Master of Arts in Teaching (MAT)
Masters Exam

Stacey Aldag

In partial fulfillment of the requirements for the Master of Arts in Teaching with a Specialization in the Teaching of Middle Level Mathematics in the Department of Mathematics.
David Fowler, Advisor

July 2007

A Monte Carlo Simulation of the Birthday Problem

Stacey Aldag

July 2007

A Monte Carlo Simulation of the Birthday Problem

Question, *how many people would you need in a group in order for there to be a 50-50 chance that at least two people will share a birthday?* **Answer,** *23 people.* But how can this be? There are 365 days in a year and half of that would be 182, so why wouldn't you need at least 182 people to have a 50-50 chance? Strangely enough the answer to this question is only 23 people are necessary to have a 50% chance at least two people in the group will share a birthday. This situation, where the answer is counter intuitive, is called a paradox, making the official name for this probability problem, the Birthday Paradox.

In 1939 Richard von Mises proposed the problem, and it has grown to be a commonly explored and talked about combinatorial probability problem. We will initially consider a simpler problem; what is the probability that two randomly selected people will have the same birthday? As calculations for this problem begin, some assumptions need to be established. First, let all days of the year be equally likely candidates for a person's birth date. Second, the date of February 29th will be disregarded so there are 365 days every year. Third, multiple births will be considered as one birthday, and fourth, two people have the same birthday if the month and day are the same, the year will be ignored.

To compute the probability that two randomly selected people have the same birthday, we note that since their birthdays are independent of each other, the two probabilities can be calculated separately and multiplied together. Using P (*E*) to represent the probability of the occurrence of event *E*, a mathematical expression can be written:

P (first person has a birthday) · P (second person's birthday is the same day).

Since it will not matter which day the first person is born on P(birthday of first person) = 100% or $\frac{365}{365}$ and P(second person's birthday is the same day) = $\frac{1}{365}$ because there is only 1 day out of 365 on which the second person can be born in order to match the first person. So the combined probability is $\frac{365}{365} \bullet \frac{1}{365} = \frac{1}{365} = 0.27\%$, meaning there is less than a 1% chance that two randomly selected individuals from a group of two people will have the same birthday.

Now what if three people are being considered? Can the probability for a group of three people be computed in the same fashion as the probability for two people as illustrated above? Assuming it can, the formula for determining the probability that a third person has the same birthday as the first person and the second person would look like this:

P (first person has a birthday) · P (second person's birthday is the same day) · P (third person's birthday is the same day) = $\frac{365}{365} \bullet \frac{1}{365} \bullet \frac{1}{365}$.

Note that the problem proposed in the opening lines of this paper only asks whether *two* people share a birthday; the third person does not need to share that birthday. How does this, the actual problem, change the calculations? Considerations need to be made for all of the combinations of choosing two people from three people. Thus the possibilities for successfully matching the birthdays of two people in a group of three have just become significantly more complicated. If computing the probability of two people sharing a birthday in a group of three people requires the consideration of several different possibilities, how can this be done for a room of twenty or thirty people?

A basic rule about probability will help develop a general formula for solving this problem. The probability that an event occurs and the probability that the event fails to occur have a sum of 100%. Symbolically, P (event happens) + P (event doesn't happen) = 1. This thinking leads to the complement rule, P (event happens) = 1 – P (event doesn't happen). So in the context of this problem, the P (at least two people from a group of people share a birthday) = 1 - P (no two people from a group of people share a birthday); meaning the event where at least two people in a group of people have the same birthday is complementary to all birthdays being different for the group of people.

Thus we can calculate P (two people share a birthday in a group of people) $=1-\frac{365}{365}\bullet\frac{364}{365}=0.27\%$. Using the complement incorporates the possible number of days for which the birthdays are *different* instead of the number of days for the birthdays are the same, and it incorporates all of the combinations of two people sharing the same day, thus making the calculations less complex. The first person has a possibility of 365 out of 365 days from which his or her birthday may be selected and the second person has 364 out of 365 days on which his or her birthday may fall in order to be different from the first person's birthday. Determining this probability from a group of three people will follow the same format as the group of two people. The probability that the third person's birthday will be different is 363 out of 365 days since there are only 363 days left from which to choose in order to be different from the other two people. Again, the birthdays are independent and thus the probabilities can be combined using multiplication. The probability equation becomes P (at least two people in a group of three people share a birthday) = $1-\frac{365}{365}\bullet\frac{364}{365}\bullet\frac{363}{365}=\frac{132132}{133225}$ or approximately 0.82% chance that at least two people will have the same birthday in a group of three people.

When considering four people: the formula can be expanded to be the P (at least two people in a group of 4 people share a birthday) = $1-\frac{365}{365}\bullet\frac{364}{365}\bullet\frac{363}{365}\bullet\frac{362}{365}=\frac{47831784}{48627125}$ or approximately 1.64% chance that at least two people out of a group of four people have the same birthday. Moving on, considerations can be made for five people, six people, and so on. The pattern continues, and a true mathematician would develop a general formula to use with any number of people. Going back to the calculations for a group of four people, a pattern has become apparent, and using this pattern a formula can be derived for a group of n people.

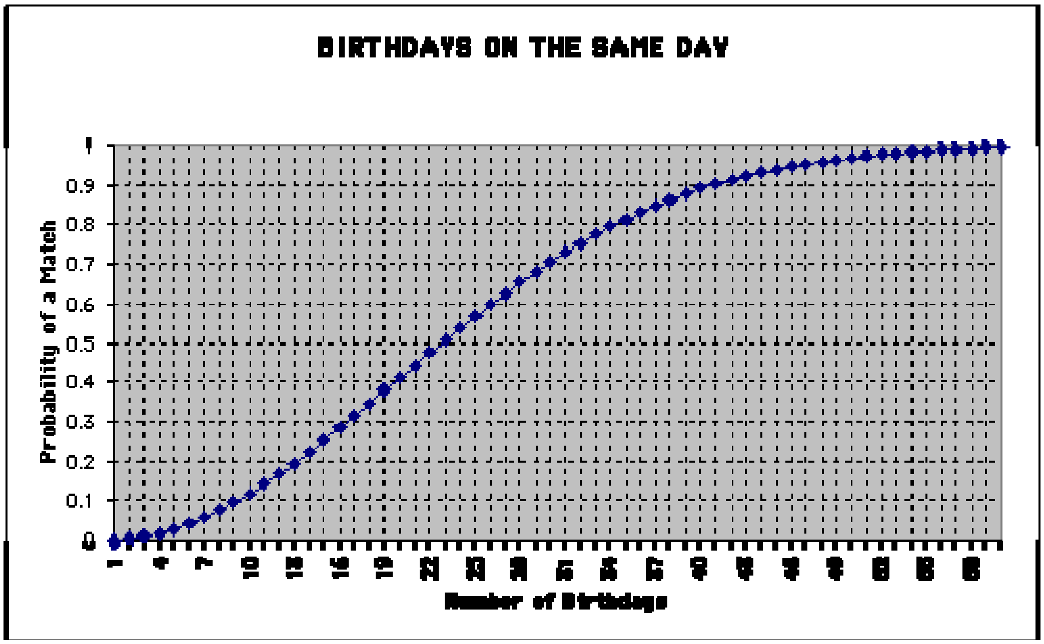
P (at least two people in a group of n people share a birthday) =

$$1-\left(\frac{365}{365}\bullet\left(\frac{365-1}{365}\right)\bullet\left(\frac{365-2}{365}\right)\bullet\left(\frac{365-3}{365}\right)\bullet\ldots\bullet\left(\frac{365-n}{365}\right)\right).$$

Better yet, permutation notation will simplify the expression to be:

P (at least two people share a birthday in a group of n people) $=1-\frac{\left({}_{365}P_n\right)}{365^n}=1-\frac{365!}{(365-n)!\bullet 365^n}$.

The graph below illustrates this formula. The x-axis represents the number of people in a group and the y-axis represents the probability that at least two people in the group will share the same birthday. Notice when 23 people are in a group, the probability of a matching birthday is approximately 50% and if 42 people are in a group, the probability of a matching birthday is just above 90%.



Source: http://www.mste.uiuc.edu/reese/birthday/explanation.html

During the research process several other mathematical formulas were found representing computations for the birthday probability problem. The table below compares three formulas. The first uses the permutation notation, developed on the previous pages; P (at least two people share a birthday in a group of n people) $=1-\frac{({}_{365}P_n)}{365^n}=1-\frac{365!}{(365-n)!\bullet 365^n}$. The second formula evaluates the probability using exponentiation and combinations of pairs of people; $1-\left(\frac{364}{365}\right)^{({}_nc_2)}$. The third method is an approximation using the Taylor series expansion of the exponential function; P(at least two people share a birthday in a group of n people) = $1-e^{-n(n-1)/2\bullet 365}$.

n	Permutation	Exponentiation	Taylor Series
10	11.695%	11.614%	11.599%
15	25.290%	25.029%	24.999%
20	41.143%	40.623%	40.581%
21	44.369%	43.793%	43.749%
22	47.570%	46.940%	46.894%
23	50.730%	50.048%	50.000%
24	53.834%	53.102%	53.054%
25	56.870%	56.091%	56.041%
30	70.631%	69.682%	69.632%
35	81.438%	80.454%	80.410%
50	97%	96.529%	96.513%
100	99.999960%	99.999873%	99.999871%
200	99.999999%	100%	100%
300	100%	100%	100%
350	100%	100%	100%
366	100%	100%	100%

The computations for each of the methods are very close (within hundredths in some cases), and it is interesting to compare the formulas and observe that all of the methods shown here calculate the probability by computing the complement of the situation first, and then subtracting that value from 1.

The derivation of the probability formula above is very interesting but how could a teacher of middle school mathematics entice students to believing that 50% of groups of 23 people probably have at least two people that share a common birthday without showing the evolution of the formulas shared in the table? One might suggest an experiment. For example, the teacher could send students out into the school to interview different classrooms in order to collect raw data and return to analyze the data during a whole class discussion. The desired outcome of the experiment would be to determine if the birthday paradox is true. An experiment is a good idea and can be fun for the students, but several factors may hinder the success of this activity. Does each of the classrooms being interviewed have 23 students? Are other teachers in the building willing to be interrupted by such an activity? Does a mathematics teacher have time to spend possibly two days collecting, sharing and analyzing data for one single activity? Most mathematics teachers would answer no to at least two of the questions stated above, making an actual experiment seem impossible. Instead of data collection outside the classroom, what about generating random numbers to represent birthdays? If each student could simulate several groups of 23 birthdays, the whole class could easily produce 100 or more pieces of data to examine. This would allow data collection and analysis to be done in one day.

Random numbers are generated by a process where the outcome is unpredictable and cannot be reproduced. Each number generated is required to be independent, so there are no correlations between successive numbers. It is considered impossible for humans to produce long strings of random numbers and justify that they are truly random. Humans may choose to include favorite numbers more often than what is probable, and avoid simple groupings of numbers which may be considered "a pattern", in order to over exaggerate randomness. Thus the study and generation of random numbers developed only as computers began to evolve.

The famous birthday paradox problem becomes a real-world mathematics application of a Monte Carlo-type simulation. The Monte Carlo simulation is defined as a problem solving technique involving organizing computer-generated random numbers as samples to model various phenomena to assist in decision making. The term Monte Carlo method was coined by Stanislaw Ulam and Nicholas Metropolis in reference to the famous casino in Monte Carlo, Monaco. Ulam states in his autobiography that the method was named in honor of his uncle, who was a gambler, and makes reference to the method being similar to the randomness and repetitive nature of the activities found in a casino.

The most famous early use of the Monte Carlo method involved a random method utilized to calculate the properties of the newly-discovered neutron in 1930 as part of the

Manhattan Project. The results using this method were limited due to the computational tools employed at that time. Once electronic computers were established, beginning around 1945, the Monte Carlo method began to be studied in more depth and in the 1950s it was used at Los Alamos, New Mexico for work relating to the development of the hydrogen bomb.

As the simulations and methods became popularized the fields of physics, physical chemistry and operations research put them to good use. The Rand Corporation and the U.S. Air Force were two of the first major organizations responsible for funding and disseminating information on the Monte Carlo method. Today the Monte Carlo method is an integral part of research in the fields of: risk analysis in the financial markets, drug testing and study of performance testing of imagery equipment for pharmaceutical and medical companies, study of traffic simulations and the development and possible production of hydrocarbon energy. Overall the Monte Carlo method is a working part of any business where risk assessment and decision making are important, particularly where an underlying formula is unknown.

So how does the Monte Carlo method fit into the middle level mathematics classroom? Simply using a graphing calculator and its random number generator, students can simulate a Monte Carlo method to solve the birthday paradox problem. The idea is to use random numbers to symbolize the birthdays of people in a group and then allow the calculator to simulate several groups of people in a few minutes. In the April 2001 issue of *Mathematics Teacher*, Matthew Whitney's article illustrates detailed instructions on how to implement the Monte Carlo simulation to solve the birthday paradox problem using the TI-83 graphing calculator. His article was an outstanding reference for the classroom activity detailed below.

First step to the Monte Carlo process is to generate sets of random integers in the lists of a TI-84+ calculator. There are several methods for generating and storing numbers into lists but for this activity the suggestion is to enter lists via the STAT menu. Two methods will be discussed. The first method illustrates the idea of assigning each birthday a number between 1 and 365. The directions for simulating the birthdays for a group of 23 people using the TI-84+ calculator are listed below:

- go to **CATALOG** (2nd key then 0)
- arrow down and select **ClrAllLists**, press **ENTER** and the calculator will read DONE
- choose the **STAT** key
- select **EDIT**, which is option 1
- position the cursor on the heading for List 1 (L1)
- select **MATH** key, the **PRB** menu and then option 5 **randint(**
- finish keying in the information to match **randint(1, 365, 23)**
- press **STAT** and select option 2 **SortA(**
- finish keying in the information to match **SortA(L1)**
  [This command will sort all of the data in ascending order from January 1st to December 31st.]
- press **STAT** and **EDIT** and look through List 1 to identify any matching dates
  [If there is at least one pair of matching birthdays the trial was a success and if there were no matching birthdays the trial was a failure.]

As students analyze the data produced with this method, it becomes apparent that it is difficult to identify the actual date of the common birthday when the birthdays have been assigned a number between 1 and 365. Thus, a second method for simulating birthdays was developed. This method generates separate random numbers for the month and the day of the birthday. The directions for simulating the birthdays for a group of 23 people using the TI-84+ calculator are listed below:

- go to **CATALOG** (2nd key then 0)
- arrow down and select **ClrAllLists**, press **ENTER** and the calculator will read DONE
- choose the **STAT** key
- select **EDIT**, which is option 1
- position the cursor on the heading for List 1 (L1)
- select **MATH** key, the **PRB** menu and then option 5 **randint(**
- finish keying in the information to match **randint(1, 12, 23)**

[The one is for January, the 12 represents December, so the months will range from 1 to 12, and the 23 is the number of random integers generated in List 1.]

- position the cursor on the heading for List 2 (L2)
- select **MATH** key, the **PRB** menu and then option 5 **randint(**
- finish keying in the information to match **randint(1, 31, 23)**
  [The one is for the first day of the month, the 31 represents the last day of the month, so the days will range from 1 to 31, and the 23 is the number of random integers generated in List 1.]
- position the cursor on the heading for List 3 (L3) and type **100*L1+L2** and press **ENTER**
  [This will display the month and day together in the same list, meaning the number 118 represents January 18th.]
- press **STAT** and select option 2 **SortA(**
- finish keying in the information to match **SortA(L3, L1, L2)**
  [This command will sort all of the data in ascending order from January 1st to December 31st.]
- press **STAT** and **EDIT** and look through List 3 to identify any matching dates
  [If there is at least one pair of matching birthdays the trial was a success and if there were no matching birthdays the trial was a failure.]

In a class of 25 students, if each student ran the simulation four times, 100 trials can be collected and analyzed quickly. This experimental data can then be compared to the theoretical probability and this classroom activity can springboard into a probability discussion and also serve as an example of where mathematics and the use of mathematical methods actually exist in the real world.

When using the second method, special attention should be paid to data considered to be "bogus", meaning dates that do not exist on the calendar. Since the instructions for the second method allow every month to have 31 days, students must be aware of this fact and discard the trials with lists containing bogus data. They will then need to generate new lists to replace the discarded trial. A teacher may also choose to have students program the calculator to run the simulation automatically, allowing for even more data to be collected. A copy of the Birthday Program to run the Monte Carlo simulation for a TI-84+ calculator has been included for both methods in the appendix.

Life is full of situations in which probability allows people to make informed decisions. Agriculture, business, engineering, medicine, social sciences and sports are just a few areas of expertise that rely on understanding what probability stands for and how outcomes affect the decision making process. Today mathematics teachers need to provide numerous opportunities to involve students in probabilistic thinking about different situations in order for students to develop an understanding of the world of chance. Technology can produce statistical samples quickly allowing the classroom focus to be on analyzing the data and thinking deeply about what the probability represents. The Birthday Paradox coupled with the Monte Carlo simulation will provide students with a hands-on experience of probabilistic thinking using a real-world simulation.

References

De Veaus, R.D., Velleman, P.F., & Bock, D.E. (2006). *Intro stats, 2nd edition.* Boston: Pearson

National Council of Teachers of Mathematics. (2007). *Illuminations: Birthday Paradox.* Retrieved on July 3, 2007 from http://illuminations.nctm.org/LessonDetail.aspx?id=L299

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics.* Reston, VA: Author.

National Council of Teachers of Mathematics. (2006). *Thinking and reasoning with data and chance: sixty-eighth yearbook.* Reston, VA: Author.

Reese, George. (2005). *The Birthday Problem: A Short Lesson in Probabilit*y. Retrieved on July 3, 2007 from http://www.mste.uiuc.edu/reese/birthday

Su, Francis E., et al. (2005). "Birthday Problem." *Mudd Math Fun Facts.* Retrieved on July 2, 2007 from http://www.math.hmc.edu/funfacts

Weisstein, E.W. (2003). *Birthday Problem.* Retrieved on July 3, 2007 from Math World: A Wolfram Web Resource. http://mathworld.wolfram.com/BirthdayProblem.html

Weisstein, E.W. (1999). *Monte Carlo Method.* Retrieved on July 3, 2007 from Math World: A Wolfram Web Resource. http://mathworld.wolfram.com/MonteCarloMethod.html

Weisstein, E.W. (1999). *Random Number*. Retrieved on July 18, 2007 from Math World: A Wolfram Web Resource. http://mathworld.wolfram.com/RandomNumber.html

Whitney, M.C. Exploring the birthday paradox using a Monte Carlo simulation and graphing calculators. Mathematics Teacher. Vol. 94, No. 4, p. 258-262. Retrieved on June 28, 2007 from http://www.hwwilson.com

Appendix

Birthday Program

The following commands should be typed into the TI-84+ calculator using the program key, PRGM, following by choosing the NEW option from the menu. Note the N variable in the program represents the size of the group that is being tested for a matching birthday:

1st Method
PROGRAM
Name=BIRTHDAY

```
:ClrAllLists
:Disp "NUMBER OF"
:Disp "PEOPLE"
:Input N
:randInt(1, 365, N)->L1
:SortA(L1)
```

2nd Method
PROGRAM
Name=BIRTHDAY

```
:ClrAllLists
:Disp "NUMBER OF"
:Disp "PEOPLE"
:Input N
:randInt(1, 12, N)->L1
:randInt(1, 31, N)->L2
:100*L1+L2->L3
:SortA(L3, L1, L2)
```