D E C A D E S
O F
J U R I M E T R I C S

Filipe J. Zabala[*1] and Fabiano F. Silveira[†2]

[1]School of Technology, PUCRS

2019-12-31

Jurimetria é harmonia
entre o exato e o direito
calibra o que é meu com os fatos
previsível, enfim, quem diria?

Contents

[*]filipe.zabala@pucrs.br
[†]fabiano.feijo.silveira@gmail.com

Abstract

Jurimetrics: decades of history, decades to-be auspicious. A Brazilian point of view on the trajectory of this forgotten concept in the quantitative approach of the law, with code and examples in free software.

Keywords: Jurimetrics, Law, Statistics, Thomas Bayes, Lee Loevinger, Artificial Intelligence, Empirical Legal Studies.

1   Old Wine in New Bottles

THE YEAR 2019 is interesting in a 'jurimetrical' point of view. Several historical decades have been completed, which makes an appropriate moment for reflections on possible deviations that may be distorting the current understanding of the matter.

The conceptual framework of jurimetrics was first presented $(2019 - 1949)/10 = 7$ decades ago by [Loevinger(1949)][1]. His manifesto begins considering that 'one of the greatest anomalies of modern times that the law, which exists as a public guide to conduct, has become such a recondite mystery that it is incomprehensible to the public and scarcely intelligible to its own votaries'. Even nowadays the scenario is not much different, but the tools available at this time allow to address this issue, rather than simply 'make the people obey the laws they do not understand'.

In Loevinger's writings from 1949 to 1992 it is possible to observe the influence of a wide variety of philosophers, like Oliver Wendell Holmes Jr., Thomas Bayes, Francis Bacon, Aristotle, among others. He was 'standing on the shoulders of giants', as the saying attributed $(2019 - 1675)/10 \approx 34$ decades ago to Isaac Newton. This idea, however, was documented $(2019 - 1159)/10 = 86$ decades ago in John of Salisbury's Metalogicon [John of Salisbury and Lejeune(2009)], which accredits to the French philosopher Bernard of Chartres. On the foreword of [Merton(1993)], Humberto Eco suggests that this concept is found in the texts of Priscian in the 5th century, around 160 decades ago.

> *As Merton himself says, Bernard's Aphorism is known through its quotation by John of Salisbury in the Metalogicon (I can confirm his answer to a question raised by Merton: it is, indeed, in III, 4), and Bernard is not the original inventor, for the concept (if not the metaphor of the dwarfs) appears in Priscian six centuries earlier.* (Humberto Eco on the foreword of [Merton(1993)], p. xiv.)

Even in modern science is not straightforward to measure the contribution of each collaborator on a complex theory, although there are proposals in the literature [Winston(1985)]. [Donoho(2015)] discusses [*t*]*he 50 years of data science*, when more than 5 decades ago John Tukey wrote *The Future of Data Analysis*, pointing an 'unrecognized science whose subject of interest was learning from data, or 'data analysis''. 'No scientific discovery is named after its original discoverer', states the *Stigler's Law of Eponymy* [Stigler(1980)] – which ironically is an eponym[2] – in the sense that

> *Laplace employed Fourier transforms in print before Fourier published on the topic, that Lagrange presented Laplace transforms before Laplace began his scientific career, that Poisson published the Cauchy distribution in 1824, 29 years before Cauchy touched on it in an incidental manner, and that Bienaymé stated and proved the Chebychev inequality a decade before and in greater generality than Chebychev's first work on the topic.* [Stigler(1980), p. 148]

In the text that assigned the term *Googol*[3] to the number 10^{100}, [Kasner and Newman(1940)] points out in the Chapter 1 – *New Names for Old* – that

> [*e*]*very once in a while there is house cleaning in mathematics. Some old names are discarded, some dusted off and refurbished; new theories, new additions to the household are assigned a place and name.* [Kasner and Newman(1940), p. 3]

Thus, we are only guardians of the old wine, passed from bottle to bottle throughout history.

---

[1] https://mncourts.libguides.com/lee_loevinger

[2] https://en.wikipedia.org/wiki/Eponym

[3] The term was coined by Edward Kasner's nephew, Milton Sirotta (1911–1981). According to [Koller(2004)] the name 'Google' is a corruption of 'Googol', misspelled by Sean Anderson in September of 1997 when the domain google.com was registered.

1.1 Recognizing the bottles

During research the authors were led to [Leibniz(1666)], that for the given purposes will be considered the milestone on the formal association of quantitative thought and law, although [Friendly(2008)] refers to early 1660s texts concerning a 'political arithmetic'. Leibniz sheds light on the subject focusing on symbolic representation of languages using basic components units. With more than thirty five decades old, his writings suggests 'the primitive terms comprise things and, as well, modes or relations', in a translation from Latin [Amunategui(2014)]. This kind of structure involving small blocks associated by a set of rules is merged on the foundations of the legal reasoning and all kinds of quantitative approach like statistics, mathematics, computer science, data science, machine learning, artificial intelligence and many other labels.
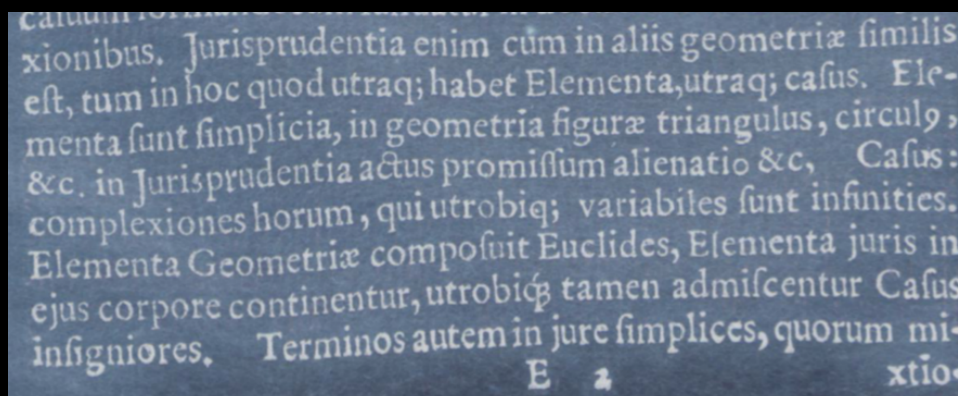

Figure 1: [Leibniz(1666)] excerpt, p. 27.

Exactly $(2019 - 1709)/10 = 31$ decades ago, Nicolaus (Nicolas/Nikolas/Niklaus) I Bernoulli (1687-1759) published his thesis *De Usu Artis Canjectandi in Jure*, or 'On the Use of the Art of Conjecturing in Law' [Bernoulli(1709)]. It is a well-documented reference of the quantitative approach in the law, which deals with everyday issues such as lottery and insurance pricings, inheritance, confidence in witnesses, probabilities of innocence and survival of people. [Kohli(1975)] and [Hald(2003)] points out he had great influence from the work of his uncle and master James (Jacques/Jakob/Jacob/Jacobi) I Bernoulli (1654-1705), *Art of Conjecture*[4], published posthumously in 1713 [Bernoulli(1713)].

> *An edition of the works of Jakob Bernoulli would be incomplete if one did not attach to it the thesis of his nephew Niklaus from the year 1709 'On the Use of the Art of Conjecturing in Law'. The spiritual father of this work is clearly Jacob. Whole sections from both the diary and the 'Ars Conjectandi' Niklaus has taken literally. At other points, questions and mere hints of Jacob were taken up and further processed.* [Kohli(1975), p. 541][5]

Also inspired on James I Bernoulli's work, [Condorcet(1785)] presents the currently known *Condorcet's jury theorem*, a milestone in problems concerning voting systems. It can be applied in a wide variety of fields, such as social sciences and machine learning. The theorem can be stated in terms of a dichotomous variable assuming the values 1 and 0, and it is considered reasonable to assign a 'correct' or 'incorrect' classification. If a decision maker – e.g. a judge or classifier – assigns correctly the 1's with probability θ greater than $1/2$, the theorem asserts that more decision makers increases the overall probability of correct assignments. With θ less than $1/2$, more decision makers decreases the overall probability of correct assignments, and for $\theta = 1/2$ the number of decisiors is indifferent. Adapted from [Berg(1996)], the theorem is formally described as follows.

---

[4] Part IV, Chapter II.

[5] *"Eine Ausgabe der Werke von Jakob Bernoulli wäre unvollständig, würde man ihr nicht die Dissertation seines Neffen Niklaus aus dem Jahre 1709 'Über den Gebrauch der Mutmaßungskunst in Fragen des Rechts' beifügen. Der geistige Vater dieses Werkes ist eindeutig Jakob. Ganze Abschnitte sowohl aus dem Tagebuch als auch aus der 'Ars Conjectandi' hat Niklaus wörtlich übernommen. An andern Stellen wurden Fragestellungen und bloße Andeutungen Jakobs aufgegriffen und weiterverarbeitet."*

Theorem 1. (***Condorcet's theorem***) *Let $(X_1, \ldots, X_n)$ be n independent binary distributed random variables such that $Pr(X_i = 1) = \theta > 1/2$ and $P_n = Pr(\sum X_i > n/2)$. Then (a) $P_n > \theta$ and (b) P_n is monotonously increasing in n and $P_n \longrightarrow 1$ as $n \longrightarrow \infty$. If $\theta < 1/2$, then $P_n < \theta$ and $P_n \longrightarrow 0$ as $n \longrightarrow \infty$. Finally, when $\theta = 1/2$, then $P_n = 1/2$ for all n.*

Example 1. *If $n = 3$ and $\theta = 0.6$ then P_3 – the probability of at least two of three decision makers agree in the correctly assignment of **1**'s – is given by*

$$P_3 = 0.6 \times 0.6 \times 0.6 + 0.6 \times 0.6 \times 0.4 + 0.6 \times 0.4 \times 0.6 + 0.4 \times 0.6 \times 0.6 = 0.648 > 0.6.$$

☺

Example 2. *If $n = 3$ and $\theta = 0.3$, then*

$$P_3 = 0.3^3 + 3 \times 0.3^2 \times 0.7 = 0.216 < 0.3.$$

☺

Example 3. *If $n = 3$ and $\theta = 0.5$, then*

$$P_3 = 0.5^3 + 3 \times 0.5^3 = 0.5.$$

☺

[Condorcet(1785)] also points James I Bernoulli and Abraham de Moivre as precursors of the idea of seeking the probability of future events according to the law of past events, even they have given no method in their works to achieve this[6]. Considering the method, Condorcet points out the work of Thomas Bayes and Richard Price [Bayes(1763)], as well as Pierre-Simon Laplace for treating the question analytically [Laplace(1774)]. The Condorcet's theorem and its consequences has been discussed and extended in literature[7] until recently.

Still in France, the 19th century brought the works of Adolphe Quetelet, André-Michel Guerry and Simeón Poisson. They were interested in Laplace's work, as well in the analysis of conviction rates. During the development of *l'homme moyen*[8] concept, from 1827 to 1835 Quetelet looked for potentially meaningful relationships in social data. According to [Stigler(1986)],

> [h]*e examined birth and death rates by month and city, by temperature, and by time of day. He calculated the month of conception from the birth month and tried to relate it to marriage statistics. He investigated mortality by age, by profession, by locality, by season, in prisons, and in hospitals. He considered other human attributes: height, weight, growth rate, and strength. Quetelet's interests also extended to moral qualities: statistics on drunkenness, insanity, suicides, and crime.* [Stigler(1986), p. 186]

Based on *Compte Général de L'administration de la Justice Criminelle en France* data [France(1829)], Quetelet worked in the analysis of conviction rates, also handled by [Guerry(1833)] and [Poisson(1837)]. According to [Stigler(1986)],

> *Quetelet had given the numbers of accused and convicted as 7,234 and 4,594; Poisson gave them as 6,652 and 4,037, respectively. Thus for Quetelet the 1825 conviction rate was 0.635, for Poisson it was 0.607. The explanation for this discrepancy is that in preparing the report for 1827 the minister of justice (Count Portalis) had discovered that in the report for 1825 the figures given had been augmented by those for accused condemned in absentia, and he had provided the needed correction. Apparently Quetelet missed this change (announced in a footnote on p. v of the Compte général for 1827), but Poisson did not.* [Stigler(1986), p. 188]

The seminal works of [Holmes Jr.(1881)] and [Holmes Jr.(1897)] lead to new perspectives in the Law, considering the precedent (the decided, or the *stare decisis*) in similar cases. Discussing early forms of liability, Holmes declares 'other tools are needed besides logic'.

---

[6] "*L'idée de chercher la probabilité des évènements futurs d'après la loi des évènements passés, parroît s'être présentée à Jacques Bernoulli & à Moivre, mais ils n'ont donné dans leurs ouvrages aucune méthode pour y parvenir. Mrs Bayes & Price en ont donné une dans les Transactions philosophiques, années 1764 & 1765, & M. de la Place est le premier qui ait traité cette question d'une maniére analytique.*"

[7] [Grofman et al.(1983)], [Boland(1989)], [Ladha(1992)], [Berg(1993)], [Berg(1996)], [Austen-Smith and Banks(1996)], [List and Goodin(2001)], [Williams(2004)], [Gehrlein(2006)], [Gehrlein and Lepelley(2011)], [Kaniovski and Zaigraev(2011)], [Zaigraev and Kaniovski(2012)] and [Gottlieb and Hussain(2015)].
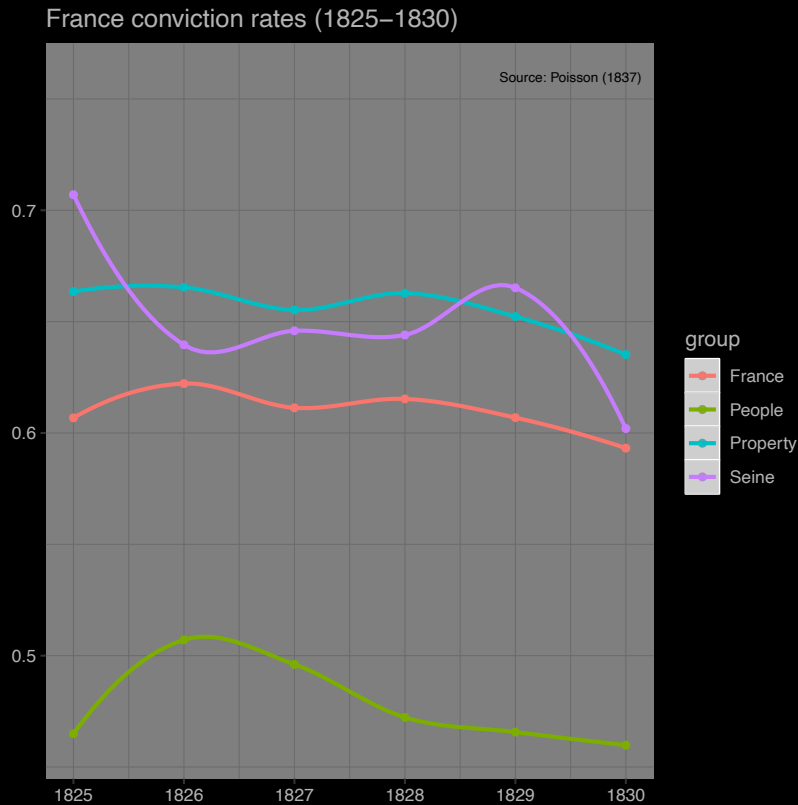
[8] Average man.

Figure 2: France conviction rates (1825-1830)

The life of the law has not been logic: it has been experience. [Holmes Jr.(1881), p. 5]

The *Bayesian* approach provides a formal mechanism to incorporate information to decision maker's experience. Such approach – discussed briefly in Sections 1.2, 3.2 and 4 – also provides tools to prediction, suggested by Holmes Jr. in stating that

[t]*he object of our study, then, is prediction, the prediction of the incidence of the public force through the instrumentality of the courts.* [Holmes Jr.(1897), p. 1]

[Llewellyn(1930b)] discusses the problem of defining law and defends the Roscoe Pound's 'precepts', in opposition to 'rules' for being a 'term sufficiently ambiguous'. At page 460 Llewellyn considers 'making the study of law a study in first instance of particularized situations', where 'generalization must come from a resynthesis of such particularized studies'.

Inspired by Oliver Wendell Holmes Jr.'s writings and considering the advances in quantitative methods, [Loevinger(1949)] is a manifesto in defense of rationality in the law. Labeled *jurimetrics*, this house cleaning occured seven decades ago and still provokes high-level discussions. Forward Llewellyn's next step, Loevinger identifies Roscoe Pound – dean of Harvard Law School – as 'the most prominent spokesman' of Legal Realists, considered by Loevinger a 'school of Sociological Jurisprudence' that 'has developed more from the stimulus of (Rudolph von) Jhering than of Holmes'. Loevinger defends the 'legal principle as the most significant aspect of law', opposed to the 'emphasis on the particular case', 'although agreeing with the Realists that law must adapt itself to social needs'.

'*The scholars whose writings have emphasized the particular case, rather than the general rule, as the basis for a study of the law, have been called Legal Realists*'. [Loevinger(1949), p. 10]

The next step forward in the long path of man's progress must be from jurisprudence (which is mere speculation about law) to jurimetrics – which is the scientific investigation of legal problems. [Loevinger(1949), p. 31]

After the 1949 seminal article, Lee Loevinger was proliferous writing about science and law in the 1950s and 1960s. [Loevinger(1950)] makes a book review of *Courts on Trial: Myth and Reality in American Justice*, published by Jerome Frank in 1949. He points out the shift of Frank's ideas between 1930 and 1949, criticizing the 'individualized cases' concept and the lack of a formal definition for 'justice'.

[Loevinger(1952)] discusses some Holmes Jr.'s ideas, like 'the life of the law has not been logic: it has been experience'. He considers this a slogan 'at least superficially misleading', but points out that Holmes Jr. wish 'a more conscious and rational recognition of the grounds of judicial decision'. Also makes a defense of the use of logic in law, referring indirectly to the concept of *coherence* [Leonard(1980)], [Loschi and Wechsler(2002)]. Loevinger retakes Jerome Frank's work, citing Mortimer Adler, Walter Wheeler Cook, Dennis Lloyd, Julius Cohen, William James, John Stuart Mill, John Dewey among others, making considerations about the foundations of probability theory.

> *While there appears to be a real difference between a 'fair preponderance of the evidence' and 'clear and convincing proof', there is no obvious indication that there is a similar differentiation between the latter phrase and 'proof beyond a reasonable doubt'. At least some of this confusion might be avoided if courts were to adopt a terminology of probability logic. The conventional mode of representation calls for the use of 0 to indicate that a proposition is untrue or impossible, the use of 1.0 to indicate that a proposition is true or certain, and the use of intervening values to indicate the relative probability (or frequency) of its being true.* [Loevinger(1952), p. 495]

Loevinger finishes *An Introduction to Legal Logic* stating

> [t]*he future of the law, and perhaps of society itself, will depend upon the ability of the legal professionals to develop and utilize patterns and forms of thinking in the law adequate to deal with the complex problems of modern life.* [Loevinger(1952), p. 522]

[Loevinger(1958)] refers again to 'legal realists' and takes up the issue of 'individualized cases', making a deep discussion in the distinction between 'evidence', 'facts'/'opinions', 'subjective'/'objective' and 'perceptions'/'conclusions' concepts. He discusses admissibility and coherence, suggesting the social policy 'be accomplished by establishing an appropriate rule relating to the degree of proof required, rather than by artificial rules excluding classes of evidence'. He ends the article declaring

> *in order to function effectively at any level, every mind must take account of the learning and thinking that has preceded and must approach contemporary problems in terms of contemporary concepts.* [Loevinger(1958), p. 175]

[Loevinger(1961)] discusses the advances of science, considering that 'science and law have been linked in man's speech and thinking for centuries'. He points out Karl Pearson's idea that 'the classification of facts, the recognition of their sequence and relative significance is the function of science'. From Oliver Wendell Holmes Jr. writings, Loevinger remembers that 'an ideal system of law should draw its postulates and its legislative justification from science' and 'for the rational study of the law the black-letter, man may be the man of the present, but the man of the future is the man of statistics and the master of economics'. The scientific advances cited by Loevinger are related largely to electronics in the field of data retrieval. He defines in a very mature way the concepts of *hardware* and *software*:

> *'Hardware' means simply the mechanical devices – the physical machines – that science has produced. 'Software' means the intellectual systems of designs and concepts that have been produced.* [Loevinger(1961), p. 258]

He alerts 'that legal prediction is an activity in which lawyers, and for that matter citizens in all occupations, are commonly engaged', and that the legal profession has no choice unless adapt to new technologies 'to retain its position of intellectual leadership'. To apply in a practical way he points out

> [t]*he branch of mathematics that appears to be of the most immediate practical utility in the fields of law and the behavioral sciences is statistics. There is much in statistics that is of present practical application in day-to-day legal problems and it has good claim to be included in every law school curriculum.* [Loevinger(1961), p. 262]

[Loevinger(1962)] discusses the principle of parsimony in scientific thinking, known as *Occam's Razor* and summarized in the maxim *'entities should not be multiplied beyond necessity'*[9]. [Loevinger(1963)]

---

[9] *'Essentia non sunt multiplicanda praeter necessitate', William of Occam, 14th century.*

retakes some ideas of Oliver Holmes Jr. and brings jurimetrics as a new science, considering jurisprudence as 'the philosophy of law'. Examining the methodology of legal inquiry, Leovinger considers the electronic data retrieval as an interesting and seminal area of work. He details some principles of text manipulation and points out an *association factor* purposed by [Stiles(1961)]. Such factor assigns a relevance to related terms, 'where A is the number of documents indexed by one term; B is the number of documents indexed by a second term; f is the number of documents indexed by the combination of both terms; and N is the total number of documents in the collection. If AB is greater than fN the association is negative'.

$$\log_{10} \frac{\left(|fN - AB| - \frac{N}{2}\right)^2 N}{AB(N - A)(N - B)} = \text{ASSOCIATION FACTOR}$$

Figure 3: Association Factor by [Stiles(1961)].

Loevinger also discusses a *probabilistic indexing* suggested by [Maron and Kuhns(1960)], as well a system named 'LEX' designed as a LEgal indeX by the Antitrust Division of the U.S. Department of Justice[10]. In the following excerpt he describes the modern approach to data manipulation:

> *Conceivably, it may become possible to record all the decisions in a field of law, or perhaps in the entire law, in such fashion that their full text can be searched electronically in microseconds.* [Loevinger(1963), p. 26]

The discussion purposed by Loevinger is very opportune, considering that substantial textual structures are easily handled today with tools written in R, Python and REGEX (REGular EXpressions)[11], for instance. The TF-IDF (Term Frequency–Inverse Document Frequency) [Spärck Jones(1972)] and its derived metrics [Lopes et al.(2016)] can be considered a new generation of relevance metrics discussed by Loevinger. Also driven by the modernization of the legal processes, [Allen and Caldwell(1963)] considers schemes and diagrams as shown in Figure 4, while [Allen(1963)] sets his thesis:

> *If*
> 1. *the written materials used in the tax field are more systematically drafted,*
>
> *Then*
> 2. *human beings will be able to 'read' and 'work with' those materials 'better', and*
> 3. *automatic devices will be able to 'read' and 'work with' those materials 'better'.*
>
> [Allen(1963), p. 714]

[Loevinger(1966)] discusses science and law as rival systems, claiming at page 535 that 'up to the present time law has made little use of science, or the empiric method, and (...) the lawyers generally have little understanding of science or its concepts and methods', pointing out a symposium issue in *Law and Contemporary Problems*[12] as one of the few significant efforts to scientific work in the law. Loevinger also discusses the empirical approach given by [Kalven et al.(1966)] about jury behavior and makes a defense of the complementarity of methods and the difficulty in recognizing the limitations of each field.

> *The point that needs to be made, however, is that the empiric and the dialectic methods are not rivals or alternatives but complementary methods adapted to different problems and applicable in different situations. (...) Much of the difficulty in the relationship between law and science has arisen from the failure of both lawyers and behavioral scientists to recognize the limitations of their respective methods and the kinds of problems to which they are appropriate.* [Loevinger(1966), pp. 541-543]

Giving preference to data production over more theories, Loevinger considers 'the records of judgements in the hundreds of trial courts' as 'an obvious rich mine of data' and indicates 'an ineluctable relationship between science and semantics'. He was concerned about the different understandings of vague terms as 'justice', 'reasonable' and 'public interest', purposing 'experimental and quantitative investigation of such semantic questions'.

---

[10] https://www.justice.gov/atr
[11] https://regex101.com/
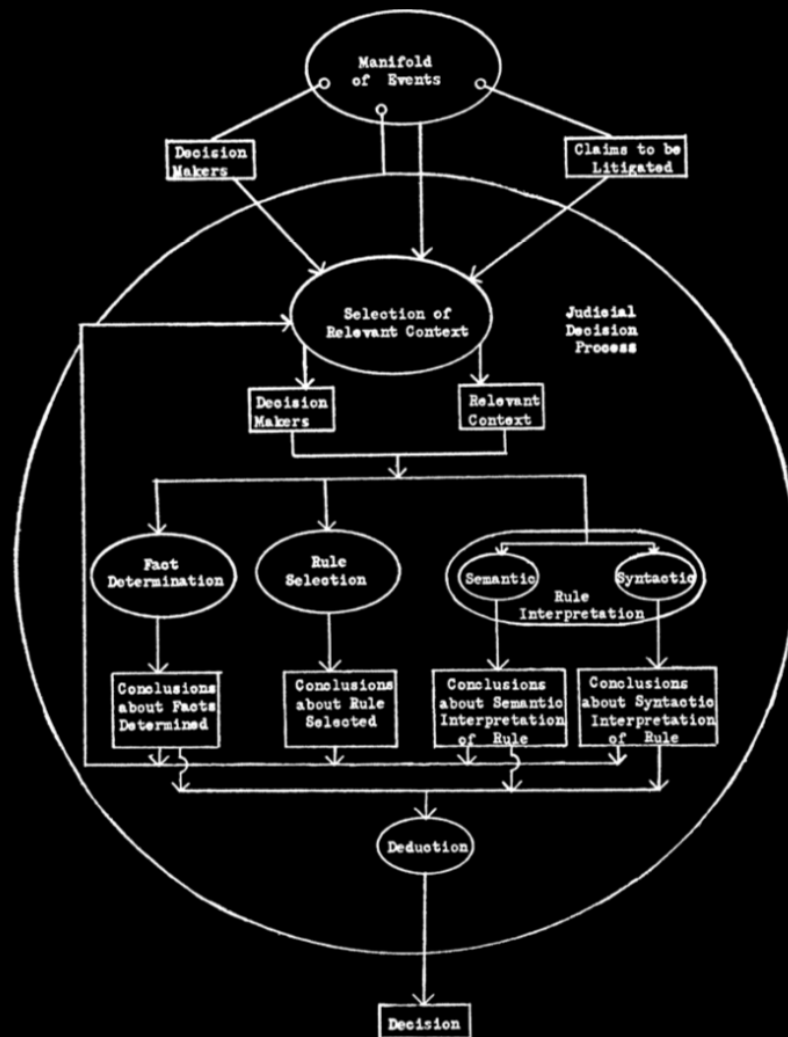[12] https://scholarship.law.duke.edu/lcp/vol28/iss1/

Figure 4: Scheme by [Allen and Caldwell(1963), p. 227].

The profession of law and those it serves can hope and demand that the law schools begin to study both of the great systems of gathering data, the dialectic and the empiric, and bring both to bear in seeking solutions of the proliferating and increasingly complex problems of government in a scientific age, and in training those who will become our future governors. [Loevinger(1966), p. 551]

[Loevinger(1985)] makes a review of the main ideas of jurimetrics, in a lecture pointing out 'the greatest developments of technology now appear to be tending toward a dispersion of power', in the sense that

the world today is divided between societies which value democracy and individual freedom and those which subjugate the individual to the state and are ruled by authoritarian regimes. Probably the most critical issue confronting us today is the distribution of power – political power – within our own society and in the community of nations. (...) In the critical area of communications new modes of transmitting and receiving information are proliferating and are segmenting audiences for all of the mass media. At the same time the new technologies provide increasing means for small groups and even individuals to disseminate views and make divergent voices heard. [Loevinger(1985), p. 19]

[Loevinger(1992b)] asserts both science and law 'rest upon subjective judgments or assumptions', and 'proof of facts in all disciplines rests upon subjective judgments of probability'. This is a Bayesian point of view, deeply supported in literature since [Bayes(1763)].

Bayes' Theorum in Boolean Notation

$$P(h|e) \;=\; \frac{P(e|h)P(h)}{P(e)}$$

The posterior probability is proportional to the prior probability multiplied by a factor resulting from revision of the prior probability on the basis of the new evidence.

P(h|e) is the posterior probability of **h** given **e**
P(e|h) is the probability of **e** given **h**
P(h) is the prior probability of **h**
h is the hypothesis
e is the new evidence

Figure 5: Bayes' Theorem by [Loevinger(1992b), p. 327]

An important reference concerning the subjectivist point of view of probability is the work of Bruno de Finetti. He purposed the *representation theorem* [de Finetti(1937)], where *exchangeable variables*[13] are *conditionally independent* given a non-observed parameter θ. This quantity is treated as a nuisance parameter, giving a predictive characterization to de Finetti's theorem. Throughout his work a strong subjectivist point of view becomes clear, as can be noticed, e.g., in the motto 'PROBABILITY DOES NOT EXIST' [de Finetti(1974), p. x] and in the following excerpt.

> *There is no way, however, in which the individual can avoid the burden of responsibility for his own evaluations. The key cannot be found that will unlock the enchanted garden wherein, among the fairy-rings and the shrubs of magic wands, beneath the trees laden with monads and noumena, blossom forth the flowers of 'Probabilitas realis'. With these fabulous blooms safely in our button-holes we would be spared the necessity of forming opinions, and the heavy loads we bear upon our necks would be rendered superfluous once and for all.* [de Finetti(1975), p. 42]

Under this point of view, it is possible to formally assign the decision maker's opinion about θ to a probability distribution called *prior*, which in turn is calibrated with data obtained from the *likelihood function*. This procedure provides a *posterior* distribution, or the decision maker's opinion about θ after the data. This approach can be naturally applied in the law, as pointed out by Loevinger.

> *Lawyers gather data, which they call 'evidence'; scientists gather evidence, which they call 'data'. Both terms mean the same thing, which is intellectual support for some conclusion or proposition.* [Loevinger(1992b), p. 323]

Loevinger retakes the Ockham's Razor concept from the Bayesian point of view [Jefferys and Berger(1992)], and points out legal rules of evidence and 'burden of proof' conditioned on the admissibility of such proofs and evidences. He indicates some cases in which probability calculations was used by courts, discussing objective and subjective approaches to probability interpretation. Finally, [Loevinger(1992a)] discusses philosophically some aspects of logic and Sociology, re-examinating 'albeit somewhat reluctantly' [Kaye(1992)], [Loevinger(1992b)] and [Jasanoff(1992)].

[Kowalski(1995)] points out the similarities between computing and the law, that 'seem to cover all areas of computing software (...) as programming, program specification, database description and query, integrity constraints, and knowledge representation in artificial intelligence'. Besides that, he indicates the similarities between computing and law are 'extend also to the problems that the two fields share of developing, maintaining and reusing large and complex bodies of linguistic texts'.

> *The linguistic style in which legislation is normally written has many similarities with the language of logic programming. (...) These extensions include the introduction of types, relative clauses, both ordinary negation and negation by failure, integrity constraints, metalevel reasoning and procedural notation.* [Kowalski(1995), p. 325]

---

[13]An exchangeable variable is characterized when the order of the observations does not alter its probability law.

At this point in the timeline it is possible to observe the non-intersection of Jurimetrics with Empirical Legal Research and Legal Realism. Several textbooks emerged indicating some point of views on the history of the quantitative empiricism in the law, to the best of our knowledge not quoting Loevinger and jurimetrics. [Schlegel(1995), $P = 432, L = 0, J = 0$][14] conceives *American Legal Realism and Empirical Social Science* in the sense that 'the point of this book is the stories'. The author declares to 'love stories' and be 'better at narrative than analytic history', offering to his readers 'a capsule summary of *the* story of Realism as it is usually told'.

> *Thus it was not until the 1920s that more than an isolated soul would claim that legal science was unscientific and so elicit clues as to why that was, and still is, the case. The group of scholars that made this claim and so brought the notion of science as an empirical inquiry if not into, then at least up against, law was the American Legal Realists.* [Schlegel(1995), p. 1]

[Kritzer(2009), $P = 50, L = 0, J = 0$] makes a bibliographic essay of the *Empirical Legal Studies Before 1940*, stating that '*Empirical Legal Studies* is a term that began to come into vogue around 2000', pointing out several studies conducted in the 1950s and early 1960s. He considers the legal realism as an essentially American movement 'in significant part' because 'almost all of the empirical legal research of the period was done by Americans focusing on the United States'. While the author do not discusses every study found, he claims to have 'provided as complete a bibliography of that research as possible'.

[Cane and Kritzer(2010), $P = 1112, L = 0, J = 1$][15] indicates in *The Oxford Handbook of Empirical Legal Research* that 'in the American legal Academy, empirical research gained contemporary prominence in the late 1990s'. The authors points out 'the genesis of empirically based studies of judicial behavior is commonly traced to the pioneering work of C. Hermann Pritchett in the late 1940s'.

In *An Introduction to Empirical Legal Research*, [Epstein and Martin(2014), $P = 324, L = 0, J = 0$] retakes some Holmes Jr.'s ideas, as 'the man of the future is the man of statistics and the master of economics', pointing out that the 'future is here'. The authors present code and databases treated in R and Stata softwares. They use a classic statistical approach, considering methods that violates the likelihood principle[16] – such as confidence intervals and p-value calculation from classical hypothesis tests – as 'close approximations to Bayesian methods', which can be noticed in the following excerpt.

> *Technically, to use population data to extrapolate to a different context or a future context, Bayesian statistics are the appropriate method. Bayesian statistics treat data as given and parameters as the things about which we have uncertainty. All the inferential tools described in this book are close approximations to Bayesian methods as long as we have sufficiently large samples, which provides justification for their use. In other words, it is appropriate to use hypothesis tests or confidence intervals as part of a process of extrapolation to different contexts or different time periods as an approximate Bayesian solution to the inferential problem.* [Epstein and Martin(2014), p. 154]

At last, the *Empirical Legal Research* by [Leeuw and Schmeets(2016), $P = 328, L = 0, J = 0$] 'covers all major fields of law, as the *Oxford Handbook of Empirical Legal Research* (Cane and Kritzer, 2010) shows'. On the other hand, Loevinger exposes in his articles 'considerable diversity among the references cited', discussing the ideas of a wide range of authors.

> *Some service has, indeed, been rendered by the modern thinkers, Bentham, Jhering, Holmes, Pound, the Realists, and others of similar views, in bringing law out of the sky and down to earth.* [Loevinger(1949), p. 15]

[Bench-Capon(2015)] refers to [Loevinger(1949)] and [Mehl(1959)] as who anticipated many of the computational systems that have been implemented and proposed. For some reason, however, the Empirical Legal Research and Legal Realism literature does not gave the appropriate credit to Lee Loevinger. Clues about this issue are found (again) in [Loevinger(1949)]. At the end of the second paragraph of page 4, he asserts that 'the trouble with law is not the public but the lawyers, that what is needed is not publicity but progress'. At pages 39-40 claims that 'advertising is no substitute for research', and complements on the footnote #85:

---

[14] P: number of (#) pages. L: # times `loevinger` term was found. J: # times `jurimetric*` radical was found.

[15] The reference found was the citation of *Robertson, D. (1982). "Judicial Ideology in the House of Lords: A **Jurimetric** Analysis," British Journal of Political Science 12: 1–25*. Robertson does not cite Loevinger, considering the 'celebrated paradigm is Schubert's *The Judicial Mind*, a classic of what has been called 'jurimetrics'.

[16] For more details see [Birnbaum(1962)], [Wechsler et al.(2008)] and [Mayo(2014)].

> *Needless to say, the term 'research' is here used in its scientific sense, and what is carelessly called 'legal research' by the average lawyer has no more relation to it than numerology has to statistics.* [Loevinger(1949), p. 40]

Cultivating a collaborative thinking and inspired by known and unknown giants, the authors support a unified approach to the topic and celebrate the dedication of so many people for so many decades in building such a vast and detailed work. Cheers!

1.2   Drinking the wine

> *Of course it is not important what term is used to indicate the scientific discipline suggested. It is important that it have a distinctive name, as well as a general program. The name suggested here seems, to the author, as good as any, since it seems to indicate the nature of the subject matter, and corresponds to other similar terms, such as biometrics and econometrics.* [Loevinger(1949), p. 31]

Jurimetrics is the application of quantitative methods in the law.[17] So, anyone who considers quantitative approaches to legal problems is making jurimetrics. The term is concise and intuitive like biometrics, chemometrics, econometrics and so on. It combines the flexibility of the Human Science with the precision of the Natural Science.

An example of this scientific association is the case *United States v. Carroll Towing Co.*, concerning the sinking of the barge Anna C on January 4, 1944. Judge Learned Hand uses for the first time a cost-benefit analysis for assigning liability and determining negligence [Hand(1947)]. Although [Feldman and Kim(2005)] discusses other approaches and indicates the called *Hand rule* might produce games with inefficient equilibria, the case is a milestone in a modern use of quantitative methods in the law. In the judge's words, the Hand rule is described as follows.

> *Since there are occasions when every vessel will break from her moorings, and since, if she does, she becomes a menace to those about her; the owner's duty, as in other similar situations, to provide against resulting injuries is a function of three variables: (1) The probability that she will break away; (2) the gravity of the resulting injury, if she does; (3) the burden of adequate precautions. Possibly it serves to bring this notion into relief to state it in algebraic terms: if the probability be called P; the injury, L; and the burden, B; liability depends upon whether B is less than L multiplied by P: i.e., whether $B < LP$.* [Hand(1947), online]

To the best of our knowledge, the first works in brazilian jurimetrics were produced in IME-USP[18]. [Montoya-Delgado(1998)] and [Montoya-Delgado et al.(2001)] presents a methodologic proposal for calculating the probability of paternity when DNA information available comes only from the accused's relatives. The proposed method allows to evaluate the probability of paternity considering the Hardy-Weinberg equilibrium law using the Bayesian approach.

[Nakano(2006)] builds 'a mathematical model for calculating the probability of paternity and implement it in software'.[19] The author also considers the Hardy-Weinberg equilibrium law, applying methods such Bayesian Networks and Full Bayesian Significance Test (FBST). Purposed by [Pereira and Stern(1999)], the FBST is a Bayesian measure of evidence for precise (sharp) hypotheses, as well as a Bayesian alternative to significance tests or, equivalently, to p-values.

[Wechsler et al.(2006)] describes a technical report concerning a civil law issue of leasing contracts dollar-indexed for vehicle purchase. The authors studies aspects related to the validity – from the vehicle buyers' financial point of view – of a legal action against leasing companies. Making use of the Bayesian approach, they build a calculator based on a 2001 jurisprudence[20], 27 variables and the opinion about the judgements. They presents the applications *Check Balance*[21] and *Decision Supporter*[22] in order to provide reference values to the decision maker. The opinion about how favorable are the legal

---

[17]Albeit [Loevinger(1963), p. 8] asserts '(i)t is unnecessary, and perhaps impossible, to give a precise definition to the field of jurimetrics'.

[18]*Instituto de Matemática e Estatística, Universidade de São Paulo* (Institute of Mathematics and Statistics, University of São Paulo).

[19]*Construir um modelo matemático para cálculo da probabilidade de paternidade e implementá-lo em software.* [Nakano(2006), p. 5]

[20]Recurso Especial do STJ n°. 473141.

[21]Verifica Saldo.

[22]Apoiador de Decisão.

decisions to the vehicle buyer or the leasing companies is given in a ordinal scale from 1 (highly favorable to vehicle buyer) to 5 (highly favorable to leasing companies), where 0 indicates ignorance about the decisions. To the levels of the ordinal scale are assigned parameters of a beta probability distribution [Johnson et al.(1995), pp. 210-275].

Figure 6: Decision Supporter window by [Wechsler et al.(2006), p. 49]

[Kadane(2012)] refers to a lecture given in the XI Brazilian Meeting on Bayesian Statistics, concerning the case of Kansas cellphone users[23] in which 'Sprint-Nextel was sued for conspiring with other cell phone providers to impose high prices for text-messaging'. Kadane discusses the probability sampling concept and its application, pointing out that in the present case 'classical statistics did not address the court's question, but Bayesian analysis did'.

Half decade ago [Zabala and Silveira(2014)] present a brief history of what was known so far about jurimetrics. Some considerations and suggestions on practical and theoretical applications of the subject are made. A division of jurimetrics into three prisms is proposed, formalizing a connection between law and quantitative thinking. Some theoretical and applied examples are presented, demonstrating part of the application in different contexts and raising fundamental questions for use in Brazilian law. In 26 April 2016 the same authors purposed a jurimetrics symbol[24] called *Juri-yang*. Presented in Figure 7, it represents the harmony between 'juri' – indicated by three dots forming a stylized balance – and 'metrics', indicated by three aligned dots suggesting a stylized scatter plot.

[Trecenti(2015)] presents the method of influence diagram applied to civil justice litigation. The data were obtained from the web, through web scraping and text mining techniques. The R packages `tjsp` and `bnr` were developed, used respectively to facilitate access to the TJSP[25] database and adjust the parameters of Bayesian networks based on a Gibbs sampler. The author discusses the importance of data extraction and manipulation, as well as the use of open-source packages that allow the reproducibility of jurimetric works.

[Nunes(2016)] discusses how statistics can reinvent the law. The author cites a series of lectures taught in Brazil in 1973 by the philosophy professor at the Universities of Milan and Turin, Mario Losano. According the author, Losano 'did not agree with the idea of the quantification of law and rejected statistics because he considered it to be incompatible with law in all its scope, including the study of general norms, principles and social values'[26], considering that 'predicting the content of a

---

[23]Quin Jackson et al. v. Sprint Nextel Corporation, Case No. 09-cv-2192 (N.D. Ill.).

[24] https://commons.wikimedia.org/wiki/File:Juri-yang.png

[25] *Tribunal de Justiça de São Paulo.*

[26] *(...) não compactuava com a ideia da quantificação do Direito e rejeitava a Estatística por considerá-la incompatível com o Direito em toda sua abrangência, incluindo o estudo das normas gerais, princípios e valores sociais.* [Nunes(2016),
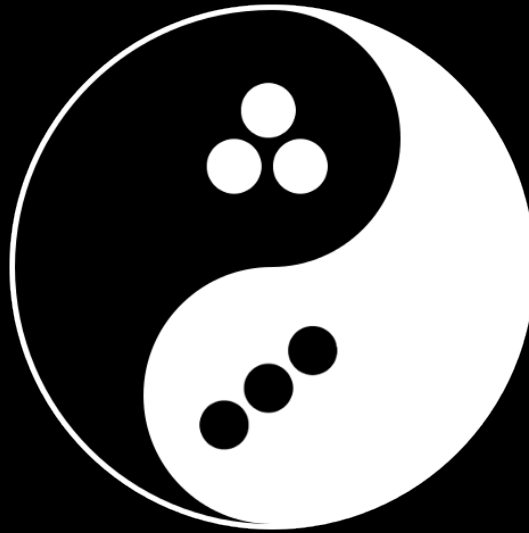
Figure 7: Juri-yang.

sentence would be impracticable'[27]. Considering the expression 'jurimetrics', Nunes points out Lee Loevinger had a 'deterministic view of knowledge'[28] and 'stuck to the concept of deterministic causality, Loevinger understood that uncertainty would deprive the scientist of the means of identifying the causes of a phenomenon, preventing him from making predictions about his future behavior.'[29]

[Marcondes et al.(2019)] discusses the necessity of 'transparency and complete auditability' in judicial systems, considering random procedures to select jury, court or judge. The authors emphasizes 'these principles are neglected by random procedures in some judicial systems, which are performed in secrecy and are not auditable by the involved parties'. As an example of such a procedure is discussed the assignment of cases in the Brazilian Supreme Court, presenting 'a review of how sortition has been historically employed by societies and discusses how Mathematical Statistics may be applied to random procedures of the judicial system'. The authors finalize the article purposing a 'statistical model for assessing randomness in case assignment', applied to the Brazilian Supreme Court.

[Aires et al.(2017)], [Aires(2019)] and [Aires et al.(2019)] develop an approach to identify potential conflicts between norms based on deontic logic. The authors evaluate the effectiveness of the presented techniques 'using a manually annotated contract conflict corpus with results close to the current state-of-the-art for conflict identification, while introducing a more complex classification task of such conflicts' for which their method 'surpasses the state-of-the art method'. [Zabala and Silveira(2019)] present general purpose tools for jurimetrics, with code and databases available in free software. Examples are given in Section 3.3 and in the Figure 2.

1.3 *Gueule de bois*

In 23 March 2019 was enacted in France the Law No. 2019-222, a regulation know as *Article 33*. It provides a penalty up to 5 years of imprisonment, a fine up to 300,000 euros and possible loss of civil rights for anyone who discloses analysis of French judiciary data [France(2019)][30]. According to Article 33,

---

p. 104]

[27] *(...) a previsão do conteúdo de uma sentença seria inviável.* [Nunes(2016), p. 105]

[28] *(...) uma visão determinista do conhecimento.* [Nunes(2016), p. 100]

[29] *Preso ao conceito de causalidade determinística, Loevinger entendia que a incerteza privaria o cientista dos meios de identificação das causas de um fenômeno, impedindo-o de formular previsões a respeito do seu comportamento futuro.* [Nunes(2016), p. 100]

[30] "*Les données d'identité des magistrats et des membres du greffe ne peuvent faire l'objet d'une réutilisation ayant pour objet ou pour effet d'évaluer, d'analyser, de comparer ou de prédire leurs pratiques professionnelles réelles ou supposées. La violation de cette interdiction est punie des peines prévues aux articles 226-18,226-24 et 226-31 du code pénal, sans préjudice des mesures et sanctions prévues par la loi n°. 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés.*"

> [t]*he identity data of magistrates and members of court can not be reuse with the aim of object or effect of evaluating, to analyze, compare or predict their practices real or assumed. Violation of this prohibition is punished by the penalties provided for in Articles 226-18,226-24 and 226-31 of the Penal Code, without prejudice to the measures and sanctions provided for by Law n°. 78-17 of 6 January 1978 relating to data processing, files and freedoms.*

The penalties provided in Article 226-18 [France(2004)][31] points out

> [t]*he collection of personal data by fraudulent, disloyal or unlawful means is punishable by five years' imprisonment and a fine of 300,000 euros.*

The Article 226-31 [France(1994)][32] governs

> [n]*atural persons guilty of one of the offenses provided for in this chapter also incur the following additional penalties: (...) [t]he ban on civil and family rights, (...) the professional or social activity in the exercise or on the occasion of the exercise of which the offense was committed, (...) [and] the confiscation of the thing that served or was intended to commit the offense or the thing that is the product of it.*

On the other hand, Brazil is much more advanced in this sense when compared to France, at least in theory. The brazilian legislation brings the Act 12.527/11, popularly known *Access to Information Act*, promulgated in 18 November 2011 [Brasil(2011)] and regulated in 16 May 2012 [Brasil(2012)]. It systematizes the data access in all public spheres in Brazil, and guarantees to Brazilian citizens the right to request access to public data. In the legal text[33] can be found the following:

> *It is the duty of public bodies and entities to promote, regardless of requirements, the disclosure in a place that is easily accessible, within the scope of their competences, of information of collective or general interest produced or guarded by them. (...) In order to comply with the caput, public bodies and entities shall use all legitimate means and instruments at their disposal, and disclosure on official websites of the World Wide Web (Internet) is mandatory. (...) Sites (...) shall (...) meet (...) the following requirements: I - contain content search tools that allow access to information in an objective, transparent, clear and in easy to understand language; II - enable the recording of reports in various electronic formats, including open and non-proprietary, such as spreadsheets and text, in order to facilitate the analysis of information; III - enable automated access by external systems in open, structured and machine readable formats; IV - disclose in detail the formats used for structuring the information; V - guarantee the authenticity and integrity of the information available for access; VI - keep updated the information available for access.* [Brasil(2011), Article 8]

The authors, however, were unsuccessful in most data requests through official brazilian channels. In 4 December 2014 at 09:46:34, the Administrative Proceeding No. 22.031/2014 has been opened regarding the Manifestation No. 211234/2014. The manifestation mentions the Act 12.527/11 to request access to a sample database for conducting academic work. The solicitation[34] requested the following:

---

[31] "*Le fait de collecter des données à caractère personnel par un moyen frauduleux, déloyal ou illicite est puni de cinq ans d'emprisonnement et de 300 000 euros d'amende.*"

[32] "*Les personnes physiques coupables de l'une des infractions prévues par le présent chapitre encourent également les peines complémentaires suivantes (...) [l]'interdiction des droits civiques, civils et de famille, (...) l'activité professionnelle ou sociale dans l'exercice ou à l'occasion de l'exercice de laquelle l'infraction a été commise, (...) [et] la confiscation de la chose qui a servi ou était destinée à commettre l'infraction ou de la chose qui en est le produit.*"

[33] "*É dever dos órgãos e entidades públicas promover, independentemente de requerimentos, a divulgação em local de fácil acesso, no âmbito de suas competências, de informações de interesse coletivo ou geral por eles produzidas ou custodiadas. (...) Para cumprimento do disposto no caput, os órgãos e entidades públicas deverão utilizar todos os meios e instrumentos legítimos de que dispuserem, sendo obrigatória a divulgação em sítios oficiais da rede mundial de computadores (internet). (...) Os sítios (...) deverão (...) atender (...) aos seguintes requisitos: I - conter ferramenta de pesquisa de conteúdo que permita o acesso à informação de forma objetiva, transparente, clara e em linguagem de fácil compreensão; II - possibilitar a gravação de relatórios em diversos formatos eletrônicos, inclusive abertos e não proprietários, tais como planilhas e texto, de modo a facilitar a análise das informações; III - possibilitar o acesso automatizado por sistemas externos em formatos abertos, estruturados e legíveis por máquina; IV - divulgar em detalhes os formatos utilizados para estruturação da informação; V - garantir a autenticidade e a integridade das informações disponíveis para acesso; VI - manter atualizadas as informações disponíveis para acesso.*"

[34] *a) a base de metadados relativa à base processual e b) a extração de dados públicos (brutos) processuais, com o maior histórico possível. Segue abaixo lista de informações que necessito para o estudo: nome do advogado, comarca, situação, número Themis, número CNJ, parte, classe/natureza, órgão julgador, última movimentação, acórdão, processo principal, processo de 1o grau, relator, data de distribuição, volumes, quantidades de folhas, notas de expediente, último julgamento, dados do 1o grau, depósitos judiciais e última atualização.*

(a) the metadata database of the court lawsuits and (b) the extraction of court lawsuits (raw) public data with the longest possible record. Below is a list of the information I need for the study: name of lawyer, county, situation, Themis number, CNJ number, part, class/nature, judging body, latest move, judgment, main proceedings, 1st degree case, rapporteur, date of distribution, volumes, sheet quantities, file notes, last judgment, 1st degree data, court deposits and last update.

Five days later a feedback was given by TJDFT-COSIST, the First Instance Systems and Statistics Coordination[35] of the Court of Justice of the Federal District and Territories[36]. The document No. 22.031/2014, sent in 9 December 2014, informed about the impossibility to fulfill the request for public databases. The allegations[37] was the following:

In this case, the request does not limit the desired data, and only indicates the extraction of public databases, which is not feasible in view of the number of records of that, and the lack of access of this Coordination to such records for extraction. In addition, there is no integration between the first and second instance computer systems, so verification of some related information here will require manual verification of the processes. Given the above, this Coordination informs the impossibility of fulfilling the request.

Anexo IV – Ofício do Tribunal de Justiça do Distrito Federal e dos Territórios
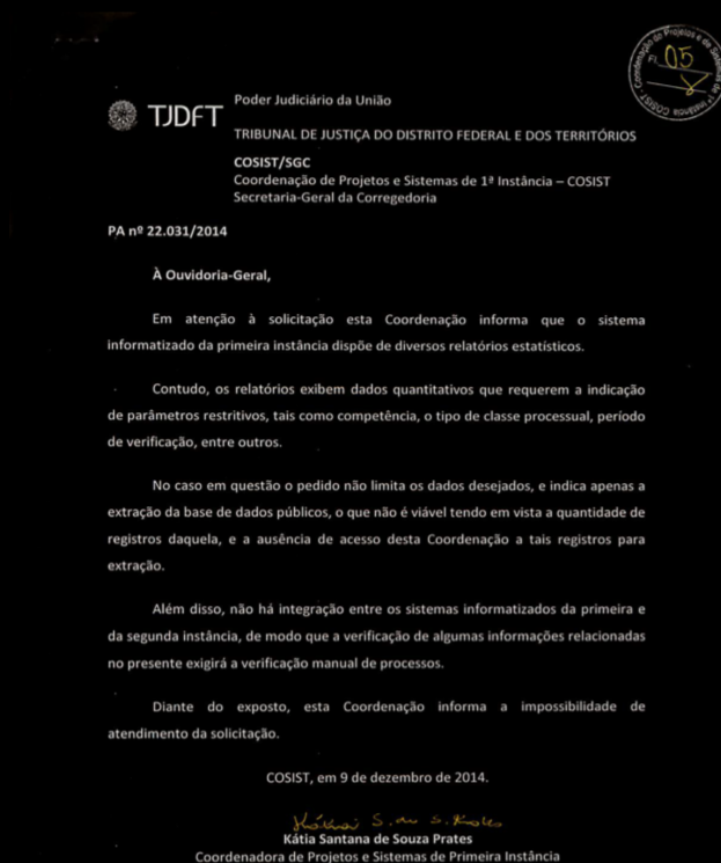

Figure 8: Administrative Proceeding No. 22.031/2014.

---

[35] COSIST – *Coordenadoria de Sistemas e Estatísticas da Primeira Instância.*

[36] TJDFT – *Tribunal de Justiça do Distrito Federal e dos Territórios.*

[37] *No caso em questão o pedido não limita os dados desejados, e indica apenas a extração de bases de dados públicos, o que não é viável tendo em vista a quantidade de registros daquela, e a ausência de acesso desta Coordenação a tais registros para extração. Além disso, não há integração entre os sistemas informatizados da primeira e da segunda instância, de modo que a verificação de algumas informações relacionadas no presente exigirá a verificação manual dos processos. Diante do exposto, esta Coordenação informa a impossibilidade de atendimento da solicitação.*

After some attempts to gain access to public databases, the only useful feedback came from TJMG-SEPLAG-CEINFO, the Institutional Management Information Center[38] and the Executive Secretariat of Planning and Quality in Institutional Management[39] of the Minas Gerais Court[40]. They made available in 5 December 2014 a sample of 11 variables for 5,080,270 TJMG court lawsuits from 2000 to 2013, analyzed by [Oliveira(2014)]. In Section 3.2 it is presented `tjmg_year`, a clean and ready-to-use version of TJMG dataset, concerning the counting of 4,236,229 lawsuits grouped by year.

2 Three Prisms of Jurimetrics

Much of the activity of government, including that of lawmakers, judges, administrators and other lawyers consists of investigating and ascertaining facts. [Loevinger(1966), p. 535]

The *Three Prisms of Jurimetrics* is an approach proposed half decade ago by [Zabala and Silveira(2014)]. The authors intended to unify the subject, in order to give a formal connection between the three main players of the judiciary: judge, lawmaker and lawyer. As the universe of study is wide, the analysis permeates all forms of legal action and needs a didactic way to understanding the role of quantitative methods in the legal framework. The prisms are, therefore, the looks or points of view of the main agents of the judiciary. [Schlag(2009)] considers the 'dedifferentiation problem', in the sense that 'more sophisticated theories of law lead us to a point where we are no longer able to distinguish law from culture, or society, or the market, or politics or anything of the sort'. Anyway, considering the separation of powers model – consecrated on the Articles 2nd and 60th of the Brazilian Federal Constitution[41][42] [Brasil(1988)] – and the Occam's Razor principle, we are still interested in delimiting the study of jurimetrics to the judiciary, although there may be associations of the latter with other entities.

2.1 The Judge

The first prism refers to the judge's view. A judge's main activity is to decide in situations under uncertainty. Considering the intersections between computing, quantitative methods and the law, it is possible to build modern calculators in order to support the magistrate's decision. [DeGroot(2004)] presents methods for *optimal statistical decisions* considering the *subjective probabilities* as numerical representations of the decision maker's beliefs and *utility* as numerical representations of his/her tastes and preferences. The approach given by Morris DeGroot maximizes the expected utility taking the primitives 'is more likely than' for probability and 'is preferred to' for utility, in order to disentangle these two components. The opinion about a latent parameter θ is updated under the Bayesian paradigm.

Judicial expertise can be brought to the judicial process, in order to resolve doubts raised by the judge. Guidelines on the reality of precedents can also be considered, demonstrating the available information on parameters of interest. [Zabala(2019)] 'presents a new tool to support the decision concerning moral damage indemnity values of the judiciary of Rio Grande do Sul, Brazil. A Bayesian approach is given, in order to allow the assignment of the magistrate's opinion about such indemnity amounts, based on historical values. The solution is delivered in free software using public data, in order to permit future audits'.

In addition to the reasons already presented, there is a possibility to consider the contribution of the *expert witness*. The 'expert testimony has become increasingly essential in a wide variety of litigated cases' [Berger(2011)], albeit 'has many virtues, and just as many negative aspects' [Cohen(2015)]. 'The probability of winning increases with the skillful presentation of evidence', argues [Matson(2012)]. [Aitken and Stoney(1991)] states 'the court is concerned with the probabilities (or odds) of guilt, and the expert witness is concerned with the probabilities of the evidence'. Once the opinion of an expert witness is widely accepted, it helps to resolve questions regarding the legal process involving the lack of immediate evidence or even in cases of difficult measurement such as, for instance, cases involving discrimination or prejudice.

---

[38] CEINFO – *Centro de Informações para Gestão Institucional.*
[39] SEPLAG – *Secretaria Executiva de Planejamento e Qualidade na Gestão Institucional.*
[40] TJMG – *Tribunal de Justiça de Minas Gerais,* http://www.tjmg.jus.br/.
[41]"*Art. 2° São Poderes da União, independentes e harmônicos entre si, o Legislativo, o Executivo e o Judiciário.*"
[42]"*Art. 60°, § 4° Não será objeto de deliberação a proposta de emenda tendente a abolir: III - a separação dos Poderes.*"

2.2 The Lawmaker

The second prism refers to the lawmaker's/legislator's view. It is foreseen as one of the main activities of the legislator to purpose data-based bills, considering technical procedures and a legislative impact study. Fortunately, the Brazilian legislation is already taking this path. Know as *Economic Freedom Act* [Brasil(2019)], it determines that proposals for editing and amending normative acts should be accompanied by a regulatory impact analysis.[43]

> *Proposals for editing and amending normative acts in the general interest of economic agents or users of the services rendered, issued by a federal government agency or entity, including municipalities and public foundations, will be preceded by a regulatory impact analysis, which will contain information and data on the possible effects of the normative act to verify the reasonableness of its economic impact.* [Brasil(2019), Art. 5]

In addition to supporting new bills and amendments to normative acts, it is required more technical forms and scientific studies to protect the population from irresponsible arbitrariness. Empirical and technical analysis helps to control the pure rhetoric and demagogy in the production of laws. This kind of approach helps to understand the areas we should prioritize efforts, focus investments and so on.

In order to helping the public administration to understand the judiciary behavior, it is possible to present realized and forecasted values from the available datasets. In Example 3.3.2 there are presented the realized values from January 2000 to December 2017, as well as the forecasting from January 2018 to July 2021. The algorithm uses only the realized values to train and test different models. Available in function `fits` from the `jurimetrics` package, it is a way to easily forecast the court proceedings volume and other univariate time series. Such analysis can be used to anticipate and improve resources allocation, such as provision of funds, agency expenses and all sort of elements related to how provisions should be made. Regarding the management of the judiciary and other public institutions, there are several examples of how the public data analysis can help in the decision making for the coming months and years.

In Brazil, there is an explicit need for data analysis in the discussion about the Civil Procedure Code (CPC) amendment proposals concerning court appeals suppression. The debate revolves around the following question: Does court appeals suppression significantly alter the timing of the process? To the best of our knowledge, there were no proposals considering technical analysis or somehow evaluating financial and social impacts of such suppression over time.

On the other hand, in 2015 the Brazilian Jurimetrics Association[44] identified, in partnership with the National Council of Justice[45], critical points in the child adoption process. The findings motivated the Amendment Project no. 5850/2016 [Brasil(2016)], approved by the Brazilian congress in 22 November 2017 as the Act 13.509/2017 [Brasil(2017)]. This Act impacts the Statute of the Child and Adolescent (ECA) and the Consolidation of Labor Laws (CLT), streamlining judicial procedures related to the dismissal of family power and the adoption of children and adolescents.

The Brazilian Jurimetrics Association also made a study called 'Biggest Litigants', an endeavor to identify the most frequent litigants and pointing out alternative ways to solve this class of legal problems. The study considered data from the website `consumidor.gov.br` and of some Brazilian courts. The authors were honored to contribute to this study by helping to clean up the TJRS databases. The main problem was the lack of a unique identification key for each litigant. In the case, it was necessary to use REGular EXpressions (REGEX) to find patterns and group the litigants by name. According the technical report[46], 'only 20 companies concentrate more than 50% of disputes', and 'in the state of São Paulo 30 companies concentrate more than 70% of the judicial processes'. The findings also indicates 'the telephone companies and financial institutions consistently grouping more than 40% of processes across all surveyed federation units. In addition, it is possible to identify that consumption ratios vary according to the location under study. Most litigation in all economic areas discuss damages for moral damages, which are often associated with improper registration in delinquent databases'.

---

[43]"*Art. 5° As propostas de edição e de alteração de atos normativos de interesse geral de agentes econômicos ou de usuários dos serviços prestados, editadas por órgão ou entidade da administração pública federal, incluídas as autarquias e as fundações públicas, serão precedidas da realização de análise de impacto regulatório, que conterá informações e dados sobre os possíveis efeitos do ato normativo para verificar a razoabilidade do seu impacto econômico.*"

[44]*Associação Brasileira de Jurimetria - ABJ.*

[45]*Conselho Nacional de Justiça - CNJ.*

[46]https://abj.org.br/cases/maiores-litigantes-2/

2.3   The Lawyer

The third prism refers to the lawyer's view. The lawyer can use technical studies to support requirements, evaluate probabilities or even forecast values. This professional can use the quantitative analysis to measure and improve his/her performance. It is currently feasible to demonstrate on court, for instance, the risks that involve a claim if a right anticipation is not granted. Note we are not talking about the end of the legal argument, but its enrichment.

> *It is difficult to imagine what difference it makes to a juror, lawyer, or scientist to be told that a defendant is 218 times more likely than some man chosen at random to be the father of a child or that there is a 99.54% probability that the defendant is the father.* [Loevinger(1992b), p. 343]

The Brazilian Civil Procedure Code observes the use of probability in decisions concerning urgency guardianship. Article 300 considers the danger of damage and risk to the useful result. Such concepts refer to the use of statistics and quantitative methods in general. The law is translated into probabilistic terms, indicating the legislator's intention to treat the legal issue in a technical and scientific manner.

> *Art. 300. The urgency guardianship will be granted when there are elements that evidence the probability of the right and the danger of damage or the risk to the useful result of the process.*[47] [Brasil(2015)]

Another utility of quantitative approach for the benefit of law is – as Holmes Jr., Loevinger and other researchers suggest – the predictability of legal decision. Judgments, systematically analyzed, can provide answers about the probability of success of a claim. Once many lawsuits are filed considering these probabilities, it is important to consider this technical perspective in order to maximize the chances to bring a financial return to the lawyer or office. It happens that in the vast majority of cases the analysis of financial viability of the demand judgment is not properly analyzed, measuring risks inappropriately and often misguiding customers. Considering the analysis of the 'obvious rich' court decisions, there are also several possibilities this kind of databases can bring, such as the analysis of judges profile, projection of indemnity values, estimated time of the process, among others.

In business law offices there is a need to estimate the provision of lawsuit funds. Estimated amount to be disbursed in lawsuits permits proper allocation of money for corporate judicial liabilities. Still considering law offices, one can mention the benefits that quantitative methods can bring to compliance. Jurimetrics can help in risk assessment, strategy development and internal controls, allowing for more objective and verifiable evaluations. This approach supports decisions to be taken, directing efforts towards legal security.

Another major point in legal action through the prism of advocacy – as well other prisms – is the possibility of alternative means for problem solving. Using state of the art technology, new platforms of agreements and arbitration tend to simplify and massify dispute resolutions. Examples of this technology are *Alternative Dispute Resolution* (ADR), *Online Dispute Resolution* (ODR) and *smart contracts*. According to [Wang(2009), p. 23], 'ODR is usually known as any of online ADR, e-ADR, iADR, virtual ADR and cyber ADR. It was technologically developed in the US and Canada, and it is still used mainly in the US'. The concept of smart contracts was coined by [Szabo(1996)], and refers to a computer protocol used to intermediate trackable and irreversible transactions, and was consecrated by [Nakamoto(2008)].

Finally, the modern lawyer must be able to use tools that allow the practitioner to describe a set of instructions to be performed. Spreadsheets and programming languages are examples of this kind of tools. A good way to start is by understanding how spreadsheets work, and what they can do to simplify daily workflows. The most spreadsheets allow programming, usually called 'macros'. The operator can record a set of instructions by using point-and-click, that is automatically converted in written language. The most widely used language in spreadsheets is BASIC and its dialects [Kemeny and Kurtz(1964)]. Besides BASIC, there are dozens of languages able to solve complex problems. A trained operator is capable to build solutions in a fast and scalable way.

---

[47] *Art. 300. A tutela de urgência será concedida quando houver elementos que evidenciem a probabilidade do direito e o perigo de dano ou o risco ao resultado útil do processo.*

3 Applications

3.1 Tools

The collection of tools used by a data analyst can be extensive. It will depends of the objective, scope and resources available. Usually this professional must be able to build their own tools, constantly adapting it to the dynamics of gather, analyze and presenting data. To allow a critical assessment to be carried out, however, it is necessary the considered data-related elements be available for any person.

The verification and reproducibility principles are pillars of science, and was recently discussed by [Pashler and Wagenmakers(2012)], [Baker(2016)], [Munafò et al.(2017)] and [Wasserstein et al.(2019)]. They points out there is as crises of confidence and reproducibility, much due to a misuse of statistics, frequently the p-value misinterpretation. In order to guarantee good practices in science, projects like *The Dryad Digital Repository*[48] defends open, easy-to-use, not-for-profit and community-governed data infrastructure for scholarly literature. On the same line is *The Science Code Manifesto*[49], in which the signatories adopt five principles:

1. **Code** All source code written specifically to process data for a published paper must be available to the reviewers and readers of the paper.

2. **Copyright** The copyright ownership and license of any released source code must be clearly stated.

3. **Citation** Researchers who use or adapt science source code in their research must credit the code's creators in resulting publications

4. **Credit** Software contributions must be included in systems of scientific assessment, credit, and recognition.

5. **Curation** Source code must remain available, linked to related materials, for the useful lifetime of the publication.

There is a plethora of projects based on free and open source software for those who wish to do science. Considering the data analysis and presentation, some examples are given on Table 1. Each software described can be extended with libraries, packages and user defined functions. For instance, recently [Gerum(2019)] presents `Pylustrator`, an open source library tool for Python to generate the code necessary to compose publication figures from single plots. In R language [R Core Team(2019)] is built `ggplot2` [Wickham(2016)], a system for 'declaratively' creating graphics based on *The Grammar of Graphics* [Wilkinson(2005)]. A general purpose tools for jurimetrics can be found at the `jurimetrics` package [Zabala and Silveira(2019)].

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of `Unix-like` platforms and Windows. It is a `GNU`[50] project which is similar to the S language and environment, developed at Bell Laboratories (formerly AT&T, now Lucent Technologies) by John Chambers and colleagues. It follows a minimalist object-oriented concept, which specifies a small standard core accompanied by language extension packages. To install R the reader may access the official website project[51]. It is recommended to always keep R at its latest version as well the using of RStudio editor[52], an R-integrated development environment. Among many tools it enables the creation of automatic presentations and reports in various formats such as pdf, html and docx, merging languages such as R, LaTeX, markdown, C++, Python, SQL and D3. It is available in Desktop and Server editions along with their respective previews, gathering R functionalities harmonically.

The packages used in this article can be installed and updated with the following R code. For Unix-like operating systems, it is recommended to run the following instructions on a terminal after executing the `sudo R` command followed by the admin password.

---

[48] https://datadryad.org
[49] http://sciencecodemanifesto.org/
[50] The GNU General Public License is a type of license used for free software that grants end users (individuals, organizations or companies) the freedom to use, study, share and modify the software.
[51] https://cloud.r-project.org/
[52] https://rstudio.com/

Application	URL
Cytoscape.js	http://js.cytoscape.org/
C++	http://www.cplusplus.com/
D3.js	https://d3js.org/
GNU Octave	https://www.gnu.org/software/octave/
GNU PSPP	https://www.gnu.org/software/pspp/
JASP	https://jasp-stats.org/
Julia	https://julialang.org/
LaTeX	https://www.latex-project.org/
LibreOffice	https://www.libreoffice.org/
Markdown	https://www.markdownguide.org/
MySQL	https://www.mysql.com/
Orange	http://orange.biolab.si/
Python	https://www.python.org/
R	https://www.r-project.org/
Scilab	https://www.scilab.org/
Stan	https://mc-stan.org/
Tabula	https://tabula.technology
TensorFlow	https://www.tensorflow.org/

Table 1: A sample of free software to handle and present data

```
> packs <- c("tidyverse","devtools")
> install.packages(packs, dep = T)
> devtools::install_github("filipezabala/jurimetrics", force = T)
> update.packages(ask = F)
```

3.2 Data

Data is a collection of tables, documents and files, usually transformed in binary format. [Breiman(2001)] asserts '*Statistics starts with data*', that are used '*to predict and to get information about the underlying data mechanism*'. The data gathering is usually laborious, taking much of the total analysis time. Tools like MySQL, Python or R – referenced on Table 1 – may help in data handling. At this point it is considered that the reader has already understood the relevance in mastering high-level languages for the solution of applied problems. In this way it is considered that the reader performs simple operations using command line[53], the reading of technical documentation and executes other eventual tasks, although it can be time-consuming.

Even not all applications needs substantial amount of data, it is commonplace to use a considerable structure of databases and processing. In a personal computer it is possible to create useful tools and deliver solutions for relevant problems. Some databases used by the authors are available in `jurimetrics` package.

---

[53] https://en.wikipedia.org/wiki/Command-line_interface

```
> data("tjmg_year")
> y <- ts(data = tjmg_year$count, start = c(2000), frequency = 1)
> head(y, 10)

Time Series:
Start = 2000
End = 2009
Frequency = 1
 [1]  38403   49560   66653   81005   92012 107442 164101 213774 280847 343614

> sum(y)
[1] 4236229
```

In opposition of data collecting, some practical situations brings the necessity to decide with sample size near to (if not) zero, considering only the previous decision maker's experience and expected utility. The Bayesian approach presents tools to address solutions involving *small* to *big* data structures, using available data to update the opinion using the *likelihood principle*. Such principle asserts the information[54] about some unknown parameter θ, related to an observed variable X, is obtained only through the *likelihood function*. The Bayesian operation calibrates the previous opinion, contained in a *prior distribution* $\pi(\theta)$, with the likelihood function $L(X|\theta)$. The Bayesian then assigns his/her opinion about θ after the data to the *posterior distribution* $\pi(\theta|X)$. The mathematical representation is given by

$$\pi(\theta|X) = \frac{\pi(\theta)L(X|\theta)}{\int_\theta \pi(\theta)L(X|\theta)d\theta} \propto \pi(\theta)L(X|\theta).$$

3.3 Examples

This section brings two examples in which the authors currently work. Code and documentation can be found at github.com/filipezabala/jurimetrics.

3.3.1 Preliminary injunction

Preliminary injunction[55] is a legal instrument used in Brazil to provide the complainant an anticipated right, based on *periculum in mora* (danger in delay) and *fumus boni iuris* (likelihood of success on the merit of the case) principles [Grubbs et al.(2003)]. To base the arguments it is common for lawyers to use rhetoric, but it is possible to consider a quantitative approach for this purpose.

Example 4. *(Protest restraint) Suppose a company has been incorrectly listed on a negative credit bureau. Consider also the customers make the purchase only if the company is not listed in negative bureaus. If a company has (i) an average of 1000 monthly credit bureau consultations, (ii) 10% of the consultations converted in customers and (iii) an average ticket of $3450.00 per client, the expected value of X: 'loss amount per business day of negative credit bureau attribution' can be calculated by*

$$E(X) = \frac{1000 \times 0.1 \times 3450}{22} = 15681.82.$$

The `jurimetrics` package brings the `exp_loss` function, that performs the purposed calculations given the parameters `average.consult`, `prob.hire` and `average.ticket`. The function returns a list of two positions: `expected.value` and `text`. Note that `text` brings a rudiment of an automatic report, generated using the input parameters only. The reader can build their own custom functions combining pre-existing functions.

---

[54] According to [Gosh(1988)], in Basu's sense '*information is what information does. It changes opinion. Only a Bayesian knows how to characterize his/her prior opinion on θ as a prior distribution $q(\theta)$. This prior opinion is changed, by the data x, to the posterior opinion $q^*(\theta) = q(\theta)L(\theta)/\sum q(\theta)L(\theta)$'*.

[55] *Antecipação de tutela*, in Portuguese.

```
> library(jurimetrics)
> (ev <- exp_loss(average.consult = 1000, prob.hire = 0.1, average.ticket = 3450))

$expected.loss
[1] 15682

$text
[1] "The estimated loss amount per business day is $15681.82."

> paste0("The half of the expected loss is ", ev$expected.loss/2, ".")
[1] "The half of the expected loss is 7840.91."
```

☺

3.3.2   Court proceedings volume forecast

In the experience of the authors, a frequently asked question is about the court proceedings volume forecast. Any law operator should consider the amount of proceedings volume in the next months and years, as it impacts the entire chain of the judiciary. There is substancial purposes in literature to address the problem of forecasting values observed over time. The theme is usually referenced as *time series*, and will be considered the treatment given by [Hyndman and Athanasopoulos(2018)] and [Hyndman(2018)].

Available in the `jurimetrics` package, the database `tjrs_year_month` contains a sample of the monthly processual volume of 5,625,666 court lawsuits in a south brazilian court[56] between January 2000 and December 2017. Based on the *Access to Information Act* [Brasil(2011)] and its regulation [Brasil(2012)], the data were obtained from the court website[57] via web scraper written in Python language. The following code attaches the data in a tibble data frame format[58] with 216 rows and 2 columns, `yearMonth` and `count`.

```
> library(jurimetrics)
> data(tjrs_year_month) # attaching data
```

In order to print the first 10 lines, just enter the name of the attached data. Note the information about the class of the object (`tibble`) is a $216 \times 2 \times 1$ dimension tensor. The tibble object also brings information about the columns classes, respectively `<date>` and `<int(eger)>`. In the last line is displayed the number of rows and columns not shown on the screen, in this case 206 rows. The option `print(n = Inf)` prints all the lines, but this procedure is not encouraged.

```
> tjrs_year_month

# A tibble: 216 x 2
    yearMonth   count
    <date>      <int>
 1  2000-01-01     12
 2  2000-02-01    222
 3  2000-03-01    576
 4  2000-04-01    647
 5  2000-05-01   1098
 6  2000-06-01   2307
 7  2000-07-01    107
 8  2000-08-01   4743
 9  2000-09-01   3301
10  2000-10-01   4255
# ... with 206 more rows

> # tjrs_year_month %>% print(n = Inf) # print all the 216 lines
```

As the data is time series-like, it is recommended to convert it in a *time-series object* through the function `ts`. The function transforms the `data` from a `start` date, in this case January 2000 indicated by `c(2000,1)`. The `frequency` of observations must be supplied, in which 12 refers to monthly, 6 to biannual, 3 to quarterly and 1 to yearly observations.

---

[56] TJRS – *Tribunal de Justiça do Rio Grande do Sul.*
[57] http://www.tjrs.jus.br/site/
[58] https://cran.r-project.org/web/packages/tibble/vignettes/tibble.html

At a glance it is possible to observe peaks of lawsuits judged in the months of March, April, August and September, while valleys occur in January. In this sense, we are looking for a model that capture this – and maybe other not so obvious – behaviors to make useful predictions. It is noteworthy, however, that all models are simplifications and idealizations of reality. The idea that complex physical, biological or social systems can be precisely described by a model is unlikely. Building idealized representations that capture important stable aspects of such systems is, however, a vital part of scientific, political and market analysis, driving decisions in situations of uncertainty. The use of modelling usually leads to helpful insights in the achievement of theoretical and practical results. [Box(1979)] points out that '*all models are wrong but some are useful*', suggesting

> [*f*]*or such a model there is no need to ask the question 'Is the model true?'. If 'truth' is to be the 'whole truth' the answer must be 'No'. The only question of interest is 'Is the model illuminating and useful?'* ([Box(1979), pp. 202-203]

```
> (y <- ts(data = tjrs_year_month$count, start = c(2000,1), frequency = 12))

        Jan    Feb    Mar    Apr    May    Jun    Jul    Aug    Sep    Oct    Nov    Dec
2000     12    222    576    647   1098   2307    107   4743   3301   4255   5017   4819
2001    220   3153   4823   4054   5868   5084    199   6271   3958   4816   5344   5366
2002    218   3185   4793   4812   5962   6059    296   5914   5057   8068   9574  13298
2003   1293   8726  12639  14259  14241  16940    953  18850  17196  20102  18526  20440
2004    754  11137  24722  20646  20489  25490   3686  23938  23300  21315  21965  25790
2005    760  11998  26251  22408  21814  29089  18033  26014  23623  22118  25293  25723
2006  15794  12219  27052  25350  28483  28662  23002  34301  27000  28537  32323  31973
2007   7171  21820  34138  29551  34852  30634  26801  36776  30171  34086  32273  29654
2008  12780  20196  32816  34032  33247  35545  34177  45412  38270  45866  39761  33548
2009  18204  17460  43022  43083  35652  41244  38571  42821  43666  40700  39835  32906
2010  22982  20864  43883  38889  44821  37571  39317  40314  38927  38610  39392  34991
2011  20010  27162  37795  36449  38173  42342  35974  44597  40687  37855  37043  34255
2012  17173  24373  42825  37704  45831  40870  32243  45656  34618  43379  37585  32612
2013  16343  21962  38792  39893  38117  39811  31737  41841  36519  43293  38288  31896
2014  16136  22415  33838  36927  36900  26766  35550  40608  38130  42740  37506  37183
2015  17096  29100  41133  39881  33891  36962  36145  38498  38599  21560  44918  31096
2016  13455  23148  35486  29451  32326  43092  37561  42904  38194  38019  37958  31796
2017   8626  27212  44352  32129  36362  33478  33534  41154  32806  37405  36977  28945

> sum(y)
[1] 5625666
```

The tidy data can then be applied in the function `fits` in order to adjust the best model from classes `ARIMA`, `ETS`, `TBATS` and `NNETAR`. The strategy is to fit/train models with a percentage corresponding to the first points of the time series and comparing the remaining last points with the prediction made by each considered model. The comparison is made using the *mean squared error* (MSE) of prediction, given by

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

where n is the number of remaining/test points of the time series, y_i is the ith tested value and $\hat{y}_i$ the ith predicted value by a given model.

The parameter `train` is set by default to $0.8 = 80\%$, remaining around 20% of the points to test. The model that leads to prediction with the smallest MSE is declared the best model. The function returns a list with four positions: `$fcast`, `$mse.pred`, `$best.model` and `$runtime`. The `$fcast` contains the predicted time series using the model that minimizes the MSE. `$mse.pred` shows the MSE for each considered model, and `$best.model` specifies the best model, in this case from the class neural network autoregressive. As suggested in Example 4, it is possible to arrange the functions to compose automatic reports.

```
> (fit <- fits(y, show.main.graph = F))

$fcast
        Jan   Feb   Mar   Apr   May   Jun   Jul   Aug   Sep   Oct   Nov   Dec
2018  14148 28343 39777 34743 37086 35660 35452 38980 35254 37454 37383 32396
2019  17767 29523 38800 36150 37260 36734 36541 38143 36520 37473 37510 34842
2020  21250 30964 38464 36825 37340 37215 37088 37798 37105 37498 37545 36256
2021  25157 33409 38250 37172 37397 37418 37344

$mse.pred
   mse.pred.aa mse.pred.ets mse.pred.tb mse.pred.nn
1     1.13e+08     68219968     68947440     31652653

$best.model
[1] "nnetar"

$runtime
Time difference of 3.019 secs
```

The parameter `steps` is set by default to `NULL`, in this case projecting the number of points corresponding to the complementary percentage defined in `train`. The user may set `steps` to a value greater or equal than `train*length(x)`, never smaller. The `max.points` limits the maximum number of points to be used, set to 500 by default. If the parameter `show.main.graph` is set to `TRUE`, the main plot is presented, and `show.sec.graph = TRUE` prints the intermediate plots. If `show.value = TRUE`, the system prints the textual results of the function. The `PI = TRUE` prints the prediction intervals used in `nnar` models, but may take long time processing. Finally, `theme_doj = TRUE` will use the Decades Of Jurimetrics theme, based on `ggplot2::theme`.

As shown in Figure 9, the black line indicates the observed points, meanwhile the blue line indicates the predicted values. The $NNAR(3, 1, 2)\,[12]$ model has inputs y_{t-1}, y_{t-2}, y_{t-3} and y_{t-12}, and two neurons in the hidden layer[59]. This means to make the predictions the model looks 1, 2, 3 and 12 months in the historical series, capturing monthly, bimonthly, quarterly and yearly behaviors.
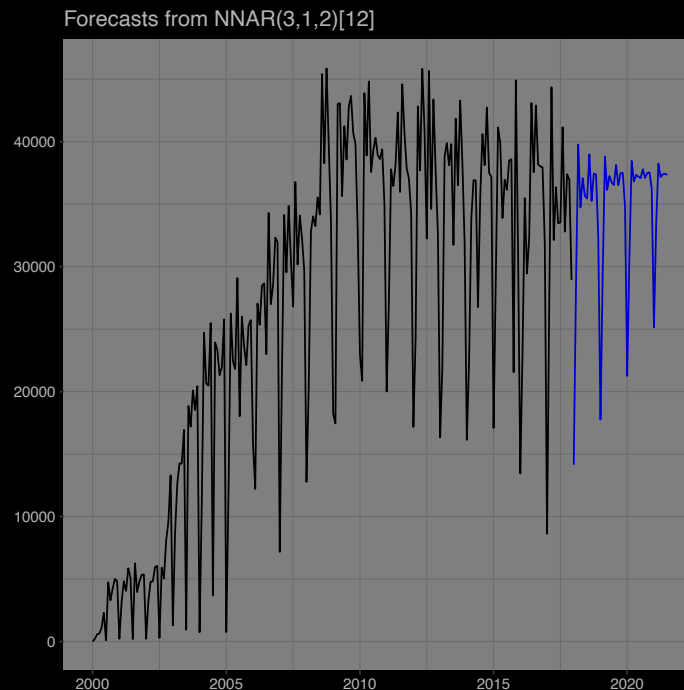

Figure 9: Monthly proceedings volume forecast in TJRS Court.

---

[59] According to [Hyndman and Athanasopoulos(2018)], a $NNAR(p, P, k)\,[m]$ model has inputs y_{t-1}, y_{t-2}, ..., y_{t-p}, y_{t-m}, y_{t-2m}, ..., y_{t-Pm} and $k = (p + P + 1)/2$ neurons in the hidden layer. A $NNAR(p, P, 0)\,[m]$ model is equivalent to an $ARIMA(p, 0, 0)(P, 0, 0)\,[m]$ model but without the restrictions on the parameters that ensure stationarity.

4 The Posterior Step

There was a man in our town
and he was wondrous wise:
he jumped into a BRAMBLE BUSH
and scratched out both his eyes—
and when he saw that he was blind,
with all his might and main
he jumped into another one
and scratched them in again.

4.1 Ignorance

In the morning a children ask: *is the man wondrous wise?*

Two states may be considered:

$$W : \text{the man } \textit{is} \text{ wondrous wise}$$
$$\overline{W} : \text{the man } \textit{is not} \text{ wondrous wise}$$

As Dennis Lindley conveniently points out on the foreword of [de Finetti(1974), p. ix], 'we shall all be Bayesian by 2020'. In advance, [Ferrie(2019)] purposes a 'babyesian' approach. So, trying to answer the children's question under this point of view, initially equal weights are assigned to the two states, let's say 'fifty-fifty'. This is called *Bayes-Laplace prior* and can be represented by

$$P(W) = \frac{50}{100} \qquad\qquad P(\overline{W}) = \frac{50}{100}$$

Under this perspective it is possible to answer 'I'm 50% sure that the man is wondrous wise'.

4.2 Learning

In the afternoon, a children ask: *is the man wondrous wise?*

Even belonging to a privileged environment, it is hard to believe that 50% of people are wondrous wise. Considering such an environment, the guess may turn around, let's say... 20-80. It can be represented by

$$P(W) = \frac{20}{100} \qquad\qquad P(\overline{W}) = \frac{80}{100}$$

Under this perspective it is possible to answer 'I'm 20% sure that the man is wondrous wise'.

Reading the poem again, it can be noted the man deliberately jumped into a bramble bush and scratched out both his eyes, going blind. Considering this information, the guess is updated to 10-90. It can be represented by

$$P(W) = \frac{10}{100} \qquad\qquad P(\overline{W}) = \frac{90}{100}$$

Under this perspective it is possible to answer 'I'm 10% sure that the man is wondrous wise'.

Continuing to read the poem, it can be noted the man jumped deliberately into a bramble bush *again*. Under this perspective, the guess is updated to 5-95. It can be represented by

$$P(W) = \frac{5}{100} \qquad\qquad P(\overline{W}) = \frac{95}{100}$$

Under this perspective it is possible to answer 'I'm 5% sure that the man is wondrous wise'.

Googling it is found a Wikipedia article[60] called 'There Was a Man in Our Town', indicating the poem is originally an English nursery rhyme[61]. The article reports '(i)t is believed to be based on the Greek myth of Bellerophon'[62]. Assuming this, the guess is updated to 40-60. It can be represented by

$$P(W) = \frac{40}{100} \qquad\qquad P(\overline{W}) = \frac{60}{100}$$

---

[60] https://en.wikipedia.org/wiki/There_Was_a_Man_in_Our_Town
[61] https://en.wikipedia.org/wiki/Nursery_rhyme
[62] https://en.wikipedia.org/wiki/Bellerophon

Under this perspective it is possible to answer 'I'm 40% sure that the man is wondrous wise'.

Continuing the reading, it is found Bellerophon bravely captured Pegasus and slaved Chimera. The guess is updated to 70-30. It can be represented by

$$P(W) = \frac{70}{100} \qquad\qquad P(\overline{W}) = \frac{30}{100}$$

Under this perspective it is possible to answer 'I'm 70% sure that the man is wondrous wise'.

Continuing to read the article, it is informed Bellerophon was arrogant, angered Zeus, lived out his life in misery and... had fallen into a thorn bush causing him to become blind. The guess is then updated to 1-99. It can be represented by

$$P(W) = \frac{1}{100} \qquad\qquad P(\overline{W}) = \frac{99}{100}$$

Under this perspective it is possible to answer 'I'm 1% sure that the man is wondrous wise'.

4.3 Knowledge

At night, a children ask: *is the man wondrous wise?*

Before you answer, you unintentionally bump into in a book called *The Bramble Bush*. The subtitle is *Some Lectures on Our Law and Its Study*. The author is *Karl N. Llewellyn*, referenced as [Llewellyn(1930a)]. What is your posterior step?

References

[Aires et al.(2017)] João Paulo Aires, Daniele Pinheiro, Vera Strube De Lima, and Felipe Meneguzzi. Norm conflict identification in contracts. Artificial Intelligence and Law, 25(4):397–428, 2017. URL https://link.springer.com/article/10.1007/s10506-017-9205-x.

[Aires et al.(2019)] João Paulo Aires, Roger Granada, Juarez Monteiro, Rodrigo Coelho Barros, and Felipe Meneguzzi. Classification of contractual conflicts via learning of semantic representations. In Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems, pages 1764–1766. International Foundation for Autonomous Agents and Multiagent Systems, 2019. URL http://www.meneguzzi.eu/felipe/pubs/aamas-conflict-classification-2019.pdf.

[Aires(2019)] João Paulo de Souza Aires. Automatic Reasoning Over Contract Clauses. PhD thesis, PUCRS, School of Technology, 2019. URL https://github.com/mir-pucrs/concon-front-end.

[Aitken and Stoney(1991)] Colin GG Aitken and David A Stoney. The Use of Statistics in Forensic Science. CRC Press, 1991. URL https://www.crcpress.com/The-Use-Of-Statistics-In-Forensic-Science/Aitken-Stoney/p/book/9780139337482.

[Allen(1963)] Layman E. Allen. Beyond document retrieval toward information retrieval. Minnesota Law Review, 47:713, 1963. URL https://digitalcommons.law.yale.edu/fss_papers/4515.

[Allen and Caldwell(1963)] Layman E. Allen and Mary Ellen Caldwell. Modern logic and judicial decision making: a sketch of one view. Law & Contemporary Problems, 28:213, 1963. URL https://digitalcommons.law.yale.edu/fss_papers/4816.

[Amunategui(2014)] Godofredo Iommi Amunategui. Symbolic Languages and Ars Combinatoria. arXiv:1411.3188, 2014. URL https://arxiv.org/abs/1411.3188.

[Austen-Smith and Banks(1996)] David Austen-Smith and Jeffrey S Banks. Information aggregation, rationality, and the condorcet jury theorem. American political science review, 90(1):34–45, 1996. URL https://authors.library.caltech.edu/67312/1/2082796.pdf.

[Baker(2016)] Monya Baker. Why scientists must share their research code. Nature News, 2016. URL https://www.nature.com/news/why-scientists-must-share-their-research-code-1.20504.

[Bayes(1763)] Thomas Bayes. An essay towards solving a problem in the doctrine of chances. by the late rev. mr. bayes, frs communicated by mr. price, in a letter to john canton, amfr s. Philosophical transactions of the Royal Society of London, (53):370–418, 1763. URL https://www.ias.ac.in/article/fulltext/reso/008/04/0080-0088.

[Bench-Capon(2015)] Trevor JM Bench-Capon. Knowledge-based systems and legal applications, volume 36. Academic Press, 2015. URL https://link.springer.com/article/10.1007/BF00150230.

[Berg(1993)] Sven Berg. Condorcet's jury theorem, dependency among jurors. Social Choice and Welfare, 10(1):87–95, 1993. URL https://link.springer.com/article/10.1007/BF00187435.

[Berg(1996)] Sven Berg. Condorcet's Jury Theorem and the reliability of majority voting. Group Decision and Negotiation, 5(3):229–238, 1996. URL https://link.springer.com/article/10.1007/BF02400945.

[Berger(2011)] Margaret A. Berger. The Admissibility of Expert Testimony. National Academies Press, 2011. URL https://www.fjc.gov/content/admissibility-expert-testimony.

[Bernoulli(1713)] Jakob Bernoulli. Ars Conjectandi. Impensis Thurnisiorum, Fratrum, 1713. URL https://archive.org/details/bub_gb_kz9nvk99EWoC.

[Bernoulli(1709)] Nicolas I Bernoulli. Dissertatio inauguralis mathematico-juridica: De Usu Artis Conjectandi in Jure. a Mechel, 1709. URL https://play.google.com/store/books/details?id=svVIAAAAcAAJ&rdid=book-svVIAAAAcAAJ&rdot=1.

[Birnbaum(1962)] Allan Birnbaum. On the Foundations of Statistical Inference. Journal of the American Statistical Association, 57(298):269–306, 1962. URL https://www.jstor.org/stable/2281640.

[Boland(1989)] Philip J Boland. Majority Systems and the Condorcet Jury Theorem. Journal of the Royal Statistical Society: Series D (The Statistician), 38(3):181–189, 1989. URL https://www.jstor.org/stable/2348873.

[Box(1979)] George E.P. Box. Robustness in the strategy of scientific model building. In Robustness in Statistics, pages 201–236. Elsevier, 1979. URL https://www.sciencedirect.com/science/article/pii/B9780124381506500182.

[Brasil(1988)] Brasil. Constituição. Brasília (DF), 1988. URL http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm.

[Brasil(2011)] Brasil. Lei n0. 12.527, de 18 de novembro de 2011, 2011. URL http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm.

[Brasil(2012)] Brasil. Decreto n0. 7.724, de 16 de maio de 2012, 2012. URL http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2012/Decreto/D7724.htm.

[Brasil(2015)] Brasil. Lei no. 13.105, de 16 de março de 2015, 2015. URL http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2015/lei/l13105.htm.

[Brasil(2016)] Brasil. Projeto de lei no. 5850/2016, de 20 de setembro de 2019, 2016. URL https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=2092189.

[Brasil(2017)] Brasil. Lei no. 13.509/2017, de 22 de novembro de 2017, 2017. URL http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2017/Lei/L13509.htm.

[Brasil(2019)] Brasil. Lei no. 13.874, de 20 de setembro de 2019, 2019. URL http://www.planalto.gov.br/ccivil_03/_ato2019-2022/2019/lei/L13874.htm.

[Breiman(2001)] Leo Breiman. Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author). Statistical Science, 16(3):199–231, 2001. URL https://projecteuclid.org/download/pdf_1/euclid.ss/1009213726.

[Cane and Kritzer(2010)] Peter Cane and Herbert Kritzer. The Oxford Handbook of Empirical Legal Research. OUP Oxford, 2010. URL https://global.oup.com/academic/product/the-oxford-handbook-of-empirical-legal-research-9780199542475.

[Cohen(2015)] Kenneth S. Cohen. Expert Witnessing and Scientific Testimony: A Guidebook. CRC Press, 2015. URL https://www.crcpress.com/Expert-Witnessing-and-Scientific-Testimony-A-Guidebook-Second-Edition/Cohen/p/book/9781498721066.

[Condorcet(1785)] Marquis de Condorcet. Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix. Paris, Imprimerie Royale, 1785. URL https://archive.org/details/essaisurlapplica00cond/page/n6.

[de Finetti(1937)] Bruno de Finetti. La prévision: ses lois logiques, ses sources subjectives. In Annales de l'institut Henri Poincaré, volume 7, pages 1–68, 1937. URL http://www.numdam.org/article/AIHP_1937__7_1_1_0.pdf.

[de Finetti(1974)] Bruno de Finetti. Theory of Probability - A critical introductory treatment, volume 1. John Wiley & Sons, 1974. URL https://www.wiley.com/en-us/Theory+of+Probability%3A+A+critical+introductory+treatment-p-9781119286370.

[de Finetti(1975)] Bruno de Finetti. Theory of Probability - A critical introductory treatment. Technical report, 1975. URL https://www.wiley.com/en-br/Theory+of+Probability%3A+A+critical+introductory+treatment-p-9781119286370.

[DeGroot(2004)] Morris H DeGroot. Optimal Statistical Decisions, volume 82. John Wiley & Sons, 2004.

[Donoho(2015)] David Donoho. 50 years of Data Science. 2015. URL http://courses.csail.mit.edu/18.337/2015/docs/50YearsDataScience.pdf.

[Epstein and Martin(2014)] Lee Epstein and Andrew D Martin. An Introduction to Empirical Legal Research. Oxford University Press, 2014. URL https://global.oup.com/academic/product/an-introduction-to-empirical-legal-research-9780199669066?cc=br&lang=en&.

[Feldman and Kim(2005)] Allan M. Feldman and Jeonghyun Kim. The Hand Rule and United States v. Carroll Towing Co. Reconsidered. American Law and Economics Review, 7(2):523–543, 10 2005. ISSN 1465-7252. doi: 10.1093/aler/ahi017. URL https://doi.org/10.1093/aler/ahi017.

[Ferrie(2019)] C. Ferrie. Bayesian Probability for Babies. Baby University. Sourcebooks, Incorporated, 2019. ISBN 9781492680796. URL https://books.google.com.br/books?id=-ee7vgEACAAJ.

[France(1994)] France. Article 226-31, Loi no. 94-653 du 29 juillet 1994 - art. 8 JORF 30 juillet 1994, 1994. URL https://www.legifrance.gouv.fr/affichCodeArticle.do?idArticle=LEGIARTI000006418012&cidTexte=LEGITEXT000006070719&dateTexte=19940730.

[France(2004)] France. Article 226-18, Loi no. 2004-801 du 6 août 2004 - art. 14 JORF 7 août 2004, 2004. URL https://www.legifrance.gouv.fr/affichCodeArticle.do?idArticle=LEGIARTI000006417968&cidTexte=LEGITEXT000006070719&dateTexte=20040807.

[France(2019)] France. Article 33, Loi no. 2019-222 du 23 mars 2019, 2019. URL https://www.legifrance.gouv.fr/eli/loi/2019/3/23/2019-222/jo/article_33.

[France(1829)] Département de la Justice France. Compte Général de L'administration de la Justice Criminelle en France. De L'imprimerie Royale, Paris, 1829. URL http://gallica.bnf.fr/ark:/12148/bpt6k54760067.

[Friendly(2008)] Michael Friendly. A.-M. Guerry's Moral Statistics of France: Challenges for Multivariable Spatial Analysis. arXiv:0801.4263, 2008. URL https://arxiv.org/abs/0801.4263.

[Gehrlein(2006)] William V Gehrlein. Condorcet's Paradox. Springer, 2006. URL https://www.springer.com/gp/book/9783540337980.

[Gehrlein and Lepelley(2011)] William V Gehrlein and Dominique Lepelley. <u>Voting Paradoxes and Group Coherence: the Condorcet Efficiency of Voting Rules</u>. Springer Science & Business Media, 2011. URL `https://www.springer.com/gp/book/9783642031069`.

[Gerum(2019)] Richard Gerum. pylustrator: code generation for reproducible figures for publication. <u>arXiv:1910.00279</u>, 2019. URL `https://arxiv.org/abs/1910.00279`.

[Gosh(1988)] JK Gosh. Statistical Information and Likelihood: A collection of Critical Essays by Dr. D. Basu. <u>Lecture Notes in Statistics</u>, 45, 1988. URL `https://www.springer.com/gp/book/9780387967516`.

[Gottlieb and Hussain(2015)] Klaus Gottlieb and Fez Hussain. Voting for image scoring and assessment (visa)-theory and application of a $2+1$ reader algorithm to improve accuracy of imaging endpoints in clinical trials. <u>BMC medical imaging</u>, 15(1):6, 2015. URL `https://www.ncbi.nlm.nih.gov/pubmed/25880066`.

[Grofman et al.(1983)] B. Grofman, G. Owen, and S.L. Feld. Thirteen theorems in search of the truth. <u>Theory and decision</u>, 15(3):261–278, 1983. URL `https://link.springer.com/article/10.1007/BF00125672`.

[Grubbs et al.(2003)] Shelby R. Grubbs et al. <u>International Civil Procedure</u>. Kluwer Law International, 2003. URL `https://books.google.com.br/books?id=67xM08c9JLgC&pg=PA108&redir_esc=y#v=onepage&q&f=false`.

[Guerry(1833)] André-Michel Guerry. <u>Essai sur la statistique morale de la France</u>. A Paris, Chez Crochard, Libraire, 1833. URL `https://archive.org/details/essaisurlastatis00gueruoft`.

[Hald(2003)] Anders Hald. <u>A History of Probability and Statistics and their applications before 1750</u>, volume 501. John Wiley & Sons, 2003. URL `https://www.wiley.com/en-us/A+History+of+Probability+and+Statistics+and+Their+Applications+before+1750-p-9780471471295`.

[Hand(1947)] Learned Hand. United States v. Carroll Towing Co., 159 f.2d 169 (2d cir. 1947). 1947. URL `https://law.justia.com/cases/federal/appellate-courts/F2/159/169/1565896/`.

[Holmes Jr.(1881)] Oliver Wendell Holmes Jr. The Common Law, 1881. URL `https://www.gutenberg.org/files/2449/2449-h/2449-h.htm`.

[Holmes Jr.(1897)] Oliver Wendell Holmes Jr. The Path of the Law. <u>10 Harvard Law Review 457</u>, 1897. URL `http://moglen.law.columbia.edu/LCS/palaw.pdf`.

[Hyndman(2018)] Rob J. Hyndman. <u>fpp2: Data for "Forecasting: Principles and Practice"</u>, 2nd edition, 2018. URL `https://CRAN.R-project.org/package=fpp2`. R package version 2.3.

[Hyndman and Athanasopoulos(2018)] Rob J. Hyndman and George Athanasopoulos. <u>Forecasting: principles and practice</u>. OTexts, 2018. URL `https://otexts.com/fpp2`.

[Jasanoff(1992)] Sheila Jasanoff. What judges should know about the sociology of science. <u>Jurimetrics Journal</u>, 32:345, 1992. URL `https://www.jstor.org/stable/29762258`.

[Jefferys and Berger(1992)] William H Jefferys and James O Berger. Ockham's Razor and Bayesian Analysis. <u>American Scientist</u>, 80(1):64–72, 1992. URL `https://www.jstor.org/stable/pdf/29774559.pdf`.

[John of Salisbury and Lejeune(2009)] Bishop of Chartres John of Salisbury and François Lejeune. <u>Metalogicon (1159)</u>. Presses de l'Université Laval, 2009. URL `https://archive.org/details/JeanDeSalisburyMetalogicon/page/n0`.

[Johnson et al.(1995)] NL Johnson, S Kotz, and N Balakrishnan. Continuous univariate distributions, vol. 2, johnson wiley & sons, 1995. URL `https://www.wiley.com/en-ag/Continuous+Univariate+Distributions,+Volume+2,+2nd+Edition-p-9780471584940`.

[Kadane(2012)] Joseph B. Kadane. Probability sampling in legal cases: Kansas cellphone users. In <u>AIP Conference Proceedings</u>, volume 1490, pages 7–12. AIP, 2012. URL `https://doi.org/10.1063/1.4759584`.

[Kalven et al.(1966)] Harry Kalven, Hans Zeisel, Thomas Callahan, and Philip Ennis. The American Jury. Little, Brown Boston, 1966. URL https://archive.org/details/americanjury00kalv.

[Kaniovski and Zaigraev(2011)] Serguei Kaniovski and Alexander Zaigraev. Optimal jury design for homogeneous juries with correlated votes. Theory and decision, 71(4):439–459, 2011. URL https://link.springer.com/article/10.1007/s11238-009-9170-2.

[Kasner and Newman(1940)] Edward Kasner and James Roy Newman. Mathematics and the Imagination. Simon and Schuster Inc., 1940. URL https://archive.org/details/mathematicsimagi00kasnrich.

[Kaye(1992)] David H. Kaye. Proof in Law and Science. Jurimetrics Journal, 32:313, 1992. URL https://elibrary.law.psu.edu/cgi/viewcontent.cgi?article=1055&context=fac_works.

[Kemeny and Kurtz(1964)] John G. Kemeny and Thomas E. Kurtz. BASIC: A Manual for BASIC, the elementary algebraic language designed for use with the Dartmouth Time Sharing System. Dartmouth College Computation Center, 1964. URL http://bitsavers.trailing-edge.com/pdf/dartmouth/BASIC_Oct64.pdf.

[Kohli(1975)] Von K. Kohli. Kommentar zur Dissertation von Niklaus Bernoulli: De Usu Artis Conjectandi in Jure. In Herausgegeben von der Natuforschenden Gesellchaft, editor, Die Werk von Jakob Bernoulli - Band 3, chapter K8, pages 541–556. Birkhäuser Verlag, Basel, 1975. URL https://www.semanticscholar.org/paper/Kommentar-zur-Dissertation-von-Niklaus-Bernoulli%3A-Kohli/248c35093047fd27faa7b18fa72dc5a970de8e2b.

[Koller(2004)] David Koller. Origin of the name "Google". 2004. URL http://graphics.stanford.edu/~dk/google_name_origin.html.

[Kowalski(1995)] Robert A Kowalski. Legislation as Logic Programs. In Informatics and the Foundations of Legal Reasoning, pages 325–356. Springer, 1995. URL https://www.doc.ic.ac.uk/~rak/papers/law.pdf.

[Kritzer(2009)] Herbert M Kritzer. Empirical Legal Studies before 1940: a bibliographic essay. Journal of Empirical Legal Studies, 6(4):925–968, 2009. URL https://scholarship.law.umn.edu/cgi/viewcontent.cgi?article=1007&context=faculty_articles.

[Ladha(1992)] Krishna K Ladha. The Condorcet Jury Theorem, Free Speech, and Correlated Votes. American Journal of Political Science, pages 617–634, 1992. URL https://www.jstor.org/stable/2111584.

[Laplace(1774)] Pierre-Simon Laplace. Mémoire sur la Probabilité des Causes par les événements. Mémoires de l'Académie royale des Sciences de Paris (Savants étrangers), 6:621–656, 1774. URL https://gallica.bnf.fr/ark:/12148/bpt6k77596b/f32.

[Leeuw and Schmeets(2016)] Frans L Leeuw and Hans Schmeets. Empirical legal research: A guidance book for lawyers, legislators and regulators. Edward Elgar Publishing, 2016. URL https://searchworks.stanford.edu/view/11663036.

[Leibniz(1666)] Gottfried Wilhelm Leibniz. Dissertatio de Arte Combinatoria. Leipzig: Johann Simon Fick and Johann Polycarp Seubold (1890), 1666. URL https://archive.org/details/ita-bnc-mag-00000844-001.

[Leonard(1980)] Tom Leonard. The roles of inductive modelling and coherence in Bayesian statistics. Trabajos de Estadistica Y de Investigacion Operativa, 31(1):537–555, 1980. URL https://link.springer.com/content/pdf/10.1007/BF02888367.pdf.

[List and Goodin(2001)] Christian List and Robert E Goodin. Epistemic Democracy: Generalizing the Condorcet Jury Theorem. Journal of political philosophy, 9(3):277–306, 2001. URL https://onlinelibrary.wiley.com/doi/full/10.1111/1467-9760.00128.

[Llewellyn(1930a)] Karl Nickerson Llewellyn. The Bramble Bush: some lectures on law and its study. Columbia University School of Law, 1930a. URL https://home.heinonline.org/titles/Legal-Classics/Bramble-Bush--Some-Lectures-on-Law-and-Its-Study/.

[Llewellyn(1930b)] Karl Nickerson Llewellyn. A Realistic Jurisprudence – The Next Step. Columbia Law Review, 30:431, 1930b. URL https://www.jstor.org/stable/1114548.

[Loevinger(1949)] Lee Loevinger. Jurimetrics – The Next Step Forward. Minnesota Law Review, 33:455–470, 1949. URL https://scholarship.law.umn.edu/cgi/viewcontent.cgi?article=2795&context=mlr.

[Loevinger(1950)] Lee Loevinger. The Semantics of Justice, 1950. URL https://www.jstor.org/stable/42581335.

[Loevinger(1952)] Lee Loevinger. An Introduction to Legal Logic. Indiana Law Journal, 27(4):1, 1952. URL https://www.repository.law.indiana.edu/cgi/viewcontent.cgi?article=2372&context=ilj.

[Loevinger(1958)] Lee Loevinger. Facts, Evidence and Legal Proof. Case Western Reserve Law Review, 9:154, 1958. URL https://scholarlycommons.law.case.edu/cgi/viewcontent.cgi?article=3731&context=caselrev.

[Loevinger(1961)] Lee Loevinger. Jurimetrics: Science and Prediction in the Field of Law. Minnesota Law Review, 46:255, 1961. URL https://www.semanticscholar.org/paper/1961-Jurimetrics-%3A-Science-and-Prediction-in-the-of-Loevinger/bb504a3ae2a072d725f2b7463aaf2ca4794adfd1.

[Loevinger(1962)] Lee Loevinger. Occam's Electric Razor. MULL: Modern Uses of Logic in Law, pages 209–214, 1962. URL https://www.jstor.org/stable/29760907.

[Loevinger(1963)] Lee Loevinger. Jurimetrics: The Methodology of Legal Inquiry. Law & Contemporary Problems, 28:5, 1963. URL https://scholarship.law.duke.edu/cgi/viewcontent.cgi?article=2945&context=lcp.

[Loevinger(1966)] Lee Loevinger. Law and Science as Rival Systems. Jurimetrics Journal, pages 63–82, 1966. URL https://www.jstor.org/stable/29761074.

[Loevinger(1985)] Lee Loevinger. Science, technology and law in modern society. Jurimetrics Journal, 26:20, 1985. URL https://www.jstor.org/stable/29761943.

[Loevinger(1992a)] Lee Loevinger. On Logic and Sociology. Jurimetrics Journal, 32:527, 6 1992a. URL https://heinonline.org/HOL/LandingPage?handle=hein.journals/juraba32&div=36&id=&page=.

[Loevinger(1992b)] Lee Loevinger. Standards of Proof in Science and Law. Jurimetrics Journal, pages 323–344, 3 1992b. URL https://www.jstor.org/stable/29762257.

[Lopes et al.(2016)] Lucelene Lopes, Paulo Fernandes, and Renata Vieira. Estimating term domain relevance through term frequency, disjoint corpora frequency - tf-dcf. Knowledge-Based Systems, 97:237–249, 2016. URL https://doi.org/10.1016/j.knosys.2015.12.015.

[Loschi and Wechsler(2002)] Rosangela H Loschi and Sergio Wechsler. Coherence, bayes's theorem and posterior distributions. Brazilian Journal of Probability and Statistics, pages 169–185, 2002. URL https://www.jstor.org/stable/43601015.

[Marcondes et al.(2019)] Diego Marcondes, Cláudia Peixoto, and Julio Michael Stern. Assessing randomness in case assignment: the case study of the Brazilian Supreme Court. Law, Probability and Risk, 18(2-3):97–114, 2019. URL https://doi.org/10.1093/lpr/mgz006.

[Maron and Kuhns(1960)] Melvin Earl Maron and John Larry Kuhns. On relevance, probabilistic indexing and information retrieval. Journal of the ACM (JACM), 7(3):216–244, 1960. URL https://dl.acm.org/citation.cfm?id=321035.

[Matson(2012)] Jack V Matson. *Effective Expert Witnessing: Practices for the 21st Century.* CRC Press, 2012. URL `https://www.crcpress.com/Effective-Expert-Witnessing-Practices-for-the-21st-Century/Matson/p/book/9781439887677`.

[Mayo(2014)] Deborah G Mayo. On the Birnbaum Argument for the Strong Likelihood Principle. *arXiv:1411.3188*, 2014. URL `https://arxiv.org/abs/1302.7021`.

[Mehl(1959)] Lucien Mehl. Automation in the legal world. In 'mechanisation of thought processes', volume 1. proceedings of a symposium held at the national physical laboratory on 24th, 25th, 26th and 27th november 1958. pages 755–788, 1959. URL `https://archive.org/details/mechanisationoft01nati`.

[Merton(1993)] Robert K Merton. *On the Shoulders of Giants: The post-Italianate edition.* University of Chicago Press, 1993. URL `https://www.press.uchicago.edu/ucp/books/book/chicago/O/bo3641858.html`.

[Montoya-Delgado et al.(2001)] Luis E Montoya-Delgado, Telba Z Irony, Carlos A de B Pereira, and Martin R Whittle. An unconditional exact test for the Hardy-Weinberg equilibrium law: sample-space ordering using the Bayes Factor. *Genetics*, 158(2):875–883, 2001. URL `https://www.genetics.org/content/genetics/158/2/875.full.pdf`.

[Montoya-Delgado(1998)] Luis Eduardo Montoya-Delgado. *Probabilidade de Paternidade – Uma proposta metodológica para o seu cálculo.* PhD thesis, IME-USP – Instituto de Matemática e Estatística da Universidade de São Paulo, 1998.

[Munafò et al.(2017)] Marcus R. Munafò, Brian A. Nosek, Dorothy V.M. Bishop, Katherine S. Button, Christopher D. Chambers, Nathalie Percie Du Sert, Uri Simonsohn, Eric-Jan Wagenmakers, Jennifer J. Ware, and John P.A. Ioannidis. A manifesto for reproducible science. *Nature Human Behaviour*, 1(1):0021, 2017. URL `https://www.nature.com/articles/s41562-%20016-0021`.

[Nakamoto(2008)] Satoshi Nakamoto. Bitcoin: A Peer-to-Peer Electronic Cash System. 2008. URL `https://bitcoin.org/bitcoin.pdf`.

[Nakano(2006)] Fábio Nakano. *Um Novo Modelo para Cálculo de Probabilidade de Paternidade - Concepção e Implementação.* PhD thesis, IME-USP – Instituto de Matemática e Estatística da Universidade de São Paulo, 2006. URL `https://teses.usp.br/teses/disponiveis/95/95131/tde-09032007-172617/en.php`.

[Nunes(2016)] Marcelo Guedes Nunes. *Jurimetria: como a Estatística pode reinventar o Direito. São Paulo: Editora Revista dos Tribunais*, 2016. URL `https://www.livrariart.com.br/jurimetria-9788520367940/p`.

[Oliveira(2014)] Islaine Sonemann Fernandes de Oliveira. Jurimetria: Uma aplicação para a base de dados do TJ-MG. Technical report, PUCRS, Mathematics Faculty, 2014. URL `http://www.aclassica.com/jurimetrics/oliveira2014jurimetria.pdf`.

[Pashler and Wagenmakers(2012)] Harold Pashler and Eric–Jan Wagenmakers. Editors' Introduction to the Special Section on Replicability in Psychological Science: A Crisis of Confidence? *Perspectives on Psychological Science*, 7(6):528–530, 2012. doi: 10.1177/1745691612465253. URL `https://doi.org/10.1177/1745691612465253`. PMID: 26168108.

[Pereira and Stern(1999)] Carlos A. de Bragança Pereira and Julio Stern. Evidence and Credibility: Full Bayesian Significance Test for Precise Hypotheses. *Entropy*, 1(4):99–110, 1999. URL `https://www.mdpi.com/1099-4300/1/4/99`.

[Poisson(1837)] Siméon Denis Poisson. *Recherches sur la probabilité des jugements en matière criminelle et en matière civile precédées des règles générales du calcul des probabilités.* Bachelier, 1837. URL `https://ia800209.us.archive.org/27/items/recherchessurlap00pois/recherchessurlap00pois.pdf`.

[R Core Team(2019)] R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2019. URL https://www.R-project.org/.

[Schlag(2009)] Pierre Schlag. The dedifferentiation problem. Continental Philosophy Review, 42(1): 35–62, 2009.

[Schlegel(1995)] John Henry Schlegel. American Legal Realism and Empirical Social Science. Univ of North Carolina Press, 1995. URL https://uncpress.org/book/9780807857533/american-legal-realism-and-empirical-social-science/.

[Spärck Jones(1972)] Karen Spärck Jones. A statistical interpretation of term specificity and its application in retrieval. Journal of documentation, 28(1):11–21, 1972. URL http://dx.doi.org/10.1108/eb026526.

[Stigler(1980)] Stephen M Stigler. Stigler's Law of Eponymy. Transactions of the New York academy of sciences, 39(1 Series II):147–157, 1980. URL http://doi.wiley.com/10.1111/j.2164-0947.1980.tb02775.x.

[Stigler(1986)] Stephen M. Stigler. The History of Statistics: The Measurement of Uncertainty before 1900. Harvard University Press, 1986. URL https://www.hup.harvard.edu/catalog.php?isbn=9780674403413.

[Stiles(1961)] H Edmund Stiles. The Association Factor in Information Retrieval. Journal of the ACM (JACM), 8(2):271–279, 1961. URL https://dl.acm.org/citation.cfm?id=321074.

[Szabo(1996)] Nick Szabo. Smart contracts: building blocks for digital markets. EXTROPY: The Journal of Transhumanist Thought,(16), 18:2, 1996. URL http://www.fon.hum.uva.nl/rob/Courses/InformationInSpeech/CDROM/Literature/LOTwinterschool2006/szabo.best.vwh.net/smart_contracts_2.html.

[Trecenti(2015)] Julio Adolfo Zucon Trecenti. Diagramas de influência: uma aplicação em Jurimetria. PhD thesis, IME-USP – Instituto de Matemática e Estatística da Universidade de São Paulo, 2015.

[Wang(2009)] Faye Fangfei Wang. Online Dispute Resolution: Technology, management and legal practice from an international perspective. Chandos, 2009. URL https://books.google.com.br/books/about/Online_Dispute_Resolution.html?id=3B9hmAEACAAJ&source=kp_book_description&redir_esc=y.

[Wasserstein et al.(2019)] Ronald L Wasserstein, Allen L Schirm, and Nicole A Lazar. Moving to a world beyond "p< 0.05", 2019. URL https://www.tandfonline.com/doi/full/10.1080/00031305.2019.1583913.

[Wechsler et al.(2006)] S. Wechsler, D.K. Colombo, F.V. Bonassi, and L.G.M. Reginato. Relatório de análise estatística sobre o projeto: 'Análise Econômica do Direito aplicada a decisões judiciais: o caso dos contratos de arrendamento mercantil para compra de veículos com cláusulas de reajuste associadas ao dólar', 2006.

[Wechsler et al.(2008)] Sérgio Wechsler, Carlos Alberto de Bragança Pereira, and P. C. F. Marques. Birnbaum's Theorem Redux, 2008. URL https://www.ime.usp.br/~pmarques/papers/redux.pdf.

[Wickham(2016)] Hadley Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016. ISBN 978-3-319-24277-4. URL https://ggplot2.tidyverse.org.

[Wilkinson(2005)] Leland Wilkinson. The Grammar of Graphics. Springer-Verlag, New York, USA, 37, 2005. URL https://www.springer.com/gp/book/9780387245447.

[Williams(2004)] David Williams. Condorcet and Modernity. Cambridge University Press, 2004. URL https://books.google.com.br/books/about/Condorcet_and_Modernity.html?id=MC1InwEACAAJ&redir_esc=y.

[Winston(1985)] Roger B Winston. A suggested procedure for determining order of authorship in research publications. Journal of Counseling and Development, 63(8):515–518, 1985. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/j.1556-6676.1985.tb02749.x.

[Zabala(2019)] Filipe J. Zabala. A Bayesian Application in Judicial Decisions. arXiv:1912.10566, 2019. URL https://arxiv.org/abs/1912.10566.

[Zabala and Silveira(2014)] Filipe J. Zabala and Fabiano F. Silveira. Jurimetrics: Statistics Applied in the Law (in portuguese). Revista Direito e Liberdade, 16(1):87–103, 2014. URL http://www.esmarn.tjrn.jus.br/revistas/index.php/revista_direito_e_liberdade/article/view/732.

[Zabala and Silveira(2019)] Filipe J. Zabala and Fabiano F. Silveira. jurimetrics: General Purpouse Tools for Jurimetrics, 2019. URL https://www.github.com/filipezabala/jurimetrics. R package version 0.1.0.

[Zaigraev and Kaniovski(2012)] Alexander Zaigraev and Serguei Kaniovski. A note on the probability of at least k successes in n correlated binary trials. Operations Research Letters, 41(1):116–120, 2012. URL https://serguei.kaniovski.wifo.ac.at/fileadmin/files_kaniovski/pdf/corr_binom.pdf.