Relative probabilities

1. Purpose

This article proposes that we should make more use of the concept of *relative probabilities*, particularly if we are using Bayesian methods. Relative probabilities are not really a new idea, though probability theory is not currently stated using them.

2. Examples of relative probabilities in everyday use

In probability theory, probabilities are numbers between 0 and 1 inclusive, but in everyday language we often express chances using numbers outside this range. For example, someone might say something is 'fifty-fifty', meaning that both of two outcomes are equally likely. In probability language this would be a distribution something like: $\{(win, 0.5), (lose, 0.5)\}$, with probabilities in the range 0 to 1 inclusive, adding up to 1 overall. However, what the person actually said was 'fifty-fifty', which would look like this: $\{(win, 50), (lose, 50)\}$.

Similarly, a bookie might give odds of 4 to 1, meaning $\{(win, 1), (lose, 4)\}$, which in probability numbers would be $\{(win, 0.2), (lose, 0.8)\}$.

From these familiar examples we can see that numbers against outcomes do not have to be in the range 0 to 1 inclusive to express the distribution. What matters is how the numbers compare to each other.

3. Relative probability distributions

Relative probabilities only make sense as part of a probability distribution of some kind. A distribution will show the relative probabilities of alternative possibilities in such a way that the relative probability numbers are in proportion to probabilities.

For example, if a fair coin is flipped it is traditional to say that the probability distribution is 0.5 for heads and 0.5 for tails. These probabilities add up to 1. Relative probability distributions conveying the same information might give 50 for heads and 50 for tails, or 1 for heads and 1 for tails, or perhaps 17 for heads and 17 for tails. The sum of the relative probabilities can change but the ratios between them do not.

3.1 Definition

The conditions for a valid relative probability distribution are that all numbers it gives must be greater than or equal to zero and at least one must be greater than zero.

If you don't like logic symbols please ignore the next paragraph, which uses established mathematical symbols in the style of Z (see Spivey, 1989). These symbols define a function that returns 'true' if a distribution meets the requirements for a relative probability distribution and false otherwise. This is a way to specify the conditions for a valid relative probability distribution.

[X]

$$isRelProbDist : (X \rightarrow \mathbb{R}) \rightarrow BOOLEAN$$

$$\forall\, f : X \rightarrow \mathbb{R} \cdot$$
$$isRelProbDist[f] \Leftrightarrow$$
$$\big((\forall\, x : X \mid x \in \mathrm{dom}[f] \cdot f[x] \geq 0\big) \wedge$$
$$(\exists\, x : X \mid x \in \mathrm{dom}[f] \cdot f[x] > 0))$$

(Translated literally the symbols of the first line mean: $isRelProbDist$ is a function (shown by the second $\rightarrow$) that takes as input a function from anything (X) to Real numbers ($\mathbb{R}$), and returns 'true' or 'false' ($BOOLEAN$). The second line says: for all functions from X to Reals ($\forall\, f : X \rightarrow \mathbb{R}$), saying $isRelProbDist[f]$ about that function means that, for all inputs to the function ($\forall\, x : X \mid x \in \mathrm{dom}[f]$), the result given by the function is greater than or equal to zero, and also, there is at least one input to the function ($\exists\ x : X \mid x \in \mathrm{dom}[f]$) for which the function gives a number greater than zero.)

3.2 Scaling

If all the relative probabilities in a relative probability distribution are multiplied by the same positive number then the result is another relative probability distribution holding the same probabilistic information as the first.

[X]

$scaled : ((X \rightarrow \mathbb{R}) \times k\ ) \rightarrow (X \rightarrow \mathbb{R})$

$\forall\ f : X \rightarrow \mathbb{R}, k : \mathbb{R} \mid isRelProbDist[f] \wedge k > 0\ \cdot$

$(\mathrm{dom}[scaled[f,k]] = \mathrm{dom}[f] \wedge$

$\forall\ x{:}\,X \mid x \in \mathrm{dom}[f] \cdot scaled[f,k][x] = f[x] \times k)$

3.3 Comparing relative probability distributions

Two relative probability distributions are equal if and only if they have the same domain (i.e. they give relative probabilities for the same things) and for each item in that domain each distribution gives the same number.

However, we are usually only concerned with the relative numbers. We can say that two distributions are similar if and only one can be scaled to equal the other.

[X]

$\_ \cong \_ : ((X \rightarrow \mathbb{R}) \times (X \rightarrow \mathbb{R})) \rightarrow BOOLEAN$

$\forall\ f, g : X \rightarrow \mathbb{R} \mid isRelProbDist[f] \wedge isRelProbDist[g]\ \cdot$

$(f \cong g \Leftrightarrow \exists\ k : \mathbb{R} \mid scaled[f,k] = g)$

The first line says: ≅ is a function that takes as input two functions from anything (X) to Real numbers, and returns 'true' or 'false'. The _≅_ notation shows that we use the ≅ sign by writing it between the two functions. The second and third line say that, for any two relative probability distributions, them being similar means it is possible to scale one to equal the other.

Because the domains of the distributions must be equal for two distributions to be similar, there is no way to scale between distributions of different types (e.g. between a probability mass function and a probability density function).

Two distributions being similar implies some properties that may be useful.

[X]

$\forall f, g : X \to \mathbb{R} \mid isRelProbDist[f] \wedge isRelProbDist[g] \cdot$

$\quad (\big(f = g \Leftrightarrow \big((\mathrm{dom}[f] = \mathrm{dom}[g]) \wedge \big((\forall x : X \mid x \in \mathrm{dom}[f] \cdot f[x] = g[x])\big)\big)\big) \wedge$

$\quad (f \cong g \Leftrightarrow$

$\qquad ((\mathrm{dom}[f] = \mathrm{dom}[g]) \wedge$

$\qquad (\forall x{:}X \mid x \in \mathrm{dom}[f] \cdot f[x] = 0 \Leftrightarrow g[x] = 0) \wedge$

$\qquad (\forall x, y : X \mid x \in \mathrm{dom}[f] \wedge y \in \mathrm{dom}[f] \wedge f[y] > 0 \cdot$

$$\frac{f[x]}{f[y]} = \frac{g[x]}{g[y]})))$$

(Again, the symbols can be translated literally. The first line introduces the remaining lines, which are continuations of the statement started on the first line. It says: for all pairs of functions from the same type of thing to Real numbers ($\forall f, g : X \to \mathbb{R}$), where those functions are both relative probability distributions... The second line then continues this by stating the first thing that is true for any such functions, which is that the two functions being equal (=) means that they both work for the same set of potential inputs and for each of those inputs the result is the same from both functions. The third and remaining lines say that: if the two functions are similar (≅) then they work for the same inputs, they give zero for the same inputs, and for every pair of those inputs that are not zero the ratio of the result for each is the same for each function.)

3.4 Normalizing

Normalizing a relative probability distribution means dividing its relative probabilities through by a number that produces either probabilities or probability densities.

There are three types of normalization for four types of distribution.

3.4.1 Probability measure ($P : (\mathbb{P}\ \Omega) \to \mathbb{R}$)

In the elementary theory of probability everything starts with a set of possibilities, which might be potential outcomes or potentially true answers to a question (depending on your perspective), combined with a set of sets of these possibilities that has special properties designed to make sure we can state probabilities for just about anything we might want to. These sets might be finite or infinite.

To this setup is added a probability measure that associates each set of possibilities with a number between 0 and 1 inclusive, in a way that ensures that the probability associated with the empty set is zero, the probability associated with the whole set of possibilities is 1, and probabilities can be added meaningfully.

If all the probabilities given by one of these probability measures were simply multiplied by a positive number (other than 0 or 1) then the ratios between these scaled probabilities for different sets would stay the same even though the scaled probability associated with the full set of possibilities would no longer be 1. Any such scaled probability measure would be a relative probability distribution giving relative probabilities.

To normalize a relative probability distribution (with the required properties for logical consistency) to a probability measure, simply divide all its relative probabilities by the relative probability associated with the set of all possibilities.

[X]

$normalizeP : (\mathbb{P}X \rightarrow \mathbb{R}) \rightarrow (\mathbb{P}X \rightarrow \mathbb{R})$

$\forall f : \mathbb{P}X \rightarrow \mathbb{R} \mid isRelProbDist[f] \ \cdot$

$$normalizeP[f] = scaled\left[f, \frac{1}{f[\cup \ \mathrm{dom}[f]]}\right]$$

3.4.2 Probability mass function ($PMF : \mathbb{R} \rightarrow \mathbb{R}$)

When possibilities are mapped to Real numbers using a function, the resulting numbers have a probability distribution implied by the underlying probability measure on the possibilities. If the number of Real numbers to which possibilities are mapped is finite or countably infinite then this implied probability distribution is called a probability mass function. It gives numbers between 0 and 1 that add up to 1.

A probability mass function is a relative probability distribution and, again, if the probabilities it gave were all scaled by a positive constant other than 0 or 1 then the resulting distribution would also be a relative probability distribution, but no longer a probability mass function.

To normalize a relative probability distribution to a probability mass function, divide all its relative probabilities by the sum of all of its relative probabilities, if you can find it.

[X]

$normalizeM : (\mathbb{R} \rightarrow \mathbb{R}) \rightarrow (\mathbb{R} \rightarrow \mathbb{R})$

$\forall f : \mathbb{R} \rightarrow \mathbb{R} \mid isRelProbDist[f] \ \cdot$

$$normalizeM[f] = scaled\left[f, \frac{1}{sum\left[y : X \mid y \in \mathrm{dom}[f] \cdot f[y]\right]}\right]$$

3.4.3 Probability density function ($PDF : \mathbb{R} \rightarrow \mathbb{R}$)

Like probability mass functions, probability density functions come into play when possibilities are mapped to Real numbers. If the number of Real numbers to which possibilities are mapped is uncountably infinite (and some other conditions are met) then the implied probability distribution is called a probability density function. It gives numbers greater than or equal to 0, and they integrate to 1.

Once again, scaling these probability densities gives a valid relative probability distribution conveying the same information, but it no longer integrates to 1, so it no longer qualifies as a probability density function.

To normalize a relative probability distribution to a probability density function, divide all its relative probabilities by the integral of its relative probabilities, if you can find it.

[X]

$$normalizeD : (\mathbb{R} \to \mathbb{R}) \to (\mathbb{R} \to \mathbb{R})$$

$$\forall\, f : \mathbb{R} \to \mathbb{R} \mid isRelProbDist[f] \;\cdot$$

$$normalizeD[f] = scaled\left[f, \frac{1}{\int_{\mathrm{dom}[f]} f}\right]$$

3.4.4 Cumulative probability function ($CPF\!: (\mathbb{P}\,\boldsymbol{\Omega}) \to \mathbb{R}$)

A cumulative probability function requires the set of possibilities to be put into an order. The cumulative probability function then takes as input a possibility in that order and returns the probability of the set of possibilities up to and including that possibility. Cumulative probability functions have useful flexibility. They can represent the information in a probability mass function, a probability density function, or a function that is a hybrid of the two with some possibilities having finite probability and others being linked only with densities.

Cumulative relative probability functions are normalized by dividing through by the relative probability assigned to the set of all possibilities (i.e. the last one in the order). This is the same as for normalizing probability measures.

3.5 Joint relative probability distributions

This is a very familiar idea. Sometimes we want to make probability models of situations where the possibilities are most easily seen as the Cartesian product of two sets. For example, if a coin is flipped twice then the possible outcomes can be constructed from the possible outcomes of one coin-flip. Instead of {H,T} (the possible outcomes for one coin-flip), the set of possibilities for two coin-flips is {(H,H),(H,T),(T,H),(T,T)} i.e. all the combinations.

$$\{H, T\} \times \{H, T\} = \{(\mathrm{H}, \mathrm{H}), (\mathrm{H}, \mathrm{T}), (\mathrm{T}, \mathrm{H}), (\mathrm{T}, \mathrm{T})\}$$

When the two parts of each possibility are probabilistically independent of each other then the compound model can be created by combining the component models. If the component probability distributions are replaced by relative probability distributions then multiplication still works just the same, producing a relative probability distribution.

[X,Y]

$$compoundI : ((X \to \mathbb{R}) \times (Y \to \mathbb{R})) \to \big((X \times Y) \to \mathbb{R}\big)$$

$$\forall f : X \to \mathbb{R}, g : Y \to \mathbb{R} \mid isRelProbDist[f] \wedge isRelProbDist[g] \cdot$$
$$(\mathrm{dom}[compoundI[f,g]] = \mathrm{dom}[f] \times \mathrm{dom}[g] \wedge$$
$$\forall x{:}X, y : Y \mid x \in \mathrm{dom}[f] \wedge y \in \mathrm{dom}[g] \cdot compoundI[f,g][(x,y)] = f[x] \times g[y])$$

This can be extended to three or more models combined.

With compound possibilities like this there are more different ways that subsets of the possibilities can be stated. For example, with two coin-flips we can talk about all possibilities where both coins gave the same result, or the possibilities where the results were different. Sometimes it is useful to be able to refer to the pairs within a domain that have a particular value on the left, or on the right. For example, the possibilities where the first flip gave H. A notation for this uses restriction of the domain or range. For example:

$$C_2 = \{(\mathrm{H},\mathrm{H}),(\mathrm{H},\mathrm{T}),(\mathrm{T},\mathrm{H}),(\mathrm{T},\mathrm{T})\}$$
$$C_2 \rhd \{\mathrm{H}\} = \{(\mathrm{H},\mathrm{H}),(\mathrm{T},\mathrm{H})\}$$
$$\{\mathrm{H}\} \lhd C_2 = \{(\mathrm{H},\mathrm{H}),(\mathrm{H},\mathrm{T})\}$$

3.6 Relative conditional probabilities and Bayes Theorem

If we take a relative probability distribution and restrict its domain to be just a subset of what it was then the remaining distribution is also a relative probability distribution, now conditional on the restricted domain. The ratios between the relative probabilities within the remaining domain are unchanged by the restriction, so information about relative probability between the remaining possibilities is preserved.

On the other hand, the increase in the probability that should be attached to each remaining possibility (now that some others have been eliminated) is not reflected in a change in the relative probabilities.

[X]

$$\_ \mid \_ : ((X \to \mathbb{R}) \times \mathbb{P}X) \to (X \to \mathbb{R})$$

$$\forall f : X \to \mathbb{R}, c : \mathbb{P}X \mid isRelProbDist[f] \cdot$$
$$((\mathrm{dom}[f|c] = \mathrm{dom}[f] \cap c) \wedge$$
$$(\forall x : X \mid x \in \mathrm{dom}[f|c] \cdot f[x|c] = f[x]))$$

In other words, a subset of a relative probability distribution is a conditional relative probability distribution. Other relative probability distributions similar to (i.e. ≅) this one would also contain the same probabilistic information.

If a relative probability distribution is normalized then it becomes a conditional probability distribution of some kind.

Some interesting details emerge with conditional relative probability distributions created from joint relative probability distributions. To take a typical example, imagine a joint relative probability distribution where the domain of the distribution is the Cartesian product of a set of hypotheses and a set of possible pieces of evidence. In other words, it is every combination of a hypothesis and a piece of evidence. This is the typical situation with statistical inference.

You can imagine this joint relative probability distribution as a table with rows for hypotheses and columns for pieces of evidence. Each cell in the table contains a number that is the relative probability for that hypothesis and piece of evidence. The table could be very large, perhaps infinitely so, but the joint relative probability distribution still gives finite numbers for every combination.

If a particular hypothesis, h, is chosen and the distribution is restricted to just those combinations of hypothesis and evidence that include h then the selected row of numbers is a conditional relative probability distribution using exactly the numbers that were there before. However, these now indicate the relative probability of each piece of evidence being seen given that h is true.

Likewise, if a particular piece of evidence, e, is chosen and the distribution is restricted to just those combinations of hypothesis and evidence that include e then the selected column of numbers is a conditional relative probability distribution showing the probability that each hypothesis is true given that e was seen.

If we want the relative probability for a particular combination of hypothesis and evidence then it does not matter if we (1) get it directly by looking for that particular combination, (2) selected the hypothesis first to get a conditional relative probability distribution and then select the pair with the relevant piece of evidence, or (3) select the evidence first to restrict the distribution and then select the pair with the relevant hypothesis. The relative probability value found will be the same.

Unfortunately, there are times when we do not have the required relative probability distribution to consult. Instead, we have a distribution for the hypotheses (ignoring evidence) and normalized distributions for the evidence given each of the hypotheses in turn. This is when Bayes Theorem is used.

To explain this with symbols, I begin with defining the basic types and abbreviating them.

$[EVIDENCE, HYPOTHESIS]$

$E == EVIDENCE$

$H == HYPOTHESIS$

Now introduce a relative probability distribution giving relative probabilities for each combination of hypothesis and evidence. (Note that this does not give relative probabilities for sets of possibilities, otherwise known as 'events' in probability theory.)

$R: (E \times H) \to \mathbb{R}$

$H_R: \mathbb{P}H$

$E_R: \mathbb{P}E$

$B : \mathbb{P}(E \times H)$

$B = H_R \times E_R$

$B = dom[R]$

$isRelProbDist[R]$

The next part says that we can get to the relative probability for a combination of hypothesis and evidence by three equivalent routes: (1) directly, (2) by restricting to just the possibilities involving the hypothesis first, or (3) by restricting to just the possibilities involving the evidence first.

$\forall e: E, h: H \mid (h, e) \in B \cdot$

$$(R[(h, e)] = R|(B \rhd e)[(h, e)] \wedge$$
$$R[(h, e)] = R|(h \lhd B)[(h, e)])$$

Hence,

$\forall e: E, h: H \mid (h, e) \in B \cdot$

$$R|(B \rhd e)[(h, e)] = R|(h \lhd B)[(h, e)]$$

The conditional relative probability distribution on the right-hand side has not been normalized. It can be normalized for each hypothesis by dividing through by a normalizing constant for that hypothesis. Let this be represented by a new relative probability distribution that has a relative probability for each hypothesis equal to the required normalizing constant multiplied by a number that is constant across all hypotheses.

$R_H: H \to \mathbb{R}$

$\mathrm{dom}[R_H] = H_R$

Inserting this we get a formula that is Bayes Theorem in slightly unusual notation but holding regardless of whether probabilities or densities are involved.

$\forall e: E, h: H, k: \mathbb{R} \mid (h, e) \in B \wedge k > 0 \cdot$

$$R|(B \rhd e)[(h, e)] = \frac{R_H[h]}{k} \times \frac{R|(h \lhd B)[(h, e)]}{\frac{R_H[h]}{k}}$$

In this the term $R_H[h]$ is a relative version of the prior distribution of the hypotheses before evidence is considered. The fraction to its right represents the normalized probabilities provided by a likelihood function i.e. the probability of seeing the evidence for each possible hypothesis.

It is not necessary to know the normalizing constant to use this formula because the likelihood function can usually be found without it and the resulting relative probability distribution will be similar to the one on the left-hand side above. This is shown as follows:

$\forall e: E, h: H, k: \mathbb{R} \mid (h, e) \in B \cdot$

$$R|(B \rhd e)[(h,e)] \cong R_H[h] \times \frac{R|(h \lhd B)[(h,e)]}{\frac{R_H[h]}{k}}$$

3.7 A proof strategy using relative probability distributions

Since probability measures, mass functions, and density functions are all relative probability distributions, a result established for relative probability distributions (i.e. using only the properties of relative probability distributions) would also be true for all the more specialized distributions.

4. Exploring a theoretical problem involving infinity

It is well established that probability distributions can provide probability numbers for each of a finite or countably infinite set of possibilities, but if the set of outcomes is uncountably infinite then this method does not work. Probability densities must be used instead.

Probability density functions show the relative probabilities of outcomes but not their probabilities. Probability density functions are normalized to integrate to 1 but the densities for all the outcomes cannot be summed.

4.1 A problem

What happens to the probabilities of each possibility in an uncountably infinite set of possibilities? I have seen it written that the probabilities of the outcomes are all zero but there are problems with this idea.

- Surely there is a difference between the impossibility of throwing 7 with an ordinary six-sided die and the possibility of being *exactly* 1.5m tall, even though both supposedly have a probability of zero.
- How can it be that we know one of infinitely many outcomes, all with probability of zero, must happen?
- How can it be that the sum of many zeros is 1, even if there are infinitely many of them? If we add up some of the zeroes the result is zero and we are where we started. It seems we can never make progress.
- Furthermore, how can it be that the probability densities successfully preserve the relative probability distribution but the probabilities themselves have somehow lost that ability?

One demonstration of the idea that all the probabilities are zero uses the cumulative probability function, $F[x]$, assumed to be continuous, and goes like this:

$$\forall x, \varepsilon : \mathbb{R} \cdot 0 \leq P(X = x) \leq P(x - \varepsilon < X \leq x)$$

$$\forall x, \varepsilon : \mathbb{R} \cdot 0 \leq P(X = x) \leq F[x] - F[x - \varepsilon]$$

$$\forall x : \mathbb{R} \cdot 0 \leq P(X = x) \leq \lim_{\varepsilon \downarrow 0} [F[x] - F[x - \varepsilon]]$$

$$\forall x : \mathbb{R} \cdot 0 \leq P(X = x) \leq 0$$

$$\forall x : \mathbb{R} \cdot P(X = x) = 0$$

The loophole in this argument occurs where the inequality goes from

$$F[x] - F[x - \varepsilon]$$

to

$$\lim_{\varepsilon \downarrow 0} [F[x] - F[x - \varepsilon]].$$

In this case the limit is actually below the lowest value of $F[x] - F[x - \varepsilon]$ as ε gets as close as you like to zero but does not actually get there, so the infinitesimal gap into which $P(X = x)$ could fall is removed incorrectly.

Does any of this make sense? I'm not sure.

4.2 Relative probabilities survive infinity

Perhaps a proper mathematician could probe these issues using relative probabilities. For example, consider the cumulative probability function above but now consider two points, x and y, again with ε. What happens to the ratio:

$$\lim_{\varepsilon \downarrow 0} \frac{F[x] - F[x - \varepsilon]}{F[y] - F[y - \varepsilon]}$$

as ε tends towards zero (but does not reach it)? The ratio moves towards a limit matching the ratio of the probability densities at x and y. Looked at this way the information in the probabilities is preserved even though the probabilities themselves become too small and undefined to work with.

Another way to probe these issues is to consider what happens to a probability distribution as additional possibilities are added to it without limit. To set the scene for this, here are some points of principle.

The first point is that limits of ratios can have stable and well-defined values even though the elements that make them up do not.

To illustrate, consider the well-known result that:

$$\lim_{x \uparrow} \frac{1}{x} = 0$$

(The $\lim$ notation means the limit 'as x rises without limit'. See Leitch 2016.)

Although zero is the limit as x grows without limit, the function never actually reaches zero. This is a typical property of limits. In this case we can get as close as we like to zero, but not quite there.

In contrast, consider this:

$$\lim_{x \uparrow} \frac{1/x}{2/x}$$

For all values of x except for zero, this fraction is equal to 0.5 and so the limit is also 0.5. In contrast to the previous function, this is a limit that the function does reach (and in fact almost never strays from).

The same basic principle can be seen with probability distributions based on sets of outcomes that are augmented without limit with additional potential outcomes. Imagine we start with a set of outcomes that contains just two outcomes, ω_a and ω_b, with positive, finite real numbers assigned to them, r_a and r_b respectively. These represent relative probabilities so they can be normalized so that:

$$p[2][\omega_a] = \frac{r_a}{r_a + r_b}, \qquad p[2][\omega_b] = \frac{r_b}{r_a + r_b}$$

The notation $p[2]$ means the probability distribution for the case with two outcomes.

The next step is to add another outcome, ω_c with its own relative probability, again finite, r_c. Once again, normalize these:

$$p[3][\omega_a] = \frac{r_a}{r_a + r_b + r_c}, \qquad p[3][\omega_b] = \frac{r_b}{r_a + r_b + r_c}, \qquad p[3][\omega_c] = \frac{r_c}{r_a + r_b + r_c}$$

All these probabilities are lower due to the larger normalizing constant. If we continue adding outcomes like this then the probabilities will drift lower and lower, and if we continue adding outcomes without limit then the probabilities will become infinitesimal. That is, they will have no definite value but will be as close as you like to zero, and yet not zero.

$$\lim_{n\uparrow} p[n][\omega_a] = 0, \text{and } \lim_{n\uparrow} p[n][\omega_b] = 0,$$

Once again, the limit is zero but zero is not actually reached.

In contrast, the *ratio* between $p[n][\omega_a]$ and $p[n][\omega_b]$ will stay fixed at ${}^{r_a}/_{r_b}$ regardless of how many outcomes are added. So, even though the probabilities will dwindle to infinitesimal values we cannot work with, the ratios between them stay fixed, preserving the information in the relative probability distribution.

$$\lim_{n\uparrow} \frac{p[n][\omega_a]}{p[n][\omega_b]} = \frac{r_a}{r_b}$$

In this description the extra outcomes are added one at a time but there is nothing restrictive about this. There is nothing special about the speed of adding outcomes. It can be done to an unlimited extent and the result is the same. There is no reason for thinking that a different situation suddenly emerges when a particular number of outcomes is reached.

Since we could do this for any pair of outcomes, provided the probability on the bottom of the fraction is not zero, the shape of the relative probability distribution is preserved in a sense as more potential outcomes are added, even though the individual probabilities dwindle towards zero.

5. References

Leitch, M. (2016). Eliminating infinity. Available online at: http://www.workinginuncertainty.co.uk/infinite.pdf

Spivey, J.M. (1989). *The Z Notation*. Prentice-Hall, Englewood Cliffs, NJ. Available online at: https://spivey.oriel.ox.ac.uk/wiki2/files/zrm/zrm.pdf

6. Version history

Version 1: 2014.

Version 2: 2019. Expansion of material on Bayes, infinity, and the outline of a potential proof of Bayes rule using relative probabilities.

Version 3: 2022. Expansion of the simpler material and a replacement of the material on Bayes that uses conditional relative probability distributions. The approach proposed in 2019 was wrong.