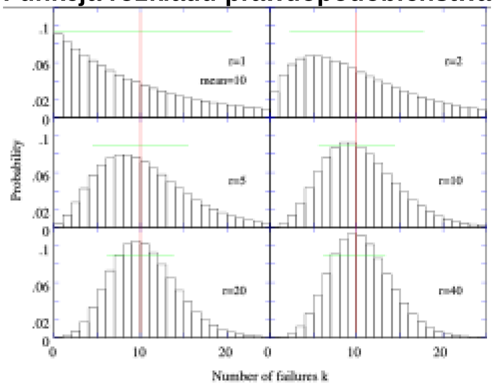
Rozkład Pascala

Rozkład Pascala (ujemny rozkład dwumianowy) – dyskretny rozkład prawdopodobieństwa opisujący m.in. liczbę sukcesów i porażek w niezależnych i posiadających równe prawdopodobieństwo sukcesu próbach Bernoulliego. Jest uogólnieniem rozkładu geometrycznego dla wielu prób.

Termin „ujemny rozkład dwumianowy” nie jest w pełni usystematyzowany. Może dotyczyć jednego z kilku wariantów funkcji opisujących te same zmienne losowe z subtelnymi różnicami w parametryzacji – liczby prób, albo sukcesów lub porażek (czasem liczonych bez ostatniego), przy określonej wartości jednej z tych zmiennych. Momenty i inne charakterystyki poszczególnych wersji rozkładu różnią o proste transformacje[1][2][3]. Nazwa „rozkład Pascala” opisuje z reguły warianty dla wartości całkowitych, liczonych bez ostatniego zdarzenia[3].

Rozkład Pascala

Funkcja rozkładu prawdopodobieństwa



Czerwona linia oznacza wartość oczekiwaną, a zielona ma w przybliżeniu długość 2σ.

Parametry	$r > 0$ (liczba rzeczywista) $0 < p < 1$ (liczba rzeczywista) Poniższe wzory dotyczą wariantu opisującego liczbę sukcesów k przed porażką r. Inne parametryzacje opisują inne wzory.
Nośnik	$k \in \{0, 1, 2, \ldots\}$
Funkcja rozkładu prawdopodobieństwa	$\frac{\Gamma(r+k)}{k!\,\Gamma(r)}\, p^k\, (1-p)^r$
Dystrybuanta	$I_p(r, k+1)$ gdzie $I_p(x, y)$ jest regularyzowaną niekompletną funkcją Beta
Wartość oczekiwana (średnia)	$\frac{rp}{(1-p)}$
Moda	$\lfloor (r-1)\,(1-p)/p \rfloor \iff r > 1$ $0 \iff r \leqslant 1$
Wariancja	$r\,\frac{p}{(1-p)^2}$
Współczynnik skośności	$\frac{1+p}{\sqrt{r\,p}}$
Kurtoza	$\frac{6}{r} + \frac{(1-p)^2}{pr}$
Entropia	$\frac{1}{2}\ln\left(\frac{2\pi erp}{(1-p)^2}\right) + O\left(\frac{1}{r}\right)$

Spis treści

Wariant dla liczby sukcesów przed r porażką

Rozważmy ciąg $X_1, X_2, \ldots$ niezależnych prób Bernoulliego z prawdopodobieństwem sukcesu równym p. Ustalmy liczbę r. Obserwujemy ten ciąg do momentu stwierdzenia r-tej porażki. Oznaczmy ten moment przez T. O zmiennej losowej $T - r$ mówimy, że ma ujemny rozkład dwumianowy NB(r,p) z parametrami r oraz p.

Funkcja tworząca momenty	$\left(\frac{1-p}{1-pe^t}\right)^r$ przy $t < -\log p$
Funkcja charakterystyczna	$\left(\frac{1-p}{1-pe^{i\,t}}\right)^r$ with $t \in \mathbb{R}$

Niech X ma rozkład NB(r,p). Wtedy $X = k$ (gdzie $k = 0, 1, 2, \ldots$) jeśli w $r + k$-tym momencie zaszła porażka oraz w ciągu $X_1, \ldots, X_{r+k-1}$ zaszło $r - 1$ porażek. Zatem

$$P(X = k) = \binom{r+k-1}{r-1}(1-p)^{r-1}p^{(r+k-1)-(r-1)}(1-p),$$

czyli

$$P(X = k) = \binom{r+k-1}{r-1}(1-p)^r p^k.$$

Na rozkład ten można spojrzeć w następujący sposób: rozważamy ciąg niezależnych zmiennych $Y_1, \ldots, Y_r$ o rozkładzie geometrycznym z parametrem sukcesu $1 - p$ odpowiadające obserwacji naszego ciągu po porażce $r - 1$ do porażki r włącznie. Niech $Y = Y_1 + \ldots + Y_r$. Wtedy zmienna losowa $X = Y - r$, zliczająca jedynie liczbę sukcesów, ma rozkład ujemny dwumianowy z parametrami r oraz p. Z tego otrzymujemy natychmiast wzór na wartość oczekiwaną zmiennej losowej o tym rozkładzie

$$E(X) = r \cdot \frac{1}{1-p} - r = \frac{rp}{1-p}.$$

W podobny sposób można wyprowadzić wzór na wariancję.

Dla porównania, w trochę innej definicji ujemnego rozkładu dwumianowego, porażkę zastępuje się sukcesem oraz nie odejmuje się parametru r od momentu zajścia r-tego sukcesu. Otrzymujemy wtedy zmienną losową X o następujący rozkładzie

$$P(X = k) = \binom{k-1}{r-1}p^r(1-p)^{k-r}, \quad k \geqslant r.$$

Zmienna ta jest sumą r niezależnych zmiennych o rozkładzie geometrycznym z parametrem sukcesu p.

Inne warianty

Rozkład był prezentowany w literaturze na kilka różnych sposobów, z subtelnymi zmianami parametryzacji[1][2][3]. Różnice w notacji dotyczą m.in. stosowania równoważności pomiędzy liczbą prób n, sukcesów k i porażek r, np. $n = k + r$, tego, czy nośnik zaczyna się od 0 czy 1, oraz z możliwości przedstawienia wzoru z użyciem różnych form symbolu Newtona, także z wykorzystaniem tożsamości kombinacji dopełniających:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \binom{n}{n-k}.$$

Poniższa tabela przedstawia niektóre spotykane formy rozkładu.

X zlicza:	Nośnik i funkcja rozkładu prawdopodobieństwa	Wzór
k sukcesów, przy danych r porażkach	dla $k \in \{0, 1, 2, \ldots\}$ $f(k; r, p) \equiv \Pr(X = k) =$	$\binom{k+r-1}{k} p^k (1-p)^r$ [4][5]
		$\binom{k+r-1}{r-1} p^k (1-p)^r$ (wariant opisany powyżej)
		$\binom{n-1}{k} p^k (1-p)^r$
n prób, przy danych r porażkach	dla $n \in \{r, r+1, r+2, \ldots\}$ $f(n; r, p) \equiv \Pr(X = n) =$	$\binom{n-1}{r-1} p^{n-r} (1-p)^r$
		$\binom{n-1}{n-r} p^{n-r} (1-p)^r$
r porażek, przy danych k sukcesach	dla $r \in \{0, 1, 2, \ldots\}$ $f(r; k, p) \equiv \Pr(X = r) =$	$\binom{k+r-1}{r} p^k (1-p)^r$ [2][6]
		$\binom{k+r-1}{k-1} p^k (1-p)^r$ [7][8][9][10]
		$\binom{n-1}{r} p^k (1-p)^r$
n prób, przy danych k sukcesach	dla $n \in \{k, k+1, k+2, \ldots\}$ $f(n; k, p) \equiv \Pr(X = n) =$	$\binom{n-1}{k-1} p^k (1-p)^{n-k}$ [2][10][11][12][13][14]
		$\binom{n-1}{n-k} p^k (1-p)^{n-k}$
k sukcesów, przy danych n próbach (rozkład dwumianowy – dla porównania)	dla $k \in \{0, 1, 2, \ldots, n\}$ $f(k; n, p) \equiv \Pr(X = k) =$	$\binom{n}{k} p^k (1-p)^r$

Wzór można także rozszerzyć dla niecałkowitych wartości r z użyciem funkcji gamma, np.:

$$\begin{aligned} f(r, k, p) &= \frac{(k+r-1)!}{k!\,(r-1)!}\, p^r\, (1-p)^k \\ &= \binom{k+r-1}{r-1}\, p^r\, (1-p)^k \\ &= \frac{\Gamma(r+k)}{k!\,\Gamma(r)}\, p^r\, (1-p)^k \end{aligned}$$

opisuje, jakie jest prawdopodobieństwo, że czas oczekiwania na r-ty sukces będzie wynosił $k + r$.

Przypisy

1. Gavin J.S. Ross, Donald Arthur Preece, *The Negative Binomial Distribution*, „The Statistician”, 34 (3), 1985, s. 323, DOI: 10.2307/2987659 (https://dx.doi.org/10.2307%2F2987659), JSTOR: 2987659 (http://www.jstor.org/stable/2987659).

2. John D. Cook, *Notes on the Negative Binomial Distribution* (http://www.johndcook.com/negative_binomial.pdf).
3. Samuel Kotz, Adrienne W. Kemp, Norman Lloyd Johnson, *Univariate discrete distributions*, wyd. 2, New York: Wiley, 1992, s. 199–213, ISBN 0-471-54897-9, OCLC 25547480 (http://worldcat.org/oclc/25547480).
4. Morris H. DeGroot, Mark J. Schervish, *Probability and statistics*, wyd. 4, Boston: Addison-Wesley, 2012, s. 297, ISBN 978-0-321-50046-5, OCLC 502674206 (http://worldcat.org/oclc/502674206).
5. William H. Beyer, *CRC standard mathematical tables*, wyd. 28, Boca Raton, Florida: CRC Press, 1987, s. 533, ISBN 0-8493-0628-0, OCLC 16167842 (http://worldcat.org/oclc/16167842).
6. *Mathworks: Negative Binomial Distribution* (http://www.mathworks.com/help/stats/negative-binomial-distribution.html).
7. Eric W. Weisstein, *Negative Binomial Distribution* (http://mathworld.wolfram.com/NegativeBinomialDistribution.html), mathworld.wolfram.com [dostęp 2019-06-17] (ang.).
8. SAS Institute, „Negative Binomial Distribution” (http://support.sas.com/documentation/cdl/en/lefunctionsref/67960/HTML/default/viewer.htm#n0n7cce4a3gfqkn1vr0p1x0of99s.htm#n1olto4j49wrc3n11tnmuq2zibj6), *SAS(R) 9.4 Functions and CALL Routines: Reference, Fourth Edition*, SAS Institute, Cary, NC, 2016.
9. Michael J. Crawley, *The R Book* (https://books.google.com/books?id=XYDl0mlH-moC), Wiley, 2012, ISBN 978-1-118-44896-0.
10. *Set theory: Section 3.2.5 – Negative Binomial Distribution* (http://www.math.ntu.edu.tw/~hchen/teaching/StatInference/notes/lecture16.pdf).
11. *Randomservices.org, Chapter 10: Bernoulli Trials, Section 4: The Negative Binomial Distribution* (http://www.randomservices.org/random/bernoulli/NegativeBinomial.html).
12. *Stat Trek: Negative Binomial Distribution* (http://stattrek.com/probability-distributions/negative-binomial.aspx).
13. Jacqueline Wroughton, *Distinguishing Between Binomial, Hypergeometric and Negative Binomial Distributions* (http://www.stat.purdue.edu/~zhanghao/STAT511/handout/Stt511%20Sec3.5.pdf).
14. Sheldon M. Ross, *A first course in probability*, wyd. 8, Upper Saddle River, N.J.: Pearson Prentice Hall, 2010, s. 157, ISBN 978-0-13-603313-4, OCLC 237199460 (http://worldcat.org/oclc/237199460).

Bibliografia

- William Feller: *Wstęp do rachunku prawdopodobieństwa*. Warszawa: PWN, 2007, s. 159–160. ISBN 978-83-01-14684-9.