Metody probabilistycznej analizy danych

SIUM 2022/23

Plan

Probability – a quick review

Basic applications of the Bayes Theorem

Bayesian Methods with PyMC

Bayesian Networks with GeNIe

Probability

A QUICK REVIEW

Probability – what is it?

Classic definition (Bernoulli, Laplace): ratio of the number of propitious outcomes to the number of all outcomes, assuming all outcomes are equally possible, $P(A) = \frac{N_A}{N}$

Frequentist: assuming infinite number of trials, it is the ratio of number of occurrences of the outcome to the number of trials (relative frequency), $P(A) = \lim_{n \to \infty} \frac{n_a}{n} = \lim_{n \to \infty} f_n(A)$

Subjectivist: one's belief in the outcome of the experiment, based on prior knowledge and experiment circumstances

Basic terms

Experiment (doświadczenie losowe): process or phenomenon that gives information

Outcome (wynik): single piece of information coming from the experiment

Sample space (przestrzeń zdarzeń): set of all possible outcomes of the experiment

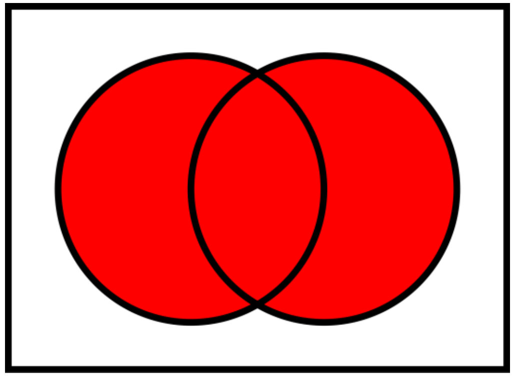
Event (zdarzenie): what comes from the experiment

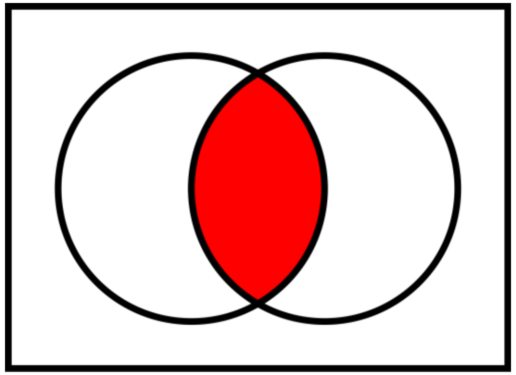
- Every outcome is an event – it is an **elementary** event
- Event may contain many outcomes – it is a **composite** event

Events and sets

Union of sets: $A \cup B$ means that A occurred, or B occurred

Intersection of sets: $A \cap B$ means that A occurred, and B occurred





Probability axioms

For every event E belonging to the sample space S: $0 \leq P(E) \leq 1$

$P(S) = 1$

If $\mathrm{E} \cap F = \emptyset$ then $P(E \cup F) = P(E) + P(F)$

Conditional probability

If E and F are events in S and $P(F) > 0$, then the conditional probability of E given F is defined by

$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$

E.g., what is the probability of passing the exam given attending classes? It is the probability of both passing and attending divided by the probability of attending.

Independence of events

Events are independent if $P(E \cap F) = P(E) \cdot P(F)$

If $P(F) > 0$ both sides can be divided: $\frac{P(E \cap F)}{P(F)} = P(E)$

Using definition of conditional probability $P(E|F) = P(E)$

Are the events of passing an exam and eating doughnuts independent?

Bayes theorem

$$P(E_j \mid A) = \frac{P(A|E_j) \cdot P(E_j)}{\sum_i P(A|E_i) \cdot P(E_i)} = \frac{P(A|E_j) \cdot P(E_j)}{P(A)}$$

Events $E_1, E_2, \ldots, E_N$ partition the sample space: $E_1, \cap E_2 \cap \cdots \cap E_N = S$, $E_i \cap E_j = \emptyset$ for $i \neq j$

Random variable

Function which assigns numerical value to every possible outcome of an experiment

- Denoted by big letters, e.g., X

For each repetition of an experiment only one value of the random variable – its **realization** – is observed

- Denoted by the corresponding small letter

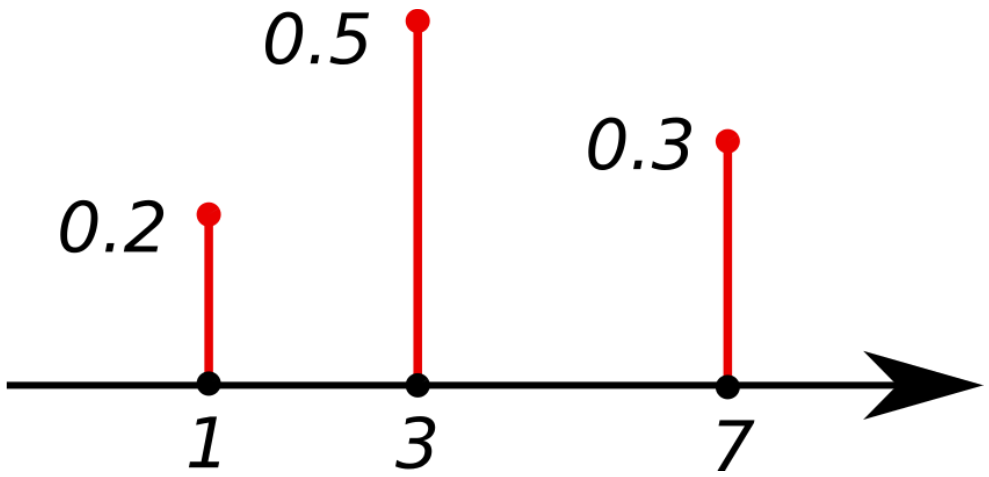
Based on the set of possible outcomes can be

- Discrete
- Continuous

Probability distribution

For a discrete random variable, it is possible for every $x \in R$ to define $p(x) = P(X = x)$

Collection of pairs $\{(x, p(x)), x \in R\}$ is called probability distribution (probability mass function)



Cumulative distribution function

For a discrete random variable, it is defined as

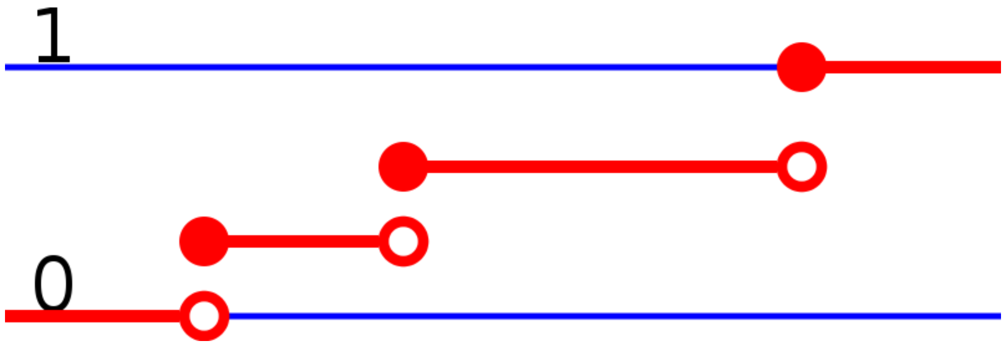
$F(x) = P(X \leqslant x) = \sum_{y \in R : y \leqslant x} p(y)$

Basic properties:

$F(-\infty) = P(X \leqslant -\infty) = 0$

$F(\infty) = P(X \leqslant \infty) = 1$

for $x_1 \leqslant x_2,\ F(x_1) \leqslant F(x_2)$



Expected value

For discrete random variable X with distribution p(x) the expected value is defined as

$E(X) = \sum_{x \in R} x \cdot p(x)$

The expected value does not have to belong to the values of the random variable!

Bernoulli distribution

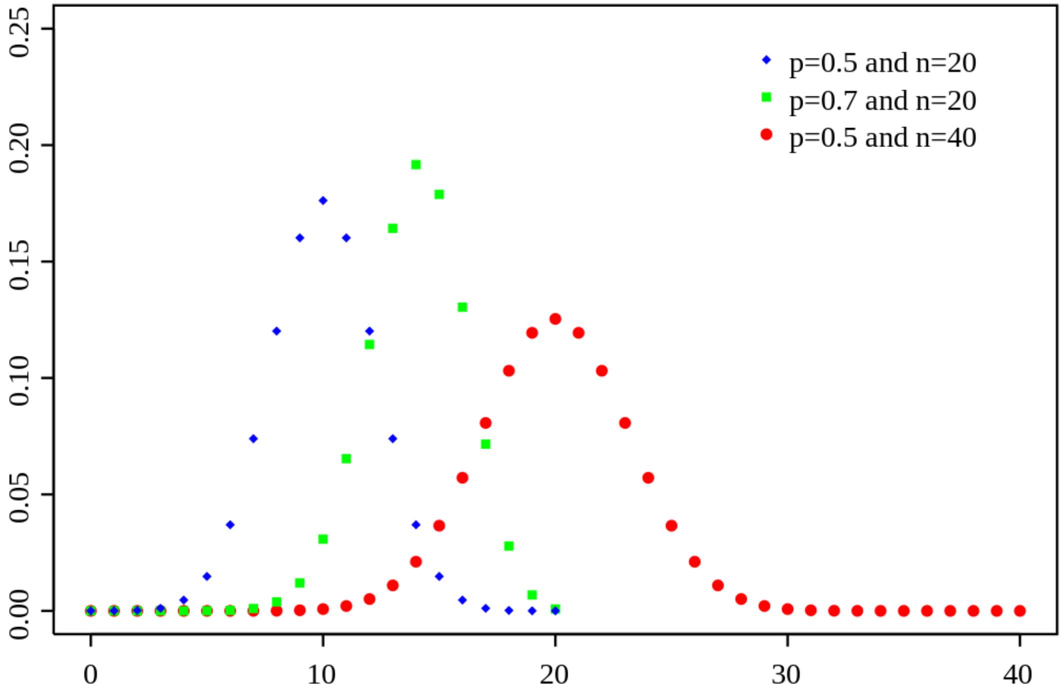
The random variable may take only two values: 0 or 1

$P(X = 1) = p$

$P(X = 0) = 1 - p$

$E(X) = p$

Binomial distribution



Results from n Bernoulli trials:

- Each trial has only two possible outcomes: success or failure
- Results of subsequent trials are independent
- In each trial probability of success is p

Random variable X describes the number of successes in n Bernoulli trials

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$E(X) = np$$

Poisson distribution

Events occur in continuous time or space

Number of events in non-overlapping periods of time or areas is independent

Events occur individually

Events occur with constant average speed or density λ

Random variable X describes number of events in time period or area of interest

$P(X = x) = e^{-\lambda}\lambda^{x}/x!$

$E(X) = \lambda$

Probability density function

For a continuous random variable, for every $x \in R$ $p(x) = 0$

Probability mass function makes no sense

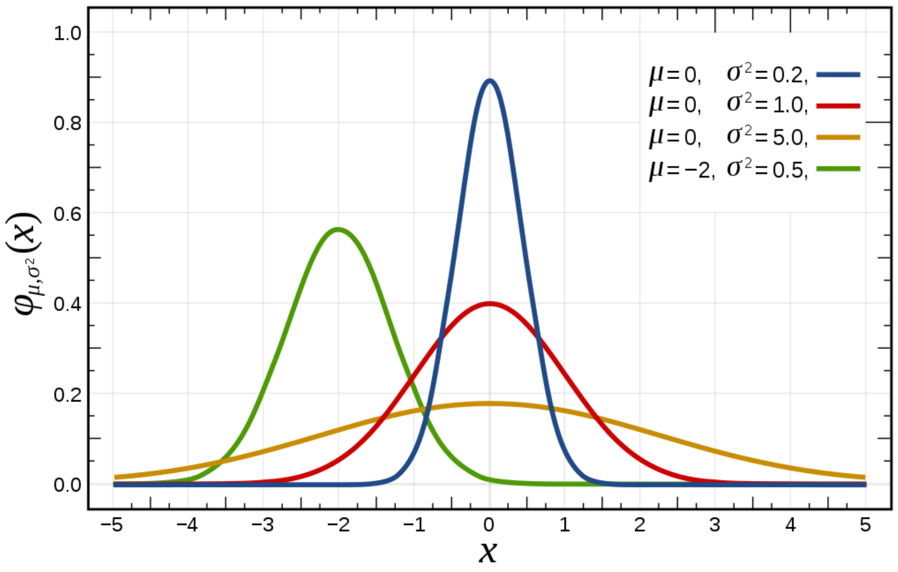
Instead, probability density function is used, derivative of the cumulative distribution function

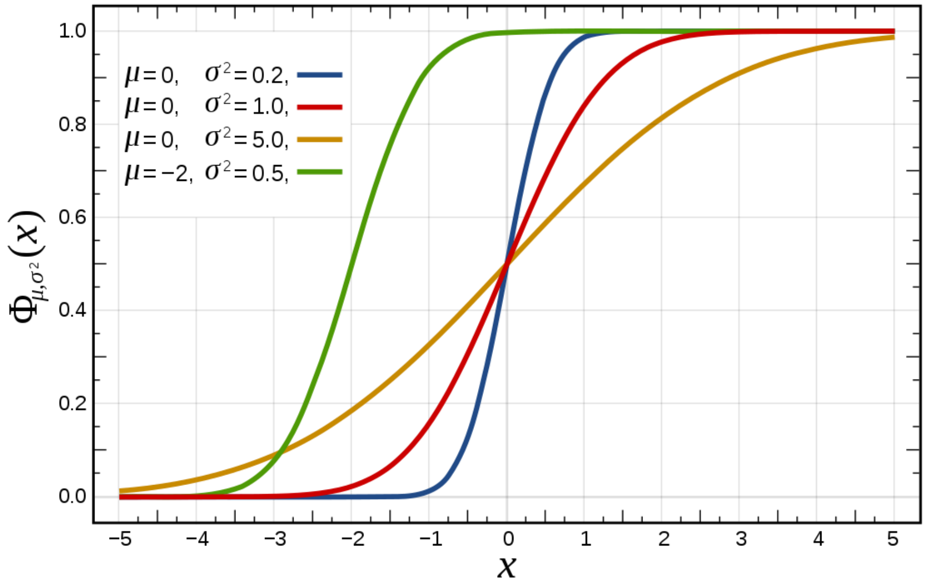
$$f(x) = \frac{d}{dx} F(x)$$

Normal (Gauss) distribution

PDF given by $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$

The parameters are mean and standard deviation (or variance)





Basic applications of the Bayes Theorem

THOMAS BAYES (1701 –1761) WAS AN ENGLISH STATISTICIAN, PHILOSOPHER AND PRESBYTERIAN MINISTER WHO IS KNOWN FOR FORMULATING A SPECIFIC CASE OF THE THEOREM THAT BEARS HIS NAME (FROM WIKIPEDIA).

T. Bayes.

Reasoning backwards

Most people, if you describe a train of events to them, will tell you what the result would be. They can put those events together in their minds, and argue from them that something will come to pass. There are few people, however, who, if you told them a result, would be able to evolve from their own inner consciousness what the steps were which led up to that result. This power is what I mean when I talk of reasoning backwards, or analytically.

Sherlock Holmes
(Arthur Conan Doyle, *A Study in Scarlet*)

Reasoning backwards

If I throw a stone at this window, what are the chances that it will break?

Here's a broken window, what are the chances it was broken by a stone?

If she flies airline X, what are the chances the flight will be delayed?

Her flight was delayed, what are the chances she was flying airline X?

If he has flu, what are the chances he has cough?

He coughs, what are the chances he has flu?

Reasoning backwards

Let there be causal relationship between events C (cause) and E (effect).

$p(E|C)p(C) = p(C|E)p(E)$

People usually can estimate $p(E|C)$ from previous experience, it can be also obtained from typically gathered data

$p(C)$ and $p(E)$ are also often available due to previous observations

$p(C|E)$ is of particular interest, as it gives knowledge about causes when effects are observed

Reasoning backwards

Loss of smell is often present during COVID-19 infection. If a person losses smell, is she/he COVID-19 positive?

C – COVID-19 infection, L – loss of smell

$p(C) = 237118/38268000 \approx 0.0062$ (from active cases in Poland)

$p(L) = 0.01$ (assumed, should be from active cases of smell loss)

$p(L|C) = 0.9$ (numerical expression of the “often present” statement)

$p(C|L) = \frac{p(L|C)p(C)}{p(L)} \approx 0.558$

Losing one’s smell dramatically rises belief in being COVID-19 positive.

Reasoning backwards

And what if loss of smell was a very common complaint?

$p(L) = 0.1$ (order of magnitude greater than before)

$p(L|C) = 0.9$ (as before)

$p(C|L) = \frac{p(L|C)p(C)}{p(L)} \approx 0.058$

When the symptom is common (i.e., occurs also due to other complaints), its presence does not rise our confidence that much.

Reasoning backwards

And what if loss of smell was present only rarely during COVID-19 infection?

$p(L) = 0.01$ (initial value, i.e., loss of smell is not common)

$p(L|C) = 0.1$ (loss of smell is only present in 10% of COVID-19 cases)

$p(C|L) = \frac{p(L|C)p(C)}{p(L)} \approx 0.062$

If the symptom is not strongly tied to the disease in question, observing it does not rise our belief that much.

Reasoning backwards

Going back to the original numbers:

$p(L) = 0.01$, consequently, $p(\neg L) = 0.99$

$p(L|C) = 0.9$, consequently, $p(\neg L|C) = 0.1$

$p(C|\neg L) = \frac{p(\neg L|C)p(C)}{p(\neg L)} \approx 0.00063$

Not having symptoms lowers our belief in the illness.

Note: “not having symptoms” is fundamentally different form “having no data about presence of symptoms”!

Multiple evidence

A burglary, a dog and a silent alarm

In my country house I have a silent burglary alarm. When intrusion is detected, this alarm does not produce any output on-site, but instead calls my phone. Moreover, my country neighbour has a very watchful dog, who is good at detecting any movements or sounds and barks loudly.

Devise a model that will help me decide whether burglary occurred at my country house, taking into account all available evidence.

Independence of evidence

In this example, we may assume independence of the evidence:

- The dog is not influenced by the alarm, as it is a silent alarm.
- Clearly, the alarm is also not influenced by the dog.

Note, that this is only true when the presence of burglary is known for sure (we know that either there is burglary, or not)!

Conditional independence

Consider the following example:

If electrician working in your house by mistake connects 400 V to your single-phase installation, it is very, very probable all of your electrical equipment will break down.

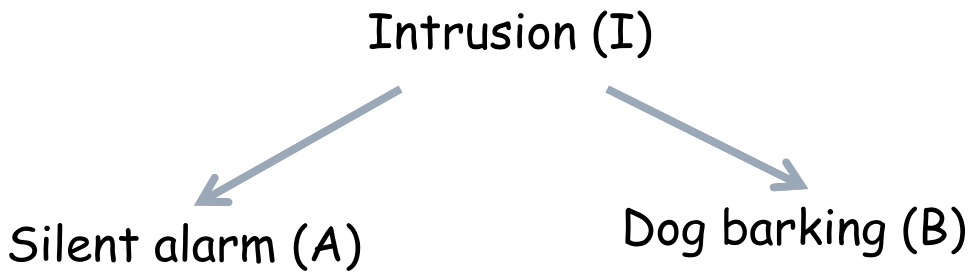
Definitely, if such a mistake is made, breaking down of, say, TV does not influence breaking down of, say, computer.

Situation A: you suspect the electrician made such a mistake (but are not sure). You observe broken TV. Does this influence your belief in the computer being broken?

Situation B: first, you learn that the electrician **did make** this mistake. Afterwards, you observe broken TV. Does this influence your belief in the computer being broken?

Depicting relationship

An informal schematic of the case may look like this:



Note lack of an arrow between (A) and (B) – this indicates conditional independence of the events.

Available data – “don’t know” is an option!

If thinking about events in term of (random) variables, each has only two possible states: *yes* and *no*

However, the model should also handle the “don’t know” situation – the variable is not observed, and we don’t know its state:

- The silent alarm calls our phone number, but we may have displaced our phone, or the battery might have died out
- It is even more obvious for the dog: we are not there, so we must call the neighbour and ask about the dog.

Model formulation

We are interested in $p(I|a,b)$ – a,b denote the observed values of events A and B

Applying definition of conditional probability $p(I|a,b) = \frac{p(I,a,b)}{p(a,b)} = \frac{p(a,b|I)p(I)}{p(a,b)}$

$p(a,b)$ – denotes the probability of events A and B *without reference to the model* – it is our **a'priori belief**. We assume that our belief in these events is independent, i.e., $p(A \cap B) = p(A)p(B)$. The values may be thought as the relative frequency of alarm going off ad dog barking, respectively.

Similarly, $p(I)$ is our a'priori belief in there being intrusion.

$p(a,b|I)$ – denotes the probability of events A, B given I (intrusion). Because A and B are conditionally independent given I, $(A \perp\!\!\!\perp \mathrm{B})|\mathrm{C}$, $p(a,b|I) = p(a|I)p(b|I)$

Model formulation

$$p(I|a,b) = \frac{p(I,a,b)}{p(a,b)} = \frac{p(a,b|I)p(I)}{p(a,b)} = \frac{p(a|I)p(b|I)p(I)}{p(a)p(b)} = \frac{p(a|I)p(b|I)}{p(a)p(b)}p(I)$$

Let’s write down conditional probability of I given *only one* piece of evidence, e.g., the alarm

$$p(I|a) = \frac{p(I,a)}{p(a)} = \frac{p(a|I)p(I)}{p(a)} = \frac{p(a|I)}{p(a)}p(I)$$

Putting into first equation $p(I|a,b) = \frac{p(b|I)}{p(b)}p(I|a)$

This means we may update our belief sequentially: we absorb one piece of evidence, then the updated belief becomes our a’priori belief for the subsequent update.

Putting in some numbers

We need a'priori probabilities: $p(a), b(b), p(i)$ and conditional probabilities $p(a|i), p(b|i)$

Sample situation might look as follows:

$p(I) = 0.002$ (for my neighbourhood: average ratio of the houses that are burgled during the night to the total houses)

$p(A) = 0.01$ (for my alarm: ratio of the nights it goes off to the nights in the assumed time period; it goes off twice a night so rarely we may treat is as a single event)

$p(B) = 0.50$ (for my neighbour's dog: ratio of the days he barks during the night to the nights in the assumed time period)

$p(A|I) = 0.8$ (for my alarm: probability of it going off when burglary takes place)

$p(B|I) = 0.98$ (for my neighbour's dog: probability of it barking when burglary takes place; the dog is *very* watchful)

Getting results

Scenario 1

My alarm phoned me this night! What is the probability of burglary taking place?

$p(I|A) = \frac{p(A|I)p(I)}{p(A)} = \frac{0.8 \cdot 0.002}{0.01} = 0.16$ – much higher than the prior of 0.002

I'm nervous so I called my neighbour: he informs me that the dog barked last night

$p(I|A,B) = \frac{p(B|I)}{p(B)} p(I|A) = \frac{0.98}{0.5} 0.16 = 0.3136$ – the probability did not go up that much, the dog is not reliable as he barks every other night

Getting results

Scenario 2

My alarm phoned me this night! What is the probability of burglary taking place?

$p(I|A) = \frac{p(A|I)p(I)}{p(A)} = \frac{0.8 \cdot 0.002}{0.01} = 0.16$ – much higher than the prior of 0.002

I'm nervous so I called my neighbour: he informs me that the dog did not bark last night.

$p(I|A, \neg B) = \frac{p(\neg B|I)}{p(\neg B)} p(I|A) = \frac{0.02}{0.5} 0.16 = 0.0064$ – the probability dropped enormously, it is very improbable the dog did not notice the burglars

Getting results

Scenario 3

I just talked to my neighbour, he tells me his dog barked during the night. Should I worry?

$p(I|B) = \frac{p(B|I)p(I)}{p(B)} = \frac{0.98 \cdot 0.002}{0.5} = 0.00392$ – not that different from the prior of 0.002

Anyway, I double-checked the status of the alarm – it did not go off.

$p(I|\neg A, B) = \frac{p(\neg A|I)}{p(\neg A)} p(I|B) = \frac{0.2}{0.99} 0.00392 \approx 0.00079$ – probability dropped, but not that much: my alarm is not really that sensitive.

Getting results

Scenario 4

I just talked to my neighbour, he tells me his dog barked during the night. Should I worry?

$p(I|B) = \frac{p(B|I)p(I)}{p(B)} = \frac{0.98 \cdot 0.002}{0.5} = 0.00392$ – not that different from the prior of 0.002

Anyway, I double-checked the status of the alarm – it did go off.

$p(I|A, B) = \frac{p(A|I)}{p(A)} p(I|B) = \frac{0.8}{0.01} 0.00392 = 0.3136$ – as in scenario 1.

Introducing PyMC3

HTTPS://DOCS.PYMC.IO/

What is it?

Python package for probabilistic programming (i.e., programming allowing to cope with distributions, random variables etc.)

Based on Theano

Employs MCMC (Markov chain Monte Carlo) methods to sample distributions – this allows to get samples from continuous variable distributions

- basic Mont Carlo methods sample distributions in completely random manner – for multi-dimensional problems this may be a problem
- MCMC is a guided way to sample distribution with emphasis on samples that most influence the computed output

Basic concepts

The basic building block is a **stochastic** variable:

```
x = pm.Normal("x", mu=0, sigma=1)
```

- defines **stochastic** variable with standard normal distribution

The variables must be contained in a model

```
with pm.Model() as model:
    x = pm.Normal("x", mu=0, sigma=1)
```

Each variable has an initial value, that is the starting point of sampling

```
print(x.tag.test_value)
```

```
0.0
```

Basic concepts

PyMC3 works with stochastic and deterministic variables:

- value of a stochastic variable is not known even if all its parameters and components are known. Instances of distribution classes (Normal, Exponential, etc.) are examples of stochastic variables.

```
x = pm.Normal("x", mu=0, sigma=1)
```

- value of a deterministic variable is known exactly once its components are known. These are variables similar to the ones used in normal programming.

```
d = pm.Deterministic("d", x + 1)
```

- deterministic variables are also created by elementary operations, however, in such case the result is not stored

```
d2 = x + 1
```

Basic concepts

Using random method, it is possible to sample the distribution:

```
print(x.random(size=10))
```

```
[-1.32036024 -0.23443528 -1.5668316   0.15234133  0.75418639 -0.24177129
  1.03009621  0.27025339 -0.37722446  0.30920738]
```

For any given argument(s) of a distribution, it is possible to obtain the log-probability value using logp method

- log-probability is just natural logarithm of probability, or, in case of continuous random variables, of the value of the pdf
- use of logarithms makes sense, as for the independent random variables $P(E \cap F) = P(E) \cdot P(F)$. Logarithm transforms multiplication into addition, $\log\big(P(E) \cdot P(F)\big) = \log\big(P(E)\big) + \log\big(P(F)\big)$

```
print(x.logp({"x":0}))
```

```
-0.9189385332046727
```

Example: investigate parameter of a distribution

You invented a new anti-viral drug. You want to test how it works for COVID-19 patients. You select a group of patients, give each patient the drug, wait 5 days and check whether the patient is healthy (success) or ill (failure). Taking into account the size of the group and obtained results, you want to know the probability the drug will cure a patient, and how sure you can be of this value.

Let's build a model. Question 1: what kind of distribution governs the process?

- for each patient there are two possible outcomes: success and failure
- patients are independent
- the drug works the same for every patient (pretty strong assumption, requires pre-selecting patients with similar background: age, sex, co-existing illnesses etc.)

Bernoulli distribution

Bernoulii distribution takes one parameter, p – the probability of success.

We need to put this distribution into our model, but don't know this parameter (in fact, this is the parameter we want to find out).

We must inform the model what range this parameter may take and how probable are the values of this parameter.

We use another distribution for this. Question 2: what kind of distribution will we use?

- the parameter is probability, so it may take any value from 0 to 1
- we have no idea about the efficiency of the drug, so we do not prefer any values

Uniform distribution

For this case $a = 0, b = 1$

Let's put it into a model:

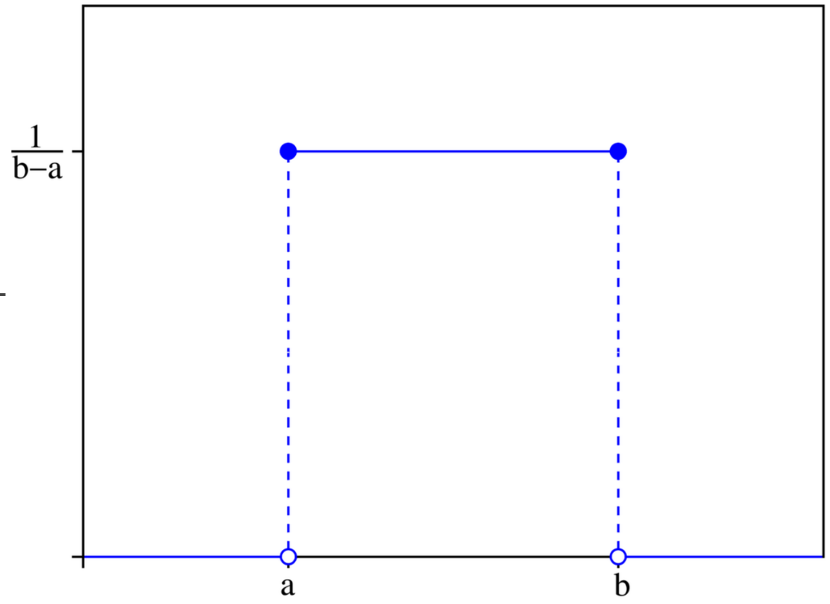
```
with pm.Model() as model:
    p = pm.Uniform('p', lower=0, upper=1)
```

This variable is **unobserved**: we don't see its value directly in our experiment

Now we will add Bernoulli. This, on the other hand, is observed: we see its values during the experiment (the patient is either cured, or not)

```
obs = pm.Bernoulli("obs", p, observed=occurrences)
```

In order for the model to work, we need the occurrences variable: a list containing results of the experiment.



Running model

The sample method performs inference

```
idata = pm.sample(2000, tune=2500)
```

- first parameter indicates number of samples to draw
- tune indicates number of samples to draw during tune-in phase (they are discarded)

One way to analyse results is to plot the so-called trace. It can be easily done using arviz package.

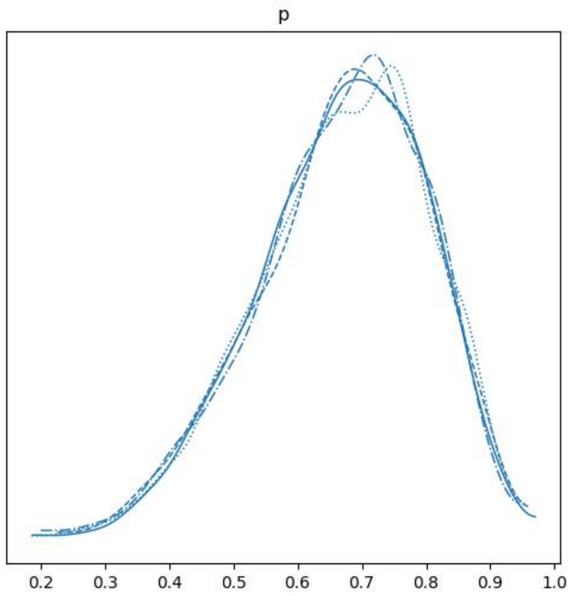
Complete code:

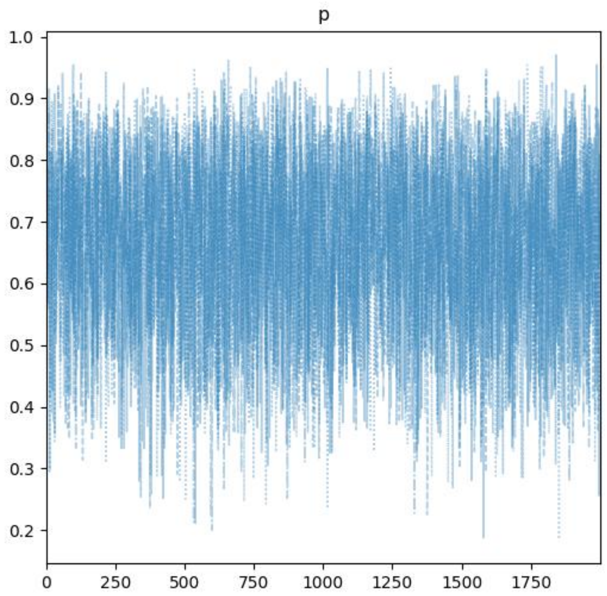
```
import arviz as az
import pymc3 as pm

with pm.Model() as model:
    p = pm.Uniform("p", lower=0, upper=1)
    obs = pm.Bernoulli("obs", p, observed=occurrences)
    idata = pm.sample(2000, tune=2500)
az.plot_trace(idata, show=True)
```

Observations

Let’s see for 10 patients, 7 of which were cured:

```
occurrences = [1, 1, 1, 1, 1, 1, 1, 0, 0, 0]
```

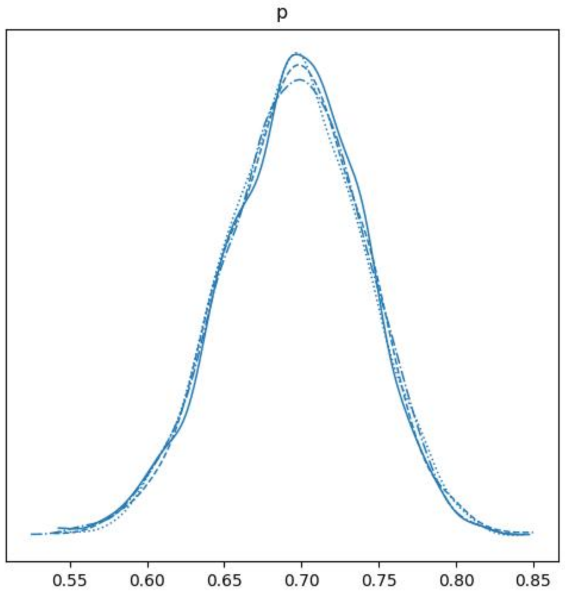


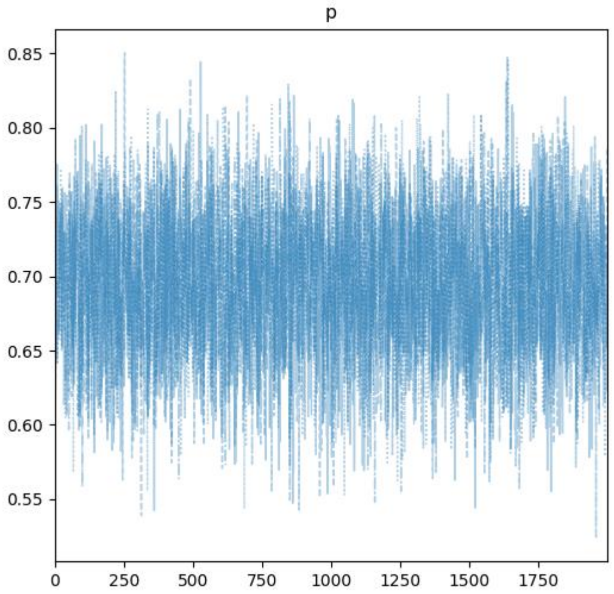


Observations

Let’s see for 100 patients, 70 of which were cured. Note that the percentage of cured remains the same.

```
ones = np.ones(70)
zeros = np.zeros(30)
occurrences = np.concatenate((ones, zeros))
```

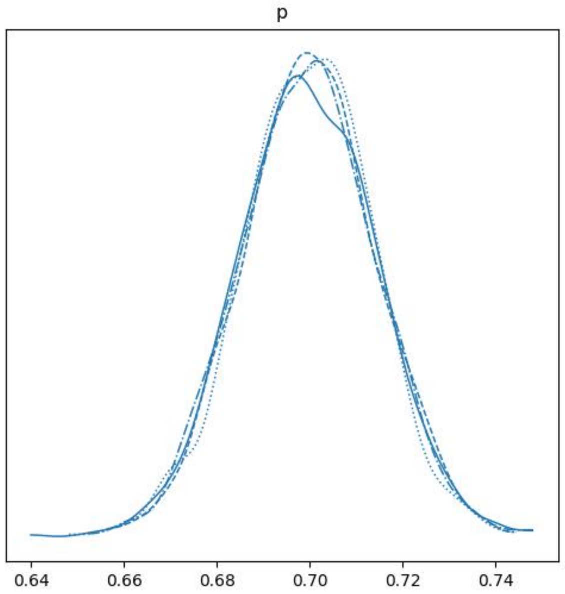


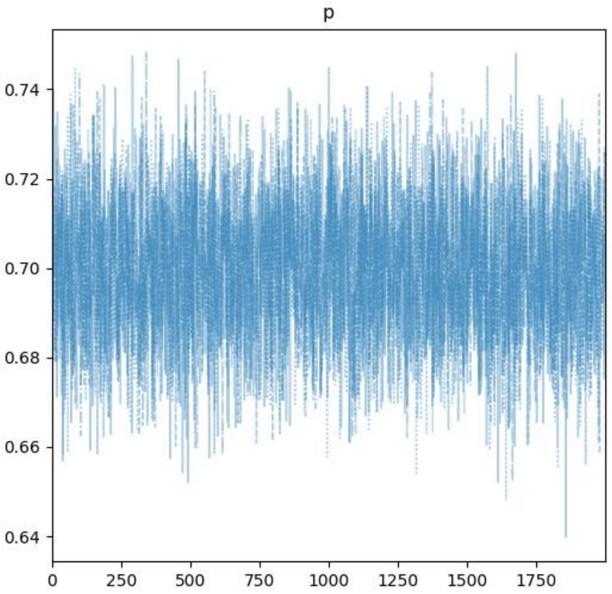


Observations

And now for 1000 patients, 700 of which were cured. The percentage of cured still remains the same.

```
ones = np.ones(700)
zeros = np.zeros(300)
occurrences = np.concatenate((ones, zeros))
```





Are you taking drugs?

You will be well paid for conducting a survey among students at the University regarding use of drugs. The condition is that the survey must be done during online meetings (e.g., in MS Teams) without any additional technical means, save the simplest everyday objects. As asking a person a point-blank question regarding drug usage is unlikely to result in credible answers, devise a probabilistic way to approximate the truth.

Possible approach

Ask the student to perform the following algorithm:

1. toss a coin. Do not tell the result. If it's heads, go to 2, if tails go to 3.

2. toss a coin. Do not tell the result. If it's heads, say that you are taking drugs, if tails, say you are not taking drugs. End.

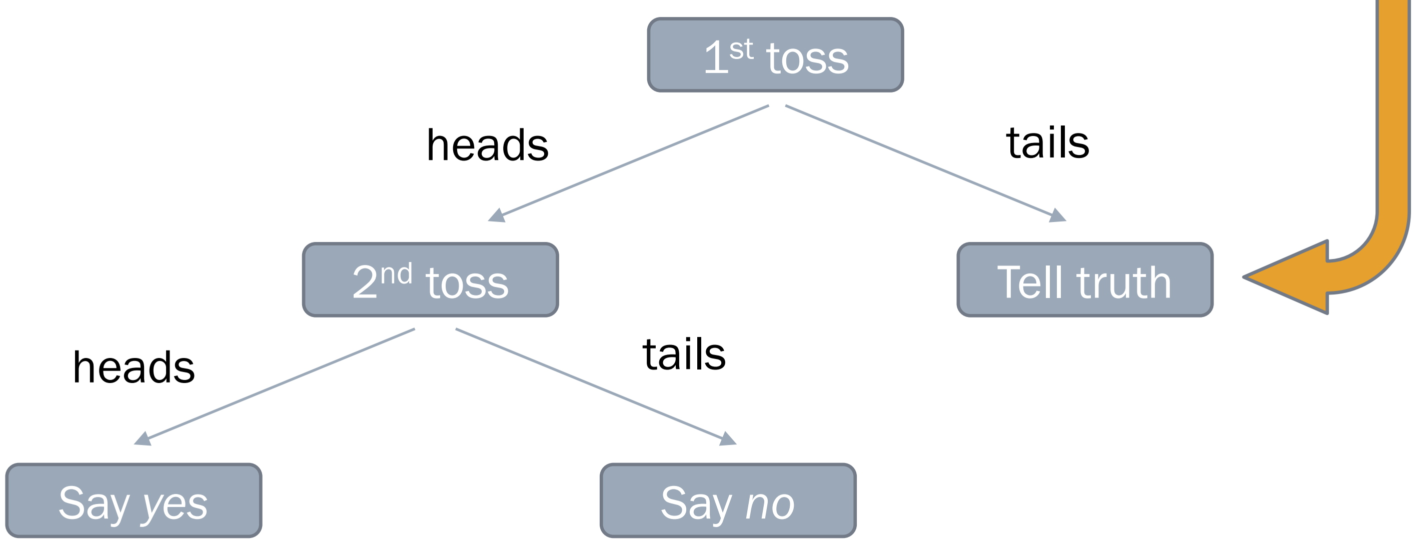
3. Say honestly whether you are taking drugs. End.

Rationale: The person conducting survey does not know whether the answer results from the person habits or is a result of the coin toss.

Encode the approach as a model

Question 1: What is the probability of obtaining “I’m taking drugs” answer?

Suppose that of the N surveyed students n_d are taking drugs. What is the probability that a randomly chosen student takes drugs? $p_d = n_d/N$



Probability of obtaining "Yes, I'm taking" answer

1st toss heads　　　　　1st toss tails

$$p(Y) = \boxed{\frac{1}{2}}\left(\boxed{\frac{1}{2}} \cdot 1 + \boxed{\frac{1}{2}} \cdot 0\right) + \boxed{\frac{1}{2}} p_d = \frac{1}{4} + \frac{1}{2} p_d$$

2nd toss heads　　　　　2nd toss tails

Put it into a model

The obtained formula “transforms” the probability we are interested in (p_d) into the probability we observe ($p(Y)$).

This is a deterministic process, so deterministic variable will be used:

```
p_observed = pm.Deterministic("p_observed", 0.25 + 0.5*p_d)
```

Prior distribution and input data

Question 2: what will be the prior distribution of p_d?

If we have no prior knowledge, uniform will do:

```
p_d = pm.Uniform("p_d", 0, 1)
```

Question 3: what kind of data will be available?

Each student will provide yes/no answer.

Students are anonymous, indistinguishable.

Total number of respondents will be known after the survey is complete.

Putting in input data

The data can be treated as Bernoulli distribution:

```
occurrences = np.concatenate((np.zeros(70), np.ones(30)))
answers = pm.Bernoulli("answers", p_observed, observed=occurrences)
```

or as binomial distribution:

```
yes_count = 70
answers = pm.Binomial("answers", 100, p_observed, observed=yes_count)
```

And now for something completely different…

In some bank operations the transaction is rounded to the whole $. Consequently, the real value of the transaction will differ from the value recorded in the account.

We will consider one year of transactions. We are interested in the impact of this policy on our bank account, i.e., how much can we gain or loose in one year due to rounding.

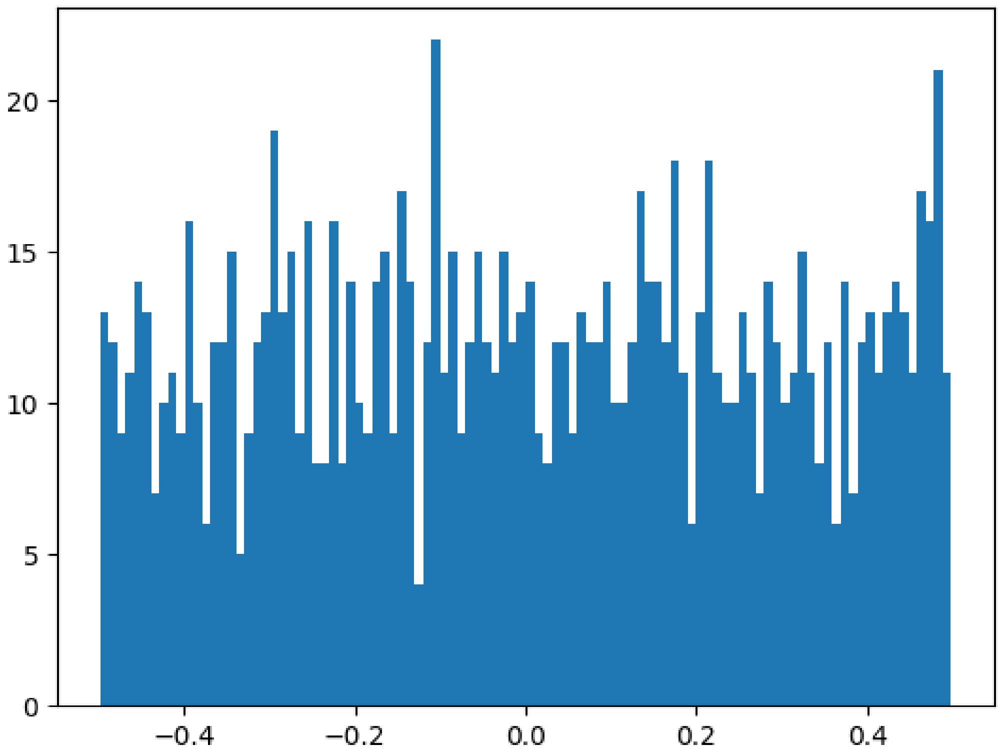
Let’s say we make yearly 1200 such transactions. What can we expect?

I WANT YOUR ESTIMATES

Underlying distribution

Question: What is the distribution of difference between real and recorded value for a single transaction?

Sample client



Write some code

```
result = []
for i in range(10000):
    u = sta.uniform.rvs(loc=-0.5, size=1200)
    u_sum = np.sum(u)
    result.append(u_sum)
plt.hist(result, bins=100)
plt.show()
```

Simulate 10 000 clients.

Use uniform distribution, it defaults to 0, 1 bounds, so use the loc parameter to shift it to −0.5, 0.5. Make 1200 draws (“transactions”), rvs method will draw random numbers.

Add transactions.

Record result.

Be astonished by the result