Chapter 2 :- Sampling and Repeated Trials

- Typically one repeats an experiment many times. (independently)

- Viewed as sampling from the population

  E.g:- A book manufacture may sample 10 books from production line every day & see how many defects are found. She will be able to assess the quality of production.

- So far we have seen experiments

Experiment	Sample Space	Event	Probability
Roll a die	$\{1,2,3,4,5,6\}$	Six appears	$P(\{\text{Six appears}\}) = 1/6$

2.1 Bernoulli trials [named after James Bernoulli (1654-1705)]

- Mathematical framework to do many trials
- Each trial - p - Probability of success. $0 \le p \le 1$
- Notation: Bernoulli(p).

$$S = \{\text{success}, \text{failure}\} \qquad \mathcal{F} = \mathcal{P}(S)$$

$$\mathbb{P}(\{\text{success}\}) = p \quad , \quad \mathbb{P}(\{\text{failure}\}) = 1-p$$

2.1.1 Example :- Roll a dice two times.

Q:- Probability that we observe exactly one six in two trials?

Approach 1 :- S - Thirty six possible outcomes.
E - exactly one six appears

Treat the experiment as Equally likely outcomes and compute $P(E) = \frac{|E|}{|S|} = \ldots$

Approach 2 :- Our concern is only with six appearing

- view two rolls as two Bernoulli$(1/6)$ trials.

Bernoulli$(1/6)$

$$\begin{cases} S = \{\text{success}, \text{failure}\} \\ \\ P(\{\text{success}\}) = 1/6 \end{cases}$$

success = {6 appears}

failure = {6 does not appear}

$$\mathbb{P}(\{\text{Success in Roll 1}, \text{Success in Roll 2}\})$$

Independence

$$= \mathbb{P}(\{\text{Success in Roll 1}\})\, P(\{\text{Success in Roll 2}\})$$

Each is Bernoulli($1/6$) trial

$$= 1/6 \cdot 1/6 = \frac{1}{36}$$

Question:- $\mathbb{P}(\text{one six appears in two trials})$

Bernoulli trials

$$= \mathbb{P}\left(\left\{\begin{matrix}\text{success,} & \text{failure}\\ \text{in} & \text{in}\\ \text{Roll 1} & \text{Roll 2}\end{matrix}\right\} \cup \left\{\begin{matrix}\text{failure,} & \text{Success}\\ \text{in} & \text{in}\\ \text{Roll 1} & \text{Roll 2}\end{matrix}\right\}\right)$$

Disjoint events

$$= \mathbb{P}\left(\left\{\begin{matrix}\text{Success} & \text{failure}\\ \text{in} & \text{in}\\ \text{Roll 1,} & \text{Roll 2}\end{matrix}\right\}\right) + \mathbb{P}\left(\left\{\begin{matrix}\text{failure} & \text{Success}\\ \text{in} & \text{in}\\ \text{Roll 1,} & \text{Roll 2}\end{matrix}\right\}\right)$$

Independence

$$= P\left(\left\{\begin{matrix}\text{Success}\\ \text{in}\\ \text{Roll 1}\end{matrix}\right\}\right) \mathbb{P}\left(\left\{\begin{matrix}\text{failure}\\ \text{in}\\ \text{Roll 2}\end{matrix}\right\}\right) + \mathbb{P}\left(\left\{\begin{matrix}\text{failure}\\ \text{in Roll 1}\end{matrix}\right\}\right) \mathbb{P}\left(\left\{\begin{matrix}\text{Success}\\ \text{in}\\ \text{Roll 2}\end{matrix}\right\}\right)$$

Bernoulli tri

$$= 1/6 \cdot 5/6 + 5/6 \cdot 1/6 = \frac{10}{36}$$

Example 2.1.2 :- let $n \geq 1$ given. We perform n independent Bernoulli(p) trials. [ie. Perform an experiment with $\langle$ success / failure]
$\mathbb{P}(\{success\}) = p$
. Each trial is independent

ⓐ What is the Probability that there are k - successes ?

Ans :- let outcomes is each trial i $1 \leq i \leq n$ be denoted by $\{\omega_i\}$. The n-trial outcome - as $\omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$

$$A_i = \{ i^{th} \text{ trial outcome is a success} \}$$

$$P(\{n \text{ successes}\}) = P(A_1 \cap A_2 \ldots \cap A_n) \underset{\text{Independent}}{=} \prod_{i=1}^{n} \mathbb{P}(A_i) = p^n$$

let E_i be any event concerned with i^{th} trial

$$P(\{\omega_1, \ldots, \omega_n\}) = P(E_1 \cap E_2 \ldots \cap E_n) = \prod_{i=1}^{n} P(E_i)$$

$B_k = \{$ there are k successes in n trials $\}$

$$\mathbb{P}(B_k) = \sum_{\omega \in B_k} \mathbb{P}(\{\omega\})$$

$$\omega \in B_k \iff \omega = \{\omega_1, \ldots, \omega_n\} = \bigcap_{i=1}^{n} E_i$$

where only k of the w_i or E_i are success'

$\omega \in B_k$, $\quad P(\{\omega\}) = P\left(\bigcap_{i=1}^{n} E_i\right) = \prod_{i=1}^{n} P(E_i) = p^k (1-p)^{n-k}$

independence

k of them are success

$n-k$ are failure

$$\forall \omega \in B_k \quad P(\{\omega\}) = p^k (1-p)^{n-k} \quad (\equiv \text{Same Probability})$$

$$\therefore P(B_k) = |B_k| \, p^k (1-p)^{n-k}$$

$|B_k| = \#$ of ways k successes occur in n trials

$= {}^nC_k$

$$\therefore P(B_k) = {}^nC_k \, p^k (1-p)^{n-k} \qquad 0 \le k \le n.$$

$$\sum_{k=0}^{n} \mathbb{P}(B_k) = \sum_{k=0}^{n} {}^nC_k\, p^k (1-p)^{n-k} = (p+1-p)^n = 1$$

B_k Disjoint

Binomial expansion

$$\mathbb{P}\left(\bigcup_{k=0}^{n} B_k\right) = \mathbb{P}(S)$$

— Set up / answer is correct upto this check.

(c) How many trials are required to obtain the first success?

let $A_i = \{ i^{th}$ trial is a success$\}$.

$C_k = \{$ first success occurs in the k^{th} trial $\}$

$$\Rightarrow C_k = \underbrace{A_1^c}_{\text{failure}} \cap A_2^c \cdots \underbrace{A_{k-1}^c}_{\text{failure}} \cap \underbrace{A_k}_{\text{success}}$$

$$\mathbb{P}(C_k) = \mathbb{P}\left(\bigcap_{i=1}^{k-1} A_i^c \cap A_k\right)$$

$$\overset{\text{independence}}{=} \prod_{i=1}^{k-1} \mathbb{P}(A_i^c)\, \mathbb{P}(A_k)$$

$$= \underbrace{(1-p)^{k-1}}_{k-\text{failures}} \underbrace{p}_{\text{success}}$$

$\boxed{\text{Geometric}(p)}$

<u>Notation</u>: • An experiment with $S = \{1,2,3\ldots\}$ $\boxed{\text{Geometric}(p)}$

$$\mathbb{P}(\{k\}) = (1-p)^{k-1} p \qquad k \geq 1$$

• An experiment with $S = \{0,1\}$ $\boxed{\text{Bernoulli}(p)}$

$$\mathbb{P}(\{1\}) = p$$

$$\mathbb{P}(\{0\}) = 1-p$$

• $n \geq 1$ An experiment with $S = \{0,1,\ldots,k,\ldots n\}$

$$\mathbb{P}(\{k\}) = {}^nC_k\, p^k (1-p)^{n-k}$$ $\boxed{\text{Binomial}(n,p)}$

2.2 Poisson Approximation

Example 2.2.1: A small college has 1460 students

- assume birth rates are constant throughout the year

What is the probability that five or more students were born on independence day?

Answer: :- $\mathbb{P}$(5 or more were born independence day)

Probability of complement

$$= 1 - \mathbb{P}(\text{atmost 4 were born on independence day})$$

Disjoint sets

$$= 1 - \sum_{k=0}^{4} \mathbb{P}(\text{k were born on independence day})$$

Setup: $n = 1460$ Bernoulli$\left(\frac{1}{365}\right)$ trials

$$= 1 - \sum_{k=0}^{4} \underbrace{{}^{1460}C_k \left(\frac{1}{365}\right)^k \left(\frac{364}{365}\right)^{1460-k}}$$

Computation difficulties: ${}^{1460}C_3 = \frac{1460!}{3!\,1457!} = \frac{(1460)(1459)(1458)}{3 \cdot 2 \cdot 1}$

- Say 100 were born on independence day
  - becomes computationally difficult.

$$\boxed{n = 1460, \quad p = \frac{1}{365} \qquad p << n}$$

Theorem 2.2.2 (Poisson Approximation)

let $\lambda > 0$, $k \geq 1$, $n \geq \lambda$ and $p = \frac{\lambda}{n}$

Consider n - Bernoulli (p) trials.

n - independent Bernoulli $\left(\frac{\lambda}{n}\right)$

dependence on n

Fix $k = 0, 1, 2, \ldots$. $A_k = \{ k$ success in n - Bernoulli (p) trials $\}$

$$\mathbb{P}(A_k) = {}^{n}C_k \, p^k (1-p)^{n-k}$$

$$\lim_{n \to \infty} \mathbb{P}(A_k) = \frac{e^{-\lambda} \lambda^k}{k!}$$

$\lim n \to \infty$

$n >>>$ $\quad p = \frac{\lambda}{n} <<<$

but $np = \lambda$