Basic Probability

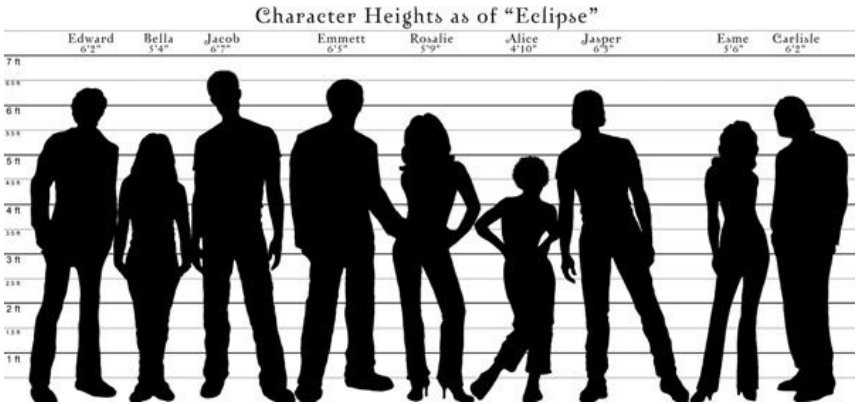
By Tejumade Afonja

AI6Lagos Cohort7
19th Sept. 2021

Probability of Large Straight in Yahtzee

Variables

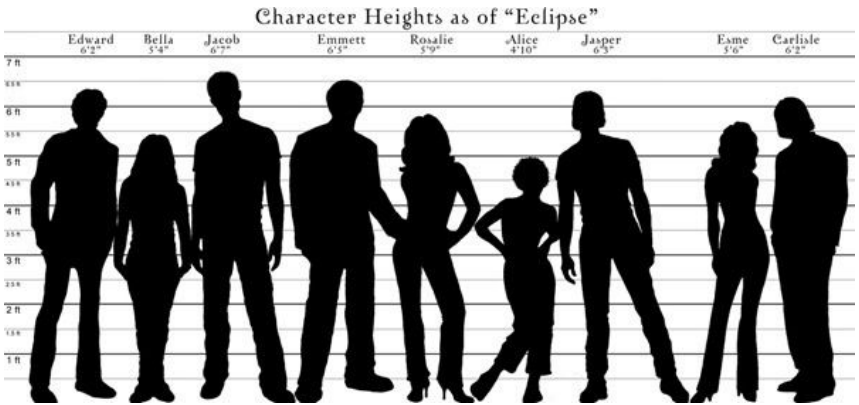
Something whose value is not *yet* known



Twilight Character Heights

Variables

A property or descriptor that can take on multiple values.



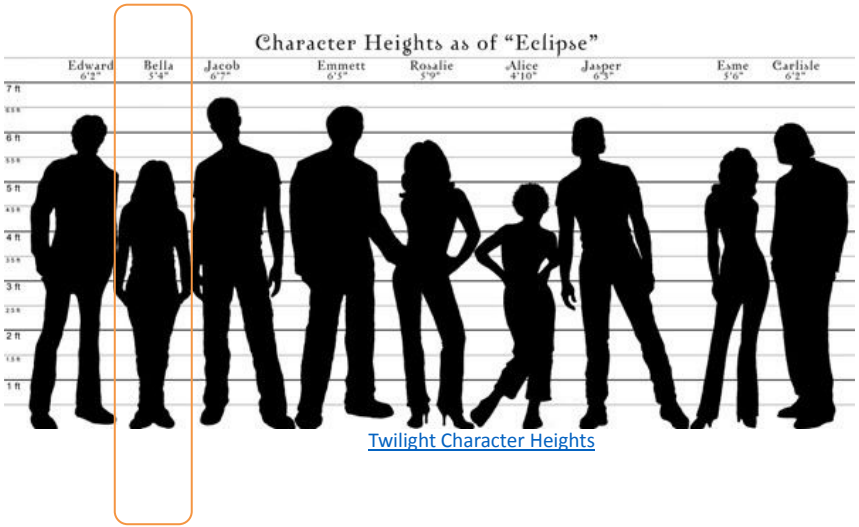
Twilight Character Heights

Name	Height	Gender	Vampire
Edward	6' 2"	Male	Yes
Bella	5'4"	Female	No
Jacob	6'7"	Male	No
Emmett	6'5"	Male	Yes
Rosalie	5'9"	Female	Yes
Alice	4'10"	Female	Yes
Jasper	6'5"	Male	Yes
Esme	5'6"	Female	Yes
Carlisle	6'2"	Male	Yes

Variables

A question, to which the value is an answer. For example,

How tall is Bella?



Twilight Character Heights

Name	Height	Gender	Vampire
Edward	6' 2"	Male	Yes
Bella	5'4"	Female	No
Jacob	6'7"	Male	No
Emmett	6'5"	Male	Yes
Rosalie	5'9"	Female	Yes
Alice	4'10"	Female	Yes
Jasper	6'5"	Male	Yes
Esme	5'6"	Female	Yes
Carlisle	6'2"	Male	Yes

Variables

The probability that Variable X takes value x

➔ P(X=x)

The probability that Variable “Height” takes value 6’2”

➔ P(Height=6’2”)

Name	Height	Gender	Vampire
Edward	6’ 2”	Male	Yes
Bella	5’4”	Female	No
Jacob	6’7”	Male	No
Emmett	6’5”	Male	Yes
Rosalie	5’9”	Female	Yes
Alice	4’10”	Female	Yes
Jasper	6’5”	Male	Yes
Esme	5’6”	Female	Yes
Carlisle	6’2”	Male	Yes

Variables

The probability that Variable X takes value x

➔ P(X=x)

The probability that Variable "Height" takes value 6'2"

➔ P(Height=6'2")

↝ = 2/9

Name	Height	Gender	Vampire
Edward	6' 2"	Male	Yes
Carlisle	6'2"	Male	Yes

Variables

How about more than one variables?

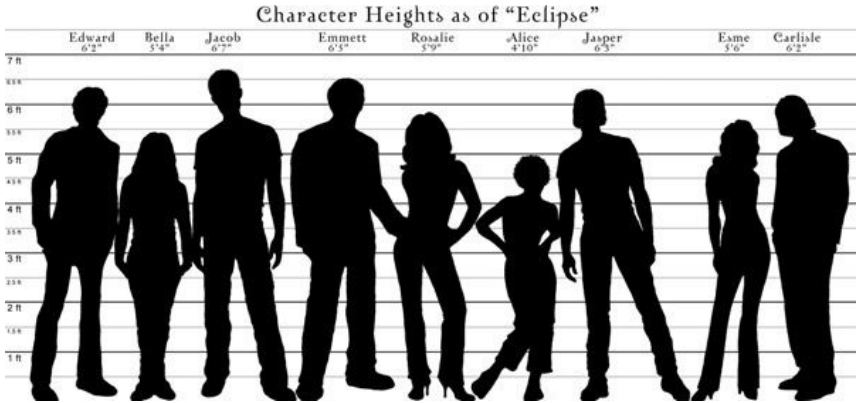
- ➔ What is the probability that an individual randomly selected from the table is Female and a Vampire?
  - P(Gender=Female, Vampire=Yes)
  - P(X=x, Y=y)
    - ⇝ $= 3/9$

Name	Height	Gender	Vampire
Rosalie	5’9”	Female	Yes
Alice	4’10”	Female	Yes
Esme	5’6”	Female	Yes

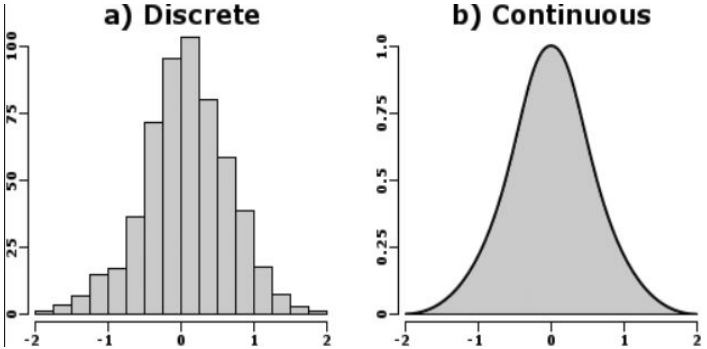
Variables

Something

- whose value is not *yet* known
- that can take on multiple values
- that ask questions
- that can take on an exact value or infinitely many values



Twilight Character Heights



Quora: The probability distribution of discrete and continuous random variables can be defined in terms of what

Random Variables

TeachResearchInnovate

Events

Assignment of values or set of values to a variable.

A declarative statement that can be true or false.

Examples

- ➔ Edward is a vampire
- ➔ Height is less than 6ft
- ➔ Name start with E
- ➔ etc

Name	Height	Gender	Vampire
Edward	6’ 2”	Male	Yes
Bella	5’4”	Female	No
Jacob	6’7”	Male	No
Emmett	6’5”	Male	Yes
Rosalie	5’9”	Female	Yes
Alice	4’10”	Female	Yes
Jasper	6’5”	Male	Yes
Esme	5’6”	Female	Yes
Carlisle	6’2”	Male	Yes

Conditional Probability

The probability that some event A occurs, giving that we know some other event B has occurred

Example

➔ What is the probability of “Gender” = “Female” given “Vampire”=”No”?

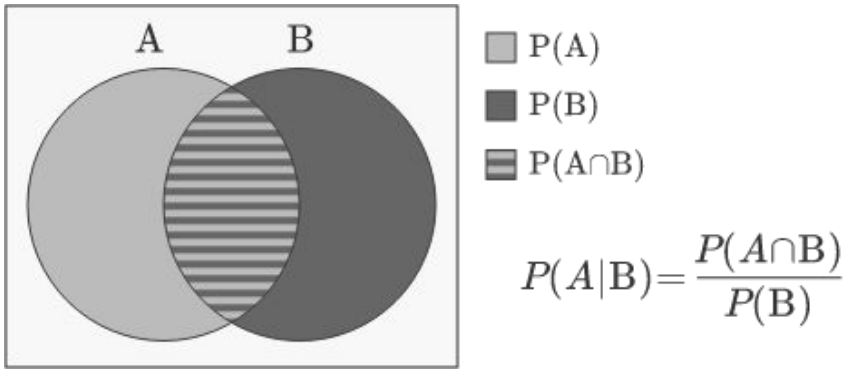
- P(A=a|B=b)
- P(Gender=Female | Vampire=No)

Name	Height	Gender	Vampire
Edward	6’ 2”	Male	Yes
Bella	5’4”	Female	No
Jacob	6’7”	Male	No
Emmett	6’5”	Male	Yes
Rosalie	5’9”	Female	Yes
Alice	4’10”	Female	Yes
Jasper	6’5”	Male	Yes
Esme	5’6”	Female	Yes
Carlisle	6’2”	Male	Yes

Conditional Probability

- P(Y=y) ⇔ P("Vampire"="No")
- P(X=x | Y=y) ⇔ P("Gender"="Female" | "Vampire"="No")
  - = 1/2

Name	Height	Gender	Vampire
Bella	5'4"	Female	No
Jacob	6'7"	Male	No



Conditional Probability

- P(No) = 2/9
- P(Female ∩ No) = 1/9
- P(Female|No) = P(Female ∩ No) / P(No)
- P(Female | No) = 1/9 ÷ 2/9
- P(Female | No) = 1/2

Independence

It might happen that the probability of one event remains unaltered with the observation of another.

Example

- P(Name=Edward | Vampire=Yes)?
- = P(Edward ∩ Yes) / P(Yes)
- P(Edward ∩ Yes) = 1/9
- P(Yes) = 7/9
- P(Edward | Yes) = 1/7
- But P(Edward) = 1/9
- P(Edward|Yes) != P(Edward)

A and B are independent if:
P(A|B) = P(A)
P(B|A) = P(B)
$P(A \cap B) = P(A)*P(B)$

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

Additional information about whether or not Edward is a Vampire does change the probability of Edward.

Thus, the two events are <u>not independent</u>

Name	Height	Gender	Vampire
Edward	6’ 2”	Male	Yes
Bella	5’4”	Female	No
Jacob	6’7”	Male	No
Emmett	6’5”	Male	Yes
Rosalie	5’9”	Female	Yes
Alice	4’10”	Female	Yes
Jasper	6’5”	Male	Yes
Esme	5’6”	Female	Yes
Carlisle	6’2”	Male	Yes

Study Question

Income and University [Khan Academy]

Researchers surveyed recent graduates of two different universities about their annual incomes. Table 1 display the data for the 300 graduates who responded to the survey.

Are the events "income under $20,000" and "attended University B" *independent*?

Annual Income	**University A**	**University B**	**Total**
Under $20,000	36	24	**60**
$20,000 to 39,999	109	56	**165**
$40,000 and over	35	40	**75**
Total	**180**	**120**	**300**

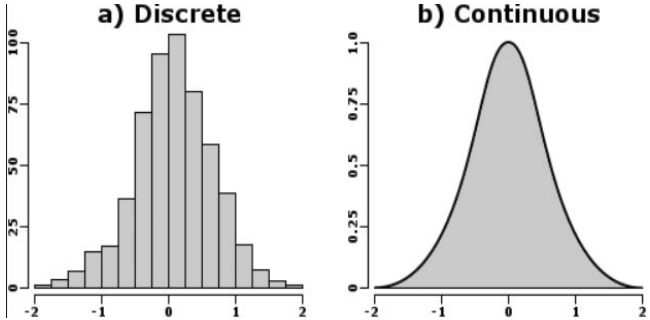
Table 1

Probability Distribution

A probability distribution for a variable X is the set of probabilities assigned to each possible value of X.

Question

➔ What type of distribution is the Gender and Age Variables?



Quora: The probability distribution of discrete and continuous random variables can be defined in terms of what

Name	Height	Gender	Vampire	Age
Edward	6’ 2”	Male	Yes	110
Bella	5’4”	Female	No	18
Jacob	6’7”	Male	No	16
Emmett	6’5”	Male	Yes	92
Rosalie	5’9”	Female	Yes	92
Alice	4’10”	Female	Yes	106
Jasper	6’5”	Male	Yes	164
Esme	5’6”	Female	Yes	76
Carlisle	6’2”	Male	Yes	300

Probability Distribution

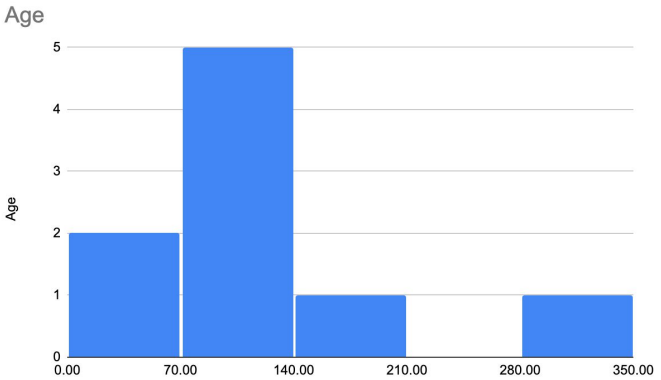
A probability distribution for a variable X is the set of probabilities assigned to each possible value of X.

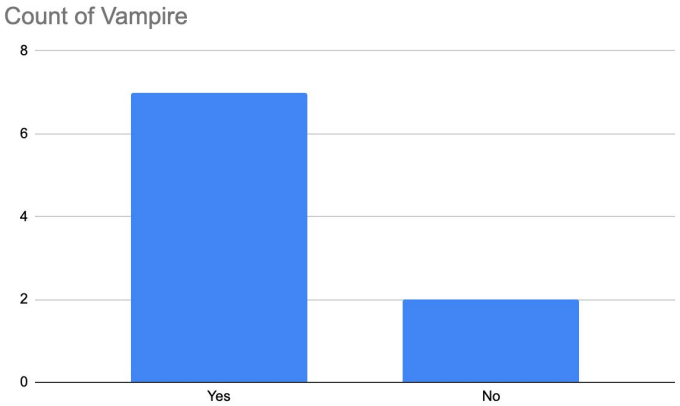
Question

➔ What type of distribution is the Vampire and Age Variables?

This is **not** a probability distribution.

Why?





Using Bayes Rule

When using Bayes' rule, we sometimes loosely refer to event A as the "hypothesis" and event B as the "evidence".

In many cases, we know or can easily determine P(B|A) (the probability that a piece of evidence will occur, given that our hypothesis is correct) but it's much harder to figure out P(A|B) (the probability of the hypothesis being correct, given that we obtain a piece of evidence) [1].

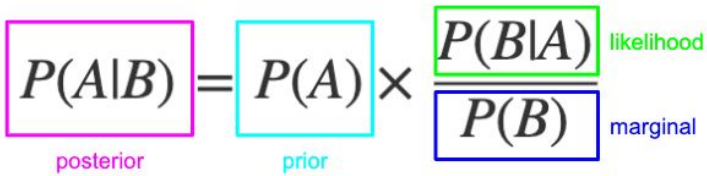
$P(A|B) = P(A \cap B) / P(B)$

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

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

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

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



Bayes' Rule - Explained for Beginners

Study Question

Doctors and Flu [datasciencecourse.org]

I want to know if I have **come** with a rare strain of flu (occurring in only 1/10,000). There is an "accurate" test for the flu (if I have the flu, it will tell me I have it 99% of the time, and if I do not have it, it will tell me I do not have it 99% of the time). I go to the doctor and test positive. What is the probability that I have this flu?

What is P(Flu=1 | Test=1)?

We know

- P(Flu=1)
- P(Test=1 | Flu=1)
- P(Test=0 | Flu=0)

→ **= 0.0098**

Expected Values

In statistics, one often deals with data set and probability distributions that are too large to effectively examine each possible combination of values.

We use statistical measures to represent, with some loss of information, meaningful features of the distribution [1]

$$E[X] = \sum_x x * p(x)$$

Dice = {1, 2, 3, 4, 5, 6}

P(Dice=1) = ⅙

E[X] = sum(x * p(x))

E[Dice] = (1 * ⅙) + (2 * ⅙) + (3*⅙)+ (4*⅙)+ (5*⅙)+ (6*⅙)

E[Dice] = 3.5

Expected Values

In statistics, one often deals with data set and probability distributions that are too large to effectively examine each possible combination of values.

We use statistical measures to represent, with some loss of information, meaningful features of the distribution [1]

$$E[X] = \sum_x x * p(x)$$

Name	Height	Gender	Gender_male	Gender_female	Vampire
Edward	6’ 2”	Male	1	0	Yes
Bella	5’4”	Female	0	1	No
Jacob	6’7”	Male	1	0	No
Emmett	6’5”	Male	1	0	Yes
Rosalie	5’9”	Female	0	1	Yes
Alice	4’10”	Female	0	1	Yes
Jasper	6’5”	Male	1	0	Yes
Esme	5’6”	Female	0	1	Yes
Carlisle	6’2”	Male	1	0	Yes
E[X]			**2.78**	**1.78**	

Study Question

Certificate Eligibility [AI6Lagos]

Researchers surveyed scholars of AI Saturdays Lagos about their eligibility to earn a certificate. Table 2 display the data for the 6 scholars who are currently enrolled in the AI6Lagos Cohort2. The organizers informed the researchers of some statistics, like the total attendance expected, total task, As well as some conditional probabilities as displayed in table 3 and 4.

Student	Gender	Attendance	# Task Completed	Project
Dara	M	2	5	Yes
Musa	M	11	7	No
Busola	F	15	5	Yes
Nenye	M	5	1	Yes
Odun	F	16	6	No
Morenike	F	10	2	Yes

Table 2

Total Attendance	16
Total Task	7

Table 3

P(Certificate | Attendance)	0.5
P(Certificate | Task)	0.45
P(Certificate | Project)	0.05

Table 4

What is the probability that Dara will receive a certificate? P(Dara, Certificate)

Will Dara earn a certificate?

Note: Probability of a person has to be > 80% to earn a certificate

References

[1] Causal Inference in Statistics: A primer (Chapter 1)

[2] Datasciencecourse.org