ON THE PROBABILITY DISTRIBUTION OF
X MODULO n

Technical Report No. 20

Department of Navy

Office of Naval Research

Contract No. Nonr-409(39)
Project No. (NR 042-212)

BIOMETRICS UNIT
DEPARTMENT OF PLANT BREEDING

NEW YORK STATE COLLEGE OF AGRICULTURE

CORNELL UNIVERSITY

ITHACA, NEW YORK

ON THE PROBABILITY DISTRIBUTION OF X MODULO n

Technical Report No. 20

Department of Navy

Office of Naval Research

Contract No. Nonr-409(39)
Project No. (NR 042-[illegible]12)

H. E. Schaffer and D. S. Robson

Biometrics Unit

New York State College of Agriculture

Cornell University

Ithaca, New York

This work was supported in part by the Office of Naval Research.

On the Probability Distribution of X modulo n

H. E. Schaffer and D. S. Robson
Cornell University

BU-170-M October, 1964

ABSTRACT

If X is a non-negative discrete random variable with probability generating function $Q_x(t)$ then

$$P(X \underline{\text{mod}}\ n = z) = \frac{1}{n} \sum_{i=0}^{n-1} \frac{Q_x(\omega_i)}{\omega_i^z}$$

where $\omega_0, \omega_1, \cdots, \omega_{n-1}$ are the n^{th} roots of unity. The form of this distribution is illustrated for the Poisson case, $Q_x(t) = \exp \lambda(1-t)$.

On the Probability Distribution of X modulo n *

H. E. Schaffer and D. S. Robson
Cornell University

BU-170-M October, 1964

The probability distribution of Z_n = X modulo n is pertinent in problems where a total of X items is portioned into lots of n items each, leaving a remainder of Z_n items. A problem of similar nature arises in computing the probability distribution of the "excess over the boundary" or "end effects" in sequential tests of hypotheses where the experimental observations are produced in lots of size n instead of singly. In this case, if the random variable X denotes the terminal sample size for a sequential test in which the observations are taken singly then, unless X mod n = 0, there will be an excess of $n - Z_n$ observations. In its simplest form this problem also arises in genetic linkage analysis; if X denotes the number of chromosomal crossovers occuring between two linked heterozygous loci then the gametic frequencies are determined by the probability distribution of X mod 2.

If $Q_X(t)$ is the probability generating function for X,

$$Q_X(t) = \sum_{x=0}^{\infty} t^x P\{X = x\} \quad , \tag{1}$$

then

$$P\{X \text{ mod } n = z\} = \frac{1}{n} \sum_{i=0}^{n-1} \frac{Q_X(\omega_i)}{\omega_i^z} \tag{2}$$

where $\omega_0, \omega_1, \cdots, \omega_{n-1}$ are the n^{th} roots of unity.

---

* Biometrics Unit, Cornell University, Ithaca, New York

Lemma: If $\omega_0,\cdots,\omega_{n-1}$ are the n^{th} roots of unity then

$$\sum_{j=0}^{n-1} \omega_j^r = 0 \quad \text{for} \quad r=1,2,\cdots,n-1 \quad .$$

The lemma follows directly from the representation $\{\omega_0,\omega_1,\cdots,\omega_{n-1}\} = \{1,\xi,\xi^2,\cdots,\xi^{n-1}\}$ where ξ is a primitive n^{th} root of unity*, for then

$$\sum_{j=0}^{n-1} \omega_j^r = \sum_{j=0}^{n-1} \xi^{rj} = \frac{1-\xi^{rn}}{1-\xi} = 0$$

Applying the lemma to (2) then gives

$$\frac{1}{n}\sum_{j=0}^{n-1} \frac{Q_X(\omega_i)}{\omega_i^z} = \sum_{\nu=0}^{n-1} \sum_{k=0}^{\infty} P\{X = kn + z + \nu\} \left[\frac{1}{n} \sum_{j=0}^{n-1} \omega_i^{kn+\nu} \right]$$

$$= \sum_{k=0}^{\infty} P\{X = kn + z\}$$

$$= P\{X \bmod n = z\} \quad .$$

---

* This proof of the lemma was called to our attention by B. L. Raktoe.

The factorial moments of Z_n may be computed from the probability generating function

$$(4) \qquad Q_{Z_n}(t) = \frac{1-t^n}{n} \sum_{i=0}^{n-1} \frac{\omega_i}{\omega_i - t} Q_x(\omega_i) \quad .$$

In particular the mean value of Z_n is given by

$$Q'_{Z_n}(1) = \frac{n-1}{2} + \sum_{i=1}^{n-1} \frac{\omega_i}{1-\omega_i} Q_x(\omega_i) \quad .$$

Thus, if X were sample size in a sequential experiment as mentioned earlier, with sequential lots of n observations each the expected excess over the boundary would be

$$n[1-Q_{Z_n}(0)] - Q'_{Z_n}(1) = \frac{n-1}{2} - \sum_{i=1}^{n-1} \frac{Q_x(\omega_i)}{1-\omega_i}$$

Poisson variable modulo n. To illustrate an explicit form of the distribution of Z_n we consider the case

$$Q_x(t) = e^{-\lambda(1-t)} \quad .$$

Using standard trigonometric identities with the n^{th} roots of unity represented by

$$\omega_k = \cos \frac{2k\pi}{n} + i \sin \frac{2k\pi}{n} = e^{\frac{2k\pi i}{n}} , \quad k=0,1,\cdots,n-1$$

and noting that $\omega_0 = 1$ and that the reciprocal of a root of unity ω_k is its complex conjugate ω_{n-k}, we obtain for n even

$$P\{Z_{2m}=z\} = \frac{1}{n}\left[1 - e^{-2\lambda} + 2e^{-\lambda}\sum_{k=1}^{m-1}\cos(\lambda \sin\frac{\pi k}{m} - z\,\frac{\pi k}{m})\, e^{\lambda\cos\frac{\pi k}{m}}\right]$$

$$E(Z_{2m}) = \frac{n-1-e^{-2\lambda}}{2} - \sum_{k=1}^{m-1} e^{-\lambda(1-\cos\frac{\pi k}{m})}\left[\frac{\sin(\lambda\sin\frac{\pi k}{m} + \frac{\pi k}{m}) + \sin(\lambda\sin\frac{\pi k}{m})}{\sin\frac{\pi k}{m}}\right]$$

and for n odd

$$P\{Z_{2m+1}=z\} = \frac{1}{n}\left[1+2e^{-\lambda}\sum_{k=1}^{m}\cos(\lambda\sin\frac{2k\pi}{n} - z\,\frac{2k\pi}{n})e^{\lambda\cos\frac{2k\pi}{n}}\right]$$

$$E(Z_{2m+1}) = \frac{n-1}{2} - \sum_{k=1}^{m} e^{-\lambda(1-\cos\frac{2\pi k}{n})}\left[\frac{\sin(\lambda\sin\frac{2\pi k}{n} + \frac{2\pi k}{n}) + \sin(\lambda\sin\frac{2\pi k}{n})}{\sin\frac{2\pi k}{n}}\right]$$

In the limit, as $\lambda \to \infty$, these reduce to

$$P\{Z_n = z\} = \frac{1}{n}$$

$$E(Z_n) = \frac{n-1}{2}$$

If the limit is taken as $n \to \infty$ then Z_n of course is distributed as X.

Acknowledgments. The authors wish to thank I. R. Savage for providing them with a substantial start on this problem. Assistance in obtaining explicit results for the Poisson case was provided by K. Choi.