Mathematics for the Life Sciences:
Calculus, Modeling, Probability, and Dynamical Systems

The Poisson Approximation for the Binomial Distribution

The derivation of the error formula for the Poisson approximation to the binomial distribution (given without justification in Problem 3.7.15) is an application of asymptotic expansion, a topic well beyond the scope of the book. The idea is to partition the factors in the binomial distribution formula, along with the extra factors e^{μ} and $e^{-\mu}$, into the Poisson distribution formula along with three additional factors that approach 1 in the limits $p \to 0$ and $n \to \infty$. Asymptotic expansion of the additional factors yields a first-order approximation for the relative error in the Poisson formula.

Using the property $\mu = np$ of the binomial distribution, we can rewrite the factor p^k as μ^k/n^k. This change, along with the introduction of factors e^{μ} and $e^{-\mu}$ allows us to identify the Poisson formula as a factor of the binomial formula. We also rewrite $(1-p)^{n-k}$ as $(1-p)^n/(1-p)^k$. With these changes, we can factor the binomial distribution formula as

$$b_{n,p}(k) = \frac{n!}{n^k(n-k)!} \cdot e^{\mu}(1-p)^n \cdot \frac{1}{(1-p)^k} \cdot \frac{\mu^k}{k!}e^{-\mu}. \qquad (1)$$

The third factor is easily approximated as a geometric expansion about $p \to 0$:

$$\frac{1}{(1-p)^k} \sim \frac{1}{1-kp} \sim 1 + kp = 1 + \frac{k\mu}{n},$$

where the symbol $\sim$ is used to indicate equality up to and including the $O(p)$ terms. The second factor requires expansion in both the limits $p \to 0$ and $n \to \infty$. Using expansion as $p \to 0$, we have

$$\ln[e^{\mu}(1-p)^n] = \mu + n\ln(1-p) \sim \mu + n\left(-p - \frac{1}{2}p^2\right) = -\frac{np^2}{2} = -\frac{\mu^2}{2n}.$$

Thus,

$$e^{\mu}(1-p)^n \sim e^{-\mu^2/2n}.$$

From here, expansion as $n \to \infty$ yields

$$e^{\mu}(1-p)^n \sim 1 - \frac{\mu^2}{2n}.$$

The first factor can be written as

$$\frac{n}{n} \cdot \frac{n-1}{n} \cdots \frac{n-k+1}{n} = 1 \cdot \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right)$$

$$\sim 1 - \frac{1+2+\cdots+(k-1)}{n} = 1 - \frac{k(k-1)}{2n}.$$

Multiplying these factors yields

$$b_{n,p}(k) \sim \left(1 + \frac{-k^2 + k + 2k\mu - \mu^2}{2n}\right) f_\mu(k) = \left(1 + \frac{k - (k-\mu)^2}{2n}\right) f_\mu(k). \tag{2}$$

Equation (2) identifies the asymptotic approximation for the relative error. Another version of the error formula is given in terms of p rather than n:

$$b_{n,p}(k) \sim \left(1 + \frac{p}{2}\left[\frac{k - (k-\mu)^2}{\mu}\right]\right) f_\mu(k). \tag{3}$$

It is also possible to estimate the largest error in using the approximation for given n and p. By straightforward computation, the maximum error is for $k = 1$ if $\mu \leq 1$ and $k = 0$ for $1 < \mu \leq 2$. For larger μ, we can obtain an approximate formula by considering the quantities

$$m(k) \equiv \frac{k - (k-\mu)^2}{\mu} \frac{\mu^k}{k!},$$

which represent the absolute error without the additional factors $p/2$ and $e^{-\mu}$. Suppose $\mu = n + r$, where n is any integer greater than 1 and $0 < r < 1$. A few experiments show that the largest error occurs for either $k = n$ or $k = n + 1$. The equality point between them seems to be roughly at $r = 1/3$, but we can determine this analytically for arbitrary n. Substituting $\mu = n + r$ into $m(n) = m(n+1)$ yields

$$[n^2 + (1 - r^2)n - r^2]\frac{\mu^{n-1}}{(n+1)!} = [n^2 + (3r - r^2)n + (2r^2 - r^3)]\frac{\mu^{n-1}}{(n+1)!},$$

which reduces to

$$(1 - 3r)n = 3r - r^3.$$

We find the cutoffs to be roughly $r = 0.29$ for $n = 2$, $r = 0.31$ for $n = 3$, and $r \to 1/3$ as $n \to \infty$. We are not far off to simply take $r = 1/3$ for all n; hence, the largest absolute error is approximately

$$\max(|b_{n,p} - f_{np}|) \approx \frac{p}{2} f_{np}(k_m), \tag{4}$$

where k_m is the integer part of $np + \frac{2}{3}$.

http://www.springer.com/978-1-4614-7275-9

Mathematics for the Life Sciences
Calculus, Modeling, Probability, and Dynamical Systems
Ledder, G.
2013, XX, 431 p. 109 illus., Hardcover
ISBN: 978-1-4614-7275-9