Probability and Conditional Probability

Introduction to Quantitative Social Science

Kosuke Imai
Harvard University / University of Tokyo

Summer 2022

Permutations and Combinations

1. Permutations
   - How to count # of ways to arrange objects?
   - {A, B, C} has 6 permutations: ABC, ACB, BAC, BCA, CAB, CBA
   - Sampling without replacement: # of permutations of n elements taken k at a time

$$ {}_nP_k = n(n-1)\cdots(n-k+1) = \frac{n!}{(n-k)!} $$

   where $k \leq n$ and $0! = 1$

2. Combinations
   - Ways to select objects without regard to their arrangement
   - # of combinations of k distinct elements from a pool of n elements

$$ {}_nC_k = \binom{n}{k} = \frac{{}_nP_k}{k!} = \frac{n!}{k!(n-k)!} $$

   - One combination yields $k!$ permutations: order does not matter

Birthday Problem

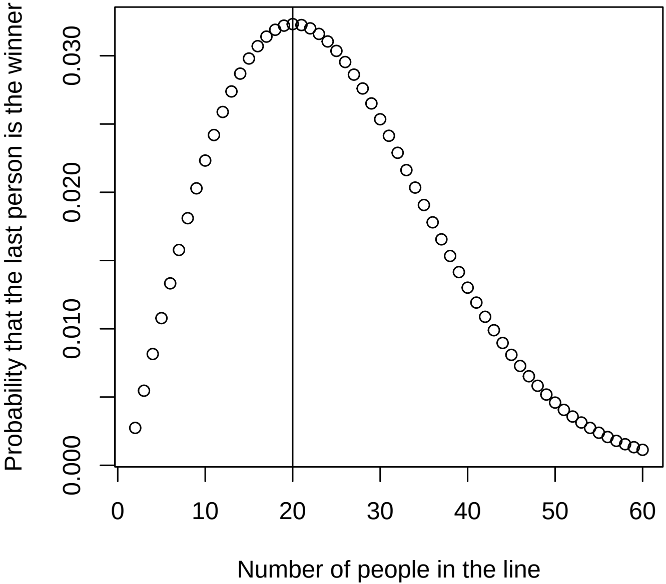
- Original Problem: How many people do you need in order for the probability that at least two people have the same birthday to exceed 0.5?

- Modified Problem: People line up. The winner is the first person who shares birthday with someone ahead in the line. How many people do you need in order to maximize the probability that the last person in the line wins?

```r
birthday.line <- function(k) {
    logdenom <- k * log(365) + lfactorial(365-k+1)
    lognumer <- lfactorial(365) + log(k - 1)
    result <- exp(lognumer - logdenom)
    return(result)
}
k <- 2:60 # number of people to examine
bday <- birthday.line(k)
names(bday) <- k # add labels
bday[bday == max(bday)] # get maximum

##     20
## 0.0323
```

```r
plot(k, bday, xlab = "Number of people in the line",
     ylab = "Probability that the last person is the winner")
abline(v = k[bday == max(bday)])
```


Probability that the last person is the winner
0.000
0.010
0.020
0.030
0
10
20
30
40
50
60
Number of people in the line

Conditional Probability and Independence

- $P(A \mid B)$ is the conditional probability of event A occurring given that event B occurs: e.g., $P(\text{support Clinton} \mid \text{support Obama})$

$$P(A \mid B) \ = \ \frac{\text{joint probability}}{\text{marginal probability}} \ = \ \frac{P(A \text{ and } B)}{P(B)}$$

- Multiplication Rule:

$$P(A \text{ and } B) \ = \ P(A \mid B)P(B) \ = \ P(B \mid A)P(A)$$

- Law of total probability:

$$P(A) \ = \ P(A \mid B)P(B) + P(A \mid \text{not } B)P(\text{not } B)$$

- Independence: Two events A and B are said to be independent if

$$P(A \text{ and } B) = P(A)P(B)$$

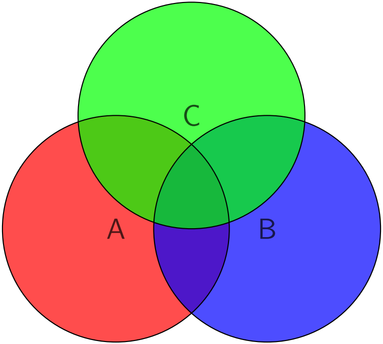
- If A and B are independent, then

$$P(A \mid B) = P(A)$$

Ph.D. Admission Problem

Getting into the Ph.D. program of the Politics Department is known to be very difficult, and there is considerable uncertainty about the admission process. Every year, a secretary types each applicant's name on a separate card together with a matching envelope. Then, he drops the pile from the window of his office on the second floor of Corwin Hall. Finally, he goes downstairs and places the cards randomly in the envelopes. If the name on a card matches with a name on the envelope in which the card is placed, the applicant will be admitted. What is the probability of nobody getting accepted? Does this probability vary as you change the number of applicants from 5 to 500 (by 25)? Try an analytical approach with 5 applicants. Use a Monte Carlo simulation for the rest.

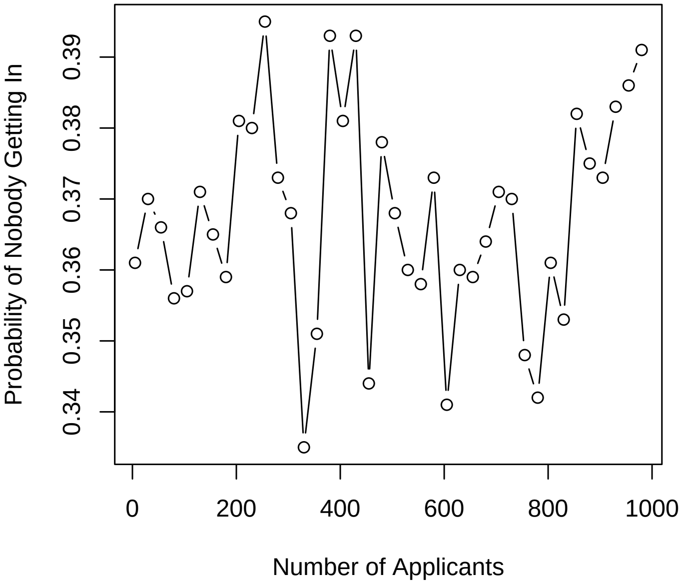
Inclusion-Exclusion Principle



$$\begin{aligned} P(A \text{ or } B \text{ or } C) &= P(A)+P(B)+P(C) \\ &-\{P(A \text{ and } B)+P(B \text{ and } C)+P(A \text{ and } C)\} \\ &+P(A \text{ and } B \text{ and } C) \end{aligned}$$

Solution via Monte Carlo Simulation

```r
sims <- 1000 ## number of simulations
num <- seq(5, 1000, by = 25) ## number of applicants
res <- rep(NA, length(num))
for (j in 1:length(num)){
  nobody <- 0 ## set counter to zero
  applicants <- 1:num[j]
  for(i in 1:sims){
    cards <- sample(applicants, num[j], replace = FALSE)
    envelopes <- sample(applicants, num[j], replace = FALSE)
    if (sum(cards == envelopes) == 0)
      nobody <- nobody+1
  }
  res[j] <- nobody / sims
}
plot(num, res, type = "b", xlab = "Number of Applicants",
     ylab = "Probability of Nobody Getting In")
```


Probability of Nobody Getting In
0.34
0.35
0.36
0.37
0.38
0.39
0
200
400
600
800
1000
Number of Applicants

Bayes' Rule

- From the conditional probability formula, we have:

$$\underbrace{\Pr(A \mid B)}_{\textit{conditional probability}} = \frac{\overbrace{\Pr(A \text{ and } B)}^{\textit{joint probability}}}{\underbrace{\Pr(B)}_{\textit{marginal probability}}}$$

$$= \frac{\Pr(B \mid A)\Pr(A)}{\Pr(B \mid A)\Pr(A) + \Pr(B \mid \text{not } A)\Pr(\text{not } A)}$$

- Knowledge of $\Pr(A)$, $\Pr(B \mid A)$, and $\Pr(B \mid \text{not } A)$ gives you $\Pr(A \mid B)$
- Bayesian updating: prior belief $\Pr(A) \rightsquigarrow$ posterior belief $\Pr(A \mid B)$

Monty Hall via Bayes' Rule

- Problem: You pick door A. Monty opens door C that has a goat. Should you switch to door C?
- YouTube explanation: http://www.youtube.com/watch?v=mhlc7peGlGg

- Prior belief: $P(A) = P(B) = P(C) = 1/3$
- Data: Monty reveals C
- Posterior belief: $P(A \mid \text{MC})$ and $P(B \mid \text{MC})$
- Question: $P(A \mid \text{MC}) < P(B \mid \text{MC})$

Prisoner's Paradox

Three prisoners, A, B, and C, are on death row. The governor decides to pardon one of the three and chooses at random the prisoner to pardon. He informs the warden of his choice but requests that the name be kept secret for a few days. The next day, A tries to get the warden to tell him who had been pardoned. The warden refuses. A then asks which of B or C will be executed. The warden thinks for a while, then tells A that B is to be executed. Did the warden give A any information about whether A will be pardoned?