Recall :-

- Bernoulli trials — $S = \{0,1\}$, $\mathcal{F} = \mathcal{P}(S)$ Bernoulli (p)

  p - Probability of success

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

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

- n repeated independent Bernoulli trials — $S = \{0,1,2,\ldots,n\}$, $\mathcal{F} = \mathcal{P}(S)$ Binomial (n,p)

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

  k - # of sucesses.

Example last time

- $n = 1460$, Binomial $\left(n, \frac{1}{365}\right)$

$$\mathbb{P}(\text{5 or more students were born on independence day}) = 1 - \underbrace{\sum_{k=0}^{4} {}^{1460}C_{k} \left(\frac{1}{365}\right)^{k} \left(\frac{364}{365}\right)^{1460-k}}_{\text{Computation can be a challenge}}$$

Theorem 2.2.2 :- let $\lambda \geq 0$, $k \geq 1$, $n \geq \lambda$ and $p = \frac{\lambda}{n}$

$A_k = \{k \text{ successes in } n \text{ Bernoulli}(p) \text{ trials}\}$

Then
$$\lim_{n \to \infty} \mathbb{P}(A_k) = \frac{e^{-\lambda} \lambda^{k}}{k!} \quad \forall\ k = 0,1,2,\ldots$$

Proof:- $n \geq 1$, Fix $k \geq 1$

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

$$= \frac{n(n-1)\cdots(n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^{k} \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

$$= \frac{\lambda^k}{k!} \, \frac{n(n-1)\cdots(n-k+1)}{n^k} \left(1-\frac{\lambda}{n}\right)^{n-k}$$

$$\mathbb{P}(A_k) = \frac{\lambda^k}{k!} \, \frac{n}{n}\left(\frac{n-1}{n}\right)\cdots\left(\frac{n-k+1}{n}\right)\cdot\left(1-\frac{\lambda}{n}\right)^{-k}\cdot\left(1-\frac{\lambda}{n}\right)^{n}$$

$$= \frac{\lambda^k}{k!} \, \underbrace{1\cdot\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{(k-1)}{n}\right)}_{k\text{ - terms}} \cdot \underbrace{\left(1-\frac{\lambda}{n}\right)^{-k}} \underbrace{\left(1-\frac{\lambda}{n}\right)^{n}}$$

Fact: $\lim_{n\to\infty} \frac{1}{n} = 0$

$\frac{k-1}{n} \to 0$ as $n\to\infty$

$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{n} = e^{-\lambda}$

$$\lim_{n\to\infty} \mathbb{P}(A_k) = \frac{\lambda^k}{k!} \, \underbrace{1 \cdots 1 \quad 1}_{k\text{ terms}} \quad \underbrace{1}_{1\text{ terms}} \quad e^{-\lambda}$$

$$= \frac{\lambda^k}{k!} e^{-\lambda} \qquad \square$$

Back to examples: $p = \frac{1}{365}$ $\quad n = 1460 \quad \lambda = np = \frac{1460}{365} = 4$

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

$$\approx 1 - \underset{(\lambda=4,\,k=1)}{e^{-4}} + 4e^{-4} + \underbrace{\frac{4^2}{2}}_{(\lambda=4,\,k=2)} e^{-4} + \frac{4^4}{24} e^{-4} \qquad \square$$

Poisson Distribution (λ) :- $S = \{0, 1, 2, \dots\}$ $\mathcal{F} = \mathcal{P}(S)$

$$\mathbb{P}(\{k\}) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k \in S.$$

- n-Independent $\underbrace{\text{- Bernoulli}(p)}$ trials $\longrightarrow$ Binomial(n, p) $\longrightarrow$ Poisson(λ).

Experiment $\equiv$ {Success & failure}
p - $\mathbb{P}(\{\text{success}\})$

2.3 Sampling with and without Replacements

- Small town of 5000 Residents.
  - 1000 of them who are under age of 18.
- choose randomly 4 residents & ask how many are under the age of 18.

Two methods:

— "with" replacement — after each selection, you select again from all residents

$\equiv$ independent Bernoulli $\left(\frac{1}{5}\right)$ trials. ✓

— "without" replacement — If an individual is chosen, the person is no longer available to be selected in a later selection.

Example 2.3.1. In the town above

$\mathbb{P}$(4 residents selected at random, from without replacement, exactly two of them are under 18) = ?

- Selecting 4 residents from 5000 residents:

$\binom{5000}{4}$ ways.

- each way is equally likely

- Selecting exactly 2 residents under the age of 18

$\binom{1000}{2}\binom{4000}{2}$ ways

$$\mathbb{P}\left(\begin{array}{l}\text{of choosing exactly 2 residents}\\ \text{under age of 18 in a random}\\ \text{sample of 4 from 5000}\\ \text{(without replacement)}\end{array}\right) = \frac{\binom{1000}{2}\binom{4000}{2}}{\binom{5000}{4}}$$

≈ 0.153592

$\square$

Compare:

- Bernoulli$(\frac{1}{5})$
- $n = 4$

$$\mathbb{P}\left(\begin{array}{l}\text{of choosing exactly 2 residents}\\ \text{under age of 18 in a random}\\ \text{sample of 4 from 5000}\\ \text{(with replacement)}\end{array}\right) = \binom{4}{2}\left(\frac{1}{5}\right)^2\left(\frac{4}{5}\right)^2$$

≈ 0.1536

Keys:
- Methods are different: without replacement / with replacement
- answer are not equal but are close

Sampling with replacement – and – Sampling without replacement

- The two methods give "similar" answers if the sample size is "much smaller" than the populations they came from.

  – key fact used in statistics –

Hypergeometric (N, r, m)

N – # of population

r – # in population with a certain characteristic – R

m – # of sample chosen without replacement.

$$S = \{\max\{0, m-(N-r)\}, \dots, \min\{m, r\}\}$$

$$\mathbb{P}(\{k\}) := \text{"}\mathbb{P}(\text{ k – with characteristic R in a sample with replacement})\text{"} = \frac{\binom{r}{k}\binom{N-r}{m-k}}{\binom{N}{m}}$$

a sample $\xrightarrow{}$ size m; replacement $\xrightarrow{}$ population N

$k \in S$.

Ex:- $\sum_{k \in S} \mathbb{P}(\{k\}) = 1$

– will ensure $\mathbb{P}$ is a Probability on S.

Observations:

with replacement $\rightarrow$ Independent Bernoulli (p) trials

without replacement $\rightarrow$ "Dependent" trials

2.3.2. Hypergeometric Distributions as a Series of Dependent Trials

Previous Example:-

- 5000 people
- 1000 of them are under the age of 18
- Choose 4 people without replacement

$E = \{$ there are exactly two of them under the age of 18 $\}$.

$j = 1, 2, 3, 4$ — index our selection

$A_j = \{ j^{th}$ selection is a person younger than 18 $\}$

$$\mathbb{P}(A_1) = \frac{1000}{5000} = \frac{1}{5}$$

— 999 under 18 are left

$$\mathbb{P}(A_2 \mid A_1) = \frac{999}{4999}$$

998 Under 18 are left

$$\mathbb{P}(A_3^c \mid A_1 \cap A_2) = \frac{4000}{4998}$$

$$\mathbb{P}(A_4^c \mid A_1 \cap A_2 \cap A_3^c) = \frac{3999}{4997}$$

$$P(A_1 \cap A_2 \cap A_3^c \cap A_4^c) \overset{\text{Theorem 1.3.8}}{=} \mathbb{P}(A_1)\, P(A_2 \mid A_1)\, P(A_3^c \mid A_1 \cap A_2) \cdot \mathbb{P}(A_4^c \mid A_1 \cap A_2 \cap A_3^c)$$

$$= \frac{1000}{5000} \frac{999}{4999} \frac{4000}{4998} \frac{3999}{4997}$$

Similarly $\mathbb{P}(A_1^c \cap A_2 \cap A_3 \cap A_4^c) = \frac{4000}{5000} \cdot \frac{1000}{4999} \cdot \frac{999}{4998} \cdot \frac{3999}{4997}$

in each ordering: the individual fractions differ but Product is same

$$\mathbb{P}(E) = \binom{4}{2} \cdot \frac{1000}{5000} \frac{999}{4999} \frac{4000}{4998} \frac{3999}{4997}$$

$$\underset{(\text{Ex.})}{=} \frac{\binom{1000}{2}\binom{4000}{2}}{\binom{5000}{4}}$$

Theorem 2.3.3: Let N, m and r be positive integers. $(m < r < N)$ $k \in \{0, 1, \ldots, m\}$.

Define $p = \frac{r}{N}$, $p_1 = \frac{r-k}{N-k}$ $p_2 = \frac{r-k}{N-m}$

$H \equiv$ Probability that Hypergeometric (N, r, m) takes value k

$$\underbrace{\binom{m}{k} p_1^k (1-p_2)^{m-k}}_{\text{If } p_1 \approx p_2 \text{ Bernoulli}(p_1)} < H < \binom{m}{k} p^k (1-p_1)^{m-k}$$

"Dependent"

If $p_1 \approx p$ then Bernoulli(p_1)

$$\boxed{N >> m > k \text{ "}\equiv\text{" } p_1 \approx p_2 \approx p}$$