Assignment 14

G HARSHA VARDHAN REDDY (CS21BTECH11017)

June 15, 2022

AI1110

Outline

Problem Statement

Papoulis Pillai Probability Random Variables and Stochastic Processes
Exercise : 8-13

We plan a pole for the purpose of estimating the probability p of Republicans in a community.We wish our estimate to be within ± 0.02 of p. How large should our sample be if the confidence coefficient of the estimate is 0.95?

Definitions

Sample Proportion

If X is a binomial random variable, then $X \sim B(n, p)$ where n is the number of trials and p is the probability of a success. To form a sample proportion, take X, the random variable for the number of successes and divide it by n, the number of trials (or the sample size). The random variable P' is the sample proportion

$$P' = \frac{X}{n} \tag{1}$$

And
p' = the estimated proportion of successes or point estimate for p

Binomial distribution

Let the random variable X be the sum of n Bernoulli random variables, i.e.

$$X = X_1 + X_2 + \ldots + X_n \tag{2}$$

Assume $E(X_i) = p$. Then,

$$\begin{align} E(X) &= E(X_1 + X_2 + \ldots + X_n) \tag{3} \\ &= E(X_1) + E(X_2) + \ldots + E(X_n) \tag{4} \\ &= p + p + \ldots + p \tag{5} \\ &= np \tag{6} \end{align}$$

Binomial Distribution

The variance is given by:

$$E((X-np)^2) = E(X^2 - 2Xnp + n^2p^2) \tag{7}$$
$$= E(X^2) - 2npE(X) + E(n^2p^2) \tag{8}$$
$$= E(X^2) - n^2p^2 \tag{9}$$

Now,

$$E(X^2) = \sum_{k=0}^{n} k^2 \times \binom{n}{k} p^k q^{n-k} \tag{10}$$
$$= \sum_{k=1}^{n} npk \times \binom{n-1}{k-1} p^{k-1} q^{n-k} \tag{11}$$

Binomial Distribution

$$\sum_{k=1}^{n} npk \times \binom{n-1}{k-1} p^{k-1} q^{n-k} = n(n-1)p^2 \times \sum_{k=2}^{n} \binom{n-2}{k-2} p^{k-2} q^{n-k} + np \times \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} q^{n-k} \tag{12}$$

$$= n(n-1)p^2(p+q)^{n-2} + np(p+q)^{n-1} \tag{13}$$

$$= n(n-1)p^2 + np \tag{14}$$

From 9,

$$\sigma^2 = (n(n-1)p^2 + np) - n^2p^2 \tag{15}$$

$$= np - np^2 = np(1-p) \tag{16}$$

$$\implies \sigma^2 = npq \tag{17}$$

Let the random variable $Y = \frac{X}{n}$. Then, Y is a binomial random variable that maps to $\frac{k}{n}$ when X maps to k. Therefore,

$$\mu_Y = \frac{\mu_X}{n} \tag{18}$$

$$= p \tag{19}$$

$$\sigma_Y^2 = \frac{\sigma_X^2}{n^2} \tag{20}$$

$$= \frac{pq}{n} \tag{21}$$

$$\implies \sigma_Y = \sqrt{\frac{pq}{n}} \tag{22}$$

Youth percentile or Z score

Z-score

Z-score indicates how much a given value differs from the standard deviation. The Z-score, or standard score, is the number of standard deviations a given data point lies above or below mean.

$$\implies Z_u = \frac{x - \mu}{\sigma} = \frac{p - p'}{\sigma_{p'}} \tag{23}$$

Where,
Z_u = Normal (Youth) percentile or Z score
x = Observed value
σ = Standard deviation

Confidence interval for a population proportion

A **Confidence Interval** is an estimate for an unknown parameter.It is governed by a number $\gamma = 1 - \delta$, which determines the accuracy of the estimation method. γ is called the **confidence coefficient**.
The confidence interval for a population proportion (p)

$$|p - p'| \leq \sigma Z_u \tag{24}$$

$$p' - \sigma Z_u \leq p \leq p' + \sigma Z_u \tag{25}$$

From (22)

$$\sigma_{p'} = \sqrt{\frac{(1 - p')(p')}{n}} \tag{26}$$

Therefore,

$$p' - Z_u \times \sqrt{\frac{(1 - p')(p')}{n}} \leq p \leq p' + Z_u \times \sqrt{\frac{(1 - p')(p')}{n}} \tag{27}$$

Solution

Given,

$$p \leq \pm 0.02 \text{ of } p' \tag{28}$$

$$\implies |p - p'| \leq 0.02 \text{ and,} \tag{29}$$

$$\text{Confidence Coefficient } (CF) = 0.95 \implies Z_u = 2 \tag{30}$$

From (27),(29) and (30),

$$\sqrt{\frac{(1 - p')(p')}{n}} \times 2 \leq 0.02 \tag{31}$$

As $n > 0$, From (31)

$$\frac{(1 - p')(p')}{n} \leq \left(\frac{1}{100}\right)^2 \tag{32}$$

$$\implies n \geq (1 - p')(p') \times 100^2 \tag{33}$$

$$A.M \geq G.M \tag{34}$$

$$\frac{(p') + (1 - p')}{2} \geq \sqrt{(p')(1 - p')} \implies p'(1 - p') \leq \frac{1}{4} \tag{35}$$

From (33),(35)

$$n \geq \frac{100^2}{4} \tag{36}$$

$$\implies n \geq 2500 \tag{37}$$

Therefore, the size of sample(n) must be greater than equal to 2500.