ANNALES
UNIVERSITATIS MARIAE CURIE - SKŁODOWSKA
LUBLIN - POLONIA
VOL. LI.1, 1 SECTIO A 1997

ROBERT BARTOSZYŃSKI and PREM K. GOEL (Columbus, Ohio)

On Probability of Evens

ABSTRACT. This paper is motivated by the observation that $P(X \text{ is even}) > 1/2$ for every Poisson random variable X which raises the question what other well known distributions may have this property. We explore broader aspects of this property which turns out to extend far beyond the class of Poisson random variables.

1. Introduction. Many introductory-calculus level probability textbooks give the following problem:

If X is a random variable with Poisson distribution what is $P(X \text{ is even})$?

The direct summation of even-numbered terms of the Poisson distribution is rather difficult, and the students are expected to invent and use the following trick: take the Taylor series-expansions of e^{λ} and of $e^{-\lambda}$, add them together (so that odd numbered terms cancel), and after some algebra one obtains $P(X \text{ is even}) = [1 + e^{-2\lambda}]/2$. To the extent that this solution requires the student to invent a trick, it is of some difficulty for mathematics majors, and is generally beyond the realm of even very good non-mathematics majors taking a course in probability.

This paper was motivated by the observation, usually missing in the textbooks, that $P(X \text{ is even}) > 1/2$ *for every Poisson random variable* X, which raises the question what other well known distributions may have this property. We explore broader aspects, as noted in [1], of this property, which turns out to extend far beyond the class of Poisson random variables.

2. The lemma. Since in the sequel we consider only integer-valued random variables, we shall omit this assumption, except in stating the Lemma.

Definition. A random variable X (or its distribution) will be called *zygophilic*, if[1]

$$P(X \text{is even}) - P(X \text{ is odd}) \geq 0 . \tag{1}$$

We shall say that X is *strictly zygophilic*, if the inequality in (1) is sharp.

The quantity on the l.h.s. of (1) will be called *even/odd* (e/o) *p-diff* for random variable X.

We start with the following key lemma.

Lemma. *Let $X = Y + Z$, where Y and Z are independent integer-valued random variables. Let*

$$p_X = P(X \text{ is even}), \qquad q_X = P(X \text{ is odd}) \tag{2}$$

and similarly for random variables Y and Z. Then

$$p_X - q_X = (p_Y - q_Y)(p_Z - q_Z). \tag{3}$$

Proof. Since $p_X = p_Y p_Z + q_Y q_Z$ and $q_X = p_Y q_Z + q_Y p_Z$, equation (3) follows from simple algebra.

The following propositions are consequences of the Lemma.

Proposition 1. *If $p_Y = p_Z$, then X is zygophilic; it is strictly zygophilic unless $p_Y = p_Z = 1/2$.*

Proposition 2. $p_X = q_X$ *iff* $p_Y = q_Y$ *or* $p_Z = q_Z$ *(or both).*

Thus, a convolution of a random variable that has equal chances for odds and evens with any other random variable, has again equal chances for odds and evens. Furthermore, we get

[1]The term derived from Greek, meaning "with a tendency towards being even" (as opposed to odd). We would like to gratefully acknowledge the contribution of those who made the pool of terms to choose from. Mr. Jerzy Bugaj of West Virginia University suggested several Latin-based possibilities, such as "amoparetic", or "paripropensic." Father Leo Stelton from Josephinum Pontifical College in Columbus, Ohio suggested the Greek-based term "dichamerizotophilic." Our final choice, "zygophilic," was suggested by Dr. Panickos Palettas of The Ohio State University.

Proposition 3. *A convolution of zygophilic random variables is again zygophilic.*

Proposition 4. *When $p_X \neq q_X$, X is zygophilic iff $p_Y - q_Y$ and $p_Z - q_Z$ are both of the same sign.*

As a consequence, we also have the following proposition.

Proposition 5. *Convolution of an even number of iid random variables is zygophilic. In particular, if X is infinitely divisible, then X is zygophilic.*

3. Analysis of families of distributions. Obviously, the Bernoulli random variable with $P(X=1) = p = 1 - P(X=0)$ is zygophilic iff $p \leq 1/2$. For binomial distributions, we obtain the following from Proposition 3.

Proposition 6. *Every binomial random variable with parameters n and $p \leq 1/2$ is zygophilic.*

On the other hand, from the first assertion of Proposition 5 we get

Proposition 7. *A binomial random variable with parameters n and p is zygophilic, if $p > 1/2$ and n is even.*

Thus, in a sense, "three-fourth majority" of binomial random variables with a fixed number of trials are zygophilic. The same conclusions as given in Propositions 6 and 7 may be obtained directly by observing that if X is a binomial random variable, then its e/o p-diff is $(q-p)^n$.

The second assertion of Proposition 5 implies

Proposition 8. *Every random variable with Poisson distribution is zygophilic.*

Note that the e/o p-diff for the Poisson distribution can also be obtained by using the Poisson approximation to binomial distribution.

Letting $p_n = \lambda/n, q_n = 1 - \lambda/n$, we easily obtain $\lim\,(q_n - p_n)^n = \lim\,(1 - 2\lambda/n)^n = e^{-2\lambda}$. While this appears simpler than "inventing the trick" mentioned in Section 1, it is not, because it requires justification for changing the order of two limits. In fact, most students may not even realize, at first glance, that two limits are involved here.

For geometric and, more generally, negative binomial distributions, the situation is somewhat complicated because of the ambiguity of definition of these random variables. According to one definition, a geometric random variable X is the number of failures preceding first success in iid Bernoulli trials with probability p of success. Then $P(X = k) = pq^k$ for $k = 0, 1, \dots$. In this case $P(X \text{ is even}) = p + q^2p + q^4p + \cdots = p/(1-q^2) = 1/(1+q)$. Thus, the e/o p-diff equals $(1-q)/(1+q)$, and we have

Proposition 9. *A geometric random variable with distribution defined above is zygophilic.*

An alternative (somewhat less often used) definition is that a geometric random variable Y is the total number of trials up to and including first success. Now we have $Y = X + 1$, where X is as defined previously. Thus, Y is even iff X is odd, and random variable Y is not zygophilic.

The ambiguity carries over to the case of negative binomial distribution. Corresponding to the first of the two geometric random variables above, we define $X^{(r)}$ as the number of failures preceding r-th success, where r is a positive integer. Thus, the probability distribution of $X^{(r)}$ is given by

$$P(X^{(r)} = k) = \binom{r+k-1}{r-1} p^r q^k, \quad k = 0, 1, 2, \dots \tag{4}$$

Since $X^{(r)}$ is a sum of r iid geometric random variables with distribution such as $X = X^{(1)}$, we have by Proposition 4:

Proposition 10. *A negative binomial random variable $X^{(r)}$, with distribution defined above, is zygophilic.*

In fact, Proposition 10 remains true even if r is not an integer. Indeed, formula (4) defines a probability distribution on the set of nonnegative integers as long as $r > 0$; moreover, $X^{(r)}$ is, for each n, a sum of n iid random variables each distributed as $X^{(r/n)}$. Consequently, $X^{(r)}$ is infinitely divisible, and Proposition 10 (without the condition that r is an integer) follows from Proposition 5.

To determine the e/o p-diff, it is easiest to use probability generating function (p.g.f.) Clearly, if X is an integer-valued random variable, and $f_X(s) = \Sigma P(X = k)s^k$, then $P(X \text{ is even}) - P(X \text{ is odd}) = f_X(-1)$. Since the p.g.f. of the negative binomial random variable $X^{(r)}$ is $\{p/(1-qs)\}^r$, we obtain

Proposition 11. *The e/o p-diff for the negative binomial distribution is* $\{p/(1+q)\}^r$.

As regards an alternative definition, one sometimes defines (for integer r) a negative binomial random variable $Y^{(r)}$ as the number of trials up to and including r-th success. Clearly, we have then $Y^{(r)} = r + X^{(r)}$, and $Y^{(r)}$ is zygophilic iff r is even.

To complete the story for the binomial random variables with parameters n and p, we now consider the situations in which the number of trials n itself is a random variable. Thus, we assume that, given $N = n$, the random variable X has a binomial distribution with parameters n and p, and that the random variable N has a Poisson distribution with parameter λ. It is well known that such a random variable has a Poisson distribution with parameter $p\lambda$. Thus, the o/e p-diff for the random variable X is equal to $e^{-2p\lambda}$, and X is obviously zygophilic.

More generally, o/e p-diff of a random variable X, representing the sum of a random number of independent random variables, i.e.,

$$X = Y_1 + Y_2 + \cdots + Y_N \tag{5}$$

where Y_i $i = 1, 2, \ldots$, is a sequence of independent random variables and N itself is a random variable, can be easily obtained from the p.g.f. of X. Let $f_Y(s)$ denote the p.g.f. of the random variable Y_i and let $g_N(s)$ denote the p.g.f. of the random variable N. It is well known, see, e.g., [2, Chapter 12], that the , p.g.f. $h_X(s)$, of X is given by the compound function $g_N(f_Y(s))$. A special case of such a random variable, in which N has a Poisson distribution with parameter λ, is said to have a *Compound Poisson distribution*. In this case, see e.g., [2],

$$h_X(s) = e^{-\lambda + \lambda f_Y(s)}. \tag{6}$$

Therefore $h_X(-1)$ is always non-negative, and we have the following proposition.

Proposition 12. *Every random variable with a compound Poisson distribution is zygophilic.*

However, since all infinitely divisible distributions, with support on the set of non-negative integers, have the p.g.f. defined by (6), see, e.g. [2], Proposition 12 is simply a more familiar version of the second assertion in Proposition 5.

Finally, let the random variable N have a negative-binomial distribution, as defined in (4), with the p.g.f. $g_N(s) = \{p/(1-qs)\}^r$. For the lack of a standard terminology, we will call a random variable X defined in (5), for which the distribution of N is given by (4), as having a *compound negative-binomial distribution*. Therefore, for a compound negative-binomial random variable X, $h_X(-1) = g_N(f_Y(-1)) = \{p/(1-qf_Y(-1))\}^r$. Now, since $f_Y(-1)$ represents the o/e p-diff of the random variable Y_i, it is clear that $|f_Y(-1)| < 1$. Thus we have the following result.

Proposition 13. *Every random variable with a compound negative binomial distribution is zygophilic.*

4. Some interesting examples and applications. Many real-life examples of phenomena that can be modeled by one of the distributions discussed in Section 3, are presented in [2]. We now present two interesting examples of real data sets in which the observed distribution possesses the zygophilic property, even though the data may not fit any of the distributions discussed in the previous Section.

Example 1. The data in Table 1 (taken from [3], p. 50) show the distribution of adults by the last digit of their age, as reported in the U.S. Census of 1880 and the U.S. Census of 1970.

Table 1. Last Digit of Age of U.S. Adult

Digit	Percentage in 1880 Census	Percentage in 1970 Census
0	16.8	10.6
1	6.7	9.9
2	9.4	10.0
3	8.6	9.6
4	8.8	9.8
5	13.4	10.0
6	9.4	9.9
7	8.5	10.2
8	10.2	10.0
9	8.2	10.1

One might expect that each of the ten possible digits should turn up in 10% of the people. However, the data show a very strong preference towards age ending in 0 or 5 in Census of 1880, and the same preference, although weaker, in 1970. In addition, the data are strongly zygophilic: in 1880 the percentage of even digits is 54.6% (despite the peak at digit 5), and in 1970, it is 50.3%.

A possible explanation lies in a combination of the following factors: (a) Some people may not know their age; the number of such people was probably much higher in 19th century than now; (b) people tend to use 0 and 5 as anchoring points; (c) There is a weaker preference towards even (over odd) endings, and (d) Some people, especially females, age in jumps: upon reaching age such as 40 they stop aging for as long as feasible, then move to $40 + X$, etc, where X is zygophylic.

Example 2. The data in Table 2 (taken from [4], p. 98) is based on a study concerning the human sex ratio. From a large study, the families with 12 children are classified according to the number of children of a given sex. (the data were collected in Germany in 19th century, when large families were common).

Table 2. Sex-ratio among families with 12 children

Number of boys	Number of girls	Observed frequency (No. of families)
0	12	3
1	11	24
2	10	104
3	9	286
4	8	670
5	7	1033
6	6	1343
7	5	1112
8	4	829
9	3	478
10	2	181
11	1	45
12	0	7
	Total	6115

According to the most common hypothesis, the distribution of the number of boys in a family of a given size n is binomial with parameters n and $p = 1/2$, so that the probability of a family of 12 having only boys is $(1/2)^{12}$, hence about 1 in 4,000. Thus, the chances of finding 7 such families among over 6,000 are so small that one should reject this hypothesis.

There were various attempts to explain this data set, by postulating that probability p of a boy varies among families and/or within a family. On the other hand, an explanation by postulating the dependence of sexes of consecutive children was rejected on non-statistical ground, since one could hardly visualize a biological system that would "remember" the sex of last child, let alone the sexes of all the preceding children.

But the curious feature of this data set is its zygophilic property: even rejecting the families with no girls, the number of families (out of 6,108) with an even number of boys (hence also even number of girls) is 3130, which is 86 above the expected 3054. The standard deviation is $\sqrt{6,108 * 0.25}$ or about 39. Thus, the excess of 86 above the mean is significant on 5% level. This propensity towards even number of boys (hence also even number of girls) is difficult to explain. It would be interesting to have the data for (say) families with 11 children, since then even number of boys would mean odd number of girls and vice versa.

The applications of our results are quite direct, and could possibly bring financial gains, although only to a rather narrow group of people, namely addicted gamblers. Imagine that one is to make a bet on the observed value in one of the following situations. In each case one may bet that the result will be even, or that the result will be odd. Correct guess will double your stake, incorrect will lead to losing the stake. The situations in question are the following:

(a) The total number of fire alarms that will occur tomorrow in your town;

(b) Total number of unsuccessful attempts that a given world-class athlete (e.g., high jumper) will make before breaking the world record;

(c) The total number of left-handed persons in a given group (e.g., passengers on a plane, customers in a restaurant, etc.);

(d) The total number of completely unsuccessful (negative response without invitation for an interview) job applications sent before one gets three invitations for a job interview.

The above four situations are real-life examples of phenomena governed, respectively, by Poisson distribution, geometric distribution, binomial distribution with $q > p$, and negative binomial distribution (with $r = 3$). What all these situations have in common is that *betting on an even outcome gives a somewhat higher chance of winning than betting on an odd outcome.*

The e/o p-diff may be small, but is positive in all four cases above. For example, in case (a), if the average number of fire alarms per day is 3, then probability of an even number of fires exceeds the probability of an odd number of fires by $e^{-2\lambda} = e^{-6}$, which is about one quarter of one percent (this is despite the fact that the most likely outcome (3) is odd). In case (b), the answer depends on how good the athlete is. For example, if his chances of breaking the world record are 5% on each attempt, then the e/o p-diff $(1-q)/(1+q) = 0.05/1.95$, or about 2.5 percent. In case (c), assuming that about 11% of people are left-handed, the answer depends on size n of the group of people in question. For example, if there are 11 people (on the plane, or in a restaurant), then the e/o p-diff $(q-p)^n = (0.89-0.11)^{11}$, or about 6.5 percent. Finally, in situation (d) the answer depends on the chances of outright rejection (as opposed to an invitation for a job interview; the chances of successful interview do not matter here). Suppose that the chances of outright rejection are 60%. Then the e/o p-diff is $\{p/(1+q)\}^r = (0.4/1.6)^3$, which is about 1.5 percent.

These advantages of even over odd outcomes may seem small, but a competent gambler with either some initial luck or huge initial pool of money to serve as a cushion, can make a fortune!

Conclusion. Nature is a zygophile: She has a distinct tendency to favor even numbers over odd numbers.

References

[1] Bartoszyński, R. and M. Niewiadomska-Bugaj, *Probability and Statistical Inference*, New York, Wiley 1996.

[2] Feller, W., *An Introduction to Probability Theory and Its Applications*, Vol. I, Third Edition. New York, Wiley 1968.

[3] Freedman, D., R. Pisani, R. Purves and Adhikari, *Statistics*, Second Edition. New York, W.W. Norton 1991.

[4] Samuels, M. L., *Statistics for the Life Sciences*, San Francisco, Dellen 1989.

Department of Statistics
The Ohio State University
Columbus, OH 43210
USA

received May 5, 1997