Ma 3/103 KC Border
Introduction to Probability and Statistics Winter 2017

Lecture 12: The Law of Small Numbers

Relevant textbook passages:
Pitman [11]: Sections 2.4,3.8, 4.2
Larsen–Marx [10]: Sections 3.8, 4.2, 4.6

12.1 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

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

See Theorem 10.5.1.

Poisson discovered another peculiar, but useful, limit of the Binomial distribution.[1] Fix $\mu > 0$ and let X_n have the Binomial distribution 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. A n gets large, for each k we have

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

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.

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 obtaned by taking the infinite Taylor series expansion of the exponential function around 0. See, e.g., Apostol [1, p. 436].

[1] von Bortkiewicz [15] gives this cryptic reference: Zu vergleichen Poisson, Recherches sur la probablité des jugements, Paris 1837. n^o 81, p. 205–207

12.2 The Poisson(μ) distribution

Larsen–Marx [10]: Section 4.3
Pitman [11]: 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

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

Table 12.1 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 12.1. The Poisson probabilities $p_\mu(k)$.

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

$$\frac{p_\mu(k+1)}{p_\mu(k)} = \begin{cases} \mu & k = 0 \\ \mu/k & k \geqslant 1. \end{cases}$$

So as long as $k \leqslant \mu$, then $p_\mu(k+1) > p_\mu(k)$, but then $p_\mu(k)$ decreases with k. See Figure 12.1.

The next set of charts show the Poisson distribution stacks up agains the Binomial.

12.3 The mean and variance of the Poisson

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

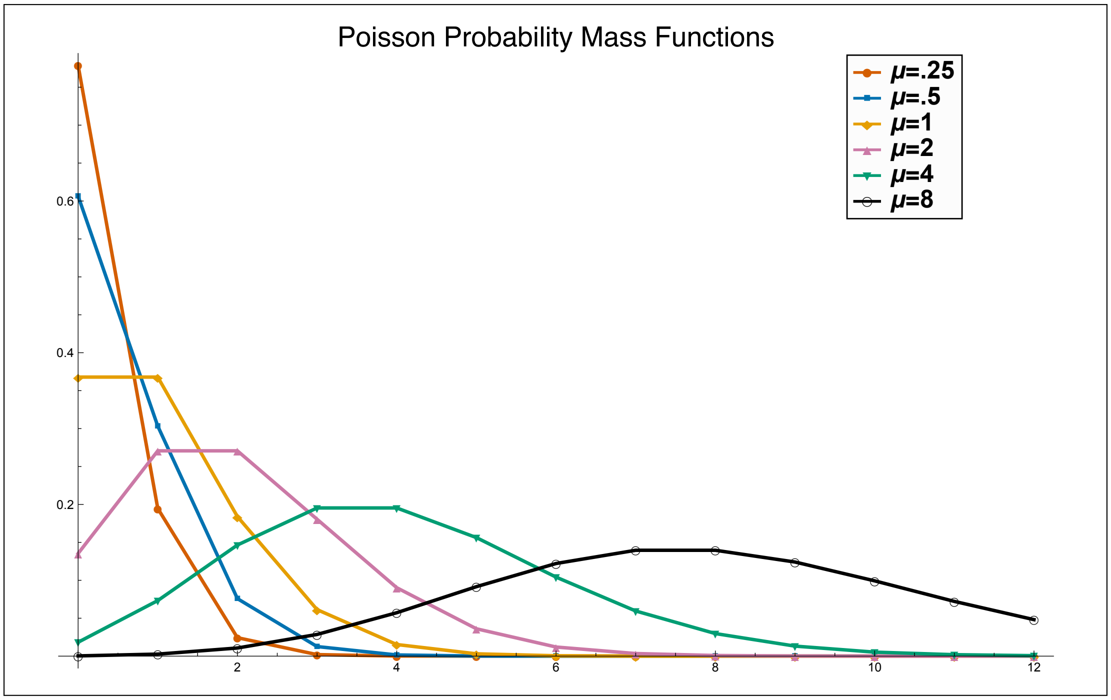

Figure 12.1.

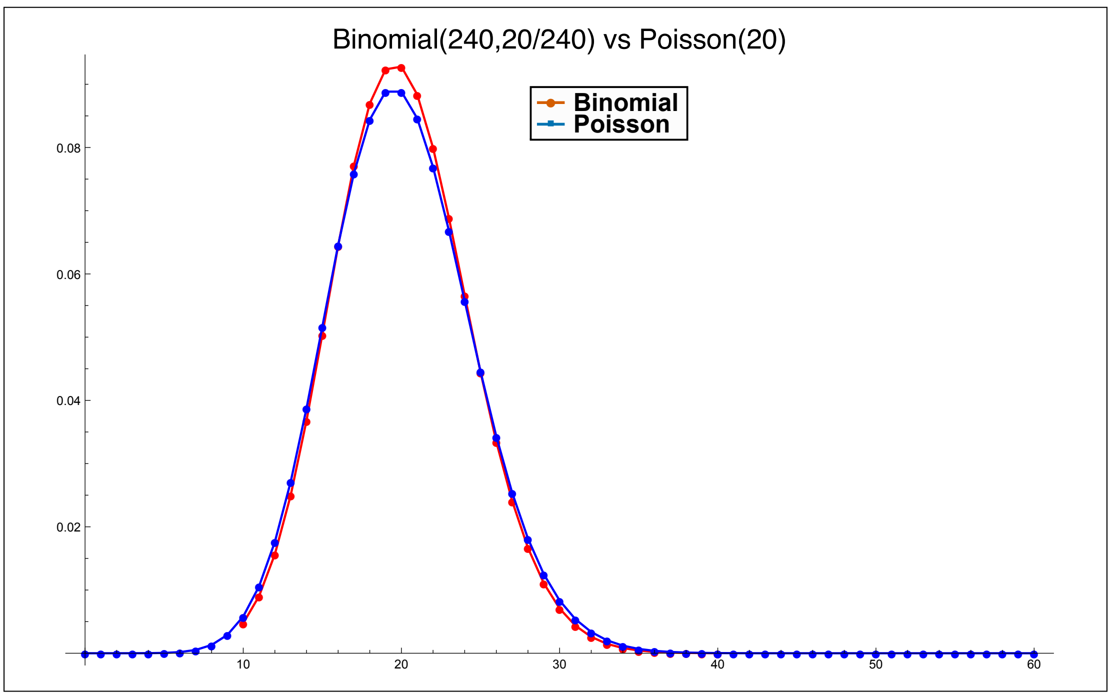

Figure 12.2.

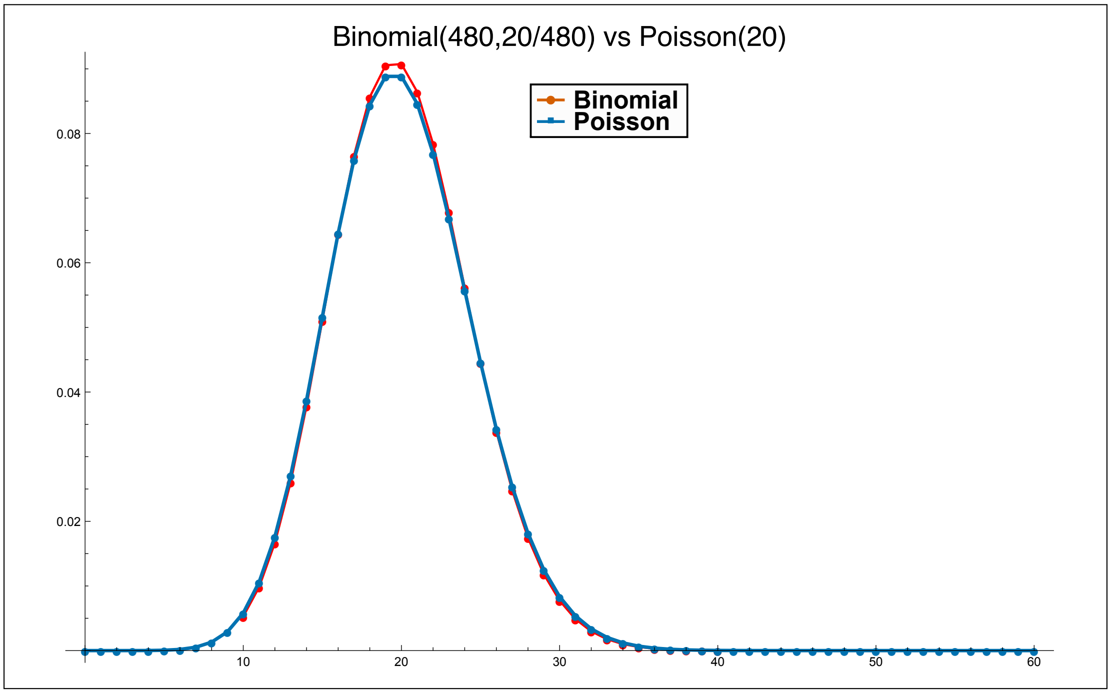

Figure 12.3.

Pitman [11, pp. 223–24] uses and write $X^2 = X(X-1) + X$. Start with $\boldsymbol{E}\big(X(X-1)\big)$:

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

So

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

12.4 The Law of Small Numbers

In 1898, Ladislaus von Bortkiewicz [15] published *Das Gesetz der kleinen Zahlen* [The Law of Small Numbers]. He described a number of observations on the frequency of occurrence of **rare events** that appear to follow a Poisson distribution. Here is a mathematical model to explain his observations.

The random experiment

The experiment is to "scatter" m numbered balls at random among n numbered urns. The average number of balls per urn is then

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

For each ball i let U_i be the number of the urn that ball i "hits." Assume that each ball i is equally likely to hit each urn b so that

$$\operatorname{Prob}(U_i = b) = \frac{1}{n}.$$

Moreover let's assume that the random variables U_i are independent. The numerical value of U_i is just a label.

For each urn b, let H_b the random variable that counts the hits on b,

$$H_b = |\{i : U_i = b\}|.$$

And let X_k be the random variable that counts the number of urns with k hits,

$$X_k = |\{b : H_b = k\}|.$$

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

12.4.1 Proposition (The Law of Small Numbers) *Fixing μ and k, if n is large enough, with high probability,*

$$X_k = \text{the number of urns with } k \text{ hits} \approx np_\mu(k). \tag{1}$$

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

12.5 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 [6] in 1950. (I have the third edition, so I will cite it.)

- During the Second World War, Nazi Germany use unmanned aircraft, the V1 Buzz Bombs, to attack London. (They weren't quite drones, since they were never designed to return or to be remote controlled. Once launched, where they came down was reasonably random.) Feller [7, pp. 160–161] cites R. D. Clarke [4] (an insurance adjuster for The Prudential), who reports that 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.

How does this fit our story? Consider each of the $m = 537$ Buzz Bombs a ball and each of the $n = 576$ sectors an urn. Then $\mu = 537/576 = 0.9323$. 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. The Law of Small Numbers then predicts that the number of districts with k hits should be approximately

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

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 smallish k. (A rule of thumb is that the model should predict a value of at least 5 sectors for it it to be a good approximation.)

- The urns 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 [3, p. 196] cites John Maynard Keynes [9, p. 402], who reports that von Bortkiewicz [15] 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 [15, p. 24]. It covers the years 1875–1894 and fourteen different Prussian Cavalry Corps. So there are $280 = 14 \times 20$CorpsYears. Each CorpsYear corresponds to an urn. There were 196 deaths, each corresponding to a ball. So $\mu = 196/280 = 0.70$, so with $n = 280$ our theoretical prediction of the number of CorpsYears with k deaths is the Poisson average $np_{.7}(k)$. Unfortunately, the numbers of expected deaths as reported by von Bortkiewicz, do not agree with with my calculations. I will look into this further. Keynes [9, 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 [9, p. 402] also reports that von Bortkiewicz [15] reports that the distribution of the annual number of the number of child suicides follows a similar pattern.

Chung [3, p. 196] also lists the following as examples of Poisson distributions.

- The number of color blind 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?[2])

[2] 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.

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?* [8]. In $n = 261$ pages there were a total $m = 43$ 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 $\mu = 43/261 = 0.165$ mistakes per page. Here is a table of actual page counts vs. rounded expected page counts $np_{0.165}(k)$ 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-values is 0.90.

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

- Rutherford, Chadwick, and Ellis [12, pp. 171–172] report the results of an experiment by Rutherford and Geiger [13] 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 was $N = 2608$. (That's a little over 5 hours total. I assume it was a poor grad student who did the key-pressing.)

These data are widely referred to in the probability and statistics literature. Feller [7, p. 160] cites their book, and also refers to Harald Cramér [5, p. 436], for some statistical analysis. Cramér in turn takes as his source a textbook by Aitken.

Table 12.2 has my reproduction of Rutherford et. al.'s table, where, like Feller, I have combined the counts for $k \geqslant 10$.[3] 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.

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 12.2. Alpha-particle emissions.

- The number of "chromosome interchanges" in cells subjected to X-ray radiation. [2]
- Telephone connections to a wrong number. (Frances Thorndike [14])
- Bacterial and blood counts.

[3] The full set of counts were:

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

The Poisson distribution describes the number of cases with k occurrences of a rare phenomenon in a large sample of independent cases.

12.6 Explanation of the Law of Small Numbers

The following story is a variation on one told by Feller [7, Section VI.6, pp. 156–159] and Pitman [11, § 3.5, pp. 228–236], where they describe the **Poisson scatter**. This is not quite that phenomenon.

Here is a mildly bogus argument to convince you that it is plausible:

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

$$P(H_b = k) = \text{Prob}\,(\text{urn } b \text{ has } k \text{ hits}) \approx p_\mu(k).$$

But this is not the Law of Small Numbers. This just says that any individual urn has a Poisson probability of k hits, but the LSN says that the for each k, the *fraction* of urns with k hits follows a Poisson distribution, $X_k/n \approx p_\mu(k)$. How do we get this?

Because the total number of balls, m, is fixed, the number of balls in urn b and urn c are not independent. (In fact, the joint distribution is one giant multinomial.) So I can't simply multiply this by n urns to get the number of urns with k hits. But random vector of the number of hits on each urn b is **exchangeable**. That is,

$$P(H_1 = k_1, H_2 = k_2, \dots, H_n = k_n) = P\left(H_{\pi(1)} = k_1, H_{\pi(2)} = k_2, \dots, H_{\pi(n)} = k_n\right)$$

Elaborate on exchangeable rvs.

for any permutation π of $\{1, 2, \dots, n\}$. We can exploit this instead of using independence.

So imagine independently replicating this experiment r times. Say the experiment itself is a "success" if urn b has k hits. The probability of a successful experiment is thus $p_\mu(k)$. By the Law of Large Numbers, the number of successes in a large number r of experiments is close to $rp_\mu(k)$.

Now exchangeability says that there is nothing special about urn b, and there are n urns, so summing over all urns and all replications one would expect that the number of urns with k hits would be n times the number of experiments in which urn b has k hits (namely, $rp_\mu(k)$). Thus all together, in the r replications there about $nrp_\mu(k)$ urns with k hits. Since all the replications are the same experiment, there should be about

$$\frac{nrp_\mu(k)}{r} = np_\mu(k)$$

urns with k hits per experiment. ∎

In this argument, I did a little handwaving (using the terms *close* and *about*). 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 (1) to hold for all values of k, just the smallish ones.

12.7 Sums of independent Poissons

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

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:

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

Bibliography

[1] T. M. Apostol. 1967. *Calculus*, 2d. ed., volume 1. Waltham, Massachusetts: Blaisdell.

[2] 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.

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

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

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

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

[7] ——— . 1968. *An introduction to probability theory and its applications*, 3d. ed., volume 1. New York: Wiley.

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

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

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

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

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

[13] 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

[14] 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

[15] 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.