Michael O. Finkelstein
Bruce Levin

Statistics for Lawyers

Second Edition

With 55 Figures

Springer

CHAPTER 3

Elements of Probability

3.1 Some fundamentals of probability calculation

Many probability puzzles can be made transparent with knowledge of a few basic rules and methods of calculation. We summarize some of the most useful ideas in this section.

Probabilities

In the classic formulation, probabilities are numbers assigned to elements of a *sample space*. The sample space consists of all possible outcomes of some conceptual experiment, such as flipping a coin or rolling dice. In making calculations, probability numbers are assigned to the "simple" (indecomposable) elements, or *events*, of the sample space. A subset of simple events is called a "compound" event, and its probability is defined as the sum of the probabilities of the simple events it contains. For example, if a coin is tossed four times, the simple events comprising the sample space are the possible sequences of heads and tails in four tosses; there are 16 such sequences, of which heads-tails-heads-tails (HTHT) is an example. A compound event is some subset of the 16 sequences, such as tossing two heads in four tails, which consists of the simple events HHTT, HTHT, HTTH, HHTT, THHT, and TTHH. The rules of the calculus of probabilities discussed in this section give some short cuts for calculating probabilities of compound events. Some probability chestnuts turn on the proper definition of the sample space, as the example below shows.

Example. Assume that boys and girls are born with equal frequency. Mr. Able says, "I have two children, and at least one of them is a boy." What is the probability that the other child is a boy? Mr. Baker says, "I went to the house

of a two-child family, and a boy answered the door." What is the probability that the other child is a boy?

The answer to the first question is 1/3, but the answer to the second question is 1/2. The reason for this small paradox is that the sample space defined by Mr. Able consists of families with three birth sequences of children: boy-girl, girl-boy, and boy-boy. Since each has the same probability, and since only in the boy-boy case is the "other" child a boy, the probability of that event is 1/3. In Mr. Baker's case, by designating a particular boy (the one who answered the door) the sample space of family types is reduced to two: the door-answering child is a boy, and the other is a girl, or the door-answering child is a boy, and the other is also a boy. Since there are only two possible family types, the probability of boy-boy is 1/2, i.e., the probability of the other child being a boy is 1/2.

The probability assigned to a simple event can be any number between 0 and 1. For purposes of applications, useful assignments represent the long-range frequency of the events, such as 1/2 for the probability of tossing heads with a coin. However, nothing in mathematical theory compels any particular assignment of probabilities to simple events, except that an event certain *not* to occur has probability 0,[1] and the probabilities of the simple events constituting a sample space must sum to 1.

The concept of probability also applies to the degree of belief in unknown events, past or future. It has been shown that the calculus of probabilities introduced in this section can consistently be applied to both interpretations so that what probability "really is" need not be resolved to make use of the mathematical theory. See Section 3.6.

Complementary events

The probability of the negation of a given event is one minus the probability of the event. In symbols,

$$\mathrm{P}[\bar{\mathrm{B}}] = 1 - \mathrm{P}[\mathrm{B}].$$

Examples.

- If there is a 1 in 5 chance of selecting a black juror from a venire, there is a 4 in 5 chance of selecting a non-black juror.
- A game-show host (Monty Hall) presents a contestant with three closed doors. Behind one of them is a prize, with equal likelihood for doors A, B, or C. The contestant is asked to select a door; say he picks A. Before opening that door, Monty opens one of the other doors, which both he and the

[1] In infinite sample spaces, an event with probability 0 may still occur if the infinitude of possibilities is non-denumerable. For example, when a dart is thrown at a target, the probability that any specified point will be hit is zero because points have zero width, but still the dart hits the target at some point.

contestant know will be empty; say B. Then Monty presents the contestant with a choice: he may either stick with A or switch to C. Question: Is it better to stick, or to switch, or does it not matter?
Answer: The contestant has a 1/3 chance that the prize is behind A and a $1 - 1/3 = 2/3$ chance that the prize is behind B or C. Monty's opening of B doesn't change the probability that the prize is behind A, because it gives no information about the correctness of that choice. But by eliminating B, the probability of the prize being at C is now 2/3. In short, the best strategy is to switch.

Disjoint unions

- The union of two events A and B is the occurrence of A or B (or both), i.e., their *inclusive disjunction*. Thus, the union of two events encompasses both the occurrence of either event without the other (their *exclusive disjunction*) or the occurrence of both together (their *exclusive conjunction*).
- A and B are *disjoint* if they are mutually exclusive (i.e., their joint occurrence, or logical conjunction, is impossible).
- The probability of a union of disjoint events A and B is the sum of their individual probabilities. In symbols,

$$P[A \text{ or } B] = P[A] + P[B].$$

Example. Suppose:

A = [Mr. F. is the girl's father and has her blood type], with $P[A] = 0.5$, and
B = [Mr. F. is *not* the girl's father, but has her blood type], with $P[B] = 0.2$.

Then:

$$P[A \text{ or } B] = P[\text{Mr. F has the girl's blood type}] = 0.7.$$

Unions in general

For events A and B, not necessarily disjoint, the probability of A or B equals the sum of the individual probabilities less the probability of their conjunction, A and B. In symbols,

$$P[A \text{ or } B] = P[A] + P[B] - P[A \text{ and } B].$$

The joint probability $P[A \text{ and } B]$ is subtracted because the sum of the probability of A and the probability of B generally overestimates $P[A \text{ and } B]$ because it counts the probability of the joint occurrence of A and B twice: once in the probability of A alone and again in the probability of B alone. This double counting is shown by the top Venn diagram in Figure 3.1, where

the areas within the circles are proportional to the probabilities of events A and B, respectively. The union of the two events is represented by the total area covered by A and B. The region labelled AB represents their intersection (conjunction).

It follows that the probability of A or B is no greater than the sum of their individual probabilities. In symbols,

$$P[A \text{ or } B] \leq P[A] + P[B].$$

The same reasoning leads to the more general inequality

$$P(A \text{ or } B \text{ or } C \ldots) \leq P(A) + P(B) + P(C) + \ldots$$

Although there are other inequalities due to Bonferroni, this is known as the Bonferroni inequality because it is the one most frequently encountered in practice.[2] See Section 6.2 for a discussion of the use of this inequality in setting levels of statistical significance for tests involving multiple comparisons.

Example. Referring to Section 2.1.4, assume that the probability of observing six or more cases of leukemia in any single census tract is 0.007925. Let $P[A_i]$ be the probability of observing such an event in census tract i. Then the probability of observing six or more cases of leukemia in at least one of six census tracts is no greater than

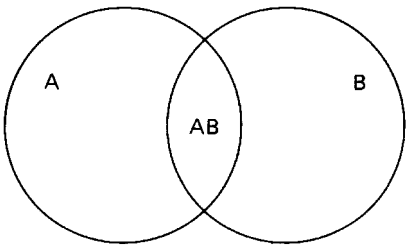
$$P[A_i] + P[A_2] + \ldots + P[A_6] = 6 \times 0.007925 = 0.04755.$$

In Section 2.1.4, where there were only 12 cases in total, the joint probability of finding six cases in two tracts is the minuscule 0.000006, and the probability of this occurrence in three or more tracts is zero. Bonferroni's first approximation is extremely accurate in this case.

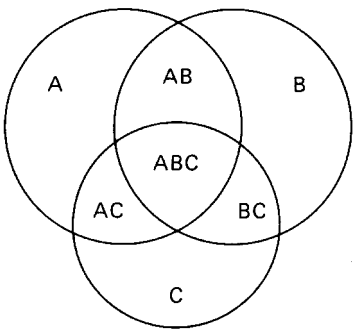
Example. Assume that a person is infected with a virus for which there are two diagnostic tests. Let A = [diagnostic test 1 is positive], with $P[A] = 0.95$. Let B = [diagnostic test 2 is positive], with $P[B] = 0.99$

What is the probability that the infected person would test positive on at least one of the tests? By Bonferroni's inequality $P[A \text{ or } B] \leq 0.95 + 0.99 = 1.94$. Clearly this is not helpful. If the probability that both tests would be positive were $0.95 \times 0.99 = 0.9405$ (independent tests), then the probability of the union would be $0.95 + 0.99 - 0.9405 = 0.9995$. In this case the Bonferroni approximation is not useful because the probability of the joint event

[2]This is Bonferroni's first inequality. More generally, if $A_1, \cdots, A_n$ are n events, the probability of the union of $A_1, \cdots, A_n$ is obtained as follows: first, take the sum of each event separately, $\sum_i P[A_i]$; second, subtract the sum of all pairwise joint probabilities $\sum_{i \neq j} P[A_i \cap A_j]$ since the previous sum overestimates the union's probability; next, add back the sum of all joint probabilities in triples, $\sum_{i \neq j \neq k} P[A_i \cap A_j \cap A_k]$, since the previous step overcorrected slightly. Continue in this way by alternately subtracting and adding sums of joint probabilities until one adds or subtracts the final term $P[A_1 \cap \cdots \cap A_n]$. At any stage, the probability of the union may be approximated by the terms included up to that point, incurring an error no larger than the magnitude of the first omitted term.



P(A or B) = P(A) + P(B) − P(AB)



P(A or B or C) = P(A) + P(B) + P(C) − P(AB) − P(AC) − P(BC) + P(ABC)

FIGURE 3.1. Venn diagrams illustrating conjunction ("and") and inclusive disjunction ("and/or")

is not negligible. In fact, Bonferroni's inequality shows that $P[A \text{and} B]$ must be substantial, as follows. The probability that one or the other test shows a false negative reading, P[$\bar{\text{A}}$ or $\bar{\text{B}}$], is, by Bonferroni's inequality, no greater than P[$\bar{\text{A}}$]+P[$\bar{\text{B}}$] = 0.05+0.01 = 0.06, which is quite accurate because the probability of two false negatives is very small ($0.01 \times 0.05 = 0.0005$, for independent tests). Taking complements implies that P[A and B] = P[not ($\bar{\text{A}}$ or $\bar{\text{B}}$)] = $1 - \text{P}[\bar{\text{A}} \text{ or } \bar{\text{B}}] \geq 1 - 0.06 = 0.94$.

Intersection

- The intersection of two events, A and B, is their conjunction, A and B.
- The probability of A and B is no greater than the probability of either event alone. This follows from the rule for mutually exclusive events, since

$$P[A] = P[A \text{ and } B] + P[A \text{ and } \bar{B}] \geq P[A \text{ and } B].$$

- If event B implies event A then $P[B] = P[A \text{ and } B]$, whence $P[B] \leq P[A]$. In words, an event is at least as likely to occur as any other event which

implies it, and an event is no more likely to occur than any other event which it implies.

Conditional probability

The conditional probability of an event A given an event B with $P[B] > 0$ is defined as $P[A|B] = P[A \text{ and } B]/P[B]$. The conditional probability of A given B is the *relative* likelihood of occurrence of A, among all times when B occurs.

Example. Referring to the boy-girl family example at p. 57, the unconditional probability of a boy-boy family (A) is P[A]= $1/4$. The probability of a family with at least one boy (B) is P[B]= $3/4$. The probability of a boy-boy family conditional on there being at least one boy (B) is P[A and B]/P[B]= $1/4 \div 3/4 = 1/3$. In Mr. Baker's case, the probability of the door-answering child being a boy (C) is P[C]= $1/2$. Hence, the probability of a boy-boy family conditional on a boy answering the door is P[A and C]/P[C]=$1/4 \div 1/2 = 1/2$.

For non-vacuous events, one can always write the joint probability of events A and B as

$$P[A \text{ and } B] = P[A|B]P[B] = P[B|A]P[A],$$

although note that in general $P[A|B] \neq P[B|A]$. In the preceding example, P[A|B] = $1/3$, but P[B|A] = 1.

Independent events

Events A and B are said to be *independent* if

$$P[A|B] = P[A] = P[A|\bar{B}].$$

In words, A and B are independent if the occurrence (or non-occurrence) of B does not affect the likelihood of occurrence of A. Thus, for independent events we have the multiplication rule

$$P[A \text{ and } B] = P[A|B] \cdot P[B] = P[A] \cdot P[B].$$

Averaging conditional probabilities

An overall probability equals a weighted average of conditional probabilities. In symbols,

$$P[A] = P[A|B] \cdot P[B] + P[A|\bar{B}] \cdot P[\bar{B}].$$

The weights are given by $P[B]$ and its complement. This follows by writing

$$\begin{aligned} P[A] &= P[(A \text{ and } B) \text{ or } (A \text{ and } \bar{B})] \\ &= P[A \text{ and } B] + P[A \text{ and } \bar{B}]. \end{aligned}$$

The conditional probabilities are often interpreted as specific rates.

Example. The overall promotion rate $P[A]$ in a company may be obtained from the specific rates for black and non-black employees (respectively $P[A|B]$ and $P[A|\bar{B}]$) by weighting the rates by the proportion of blacks and non-blacks, $P[B]$ and $P[\bar{B}]$.

3.1.1 Interracial couple in yellow car

In *People v. Collins*, 68 Cal. 2d 319, 438 P.2d 33 (1968)(*en banc*) an elderly woman, while walking through an alley in the San Pedro area of Los Angeles, was assaulted from behind and robbed. The victim said that she managed to see a young woman with blond hair run from the scene. Another witness said that a Caucasian woman with dark-blond hair and a ponytail ran out of the alley and entered a yellow automobile driven by a black man with a mustache and a beard.

A few days later, officers investigating the robbery arrested a couple on the strength of these descriptions and charged them with the crime.[3] It is not clear what led police to the Collins couple. A police officer investigating the robbery went to their house and took them to the police station, where they were questioned, photographed, and then released. When police officers came a second time to their house, apparently to arrest them, Malcolm was observed running out the back and was found hiding in a closet in a neighboring house. The officer found two receipts in Malcolm's pocket, indicating that he had just paid two traffic fines in a total amount equal to the amount of money stolen. Questioned as to the source of the money, Malcolm and Janet gave conflicting accounts. Finally, first Janet alone, and then the two together, engaged in a bargaining session with police in an effort to have charges dismissed against Malcolm, in particular, because he had a criminal record. Although no admissions were made, the tone of the conversion, according to the appellate court, "evidenced a strong consciousness of guilt on the part of both defendants who appeared to be seeking the most advantageous way out."

At their trial, the prosecution called an instructor of mathematics to establish that, assuming the robbery was committed by a Caucasian blonde with a ponytail who left the scene in a yellow car driven by a black man with a beard and mustache, the probability was overwhelming that the accused were guilty because they answered to this unusual description. The mathematician testified to the "product rule" of elementary probability theory, which states that the probability of the joint occurrence of a number of mutually independent events equals the product of their individual probabilities. The prosecutor then had the witness assume the following individual probabilities for the relevant characteristics:

---

[3] When defendants were arrested, the woman's hair was light, not dark, blond and the man did not have a beard. There was some evidence that the man had altered his appearance after the date on which the offense had been committed. The car was only partly yellow.

(a)	Yellow automobile	1/10
(b)	Man with mustache	1/4
(c)	Girl with ponytail	1/10
(d)	Girl with blond hair	1/3
(e)	Black man with beard	1/10
(f)	Interracial couple in car	1/1000

Applying the product rule to the assumed values, the prosecutor concluded that there was but one chance in twelve million that a couple selected at random would possess all these incriminating characteristics. The prosecutor gratuitously added his estimation that the "chances of anyone else besides these defendants being there ... having every similarity ... is somewhat like one in a billion." The jury convicted the defendants. On appeal, the Supreme Court of California reversed, holding that the trial court should not have admitted the evidence pertaining to probability.

In an appendix to the opinion, the court proposed a mathematical model to prove that, even on the prosecution's assumption that the probability that a random couple would answer to the description of the Collins couple (a "C-couple") was 1 in 12,000,000, there was a 41% probability of there being at least a second C-couple, if 12,000,000 selections were made. One way of describing the court's model involves imagining that there is a very large population of couples in cars, with the rate of C-couples in that population being 1 in 12,000,000. Out of this population one picks a couple at random, checks whether it is a C-couple, returns it to the population, and picks again. Assuming that one makes 12,000,000 selections to simulate the creation of the population, count the number of C-couples picked. Repeat this process many times and one finds that, in about 41% of these hypothetical populations in which there appeared at least one C-couple, there was more than one. The appendix concludes:

> Hence, even if we should accept the prosecution's figures without question, we would derive a probability of over 40 percent that the couple observed by the witnesses could be "duplicated" by at least one other equally distinctive interracial couple in the area, including a Negro with a beard and mustache, driving a partly yellow car in the company of a blonde with a ponytail. Thus, the prosecution's computations, far from establishing beyond a reasonable doubt that the Collinses were the couple described by the prosecution's witnesses, imply a very substantial likelihood that the area contained more than one such couple, and that a couple other than the Collinses was the one observed at the scene of the robbery.

Id. 438 P. 2d at 42.

Questions

1. Are the identifying factors listed likely to be statistically independent?

2. Suppose the frequency of identifying factors had been determined as follows: A survey was made of couples in cars and one in a thousand of them was interracial. In one of ten of those interracial couples there was a black man with a beard. In one of three interracial couples in which the black man had a beard the woman had blond hair. And so forth for the rest of the factors. In those circumstances would multiplication together of the frequencies have been correct?

3. Assuming the 1 in 12,000,000 figure were correct as the probability of selecting at random a couple with the specified characteristics, what objections do you have to the prosecutor's argument?

4. Assume, as in the appendix to the court's opinion, that the couples who might conceivably have been at the scene of the crime were drawn from some larger population in which the rate of C-couples was 1 in 12,000,000. Does it follow that the 41% probability computed in the appendix is relevant to the identification issue?

5. Does the conclusion of the court's appendix that the prosecutor's computation implies a very substantial likelihood that a couple other than the Collinses was the one observed at the scene of the robbery follow from the preceding mathematical demonstration?

Source

Michael O. Finkelstein and William Fairley, *A Bayesian Approach to Identification Evidence*, 83 Harv. L. Rev. 489 (1970), reprinted in *Quantitative Methods In Law*, ch. 3 (1978); see also, William Fairley and Frederick Mosteller, *A Conversation About Collins*, 41 U. Chi. L. Rev. 242 (1974).

Notes

The product rule has been involved in a number of cases both before and after *Collins*.

In *People v. Risley,* 214 N.Y. 75, 108 N.E. 200 (1915), the issue was whether defendant had altered a court document by typing in the words "the same." Defendant was a lawyer and the alteration helped his case. Eleven defects in the typewritten letters on the court document were similar to those produced by the defendant's machine. The prosecution called a professor of mathematics to testify to the chances of a random typewriter producing each of the defects found in the words. The witness multiplied these component probabilities together to conclude that the joint probability of all the defects was one in four billion. On appeal the court reversed, objecting that the testimony was "not based upon actual observed data, but was simply speculative, and an attempt to make inferences deduced from a general theory in no way connected with

the matter under consideration supply [sic] the usual method of proof." *Id.* at 85, 108 N.E. at 203.

In *Miller v. State,* 240 Ark. 340, 399 S.W.2d 268 (1966), the expert testified that dirt found on defendant's clothing matched dirt at the burglary site as to color, texture, and density and that the probability of a color match was 1/10, a texture match was 1/100, and a density match was 1/1000. Multiplying these together, the expert concluded that the probability of an overall match was 1/1,000,000. On appeal, the conviction was reversed. The expert's testimony as to probability was inadmissible because he had neither performed any tests, nor relied on the tests of others, in formulating his probability estimates.

On the other hand, in *Coolidge v. State,* 109 N.H. 403, 260 A.2d 547 (1969), the New Hamphire Supreme Court cited *Collins*, but came to a different conclusion. The expert in that case obtained particles by vacuuming the victim's clothes and the defendant's clothes and automobile (where the crime was believed to have taken place). Forty sets of particles (one from the victim and the other from the defendant) were selected for further testing on the basis of visual similarity under a microscope. In these further tests the particles in 27 of the 40 sets could not be distinguished. Previous studies made by the expert indicated that "the probability of finding similar particles in sweepings from a series of automobiles was one in ten." The expert concluded that the probability of finding 27 similar particles in sweepings from independent sources would be only one in ten to the 27th power. On cross-examination, the expert conceded that all 27 sets may not have been independent of one another, but the court found that this went to weight rather than admissibility and affirmed the conviction.

3.1.2 Independence assumption in DNA profiles

We continue the discussion of DNA profiles that was begun in Section 2.1.1.

In forensic science applications, a sample of DNA that is connected with the crime is compared with the suspect's DNA. For example, in a rape case, DNA in semen found in the victim may match the suspect's DNA (which would be incriminating) or may not (which would be exculpatory). If there is a match, its forensic power depends on the probability that such a profile would have been left by a randomly selected person, if the suspect was not responsible. This in turn depends on the frequency of matching profiles in the population of persons who could have left the trace.

To estimate such frequencies, DNA laboratories begin with the frequency of the observed allele at each locus included in the profile. Each heterozygous genotype frequency is determined by multiplying together the frequencies of the maternal and paternal alleles that constitute the genotype, and multiplying that product by two. For homozygous genotypes, the frequency of the observed allele should be squared. The independence assumption that underlies the multiplications is justified on the belief that people mate at random, at least with respect to VNTR alleles, which are not known to correspond to any observable

trait. However, the same assumption is made for PCR alleles, which may involve observable traits. A population whose genotypes are in these proportions is said to be in Hardy-Weinberg (HW) equilibrium.

Genes that are on the same chromosome are *linked*, that is, they tend to be inherited together. However, during the formation of a sperm or egg, the two members of a chromosomal pair lined up side by side can randomly exchange parts, a process called *crossing over* or *recombination*. Genes that are very close together on the same chromosome may remain associated for many generations while genes that are far apart on the same chromosome or on different chromosomes become randomized more rapidly.

To arrive at an overall frequency for a multilocus genotype, it is usual to take the product of the frequencies of the genotypes at the separate loci. This is justified on the assumption that genotypes at different loci are independent. A population in such a state is said to be in *linkage equilibrium* (LE). The state of LE, like HW, is the result of random mating, but a population only arrives at LE after several generations, whereas HW is arrived at in one generation. Because of recombination, loci that are close together on the same chromosomal pair approach LE more slowly than those far apart on the same pair or on different pairs. Departure from LE is called *linkage disequilibrium*, and is an important tool for locating marker genes close to true disease genes.

It has been objected that frequencies of apparent homozygotes will be greater than expected under HW if either (i) there are subgroups in the population that tend to in-breed and have higher rates of the particular alleles observed as homozygotes, or (ii) only a single band is found at a locus because the autorad band for the other allele erroneously has been missed. To protect against these possibilities, some laboratories conservatively estimate the frequency of a homozygote as twice the frequency of the observed allele, instead of the square of its frequency. This change will generally favor defendants by increasing the estimated frequency of matching homozygous genotypes in the population.

To test the HW assumption, Table 3.1.2 shows allele frequencies for 3 out of 28 different alleles found at locus D2A44 in samples of varying sizes from four

TABLE 3.1.2. Numbers of three alleles at locus D2S44 in samples from four populations

Allele type (i)	Canadian	Swiss	French	Spanish	Total alleles by type
⋮					
9	130	100	68	52	350
10	78	73	67	43	261
11	72	67	35	48	222
⋮					
Total alleles/sample	916	804	616	508	2844

white populations–Canadian, Swiss, French, and Spanish. Presumably there is not extensive mating at random across these populations so that hypothetical combined populations (e.g., the U.S. white population) would be vulnerable to departures from HW.

Questions

1. To arrive at an average figure for a total population consisting of the four subpopulations combined, compute the weighted average frequency of the homozygous genotype consisting of allele 9 (without adjustment for the risk of missing a band) and the heterozygous genotypes consisting of alleles 9 and 10 across the four subpopulations, with weights proportional to the sizes of the subpopulation samples. (These calculations assume that there is mating at random *within* each subpopulation, but not necessarily across subpopulations.)

2. Compute the frequency of the same alleles using the total population figures. (These calculations assume that there is mating at random within and across the subpopulations.)

3. Compare the results. Is HW justified?

4. Since there is not, in fact, mating at random across these subpopulations, what is a sufficient condition for HW in the total population, given HW in each of the subpopulations?

5. Consider a hypothetical population comprised of Canadians and non-Canadians in equal numbers. The Canadians have allele 9 frequency 130/916 as in Table 3.1.2, but the non-Canadians have allele frequency 786/916. Assuming that HW holds for Canadians and non-Canadians separately and they don't intermarry, does HW hold in the combined population?

Source

Federal Bureau of Investigation, *VNTR population data: a worldwide survey*, at 461, 464-468, *reprinted in* National Research Council, *The Evaluation of Forensic DNA Evidence*, Table 4.5 at 101 (1996).

Notes

DNA profiling has been subject to searching criticism by professional groups and in the courts, but its value is now well established. The National Research Council Report is one of many discussions of this subject. See, e.g., Faigman, et al., *Modern Scientific Evidence*, ch. 47 (1997).

3.1.3 Telltale fibers

Defendant Wayne Williams was charged with the murders of two young black males in Atlanta, Georgia. There had been ten other similar murders. Critical evidence against Williams consisted of a number of fibers found on the bodies that resembled fibers taken from his environment, in particular, certain unusual trilobal Wellman 181-b carpet fibers dyed English Olive. A prosecution expert testified that this type of fiber had been discontinued and that, on conservative assumptions, there had been only enough sold in a ten-state area to carpet 820 rooms. Assuming that sales had been equal in each of the ten states, that all Georgia carpet had been sold in Atlanta, and that only one room per house was carpeted, 81 Atlanta homes had carpet containing this fiber. Because, according to the expert, there were 638,992 occupied housing units in Atlanta, the probability that a home selected at random would have such carpeting was less than 81/638,992 or 1 in 7,792. Wayne Williams's bedroom had carpet with this fiber (although defendant subsequently disputed this).

Based on this testimony, the prosecutor argued in summation that "there would be only one chance in eight thousand that there would be another house in Atlanta that would have the same kind of carpeting as the Williams home." Williams was convicted. On appeal, the Georgia Court of Appeals held that the state's expert was entitled to discuss mathematical probabilities, that counsel in closing argument was not prohibited from suggesting inferences to be drawn from the evidence, and that such inferences might include mathematical probabilities.

Questions

1. Is the prosecutor's argument correct?

2. Does the 1 in 7,792 figure imply that there is 1 chance in 7,792 that the fibers did not come from Williams's home?

3. Should the evidence have been excluded because by itself (no other evidence being considered), the probability of guilt it implies is no more than 1 in 81?

Source

Williams v. State, 251 Ga. 749, 312 S.E.2d 40 (1983).

3.1.4 Telltale hairs

State v. Carlson, 267 N.W.2d 170 (Minn. 1978).

Defendant Carlson was charged with murdering a 12-year old girl who had last been seen in his company. Investigating officers found two foreign pubic hairs stuck to the skin of the deceased in her groin area and head hairs clutched in

her hand. Gaudette, an expert on hair comparisons, testified that the pubic and head hairs found on the victim microscopically matched those of the accused. Based on a study he had done a few years earlier [the Gaudette-Keeping study described below] for the pubic hair the "chances those hairs did not come from David Carlson would be on the order of 1 chance in 800 for each hair," and for the head hair the figure was 1 in 4,500. Carlson was convicted. On appeal, the Supreme Court of Minnesota found that Gaudette's testimony on mathematical probabilities was improperly received because of "its potentially exaggerated impact on the trier of fact," but affirmed the conviction because the evidence was merely "cumulative and thus nonprejudicial on the facts of the case."

State v. Massey, 594 F.2d 676 (8th Cir. 1979).

Defendant Massey was charged with bank robbery. The robber wore a blue ski mask and a similar mask was recovered from the house of Massey's associate. At his trial, an FBI expert testified that three out of five hairs found in the mask were microscopically similar to one or more of nine mutually dissimilar hairs taken from Massey's scalp.[4] Under questioning by the judge, the expert testified that he had examined over 2,000 cases and that "only on a couple of occasions" had he seen hairs from two different individuals that he "could not distinguish." He also made reference to the Gaudette-Keeping study, which found, as he described it, "that a possibility that a hair which you have done or matched in the manner which I have set forth, there's a chance of 1 in 4,500 these hairs could have come from another individual." In summation, the prosecutor argued that, assuming there were as many as 5 instances out of 2,000 in which hairs from different individuals could not be distinguished, the accuracy was better than 99.44% and thus constituted proof of guilt beyond a reasonable doubt. Massey was convicted. The court of appeals reversed the conviction, holding that the prosecutor had "confused the probability of occurrence of the identifying marks with the probability of mistaken identification of the bank robber." It also followed *Carlson* in objecting to the evidence because of its "potentially exaggerated impact upon the trier of fact."

Gaudette-Keeping study

The unwillingness of the courts to accept population frequency evidence in *Carlson* and *Massey* may have been due in part to, or at least justified by, the weaknesses of the underlying studies on which the estimates were based. The Gaudette-Keeping study to which the experts referred had been conducted several years earlier and used the following methodology. A sample of 80 to

[4]It is unclear how a hair could be similar to more than one of nine mutually dissimilar hairs.

100 hairs "randomly selected" from various parts of the scalps of 100 subjects was reduced to a subsample of 6 to 11 representative hairs from each subject (861 in all). Investigators examined every inter-person pair of hairs macroscopically and microscopically. Only 9 inter-person pairs were found indistinguishable. The investigators knew, however, that hairs from different people were involved. According to Gaudette, hair comparisons are somewhat subjective, and when experiments included "common featureless hairs," investigators were unable to distinguish a much higher proportion of hairs than in the original study. Nevertheless, Gaudette concluded in testimony in *Carlson* that "if nine dissimilar hairs are independently chosen to represent the hair on the scalp of Individual B, the chance that the single hair from A is distinguishable from all nine of B's may be taken as $(1 - (1/40{,}737))^9$, which is approximately $1 - (1/4500)$."

Questions

1. Do you see how Gaudette-Keeping derived their estimate of 1/4500 as the probability of being unable to distinguish a hair selected at random from any of 9 selected from a subject? What assumptions underlie the method of calculation?

2. Assuming the study results were accurate and representative, what two possible meanings are attributable to the expert's conclusion? Which is validly deducible without other assumptions from the study?

3. What issues would you explore on cross-examination or in rebuttal testimony with respect to the validity of the study?

Source

Gaudette's studies were reported in B. D. Gaudette and E. S. Keeping, *An Attempt at Determining Probabilities in Human Scalp Comparison*, 19 J. Forensic Sci. 599 (1974); *Probabilities and Human Pubic Hair Comparisons*, 21 *id*. 514 (1976); *Some Further Thoughts on Probabilities and Human Hair Comparisons*, 23 *id*. 758 (1978). Gaudette's work was criticized by P. D. Barnett and R. R. Ogle in *Probabilities and Human Hair Comparison*, 27 *id*. 272 (1982). Despite the weaknesses of the underlying studies, forensic hair analysis has rarely been rejected by the courts. For a discussion of the cases and the studies, see Clive A. Stafford Smith & Patrick Goodman, *Forensic Hair Comparison Analysis: Nineteenth Century Science or Twentieth Century Sham?*, 27 Colum. Hum. Rts. L. Rev. 227 (1996).

Notes

Carlson was not a fluke—in Minnesota. In *State v. Boyd*, 331 N.W. 2d 480 (1983), the Minnesota Supreme Court followed *Carlson*, but added some new

reasons for excluding population frequency statistics. It held that a population frequency statistic of less than 1 in 1,000 should not have been admitted in evidence because of a "real danger that the jury will use the evidence as a measure of the probability of defendant's guilt or innocence, and that evidence will thereby undermine the presumption of innocence, erode the values served by the reasonable doubt standard, and dehumanize our system of justice." *Id.* at 483. *Boyd* was followed in *State v. Kim*, 398 N.W.2d 544 (1987) (population frequency less than 3.6%) and, with specific reference to DNA profiling, in *State v. Schwartz*, 447 N.W.2d 422 (1989).

The Minnesota legislature responded to the trilogy of cases ending with *Kim* by passing an act providing that: "In a civil or criminal trial or hearing, statistical population frequency evidence, based on genetic or blood test results, is admissible to demonstrate the fraction of the population that would have the same combination of genetic marks as was found in a specific human biological specimen." Minn. Stat. §634.26 (1992). In subsequent rape cases, the Minnesota Supreme Court ignored the statute and opted for a "black box" approach: quantification of random match probabilities for DNA profiles may not be presented to the jury, although an expert may use them as the basis for testifying that, to a reasonable scientific certainty, the defendant is (or is not) the source of the bodily evidence found at the crime scene. See, e.g., *State v. Bloom*, 516 N.W.2d 159 (1994). Is this a reasonable solution to the problems of misinterpretation noted by the Minnesota Supreme Court?

The position of the Minnesota Supreme Court, as articulated in *Boyd*, is an extension to population frequency statistics of an argument by Professor Laurence Tribe against the use of Bayes's theorem in evidence. See Section 3.3.2 at p. 79. In his article, Professor Tribe objected to quantification of guilt, which Bayes's theorem could in some applications require, but did not go so far as to advocate the exclusion of population frequency statistics.[5] Most courts have not followed the Minnesota Supreme Court on this issue. The conclusion of Judge Easterbrook in his opinion in *Branion v. Gramly*, 855 F.2d 1256 (7th Cir. 1988), seems more reasonable and probably represents the dominant view:

> Statistical methods, properly employed, have substantial value. Much of the evidence we think of as most reliable is just a compendium of statistical inferences. Take fingerprints. The first serious analysis of fingerprints was conducted by Sir Francis Galton, one of the pioneering statisticians, and his demonstration that fingerprints are unique depends entirely on statistical methods. Proof based on genetic markers (critical in rape and paternity litigation) is useful though altogether statistical. So, too, is evidence that, for example, the defendant's hair matched hair

[5]The introduction of population frequency statistics, without Bayes's theorem, does not require jurors to come up with a numerical probability of guilt, but the Minnesota Supreme Court equated the Bayesian and non-Bayesian scenarios by focusing on the risk that jurors would misread the population statistics as just such a quantification.

> found at the scene of the crime. None of these techniques leads to inaccurate verdicts or calls into question the ability of the jury to make an independent decision. Nothing about the nature of litigation in general, or the criminal process in particular, makes anathema of additional information, whether or not that knowledge has numbers attached. After all, even eyewitnesses are testifying only to probabilities (though they obscure the methods by which they generate those probabilities) – often rather lower probabilities than statistical work insists on.

Id. at 1263–1264 (citations omitted).

3.2 Selection effect

Suppose that in *Williams v. State*, section 3.1.3, a fiber found on one of the bodies was compared with fibers found in the apartment of the defendant. If the defendant did not leave the fiber, the chance of matching a pre-selected fiber is 1/100. But if there are 50 distinct fibers in defendant's apartment (each with a 1/100 chance of matching), and the fiber found on the body is compared with each, the probability of one or more matches is $1-0.99^{50} = 0.395$. In *Coolidge v. State*, p. 66, since the 40 pairs of particles were apparently chosen from a larger group on the basis of visual (microscopic) similarity, the probability of a random match in the data might be much higher than the 10% rate reported for studies that did not use a visual similarity screening criterion. A great deal may thus depend on the way attempts to match characteristics are made and reported, a phenomenon sometimes referred to as "selection effect."

3.2.1 L'affaire Dreyfus

In the 1899 retrial of the 1894 secret court-martial of Alfred Dreyfus, Captain in the French General Staff, the prosecution again sought to prove that Dreyfus was the author of a handwritten *bordereau* (note) that transmitted five memoranda purportedly containing secret military information to the German ambassador in Paris. This bordereau was among a package of papers that a charwoman, in the pay of French intelligence, delivered to her employers, claiming that she discovered it (the *bordereau* torn to pieces) in the ambassador's wastebasket. The famous criminologist Alphonse Bertillon testified that there were suspicious coincidences of the initial and final letters in four of the thirteen polysyllabic words in the bordereau. Evaluating the probability of such a coincidence in a single word in normal writing as 0.2, Bertillon argued that the probability of four sets of coincidences was $0.2^4 = 0.0016$ in normal writing. This suggested to him that the handwriting of the document was not normal, which connected with a prosecution theory that Dreyfus had disguised his own handwriting to conceal his authorship. A divided military court again found Dreyfus guilty of treason, but this time with extenuating circumstances.

Questions

1. Given the assumed probability of coincidence in a single word as 0.2, exactly what probability did the expert compute?

2. If a high number of coincidences in the thirteen polysyllabic words were somehow indicative of contrived handwriting, compute a relevant probability under the assumption that the handwriting was not contrived.

Source

Laurence H. Tribe, *Trial by Mathematics: Precision and Ritual in the Legal Process*, 84 Harv. L. Rev. 1329, 1333-34 (1971); Rapport de MM. Darboux, Appell et Poincaré, in *L'affaire Dreyfus: La Révision du Procés de Rennes, Enquête* 3 at 501 (1909).

Notes

The 1899 verdict was widely recognized as a manifest injustice; there was an international outcry and Dreyfus was immediately pardoned on health grounds. The Dreyfus family eventually obtained review of the court martial by the civil court of appeals. As part of its investigation, the court requested the Academie des Sciences to appoint an expert panel to examine and report on the expert evidence. The panel–which included Henri Poincaré, a famous professor of the calculus of probabilities at the Sorbonne–pronounced it worthless. They added that "its only defense against criticism was its obscurity, even as the cuttle-fish cloaks itself in a cloud of ink in order to elude its foes." Armand Charpentier, *The Dreyfus Case* 226 (J. Lewis May translated 1935). In 1906, the court exonerated Dreyfus and annulled the 1899 verdict. In the end, having endured five years on Devil's Island, Dreyfus was restored to the army, promoted to major, and decorated with the *Legion d'Honneur*.

3.2.2 Searching DNA databases

According to the 1996 Report of the Committee on DNA of the National Research Council, in criminal investigations more than 20 suspects have already been initially identified by computerized searches through DNA databases maintained by various states. As the number and size of such databases increase, it is likely that initial identifications will more frequently be made on this basis. In its report, the Committee on Forensic DNA Science of the National Research Council stated that in such cases the usual calculation of match probability had to be modified. It recommended as one of two possibilities that the calculated match probability be multiplied by the size of the data base searched.

Questions

1. Explain the theory of such a calculation by reference to Bonferroni's inequality (see Section 3.1 at p. 57).

2. What is the probability computed with such an adjustment? Is it relevant to the identification issue?

3. What is the difference between this case and the fiber matching problem referred to in Section 3.2?

4. After reading Section 3.3 on Bayes's theorem, consider the following: Suppose that after the initial identification based on a computerized search, specific evidence is discovered that would have been sufficient to have the suspect's DNA tested if the specific evidence had been discovered first. Does it make any difference to the strength of the statistical evidence that it was used before or after the specific evidence was discovered? Suppose the specific evidence was very weak or even exculpatory so that it would not have led to a testing of the suspect's DNA. How would that affect the probative force of the statistical evidence?

5. Should the adjustment recommended by the Committee be made?

Source

National Research Council, *The Evaluation of Forensic DNA Evidence* 32 (1996). For criticisms of the Committee's position see Peter Donnelly and Richard D. Freedman, *DNA Database Searches and the Legal Consumption of Scientific Evidence*, 97 Mich. L. Rev. 931 (1999).

3.3 Bayes's theorem

Bayes's theorem is a fundamental tool of inductive inference. In science, as in law, there are competing hypotheses about the true but unknown state of nature and evidence that is more or less probable depending on the hypothesis adopted. Bayes's theorem provides a way of combining our initial views of the probabilities of the possible states of nature, with the probabilities of the evidence to arrive at posterior probabilities of the states of nature, given the evidence. It is thus a way of reasoning "backward" from effects to their causes.

In mathematical notation, Bayes's theorem shows how a set of conditional probabilities of the form $P(B_i|A_j)$ may be combined with initial or prior probabilities $P(A_i)$ to arrive at final or posterior probabilities of the form $P(A_i|B_j)$, wherein the roles of conditioning event and outcome event have

been interchanged.[6] In the case of discrete events, Bayes's theorem is easily derived. By definition, $P(A_i|B_j) = P(A_i \text{ and } B_j)/P(B_j)$. The joint probability $P(A_i \text{ and } B_j)$ may be written $P(B_j|A_i)P(A_i)$ and, similarly, the marginal probability $P(B_j)$ may be written as $P(B_j) = \sum_i P(B_j|A_i)P(A_i)$, the sum taken over all possible states of nature A_i (see Section 3.1). Thus we have

$$P(A_i|B_j) = \frac{P(B_j|A_i)P(A_i)}{\sum_i P(B_j|A_i)P(A_i)}.$$

In the case of only two states of nature, say A and not-A ($\bar{A}$), the result is:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\bar{A})P(\bar{A})}.$$

A more enlightening formulation is in terms of odds:

$$\underset{(1)}{\frac{P(A|B)}{P(\bar{A}|B)}} = \underset{(2)}{\frac{P(A)}{P(\bar{A})}} \times \underset{(3)}{\frac{P(B|A)}{P(B|\bar{A})}}.$$

In words, this says that (1) the posterior odds on the truth of state A as opposed to not-A given evidence B are equal to (2) the prior odds on A times (3) the likelihood ratio for B, i.e., the ratio of the probability of B given A and not-A. Thus, the probative force of evidence is an increasing function of both the prior odds and the likelihood ratio.

Bayes's theorem honors the Rev. Thomas Bayes (1702-61), whose result was published posthumously in 1762 by Richard Price in the Philosophical Transactions. In Bayes's paper, the prior probability distribution was the uniform distribution of a ball thrown at random on a billiard table. While Bayes's original example utilized a physical prior probability distribution, some more controversial applications of Bayes's theorem have involved subjective prior probability distributions. See Section 3.3.2.

Although prior odds are usually subjective, sometimes they are objective and can be estimated from data. An intriguing example of objective calculation of a prior probability was Hugo Steinhaus's computation for paternity cases. See Steinhaus, *The Establishment of Paternity*, Prace Wroclawskiego Towarzystwa Naukowego, ser. A., No. 32, at 5 (1954).

The background, or prior, probability computed by Steinhaus was the probability that the accused was the father after intercourse had been established, but before serological test results were known. The posterior probability was

[6]The difference between these events can be made clear from an example attributed to Keynes. If the Archbishop of Canterbury were playing in a poker game, the probability that he would deal himself a straight flush, given honest play on his part, is not the same as the probability of honest play on his part, given that he has dealt himself a straight flush. The first is 36 in 2,598,960; the second most people think would be much larger, perhaps close to 1. Would the same result apply to the cardsharping Cardinal Riario?

the probability of paternity given the test results. A significant aspect of Steinhaus's procedure was his use of population statistics to estimate the proportion of guilty fathers among those designated for the test, even though no individuals (except those subsequently exonerated by the test) could be identified as guilty or innocent. For the sake of clarifying his theory, we simplify it slightly.

Different blood types occur with different frequencies in the population. Let the type in question be called "A" and have frequency f; the frequency of those who do not have this type is $1 - f$. Consider the group of accused fathers who take the serological test because the child has blood type A, one not shared by the mother. If the mothers' accusations were always right, the serological test would show that every member of this group had type "A" blood (although the converse, of course, is not true). If the mothers' accusations were always wrong, the members of this group would constitute a random sample from the population with respect to blood type, and the expected frequency of those with blood types other than A would be $1 - f$. The disparity between the actual rate of type A blood in this accused group and the population rate measures the overall accuracy of the accusations. The higher the proportion of men with type A blood, the more correct the accusations.

Let p be the proportion of the accused group who are fathers. Then $1 - p$ is the proportion of unjustly accused men and $(1 - p)(1 - f)$ is the expected proportion of those unjustly accused whom the test will exonerate. The ratio of the expected proportion of the exonerated group to the proportion of the general population who do not have blood type A is $(1 - p)(1 - f)/(1 - f)$, or simply $1 - p$, the prior probability of a false accusation. The importance of this ratio is that both its numerator and denominator can be estimated from objective sample and population statistics.

Using the results of 1,515 Polish paternity cases in which serological tests had been administered, Steinhaus concluded that the prior probability of a true accusation was about 70 percent. (With perhaps less than complete fairness, this factor has been called "the veracity measure of women.") The 70 percent figure may be regarded as the background probability in paternity cases. It was, however, computed from a subgroup of paternity cases, including only those cases in which the child did not share the blood type of the mother, requiring a serological test to establish paternity. Nevertheless, it seems fair to test the attributes of various decision rules by this subgroup because it is probably a random sample with respect to the fact of paternity; at the very least there are not more paternities among defendants in this group than in the larger group.

3.3.1 Rogue bus

On a rainy night, a driver is forced into a collision with a parked car by a swerving bus that does not stop. There are two bus lines that travel the street: Company A has 80% of the buses; company B has 20% of the buses. Their

schedules shed no light on the culprit company. An eyewitness says it was company B's bus, but eyewitness testimony under sub-optimal conditions such as those prevailing here (rainy night, speeding bus) is known to have a high error rate.

Questions

1. In a civil suit by the injured driver against company A, is the statistical evidence that company A has 80% of the buses sufficient to satisfy plaintiff's burden of proof by a preponderance of the evidence that it was company A's bus? In the *Smith* case cited below, the court held (quoting *Sargent v. Massachusetts Accident Co*., 307 Mass. 246, 250, 29 N.E.2d 825, 827 (1940)) that the evidence must be such as to produce "actual belief" in the event by the jury, and statistical evidence could only produce probabilities, not actual belief. Do you agree that this is a valid reason for concluding that statistical evidence is per se insufficient?

2. If suit is also brought against company B, is the eyewitness's testimony sufficient to satisfy plaintiff's burden of proof that it was company B's bus?

3. If the statistical evidence is insufficient but the eyewitness testimony is sufficient, how do you reconcile those results?

4. Assume that the eyewitness testimony has a 30% error rate. Treating the statistical evidence as furnishing the prior odds, and the eyewitness testimony as supplying the likelihood ratio, use Bayes's theorem to combine the statistical and eyewitness evidence to determine the probability, given the evidence, that it was company B's bus.

Source

Cf. Smith v. Rapid Transit, Inc., 317 Mass. 469, 58 N.E.2d 754 (1945). For a discussion of some other early cases on this subject, see *Quantitative Methods in Law* 60-69.

Notes

Whether "naked" statistical evidence (i.e., statistical evidence without case-specific facts) can be sufficient proof of causation in a civil or criminal case has provoked extensive academic discussion, with the professorial verdict usually being negative, at least in criminal cases. The arguments are criticized in Daniel Shaviro, *Statistical-Probability Evidence and the Appearance of Justice*, 103 Harv. L. Rev. 530 (1989). For criminal cases, the discussion has been conducted on the level of law-school hypotheticals, because in real cases there is always case-specific evidence to supplement the statistics. For civil cases, statistics have been held to be sufficient evidence of causation when

that has seemed necessary to do justice. The outstanding example is the diethylstilbestrol (DES) litigation, in which the DES manufacturers were held proportionately liable, based on their market shares, to plaintiffs whose mothers had taken DES during pregnancy, even though there was no case-specific evidence of which company's DES the mother had taken. *Sindell v. Abbott Labs, Inc,* 26 Cal.3d 588, 607 P.2d 924 (1980), *cert. denied*, 449 U.S. 912 (1980); *Hymowitz v. Lilly & Co.,* 73 N.Y.2d 487 (1989).

3.3.2 Bayesian proof of paternity

A New Jersey statute criminalizes sexual penetration when the defendant has supervisory or disciplinary power by virtue of his "legal, professional or occupational status" and the victim is on "probation or parole or is detained in a hospital, prison or other institution." Defendant was a black male corrections officer at the Salem County jail where the female victim was incarcerated on a detainer from the Immigration and Naturalization Service. The victim conceived a child while in custody. If defendant was the father he was guilty of a crime, irrespective of the victim's consent.

In contested paternity proceedings, prior to the advent of DNA testing, the parties were frequently given Human Leukocyte Antigen (HLA) tests to identify certain gene-controlled antigens in the blood. After making HLA tests, an expert witness for the state testified that the child had a particular set of genes that was also possessed by the defendant, but not by the mother. She further testified that the frequency of this particular set of genes was 1% in the North American black male population. The expert assumed that the odds of defendant being the father, quite apart from the HLA tests, were 50-50 and, based on that assumption and the 1% frequency of the gene type, concluded that "the likelihood of this woman and this man producing this child with all of the genetic makeup versus this woman with a random male out of the black population . . . [results in] a probability of paternity of 96.55 percent."

Questions

1. Was the expert's testimony on the probability of paternity properly admitted?

2. Was the restriction to the rate of the haplotype in the black population warranted?

3. If the expert had proposed to give the jurors a hypothetical range of prior probabilities and the posterior probability associated with each prior, should her testimony have been admitted?

Source

State v. Spann, 130 N.J. 484, 617 A.2d 247 (Sup. Ct. N.J. 1993).

Notes

Whether Bayes's theorem should be explicitly used as suggested in Question 3 has been the subject of considerable academic and some judicial debate. On the academic side, among the first articles are Michael O Finkelstein & William Fairley, *A Bayesian Approach to Identification Evidence*, 83 Harv. L. Rev. 489 (1970)(proposing the use of Bayes's theorem); Laurence H. Tribe, *Trial by Mathematics:Precision and Ritual in the Legal Process*, 84 Harv. L. Rev. 1329 (1971) (criticizing the proposal); Finkelstein & Fairley, *A Comment on "Trial by Mathematics,"* 84 Harv. L. Rev. 1801 (responding to Tribe); and Tribe, *A Further Critique of Mathematical Proof*, 84 Harv. L. Rev. 1810 (1971) (rejoinder). A further critique appears in L. Brilmayer & L. Kornhauser, *Review: Quantitative Methods and Legal Decisions*, 46 U. Chi. L. Rev. 116 (1978). See generally two symposia: *Probability and Inference in the Law of Evidence*, 66 B. U. L. Rev. 377-952 (1986) and *Decision and Inference in Litigation*, 13 Cardozo L. Rev. 253-1079 (1991). On the judicial side, compare *Plemel v. Walter*, 303 Ore. 262, 735 P.2d 1209 (1987) and *State v. Spann, supra* (both approving an explicit use) with *Connecticut v. Skipper*, 228 Conn. 610, 637 A.2d 1104 (1994) (disapproving an explicit use).

Those approving an explicit use in a criminal case argue that jurors tend to underestimate the probative force of background statistical evidence. Such insensitivity to prior probability of outcomes appears to be a general phenomenon in subjective probability estimation. See, e.g., *Judgement Under Uncertainty: Heuristics and Biases*, at 4-5 (Daniel Kahneman, Paul Slovic & Amos Tversky, eds., 1982). Empirical studies based on simulated trials tend to support this. See, e.g., Jane Goodman, *Jurors' Comprehension and Assessment of Probabilistic Evidence*, 16 Am. J. Trial Advocacy 361 (1992). They also point to what is called the prosecutor's fallacy: the risk that the jury will misinterpret the low population frequency of the blood type as the probability of innocence. Those opposed to explicit use object that jurors would be invited to estimate a probability of guilt before hearing all the evidence, which they view as inconsistent with the presumption of innocence and the instruction commonly given to jurors to withhold judgement until all the evidence is heard. On the other hand, if the jurors wait until they hear all the evidence before estimating their priors, the statistics are likely to influence those estimates. Some scholars further object to any juror quantification of the probability of guilt as inconsistent with the "beyond a reasonable doubt" standard for criminal cases. Since conviction is proper despite some doubt, it is not clear why quantification of that doubt by a juror would be per se objectionable. There is some evidence that quantification of the burden of proof influences verdicts in an appropriate direction. Dorothy K. Kagehiro & W. Clark Stanton, *Legal vs. Quantified Definitions of Standards of Proof*, 9 L. & Hum. Behav. 159 (1985).

Perhaps the strongest case for an explicit use by the prosecution arises if the defense argues that the trace evidence does no more than place defendant in a group consisting of those in the source population with the trace in question.

Known as the defense fallacy, the argument assumes that without the trace defendant is no more likely to be guilty than anyone else in the source population. (This is an unlikely scenario since there is almost always other evidence that implicates the defendant.) The prosecution might then be justified in using Bayes's theorem to show what the probabilities of guilt would be if the jurors believed at least some of the other evidence. Conversely, the prosecutor's fallacy (that the frequency of the trace in the population is the probability of innocence) assumes that the prior probability of defendant's guilt is 50%. If the prosecutor makes such an argument, the defense should then be justified, using Bayes's theorem, to demonstrate what the probabilities would be if some or all of the other evidence were disbelieved.

Another set of issues is presented if identifying the source of the trace does not necessarily imply guilt. A thumb print on a kitchen knife, used as a murder weapon, may have been left there innocently. The complication here is that the same facts suggesting guilt that are used to form the prior probability of authorship of the print would also be used to draw an inference from authorship of the print to guilt. If this is an impermissible double use, it would be hard or impossible to partition the non-statistical evidence among uses.

Whether an explicit use of Bayes's theorem is allowed in the courtroom may stir legal academics more than jurors. In one empirical study the jurors simply disregarded the expert's Bayesian explanations of the statistics. See David L. Faigman & A. J. Baglioni, Jr., *Bayes' Theorem in the Trial Process*: *Instructing Jurors on the Value of Statistical Evidence*, 12 Law & Hum. Behav. 1 (1988). The more important (and often ignored) teaching of Bayes's theorem is that one need not assert that a matching trace is unique or nearly unique in a suspect population to justify its admission as powerful evidence of guilt.

3.4 Screening devices and diagnostic tests

Screening devices and diagnostic tests are procedures used to classify individuals into two or more groups, utilizing some observable characteristic or set of characteristics. Most familiar examples come from medical diagnosis of patients as "affected" or "not affected " by some disease. For our discussion we adopt the clinical paradigm, but the central ideas are by no means limited to that context.

False positives and negatives

No diagnostic test or screening device is perfect. Errors of omission and commission occur, so we need to distinguish between the *true* status (say, A = affected or U = unaffected) and the *apparent* status based on the test (say, + = test positive or − = test negative). A *false positive* diagnosis is the occurrence of a positive outcome (+) in an unaffected person (U); it is denoted