# A User-Friendly Probability Machine with Some Applications to Philosophy of Science

Branden Fitelson

Department of Philosophy

San José State University

branden@fitelson.org

http://fitelson.org/





## Brief Overview of Talk

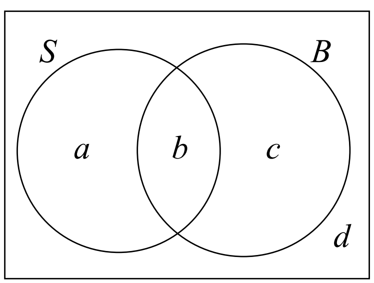
- Probabilities and Venn Diagrams
  - Using Venn Diagrams to visualize probability models
  - Translation from probability to algebra using Venn Diagrams
- The Probability Machine
  - Tarski's decision procedure for algebra now applies to probabilities
  - A user-friendly *Mathematica* function for reasoning about probabilities
- Applications to Bayesian Confirmation Theory
  - Most problems in Bayesian confirmation are solvable with our Machine
  - Some well-known historical examples — very easy for our Machine





## Probability Calculus & Venn Diagrams I

- Example: We draw a card at random from a standard 52-card deck. Let $S$ be the proposition that the card is a spade, and let $B$ = the card is black.
- We can represent probabilities of $S$ and $B$ (and all their logical combinations) using the following Venn Diagram [1] (not drawn to scale):




$a = \Pr(S \text{ and not-}B) = 0$

$b = \Pr(S \text{ and } B) = \Pr(S) = 1/4$

$c = \Pr(\text{not-S and B}) = 1/4$

$d = \Pr(\text{not-}S \text{ and not-}B) = \Pr(\text{not-}B) = 1/2$

$a + b + c + d = 1$

$$\Pr(B \text{ given } S) = \frac{\Pr(S \text{ and } B)}{\Pr(S)} = \frac{b}{a+b} = 1$$

$$\Pr(S \text{ given } B) = \frac{\Pr(S \text{ and } B)}{\Pr(B)} = \frac{b}{b+c} = \frac{1}{2}$$

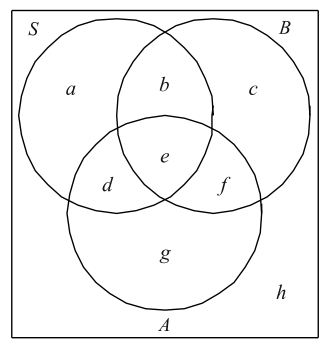
- Box =“all possible worlds”. circles = “worlds in which $S$, $B$ are true”. $a$–$d$ = “proportions of worlds” in which combinations of $S$, $B$ are true.





## Probability Calculus & Venn Diagrams II

- If we add a third proposition $A$ = “the card is the ace of clubs,” then we will have $2^3 = 8$ combinations of $S$, $B$, and $A$. We can still think of the probabilities as “areas of regions in a Venn Diagram” — with 3 circles:




$a = \Pr(\textit{not-S and not-B and A}) = 0$

$h = 1 - (a + b + c + d + e + f + g)$

$\Pr(S) = a + b + d + e$

$\Pr(B) = b + c + e + f$

$\Pr(A) = d + e + f + g$

$\Pr(A \text{ or } B) = b + c + d + e + f + g$

$$\Pr(A \text{ given } (S \text{ or } B)) = \frac{\Pr(A \text{ and } (S \text{ or } B))}{\Pr(S \text{ or } B)} = \frac{d+e+f}{a+b+c+d+e+f}$$

- Thus, we can translate any *probabilistic* claims about the *propositions* $S$, $B$, and $A$, into *algebraic* claims about the (seven) *real numbers a–g*.

## The Probability Machine I

- Generalizing this Venn Diagram technique [10], we can translate any *probabilistic* equation or inequality in any *n*-proposition probability model into an *algebraic* formula in terms of $2^n - 1$ real variables.
- As a result, a great many arguments involving the probability calculus can be expressed as sets of equations and inequalities in the theory of real closed fields (see [9], pp. 273 *ff* for a precise definition of TRCF).
- Tarski [14] described a decision procedure for the theory of real closed fields. Thus, in principle, Tarski's method gives us a way to determine which of these arguments couched in the probability calculus are valid.
- Unfortunately, Tarski's procedure is quite complex (exponential [6], and sometimes double-exponential [11], in the number of variables).
- Even for the simple 3-proposition case (with 7 variables), Tarski's method is (in general) intractable (even on today's computers).





## The Probability Machine II

- George Collins invented an improved quantifier-elimination method known as *cylindric algebraic decomposition* (CAD) [5].
- Although Collins's CAD method is double-exponential in the number of variables, it is polynomial for a *fixed* number of variables ([5], p. 86).
- This suggests that it might be possible to implement a version of CAD which can do many of the problems (*e.g.*, the 3-event ones) in Bayesian confirmation theory, given reasonable time and memory constraints.
- Hong [8] improved upon the CAD algorithm, calling his enhanced version *partial* CAD, and implemented it in a program called `qepcad`.
- This program has subsequently been further improved upon by many others, and is now publicly available on the Web [2] (linux/unix/PC).
- Some of `qepcad`'s functionality has been implemented (by Strzebonski [13]) in *Mathematica* 4.1 and 4.2, in the `Experimental` package.





## The Probability Machine III

- The function `InequalityInstance` in the `Experimental` package of *Mathematica* 4.2 is particularly useful for our purposes.
- `InequalityInstance` takes as its argument a finite set $\mathcal{S}$ of equations, inequations, inequalities (in TRCF) over a finite set of real variables $\mathcal{V}$.
- `InequalityInstance` outputs an assignment of real numbers (if one exists) to the variables in $\mathcal{V}$, which satisfies all of the members of $\mathcal{S}$. If $\mathcal{S}$ is unsatisfiable, then `InequalityInstance` outputs "{ }".
- Using the translation procedure above, I have written a user-friendly *Mathematica* function called `PrSAT`, which takes as input a set $\mathcal{S}$ of equations, inequations, or inequalities in the probability calculus.
- If $\mathcal{S}$ is satisfiable (in TRCF), then `PrSAT` returns a probability model satisfying all the members of $\mathcal{S}$. If not, `PrSAT` returns "`Unsatisfiable`".





## The Probability Machine IV

- `PrSAT` is quite useful for probabilists! I used it extensively in my dissertation (which was on probabilistic theories of evidence [7]).
- In particular, I had over 100 open questions in probability calculus that I needed to answer. I was able to solve *all* of these questions, using `PrSAT`!
- `PrSAT` is particularly useful for finding models of complicated sets of probabilistic equations, inequations, and inequalities. While `PrSAT` does not generate *proofs*, it does *verify* theoremhood (which is also useful!).
- I will now demonstrate `PrSAT` on a well-known example (see [12], p. 85). Fact: $X$ can be (probabilistically) independent of each of $Y$ and $Z$, but *dependent* on their conjunction $Y$ & $Z$; *i.e.*, there are probability models in which all three of the following conditions are satisfied for $X$, $Y$, and $Z$:
  1. $\Pr(X \textit{ and } Y) = \Pr(X) \cdot \Pr(Y)$
  2. $\Pr(X \textit{ and } Z) = \Pr(X) \cdot \Pr(Z)$
  3. $\Pr(X \textit{ and } (Y \textit{ and } Z)) \neq \Pr(X) \cdot \Pr(Y \textit{ and } Z)$

## The Probability Machine V

- PrSAT is effective for 3-event spaces (I've not seen many real problems in 3-event probability calculus that cannot be solved with PrSAT).
- The translation procedure is easily generalized to *n*-events. But, PrSAT becomes inefficient, even for 4-event spaces (with 15 primitive variables).
- It would be nice if we could come up with ways of optimizing InequalityInstance for applications to the probability calculus.
- This is one of my current research projects in "logical methods".
- My interest in such methods is purely instrumental. It grew out of a desire to be able to reconstruct, understand, and improve upon arguments involving the probability calculus that appear in the PoS literature.
- In general, I am always on the lookout for methods that can take the drudgery out of argument reconstruction and analysis in PoS. The nice thing about these methods is that they free us up to do more *philosophy*!





## Applications to Bayesian Confirmation Theory

- According to Bayesianism, $E$ confirms (or supports) $H$ if $\Pr(H \textit{ given } E) > \Pr(H)$. $E$ disconfirms $H$ if $\Pr(H \textit{ given } E) < \Pr(H)$.
- In our first example, $B$ confirms $S$, since $\Pr(S \textit{ given } B) = \frac{1}{2} > \frac{1}{4} = \Pr(S)$ (*i.e.*, the card's being black confirms its being a spade — makes sense).
- Question: is it possible for $E_1$ to confirm $H$, but at the same time for "$E_1$ *and* $E_2$" to *dis*confirm $H$? Answer: Yes [3]. Easy for PrSAT.
- Question: is it possible for $E_1$ to confirm $H$, but at the same time for "$E_1$ *or* $E_2$" to *dis*confirm $H$? Answer: Yes [3]. Easy for PrSAT.
- Question: is it possible for $E$ to confirm $H_1$, but at the same time for $E$ to *dis*confirm "$H_1$ *or* $H_2$? Answer: Yes [3]. Easy for PrSAT.
- Question: is it possible for $E$ to confirm $H_1$, but at the same time for $E$ to *dis*confirm "$H_1$ *and* $H_2$? Answer: Yes [3]. Easy for PrSAT.





# References

[1] E. Adams, *A primer of probability logic*, CSLI Lecture Notes, vol. 68, 1998.

[2] C. Brown, *Qepcad-b, a program for computing with semialgebraic sets using cads.*, to appear, preprint available at http://www.cs.usna.edu/wcbrown/research/MOTS2002.2.pdf. The qepcad program itself is available from http://www.cs.usna.edu/~qepcad/B/QEPCAD.html.

[3] R. Carnap, *Logical Foundations of Probability*, 1 ed., University of Chicago Press, 1950.

[4] B. Caviness and J. Johnson (eds.), *Quantifier elimination and cylindrical algebraic decomposition*, Texts and Monographs in Symbolic Computation, Springer-Verlag, 1998.

[5] G. Collins, *Quantifier elimination for real closed fields by cylindrical algebraic decomposition*, reprinted in [4], pp. 85–121.

[6] M. Fischer and M. Rabin, *Super-exponential complexity of Presburger arithmetic*, Complexity of computation (Proc. SIAM-AMS Sympos., New York,





1973), Amer. Math. Soc., 1974, pp. 27–41. SIAM–AMS Proc., Vol. VII.

[7] B. Fitelson, *Studies in Bayesian confirmation theory*, Ph.D. thesis, University of Wisconsin–Madison (Philosophy), 2001, http://fitelson.org/thesis.pdf.

[8] H. Hong, *An improvement of the projection operator in cylindrical algebraic decomposition*, reprinted in [4], pp. 166–173.

[9] S. Lang, *Algebra*, Addison-Wesley Publishing Co., 1965.

[10] T. More, Jr., *On the construction of Venn diagrams*, J. Symb. Log. **24** (1959), 303–304.

[11] J. Renegar, *Recent progress on the complexity of the decision problem for the reals*, reprinted in [4], pp. 220–241.

[12] S. Ross, *A first course in probability*, fourth ed., Prentice Hall, 1994.

[13] A. Strzebonski, *Solving algebraic inequalities with version 4*, The Mathematica Journal **7** (2000), 525–541.

[14] A. Tarski, *A decision method for elementary algebra and geometry*, Univ. of Calif. Press, 1951.