Probability, 1

Frank Edwards

How often, on average, does an event occur?

Probability is a set of tools for describing randomness.

Probability helps us sort signal (patterns) from noise.

Two core theories

Frequentist: Probability is the proportion of times an event occurs if we repeat an experiment under the same conditions many times

Consider a simple experiment

What is the probability of obtaining a heads on a single fair coin flip?

```
flip_n_coins<-function(x){
  flip<-rbinom(x, 1, 0.5)
  flip<-ifelse(flip==1, "Heads", "Tails")
  return(flip)
}
```

Consider a simple experiment

What is the probability of obtaining a heads on a single fair coin flip?

```
flip_n_coins<-function(x){
  flip<-rbinom(x, 1, 0.5)
  flip<-ifelse(flip==1, "Heads", "Tails")
  return(flip)
}
```

```
flip_n_coins(1)
```

```
## [1] "Heads"
```

Consider a simple experiment

What is the probability of obtaining a heads on a single fair coin flip?

```
flip_n_coins<-function(x){
  flip<-rbinom(x, 1, 0.5)
  flip<-ifelse(flip==1, "Heads", "Tails")
  return(flip)
}
```

```
flip_n_coins(1)
```

```
## [1] "Heads"
```

```
flip_n_coins(1)
```

```
## [1] "Heads"
```

Consider a simple experiment

What is the probability of obtaining a heads on a single fair coin flip?

```
flip_n_coins<-function(x){
  flip<-rbinom(x, 1, 0.5)
  flip<-ifelse(flip==1, "Heads", "Tails")
  return(flip)
}
```

```
flip_n_coins(1)
```

```
## [1] "Heads"
```

```
flip_n_coins(1)
```

```
## [1] "Heads"
```

```
flip_n_coins(1)
```

```
## [1] "Heads"
```

Consider a simple experiment

What is the probability of obtaining a heads on a single fair coin flip?

```
flip_n_coins<-function(x){
  flip<-rbinom(x, 1, 0.5)
  flip<-ifelse(flip==1, "Heads", "Tails")
  return(flip)
}
```

```
flip_n_coins(1)
```

```
## [1] "Heads"
```

```
flip_n_coins(1)
```

```
## [1] "Heads"
```

```
flip_n_coins(1)
```

```
## [1] "Heads"
```

What is $\frac{\sum_1^n x_i}{n}$ where $x_i = 1$ when the coin is heads?

Probability as the average outcome under repeat experiments

```
sum((flip_n_coins(5)=="Heads")/5)
```

```
## [1] 0.2
```

Probability as the average outcome under repeat experiments

```
sum((flip_n_coins(5)=="Heads")/5)
```

```
## [1] 0.2
```

```
sum((flip_n_coins(20)=="Heads")/20)
```

```
## [1] 0.5
```

Probability as the average outcome under repeat experiments

```
sum((flip_n_coins(5)=="Heads")/5)
```

```
## [1] 0.2
```

```
sum((flip_n_coins(20)=="Heads")/20)
```

```
## [1] 0.5
```

```
sum((flip_n_coins(50)=="Heads")/50)
```

```
## [1] 0.6
```

Probability as the average outcome under repeat experiments

```
sum((flip_n_coins(5)=="Heads")/5)
```

```
## [1] 0.2
```

```
sum((flip_n_coins(20)=="Heads")/20)
```

```
## [1] 0.5
```

```
sum((flip_n_coins(50)=="Heads")/50)
```

```
## [1] 0.6
```

```
sum((flip_n_coins(1000)=="Heads")/1000)
```

```
## [1] 0.478
```

Probability as the average outcome under repeat experiments

```r
sum((flip_n_coins(5)=="Heads")/5)
```

```
## [1] 0.2
```

```r
sum((flip_n_coins(20)=="Heads")/20)
```

```
## [1] 0.5
```

```r
sum((flip_n_coins(50)=="Heads")/50)
```

```
## [1] 0.6
```

```r
sum((flip_n_coins(1000)=="Heads")/1000)
```

```
## [1] 0.478
```

```r
sum((flip_n_coins(100000)=="Heads")/100000)
```

```
## [1] 0.4981
```

Frequentist probability

If n_x is the number of heads, and n_t is the number of coin flips, then the probability of heads is

$$P(x) \approx \frac{n_x}{n_t}$$

$$P(x) = \lim_{n_t \to \infty} \frac{n_x}{n_t}$$

Bayesian: Probability is a subjective judgment about the likelihood that an event occurs, with endpoints at 0 (never occurs) and 1 (always occurs). Repeat experiments are often nonsensical.

Bayesian: Probability is a subjective judgment about the likelihood that an event occurs, with endpoints at 0 (never occurs) and 1 (always occurs). Repeat experiments are often nonsensical.

I have strong prior information that a fair coin will be heads half of the time, and tails half of the time. If I flip coins and see different patterns, I may change my beliefs about the likelihood of a heads.

Deterministic processes do not include randomness. For example, if I drop a ball, it will fall (because gravity).

Definitions

Deterministic processes do not include randomness. For example, if I drop a ball, it will fall (because gravity).

Stochastic events include randomness. For example, if I flip a fair coin, it will be heads half of the time.

Definitions

Deterministic processes do not include randomness. For example, if I drop a ball, it will fall (because gravity).

Stochastic events include randomness. For example, if I flip a fair coin, it will be heads half of the time.

Nearly all social processes are have random components, and can be treated as *stochastic*.

Definitions and axioms

- Experiment: an action that produces stochastic events

Definitions and axioms

- Experiment: an action that produces stochastic events
- Sample space (Ω): a set of all possible outcomes of the experiment

Definitions and axioms

- Experiment: an action that produces stochastic events
- Sample space (Ω): a set of all possible outcomes of the experiment
- Event: a subset of the sample space

Example: coin flips

- Experiment
    1. flip a coin
    2. roll a dice

Example: coin flips

- Experiment
    1. flip a coin
    2. roll a dice
- Sample space Ω
    1. {Heads, Tails}
    2. {1,2,3,4,5,6}

Example: coin flips

- Experiment
  1. flip a coin
  2. roll a dice

- Sample space Ω
  1. {Heads, Tails}
  2. {1,2,3,4,5,6}

- Event
  1. Heads, tails, not heads, heads or tails, heads and tails
  2. 3, even number, anything but 6

Probability with equal likelihood of events

If all outcomes are equally likely, and n represents the number of elements in a given set, then probability P of event A is:

$$P(A) = \frac{n_A}{n_\Omega}$$

Probability axioms

1. The probability of any event A is non-negative: $P(A) \geq 0$

Probability axioms

1. The probability of any event A is non-negative: $P(A) \geq 0$
2. If an experiment is conducted, the probability that one of the outcomes in the sample space occurs is 1: $P(\Omega) = 1$

Probability axioms

1. The probability of any event A is non-negative: $P(A) \geq 0$
2. If an experiment is conducted, the probability that one of the outcomes in the sample space occurs is 1: $P(\Omega) = 1$
3. Addition rule: If events A and B are mutually exclusive:

$$P(A \text{ or } B) = P(A) + P(B)$$

Probability that an event doesn't occur

$$1 - P(\text{not } A) = P(A)$$

Probability that an event doesn't occur

$$1 - P(\text{not } A) = P(A)$$

If $P(\text{rolling a 6}) = \frac{1}{6}$ then $P(\text{not rolling a 6}) = 1 - \frac{1}{6} = \frac{5}{6}$

Law of total probability

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

Law of total probability

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

Assume the sample space for A is [eats pizza, does not eat pizza], and the sample space for B is [happy, unhappy].

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

Assume the sample space for A is [eats pizza, does not eat pizza], and the sample space for B is [happy, unhappy].

If $P(\text{eats pizza}) = 0.5$ and $P(\text{eats pizza, happy}) = 0.4$, then $P(\text{eats pizza, unhappy}) = 0.1$

General addition rule

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

If $P(\text{happy}) = 0.5$, then $P(\text{eats pizza or happy}) = 0.6$

Three kinds of probability

The joint probability of two events (A and B) occurring is expressed as

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

Marginal probability

The marginal probability of an event B is

$$P(B)$$

The conditional probability of event A occurring given that event B occurred is the ratio of the joint probability of A and B divided by the marginal probability of B

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

Working with some real data

Voter files

```r
data("FLVoters")
voters<-na.omit(FLVoters)
head(voters)
```

```
##       surname county VTD age gender  race
## 1      PIEDRA    115  66  58      f white
## 2       LYNCH    115  13  51      m white
## 4     LATHROP    115  80  54      m white
## 5      HUMMEL    115   8  77      f white
## 6 CHRISTISON    115  55  49      m white
## 7       HOMAN    115  84  77      f white
```

Marginal probability

What is the probability that a randomly sampled voter in the population is Black: $P(Black) = ?$

```
voters %>%
  count(race, name = "voters") %>%
  mutate(p = voters/sum(voters))
```

```
##       race voters           p
## 1    asian    175 0.019203336
## 2    black   1194 0.131021617
## 3 hispanic   1192 0.130802151
## 4   native     29 0.003182267
## 5    other    310 0.034017338
## 6    white   6213 0.681773291
```

Is a woman: $P(Woman) = ?$

```
voters %>%
  count(gender) %>%
  mutate(n = n/sum(n))
```

```
##   gender         n
## 1      f 0.5358279
## 2      m 0.4641721
```

Joint probability

What is the probability that a voter is a Black woman:

$P(\text{Black and woman}) = ?$

```
voters %>%
  count(gender, race) %>%
  mutate(n = n/sum(n)) %>%
  pivot_wider(names_from = gender, values_from = n)
```

```
## # A tibble: 6 x 3
##   race           f       m
##   <chr>      <dbl>   <dbl>
## 1 asian    0.00911 0.0101
## 2 black    0.0744  0.0566
## 3 hispanic 0.0731  0.0577
## 4 native   0.00187 0.00132
## 5 other    0.0173  0.0167
## 6 white    0.360   0.322
```

What is the probability that a voter is a woman?

```
## # A tibble: 6 x 3
##   race           f       m
##   <chr>      <dbl>   <dbl>
## 1 asian    0.00911 0.0101
## 2 black    0.0744  0.0566
## 3 hispanic 0.0731  0.0577
## 4 native   0.00187 0.00132
## 5 other    0.0173  0.0167
## 6 white    0.360   0.322
```

Use the law of total probability:

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

put differently, for all categories of B i:

$$P(A) = \sum_{i=1}^{n} P(A \text{ and } B_i)$$

Conditional probability

If a voter is a man, what is the probability that he is Asian:
$P(\text{Asian}|\text{man}) = ?$

```
voters %>%
  filter(gender=="m") %>%
  count(race) %>%
  mutate(n=n/sum(n))
```

```
##       race           n
## 1    asian 0.021749409
## 2    black 0.121985816
## 3 hispanic 0.124349882
## 4   native 0.002836879
## 5    other 0.035933806
## 6    white 0.693144208
```

Conditional probability

Alternatively, we can use the definition of conditional probability as the ratio of the joint probability to the marginal probability:

$$P(\text{Asian}|\text{man}) = \frac{P(\text{Asian and man})}{P(\text{man})}$$

```
voters %>%
  count(gender, race) %>%
  mutate(n = n/sum(n))%>%
  pivot_wider(names_from = gender, values_from = n)
```

```
## # A tibble: 6 x 3
##   race           f       m
##   <chr>      <dbl>   <dbl>
## 1 asian    0.00911 0.0101
## 2 black    0.0744  0.0566
## 3 hispanic 0.0731  0.0577
## 4 native   0.00187 0.00132
## 5 other    0.0173  0.0167
## 6 white    0.360   0.322
```

Conditioning on more than one variable

What is the probability that a male voter over age 60 is white?

$$P(\text{white}|\text{male and over 60})$$

```
voters %>%
  mutate(over60=age>60) %>%
  count(over60, gender, race) %>%
  mutate(n=n/sum(n)) %>%
  pivot_wider(names_from = gender, values_from = n)
```

```
## # A tibble: 12 x 4
##    over60 race            f        m
##    <lgl>  <chr>       <dbl>    <dbl>
##  1 FALSE  asian    0.00691  0.00823
##  2 FALSE  black    0.0555   0.0435
##  3 FALSE  hispanic 0.0549   0.0436
##  4 FALSE  native   0.00121  0.000768
##  5 FALSE  other    0.0124   0.0129
##  6 FALSE  white    0.212    0.198
##  7 TRUE   asian    0.00219  0.00187
##  8 TRUE   black    0.0189   0.0132
##  9 TRUE   hispanic 0.0182   0.0142
## 10 TRUE   native   0.000658 0.000549
## 11 TRUE   other    0.00494  0.00373
## 12 TRUE   white    0.148    0.124
```

Conditioning on more than one variable

In general:

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

and

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

Two events are independent if knowledge of one event gives us no information about the other event.

$P(A|B) = P(A)$ and $P(B|A) = P(B)$

$$A \perp B$$

if and only if

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

Recall that a Bayesian perspective treats probability as a subjective opinion about how likely an event is. How should we change our beliefs after we make observations about the world?

Bayes' rule

Recall that a Bayesian perspective treats probability as a subjective opinion about how likely an event is. How should we change our beliefs after we make observations about the world?

Bayes' rule formalizes how we should update our beliefs based on evidence:

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

Bayes' rule

Recall that a Bayesian perspective treats probability as a subjective opinion about how likely an event is. How should we change our beliefs after we make observations about the world?

Bayes' rule formalizes how we should update our beliefs based on evidence:

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

Prior beliefs and evidence

If we have a *prior* belief that event A has $P(A)$ chance of occurring, then we observe some data, represented as event B, we update our beliefs and obtain a *posterior probability* $P(A|B)$.

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

Example: Detecting breast cancer

How good is a mammogram at detecting breast cancer?

What we know: One percent of women have breast cancer. 80 percent of people who have cancer and take a mammogram test positive. 9.6 percent of people who take a mammogram get a positive result when they do not have breast cancer.

Example: Detecting breast cancer

How good is a mammogram at detecting breast cancer?

What we know: One percent of women have breast cancer. 80 percent of people who have cancer and take a mammogram test positive. 9.6 percent of people who take a mammogram get a positive result when they do not have breast cancer.

If you take a mammogram and get a positive result, what is the probability that you have breast cancer?

Rewriting as probabilities

One percent of women have breast cancer

$$P(\text{Cancer}) = 0.01$$

Rewriting as probabilities

One percent of women have breast cancer

$$P(\text{Cancer}) = 0.01$$

80 percent of people who have cancer and take a mammogram test positive

$$P(\text{Test positive}|\text{Cancer}) = 0.8$$

Rewriting as probabilities

One percent of women have breast cancer

$$P(\text{Cancer}) = 0.01$$

80 percent of people who have cancer and take a mammogram test positive

$$P(\text{Test positive}|\text{Cancer}) = 0.8$$

9.6 percent of people who take a mammogram get a positive result when they do not have breast cancer

$$P(\text{Test positive}|\text{No cancer}) = 0.096$$

Using Bayes' rule

The prior probability of having cancer is 0.01. How should we update our belief that someone has cancer based on a positive test?

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

Using the law of total probability, we can rewrite the denominator as:

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

Using Bayes' rule

We can apply Bayes' rule for A = Cancer, B = positive test:

$$P(\text{Cancer}|\text{Test positive}) =$$

$$\frac{P(\text{Test positive}|\text{Cancer})P(\text{Cancer})}{P(\text{Test positive})}$$

$$P(\text{Cancer}|\text{Test positive}) = \frac{0.8 \times 0.01}{0.8 \times 0.01 + 0.096 \times 0.99}$$

```
(0.8 *  0.01)/(0.8 *  0.01 + 0.096 * 0.99)
```

```
## [1] 0.07763975
```

The probability that someone has cancer given a prior probability of one percent and a positive test is about 0.078