Journal of Criminal Law and Criminology

Volume 57 | Issue 1 Article 13

1966

# Probability and Legal Proceedings

Charles R. Kingston

THE JOURNAL OF CRIMINAL LAW, CRIMINOLOGY AND POLICE SCIENCE
Vol. 57, No. 1


# PROBABILITY AND LEGAL PROCEEDINGS

CHARLES R. KINGSTON

Charles R. Kingston, D.Crim., is Senior Research Criminalist, New York State Identification and Intelligence System, Albany, New York. The author received his doctorate in criminology from the University of California, Berkeley, and served as an assistant to Prof. Kirk during his graduate studies. His previous contribution to this Journal, "The Use of Statistics in Criminalistics" appeared in Vol. 55.—EDITOR.

One of the controversial aspects of applied probability is its explicit use in legal proceedings in order to assess the significance of various types of evidence. The ubiquitousness of probability in legal proceedings has long been recognized, but little has been done to clarify the precise role and usage of mathematical probability in our courts.[1] One of Rabelais' characters, Judge Bridlegoose, solved the problems quite simply by allowing each of the parties to a civil suit to roll a pair of dice after the evidence had been collected and allowed to ripen for a suitable period of time.

> "And when you have done that," said Bluster [to Bridlegoose], "how do you set about handing down a decision?"
>
> "Just like you other gentlemen," replied Bridlegoose, "I give the decision to the one who wins the throw, in accordance with the judicial, tribunary, praetorian, first-come-first-served dice."[2]

Although this may not be the ideal way to apply the concepts of probability in court, it must be realized that probability, and consequently probability theory *does* have a place in court. It is difficult to imagine that any reasonable assessment of circumstantial evidence can be made without the elementary principles of probability theory being used, whether this use is made explicit or not. Generally the basic probabilistic ideas remain unexpressed, and serve only as mental pathways guiding the jurors' thought processes to their conclusions. But every so often these ideas are made explicit and are cast into mathematical terms.

When this occurs, it is essential that the manipulation and interpretation of these mathematical terms be correct. Unfortunately, in many reported instances where probability figures have been expressed in criminal trials, or discussed in connection with such proceedings, gross errors have been made. Some of these errors are exemplified in a newspaper report of a recent criminal case which was afforded considerable attention in the national press under such titles as, "Probability Law in Court," and "Law of Probability Foils 2 Robbers In Tough Case." The remainder of this article will be devoted to a discussion of some of the errors in the application of probability theory that are exemplified in the newspaper report.

Part of the newspaper item which reported some of the proceedings of the trial read as follows:[3]

> The questions and answers went something like this:
>
> Q. What is the probability of a man having a mustache? (1)
>
> A. One in three.
>
> Q. What is the probability of a young woman having blonde hair? (2)
>
> A. One in four.
>
> Q. What is the probability of seeing a yellow car on a Los Angeles county—or even California—street? (3)
>
> A. One in ten.
>
> Q. What is the probability of a man having a beard? (4)
>
> A. One in ten.
>
> Q. What is the probability of a Caucasian woman being seen with a Negro man? (5)
>
> A. One in a thousand.
>
> Q.? Then—what probability is there that you will find, at a given time and spot, a Negro male with a beard and a mustache driving a yellow car in which a blonde woman with a pony tail is riding? (6)
>
> A. Multiplying the probabilities of the

[1] For a general discussion of this, see: BALL, V. C., "The Moment of Truth: Probability Theory and Standards of Proof", VANDERBILT LAW REVIEW *14* (1961) 807–830.

[2] PUTNAM, SAMUAL (transl. & ed.), THE PORTABLE RABELAIS, The Viking Press, New York, 1946, p. 495.

[3] SAN FRANCISCO SUNDAY CHRONICLE, December 13, 1964, p. 13.

six factors together, the chance would be one in 12 million.

That, plus other circumstantial evidence, was enough for the jurors. They returned a guilty verdict. . . ."

Aside from the omission of the sixth item—the probability that a woman has a pony tail, which must be one in ten to arrive at the figure of one in 12 million—it is difficult to know how close the above report is to the actual testimony during the trial. However, since the type of report quoted represents the kind of information that the public is given, the following discussion will be presented on the basis of the quoted questions and answers.

## Probability

Since the questions all begin with, "What is the probability of . . .," or, ". . . what probability is there . . .," it seems reasonable to ask just what this means. Although several definitions of probability might be given[4], the most applicable to the situation in this case would seem to be that which is first presented in most elementary textbooks on probability, namely, the frequency concept. If we are considering a population of n objects, m of which have some property, say A, then $p = m/n$ is the probability of selecting an object with property A if the selection is a random one from the $n$ objects.[5] (A random selection is defined as one where each of the n objects has an equal chance of being selected.)

The value of p is quite often estimated from a sample of the n objects. If the random sample size is s, and if there are $r$ objects with property A in the sample, then r/s is a reasonable estimate of p. When such an estimate is made, it must be realized that we are dealing with a variable quantity that has certain statistical properties of its own, although these will not be considered in this article.

It is clear that in order to express a probability as defined above, it is necessary to relate it to some particular population. Although the general population being considered for the probability estimates in the example is not made explicit, it is presumably the entire adult or near adult population of persons either in California or in Los Angeles county. At times only a subpopulation may be considered, such as the entire *adult, male* population of an area. This is equivalent in many respects to the use of conditional probability, and should be handled accordingly.

In the third question, the population being considered is not at all clear. Does it consist of people who "see" cars, or of automobiles, or perhaps of people who own or drive cars. Similar problems about the population arise in connection with the fifth question. In order to relate the final calculations to the same general population, the probabilities involved in the third and fifth questions will be considered to apply to the general population in the third question (i.e., the probability that a randomly selected person owns or drives a yellow car), and to the Negro male subpopulation in the fifth question (i.e., the probability that a randomly selected Negro male would "be seen" with a Caucasian female).

It will be convenient to use some notation to represent the exact probability that is being used. The symbol Pr(A) will be used (instead of p) to represent the probability of A occurring (i.e., that a randomly selected member of the population has property A) with respect to the entire population of the area, and $\Pr(A \mid B)$ will be used to represent the probability of A occurring with respect to that portion of the population possessing property B. The symbol AB will mean the occurrence of both A and B.

## Statistical Dependence

The first error to be discussed is that of ignoring the existence of any statistical dependence that might exist among the various properties of interest. Consider the first question: "What is the probability of a man having a mustache?" If A represents the event of having a mustache, and B is the event of being a man, the desired probability is $\Pr(A \mid B)$. This is estimated as 1/3 according to the quoted answer. Now consider the fourth question: "What is the probability of a man having a beard?" If we let C be the event of having a beard, the probability is, in notation form, $\Pr(C \mid B)$, which is estimated as 1/10. Suppose we now ask: What is the probability that a man has a mustache *and* a beard? In notation form, this is $\Pr(AC \mid B)$. Using the reasoning represented in the news item, the answer would be $\Pr(A \mid B) \times \Pr(C \mid B)$, which is 1/30.

Suppose a person has been selected from the population at random (any population that the

[4] For example, see: Good, I. J., "Kinds of Probability", Science, *129* (1959) 443–447.

[5] This definition will be adhered to in connection with the basic figures, but will be generalized for further derivations.

reader is familiar with), and you are told that the person is a man. Using your own experience as a guide, estimate the probability that the person has a mustache (i.e., $Pr(A \mid B)$). Now suppose that another person is selected at random, and you are told that this person is a man and that he also has a beard. What is your estimate now of the probability that the person has a mustache (i.e., $Pr(A \mid BC)$)? The two estimates will probably not be the same; if not, it indicates that having a beard and having a mustache are not independent events. Under general conditions, it is easily shown that $Pr(AC \mid B) = Pr(A \mid BC) \times Pr(C \mid B)$, which is true regardless of any dependence between A and C. Now if $Pr(A \mid B)$ is not the same as $Pr(A \mid BC)$, the calculation of $Pr(AC \mid B)$ in the previous paragraph must be in error. Individual probabilities over the same population can be multiplied together only if they are independent.

Other questions that might be asked about the properties considered in the example are: Do blondes tend to wear pony tails more than brunettes, redheads, etc.? Is the probability of a Negro driving or owning a yellow car greater than that of a Caucasian driving or owning one? Is the probability of a person with a beard and mustache driving or owning a yellow car greater than is the case with the general population? Would a blonde Caucasian girl be more likely to go with a Negro male than would a Caucasian girl with some other hair color?

What do you do when the answer to these or similar questions is yes? Aside from mathematical manipulation for dependent probabilities, perhaps the easiest thing to do is to consider the dependent properties together. For instance, instead of asking about a man having a beard and then about a man having a mustache, ask about the chance of a man having both a beard and mustache, and estimate this as a single probability figure. If too many items are interdependent, this may not be practical, and mathematical methods will have to be used. For example, suppose the probability of a bearded man having a mustache is 9/10; then $Pr(A \mid BC) = 9/10$. Using the estimate of 1/10 for $Pr(C \mid B)$, and using the formula $Pr(AC \mid B) = Pr(A \mid BC) \times Pr(C \mid B)$, we find that $Pr(AC \mid B) = 1/10 \times 9/10 = 9/100$.

## Probabilities of Duplication and Error

The last question in the news item essentially asks about the probability of the combined occurrence of all the mentioned properties in a single random selection from some population or subpopulation, which is not clearly defined. The phrase "... at a given time and spot ...," is essentially irrelevant to the problem, and, in addition, involves properties beyond those already considered. If the time and place are those related to the crime, the probability requested is one, provided the information that the witnesses have given is correct. If they are not related to the crime, their consideration seems immaterial. The main probability which is of immediate interest is that for the occurrence of the logical product of all listed events in a single random selection from the population. Call this probability $Pr(T)$. However, this still does not present the situation in a form that is readily usable, and the second error to be noted is the use of $Pr(T)$ as a guide to the conclusion that the evidence and facts do or do not warrant a guilty verdict, or whatever verdict may be at issue.

The situation that exists is somewhat as follows. The investigating agency had information about certain properties that the guilty couple would possess. An intensive, non-random search was made, which was concluded when a couple possessing these properties was found. The important considerations now are how likely it is that: first, another couple with these same properties exists in the population, and second, the related probability that the conclusion the couple are guilty would be in error. If the probability that another couple with the same properties exists is high, could anyone be certain "beyond a reasonable doubt" that the couple on trial are the guilty couple, even if $Pr(T)$ is very small? A probability related to that of duplication is the probability of error, $Pr(Er)$, which is defined as the probability that a guilty verdict would be in error (based solely of course on the evidence used to derive the probability), and this figure offers a fairly clear basis to guide the decision making process.

A derivation of $Pr(Er)$ was made in a previous article[6], where it was shown that the probability of error could be related to $Pr(T)$ by one of two formulas, depending upon the actual model used, namely, $Pr(Er) \doteq \lambda/4$ or $\lambda/2$, where $\lambda = Pr(T) \times n$, $n$ being the total size of the population, and provided that $\lambda$ is less than one. This latter provision

[6] Kingston, C. R., "Applications of Probability Theory in Criminalistics", J. Am. Stat. Asso., *60* (1965) 70–80.

will certainly be true if the evidence has any significance at all. Although the actual derivation was made with reference to a particular kind of physical evidence, its extension to the type of evidence considered in the present case is not unreasonable.

### Probability of Chance Occurrence

Now that the probability of chance occurrence of the properties, Pr(T), has been shown to be a necessary intermediary to a usable figure, its derivation from the information given in the news item will be examined next. We want to find the probability that a randomly chosen member of the total population will have certain properties associated with him or her. In order to do this, it will be necessary to adjust the probabilities so that they refer to the same population. The third error, which is more of a confusion factor rather than an actual error, is the combination of probabilities that are based upon different populations or subpopulations without taking this into account and without clarifying the final reference population. So even though it may be permissable to multiply the individual probabilities which are independent in the news item together, the interpretation of the resulting probability is not clear.

A sample calculation of Pr(T) will be made without trying to assess the individual probabilities accurately. For convenience in presentation, letter symbols will be assigned to each of the properties; these are listed in Table 1.

For simplicity, certain assumptions will be made: (1) That no dependencies which may exist between the properties, except for A and C, would significantly change the results if they were taken into account in the calculations; (2) that the male member of the couple was the owner or provider of the yellow car; (3) that the adult or near adult population of Los Angeles county is the population being considered. Also, some probabilities not mentioned in the news item will have to be estimated. It will be assumed for purposes of the subsequent calculations that $\Pr(B) = \Pr(D) = 1/2$, $\Pr(E) = 14/100$, and $\Pr(I) = 75/100$.

TABLE 1

| Symbol | Property |
|---|---|
| A | having a mustache |
| B | being male |
| C | having a beard |
| D | being female |
| E | being a Negro |
| F | having blonde hair |
| G | wearing a pony tail |
| H | owning or driving a yellow car |
| I | being a Caucasian |
| J | associating with a Negro male |
| K | associating with a Caucasian female |

Properties J and K will need some attention to adjust them to the other properties and subpopulations considered. The figure of 1 in 1000 in the answer to the fifth question is not clearly related to any subpopulation, as previously noted. Since this article is concerned with the manipulation and interpretation of such figures, rather than the correctness of the individual estimates, 1/1000 will be taken as an estimate of the ratio of Negro males who associate with Caucasian females, where the association is something more than just casual. Thus $\Pr(K \mid BE) = 1/1000$. Assuming a ratio of about 1:5 for Negro males to Caucasian females in the population, and keeping in mind the definition of probability given earlier, we have $\Pr(J \mid DI) = 1/5000$.

With the simplifying assumptions, there will essentially be two possibilities for a random selection of one person from the population to have all properties of interest. One is for the person to be the male half of a couple possessing all properties (call this event Q), and the other is for the person to be the female half (call this event R). Since these are mutually exclusive events, the *sum* of their individual probabilities will equal Pr(T).

We will now need the probabilities of ABCEH and DFGI. $\Pr(ABCEH) = \Pr(AC) \times \Pr(BEH) = 9/100 \times (1/2 \times 14/100 \times 1/10) = 63/10^5$. The factorization was made to emphasize the handling of the dependency between A and C in connection with the other properties. $\Pr(DFGI) = 1/2 \times 1/4 \times 1/10 \times 75/100 = 75/(8 \times 10^3)$. Since dependencies other than those already specified or included in the estimates are being ignored, we have $\Pr(K \mid BE) = \Pr(K \mid ABCEH)$ and $\Pr(J \mid DI) = \Pr(J \mid DFGI)$. However, the additional properties in the other direction cannot be tossed in so casually. Let J′ be the property of associating with a Negro male who also has properties A, C, and H, and let K′ be the property of associating with a Caucasion female who also has properties F and G. A simple calculation shows that $\Pr(K' \mid ABCEH) = 9/10^6$ and $\Pr(J' \mid DFGI) = 1/(2 \times 10^5)$.

Using the general formula for conditional probabilities, we have $\Pr(Q) = \Pr(K' \mid ABCEH) \times \Pr(ABCEH) = 5.67/10^9$, and $\Pr(R) = \Pr(J' \mid DFGI) \times \Pr(DFGI) = 1/(2 \times 10^5)$. Thus $\Pr(T) = \Pr(Q) + \Pr(R) = 5.26/10^8$, which is the calculated probability of a chance occurrence of T in a single random selection from the population. As it happened, the more extensive breakdown of the properties involved in the calculations, which tended to decrease the value of Pr(T), was balanced by the consideration of dependencies and other factors which tended to increase the value of Pr(T). Consequently, the result arrived at in the news item, $8.34/10^8$ (another way of writing 1 in 12 million), is about the same as that calculated here, $5.26/10^8$. Therefore, the most significant error is not in the actual value of Pr(T) obtained, but rather in the use of Pr(T) instead of Pr(Er) in trying to convey the significance of the evidence to the jury.

## The Final Probability of Error

We now have the basic information to compute Pr(Er). Estimating the population of interest to be about two million, we have $\lambda = \Pr(T) \times n = .1051$. Using the formula for Pr(Er) favoring the accused, $\Pr(Er) = \lambda/2 = .0526$. This is interpreted to mean that, in the long run, about 5 percent of the conclusions of guilt, which are based on evidence giving rise to approximately the same value of Pr(Er), can be expected to be in error.

The effect of using Pr(T), and saying, "The chance of this evidence occurring is one in twelve million," should carefully be compared to the effect of using Pr(Er), and saying, "The chance of error in convicting these people on the basis of this evidence is about one in twenty." Although both statements are technically proper, the former omits some of the facts and conditions that are vital to the evaluation of the figure given, and is therefore misleading.

It is again emphasized that the figures are not presented as accurate or even approximate evaluations of the actual probabilities involved in the case under discussion. They only apply to the data presented in the newspaper report, plus the assumptions and additional probability estimates that were necessary. Further, the probability of one in twenty would apply only to the descriptive evidence; it is stated in the news item that there was other circumstantial evidence to be considered. This additional evidence would, provided it was incriminating, reduce the probability of error beyond that which might be calculated for the descriptive evidence alone.

## Some Other Problems

There are a few additional factors that should be mentioned in connection with such applications of probability theory.

Even though the analysis given above is internally consistent, the resulting figures are only as good as the raw figures fed into the mathematical machinery. The estimates of the individual probabilities must be reasonably accurate to expect valid results from the calculations. There is no indication of how the estimates quoted were derived, and consequently no complaint can be registered against them. Such estimates should be based on an actual sample of reasonable size if they are going to be explicitly presented in court. The other alternative, which is far too common in the legal-probabilistic setting, is to make a "conservative guess". That is, to select a probability of chance occurrence that is clearly much higher than the true one, using the justification that the final calculation will then be higher than it should, thus underemphasizing its importance. If such conservative guesses are indeed conservative, this is a perfectly good procedure, as long as the limitations are not forgotten. It would be instructive to perform some simple experiments to determine the relationships between the figures arrived at by the two alternative procedures.

Another factor that cannot be lost sight of is the probability that the witnesses are in error. This affects the final result by adding a possibility that the guilty persons do not possess all of the properties that form the description. In the above analysis, the probability that the guilty persons do possess all of the properties is assumed to be one. The mathematical treatment of the same situation when this probability is not one has not yet been theoretically derived in a form suitable for direct application.

Another factor to be considered is the probability that the guilty couple are no longer in the area being searched. The probability of error can be no smaller than the probability of the couple not being in the population of size $n$. One way of minimizing this probability is to consider the entire population of the world. If the evidence has a low enough probability of chance occurrence, this can be a useful procedure, but if minimal amounts of evidence are being worked with, a balance must be made between the probability of error and the probability that the guilty persons are in a particular population or area.

## SUMMARY AND CONCLUSIONS

It is hoped that the above discussion of the Collins case will create an awareness of the value, as well as the pitfalls, of an explicit probabilistic approach to the evaluation of evidence in those persons—judges, attorneys, witnesses (both lay and expert), jurors, etc.—who may have a directing force, or indeed any interest, in our system of justice. The applicability of the evidence to the case at hand can be determined far more accurately by making the varous steps in the evaluation explicit and analyzing them with *valid* logical procedures. We have seen how an impressively small probability figure takes on an entirely different significance when evaluated with respect to its true logical content.

The three major errors that were discussed are:

1. Failure to consider all possible statistical dependencies that may exist between individually estimated probability figures.

2. Using a probability of a chance occurrence of a set of properties as a guide to the significance of evidence rather than a probability related to the duplication of the set of properties in the population.

3. Combining probabilities formulated over different populations or subpopulations, and not clarifying the final population or the interpretation of the results.

It will only be by recognizing the above, and similar, errors and problems that the explicit application of probability theory in legal matters will attain its proper recognition and its full potential.

NOTE: After completing this article, this author had the opportunity of discussing the relevant aspects of the case with the defending attorney. It seems that the presentation of the situation is somewhat clearer in the TIME magazine write-up (January 8, 1965, page 42). The questions and answers quoted earlier in this article were apparently not involved in the expert's testimony, but were rather hypothetical questions composed by the newspaper reporter on the basis of the prosecutor's summary to the jury. It appears that the expert testified only as to the manipulation of independent probabilities, such as in dice throwing situations, and had no opportunity to testify about the factors relative to the particular case or to be subjected to cross-examination on these issues. It also appears that there was sufficient circumstantial evidence other than that which the probability calculations were based upon to seriously question the assertion that these probability calculations played the dominant role.