Chapter 9

Poisson approximations

9.1 Overview

The Bin(n, p) can be thought of as the distribution of a sum of independent indicator random variables $X_1 + \dots + X_n$, with $\{X_i = 1\}$ denoting a head on the ith toss of a coin that lands heads with probability p. Each X_i has a Ber(p) distribution. The normal approximation to the Binomial works best when the variance $np(1-p)$ is large, for then each of the standardized summands $(X_i - p)/\sqrt{np(1-p)}$ makes a relatively small contribution to the standardized sum.

When n is large but p is small, in such a way that $\lambda := np$ is not too large, a different type of approximation to the Binomial is better. The traditional explanation uses an approximation to

$$\mathbb{P}\{X = k\} = \binom{n}{k} p^k (1-p)^{n-k}$$

for a fixed k. For a k that is small compared with n, consider the contributions $\binom{n}{k} p^k$ and $(1-p)^{n-k}$ separately.

$$\begin{aligned}\binom{n}{k} p^k &= \frac{n(n-1)\dots(n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \\ &= 1 \times \left(1 - \frac{1}{n}\right) \times \dots \left(1 - \frac{k-1}{n}\right) \frac{\lambda^k}{k!} \approx \frac{\lambda^k}{k!}\end{aligned}$$

and

$$\log(1-p)^{n-k} = (n-k)\log(1 - \lambda/n) \approx n(-\lambda/n).$$

That is, $(1-p)^{n-k} \approx e^{-\lambda}$. Together the two approximations give

$$\binom{n}{k} p^k (1-p)^{n-k} \approx \frac{\lambda^k}{k!} e^{-\lambda}.$$

For large k, both $\mathbb{P}\{X = k\}$ and $p_k' := e^{-\lambda}\lambda^k/k!$ are small. The p_k' define a new distribution.

Definition. A random variable Y is said to have a ***Poisson distribution*** with parameter λ if it can take values in $\mathbb{N}_0$, the set of nonnegative integers, with probabilities

$$\mathbb{P}\{Y = k\} = \frac{e^{-\lambda}\lambda^k}{k!} \qquad \text{for } k = 0, 1, 2, \ldots$$

The parameter λ must be positive. The distribution is denoted by Poisson(λ).

That is , for $\lambda = np$ not too large, the Bin(n, p) is (well?) approximated by the Poisson(λ).

Remark. Counts of rare events—such as the number of atoms undergoing radioactive decay during a short period of time, or the number of aphids on a leaf—are often modeled by Poisson distributions, at least as a first approximation.

The Poisson inherits several properties from the Binomial. For example, the Bin(n, p) has expected value np and variance $np(1-p)$. One might suspect that the Poisson(λ) should therefore have expected value $\lambda = n(\lambda/n)$ and variance $\lambda = \lim_{n\to\infty} n(\lambda/n)(1-\lambda/n)$. Also, the coin-tossing origins of the Binomial show that if X has a Bin(m, p) distribution and Y has a Bin(n, p) distribution independent of X, then $X + Y$ has a Bin($n + m, p$) distribution. Putting $\lambda = mp$ and $\mu = np$ one might then suspect that the sum of independent Poisson(λ) and Poisson(μ) distributed random variables is Poisson($\lambda + \mu$) distributed. These suspicions are correct.

Example <9.1> If X has a Poisson(λ) distribution, then $\mathbb{E}X = \text{var}(X) = \lambda$. If also Y has a Poisson(μ) distribution, and Y is independent of X, then $X + Y$ has a Poisson($\lambda + \mu$) distribution.

There is a clever way to simplify some of the calculations in the last Example using ***generating functions***, a way to code all the Poisson probabilities into a single function on $[0, 1]$.

Example <9.2> Calculate moments of the Poisson(λ) distribution using its generating function.

9.2 A more precise Poisson approximation

Modern probability methods have improved this rough approximation of the Binomial by the Poisson by giving useful upper bounds for the error of approximation. Using a technique known as the ***Chen-Stein method*** one can show that

$$d_{TV}(\text{Bin}(n,p), \text{Poisson}(np)) := \frac{1}{2}\sum\nolimits_{k\geq 0}\left|\mathbb{P}\{S=k\} - e^{-\lambda}\frac{\lambda^k}{k!}\right| \leq \min\left(p, np^2\right),$$

which makes precise the traditional advice that the Poisson approximation is good "when p is small and np is not too big". (In fact, the tradition was a bit conservative.)

Remark. The quantity $d_{TV}(P,Q)$ is called the ***total variation distance*** between two probabilities P and Q. It is also equal to $\max_A |PA - QA|$ where the maximum runs over all subsets A of the set where both P and Q are defined. For $P = \text{Bin}(n,p)$ and $Q = \text{Poisson}(np)$, the A runs over all subsets of the nonnegative integers.

The Chen-Stein method of approximation also works in situations where the rare events do not all have the same probability of occurrence. For example, suppose $S = X_1 + X_2 + \cdots + X_n$, a sum of independent random variables where X_i has a Ber(p_i) distribution, for constants $p_1, p_2, \ldots, p_n$ that are not necessarily all the same. The sum S has expected value $\lambda = p_1 + \cdots + p_n$. Using Chen-Stein it can also be shown that that

$$\frac{1}{2}\sum\nolimits_{k\geq 0}\left|\mathbb{P}\{S=k\} - e^{-\lambda}\frac{\lambda^k}{k!}\right| \leq \min\left(1, \frac{1}{\lambda}\right)\sum\nolimits_{i=1}^{n} p_i^2.$$

The Chen-Stein method of proof is elementary—in the sense that it makes use of probabilistic techniques only at the level of Statistics 241—but extremely subtle. See Barbour et al. (1992) for an extensive discussion of the method.

9.3 Poisson approximations under dependence

The Poisson approximation also applies in many settings where the trials are "almost independent", but not quite. Again the Chen-Stein method delivers impressively good bounds on the errors of approximation. For example, the method works well in two cases where the dependence takes an a simple form.

Once again suppose $S = X_1 + X_2 + \cdots + X_n$, where X_i has a Ber(p_i) distribution, for constants $p_1, p_2, \ldots, p_n$ that are not necessarily all the same. Often X_i is interpreted as the indicator function for success in the ith in some finite set of trials. Define $S_{-i} = S - X_i = \sum_{1 \le j \le n} \mathbb{I}\{j \ne i\} X_j$. The random variables $X_1, \ldots, X_n$ are said to be ***positively associated*** if

$$\mathbb{P}\{S_{-i} \ge k \mid X_i = 1\} \ge \mathbb{P}\{S_{-i} \ge k \mid X_i = 0\} \qquad \text{for each } i \text{ and } k = 0, 1, 2, \ldots$$

and ***negatively associated*** if

$$\mathbb{P}\{S_{-i} \ge k \mid X_i = 1\} \le \mathbb{P}\{S_{-i} \ge k \mid X_i = 0\} \qquad \text{for each } i \text{ and } k = 0, 1, 2, \ldots.$$

Intuitively, positive association means that success in the ith trial makes success in the other trials more likely; negative association means that success in the ith trial makes success in the other trials less likely.

With some work it can be shown (Barbour et al., 1992, page 20) that

$$\frac{1}{2} \sum_{k \ge 0} \left| \mathbb{P}\{S = k\} - e^{-\lambda} \frac{\lambda^k}{k!} \right| \le \min\left(1, \frac{1}{\lambda}\right) \times \begin{cases} \left(\text{var}(S) - \lambda + 2\sum_{i=1}^n p_i^2\right) & \text{under positive association} \\ (\lambda - \text{var}(S)) & \text{under negative association} \end{cases}.$$

These bounds take advantage of the fact that var(S) would be exactly equal to λ if S had a Poisson(λ) distribution.

The next Example illustrates both the classical approach and the Chen-Stein approach (via positive association) to deriving a Poisson approximation for a matching problem.

Example <9.3> Poisson approximation for a matching problem: assignment of n letters at random to n envelopes, one per envelope.

9.4 Examples for Chapter 9

<9.1> **Example.** *If X has a* Poisson(λ) *distribution, then* $\mathbb{E}X = \text{var}(X) = \lambda$. *If also Y has a* Poisson(μ) *distribution, and Y is independent of X, then $X+Y$ has a* Poisson($\lambda+\mu$) *distribution.*

Assertion (i) comes from a routine application of the formula for the expectation of a random variable with a discrete distribution.

$$\begin{aligned}\mathbb{E}X = \sum\nolimits_{k=0}^{\infty} k\mathbb{P}\{X=k\} &= \sum\nolimits_{k=1}^{\infty} k\frac{e^{-\lambda}\lambda^k}{k!} \qquad \text{What happens to } k=0?\\ &= e^{-\lambda}\lambda \sum_{k-1=0}^{\infty} \frac{\lambda^{k-1}}{(k-1)!}\\ &= e^{-\lambda}\lambda e^{\lambda}\\ &= \lambda.\end{aligned}$$

Notice how the k cancelled out one factor from the $k!$ in the denominator.

If I were to calculate $\mathbb{E}(X^2)$ in the same way, one factor in the k^2 would cancel the leading k from the $k!$, but would leave an unpleasant $k/(k-1)!$ in the sum. Too bad the k^2 cannot be replaced by $k(k-1)$. Well, why not?

$$\begin{aligned}\mathbb{E}(X^2-X) &= \sum_{k=0}^{\infty} k(k-1)\mathbb{P}\{X=k\}\\ &= e^{-\lambda}\sum_{k=2}^{\infty} k(k-1)\frac{\lambda^k}{k!} \qquad \text{What happens to } k=0 \text{ and } k=1?\\ &= e^{-\lambda}\lambda^2 \sum_{k-2=0}^{\infty}\frac{\lambda^{k-2}}{(k-2)!} = \lambda^2.\end{aligned}$$

Now calculate the variance.

$$\text{var}(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2 = \mathbb{E}(X^2-X) + \mathbb{E}X - (\mathbb{E}X)^2 = \lambda.$$

For assertion (iii), first note that $X+Y$ can take only values $0,1,2\dots$. For a fixed k in this range, decompose the event $\{X+Y=k\}$ into disjoint pieces whose probabilities can be simplified by means of the independence between X and Y.

$$
\begin{aligned}
&\mathbb{P}\{X+Y=k\} = \\
&\quad \mathbb{P}\{X=0, Y=k\} + \mathbb{P}\{X=1, Y=k-1\} + \dots + \mathbb{P}\{X=k, Y=0\} \\
&\quad = \mathbb{P}\{X=0\}\mathbb{P}\{Y=k\} + \dots + \mathbb{P}\{X=k\}\mathbb{P}\{Y=0\} \\
&\quad = \frac{e^{-\lambda}\lambda^0}{0!}\frac{e^{-\mu}\mu^k}{k!} + \dots + \frac{e^{-\lambda}\lambda^k}{k!}\frac{e^{-\mu}\mu^0}{0!} \\
&\quad = \frac{e^{-\lambda-\mu}}{k!}\left(\frac{k!}{0!k!}\lambda^0\mu^k + \frac{k!}{1!(k-1)!}\lambda^1\mu^{k-1} + \dots + \frac{k!}{k!0!}\lambda^k\mu^0\right) \\
&\quad = \frac{e^{-\lambda-\mu}}{k!}(\lambda+\mu)^k.
\end{aligned}
$$

The bracketed sum in the second last line is just the binomial expansion of $(\lambda+\mu)^k$. □

> **Remark.** How do you interpret the notation in the last calculation when $k=0$? I always feel slightly awkward about a contribution from $k-1$ if $k=0$.

<9.2> **Example.** There is a sneakier way to calculate $\mathbb{E}X^m$ for $m = 1, 2, \dots$ when X has a Poisson(λ) distribution. Code the whole distribution into a function (the ***probability generating function***) of a dummy variable s:

$$
g(s) := \mathbb{E}s^X = \sum\nolimits_{k\ge 0} s^k e^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}\sum\nolimits_{k\ge 0}\frac{(s\lambda)^k}{k!} = e^{-\lambda}e^{\lambda s}.
$$

Given g, the individual probabilities $\mathbb{P}\{X=k\}$ could be recovered by expanding the function as a power series in s.

Other facts about the distribution can also be obtained from g. For example,

$$
\frac{d}{ds}g(s) = \lim_{h\to 0}\mathbb{E}\left(\frac{(s+h)^X - s^X}{h}\right) = \mathbb{E}\frac{\partial}{\partial s}s^X = \mathbb{E}Xs^{X-1}
$$

and, by direct calculation, $g'(s) = e^{-\lambda}\lambda e^{\lambda s}$. Put $s = 1$ in both expressions to deduce that $\mathbb{E}X = g'(1) = \lambda$.

Similarly, repeated differentiation inside the expectation sign gives

$$
g^{(m)}(s) = \frac{\partial^m}{\partial s^m}\mathbb{E}(s^X) = \mathbb{E}\left(X(X-1)\dots(X-m+1)s^{X-m}\right),
$$

and direct differentiation of g gives $g^{(m)}(s) = e^{-\lambda}\lambda^m e^{\lambda s}$. Again put $s = 1$ to deduce that

$$
\lambda^m = g^{(m)}(1) = \mathbb{E}\left(X(X-1)\dots(X-m+1)\right) \qquad \text{for } m = 1, 2, \dots
$$

□

<9.3> **Example.** Suppose n letters are placed at random into n envelopes, one letter per envelope. The total number of correct matches, S, can be written as a sum $X_1 + \cdots + X_n$ of indicators,

$$X_i = \begin{cases} 1 & \text{if letter } i \text{ is placed in envelope } i, \\ 0 & \text{otherwise.} \end{cases}$$

The X_i are dependent on each other. For example, symmetry implies that

$$p_i = \mathbb{P}\{X_i = 1\} = 1/n \qquad \text{for each } i$$

and

$$\mathbb{P}\{X_i = 1 \mid X_1 = X_2 = \cdots = X_{i-1} = 1\} = \frac{1}{n-i+1}$$

Remark. If we eliminated the dependence by relaxing the requirement of only one letter per envelope, the number of letters placed in the correct envelope (possibly together with other, incorrect letters) would then have a Bin$(n, 1/n)$ distribution, which is approximated by Poisson(1) if n is large.

We can get some supporting evidence for S having something close to a Poisson(1) distribution under the original assumption (one letter per envelope) by calculating some moments.

$$\mathbb{E}S = \sum\nolimits_{i \le n} \mathbb{E}X_i = n\mathbb{P}\{X_i = 1\} = 1$$

and

$$\begin{aligned} \mathbb{E}S^2 &= \mathbb{E}\left(X_1^2 + \cdots + X_n^2 + 2\sum_{i<j} X_i X_j \right) \\ &= n\mathbb{E}X_1^2 + 2\binom{n}{2}\mathbb{E}X_1X_2 \qquad \text{by symmetry} \\ &= n\mathbb{P}\{X_1 = 1\} + (n^2 - n)\mathbb{P}\{X_1 = 1,\, X_2 = 1\} \\ &= \left(n \times \frac{1}{n}\right) + (n^2 - n) \times \frac{1}{n(n-1)} \\ &= 2. \end{aligned}$$

Thus $\text{var}(S) = \mathbb{E}S^2 - (\mathbb{E}S)^2 = 1$. Compare with Example <9.1>, which gives $\mathbb{E}Y = 1$ and $\text{var}(Y) = 1$ for a Y distributed Poisson(1).

Using the ***method of inclusion and exclusion***, it is possible (Feller, 1968, Chapter 4) to calculate the exact distribution of the number of correct matches,

$$\mathbb{P}\{S = k\} = \frac{1}{k!}\left(1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} - \cdots \pm \frac{1}{(n-k)!}\right) \qquad \text{for } k = 0, 1, \ldots, n.$$

For fixed k, as $n \to \infty$ the probability converges to

$$\frac{1}{k!}\left(1 - 1 + \frac{1}{2!} - \frac{1}{3!} - \ldots\right) = \frac{e^{-1}}{k!},$$

which is the probability that $Y = k$ if Y has a Poisson(1) distribution.

The Chen-Stein method is also effective in this problem. I claim that it is intuitively clear (although a rigorous proof might be tricky) that the X_i's are positively associated:

$$\mathbb{P}\{S_{-i} \geq k \mid X_i = 1\} \geq \mathbb{P}\{S_{-i} \geq k \mid X_i = 0\} \qquad \text{for each } i \text{ and each } k \in \mathbb{N}_0.$$

I feel that if $X_i = 1$, then it is more likely for the other letters to find their matching envelopes than if $X_i = 0$, which makes things harder by filling one of the envelopes with the incorrect letter i. Positive association gives

$$\frac{1}{2}\sum\nolimits_{k\geq 0}\left|\mathbb{P}\{S = k\} - e^{-\lambda}\frac{\lambda^k}{k!}\right| \leq 2\sum\nolimits_{i=1}^{n} p_i^2 + \text{var}(S) - 1 = 2/n.$$

As n gets large, the distribution of S does get close to the Poisson(1) in the strong, total variation sense. □

References

Barbour, A. D., L. Holst, and S. Janson (1992). *Poisson Approximation.* Oxford University Press.

Feller, W. (1968). *An Introduction to Probability Theory and Its Applications* (third ed.), Volume 1. New York: Wiley.