# ON THE BERNOULLI DISTRIBUTION*

BY SOLOMON KULLBACK

1. *Introduction.* The Bernoulli or binomial distribution, of central importance in the theory of mathematical probability and statistics, has been the subject of considerable study. The derivation of the moments, or recursion formulas for the moments, of this distribution, as usually presented, involve the use of moment-generating or characteristic functions, or the explicit form of the distribution itself.† In the following we will derive these moments in an elementary manner and extend the results to the Poisson exponential distribution, distributions of the Lexis and Poisson types, and the multinomial distribution.

2. *Preliminary Notions.* We need and use the following assumption, definitions, and theorem.

EMPIRICAL ASSUMPTION. If an event which can happen in two different ways be repeated a great number of times under the same essential conditions, the ratio of the number of times that it happens in one way, to the total number of trials, will approach a definite limit, as the latter number increases indefinitely.‡

DEFINITION. The limit described in the empirical assumption shall be called the probability that the event shall happen in the first way under those conditions.§ We shall express the fact

* Presented to the Society, November 30, 1934.

† See, for example, A. Fisher, *The Mathematical Theory of Probabilities*, 2d ed., p. 104 ff.; H. L. Rietz, *Mathematical Statistics*, 1927, p. 26 ff.; V. Mises, *Wahrscheinlichkeitsrechnung*, 1931, pp. 131–133; Risser and Traynard, *Les Principes de la Statistique Mathématique*, 1933, pp. 39–40 and 320–321; V. Romanovsky, *Note on the moments of the binomial* $(q+p)^n$ *about its mean*, Biometrika, vol. 15 (1923), pp. 410–412; A. T. Craig, *Note on the moments of a Bernoulli distribution*, this Bulletin, vol. 40 (1934), pp. 262–264; A. R. Crathorne, *Moments de la binomiale par rapport à l'origine*, Comptes Rendus, vol. 198 (1934), p. 1202; A. A. Krisknasuami Aygangar, *Note on the recurrence formulae for the moments of the point binomial*, Biometrika, vol. 26 (1934), pp. 262–264.

‡ J. L. Coolidge, *An Introduction to Mathematical Probability*, 1925, p. 4.

§ J. L. Coolidge, op. cit., p. 4.

that the relative frequency $f/s$ (the ratio of the number of times the event happens in one way to the total number of trials) is assumed to have the probability $p$ as a limit when $s\to\infty$, in abbreviated form by writing $E(f/s)=p$, where $E(f/s)$ is read "expected value of $f/s$."*

THEOREM OF COMPOUND PROBABILITY. If a compound event consists in the conjunction of any number of independent events, the probability of the compound event is the product of the probabilities for the individual events.†

3. *Bernoulli Distribution.* Suppose that in $n$ independent trials we observe an event to occur $x$ times. Then, applying the notions of §2, we have

$$
(1)\qquad
\begin{aligned}
&E\left(\frac{x}{n}\right)=p, \qquad E\left(\frac{x(x-1)}{n(n-1)}\right)=p^2, \cdots,\\
&E\left(\frac{x(x-1)\cdots(x-r+1)}{n(n-1)\cdots(n-r+1)}\right)=p^r, \cdots,\\
&E\left(\frac{n-x}{n}\right)=q, \qquad E\left(\frac{(n-x)(n-x+1)}{n(n-1)}\right)=q^2, \cdots,\\
&E\left(\frac{n-x)(n-x-1)\cdots(n-x-r+1)}{n(n-1)\cdots(n-r+1)}\right)=q^r, \cdots,\\
&E\left(\frac{x(x-1)\cdots(x-r+1)(n-x)(n-x-1)\cdots(n-x-s+1)}{n(n-1)(n-2)\cdots(n-r-s+1)}\right)\\
&\qquad\qquad=p^rq^s
\end{aligned}
$$

for $r+s\leqq n$. Each fraction in (1) represents the ratio of observed $r$-plets to the total possible number of such $r$-plets. We may also write (1) as

$$
(2)\qquad \{E(x/n)\}^r\{E((n-x)/n)\}^s = E([x]^r[n-x]^s/[n]^{r+s}),
$$

where $[x]^r$ is the factorial function‡ $x(x-1)(x-2)\cdots(x-r+1)$

* H. L. Rietz, op. cit., p. 9.

† J. L. Coolidge, op. cit., p. 18.

‡ Whittaker and Robinson, *The Calculus of Observations*, 1924, p. 8. See also J. F. Steffensen, *Factorial moments and discontinuous frequency functions*, Skandinavisk Aktuarietidsskrift, vol. 6 (1923), pp. 73–89.

and $\{E(x/n)\}^r$ is the $r$th power of the expected value of $x/n$, and $r+s \leqq n$.

In order to find the expected value of the $r$th power of the variable (or the $r$th moment about the origin), we avail ourselves of the fact that

$$(3) \qquad x^r = a_{r,0}[x]^r + a_{r-1,1}[x]^{r-1} + \cdots + a_{0,r},$$

where $a_{i,j}$, $(i, j = 0, 1, 2, \cdots)$, is the element in the $i$th row and $j$th column of the table*

| | | | | | |
|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | $0 \cdots$ |
| 1 | 1 | 1 | 1 | 1 | $1 \cdots$ |
| 1 | 3 | 7 | 15 | 31 | $63 \cdots$ |
| 1 | 6 | 25 | 90 | 301 | $966 \cdots$ |
| 1 | 10 | 65 | 350 | $1701 \cdots$ | |
| 1 | 15 | 140 | $1050 \cdots$ | | |
| $\cdots$ | $\cdots$ | $\cdots$ | $\cdots$ | $\cdots$ | $\cdots$ |

in which $a_{i,j} = ia_{i,j-1} + a_{i-1,j}$, $a_{0,j} = 0$, $(j = 0, 1, \cdots)$, $a_{i,0} = 1$, $(i = 1, 2, \cdots)$. We thus have

$$(4) \qquad \begin{aligned} E(x^r) = a_{r,0}\frac{[n]^r}{n^r}\{E(x)\}^r + a_{r-1,1}\frac{[n]^{r-1}}{n^{r-1}}\{E(x)\}^{r-1} \\ + \cdots + a_{1,r-1}E(x), \end{aligned}$$

or

$$(5) \quad E(x^r) = a_{r,0}[n]^r p^r + a_{r-1,1}[n]^{r-1}p^{r-1} + \cdots + a_{1,r-1}np.$$

For the expected value of the $r$th power of the deviations of $x$ from its expected value (or the $r$th moment about the expected value), we have

$$(6) \qquad E\{(x - E(x))^r\} = \sum_{i=0}^{r} (-1)^i C_{r,i} E(x^{r-i})\{E(x)\}^i,$$

in which we replace $E(x^r)$ by its value as given in (5). This yields

* Whittaker and Robinson, op. cit., p. 9.

$$(7)\qquad E\{(x - E(x))^r\} = B_r p^r + B_{r-1}p^{r-1} + \cdots + B_1 p,$$

where

$$B_r = \sum_{i=0}^{r} (-1)^i C_{r,i} a_{r-i,0} n^i [n]^{r-i},$$

$$B_{r-j} = \sum_{i=0}^{r-j-1} (-1)^i C_{r,i} a_{r-j-i,j} n^i [n]^{r-j-i}.$$

4. *Poisson Exponential Distribution.* The moments of the Poisson exponential distribution are derived from those of the Bernoulli distribution by letting $n \to \infty$ with $\lim_{n\to\infty} E(x) = \lim_{n\to\infty} np = m$, where $m$ is finite. Accordingly, we have from (4) and (7), respectively,

$$(8)\qquad E(x^r) = a_{r,0}m^r + a_{r-1,1}m^{r-1} + \cdots + a_{1,r-1}m,$$

$$(9)\qquad E\{(x - E(x))^r\} = A_r m^r + A_{r-1}m^{r-1} + \cdots + A_1 m,$$

where

$$A_r = \sum_{i=0}^{r} (-1)^i C_{r,i} a_{r-i,0}, \qquad A_{r-j} = \sum_{i=0}^{r-j-1} (-1)^i C_{r,i} a_{r-j-i,j}.$$

5. *Lexis Type Distribution.** Consider $N$ independent sets of $n$ independent trials each, with the probability for the occurrence of an event constant for a set, but varying from set to set. Let the varying probability be $p_1, p_2, \cdots, p_N$, and the number of occurrences observed, respectively, $x_1, x_2, \cdots, x_N$.

Then, in a manner similar to that of §3, we have

$$(10)\quad E([x]^{(r)}/[n]^r) = \frac{1}{N}\sum_{i=1}^{N} p_i^r, \quad \text{where} \quad [x]^{(r)} = \frac{1}{N}\sum_{i=1}^{N} [x_i]^r.$$

If we set $p^{(r)} = \sum_{i=1}^{N} p_i^r/N$, we have, corresponding to (5),

$$(11)\qquad \begin{aligned} E(x^{(r)}) = {} & a_{r,0}[n]^r p^{(r)} + a_{r-1,1}[n]^{r-1} p^{(r-1)} + \cdots \\ & + a_{1,r-1} n p^{(1)}, \end{aligned}$$

where $x^{(r)} = \sum_{i=1}^{N} x_i^r/N$, and corresponding to (7),

$$(12)\quad E\{(x^{(1)} - E(x^{(1)}))^r\} = B_r p^{(r)} + B_{r-1}p^{(r-1)} + \cdots + B_1 p^{(1)}.$$

Since†

---

* H. L. Rietz, op. cit., Chap. 6.

† Chrystal, *Textbook of Algebra*, 2nd ed., vol. 2, p. 49.

$$\frac{1}{N}\sum_{i=1}^{N} p_i^r \geqq \left(\frac{1}{N}\sum_{i=1}^{N} p_i\right)^r,$$

we see, by comparing (5) and (11), that

$$E_L(x^r) \geqq E_B(x^r), \tag{13}$$

where $E_L$ represents the expected value from a Lexis type distribution and $E_B$ the corresponding expected value from a Bernoulli distribution with $p=\sum_{i=1}^{N} p_i/N$. Indeed $E_L\{(x+a)^r\} \geqq E_B\{(x+a)^r\}$ for $a \geqq 0$.

6. *Poisson Type Distribution.** Consider a set of $n$ independent trials with the probability for the occurrence of an event varying from trial to trial. Let these probabilities be $p_1, p_2, \cdots, p_n$. The study of the moments for this case is really the consideration of the moments for the distribution of the *total* of a Lexis type distribution of $n$ independent sets of one trial each. Thus, if the number of observed occurrences of an event in $x$ trials is $x=x_1+x_2+\cdots+x_n$, where $x_i=$ zero or one, then

$$E(x) = \sum_{i=1}^{n} E(x_i) = \sum_{i=1}^{n} p_i, \tag{14}$$

$$E(x^r) = a_1^{(r)}\sum p_i + a_2^{(r)}\sum p_i p_j + a_3^{(r)}\sum p_i p_j p_k + \cdots + a_r^{(r)}\sum p_i p_j \cdots p_r, \tag{15}$$

where $a_k^{(r)} = \sum r!/(s_1!s_2!\cdots s_k!)$ for $s_1+s_2+\cdots+s_k=r$, $s_i \neq 0$, $(i=1, 2, \cdots, k)$. The right member of (15) follows from the fact that $E(x_i^r)=p_i$, as may be readily seen from (4) for $n=1$ and $r \geqq 2$.

For the moments about the expected value, we have

$$E\{(x - E(x))^r\} = \sum_{i=0}^{r} (-1)^i C_{r,i} E(x^{r-i})\{E(x)\}^i, \tag{16}$$

where $E(x^r)$ is given by (15) and $\{E(x)\}^r = (p_1+p_2+\cdots+p_n)^r$.

If we repeat the preceding argument with constant probability $p=\sum_{i=1}^{n} p_i/n$, we find that

$$E(x^r) = a_1^{(r)} np + a_2^{(r)} C_{n,2} p^2 + \cdots + a_r^{(r)} C_{n,r} p^r. \tag{17}$$

* H. L. Rietz, op. cit., Chap. 6.

Since*

$$C_{n,r}\left(\frac{1}{n}\sum_{i=1}^{n} p_i\right)^r \geqq \sum p_i p_j \cdots p_r,$$

we see by comparing (17) and (15) that

$$E_B(x^r) \geqq E_P(x^r), \tag{18}$$

where $E_P$ represents the expected value from a Poisson type distribution and $E_B$ the corresponding expected value from a Bernoulli distribution with $p=\sum_{1=i}^{n} p_i/n$. Indeed, $E_B\{(x+a)^r\} \geqq E_P\{(x+a)^r\}$ for $a \geqq 0$.

A comparison of (17) and (5) shows that

$$C_{n,k}a_k^{(r)} = [n]^k a_{k,r-k}, \qquad \text{or} \qquad a_k^{(r)} = k!a_{k,r-k}. \tag{19}$$

7. *The Multinomial Distribution.* Consider a trial in which one of $r$ mutually exclusive events may occur. Let the respective probabilities of occurrence be $p_1, p_2, \cdots, p_r$ with $p_1+p_2+\cdots+p_r=1$. If in $n$ independent trials there are observed $x_1, x_2, \cdots, x_r$ occurrences, respectively, of the $r$ events, then

$$E\{[x_1]^\alpha [x_2]^\beta \cdots [x_r]^\rho / [n]^{\alpha+\beta+\cdots+\rho}\} = p_1^\alpha p_2^\beta \cdots p_r^\rho. \tag{20}$$

Thus, for example, $E(x_i x_j)=n(n-1)p_i p_j$ and $E(x_i x_j)-E(x_i)E(x_j) = n(n-1)p_i p_j - np_i np_j = -np_i p_j$.

Now we may write†

$$x_i^l = \sum_{\lambda=0}^{l} \Delta^\lambda 0^l [x_i]^\lambda / \lambda!,$$

where $\Delta^\lambda 0^l$ are the differences of zero and

$$\Delta^\lambda 0^l = \lambda^l - \lambda(\lambda-1)^l + \frac{1}{2!}\lambda(\lambda-1)(\lambda-2)^l - \cdots \pm \lambda\cdot 1^l.$$

* This can be shown by using the fact that $f(x_1, x_2, \cdots, x_n)=\sum x_1x_2\cdots x_r$ subject to the condition $x_1+x_2+\cdots+x_n=$constant, has its maximum value for $x_1=x_2=\cdots=x_n$.

† J. F. Steffensen, *Interpolation*, 1927, p. 13, equation (3).

Accordingly,

$$(21) \qquad x_1^k x_2^l \cdots x_r^m = \sum_{\kappa=0}^{l} \frac{\Delta^\kappa 0^k [x_1]^\kappa}{\kappa!} \sum_{\lambda=0}^{l} \frac{\Delta^\lambda 0^l [x_2]^\lambda}{\lambda!} \cdots \sum_{\mu=0}^{m} \frac{\Delta^\mu 0^m [x_r]^\mu}{\mu!},$$

and

$$(22) \qquad E(x_1^k x_2^l \cdots x_r^m) = \sum_{\kappa=0,\lambda=0,\cdots,\mu=0}^{k,l,\cdots,m} \frac{\Delta^\kappa 0^k \cdot \Delta^\lambda 0^l \cdots \Delta^\mu 0^m [n]^{\kappa+\lambda+\cdots+\mu} p_1^\kappa p_2^\lambda \cdots p_r^\mu}{\kappa!\lambda! \cdots \mu!}$$
$$= (1 + p_1\Delta_1 + p_2\Delta_2 + \cdots + p_r\Delta_r)^n \cdot 0_1^k 0_2^l \cdots 0_r^m,$$

where $\Delta_i^k 0_j^l = 1$ for $i \neq j$ and $\Delta_i^k 0_i^l = \Delta^k 0^l$. The latter result is true for any $n$ and $k, l, \cdots, m$ since $[n]^{\kappa+\lambda+\cdots+\mu} = 0$ for $\kappa+\lambda+\cdots+\mu > n$ and $\Delta^\alpha 0^r = 0$ for $\alpha > r$.

The procedure with respect to the product moments about any constant is similar to that outlined. Thus, since* $(x_i - a_i)^l = \sum_{\lambda=0}^{l} \Delta^\lambda(-a_i)^l [x_i]^\lambda / \lambda!$, we have

$$(23) \qquad (x_1 - a_1)^k (x_2 - a_2)^l \cdots (x_r - a_r)^m = \sum_{\kappa=0,\lambda=0,\cdots,\mu=0}^{k,l,\cdots,m} \frac{\Delta^\kappa(-a_1)^k \Delta^\lambda(-a_2)^l \cdots \Delta^\mu(-a_r)^m [x_1]^\kappa [x_2]^\lambda \cdots [x_r]^\mu}{\kappa!\lambda! \cdots \mu!},$$

and

$$(24) \qquad E\{(x_1 - a_1)^k (x_2 - a_2)^l \cdots (x_r - a_r)^m\} = (1 + p_1\Delta_1 + p_2\Delta_2 + \cdots + p_r\Delta_r)^n \cdot (-a_1)^k (-a_2)^l \cdots (-a_r)^m,$$

where $a_1, a_2, \cdots, a_r$ are any constants and $\Delta_i^k(-a_j)^l = 1$ for $i \neq j$ and $\Delta_i^k(-a_i)^l = \Delta^k(-a_i)^l$. Since† $\Delta = e^D - 1$, where $D = d/dx$,

* J. F. Steffensen, op. cit., p. 13.

† Whittaker and Robinson, op. cit., p. 63.

$$
\begin{aligned}
(25)\qquad & E\{(x_1 - a_1)^k (x_2 - a_2)^l \cdots (x_r - a_r)^m\} \\
& = \left.(p_1 e^{D_1} + p_2 e^{D_2} + \cdots + p_r e^{D_r})^n \cdot x_1^k x_2^l \cdots x_r^m \right]_{\substack{x_1 = -a_1 \\ \vdots \\ x_r = -a_r}},
\end{aligned}
$$

where $D_1 = \partial/\partial x_1$.

WASHINGTON, D. C.

---

# ON A RESULTANT CONNECTED WITH FERMAT'S LAST THEOREM

BY EMMA LEHMER

E. Wendt* seems to have been the first to introduce the resultant of $x^n = 1$ and $(x+1)^n = 1$ in connection with Fermat's Last Theorem. This resultant can be expressed by means of the following circulant of binomial coefficients

$$
\Delta_n = \begin{vmatrix} 1 & C_{n,1} & C_{n,2} & \cdots & C_{n,n-1} \\ C_{n,n-1} & 1 & C_{n,1} & \cdots & C_{n,n-2} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ C_{n,1} & C_{n,2} & C_{n,3} & \cdots & 1 \end{vmatrix}.
$$

In his book on Fermat's Last Theorem Bachmann† proved that if $p$ is an odd prime and if $\Delta_{p-1}$ is not divisible by $p^3$, then Fermat's equation $x^p + y^p + z^p = 0$ has no solution $(x, y, z)$ prime to $p$.

S. Lubelsky‡ proved in a recent paper, using the distribution of quadratic residues, that if $p \geqq 7$, $\Delta_{p-1}$ is not only divisible by $p^3$, but by $p^8$, thus annulling Bachmann's criterion except for $p = 3$ and $p = 5$.

We shall now show how, by a straightforward manipulation with the above determinant, one can prove much more.

THEOREM 1. *$\Delta_{p-1}$ is divisible by $p^{p-2} q_2$ for every prime $p$, where $q_2$ is the Fermat quotient $(2^{p-1} - 1)/p$.*

---

* Journal für Mathematik, vol. 113 (1894), pp. 335–347.

† *Das Fermatproblem*, 1919, p. 59.

‡ Prace Matematyczno-Fizyczne, vol. 42 (1935), pp. 11–44.