The Well-Posed Problem

Edwin T. Jaynes

Department of Physics, Washington University, St. Louis, Missouri

Received June 1, 1973

Many statistical problems, including some of the most important for physical applications, have long been regarded as underdetermined from the standpoint of a strict frequency definition of probability; yet they may appear well posed or even overdetermined by the principles of maximum entropy and transformation groups. Furthermore, the distributions found by these methods turn out to have a definite frequency correspondence; the distribution obtained by invariance under a transformation group is by far the most likely to be observed experimentally, in the sense that it requires by far the least "skill." These properties are illustrated by analyzing the famous Bertrand paradox. On the viewpoint advocated here, Bertrand's problem turns out to be well posed after all, and the unique solution has been verified experimentally. We conclude that probability theory has a wider range of useful applications than would be supposed from the standpoint of the usual frequency definitions.

1. Background

In a previous article (Jaynes 1968) we discussed two formal principles—maximum entropy and transformation groups—that are available for setting up probability distributions in the absence of frequency data. The resulting distributions may be used as prior distributions in Bayesian inference; or they may be used directly for certain physical predictions. The exact sense in which distributions found by maximum entropy correspond to observable frequencies was given in the previous article; here we demonstrate a similar correspondence property for distributions obtained from transformation groups, using as our main example the famous paradox of Bertrand.

Bertrand's problem (Bertrand, 1889) was stated originally in terms of drawing a straight line "at random" intersecting a circle. It will be helpful to think of this in a more concrete way; presumably, we do no violence to the problem (*i.e.,* it is still just as "random") if we suppose that we are tossing straws onto the circle, without specifying how they are tossed. We therefore formulate the problem as follows.

A long straw is tossed at random onto a circle; given that it falls so that it intersects the circle, what is the probability that the chord thus defined is longer than a side of the inscribed equilateral triangle? Since Bertrand proposed it in 1889 this problem has been cited to generations of students to demonstrate that Laplace's "principle of indifference" contains logical inconsistencies. For, there appear to be many ways of defining "equally possible" situations, and they lead to different results. Three of these are: Assign uniform probability density to (A) the linear distance between centers of chord and circle, (B) angles of intersections of the chord on the circumference, (C) the center of the chord over the interior area of the circle. These assignments lead to the results $p_A = 1/2$, $p_B = 1/3$, and $p_C = 1/4$, respectively.

Which solution is correct? Of the ten authors cited (Bertrand 1889, Borel 1909, Poincaré 1912, Uspensky 1937, Nortrup 1944, Gnedenko 1962, Kendell and Moran 1963, von Mises 1957 and 1964, and Mosteller 1965) with short quotations, in the appendix only Borel is willing to express a definite preference, although he does not support it by any proof. Von Mises takes the opposite extreme, declaring that such problems (including the similar Buffon needle problem) do not belong to the field of probability theory at all. The others including Bertrand, take the intermediate position of saying simply that the problem has no definite solution because it is ill posed, the phrase "at random" being undefined.

In works on probability theory this state of affairs has been interpreted, almost universally, as showing that the principle of indifference must be totally rejected. Usually, there is the further conclusion that the only valid basis for assigning probabilities is frequency in some random experiment. It would appear, then, that the only way of answering Bertrand's question is to perform the experiment.

But do we really believe that it is beyond our power to predict by "pure thought" the result of such a simple experiment? The point at issue is far more important than merely resolving a geometric puzzle; for, as discussed further in Section 7, applications of probability theory to physical experiments usually lead to problems of just this type; *i.e.,* they appear at first to be undetermined, allowing may different solutions with nothing to choose among them. For example, given the average particle density and total energy of a gas, predict its viscosity. The answer, evidently, depends on the exact spatial and velocity distributions of the molecules (in fact, it depends critically on position-velocity correlations), and nothing in the given data seems to tell us which distribution to assume. Yet physicists *have* made definite choices, guided by the principle of indifference, and they *have* led us to correct and nontrivial predictions of viscosity and many other physical phenomena.

Thus, while in some problems the principle of indifference has led us to paradoxes, in others it has produced some of the most important and successful applications of probability theory. To reject the principle without having anything better to put in its place would lead to consequences so unacceptable that for many years even those who profess the most faithful adherence to the strict frequency definition of probability have managed to overlook these logical difficulties in order to preserve some very useful solutions.

Evidently, we ought to examine the apparent paradoxes such as Bertrand's more closely; there is an important point to be learned about the application of probability theory to real physical situations.

It is evident that if the circle becomes sufficiently large, and the tosser sufficiently skilled, various results could be obtained at will. However, in the limit where the skill of the tosser must be described by a "region of uncertainty" large compared to the circle, the distribution of chord lengths must surely go into one unique function obtainable by "pure thought." A viewpoint toward probability theory which cannot show us how to calculate this function from first principles, or even denies the possibility of doing this, would imply severe—and, to a physicist, intolerable—restrictions on the range of useful applications of probability theory.

An invariance argument was applied to problems of this type by Poincaré (1912), and cited more recently by Kendall and Moran (1963). In this treatment we consider straight lines drawn "at random" in the xy plane. Each line is located by specifying two parameters (u, v) such that the equation of the line is $ux + uy = 1$, and one can ask: Which probability density $p(u, v)\, du\, dv$ has the property that it is invariant in *form* under the group of Euclidean transformations (rotations and translations) of the plane? This is a readily solvable problem (Kendall and Moran 1963), with the answer $p(u, v) = (u^2 + v^2)^{-3/2}$.

Yet evidently this has not seemed convincing; for later authors have ignored Poincaré's invariance argument, and adhered to Bertrand's original judgment that the problem has no definite solution. This is understandable, for the statement of the problem does not specify that the distribution of straight lines is to have this invariance property, and we do not see any compelling reason to expect that a rain of straws produced in a real experiment would have it. To assume this would seem to be an intuitive judgment resting on no stronger grounds than the one which led to the three different solutions above. All of these amount to trying to guess what properties a "random" rain of straws should have, by specifying the intuitively "equally possible" events; and the fact remains that different intuitive judgments lead to different results.

The viewpoint just expressed, which is by far the most common in the literature, clearly represents one valid way of interpreting the problem. If we can find another viewpoint according to which such problems *do* have definite solutions, *and define the conditions under which these solutions are experimentally verifiable,* then while it would perhaps be overstating the case to say that this new viewpoint is more "correct" in principle than the conventional one, it will surely be more useful in practice.

We now suggest such a viewpoint, and we understand from the start that we are not concerned at this stage with *frequencies* of various events. We ask rather: Which probability distribution describes our *state of knowledge* when the only information available is that given in the above statement of the problem? Such a distribution must conform to the desideratum of consistency formulated previously (Jaynes 1968): In two problems where we have the same state of knowledge we must assign the same subjective probabilities. The essential point is this: If we start with the assumption that Bertrand's problem has a definite solution *in spite of the many things left unspecified,* then the statement of the problem automatically implies certain invariance properties, which in no way depend on our intuitive judgments. After the subjective solution is found, it may be used as a prior for Bayesian inference whether or not it has any correspondence with frequencies; any frequency connections that may emerge will be regarded as an additional bonus, which justify its use also for direct physical prediction.

Bertrand's problem has an obvious element of rotational symmetry, recognized in all the proposed solutions; however, this symmetry is irrelevant to the distribution of chord lengths. There are two other "symmetries" which are highly relevant: Neither Bertrand's original statement nor our restatement in terms of straws specified the exact size of the circle, or its exact location. If, therefore, the problem is to have any definite solution at all, it must be "indifferent" to these circumstances; *i.e.,* it must be unchanged by a small change in the size or position of the circle. This seemingly trivial statement, as we will see, fully determines the solution.

It would be possible to consider all these invariance requirements simultaneously by defining a four-parameter transformation group, whereupon the complete solution would appear suddenly, as if by magic. However, it will be more instructive to analyze the effects of these invariances separately, and see how each places its own restrictions on the form of the solution.

2. Rotational Invariance

Let the circle have radius R. The position of the chord is determined by giving the polar coordinates (r, θ) of its center. We seek to answer a more detailed question than Bertrand's: What probability density $f(r,\theta)dA = f(r,\theta)\, r\, dr\, d\theta$ should we assign over the interior area of the circle? The dependence on θ is actually irrelevant to Bertrand's question, since the distribution of chord lengths depends only on the radial distribution

$$g(r) = \int_0^{2\pi} f(r,\theta) d\theta.$$

However, intuition suggests that $f(r,\theta)$ should be independent of θ, and the formal transformation group argument deals with the rotational symmetry as follows.

The starting point is the observation that the statement of the problem does not specify whether the observer is facing north or east; therefore if there is a definite solution, it must not depend on the direction of the observer's line of sight. Suppose, therefore, that two different observers, Mr. X and Mr. Y, are watching this experiment. They view the experiment from different directions, their lines of sight making an angle α. Each uses a coordinate system oriented along his line of sight. Mr. X assigns the probability density $f(r, \theta)$ in his coordinate system S; and Mr. Y assigns $g(r, \theta)$ in his system S_α. Evidently, if they are describing the same situation, then it must be true that

$$f(r, \theta) = g(r, \theta - \alpha) \tag{1}$$

which expresses a simple change of variables, transforming a fixed distribution f to a new coordinate system; this relation will hold whether or not the problem has rotational symmetry.

But now we recognize that, because of the rotational symmetry, the problem appears exactly the same to Mr. X in his coordinate system as it does to Mr. Y in his. Since they are in the same state of knowledge, our desideratum of consistency demands that they assign the same probability distribution; and so f and g must be the same function:

$$f(r, \theta) = g(r, \theta). \tag{2}$$

These relations must hold for all α in $0 \le \alpha \le 2\pi$; and so the only possibility is $f(r, \theta) = f(r)$.

This formal argument may appear cumbersome when compared to our obvious flash of intuition; and of course it is, when applied to such a trivial problem. However, as Wigner (1931) and Weyl (1946) have shown in other physical problems, it is this cumbersome argument that generalizes at once to nontrivial cases where our intuition fails us. It always consists of two steps: We first find a transformation equation like (1) which shows how two problems are related to each other, irrespective of symmetry; then a symmetry relation like (2) which sates that we have formulated two equivalent *problems.* Combining them leads in most cases to a functional equation which imposes some restriction on the form of the distribution.

3. Scale Invariance

The problem is reduced, by rotational symmetry, to determining a function $f(r)$, normalized according to

$$\int_0^{2\pi} \int_0^R f(r)\, r\, dr\, d\theta = 1. \tag{3}$$

Again, we consider two different problems; concentric with a circle of radius R, there is a circle of radius aR, $0 < a \le 1$. Within the smaller circle there is a probability $h(r)\, r\, dr\, d\theta$ which answers the question: Given that a straw intersects the smaller circle, what is the probability that the center of its chord lies in the area $dA = r\, dr\, d\theta$?

Any straw that intersects the small circle will also define a chord on the larger one; and so, within the small circle $f(r)$ must be proportional to $h(r)$. This proportionality is, of course, given by the standard formula for a conditional probability, which in this case takes the form

$$f(r) = 2\pi h(r) \int_0^{\alpha R} f(r)\, r\, dr, \quad 0 < a \le 1, \;\; 0 \le r \le aR. \tag{4}$$

This transformation equation will hold whether or not the problem has scale invariance.

But we now invoke scale invariance; to two different observers with different size eyeballs, the problems of the large and small circles would appear exactly the same. If there is any unique solution independent of the size of the circle, there must be another relation between $f(r)$ and $h(r)$, which expresses the fact that one problem is merely a scaled-down version of the other. Two elements of area $r\, dr\, d\theta$ and $(ar) d(ar) d\theta$ are related to the large and small circles respectively in the same way; and so they must be assigned the same probabilities by the distributions $f(r)$ and $h(r)$, respectively:

$$h(ar)\,(ar)\,d(ar)\,d\theta = f(r)\,r\,dr\,d\theta$$

or

$$a^2 h(ar) = f(r) \tag{5}$$

which is the symmetry equation. Combining (4) and (5), we see that invariance under change of scale requires that the probability density satisfy the functional equation

$$a^2 f(ar) = 2\pi f(r) \int_0^{aR} f(u)\,u\,du, \quad 0 < a \leq 1, \quad 0 \leq r \leq R. \tag{6}$$

Differentiating with respect to a, setting $a = 1$, and solving the resulting differential equation, we find that the most general solution of (6) satisfying the normalization condition (3) is

$$f(r) = \frac{qr^{q-2}}{2\pi R^q} \tag{7}$$

where q is a constant in the range $0 < q < \infty$, not further determined by scale invariance.

We note that the proposed solution B in the introduction has now been eliminated, for it corresponds to the choice $f(r) \sim (R^2 - r^2)^{-1/2}$, which is not of the form (7). This means that if the intersections of chords on the circumference were distributed in angle uniformly and independently on one circle, this would not be true for a smaller circle inscribed in it; *i.e.,* the probability assignment of B could be true for, at most, only one size of circle. However, solutions A and C are still compatible with scale invariance, corresponding to the choices $q = 1$ and $q = 2$ respectively.

4. Translational Invariance

We now investigate the consequences of the fact that a given straw S can intersect two circles C, C' of the same radius R, but with a relative displacement b. Referring to Fig. 1, the midpoint of the chord with respect to circle C is the point P, with coordinates (r, θ); while the same straw defines a midpoint of the chord with respect to C' at the point P' whose coordinates are (r', θ'). From Fig. 1 the coordinate transformation $(r, \theta) \to (r', \theta')$ is given by

$$r' = |r - b\cos\theta| \tag{8}$$

$$\theta' = \begin{cases} \theta, & r > b\cos\theta \\ \theta + \pi, & r < b\cos\theta \end{cases} \tag{9}$$

As P varies over the region Γ, P' varies over Γ', and vice versa; thus the straws define a 1:1 mapping of Γ onto Γ'.

Now we note the translational symmetry; since the statement of the problem gave no information about the location of the circle, the problems of C and C' appear exactly the same to two slightly displaced observers O and O'. Our desideratum of consistency then demands that they assign probability density C and C' respectively which have the same form (7) with the same value of q.

It is further necessary that these two observers assign equal probabilities to the regions Γ and Γ', respectively, since (a) they are probabilities of the same event, and (b) the probability that a

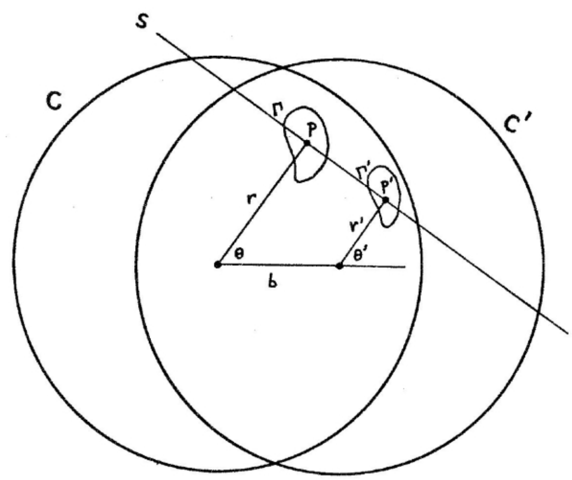

Fig. 1 A Straw S intersects two slightly displaced circles C and C'.

straw which intersects one circle will also intersect the other, thus setting up this correspondence, is also the same in the two problems. Let us see whether these two requirements are compatible.

The probability that a chord intersecting C will have its midpoint in Γ is

$$\int_\Gamma f(r)\, r\, dr\, d\theta = \frac{q}{2\pi R^q} \int_\Gamma r^{q-1} dr\, d\theta. \tag{10}$$

The probability that a chord intersecting C' will have its midpoint in Γ' is

$$\frac{q}{2\pi R^q} \int_{\Gamma'} (r')^{q-1}\, dr'\, d\theta' = \frac{q}{2\pi R^q} \int_\Gamma |r - b\cos\theta|^{q-1}\; dr\, d\theta \tag{11}$$

where we have transformed the integral back to the variables (r, θ) by use of (8) and (9), noting that the Jacobian is unity. Evidently, (10) and (11) will be equal for arbitrary Γ if and only if $q = 1$; and so our distribution $f(r)$ is now uniquely determined.

The proposed solution C in the introduction is thus eliminated for lack of translational invariance; a rain of straws which had the property assumed with respect to one circle, could not have the same property with respect to a slightly displaced one.

5. Final Results

We have found the invariance requirements determine the probability density

$$f(r, \theta) = \frac{1}{2\pi R r}, \quad 0 \le r \le R, \quad 0 \le \theta \le 2\pi \tag{12}$$

corresponding to solution A in the introduction. It is interesting that this has a singularity at the center, the need for which can be understood as follows. The condition that the midpoint (r, θ) falls within a small region Δ imposes restrictions on the possible directions of the chord. But as Δ moves inward, as soon as it includes the center of the circle all angles are suddenly allowed. Thus there is an infinitely rapid change in the "manifold of possibilities."

Further analysis (almost obvious from contemplation of Fig. 1) shows that the requirement of translational invariance is so stringent that it already determines the result (12) uniquely; thus the proposed solution B is incompatible with either scale or translational invariance, and in order to find (12), it was not really necessary to consider scale invariance. However, the solution (12) would in any event have to be tested for scale invariance, and if it failed to pass that test, we would conclude that the problem as stated has *no* solution; *i.e.,* although at first glance it appears underdetermined, it would have to be regarded, from the standpoint of transformation groups, as overdetermined. As luck would have it, these requirements *are* compatible; and so the problem has one unique solution.

The distribution of chord lengths follows at once from (12). A chord whose midpoint is at (r, θ) has a length $L = 2(R^2 - r^2)^{1/2}$. In terms of the reduced chord lengths, $x \equiv L/2R$, we obtain the universal distribution law

$$p(x)dx = \frac{x\,dx}{(1-x^2)^{1/2}}, \quad 0 \le x \le 1 \tag{13}$$

in agreement with Borel's conjecture (1909).

6. Frequency Correspondence

From the manner of its derivation, the distribution (13) would appear to have only a subjective meaning; while it describes the only possible state of knowledge corresponding to a unique solution in view of the many things left unspecified in the statement of Bertrand's problem, we have as yet given no reason to suppose that it has any relation to frequencies observed in the actual experiment. In general, of course, no such claim can be made; the mere fact that my state of knowledge gives me no reason to prefer one event over another is not enough to make them occur equally often! Indeed, it is clear that no "pure thought" argument, whether based on transformation groups or any other principle, can predict with certainty what must happen in a real experiment. And we can easily imagine a very precise machine which tosses straws in such a way as to produce any distribution of chord lengths we please on a given circle.

Nevertheless, we are entitled to claim a definite frequency correspondence for the result (13). For there is one "objective fact" which *has* been proved by the above derivation: Any rain of straws which does *not* produce a frequency distribution agreeing with (13) will necessarily produce different distributions on different circles.

But this is all we need in order to predict with confidence that the distribution (13) *will* be observed in any experiment where the "region of uncertainty" is large compared to the circle. For, if we lack the skill to toss straws so that, with certainty, they intersect a given circle, then surely we lack *a fortiori* the skill consistently to produce different distributions on different circles *within* this region of uncertainty!

It is for this reason that distributions predicted by the method of transformation groups turn out to have a frequency correspondence after all. Strictly speaking, this result holds only in the limiting case of "zero skill," but as a moment's thought will show, the skill required to produce any appreciable deviation from (13) is so great that in practice it would be difficult to achieve even with a machine.

Of course, the above arguments have demonstrated this frequency correspondence in only one case. In the following section we adduce arguments indicating that it is a general property of the transformation group method.

These conclusions seem to be in direct contradiction to those of von Mises (1957, 1964), who denied that such problems belong to the field of probability theory at all. It appears to us that if we were to adopt von Mises' philosophy of probability theory strictly and consistently, the range of legitimate physical applications of probability theory would be reduced almost to the vanishing point. Since we have made a definite, unequivocal prediction, this issue has now been removed from the realm of philosophy into that of verifiable fact. The predictive power of the transformation group method can be put to the test quite easily in this and other problems by performing the experiments.

The Bertrand experiment has, in fact, been performed by the writer and Dr. Charles E. Tyler, tossing broom straws from a standing position onto a 5-in.-diameter circle drawn on the floor. Grouping the range of chord lengths into ten categories, 128 successful tosses confirmed Eq. (13) with an embarrassingly low value of chi-squared. However, experimental results will no doubt be more convincing if reported by others.

7. Discussion

Bertrand's paradox has a greater importance than appears at first glance, because it is a simple crystallization of a deeper paradox which has permeated much of probability theory from its beginnings. In "real" physical applications when we try to formulate the problem of interest in probability terms we find almost always that a statement emerges which, like Bertrand's, appears too vague to determine any definite solution, because apparently essential things are left unspecified.

We elaborate the example noted in the introduction: Given a gas of N molecules in a volume V, with known intermolecular forces, total energy E, predict from this its molecular velocity distribution, pressure, distribution of pressure fluctuations, viscosity, thermal conductivity, and diffusion constant. Here again the viewpoint expressed by most writers on probability theory would lead one to conclude that the problem has no definite solution because it is ill posed; the things specified are grossly inadequate to determine any unique probability distribution over microstates. If we reject the principle of indifference, and insist that the only valid basis for assigning probabilities is frequency in some random experiment, it would again appear that the only way of determining these quantities is to perform the experiment.

It is, however, a matter of record that over a century ago, without benefit of any frequency data on positions and velocity of molecules, James Clark Maxwell was able to predict all these quantities correctly by a "pure thought" probability analysis which amounted to recognizing the "equally possible" cases. In the case of viscosity the predicted dependence on density appeared at first to contradict common sense, casting doubt on Maxwell's analysis. But when the experiments were performed they confirmed Maxwell's prediction, leading to the first great triumph of kinetic theory. These are solid, positive accomplishments; and they cannot be made to appear otherwise merely be deploring his use of the principle of indifference.

Likewise, we calculate the probability of obtaining various hands at poker; and we are so confident of the results that we are willing to risk money on bets which the calculations indicate are favorable to us. But underlying these calculations is the intuitive judgment that all distributions of cards are equally likely; and with a different judgment our calculations would give different results. Once again we are predicting definite, verifiable facts by "pure thought" arguments based ultimately on recognizing the "equally possible" cases; and yet present statistical doctrine, both orthodox and personalistic, denies that this is a valid basis for assigning probabilities!

The dilemma is thus apparent; on the one hand, one cannot deny the force of arguments which, by pointing to such things as Bertrand's paradox, demonstrate the ambiguities and dangers in the principle of indifference. But on the other hand, it is equally undeniable that use of this principle has, over and over again, led to correct, nontrivial, and useful predictions. Thus it appears that while we cannot wholly accept the principle of indifference, we cannot wholly reject it either; to do so would be to cast out some of the most important and successful applications of probability theory.

The transformation group method grew out of the writer's conviction, based on pondering this situation, that the principle of indifference has been unjustly maligned in the past; what it has needed was not blanket condemnation, but recognition of the proper way to apply it. We agree with most other writers on probability theory that it is dangerous to apply this principle at the level of indifference between *events,* because our intuition is a very unreliable guide in such matters, as Bertrand's paradox illustrates.

However, the principle of indifference may, in our view, be applied legitimately at the more abstract level of indifference between *problems;* because that is a matter that is definitely determined by the statement of a problem, independently of our intuition. Every circumstance left unspecified in the statement of a problem defines an invariance property which the solution must have if there is to be any definite solution at all. The transformation group, which expresses these invariances mathematically, imposes definite restrictions on the form of the solution, and in many cases fully determines it.

Of course, not all invariances are useful. For example, the statement of Bertrand's problem does not specify the time of day at which the straws are tossed, the color of the circle, the luminosity of Betelgeuse, or the number of oysters in Chesapeake Bay; from which we infer, correctly, that if the problem as stated is to have a unique solution, it must not depend on these circumstances. But this would not help us unless we had previously thought that these things might be germane.

Study of a number of cases makes it appear that the aforementioned dilemma can now be resolved as follows. We suggest that the cases in which the principle of indifference has been applied successfully in the past are just the ones in which the solution can be "reverbalized" so that the actual calculations used are seen as an application of indifference between problems, rather than events.

For example, in the case of poker hands the statement of the problem does not specify the order of cards in the deck before shuffling; therefore if the problem is to have any definite solution, it must not depend on this circumstance; *i.e.,* it must be invariant under the group of 52! permutations of cards, each of which transforms the problem into an equivalent one. Whether we verbalize the solution by asserting that all distributions of cards in the final hands are "equally likely," or by saying that the solution shall have this invariance property, we shall evidently do just the same calculation and obtain the same final result.

There remains, however, a difference in the logical situation. After having applied the transformation group argument in this way we are not entitled to assert that the predicted distribution of poker hands *must* be observed in practice. The only thing that can be proved by transformation groups is that if this distribution is *not* forthcoming then the probability of obtaining a given hand will necessarily be different for different initial orders of cards; or, as we would state it colloquially, the cards are not being "properly" shuffled. This is, of course, just the conclusion we do draw in practice, whatever our philosophy about the "meaning of probability."

Once again it is clear that the invariant solution is overwhelmingly the most likely one to be produced by a person of ordinary skill; to shuffle cards in such a way that one particular aspect of the initial order is retained consistently in the final order requires a "microscopic" degree of control over the exact details of shuffling (in this case, however, the possession of such skill is generally regarded as dishonest, rather than impossible).

We have not found any general proof that the method of transformation groups will always lead to solutions which this frequency correspondence property; however, analysis of some dozen problems like the above has failed to produce any counterexample, and its general validity is rendered plausible as follows.

In the first place, we recognize that every circumstance which our common sense tells us may exert some influence on the result of an experiment ought to be given explicitly in the statement of a problem. If we fail to do that, then of course we have no right to expect agreement between prediction and observation; but this is not a failure of probability theory, but rather a failure on our part to state the full problem. If the statement of a problem *does* properly include all such information, then it would appear that any circumstances that are still left unspecified must correspond to some lack of control over the conditions of the experiment, which makes it impossible for us to state them. But invariance under the corresponding transformation group is just the formal expression of this lack of control, or lack of skill.

One has the feeling that this situation can be formalized more completely; perhaps one can define some "space" corresponding to all possible degrees of skill and define a measure in this space, which proves to be concentrated overwhelmingly on those regions leading to the invariant solution. Up to the present, however, we have not seen how to carry out such a program; perhaps others will.

8. Conjectures

There remains the interesting, and still unanswered, question of how to define precisely the class of problems which can be solved by the method illustrated here. There are many problems in which we do not see how to apply it unambiguously; von Mises' water-and-wine problem is a good example. Here we are told that a mixture of water and wine contains at least half wine, and are asked: What is the probability that it contains at least three-quarters wine? On the usual viewpoint this problem is underdetermined; nothing tells us which quantity should be regarded as uniformly distributed. However, from the standpoint of the invariance group, it may be more useful to regard such problems as *overdetermined;* so many things are left unspecified that the invariance group is too large, and no solution can conform to it.

It thus appears that the "higher-level problem" of how to formulate statistical problems in such a way that they are neither underdetermined nor overdetermined may itself be capable of mathematical analysis. In the writer's opinion it is one of the major weaknesses of present statistical practice that we do not seem to know how to formulate statistical problems in this way, or even how to judge whether a given problem is well posed. Again, the Bertrand paradox is a good illustration of this difficulty, for it was long thought that not enough was specified to determine any unique solution, but from the viewpoint which recognizes the full invariance group implied by the above statement of the problem, it now appears that it was well posed after all.

In many cases, evidently, the difficulty has been simply that we have not been reading out all that is implied by the statement of a problem; the things left unspecified must be taken into account just as carefully as the ones that are specified. Presumably, a person would not seriously propose a problem unless he supposed that it had a definite solution. Therefore, as a matter of courtesy and in keeping with a worthy principle of law, we might take the view that *a problem shall be presumed to have a definite solution until the contrary has been proved.* If we accept this as a reasonable attitude, then we must recognize that we are not in a position to judge whether a problem is well posed until we have carried out a transformation group analysis of all the invariances implied by its statement.

The question whether a problem is well posed is thus more subtle in probability theory than in other branches of mathematics, and any results which could be obtained by study of the "higher-level problem" might be of immediate use in applied statistics.

Appendix: Comments on Bertrand's Problem

Bertrand (1889, pp. 4-5); "Aucune de trois n'est fausse, aucune n'est exacte, la question est mal posée."

Borel (1909, pp. 110-113): "... il est aisé de voir que la plupart des procédés naturels que l'on peut imaginer conduiser à la première."

Poincaré (1912, pp. 118-130): "... nous avons definie la probabilité de deux manières différentes."

Uspensky (1937, p. 251): "... we are really dealing with two different problems."

Northrup (1944, pp. 181-183): "One guess is as good as another."

Gnedenko (1962, pp. 40-41): The three results "would be appropriate" in three different experiments.

Kendall and Moran (1963, p. 10): "All three solutions are correct, but they really refer to different problems."

Weaver (1963, pp. 356-357): "... you have to watch your step."

Von Mises (1964, pp. 160-166): "Which one of these or many other assumptions should be made is a question of fact and depends on how the needles are thrown. It is not a problem of probability calculus to decide which distribution prevails ..." Von Mises, in the preface to (1957), also charges that, "Neither Laplace nor any of his followers, including Poincaré, ever reveals how, starting with *a priori* premises concerning equally possible cases, the sudden transition to the description of real statistical events is to be made." It appears to us that this had already been accomplished in large part by James Bernoulli (1703) in his demonstration of the weak law of large numbers, the first theorem establishing a connection between probability and frequency, Jaynes (1968), and the present article may be regarded as further contributions toward answering von Mises' objections.

Mosteller (1965, p. 40): "Until the expression 'at random' is made more specific, the question does not have a definite answer .... We cannot guarantee that any of these results would agree with those obtained from some physical process ...."

References

Bertrand, J. (1889), *Calcul des probabilités,* Gauthier-Villars, Paris, pp. 4-5.

Borel, E. (1909), "*Éléments de la théorie des probabilités,* Hermann et Fils, Paris, pp. 110-113.

Gnedenko, B. V. (1962), *The Theory of Probability,* Chelsea Publ. Co., New York, pp. 40-41.

Jaynes, E. T. (1968), "Prior Probabilities," *IEEE Trans. Systems Sci. Cybernetics,* **SSC-4,** (3), pp. 227-241.

Kendall, M. G. and Moran, P. A. P. (1963), *Geometrical Probability,* Hafner Publ. Co., New York, p. 10.

von Mises, R. (1957), *Probability, Statistics and Truth,* Macmillan, New York.

von Mises, R. (1964), in *Mathematical Theory of Probability and Statistics,* H. Geiringer *ed.,* Academic Press, New York, pp. 160-166.

Mosteller, F. (1965), *Fifty Challenging Problems in Probability,* Addison-Wesley, Reading, Massachusetts, p. 40.

Northrup, E. P. (1944), *Riddles in mathematics,* van Nostrand, New York, pp.181-183.

Poincaré H. (1912), *Calcul des probabilitiés,* Paris, pp. 118-130.

Uspensky, J. V. (1937), *Introduction to Mathematical Probability,* McGraw-Hill, New York, p. 251.

Weaver, W. (1963), *Lady Luck: the Theory of Probability,* Doubleday-Anchor, Garden City, New York, pp. 356-357.

Weyl, H. (1946), *The Classical Groups,* Princeton University Press, Princeton, New Jersey.

Wigner, E. P. (1931), *Gruppentheorie und ihre Anwendung auf die Quantenmechanik der Atomspektren,* Fr. Vieweg, Braunschweig.