The Poisson Distribution as an Approximation to the Binomial Distribution

Prerequisites

You should be familiar with both the binomial and the Poisson distributions.

Example (1)

A company sells memory sticks to shop retailers in boxes of 250. On average 1% of these are faulty. What is the chance of getting less than 4 defective memory sticks in a box?

(*a*) Solve this problem using the binomial distribution.

(*b*) It is stated that the Poisson distribution $Y \sim Po(2.5)$ is an approximation to the binomial distribution for this problem. Solve the problem again using this Poisson distribution and comment on the accuracy of the approximation.

Solution

(*a*) The probability that a component is faulty equals $p = 0.01$. Let X denote the number of faulty components in a box. X is binomial with parameters $n = 250, p = 0.01$.

$X \sim B(250, 0.01)$

$$\begin{aligned} P(X < 4) &= P(X \le 3) \\ &= P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) \\ &= \binom{250}{0}(0.01)^0(0.99)^{250} + \binom{250}{1}(0.01)^1(0.99)^{249} \\ &\quad + \binom{250}{2}(0.01)^2(0.99)^{248} + \binom{250}{3}(0.01)^3(0.99)^{247} \\ &= 0.081059 + 0.204693 + 0.257417 + 0.214948 \\ &= 0.7581 \text{ (4 d.p.)} \end{aligned}$$

(*b*) Let Y denote the number of faulty components in a box and we are given $Y \sim Po(2.5)$. From tables

$P(X<4)=P(X\le 3)=0.7576$

The absolute error is $|0.7581-0.7576|=0.0005$ and the percentage relative error is

$$\%\text{ relative error}=\frac{0.0005}{0.7581}\times 100=0.066\ (2\text{ s.f.})$$

This is very small making the approximation a good one.

Approximating the binomial distribution

The Poisson distribution may be viewed as a special form of the binomial distribution and it is not surprising, therefore, that it can be used as an approximation to it, under certain conditions.

The Poisson approximation to the binomial distribution

Given

$X\sim B(n,p)$

$n>50$

$np<5$

Then

X may be approximated by $Y\sim Po(\lambda)$ where $\lambda=np$

λ is the mean of the Poisson distribution.

The Poisson distribution becomes closer and closer to the binomial distribution, as the sample size (number of trials) gets larger and larger. In the limit as $n\to\infty$ the approximation becomes exact. When the population is not infinite the Poisson distribution is a better approximation to the binomial the smaller $\lambda=np$ is. It is a rule of thumb that $\lambda=np<5$ and $n>50$. Experience has shown that this gives a good approximation, though the view of what makes a "good approximation" may differ from context to context. For this reason some texts give a different value for *n*, the number of trials in the sample, for the binomial distribution to be approximated by the Poison distribution. The different values represent the different levels of closeness of the two distributions. As *n* gets larger and larger the binomial distribution, where *np* is small, gets closer and closer to the Poisson. At a certain level the two are so close that the difference does not matter. But some theorists are more cautious about just how close the approximations need to be than others. Hence, the difference in quoted values.

Example (2)

95% of laboratory mice that have been treated with a new drug are resistant to a strain of diphtheria, which is a bacterial infection. Let X be the number of mice in a sample of 70 mice that are resistant, and Y be the number that are not resistant. State which of X or Y can be approximated by a Poisson distribution. Use the Poisson distribution to find (*a*) the probability that exactly 67 mice are resistant to the bacterium and (*b*) the probability that at least 67 mice are resistant to the bacterium.

Solution

$X \sim (70, 0.95)$

$Y \sim (70, 0.05)$

For X, $n > 50$ but $np = 70 \times 0.95 = 66.5$ so clearly $np \gg 5$.

Thus, X cannot be approximated by a Poisson distribution.

However, for Y, $n > 50$, and $np = 70 \times 0.05 = 3.5$, so $np < 5$.

Y can be approximated by a Poisson distribution Y.

$Y \sim Po(3.5)$

(*a*) $P(X = 67) = P(Y = 3) = e^{-3.5}\dfrac{(3.5)^3}{3!} = 0.2158 \text{ (4 s.f.)}$.

(*b*) From tables $P(X \geq 67) = P(Y \leq 3) = 0.5366$.