THE POISSON APPROXIMATION TO THE POISSON BINOMIAL DISTRIBUTION[1]

BY J. L. HODGES, JR. AND LUCIEN LE CAM

University of California, Berkeley

1. Introduction. It has been observed empirically that in many situations the number S of events of a specified kind has approximately a Poisson distribution. As examples we may mention the number of telephone calls, accidents, suicides, bacteria, wars, Geiger counts, Supreme Court vacancies, and soldiers killed by the kick of a horse.

Many textbooks in probability content themselves with an explanation of this phenomenon that runs something like this: There is a large number, say n, of events that might occur—for example, there are many telephone subscribers who might place a call during a given minute. The chance, say p, that any specified one of these events will occur (e.g., that a specified telephone subscriber will call), is small. Assuming that the events are independent, S has exactly the binomial distribution, say $\mathcal{B}(n, p)$. If we now let $n \to \infty$ and $p \to 0$, so that $np \to \lambda$ where λ is fixed and $0 < \lambda < \infty$, it is shown that $\mathcal{B}(n, p)$ tends to the Poisson distribution $\mathcal{P}(\lambda)$ with expectation λ.

As was pointed out by von Mises [4], such an explanation is often not satisfactory because the various trials cannot in many applications reasonably be regarded as equally likely to succeed. Let p_i denote the success probability of the ith trial, $i = 1, 2, \cdots, n$. Then S has the distribution sometimes called "Poisson binomial." Starting from this more realistic model von Mises shows that S has in the limit the distribution $\mathcal{P}(\lambda)$, provided $n \to \infty$ and the p_i vary with n in such a way that $\Sigma p_i = \lambda$ is fixed and $\alpha = \max\{p_1, p_2, \cdots, p_n\}$ tends to 0. This result is given in a few textbooks [1], [5].

The limit theorem of von Mises suggests that the Poisson approximation will be reliable provided that n is large, α is small, and λ is moderate. But even these requirements are unnecessarily restrictive, as may be seen from a general approximation theorem of Kolmogorov [2]. When this theorem is applied to our problem, it asserts that there is some constant C, independent of n and the p_i, such that the maximum absolute difference D between the cumulative distributions of S and of $\mathcal{P}(\Sigma p_i)$ satisfies the bound $D \leqq C \sqrt[5]{\alpha}$. Thus, the Poisson approximation will be good provided only α is small, whether n is small or large, and whatever value Σp_i may have. It seems to us that this is the type of theorem that best "explains" the empirical phenomenon of the "law of small numbers."

The purpose of our note is to present an elementary and relatively simple

Received November 6, 1959.

[1] This paper was prepared with the partial support of the Office of Naval Research (Nonr-222-43). This paper in whole or in part may be reproduced for any purpose of the United States Government.

proof of a bound of Kolmogorov's type. By using special features of the Poisson distribution, we are able to get the improved bound $3\sqrt[3]{\alpha}$ for D, and to accomplish this in a good deal simpler way than is required for the general result. We believe that our proof is suitable for presentation to an introductory class in probability theory.

2. The approximation theorems. Let X_i indicate success on the ith trial, so that $P(X_i = 1) = p_i$ and $P(X_i = 0) = 1 - p_i$. Our proofs will be based on the device of introducing random variables Y_i that have the Poisson distribution with $E(Y_i) = p_i$, and are such that $P(X_i = Y_i)$ is as large as possible. Specifically, we give to X_i and Y_i the joint distribution according to which

$$P(X_i = Y_i = 1) = p_i e^{-p_i}, \qquad P(X_i = 1, Y_i = 0) = p_i(1 - e^{-p_i}),$$
$$P(X_i = Y_i = 0) = e^{-p_i} - p_i(1 - e^{-p_i}),$$

and

$$P(X_i = 0, Y_i = y) = p_i^y e^{-p_i}/y! \qquad \text{for } y = 2, 3, \cdots .$$

We let the Y_i be independent of each other. (The construction is valid if $p_i \leqq 0.8$, insuring $P(X_i = Y_i = 0) \geqq 0$. For $p_i > 0.8$ the results below are trivially correct.)

From the familiar additive property of Poisson variables, we know that $T = \Sigma Y_i$ has exactly the Poisson distribution $\mathcal{P}(\Sigma p_i)$. Our objective is to show that $S = \Sigma X_i$ has nearly this distribution. Specifically, if we let

$$D = \sup_u | P(S \leqq u) - P(T \leqq u) |$$

denote the maximum absolute difference between the cumulatives of S and T, we want to find conditions under which D is small.

THEOREM 1. $D \leqq 2\Sigma p_i^2$.

Using the inequality $e^{-p_i} \geqq 1 - p_i$, it is easy to check that

$$P(X_i \neq Y_i) = 1 + p_i - (1 + 2p_i)e^{-p_i} \leqq 2p_i^2.$$

Therefore, by Boole's inequality, $P(S \neq T) \leqq \Sigma P(X_i \neq Y_i) \leqq 2\Sigma p_i^2$. But since $| P(S \leqq u) - P(T \leqq u) | \leqq P(S \neq T)$, the theorem follows.

In order to prove our next theorem, we shall need a uniform bound on the individual terms $p(k, \lambda) = e^{-\lambda}\lambda^k/k!$ of the Poisson distribution. It is well known that for large λ, the maximum term is of the order $\lambda^{-\frac{1}{2}}$, but we will give a specific upper bound.

LEMMA. *The maximum term of the distribution* $\mathcal{P}(\lambda)$ *is less than* $(1 + 1/12\lambda)/(2\pi\lambda)^{\frac{1}{2}}$.

PROOF. Suppose $k \leqq \lambda < k + 1$. The maximum term is then $e^{-\lambda}\lambda^k/k!$, as may be seen by looking at the ratio of successive terms. Since $(\lambda)^{\frac{1}{2}}e^{-\lambda}\lambda^k$ is maximized at $\lambda = k + \frac{1}{2}$, and since $1 + 1/12\lambda > 1 + 1/12(k + 1)$, it will suffice to show that

$$(2\pi)^{\frac{1}{2}}(k + \tfrac{1}{2})^{k+\frac{1}{2}}e^{-k-\frac{1}{2}} < k![1 + 1/12(k + 1)]$$

for $k = 0, 1, 2, \cdots$. This inequality may easily be checked by direct computation for $k = 0$, 1, and 2, and for $k \geqq 3$ by using the Stirling bound

$$k! > (2\pi)^{\frac{1}{2}} k^{k+\frac{1}{2}} e^{-k+(1/12k)-(1/360k^3)}.$$

Let us denote Σp_i by λ and Σp_i^2 by μ.

THEOREM 2. $D \leqq (3\mu/a^2) + (a+1)(1+1/12\lambda)/(2\pi\lambda)^{\frac{1}{2}}$.

To prove this, we shall consider the random variables $Z_i = Y_i - X_i$.

$$E(Z_i) = 0,$$

while

$$\mathrm{Var}(Z_i) = E(Z_i^2) = p_i(1-e^{-p_i}) + \sum_{k=2}^{\infty} k^2(p_i^k e^{-p_i})/k!$$

$$= p_i(1-e^{-p_i}) + E(Y_i^2) - p_i e^{-p_i} = p_i^2 + 2p_i(1-e^{-p_i}) \leqq 3p_i^2.$$

Let $\Sigma Z_i = U$. Then $E(U) = 0$ and $\mathrm{Var}(U) \leqq 3\mu$.

Let a be any positive number. If $T = S + U \leqq v - a$, then either $S \leqq v$ or $U \leqq -a$, so that $P(T \leqq v - a) \leqq P(S \leqq v) + P(U \leqq -a)$ and

$$P(T \leqq v) - P(S \leqq v) \leqq P(v - a \leqq T \leqq v) + P(U \leqq -a).$$

Similarly, if $S = T - U \leqq v$, then either $T \leqq v + a$ or $U \geqq a$, so that

$$P(S \leqq v) \leqq P(T \leqq v + a) + P(U \geqq a)$$

and $P(S \leqq v) - P(T \leqq v) \leqq P(v \leqq T \leqq v + a) + P(U \geqq a)$. Combining, we see that

$$D = \sup_v |P(S \leqq v) - P(T \leqq v)|$$

$$\leqq \sup_v P(v \leqq T \leqq v + a) + P(|U| \geqq a).$$

By the Chebycheff inequality, $P(|U| \geqq a) \leqq \mathrm{Var}(U)/a^2 \leqq 3\mu/a^2$. Using the lemma, we see that

$$\sup_v P(v \leqq T \leqq v + a) \leqq (a+1)(1+(1/12\lambda))/(2\pi\lambda)^{\frac{1}{2}},$$

since there are at most $a + 1$ Poisson terms in the interval from v to $v + a$. This completes the proof.

We now combine Theorems 1 and 2 to obtain our main result.

THEOREM 3. $D \leqq 3\sqrt[3]{\alpha}$.

We prove this by considering two cases. If $2\mu \leqq 3\sqrt[3]{\alpha}$, the theorem is an immediate consequence of Theorem 1. On the other hand, if $2\mu > 3\sqrt[3]{\alpha}$, we have by virtue of $\mu \leqq \alpha\lambda$ the inequality $\lambda > 3/2\alpha^{2/3} > 1$. Now suppose that $a \geqq 1$. Then

$$[(a+1)(1+(1/12\lambda))]/[(2\pi\lambda)^{\frac{1}{2}}] < a/\lambda^{\frac{1}{2}}$$

and

$$D \leqq (3\mu/a^2) + (a/\lambda^{\frac{1}{2}}).$$

This is minimized when $a = \sqrt[3]{6\mu(\lambda)^{\frac{1}{2}}} = a_0$. Since $\lambda > \mu/\alpha$ and $\mu > 3(\alpha)^{\frac{1}{2}}/2$, we see that $\mu(\lambda)^{\frac{1}{2}} > (\frac{3}{2})^{3/2}$ or $a_0 > (3^5/2)^{1/6} > 1$, so the restriction $a \geqq 1$ is satisfied, and the theorem is proved.

3. Remarks.

(i) We have presented our results as approximation theorems rather than as limit theorems. We believe it is better pedagogy to do so, since in the applications there will be definite values of n and the p_i, which are not "tending" to anything. However, if limit theorems are desired they follow at once. For example, Theorem 1 implies that $D \to 0$ as $\mu = \Sigma p_i^2 \to 0$, whereas Theorem 3 implies that $D \to 0$ as $\alpha = \max \{p_1, \cdots, p_n\} \to 0$.

(ii) Our Theorem 1 gives a simple and elementary proof of the standard textbook result that $\mathcal{B}(n, p) \to \mathcal{P}(\lambda)$ as $n \to \infty$, $p \to 0$, and $np \to \lambda$, since under these conditions $\Sigma p_i^2 = \lambda p \to 0$. Furthermore, Theorem 1 implies the more realistic theorem of von Mises, since if $\Sigma p_i = \lambda$ is fixed while $\alpha \to 0$, we must have $\Sigma p_i^2 \leqq \alpha \Sigma p_i = \alpha\lambda \to 0$.

(iii) As is customarily the case with bounds for the accuracy of approximations, our bound has only theoretical interest, being much too crude for practical usefulness. By pushing the method of proof, the constant factor 3 in the inequality $D \leqq 3\sqrt[3]{\alpha}$ can be reduced, but the result would still be of only theoretical value. It can be shown [3], using a much less simple argument, that $D \leqq 9\alpha$. While it is clearly a theoretical improvement to have a bound of order α rather than one of order $\sqrt[3]{\alpha}$, even the bound 9α is of limited applicational use. Fortunately, approximations are usually found in practice to be much better than the known bounds would indicate them to be.

(iv) The condition that $\alpha \to 0$ is sufficient but not necessary for $D \to 0$. It is easy to see that S will have approximately a Poisson distribution even if a few of the p_i are quite large, provided these values contribute only a small part of the total Σp_i.

REFERENCES

[1] William Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 1, 2nd ed., New York, John Wiley and Sons, 1957.

[2] A. N. Kolmogorov, "Deux théorèmes asymptotiques uniformes pour sommes des variables aléatoires," *Teor. Veroyatnost. i Primenen.*, Vol. 1 (1956), pp. 426–436.

[3] Lucien Le Cam, "An approximation theorem for Poisson binomial distributions," to be published in the *Pacific J. Math.*

[4] R. v. Mises, "Über die Wahrscheinlichkeit seltener Ereignisse," *Z. Angew. Math. Mech.*, Vol. 1 (1921), pp. 121–124.

[5] Richard v. Mises, *Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik*, Wien, Deuticke, 1931.