Calculating the Theoretical Probability of the Birthday Problem Using R
Stat 341 - Fall 2008

This is the R code that looks at the probability of the birthday problem, the probability of at least 2 people in a room of n people sharing a birthday. In this handout, we will first look at the problem using $n = 26$ since this was the number of people in Stat 341 on the first day we began discussion of this problem. Then we will look at calculating the probability for a general value of n.

To begin calculating the theoretical probability, we need to first set up the problem. There are 365 different possible birthdays (ignoring Feb. 29th). We need to calculate all the possible ways a group of 26 people can have birthdays. In R, this is

```
allpossbd26<- 365^(26)
```

To calculate the probability of the birthday problem, we are going to look at the "opposite" situation (called the complement event), where no one in the group of 26 people shares a birthday. How many ways can this happen? In R this is

```
noshare26<- prod(c(340:365))
```

We can also write this as

```
noshare26<- choose(365,26)*factorial(26)
```

The theoretical probability of the "opposite" situation is

```
noshare26/allpossbd26
```

This value is 0.4017592. The probability that we want is

```
1 - noshare26/allpossbd26
```

This value is 0.5982408. Notice how close the emperical probability of 0.5960 is to the theoretical probability of 0.5982408.

How could we calculate this theoretical probability with a general value of n? Let's assign n values from 1 to 100. In R, the command is

```
n <- c(1:100)
```

To get the number of all possible ways to assign birthdays to n people, we need to calculate

```
allpossbd<- 365^n
```

To get the number of ways that no one in the group of n people will share a birthday, we need to calculate

```
noshare<- choose(365,n)*factorial(n)
```

Then the theoretical probabilities for the birthday problem with n people is

```
bdprobs<- 1 - noshare/allpossbd
```

For n from 1 to 100, the theoretical probabilities are

```
  [1] 0.000000000 0.002739726 0.008204166 0.016355912 0.027135574 0.040462484
  [7] 0.056235703 0.074335292 0.094623834 0.116948178 0.141141378 0.167024789
 [13] 0.194410275 0.223102512 0.252901320 0.283604005 0.315007665 0.346911418
 [19] 0.379118526 0.411438384 0.443688335 0.475695308 0.507297234 0.538344258
 [25] 0.568699704 0.598240820 0.626859282 0.654461472 0.680968537 0.706316243
 [31] 0.730454634 0.753347528 0.774971854 0.795316865 0.814383239 0.832182106
 [37] 0.848734008 0.864067821 0.878219664 0.891231810 0.903151611 0.914030472
 [43] 0.923922856 0.932885369 0.940975899 0.948252843 0.954774403 0.960597973
 [49] 0.965779609 0.970373580 0.974431993 0.978004509 0.981138113 0.983876963
 [55] 0.986262289 0.988332355 0.990122459 0.991664979 0.992989448 0.994122661
 [61] 0.995088799 0.995909575 0.996604387 0.997190479 0.997683107 0.998095705
 [67] 0.998440043 0.998726391 0.998963666 0.999159576 0.999320753 0.999452881
 [73] 0.999560806 0.999648644 0.999719878 0.999777437 0.999823779 0.999860955
 [79] 0.999890668 0.999914332 0.999933109 0.999947953 0.999959646 0.999968822
 [85] 0.999975997 0.999981587 0.999985925 0.999989280 0.999991865 0.999993848
 [91] 0.999995365 0.999996521 0.999997398 0.999998061 0.999998560 0.999998935
 [97] 0.999999215 0.999999424 0.999999578 0.999999693
```

We can plot the number of people in the room versus the probability that at least two people in the room will share a birthday. The R command for this plot is below and the plot can be found on the next page.

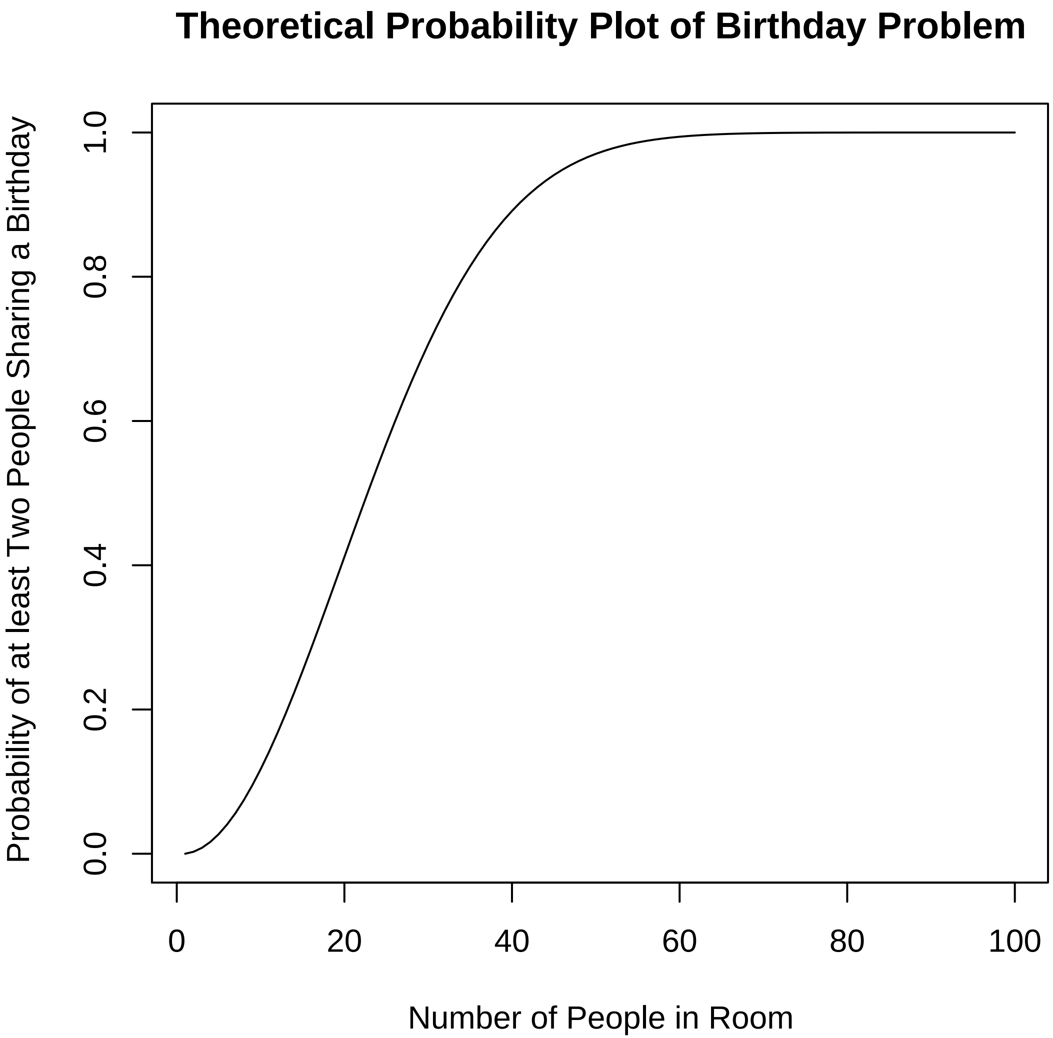
```
plot(n,bdprobs, type = "l")
```

People are generally surprised that the probability of the birthday problem is just over 0.5 for only 23 people in the room and that the probability reaches 0.75 with only 32 people in the room. This probability will continue to grow closer to 1 until $n = 366$. When there are 366 people in the room, there must be at least 2 people who share a birthday, making the probability exactly 1.

In conducting the simulation and calculating the theoretical probability of the birthday problem, we have made several assumptions. These assumptions are necessary to simplify the problem in order to be able to either simulate or calculate the probability. Whether these assumptions are reasonable is open for debate.

Assumptions in Calculating or Simulating the Birthday Problem

- No one has a birthday on February 29th. While this assumption is clearly not true, violating the assumption won't change the probabilities very much.
- Birthdays are evenly distributed throughout the 365 days of the year.
- No twins, triplets, etc. are in the room.


Theoretical Probability Plot of Birthday Problem
Probability of at least Two People Sharing a Birthday
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100
Number of People in Room