Lecture 3 - Axioms of Probability

Sta102 / BME102

Colin Rundel

September 1, 2014

Mean, Median and Mode

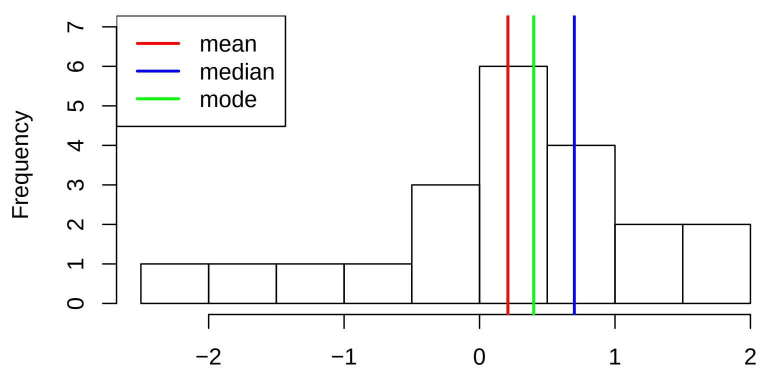
We start with a set of 21 numbers,

```
##  [1] -2.2 -1.6 -1.0 -0.5 -0.4 -0.3 -0.2  0.1  0.1  0.2  0.4
## [12]  0.4  0.5  0.7  0.7  0.7  0.9  1.2  1.2  1.7  1.8
```

and calculate the mean, median, and mode:

```
mean(x)

## [1] 0.2095

median(x)

## [1] 0.4

Mode(x)

## [1] 0.7
```


Graphically



Where do they come from?

Imagine we didn't know about the mean, median, or mode - how should we choose a single number s that best summarizes a set of numbers (data)?

There are a couple of different ways we could think about doing this by defining different discrepancy functions

$$L_0 = \sum_i |x_i - s|^0 \qquad \text{assume, } n^0 = \begin{cases} 0 & \text{if n=0} \\ 1 & \text{otherwise} \end{cases}$$

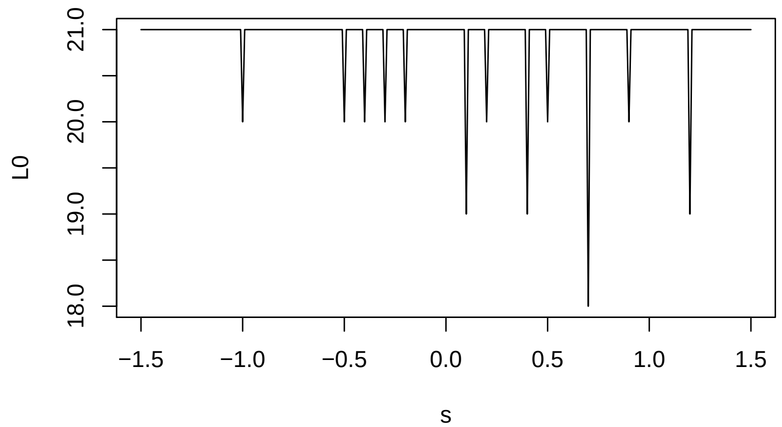
$$L_1 = \sum_i |x_i - s|^1$$

$$L_2 = \sum_i |x_i - s|^2$$

we want to find the values of s that minimizes L_0, L_1, L_2 for any given data set x.

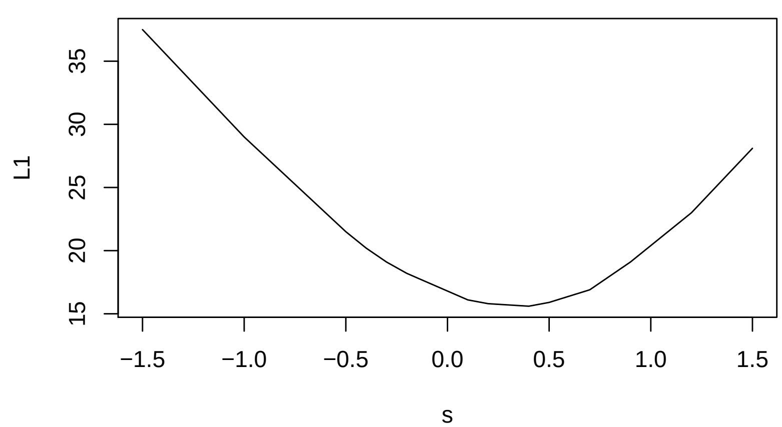
Minimizing L_0

$$L_0 = \sum_i |x_i - s|^0$$



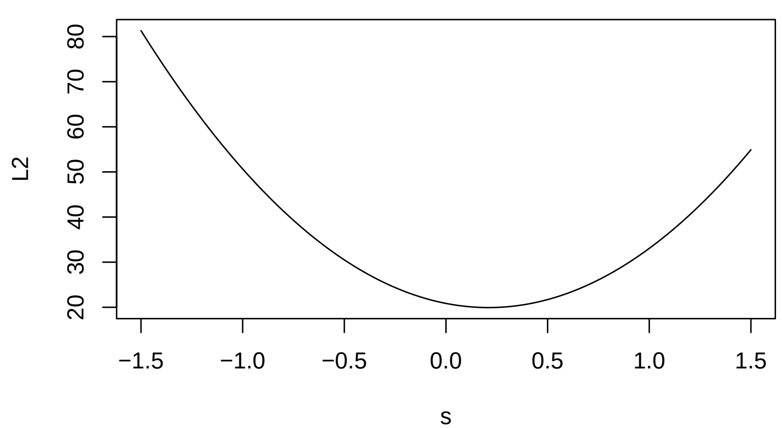
Minimizing L_1

$$L_1 = \sum_i |x_i - s|^1$$



Minimizing L_2

$$L_2 = \sum_i |x_i - s|^2$$



What have we learned?

L_0, L_1, and L_2 are examples of what we call loss functions. These come up all the time in higher level statistics.

What we have just seen is that:

L_0 is minimized when s is the mode.

L_1 is minimized when s is the median.

L_2 is minimized when s is the mean.

What does it mean to say that:

- The probability of rolling snake eyes is $P(S) = 1/36$?
- The probability of flipping a coin and getting heads is $P(H) = 1/2$?
- The probability Apple's stock price goes up today is $P(+) = 3/4$?

Interpretations:

- Symmetry: If there are k equally-likely outcomes, each has
$$P(E) = 1/k$$
- Frequency: If you can repeat an experiment indefinitely,
$$P(E) = \lim_{n\to\infty} \frac{\#E}{n}$$
- Belief: If you are indifferent between winning \$1 if E occurs or winning \$1 if you draw a blue chip from a box with $100 \times p$ blue chips, rest red,
$$P(E) = p$$

Terminology

Outcome space (Ω) - set of all possible outcomes (ω).

Examples:	3 coin tosses	{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}
	One die roll	{1,2,3,4,5,6}
	Sum of two rolls	{2,3,...,11,12}
	Seconds waiting for bus	$[0, \infty)$

Event (E) - subset of Ω ($E \subseteq \Omega$) that might happen, or might not

Examples:	2 heads	{HHT, HTH, THH}
	Roll an even number	{2,4,6}
	Wait < 2 minutes	$[0, 120)$

Random Variable (X) - a value that depends somehow on chance

Examples:	# of heads	{3, 2, 2, 1, 2, 1, 1, 0}
	# flips until heads	{3, 2, 1, 1, 0, 0, 0, 0}
	2^die	{2, 4, 8, 16, 32, 64}

Set Operations and Definitions

Intersection	E and F, EF, $E \cap F$
Union	E or F, $E \cup F$
Complement	not E, E^c
Disjoint	$E \cap F = \emptyset$
Difference	$E \backslash F = E$ and F^c
Symmetric Difference	$E \triangle F = (E$ and $F^c)$ or $(E^c$ and $F)$

Rules of Probability (Kolmogorov axioms)

1. $P(E) \geq 0$
2. $P(\Omega) = P(\omega_1 \text{ or } \omega_2 \text{ or } \cdots \text{ or } \omega_n) = 1$
3. $P(E \text{ or } F) = P(E) + P(F)$
   if E and F are disjoint, i.e. $P(E \text{ and } F) = 0$

Useful Identities

Complement Rule:

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

Difference Rule:

$$P(B \text{ and } A^c) = P(B) - P(A) \text{ if } A \subseteq B$$

Inclusion-Exclusion:

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

Useful Identities (cont)

Commutativity & Associativity:

$$A \text{ or } B = B \text{ or } A \qquad A \text{ and } B = B \text{ and } A$$
$$(A \text{ or } B) \text{ or } C = A \text{ or } (B \text{ or } C) \qquad (A \text{ and } B) \text{ and } C = A \text{ and } (B \text{ and } C)$$

$$(A \text{ or } B) \text{ and } C = (A \text{ and } C) \text{ or } (B \text{ and } C)$$

*Think of union as addition and intersection as multiplication: $(A + B) \times C = AC + BC$

DeMorgan's Rules:

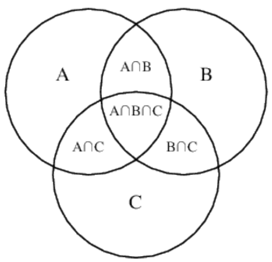
$$\text{not } (A \text{ and } B) = (\text{not } A) \text{ or } (\text{not } B)$$
$$\text{not } (A \text{ or } B) = (\text{not } A) \text{ and } (\text{not } B)$$

Generalized Inclusion-Exclusion

$$P(\bigcup_{i=1}^{n} E_i) = \sum_{i \le n} P(E_i) - \sum_{i<j\le n} P(E_i E_j) + \sum_{i<j<k\le n} P(E_i E_j E_k) - \ldots + (-1)^{n+1} P(E_1 \ldots E_n)$$

For the case of $n = 3$:

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$



Equally Likely Outcomes

$$P(E) = \frac{\#(E)}{\#(\Omega)} = \frac{1}{\#(\Omega)} \sum_i 1_{\omega_i \in E}$$

Notation:

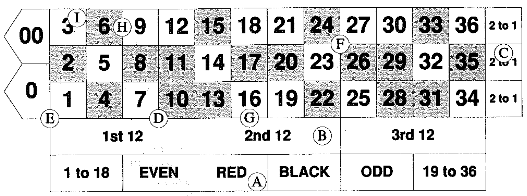
Cardinality - $\#(S)$ = number of elements in set S

Indicator function - $1_{x \in S} = \begin{cases} 1 & \text{if } x \in S \\ 0 & \text{if } x \notin S \end{cases}$

Probability of rolling an even number with a six sided die?

Roulette

FIGURE 1. Layout of a Nevada roulette table. Key to colors: 0 and 00 = Green, *unshaded numbers* = Red, *shaded numbers* = Black.



Play	Set of winning numbers	Payoff odds
A. Even money play	Group of 18 numbers as marked in the box	1 to 1
B. Dozen play	12 numbers marked in the box	2 to 1
C. Column play	12 numbers in column (shown here as a row)	2 to 1
D. Line play	Six numbers above	5 to 1
E. House special	0, 00, 1, 2, 3	6 to 1
F. Quarter play	Four numbers in square	8 to 1
G. Street play	Three numbers above	11 to 1
H. Split play	Two adjoining numbers	17 to 1
I. Straight play	Single number	35 to 1

Conditional Probability

This is the probability an event will occur when another event is known to have already occurred.

With equally likely outcomes we define the probability of A *given* B as

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

(the proportion of outcomes in B that are also in A)

Conditional Probability, cont.

We can rewrite the counting definition of conditional probability as

$$\begin{aligned} P(A|B) &= \frac{\#(A \cap B)}{\#(B)} \\ &= \frac{\#(A \cap B)/\#(\Omega)}{\#(B)/\#(\Omega)} \\ &= \frac{P(A \cap B)}{P(B)} \end{aligned}$$

which is the general definition of conditional probability.

Note that $P(A|B)$ is undefined if $P(B) = 0$.

Useful Rules

Very often we may know the probability of events and their conditional probabilities but not probabilities of the events together, in which case we can use

Multiplication rule:

$$P(A \cap B) = P(A|B)P(B)$$

Other cases where we do not have the probability of one of the events, we can use

Rule of total probability:

For a partition $B_1, \ldots, B_n$ of Ω, with $B_i \cap B_j = \emptyset$ for all $i \neq j$.

$$\begin{aligned} P(A) &= P(A \cap B_1) + P(A \cap B_2) + \ldots + P(A \cap B_n) \\ &= P(A|B_1)P(B_1) + \ldots + P(A|B_n)P(B_n) \end{aligned}$$

Example - Hiking

A quick example of the application of the rule of total probability:

Whether or not I go hiking depends on the weather, if it is sunny there is a 60% chance I will go for a hike, while there is only a 10% chance if it is raining and a 30% chance if it is snowing.

The weather forecast for tomorrow calls for 50% chance of sunshine, 40% chance of rain, and a 10% chance of snow.

What is the probability I go for a hike tomorrow?

Independence

We defined events A and B to be independent when

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

which also implies that

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

This should *not* to be confused with disjoint (mutually exclusive) events where

$$P(A \cap B) = 0$$

Example - Eye and hair color

Table 3.3.1 Hair color and eye color

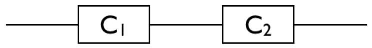
		Hair color			
		Brown	Black	Red	Total
Eye color	Brown	400	300	20	720
	Blue	800	200	50	1,050
	Total	1,200	500	70	1,770

1. Are brown and black hair disjoint?
2. Are brown and black hair independent?
3. Are brown eyes and red hair disjoint?
4. Are brown eyes and red hair independent?

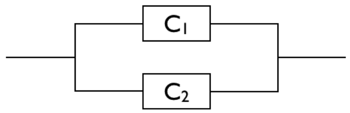
Example - Circuit Reliability

If the probability that C_1 will fail in the next week is 0.2, the probability C_2 will fail is 0.4, and component failure is independent which circuit configuration is more reliable? (has greater probability of being functional next week)

Series:



Parallel:



Bayes' Rule

Expands on the definition of conditional probability to give a relationship between $P(B|A)$ and $P(A|B)$

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

In the case where $P(A)$ is not known we can extend this using the law of total probability

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

Example - *House*

If you've ever watched the TV show *House* on Fox, you know that Dr. House regularly states, "It's never lupus."

Lupus is a medical phenomenon where antibodies that are supposed to attack foreign cells to prevent infections instead see plasma proteins as foreign bodies, leading to a high risk of blood clotting. It is believed that 2% of the population suffer from this disease.

The test for lupus is very accurate if the person actually has lupus, however is very inaccurate if the person does not. More specifically, the test is 98% accurate if a person actually has the disease. The test is 74% accurate if a person does not have the disease.

Is Dr. House correct even if someone tests positive for Lupus?