Why Square Roots of Probabilities?

Kevin H. Knuth
Departments of Physics and Informatics
University at Albany (SUNY)
Albany NY 12222, USA

March 29, 2016

Abstract

Square roots of probabilities appear in several contexts, which suggests that they are somehow more fundamental than probabilities. Square roots of probabilities appear in expressions of the Fisher-Rao Metric and the Hellinger-Bhattacharyya distance. They also come into play in Quantum Mechanics via the Born rule where probabilities are found by taking the squared modulus of the quantum amplitude. Why should this be the case and why do these square roots not arise in the various formulations of probability theory?

In this short, inconclusive exploration, I consider quantifying a logical statement with a vector defined by a set of components each quantifying one of the atomic statements defining the hypothesis space. I show that conditional probabilities (bi-valuations), such as $P(x|y)$, can be written as the dot product of the two vectors quantifying the logical statements $x$ and $y$ each normalized with respect to the vector quantifying the conditional $y$. The components of the vectors are proportional to the square root of the probability. As a result, this formulation is shown to be consistent with a concept of orthogonality applied to the set of mutually exclusive atomic statements such that the sum rule is represented as the sum of the squares of the square roots of probability.

Key Words: probability, orthogonality, information, information geometry, quantum mechanics, Fisher-Rao metric, Hellinger-Bhattacharyya distance

1 Introduction

Square roots of probabilities appear in a variety of interesting contexts ranging from information geometry and statistical manifolds in the form of the Fisher-Rao metric

$$g_{ij}(\theta) = 4 \int \frac{\partial \sqrt{p(x|\theta)}}{\partial \theta_i} \frac{\partial \sqrt{p(x|\theta)}}{\partial \theta_j} dx$$

and the Hellinger-Bhattacharyya Distance

$$H^2(P,Q) = \frac{1}{2} \int \left( \sqrt{\frac{dP}{d\lambda}} - \sqrt{\frac{dQ}{d\lambda}} \right)^2 d\lambda$$
to quantum mechanics via the Born rule where probabilities are found by taking the squared modulus of complex-valued quantum amplitudes
\[ p(x) = |\psi(x)|^2. \] (3)

While quantum amplitudes are clearly fundamental, and in some sense foundational understood in terms of symmetries[10][9], the roles that square roots (or roots) of probabilities play in various applications[6][3][4][14] have an unclear and less well-understood foundational significance. This includes the area where quantum mechanics and information geometry intersect[8].

This paper represents a short, inconclusive exploration into the fact that square roots of probabilities appear to be somehow fundamental, but do not appear to play a role in any of the major foundational approaches to probability theory[7][13][5][12].

2 Quantification with Vectors

Let me propose—out of the blue—that we quantify statements with a vector. This is not a derivation—it is instead an exploration.

Consider a hypothesis space based on three atomic (mutually exclusive) statements \( a_1, a_2, \) and \( a_3 \), and let us quantify each statement with a vector, such that the statement \( a_1 \) is quantified by the vector \( (q(a_1), 0, 0) \), where \( q \) is a function that takes its argument, which is a statement, to a real number. We then have the following:

\[ \begin{align*}
    a_1 & \text{ is quantified by } v(a_1) := (q(a_1), 0, 0) \\
    a_2 & \text{ is quantified by } v(a_2) := (0, q(a_2), 0) \\
    a_3 & \text{ is quantified by } v(a_3) := (0, 0, q(a_3)).
\end{align*} \]

The join, or disjunction, of two statements, denoted by \( \lor \), is a statement found by applying the logical OR. As such, the join operation is associative and commutative. Since we want to ultimately rank the statements, given the properties of the join operation we can expect that one should sum the vectors of atomic statements: \[ v(a_1 \lor a_2) := v(a_1) + v(a_2) = (q(a_1), q(a_2), 0). \]

Vectors representing non-atomic statements should be summed using the inclusion-exclusion formula to avoid overcounting[11].

3 Bi-Valuations and Inner Products

Let us consider bi-valuations, such as \( P(x|y) \), to represent a projection from one statement \( x \) onto another statement \( y \), which is called the context, defined by a context-normalized inner product where
\[ P(x|y) := \frac{v(x) \cdot v(y)}{|v(y)|^2}. \] (4)

\footnote{1At the very least, this is certainly less well-understood by me.}

\footnote{2Again, this is not a derivation. However, Skilling has an unpublished derivation that could be employed to make this assertion of additivity rigorous.}
Note that the fact that the inner product is normalized with respect to the
squared magnitude of the vector representing the context (second argument)
introduces an important asymmetry.

Next we show, via several examples, that this forces the vector components
to be square roots of probabilities. For example, since the atoms \(a_1, a_2,\) and \(a_3\)
are assumed to be mutually exclusive and exhaustive, the join of all three atoms
is the truism. This means that the probability of any one of the atoms can be
expressed as the conditional probability (bi-valuation) \(p_i = P(a_i | a_1 \lor a_2 \lor a_3).\)
For \(a_i = a_1,\) we then have that:

\[
P(a_1 | a_1 \lor a_2 \lor a_3) = v(a_1) \cdot \frac{v(a_1 \lor a_2 \lor a_3)}{|v(a_1 \lor a_2 \lor a_3)|^2}
= (q(a_1), 0, 0) \cdot \frac{(q(a_1), q(a_2), q(a_3))}{q(a_1)^2 + q(a_2)^2 + q(a_3)^2}
= \frac{q(a_1)^2}{q(a_1)^2 + q(a_2)^2 + q(a_3)^2}
= \frac{m(a_1)}{m(a_1) + m(a_2) + m(a_3)},
\]

where \(m(a_1) + m(a_2) + m(a_3) = 1\) and \(m(a_i) = q(a_i)^2\) is proportional to the
probability of \(a_1.\) This implies that the vector components \(q(a_i)\) are square
roots of probabilities: \(q(a_i) = \sqrt{p_i}.\)

To ensure that this concept of inner products is consistent with probability
theory, we can consider switching the arguments of the bi-valuation above. This
results in

\[
P(a_1 \lor a_2 \lor a_3 | a_1) = v(a_1 \lor a_2 \lor a_3) \cdot \frac{v(a_1)}{|v(a_1)|^2}
= (q(a_1), q(a_2), q(a_3)) \cdot \frac{(q(a_1), 0, 0)}{q(a_1)^2}
= \frac{q(a_1)^2}{q(a_1)^2}
= 1,
\]

which is as expected since \(a_1\) implies \(a_1 \lor a_2 \lor a_3.\)

Let us now consider two statements, \(a_1\) and \(a_2,\) which are mutually exclusive.
Note that their vectors are orthogonal, which results in a zero projection

\[
P(a_1 | a_2) = v(a_1) \cdot \frac{v(a_2)}{|v(a_2)|^2}
= (q(a_1), 0, 0) \cdot \frac{(0, q(a_2), 0)}{q(a_2)^2}
= \frac{0}{q(a_2)^2}
= 0.
\]
Last, let us consider a more limited context

\[ P(a_1|a_1 \vee a_2) = v(a_1) \cdot \frac{v(a_1 \vee a_2)}{|v(a_1 \vee a_2)|^2} \]  

\[ = (q(a_1), 0, 0) \cdot \frac{(q(a_1), q(a_2), 0)}{q(a_1)^2 + q(a_2)^2} \]  

\[ = \frac{q(a_1)^2}{q(a_1)^2 + q(a_2)^2} \]  

\[ = \frac{m(a_1)}{m(a_1) + m(a_2)}. \]  

which is as expected.

4 Determinants and Volume

Packing the vectors quantifying the atomic statements into a matrix

\[ g = \begin{pmatrix} q(a_1) & 0 & 0 \\ 0 & q(a_2) & 0 \\ 0 & 0 & q(a_3) \end{pmatrix} \]

and taking the determinant

\[ g = \begin{vmatrix} q(a_1) & 0 & 0 \\ 0 & q(a_2) & 0 \\ 0 & 0 & q(a_3) \end{vmatrix} \]

yields a volume given by

\[ \det(g) = q(a_1)q(a_2)q(a_3) \]

\[ = \sqrt{p_1p_2p_3}. \]  

Assigning a prior probability based on the inverse volume\(^3\) yields the non-informative prior for the multinomial. \(^2\)

5 Conclusions

Statements are quantified by vectors with a dimensionality given by the number of atomic statements. The vectors quantifying the atomic statements form a basis, such that each of the atomic vectors has one and only one non-zero component proportional to the square root of the probability of that statement. In general, mutually exclusive statements are quantified by vectors that are orthogonal to one another. The probability of one statement given another is proportional to the inner product of the two vectors. This results in a real-valued Hilbert space.

In general, entities that are orthogonal to one another are often quantified by quantities whose squares sum when the entities are joined. Azcél and Dhombres \(^1\) present the following theorem which establishes the precise relation

\(^3\)This is true regardless of whether such an assignment would be appropriate.
Theorem 1 (Azcél and Dhombres: Theorem 11.17) Let $\langle E, <, > \rangle$ be a real inner product space of dimension at least 2. A continuous function $f : E \to \mathbb{R}$ is orthogonally additive if, and only if, there exists a real constant $a$ and a continuous linear function $h : E \to \mathbb{R}$ such that

$$f(x) = a\|x\|^2 + h(x) \quad (x \in E).$$

(22)

While this exposition does not constitute a rigorous treatment, it appears that since we can represent statements as vectors with a well-defined inner product, this results in conditional probabilities, which quantify these vectors in a way (represented by the function $f$), that depends, at least in part on the square modulus of the vector. Here the sum rule of probability in the case of two mutually exclusive statements becomes a sum of squares of square roots of probabilities as one would expect for orthogonal entities:

$$P(a_1 \lor a_2 | a_1 \lor a_2 \lor a_3) = P(a_1 | a_1 \lor a_2 \lor a_3) + P(a_2 | a_1 \lor a_2 \lor a_3)$$

(23)

$$= \frac{q(a_1)^2}{Z} + \frac{q(a_2)^2}{Z}$$

(24)

$$= \frac{q(a_1)^2 + q(a_2)^2}{Z},$$

(25)

where $Z = q(a_1)^2 + q(a_2)^2 + q(a_3)^2$. So it could be that the square roots of probabilities are fundamental in the sense that they allow one to quantify the combination of orthogonal entities via the sum of squares.

Such a formulation of probability theory may promise to resolve some of these mysteries surrounding square roots of probability, tie in more closely with some of the geometric concepts discussed here and elsewhere, and perhaps shed more light on quantum mechanics and the fact that the Born rule involves taking the square of the modulus of the quantum amplitude.

6 Acknowledgements

I would like to thank Ariel Caticha and James Walsh for their careful reading of this manuscript. I would also like to thank Fredrick Daum for bringing several of the applications of square roots of probabilities and densities to my attention.

References

[1] J. Aczél and J. G Dhombres, *Functional equations in several variables*, vol. 31, Cambridge University Press, 1989.

[2] G. E. P. Box and G. C. Tiao, *Bayesian inference in statistical analysis*, John Wiley & Sons, New York, 1992.

[3] D. Brigo, B. Hanzon, and F. LeGland, *A differential geometric approach to nonlinear filtering: the projection filter*, IEEE T. Automat. Contr. 43 (1998), no. 2, 247–252.

[4] L. Chizat, B. Schmitzer, G. Peyré, and F.-X. Vialard, *An interpolating distance between optimal transport and Fischer-Rao*, 2015, arXiv:1506.06430 [math.AP].
[5] R. T. Cox, *Probability, frequency, and reasonable expectation*, Am. J. Physics 14 (1946), 1–13.

[6] F. E. Daum, *Exact finite-dimensional nonlinear filters*, IEEE T. Automat. Contr. 31 (1986), no. 7, 616–622.

[7] B. de Finetti, *Probabilism*, Erkenntnis 31 (1989), no. 2, 169–223, English translation and reprinting of de Finetti, B.: 1931c, ‘Probabilismo’, Logos (Napoli), pp. 163-219.

[8] P. Goyal, *Information-geometric reconstruction of quantum theory*, Phys. Rev. A 78 (2008), no. 5, 052120.

[9] P. Goyal and K. H. Knuth, *Quantum theory and probability theory: their relationship and origin in symmetry*, Symmetry 3 (2011), no. 2, 171–206.

[10] P. Goyal, K. H. Knuth, and J. Skilling, *Origin of complex quantum amplitudes and Feynman’s rules*, Phys. Rev. A 81 (2010), 022109, (arXiv:0907.0909 [quant-ph]).

[11] K. H. Knuth, *The deeper roles of mathematics in physical laws*, 2016, In press, FQXi 2015 Essay Contest (Third Prize),(arXiv:1504.06686 [math.HO]).

[12] K. H. Knuth and J. Skilling, *Foundations of inference*, Axioms 1 (2012), no. 1, 38–73.

[13] A. N. Kolmogorov, *Foundations of the theory of probability*, Chelsea, New York, NY, USA, 1950, English translation and reprinting of Kolmogorov, A., 1933, Grundbegriffe der Wahrscheinlichkeitsrechnung (Springer, Berlin).

[14] F. Pfaff, G. Kurz, and U. D. Hanebeck, *Multimodal circular filtering using fourier series*, 18th International Conference on Information Fusion (Fusion), 2015, IEEE, 2015, pp. 711–718.