Data Analysis in Astronomy and Physics

Lecture 4 - 2: Conditional Probability

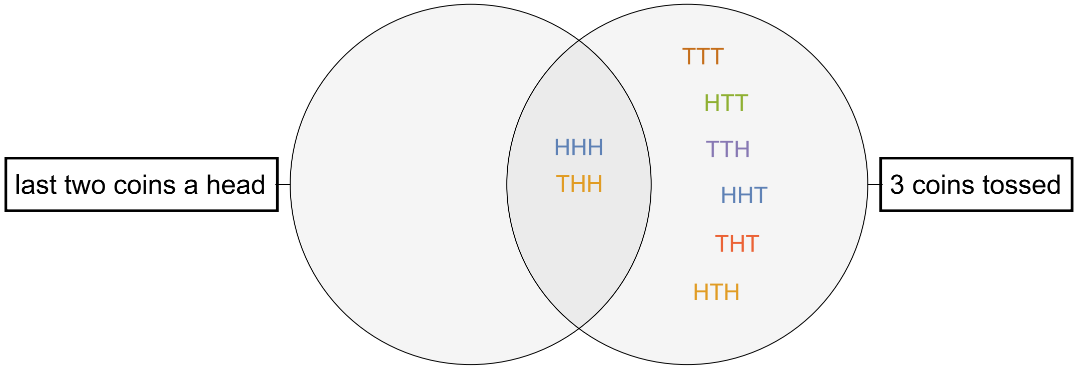
M. Röllig

Introduction

Imagine you toss a fair coin three times. What is the probability to get three "heads"? It is 1/9.

Now imagine you already tossed the coins and you know that the last two results were "head". What is the probability to get three "heads" now? It is 1/2.

Out[ ]=



Conditionality and independence

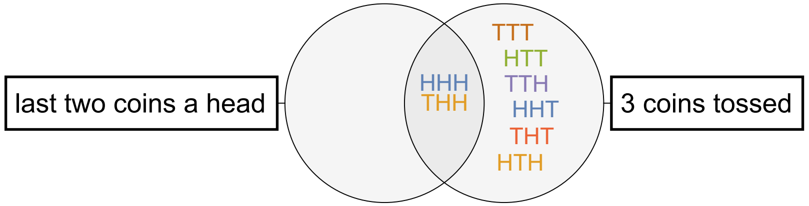
- Two events $\mathcal{A}$ and $\mathcal{B}$ are independent if the probability of one is unaffected by the other.
- It then follows: $\mathcal{P}(\mathcal{A} \text{ and } \mathcal{B}) = \mathcal{P}(\mathcal{A})\,\mathcal{P}(\mathcal{B})$

  Example: You toss a coin twice, what is the probability of getting two tails in a row?

  Answer: $\mathcal{P}$(two T in a row)=$\mathcal{P}$(T on the 1st toss)×$\mathcal{P}$(T on the 2nd toss)=1/2×1/2=1/4
- Conditional probability: the probability of $\mathcal{A}$, given that we know $\mathcal{B}$

$$\mathcal{P}(\mathcal{A} \mid \mathcal{B}) = \frac{\mathcal{P}(\mathcal{A} \text{ and } \mathcal{B})}{\mathcal{P}(\mathcal{B})}$$

Out[∘]=



Dividing by $\mathcal{P}(\mathcal{B})$ normalizes to updates sample space $\mathcal{B}$.

- If $\mathcal{A}$ and $\mathcal{B}$ are independent this reduces to $\mathcal{P}(\mathcal{A})$:

Conditionality and independence

- Several possibilities for $\mathcal{B}$:

$$\mathcal{P}(\mathcal{A}) = \sum_i \mathcal{P}(\mathcal{A} \mid \mathcal{B}_i)\,\mathcal{P}(\mathcal{B}_i)$$

- $\mathcal{A}$ might be a parameter of interest, while the $\mathcal{B}_i$ are not (maybe instrumental parameters). Knowing the $\mathcal{P}(\mathcal{B}_i)$ we can get rid of them by summation (integration). This is called *marginalization*.

Example: Conditional probability

An Urn contains 100 balls.

70 balls are made from wood, 30 balls are made from plastic.

25 wooden balls are red, 45 are green.

10 of the plastic balls are red, 20 are green.

We define the following events:

$\mathcal{A}$: the balls is made from wood

$\overline{\mathcal{A}}$: the ball is made from plastic

$\mathcal{B}$: the ball is red

$\overline{\mathcal{B}}$: the ball is green

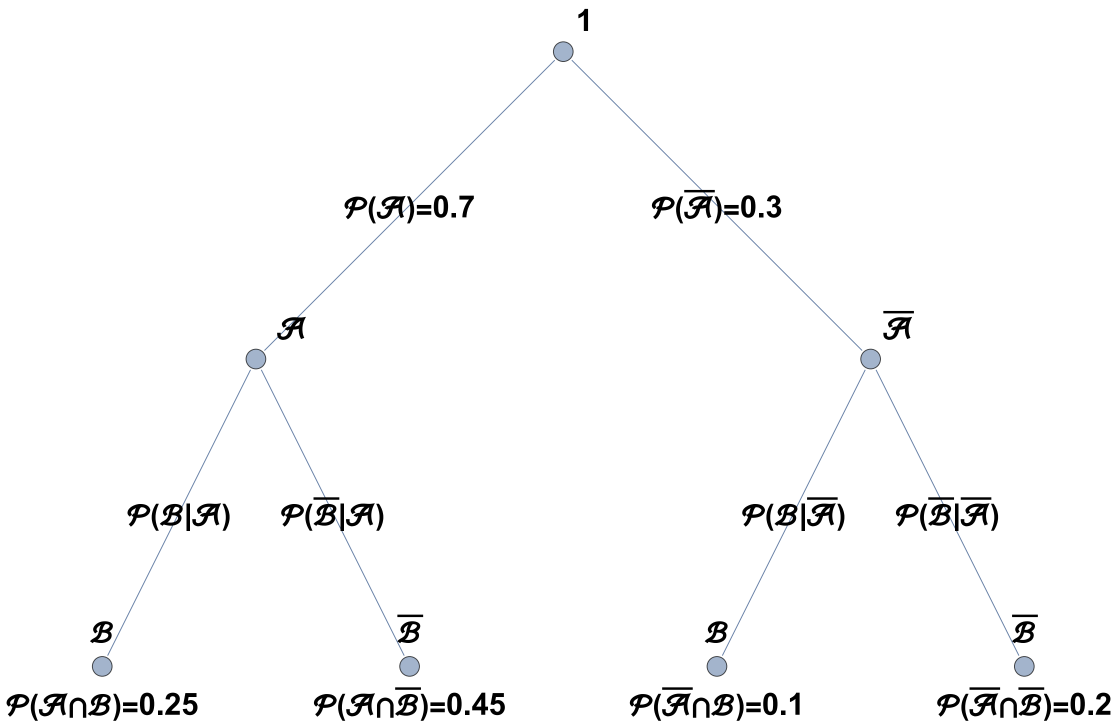
The probabilities are summarized in the following table:

	$\mathcal{B}$	$\overline{\mathcal{B}}$	sum
$\mathcal{A}$	$\mathcal{P}(\mathcal{A}\cap\mathcal{B}) = 0.25$	$\mathcal{P}(\mathcal{A}\cap\overline{\mathcal{B}}) = 0.45$	$\mathcal{P}(\mathcal{A}) = 0.7$
$\overline{\mathcal{A}}$	$\mathcal{P}(\overline{\mathcal{A}}\cap\mathcal{B}) = 0.1$	$\mathcal{P}(\overline{\mathcal{A}}\cap\overline{\mathcal{B}}) = 0.2$	$\mathcal{P}(\overline{\mathcal{A}}) = 0.3$
sum	$\mathcal{P}(\mathcal{B}) = 0.35$	$\mathcal{P}(\overline{\mathcal{B}}) = 0.65$	1

Example: Conditional probability

	$\mathcal{B}$	$\overline{\mathcal{B}}$	sum
$\mathcal{A}$	$\mathcal{P}(\mathcal{A}\cap\mathcal{B}) = 0.25$	$\mathcal{P}(\mathcal{A}\cap\overline{\mathcal{B}}) = 0.45$	$\mathcal{P}(\mathcal{A}) = 0.7$
$\overline{\mathcal{A}}$	$\mathcal{P}(\overline{\mathcal{A}}\cap\mathcal{B}) = 0.1$	$\mathcal{P}(\overline{\mathcal{A}}\cap\overline{\mathcal{B}}) = 0.2$	$\mathcal{P}(\overline{\mathcal{A}}) = 0.3$
sum	$\mathcal{P}(\mathcal{B}) = 0.35$	$\mathcal{P}(\overline{\mathcal{B}}) = 0.65$	1

Out[ ]=



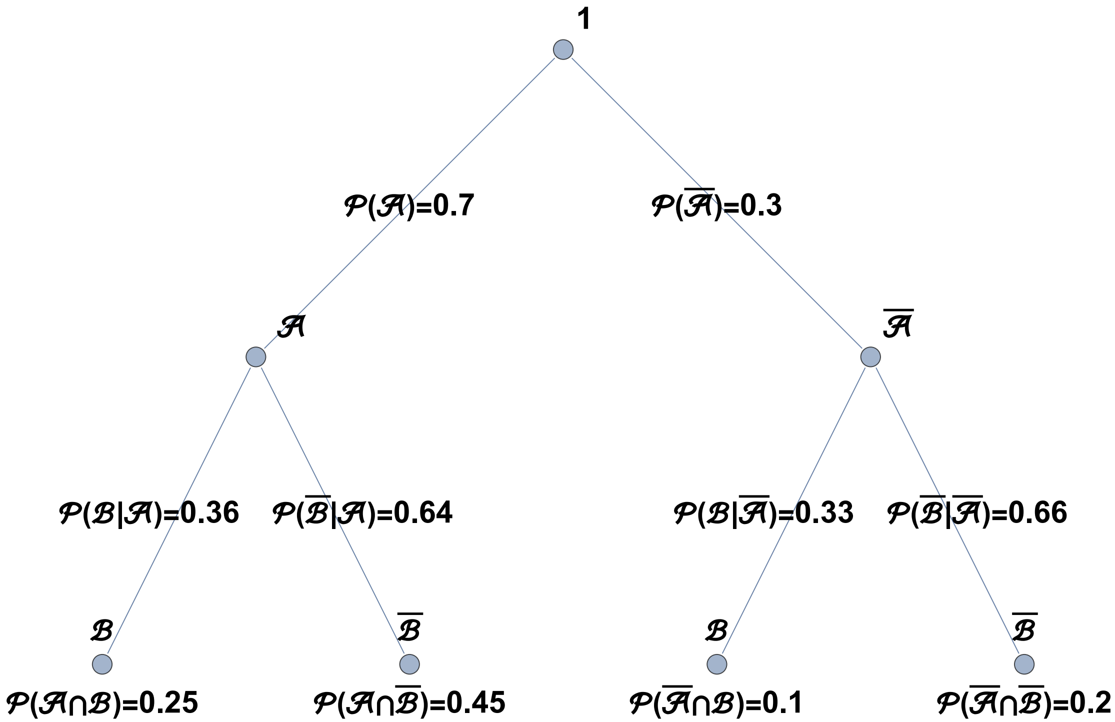
Example: Conditional probability

We know that $\mathcal{P}(\mathcal{B} \mid \mathcal{A}) = \frac{\mathcal{P}(\mathcal{A} \cap \mathcal{B})}{\mathcal{P}(\mathcal{A})}$, hence:

$$\mathcal{P}(\mathcal{B} \mid \mathcal{A}) = \frac{\mathcal{P}(\mathcal{A} \cap \mathcal{B})}{\mathcal{P}(\mathcal{A})} = \frac{5}{20} \Big/ \frac{7}{10} = \frac{5}{14} = 0.36 \qquad \mathcal{P}(\overline{\mathcal{B}} \mid \mathcal{A}) = \frac{\mathcal{P}(\mathcal{A} \cap \overline{\mathcal{B}})}{\mathcal{P}(\mathcal{A})} = \frac{9}{20} \Big/ \frac{7}{10} = \frac{9}{14} = 0.64$$

$$\mathcal{P}(\mathcal{B} \mid \overline{\mathcal{A}}) = \frac{\mathcal{P}(\overline{\mathcal{A}} \cap \mathcal{B})}{\mathcal{P}(\overline{\mathcal{A}})} = \frac{2}{20} \Big/ \frac{3}{10} = \frac{1}{3} = 0.33 \qquad \mathcal{P}(\overline{\mathcal{B}} \mid \overline{\mathcal{A}}) = \frac{\mathcal{P}(\overline{\mathcal{A}} \cap \overline{\mathcal{B}})}{\mathcal{P}(\overline{\mathcal{A}})} = \frac{4}{20} \Big/ \frac{3}{10} = \frac{2}{3} = 0.66$$

Out[ ]=



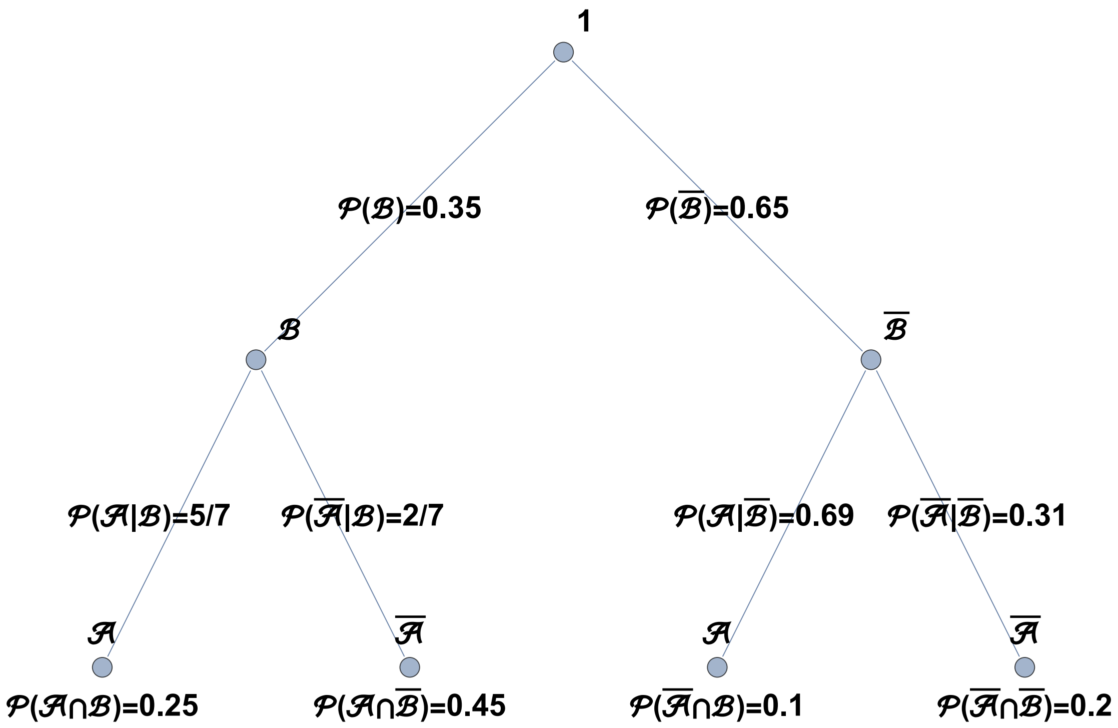
Example : Conditional probability

Now imagine that you draw a red ball (event $\mathcal{B}$). What is the probability that it is made from wood? We are looking for $\mathcal{P}(\mathcal{A}|\mathcal{B})$. How does it relate to $\mathcal{P}(\mathcal{B}|\mathcal{A})$? We can write:

$$\mathcal{P}(\mathcal{A} \mid \mathcal{B}) = \frac{\mathcal{P}(\mathcal{B} \cap \mathcal{A})}{\mathcal{P}(\mathcal{B})} = \frac{\mathcal{P}(\mathcal{B} \cap \mathcal{A})}{\mathcal{P}(\mathcal{B} \cap \mathcal{A})+\mathcal{P}(\mathcal{B} \cap \overline{\mathcal{A}})} = \frac{5/20}{5/20+0.1} = 5/7 \qquad \mathcal{P}(\overline{\mathcal{A}} \mid \mathcal{B}) = \frac{\mathcal{P}(\mathcal{B} \cap \overline{\mathcal{A}})}{\mathcal{P}(\mathcal{B})} = \frac{2/20}{5/20+0.1} = 2/7$$

$$\mathcal{P}(\mathcal{A} \mid \overline{\mathcal{B}}) = \frac{\mathcal{P}(\overline{\mathcal{B}} \cap \mathcal{A})}{\mathcal{P}(\overline{\mathcal{B}})} = \frac{\mathcal{P}(\overline{\mathcal{B}} \cap \mathcal{A})}{\mathcal{P}(\overline{\mathcal{B}} \cap \mathcal{A})+\mathcal{P}(\overline{\mathcal{B}} \cap \overline{\mathcal{A}})} = \frac{0.45}{0.45+0.2} = 0.69 \qquad \mathcal{P}(\overline{\mathcal{A}} \mid \overline{\mathcal{B}}) = \frac{\mathcal{P}(\overline{\mathcal{B}} \cap \overline{\mathcal{A}})}{\mathcal{P}(\overline{\mathcal{B}})} = \frac{0.2}{0.45+0.2} = 0.31$$

Out[ ]=



We calculated the inverse probability, also called *Bayes probability*.

Bayes' Theorem

The following is trivial:

Out[ ]//TraditionalForm=

$$\boxed{\mathcal{P}(\mathcal{A} \wedge \mathcal{B}) = \mathcal{P}(\mathcal{B} \wedge \mathcal{A})}$$

it follows

Out[ ]//TraditionalForm=

$$\boxed{\mathcal{P}(\mathcal{A} \mid \mathcal{B})\, \mathcal{P}(\mathcal{B}) = \mathcal{P}(\mathcal{B} \mid \mathcal{A})\, \mathcal{P}(\mathcal{A})}$$

this gives the **Bayes theorem**

Out[ ]//TraditionalForm=

$$\boxed{\mathcal{P}(\mathcal{B} \mid \mathcal{A}) = \frac{\mathcal{P}(\mathcal{A} \mid \mathcal{B})\, \mathcal{P}(\mathcal{B})}{\mathcal{P}(\mathcal{A})}}$$

Bayes' Theorem

Out[ ]//TraditionalForm=

$$\boxed{\mathcal{P}(\mathcal{B} \mid \mathcal{A}) = \frac{\mathcal{P}(\mathcal{A} \mid \mathcal{B})\, \mathcal{P}(\mathcal{B})}{\mathcal{P}(\mathcal{A})}}$$

- denominator $\mathcal{P}(\mathcal{A})$ is a **normalizing factor**
- useful interpretation: **the data, event $\mathcal{A}$,** are regarded as succeeding $\mathcal{B}$, the state of belief preceding the experiment.
- $\mathcal{P}(\mathcal{B})$ is the ***prior probability*** which will be altered by experience
- this experience is expressed by the ***likelihood*** $\mathcal{P}(\mathcal{A} \mid \mathcal{B})$
- the ***posterior probability*** $\mathcal{P}(\mathcal{B} \mid \mathcal{A})$ is the state of belief after the data has been analyzed

Bayes’ Theorem

We can reformulate the previous slides:

Suppose we have a set of events $\mathcal{H}_1, \mathcal{H}_2, \ldots, \mathcal{H}_m$ that are pairwise disjoint and such that the sample space $\mathcal{S}$ satisfies the equation

Out[ ]//TraditionalForm=

$$\mathcal{S} = \mathcal{H}_1 \cup \mathcal{H}_1 \cup \cdots \cup \mathcal{H}_m$$

Let's call these events *hypotheses.* We also have an event $\mathcal{E}$ that gives us some information about which hypothesis is correct. We call $\mathcal{E}$ *evidence*.

Before we receive the evidence, we have a set of *prior probabilities* $\mathcal{P}(\mathcal{H}_1), \mathcal{P}(\mathcal{H}_2), \ldots, \mathcal{P}(\mathcal{H}_m)$. That is, we know $\mathcal{P}(\mathcal{E} \mid \mathcal{H}_i)$ for all i. We want to find the probabilities for the hypotheses given the evidence, i.e. we want to find the conditional probabilities $\mathcal{P}(\mathcal{H}_i \mid \mathcal{E})$. That probability is called the *posterior probability.*

Out[ ]//TraditionalForm=

$$\boxed{\mathcal{P}(\mathcal{H}_i \mid \mathcal{E}) = \frac{\mathcal{P}(\mathcal{E} \mid \mathcal{H}_i)\,\mathcal{P}(\mathcal{H}_i)}{\sum_{k=1}^{m} \mathcal{P}(\mathcal{E} \mid \mathcal{H}_k)\,\mathcal{P}(\mathcal{H}_k)}}$$

Example: drawing coloured balls from an urn

If there are ***R* red balls** and ***W* white balls** in the urn, the probability of drawing three red balls and two white balls with replacement is:

Out[ ]//TraditionalForm=

$$\mathcal{P}(3\,W \wedge 2\,R) = \left(\frac{W}{W+R}\right)^3 \left(\frac{R}{W+R}\right)^2$$

However, this is usually not the problem a researcher faces. Usually we have the inverse problem, e.g. we drew *R* red balls and *W* white balls from an urn with unknown contents. (**Problem of inverse probability**) A similar, more astronomical, problem would be to draw a sample of objects from a sky survey and infer from this sample how these objects are distributed throughout the Universe.

Example: Bayes’ interpretation

Out[ ]//TraditionalForm=

$$\mathcal{P}(\text{contents of urn} \mid \text{data}) \propto \mathcal{P}(\text{data} \mid \text{contents of urn})$$

Given some assumptions, we can calculate the right-hand side.

Take *R* red balls and *W* white balls in an urn, with a total of $R + W = 10$. We draw $T = 3$ times with replacement and get $r = 2$ red balls. How many red balls are in the urn? The probability of drawing a red ball is

Out[ ]//TraditionalForm=

$$\left(\frac{R}{W+R}\right)$$

The probability of getting r red balls, the *likelihood*, is

Out[ ]//TraditionalForm=

$$\binom{T}{r}\left(\frac{R}{W+R}\right)^{r}\left(\frac{W}{W+R}\right)^{T-r}$$

This is the number of permutations of **r red balls amongst T draws** multiplied by the probability that **r balls will be red** and **T-r balls will not be red**. This is $P(\text{data} \mid \text{model})$ of the right-hand side of the Bayes’ theorem. We still need the *prior* $P(\text{model})$.

Example: Bayes' interpretation

We assume that R is the only uncertainty of the model. We start by assuming it is uniformly likely between 0 and $R + W$. This is just a constant factor of $1/(R + W)$. So our posterior probability is

Out[•]//TraditionalForm=

$$\mathcal{P}(R \mid \text{data}) \propto \binom{T}{r}\left(\frac{R}{W+R}\right)^{r}\left(\frac{W}{W+R}\right)^{T-r} \times \text{prior on R}$$

We still need to normalize the posterior probability:

Out[•]//TraditionalForm=

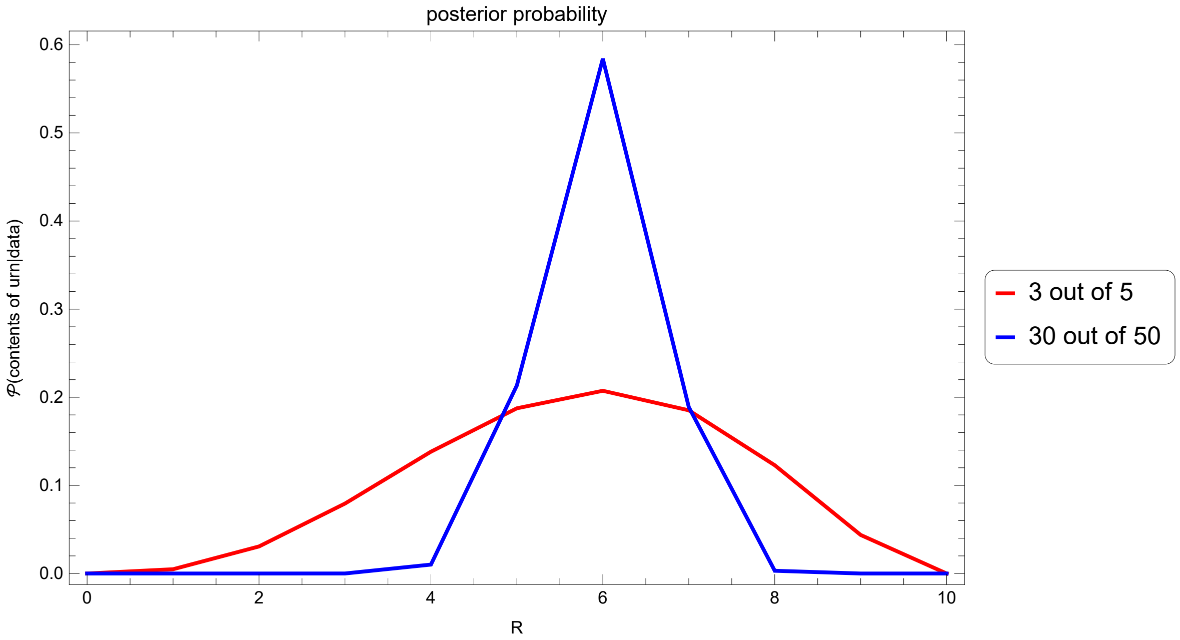
$$\int_0^{10} \mathcal{P}(R \mid \text{data})\, dR \propto \int_0^{10} \binom{T}{r}\left(\frac{R}{W+R}\right)^{r}\left(\frac{W}{W+R}\right)^{T-r} dR$$

Example: Bayes' interpretation

In[ ]:=
```
prob[{r_, T_}, {R_, tot_}] := Binomial[T, r] (R/tot)^r ((tot - R)/tot)^(T-r)
Integrate[prob[{r, T}, {R, 0, tot}], {R, 0, tot}, Assumptions → T ≥ r > 0 && tot > 0]
```

Out[ ]= $\frac{\text{tot}}{1 + T}$

With this we can compute the posterior probability that the urn contains *R* red balls (with 10 ball total) after drawing draw *r* red balls in *T* draws.

Out[ ]=


Example: Clinical tests

Consider the clinical test:

Suppose that 1 in 1000 of the population is a carrier of the disease.

Suppose also that the probability that a carrier tests negative is 1%, while the probability that a non-carrier tests positive is 5%. (A test achieving these values would be regarded as very successful.)

Let $\mathcal{A}$ be the event ‘the patient is a carrier’, and $\mathcal{B}$ the event ‘the test result is positive’. We are given that $\mathcal{P}(\mathcal{A}) = 0.001$ (so that $\mathcal{P}(\mathcal{A}') = 0.999$), and that

$$\mathcal{P}(\mathcal{B} \mid \mathcal{A}) = 0.99, \qquad \mathcal{P}(\mathcal{B} \mid \mathcal{A}') = 0.05$$

- A patient has just had a **positive test result**. What is the **probability that the patient is a carrier**?

$$\begin{aligned}\mathcal{P}(\mathcal{A} \mid \mathcal{B}) &= \frac{\mathcal{P}(\mathcal{B} \mid \mathcal{A})\,\mathcal{P}(\mathcal{A})}{\mathcal{P}(\mathcal{B} \mid \mathcal{A})\,\mathcal{P}(\mathcal{A}) + \mathcal{P}(\mathcal{B} \mid \mathcal{A}')\,\mathcal{P}(\mathcal{A}')} \\ &= \frac{0.99\times 0.001}{(0.99\times 0.001)+(0.05\times 0.999)} \\ &= \frac{0.00099}{0.05094} = 0.0194\end{aligned}$$

Example: Clinical tests

- A patient has just had a **negative test result**. What is the **probability that the patient is a carrier**?

$$
\begin{aligned}
\mathcal{P}(\mathcal{A} \mid \mathcal{B}') &= \frac{\mathcal{P}(\mathcal{B}' \mid \mathcal{A})\,\mathcal{P}(\mathcal{A})}{\mathcal{P}(\mathcal{B}' \mid \mathcal{A})\,\mathcal{P}(\mathcal{A})+\mathcal{P}(\mathcal{B}' \mid \mathcal{A}')\,\mathcal{P}(\mathcal{A}')} \\
&= \frac{0.01\times 0.001}{(0.01\times 0.001)+(0.95\times 0.999)} \\
&= \frac{0.00001}{0.94095} = 0.00001
\end{aligned}
$$

So a patient with a negative test result can be reassured; but a patient with a positive test result still has less than 2% chance of being a carrier, so is likely to worry unnecessarily.

Example: SN1987 revisited again

Return to the question of supernova rate per century ρ. How do we estimate this?

$\mathcal{P}(\text{data}\|\rho)$:	**likelihood**
	Probability to observe data given that the SN rate is ρ
$\mathcal{P}(\rho)$:	**prior**
	Probability distribution for ρ
$\mathcal{P}(\rho\|\text{data})$:	**posterior probability**
	Probability for the SN rate to be ρ, given that we observed data
	Our state of belief after we analyzed data → **wanted result of our analysis**
$\mathcal{P}(\text{data})$:	**evidence**
	accumulated within the normalizing factor

$$\mathcal{P}(\rho|\text{data})\,\mathcal{P}(\text{data}) = \mathcal{P}(\text{data}|\rho)\,\mathcal{P}(\rho) \qquad \Rightarrow \qquad \mathcal{P}(\rho|\text{data}) = \frac{\mathcal{P}(\text{data}\,|\,\rho)\,\mathcal{P}(\rho)}{\mathcal{P}(\text{data})}$$

Example: SN1987 revisited again

Suitable **model for $\mathcal{P}$(data | ρ) is the binomial distribution**, because in any century we either get a supernova or we do not (neglecting multiple SNs).

$\mathcal{P}_r$(4 out of 10 centuries with SN) = $\rho^4(1-\rho)^6$ (ρ: probability for a supernova)

$\mathcal{P}$(data | ρ) = $\mathcal{P}_r$(4 out of 10 centuries with SN)×(number of all possible combinations) = $\binom{10}{4}\rho^4(1-\rho)^6$

Our **posterior probability** is then

Out[ ]//TraditionalForm=

$$\mathcal{P}(\rho\,|\,\text{data}) \propto \binom{10}{4}\rho^4(1-\rho)^6 \times \text{prior on } \rho$$

We assume the prior to be uniform in the range of 0 to 1 (minimum knowledge, without any preconceptions about ρ), i.e. $\mathcal{P}(\rho) = \frac{1}{\text{max}-\text{min}} = 1/(1-0) = 1$.

Then to normalize the posterior probability we need:

Out[ ]//TraditionalForm=

$$\int_0^1 \mathcal{P}(\text{data}\,|\,\rho)\,\mathcal{P}(\rho)\,d\rho = 1$$

Example: SN1987 revisited again

This gives us the normalizing constant

Out[ ]//TraditionalForm=

$$\int_0^1 \mathcal{P}(\text{data} \mid \rho)\, \mathcal{P}(\rho)\, d\rho = \int_0^1 \binom{10}{4} \rho^4 (1-\rho)^6\, d\rho = \binom{10}{4} \frac{\Gamma(7)\,\Gamma(5)}{\Gamma(12)} = \binom{10}{4} \mathrm{B}(5,\ 7)$$

where Γ is the Gamma function (Γ(n)=(n-1)! for integer n) and B is the beta function. In general, for *n* supernovae in *m* centuries, the distribution is:

Out[ ]//TraditionalForm=

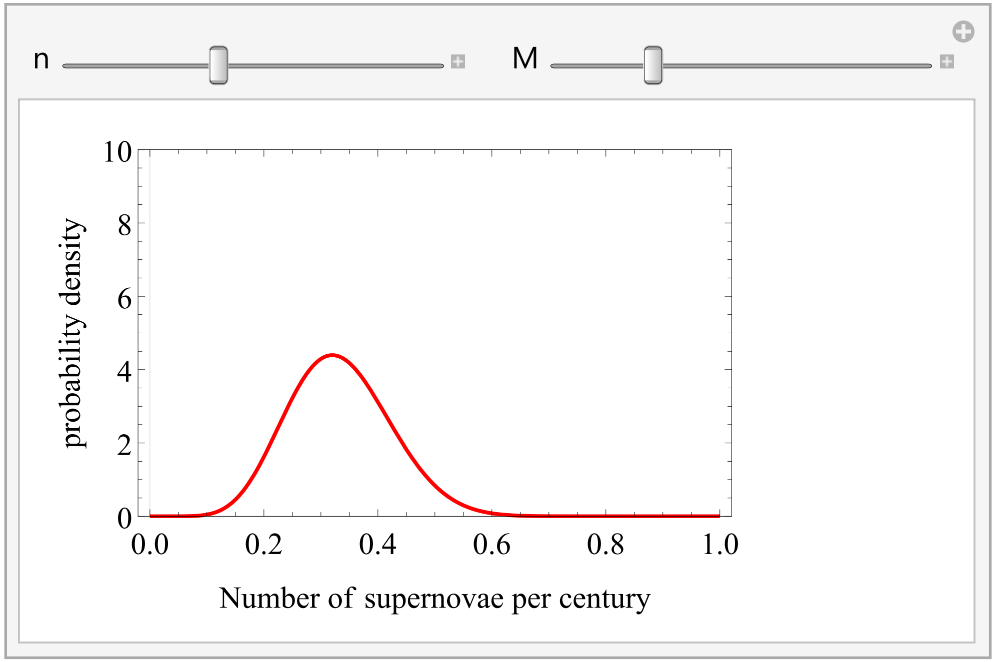
$$\boxed{\mathcal{P}(\rho \mid \text{data}) = \frac{\rho^n (1-\rho)^{m-n}}{\mathrm{B}(n+1,\ m-n+1)}}$$

Example: SN1987 revisited again

Out[ ]//TraditionalForm=

$$\mathcal{P}(\rho \mid \text{data}) = \frac{\rho^n (1-\rho)^{m-n}}{\mathrm{B}(n+1,\, m-n+1)}$$

Out[ ]=



prior = “knowledge”

Assume we establish our posterior distribution at the end of the 19th century, so that it is

Out[ ]//TraditionalForm=

$$\frac{\rho^4 (1-\rho)^6}{\mathrm{B}(5,\ 7)}$$

as shown earlier. At this stage, our data are 4 supernovae in 10 centuries.

Reviewing the situation at the end of the 20th century, **we take this as our prior**. Available new data consist of one additional supernova, so that the likelihood is simply the probability of observing exactly one event of probability ρ, namely ρ. The updated posterior distribution is

Out[ ]//TraditionalForm=

$$\mathcal{P}(\rho \mid \text{data}) = \frac{\binom{10}{5} \rho^5 (1-\rho)^5 \left(\frac{\rho^4 (1-\rho)^6}{\text{Beta}[5,7]}\right)}{\int_0^1 \binom{10}{5} \rho^5 (1-\rho)^5 \left(\frac{\rho^4 (1-\rho)^6}{\text{Beta}[5,7]}\right) d\rho} = \frac{\rho^5 (1-\rho)^6}{\mathrm{B}(6,\ 7)}$$

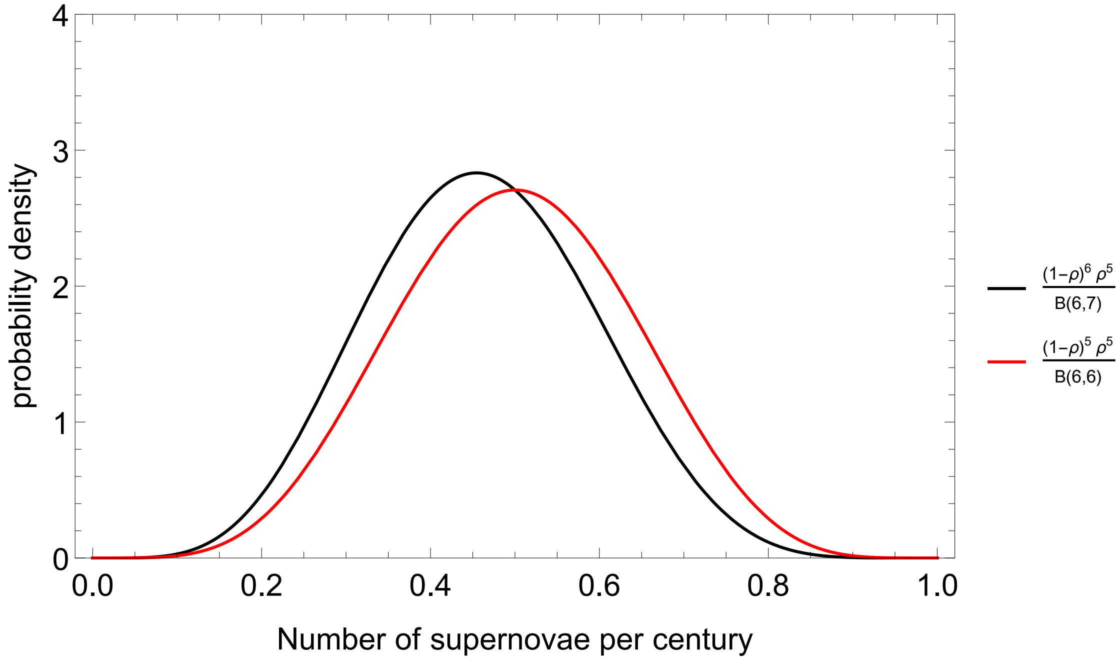
prior = "knowledge"

Out[∘]//TraditionalForm=

$$\mathcal{P}(\rho \mid \text{data}) = \frac{\binom{10}{5} \rho^5 (1-\rho)^5 \left(\frac{\rho^4 (1-\rho)^6}{\text{Beta}[5,7]}\right)}{\int_0^1 \binom{10}{5} \rho^5 (1-\rho)^5 \left(\frac{\rho^4 (1-\rho)^6}{\text{Beta}[5,7]}\right) d\rho} = \frac{\rho^5 (1-\rho)^6}{\text{B}(6,\ 7)}$$

This is not the same as calculating the posterior distribution at the end of the 20th century with a uniform prior, giving $\frac{(1-\rho)^5 \rho^5}{\text{Beta}[6,6]}$:

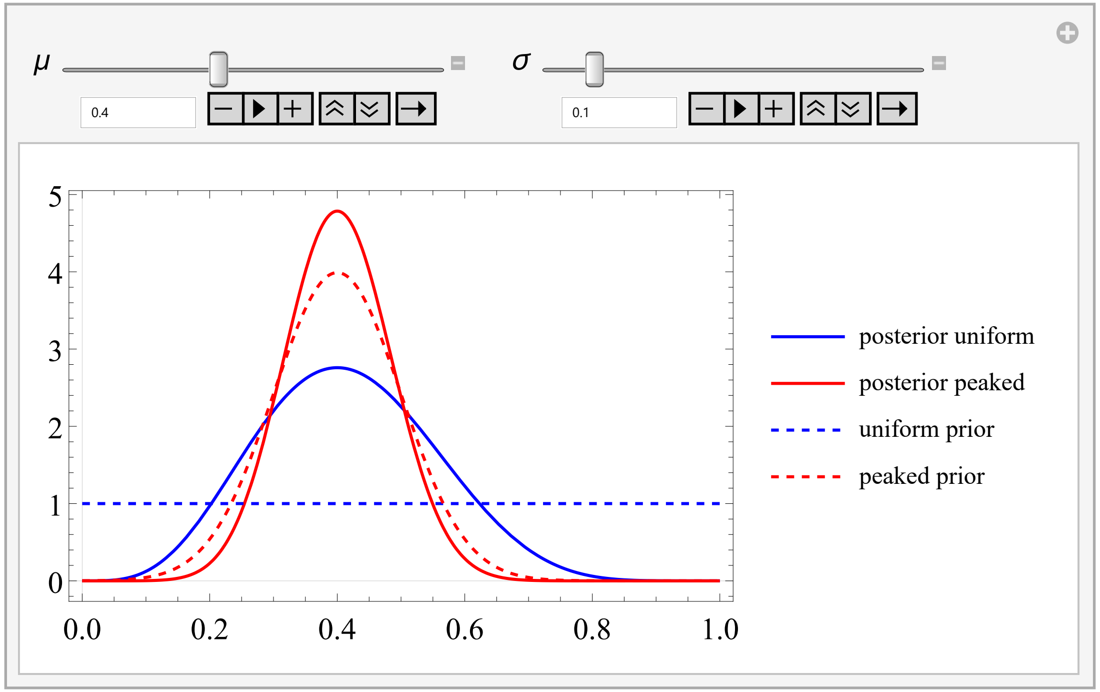
Out[∘]=



Choice of prior?

On common criticism of the Bayesian approach is the subjective choice of priors. In our supernova example, assume we chose a peaked prior instead of a uniform. The posterior probability is now more peaked.

Out[ ]=



Different priors

In many cases, a uniform prior is not the best choice, for example if we plan to measure a physical effect. Both Jeffreys (1961) and Jaynes (1986) discuss other possibilities:

The Jeffrey's prior:

Out[∘]//TraditionalForm=

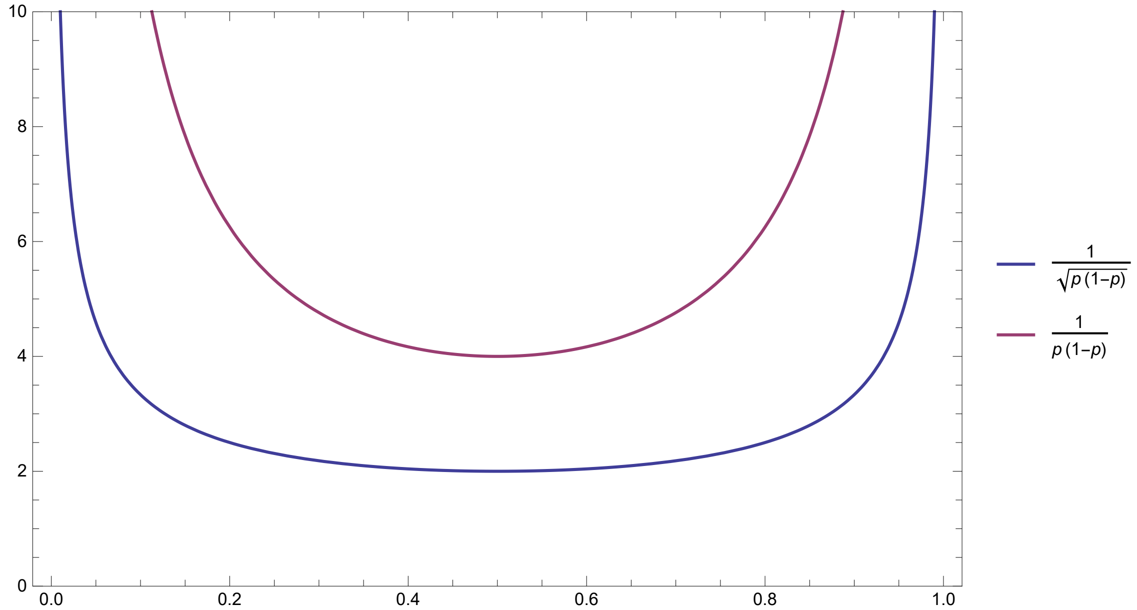
$$\mathcal{P}(\rho) = \frac{1}{\sqrt{\rho\,(1-\rho)}}$$

and the Haldane's prior:

Out[∘]//TraditionalForm=

$$\mathcal{P}(\rho) = \frac{1}{\rho\,(1-\rho)}$$

Different priors



These are intended to reflect the fact that in most experiments we are expecting a yes or no answer.

Jeffreys H., 1961, *Theory of Probability.* Clarendon Press

Jaynes, Edwin T. (Sep 1968). "Prior Probabilities" (PDF). IEEE Transactions on Systems Science and Cybernetics 4 (3): 227–241. doi:10.1109/TSSC.1968.300117.

Example:

Suppose we make an observation with a radio telescope at a randomly selected position in the sky. Our model of the data (an event labelled $\mathcal{D}$, consisting of the single measured flux density *f*) is that it is distributed in a Gaussian way about the true flux density S with a variance σ^2. The extensive body of radio source counts also tells us the a-priori distribution of S; for the purposes of this example, we approximate this information by the simple prior

Out[ ]//TraditionalForm=

$$\mathcal{P}(S) = K\, S^{-5/2}$$

describing our **prior** state of knowledge. (*K* normalizes the counts to unity; there is presumed to be one source in the beam at some flux density level.) The probability of observing *f* when the true value is *S* we take to be

Out[ ]//TraditionalForm=

$$\exp\left(-\frac{1}{2\,\sigma^2}\,(f - S)^2\right)$$

Example:

Bayes' theorem then tells us

Out[ ]//TraditionalForm=

$$\mathcal{P}r(S \mid \mathcal{D}) = K' \exp\left(-\frac{1}{2\,\sigma^2}\,(f - S)^2\right) S^{-5/2}$$

with the normalization K'. For n independent flux measurements f_i, the result would be

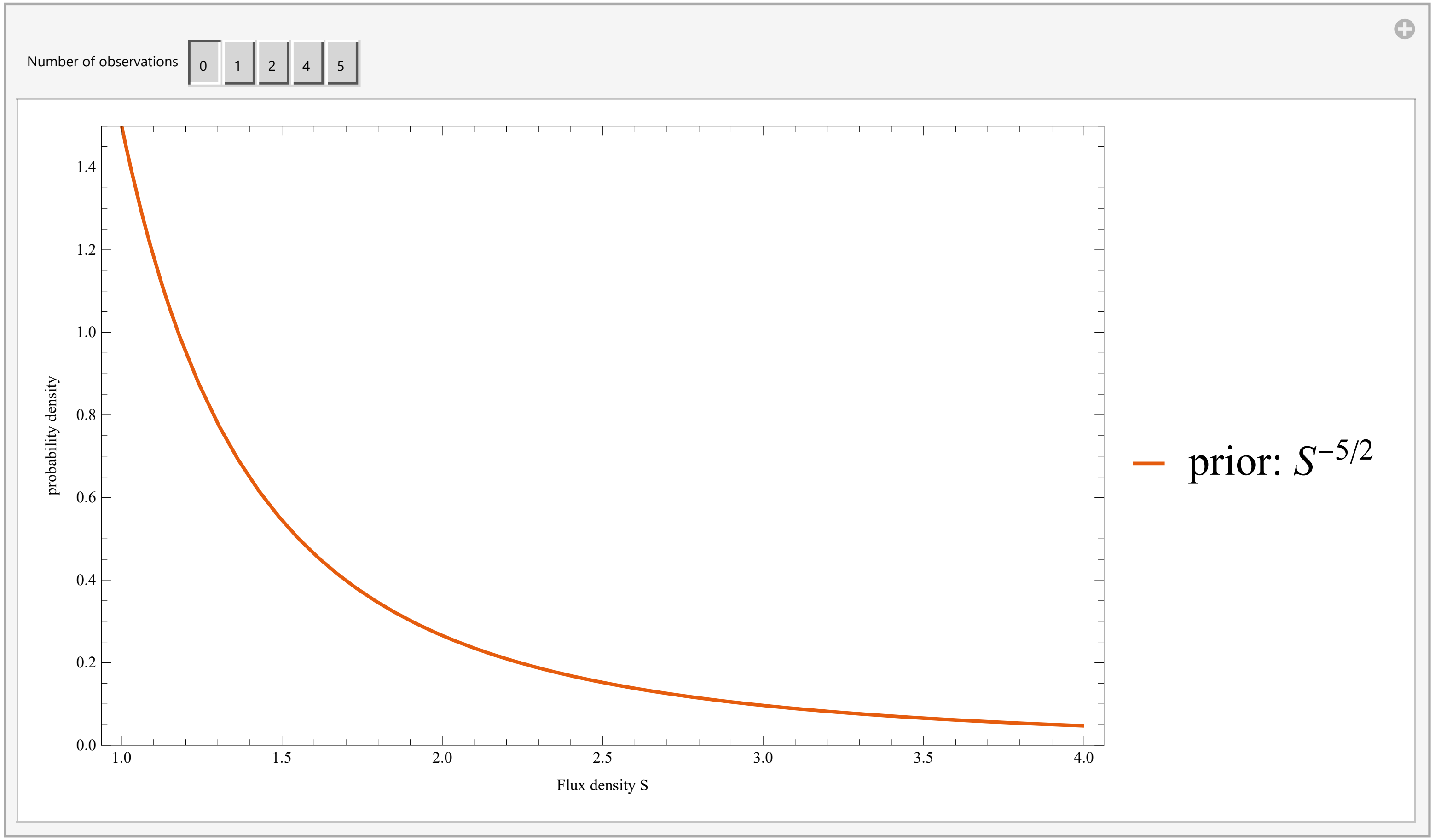
Out[ ]//TraditionalForm=

$$\mathcal{P}r(S \mid \mathcal{D}) = K'' \exp\left(-\frac{1}{2\,\sigma^2}\sum_{i=1}^{n}(f_i - S)^2\right) S^{-5/2}$$

Example:

To compute the normalization we assume that source counts extend between 1 and 100 units, the noise level σ=1. The data were 2, 1.3, 3, 1.5, 2, 1.8. The following sequence of figures demonstrates how successive addition of measurements gradually overcomes the initial prior, which dominates any conclusion for few measurements.

Out[ ]=



Laplace’s rule of succession

Suppose that every time there is an opportunity for an event to happen, then it occurs with unknown probability *p*. Laplace's law of succession states that, **if before we observed any events we thought all values of *p* were equally likely, then after observing *r* events out of *n* opportunities a good estimate of *p* is**:

Out[ ]//TraditionalForm=

$$\hat{p} = \frac{(r+1)}{(n+2)}$$

Example: before we start an observation, i.e. $(r = 0, n = 0) \Rightarrow \hat{p} = 1/2$, both outcomes equally likely. Suppose we observe the event at the first opportunity: $(r = 1, n = 1) \Rightarrow \hat{p} = 2/3$, ... (see sunrise problem)

When the event keeps on occurring

Suppose we have observed n events in n opportunities. Then the probability of this sequence is p^n. We need to turn this into a distribution for p that represents reasonable belief: this requires **Bayes theorem** but, if we believe all values of p were equally plausible before making any observations, the distribution for p is just proportional to p^n. Probability distributions must integrate up to 1, and

```
Integrate[p^n, {p, 0, 1}]
```

```
ConditionalExpression[1/(1 + n), Re[n] > -1]
```

Out[•]//TraditionalForm=

$$\int_0^1 p^n \, d\, p = \frac{1}{(n+1)}$$

and so the full probability distribution for p must be of the form

Out[•]//TraditionalForm=

$$f(p) = (n+1)\, p^n$$

Example: Sunrise problem

The sun rose n-times in a row. What is the probability that it will also rise tomorrow?

We assume that the a priori probability for the sunrise is a constant that we don’t know. Because of our total lack of knowledge we assume that all possible values ξ are equally probably → uniform prob. distribution. ξ is a probability, but it also has a probability density $f(p) = 1$ for $0 \leq p \leq 1$.

Out[ ]//TraditionalForm=

$$\mathcal{P}_{\text{pr}}\,(p \leq \xi \leq p + \text{dp}) = \text{dp}\,, \qquad 0\leq\text{p}\leq 1$$

If the true value of ξ is p, then the probability of n sunrises in a row is p^n. The event $\mathcal{S}^n$ of “The sun rises n-times.” has the probability:

Out[ ]//TraditionalForm=

$$\mathcal{P}_{\text{pr}}\,(\mathcal{S}^n \mid \xi = p) = p^n$$

Example: Sunrise problem

Out[ ]//TraditionalForm=

$$\mathcal{P}_{\text{pr}}\left(\mathcal{S}^n \mid \xi = p\right) = p^n$$

We can use the Bayes theorem to sum over all possible values of p

Out[ ]//TraditionalForm=

$$\mathcal{P}_{\text{pr}}\left(\mathcal{S}^n\right) = \sum_{p=0}^{1} \mathcal{P}_{\text{pr}}\left(\xi = p\right) \mathcal{P}_{\text{pr}}\left(\mathcal{S}^n \mid \xi = p\right)$$

Then moving to the continuous form

Out[ ]//TraditionalForm=

$$\mathcal{P}_{\text{pr}}\left(\mathcal{S}^n\right) = \int_0^1 \mathcal{P}_{\text{pr}}\left(\mathcal{S}^n \mid \xi = p\right) dp = \int_0^1 p^n \, dp = \frac{1}{n+1}$$

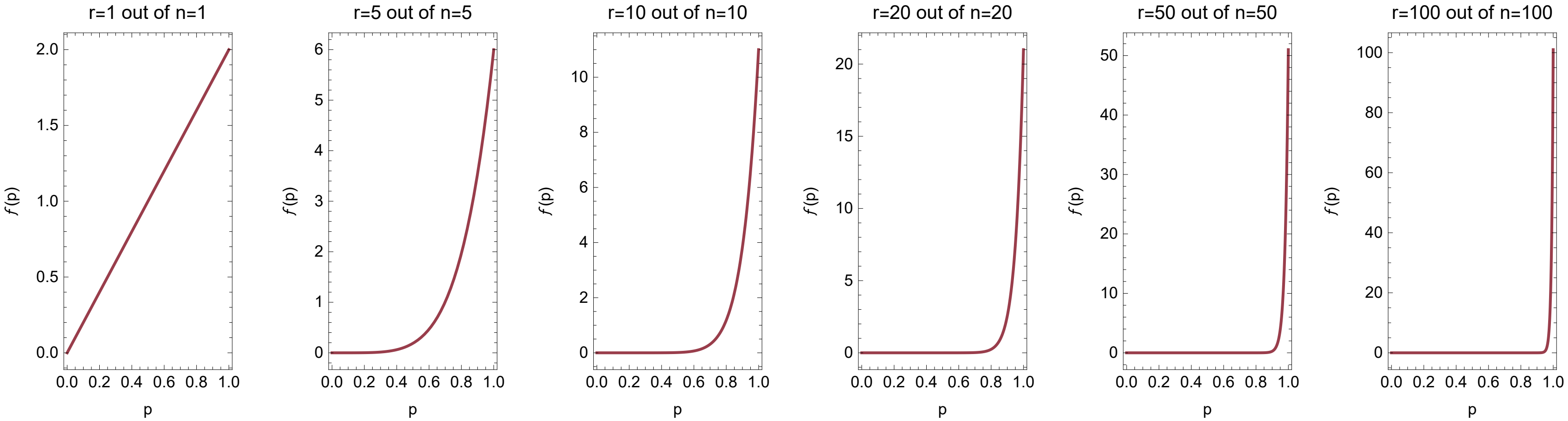
So for tomorrow's probability we find

Out[ ]//TraditionalForm=

$$\mathcal{P}_{\text{pr}}\left(\mathcal{S}^{n+1} \mid \mathcal{S}^n\right) = \frac{\mathcal{P}_{\text{pr}}\left(\mathcal{S}^{n+1}\right)}{\mathcal{P}_{\text{pr}}\left(\mathcal{S}^n\right)} = \frac{\left(\frac{1}{n+2}\right)}{\left(\frac{1}{n+1}\right)} = \frac{n+1}{n+2}$$

Laplace’s rule of succession

Out[•]=



We can see that as the event occurs again and again, we become more confident that the underlying chance is near 1. To find the expectation of this distribution, we need to work out

Out[•]//TraditionalForm=

$$\int_0^1 p\, f(p)\, dp = (n+1) \int_0^1 p^{n+1}\, dp = \frac{(n+1)}{(n+2)}$$

Laplace’s rule of succession

When the event does not occur every time

Suppose we have observed r events in n opportunities. Then the probability of this particular sequence is $p^r(1-p)^{n-r}$. Again, if we believe all value of p were equally plausible before making any observations, the distribution for p is just proportional to $p^r(1-p)^{n-r}$. Probability distributions must integrate up to 1, and a standard result is that

Out[ ]//TraditionalForm=

$$\int_0^1 p^r (1-p)^{n-r} \, dp = \frac{r!\,(n-r)!}{(n+1)!}$$

So the full probability distribution for p must be of the form

Out[ ]//TraditionalForm=

$$f(p) = \frac{(n+1)!}{r!\,(n-r)!} p^r (1-p)^{n-r}$$

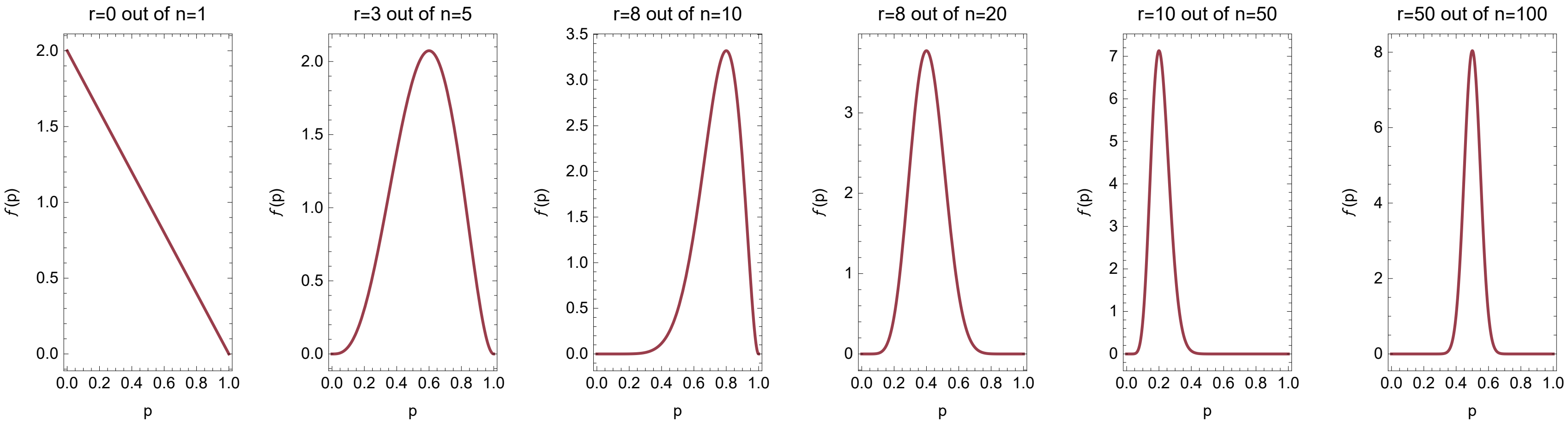
Laplace’s rule of succession

When the event does not occur every time

Out[]//TraditionalForm=

$$f(p) = \frac{(n+1)!}{r!\,(n-r)!}\, p^r\,(1-p)^{n-r}$$

Out[]=



To find the expectation of this distribution, we need to work out

Out[]//TraditionalForm=

$$\int_0^1 p\, f(p)\, dp = \frac{(n+1)!}{r!\,(n-r)!} \int_0^1 p^{r+1}\,(1-p)^{n-r}\, dp = \frac{(n+1)!}{r!\,(n-r)!}\, \frac{(r+1)!\,(n-r)!}{(n+2)!} = \frac{(r+1)}{(n+2)}$$

Code

Init