PROB CLASS 24

M 4/11/22

CHAP 8 STATISTICS

A POPULATION OF WIDGETS IS BEING GENERATED BY RANDOM DISTRIBUTION WITH UNKNOWN PARAMS.

E.G. GENERATE RANDOM UNIFORM VARS X_i IN RANGE $[L, H]$

WE KNOW IT'S UNIFORM BUT DON'T KNOW L, H.

OBSERVE 100 X_i

$X_1 \dots X_{100}$

ESTIMATE L, H.

$X_i \sim U(L, H)$

SOME x_i: 3, 10, 7, 20, 5, 1, 7, 14

IS RANGE [3, 10] POSSIBLE?

GOOD VALUE FOR $L = 1$.

L CANNOT BE > 1

$H \geq 20$

COMMONSENSE GUESS

$L = MIN(X_i)$

$H = MAX(X_i)$

NOT BEST PERHAPS BUT PRETTY GOOD.

Q1: WHAT ARE UNKNOWN PARAMS.

Q2: IS MY ASSUMED DISTN CORRECT?

X_n: 1, 5, 2, 1, 7, 3, 2, 1

PERHAPS NOT UNIFORM.

Q3: GUESS THE PARAM.

IS THIS GUESS REASONABLE?

THERE IS A PARTICULAR WAY THESE QUESTIONS ARE SOMETIMES WORDED.

THIS IS FOR YES/NO Q.

HAVE A COIN. MAYBE IT'S FAIR. MAYBE NO. TOSS IT N=100 TIMES

X_n = 0, 1, 0 ... SEE 60 H 40 T

IS IT FAIR?

IF IT IS FAIR, WHAT'S PROBABILITY OF SEEING THIS OUTCOME?

P(SEEING 60 H) BETTER OUTCOME TO ANALYSE.

RUN WITH THIS COIN EXAMPLE.

EXP: TOSS 100 TIMES.

R.V. OBSERVE: X: # HEADS.

$\mu = E[X] = 50 = np$ $(100 \cdot 1/2)$

$\sigma = STD[X] = \sqrt{npq} = \sqrt{100 \cdot \frac{1}{2} \cdot \frac{1}{2}} = 5$

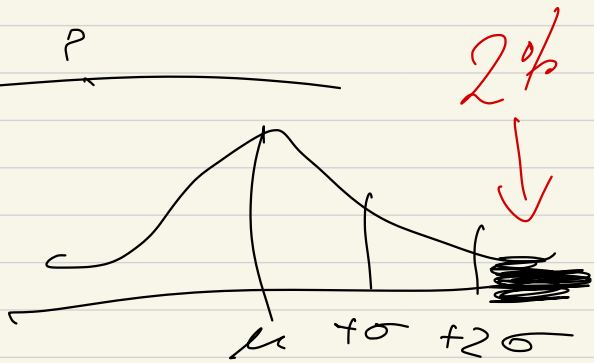
$P[\mu - \sigma < X < \mu + \sigma] \approx 2/3$

$P[45 < X < 55] \approx 2/3$

$= \Phi(1) - \Phi(-1)$?

$P[X > 60]$

$\mu + 2\sigma$



IF THIS COIN IS FAIR AND WE TOSS IT 100 TIMES, PROB WE'LL SEE ≥60 HEADS IS 2%

THAT'S MATH.

IS IT FAIR? THAT'S POLICY.

"COIN IS FAIR" | NULL HYPOTHESIS |

"COIN NOT FAIR" | ALTERNATIVE HYP |

YOU HAVE FREEDOM IN SELECTING ALTERNATIVE. I MIGHT HAVE ASKED P[#HEADS IS MORE THAN 10 OFF FAIR EITHER WAY:]

$P[|\#-50| > 10)$: >60 or <40

HAVE COIN WITH UNKOWN π

TOSS IT N TIMES, SEE X HEADS. ESTIMATE π.

$\pi = \frac{X}{N}$ IS GOOD.

IS BEST.

WE WANT **ESTIMATOR FUNCTIONS** FOR DISTRIBUTION.

§8.2 PARAMETER ESTIMATION

SOME ESTIMATORS ARE BETTER.

WHAT'S BETTER?

DISTRIBUTION IS $N(\mu, 1)$

GAUSSIAN UNKNOWN μ $\sigma = 1$

SAMPLE 100 X_i

WANT TO ESTIMATE μ MEAN

CHOICES FOR ESTIMATOR FUNCTION

C1. $\frac{\Sigma X_i}{N}$ SAMPLE MEAN

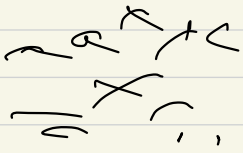
C2: MEDIAN(X_i)

C3: $\frac{MAX(X_i) \& MIN(X_i)}{2}$

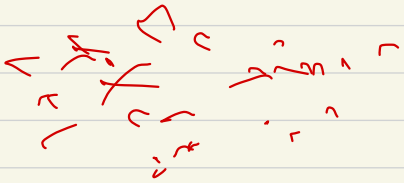
THESE ESTIMATORS ARE RANDOM FUNCTIONS OF POPULATION.

THEY HAVE MEANS, STD THEMSELVES.

THE STD MEASURES HOW MUCH THE ESTIMATOR JUMPS AROUND WITH DIFFERENT SAMPLES.

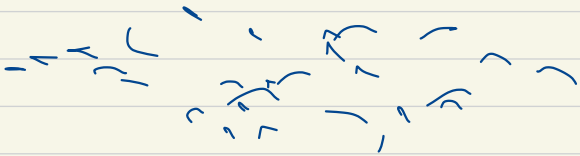


MEANS



MEDIANS

$\frac{\max + \min}{2}$



ESTIMATOR WHOSE OWN STD IS SMALL IS BETTER.

IN THIS CASE MEAN IS BEST.

BUT WHAT IF $\Delta(STN)$ IS NOT GAUSSIAN?

~~THE~~ MAYBE POISSON.

BIASSED LIKE

THEN MEDIAN MIGHT BE BETTER.