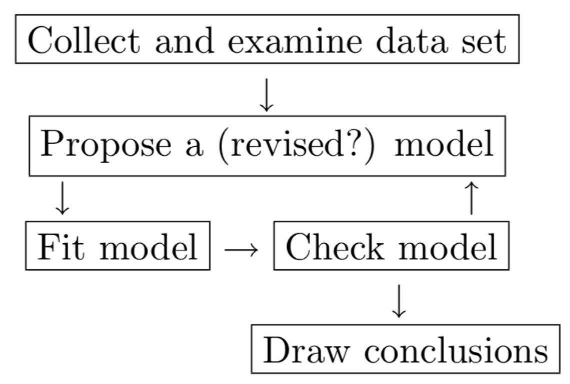

Point estimate: the value of a function of observed data {y1, y2, … y_n}

- Mean, standard deviation
- Does not require any unknowns to calculate

Likelihood Functions

Let Y be a discrete random variable; let θ be some characteristic of a population described by Y. We are trying to estimate θ.

The *likelihood function* of θ, L(θ), is defined as P(Y = y; θ)

- The probability of observing Y=y while trying to estimate θ

Ex. Suppose on January 8, 2024, I sampled 1000 Canadians and asked whether they drink coffee on a regular basis. Of these individuals, 684 said yes. How can we model the likelihood function based on these data? What is our maximum likelihood estimate for the proportion of Canadians who drink coffee on a regular basis?

Let Y be the number of people in a sample of n that have a characteristic of interest

Let θ be the proportion of people that have that characteristic of interest; in this case, drinks coffee

The probability that Y = y people drink coffee follows a binomial distribution

$$L(\theta) = L(\theta; y) = \binom{1000}{684}\theta^{684}(1-\theta)^{316}$$

Maximizing L(θ) is an optimization problem (find point when derivative is 0)

The value of θ that maximizes L(θ) is the *maximum likelihood estimate*

In this case, the most likely value of θ is 684/1000, because of course it is. MLE notation: $\hat{\theta}$ = 684/1000.

And, plugging in θ=0.684 into the above equation does indeed maximize L(θ).

Conversely, values of θ further away from the measured proportion of 0.684 will have smaller L(θ).

Ex. How would you interpret L(θ_1) / L(θ_2) = 0.8?

Rearranging, we have L(θ_1) = 0.8 • L(θ_2)

This means that θ_2 is more likely

Relative Likelihood Function

- The **relative likelihood function** is defined as follows:

$$R(\theta) = \frac{L(\theta)}{L(\hat{\theta})}, \; \theta \in \Omega$$

- Note: $0 \le R(\theta) \le 1$ for all θ
- The **log likelihood function** is defined as follows:

$$\log L(\theta) = \ln L(\theta) = l(\theta)$$

- Sometimes it's easier working with the log likelihood when trying to calculate the MLE, i.e. we can use $\frac{dl(\theta)}{d\theta}$ rather than $\frac{dL(\theta)}{d\theta}$

→ $L(\hat{\theta})$: MLE

so this is the likelihood of θ relative to the highest value $\hat{\theta}$

Going back to the coffee example:

$$L(\theta) = \binom{1000}{684} \theta^{684} (1-\theta)^{316}$$

$$\Rightarrow l(\theta) = \ln(L(\theta)) = \ln\binom{1000}{684} + 684\ln(\theta) + 316\ln(1-\theta)$$

Getting the derivative, we have $\hat{\theta}$=0.684

Ex. Let Y_i be an indicator random variable indicating whether someone has tested positive for COVID.

Solving for the likelihood function:

L(θ) = P(Y = y; θ)

5. In modelling the number of transactions of a certain type received by a central computer for a company with many on-line terminals the Poisson distribution can be used. If the transactions arrive at random at the rate of θ per minute then the probability of y transactions in a time interval of length t minutes is

$$P(Y=y;\theta) = f(y;\theta) = \frac{(\theta t)^y}{y!} e^{-\theta t} \quad \text{for } y = 0, 1, \ldots \text{ and } \theta \geq 0$$

(a) The numbers of transactions received in 10 separate one minute intervals were

8 3 2 4 5 3 6 5 4 1

Find the likelihood function $L(\theta)$, the log likelihood function $l(\theta)$, and the maximum likelihood estimate $\hat{\theta}$ for these data.

$$L(\theta) = P(Y_1 = 8)\, P(Y_2 = 3) \cdots P(Y_{10} = 1) \qquad \text{Sum}: 41$$

$$= \frac{\theta^{41} e^{-10\theta}}{8!\,3!\cdots 1!}$$

$$= \theta^{41} e^{-10\theta} \cdot \left(\prod_{i=1}^{10} y_i!\right)$$

$$l(\theta) = 41\log(\theta) - 10\theta + \sum_{i=1}^{10} y_i!$$

$$\frac{dl}{d\theta} = \frac{41}{\theta} - 10 = 0$$

$$\frac{41}{\theta} = 10$$

$$10\theta = 41$$
$$\hat{\theta} = 4.1$$

Ex. Suppose that the random variable Y represents the number of persons infected with the human immunodeficiency virus (HIV) in a randomly selected group of n persons.

Y is binomial, so

$$f(y;\theta) = \binom{n}{y} \theta^{y} (1-\theta)^{n-y}$$

$\Rightarrow L(\theta) = \theta^y (1-\theta)^{n-y}$ remove $\binom{n}{y}$ since it is constant wrt θ

$\Rightarrow \hat{\theta} = \frac{y}{n}$

Alternatively, we could use indicator random variables, with Y_i corresponding to whether person i tests positive for HIV. Then:

$$L(\theta) = \prod_{i=1}^{n} f(y_i ; \theta)$$

$\mapsto n \to 1 : Y_i \sim Bin(1, \theta)$

$$= \prod_{i=1}^{n} \theta^{y_i} (1-\theta)^{1-y_i}$$

$$= \theta^{\sum_{i=1}^{n} y_i} (1-\theta)^{\sum_{i=1}^{n} 1 - y_i}$$

Note that $\sum_{i=1}^{n} y_i = y$ since y_i is an indicator RV

$\Rightarrow 0 + 1 + 0 + 0 + \cdots$

$$= \theta^y (1-\theta)^{n-y}$$

$\Rightarrow \hat{\theta} = \frac{y}{n}$ ✓