Article

Computational Simulations of Similar Probabilistic Distributions to the Binomial and Poisson Distributions

Terman Frometa-Castillo [1,*], Anil Pyakuryal [2], Amadeo Wals-Zurita [3], Asghar Mesbahi [4]

[1.] 6134 N Oakley Ave Unit 2, Chicago, IL, 60659, USA

[2.] University of District of Columbia, Division of Science and Mathematics. Email: anil.pyakuryal@udc.edu

[3.] Hospital Universitario Virgen Macarena. Spain. Email: amadeo.wals@gmail.com

[4.] Tabriz University of Medical Sciences, Iran. Email: mesbahiiran@yahoo.com

[*] Correspondence: terman.frometa@gmail.com； Tel: 312-687-6422.

Abstract

This study has developed a Matlab application for simulating statistical models project (SMp) probabilistic distributions that are similar to binomial and Poisson, which were created by mathematical procedures. The simulated distributions are graphically compared with these popular distributions. The application allows to obtain many probabilistic distributions, and shows the trend (τ) for n trials with success probability p, i.e. the maximum probability as $\tau=np$. While the Poisson distribution $PD(x;\mu)$ is a unique probabilistic distribution, where PD=0 in $x=+\infty$, the application simulates many $SMp(x;\mu,Xmax)$ distributions, where μ is the Poisson parameter and value of x with generally the maximum probability, and Xmax is the upper limit of x with $SMp(x;\mu,Xmax) \geq 0$ and limit of the stochastic region of a random discrete variable. It is shown that by simulation via, one can get many and better probabilistic distributions than by mathematical one.

Keywords: simulation; binomial distribution; Poisson distribution; stochastic process; modelling

1. Introduction

Nowadays, the probabilities of some stochastic processes/effects are calculated with experimental/observational data, complex phenomenological/mechanistic models, or analytical expressions, such as the binomial and Poisson, which are results of mathematical derivations. Also, probability can be calculated using computational simulators. This method could become in the most important way of calculation.

The binomial theorem in its most general case, i.e., the binomial series of [1] states that for any positive integer *n*, the nth power of the sum of two numbers *a* and *b* may be expressed as follows

$$(a + b)^n = \binom{n}{0}a^0 b^{n-0} + \binom{n}{1}a^1 b^{n-1} + \cdots + \binom{n}{n}a^n b^{n-n} \tag{1}$$

The binomial distribution (BD) is used to calculate the probability of k successes in n trials with a parameter p (success probability in each trial) of a stochastic process/effect. The BD(k;n,p) expression was obtained with the

substitution in the Eq. (1) of a=p and b=1-p, which let obtaining n+1 BD elements in the right side   and 1 in left side as the sum of all BD elements.

"The Poisson distribution (PD) is a probability distribution of a discrete random variable that stands for the number (count) of statistically independent events, occurring within a unit of time or space" [2].

The PD is the limiting case of the BD, and was determined as follows

$$\lim_{n\to\infty} BD(k;n,p) = PD(k;\lambda) = \frac{e^{-\lambda}\lambda^k}{k!} \tag{2}$$

and used for calculating probabilities of random discrete variable X, P(X=x) as

$$PD(x;\mu) = \frac{e^{-\mu}\mu^x}{x!} \tag{2.1}$$

The BD and PD have been overestimated and considered almost exclusive for calculating their respective probabilities of some stochastic processes/effects (SP/Es), even they have been irrationally used in the ionizing radiation field, such as in the formulation of the tumor control probability (TCP) models and radiation interactions with living tissues.

The statistical models project (SMp) calculates probabilities of many SP/Es by means of its probabilistic-mechanistic models and simulators. These models are based on SMp function of [3]. The radiobiological simulator developed in [4] is a new and interesting way of calculating probabilities.

It was clearly described the mathematical origin and limitations of the popular BD and PD in [3]. Although it was shown that SMp function can play role of BD, this was done only for showing the diversity of SMp. It was well-explained the little or none importance of BD that is more 300 years old of created and widely used in different scientific fields.

In this study, the SMp proposes a simulator based on probabilistic methodology of probabilistic distributions, which are similar to BD and PD, where is shown that is more interesting to generate probabilistic discrete distributions via computational simulations than by mathematical procedures.

2. Material and methods

2.1 The "Binomial" module

For simulating probabilistic distributions that are similar to binomial, the Matlab function rand is used for generating n random positive numbers ≤ 1, which are compared with the binomial parameter p. If generated number is < p there is a success, but a failure. This allows to determine the k successes (kSs). The n generations, i.e. "the n trails" are repeated rptrial time. The probability for each value k, the SMp(k;n,p) simulated is calculated as

$$SMp(k;n,p) = kSs/rptrial \tag{3}$$

The quality of the simulations is checked using the sum of the simulated distributions; i.e. $\sum_0^k SMp(k; n, p)$, which should be around 1 for a good simulation.

2.2 The "Poisson" module

For simulating probabilistic distributions that are similar to Poisson distribution, the Matlab function rand is used for generating Xmax random positive numbers ≤ 1, which are compared with success probability p that is calculated as p=μ/Xmax. If generated number is < p there is a success, but a failure. This lets to determine the k successes (kSs). The Xmax generations, i.e. "the Xmax trails" are repeated rptrial time. The probability for each value x, the SMp(x;μ,Xmax) simulated is determining as

$$SMp(x; μ, Xmax) = kSs/rptrial \qquad (4)$$

where μ is the Poisson parameter and value of x with generally the maximum probability, and Xmax is upper limit of x with SMp(x;μ,Xmax) ≥ 0 and limit of the stochastic region for random discrete variable X.

The quality of the simulations is checked using the sum of the simulated distributions; i.e. $\sum_0^{Xmax} SMp(x; μ, Xmax)$, which should be around 1 for a good simulation.

3. Results

3.1 Description of the codes

The MatLab application SimPD developed in this study is available in https://gitlab.com/tfrometa/simpd. This application simulates the process of k successes of n trials of a SP/E with a success probability p, and obtains SMp(k;n,p) distributions. Also, this obtains the probabilistic distributions for a random discrete variable X with maximum probability in x=μ or x≈μ, and equal to zero in x $\geq$ Xmax, where μ is the Poisson parameter, and Xmax is upper limit of x with probabilty ≥ 0. The probabilistic distributions simulated are graphically compared with the binomial and Poisson distributions.

The SimPD.m, Binomial.m and Poisson.m functions compose the MatLab application. The former is the main code, and its execution allows accessing to two others through their respective buttons, "P(k;n,p) distributions" and "P(x;μ,Xmax) distributions".

3.2 Reproducibility

3.2.1 The "Binomial" module

At the application, the input values (IVs) p and n appear in yellow color, while outcome of the sum of simulated probabilities appears in green. One should press the "Enter" key placed at each field for introducing the IVs into the application.

The steps for the execution of this module are:
    a) Introduce binomial parameters p and n
    b) After the complete introduction of the input values, press the "Simulate" button

3.2.1 The "Poisson" module

The input values (IVs) μ and Xmax appear in yellow color, while outcome of the sum of simulated probabilities appears in green. One should press the "Enter" key placed at each field for introducing the IVs into the application.

The steps for the execution of this module are:
 a) Introduce Poisson parameter μ, and value Xmax.
 b) After the complete introduction of the input values, press the "Simulate" button

4. Discussion

Like is shown by our simulator, there are many possible distributions for k successes in n trials of a SP/E with success probability p; and their trends calculated as np is always the value of k for the maximum probability.

While the binomial distribution (BD(k;n,p)) was derived of a mathematical theorem, the application developed in this study lets obtaining several distributions SMp(k;n,p). The BD(k;n,p) is only one of the possible distributions P(k;n,p). This can be seen when one repeats the simulations for a determined value of p and n. The binomial is result of a power expression (the binomial theorem) whose sum is always 1 of summing p and q=1-p. The most important is that each trial is will be performed with a success probability p and 1-p of failing.

Generally, the SP/Es are characterized with a success probability p as result of previous N experiments or observations with K successes, where p=K/N. For this reason, the trend (τ) for n further trials is τ=n*p. τ is equal to or approximately equal to number of trials with the most likelihood of success trials.

If a SP/E is characterized with a success probability p, it is not necessary to create an additional second probability BD(k;n,p), which was result of a mathematic theorem and satisfies properties of only one of the P(k;n,p) distributions.

The creation of new probabilities depending of p, like BD(k;n,p) with a complex-mathematic expression, have little or none probabilistic importance and has generated confusions, how is shown in [5] where the tumour control probability (TCP) model is irrationally defined as a binomial function.

The Poisson distribution (PD) with parameter μ=4, where PD(4;3)=P(4;4), it means that there is not maximum of probability in x=μ=4. The Figure 1 shows a simulated distribution SMp(x;4,7) and PD(x;4). The former clearly has its maximum in x=4.

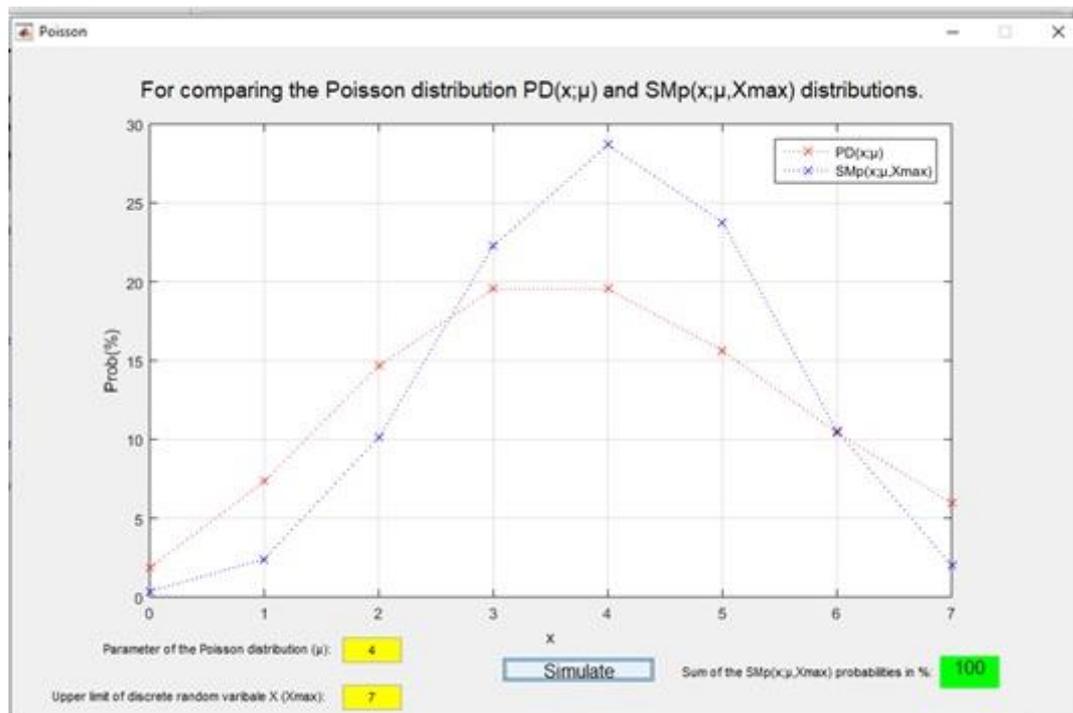Figure 1 Graphical representation of a simulated distribution
SMp(x;4,7) and the Poisson distribution PD(x;4)

According to [6]-[7], the use of Poisson statistics (PS) in TCP models has led to a negative exponential expression. Also, the cell survival (S) has been described with PS in these same references. The TCP exponential expression does not have a strong radiobiological-probabilistic foundation, and the ways of describing S with the PS is probabilistically very complicated. Indeed, S is a complement of the cell kill (K); i.e. probabilistically K=1-S, and K can be modelled with SMp function of [3] as a stochastic effect type SMp P1.

5. Conclusions

The simulation of the binomial distribution (BD) is only a computational exercise, that shows the expected result for n trails of a SP/E with success probability p, the trend is equal to n*p. The BD is only a mathematical exercise without probabilistic importance.

The simulator developed in this work probabilistically generates many distributions that are similar to BD. Additional to each trials is performed with a success probability p and of failure 1-p, the simulations show that trend (τ) for n further trials is τ=n*p.

Our application generates many probabilistic distributions SMp(x;μ,Xmax), and better compared with the PD, which due to its mathematical origin, is unique and with some limitations. Also, the simulator gives possibility of choosing a upper limit value of x, where SMp(x;μ,Xmax) $\geq$ 0; while PD=0 for x=+∞. The Xmax represent the limit of the stochastic region of a random discrete variable.

The computational tool of this study can be used for whatever of the current PD applications. The generated similar-Poisson probabilistic distributions can be used for assuming the description of possible distributions of

the normal tissue effects of the radiation in whatever stochastic activity, such as radiation therapies and other radiation activities.

Nowadays, as result of not considered that BD and PD are discrete probabilistic functions (PFs) , respectively function of k and x, and its mathematical relationship, these PFs have been irrationally used for modelling some SP/Es, like tumor control probability (TCP) that is a continuous PF.

Author Contributions: Conceptualization, Frometa-Castillo and Pyakuryal; Methodology, Wals-Zurita; software, Frometa-Castillo; validation, Pyakuryal, Wals-Zurita and Mesbahi; writing—original draft preparation, Frometa-Castillo; writing—review and editing, Pyakuryal, Wals-Zurita and Mesbahi.

Conflicts of Interest: None

References

1.   Weisstein, E.W.: Binomial Theorem. Wolfram MathWorld (2019) http://mathworld.wolfram.com/BinomialTheorem.html. Accessed 11 June 2019.

2. Letkowski, J.Applications of the Poisson probability distribution, Western New England University (2019). https://www.aabri.com/manuscripts/121246.pdf. Accessed 11 June 2019.

3. Frometa-Castillo, T. (2018). The SMp(x or y;PXmin,Xmax,ML,p1,p2,Max) a probabilistic distribution, or a probability density function of a random variable X, Chapter 48 of 16th International Conference on Information Technology-New Generations (ITNG 2019), Springer Nature Switzerland AG.

4. Frometa-Castillo, T; Pyakuryal, A; Piseaux-Aillon; R. Simulator of radiation biological effects in tumor in order to determinate the tumor control probability; Informatics in Medicine Unlocked; 16; 2019). Available in https://doi.org/10.1016/j.imu.2019.100217

5. Gong, J; Dos Santos, MM; Finlay, C; Haillen, T. Are more complicated tumour control probability models better? Mathematical Medicine and Biology, Volume 30, Issue 1 ,1-19 2011.

6. Chapman, JD; Nahum, AE. Radiotherapy treatment planning Linear-Quadratic Radiobiology. Taylor & Francis Group, LLC; 2015.

7. Dawson, A; Hillen, T. Derivation of the tumour control probability (TCP) from a cell cycle model, Computational and Mathematical Methods in Medicine, 7:2-3, 121-141, DOI: 10.1080/10273660600968937. 2006