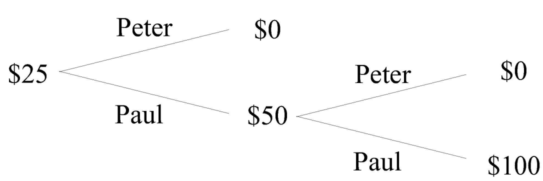
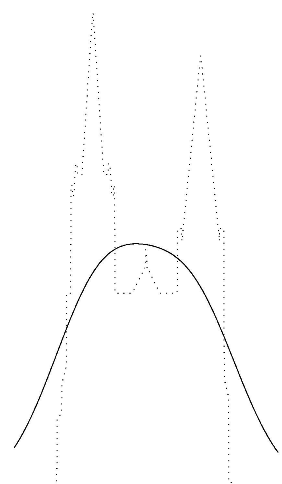
When to call a variable random

Glenn Shafer

gshafer@business.rutgers.edu
www.glennshafer.com





The Game-Theoretic Probability and Finance Project

Working Paper #41

First posted June 21, 2015. Last revised March 5, 2018.

Project web site:
http://www.probabilityandfinance.com

Abstract

Kolmogorov's axioms have enabled probability to flourish as pure mathematics for more than half a century. But today there is substantial interest in broadening Kolmogorov's framework to accommodate applications involving phenomena that are not fully probabilized. This raises questions of interpretation and terminology. Can the different interpretations of probability, objective and subjective, that have accepted Kolmogorov's framework also be accommodated in a broader framework? How should the terms *event*, *variable*, *expectation*, and *process* be used in a broader framework? Should all uncertain quantities be called *random variables*? Should uncertain quantities that change over time be called *stochastic processes*?

In this working paper, I look at how the historical record can help us with these questions. This record shows how Kolmogorov's framework developed out of the mathematics of betting in games of chance, and how its flexible vocabulary emerged from a two complimentary and sometimes competing projects: the project of making mathematical probability into pure mathematics, and the project of using it to understand statistical realities.

A broader framework, which again begins with the mathematics of betting but treats it with the rigor of modern game theory, can also flourish as pure mathematics while accommodating diverse applications, involving both subjective and objective interpretations of probability. The terms *event*, *variable*, and *process* can be retained. Whereas Kolmogorov's framework assumes that enough bets are offered to define probabilities for all events and expectations for all variables, a broader framework can allow fewer bets to be offered, so that some events have only upper and lower probabilities rather than probabilities, and some variables have only upper and lower expectations rather than expectations. I argue that in order to maintain continuity with the accomplishments and wisdom accumulated within Kolmogorov's framework, we should call variables *random* and processes *stochastic* only when they are fully probabilized.

Subjective and objective interpretations of the bets offered can be associated with the viewpoints of two different players. Taking the viewpoint of the player who offers the bets, we can regard them as expressions of belief. Taking the viewpoint of the player to whom they are offered, we can regard them as a theory to be challenged and tested by betting strategies that attempt to multiply existing capital.

In spite of its length, this remains a working paper. Its scope as a historical study is so vast that I have surely made significant errors of interpretation and omission. I am hoping that other historians of probability and statistics will help me make it more accurate. I have included in red many reminders to myself about points that need further research or consideration.

Contents

Part I
Introduction

In the first half of the twentieth century, interactions between mathematicians and statisticians produced an enduring framework for mathematical probability. Based on measure theory and functional analysis, this framework was put in its definitive form by Andrei Kolmogorov in 1933 [219] and elaborated, in a way that allowed it to handle continuous time successfully, by Joseph Doob in 1953 [127].

Kolmogorov's framework has enabled probability theory to flourish as a purely mathematical enterprise for more than half a century. Why? Because it gave purely mathematical form to the rich language of probability, allowing mathematicians to use this language without falling into the philosophical conundrums that arise when it is used outside mathematics.[1][1]

The need for such mathematical autonomy is clear, for the philosophical conundrums that arise in applications of mathematical probability are themselves durable. For three centuries, the most penetrating philosophers of probability have recognized its dual meaning: subjective and objective.[2] We encounter this duality even in games of chance, where the mathematical theory originated. The probability of drawing a winning hand ten times in a row in a certain game may be very small. This means we do not believe it will happen. But this small probability surely also has some objective meaning. Is it not a fact about the game, a feature of the world external to our beliefs? And what is the nature of this objectivity?[3] Is it a material fact or some more hidden reality? To do mathematics, we must find a way to set these questions aside.

Yet the boundary that Kolmogorov and Doob drew around their mathematics may have been too tight. While fencing out philosophy, they also excluded probabilistic language and ideas that are important in applications and can be mathematized.

Variables within Kolmogorov's framework all have probability distributions; they are *random variables*. Variables that do not have probability distributions remain outside the framework; they are *not random*. Consequently, most work that uses probability, even if it remains theoretical, lies only partly within the framework. In mathematical statistics, some dependent variables may be random only conditional on the values of independent variables that cannot be considered random for a variety of reasons. In control theory and Markov decision theory, some variables may be random only conditional on a policy for making control decisions. In quantum mechanics, the randomness (and even existence) of some measurements depends on what an observer decides to observe. Statisticians, engineers, physicists, and other scientists have found various ways to bring mathematical results legitimized within Kolmogorov's framework to bear on the settings where such non-random variables also play a role, but it cannot be said that the framework fully embraces the mathematical structures

[1]Endnotes begin on page 102.

these researchers are using.

Although this status quo has proven workable, there is now substantial interest in broader frameworks for probability, frameworks that accommodate both fully probabilized events and variables and others that are probabilized only partially or not at all. Some work in this direction, such as the work on "outer contents" in functional analysis [186], has avoided probabilistic language and ideas. Other proposals for broader frameworks have been tied to particular philosophies of probability or particular applications. The most popular of these, perhaps, are Peter Walley's framework for "imprecise probabilities", which is based on a subjective interpretation of probability and has inspired a growing literature [374, 359, 12], and the work on "coherent measures of risk" in finance [11].

As mathematics, a broader framework needs the same mathematical autonomy that Kolmogorov's framework has enjoyed. But it also needs to be framed in a way that permits it to be used in applications by people who bring different philosophies of probability and different purposes to the table. Like Kolmogorov's framework, it needs a language that can be understood in different ways by different philosophies.

My 2001 book with Vladimir Vovk, *Probability and Finance; It's Only a Game*, proposed a broadening of Kolmogorov's framework based on the modern mathematical theory of games. This theory was not available to Kolmogorov in 1933; it emerged only with the publication of a classical book by von Neumann and Morgenstern in 1944 [365]. But it is now recognized as a legitimate branch of pure mathematics, fully as rigorous even if less prestigious than older and more austere branches such as functional analysis. In the 2001 book and in subsequent working papers (see `www.probabilityandfinance.com`), Vovk and other researchers, including myself, have shown that game theory can replace Kolmogorov's axioms as a basis for the classical limit theorems that are used to model repetitive phenomena and test statistical hypotheses.

Much remains to be done to make this game-theoretic framework as accessible and usable as possible both to mathematicians and to those who use probablity theory in other domains. The framework's mathematical potential is yet to be fully realized, and its language needs to be developed in a way that permits both mathematicians and users with different philosophies and applied problems to see how it embraces their own philosophical and practical understandings of probability.

This working paper is based premised on two convictions:

- We can best see how our diverse conceptions of probability can be interpreted in modern game theory if we come with some perspective on how these conceptions evolved from probability's roots in calculations of value in games of chance.

- An effective way to understand this historical development without losing our way in mathematical detail is to focus on the framework's most basic vocabulary: how this vocabulary arose and how it developed into the central concepts of Kolmogorov's and Doob's framework: *random variable*, *expectation*, and *stochastic process*.

Part II sketches this historical development. The story crosses countries and languages as well as different philosophies of probability. The meaning of *expectation* evolved over three centuries, beginning with Christian Huygens's use of the Latin *expectatio* to refer to a player's chance or chances. The term *random variable* emerged from similar terms in Italian, Russian, German, and French, often used by authors who migrated from one country or language to another. The adjective *stochastic* was first used in English by a Russian, for whom it denoted a particular way of thinking about objective probability.

In Part III, I draw lessons from this history for game-theoretic generalization of Kolmogorov's framework. I argue that we can use a language of betting in a way that preserves the openness to a wide range of philosophical views on probability that has distinguished Kolmogorov's framework and retains the wisdom accumulated over the past century by those who have used random and non-random variables together in applied work.

Part II
History

> Le calcul de probabilités ... est le premier pas des mathématiques hors du domaine de la vérité absolue.
>
> Irenée Jules Bienaymé, 1853[4]

To understand fully the evolution that produced Kolmogorov's framework, we must begin long before Kolmogorov's time. In this working paper, I begin, as most histories of probabilities do, in the seventeenth century, when Christian Huygens first used the Latin *expectatio* for a gambler's expectation. The ensuing story will take us across many countries and many languages. Figures familiar from other histories of probability and statistics, such as Laplace, Galton, Kolmogorov, and Neyman, and Doob, will play important roles. Other figures, somewhat less familiar, will also star, including Crofton, Chuprov, Slutsky, and Darmois.

In spite of its chronological breadth, this will not be a comprehensive history of probability and statistics. We will scarcely touch, for example, on the development and use of Bayes's theorem, because this important theme was not central to the development of Kolmogorov's framework.[5] Nor will we dwell on the details of Kolmogorov's axioms or on other early twentieth-century efforts to axiomatize probability.[6] Instead, we will focus on the basic concepts of Kolmogorov's framework: *random variable*, *expectation*, and *stochastic process*—on the words themselves and on the conceptual issues involved in their adoption by Kolmogorov's predecessors and contemporaries.

Although this emphasis on vocabulary may seem superficial, it will lead us into conceptual issues that have continued relevance in the twenty-first century. Moreover, by looking at what earlier scholars actually wrote, we will gain insights that are difficult for historians to convey when they seek to explain mathematical details of older work to contemporary readers. Hans Fischer acknowledges this difficulty as follows ([147], page 12):

> One particular difficulty in completing studies on the history of probability theory and statistics consists in the fact that, for the sake of succinctness and clarity, some 18th and 19th century contributions must be presented in the modern terminology to which the reader is accustomed. In particular, this relates to the use of stochastic terms like "random variable,"...

In this working paper, since we are not looking at mathematical details, we can avoid taking the modern concept of a random variable for granted in discussing its creation. In order to avoid creating any misunderstanding about what words particular scholars used when referring to what we might now call a random variable, I will call such an object—a quantity that takes different values with different probabilities (i.e., has a probability distribution) a *probabilized quantity.* To my knowledge, this is a term no one has previously used, and I propose to use it only in this historical exercise.

I begin with *expectation* in the seventeenth century because this concept, including an understanding of its additivity, is older than the concept of mathematical probability or any concept of a probabilized quantity. We study the early understanding of *expectation* in §1, with attention to Christian Huygens, Gabriel Cramer, and Sylvestre Lacroix.

The concept of a probabilized quantity emerged in the eighteenth century, in the study of the theory of errors. In §2 we look how this concept was discussed by the celebrated French scholars Laplace, Poisson, and Cournot and by a scarcely remembered German scholar, Carl Friedrich Hauber. They all distinguished between a probabilized quantity and its values. It was the values, not the probabilized quantity, that they called *fortuit* in French and *zufällig* in German—words that would be translated into English, then and now, as *random.*

In §3, we look at how the notion of expectation expanded in the nineteenth century. By the end of the century, Russian mathematicians were using *matematicheskoe ozhidanie* (mathematical expectation) as a general name for the theoretical mean of a probabilized quantity. A few of them also began applying the adjective случайно (random) to a probabilized quantity (rather than merely to its values). The most prominent of these was Pavel Nekrosov, who tied probability to a religious conception of free will.

In §4, we look at the struggle of statistics in the late nineteenth century to deal with the aftermath of Poisson's attempt to explain statistical stability. As we will see, Bienaymé, Lexis, and others who tried to use Poisson's work on the variability of chances in the study of statistical series were not able to fashion a satisfactory bridge between probability theory and statistics. As Stephen M.

Stigler showed in *History of Statistics* (1986 [352]), two were first put together successfully in Fechner's psychophysics and the biometry of the British school led by Galton and Pearson. One key to there success, perhaps, was the fact that they were dealing primarily with the concept of a random sample, which can be seen as a collection of random values from a single probabilized quantity with a stable probability distribution, rather than the concept of statistical series, a sequence of demographic or economic counts or measurements that have no such stability.

This brings us to the twentieth century, when name *random variable* and its modern definitions finally emerged in English. As we will see, it emerged from the interplay of many cross currents and competing goals. Two competing and sometimes conflicting projects stand out:

1. A project associated with the mathematical statisticians Ladislaus von Bortkiewicz and Aleksandr Chuprov and later with Harald Cramér and Jerzy Neyman: to make full use of mathematical probability in statistics.

2. A project associated with the mathematician Andrei Kolmogorov and later with Joseph Doob: to make probability a branch of pure mathematics that can be pursued without attention to issues involved in its applications.

Bortkiewicz and Chuprov believed that statistical theory needed to develop into a new science of stochastics, which would bring to statistical practice tools that mathematicians had developed to make probability theory more rigorous and to better understand the limiting behavior of averages when the number of observations grows to infinity. For mathematicians studying probability, on the other hand, the challenge was to take advantage of emerging mathematical tools, such as functional analysis and measure theory, to further explore the infinite frontier, both the frontier as the number of observations grows through time and the frontier as we observe changes on so fine a scale that we consider time continuous. In order to pursue this project, Kolmogorov and many of his fellow mathematicians sought a purely mathematical framework for probability, which would isolate their reasoning from questions about the applicability of its assumptions to practice.

Our discussion of these twentieth century developments is organized by language. Beginning with Cantelli's introduction of the name *variabile casuale* in the 1910s, which I discuss in §5, successive languages acquired standard names for probabilized quantities. In §6, we look at Chuprov introduction of *zufällige Variable* in German; in §7 we look at Darmois's introduction of *variable aléatoire* in French. In §8, we look at how Kolmogorov's *Grundbegriffe*, the German monograph in which he introduced his axiomatic framework, was situated in this process, and in §9, we look at how *random variable* won out in English over alternatives that make seem more logical.

In §10, I review the results of this evolution, especially the vocabulary now used in Kolmogorov's framework. Having issued from the dialectic between these two projects, this vocabulary retains some old contradictions and has produced some new ones.

Appendix 11 provides some perspective on the linguistic diversity of our story by counting the references to work in different languages in some of the most important histories of probability and statistics.

1 Expectation in games of chance

Many authors in probability and mathematical statistics, even in the twentieth century, did their work without using any general name for probabilized quantities.[2] In applications, context usually provided names for the quantities being considered: gamblers' gains, errors in measurement, lengths of life, etc. When someone did assign a name, formally or informally, to a class of probabilized quantities, they were usually looking for a way to define and calculate an expected or typical or mean value for the quantities in the class. What is the value now of my uncertain future gain? How long can I expect to live? How great is the probable error of a measurement or estimate?

1.1 Huygens's expectations and their values

As we know, our mathematical theory of probability began with games of chance, and it was here that mean values were first calculated. In his *De ratiociniis in ludo aleae*, published in Latin in 1657, Christiaan Huygens (1629–1695) showed how to find the value to a player of a game in which he has certain chances to win or lose money. Today we might say that the player's gain is a random variable and that Huygens calculated its expected value. But Huygens was actually doing something entirely different and much more principled. In *De ratiociniis in ludo aleae*, the player's situation is called his *sors* or his *expectatio*. The question was how much this *sors* or *expectatio* is worth. Huygens answered that it is worth the amount of money it would cost to construct it from bets that are clearly fair because they treat the players symmetrically. Huygens showed, using arguments like those Pascal had used a few years earlier in his letters to Fermat, that such constructions are possible and give a unique price.

De ratiociniis in ludo aleae was reproduced and discussed by James Bernoulli (1654–1705) in his *Ars Conjectandi*, published in 1713.[3] Bernoulli introduced the notion of numerical probability into the discussion, and then it became

[2]In this working paper, I use *probabilized quantity* to mean any quantity that takes different values with different probabilities or is thought to do so, regardless of whether those values or probabilities are known or how they might be specified. To my knowledge, this term has not been used by previous authors, and I am not recommending it to future authors. But it will help me discuss the work of various authors without inadvertently suggesting that they made assumptions associated with any term that was or is in use, such *random variable*. In the same spirit, I will use the term *theoretical mean* for the expected value of the probabilized quantity.

[3]Huygens wrote *De ratiociniis in ludo aleae* in Dutch; it was translated into Latin and published by Frans van Schooten. Huygens suggested the Latin term *expectatio* to van Schooten, who used both it and *sors*. For more detail on the terminology used by Huygens and Bernoulli, see Edith Sylla's introduction to her translation of *Ars Conjectandi* [23], especially pages 112–123.

possible to say that the values Huygens had determined were equal to the sum of the different possible values of an unknown quantity, each multiplied by its probability. Today it seems natural, even obligatory, to generalize this formula beyond the case where the unknown quantity is a player's gain or loss. In order to understand the subsequent history, we must recognize that this was not natural in the seventeenth and eighteenth centuries, because the calculation had no meaning outside the gambling picture where Huygens's constructions could be carried out.

Van Schooten's *sors* and *expectatio* were translated into the European vernaculars in various ways. The most influential translation was probably that of Abraham De Moivre (1667–1754), who settled on *expectation* as a translation of *expectatio* in his *Doctrine of Chances* [118], first published in 1718. Like Huygens, De Moivre distinguished between an *expectation* and its value, but he often simplified *value of an expectation* to *expectation*, so that *expectation* became a number. This double meaning of *expectation*, as a number and the bargain valued by that number, was common in English well into the nineteenth century.

The first author to follow Huygens with a book in French was Pierre de Montmort (1678–1719), who published his *Essay d'analyse sur les jeux de hazard* in 1708 [267]. Montmort mainly used the French *sort*, evidently a translation of *sors*, though he also used *espérance*, which can be considered a translation of *expectatio*.[4]

1.2 Cramer's mathematical and moral expectation

In 1713, Jacob Bernoulli's nephew Nicolas Bernoullli (1687–1759) wrote to Montmort with a counterexample to Huygens's method later called the *Saint Petersburg paradox*: a contract that pays 2^{n-1} with probability $1/2^n$ for $n = 1, 2, \ldots$ has infinite value according to Huygens's principles, but no one would pay an unlimited price for it.[5] The ensuing discussion eventually led to the celebrated *Specimen theoriae novae de mensura sortis* [21], published in Latin by Daniel Bernoulli (1700–1782) in the journal of the imperial academy in Saint Petersburg. Bernoulli took account of aversion to large losses and the diminishing utility of money by calculating expectations on a logarithmic scale, applied not to net gains and losses but to the player's resulting total fortune. At the end of his article, Bernoulli quoted from a letter written in French in 1728 by Gabriel Cramer (1704–1752), who had called Huygens's *expectatio* the *espérance mathématique* and contrasted it with the *espérance morale*, which takes utility into account.

The terms *espérance mathématique* and *espérance morale* were taken up by in the 1770s by Pierre Simon Laplace (1749–1827). He apparently first used the terms in a memoir that he read to the Royal Academy in Paris in 1773 and

[4] We might render *sors* and *sort* into English as "lot" or "fate". The most natural translation of *espérance* is "hope".

[5] I need to check what words were used by Nicolas and Montmort in their correspondence, in *Die Werke von Jakob Bernoulli*, Band 3.

that was published in 1776 [230] ([236], page 147). He used them consistently thereafter; they appear on page 187 of the first edition of his *Théorie analytique des probabilités* [232]. Other French authors on probability followed his lead.

The task of understanding, clarifying, and simplifying Laplace's great work on probability was undertaken across Europe in the first half of the nineteenth century, and those undertaking this task translated the terms *espérance mathématique* and *espérance morale* into their own vernaculars.

In his 1837 treatise in English, which appeared in the *Encyclopedia Metropolitana*, Augustus De Morgan (1806–1871) made the translation that was most natural, given that *expectation* had already been established by De Moivre: *mathematical expectation* and *moral expectation*. (See [119], 1849 edition, page 408, and also [120], page 97.) The article on probability in the seventh edition of the *Encyclopedia Britannica* also used these terms; it was published separately in 1839 by its author, Thomas Galloway (1796–1851) [162].

The cultural transmission from France to Germany was implemented in part by the translation of French treatises [313], in which *espérance* was translated more literally, as *Hoffnung* (hope). We see *mathematische Hoffung* and *moralische Hoffnung* in the translations of Sylvestre Lacroix's *Traitée* in 1818 ([227], page 127), Laplace's *Essai philosophique* in 1819 [234], Siméon-Denis Poisson's *Recherches* in 1841 ([301], page), and Augustin Cournot's *Exposition* in 1849 ([79], pages 76–77). These terms, *mathematische Hoffung* and *moralische Hoffnung*, persisted in the most influential German treatises on probability into the twentieth century.

Ludwig Oettinger (1797–1869), the professor at the University of Freiburg whose treatise on probability was published in 1852 [286] after having appeared in installments in the *Journal für die reine und angewandte Mathematik* in the 1840s, cited De Moivre and Poisson as well as Laplace and mostly followed De Moivre by writing consistently about a player's *Erwartung* and its *Werth* (value), explaining as follows (page 191):

> Der *Werth der Erwartung* wird auch mit dem Namen *"mathematische Hoffnung"* bezeichnet und diese der *"moralische"* gegenübergestellt. Zweckmässsiger würde der Name *"objective Hoffnung"* sein. Auch liesse sich der Werth der Erwartung durch *"mittlerer Werth"* oder *"Durchschnitts-Werth"* bezeichnen. Für viele Fälle passt die Benennung *mittlerer Werth* recht gut.

Occasionally (pages 207–209), *Werth der Erwartung* became *Erwartungswerth*.

The Russian terminology appears to have been set by Viktor Yakovlevich Bunyakovskii (1804–1889), in his Основания Математическоѣ Теории Вероятностеѣ, published in 1846 in Saint Petersburg [46]. In his preface, Bunyakovskii mentions that writing the book had required him to coin many new Russian terms. He used математическое ожидание for *mathematical expectation* and нравственное ожидание for *moral expectation*. The Russian ожидание is the most natural translation of the English *expectation*. Since Bunyakovskii had studied in France with Cauchy, it is notable that he did not instead use надежда, the Russian word for *hope*. He may have been influenced in the matter by De Morgan, with whom he maintained an extensive correspondence [331, 317, 319].[6]

Laplace and his nineteenth-century authors introduced mathematical expectation by considering first the case where there is only one possible value for the gain: the mathematical expectation is this possible gain multiplied by the probability of obtaining it. In cases where different gains are possible with different probability, we add the individual mathematical expectations to obtain a total mathematical expectation. This way of proceeding, which shows that mathematical expectation is not being conceived of as a property of probabilized quantities, persisted into the early twentieth century. We still see it, for example, in treatises in German by Emanuel Czuber in 1903 ([103], page 168) and in English by Arne Fisher in 1915 [148]. The conceptual distance between mathematical expectation and the notion of a probabilized quantity remained even after mathematicians became accustomed to the general idea of the theoretical mean of a probabilized quantity. As we will see, the two ideas were conflated in the nineteenth century only in Russia.

1.3 Lacroix's *somme éventuelle*

Because we are interested first of all in antecedents for the term *random variable*, we should pause over the paragraph where Sylvestre François Lacroix (1765–1843) introduces the term *espérance mathématique* in his *Traitée* ([226], §60, page 98). he writes as follows:

> La considération des sommes éventuelles, c'est-à-dire dépendantes du hasard, a introduit dans le calcul des probabilités une expression qui mérite un examen particulier, celle de d'*espérance mathématique*, par laquelle on désigne le *produit d'une somme éventuelle par la probabilité de l'obtenir.*[7]

Lacroix's *somme éventuel* was a single amount of money, which a player might or might not receive. The French adjective *éventuel*, which we will encounter again in our story, means *depending on the circumstances—on how events turn out.* Lacroix was not the first to use it in writing about probability. Condorçet, for example, had written about *droits éventuelles* (contingent claims) in 1785 ([81], page 310).

Should we translate *somme éventuel* into English as *random amount*? Surely not, for this suggests not a single amount that one may or may not receive but an amount that could take several different values. If I reach into my wallet and pull out a random bill, it might be $1, $5, or more.

[6] Check Zernov's pamphlet, which appeared in 1843, before Bunyakovsky's book. See Maistrov, [256], page 169.

[7] English translation: The consideration of gains that are uncertain, i.e., depend on chance, has introduced an expression that merits examination into the probability calculus, that of *mathematical expectation*, by which we mean the *product of an uncertain gain by the probability of obtaining it.*

Lacroix's German translator, the Erfurt mathematician Ephraim Salomon Unger (1789–1870), rendered the quoted sentence as follows ([227], §60, pages 126–127):

> Die Untersuchung der ungewissen Summen, d. h. solcher, die vom Zufalle abhängen, hat in der Wahrscheinlichkeitsrechnung einen Ausdruck eingeführt, der einer bei sondern Prüfung verdient, nämlich *mathematische Hoffnung* worunter *das Product einer ungewissen Summe mit der Wahrscheinlichkeit sie zu erhalten multiplicirt*, verstanden wird.

Notice that Unger translated *éventuelles* by *ungewissen* (uncertain), an adjective that is less loaded, at least for us today. The German phrase *die vom Zufalle abhängen*, a straightforward translation of Lacroix's *dépendantes du hasard*, is also notable. Well into the twentieth century, similar German phrases were used where we might now say *random*.

2 Random errors and random values

When should we say that a new concept has emerged? A reasonable answer is that the concept should have been named. Modern mathematical eyes can see a probability distribution and hence a *random variable* with that distribution in any problem where a gambler receives different amounts of money with different probabilities or in any life table. But by 1750, there were still no names for these concepts: no name for a probability distribution, and no name for a quantity that possesses one. The concept of a probability distribution emerged from the theory of errors.

2.1 The concept of a probability distribution

When did the notion of a probability distribution first acquire a name? We can say with some confidence that the concept and names for it emerged in the period from 1755 to 1780, in work on errors of measurement by Thomas Simpson (1710–1761), Johann Heinrich Lambert (1728–1777), Joseph-Louis Lagrange (1736–1813), Daniel Bernoulli, and Laplace [178].

In his insightful book, *The History of Statistics*, Stephen M. Stigler states, with justice, that "a random distribution of errors was first conceived, perhaps independently, by Simpson in 1755, Lambert in 1760, and Lagrange in about 1769" ([352], page 100). This is not to say, however, that these authors all named the concept. Simpson suggested some hypotheses for the chances with which different errors might occur and pointed out that his suggested chances were the same as the chances for obtaining different numbers when throwing a certain number of dice with a certain number of faces [345]. Here he was conceiving of a probability distribution for errors only in the sense in which De Moivre before him had conceived of a probability distribution for results from dice throwing; he was not naming the concept. Even when authors began drawing probability curves, we may question whether they had yet named the concept of a probability distribution.

By the late 1770s, in any case, the concept did have a name: *loi de facilité* (law of facility). We find this term in Lagrange's article on the utility of taking averages of observations, which appeared in 1776 [228]. The following year, Laplace used this same term in a memoir that he read to the French Academy of Sciences (published only two centuries later, by Charles Gillispie [49]). By 1781, Laplace was using both this term and *loi de possibilité* (law of possibility) [231]. He did not explain any difference in meaning between the two terms,[8] but he tended to use *loi de facilité* for possibly unknown probabilities and *loi de possibilité* for probabilities based on judgment. In the table of contents of his monumental *Théorie analytique des probabilités*, published in 1812, he calls a probability distribution a *loi de probabilité* [232].

Laplace's unexplained contrast between *loi de facilité* and *loi de possibilité* was a manifestation of a deeper duality that would re-emerge throughout the nineteenth and twentieth centuries. At the beginning of the 1781 memoir, Laplace lists three ways of finding the probability of an event ([237], pages 384–385):

1. *a priori*, as when we know that a coin or die is balanced,
2. *a posteriori*, as when we observe repeated trials,
3. from consideration of the reasons we might have to say that the event will occur (in his words: *par la considération des motifs qui peuvent nous déterminer à prononcer sur l'existence de cet événement*).

Later authors would see here distinction between subjective and objective probability: probabilities found in the first two ways are objective; those found in the third way are subjective. Discuss Daston [109]

2.2 Laplace's *variables*

Laplace's 1781 memoir may contain the first instance in which a probabilized quantity is called a *variable*. This is in a passage at the beginning of §VII ([237], page 396):[9]

> Soient n quantités variables et positives $t, t_1, t_2, \ldots, t_{n-1}$, dont la somme soit s et dont la loi de possibilité soit connue; on propose de trouver la somme des produits de chaque valeur que peut recevoir une fonction donnée $\psi(t, t_1, t_2, \ldots, t_{n-1})$ de ces variables, multipliée par la probabilité correspondante à cette valeur.[10]

[8] What have other commentators said about this?

[9] This passage was called to my attention by Hans Fischer. It reappears, with some variation in the wording, at the beginning of §15 of Volume II of Laplace's *Théorie analytique* [232].

[10] *English translation:* Let $t, t_1, t_2, \ldots, t_{n-1}$ be n variable and positive quantities, whose sum s and law of possibility are known. We propose to find the sum of the products of each value that can be taken by a given function $\psi(t, t_1, t_2, \ldots, t_{n-1})$ of these variables, multiplied by the probability corresponding to that value.

The problem posed in this passage is one of numerical analysis. Laplace makes weak assumptions about the laws of possibility of the variables, which he allows to vary over continuous ranges and implicitly assumes to be independent, so that it is an integral that he seeks (even though he calls it a *somme*), and he explains how this integral can be approximated with the tools he had at the time. He uses the abstract term *quantité variable*, which he then simplifies to *variable*, rather than a more specific term such as "error" or "advantage for a player" because he is interested in a number of different applications. In the case where the t are individual errors of measurement, he wants to find the probability distribution of their sum s; he accomplishes this by setting the function $\psi(t, t_1, t_2, \ldots, t_{n-1})$ identically equal to one, so that he is merely adding the probabilities for values of $t, t_1, t_2, \ldots, t_{n-1}$ that add to s. In other cases he proposes to carry out integrations required for what we would now call a Bayesian or inverse probability analysis.

Well into the twentieth century, we see echoes of Laplace's way of introducing the word "variable" as a general name for a probabilized quantity. First one uses "variable" as an adjective (as in "variable quantity" or "variable magnitude"); then one simplifies to "variable", which thereby becomes a noun.[11]

2.3 Poisson's explanation of statistical stability

Inasmuch as he used the terms *loi de possibilité* and *loi de facilité* across an array of applications, we can say that Laplace had a general concept of a probabilized quantity. His use of the terms *magnitude variable* and *variable* to name the concept seems to have been limited, however, to the passage just cited. He usually relied on the context to name the particular quantities with which he was working.

In his *Ars conjectandi*, Jacob Bernoulli had proven that the relative frequency of an event in a large number of independent trials can be expected to approximate the event's probability. In Laplace's work, mainly but not wholly in his study of errors, we can discern generalizations of Bernoulli's theorem to probabilized quantities: the average of a large number of probabilized quantities should approximate their mean (or the average of their means).[12] Siméon-Denis Poisson (1781–1840), Laplace's successor as the leader of French mathematics in the councils of French governments [41], made the generalization more explicit, in the course of explaining what he called *la loi des grands nombres* the (law of large numbers).[13]

---

[11] Discuss how *variable* was introduced in mathematics in other contexts. Newton for example, used *variable Quantity*. What was the practice of Leibniz? Johann Bernoulli? Consult Kline. Other sources?

[12] See [147], page 30, for one occasion when Laplace considered the average of a large number of probabilized quantities with different means.

[13] See Sheynin ([329], pages 270–275), Bru ([43], pages 11–15), Stigler ([352], pages 182–186), Fischer ([147], pages 35–36).

2.3.1 The law of large numbers

Poisson announced his law of large numbers with great enthusiasm in 1835, writing as follows ([299], page 478):

> Les choses de toute nature sont soumises à une loi universelle qu'on peut appeler *la loi des grandes nombres*. Elle consiste en ce que, si l'on observe des nombres très considérables d'événements d'un même nature, dépendants de causes qui varient irrégulièrement, tantôt dans un sens, tantôt dans l'autre, sans que leur variation soit progressive dans aucun sens déterminé, on trouvera, entre ces nombres, des rapports à peu près constants.

As this quotation makes clear, Poisson's law of large numbers was a statement about what is observed empirically, not a statement about theoretical means.

In his *Recherches sur la probabilité des jugements*, published in 1837 ([300], §§52–54, pages 138–145), Poisson considered both events (e.g., a coin falls heads) and quantities (e.g., how long a person lives). In both cases, he asserted that average outcomes are stable from one set of observations to another, provided that there are many observations in each set.

- In the case of events, Poisson wrote

  $$\frac{m}{\mu} = \frac{m'}{\mu'}, \tag{1}$$

  where μ is the number of trials (e.g., coin tosses), m is the number of times the event happens (e.g., the coin falls heads) in that set of trials, and μ' and m' are the corresponding numbers for a second set of trials. Today we might state (1) by saying that the relative frequency is the same in the two sets of trials.

- In the case of quantities, Poisson wrote

  $$\frac{s}{\mu} = \frac{s'}{\mu'}, \tag{2}$$

  where μ and μ' are the numbers of observations, and s and s' are their sums (e.g, s is the sum of lifetimes for one group of μ people, and s' is the sum of the lifetimes for a different group of μ' people). We might state (2) by saying that the empirical average is the same in the two sets of observations.

2.3.2 Three general propositions

Poisson provided mathematical proofs of his two laws of large numbers, which he saw as confirmed by experience, using three *propositions générales* (general propositions), which he proved in his nineteenth-century fashion. Continuing to use Poisson's symbols (but not always his words), we can state these propositions as follows:

First general proposition Assuming that the probability of an event E varies over the μ trials, where μ is large, let us write p_i for its probability on the ith trial, and let us write p' for the average probability:

$$p' = \frac{1}{\mu}(p_1 + p_2 + p_e + \ldots + p_\mu).$$

Then if μ is very large, then the relative frequency m/μ will be equal, *à très peu près et avec une très grande probabilité* (very nearly and with a very great probability), to p'. This generalizes Bernoulli's theorem, in which the probability is constant.

Second general proposition Consider successive trials that can be governed by various "causes". Only one of the causes is in effect on each trial, and the several causes have definite probabilities, constant from trial to trial, of being in effect on a given trial. Suppose that the probability p_i for an event E happening on the ith trial is depends on which of the causes is in effect on that trial. Define γ by averaging the probabilities given to E by each cause, weighted by the probability of that cause being in effect. (We may call γ E's theoretical mean probability, though Poisson used no such name.) Then in a long sequence of trials, p' will equal γ, again very nearly and with a very great probability.

Third general proposition In the case of μ observations of a probabilized quantity and the observed sum s of these observations, we will have, again very nearly and with very great probability,

$$\frac{s}{\mu} = a_1\alpha_1 + a_2\alpha_2 + a_3\alpha_3 + \ldots + a_\lambda\alpha_\lambda, \tag{3}$$

where $a_1, a_2, a_3, \ldots, a_\lambda$ are the possible values for the probabilized quantity, and α_i is the theoretical mean probability for the event that the probabilized quantity comes out equal to a_i. (This theoretical mean probability is defined just as γ is defined for the event E in the second general proposition, and it is the same for all μ trials.)

As Poisson noted (pages 139–140), the first two general propositions bring us back to Bernoulli's theorem. By the first proposition, m/μ approximates p', and by the second, p' approximates γ. So m/μ approximates γ. (All these asserted inequalities being, of course, only very close approximations with very high probability, μ being large.) But γ is the probability of E on each trial, and Bernoulli had already told us that m/μ will approximate that probability. On the other hand, the third general proposition is more novel. It generalizes Bernoulli's theorem to probabilized quantities.

2.3.3 *Une chose d'une nature quelconque*

As I have already emphasized, Poisson did not have a name for a probabilized quantity or its theoretical mean. In his 1837 book, he introduced what I have been calling a probabilized quantity as follows ([300], page 141; typographical error corrected):

> ...on considère une chose A d'un nature quelconque, susceptible d'un nombre λ de valeurs, connues ou inconnues, que je représenterai par $a_1, a_2, a_3, \ldots a_\lambda$, et parmi lesquelles une seule devra avoir lieu à chaque épreuve, de sorte que celle qui sera arrivée ou qui arrivera sera, dans cette question, l'évènement futur. Soit aussi $c_{i,i'}$ la chance que la cause C_i, si elle était certaine, donnerait à la valeur $a_{i'}$ de A. Les valeurs de $c_{i,i'}$, relatives aux divers indices i et i', depuis $i = 1$ jusqu'a $i = \nu$ et depuis $i' = 1$ jusqu'a $i' = \lambda$, seront connues ou inconnues; mais pour chaque indice i, on devra avoir
>
> $$c_{i,1} + c_{i,2} + c_{i,3} + \cdots + c_{i,\lambda} = 1;$$
>
> car si la cause C_i était certaine, l'une des valeurs $a_1, a_2, a_3, \ldots a_\lambda$, arriverait certainement en vertu de cette cause.[14]

The context makes it clear that $a_1, \ldots a_\lambda$ are real numbers, so that Poisson's *chose* (thing) A is what we now call a *random variable*,[15] with a probability distribution that depends on the cause (we might now call it the value of a parameter or an independent variable) C_i.

Several historians have emphasized Poisson's originality in this passage. According to Oscar Sheynin, it shows that Poisson "was the first to introduce the concept of a random quantity" (1978 [329], page 250). Bernard Bru, citing Sheynin, agrees that Poisson was the first to give a definition of a random quantity ([43], page 7). More recently, Hans Fischer states that Poisson created an early concept of random variables ([147], page 31), and Prakash Gorroochurn states that Poisson discovered the concept of a random variable ([169], page 75).

Poisson's third general proposition says that the average value of the thing A over a large number of independent trials is very likely to come very close to the mean value calculated from the possible values and their probabilities. Poisson proved this law, in his nineteenth-century fashion, not only for the case where A has a finite number λ of possible values but also for the case where it has indefinitely many possible values, so that its probability distribution can be approximated by a density over a continuous range. But his statement of the

[14] *English translation:* ...we consider a thing A, of any nature, that can take a number λ of values, known or unknown, which I will designate by $a_1, a_2, a_3, \ldots a_\lambda$, and among which one must happen on each trial, so that the one that happens or will happen will be, with respect to this question, the future event. Also, let $c_{i,i'}$ be the chance that the cause C_i, if it were certain, would give rise to the value $a_{i'}$ for A. The values of $c_{i,i'}$, for the different indices i et i', from $i = 1$ to $i = \nu$ and from $i' = 1$ to $i' = \lambda$, will be known or unknown; but for every index i, we should have

$$c_{i,1} + c_{i,2} + c_{i,3} + \cdots + c_{i,\lambda} = 1;$$

because if the cause C_i were certain, one of the values $a_1, a_2, a_3, \ldots a_\lambda$ will surely happen because of that cause.

[15] Some authors have used the term *random variable* in cases where the values are vectors or other mathematical objects rather than real numbers. See, for example, [89]. But this has not been common.

law falls short of defining the general concept of a theoretical mean or expected value. He writes as follows (page 155):

> ...si A est une chose quelconque, qui soit susceptible de différents valeurs à chaque épreuve, la somme de ses valeurs que l'on observera dans une longue série d'épreuves, sera à très peu près et très probablement, proportionnelle à leur nombre. Le rapport de cette somme à ce nombre, pour une chose déterminée A, convergera indéfiniment vers une valeur spéciale qu'il atteindrait si ce nombre pouvait devenir infini, et qui dépend de la loi de probabilité des diverse valeurs possibles de A.[16]

We might call the ratio of which he writes the mean of the random variable A. Poisson writes it as an integral:

$$\frac{s}{\mu} = \int_{l}^{l'} Zzdz, \tag{4}$$

where s is the sum of the observed values, μ is their number (so that s/μ is the average observed value), the equals sign indicates the approximation when μ is large, $[l, l']$ is an interval within which A falls except with neglibly small probability, and Z is what we would now call the probability density, so that the integral is the theoretical mean. Poisson asserts that the formula (4) has numerous useful applications. But he does not give the quantity a general name. It is simply a *valeur spéciale* (special value), its meaning and denomination depending on the application. When the A are measurements of a quantity α, and the measuring instrument is well designed, the special value will be close to α (pages 155–157). When the A are lifespans of a large number of individuals from a stable population, the special value will be what is called the *vie moyenne* (average lifetime) (pages 158–159).

We now take the notion of a theoretical mean value so for granted that Poisson's failure to forthrightly name it as such seems inexplicably reticent. He fully understood that his *valeur spéciale* was calculated in the same way as a mathematical expectation; there is no room for doubt about this. But the justification of the definition of the value of a expectation (the amount it would cost to construct the expectation's payoff from fair bets) obviously did not extend to justifying a definite value for an arbitrary thing A that takes various values with various probabilities. And unlike Hauber, Poisson evidently did not see his way towards an arbitrary definition with no justification. His third general proposition was the justification.

Poisson's enthusiasm about his law of large numbers was influential, but his attempt to pull apart Laplace's subjective and objective aspects of probability

[16] *English translation:* ...if A is anything that can take different values on different trials, the sum of the values we observe in a long series of trials will very likely, to a very good approximation, be proportional to their number. The ratio of the sum to the number, for a definite thing A, will converge indefinitely to a special value that it would attain if the sum became infinite and that depends on the probability law of A's various possible values.

and to develop Laplace's application of probability to judicial judgements contributed to what Herbert Weisberg has called the great unraveling of classical probability ([377], page 216). Nineteenth-century mathematicians and statisticians could not agree on how his law of large numbers should be interpreted and what, exactly, he had proven ([352], pages 182–186). Irenée Jules Bienaymé (1796–1878) did not think he had added much to Bernoulli's theorem (see [185]). Bienaymé and the German statistician Wilhelm Lexis tried to use Poisson's concept of variable chances to understand statistical series, but without substantial success. Many mathematicians adopted Poisson's term, *law of large numbers*, as a new name for Bernoulli's theorem rather than as a name for an empirically observed regularity.[17]

2.4 Hauber's *zufälligen Werthe*

The difference between our perspective and the perspective of Poisson's time is further underlined by the reception, or lack of reception, given at the time to what now looks like a more modern treatment of mean values of probabilized quantities by a young German scholar, Carl Friedrich Hauber (1804–1831).[18] Inspired by articles Poisson had published in 1824 and 1829 [297, 298], Hauber made his own effort to streamline the theory of observations in a lengthy article published in installments in *Zeitschrift für Physik und Mathematik* in Vienna from 1830 to 1832 [182]. In this article, he introduced an explicit concept of *mittlere Werth* (mean value), while arguing that the concept had already been used to assess profits and losses, in actuarial work, and by Laplace, Gauss, and Poisson in their theory of observations. As he put it (page 25):

> Da aber in diesen Schriften die Sätze nur speciell für die jedesmaligen Gegenstände der Andwendung gegeben sind, so will ich allgemeine, auf alle jene Gegenstände anwendbare, Sätze über die mittleren Werthe aufstellen und beweisen, und Bermerkungen über die Anwendungen dieser Sätze hinzufügen.[19]

Perhaps following Poisson, Hauber begins by calling a probabilized quantity *eine Sache* (a thing). But then he calls it *eine unbestimmter Größe* (indeterminate quantity), which has *möglichen Werthe* (possible values) with various probabilities. He then considers several such indeterminate quantities and their

[17] Bicentenary in St. Petersburg: [322]. In *Die Iterationen*, Bortkiewicz complains that Markov has reduced the law of large numbers to Bernoulli.

[18] A native of Würtemberg, where his father was a well known theologian and mathematician, Hauber did his work on the theory of errors while a student in Vienna and began publishing on the topic in 1829 [180, 181]. He fell ill while still in Vienna and died a few days after returning home, on 12 April 1831. See the note at the end of the last installment of [182] and the obituary in [30], pages 318–322. Hauber had been slated to teach at the University of Tübingen; according to Scharlau ([312], page 256), his *Habilitation* in mathematics was recorded there in 1831.

[19] *English translation:* But since the results given in these writings are always specific to the subject in the particular application, I will state and prove general results about mean values, applicable to all these subjects, and add remarks about the application of these results.

wirklichen zufälligen Werthe (actual random values), which he explicitly assumes to be independent. He defines the *mittlere Werth* (mean value) of each indeterminate quantity in the usual way and then shows that the mean value of a sum of indeterminate quantities is the sum of their mean values. Similarly, he shows that the mean value of the product of indeterminate quantities is the product of their mean values. These results[20] now appear in every probability textbook, but Poisson did not take them up in his 1837 book, and to the best of my knowledge, Hauber's work was never cited by his contemporaries.[21]

2.5 Cournot's *variables* and their *valeurs fortuites*

Exposition de la thórie des chances et des probabilités [78], published by Antoine-Augustin Cournot (1801–1877) in 1843, was probably the most influential probability textbook of the nineteenth century.[22] Cournot combined an innovative introduction to the basics with a readable philosophical treatment of probability, and he painted a broad canvas of applications, giving as much attention to the theory of errors, demography, and insurance as to the treacherous project of applying probability to judicial judgements, which had figured so prominently in Poisson's book. He avoided Poisson's phrase "law of large numbers", but he took advantage of work by his friend Irenée Jules Bienaymé (1796–1878) to discuss how different hypotheses about the variability of chances imply different rates at which observed proportions might approximate an average chance. Coming at a time when European universities were beginning to teach probability, Cournot's book had a lasting influence on the subject's terminology, not only in France but in other European countries as well. Some of this influence came through the translation into German by Christian Schnuse [79], which omitted only the chapters on applications.

2.5.1 Variables

In the course of the book, Cournot uses several terms to refer to a probabilized quantity. He begins by calling it a *quantité variable* or a *grandeur variable*, both of which we would translate into English as "variable quantity". But then, in Chapter VI, where he explains how a continuous probability distribution for a *grandeur* x can be represented geometrically (by what probability teachers of probability now call a probability density), he slips into calling x simply a *variable*.[23] This is exceedingly natural, because the geometric picture comes

[20]Or at least the result about sums, which does not require the assumption of independence.

[21]Hauber's 1830 article was listed in Laurent's 1873 bibliography [238] and Merriman's 1877 bibliography [262]. In his 1899 history of probability [101], Czuber twice remarks that Hauber's work merits attention, a comment that seems to confirm that it had not received any. The only other citations of which I am aware are recent citations by historians: [214, 147].

[22]Bernard Bru discusses the origins, accomplishments, and influence of the book in the introduction to its republication as the first volume of Cournot's complete works [81]. This introduction, along with the notes and index provided by Bru and his collaborators, are very informative. Oscar Sheynin's translation of the book into English [82] is also useful.

[23]In §65, he writes about "la loi de probabilité des diverse valeurs de la variable". He quickly reverts to using *variable* as an adjective, but he again uses *variable* as a noun meaning a

right out of the differential and integral calculus, where the name for x had already long been simplified from "variable quantity" to "variable".

Although Cournot's use of *variable* was doubtlessly influential, it was not adopted by other leading French mathematicians in the nineteenth century. We do not find *variable* for a probabilized quantity in the treatises by Joseph Bertrand (1889 [26]) or Henri Poincaré (1896 [295, 296]). Poincaré vacillated between *grandeur* and *quantité*.[24]

Cournot gave a general definition of the mean of variable x, which he called its *moyenne* (§67):

> ... la *moyenne* de toute les valeurs que peut prendre fortuitement la grandeur x.[25]

He also used *moyenne* or *valeur moyenne* as the name of the average of a list of numbers. When confusion threatened, and he needed to distinguish between the *moyenne* of x and the *moyenne* of a set of observed values of x, he called the former the *moyenne absolue* (§69). He also defined the *médiane* (median) of a variable and criticized previous authors who had called this quantity the *valeur probable* (§§34, 68).

Cournot noted that the mean of a linear combination of variables is the same linear combination of their means (§74). As we have seen, Hauber had belabored this point, and mathematicians had been using it, starting at least with Laplace, for fifty years. But Cournot can probably be credited with first making it explicit in a book.

Cournot used the concept of an *espérance mathématique* (mathematical expectation) in the same way as Laplace, Lacroix, and Poisson had done. He introduced the concept with these words (§50):

> Par une association de mots assez bizarre, on appelle *espérance mathematique* le produit qu'on obtient en multipliant la valeur d'une chose en unités monétaires, par la fraction qui exprime that probabilité mathématique du gain de cette chose.[26]

Mathematical expectations can be added to obtain a total mathematical expectation (§61). It is important to recognize that this is not the same as calculating the mean of a probabilized quantity. When mathematical expectations are added, no probabilized quantity is being identified, and it may be problematic to identify one. If I have a 50-50 chance of being invited by a friend to a meal worth \$20 and a 50-50 chance of inheriting a painting worth \$200 from my aunt, I can add my expectations ($0.5 \times \$20 + 0.5 \times \$200 = \$110$) without making any

probabilized quantity in §§84,85,91,125.

[24] On page 169 of his second edition (1912), Poincaré writes, "Je suppose qu'on a effectué différentes mesures $x_1, x_2, \ldots, x_n$ d'une même grandeur...". On page 234, he writes about "les quantités observées...".

[25] *English translation:* ...the mean of all the values that the quantity x can take randomly.

[26] English translation: By a rather bizarre association of words, we call the product of the value of a thing, in monetary units, times the fraction that expresses the mathematical probability of gaining the thing, *mathematical expectation*.

Table 1: Nouns that Cournot used most often with *fortuit* in his 1843 *Exposition* [78], with approximate counts of the number of sections where the terms appear, translations used by Schnuse in his 1849 German translation [79], and translations that might be used in teaching probability in English today. (The book had 240 sections altogether.) Schnuse did not translate the entire book, and the omitted passages included those where Cournot used *manière fortuite*.

	Cournot 1843	*Schnuse 1849*	*Possible English*
18	cause fortuite	zufällige Ursache	random cause
10	écart fortuit	zufällige Abweichung	random deviation
8	anomalie fortuite	zufällige Anomalie	random anomaly
8	combinaison fortuite	zufällige Verbindung	random combination
4	rencontre fortuite	zufällig Zusammentreffen	random encounter
4	tirage fortuit	zufällig Zug	random draw
3	épreuve fortuite	zufällig Versuch	trial
3	erreur fortuite	zufällig Fehler	random error
3	événement fortuit	zufällig Ereignis	random event
3	manière fortuite	—	random manner
3	valeur fortuite	zufällig Wert	random value

hypothesis about my chance of getting both. But I need to know this chance in order to define a probabilized quantity equal to the total value I will receive. If there is no chance that I will get both, the probabilized quantity would take the value \$20 with probability 1/2 and \$200 with probability 1/2; if I am sure to get both or neither, the probabilized quantity would take the value \$220 with probability 1/2 and \$0 with probability 1/2; etc.

2.5.2 Randomness

Cournot used several adjectives that we might translate into English as *random*. The one he used the most was *fortuit*, which had long been used by French authors on games of chance. Table 1 lists some of the nouns with which Cournot used this adjective. He also used *fortuit*, though less often, with *accumulation*, *cas*, *choix*, *concours*, *détermintion*, *estime*, *extraction*, *gain influence*, *oscillation*, *perte*, *phénomène*, *resultat*, and *variation*. In his translation into German, Schnuse almost always translated *fortuit* as *zufällig*.

For our theme, the most notable entry in Table 1 is *valeur fortuite*. Today we call variables random. Cournot did not call his *variables* random, but he called their values random. They are *valeurs fortuites*. This makes some sense. The *variable*, which has a *loi de probabilité* (probability law, or distribution) and a *moyenne* (mean) is not being chosen at random. Its *valeurs* (values) are determined at random. But we can also see a certain instability in the distinction, which will emerge with writers and teachers less verbally agile than Cournot. Cournot frequently refers to the probability that an *écart fortuit* or an *erreur fortuite* will fall between certain limits. So what has the probability law, the variable or the random values?

Another adjective that Cournot used fairly often was *aléatoire*. The words *aléatoire* in French and *aleatory* in English derive from the Latin *alea*, which means a die, a game of chance, or a chance taken, as in Cesar's *iacta alea est* on crossing the Rubicon. The law (Roman and now French and English) calls an insurance policy or an annuity an aleatory contract (*contrat aléatoire* in French). Today the French equivalent of *random variable* is *variable aléatoire*, but as we are learning, both are twentieth-century terms. A close examination of Cournot's use of *aléatoire* makes it clear that for him it was confined to the same circle of ideas as *espérance mathématique*: it meant having something to do with an aleatory contract—with an gamble or a gambling device or some business agreement that provides a contingent payoff.[27]

Table 2 lists the nouns with which Cournot used the adjective *aléatoire*. As the list shows, Schnuse did not use *zufällig* when translating these uses into German, and we would not use *random* in translating them into English. Sometimes Schnuse translated *aléatoire* as *aleatorische*, but he does not appear to have been fond of this neologism, for he often avoided it, sometimes by using no adjective with the noun and sometime by using a circumlocution that avoided the noun as well.

In using the term *instrument aléatoire*, Cournot was following Lacroix ([226], pages 8, 62, and 262). I have not seen other earlier uses of *aléatoire* by authors on mathematical probability.

Three nouns in Table 2 merit comment because Cournot used them with both *fortuit* and *zufällig*: *événement*, *gain*, *épreuve*.

- Cournot uses *événement aléatoire* for an event that we focus on because it defines a gamble; it is the event that a certain player wins (§§12,20,55,87). For example:

  > In §12: ...lorsqu'il s'agit d'un événement aléatoire, ce qu'on a intérêt à connaître...
  >
  > In §20: ...l'événement aléatoire, le gain d'un joueur...[28]

  He uses *événement fortuit* for an unpredictable event that has no part in a contract (§§40, 41, 86). In §40, for example, he is explaining what he means by chance:

[27] As Bernard Bru points out ([81], page 301), French authors were beginning to use *aléatoire* to mean *uncertain* around the time Cournot was writing. French dictionaries provide these two literary examples:

- Honoré de Balzac in *Les Employés* in 1837: certaines femmes dont la fortune est aléatoire.
- Jules Michelet in *Journal intime*, March 1839, page 293: Tout ce que vous voyez ici, c'est la terre froide et triste, c'est le sillon entre les vignes, c'est la vigne et son vigneron infatigable, ôtant l'échalas à l'automne pour le remettre au printemps. Grand travail, profit incertain. Rien n'est plus aléatoire que cette culture.

Even in these examples, however, the uncertainty concerns material or monetary gains.

[28] English translations: ...when it is a matter of an aleatory event, one which we have an interest in knowing...the aleatory event, the win by a player....

Table 2: Nouns that Cournot used with *aléatoire* in his 1843 *Exposition* [78], with approximate counts of the number of sections where the terms appear, translations used by Schnuse in his 1849 German translation [79], and translations that might be used in teaching probability in English today.

	Cournot 1843	*Schnuse 1849*	*Possible English*
7	épreuve aléatoire	Versuch	trial
4	événement aléatoire	aleatorisch Ereignis	event
4	spéculation aléatoire	aleatorische Spekulation	financial speculation
3	prime aléatoire	Prämie	risk premium
3	marché aléatoire	aleatorische Handel	financial market
2	condition aléatoire	Umstand	condition
2	droit aléatoire	eventuell Recht	contingent claim
2	instrument aléatoire	aleatorisch Instrument	chance device
1	convention aléatoire	aleatorische Uebereinkunft	bet
1	gain aléatoire	eventuell Gewinn	possible gain

> Les événements amenés par la combinaison ou la rencontre de phénomènes qui appartiennent à des séries indépendantes, dans l'ordre la causalité, sont ce qu'on nomme des événements *fortuits* ou des résultats du *hasard*.

In §41, he discusses a specific *événement fortuit*; a man is struck by lightening.

- In §61, Cournot lists the possible gains the player in the Saint Petersburg problem: 1,2,4,8, etc. These are the *gains aléatoires*. In §183, in contrast, when he is discussing the variation in profits and losses a company might experience over a span of years, he uses *fortuit*: *les limites de la perte et du gain fortuits moyens* (the limits on the average profit or loss). Here money is involved, but there is no contract that specifies events on which specific gains or losses are contingent.

- Cournot uses *épreuve aléatoire* when he is talking about games of chance (§§23,30,33,36,58,63,89) and *épreuve fortuite* in more abstract discussions or applications that do not involve money (§§67,73,79). Poisson had introduce the term *épreuve*, but he did not attach an adjective to it. In most cases, Cournot does not either.

Table 3 shows two other adjectives that Cournot also used: *éventuel* and *accidentel*. We have already encountered *éventuel* in Lacroix's *somme éventuelle* and Condorçet's *droit éventuel*. In Cournot's use, as well as these previous uses, *éventuel* is more or less a synonym for *aléatoire*. Similarly, we can consider *accidentelle* a synonym for *fortuit*.

Table 3: Nouns that Cournot used with *éventuel* and *accidentel* in his 1843 *Exposition* [78], with approximate counts of the number of sections where the terms appear, translations used by Schnuse in his 1849 German translation [79], and translations that might be used in teaching probability in English today.

	Cournot 1843	*Schnuse 1849*	*Possible English*
1	droit éventuel	eventuell Recht	contingent claim
1	gain éventuel	eventuell Gewinn	possible gain
1	perte éventuelle	eventuell Verlust	possible loss
3	cause accidentellle	unregelmässige Ursache	random cause
1	trouble accidentelle	zufällige Störung	random disturbance
1	circonstance accidentelle	—	random circumstance

2.5.3 Influence

We can trace Cournot's vocabulary in other 19th century textbooks in French, particularly those of Jean-Baptiste-Joseph Liagre (1815–1891) and Hermann Laurent ((1841–1908).

Liagre, an officer in the Belgian military, published his *Calcul des probabilités* [248] in 1852, with a second edition in 1879. He states in his preface that he borrowed a great deal from Lacroix, Cournot, and Quetelet, and Cournot's influence on his vocabulary is striking. Like Cournot, he uses *fortuit* and *accidentel* to mean *random*; *erreur*, *événement*, *cas*, *phénomène*, *cause*, *reproduction*, *triage* (probably a typographical error; *tirage* being meant), and *coïncidence* can be *aléatoire*; *erreur*, *cause*, emphvariation, *circonstance*, *négligence*, *déviation* can be *accidentel*. And like Cournot, he uses *aléatoire* and *éventuel* in connection with betting: *spéculation*, *épreuve*, *événement*, and *marché* can be *aléatoire*, while *gain*, *droit*, *somme*, *fortune* can be *éventuel*.

We also see some of Cournot's vocabulary in Laurent's *Traité du calcul des probabilités*, published in 1873 [238]. For Laurent, the law of large numbers justified a frequentist interpretation of probabilities that replaced Huygens's constructive argument for mathematical expectation in games of chance. But since Laurent understood the law only for events, not for probabilized quantities in general, he still made the argument in two steps. Like all his predecessors, he first considered a single amount of money a person might receive, which he called a *somme éventuelle* (page 124):

> Lorsqu'une personne attend une somme d'argent qui dépend de l'arrivée d'un événemen t incertain, on dit que cette somme est *éventuelle*.[29]

Suppose the *somme éventuelle* is a, the event is E, and E's probability is p. Then the *espérance mathématique*, as we know, is ap. Laurent justified this on

[29] *English translation:* When a person expects a sum of money that depends on the happening of an uncertain event, we say that this amount is *chancy*.

the grounds that E will happen approximately p of the time in a long sequence of trials, and hence you will obtain ap on average. The *espérance mathématique* is the appropriate price for the *somme éventuelle* because it is equal to this average payoff. He then took the second step just as Laplace and his other predecessors had: when a person can get any of several different *sommes éventuelles* with different probabilities, we add the different *espérances mathématiques* to get a total *espérance mathématique*.

3 The expansion of expectation

During the second half of the nineteenth century, probability theory's notion of a theoretical mean came into contact with two other branches of mathematics: mathematical physics and geometry. On the one hand, the Russian mathematician Pafnutii Chebyshev brought his work in the theory of moments, originally motivated by physics, to bear on the theory of errors. On the other hand, French and British mathematicians, especially the Irish mathematician Morgan Crofton, developed geometric probability, in which points, lines, and other geometric objects are chosen at random. These two lines of work remained largely separate, but both brought increased attention to the notion of the mean of a probabilized quantity, and by the end of the century this new attention had produced two innovations:

1. The extension of the term *expectation* (or *Erwartung* in German, *espérance mathématique* in French, or математическое ожидание in Russian), previously used only to name of a player's payoff in a game of chance, to designate the mean of any probabilized quantity.

2. The symbolic representation of this mean or expectation as an operator—i.e., the use of symbols such as $\mathrm{M}(x)$ or $E(x)$ to designate the expectation of a probabilized quantity x.

Chebyshev was using математическое ожидание and *espérance mathématique* in its broad sense in 1867. Crofton was using M() in 1869, and by the 1890s William Whitworth and Ladislaus von Bortkiewicz had replaced the M with an E. Eventually, well into the twentieth century, these innovations were adopted widely.

By 1900, an important role is this story was played by the Saint Petersburg mathematician Andrei Markov, who elaborated Chebyshev's mathematical methodology, calling it the метод математических ожиданий (method of mathematical expectation) and applying it to strengthen Chebyshev's work on the theory of errors and to create what we now call the theory of Markov chains. We will be particularly interested in the impact of the probability textbook that Markov published in Russian in 1900 and in German in 1912. This textbook was celebrated for its rigor and clarity, and the German edition was influential in making the the concept of expectation in the broad sense, and hence the concept of a probabilized quantity, basic in probability theory. We will trace this influence by looking at the spread of one of the book's notational innovations: the use of upper case letters $X, Y, \ldots$ to designate probabilized quantities and the corresponding lower case letters $x, y, \ldots$ to designate their possible values.

We will see some evidence during this period, both in the theory of errors and in geometric probability, of greater use of the word *variable* to designate probabilized quantities, and we will also see wider use of *random* as an adjective, especially in geometric probability in English. This was hardly, however, a step towards the modern use of *random variable* or *random quantity* as a general name for probabilized quantities, because *random* carried (and still carries) a restrictive connotation in geometric probability; when a point is chosen from a geometric set, all the points in that set are supposed to have an equal chance of being chosen.

In the course of discussing these developments, we will encounter several other prominent mathematicians, including Felix Hausdorff in Germany and Aleksandr Liapunov in Russia. We will also encounter a prominent Austrian mathematician, Emanuel Czuber, whose many books on probability tracked the evolution of its vocabulary in Germany in the late nineteenth and early twentieth centuries.

3.1 Chebyshev's математическое ожидание

Chebyshev's first contribution to probability theory was an 1846 article in French [64], drawn from his master's work in Moscow, in which he gave an elementary but rigorous proof of the first of the three statements that Poisson had called the law of large numbers.[30] Chebyshev formulated the statement in French as follows:

> On peut toujours assigner un nombre d'épreuves tel, qu'il donne un probabilité, aussi approchante de la certitude qu'on voudra, et que le rapport du nombre de répétitions de l'événement E à celui des épreuves ne s'écartera pas de la moyenne des chances de E au delà des limites données, quelques resserées que soient ces limites.

Here is how Chebyshev put this into symbols: Suppose that in μ consecutive trials, the event E has probabilities $p_1, p_2, p_3, \ldots, p_\mu$, respectively, and let S be the sum $p_1 + p_2 + p_3 + \cdots + p_\mu$. For arbitrarily small positive numbers z and Q, Chebyshev found a value of μ sufficiently large that the number of times E happens divided by μ will be between $S/\mu - z$ and $S/\mu + z$ with probability at least $1 - Q$. This was an advance on Poisson, because Poisson had not specified how large μ would need to be for given z and Q.

Twenty years later, after working extensively in various areas of mathematics, including number theory and mechanics, Chebyshev returned to probability with a much more important and celebrated contribution. This article, published in 1867 in both Russian and French [65], used what he had learned about

[30] The master's thesis, in Russian, was published in Chebyshev's collected works in 1951. Take a look at that text. Add English translation to bibliography.

the theory of moments to prove a version of the third statement in Poisson's law of large numbers.[31]

Chebyshev seems never to have used *variable* for a probabilized quantity as Cournot had. Instead, he always used words that we would translate into English as *quantity*: величина in Russian and *grandeur* and *quantité* in French. But his terminology was innovative in a different way. In spite of its title (О среднихъ величинахъ in Russian and *Sur les valeurs moyennes* in French), Chebyshev did not call call the mean of a probabilized quantity a *valeur moyenne* or a средниа величина (mean value). Instead he called it an *espérance mathématique* or математическое ожидание (mathematical expectation). He did not give any argument for this innovation.

As we have seen, математическое ожидание was introduced into Russian mathematics by Bunyakovskii, who had been teaching in Saint Petersburg for many years when he published his textbook on probability in 1846. Chebyshev had come to Saint Petersburg in 1847, after completing his studies in Moscow. In 1849, he collaborated with Bunyakovskii in publishing Euler's papers on number theory, and in 1860 he took over teaching probability at the university from Bunyakovskii ([389], page 223). He clearly borrowed the Russian term математическое ожидание from Bunyakovskii, but he seems to have been the first to use it in print to designate the mean of any probabilized quantity. His example firmly established this practice in Russia. Eventually, in the twentieth century, the Russian influence helped establish the use of cognate terms in other languages as well.

While Bunyakovskii had not used математическое ожидание for all probabilized quantities in his book, there is one passage, on pages 278–279, that may have nudged Chebyshev in that direction. In this passage, Bunyakovskii uses the concept of математическое ожидание to argue for the method of estimation that minimizes the sum of absolute errors. The errors can be thought of, he suggests, as losses in a game, and positive and negative errors are equally bad. Since there is no real monetary loss, one should minimize the математическое ожидание rather than the нравственное ожидание. It is also possible, of course, that Bunyakovskii himself drifted into using математическое ожидание for probabilized quantities in his teaching during the period from 1846 to 1860, and that Chebyshev picked the usage up from him. Or he might even have picked it up from his Moscow probability teacher, Nikolai Efimovich Zernov (1804–1862).

The French version of Chebyshev's 1867 article begins as follows:

> Si nous convenons d'appeler *espérance mathématique* d'une grandeur quelconque, la somme de toutes les valeurs qu'elle est susceptible de

[31] Additional points to make: (1) The concept of a moment in physics goes back to Archimedes's study of levers, and what we now call the center of gravity of a body was discussed in the middle ages. The study of moments was thus part of the integral calculus from its beginnings in the seventeenth century. (2) Among students, Chebyshev is best remembered for his inequality, sometimes called the *Bienaymé-Chebyshev inequality* (see [185]) used in the paper. (3) Comment on Chebyshev's visit to Germany, France, and London in 1852

> prendre, multipliées par leurs probabilités respectives, il nous sera aisé d'établir un théorème très-simple sur les limites entres lesquelles restera renfermée une somme de grandeurs quelconques.[32]

Because he put this definition at the beginning of his exposition, Chebyshev was able to state the third statement of Poisson's law of large numbers with a clarity that had eluded Poisson himself.

Aside from its proof, which used Chebyshev's earlier work on moments and continued fractions, the great accomplishment of Chebyshev's result was that it provided precise bounds on the difference between the observed average of a sequence of variables and the average of their means, with a bound also on the probability that the difference will be within these bounds. But rather than quote these bounds, I will quote his less detailed statement of the law, which he gave after stating and deriving his precise bounds:

> Si les espérances mathématiques des quantités $U_1, U_2, U_3, \ldots$ et leurs carrés $U_1^2, U_2^2, U_3^2, \ldots$ ne dépassent pas une limite finie quelconque, la probabilité que la différence entre la moyenne arithmétique d'un nombre N de ces quantités, et la moyenne artihmétique de leurs espérances mathématiques, sera moindre qu'une quantité donnée, se réduit à l'unité, quand N devient infini.

Notice that the quantities can have different probability distributions and different means. One might argue that this statement was already in Poisson, but Chebyshev stated it much more clearly than Poisson had. Because of Chebyshev's clarity, and because Chebyshev did not hesitate to treat the squares of probabilized quantities as probabilized quantities themselves, with their own means, Fischer concludes ([147], page 185) that Chebyshev's "conception of a random variable was much more general than Poisson's."

Nekrasov, Liapunov, Markov, and other Russian mathematicians followed Chebyshev in calling the mean of any probabilized quantity its математическое ожидание,[33] but it was some time before ожидание's English, French, and German equivalents (*expectation, espérance*, and *Erwartung*) acquired similar breadth.

3.2 Geometric expectations

Historians agree that geometric probability began with Buffon's needle problem. In 1777, the French naturalist Georges Buffon (1707–1788) investigated the probability that a thin rod, thrown randomly onto a surface ruled with evenly spaced parallel lines, will hit one of the lines. Buffon showed that if the length of the rod, say d, is less than the distance between the parallel lines, say l, then

[32] *English translation:* If we agree to give the name *mathematical expectation* to the sum of all the values an arbitrary quantity can take, multiplied their respective probabilities, it will be easy for us to establish a very simple theorem concerning the limits within which any sum of quantities will be found.

[33] Check Zernov, Maksimovich, Vasilev.

this probability is $2l/\pi d$. Laplace took up the problem in the first edition of his *Théorie analytique* in 1812 ([232], pages 359–362), considering also the case where there is a second ruling by parallel lines perpendicular to the first. The subsequent history of geometric probability in the nineteenth century has been informatively reviewed by Seneta, Parshall, and Jongmans [323]. Starting in the 1860s, as they note, interest in geometric probability took off in both England and France, involving many prominent mathematicians. All sorts of problems were considered and solved, and means quickly entered the picture.

3.2.1 Barbier's convex disk

Particularly striking was the use of expectation by the French mathematician Joseph-Emile Barbier (1839–1889). In 1860 [14], Barbier considered a generalization of Buffon's problem in which a regular polygon or some other convex disk is tossed randomly onto the ruled surface. What is the probability it will hit one of the lines? At the suggestion of his teacher Joseph Bertrand, Barbier used the additivity of expectation to solve the problem. Writing a for the distance between the parallel lines, Barbier first considered a convex polygon[34] with m sides, each of length c, and a diameter less than a. He brought *mathematical expectation* into the picture this way:

> It est évident que tous les côtés du disque ont la même chance de rencontre; par conséquent, si le premier côté appartient à un premier joueur, le second côté à un second joueur, etc., et si un côté, coupé par une ligne du plan, amène un gain fixe au joueur qui le possède, tous les joueurs ont la même espérance mathématique E avant chaque coup.
>
> L'espérance mathématique d'une personne A qui aurait acheté toutes ces espérances en nombre m, serait égale á mE.

If the fixed gain is one monetary unit, then each player's mathematical expectation is the same as the probability of his side hitting one of the lines, and this is equal, as Barbier knew from Laplace's book, to $2c/\pi a$. So the mathematical expectation of the player who buys all m sides is $2cm/\pi a$. Neglecting the theoretically possible but infinitely unlikely event that an apex of the polygon touches a line, the polygon will fall on a line only if the line crosses two of its sides. So $2cm/\pi a$ is twice the mathematical expectation of a player who gets one monetary unit when the polygon hits a line. In other words, the probability of the polygon hitting the line is $cm/\pi a$. Since cm is the perimeter of the polygon, this shows that the probability is equal to the perimeter divided by πa, and the result generalizes to any convex shape, since it can always be approximated by a convex polygon with sides of equal length. Barbier's argument became classic, finding its way into Bertrand's textbook (with acknowledgement to Barbier) and Poincaré's textbook (with no acknowledgement to Barbier).

[34] A polygon is convex if no line intersects it more than twice.

3.2.2 Crofton's notion of randomness

In Britain, geometric probability became a popular topic in the *Mathematical Questions* column of the *Educational Times*, a British forum where mathematicians high and low could publish mathematical problems and their solutions. Starting in 1864, these problems were collected in a separate annual publication entitled *Mathematical Questions, with their Solutions*, where they were numbered.[35] The contributors to *Mathematical Questions* sometimes debated their solutions, and the volume for the first half of 1867 [266] is notable for its debate about the word *random*. Hugh Godfray (1821–1877), a Cambridge graduate from the Isle of Jersey now remembered as a prominent chess player, was one of those who found the word ambiguous. He offered this example of how *random line* might be interpreted in two different ways ([167], page 65):

> Let us consider a limited area—say a circle— across which a random line has to be drawn.
>
> *Firstly:* We may define a random line to be a line joining two random points of the area,—a random point being, as above, one whose chance of falling on any area is proportional to that area.
>
> *Secondly:* We may define a random line to be one that crosses the circumference in two random points,—a random point on the circumference being one whose chance of falling on any arc is proportional to the length of that arc.

The Irish mathematician Morgan William Crofton (1826–1915), along with others, disagreed, contending that the notion of drawing a geometric object at random from an assemblage is clear once the assemblage is clearly specified. Crofton wrote ([93], page 85),

> . . . the want of definiteness, in my opinion, lies in the different senses which may be given to the expression "drawing a chord." If it means "a line is drawn (at random) cutting the circle," it will give . . . the second case Mr. Godfray mentions, all distances from the centre being equally probable. If it menas "a (random) line drawn from a (random) point on the circle," or "a line joining two (random) points on the circle," it will give the first case. . . .

In a footnote after "a line is drawn (at random) cutting the circle," Crofton added

> There is no contradiction in subjecting a random entity to law, provided in some respect it be left free. Thus a random line may be subjected to a law, viz., it must meet the circle, but otherwise it follows no law—so that the above expression will mean, "A line taken at random from the random lines which meet the circle."

[35] By the time publication ceased in 1918, a total of 18,769 problems had appeared [171]. Contributors included J. L. Sylvester, G. H. Hardy, and many other prominent English, French, German, and American mathematicians.

The following year, Crofton expanded his note in *Mathematical Questions* into an article for the Royal Society's *Philosophical Transactions* [94] that earned him a reputation as the leading scholar in geometric probability (or *local probability*, as he preferred to call it).[36] On pages 183–184 of that article, we find a passage that gives more insight into how Crofton thought about *randomness*:

> ...if a point be taken at random in a plane, the total number of cases is of an inconceivable nature, inasmuch as a plane cannot be filled with mathematical points, any infinitesimal element of the plane containing an unlimited number of points. We see, however, ...that we may consider the assemblage we are dealing with as *an infinity of points all taken at random in the plane*.
>
> Let us examine the nature of this assemblage. As the points continue to be scattered at random over the plane, their density tends to become uniform. It is evident, in fact, that a random point is as likely to be in any element dS of the surface, as in any equal element dS$'$; and therefore by continuing to multiply points, the number in dS will be equal ... to that in dS$'$. Of course, though the density tends to become uniform, the disposition of the points does not tend to become symmetrical; those within any element dS will be dispersed in the most irregular manner over that element.[37] However, it is important to remark that, *for all purposes of calculation*, the ultimate disposition may be supposed symmetrical; for as the position of any point is determined by that of the element dS, within which it falls, it matters not what arbitrary arrangement we assume for the points within that element. Hence we may, if we please, assume that, when a point is taken at random in a plane, those from which it is taken are an infinite number symmetrically disposed over the plane.
>
> Likewise, points taken at random in a line may be supposed equidistant. And if random values be taken for any *quantity*, they may be supposed to form an arithmetical series, with an infinitesimal difference.

The final sentence quoted must catch the eye of anyone searching for antecedents to the modern meaning of *random* in probability theory, but Crofton's insistence on identifying randomness with uniformity was not a step towards *random variable*. Quite to the contrary, it created or at least marked an obstacle that had to be overcome in order for the term to be adopted. It remained impossible to call all probabilized quantities or their values *random* so long as the adjective implied

[36]In the preface to his 1884 treatise on geometric probability ([99], page iv), Czuber listed a number of French and English contributors to the topic and then added: "Allen voran aber muss M. W. Crofton, Professor an der Kriegsakademie in Woolwich, genannt werden, der abgesehen von mehreren kleineren Aufsätzen in einer grundlegenden Abhandlung (in den Philosophical Transactions 1868) diejenigen Probleme der geometrischen Wahrscheinlichkeit behandelte, welche auf willkürlich gezogene Gerade und willkürlich Ebenen sich beziehen."

[37]Here Crofton inserts a footnote that quotes Laplace concerning the emergence of order from disorder and notes that drops of rain on a pavement look more and more uniform as their number multiplies.

a uniform distribution. It is notable, in this connection, that Thornton C. Fry, in his 1928 textbook, *Probability and its Engineering Uses*, called probabilized quantities *variables* but interpreted *random* to mean uniformly distributed. He wrote ([160], page 141):

> Just as "equally likely events," in spite of their theoretical importance in giving us a starting point for the discussion of discrete groups of events, are comparatively rare in practical studies, so "randomness" (or, if we prefer the phrase, "equally likely intervals") is also of greater theoretical than practical importance.

And he proceeded to give two examples of variables that are not distributed at random.

In geometric probability *random* carries to this day the connotation of uniformity that Crofton gave it; see [350]. But the ambiguities that Godfray and others had identified continued to trouble others in the nineteenth century. For Joseph Bertrand, they were among many paradoxes that cast doubt on the coherence of probability theory. (Give citations and perhaps examples.) As Francis Edgeworth wrote in 1908, there is an

> ... indeterminateness which baffles us when we try to define a "random line" on a plane, or a "random chord" of a circle.

So far as I have seen, authors in other languages did not follow Crofton in applying an adjective such as *random* to points, lines, chords, etc. In his 1884 German treatise on geometric probability [99], Czuber singled Crofton out for the praise due him. But instead of writing about that points, lines, and other geometric objects being *zufällig* (random), he wrote about them being chosen *willkürlich* (arbitrarily). In his introduction (pages 1–8), he explained that this meant that their probabilities were determined by the measures of the regions covered by the *unabhängigen Variablen* (independent variables) in an analytic description. The nineteenth French contributions to geometric probability also tended to avoid calling geometric objects random (see, e.g., [229]); the word *fortuit* does not appear, but *variable* is everywhere.

Crofton also made another remarkable contribution to our story: he represented mean value as an operator. We find this innovation in a paper on the theory of errors he read to the Royal Society of London in 1869, where in a footnote ([95], page 185) he adopts

> ... for shortness the symbol M(K) for "the mean value of K," ...

He uses the same notation more extensively a few years later, in a paper he read to the London Mathematical Society in 1877, entitled "Geometrical theorems relating to mean values" [97]. There he writes M(AB) for the mean value of a random arc AB (page 306), M(ρ) for the mean value of the distance ρ between two points (page 308), etc. We also find the symbol in Crofton's article on probability in the 9th edition of the *Encyclopædia Britannica* [98], which first appeared in 1885. There he writes $\mathrm{M}(x)$ for the mean value of an error x and $\mathrm{M}(x^2)$ for the mean value of its square (page 782).[38]

Czuber duly brought Crofton's notational innovation into the German literature, using $M()$ in 1884 in his book on geometric probability ([99], pages 212–244) and his 1891 book on the theory of errors [100], where he writes $M(u)$ for the "Mittelwert einer Grösse u". In the first edition of his general treatise on probability [103], which appeared in 1903, Czuber...

3.2.3 Whitworth's *expectation of parts*

Another very active contributor to *Mathematical Questions* was the British mathematician and clergyman William Allen Whitworth (1840–1905). Although not known for publications in research journals, Whitworth was relatively prominent in British mathematics. As a student at Cambridge in the 1860s, he was involved in launching the *Oxford, Cambridge and Dublin Messenger of Mathematics*, which aimed to give students an opportunity to publish their work, and he was the founding editor-in-chief when it was relaunched, with a more ambitious agenda, as the *Messenger of Mathematics* in 1872. He left this editorship in 1880, but he continued to study and publish mathematical problems even as he became increasingly active as a clergyman. In 1867, Whitman published a short and elementary book, *Choice and Chance*, based on lectures on combinatorics and probability that he had given at Queens College in Liverpool [379]. He expanded the text and added exercises in a second edition in 1870, and both exercises and text grew in subsequent editions, which appeared in 1878, 1886, and 1901.

On pages 199–200 of the fourth edition of *Choice and Chance*, which appeared in 1886, this passage appears:

> The term *Expectation* is usually limited to cases in which a person is to receive a sum of money contingent on the issue of some doubtful event. ...But we may well speak of expectation independently of money, and say that the player has an expectation of seven shots. Similarly, if a man tosses a coin till he gets a sequence of 4 heads, we may say that his expectation is 30 tosses.

This may be the first time that an author writing in English suggested that the use of *expectation* should be broadened in this way. It might even be the first time such a suggestion was made in print in any language. So far as we have seen, Chebyshev did not talk about broadening the use of математическое ожидание and *espérance mathématique*; he just did it.

The fourth edition of *Choice and Chance* also included, for the first time, exercises in geometric probability. Eleven years later, in 1897, Whitworth published a 237-page companion to *Choice and Chance* with the title *DCC Exercises*

[38] In 1853 ([27], pages 314–315), Bienaymé used $S.b\alpha$ for the mean of a probabilized quantity that takes values $\alpha_1, \alpha_2, \ldots$ with probabilities $b_1, b_2, \ldots$; in other words $S.b\alpha = \sum_i b_i\alpha_i$. This can be considered an earlier instance in which mean value is treated as an operator, but Crofton's notation comes closer to modern practice. Crofton was also interested in operators in the differential calculus; see [96].

[380]. It contained, as its title proclaimed, 700 exercises with solutions. Exercise 644, on page 183, reads as follows:

> 644. From a point P within the rectangle $OADB$, perpendiculars OM, ON are let fall on OA, OB. Shew that if P be taken at random within the rectangle, $\mathcal{E}(OM.ON) = \frac{1}{4}OA.OB$. But if P be restricted to lie on the diagonal OD, then $\mathcal{E}(OM.ON) = \frac{1}{3}OA.OB$; and if on the diagonal AB then $\mathcal{E}(OM.ON) = \frac{1}{6}0A.OB$.
>
> NOTE. We use the symbol $\mathcal{E}(X)$ to denote the expectation or average value of a variable quantity X. If X can take n different values all equally likely and $\Sigma(X)$ denote their sum it follows that $\Sigma(X) = n\mathcal{E}(X)$.

Given Crofton's prominence in Whitworth's mathematical world, we may assume that Whitworth's $\mathcal{E}$ was inspired by Crofton's M; Whitworth was making the obvious change from M for mean to E for expectation, and for some reason he fancied the calligraphic font.

The following year, 1898, Whitworth published a 23-page pamphlet with the lengthy title *The expectation of parts into which a magnitude is divided at random investigated mainly by algebraic methods* [381], promising it would become a chapter in *Choice and Chance* if there were another edition, a promise kept when the fifth and final edition appeared in 1901. On page 6, following his preface and preceding the text, he inserts a note that begins as follows:

> We use the symbol $\mathcal{E}(x)$, as explained in DCC Exercises, 644, to denote the expectation or mean value of a variable magnitude x, the variation depending upon chance.
>
> It should be noted that
>
> $$\mathcal{E}(x) + \mathcal{E}(y) = \mathcal{E}(x + y), \tag{5}$$
>
> and if a is constant
>
> $$\mathcal{E}(ax) = a\mathcal{E}(x). \tag{6}$$
>
> But $\mathcal{E}(x) \times \mathcal{E}(y)$ is not the same thing as $\mathcal{E}(xy)$ unless x and y are quite independent so that every value of x can occur with any value of y, and every value of y with any value of x.

In the following pages, Whitworth applies the adjective *random* to his variable magnitudes. His first proposition, on page 7, reads as follows:

> If n random magnitudes be subject only to the condition that their sum is s, the expectation of any one of them is s/n.

On the following page, he uses both *random magnitudes* and *random quantities*. As the quoted sentence indicates, Whitworth is here using *random* in Crofton's sense.

3.2.4 Geometry and logic

Significance of the use of *random* in English goes deeper [395]. Boole's enterprise [31]. Importance of fact that the other European languages did not use *random* in the British sense.

It is notable that JSTOR reports no nineteenth-century instances of *random event* and only one instance of *random error*, and it is in the context of geometric probability [2]; the random error is that of someone shooting an arrow.

3.3 Moscow and Saint Petersburg

At the turn of the twentieth century, the Russian empire was a center of innovation in the arts and sciences. Evoking the golden age of Russian literature, the period from 1810 to 1830 when Alexander Pushkin, Mikhail Lermontov, and Nikolai Gogol had flourished, many of the writers, artists, and scientists of the period from 1890 to 1910 called theirs the "Silver Age" [39]. The Silver Age was an exhilerating time for those with means in the two great cities of the empire, Moscow and Saint Petersburg. By 1900, both had populations over a million.

The most prestigious scientific establishment was the Imperial Saint Petersburg Academy of Sciences, which had been established by Peter the Great in 1724. The two most important Saint Petersburg mathematicians in our story, Chebyshev and Markov, were both prominent members of the academy. Chebyshev became a full member (ординарный академик) in 1859, Markov in 1896. Chebyshev and Markov were also professors at the University of Saint Petersburg, and there they trained mathematicians who took places in many other universities in the empire. As a center for training mathematicians, however, Saint Petersburg was rivalled by Moscow, and an important part of our story is played by the Moscow Mathematical Society.

3.3.1 Bugaev's Moscow school

During the Silver Age, a group of Moscow mathematicians developed a philosophy that supported both their tastes in mathematics and their right-wing political and religious views. Followers of Nikolai Vasilevich Bugaev (1837–1903), who became president of the Moscow Mathematical Society in 1891, they became known as the Moscow school of mathematics.[39]

Bugaev believed that mathematical analysis, then devoted primarily to the study of continuous functions, should be balanced by number theory and what he called "arithmology"—the study of discontinuous or highly irregular func-

[39] Recent historical work concerning this group includes books in English by Loren Graham and Jean-Michel Kantor [170] and by Ilona Svetlikova [356]. Graham and Kantor emphasize the group's religious views and their influence on Russian mathematicians of a later generation, who were able to broaden mathematical analysis because they did not share their French colleagues' unwillingness to consider highly irregular functions. Svetlikova emphasizes the group's monarchism and anti-Semitism and its literary influences.

tions.[40] As he explained in his contribution to the International Congress of Mathematicians in Zurich in 1897 [45], discontinuities occur in music, in atomic physics, and in the actions of independent and autonomous individuals. Writing in French, he argued that these actions bring probability into the world (page 219):

> Une certaine part de hasard, qui apparaît dans nos actions, introduit un élément d'éventualité dans la nature même. L'éventualité entre ainsi en scène, comme une propriété essentielle de certains phénomènes du monde. Dans le monde il n'y a pas que le règne de la certitude seule. La probabilité y a aussi son empire.[41]

Mathematics, according to Bugaev, is the study of phenomena that change in quantity, and the different modes of change define two great branches of mathematics (pages 207–208):

> ... Une quantité susceptible de modification s'appelle une grandeur variable. Les quantités variables peuvent changer indépendamment de la modification d'autres quantités ou en dépendre. Selon cette modification, on les nomme quantités variable indépendantes ou dépendantes. Les quantités variables dépendantes s'appellent aussi des fonctions. Par conséquent les mathématiques apparaissent comme la théorie des fonctions. ...
>
> Les quantités peuvent se modifier d'une façon continue ou discontinue. D'après ces deux moyens de modifications des quantités, les fonctions se subdivisent en fonctions continues et discontinues, et les mathématiques pures se divisent, á leur tour, en deux grandes parties: la théorie des fonctions continues et la théorie des fonctions discontinues. On appelle généralement analyse mathématique la théorie des fonctions continues et arithmologie la théorie des fonctions discontinues. ...

Elsewhere in this same article, Bugaev mentions that original mathematical methods are not encountered in probability theory,[42] and this is sometimes taken to mean that he and his followers were not very interested in probability. But it was a commonplace of his time that probability theory is an application of mathematics, not a domain where new mathematics originates. The passage just quoted, when coupled with his explanation of how probability comes into the world, makes it clear that the philosophy of probability was in fact an important motivation for his study of "arithmology". It also foreshadows how

[40] Comment on the meaning of "continuous" and "discontinuous" in the mathematics at this time.

[41] English translation: A certain element of chance, which appears in our actions, introduces an element of randomness into nature itself. Randomness thus enters the scene as an essential property of certain phenomena in the world. Certainty does not reign alone in the world. Probability also has its empire there.

[42] Page 210: Dans la théorie des probabilités on ne rencontre pas des méthods mathématiques originales.

probabilized quantities will finally come to be understood in the 1930s: they are functions.

The Moscow Mathematical Society was the most prestigious mathematical organization in pre-revolutionary Russia. It had been organized in 1864 by Nikolaus Braschmann (1796–1866), an Austrian mathematician who had taught at Moscow University from 1834 to 1859. Chebyshev, who had studied with Braschmann in Moscow in the 1840s, was the only founding member who did not then live in Moscow. But by 1900 the majority of its members lived outside Moscow. It is not clear how active Chebyshev was in the Society or how interested he might have been in Bugaev's ideas, but Bugaev's followers counted him as a member of their "school". On the occasion of Bugaev's death, for example, Vissarion Grigorevich Alekseev (1866–1943),[43] published an article in German [6] that opened with these words:

> Noch sind nicht 9 Jahre seit dem Tode des genialen russischen Mathematikers, Mitgliedes der Moskauer mathematischen Schule, P. L. TSCHEBISCHEW († 26. Nov. 1894 a. S.) verflossen und schon hat ein neues Grab den unvergesslichen N. W. BUGAJEW († 29. Mai 1903 a. S.) verschlungen...[44]

On the same occasion, Pavel Alekseevich Nekrasov (1853–1924), Bugaev's successor as president of the Society, also published a book-length tribute to the school in the Society's journal, Математический Сборник. He called it the Moscow school of philosophy and mathematics (Московская философско-математическая школа) [269].

Probability had been one of the topics of Nekrasov's research and teaching before the turn of the century, and when he turned to philosophy he elaborated on the relation Bugaev had seen between randomness and free will. Free will, Nekrasov, argued, explains the independence between trials assumed in Chebyshev's proof of the law of large numbers. The stability predicted by the law of large numbers being confirmed by observation, free will is also confirmed. Nekrasov had become Rector of Moscow University in 1893. He became an official in the Ministry of Education in Saint Petersburg in 1905 and a partisan of the repression that followed Russia's defeat by Japan and the abortive revolution of that year. Even in the 1890s his published work on probability was more prolix and less disciplined than might have been permitted someone with less authority, and his philosophical writings after Bugaev's death have been described as particularly obscure and incoherent, even unhinged.[45]

[43] The exact date of death seems uncertain. Alekseev lived and taught in Estonia, then in the Russian empire. After the Russian Revolution, when other followers of Bugaev had to adapt to the Soviet regime, Estonia became independent and Alekseev was able to continue developing the implications of Bugaev's philosophy; in 1926 he published a pamphlet entitled *Goethe, Schiller, Herbart im Lichte des Moskauer exakten Idealismus* [7].

[44] English translation: Not nine years have passed since the death of the brilliant Russian mathematician, member of the Moscow mathematical school, P. L. TSCHEBISCHEW († 26 Nov. 1894 A.D.), and already a new grave has swallowed up the unforgettable N. W. BUGAJEW († 29 May 1903 A.D.)

[45] See Svetlikova [356], Chapter 3.

Nekrasov was particularly resented by Markov, who was known for publicly pointing out the errors of others, and who led the opposition to Nekrasov's proposal to teach probability in secondary schools.[46] Markov explained in a letter to Chuprov in 1910 ([288], page 5) that Nekrasov's argument for free will, which began with the notion that independence was an essential assumption in Chebyshev's proof of the law of large numbers, was the impetus for Markov's work showing that the law holds for many sequences of dependent quantities—those we now call Markov chains.

Nekrasov's argument that from the stability to statistics to free will also attracted attention in the West. Friedrich Maria Urban (1878–1964), a Czech mathematician then working at the University of Pennsylvania, discussed it in his 1908 book on the use of statistics in psychophysics. According to Urban, the argument was supported by Chebyshev, Nekrasov, Alekseev, and Bugaev, head of the "Moscow school of idealism". Urban felt that the argument deserved consideration, having attracted so many distinguished supporters, but that it proved too much; physical phenomena such as the tides obey the law of large numbers just as reliably as human phenomena such as births and deaths ([361], page 166). Nekrasov's and Urban's arguments were discussed by Chuprov in [68], pages 260–261.

3.3.2 Chebyshev's Saint Petersburg school

For information on Chebyshev's other mathematical work: [48]. Concerning the Moscow school: [168].

One of the most religious and most accomplished mathematicians in Bugaev's group was the analyst Dmitrii Fedorovich Egorov (1869–1931). His student Nikolai Nikolaevich Luzin (1883–1950), also religious and even more accomplished mathematically, trained many of the mathematicians who became the pride of the Soviet Academy of Sciences in the 1930s, including Aleksandr Khinchin and Andrei Kolmogorov.[47] This genealogy, obviously unsuited to Soviet propaganda, became more and more dangerous to the young Moscow mathematicians as Stalin's purges loomed in the 1930s.

The outspoken atheist Andrei Markov was a more suitable forebear for Soviet mathematicians, especially those studying probability, and Chebyshev needed to be rescued from the Moscow school's claims on him. The natural solution was to invent a "Saint Petersburg school of probability". As Sergei Natanovich Bernstein (1880–1968) told the story in 1940 [25], this school had just three members:

[46] In a letter to Markov ([288], pages 3–4), Aleksandr Chuprov, who often defended other scholars against Markov's criticism, agreed with Markov's low opinion of Nekrasov's later writing, but this does not tell the whole story. Chuprov was a gentle correspondent, who recognized the elements of truth in almost everyone's views. Seneta and Fischer give relatively positive evaluations of Nekrasov's earlier contributions [316, 332, 147]. Sheynin provides translations of some of his work [336], and Fischer looks closely at his contributions to the central limit theorem.

[47] See the chart on page 163 of [170].

1. Chebyshev, whose principle contributions to probability were his 1847 and 1867 articles on the law of large numbers and his 1887 article on the central limit theorem [66].

2. Liapunov, who also worked very little in probability but published two outstanding articles on the central limit theorem in 1900 and 1901.

3. Markov, who published a textbook on probability in 1900 and put most of his energy into the subject during the following 20 years, deepening Chebyshev's method of mathematical expectation, as Markov called it, and developing what we now call the theory of Markov chains.

As Hans Fischer has pointed out ([147], pages 140, 159), the Saint Petersburg school of probability never had the scale and cohesion that the name was chosen to suggest. Markov and Liapunov attended Chebyshev's course in probability as students at Saint Petersburg, but neither of them worked seriously on probability until after Chebyshev's death. Probability eventually became Markov's main topic of research, but this was never the case for Chebyshev or Liapunov, and Markov often approached Chebyshev's work, like that of most mathematicians, more as a critic than as a disciple. The three mathematicians did, however, have a profound effect on the theory of probability, which was felt throughout the first half of the twentieth century. Their contributions have been recounted in detail by Adams, Fischer, Seneta, and Sheynin [1, 147, 320, 330, 332, 336].[48]

3.3.3 *Sluchainaya velichina* and *peremennaya*

As we have seen, Chebyshev used величина, Russian for *quantity* as his name for a probabilized quantity, never adding an adjective to signal its randomness. But according to Sheynin ([330], page 350, and [341], page 151), several Russian authors used the term случайная величина (random quantity) in lectures notes and local publications in the 1880s:

- The Kazan mathematician Aleksandr Vasilevich Vasilev (Александр Васильевич Васильев, 1853–1929), in Теориа Вероиатностей, published in Kazan in 1885. I have not seen this publication. It is not listed in WorldCat, although other Kazan publications by Vasilev, including his tribute to Lobachevskii, are listed there. It is also not in the list of his works in his biography in Russian Wikipedia. According to that biography, Vasilev was at odds with the Soviet regime and apparently emigrated and died in Paris in 1929. He was married to Alexandra Pavlovna Maximovich. His last name is common and is often transliterated as Vasiliev.

- The Kazan mathematician Vladimir Pavlovich Maksimovich (Владимир Павлович Максимович, 1850–1889). Maksimovich, who died young after falling mentally ill, published in 1888 an article entitled О законе

[48] Discuss also Kolmogorov's exaggerated praise of Chebyshev.

вероятностей случайных величин и применение его к одному вопросу учебной статистики (On the law of probabilities of random magnitudes and its application to a problem of educational statistics). Russian Wikipedia gives the citation as Реч., чит. на унив. акте 8 янв. 1888 г./[Соч.] Проф. В. Максимовича Киев: Унив. тип., 1888. Sheynin gives it as *Univ. Izv.* (Kiev), year 28, No. 1, pages 1–21.

- The Moscow mathemtician Pavel Alekseevich Nekrasov (1853–1924), already mentioned, in lithographed lecture notes for his probability course in Moscow in 1887/1888, entitled Теориа Вероятностей.

In a series of articles on probability in Математический Сборник starting in 1898 [268], Nekrasov occasionally used both случайная величина (random quantity) and случайная переменная (random variable). He alternated the use of the terms just as Cournot had sometimes alternated between *quantité* and *variable*, clearly using the two as synonyms. He also referred to the случайная сумма of several such quantities, and he followed Chebyshev in the general use of математическое ожидание.[49]

3.3.4 Liapunov's variables

Liapunov's celebrated contribution to probability theory was contained in two articles on the central limit theorem, with proofs that were rigorous by the standards of his time and, in Fischer's words, were modern insofar as they brought "full mathematical autonomy" to the problem. They were inspired by Chebyshev's 1887 article, but whereas Chebyshev relied on the theory of moments just as in his 1867 article, Liapunov used characteristic functions.

Liapunov's articles did not appear in Russian. They were published in French in Russian mathematical journals dated 1900 and 1901, respectively [249, 250]; they were announced in the *Comptes rendus* in January and April of 1901 [251, 252].

The four publications all use *variable* to name the probabilized quantities Liapunov is considering. For example, the 1900 paper in the *Bulletin de l'Académie Impériale des Sciences* [249] opens as follows:

> Tchebychef, dans un de ses Mémoires, a montré comment les résultats de ses recherches sur les valeurs limites des intégrales peuvent conduire à la démonstration du théorème fameux de Laplace et Poisson sur la probabilité pour que la somme d'un grand nombre de variables indépendantes soumises au hasard soit comprise entre des limites données.

Liapunov may have been the first author in French to use *variable* to mean a probabilized quantity without beginning with a term in which *variable* is an adjective, as Laplace and Cournot had done. Later in his articles, however,

[49] Sheynin ([330], page 350) reports that he also used both terms in his 1888 lithographed lecture notes.

when writing about powers of the variables he is considering, Liapunov uses *quantité*.

3.3.5 Markov's method of mathematical expectation

Markov began teaching probability at the university in Saint Petersburg in the academic year 1882/1883, after Chebyshev's retirement from teaching [256, 320]. Lithographed notes of his lectures were circulated beginning that year,[50] and he published a Russian textbook on probability, Исчисление Вероянoцтей, in 1900 [258].

Like Chebyshev, Markov called a probabilized quantity simply a величина (quantity). In accord with everyone outside Russia, but in contrast with Vasilev, Maksimov, and Nekrasov, he did not add an adjective to indicate randomness. In his treatment of geometric probability in his book, he used the phrase на удачы, meaning *at random*, when writing about the random choice of a point from a line segment.[51] But in a letter to Aleksandr Chuprov in 1912 ([288], page 65; [314]), Markov stated that he considered both the adjective случайно (random) and the adverb наудачу (at random) completely undefined and avoided them whenever possible. Check whether he avoided наудачу in the 1924 edition of his book.

Markov also followed Chebyshev in calling any probabilized quantity's mean its математическое ожидание, and Исчисление Вероянoцтей emphasized this notion. As Oscar Anderson wrote in 1935 ([10], page 172):

> Markoff hat auf den Begriffe der mathematischen Erwartung sein ganzes Lehrbuch der Wahrscheinlichkeitsrechnung aufgebaut und hierbei eine außerordentliche Einheitlichkeit des logischen Aufbaues der Materie und die größte Eleganz der Beweise erzielt. In dieser Hinsicht steht sein Werk ganz einzigartig da.[52]

It is reasonable to conjecture that Markov had noticed, by the time he was writing his book, Czuber's use of the letter M to designate the expectation operator. In any case, Markov's notation for the operator was м. о., for математическое ожидание. We see this in a letter he wrote to Vasilev in September 1898, which Vasilev published in 1899 [257]. Then we see it on page 59 of the book, where we writes

$$\text{м. о. } (X+Y+\ldots+W) = \text{м. о. } X + \text{м. о. } Y + \ldots + \text{м. о. } W$$

[50] See Sheynin [330]. Sheynin has called my attention to the appendix to Markov's selected works where several versions of the notes are listed: [261], page 710, items 153–158.

[51] See page 163, for example. At the end of the chapter where he treats geometric probability (page 187), he cites Czuber's book [99]. In the 1912 German edition, he also cites Crofton [94]. The German translator renders на удачы as *beliebig*, which might be translated into English as *arbitrary*.

[52] *English translation:* Markov built his entire probability textbook on the idea of mathematical expectation, thus achieving an extraordinary unity in the logical construction of the topic and the greatest elegance in the proofs. In this respect, his work is unique

for quantities $X, Y, \ldots, W$, and on page 63, where he writes

$$\text{м. о. } XY = \text{м. о. } X \times \text{м. о. } Y$$

for independent quantities X and Y.[53] Mathematicians had made use of these properties since Laplace, and they had been noted by Hauber, Crofton, and Whitworth, but this was the first time they had been presented in a textbook as part of the elements of probability.

Following in Chebyshev's footsteps, Markov had worked on the theory of moments before considering its applications to probability, and his book showcased the use of moments to study sums of probabilized quantities (see [147], Chapter 4). In his forward to the German edition of his book, he called this the *Methode von* BIENAYMÉ-TSCHEBYSCHEFF. In a 1911 article [259], he called it the метод математических ожиданий (method of mathematical expectation).[54]

Исчисление Верояноцтей was reviewed favorably in Germany and France. In the *Jahrbuch über die Fortschritte der Mathematik*; the Russian mathematician Dmitrii Matveevich Sintsov wrote ([346]):

> Das Buch von Markow zeichnet sich durch eine eigenartige Anordnung des Stoffes und durch die Strenge der Darstellung aus.

In *L'Enseignement mathématique*, the French mathematician G. Papelier wrote ([290]):

> Il existe un grand nombre d'ouvrages sur le calcul de probabilités, mais la matière est tellement vaste que tous ces ouvrages présentent de grandes dissemblances. Le livre de M. Markoff peut être rangé parmi les plus clairs, les mieux ordonnés et les plus intéressants.

Papelier concluded that the book could be recommended to mathematicians who knew Russian. There being few of these outside Russia, the book was not immediately influential. As we will see, this changed in 1912, when a German translation appeared [260].

It is notable that Papelier, in describing the content of Markov's book, follows his lead by using *espérance mathématique* broadly, to designate the mean of any probabilized quantity, without any particular comment. This is typical of how non-Russian authors treated the broad Russian use of математическое ожидание at the turn of the century. They accepted it readily, using the equivalent expressions in their own languages without comment when describing the Russian work, but they did not choose to use it in their own work. Czuber, for example, writes *mathematischen Erwartungen oder Mittelwerte* for the means

[53]In the 1912 German translation [260], математическое and ожидание becomes *mathematische Hoffnung* (mathematical hope), and м. о. becomes m. H.

[54]In the talk he gave in Saint Petersburg to commemorate the 200th anniversary of the appearance of Jakob Bernoulli's *Ars conjectandi* ([287], pages 171–177), Markov used both these names, and also the more established term метход моментов (method of moments). In 1927, Sergei Bernshtein used способ математических ожиданий as the tile of the corresponding chapter of his textbook ([24], Chapter 3).

of the quantities under consideration when he is describing Chebyshev's 1867 results in his 1903 treatise but does not use *mathematischen Erwartungen* in this broad sense elsewhere.

3.3.6 Markov's inequality

While emphasizing Chebyshev's inequality, which bounds probabilities for a quantity's deviation from its mathematical expectation, Markov derived the more fundamental inequality we now call *Markov's inequality.* He stated it on page 63 of his book as follows:

> **Лемма.** Если A означаеть математическое ожидание величины U, все значения которой числа положительныя, а t число произвольное; то вероятность неравенства
>
> $$U \leq At^2$$
>
> больше
>
> $$1 - \frac{1}{t^2}.$$ [55]

Comment on the connection between Markov's inequality and Cournot's principle.

3.4 The German crossroads

As we have seen, the vocabulary of probabilty in Germany was strongly influenced in the mid-nineteenth century by French authors and their translators. Beginning at the end of the century, the German vocabulary also began to experience an influence from the Russia, which is passed on to the west. Much of that influence was mediated by Ladislaus von Bortkiewicz and Emanuel Czuber.

3.4.1 The role of Emanuel Czuber

The Austrian mathematician Emanuel Czuber (1851–1925) was prominent socially, having family connections with the Emperor in Vienna. He was not known for innovative contributions to mathematics, but he was very influential through the many books in which he exposited probability theory at the end of the nineteenth century and the beginning of the twentieth. These books appeared over the course of more than four decades:

Czuber's first book in probability was his translation into German, in 1879 [264], of the lectures in probability of the Belgian poet and mathematician Antoine Meyer (1801–1857); he had obtained the manuscript of the lectures, written in French, from Meyer's widow.

His later books, all published by Teubner in Leipzig:

[55] English translation: **Lemma.** If A is the mathematical expectation of a quantity U whose values are all positive, and t is any number, then the probability of the inequality $U \leq At^2$ is greater than $1 - 1/t^2$.

1. *Vorlesung über Wahrscheinlichkeitsrechnung* in 1879 [264]. This was Czuber's translation from French into German of lectures on probability by the Belgian poet and mathematician Antoine Meyer (1801–1857). Czuber obtained the manuscript of the lectures from Meyer's widow.[56]

2. *Geometrische Wahrscheinlichkeiten und Mittelwerte* in 1884 [99].

3. *Theorie der Beobachtungsfehler* in 1891 [100].

4. *Die Entwickung der Wahrscheinlichkeitstheorie und ihrere Anwendungen* in 1899 in [101].

5. *Wahrscheinlichkeitsrechnung und ihre Anwendung auf Fehlerausgleichung, Statistik und Lebensversicherung* in 1903 [103]. (The preface is dated November 1902. Later editions were in two volumes. The two volumes for the second edition appeared in 1908 and 1910, respectively. The third edition of the first volume appeared in 1914, with a preface dated November 1912.

6. *Die philosophischen Grundlagen der Wahrscheinlichkeitsrechnung* in 1923 [104].

In this sequence of books, Czuber sought to capture the summarize the best work on probability of his time, and they in turn influenced those writing about probability in German and in other languages as well. They provide, therefore an important set of mileposts for the diversity and evolution of the German vocabulary.

In his translation of Meyer's *Vorlesung*, used both *Hoffnung* and *Erwartung*, sometimes interchangeably. The book does not introduce the general concept of a probabilized quantity, but it does study both games of change and actuarial topics. At one point, the translator contrasts *Hoffnung* (hope) with *Furcht* (fear), explaining that the *mathematische Hoffnung* balances the hope of gain with the fear of loss. But often *mathematische Erwartung* is used instead, or the *Wert* of the *Erwartung* or simply the *Erwartung*. When explaining how a *mittlere Lebensdauer* (mean lifetime) is calculated as an integral, he notes that the quantities being calculating are in a certain sense *mathematische Erwartung* and *mathematische Hoffnung*. [57]

[56] I do not know if the French original survives. However, Luxembourgian Wikipedia lists a shorter book on probability by Meyer that appeared in 1857: Essai sur une exposition nouvelle de la théorie analytique des probabilités à postériori. Liège : H. Dessain. 122 p.

[57] Nun ist offenbar

$$\phi(x)dx = \frac{\psi(x)dx}{F(o)} = \frac{t_x}{F(o)}$$

die Wahrscheinlichkeit, zwischen den Altersgrenzen x und $x + dx$ zu sterben, oder die Wahrscheinlichkeit, ein zwischen den unendlich nahen Grenzen x und $x + dx$ liegendes Alter zu erreichen, das Product $x\phi(x)dx$ daher die auf dieses Alter bezügliche mathematische Erwartung; das Alter wird nämlich mit einer Summe verglichen, für deren Erlangung die Wahrscheinlichkeit $\phi(x)dx$ besteht. In diesem Sinne erscheint also die mittlere Lebensdauer als *mathematische Hoffnung*. (Page 345)

As we have already noted, *Geometrische Wahrscheinlichkeiten und Mittelwerte* brought Crofton's work in geometric probability into the German literature, translating his *mean value* as *Mittelwert* and using his $M()$. Here Czuber explained the analogy to discrete probability this way:

> Der Mittelwert einer stetigen Gesamtheit geometrischer Grössen steht zu dem Mittelwert einer diskreten Grössenreihe in derselben Beziehung wie die Wahrscheinlichkeit, die auf eine stetige Gesamtheit möglicher Fälle sich bezieht zu einer solchen, der eine diskrete Reihe möglicher Fälle zu Grunde liegt.[58]

To be continued.

Note: "dem Zufall unterworfene Größe" ([103], page 234 of Volume 1 of the 3rd edition of Czuber's *Wahrscheinlichkeitsrechnung*)

3.4.2 Bortkiewicz's *mathematische Erwartung*

Ladislaus von Bortkiewicz (1868–1931) is remembered today as an economist and a statistician, but he worked at a time when statistics was not a separate discipline from economics, both being taught in Europe primarily in faculties of law. Growing up in Saint Petersburg in a family of Polish heritage, he studied in the faculty of law in his home town and then went to Germany in 1890, studying with the statisticians Wilhelm Lexis in Göttingen and Georg Friedrich Knapp in Straßburg. Within a few years he was widely published in statistics ([33, 34, 35]) and already known for his penchant for controversy, having debated economics with Léon Walras and statistics with Frances Edgeworth. Not having found a permanent position in Germany, he returned to Saint Petersburg in 1897 to work as an actuary for the national railways. He had time to continue his scholarly work, and starting in 1899 he also taught in a prestigious secondary school. But in 1901 he accepted a professorship in Berlin. He then adopted a German identity, Germanizing his name[59] and teaching in Berlin until his death in 1931.[60]

Bortkiewicz brought two innovations into German in the 1890s:

1. Calling the mean of an arbitrary probabilized quantity, not necessarily a gain in money or the duration of life, its *mathematische Erwartung*. Bortkiewicz does this already in an article on statistics that he published in German in 1895 ([34], page 334), where he writes about the *abstrakten Durchschnittswerten* (abstract average values) of statistical quantities and then explains that these are called *mathematische Erwartung* in probability theory:

---

[58] English translation: The mean value of a continuous body of geometric quantities is related to the mean value of a discrete series of numbers in the same way as the probability of choosing from a continuous body of possible cases is related to the probability of choosing from a discrete series of possible cases.

[59] His name in Russian was Владислав Иосифович Борткевич, or Vladislav Yosifovich Bortkevich. In Polish: Władysław Bortkiewicz.

[60] For futher biographical information, see [179] and §7.3 of [68].

> Derjenige der Wahrscheinlichkeitsrechnung geläufige Begriffe, worunter jede benannte Durchschnittszahl der Statistik zu subsumieren ist, führt die Bezeichnung "mathematische Erwartung". Letztere wird definiert als die Summe aus all den Werten, die eine Größe annehmen kann, multipliziert mit den entsprechenden Wahrscheinlichkeiten jener Werte.[61]

2. Using $E()$ as the expectation operator. He does this near the beginning of his 1898 book *Das Gesetz der kleinen Zahlen* ([36], page 2):[62]

   > Verabredet man sich, die mathematischen Erwartung der Größe a in Folgendem mit $E(a)$ zu bezeichnen,...[63]

   This comes just one year after Whitworth used $\mathcal{E}()$ in English.

As we have seen, both *mathematische Erwartung* and *mathematische Hoffnung* had been used in German as the equivalent of *espérance mathématique* in French and *mathematical expectation* in English—i.e., as the name of the mean value of a gambler's gain or person's duration of life. But Bortkewicz seems to have been the first author in German to use it as the name for the theoretical mean of an arbitrary probabilized quantity. Other authors used instead *Mittelwert* or *Durchschnittswert*. So we might interpret Bortkiewicz's quotation marks almost as an acknowledgement that he was translating from the Russian.

Thanks to Oscar Sheynin's compilation of letters between Bortkiewicz and his friend Aleksandr Aleksandrovich Chuprov [334], we know that Bortkiewicz had attended Markov's course when he was a student in the law faculty at Saint Petersburg, and that he was already using $E()$ in 1896, in a draft of *Das Gesetz der kleinen Zahlen* that he asked Chuprov to criticize (Letter 6). We may assume that he was familiar both with Crofton's operator $M()$, which had been used by Czuber in German, and with Markov's м. о.. Nothing more natural than replacing M with E once he had replaced *Mittelwert* with *Erwartung*.[64] This was before Whitworth's first publication of $\mathcal{E}()$ and hence clearly independent of it.

In the fall of 1897, when Bortkiewicz was back in Saint Petersburg, he sought advice from Markov on the manuscript and asked Markov for a recommendation. An irritating three-hour conversation, he reported to Chuprov; but Markov had agreed to write a recommendation stating that his mathematics was sound (Letter 27, translated into English on page 60 of [341]).

[61] English translation: The idea in probability theory under which these average values in statistics are subsumed goes by the name "mathematical expectation". This is defined as the sum of all the values the quantity can take, each multiplied by its corresponding probability.

[62] The title translates into English as "The law of small numbers"; see Quine and Seneta [304].

[63] English translation: Let us agree to designate the mathematical expectation of the quantity a by $E(a)$ in the following,...

[64] It is notable that in their letters, Bortkiewicz and Chuprov shorten математическое ожидание to ожидание.

3.4.3 Hausdorff's *Durschnittswerth*

The German mathematician Felix Hausdorff (1868–1942) enters our tale through an article on probability that he published in 1901 [183]. The article consisted of three unrelated contributions; the first argued for a general concept of relative or conditional probability and introduced the notation $P_F(E)$ for the probability of E given F,[65] the second was concerned with estimating the precision in the Gaussian law, and the third was devoted to clarifying some of the arguments for the Gaussian law.

Hausdorff introduced what we now recognize as the expectation operator in the third section of the article. Following standard German practice at the time, he called a probabilized quantity simply a *Größe*. Considering a *Größe* x that was either discrete (taking values $x_1x_2x_3\ldots$ with probabilities $p_1p_2p_3\ldots$) or continuous (with a *Fehlerfunction* $\phi(x)$), he introduced his operator with these words:

> Das Zeichen D soll den mit Rücksicht auf die Wahrscheinlichkeit der einzelnen Werthe gebildeten Durchschnittswerth (Mittelwerth, valeur moyenne oder valeur probable) irgendwelcher Function von x bedeuten, ist also durch
>
> $$Df(x) = \sum_i p_i f(x_i) \quad \text{oder} \quad Df(x) = \int_{-\infty}^{\infty} \varphi(x) f(x) dx$$
>
> definirt...

In the ensuing analysis, Hausdorff rediscovered Thiele's semi-invariants.

Hausdorff says little about the work of other authors; the only precise citations are to Gauss and Bruns. It seems plausible and even likely that he introduced his operator D without ever having been aware of Bortkiewicz's E, Crofton and Czuber's D, Whitworth's $\mathcal{E}$ or Markov's м. о..

Fischer ([147], pages 116–118) reports that Hausdorff's contribution did not influence subsequent authors working on the central limit theorem. All three of Hausdorff's contributions in the 1901 article did receive attention, however, in the the first volume of the second edition of Emanuel Czuber's widely read *Wahrscheinlichsrechnung* [103], which appeared in 1908. In his first edition, published in 1903, Czuber had used $\mu(S)$ for the *Mittelwert einer vom Zufall abhängigen Größe* S. In the second edition, which cites Hausdorff's paper, he used $\mathfrak{W}_F(E)$ for the relative probability of E given F, with an explicit acknowledgement to Hausdorff (page 45), and he replaced $\mu(S)$ with $\mathfrak{D}(S)$ (page 91).

3.4.4 Markov's *Wahrscheinlichkeitsrechnung*

In spite of positive reviews, Markov's Исчисление Верояноцтей had little impact outside Russia. But the influence of its 1912 German edition, *Wahrschein-*

[65] Hausdorff wrote $P_F(E)$ for the probability of E given F. Markov had written (A, B) for the probability of B given A ([258], page 16). For other early notations for this concept, see [325], Appendix I.

lichkeitsrechnung, was very substantial. Oscar Anderson was probably correct when he suggested that the most influential features of the book were its clarity and overall organization, which emphasized expectation and the law of large numbers. But the book's influence is illustrated most vividly by the subsequent adoption of a more discrete innovation: the use of upper case Roman letters (e.g., X and Y) for probabilized quantities and the corresponding lower case letters (e.g., x and y) for their values.

Table 4 attempts to list all the books on probability, in languages other than Russian, that picked up this innovation in the first half of the twentieth century. There were other books on probability during this period, some of which were influenced by Markov in other ways, but the stature of the works in Table 4 suffices to demonstrate Markov's influence, and the debts they owed each other tells us something about the path that influence followed. The table also provides some glimpses of later chapters in our story, by listing the names these authors used for a probabilized quantity and its mean, as well the symbols they used for the expectation operator. Here are some further details:

- In Czuber's case, we see a sharp change from the first volume of the second edition of his treatise, which appeared in 1908, to the corresponding volume for the third edition, which appeared in 1914. In 1908, Czuber had written $S_1, \ldots, S_\nu$ for the possible values of a *veränderliche Größe*. As I have already mentioned, he used $\mathcal{D}(S)$ in 1908, for S's *Durschnittwert* or *Mittelwert*.

- Castelnuovo followed Cantelli. I need to check the first edition. In the second edition, he states that he has used the earlier textbooks by Bertrand, Poincaré, Czuber, Markoff, Bachelier, Fisher, and Coolidge.

- In his preface, Coolidge mentions "the encyclopaedic but readable text of Czuber", "the translation of Markhoff, with its unusual attention to rigour", and "the recent work of Castelnuovo, careful, critical, and judicious". The references are to the second edition of [103], [260], and [59].

- Lévy praised Castelnuovo's book in the first paragraph of the preface of his 1925 book, *Calcul des probabilités*. Although he used the $X, Y \ldots, x, y, \ldots$ convention in his books, and also in one article prior to the book ([241], a note in the *Comptes rendus*), he often did not use it in his articles.

- In the bibliography of Darmois's *Statistique matématique*, we find of course all the preceding books in this list: Markov, Czuber, Castelnuovo, Coolidge, and Lévy.

- Starting in 1931, Lévy used $\mathcal{E}\{\}$ in some of his articles (e.g., [243]). In his 1937 book, *Théorie de l'addition des variables aléatoires*, he states that he is switching from $\mathcal{E}\{\}$ to $\mathcal{M}\{\}$ in order to agree with Fréchet.

- By 1937, Markov's own book was no longer fresh, but Cramér includes Castelnuovo's book and Lévy's 1925 book in his bibliography.

One book that tragically misses this list is Aleksander Chuprov's book on the theory of correlation, which appeared in German in 1925. Chuprov left it to his student Nikolai Chetverikov to translate the book into Russian. The German original, *Gurdbegriffe und Grundprobleme der Korrelationstheorie*, appeared in 1925 [75], the Russian translation in 1926. The Russian translation follows Markov's $X, Y \ldots, x, y, \ldots$ convention, but German original is inconsistent, even garbled.[66] We know that Chuprov was in great difficulties during this period, falling ill in Prague, where he was seeking employment, in July 1925 and then traveling to Italy in an effort to improve his health before dying at the home of a friend in Geneva in April 1926 ([341], pages 44–45). It seems likely that the failure of the German text to follow Markov's convention was due to Chuprov's not being able to correct its proofs.

Markov's convention had to compete with George Udny Yule's convention, which used X for the values of a variable and x for those values' deviations from the empirical mean ([393], page 134). But after Feller used Markov's convention in his immensely popular textbook in 1950, it became popular even in mathematical statistics.

3.5 *Expectation* on the eve of Great War

As we can see already in Table 4, the vocabulary of probability theory would change in three basic ways between the two world wars. First, the word *variable* would be generally accepted as the general name for probabilized quantities. Second, *expectation* (or *espérance* or *Erwartung*) would become acceptable as a name for a theoretical mean of arbitrary probabilized quantities. Third, the notation $E()$ would be adopted.

In the years before the World War I, these changes had not yet taken place, but they were in the air in various ways. Among authors in probability theory, the word *variable* was used occasionally as a general name for a probabilized quantity, perhaps most intensely in geometric probability, but this was hardly new. Such occasional use had started with Laplace. The broadening of the use of *expectation* was more in the air.

It is not easy to judge how important Whitworth's example was to the eventual adoption of *expectation* for *mean value* and the use of E to designate the expectation operator. The elementary tone and pedagogical purpose of Whitworth's publications left them mostly outside the chain of citations and acknowledgements in the theoretical literature, but they were seen by many students.[67] *Choice and Chance* was widely distributed and widely used, at least in England and the United States. Frances Edgeworth cited it repeatedly in his article on probability in the 11th edition of the Encyclopædia Britannica, published in 1911. In an article commemorating the hundredth anniversary of its appearance, J. O. Irwin recalled that he had used *Choice and Chance* when lecturing

[66] The later English translation [76] followed the German version.

[67] *Choice and Chance* and *Expectation of Parts* did appear in Keynes's bibliography in 1921 [200], and Whitworth has been credited with several substantive contributions. See for example [173].

on probability in Karl Pearson's department at University College London in 1921 [194], and it was cited as supplementary reading by a number of probability textbooks used in the United States in the first half of the twentieth century, including the 1928 text by Thornton E. Fry [160] and the 1937 text by James V. Uspensky [362].

Edgeworth: used *expectation* in the section heading but talked about *means* in the text.

In France, Bachelier and Bertrand: acknowledged that *moyenne* and *espérance mathématique* are the same thing, but insisted on the distinction.

- Bertrand, [26], page 61: La valeur probable d'une grandeur inconnue a est, par définition, l'espérance mathématique de celui qui devrait recevoir une somme égale à a.

- Bachelier [13], page 58: Les notions de valeur moyenne et d'espérance mathématique sont analogues et même identiques. L'espérance est la valeur moyenne d'un gain.

Lexis had used *Erwartungswert* for the value of a gambler's expectation, just as Oettinger had done in 1852. See for example page 438 of [247].. Arne Fisher's used *expected value* in [148], presumably inspired by *Erwartungswert*. Was this the first use in English?

Discuss vocabulary of Czuber [102], Bortkiewicz [37], and Bohlmann [29] in Wilhem Franz Meyer's *Encyklopädie der Mathematischen Wissenschaften* [265].

Table 4: Probability books published in the first half of the twentieth century in languages other than Russian that followed Markov in using $X, Y, \ldots$ for probabilized quantities and $x, y, \ldots$ for their possible values. (This list is meant to be exhaustive; please let me know about any others that I have missed.)

	Name of quantity	*Name of mean*	*Operator*
1914	Emanuel Czuber in German [103], third edition of volume 1.		
	Größe	*Mittelwert*	$M(X)$
1919	Guido Castelnuovo in Italian [59].		
	variabile causale	*valore medio teorico*	$M(X)$
1925	Julian Coolidge in English [77].		
	variable	*mean value*	None
1925	Paul Lévy in French [242].		
	variable éventuelle	*valeur probable*	None
1928	Georges Darmois in French [106].		
	variable aléatoire	*valeur probable* *espérance mathématique*	$E(X)$
1937	Maurice Fréchet in French [154].		
	variable aléatoire	*valeur moyenne*	$\mathcal{M}\{X\}$
1937	Paul Lévy in French [245].		
	variable aléatoire	*valeur probable*	$\mathcal{M}\{X\}$
1937	Harald Cramér in English [89].		
	random variable	*mean value* *mathematical expectation*	$E(X)$
1949	Harald Cramér in Swedish [90]. (I have not yet seen the 1927 version of this book.)		
	tillfällig variabel *stokastisk variabel*	*medelvärdet*	$E\{X\}$
1950	William Feller in English [145].		
	random variable	*expectation* *mean* *mathematical expectation* *expected value* *average*	$\mathbf{E}(X)$

4 Statistical series vs. random samples

Historians agree that the term *mathematische Statistik* was coined by the German actuary Theodor Wittstein in the 1860s, and that it emphasized the application of probability to statistics ([303], page 245; [101], page 231).

Discuss Pearson's idealism [291], its criticism by Lenin, its difference from the idealism of the Moscow school.

4.1 Kollektivmasslehre

Look at vocabulary of Lexis [246, 247], Bruns [44], and Fechner [136].

4.2 Galton's *variables*

Francis Edgeworth (1845–1926) may have been the first to use *variable* in English to mean a probabilized quantity. Here is a passage from one of his discussions of the conditions for the "law of error"—the law that says that errors are normally distributed:

> Each observation must be one and the same definite and constant function of a number of variable elements. Each variable assumes for different observations different values according to some law of facility, and the number of the variables on which an observation depends must be large. (1887 [132])

The influence of this passage may have been limited, for in most of Edgeworth's discussions of the law of error he shortened *variable element* to *element* rather than *variable* ([147], page 122). See for example [130].

And Edgeworth [131] page 226: the quantity of advantage that Laplace calls *espérance*

Far more significant and influential was the systematic use of *variable* by Francis Galton in his 1888 article in the *Proceedings of the Royal Society of London*, where he studied the "co-relation" between variables [163] and by Edgeworth in subsequent work on measuring such correlations when there are more than two variables [133].[68]

Karl Pearson (1857–1936) was far more successful than Edgeworth in making Galton's eugenics mathematical; he transformed Galton's ideas into an international mathematical project [354]. But he was not as quick as Edgeworth to use the term *variable* in this connection. In his earliest articles, Pearson used more concrete terms, such as *measurement*, *organ*, *character*, or *observation*. But as his work became more mathematical, *variable* crept in. George Udny Yule (1871–1951), who began his career in 1893 as a teaching assistant to Pearson at University College, may have been responsible for the shift. In 1895, in his first publication in statistics, Yule wrote:

[68] In addition to Cournot and Lyanpunov, other mathematicians occasionally used cognate terms in various languages. One example is T. N. Thiele, who used the Danish *variabel* occasionally in his 1889 book; see [239]. Pearson, who knew Norwegian, appears to have read and appreciated the book [176].

> ...if one variable be correlated with the rate of change of another the two variables will, in general, also be correlated. [390]

Then, in a note read to the Royal Society of London in February 1897 [391], Yule used *variable* systematically, and this article was followed by comments by Pearson [292] and Galton [164] that also used the term. Yule used *variable* systematically again in another article published that year in the *Journal of the Royal Statistical Society* [392]. It is not far-fetched to suppose that Liapunov's use of *variable* in French might have owed something to these English examples.

Yule's textbook, *An Introduction to the Theory of Statistics* [393], appeared in 1911 and ran to fourteen editions during his lifetime. Its readers learned to call dichotomous characters *attributes* and numerical characters *variables*,[69] and they learned about frequency distributions, dispersion, correlation, regression, and sampling. Probability was relegated to an appendix.

4.3 Davenport's *variates*

Pearson used *variable* regularly after 1897; the famous article in which he introduced the χ^2-test (1900 [293]) had *variable* in its lengthy title:

> On the Criterion that a given System of Deviations from the Probable in the case of a Correlated System of Variables is such that it can be reasonably supposed to arisen from Random Sampling

But in the following decades he and some of his British colleagues began to use, alongside or in place of *variable*, a neologism: *variate*.

The earliest use of *variate* as a noun that the Oxford English Dictionary has located is in a small manual published in 1899 [110] by Charles Benedict Davenport (1866–1944), an American who later became prominent as a eugenicist [289] and even as an American supporter of the Nazis. Davenport was then an instructor in Zoology at Harvard and shortly became one of the founding editors of *Biometrika*. He gives these definitions at the beginning of the manual:

> A *character* is any quality common to a number of individuals.
>
> The *magnitude* of a character is a quantitative expression of the character.
>
> A *variate* is a single magnitude-determination of a character.

For Davenport, *variates* were actual numbers in the notes of a field biologist. Roughly speaking, at least, his *magnitude of a character* is synonymous with Yule's *variable*. If we randomly sample ten plants from a certain population and measure their height, then height is the variable, and the ten heights are variates.

I have not seen *variate* in Yule's writings;[70] nor does it seem to have been used by William Sealy Gosset (1876–1937), the statistician for the Guinness

[69] Arthur Lyon Bowley's more elementary *Elements of Statistics*, published in 1901 [40], had not used the term *variable*.

[70] Yule and Pearson quarreled over other issues around the time Pearson began using *variate*. See [353], Chapter 1, and [255].

brewery who made important contributions to *Biometrika* under the pen name *Student*. But Pearson became fond of *variate*, perhaps because it seemed so practical, as did Ronald Aylmer Fisher (1890–1962), whose influence kept it current in English for over half a century. English-language textbooks in statistics that used *character* for what we would now call a *variable* and *variate* for its values included *Researches into the theory of probability*, published by Carl V. L. Charlier in Sweden in 1906 [63], and *Introduction to Mathematical Statistics*, published by J. L. West, an assistant professor at Ohio State University, in 1918 [378]. In spite of their names, these books were primarily teaching descriptive statistics, and this context the *character/variate* distinction works very well.

But as eager as they were to address research workers, Pearson and Fisher were primarily mathematicians, and in mathematical discourse the distinction between *character* (or *variable*) and *variate* proved unstable. Over time, *variate* tended to became a synonym for *variable*. We can see this tendency already in 1905, in the article where, according to JSTOR, *variate* first appears in *Biometrika* [294]. In this article, Pearson discusses the topic on which we quoted Edgeworth: the argument that observations will follow a normal distribution when each is the same function of many small independent random influences. The passage we quoted from Edgeworth omitted the condition that the influences (which Edgeworth was calling variables) be independent. If they are not independent, non-normal distributions can result. Pearson discusses this as follows (here the observations, not the small influences, are the variables):

> ... all these attempts ... amount to abolishing the third of the Gaussian assumptions, namely that small increments of the variable or the character are independent of the total already reached. That is to say that they amount to saying that increments of the variate are *correlated* with the value of the variate already reached. (pages 203–204)[71]

The distinction between "increments of the variable" and "increments of the variate" is elusive.

Here are a few examples of Fisher's use of the two words in his *Statistical Methods for Research Workers* (1925, [150]):

> The type of diagram in most frequent use consists in plotting the values of a variable, such as the weight of an animal or of a sample of plants against its age, or the size of a population at successive intervals of time. (page 27)
>
> A variate is said to be normally distributed when... (page 45)
>
> The two variates bear very different relations to the regression line. (page 147)

[71] The three assumptions are that the influences be small, numerous, and independent. These assumptions were implicit in Laplace's work and explicit in later work by Bessel and Hagen; see [147], Chapter 3. They were not discussed by Gauss. Pearson's careless reference to Gauss may have been a courtesy towards the German critics to whom he was responding.

Here is another example, from Fisher's 1924 article on the distribution of the partial correlation coefficient [149]:

> Let $x_1, x_2, \ldots, x_n$ represent the n values of one variate in the sample, and $y_1, y_2, \ldots, y_n$ the n values of the second variate...

It is difficult to reconcile all these examples with the idea that variates are values of a variable. In the last example, at least, *variate* seems to have replaced *variable*. In any case, the only way to translate it into another language would be to use your usual translation for *variable*: *variable* in French, *Veränderliche* or perhaps *Variable* in German, *variabile* in Italian, переменная in Russian.

In 1930, B. L. Shook distinguished between *variable* and *variate* in this way:

> *Variates.* Practically all statistical data is obtained as the result of observations that endeavor to establish the magnitudes of certain variables. The individual magnitudes that are recorded are termed variates. Thus in computing the average rainfall of a region, the variable is rainfall, and the amount of rainfall for any single year is a variate. Likewise, if the bank clearings for the city of New York be under consideration, then the variable is bank clearings, and the clearings for any specified interval is a variate. [342]

Comment also on Galton's use of *variate* and *deviate* in 1907 [165].

Comment on Sam Wilks's use of mean value or expected value; variates of fixed and stochastic types ([382] 1937).

4.4 Gossett and Anderson: Back to time series

Pearson's conceptual starting point was the notion of random sampling from a fixed population. Each variable had a frequency distribution in the population; each pair of variables had a joint frequency distribution. If the population were very large, or the sampling was with replacement, then variates (values for a variable in a random sample) would be what we now call independent random variables with that distribution, which we now call a probability distribution rather than a frequency distribution. Working in an English tradition that nearly equated probability with frequency, and wanting to make his mathematics as accessible as possible to biologists, Pearson avoided talk about probability when he could, thus creating puzzles for Continental mathematicians. Were the probabilities attached to the variable or to the variate? Only the variate could be considered a *random* object, but it might make sense to call the variable a *chance* variable.

The most important limitation of the picture of random sampling was its doubtful applicability to observations over time. As we have just seen in the words of Yule and Fisher, time series were prominent in the work of the biometricians from the outset. In the early years of the century, a number of authors proposed to bring time series into the random-sampling picture by taking differences of successive observations. Perhaps the changes from one time period to another would look like independent draws from some frequency distribution. As Student pointed out in *Biometrika* in 1914 ([355], page 180), it might be necessary to take differences more than once in order to obtain "random variables independent of time". This is the earliest use of *random variable* I have seen. In the next issue of *Biometrika*, [62], Beatrice M. Cave and Pearson wrote about "random variates uncorrelated to each other" and dubbed the differencing of successive observations the *variate difference method*. These isolated instances did not establish *random variable* or *random variate* as technical terms, but the variate difference method attracted enduring international attention, and we know that these articles were read by many people who wrote in French, German, Italian, or Russian.

In the same issue as Cave and Pearson's article, *Biometrika* published an article in German [8] by Oskar Anderson (1887–1960), a young Russian of Baltic German parentage, who explained that Student's results could be obtained using the method of *mathematische Erwartung* (mathematical expectation), which had been developed by Russian and German authors but neglected by the English. Here is how he said this and how he introduced the concept of mathematical expectation:

> *Methode.* Die englische statistische Schule vernachlässigt in ihren Untersuchungen ein Verfahren, das von russischen und deutschen Gelehrten oft angewandt wird (Tchebycheff, Markoff, v. Bortkiewicz, u.s.w.) und neben großer Strenge und Exaktheit noch den Vorzug hat recht elementar zu sein—die Methode der mathematischen Erwartung nämlich. Mathematische Erwartung einer Größe (A) heißt bekanntlich soviel als das Produkt aus dieser Größe und ihrer Wahrscheinlichkeit (w), also Aw. Wenn eine Variable eine Reihe einander ausschließender Größen annehmen kann, so ist deren math. Erwartung als die Summe der Erwartungen aller dieser Größen definiert. Wir werden hier die mathem. Erwartung überall durch das Symbol $E(\ )$ bezeichnen. $E(A)$ is also, z. B., gleich Aw.

If you get the amount A with probability w, the value of your expectation is Aw, and if there are several different amounts you might get, with different probabilities, you add up the mathematical expectations to get your total mathematical expectation. This explanation, which seems roundabout today, was standard at the time. We see it in many probability textbooks of the day.[72]

But Anderson did something that was not standard: he used *Variable* as the name of a probabilized quantity. The standard German term at that time was

[72]In French, Poincaré: [296], 1912, page 64. In German, Emanuel Czuber (1851–1925): [103], third edition, 1914, pages 72 and 226 of Volume 1. In English (the year after Anderson's article), Arne Fisher (1874–1944): [148], first edition, 1915, pages 49–50. Fisher was a Danish-American actuary. The first edition of his book, based on original notes in Danish, can be considered the first twentieth-century textbook on probability in English. He used *variable quantity* for a probabilized quantity but did not simplify this to *variable*. His second edition, which appeared in 1922, included material on mathematical statistics, drawing especially on work of the Scandinavian statisticians, that he had originally intended for a second volume.

Größe (quantity).[73]

5 *Variabile casuale*

In the early decades of the twentieth century, Italy emerged as a leader in probability and statistics. As in other continental countries, two major academic forces were in play: mathematicians, often responsible for teaching probability to students preparing for actuarial work, found ways to apply their mathematical results to probability theory, and economists, often working in faculties of law, developed the theory of statistics. Major figures included the mathematician Vito Volterra (1860–1940), who had close ties with the French functional analysts, Corrado Gini (1884–1965), the statistician best remembered for his indices of inequality, and the mathematician and actualry Francesco Paolo Cantelli (1875–1966). As Eugenio Regazzini [305] has put it, the Italian panorama of probability and statistics from 1910 to 1930 was dominated by Gini and Cantelli.

Cantelli, the youngest of these individuals, is remembered for many contributions, including an early attempt to axiomatize probability and his discovery of what we now call the law of large numbers [19, 315, 305, 20]. In our story, he stands out as the first mathematician to bring together the work of the English statisticians and the Russian mathematicians. From this fusion emerged the Italian term *variabile casuale*, which led, tortuously, to the English *random variable*.

5.1 The evolution of Cantelli's terminology

Cantelli's Italian terminology was influenced both by the English statisticians' use of *variable* and by Liapunov's use of *variable* and *espérance mathématique* in French. We see this in the evolution of the terminology in his published papers:

- In an introduction to probability theory that Cantelli published in the 1905–1906 volume of a journal for secondary school graduates, he called a probabilized quantity a *grandezza sconosciuta* (unknown quantity) and called its mean its *valore probabile* ([51], page 69). (Cantelli had taught in secondary schools in Palermo from 1899 to 1903.)

- In a 1905 article in which he proposed an alternative to Pearson's method of moments [50], he used in passing the word *variabile* to refer to the variables involved. (He had began working in the Italian government's actuarial institute in Rome, the Istituti di Previdenza, in 1903.)

- In a more theoretical article in 1910 [52], on the estimation of probabilities from moments, he cites work in French by Chebyshev and Liapunov, and he adopts two terminological innovations: he uses *quantità casuale* for a probabilized quantity, and following Liapunov, he adopts *speranza*

[73] When Cournot's book was translated into German in 1849 [79], the translator rendered his *variable* as *veränderliche Größe* and his *moyenne* as *Mittel*.

matematica for its mean. (The first appearance of *speranza matematica* in the article is in a quotation of Liapunov, translated from the French into Italian.)

- Finally, in 1913 [53], in an article in which he proves a simple formula that Gini had discovered but not rigorously proven, he uses *variabile casuale* as his name for a probabilized quantity. The formula gave the mean value of the absolute value for the difference between two independent observations from a binomial distribution; Gini had published it in his pathbreaking book on indices of inequality, which appeared in 1912 [166].[74]

The Italian adjective *casual* can mean *informal*, as in English, but its more usual meaning, now as in Cantelli's time and before, is *random*.[75] Gini had used it with this meaning, along with *accidentale*, in his 1912 book.But applying this adjective to *variabile* (or even to *quantità*) was an important innovation. So far as we know, the only precedents for such a usage were those in Russian by Maksimov, Vasilev, and Nekrasov, and it seems unlikely that Cantelli would have seen them at this point, if ever. Very likely he wrote *variabile casuale* in the 1913 article simply to emphasize that the *variabile* had probabilities for its possible values. Such a clarification might have seemed needed, because Cantelli was responding to Gini, a statistician. When Gini wrote about a *variabile*, as when Pearson or Galton wrote about a *variable*, it was not always clear that there were probabilities in the background.

During World War I, while still working at the Istituti di Previdenza, Cantelli began working with a more senior mathematician, Guido Castelnuovo (1865–1952), to develop a program in probability and statistics at the University of Rome. Responding to questions raised by Castelnuovo, he began studying the notion of convergence for probabilized quantities. His articles on this topic, beginning with two in 1916 [54, 55], all used *variabile casuale*. The second of the 1916 articles [54], on the law of large numbers, cites both the German edition of Markov's textbook and the second edition of Yule's textbook, both of which appeared in 1912. Here Cantelli adopts Markov's X, x convention for the *variabile casuale* and its values, and he uses $E(X)$ for the *valore medio* or *speranza matematica* of the *variabile casuale* X. He does not mention his source for $E(X)$; presumably it comes either directly or indirectly from Whitworth or Bortkiewicz or both. He had surely seen Bortkiewicz's 1898 book, and he quite likely would have seen Anderson's use of $E(X)$ in *Biometrika*.

Discuss Cantelli's strong law of large numbers later [56].

Cantelli continued publishing in the 1940s, and he always used *variabile casuale*. After his 1910 article, he generally uses *valore medio* for the mean, but he occasionally acknowledges that *speranza matematica* is an alternative name for it. He sometimes used $E(X)$ but also sometimes used $M(X)$.

[74]In 1935 ([57], page 43), in a review of his work in this period, Cantelli forgot about his use of *variabile casuale* in 1913 and stated that he had first used the term in 1916. This statement recurs in some secondary sources.

[75]Like the English words *case* and *chance*, it derives from the Latin *casus*, meaning *fall*.

Castelnuovo is remembered for research in algebraic geometry, not for research in probability. But in 1919, he published a treatise on probability, *Calcolo delle probabilità*, which adopted Cantelli's terminology and notation and was widely read in Europe and the United States the 1920s. After explaining that Cantelli had introduced the term *variabile casuale*, Castelnuovo defined it in the discrete case as follows:

> *Chiameremo variabile casuale una quantità variablile X che può assumere vari valori reali $x_1, x_2, \ldots, x_h$, secondo che si presenti uno degli eventi incompatibili $E_1, E_2, \ldots, E_h$, di probabilità note $p_1, p_2, \ldots, p_h$ aventi la somma*
>
> $$p_1 + p_2 + \cdots + p_h = 1.$$
>
> ([59] 1919, page 30)

As we noted in Table!4, Castelnuovo called the mean of a *variabile casuale* X its *valore medio* (mean value) and designated it by $M(X)$. He discussed altenative names this way (page 40):

> Il valor medio da alcuni autori è detto *valore probabile* (con una locuzione che si presta ad equivoci, perchè talvolta usata in senso diverso); da altri *speranza matematica*. Effettivamente la speranza matematica inerente a guadagni aleatori è il valor medio dei detti guadagni, quando si tenga conto di tutte la alee che può correre il giuocatore (vincite e perdite).[76]

Less encyclopedic than Czuber's *Wahrscheinlichkeitsrechnung*, written with the same clarity and organized in a similar way as Markov's *Wahrscheinlichkeitsrechnung* and in a language more congenial to the French, Castelnuovo's book was extremely influential.

5.2 Translating *variabile casuale*

In the 1920s, the decade following the first appearance of Castelnuovo's *Calcolo delle probabilità*, *variable* became the standard name for a probabilized quantity in French and English. Some authors used it, as Cournot had done, without any adjective. Others found ways of translating Cantelli's *casual*.

5.2.1 French translations

The most direct translation of *variabile casuale* into French is *variable casuelle*. The Swiss mathematician Louis-Gustave Du Pasquier (1875–1957) used this term in his 1926 book, *Le calcul des probabilités, son évolution mathématique et philosophique* ([128], page 150). Du Pasquier also used *grandeur variable* and *quantité casuelle*. He wrote:

[76] English translation: Some authors call the mean value the *probable value* (a term that lends itself to misunderstandings, because it is sometimes used in a different sense); others call it the *mathematical expectation*. The mathematical expectation of uncertain gains is in fact the mean value of those gains, when account is taken of all the gambler's chances for winning and losing.

> Par "quantité *casuelle*", nous entendons une variable, continue ou non, dont la valeur dépend du hasard ou d'événements aléatoires.

Du Pasquier wrote $M(X)$ for the *valeur moyenne* (mean value) of a *quantité casuelle* X, noting that it is *théorique* and distingushing it from any *valeur moyenne empirique* that might be obtained from data. He noted, as Castelnuovo had done, that the theoretical mean value of X is the same as the *espérance mathématique* of a person who receives X:

> La valeur moyenne d'une variable casuelle est égale à l'espérance mathématique qui s'y rattache.

But, following all the earlier authors in both French and Italian, he still kept the concepts separate.

The French mathematician Paul Lévy, who was much more influential, used instead *éventuel*, which, as we have seen, had been present in probability theory since Condorçet. In his *Calcul des probabilités*, published in 1925 [242], Lévy introduced the concept of a probabilized quantity with these words (page 54):

> On appelle *variable éventuelle* une variable obéissant à une loi de probabilité.

Lévy's example was noticed and followed. The Russian statistician Evgenii Slutskii used *éventuel* in a number of notes in *Comptes rendus* starting in 1927; in one in 1928 [348], he uses *variable éventuelle* and *valeur éventuelle* to make the variable/variate distinction: a *valeur éventuelle* is the value of a *variable éventuelle*. Other scholars who used *variable éventuelle* in French included the Polish mathematicians Antoni Łomnicki and Stanisław Ulam [254] and Cantelli himself [57].

The term *variable accidentelle* can also be found—e.g., in a note by Jerzy Neyman in 1929 [272] and later in Risser and Traynard's 1933 statistics textbook ([308], page 120).

In all these instances, the authors in French used the adjective (*casuel*, *éventuel* or *accidentel*) sparingly. Generally they used the adjective in introducing the concept and then dropped it, writing merely *variable*.

A more innovative translation of *casual* into French was *aléatoire*. I call this innovative because in the early twentieth century the adjective *aléatoire* still connoted contingencies that determined gains and losses, in gambling, contracts, or finance. So far as I have seen, it was not yet used to indicate a fortuitous element in other domains. But as we will see in §7, two French mathematicians, Georges Darmois and Maurice Fréchet, did begin to use *aléatoire* in their teaching in the 1920s.

5.2.2 English translations

Why not translate *variabile casuale* into English as *random variable*? Some authors did just this.

- In 1926 [309], the Russian mathematician Vsevolod Ivanovich Romanovskii (1879–1954), graduate of Saint Petersburg University and professor at Tashkent, used the term in an article he published in English the *Bulletin de l'Académie des Sciences de l'URSS*. He wrote:

  > Denoting with $E(t)$ the mathematical expectation of any random variable t....

  In 1929, Romanovskii even published an article, probably the first, with *random variable* is its title "On the moments of means of functions of one or more random variables" [310].

- The American economist and statistician Harold Hotelling, then at Stanford, used *random variable* in passing in 1927 [190].

- In 1928, in an article devoted to explaining Italian work on probability to British actuaries [360], Charles Trustam began his introduction to Cantelli's 1916 paper on the law of large numbers with these words:

  > ... Prof. Cantelli investigates fundamental conceptions introduced by Tchebycheff and now forming an integral part of the classical structure of the Calculus of Probabilities. The abstract notion of a "casual" or "random" variable plays an important part in the paper ...

  After putting "random variable" in quotation marks one more time, he used it freely.

This translation was not taken up, however, in books and research articles on probability theory. For the experts, the word random was still too closely tied to the notion of uniform distribution in geometric probability.

A few authors began to use *chance variable*, which is a natural translation into English of *variabile casuale* and *variable éventuelle*. The first instance I have found is Chuprov's 1925 article in the *Journal of the Royal Statistical Society* [75]. During the 1930s, three authors used *chance variable* in the newly launched *Annals of Mathematical Statistics*: Edward L. Dodd in 1930 [122], William Dowell Baten in 1930 and 1934 [17, 18], and Joseph L. Doob, in 1935 [124].[77]

Julian Lowell Coolidge, professor mathematics at Harvard, used *variable* without any adjective. Although Yule's popular textbook, which used *variable* in a statistical context, had been popular since 1911, Coolidge's 1925 textbook, *An Introduction to Mathematical Probability*, was the first book in English that systematically used *variable* as a general name for a probabilized quantity [77]. On page 60, Coolidge wrote:

[77] Of the 17 American mathematicians who founded the Institute of Mathematical Statistics in 1935, Dodd was one of the oldest (Ph.D., Yale, 1904), and Doob was one of the youngest (Ph.D., Harvard, 1932) [193].

> If a variable take the different values $V_1 V_2 \dots V_n$ with the respective probabilities $p_1 p_2 \dots p_n$, and these are all the possible values for that variable, then the expression $\sum_{i=1}^{i=n} p_i V_i$ is called the *mean value* of that variable.

In his preface, Coolidge listed Castelnuovo's Italian text as one of his models, calling it "careful, critical, and judicious". Coolidge's example was followed by Thornton C. Fry, in his 1928 book, *Probability and its Engineering Uses* ([160], page 117).

By the end of the 1920s, it seems fair to say, *variable* was as standard among mathematicians doing probability theory in English as it was among statisticians writing in English. William Burnside, in his 1928 *Theory of Probability* [47], William Burnside used *number* instead:

> If a number can take any one of the distinct values $a_i, (i = 1, 2, \dots, n)$, and if the probability that the number takes the value a_i is $p_i \dots$

But Burnside was as out of date as a mathematician of probability as he was as a statistician. (See John Aldrich's account of his interaction with the English statisticians [4].)

Discuss also the lag in the British use of *expectation* [3]

6 *Zufällige Variable*

In 1917, a few years after Cantelli had introduced *variabile casuale* in Italian, Bortkiewicz introduced *zufällige Größe* into German, and the following year Chuprov made this into the *zufällige Variable*. Bortkiewicz and Chuprov had undoubtedly seen the Russian equivalents of these terms in Nekrasov's writings, and although they were not admirers of Nekrasov, he has clearly an influence. Were they also influenced by Cantelli? Germany and Italy were at war with Italy and Russia beginning in 1914, and although Borkkiewicz and Chuprov were always widely read, we have no direct evidence that either of them saw Cantelli's Italian articles of 1913, 1916, and 1917.

Bortkiewicz and especially Chuprov were also responsible for the introduction of the term *stochastic* into twentieth-century probability theory. For Chuprov, at least, the term carried a very strong idealistic connotation, revealing another influence of the Moscow school of mathematics.

6.1 Chuprov's innovations in English

When Oscar Anderson published his response to Gossett in *Biometrika*, in 1914, he was a student in Saint Petersburg of Aleksandr Aleksandrovich Chuprov (1874–1926), an erudite and energetic Russian mathematician who admired the English biometricians and made it his mission to help them and the Continental mathematicians learn from each other. In 1918, Chuprov made his case in an article in English in *Biometrika* [69].[78] Because his opening words lay out out his vision so clearly, I will repeat them at length:

> One of my pupils, O. Anderson, in a brief exposition of his researches on the Variate Difference Correlation Method in *Biometrika* (1914), draws attention to the superiority of the method of mathematical expectation over the methods usually employed by English statisticians. The small popularity enjoyed by the method of mathematical expectation in England is not of course accidental.
>
> English scientific tradition rejects the concept of "mathematical probability."
>
> From the time of R. L. Ellis and of the first edition of John Stuart Mill's *System of Logic*, the logician's basis of probability has, in England, been the notion of empirical frequency. English mathematicians have followed the lead of the writers on logic in their preference for the idea of statistical frequency, and the method of mathematical expectation has naturally shared the fate of the concept of mathematical probability on which it rests.
>
> Notwithstanding its deep-rooted historical basis, English statisticians should break with tradition. The substitution of statistical frequency for mathematical probability does not obviate the logical difficulties in laying the foundations for a statistical study of Causation, but merely shifts them elsewhere. The gain from the point of view of philosophical representation is sufficiently doubtful, while from the purely mathematical point of view the rejection of the ideas of mathematical probability and mathematical expectation is accompanied by very substantial disadvantages. Verbal formulation becomes very complicated, leading to loss of economy of attention: it is continually necessary to speak of "the statistical frequencies which would become established if the number of occurrences were infinitely great." The absence of a sharp distinction in terminology between statistical frequency in the exact meaning of the term and those quasi-empirical "frequencies which would become established in an indefinitely great number of occurrences" often fails to make the very statement of the problem clear to the reader, and occasionally it would appear, to the author: when reading published papers one not infrequently feels that the author does not give himself a full account as what he is really calculating.

Pearson, of course, knew very well how he was using probability theory; Fisher later showed that he had erred in many calculations, but his method had been sound [354]. He saw mostly misunderstanding in Chuprov's misgivings, but he published Chuprov's and Anderson's articles. Chuprov's calculations, though laborious and soon superseded, did add to the understanding of correlation.

Here is how Chuprov defined mathematical expectation in the 1918 article:

[78] Discuss [172] and [3, 5].

If the variable magnitude X can take the values $\xi_1, \xi_2, \ldots, \xi_k$ with probabilities $p_1, p_2, \ldots, p_k$, I call the system of values $\xi_1, \xi_2, \ldots, \xi_k$ and the values $p_1, p_2, \ldots, p_k$ associated with them "*the law of distribution of the values of the variable X*." The law of distribution of values lies at the base of empirical "*frequency curves*," just as the mathematical probability of an event lies at the base of its statistically established frequency.

Denoting by the symbol EX the mathematical expectation of the variable magnitude X, we have as is well known:

$$EX = \sum_{i=1}^{k} p_i \xi_i,$$

where

$$\sum_{i=1}^{k} p_i = 1.$$

In substance, this hardly goes beyond Cournot's one-sentence definition of a gambler's expectation (page 107): "La valeur de son espérance est la somme de ces gains aléatoire, multipliés par les probabilités correspondantes." But it was new to the English language in its forthrightness (not beginning with a reference to gambling), its use of *variable* to name a probabilized quantity, and its use of E as an operator.

When Bortkiewicz had first used E in his letters to Chuprov in 1896, Chuprov had commented that he did not like expectations, as he thought this brought subjective connotations into probability theory, but he eventually made his peace with this recognition of the primal role of games of chance.

The operators E and м. о. appear in letters between Chuprov and Markov that were collected and published by Kh. O. Ondar in 1977 [287]. In a series of letters in 1910 (numbered 13, 15, 16, 17, 20, and 25 by Ondar) and in later letters in 1912 and 1916 (numbered 45 and 74), Markov uses his м. о.. Chuprov not use either symbol during most of this period, but uses E in a response to Markov in March 1916 (numbered 76). Markov then uses E in his reply and continues to use it most of the time in the remaining correspondence in 1916 and 1917. The English translation of the letters published by Charles and Margaret Stein in 1981 [288] obscures this by changing all the occurrences of м. о. to E.

6.2 Bortkiewicz and Chuprov's *zufällig*

Bortkiewicz apparently first used *zufällige Größe* (random quantity) in his his 1917 book *Die Iterationen*. He used it there systematically.

Aside from Bernoulli and Markov, the authors that Bortkiewicz cited most often in his 1917 book could be called statisticians: Bruns, Lexis, Pearson, and Chuprov. He cited Chebyshev but not Liapunov or Cantelli. His foreword is dated 18 October 1916.

Chuprov, who was steeped in the history of probability and statistics, as evidenced by his 1911 book on the philosophy of statistics [68], and a far better expositor than Bortkiewicz, enthusiastically adopted his friend's innovation, with a twist that reflected a respect for the English statisticians that Bortkiewicz did not fully share. Instead of *zufällige Größe*, Chuprov used *zufällige Variable.* From 1918 onward, he used this term consistently when he wrote in German. He used it in the *Skandinavisk Aktuarietidskrift* in 1918 [70], in *Metron* in 1921 [72], and in his 1925 book *Grundbegriffe und Grundprobleme der Korrelationstheorie* [74]. In the Russian version of the book, which appeared in 1926, he used the Russian equivalent, случайная переменная.

In English, Chuprov was uncertain about what name to use. As we have already seen, he used *variable-magnitude*, simplified to *variable*, in 1918 in *Biometrika* [69]. In *Biometrika* in 1921 [71], he used *variate.* In *Metron* in 1923 [73], he used *variable*, *correlated quantities*, and *observations* (but *zufällige Variable* in the summary in German). In the *Journal of the Royal Statistical Society* in 1925 [75], as we have already seen, he used *chance variable.* Much later, in 1939 [76], the translator of his 1925 book on correlation again used *chance variable.*

Chuprov died in Geneva in 1926, at the home of a friend [341]. He had left Saint Petersburg in 1917 to spend the summer doing research in Oslo and Stockholm, and he never returned to Russia. In the period we have been discussing, from 1918 to 1926, he often lived reclusively, mainly on his royalties, mostly in Germany. Having avoided any public criticism of the new authorities in Russia, he was able to publish there as well as in western Europe, and he continued to be influential, through correspondence and occasional lectures as well as through his publications. The term *zufällig Variable* did not become standard in the 1920s, but it did become familiar, and this was undoubtedly due to Chuprov.

On the substance, Chuprov's influence was also felt in English. We can see it in the small but influential book *Mathematical Statistics* published by Henry Lewis Rietz, professor at the University of Iowa, in 1927 [307]. Rietz distinguishes straightaway between the arithmetic mean and mathematical expectation, introducing mathematical expectation with these words:

> The *mathematical expectation* of the experimenter or the *expected value* of the variable is a concept that has been much used by various continental European writers on mathematical statistics. (page 16)

Towards the end of the book, in the course of acknowledging that he has not given a full account of regression and correlation because he has not brought in probabilities or idealized "actual distributions into theoretical distributions or laws of frequency", Rietz acknowledges Chuprov's *Grundbegriffe und Grundprobleme der Korrelationstheorie*, which had just appeared, with these words:

> In a recent book by the Russian mathematician A. A. Tschuprow, an important step has been taken toward connecting the regression method of dealing with correlation more closely with the theory of probability. (pages 102–103)

Chuprov's use of случайная переменная and *zufällige Variable* was enthusiastically followed by Oskar Anderson, whom we have already encountered, and the Russian economist and statistician Evgeniĭ Evgen′evich Slutskii (1880–1948). Both shared Chuprov's admiration for Cournot's philosophy of probability and the work of the British statisticians. The Norwegian mathematician Alf Guldberg (1866–1936), another of Churpov's many correspondents, also used *zufällige Variable* early on, in the Italian statistics journal *Metron* in 1923 [174].

In a 1925 article in *Metron*, Slutskii considered both variables that are *zufällige* and variables that are not. The probability distribution for a *zufällige Variable* might depend on an independent (*unabhängige*) variable that is *nicht-zufällige*. Anderson used *zufällige Variable* systematically in his 1935 German textbook on mathematical statistics [10].[79] Both Slutskii and Anderson acknowledged that Nekrasov had used the term earlier but insisted that Chuprov had made it a fundamental idea in theoretical statistics. In 1925 ([347], page 6), Slutskii wrote:

> Aber zum Grundpfeiler aller Begriffskonstruktionen der theoretischen Statistik wurde dieser Begriff erst von Prof. A. A. Tschuprow erhoben.

Anderson echoed this ([10], page 168): "als Grundpfeiler der statistischen Begriffskonstruktionen tritt er aber erst in den Arbeiten von Tschuprow und Bortkiewicz auf."

6.3 Stochasticity

We can also credit Chuprov, along with Slutskii, with the successful popularization of another terminological innovation in Bortkiewicz's 1917 book: the revival of Jakob Bernoulli's use of the Greek *stokhastikos*.

Bortkiewicz wrote ([38], page 3):

> Die an der Wahrscheinlichkeitstheorie orientierte, somit auf "das Gesetz der Grossen Zahlen" sich gründende Betrachtung empirischer Vielheiten möge als Stochastik ($\sigma\tau o\chi\alpha\zeta\varepsilon\sigma\vartheta\alpha\iota =$ zielen, mutmaßen)bezeichnet werden. Die Stochastik is nicht sowohl Wahrscheinlichkeitstheorie schlechthin, also vielmehr Wahrscheinlichkeitstheorie in ihrer Anwendung, sei es auf empirische Vielheiten überhaupt, sei es auf empirische Vielheiten einer bestimmten Art. ...

Chuprov added that the new stochastic science was more than statistics: it was about the stochastic connection between variables. As we shall see, *stochastic* took hold among mathematicians in the 1930s.

Chuprov opened an article in *Metron* in English in 1923 with these words ([73] page 461, typographical errors corrected):

[79] By this time, Anderson was working in Bulgaria. He had fled Russia in 1920. He ended his career as a professor in Munich, where he launched what is now the Department of Statistics of the Ludwig-Maximilians-University Munich.

> Every stochastical[80] theory of statistics sees in the empirical statistical numbers images of certain really significant quantities—reflected confused images blurred more or less by the Chance. Behind the statistical frequency of an event it discerns the corresponding mathematical probability or, as the English school does, the meta-empirical frequency which would become established in an indefinitely long run, if the observations could be carried out under unaltered conditions. Behind the average of the observed data it perceives the corresponding mathematical expectation; behind the "frequency-curve"—the law of the distribution of the values of the variable; behind the "correlation-table"—the law of interdependence of the variables.

7 *Variable aléatoire*

After the appearance of Lévy's book in 1925, the term *variable éventuelle* had some currency in French. It appeared in multiple notes in the *Comptes rendus*, for example. How did *variable aléatoire* displace it?

The key figures in the shift were two French mathematicians, Maurice Fréchet (1878–1973) and Georges Darmois (1888–1960), who both began teaching probability and statistics in the early 1920s, Fréchet at the University of Strasbourg, and Darmois at the University of Nancy. Fréchet was very well known for his work in functional analysis. Darmois, though he had done work in pure mathematics and on the theory of relativity, was less well known and always remained less published.

7.1 Fréchet and Halbwachs: *nombres aléatoires*

To my knowledge, the first time the adjective *aléatoire* was applied to an arbitrary probabilized quantity—and indeed, the first time is was used in any way in mathematical probability theory outside its traditional domain of gambling and finance) was in *Le calcul des probabilités à la portée de tous* [158], published by Maurice Fréchet and Maurice Halbwachs in 1924. This was an unusual book, based on an unusual pedagogical philosophy and an unusual sort of collaboration. Fréchet, as already noted, was a well known mathematician. Maurice Halbwachs (1877–1945) was a well known sociologist, who died in the concentration camp at Buchenwald.

As the authors explain in the preface to their book, the book was prepared primarily by Halbwachs from notes from Fréchet's 1921 lectures on probability at the University of Strasbourg. The two authors conceived of the project later, when they were both teaching at a business school in Strasbourg, Fréchet teaching insurance mathematics and Halbwachs teaching statistics. The goal of the book was to introduce probability to students whose mathematics was at the secondary-school level, teaching more by example than by definition and proof. The book is relatively traditional; the term *événement fortuit* is prominent in

[80]Here Chuprov inserted a footnote citing Bernoulli and Bortkiewicz.

its opening pages, and *espérance mathématique* is used only for gamblers' expectations. The word *aléatoire* first appears in its traditional role (*paiements aléatoires* on page 99), but after probabilized quantities are introduced by example, they are called *nombres aléatoires* (page 104);[81] the theoretical mean of a *nombre aléatoire* is called its *valeur moyenne*.

7.2 Darmois: *Statistique mathématique*

It appears that Darmois was the first author to use *variable aléatoire* in print.

In 1925, Darmois, while still a professor at Nancy, began, at Borel's request, to teach mathematical statistics at the Institut de Statistique at the University of Paris [61, 343]. His *Statistique mathématique* [106], based on his lectures there, appeared in 1928[82] and used *variable aléatoire* extensively. The book is impressive and does full justice to work in England, Scandinavia, and Russia. The authors cited most in its index are Pearson (23 times), Charlier (12 times), and Chuprov (9 times). The way in which *variable aléatoire* appears suggests that the author had been using *grandeur aléatoire* when he began writing the book, for this term appears first but then fades away in favor of *variable aléatoire*. So we cannot say quite when Darmois began using the term.

It is possible that Darmois's use of *aléatoire* was influenced by Fréchet and Halbwachs, but we do not have any direct evidence of this. He does not cite their book in his bibliography, but this tells us nothing, because it is too elementary to belong there.

Charles Jordan, at Budapest, in his *Statistique mathématique*, published in 1927, used "espérance mathématique" for mean value (p. 82), but does not use this consistently in book Also uses the term *moment* [198]. Cites recent progress by Edgeworth, Bruns, Kapteyn, Charlier, and especially Pearson. Uses *grandeur* but also *variable*.

7.3 Bologna 1928

The quick adoption of *variable aléatoire* following the publication of Darmois's textbook owes something to the intellectual ferment at and around the 6th International Congress of Mathematicians in Bologna in 1928. In his wide-ranging and insightful *Souvenirs de Bologne* [42], Bernard Bru has described how this Congress brought together most of Europe's leaders in probability and mathematical statistics—an unprecedented event. Those in attendance included:

- Georges Darmois, Maurice Fréchet, Jacques Hadamard, and Paul Lévy from Paris,
- Aleksandr Khinchin and Evgenii Slutskii from Moscow, Sergei Bernshtein from Kharkov, and Vsevolod Romanovskii from Tashkent,

[81] Occasionally, we see *variable*, but without the adjective, as on page 112. On one occasion (page 155), the term *valeurs aléatoires* appears.

[82] While giving its publication date as 1928, the books states that it was copyrighted in 1927.

- Franciso Cantelli from Naples and Bruno de Finetti and Corrado Gini from Rome,
- Antoni Łomnicki from Lwów and Jerzy Neyman from Warsaw,
- R. A. Fisher from England,
- Emil Gumbel from Heidelberg,
- Bohuslav Hostinský from Brno, and
- Georg Pólya from Geneva.

Only Harald Cramér at Stockholm and Richard von Mises at Berlin were conspicuous by their absence; Andrei Kolmogorov was not yet prominent.

The proceedings of the Congress were not published until 1932 and so do not tell us what words were actually used in 1928. But most of the conversations would have been in French and German, and the participants were surely familiar with Cantelli and Castelnuovo's *variabile casuale*, with the currency of *variable* in English and French, and with the more recent use of *zufällige Variable* by Chuprov, Anderson, and Slutskii. The word *variable* would have been unavoidable, and when you heard *zufällig* in German, you would probably think *aléatoire* in French, not *éventuel.* It is easy to imagine many of these mathematicians saying *variable aléatoire* as they switched from German to French, without even realizing that they were coining a new term.

This we know: It is difficult to find *variable aléatoire* before the meeting in Bologna in September 1928. From then on, we find it everywhere.

A second appearance of *variable aléatoire* in 1928 is in *Sur les transformations itérées des variables aléatoires* [187], written in French by the Czech mathematician Bohuslav Hostinský (1884–1951) and published in 1928 by the faculty of sciences of his university in Brno. Now far less remembered than many of the other attendees, Hostinký was at the forefront of research on a topic that then occupied the attention of many of them, including Fréchet, Hadamard, de Finetti, and Khinchin—the emerging theory of Markov chains. Hostinský used *variable aléatoire* to refer not to what we now call a random variable but rather to a random function of time. In November 1928, Slutskiĭ called this a *fonction éventuelle*. We now call it a stochastic process. Hostinský called it a variable because it varied with time. He also used *variable aléatoire* with this meaning in a note in the *Comptes rendus* for the week of 8 July 1929 [188] and in a book that appeared in Paris in 1931 [189]. Harald Cramér used *variable aléatoire* with Hostinký's meaning in 1935 [88]. But the meaning that became standard was Darmois's. The arbiter in this question of usage, as in many others, was Maurice Fréchet.

Before World War I, Maurice Fréchet (1878-1973) had been one of the creators of functional analysis. Throughout the period between the two wars, he was the leading international diplomat for French mathematics, constantly organizing visits, colloquia, and publications aimed at making France, rather than

Germany, the center of mathematics and French, rather than German, the international language of mathematics. He had been in close contact with Hostinský, for example, since 1920, when he wrote to Prague to encourage connections between French and Czech mathematicians [184]. At that time, he was one of the leaders charged with reconstituting the University of Strasbourg as a French institution. In the fall of 1928, Emile Borel brought him back to Paris to serve as his lieutenant for probability at the newly founded Institut Henri Poincaré and to be in charge of probability at the *Ecole Normale*, just as he had earlier recruited Darmois to take charge of mathematical statistics in Paris.[83]

Fréchet saw probability as an application of mathematics, not a field of mathematical research; in a lecture in 1925, he had even argued that probability should not be axiomatized ([157], pages 1–10). Now probability was flowering as mathematics, and he was in charge of training the young normalians who could keep Paris its forefront. This new teaching assignment, as he explained in November 1929, led him to transpose his previous work on the convergence of functions into the language of probability, thus generalizing the results on probabilistic convergence obtained by Cantelli, Slutskiĭ, and others.[84]

Fréchet's first report on this research was a short note in the *Comptes rendus* for 14 January 1929 [151]. Here is a passage near the beginning:

> ... si l'on se donne une fonction mesurable $f(x)$ sur le segment 0-1, et si l'on choisit au hasard un point x sur ce segment, on peut considérer $f(x)$ comme *un nombre aléatoire*. On peut alors interpréter les notions de convergence en mesure et de distance de deux fonctions *dans le langage de probabilités*, lorsque on se place dans le cas d'une distribution uniforme.
>
> On peut ensuite donner *un sens plus étendu* à cette interprétation, en considérant le cas plus général où x est le résultat fortuit d'une épreuve. On arrive ainsi, tout naturellement, à la notion de "convergence probable"...

Here Fréchet uses *nombre aléatoire*, which he had used extensively in his book with Halbwachs five years earlier, but on the second and last page, he uses *variable aléatoire* instead. Apparently he had decided to switch to *variable aléatoire* as he completed the note and forgot to make the change at the beginning, for he used only *variable aléatoire* in a sequel in the *Comptes rendus* two weeks later [152]. In the article in which he proved the results announced in the two *Comptes rendus* notes, which appeared in *Metron* in 1930 [153] (and where he

[83] Statistics being less prestigious than probability mathematics, Darmois remained professor at Nancy, moving full time to Paris only when he was appointed a professor at the University of Paris in 1933.

[84] "Au moment où notre enseignement nouveau nous amenait à concentrer nos pensées beaucoup plus qu'antérieurement sur le Calcul des Probabilités, nous avons été tout naturellement conduit á *transposer dans le langage des Probabilités* avec les modifications et les précautions convenables *un de nos précedents mémoires "Sur diverses modes de convergences d'une suite de fonctions d'une variable"*. ([153], page 4). Although published in 1930, the article where Fréchet makes this statement is dated November 1929. He also states in the article that he had presented the results at the end of his 1928–29 course at the Sorbonne.

made his remark about transposing his earlier work into the language of probability), he continued to be somewhat careless. He makes it clear that *variable aléatoire* is his formal name for a probabilized quantity, but he also uses *nombre alátoire* and *valeur aléatoire* as synonyms.

Fréchet's January 1929 notes and 1930 article were remarkable not only for the term *variable aléatoire* but also for their definition of the term: a *variable aléatoire* is a function, a function that maps each possible outcome of a random trial to a number. This makes the convergence that Cantelli studied the convergence of a sequence of functions. Cantelli had considered only convergence to a constant; Fréchet proposed to also consider convergence to another function—i.e., another *variable aléatoire*. From the viewpoint of statistics, it was no revelation that the variables in a particular study are functions on a common space; the space is the population from which we are sampling, and the variable varies from individual to individual in that population. But it was novel to take this aspect of the matter as a starting point for probability mathematics.[85]

Probably no explanation is needed for Fréchet's use of *aléatoire* instead of Lévy's *éventuel*; it was already the more common translation of the German *zufällig* and the English *random*, and as Fréchet's book with Halbwachs shows, he was already accustomed to using it. But *aléatoire* also has a relevant meaning that *éventuel*, *random*, and *zufällig* do not share. This is its more ancient meaning, still essentially the only meaning of *aleatory* in English: when we apply this English adjective to a method or process, we mean that its results will depend on uncertain events, not that it itself has been chosen haphazardly or at random. The law (Roman and now French and English) calls an insurance policy or an annuity an aleatory contract (*contrat aléatoire* in French). This connotation made it easy for Fréchet to apply the adjective to functions that are not themselves random.

It would be several more years before Andrei Kolmogorov made Fréchet's vision into an axiomatic foundation for probability and others began thinking of *variables aléatoires* as functions, but Fréchet was immediately followed in his use of the name *variable aléatoire* for a probabilized quantity by two other Russian mathematicians: Alexander Khinchin (1894–1959) and Vsevolod Ivanovich Romanovskiĭ (1879–1954). Khinchin used *variable aléatoire* in two notes [203, 204] that appeared in February in the same volume of the *Comptes rendus* as Fréchet's January notes. He had used *variable éventuelle* in a January 1928 *Comptes rendus* note in which he had coined the term *loi forte des grands nombres* [202]. Romanovskiĭ, a student of Markov's, used the term in 1929, in *Bulletin de l'Académie des Sciences de l'URSS* [311].[86] (Romanovskiĭ also used *variable aléatoire* in two contributions to the proceedings of the Bologna meeting, but these were not published until 1932.)

[85]For more information on the context, see §4.3 of [15].

[86]Soviet mathematicians were still publishing their best research in French, German, and English in these years. Leading Russian mathematics journals, including the *Bulletin de l'Académie des Sciences de l'URSS* and Математический Сборник, were multilingual. Alongside articles in Russian, they published articles in other languages, by both foreign and domestic authors.

Fréchet's example, and perhaps his editing, quickly made *variable aléatoire* popular in France. Indeed, a whole series of prominent scholars whom Fréchet invited to lecture at the Institut Henri Poincaré turned their lectures into articles for the institute's *Annales* that used *variable aléatoire*. These included:

- Georg Pólya (1887–1985), one of the first lecturers at the institute in 1929. His lecture appeared in Volume 1 of the *Annales* in 1930 [302].

- Richard von Mises (1883–1953), whose November 1931 lecture appeared in the third volume in 1932 [364]. Arguing for his theory of collectives, von Mises framed his case as an attack on *variables aléatoires* (pages 156–157):

  > Il faut mentionner d'abord les recherches abordées by M. Cantelli, continuées par M. Fréchet et quelques autres qui s'occupent d'une nouvelle notion de "variable aléatoire" et essayent de définir une certaine "convergence au sens du calcul des probabilités". Mais qu'est-ce que c'est qu'une variable aléatoire?

- Fréchet's colleague Georges Darmois (1888–1960), who was in charge of statistics in Paris. He lectured in 1929; his manuscript was received in 1931 and appeared in 1932 [108]. Darmois was also using *variable aléatoire* in *Metron* in 1929 [107].

- The Dane Johan Steffensen (1873–1873). Lecture 1931, manuscript received 1932 and published 1933 [351].

- The Norwegian Alf Guldberg (1866–1936). Lecture 1931, manuscript received 1932 and published 1933 [175].

- Guido Castelnuovo (1865–1952), whom we have already encountered. Lecture 1932, manuscript received 1932 and published 1933 [60].

The matter was settled when Paul Lévy followed Fréchet's lead. The last article in which Lévy used *variable éventuelle* appeared in 1931 [243]. In 1932 [244] and thereafter, he used *variable aléatoire*.

7.4 German and Russian

Aleksandr Yakovlevich Khinchin (1894–1959) followed Fréchet's lead not only in French but also in German, where he was the most prolific and influential mathematician of probability at the time. He was still using *Wahrscheinlichkeitsgröße* and *Größe* in 1929 [205, 206, 207, 208], but by 1933, he was using *zufällige Variable* [209, 210, 211, 212], as Slutskiĭ, following Chuprov, had been doing since 1925.

In Russian, Khinchin was less radical. He had already used случайная величина in his 1927 Russian book on probability [201]—a departure from Chebyshev and Markov's simple величина but not so radical as Chuprov's and Slutskiĭ's случайная переменная. He continued to use случайная величина throughout the remainder of his career. Kolmogorov and others followed his lead, and случайная величина was the standard Russian equivalent of *random variable* from the 1930s through the end of the twentieth century. Only in very recent years, as English has become the language of mathematical research even in Russia, has случайная переменная become current.

Khinchin's stopping short of случайная переменная could be interpreted simply as loyalty to Russian tradition, but by avoiding the Russian equivalent of *variable* he was surely also making clear that he was not a statistician. The statisticians, prone to reporting numbers that did not suit the authorities, were early targets of Stalin's terror; the Soviet statistics journal вестник статистики was closed in 1930, and the head of the statistical institute in Moscow where Slutskiĭ worked was arrested in 1931.

8 The language of the *Grundbegriffe*

Kolmogorov, Feller, Doob

8.1 Kolmogorov's *Grundbegriffe*

Andrei Nikolaevich Kolmogorov (1903–1987) was a leading figure among the new generation of Russian mathematicians who came of age in the Soviet Union. He began publishing mathematics in international journals in 1923, at the age of 20. His first article on probability, co-authored with Khinchin in 1925, concerned sums of independent probabilized quantities. In the late 1920s, he made important contributions to the law of large numbers and the law of the iterated logarithm. The 1925 article with Khinchin used the established German terminology that Khinchin was still following at that time: they discussed quantities (*Größen*) that were determined by chance (*durch den Zufall bestimmt*) [213]. Kolmogorov was still using this terminology in 1928 [215], but in 1929 he switched to *zufällige Größe* [216]. In the *Comptes rendus* in 1930 [217], he used *grandeur éventuelle*.

Kolmogorov's most important contributions to probability came after his visit to Germany and France in the summer of 1930, when he had an extended conversation with Fréchet. The most celebrated of these contributions was his 1933 monograph *Grundbegriffe der Wahrscheinlichskeitsrechnung* ([219], now considered the definitive formulation of the measure-theoretic foundation for probability theory. But this monograph was preceded by two more substantive and very influential articles on Markov processes, published in 1931 and 1933 [218, 220]. This articles did not yet use the 1933 framework. Instead, each time instant had its own probability space, and the probabilities were related across time by differential equations.

In the *Grundbegriffe* in 1933, Kolmogorov adopted Fréchet's view of probability as an application of functional analysis but deftly gave it boundaries as an independent mathematical field, fully axiomatized on its own in a way that Fréchet had been loath to do [366, 327, 328]. The axiomatization stepped back from functional analysis, focusing on the more basic assignment of probabilities to events—i.e., on measure theory. The starting point was a set E whose elements Kolmogorov called elementary events (*elementare Ereignisse*) and a class of subsets of E that he called random events (*zufällige Ereignisse*). Each random event A was assigned a number $\mathsf{P}(A)$ called its probability. Measure-theoretic axioms already familiar to functional analysts were given for the set of random events and their probabilities. A *zufällige Größe* (Fréchet's *variable aléatoire*) was a measurable real-valued function on E, and its *mathematische Erwartung* was its Lebesgue integral with respect to P. This set the stage for using functional analysis in the way Fréchet had been doing. But now the difficult question of whether assumptions being made were fulfilled in problems where probability might be applied was set aside. Probability itself was now a field of mathematics.

Khinchin and Kolmogorov were both still in tandem at the university in Moscow at this point. So it is notable that they were diverging on one point of German terminology. Khinchin was switching to Chuprov's and Slutskiĭ's *zufällige Variable*. In fact, he used *zufällige Variable* in *Asymptotische Gesetze der Wahrscheinlichkeitsrechnung* [209], which appeared in the same volume of monographs as the *Grundbegriffe*. But Kolmogorov continued to avoid *Variable* in German, just as he and Khinchin both continued to avoid its Russian equivalent, переменная. Their divergence on this point of German terminology became moot by 1938, when Soviet mathematicians no longer dared publish in languages other than Russian [318].

The ideas in the *Grundbegriffe*'s axiomatization of probability were not news to Fréchet and Lévy, and letters from Lévy to Fréchet ([15], letters 26 and 27) suggest that they did not pay much attention to it until 1936. But the axiomatization proved attractive and useful for mathematicians who were more familiar with functional analysis than with the language of probability. These newcomers to probability could now put aside the philosophical and practical issues that had traditionally barred the way and plunge into probability as a purely mathematical enterprise. Two powerful mathematicians who quickly took advantage of this opportunity were William Feller (1906–1970), then at Stockholm, and Joseph L. Doob (1910–2004), then at Columbia in New York.

8.2 Feller vacillates

Born into a German-speaking family in Zagreb, Feller had gone to Göttingen to study in 1925 and then to Kiel to teach in 1928 [321]. He fled when Hitler came to power in 1933, spending a year in Copenhagen and then relocating to Stockholm in 1935, where he worked with Marcel Riesz and Harold Cramer. His review in the *Zentralblatt für Mathematik und ihre Grenzegebiete* [137], written while he was still at Kiel, was probably the most enthusiastic and influential early published comment on the *Grundbegriffe* [327]:

> Die Wahrscheinlichkeitsrechnung wird in größter Allgemeinheit lückenlos axiomatisch aufgebaut und erstmalig ganz systematisch und sehr naturgemß̈ in die abstrakte Maßtheorie eingeordnet. Das

> Axiomsystem ist wohl das denkbar einfachste. ... Bemerkenswert ist hier die große Allgemeinheit: es werden Wahrscheinlichkeiten auch in unendlich dimensionalen Räumen beliebiger Mächtigkeit behandelt. ... Die Darstellung ist sehr präzise, aber etwas knapp, und wendet sich an Leser, denen die Materie nicht fremd ist. Die Maßtheorie wird vorausgesetzt.[87]

After arriving in Stockholm, Feller began to work on probability himself. Still writing in German, he published on the central limit theorem in 1935 [138] and on stochastic processes in 1936 [139]. He cast both articles as pure mathematics, while noting how they could be translated into probability language. In the 1936 article, Feller again praised Kolmogorov's axiomatization, noting that it allowed a mathematician to use probability language while still doing pure mathematics, independently of any applied problem. But he did not use Kolmogorov's term *zufällige Größe*. In 1935, he used *statistische Variable*. In 1936, he used *stochastische Veränderliche*, adding this footnote:

> D. i. "zufällige Größe" in der Terminologie bei Kolmogoroff (variable aléatoire), also einfach eine auf der Grundmenge meßbare reele Funktion.

Perhaps we can attribute Feller's hesitance to use Kolmogorov's *zufällig* partly to the fact that the object was no longer an random object; it was a determinate real-valued function. Like others, he would eventually overcome any such scruples.

The German terminology remained unsettled. There was no Fréchet in Germany to arbitrate. German mathematics never had a center playing the role of Paris in France, and nearly all of the important German-language work in probability in the first half of the twentieth century had been done by foreigners or by Germans who were soon fleeing the Nazis. In his 1956 *Wahrscheinlichkeitstheorie* [306], the first textbook on measure-theoretic probability after World War II, Hans Richter introduced the concept of a random variable with these words:

> ... Man nennt daher $a(x)$ eine *zufällige Größe*, eine zufällige Variable, Zufallsvariable, *stochastische* Variable oder auch *aleatorische* Größe.

In body of the book, Richter used *zufällige Variable* more than any of the other choices. But now that German is no longer an international language, serving instead as a language in which Germans address Germans, its preference for compound nouns has reasserted itself, and *Zufallsvariable* has become more common.

---

[87] The calculus of probabilities is constructed axiomatically, with no gaps and in the greatest generality, and for the first time systematically integrated, fully and naturally, with abstract measure theory. The axiom system is certainly the simplest imaginable. ... The great generality is noteworthy; probabilities in infinite dimensional spaces of arbitrary cardinality are dealt with. ... The presentation is very precise, but rather terse, directed to the reader who is not unfamiliar with the material. Measure theory is assumed.

8.3 Doob: *Chance variable*

The American Joseph L. Doob (1910–2004) was even quicker than Feller to put the *Grundbegriffe* framework to work. After completing a doctorate in complex variables at Harvard in 1932, Doob spent two years as a postdoctoral student in mathematics at Columbia. As he recalled in an interview with J. Laurie Snell in 1997, this did not go well, and so he talked with Harold Hotelling about working in statistics, where he thought he might have a better shot at an academic position. Hotelling obtained a Carnegie fellowship that would allow Doob to stay at Columbia for another year. Looking for how he could put his skills as a mathematician to work in statistics, Doob quickly focused on Khinchin's treatment of stochastic processes within Kolmogorov's framework. His first publication in probability was a short but ambitious (and flawed) note on stochastic processes that he published in the *Proceedings of the National Academy of Sciences* in June 1934 as a National Research Fellow [123]. Its first sentence reads:

> A stochastic process is defined by Khinchin to be a one parameter set of chance variables: $\mathbf{x}(t), -\infty < t < \infty$.

A footnote directs the reader to Khinchin's 1934 article, which had introduced the term *stochasticher Prozeß* and used *zufällige Variable*. Why translate Khinchin's *zufällige Variable* as *chance variable*? Perhaps the concatenation of nouns seemed natural to Doob's American ear, just as *Zufallsvariable* sounds natural to the German ear. Perhaps he had seen Dodd and Baten's articles in the *Annals of Mathematical Statistics*. But given that Doob persisted with *chance variable* for fifteen years, we might speculate that he, like Feller, was uncomfortable calling a determinate function random. As an adjective, *chance* can function as a synonym for *stochastic* or *aleatory*. It says that the noun it is modifying has something to do with probability theory, but not that it is random.

8.4 Stochastik

Interaction between Kolmogorov and Khinchin was responsible during this period for another terminological innovation: the term *stochastic process*. In his two articles on continuous processes (1931, [218], and 1932, [220]), Kolmogorov used the name *stochastischer-definiter Prozeß* (stochastically definite process) for a Markov process. This adverbial use is consistent with the higher-level meaning that Chuprov and Slutskiĭ had given to the adjective *stochastic*. Calling a process stochastically definite did not mean that it is was a random object. It meant that when you know the current value of the process, you can use probability theory to make predictions about future values. But in 1934 [212], when Khinchin put the notion of a random process into the framework of Kolmogorov's *Grundbegriffe*, he called any random process a *stochastischer Prozeß*. Perhaps Khinchin was simply using *stochastische* as a synonym for *zufällig*, contrary to its then established meaning. Or perhaps he felt that in

Kolmogorov's framework the process was not a random object. It was now a set of functions indexed by time. For historical reasons, he was calling these functions *zufällige Variablen*, but they were not really random; they were all functions on a single probability space, and the randomness lie entirely in the choice of an *elementar Ereignisse* from that space.

9 *Random variable*

Once *variable aléatoire* was standard and *zufällige Variable* was current, it was natural for those familiar with these terms to translate them into English as *random variable*, but it took time for this term to become dominant. The process can be seen as one in which the habits of the British statisticians were overwhelmed by the inclinations of a powerful cohort of mathematicians whose native language was not English: some of them migrating to the United States, others remaining in Europe but shifting to English as the new international language of mathematics. Harald Cramér (1893–1985) and Jerzy Neyman (1894–1981) were leading and archetypal figures in this process.

Pearson uses "random error" on page 161 of [293].

9.1 Precursors

As we have already seen, Student used *random variable* in passing in *Biometrika* in 1914. Most of the other early occurrences of the term also came from statisticians. Here are all the occurrences I have seen up to 1935:

- In 1926 [309], Romanovskiĭ used the term in the *Bulletin de l'Académie des Sciences de l'URSS*, writing:

  > Denoting with $E(t)$ the mathematical expectation of any random variable t....

  In 1929, the same year Romanovskiĭ first used *variable aléatoire*, he published an article entitled "On the moments of means of functions of one or more random variables" [310] in *Metron*. This appears to have been the first article with *random variable* in its title.

- The American economist and statistician Harold Hotelling, then at Stanford, used *random variable* in passing in 1927 [190].

- In 1928, C. F. Trustam used *random variable* in a British journal for actuaries [360]. He explained that

  > ... Prof. Cantelli investigates fundamental conceptions introduced by Tchebycheff and now forming an integral part of the classical structure of the Calculus of Probabilities. The abstract notion of a "casual" or "random" variable plays an important part in the paper ...

After putting "random variable" in quotation marks one more time, they use it freely.

- In 1930, Cramér used *random variable* in a short book on the mathematical theory of risk, apparently sponsored by the insurance company where he worked as an actuary at the time [87].
- In 1931, the prominent Norwegian economist Ragnar Frisch (1895–1973), an associate of Slutskiĭ [16, 28], used it in the *Journal of the American Statistical Association* [159].
- In 1934, the Hungarian-American mathematician Aurel Wintner (1903–1958), used it in the *American Journal of Mathematics.* Trained in Leipzig, Wintner had immigrated to the United States, to a position at Johns Hopkins, in 1930 [385]. Wintner is the only individual on this list whom we could not call a statistician.
- In 1935, Neyman and two Polish colleagues used it in an article on agricultural experimentation in the *Journal of the Royal Statistical Society* [283].

Cramér and Neyman do not come at the head of this list, but they were the first to use *random variable* in a sustained way.

9.2 Harald Cramér

Harald Cramér (1893-1985) completed his Ph.D. in mathematics and began teaching at the university in Stockholm in 1917. In 1920, he also took a position as an actuary in an insurance company. More than any of the other individuals in our story, he embodied both the ambitions that drive our story—Kolmogorov's ambition to find a sound mathematical foundation for probability and the Chuprov's ambition to use the accomplishments of probability better in statistics. In 1976 [91], he recalled a passage from an article that he had published in 1926:

> The probability concept should be introduced by a purely mathematical definition, from which its fundamental properties and the classical theorems are deduced by purely mathematical operations. ... Against such a mathematical theory, no objection can be valid except on mathematical grounds. On the other hand, it should be emphasized that the mathematical theory does not prove anything about the real events that will occur. (page 517)

Guttorp and Lindgren [176] quote the following from an unpublished memoir in Swedish dated 1978:

> During the 20s and 30s so many new findings regarding statistical methodology had been published,particularly in England, where

> Karl and Egon Pearson (father and son), R. A. Fisher and Jerzy Neyman had an intensive production of novelties. I realized their great importance for applications, but felt very critical of their mathematics. Both Fisher and the two Pearsons seemed completely alien to the new probability theory which was founded upon the work of Russian and French mathematicians. I was tempted to try to produce a synthesis of the two lines of development. (page 69)

In another telling recollection, in an article Cramér published in 1981 [92], concerns an incident at the 1937 colloquium in Geneva:

> Feller and I attended from Stockholm, and for me it was particularly interesting to meet Neyman, who gave an account of his new method of estimation by confidence sets. In the middle of his talk he was interrupted by Fréchet and Lévy, who wanted to criticize. I happened to be chairman of that meeting, and had to use my poor ability of talking French to quite them down and let him finish his talk. Having previously read his main paper in the *Proceedings of the Royal Society*, I was convinced that his ideas were sound, and I believe that his French opponents afterwards came to the same conclusion. (page 314)

Cramér's early mathematical work was mostly in French and German. His first publication on mathematical statistics, in 1923 [83], was in German, but thereafter he published most of his work in English.[88]

In his 1923 article in German, Cramér used a formulation similar to ones sometimes used by Czuber:

> Es seien $x_1, x_2, \ldots, x_n$ Grössen, die vom Zufall abhängen...

In English he always used *variable*, but he hesitated about how to distinguish probabilized variables from others. In 1924 and 1925, he used *statistical variable* [84, 85]. In 1928, he used *variable in the sense of the Theory of Probability* [86]. In 1930, as I already noted, he settled on *random variable*, with these words:

> According to the ordinary method, we may begin by considering the gain or loss arising during a certain period on one insurance policy. This is a quantity capable of assuming certain values with certain probabilities, which may be calculated. In the mathematical theory of Probability, such a quantity is generally denoted as a *variable*, or a *random variable*.

[88] This shift to English was not unusual in Scandinavia at the time, a point that can be documented by looking at the languages used in *Skandinavisk Aktuarietidskrift*, a journal for which Cramér became the Swedish editor in April 1921. This journal was launched in 1918 by the Swedish, Danish, and Norweigen actuarial societies and served as the main Scandanavian outlet for work in mathematical statistics until 1974, when the *Scandinavian Journal of Statistics* began publication. In its first decade, 1918–1927, *Skandinavisk Aktuarietidskrift* published 111 articles: 49 in German, 42 in English, and 20 in French. In its second decade, 1928–1937, it published 104 articles: 37 in German, 60 in English, and 7 in French.

He used *random variable* throughout the rest of his career, though he continued to drop the adjective once his meaning was clear.

9.3 Jerzy Neyman

The individual who most embodied Chuprov's vision of basing statistics on mathematical probability, Jerzy Neyman (1894–1981) studied with Sergei Bernstein in Kharkov during World War I and completed a doctoral degree in Warsaw in 1924. He began to work with Karl Pearson's son Egon during a fellowship year in London and Paris in 1926–1927 and continued to do so while back in Poland and after his return to London in 1934, when Egon had become Head of the Department of Applied Statistics at University College. In his first *Biometrika* articles in 1925 and 1926, Neyman was citing Chuprov while elegantly deploying the British terms *character* and *variate*: when X is a character, its values $x_1, \ldots, x_n$ are variates [270, 271]. In French he used *variable*, of course; I have already mentioned the 1929 note in which he used *variable accidentelle* [272]. In 1935, he used *variable aléatoire* in French [274], *random discontinuous variate* in the *Annals of Mathematical Statistics* [273], and, as I have already noted, *random variable* in the *Journal of the Royal Statistical Society*.[89]

By 1936, the year Karl Pearson died, Neyman and the younger Pearson had settled on *random variable*. In an article they published that year [284] in their department's newly launched *Statistical Research Memoirs*,[90] they wrote:

> If the variables $x_1, \ldots x_n$ have the property that, whatever the region w in the sample space, there exists a number, $P\{E \in w\}$, representing the probability that the sample point E will fall within the region w, then we shall describe these x's as random variables.

In 1937, Neyman was still defining *random variable* in a way that marked the term as newly adopted [275, 276]. But in that same year Cambridge University Press published Cramér's *Random Variables and Probability Distributions*, the first book with *random variable* in its title. Neyman, Pearson, and their disciples would now use the term without explanation and no concern that it had only recently become current.

The year 1937 was a watershed both for the adoption of *random variable* in English and also for the acceptance of Kolmogorov's axiomatization, but the two were loosely coupled. In his 1937 book, Cramér declared Kolmogorov's axioms to be his starting point but did not actually treat random variables as functions on an underlying probability space; instead he represented independent random variables by constructing the appropriate measures on n-dimensional Euclidean space. Neyman, in his 1937 article in the *Philosophical Transactions of the Royal Society of London*, declined to choose among the different foundations of probability advanced by Borel, Levy, Kolmogorov, and Fréchet. As the title of

[89] Neyman's early statistical papers were collected by the University of California Press in 1967 [282, 285].

[90] Two volumes appeared, one in 1936 and one in 1938. The journal ceased publication when Neyman left University College for Berkeley ([240], page 41).

his article indicated, he believed that he was simply using the "classical theory of probability". But he did represent random variables as real-valued functions on an underlying probability space. In his contribution to the colloquium on probability held in Geneva in October 1937 [277], he noted that the classical theory had been modernized:

> La signification de la forme nouvelle du problème d'estimation sera plus claire si on s'appuie rigidement sur le point de vue de la théorie classique modernisée de probabilités.

The 1937 Geneva colloquium was decisive for the recognition of Kolmogorov's *Grundbegriffe*. The most prominent theme of the colloquium was the opposition between the axiomatic approach, which was now represented by the *Grundbegriffe*, and von Mises's collectives. Feller was once again Kolmogorov's most enthusiastic champion [140], but his view was now close to a consensus. As Fréchet explained in his keynote address [156], Abraham Wald had resolved the apparent contradictions in von Mises's approach [369], but mathematical probability was best served by Kolmogorov's approach.

9.4 Alternatives

The most serious rival to *random variable* after 1937 was *variate*, which remained anchored in the language of many British and American statisticians. In the 1936 and 1937 articles just discussed, Neyman tried to give *variate* what he saw as its traditional and proper place: a *variate* is the value of a *character* for an individual in a population. Only when we attach probabilities to a variate's possible values does it become a *random variable*. But for many of the British and American disciples of Karl Pearson and R. A. Fisher, this was a needless proliferation of terminology. For Pearson, whether a frequency distribution was available for a character was a practical matter, and there was no need to talk about assigning probabilities to characters or variables or variates. Fisher and others of his generation took probabilities more seriously, but many saw no reason to invent another word when *variate* was handy. The American economist and statistician Harold Hotelling (1895–1973), well known for his enthusiasm for the work of R. A. Fisher [240], is one example. As we have just seen, he used the term *random variable* in passing in 1927, but from 1936 [192] to 1953 [191], he consistently used *variate* where we would now use *random variable*. The persistence of *variate*, especially in Britain, is evident from its presence in *A Dictionary of Statistical Terms*, published by Maurice G. Kendall and William R. Buckland in 1957. The precariousness of its role and the confusion about its meaning are also evident, however, in the explanation in their preface ([199], pages viii–ix):

> We use the word "variate" to denote a random variable and reserve the word "variable" for a mathematical variable or a varying quantity where the nature of the variation is unspecified. It is possible to distinguish between the "random variable", which is the totality of

possible values, and the "variate", which is the value it assumes in particular instances; but we have not attempted to preserve this distinction.

Aside from *variable* and *variate*, other rivals to *random variable* in the 1930s were *chance variable*, *statistical variable*, *stochastic variable*, and *random variate*. The table on page 91 shows the popularity of these terms in some prominent journals in the 1930s and the three following decades. Each of these terms merits further comment.

Chance variable As we have already seen, *chance variable* was used by Chuprov in 1925 and had a toehold among American mathematical statisticians in the early 1930s. Its persistence into the 1940s, as we shall see, was due in large part to J. L. Doob.

Statistical variable This term also dates back to the 1920s; we have seen Cramér using it in 1924 and 1925. Its German counterparts, *statistische Variable* and *statistische Veränderliche*, were already being used by K. G. Hagström in *Skandinavisk Aktuarietidskrift* in 1919 [177]. The earliest citations in JSTOR are from G. Udny Yule in 1926 [394] and Oskar Anderson in 1927 [9].

The meaning of *statistical variable* can vary. Alf Guldberg, in his lecture at the Institute Henri Poincaré [175], distinguished between a *variable aléatoire*, which has probabilities, a *variable statistique*, one for which we are tabulating frequencies. *Statistical variable* is now used more often in various areas of applied statistics than in mathematical statistics.

Stochastic variable Like *statistical variable*, *stochastic variable* moves easily across languages. As we have seen, Feller used *stochastische Veränderliche* in 1936 [139]. The Russian-American mathematician J. V. Uspensky (1883–1947) may have been the first to use *stochastic variable* in English, in the textbook he published in 1937 and his earlier lectures at Stanford, where he had taught since 1929. He had studied and worked in Saint Petersburg, becoming a member of the Russian Academy of Sciences in 1921 before immigrating to the United States. Although his textbook was well received, his research was in number theory and other areas of mathematics, not in probability and mathematical statistics. Had the Neyman-Pearson theory been the Uspensky-Pearson theory, random variables might be stochastic variables.

The earliest occurrence of *stochastic variable* in JSTOR comes in 1939, in the first joint article by Abraham Wald and Jacob Wolfowitz. They continued using it in their joint articles for several years, and Wolfowitz used it a sole-authored article in 1942. Both eventually switched to *chance variable*.

Random variate *Random variate* is still in use in mathematical statistics. But whereas it was a synonym for *random variable* or *variate* for most of those using it in the 1940s and 1950s, it is now usually a synonym for *pseudo-random number*.

In 1938, as Neyman was leaving Britain for a position at the University of California at Berkeley, *random variable* was rapidly becoming current in the United States as well as Britain. Its first appearance in the *Annals of Mathematical Statistics* was in an article by Neyman that appeared in June 1938 [278]. Within a year, it was already more popular than *chance variable*. All 12 occurrences of *random variable* in *Annals of Mathematical Statistics* in the 1930s came in 1938 and 1939, whereas the 4 occurrences of *chance variable* had all been in 1935 or earlier.

When an English translation of Chuprov's 1925 book on correlation theory finally appeared in 1939 and used *chance variable* to translate Chuprov's *zufällige Variable* and случайная переменная, the English statistician Florence Nightingale David, who had completed her doctorate with Neyman and Pearson in 1938, considered it

> ... a little astonishing to find that the translator makes no use of the term "random variable", which has long passed into common use.

The review in which she expressed this sentiment [112] appeared in the March 1940 issue of *Biometrika* and was only about the fourth article in that journal in which *random variable* had appeared. She had not expressed any similar astonishment two years earlier in her *Biometrika* review of Uspensky's book [111]; instead she had repeated his term *stochastic variable* without comment.

Although Hotelling, long a champion of R. A. Fisher, never rallied to Neyman's terminology, many did. One important convert was Samuel S. Wilks (1906–1964), in charge of statistics at Princeton. A Texan who had completed his Ph.D. at Iowa in 1931, Wilks had spent fellowship years at Columbia in 1931–1932, where he worked with Hotelling, and in London and Cambridge in 1932–1933, where he collaborated with the younger Pearson. In the March 1938 issue of the *Annals of Mathematical Statistics*, Wilks was still using *variate* [383]; in the September 1938 issue, he was using *random variable* [384]. For Wilks, as for many others who were whole-heartedly developing the Neyman-Pearson theory, it was natural to use the Neyman-Pearson terminology.

9.5 Abraham Wald

Abraham Wald (1902–1950) was less constant. Wald's reputation in Vienna in the 1930s was primarily in economics, but he was also accomplished as a statistician, having published a book on time series and seasonal adjustment in 1936 [367] in which he used *zufällige Variable*,[91] and as a probabilist, having largely

[91] On page 36, discussing Oscar Anderson's contribution to the variance difference method, Wald writes: Er geht von der Grundannahme aus, daß jedes Glied der Zeitreihe eine zufällige Variable im Sinne der Wahrscheinlichkeitstheorie ist, d. h. daß es verschiedene Werte mit verschiedenen Wahrscheinlichkeiten annehmen kann, ferner, daß die mathematischen Erwartungen aller Glieder der Reihe endliche Größen sind.

settled the controversy over von Mises's concept of a collective [368]. After fleeing from Vienna to the United States in the summer 1938, he quickly began working with Hotelling and with Jacob Wolfowitz (1910–1981) at Columbia. He was the most influential mathematical statistician of the 1940s, producing over 70 articles, two books, and a torrent of ideas until his tragic death in an airplane accident in December 1950. In his very first publication in English, an article in the *Annals of Mathematical Statistics* in December 1938 [369], Wald used *random variable*, but in the 41 articles that appeared in that journal under his name from then to March 1951 (listed in [372]), he covered the territory, using *variate* in 15 of them, *random variable* in 13, *chance variable* in 12, and *stochastic variable* in 3. In only two of the articles did none of these terms appear. These numbers add to more than 41, because in some articles he used more than one of the terms, with no difference of meaning. He was working too fast to pay too much attention to this terminological nicety.

Wald was relying on Hotelling to secure his appointments and promotions in the economics department at Columbia, and so it is not surprising that *variate* was his dominant choice until 1944, when he was promoted to full professor [388]. But he also quickly developed a very productive collaboration with the younger Jacob Wolfowitz, who was teaching high school and working on his Ph.D. at New York University when Wald arrived in New York. Wolfowitz had been born in Poland and had emigrated to the United States in 1920, at the age of 10. Their first joint article, in 1939, was Wolfowitz's first published article, and as we have already noted, it used *stochastic variable*. This term persisted in their joint articles in 1940 and 1941. Wolfowitz again used *stochastic* in a sole-authored article in September 1942 [386], but beginning in 1943, probably because of the example of Joseph Doob, he switched to *chance variable* [387]. In the mid-1940s, Wald deferred to Wolfowitz's preference for *chance variable* when writing with Wolfowitz but increasingly used *random variable* in his other articles.

In 1946, when Hotelling left Columbia to build a statistics department at the University of North Carolina, Columbia created its own statistics department in order to retain Wald [396]. The new department hosted two very accomplished visitors in its first year: Neyman in Fall 1946, and Doob in Spring 1947. Neyman, then in his 50s, was increasingly occupied with applications rather than theory; Doob, in his 30s, was in his most intellectually productive period. It seems likely that Wald found more intellectual promise in Doob's current work than in Neyman's. In any case, by 1948, Wald had shifted from *random variable* to *chance variable*. He used *random variable* in his 1947 book on sequential analysis [370], but he used *chance variable* in his 1950 book on statistical decision functions [371]. Wolfowitz continued to use *chance variable* until the end of his career in the 1970s.

9.6 Endgame

Doob, Wolfowitz, and Wald notwithstanding, *random variable* outpolled *chance variable* in the statistics journals in the 1940s by about three to one; see the table on page 91. We see the same ratio in the celebrated Berkeley Symposium on Mathematical Statistics and Probability that Neyman organized in 1945 and 1946. In its proceedings [280], which appeared in 1949, we find six authors using *random variable*: Evelyn Fix, Neyman himself, P. L. Hsu, Feller, Edward Barankin, and Erich Lehmann. Two authors use *chance variable*: Wolfowitz and Doob. Three others use *variate* to mean random variable: Arthur Copeland, Hotelling, and Carl Kossack.

Feller had published his first article in English in 1938, a discussion of the Neyman-Pearson concept of similar regions in Neyman and Pearson's *Statistical Research Memoirs*; there he used *variate* and *random variable*. In 1939, Feller immigrated to the United States to join the faculty at Brown University and launch *Mathematical Reviews*. In 1940, in his first substantial article on probability in English [141], he used *chance variable*, but by 1943, he was consistently using *random variable* [143, 142, 144].

Doob was a late but influential adopter of *random variable*. As we have seen, he used *chance variable* beginning with his first publication on probability in 1934 [123], in which he presented it as a translation of Khinchin's *zufäillge Variable*. He was still using *chance variable* at the symposium in Lyon in the summer of 1948 [125, 253]. But by the time of the second Berkeley Symposium in the summer of 1950, or at least when its proceedings appeared in 1951 [126], he had switched to *random variable*. In a conversation with J. Laurie Snell in 1997, he explained the switch this way:

> While writing my book I had an argument with Feller. He asserted that everyone said "random variable" and I asserted that everyone said "chance variable." We obviously had to use the same name in our books, so we decided the issue by a stochastic procedure. That is, we tossed for it and he won.

Feller's book appeared in 1950, Doob's in 1953 [145, 127].[92] The coin toss that Doob remembered nearly 50 years later may have happened, but it seems doubtful that its outcome could have moved Feller back to *chance variable*. As editor of *Mathematical Reviews*, Feller knew the score, and after Wald's death Doob was outnumbered, both literally and in intellectual heft, by the flourishing Neyman-Pearson tribe. Of the 14 articles using *chance variable* that appeared in the *Transactions of the American Mathematical Society* in the 1930s and 1940s, 9 were by Doob himself.

10 Synthesis

The consensus in favor of the name *random variable* represented a marriage between mathematical statisticians and mathematicians studying probability abstractly. The two groups were not using the term in the same way. For the

[92] Feller's book became Volume I of a two-volume set when Volume II appeared in 1966 [146]. In the preface to Volume II, Feller states that Volume I was written between 1941 and 1948. In his interview with Snell, Doob states that he began writing his book in 1945.

mathematical statisticians, a *random variable* was a quantity in the world to which probabilities are ascribed. For the mathematicians following Kolmogorov and Doob, a *random variable* was a well-defined mathematical object: a real-valued function on a probability space. But a shared vocabulary—*probability*, *event*, *random variable*, *independence*, etc.—helped the two groups benefit from each other's accomplishments. The mathematicians could draw on intuition and new mathematical questions emerging from applied problems, while the statisticians could draw on the mathematicians' results when they were useful. The mathematicians might gain enhanced recognition and even funding from clearer recognition of the relevance of their work to applied problems; the statisticians gained some protection against suspicions that their work was not mathematically sound.

Kolmogorov's basic framework is very simple: a set Ω and a measure P on Ω that gives it total measure one. The synthesis allowed mathematicians to use the language of probability and randomness in talking about these objects. As Feller observed, this does not burden the mathematics with problems involved in applications.[93] We are allowed to call Ω the *sample space*, to call subsets of Ω *events*, and to call real-valued functions on Ω *random variables*.

The statisticians continued to explain the meaning of *random variable* in a different way. Here is a typical explanation, from the opening passage of Wald's *Sequential Analysis*, published in 1947:

> The outcome of an experiment or the reading of a measurement is usually a variable quantity or, more briefly, a variable, since generally it can take different values. For example, repeated measurements on the length of a bar will yield, in general, different values. Frequently, it will be possible to make probability statements concerning the outcome of an experiment or the reading of a measurement....
>
> A variable x is called a random variable if for any given value c a definite probability can be ascribed to the event that x will take a value less than c.... ([370], p. 5)

When studying multiple variables, statisticians also required joint probabilities. As Cramér and Neyman explained repeatedly, a probability must be defined for every (measurable) set of possible values for the vector of variables $(x_1, x_2, \ldots, x_n)$ ([89], page 9; [275], page 342; [279], page 86). To the mathematician's ear, this is not a definition. A joint probability distribution is a well-defined mathematical object, but this does not make the variables mathematical objects. So be it. For statisticians, the variables have a reality outside mathematics, and most of the work takes place outside Kolmogorov's framework. This was true even for Feller's two-volume treatise, which dealt extensively with applied problems. Only one chapter in the two volumes used Kolmogorov's framework.

---

[93] [139], page 115: ...die beiläufig erwähnten wahrscheinlichkeitstheoretischen Begriffe unabhängig von jedem Anwendungsproblem mengentheoretisch exakt gefaßt wurden, insbesondere in der durchsichtigen Axiomatik von Kolmogoroff...

Having the same vocabulary used in two different ways, in two overlapping but different discourses, does engender tensions and contradictions. One concerns the idea of sampling. In Kolmogorov's framework, repeated sampling has disappeared, but the language of sampling is sometimes still used. We are allowed to say that P describes the randomness involved in sampling an element ω from Ω. This single sampling determines all the outcomes that we call random: an event E happens if ω is in E, and the value of a random variable X is $X(\omega)$. In the course of making the picture work for continuous stochastic processes, Doob brought a sense of time back into the picture, along with some complexity, by means of what we now call a filtration, which tells us which events are settled at each time t. But everything is still determined by the single ω.

The concentration of the randomness in the single random ω painfully contradicted intuitions shared by mathematicians and statisticians. As Paul Lévy complained in 1954 ([245], second edition, note II; [15], page 22), we want to think of chance as intervening at every instant, yet ω encodes all the successive influences of chance, expressing the perpetual future as if it were born in one instant. This also destabilized the near-identification of probability with frequency that drove the development of statistics, on the Continent as well as in Britain, in the nineteenth and twentieth centuries. If the choice of ω is not repeated in our world, then what do the probabilities for its values mean? Must we imagine, as some physicists and philosophers have tried to do, multiple invisible worlds that have all the different possible values for ω?

Another contradiction, perhaps more important for statistics and other applications of probability, arose from probability itself not being contingent within Kolmogorov's and Doob's picture. Statisticians took it for granted—and still usually take it for granted—that only some aspects of our experience can be probabilized. In the nineteenth-century picture, errors of measurement might be considered random, but perhaps not the equally unknown quantity being measured. The independent variables in a statistical regression study are sometimes random variables, sometimes not. As Slutskiĭ said, they may be *nichtzufällige*. Perhaps they arise from some process that is indeterminate but not random. Perhaps we choose their values as we go along, making them *variables* rather than *variants* as William Stanley Jevons would say [197]. The picture of a closed world with everything determined by ω has no place for these *nichtzufällige Variablen* unless they are fixed in advance and thus really *zufällige* in their own way, having the values they have with probability one.

Neyman was always careful to limit the scope of probability theory. In the 1937 article where he emphasized the classical theory of probability [275], he wrote

> I want to emphasize at the outset that the definition of probability as given below is applicable only to certain objects A and to certain of their properties B—not to all possible. (page 336)

Kolmogorov later made the same point at greater length in an encyclopedia article:

> It is far from true that every event whose occurrence is not unambiguously determined under given conditions has a definite probability under those conditions. The assumption that a definite probability (i.e. a completely defined fraction of the number of occurrences of an event if the conditions are repeated a large number of times) in fact exists for a given event under given conditions is a *hypothesis* which must be verified or justified in each individual case. ([221], 1951)

Neyman and Kolmogorov were only stating the obvious; everyone agreed with them at the time. It may have been the closed appearance of the new framework that made these comments feel needed.

Today the common knowledge of the 1930s and 1950s has faded. The applications of probability have mushroomed far beyond statistics, and many users of probability do not have have the experience or training with data that cautions against equating the indeterminate with the probabilistic. It is now commonplace, in many domains, to see a general assumption that everything does have a probability. This is especially true in some areas of economics, including finance. Academics and practitioners in finance routinely and fearlessly assume the existence of unknown "physical probabilities" for non-repeatable events in financial markets. Nassim Taleb, their most iconoclastic critic, challenges them in his bestselling book, *The Black Swan*, not for assigning probabilities to events that do not have probabilities, but for getting the probabilities wrong. This unquestioning acceptance of mysterious probabilities may have many sources, but the authority and closed appearance of Kolmogorov's framework is surely one of them.

Numbers of articles in five leading journals that used different terms for *random variable*.

	1930s	**1940s**	**1950s**	**1960s**
Annals of Mathematical Statistics				
chance variable	4	37	53	9
random variable	12	108	310	776
stochastic variable	2	11	3	8
statistical variable	2	4	0	0
random variate	0	2	6	4
Biometrika				
chance variable	3	1	2	4
random variable	2	23	83	222
stochastic variable	1	2	3	0
statistical variable	2	3	0	1
random variate	0	0	4	7
Journal of the American Statistical Association				
chance variable	0	6	8	4
random variable	1	13	94	284
stochastic variable	0	2	3	6
statistical variable	1	1	1	0
random variate	0	1	3	2
Journal of the Royal Statisical Society				
chance variable	1	2	0	0
random variable	4	10	18	59
stochastic variable	1	1	1	1
statistical variable	2	0	2	1
random variate	0	0	0	3
Transactions of the American Mathematical Society				
chance variable	6	8	2	0
random variable	0	5	40	72
stochastic variable	0	0	1	0
statistical variable	0	0	0	0
random variate	0	0	0	0

These counts, obtained from JSTOR, should be considered approximate. JSTOR's indexing is imperfect; it does not, for example, detect Student's use of *random variable* in *Biometrika* in 1914. In some cases the counts include occurrences in abstracts and indexes, but I have tried to exclude these when the counts are small.

During the period studied, the *Annals of Mathematical Statistics* and the *Journal of the American Statistical Association* were the most prominent journals in statistics in the United States, and *Biometrika* and the *Journal of the Royal Statistical Society* were the most prominent ones in Britain. The *Transactions of the American Mathematical Society* was the most prominent mathematics journal that published a substantial number of articles on probability in English before 1970.

The most common names for a probabilized quantity in the 1930s and 1940s were simply *variable* and *variate*. I have not tried to count the numbers of articles using these names. Doing so would be both laborious (these words occur within and alongside our two-word names) and difficult (it is not always clear whether an author's *variable* or *variate* is probabilized).

11 Appendix: Linguistic diversity

Many assessments in the preceding narrative hinged judgements about how the vocabulary of probability in one language influenced the vocabulary of probability in another. In most cases, scholars over the period we have studied paid most attention to what was written in their own language, but most also kept abreast of developments in multiple languages. Latin was the international language of science in the seventeenth century, but English and French became more important in the eighteenth century. In the nineteenth, German became important and sometimes dominant. In the nineteen century and the first half of the twentieth century, most European and American mathematicians had at least a reading knowledge of German, French, English, and Italian, and many published in more than one language. Knowledge of Russian was less common. In the early twentieth century, until the Soviet regime stopped the practice in the late 1930s, Russian mathematicians published their best work in other languages, sometimes in journals published outside Russia, sometimes in journals published in Russia. British and American mathematical statisticians were less interested in work on the Continent, but they could read German and French when they wanted.

The linguistic diversity of probability at the turn of the twentieth century is illustrated by Frances Edgeworth's bibliography at the end of his article on probability in the 11th edition of the Encyclopædia Britannica (1911, [135]). In addition to articles by Crofton and Pearson, Edgeworth lists eleven books:

- three in French: Laplace's *Théorie analytique*, Bertrand's *Calcul des probabilités*, and Poincaré's *Calcul des probabilités*;
- four in German: von Kries's *Principien* and three by Czuber;
- four in English: Venn's *Logic of Chance*, Bowley's *Elements of Statistics*, and Whitworth's *Choice and Chance* and *DCC exercises.*

Another illustration of this diversity is the bibliography on least squares published by the Yale professor Mansfield Merriman in 1877 ([262], page 184). Merriman listed 408 publications that appeared from 1722 through 1876, only 22 of which appeared before 1805. He classified these articles by the language in which they were written and the country where they were published, obtaining the counts shown in Table 5, on page 93.

Table 6, on page 94, breaks Merriman's counts into three time periods and provides similar counts for bibliographies given by four other early authors: the British mathematician Isaac Todhunter, the French mathematician Hermann Laurent, the Austrian mathematician Emanuel Czuber, and the British economist John Maynard Keynes. Table 7, on page 95, gives similar information on the bibliographies given by five more contemporary historians: the American Stephen M. Stigler, the Dane Anders Hald, the Russian Oscar Sheynin, and the German Hans Fischer.

Table 5: Mansfield Merriman's counts of the languages and locations of 408 articles on least squares published before 1876.

Language of article		Location of publication	
Latin	16		
German	167	Germany	153
		Austria	10
French	110	France	78
		Belgium	19
English	90	Great Britain	56
		United States	34
Italian	9	Italy	14
Dutch	7	Holland	7
Danish	5	Denmark	5
Swedish	4	Sweden	7
		Switzerland	9
		Russia	16

Table 6: Distribution across languages for five early bibliographies. The numbers in boldface are percentages. For example, 21% of the citations in Laurent's bibliography for the period before 1800 were in Latin.

	Todhunter 1865 [358]		Laurent 1873 [238]		Merriman 1877 [262]		Czuber 1899 [101]		Keynes 1921 [200]	
		%		%		%		%		%
Before 1800										
Latin			17	**21**			23	**26**		
English			17	**21**			11	**12**		
French			41	**51**			50	**56**		
German			1	**1**			4	**4**		
Italian			1	**1**			1	**1**		
Dutch			3	**4**			1	**1**		
Swedish			1	**1**						
1800–1849										
Latin			6	**5**			3	**3**		
English			11	**9**			11	**12**		
French			67	**55**			46	**48**		
German			33	**27**			34	**36**		
Italian			3	**2**						
Danish							1	**1**		
Dutch			1	**1**						
Swedish										
Russian			1	**1**						
1850–1899										
Latin										
English			16	**13**			53	**17**		
French			62	**51**			53	**17**		
German			36	**30**			165	**53**		
Italian			2	**2**			14	**4**		
Danish			1	**1**			7	**2**		
Dutch			1	**1**			9	**3**		
Polish							2	**1**		
Russian							3	**1**		
Swedish			1	**1**			8	**3**		
Finnish			1	**1**						

Table 7: Distribution across languages for five more recent bibliographies.

	Stigler 1986	%	Hald 1998 %	Sheynin 2009 %	Fischer 2001	%
Before 1800						
Latin	12	**17**			12	**60**
English	24	**38**			1	**5**
French	24	**38**			6	**30**
German	2	**3**			0	**0**
Italian	0	**0**			0	**0**
Dutch	1	**2**			1	**5**
Swedish	0	**0**			0	**0**
1800–1849						
Latin	3	**5**			9	**13**
English	13	**20**			2	**3**
French	42	**67**			27	**40**
German	6	**9**			28	**42**
Italian	0	**0**			0	**0**
Dutch	0	**0**			0	**0**
Swedish	0	**0**			0	**0**
Russian	0	**0**			1	**1**
1850–1899						
Latin	0	**0**			0	**0**
English	112	**77**			28	**21**
French	16	**11**			57	**43**
German	17	**12**			32	**24**
Italian	1	**1**			1	**1**
Dutch	0	**0**			0	**0**
Swedish	0	**0**			0	**0**
Danish	0	**0**			4	**3**
Finnish	0	**0**			0	**0**
Russian	0	**0**			12	**9**
1900–1949						
Latin	0	**0**			0	**0**
English	48	**89**			48	**23**
French	3	**6**			61	**29**
German	3	**6**			68	**32**
Italian	0	**0**			10	**5**
Dutch	0	**0**			1	**0**
Swedish	0	**0**			0	**0**
Danish	0	**0**			0	**0**
Finnish	0	**0**			1	**0**
Russian	0	**0**			24	**11**

Part III
Going Forward

> Далеко не всякое событие, наступление которого при заданнвых условиях не является однозначно определённым, имеет при этом комнплексе условий определённую вероятность.
>
> Andrei Andreevich Kolmogorov, 1951 [221][8]

This working paper has been written to provide perspective for a contemporary project, that of generalizing Kolmogorov's framework so that we can consider within it events and variables for which we do not have probabilities. There are two major ways of thinking about this project, corresponding to what Jerzy Neyman called the classical and subjective views of probability. These views put different interpretations on the word *random*, but regardless of which view one adopts, much will be gained, in communicating with each other and in drawing on the accomplishments of the past, if we reserve the name *random variable* for those variables for which we do have probability distributions.[94]

12 The two views in terms of betting

The view that Neyman called *classical* insisted that the application of probability theory involves hypotheses about the world, predictions based on these hypotheses, and empirical tests of the predictions. It would be a caricature to equate this classical view with the views of John Venn and R. L. Ellis, who equated probability with frequency. But Neyman called it *frequentist*, because it saw the approximation of probabilities by frequencies, as predicted by Chebychëv's law of large numbers, as the archetype of testable probabilistic prediction. Nearly all the individuals in the history just recounted subscribed to the classical view. But the subjective view was always in the wings.

Neyman described the situation of the subjective view of probability at the end of the 1930s as follows:

> ...There is considerable variation of, say, radicalism, among the present proponents of this theory. The most radical seem to be the writings of Jeffreys for whom to any proposition, on any amount of information, there corresponds a perfectly determined probability, which has nothing to do with frequencies. However, I could not quite follow the method of actually determining these probabilities. A less radical point of view, which I think I share, is represented by

[94]Recall comments by John Stuart Mill and Francis Edgeworth concerning the "evil consequences of casting off any portion of the customary connotation of words" [134], page 382).

> Borel, for whom, whenever possible, the probability does represent a mathematical correspondent of the observable relative frequency but who does not deny the possibility of usefully considering the probabilities of isolated instances. ... ([279], page 83)[95]

The subjective view has grown in strength and in radicalism since Neyman wrote this passage. As I have suggested, this stems in part from the contradictions built into the synthesis of the 1930s, which were eventually felt as shortcomings of the classical view.

By the end of the twentieth century, most proponents of the subjective view were anchoring their thinking in the betting picture with which mathematical probability theory began in the seventeenth century. As Bruno de Finetti had explained in the early 1930s[96] a subjective probability can be considered a guarded willingness to bet. When you have thought about the evidence for E and assigned it probability p, you will be inclined to accept modest bets on or against E at odds p to $(1-p)$. It might be a bad idea to bet on E with someone else who might know more than you. But perhaps you will act in other ways, not involving an adversary, that will reveal $p:(1-p)$ to be your odds on E.

As Vladimir Vovk and I argued in our 2001 book [326], the classical view can also be reconstructed with betting as its foundation. In this case, proffered odds are not necessarily expressions of anyone's opinion. They are conjectures, to be tested by strategies for betting. Classical tests that do not average over irrelevant possibilities[97] can always be recast in this way. In the case of stochastic processes, the betting can be implemented over time as conjectures and observations are made.

13 Generalizing Kolmogorov subjectively

As Neyman emphasized, it was a tenet of the subjective view of probability that one can provide probabilities for any proposition, on any amount of information. But de Finetti's picture left open one way in which you might fail to have a probability. You might not have taken the time to think about the proposition. For contemporary subjectivists, probability is a norm, not a psychological theory. We do not always have probabilities in our heads, but when we need to make a decision, we should try to assess relevant evidence by making probability judgements. In complicated situations we are unlikely ever to complete this exercise,

[95]This quotation is drawn from Neyman's contribution to a colloquium on the applications of probability held in Geneva in July 1939, which was conceived as a sequel to the celebrated colloquium on probability held there in 1937. Because the war intervened, the proceedings of the 1939 meeting were not published until 1946, but Neyman's contribution does not seem to have been revised after 1939. He cites Harold Jeffreys's 1931 *Scientific Inference* [195] and a book just published by Emile Borel [32], but not Jeffreys's 1939 *Theory of Probability* [196].

[96]De Finetti's most important Italian articles from this period are now available in English translation in [117]. His earliest article on the subjective foundations of probability, written in 1929 and published in 1931, was translated into English in a special issue of *Erkenntnis* devoted to his work [116]. His 1937 presentation in French at the Institut Henri Poincaré [114] has been available in English since the 1960s [225] and is widely cited. See also [161].

[97]See [113] for a discussion of such averaging.

and so at any point we will have only incomplete probability assessments. This opening was explored at length by Peter Walley in his *Statistical Reasoning with Imprecise Probabilities* [374], which stimulated ongoing interest in generalizing Kolmogorov's framework in a way that permits subjective interpretation.

What role should the term *random variable* play in this generalization? This was a vexed question for Walley, because many subjectivisits are uncomfortable with the term. For the most part, subjectivists have adopted *random* as their own; an event or quantity is random for you if it is unknown and you have assessed its subjective probability. But some, including de Finetti, rejected *variable*. For de Finetti, ([115], §1.10) there was nothing variable about an unknown quantity; it is what it is. So, following Markov and Kolmogorov, we should call it a random quantity, not a random variable. (After all, de Finetti adds as a debating point, we now talk about random vectors, random matrices, and so on; why should a quantity be called a variable when it is random, while these other objects are not?) Statisticians must study statistical variables, of course—quantities that vary from individual to individual in a population, but these are not random quantities (see also Dempster [121], page 4).

Walley avoided this argument by calling a real-valued function on Ω a *gamble* rather than a *random variable* or *random quantity*. Some authors have followed Walley in his use of *gamble*, but this novel usage is a barrier for those trained only in standard terminology, and so other authors have reverted to *random variable*. The adjective *random* is defended on the grounds that you potentially have a probability distribution for the variable; you just need to give the matter some thought.

14 Generalizing Kolmogorov classically

The game-theoretic reconstruction of classical probability in my book with Vovk formalizes the betting picture, which remained relatively informal in de Finetti's work. This formalization typically distinguishes three players: one who sets odds (this player might represent a theory or an observer who sets up the conditions for the application of a theory), one who decides what bets to make at these odds in order to test them, and one who decides the outcomes (this player might represent the interaction of many people or many physical influences). This obviously allows a generalization of Kolmogorov's framework to a partial probability structure: we simply allow the player who sets odds to do so for all events within the framework, or even to offer to take either side of every bet offered.

In fact, so long as we do not impose a fixed strategy on the player setting the odds (thereby removing him as an active participant in the game), this game-theoretic probabilistic structure is necessarily partial as compared to the full probability structure assumed by Kolmogorov. This is because from the perspective of the player testing the conjectured odds by betting, the observations include the moves of both the other players—the one giving the odds and the one deciding the outcomes. The set Ω of possible observations is the set of all sequences of moves by these two players. This puts the probability structure inside Ω and makes it partial; the observer can only bet on the relationship between the moves of his two opponents. This aspect of the matter, together with the fact that the observer is also inside the game-theoretic formalism, makes the whole structure both very open and very flexible.

Borrowing the title of Jacob Bernoulli's celebrated *Ars Conjectandi* [23], we can call this game-theoretic set-up a framework for conjecture. To understand how it accommodates the classical testing of statistical hypotheses that interested Neyman, consider Chebyshev's law of large numbers, quoted on page 30 above. This theorem says that when n random variables $U_1, \ldots U_n$ are compared with their expected values $E(U_1), \ldots E(U_n)$, the difference

$$\frac{\sum_{i=1}^n U_i}{n} - \frac{\sum_{i=1}^n E(U_i)}{n} \tag{7}$$

will be arbitrarily small with arbitrarily high probability for n sufficiently large. The game-theoretic way of proving that an event happens with "arbitrarily high probability" is to prove the existence of a strategy that mulitplies the capital it risks by an arbitrarily large factor if the event does not happen. This strategy functions both as a proof and as a test. The expected values or prices $E(U_i)$ can be rejected if the strategy succeeds. To accommodate stochastic processes, which Neyman considered the fundamental and essential contribution of modernized classical probability to science [281], we use the "martingale" version of Chebyshev's theorem, which replaces $E(U_i)$ in (7) with the conditional expected value $E_{i-1}(U_i)$, the expected value of U_i after the previous quantities are observed—game-theoretically, the price for U_i given on the ith round by the player setting the odds.

When U_i is the indicator function for an event A_i, so that $E_{i-1}(A_i)$ is the event's probability $P_{i-1}(A_i)$, the test statistic (7) becomes the difference

$$\begin{array}{l}\text{fraction of the } A_i \\ \text{that happen}\end{array} \quad - \quad \frac{\sum_{i=1}^n P_{i-1}(A_i)}{n}. \tag{8}$$

Thus we reject the purported probabilities if their average does not approximate the frequency with which the events happen. This justifies Neyman's calling his philosophy "frequentist", but the label can be misleading. The essential point is the odds or probabilities or expected values are tested by observations. Moreover, our framework for conjecture generalizes to cases where the $E_{i-1}(U_i)$ are replaced by upper or lower numbers and the tests are only one-sided.

As an acknowledgement that our framework for conjecture incorporates Kolmogorov's framework and classical statistical testing, Vovk and I retained the symbol Ω and the name *sample space*. We also called subsets of Ω *events* and real-valued functions on Ω *variables*.

The idea of testing odds or probabilities does not exclude giving them a subjective interpretation. Indeed, a framework with three players has plenty of room for multiple perspectives; the player giving the odds can play the role of the believer while the player testing them plays the role of the skeptic. But our emphasis on testing does not fit well with the assumption that partial probabilities are on their way to becoming complete probabilities, for when the observations suggest that your partial probabilities should be rejected, you may want to change them before ever completing them. So it would be very misleading for us to call variables to which we have not already assigned a complete probability distribution *random*.

15 Conclusion

The mathematical framework based on sequential betting offers that was laid out in my 2001 book with Vovk can be interpreted in classical or subjective terms. When the moves by the player who offers the odds are fixed, it is more or less isomorphic to the framework for imprecise probabilities developed by Walley [374], Troffaes and de Cooman [359] and other authors [12]. This mathematical similarity can be obscured by terminology. Just as the mathematicians and statisticians of the 1930s adopted a common vocabulary in order to make communication possible, the proponents of different interpretations of a generalized Kolmogorovian framework should adopt as far as possible a vocabulary that helps us talk with each other and preserve our ties to the accomplishments of the past.

An important priority should be to keep contact with as much of the heritage of probability and statistics as possible, including historical compromises that may look like contradictions. So yes, let us call Ω our *sample space*, even though we will not repeatedly sample from it. Like Cournot 170 years ago, let us talk about *variables*. But let us call variables *random* only when they have probability distributions.

Notes

[1] William Feller, along with Doob one of the leading advocates of Kolmogorov's framework made the point with these words in his 1936 paper on stochastic processes ([139], page 115):

> ...the beiläufig erwähnten wahrscheinlichkeitstheoretisch Begriffe unabhängig von jedem Anwendungsproblem mengentheoretisch exakt gefaßt wurden, insbesondere in der durchsichtigen Axiomatik von Kolmogoroff...

English translation:

> ...the just mentioned theoretical probability ideas are made set-theoretically precise independently of any applied problem, especially in Kolmogorov's thorough axiomatization ...

[2] In his *Ars conjectandi* [22], published posthumously in 1713, Jakob Bernoulli wrote (pages 210–122):

> *Certitudo* rei cujusvis spectatur vel *objectivè* & in re; nec aliud significat, quàm ipsam veritatem existentiae aut futuritionis illius rei: vel *subjective* & in ordine ad nos; & consistit in mensura cognitionis nostræ circa hanc veritatem. ...*Probabilitas* enim est gradus certitudinis, & ab hac differt ut pars à toto.

English translation by Edith Sylla ([23], page 315):

> The *certainty* of anything is considered either *objectively* and in itself or *subjectively* and in relation to us. Objectively, certainty means nothing else than the truth of the present or future existence of the thing. Subjectively, certainty is the measure of our knowledge concerning this truth. ... *Probability*, indeed, is degree of certainty, and differs rom the latter as a part from the whole.

In the preface to his *Exposition* [78], 1843, Antoine-Augustin Cournot wrote:

> J'ai craint qu'on ne me reprochât, si j'y insistais davantage ici, de trop mêler la métaphysique à la géométrie. *Est modus in rebus.* C'est pourtant à la langue des métaphysiciens que j'ai emprunté sans scruple les deux épithètes d'*objective* et de *subjective*, qui m'étaient nécessaire pour distinguer radicalement les deux acceptions du terme de *probabilité*, auxquelles s'appliquent les combinaisons du calcul; mais j'y étais autorisé par l'exemple de Jacques Bernoulli.

English translation:

> Were I to insist further on this point, I fear I would be criticized for injecting too much metaphysics into the mathematics. *Est modus in rebus.* I have nevertheless borrowed from the metaphysicians the two terms *objective* and *subjective*, which I need in order to distinguish sharply between the two meanings of the term *probability* to which the theory's calculations apply. But the example of Jakob Bernoulli authorized me to do this.

[3]In 1886, the nineteenth-century German statistician Wilhem Lexis asked this question as follows ([247], first sentence):

> Wenn die Wahrscheinlichkeitsrechnung sich als rein mathematischer Lehrzweig in voller Sicherheit außerhalb aller Meinungsverschiedenheiten über ihre praktische Anwendbarkeit begründen und ausbauen läßt, so finden wir andererseits mannigfaltige und wesentlich verschiedene Anschauungen über die objektive Bedeutung des Begriffs der mathematischen "Wahrscheinlichkeit" und über die Rolle, die ihr innerhalb des allgemeinen Kausalzusammenhanges der wirklichen Erscheinungen zuzuschreiben ist.

[4][27], page 310. English translation: The probability calculus is mathematics' first step outside the domain of absolute truth.

[5]The most comprehensive historical study of this topic is by Andrew Dale [105]. Chapter 3 of Stigler's *History of Statistics* [352] may be the most insightful treatment of the crucial contributions by Laplace and Bayes. I discussed an important aspect of Bayes's argument in [325].

[6]See [327, 328].

[7]Fuller quotation from Anderson ([10], pages 171–172):

> Dei Ausarbeitung einer verzweigten Methode der mathematischen Erwartung is das Werk der neuesten Zeit, und zwar haben sich darum vorwiegend russische Mathematiker verdient gemacht. Abgesehen von Tschebyscheff, der sich in der zweiten Hälfte des 19. Jahrhunderts höchst erfolgreich, aber doch nur vorübrigehend damit beschäftigte, sind hier in erster Linie A. Tschuprow, A. Markoff und in Deutschland L. Bortkiewicz zu nennen. Markoff hat auf den Begriffe der mathematischen Erwartung sein ganzes Lehrbuch der Wahrscheinlichkeitsrechnung aufgebaut und hierbei eine außerordentliche Einheitlichkeit des logischen Aufbaues der Materie und die größte Eleganz der Beweise erzielt. In dieser Hinsicht steht sein Werk ganz einzigartig da. Die zweite Auflage desselben ist von H. Liebmann deutsch herausgegeben worden (Leipzig 1912). Doch bleibt die Übersetzung weit hinter dem Original zurück, insbesondere hinter der posthumen 4. Augl. (Moskau 1924). Es sei ferner auf die bereits mehrmals zitierten "Iterationen" Bortkiewicz' hingewiesen, in welchen der Methode der mathematischen Erwartungen das zweite Kapitel gewidmet ist, und auf die "Wahrscheinlichkeitsrechnung" von Mises (S. 37–127 passim). Eine elementar gehaltene Darstellung einiger Hauptsätze der Methode der mathematischen Erwartungen findet man auch in den "Grundzügen der Theorie der Statistik" von Westergaard und Nybølle (S. 188ff.). Vgl. ferner noch E. Czuber, Wahrscheinlichkeitsrechnung, I. Bd., S. 72–80.

[8]English translation: It is far from true that every event whose occurrence is not unambiguously determined under given conditions has a definite probability under those conditions.

Acknowledgements

I have benefited immensely from suggestions and comments on earlier drafts by a number of people, including Thomas Augustin, Bernard Bru, Gert de Cooman, Hans Fischer, Prakash Gorroochurn, Alex Kogan, Steffen Lauritzen, Laurent Mazliak, Nell Painter, Eugene Seneta, Oscar Sheynin, and Steve Stigler, and Volodya Vovk. I thank them for their generosity. I also want to emphasize that this draft remains partial and tentative. None of the individuals just named have seen it, and many of their suggestions and corrections are not yet incorporated in it.

Notes concerning the list of references

The names of Russian authors were often transliterated in different ways in the different languages in which they published. Here their names are transliterated into English following the Library of Congress system, except that diacritics and hard and soft signs are omitted. The Library of Congress system is used by most libraries in the United States and Great Britain.

Some other names are also given in a fuller or different form than in the publication cited. In the case of authors who changed their nationality, I have used the name the author used in his adopted language, even when the publication occurred earlier: Oskar N. Anderson rather than Oskar Nikolaevich Anderson, Laudislaus von Bortkiewicz rather than Vladislav Iocifovich Bortkevich, William Feller rather than Willy Feller, James Victor Uspenksy rather than Jakob Viktorovich Uspenskii.

For my own convenience, I have listed urls when the cited book or article is directly available on-line on an open access basis. Many or most of these urls may be ephemeral and presumably will be omitted in any published version of the paper.

References

[1] William J. Adams. *The Life and Times of the Central Limit Theorem.* American Mathematical Society, Providence, second edition, 2009. 41, 120

[2] R. J. Adcock. Law of random errors. *The Analyst*, 10(5):140–141, 1883. 37

[3] John Aldrich. The language of the English biometric school. *International Statistical Review*, 71(1):109–129, 2003. 64, 65

[4] John Aldrich. Burnside's engagement with the "modern theory of statistics". *Archive for History of Exact Sciences*, 63(1):51–79, 2009. 64

[5] John Aldrich. England and probability in the inter-war years. *Electronic Journal for History of Probability and Statistics*, 5(2), 2009. 65

[6] Vissarion Grigorevich Alekseev. N. W. Bugajew und die idealistischen Probleme der Moskauer mathematischen Schule. *Vierteljahrsschrift für wissenschaftliche Philosophie und Soziologie*, 29:335–367, 1905. 39

[7] Vissarion Grigorevich Alekseev. *Goethe, Schiller, Herbart im Lichte des Moskauer exakten Idealismus*. Dorpater Zeitung, Dorpat, 1926. (OCoLC)669405452. 39

[8] Oskar N. Anderson. Nochmals über "The elimination of spurious correlation due to partition in time or space". *Biometrika*, 10:269–279, 1914. 58

[9] Oskar N. Anderson. On the logic of the decomposition of statistical series into separate components. *Journal of the Royal Statistical Society*, 90:548–569, 1927. 84

[10] Oskar N. Anderson. *Einführung in die mathematische Statistik*. Springer, Vienna, 1935. 43, 68, 101

[11] Philippe Artzner, Freddy Delbaen, Jean-Marc Eber, and David Heath. Coherent measures of risk. *Mathematical Finance*, 9:203–228, 1999. 5

[12] Thomas Augustin, Frank P. A. Coolen, Gert de Cooman, and Matthias C. M. Troffaes, editors. *Introduction to Imprecise Probabilities*. Wiley, Chichester, 2014. 5, 100

[13] Louis Bachelier. *Le jeu, la Chance, et le Hasard*. Flammarion, Paris, 1914. 52

[14] E. Barbier. Note sur le problème de l'aiguille et le jeu du joint couvert. *Journal de Mathématiques Pures et Appliquées*, pages 273—286, 1860. http://eudml.org/doc/234201. 31

[15] Marc Barbut, Bernard Locker, and Laurent Mazliak. *Paul Lévy and Maurice Fréchet, 50 Years of Correspondence in 107 letters*. Springer, London, 2014. Translation from the French edition, published in 2004 by Hermann in Paris. 73, 76, 89

[16] Vincent Barnett. *E. E. Slutsky as Economist and Mathematician*. Routledge, London, 2011. 80

[17] William Dowell Baten. Simultaneous treatment of discrete and continuous probability by use of Stieltjes integrals. *Annals of Mathematical Statistics*, 1:95–100, 1930. https://projecteuclid.org/download/pdf_1/euclid.aoms/1177733262. 63

[18] William Dowell Baten. Combining two probability functions. *Annals of Mathematical Statistics*, 5:13–20, 1934. https://projecteuclid.org/download/pdf_1/euclid.aoms/1177732732. 63

[19] Margherita Benzi. Un "probabilista neoclassico": Francesco Paolo Cantelli. *Historia Mathematica*, 15:53–72, 1988. http://ac.els-cdn.com/0315086088900493/1-s2.0-0315086088900493-main.pdf?_tid=

dfb4b442-0d76-11e5-98fe-00000aacb361&acdnat=1433724171_d3ef81ccfcf8c00f620adfa4c3dbb58b. 59

[20] Margherita Benzi, Michele Benzi, and Eugene Seneta. Francesco Paolo Cantelli. *International Statistical Review*, 75:127–130, 2007. 59

[21] Daniel Bernoulli. Specimen theoriae novae de mensura sortis. *Commentarii Academiae Scientiarum Imperialis Petropolitanae*, 5:175–192, 1738. 10

[22] Jacob Bernoulli. *Ars Conjectandi*. Thurnisius, Basel, 1713. 100, 104

[23] Jacob Bernoulli. *The Art of Conjecturing, together with Letter to a Friend on Sets in Court Tennis*. Johns Hopkins University Press, Baltimore, 2006. Translation of [22] and commentary by Edith Sylla. 9, 99, 100

[24] Sergei N. Bernshtein. Теория вероятностей *(Theory of Probability)*. Государственное издателство (State Publisher), Moscow, 1927. 44

[25] Sergei N. Bernshtein. Петербургская школа теории вероятностей (the Petersburg school of the theory of probability). *Uchenye Zapiski Leningradsk. Gosudarstven. Univ.*, 55:3–11, 1940. Translation at www.sheynin.de. 40

[26] Joseph Bertrand. *Calcul des Probabilités*. Gauthier-Villars, Paris, 1889. https://archive.org/details/calculdesprobab00bertgoog. Some copies of the first edition are dated 1888. A second edition appeared in 1907. 22, 52

[27] Irenée Jules Bienaymé. Considérations à l'appui de la découverte de Laplace sur la loi de probabilité dans la méthode des moindres carrés. *Comptes rendus hebdomadaires des séances de l'Académie des Sciences*, 37:309–324, 1853. Session of 29 August 1853. http://gallica.bnf.fr/ark:/12148/bpt6k29948/f313.image.langEN. 35, 101

[28] Olav Bjerkholt. Ragnar Frisch and the foundation of the Econometric Society and Econometrica. Technical report, Statistics Norway, Research Department, Oslo, 1995. Document 95/9. www.ssb.no/a/histstat/doc/doc_199509.pdf. 80

[29] Georg Bohlmann. Lebensversicherungs-Mathematik. In *Encyklopädie der mathematischen Wissenschaften, Bd. I, Teil 2*, pages 852–917. Teubner, Leipzig, 1901. 52

[30] Bernhard Friedrich Boigt, editor. *Neuer Nekrolog der Deutschen*, volume 1. Boigt, Flemenau, 1833. https://books.google.com/books/reader?id=JPdQAAAAcAAJ&printsec=frontcover&output=reader. 20

[31] George Boole. *An investigation of the laws of thought, on which are founded the mathematical theories of logic and probabilities*. Macmillan, London, 1854. Reprinted by Dover, New York, 1958. 37

[32] Émile Borel. *Valeur pratique et philosophie des probabilités*. Gauthier-Villars, Paris, 1939. Reprinted in 1991 by Éditions Jacques Gabay. 97

[33] Ladislaus von Bortkiewicz. Kritische Betrachtungen zur theoretischen Statistik. Erster Artikel. *Jahrbücher für Nationalökonomie und Statistik*, 63:641–680, 1894. 47

[34] Ladislaus von Bortkiewicz. Kritische Betrachtungen zur theoretischen Statistik. Zweiter Artikel. *Jahrbücher für Nationalökonomie und Statistik*, 65:321–360, 1895. 47

[35] Ladislaus von Bortkiewicz. Kritische Betrachtungen zur theoretischen Statistik. Dritter Artikel. *Jahrbücher für Nationalökonomie und Statistik*, 66:671–705, 1896. 47

[36] Ladislaus von Bortkiewicz. *Das Gesetz der kleinen Zahlen*. Teubner, Leipzig, 1898. 48

[37] Ladislaus von Bortkiewicz. Anwendungen der Wahrscheinlichkeitsrechnung auf Statistik. In *Encyklopädie der mathematischen Wissenschaften, Bd. I, Teil 2*, pages 821–851. Teubner, Leipzig, 1901. 52

[38] Ladislaus von Bortkiewicz. *Die Iterationen. Ein Beitrag zur Wahrscheinlichkeitstheorie*. Springer, Berlin, 1917. `https://ia700409.us.archive.org/10/items/dieiterationene00bortgoog/dieiterationene00bortgoog.pdf`. 68

[39] John E. Bowit. *Moscow and St. Petersburg in Russia's Silver Age, 1900–1920*. Thames & Hudson, London, 2008. 37

[40] Arthur Lyon Bowley. *Elements of Statistics*. King, Westminster, 1901. `https://archive.org/details/elementsstatist03bowlgoog`. Later editions appeared in 1902, 1907, 1920, 1925, and 1937. 55

[41] Bernard Bru. Poisson, le calcul des probabilités, and l'instruction public. In Piere Costabel, Pierre Dugac, and Michel Métiver, editors, *Siméon-Denis Poisson et la science de son temps*, pages 51–94. École Polytechnique, Palaiseau, 1981. English translation in [43]. 15, 105

[42] Bernard Bru. Souvenirs de Bologne. *Journal de la Société Française de Statistique*, 144(1–2):134–226, 2003a. `http://archive.numdam.org/ARCHIVE/JSFS/JSFS_2003__144_1-2/JSFS_2003__144_1-2_135_0/JSFS_2003__144_1-2_135_0.pdf`. 70

[43] Bernard Bru. Poisson, the probability calculus, and public education. *Electronic Journal for History of Probability and Statistics*, 1(2), November 2005. Translation of [41]. `http://www.jehps.net/`. 15, 18, 105

[44] Heinrich Bruns. *Wahrscheinlichkeitsrechnung und Kollektivmasslehre*. Teubner, Leipzig and Berlin, 1906. 54

[45] Nikolai Vasilevich Bugaev. Les mathématiques et la conception du monde au point de vue de la philosophie scientifque. In Ferdinand Rudo, editor, *Verhandlungen des Ersten Internationalen Mathematiker-Kongresses in Zürich vom 9. bis 11. August 1897*, pages 106–223. Teubner, Leipzig, 1898. 38

[46] Viktor Yakovlevich Bunyakovskii. Основания Математическоѝ Теории Вероятностеѝ *(Foundations of the Mathematical Theory of Probabilities)*. Imperial Academy of Sciences, Saint Petersburg, 1846. 11

[47] William Burnside. *Theory of Probability*. Cambridge University Press, Cambridge, 1928. `https://ia700404.us.archive.org/11/items/theoryofprobabil00burn/theoryofprobabil00burn.pdf`. 64

[48] Paul Butzer and François Jongmans. P. l. chebyshev (1821–1894), a guide to his life and work. *Journal of Approximation Theory*, 96:111–138, 1999. 40

[49] Gillispie Charles C. Mémoires inédits ou anonymes de Laplace sur la théorie des erreurs, les polynômes de Legendre et la philosophie des probabilités. *Revue d'histoire des sciences*, 32(3):223–279, 1979. `http://www.persee.fr/web/revues/home/prescript/article/rhs_0151-4105_1979_num_32_3_1633`. 14

[50] Francesco Paolo Cantelli. *Sull'adattamento delle curve ad una serie di misure o di osservazioni*. Pubblicato dalla Officina Tipografica Bodoni, Rome, 1905. Reprinted in [58], pages 46–65. 59

[51] Francesco Paolo Cantelli. Calcolo delle probabilità. *Il Pitagora. Giornale di matematica per gli alunni delle scuole secondarie*, 12:33–39 and 68–74, 1905–1906. In issue 3–4, for December 1905 and Janaury 1906, continued in issue 5–6–7, for February, March, and April 1906. 59

[52] Francesco Paolo Cantelli. Intorno ad un teorema fondamentale della teoria del rischio. *Bolletino dell'Associazione degli Attuari Italiani*, 24:1–23, 1910. Reprinted in [58], pages 66–85. 59

[53] Francesco Paolo Cantelli. Sulla differenza media con ripetizione. *Giornale degli Economisti e Rivista di Statistica*, 46(2):194–199, 1913. 60

[54] Francesco Paolo Cantelli. La tendenza ad un limite nel senso del calcolo delle probabilità. *Rendiconti del Circolo Matematico di Palermo*, 41:191–201, 1916. Reprinted in [58], pp. 175–188. 60

[55] Francesco Paolo Cantelli. Sulla legge dei grandi numeri. *Memorie, Accademia dei Lincei*, V, 11:329–349, 1916. Reprinted in [58], pp. 189–213. 60

[56] Francesco Paolo Cantelli. Sulla probabilità come limite della frequenza. *Atti Reale Accademia Nazionale dei Lincei. Rendiconti*, 26:39–45, 1917. Reprinted in [58], pp. 214–221. 60

[57] Francesco Paolo Cantelli. Considérations sur la convergence dans le calcul des probabilités. *Annales de l'Institut Henri Poincaré*, 5:3–50, 1935. `http://eudml.org/doc/78991`. Reprinted in [58], pp. 322–372. 60, 62

[58] Francesco Paolo Cantelli. *Alcune Memorie Matematiche*. Giuffrè, Milan, 1958. 106

[59] Guido Castelnuovo. *Calcolo delle probabilitá*. Albrighi e Segati, Milan, Rome, and Naples, 1919. Second edition in two volumes, 1926 and 1928. Third edition 1948. 50, 53, 61

[60] Guido Castelnuovo. Sur quelques problèmes se rattachant au calcul des probabilités. *Annales de l'Institut Henri Poincaré*, 3:465–490, 1933. `http://www.numdam.org/item?id=AIHP_1933__3_4_465_0`. 74

[61] Rémi Catellier and Laurent Mazliak. The emergence of French probabilistic statistics. Borel and the Institut Henri Poincaré around the 1920s. *Revue d'histoire des mathématiques*, 18:271–335, 2012. 70

[62] Beatrice M. Cave and Karl Pearson. Numerical illustrations of the variate difference correlation method. *Biometrika*, 10:340–355, 1914. 58

[63] Carl V. L. Charlier. *Researches into the theory of probability.* Ohlsson, Lund, 1906. 56

[64] Pafnutii Lvovich Chebyshev. Démonstration élementaire d'une proposition générale de la théorie des probabilités. *Journal für die reine und angewandte Mathematik*, 33:259–267, 1846. `https://eudml.org/doc/183251`. 28

[65] Pafnutii Lvovich Chebyshev. Des valeurs moyennes (translation from the Russian by N. de Khanikof). *Journal de Mathématiques Pures et Appliquées*, pages 177–184, 1867. `http://eudml.org/doc/234989`. The Russian version is in Математический Сборник, volume II, 1867, pages 1–9, `http://www.mathnet.ru/php/archive.phtml?wshow=paper&jrnid=sm&paperid=6922&option_lang=eng`. 28

[66] Pafnutii Lvovich Chebyshev. О двух теоремах относительно вероыаностей. Записки Академии Наук, 55:pages numbers not usually given, because the volume is so little available, 1887. Published in French as[67]. 41, 107

[67] Pafnutii Lvovich Chebyshev. Sur deux théorèmes relatifs aux probabilités. *Acta Mathematica*, 14:305–315, 1890. Translation of [66]. 107

[68] Aleksandr Aleksandrovich Chuprov. Очерки по теории статистики *(Essays on the theory of statistics).* Sabashnikov, Saint Petersburg, second edition, 1910. The first edition appeared in 1909. The second edition was reprinted by the State Publishing House, Moscow, in 1959. 40, 47, 67

[69] Aleksandr Aleksandrovich Chuprov. On the mathematical expectation of the moments of frequency distributions. *Biometrika*, 12:140–169, 1918. Continued in Volume 19, 1919, pages 185–210. 65, 67

[70] Aleksandr Aleksandrovich Chuprov. Zur Theorie der Stabilität statistischer Reihen. *Skandinavisk Aktuarietidskrift*, 1:199–256, 1918. Continued on pages 80–133 of Volume 2. 67

[71] Aleksandr Aleksandrovich Chuprov. On the mathematical expectation of the moments of frequency distributions. *Biometrika*, 13:283–295, 1921. 67

[72] Aleksandr Aleksandrovich Chuprov. Über die Korrelationsfläche der arithmetischen Durschnitte (ein Grenztheorem). *Metron*, 1:41–48, 1921. 67

[73] Aleksandr Aleksandrovich Chuprov. On the mathematical expectation of the moments of frequency distributions in the case of correlated observations. *Metron*, 3:461–493 and 646–683, 1923. 67, 68

[74] Aleksandr Aleksandrovich Chuprov. *Grundbegriffe und Grundprobleme der Korrelationstheorie.* Teubner, Leipzig and Berlin, 1925. A Russian edition was published in Moscow in 1926. An English translation [76] appeared in 1939. 67, 107

[75] Aleksandr Aleksandrovich Chuprov. On the asymptotic frequency distribution of the arithmetic means of n correlated observations for very great values of n. *Journal of the Royal Statistical Society*, 88:91–104, 1925. 51, 63, 67

[76] Aleksandr Aleksandrovich Chuprov. *Principles of the Mathematical Theory of Correlation.* Hodge, London, 1939. Translation of [74]. 51, 67, 107, 110

[77] Julian Lowell Coolidge. *An Introduction to Mathematical Probability*. Oxford University Press, London, 1925. German translation by F. M. Urban published in 1927 as *Einführung in die Wahrscheinlichkeitsrechnung* by B. G. Teubner, Leipzig and Berlin. 53, 63

[78] Antoine-Augustin Cournot. *Exposition de la théorie des chances et des probabilités*. Hachette, Paris, 1843. `http://gallica.bnf.fr/ark:/12148/bpt6k679998`. Reprinted in 1984 as Volume I (B. Bru, editor) of [80]. 21, 23, 25, 26, 101, 108

[79] Antoine-Augustin Cournot. *Die Grundlehren der Wahrscheinlichkeitsrechnung, leichtfaßlich dargestellt für Philosophen, Staatsmänner, Juristen, Kameralisten und Gebildete überhaupt*. Leibrod, Braunschweig, 1849. Abridged translation of [78] by Christian Heinrich Schnuse. `http://catalog.hathitrust.org/Record/011436243`. 11, 21, 23, 25, 26, 59

[80] Antoine-Augustin Cournot. *Œuvres complètes*. Vrin, Paris, 1973–2010. 11 volumes. 108

[81] Antoine-Augustin Cournot. *A. A. Cournot: Œurvres Complètes. Tome I: Exposition de la théorie des chances et des probabilités*. Vrin, Paris, 1984. Republication, with introduction, notes, and index, of [78]. Edited by Bernard Bru. 12, 21, 24

[82] Antoine-Augustin Cournot. *Exposition of the Theory of Chances and Probabilities*. NG Verlag, Berlin, 2013. Tanslation of [78] by Oscar Sheynin. `www.sheynin.de`. 21

[83] Harald Cramér. Das Gesetz von Gauss und die Theorie des Risikos. *Skandinavisk Aktuarietidskrift*, 6:209–273, 1923. 81

[84] Harald Cramér. Remarks on correlation. *Skandinavisk Aktuarietidskrift*, 7:220–240, 1924. 81

[85] Harald Cramér. On some classes of series used in mathematical statistics. In *Proc. 6th Scand. Math. Congr. Copenhagen*, pages 399–425, 1925. 81

[86] Harald Cramér. On the composition of elementary errors. *Skandinavisk Aktuarietidskrift*, 11:13–74 and 141–180, 1928. 81

[87] Harald Cramér. *On the Mathematical Theory of Risk*. Centraltryckeriet (Sponsored by Försäkringsaktiebolaget Skandia), Stockholm, 1930. 80

[88] Harald Cramér. Sur les propriétés asymptotiques d'une classe de variables aléatoires. *Comptes rendus hebdomadaires des séances de l'Académie des Sciences*, 201:441–443, 1935. Session of 19 August. `http://gallica.bnf.fr/ark:/12148/bpt6k31534/f441.image.langEN`. 71

[89] Harald Cramér. *Random Variables and Probability Distributions*. Cambridge University Press, Cambridge, 1937. 18, 53, 88

[90] Harald Cramér. *Sannolikhetskalkylen och Några av dess Anvandningär*. Almqvist & Wiksell, Stockholm, Göteborg, Uppsala, 1949. 53

[91] Harald Cramér. Half a century with probability theory: Some personal recollections. *Annals of Probability*, 4:509–546, 1976. `https://projecteuclid.org/download/pdf_1/euclid.aop/1176996025`. 80

[92] Harald Cramér. Mathematical probability and statistical inference: Some personal recollections from an important phase of scientific development. *International Statistical Review*, 49:309–317, 1981. 81

[93] William Morgan Crofton. Note on local probability. In Miller [266], pages 84–86. `https://play.google.com/store/books/details?id=QU4hAQAAIAAJ&rdid=book-QU4hAQAAIAAJ&rdot=1`. 32

[94] William Morgan Crofton. On the theory of local probability, applied to straight lines drawn at random in a plane; the methods used being also extended to the proof of certain new theorems in the integral calculus. *Philosophical Transactions of the Royal Society of London*, 158:181–199, 1868. 33, 43

[95] William Morgan Crofton. On the proof of the law of errors of observations. *Philosophical Transactions of the Royal Society of London*, 160:175–187, 1870. 34

[96] William Morgan Crofton. On operative symbols in the differential calculus. *Proceedings of the London Mathematical Society*, 12:122–134, 1880/1881. Read April 14, 1881. 35

[97] William Morgan Crofton. Geometrical theorems relating to mean values. *Proceedings of the London Mathematical Society*, 8:304–309, 1886/1887. Read June 14, 1877. 34

[98] William Morgan Crofton. Probability. *Encyclopædia Brittanica, Ninth Edition*, XIX:768–788, 1890. Czuber [101] and Fischer [147] give the year 1885 for this article. The copy of volume XIX that I have consulted is an American reprint by Allen of New York dated 1890. 34

[99] Emanuel Czuber. *Geometrische Wahrscheinlichkeiten und Mittelwerte*. Teubner, Leipzig, 1884. 33, 34, 35, 43, 46

[100] Emanuel Czuber. *Theorie der Beobachtungsfehler*. Teubner, Leipzig, 1891. 35, 46

[101] Emanuel Czuber. *Die Entwickung der Wahrscheinlichkeitstheorie und ihrere Anwendungen*. Teubner, Leipzig, 1899. This book was issued as part two of volume 7 of the *Jahresbericht der Deutschen Mathematiker-Vereinigung*. 21, 46, 54, 94, 109

[102] Emanuel Czuber. Wahrscheinlichkeitsrechnung. In *Encyklopädie der mathematischen Wissenschaften, Band I, Teil 2*, pages 733–767. Teubner, Leipzig, 1900. 52

[103] Emanuel Czuber. *Wahrscheinlichkeitsrechnung und ihre Anwendung auf Fehlerausgleichung, Statistik und Lebensversicherung*. Teubner, Leipzig, 1903. The preface is dated November 1902. Later editions were in two volumes. The two volumes for the second edition appeared in 1908 and 1910, respectively. The third edition of the first volume appeared in 1914 (with a preface dated November 1912. 12, 35, 46, 47, 49, 50, 53, 58

[104] Emanuel Czuber. *Die philosophischen Grundlagen der Wahrscheinlichkeitsrechnung*. Teubner, Leipzig and Berlin, 1923. 46

[105] Andrew I. Dale. *A History of Inverse Probability From Thomsas Bayes to Karl Pearson*. Springer, New York, second edition, 1999. 101

[106] Georges Darmois. *Statistique mathématique*. Doin, Paris, 1928. 53, 70

[107] Georges Darmois. Analyse et comparaison des series statistiques qui se developpent dans le temps (the time correlation problem). *Metron*, 8:211–250, 1929. 74

[108] Georges Darmois. La méthode statistique dans les sciences d'observation. *Annales de l'Institut Henri Poincaré*, 3:191–228, 1932. http://www.numdam.org/item?id=AIHP_1932__3_2_191_0. 74

[109] Lorraine Daston. *Classical Probability in the Enlightenment*. Princeton University Press, Princeton, NJ, 1988. 14

[110] Charles Benedict Davenport. *Statistical methods with special reference to biological variation*. Wiley and Chapman & Hall, New York and London, 1899. 55

[111] Florence Nightingale David. Review of [362]. *Biometrika*, 30:194–195, 1938. 85

[112] Florence Nightingale David. Review of [76]. *Biometrika*, 31:396, 1940. 85

[113] A. Philip Dawid and Vladimir G. Vovk. Prequential probability: Principles and properties. *Bernoulli*, 5:125–162, 1999. 97

[114] Bruno de Finetti. La prévision, ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré*, 7:1–68, 1937. An English translation is included in both editions of [225]. 97

[115] Bruno de Finetti. *Teoria Delle Probabilità*. Einaudi, Turin, 1970. An English translation, by Antonio Machi and Adrian Smith, was published as *Theory of Probability* by Wiley (London, England) in two volumes in 1974 and 1975. 98

[116] Bruno de Finetti. Probabilism: A critical essay on the theory of probability and on the value of science. *Erkenntnis*, 31:169–223, 1989. 97

[117] Bruno de Finetti. *Probabilità e Induzione; Probability and Induction*. Biblioteca de STATISTICA, CLUEB (Cooperative Libraria Universitaria Editrice Bologna), Bologna, 1993. A collection of Italian articles, all with English translations by Mara Khale and Antonella Ansani. The collection is introduced by review essays by Richard Jeffrey and Dario Fürst. 97

[118] Abraham De Moivre. *The Doctrine of Chances: or, A Method of Calculating the Probabilities of Events in Play*. Pearson, London, 1718. Second edition 1738, third 1756. 10

[119] Augustus De Morgan. A treatise on the theory of probabilities. In Edward Smedley, Hugh James Rose, and Henry John Rose, editors, *Encyclopaedia Metropolitana*, volume 2, pages 393–490. Griffin, London, 1837. 1849 edition available at https://archive.org/details/encyclopaediamet02smeduoft. 11

[120] Augustus De Morgan. *An Essay on Probabilities, and on their application to Life Contingencies and Insurance Offices*. Longman, Orme, Brown, Green & Longmans, London, 1838. https://play.google.com/books/reader?id=NtA3AAAAMAAJ&printsec=frontcover&output=reader&hl=en. 11

[121] Arthur P. Dempster. *Elements of Continuous Multivariate Analysis*. Addison-Wesley, Reading, Massachusetts, 1969. 98

[122] Edward L. Dodd. The use of linear functions to detect hidden periods in data separated into small sets. *Annals of Mathematical Statistics*, 1:205–223, 1930. https://projecteuclid.org/download/pdf_1/euclid.aoms/1177733089. 63

[123] Joseph L. Doob. Stochastic processes and statistics. *Proceedings of the National Academy of Sciences of the United States*, 20:376–379, 1934. http://www.pnas.org/content/20/6/376.full.pdf. 78, 87

[124] Joseph L. Doob. The limiting distributions of certain statistics. *Annals of Mathematical Statistics*, 6:160–169, 1935. https://projecteuclid.org/download/pdf_1/euclid.aoms/1177732594. 63

[125] Joseph L. Doob. Application of the theory of martingales. In *Le Calcul des Probabilités et ses Applications*, Colloques Internationaux, pages 23–27. Centre National de la Recherche Scientifique, Paris, 1949. 87

[126] Joseph L. Doob. Continuous parameter martingales. In Jerzy Neyman, editor, *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pages 269–277. University of California Press, Berkeley, 1951. https://projecteuclid.org/download/pdf_1/euclid.bsmsp/1200500234. 87

[127] Joseph L. Doob. *Stochastic Processes*. Wiley, New York, 1953. 4, 87

[128] Louis-Gustave Du Pasquier. *Le calcul des probabilités, son évolution mathématique et philsophique*. Hermann, Paris, 1926. 61

[129] Didier Dubois and Henri Prade. Possibility theory, probability theory and multiple-valued logics: A clarification. *Annals of Mathematics and Artificial Intelligence*, 32:35–66, 2001.

[130] Francis Edgeworth. The law of error. *Philosophical Magazine*, 16:300–309, 1883. 54

[131] Francis Edgeworth. The philosophy of chance. *Mind*, 9(34):223–235, 1884. 54

[132] Francis Edgeworth. The empirical proof of the law of error. *Philosophical Magazine*, 24:330–342, 1887. 54

[133] Francis Edgeworth. Correlated averages. *Philosophical Magazine*, 34:190–204, 1892. 54

[134] Francis Edgeworth. On the probable errors of frequency-constants. *Journal of the Royal Statistical Society*, 71:381–397, 499–512, 651–668, 1908. 96

[135] Francis Edgeworth. Probability. *Encyclopædia Brittanica, Eleventh Edition*, 22:376–403, 1911. 92

[136] Gustav Theodor Fechner. *Kollektivmasslehre*. Engelmann, Leipzig, 1897. Edited by G. F. Lipps. https://archive.org/details/kollektivmassle00fechgoog. 54

[137] William Feller. Review of [219]. *Zentralblatt für Mathematik und ihre Grenzegebiete*, 7:216, 1934. http://gdz.sub.uni-goettingen.de/dms/load/img/?PPN=PPN245319514_0007&DMDID=DMDLOG_0047&LOGID=LOG_0047&PHYSID=PHYS_0205. 76

[138] William Feller. Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, 40:521–559, 1935. http://www.digizeitschriften.de/download/PPN266833020_0040/log44.pdf. 77

[139] William Feller. Zur Theorie der stochastischen Prozesse (Existenz- und Eindeutigkeitssätze). *Mathematische Annalen*, 113:113–160,

1936. http://www.digizeitschriften.de/download/PPN235181684_0113/log12.pdf. 77, 84, 88, 100

[140] William Feller. Sur les axiomatiques du calcul des probabilités et leurs relations avec les expériences. In Wavre [375], pages 7–21 of the second fascicle, number 735 *Les fondements du calcul des probabilités*. 83

[141] William Feller. On the integro-differential equations of purely discontinuous Markoff processes. *Transactions of the American Mathematical Society*, 48:488–515, 1940. http://www.ams.org/journals/tran/1940-048-03/S0002-9947-1940-0002697-3/S0002-9947-1940-0002697-3.pdf. 87

[142] William Feller. The general form of the so-called law of the iterated logarithm. *Transactions of the American Mathematical Society*, 54:373–402, 1943. http://www.ams.org/journals/tran/1943-054-03/S0002-9947-1943-0009263-7/S0002-9947-1943-0009263-7.pdf. 87

[143] William Feller. Generalization of a probability limit theorem of Cramér. *Transactions of the American Mathematical Society*, 54:361–372, 1943. http://www.ams.org/journals/tran/1943-054-03/S0002-9947-1943-0009262-5/S0002-9947-1943-0009262-5.pdf. 87

[144] William Feller. On a general class of "contagious distributions". *Annals of Mathematical Statistics*, 14:389–400, 1943. https://projecteuclid.org/download/pdf_1/euclid.aoms/1177731359. 87

[145] William Feller. *An Introduction to Probability Theory and Its Applications*, volume 1. Wiley, New York, 1950. Second edition 1957, third edition 1968. 53, 87

[146] William Feller. *An Introduction to Probability Theory and Its Applications*, volume 2. Wiley, New York, 1966. Second edition 1971. 87

[147] Hans Fischer. *A History of the Central Limit Theorem from Classical to Modern Probability Theory*. Springer, New York, 2011. 7, 15, 18, 21, 30, 40, 41, 44, 49, 54, 56, 109

[148] Arne Fisher. *The Mathematical Theory of Probabilities and Its Application to Frequency Curves and Statistical Methods*. Macmillan, New York, 1915. The first edition was "translated and edited from the author's original Danish notes with the assistance of William Bonynge". Second edition 1922 at https://play.google.com/books/reader?id=p7Vv3mtT9XMC&printsec=frontcover&output=reader&hl=en&pg=GBS.PP1. 12, 52, 58

[149] Ronald Aylmer Fisher. The distribution of the partial correlation coefficient. *Metron*, 3:329–333, 1924. 57

[150] Ronald Aylmer Fisher. *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburgh, first edition, 1925. Thirteenth edition 1958. 56

[151] Maurice Fréchet. Sur la convergence probable. *Comptes rendus hebdomadaires des séances de l'Académie des Sciences*, 188:213–214, 1929. Session of 14 January. http://gallica.bnf.fr/ark:/12148/bpt6k31417/f213.image.langEN. 72

[152] Maurice Fréchet. Sur la distance de deux variables aléatoires. *Comptes rendus hebdomadaires des séances de l'Académie des Sciences*, 188:368–

370, 1929. Session of 28 January. `http://gallica.bnf.fr/ark:/12148/bpt6k31417/f368.image.langEN`. 72

[153] Maurice Fréchet. Sur la convergence "en probabilité". *Metron*, 8:3–50, 1930. 72

[154] Maurice Fréchet. *Généralités sur les probabilités. Variables aléatoires.* Gauthier-Villars, Paris, 1937. This is Book 1 of Fréchet (1937–1938). The second edition (1950) has a slightly different title: *Généralités sur les probabilités. Éléments aléatoires.* 53

[155] Maurice Fréchet. *Recherches théoriques modernes sur la théorie des probabilités.* Gauthier-Villars, Paris, 1937–1938. This work is listed in the bibliography of the *Grundbegriffe* as in preparation. It consists of two books, Fréchet (1937) and Fréchet (1938a). The two books together constitute Fascicle 3 of Volume 1 of Émile Borel's *Traité du calcul des probabilités et ses applications.*

[156] Maurice Fréchet. Exposé et discussion de quelques recherches récentes sur les fondements du calcul des probabilités. In Wavre [375], pages 23–55 of the second fascicle, number 735, *Les fondements du calcul des probabilités.* 83

[157] Maurice Fréchet. *Les mathématiques et le concret.* Presses Universitaires de France, Paris, 1955. 72

[158] Maurice Fréchet and Maurice Halbwachs. *Le calcul des probabilités à la portée de tous.* Dunod, Paris, 1924. 69

[159] Ragnar Frisch. A method of decomposing an empirical series into its cyclical and progressive components. *Journal of the American Statistical Association, Supplement: Proceedings of the American Statistical Association*, 26:73–78, 1931. 80

[160] Thornton C. Fry. *Probability and Its Engineering Uses.* Van Nostrand, New York, 1928. 34, 52, 64

[161] Maria Carla Galavotti. *Bruno de Finetti. Radical Probabilist.* College Publications, London, 2008. 97

[162] Thomas Galloway. *A Treatise on Probability, forming the article under that head in the seventh edition of the Encyclopædia Britannica.* Black, North Bridge, 1839. 11

[163] Francis Galton. Co-relations and their measurement, chiefly from anthropometric data. *Proceedings of the Royal Society of London*, 45:135–145, 1888. 54

[164] Francis Galton. Note to the memoir by Professor Karl Pearson, F.R.S., on spurious correlation. *Proceedings of the Royal Society of London*, 60:489–502, 1897. `http://rspl.royalsocietypublishing.org/content/60/359-367/498.full.pdf+html`. 55

[165] Francis Galton. *Probability, the Foundation of Eugenics.* Oxford Clarendon Press, London, 1907. `www.galton.org`. 57

[166] Corrado Gini. *Variabilità e mutabilità: contributo allo studio delle distribuzioni e delle relazioni statistiche.* Cuppini, Bologna, 1912. Studi Economico-Giurdici della Facoltà di Giurisprudenza della Regia Università di Cagliari, 3 (part 2). 60

[167] Hugh Godfray. At random. In Miller [266], pages 65–67. https://play.google.com/store/books/details?id=QU4hAQAAIAAJ&rdid=book-QU4hAQAAIAAJ&rdot=1. 32

[168] Evgenii Alekseevich Godin. Развитие идей Московской философско-математической сколы *(Development of the ideas of the Moscow philosophical and mathematical school).* Красный свет, Moscow, 2006. ISBN 5-902967-05-8. 40

[169] Prakash Gorroochurn. *Classic Problems of Probability.* Wiley, Hoboken, 2012. 18

[170] Loren Graham and Jean-Michel Kantor. *Naming Infinity: A True Story of Religious Mysticism and Mathematical Creativity.* Belknap, Harvard, Cambridge, London, 2009. 37, 40

[171] Ivor Grattan-Guinness. A note on *The Educational Times* and *Mathematical Questions*. *Historia Mathematica*, 19:76–78, 1992. 32

[172] Major Greenwood. Professor Tschouprow on the theory of correlation. *Journal of the Royal Statistical Society*, 89:320–325, 1926. 65

[173] Major Greenwood. The statistical study of infectious diseases. *Journal of the Royal Statistical Society*, 109(2):85–110, 1946. 51

[174] Alf Guldberg. Ueber Markoffs Ungleichung. *Metron*, 3:3–5, 1923. 68

[175] Alf Guldberg. Les fonctions de fréquence discontinues et les séries statistiques. *Annales de l'Institut Henri Poincaré*, 3:229–278, 1933. http://www.numdam.org/item?id=AIHP_1933__3_3_229_0. 74, 84

[176] Peter Guttorp and Georg Lindgren. Karl Pearson and the Scandinavian school of statistics. *International Statistical Review*, 77:64–71, 2009. 54, 80

[177] K. G. Hagström. Der Begriff der statistischen Funktion. *Skandinavisk Aktuarietidskrift*, 2:1–52, 1919. 84

[178] Anders Hald. *A History of Mathematical Statistics from 1750 to 1930.* Wiley, New York, 1998. 13

[179] Wolfgang Karl Härdle and Annette B. Vogt. Ladislaus von Bortkiewicz—Statistician, economist and a European intellectual. *International Statistical Review*, 83(1):17–35, 2015. 47

[180] Carl Friedrich Hauber. Über die Bestimmung der Genauigkeit der Beobachtungen. *Zeitschrift für Physik und Mathematik*, 7:286–314, 1829. https://ia801407.us.archive.org/20/items/zeitschriftfurp00ettigoog/zeitschriftfurp00ettigoog.pdf. 20

[181] Carl Friedrich Hauber. Verallgermeinerung der Poisson'schen Untersuchungen über die Wahrscheinlichkeit der mittlern Resultate der Beobachtungen in den "Additions à la *Connaiss. des Tems* de 1827". *Zeitschrift für Physik und Mathematik*, 7:406–429, 1829. https://ia801407.us.archive.org/20/items/zeitschriftfurp00ettigoog/zeitschriftfurp00ettigoog.pdf. 20

[182] Carl Friedrich Hauber. Theorie der mittleren Werthe. *Zeitschrift für Physik und Mathematik*, 8:25–56, 1830. Continued on pages 147–179, 295–316, 443–455, then on pages 302–322 of Volume 9 (1831) and pages 425–457 of Volume 10 (1832). http://reader.digitale-sammlungen.

de/de/fs1/object/display/bsb10130666_00037.html, https://play.google.com/books/reader?id=GA85AAAAcAAJ&printsec=frontcover&output=reader&hl=en&pg=GBS.PR1. 20

[183] Felix Hausdorff. Beiträge zur Wahrscheinlichkeitsrechnung. *Berichte der Königlich Sächsischen Gesellschaft der Wissenschaften zu Leipzig, Mathematisch-Physische Klasse*, 53:152–178, 1901. http://babel.hathitrust.org/cgi/pt?id=mdp.39015049033569;view=1up;seq=228. 49

[184] Veronika Havlová, Laurent Mazliak, and Pavel Sisma. Le début des relations mathématiques franco-tchécoslovaques vu à travers la correspondance Hostinský-Fréchet. *Electronic Journal for History of Probability and Statistics*, 1, 2005. http://www.jehps.net/Mars2005/HavlovaMazliakSisma.pdf. 72

[185] Christopher C. Heyde and Eugene Seneta. *I. J. Bienaymé: Statistical theory anticipated*. Springer, New York, 1977. 20, 29

[186] Jørgen Hoffmann-Jørgensen. The general marginal problem. In Svetozar Kurepa, Hrvoje Kraljević, and Davor Butković, editors, *Functional Analysis II*, volume 1242 of *Lecture Notes in Mathematics*, pages 77–367. Springer, Berlin, 1987. 5

[187] Bohuslav Hostinský. *Sur les transformations itérées des variables aléatoires*. Publications de la Faculté des Sciences de l'Université Masaryck, No. 93, Brno, 1928. 24 pages. 71

[188] Bohuslav Hostinský. Sur les probabilités des phénomènes liés en chaîne de Markoff. *Comptes rendus hebdomadaires des séances de l'Académie des Sciences*, 189:78–80, 1929. Session of 8 July. http://gallica.bnf.fr/ark:/12148/bpt6k3142j/f78.image.langEN. 71

[189] Bohuslav Hostinský. *Méthodes générales du calcul des probabilités*. Gauthier-Villars, Paris, 1931. 71

[190] Harold Hotelling. Differential equations subject to error, and population estimates. *Journal of the American Statistical Association*, 22:283–314, 1927. 63, 79

[191] Harold Hotelling. New light on the correlation coefficient and its transforms. *Journal of the Royal Statistical Society. Series B (Methodological)*, 15:193–232, 1953. 83

[192] Harold Hotelling and Margaret Richards Pabst. Rank correlation and tests of significance involving no assumption of normality. *Annals of Mathematical Statistics*, 7:29–43, 1936. 83

[193] Patti Wilger Hunter. An unofficial community: American mathematical statisticians before 1935. *Annals of Science*, 56:47–68, 1999. 63

[194] Joseph Oscar Irwin. William Allen Whitworth and a hundred years of probability. *Journal of the Royal Statistical Society. Series A*, 130:147–176, 1967. 52

[195] Harold Jeffreys. *Scientific Inference*. Cambridge University Press, Cambridge, 1931. Second edition 1957, third 1973. 97

[196] Harold Jeffreys. *Theory of Probability*. Oxford University Press, Oxford, 1939. Second edition 1948, third 1961. 97

[197] William Stanley Jevons. *Principles of Science*. Macmillan, London, 1874. 89

[198] Charles Jordan. *Statistique mathématique*. Gauthier-Villars, Paris, 1927. Also published in the same year in Hungarian, in Budapest, as *Matematikai statisztika* by Jordán Károly, OCLC 909440965. 70

[199] Maurice G. Kendall and William R. Buckland. *A Dictionary of Statistical Terms*. Oliver and Boyd and Hafner, London and New York, 1957. 83

[200] John Maynard Keynes. *A Treatise on Probability*. Macmillan, London, 1921. 51, 94

[201] Aleksandr Yakovlevich Khinchin. Über das Gesetz der großen Zahlen. *Mathematische Annalen*, 96:152–168, 1927. `http://eudml.org/doc/159161`. 74

[202] Aleksandr Yakovlevich Khinchin. Sur la loi forte des grands nombres. *Comptes rendus hebdomadaires des séances de l'Académie des Sciences*, 186:285–287, 1928. Session of 30 January. `http://gallica.bnf.fr/ark:/12148/bpt6k3139f/f285.image.langEN`. 73

[203] Aleksandr Yakovlevich Khinchin. Sur la loi des grands nombres. *Comptes rendus hebdomadaires des séances de l'Académie des Sciences*, 188:477–479, 1929. Session of 11 February. `http://gallica.bnf.fr/ark:/12148/bpt6k31417/f477.image.langEN`. 73

[204] Aleksandr Yakovlevich Khinchin. Sur une généralisation de quelques formules classiques. *Comptes rendus hebdomadaires des séances de l'Académie des Sciences*, 188:532–534, 1929. Session of 18 February. `http://gallica.bnf.fr/ark:/12148/bpt6k31417/f532.image.langEN`. 73

[205] Aleksandr Yakovlevich Khinchin. Über Anwendbarkeitskriterien für das Gesetz der grossen Zahlen. Математический Сборник, 36(1):78–80, 1929. `http://mi.mathnet.ru/msb7342`. 74

[206] Aleksandr Yakovlevich Khinchin. Über die positiven und negativen Abweichungen des arithmetischen Mittels. *Mathematische Annalen*, 101:381–385, 1929. `http://eudml.org/doc/159342`. 74

[207] Aleksandr Yakovlevich Khinchin. Über ein Problem der Wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, 29:746–752, 1929. `http://eudml.org/doc/168102`. 74

[208] Aleksandr Yakovlevich Khinchin. Über einen neuen Grenzwertsatz der Wahrscheinlichkeitsrechnung. *Mathematische Annalen*, 101:745–752, 1929. `http://eudml.org/doc/159367`. 74

[209] Aleksandr Yakovlevich Khinchin. Asymptotische Gesetze der Wahrscheinlichkeitsrechnung. In *Ergebnisse der Mathematik*, volume 2, pages 263–345. Springer (for *Zentralblatt für Mathematik*), Berlin, 1933. 74, 76

[210] Aleksandr Yakovlevich Khinchin. Über stationäre Reihen zufälliger Variablen. Математический Сборник, 40:124–128, 1933. `http://mi.mathnet.ru/msb6772`. 74

[211] Aleksandr Yakovlevich Khinchin. Zur mathematischen Begründung der statistichen Mechanik. *Zeitschrift für Angewandte Mathematik und Mechanik*, 13:101–103, 1933. 74

[212] Aleksandr Yakovlevich Khinchin. Korrelationstheorie der stationären stochastischen Prozesse. *Mathematische Annalen*, 109:604–615, 1934. http://eudml.org/doc/159698. 74, 78

[213] Aleksandr Yakovlevich Khinchin and Andrei Nikolaevich Kolmogorov. Über Konvergenz von Reihen, deren Glieder durch den Zufall bestimmt werden. Математический сборник. *(Sbornik: Mathematics)*, 32:668–677, 1925. http://mi.mathnet.ru/msb7426. Translated into Russian on pp. 7–16 of [223] and thence into English on pp. 1–10 of [224]. 75

[214] Eberhard Knobloch. Historical aspects of the foundations of error theory. In J. Echeverria, A. Ibarra, and Th. Mormann, editors, *The Space of Mathematics*, pages 253–279. Gruyter, Berlin–New York, 1992. 21

[215] Andrei Nikolaevich Kolmogorov. Über die Summen durch den Zufall bestimmter unabhängiger Grössen. *Mathematische Annalen*, 99:309–319, 1928. http://www.digizeitschriften.de/download/PPN235181684_0099/log19.pdf. An addendum appears in 1929: Volume 102, pp. 484–488. The article and the addendum are translated into Russian on pp. 20–34 of [223] and thence into English on pp. 15–31 of [224]. 75

[216] Andrei Nikolaevich Kolmogorov. Über das Gesetz der iterierten Logarithmus. *Mathematische Annalen*, 101:126–135, 1929. http://www.digizeitschriften.de/download/PPN235181684_0101/log9.pdf. Appears in English in [224], pp. 32–42. 75

[217] Andrei Nikolaevich Kolmogorov. Sur la loi forte des grands nombres. *Comptes rendus hebdomadaires des séances de l'Académie des Sciences*, 191:910–912, 1930. Session of 17 November. http://gallica.bnf.fr/ark:/12148/bpt6k31445/f910.image.langEN. Translated into Russian on pp. 44–47 of [223] and thence into English on pp. 60–61 of [224]. 75

[218] Andrei Nikolaevich Kolmogorov. Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung. *Mathematische Annalen*, 104:415–458, 1931. http://www.digizeitschriften.de/download/PPN235181684_0104/log33.pdf. Translated into Russian on pp. 60–105 of [222] and thence into English on pp. 62–108 of [224]. 75, 78

[219] Andrei Nikolaevich Kolmogorov. Grundbegriffe der Wahrscheinlichkeitsrechnung. In *Ergebnisse der Mathematik*, volume 2, pages 195–262. Springer (for *Zentralblatt für Mathematik*), Berlin, 1933. A Russian translation, by Grigory M. Bavli, appeared under the title Основные понятия теории вероятностей (Nauka, Moscow) in 1936, with a second edition, slightly expanded by Kolmogorov with the assistance of Albert N. Shiryaev, in 1974, and a third edition (ФАЗИС, Moscow) in 1998. An English translation by Nathan Morrison appeared under the title *Foundations of the Theory of Probability* (Chelsea, New York) in 1950, with a second edition in 1956. 4, 75, 111

[220] Andrei Nikolaevich Kolmogorov. Zur Theorie der stetigen zufälligen Prozesse. *Mathematische Annalen*, 108:149–160, 1933. http://www.digizeitschriften.de/download/PPN235181684_0108/log14.pdf. 75, 78

[221] Andrei Nikolaevich Kolmogorov. Вероятность (Probability). In Большая Советская Энциклопедия *(Great Soviet Encyclopedia)*, volume 7, pages 508–510. Soviet Encyclopedia Publishing House, Moscow, second edition, 1951. The same article appears on p. 544 of Vol. 4 of the third edition of the encyclopedia, published in 1971, and in a subsidiary work, the Математическая энциклопедия, published in 1987. English translations of both encyclopedias exist. 90, 96

[222] Andrei Nikolaevich Kolmogorov. Избранные труды. Математика и механика. Nauka, Moscow, 1985. 117

[223] Andrei Nikolaevich Kolmogorov. Избранные труды. Теория вероятностей и математическая статистика. Nauka, Moscow, 1986. 117, 118

[224] Andrei Nikolaevich Kolmogorov. *Selected Works of A. N. Kolmogorov. Volume II: Probability Theory and Mathematical Statistics.* Kluwer, Dordrecht, 1992. Translation of [223] from the Russian. 117

[225] Henry E. Kyburg Jr. and Howard E. Smokler, editors. *Studies in Subjective Probability.* Wiley, New York, 1964. A second edition, with a slightly different selection of readings, was published by Krieger, New York, in 1980. 97, 110

[226] Sylvestre François Lacroix. *Traité élémentaire du calcul des probabilités.* Courcier, Paris, 1816. Second edition 1822. 12, 24, 118

[227] Sylvestre François Lacroix. *Lehrbuch der Wahrscheinlichkeitsrechnung.* Keyser, Erfurt, 1818. Translation of [226] by Ephraim Salomon Unger. `http://www.mdz-nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:bvb:12-bsb10082213-0`. 11, 13

[228] Joseph-Louis Lagrange. Mémoire sur l'utilité de la méthode de prendre le milieu entre les résultats de plusieurs observations, dans lequel on examine les avatages de cette méthode par le calcul des probabilités, et ou l'on résout différents problèmes relatifs à cette matière. *Miscellanea Taurinesia*, 5:167–232, 1776. Reprinted in Volume 2 of *Oeuvres de Lagrange*, pages 173–236, `http://gallica.bnf.fr/ark:/12148/bpt6k215570z/f174`. 14

[229] Léon Lalanne. De l'emploi de la géométrie pour résoudre certaines questions de moyennes et de probabilités. *Journal de Mathḿatiques Pures et Appliqués*, pages 107–130, 1879. `http://eudml.org/doc/234323`. 34

[230] Pierre Simon de Laplace. Recherches sur l'intégration des équations différentielles aux différences finies et sur leur usage dans la théorie des hasards. *Mémoires de l'Académie royale des sciences de Paris 1773*, pages 37–163, 1776. Reprinted on pages 69–197 of [236]. 11

[231] Pierre Simon de Laplace. Mémoire sur les probabilités. *Mémoires de l'Académie royale des sciences de Paris 1778*, pages 227–332, 1781. Reprinted on pages 383–485 of [237]. 14

[232] Pierre Simon de Laplace. *Théorie analytique des probabilités.* Courcier, Paris, first edition, 1812. Second edition 1814, third 1820. Third edition reprinted in [235]. 11, 14, 31

[233] Pierre Simon de Laplace. *Essai philosophique sur les probabilités.* Courcier, Paris, first edition, 1814. The fifth and definitive edition was published

in 1825. A modern edition with commentary was published byChristian Bourgoi s, Paris, in 1986. An English translation was published in 1994 [?]. 119

[234] Pierre Simon de Laplace. *Philosophische Versuch über Wahrscheinlichkeiten*. Translation of the third edition of [233] by Friederich Wilhelm Tönnies, Heidelberg, 1819. 11

[235] Pierre Simon de Laplace. *Œuvres complètes de Laplace, Volume VII*. Gauthier-Villars, Paris, 1886. 118

[236] Pierre Simon de Laplace. *Œuvres complètes de Laplace, Volume VIII*. Gauthier-Villars, Paris, 1891. 11, 118

[237] Pierre Simon de Laplace. *Œuvres complètes de Laplace, Volume IX*. Gauthier-Villars, Paris, 1893. 14, 118

[238] Hermann Laurent. *Traité du calcul des probabilités*. Gauthier-Villars, Paris, 1873. 21, 26, 94

[239] Steffen L. Lauritzen. *Thiele. Pioneer in Statistics*. Oxford University Press, Oxford, 2002. 54

[240] Erich L. Lehmann. *Fisher, Neyman, and the Creation of Classical Statistics*. Springer, New York, 2011. 82, 83

[241] Paul Lévy. Sur la détermination des lois de probabilité par leurs fonctions caractéristiques. *Comptes rendus hebdomadaires des séances de l'Académie des Sciences*, 175:854–856, 1922. Session of 13 November. `http://http://gallica.bnf.fr/ark:/12148/bpt6k3128v/f854.image`. 50

[242] Paul Lévy. *Calcul de probabilités*. Gauthier-Villars, Paris, 1925. This pathbreaking book includes the first theory of stable distributions. The first few chapters provide an eclectic understanding of the subjective and objective sides of probability. An appendix discusses the laws of probability in abstract spaces. 53, 62

[243] Paul Lévy. Sur les séries dont les termes sont des variables éventuelles indépendantes. *Studia Mathematica*, III:119–155, 1931. `http://matwbn.icm.edu.pl/ksiazki/sm/sm3/sm318.pdf`. 50, 74

[244] Paul Lévy. Sur quelques questions de calcul des probabilités. *Prace Matematyczno-Fizyczne*, 39(1):19–28, 1932. `http://eudml.org/doc/215494`. 74

[245] Paul Lévy. *Théorie de l'addition des variables aléatoires*. Gauthier-Villars, Paris, 1937. Second edition: 1954. 53, 89

[246] Wilhelm Lexis. *Zur Theorie der Massenerscheinungen in der Menschlichen Gesellschaft*. Wagner, Freiburg, 1877. `https://ia801408.us.archive.org/10/items/zurtheoriederma00lexigoog/zurtheoriederma00lexigoog.pdf`. 54

[247] Wilhelm Lexis. Über die Wahrscheinlichkeitsrechnung und deren Anwendung auf die Statistik. *Jahrbücher für Nationalökonomie und Statistik*, 13 (47)(5), 1886. 52, 54, 101

[248] Jean-Baptiste-Joseph Liagre. *Calcul des probabilités et théorie des erreurs avec des applications aux sciences d'observation en général et à la*

géodésie. Muquardt, Brussels, 1852. Second edition, 1879, prepared with the assistance of Camille Peny. 26

[249] Aleksander Mikhailovich Liapunov. Sur une proposition de la théorie des probabilités. *Bulletin de l'Académie Impériale des Sciences*, 13:359–386, 1900. http://mi.mathnet.ru/eng/izv4765. English translation in [1], pages 151–171. 42

[250] Aleksander Mikhailovich Liapunov. Nouvelle forme du théorème sur la limite de probabilité. *Mémoires de l'Académie Impériale des Sciences de St.-Pétersbourg VIIIe Série, Class Physico-Mathématique*, 12:1–24, 1901. English translation in [1], pages 175–191. 42

[251] Aleksander Mikhailovich Liapunov. Sur un théoreme du calcul des probabilités. *Comptes rendus hebdomadaires des séances de l'Académie des Sciences*, 132:126–128, 1901. Session of 12 January. http://gallica.bnf.fr/ark:/12148/bpt6k30888/f138.image. English translation in [1], pages 149–150. 42

[252] Aleksander Mikhailovich Liapunov. Une propostion générale du calcul des probabilités. *Comptes rendus hebdomadaires des séances de l'Académie des Sciences*, 132:814–815, 1901. Session of 1 April. http://gallica.bnf.fr/ark:/12148/bpt6k30888/f866.image. English translation in [1], pages 173–174. 42

[253] Bernard Locker. Doob at Lyon. *Electronic Journal for History of Probability and Statistics*, 2009. http://www.jehps.net/juin2009/Locker.pdf. 87

[254] Zbigniew Łomnicki and Stanisław Ulam. Sur la théorie de la mesure dans les espaces combinatoires et son application au calcul des probabilités. I. Variables indépendantes. *Fundamenta Mathematicae*, 23:237–278, 1934. http://matwbn.icm.edu.pl/ksiazki/fm/fm23/fm23121.pdf. 62

[255] Donald A. MacKenzie. *Statistics in Britain 1865–1930*. Edinburgh University Press, Edinburgh, 1981. 55

[256] Leonid E. Maistrov. *Probability Theory: A Historical Sketch*. Academic Press, New York, 1974. Translated and edited by Samuel Kotz. 12, 43

[257] Andrei Andreevich Markov. закон болших чисел и способ наименших квадратов. Известия физ.-мат. общества Казан унив., 8(2):110–128, 1899. English translation at www.sheynin.de. 43

[258] Andrei Andreevich Markov. Исчисление вероятностей *(Calculus of Probability)*. Типография Императорской Академии Наук, Saint Petersburg, 1900. Second edition 1908, fourth 1924. 43, 49, 120, 123, 126

[259] Andrei Andreevich Markov. о связанных величинах не образующих настоящеи цепи. *Bulletin de l'Académie Impériale des Sciences*, 5:113–126, 1911. http://mi.mathnet.ru/izv6867. German translation in [260], pages 298–311. 44

[260] Andrei Andreevich Markov. *Wahrscheinlichkeitsrechnung*. Teubner, Leipzig, 1912. Translation of second edition of [258]. https://ia902306.us.archive.org/34/items/wahrscheinlich00markrich/wahrscheinlich00markrich.pdf. 44, 50, 120

[261] Andrei Andreevich Markov. Избранные Труды, Теория чисел: Теория верояноцтей *(Selected works: Number theory, probability theory)*. Издательство Ададемии Наук СССР, Moscow, 1951. 43

[262] Mansfield Merriman. Least squares: A list of writings relating to the method, with historical and critical notes. *Transactions of the Connecticut Academy of Arts and Sciences*, 4:151–232, 1877. http://babel.hathitrust.org/cgi/pt?id=hvd.32044091870683;view=1up;seq=1. 21, 92, 94

[263] Mansfield Merriman. *A Text-Book on the Method of Least Squares*. Wiley, New York, eighth edition, 1911. https://ia802205.us.archive.org/16/items/atextbookonmeth03merrgoog/atextbookonmeth03merrgoog.pdf.

[264] Antoine Meyer. *Vorlesung über Wahrscheinlichkeitsrechnung*. Leipzig, Teubner, 1879. Translated by Emanuel Czuber. 45, 46

[265] Wilhelm Franz Meyer, editor. *Encyklopädie der mathematischen Wissenschaften, Band I, Teil 2*. Teubner, Leipzig, 1900–1904. 52

[266] W. J. Miller, editor. *Mathematical Questions, with their Solutions. From the "Educational Times." With many Papers and Solutions not published in the "Educational Times"*, volume VII. Hodgson, London, 1867. https://play.google.com/store/books/details?id=QU4hAQAAIAAJ&rdid=book-QU4hAQAAIAAJ&rdot=1. 32, 109, 114

[267] Pierre de Montmort. *Essay d'analyse sur les jeux de hazard*. Quillau, Paris, 1708. Second edition 1713. 10

[268] Pavel Alekseevich Nekrasov. Propriétés générales des phénomènes nombreux indépendants dans leur rapport avec le calcul approché des fonctions des grands nombres (in Russian). Математический Сборник, 20(3):431–442, 1898. http://mi.mathnet.ru/msb8005. English translation at www.sheynin.de. 42

[269] Pavel Alekseevich Nekrasov. Московская философско-математическая школа и ея основатели. Universitetskaya tipografiya, Moscow, 1904. http://babel.hathitrust.org/cgi/pt?id=hvd.hnyi6a;view=1up;seq=9. Also published in the first issue of Volume 25 of Математический Сборник, 1904. 39

[270] Jerzy Neyman. Contributions to the theory of small samples drawn from a finite population. *Biometrika*, 17:472–479, 1925. 82

[271] Jerzy Neyman. Further notes on non-linear regression. *Biometrika*, 18:257–262, 1926. 82

[272] Jerzy Neyman. Méthodes nouvelles de vérification des hypothèses. In *Comptes Rendus du 1 Congrès de Mathématiciens des Pays Slavs*, pages 355–366, Warsaw, 1929. 62, 82

[273] Jerzy Neyman. On the problem of confidence intervals. *Annals of Mathematical Statistics*, 6:111–116, 1935. https://projecteuclid.org/download/pdf_1/euclid.aoms/1177732585. 82

[274] Jerzy Neyman. Sur la vérification des hypothèses statistiques composés. *Bulletin de la Société Mathématique de France*, 63:246–266,

1935. http://archive.numdam.org/ARCHIVE/BSMF/BSMF_1935__63_/BSMF_1935__63__246_0/BSMF_1935__63__246_0.pdf. 82

[275] Jerzy Neyman. Outline of a theory of statistical estimation based on the classical theory of probability. *Philosophical Transactions Royal Society of London, Series A*, 236:333–380, 1937. http://rsta.royalsocietypublishing.org/content/roypta/236/767/333.full.pdf. 82, 88, 89

[276] Jerzy Neyman. 'Smooth' test for goodness of fit. *Skandinavisk Aktuarietidskrift*, 20:149–199, 1937. 82

[277] Jerzy Neyman. L'estimation statistique, traitée comme un problème classique de probabilités. In Wavre [375], pages 25–57 of the sixth fascicle, number 739, *Conceptions diverses*. 83

[278] Jerzy Neyman. Tests of statistical hypotheses which are unbiased in the limit. *Annals of Mathematical Statistics*, 9:69–86, 1938. https://projecteuclid.org/download/pdf_1/euclid.aoms/1177732329. 85

[279] Jerzy Neyman. Basic ideas and some recent results of the theory of testing statistical hypotheses. In Wavre [376], pages 81–127. Because of the war, these proceedings did not appear until 1946. 88, 97

[280] Jerzy Neyman, editor. *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press, Berkeley and Los Angeles, 1949. https://projecteuclid.org/euclid.bsmsp/1166219193/ Held at the Statistical Laboratory, Department of Mathematics, University of California, August 13–18, 1945, January 27-29, 1946. 87

[281] Jerzy Neyman. Indeterminism in science and new demands on statisticians. *Journal of the American Statistical Association*, 55:625–639, 1960. 99

[282] Jerzy Neyman. *A Selection of Early Statistical Papers of J. Neyman*. University of California Press, Berkeley and Los Angeles, 1967. 82

[283] Jerzy Neyman, K. Iwaszkiewicz, and St. Kolodziejczyk. Statistical problems in agricultural experimentation. *Supplement to the Journal of the Royal Statistical Society*, 2:107–180, 1935. 80

[284] Jerzy Neyman and Egon S. Pearson. Contributions to the theory of testing statistical hypotheses. *Statistical Research Memoirs*, 1:1–17, 1936. 82

[285] Jerzy Neyman and Egon S. Pearson. *Joint Statistical Papers*. University of California Press, Berkeley and Los Angeles, 1967. 82

[286] Ludwig Oettinger. *Die Wahrscheinlichkeits-Rechnung*. Reimer, Berlin, 1852. The contents had already been published in five installments in the *Journal für die reine und angewandte Mathematik* over the period from 1843 to 1848. In the preface, the author states that he had already written the material in 1842 http://babel.hathitrust.org/cgi/pt?id=hvd.32044091868638;view=1up;seq=7. 11

[287] Kh. O. Ondar, editor. О теории вероятностей и математической статистике (переписка А. А. Маркова и А. А. Чупрова). Nauk, Moscow, 1977. See [288] for English translation. 44, 66, 122

[288] Kh. O. Ondar, editor. *The Correspondence Between A. A. Markov and A. A. Chuprov on the Theory of Probability and Mathematical Statistics.* Springer, New York, 1981. Translation of [287] by Charles M. and Margaret Stein. Additional letters between Markov are provided in translation by Sheynin in [341], Chapter 8. 40, 43, 66, 122, 124

[289] Nell Irvin Painter. *The History of White People.* Norton, New York, 2010. 55

[290] G. Papelier. Review of [258]. *L'Enseignement Mathématique*, 3(1):229–230, 1901. `http://retro.seals.ch/digbib/view2?pid=ensmat-001:1901:3::443`. 44

[291] Karl Pearson. *The Grammar of Science.* Walter Scott, London, 1892. Second edition in 1900 and third edition in 1911 published by Adam & Charles Black. 54

[292] Karl Pearson. On a form of spurious correlation which may arise when indices are used in the measurement of organs. *Proceedings of the Royal Society of London*, 60:489–498, 1897. `http://rspl.royalsocietypublishing.org/content/60/359-367/489.full.pdf+html`. 55

[293] Karl Pearson. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburg and Dublin Philosophical Magazine and Journal of Science, fifth series*, pages 157–175, 1900. 55, 79

[294] Karl Pearson. "Das Fehlergesetz und seine Verallgemeinerungen durch Fechner und Pearson." A rejoinder. *Biometrika*, pages 169–212, 1905. 56

[295] Henri Poincaré. *Calcul des probabilités. Leçons professées pendant le deuxième semestre 1893–1894.* Gauthier-Villars, Paris, 1896. `https://archive.org/details/calculdeprobabil00poinrich`. Second edition 1912. 22, 123

[296] Henri Poincaré. *Calcul des probabilités.* Gauthier-Villars, Paris, 1912. Second edition of [295]. 22, 58

[297] Siméon-Denis Poisson. Sur la probabilité du résultat moyen des observations. *Connaissance des Tems pour l'an 1827*, pages 273–302 of section for articles, 1824. `http://gallica.bnf.fr/ark:/12148/bpt6k65067080/f281.image.langEN`. 20

[298] Siméon-Denis Poisson. Suite du mémoire sur la probabilité du résultat moyen des observations, inséré dans la Connaissance des Tems de l'année 1827. *Connaissance des Tems pour l'an 1932*, pages 3–22 of section for articles, 1829. `https://archive.org/details/connaissancedes13longgoog`. 20

[299] Siméon-Denis Poisson. Recherches sur la probabilité des jugements, principalement en matière criminelle. *Comptes rendus hebdomadaires des séances de l'Académie des Sciences*, 1:473–494, 1835. Session of 14 December 1835. `http://gallica.bnf.fr/ark:/12148/bpt6k29606/f473.image.langEN`. 16

[300] Siméon-Denis Poisson. *Recherches sur la probabilité des judgments en matière criminelle et en matière civile, précédés des règles générales du calcul des probabilités*. Bachelier, Paris, 1837. http://gallica.bnf.fr/ark:/12148/bpt6k110193z.r=.langFR. 16, 18, 124

[301] Siméon-Denis Poisson. *Lehrbuch der Wahrscheinlichskeitrechnung und deren wichtigsten Anwendungen*. Meyer, Braunschweig, 1841. Translation of [300] by Christian Heinrich Schnuse. 11

[302] Georg Pólya. Sur quelques points de la théorie des probabilités. *Annales de l'Institut Henri Poincaré*, 1:117–161, 1930. 74

[303] Theodore Porter. *The Rise of Statistical Thinking, 1820–1900*. Princeton University Press, Princeton, NJ, 1986. 54

[304] M. P. Quine and Eugene Seneta. Bortkiewicz's data and the law of small numbers. *International Statistical Review*, 55(2):173–181, 1987. http://www.jstor.org/stable/1403193. 48

[305] Eugenio Regazzini. Probability and statistics in Italy during the First World War. I: Cantelli and the laws of large numbers. *Electronic Journal for History of Probability and Statistics*, 1(1), 2005. 59

[306] Hans Richter. *Wahrscheinlichkeitstheorie*. Springer, Berlin, 1956. 77

[307] Henry Lewis Rietz. *Mathematical Statistics*. Open Court for the Mathematical Association of American, Chicago, 1927. https://ia700400.us.archive.org/34/items/mathematicalstat00riet/mathematicalstat00riet.pdf. 67

[308] Réné Risser and Claude-Emile Traynard. *Les principes de la statistique mathématique*. Gauthier-Villars, Paris, 1933. 62

[309] Vsevolod Ivanovich Romanovskii. On the distribution of the regression coefficient in samples from normal population. *Bulletin de l'Académie des Sciences de l'URSS. VI série*, 20:643–648, 1926. http://mi.mathnet.ru/izv5578. 63, 79

[310] Vsevolod Ivanovich Romanovskii. On the moments of means of functions of one or more random variables. *Metron*, 8:251–, 1929. 63, 79

[311] Vsevolod Ivanovich Romanovskii. Sur une extension du théorème de A. Liapounoff sur la limite de probabilité. *Bulletin de l'Académie des Sciences de l'URSS. VII série*, 2:209–225, 1929. http://mi.mathnet.ru/izv5336. 73

[312] Winfried Scharlau. *Mathematische Institute in Deutschland 1800–1945*. Vieweg, Braunschweig, 1990. 20

[313] Ivo Schneider. Christian Heinrich Schnuse als übersetzer mathematischer, naturwissenschaftlicher und technischer literatur. *Aus dem Antiquariat*, pages A205–A221, A256–A261, 1982. 11

[314] Eugene Seneta. Review of [288]. *Annals of Science*, 39:614–617, 1982. 43

[315] Eugene Seneta. On the history of the strong law of large numbers and Boole's inequality. *Historia Mathematica*, 19:24–39, 1992. http://ac.els-cdn.com/031508609290053E/1-s2.0-031508609290053E-main.pdf?_tid=71bec798-0d76-11e5-8586-00000aab0f02&acdnat=1433723987_7e28add3cab56ecd4ec6d8e0164b98c7. 59

[316] Eugene Seneta. Statistical regularity and free will: L.A.J. Quetelet and P.A. Nekrasov. *International Statistical Review*, pages 319–324, 2003. 40

[317] Eugene Seneta. Buniakovsky's probability book. Reviews. Quality control. Regularly varying sequences. In H. Syta, M. Horbachuk, and A. Yurachkivsky, editors, *Viktor Yakovych Bunyakovsky: On the 200th anniversary of his birth*, pages 149–164. Institute of Mathematics, Ukrainian Academy of Sciences, Kiev, 2004. ISBN 966-02-3380-9. 12

[318] Eugene Seneta. Mathematics, religion and Marxism in the Soviet Union in the 1930s. *Historia Mathematica*, 31:337–367, 2004. `http://ac.els-cdn.com/S0315086003000466/1-s2.0-S0315086003000466-main.pdf?_tid=e8bc4838-0d22-11e5-9bd1-00000aacb360&acdnat=1433688109_fd3b312539c7d2a845272079acb85a79`. 76

[319] Eugene Seneta. V.Ya. Buniakovsky: A sketch of life and work. In H. Syta, M. Horbachuk, and A. Yurachkivsky, editors, *Viktor Yakovych Bunyakovsky: On the 200th anniversary of his birth*, pages 61–70. Institute of Mathematics, Ukrainian Academy of Sciences, Kiev, 2004. ISBN 966-02-3380-9. 12

[320] Eugene Seneta. Markov and the creation of Markov chains. In Amy N. Langville and William J. Stewart, editors, *MAM 2006: Markov Anniversary Meeting*. Boson Books, Raleigh, North Carolina, 2006. ISBN 1-932482-34-2. 41, 43

[321] Eugene Seneta. Review of [344]. *Historia Mathematica*, 37:716–722, 2010. `http://ac.els-cdn.com/S0315086009001335/1-s2.0-S0315086009001335-main.pdf?_tid=a82e94f2-0d76-11e5-8704-00000aacb35d&acdnat=1433724078_b103ae5e4db688258b898659a3b39427`. 76

[322] Eugene Seneta. On the Bicentenary in St. Petersburg of Jacob Bernoulli's Theorem. *International Statistical Review*, 82(1):17–26, 2014. 20

[323] Eugene Seneta, Karen Hunger Parshall, and François Jongmans. Nineteenth-century developments in geometric probability: J. J. Sylvester, M. W. crofton, J.-E. Barbier, and J. Bertrand. *Archive for History of Exact Sciences*, 55(6):501–524, 2001. `http://www.jstor.org/stable/41134124`. 31

[324] Glenn Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, NJ, 1976.

[325] Glenn Shafer. Bayes's two arguments for the rule of conditioning. *Annals of Statistics*, 10:1075–1089, 1982. An examination of the role of time in Thomas Bayes's theory of probability. 49, 101

[326] Glenn Shafer and Vladimir Vovk. *Probability and Finance: It's Only a Game!* Wiley, New York, 2001. 97

[327] Glenn Shafer and Vladimir Vovk. The origins and legacy of Kolmogorov's *Grundbegriffe*, 2004. Working Paper 4, `http://www.probabilityandfinance.com`. 75, 76, 101

[328] Glenn Shafer and Vladimir Vovk. The sources of Kolmogorov's *Grundbegriffe*. *Statistical Science*, 21:70–98, 2006. `https://projecteuclid.org/download/pdfview_1/euclid.ss/1149600847`. 75, 101

[329] Oscar Sheynin. S. D. Poisson's work in probability. *Archive for History of Exact Sciences*, 18:245–300, 1978. 15, 18

[330] Oscar Sheynin. A. A. Markov's work on probability. *Archive for History of Exact Sciences*, 39:337–377, 1989. 41, 42, 43

[331] Oscar Sheynin. On V. Ya. Buniakovsky's work in the theory of probability. *Archive for History of Exact Sciences*, 43(3):199–223, 1991. 12

[332] Oscar Sheynin. Nekrasov's work on probability: The background. *Archive for History of Exact Sciences*, 57:337–353, 2003. 40, 41

[333] Oscar Sheynin, editor. *Alexandr Chuprov, Statistical Papers and Memorial Publications*. NG Verlag, #2 at `www.sheynin.de`, Berlin, 2004.

[334] Oscar Sheynin, editor. В.И. Борквевич, А.А. Чиров, Перепицска (1895–1926) *(V. I. Bortkevich, A. A. Chuprov, Letters (1895–1926))*. NG Verlag, #9 at `www.sheynin.de`, Berlin, 2004. 48

[335] Oscar Sheynin, editor. *P. A. Nekrasov: The Theory of Probability*. NG Verlag, #4 at `http://www.sheynin.de`, Berlin, 2004.

[336] Oscar Sheynin. *Russian papers on the history of probability and statistics, translated by the author*. NG Verlag, #1 at `www.sheynin.de`, Berlin, 2004. 40, 41

[337] Oscar Sheynin, editor. *Russian papers of the Soviet period*. NG Verlag, #7 at `www.sheynin.de`, Berlin, 2005.

[338] Oscar Sheynin. *Theory of Probability. A Historical Essay*. #10 at `www.sheynin.de`, Berlin, 2009. Second revised and enlarged edition, ISBN 3-938417-88-9.

[339] Oscar Sheynin, editor. *Evgeny Slutsky: Collected Statistical Papers*. NG Verlag, #40 at `www.sheynin.de`, Berlin, 2010.

[340] Oscar Sheynin, editor. *Russian papers of the Soviet period*. NG Verlag, #7 at `www.sheynin.de`, Berlin, 2010.

[341] Oscar Sheynin. *Aleksandr A. Chuprov: Life, Work, Correspondence. The making of mathematical statistics*. V&R unipress, Goettingen, 2011. Second revised edition, edited by Heinrich Strecker. The first edition appeared in 1996. 41, 48, 51, 67, 122

[342] B. L. Shook. Synopsis of elementary mathematical statistics. *Annals of Mathematical Statistics*, 1:14–41, 1930. `https://projecteuclid.org/download/pdf_1/euclid.aoms/1177733258`. 57

[343] Reinhard Siegmund-Schultze. *Rockefeller and the internationalization of mathematics between the two world wars*. Birkhäuser, Basel, 2001. 70

[344] Reinhard Siegmund-Schultze. *Mathematicians Fleeing from Nazi Germany: Individual Fates and Global Impact*. Princeton University Press, Princeton, 2009. 125

[345] Thomas Simpson. A letter to the Right Honorable George Earl of Macclesfield, President of the Royal Society, on the advantage of taking the mean of a number of observations, in practical astronomy. *Philosophical Transactions of hte Royal Society of London*, 49:82–93, 1755. `http://rstl.royalsocietypublishing.org/content/49/82.full.pdf+html`. 13

[346] Dmitrii Matveevich Sintsov. Review of [258]. *Jahrbuch über die Fortschritte der Mathematik for 1899)*, 30:228–229, 1901. JFM 31.0228.03. 44

[347] Evgenii Evgenevich Slutskii. Über stochastische Asymptoten und Grenzwerte. *Metron*, 5:3–89, 1925. 68

[348] Evgenii Evgenevich Slutskii. Sur un critérium de la convergence stochastique des ensembles de valeurs éventuelles. *Comptes rendus hebdomadaires des séances de l'Académie des Sciences*, 187:370–372, 1928. Session of 13 August. `http://gallica.bnf.fr/ark:/12148/bpt6k3140x/f370.image.langEN`. 62

[349] J. Laurie Snell. A conversation with Joe Doob. *Statistical Science*, 12:301–311, 1997. `https://projecteuclid.org/download/pdf_1/euclid.ss/1030037961`.

[350] Herbert Solomon. *Geometric Probability*. SIAM, Philadelphia, 1978. 34

[351] Johan Steffensen. Deux problèmes du calcul des probabilités. *Annales de l'Institut Henri Poincaré*, 3:319–344, 1933. `http://www.numdam.org/item?id=AIHP_1933__3_3_319_0`. 74

[352] Stephen M. Stigler. *The History of Statistics: The Measurement of Uncertainty before 1900*. Harvard University Press, Cambridge, MA, 1986. 8, 13, 15, 20, 101

[353] Stephen M. Stigler. *Statistics on the Table: The History of Statistical Concepts and Methods*. Harvard University Press, Cambridge, MA, 1999. 55

[354] Stephen M. Stigler. Studies in the history of probability and statistics, L: Karl Pearson and the Rule of Three. *Biometrika*, 99:1–14, 2012. 54, 65

[355] Student. The elimination of spurious correlation due to position in time or space. *Biometrika*, 10:179–180, 1914. 58

[356] Ilona Svetlikova. *The Moscow Pythagoreans : mathematics, mysticism, and anti-semitism in Russian symbolism*. Palgrave Macmillan, New York, 2013. 37, 39

[357] Nassim Taleb. *The Black Swan: The Impact of the Highly Improbable*. Random House and Penguin, New York, 2007.

[358] Isaac Todhunter. *A History of the Mathematical Theory of Probability from the Time of Pascal to that of Laplace*. Macmillan, London, 1865. `https://ia601409.us.archive.org/27/items/historyofthemath000979mbp/historyofthemath000979mbp.pdf`. 94

[359] Matthias C. M. Troffaes and Gert de Cooman. *Lower Previsions*. Wiley, Chichester, 2014. 5, 100

[360] Charles F. Trustam. Some recent Italian work on the mathematical basis of actuarial science. *Journal of the Institute of Actuaries*, 59:51–66, 1928. 63, 79

[361] Friedrich Maria Urban. *The Application of Statistical Methods to the Problems of Psychophysics*. Psychological Clinic Press, Philadelphia, 1908. 40

[362] James V. Uspensky. *Introduction to Mathematical Probability*. McGraw-Hill, New York, 1937. 52, 110

[363] Johannes von Kries. *Die Principien der Wahrscheinlichkeitsrechnung. Eine logische Untersuchung*. Mohr, Freiburg, 1886. https://ia802708.us.archive.org/1/items/dieprincipiende00kriegoog/dieprincipiende00kriegoog.pdf. The second edition, which appeared in 1927, reproduced the first without change and added a new 12-page forward.

[364] Richard von Mises. Théorie des probabilités. Fondements et applications. *Annales de l'Institut Henri Poincaré*, 3:137–190, 1932. http://www.numdam.org/item?id=AIHP_1932__3_2_137_0. 74

[365] John von Neumann and Oscar Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, Princeton, NJ, third edition, 1953. Previous editions appeared in 1944 and 1947. 5

[366] Jan von Plato. *Creating Modern Probability: Its Mathematics, Physics, and Philosophy in Historical Perspective*. Cambridge University Press, Cambridge, 1994. 75

[367] Abraham Wald. *Berechnung und Ausschaltung von Saisonschwankungen*. Springer, Vienna, 1936. 85

[368] Abraham Wald. Die Widerspruchfreiheit des Kollectivbegriffes der Wahrscheinlichkeitsrechnung. *Ergebnisse eines Mathematischen Kolloquiums*, 8:38–72, 1937. 86

[369] Abraham Wald. Generalization of the inequality of Markoff. *Annals of Mathematical Statistics*, 9:244–255, 1939. https://projecteuclid.org/download/pdf_1/euclid.aoms/1177732281. 83, 86

[370] Abraham Wald. *Sequential Analysis*. Wiley, New York, 1947. 86, 88

[371] Abraham Wald. *Statistical Decision Functions*. Wiley, New York, 1950. 86

[372] Abraham Wald. The publications of Abraham Wald. *Annals of Mathematical Statistics*, 23:29–33, 1952. https://projecteuclid.org/download/pdf_1/euclid.aoms/1177729483. 86

[373] Abraham Wald and Jacob Wolfowitz. Confidence limits for continuous distribution functions. *Annals of Mathematical Statistics*, 10:105–118, 1939. https://projecteuclid.org/download/pdf_1/euclid.aoms/1177732209.

[374] Peter Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, 1991. 5, 98, 100

[375] Rolin Wavre, editor. *Colloque consacré à la théorie des probabilités*. Hermann, Paris, 1938–1939. 112, 113, 122

[376] Rolin Wavre, editor. *L'application du calcul des probabilités. Genève, 12-15 juillet 1939*, Geneva, Paris, 1946. Faculté des Sciences de l'Université de Genève, Institut International de Coopération Intellectuelle. Because of the war, these proceedings did not appear until 1946. 122

[377] Herbert I. Weisberg. *Willful Ignorance: The mismeasure of uncertainty*. Wiley, Hoboken, 2014. 20

[378] Carl Joseph West. *Introduction to Mathematical Statistics*. Adams, Columbus, 1918. https://books.google.com/books/reader?id=nsoUAAAAYAAJ&printsec=frontcover&output=reader. 56

[379] William Allen Whitworth. *Choice and Chance*. Cambridge University Press, Cambridge, 1867. Later editions appeared in 1870, 1878, 1886, and 1901. 35

[380] William Allen Whitworth. *DCC Exercises, Including Hints for the Solution of All the Questions in Choice and Chance*. Deighton Bell, Cambridge, 1897. Here DCC is a Roman numeral, meaning 700. 36

[381] William Allen Whitworth. *The expectation of parts into which a magnitude is divided at random investigated mainly by algebraic methods*. Deighton Bell, Cambridge, 1898. http://babel.hathitrust.org/cgi/pt?id=mdp.39015017398911;view=1up;seq=8. 36

[382] Samuel S. Wilks. *The Theory of Statistical Inference*. Edwards, Ann Arbor, 1937. Planographed notes from a one-semester graduate course at Princeton University. 57

[383] Samuel S. Wilks. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Annals of Mathematical Statistics*, 9:60–62, 1938. https://projecteuclid.org/download/pdf_1/euclid.aoms/1177732360. 85

[384] Samuel S. Wilks. Shortest average confidence intervals from large samples. *Annals of Mathematical Statistics*, 9:166–175, 1938. https://projecteuclid.org/download/pdf_1/euclid.aoms/1177732308. 85

[385] Aurel Winter. On analytic convolutions of Bernoulli distributions. *American Journal of Mathematics*, 56:659–663, 1934. 80

[386] Jacob Wolfowitz. Additive partition functions and a class of statistical hypotheses. *Annals of Mathematical Statistics*, 13:247–279, 1942. https://projecteuclid.org/download/pdf_1/euclid.aoms/1177731566. 86

[387] Jacob Wolfowitz. On the theory of runs with some applications to quality control. *Annals of Mathematical Statistics*, 14:280–288, 1943. https://projecteuclid.org/download/pdf_1/euclid.aoms/1177731421. 86

[388] Jacob Wolfowitz. Abraham Wald, 1902–1950. *Annals of Mathematical Statistics*, 23:1–13, 1952. https://projecteuclid.org/download/pdf_1/euclid.aoms/1177729480. 86

[389] A. P. Youschkevitch. Chebyshev, Pafnuty Lvovich. In *Complete Dictionary of Scientific Biography*, volume 3, pages 222–232. Charles Scribner's Sons, New York, 2008. http://www.encyclopedia.com/doc/1G2-2830900876.html. 29

[390] George Udny Yule. On the correlation of total pauperism with proportion of out-relief. *The Economic Journal*, 5:603–611, 1895. 55

[391] George Udny Yule. On the significance of Bravais' formulae for regression, &c., in the case of skew correlation. *Proceedings of the Royal Society of London*, 60:477–489, 1897. http://rspl.royalsocietypublishing.org/content/60/359-367/477.full.pdf+html. 55

[392] George Udny Yule. On the theory of correlation. *Journal of the Royal Statistical Society*, 60:812–854, 1897. 55

[393] George Udny Yule. *An Introduction to the Theory of Statistics*. Griffin, London, 1911. https://ia801405.us.archive.org/25/items/anintroductiont00yulegoog/anintroductiont00yulegoog.pdf. 51, 55

[394] George Udny Yule. Why do we sometimes get nonsense-correlations between time-series?–A study in sampling and the nature of time-series. *Journal of the Royal Statistical Society*, 89:1–63, 1926. 84

[395] Sandy L. Zabell. *Symmetry and its Discontents*. Cambridge, Cambridge, 2005. 37

[396] Tian Zheng and Zhiliang Ying. Columbia University statistics. In Alan Agresti and Meng Xiao-Lin, editors, *Stength in Numbers: The Rising of Academic Statistics Departments in the U.S.*, pages 27–38. Springer, New York, 2013. `http://stat.columbia.edu/wp-content/uploads/2014/02/StatDeptHistory.pdf`. 86