Indifference, Sample Space, and the Wine/Water Paradox

Abstract

Von Mises' wine/water paradox has served as a foundation for detractors of the Principle of Indifference and logical probability. Mikkelson recently proposed a first solution, and here several additional solutions to the paradox are explained. Learning from the wine/water paradox, I will argue that it is meaningless to consider a particular probability apart from the sample space containing the probabilistic event in question.

Introduction

Although most current probability theorists find logical formulations of probability out of vogue, Mikkelson's [2004] defense of the Principle of Indifference (PI) begs us to critically examine the content of the classical theory. Indifference with respect to probability is the notion that given a sample space of outcomes and insufficient reason to consider one outcome more likely than another, we should assign equal probability to each of the outcomes. PI underlies the enticing urge to assign equal probability to the faces of a six-sided die, even though very few of us have ever empirically addressed the issue. The most quoted lines of attack against PI come in the form of apparent paradoxes to the principle. Because of these paradoxes, few probability theorists believe PI to be a logical principle, and at most they may agree with Gillies ([2000a]) that it has had 'heuristic successes'. One particular example, von Mises' wine/water paradox, is a favorite weapon against indifference because until Mikkelson's recent proposal no clear solution to the paradox had been forwarded. In his paper Mikkelson [2004] gives an introduction to the problems surrounding indifference that need not be repeated here, as well he outlines statements by prominent voices against PI with regard to this resistant paradox. Gillies [2000a] provides thoughtful historical background to the problem of PI, crediting Keynes for presenting the various paradoxes to PI while defending it nonetheless ([Keynes 1921], pp 41-64). Bertrand [1889] is famous for establishing an entire family of geometrical PI paradoxes that persist today. Somewhat more recent support of PI can be found in Bartha and Johns [2001] who appeal to symmetry as underlying PI, and in Jaynes [1973] who defends application of PI to physical probabilities.

In this paper I first present the wine/water paradox, and follow with Mikkelson's proposal of the solution. Although Mikkelson and I are on the same side, I do not agree with his solution of the original paradox as it assumes information not given in the original paradox. I then propose a solution to the wine/water paradox by applying indifference to the joint sample space of a ratio and its inverse, and show there is no paradox when indifference is applied within the joint sample space.

Learning from the wine/water paradox, I suggest that a closer look into the meaning of a sample space may be useful. It seems as though the sample space is a neglected object in almost all discussions on the philosophy of probability. When this set of outcomes is mentioned at all, it is only in order establish its existence. For instance, in asking about the probability that a coin lands heads, we assume that the possible outcomes are heads, tails, and perhaps edge. But by what principle is this sample space entailed, and does it matter which space we choose? I suggest that it is meaningless to consider a particular probability apart from the sample space containing the probabilistic event in question. I will argue, through examples and first axioms, that all probabilities are conditional on the sample space of outcomes, and that all so-called paradoxes to PI are simply disagreements and ambiguity in relation to sample space identification, saying nothing against PI as a logical principle. In other words, by what reasoning does ambiguity about a sample space entail that indifference applied within a sample space is a faulty principle?

The Problem

I first support PI by offering two new solutions to von Mises' wine/water paradox. I say new because Mikkelson has only recently produced a neat first solution. My first solution is similar in essence to Mikkelson's solution in that I transform the original problem into a problem that the PI can handle more easily. In fact, I only thought to derive this solution after noting Mikkelson's proposal. The second solution is quite different from the other two but requires no transformation of the original problem. It is this solution that I will defend as the entailed solution to the wine/water paradox as it is classically stated. I agree that Mikkelson has produced a solution, although it is a solution that applies to a different question of probability. Let us go forward the paradox:

> Suppose we have a mixture of liquid. All that we know about the liquid is that it is composed entirely of wine and water, and that at most there is 3 times as much of one as of the other. What is the probability that the ratio of wine to water is less than or equal to 2?

Briefly, let x=wine/water (the ratio of wine to water). It follows directly from the problem statement that the ratio x can take on values over the interval $1/3 \le x \le 3$, and that an application of PI yields $P(x \le 2) = 5/8$. [1] However, one might also phrase the problem in terms of y=1/x=water/wine (the ratio of water to wine). Similarly, $1/3 \le y \le 3$, and we recognize that the event $x \le 2$ corresponds to the event $y \ge 1/2$. Applying PI to y yields $P(y \ge 1/2) = 15/16$. Indifference has given us two different answers for the same outcome of interest. The paradox is born.

Mikkelson derives a solution to the paradox by rephrasing the original question. He asks us to imagine pouring the mixture into a graduated cylinder, and to assume that the two substances do not mix with each other so that they form a definite boundary. Depending upon the actual quantities of wine and water, the border will form between $1/4$ and $3/4$ of the total volume of liquid, corresponding to a volume ratio between $1/3$ and 3 as stated in the original problem. Call the position of this border b, and note that $1/4 \le b \le 3/4$. Given this description, Mikkelson applies PI to b and suggests that

$$P(x \le 2) = P(y \ge 1/2) = P(b \le 2/3) = 5/6$$

I encourage you to see his diagram for a better description of the solution. This is a good bit, because now we have a solution to the paradox that has the appeal of symmetry and simplicity. I am; however, somewhat frightened that this solution does not necessarily correspond to the original problem. What does all of this talk about volumes of liquid that form definite borders have to do with my original ratios? What if the water and wine mixed uniformly and could not separate even in theory; would the solution still make sense? Consider this example:

> Suppose we have a wire. All we know about the wire is that a voltage and a current can be measured across it, and that at most there is 3 times as much of one as of the other. What is the probability that the resistance across the wire is less than or equal to 2?

Unfortunately, it makes no sense to try to apply Mikkelson's solution to this example. Resistance is the ratio of voltage to current. Since voltage and current have completely different units we cannot measure an analogous 'volume' for both of them and draw a border between. Nor does mixing in a graduated cylinder or otherwise make sense. Intuition suggests that both the wine/water problem and the resistance problem should have the same answer, but I see no way to reasonably apply Mikkelson's technique to this new problem.

Before I give a solution that hopefully chases away this new paradox as well, I will present a similar solution to Mikkelson's and justify how one might apply his solution to the resistance problem. Let x correspond to the ratio of wine to water as above. Consider the transform $x^* = \log(x)$. The transformed ratio can take on values on the interval $-\log(3) \le x^* \le \log(3)$, and similarly $y^* = \log(y)$, giving $-\log(3) \le y^* \le \log(3)$. After applying PI, the probability of the event $x \le 2$ corresponds to

$$P(x^* \le \log(2)) = \bigl(\log(2) + \log(3)\bigr)/\,2\log(3) = \log(6)/\log(9) \approx 0.8155$$

It follows quite readily that after applying PI to y^*, the event $y \ge \frac{1}{2}$ corresponds to $P(y^* \ge -\log(2)) = \log(6)/\log(9)$. I have discovered another solution to the wine/water paradox, although I grow suspicious that solutions are now too easy to come by. This solution is symmetric and does not create a paradox under PI no matter how the question is asked, although it gives a different answer than Mikkelson's solution. On the surface it would seem that we could apply this general procedure to the resistance paradox or any other abstract ratio paradox and generate an answer. Nothing would stop us from applying the log transform. What then stops us from applying Mikkelson's solution to the problem? The answer is also nothing, but we must strip away his physical analogy in order to make sense of it all. Mikkelson's solution is also a mathematical transform of the original ratios presented in the problem. Again, let x be the ratio wine/water and b the 'border' in Mikkelson's graduated cylinder. The mathematical transform relating the two is

$$b = x/(1+x)$$

For instance, the ratio $x = \frac{1}{3}$ transforms to $b = \frac{1}{4}$, the lower bound in the transformed problem. This transform assumes that wine is on the bottom and water is on the top. As Mikkelson argues, his method is invariant to the way in which the question is asked, and therefore we must be able to transform the inverse ratio, y=1/x in the same manner. Replacing x with 1/y gives us the 'reverse' border r

$$r = (1/y)/(1+1/y) = 1/(1+y)$$

In other words, if b is the fraction of wine then r is the fraction of water and vice-versa. While this solution does result in an answer that is symmetric, the procedure for generating this answer is clearly asymmetric, at least in some sense. What could possibly justify a priori a transform that treats a ratio and its inverse differently? With regard to the wine/water problem it is to say, in order to apply the PI you must transform one ratio according to *b* and transform the inverse ratio according to *r*. Further, why should this transform be preferred over any other? In the resistance problem it does not appear that we have any reason to suppose that this transform is correct. At least in using the log transform I treat a ratio and its inverse exactly the same. Since I have no reason to treat them differently, the log solution seems to preserve my intuition about the problem a bit better. There are in fact *infinitely* many transforms that produce solutions with the desired symmetry, although each solution is in general different (Appendix A). The question remains as to how one might choose one transform over another. Although I believe the notion of symmetry is essential to probability, I do not support the use of symmetry arguments to resolve the various paradoxes to PI.[2] Even more damaging, what could possibly justify any transform at all, especially considering that the transformed problem has its own epistemic status as a distinct probabilistic question? I will avoid this discussion by illustrating a solution that obviates it.

Higher-Order Indifference

Applying PI to a ratio and then its inverse produces inconsistent results. PI can be consistently applied to the wine/water example if we transform the original ratios with a mathematical transform; however, there are infinitely many transforms that result in symmetric answers and no obvious way to choose between them. Worse, there is no a priori reason to transform the problem in the first place. These attempts do not work, although another procedure holds promise – remain indifferent to both *x* and *y* simultaneously. On the outset it seems as though this has been the standard procedure for decades. We have always attempted to apply PI to both *x* and *y* and it has not worked. Actually, the way the paradox is generated is by applying PI to *x* and *y* in sequence. Applying PI to *x* and *y jointly* is an entirely different matter.

Let us make clear what is meant by jointly. Probability theory applies to multiple variables just as well to a single variable. Consider tossing two fair dice, A and B. If we are interested in the probability that A=6 and B=6, then we must consider the joint probability $P(\mathrm{A}=6,\mathrm{B}=6)$. If A and B are independent, then I do not need to know about this joint description because everything can be derived from the marginal probabilities noting $P(\mathrm{A}=6,\mathrm{B}=6)$ is equal to $P(\mathrm{A}=6)P(\mathrm{B}=6)$. When A and B are correlated, this relationship no longer holds, and I need to know how A and B are jointly related. The wine-water paradox requires a joint description because I am indifferent to two simultaneous quantities, a ratio and its inverse. Further, these two quantities are necessarily dependent; therefore I can only describe the probability between them using a joint probability function. I cannot assume that x is uniformly distributed and that y is uniformly distributed because any probabilistic assumption I make about one necessarily effects the distribution of the other through their dependence relationship. Applying PI in sequence ignores this dependence and is a direct violation of probability theory.

I can apply PI to x and y jointly by being indifferent to the joint probability density function $p(x, y)$. Instead of assuming that the marginal densities of x and y are uniform, I suppose that $p(x, y)$ has a uniform density. That is, I have no reason to prefer any pair (x, y) over any other pair. Since a ratio and its inverse always come as a pair by logical necessity, it seems appropriate to be indifferent to this joint object. Is this the correct way to apply my indifference to the problem? As we saw above, there are infinitely many ways using transforms to express indifference and still derive a consistent solution. The joint method of expressing indifference is unique to the wine/water problem in that it does not require any transform of the quantities presented in the example. In other words, it answers the problem without assuming any information that is not given. While we do use the notion of an inverse ratio, this object is logically entailed by the example as the bounds are given such that they apply to both a ratio and its inverse and not to either alone.

It is unfortunate that the technical machinery needed to rigorously deal with this example goes somewhat beyond the mathematics typically presented in philosophical discussions. I derive a resolution to the paradox using the joint space in Appendix B. The joint density $p(x, y)$ has a simple form, but applying PI correctly to this space requires some calculus of two-dimensional functions. With these precautions in mind, the reader is encouraged to look at the derivation.

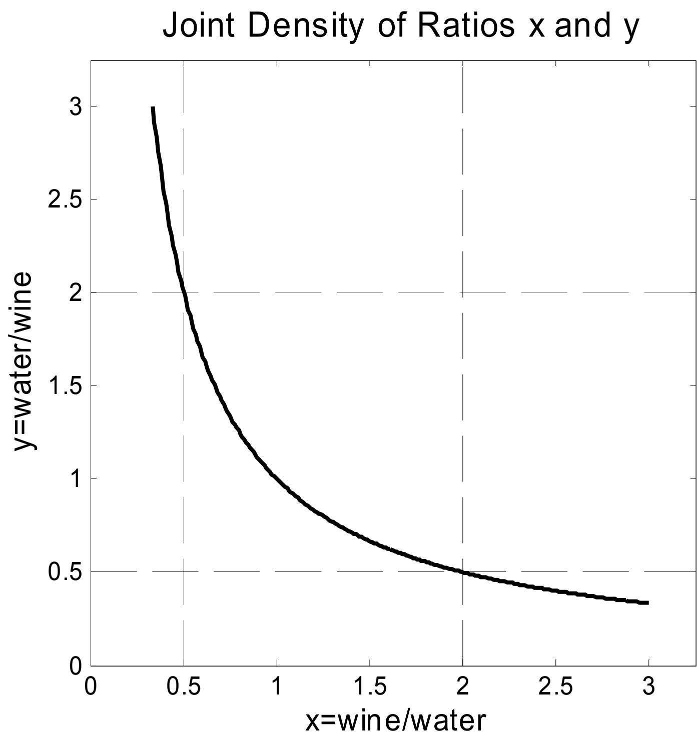

Figure 1. A two-dimensional view of the wine/water paradox. The density function would come out of the page and be a uniform height. Dashed lines are drawn for reference to the probabilities of interest.

Although the mathematics provides some additional insight, the solution to the paradox and the appeal to joint indifference can be easily appreciated when considering the graph of $p(x, y)$ (Figure 1). The x-axis corresponds to x values and the y-axis to y values. The curve is the set of allowable points or domain of $p(x, y)$. If we are indifferent about pairs (x, y) then the height or density of $p(x, y)$ is constant and would come out of the page. Formally, the joint density is

$$p(x, y) = k \ , \ \tfrac{1}{3} \leq x, y \leq 3; \ y = 1/x$$

where k is a constant such that the total probability is equal to one. This formal definition represents the only given information being used to solve the problem. Noting the dashed lines for reference, it is immediately obvious that $P(x \leq 2)$ is the length of curve for $x \leq 2$ divided by the total length of the curve and that

$$P(x \leq 2) = P(x \geq \tfrac{1}{2}) = P(y \leq 2) = P(y \geq \tfrac{1}{2})$$

The specific value of this probability using the joint density is about 0.764 which is not equal to either of the other two solutions above. This solution is unique in that it is the only solution that corresponds to joint indifference over the ratios. The symmetry is obvious in the figure, although symmetry does not justify the solution. The paradox arises when we try to consider a ratio separately from its inverse, but clearly these objects are dependent, related by the expression $y = 1/x$. Applying PI sequentially to x and then to y violates this dependency. As well there is no need to transform the original problem into something more suitable. Transforming the original example and applying PI to this transformed problem yields a solution to the transformed problem, not to the original example. It is interesting to note that the naïve average of the two paradoxical solutions $5/8$ and $15/16$ is 0.781, which is quite close but not identical to the joint solution.

Joint indifference is the natural measure for numerous probabilistic situations, and it is not difficult to imagine examples where multidimensional thinking is necessary.[3] Tossing two dice is a physical example. If the two dice are independent then I can appeal to marginal indifference. When the dice are correlated in some way then I must appeal to the joint description. Here is another example that would create a paradox or at least much confusion under standard analysis. As well, none of the transform methods above can resolve it.

> Suppose I have two four-sided dice, A and B, that can take on values between 1 and 4. The dice are related such that B≤5-A when tossed together. I toss both dice. What is the probability that at least one of the dice lands on 1?

One can analyze this example in so many wrong ways using standard analysis and thus I do not wish to open the door. Using joint indifference it is simple. Imagine a 4x4 grid corresponding to the possible (A, B) pairs. Of these, the constraint B≤5-A allows 10 of these 16 pairs. Of these 10 allowed pairs (A, B), 7 of them contain a 1. The answer is $7/10$. This probability seems entailed by the problem if we are not given any additional information.[4]

It is doubtful that any practitioner of applied probability would see the joint solution to the wine/water paradox as exotic, although it is somewhat odd that joint indifference has not been applied to the wine/water example before. If we are concerned about a ratio and its inverse, and these two objects are observed as a pair like tossing two dice, then a multidimensional analysis is demanded by the problem. If a ratio and its inverse were probabilistically independent then joint indifference would not have been required. This is not the case.

Sample Space

The wine/water paradox is resolved without transformation of the problem by appealing to the appropriate *attribute space* or *sample space* for the problem as stated. The example tells us that there is a mixture of wine and water, and that there is 3 times as much of one as the other. I am indifferent to the pair (x, $1/x$) because I have no reason to prefer one ratio and its inverse over any other pair. As well a ratio and its inverse are necessarily related – I cannot assign probability to one without fixing the probability of the other. These two reasons along with a unique solution that requires no transformation of the original problem suggests that PI works fine so long as I follow the constraints I am given in the problem. The sample space is formally described in set theoretic notation as

$$\Omega = \{(x, 1/x) \in \mathbf{R} \times \mathbf{R} : 1/3 \le x \le 3\}$$

where $\mathbf{\Omega}$ is the sample space and $\mathbf{R} \times \mathbf{R} = \mathbf{R}^2$ is the real plane. In words, this set contains the pairs (x, $1/x$) for x between the real numbers $\frac{1}{3}$ and 3. The paradox results when one attempts to consider a different sample space such as the 1-dimensonial range of wine/water ratios. You might suppose that two 1-dimensonial sample spaces are an appropriate model for this 2-dimensional entity. In probability theory, that situation only arises when two variables are independent. In the wine/water paradox the ratio and its inverse are clearly dependent. Only a 2-dimenional sample space will do.

The wine/water paradox derives its mettle from an inappropriately specified sample space; however, sample space dilemmas seem to be the crux of all other counterexamples to PI. By this I mean, the common feature unifying counterexamples to PI is uncertainty in identifying the 'correct' sample space for the problem. Certainly Bertrand's geometrical paradoxes share sample space specification difficulties. In the famous chord paradox, Bertrand asks us to 'select a chord at random', although this command is not further clarified. Given the vagueness of this question, Bertrand draws us into the paradox by illustrating three different ways of selecting chords. In other words, he suggests three distinct sample spaces from which to select elements. After selecting an element from the given sample space, he maps this element to the appropriate chord length. Depending upon the sample space from which we choose outcomes, we compute a different probability for the quantity of interest. This dependency on sample space has nothing to do with a deficiency in PI; rather, it suggests that *the probability of an outcome is* *fundamentally dependent upon the specific sample space in which the outcome is contained*. Accepting this last point makes most paradoxes to PI seem much less interesting, but only if such a position is justified.

The notion that conditional probability is primitive is not new, although the prevailing view is that conditional probabilities should be understood as ratios in accordance with Kolmogorov's axiomatization of probability. Notably, Hajek [2003a] rejects the ratio formulation of conditional probability, proving that any probability assignment has uncountably many 'trouble spots' which create inconsistencies under the ratio analysis of conditional probability. He suggests that we must take conditional probability as the primitive notion, and analyze unconditional probability in terms of it. Other authors such as de Finneti [1990] contend that all probabilities are conditional. Hajek [2003a] reverently accuses the dogmatization of Kolmogorov's axioms as the source of confusion surrounding notions of conditional probability. This may be so; however, it appears that Kolmogorov and Hajek are in implicit agreement, at least in some sense.

While the Kolmogorov axiomatization is not immutable law, most work in probability theory has taken Kolmogorov's approach for its foundation and I will proceed likewise. Kolmogorov begins with a 'universal set' Ω which is often called the 'sample space' or 'attribute space' as above. In abstract mathematical probability theory, Ω is taken to be an arbitrary non-empty set. Any set will do. From Ω we define a field F of subsets of Ω and a probability measure P on this field, completing the so-called probability space three-tuple (Ω, F, P). While the set Ω is arbitrary, it is not inconsequential. Specifically, the probability of two 'equivalent' events may differ given they are elements of different fields derived from different sample spaces.

A simple finite sample space example is sufficient to illustrate the situation. Suppose we have a set $\Omega_1 = \{x \in \mathbf{Z} : 1 \leq x \leq 10\}$ and a set $\Omega_2 = \{x \in \mathbf{Z} : 1 \leq x \leq 100\}$. What is the probability of randomly selecting the element x=8, which we will call event A? Without defining what a random selection is exactly, a reasonable response for the probability is 1/10 or 1/100, depending upon the set you use for selection. While one may disagree with the use of indifference to analyze this problem, I cannot imagine a persuasive argument that insists $P(A)$ *must* be the same in each sample space. If you do maintain that $P(A|\Omega_1) = P(A|\Omega_2)$, then you should demand that all probabilities in Ω_1 transfer over to Ω_2 as well. However, then we have just made any definition of Ω_2 obsolete, as all members of Ω_2 that are not members of Ω_1 will be assigned a probability of zero. Further, by what principle should probabilities in Ω_1 transfer over to Ω_2 and not the other way around? Perhaps this situation seems possible in abstract set theory, but as soon as we map the abstract problem to a physical example it becomes even less tenable. It's as though we started with a bag of ten marbles labeled 1 to 10 and assigned a probability function for the selection of various marbles, then added ninety more marbles to the bag and claim that these additional marbles in no way effect the probability function on the original ten marbles. While everyone will reply 'the experiment has changed', and I agree; one type of change can be represented in abstract probability theory and the physical situation equally well – a change in the set of fundamentally possible options.

Conditioning all probabilities on a given sample space is entirely consistent within Kolmogorov's axioms, including his ratio definition of conditional probability. He tells us that $P(\Omega) = 1$ by definition and that for an arbitrary event A in $\mathbf{\Omega}$

$$P(A \mid \Omega) = P(A \,\&\, \Omega) / P(\Omega) = P(A \,\&\, \Omega) = P(A)$$

by the ratio analysis of conditional probability. Usually we drop all reference to the sample space when using mathematical symbols to manipulate probability expressions. I believe that this is notational non-rigor that opens the door to paradoxes. Using standard notation, suppose that A is a member of both Ω_1 and Ω_2 as in the example above, then $P(A \mid \Omega_1) = P(A)$ and $P(A \mid \Omega_2) = P(A)$ because $P(\Omega_1) = P(\Omega_2) = 1$. As you can see this notation leads to confusion, for the probability of event A can differ depending upon the sample space being considered. Gilles [2000b, p. 828] makes a similar observation, conditioning each event A upon a set **S** of repeatable conditions under a propensity interpretation of probability, and even earlier von Mises did the same for the collective under a frequency interpretation. [5] To avoid confusion with standard conditional probabilities, it may be clearer to use the representation $P_\Omega(A)$ and call this 'the probability of A in Ω'. Here the sample space is the conditional quantity of interest; an object that potentially has relevance under all interpretations of probability. In the above example there is no reason to think that $P_{\Omega_1}(A)$ will be equal to $P_{\Omega_2}(A)$, nor is there any reason to suppose that one sample space is correct and the other is incorrect. The spaces correspond to different questions of probability. The same event may be contained within a space of fundamentally different possible outcomes.

While I have focused on Kolmogorov's axiomatization of probability to make this argument, it seems as though any notion of probability that starts with a universal set or language **Ω** and demands that $P(\Omega)=1$ must condition all probabilities on the object **Ω**. If a particular event can belong to two or more sample spaces, then it is possible that the probability of this event may differ depending upon the sample space in which it is contained.

The negated probabilistic event

Every question of probability entails at least one well-defined sample space. When I ask for the probability of event *A*, the sample space $\Omega=\{A,\neg A\}$ logically follows. While the event *A* may not cause any problems, the event $\neg A$ requires careful consideration. On the surface $\neg A$ may appear to be a well-defined object in no need of clarification, but a simple example suggests this is not the case. Suppose event *H* refers to a coin landing on heads. Let us take a propensity interpretation of probability where the experimental coin-toss setup is given by **S**, and the resulting propensity of *H* is given by $P(H\,|\,\mathbf{S})=p$. For consistency the propensity interpretation must acknowledge that $P(\neg H\,|\,\mathbf{S})=1-p$. Since $\neg H$ has non-zero probability, then it makes sense to ask for instance, given experimental coin-toss setup **S**, what is the probability of 'Mona Lisa'? Since 'Mona Lisa' is not *H*, and $\neg H$ has non-zero probability for *p* less than one, we cannot logically exclude 'Mona Lisa' as an outcome of the experiment.

One may reply that 'Mona Lisa' is not associated with **S** in anyway, and therefore the probability of 'Mona Lisa' is zero. This would suggest that $\neg H$ is not a purely logical negation of *H*, but rather a conditional logical negation constrained by **S**. We cannot allow $\neg H$ to refer to Mona Lisa because this association is not possible given **S**. Consider then an event which is not *H* that is also associated with **S**, perhaps the event *L*='the coin lands to the left of its original position'. It seems reasonable to claim that *L* is logically $\neg H$, as well *L* is a possibility given **S**. However, a coin can certainly land on heads and to the left of its original position, but this would imply that both *H* and $\neg H$ have occurred, a certain contradiction.

Our inclination is to consider the event *T*='the coin lands tails' and demand that $T=\neg H$, but how does the original propensity proposition that $P(H\,|\,\mathbf{S})=p$ entail such an equivalence?

While it does demand that $P(\neg H \mid \mathbf{S}) = 1 - p$, specifying the propensity of H does not clarify the nature of $\neg H$. Logically $\neg H$ may be a vast collection of events associated with **S** which are not *H*. To avoid such difficulties one often assumes that *H* is a member of a set of outcomes **Ω**; a set that we assume exists and has select membership but needs no justification. I find this cursory attitude toward the sample space troubling. In abstract probability theory it may be sufficient to state that **Ω** is arbitrary yet necessary, but a careful analysis of probability requires a bit more. So I ask, is **Ω** given and arbitrary, or is it somehow entailed?

To construct a sample space from scratch we require a theoretical framework in which to interpret the negated event. Under a propensity interpretation of probability, the event $\neg H$ refers to an event that is logically not *H* that cannot occur in simultaneous conjunction with *H* and is possible given experimental setup **S**. Let us build up this object in pieces for a coin toss. Define $\Omega_{\mathbf{S}}$ to be the set of nomologically possible attributes of a coin given an experimental setup **S**. For example, certain outcomes like Mona Lisa should not be allowed given a typical coin toss, although the position of the coin in reference to a particular axis should. The set $\Omega_{\mathbf{S}}$ is much larger than $\{H, T\}$, including continuous and discrete measurements on any conceivable quantity associated with a coin toss. The elements of $\Omega_{\mathbf{S}}$ must have the same epistemic dimensionality as the given event *H*. For instance, the observation of a coin landing on tails and to the left of midline has an epistemic dimensionality of two. As such, if *R* and *L* represent right and left of midline, then for two-dimensional observations $\Omega_{\mathbf{S}}$ is written as $\{(H,R),(H,L),(T,R),(T,L)...\}$ rather than $\{H,T,R,L,...\}$ as for one-dimensional observations. Given the universal experimental set $\Omega_{\mathbf{S}}$, events which are logically not *H* correspond to the complement H^{c} in $\Omega_{\mathbf{S}}$ which may be written $\Omega_{\mathbf{S}\setminus H}$. Lastly, we must exclude all elements in $\Omega_{\mathbf{S}\setminus H}$ that may occur in simultaneous conjunction with *H*, giving $\neg H = \{x \in \Omega_{\mathbf{S}\setminus H} : x \,\&\, H \text{ is logically impossible}\}$. For a particular experimental setup **S** and attribute *H*, the sample space of outcomes is

$$\Omega_{\mathbf{S};H} = \{H \cap \Omega_{\mathbf{S}}\} \cup \{x \in \Omega_{\mathbf{S}\setminus H} : x \,\&\, H \text{ is logically impossible}\}$$

This sample space is entailed by the given probabilistic event in question and the experimental setup. When *H* is not a member of $\Omega_{\mathbf{S}}$, $\Omega_{\mathbf{S};H} = \varnothing$. For a coin toss experiment, it makes no sense to ask about the probability of Mona Lisa. Rather than say this probability is zero, we say that the sample space is the empty set and that the probability of Mona Lisa given **S** is undefined.

The primary intent of this construction is to show that as soon as one asks for the probability of a physical event, a particular sample space of outcomes is entailed. I have done this by specifying the meaning of the negated event under a propensity interpretation of probability. A particular *H* and **S** should uniquely identify $\Omega_{\mathbf{S};H}$ as the objects in question are presumed to be unique. One might take up issue with $\Omega_{\mathbf{S}}$ since we do not necessarily know what outcomes are nomologically possible. For instance, is 'edge' a possible outcome for a coin toss? As well, attributes like color are troublesome in that the set of physically possible colors is somewhat difficult to pin down and suffers from subjectivity. Nonetheless, one must concede that the negated event $\neg H$ in a propensity based probability theory is a logical negation constrained by physical and logical possibility. It is interesting that the sample space $\Omega_{\mathbf{S};H}$ may be thought of as a determinable generated by the determinate *H*.[6] In other words, by identifying the event 'lands on heads', we automatically generate the sample space $\Omega_{\mathbf{S};H}=\{\text{heads, tails}\}$ using the construction above. In English we might designate the set $\Omega_{\mathbf{S};H}$ 'the side of the coin'.

These arguments share much in common with Hajek [2003b] when he discusses the reference class problem and how it is a problem for all meaningful interpretations of probability. Like he, I feel that the reference class problem is not a problem but a fundamental aspect of probability. In addition to his analysis I posit that the sample space **Ω** is *a* source of all conditional probabilities, and that Kolomogorov himself would approve of conditioning all probabilities on this sample space.

It may not be obvious but a reference class and a sample space are not necessarily equivalent objects. Take the classic reference class problem involving a consumptive Englishman named John Smith who is fifty years old: what is the probability that he lives to sixty-one? As the problem goes for frequentist, to compute the probability that Mr. Smith lives to sixty-one we may take the reference class to be all men, fifty year-old men, fifty year-old Englishmen, and so forth. The given set of attributes that a collection of objects share in common is a reference class. The probabilistic sample space is an entirely different object. For this problem, the sample space is **Ω**={'John Smith lives to sixty-one', 'John Smith dies before sixty-one'}. Clearly $P(\Omega)=1$, although I am unsure how to assign a probabilistic measure to the elements in this Ω. No one would necessarily expect the sample space Ω={'An Englishman lives to sixty-one', 'An Englishman dies before sixty-one'} to have the same probabilistic measure. Although the sample spaces have the same number elements, they do not identify the same events and therefore the probabilities need not be the same. Nor is one sample space correct and the other incorrect. The sample spaces correspond to different questions of probability.

Another example

One cannot separate a question of probability from the sample space in which the question is posed. The probability of a specific event depends upon the sample space in which that event is contained, where the sample space is dependent upon the experimental set-up or method of generating outcomes. The construction above illustrated this connection. One must be careful, though, to include the observing process in these objective physical situations. Here is an example that makes use of ideas in this paper – joint indifference, dependence, and careful sample space characterization – and adds observation. Suppose there is a bag containing 5 coins. Each side of each coin is labeled with a unique number between 1 and 10 such that the numbering is in order on each coin. In other words, one coin is labeled with 1 on one side and 2 on the other; the next coin is labeled with 3 on one side and 4 on the other, and so on. In the first experiment designated Ω_1 choose a coin from the bag without looking and then observe the number on the first side that you see; do not look at the other side. Let the event A='a number greater than 7 is observed' so that $P(A)$ is the corresponding probability. In the second experiment Ω_2 choose a coin from the bag and look at both sides of the coin. Again $P(A)$ is the probability that 'a number greater than 7 is observed'. Appealing to indifference, I suggest that $P(A) = 3/10$ in the first experiment and $P(A) = 4/10$ in the second experiment, yet the event A corresponds to the same outcome 'a number greater than 7 is observed'. As well, are not the sample spaces identical in each experiment?

The two experiments above are fundamentally different in that they consider the same outcomes of interest but refer to two perhaps subtlety different sample spaces. We can formally describe the sample spaces as

$$\Omega_1 = \{1,2,3,4,5,6,7,8,9,10\} \text{ and } \Omega_2 = \{(1,2),(3,4),(5,6),(7,8),(9,10)\}$$

In the first experiment we select one element in Ω_1 and ask the probability that we selected (or observed) a number greater than 7. In the second experiment we select pairs of sides such as (3, 4) and ask about the probability of a number greater than 7 in that pair. One can run the actual experiment and verify the probabilities via a frequentist interpretation. Again, joint indifference is needed to fully explain the example. The probability $P(A)$ in the sample space Ω_1 is $3/10$ while $P(A)$ in Ω_2 is $4/10$. A sample space of two-dimensional elements is not identical to the one-dimensional analog, even though the elements have much in common. There is no reason to believe that one answer is the correct answer. The two probabilities correspond to different problems.

This example is physically simple so that you can easily see that the two experiments are fundamentally different. Perhaps your intuition easily resolved the coin problem, and at no point did you ever think that a paradox existed. Of course the probabilities do not coincide; in the first experiment only one observation was made and in the second experiment two observations were made on the coin. However, the probability in the second experiment is not double the probability of the first experiment. To resolve and understand this example we need to carefully specify the sample spaces of interest. This problem is about comparing a one-dimensional sample space to a two-dimensional sample space and noting that the probability of a particular event is changed. Bertrand's classic geometrical counterexamples to PI also contain comparisons of one sample space to another. Whenever a particular event is evaluated in a different sample space, the probability of that event may differ. Each sample space may have relevance in the sense that each corresponds to a different physical example or a different way of observing outcomes.

Bertrand and friends

I would like to return to Bertrand's chord paradox as stated in Gilles [2000a]

> Consider a fixed circle and select a chord at random. What is the probability that this random chord is longer than the side of the equilateral triangle inscribed in the circle?

Given this statement of the problem and no other information, what sample space is entailed by the problem? This sample space question is the source of confusion in Bertrand's paradox; not the applicability of PI. Just as an ambiguously defined abstract set invites confusion in sample space specification, Bertrand's paradox does the same for more physical sets. Considering

Bertrand's question as stated and appealing to no other information, two entailed sample spaces seem reasonable. The two possible samples spaces are (1) the set of all chord lengths $C = \{x : 0 \le x \le 2\mathrm{R}\}$ where R is the radius, in which case the solution to the paradox is trivial, and (2) the collection of all possible sets *A* defined on a circle of radius R such that for any $c \in C$ there exists an $a \in A$ for which $c = f(a)$. This particular definition of the set *A* in (2) guarantees that every chord length in *C* is *selectable* via one or more elements in *A*. Here we have made rigorous the notion of selecting a chord. Bertrand tells us to select a chord, but he does not say how. Note that (2) demands that the collection of selectable sets is logically entailed but not any individual set *A* by itself. Reference to chord lengths is given in the problem, making (1) a logical possibility. The three conflicting probabilities typically given to this question correspond to three arbitrary sets *A* belonging to the *infinite* collection of sets in (2) that allow for selection.[7] To be fair, these sets are not wholly arbitrary; they were chosen because of computational ease for the probability of interest.[8]

Another Bertrand-type example (Hajek [2003b] adapted from van Frasssen [1989]) illustrates that opponents of PI may be ignoring basic theorems in probability theory and the importance of a given sample space. In this example, a factory produces cubes with side-lengths between 0 and 1 foot: what is the probability that a randomly chosen cube has side-length between 0 and ½ foot? Your intuitive response is ½, as you assumed that the factory produces sides with a uniform distribution in length. The paradox ensues by giving an 'equivalent' restatement of the problem: suppose the factory produces cubes using a process by which the face-area of a cube is uniformly distributed; what is the probability that a randomly chosen cube has a face area between 0 and ¼ feet? Again, using intuition your response is ¼, as you assumed that face-area was uniformly distributed in production. Since a face-area less than ¼ and a side-length less than ½ both correspond to a cube-volume less than ⅛, it seems as though these two probabilities should be identical, but somehow intuition has led us astray. Actually, both answers may be correct; we must only prove that a production process uniformly distributed in side-lengths produces a different distribution of cube volumes than a process uniformly distributed in face-areas. We may do this rigorously, although a figure illustrates the point cleanly (Figure 2).

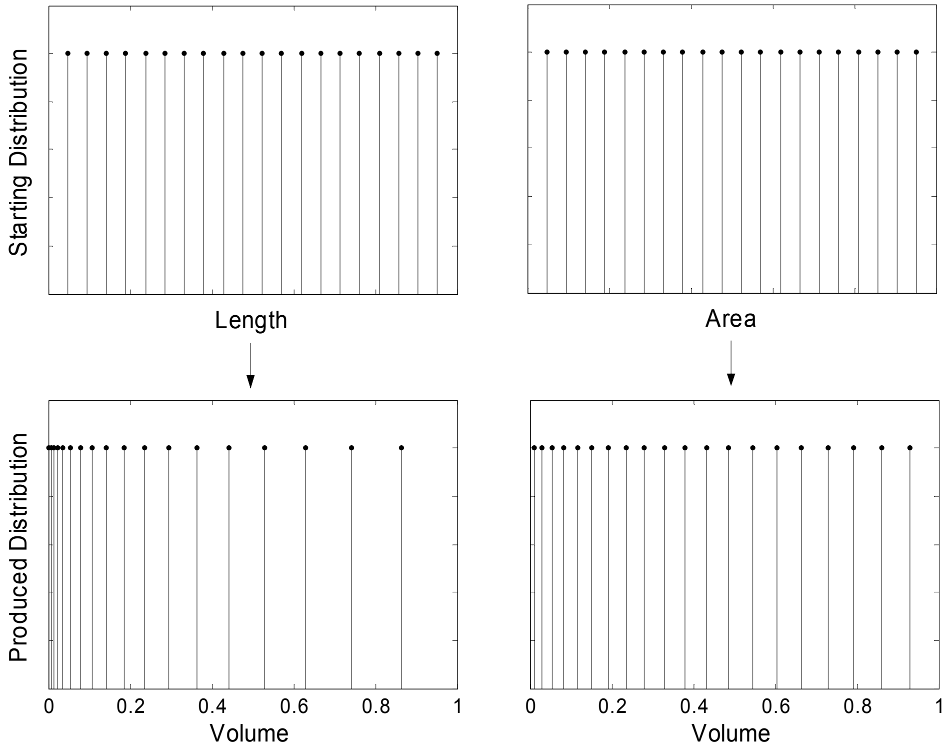

Figure 2. The top two panels represent uniformity in length and area. The bottom two panels show how these uniform distributions change when mapped to volume. A process uniform in length produces a different set of volumes than a process uniform in area.

Suppose a factory produced cubes using 20 different side-lengths uniformly spread over the interval (0,1), while a second factory used 20 different face-areas also uniformly spread. The lines in the two top panels correspond to the 20 side-lengths and face-areas used in the process. Given those starting quantities, the bottom panels show the 20 cube volumes produced using those 20 lengths or areas. Clearly uniformity over side-length or face-area does not result in uniformity of cube volume. Further, the spread of cube volumes is different given we started with uniformity in side-length or face-area. The probability that a randomly chosen cube has a volume less than ⅛ can be estimated by the proportion of cube volumes less than ⅛, giving 10/20 for uniformly distributed lengths and 5/20 for uniformly distributed face-areas (equivalence to the exact answers is coincidental). It is straightforward to calculate the exact continuous density of cube volumes given we started with a uniform density over side-lengths or face-areas.[9] If v is a random variable representing cube volume; then the density of cube volumes given uniform side-lengths is $p_{L\to V}(v) = \frac{1}{3} v^{-\frac{2}{3}}$, while the density of cube volumes given uniform face-areas is $p_{A\to V}(v) = \frac{2}{3}v^{-\frac{1}{3}}$. The subscript notation $L \to V$ on the density function denotes a mapping from the sample space of side-lengths *L* to that of cube volumes *V*. The probability that a randomly selected cube has a volume less than ⅛ is calculated by integrating each density over the interval $0 \le v \le \frac{1}{8}$, giving $P_{L\to V}(v \le \frac{1}{8}) = \frac{1}{2}$ and $P_{A\to V}(v \le \frac{1}{8}) = \frac{1}{4}$. How do we determine whether an actual factory uses a process uniform in side-lengths or face-areas? That is a good question, but it has nothing to do with the admissibility of PI.

This example as stated does not create a logical inconsistency under PI. Of course, we have not uniquely identified the unqualified probability in question – the probability that a randomly selected cube has a volume less than ⅛. We have instead proved that the probability will be dependent upon how the factory makes cubes. If it makes cubes through a process uniform in side-lengths, we know the answer is ½, whereas a process uniform in face-areas gives an answer of ¼. There is no mathematical trickery. Length and area are related to volume in different ways, and we must appropriately account for these differences when computing probabilities.

For physical experiments, the method of generating and observing outcomes determines the sample space in a question of probability, whereas in abstract examples of probability the sample space is logically entailed by the specifics of the question. Numerous authors have already suggested – and I agree – that Bertrand's chord paradox has no definitive answer because it harbors inherent ambiguity.[10] For instance, von Mises [1957, pp 80-1] states that

> "…the assumption of a 'uniform distribution' means something different in different co-ordinate systems. No general prescription for selecting 'correct' co-ordinates can be given, and there can therefore be no general preference for one of the many possible uniform distributions."

What von Mises calls coordinates I call the sample space. By identifying a different coordinate system we identify a different continuous sample space. I have suggested that all probabilities are fundamentally contingent upon the sample space in which they reside, and therefore it is no surprise that applying PI to these different sample spaces yields different results. What does not follow is the claim that PI, when applied within an identified sample space, is somehow inconsistent or unjustified. The real meat of Bertrand's paradox is in envisioning a very specific physical method of chord selection and trying to map that selection process to the appropriate sample space (or coordinate system) which encompasses our indifference appropriately. Selecting a chord at random may mean different things to different individuals, and thus probability assignments may vary. Once we agree upon the meaning of chord selection, or agree upon a constrained physical experiment; we should all derive the same probability.

As von Mises states, there is 'no general preference' for mapping a physical process to the appropriate sample space, but certainly there are rational ways of thinking that point us toward one sample space versus another. For instance, given a random wheel spin of continuous values labeled between 0 and 1, I should choose the interval $0 \le x \le 1$ as my abstract sample space and spread probability uniformly over those values. I could have chosen the interval $\exp(0) \le y \le \exp(1)$, and spread probability uniformly over those values as well. When you ask me to calculate $P(x > 0.5)$, I could tell you this quantity corresponds to $P(y > \exp(0.5))$ and present to you a probability that is not equal to one half. You might be inclined to call this a paradox or an inadequacy of PI, for there may be no reason to prefer the sample space *x* over *y*. However, there is good reason to prefer one sample space over the other – the sample space *x* is the given object in the probabilistic problem while *y* is an arbitrary addition corresponding to a different question of probability altogether.

To summarize how paradoxes to PI are born. We start with an event *A*. One then states that *A* may be a member of numerous different sample spaces Ω_1, Ω_2, and so on. Indifference is applied within each sample space giving $P(A \mid \Omega_1) \ne P(A \mid \Omega_2)$ and the like. Since the probabilities do not agree we say that PI is inconsistent. As I have shown, the probabilities should not be expected to be equal; rather, we should be surprised if they agree. The result that $P(A \mid \Omega_1) \ne P(A \mid \Omega_2)$ says nothing about the content of PI, nor should it send us scurrying for the 'true' sample space. An example of probability may harbor irreducible ambiguity. One may identify a single sample space in these questions, but to do so requires additional assumptions or information not given in the original example. If the example does not entail this additional information, then you are not answering the original question. Whether a particular abstract sample space is an appropriate object for our indifference over the outcomes of a well-characterized physical experiment is another question entirely and seems to be fertile ground for philosophical exploration.

The Meaning of a Probabilistic Event

Above I have taken the entire sample space as a conditional object for all probabilities, in part because this notion is supported by Kolmogorov's axioms which are widely accepted. A related idea is that all probabilities are dependent upon the event in question, where the event is an element of the field *F* in the probability space (Ω, *F*, *P*), coupled to the meaning of that event. For instance, when you tell me that the probability of randomly selecting a red book is ½, your idea of red and my idea of red (and books) had better be epistemically identical or that probability is potentially meaningless. In other words, the probability $P(\mathit{red\ book}) = \frac{1}{2}$ means nothing, or not necessarily the same thing to you if we do not agree upon the meaning of red books. Let us just take the red part and see the trouble. Suppose for simplicity that red to me, RED_1, is any object that gives off wavelengths between 650nm and 750nm, while red to you, RED_2, is any object that gives off wavelengths between 625nm and 725nm. Clearly, $P(\mathit{red\ book} \mid RED_1)$ need not equal $P(\mathit{red\ book} \mid RED_2)$; rather they are almost certainly different. As another example, suppose a bag is filled with one hundred marbles colored with different shades of reds, oranges, yellows, blues, etc., and that we wish to enquire about the probability of selecting a red marble. To me, ten of the one hundred marbles are red, while to you fifteen of the one hundred marbles are red – I think some of your red marbles are orange or purple. Now we have that $P(\mathit{red}) = \frac{10}{100}$ for me and that $P(\mathit{red}) = \frac{15}{100}$ for you. Who is correct? Neither of course. Probabilities only become useful between individuals when we agree upon the meaning of the event in question. The meaning need not be objective; it need only be shared.

I beg ignorance on many topics in the philosophy of language, but it seems unavoidable that all probabilities are of the form $P(A \mid m_i(A))$ where *A* is an event and $m_i(A)$ is the meaning of the event for the i^{th} individual. Exact knowledge about the probability of an event *A* is only communicated when for two individuals *i* and *j*, $m_i(A) = m_j(A)$. Perhaps my example involving colors seems unfairly subjective. What then is the probability of rain tomorrow? If your idea of rain is only one drop of water falling from the sky, while mine is a continuous downpour for at least 1 second, then our probabilities for the event in question will not coincide. I am not suggesting that probabilities are hopelessly subjective. Rather, I hope to reinforce the notion that questions of probability are extremely particular.

I am sure that all of us assume we are speaking of the same event when we talk about the probability of a particular event *A*; however, this implicit assumption seems certain to cause trouble if not addressed. The assumption becomes especially problematic when the probabilistic question is vague at the onset. If we cannot be certain of the shared meaning for seemingly concrete examples involving a selection of colored marbles from a bag or the probability of rain, then ambiguous problems such as Bertrand's chord paradox are sure to elicit a variety of interpretations. The sample space points us toward a common meaning for the event in question. Once we agree upon a given sample space, we may begin to speak about the probability of an event in that sample space intelligibly. Otherwise we are certain to find inconsistency and paradoxes all around.

I am not going to further address the important question of how to logically determine the appropriate sample space given a particular question of probability or physical experiment. Certain questions of probability, such as the wine/water example and the dice example, entail a multidimensional sample space. Other examples, such as the chord paradox, may imply a collection of selectable sets or sets with some other characteristic and not single out a particular set as the true and only set. While one may appeal to symmetry for guidance in choosing an individual set, it is not obvious that symmetry arguments follow from questions of probability, nor does symmetry always lead to unique answers as we have shown.

Appendix A: Transforms of the wine/water paradox

The transforms $f(u)$ that lend themselves to an application of PI and result in consistent solutions to the wine/water paradox have the following relation

$$f(u) + f(1/u) = \text{constant}$$

where u is a ratio. Any transform satisfying this condition will assure that PI can be applied consistently without paradox. For the log transform, $f(u) = \log(u)$ and $f(1/u) = \log(1/u)$, giving the sum

$$\log(u) + \log(1/u) = 0$$

Mikkelson's solution belongs to an infinite family of solutions of the form

$$f(u) = u^k/(1+u^k) \text{ such that } u^k/(1+u^k) + 1/(1+u^k) = 1$$

where k is an arbitrary positive constant equal to one for Mikkelson's solution. The proof of this statement follows directly from an observation in probability theory – if the sum of 2 random variables is constant, and the density of one of the variables is uniform, then the density of the other random variable will also be uniform. Every value of k generates a different but consistent solution to the wine-water paradox. A formal proof is beyond the scope of this work, although another example may further suggest it is true. Consider the case where k=2. Recall that the original ratios occupy the interval $\frac{1}{3} \le x \le 3$. Transform this interval to the new interval with the formula $b = f(x) = x^2/(1+x^2)$, giving $\frac{1}{10} \le b \le \frac{9}{10}$. The probability $P(x \le 2)$ corresponds to $P(b \le \frac{8}{10})$. Then

$$P(b \le \tfrac{8}{10}) = (\tfrac{8}{10} - \tfrac{1}{10})/(\tfrac{9}{10} - \tfrac{1}{10}) = \tfrac{7}{8}$$

while $P(x \ge \frac{1}{2})$ corresponds to $P(b \ge \frac{2}{10})$ giving

$$P(b \ge \tfrac{2}{10}) = (\tfrac{9}{10} - \tfrac{2}{10})/(\tfrac{9}{10} - \tfrac{1}{10}) = \tfrac{7}{8}$$

As you can see, this solution does not create a paradox under PI. The answer is invariant regardless of how the question is posed. Symmetric solutions are plentiful, so long as we transform the original problem appropriately. The probability here corresponds to the probability of the transformed question, not to the original problem.

Appendix B: Wine/water Paradox – A 2D resolution

The Principle of Indifference gives a satisfying solution to the wine-water example when our indifference is jointly distributed over the possible values for all of the ratios. I show the math to be rigorous and to calculate a specific probability, yet I envision criticism that this solution is too complicated to be correct. Going ahead, let x=wine/water and y=1/x=water/wine. We are interested in the joint density of x and y, which when using PI will be of the form

$$p(x,y) = k\ ,\ \tfrac{1}{3} \le x, y \le 3;\ y = 1/x$$

where k is a constant density chosen so that the total area under $p(x,y)$ is equal to one. This is analogous to the 1-D example where a constant density is assigned to either the marginal density of x or y. The value of k above is not equal to the constant in the 1-D example. We must calculate k by first computing the arc-length of the 2-D curve swept out in the x-y plane, where the curve is given by the boundary conditions in the problem statement (Figure 1). As a physical analogy, this is equivalent to calculating the length of string on the ground that curves with the function y=1/x. The general formula for arc length is

$$L = \int_a^b \sqrt{1 + [f'(u)]^2}\, du$$

where $f'(u)$ is the derivative of our curve. Substituting values for the current example gives

$$L = \int_{1/3}^{3} \sqrt{1 + \left(-\tfrac{1}{x^2}\right)^2}\, dx \approx 4.2932$$

which gives k=1/L≈0.2329, the height of our constant density function. Now we can compute the various probabilities. For instance, $P(x \le 2)$ is all of the probability of $p(x, y)$ for which x≤2. This in turn is given by the constant density k multiplied by the curve length for x≤2. Therefore

$$P(x \le 2) = k\int_{1/3}^{2} \sqrt{1 + \left(-\frac{1}{x^2}\right)^2}\, dx \approx 0.7637$$

which was evaluated numerically. Similarly,

$$P(x \ge 1/2) = k\int_{1/2}^{3} \sqrt{1 + \left(-\frac{1}{x^2}\right)^2}\, dx \approx 0.7637$$

The symmetry of the problem when viewed from the 2-D perspective readily gives

$$P(x \le 2) = P(x \ge 1/2) = P(y \le 2) = P(y \ge 1/2) \approx 0.7637$$

The only information we are given at the start of this example is that there is a mixture of wine and water and that at most there is 3 times as much of one as of the other. This answer is unique in that it is the only solution corresponding to joint indifference of a ratio and its inverse. The answer is symmetric as intuition suggests it should be, although symmetry does not justify the solution.

References

Bartha, P. and Johns, R. [2001]: 'Probability and Symmetry', *Philosophy of Science* 68 (Proceedings), pp. S109-S122

Carnap, R. [1950]: *Logical Foundations of Probability*, Chicago: University of Chicago Press.

de Finetti, B. [1990 (originally published 1974)]: *Theory of Probability*, Vol. 1, Wiley Classics Library, John Wiley & Sons.

Johnson, W. E. [1921]: *Logic*, Part I, Cambridge: Cambridge U.P.

Keynes, J. M. [1921]: *A Treatise on Probability*, London/New York: Routledge.

Gillies, D. [2000a]: *Philosophical Theories of Probability*, London/New York: Routledge.

Gillies, D. [2000b]: 'Varieties of Propensity', *Brit. J. Phil. Sci.*, 51, pp. 807-835.

Jaynes, E. T. [1973]: 'The Well-Posed Problem', *Foundations of Physics*, 3, pp. 477-92.

Hajek, A. [2003a]: 'What Conditional Probability Could Not Be', *Synthese*, 137, No. 3, pp. 273-323.

Hajek, A. [2003b]: 'Conditional Probability is the Very Guide of Life", in *Probability is the Very Guide of Life: The Philosophical Uses of Chance*, eds. Henry Kyburg, Jr. and Mariam Thalos, Open Court, pp. 183-203.

Mikkelson, J. M. [2004]: 'Dissolving the Wine/Water Paradox', *Brit. J. Phil. Sci.*, 55, pp. 137-145.

Bertrand, J. [1889]: *Calcul des probabilitiés*, Paris: Gauthier-Villars.

Papoulis, A. and Pillai, S. U. [2002] *Probability, Random Variables, and Stochastic Processes*, 4th ed., McGraw-Hill, Inc.

von Mises R. [1957]: *Probability, Statistics and Truth*, revised English edition, New York: Macmillan

van Frasssen, B. [1989]: *Laws and Symmetry*, Oxford: Clarendon Press.

Kendall M.G., Moran P.A.P. [1963]: *Geometrical Probability*, London: Charles Giffin & Co. Ltd.

Footnotes

[1] The probability that the ratio x is less than two is calculated assuming a uniform distribution over the interval $1/3 \leq x \leq 3$, such that $P(x \leq 2)=(2-1/3)/(3-1/3)=5/8$. The analogous quantity for the inverse ratio y is given by $P(y \geq 1/2)=(3-1/2)/(3-1/3)=15/16$, again assuming a uniform distribution.

[2] See Bartha and Johns [2001] for an approach to probability that appeals fundamentally to symmetries. They argue that PI results from the symmetrical treatment of possible outcomes.

[3] Suppose two wheels x and y are labeled with values between 0 and 1. When both wheels are spun, it so happens that the value of wheel x is always less than that of wheel y. What are $P(x \leq 0.5)$ and $P(y \leq 0.5)$? By drawing the graph of the allowed values of (x,y), one can show that $P(x \leq 0.5)=3/4$ and $P(y \leq 0.5)=1/4$.

[4] I further explain the two dice example. Define the sets $A=\{1, 2, 3, 4\}$ and $B=\{1, 2, 3, 4\}$ representing the dice. Formally, the relation B≤5-A over a set A and a set B is a subset of the Cartesian product $A \times B$. In the dice example, the relation abbreviated as R may be written as

$$R = \{(1,1),(1,2),(1,3),(1,4),(2,1),(2,2),(2,3),(3,1),(3,2),(4,1)\} \subset A \times B$$

Although 16 dice pairs are conceivable, only ten of these pairs are logical alternatives given the constraint of the problem. Appealing to indifference, the probability that at least one of the dice lands on 1 is the number of elements in R that contain at least one 1 divided by the total number of elements in R.

[5] von Mises [1957] was the first to draw this distinction. In his frequency interpretation of probability he starts with an empirical collective C and defines the probability of the attribute A in C as the limiting frequency. I am suggesting that we take the attribute or sample space Ω as our conditional quantity.

[6] Johnson [1921] introduced the terms determinable and determinate to apply to the idea that being red, for instance, is a determinate of the determinable color. We might say that the head of a coin is a determinate of the determinable side.

[7] Here is a relatively simple infinite collection of selectable sets. Choose N≥2 random points on the perimeter of the circle with radius R. Draw the line of the best fitting linear regression through that set of N points. This line is unique, always intersects the circle, and creates chords lengths between 0 and 2R. Each value of N will generate a different probability in Bertrand's chord example, although calculating these probabilities seems non-trivial.

[8] See Gilles [2000: 38-42] for a brief but clear presentation of Bertrand's chord paradox.

[9] There is a theorem in probability theory that tells one how to derive the density of a random variable given the density of a related random variable. Suppose x and y are random variables, $y=g(x)$, and that g is a one-to-one mapping. If the density of x is given, then the density of y is computed by the formula $p(y)=p(x)*|dx/dy|$. Take y to be cube volume and x to be either side-length or face-area. For x as uniformly distributed side-length, $p(y)=1/3*y$^(-2/3), while for x as uniform face-area, $p(y)=2/3*y$^(-1/3). The distribution of cube volumes depends upon our given sample space *and* indifference. See Papoulis and Pillai [2002, pp123-137] for a far more rigorous presentation of this theorem.

[10] See Jaynes [1973, p 1] for a list of authors that believe no resolution exists. Included are Poincaré, Bertrand, and von Mises. Jaynes himself points to a specific answer to the paradox with an appeal to symmetry arguments.