The Poisson Approximation to Binomial Probabilities

The Poisson approximation is useful for estimating binomial probabilities with a large number of trials (more than 50, say), but the expected number of "successes" is small (less than 5, say.) It is sometimes called the law of "rare events".

Example (1) A certain disease occurs in 1.2% of the population. 100 people are selected at random from the population. This is a binomial experiment, where the number of trials 100 is large, and the expected number of people with the disease $\mu = 100(.012) = 1.2$ is small.

Recall that for binomial experiments, the expected number of successes is

$$\mu = n\,p$$

Poisson Approximation. Consider a binomial experiment with a large number of trials. Let μ be the expected number of successes, and suppose that μ is small. Then the Poisson approximation (to the binomial experiment) asserts that:

$$P(\text{exactly } k \text{ successes}) \approx \frac{\mu^k}{k!}\, e^{-\mu}$$

Note: To apply this to find the probability of $k = 0$ successes, remember that $\mu^0 = 1$, and that $0! = 1$. Thus, the Poisson approximation says:

$$P(0 \text{ successes}) \approx e^{-\mu}$$

Example (1) (continued) Then $\mu = 1.2$. The Poisson approximation says that, in our sample of 100 people, the probability that none of them have the disease is

$$P(0 \text{ have disease}) \approx e^{-1.2} \approx .301$$

The probability that exactly 2 have the disease is

$$P(\text{exactly 2 have disease}) \approx \frac{(1.2)^2}{2!}\, e^{-1.2} \approx .217$$

The Poisson approximation uses the fact that, if n is large, and μ is not large, then

$$\left(1 - \frac{\mu}{n}\right)^n \approx e^{-\mu}$$

For example,

$$\left(1 - \frac{1.2}{100}\right)^{100} = .299016\,, \qquad e^{-1.2} = .301194$$

To see how this applies to **Example (1)**, note that the probability of any 1 person having the disease is 1.2/100. Then the binomial probability formula says that the probability of 0 people having the disease in a sample of 100 is

$$C(100,0)\left(1 - \frac{1.2}{100}\right)^{100}\left(\frac{1.2}{100}\right)^{0} = \left(1 - \frac{1.2}{100}\right)^{100} \approx e^{-1.2}$$

The probability that exactly 2 people in the sample of 100 have the disease is

$$C(100,2)\left(1-\frac{1.2}{100}\right)^{98}\left(\frac{1.2}{100}\right)^{2}=\frac{100\cdot 99}{2!}\left(1-\frac{1.2}{100}\right)^{98}\frac{(1.2)^2}{(100)^2}$$

If we replace the 98 and the 99 by 100, the fractions clear to give

$$\frac{(1.2)^2}{2!}\left(1-\frac{1.2}{100}\right)^{100}\approx\frac{(1.2)^2}{2!}\,e^{-1.2}$$

which is the Poisson approximation.

An important application of the Poisson approximation in life sciences is to what are known as *scatter plots,* as illustrated in the following example.

Example (2) A 10 meter by 10 meter plot of land, is divided into a grid of 100 squares. 300 seeds are scattered on this plot. Assume that each seed falls at random, so that it is equally likely to fall on any of the 100 squares. Consider the square in the upper left hand corner. What is the probability that exactly 4 seeds fall in it? What is the probability that 0 seeds fall in it?

Solution (2) To see that this is a binomial experiment, think of this as dropping 300 seeds, one after the other, and recording whether the seed falls into the upper left hand square or not. "Success" means falling into the square, and that happens with probability $1/100 = .01$ (since there are 100 squares.) So, $n = 300$, and $p = .01$. The expected number of successes is $\mu = 500(.01) = 3$. By the Poisson approximation,

$$P(\text{exactly 4 seeds})\approx\frac{3^4}{4!}\,e^{-3}\approx .17$$

And,

$$P(0\text{ seeds})\approx e^{-3}\approx .05$$

A model of this experiment was done on the computer, choosing 300 random points in the square, and counting the number of points that land in each square.

1	1	1	3	6	4	4	2	2	3
3	3	1	2	4	4	3	4	6	1
3	5	6	2	3	3	3	2	2	0
5	5	5	3	4	3	4	2	2	4
3	1	1	3	2	0	1	6	1	2
5	3	1	4	5	3	0	6	1	5
2	1	6	5	0	2	0	2	2	2
4	3	3	4	3	3	4	4	2	4
1	8	5	4	2	3	2	3	1	4
3	6	2	3	3	2	4	3	5	3

The histogram shows the number of squares $S(k)$ containing k seeds, for $k = 0, \ldots, k = 10$.

k	0	1	2	3	4	5	6	7	8	9	10
$S(k)$	5	14	20	26	17	10	7	0	1	0	0

Since each square has probability e^{-3} of containing 0 seeds, we expect $e^{-3} \approx 5\%$ of the squares to contain 0 seeds. In this experiment we did get exactly 5 squares with 0 seeds.

Compare the histogram to the expected Poisson distribution. $P(k) \approx \dfrac{3^k}{k!}e^{-3}$

k	0	1	2	3	4	5	6	7	8	9	10
$P(k)$	.05	.15	.22	.22	.17	.10	.05	.02	.01	.002	.001

Example (3) A certain nucleus contains 1×10^6 genes. Each gene has probability 3.2×10^{-6} of being mutated. What is the probability that the nucleus contains at most 2 mutated genes?

Solution (3) This is binomial, where $P(\text{mutated}) = .0000032\,, \quad P(\text{not mutated}) = .9999968\,,$ and we run 1×10^6 trials. The expected number of mutated cells is

$$\mu = 1 \times 10^6\,(3.2 \times 10^{-6}) = 3.2$$

To find the probability of at most 2 mutated cells, we take $P(0) + P(1) + P(2)$, where

$$\begin{aligned} P(0\text{ mutated cells}) &\approx e^{-3.2} \approx .0408 \\ P(\text{exactly 1 mutated cell}) &\approx 3.2\,e^{-3.2} \approx .1304 \\ P(\text{exactly 2 mutated cells}) &\approx \frac{(3.2)^2}{2!}\,e^{-3.2} \approx .2087 \\ \text{So, } P(\text{at most 2 mutated cells}) &\approx .3799 \end{aligned}$$

Example (4) A computer chip contains 1000 transistors. Each transistor has probability .0025 of being defective. What is the probability that the chip contains at most 4 defective transistors?

Solution (4) When we make a chip we are running 1000 trials (trial = make 1 transistor) and for each trial $P(\text{defective}) = .0025$. The expected number of defective transistors is $\mu = 1000(.0025) = 2.5$. At most 4 defective means 0, 1, 2, 3, or 4 defective. If we add up the probabilities of these cases, we get

$$e^{-2.5} + 2.5\,e^{-2.5} + \frac{(2.5)^2}{2!}\,e^{-2.5} + \frac{(2.5)^3}{3!}\,e^{-2.5} + \frac{(2.5)^4}{4!}\,e^{-2.5} \approx .891$$

Example (5) During the noon lunch hour, 47 customers will walk through the door of the central post office. Assume that each person arrives at a random time, independent of the other customers. What is the probability that more than one person walks through the door during the first minute?

Solution (5) To see why this is binomial, think of each of the 47 persons choosing at random a minute during which to arrive. The probability that they choose the first minute is then $\frac{1}{60}$. Thus, we have 47 repetitions of a trial, where the probability of success is $\frac{1}{60}$. The expected number of successes is

$$\mu = 47\left(\frac{1}{60}\right) \approx .783$$

To find the probability, we say $P(\text{more than 1 arrival}) = 1 - P(\text{0 or 1 arrival})$.

$$\begin{aligned} P(\text{0 or 1 arrival in 1st minute}) &\approx e^{-.783} + .783\, e^{-.783} \approx .815 \\ P(\text{more than 1 arrival in 1st minute}) &\approx 1 - .815 \\ &\approx .185 \end{aligned}$$