Probability

4 min

Catherine is going to visit the Inference and Machine Learning Teams this week. She knows that both of these specialists work with Probability, and wants to brush up on her skills before she goes, so she's going to explore a famous problem – about birthdays.

Calculating the probability of an event is sometimes dependent on external factors. For instance, in the birthday problem "What is the probability that two people in a room have the same birthday?" the probability is dependent on the number of people in the room.

Other times, the probability of something is constant. For instance, the probability of flipping a coin and it landing heads will always be 50%.

Instructions

1. Checkpoint 1 Passed

1.

In data science, probability is often used to simulate scenarios.

The code on the right simulates the birthday problem. Right now the code simulates a room with only 2 people that get random birthdays, and the probability that those 2 people have the same birthday is really low.

Change the number 2 to a higher number of your choosing where it says #Change This Number and run the code.

Is there a match in the simulation? What's the probability that there would be a match?

Keep changing the number to test out different simulations.

Note that if you make your number too big, the program will throw an error due to the way we have implemented some of the math. This is a great example of needing to be mindful of possible inputs to your program!

Did you change the number in line 3?

script.py

```
# We have hidden code in another file. If you're curious, open the folder to the left and inspect the simulate.py file


from simulate import simulate


num_people_in_room = 30     #Change This Number (keep it smaller than 100 to save processing power)


simulate(num_people_in_room)
```

simulate.py

```
import random

#Simulate a room with a certain number of people

def simulate(num_people):

  birthdays = []

  print("Here's what our room looks like:\n")

  months = ["January", "February", "March", "April", "May", "June", "July", "August", "September", "October", "November", "December"]

  #Assign a random birthday to each person

  for i in range(0, num_people):

    #Choose a random month

    month_choice = random.choice(months)

    #Choose a random day based on month

    if month_choice == "February":

      day = random.randint(1, 29)

    elif month_choice == "April" or month_choice == "June" or month_choice == "September" or month_choice == "November":

      day = random.randint(1, 30)

    else:

      day = random.randint(1, 31)

    birthday = month_choice + " " + str(day)

    #Store the birthday

    birthdays.append(birthday)

    print("Person {0}'s birthday: {1}".format(i + 1, birthday))

  calculate_probability(num_people)

  match = False

  #Check for matching birthdays

  for i in range(len(birthdays)):

    if find_duplicates(birthdays, birthdays[i], i):

      match = True

      break

  if not match:

    print("\n\nIn our simulation, no two people have the same birthday")
```

```
#Calculate the probability of there being 2 people with the same birthday
def calculate_probability(num_people):
  #Check there is at least 2 people in the room
  if num_people < 2:
    print("\n\nNot enough people in the room!")
    return
  else:
    #Calculate the probability
    numerator = 365
    countdown = 364
    for i in range(2, num_people + 1):
      numerator = numerator * countdown
      countdown -= 1
    denominator = 365 ** num_people
    probability = 1 - numerator/float(denominator)
    #Change probability to percentage
    rounded = round(probability*100, 2)
    print("\n\nThe probability that two people in a room of {0} people have the same birthday is nearly
{1}%".format(num_people, rounded))



#Find the same birthday within our list of birthdays
def find_duplicates(birthdays_list, birthday, index):
  people = []
  for i in range(len(birthdays_list)):
    if birthdays_list[i] == birthday and i != index:
      people.append(i + 1)
  if people:
    people.append(index + 1)
    print("\n\nIn our simulation, the following people have the same birthdays: ")
    for person in people:
      print("Person {0}".format(person))
```

```
    return True
else:
    return False
```