Lecture 13: The Law of Small Numbers

Relevant textbook passages:

Pitman [15]: Sections 2.4, 3.5, 3.8, 4.2

Larsen–Marx [14]: Sections 3.8, 4.2, 4.6

13.1 Poisson's Limit

The French mathematician Siméon Denis Poisson (1781–1840) is known for a number of contributions to mathematical physics. But what we care about today is his discovery regarding the binomial distribution. We have seen that the Standard Normal density can be use to approximate the binomial probability mass function. In fact, if X has a Binomial(n,p) distribution, which has mean $\mu = np$ and standard deviation $\sigma_n = \sqrt{np(1-p)}$, then for each k, letting $z_n = (k-\mu)/\sigma_n$, we have

$$\lim_{n\to\infty} \left| P(X=k) - \frac{1}{\sqrt{2\pi}\sigma_n} e^{-\frac{z^2}{2}} \right| = 0.$$

See Theorem 10.5.1.

Poisson discovered another peculiar, but useful, limit of the binomial distribution.[1] Fix $\mu > 0$ and let

$$X_n \sim \text{Binomial}(n, \mu/n).$$

Then $\boldsymbol{E}\, X_n = \mu$ for each n, but the probability of success is μ/n, which is converging to zero. As n gets large, for fixed k we have

$$\begin{aligned}
P(X_n = k) &= \binom{n}{k} \left(\frac{\mu}{n}\right)^k \left(1-\frac{\mu}{n}\right)^{n-k} \\
&= \frac{\overbrace{n(n-1)(n-2)\cdots(n-k+1)}^{k \text{ terms}}}{k!} \left(\frac{\mu}{n}\right)^k \left(1-\frac{\mu}{n}\right)^{n-k} \\
&= \frac{\frac{n}{n}\frac{n-1}{n}\cdots\frac{n-k+1}{n}}{k!} \mu^k \left(1-\frac{\mu}{n}\right)^{n-k} \\
&= \frac{(1-\frac{1}{n})(1-\frac{2}{n})\cdots(1-\frac{k-1}{n})}{k!} \mu^k \left(1-\frac{\mu}{n}\right)^{-k} \left(1-\frac{\mu}{n}\right)^{n} \\
&\xrightarrow[n\to\infty]{} \frac{1}{k!} \cdot \mu^k \cdot 1 \cdot e^{-\mu} \\
&= e^{-\mu}\frac{\mu^k}{k!}.
\end{aligned}$$

This result was known for a century or so as **Poisson's limit**. Note that if $k > n$, the binomial random variable is equal to k with probability zero. But the above is still a good approximation of zero.

[1] von Bortkiewicz [22] gives this reference: Zu vergleichen Poisson, Recherches sur la probablités des jugements, Paris 1837. n^o 81, p. 205–207. Crathorne [8] gives a little more detail on Poisson's contribution.

You may recognize the expression $\frac{\mu^k}{k!}$ from the well known expression

$$\sum_{k=1}^{\infty} \frac{\mu^k}{k!} = e^{\mu},$$

which is obtained by taking the infinite Taylor series expansion of the exponential function around 0. See, e.g., Apostol [2, p. 436].

13.2 The Poisson(μ) distribution

Larsen–Marx [14]: Section 4.3
Pitman [15]: p. 121

The **Poisson(μ) distribution** is a discrete distribution that is supported on the nonnegative integers, which is based on the Poisson limit. For a random variable X with the Poisson(μ) distribution, where $\mu > 0$, the probability mass function is

$$P(X = k) = p_\mu(k) = e^{-\mu}\frac{\mu^k}{k!}, \qquad (k = 0, 1, 2, 3, \dots).$$

Table 13.1 gives a sample of the values for various μ and k.

k	$\mu = .25$	$\mu = .5$	$\mu = 1$	$\mu = 2$	$\mu = 4$	$\mu = 8$
0	0.7788	0.6065	0.3679	0.1353	0.01832	0.0003355
1	0.1947	0.3033	0.3679	0.2707	0.07326	0.002684
2	0.02434	0.07582	0.1839	0.2707	0.1465	0.01073
3	0.002028	0.01264	0.06131	0.1804	0.1954	0.02863
4	0.0001268	0.001580	0.01533	0.09022	0.1954	0.05725
5	$6.338 * 10^{-6}$	0.0001580	0.003066	0.03609	0.1563	0.09160
6	$2.641 * 10^{-7}$	0.00001316	0.0005109	0.01203	0.1042	0.1221
7	$9.431 * 10^{-9}$	$9.402 * 10^{-}7$	0.00007299	0.003437	0.05954	0.1396
8	$2.947 * 10^{-10}$	$5.876 * 10^{-8}$	$9.124 * 10^{-6}$	0.0008593	0.02977	0.1396
9	$8.187 * 10^{-12}$	$3.264 * 10^{-9}$	$1.014 * 10^{-6}$	0.0001909	0.01323	0.1241
10	$2.047 * 10^{-13}$	$1.632 * 10^{-10}$	$1.014 * 10^{-7}$	0.00003819	0.005292	0.09926
11	$4.652 * 10^{-15}$	$7.419 * 10^{-12}$	$9.216 * 10^{-9}$	$6.944 * 10^{-6}$	0.001925	0.07219
12	$9.691 * 10^{-17}$	$3.091 * 10^{-13}$	$7.680 * 10^{-10}$	$1.157 * 10^{-6}$	0.0006415	0.04813

Table 13.1. The Poisson probabilities $p_\mu(k)$.

13.2.1 Remark The ratio of successive probabilities $p_\mu(k+1)/p_\mu(k)$ is easy to compute.

$$\frac{p_\mu(k+1)}{p_\mu(k)} = \frac{\mu}{k+1}.$$

So as long as $k+1 < \mu$, then $p_\mu(k+1) > p_\mu(k)$, but then $p_\mu(k)$ decreases with k. See Figure 13.1.

One can show that if $\mu > \lambda$, then the Poisson(μ) distribution stochastically dominates the Poisson(λ) distribution. (Stochastic dominance is defined and discussed in Section 12.2⋆.)

The next set of charts show how the Poisson distribution stacks up against the binomial.

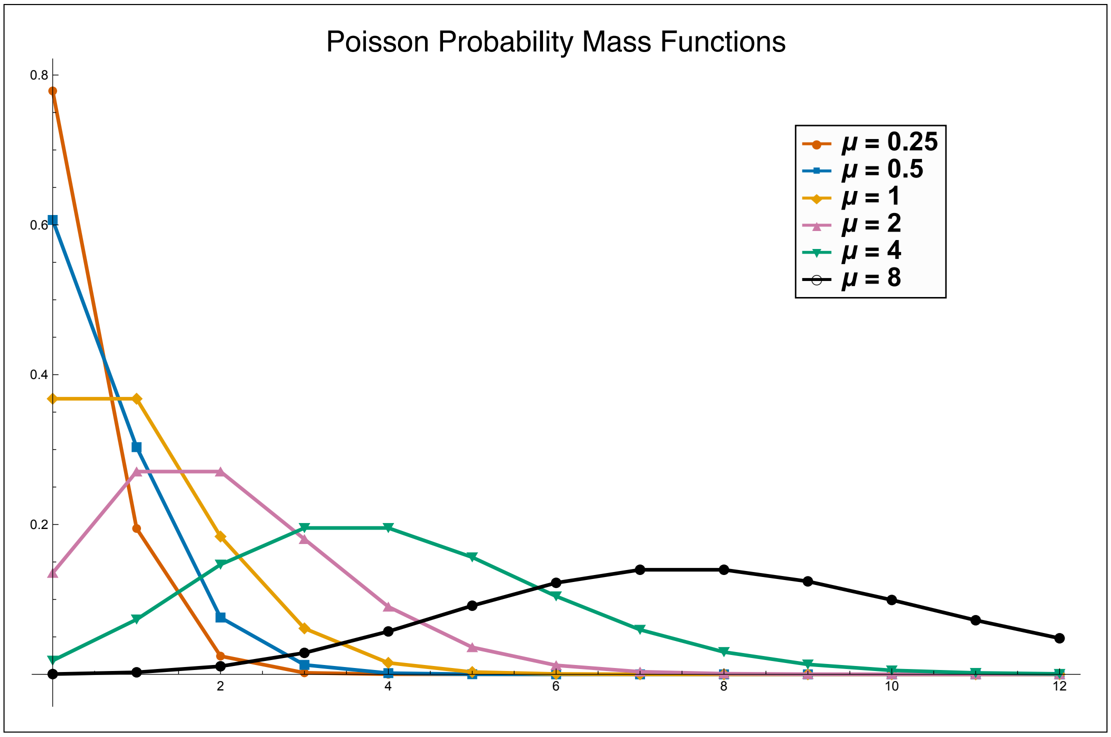

Figure 13.1.

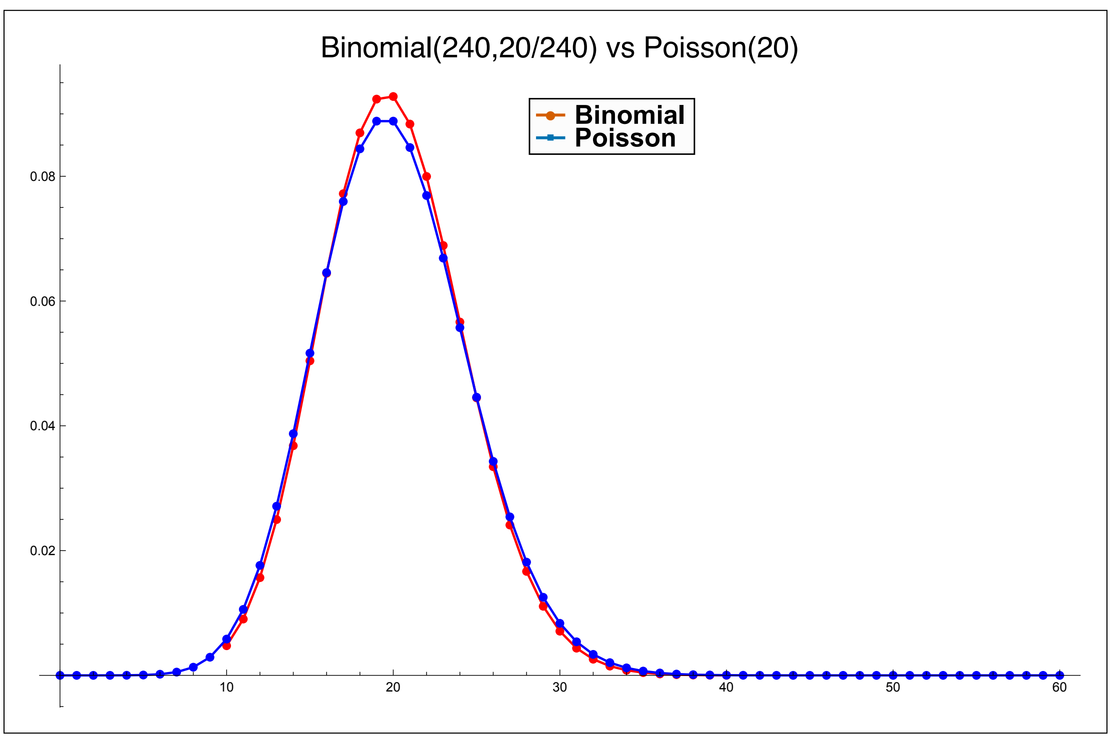

Figure 13.2.

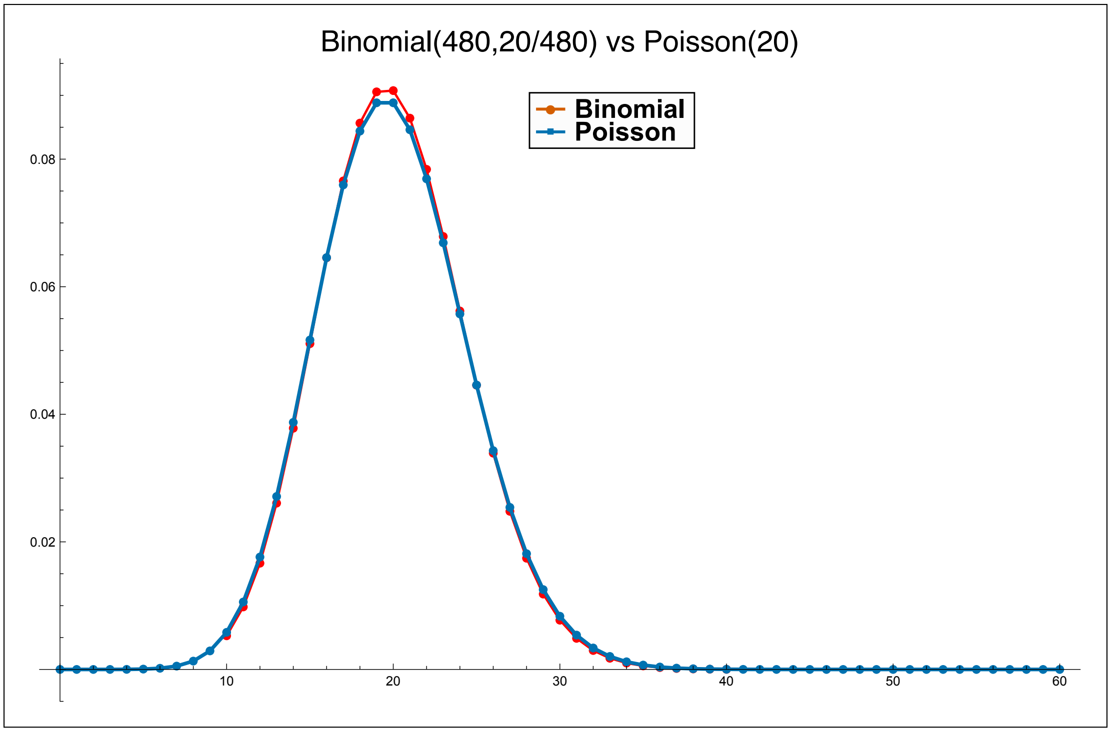

Figure 13.3.

13.3 The mean and variance of the Poisson distribution

If X has a Poisson(μ) distribution, then

$$\begin{aligned}\boldsymbol{E}\,X &= e^{-\mu}\sum_{k=0}^{\infty} k\frac{\mu^k}{k!} = e^{-\mu}\sum_{k=1}^{\infty} k\frac{\mu^k}{k!}\\ &= \mu e^{-\mu}\sum_{k=1}^{\infty}\frac{\mu^{k-1}}{(k-1)!} = \mu e^{-\mu}\sum_{j=0}^{\infty}\frac{\mu^j}{j!} = \mu.\end{aligned}$$

To compute the variance, let's use the identity $\boldsymbol{Var}\,X = \boldsymbol{E}(X^2) - (\boldsymbol{E}\,X)^2$. To compute this we first need $E(X^2)$. This leads to the awkward sum $\sum_{k=0}^{\infty} k^2 e^{-\mu}\frac{\mu^k}{k!}$, so we'll use the trick that Pitman [15, pp. 223–24] uses and write $X^2 = X(X-1) + X$. Start with $\boldsymbol{E}\big(X(X-1)\big)$:

$$\begin{aligned}\boldsymbol{E}\big(X(X-1)\big) &= e^{-\mu}\sum_{k=0}^{\infty} k(k-1)\frac{\mu^k}{k!} = e^{-\mu}\sum_{k=2}^{\infty} k(k-1)\frac{\mu^k}{k!}\\ &= \mu^2 e^{-\mu}\sum_{k=2}^{\infty}\frac{\mu^{k-2}}{(k-2)!} = \mu^2 e^{-\mu}\sum_{j=0}^{\infty}\frac{\mu^j}{j!} = \mu^2.\end{aligned}$$

So

$$\boldsymbol{Var}\,X = \boldsymbol{E}(X^2) - (\boldsymbol{E}\,X)^2 = \Big(\boldsymbol{E}\,X(X-1) + \boldsymbol{E}\,X\Big) - (\boldsymbol{E}\,X)^2 = (\mu^2+\mu) - \mu^2 = \mu.$$

13.4 Sums of independent Poisson random variables

13.4.1 Proposition *Let X be Poisson(μ) and Y be Poisson(λ) and independent. Then $X+Y$ is Poisson($\mu+\lambda$).*

Proof: Convolution:

$$\begin{aligned} P(X+Y=n) &= \sum_{j=0}^{n} P(X=j,\ Y=n-j) \\ &= \sum_{j=0}^{n} P(X=j)P(Y=n-j) \\ &= \sum_{j=0}^{n} e^{-\mu}\frac{\mu^j}{j!}e^{-\lambda}\frac{\lambda^{n-j}}{(n-j)!} \\ &= e^{-(\mu+\lambda)}\frac{(\mu+\lambda)^n}{n!}, \end{aligned}$$

where the last step comes from the binomial theorem:

$$(\mu+\lambda)^n = \sum_{j=0}^{n} \frac{n!}{j!(n-j)!}\mu^j\lambda^{n-j}.$$

∎

13.5 Random scattering

Pitman [15]: pp. 238–233

Pitman [15, pp. 238–233] discusses a remarkable result that shows how the Poisson distribution arises when points are scattered at random throughout a some region in space or time. Let $A = I_1 \times \cdots \times I_k$ be a bounded rectangle in $\boldsymbol{R}^{\mathrm{k}}$. A **scatter** is a random experiment that "scatters/hits/chooses" a random finite number of points at random throughout A.

For each n the n^{th} subdivision of A partitions A into $m = 2^{kn}$ subrectangles by dividing each interval I_j into 2^n subintervals of equal length, $j = 1,\dots,k$.

For each Borel subset B of A, let N_B be the number of points in B, and let $\operatorname{vol}(B)$ denote the k-dimensional volume of B.

13.5.1 Poisson Scatter Theorem *Assume the scatter satisfies these three conditions:*

1. ***Nondegeneracy****: The scatter is nondegenerate, that is, there is strictly positive probability that $N_A > 0$.*

2. ***Distinctness****: The probability is zero that two points coincide.*

3. ***Random scattering****: For each n, in the n^{th} subdivision, the events $E_j^n = (\textit{subrectangle } j \textit{ is hit})$, $j = 1,\dots,2^{kn}$, are independent and all have the same probability.*

Then there is a $0 < \lambda < \infty$ such that

1. *for every Borel subset of A, the random variable N_B has a Poisson distribution with parameter $\lambda \times \operatorname{vol}(B)$, and*

2. *if $B_1,\dots,B_m$ are disjoint, then the random variables $N_{B_1},\dots,N_{B_m}$ are stochastically independent.*

13.5.2 Definition *A scatter satisfying the conclusion of Theorem 13.5.1 is called a* ***Poisson scatter****.*

13.5.3 Remark While the proof of the theorem relies on Poisson's approximation, the result is an exact one. It says that each N_B has a Poisson distribution, not that its distribution is approximately Poisson.

13.5.4 Remark There are technical mathematical issues that Pitman ignores, largely because the answers all lie outside the scope of this course. Here are a few.

What is the sample space for the scatter experiment? Well, if one point is scattered, the outcome is just an element of A. It two are scattered, then the outcome is an element of A^2, etc. But wait, what if the scatter is empty? That is, there are no points. We need to append some outcome to the sample to indicate that, and we might as well call it A^0, so the natural candidate for the sample space is $S = \bigcup_{n=0}^{\infty} A^n$. This is awkward to work with, but mathematicians have invented ways beyond the scope of this class to deal with it.

The next question is to decide that the events are. We already know that for each Borel set we want N_B, the number of scattered points in B to be a random variable, so we need enough events to make that true.

Now the question arises, is there any probability measure on the set of events that satisfies the hypotheses of the theorem. I hate say once more that the answer is yes, but the proof of it is beyond the scope of this course.

The theorem guarantees that if we can find such a probability on the sample space, it must then have the additional properties stated in the conclusion.

Outline of a proof of Theorem 13.5.1: (Cf. Pitman [15, p. 233])

Assume we have solved the technical issues have and have found a probability measure with the desired properties. By throwing out the event of probability zero that two scatter points coincide, we may assume that all the points are distinct.

Or should I just do this a.s.?

Let the random variable N be the total number of "hits" on the rectangle A. For each n, let H_n be the number of subrectangles in the n^{th} subdivision that have been hit. Note that

$$H_{n+1} \geqslant H_n, \tag{1}$$

since the hits in any n^{th}-subdivision rectangle E will also belong to one of E's 2^k subrectangles in the $n+1^{\text{th}}$ subdivision.

Since the points are distinct, for large enough n all the points will be in distinct n^{th}-subdivision rectangles, $H_n \uparrow N$ on the sample space. Thus

$$(N \leqslant t) = \bigcap_{n=1}^{\infty} (H_n \leqslant t),$$

so by Proposition 2.4.3,

$$P(N \leqslant t) = \lim_{n\to\infty} P(H_n \leqslant t).$$

So the distribution of N can be found by taking the limit as $n \to \infty$ of the distribution of H_n.

By assumption (3) of random scattering, each event E_j^n = (subrectangle j is hit), $j = 1, \dots, 2^{kn}$, are independent and all have the same probability. Call this probability p_n. Then H_n is just the number of the events E_j^n, that have occurred, which has a Binomial$(2^{nk}, p_n)$ distribution with mean $\boldsymbol{E}\, H_n = 2^{kn} p_n$.

Since **expectation is a positive linear operator**, (1) tells us that $\boldsymbol{E}\, H_n = 2^{kn} p_n$ is an increasing sequence. So the sequence is either unbounded or converges to a limit $\mu < \infty$.

Do Pitman's exercise.

Pitman leaves it as an exercise to show that the limit μ is finite. In this case, the distribution of the limit random variable H must be Poisson with parameter μ:

$$P(N \leqslant k) = \lim_{n\to\infty} P(H_n \leqslant k = F_\mu(k)),$$

where F_μ is the Poisson(μ) cumulative distribution function—the last equality is Poisson's limit. Now set $\lambda = \mu/\operatorname{vol}(A)$. Then $N = N_A$ has a Poisson distribution with parameter $\lambda \operatorname{vol}(A)$.

The remainder of the theorem proceeds by showing that for any subrectangle E in a subdivision, the random variable N_E has a Poisson distribution with parameter $\lambda \operatorname{vol}(E)$. The proof is the same as for A.

Since the events that subrectangles in a given subdivision are independent, any disjoint events $B_1, \dots, B_m$ that are unions of finitely many subrectangles from the same have the property that the random variables N_{B_i} are independent. Since the sum of independent Poisson random variables is also Poisson, we see that for each B_i has a Poisson distribution with parameter $\lambda \operatorname{vol}(B_i)$.

To get the result for any Borel set involves reasonably standard measure-theoretic techniques (e.g., the Carathéodory extension theorem) that are beyond the scope of this class. ∎

13.6 The Law of Small Numbers

The term "Law of Small Numbers" does not have the same universal agreement on it meaning as the Law of Large Numbers. The expression was coined in 1898, when Ladislaus von Bortkiewicz[2] published *Das Gesetz der kleinen Zahlen* [The Law of Small Numbers] [22]. He described a number of observations on the frequency of occurrence of **rare events** that appear to follow a Poisson distribution. Crathorne [8] discusses controversies surrounding Bortkiewicz's use of the term "Law of Small Numbers" and alternative descriptions, such as the Law of Rare Events. Feller [10, p. 159n] says, "These are misnomers which proved detrimental to the realization of the fundamental role of the Poisson distribution." Nowadays some of my colleagues have used the term at Faculty Board meetings to refer to the fact that when sample sizes are small, confidence intervals are wide (see Lecture 19). The psychologists Tversky and Kahneman [21] use the term to describe the mistaken intuition of many psychologists that small sample sizes are representative of large populations. But we shall use the term in the same way as von Bortkiewicz.

Here is a mathematical model to explain his observations. It is closely related to the Poisson scatter, but is a little different.

The random experiment

The experiment is to "scatter" a *known* number m of numbered balls at random among n bins, labeled $1, \dots, m$.[3] We can think of the outcome of this experiment as a multinomial random vector with m different *types* of outcomes (namely, the bins) and n independent trials (one for each ball). (Recall Section 3.9.) The components N_b, $b = 1, \dots, m$, of the vector

$$\boldsymbol{N} = (N_1, \dots, N_m)$$

count how many of the experiments (balls) had outcome b (hitting bin b). Each bin is equally likely to be hit, so

$$N_b \sim \operatorname{Binomial}\Bigl(n, \frac{1}{m}\Bigr),$$

but the N_b's are not independent, since

$$\sum_{b=1}^{m} N_b = n.$$

The average number μ of balls per bin is

$$\mu = \frac{n}{m},$$

[2] As an interesting aside, the OED [19] credits von Bortkiewicz in 1917 for first using the term *stochastic* [German, *stochastik*] to mean *random.* Before that, its meaning was *conjectural.*

[3] Alternatively we could treat the number n as a random variable, and proceed conditional on its value.

Poisson's limit tells us that, fixing $\mu = n/m$, if n (and m) are large, then each component N_b of $\boldsymbol{N}$ has an approximate Poison distribution with intensity μ.

But the Law of Small Numbers is not concerned with the distribution of N_b. Instead we are interested in X_k, $k = 0, \dots, m$, the random variables that count the number of bins with k hits,

$$X_k = \#\{b : N_b = k\},$$

so

$$\sum_{k=0}^{m} X_k = m.$$

Let $p_\mu(k)$ be the Poisson(μ) probability mass function.

13.6.1 Proposition (The Law of Small Numbers) *Fixing the ratio $\mu = n/m$ and the count k, if n and m are large enough, then with high probability,*

$$X_k = \text{the number of bins with } k \text{ hits} \approx m p_\mu(k). \tag{2}$$

Before I explain why the Law of Small Numbers is true, let me give some examples of its application.

13.7 The Law of Small Numbers in practice

There are many stories of data that fit this model, and many are told without any attribution. Many of these examples can ultimately be traced back to the very carefully written book by William Feller [9] in 1950. (I have the third edition, so I will cite it.) Feller attributes some of his examples to John Maynard Keynes' [13] monograph.

• During the Second World War, Nazi Germany used unmanned aircraft, the V1 Buzz Bombs, to attack London. (They weren't drones, since they were never designed to return or to be remote controlled. Once launched, where they came down was moderately random.) Feller [10, pp. 160–161] cites R. D. Clarke [6] (an insurance adjuster for The Prudential), who reports that a region of 144 square kilometres of South London was divided into 576 sectors of about 1/4 square kilometre, and the number of hits in each sector was recorded. The total for the region was 537 Buzz Bombs. For more details, see the article by Shaw and Shaw [18].

How does this fit our story? Consider each of the $n = 537$ Buzz Bombs as a ball and each of the $m = 576$ sectors as a bin. Then $\mu = n/m = 537/576 = 0.9323$. Our model requires that each Buzz bomb is equally likely to hit each sector. I don't know if that is true, but that never stops an economist from proceeding as if it might be true. In fact Shaw and Shaw [18] suggest that Clarke was investigating whether there were suspicious clusters of hits. The Law of Small Numbers then predicts that the number of districts with k hits should be approximately

$$576 p_{0.9323}(k).$$

Here is the actual data compared to the Law's prediction:

No. of Hits k:	0	1	2	3	4	$\geqslant 5$
No. of Sectors with k hits:	229	211	93	35	7	1
Law prediction:	226.7	211.4	98.5	30.6	7.1	1.6

That looks amazingly close. Later on in Lecture 23 we will learn about the χ^2-test, which gives a quantitive measure of how well the data conform to the Poisson distribution, and the answer will turn out to be, "very." (The p-value of the χ^2-test statistic is 0.95. For now, you may think of the p-value as a measure of goodness-of-fit with 1 being perfect.)

One of the things you should note here is that there is a category labeled $\geqslant 5$ hits. What should prediction be for that category? It should be $n\sum_{k=5}^{\infty} p_\mu(k)$, which it is. On the other hand, you can count and figure out that there is exactly one sector in that category and it had seven hits. So the extended table should read as follows

No. of Hits k:	0	1	2	3	4	5	6	7	$\geqslant 8$
No. of Sectors with k hits:	229	211	93	35	7	0	0	1	0
Law prediction:	226.7	211.4	98.5	30.6	7.1	1.3	0.2	0.03	0.004

As you can see, it doesn't look quite so nice. The reason is that Poisson approximation is for valid for smallish k. (A rule of thumb is that the model should predict a value of at least 5 sectors for it to be a good approximation.)

- The bins don't have to be geographical, they can be temporal. So distributing a fixed average number of events per time period over many independent time periods, should also give a Poisson distribution. Indeed Chung [5, p. 196] cites John Maynard Keynes [13, p. 402], who reports that von Bortkiewicz [22] reports that the distribution of the number of cavalrymen killed from being kicked by horses is described by a Poisson distribution! Here is the table from von Bortkiewicz's book [22, p. 24]. It covers the years 1875–1894 and fourteen different Prussian Cavalry Corps. So there are $m = 280 = 14 \times 20$ CorpsYears (bins). Each CorpsYear corresponds to a bin. There were $n = 196$ deaths (balls). So $\mu = n/m = 196/280 = 0.70$, so with $n = 280$ our theoretical prediction of the number of CorpsYears with k deaths is $280p_{.7}(k)$. Unfortunately, the numbers of expected deaths as reported by von Bortkiewicz, do not agree with my calculations. I will look into this further. Keynes [13, p. 404] complains about von Bortkiewicz and his reluctance to describe his results in "plain language," writing, "But like many other students of Probability, he is eccentric, preferring algebra to earth."

Number of CorpsYears with N deaths			
N	Actual	Bortkiewicz's Theoretical	My Theoretical
0	144	143.1	139.0
1	91	92.1	97.3
2	32	33.3	34.1
3	11	8.9	7.9
4	2	2.0	1.4
5 +	—	0.6	0.2

(By the way, the p-value of the χ^2-statistic for my predictions is 0.80.)

- Keynes [13, p. 402] also reports that von Bortkiewicz [22] reports that the distribution of the annual number of the number of child suicides follows a similar pattern.

Chung [5, p. 196] also lists the following as examples of Poisson distributions.

- The number of colorblind people in a large group.
- The number of raisins in cookies. (Who did this research?)
- The number of misprints on a page. (Again who did the counting?[4])

It turns out that just as class ended in 2015, my colleague Phil Hoffman, finished correcting the page proofs for his new book, *Why Did Europe Conquer the World?* [11]. In $m = 261$ pages

[4] According to my late coauthor Roko Aliprantis, Apostol's Law states there are an infinite number of misprints in any book. The proof is that every time you open a book that you wrote, you find another misprint.

there were a total $n = 43$ mistakes. There were no mistakes on 222 pages, 1 mistake on 35 pages, and 2 mistakes on 4 pages. This is an average rate of $\mu = n/m = 43/261 = 0.165$ mistakes per page. Here is a table of actual page counts vs. rounded expected page counts $mp_{0.165}(k)$ with k errors, based on the Poisson(0.165) distribution:

	0	1	2	$\geqslant 3$
Actual	222	35	4	0
Model	221.4	36.5	3.0	0.17

As you can see this looks like a good fit. The p-value is 0.90.

Feller [10, § VI.7, pp. 159–164] lists these additional phenomena, and supplies citations to back up his claims.

- Ernest Rutherford (1871–1937) was awarded the 1908 Nobel Prize in Chemistry "for his investigations into the disintegration of the elements, and the chemistry of radioactive substances" (Nobel Prize website). Rutherford, Chadwick,[5] and Ellis [16, pp. 171–172] report the results of an experiment by Rutherford and Geiger[6] [17] in 1910 where they recorded the time of scintillations caused by α-particles emitted from a film of polonium. "[T]he time of appearance of each scintillation was recorded on a moving tape by pressing an electric key. ... The number of α particles counted was 10,097 and the average number appearing in the interval under consideration, namely 1/8 minute, was 3.87." The number of 7.5-second intervals (bins) was $m = 2608$. (That's a little over 5 hours total, but it was spread out over five days in three to five minute intervals [17, p. 699]. I assume it was a poor grad student who did the key-pressing.) The number of emissions (balls) was, as stated above, $n = 10,097$. Rutherford and Geiger enlisted the aid of Bateman[7] [17] who "worked out ... [t]he distribution of α particles according to the law of probability."

These data are widely referred to in the probability and statistics literature. Feller [10, p. 160] cites their book, and also refers to Harald Cramér [7, pp. 435–437], for some statistical analysis. Cramér in turn takes as his source a textbook by Aitken [1, pp. 77–78].

Table 13.2 has my reproduction of Rutherford et. al.'s table, where, like Feller, I have combined the counts for $k \geqslant 10$.[8] I have also recalculated the model predictions for $\mu = 10097/2608 = 3.87$, which differ from Cramér's numbers by no more than 0.1. (Rutherford, et. al. rounded to integers.) I calculate the p-value measure of fit to be 0.23, but Cramér reported 0.17.

- The number of "chromosome interchanges" in cells subjected to X-ray radiation. [4]
- Telephone connections to a wrong number. (Frances Thorndike [20])
- Bacterial and blood counts.
- Too late for mention by Feller, Keller [12] found the Poisson distribution was a good model for the number of goals scored in a soccer game and the number of runs in a baseball game.

> The Poisson distribution describes the number of cases with k occurrences of a rare phenomenon in a large sample of independent cases.
>
> It can be used to estimate the probability of the occurence of a cluster of rare events.

[5] James Chadwick (1891–1974) won the 1935 Nobel Prize in Physics "for the discovery of the neutron" (Nobel Prize website), a particle that had been proposed by Rutherford.

[6] Hans Geiger was the inventor of the eponymous counter (wp). The famous Rutherford "gold foil" experiment that demonstrated the existence of atomic nuclei was actually conducted by Geiger and Marsden under Rutherford's direction (wp).

[7] Harry Bateman (1882–1946) (wp) was an English mathematician who later moved to Caltech. There is still a Harry Bateman Research Instructorship in the math department.

[8] The full set of counts were:

k	10	11	12	13	14
count	10	4	0	1	1

k	Actual	Model
0	57	54.3
1	203	210.3
2	383	407.1
3	525	525.3
4	532	508.4
5	408	393.7
6	273	254.0
7	139	140.5
8	45	68.0
9	27	29.2
10+	16	17.1

Table 13.2. Alpha-particle emissions.

13.8 Explanation of the Law of Small Numbers

Recall the experiment described in Subsection 13.6. There were

- n balls, dropped at random into
- m bins, so
- $\mu = n/m$ average number of balls per bin.
- N_b is the random variable that counts the hits on bin b.
- X_k is the random variable that counts the number of bins with k hits,

$$X_k = |\{b : N_b = k\}|, \qquad k = 0, \dots, m.$$

I asserted the following:

13.8.1 The Law of Small Numbers *Fixing μ and k, if n is large enough, with high probability,*

$$X_k = \text{the number of bins with } k \text{ hits} \approx m p_\mu(k). \tag{3}$$

Here is a mildly bogus argument to convince you that it is plausible. It is a variation on the arguments of Feller [10, Section VI.6, pp. 156–159] and Pitman [15, § 3.5, pp. 228–236], where they describe the **Poisson scatter**. The LSN though is not quite that phenomenon.

Quasi-proof: Pick any bin, say bin b and pick some number k of hits. The probability that ball i hits bin b is $1/m = \mu/n$. So the number of hits on bin b, has a Binomial$(n, \mu/n)$ distribution, which for fixed μ and large n is approximated by the Poisson(μ) distribution, so

$$P(N_b = k) = \operatorname{Prob}(\text{bin } b \text{ has } k \text{ balls}) \approx p_\mu(k).$$

But this is not the Law of Small Numbers. This just says that any individual bin has an approximate Poisson probability of k hits, but the LSN says that the for each k, the *fraction* of bins with k hits follows a Poisson distribution, $X_k/m \approx p_\mu(k)$. You might think that this is just an application of the Law of Large Numbers: If getting k hits is counted as a success, and if the number of bins m is large, the number of successes should be close to its expectation $m p_\mu(k)$. This argument, while giving the right answer, is invalid, because to apply the LLN, the trials have to be independent. That is, the event that bin b has k hits has to be independent of the event that bin c, has k hits, but it isn't. For a stark example, let $n = m = 2$, so if bin 1 has two hits, then bin 2 must have zero. So how do we justify the LSN?

Imagine independently replicating this entire experiment r times. Say a replication is a "success" if bin b has k balls. The probability of a successful experiment is thus approximately $p_\mu(k)$. By the Law of Large Numbers, the number of successful experiments in a large number r of independent replications is close to $rp_\mu(k)$ with high probability.

That is, for large n, for large enough r, with high probability, the number of replications in which bin b has k hits is approximately $rp_\mu(k)$.

Now symmetry says that there is nothing special about bin b, and since there are m bins, summing over all bins and all replications one would expect that the number of bins with k hits would be approximately m times the number of replications in which bin b has k hits (which is approximately $rp_\mu(k)$). Thus all together, in the r replications there are approximately $mrp_\mu(k)$ bins with k hits. Since all the replications are of the same experiment, there should be approximately

$$\frac{mrp_\mu(k)}{r} = mp_\mu(k)$$

bins with k hits per replication. ∎

In this argument, I did a little(?) handwaving (using the terms *close* and *approximately*). To make it rigorous would require a careful analysis of the size of the deviations of the results from their expected values. Note though that r has to be chosen after k, so we don't expect (2) to hold for all values of k, just the smallish ones. In particular, if $\mu = n/m$, then I cannot fix μ and m and let n get large. I may need to go back to the drawing board.

So how big does n have to be? Breiman [3, p. 35] suggests this approximation works well when $\mu^2/n = n/m^2 \ll 1$.

Bibliography

[1] A. C. Aitken. 1944. *Statistical mathematics*, 3d. ed. University Mathematical Texts. Edinburgh and London: Oliver and Boyd.

[2] T. M. Apostol. 1967. *Calculus, Volume I: One-variable calculus with an introduction to linear algebra*, 2d. ed. New York: John Wiley & Sons.

[3] L. Breiman. 1986. *Probability and stochastic processes: With a view toward applications*, 2d. ed. Palo Alto, California: Scientific Press.

[4] D. G. Catchside, D. E. Lea, and J. M. Thoday. 1945–46. Types of chromosomal structural change induced by the irradiation of *Tradescantia* microspores. *Journal of Genetics* 47:113–136.

[5] K. L. Chung. 1979. *Elementary probability theory with stochastic processes*. Undergraduate Texts in Mathematics. New York, Heidelberg, and Berlin: Springer–Verlag.

[6] R. D. Clarke. 1946. An application of the Poisson distribution. *Journal of the Institute of Actuaries* 72:481.

[7] H. Cramér. 1946. *Mathematical methods of statistics*. Number 34 in Princeton Mathematical Series. Princeton, New Jersey: Princeton University Press. Reprinted 1974.

[8] A. R. Crathorne. 1928. The law of small numbers. *American Mathematical Monthly* 35(4):169–175. http://www.jstor.org/stable/2300002

[9] W. Feller. 1950. *An introduction to probability theory and its applications*, 1st. ed., volume 1. New York: Wiley.

[10] ———. 1968. *An introduction to probability theory and its applications*, 3d. ed., volume 1. New York: Wiley.

[11] P. T. Hoffman. 2015. *Why did Europe conquer the world?* Princeton, New Jersey: Princeton University Press. http://press.princeton.edu/titles/10452.html

[12] J. B. Keller. 1994. A characterization of the Poisson distribution and the probability of winning a game. *AmStat* 48(4):294–298. DOI: 10.1080/00031305.1994.10476084

[13] J. M. Keynes. 1921. *A treatise on probability*. London: Macmillan and Co.

[14] R. J. Larsen and M. L. Marx. 2012. *An introduction to mathematical statistics and its applications*, fifth ed. Boston: Prentice Hall.

[15] J. Pitman. 1993. *Probability*. Springer Texts in Statistics. New York, Berlin, and Heidelberg: Springer.

[16] E. Rutherford, J. Chadwick, and C. D. Ellis. 1930. *Radiations from radioactive substances*. Cambridge: Cambridge University Press.

[17] E. Rutherford, H. Geiger, and H. Bateman. 1910. The probability variations in the distribution of α particles. *Philosophical Magazine Series 6* 20(118):698–707. DOI: 10.1080/14786441008636955

[18] L. P. Shaw and L. F. Shaw. 2019. The flying bomb and the actuary. *Significance* 16(5):12–17.

[19] J. A. Simpson and E. S. C. Weiner, eds. 1989. *The Oxford English Dictionary*, 2d. ed. Oxford: Oxford University Press.

[20] F. Thorndike. 1926. Applications of Poisson's probability summation. *Bell System Technical Journal* 5(4):604–624. http://www3.alcatel-lucent.com/bstj/vol05-1926/articles/bstj5-4-604.pdf

[21] A. Tversky and D. Kahneman. 1971. Belief in the law of small numbers. *Psychological Bulletin* 76(2):105–110.

[22] L. von Bortkiewicz. 1898. *Das Gesetz der kleinen Zahlen [The law of small numbers]*. Leipzig: B.G. Teubner. The imprint lists the author as Dr. L. von Bortkewitsch.