Bush 631-603: Quantitative Methods

Lecture 9 (03.22.2022): Probability vol. I

Rotem Dvir

The Bush school of Government and Public Policy

Texas A&M University

Spring 2022

What is today's plan?

- Calculating uncertainty: **probability**
- What is probability? why should we learn it?
- Probability theory (some equations. . .)
- How we use probability in the real world?
- R work: prop.table(), addmargins().
- Final project: the data report.

Learning from data

Our 8-week quest:

- How to estimate causal effects.
- Understand measurement challenges.
- Assess correlations.
- Generate prediction about unknown quantities.

The question now?

How do we know our estimates are 'real' or just due to random chance?

We have findings!!!

- Data patters are systematic? Or noise?
- Our estimates $\rightarrow$ real relationship or random?

Solutions:

- Select (at random) a different treatment/control group.
- Select (at random) a different sample.
- Method to **quantify the degree of statistical uncertainty** of empirical findings.

Probability

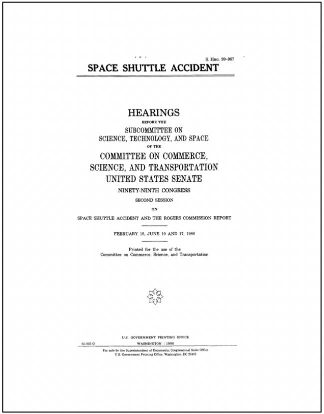
Intro to probabilities

Probability:

- Set of tools to measure uncertainty in world (and our data).
- Method to formalize uncertainty or chance variation.
- Define odds for all (defined) possible outcomes.

What's the chance?

January 28, 1986: Challenger shuttle



S. Hrg. 99-967

SPACE SHUTTLE ACCIDENT

HEARINGS
BEFORE THE
SUBCOMMITTEE ON
SCIENCE, TECHNOLOGY, AND SPACE
OF THE
COMMITTEE ON COMMERCE,
SCIENCE, AND TRANSPORTATION
UNITED STATES SENATE
NINETY-NINTH CONGRESS
SECOND SESSION
ON
SPACE SHUTTLE ACCIDENT AND THE ROGERS COMMISSION REPORT

FEBRUARY 18, JUNE 10 AND 17, 1986

Printed for the use of the
Committee on Commerce, Science, and Transportation

U.S. GOVERNMENT PRINTING OFFICE
52-885 O WASHINGTON : 1986

For sale by the Superintendent of Documents, Congressional Sales Office
U.S. Government Printing Office, Washington, DC 20402

Probabilites translated

Challenger accident - chance of failure?

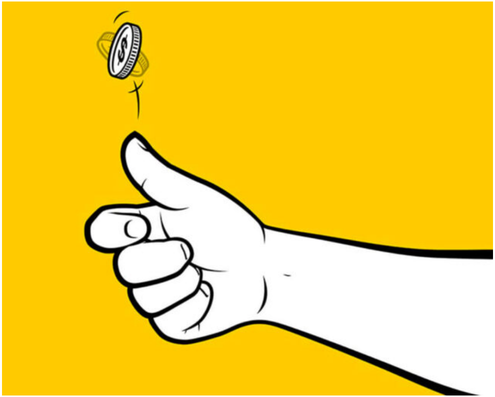
- Experts: 100-1.
- NASA management: 100,000-1.

- What is 100,000 in 1?
- Repeated testing and odds of event (failure).
- Enough events? we can calculate probabilities...

Probability explained

- Probability $\rightarrow$ measure randomness.
- Random $\neq$ complete unpredictability:
  - Short-term: unpredictable (very hard to calculate).
  - Long-term: predictable (multiple repetitions).

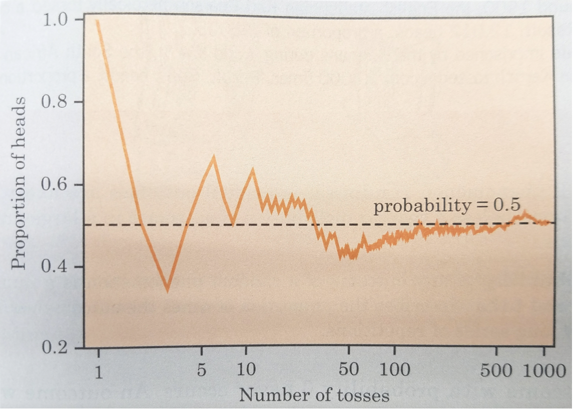
Probability explained



- Odds for heads? and tails?
- Overall: 0.5 probability H/T.

Coin toss chances

- 5 flips: HHHHT
- How 0.5 exactly?



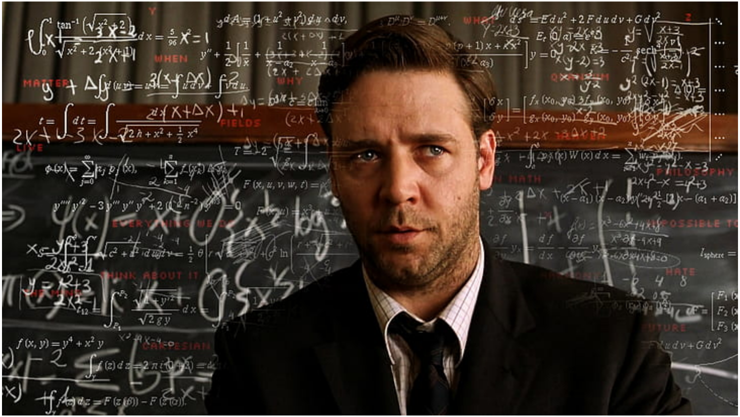
The secret?

Repetition - multiple iterations

- Estimate probability.
- Why only estimate? “toss again. . . ”
- Mathematical probability - ideal in infinite series of trails.
- Explain long-term regularity of random event (behavior).

Figuring the odds

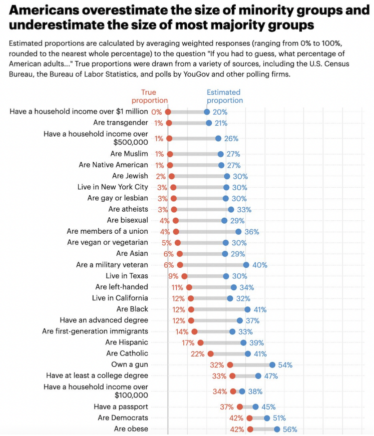
Why?



Figuring the odds

Can we estimate the odds?

- Religions?
- Place of residence: TX, NY?
- Other groups/identities?



Figuring the odds

Law of Averages: When?

Figuring the odds

Rare event and our behavior

TABLE 1.1 How Dangerous Is Terrorism?

Cause of Death	*Times more likely to kill an American compared to a terrorist attack*
Heart disease	35,079
Cancer	33,842
Alcohol-related death	4,706
Car accident	1,048
Risky sexual behavior	452
Fall	353
Starvation	187
Drowning	87
Railway accident	13
Accidental suffocation in bed	12
Lethal force by a law enforcement officer	8
Accidental electrocution	8
Hot weather	6

	% Critical threat	% Important but not critical threat
International terrorism	79	18
Development of nuclear weapons by Iran	75	18

Figuring the odds

Solve this:

Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.

Schools of thought

Frequentist

- The *limit* of relative frequency.
- Ratio of number of events occur and total number of trails.
- Challenge: same conditions??

Bayesian

- Measure of *subjective* belief about an event occurring.
- Challenge: how to conduct science?

Probability theory

Concepts, axions and definitions

- Sample space (Ω): set of all possible outcomes.
- Event: any subset of outcomes in sample space.
- Card deck: 52 cards (13 rank) x (4 suits)
- Trial: pick a card at-random

Sample space:

$2\clubsuit$ $3\clubsuit$ $4\clubsuit$ $5\clubsuit$ $6\clubsuit$ $7\clubsuit$ $8\clubsuit$ $9\clubsuit$ $10\clubsuit$ $J\clubsuit$ $Q\clubsuit$ $K\clubsuit$ $A\clubsuit$
$2\spadesuit$ $3\spadesuit$ $4\spadesuit$ $5\spadesuit$ $6\spadesuit$ $7\spadesuit$ $8\spadesuit$ $9\spadesuit$ $10\spadesuit$ $J\spadesuit$ $Q\spadesuit$ $K\spadesuit$ $A\spadesuit$
$2\heartsuit$ $3\heartsuit$ $4\heartsuit$ $5\heartsuit$ $6\heartsuit$ $7\heartsuit$ $8\heartsuit$ $9\heartsuit$ $10\heartsuit$ $J\heartsuit$ $Q\heartsuit$ $K\heartsuit$ $A\heartsuit$
$2\diamondsuit$ $3\diamondsuit$ $4\diamondsuit$ $5\diamondsuit$ $6\diamondsuit$ $7\diamondsuit$ $8\diamondsuit$ $9\diamondsuit$ $10\diamondsuit$ $J\diamondsuit$ $Q\diamondsuit$ $K\diamondsuit$ $A\diamondsuit$

An event: picking a Queen, $\{Q\clubsuit, Q\spadesuit, Q\heartsuit, Q\diamondsuit\}$

Probability

Calculate probability of event:

$$P(A) = \frac{Elements(A)}{Elements(\Omega)}$$

Example: coin toss x 3

Sample space (Ω): {HHH,HHT,HTH,HTT,THH.THT,TTH,TTT}.

Get an least two heads?

Event A: {HHH,HHT,HTH,THH}.

Probability: $P(A) = \frac{4}{8} = 0.5$

Probability

- Define how likely/unlikely events are.
- Based on three axioms:
    1. Probability of any event A is nonnegative ($P(A) >= 0$).
    2. Normalization ($P(\Omega) = 1$).
    3. Addition rule - If events A and B are mutually exclusive then $P(AorB) = P(A) + P(B)$
- Axioms 1&2 $\rightarrow 1 > P(event) > 0$

Gambling 101

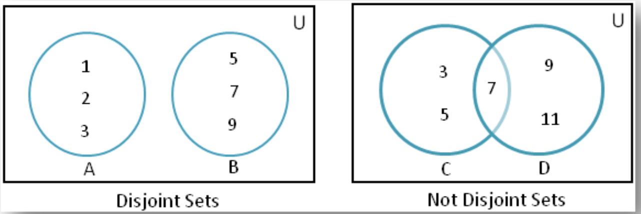
Probability of mutually exclusive events

- What is P(A) $\rightarrow$ select Queen card at-random?

- Any card selection: $1/52$.
- Select queen event: $\{Q\clubsuit, Q\diamondsuit, Q\heartsuit, Q\spadesuit\}$.
- P (event) = union of mutually exclusive events $\rightarrow$ addition rule
- $P(Q) = P(Q\clubsuit) + P(Q\diamondsuit) + P(Q\heartsuit) + P(Q\spadesuit) = \frac{4}{52} \approx 7.7\%$

Events relationships

Mutually & not Mutually exclusive events

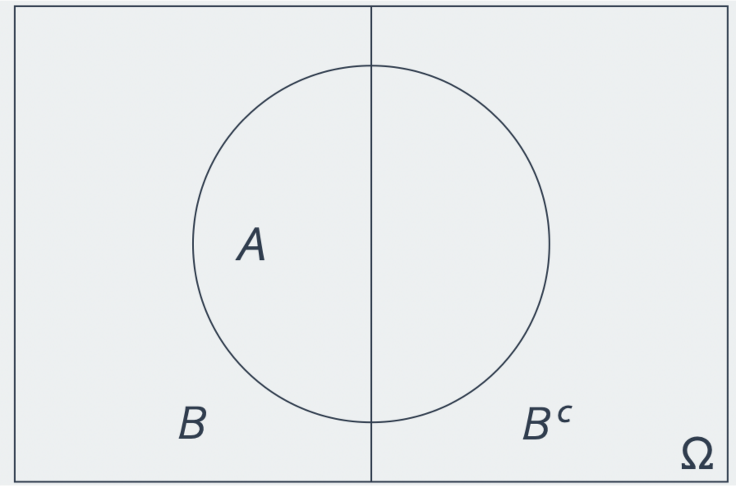


Probability facts

- Probability of complement: $P(A^C) = P(notA) = 1 - P(A)$
- Probability of **not drawing** a Queen: $1 - \frac{4}{52} = \frac{48}{52}$
- Probability of events (not disjointed): the presidential race with 3rd candidate.
- General addition rule: $P(AorB) = P(A) + P(B) - P(A\&B)$
- Cards example: probability of Queen or ♣?
- Queen ($\frac{4}{52}$) + ♣ ($\frac{13}{52}$) - Q♣ ($\frac{1}{52}$) $= \frac{16}{52}$

Probability facts

- Law of total probability: $P(A) = P(A\&B) + P(A\&B^C)$



Calculating outcomes

- **Permutations**: enumerating all possible outcomes.
- Ordering three events (A/B/C): {ABC,ACB,BAC,BCA,CAB,CBA}.
- A short-cut??

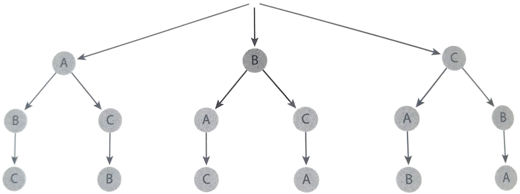


Figure 6.3. A Tree Diagram for Permutations. There are 6 ways to arrange 3 unique objects. Source: Adapted from example by Madit, http://texample.net.

Calculating outcomes

- General permutation formula:

$$ {}_nP_k = n * (n-1) * ... * (n-k+2) * (n-k+1) = \frac{n!}{(n-k)!} $$

- How many ways to sit 5 students in our class?

```
# Use permutations formula
factorial(25)/factorial(20)
```

```
## [1] 6375600
```

Permutations

- *The birthday problem*:
  - What n so P(two people share birthday) > 0.5?
  - Easier route by looking at complement.
  - Find $\rightarrow$ 1 - P(nobody has the same birthday).

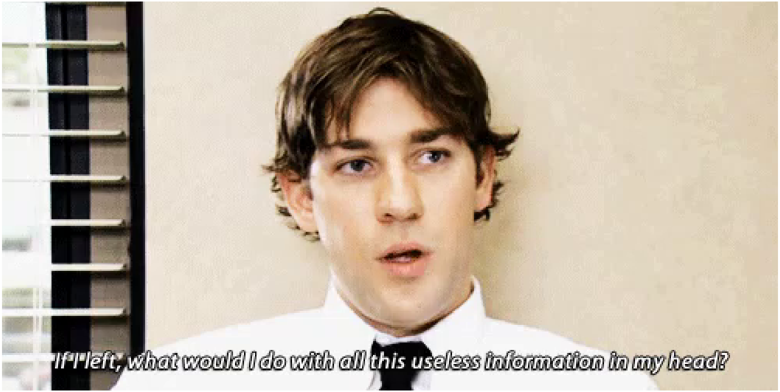
```
bday <- function(k){
  logdenom <- k * log(365) + lfactorial(365-k)
  lognumer <- lfactorial(365)
  pr <- 1 - exp(lognumer - logdenom)
  return(pr)
}

k <- 1:25
test_bday <- bday(k)
names(test_bday) <- k

test_bday[19:25]
```

```
##        19        20        21        22        23        24        25
## 0.3791185 0.4114384 0.4436883 0.4756953 0.5072972 0.5383443 0.5686997
```

Sampling procedures

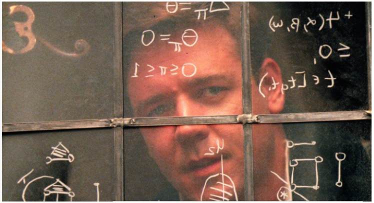
- **With replacement**:
  - Same unit can be ‘selected’ repeatedly.
  - Replace card in stack after draw.
  - Two people born on the same day.
- **Without replacement**:
  - Each unit can be sampled at most once.
  - Card removed after draw.
- Procedure matters for probability calculations.
- **Combinations**: another counting method (ignore ordering).

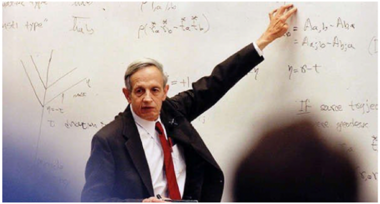
And. . .


If I left, what would I do with all this useless information in my head?

And. . .

Probabilities and the real-world





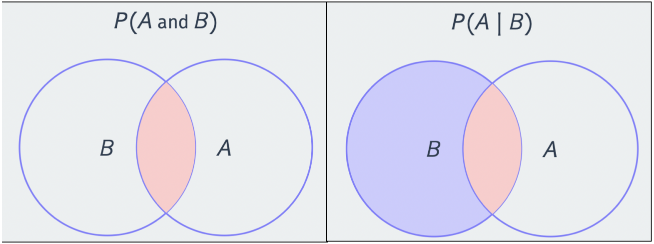
Using probability in bars?

- Setting: 5 men, 5 women.
- Objective: get a dance.
- All go for blonde $\rightarrow$ P(dance) $= \frac{1}{4}$
- Each man $\rightarrow$ non-blonde: P(dance) $= \frac{1}{1}$
- *Nash equilibrium*: no incentive to deviate.
- Mutual cooperation: global trade, negotiations (prisoner's dilemma).

Conditional probability

- We know event B occurred, what is the probability of event A?
- Examples:
  - What is the probability of two states going to war if they are both democracies?
  - What is the probability of a judge ruling in a pro-choice direction conditional on having daughters?
  - What is the probability that there will be a coup in a country conditional on having a presidential system?

Conditional probability



$$P(A|B) = \frac{P(A\&B)}{P(B)}$$

Conditional probability

- Conditioning information matters!
- Twins:
  - Sample space: $\Omega = \{\text{GG,GB,BG,BB}\}$.
  - $\text{P(BB} \mid \text{at least one boy)} = \text{P(BB} \mid \text{elder is a boy)}$??

$$\text{P(BB} \mid \text{at least one boy)} = \frac{P(BB\&(BB|BG|GB))}{P(BB|BG|GB)} = \frac{P(BB)}{P(BB|BG|GB)} = \frac{1/4}{3/4} = \frac{1}{3}$$

$$\text{P(BB} \mid \text{elder is a boy)} = \frac{P(BB\&(BB|BG))}{P(BB|BG)} = \frac{P(BB)}{P(BB|BG)} = \frac{1/4}{1/2} = \frac{1}{2}$$

Conditioning info: numbers and Aggs

Aggies in the NFL: position groups and conferences

```
head(Ags)
```

```
## # A tibble: 6 x 5
##   Player            Team              Position Group Conference
##   <chr>             <chr>             <chr>    <chr> <chr>
## 1 Christian Kirk    Arizona Cardinals WR       OF    NFC
## 2 Jake Matthews     Atlanta Falcons   OT       OF    NFC
## 3 Otaro Alaka       Baltimore Ravens  LB       DF    AFC
## 4 Justin Madubuike Baltimore Ravens  DT       DF    AFC
## 5 Tyrel Dodson      Buffalo Bills     LB       DF    AFC
## 6 Germain Ifedi     Chicago Bears     OG       OF    NFC
```

Conditioning info: numbers and Aggs

```
# Tabulate data
t <- table(Conf = Ags$Conference, Pos.Grp = Ags$Group)
addmargins(t)
```

```
##      Pos.Grp
## Conf  DF OF ST Sum
##   AFC  8 10  2  20
##   NFC  4 12  2  18
##   Sum 12 22  4  38
```

- Choose one at-random.
- What is probability of choosing Offense?
  - $P(OF) = \frac{22}{38} = 0.57$
- What is probability of choosing Offense & NFC?
  - $P(OF \ \& \ NFC) = \frac{12}{38} = 0.31$
- What is probability that randomly selected NFC is offense?
  - $P(OF \mid NFC) = \frac{P(OF\&NFC)}{P(NFC)} = \frac{12/38}{18/38} = 0.66$

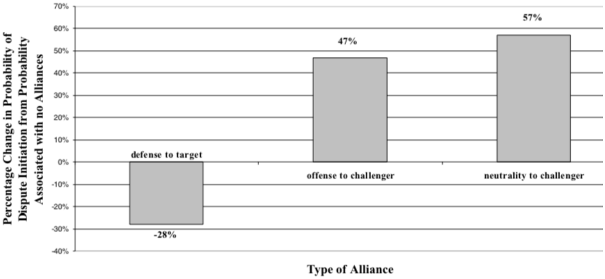
Conditional probability in Global affairs

Military alliances: a contract

Global military alliances

Leeds (2003):

- Defensive cooperation.
- Offensive cooperation.
- Neutrality.
- Non-aggression.
- Consultation.



Probability and data

Military Alliances (ATOP) data (1815-2018)

```
## # A tibble: 6 x 24
##   atopid member yrent moent ineffect estmode pubsecr secrart length
##    <dbl>  <dbl> <dbl> <dbl>    <dbl>   <dbl>   <dbl>   <dbl>  <dbl>
## 1   3150     58  1981    12        1       1       0       0      0
## 2   3900     58  1981     6        1       1       0       0      0
## 3   4778     58  1996     3        1       1       0       0      0
## 4   2075    700  1921     3        0       1       0       0      0
## 5   2090    700  1921     6        0       1       0       0      0
## 6   2170    700  1926     8        0       1       0       0     36
## # ... with 14 more variables: offense <dbl>, neutral <dbl>, nonagg <
## #   consul <dbl>, active <dbl>, notaiden <dbl>, terrres <dbl>, spect
## #   milaid <dbl>, base <dbl>, armred <dbl>, ecaid <dbl>, StateAbb <c
## #   StateName <chr>
```

Probability in R

- Alliance & domestic ratification

```
# Probabilities for domestic ratification
prop.table(table(Ratification = atop2$estmode))
```

```
## Ratification
##         0         1
## 0.2187919 0.7812081
```

```
# Probabilities for secret provisions
prop.table(table(publicity = atop2$pubsecr))
```

```
## publicity
##          0          1          2
## 0.92557828 0.01709688 0.05732484
```

Probability in R

- Alliance $\rightarrow$ commitment.
- US guarantee military assistance?

```
# Subset data (tidyverse): US alliances only
atop.us <- atop2 %>%
  filter(member == 2)

# Probability of military commitment
prop.table(table(atop.us$defense))
```

```
##
##         0         1
## 0.4210526 0.5789474
```

- Conditional probability

```
## Types of military aid given that alliance has defensive provision
prop.table(table(atop2$milaid[atop2$defense == 1]))
```

```
##
##          0          1          2          3          4
## 0.81632653 0.03755102 0.01551020 0.11183673 0.01877551
```

Probability in R

- Joint probability tables
- Marginal probabilities → sum of rows/columns

```
# Defense and Offense provisions
j1 <- prop.table(table(def = atop2$defense, off = atop2$offense))
addmargins(j1)
```

```
##      off
## def            0          1        Sum
##   0   0.56569709 0.01738549 0.58308258
##   1   0.33132732 0.08559010 0.41691742
##   Sum 0.89702441 0.10297559 1.00000000
```

```
# Offensive and secret provisions
j2 <- prop.table(table(secret = atop2$secrart, off = atop2$offense))
addmargins(j2)
```

```
##       off
## secret           0           1         Sum
##    0   0.849480389 0.076097888 0.925578277
##    1   0.003687563 0.000000000 0.003687563
##    3   0.003687563 0.000000000 0.003687563
##    4   0.004022796 0.001005699 0.005028495
##    5   0.001005699 0.000000000 0.001005699
##    6   0.000000000 0.001005699 0.001005699
##    7   0.002681864 0.000000000 0.002681864
##    8   0.034193765 0.023131076 0.057324841
##    Sum 0.898759638 0.101240362 1.000000000
```

Independence

- Events are not related.
- Knowing the A occurred does not affect the probability of B occurring.
- Marginal probability of B (knowing A occurred) remains P(B).
- Formally:
  - $P(A \& B) = P(A) * P(B)$
  - $P(A|B) = P(A)$
  - $P(B|A) = P(B)$

Independence in ATOP data

- Defense treaties & Economic aid: related?

```
# Marginal probability: levels of economic aid
prop.table(table(EconAid = atop2$ecaid))
```

```
## EconAid
##          0          1          2          3
## 0.88870037 0.01798439 0.02341364 0.06990159
```

```
# Marginal probability: defense alliance
prop.table(table(Defense = atop2$defense))
```

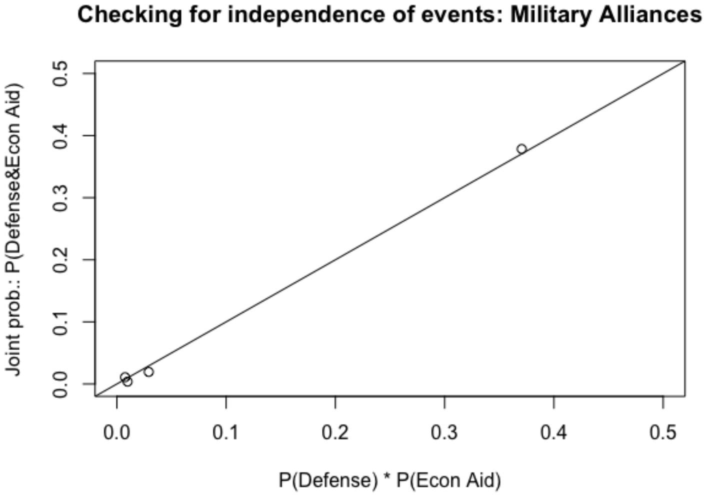
```
## Defense
##         0         1
## 0.5830826 0.4169174
```

```
# Joint probability: defense and econ aid
prop.table(table(Defense = atop2$defense, EconAid = atop2$ecaid))
```

```
##        EconAid
## Defense           0           1           2           3
##       0 0.510349508 0.007125891 0.019681032 0.050559891
##       1 0.378350865 0.010858500 0.003732609 0.019341703
```

Plotting independence

Defense treaties & Economic aid



Independence

- Throw conditional probability into the mix.
- The *Monty Hall problem*:

Bayesian probability

- The subjective side of probability estimates.
- How prior knowledge and new evidence shape our behavior?

- Bayes rule: mathematical solution to update our beliefs.

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)} = \frac{P(B|A) * P(A)}{P(B|A) * P(A) + P(B|A^C) * P(A^C)}$$

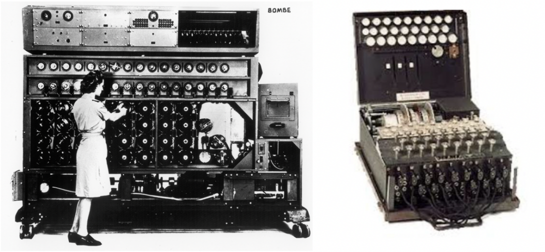
- P(A): *prior probability*.
- Event B occur.
- P(A|B) = *posterior probability*

Bayes in real life

- Where is my phone/pager?
- Health diagnosis.
- Monetary policy.
- Insurance premiums and hazard events.

Bayes and the British code breakers

Alan Touring and Enigma Machine



- Near-infinite potential code translations.
- Solutions $\rightarrow$ previous encrypted messages.
- U-Boats $\rightarrow$ weather and shipping phrases.

Wrapping up week 9

Summary:

- Probability: tool to measure uncertainty in events.
- What is it good for?
- Conditional probability: importance of information.
- Independence of events.
- Bayesian reasoning.

Final project: the data report

Final research project

- Data report document.
- Why? How does it help?
- Describe variables, measures.
- Explore the data - add visuals.