Introduction a first course in Programming to Data Analysis

Using python.

Iván Andrés Trujilllo Abella

Facultad de Ingenieria
Pontificia Universidad Javeriana

trujilloiv@javeriana.edu.co

Contingency table

		Diagnose	
		Disease	No-Disease
Risk Factor	Smoke	a	b
	Not Smoke	c	d

What it is $P(Disase \mid Smoke) = \frac{a}{a+b}$ note that marginal distribution. Note that $P(Disease \cap Smoke) = \frac{a}{(a+b+c+d)}$. Also note that $P(Smoke) = \frac{a+b}{(a+b+c+d)}$. Note the result of divide the last two probabilities.

Conditional Probability

The probability of event given a "information". *Probability of A occur given B occurs.*

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \tag{1}$$

Note that $P(A \cap B)$ it is equal to $P(B \cap A)$.

$$P(A \mid B)P(B) = P(B \mid A)P(A) \tag{2}$$

Pontificia Universidad JAVERIANA Bogotá

Bayes theorem

Notice that not is same: *The probability that occur A given B, that occur B given A.* However we can compute one of the another.

$$P(A \mid B) = \frac{P(A)P(B \mid A)}{P(B)} \tag{3}$$

Bowls problems

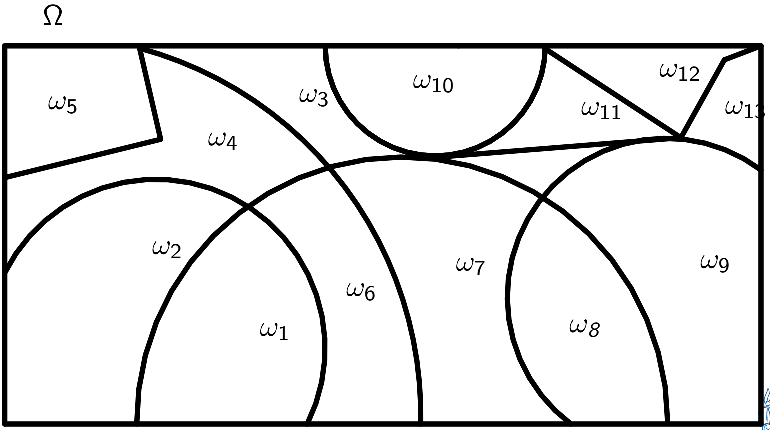
Derive the following problem.

Law of total probability

The sample space defined as Ω if we split omega in ω k subsets in order that each *subset* no overlap with others. $\bigcup_{i=1}^{k} \omega_i = \Omega$ y $\bigcap_{i=1}^{k} \omega_i = \emptyset$ For instance the sample space defined as $\Omega = \{a, b, c, d, e, f\}$ $\omega_1 = \{a, f\}$ $\omega_2 = \{b, c, d\}$ $\omega_3 = \{e\}$.

Pontificia Universidad JAVERIANA Bogotá

Split Ω



$P(A)$

total law probability

We need remember by set theory that a event A could be rewrite as $A = (A \cap B) \cup (A \cap C)$. if $(B \cup C) = \Omega$ for this case we can rewrite

$$\begin{aligned} A &= (A \cap \omega_1) \cup (A \cap \omega_2) ... (A \cap \omega_k) \\ P(A) &= P(A \cap \omega_1) + ... + P(A \cap \omega_k) \\ P(A) &= P(A \mid \omega_1) P(\omega_1) + P(A \mid \omega_2) P(\omega_2) + ... + P(A \mid \omega_k) P(\omega_k) \end{aligned} \tag{4}$$

note that by total law $P(A)$

Monty hall

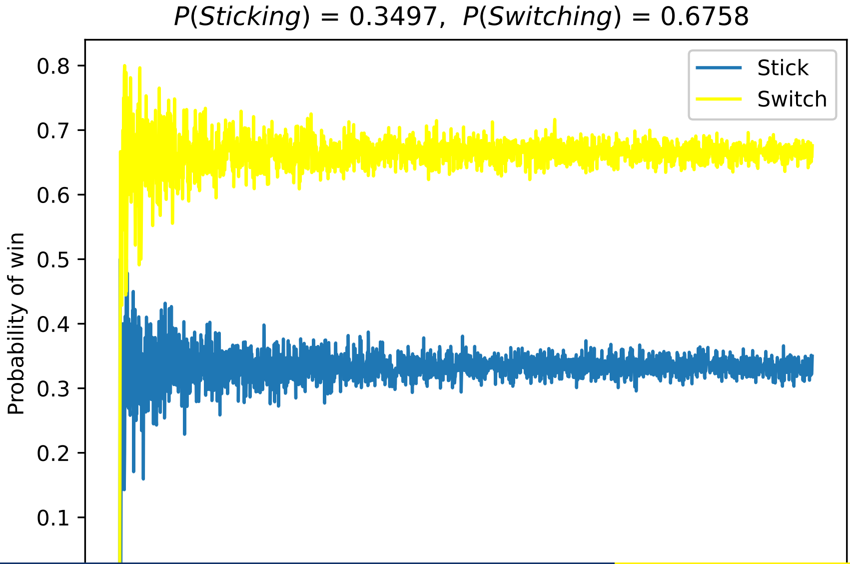
Reach the famous in 1990

There are three closed doors, and you must select one to win a car, behind of the only one, there is a car, and behind the other two there are goats. and after select the door, another door with a goat it is showed, you must remain in the selected door or switch?
initial probality it is the $1/3$.

Monty Hall simulation

This so not intuitive that generated controversy in the academic community.

The following Figure 16 show the relation among the probability of two coincidences among two person in the sample.



Monty Hall problem

Python code

```
def monty_game():
  doors=[1,2,3]
  doors_variable = doors.copy()
  winner = random.randint(1,3)
  select_one = random.randint(1,3)
  values = [winner,select_one]
  switch = list( set(doors) - set(values))
  select_two = random.randint(switch[0], switch[-1])
  s1,s2 = 0,0
  if winner == select_one:
    s1 += 1
  else:
    s2 +=1
  return [s1,s2]
```

Monty Hall problem

python code

```
games=1000
s1_total = []
s2_total = []
for trials in range(1,games):
  values_s1 = sum([monty_game()[0] for _ in range(trials)]) /
      trials
  s1_total.append(values_s1)
  values_s2 = sum(monty_game()[1] for _ in range(trials)) /
      trials
  s2_total.append(values_s2)
plt.plot(np.arange(1,games),s1_total)
plt.plot(np.arange(1,games),s2_total)
print(s1_total[-1], s2_total[-1])
```

Monty Hall

Bayesian solution

The event D_i the i winner door and M_j monty open j door, for $i, j = 1, 2, 3$.

$$P(D_i \mid M_j) = \frac{P(M_j \mid D_i)P(D_i)}{P(M_j)} \tag{5}$$

you select the first door and monty the second, therefore the question is $P(D_3 \mid M_2)$. Notice that $P(M_2) = \sum_{i=1}^{3} P(M_2 \mid D_i)$. given the rules of games, $P(M_2 \mid D_2) = 0$, and $P(M_2 \mid D_3) = 1$, if monty could select random in two choices $P(M_2 \mid C_1) = 1/2$ and finally, switch strategy have a probability of $2/3$.

Pontificia Universidad JAVERIANA Bogotá

Buffon Needled

Here we have a column

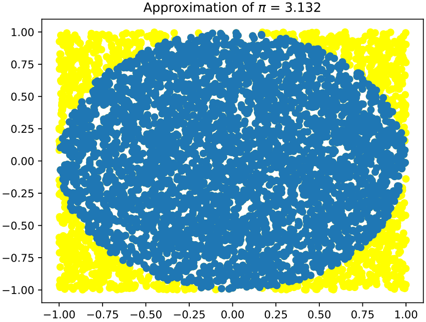
we have another column

π Number

```
import numpy as np
import matplotlib.pyplot as plt
dots = 5000
c1,c2=-1,1
x = np.random.uniform(c1,c2,
    size=dots)
y = np.random.uniform(c1,c2,
    size=dots)
coordenates_circle =
    (x**2)+(y**2) < 1
circle_y=y[coordenates_circle]
circle_x=x[coordenates_circle]
pi = 4*sum(coordenates_circle)
    / dots
plt.scatter(x,y, color='yellow')
plt.scatter(circle_x,circle_y)
```

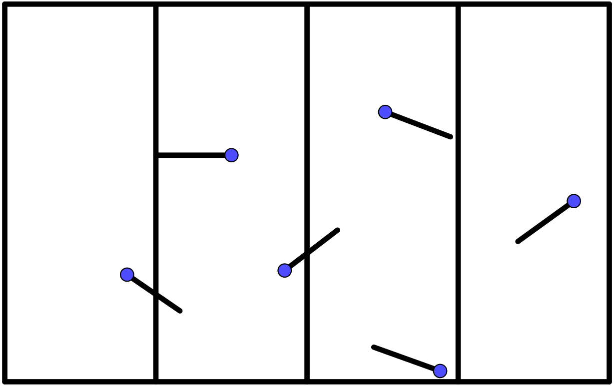
What is the probability of a drop lands in the circle?

$$P(hit) = \frac{\pi r^2}{4r^2} \tag{6}$$



Needles

Proposed by and resolved by the naturalist. take in mind the



Uniform distribution

$x \sim U(a,b)$ in the interval(a,b).

$$f(x) = \frac{1}{b-a} \tag{7}$$

the function is defined in the open interval $a < x < b$. Remember that:

$$F(x) = \int_{-\infty}^{x} f(u)du \tag{8}$$

```
import numpy as np
np.random.uniform(a,b
    ,size=(k,p)) # []

#Draw k list with p elements
    with numbers [a,b)
```

Choose a point in the interval (a,b), you can calculate what it is the probability that a point is in (c,d)

Pontificia Universidad JAVERIANA Bogotá