How Big is your Mathematics?

Frank Lad

Department of Mathematics and Statistics
University of Canterbury, Christchurch, New Zealand
email: F.Lad@math.canterbury.ac.nz

Dedicated to the memory of Derrick Breach, 1933-1996.

Abstract

Requiring the reader's participation, this article provides a cautionary tale on the use of combinatoric counting procedures for solving problems in probability. While their value is incontrovertible, it is crucial to pay attention to what is being counted. In standard cases, too much is typically presumed, and this limits unduly the size of the coherent solution. The operational subjective theory of probability requires that the entire range of solutions to a problem cohering with the assertions it specifies be admitted as "the solution." This article is presented as a tribute to the late Derrick Breach, who had proposed the problem it assesses for discussion in our Department Newsletter.

The Problem. A bag contains 16 billiard balls, some white and the remainder red. Two balls are drawn at the same time. It is equally likely that the two balls will be of the same colour as it is that they are of different colours. How many of each colour are there?

Before continuing with this article, the reader is requested to solve this problem, precisely as it is posed, for yourself. The content of this article concerns three solutions to the problem: a meritorious high school level solution, with a concluding deliberation posed by a thoughtful student at this level; a second-year university level solution by an honours student of physics; and a second-year honours solution by a student of probability as it is conceived within the operational subjective tradition developed by Bruno de Finetti (1937) and his followers.

My motivation in this presentation of university level concepts is to encourage serious thinking into the mathematical formulation of probability as it has been produced by constructive subjectivism. In many applied problems of science and commerce, practitioners are unable to assess completely the vast array of probabilities required to specify precisely a joint distribution over all the unknown quantities it concerns. Rather than completing the distribution uniquely by an arbitrary extension of "assumptions," the subjectivist approach allows a computation of the bounds on any probability assertions that would cohere with the limited array that has been specified. Thus, a second motivation for this article is my desire to extend the audience of mathematicians who are familiar with the "fundamental theorem of prevision," which formalises this computational procedure and its logic. I shall close this introduction with a relevant quotation from de Finetti (1974, Section 6.3.3):

> Whether one solution is more useful than another depends on further analysis, which should be done case by case, motivated by issues of substance, and not – as I confess to having the impression – by a preconceived preference for that which yields a unique and elegant answer even when the exact answer should instead be 'any value lying between specifiable limits.'

Have you completed your solution? Let's address the Problem as it has been posed.

The High School Level Solution. When two balls are selected from a bag containing 16 balls, the observable quantity W_2, defined as the number of white balls among two balls withdrawn, is distributed Hypergeometric, specifically with the parameters $N = 2$ draws, W white balls in the bag, and $R = (16 - W)$ red balls in the bag. We can write that $W_2 \sim H(N = 2, W, R = 16 - W)$. Using the notation ${}^{X}C_{Y}$ to denote the combinatorial expression $X!/[Y!(X - Y)!]$, it is clear that the probability that only one of the selected balls is white equals $P(W_2 = 1) = {}^{W}C_1\ {}^{(16-W)}C_1\ /\ {}^{16}C_2$, whereas the probability that both selected balls are the same colour equals $P(W_2 = 0\ or\ W_2 = 2) = [\ {}^{W}C_2\ +\ {}^{(16-W)}C_2]\ /\ {}^{16}C_2$. Equating these two probabilities as the Problem prescribes yields a quadratic equation in W of the form $W(16-W) = (1/2)\ [W(W-1) + (16-W)(15-W)]$, which has the solutions $W = 6$ and $W = 10$. A precocious high school student might conclude this answer with the comment, "Thus, $P(W = 6) = P(W = 10) = 1/2$ is the correct solution." Upon further consideration, the student may reflect that "In fact, any probability distribution of the form $P(W = 6) = q$ and $P(W = 10) = 1 - q$ satisfies the conditions of the problem. Thus, the solution to the problem is the interval of points on the closed line segment $(P(W = 6), P(W = 10)) \in \{\ (q, 1 - q)\ |\ q \in [0, 1]\ \}$, a rather BIG solution!"

The BIG Mistake. While any teacher would be thrilled to read such a solution from a high school student, further reflection would allow that it actually partakes in the BIG mistake that has plagued probabilistic and statistical thinking throughout this century. Did you make the mistake yourself? The source of the mistake is that the specification of the problem does not identify the distribution of W_2 as Hypergeometric, but rather the conditional distribution of $W_2|W = w$ as Hypergeometric. After all, W is unknown. Its unknown value is, in fact, the question that the Problem poses for us! Recognising that the assertive statement $W_2|W = w \sim H(N = 2, w, R = 16 - w)$ is a specification of 17 different conditional distributions for W_2 given $W = 0, 1, ..., 16$, we need to rely on the Theorem of Total Probability to represent assertions regarding the unconditional probabilities

$P(W_2 = 1)$ and $P(W_2 = 0 \text{ or } W_2 = 2)$. That theorem, remember, tells us that if $H_1, H_2, ..., H_N$ are events constituting a partition, then for any event D,

$$P(D) = \sum_{i=1}^{N} P(D|H_i)P(H_i).$$

Let us follow the thinking of a university student to see where these considerations lead us.

The University Level Honours Physics Solution. Since any number of white balls from 0 through 16 is a possible state of the bag, the solution will not be merely a number, but a probability distribution over these possibilities. Suppose we write $f(w) \equiv P(W = w)$ for $w = 0, 1, ..., 16$ to represent an unconditional probability mass function for W; the equality specified in the Problem can then be represented as

$$\sum_{w=0}^{16} [{}^{w}C_2 \; + \; {}^{(16-w)}C_2] \; f(w) \; = \; \sum_{w=0}^{16} [{}^{w}C_1 \; {}^{(16-w)}C_1] \; f(w),$$

which is derived by applying the Theorem of Total Probability to the specified conditional distributions for $W_2|W = w$ and the unspecified unconditional distribution for W identified by $f(\cdot)$. The denominator ${}^{16}C_2$ has been eliminated from both sides of this expression. Algebraically, this summation equality can be reduced to the condition that

$$\sum_{w=0}^{16} w^2 \; f(w) \; - \; 16 \sum_{w=0}^{16} w \; f(w) \; + \; 60 \; = \; 0,$$

which is a linear condition on the first two moments of the unspecified distribution for W: $\mathrm{M}_2(W) - 16 \; \mathrm{M}_1(W) + 60 \; = \; 0$. Thus, the physics student concludes this solution with the statement: "Any probability distribution representable by a mass function vector $[P(W = 0), P(W = 1), ..., P(W = 16)]$ lying within the 16-dimensional unit-simplex, $\boldsymbol{S}^{16}$, for which this moment condition holds is a solution to the Problem."

Our creditable physics student's solution is clearly much BIGGER than the high school solution. This solution is a 15-dimensional convex subset of the 16-dimensional unit-simplex, a BIG solution indeed. This student would surely make the teacher proud and even happy to enjoy a last minute intriguing waffle scribbled into the margin of the solution paper, saying "Maybe $P(W = 0)$ and $P(W = 16)$ must equal 0. I don't know."

The University Level Honours Mathematical Probability Solution. Reading the Problem as posed, carefully, the student realises that no statement has been made regarding how the two balls defining the quantity W_2 have been drawn from the bag, other than that they "are drawn at the same time." The only probabilistic assertion formally specified in the Problem statement is that $P(W_2 = 1) = P(W_2 = 0 \text{ or } W_2 = 2)$. There are many procedures of "drawing two balls at the same time" that would allow this as a reasonable assertion. The physics student's assertion of the conditional hypergeometric distributions would be merited automatically as input to the problem only if it were specified further that the balls in the bag are scrambled in such a way that equal probabilities are also asserted for each of the ${}^{16}C_2$ events defined by the draw of two specific balls from the bag.

The following scenario can provide for you one simple alternative which would also merit the assertion of the only probability that has been formally admitted in the Problem statement. Into a bag of 14 balls, some white and the remainder red, the person drawing the two balls at the same time might insert a hand containing a white and a red ball, and then exchange (unseen, inside the bag) one of these with one of the balls already in the bag, and draw out the two balls so determined. Nothing in the Problem statement precludes this as the procedure by which the two balls were drawn at the same time. You might imagine other possibilities now as well. The "solution" to the problem for which we shall search is the largest possible array of solutions that would cohere with the assertions formally specified in the Problem as it has been posed. Any further restriction on the solution must be regarded as an arbitrary imposition onto the problem that is not required by its formal statement.

The following solution relies on the constructive logic of subjective probability in the tradition begun by Bruno de Finetti, which is described in the sophisticated undergraduate and graduate level text of Lad (1996), undoubtedly new to many a reader. I invite your examination of that text and solicit your thoughts about it. The presentation of the solution here is couched in the language of "the fundamental theorem of prevision" which can be studied specifically in the text, but whose operation hopefully will be understandable in its direct application to our Problem in the remainder of this article.

As a prelude to your understanding the direction we follow, let me describe somewhat informally, de Finetti's theory of coherent prevision and its fundamental theorem, in the context of quantities whose measurement value possibilities are discrete and finite. To begin, a prevision assertion is a numerically specified expectation for a quantity. Since probabilities are merely expectations of events (quantities that must equal either 0 or 1), the coherent (non-contradictory) operation of the prevision symbol, P, unifies the theory of expectation and probability. The fundamental theorem states that if you assert your previsions (expectations) for any N quantities whatsoever, then a cohering prevision for any further quantity must lie within a specified interval whose endpoints are calculable by a linear programming routine. For a coherent prevision vector must lie within the convex hull of the vectors of possible values for the unknown quantities in question. Asserted previsions for some of the quantities place linear reductions on the subset of cohering assertions within this hull. Without further discussion, let us turn to an application of this theorem to our Problem.

Three observable but unknown quantities are involved in the statement of the Problem: $W, R = 16 - W$, and W_2. The realm matrix of possibilities for this vector of quantities along with a fourth quantity which I shall denote for the moment by the symbol D, and which I shall define shortly, is a 4×45 matrix, $\boldsymbol{R}(W, R, W_2, D)^T =$

$$\begin{pmatrix} 0 & 1 & 1 & 2 & 2 & 2 & 3 & 3 & 3 & \cdots & 14 & 14 & 14 & 15 & 15 & 16 \\ 16 & 15 & 15 & 14 & 14 & 14 & 13 & 13 & 13 & \cdots & 2 & 2 & 2 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 2 & 0 & 1 & 2 & \cdots & 0 & 1 & 2 & 1 & 2 & 2 \\ -1 & -1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 & \cdots & -1 & 1 & -1 & 1 & -1 & -1 \end{pmatrix} .$$

The columns of this matrix exhaustively list the possible values of the unknown quantity vector, $(W, R, W_2, D)^T$. Ignoring the role of the quantity D for the moment, it is important to recognise that the actual quantity vector $(W, R, W_2, D)^T$, unknown though it may be, is representable algebraically in the form

$$\begin{pmatrix} W \\ R \\ W_2 \\ D \end{pmatrix} = \mathbf{R} \begin{pmatrix} W \\ R \\ W_2 \\ D \end{pmatrix} \mathbf{Q} \begin{pmatrix} W \\ R \\ W_2 \\ D \end{pmatrix} ,$$

where $\boldsymbol{Q}(\cdot)$ is a 45×1 vector of events constituting the partition generated by the vector of unknown quantities, $(W, R, W_2)^T$. Its i^{th} component is the event of the form $Q_i \equiv ((W, R, W_2, D)^T = \text{column } i \text{ of the realm matrix } \boldsymbol{R})$.

Whereas the matrix $\boldsymbol{R}$ in this representation is a known matrix of numbers, defined by the operational definitions of the measurements described in the Problem, the vector $\boldsymbol{Q}$ is an unknown vector of events. Since it constitutes a partition, however, the sum of its components must equal 1. We know that one of its components must equal 1, and the remainder must equal 0, but we do not know which component is the 1, since we avowedly do not know the values of W, R, and W_2. Thus, any coherent uncertainty regarding these three quantities must specify a prevision (expectation) vector that lies within a linear transformation of the 44-dimensional unit-simplex, for coherent prevision assertions must be linear: $P(W, R, W_2)^T = \boldsymbol{R}\ P(\boldsymbol{Q})$. Since your prevision for the sum of the components of $\boldsymbol{Q}$ must equal the sum of your previsions (probabilities in this case) for its components, the vector $P(\boldsymbol{Q})$ must lie within $\boldsymbol{S}^{44}$. Without specifying a particular vector of such probabilities, we would say that your uncertainty about the possibilities for (W, R, W_2) is represented by this entire simplex.

Now although a complete probability distribution has not been asserted in our Problem, one specific probability assertion has been admitted: $P(W_2 = 1) = P(W_2 = 0 \textit{ or } W_2 = 2)$. In the arithmetic notation used here, the linearity of coherent prevision again requires that this condition is equivalent to the assertion $P[(W_2 = 1) - (W_2 = 0) - (W_2 = 2)] = 0$. You should notice that whereas each of the events denoted by parenthetical expressions within this bracketed argument, are just that, events (numbers equal to either 0 or 1) the entire bracketed expression is not an event. For its possible numerical values are -1 and 1. Nonetheless, in the mathematics of coherent prevision assertion such quantities are not treated in any way differently than are events. Prevision for a linear combination of quantities must as always equal the same linear combination of previsions for the quantities in question.

At any rate, the numerical value of the bracketed expression, $[(W_2 = 1) - (W_2 = 0) - (W_2 = 2)]$, is defined by a function of the numerical value of W_2. Thus, the value of this quantity has been denoted by D in our notation to this point. Moreover, it has been appended to the vector of quantities considered, yielding $(W, R, W_2, D)^T$; and its possible values have been appropriately represented as another row in the realm matrix $\boldsymbol{R}$ we have described. We shall denote this row vector by $\boldsymbol{r}(D)$ in the solution statement below.

Now what are the cohering prevision (probability) assertions $P(\boldsymbol{W}_{17}) \equiv P[(W = 0), (W = 1), (W = 2), ..., (W = 16)]$ that could possibly support this one single assertion that has been admitted in the Problem? The answer is straightforward, and relies only on the linearity of coherent prevision which is fundamental to the operational subjective characterisation of coherent probability. "The coherent solution to this problem is

$$P(\boldsymbol{W}_{17}) \in \{\boldsymbol{R}(\boldsymbol{W}_{17})\boldsymbol{q}_{45} \mid \boldsymbol{q}_{45} \in \boldsymbol{S}^{44},\ and\ \ \boldsymbol{r}(D)\ \boldsymbol{q}_{45} = 0\},$$

where $\boldsymbol{R}(\boldsymbol{W}_{17})$ is the matrix whose each row is determined by applying the event functions $(W = i)$ for $i = 0, 1, ..., 16$ to the columns of the displayed realm matrix $\boldsymbol{R}$, and $\boldsymbol{S}^{44}$ denotes the 44-dimensional unit-simplex." (Furthermore, the coherent bounds on the probability of any specific event that is defined by a linear combination of the partition of events, $\boldsymbol{Q}_{45}$, say $E \equiv \boldsymbol{c}^T\boldsymbol{Q}_{45}$, are specified by the linear programming solutions that minimise and maximise $\boldsymbol{c}^T\boldsymbol{q}_{45}$ subject to the linear restrictions that $\boldsymbol{q}_{45} \in \boldsymbol{S}^{44}$, and $\boldsymbol{r}(D)\ \boldsymbol{q}_{45} = 0$.)

This solution is actually a 16-dimensional convex and proper subset of $\boldsymbol{S}^{16}$ which properly contains the 15-dimensional subspace that was the physics student's solution. Thus, this solution is the BIGGEST of the three we have considered. The only mass function vectors in $\boldsymbol{S}^{16}$ that would not be allowed would be those that give too much probability to the events $(W = 0)$ and $(W = 16)$.

Thinking about this solution, this student too scribbles a last minute waffle into the margin. "Perhaps the statement in the Problem posed that 'some of them are white' is meant to preclude the possiblity that $W = 0$, I don't know, since zero white balls might not be considered to constitute 'some'. But $W = 16$ must surely be allowed since the number 0 is valid as a 'remainder'."

This quibble covers fine points of language usage, which could only be resolved by clarification. Notice that it would be another matter entirely to assert additionally in the Problem that $P(W_2 = 1|W = w) = P[(W_2 = 0) + (W_2 = 2)|W = w]$ for conditioning values of $w = 1, 2, ..., 15$. (This assertion would be incoherent for $w = 0$ or $w = 16$, since $P(W_2 = 1|W = 0) = P(W_2 = 1|W = 16) = 0$, whereas $P(W_2 = 0|W = 0) = P(W_2 = 2|W = 16) = 1$.) Without showing all the algebraic detail here, these assertions would put 15 more independent linear restrictions on $\boldsymbol{q}_{45}$. Their implications would be that $q_2 = q_3, q_{43} = q_{44}, q_4 + q_6 = q_5, q_7 + q_9 = q_8, q_{10} + q_{12} = q_{11}$, ... , and $q_{40} + q_{42} = q_{41}$. Interestingly, the resulting polytope of cohering vectors for $P(\boldsymbol{W}_{17})$ would be unaffected by this restriction.

However, in the context of the physicist's solution (which adds 28 more linear restrictions to the problem based on unmotivated conditionally hypergeometric specifications) these 15 assertions would be sufficient to reduce all the components of $\boldsymbol{q}_{45}$ to 0 except for $q_{15} + q_{17} = q_{16}$ and $q_{27} + q_{29} = q_{28}$. These are the components corresponding to the events $(W = 6)$ and $(W = 10)$, whose probabilities then need only sum to 1. Thus, these are the assertions, unrequired by the problem as stated, that would reduce the physics student's solution to the high school solution.

Concluding Comments. The setup of the problem we have discussed is admittedly contrived, but the issues arising in its resolution are central to contemporary developments in computational mathematics applied to problems of uncertainty. In many instances when someone is faced with making a decision, the myriad details of the unknowns are so complex that a complete distribution over the entire space of possibilities is virtually impossible to assess. Two different strategies can lead one out of this impasse. The one, followed routinely in the past, involves doing your best to conjur up a joint distribution that approximates your personal joint probabilities as well as you can imagine them. If the approximating distribution is tractable for computing the required expected utilities in the problem, an appropriate decision can be made. An alternative strategy with increasingly viable applicability has emerged from recent developments in speedy and large computers. Its logic is grounded in the operational subjective theory of probability, as encoded in the fundamental theorem of prevision, and has been exemplified in the BIGGEST solution to the Problem posed in the present exposition.

This second approach allows a practitioner to specify your assertions in whatever form is most accessible to your intuitions, as minimally or extensively as you are able. You may then compute the tightest bounds on a further assertion regarding any other quantity that is logically consistent with (coheres with) your assertions. Rather than force the reporting of a "unique solution" that is achieved by introducing unmotivated arbitrary assumptions into a problem, the spirit of this minimalist mathematics is to express the entire range of coherent options that are available to you when you express your uncertainties in the limited realistic fashion that you typically do.

The freedom allowed by this solution is what underlies the abstract and sometimes abstruse arguments in the mathematical foundations of probability concerning the requirement merely of finite additivity, as opposed to complete additivity, as the requisite property of coherent probability functions in their axiomatic construction.

References

de Finetti, B. (1937) La prévision, ses lois logiques, ses sources subjectives, *Annales de L'Institute Henri Poincaré*, **7**, 1-68. H. Kyburg (tr.) Foresight, its logical laws, its subjective sources, in *Studies in Subjective Probability*, H. Kyburg and H. Smokler (eds.) 1964, New York: John Wiley; second edition 1980, New York: Krieger.

de Finetti, B. (1974,75) *Probability Theory*, two volumes, New York: John Wiley.

Lad, F. (1996) *Operational Subjective Statistical Methods: a mathematical, philosophical, and historical introduction*, New York: John Wiley.

(Received 20/10/97)