A4 From binomial to normal distributions

A4.1 The binomial distribution

Consider a case where the outcome of an observation x can be either 0 or 1 (or tail or head, or no or yes, or false or true, or absent or present, or whatever binary choice you wish to define). Let the probability of obtaining 1 be equal to p, meaning $E[x] = p$. Then for two observations the following combinations may occur: 00, 01, 10, 11. Assuming that successive observations are independent, the probability $f(k) (k = 0, 1, 2)$ that exactly k times a 1 is observed is

$$\begin{aligned} f(0) &= (1-p)^2 \\ f(1) &= 2p(1-p) \\ f(2) &= p^2. \end{aligned} \tag{A4.1}$$

In general: the probability $f(k; n)$ that in n independent observations exactly k times a 1 is observed equals

$$f(k; n) = \binom{n}{k} p^k (1-p)^{(n-k)}, \tag{A4.2}$$

where

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \tag{A4.3}$$

is the *binomial n over k*, i.e., the number of ways k items can be chosen from a collection of n items. For the case considered above: $n = 2$, the three binomial coefficients $(k = 0, 1, 2)$ are respectively 1, 2 and 1; these are the coefficients in $f(k)$ of (A4.1). This is the *binomial distribution*.

Note that the sum of all probabilities equals 1:

$$\sum_{k=0}^{n} f(k; n) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{(n-k)} = (p + 1 - p)^n = 1. \tag{A4.4}$$

The mean $E[k]$ is defined by the sum

$$E[k] = \sum_{k=0}^{n} kf(k;n); \tag{A4.5}$$

this can be worked out as

$$E[k] = pn \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1}(1-p)^{\{n-1-(k-1)\}} = pn. \tag{A4.6}$$

Similarly (details are left to the reader), the variance equals

$$\begin{aligned} E[(k-pn)^2] &= E[k^2] - 2pnE[k] + (pn)^2 \\ &= \sum_{k=1}^{n} k^2 f(k;n) - (pn)^2 = p(1-p)n. \end{aligned} \tag{A4.7}$$

A4.2 The multinomial distribution

When there are more than one (e.g. m) possible values for the sampled variable, with probabilities $p_1, p_2, \ldots, p_m$ ($\sum_i p_i = 1$), then the distribution is a *multinomial distribution*:

$$f(k_1, k_2, \ldots, k_m; n) = \frac{n!}{k_1!k_2!\ldots k_m!} \Pi_{i=1}^{m} p_i^{k_i}; \quad \sum_i k_i = n. \tag{A4.8}$$

This is an example of a multidimensional *joint* probability, meaning the probability that event 1 occurs k_1 times *and* event 2 occurs k_2 times *and* etc. The means and variances for each of the number of occurrences are the same as for the binomial distribution:

$$E[k_i] = \mu_i = np_i, \tag{A4.9}$$

$$E[(k_i - \mu_i)^2] = \sigma_i^2 = np_i(1-p_i). \tag{A4.10}$$

The fact that the sum of all k_i is constrained causes a covariance between k_i and $k_j (i \neq j)$:

$$covar(k_i, k_j) = E[(k_i - \mu_i)(k_j - \mu_j)] = -np_i p_j. \tag{A4.11}$$

The covariance matrix is a symmetric matrix of which the diagonal elements are the variances and the non-diagonal elements are the covariances.

A4.3 The Poisson distribution

From binomial to Poisson

Consider a suspension of small particles. You want to determine the average number of particles per unit volume by counting the number of particles under a microscope in a sample of $0.1 \times 0.1 \times 0.1$ mm (10^{-6} cm^3). If the number density is known, and thus the average number of particles in the sample volume is known, what then is the probability of finding exactly k particles in the small volume?

Let the average number of particles in the sample volume be μ. Divide the sample volume into a large number n of cells, small enough to contain no more than one particle. The probability that a specified cell contains a particle equals $p = \mu/n$. The probability that precisely k particles will be found in the sample volume equals the binomial distribution $f(k; n)$ with $p = \mu/n$.

Equivalently you may consider another example: Electrical impulses (or photons, or gamma quanta, or any other short events) occur randomly and independently of each other. You observe the events during a given time span T. If the average number of events within a time T is known, what then is the probability that precisely k events are observed in a time span of length T? In this case we divide the interval T into n short time intervals. Let the average number of events in a time T be μ. The probability that precisely k events will be counted in the interval T equals the binomial distribution $f(k; n)$ with $p = \mu/n$.

Now let the number of cells, or the number of time intervals, n go to infinity, while $pn = \mu$ is kept constant. This means that $p \to 0$, but in such a way that $pn = \mu$ remains the same. Thus $k \ll n$. The binomial coefficient then approaches $n^k/k!$:

$$\frac{n!}{k!(n-k)!} = \frac{n(n-1)\dots(n-k+1)}{k!} \approx \frac{n^k}{k!}, \tag{A4.12}$$

so that

$$p(k) \to \frac{n^k}{k!}\left(\frac{\mu}{n}\right)^k \left(1 - \frac{\mu}{n}\right)^{n-k}.$$

The term on the right approaches $e^{-\mu}$ because $n - k \to n$ and

$$\lim_{n\to\infty}\left(1 - \frac{\mu}{n}\right)^n = e^{-\mu}, \tag{A4.13}$$

from which it follows that

$$f(k) = \frac{\mu^k e^{-\mu}}{k!}. \tag{A4.14}$$

This is the probability mass function of the *Poisson distribution* for k, given the average μ.

The Poisson distribution is a discrete distribution: the observed number k can only assume positive integer values $0, 1, 2, \ldots$ The mean μ is a parameter of the distribution and can be any positive real number.

Properties of the Poisson distribution

It is easy to show that the Poisson distribution is normalized and that its mean equals μ. Prove this by using the series expansion

$$e^{\mu} = \sum_{k=0}^{\infty} \frac{\mu^k}{k!}. \tag{A4.15}$$

The variance of the distribution is

$$\mathrm{var}\,(k) = \sigma^2 = E[(k-\mu)^2] = \mu. \tag{A4.16}$$

This follows from $\sum_{k=0}^{\infty} k^2 \mu^k / k! = \mu^2 + \mu$, but it is also the limit of (A4.10) for $p \to 0$.

A4.4 The normal distribution

From Poisson to normal

For large values of μ the Poisson distribution approaches a normal distribution with mean μ and s.d. $\sqrt{\mu}$. When we attempt to derive this limit we must be very careful to retain a sufficiently high order in the approximations as terms tend to compensate each other.

Let both k and μ go simultaneously to ∞, but in a coordinated way. Define

$$x = \frac{k-\mu}{\sqrt{\mu}}; \quad k = \mu + x\sqrt{\mu},$$

and use the Stirling approximation of the factorial $k!$:

$$k! = k^k e^{-k} \sqrt{2\pi k}\,[1 + O(k^{-1})]. \tag{A4.17}$$

The logarithm of the Poisson probability (A4.14) expands in orders of k^{-1} as follows:

$$\begin{aligned} \ln f(k) &= k - \mu - k\ln(k/\mu) - \frac{1}{2}\ln(2\pi k) + O(k^{-1}) \\ &= x\sqrt{\mu} - (\mu + x\sqrt{\mu})\ln\left(1 + \frac{x}{\sqrt{\mu}}\right) - \frac{1}{2}\ln\left[2\pi\mu\left(1 + \frac{x}{\sqrt{\mu}}\right)\right]. \end{aligned}$$

Because $\ln\mu \to \infty$, the whole expression goes to $-\infty$! This is what we expect because we calculate the probability of finding precisely one (integer) value of k (which obviously goes to zero) and *not* the probability density of $f(x)$. When we expand the logarithm

$$\ln(1+z) = z - \frac{1}{2}z^2 + O(z^3), \tag{A4.18}$$

we find eventually that

$$\lim_{k\to\infty} \ln f(k) = -\frac{1}{2}x^2 - \frac{1}{2}\ln(2\pi\mu).$$

The distance between two successive discrete values of x is

$$\Delta x = \frac{k+1-\mu}{\sqrt{\mu}} - \frac{k-\mu}{\sqrt{\mu}} = \frac{1}{\sqrt{\mu}};$$

therefore there are $\sqrt{\mu}\,dx$ discrete values between x and $x+dx$. It follows that

$$f(x)\,dx = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)dx, \tag{A4.19}$$

what we set out to prove.