Chapter 8

Poisson approximations

The Bin(n, p) can be thought of as the distribution of a sum of independent indicator random variables $X_1 + \ldots + X_n$, with $\{X_i = 1\}$ denoting a head on the ith toss of a coin. 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 np is not large, a different type of approximation to the Binomial is better.

•Poisson distribution

<8.1> **Definition.** *A random variable Y is said to have a* POISSON DISTRIBUTION *with parameter λ, abbreviated to* Poisson(λ)*, if it can take values* $0, 1, 2, \ldots$ *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 Poisson(λ) appears as an approximation to the Bin(n, p) when n is large, p is small, and $\lambda = np$:

$$\begin{aligned}\binom{n}{k} p^k (1-p)^{n-k} &= \frac{n(n-1)\ldots(n-k+1)}{k!}\left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &\approx \frac{\lambda^k}{k!}\left(1 - \frac{\lambda}{n}\right)^n \qquad \text{if } k \text{ small relative to } n \\ &\approx \frac{\lambda^k}{k!} e^{-\lambda} \qquad \text{if } n \text{ is large}\end{aligned}$$

The exponential factor comes from:

$$\log\left(1 - \frac{\lambda}{n}\right)^n = n \log\left(1 - \frac{\lambda}{n}\right) = n\left(-\frac{\lambda}{n} - \frac{1}{2}\frac{\lambda^2}{n^2} - \ldots\right) \approx -\lambda \qquad \text{if } \lambda/n \approx 0.$$

By careful consideration of the error terms, one can give explicit bounds for the error of approximation. For example, there exists a constant C, such that, if X is distributed Bin(n, p) and Y is distributed Poisson(np) then

$$\sum_{k=0}^{\infty} |\mathbb{P}\{X = k\} - \mathbb{P}\{Y = k\}| \le Cp$$

Le Cam[1] has sketched a proof showing that C can be taken equal to 4. Clearly the Poisson is an excellent approximation when p is small.

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

[1] Page 187 of "On the distribution of sums of independent random variables", in *Bernouilli, Bayes, Laplace: Anniversary Volume*, J. Neyman and L Le Cam, eds., Springer-Verlag 1965.

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 X' has Bin(n, p) distribution independent of X, then $X + X'$ has a Bin$(n + m, p)$ distribution. Putting $\lambda = mp$ and $\mu = np$ one would then suspect that the sum of independent Poisson(λ) and Poisson(μ) distributed random variables is Poisson$(\lambda + \mu)$ distributed.

<8.2> **Exercise.** Verify the properties of the Poisson distribution suggested by the Binomial analogy: If Y has a Poisson(λ) distribution, show that

(i) $\mathbb{E}Y = \lambda$

(ii) $\text{var}(Y) = \lambda$

Also, if Y' has a Poisson(μ) distribution independent of Y, show that

(iii) $Y + Y'$ has a Poisson$(\lambda + \mu)$ distribution

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

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

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

If we were to calculate $\mathbb{E}(Y^2)$ 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?

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

Now calculate the variance.

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

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

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

$$= \frac{e^{-\lambda-\mu}}{k!}(\lambda+\mu)^k.$$

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

Question: How should you interpret the notation in the last calculation when $k = 0$? I always feel slightly awkward about a contribution from $k-1$ when $k = 0$.

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 modelled by Poisson distributions, at least as a first approximation. In some situations it makes sense to think of the counts as the number of successes in a large number of independent trials, with the chance of a success on any particular trial being very small ("rare events"). In such a setting, the Poisson arises as an approximation for the Binomial. The Poisson approximation also applies in many settings where the trials are "almost independent" but not quite.

<8.3> **Example.** Suppose N letters are placed at random into N envelopes, one letter per envelope. The total number of correct matches, X, can be written as a sum $X_1 + \ldots + X_N$ 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

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

and

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

We could eliminate 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 $\text{Bin}(N, 1/N)$ distribution, which approximates Poisson(1) if N is large.

We can get some supporting evidence for X having something close to a Poisson(1) distribution by calculating some MOMENTS:

$$\mathbb{E}X = \sum_{i\le N}\mathbb{E}X_i = N\mathbb{P}\{X_i = 1\} = 1$$

and

$$\begin{aligned}\mathbb{E}X^2 &= \mathbb{E}\left(X_1^2 + \ldots + X_N^2 + 2\sum_{i<j}X_iX_j\right)\\ &= N\mathbb{E}X_1^2 + 2\binom{N}{2}\mathbb{E}X_1X_2 \qquad \text{by symmetry}\\ &= N\mathbb{P}\{X_1 = 1\} + (N^2-N)\mathbb{P}\{X_1 = 1,\ X_2 = 1\}\\ &= N \times \frac{1}{N} + (N^2-N)\times\frac{1}{N(N-1)}\\ &= 2\end{aligned}$$

Compare with Exercise <8.2>, which implies $\mathbb{E}Y = 1$ and $\mathbb{E}Y^2 = 2$ for a Y distributed Poisson(1). □

Using the METHOD OF INCLUSION AND EXCLUSION, it is possible[2] to calculate the exact distribution of the random variable X from the previous Example:

<8.4> $$\mathbb{P}\{X = k\} = \frac{1}{k!}\left(1 - 1 + \frac{1}{2!} - \frac{1}{3!} - \ldots \pm \frac{1}{(N-k)!}\right)$$

[2] Feller Vol 1, Chapter 4

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

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

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

✠ ✠ ✠ ✠ ✠ ✠ ✠ ✠ ✠ ✠ ✠ ✠ ✠ ✠ ✠ ✠

One of the most elegant developments in modern probability theory is a general method for establishing approximation results, due principally to Charles Stein. It has been developed by Chen and Stein to derive Poisson approximations for many situations[3] The method is elementary—in the sense that it makes use of probabilistic techniques at the level of Statistics 241—but extremely subtle. The next Example illustrates the Chen-Stein method by deriving a Poisson approximation for the matching problem.

<8.5> **Example.** Consider once more the matching problem described in Example <8.3>. Use the Chen-Stein method to establish the approximation

$$\mathbb{P}\{X = k\} \approx \frac{e^{-1}}{k!} \qquad \text{for } k = 0, 1, 2, \dots$$

starting point is a curious connection between the Poisson(1) and the function $g(\cdot)$ defined by $g(0) = 0$ and

$$g(j) = \int_0^1 e^{-t} t^{j-1} dt \qquad \text{for } j = 1, 2, \dots$$

Notice that $0 \le g(j) \le 1$ for all j. Also, integration by parts shows that

$$g(j+1) = jg(j) - e^{-1} \qquad \text{for } j = 0, 1, 2, \dots$$

and direct calculation gives

$$g(1) = 1 - e^{-1}$$

More succinctly,

<8.6> $$g(j+1) - jg(j) = 1\{j = 0\} - e^{-1} \qquad \text{for } j = 0, 1, \dots$$

Actually the definition of $g(0)$ has no effect on the validity of the assertion when $j = 0$; you could give $g(0)$ any value you liked.

Suppose Y has a Poisson(1) distribution. Substitute Y for j in <8.6>, then take expectations to get

$$\mathbb{E}\left(g(Y+1) - Yg(Y)\right) = \mathbb{E}1\{Y = 0\} - e^{-1} = \mathbb{P}\{Y = 0\} - e^{-1} = 0.$$

A similar calculation with X in place of Y gives

<8.7> $$\mathbb{P}\{X = 0\} - e^{-1} = \mathbb{E}\left(g(X+1) - Xg(X)\right).$$

If we can show that the right-hand side is close to zero then we will have

$$\mathbb{P}\{X = 0\} \approx e^{-1},$$

which is the desired Poisson approximation for $\mathbb{P}\{X = k\}$ when $k = 0$. A simple symmetry argument will then give the approximation for other k values.

There is a beautiful probabilistic trick for approximating the right-hand side of <8.7>. Write the $Xg(X)$ contribution as

<8.8> $$\mathbb{E}Xg(X) = \mathbb{E}\sum_{i=1}^{N} X_i g(X) = \sum_{i=1}^{N} \mathbb{E}X_i g(X) = N\mathbb{E}X_1 g(X)$$

[3] See the 1992 book by Barbour, Holst, and Janson, *Poisson Approximation*, for a detailed discussion of the Chen-Stein method for deriving Poisson approximations.

The trick consists of a special two-step method for allocating letters at random to envelopes, which initially gives letter 1 a special role.

(1) Put letter 1 in envelope 1, then allocate letters $2, \ldots, N$ to envelopes $2, \ldots, N$ in random order, one letter per envelope. Write $1 + Z$ for the total number of matches of letters to correct envelopes. (The 1 comes from the forced matching of letter 1 and envelope 1.) Notice that $\mathbb{E}Z = 1$, as shown in Example <8.3>.

(2) Choose an envelope R at random (probability $1/N$ for each envelope), then swap letter 1 with the letter in the chosen envelope.

Notice that X_1 is independent of Z, because of step 2. Notice also that $X \neq Z$ if and only if the envelope R chosen in step 2 does not contain its correct letter. Thus

$$\mathbb{P}\{X \neq Z \mid Z = k\} = \frac{k+1}{N}$$

and

$$\mathbb{P}\{X \neq Z\} = \sum_k \frac{k+1}{N}\mathbb{P}\{Z = k\} = \frac{\mathbb{E}Z + 1}{N} = \frac{2}{N}$$

That is, the construction gives $X = Z$ with high probability.

From the fact that when $X_1 = 1$ (that is, $R = 1$) we have $X = Z + 1$, deduce that

<8.9> $$X_1 g(X) = X_1 g(1 + Z)$$

The asserted equality holds trivially when $X_1 = 0$. Take expectations. Then argue that

$$\begin{aligned} \mathbb{E}Xg(X) &= N\mathbb{E}X_1 g(X) && \text{by } <8.8> \\ &= N\mathbb{E}X_1 g(1+Z) && \text{by } <8.9> \\ &= N\mathbb{E}X_1 \mathbb{E}g(1+Z) && \text{by independence of } X_1 \text{ and } Z \\ &= \mathbb{E}g(1+Z) \end{aligned}$$

The right-hand side of <8.7> therefore equals $\mathbb{P}\left(g(X+1) - g(Z+1)\right)$. On the part of the sample space where $X = Z$ the two terms cancel; on the part where $X \neq Z$, the contribution lies between ± 1 because $0 \leq g(j) \leq 1$ for $j = 1, 2, \ldots$. Thus

$$|\mathbb{P}\left(g(X+1) - g(Z+1)\right)| \leq 1 \times \mathbb{P}\{X \neq Z\} \leq \frac{2}{N}$$

and

<8.10> $$|\mathbb{P}\{X = 0\} - e^{-1}| = |\mathbb{P}\left(g(X+1) - Xg(X)\right)| \leq 2/N$$

The exact expression for $\mathbb{P}\{X = 0\}$ from <8.4> shows that $2/N$ greatly overestimates the error of approximation, but at least it tends to zero as N gets large.

After all that work to justify the Poisson approximation to $\mathbb{P}\{X = k\}$ for $k = 0$, you might be forgiven for shrinking from the prospect of extending the approximation to larger k. Fear not! The worst is over.

For $k = 1, 2, \ldots$ the event $\{X = k\}$ specifies exactly k matches. There are $\binom{N}{k}$ choices for the matching envelopes. By symmetry, the probability of matches only in a particular set of k envelopes is the same for each specific choice of the set of k envelopes. It follows that

$$\mathbb{P}\{X = k\} = \binom{N}{k}\mathbb{P}\{\text{envelopes } 1, \ldots, k \text{ match; the rest don't}\}$$

The probability of getting matches in envelopes $1, \ldots, k$ equals

$$\frac{1}{N(N-1)\ldots(N-k+1)}$$

The conditional probability

$$\mathbb{P}\{\text{envelopes } k+1, \ldots, N \text{ don't match } \mid \text{ envelopes } 1, \ldots, k \text{ match}\}$$

is equal to the probability of zero matches when $N - k$ letters are placed at random into their envelopes. If N is much larger than k, this probability is close to e^{-1}, as shown above. Thus

$$\mathbb{P}\{X = k\} \approx \frac{N!}{k!(N-k)!} \frac{1}{N(N-1)(N-2)\dots(N-k+1)} e^{-1} = \frac{e^{-1}}{k!}$$

More formally, for each fixed k,

$$\mathbb{P}\{X = k\} \to \frac{e^{-1}}{k!} = \mathbb{P}\{Y = k\} \qquad \text{as } N \to \infty,$$

where Y has the Poisson(1) distribution. $\square$