Negative evidence and quantum mechanics

Mark J. Wierman
Creighton University - wierman@creighton.edu

Abstract

Quantum mechanics is an extremely successful theory in physics for the prediction of the microverse. The Wigner function which corresponds to a probability distribution is unfortunately negative in certain regions. Evidence theory, as originated by Dempster, corresponds in some cases to a negative basic probability assignment, a result that Shafer disallows. This paper shows that allowing negative evidence permits the modeling of interference effects in the two slit experiment.

1 Quantum mechanics and negative probability

Quantum mechanics is one of the most successful physical theories in the history of science. Quantum mechanics predictions about the results of an experiment are unerringly accurate. However, quantum mechanical predictions about the result of an experiment are probabilistic in nature. Quantum mechanics cannot predict whether or not a particle will decay at a specific moment. Quantum mechanics can predict what percentage of a mass of particles will decay over a given time period with spectacular accuracy.

One of the many interesting results in quantum mechanics are negative probabilities [1]. Feynman states that the probability of an event that can actually occur never turns out to be negative, but that the probabilities in the intermediate calculations can have negative values.

For example suppose an apples vendor starts with 6 apples. At noon he is resupplied with 5 additional apples. During the day he sells a total of 8 apples. The result is 6-8+5 apples, or 3 apples. An intermediate calculation if we group the terms appropriately gives $6 - 8 = -2$ apples or negative two apples. Now, no one can have a negative number of apples since this is simply the result of a bookkeeper, starting with an initial stock, first subtracting the sales, and then adding the restock. The bookkeeper's intermediate results are negative but this is not reflected in the real world. Another example, from statistical signal processing, occurs when we take a function g and use its Fourier transform G (which contains sin and cos terms which may be negative in regions) in processing to remove noise. In the end we inverse-transform the composition of G and the filter F to recover the source minus the noise.

A very popular and accessible review of Feynman's paper was published in [2]. It gives a simplified example of the troubles probability theory has when applied to quantum mechanical applications. Imagine we have two coins. If either coin is flipped independently in a black box, then the ratio of heads to tails is 1:2. For either coin we have that the probability of heads is one-third and the probability of tails is two-thirds, or $P(H_i) = \frac{1}{3}$ and $P(T_i) = \frac{2}{3}$ for $i = 1, 2$. However when we flip both coins together in the black box it is always the case that the coin one is heads and coin two is tails or coin one is tails and coin two is heads. That is, the events that both coins are heads or both coins are tails never occurs. Thus the probabilities of the events HH, HT, TH, and TT, are such that $P(HH, TT) = 0$ and $P(HT, TH) = 1$. Table 1 shows the results of various thinking. The marginals induce the probability distribution shown in Table (1.A) where α is in the interval 0 to $\frac{1}{3}$. The Table (1.B) shows the noninteractive solution where $P(HH) = P(H_1) * P(H_2)$, etc. The Table (1.C) shows a probability distribution that satisfies the requirements that HH and TT never occur together. The only way to get a solution to both requirements is to abandon positivity, a result shown in Table (1.D).

Lowe [3] mentions that negative probabilities also appear in neural networks. Using a probabilistic learning rule that conditions individual neurons to learn, it turns out that some inner nodes have negative weights but that the output nodes always end up producing positive results. He also mentions that it is impossible, see [4] and [5], to generate approxima-

A	H	T
H	α	$\frac{1}{3}-\alpha$
T	$\frac{1}{3}-\alpha$	$\frac{1}{3}+\alpha$

B	H	T
H	$\frac{1}{9}$	$\frac{2}{9}$
T	$\frac{2}{9}$	$\frac{4}{9}$

C	H	T
H	0	p
T	$1-p$	0

D	H	T
H	$\frac{1}{6}$	$\frac{1}{2}$
T	$\frac{1}{2}$	$\frac{-1}{6}$

Table 1: Joint probabilities of the outcomesd of tossing two coins. (A) Joint probailities induced by the marginal ditributions of $P(H) = \frac{1}{3}$ and $P(T) = \frac{2}{3}$ where α ranges from 0 to $\frac{1}{3}$. (B) The noninteractive solution of (A). (C) Marginal probabilites where the $P(H) = P(T) = 0$ (C) Pseudo probabilites that satisfy both the requirements of both (A) and (C).

tions to (unknown) pdfs which simultaneously satisfy the three 'axioms':

- Realness
- Positivity
- Reproduce correct expectation values.

If Lowe's proposition is correct then the first (the values are real numbers) and the third (the results correspond to the experiment) properties are not disposable, and positivity is the property we must abandon.

2 Dempster and negative evidence

Wierman, in [6] reexamines the connection between the theory of upper and lower probabilities promulgated by Dempster [7][8] and evidence theory as proposed by Shafer [9]. In Dempster's original work it is assumed that we know certain facts that can be expressed as expected values of events on some unknown probability distribution. These facts can be expressed as constraints upon the unknown probability distribution. Dempster then defines an upper and a lower probability measure upon events that are the inf and sup of their expected values of the event over all probability distributions compatible with the given constraints.

Example 1 *Let $X = \{a, b, c, d\}$ with given data $P(a, b) = P(b, c) = \frac{2}{3}$. It is not to difficult to derive that the distributions on X that could reproduce the data are of the form $\langle \frac{2}{3}-\alpha, \alpha, \frac{2}{3}-\alpha, \alpha-\frac{1}{3}\rangle$ where $\alpha \in \left[\frac{1}{3}, \frac{2}{3}\right]$. Given these distributions, maximal and minimal probabilities for all events can be calculated. The results are presented in Table (2).*

a	b	c	d	P	LP	UP	$m?$
0	0	0	0	0	0	0	.
0	0	0	1	$\alpha-\frac{1}{3}$	0	$\frac{1}{3}$	0
0	0	1	0	$\frac{2}{3}-\alpha$	0	$\frac{1}{3}$	0
0	0	1	1	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
0	1	0	0	α	$\frac{1}{3}$	$\frac{2}{3}$	$\frac{1}{3}$
0	1	0	1	$2\alpha-\frac{1}{3}$	$\frac{1}{3}$	1	0
0	1	1	0	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{1}{3}$
0	1	1	1	$\frac{1}{3}+\alpha$	$\frac{2}{3}$	1	$\frac{-1}{3}$
1	0	0	0	$\frac{2}{3}-\alpha$	0	$\frac{1}{3}$	0
1	0	0	1	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
1	0	1	0	$\frac{4}{3}-2\alpha$	0	$\frac{2}{3}$	0
1	0	1	1	$1-\alpha$	$\frac{1}{3}$	$\frac{2}{3}$	$\frac{-1}{3}$
1	1	0	0	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{1}{3}$
1	1	0	1	$\frac{1}{3}+\alpha$	$\frac{2}{3}$	1	$\frac{-1}{3}$
1	1	1	0	$\frac{4}{3}-\alpha$	$\frac{2}{3}$	1	$\frac{-1}{3}$
1	1	1	1	1	1	1	$\frac{2}{3}$

Table 2: The Basic Probability Assignment, Lower and Upper probabilities of Example 1

When Shafer recast Dempster's work in his book Evidence Theory he retitled lower probabilities as belief, or Bel, and the upper probability as plausibility, or Pl.. He also claimed that they could be calculated from a Basic Probability Assignment (BPA), labeled m, upon the power set of X, $m : \mathcal{P}(X) \to [0, 1]$,which is required to satisfy two conditions: (i) $m(\emptyset) = 0$ and (ii) $\sum_{A\in\mathcal{P}(X)} m(A) = 1$.

For each set $A \in \mathcal{P}(X)$, the value $m(A)$ expresses the proportion to which all available and relevant evidence supports the claim that a particular element of X, whose characterization in terms of relevant attributes is deficient, belongs to the set A. This value, $m(A)$, pertains solely to one set, set A; it does not imply any additional claims regarding subsets of A. If there is some additional evidence supporting the claim that the element belongs to a subset of A, say $B \subseteq A$, it must be expressed by another value $m(B)$.

Given a BPA m, the corresponding belief measure and plausibility measure are determined for all sets $A \in \mathcal{P}(X)$ by the formulas $\mathrm{Bel}(A) = \sum_{B\subseteq A} m(B)$,and $\mathrm{Pl}(A) = \sum_{B\cap A\neq\emptyset} m(B)$. It can be shown that the plausibility of an event is one minus the belief of the complement of that event, that is

$\mathrm{Pl}(A) = 1 - \mathrm{Bel}(A^c)$. Inverse procedures are also possible. Given, for example, a belief measure Bel, the corresponding basic probability assignment m is determined for all $A \in \mathcal{P}(X)$ by the formula

$$m(A) = \sum_{B \subseteq A} (-1)^{|A-B|} \mathrm{Bel}(B), \qquad (1)$$

where $|A - B|$ is the cardinality of the set difference of A and B, as proven by Shafer. Thus each of the three function, m, Bel and Pl, is sufficient to determine the other two.

Using the formula in Eq. (1) and associating the lower probability of Dempster with the belief measure of Shafer, $\mathrm{Bel} \equiv LP$, we calculate the basic probability assignment in the last column of Table (2). This gives, for example, $m(b, c, d) = \frac{-1}{3}$, a great difficulty.

We can immediately conclude that Shafer's version of evidence theory does not completely capture Dempster's original model. Not unless we abandon positivity, that is, unless we abandon the requirement that $m : \mathcal{P}(X) \to [0, 1]$.

Remark 2 *If we abandon the requirement that the empty set can contain no information, that* $m(\emptyset) = 0$ *since the empty set has no elements, then we get a variation on evidence theory called O-Theory [10].*

A revised theory of evidence that allows for a negative BPA does not allow for negative beliefs. Here we have a terminology problem, beliefs and plausibilities are Lower and Upper Probabilities respectively and cannot be negative. While termed a BPA by Shafer, the function m that can be calculated from Bel using the formula in Eq. (1) is simply the combinatorial result of the Mobius inversion formula. A better name for the function m then is a basic evidential weight, or BEW, which in our case is a function $m : \mathcal{P}(X) \to [-1, 1]$,which is required to satisfy three conditions: (i) $m(\emptyset) = 0$, (ii) $\sum_{A \in \mathcal{P}(X)} m(A) = 1$ and (iii) for all $A \subseteq X$ we have $0 \leq \sum_{B | B \subseteq A} m(B) \leq 1$. Under these new restrictions both Bel and Pl are still fuzzy measures, that is monotone, positive set functions.

It is not always easy to create BEWs that satisfy requirement (iii). However, if we have start from a collection of lower probabilities, LPs, as in Dempster and Example (1), we can always calculate a BEW using Eq. (1) with the assumption that $\mathrm{Bel} \equiv LP$ that satisfies all the requirements of a revised evidence theory.

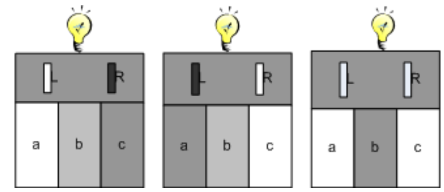

Figure 1: The situations (A) Left slit only; (B) Right slit only; and (C) Both slits open.

3 Quantum mechanics and negative evidence

If we abandon the positivity requirement for evidence theory and instead use a BEW that ranges over the interval $[-1, 1]$ then we arrive at a theory that has some interesting applications. Consider for example the two slit experiment in physics where a polarized light is behind a screen containing two slits.

If only the left slit L is open then the photons spread out from it with a distribution that depends only on distance from the opening. Similarly if only the right slit R is open then the photons spread out from it with a distribution that depends only on distance from the opening. However if both slits are open then a curious interaction in the wave nature of particles (and the particle-wave model is at the heart of quantum mechanics) causes dark and light bands. If we consider a dark region, b, in between the two slits then the probabilities of a particles landing there when either of the two slits is open is some positive probability β, or $P(b|L) = P(b|R) = \beta$. However the probability of a photon landing in b if both the slits are open is zero, so that $P(b|L, R) = 0$. This defies the monotone nature of probabilities. If we abandon positivity in evidence theory we can give a fascinating model of the interaction of the photons as presented in Table (3). The BEW still sums to one, and note that all the lower probabilities or Beliefs are positive. The interesting BEWs are the values $m(b|L) = m(b|R) = 0.1$ and $m(b|L, R) = -0.2$. We can interpret $m(b|L, R) = -0.2$ as saying that when slits L and R are both open then the weight that flows to event $(b|L, R)$ from $m(b|L)$ and $m(b|L)$ is canceled out. The corresponding Beliefs are $Bel(b|L) = Bel(b|R) = 0.1$ and $Bel(b|L, R) = 0.0$. All of the BEWs and Bel$s$ are spelled out in Table (3).

Remark 3 *It is not possible to get a better model of the coin flipping example using evidence theory. The*

Set	Slits	BEW	Bel
a	L	0.2	0.2
b	L	0.1	0.1
c	L	0.05	0.05
a,b	L	0	0.3
a,c	L	0	0.25
b,c	L	0	0.15
a,b,c	L	0	0.35
a	R	0.05	0.05
b	R	0.1	0.1
c	R	0.2	0.2
a,b	R	0	0.15
a,c	R	0	0.25
b,c	R	0	0.3
a,b,c	R	0	0.35
a	L,R	0.25	0.5
b	L,R	-0.2	0
c	L,R	0.25	0.5
a,b	L,R	0	0.5
a,c	L,R	0	1
b,c	L,R	0	0.5
a,b,c	L,R	0	1

Table 3: Evidence ditribution that produces the desired lower Beliefs

marginal constraints combined with the double flipping constraints allow for a single probability distribution, the one given in Table (1.D). For this example the upper and lower probabilities are simply the probabilities $P \equiv UP = LP$ and m is identical to the single quasi-probaility distribution given in Table (1.D).

4 Conclusions

In Dempster's original papers, lower probabilities are the inf of the probability of an event over a constrained set of probability distributions. We have seen that a Mobius inversion of the lower probabilities can produce a function m that contains values that are negative. Negative values in this function can be interpreted as restrictions on the flow of probability weights onto an event.

Negative evidential weights may be useful in modeling other events where there are simultaneity restrictions. For example, the Heisenberg Uncertainty Principle states that it is impossible to simultaneously measure the mass and momentum of a particle to arbitrary degree of precision.

We should mention that there is much disagreement in evidence theory about the process of conditioning or information fusion, see for example [11].

References

[1] R. A. Feynman, "Negative probability," in *Quantum implications : essays in honour of David Bohm*, 1987, pp. 3–18.

[2] T. P. Dietrich Leibfried and C. Monroe, "Shadows and mirrors: Reconstructing quantum states of atom motion," *Physics Today*, vol. 51, no. 4, pp. 22–28, 1998.

[3] D. Lowe, "Negative probability," *NCRG Seminar*, 1998.

[4] M. Rosenblatt, "Remarks on some nonparametric estimates of a density function," *Annals of Mathematical Statistics*, vol. 27, pp. 832–835, 1956.

[5] H. Yamato, "Sequential estimation of a continuous probability density function and the mode," *Bulletin of Mathematical Statistics*, vol. 14, pp. 1–12, 1971.

[6] M. J. Wierman, "Dempster $\neq$ shafer," in *JCIS 2002*, 2002.

[7] A. P. Dempster, "Upper and lower probabilities induced by a multivalued mapping," *Annals of Mathematical Statistics*, vol. 38, no. 2, pp. 325–339, 1967.

[8] ——, "Upper and lower probability inferences based on a sample from a finite univariate population," *Biometrika*, vol. 54, no. 3,4, pp. 515–528, 1967.

[9] G. Shafer, *A Mathematical Theory of Evidence.* Princeton, New Jersey: Princeton University Press, 1976.

[10] E. M. Oblow, "O-theory: A probabilistic alternative to fuzzy set theory." in *IPMU*, ser. Lecture Notes in Computer Science, B. Bouchon-Meunier and R. R. Yager, Eds., vol. 286. Springer, 1986, pp. 111–119.

[11] N. Friedman and J. Y. Halpern, "Belief revision: a critique," *Journal of Logic, Language and Information*, vol. 8, no. 4, pp. 401–20, 1999, publisher: Kluwer Academic Publishers, Netherlands.

[12] M. S. Bartlett, "Negative probability," *Proc. Cambridge Philos. Soc.*, vol. 41, pp. 71–73, 1945.

[13] J. H. Halton, "An interpretation of negative probabilities," *Proc. Cambridge Philos. Soc.*, vol. 62, pp. 83–86, 1966.