International Journal of Scientific Research in Mathematical and Statistical Sciences
Research Paper
Vol.6, Issue.3, pp.97-99, June (2019)
E-ISSN: 2348-4519
DOI: https://doi.org/10.26438/ijsrmss/v6i3.9799

Verification of Birthday Paradox

Yogesh R Yewale

Department of Statistics, Ahmednagar College, Ahmednagar, Maharashtra, India

**Corresponding Author: yryewale@gmail.com Tel.: 9421165236*

Available online at: www.isroset.org

Received: 26/Apr/2019, Accepted: 20/Jun/2019, Online: 30/Jun/2019

Abstract— Birthday problem is one of the famous problem in the probability theory. It is also known as birthday paradox in some literature. In this paper attempt has been made to discuss the verification of the birthday paradox by collecting real life data. Birthdates of one thousand individuals are considered for this study. These individuals are divided into the various groups of size n. In each group matching pairs of birthdates were observed and the number of groups with matched pairs of birthdates were counted. This procedure is then repeated for different values of group size n. Chi-square test of goodness of fit is used to decide whether the discrepancies between the observed and expected values of the number of groups with matched birthdates is insignificant or not. It is found that the discrepancies between observed and expected values of the number of matched pairs is insignificant at 5% level of significance, which verifies the statement of the birthday paradox.

Keywords*—Birthday paradox, Chisquare test of goodness of fit*

I.INTRODUCTION

Richard Von Mises in 1939 posed the birthday problem. It states that "In a group of containing n people, what is probability that at least two will have same birthday? It can be stated in other way as "How many individuals one should consider in a group to ensure that the probability of at least two of them having same birthday exceeds one half ?" [1] The solution to this problem comes out to be 23. Thus according to this problem if there are ten groups of 23 individuals then 5 of them should have at least one pair of students with matching birthdates. If the problem is "How many individuals one should consider in a group to ensure that the probability of at least one matched pair is 0.99".[1] The solution to this problem comes out to be 57. This seems in contradiction with the pigeonhole principle which states that "If n units are placed into r boxes, and n > r, then at least two units will go into the same box". [2] According to this principle, to get matched pair of birthdates in a year of 365 days it is required to have 366 people.

Thus it becomes interesting to verify this paradox for the real life data. In this study, birthdates of these 1000 individuals are considered. These individuals are divided into the group of size n. Thus total number of groups (k) is an integer part of the number 1000 divided by n. In each group, matching pairs of birthdates were observed and the number of such groups with at least one matched pair of birthdate was counted. This is termed as the observed value of the number of groups with matched pair of birthdates out of the k. This procedure is then repeated for various values of n and for

each value of n number of groups with at least one matched pair of birthdate was observed .This is displayed in **Table 2**.

The calculation of probabilities of getting a matched pair and hence expected number of groups with matched pair out of k is discussed in the methodology section. These values are calculated at different values of 'n' such as 5, 10, 20, 23, 30, 40, 50, and 57 and are shown in **Table 1**. The result section discusses about the Chi-square test of goodness of fit which is used to check whether the difference between the observed and expected value is insignificant or not. After performing the testing of hypothesis procedure, it is found that the difference between observed and expected value is insignificant at 5% level of significance. Hence it can be concluded that the birthday paradox is verified.

II.RELATED WORK

John M. Johnson in his paper titled 'THE BIRTHDAY PROBLEM EXPLAINED' discussed about the computation of probabilities in a group of five persons for various events such as exactly two matched birthdates, exactly three matched birthdates, and exactly four matched birthdates, exactly five matched birthdates, two pairs with matched birthdates, one matched pair of birthdate and other three have same birthdates (full house problem) [1] . Joe Dan Austin in his paper titled 'The Birthday Problem Revisited 'discuss the expected number of matched pairs (birth mates) in a room with n people.[3]. Kevin S. Jones in his paper titled 'The Birthday Problem Again?' discuss about the variant of the birthday problem i.e. to calculate the probability of at least two of the k people (group size) have birthday occurring within r days. [4]

III. METHODOLOGY

Twenty students of Ahmednagar College were given the task of collecting the dataset of the birthdates of 50 individuals each. Here only day and month was taken into consideration and not the year of birth. These are then divided into groups of size n. Thus there are k groups each of size equal to integer part of 1000/n. Each of this group is checked for matched pair of birthdates. Then number of such groups out of k with at least one matched pair is counted. This value is considered as an observed value of the groups with matched pair with n as size of each group. This procedure is repeated for different values of n.

Let A: The event that birthdates of any two students is same

A_1: In a group of n, birthdate of first student will match with any of the remaining in the group.

A_2: In a group of size n, the birthdate of second student matches with that of first.

A_3: In a group of size n, the birthdate of third student matches with that of first or second

A_n: In a group of size n, the birthdates of n^{th} student matches with that of first or second ……or $(n-1)^{th}$ student.

Thus

$$P(A_1') = \frac{364}{365} \quad ,P(A_2') = \frac{364}{365}\times\frac{363}{365} \quad ,P(A_3') = \frac{364}{365}\times\frac{363}{365}\times\frac{362}{365}$$

$$P(A') = P(A_1' \cap A_2' \cap A_3' \cap \ldots A_n')$$

$$P(A') = P(A_1') \times P(A_2') \times P(A_3') \ldots P(A'_n)$$

$$P(A') = \left[\frac{364}{365}\right] \times \left[\frac{363}{365}\right] \times \left[\frac{362}{364}\right] \times \ldots \left[\frac{365-(n-1)}{364}\right]$$

$$P(A') = \frac{n!\binom{365}{n}}{365^n} \Rightarrow P(A) = 1 - \frac{n!\binom{365}{n}}{365^n}$$

For different values of n the value of probability of event A (getting at least one matched pair) is given in the Table 1. This table also gives the expected number of matched pairs out of k groups.

Table 1 Probability of Getting Matched Pair

Group size (n)	P(A)	Expected number of groups with matched pairs
5	0.0270	5
10	0.1170	11
20	0.4110	20
23	0.5072	21
30	0.7060	23
40	0.8910	22
50	0.9760	19
57	0.9900	16

represents the plot of probability of getting a matched pair (Y-axis) against a group of size n (X-axis). From this graph it can be observed that the probability of getting a matched pair exceeds 0.5 for group size 23 and it is 0.99 for a group of size 57.

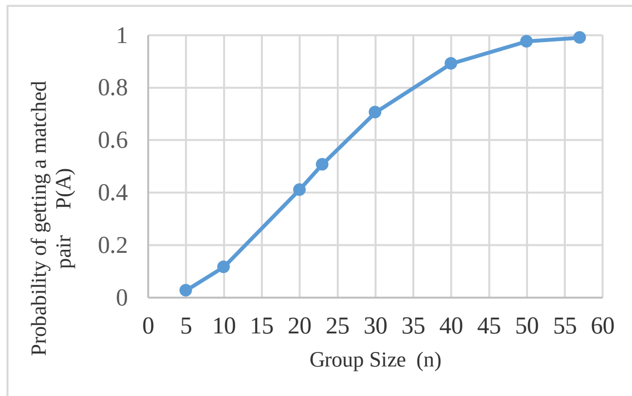

Figure 1 : Probability of getting a matched pair against group size n

IV. RESULTS AND DISCUSSION

One thousand observations are divided into k groups (where k denotes the integer part of 1000/n) each of size equal to n (size of group).

Table 2 represents the group size and corresponding value of the number of matched pairs out of k groups.

Table 2 Observed number of groups with matched pairs

Group size (n)	Observed number of Groups with matched pairs
5	4
10	8
20	11
23	18
30	15
40	16
50	20
57	16

To test whether the discrepancies between expected values and observed values of matched pairs is significant or not the Chi-square test of goodness of fit is used. Following are the steps to be followed in testing this claim.

Step 1: Define hypothesis
Null Hypothesis: The difference between observed number of matched pairs and expected values of matched pairs is insignificant.
Alternative Hypothesis: The difference between observed number of matched pairs and expected number of matched pairs is significant.

Step 2: Define test Statistic:
Assuming that null hypothesis is true the test statistic is defined by

$$\sum_{i=1}^{i=8} \frac{(O_i - E_i)^2}{E_i} \rightarrow \chi_7^2$$
$$\chi_{cal}^2 = 9.968357$$

Where
Oi: Observed values of the matched pairs.
Ei: Expected values of the matched pairs.
As eight groups of size 5, 10, 20, 23, 30, 40, 50, 57 are considered therefore the Chi-square statistic has 7 degrees of freedom.
Test Criterion: Reject null hypothesis at 5% level of significance if

$$\chi_{cal}^2 > {\chi^2}_{7,0.05} \quad where \quad P({\chi^2}_7 > {\chi^2}_{7,0.05}) = 0.05$$

From statistical table of Chi-square probabilities

$$\chi_{7,0.05}^2 = 14.06714$$
$$9.968357 = \chi_{cal}^2 < {\chi^2}_{7,0.05} = 14.06714$$

Decision: There is no sufficient evidence from the given data to reject the null hypothesis.
Conclusion: Statement of birthday paradox holds true.

V. CONCLUSION AND FUTURE SCOPE

From this study it can be concluded that the difference between observed and expected values of the matched pairs is insignificant at 5% level of significance. Thus it can be concluded that the birthday paradox holds true.

One can also calculate the probabilities that the difference between the birthdates of two individuals is ‘p’ days, where p is a positive integer and verify it by collecting the real life data.

REFERENCES

[1] J. M. Johnson, “THE BIRTHDAY PROBLEM EXPLAINED,” *Math. Teach.*, vol. 90, no. 1, pp. 20–22, 1997.
[2] K. R. Rebman, “The Pigeonhole Principle (What It Is , How It Works , and How It Applies to Map Coloring),” *Two-Year Coll. Math. J.*, vol. 10, no. 1, pp. 3–13, 1979.
[3] J. D. Austin, “The birthday problem revisited,” *Two-Year Coll. Math. J.*, vol. 7, no. 4, pp. 39–42, 1976.
[4] K. S. Jones, “The Birthday Problem Again ?,” *Math. Teach.*, vol. 86, no. 5, pp. 373–377, 1993.