Data Analysis in Astronomy and Physics

Lecture 4: Probability

M. Röllig

Introduction

There are various ways to describe what probability is.

It is motivated by the question: "**Will an event occur?**"

More subjectively, it tries to answer the question: "How certain are we that a particular event will occur?"

Generally speaking, **probability is a numerical formalization of our degree or intensity of belief**.

Introduction

There are various ways to describe what probability is.

It is motivated by the question: "**Will an event occur?**"

More subjectively, it tries to answer the question: "How certain are we that a particular event will occur?"

Generally speaking, **probability is a numerical formalization of our degree or intensity of belief**.

Quiz

Which of the following events would you be most surprised by?

1. exactly 3 heads in 10 coin flips
2. exactly 3 heads in 100 coin flips
3. exactly 3 heads in 1000 coin flips

Introduction

In a purely theoretical random experiment, with a well defined setting, like throwing a coin or casting a dice, probabilities describe the statistical number of outcomes considered divided by the number of all outcomes.
For example: tossing a fair coin twice will yield heads-heads with probability 1/4, because the four outcomes heads-heads, heads-tails, tails-heads and tails-tails are **equally likely** to occur.

Practically, there are two major competing categories of probability interpretations:

- Objectivists assign numbers to describe some objective or physical state of affairs. Most frequent view: **frequentist probability**, which claims that the probability of a random event denotes the relative frequency of occurrence of an experiment’s outcome, when repeating the experiment. This interpretation considers probability to be the relative frequency “in the long run” of outcomes.
- Subjectivists assign numbers per subjective probability, i.e., as a degree of belief. Interpretation: the price at which you would buy or sell a bet that pays 1 unit of utility if E, 0 if not E.
  Most popular view: **Bayesian probability**, which includes expert knowledge as well as experimental data to produce probabilities. The expert knowledge is represented by some (subjective) prior probability distribution.

Probability in astronomical context

- Astronomical measurements are subject to random measurement error, particularly because of
  - the inability to carefully **set up an experiment**.
  - the inability to **re-run an experiment**/observation
  - and the perpetual wish to observe at the **extreme limits** of instrumental capabilities.
- We can not do experiments on our subject matter (e.g. on all low-mass stars).
  We do not know the population properties.
  - We always observe, i.e. do experiments, on very small **samples**.
  - We conclude by comparing or contrasting these samples to other samples.

Sampling/Combinatorics

Draw a pen from a drawer containing 4 different pens (red, green, blue, purple).

$\mathcal{S} = \{R, G, B, P\}$

Two possibilities:

- **sampling with replacement**

  Example: draw 2 pens from the drawer

  $$\mathcal{S} = \{ \begin{array}{llll} RR, & RG, & RB, & RP, \\ GR, & GG, & GB, & GP, \\ BR, & BG, & BB, & BP, \\ PR, & PG, & PB, & PP\} \end{array}$$

  The event: “at least one red pen” is {RR,RG,RB,RP,GR,BR,PR} with probability 7/16.

Sampling/Combinatorics

Draw a pen from a drawer containing 4 different pens (red, green, blue, purple).

$\mathcal{S} = \{R, G, B, P\}$

Two possibilities:

- **sampling without replacement**

  Example: draw 2 pens from the drawer

$$\mathcal{S} = \left\{ \begin{array}{llll} & RG, & RB, & RP, \\ GR, & & GB, & GP, \\ BR, & BG, & & BP, \\ PR, & PG, & PB, & \end{array} \right\}$$

The event: “at least one red pen” is {RG,RB,RP,GR,BR,PR} with probability 6/12=1/2.

Ordering

If we consider RG and GR to be different events → **ordered sample**.

If order does not matter, e.g. because we draw two pens together → **unordered sample**

Example: draw 2 pens from the drawer at the same time

$$\mathcal{S} = \{\{R, G\}, \{R, B\}, \{R, P\}, \{G, B\}, \{G, P\}, \{B, P\}\}$$

The event “at least one red pen” is {{R,G}, {R,B}, {R,P}} with probability 3/6=1/2.

Sample size

The sample sizes in the three cases above can be calculated according to:

$$n! = n((n-1)(n-2)\ldots 1$$

$$^{n}P_{k} = n(n-1)(n-2)\ldots(n-k+1) = \frac{n!}{(n-k)!}$$

$$^{n}C_{k} = {}^{n}P_{k}/k! = \frac{n!}{k!\,(n-k)!} = \binom{n}{k}$$

	with replacement	without replacement
ordered sample	n^k	$^{n}P_{k}$
unordered sample	$^{n+k-1}C_{k}$	$^{n}C_{k}$

  The number of selections of k objects from a set of n objects

Sample size

Example:

A box contains 20 balls, of which **10 are red** and **10 are blue**. We draw ten balls from the box, and we are interested in the event that exactly 5 of the balls are red and 5 are blue. Do you think that this is more likely to occur if the draws are made with or without replacement? Let $\mathcal{S}$ be the sample space, and $\mathcal{A}$ the event that five balls are red and five are blue.

	with replacement	without replacement
ordered sample	n^k	${}^nP_k$
unordered sample	${}^{n+k-1}C_k$	${}^nC_k$

The number of selections of k objects from a set of n objects

Definitions

Originally, probability was used to analyze games of chance.

Probability then interpreted as the limiting case of a frequency.

- probability $= \frac{\text{number of favourable events}}{\text{total number of events}}$
- the *sample space* $\mathcal{S}$ (also called *outcome space*) is the **set of all possible outcomes** of a random experiment

  Example: tossing a coin three times: $\mathcal{S}$={HHH,HHT,HTH,HTT,THH,THT,TTH,TTT}
- an **event $\mathcal{E}$ is a subset of $\mathcal{S}$**

  Example: event $\mathcal{A}$ = “more tails than heads” = {HHH,HHT,HTH,THH} prob = 1/2
  event $\mathcal{B}$ = “heads on last throw” = {HHH,HTH,THH,TTH} prob = 1/2

Sample size

Example:

- **Sampling with replacement**
  - $| \mathcal{S} | = 20^{10}$ (1st draw: 20 possibilities, 2nd draw 20 possibilities, ...)
  - $| \mathcal{A} | = ?$
    - number of ways to draw {RRRRRBBBBB} is $10^5 \times 10^5 = 10^{10}$
    - many other combinations possible. There are ${}^{10}C_5 = 252$ different patterns of five R's and five B's
    - $| \mathcal{A} | = 252 \times 10^{10}$
  - $P(\mathcal{A}) = \frac{252 \times 10^{10}}{20^{10}} = 0.246 \ldots$

Sample size

Example:

- **Sampling without replacement**
  - $|\ \mathcal{S}\ | = {}^{20}P_{10}$ (${}^{10}P_5$ ways of choosing 5 of the ten red balls, same for blue)
  - $|\ \mathcal{A}\ | = \left({}^{10}P_5\right)^2 . {}^{10}C_5$
  - $P(\mathcal{A}) = \frac{\left({}^{10}P_5\right)^2 . {}^{10}C_5}{{}^{20}P_{10}} = 0.343 \ldots$

Sample size

Example:

- **Unordered sampling**
  - $|\mathcal{S}| = {}^{20}C_{10}$
  - $|\mathcal{A}| = \left({}^{10}C_5\right)^2$
  - $P(\mathcal{A}) = \frac{\left({}^{10}C_5\right)^2}{{}^{20}C_{10}} = 0.343\ldots$

Definitions

- we can build new events from old ones:
  - $\mathcal{A} \cup \mathcal{B}$ ($\mathcal{A}$ **union** $\mathcal{B}$) consists of all outcomes in $\mathcal{A}$ or in $\mathcal{B}$, or both
    Example: $\mathcal{A} \cup \mathcal{B}$ = {HHH,HHT,HTH,THH,HTH,TTH}
  - $\mathcal{A} \cap \mathcal{B}$ ($\mathcal{A}$ **intersection** $\mathcal{B}$) consists of all outcomes in $\mathcal{A}$ and $\mathcal{B}$
    Example: $\mathcal{A} \cap \mathcal{B}$ = {HHH,THH,HTH}
  - $\mathcal{A} \setminus \mathcal{B}$ ($\mathcal{A}$ **minus** $\mathcal{B}$) consists of all outcomes in $\mathcal{A}$ but not in $\mathcal{B}$ (also called relative complement)
    Example: $\mathcal{A} \setminus \mathcal{B}$ = {HHT}
  - $\mathcal{A}'$ ($\mathcal{A}$ **complement**) consists of all outcomes not in $\mathcal{A}$ (i.e. $\mathcal{S} \setminus \mathcal{A}$) (absolute complement)
    Example: $\mathcal{A}'$ = {HTT,THT,TTH,TTT}
  - $\emptyset$ (**empty set**) for the event which doesn’t contain any outcomes.
    Example $\emptyset$ = {}

Kolmogorov's Axioms

1. For any event $\mathcal{A}$, we have the probability $\mathcal{P}(\mathcal{A})\geq 0$
2. $\mathcal{P}(\mathcal{S})=1$
3. If events $\mathcal{A}_1$, $\mathcal{A}_2$, … are pairwise disjoint, then

$$\mathcal{P}(\mathcal{A}_1 \cup \mathcal{A}_2 \cup \ldots) = \mathcal{P}(\mathcal{A}_1) + \mathcal{P}(\mathcal{A}_2) + \ldots$$

Usually, in absence of any further information we assume equal probability for all events. (**Laplace's principle of indifference**).

Example: Tossing a coin has two possible results: head and tail. $\mathcal{P}(\text{head}) = \frac{1}{2}$

Example: Casting a 20-sided dice and getting a prime number: there are 8 possible primes: {2,3,5,7,11,13,17,19}
$\mathcal{P}(\text{prime}) = \frac{8}{20} = \frac{2}{5}$

- to define a probability we have to assume that each side of the dice is equally probable.
- if we identify equally probable cases, calculating a probability means simply enumerating them!
  (identifying the equally probable cases is the difficult part)

Proving things from the axioms

If the event A contains only a finite number of outcomes, say $\mathcal{A} = \{a_1, a_2, \ldots, a_n\}$, then $\mathcal{P}(\mathcal{A}) = \mathcal{P}(a_1) + \mathcal{P}(a_1) + \ldots + \mathcal{P}(a_n)$

Define a new event $\mathcal{A}_i$=$\{a_i\}$ for $i = 1, \ldots, n$. Then $\mathcal{A}_1,\ldots,\mathcal{A}_n$ are mutually disjoint (each contains only one element which is none of the others) and

$$\mathcal{A}_1 \cup \mathcal{A}_2 \cup \ldots \cup \mathcal{A}_n = \mathcal{P}(a_1) + \mathcal{P}(a_2) + \ldots + \mathcal{P}(a_n)$$

Example: We draw a card from a full deck of cards. What is the probability to draw a “3”?

Solution: The event of interest $\mathcal{A} = \{a_1, a_2, a_3, a_4\} = \{\diamondsuit 3, \clubsuit 3, \heartsuit 3, \spadesuit 3\}$. Each card is equally probable

$$\mathcal{P}_r(a_i) = \text{number of favorable outcomes} / \text{total number of events} = 1/52$$

$$\mathcal{P}_r(\mathcal{A}) = \mathcal{P}_r(a_1) + \mathcal{P}_r(a_2) + \mathcal{P}_r(a_3) + \mathcal{P}_r(a_4) = 4/52 = \frac{1}{13}$$

Proving things from the axioms

If the sample space $\mathcal{S}$ is finite, say $\mathcal{S} = \{a_1, a_2, \ldots, a_n\}$, then $\mathcal{P}(a_1) + \mathcal{P}(a_1) + \ldots + \mathcal{P}(a_n) = 1$

From the proposition above it follows $\mathcal{P}(a_1) + \mathcal{P}(a_1) + \ldots + \mathcal{P}(a_n) = \mathcal{P}(\mathcal{S})$. From Axiom 2 follows $\mathcal{P}(\mathcal{S})$=1. From this we also see, that if all single events a_i are equally probably and their probabilities sum is 1, then each has probability $1/n$.

Example: Consider a full deck of cards.

Solution: $\mathcal{S} = \{a_1, a_2, \ldots, a_{52}\} = \{\diamondsuit 2, \clubsuit 2, \ldots, \spadesuit A\}$

Each card is equally probable

$$\mathcal{P}(\mathcal{S}) = \mathcal{P}(a_1) + \mathcal{P}(a_2) + \ldots + \mathcal{P}(a_{52}) = \sum_{i=1}^{52} 1/52 = 1$$

Proving things from the axioms

$\mathcal{P}(\mathcal{A}') = 1 - \mathcal{P}(\mathcal{A})$ for any event $\mathcal{A}$

Let $\mathcal{A}_1 = \mathcal{A}$ and $\mathcal{A}_2 = \mathcal{A}'$ then $\mathcal{A}_1 \cap \mathcal{A}_2 = \varnothing$ ($\mathcal{A}_1$ and $\mathcal{A}_2$ are disjoint) and $\mathcal{A}_1 \cup \mathcal{A}_2 = \mathcal{S}$. So

$$\mathcal{P}(\mathcal{A}_1) + \mathcal{P}(\mathcal{A}_2) = \mathcal{P}(\mathcal{A}_1 \cup \mathcal{A}_2) \qquad \text{(Axiom 3)}$$

$$= \mathcal{P}(\mathcal{S})$$
$$= 1 \qquad \text{(Axiom 2)}$$

So,

$$\mathcal{P}(\mathcal{A}) = \mathcal{P}(\mathcal{A}_1) = 1 - \mathcal{P}(\mathcal{A}_2)$$

Proving things from the axioms

$\mathcal{P}(\mathcal{A}') = 1 - \mathcal{P}(\mathcal{A})$ for any event $\mathcal{A}$

Example: We draw a card from a full deck of cards. What is the probability NOT to draw a “3”?

Solution: The event of interest

$$\mathcal{A} = \{a_1, a_2, \ldots, a_{48}\} = \{\diamondsuit 2, \clubsuit 2, \heartsuit 2, \spadesuit 2, \diamondsuit 4, \clubsuit 4, \ldots, \diamondsuit A, \clubsuit A\}$$

$$\mathcal{A}' = \{\diamondsuit 3, \clubsuit 3, \heartsuit 3, \spadesuit 3\}$$

$$\Rightarrow \mathcal{P}_r(\mathcal{A}) = 1 - \mathcal{P}_r(\mathcal{A}') = 1 - \frac{1}{13} = \frac{12}{13}.$$

Proving things from the axioms

$\mathcal{P}(\mathcal{A}) \leq 1$ for any event $\mathcal{A}$

For $\mathcal{P}(\mathcal{A}') = 1 - \mathcal{P}(\mathcal{A})$ (above Proposition), and $\mathcal{P}(\mathcal{A}') \geq 0$ (Axiom 1), so $1 - \mathcal{P}(\mathcal{A}) \geq 0$ or $\mathcal{P}(\mathcal{A}) \leq 1$.

Probabilities are always numbers between 0 and 1!

Quiz

Given that 9.1% of the German population has a tattoo, what is the probability that in a random sample of 3 Germans at least one person has a tattoo? (Source: Trampisch & Branda, Tattoos und Piercings in Deutschland)

(a) $1 - 0.091 = 0.909$

(b) $1 - 0.091^3 = 0.999246$

(c) $1 - (1 - 0.091)^3 \approx 0.248911$

(d) $(1 - 0.091)^3 \approx 0.751089$

(e) $0.091 \times 3 = 0.273$

Proving things from the axioms

$\mathcal{P}(\varnothing) = 0$

For $\varnothing = \mathcal{S}'$, $\mathcal{P}(\varnothing) = 1 - \mathcal{P}(\mathcal{S})$ (by the third Proposition) and $\mathcal{P}(\mathcal{S}) = 1$ (Axiom 2), so $\mathcal{P}(\varnothing) = 0$.

Proving things from the axioms

If $\mathcal{A} \subseteq \mathcal{B}$, then $\mathcal{P}(\mathcal{A}) \leq \mathcal{P}(\mathcal{B})$ (*subset equal*)

The notation $\mathcal{A} \subseteq \mathcal{B}$ means that $\mathcal{A}$ is contained in $\mathcal{B}$, i.e., every outcome in $\mathcal{A}$ also belongs to $\mathcal{B}$.

Take $\mathcal{A}_1 = \mathcal{A}$, $\mathcal{A}_2 = \mathcal{B} \setminus \mathcal{A}$ (i.e. $\mathcal{B}$ **minus** $\mathcal{A}$ **).**Then $\mathcal{A}_1 \cap \mathcal{A}_2 = \emptyset$ (since the elements of $\mathcal{B} \setminus \mathcal{A}$ are, by definition not in $\mathcal{A}$), and $\mathcal{A}_1 \cup \mathcal{A}_2 = \mathcal{B}$. So by Axiom 3,

$$\mathcal{P}(\mathcal{A}_1) + \mathcal{P}(\mathcal{A}_2) = \mathcal{P}(\mathcal{A}_1 \cup \mathcal{A}_2) = \mathcal{P}(\mathcal{B})$$

in other words: $\mathcal{P}(\mathcal{A}_1) + \mathcal{P}(\mathcal{B} \setminus \mathcal{A}) = \mathcal{P}(\mathcal{B})$

Now $\mathcal{P}(\mathcal{B} \setminus \mathcal{A}) \geq 0$ by Axiom 1, so

$$\mathcal{P}(\mathcal{A}) \leq \mathcal{P}(\mathcal{B})$$

General Addition Rule

If $\mathcal{A}$ and $\mathcal{B}$ are any two events, disjoint or not, then the probability that at least one of them will occur is

Out[ ]//TraditionalForm=

$$\boxed{\mathcal{P}_r(\mathcal{A} \text{ or } \mathcal{B}) = \mathcal{P}_r(\mathcal{A}) + \mathcal{P}_r(\mathcal{B}) - \mathcal{P}_r(\mathcal{A} \text{ and } \mathcal{B})}$$

where $\mathcal{P}_r(\mathcal{A}$ and $\mathcal{B})$ is the probability that both events will occur.

If $\mathcal{A}$ and $\mathcal{B}$ are disjoint events, then $\mathcal{P}_r(\mathcal{A}$ and $\mathcal{B})=0$ and the addition rule simplifies to

The addition rule

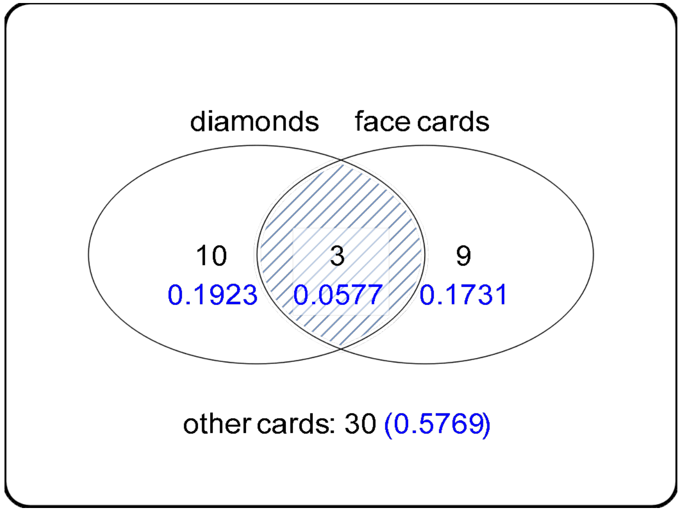
Out[ ]//TraditionalForm=

$$\boxed{\mathcal{P}_r(\mathcal{A} \text{ or } \mathcal{B}) = \mathcal{P}_r(\mathcal{A}) + \mathcal{P}_r(\mathcal{B})}$$

General Addition Rule

Example: Drawing from a full deck of cards, $\mathcal{P}_r(\Diamond)$ is the probability to draw a diamond card, $\mathcal{P}_r$(face card) is the probability to draw a face card. The following diagram visualizes the event space (it is called a **Venn diagram**):

Out[∘]=



There are 13 diamond ($\Diamond$) cards and 12 face cards.

$\mathcal{P}_r(\Diamond) = \frac{13}{52} = 0.25$, $\mathcal{P}_r(\text{face card}) = \frac{12}{52} = 0.230769$

General Addition Rule

Events ◊ and “ face card” are not disjoint! ◊J, ◊Q, ◊K fall in both categories. What is the probability to draw ◊ OR a face card? If this were disjoint events, we could use the addition rule

$\mathcal{P}_r$(◊ or face card)=$\mathcal{P}_r$(◊) + $\mathcal{P}_r$(face card)= $\frac{13}{52} + \frac{12}{52}$

$\mathcal{P}_r(\mathcal{A}$ or $\mathcal{B})$ = $\mathcal{P}_r$(◊ or face card)

= $\mathcal{P}_r$(◊) + $\mathcal{P}_r$(face card) **– $\mathcal{P}_r$(◊ and face card)**

= 13 / 52 + 12 / 52 – 3 / 52

= 22 / 52 = 11 / 26

3 cards that are in both events were counted twice, once in each probability. We must correct this double counting.

Quiz - Disjoint events

What is the probability that a randomly sampled student thinks marijuana should be legalized or they agree with their parents' political views?

	Share Parents' Politics		
Legalize MJ	No	Yes	Total
No	11	40	51
Yes	36	78	114
Total	47	118	165

(a) (40+36-78) / 165

(b) (114+118-78) / 165

(c) 78 / 165

(d) 78 / 188

(e) 11 / 47

Estimate probability from data

- In most cases we don’t know the probability of an event.
- Usually identifying equally likely cases is difficult
- When estimating a probability from data we are limited by the sample size (limited data)

Example : The probability of our latest observing run being clouded out is estimated by

number of cloudy nights last year / 365

1. issue: limited data; the same statistic over the last ten years will give a different, maybe more accurate result.
2. issue: identification of equally likely cases; What are equally clouded nights? How are these nights correlated with season? Are they independent of each other or are two cloudy nights in a row more likely?

After identifying the equally likely cases, which is **subjective!**, we can estimate probabilities.

Estimate probability from data

Example: Pick a prime number from a uniformly distributed set of random numbers between 1 and 10000.

In[ ]:=
```
Select[Range[100], PrimeQ]
Length@Select[Range[100], PrimeQ]
---------------------------------
              100
```

Out[ ]= {2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71, 73, 79, 83, 89, 97}

Out[ ]= $\frac{1}{4}$

In[ ]:=
```
theo = Length@Select[Range[10 000], PrimeQ] / 10 000
```

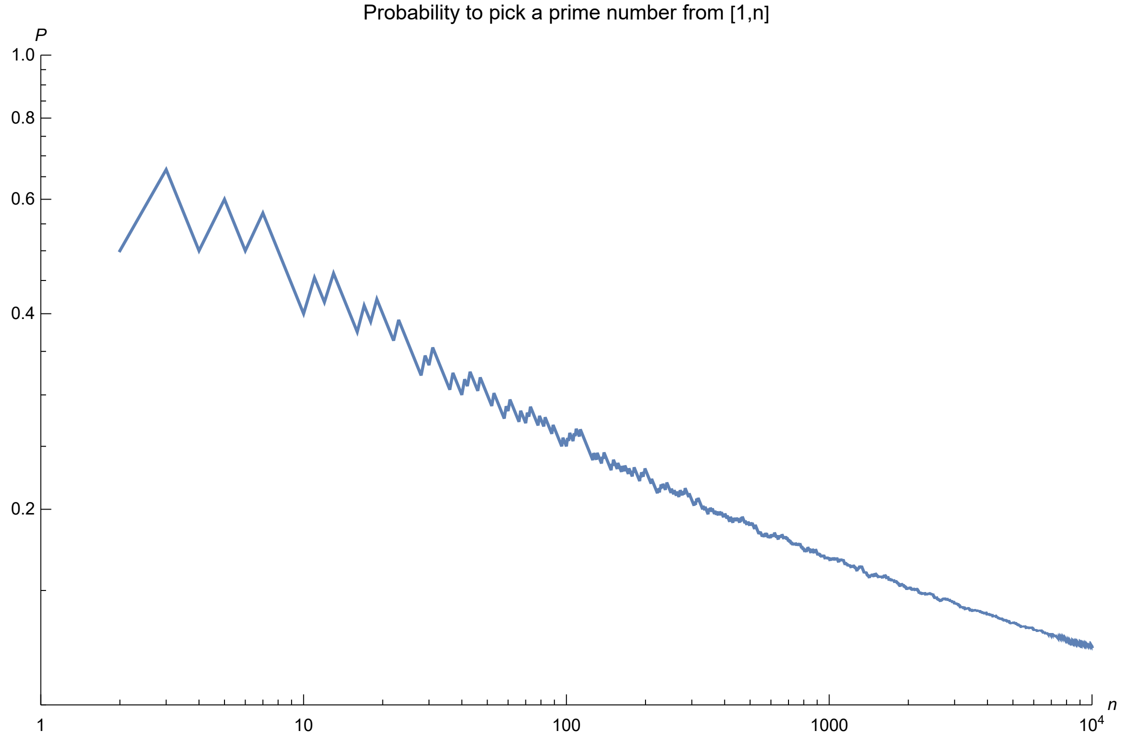
Out[ ]= $\frac{1229}{10\,000}$

Estimate probability from data

Example: Pick a prime number from a uniformly distributed set of random numbers between 1 and 10000.

In[ ]:=
```
set1 = Table[{n, 1/n Length@Select[Range[n], PrimeQ] // N}, {n, 2, 10000}];
ListLogLogPlot[set1, Joined → True, PlotRange → {{1, 10000}, {0.1, 1}},
 AxesLabel → {n, P}, PlotLabel → "Probability to pick a prime number from [1,n]", ImageSize → Large]
```

Out[ ]=



Estimate probability from data

Experiment: Pick a number from the range 1-10000 and check for prime. From the above estimation we expect about one out of ten picks to yield a prime. We repeat this 100 times and calculate the mean value of found prime numbers.

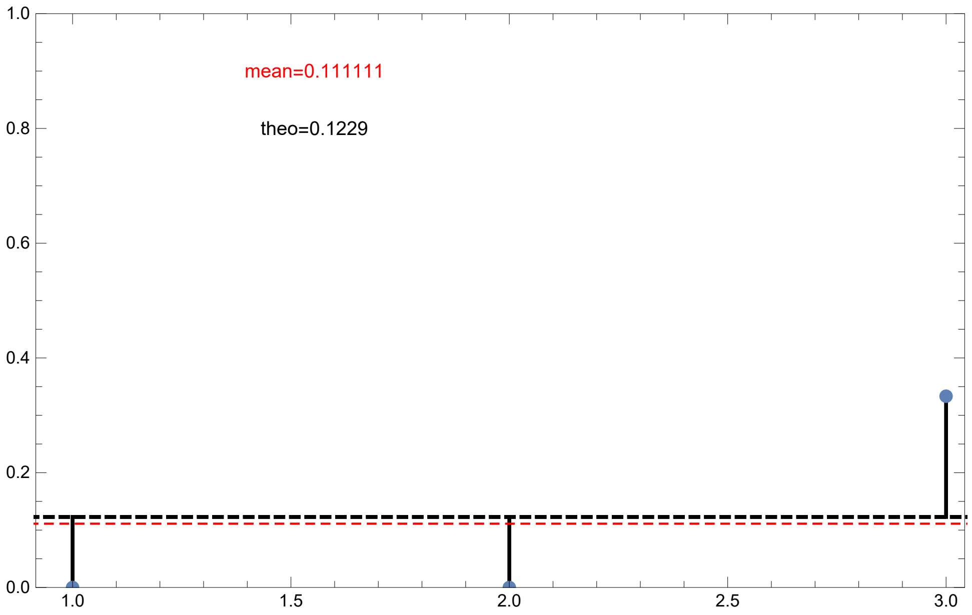
```
pickOne := RandomChoice[Range[10000]]
```

In[ ]:=
```
Mean[Table[Count[PrimeQ /@ Table[pickOne, {10}], True], {100}]]
```

Out[ ]= $\frac{11}{10}$

In[ ]:=
```
primeExperiment[3]
```

Out[ ]=

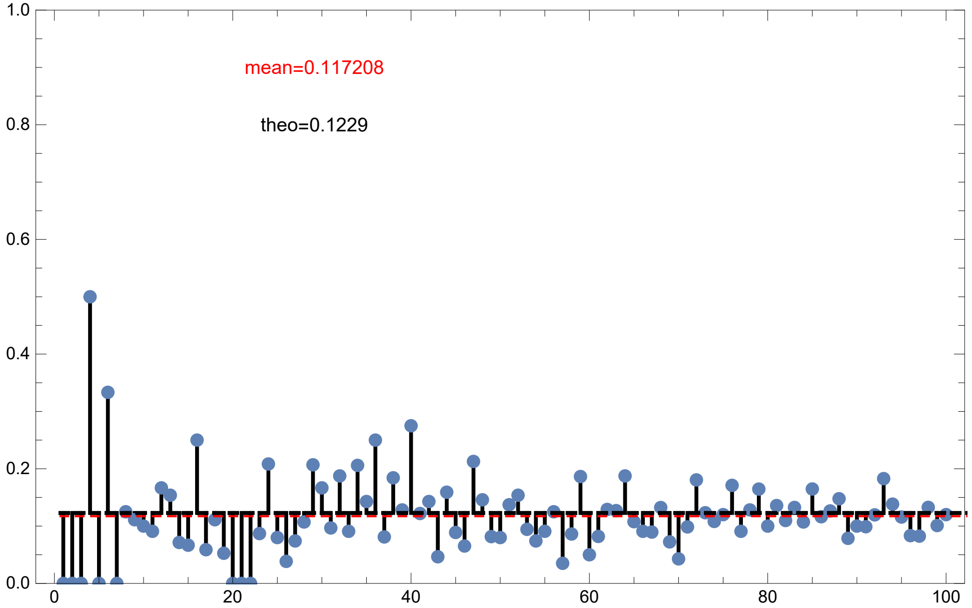


After randomly drawing 1,2 and 3 numbers from 1 to 10000 we got 0, 0 and 1 prime, i.e. a ratio of 0,0, and 1/3. The mean of all three results is 1/3*(0+0+1/3)=1/9=0.1111111.

Estimate probability from data

Experiment: Pick a number from the range 1-10000 and check for prime. From the above estimation we expect about one out of ten picks to yield a prime. We repeat this 100 times and calculate the mean value of found prime numbers.

In[∘]:=
```
primeExperiment[100]
```

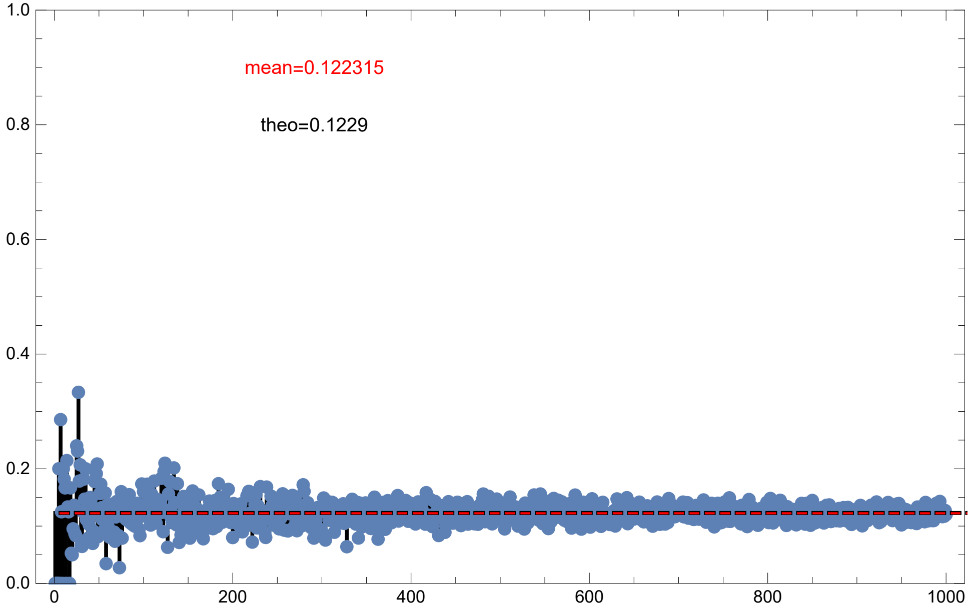
Out[∘]=


Estimate probability from data

Experiment: Pick a number from the range 1-10000 and check for prime. From the above estimation we expect about one out of ten picks to yield a prime. We repeat this 100 times and calculate the mean value of found prime numbers.

In[ ]:=
```
primeExperiment[1000]
```

Out[ ]=



Estimate probability from data

Example: SN1987

Before 1987, only 4 naked-eye supernovae had been recorded in 1000 years. If asked before 1987, what was the probability of a bright supernova happening in the 20th century?

1. A supernovae is not a random process, but a physically determined event. In principle it can be calculated when they are going to happen and the concept of probability is meaningless.
2. Given the past frequency we find 4/10 supernovae/century. (Assuming they are equally likely to be recorded during the last 10 centuries....)
3. We could do an a-priory assignment. Given a stellar mass function, stellar evolution as a function of mass, and star-formation rate plus some detection efficiency, we could calculate an expectation value of the supernovae in 1987. This would involve putting error bars around this number reflecting uncertainties in our assumptions.

Estimate probability from data

Example: SN1987

Suppose now we observe SN1987A. How does this sighting affect the probability of there being a supernova later in the 20th century?

1. It does not. One supernova does not influence the other.
2. Since the probability just reflects what we already know, we update the probability to 5/10.
3. Will probably need to revise its estimate of the probability, because SN1987A might lead to deeper understanding.

Probabilities reflect what we know! They are not properties of supernovae!

Code

Init