PROB CLASS 26
M 4/18/22

PROPERTIES OF ESTIMATORS p416

- YOU HAVE A POPULATION
- YOU DRAW N RANDOM VARS X_i
- YOU WANT TO INFER UNKNOWN PARAMS OF THAT DISTRIBUTION.

eg. μ UNKNOWN, MAYBE σ.

$$\bar{X}_n = \frac{\sum x_i}{N}$$

IS A GOOD ESTIMATOR OF μ.
IT'S UNBIASED.

SUPPOSE σ^2 IS UNKNOWN.

OBVIOUS ESTIMATOR IS

$$\hat{S}^2 = \frac{1}{N} \sum (X_i - \bar{X}_N)^2$$

PROBLEM: IT'S BIASED.

$$E[\hat{S}^2] \neq \sigma^2$$

$$\hat{S}^2 = \frac{1}{N} \sum (X_i - \bar{X})^2$$

$-\mu + \mu$

$$= \frac{1}{N} \sum \left[(X_i - \mu) + (\mu - \bar{X}) \right]^2$$

$$= \frac{1}{N} \sum \left[(X_i - \mu)^2 + 2(X_i - \mu)(\mu - \bar{X}) + (\mu - \bar{X})^2 \right]$$

$$= \frac{1}{N} \left[\sum (X_i - \mu)^2 + 2(\mu - \bar{X}_N) \sum (X_i - \mu) + N(\mu - \bar{X})^2 \right]$$

$n\bar{X} - n\mu$

$$= \frac{1}{N}\sum (x_i-\mu)^2 + \underbrace{\underbrace{2(\mu-\bar{x})(\bar{x}-\mu)}_{-2(\mu-\bar{x})^2} + (\mu-\bar{x})^2}_{-(\mu-\bar{x})^2}$$

$$\hat{S}^2 = \frac{1}{N}\sum (x_i-\mu)^2 - (\bar{x}-\mu)^2$$

$$E[S^2] = \underbrace{\sum E[(x_i-\mu)^2]}_{\sigma_x^2} - \underbrace{E[(\bar{x}-\mu)^2]}_{\frac{\sigma_x^2}{N}}$$

(SKIPPING)

$$= \frac{N-1}{N}\sigma_x^2$$

THIS OBVIOUS ESTIMATOR IS TOO SMALL.

UNBIASED ESTIMATOR

$$\hat{\sigma}^2 = \frac{1}{N-1}\sum (x_i-\bar{x}_N)^2$$

p430 §8.4 CONFIDENCE INTERVALS.

- WE HAVE A POPULATION WITH UNKNOWN μ
- TAKE N OBSERVATIONS X_i.
- COMPUTE ESTIMATOR OF μ

$$\overline{X}_n = \frac{1}{n} \sum X_i$$

WHAT DOES THIS SAY ABOUT REAL μ?

WE'LL COMPUTE c

$$\mathbb{P}[\overline{X}_n - c \le \mu \le \overline{X}_n + c]$$

$$= .95 \quad \text{(SAY)}$$

SEVERAL DIFFERENT CASES.

1. POP IS GAUSSIAN, WE KNOW σ, DON'T KNOW μ.

$$X_n \sim \mathcal{N}(\mu, 1) \quad \text{- GAUSSIAN}$$

$$\overline{X}_n = \frac{\sum X_i}{N} : \mathcal{N}(\mu, \tfrac{1}{\sqrt{N}})$$

USE TABLES FOR CDF OF GAUSSIAN -

LET Y BE $N(0,1)$

$P(Y > 2) = 4\%$ (?)

$P(|Y| > 2) \sim 8\%$ ← APPROX ONLY.

$$P(\bar{X}_L - c \le \mu \le \bar{X}_L + c) = 95\%$$

$$= P(\mu - c \le \bar{X}_L \le \mu + c)$$

$$\bar{X}_L \sim N\left(\mu, \frac{1}{\sqrt{N}}\right)$$

Z VALUE. $c\sqrt{N}$ - LOOK INTO TABLE. GET ANSWER.

CASE 2 WE DON'T KNOW μ OR σ.

USE SOMETHING CALLED A T-TEST.

IN PREVIOUS CASE, I NORMALIZED X_L

$$\frac{X_L - \bar{X}_L}{\sigma/\sqrt{N}}$$ ← THIS IS $N(0,1)$.

BUT NOW I DON'T KNOW σ EITHER.

(STANDARDIZE MY OBSERVATIONS

$$T = \frac{\bar{X}_N - \mu}{\hat{\sigma}_N / \sqrt{N}}$$

(P433)

THIS IS NOT GAUSSIAN,

IT'S A STUDENT-T DISTN -

THERE ARE TABLES FOR IT (OR BUILTIN FUNCTIONS)

YOU CAN FIND CONFIDENCE INTERVAL -

FOR $N \geq 10$ GAUSSIAN IS EXCELLENT

5 PRETTY GOOD.

SEE PLOT COMPARING GAUSSIAN AND T

PAGE 434

CASE 3 PAGE 435

DISTN IS NOT GAUSSIAN.

E.G. DEVICE LIFETIME.

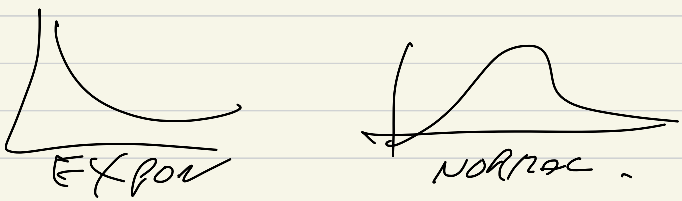
IT'S AN EXPONENTIAL R.V.

WE DON'T KNOW MEAN.

OR $\frac{1}{2}$ LIFE OF RADIOACTIVE ELEMENT.

OR # COSMIC RAYS.

OR DEFECTS ON CHIP.



NICE PROPERTY: SUM OF R.V. FROM EXPONENTIAL DIST STARTS TO LOOK GAUSSIAN QUICKLY.

SO DO SOMETHING CALLED BATCH MEAN.

SAY $N = 200$. GROUP INTO 10 BATCHES OF 20 SAMPLES EACH

MEAN OF 20 EXPONENTIAL R.V. IS GAUSSIAN.

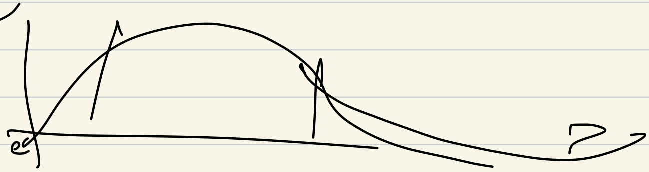
SO: COMPUTE CONFIDENCE INTERVAL FOR BATCH MEAN.

CASE 4: CONFIDENCE INTERVAL FOR σ^2

OUR ESTIMATOR $\hat{S}^2 = \frac{1}{N-1} \sum (\bar{X} - X_i)^2$

S^2 IS NOT GAUSSIAN

EG, $S^2 \geq 0$



IT'S A CHI-SQUARED DIST.

§8.7 p462

TEST THE FIT OF A DIST

I DON'T EVEN KNOW THE DISTN BUT THINK IT'S WHATEVER. SAY GAUSSIAN.

TAKE N=100 OBSERVATIONS.

CHI-SQUARE TEST WILL HELP.

ALSO USEFUL TO TELL IF DIE IS FAIR

TOSS A DIE 60 TIMES.

TOP FACE	# TIMES	EXPECTED # TIMES
1	15	10
2	8	10
3	12	10
4	9	10
5	10	10
6	6	10
	60	60

~~k~~ $k=6$ POSSIBLE OUTCOMES.

$N=60$ OBSERVATIONS.

N_i = # TIMES WE SAW i 15, 8, 12, 9, 10, 6

M_i = EXPECTED # $=10$

COMPUTE

$$D^2 = \sum_{i=1}^{k} \frac{(N_i - M_i)^2}{M_i}$$

THIS IS A χ^2 (CHI-SQUARED)

DISTN WITH N D.F.

THIS WORKS BETTER IF ALL M_i SIMILAR

AND $M_i \geq 5$.

TEST HYPOTHESIS THAT DIST IS UNIFORM.