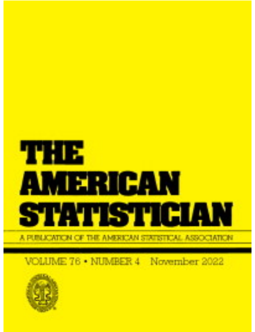

The American Statistician

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/utas20

The “Poisson” Distribution: History, Reenactments, Adaptations

James A. Hanley & Sahir Bhatnagar

To cite this article: James A. Hanley & Sahir Bhatnagar (2022) The “Poisson” Distribution: History, Reenactments, Adaptations, The American Statistician, 76:4, 363-371, DOI: 10.1080/00031305.2022.2046159

To link to this article: https://doi.org/10.1080/00031305.2022.2046159

Published online: 23 Mar 2022.

THE AMERICAN STATISTICIAN
2022, VOL. 76, NO. 4, 363–371: History Corner
https://doi.org/10.1080/00031305.2022.2046159

HISTORY CORNER

The "Poisson" Distribution: History, Reenactments, Adaptations

James A. Hanley[a] and Sahir Bhatnagar[a,b]

[a]Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montreal, QC, Canada; [b]Department of Radiology, McGill University, Montreal, QC, Canada

ABSTRACT
Although it is a widely used—and misused—discrete distribution, textbooks tend to give the history of the Poisson distribution short shrift, typically deriving it in the abstract as a limiting case of a binomial. The biological and physical scientists who independently derived it using space and time considerations and used it in their work are seldom mentioned. Nor are the difficulties of applying it to counts involving human activities/behavior. We (a) sketch the early history of the Poisson distribution (b) illustrate principles of the Poisson distribution involving space and time using the original biological and physical applications, as well as modern multimedia reenactments of them, and (c) motivate count distributions accounting for extra-Poisson variation. The replayed historical applications can help today's students, teachers and practitioners to see or hear what randomness looks or sounds like, to get practice in the practicalities of "counting statistics," to distinguish situations where the pure Poisson distribution does and doesn't hold—and to think about what one might do when it doesn't.

ARTICLE HISTORY
Received June 2021
Accepted February 2022

KEYWORDS
Animation; Brewing; Counting statistics; Derivation; Heterogeneity; History; Physics; Space; Telegraphy; Time; Visualization

1. Introduction

Although it is a widely used—and misused—discrete distribution, textbooks tend to give the history of the Poisson distribution short shrift, typically deriving it in the abstract as a limiting case of a binomial. Moreover, some of the generic examples (such as the numbers of patients arriving in an emergency room between certain hours; daily tweets from a certain politician; phone calls received by a call center in a day; daily fatalities on highways) do not necessarily meet the conditions for its proper application.

Poisson himself never gave an example of, or applied the distribution now called after him. It ended up being independently rediscovered and used by early-20th century biological and physical scientists who derived it using space and time considerations. Some of them also linked it with what we refer to today as the exponential distribution. Quite early on, it was also recognized that counts involving human activities/behavior were not always well fitted by the Poisson distribution, and that extensions were needed.

We (a) sketch the early history of the Poisson distribution (b) illustrate its space and time aspects using both the original biological and physical applications, and modern multimedia reenactments of them, and (c) motivate count distributions that allow for recognized and unrecognized sources of heterogeneity, that is, "extra-Poisson" variation.

These sources of heterogeneity are, of course, not limited to this specialized context of count data. Examples are the use of 2-level (hierarchical) Gaussian models for the sizes of apples from trees in an orchard or of offspring of different families, performance of children from different schools; and beta-geometric models for the number of cycles different couples require to achieve pregnancy. Even in introductory courses that are limited to 1-level models, such examples serve as a reminder that more complex models are needed for applications involving human activities/behavior or environmental factors.

The early 1900s are when the Poisson distribution began to be rediscovered, and to come of age. Thus, we divide our retelling accordingly. Section 2 deals with some pre-1900 derivations, and the limited applications, while Section 3 describes—and uses 21st century computer power to animate—the application-based rediscoveries of the first decade of the 20th century. The animations can help today's students, teachers and practitioners to see and hear what randomness looks like, and introduce them to the practicalities and analysis of "counting statistics." Section 4 deals with its formal entry into statistics in the second decade, and the teaching of it in the latter half of the 20th century, while Section 5 explains why extensions are typically needed when it is applied outside biological and physical laboratories.

2. Before 1900

2.1. deMoivre 1718, and Poisson 1837

Stigler's (1982) article "Poisson on the Poisson Distribution" is an excellent point of departure. It includes a translation of the totality of Poisson's own (1837, p. 205) discussion of the Poisson distribution. It describes how "Poisson reexpressed the upper tail cumulative binomial probability as a lower tail cumulative

CONTACT James A. Hanley james.hanley@mcgill.ca Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, 2001 McGill College Avenue, Montreal, Quebec H3A 1G1, Canada.

negative binomial probability. Consequently, we will have, very nearly, (in today's notation)

$$\text{“}\ P = \left(1 + \mu + \frac{\mu^2}{2!} + \frac{\mu^3}{3!} + \cdots + \frac{\mu^y}{y!}\right) e^{-\mu}$$

for the probability that an event, whose chance on each trial is the very small fraction μ/n, will not happen more than y times in a large number n of trials. In the case $y = 0$, this value of P reduces to $e^{-\mu}$, [..] the probability [...] that the event we are concerned with will not happen a single time in the n trials."

Stigler also describes how Poisson's distribution was foreshadowed by de Moivre's (1718, pp. 14–18) analysis—involving the same series that defines P—that sought the value n for which $P = 1/2$, for various values of y. Stigler addressed authors such as Newbold (1927) and David (1962), who have "felt the distribution should be attributed to De Moivre." While admitting that "Poisson added little to de Moivre's mathematical approximation, with which he was quite familiar," Stigler thinks "one would have to stretch the point to claim the discrete distribution $e^{-\mu}\mu^y/y!$ is found in De Moivre."

2.2. Newcomb 1860

For a further history of the distribution, Stigler suggests chap. 9 of Haight (1967). He notes, however, that it overlooks the (1860) work of Newcomb "where the distribution is suggested as a fit to data for perhaps the first time." The U.S.-based Newcomb derived the $e^{-\mu}\mu^y/y!$ distribution as a limiting case of the binomial, and used it to determine the "probability that, if the stars were scattered at random over the heavens, *any small space selected at random* would contain s stars." The arguments focused on the six brightest stars in the Pleiades. He considered a space of 1 square degree and supposed that there were approximately 1500 large stars spread at random over the entire 41,253 square degrees of heavens, substituting μ = 1500/41,253 = 0.03636 into $e^{-\mu}\mu^6/6!$ gave the probability that any square degree selected at random contains six stars. Multiplying this by 41,253, he obtained the probability of 0.000000128 that, if the heavens were divided at random into square degrees, *some* one of those square degrees would contain six stars.

The distinction between "one selected *at random*" and "*some* one" is lost on many amateurs who calculate probabilities after the fact (Hanley 1984). In a message that is especially relevant to today's p-hackers and "Texas sharpshooters" (Hanley 2018) with ready access to Big Data and computing power, Newcomb went further, cautioning that "this, however, is evidently rather smaller than the probability that six stars should be found so near together that a square degree could be fitted on so as to include them." Evaluating the latter probability by simulation could provide a vivid lesson—and a computing challenge—for today's students.

2.3. Clausius to Bortkewitsch, 1858–1898

Several other early authors, such as Clausius in 1858, and Whitworth in 1870 had derived the probability, $e^{-\mu}$, of observing no events.

To help users of a new Zeiss microscope count the number of blood cells in a sample, Abbe (1879) determined the number of cells one needed to count in order to achieve a sufficiently small "probable error" for an estimated blood cell concentration. He began by stating, without citing a source, the $e^{-\mu}\mu^y/y!$ distribution, but immediately moved to a Normal approximation, which he claimed was sufficient when μ is larger than 30.

Bortkewitsch's (1898) book, entitled the Law of Small Numbers, begins by mentioning Poisson, and by stating that Poisson's Law can be derived as a limiting case of the binomial distribution; it ends with extensive tables. In between, it worked through four data series of very small counts, each subdivided by sex, state, or work unit. The best-known is the yearly numbers of deaths from horse-kicks over 20 years in each of 14 Prussian cavalry corps. He found that "the agreement of theory with observation leaves nothing to be desired." whether one examines all 280 counts, or the 200 from the 10 normal corps. Unfortunately, unless one studies the original, as Winsor (1947) did, and his sources, as Stigler (2019) did, it is easy to misinterpret Bortkewitsch's work. Some authors took the title of his book as a synonym for, and endorsement of, the Poisson distribution, and the good fit as evidence of uniformity. Others criticized him for not looking more intensely for evidence of nonuniformity. In fact, he was concerned about the possibility of heterogeneity, and his results can also be taken as a warning that it is very difficult to detect nonuniformity in series of counts where μ's are small. His arguments were both theoretical and empirical; the 20th century produced more sensitive methods to demonstrate the nonuniformity in his empirical series (Fisher, Thornton, and Mackenzie 1922; Winsor 1947; Preece, Ross, and Kirby 1988).

Today, some textbooks and online sources continue to oversell the Poisson law, as if numbers of suicides, or accidents, or infections obeyed the law in the same way that the counts of yeast cells or atomic disintegrations do. In Section 5 we explain why counts of events generated by humans seldom follow pure Poisson distributions. But, for now, we turn to the laboratory scientists, and to their considerable—but little known—early-1900 involvement with the "almost pure" Poisson distribution.

3. Early 1900s

3.1. The Distribution of Objects in Space/Volumes, Gosset 1907

In 1899, the Guinness company hired Oxford science graduate William Gosset to help with its "scientific" brewing and compete with its European competitors Hanley, Julien, and Moodie (2008). Gosset's lesser-known 1907 article "On the error of counting with a haemocytometer" (Student 1907), is a rigorous, practical and readable introduction to the use of (what we now know as) the Poisson distribution for the "counting statistics" generated when "counting yeast cells or blood corpuscles." It was his first article, and published under the pen-name "Student," a choice which may be explained by viewing the cover of the notebook used to record the counts (Gosset 1907). To promote this classic but under-valued work, and use today's media to visualize spatial randomness (and counting) in action, we have created a dedicated *website http://www.biostat.mcgill.ca/hanley/Gosset/*.

The homepage begins with his motivation: since the distribution over the area which is examined is "never absolutely uniform," there is an unavoidable "error of random sampling."

Thus, his concern was "the distribution of particles throughout a drop of liquid, as shewn by spreading it in a thin layer over a measured surface and counting the particles per unit area."

The homepage also highlights the sections of his article, beginning with his derivation of the theoretical distribution from first principles, with no references to prior work. He imagined the whole liquid (i.e., the liquid sampled from) to have been well mixed and spread out in a thin layer over a very large number of units of area and the particles allowed to subside. Then assuming the liquid has been properly mixed, a given particle will have an equal chance of falling on any unit area. A little algebra showed that, since the chance is small, the (binomial) probabilities of y particles falling in a given unit area converge to what we now call the Poisson distribution.

He showed that its variance equaled its mean and that it follows "the normal curve" as the mean becomes infinite. The latter allowed him to offer guidance on how much one needs to count to ensure a small enough margin of error for use in a (say 50%) z-based confidence interval. Today, it continues to be a useful exercise for students to show that what matters is the number of particles counted, not the dilution: "the most accurate way is to dilute the solution to the point at which the particles may be counted most rapidly, and to count as many as time permits." With "particles" replaced by events and "fields" by person-time, the same principle underlies today's epidemiologic and vaccine research: the precision of an estimated hazard ratio or vaccine efficacy depends on the numbers of events, not the person time.

He was, however, also concerned with a second possible source of error, the difficulty of "obtaining a drop representative of the bulk of the liquid." Thus, "it is usual to take several drops: if two of these differ in their means by a significant amount compared with the [SE] of their difference, it is probable that one at least of the drops does not represent the bulk of the solution."

He tested his law on four distributions (from four different concentrations) which had been counted over the whole 400 squares of the hemacytometer. He did so by comparing the first two moments, and by χ^2 goodness of fit statistics. We urge modern-day students to read his clear description of the laboratory methods, and how he addressed the several practical issues. His careful work produced quite satisfactory fits, even with the budding (cell division) of some of the yeast cells and "a tendency to stick together in groups which was not altogether abolished even by vigorous shaking." Using the raw data Gosset provided on the four distributions, we also encourage students to reproduce his fits and statistics, and teachers to explain why we would compute and interpret them a little differently today. Back then, when all one had was a series of counts, it was common to use the method of moments to fit the "binomial" parameters n and q from their first two empirical moments m_1 and m_2. Thus, in his series I, where $m_1 = 0.6825$ and $m_2 = 0.8117$, he used the m_2/m_1 ratio of 1.1893 to obtain the "best fitting (here, negative) binomial" frequencies $(1.1893 - 0.1893)^{-3.6054} \times 400$. In 1907, there were no warnings about small expected numbers, and the appropriate number of degrees of freedom for the goodness of fit statistic (Stigler 2008) had not yet been established. However, Gosset did appreciate that the better fit of the binomial distribution with two free parameters n and q than the one-parameter $400e^{-\mu}\mu^y/y!$ "Poisson" distribution was "only to be expected."

Even though we are not able to emulate his critical wet-lab procedure, our interactive website does let students "see" what the resulting randomness looks like, either by hovering over a portion of the grid to magnify it, or in a video which proceeds sequentially through the $20 \times 20 = 400$ squares. Gosset was well aware of the "difficulty in putting the drop on to the slide so as to be able to count at any point and in any order." Our numbered squares allow students to test whether, "as good a result may be expected from counting a column as from counting the same number of squares at random." Today's students have many more options for testing this than "Student" did.

We provide visualizations of the "cells" in 5 drops from each of four concentrations. In some we attempt to match the reported distributions, and in others we create drops that are not representative of the bulk, or contain some sticky cells. Identifying which drop is which is left as an exercise. This task leads naturally to a discussion about the appropriate standard error of the mean (SEM) to accompany the estimated concentration, in cases where the model-based one is too tight. To focus attention on the assumptions, we also encourage readers to think about how they might simulate *violations* of the assumptions of the Poisson distribution, and are happy to privately tell interested teachers what we did with some of our drops. Lastly, with M denoting the number of unit volumes counted, we ask readers if they agree with Gosset's concluding statement "the criterion of whether two *solutions* contain different numbers of cells is whether $\bar{y}_1 - \bar{y}_2$ is significant compared with the (model-based) $\sqrt{\bar{y}_1/M_1 + \bar{y}_2/M_2}$" SE of the difference, or whether one based on between-drop variations is a better measure of the noise.

3.2. Counting Events in Time, Rutherford, Geiger, and Bateman 1910

The early 1900s saw rapid developments in our understanding of the atom, and some important related, but lesser known, statistical developments. In 1905 already, Schweidler had addressed the sampling errors involved in counting a sufficient number of radioactive transformations within a feasible amount of time. He showed that at the limit, with small short term probabilities in a binomial model with a large "N," the variance equals its mean. He did not derive the exact discrete distribution of individual counts in small time intervals; instead, because of the large overall count required, he used a Normal distribution whose variance equaled its mean. Thus, for example, counting 10^4 radioactive transformations involves a coefficient of variation (CV) of $10^2/10^4$ or 1%, counting $(1/36) \times 10^6$ a CV of 0.6%.

Like others, Manchester physicists Rutherford, Geiger, and Bateman (1910) had noticed that when

> counting the α particles emitted from radioactive substances either by the scintillation or electric method, while the average number of particles from a steady source is nearly constant, when a large number is counted, the number appearing in a given short interval is subject to wide fluctuations. These variations are especially noticeable when only a few scintillations appear per minute. For example, during a considerable interval it may happen that no α particle appears;

then follows a group of α particles in rapid succession; then an occasional a particle, and so on.

To them, it was "of importance to settle whether these variations in distribution are in agreement with the laws of probability," that is,

> whether the distribution of α particles on an average is that to be anticipated if the particles are *expelled at random both in regard to space and time*. It might be conceived, for example, that the emission of an α particle might *precipitate the disintegration of neighbouring atoms*, and so lead to a distribution of α particles at variance with the simple probability law.

In 1910 they directly tested the probability variations. They had intended initially to determine the distribution of α particles in time by the electric method. Photographs were readily obtained on a revolving film ; but it was found to be a long and tedious matter to obtain records of the large number of α particles required. Thus, it was simpler, if not quite so accurate, to count them by the scintillation method. Like Gosset's, their article is a model of clarity that today's authors could profitably study and emulate. Since the newer generation likes to engage with video/audio as well as with words, we have created a multimedia *website http://www.biostat.mcgill.ca/hanley/Rutherford/*.

In addition to highlighting the study background, experimental arrangement, and results, the website includes a re-enactment of their five days of data-acquisition, and allows students to see and hear the time-distribution of the disintegrations, and to carry out their own counting and data-analysis. The website also repeats Bateman's unusual derivation, using a time-based argument that led to differential equations, of the theoretical law (now attributed to Poisson) that closely matched their empirical distribution.

Several features of their experiment were critical to its success: the daily moving of polonium closer to the screen to maintain a nearly constant average number of α particles impinging on the screen, time-stamping the scintillations recorded on a chronograph tape, and resting of the eyes after every 3–5 min of counting. Particularly important was the wise combination of the source distances and the 1/8 min time-bins into which the 10,097 scintillations (recorded over a total of $2608 \div 8 \div 60 =$ 5.4 hr) were counted. These produced a mean of $10,097 \div 2608$ = 3.87 scintillations per bin, with a range of 0–14; 10 of the 15 frequencies exceeded 25. Had the mean per bin exceeded 10, the Central Limit theorem would have made it difficult to distinguish the distribution from a Gaussian-like one. Had it been below 2 (it was 0.68, 1.32 and 1.8 in three of Gosset's four distributions, and 0.78 in Bortkewitsch's best known one), the limited range and number of degrees of freedom would have made it difficult to reliably distinguish among several candidate distributions. (The authors did also combine every second 1/8 min interval to make larger intervals, with a mean of 7.74 particles per bin.)

Even if our conditions are less difficult than counters endured in 1910, we have further simplified the task: we limit our video and audio clips to 1 min each, with cues separating the 7.5 sec intervals, thus allowing for eight separate hand tallies. Some may prefer to timestamp the flashes on one of their devices using their own app or the stopwatch/lap-split app we include.

The authors concluded that, "on the whole, theory and experiment are in excellent accord," but seemed unaware of Pearson's goodness of fit (G.o.F) statistic(1900). The fits may be compared with the multi-parameter (Pearson) "ideal frequency curves" fitted a year later (Snow 1911). Snow's use of Pearson's catalog of distributions, and his zeal to improve the p-value of the G.o.F. statistic, without correcting for the additional degrees of freedom expended, or the number of models "tried-on," can be made into a valuable lesson for today's even-more-zealous "data" scientists.

3.3. Connecting the Poisson and Exponential Distributions, Marsden and Barratt 1910–1911

In 1910–1911, then-London-based physicists Marsden and Barratt (1911a, 1911b) made a very important connection between the discrete "number of events in a time interval" (Poisson) random variable and the continuous "interval between events" (negative-exponential) random variable. They too had been keen to understand the observation that, when "counting the alpha particles from radio-active substances by the scintillation method, the great preponderance of short intervals is very noticeable, especially when the scintillations are appearing at a slow rate. In fact this preponderance leads one at first sight to consider the particles as coming in groups and not distributed according to a simple probability law." That the counts seemed to obey Rutherford and Geiger's law was not sufficient for them.[1] "Were the particles given off in equal groups and the groups distributed in time according to probability it is conceivable that that law would still hold." To them, it therefore, seemed "preferable in many respects to test the application of the probability laws to *actual time intervals between successive alpha particles*."

As it may have been the first time the distribution of time intervals was approached in this way, we repeat their derivation here.

> Let the average time interval between successive α particles falling on a zinc-sulphide screen from a radio-active source be $1/\mu$, and let us assume that the time under consideration is small compared with the time period of the radio-active substance. Assuming that at time 0 the observer sees a scintillation, let us find the probability that a time interval, t, elapses without the occurrence of a second scintillation but that within a further very small interval, δt, a scintillation occurs. This problem corresponds exactly to the method employed in practice. The probability that no scintillation occurs in the time interval t is $e^{-\mu t}$. The probability that a scintillation occurs in the time interval from t to $t + \delta t$ is independent of t and is $\mu\delta t$. As these events are independent of one another the probability of an interval from t to $t + \delta t$ is the product of their separate probabilities and is therefore, $\mu e^{-\mu t} dt$; or in a large number of N intervals the probable number of intervals larger than t and smaller than $t + \delta t$ is $\mathrm{N}\mu e^{-\mu t} dt$.

[1] Nor was it for Fisher, Thornton, and Mackenzie (1922): "The distribution of 2608 counts shows a general agreement with expectation though there are discrepancies not easily to be explained by chance. The observations are certainly not adequate, as these authors suggest, as 'a method of testing the laws of probability.' "

> This result is at first sight somewhat surprising, for it indicates that whatever the value of μ small intervals are more probable than large ones, whereas one would at first sight expect that the intervals would be distributed according to a law somewhat similar to that of *Maxwell* for the distribution of the velocities of the molecules of a gas.

They applied this formula to their several sets of observations on the α particles from polonium and uranium and found it "to agree well with experiment." Our site shows their table and figure, based on 319 intervals, as well as their table showing the excellent fit to the 7564 intervals measured from time marks on the paper tapes from Rutherford and Geiger's experiments.

In an effort to connect the discrete Poisson and the continuous exponential distributions, our website includes a second version of each of the recreated minutes from Rutherford and Geiger's experiments, in which frame numbers are displayed in the frames in which flashes occur. The 24 frames per second are "sufficiently continuous" to be matched up with Marsden and Barratt's tabulated frequency distribution. The video-savvy newer generation will have little trouble extracting these timestamps.

3.4. *The Minimalist Derivation by Danish Engineer Erlang 1909*

In 1909, a year before Rutherford and Geiger, the Danish mathematician, scientist and telephone engineer Erlang began to describe how the Telephone Company of Copenhagen had, for several years already, been applying the theory of probabilities to their field (Erlang 1909). He began with "the probability of a certain number of (say y) calls being originated during a certain interval of time. He assumed *that there is no greater probability of a call being attempted at one particular moment than at any other moment*, and divided the unit of time in question into an infinitely large number of tiny but equal moments. Like Student, and without any reference to others, he arrived at the binomial limit, namely $(\mu^y/y!)e^{-\mu}$, where μ is the average number of calls arriving during the interval of time.

Erlang's "simple proof" of the Poisson formula—in an appendix to an article on telephone waiting times (Erlang 1920, sec. 12)—provides an elegant and minimalist way we might today derive and teach the $e^{-\mu}\mu^{/}y!$ formula (earlier in that piece, he had credited the formula "to the important theorem, the mathematical content of which was found by Poisson.")

> The proof of the theorem: *when, during a given time, the average number of calls is* μ*, the probability* (P_y) *of y calls being originated will be* $e^{-\mu}\mu^y/y!$ Let it be assumed that the time in consideration represents a portion of a very long time over which a correspondingly great number of calls is dispersed so that μ calls, at an average, fall within the time portion considered. The duration of the latter can be called μ, the unit of time being chosen in such a manner as to give an average of 1 call per unit of time. Let us suppose that, in a certain case, say, 5 calls occur within the time μ, and let us move μ a short distance $d\mu$; then there will be a probability $\frac{5d\mu}{\mu}$ that 1 of the 5 calls is shut out so that the number is reduced to 4. Vice versa, if we had 4 calls before μ was moved, there will be a probability $d\mu$ of gaining 1 new call by the movement. But the transitions from 5 to 4 and vice versa must neutralize each other, and so
>
> $$P_5 \times \frac{5d\mu}{\mu} = P_4 \times d\mu;\ \text{ that is, } P_5\frac{5}{\mu} = P_4,\ \text{ or } P_5 = P_4 \times \frac{\mu}{5}.$$

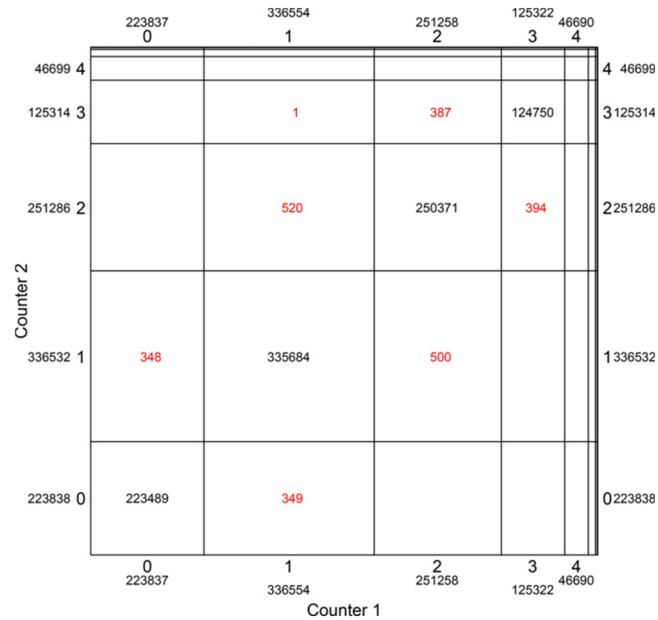

Figure 1. Illustration of Erlang's "statistical equilibrium" derivation of the Poisson distribution, using a simulation of events occurring randomly during 1,000,000 time units, with "no greater probability of an event at one particular moment than at any other." The time unit is chosen so that the mean number of events per (1.5 time-units wide) bin is $\mu = 1.5$. The numbers inside the square give the joint frequency distribution of the number of events (e.g., scintillations, or telephone calls initiated) per bin if counter 2 begins counting $d\mu = 0.0015$ sec after counter 1, and both count for 1,000,000 time units. The integers $y = 0, 1, 2, \ldots$ in a larger font along the margins are the numbers of events in a bin, and the numbers in a smaller font near them are the observed marginal frequencies (proportional to $e^{-1.5}1.5^y/y!$). To within Monte Carlo simulation error, the off-diagonal frequencies (in red) balance each other. For example, $500 \approx 251{,}300 \times \frac{2\times 0.0015}{1.5} \approx 336{,}500 \times 0.0015 \approx 520$. Likewise, there are equal numbers transitions between bins containing 1 and 2 calls, and between ones containing 2 and 3.

We leave to an exercise the remaining steps: they rely on the ratios between successive members of the sequence $P_0, P_1, \ldots,$ and the fact that the P_y's must add to 1. To further illustrate the equilibrium principle that Erlang used quite often, Figure 1 shows the results of a large number of moments, where $\mu = 1.75$, where "counter 1" began counting $d\mu = 0.01$ sec after "counter 2," but where both counted for the same total length of time.

4. Formal Entry | Reception | Warnings

4.1. *Soper 1914*

The distribution formally entered the English-language statistical literature in 1914, with the Biometrika publication of two back-to-back articles, each with Poisson's name in the title. Soper's Tables of Poisson's Exponential Limit contained nine pages of tables for μ values from 0.1 to 15 in steps of 0.1. These were preceded by just two pages of text that cited, in turn, Poisson, Student, Rutherford and Geiger, and Bortkewitsch. It also warned that "In vital statistics the sample may be an individual or house or community and the event an accident or disease and so on. But it must be borne in mind that for such series as the above to be applicable the occurrence of one event in the sample must not preclude or influence in any way the occurrence of a second."

4.2. Objections: Whitaker 1914, Student 1919, Keynes 1921

Soper's piece was followed by a 36-page article 'On the Poisson Law of Small Numbers' by Lucy Whitaker, a student of Karl Pearson. Rather than fit a single Poisson parameter μ, she used the method of moments to fit the two (positive or negative) binomial parameters n and q to the laboratory series of Gosset and the human series of Bortkewitsch and Mortara. She found that in a considerable number of instances both the fitted n and the fitted q were negative; in many others the fitted positive n was insufficiently large and the positive q insufficiently small for the "large n small q" prerequisites for the Poisson Law. Today this approach seems quite dated and unhelpful: the negative binomial distribution is still widely used today, but we no longer parameterize it using these two uninterpretable parameters.

Keynes (1921), who appears to be the first to reprint the full horse kicks dataset in an English text, was equally critical of Bortkewitsch's "so-called Law of Small Numbers." His reasons were more visceral and more Bayesian: "I should add that there is one other element which may contribute to the total psychological reaction of the reader's mind to the Law of Small Numbers, namely, the surprising and piquant examples which are cited in support of it. It is startling and even amusing to be told that horses kick cavalrymen with the same sort of regularity as characterizes the rainfall."

4.3. "Extra-Poisson" Variation in Bacteriology: Fisher 1922

Gosset (1907) has already alerted readers to the possibility of lumpiness and inhomogeneity when measuring concentrations in biological material. Fifteen years later, in an important technical article aimed at bacteriologists, Fisher (1922) noted how Gosset had arrived independently at the Poisson formula "as a theoretical result under technically perfect conditions." Furthermore, Gosset "was able to show that, i*n some instances* (italics ours), counts of 400 squares agreed with the theoretical distribution" and moreover, that "*when this is the case* the accuracy of the count is known with precision and depends only on the number of cells counted." He went on to explain that

> However, whereas the ideal conditions for bacterial counts made by the dilution method, are closely parallel to those found necessary in the case of the haemocytometer, the chief practical difference lies in the fact that instead of 400 squares with only a few yeast cells in each, we have some five plates with perhaps 200 colonies apiece. The agreement of the results with the theoretical distribution cannot, therefore, be demonstrated from a single count.

He summarized his extensive empirical investigations reported in this important but seldom-cited article with this advice

> Under ideal conditions the bacterial counts on parallel plates will vary in the same manner as samples from a Poisson Series. When these conditions are fulfilled the mean count of a number of plates is a direct measure of the density of the bacterial population considered; and the accuracy of such an estimate is known with precision. However, any significant departure from the theoretical distribution is a sign that the mean may be wholly unreliable.

4.4. Extra Variation in Human Counts: Erlang 1909; Student 1919

Before moving on to his main concern, the delays before the switch board operators could answer calls, Erlang (1909)—like Gosset—warned about (possibly unrecognized) sources of heterogeneity that would subvert the probability law and make the variance larger than μ. Since this may well be the first formal description of a mixture model for the Poisson distribution, and since deriving the excess variance can also be of didactic value, it is worth repeating.

> The simple suppositions leading to the simple formula will not, of course, always be satisfied in practice. Let us suppose, for example, that a business firm has certain busy days every week corresponding to a mean value μ_1, and certain less busy days corresponding to a mean value μ_2. Let the busy portion of the week be π_1, and the less busy portion $\pi_2 = 1 - \pi$. If it is desired here to express the variations in the number of calls from day to day in terms of one single law of distribution, we find that the mean value is $\pi_1\mu_1 + \pi_2\mu_2$; but the mean square error is $\pi_1\pi_2(\mu_1 - \mu_2)^2$ greater than the mean value.

His remark serves as a warning that whenever we count events arising out of human activity, we should not expect the tight $100 \times \sqrt{\mu}/\mu = 100/\sqrt{\mu}$ % coefficient of variation that Abbe, Gosset and Fisher hoped for. Human contexts where this constancy does seem to prevail [e.g., daily deaths in hospitals (Zweig and Csank 1978) and yearly vacancies in the U.S. Supreme Court (Cole 2010)] are the exception, but may well be examples of the illusion (produced by unexamined or unrecognized nonuniformity) that Bortkewitsch tried to describe by his Law. Daily numbers of deaths (Phillips, Brewer, and Wadensweiler 2010); Office of National Statistics 2020), births (Alam and Hanley 2018), and traffic accidents (Palayew, Harper, and Hanley 2021) in entire populations, and hourly numbers of births within a day (Hanley 2022) are examples of the more common pattern of "extra-Poisson" variation.

In 1919 Gosset offered an explanation for those "demographic cases" in which Whitaker's fits to the frequencies of small numbers "appear[ed] far more often to correspond to a negative than to a positive binomial: "in vital and demographic statistics, divisions either of space or time are generally governed by different environments which will vary the chances of an individual falling into them, and so we may expect that as a rule negative binomials will occur in place of the Poisson."

4.5. An Early Extra-Poisson Model: Greenwood and Yule 1920

The pioneering work of Greenwood and Yule (1920) was motivated by the challenge of describing the distribution of multiple attacks of disease or of repeated accidents in the same persons. It was the first to move away from the negative signs of n and q and instead viewed the "negative" binomial in a more positive and helpful way: *as a mixture of Poisson distributions with μ's that vary according to a gamma distribution*. Sadly, some software packages that fit negative binomial models use parameterizations that do not sufficiently highlight this heterogeneity.

We end our account with examples where Greenwood and Yule's or other forms of extra-Poisson variation should have been considered a priori.

5. A Broader View: Applications Involving Human Activities/Behavior

5.1. Feller's 1950 Story, Revised by the More Recent "Bigger" Picture

The first example concerns a randomness story that older statisticians were brought up on, but that in light of the bigger-picture data recently retrieved from the archives, has had to be revised.

Feller's classic textbook (1950) showed data from four contexts where where the Poisson distribution seemed to fit well: Rutherford and Geiger's radioactive disintegrations, chromosome interchanges produced by X-rays, telephone connections to a wrong number; and "flying-bomb hits on London." The spatial distribution of 537 of these over 576 squares of 1/4 square kilometer each was published by Clarke in an actuarial journal in 1946. He took the close agreement between the observed and expected and numbers of squares containing 0, 1, 2, 3, ... etc. flying bombs as "a very neat example of conformity to the Poisson law [that] might afford material to future writers of statistical text-books." Feller immortalized the surprisingly good fit: "It is interesting to note that most people believed in a tendency of the points of impact to cluster. If this were true, there would be a higher frequency of areas with either many hits or no hit and a deficiency in the intermediate classes. The Table indicates *perfect randomness* and homogeneity of the area; we have here an instructive illustration of the established fact that to the untrained eye randomness appears as regularity or tendency to cluster."

However, recent work by Shaw and Shaw (2019), using newly-assembled data from a larger area, showed that "overall, the distribution of V-1s was not uniformly random over London but was skewed to the south. We have found that Clarke's analysis is correct, but only within this southern region. He would have been well aware of this. During late 1944, British military intelligence had noticed that the V-1s tended to fall short of the center of London, on less densely populated suburbs. To save lives, they therefore, attempted to increase this shortfall using misinformation."

An even greater degree of nonuniformity emerges from Figure 5 of Evans and Delaney (2018), based on their new and independently collected data from an even wider area. One wonders if Clarke was being less innocent and more selective than he let on. In addition, statisticians must now ask themselves how they could have seduced into accepting that it was truly random. In hindsight, it is obvious that the data were not extensive enough to show the nonuniformity—a classic example where because of lack of power, a null hypothesis that is at variance with logic is nevertheless "accepted." The data are another example of what Bortkewitsch meant to convey by his "Law": with an overall mean of just 537/576 = 0.93 hits per square, any real but subtle geographic variation would be masked by the large Poisson variation.

5.2. Avoiding "Fake Standard Errors"

A careful rereading of the writings of Gosset, Erlang, Whitaker, Keynes, Yule and Greenwood, and Fisher should dampen the modern-day fervor to analyze *any* series of counts—or concentrations of CD4 and white blood cells that masquerade as "counts"—with (what we prefer to refer to as) Poisson's "Law of *Tight* Numbers."

Although the authors from more than a century ago have already warned that extra-Poisson variation should be the default model, the numbers of instances where modern-day investigators cling to too-tight "pure-Poisson" standard errors continue to grow. Worse still, since investigators have increasing access to Big Data, the artificially-small standard errors from these overly-tight models increase the risk of making noise look like signal.

One example is a common study design used in motor-vehicle epidemiology, where the numbers of motor vehicle crashes occurring on days when a certain factor or condition of concern is present are compared with those seven days before and seven days after that day of concern. For example, if the day of concern were St. Patrick's Day (March 17), the two comparison days would be March 10 and 24. The prevailing statistical model assumes that in year i, the numbers of events on March 10 and 24 are two realizations of the *same* pure Poisson random variable with expectation μ_i, and that the observed number of events on March 17 is a realization of a Poisson random variable with expectation $\theta \times \mu_i$, where θ is the rate ratio parameter of interest. This model treats the numbers of crashes on the two comparison days in the same year as if they were the cell counts for two drops of the same well-mixed liquid studied by Gosset, or the radioactivity counts in two one-second intervals, just 14 sec apart. It treats the number on the day of concern as if Rutherford has taken a middle count when the radioactivity source was moved by a distance that produced θ times as much radioactivity. In traffic crash epidemiology, however, the large number of environmental and social conditions cannot be kept constant the way they are in Gosset's and Rutherford's laboratories.

Over 40 years in the USA, some 4494 fatal motorcycle crashes occurred on the 494 nights with a full moon and 8535 on the 988 control nights without a full moon (seven days before/after), a rate ratio of 1.05 (Redelmeier and Shafir 2017). If, as the authors did, we treat the 494 pairs of counts that summed to 8535 as realizations from 494 different Poisson distributions with expectations μ_1 to μ_{494}, and the 494 counts that summed to 4494 as realizations from 494 Poisson distributions with expectations $\mu_1 \times \theta$ to $\mu_{494} \times \theta$, then the standard error for $\log \hat{\theta}$ would be $(1/4494 + 1/8535)^{1/2} = 0.018$. It would produce a 95% confidence interval for θ of 1.02 to 1.09, and a p-value of approx. 0.005. If however, instead of imposing this overly tight pure-Poisson model, we go by the data rather than the model (Palayew, Harper, and Hanley 2021), the real standard error is 1/3rd larger, and thus, the confidence interval is 1/3rd wider.

In this instance, the more realistic SE weakens the evidence for the lunacy claim, that is, that "the full moon is associated with an increased risk of fatal motorcycle crashes." However, the choice of model has much bigger implications if, instead of fatal motorcycle crashes, one uses the same design to study the intensity with which a well-known political figure has tweeted. The dataset, *http://www.biostat.mcgill.ca/hanley/Gosset/tweetsOnFullMoonDays.txt*, compiled from various archives, covers a total of 148 moons (and 56,214 tweets) from 2009 to the end of 2020. Some 2132 tweets were emitted on the 148 days with a full moon and 3708 on the 296 control days without a full moon (seven days before/after), a rate ratio of 1.15. Its log is 0.14 and the associated "pure-Poisson" standard error is $(1/2{,}132 + 1/3{,}708)^{1/2}= 0.027$, yielding a z-statistic of 5.14, and a 2-sided p-value of 2.7×10^{-7}. Based on this model-based SE, the 95% CI for the rate ratio is 1.09–1.21. Five different robust standard errors (Palayew, Harper, and Hanley 2021) that respect the lunar "matching" range from 0.08 to 0.10 and suggest that the rate ratio of 1.15 could have arisen from natural fluctuations within the same lunar month.

We leave it to readers to decide which is/are the fake SEs, the "pure-Poisson" one or the robust ones that are at least three times larger. Instead of matching on the lunar month, some may wish to fit regression models for extra-Poisson variation, such as the quasi-Poisson model, and the negative-binomial model of Greenwood and Yule; others may conclude that this is an example of the occasional dataset that defies all statistical regression models.

6. In Conclusion

Few of today's counts arise from controlled environments at brewing and physics laboratories. Thus, we should probably dampen the modern-day fervor to analyze *any* series of counts—or concentrations of CD4 and white blood cells that masquerade as "counts"—with (what we prefer to refer to as) Poisson's "Law of *Tight* Numbers." One hundred years ago, it was not through black box model diagnostics or computing power that the early users of the Poisson distribution understood its limitations and the need for broader distributions. Rather, it was because they were close to their data, and understood the biological and physical processes that produced these counts.

Recognizing and dealing with extra-Poisson—and indeed "extra-binomial" and "extra-Gaussian"—variation is increasingly important. The advice by authors such as McCullagh and Nelder (1989) and Gelman (2007, 2022) that "overdispersion is the rule rather than the exception" is welcome, and needs to be practiced. Otherwise, with greater access to Big Data, the artificially-small ("fake") standard errors from overly-tight models increase the risk of making noise look like signal (Palayew, Harper, and Hanley 2021).

Today's real-world applications involving counts relating to human behavior are a good deal more complex than the typical textbook examples, and the data are far less obedient than those generated under the tightly controlled conditions in science laboratories. The "pure-Poisson" models of Gosset, Rutherford, and Erlang can no longer be considered the "default." They are however, a natural and historical point of departure. And, thanks to the availability of modern communication media, the teaching of them and their derivations can be made more engaging and realistic.

Acknowledgments

The authors thank the editor, the associate editor and the two referees for their very helpful comments on an earlier version of this article.

References

Abbe, E. (1879), "Ueber Blutkörper-Zählung," *Sitzungsberichte der Jenaischen Gesellschaft fur Medicin und Naturwissenschaft fur das Jahr 1878*, Jena: Verlag von Gustav Fischer. [364]

Alam, S., and Hanley, J. (2018), "Betting on a Royal Birth: How Many Babies Will be Born the Same Day?" *Significance Magazine*, available at *https://www.significancemagazine.com/culture/585-betting-on-a-royal-birth-how-many-babies-will-be-born-the-same-day* [368]

Bortkewitsch, L. (1898), *Das Gesetz der kleinen Zahlen*, Leipzig: B.G. Teubner. [364]

Clarke, R. D. (1946), "An Application of the Poisson Distribution," *Journal of the Institute of Actuaries*, 72, 481–481. [369]

Clausius, R. (1858), "Ueber die mittlere Länge der Wege, welche bei der Molecularbewegung gasförmiger Körper con den einzelnen Molecülen zurückgehgt werden; nebst einigen anderen Bemerkungen über die mechanische Wärmetheorie," *Annals of Physics*, 105, 239–258. [364]

Cole, J. H. (2010), "Updating a Classic: 'The Poisson Distribution and the Supreme Court' Revisited," *Teaching Statistics*, 32, 78–80. [368]

David, F. N. (1962), *Games, Gods and Gambling*, London: Griffin. [364]

de Moivre, A. (1718), *The Doctrine of Chances*, Problems 5-8, 14-21, available at *https://books.google.ca/books?id=3EPac6QpbuMC*. [363,364]

Erlang, A. K. (1909), "The Theory of Probabilities and Telephone Conversations." First published as 'Sandsynlighedsregning og Telefonsamtaler' in *Nyt Tidsskrift for Matematic, B*, 20, 33–39. English translation in: The Life and Works of A.K. Erlang, Transactions of the Danish Academy of Technical Sciences, eds. E. Brockmeyer, H. L. Halstrøm, and A. Jensen, available at *https://web.archive.org/web/20110719123704/http://oldwww.com.dtu.dk/teletraffic/erlangbook/pps131-137.pdf* [367,368]

——— (1920), "Telephone Waiting Times." First published as 'Telefon-Ventetider. Et Stykke Sandsynlighedsregning' in *Matematisk Tidsskrift B*, 31, 25–42. English translation in: The Life and Works of A. K. Erlang, Transactions of the Danish Academy of Technical Sciences, eds. E. Brockmeyer, H. L. Halstrøm, and A. Jensen, available at *https://web.archive.org/web/20110719123704/http://oldwww.com.dtu.dk/teletraffic/erlangbook/pps156-171.pdf*. Note: There is a 'cut-and-paste' error in the English translation in the sentence that begins *Vice versa*. [367]

Evans, S. G., and Delaney, K. B. (2018), "The V1 (Flying Bomb) Attack on London (1944–1945); The Applied Geography of Early Cruise Missile Accuracy," *Applied Geography*, 99, 44–53. [369]

Feller, W. (1950), *An Introduction to Probability Theory and its Applications* (Vol. I), New York: Wiley. [369]

Fisher, R. A., Thornton, H. G., Mackenzie W. A. (1922), "The Accuracy of the Plating Method of Estimating the Density of Bacterial Populations," *Annals of Applied Biology*, ix, 325–359. [364,366,368]

Gelman, A. (2022), "Statistical Modeling, Causal Inference and Social Science," Blog, available at *https://statmodeling.stat.columbia.edu*. [370]

Gelman, A., and Jennifer Hill. J. (2007), *Data Analysis Using Regression and Multilevel/Hierarchical Models*," Cambridge: Cambridge University Press. [370]

Gosset, W. S. (1907), "Haemacytometer Paper: Yeast-Cell Counts" (Front cover of notebook). See "Why Gosset Chose the Pen-name 'Student,'" Available at *http://www.biostat.mcgill.ca/hanley/Student/750samplesOf4/* [364,368]

Greenwood, M„ and Yule, G. U. (1920), "An Inquiry into the Nature of Frequency Distributions Representative of Multiple Happenings with Particular Reference to the Occurrence of Multiple Attacks of Disease or of Repeated Accidents," *Journal of the Royal Statistical Society*, 83, 255–279. [369]

Haight, F. A. (1967), *Handbook of the Poisson Distribution*, New York: Wiley. [364]

Hanley, J. A. (1984), "Lotteries and Probability: Three Case Reports," *Teaching Statistics*, 1984, 88–92. [364]

——— (2018), "Statistical Behaviors: Personal and Computer-Aided Observations," *SLAS Discovery*, 23, 440–447. [364]

——— (2022), "What Events Are These?," available at *http://www.biostat.mcgill.ca/hanley/mysteryData/* [368]

Hanley, J. A., Julien, M., and Moodie, E. M. M. (2008), "Student's z, t, and s: What if Gosset had R?" *The American Statistician*, 62, 64–69. [364]

Keynes, J. M. (1921), *Treatise on Probability*, London: Macmillan and Co. [368]

Marsden, E., and Barratt, T. (1911a), "The Probability Distribution of the Time Intervals of α Particles with Application to the Number of α Particles Emitted by Uranium," *Proceedings of the Physical Society of London*, 23, 367–373. [366]

——— (1911b), "The α Particles Emitted by the Active Deposits of Thorium and Actinium," *Proceedings of the Physical Society of London*, 24, 50–61. [366]

McCullagh P., and Nelder, J. A. (1989), *Generalized Linear Models*, Chapman and Hall, Boca Raton. [370]

Newbold, E. M. (1927), "Practical Applications of the Statistics of Repeated Events, Particularly to Industrial Accidents," *Journal of the Royal Statistical Society*, 90, 487–547. [364]

Newcomb, S. (1860), "Notes on the Theory of Probabilities," *The Mathematical Monthly*, 2, 134–140. [364]

Office of National Statistics (2020), "Excess Winter Mortality in England and Wales: 2019 to 2020 (provisional) and 2018 to 2019 (final)," available at *https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/bulletins/excesswintermortalityinenglandandwales/2019to2020provisionaland2018to2019final* [368]

Palayew, A., Harper, S., and Hanley, J. A. (2021), "Towards Reducing the Possibility of False Positive Results in Epidemiologic Studies of Traffic Crashes," *Chance*, 34, 44–52. [368,370]

Pearson, K. (1900), "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, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, Series 5*, 50, 157–175. [366]

Phillips, D. P., Brewer, K. M., and Wadensweiler, P. (2010), "Alcohol as a Risk Factor for Sudden Infant Death Syndrome (SIDS)," *Addiction*, 106, 516–525. [368]

Poisson, S. (1837), *Recherches sur la Probabilité des Jugements en Matière Criminelle et en Matière Civile Précédés des Règles Générales du Calcul des Probabilités* (Bachelier, Paris), available at *https://gallica.bnf.fr/ark:/12148/bpt6k110193z/* [363]

Preece, D. A., Ross, G. J. S., and Kirby, S. P. (1988), "Bortkewitsch's Horse-kicks and the Generalised Linear Model," *The Statistician*, 37, 313–318. [364]

Redelmeier, D. A., and Shafir, E. (2017), "The Full Moon and Motorcycle Related Mortality: Population Based Double Control Study," *British Medical Journal*, 359, 1–7. [369]

Rutherford, E., Geiger, H., and Bateman H. (1910), "The Probability Variations in the Distribution of α Particles," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 20, 698–70. [365]

Schweidler, E. (1905), "Uber Schwankungen der radioaktiven Umwandlung," in P̂remier Congrès international pour l'étude de la radiologie et de l'ionisation, avaialble at *https://archive.org/details/premiercongrsin03unkngoog/page/*. [365]

Shaw, L. P., and Shaw, L. F. (2019), "The Flying Bomb and the Actuary," *Significance Magazine*, available at *https://doi.org/10.1111/j.1740-9713.2019.01315.x* [369]

Snow, E. C. (1911), "Note on the Paper by Prof. E. Rutherford, F.R.S., and H. Geiger, Ph.D., on 'The Probability Variations in distribution of α Particles,'" *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 22, 198–200. [366]

Soper, H. E. (1914), "Poisson's Exponential Binomial Limit," *Biometrika*, 10, 25–35. [367]

Stigler, S. M. (1982), "Poisson on the Poisson Distribution," *Statistics and Probability Letters*, 1, 33–35. [363]

——— (2008), "Karl Pearson's Theoretical Errors and the Advances They Inspired," *Statistical Science*, 23, 261–271. [365]

——— (2019), "Data have a Limited Shelf Life," *Harvard Data Science Review* 1.2, avaialble at *https://hdsr.mitpress.mit.edu/pub/iu26pfw1*. [364]

Student (1907), "On the Error of Counting with a Haemacytometer," *Biometrika*, 5, 351–360. [364]

——— (1919), "An Explanation of Deviations from Poisson's Law in Practice," *Biometrika*, 12, 211–215. [368]

Whitaker, L. (1914), "On the Poisson Law of Small Numbers," *Biometrika*, 10, 36–71. [368]

Whitworth, W. A. (1870), *Choice and Chance* (2nd ed.), London: Bell and Daldy. [364]

Winsor, C. P. (1947), "Das gesetz der kleinen zahlen," *Human Biology*, 19, 154–161. [364]

Zweig, J. P., and Csank, J. Z. (1978), "The Application of a Poisson Model to the Annual Distribution of Daily Mortality at Six Montreal Hospitals," *Journal of Epidemiology and Community Health*, 32, 206–211. [368]