Stat00-Probability.wxmx

TABLE OF CONTENTS


```
(%i6) load(draw)$
set_draw_defaults(line_width=2, grid = [2,2], point_type = filled_circle,
        head_type = 'nofilled, head_angle = 20, head_length = 0.5,
        background_color = light_gray, draw_realpart=false)$
load (descriptive)$  load (distrib)$
 fpprintprec : 5$ ratprint : false$
```

Homemade functions fll, head, tail, Lsum are useful for looking at long lists.

```
(%i10) fll ( aL) := [ first (aL), last (aL), length (aL) ]$
head(L) := if listp (L) then rest (L, - (length (L) - 3) ) else
     error("Input to 'head' must be a list of expressions ")$
tail (L) := if listp (L) then rest (L, length (L) - 3 )  else
      error("Input to 'tail' must be a list of expressions ")$
Lsum (aList) := apply ("+", aList)$
```

1 Preface

In Stat00-Probability.wxmx we discuss the axioms of probability, the different basic types of probability needed, and illustrate with a selection of problems taken from the literature. This is the first chapter in the series Statistics with Maxima.

Edwin L. (Ted) Woollett
https://home.csulb.edu/~woollett/
Nov. 28, 2024

2 References

R. J. Barlow, Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences (Manchester Physics Series), 1993, Wiley

Louis Lyons, Statistics for Nuclear and Particle Physics, 1986, Cambridge Univ. Press,

Luca Lista, 'Statistical Methods for Data Analysis in Particle Physics', Lecture Notes in Physics 909, 2016, Springer-Verlag,

Glen Cowan , Statistical Data Analysis, Clarendon Press, Oxford, 1998

Glen Cowan, Statistical Methods for Particle Physics, 2018 Trieste Summer School on Particle Physics and Detectors, https://www.pp.rhul.ac.uk/~cowan/stat/cowan_trieste2018_1.pdf

Harry Lass, Elements of Pure and Applied Mathematics, McGraw-Hill, 1957

PSU stat course: https://online.stat.psu.edu/stat414/

3 Probability in General

Here we quote some paragraphs from Lyons, Ch. 2.

"In many situations we deal with experiments in which the essential circumstances are kept constant, and yet repetitions of the experiment produce different results. Thus the result of an individual measurement or trial may be unpredictable, and yet the possible results of a series of such measurements have a well defined distribution. The probability p of obtaining a certain specified result on performing one of these measurements is then simplest to visualize as the ratio

p = (number of occasions on which that result occurs)/(total number of measurements)."

"Example (a): We throw a die and want to score four. Because the die has six faces and because of the symmetry of the situation (assuming the die is unbiased) we expect to be successful in 1/6 of our attempts. In this case our experiment results in a discrete distribution with only six different possibilities (1, 2, 3, 4, 5, 6), each equally probable."

"Example (b): A radioactive source situated in a magnetic field decays isotropically in space. A counter which detects one of the decay products subtends a solid angle of $d\Omega$ steradians as seen by the source, and can be set at an angle θ with respect to the direction of the magnetic field. The probability that, in any given decay, the decay product will pass through the detector is $d\Omega/(4*\pi)$, independent of θ. Here the distribution in θ is continuous, and constant."

"Example (c): We count the number n of throws of a die required until we score six. Clearly, lower values of n are more likely than higher values. This time we have a discrete distribution with an infinite number of possibilities (n can take any integral value from one upwards), and the relative probabilities are different."

" Example (d): An excited state is produced by a nuclear reaction, but it has a very short lifetime. Thus its mass m is not uniquely determined, but is characterised by the Breit-Wigner distribution B (m, Mo, Γ) whose parameters are the mass Mo and width Γ of the resonant state."

We will show the Breit-Wigner distribution, using the definition on pages 55, 56 of Luca Lista, Statistical Methods for Data Analysis, with Applications in Particle Physics, 3rd edition.

3.1 Relativistic Breit-Wigner Distribution B (m, Mo, Γ)

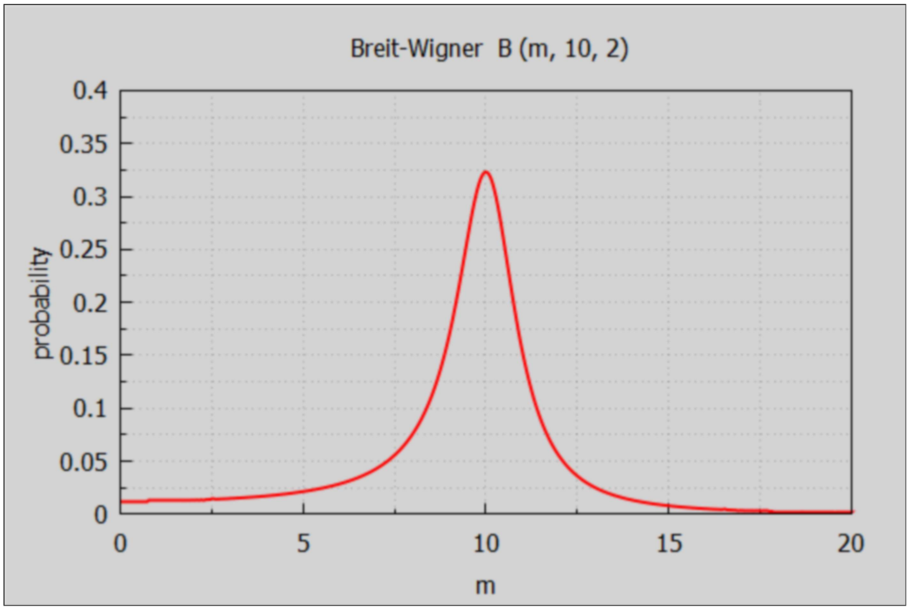
We are only going to use this function for getting numerical values. In our function definition, xx refers to the input parameter m, mm refers to the input parameter Mo, and gam refers to the input parameter Γ. B (m, Mo, Γ), as defined here, is normalized to unity over the mass interval [0, ∞], as we show below, and is a probability distribution function (pdf) which gives the probabilities of observing different values of the mass of the excited state. This is an example of a continuous distribution (in contrast with a discrete distribution).

```
(%i11) B (xx, mm, gam) :=
       block ([NN, kk],
          kk : sqrt (mm^2 * (mm^2 + gam^2) ),
          NN : 2*sqrt(2)*mm*gam*kk / (%pi*sqrt (mm^2 + kk) ),
          NN / ( (xx^2 - mm^2)^2  +  mm^2 * gam^2 ) )$

(%i12) float (integrate (B (m, 10, 2), m, 0, inf) );
(%o12) 1.0

(%i13) B (10, 10, 2), numer;
(%o13) 0.32302
```

(%i14) wxdraw2d (xrange = [0, 20], yrange = [0, 0.4], xlabel = "m", ylabel = "probability",
title = " Breit-Wigner B (m, 10, 2)",
color = red, line_width = 2, explicit (B (m, 10, 2), m, 0, 20))$

(%t14) 

"The probability of obtaining a mass within a small range of any specific value m is proportional to the height of the curve at that particular point. Thus we are most likely to observe excited state masses near the central value Mo; masses that differ from Mo by much more than Γ are unlikely."

What fraction of the area under the curve exists in the m interval [8, 12]? That is in the interval [Mo - Γ, Mo + Γ].

Using integrate for finding a definite integral results in a long complicated expression. Using float (integrate (...)) returns a divide by zero type error: (expt: undefined: 0 to a negative exponent) with either B (m, 10, 2) as the integrand, or a numerical approximation to B (m, 10, 2) as the integrand.

Using Maxima's big float abilities (arbitrarily many digits of precision), with 200 digit arithmetic, we can evaluate the numerical value of the integral we want via integrate.

(%i15) bfloat (integrate (B (m, 10, 2), m, 8, 12)), fpprec : 200;

(%o15) 7.1668b−1

The big float number representation 7.16678b-1 translates to 0.716678.

Maxima has the numerical integration function quad_qags (f(x), x, a, b, [options]) for numerical integration of a smooth function over a finite interval [a, b]. If I try with B (m, 10, 2) [as defined above], we get the same answer (~ 0.72) as when using bfloat. (See the manual for the meaning of the other numbers returned in the list returned by quad_qags.)

One of the Maxima functions available for numerical integration is
quad_qags (f, x, a, b, [epsrel, epsabs, limit]), designed for smooth functions over a finite interval. The first number in the returned list is the approximate numerical value of the integral requested.

```
(%i16) quad_qags (B (m, 10, 2), m, 8, 12, 'epsrel=1d-10 );
```

(%o16) $[0.71668, 5.653\ 10^{-11}, 63, 0]$

Thus we have about a 72% chance of observing a mass value (for the excited state) in the range [8, 12] and about a 28% chance of observing a mass value outside of this range.

4 Rules of Probability

[MW]: "Let P(A) be the probability of something (called A) occurring when an experiment is performed. 0 <= P(A) <= 1. If A is certain to happen, then P(A) = 1. If A will certainly not happen, then P(A)=0."

4.1 Rule 1

Lyons Sec. 2.2:

"Rule 1: The probability of any particular event occurring is a number between zero and one inclusive. A probability of zero implies that that particular event never happens, while a value of one implies that it always occurs. Thus for a single six sided die, the probability of throwing a seven is zero, of obtaining any number less than 10 is unity, and of obtaining an even number is 1/2."

4.2 Notation: P (A + B), P (A, B)

4.2.1 P (A + B) = P (A or B)

The probability of either A or B (or both) occurring is denoted as P (A + B), Barlow uses P (A or B).

In set theory (Harry Lass, pp 342-350, 388), if A1 and A2 are subsets of some set of events (elements) S, the union or sum of all events (elements) which belong to either A1 or A2 or both is written A1 + A2.

4.2.2 Joint Probability P (A, B) = P (A and B)

The probability of both A and B occurring is called the 'joint probability' of A and B and we denote it by the symbol P (A, B) or P (A and B). Lyons uses the notation P (A B) for the joint probability. Barlow uses P (A and B) for the joint probability.

In set theory (Lass) the intersection or product of A1 and A2, written A1 A2 is the set of elements belonging to both A1 and A2. If A1 and A2 have no elements in common, their intersection is the null set, A1 A2 = 0.

4.3 Rule 2 P (A + B) = P (A or B)

Lyons: "Rule 2: The probability P (A + B) that at least one of the events A or B occurs is given in terms of the individual probabilities P(A) and P(B) by
P (A + B) <= P(A) + P(B).
The equality applies if the events A and B are mutually exclusive; i.e. if the occurrence of A precludes that of B." Thus for mutually exclusive events A and B, P(A + B) = P(A) + P(B).

It can be shown that in the general case, P (A + B) = P(A) + P(B) - P (A, B), equivalent to
P (A or B) = P (A) + P (B) - P (A and B).

4.4 Rule 3 Conditional Probability P (A | B), and Multiplication Rule

The probability that A occurs, given that B occurs, is denoted as: P(A | B).

Rule 3: The [joint] probability P (A, B) of obtaining both A and B can be decomposed in two ways: ('Multiplication Rule'):

P (A, B) = P(B) * P (A | B) = (the probability B occurs)*(the probability that A occurs given B),
or as
P (A, B) = P(A) * P (B | A). (2.3)

From the first line, we get the "conditional probability formula" :

P (A | B) = P (A, B) / P (B), [if P (B) # 0] (2.4)

"The conditional probability P (A | B) is obtained by dividing the number of times that both A and B are observed together by the total number of times that B occurs."

"If the occurrence of B does not affect whether or not A occurs, then
P (A | B) = P(A) (2.5)
and A and B are said to be independent. From (2.3) and (2.5) we then get, for independent events:
P (A, B) = P(A) * P(B) (2.6) "

4.5 Lyons' Examples

Lyons Example (a)
" A = It is Sunday.
B = It is raining.
The probability of its being rainy on a Sunday is the same as that for any other specific day of the week, and so A and B are independent."

Lyons Example (b)
A and B correspond to two different beam tracks interacting in a target. Whether the first track interacts [with the target] or not, does not affect whether the second track interacts with the target. This is true even if the [first and second] beam particles are of different types with different interaction probabilities.

Lyons Example (c)
"If each week you spend 42 hours in your laboratory, the probability at any moment that your head is in the laboratory is 0.25. [42 / (7*24) = 42 / 168 = 0.25] Similarly, the probability of your feet being in the laboratory is also 0.25. But the probability of both your head and your feet being there is again 0.25, and not 1/16, as would be obtained from eqn (2.6)."
The events A = your head is in the lab, B = your feet are in the lab, are not independent:
"Here the non-independence rises from the correlation between your head and your feet: they are constrained to be less than about 1.5 meters apart."

Lyons Example (d)
"Abram and Lot were standing at a road junction. The probability of Lot taking the westward road was 0.5. The probability of Abram taking the westward road was also 0.5. But the probability of both of them taking the westward road was zero. (The reader interested in further details should consult Genesis 13.9.)

Lyons Example (e)

[In nuclear physics, beta decay (β-decay) is a type of radioactive decay in which an atomic nucleus emits a beta particle (fast energetic electron or positron), transforming into an isobar of that nuclide.]

"The Fermi theory allows us to calculate the electron spectrum in β decay, e.g. the decay:
neutron --> proton + electron + anti-neutrino. [The 'electron spectrum' is usually shown by plotting the number of beta particles emitted with energy E as a function of E.]"

"So we can obtain the probability of the electron having a certain high fraction (say >= 3/4) of the available energy. We can also do a similar calculation for the anti-neutrino. But the probability of their *both* having high energies is zero, since they are constrained by energy conservation to share the total available energy."

Lyons Example (f)
"As an example of almost independent variables, consider a beam particle entering a region of target material, traveling through the material a distance x, at which position interaction with the target material causes the beam particle to be scattered away from the beam direction by an angle θ. Because of the energy lost by the beam particle while traversing the target material, the strength of interaction with the material at the far end of the target material will occur at a lower beam particle energy than for a beam particle just entering the target. This difference of energy will cause a slight difference in the scattering angular distribution P(θ) at different positions withing the target material, since P(θ) depends on the energy.
If A = position of interaction in the target, B = scattering angle in that interaction,
then if the beam momentum is not too high, then the energy loss as the beam passes through the target may be significant. Thus the scattering angle and the interaction position are not completely independent, although at high beam energy and/or for thin targets this lack of independence may in practice be unimportant."

5 *Conditional Probability Examples*

https://online.stat.psu.edu/stat414/lesson/4/4.1

5.1 PSU Ex. 4.1

"A researcher is interested in evaluating how well a diagnostic test works for detecting renal disease (kidney disease) in patients with high blood pressure. She performs the diagnostic test on 137 patients, 67 with known renal disease and 70 who are known to be healthy. The diagnostic test comes back either positive (the patient has renal disease) or negative (the patient does not have renal disease). Here are the results of her experiment:

DIAGNOSTIC RENAL TEST RESULTS

TRUTH	POSITIVE	NEGATIVE	TOTAL	
Renal disease	44	23	67	Sample space D
Healthy	10	60	70	Sample space H
Total	54	83	137	Sample space S = D + H

" If we let T+ be the event that the person tests positive we can use the relative frequency approach to assigning probability to determine that:

P (T+) = 54/137

because, of the 137 patients, 54 tested positive.

If we let D be the event that the person is truly diseased, we determine that:

P (D) = 67/137

because, of the 137 patients, 67 are truly diseased. That's all well and good, but the question that the researcher is really interested in is this: If a person has renal disease, what is the probability P (T+ | D) that he/she tests positive for the disease? The first clause is the 'conditional' part of the question, the second clause is the 'probability'.

We can again use the relative frequency approach and the data the researcher collected to determine:

P (T+ | D) = 44/67 = 0.657.

That is, the probability a person tests positive, given he/she has renal disease, is 0.657. There are a couple of things to note here. First, the notation P (T+ | D) is standard conditional probability notation. It is read as "the probability a person tests positive given he/she has renal disease." The bar (|) is always read as "given." The probability we are looking for precedes the bar, and the conditional follows the bar.

Second, note that determining the conditional probability involves a two-step process. In the first step, we restrict the sample space to only those (67) who are diseased, that is sample space D. Then, in the second step, we determine the number of interest (44) based on the same sample space.

To derive a general formula we follow what we did above:

P (T+ | D) = (number known to have renal disease who tested positive) /
(number known to have renal disease)
= 44/67
= N (T+, D)/N(D) = [N(T+,D) / N(S)] / [N(D) / N(S)] = P (T+, D) / P (D)

In summary

We can abstract this formula to a general situation if we denote T+ by A and D by B.
then the conditional probability of event A, given event B, is calculated using:
P (A | B) = P (A, B) / P (B)
= (number of times both A and B occur together) / (number of times B occurs).

"Now, when a researcher is developing a diagnostic test, the question she cares about is the one we investigated previously, namely: If a person has renal disease, what is the probability of testing positive? This quantity is what we would call the 'sensitivity' of a diagnostic test.
As patients, we are interested in knowing what is called the 'positive predictive value' of a diagnostic test. That is, we are interested in this question: If I receive a positive test, what is the probability P (D | T+) that I actually have the disease?"

"We would hope, of course, that the probability is 1. But, only rarely is a diagnostic test perfect. The collected data suggest that the renal disease test [example above] is not perfect. How good is it? That is, what is the positive predictive value of the test?"

"The definition of conditional probability we found above says that the conditional probability of actually having the disease, given you have tested positive [with the above diagnostic test, and based on the experimental data] is
P (D | T+) = P (D, T+) / P(T+)
= [44/137] / [54 / 137]
= 44 / 54 = 0.815
So you have about an 81% chance of having the disease if you test positive - pretty high."

"Now above we found the data for this hypothetical diagnostic test and hypothetical group tested implied that P (T+ | D) = 0.657, which illustrates that in general
P (D | T+) # P (T+ | D)."

5.1.1 Properties of Conditional Probability

https://online.stat.psu.edu/stat414/lesson/4/4.2

As long as P(B) > 0, we have the theorems:
1. P (A | B) >= 0
2. P (B | B) = 1
3. If A_1, A_2, A_3, ..., A_k are mutually exclusive events, then
P (A_1 + A_2 + A_3 + ... + A_k | B) = P (A_1 | B) + P (A_2 | B) + ... + P (A_k | B).

5.2 PSU Example 4.3

https://online.stat.psu.edu/stat414/lesson/4/4.2

A box contains 6 white balls and 4 red balls. We randomly (and without replacement) draw two balls [in succession] from the box.

A.) What is the probability that the second ball selected is red, given that the first ball selected is white?

We are looking for P (R2 | W1) = probability of drawing a red ball on the second draw from the box, given that we drew a white ball on the first draw. After the first draw we have 5 white balls and 4 red balls in the box, and in this 'reduced sample space' of 9 balls, the chance of drawing a red ball on the second draw is 4/9.

B.) What is the joint probability [P (R1, R2)] the first two balls selected are both red?

If we solve for the joint probability P (R1, R2) using the conditional probability formula
P (R2 | R1) = P (R1, R2) / P (R1), we can use

P (R1, R2) = P (R2 | R1) * P (R1), known as the 'multiplication rule.'

Now P (R1) = 4/10 since the box starts with 10 balls, 4 of them red.

In the reduced sample space after the first draw, we have a box with 3 red balls out of 9 balls. So P (R2 | R1) = 1/3.

Hence P (R1, R2) = (1/3) * (4/10) = 2/15.

An alternative derivation uses the total number of ways 2 objects can be chosen from among 10 objects, without regard for order, given by C (n, r) = number of combinations of n objects taken r at a time, given by C (n, r) = n! / (r! * (n - r)!).

Then C (10, 2) = 45. The number of ways two red balls can be drawn from 4 red balls is C (4, 2) = 6. The probability of drawing 2 red balls from a box containing 10 balls with 4 of them red is then the ratio C (4,2) / C (10,2) = 6 / 45 = 2/15

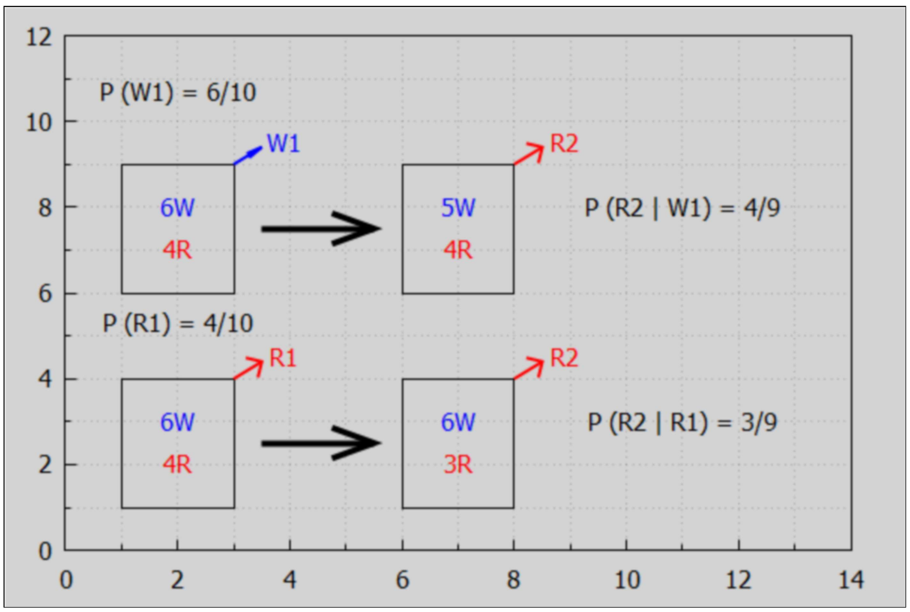
5.3 PSU Ex. 4.4 The Multiplication Rule

https://online.stat.psu.edu/stat414/lesson/4/4.3

A box contains 6 white balls and 4 red balls. We randomly (and without replacement) draw two balls from the box. What is the probability that the second ball selected is red?

We use wxdraw2d to show two different paths possible. The top path and the bottom path have different probabilities.

```
(%i17) wxdraw2d (xrange = [0,14], yrange = [0, 12],  color = black,line_width = 1,
    transparent = true, rectangle ([1,6],[3,9] ), color = blue, label ( ["6W", 2, 8] ),
    color = red, label (["4R",2,7] ), color = blue,  head_angle = 5,
    line_width = 2, vector ([3,9],[0.5,0.4]), label ( ["W1",3.9,9.5]), color = black,
    line_width = 4, head_angle = 20,  head_length =  0.8, vector ([3.5,7.5],[2,0]),
    line_width = 1,
       rectangle ( [6,6],[8,9] ), color = blue, label( ["5W", 7,8]), color = red,
   label (["4R",7,7]), line_width = 2, vector ([8,9],[0.5,0.4] ), label ( ["R2",8.9,9.5 ]),
    color = black, label ( ["P (R2 | W1) = 4/9", 11,8]), line_width = 1,
    rectangle ([1,1],[3,4]), color = blue, label( ["6W", 2,3]), color = red,
    label (["4R",2,2] ), line_width = 2,  vector ([3,4],[0.5,0.4]), label ( ["R1",3.9,4.5 ]),
    color = black, line_width = 4, vector ([3.5,2.5],[2,0] ), line_width = 1,
     rectangle ([6,1],[8,4] ), color = blue, label( ["6W", 7,3]), color = red,
      label (["3R",7,2]), line_width = 2, vector ([8,4],[0.5,0.4]),   label ( ["R2",8.9,4.5]),
     color = black, label ( ["P (R2 | R1) = 3/9", 11, 3]), label ( ["P (W1) = 6/10",2, 10.7],
       ["P (R1) = 4/10", 2, 5.3] )   )$
```

(%t17) 

The key to this question is recognising there are two ways the second ball drawn from the box is red. The first way (path) is the first ball drawn is white. The second way (path) is the first ball drawn is red. These two ways are mutually exclusive.

P (R2) = P [(W1 and R2) or (R1 and R2)]

= P (W1 and R2) + P (R1 and R2), since the two paths are mutually exclusive.

Now the "multiplication rule" says that

P (W1,R2) == P (W1 and R2) = P (R2 | W1) * P (W1) = (4/9) * (6/10) = 24/90

and P (R1, R2) == P (R1 and R2) = P (R2 | R1) * P(R1) = (3/9) * (4/10) = 12/90.

So
P(R2) = 36/90 = 2/5 = 0.4

5.4 PSU Ex. 4-5, The Multiplication Rule Extended

"The multiplication rule can be extended to three or more events. In the case of three events, the rule looks like this:"

P (A,B,C) = P (A and B and C) = P [(A and B) and C] = P [C and (A and B)] * P (A and B).

But P (A and B) = P (B | A) * P (A), so

P (A, B, C) = P [C and (A and B)] * P (B | A) * P (A).

Three cards are dealt successively at random and without replacement from a 'standard deck' of 52 playing cards. What is the probability of receiving, in order, a king, a queen, and a jack?

A "standard" deck of playing cards consists of 52 Cards in each of the 4 suits of Spades, Hearts, Diamonds, and Clubs. Each suit contains 13 cards: Ace, 2, 3, 4, 5, 6, 7, 8, 9, 10, Jack, Queen, King. Modern decks also usually include two Jokers.

P (K1 and Q2 and J3) = P(K1)*P (Q2 | K1) * P [J3 | (K1 and Q2)]
= (4/52) * (4/51) * (4/50) = 64/ (132,600).

```
(%i18) float ( 64/132600 );
```

(%o18) $4.8265\ 10^{-4}$

A pretty low probability.

5.5 PSU Ex. 4-6

A drawer contains 18 socks:
4 red socks
6 brown socks
8 green socks

A man is getting dressed one morning and barely awake when he randomly selects 2 socks from the drawer (without replacement, of course). Let's assume the room is very dark and all socks are made from the same material.

Given he ends up with two socks of the same color, what is the probability that both of the socks he selects are green? If we define four events as such:

Let R_i = the event the man selects a red sock on selection i for i = 1,2,
Let B_i = the event the man selects a brown sock on selection i for i = 1,2,
Let G_i = the event the man selects a green sock on selection i for i = 1,2,
Let S = the event that the 2 socks selected are the same color, irrespective of color,

then we are looking for the conditional probability: P (G1 and G2 | S).

Using the conditional probability formula:

P (G1 and G2 | S) = P (G1 and G2 and S) / P(S)

Now the numerator P (G1 and G2 and S) = P (G1 and G2) since if you have two greens you obviously have the same color, and using the multiplication formula,

P (G1 and G2) = P (G1) * P (G2 | G1) = (8/18) * (7/17) = 56/306.

For the denominator, the probability of getting two socks of the same color
P (S) = P [(G1 and G2) or (B1 and B2) or (R1 and R2)].
Now use: getting two green socks is a mutually exclusive event from getting two brown socks, etc., to write
P (S) = P (G1 and G2) + P (B1 and B2) + P (R1 and R2),
then use the multiplication formula three times:
P (S) = P (G1) * P (G2 | G1) + P (B1) * P (B2 | B1) + P (R1) * P (R2 | R1)
= (8/18) * (7/17) + (6/18) * (5/17) + (4/18) * (3/17) = 98/306.

Finally P (G1 and G2 | S) = P (G1 and G2 and S) / P(S) = P (G1 and G2) / P(S)
= (56/306) / (98/306) = 56/98 = 0.571,
so the man, given he has pulled two pair of socks of the same color from his drawer, has about a 57% chance the two pair are green.

5.6 PSU Ex. 4-7

https://online.stat.psu.edu/stat414/lesson/4/4.4

Data:
Medical records reveal that of the 937 men who died in a particular region in 1999:

212 of the 937 men died of heart disease (hd).
312 of the 937 men had at least one parent who died of hd, and 102 out of this 312 died of hd.

Problem:
If we randomly select one man from this sample of 937 males, what is the probability that he dies of hd, given that *neither* of his parents died from hd?

Solution:
If we define two events, D and H, as:
D = the event that a randomly selected man died of hd.

H = the event that at least one of the parents of a randomly selected man died of hd.

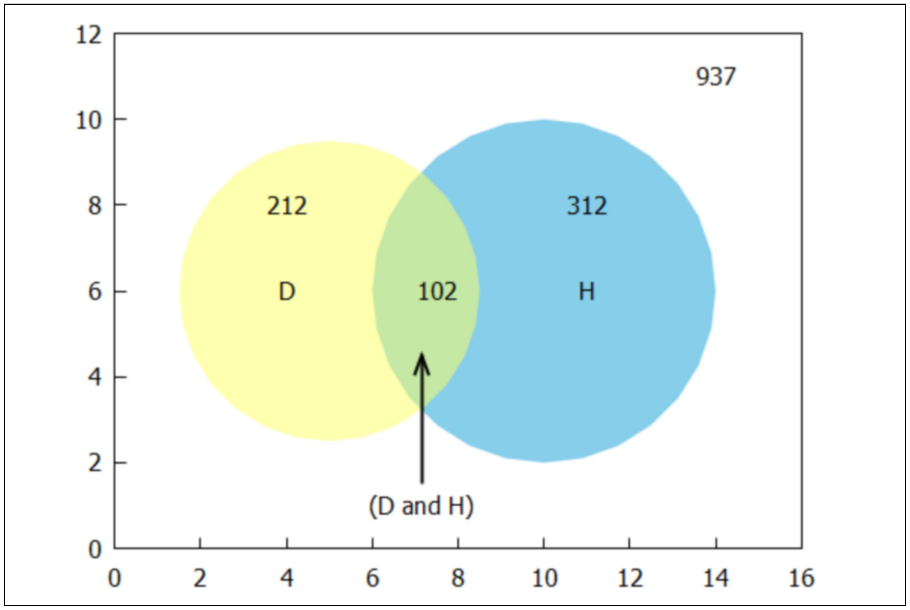
Then we are looking for the following conditional probability:
P (D | not H) == P (D | ~H)

5.6.1 Venn Diagram

We can use a Venn diagram to think about this problem. We draw a large square which represents the sample of 937 men who died in the particular region in 1999. We then draw two circles inside this square: one circle "D" representing the 212 men from this sample who died from hd, and a second circle "H" representing the 312 men from this sample who had parental hd death. The two circles overlap with an area representing the 102 men who both died from hd and had parental hd death.

(%i19) wxdraw2d (xrange = [0,16], yrange = [0, 12], proportional_axes = 'xy, border = false,
background_color = white, grid = false, fill_color = "#87ceeb", ellipse (10,6,4,4,0,360),
fill_color = "#ffff6678",ellipse (5,6,3.5,3.5,0,360), color = black,
label (["937",14,11], ["102",7.5,6], ["312",11,8], ["H", 11,6], ["212",4,8],
["D",4, 6], ["(D and H)", 7.15, 1]), head_angle = 20, vector ([7.15,1.5],[0,3]))$

(%t19)



In the above plot we used hex color codes in the form "#rrggbb" (two hex digits for red, two hex digits for green, and two hex digits for blue) and the eight hex digit form "#rrggbbaa". In the latter form, the last two hex digits aa are called the "alpha channel" and govern the amount of transparency the color has. The color "#ffff6678" is the eight hex digit form of yellow, with aa = 78.

Transparency values near ff mean "very transparent" and values near 00 mean "almost opaque". 78 (hex) = 120 (decimal) = 0.47*255. ff (hex) = 255 (decimal), so 78 (hex) is about 47% of ff.

5.6.2 Two-Way Table Method

Numbers we can infer from the given data are placed in a two-way table. Rows are labeled by the cause of death, either heart disease (hd) (D), or causes other than hd (~D). Columns are labeled by whether a random man from the sample had at least one parent who died from hd (H), or if a random man from the sample had no parent who died from hd (~H).

(%i20) matrix ([" ", "|", "H","|", "~H", "|", "Total"], ["--", "|", "--", "|", "--", "|", "--"],
[" (D) ", "|",102, "|", 110, "|", 212], ["--", "|", "--", "|", "--", "|", "--"],
[" (~D) ", "|", " ", "|", " ", "|", 725],
["--", "|", "--", "|", "--", "|", "--"], [" ", "|", 312, "|", 625, "|", 937]);

(%o20)

	|	*H*	|	*~H*	|	*Total*
--	|	--	|	--	|	--
(D)	|	102	|	110	|	212
--	|	--	|	--	|	--
(~D)	|		|		|	725
--	|	--	|	--	|	--
	|	312	|	625	|	937

For example, in row 2 and column 2 we have 102 = (D and H).
In row 2, col 3 we have 110 = 212 - 102 = (D and ~H).

Using the conditional probability formula (Bayes' theorem):

P (D | ~H) = P (D and ~H) / P (~H)

From the table, P (~H) = 625/937, P(D and ~H) = 110/937, so
P (D | ~H) = (110/937) / (625/937) = 110/625 = 0.176.

6 Mathematical Probability

Some paragraphs [with judicious editing] taken from Barlow Ch. 7.

"Discrete events example."

"S = {E1, E2, E3, ..,} is the set of possible results of an experiment ("events"). Events are said to be 'mutually exclusive' if it is impossible for both to occur in a measurement result. For every event E there is a probability P(E), a real number, which satisfies three axioms:"

1. P(E) >= 0.
2. P (E1 + E2) == P(E1 or E2 or both) = P(E1) + P(E2) if E1 and E2 are mutually exclusive so P (E1 and E2) = 0.
3. ∑ P(E_i) = 1, where the sum is over all possible *mutually exclusive* events.

"These three axioms lead to other results of probability theory. For example, Axiom 3 implies P (~E) == P(not E) = 1 - P(E), and this implies that P(E) <= 1."

"These axioms can be extended to the results of measuring a continuous variable, in which case the sample space is infinite, with no essential difficulty. Integrals replace sums, and P(x)*dx is the probability the variable value is measured to be in the interval [x, x + dx]."

"These axioms do not constrain the physical meaning of the number you calculate - the axioms do not tell you what meaning to ascribe to that number."

7 Empirical Probability

Some paragraphs [with judicious editing] taken from Barlow Ch. 7.

In this section (and in other sections), Barlow frequently uses casual language such as 'The probability of event A is 20%' or 'the fraction occurring is 2%', etc. Since a 'fraction' is strictly a number less than one, such as 1/4 = 0.25, and not 25%, and we have defined probability P(E) as having the property 0 <= P(E) <= 1, we try to use more careful language, such as if P(E) = 0.25, then 'the chance of event E occurring is 25%', etc. Thus Barlow's statement 'the probability is 20%' means 'the probability is 0.2 and the chance is 20%.' Once you have a firm understanding of the mathematical meaning of probability, you can loosen up your language when talking either to other 'experts' or to lay people who are not interested in strict language use.

"Scientists almost always assign a meaning to a probability using a frequency [frequentist] approach. If an experiment is repeated, under identical conditions, N times, and a particular outcome A occurs in M of these cases, as N → ∞ the ratio M/N tends to a limit defined as the probability P(A). "

"The N performances may be done by repeating the same experiment N times, one after the other, or by simultaneous measurements on N identical experiments. The set of all N cases is called the 'collective' or 'ensemble'."

"The ‘probability’ of an event is not a property just of the particular experiment, but a joint property of the experiment and the collective. This is nicely illustrated by an example (taken from von Mises). It has been found by the German insurance companies that the fraction of their male clients dying when aged 40 is 1.1%. However, we cannot then say that a particular Herr Schmidt has a probability of 1.1% of dying (or, more cheerfully, 98.9% of surviving) between his 40th and 41st birthdays."

"If data had been collected not from German insured men but from any other sample to which Herr Schmidt belongs (all German men, all men, all Germans, all German insured nonsmoking men, all German hang-glider pilots,…) we would have got a different fraction. All these different numbers can be considered as the ‘probability’ of his passing away prematurely, so the quoted ‘probability’ depends not just on the individual but on the collective to which it is considered to belong."

"Secondly, the experiment must be repeatable, under identical conditions, with different possible outcomes. This is a very strong requirement. Consider the harmless-sounding phrase. ‘It will probably rain tomorrow.’ What can we mean by this? There is only one tomorrow, we can wait and see what happens, but we can only do that once."

"Von Mises faces up to this squarely: in his view, such a use of the word ‘probability’ is unscientific, in the same way that ‘work’ and ‘energy’ are often so (mis)used. However, one would think that if the barometer falls, the clouds gather, and satellite pictures show a cold front approaching, then to say that it will probably rain is a meaningful and sensible statement, and that any definition of probability which forbids it is unduly restrictive."

8 Relative Frequency Interpretation

Quoting Cowan's text, pdf p.18:

"The relative frequency interpretation is consistent with the axioms of probability, since the fraction of occurrences is always greater than or equal to zero, the frequency of any out of a disjoint set of outcomes is the sum of the individual frequencies, and the measurement must by definition yield some outcome (i.e. P(S) = 1). The conditional probability P(A I B) is thus the number of cases where both A and B occur divided by the number of cases in which B occurs, regardless of whether A occurs. That is, P(A I B) gives the frequency of A with the subset B taken as the sample space. "

"Clearly the probabilities based on such a model can never be determined experimentally with perfect precision. The basic tasks of classical statistics are to estimate the probabilities (assumed to have some definite but unknown values) given a finite amount of experimental data, and to test to what extent a particular model or theory that predicts probabilities is compatible with the observed data."

"The relative frequency interpretation is straightforward when studying physical laws, which are assumed to act the same way in repeated experiments. The validity of the assigned probability values can be experimentally tested. This point of view is appropriate, for example, in particle physics, where repeated collisions of particles constitute repetitions of an experiment. The concept of relative frequency is more problematic for unique phenomena such as the big bang. Here one can attempt to rescue the frequency interpretation by imagining a large number of similar universes, in some fraction of which a certain event occurs. Since, however, this is not even in principle realizable, the frequency here must be considered as a mental construct to assist in expressing a degree of belief about the single universe in which we live."

9 Subjective Probability, Bayesian Statistics

Quoting Cowan (Trieste Summer School, 2018) slides:

"Assume A, B, ... are hypotheses (statements that are true or false) .
Let P(A) = degree of belief that A is true.

Subjective probability can provide a more natural treatment of non-repeatable phenomena, such as systematic uncertainties, the probability that the Higgs boson exists, etc."

Quoting Cowan's text, pdf. p. 19:

"The statement that a measurement will yield a given outcome a certain fraction of the time can be regarded as a hypothesis, so the framework of subjective probability includes the relative frequency interpretation. In addition, however, subjective probability can be associated with, for example, the value of an unknown constant; this reflects one's confidence that its value lies in a certain fixed interval. A probability for an unknown constant is not meaningful with the frequency interpretation, since if we repeat an experiment depending on a physical parameter whose exact value is not certain (e.g. the mass of the electron), then its value is either never or always in a given fixed interval. The corresponding probability would be either zero or one, but we do not know which. With subjective probability, however, a probability of 95% that the electron mass is contained in a given interval is a reflection of one's state of knowledge."

"The use of subjective probability is closely related to Bayes' theorem and forms the basis of Bayesian (as opposed to classical) statistics. The subset A appearing in Bayes' theorem (equation (1.6)) can be interpreted as the hypothesis that a certain theory is true, and the subset B can be the hypothesis that an experiment will yield a particular result (i.e. data). Bayes' theorem then takes on the form

P(theory I data) [is proportional to] P (data I theory) * P (theory).

Here P (theory) represents the prior probability that the theory is true, and P (data I theory), called the likelihood, is the probability, under the assumption of the theory, to observe the data which were actually obtained."

"The posterior probability that the theory is correct after seeing the result of the experiment is then given by P (theory I data). Here the prior probability for the data P(data) does not appear explicitly, and the equation is expressed as, a proportionality."

"Bayesian statistics provides no fundamental rule for assigning the prior probability to a theory, but once this has been done, it says how one's degree of belief should change in the light of experimental data."

10 Bayes' Theorem: p (a | b) = p (b | a)*p(a) / p(b)

Bayes' theorem relates the conditional probabilities p (a | b) and p (b | a). Some paragraphs [with judicious editing] taken from Barlow Ch. 7.

"This view of probability [Subjective Probability] is also known as Bayesian statistics, and to discuss this we need Bayes’ theorem, and for that we need to define conditional probability. The conditional probability p(a | b) is the probability of 'a', given that 'b' is true."

"For example, if a day is chosen at random then p (Monday) = 1/7, However, if we know that it is a weekday, this becomes p (Monday | weekday) = 1/5."

If we write p (a, b) == p (a and b) = p(a | b) * p(b) = p (b | a)*p(a), and solve the last equality for p (a | b),

p (a | b) = p (b | a)*p(a) / p(b)   (Bayes' Theorem)

In this expression, the value of p(a) is referred to as 'the prior probability for a' and likewise p(b) is the 'prior probability for b'. Priors can come either prior knowledge, other experiments, and/or our prejudices.

10.1 P (~A) = P (not A)

Let P (~A) = P (not A) = 1 - P (A).

Since p (b) = the probability b will happen whether a is true or not and hence:
p(b) = p (b | a)*p(a) + p (b | ~a) * p (~a), and if we set p (~a) = 1 - p (a),
then
p (b) = p (b | a) * p(a) + p (b | ~a) * (1 - p(a)),

and we can recast Bayes' theorem in the form

p (a | b) = p (b | a)*p(a) / [p (b | a) * p(a) + p (b | ~a) * (1 - p(a))].

11 Counting Pions with a Cherenkov Counter

This is Example A in Barlow's Ch. 7 Section on 'Subjective Probability'.

Suppose one is using a Cherenkov counter, placed in the path of a beam of mesons, composed of 90% pions and 10% kaons. "In principle, the counter gives a signal for pions but not for kaons, thereby identifying any particular meson
[pion: gives signal, kaon: gives no signal].
In practice [the counter] is 95% efficient at giving a signal for [any beam] pion, and also has a 6% probability of giving an accidental signal from a [any given beam] kaon. "

"If a meson 'gives a signal', we can use Bayes' theorem to say that the probability of its being a pion is [with 'a' = π, and 'b' = signal]:"

p (π | signal) = p (signal | π) * p (π) / p (signal) = (0.95)*(0.9)/p(signal) = 0.855/p (signal).

Now using the expanded alternative form for p (b),

p (signal) = p (signal | π) * p (π) + p (signal | ~π) * (1 - p (π))
= (0.95) * (0.9) + (0.06) * (1 - 0.9)
= 0.855 + 0.006
= 0.8610.

Finally,
p (π | signal) = 0.855 / p (signal) = 0.855 / 0.861 = 0.993.
Hence the presence of a signal indicates a 99.3% chance the meson is a pion.

"The probability of its being a kaon is the complement of this:
p (K | signal) = 1 - p (π | signal) = 1 - 0.993 = 0.007 "
Thus the presence of a signal indicates a 0.7% chance the meson is a kaon.

Using Bayes' theorem to calculate p (K | signal) de novo: ('a' = K, 'b' = signal)

p (K | signal) = p (signal | K) * p (K) / p (signal) = (0.06) * (0.1) / 0.861
= 0.006/0.861 = 0.007, which confirms the previous calculation using
p (K | signal) = 1 - p (π | signal).

"If there is no signal, we have [with 'a' = K, 'b' = no signal == ~signal]

p (K | ~signal) = p (~signal | K) * p (K) / p(~signal) = (0.94) * (0.1) / p(~signal)
= 0.094 / p(~signal).

Now
p(~signal) = p (~signal | π) * p (π) + p (~signal | K) * p (K)
= (0.05) * (0.09) + (0.94) * (0.1) = 0.045 + 0.094 = 0.139.

Hence
p (K | ~signal) = 0.094 / p(~signal) = 0.094 / 0.139 = 0.6763.

"So the presence of a signal indicates an almost certain π; its absence indicates a probable but not certain K."

12 *Fraction of Real Muons*

This example is taken from Example 5.1 - Purity and Contamination, in Luca Lista, Ch. 5, p. 97.

"A proton beam hits a target and produces a sample of particles that is composed of muons (8%), kaons (10%), and pions (82%). A particle identification detector (ID) correctly identifies 95% of muons. It also incorrectly identifies as muons 7% of all kaons and 4% of all pions. What is the expected fraction of real muons in the sample of particles identified as such by the detector?"

"Our input data is the following:

P (μ) = 0.08 = probability a sample particle is a muon,
P (π) = 0.82 = probability a sample particle is a pion,
P (K) = 0.10 = probability a sample particle is a kaon,
P (ID | μ) = 0.95 = probability that, presented with a real muon, the detector records the muon,
P (ID | π) = 0.04 = probability that, presented with a real pion, the detector records a muon,
P (ID | K) = 0.07 = probability that, presented with a real kaon, the detector records a muon."

"We wish to calculate P (μ | ID), and we can use Bayes' theorem:
P (μ | ID) = P (ID | μ) * P (μ) / P (ID) = probability that a detector record of a muon corresponds to a real muon
= fraction of real muons in the sample of particles identified as such by the detector."

"The denominator in Bayes' theorem can be decomposed as
P (ID) = P (ID | μ) * P (μ) + P (ID | K) * P (K) + P (ID | π) * P (π)
= 0.95 * 0.08 + 0.07 * 0.1 + 0.04 * 0.82
= 0.076 + 0.007 + 0.0328 = 0.1158 = probability that, given a sample particle, the detector records a muon."

We then get
P (μ | ID) = P (ID | μ) * P (μ) / P (ID) = 0.95 * 0.08 / 0.1158 = 0.6563, which means that roughly 66% of the muon 'counts' correspond to real muons in a given sample.

"The fraction of properly identified particles, muons in this case, in a selected sample is called 'purity'. The purity of the selected sample is 65.63%."

We can also calculate
P (π | ID) = P (ID | π) * P (π) / P (ID) = 0.04 * 0.82 / 0.1158 = 0.2832 = probability that a detector record of a muon corresponds to a real pion
= fraction of real pions in the detector muon record.

Finally,
P (K | ID) = P (ID | K) * P (K) / P (ID) = 0.07 * 0.10 / 0.1158 = 0.0604 = probability that a detector record of a muon corresponds to a real kaon
= fraction of real kaons in the detector muon record.

13 'Posterior Odds'

If we form the ratio of P (μ | ID) to P (π | ID), using Bayes' theorem, the denominator P (ID) cancels, and we have

[P (μ | ID) / P (π | ID)] = (P (ID | μ) * P (μ)) / (P (ID | π) * P (π))

The ratio on the left is then known from the prior knowledge of P (μ) and P (π) and from the prior knowledge of the calibration of the detector relevant to muons and pions, and one doesn't need to know the total number of particle types present in the sample.

14 Example of Bayes' Theorem in Epidemiology

This example is from Cowan's Trieste slides, pdf. p. 10, and from his textbook, pdf. p 16.

"As an example [of the use of Bayes' theorem], consider a disease which is known to be carried by 0.1 % of the population, i.e. the prior probabilities to have the disease or not are:

P(disease) = 0.001,
P(no disease) = 0.999.

A test is developed which yields a positive result with a probability of 98%, given that the person carries the disease, i.e.

P(+ I disease) = 0.98 = probability of correct disease detection by the test (comes back positive) for a person with the disease,
P(- I disease) = 0.02 = probability the test will come back negative for a person who has the disease.
The probability of having a wrong result if you carry the disease is 2%.

Suppose there is also a 3% probability, however, to obtain a positive result for a person without the disease,

P(+ I no disease) = 0.03, which says the probability of having a wrong result if you have no disease is 3%,
P(- I no disease) = 0.97 = probability test will come back negative for a person without the disease.

What is the probability that you have the disease if your test result is positive? We need to calculate P (disease | +).

According to Bayes' theorem, this is given by

P (disease | +) = P (+ | disease)*P (disease)/P (+) = 0.98 * 0.001/P (+) = 0.00098/P (+).

The denominator is
P (+) = P (+ | disease) * P (disease) + P (+ | no disease) * P (no disease)
= 0.98 * 0.001 + 0.03 * 0.999 = 0.03095

Finally, P (disease | +) = 0.00098 / 0.03095 = 0.0317 ~ 0.032.

"The probability that you have the disease, given a positive test result, is only 3.2%. This may be surprising, since the probability of having a wrong result is only 2% if you carry the disease [P(- I disease) = 0.02] and 3% if you do not [P(+ I no disease) = 0.03] .

But the prior probability is very low, 0.1 % [P(disease) = 0.001], which leads to a posterior probability of only 3.2%."