Binomial and Poisson distributions: rationale

The Binomial distribution is used to model the number of successes occuring in n independent Bernoulli trials. The Poisson distribution is used to model a phenomenon that represents the count of events in a fixed time interval (or space unit). The r.v. X represents then the number of successes in this interval.

Suppose that λ denotes the number of independent observations of an event over a given time interval. If we divide this interval into n smaller intervals, then we find $p = \frac{\lambda}{n} \Leftrightarrow \lambda = np$, the probability of observing a success over a time interval n times smaller. Conditions for a Poisson approximation to the Binomial distribution:

$$\begin{array}{ll} (i) & n \text{ is large} \\ (ii) & n * p \text{ is small enough} \end{array}$$

Poisson distribution as an asymptotic case of the Binomial distribution

$$
\begin{aligned}
P(X=x) &= \binom{n}{x} p^x (1-p)^{n-x} \\
&= \binom{n}{x} \left(\frac{\lambda}{n}\right)^x \left(1-\frac{\lambda}{n}\right)^{n-x} \\
&= \binom{n}{x} \frac{\lambda^x}{n^x} \left(1-\frac{\lambda}{n}\right)^n \left(1-\frac{\lambda}{n}\right)^{-x} \\
&= \frac{n(n-1)...(n-x+1)}{x!} \frac{\lambda^x}{n^x} \left(1-\frac{\lambda}{n}\right)^n \left(1-\frac{\lambda}{n}\right)^{-x} \\
&= \frac{n(n-1)...(n-x+1)}{n^x} \frac{\lambda^x}{x!} \left(1-\frac{\lambda}{n}\right)^n \left(1-\frac{\lambda}{n}\right)^{-x} \\
&= \underbrace{1\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)...\left(1-\frac{x-1}{n}\right)}_{\approx 1} \frac{\lambda^x}{x!} \underbrace{\left(1-\frac{\lambda}{n}\right)^n}_{\approx e^{-\lambda}} \left(1-\underbrace{\frac{\lambda}{n}}_{\approx 0}\right)^{-x} \\
&\approx \frac{\lambda^x}{x!} e^{-\lambda} \qquad \text{as } n \to \infty
\end{aligned}
$$

Example of application

A machine tool produces $20,000$ pieces a day. The probability of one piece being defective is 0.1%. The probability that this machine produces exactly 20 defective pieces is given by the Binomial distribution and equals $f_X(20) = P(X = 20) = \binom{20000}{20}\ 0.001^{20}\ 0.999^{20000-20} = 0.0889$.

In the R programming language:

```r
dbinom(x = 20, size = 20000, prob = 0.001)
[1] 0.08887977
```

In the Python language:

```python
from scipy.stats import binom
binom.pmf(20, n = 20000, p = 0.001)
0.08887976802110753
```

Poisson approximation

A machine tool produces $20,000$ pieces a day. The probability of one piece being defective is 0.1%. The probability that this machine produces exactly 20 defective pieces can be approximated by a Poisson distribution. If we let $\lambda = np$, that is, in our example $\lambda = 20000 * 0.001 = 20$, we can thus write $f_X(20) = P(X = 20) = \frac{20^{20}e^{-20}}{20!} = 0.0888$.

In the R programming language:

```
dpois(x = 20, lambda = 20000*0.001)
[1] 0.08883532
```

In the Python language:

```
from scipy.stats import poisson
poisson.pmf(20, mu = 20000*0.001)
0.0888353173920848
```

References

Rizzo, M.L. (2019). Statistical Computing with R, Second Edition (2nd ed.). Chapman and Hall/CRC.
https://doi.org/10.1201/9780429192760

The R Project for Statistical Computing:
https://www.r-project.org/

Python:
https://www.python.org/

course notes