Statistics and Probability

MAE301 Applied Experimental Statistics

Max Yi Ren

School for Engineering of Matter, Transport Energy
Arizona State University

September 2, 2015

Outline

the birthday problem

What's the chance for multiple people in this class to have the same birthday?
We can use probability for *deduction*, based on the following assumptions:

- Each year has 365 days;
- Each date has the same chance to be your DOB;
- Everyone's DOB is *independent*;

a solution

For n people, the number of DOB combinations is: 365^n; the number of combinations with all different DOBs is: ${}^{365}P_n$. So the chance for n people's DOB to be different (denoted as *event* or (**random variable**) A) is

$$P(A) = \frac{{}^{365}P_n}{365^n} = \frac{365!}{(365-n)!365^n}. \tag{1}$$

Notice that we have some large numbers here, so Taylor expansion! The function $P(n) = \frac{365!}{(365-n)!365^n}$ can be approximated as

$$P(A) \approx e^{-n^2/(2\times 365)} \tag{2}$$

See for the approximation technique.

probability vs statistics

Probability

To predict chances of an unknown event using assumptions - deduction

Statistics

To analyze observed data and make sense of it - induction and deduction

Example

Abraham Wald's Memo

Abraham is tasked with reviewing damaged planes coming back from Germany in the Second World War. He has to review the damage of the planes to see which areas must be protected even more. Abraham finds that the fuselage and fuel system of returned

planes are much more likely to be damaged by bullets or flak than the engines. What should he recommend to his superiors?

Abraham Wald's case study

Consider 100 planes sent out and only 50 came back, 30 with fuselage damaged, 20 with engine damaged, and 10 without damage.

What is the chance of survival?

What is the chance of survival under the two damages?

What is the chance of survival under both damages?

What is the chance of engine damage under fatality?

discrete random variable

Consider the event "plane survived" to have two states: happened and not. We can use a random variable S to describe the event: The plane survived when $S = 1$; and it did not when $S = 0$.

We use $P(S = 1)$ and $P(S = 0)$ to describe the probabilities of survival and fatality. Since these are the only two states the random variable can take, we have $P(S = 1) + P(S = 0) = 1$.

probability distribution: A non-negative function defined on the set of states of a random variable.

joint probability

Now introduce another random variables F and E for fuselage and engine damage: $F = 1$ for fuselage damage and $E = 1$ for engine damage.

The **joint probability** of a plane survived and got both fuselage and engine damage is denoted as $P(S = 1, F = 1, E = 1)$. How do we calculate this number?

conditional probability

The **conditional probability** of both damages under survival is denoted as $P(F = 1, E = 1 | S = 1)$.

Recall that among 50 planes survived, 30 have fuselage, 20 have engine damaged and 10 undamaged. So $P(F = 1, E = 1 | S = 1) = (30 + 20 + 10 - 50)/50 = 0.2$.

The relationship between conditional and joint probabilities:

$$P(S = 1, F = 1, E = 1) = P(F = 1, E = 1 | S = 1) P(S = 1). \quad (3)$$

Recall that among the 100 planes, 50 survived. So $P(S = 1, F = 1, E = 1) = P(F = 1, E = 1 | S = 1) P(S = 1) = 0.2 \times 0.5 = 0.1$.

marginal probability

Given joint probabilities for multiple random variables, we can calculate the **marginal probability** for each random variable.
In this case, the probability of survival can be considered as a marginal probability:

$$\begin{aligned} P(S=1) = P(S=1, F=1, E=1) + P(S=1, F=0, E=1) \\ + P(S=1, F=1, E=0) + P(S=1, F=0, E=0). \end{aligned} \tag{4}$$

The probability of survival under fuselage damage is a marginal probability conditioned on $F = 1$:

$$P(S=1|F=1) = P(S=1, E=1|F=1) + P(S=1, E=0|F=1). \tag{5}$$

Note that we are missing some information to calculate this.

exercise

Consider two discrete random variables A and B, each takes four states. The joint probabilities are as follows:

A,B	1	2	3	4
1	0.04	0.08	0.12	0.16
2	0.03	0.06	0.09	0.12
3	0.02	0.04	0.06	0.08
4	0.01	0.02	0.03	0.04

Calculate $P(A|B=1)$ and $P(B)$.

independence

Two random variables are **independent** if the realization of one does not affect the probability distribution of the other, i.e., for variables A and B, if and only if $P(A|B) = P(A)$ and $P(B|A) = P(B)$, A and B are independent.

Exercise: Show that two variables A and B are independent, if and only if $P(A, B) = P(A)P(B)$.

More than two random variables: $P(\bigcap A_i) = \prod P(A_i)$ if and only if A_is are *mutually independent*, i.e., $P(A_i | A_{i'} \forall i' \neq i) = P(A_i)$.

mutually exclusive

If two events are *mutually exclusive*, then
$P(A = 1 \text{ OR } B = 1) = P(A = 1) + P(B = 1)$ and
$P(A = 1, B = 1) = 0$.

What is the relationship between *mutually exclusive* and independent?

Bayes' theorem

Theorem

For two random variables A and B with probability distributions $P(A)$ and $P(B)$, the Bayes' theorem is as follows

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}. \tag{6}$$

revisit: Abraham Wald's case study

Consider 100 planes sent out and only 50 came back, 30 with fuselage and 20 with engine damaged, 10 undamaged. Further assume that the probabilities of fuselage and engine damages are 0.7 and 0.4, respectively.

What is the chance of survival under the two damages? (fuselage 3/7, engine 0.5)

What is the chance of survival under both damages? (0.5)

What is the chance of damage under fatality? (fuselage 0.8, engine 0.4)

Are engine and fuselage damage events independent? (no)

the Monty Hall problem

Suppose you're on a game show, and you're given the choice of three doors: Behind one door is a car; behind the others, goats. You pick a door, say No. 1, and the host, who knows what's behind the doors, opens another door, say No. 2, which has a goat. He then says to you, "Do you want to pick door No. 3?" Is it to your advantage to switch your choice?

Solution to Monty Hall using Bayes' theorem (1/2)

Introduce a random variable C that takes three discrete states: car behind door 1, 2 and 3. At the beginning, without knowing anything else, we have $P(C = 1) = P(C = 2) = P(C = 3) = 1/3$. Let random variable B be the door the host opened.

Without loss of generality, let's assume that the door we picked is door 1, the door the host opened is door 2. If the car is behind door 1, the chance for the host to open door 2 is 0.5, i.e., $P(B = 2|C = 1) = 1/2$; if the car is behind door 2 or 3, we have $P(B = 2|C = 2) = 0$ and $P(B = 2|C = 3) = 1$.

Solution to the Monty Hall problem using Bayes' theorem (2/2)

To make a decision, we need to calculate $P(C = 1|B = 2)$, $P(C = 2|B = 2)$ and $P(C = 3|B = 2)$. Obviously $P(C = 2|B = 2) = 0$.

$$\begin{aligned} P(C = j|B = 2) &= \frac{P(B = 2|C = j)P(C = j)}{P(B = 2)} \\ &= \frac{P(B = 2|C = j)P(C = j)}{\sum_{i=1}^{3} P(B = 2|C = i)P(C = i)} \end{aligned} \tag{7}$$

What is your conclusion?

Summary of the class

- Discrete random variable: States and probabilities
- Joint probability
- Conditional probability
- Marginal probability
- Independence
- Bayes' theorem

Exercise

- A die is loaded in such a way that an even number is twice as likely to occur as an odd number. If A is the event that a number less than 4 occurs on a single toss of the die, find $P(A)$.
- For three binary random variables A, B and C, how do you calculate $P(A = 1 \text{ OR } B = 1 \text{ OR } C = 1)$?
- An automobile manufacturer is concerned about a possible recall of its best-selling four door sedan. If there were a recall, there is 0.25 probability that a defect is in the brake system, 0.18 in the transmission,0.17 in the fuel system, and 0.40 in some other area. (1) Determine the probability that the defect is the brakes or fueling system if the probability of defects in both systems simultaneously is 0.15; (2) Determine the probability that there are no defects in either the brakes or the fueling system.

Exercise (cont.)

- The probability that a flight departs on time is 0.83; the probability that it arrives on time is 0.82; and the probability that it departs and arrives on time is 0.78. Find the probability that a plane arrives on time given that it departed on time.
- A company buys tires from two suppliers, 1 and 2. Supplier 1 has a record of delivering tires containing 10% defectives, whereas supplier 2 has a defective rate of only 5%. Suppose 40% of the current supply came from supplier 1. If a tire is selected randomly from this supply and observed to be defective, find the probability that it came from supplier 1.
- A coin is flipped twice. Determine the probability that the second flip is heads if the first flip was tails. Are the two flips independent?
- A town has 2 fire engines operating independently. The probability that a specific engine is available when needed is 0.96. Determine the probability that a fire engine is available when needed.

Python code for the birthday problem

```
import numpy as np
import matplotlib.pyplot as plt
# set number of people
n = np.linspace(1, 50, 50).astype(int) # 50 numbers from 1 to 50
# calculate probabilities of everyone having a different birthday
p = np.exp(-n*n/2.0/365.0)
plt.plot(n, p)
plt.show()
```