FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 4 No. 4, December, 2020, pp 415 – 424
DOI: https://doi.org/10.33003/fjs-2020-0404-498

APPLICATIONS OF TWO NON-CENTRAL HYPERGEOMETRIC DISTRIBUTIONS OF BIASED SAMPLING STATISTICAL MODELS

[1]Adetunji, K. O., [*2]Issa, A. A., [1]Alanamu, T., [1]Adefila, E. J. and [1]Muhammed, K. A.

[1]Department of Mathematics, Kwara State College of Education, Ilorin, Nigeria
[2]Department of Mathematical Sciences, Abubakar Tafawa Balewa University (ATBU), Bauchi, Bauchi State, Nigeria.

[*]Corresponding author's email: kunlemumeen@gmail.com *Tel.: 08068954589*

ABSTRACT
Statistical models of biased sampling of two non-central hypergeometric distributions Wallenius' and Fisher's distribution has been extensively used in the literature, however, not many of the logic of hypergeometric distribution have been investigated by different techniques. This research work examined the procedure of the two non-central hypergeometric distributions and investigates the statistical properties which includes the mean and variance that were obtained. The parameters of the distribution were estimated using the direct inversion method of hyper simulation of biased urn model in the environment of R statistical software, with varying odd ratios (w) and group sizes (m_i). It was discovered that the two non - central hypergeometric are approximately equal in mean, variance and coefficient of variation and differ as odds ratios (w) becomes higher and differ from the central hypergeometric distribution with ω = 1. Furthermore, in univariate situation we observed that Fisher distribution at (ω = 0.2, 0.5, 0.7, 0.9) is more consistent than Wallenius distribution, although central hypergeometric is more consistent than any of them. Also, in multinomial situation, it was observed that Fisher distribution is more consistent at (ω = 0.2, 0.5), Wallenius distribution at (ω = 0.7, 0.9) and central hypergeometric at (ω = 0.2)

Keywords: Non-central hypergeometric, Wallenius distribution, Fisher distribution, univarate situation

INRODUCTION

The hyper geometric distribution occupies a place of great significance in statistic theory.
It applies to sampling without replacement from a finite population whose element can be classified into two categories, one which possesses certain characteristics. The category could be male or female, employed or unemployed etc. When random selections are made without replacement from the population, each subsequent drawn is dependent on the outcome of the previous draws and the probability of success change, consequently. The conditions underlying hypergeometric distribution are as follows:

- The result of each draw can be classified into one of two categories
- The probability of success change in each draw.

In probability theory and statistics, the hypergeometric distribution is a discrete probability that describes the probability of number of successes in n draws, without replacement, from a finite population of size N containing K successes.

A hypergeometric random variable with parameter W+B, W and n, give a set consisting of W element of first kind and B element of the second kind, a number of element of the first kind appearing in a randomly chosen subset of n element, where every of such subset are equally likely. For a hypergeometric random variable X i.e. $X \sim H(W + B, W, n)$

- The sample space is the set of integers that meet $Max\ (0, n-B) < x < Min\ (n, w)$ and
- The probability mass function
  $P(X = x/(W+B), W, n)$ or simply $P(X = x)$ is define thus:

$$P(X = x) = \frac{\binom{W}{x}\binom{B}{n-x}}{\binom{W+B}{n}} \quad \ldots\ (1)$$

Various generalizations to this distribution exist. One instance could be a case of picking from an urn containing biasedly colored tagged balls, so that balls of one color are more likely to be picked than balls of another color. Another instance could be a case of an opinion poll, conducted by calling random telephone numbers and it is assumed that unemployed people are more likely to be home and answer the phone than employed people therefore, an unemployed respondent are likely to be over-represented in the sample.

The probability distribution of employed versus unemployed respondents in a sample of n respondents could be described as a non-central hypergeometric distribution. The description of biased urn models is complicated by the fact that there is more than one non central hypergeometric distribution, depending on whether items (e.g. coloured tagged balls) are sampled in a manner where there is competition between the items or they are sampled independently of each other. There is widespread confusion about this fact.

The name non-central hypergeometric distribution has been used for two different distribution and several scientists have either used the distributions wrongly or erroneously believed that the two distributions were identical. The use of the same name for two different distributions has been possible because these two distributions were studied by two different groups of scientist who are hardly have any contact with each other.

Anger Fog (2008) has suggested that the best way to avoid confusion is to use the name Wallenius non-central hypergeometric distribution of a biased urn model where a predetermined number of items are drawn one by one in a competitive manner while the name Fisher's non-central hypergeometric distribution is used where items are drawn independently of each other. This is done so that the total number of items drawn is known only after the experiment. There come the names K.T. Wallenius and R.A. Fisher, been those who first describe the respective distribution.

DESCRIPTION OF WALLENIUS' NON-CENTRAL HYPERGEOMETRIC DISTRIBUTION

For Wallenius' distribution, let assume that an urn contains m_1 red balls and m_2 white balls and $N = m_1 + m_2$, the totally number of balls in the urn. n balls are drawn at random from the urn one by one without replacement. Each red ball has the weight w_1, and each white ball has the weight w_2. We assume that the probability of taking a particular ball is proportional to its weight. The physical property that determines the odds may be something else than weight, such as size or slipperiness or whatever, but it is convenient to use the word weight for the odds parameter.

The probability that the first ball picked is red is equal to the weight fraction of the red balls:

$$P_1 = \frac{m_1 w_1}{m_1 w_1 - m_2 w_2} \quad \dots \quad (2)$$

The probability that the second ball picked is red depends on whether the first ball was red or white. If the first ball was red then (2) above is used with m_1 reduced by one. If the first ball was white then (2) is used with m_2 reduced by one.

The important fact that distinguishes Wallenius' distribution is that there is competition between the balls. The probability that a particular ball is taken in a particular draw depends not only on its own weight, but also on the total weight of the competing balls that remain in the urn at that moment. And the weight of the competing balls depend on the outcomes of all preceding draws.

The distribution of the balls that are not drawn is a complementary Wallenius' non-central hypergeometric distribution.

DESCRIPTION OF FISHER'S NON - CENTRAL HYPERGEOMETRIC DISTRIBUTION

In the Fisher model, the balls are independent and there is no dependence between draws. We may as well take all *n* balls at the same time. Each ball has no "knowledge" of what happens to the other balls. For the same reason, it is impossible to know the value of *n* before the experiment. If we try to fix the value of *n* then we would have no way of preventing ball number *n*+1 from being taken without violating the principle of independence between balls. *n* is therefore a random variable, and the Fisher distribution is a conditional distribution which can only be determined after the experiment, when *n* is known. The unconditional distribution is two independent binomials, one for each color.

Fisher's distribution can simply be defined as the conditional distribution of two or more independent binomial varieties dependent upon their sum. A multinomial version of the Fisher's distribution is used if there are more than two colors in the urn.

NON-CENTRAL HYPERGEOMETRIC DISTRIBUTION

The known standard hypergeometric distribution shows no dependence between the colour of a ball in the urn and its probability of been drawn. The only influencing parameter is the number of balls of the different colours in the urn.

Should one want to model the preferences in drawing balls of different colours, weight parameters are introduced, and the resulting distribution is called the NON-CENTRAL HYPERGEOMETRIC DISTRIBUTION and was developed by Wallenius and Fishers for the univariate cases and extended to a multinomial distribution.

In general, the non - central hypergeometric distribution has a number of important practical applications which includes:

1) Industrial quality control: lot of size N containing a proportion p of defectives are sampled using samples of fixed size n. The number of defectives x per sample is then a non-central hypergeometric random variable.
2) Estimation of the size of animal and other populations from capture - receptive data.
3) Estimation of a target population N in epidemiological studies, can be achieved by counting the number of cases that appear on both of two lists of sizes n and m
4) Opinion surveys: a random sample of size n, of respondents, is drawn without replacement from a finite population of size N.
5) Analysis of 2x2 contingency tables with both sets of marginal frequencies fixed. The probability of a result as extreme as the observed result is the task probability for the resulting classical hypergeometric distribution.

Statistical models of biased sampling in the two non - central hypergeometric distribution occur when taking coloured balls from a bias urn without replacement. The univariate is used when there are two colours of balls. While the multinomial is used when there are more than two colours of balls.

The traditional procedure is to use unbiased sampling but a model of biased sampling may be used if bias is unavoidable or if bias is desired in order to increase the probability of detection.

This study is aimed at applying the two non-central hypergeometric distributions under statistical models of biased sampling. It is therefore designed in line with the following objectives: to examine the procedure of applying the two non-central hypergeometric distributions (Wallenius' and Fisher's distributions); to investigate the statistical properties of the two non – central hypergeometric distributions such as (mean, variance and coefficient of variation), and to compare the coefficient of variation of the two non – central hypergeometric distributions in univariate and multinomial cases

APPLICATIONS OF THE TWO DISTRIBUTIONS, IN CONSIDERATION

Wallenius	**Fishers'**
1. Wallenius distribution is used in models of natural selection and biased sampling	1. Fisher's non-central hyper-geometric distribution is useful for models of biased sampling or biased.
2. Wallenius' non-central hyper-geometric distribution is used when items are sampled one by one with competition.	2. Fisher's non-central hyper-geometric distribution can also be used to select on items sampled.
3. The distribution is applicable in random number theory.	3. The distribution can also be used for test in contingency tables where a conditional distribution for fixed margin is desired.

CONDITIONS UNDER WHICH EACH OF THE TWO DISTRUNTIONS CAN BE USED	
Wallenius	**Fishers'**
1. Items are taken randomly from a finite source containing different kinds of items without replacement.	1. Items are taken randomly from a finite source containing different kinds of items without replacement.
2. Items are drawn one by one.	2. Items are taken independently of each other. Whether one item is taken is independent of whether another item is taken. Whether one item is taken before, after, or simultaneously with another item is irrelevant.
3. The probability of taking a particular item at a particular draw is equal to its fraction of the total weight of all items that have not yet been taken at that moment. The weight of an item depends only on its kind.	3. The probability of taking a particular item is proportional to its weight. The weight of an item depends only on its kind.
4. The total number *n* of items to take is fixed and independent of which items happen to be taken first.	4. The total number *n* of items that will be taken is not known before the experiment.
	5. *n* is determined after the experiment and the conditional distribution for *n* known is desired.

METHODOLOGY

Central Hypergeometric Distribution

Suppose X_1 and X_2 represent two independent binomial random variables with parameter (m_1, π) and (m_2, π) respectively. Then $X_1 - X_2$ has a binomial distribution with parameters $N = m_1 - m_2$ and π. The conditional distribution of X_1 given $X_1 - X_2 = n$ is the univariate central hypergeometric distribution and is derived as follows;

$$P(X_1 = x_1) = \binom{m_1}{x_1}\pi^{x_1}(1-\pi)^{m_1 - x_1} \quad \ldots \quad (3)$$

$$P(X_2 = x_2) = \binom{m_2}{x_2}\pi^{x_2}(1-\pi)^{m_2 - x_2} \quad \ldots (4)$$

$$P(X_1 - X_2 = n)$$

$$= \binom{m_1 - m_2}{n}\pi^{n}(1-\pi)^{m_1 - m_2 - n} \quad \ldots (5)$$

Now if we let $x_1 = x$ and $x_2 = n - x$ in (3) and (4) then the conditional distribution becomes;

$$P(X_1 = x/n) = \frac{\binom{m_1}{x}\pi^{x}(1-\pi)^{m_1 - x}\binom{m_2}{n-x}\pi^{n-x}(1-\pi)^{m_2 - n + x}}{\binom{m_1 - m_2}{n}\pi^{n}(1-\pi)^{m_1 - m_2 - n}}$$

$$\ldots (6)$$

Collecting the exponent involving x together, (6) becomes:

$$P(X_1 = x/n) = \frac{\binom{m_1}{x}\binom{m_2}{n-x}}{\binom{m_1 - m_2}{n}} \quad \ldots \quad (7)$$

Therefore, *(7)* is the required probability mass function *(pmf)* for the Central Hypergeometric Distribution.

For univariate central hypergeometric distribution, the *pmf* in *(7)* above has the corresponding mean and variance respectively.

$$E(x) = \frac{nm_1}{N} \quad \ldots \quad (8)$$

$$V(x) = \frac{nm_1(N - m_1)(N - n)}{N^2(N-1)} \quad \ldots \quad (9)$$

For the multinomial hypergeometric distribution, the p*mf* is as follow;

$$\prod_{i=1}^{c}\frac{\binom{m_i}{x_i}}{\binom{N}{n}} \quad \ldots \quad (10)$$

The corresponding mean and variance are;

$$E(x) = \frac{nm_i}{N} \quad \ldots \quad (11)$$

$$V(x) = \frac{nm_i(N - m_i)(N - n)}{N^2(N-1)} \quad \ldots \quad (12)$$

The *pmf* in *(7)* can be extended based on the violation of equal probability assumption of the two binomial random variables. If the assumption of equal probability is violated, then, the hypergeometric distribution becomes non-central hypergeometric distribution, since the distribution of the sum is no longer binomial. The proof according to Lawal (2003) is:

$$P(X_1 = x/n) = \frac{\binom{m_1}{x}\pi_1^x(1-\pi_1)^{m_1-x}\binom{m_2}{n-x}\pi_2^{n-x}(1-\pi_2)^{m_2-n-x}}{\sum_j \binom{m_1}{j}\pi_1^j(1-\pi_1)^{m_1-j}\binom{m_2}{n-j}\pi_2^{n-j}(1-\pi_2)^{m_2-n-j}} \quad \text{... (13)}$$

Collecting the exponent involving x and that involving j together, *(13)* becomes:

$$P(X_1 = x/n) = \frac{\binom{m_1}{x}\binom{m_2}{n-x}\left(\frac{\pi_1(1-\pi_2)}{\pi_2(1-\pi_1)}\right)^x}{\sum_j \binom{m_1}{j}\binom{m_2}{n-j}\left(\frac{\pi_1(1-\pi_2)}{\pi_2(1-\pi_1)}\right)^j} \quad \text{... (14)}$$

Let $\omega = \left(\frac{\pi_1(1-\pi_2)}{\pi_2(1-\pi_1)}\right)$ in *(14)*, then,

$$P(X_1 = x/n) = \frac{\binom{m_1}{x}\binom{m_2}{n-x}\omega^x}{\sum_j \binom{m_1}{j}\binom{m_2}{n-j}\omega^j} \quad \text{... (15)}$$

Therefore, *(15)* is the required *pmf* of the non-central hypergeometric distribution.

Univariate fishers' non-central hypergeometric distribution

The *pmf* in *(15)* was referred to as the extended hypergeometric distribution, by fishers (1935), where ω is the non-centrality parameter. However, the above non-central hypergeometric was, according to Fog (2008), also referred to as fisher's hypergeometric distribution, by fog (2008).

The corresponding mean and variance of the *pmf* in *(15)*, according to McCullagh and Nelder (1989), are:

$$E(X;\omega) = \frac{P_1(\omega)}{P_0(\omega)} \quad \text{... (16)}$$

And

$$V(X;\omega) = \frac{P_2(\omega)}{P_0(\omega)} - \left(\frac{P_1(\omega)}{P_0(\omega)}\right)^2 \quad \text{... (17)}$$

Where,

$$P_r(\omega) = \sum_j \binom{m_1}{j}\binom{m_2}{n-j} j^r \omega^j \quad \text{... (18)}$$

The moments about the origin is expressible as:

$$\mu_r(\omega) = \frac{P_r(\omega)}{P_0(\omega)} \quad \text{... (19)}$$

Multinomial Fisher Non-Central Hypergeometric Distribution

Suppose that $X_1 \sim M(m_1, \pi_1)$ and $X_2 \sim M(m_2, \pi_2)$ are two independent random variables of k categories each. Then the conditional distribution of $X_1 = x_1$ given $X_1 - X_2 = n$, is as follows;

In equation *(7)* above, equal probability was assumed and thus the probabilities canceled out.

For the **multinomial fisher's hypergeometric** case, where the equal probability is not assumed, the resultant *pmf* follows the pattern of the univariate case in *(15)* above and is defined, according to McCullagh and Nelder (1989), as:

$$P(X_1 = x/n) = \frac{\binom{m_1}{x}\binom{m_2}{n-x}\omega_1^{x_1}\dots\omega_k^{x_k}}{\sum_j \binom{m_1}{j}\binom{m_2}{n-j}\omega_1^{j_1}\dots\omega_k^{j_k}} \quad \text{... (20)}$$

where,

$$\omega_j = \frac{\pi_{1j}\pi_{2k}}{\pi_{2j}\pi_{1k}}$$

The corresponding approximate relationship between the mean and variance, according to McCullagh and Nelder (1989), is as follows:

$$\frac{E(X_{1j}, X_{2k})}{E(X_{2j}, X_{1k})} = \omega_j = \frac{\mu_{1j}\mu_{2k} - \sigma_{jk}}{\mu_{2j}\mu_{1k} - \sigma_{jk}}$$

Where,

$$\sigma_{jk} = cov(X_{1j}, X_{1k}) = -cov(X_{1j}, X_{2k})$$

It is negative for $j < k$

The covariance matrix $\sum$ may be approximated quite accurately as follows. If we let the vector ζ with components ζ_j given by: $\frac{1}{\zeta_j} = \frac{1}{\mu_{1j}} + \frac{1}{\mu_{2j}}$

The approximate value of $\sum$ is then given, in terms of ζ, as:

$$\hat{\Sigma} = \frac{n}{n-1}\{diag(\zeta) - \zeta\zeta'/\zeta\} \quad \text{... (21)}$$

Univariate Wallenius Non-Central Hypergeometric Distribution

The Wallenius non-central hypergeometric distribution is the name given by Wallenius (1963) to a distribution constructed by supposing that, in sampling without replacement, the probability of drawing a white ball given that there are m_1 white and m_2 black balls is not p but $p/[p-\omega(1-p)]$, $\omega \neq 1$. The mathematical analysis that follows from this assumption is very complicated.

Starting from the recurrence relation:

$$P(X=x|m_1,m_2,N) = \frac{pP(X=x-1|m_1-1,m_2-1,N-1)}{p-\omega(1-p)} - \frac{\omega(1-p)P(X=x|m_1,m_2-1,N-1)}{p-\omega(1-p)} \quad \ldots (22)$$

Wallenius obtained the formula for his Non-central Hypergeometric distribution as:

$$P(X=x) = \binom{m_1}{x}\binom{m_2}{n-x}\int_0^1 (1-t^c)^x (1-t^{\omega/c})^{n-x}\,dt \quad \ldots \quad (23)$$

where,

$$c = [m_1 - x - \omega(N-m_1-n-x)]^{-1}$$

For the univariate case with $\omega > 1$, Fog (2008) indicated that it may be more efficient to solve:

$$\left(1-\frac{\mu_1}{m_1}\right)^{1/\omega_1} = \left(1-\frac{\mu_2}{m_2}\right)^{1/\omega_2} = \cdots = \left(1-\frac{\mu_k}{m_k}\right)^{1/\omega_k}$$

Levin (1984) proposed an approximation formula for approximating the variance of Wallenius' non-central hypergeometric distribution using the Fisher's non-central hypergeometric distribution with the same mean. His approximation formula is given as:

$$\sigma^2 \cong \sigma_F^2 = \frac{Nab}{(N-1)[mb-(N-m)a]} \quad \ldots \quad (24)$$

Where $a = \mu(m-\mu)$ and

$b = (n-\mu)(\mu - N - n - m$

This approximation is good when ω is closer to 1 and n is far from N

Multinomial Wallenius Non-central Hypergeometric Distribution

Accordingly, Chesson (1976) extended the univariate wallenius distribution to the multinomial case by defining:

$$P(X=x) = \prod_{i=1}^{k}\binom{m_i}{x_i}\int_0^1 \prod_{i=1}^{k}\left(1-t^{\frac{\omega_i}{d}}\right)^{x_i} dt$$

where, $d = \omega(m-x)$,

$m = (m_1, m_2, \ldots, m_k)$,

$x = (x_1, x_2, \ldots, x_k)$ and

$\omega = (\omega_1, \omega_2, \ldots, \omega_k)$

The approximation formula for mean and variance for Wallenius, as derived by Fog (2008) is as follows:

$\mu_{i,} = \mu_{i(i-1,} - p_{i,}(\mu_{i-1}$

where, $\mu_{i,} = (\mu_1, \mu_2, \ldots, \mu_k$

Notations

$x = (x_1, x_2, \ldots, x_k)$ is the number of balls drawn of each color

$\omega = (\omega_1, \omega_2, \ldots, \omega_k)$ is the initial number of balls of each color in the urn

$m = (m_1, m_2, \ldots, m_k)$ is the weight or odds of balls of each Color

n is the total number of balls drawn

k is the number of colors

N is the total number of balls in urn before sampling

SIMULATION STUDY

In this study, the direct inversion method of R statistical software was used with varying Odds ratio and group sizes. The aim of the research is to investigate and compare the statistical properties (mean, variance and coefficient of

variation) of the two non-central hypergeometric distributions in relation to the central hypergeometric distribution. Also to investigate the consistency nature of the distributions on the basis of five random number *(10, 50, 100, 500, 1000)*, that were considered in the study. The samples generated were replicated 1000 times to ensure stability of the results.

RESULTS AND DISCUSSIONS

Table 1: Univariate Case: ***Mean*** of simulated data, based on; $m_1 = 80$; $m_2 = 20$; n=20; $0 < \omega < 1$

Random numb generated	Odds ratio ω	Wallenius Distributn	Fisher Distributi on	Central Hypergeomet ric Distributn
10	0.2	10.12	11.01	16
	0.5	13.71	14.01	16
	0.7	14.89	15.02	16
	0.9	15.69	15.72	16
50	0.2	10.12	11.01	16
	0.5	13.71	14.01	16
	0.7	14.89	15.02	16
	0.9	15.69	15.72	16
100	0.2	10.12	11.01	16
	0.5	13.71	14.01	16
	0.7	14.89	15.02	16
	0.9	15.69	15.72	16
500	0.2	10.12	11.01	16
	0.5	13.71	14.01	16
	0.7	14.89	15.02	16
	0.9	15.69	15.72	16
1000	0.2	10.12	11.01	16
	0.5	13.71	14.01	16
	0.7	14.89	15.02	16
	0.9	15.69	15.72	16

TABLE 2: Univariate Case: ***Variance*** of the simulated data, based on $m_1 = 80$; $m_2 = 20$; n=20; $0 < \omega < 1$

Rand. numb generat	Odds ratio ω	Wallenius Distrib.	Fisher Distributi on	Central Hyperge o. Distribut n
10	0.2	3.30	3.31	2.59
	0.5	3.17	3.12	2.59
	0.7	2.93	2.90	2.59
	0.9	2.69	2.68	2.59
50	0.2	3.30	3.31	2.59
	0.5	3.17	3.12	2.59
	0.7	2.93	2.90	2.59
	0.9	2.69	2.68	2.59
100	0.2	3.30	3.31	2.59
	0.5	3.17	3.12	2.59
	0.7	2.93	2.90	2.59
	0.9	2.69	2.68	2.59
500	0.2	3.30	3.31	2.59
	0.5	3.17	3.12	2.59
	0.7	2.93	2.90	2.59
	0.9	2.69	2.68	2.59
1000	0.2	3.30	3.31	2.59
	0.5	3.17	3.12	2.59
	0.7	2.93	2.90	2.59
	0.9	2.69	2.68	2.59

TABLE 3: Univariate Case: ***Coefficient of variation*** of the simulated data, based on $m_1 = 80$; $m_2 = 20$; n=20; $0 < \omega < 1$

Rand. numb generat	Odds ratio ω	Wallenius Distrib.	Fisher Distributio n	Central Hypergeo. Distribtn
10	0.2	17.9505	16.5244	10.0584
	0.5	12.9858	12.6078	10.0584
	0.7	11.4958	11.3378	10.0584
	0.9	10.4533	10.4139	10.0584
50	0.2	17.9505	16.5244	10.0584
	0.5	12.9858	12.6078	10.0584
	0.7	11.4958	11.3378	10.0584
	0.9	10.4533	10.4139	10.0584
100	0.2	17.9505	16.5244	10.0584
	0.5	12.9858	12.6078	10.0584
	0.7	11.4958	11.3378	10.0584
	0.9	10.4533	10.4139	10.0584
500	0.2	17.9505	16.5244	10.0584
	0.5	12.9858	12.6078	10.0584
	0.7	11.4958	11.3378	10.0584
	0.9	10.4533	10.4139	10.0584
1000	0.2	17.9505	16.5244	10.0584
	0.5	12.9858	12.6078	10.0584
	0.7	11.4958	11.3378	10.0584
	0.9	10.4533	10.4139	10.0584

From the information in table 3, the plot of ***Coefficient of Variation (C.V)*** value, at each values of $\omega = (0.2, 0.5, 0.7, 0.9)$, is as follows

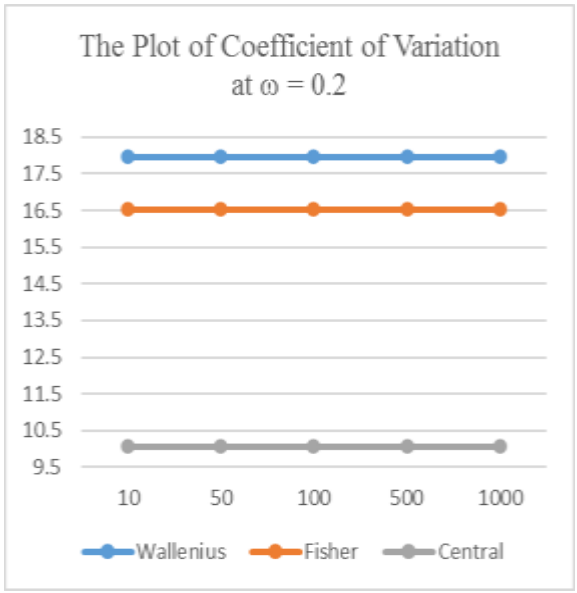

Fig. 1: the plot of *C.V* across the three distributions at $\omega = 0.2$

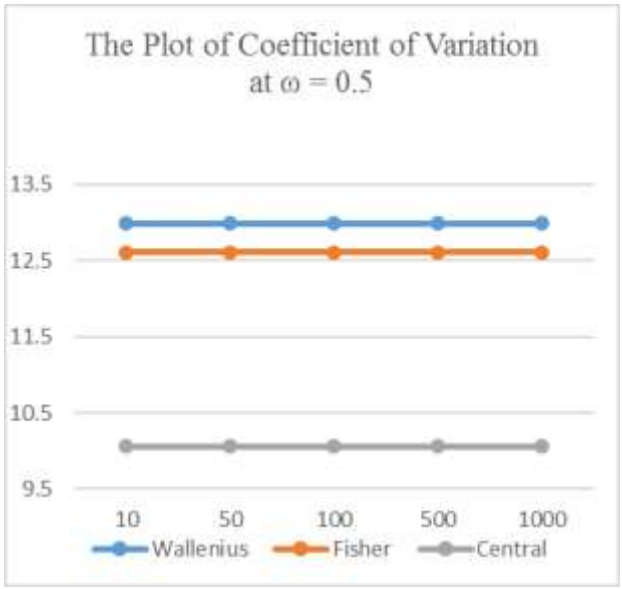

Fig. 2: the plot of *C.V* across the three distributions at $\omega = 0.5$

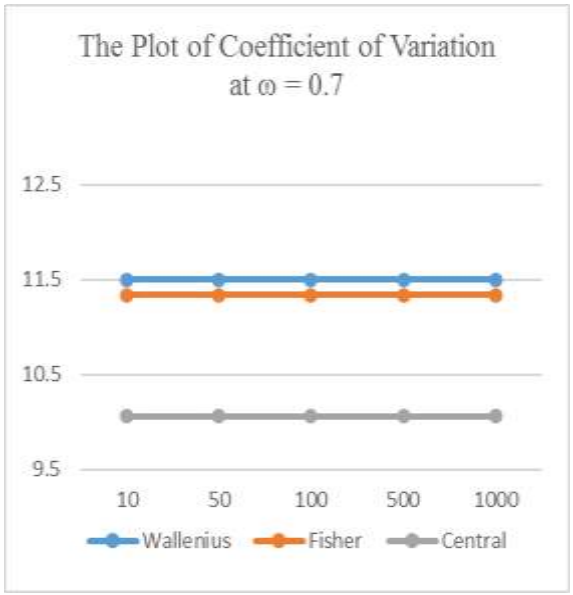

Fig. 3: the plot of *C.V* across the three distributions at $\omega = 0.7$

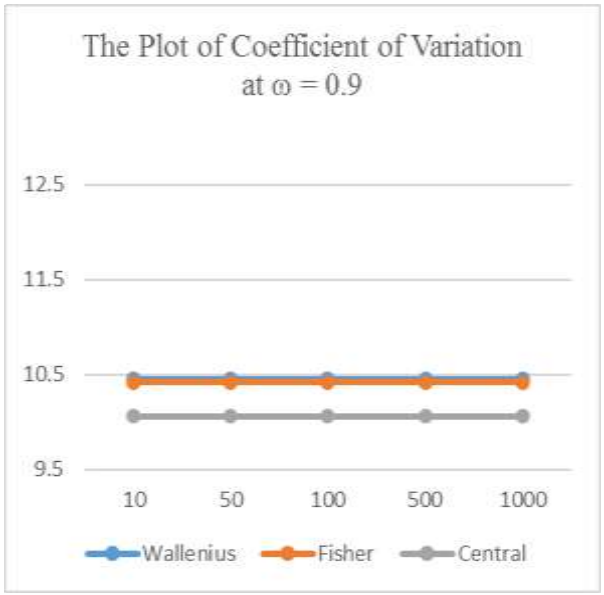

Fig. 4: the plot of *C.V* across the three distributions at $\omega = 0.9$

Table 4: Multinomial case: ***Mean*** of the simulated data based on odd ratio, $\omega = (0.2, 0.5, 0.7, 0.9)$; $m_1 = (80, 20.30, 40)$ and $n = 20$

Random Num. Generated	Odds Ratio ω	Wallenius Distribution	Fisher's Distribution	Central Hypergeometrc Distribution
10	0.2	4.08	4.3	9.41
	0.5	2.46	2.5	2.35
	0.7	5.04	5	3.53
	0.9	8.43	8.2	4.71
50	0.2	4.08	4.3	9.41
	0.5	2.46	2.5	2.35
	0.7	5.04	5	3.53
	0.9	8.43	8.2	4.71
100	0.2	4.08	4.3	9.41
	0.5	2.46	2.5	2.35
	0.7	5.04	5	3.53
	0.9	8.43	8.2	4.71
1000	0.2	4.08	4.3	9.41
	0.5	2.46	2.5	2.35
	0.7	5.04	5	3.53
	0.9	8.43	8.2	4.71

Table 5: Multinomial case: ***Variance*** of the simulated data based on odd ratio, $\omega = (0.2, 0.5, 0.7, 0.9)$; $m_1 = (80, 20.30, 40)$ and $n = 20$

Random Num. Generated	Odds Ratio ω	Wallenius Distribution	Fisher's Distribution	Central Hypergeometrc Distribution
10	0.2	3.0	3.11	4.42
	0.5	1.89	1.92	1.84
	0.7	3.18	3.17	2.58
	0.9	4.07	4.04	3.19
50	0.2	3.0	3.11	4.42
	0.5	1.89	1.92	1.84
	0.7	3.18	3.17	2.58
	0.9	4.07	4.04	3.19
100	0.2	3.0	3.11	4.42
	0.5	1.89	1.92	1.84
	0.7	3.18	3.17	2.58
	0.9	4.07	4.04	3.19
1000	0.2	3.0	3.11	4.42
	0.5	1.89	1.92	1.84
	0.7	3.18	3.17	2.58
	0.9	4.07	4.04	3.19

Table 6: Multinomial case: *Coefficient of Variation (C.V)* of the simulated data based on odd ratio, $\omega = (0.2, 0.5, 0.7, 0.9)$; $m_1 = (80, 20.30, 40)$ **and** $n = 20$

Rand. numb generated	Odds ratio ω	Wallenius Distrib.	Fisher Distribution	Central Hypergeo. Distribution
10	0.2	42.4522	41.0121	22.3420
	0.5	55.8851	55.4256	57.7220
	0.7	35.3821	35.6090	45.5025
	0.9	23.9315	24.5119	37.9205
50	0.2	42.4522	41.0121	22.3420
	0.5	55.8851	55.4256	57.7220
	0.7	35.3821	35.6090	45.5025
	0.9	23.9315	24.5119	37.9205
100	0.2	42.4522	41.0121	22.3420
	0.5	55.8851	55.4256	57.7220
	0.7	35.3821	35.6090	45.5025
	0.9	23.9315	24.5119	37.9205
1000	0.2	42.4522	41.0121	22.3420
	0.5	55.8851	55.4256	57.7220
	0.7	35.3821	35.6090	45.5025
	0.9	23.9315	24.5119	37.9205

The corresponding plot of ***Coefficient of Variation (C.V)*** value in table 6, at each values of $\omega = (0.2, 0.5, 0.7, 0.9)$, is as follows:

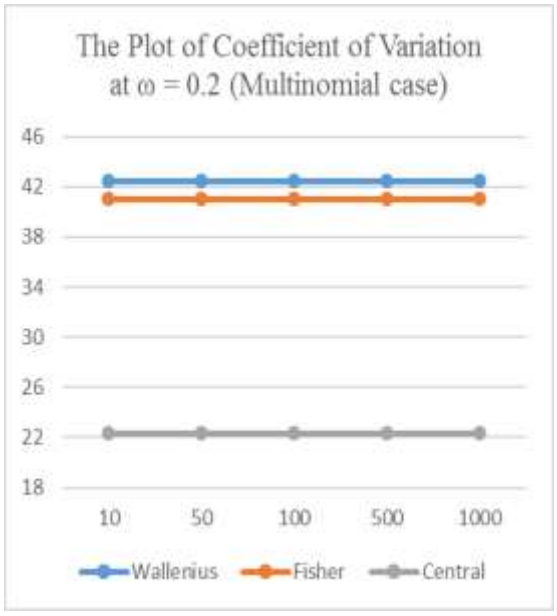

Fig. 5: the plot of *C.V* across the three distributions at $\omega = 0.2$

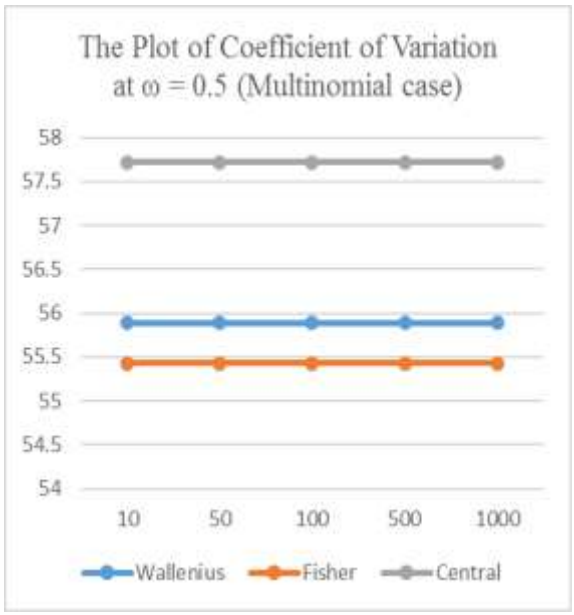

Fig. 6: the plot of *C.V* across the three distributions at $\omega = 0.5$

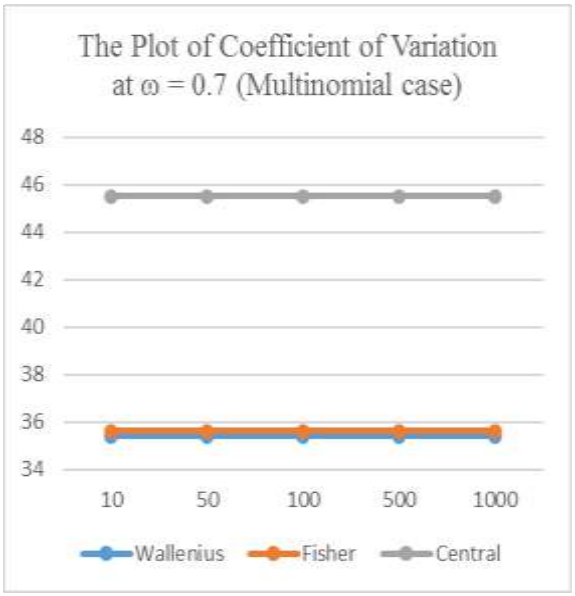

Fig. 7: the plot of *C.V* across the three distributions at $\omega = 0.7$

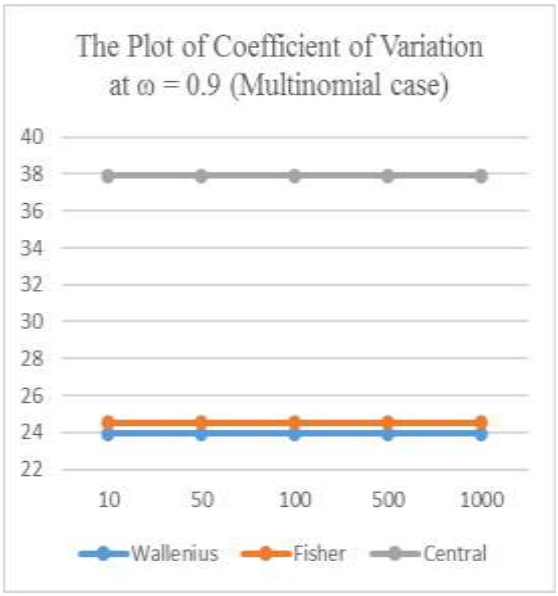

Fig. 8: the plot of *C.V* across the three distributions at $\omega = 0.5$

DISCUSSION

The simulation result in table 1 – 6 show, on the based random sample generated numbers (10, 50, 100, 500, 1000), that the estimated mean, the variance and coefficient of variation are approximately the same for the two non – central distributions with varying odds ratios $(\omega = 0.2, 0.5, 0.7, 0.9)$.

It could also be seen from these tables that the non – central hypergeometric distribution (Wallenius and Fishers') possess a closely approximate estimate of mean, variance and that the coefficient of variation differ from that given by the central hypergeometric distribution.

In univariate cases, table 1 – 3 and figure 1 – 4, it was observed that Fishers distribution at (ω = 0.2, 0.5, 0.7, 0.9) is more consistent than Wallenius distribution although central hypergeometric is better.

In multinomial cases, table 3 – 6 and figure 5 - 8, it was observed that Fisher distribution is more consistent at ω =0.5, Wallenius distribution at ω = 0.7, 0.9 and central hypergeometric distribution at ω = 0.2.

CONCLUSION

- Base on the aforementioned, it can be concluded that:
- The two non – central hypergeometric distributions (Wallenius and Fishers') are approximately equal in the estimate mean, variance and coefficient of variation across all the five random samples generated.
- The difference between the two non – central hypergeometric distributions becomes higher when the odd ratio is closer to 1

- The two non – central hypergeometric distributions differ from the central hypergeometric when odd ratio is closer to 1
- The two non – central hypergeometric distributions approximately equal to each other when they have same mean than when they have same ratio
- In univariate case, Fisher distribution are more consistent than Wallenius distribution while in multinomial case, both distributions perform differently.

REFERENCES

Chesson J. (1976), A non-central multivariate hypergeometric distribution arising from biased sampling with application to selective predation. *Journal of Appl. Probability 13(4): 795 - 797*

Fishers R. A. (1935), "The mathematical theory of probabilities and its application to frequency curves and statistical method" *Vol 1, Second Edition, New York; Macmillan.*

Fog A. (2008), "Calculation methods for Wallenius Non-centrall hypergeometric Distribution" *Communication Statistics Simulation and Computation 37 (2): 258 - 273.*

Lawal H.B (2003), "Categorical Data Analysis with SAS and SPSS Applications". *St Cloud University.*

Levin B. (2007), "Compound multinomial likelihood functions are uni - model: proof of a conjecture of I. J. Good". *Animals of statistics, 5, 79 - 87 [5.8.5]*

Mc cullagh P. and Nelder J. A. (1989), "Generalized linear models". *London; Chappman & Hall (11.1.1).*

Walleniius K.T. (1963) Biased Sampling. The noncentral hypergeometric probability distribution. *Technical report, Department of Statistics, Stanford University, Stanford, CA.*