STAT 231.

Roadmap

- Statistical modelling
- The Theory of Estimation:
  - Method of Maximum Likelihood.
  - Likelihood function, and

    the MLE

    Maximum Likelihood Estimate.
  - Relative Likelihood function

STATISTICAL MODEL

θ There is some parameter (attribute of the population) that we are interested in.

→ Greek letters; π ; γ, θ, σ.

Collect a sample (independent) from this population

$\{y_1, \ldots y_n\}$ → SAMPLE.

Insight: The sample values can be thought of as outcomes of a random experiment.

A statistical model is the "identification" of the random variable from which y_i is drawn $i=1,\dots n$

$$\boxed{Y_i \sim f(y_i;\theta)} \quad i=1,\dots n.$$

f = distribution of the random variable.

(ii) The unknown population attribute is a parameter in the statistical model constructed.

Example

Canadian
%age of seniors who has an Instagram Account. Sample.

200 seniors, $y = \#$ of successes
$= 40$

(i) What model is appropriate here?

(ii) What is our variate of interest?

Attribute of interest ? PARAMETER?

(1) • Population Mean μ

(2) • Population Proportion π

(3) Population Variance σ

• What distribution is appropriate?

(i) • Binomial.

(ii) • Poisson

(iii) • Gaussian

$$Y \sim Bin(200, \pi) \quad \text{--- (1)}$$

π = Probability of success for each trial.

40 = y = outcome of the experiment in (1).

$\{y_1, \ldots y_n\}$

y_i = # of accidents observed in day i

Parameter of interest = POP. MEAN $= \mu$.

$$Y_i \sim \text{Poi}(\mu) \quad \underline{\text{indep}} \quad i = 1, \ldots n.$$

Objective: To find our "best guess" of μ based on $\{y_1, \ldots y_n\}$

How does one estimate unknown parameters?

θ = unknown parameter.

$\hat{\theta}$ = a # constructed from our data set $f(y_1, \ldots y_n)$ which "estimates" θ.

Example

A coin is tossed.

$\pi = P(H) =$ unknown.

$$\pi = \begin{cases} 1/3 \\ 2/3 \end{cases}$$

There are no other possibilities

Experiment: Toss the coin 20 times

Record the # of heads: y

$y = 4$

It seems like $\pi = 1/3$ is "MORE LIKELY" given the sample.

$P(\text{obs. my sample} : \pi = 1/3)$ ——— ①

$$= {}^{20}C_4 \left(\frac{1}{3}\right)^4 \left(2/3\right)^{16}$$

$P(\text{obs. my sample} : \pi = 2/3)$ ——— ②

$$= {}^{20}C_4 \left(\frac{2}{3}\right)^4 \left(\frac{1}{3}\right)^{16}$$

Since $1 > 2$, $\pi = 1/3$ is "MORE LIKELY" than $\pi = 2/3$.

Example

π = Proportion of Trump supporters standing in line at Walmart

$\{y_1, \ldots, y_n\}$

$n = 10$

Y Y N N N N N N N Y

$\pi \quad \pi \quad (1-\pi) \quad \ldots \quad \pi$

What is $\hat{\pi}$? MLE for π.

$L(\pi) = \pi^3 (1-\pi)^7$

We will choose the value of π that maximizes this probability

SOLUTION: MAXIMUM LIKELIHOOD ESTIMATE

$$L(\pi) = \pi^3(1-\pi)^7$$

Log-likelihood function (base e)

$$\ell(\pi) = 3 \ln \pi + 7 \ln(1-\pi)$$

$$d\ell/d\pi = 0 \Rightarrow \frac{3}{\pi} - \frac{7}{1-\pi} = 0$$

$$\boxed{\hat{\pi} = 3/10}$$

Example.

Objective: To estimate μ (POPULATION AVERAGE) = of accidents on HWY 401.

$y_i = \{2, 1, 0, 3, 1, 5, 7\}$

$Y_i \sim Poi(\mu) \quad i = 1, \ldots n.$

indep.

$$L(\mu) = \frac{e^{-\mu}\mu^{2}}{2!} \cdot \frac{e^{-\mu}\mu^{1}}{1!} \cdots \frac{e^{-\mu}\mu^{7}}{7!}$$

$$L(\mu) = \frac{e^{-7\mu} \cdot \mu^{19}}{2!\, 1! \ldots 7!}$$

$$l(\pi) = -7\mu + 19 \ln \mu - \ln(2!\, 1! \ldots 7!)$$

$$dl/d\pi = -7 + \frac{19}{\mu} = 0$$

$$\boxed{\hat{\mu} = 19/7}$$