

## EXACT LOWER BOUND ON AN ‘EXACTLY ONE’ PROBABILITY

IOSIF PINELIS 

(Received 23 November 2020; accepted 14 January 2021; first published online 23 February 2021)

### Abstract

We obtain the exact lower bound on the probability of the occurrence of exactly one of  $n$  random events each of probability  $p$ .

2020 Mathematics subject classification: primary 60E15.

Keywords and phrases: exact lower bound, independence, pairwise independence, probability inequalities.

### 1. Introduction, summary and discussion

Suppose that  $A_1, \dots, A_n$  are random events each of probability  $p$ . Let  $E$  denote the event that exactly one of the events  $A_1, \dots, A_n$  occurs. We aim to provide the exact lower bound for the probability,  $P(E)$ , of the event  $E$ .

If the  $A_i$ 's are independent, then, by the binomial probability mass function formula (see, for example, [2, Section 1.3]),  $P(E) = npq^{n-1}$ , where  $q := 1 - p$ . So, in the ‘independent’ case,  $P(E)$  attains its maximum,  $(1 - 1/n)^{n-1} [\rightarrow 1/e \text{ as } n \rightarrow \infty]$ , at  $p = 1/n$ .

What will happen with  $P(E)$  when the  $A_i$ 's are only assumed to be pairwise independent? Already for  $n = 3$ , the pairwise independence of the  $A_i$ 's does not imply their ‘complete’ independence. Feller [3, page 126] wrote: ‘Actually such occurrences [of pairwise independence but not “complete” independence] are so rare that their possibility passed unnoticed until S. Bernstein constructed an artificial example. It still takes some search to find a plausible natural example.’ This is followed [3, page 127] by an example of three pairwise independent events that are not ‘completely’ independent. Another such example, [2, Example 2.3.3], ascribed in [2] to Bernstein, actually appears more common and natural than the example on page 127 in [3].

One may dispute the assertion that occurrences of pairwise independence without ‘complete’ independence are rare. Indeed, the definition of the independence of three events  $A, B, C$  consists of the four equations  $P(A \cap B) = P(A)P(B)$ ,  $P(B \cap C) = P(B)P(C)$ ,  $P(A \cap C) = P(A)P(C)$  and  $P(A \cap B \cap C) = P(A)P(B)P(C)$ . The first three of these four equations define the pairwise independence. The probabilities of the events  $A, B, C$  and of their pairwise and triple intersections can all be expressed as

the sums of the probabilities of certain pieces of the partition of the sample space (say  $\Omega$ ) generated by the events  $A, B, C$ . There are  $2^3 = 8$  pieces of this partition, with eight corresponding probabilities, which may be considered as nonnegative real variables tied just by one equation stating that the sum of these eight probabilities is 1. Thus, we have  $4 + 1 = 5$  equations with eight unknowns, which leaves us  $8 - 5 = 3$  degrees of freedom, which one can easily use to show that none of the four equations defining the independence of the events  $A, B, C$  may be dropped without altering the notion of independence. In particular, in this way it is easy to see that the pairwise independence does not imply the ‘complete’ independence. Moreover, it seems plausible that, in the case of three events  $A, B, C$ , the dimension of the semi-algebraic set [1] (in the eight variables mentioned above) corresponding to the ‘complete’ independence is one less than the dimension of the semi-algebraic set corresponding to the pairwise independence. More generally, for any natural number  $n$  of events  $A_1, \dots, A_n$ , the difference between the corresponding dimensions appears to be  $2^n - 1 - n - n(n-1)/2 \sim 2^n$  (as  $n \rightarrow \infty$ ). Here,  $2^n$  appears as the number of all equations of the form  $P(\bigcap_{j \in J} A_j) = \prod_{j \in J} P(A_j)$  for  $J \subseteq [n] := \{1, \dots, n\}$ . Of these  $2^n$  equations,  $1 + n$  equations (for the sets  $J \subseteq [n]$  of cardinalities 0 and 1) are trivial and  $n(n-1)/2$  of the  $2^n$  equations define the pairwise independence of the  $n$  events. Thus, there are  $2^n - 1 - n - n(n-1)/2$  ‘nontrivial’ equations defining the independence of the  $n$  events, in addition to the  $n(n-1)/2$  ‘nontrivial’ equations defining the pairwise independence.

From this viewpoint, the occurrences of ‘complete’ independence constitute an infinitesimally thin slice among the occurrences of pairwise independence. Therefore, it may seem very surprising that the strong law of large numbers (SLLN) for identically distributed random variables with a finite mean turns out to hold assuming only pairwise independence, as was demonstrated comparatively very recently by Tao [4, Remark 2].

In this note it will be shown that, in contrast with the SLLN result, the ‘exactly one’ probability  $P(E)$  may be quite sensitive to the distinction between pairwise independence and ‘complete’ independence.

**THEOREM 1.1.** *For each natural number  $n$  and each  $p \in [0, 1]$ ,*

$$\min P(E) = P_{n,p} := np(1 - (n-1)p)_+, \quad (1.1)$$

where the minimum is taken over all pairwise independent events  $A_1, \dots, A_n$  each of probability  $p$ , and  $x_+ := \max(0, x)$  for real  $x$ .

We see that, in contrast to the ‘completely independent’ case, for just pairwise independent events  $A_1, \dots, A_n$  the probability  $P(E)$  can be 0 for any  $n \geq 2$  and any  $p \geq 1/(n-1)$ . If we consider the special value  $p = 1/n$ , at which, as noted above,  $P(E)$  attains its maximum value  $(1 - 1/n)^{n-1} \approx 1/e$  in the ‘completely independent’ case, then for just pairwise independent events  $A_1, \dots, A_n$  we have  $\min P(E) = 1/n \rightarrow 0$ . If, for example,  $p = c/n$  with a fixed  $c \in (0, 1)$ , then in both cases of the ‘complete’ independence and of the pairwise independence, the probability  $P(E)$  stays away

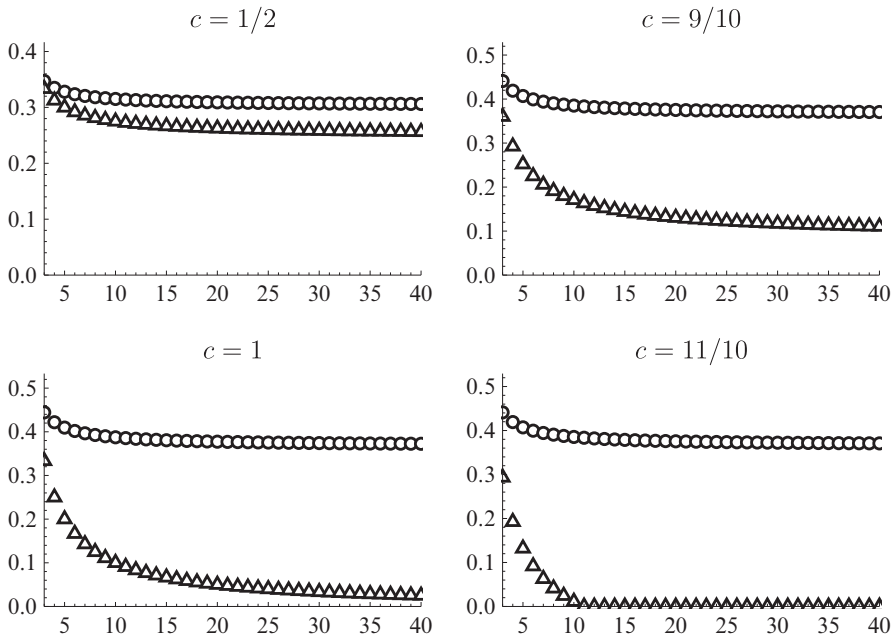


FIGURE 1. Graphs of the values of  $P(E)$  for  $p = c/n$ .

from 0. So,  $P(E)$  will necessarily be of the same order of magnitude (for large  $n$ ) in both cases only if  $p$  is small—more specifically, if  $p$  stays below  $c/n$  for some fixed  $c \in (0, 1)$ .

This is illustrated in Figure 1, which shows the graphs of the values of  $P(E)$  (the vertical axis) in the ‘completely independent’ case (circles) and in the ‘pairwise independent’ case (triangles) for  $n \in \{3, \dots, 40\}$  (the horizontal axis),  $p = c/n$  and  $c \in \{1/2, 9/10, 1, 11/10\}$ .

### 2. Proof of Theorem 1.1

For  $n = 1$ , Theorem 1.1 is trivial. So, in what follows we assume that  $n \geq 2$ .

For each  $j \in [n]$ , let

$$X_j := 1_{A_j},$$

the indicator of the event  $A_j$ . Let

$$N := X_1 + \dots + X_n,$$

the number of the events  $A_1, \dots, A_n$  that occurred. Then

$$E = \{N = 1\}. \tag{2.1}$$

Note that  $E X_j = p$  and (by the pairwise independence)  $E X_j X_k = p^2 + pq 1_{(j=k)}$  for all  $j$  and  $k$  in  $[n]$ . Now we have a perhaps unexpected use of the Chebyshev–Markov

inequality (see, for example, [2, Theorem 4.7.4]):

$$\begin{aligned}
 P(N \neq 1) &= P((N - 1)^2 \geq 1) \\
 &\leq E(N - 1)^2 \\
 &= EN^2 - 2EN + 1 \\
 &= \sum_{j,k \in [n]} EX_j X_k - 2 \sum_{j \in [n]} EX_j + 1 \\
 &= n^2 p^2 + npq - 2np + 1 \\
 &= 1 - np(1 - (n - 1)p).
 \end{aligned}$$

Therefore, and because  $P(N = 1) \geq 0$ ,

$$P(N = 1) \geq np(1 - (n - 1)p)_+ = P_{n,p}.$$

So, in view of (2.1),  $P_{n,p}$  is a lower bound on  $P(E)$  (compare (1.1)).

It remains to show that this lower bound is attained for each natural  $n \geq 2$  and each  $p \in [0, 1]$ . To do this, introduce the events

$$C_J := \left( \bigcap_{j \in J} A_j \right) \cap \left( \bigcap_{j \in [n] \setminus J} (\Omega \setminus A_j) \right)$$

for  $J \subseteq [n]$ . These events constitute a partition of the sample space  $\Omega$ . Moreover, for each  $m \in \{0\} \cup [n]$ ,

$$\{N = m\} = \bigcup_{\substack{J \subseteq [n], \\ |J|=m}} C_J, \quad (2.2)$$

where  $|J|$  denotes the cardinality of the set  $J$ . Also,

$$A_1 = \bigcup_{\substack{J \subseteq [n], \\ J \ni \{1\}}} C_J \quad \text{and} \quad A_1 \cap A_2 = \bigcup_{\substack{J \subseteq [n], \\ J \ni \{1,2\}}} C_J. \quad (2.3)$$

For each  $m \in \{0\} \cup [n]$ , let us assign the same probability, say  $x_m$ , to each event  $C_J$  with  $J \subseteq [n]$  such that  $|J| = m$ . Then, by (2.2),

$$P(N = m) = \binom{n}{m} x_m. \quad (2.4)$$

So, there will exist a probability space supporting such an assignment of probabilities to the  $C_J$ 's if and only if  $x_m \geq 0$  for all  $m \in \{0\} \cup [n]$  and

$$\sum_{m=0}^n \binom{n}{m} x_m = 1. \quad (2.5)$$

This follows because the set of values of the random variable  $N$  is the set  $\{0\} \cup [n]$ .

In view of (2.3),

$$P(A_1) = \sum_{m=1}^n \sum_{\substack{J \subseteq [n], \\ J \ni \{1\}, \\ |J|=m}} P(C_J) = \sum_{m=1}^n \binom{n-1}{m-1} x_m$$

(which is actually the value of  $P(A_j)$  for all  $j \in [n]$ ) and

$$P(A_1 \cap A_2) = \sum_{m=1}^n \sum_{\substack{J \subseteq [n], \\ J \ni \{1,2\}, \\ |J|=m}} P(C_J) = \sum_{m=1}^n \binom{n-2}{m-2} x_m$$

(which is actually the value of  $P(A_i \cap A_j)$  for all distinct  $i$  and  $j$  in the set  $[n]$ ). Now the conditions that  $P(A_j) = p$  for all  $j \in [n]$  and the  $A_j$ 's are pairwise independent can be rewritten as

$$\sum_{m=1}^n \binom{n-1}{m-1} x_m = p \quad \text{and} \quad \sum_{m=1}^n \binom{n-2}{m-2} x_m = p^2. \tag{2.6}$$

Take any  $p \in [0, 1]$ . Then there is some  $k \in [n - 1]$  such that

$$\frac{k-1}{n-1} \leq p \leq \frac{k}{n-1}. \tag{2.7}$$

For such a number  $k \in [n - 1]$ , let

$$x_m := \begin{cases} \frac{np}{k} (k - (n-1)p) / \binom{n}{k} & \text{if } m = k, \\ \frac{np}{k+1} ((n-1)p - (k-1)) / \binom{n}{k+1} & \text{if } m = k+1, \\ 0 & \text{if } m \in [n] \setminus \{k, k+1\}. \end{cases} \tag{2.8}$$

In view of (2.7),  $x_m \geq 0$  for all  $m \in [n]$ . Also, straightforward calculations show that the conditions (2.6) hold and

$$s := \sum_{m=1}^n \binom{n}{m} x_m = \frac{np(2k - (n-1)p)}{k(k+1)} \leq 1. \tag{2.9}$$

(The latter inequality is elementary. To prove it, one may first note that the maximum in  $p$  of the ratio in (2.9) is  $kn/(k+1)(n-1)$ , which increases in  $k \in [n - 1]$  to 1.) Therefore, one can satisfy (2.5) by letting  $x_0 := 1 - s \geq 0$ , so that the condition  $x_m \geq 0$  for all  $m \in \{0\} \cup [n]$  holds as well.

Furthermore, it follows from (2.1), (2.4), (2.7), (2.8) and the definition of  $P_{n,p}$  in (1.1) that

$$\begin{aligned} P(E) = P(N = 1) = nx_1 &= \begin{cases} np(1 - (n-1)p) & \text{if } 0 \leq p \leq \frac{1}{n-1}, \\ 0 & \text{otherwise} \end{cases} \\ &= np(1 - (n-1)p)_+ = P_{n,p}. \end{aligned}$$

This shows that the lower bound  $P_{n,p}$  on  $P(E)$  is indeed attained. This completes the proof of Theorem 1.1.  $\square$

We have the following easy corollary of Theorem 1.1.

**COROLLARY 2.1.** *Under the conditions of Theorem 1.1, the best lower bound on  $P(N = n-1)$  is  $P_{n,q}$  (compare (2.1)).*

To see why this corollary holds, switch from the ‘successes’  $A_j$  to the ‘failures’  $\Omega \setminus A_j$ , and also interchange the roles of  $p$  and  $q = 1 - p$ .

There are a number of further questions that one may ask concerning Theorem 1.1, including the following ones.

- (1) Assuming still that  $A_1, \dots, A_n$  are pairwise independent events each of probability  $p$ , what is the best *upper* bound on  $P(E) = P(N = 1)$ ? More generally, for each  $m \in \{0\} \cup [n]$ , under the same conditions on the  $A_j$ ’s, what are the best lower and upper bounds on  $P(N = m)$ ?
- (2) The same questions as above, but assuming, more generally, that the  $A_j$ ’s are  $r$ -independent for some  $r \in \{2, \dots, n-1\}$ , that is, assuming that for any  $J \subseteq [n]$  with  $|J| = r$ , the family  $(A_j)_{j \in J}$  is independent.
- (3) The same questions as above, but assuming, more generally, that the probabilities  $P(A_j)$  have possibly different prescribed values  $p_j$  for  $j \in [n]$ .
- (4) Yet more generally, let  $\mathcal{B}$  be any subset of the algebra (say  $\mathcal{A}$ ) generated by events  $A_1, \dots, A_n$ . Suppose that the probabilities  $P(B)$  have prescribed values, say  $p_B$ , for all  $B \in \mathcal{B}$ . Take any  $A \in \mathcal{A}$ . What are the best lower and upper bounds on  $P(A)$  in terms of the  $p_B$ ’s?

Looking back at the proof of Theorem 1.1 and recalling the discussion in Section 1, one can see that all the further problems listed above are ones of linear programming in a space of dimension exponentially growing with  $n$ , with the values of the  $P(C_J)$ ’s for  $J \subseteq [n]$  as the variables. Because the proof of Theorem 1.1, with all its parts fitting together quite tightly, was not easy to devise, all these problems seem hard to tackle theoretically or even computationally.

### 3. Conclusion

As we saw in Section 1, the condition of the ‘complete’ independence of  $n$  events, oftentimes assumed quite casually, actually involves  $\sim 2^n$  equations, which are practically impossible to test well even for rather moderate values of  $n$ , such as  $n = 40$ .

In contrast, the pairwise independence of  $n$  events involves only  $n(n-1)/2$  conditions. It may therefore be of value and interest to know how much the consequences of these two kinds of independence may differ from each other in various settings. It was noted in Section 1 that, at least as far as the most common version of the strong law of large numbers (for identically distributed random variables with a finite mean) is concerned, the pairwise independence is just as good as the ‘complete’ independence of the random variables. In stark contrast with that, the ‘exactly one’ probability may be quite sensitive to the distinction between the pairwise independence and the ‘complete’ independence, as shown in this note.

It is hoped that this small study may stimulate further research into other aspects of the difference between ‘complete’ independence and, on the other hand, pairwise independence (or, more generally,  $r$ -independence for some  $r \in \{2, \dots, n-1\}$ ). Also, perhaps some of the further questions enumerated at the end of Section 2 will attract attention. Finally, the methods presented in this note might turn out to be of use in other optimisation problems in probability, statistics and, perhaps, elsewhere, especially where the ‘complete’ independence is in doubt.

### References

- [1] J. Bochnak, M. Coste and M.-F. Roy, *Real Algebraic Geometry*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3), 36 (Springer, Berlin, 1998). Translated and updated from the 1987 French original.
- [2] A. A. Borovkov, *Probability Theory*, Universitext (Springer, London, 2013). Translated from the 2009 Russian fifth edition by O. B. Borovkova and P. S. Ruzankin.
- [3] W. Feller, *An Introduction to Probability Theory and its Applications*, 3rd edn, Vol. I (John Wiley, New York–London–Sydney, 1968).
- [4] T. Tao, ‘The strong law of large numbers’, <https://terrytao.wordpress.com/2008/06/18/the-strong-law-of-large-numbers/>, November 2008.

IOSIF PINELIS, Department of Mathematical Sciences,  
Michigan Technological University,  
Houghton, Michigan 49931, USA  
e-mail: [ipinelis@mtu.edu](mailto:ipinelis@mtu.edu)