NOTES ABOUT STATISTICAL EVIDENCE

Notas sobre a prova estatística

Michele Taruffo
Catedra de Cultura Jurídica, Universitat de Girona, Espanha. Università degli Studi di Pavia, Itália.
michelino.taruffo@unipv.it

Autor convidado.

ABSTRACT

The essay deals with some of the problems concerning the use of statistics in evidentiary inferences. Limits and conditions of such a use are explained, with reference to the modern theory of evidence and proof.

Keywords: Evidence Law, Legal Proof, Statistics.

Resumo

Este artigo aborda alguns dos problemas concernentes ao uso da estatística em inferências probatórias. Os limites e as condições de tal uso são explicadas, com referência à teoria moderna da prova.

Palavras-chave: Direito Probatório, Prova Legal, Estatística.

1. INFERENCES AND DECISION

These notes do not deal with the whole complex problem of "forensic statistics", on which there is a relevant amount of literature[1]. Rather, they are aimed at stressing some doubtful and disputable topics that should be considered when the use of statistics in judicial processes is analyzed.

A first set of relevant questions arises when one deals with the basic problem of any judicial decision, that is: the inference – or, more frequently – the set of inferences that are based upon the relevant evidence at hand and the conclusion concerning the truth or

1 See mainly FROSINI, 2002.

falseness of the statements concerning the facts in issue. However, before dealing with such basic problems a couple of premises should be clarified. One of these premises is that the decision about those facts is not conceived as an irrational and merely subjective spiritual act, as it is with the *intime conviction* of the French criminal system[2], but – rather – as a reasoning made according with rational models and logical principles. Then it may be interpreted as a set of logical inferences[3]. The second prem is that when one speaks of *judicial truth* there is no reference to any idea of "absolute" truth, that is of something that cannot be achieved in any judicial process (as well as in other areas of human experience and even of science). Then the judicial truth has to be intended as a *relative* truth, not in the subjective sense that each individual has his own personal truth, but in the objective sense that any conclusion about the facts in issue is based upon the evidence that is available in each specific case. Then such a truth is a *matter of degree* depending on such evidence, or – as it may be said – a matter of *approximation* to the unattainable "true" truth of those facts. That's why this idea may also be expressed in terms of *probable truth* or of *probability of truthfulness* of the factual statements[4].

On the basis of these premises the fundamental problem may be stated in these terms: whether, and if yes in which way, the reference to statistics may be used as evidence concerning a specific individual fact (for instance: a particular causal connection), that is usually the subject matter of a judicial case[5]. Of course this problem arises because the common opinion is that statistics refer to a frequency of a situation within a given relevant population or set of cases and are useful for predictions, although predicting a specific fact is extremely difficult[6], but they cannot say anything about individual past events[7]. Then the problem may be interpreted as dealing with the structure of the legal connection between statistics and conclusions concerning the specific fact in issue.

2 See TARUFFO, 2009, p. 87, 161 and 244.

3 See TARUFFO, 2009, p. 207.

4 On these complex topics see again TARUFFO, 2009, p. 74 and 90.

5 But for different situations see infra, 3.

6 See FROSINI, 2002, p. 22.

7 See FROSINI, 2002, p. 9 and 149.

1.1 Naked Statistical Evidence?

The problem of the inferences connecting rationally the evidence at hand with a conclusion concerning the facts in issue may be interpreted – and actually is interpreted – on the basis of various conceptual models. A complete analysis of all these models cannot be made here, then a specific attention will be paid only to those models in which statistics may be or are actually used.

The judge does not play with paradoxes: he has to deal with specific and concrete empirical facts that occurred in the past.

First of all, however, one of these models may be set aside immediately, that is: the theory according to which the so-called *naked statistical evidence* may support a conclusion about the facts in issue even when there is no other evidence. It is a well-known and disputed theory[8], but a full discussion of it is not relevant here, because of at least two reasons. One is that in the administration of justice there is no interest in paradoxes as those of the blue bus or the public of a rodeo[9]. The judge does not play with paradoxes: he has to deal with specific and concrete empirical facts that occurred in the past. Another reason is that – as it is commonly said – statistics have nothing to say about specific past facts, since they deal with populations or sets of events and – moreover – are oriented towards the future rather than towards the past[10]. This does not prevent, of course, the reference to statistics in the analysis of the evidence, but it shows that *naked* statistics cannot be taken as an autonomous and sufficient item of evidence.

1.2 Bayesian Probability?

Among the approaches to the problem of judicial decision on facts there is a very well-known theory that has been developed in the last decades mainly – but not only[11] – in the United States, according to which the evidentiary inferences should be made in terms of *quantitative probability,* and mainly by applying the so-called *Bayes' theorem.* In this way, it is said, the judge may arrive at a conclusion in terms of a percent degree of probability of the statement concerning the fact in issue. The problem that would be solved in this way is to establish which one, among various hypotheses (or among several

8 See FROSINI, 2002, p. 12 and 65.

9 About these paradoxes see FROSINI, 2002, p. 86, also for other references.

10 See e.g. FROSINI, 2002, p.17.

11 In the Italian literature see mainly GARBOLINO. See also FROSINI, 2002, p. 45; TARUFFO, 1992, p. 169.

causes) is most probably connected with this statement. In this model, the function of statistics is to provide *a priori* probabilities concerning the different hypotheses (or causes) about an event, and the theorem permits to determine the *a posteriori* probability of this event in connection with the relatively best hypothesis or cause[12].

There is no doubt about the mathematical validity of the Bayes' theorem that actually is used in a broad variety of contexts. The problem, however, is to understand whether the Bayesian calculus may be taken – as its supporters say – as the general model of the judge's reasoning about the evidence. If the question is stated in such general terms, the answer cannot be but negative, for several reasons. One of these reasons is that only sometimes – but for sure not always – the judge has to decide which is the more probable evidence or the more probable cause of an event. When, as it more often happens – the judge has just to establish whether a fact occurred or did not occur on the basis of some items of evidence at hand, there is no problem about alternative hypotheses or various possible causes of that fact. The problem may simply be to establish whether a fact X (the only possible cause) provoked or did not provoke the effect Y (i.e. the fact in issue). In all these cases, it seems that there is no possible use of the Bayes' theorem[13].

But even assuming that the theorem could be hypothetically applied, its use may be impossible. Here the problem is that in many – or most – cases in a judicial context there are no statistics to be used as probabilities *a priori*, because of the simple fact that such information is lacking. Then the Bayesian calculus cannot be performed[14]. In order to save this possibility even when an objective statistical information is not at hand, it is sometimes said that one should use *subjective* evaluations of the frequency of a given hypothesis or cause, and such estimates should be taken as probabilities *a priori* upon which the calculus could be based[15]. But even admitting hypothetically that in some contexts such a subjective version of probability may be accepted, this is not the case of judicial contexts. On the one hand, in fact, the decision has to be made by the judge on the basis of *objective*

12 See e.g. GARBOLINO, 2014, p. 88; FROSINI, 2002, p. 49.

13 See FROSINI, 2002, p. 99.

14 More broadly see TARUFFO, 1992, p. 175.

15 See e.g. GARBOLINO, 2014, p. 92 and 95; FROSINI, 2002, p. 53 and 98.

items of evidence assessed in a rational and controllable way, not on the basis of the judge's subjective biases and personal evaluations of any merely supposed frequency of any event. On the other hand, it is well known that one of the most frequent *fallacies* in the common reasoning is just that of quantifying arbitrarily and subjectively any supposed frequency of anything[16]. In a word: inventing nonexistent and fictitious statistics is a very bad kind of reasoning.

Moreover, one of the last supporters of the Bayesian method for judicial decisions underlines that up to date the development of the theory has reached only the situation in which there is just one specific item of evidence, but that the theory is not yet applicable when there are several items of evidence[17]. Unfortunately for the theory, however, the reality of judicial processes is that usually there are several – and in some cases many – different items of evidence, with the further problem that some of them are favorable and some are contrary to a specific conclusion about the facts in issue. This means that even if one accepts the Bayesian theory, he cannot apply it in a large number of cases: therefore, once again it cannot be taken as a general model for judicial decisions.

A more positive and fruitful approach to the problem of the judicial use of statistics requires a due consideration of the inferences by which evidence is connected to a conclusion concerning the facts in issue.

2. INFERENTIAL MODELS

A more positive and fruitful approach to the problem of the judicial use of statistics requires a due consideration of the inferences by which evidence is connected to a conclusion concerning the facts in issue. The set of such inferences is sometimes very complex and may be analyzed by means of different logical models. However, such an analysis cannot be properly developed here. Therefore, for sake of simplicity, attention may be paid just to two of the main models of inferences in which statistics may be involved.

One of these models is the Hempel's model of a *nomological-deductive* inference. It is relevant in the theory of judicial decision mainly because of the reference made by Federico Stella[18]. This kind of inference is so called because it connects a premise with a conclusion on the basis of a *general covering law*, and therefore

16 See NISBETT and ROSS, 1989, p. 55, 123, 159 and 228; TARUFFO, 1992, p. 176.

17 See GARBOLINO, 2014, p. 310.

18 See STELLA, 2003, p. 239. See also TARUFFO, 2012, p. 226 and FROSINI, 2006, p. 307.

the conclusion is certain in a *deductive* way. So far, however, we are simply dealing with a modern version of the Aristotelian syllogism, and there is no problem of statistics. The problem arises when the reference is made to a *quasi* nomological-deductive model, which is to a *probabilistic* version of the original model. It happens when there is not a covering general law, but there is a *statistical frequency* of the connection between premise and conclusion, and such a frequency has an especially high value (of 90% or even more). In such a case it is said that the conclusion may be considered as *practically certain,* since its truth is highly probable. There are, however, some criticisms that can be addressed to this theory. On the one hand, it may be said that it does not represent what normally happens in judicial contexts, where the reference to general laws, but also to very high probabilities, is not impossible but is not frequent. Then this model cannot be taken as a general model of judicial inferences.

On the other hand, the role of statistics in such a model deserves to be properly defined. It seems clear that if there are statistics suggesting that A provokes B in 95% of the cases, it provides a good reasonable justification for believing that most probably A provoked B also in the specific case. But it would be incorrect to say that in such a case the occurrence of B has a 95% probability, since statistics provide frequencies but do not say anything about a specific instance. Rather, it could be said that in such a case the statistical frequency offers a nice justification for a *practical decision*. In other words, a judge would be reasonably justified in taking the conclusion of the inference *as if* such a statement were highly probable, and to behave *as if* it were true. However, it is said that to speak of practical certainty is possible only when the probability is higher than 99%[19].

It seems, therefore, that the reference to statistics may play a relevant role in providing a rational justification for judicial decisions, but here a further problem arises. Actually such a justification may be rational when the probability at stake is very high (that is when the statistical information is *quasi general*) mainly because in such cases the rate of error is very low, and then the probability of a wrong decision is also very low, or at any rate *tolerable*. But what about the much more frequent case in which the statistical frequency is

19 See FROSINI, 2002, p. 130.

lower (for instance, of 80% or 70%), and then taking the conclusion of the inference "as if it were true" has a much higher probability to be wrong? Moreover: what about the case in which the statistical frequency is low or very low (for instance 30% or 20%), with the corresponding high probability of error concerning the conclusion[20]?

In a sense, and in general, it could be roughly said that the *degree of confidence* (or of *belief*) in the truthfulness of the conclusion depends on the degree of probability of the statistics that are used as the basis of the inference. Then, if this degree is low, the conclusion of the inference cannot be taken as reliable, and the inference cannot be taken as an independent and sufficient item of evidence.

Here the main problem is to establish whether or not several statistics, each one not sufficient per se to support a conclusion, may be *combined* in a coherent set of inferences, so that a conclusion about a causal connection may be finally justified.

This does not mean that only statistics with high probabilities should be used, since also low probabilities may be useful[21]. However, an important aspect of the problem is whether and when statistics may or may not be sufficient to achieve the standard of proof that is required in each specific case[22], although it may be admitted that even "low" statistics may be sometimes relevant as a support of a conclusion about the facts in issue[23]. Here the main problem is to establish whether or not several statistics, each one not sufficient *per se* to support a conclusion, may be *combined* in a coherent set of inferences, so that a conclusion about a causal connection may be finally justified[24].

Similar remarks can be made about another logical model of inference that is used in epistemology and also in the analysis of judicial decisions[25]. It is the type of inference studied by Stephen Toulmin[26], according to which a hypothesis H may be supported by an evidence E on the basis of a warrant W. This type of inference may be considered as the simple and atomic element that – in various combinations – forms complex sets of inferences, in various contexts and even in the judicial reasoning about evidence and factual hypotheses[27]. It is clear that the most important

20 About the medium-low frequencies see STELLA, 2003, p. 350.

21 But see *infra*, 3.1., and STELLA, 2003, p. 350.

22 About the standards of proof see *infra*, 4.

23 See STELLA, 2003, p. 350.

24 For a positive answer see HAACK, 2014, p. 209, 218, 222, 225.

25 See e.g. HAACK, 2003, p. 60; GONZÁLEZ LAGIER, 2013, 55; TARUFFO, 2009, p. 207.

26 See TOULMIN, 2008, p. 91.

27 See mainly TARUFFO, 2009, p. 209.

factor of the inference is the warrant, since the conclusion about H depends essentially on it[28]. When the W is a general law, once again we have a version of the syllogism and the conclusion is that H is deductively certain. But at least two problems arise when W is something different.

One of these problems deals with the case in which W is a *generalization.* If, following Frederick Schauer, there is a generalization that corresponds with a universal law[29], once again the inference has a deductive character. In other cases, there may be a generalization that is not properly general but that corresponds to what is considered as "normal", more frequent and familiar, in the given situation. Here the inference is not deductive but if such generalizations are used as W in the Toulmin model, the conclusion may be that H is reasonably justified[30]. But of course this does not happen when the generalization is *spurious,* being devoid of any epistemic meaning and of any correspondence with the reality[31]. In such a case, of course, this kind of warrant cannot give any outcome in terms of rational justification of H. Then the problem is that in each case in which *common sense generalization* (corresponding to the Italian notion of *massime d'esperienza*)[32] are used as W, their nature and meaning, and their epistemic foundation, has to be carefully checked and eventually denied, since – as William Twining writes[33] – the "stock of knowledge" that is typical of common sense is an "ill-defined agglomerate of beliefs", that is a "complex soup of more or less well-grounded information, sophisticated models, anecdotal memories, impressions, stories, myths, proverbs, wishes, stereotypes, speculations and prejudices".

The situation may be less complex and uncertain when there are statistics that may be used as W, but here again a relevant problem arises. As it was said above, a statistical frequency with a high (how much?) probability may provide a good *practical* reason to take H as confirmed "as if it were true", and then it may be used as a W. But

28 See e.g. TARUFFO, 2009, p. 208.

29 See SCHAUER, 2003, p. 7.

30 See SCHAUER, 2003, p. 7, 10, 55 and 108.

31 See SCHAUER, 2003, p. 7, 12, 17, 137 and 152.

32 About such a notion, which is common also to other procedural cultures, see TARUFFO, 2012, p. 225; TARUFFO, 2009, p. 210.

33 See TWINING, 2006, p. 338.

then an open question deals once more with frequencies lower than "quasi 100%", and also, in general, with low frequencies. In these cases, in fact, the best that can be said is that such statistics provide a low level of W, and then they do not offer a sufficient confirmation – if taken alone – of H.

3. SCIENTIFIC EVIDENCE

The domain in which the use of statistics is by far more frequent is that of the so-called scientific evidence.

The domain in which the use of statistics is by far more frequent is that of the so-called *scientific evidence*. Actually in the last decades, and mainly after the *Daubert* decision issued in 1993 by the Supreme Court of the United States[34], this kind of evidence as become almost usual in the civil and criminal justice of several countries. The need for the trier of fact (judge or juror) to use scientific knowledge that does not belong to his culture, since he is only an "average man" from this point of view, has become a daily duty when the case needs to be properly and carefully decided with reference to the facts in issue. On the other hand, the great development of many branches of science, mainly – but not only – in various areas of medicine and genetics, provides a wide array of techniques that may be useful, and then should be used, in order to make such a decision.

The literature concerning such a complex and evolving phenomenon is immense, in the United States as well as in other countries, but a fair overview may be achieved by looking at the 1.000 pages of a basic text in which several essays are collected about the main examples of scientific evidence and some of the problems they raise. It is the *Reference Manual on Scientific Evidence.* It includes essays concerning several types of scientific evidence, such as DNA test[35], exposure science[36], epidemiology[37], toxicology[38], neurosciences[39] and mental health[40], but also about evidence of economic damages[41]

34 The literature about Daubert and its effect is immense and includes thousands of essays and dozens of books. In the Italian literature see e.g. STELLA, 2003, p. 458; FROSINI, 2002, p. 41; DOMINIONI, 2005, p. 137; TARUFFO, 1996; DONDI, 1996.

35 KAYE and SENSABAUG, 2011, p. 129.

36 RODRIKS, 2011, p. 503.

37 GREEN, FRIEDMAN and GORDIS, 2011, p. 549.

38 GOLDSTEIN and HENIFIN, 2011, p. 633.

39 GREELY and WAGNER, 2011, p. 747.

40 APPELBAUM, 2011, p. 813.

41 ALLEN, HALL and LAZEAR, 2011, p. 426.

and engineering[42]. Moreover, there are also essays concerning the use of statistics[43] and multiple regressions[44], considering the great importance of such calculations in all the areas of scientific evidence. However, in the modern idea of science other areas of knowledge that are not considered in the *Manual* are included, such as psychology, sociology, economy, anthropology, and so forth.

Considering the variety and the dimensions of what we now call scientific evidence, it is clearly impossible to try to develop here a detailed analysis of the role that statistics have in all these different domains. However, some general aspects of the phenomenon deserve to be stressed.

A first important aspect is that each science – hard or human – and even each specific branch of every science, has its own particular paradigms, protocols and methods of inquiry, with the obvious consequence that they provide different kinds of knowledge, with different degrees of epistemic support, and then different kinds of evidence. There is no unique and common scientific method, and therefore there is no unique idea or theory of scientific knowledge: correspondingly, even the idea of scientific evidence is fragmented into a lot of different kinds of information. This is the fundamental problem of scientific evidence, at least after *Daubert*, considering that – according with the principles affirmed in this judgment – any item of evidence is admitted into a judicial process only if it is scientifically valid and fits with the particular facts in issue as a possibly useful information about them. When statistics are used as an evidence, or as a part of the inquiry leading to any kind of scientific evidence, they must be reliable, properly calculated, correctly interpreted and carefully analyzed from the point of view of their relevance for a decision about the facts in issue[45].

The necessary reference to the facts of the case leads to another important problem, which deals with the evidentiary use of statistics. A good example of this problem is the use of statistics in epidemiology, mainly when the decision deals with the causation

42 ROBERTSON, MOALLI and BLACK, 2011, p. 897.

43 KAYE and FREEDMAN, 2011, p. 211.

44 RUBINFELD, 2011, p. 303.

45 See mainly KAYE and FREEDMAN, 2011, p. 216, 230, 240 and 260, but these aspects are analyzed in all the specific essays included in the Manual.

between facts. In most cases, actually, a judicial case deals with a specific and individual factual situation, such as: "at the time t in the place p the fact f provoked the effect e". Then this would be a case of *individual* or *specific causation* in which a particular event is taken as the cause of a particular consequence. But epidemiology deals essentially with the so-called *general causation,* that is with the statistical frequency of the occurrence of a "type" of event (that is "facts of the kind A provoke the effect of the type E in the X% of cases representing the relevant population")[46]. Then the problem is that (except what we shall see later….) an epidemiological statistical frequency has nothing to say about the probability of the causal connection in a case of specific causation. This is not to say that such a frequency is completely useless, but it means that it is no sufficient, *per se,* to prove the individual causation[47].

The necessary reference to the facts of the case leads to another important problem, which deals with the evidentiary use of statistics.

On the other hand, the statistical evidence of a general causation may be used *directly* – so to speak – as an evidence when the subject matter of the case deals with the *risk* or the *increasing risk* of damages provoked for instance by the exposure to dangerous materials, by the use of toxic medicament or by environmental pollution. Such a risk may be the subject matter of the case, for instance when a factory is sanctioned for not introducing protective measures preventing pollution or intoxication. In such cases the risk of damage – an example of general causation – is properly the main fact in issue, and then the epidemiological evidence may provide directly the proof of this fact. In general terms, so to speak, statistics may be properly used as evidence when the case does not deal with specific *past* events of individual causation, but deals with *future* probabilities of the occurrence of given events in a given population, since in these cases the frequency of such events in just a relevant aspect of the decision. Something similar happens when the case is a *class action* dealing with *mass torts*, when the number of individual harms cannot be established in a specific way, and then the dimension of the damages has to be determined by means of statistics concerning the population exposed to the risk[48].

46 See e.g. KAYE and FREEDMAN, 2011, p. 551 and 597, FROSINI, 2002, p. 37; STELLA, 2003, p. 294.

47 About low epidemiological frequencies see STELLA, 2003, p. 352.

48 See TARUFFO, 2016; GIUSSANI, 2016.

3.1 A Doubtful Example: Toxic Torts

Toxic torts are the domain in which the reference to statistics, mainly provided by epidemiology, is most frequent. However, it is also the domain in which the use of statistical evidence raises several problems[49].

First of all, it is commonly said that in the cases concerning toxic torts the general causation of the toxic effects of the use of dangerous medicaments or of the exposure to dangerous materials needs to be properly demonstrated, and then the specific causation of such effects in individual cases[50]. As to the proof of general causation there are no special problems since – as abovementioned – statistics may provide such a proof. The problem arises concerning the proof of specific causation: it is usually said that statistical probabilities have nothing to say about specific causation[51], but sometimes it is also said that statistics *may* prove such a causation, since they could provide a proof that achieves the civil standard of the preponderance of evidence, that is a probability of at least 51%[52].

Several American courts have accepted this theory[53]. The main argument is – in extreme synthesis – the following: if the *relative risk* of disease of those who used the medicament or were exposed to the dangerous material[54] is *two times* the risk of the non-users or unexposed, *therefore* in such cases there would be a proof of the specific causation in the individual cases, because the standard of the more probable than not has been achieved. Moreover, sometimes it is said that the statistics showing the *double risk* are a *sufficient* proof of the specific causation, and sometimes it is even said that such statistics are *necessary* to prove such causation[55].

49 See e.g. FROSINI, 2002, p. 59.

50 See GREEN, FREEDMAN and GORDIS, 2011, p. 552.

51 See e.g. RUBINFELD, 2011, p. 319.

52 About this standard see *infra*, 4.

53 See GREEN, FREEDMAN and GORDIS, 2011, p. 608, and broadly HAACK, 2014, p. 264, also for specific and analytical references.

54 The relative risk is the ratio between the rate of diseases in those who used the medicament or were exposed and the rate of the same disease among non-users or non-exposed. See GREEN, FREEDMAN and GORDIS, 2011, p. 566.

55 For analytical references and an analysis of such cases see HAACK, 2014.

TEORIA JURÍDICA CONTEMPORÂNEA
1:1-1, janeiro-junho 2016 © 2016 PPGD/UFRJ, p. 161-180

There is no need to develop here a thorough analysis of this argument, but some critical remarks are necessary, notwithstanding the positive opinion shared by several courts and by some writers[56].

First of all, one may be inclined to believe that if – for instance – the non-users of the medicament or the nonexposed suffer the disease in the proportion of 5% of the relevant population, and the users or the exposed suffer the same disease two times more (that is with a risk of 2), the outcome would be that for the users or the exposed the risk of such a disease is of 10%, but this would not say anything about the specific causation concerning particular individuals. It would just be information about the general causation in the population of the users or of the exposed, but nothing more. After all, a probability of 10% of risk for users and exposed may be relevant within the general assessment of evidence, but it is in no way equivalent to a probability of 50% in any case of specific causation.

(...) sometimes it is said that the statistics showing the double risk are a sufficient proof of the specific causation, and sometimes it is even said that such statistics are necessary to prove such causation.

On the other hand, even admitting that the double risk produces a probability of 50% in specific cases, this does not mean that the standard of the preponderance of evidence (or of the more probable than not) is achieved: 50% is not preponderant upon another 50%, then with 50% of probability the proof is not reached.

Moreover, in a recent essay[57] Susan Haack provides various epistemological arguments showing that the double risk is neither necessary nor sufficient to prove a specific causation. It is not necessary because such causation may be demonstrated, by any other kind of evidence, even when the double risk does not exist, as when, for instance, it is of 1,9 or even lower. In such cases the reference to the ratio of the risk may be useful together with other evidence, but it is never decisive and should not necessarily be of 2. The double risk is not sufficient because the statistical frequency of the risk is not *per se* a proof of what happened in specific cases. Therefore, once again, the knowledge of such a frequency may be useful if considered as a concurring information in the context of all the evidence available, but individual causation is not demonstrated only by the double risk.

56 See again GREEN, FREEDMAN and GORDIS, 2011, p. 611; HAACK, 2014, p. 269.

57 See HAACK, 2014, p. 264 and 285.

4. STANDARDS OF PROOF

Every time in which one speaks of the "sufficiency" of the evidence to support a conclusion about the facts in issue the problem is to establish when the evidence available is or is not sufficient for this purpose. In the legal language it is the problem of the so called s*tandard of proof,* that is of the threshold that should be achieved – by the evidentiary inferences – in order to conclude that the fact has been duly proven, that is: that the statement concerning such a fact has received a proper degree of logical confirmation (or justification) by the evidence at hand. Here the problem is relevant because in several cases the standards of proof are defined in terms of statistical probabilities.

However, as a preliminary remark it should be underlined that in several procedural systems – mainly in the *civil law* area – there are no numerical definitions of these standards, and the decision about whether or not a sufficient degree of proof of the facts in issue has been reached is left to a discretionary assessment of the trier of fact (usually a professional judge). In some cases, and the main example is the French criminal procedure, such a judgment is left to so-called *intime conviction* of the judge or of the jurors, that is to a merely subjective and basically irrational decision. In other cases, such as in the Italian and Spanish civil procedure, it is assumed that such an evaluation is made rationally[58], i.e. by applying rational criteria (such as the Spanish *sana crítica*)[59], but no numerical or probabilistic standards are established.

Correspondingly, the homeland of statistical standards are the common law systems, and specially the American one, but they are a very common point of reference even for other systems, and then some remarks are deserved here.

A very well-known standard for a criminal judgment, which is followed in more or less explicit ways also in other systems, as for instance in Italy after a recent reform[60], says that a criminal condemnation is possible if and only if the defendant's liability is established *beyond any reasonable doubt* (BARD). Such a standard

58 About the Italian system see TARUFFO, 2011, p. 519.

59 See ABEL LLUCH, 2015, p. 54, 86 and 113.

60 See e.g. STELLA, 2003, p. 154 and 195; FROSINI, 2002, p. 121; GARBOLINO, 2014, p. 461.

belongs to the history of common law[61] and it is used in several judgments of the American Supreme Court, beginning mainly in the Seventies[62]. In the present context it is not relevant in itself, since it may be interpreted simply by saying that it is better to acquit a guilty defendant than to condemn an innocent, but mainly because it is often interpreted in probabilistic terms, by saying for instance that it establishes a 90% or a 95% degree of proof of guilt as a minimum threshold required for condemnation. This interpretation of the BARD standard is very common but it is very doubtful, and there are various reasons to reject it. One of these reasons, argued very convincingly by Larry Laudan[63], is that any numerical quantification of the standard is meaningless. In particular, what seems impossible is *weighing* in a numerical or statistical scale the *reasonability* of the doubt concerning the guilt of the defendant. Therefore, as it argued by Jordi Ferrer[64], whether a doubt is reasonable should be established according to *rational*, not numerical, criteria.

In civil cases the common law, and mainly the American system, apply the standard of the *preponderance of evidence*, according to which the factual hypothesis proposed by the plaintiff has to be accepted if its degree of proof is higher than the degree of proof of the defendant's hypothesis. This standard is interesting in the present context because it is commonly said that it means that one of the two hypotheses should have at least a 0.51 (or a 51%) of probability, and consequently the opposed hypothesis should have at best a probability of 0.49 (or of 49%). For this reason it is usually said that the standard is of the *most probable than not*[65].

This theory is widely accepted by the American courts (not by the English ones)[66], but it seems almost meaningless to a thorough analysis. On the one hand, it is based on the assumption – which is typical of the American culture of the adversarial process – according to which in any case the judge has always to make a choice between

61 See SHAPIRO, 1999, p. 1.

62 See mainly STELLA, 2003, p. 156.

63 See LAUDAN, 2006, p. 29.

64 See FERRER BELTRÁN, 2007, p. 144.

65 In the wide literature about the subject see e.g. MORGAN, 1962, p. 21; CLERMONT, 2013, p. 16; CLERMONT, 2012, p. 60, LILLY, 1996, p. 55; REDMAYNE, 1999, p. 167; JAMES, HAZARD and LEUBSDORF, 1992, p. 339.

66 See REDMAYNE, 1999, p. 174.

the plaintiff's and the defendant's factual hypotheses. This does not work in other systems (or in other cultures), in which other hypotheses about the facts in issue may be considered, including also the judge's own "third" hypothesis[67]. But the most important problem of this theory is the implicit assumption that the plaintiff's and the defendant's versions of the facts are *complementary*, so that if one version is more probable the other has to be *correspondingly* less probable, and vice versa. Only assuming this premise, in fact, it may be said that if the plaintiff's version is proven at 0.51 *then* the defendant's one has only a probability of 0.49, and therefore should be rejected. But this statement is incorrect: actually in many cases the two hypotheses are not complementary and are simply *different.* Imagine, for instance, that the plaintiff says (A) "Peter provoked to me a damage of 1,000", and Peter says (B) "yes, but I paid 1,000 to the plaintiff". Both A and B may be true, or both may be false, and therefore the degree of probability of one hypothesis does not depend on the degree of probability of the other. Then, if one wants to think in terms of complementary probabilities he has to think of a *positive* hypothesis (A is true) and the corresponding *negative* hypothesis (A is false), since only in this case if A is true at 0.51 then A is false at 0.49, or vice versa. Then the correct way of interpreting the standard is the rule of the *more probable than not*, as it is said to be – but in a doubtful way – the real meaning of the preponderance of evidence.

Even in this case, however, things are not easy. On the one hand, and using the same example, A may be more probable than its negative, and then should be accepted, but also B may be more probable than its negative, and then should also be accepted. And so what? The decision has to be taken according to *legal rules* determining the legal consequences of such a situation, not according to complementary probabilities. Only when A has a probability at least of 0,51 or higher upon its negative, and B's negative has a probability of more than 0.51 A will prevail on B, but only because of legal reasons.

On the other hand, all this way of reasoning is based upon the assumption that probabilities from 0 to 1 may be ascribed to any factual hypothesis, but it is not so. Notwithstanding the

67 See TARUFFO, 2009, p. 32.

efforts of American courts and common law scholars, it has to be acknowledged that there are no methods that can be properly used in order to determine the numerical values of the preponderance of evidence or of the more probable than not of one or another factual hypothesis[68].

At any rate, determining standards of proof is a fundamental need in any rational theory of judicial decisions, since the standard establishes whether and when the proof of the facts in issue has been achieved. This does not mean that the American standards should be taken as models, and also it does not mean that there should be different standards for criminal cases and for civil cases. Rather, it may be admitted that several standards of proof may be determined, mainly depending on the kind of error distribution that is accepted in any particular situation, and on the basis of several factors such as the value of the amount or the kind of the sanction at stake, the different steps of the judicial proceedings, and so forth. The most important point to stress here is that this is not a theoretical problem: this is a matter for policy options that the lawgivers should make for the various situations in which a decision about the facts in issue has to be made.

(...) determining standards of proof is a fundamental need in any rational theory of judicial decisions, since the standard establishes whether and when the proof of the facts in issue has been achieved.

68 See JAMES, HAZARD and LEUBSDORF, 1992, p. 339.

REFERENCES

ABEL LLUCH, Xavier. *Las reglas de la sana crítica.* Madrid: La Ley, 2015.

ALLEN, Mark; HALL, Robert; LAZEAR, Victoria. Reference guide on Estimation of Economic Damages. In: Federal Judicial Center et al. *Reference Manual on Scientific Evidence.* 3rd ed. Washington, D.C.: The National Academic Press, 2011. p. 425-502.

APPELBAUM, Paul. Reference Guide on Mental Health Evidence. In: Federal Judicial Center et al. *Reference Manual on Scientific Evidence.* 3rd ed. Washington, D.C.: The National Academic Press, 2011. p. 813-896.

CLERMONT, Kevin. *Standards of Decision in Law:* Psychological and Logical Bases for the Standard of Proof, Here and Abroad. Durham: N.C., 2013.

______. *Death of Paradox:* The Killer Logic Beneath the Standard of Proof. 88 *Notre Dame L.Rev.*, v. 88, n. 3, p. 1061-1138, fev. 2013.

DOMINIONI, Oreste. *La prova penale scientifica.* Milano: Giuffrè, 2005.

DONDI, Angelo. *Paradigmi processuali ed "expert witness testimony" nel diritto statunitense. Riv.trim.dir proc.civ.,*p. 261-285, 1996.

FERRER BELTRÁN, Jordi. *La valoración racional de la prueba.* Madrid-Barcelona-Buenos Aires: Marcial Pons, 2007.

FROSINI, Benito Vittorio. *Causality and Causal Models:* A Conceptual Perspective. *International Statistical Review*, vol. 74, n. 3, p. 305-334, dez. 2006.

______. *Le prove statistiche nel processo civile e nel processo penale.* Milano: Giuffrè, 2002.

GARBOLINO, Paolo. *Probabilità e logica della prova*. Milano: Giuffrè, 2014.

GIUSSANI, Andrea. Proof of causation in group litigation. In: MartÍn-Casals, M.; Papayannis, D. M. (Ed.). *Uncertain Causation in Tort Law.* Cambridge, UK: Cambridge University Press, 2016. p. 240-249.

GOLDSTEIN, Bernard; HENIFLIN, Mary Sue. Reference Guide on Toxicology. In: Federal Judicial Center et al. *Reference Manual on Scientific Evidence.* 3rd ed. Washington, D.C.: The National Academic Press, 2011. p. 633-686.

GONZÁLEZ LAGIER, Daniel. *Questio facti*: ensayos sobre prueba, causalidad y acción. México: Fontamara, 2013.

GREELY, Henry; WAGNER, Anthony. Reference Guide on Neurosciences. In: Federal Judicial Center et al. *Reference Manual on Scientific Evidence.* 3rd ed. Washington, D.C.: The National Academic Press, 2011. p. 747-812.

GREEN, Michael; FREEDMAN, D. Michal; GORDIS, Leon. Reference Guide on Epidemiology. In: Federal Judicial Center et al. *Reference Manual on Scientific Evidence.* 3rd ed. Washington, D.C.: The National Academic Press, 2011. p. 549-632.

HAACK, Susan. *Evidence Matters*: Science, Proof and Truth in the Law. New York: Cambridge University Press, 2014.

______. *Defending science – within reason. Between Scientism and Cynicism,* Amherst, NY: Prometheus Books, 2003.

JAMES, Fleming; HAZARD, Geoffrey C; LEUBSDORF, John. *Civil Procedure.* 4th ed. Boston: Little Brown, 1992.

KAYE, David; FREEDMAN, David. Reference Guide on Statistics. In: Federal Judicial Center et al. *Reference Manual on Scientific Evidence.* 3rd ed. Washington, D.C.: The National Academic Press, 2011. p. 211-302.

______; SENSABAUGH, George. Reference Guide on DNA Identification Evidence. In: Federal Judicial Center et al. *Reference Manual on Scientific Evidence.* 3rd ed. Washington, D.C.: The National Academic Press, 2011. p. 129-210.

LAUDAN, Larry. *Truth, Error, and Criminal Law:* An Essay in Legal Epistemology. New York: Cambridge University Press, 2006.

LILLY, Graham. *An Introduction to the Law of Evidence.* 3rd ed. St. Paul: West Publishing Company, 1996.

MORGAN, Edmund. *Basic Problems of Evidence.* Philadelphia: American Law Institute, 1962.

NISBETT, Richard; ROSS, Lee. *L'inferenza umana. Strategie e lacune del giudizio sociale.* Bologna: Il Mulino, 1989.

REDMAYNE, Mike. Standards of Proof in Civil Litigation*. The Modern Law Review*, vol. 62, n. 2, p. 167-195, mar. 1999.

FEDERAL JUDICIAL CENTER et al. *Reference Manual on Scientific Evidence.* 3rd ed. Washington, D.C.: The National Academic Press, 2011.

ROBERTSON, Channing; MOALLI, John; BLACK, David. Reference Guide on Engineering. In: Federal Judicial Center et al. *Reference Manual on Scientific Evidence.* 3rd ed. Washington, D.C.: The National Academic Press, 2011. p. 897-960.

RODRICKS, Joseph. Reference Guide on Exposure Science. In: Federal Judicial Center et al. *Reference Manual on Scientific Evidence.* 3rd ed. Washington, D.C.: The National Academic Press, 2011. p. 503-548.

RUBINFELD, Daniel. Reference Guide on Multiple Regression. In: Federal Judicial Center et al. *Reference Manual on Scientific Evidence.* 3rd ed. Washington, D.C.: The National Academic Press, 2011. p. 303-358.

SCHAUER, Frederick. *Profiles, Probabilities and Stereotypes*. Cambridge, MA: The Belknap Press of Harvard University Press, 2003.

SHAPIRO, Barbara. *Beyond Reasonable Doubt and Probable Cause:* Historical Perspectives on the Anglo-American Law of Evidence. Berkeley: University of California Press, 1999.

STELLA, Federico. *Giustizia e modernità:* La protezione dell'innocente e la tutela delle vittime. III ed. Milano: Giuffrè, 2003.

TARUFFO, Michele. Proving Complex Facts: The Case Of Mass Torts. In: MartÍn-Casals, M.; Papayannis, D. M. (Ed.). *Uncertain Causation in Tort Law*. Cambridge, UK: Cambridge University Press, 2016. p. 165-175.

______. *La valutazione delle prove.* In: TARUFFO, M (Ed.). *La prova nel processo civile*. Milano: Giuffrè, 2012.

______. *Art.116*. In: Taruffo-Carratta. *Dei poteri del giudice:* Art.112-120. Bologna: Zanichelli, 2011. p. 519.

______. *La semplice verità:* Il giudice e la ricostruzione dei fatti. Bari: Laterza, 2009.

______. *Le prove scientifiche nella recente esperienza statunitense*. *Riv.trim.dir.proc.civ.,* 1996. p. 219.

TOULMIN, Stephen. *The Uses of Argument.* upd. ed. New York: Cambridge University Press, 2008.

TWINING, William. *Rethinking Evidence:* Exploratory Essays. 2nd ed. New York: Cambridge University Press, 2006.