Subject	Statistics
Year	1st Year B.Sc
Paper no	05
Paper Name	Probability Distributions-1
Topic no	21
Topic name	Trinomial Distribution
SME	Suguna Nayak
ID	Syeda Shafia Kausar

E-Learning Module on Trinomial Distribution

Learning Objectives

By the end of this session, you will be able to:

- Define trinomial distribution
- List the application of trinomial distribution
- List the properties of trinomial distribution
- Solve the numerical on trinomial distribution

Introduction

We have already studied binomial distribution, which applies for entities having two states. For example, boy-girl, head-tail, working-broken and so on that are classified as success and failure.

There are circumstances when three states are appropriate. For example,

- A bicycle has three principal states namely, Parked, Ridden or Pushed.
- The traffic lights can be Red, Green or Changing.
- There was brief period when ternary computers were thought worth exploring: voltage would have been positive, zero or negative.

There are many three-state examples in genetics. It would be fanciful to imagine that children came in three sexes, but if both parents have blood group AB, then each of offspring will necessarily have blood group, AA, AB or BB and there are known probabilities for each.

One can reasonably ask the probability of such parents with four children having two children AA, one AB and one BB. Problems involving entities which have three states lead naturally to a discussion of the Trinomial distribution.

Before proceeding, consider a summary of the case of a family of 4 children and the two-state boy-girl analysis.

- There are two salient probabilities p and q; these are the probabilities of a child being a boy or a girl respectively. Necessarily $p+q=1$.

- If order is taken into account, there are $2^4=16$ ways of having four children being GGGG, GGGB, … , BBBB.

- If order is not taken into the account, there are 4+1=5 ways of having four children, which can be listed as 0 boys, 1 boy, 2 boys, 3 boys and 4 boys.

- The probability of there being r boys is,

$$\binom{4}{r} p^r q^{4-r} = \frac{4!}{r!(4-r)!} p^r q^{4-r}$$

Suppose, the three bold groups are labeled a, b and c and are regarded as three states,

- There are three salient probabilities; call these p_a, p_b and p_c and note that necessarily $p_a+p_b+p_c=1$.
- If order is taken into account, there are $3^4=81$ ways of having 4 children because each may be one of the three possibilities.
- If order is not taken into the account, the number of ways of having four children turns out to be 15 and these ways are most easily presented in a triangular array.

aaaa

aaab aaac

aabb aabc aacc

abbb abbc abcc accc

bbbb bbbc bbcc bccc cccc

There is no direct parallel to ‘the probability of there being r boys’ because of the complication introduced by having three possibilities.

Consider a rewrite of the expression for the probability of there being r boys in the Binomial case:

$$\frac{4!}{r!(4-r)!}p^r q^{4-r} = \frac{4!}{r_a!r_b!}{p_a}^{r_a}{p_b}^{r_b}$$

The boy-girl probabilities p and q have been replaced by p_a and p_b and the boy-girl numbers r and 4-r have been replaced by r_a and r_b. Clearly $p_a+p_b=1$ and $r_a+r_b=4$. This expression can be generalised to the Trinomial case:

$$\frac{4!}{r_a!r_b!r_c!}{p_a}^{r_a}{p_b}^{r_b}{p_c}^{r_c}$$

Where, $p_a+p_b+p_c=1$ and $r_a+r_b+r_c=4$

Definition

Suppose we define the sample space consisting of all sequences of length n such that $\omega=(i_1, i_2, \ldots, i_n)$, where each i_j take values 1, 0 and -1 with probabilities $P(i_j=1)=p$, $P(i_j=0)=\theta$, $P(i_j=-1)=1-p-\theta$.

If a specific sequence of ω has x 'successes'(1's), and y 'failures' (0's)
Let X be the number of trails where 1 occurs, and Y be the number of trials where 0 occurs. The joint distribution of pair (X, Y) is called the trinomial distribution.

Theorem

The joint pmf for (X, Y) is given by,

$$f_{XY}(x,y) = P(X = x, Y = y) = \frac{n!}{x!\,y!\,(n-x-y)!} p^x \theta^y (1-p-\theta)^{n-x-y}$$

Where x, y ≥0 and x+y≤n

Proof:

The sample space consists of all sequences of length n such that $\omega=(i_1, i_2, \ldots, i_n)$, where each i_j take values 1, 0 and -1 with probabilities, $P(i_j=1)=p$, $P(i_j=0)=\theta$, $P(i_j=-1)=1-p-\theta$.

If a specific sequence of ω has x 'successes'(1's), and y 'failures' (0's), then $P(\omega)=p^x\theta^y(1-p-\theta)^{n-x-y}$

There are $\binom{n}{x}\binom{n-x}{y} = \frac{n!}{x!y!(n-x-y)!}$ different sequences with x successes and y failures. Hence,

$$P(X = x, Y = y) = \frac{n!}{x!y!(n-x-y)!} p^x \theta^Y (1-p-\theta)^{n-x-y}$$

The name of the distribution comes from the trinomial expansion, i.e.

$(a+b+c)^n = [a+(b+c)]^n$

$$= \sum_{x=0}^{n} \binom{n}{x} a^x (b+c)^{n-x}$$

$$= \sum_{x=0}^{n} \sum_{y=0}^{n-x} \binom{n}{x}\binom{n-x}{y} a^x b^y c^{n-x-y}$$

$$= \sum_{x=0}^{n} \sum_{y=0}^{n-x} \frac{n!}{x!y!(n-x-y)!} a^x b^y c^{n-x-y}$$

Properties

1. The marginal distributions of X and Y are just X~ Binomial(n,p) and Y~ Binomial(n,θ). This follows the fact that X is the number of 'successes' in n independent trials with p being the probability of 'successes' in each trial. Similar argument works with Y.

Therefore, $E(X)=np$, $E(Y)=n\theta$

$V(X)=np(1-p)$ and $V(Y)=n\theta(1-\theta)$

$E(X^2)=V(X)+[E(X)]^2 =np(1-p)+n^2p^2$ and

$E(Y^2)=V(Y)+[E(Y)\}^2=n\theta(1-\theta)+n^2\theta^2$

2. If Y=y, then the conditional distribution of X|(Y=y) is Binomial (n-1, $^p/_{1-\theta}$)

Proof

$$P(X=x \mid Y=y) = \frac{P(X=x, Y=y)}{P(Y=y)}$$

$$= \frac{\dfrac{n!}{x!\,y!\,(n-x-y)!} p^x \theta^y (1-p-\theta)^{n-x-y}}{\dfrac{n!}{y!\,(n-y)!} \theta^y (1-\theta)^{n-y}}$$

$$= \binom{n-1}{x} \left(\frac{p}{1-\theta}\right)^x \left(1-\frac{p}{1-\theta}\right)^{n-y-x}$$

for x=0, 1, 2, …, n-y. Hence, X|(Y = y) is Binomial (n-1, $^p/_{1-\theta}$)

Note

This is intuitively obvious. Consider those trials for which 'failure' (or 0) did not occur. There are n-y trials, for which the probability that 1 occurs is actually the conditional probability of 1 given that 0 has not occurred. i.e. $^{p}/_{1-\theta}$. So, we have the standard binomial set up.

3. Now, using the results on conditional distributions and above properties, let us find Cov(X, Y) and the coefficient of correlation ρ_{XY}.

We know that for any two random variables, E[XY]=E[Y.E(X|Y)]. And according to the property 2, X|(Y = y) is a Binomial (n-1, $^p/_{1-\theta}$) and hence E[X|Y=y]=(n-y).[$^p/_{1-\theta}$] and thus E[X|Y]=(n-Y).[$^p/_{1-\theta}$] . Hence,

$$E[XY] = E\left[Y \times (n - Y)\frac{p}{1-\theta}\right] = \frac{p}{1-\theta}E(nY - Y^2)$$

$$= \frac{p}{1-\theta}(n^2\theta - n\theta(1-\theta) - n^2\theta^2) = \frac{p \times n\theta}{1-\theta}[n-1-\theta(n-1)]$$

$$= \frac{p \times n\theta}{1-\theta}[(1-\theta)(n-1)] = n(n-1)p\theta$$

Therefore ,

Cov(X,Y)=E(XY)-E(X)E(Y)

$=n(n-1)p\theta - n^2 p\theta = -np\theta$

And hence

$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{V(X)V(Y)}} = \frac{-np\theta}{\sqrt{n^2 p(1-p)\theta(1-\theta)}} = -\left(\frac{p\theta}{(1-p)(1-\theta)}\right)^{1/2}$$

Note that if $p+\theta=1$, then $Y=n-X$ and there is an exact linear relation between Y and X. In this case, it is easily seen that $\rho_{XY}=-1$

Generalization

Consider that there are k outcomes possible at each of the n independent trials. Denote the outcomes $A_1, A_2, \ldots, A_k$ and the corresponding probabilities $p_1, \ldots, p_k$ where $\Sigma_j p_j=1$, $j=1, 2, \ldots, k$. Let X_j count the number of A_j occurs. Then,

$$P(X_1 = x_1, \ldots, X_{k-1} = x_{k-1}) = \frac{n!}{x_1! \ldots x_{k-1}!(n - \sum_{j=1}^{k-1} x_j)!} p_1^{x_1} p_2^{x_2} \ldots p_{k-1}^{x_{k-1}} p_k^{n - \sum_{j=1}^{k-1} x_j}$$

Where, $x_1, x_2, \ldots, x_{k-1}$ are non-negative integers with $\Sigma_j x_j \leq n$

Illustration

In a recent three-way election for a large country, candidate A received 20% of the votes, candidate B received 30% of the votes, and candidate C received 50% of the votes. If six voters are selected randomly, what is the probability that there will be exactly one supporter for candidate A, two supporters for candidate B and three supporters for candidate C in the sample?

Solution:

Since we’re assuming that the voting population is large, it is reasonable and permissible to think of the probabilities as unchanging once a voter is selected for the sample.

Therefore, we can use trinomial distribution to find the probability. i.e. we can write the pmf as,

$$P(X = x, Y = y, Z = z) = \frac{n!}{x!\,y!\,z!} p_1^x p_2^y p_3^z$$

Where, $x+y+z=n$ and $p_1+p_2+p_3=1$

Here $p_1=0.2$, $p_2=0.3$ and $p_3=0.5$

$$P(X = 1, Y = 2, Z = 3) = \frac{6!}{1!2!3!}(0.2)^1(0.3)^2(0.5)^3 = 0.135$$