Math 3070 § 1.	**Birthday Example:**	Name: Example
Treibergs	**Use of Conditional Probability.**	June 12, 2011

The Birthday problem asks for how many people must be in a room before the chances are better than $50\% - 50\%$ that at least two of the people share the same birthdate. The elementary solutions assume that there are $n_1 = 366$ days or $n_2 = 365$ days in the year and that any one of these n_i days is equally likely to be somebodies birthdate. Assuming peoples birthdates are independent, then the probability that n have distinct birthdates is

$$P = \frac{n_i - 1}{n_i} \cdot \frac{n_i - 2}{n_i} \cdots \frac{n_i - n + 1}{n_i}.$$

Hence the probability that at least two have the same birthdate is $1 - P$. So what difference does the n_i make? Also, what difference does it make to include leap years? It turns out that the probabilities are very close, and all three approximations give $k = 23$ as the least number required for the probability that some birthdates match to exceed one half.

Let me describe the computation that includes leap years. In this case, we assume that every fourth year is a leap year. This is not quite true since on most century years the leap day is omitted. We continue to assume that the birthdates are independent. The number of days in four years is

$$s = 365 + 365 + 365 + 366 = 1461.$$

Assuming each day is equally likely, the probability that somebody's bithdate is Feb. 29, "leap day," is $1/s$. The probability that any other date is somebody's birthdate is $4/s$, which is between $1/366$ and $1/365$.

The computation may be done in two ways, which turn out to be equivalent. Both compute the probability that for n people all birthdates are distinct. Let X be the number of leap days among the n. Then X is a binomial variable $X \sim \text{bin}\left(n, \frac{1}{s}\right)$. Let A_n denote the event that all n people have distinct birthdates. Then by the total probability formula

$$P(A_n) = \sum_{i=0}^{n} P\left(A_n \cap \{X = i\}\right) = \sum_{i=0}^{n} P\left(A_n | \{X = i\}\right) P\left(\{X = i\}\right).$$

If more than two people have leap day as their birthday, then the probability of distinct dates is zero $P\left(A_n | X = i\}\right) = 0$ if $i > 1$. Since $P\left(\{X = i\}\right) = \text{bin}(i, n, 1/s)$,

$$P(A_n) = \left(\frac{s-1}{s}\right)^n P\left(A_n | X = 0\right) + \frac{n}{s}\left(\frac{s-1}{s}\right)^{n-1} P\left(A_n | X = 1\right) \tag{1}$$

Now conditioning on $X = 0$ implies that all dates are non leap days so for $n \geq 1$, we compute as for 365 days: the second can be any day but the first, the third can be any day but one of the first two and so on, so for $n \geq 1$,

$$P\left(A_n | X = 0\right) = \frac{366 - 1}{365} \cdot \frac{366 - 2}{365} \cdots \frac{366 - n}{365}.$$

Similarly, conditioning on $X = 1$ implies that there is one leap day and all other dates are non leap days. One day is unique so $P\left(A_1 | X = 1\right) = 1$. Also $P\left(A_2 | X = 1\right) = 1$ because at least one day is a leap day and other day is not, hence distinct. So for $n \geq 2$, we compute as for 365 days: for three days, there are two non leap days, the second of which can be any day but the first non leap day, and so on. So for $n \geq 2$,

$$P\left(A_n | X = 1\right) = \frac{367 - 1}{365} \cdot \frac{367 - 2}{365} \cdots \frac{367 - n}{365}.$$

Now let me describe the second way to do this computation. Let B denote the event that the leap day occurs among the n and B^c its complement that none of the birthdays are leap days. Denote for $n \geq 1$ the probabilities

$$p_n = P(A_n \cap B); \qquad q_n = P(A_n \cap B^c).$$

It follows that $P(A_n) = p_n + q_n$. We deduce the recursion formulas for p_n and q_n. Again by the total probability formula

$$P(A_{n+1} \cap B) = P(A_n \cap B)P(A_{n+1} \cap B | A_n \cap B) + P(A_n \cap B^c)P(A_{n+1} \cap B | A_n \cap B^c)$$

Since $P(A_{n+1} \cap B | A_n \cap B^c)$ is the probability that the $(n+1)$-st person changes n people without leap day to $n+1$ with a leap day, in other words that her birthday is leap day. The $n+1$ days are automatically distinct, thus

$$P(A_{n+1} \cap B | A_n \cap B^c) = \frac{1}{s}.$$

Similarly, $P(A_{n+1} \cap B | A_n \cap B)$ is the probability that $(n+1)$-st person's birthdate is not a leap day and is not on of the previous non-leap days, so

$$P(A_{n+1} \cap B | A_n \cap B) = \frac{s - 1 - 4(n-1)}{s}.$$

Again by the total probability formula

$$P(A_{n+1} \cap B^c) = P(A_n \cap B)P(A_{n+1} \cap B^c | A_n \cap B) + P(A_n \cap B^c)P(A_{n+1} \cap B^c | A_n \cap B^c)$$

Since $P(A_{n+1} \cap B^c | A_n \cap B^c)$ is the probability that the $(n+1)$-st person was not born on leap day and has a date different than the previous non leap days, we have

$$P(A_{n+1} \cap B^c | A_n \cap B^c) = \frac{s - 1 - 4n}{s}.$$

Similarly, $P(A_{n+1} \cap B^c | A_n \cap B) = 0$ because once leap day is included among the first n it cannot be excluded by adding another person.

Hence the second method is the recursion $p_1 = \frac{1}{s}$, $q_1 = \frac{s-1}{s}$, and for $n \geq 1$,

$$\begin{aligned} p_{n+1} &= \frac{s + 3 - 4n}{s} p_n \ + \ \frac{1}{s} q_n, \\ q_{n+1} &= \qquad\qquad \frac{s - 1 - 4n}{s} q_n. \end{aligned}$$

Of course, both methods give the same result. To see it, (1) is for $n \geq 2$,

$$\begin{aligned} P(A_1) &= \frac{1}{s} \cdot 1 + \frac{s-1}{s} \cdot 1, \\ P(A_n) &= \frac{n}{s} \left(\frac{s-1}{s} \right)^{n-1} \frac{367 - 1}{365} \cdots \frac{367 - n}{365} + \left(\frac{s-1}{s} \right)^{n} \frac{366 - 1}{365} \cdots \frac{366 - n}{365}. \end{aligned}$$

But it turns out that $p_1 = 1/s$, $q_1 = (s-1)/2$ and for $n > 1$,

$$p_n = \frac{n}{s} \left(\frac{s-1}{s} \right)^{n-1} \frac{367 - 2}{365} \cdots \frac{367 - n}{365}; \qquad q_n = \left(\frac{s-1}{s} \right)^{n} \frac{366 - 2}{365} \cdots \frac{366 - n}{365}.$$

Evidently these quantities satisfy

$$\begin{aligned} p_{n=1} &= \frac{s-1}{s} \cdot \frac{366 - n}{365} p_n \ + \ \frac{1}{s} \cdot q_n \\ q_{n+1} &= \qquad\qquad \frac{s-1}{s} \cdot \frac{365 - n}{365} q_n \end{aligned}$$

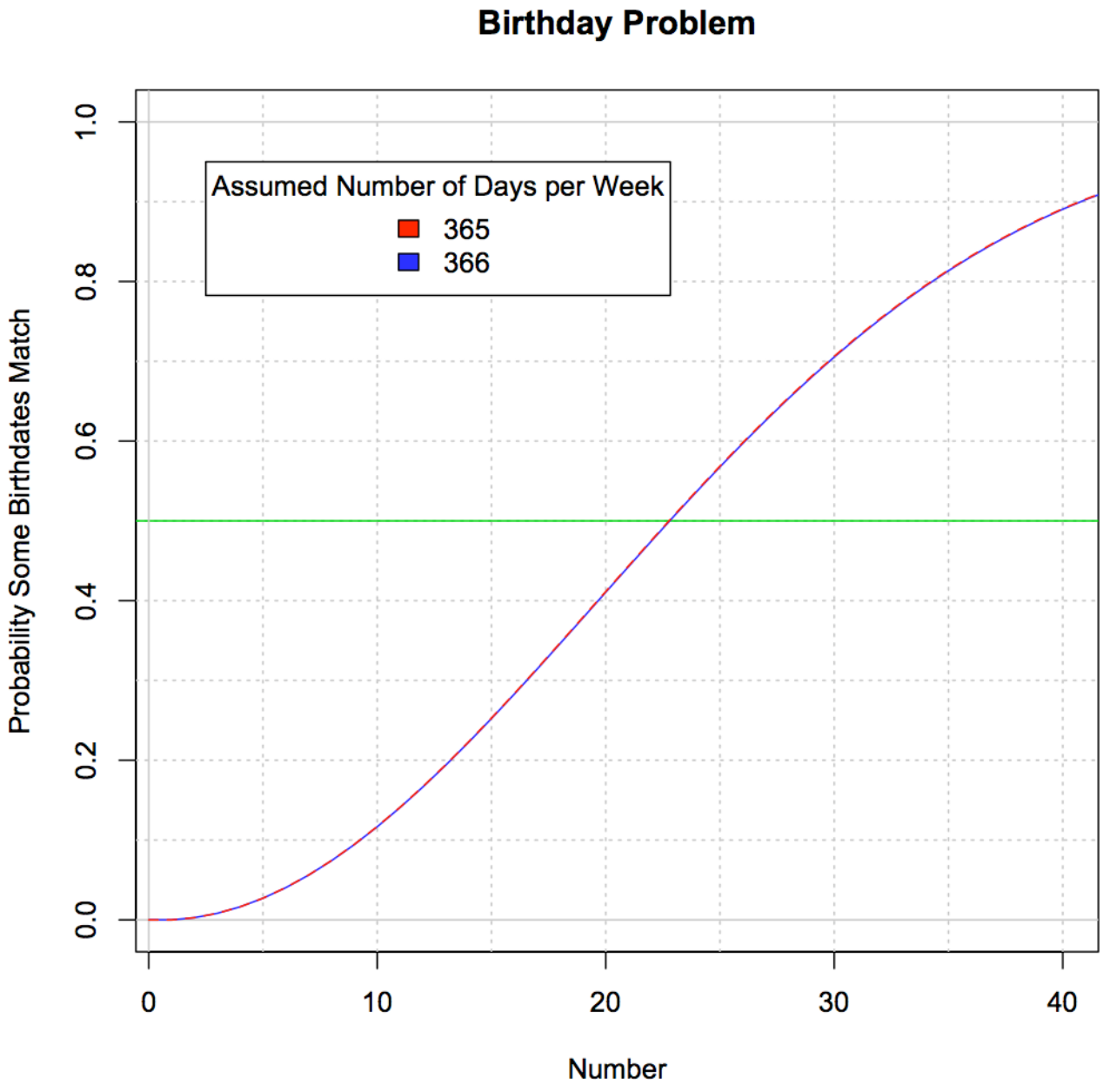
which is the same as our recursion because $s-1=4\cdot 365$ so

$$\frac{s-1}{s}\cdot\frac{366-n}{365}=\frac{s-1}{s}\cdot\frac{4\cdot 366-4n}{4\cdot 365}=\frac{s+3-4n}{s};$$
$$\frac{s-1}{s}\cdot\frac{365-n}{365}=\frac{s-1}{s}\cdot\frac{4\cdot 365-4n}{4\cdot 365}=\frac{s-1-4n}{s}.$$

R Session:

```
R version 2.10.1 (2009-12-14)
Copyright (C) 2009 The R Foundation for Statistical Computing
ISBN 3-900051-07-0

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[R.app GUI 1.31 (5538) powerpc-apple-darwin8.11.1]

[Workspace restored from /Users/andrejstreibergs/.RData]

> ################### COMPUTE PROB. ASSUMING ALL DAYS EQUALLY LIKELY ############
> n1 <- 365
> n2 <- 366
> x <- rep(1,n2); y <- rep(1,n2)
> for (i in 2:n2) {
+                  x[i] <- x[i-1]*(n1+1-i)/n1
+                  y[i] <- y[i-1]*(n2+1-i)/n2
+                 }
> x[1:20]-y[1:20]
 [1]  0.000000e+00 -7.485590e-06 -2.237485e-05
 [4] -4.446398e-05 -7.343066e-05 -1.088398e-04
 [7] -1.501518e-04 -1.967325e-04 -2.478655e-04
[10] -3.027659e-04 -3.605953e-04 -4.204774e-04
[13] -4.815150e-04 -5.428061e-04 -6.034605e-04
[16] -6.626156e-04 -7.194512e-04 -7.732028e-04
[19] -8.231740e-04 -8.687465e-04
> # Probs very close!
> ################## PLOT PROBABILITY OF MATCH VS. N ############################
> plot(x[1:40], type="n", main="Birthday Problem", ylim=0:1,
+ ylab = "Probability Some Birthdates Match", xlab = "Number")
```

```
> abline(h=0:1,col=8)
> abline(h=(1:9)/10,lty=3,col=8)
> abline(h=.5,col=3)
> abline(v=0,col=8)
> abline(v=seq(5,40,5),lty=3,col=8)
> lines(0:44,c(0,1-y[1:44]),col=4)
> lines(0:44,c(0,1-x[1:44]),col=2,lty=2)
> legend(2.5, .95, fill = c(2,4), legend = c(n1,n2),
+ title = "Assumed Number of Days per Week", bg="white")
```




```
>
>
```

```
> x[1:30]
 [1] 1.0000000 0.9972603 0.9917958 0.9836441 0.9728644
 [6] 0.9595375 0.9437643 0.9256647 0.9053762 0.8830518
[11] 0.8588586 0.8329752 0.8055897 0.7768975 0.7470987
[16] 0.7163960 0.6849923 0.6530886 0.6208815 0.5885616
[21] 0.5563117 0.5243047 0.4927028 0.4616557 0.4313003
[26] 0.4017592 0.3731407 0.3455385 0.3190315 0.2936838

> ################## HOW MANY SO PROB. MATCH > 50%? ############################
> k <- 1
> while(x[k] >= .5)k <- k+1
> k
[1] 23
> 1-x[k-1];1-x[k]
[1] 0.4756953
[1] 0.5072972
>
> cat("\n Assuming n=", n1,
+ "days per year,\n the smallest number needed so that P(match)>.5 is ",
+ k, ".\n The probability that a match occurs is\n P(match | n =",k-1,") =",
+ 1-x[k-1], "\n P(match | n =",k,") =", 1-x[k])

 Assuming n= 365 days per year,
 the smallest number needed so that P(match)>.5 is  23 .
 The probability that a match occurs is
 P(match | n = 22 ) = 0.4756953
 P(match | n = 23 ) = 0.5072972
>
> k <- 1
> while(y[k] >= .5)k <- k+1
> cat("\n Assuming n=", n2,
+ "days per year,\n the smallest number needed so that P(match)>.5 is ",
+ k, ".\n The probability that a match occurs is\n P(match | n =",k-1,") =",
+ 1-y[k-1], "\n P(match | n =",k,") =", 1-y[k])

 Assuming n= 366 days per year,
 the smallest number needed so that P(match)>.5 is  23 .
 The probability that a match occurs is
 P(match | n = 22 ) = 0.4747506
 P(match | n = 23 ) = 0.506323
>
> ############### COMPUTATION THAT TAKES LEAP YEAR INTO ACCOUNT ###############
> # No days in four years. Prob b-day is leap day is  pl
> s <- 365+365+365+366; s
[1] 1461

> pl <- 1/s
> po <- (s-1)/s
> # First two probs by hand (as a reality check!)
> zq <- 2*(s-1)/s^2 + (s-1)^2*(364/365)/s^2
> c(x[2],zq,y[2])
```

```
[1] 0.9972603 0.9972636 0.9972678
> zq2 <- 2*(s-1)^2/s^3*(364/365) + (s-1)^3*(364/365)*(363/365)/s^3 + (s-1)^2*(364/365)/s^3
> c(x[3],zq2,y[3])
[1] 0.9917958 0.9918056 0.9918182


> ###################### DO THE LEAP YEAR RECURSION ############################
> z <- rep(1, 400)
> pi <- pl; qi <- po
> for (i in 2:400) {
+                   pi <- pi*(s+7-4*i)/s + qi*pl
+                   qi <- qi*(s+3-4*i)/s
+                   z[i] <- pi + qi
+                  }


> ######################### LEAP YEAR 50% CHANCE OF MATCH NUMBER ###############
> k<-1
> while(z[k] >= .5)k <- k+1
> k
[1] 23

>  cat("\n Assuming that Every Fourth Year is a Leap Year,",
+ "\n the smallest number needed so that P(match)>.5 is ", k,
+ ".\n The probability that a match occurs is\n P(match | n =",k-1,") =",
+ 1-z[k-1], "\n P(match | n =",k,") =", 1-z[k])



 Assuming that Every Fourth Year is a Leap Year,
 the smallest number needed so that P(match)>.5 is  23 .
 The probability that a match occurs is
 P(match | n = 22 ) = 0.4752764
 P(match | n = 23 ) = 0.506865
```

```
> ######################### PRINT TABLE OF THREE PROBABILITIES ################
> m<-cbind(x[1:30],z[1:30],y[1:30])
> dimnames(m) <- list(1:30, c(paste(n1, "Days/Yr"), "Leap Year",
+ paste(n1, "Days/Yr")))

> cat("Probability that Birthdates Match\n\n   Assuming    Including   Assuming\n")
> m
Probability that Birthdates Match

   Assuming    Including   Assuming
   365 Days/Yr Leap.Yr     366 Days/Yr
1  0.000000000 0.000000000 0.000000000
2  0.002739726 0.002736445 0.002732240
3  0.008204166 0.008194354 0.008181791
4  0.016355912 0.016336402 0.016311448
5  0.027135574 0.027103335 0.027062143
6  0.040462484 0.040414671 0.040353644
7  0.056235703 0.056169704 0.056085551
8  0.074335292 0.074248768 0.074138560
9  0.094623834 0.094514757 0.094375968
10 0.116948178 0.116814863 0.116645412
11 0.141141378 0.140982508 0.140780783
12 0.167024789 0.166839428 0.166604311
13 0.194410275 0.194197883 0.193928760
14 0.223102512 0.222862946 0.222559706
15 0.252901320 0.252634829 0.252297859
16 0.283604005 0.283311222 0.282941390
17 0.315007665 0.314689584 0.314288214
18 0.346911418 0.346569374 0.346138215
19 0.379118526 0.378754165 0.378295352
20 0.411438384 0.411053628 0.410569637
21 0.443688335 0.443285347 0.442778947
22 0.475695308 0.475276447 0.474750646
23 0.507297234 0.506865017 0.506323012
24 0.538344258 0.537901311 0.537346429
25 0.568699704 0.568248726 0.567684368
26 0.598240820 0.597784533 0.597214124
27 0.626859282 0.626400394 0.625827329
28 0.654461472 0.654002637 0.653430231
29 0.680968537 0.680512318 0.679943765
30 0.706316243 0.705865080 0.705303412
```