Chapter 8

Poisson approximations

The Binomial distribution 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 $\lambda$ if it can take values in $\mathbb{N}_0$, the set of nonnegative integers, with probabilities

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

The parameter $\lambda$ must be positive. The distribution is denoted by Poisson$(\lambda)$.

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

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

Example <8.1>: Poisson$(np)$ approximation to the Binomial$(n, p)$

The Poisson inherits several properties from the Binomial. For example, the Binomial$(n, p)$ has expected value $np$ and variance $np(1-p)$. One might suspect that the Poisson$(\lambda)$ 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 Binomial$(m, p)$ distribution and $Y$ has a Binomial$(n, p)$ distribution independent of $X$, then $X + Y$ has a Binomial$(n + m, p)$ distribution. Putting $\lambda = mp$ and $\mu = np$ one would then suspect that the sum of independent Poisson$(\lambda)$ and Poisson$(\mu)$ distributed random variables is Poisson$(\lambda + \mu)$ distributed.

Example <8.2>: If $X$ has a Poisson$(\lambda)$ distribution, then $E X = \text{var}(X) = \lambda$. If also $Y$ has a Poisson$(\mu)$ 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 $S = X_1 + X_2 + \ldots + X_n$, where $X_i$
has a Binomial$(1, p_i)$ 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

\[
\max_A |\mathbb{P}[S \in A] - Q_\lambda(A)| \leq \left(1 - e^{-\lambda}\right)/\lambda \sum_{i=1}^{n} p_i^2 \quad \text{where } \lambda = p_1 + \ldots + p_n.
\]

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 $A = \{k \in \mathbb{N}_0 : \mathbb{P}[S = k] \geq Q_\lambda(k)\}$, 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 $Q_\lambda$ 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 $\lambda$ 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 $p$ is 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 $S = X_1 + X_2 + \ldots + X_n$, where $X_i$ has a Binomial$(1, p_i)$ distribution, for constants $p_1, p_2, \ldots, p_n$ that are not necessarily all the same. Define $S_{-i} = S - X_i = \sum_{j \neq i} X_j$. The random variables $X_1, \ldots, X_n$ are said to be **positively associated**\(^1\) if

\[
\mathbb{P}[S_{-i} \geq k | X_i = 1] \geq \mathbb{P}[S_{-i} \geq k | X_i = 0] \quad \text{for each } i \text{ and each } k \in \mathbb{N}_0;
\]

they are said to be **negatively associated**\(^2\) if

\[
\mathbb{P}[S_{-i} \geq k | X_i = 1] \leq \mathbb{P}[S_{-i} \geq k | X_i = 0] \quad \text{for each } i \text{ and each } k \in \mathbb{N}_0;
\]

With some work it can be shown that

\[
\max_A |\mathbb{P}[S \in A] - Q_\lambda(A)| \leq \left(1 - e^{-\lambda}\right)/\lambda \left\{\begin{array}{ll}
2 \sum_{i=1}^{n} p_i^2 + \text{var}(S) - \lambda & \text{under positive association} \\
\lambda - \text{var}(S) & \text{under negative association}
\end{array}\right.
\]

These bounds take advantage of the fact that $\text{var}(S)$ would be exactly equal to $\lambda$ if $S$ had a Poisson$(\lambda)$ 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 <8.3>:** Poisson approximation for a matching problem: assignment of $n$ letters at random to $n$ envelopes, one per envelope.

---

\(^1\) 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 λ = np:

\[
\binom{n}{k} p^k (1 - p)^{n-k} = \frac{n(n-1) \ldots (n-k+1)}{k!} \left( \frac{\lambda}{n} \right)^k \left( 1 - \frac{\lambda}{n} \right)^{n-k} = 1 \times \left( 1 - \frac{1}{n} \right) \times \ldots \left( 1 - \frac{k-1}{n} \right) \left( 1 - \frac{\lambda}{n} \right)^{-k} \frac{\lambda^k}{k!} \left( 1 - \frac{\lambda}{n} \right)^n \approx \frac{\lambda^k}{k!} \left( 1 - \frac{\lambda}{n} \right)^n \text{ if } k \text{ is small relative to } n
\]

\[\approx \frac{\lambda^k}{k!} e^{-\lambda} \quad \text{if } n \text{ is large.}\]

The final e^{-λ} 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) \( \text{var}(X) = \lambda \)

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

(iii) \( X + Y \) has a Poisson(λ + µ) distribution

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

\[\mathbb{E}X = \sum_{k=0}^{\infty} k \mathbb{P}(X = k) = \sum_{k=1}^{\infty} k \frac{e^{-\lambda} \lambda^k}{k!} = e^{-\lambda} \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = e^{-\lambda} \lambda e^\lambda = \lambda.\]

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?

\[\mathbb{E}(X^2 - X) = \sum_{k=0}^{\infty} k(k-1) \mathbb{P}(X = k) = e^{-\lambda} \sum_{k=2}^{\infty} k(k-1) \frac{\lambda^k}{k!} \quad \text{What happens to } k = 0 \text{ and } k = 1? \]

\[= e^{-\lambda} \lambda^2 \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} = \lambda^2.\]

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 $0, 1, 2, \ldots$. 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 $Y$.

\[
\mathbb{P}(X + Y = k) = \mathbb{P}(X = 0, Y = k) + \mathbb{P}(X = 1, Y = k - 1) + \ldots + \mathbb{P}(X = k, Y = 0)
\]

\[
= \mathbb{P}(X = 0) \mathbb{P}(Y = k) + \mathbb{P}(X = 1) \mathbb{P}(Y = k - 1) + \ldots + \mathbb{P}(X = k) \mathbb{P}(Y = 0)
\]

\[
= e^{-\lambda} \frac{\lambda^0}{0!} e^{-\mu} \frac{\mu^k}{k!} + \ldots + e^{-\lambda} \frac{\lambda^k}{k!} e^{-\mu} \frac{\mu^0}{0!}
\]

\[
= e^{-\lambda - \mu} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \frac{\mu^k}{k!} + \ldots + \frac{k!}{1!(k-1)!} e^{-\lambda - \mu} \frac{\lambda^{k-1}}{k-1!} \frac{\mu^k}{k!} + \ldots + \frac{k!}{k!} e^{-\lambda - \mu} \frac{\lambda^0}{0!}
\]

\[
= e^{-\lambda - \mu} (\lambda + \mu)^k.
\]

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

**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

\[
p_i = \mathbb{P}(X_i = 1) = 1/n \quad \text{for each } i
\]

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 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.

\[
\mathbb{E}S = \sum_{i \leq n} \mathbb{E}X_i = n \mathbb{P}(X_i = 1) = n
\]

and

\[
\mathbb{E}S^2 = \mathbb{E}
\left(n^2 - n\right) X_1 X_2 \quad \text{by symmetry}
\]

\[
= n \mathbb{P}(X_1 = 1) + (n^2 - n) \mathbb{P}(X_1 = 1, X_2 = 1)
\]

\[
= \left(n \times \frac{1}{n}\right) + (n^2 - n) \times \frac{1}{n(n-1)}
\]

\[
= 2.
\]

Thus $\text{var}(S) = \mathbb{E}S^2 - (\mathbb{E}S)^2 = 1$. Compare with Example <8.2>, 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,

\[ P(S = k) = \frac{1}{k!} \left( 1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} - \ldots \pm \frac{1}{(n-k)!} \right) \]

for \( k = 0, 1, \ldots, n \).

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

\[ \frac{1}{k!} \left( 1 - 1 + \frac{1}{2!} - \frac{1}{3!} - \ldots \right) = e^{-1} \frac{1}{k!} = Q_1(k), \]

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:

\[ P(S_i \geq k \mid X_i = 1) \geq P(S_i \geq k \mid X_i = 0) \]

for each \( i \) 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

\[ \max_A |P(S \in A) - Q_1(A)| \leq 2 \sum_{i=1}^n p_i^2 + \text{var}(S) - 1 = 2/n. \]

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

REFERENCES


APPENDIX: THE CHEN-STEIN METHOD FOR THE MATCHING PROBLEM

You might actually find the argument leading to the final bound of Example <8.3> 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

\[ P(S = k) \approx e^{-1} \frac{1}{k!} \]

for \( k = 0, 1, 2, \ldots \).

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 \( j = 1, 2, \ldots \).

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

\[ g(j + 1) = jg(j) - e^{-1} \]

and direct calculation gives

\[ g(1) = 1 - e^{-1}. \]

More succinctly,

\[ g(j + 1) - jg(j) = 1(j = 0) - e^{-1} \]

for \( j = 0, 1, \ldots \).
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

\[
E(g(Y + 1) - Y g(Y)) = E1\{Y = 0\} - e^{-1} = P\{Y = 0\} = e^{-1} = 0.
\]

A similar calculation with \( S \) in place of \( Y \) gives

\[
<8.6>

P\{S = 0\} - e^{-1} = E(g(S + 1) - S g(S)).
\]

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

\[P\{S = 0\} \approx e^{-1},\]

which is the desired Poisson approximation for \( P\{S = k\} \) 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 \( <8.6> \). Write the \( S g(S) \) contribution as

\[
<8.7>

E S g(S) = E \sum_{i=1}^{n} X_i g(S) = \sum_{i=1}^{n} E X_i g(S) = n 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.

1. 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 \( E Z = 1 \), 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. Indeed,

\[P\{X_1 = 1 \mid Z = k\} = P\{R = 1 \mid Z = k\} = 1/n \quad \text{for each} \; k.
\]

Notice also that

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

Thus

\[
P\{S \neq Z \mid Z = k\} = P\{R = 1\} + \sum_{j \geq 2} P\{R = j, Y_j = j \mid Z = k\}
\]

\[
= \frac{1}{n} + \frac{1}{n} \sum_{j \geq 2} P\{Y_j = j \mid Z = k\}
\]

\[
= \frac{k + 1}{n}
\]

and

\[
P\{S \neq Z\} = \sum_k \frac{k + 1}{n} P\{Z = k\} = \frac{EZ + 1}{n} = \frac{2}{n}.
\]

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_1 g(S) = X_1 g(1 + Z)
\]
The same equality holds trivially when $X_1 = 0$. Take expectations. Then argue that
\[
\mathbb{E}g(S) = n \mathbb{E}X_1g(S) \quad \text{by } \langle 8.7 \rangle \\
= n \mathbb{E}X_1g(1 + Z) \quad \text{by } \langle 8.8 \rangle \\
= n \mathbb{E}X_1 \mathbb{E}g(1 + Z) \quad \text{by independence of } X_1 \text{ and } Z \\
= \mathbb{E}g(1 + Z)
\]
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 $\pm 1$ because $0 \leq g(j) \leq 1$ for $j = 1, 2, \ldots$. Combining these two contributions, we get
\[
|\mathbb{P}(g(S + 1) - g(Z + 1))| \leq 1 \times \mathbb{P}(S \neq Z) \leq \frac{2}{n}
\]
and
\[
|\mathbb{P}(S = 0) - e^{-1}| = |\mathbb{P}(g(S + 1) - Sg(S))| \leq \frac{2}{n}.
\]
The exact expression for $\mathbb{P}(S = 0)$ from $\langle 8.4 \rangle$ 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 $k = 1, 2, \ldots$ 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
\[
\mathbb{P}(S = k) = \binom{n}{k} \mathbb{P}\{\text{envelopes 1, \ldots, } k \text{ match; the rest don’t}\}
\]
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
\[
\mathbb{P}(S = k) \approx \frac{n!}{k!(n-k)!} \frac{1}{n(n-1)(n-2)\ldots(n-k+1)} e^{-1} = \frac{e^{-1}}{k!}.
\]
More formally, for each fixed $k$,
\[
\mathbb{P}(S = k) \rightarrow \frac{e^{-1}}{k!} = \mathbb{P}(Y = k) \quad \text{as } n \rightarrow \infty,
\]
where $Y$ has the Poisson(1) distribution. □