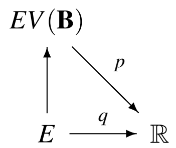
On the Second Half of Boole's Laws of Thought

Jay Sulzberger
jays@panix.com

January 4, 2009

George Boole in his Laws of Thought [3], published in 1854, defined the following problem of estimating an unknown probability given certain data. We are given $\mathbf{B}$, a finite Boolean algebra and a subset E of the events of $\mathbf{B}$. The pair $(\mathbf{B}, E)$ is called a Venn diagram. In general E will not be a Boolean subalgebra of $\mathbf{B}$. We are also given a function $q : E \to \mathbb{R}$, that is, from E to the reals. We wish now to find a probability p on $\mathbf{B}$ which extends q. Diagrammatically p is such that the following is commutative:



The up arrow is the embedding of E into the set of events of $\mathbf{B}$. In case there does exist such a probability p we say that q is consistent. The set of all consistent q we call Boole's polyhedron $BP(\mathbf{B}, E)$. In the cases we are interested in there are many different probabilities which extend q, and we wish to find a p which is, in some sense, the natural extension of q. An estimator for Boole's problem is a function *est* defined on the collection of all consistent data triples $(\mathbf{B}, E, q)$ such that $est(\mathbf{B}, E, q)$ is a probability space $(\mathbf{B}, p)$ with p an extension of q. Now if probability spaces formed a variety of algebras, in the sense of universal algebra, then we could define our estimate p to be the unique probability given by considering our data $(\mathbf{B}, E, q)$ to be generators and relations for p. Although in 1854 no mathematician had as yet attained to a clear understanding of the concept of a variety of algebras, Boole gave a successful definition of what it means to say "$(\mathbf{B}, p)$ is given by the generators and relations $(\mathbf{B}, E, q)$". In 1862 [4] he published a proof that, assuming the data $(\mathbf{B}, E, q)$ are consistent, p exists and is unique and restricts to q on E. Thus Boole was able to define his estimator.

Boole's estimator is, in extension, the maximum entropy estimator. But Boole did not have the entropy function H. Rather he used a polynomial form of the partition function.

Boole starts from the special case $(\mathbf{B}, E)$ in which E generates $\mathbf{B}$ and E is a set of "unconditioned events". Boole defines E to be a set of unconditioned events when there are no Boolean relations among the events in E. In this case, $BP(\mathbf{B}, E)$ is the unit cube, consisting of all real valued functions q defined on E, such that for all e in E, $0 \leq q(e) \leq 1$. Boole's estimate $est(\mathbf{B}, E, q)$, for any q in $BP(\mathbf{B}, E)$, is the unique probability p on $\mathbf{B}$ which makes E a set of independent events.

In case $(\mathbf{B}, E)$ is not a Venn diagram of events where E generates $\mathbf{B}$, Boole describes what further information might be furnished to get a Venn diagram in which the events do generate the Boolean algebra. Boole then deals with the remaining case, in which E does generate $\mathbf{B}$. One further simplification is made without loss of generality, and we get to the case in which E generates $\mathbf{B}$, and when E is considered as a set of functions on $PROB(\mathbf{B})$, and the constant function with value 1 is adjoined, we get a set of linearly independent functions. We now assume that $(\mathbf{B},E)$ satisfies these conditions.

Because E generates $\mathbf{B}$, we may define a quotient of Boolean algebras $\gamma : \mathbf{A} \to \mathbf{B}$. $\mathbf{A}$ is the free Boolean algebra generated by the set F of unconditioned events. F is a set of events created in one to one correspondence with E. In Boole's terms, F is the unconditioned version of E. More precisely, $(\mathbf{A},F)$ is the unconditioned Venn diagram corresponding to the conditioned Venn diagram $(\mathbf{B},E)$. γ is the surjection of Boolean algebras defined by the condition that for every f in F, $\gamma(f) = e$, where e is the conditioned event corresponding to f. γ exists and is unique by construction, because $(\mathbf{A},F)$ is a Venn diagram of unconditioned events which generate $\mathbf{A}$.

Boole now defines the conditionalization functor. The conditionalization functor is defined on the category of all surjections between finite Boolean algebras, with the one element Boolean algebra excluded from the category. The conditionalization functor takes values in a category which has one object for every finite Boolean algebra, the object being the polyhedron of all probabilities on the underlying Boolean algebra. Again the one element Boolean algebra is excluded. The image of γ under the conditionalization functor is the usual conditionalization of probabilities

$$\Gamma : PROB(\mathbf{A}) \to PROB(\mathbf{B})$$

Note that Γ is not, in general, an everywhere defined function. For all proper quotients γ, there will be probabilities r in $PROB(\mathbf{A})$ for which $\Gamma(r)$ is not defined. But if r is non-zero on all atoms of $\mathbf{A}$, then $\Gamma(r)$ is defined.

Boole uses the conditionalization functor to transport his estimator from $BP(\mathbf{A},F)$ to $BP(\mathbf{B},E)$. Note that $BP(\mathbf{A},F)$ is just the n cube, where n is the cardinality of F, also, of course, the cardinality of E. Below we write $BP(\mathbf{A},F)$ as $CUBE(F)$. Here is the diagram which gives the transport:

$$\begin{array}{ccc} PROB(\mathbf{A}) & \xrightarrow{\Gamma} & PROB(\mathbf{B}) \\ {\scriptstyle est}\uparrow & & \downarrow{\scriptstyle M} \\ CUBE(F) & \xrightarrow{brm} & BP(\mathbf{B},E) \end{array}$$

The left up going arrow from $CUBE(F)$, that is, $BP(\mathbf{A},F)$, is Boole's estimator, which is already defined in this case. The top arrow from $PROB(\mathbf{A})$ to $PROB(\mathbf{B})$ is the conditionalization Γ. The down arrow M is the restriction of probabilities over $\mathbf{B}$ to the set of events E. Finally the bottom arrow *brm*, which we call *Boole's rational map*, is defined as the composition $M \circ \Gamma \circ est$, so the diagram is commutative. We repeat: *est* is everywhere defined on $CUBE(F)$, Γ is not usually everywhere defined on $PROB(\mathbf{A})$, and M is everywhere defined on $PROB(\mathbf{B})$. So Boole's rational map is usually not defined on all of $CUBE(F)$, but in certain important special cases, it is. In equally important cases, it is not. Boole's rational map is always defined on the interior of $CUBE(F)$.

Boole defines his estimator

$$est : BP(\mathbf{B},E) \to PROB(\mathbf{B})$$

by means of the following theorem:

Theorem 1. *(Boole's Theorem)* Let q be strictly consistent. Then there exists a unique s in the interior of $CUBE(F)$ such that

$$M(\Gamma(est(s))) = q$$

Boole's Theorem may be rephrased by saying that for any strictly consistent q, the set of preimages of q in $PROB(\mathbf{B})$ under M and the image of the interior of $CUBE(F)$ under $\Gamma \circ est$ in $PROB(\mathbf{B})$ intersect in exactly one point. In statistics, this image of $CUBE(F)$ in $PROB(\mathbf{B})$, modulo boundary difficulties, is called "the exponential family for $(\mathbf{B}, E)$".

Definition 1. *(Boole's Estimator)* Let $(\mathbf{B}, E)$ be as above. Let q be strictly consistent for $(\mathbf{B}, E)$. Let s be the unique s guaranteed by Boole's Theorem. The value of *Boole's estimator est* at q is

$$est(q) = \Gamma(est(s))$$

Definition 2. *(Boole's Partition Function)* Let $(\mathbf{B}, E)$ be as above. Then there is a unique least event v in $\mathbf{A}$ for which

$$\gamma(v) = 1$$

where the 1 on the right is the top element of $\mathbf{B}$. *Boole's Partition Function* is the function

$$Z : CUBE(F) \to \mathbb{R}$$

defined for all s in $CUBE(F)$ by

$$Z(s) = [est(s)](v)$$

Let s lie in $CUBE(F)$. For every atom b of $\mathbf{B}$ there is a unique atom a in $\mathbf{A}$ such that $\gamma(a) = b$, and the union of all a's which get sent to atoms of $\mathbf{B}$ is the conditioning event v. The other atoms of $\mathbf{A}$ are sent to the zero of $\mathbf{B}$. (This is just Stone duality for the special case of finite Boolean algebras and quotient maps.) Let us compute $[est(s)](a)$:

$$[est(s)](a) = \prod_{a \leq f} s(f) \prod_{\text{not } a \leq f} (1 - s(f))$$

$Z(s)$ is the sum of the probabilities of all atoms a in $\mathbf{A}$ which lie below v:

$$\begin{aligned} Z(s) &= \sum_{a \leq v} [est(s)](a) \\ &= \sum_{a \leq v} \left[\prod_{a \leq f} s(f) \prod_{\text{not } a \leq f} (1 - s(f)) \right] \end{aligned}$$

Definition 3. *(Boole's Rational Map, version 1)* Let $(\mathbf{B}, E)$ be as above. *Boole's Rational Map* is the function

$$brm : CUBE^o(F) \to BP^o(\mathbf{B}, E)$$

which, for all s in the interior $CUBE^o(F)$ of $CUBE(F)$,

$$brm(s) = M(\Gamma(est(s)))$$

At the event e in E, we have

$$\begin{aligned} [brm(s)](e) &= [M(\Gamma(est(s)))](e) \\ &= \frac{1}{Z(s)} \sum_{\substack{a \text{ an atom of } \mathbf{A} \\ 0 < \gamma(a) \leq e}} \left[\prod_{a \leq f} s(f) \prod_{\text{not } a \leq f} (1 - s(f)) \right] \end{aligned}$$

Boole's work in probability was not well understood in his lifetime, but it was sometimes taken seriously. After Boole's death this work was mostly ignored. When not ignored it was almost always misunderstood. In 1986 Theodore Hailperin published the second edition of "Boole's Logic and Probability" [7]. Hailperin's book is an extraordinary work of mathematical archaeology. Boole's "Laws of Thought" and his papers on probability are written in a language that by 1900 was no longer spoken by any mathematician on Earth. But Hailperin's presentation of Boole's work is both complete and absolutely clear. We present Boole's work, with the help of Hailperin's translation, and we explain some of the connections between Boole's estimator, Boltzmann's entropy, Fisher's information matrix, Kolmogorov's consistency conditions, the Einstein-Podolsky-Rosen-Bohm circuit, John Bell's first theorem, quantum field theory, the degree of the moment map, and acyclic database schemes.

References

[1] M. Atiyah. Convexity and Commuting Hamiltonians. *Bulletin of the London Mathematical Society*, 14:1–15, 1982. Boole's main theorem is proved as 'convexity of the image of the moment map' when one notes that the image is actually exactly presented.

[2] J. S. Bell. On the Einstein Podolsky Rosen Paradox. In *Speakable and Unspeakable in Quantum Mechanics*. Cambridge University Press, 1987. There are two theorems in the paper. The first theorem is that the set of possible behaviors of all classical Einstein-Podolsky-Rosen-Bohm circuits is naturally equivalent to the set of consistent data for a certain Venn diagram. The second theorem is that an inequality satisfied by all classical behaviors is violated by the behavior of one particular quantum Einstein-Podolsky-Rosen-Bohm circuit. The first theorem is not as celebrated as the second.

[3] G. Boole. *An Investigation of the Laws of Thought, on which are founded the Mathematical Theories of Logic and Probability*. Walton and Maberly, London, England, 1854.

[4] G. Boole. On the Theory of Probabilities. In R. Rhees, editor, *Studies in Logic and Probability*. Watt and Co., 1952.

[5] V. Guillemin and S. Sternberg. Convexity Properties of the Moment Mapping. *Inventiones Mathematicae*, 67:491–513, 1982. Boole's main theorem is proved here too.

[6] R. Haag. *Local Quantum Physics*. Springer-Verlag, Berlin, Germany, 1992. This is a presentation of quantum field theory based on the idea that a quantum field is a circuit of quantum gates, and implicitly relies on the fact that we derive the behavior of the whole circuit from the behavior of its gates by applying Boole's estimator. The presentation is suitable for mathematicians. That is, it is slow, clear, and precise.

[7] T. Hailperin. *Boole's Logic and Probability*. North-Holland, Amsterdam, Holland, second edition, 1986.