Why Product of Probabilities (Masses) for Independent Events? A Theorem

Vladik Kreinovich[1] and Scott Ferson[2]
[1]Department of Computer Science
University of Texas at El Paso
El Paso, TX 79968, USA, vladik@cs.utep.edu
[2]Applied Biomathematics
100 North Country Road
Setauket, NY 11733, USA, scott@ramas.com

Abstract

For independent events A and B, the probability $P(A\,\&\,B)$ is equal to the product of the corresponding probabilities: $P(A\&B)=P(A)\cdot P(B)$. It is well known that the product $f(a,b)=a\cdot b$ has the following property: once $\sum\limits_{i=1}^{n} P(A_i)=1$ and $\sum\limits_{j=1}^{m} P(B_j)=1$, the probabilities $P(A_i\,\&\,B_j)=f(P(A_i),P(B_j))$ also add to 1: $\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m} f(P(A_i),P(B_j))=1$. We prove that the product is the only function that satisfies this property, i.e., that if, vice versa, this property holds for some function $f(a,b)$, then this function f is the product. Thus, we provide an additional explanation of why for independent events, we multiply probabilities (or, in the Dempster-Shafer case, masses).

Product is normally used as a combination rule for independent events. For independent events A and B, the probability $P(A\,\&\,B)$ is equal to the product of the corresponding probabilities: $P(A\&B)=f(P(A),P(B))$, where the combination function is the product $f(a,b)=a\cdot b$; see, e.g., [5].

Similarly, in Dempster-Shafer theory (see, e.g., [2, 6]) one of the ways to combine the masses from two independent knowledge bases is to multiply them.

A reasonable property of the combination rule. Due to the additivity property of probability, if the events $A_1,\ldots,A_n$ form a partition of the universal set, i.e., if one of these events always occurs and no two can occur at the same time, then $\sum\limits_{i=1}^{n} P(A_i)=1$. If the events A_i form a partition and the events B_j

form a partition, then their combinations $A_i \,\&\, B_j$ also form a partition; indeed:

- since A_i and B_j form a partition, any situation belongs to one of A_i and to one of B_j, thus, for this situation, the corresponding event $A_i \,\&\, B_j$ holds;
- similarly, since the events A_i are mutually exclusive and the events B_j are mutually exclusive, the combinations $A_i \,\&\, B_j$ are also mutually exclusive.

It is therefore reasonable to expect that if the events A_i form a partition, i.e., $\sum\limits_{i=1}^{n} P(A_i) = 1$, and if events B_j form a partition, i.e., $\sum\limits_{j=1}^{m} P(B_j) = 1$, then the events $A_i \,\&\, B_j$ should also form a partition, i.e., $\sum\limits_{i=1}^{n} \sum\limits_{j=1}^{m} f(P(A_i), P(B_j)) = 1$.

In formal terms, the function $f : [0,1] \times [0,1] \to [0,1]$ that describes the combination rule should satisfy the following property:

$$\begin{gathered}\text{For every two finite sequences}\\ \text{of non-negative real numbers } (a_1, \ldots, a_n) \text{ and } (b_1, \ldots, b_m),\\ \text{if } \sum_{i=1}^{n} a_i = 1 \text{ and } \sum_{j=1}^{m} b_j = 1, \text{ then } \sum_{i=1}^{n} \sum_{j=1}^{m} f(a_i, b_j) = 1.\end{gathered} \tag{1}$$

What is known. It is well known that the product function $f(a,b) = a \cdot b$ satisfies the property (1). It is also known that many other possible combination functions, e.g., many t-norms that are different from the product (see, e.g., [3, 4]), do not satisfy this property.

What we will prove. In this paper, we prove that the product function is the only function that satisfies the above property. We will also prove a similar result for the case when we combine more than two events.

Theorem 1. *If a function $f : [0,1] \times [0,1] \to [0,1]$ satisfies the property* (1), *then this function is the product: $f(a,b) = a \cdot b$ for all a and b.*

Conclusion. Thus, we provide an additional explanation of why for independent events, we multiply probabilities (or, in the Dempster-Shafer case, masses).

Case of several events. Let $k \geq 2$ be an integer, and let $f : [0,1]^k \to [0,1]$ be a function of k variables. For such functions, we will consider the following property:

For every k finite sequences

of non-negative real numbers $(a^{(1)}_1, \ldots, a^{(1)}_{n_1}), \ldots, (a^{(k)}_1, \ldots, a^{(k)}_{n_k})$,

$$\text{if } \sum_{i_1=1}^{n_1} a^{(1)}_{i_1} = 1 \text{ and } \ldots \text{ and } \sum_{i_k=1}^{n_k} a^{(k)}_{i_k} = 1, \tag{2}$$

$$\text{then } \sum_{i_1=1}^{n_1} \ldots \sum_{i_k=1}^{n_k} f(a^{(1)}_{i_1}, \ldots, a^{(k)}_{i_k}) = 1.$$

Theorem 2. *If a function $f : [0,1]^k \to [0,1]$ satisfies the property (2), then this function is the product: $f(a_1, \ldots, a_k) = a_1 \cdot \ldots \cdot a_k$ for all $a_1, \ldots, a_k$.*

Proof of the Theorems. The proof of Theorems 1 and 2 is based on the following Lemma:

Lemma. *Let a function $g : [0,1] \to R^+_0 \stackrel{\text{def}}{=} [0, \infty)$ satisfy the following property:*

For every finite sequence of non-negative real numbers $(a_1, \ldots, a_n)$,

$$\textit{if } \sum_{i=1}^{n} a_i = 1, \textit{ then } \sum_{i=1}^{n} g(a_i) = 1. \tag{3}$$

Then, $g(a) = a$ for every real number a.

Proof of the Lemma. Let us first consider the case when $n = 2$. In this case, the condition of the Lemma means that $a_1 + a_2 = 1$ implies $g(a_1) + g(a_2) = 1$, i.e., that $g(a_2) = 1 - g(a_1)$. The equality $a_1 + a_2 = 1$ means that $a_2 = 1 - a_1$, so the condition of the Lemma means that

$$g(1 - a_1) = 1 - g(a_1) \tag{4}$$

for all $a_1 \in [0,1]$.

For $n = 3$, we similarly conclude that $g(a_1) + g(a_2) + g(1 - (a_1 + a_2)) = 1$ for all $a_1 \geq 0$ and $a_2 \geq 0$ for which $a_1 + a_2 \leq 1$. Therefore, $g(a_1) + g(a_2) = 1 - g(1 - (a_1 + a_2))$. Due to (4), we have $1 - g(1 - (a_1 + a_2)) = g(a_1 + a_2)$, so the above property reads $g(a_1 + a_2) = g(a_1) + g(a_2)$. It is known (see, e.g., [1]) that every function g whose values are non-negative and which satisfies the above *additivity* property is linear, i.e., $g(a) = k \cdot a$ for some real number k. Substituting this expression for $g(a)$ into both sides of the formula (4), we conclude that $k = 1$, i.e., that $g(a) = a$. The Lemma is proven.

Completing the proof. Let us first prove Theorem 1. Let b_j be a sequence for which $\sum\limits_{j=1}^{m} b_j = 1$. For this sequence, let us introduce an auxiliary function $g(a) \stackrel{\rm def}{=} \sum\limits_{j=1}^{m} f(a, b_j)$. In terms of this function, the double sum in (1) takes the form $\sum\limits_{i=1}^{n} g(a_i)$, so the property (1) takes the form (3).

Since the values of the function f are non-negative, the new auxiliary function $g(a)$ has non-negative values as well. Due to Lemma, we now conclude that $g(a) = a$, i.e., that for every a, we have

$$\sum_{j=1}^{m} f(a, b_j) = a. \tag{5}$$

When $a = 0$, then, from the fact that $f(a, b) \ge 0$ for all b, we conclude that $f(a, b_j) = 0$ for all j – since the only way for a sum of non-negative numbers to be 0 is when each of these numbers is equal to 0. Thus, we conclude that $f(0, b) = 0$ for all b, i.e., that $f(a, b) = a \cdot b$ for $a = 0$.

When $a > 0$, then we can divide both sides of the formula (5) by a and get the following formula:

$$\sum_{j=1}^{m} \frac{f(a, b_j)}{a} = 1.$$

So, for every $a > 0$, the new auxiliary function $g(b) \stackrel{\rm def}{=} \dfrac{f(a, b)}{a}$ satisfies the following property:

For every finite sequence of non-negative real numbers $(b_1, \ldots, b_m)$,

$$\text{if } \sum_{j=1}^{m} b_j = 1, \text{ then } \sum_{j=1}^{m} g(b_j) = 1.$$

This is exactly the property (3), so, due to Lemma, $g(b) = b$ for every real number b. Since $g(a) = f(a, b)/a$, we conclude that $f(a, b) = a \cdot b$ for all a and b.

Theorem 2 can be now proved by induction over k. We have already proven this theorem for $k = 2$ – this case corresponds exactly to Theorem 1. Let us now assume that we have proved this result for $k - 1$, let us show how to prove it for k. For that, we first fix $k - 1$ sequences $(a^{(2)}_1, \ldots, a^{(2)}_{n_2}), \ldots, (a^{(k)}_1, \ldots, a^{(k)}_{n_k})$, and consider an auxiliary function $g(a) \stackrel{\rm def}{=} \sum\limits_{i_2=1}^{n_2} \ldots \sum\limits_{i_k=1}^{n_k} f(a, a^{(2)}_{i_2}, \ldots, a^{(k)}_{i_k})$. For this function, the condition (2) turns into (3), so, due to Lemma, we conclude that $g(a) = \sum\limits_{i_2=1}^{n_2} \ldots \sum\limits_{i_k=1}^{n_k} f(a, a^{(2)}_{i_2}, \ldots, a^{(k)}_{i_k}) = a$ for all a. Thus, for every a, the new function $f'(a_2,\ldots,a_k) \stackrel{\rm def}{=} f(a,a_2,\ldots,a_k)a$ of $k-1$ variables satisfies the following property:

$$\text{For every } k-1 \text{ finite sequences}$$

$$\text{of non-negative real numbers } (a^{(2)}_1,\ldots,a^{(2)}_{n_2}),\ldots,(a^{(k)}_1,\ldots,a^{(k)}_{n_k}),$$

$$\text{if } \sum_{i_2=1}^{n_2} a^{(2)}_{i_1} = 1 \text{ and } \ldots \text{ and } \sum_{i_k=1}^{n_k} a^{(k)}_{i_k} = 1,$$

$$\text{then } \sum_{i_2=1}^{n_2} \ldots \sum_{i_k=1}^{n_k} f'(a^{(2)}_{i_2},\ldots,a^{(k)}_{i_k}) = 1.$$

This is exactly the property (2) for $k-1$, so, due to induction assumption, we conclude that $f'(a_2,\ldots,a_k) = a_2\cdot\ldots\cdot a_k$. Since $f'(a_2,\ldots,a_k) = f(a,a_2,\ldots,a_k)/a$, we thus conclude that $f(a,a_2,\ldots,a_n) = a\cdot f'(a_2,\ldots,a_k) = a\cdot a_2\cdot\ldots\cdot a_k$. The induction step is proven, and so is the theorem.

Acknowledgments.

This work was supported in part by NASA under cooperative agreement NCC5-209, by the Future Aerospace Science and Technology Program (FAST) Center for Structural Integrity of Aerospace Systems, effort sponsored by the Air Force Office of Scientific Research, Air Force Materiel Command, USAF, under grant F49620-00-1-0365, by NSF grants EAR-0112968, EAR-0225670, and EIA-0321328, and by the Army Research Laboratories grant DATM-05-02-C-0046.

This work was also supported, in part, by Small Business Innovation Research grant 9R44CA81741 to Applied Biomathematics from the National Cancer Institute (NCI), a component of the National Institutes of Health (NIH), and by a research grant from Sandia National Laboratories as part of the Department of Energy Accelerated Strategic Computing Initiative (ASCI).

References

[1] J. Aczél, *Lectures on functional equations and their applications,* Academic Press, N.Y., London, 1966.

[2] S. A. Ferson, V. Kreinovich, L. Ginzburg, D. S. Myers, and K. Sentz, *Constructing Probability Boxes and Dempster-Shafer Structures*, Sandia National Laboratories, Report SAND2002-4015, January 2003.

[3] G. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Prentice Hall, Upper Saddle River, New Jersey, 1995.

[4] H. T. Nguyen and E. Walker, *First Course in Fuzzy Logic*, CRC Press, Boca Raton, Florida, 1999.

[5] H. M. Wadsworth Jr., *Handbook of statistical methods for engineers and scientists*, McGraw-Hill, N.Y., 1990.

[6] R. R. Yager, J. Kacprzyk, and M. Pedrizzi (eds.), *Advances in the Dempster-Shafer Theory of Evidence*, Wiley, N.Y., 1994