MOMENTS ABOUT THE ARITHMETIC MEAN OF A BINOMIAL FREQUENCY DISTRIBUTION

W. J. Kirkham, *Oregon State College*

Although the most useful moments of a binomial distribution have been derived as a function of the parameters of the generating binomial for any binomial frequency series, a generalization of notation and procedure is well worth our consideration. The problem attempted in this paper is the calculation of the moments about the mean for the general frequency series of Table I.

TABLE I

The generalized binomial frequency series

x (item)	f (frequency)
0	$N \cdot {}_nC_0 p^0 q^n$
1	$N \cdot {}_nC_1 p^1 q^{n-1}$
2	$N \cdot {}_nC_2 p^2 q^{n-2}$
.	....
.	....
n	$N \cdot {}_nC_n p^n q^0$

In calculating the moments of a set of data about any value, it is often found convenient to use an arbitrary origin, define the moments about this value, and represent the desired moments in terms of those calculated. In the general binomial series, the origin of the x's is found to be the best arbitrary origin. These intermediate moments are

$$\nu_1 = \frac{\sum fx}{N} = M, \text{ arithmetic mean};$$

$$\nu_2 = \frac{\sum fx^2}{N}; \qquad (1)$$

$$\cdots\cdots\cdots\cdots\cdots$$

$$\cdots\cdots\cdots\cdots\cdots$$

$$\nu_n = \frac{\sum fx^n}{N}$$

where ν_i is the i^{th} moment.

The moments (μ's) about the mean are easily defined as functions of the ν's from fundamental definitions of the μ's. Denoting the i^{th} moment by μ_i, we have

$$\begin{aligned} \mu_1 &= \frac{\sum f(x-M)}{N} = 0\,, \\ \mu_2 &= \frac{\sum f(x-M)^2}{N} = \nu_2 - \nu_1^2\,, \\ \mu_3 &= \frac{\sum f(x-M)^3}{N} = \nu_3 - 3\nu_2\nu_1 + 2\nu_1^3\,, \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots \end{aligned} \tag{2}$$

In general,

$$\mu_n = \nu_n - {}_nC_1\nu_{n-1}\nu_1 + {}_nC_2\nu_{n-2}\nu_1^2 + \cdots + (-1)^{n-1}({}_nC_{n-1} - 1)\nu_1^n\,. \tag{3}$$

Or, if we let $\{\nu\}^n = \nu_n$, we may express the n^{th} moment by a simple notation.

$$\mu_n = \{\mu\}^n = \{\nu\}^n - {}_nC_1\{\nu\}^{n-1}\nu_1 + {}_nC_2\{\nu\}^{n-2}\nu_1^2 + \cdots = (\{\nu\} - \nu_1)^n\,. \tag{4}$$

Solving the equation for $\{\nu\}$,

$$\{\nu\} = \{\mu\} + \nu_1\,.$$

Raising both sides to the n^{th} power and substituting for the "brace" notation,

$$\nu_n = \mu_n + {}_nC_1\mu_{n-1}\nu_1 + {}_nC_2\mu_{n-2}\nu_1^2 + \cdots + \nu_1^n\,.$$

Whence

$$\mu_n = \nu_n - {}_nC_1\mu_{n-1}\nu_1 - {}_nC_2\mu_{n-2}\nu_1^2 - \cdots - \nu_1^n\,, \tag{5}$$

a semi-recursion formula.

The original formula for μ_n contained n moments or variables; and since there are only $(n - 2)$ of the μ's which are of lower order than μ_n, it is necessary to retain ν_n and ν_1 in (5). Since $\mu_1 = 0$, one term in the expansion of μ_n is zero. For instance, when $n = 5$, we have

$$\mu_5 = \nu_5 - 5\mu_4\nu_1 - 10\mu_3\nu_1^2 - 10\mu_2\nu_1^3 - \nu_1^5\,.$$

To calculate μ_k, it is necessary to calculate the ν's from ν_1 to ν_k. For the binomial series, these ν's are

$$\begin{aligned} \nu_1 &= 1npq^{n-1} + \frac{2(n)(n-1)}{1\cdot 2}p^2q^{n-2} + \frac{3(n)(n-1)(n-2)}{1\cdot 2\cdot 3}p^3q^{n-3} + \cdots \\ &= np\left[q^{n-1} + \frac{(n-1)}{1}p^1q^{n-2} + \frac{(n-1)(n-2)}{2!}p^2q^{n-3} + \cdots + p^{n-1}\right] \\ &= np(q+p)^{n-1} = np\,, \\ \nu_2 &= np\left[1\cdot q^{n-1} + \frac{2(n-1)}{1!}p^1q^{n-2} + 3\,\frac{(n-1)(n-2)}{2!}p^2q^{n-3} + \cdots + np^{n-1}\right], \end{aligned}$$

$$\nu_3 = np\left[1^2\cdot q^{n-1} + 2^2\frac{(n-1)}{1!}p^1q^{n-2} + 3^2\frac{(n-1)(n-2)}{2!}p^2q^{n-3} + \cdots + n^2p^{n-1}\right],$$

....

....

$$\nu_k = np\left[1^{k-1}q^{n-1} + 2^{k-1}\frac{(n-1)}{1!}p^1q^{n-2} + 3^{k-1}\frac{(n-1)(n-2)}{2!}p^2q^{n-3} + \cdots + n^{k-1}p^{n-1}\right].$$

In the simplified form of ν_k, the [ ] is the $(k-1)^{\text{th}}$ moment about -1 of the binomial series generated by the binomial $(q+p)^{n-1}$. Denoting this [ ] by $\nu'_{k-1}(n-1)$, the ν's can be expressed by the formula

$$\nu_k = np\nu'_{k-1}(n-1), \tag{6}$$

where ν' is a function of $(n-1)$ and $(k-1)$ while ν_k was a function of n and k.

Let us see how a ν' in ν_k can be defined in terms of the ν's of lower order than k. In finding this relationship, a consideration of the two series of Table II will be helpful.

TABLE II

x'	f	x	f
1	$N_{n-1}C_0p^0q^{n-1}$	0	$N_{n-1}C_0p^0q^{n-1}$
2	$N_{n-1}C_1p^1q^{n-2}$	1	$N_{n-1}C_1p^1q^{n-2}$
.	....	.	....
.	....	.	....
n	$N_{n-1}C_{n-1}p^nq^0$	$n-1$	$N_{n-1}C_{n-1}p^{n-1}q^0$

The [ ] in ν_k for Table I is equal to the $(k-1)^{\text{th}}$ moment of x' about $x' = 0$. Or

$$\nu_{k-1}, \text{ Table II}, x', = \nu_{k-1}, \text{ Table I}, = \nu'_{k-1}(n-1).$$

Also ν_{k-1} for x, Table II, is ν_{k-1} for the series generated by $(q-p)^{n-1}$.

The desired relationship between the ν's for the two series of Table II can be found by making use of the equations expressing the equality of the μ's for x and x'. Dropping the variable which shows the number of items, the same for the two series of Table II, in the notation, we have

$$\mu_2 = \mu'_2 = \nu_2 - \nu_1^2 = \nu'_2 - \nu'^2_2, \qquad \nu'_2 = \nu_2 - 2\nu_1(\nu_1 - \nu'_1) + (\nu_1 - \nu'_1)^2;$$

$$\mu_3 = \mu'_3 = \nu_3 - 3\nu_2\nu_1 + 2\nu_1^3 = \nu'_3 - 3\nu'_2\nu'_1 + 2\nu'^3_1,$$

$$\nu'_3 = \nu_3 - 3\nu_2\nu_1 + 2\nu_1^3 + 3\nu'_2\nu'_1 - 2\nu'^3_1.$$

Substituting the value of ν'_2 in the right member of ν'_3,

$$\nu'_3 = \nu_3 - 3\nu_2(\nu_1 - \nu'_1) + 3\nu_1(\nu_1 - \nu'_1)^2 - (\nu_1 - \nu'_1)^3.$$

In general,

$$\nu_k' = \nu_k - {}_kC_1\nu_{k-1}(\nu_1 - \nu_1') + {}_kC_2\nu_{k-2}(\nu_1 - \nu_1')^2 + \cdots + (-1)^k(\nu_1 - \nu_1')^k. \tag{7}$$

The formula just derived may be used to define the moments about any origin in terms of those about the original zero of the x's. For our immediate use, the formula simplifies since $\nu_1' = \nu_1 + \nu_0 = \nu_1 + 1$. Then

$$\nu_k' = \nu_k + {}_kC_1\nu_{k-1} + {}_kC_2\nu_{k-2} + {}_kC_3\nu_{k-3} + \cdots \tag{8}$$

By simple analysis we found the value of ν_1 to be np. By the method of continuation, we are able to extend the list of ν's to any number. ν' from (8) is used in (6) with n replaced by $(n - 1)$ in the ν's.

$\nu_0 = 1.$

$\nu_1 = np.$

$$\begin{aligned}
\nu_2 &= np\nu_1'(n-1) = np[\nu_1(n-1) + \nu_0(n-1)] \\
&= np[(n-1)p + 1] = n(n-1)p^2 + np.
\end{aligned}$$

$$\begin{aligned}
\nu_3 &= np\nu_2'(n-1) = np[\nu_2(n-1) + 2\nu_1(n-1) + \nu_0(n-1)] \\
&= n(n-1)(n-2)p^3 + 3n(n-1)p^2 + np.
\end{aligned}$$

$$\begin{aligned}
\nu_4 &= np\nu_3'(n-1) = np[\nu_3(n-1) + 3\nu_2(n-1) + 3\nu_1(n-1) + \nu_0(n-1)] \\
&= np\{[(n-1)(n-2)(n-3)p^3 + 3(n-1)(n-2)p^2 + (n-1)p] \\
&\qquad + 3\,[(n-1)(n-2)p^2 + (n-1)p] + 3\,[(n-1)p] + 1\}. \\
&= n(n-1)(n-2)(n-3)p^4 + 6(n)(n-1)(n-2)p^3 + 7(n)(n-1)p^2 + np.
\end{aligned}$$

....

If the order of the terms in the expansion is reversed, ν_n is an ascending power series in p. The pure numerical coefficients in some of these ν's are

$\nu_1 = (1)$

$\nu_2 = (1, 1)$

$\nu_3 = (1, 3, 1)$

$\nu_4 = (1, 7, 6, 1)$

$\nu_5 = (1, 15, 25, 10, 1)$

$\nu_6 = (1, 31, 90, 65, 15, 1)$

$\nu_7 = (1, 63, 301, 350, 140, 21, 1)$

$\nu_8 = (1, 127, 966, 1701, 1050, 266, 28, 1).$

In general,

$$\nu_{n+1} = \left(1, \sum_1^n {}_nC_i, \sum_2^n \left({}_nC_i \sum_1^{i-1} {}_{i-1}C_j\right), \right.$$
$$\left. \sum_3^n \left({}_nC_i \sum_2^{i-1} \left({}_{i-1}C_j \sum_1^{j-1} {}_{j-1}C_k\right)\right), \cdots\right). \tag{9}$$

Using the foregoing ν's, and the semi-recursion formula, we are able to determine the μ's.

$$\begin{aligned}
\mu_2 &= \nu_2 - \nu_1^2 \\
&= [np + (n)(n-1)p^2] - (np)^2 \\
&= np(1-p) \\
&= npq.
\end{aligned}$$

$$\begin{aligned}
\mu_3 &= \nu_3 - 3\nu_1\mu_2 - \nu_1^3 \\
&= [np + 3n(n-1)p^2 + (n)(n-1)(n-2)p^3] - 3(np)[np(1-p)] - [np]^3. \\
&= np + (-3n)p^2 + (2n)p^3 = np(1 - 3p + 2p^2) \\
&= np(1-p)(1-2p) \\
&= npq(q-p).
\end{aligned}$$

$$\begin{aligned}
\mu_4 &= [np + 7(n)(n-1)p^2 + 6(n)(n-1)(n-2)p^3 + (n)(n-1)(n-2) \\
&\qquad (n-3)p^4] - 4(np)(np)(1 - 3p + 2p^2) - 6(np)^2(np)(1-p) - (np)^4 \\
&= np + (-7n + 3n^2)p^2 + (12n - 6n^2)p^3 + (-6n + 3n^2)p^4 \\
&= np(1 - 7p + 12p^2 - 6p^3) + 3n^2p^2(1 - 2p + p^2) \\
&= npq - 6np^2q^2 + 3n^2p^2q^2.
\end{aligned}$$

$$\begin{aligned}
\mu_5 &= np(1 - 15p + 50p^2 - 60p^3 + 24p^4) + 10n^2p^2(1 - 4p + 5p^2 - 2p^3) \\
&= (q-p)(npq - 12np^2q^2 + 10n^2p^2q^2).
\end{aligned}$$

$$\begin{aligned}
\mu_6 &= np(1 - 31p + 180p^2 - 390p^3 + 360p^4 - 120p^5) + 5n^2p^2(5 - 36p \\
&\qquad + 83p^2 - 78p^3 + 26p^4) + 15n^3p^3(1 - 3p + 3p^2 - p^3) \\
&= npq - 30np^2q^2(q-p)^2 + 25n^2p^2q^2 - 130n^2p^3q^3 + 15n^3p^3q^3.
\end{aligned}$$

$$\begin{aligned}
\mu_7 &= np(1 - 63p + 602p^2 - 2100p^3 + 3360p^4 - 2520p^5 + 720p^6) \\
&\qquad + n^2p^2(56 - 686p + 2590p^2 - 4270p^3 + 3234p^4 - 924p^5) + n^3p^3(105 \\
&\qquad - 525p + 945p^2 - 735p^3 + 210p^4)
\end{aligned}$$

$$= (q - p)(npq - 60np^2q^2 + 360np^3q^3 + 56n^2p^2q^2 - 462n^2p^3q^3 + 105n^3p^3q^3).$$

$$\mu_8 = np(1 - 127p + 1932p^2 - 10206p^3 + 25200p^4 - 31920p^5 + 20160p^6$$
$$- 5040p^7) + n^2p^2(119 - 2394p + 13895p^2 - 35700p^3 + 46004p^4$$
$$- 29232p^5 + 7308p^6) + n^3p^3(490 - 3850p + 10990p^2 - 14770p^3$$
$$+ 9520p^4 - 2380p^5) + n^4p^4(105 - 420p + 630p^2 - 420p^3 + 105p^4)$$
$$= npq(1 - 42pq(3 - 40pq(1 - 3pq))) + 7n^2p^2q^2(17 - 4pq(77 - 261pq))$$
$$+ 70n^3p^3q^3(7 - 34pq) + 105n^4p^4q^4.$$