Examples and Counterexamples 8 (2025) 100196

Contents lists available at ScienceDirect

Examples and Counterexamples

journal homepage: www.elsevier.com/locate/exco

A catch-all discrete distribution: Short description and main properties

Nuno Fachada [a,c,*], João P. Matos-Carvalho [b,c], Carlos M. Fernandes [a,c]

[a] Lusófona University, Campo Grande, 376, Lisboa, 1749-024, Portugal
[b] LASIGE and Departamento de Informática, Faculdade de Ciências, University of Lisbon, Campo Grande, Lisboa, 1749-016, Portugal
[c] Center of Technology and Systems (UNINOVA-CTS) and Associated Lab of Intelligent Systems (LASI), Caparica, 2829-516, Portugal

ARTICLE INFO

Communicated by F. Nestler

Dataset link: https://zenodo.org/doi/10.5281/zenodo.13357229

Keywords:
Discrete distribution
All successes
Empirical validation
Simulation analysis

ABSTRACT

Consider a population of size N where each element has an independent probability p of being a success. By sampling this population without replacement, how many elements need to be drawn to find all successes? This paper describes this discrete distribution, derives its main properties, and validates the results through simulation.

1. Introduction

A population of N elements is considered, where each element has an independent probability p of being a *success* and a probability $1 - p$ of being a *failure*. Elements are drawn individually without replacement until all successful elements are identified, resulting in a total of k draws. The proposed *catch-all* distribution, denoted as $\mathcal{CA}(N, p)$, represents the discrete distribution of k.

The description of this distribution is motivated by the *Trash Picker* game [1], which itself is inspired by Mitchell's Robby's World [2]. In this game, a robot collects trash on a rectangular grid with N cells, where each cell has a probability p of containing trash. Assuming the robot can observe and pick up trash from any cell without revisiting the same cell (a simplification of the actual game rules), the catch-all distribution models the number of steps required for the robot to collect all the trash. Since the game is turn-based, the expected value of the random variable $K \sim \mathcal{CA}(N, p)$ can, for example, help determine the maximum or optimal number of turns for the game.

The proposed distribution models scenarios where the goal is to sample from a finite population until all individuals with a specific binary characteristic are identified. This includes computational problems such as binary cellular automata and practical applications like disease surveillance—e.g., determining the number of individuals that must be tested in a population until all infected individuals are identified, assuming each individual is infected with probability p—and defect detection—e.g., calculating the number of products that must be inspected in a batch until all defective items are found, given that each product has a defect with probability p. Understanding the properties of this distribution is therefore important for accurately analyzing such processes and supporting decision-making in these and similar practical scenarios.

The remainder of this paper is organized as follows: Section 2 presents the distribution's properties, with an alternative derivation described in Section 3. Section 4 details the distribution's empirical validation, and Section 5 concludes with a brief discussion, potential applications, and open questions.

2. Properties

2.1. Cumulative distribution function (CDF)

The probability that all successes are included within the first k observations is equivalent to the probability that the final $N - k$ observations contain no successes. This probability is given by:

$$F_K(k) = (1 - p)^{N-k}$$

However, this expression is not valid when $k = N$ and $p = 1$ simultaneously. While the expression holds for $p < 1$, the edge case where $k = N$ requires an adjustment, as exactly N draws are needed to capture all N successes. Thus, the expression can be generalized as follows:

$$F_K(k) = \begin{cases} (1 - p)^{N-k} & \text{for } 0 \leq k < N \\ 1 & \text{for } k = N \end{cases} \quad (1)$$

* Corresponding author at: Lusófona University, Campo Grande, 376, Lisboa, 1749-024, Portugal.
E-mail address: nuno.fachada@ulusofona.pt (N. Fachada).

https://doi.org/10.1016/j.exco.2025.100196
Received 25 September 2024; Received in revised form 12 May 2025; Accepted 12 August 2025
Available online 20 August 2025
2666-657X/© 2025 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Expression (1) represents the CDF of the distribution, indicating the probability that the random variable takes a value less than or equal to k.

2.2. Probability mass function (PMF)

The PMF, $f_K(k)$, defines the probability of obtaining exactly k draws to capture all successes. It can be derived from the CDF using finite differencing [3], specifically, $F_K(k) - F_K(k-1)$. Two edge cases must be considered: (1) For $k = 0$, we have $F_K(k-1) = F_K(-1)$, which is assumed to be zero, as the probability of drawing all successes in fewer than zero draws is zero—it is impossible to have a negative number of draws; and, (2) For $k = N$, which is an edge case in the CDF itself. Therefore, the PMF is given by:

$$f_K(k) = \begin{cases} F_K(0) - 0 & \text{for } k = 0 \\ F_K(k) - F_K(k-1) & \text{for } 0 < k < N \\ F_K(N) - F_K(N-1) & \text{for } k = N \end{cases}$$

Substituting the terms, we obtain:

$$f_K(k) = \begin{cases} (1-p)^N - 0 & \text{for } k = 0 \\ (1-p)^{N-k} - (1-p)^{N-(k-1)} & \text{for } 0 < k < N \\ 1 - (1-p)^{N-(N-1)} & \text{for } k = N \end{cases}$$

This leads to the final expression for the PMF:

$$f_K(k) = \begin{cases} (1-p)^N & \text{for } k = 0 \\ p(1-p)^{N-k} & \text{for } 0 < k < N \\ p & \text{for } k = N \end{cases} \tag{2}$$

For $k = 0$, the PMF simplifies to $(1-p)^N$, which is the probability that all elements in the population are non-successes. For $k = N$, the PMF equals p, reflecting the probability that the last drawn element is a success. Additionally, the expression $p(1-p)^{N-k}$ reduces to p when $k = N$ and $p < 1$.

2.3. Hazard rate function (HRF)

The HRF represents the conditional probability that the process ends at step k, given it has persisted until step $k-1$ [4]. For the Catch-All distribution, the HRF denotes the conditional probability that the final success is obtained on the kth draw, given that it has not yet occurred by the $(k-1)$th draw. This characterization is particularly relevant for potential uses cases of the Catch-All distribution. For example, the HRF can inform on the expected effort or stopping rules in processes such as disease testing or defect detection.

As discussed by Barlow and Proschan [4], the HRF for a discrete random variable K can be defined in terms of its PMF and CDF, as:

$$h_K(k) = \frac{f_K(k)}{S_K(k-1)} = \frac{f_K(k)}{1 - F_K(k-1)}$$

where $S_K(k-1) = 1 - F_K(k-1)$ is the survival function at $k-1$, i.e., the probability that all successes have not been caught before k.

Given the piecewise definitions for the PMF and CDF, the HRF must be likewise determined for different values of k. For $k = 0$:

$$h_K(0) = \frac{f_K(0)}{1 - F_K(0-1)} = \frac{(1-p)^N}{1-0} = (1-p)^N \tag{3}$$

For $0 < k < N$, we have:

$$\underset{0<k<N}{h_K(k)} = \frac{f_K(k)}{1 - F_K(k-1)} = \frac{p(1-p)^{N-k}}{1-(1-p)^{N-(k-1)}} = \frac{p(1-p)^{N-k}}{1-(1-p)^{N-k+1}} \tag{4}$$

Finally, for $k = N$:

$$h_K(N) = \frac{f_K(N)}{1 - F_K(N-1)} = \frac{p}{1-(1-p)^{N-(N-1)}} = \frac{p}{1-(1-p)^1} = 1 \tag{5}$$

However, the HRF also yields 1 when setting $k = N$ in (4):

$$\underset{k=N}{h_K(k)} = \frac{p(1-p)^{N-k}}{1-(1-p)^{N-k+1}} = \frac{p(1-p)^0}{1-(1-p)^1} = \frac{p}{p} = 1$$

Thus, considering that (4) and (5) are equivalent for $k = N$, the final expression for the HRF can be obtained by combining (3) and (4):

$$h_K(k) = \begin{cases} (1-p)^N & \text{for } k = 0 \\ \frac{p(1-p)^{N-k}}{1-(1-p)^{N-k+1}} & \text{for } 0 < k \leq N \end{cases} \tag{6}$$

Note that the HRF is undefined at $p = 0$, as the survival function in the denominator, $S_K(k-1)$, is zero—i.e., the probability that all successes have not been caught before k is zero, as there are no successes to catch. Although it is possible to obtain an expression for the HRF at $p = 0$, in particular by taking the limit $p \to 0$ for the general case $(0 < k \leq N)$, such expression does not represent a physically or empirically meaningful quantity.

2.4. Expected value

The expected value (mean) of $K \sim \mathcal{CA}(N, p)$ can be determined using the standard formula for the expected value [3]:

$$E[K] = \sum_{k=1}^{N} k \cdot f_K(K = k)$$

For the edge case $k = N$, the PMF expression $p(1-p)^{N-k}$ is used, since it is valid for $k = N$ when $p < 1$. The case where $p = 1$ will be analyzed separately at the end of this section. Substituting the terms specific to this problem:

$$E[K] = \sum_{k=1}^{N} kp(1-p)^{N-k} \tag{7}$$

$$= p(1-p)^N \sum_{k=1}^{N} k(1-p)^{-k} \tag{8}$$

Defining $r = (1-p)^{-1}$, the summation in (8) becomes:

$$\sum_{k=1}^{N} kr^k$$

To simplify this summation, the sum of the first n terms of the arithmetico-geometric series can be used [3]:

$$\sum_{k=1}^{N} [a + (k-1)d]\, br^{k-1} = \frac{ab - (a + Nd)br^N}{1-r} + \frac{dbr(1-r^N)}{(1-r)^2}$$

Setting $a = d = 1$ and $b = r$ aligns the summations:

$$\sum_{k=1}^{N} kr^k = \frac{r - (1+N)r^{N+1}}{1-r} + \frac{r^2(1-r^N)}{(1-r)^2}$$

Substituting this into the expected value expression, with $r = (1-p)^{-1}$, gives:

$$E[K] = p(1-p)^N \left[\frac{(1-p)^{-1} - (1+N)(1-p)^{-N-1}}{1-(1-p)^{-1}} + \frac{(1-p)^{-2}(1-(1-p)^{-N})}{(1-(1-p)^{-1})^2} \right]$$

Simplifying further:

$$\begin{aligned} E[K] &= \frac{p\left(N - (1-p)^N + 1\right) + (1-p)^N - 1}{p} \\ &= N - (1-p)^N + 1 + \frac{(1-p)^N - 1}{p} \\ &= N + \frac{p - p(1-p)^N + (1-p)^N - 1}{p} \\ &= N + \frac{(1-p)^N(1-p) - (1-p)}{p} \\ &= N + \frac{(1-p)^{N+1} - (1-p)}{p} \end{aligned}$$

For $p = 1$, this expression simplifies to N, which is consistent, as $p = 1$ implies that N successes exist, and $k = N$ draws are thus necessary to capture all successes.

However, $E[K]$ cannot be directly obtained for $p = 0$ using the previous expression. This can be determined by setting $p = 0$ in (7) or (8), or by taking the limit of the previous expression as $p \to 0$, as demonstrated in the `moments` notebook in the supplementary material, yielding $E[K] = 0$ in both cases. Empirically, $p = 0$ implies no successes exist, so all successes are considered caught at $k = 0$, i.e., before any draw occurs. Therefore, the expected value for the catch-all distribution is given by (9):

$$E[K] = \begin{cases} 0 & \text{for } p = 0 \\ N + \frac{(1-p)^{N+1} - (1-p)}{p} & \text{for } 0 < p \leq 1 \end{cases} \tag{9}$$

2.5. Variance

To determine the variance of $K \sim \mathcal{CA}(N, p)$, the general expression for variance can be used [3]:

$$\mathrm{Var}[K] = E[K^2] - E[K]^2 \tag{10}$$

For $p > 0$, this results in:

$$\begin{aligned} \mathrm{Var}[K] &= \sum_{k=1}^{N} k^2 p(1-p)^{N-k} - \left(N + \frac{(1-p)^{N+1} - (1-p)}{p} \right)^2 \\ &= \frac{(p-1)\left(p(2N+1)(1-p)^N + (1-p)^{2N+1} - 1\right)}{p^2} \end{aligned} \tag{11}$$

The complete derivation of (11) is presented in the `moments` notebook in the supplementary material. As with the expected value, (11) is not valid for $p = 0$. However, both terms in (10) are zero when $p = 0$, and taking the limit in (11) as $p \to 0$ also yields zero (as calculated in the `moments` notebook). Therefore, the variance for the catch-all distribution is given by (12):

$$\mathrm{Var}[K] = \begin{cases} 0 & \text{for } p = 0 \\ \frac{(p-1)\left(p(2N+1)(1-p)^N + (1-p)^{2N+1} - 1\right)}{p^2} & \text{for } 0 < p \leq 1 \end{cases} \tag{12}$$

3. Alternative PMF derivation

The catch-all distribution PMF can be derived by combining a negative hypergeometric-like distribution with the binomial distribution. The former assumes a fixed number of successes m, while the latter models m based on the probability of achieving exactly k successes in n independent Bernoulli trials, each with an independent probability p.

The probability of drawing all successes within exactly k draws, with the final draw being the last success, relates to the negative hypergeometric distribution, which counts the number of failures before reaching a specified number of successes. Although the problems are similar, deriving this negative hypergeometric-like distribution from first principles can be simpler.

To generalize the solution for a population of N elements, where m are successes and the remaining $N - m$ are non-successes, the goal is to determine the probability that all m successes are drawn in exactly k draws, with the kth draw being the final success. Assuming the actual number of successes, M, is a constant m:

$$\begin{aligned} &\Pr(m \text{ successes in } k \text{ draws} \mid M = m) = \\ &\quad \Pr(m-1 \text{ successes in } \leq k-1 \text{ draws}) \cdot \Pr(k\text{th draw is a success}) \end{aligned}$$

To calculate $\Pr(m-1$ successes in $\leq k-1$ draws), i.e., the probability of drawing exactly $m-1$ successes in the first $k-1$ draws, the following formula is used, assuming $k \geq 1$:

$$\Pr(m-1 \text{ successes in } \leq k-1 \text{ draws}) = \frac{m \cdot \binom{N-m}{k-m}}{\binom{N}{k-1}}$$

where:

- $\binom{m}{m-1} = \binom{m}{1} = m$: Number of ways to choose $m-1$ successes from m;
- $\binom{N-m}{k-m}$: Number of ways to choose $k-m$ failures from $N-m$;
- $\binom{N}{k-1}$: Number of ways to choose $k-1$ elements from N, as the kth draw is not yet included, assuming $k \geq 1$.

After drawing $m-1$ successes and $k-m$ failures, there is exactly one success left among the $N-k+1$ remaining elements. Thus, the probability that the kth draw is a success is:

$$\Pr(k\text{th draw is a success}) = \frac{1}{N-k+1}$$

The overall probability is then determined as:

$$\Pr(m \text{ successes in } k \text{ draws} \mid M = m) = \frac{m\binom{N-m}{k-m}}{(N-k+1)\binom{N}{k-1}}$$

This expression can be simplified as follows:

$$\begin{aligned} \frac{m\binom{N-m}{k-m}}{(N-k+1)\binom{N}{k-1}} &= \frac{m\frac{(N-m)!}{(k-m)!(N-m-(k-m))!}}{\frac{N!(N-k+1)}{(k-1)!(N-(k-1))!}} \\ &= \frac{m(N-m)!(k-1)!}{N!(k-m)!} \end{aligned}$$

Thus, the final expression for the probability is:

$$\Pr(m \text{ successes in } k \text{ draws} \mid M = m) = \frac{m(N-m)!(k-1)!}{N!(k-m)!} \tag{13}$$

Expression (13) represents the PMF for the event where all successes are drawn in exactly k draws, with the last draw being a success. This calculation assumes $m \leq k \leq N$, meaning at least m draws are required to obtain all successes, and no more elements can be drawn than the total number of elements.

However, this distribution assumes a fixed m, while in the catch-all distribution, m is a random variable, with each element having an independent probability p of being a success when drawn. Thus, m follows a binomial distribution, where each draw is a Bernoulli trial. In this case, m is replaced by $M \sim \text{Binomial}(N, p)$, and the probability of exactly m successes is:

$$\Pr(M = m) = \binom{N}{m} p^m (1-p)^{N-m} = \frac{N!}{m!(N-m)!} \cdot p^m (1-p)^{N-m}$$

Combining these using the law of total probability [3]:

$$\begin{aligned} &\Pr(m \text{ successes in } k \text{ draws}) = \\ &\quad \sum_{m=0}^{N} \Pr(m \text{ successes in } k \text{ draws} \mid M = m) \times \Pr(M = m) \end{aligned}$$

Considering that the expression inside the sum is zero when $m = 0$ and that $m \leq k$, the overall expression expands to:

$$\begin{aligned} &\Pr(m \text{ successes in } k \text{ draws}) = \Pr(m \text{ in } k) = \\ &\quad \sum_{m=1}^{k} \left(\frac{m(N-m)!(k-1)!}{N!(k-m)!} \right) \cdot \frac{N!}{m!(N-m)!} \cdot p^m (1-p)^{N-m} \end{aligned}$$

This can be simplified to the PMF presented earlier:

$$\begin{aligned} \Pr(m \text{ in } k) &= \sum_{m=1}^{k} \left(\frac{(k-1)!}{(m-1)!(k-m)!} \right) \cdot p^m (1-p)^{N-m} \\ &= \sum_{m=1}^{k} \binom{k-1}{m-1} p^m (1-p)^{N-m} \\ &= \sum_{m'=0}^{k-1} \binom{k-1}{m'} p^{m'+1} (1-p)^{N-(m'+1)} \quad \text{(assuming } m' = m-1) \\ &= p(1-p)^{N-1} \sum_{m'=0}^{k-1} \binom{k-1}{m'} \left(\frac{p}{1-p} \right)^{m'} \end{aligned}$$

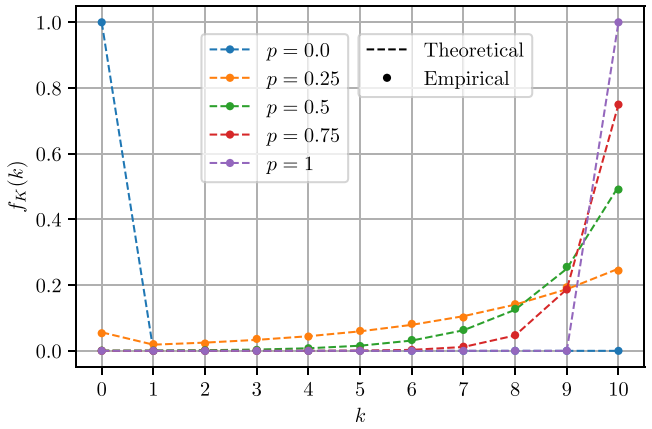

Fig. 1. Theoretical (dashed line) and empirical (circle markers) catch-all PMF for $N = 10$ and different values of p (depicted in different colors). Theoretical values were calculated with (2), while empirical results were obtained via simulation.

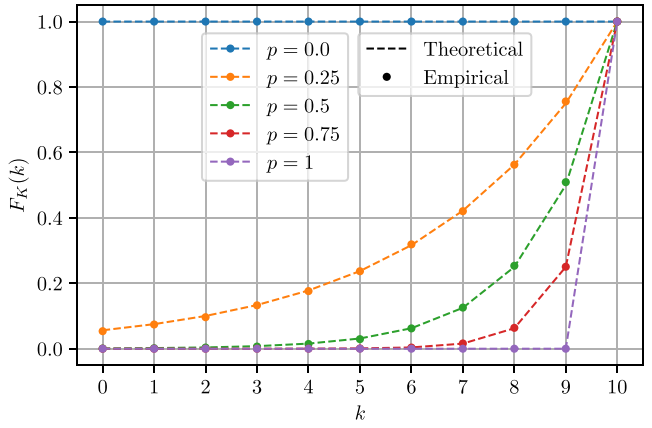

Fig. 2. Theoretical (dashed line) and empirical (circle markers) catch-all CDF for $N = 10$ and different values of p (depicted in different colors). Theoretical values were calculated with (1), while empirical results were obtained via simulation.

Using the binomial expansion [3], $(1+x)^n = \sum_{i=0}^{n} \binom{n}{i} x^i$, with $n = k-1$, $i = m'$, and $x = p/(1-p)$, we have:

$$\begin{aligned}\Pr(m \text{ in } k) &= p(1-p)^{N-1}\left(1+\frac{p}{1-p}\right)^{k-1}\\ &= p(1-p)^{N-1}\left(\frac{1}{1-p}\right)^{k-1}\\ &= p(1-p)^{N-1}(1-p)^{1-k}\\ &= p(1-p)^{N-k}\end{aligned}$$

which matches the PMF in (2) for $0 < k < N$.

4. Empirical validation

The catch-all distribution was empirically validated through simulation, where elements of a population were drawn until all successes were found. The complete simulation is available in the `simulation` notebook, included in the supplementary material. This section summarizes the main results.

The first step in empirical validation involved running 10 000 simulation trials for various combinations of p and N. From these trials, empirical values for the PMF, CDF, HRF, expected value, and variance were calculated. Selected combinations are presented in Figs. 1–5, along with the corresponding theoretical values. As can be observed, the empirical and theoretical results are virtually identical.

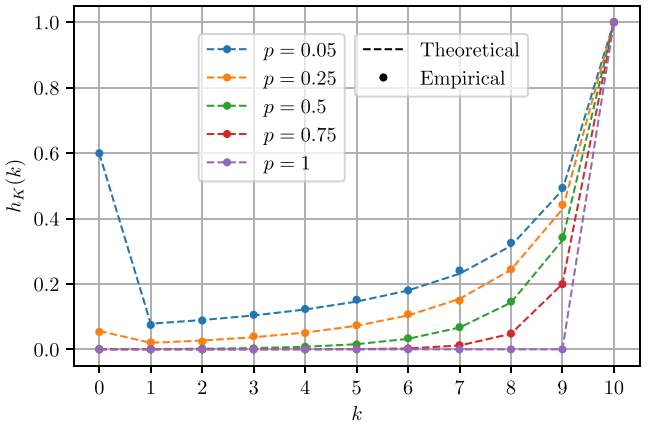

Fig. 3. Theoretical (dashed line) and empirical (circle markers) catch-all HRF for $N = 10$ and different values of p (depicted in different colors). Theoretical values were calculated with (6), while empirical results were obtained via simulation.

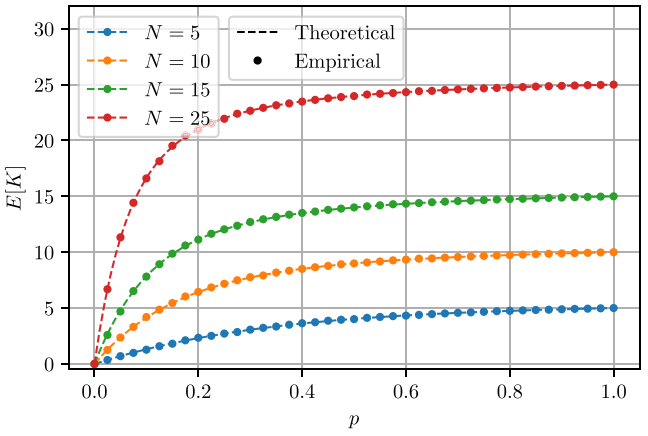

Fig. 4. Theoretical (dashed line) and empirical (circle markers) expected values for the catch-all distribution with different values of N (shown in different colors) and p (along the x-axis). Theoretical values were calculated with (9), while empirical results were obtained via simulation.

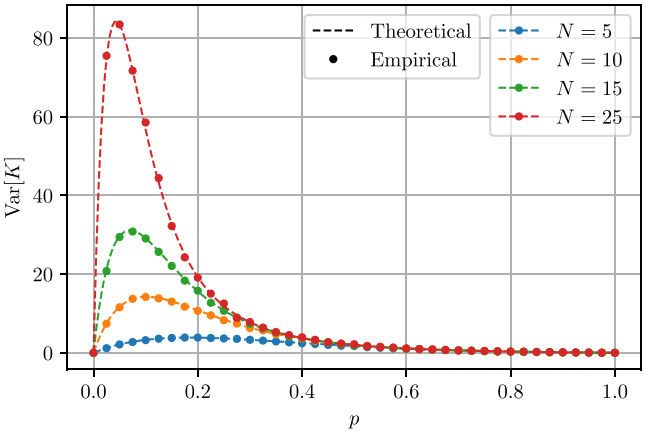

Fig. 5. Theoretical (dashed line) and empirical (circle markers) variances for the catch-all distribution with different values of N (shown in different colors) and p (along the x-axis). Theoretical values were calculated with (12), while empirical results were obtained via simulation.

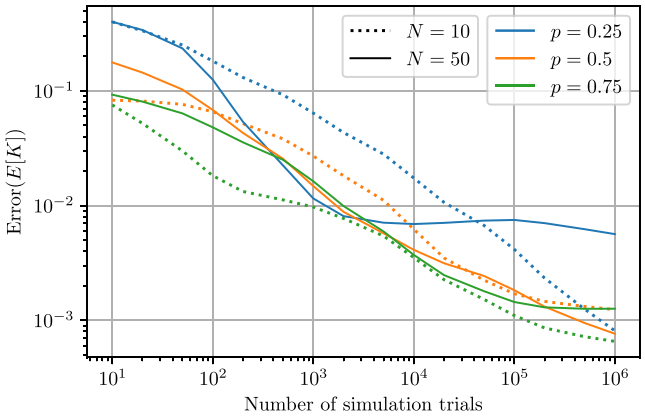

Fig. 6. Error between theoretical and empirical expected values for various values of N and p, obtained with expression (9) and simulation, respectively. Gaussian filtering with $\sigma = 2$ was applied in the log domain to highlight the decreasing error tendency.

In the second validation step, simulations with an increasing number of trials were conducted for various values of N and p, and the resulting empirical expected value was compared to its theoretical counterpart. According to the law of large numbers, as the number of simulation trials increases, the empirical results should converge to the theoretical predictions, resulting in a decreasing error between them. This expectation is confirmed by the results depicted in Fig. 6, which illustrate a general trend of decreasing error with an increasing number of trials. Since the error decrease is not strictly monotonic, one-dimensional Gaussian filtering (with $\sigma = 2$) was applied in the log domain to smooth out non-monotonic variations, thereby emphasizing the overall decreasing trend.

5. Conclusions

Theoretical results for the proposed catch-all distribution were derived and validated through simulation, with empirical observations closely matching the analytical expressions. This confirms the suitability of the distribution to model scenarios involving sampling without replacement until all successes are identified. Potential applications include areas such as disease surveillance, defect detection in manufacturing, and related processes. Future work may explore additional properties, including higher-order moments, skewness, and kurtosis, as well as extend the study to further application domains. Investigating its practical deployment and performance with real-world data also remains an open and relevant direction.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was partially funded by: Fundação para a Ciência e a Tecnologia, Portugal (FCT, https://ror.org/00snfqn58) under Grants Copelabs ref. UIDB/04111/2020, Centro de Tecnologias e Sistemas (CTS) ref. UIDB/00066/2020, LASIGE Research Unit ref. UIDB/00408/2025, and COFAC ref. CEECINST/00002/2021/CP2788/CT0001; Instituto Lusófono de Investigação e Desenvolvimento (ILIND) under Project COFAC/ILIND/COPELABS/1/2024; and, Ministerio de Ciencia, Innovación *y* Universidades (MICIU/AEI/10.13039/501100011033, https://ror.org/05r0vyz12) under Project PID2023-147409NB-C21.

Data availability

All supplementary material, including Jupyter notebooks containing symbolic derivations, simulations, and additional plots, is available at https://zenodo.org/doi/10.5281/zenodo.13357229. These notebooks were executed using Python 3.12 and make use of standard scientific libraries specified in the accompanying `requirements.txt` file.

References

[1] J. Inácio, N. Fachada, J.P. Matos-Carvalho, C.M. Fernandes, Humans vs ai: An exploratory study with online and offline learners, in: Videogame Sciences and Arts, Springer Nature Switzerland, 2024, pp. 272–286, http://dx.doi.org/10.1007/978-3-031-51452-4_19.
[2] M. Mitchell, Complexity: A Guided Tour, Oxford University Press, 2009.
[3] K.F. Riley, M.P. Hobson, S.J. Bence, Mathematical Methods for Physics and Engineering, third ed., Cambridge University Press, 2006.
[4] R.E. Barlow, F. Proschan, Statistical theory of reliability and life testing, 1981, To Begin With.