}{f_X(x \mid H_0)} = e^{-\frac{1}{2}\left(\sum_{i=1}^{n}\left[(x_i-1)^2-(x_i+1)^2\right]\right)}$$

$$= e^{-\frac{1}{2}\sum_{i=1}^{n} -4x_i} = e^{2\sum_{i=1}^{n} x_i}$$

$$= e^{2S_n}$$

$$e^{2S_n} \underset{H_0}{\overset{H_1}{\gtrless}} \tau$$

THRESHOLD

WE CAN EQUIVALENTLY THINK OF THE TEST AS OPERATING ON THE log LIKELIHOOD.

$$2S_n \underset{H_0}{\overset{H_1}{\gtrless}} \log\left(\frac{P(H_0)}{P(H_1)} \cdot \frac{C_{01}-C_{00}}{C_{10}-C_{11}}\right)$$

$$= \log P(H_0) - \log P(H_1)$$

$$+ \log(\text{COSTS})$$

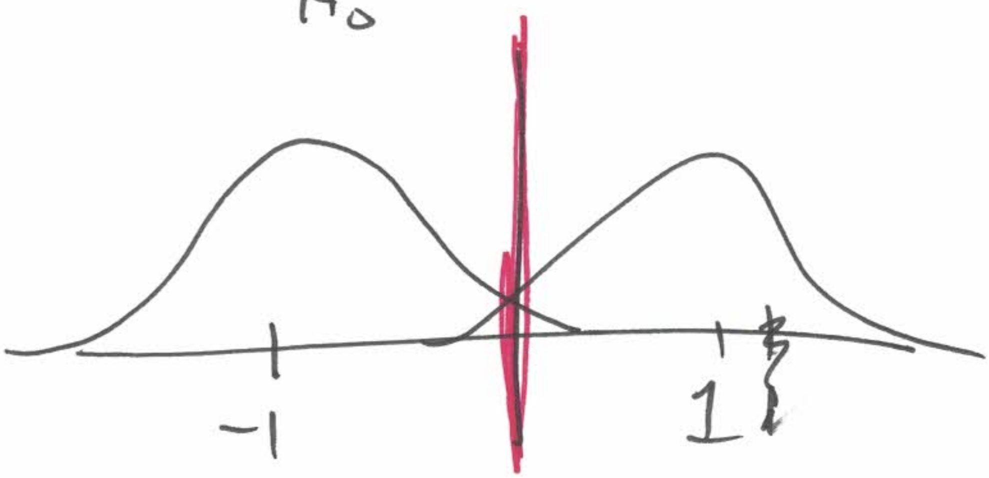
IF WE USE THE USUAL COSTS

$$(C_{00} = C_{11} = 0, \quad C_{10} = C_{01} = C)$$

$$2S_n \underset{H_0}{\overset{H_1}{\gtrless}} \log P(H_0) - \log P(H_1)$$

IF INPUTS ARE EQUALLY PROBABLE, $P(H_0) = P(H_1) = \frac{1}{2}$

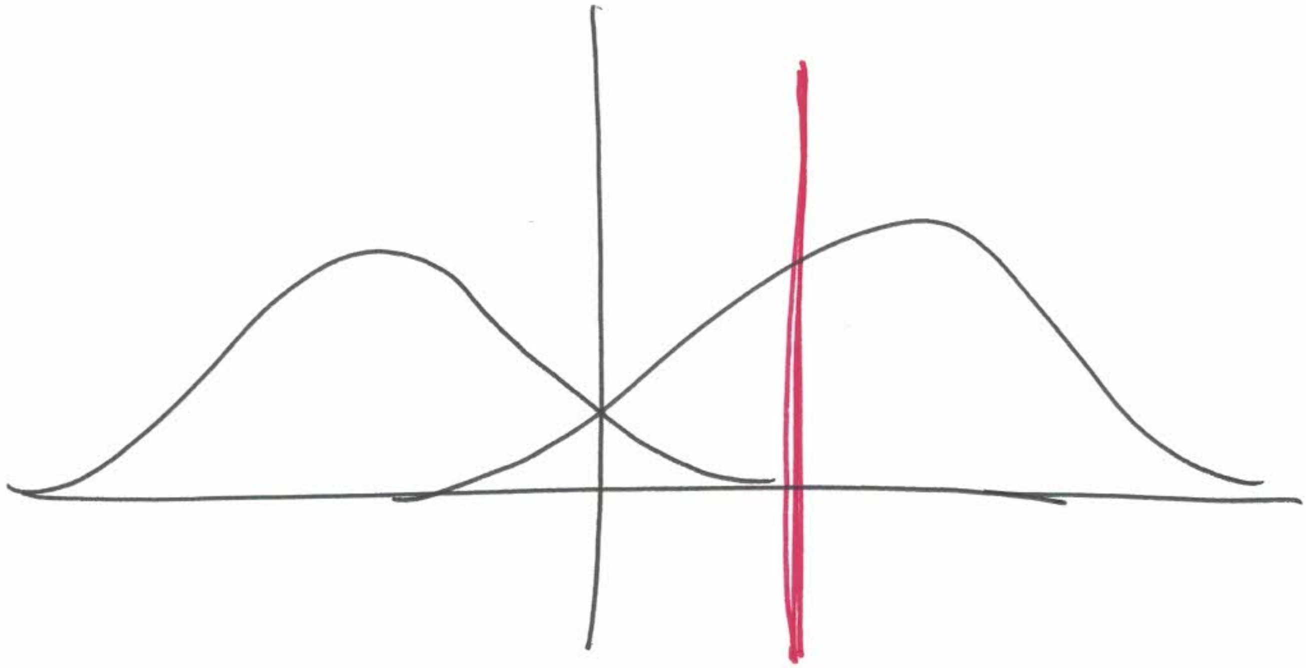
$$S_n \underset{H_0}{\overset{H_1}{\gtrless}} 0$$



IF $P(H_0) \gg P(H_1)$

-1 IS MORE LIKELY.

$$\log P(H_0) > \log P(H_1)$$



S_n WOULD NEED TO BE > 0 TO ACCEPT H_1.

NOTE:

$$\frac{1}{n}\sum_{i=1}^{n} X_i \underset{H_0}{\overset{H_1}{\gtrless}} \frac{1}{2n}\log\frac{P(H_0)}{P(H_1)}$$

AS $n \to \infty$, THRESHOLD GOES BACK TO 0.

COMPOSITE HYPOTHESIS

TEST A NULL HYPOTHESIS AGAINST A FAMILY OF ALTERNATE HYPOTHESIS, e.g.,

H_0: MEAN OF DISTRIBUTION (θ_0) IS 0

H_1: MEAN OF DISTRIBUTION (θ) IS > 0.

IN THIS CASE, MEASURE THE POWER OF A TEST.

SPECIFY TYPE I ERROR α.

CREATE A CURVE

TYPE II ERROR

$$1 - \beta(\theta) =$$

THE PROBABILITY OF REJECTING H_0 WHEN THE TRUE PARAMETER IN H_1 IS θ.

WE WANT $1 - \beta(\theta) \approx 0$ WHEN $\theta \approx \theta_0$, BIGGER AS θ GETS FAR FROM θ_0.