Probability and Conditional Probability

Introduction to Quantitative Social Science

Xiaolong Yang University of Tokyo

Summer 2022

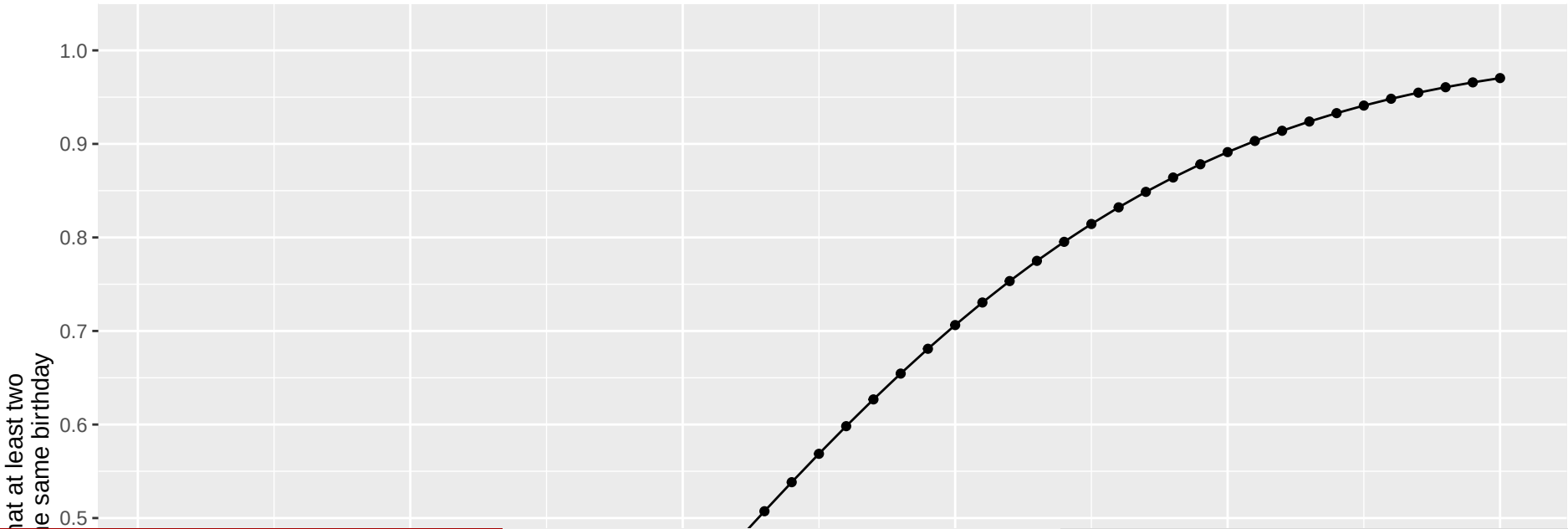
Birthday Problem

```
## write the birthday function
birthday <- function(k) {
logdenom <- k * log(365) + lfactorial(365 - k)
lognumer <- lfactorial(365)
pr <- 1 - exp(lognumer - logdenom)
pr
}
```

```
## create a tibble with the k and pr per k
bday <- tibble(k = 1:50, pr = birthday(k))
```

Birthday Problem

```
## plot (need to adjust size)
ggplot(bday, aes(x = k, y = pr)) +
geom_line() +
geom_point() +
scale_y_continuous(str_c("Probability that at least two",
                         "people have the same birthday", sep
                   limits = c(0, 1), breaks = seq(0, 1, by = 0
labs(x = "Number of people")
```



Ph.D. Admission Problem Solution via Monte Carlo Simulation

```
## setting seed for replication
set.seed(4444)
k <- 23 # number of people
sims <- 1000 # number of simulations
event <- 0 # counter
for (i in 1:sims) {
days <- sample(1:365, k, replace = TRUE)
days.unique <- unique(days) # unique birthdays
## if there are duplicates, the number of unique birthdays
## will be less than the number of birthdays, which is `k'
if (length(days.unique) < k) {
event <- event + 1
}
}
```

Solution via Monte Carlo Simulation

```
## fraction of trials where at least two bdays are the same
answer <- event / sims
answer
```

Need to make a plot here myself. qss tidyverse p280