

Computation: Three Birthday Problem

Austin Pesina

6/4/2021

```
library(ggplot2)
```

What is the probability that three or more people have a common birthday?

Let k be the number of people. Exclude February 29 and assume the other 365 days are equally likely. Assume independence of births.

Let $k = 365$.

$$P(\text{nomatch}) = \frac{365 \cdot 364 \cdot 363 \cdot \dots \cdot (365 - k + 1)}{365^k}$$

$$P(\text{match}) = \binom{k}{3}$$

The probability of multiple people sharing a birthday will depend on the number of people in the room. R has a built in function under the `stats` library called `pbirthday` that we will use to compute the probability of coincidences.

Note: This function uses n instead of k . It also allows us to change the number of coincidences.

```
x <- seq(0, 300, by = 25)
for (i in x)

print(paste0(pbirthday(i, coincident = 3)), sep = "")
```

```
## [1] "0"
## [1] "0.0183997270354577"
## [1] "0.131678633857538"
## [1] "0.36429055243294"
## [1] "0.639821738788553"
## [1] "0.850025464170109"
## [1] "0.95583857412604"
## [1] "0.991051977355366"
## [1] "0.998774602349259"
## [1] "0.99988772273296"
## [1] "0.99999314260916"
## [1] "0.99999972040881"
## [1] "0.999999992347181"
```

```
pbirthday(88, coincident = 3)
```

```
## [1] 0.509759
```

By using a for loop, we see that when $k = 75$, our probability of 3 people sharing a birthday is 36.4%. When we increase to $k = 100$, our probability jumps up to 63.9%. When $k = 88$, we get our first value over 50%.

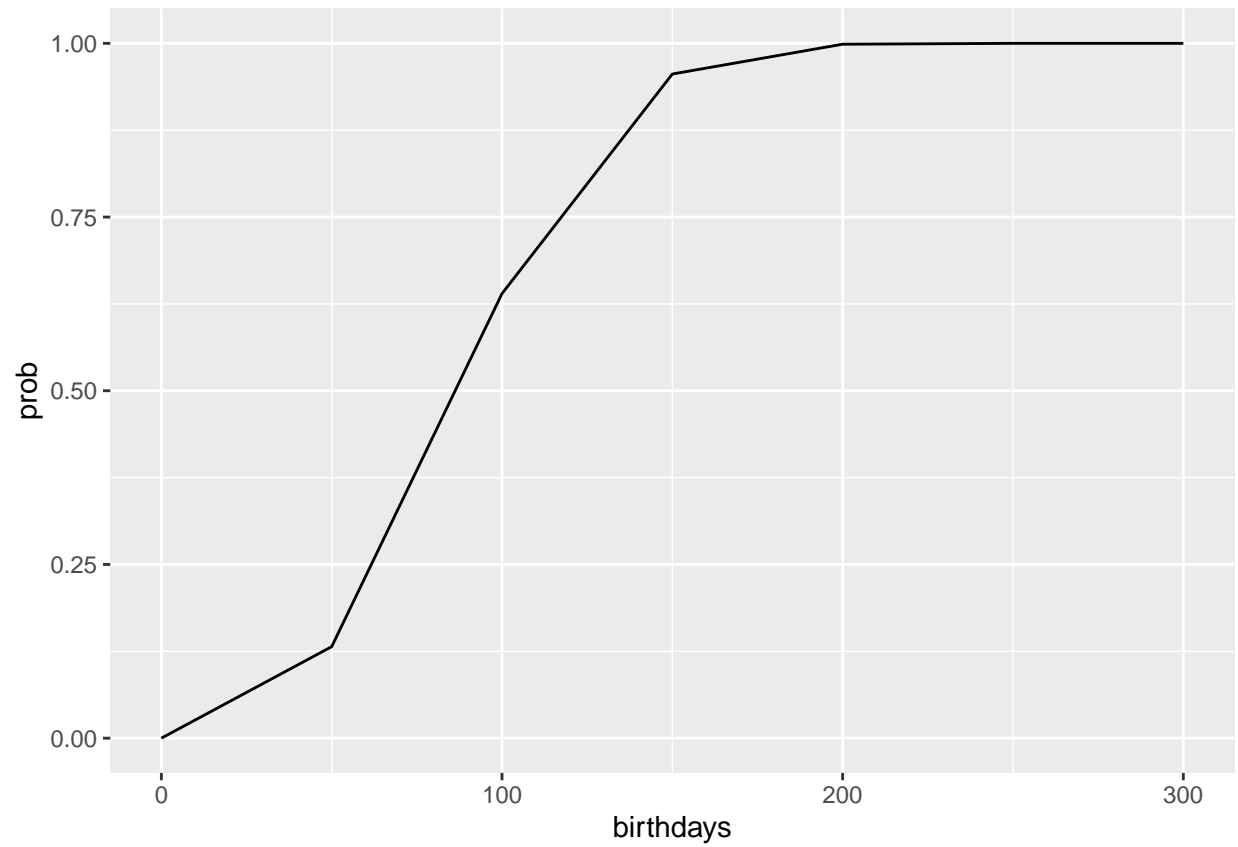
```
a <- pbirthday(0, coincident = 3)
b <- pbirthday(50, coincident = 3)
c <- pbirthday(100, coincident = 3)
d <- pbirthday(150, coincident = 3)
e <- pbirthday(200, coincident = 3)
f <- pbirthday(250, coincident = 3)
g <- pbirthday(300, coincident = 3)

tab <- matrix(c(0, 50, 100, 150, 200, 250, 300, a, b, c, d, e, f, g), ncol = 2, byrow = F)
tab
```

```
##      [,1]      [,2]
## [1,]    0 0.0000000
## [2,]   50 0.1316786
## [3,]  100 0.6398217
## [4,]  150 0.9558386
## [5,]  200 0.9987746
## [6,]  250 0.9999931
## [7,]  300 1.0000000
```

```
birthdays <- seq(0, 300, by = 50)
prob <- c(a,b,c,d,e,f,g)
df <- data.frame(birthdays, prob)

ggplot(data = df) +
  geom_line(mapping = aes(x = birthdays, y = prob))
```



As we can see, once we have $k = 200$, we have an almost 100% probability of having 3 people with the same birthday.