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 *i*th 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 too large, a different type of approximation to the Binomial is better.

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

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

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

Throughout this Chapter I will use Q_{λ} to denote the Poisson(λ) distribution. That is, Q_{λ} is a probability distribution concentrated on \mathbb{N}_0 for which

$$Q_{\lambda}\{k\} = \frac{e^{-\lambda}\lambda^k}{k!}$$
 for $k = 0, 1, 2, ...$

Example $\langle 8.1 \rangle$: Poisson(*np*) approximation to the Binomial(*n*, *p*)

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 would then suspect that the sum of independent Poisson (λ) and Poisson (μ) distributed random variables is Poisson $(\lambda + \mu)$ distributed.

Example <8.2>: If X has a Poisson(λ) distribution, then $\mathbb{E}X = \operatorname{var}(X) = \lambda$. If also Y has a Poisson(μ) distribution, and Y is independent of X, then X + Y has a Poisson($\lambda + \mu$) distribution.

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 a sum of independent counts.

In fact, modern probability methods can handle situations much more general than approximation to the Binomial. For example, suppose $S = X_1 + X_2 + ... + X_n$, where X_i

has a Bin $(1, p_i)$ distribution, for constants p_1, p_2, \ldots, p_n that are not necessarily all the same. Suppose the X_i 's are independent. If the p_i 's are not all the same then S does not have a Binomial distribution. Nevertheless, the **Chen-Stein method** (see Barbour, Holst & Janson 1992 for an extensive discussion of the method) can be used to show that

 $\max_{A} |\mathbb{P}\{S \in A\} - Q_{\lambda}(A)| \le (1 - e^{-\lambda})/\lambda \sum_{i=1}^{n} p_i^2 \quad \text{where } \lambda = p_1 + \ldots + p_n.$

The method of proof is elementary—in the sense that it makes use of probabilistic techniques at the level of Statistics 241—but extremely subtle.

REMARK. The maximum here runs over all subsets A of \mathbb{N}_0 . In fact the maximum is achieved when $A = \{k \in \mathbb{N}_0 : \mathbb{P}\{S = k\} \ge Q_\lambda\{k\}\}$, in which case $|\mathbb{P}\{S \in A\} - Q_\lambda(A)|$ equals $\frac{1}{2} \sum_{k=0}^{\infty} |\mathbb{P}\{S = k\} - Q_\lambda\{k\}|$. This quantity is called the **total variation distance** between Q_λ and the distribution of X; it gives a very strong control over the errors in the approximation.

Note also that $(1 - e^{-\lambda})/\lambda \leq \min(1, 1/\lambda)$. Indeed, the left-hand side is close to 1 when $\lambda \approx 0$ and it behaves like $1/\lambda$ when λ is large.

When all the p_i are equal to some small p, we get a bound on the total variation distance between the Binomial(n, p) and the Poisson(np) smaller than min (p, np^2) . This bound makes precise the traditional advice that the Poisson approximation is good "when pis small and np is not too big". (In fact, the tradition was a bit conservative.)

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 $S = X_1 + X_2 + ... + X_n$, where X_i has a Bin $(1, p_i)$ distribution, for constants $p_1, p_2, ..., p_n$ that are not necessarily all the same. Define $S_{-i} = S - X_i = \sum_{j:j \neq i} X_j$. The random variables $X_1, ..., X_n$ are said to be **positively associated**¹ if

 $\mathbb{P}\{S_{-i} \ge k \mid X_i = 1\} \ge \mathbb{P}\{S_{-i} \ge k \mid X_i = 0\}$ for each *i* and each $k \in \mathbb{N}_0$; they are said to be **negatively associated**² if

 $\mathbb{P}\{S_{-i} \ge k \mid X_i = 1\} \le \mathbb{P}\{S_{-i} \ge k \mid X_i = 0\}$ for each *i* and each $k \in \mathbb{N}_0$; With some work it can be shown that

The solide work it can be shown

 $\max_{A} |\mathbb{P}\{S \in A\} - Q_{\lambda}(A)|$

$$\leq (1 - e^{-\lambda})/\lambda \begin{cases} \left(2\sum_{i=1}^{n} p_i^2 + \operatorname{var}(S) - \lambda\right) & \text{under positive association} \\ \left(\lambda - \operatorname{var}(S)\right) & \text{under negative association} \end{cases}$$

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 $\langle 8.3 \rangle$: Poisson approximation for a matching problem: assignment of *n* letters at random to *n* envelopes, one per envelope.

The Appendix to this Chapter provides a more detailed introduction to the Chen-Stein method, as applied to another aspect of the matching problem. (I have taken advantage of a few special features of the matching problem to simplify the exposition.) You could safely skip this Appendix. For more details, see the monograph by Barbour et al. (1992).

¹ not standard terminology

² not standard terminology

Examples for Chapter 8

<8.1> **Example.** The Poisson(λ) appears as an approximation to the Bin(n, p) when n is large, p is small, and $\lambda = np$:

$$\binom{n}{k}p^{k}(1-p)^{n-k} = \frac{n(n-1)\dots(n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^{k} \left(1-\frac{\lambda}{n}\right)^{n-k}$$
$$= 1 \times \left(1-\frac{1}{n}\right) \times \dots \left(1-\frac{k-1}{n}\right) \left(1-\frac{\lambda}{n}\right)^{-k} \frac{\lambda^{k}}{k!} \left(1-\frac{\lambda}{n}\right)^{n}$$
$$\approx \frac{\lambda^{k}}{k!} \left(1-\frac{\lambda}{n}\right)^{n} \qquad \text{if } k \text{ is small relative to } n$$
$$\approx \frac{\lambda^{k}}{k!}e^{-\lambda} \qquad \text{if } n \text{ is large.}$$

The final $e^{-\lambda}$ comes from an approximation to the logarithm,

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

- <8.2> **Example.** Verify the properties of the Poisson distribution suggested by the Binomial analogy: If X has a Poisson(λ) distribution, show that
 - (i) $\mathbb{E}X = \lambda$
 - (ii) $var(X) = \lambda$

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

(iii) X + 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.

$$\mathbb{E}X = \sum_{k=0}^{\infty} k \mathbb{P}\{X = k\} = \sum_{k=1}^{\infty} k \frac{e^{-\lambda} \lambda^k}{k!} \quad \text{What happens to } k = 0?$$
$$= e^{-\lambda} \lambda \sum_{k-1=0}^{\infty} \frac{\lambda^{k-1}}{(k-1)!}$$
$$= e^{-\lambda} \lambda e^{\lambda}$$
$$= \lambda.$$

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

If we 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?

$$\mathbb{E}(X^2 - X) = \sum_{k=0}^{\infty} k(k-1) \mathbb{P}\{X = k\}$$

= $e^{-\lambda} \sum_{k=2}^{\infty} k(k-1) \frac{\lambda^k}{k!}$ What happens to $k = 0$ and $k = 1$?
= $e^{-\lambda} \lambda^2 \sum_{k-2=0}^{\infty} \frac{\lambda^{k-2}}{(k-2)!}$
= λ^2 .

Now calculate the variance.

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

Statistics 241: 18 October 2005

© David Pollard

Chapter 8

For assertion (iii), first note that X + Y can take only values 0, 1, 2.... 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.

$$\mathbb{P}\{X+Y=k\} = \mathbb{P}\{X=0, Y=k\} + \mathbb{P}\{X=1, Y=k-1\} + \dots + \mathbb{P}\{X=k, Y=0\}$$

$$= \mathbb{P}\{X=0\}\mathbb{P}\{Y=k\} + \mathbb{P}\{X=1\}\mathbb{P}\{Y=k-1\} + \dots + \mathbb{P}\{X=k\}\mathbb{P}\{Y=0\}$$

$$= \frac{e^{-\lambda}\lambda^{0}}{0!}\frac{e^{-\mu}\mu^{k}}{k!} + \dots + \frac{e^{-\lambda}\lambda^{k}}{k!}\frac{e^{-\mu}\mu^{0}}{0!}$$

$$= \frac{e^{-\lambda-\mu}}{k!}\left(\frac{k!}{0!k!}\lambda^{0}\mu^{k} + \frac{k!}{1!(k-1)!}\lambda^{1}\mu^{k-1} + \dots + \frac{k!}{k!0!}\lambda^{k}\mu^{0}\right)$$

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

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

Question: How 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.

<8.3> **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 $X_1 + \ldots + X_n$ 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

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

and

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

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 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.

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

and

$$\mathbb{E}S^{2} = \mathbb{E}\left(X_{1}^{2} + \ldots + X_{n}^{2} + 2\sum_{i < j} X_{i}X_{j}\right)$$

= $n\mathbb{E}X_{1}^{2} + 2\binom{n}{2}\mathbb{E}X_{1}X_{2}$ 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.

Thus $var(S) = \mathbb{E}S^2 - (\mathbb{E}S)^2 = 1$. Compare with Example <8.2>, which gives $\mathbb{E}Y = 1$ and var(Y) = 1 for a Y distributed Poisson(1).

© David Pollard

DP: check result and Feller citation 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,

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

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!}=Q_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:

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

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*. We therefore have

$$\max_{i=1} |\mathbb{P}\{S \in A\} - Q_1(A)| \le 2\sum_{i=1}^n p_i^2 + \operatorname{var}(S) - 1 = 2/n.$$

As *n* gets large, the distribution of *S* does get close to the Poisson(1) in the strong, total variation sense. However, it is possible (see Barbour et al. 1992, page 73) to get a better bound by working directly from <8.4>

References

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

Feller, W. (1968), An Introduction to Probability Theory and Its Applications, Vol. 1, third edn, Wiley, New York.

APPENDIX: THE CHEN-STEIN METHOD FOR THE MATCHING PROBLEM

You might actually find the argument leading to the final bound of Example $\langle 8.3 \rangle$ more enlightening than the condensed exposition that follows. In any case, you can safely stop reading this chapter right now without suffering major probabilistic deprivation.

You were warned.

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

$$\mathbb{P}{S = k} \approx \frac{e^{-1}}{k!}$$
 for $k = 0, 1, 2, ...$

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

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

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

$$g(j+1) = jg(j) - e^{-1}$$
 for $j = 1, 2, ...$

and direct calculation gives

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

More succinctly,

$$g(j+1) - jg(j) = 1\{j=0\} - e^{-1}$$
 for $j = 0, 1, ...$

Statistics 241: 18 October 2005

© David Pollard

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 <8.5>, then take expectations to get

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

A similar calculation with S in place of Y gives

$$\mathbb{P}\{S=0\} - e^{-1} = \mathbb{E}\left(g(S+1) - Sg(S)\right)$$

<8.6>

<8.7>

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

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

which is the desired Poisson approximation for $\mathbb{P}{S = k}$ 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 $\langle 8.6 \rangle$. Write the Sg(S) contribution as

$$\mathbb{E}Sg(S) = \mathbb{E}\sum_{i=1}^{n} X_i g(S) = \sum_{i=1}^{n} \mathbb{E}X_i g(S) = n\mathbb{E}X_1 g(S)$$

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

- Put letter 1 in envelope 1, then allocate letters 2,..., n to envelopes 2,..., 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.) Write Y_j for the letter that goes into envelope j. Notice that EZ = 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. Indeed,

$$\mathbb{P}\{X_1 = 1 \mid Z = k\} = \mathbb{P}\{R = 1 \mid Z = k\} = 1/n \quad \text{for each } k.$$

Notice also that

$$S = \begin{cases} 1+Z & \text{if } R = 1\\ Z-1 & \text{if } R \ge 2 \text{ and } Y_R = R\\ Z & \text{if } R \ge 2 \text{ and } Y_R \neq R \end{cases}$$

Thus

$$\mathbb{P}\{S \neq Z \mid Z = k\} = \mathbb{P}\{R = 1\} + \sum_{j \ge 2} \mathbb{P}\{R = j, Y_j = j \mid Z = k\}$$
$$= \frac{1}{n} + \frac{1}{n} \sum_{j \ge 2} \mathbb{P}\{Y_j = j \mid Z = k\}$$
$$= \frac{k+1}{n}$$

and

$$\mathbb{P}\{S \neq Z\} = \sum_{k} \frac{k+1}{n} \mathbb{P}\{Z=k\} = \frac{\mathbb{E}Z+1}{n} = \frac{2}{n}$$

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

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

$$<8.8>$$
 $X_1g(S) = X_1g(1+Z)$

Statistics 241: 18 October 2005

$$= n\mathbb{E}X_1g(1+Z) \qquad \text{by } <8.8>$$

= $n\mathbb{E}X_1\mathbb{E}g(1+Z) \qquad \text{by independence of } X_1 \text{ and } Z$
= $\mathbb{E}g(1+Z)$

Thus the right-hand side of $\langle 8.6 \rangle$ equals $\mathbb{E}(g(S+1) - g(Z+1))$. On the event $\{S = Z\}$ the two terms cancel; on the event $\{S \neq Z\}$, the difference g(S+1) - g(Z+1) lies between ± 1 because $0 \leq g(j) \leq 1$ for j = 1, 2, ... Combining these two contributions, we get

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

 $\mathbb{E}Sg(S) = n\mathbb{E}X_1g(S)$ by <8.7>

$$\mathbb{P}\left(g(S+1) - g(Z+1)\right) \mid \le 1 \times \mathbb{P}\{S \neq Z\} \le \frac{2}{n}$$

and

Chapter 8

$$|\mathbb{P}\{S=0\} - e^{-1}| = |\mathbb{P}(g(S+1) - Sg(S))| \le 2/n.$$

The exact expression for $\mathbb{P}{S = 0}$ 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 $\mathbb{P}{S = k}$ 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, ... the event $\{S = k\}$ 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

$$\mathbb{P}{S = k} = \binom{n}{k} \mathbb{P}{\text{envelopes } 1, \dots, k \text{ match; the rest don't}}$$

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

$$\frac{1}{n(n-1)\dots(n-k+1)}.$$

The conditional probability

 \mathbb{P} {envelopes k + 1, ..., n don't match | envelopes 1, ..., k 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

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

More formally, for each fixed k,

$$\mathbb{P}\{S=k\} \to \frac{e^{-1}}{k!} = \mathbb{P}\{Y=k\} \quad \text{as } n \to \infty,$$

where Y has the Poisson(1) distribution.

© David Pollard

<8.9>