Lecture 19

PHYSICS OF "RANDOM" EXPERIMENTS

As we have already noted several times in these lectures, the idea that probability assignments must be based ultimately on observed frequencies in random experiments is fundamental to almost all recent expositions of probability theory; which would seem to make it a branch of experimental science. At the end of Lecture 9 we saw some of the difficulties that this view leads us to, in that in some real physical experiments the distinction between random and nonrandom quantities is so obscure and artificial that you have to resort to black magic in order to force this distinction into the problem. But in that discussion we didn't really get into the serious physics of the situation. In this lecture, I want to take time off from development of probability theory, and have a little interlude of more physical considerations that show the fundamental difficulty with the notion of "random" experiments--even the ones, such as coin tossing, which at first glance seem most appropriately regarded as "random."

We have also noted that there have always been dissenters from the orthodox view who have maintained, with Laplace, that probability theory is properly regarded as the "calculas of inductive reasoning," and is not fundamentally related to random experiments at all. According to this second view, consideration of random experiments is only one particular <u>application</u> of probability theory (and not even the most important one); for probability theory accounts equally well for general inductive inferences where no random experiment is involved. But we haven't yet noted that there is an interesting correlation; those who have advocated the second view have tended to be physicists rather than mathematicians. So, it will be of interest to examine the hsstorical background of this question with particular emphasis on the physics of the situation.

With the rise of the "Neo-Bayesian" school of thought, this question has flared up again in the recent literature of statistics. Several participants have recognized that the issue is not merely one of philosophy or mathematics; in some way not yet made entirely clear, it also involves physics. The mathematician tends to think of a random experiment as an abstraction--really nothing more than a sequence of numbers. To define the "nature" of of the random experiment he introduces statements--variously termed assumptions, postulates, or axioms--which specify the sample space and assert the existence, and certain other properties, of limiting frequencies. In real life, however, a random experiment is not an abstraction whose properties can be defined at will; it is surely subject to the laws of physics.

As soon as a specific random experiment is described, it is the nature of a physicist to start thinking, not about the abstract sample space thus defined, but about the physical mechanism of the phenomenon being observed. The question whether the usual postulates of probability theory are compatible with the known laws of physics is capable of logical analysis, with results that have a direct bearing on the question, not of the mathematical validity of frequency and non-frequency theories of probability, but of their <u>applicability</u> to real situations. Any such conclusions have, evidently, a relevance to the question of orthodox vs. Bayesian statistical methods.

In a recent discussion of these questions Professor G. E. P. Box (196) has remarked, "I believe, for instance, that it would be very difficult to persuade an intelligent physicist that current statistical practice was sensible, but that there would be much less difficulty with an approach via likelihood and Bayes' theorem." Let's analyze this statement in the light both of history and of physics.

19.1. Historical Background.

As we know, probability theory started in consideration of gambling devices by Cardano, Pascal, and Fermat; but its development beyond that level, in the 18'th and 19'th centuries, was stimulated by applications in physics and astronomy, and was the work of people--Jacob and Daniel Bernoulli, Laplace, Poisson, Legendre, Gauss--most of whom we would describe today as mathematical physicists.

In the nineteenth century a knowledge of statistical analysis, consisting largely of the work of Laplace, Legendre, and Gauss, was considered an essential part of the training of a scientist. For example, as a young man J. Willard Gibbs spent three years (1866-69) in post-doctoral study at the Universities of Paris, Berlin, and Heidelberg; and the most prominent topic mentioned in the list of lectures he attended was statistical analysis. This study undoubtedly contributed to his discovery, 33 years later, of the basic "canonical ensemble" formalism of statistical mechanics.

A radical change took place early in this century when a new group of workers, not physicists, entered the field. They proceeded to reject virtually everything done by Laplace, and sought to develop statistics anew based on entirely different principles. This extremely aggressive school soon dominated the field so completely that its methods have come to be known as "orthodox" statistics.

Simultaneously with this development, the physicists--with Sir Harold Jeffreys as almost the sole exception--quietly retired from the field, and statistical analysis disappeared from the physics curriculum. This disappearance has been so complete that, if today someone were to take a poll of physicists, I think he would find that not one in a hundred could identify such names as Fisher, Neyman, Wald; or such terms as maximum likelihood, confidence interval, analysis of variance.

This course of events--the leading role of physicists in development of the original Bayesian methods, and their later withdrawal from orthodox statistics--was no accident. As further evidence that there is some kind of basic conflict between orthodox statistical doctrine and physics, we may note that two of the most eloquent proponents of non-frequency definitions in this century--Poincaré and Jeffreys--have been mathematical physicists of the very highest competence, as was Laplace. Professor Box's statement thus has a clear basis in historical fact.

But what is the nature of this conflict? What is there in the physicist's knowledge that has led him to reject the very thing that the orthodox statistician regards as conferring "objectivity" on his methods? To see where the difficulty lies, we examine a few simple random experiments from the physicist's viewpoint. The facts I want to point out are so elementary that you can't believe they are really unknown to modern writers on probability theory. The continual appearance of new statistical textbooks which ignore them merely illustrates what we physics teachers have always known; you can teach a student the laws of physics, but you cannot teach him the art of recognizing the <u>relevance</u> of this knowledge, much less the habit of <u>applying</u> it, in his everyday problems.

<u>19.2. How to Cheat at Coin and Die Tossing.</u>

Cramér (1946) takes it as an axiom that "Any random variable has a unique probability distribution." From the later context, it is clear that what he really means is that it has a unique <u>frequency</u> distribution. If one assumes that the number obtained by tossing a die is a random variable, this leads to the conclusion that the frequency with which a certain face comes up is a physical property of the die; just as much so as its mass, moment of inertia, or chemical composition. Thus, Cramér (*loc. cit.*, p. 154) states, "The numbers p_r should, in fact, be regarded as physical constants of the particular die that we are using, and the question as to their numerical values cannot be answered by the axioms of probability theory, any more than the size and the weight of the die are determined by the geometrical and mechanical axioms. However, experience shows that in a well-made die the frequency of any event r in a long series of throws usually approaches 1/6, and accordingly we shall often assume that all the p_r are equal to 1/6 . . . ."

To a physicist, such an attitude seems to show utter contempt for the known laws of mechanics. The results of tossing a die many times do *not* tell us any definite number characteristic of the die. Thyy tell us something about the way the die was tossed. If you toss "loaded" dice in different ways, you can easily alter the relative frequencies of the faces. With more difficulty, and over a smaller range, you can even do this if the die is perfectly "honest."

Although the principles will be just the same, it will be simpler to discuss a random experiment with only two possible outcomes per trial. Consider, therefore, a "biased" coin, about which I. J. Good has remarked (Savage, 1962): "Most of us probably think about a biased coin as if it had a physical probability. Now whether it is defined in terms of frequency or just falls out of another type of theory, I think we do argue that way. I suspect that even the most extreme subjectivist such as de Finetti would have to agree that he did sometimes think that way, though he would perhaps avoid doing it in print." It is, of course, just the famous theorem of de Finetti that we studied in Lecture 17, which shows us how to carry out a probability analysis of the biased coin without thinking in the manner suggested (it does not follow, however, that this analysis is applicable to a real biased coin). In any event, it is quite easy to show how a physicist would analyze the problem. Let us suppose that the center of gravity of this coin lies on its axis, but displaced a distance x from its geometrical center. If we agree that the result of tossing this coin is a "random variable," then according to the axiom stated by Cramér and hinted at by Good, there must exist a definite functional relationship between the frequency of heads and x:

$$p_H = f(x)$$

But this assertion goes far beyond the mathematician's traditional range of freedom to invent arbitrary axioms, and encroaches on the domain of physics; for the laws of mechanics are quite competent to tell us whether such a functional relationship does or does not exist.

The easiest game to analyze turns out to be just the one most often played to decide such practical matters as the starting side in a football game. Your opponent first calls "heads" or "tails" at will. You then toss the coin into the air, catch it in your hand, and without looking at it, show it first to your opponent, who wins if he has called correctly. It is further agreed that a "fair" toss is one in which the coin rises at least nine feet into the air, and thus spends at least 1.5 seconds in free flight.

The laws of mechanics now tell us the following. The ellipsoid of inertia of a thin disc is an oblate spheroid of eccentricity $1/\sqrt{2}$. The displacement x does not affect the symmetry of this ellipsoid, and so according to the Poinsot construction, as found in textbooks on rigid dynamics [such as Routh (19)], the polhodes remain circles concentric with the axis of the coin. In consequence, the character of the tumbling motion of the coin while in flight is exactly the same for a biased as an unbiased coin, except that for the biased one it is the center of gravity, rather than the geometrical center, which describes the parabolic "free particle" trajectory.

An important feature of this tumbling motion is conservation of angular momentum; during its flight the angular momentum of the coin maintains a fixed direction in space (but the angular velocity does not; and so the tumbling may appear chaotic to the eye). Let us denote this direction by the unit vector n; it can be any direction you choose, and it is determined by the particular kind of twist you give the coin at the instant of launching. Whether the coin is biased or not, it will show the same face throughout the motion if viewed from this direction (unless, of course, n is exactly perpendicular to the axis of the coin, in which case it shows no face at all).

Therefore, in order to know which face will be uppermost in your hand, you have only to carry out the following procedure. Denote by k a unit vector passing through the coin along its axis, with its point on the "heads" side. Now toss the coin with a twist so that k and n make an acute angle, then catch it with your palm held flat, in a plane normal to n. On successive tosses, you can let the direction of n, the magnitude of the angular momentum, and the angle between n and k, vary widely; the tumbling motion will then appear entirely different to the eye on different tosses, and it would require almost superhuman powers of observation to discover your strategy.

Thus, anyone familiar with the law of conservation of angular momentum can, after some practice, cheat at the usual coin-toss game and call his shots with 100 per cent accuracy. You can obtain any frequency of heads you want; and the bias of the coin has no influence at all on the results!

Of course, as soon as this result is out, someone will object that the experiment analyzed is too "simple." In other words, those who have postulated a "physical" probability for the biased coin have, without stating so, really had in mind a more complicated experiment in which some kind of "randomness" has more opportunity to make itself felt.

While accepting this criticism, I can't suppress the obvious comment: scanning the literature of probability theory, isn't it curious that so many mathematicians, usually far more careful than physicists to list all the qualifications needed to make a statement correct, should have failed to see the need for any qualifications here? However, to be more constructive, we can just as well analyze a more complicated experiment.

Suppose that now, instead of catching the coin in our hand, we toss it onto a table, and let it spin and bounce in various ways until it comes to rest. Is this experiment sufficiently "random" so that the true "physical probability" will manifest itself? No doubt, the answer will be that it is _not_ sufficiently random if the coin is merely tossed up two inches starting at the table level, but it will become a "fair" experiment if we toss it up higher.

Exactly how high, then, must we toss it before the true "physical probability" can be measured? This is not an easy question to answer, and I certainly won't make any attempt to answer it here. It would appear, however, that anyone who asserts the existence of a "physical" probability for the coin ought to be prepared to answer it; otherwise it is hard to see what content the assertion has (in the sense of operational verifiability).

I don't deny that the bias of the coin will now have some influence on the frequency of heads; I claim only that the _amount_ of that influence depends very much on how you toss the coin so that, again in this experiment, there is no definite number $p_H = f(x)$ describing a physical property of the coin. Indeed, even the _direction_ of this influence can be reversed by different methods of tossing, as follows.

However high we toss the coin, we still have the law of conservation of angular momentum; and so we can toss it by _Method A_: to ensure that heads will be uppermost when the coin first strikes the table, we have only to hold it heads up, and toss it so that the total angular momentum is directed vertically. Again, we can vary the magnitude of the angular momentum, and the angle between n and k, so that the motion appears quite different to the eye on different tosses, and it would require very close observation to notice that heads remains uppermost throughout the free flight. Although what happens after the coin strikes the table is complicated, the fact that heads is uppermost at first has a strong influence on the result, which is more pronounced for large angular momentum.

Many people have developed the knack of tossing a coin by Method B: it goes through a phase of standing on edge and spinning rapidly about a vertical axis, before finally falling to one side or the other. If you toss the coin this way, the eccentric position of the center of gravity will have a dominating influence, and render it practically certain that it will fall always showing the same face. Ordinarily, one would suppose that the coin prefers to fall in the position which gives it the lowest center of gravity; i.e., if the center of gravity is displaced toward tails, then the coin should have a tendency to show heads. However, for an interesting mechanical reason, which I leave for you to work out, method B produces the opposite influence, the coin strongly preferring to fall so that its center of gravity is high.

On the other hand, the bias of the coin has a rather small influence in the opposite direction if we toss it by Method C: the coin rotates about a horizontal axis which is perpendicular to the axis of the coin, and so bounces until it can no longer turn over.

In this experiment also, therefore, a person familiar with the laws of mechanics can toss a biased coin so that it will produce predominantly either heads or tails, at will. Furthermore, the effect of method A persists whether the coin is biased or not; and so one can even do this with a perfectly "honest" coin. Finally, although we have been considering only coins, essentially the same mechanical considerations apply to the tossing of any other object, such as a die.

From the fact that we have seen a strong preponderance of heads, we cannot legitimately conclude that the coin is biased; it may be biased, or it may have been tossed in a way that systematically favors heads. Likewise, from the fact that we have seen equal numbers of heads and tails, we cannot legitimately conclude that the coin is "honest." It may be honest, or it may have been tossed in a way that nullifies the effect of its bias.

19.3. Experimental Evidence.

Since the conclusions just stated are in direct contradiction to what is postulated, almost universally, in expositions of probability theory, it is worth noting that anyone can easily verify them for himself, in a few minutes of experimentation in his kitchen. An excellent "biased coin" is provided by the metal lid of a small pickle jar, of the type which is not knurled on the outside, and has the edge rolled inward rather than outward, so that the outside surface is accurately round and smooth, and so symmetrical that on an edge view one cannot tell which is the top side.

Suspecting that many people simply would not believe the things just claimed without experimental proof, I have performed these experiments with a jar lid of diameter $d = 2\ 5/8"$, height $h = 3/8"$. Assuming a uniform thickness for the metal, the center of gravity should be displaced from the geometrical center by a distance $x = dh/(2d+8h) = 0.120$ inches; and this was confirmed by hanging the lid by its edge and measuring the angle at which it comes to rest. Ordinarily, one expects this bias to make the lid prefer to fall bottom side up; and so this side will be called "heads." The lid was tossed up about 6 feet, and fell onto a smooth linoleum floor. I allowed myself ten practice tosses by each of the three methods described, and then recorded the results of a number of tosses by: method A deliberately favoring heads, method A deliberately favoring tails, method B, and method C, as given in Table 19.1.

Method	No. of Tosses	No. of Heads
A(H)	100	99
A(T)	50	0
B	100	0
C	100	54

Table 19.1. Results of tossing a "biased coin" in four different ways.

In method A the mode of tossing completely dominated the result (the effect of bias would, presumably, have been much greater is the "coin" were tossed onto a surface with a greater coefficient of friction). In method B, the bias completely dominated the result (in about thirty of these tosses it looked for a while as if the result were going to be heads, as one might naively expect; but each time the "coin" eventually righted itself and turned over, as predicted by the laws of rigid dynamics). In method C, there was no significant evidence for any effect of bias.

One can, of course, always claim that tossing the coin in any of the four specific ways described is "cheating," and that there exists a "fair" way of tossing it, such that the "true" probabilities will emerge from the experiment. But again, the person who asserts this ought to be prepared to define precisely what this fair method is, otherwise the assertion is without content. Presumably, a fair method of tossing ought to be some kind of random mixture of methods A(H), A(T), B, C, and others; but what is a "fair" relative weighting to give them? It is difficult to see how one could define a "fair" method of tossing except by the condition that it should result in a certain frequency of heads; and so we are involved in a circular argument.

This analysis can be carried much further than we have done here, and I want to go into it some more in a minute; but it is perhaps sufficiently clear already that analysis of coin and die tossing is not a problem of abstract statistics, in which one is free to introduce postulates about "physical" probabilities which ignore the laws of physics. It is a problem of mechanics, highly complicated and irrelevant to probability theory except insofar as it forces us to think a little more carefully about how probability theory must be formulated if it is to be applicable to real situations. Performing a random experiment with a coin does not tell us what the "physical" probability of heads is; it may tell us something about the bias, but it also tells us something about how the coin is being tossed. Indeed, unless we know how it is being tossed, we cannot draw any inferences about its bias from the experiment.

It may not, however, be clear from the above that conclusions of this type hold quite generally for random experiments, and in no way depend on the particular mechanical properties of coins and dies. In order to illustrate this, let's consider an entirely different kind of random experiment.

19.4. Bridge Hands.

In Lectures 5 and 13, we have already quoted Professor Wm. Feller's pronouncements on the use of Bayes' theorem in quality control testing, about Laplace's rule of succession, and about Daniel Bernoulli's conception of the utility function for decision theory. He does not fail us here either; in this interesting textbook (Feller, 1950), he writes: "The number of possible distributions of cards in bridge is almost 10^{30}. Usually, we agree to consider them as equally probable. For a check of this convention more than 10^{30}

experiments would be required ...." Here again, we have the view that bridge hands possess "physical" probabilities, that the uniform probability assignment is a "convention," and that the ultimate criterion for its correctness must be observed frequencies in a random experiment.

The thing which is wrong here is that none of us would be willing to use this criterion in a real-life situation because, if we know that the deck is an "honest" one, our common sense tells us something which carries more weight than 10^{30} random experiments do. We would, in fact, be willing to accept the result of the random experiment <u>only if it agreed with our preconceived notion that all distributions are equally likely</u>.

To many of you this last statement may seem like pure blasphemy--it stands in violent contradiction to what we have all been taught. Yet in order to see why it is true, we have only to imagine that those 10^{30} experiments <u>had</u> been performed, and the uniform distribution was not forthcoming. We expect, if all distributions of cards have equal frequencies, that any combination of two specified cards will appear together in a given hand, on the average, once in $52 \cdot 51/13 \cdot 12 = 17$ deals. But suppose that the particular combination (Jack of hearts--Seven of clubs) appeared together in each hand three times as often as this. <u>Would we then accept it as an established fact that this particular combination is inherently more likely than others?</u>

We would not. We would say that the cards had not been properly shuffled. But once again we are involved in a circular argument; because there is no way to <u>define</u> a "proper" method of shuffling except by the condition that it should produce all distributions with equal frequency!

In carrying out a probability analysis of bridge hands, are we really concerned with physical probabilities; or with inductive reasoning? In order to help answer this, consider the following scenario: I tell an orthodox statistician that I have dealt at bridge 1000 times, shuffling "fairly" each time; and that in every case the seven of clubs was in my own hand. What will his reaction be? He will, I think, mentally visualize the number

$$\left(\frac{1}{4}\right)^{1000} \simeq 10^{-602}$$

and conclude instantly that I have not told the truth; and no amount of persuasion on my part will shake that judgment. But what accounts for the strength of his belief? Obviously, it cannot be justified if our assignment of equal probabilities to all distributions of cards is merely a "convention," subject to change in the light of experimental evidence. Even more obviously, he is not making use of any knowledge about the outcome of an experiment involving 10^{30} bridge hands.

What is the extra evidence he has, which his common sense tells him carries more weight than any number of random experiments; but whose help he refuses to acknowledge in expounding probability theory? In order to maintain the claim that probability theory is an experimental science, based fundamentally not on inductive inference but on frequency in a random experiment, it is necessary to suppress some of the information which is available. This suppressed information, however, is just what enables inductive reasoning to approach the certainty of deductive reasoning in this example.

The suppressed evidence is, of course, simply our recognition of the symmetry of the situation. The only difference between a seven and an eight is that there is a different number printed on the face of the card. Our common sense tells us that where a card goes in shuffling depends only on the mechanical forces that are applied to it; and not on which number is printed on its face. If we observe any systematic tendency for one card to appear in the dealer's hand, which persists on indefinite repetitions of the experiment, we can infer from this only that there is some systematic tendency in the procedure of shuffling, which alone determines the outcome of the experiment.

Once again, therefore, performing the experiment tells you nothing about the "physical" probabilities of different hands. It tells you something about how the cards are being shuffled.

19.5. General Random Experiments.

In the face of the foregoing arguments, one can still take the following position (as a member of the audience did after one of my recent lectures): "You have shown only that coins, dies, and cards represent exceptional cases, where mechanical considerations obviate the usual probability postulates; i.e., they are not really 'random experiments.' But that is of no importance because these devices are used only for illustrative purposes; in the more dignified random experiments which merit the serious attention of the scientist or engineer, there is a physical probability."

To answer this, note that any specific experiment for which the existence of a physical probability is asserted, is subject to physical analysis like the ones just given, which will lead eventually to an understanding of its mechanism. But as soon as this understanding is reached, then this new experiment will also appear as an exceptional case where physical considerations obviate the usual postulates of "physical" probabilities. For, as soon as we have understood the mechanism of any experiment E, then there is logically no room for any postulate that various outcomes possess physical probabilities; for the question: "What are the probabilities of various outcomes O_1, O_2,...?" then reduces immediately to the question: "What are the probabilities of the corresponding initial conditions I_1, I_2, ... that lead to these outcomes?"

We might suppose that the possible initial conditions of experiment E themselves possess physical probabilities. But then we are considering an antecedent random experiment E', which produces conditions I_k as its possible outcomes. We can analyze the physical mechanism of E' and as soon as this is understood, the question will revert to: "What are the probabilities of the various initial conditions I_k' for experiment E'?" Evidently, we are involved in an infinite regress {E, E', E", ...}; the attempt to introduce a physical probability will be frustrated at every level where our knowledge of physical law permits us to analyze the mechanism involved. The notion of "physical probability" must retreat continually from one level to the next, as knowledge advances.

We are, therefore, in a situation very much like the "warfare between science and theology" of earlier times. For several centuries, theologians insisted on making factual assertions which encroached on the domains of astronomy, physics, biology, and geology--and which they were later forced to retract one by one in the face of advancing knowledge.

Clearly, probability theory ought to be formulated in a way that avoids factual assertions properly belonging to other fields, and which will later need to be retracted (as is now the case for many assertions in the literature concerning coins, dies, and cards). It appears to me that the only formulation which accomplishes this is the original one given by Laplace and expounded by Poincaré and Jeffreys, in which probability theory is regarded as the general "calculus of inductive reasoning," whose validity does not depend on any assumptions about properties of physical experiments. As we saw back in Lecture 3, a very important contribution to the logical foundations of this approach was made recently by R. T. Cox (1946), (1961), who showed that, if we represent degrees of plausibility by real numbers, then the mathematical rules for inductive inference are restricied by elementary conditions of consistency, stated in the form of functional equations whose general solutions are readily found. As already noted, it is no accident that all the aforementioned gentlemen are to be classed as physicists, to whom the things I have pointed out in this lecture would be obvious from the start.

The Laplace-Poincaré-Jeffreys-Cox formulation of probability theory does not require us to take one reluctant step after another down that infinite regress; it recognizes that anything which continually recedes from the light of detailed analysis can exist only in our imagination. Performing any of the so-called random experiments will not tell us what the "physical" probabilities are, because there is no such thing as a "physical" probability. The experiment tells us, in a very crude and incomplete way, something about how the initial conditions are varying from one repetition to another.

A much more efficient way of obtaining this information would be to study the initial conditions directly. However, in many cases this is beyond our present abilities; as in determining the safety and effectiveness of a new medicine. Here the only fully satisfactory approach would be to analyze the detailed sequence of chemical reactions that follow the taking of this medicine, in persons of every conceivable state of health. Having this analysis one could then predict, for each individual patient, exactly what the effect of the medicine will be.

Such an analysis being entirely out of the question at present, the only feasible way of obtaining the information we want is to perform a "random" experiment. No two patients are in exactly the same state of health; and for a given dose, the unknown variations in this factor constitute the variable initial conditions of the experiment, while the sample space comprises the set of distinguishable reactions to the medicine.

Our use of probability theory in this case is an example of inductive reasoning which amounts to the following: "If the initial conditions of the experiment continue in the future to vary over the same unknown range as they have in the past, then I expect that the relative frequencies of various outcomes will, in the future, approximate those which I have observed in the past. In the absence of positive evidence giving a reason why there should be some change in the future, and indicating in which direction this change should go, I can only suppose that things will continue in more or less the same way. As I observe the relative frequencies to remain stable over longer and longer times, I become more and more confident about this conclusion. But still, I am doing only inductive reasoning--there is no deductive proof that frequencies in the future will not be entirely different than those in the past.