**Parabola Volume 57, Issue 2 (2021)**

# Using mathematics to study the spread of COVID-19

**John Pollard**[1]

There has been a lot of discussion about eliminating or extinguishing the COVID-19 virus, and basic mathematics can be very helpful in understanding whether this is possible. Let us imagine that there is a single case of the virus and that this case has probability $p_0$ of infecting nobody, $p_1$ of infecting 1 person, probability $p_2$ of infecting exactly 2 persons, and so on. We note that the sum of all these probabilities equals 1,

$$\sum_{j=0}^{\infty} p_j = 1\,,$$

and then ask the question:

*What is the probability of complete elimination of the virus?*

For complete elimination of the virus, either the initially infected person must infect no-one else (with probability $p_0$) or, if that person infects exactly one person (with probability $p_1$), then the case sequence from that single case must eventually die out or, if that person produces infects exactly 2 people (with probability $p_2$), then the sequences of cases produced by both of those people must eventually die out, and so on. Let us assume that the probability of complete elimination is $x$. Then the reasoning of the previous paragraph implies that

$$x = p_0 + p_1 x + p_2 x^2 + p_3 x^3 + \cdots\,. \tag{1}$$

The right-hand side of equation (1) is known as the *probability-generating function* of the probability distribution of $x$.

**Remark 1.** *Note that*

*if* $p_0 = 1$*, then the virus is eliminated immediately since only the first person becomes infected;*

*if* $p_0 = 0$*, then complete elimination is not possible since each infected person has zero probability of not infecting further people;*

*if* $p_1 = 1$*, then complete elimination is not possible since each infected person must infect one other person;*

*if* $p_0$ *and* $p_1$ *are the only non-zero probabilities, then complete elimination is certain since each case of infection will on average lead to less than one new case; and*

*if there are* $n \geq 2$ *initial infection cases whose infection sequences subsequently behave completely independently of each other, then the probability of complete elimination is* $xn$*.*

---

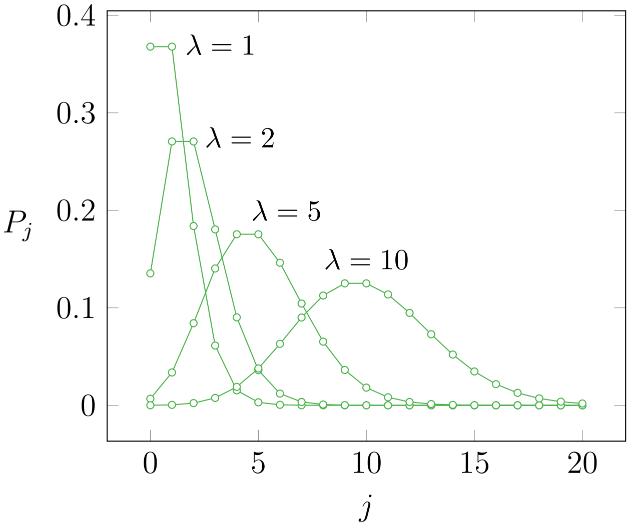
[1] John Pollard is Emeritus Professor and former Professor of Actuarial Studies at Macquarie University.

# Poisson distribution for infections

The *Poisson distributions* is one of the fundamental probability distributions in statistics and finds application in a wide range of areas ranging from blood counts and epidemiology to motor accidents and insurance, and many other areas. According to this discrete distribution, the probability of exactly $j \geq 0$ events happening is as follows:

$$P_j = e^{-\lambda}\frac{\lambda^j}{j!} \tag{2}$$

where $\lambda$ is the mean of the distribution. To get a feel for the Poisson distribution let us look at the numerical values of the probabilities for different values of the mean $\lambda$, given by the graphs of $P_j$ below.



From these graphs, we might intuitively expect the probability of complete elimination to be low when the mean number $\lambda$ of infections produced by a single case is high, and conversely when the mean number $\lambda$ is low.

# Calculating the probability of complete elimination

If the probabilities of the number of infections caused by a single infectious person follows the Poisson distribution, a not an unreasonable assumption, then the probability of the virus being completely eliminated according to equation (1) must be the solution to the equation

$$x = \frac{e^{-\lambda}\lambda^0}{0!}x^0 + \frac{e^{-\lambda}\lambda^1}{1!}x^1 + \frac{e^{-\lambda}\lambda^2}{2!}x^2 + \cdots . \tag{3}$$

The powers of $x$ and $\lambda$ in each term are the same as each other, so from our knowledge of the exponential series, the sum of the infinite series is $e^{-\lambda}e^{\lambda x}$ or, in other words,

$e^{-\lambda(1-x)}$. To determine the probability of complete elimination, we therefore need to solve the equation

$$x = e^{-\lambda(1-x)}\,. \tag{4}$$

This is a non-linear equation, and there are several ways in which it can be solved. The simplest is to use mathematical software such as the *Solver* add-on for *Excel*; other methods are mentioned in Appendix 1 below. Results for various Poisson mean values $\lambda$ are as follows, where $P(CE|1)$ and $P(CE|10)$ denote the probability the virus is eventually completely eliminated, given 1 initial case and 10 initial cases, respectively.

| $\lambda$ | $P_0$ | $P_1$ | $P_2$ | $P(CE\|1)$ | $P(CE\|10)$ |
|---|---|---|---|---|---|
| 1 | 0.368 | 0.368 | 0.264 | 0.9950 | 0.9511 |
| 2 | 0.135 | 0.271 | 0.594 | 0.2032 | $0.1200 \times 10^{-6}$ |
| 3 | 0.050 | 0.149 | 0.801 | $0.5952 \times 10^{-1}$ | $0.5580 \times 10^{-12}$ |
| 4 | 0.018 | 0.073 | 0.908 | $0.1983 \times 10^{-1}$ | |
| 5 | 0.007 | 0.034 | 0.960 | $0.6977 \times 10^{-2}$ | |
| 7 | 0.001 | 0.006 | 0.993 | $0.3364 \times 10^{-3}$ | |
| 10 | $< 0.001$ | $< 0.001$ | $> 0.999$ | $< 0.0001 \times 10^{-4}$ | |

As we argued above intuitively, the probability of complete elimination is higher when the average number of infections per single infectious person is small and conversely when the average number is high. What we were not able to guess intuitively was the speed with which the probability of complete elimination would fall as the mean number of infections per initial single case increased beyond 1 or 2. The drop in complete elimination probability with 10 initial cases (the 10th power of the previous column) is frightening. One can see why our health authorities are concerned to keep locally acquired COVID-19 cases as close to zero as possible!

## Other models

The model we have described is only one possible model for studying the epidemiology of COVID-19. It addresses only one important aspect - extinction. A multi-dimensional version exists which would allow study of mutations. Other models are needed to answer different questions and usually require computer simulation.

# Appendix 1. Other methods to solve non-linear equations

The *Newton-Raphson method* uses the first two terms of the Taylor series expansion of a function, or more simply, by using the value of the function and its slope:

$$f(a+h) \sim f(a) + hf'(a)\,.$$

If the equation we want to solve is $f(x) = 0$, then we choose a suitable starting value $a$ and calculate

$$h = -\frac{f(a)}{f'(a)}$$

which gives us the next approximate solution $a + h$. We then repeat the process with this new starting value and continue iteratively until until convergence to the solution required. Using this approach with

$$f(x) = e^{-\lambda(1-x)} - x$$

and mean $\lambda = 2$, and starting with the trial value $x = 0.15$, the following sequence of values emerges:

| trial value | improved value |
|---|---|
| 0.15 | 0.2015 |
| 0.2015 | 0.2032 |
| 0.2032 | 0.2032 |

Convergence was fast to a value which accords with the probability calculated earlier by *Excel Solver*.

Another approach which sometimes works and might be tried is to use the actual non-linear equation (or a rearranged version of it) and apply it iteratively. This will not always work or may converge slowly. Substituting a starting trial value of 0.15 in the right-hand side of (4) for the case $\lambda = 2$, the following slow sequence to convergence emerges:

| trial value | improved value |
|---|---|
| 0.15 | 0.1827 |
| 0.1827 | 0.1950 |
| 0.1950 | 0.1999 |
| 0.1999 | 0.2019 |
| 0.2019 | 0.2027 |
| 0.2027 | 0.2030 |
| 0.2030 | 0.2031 |
| 0.2031 | 0.2032 |
| 0.2032 | 0.2032 |

# Appendix 2. Historical notes

The process that we have studied above is known as a *branching process* or a *Galton-Watson*. Francis Galton was a 19th century Victorian polymath who is now infamous for inventing and promoting eugenics. He became interested in the extinction of aristocratic family names and, in 1873, posed the following problem [1]:

**Problem 4001**: *A large nation, of whom we will only concern ourselves with adult males, $N$ in number, and who each bear separate surnames colonise a district. Their law of population is such that, in each generation, $a_0$ per cent of the adult males have no male children who reach adult life; $a_1$ have one such male child; $a_2$ have two; and so on up to $a_5$ who have five. Find*

*(1) what proportion of their surnames will have become extinct after $r$ generations;*

*(2) how many instances there will be of the surname being held by $m$ persons.*

As you might notice, the problem bears strong resemblance to the problem of this present article and involves a branching of special cases. Henry William Watson, a priest and prolific mathematician, published a solution [2], and in 1875 he and Galton together published a mathematics research paper [3] on the general problem and its method of solution. It was later discovered that Watson's solution contained an error. In 1977, two Australian mathematicians Heyde and Seneta [5] uncovered that the correct extinction result was known to the French statistician Irénée-Jules Bienaymé in as early as 1845.

# References

[1] F. Galton, Problem 4001, *Educational Times* **25 (143)** (1873), 300.

[2] H.W. Watson, Problem 4001, *Educational Times* **26 (148)** (1873), 115.

[3] F. Galton and H.W. Watson, On the probability of the extinction of families, *Journal of the Royal Anthropological Institute* **4** (1875), 138–144.

[4] T.E. Harris, *The Theory of Branching Processes*, Springer, New York, 1963.

[5] C.C. Heyde and E. Seneta, *I.J. Bienaymé: Statistical Theory Anticipated*, Springer-Verlag, New York, 1977.

[6] D.G. Kendall, Branching processes since 1873, *Journal of the London Mathematical Society* **41** (1966), 385–406.