Reference Manual on Scientific Evidence

Second Edition

Federal Judicial Center 2000

Summary Table of Contents

A detailed Table of Contents appears at the front of each chapter

Multiple authors of independent chapters, so no guarantee of uniform thought.

Extracts from two chapters.

D. Posterior Probabilities

Standard statistics

Standard errors, *p*-values, and significance tests are common techniques for assessing random error. These procedures rely on the sample data, and are justified in terms of the "operating characteristics" of the statistical procedures.[164] However, this frequentist approach does not permit the statistician to compute the probability that a particular hypothesis is correct, given the data.[165] For instance, a frequentist may postulate that a coin is fair: it has a 50-50 chance of landing heads, and successive tosses are independent; this is viewed as an empirical statement—potentially falsifiable—about the coin. On this basis, it is easy to calculate the chance that the coin will turn up heads in the next ten tosses:[166] the answer is 1/1,024. Therefore, observing ten heads in a row brings into serious question the initial hypothesis of fairness. Rejecting the hypothesis of fairness when there are ten heads in ten tosses gives the wrong result—when the coin is fair—only one time in 1,024. That is an example of an operating characteristic of a statistical procedure.

Bayesian statistic

But what of the converse probability: if a coin lands heads ten times in a row, what is the chance that it is fair?[167] To compute such converse probabilities, it is necessary to postulate initial probabilities that the coin is fair, as well as probabilities of unfairness to various degrees.[168] And that is beyond the scope of frequentist statistics.[169]

164. "Operating characteristics" are the expected value and standard error of estimators, probabilities of error for statistical tests, and related quantities.

165. *See supra* § IV.B.1; *infra* Appendix. Consequently, quantities such as *p*-values or confidence levels cannot be compared directly to numbers like .95 or .50 that might be thought to quantify the burden of persuasion in criminal or civil cases. *See* Kaye, *supra* note 144; D.H. Kaye, *Apples and Oranges: Confidence Coefficients and the Burden of Persuasion*, 73 Cornell L. Rev. 54 (1987).

166. Stated slightly more formally, if the coin is fair and each outcome is independent (the hypothesis), then the probability of observing ten heads (the data) is $Pr(data | H_0) = (1/2)^{10} = 1/1{,}024$, where H_0 stands for the hypothesis that the coin is fair.

167. We call this a "converse probability" because it is of the form $Pr(H_0 | data)$ rather than $Pr(data | H_0)$; an equivalent phrase, "inverse probability," also is used. The tendency to think of $Pr(data | H_0)$ as if it were the converse probability $Pr(H_0 | data)$ is the "transposition fallacy." For instance, most United States senators are men, but very few men are senators. Consequently, there is a high probability that an individual who is a senator is a man, but the probability that an individual who is a man is a senator is practically zero. For examples of the transposition fallacy in court opinions, see cases cited *supra* note 142. *See also* Committee on DNA Forensic Science: An Update, *supra* note 60, at 133 (describing the fallacy in cases involving DNA identification evidence as the "prosecutor's fallacy"). The frequentist *p*-value, $Pr(data | H_0)$, is generally not a good approximation to the Bayesian $Pr(H_0 | data)$; the latter includes considerations of power and base rates.

168. *See infra* Appendix.

169. In some situations, the probability of an event on which a case depends can be computed with objective methods. However, these events are measurable outcomes (like the number of heads in a series of tosses of a coin) rather than hypotheses about the process that generated the data (like the claim that the coin is fair). For example, in *United States v. Shonubi*, 895 F. Supp. 460 (E.D.N.Y. 1995), *rev'd*,

In the Bayesian or subjectivist approach, probabilities represent subjective degrees of belief rather than objective facts. The observer's confidence in the hypothesis that a coin is fair, for example, is expressed as a number between zero and one;[170] likewise, the observer must quantify beliefs about the chance that the coin is unfair to various degrees—all in advance of seeing the data.[171] These subjective probabilities, like the probabilities governing the tosses of the coin, are set up to obey the axioms of probability theory. The probabilities for the various hypotheses about the coin, specified before data collection, are called prior probabilities.

These prior probabilities can then be updated, using "Bayes' rule," given data on how the coin actually falls.[172] In short, Bayesian statisticians can compute posterior probabilities for various hypotheses about the coin, given the data.[173] Although such posterior probabilities can pertain directly to hypotheses of legal interest, they are necessarily subjective, for they reflect not just the data but also

The worry is subjectivity.

103 F.3d 1085 (2d Cir. 1997), a government expert estimated for sentencing purposes the total quantity of heroin that a Nigerian defendant living in New Jersey had smuggled (by swallowing heroin-filled balloons) in the course of eight trips to and from Nigeria. He applied a method known as "resampling" or "bootstrapping." Specifically, he drew 100,000 independent simple random samples of size seven from a population of weights distributed as in customs data on 117 other balloon swallowers caught in the same airport during the same time period; he discovered that for 99% of these samples, the total weight was at least 2090.2 grams. 895 F. Supp. at 504. Thus, the researcher reported that "there is a 99% chance that Shonubi carried at least 2090.2 grams of heroin on the seven [prior] trips . . . ." *Id*. However, the Second Circuit reversed this finding for want of "specific evidence of what Shonubi had done." 103 F.3d at 1090. Although the logical basis for this "specific evidence" requirement is unclear, a difficulty with the expert's analysis is apparent. Statistical inference generally involves an extrapolation from the units sampled to the population of all units. Thus, the sample needs to be representative. In *Shonubi*, the government used a sample of weights, one for each courier on the trip at which that courier was caught. It sought to extrapolate from these data to many trips taken by a single courier—trips on which that other courier was not caught.

170. Here "confidence" has the meaning ordinarily ascribed to it rather than the technical interpretation applicable to a frequentist "confidence interval." Consequently, it can be related to the burden of persuasion. *See* Kaye, *supra* note 165.

171. For instance, let p be the unknown probability that coin lands heads: What is the chance that p exceeds .6? The Bayesian statistician must be prepared to answer all such questions. Bayesian procedures are sometimes defended on the ground that the beliefs of any rational observer must conform to the Bayesian rules. However, the definition of "rational" is purely formal. *See* Peter C. Fishburn, *The Axioms of Subjective Probability*, 1 Stat. Sci. 335 (1986); David Kaye, *The Laws of Probability and the Law of the Land*, 47 U. Chi. L. Rev. 34 (1979).

172. *See infra* Appendix.

173. *See generally* George E.P. Box & George C. Tiao, Bayesian Inference in Statistical Analysis (Wiley Classics Library ed., John Wiley & Sons, Inc. 1992) (1973). For applications to legal issues, see, e.g., Aitken et al., *supra* note 45, at 337–48; David H. Kaye, *DNA Evidence: Probability, Population Genetics, and the Courts*, 7 Harv. J.L. & Tech. 101 (1993).

the subjective prior probabilities—that is, the degrees of belief about the various hypotheses concerning the coin specified prior to obtaining the data.[174]

Such analyses have rarely been used in court,[175] and the question of their forensic value has been aired primarily in the academic literature.[176] Some statisticians favor Bayesian methods,[177] and some legal commentators have proposed their use in certain kinds of cases in certain circumstances.[178]

Very guarded summary.

V. Correlation and Regression

Regression models are often used to infer causation from association; for example, such models are frequently introduced to prove disparate treatment in discrimination cases, or to estimate damages in antitrust actions. Section V.D explains the ideas and some of the pitfalls. Sections V.A–C cover some preliminary material, showing how scatter diagrams, correlation coefficients, and regression lines can be used to summarize relationships between variables.

174. In this framework, the question arises of whose beliefs to use—the statistician's or the factfinder's. *See, e.g.*, Michael O. Finkelstein & William B. Fairley, *A Bayesian Approach to Identification Evidence*, 83 Harv. L. Rev. 489 (1970) (proposing that experts give posterior probabilities for a wide range of prior probabilities, to allow jurors to use their own prior probabilities or just to judge the impact of the data on possible values of the prior probabilities). *But see* Laurence H. Tribe, *Trial by Mathematics: Precision and Ritual in the Legal Process*, 84 Harv. L. Rev. 1329 (1971) (arguing that efforts to describe the impact of evidence on a juror's subjective probabilities would unduly impress jurors and undermine the presumption of innocence and other legal values).

Dueling references, one for Bayes, one against.

175. The exception is paternity litigation; when genetic tests are indicative of paternity, testimony as to a posterior "probability of paternity" is common. *See, e.g.*, 1 Modern Scientific Evidence: The Law and Science of Expert Testimony, *supra* note 3, § 19–2.5.

176. *See, e.g.*, Probability and Inference in the Law of Evidence: The Uses and Limits of Bayesianism (Peter Tillers & Eric D. Green eds., 1988); Symposium, *Decision and Inference in Litigation*, 13 Cardozo L. Rev. 253 (1991). The Bayesian framework probably has received more acceptance in explicating legal concepts such as the relevance of evidence, the nature of prejudicial evidence, probative value, and burdens of persuasion. *See, e.g.*, Richard D. Friedman, *Assessing Evidence*, 94 Mich. L. Rev. 1810 (1996) (book review); Richard O. Lempert, *Modeling Relevance*, 75 Mich. L. Rev. 1021 (1977); D.H. Kaye, *Clarifying the Burden of Persuasion: What Bayesian Decision Rules Do and Do Not Do*, 3 Int'l J. Evidence & Proof 1 (1999).

177. *E.g.*, Donald A. Berry, *Inferences Using DNA Profiling in Forensic Identification and Paternity Cases*, 6 Stat. Sci. 175, 180 (1991); Stephen E. Fienberg & Mark J. Schervish, *The Relevance of Bayesian Inference for the Presentation of Statistical Evidence and for Legal Decisionmaking*, 66 B.U. L. Rev. 771 (1986). Nevertheless, many statisticians question the general applicability of Bayesian techniques: The results of the analysis may be substantially influenced by the prior probabilities, which in turn may be quite arbitrary. *See, e.g.*, Freedman, *supra* note 112.

178. *E.g.,* Joseph C. Bright, Jr. et al., *Statistical Sampling in Tax Audits*, 13 L. & Soc. Inquiry 305 (1988); Ira Mark Ellman & David Kaye, *Probabilities and Proof: Can HLA and Blood Group Testing Prove Paternity?*, 54 N.Y.U. L. Rev. 1131 (1979); Finkelstein & Fairley, *supra* note 174; Kaye, *supra* note 173.

Notwithstanding the lack of adequate empirical research, other commentators believe that the danger of prejudice (in the form of the transposition fallacy) warrants the exclusion of likelihood ratios.[261]

Bayes assessed again concerning DNA evidence.

3. Should Posterior Probabilities Be Excluded?

Match probabilities state the chance that certain genotypes would be present conditioned on specific hypotheses about the source of the DNA (a specified relative, or an unrelated individual in a population or subpopulation). Likelihood ratios express the relative support that the presence of the genotypes in the defendant gives to these hypotheses compared to the claim that the defendant is the source. Posterior probabilities or odds express the chance that the defendant is the source (conditioned on various assumptions). These probabilities, if they are meaningful and accurate, would be of great value to the jury.

"…if they are meaningful…"

Experts have been heard to testify to posterior probabilities. In *Smith v. Deppish*,[262] for example, the state's "DNA experts informed the jury that . . . there was more than a 99 percent probability that Smith was a contributor of the semen,"[263] but how such numbers are obtained is not apparent. If they are instances of the transposition fallacy, then they are scientifically invalid (and objectionable under Rule 702) and unfairly prejudicial (under Rule 403).

However, a meaningful posterior probability can be computed with Bayes' theorem.[264] Ideally, one would enumerate every person in the suspect population, specify the prior odds that each is the source of the forensic DNA and weight those prior odds by the likelihoods (taking into account the familial relationship of each possible suspect to the defendant) to arrive at the posterior odds that the defendant is the source of the forensic sample. But this hardly seems practical. The 1996 NRC Report therefore discusses a somewhat different implementation of Bayes' theorem. Assuming that the hypotheses of kinship and error could be dismissed on the basis of other evidence, the report focuses on "the variable-prior-odds method," by which:

Determining appropriate priors "hardly seems practical."

> an expert neither uses his or her own prior odds nor demands that jurors formulate their prior odds for substitution into Bayes's rule. Rather, the expert presents the jury with a

"an expert neither uses his or her own prior odds nor demands that jurors formulate their prior odds for substitution into Bayes's rule. Rather, the expert presents the jury with a table or graph showing how the posterior probability changes as a function of the prior probability"

261. *See* Koehler, *supra* note 168, at 880; Thompson, *supra* note 168, at 850; *cf.* Koehler et al., *supra* note 246 (proposing the use of a likelihood ratio that incorporates laboratory error).

262. 807 P.2d 144 (Kan. 1991).

263. *See also* Thomas v. State, 830 S.W.2d 546, 550 (Mo. Ct. App. 1992) (a geneticist testified that "the likelihood that the DNA found in Marion's panties came from the defendant was higher than 99.99%"); Commonwealth v. Crews, 640 A.2d 395, 402 (Pa. 1994) (an FBI examiner who at a preliminary hearing had estimated a coincidental-match probability for a VNTR match "at three of four loci" reported at trial that the match made identity "more probable than not").

264. *See supra* § VII.C.2.

> table or graph showing how the posterior probability changes as a function of the prior probability.[265]

This procedure, it observes, "has garnered the most support among legal scholars and is used in some civil cases."[266] Nevertheless, "very few courts have considered its merits in criminal cases."[267] In the end, the report concludes:

> How much it would contribute to jury comprehension remains an open question, especially considering the fact that for most DNA evidence, computed values of the likelihood ratio (conditioned on the assumption that the reported match is a true match) would swamp any plausible prior probability and result in a graph or table that would show a posterior probability approaching 1 except for very tiny prior probabilities.[268]

E. Which Verbal Expressions of Probative Value Should Be Presented?

Having surveyed various views about the admissibility of the probabilities and statistics indicative of the probative value of DNA evidence, we turn to a related issue that can arise under Rules 702 and 403: Should an expert be permitted to offer a non-numerical judgment about the DNA profiles?

Inasmuch as most forms of expert testimony involve qualitative rather than quantitative testimony, this may seem an odd question. Yet, many courts have held that a DNA match is inadmissible unless the expert attaches a scientifically valid number to the figure.[269] In reaching this result, some courts cite the statement in the 1992 NRC report that "[t]o say that two patterns match, without providing any scientifically valid estimate (or, at least, an upper bound) of the frequency with which such matches might occur by chance, is meaningless."[270]

"…likelihood ratio … would swamp any plausible prior probability and result in a graph or table that would show a posterior probability approaching 1 except for very tiny prior probabilities."

Footnote 268:
"Id. For arguments said to show that the variable-prior-odds proposal is "a bad idea," see Thompson, supra note 69, at 422–23."

265. NRC II, *supra* note 1, at 202 (footnote omitted).
266. *Id.*
267. *Id.* (footnote omitted).
268. *Id.* For arguments said to show that the variable-prior-odds proposal is "a bad idea," see Thompson, *supra* note 69, at 422–23.
269. *E.g.*, Commonwealth v. Daggett, 622 N.E.2d 272, 275 n.4 (Mass. 1993) (plurality opinion insisting that "[t]he point is not that this court should require a numerical frequency, but that the scientific community clearly does"); State v. Carter, 524 N.W.2d 763, 783 (Neb. 1994) ("evidence of a DNA match will not be admissible if it has not been accompanied by statistical probability evidence that has been calculated from a generally accepted method"); State v. Cauthron, 846 P.2d 502 (Wash. 1993) ("probability statistics" must accompany testimony of a match); *cf.* Commonwealth v. Crews, 640 A.2d 395, 402 (Pa. 1994) ("The factual evidence of the physical testing of the DNA samples and the matching alleles, even without statistical conclusions, tended to make appellant's presence more likely than it would have been without the evidence, and was therefore relevant.").
270. NRC I, *supra* note 1, at 74. For criticism of this statement, see Kaye, *supra* note 195, at 381–82 (footnote omitted):

> [I]t would not be 'meaningless' to inform the jury that two samples match and that this match makes it more probable, in an amount that is not precisely known, that the DNA in the samples comes from the same person. Nor, when all estimates of the frequency are in the millionths or billionths, would it be meaningless