Class09-11

September 13, 2020

- Conditional probability
- Independence
- Multiplication rule
- Random variables ** Expected Value ** Variance ** Linear combinations
- Bernoulli distribution
- Binomial distribution ** Permutations ** Combinations
- Geometric distribution
- Hypergeometric Distribution
- Normal distribution
- Poisson Distribution **

0.1 Conditional probability

0.1.1 Definition

A **conditional probability** of an event A given B describes the chances that the event A occurs, having already observed an event B.

The conditional probability above can be represented in mathematical notation as

$$p(A|B) \tag{1}$$

For example, the probability of being admitted to the hospital given a patient tested positive for the novel coronavirus (COVID-19). This could be written

$$p(\text{Admitted to the hospital}|\text{Positive test}) \tag{2}$$

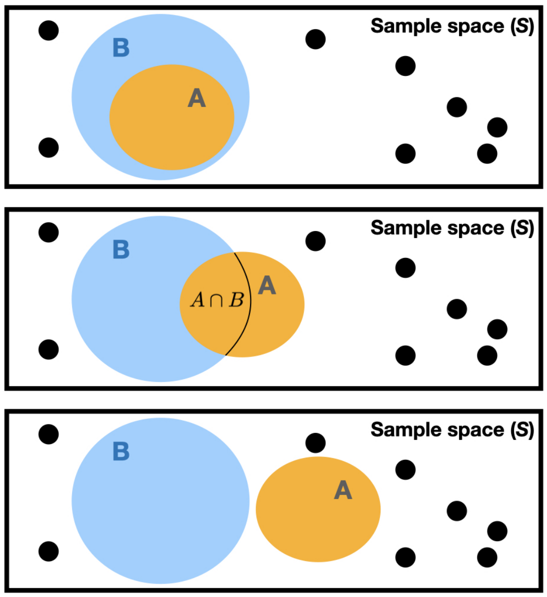
0.1.2 Computation

We can compute a conditional probability by

$$p(A|B) = \frac{p(A \cap B)}{p(B)} \tag{3}$$

The conditional probability is the probability that the events A and B occur simultaneously divided by the probability of event B.

There are three ways two events like A and B can interact to help us understand why we would compute conditional probabilities like this.



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

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

$$p(A|B) = \frac{p(A \cap B)}{p(B)} = \frac{0}{p(B)} = 0$$

In the top panel, the event A only occurs if B occurs. The conditional probability computes the proportion of times A occurs relative to B The bottom panel shows the events A and B never occurring together. Since they never occur at the same time, if the event B occurs the event A will never occur: the conditional probability of A given B is zero. Te middle panel shows a common scenario. There is a subset of outcomes where A occurs when B happens. The conditional probability asks "how many outcomes include the event A and B relative to the number of times B occurs?"

0.1.3 Application

Below are two examples of conditional probabilities, the first more obvious than the second. Suppose we wanted to compute the probability of having SARS-COV-2 given a positive test. We estimate that the probability of having SARS-COV-2 **and** a test returning positive is 0.10. Next, suppose we estimate the probability of a test returning positive whether or not you have SARS-COV-2 is 0.50.

The conditional probability

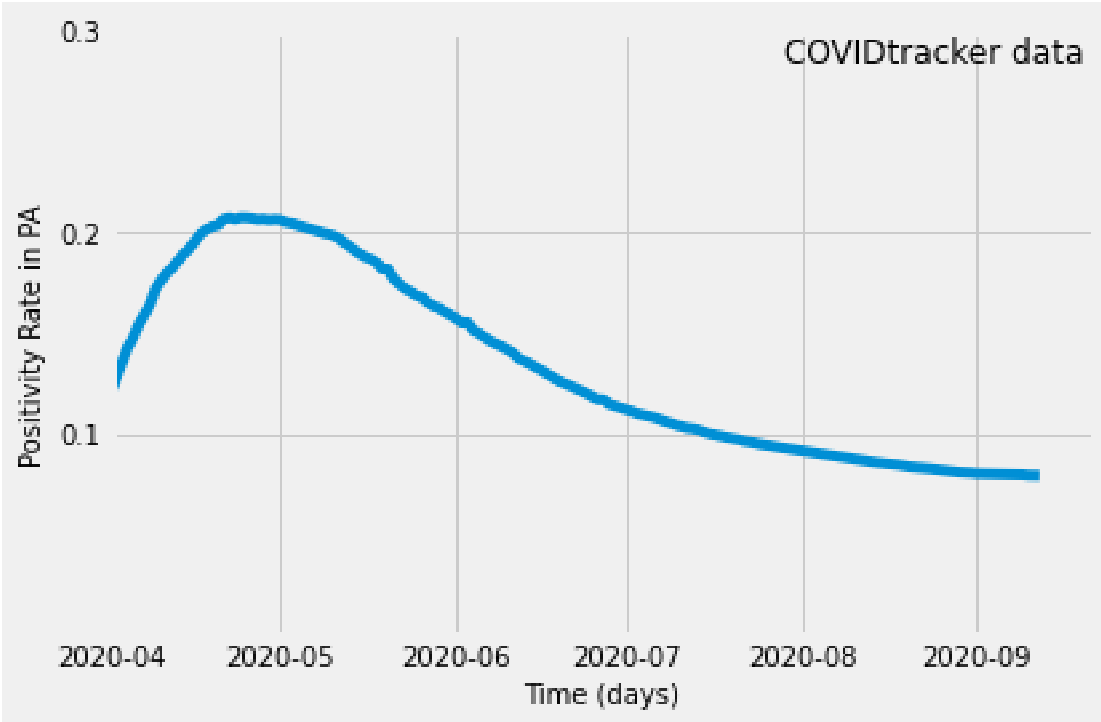
$$\begin{aligned} p(\text{ SARS-COV-2 } \cap \text{ Test Pos. }) &= 0.10 && (4)\\ p(\text{ Test Pos. }) &= 0.50 && (5)\\ p(\text{ SARS-COV-2 } | \text{ Test Pos. }) &= 0.10/0.50 = 20\% && (6)\\ & && (7) \end{aligned}$$

A second example is below and a more subtle use of conditional probabilities. Data on COVID-19 positive rates, the probability of testing positive for SARS-COV-2, was taken from the COVID

Tracking Project. The COVID tracking project is hostsed by the Atlantic. They scour as many news and iinformation sources on COVID-19 as possible to provide best possible estimates of SARS-COV-2/COVID-19 in the US.

Below is a plot of the number of positive tests divided by the total number of tests administered over time (in days) for the state of Pennsylvania. What is this proportion measuring?

[26]:
```
covidData = pd.read_csv("https://covidtracking.com/data/download/
 ↪all-states-history.csv") # downlaod data from the Covidtracker
covidData["positiveRate"] = covidData.positive/covidData.totalTestResults #␣
 ↪compute positivity rate
covidData["date"] = [pd.to_datetime(x,format="%Y%m%d") for x in covidData.date]␣
 ↪# convert integer date to date obj.

paData = covidData[covidData.state=="PA"] # subset to PA

plt.style.use("fivethirtyeight")
fig,ax = plt.subplots() # setup a plotting space

ax.plot(paData.date, paData.positiveRate ) # plot the date by positivity rate

# Format the x and ylimits
ax.set_xlim(pd.to_datetime("2020-04-01"),ax.get_xlim()[-1])
ax.set_ylim(0,0.30)

ax.set_ylabel("Positivity Rate in PA", fontsize=10)
ax.set_xlabel("Time (days)", fontsize=10)

ax.tick_params(labelsize=10)

ax.set_yticks([0.1,0.2,0.3])

ax.text(0.99,0.99,"COVIDtracker data",fontsize=12,transform=ax.
 ↪transAxes,ha='right',va='top')

# a tightlayout asks python to move around objects on the graph for the "best"␣
 ↪possible layout
fig.set_tight_layout(True)
plt.show()
```



0.2 Marginal probs from conditional probs

We can compute marginal probabilities (for example $p(A)$) by first finding a second set of events $(B_1, B_2, \cdots, B_N)$ that is a **partition** of A.

A **partition** of an event A is a collection of sets such that their union equals A

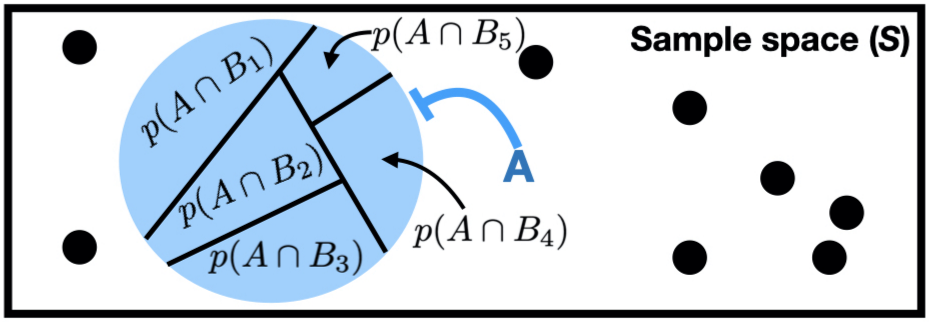
if

$$B_1 \cup B_2 \cup \cdots \cup B_N = A \tag{8}$$

then the collection of events B is a partition for A. We can compute $p(A)$ using a partition as

$$p(A) = p(A|B_1)p(B_1) + p(A|B_2)p(B_2) + \cdots p(A|B_N)p(B_N) \tag{9}$$

$$p(A) = \sum_{i=1}^{N} p(A|B_i)p(B_i) \tag{10}$$



$$p(A) = p(A \cap B_1) + p(A \cap B_2) + p(A \cap B_3) + p(A \cap B_4) + p(A \cap B_5)$$

$$\begin{aligned} p(A) = p(A \cap B_1)\frac{p(B_1)}{p(B_1)} + p(A \cap B_2)\frac{p(B_2)}{p(B_2)} + p(A \cap B_3)\frac{p(B_3)}{p(B_3)} \\ + p(A \cap B_4)\frac{p(B_4)}{p(B_4)} + p(A \cap B_5)\frac{p(B_5)}{p(B_5)} \end{aligned}$$

$$\begin{aligned} p(A) = p(A|B_1)p(B_1) + p(A|B_2)p(B_2) + p(A|B_3)p(B_3) \\ + p(A|B_4)p(B_4) + p(A|B_5)p(B_5) \end{aligned}$$

This equation can come in handy when there is more information about a set of conditional probabilities that partition an event A. A common case is when you know * the probability the event B occurs * the conditional probability of A when B occurs * the conditional probability of A when B does not occurs

One way we could compute the probability of SAR-COV-2 could be to estimate * the probability the a SARS-COV-2 test returns a positive result * the conditional probability of SARS-COV-2 when a test returns a positive result * the conditional probability of SARS-COV-2 when a test returns a negative result

$$\begin{aligned} p(\text{SARS-COV-2}) &= p(\text{SARS-COV-2}|+)p(+) + p(\text{SARS-COV-2}|-)p(-) && (11) \\ &= p(\text{SARS-COV-2}|+)p(+) + p(\text{SARS-COV-2}|-)(1 - p(+)) && (12) \\ & && (13) \end{aligned}$$

and it may be easier to find the probability of a positive and negative test in order to compute the probability of SARS-COV-2. We can use another event that we have data on to compute an event we're interested in.

0.3 Independence and the multiplication rule

We can rearrange the conditional probability of A given B

$$p(A|B) = \frac{p(A \cap B)}{p(B)} \tag{14}$$

$$p(A \cap B) = p(A|B)p(B) \tag{15}$$

to compute the probability of A and B. This is called the **general multiplication rule**.

Two events are called **independent** when the occurrence of one event does not impact the probability of a second event occurring.

$$p(A|B) = p(A) \tag{16}$$

Given that B occurred does not change the probability of A. If two event are independent then computing the general multiplication rule is easier

$$p(A|B) = \frac{p(A \cap B)}{p(B)} \tag{17}$$

$$p(A \cap B) = p(A|B)p(B) \tag{18}$$

$$p(A \cap B) = p(A)p(B) \tag{19}$$

0.4 Baye's Theorem

0.5 Random variables

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]: