A brief introduction to the Pascal distribution

Radoslav Harman

May 8, 2022

1 Introduction

I did not use any particular book or paper to prepare this classroom material; everything here is simple to derive from the basic theorems of combinatorics, calculus and probability. However, it is possible to find much published information on the Pascal distribution;[1] see, for instance, the book by Balakrishnan and Nevzorov (ISBN: 978-0-471-42798-8) for an introduction and the book by Johnson, Kemp and Kotz (ISBN: 978-0-471-27246-5) for details.

2 Definition

We perform a sequence of independent trials. Each trial results in an outcome $\mathcal{P}$ (called success) or an outcome $\mathcal{Q}$ (called failure). The probability that an individual trial ends up as $\mathcal{P}$ is denoted by p.[2] The probability that an individual trial ends up as $\mathcal{Q}$ is denoted by $q := 1 - p$. We know that if we fix the number of trials to be n,[3] the number of outcomes $\mathcal{P}$ is a $Bin(n, p)$ random variable. However, here we will focus on an experiment organized in a different manner:

We will repeat the independent trials until we get the n-th outcome $\mathcal{Q}$, that is, we will repeat the trials until the n-th failure. Let X be the number of outcomes $\mathcal{P}$, i.e., the number of successes. The distribution of X is then called the Pascal distribution.[4] This will be denoted $X \sim Pas(n, p)$.

The interpretation above and some basic combinatorics yield:

---

[1] Or the negative binomial distribution, which is *usually* defined as a generalization of the Pascal distribution.

[2] We will only assume that $p \in (0, 1)$, and avoid the pathological cases of $p = 0$ and $p = 1$.

[3] We will assume that $n \in \mathbb{N}$, i.e., $n \neq 0$.

[4] The definition of the Pascal distribution is annoyingly inconsistent throughout the literature. We will not discuss the plethora of definitions, but let us mention that the Pascal distribution, as defined here, is sometimes called the negative binomial distribution. Usually, however, the negative binomial distribution is a generalization of the Pascal distribution, where n is allowed to be non-integer. Note that here we adopt the logic always used for the binomial distribution; that is, the Pascal random variable counts the number of "successes" (not "failures" nor all trials), and the probability of each individual success is denoted by p.

Theorem 1. $X \sim Pas(n,p)$ if and only if

$$P[X = k] = \binom{n+k-1}{n-1} p^k q^n \text{ for } k = 0, 1, 2, \ldots\ . \tag{1}$$

The most important special case of the Pascal distribution is the geometric distribution:[5]

Theorem 2. $Pas(1,p)$ is the geometric distribution.[6]

The support of $X \sim Pas(n,p)$ is the entire $\mathbb{N}_0$, similarly to the distribution $Poi(\lambda)$ for $\lambda > 0$. In fact, as we will see, the class of Pascal distributions contains the Poisson distribution as a limit case. However, the two-parametric class of Pascal distributions is more flexible.

3 Generating functions

The form of the probability generating function $P_X = E(s^X)$ (and the moment generating function $M_X(t) = P_X(e^t)$) of $X \sim Pas(n,p)$ is easy to derive from the following Taylor series, which is valid for all $x \in (-1,1)$:[7]

$$\frac{1}{(1-x)^n} = \sum_{k=0}^{\infty} \binom{k+n-1}{n-1} x^k.$$

Theorem 3. Let $X \sim Pas(n,p)$. Then

$$P_X(s) = \left(\frac{q}{1-ps}\right)^n \text{ for } s \in (-1/p, 1/p), \tag{2}$$

$$M_X(t) = \left(\frac{q}{1-pe^t}\right)^n \text{ for } t \in (\ln(p), -\ln(p)). \tag{3}$$

In particular, the probability generating function of the geometric distribution $Pas(1,p)$ is $q/(1-ps)$ for $s \in (-1,1)$. Consequently, the sum X of iid random variables $X_1, \ldots, X_n \sim Pas(1,p)$ has the generating function

$$P_X(s) = \prod_{i=1}^{n} P_{X_i}(s) = \left(\frac{q}{1-ps}\right)^n$$

implying[8]

[5] Therefore, this text also covers the geometric distribution. For instance, we will obtain the generating functions and the higher-order moments of the geometric distribution as a special case of the Pascal distribution. However, some interesting properties of the geometric distribution, such as its "memorylessness", do not generalize to the Pascal distribution.

[6] With parameter p or with parameter q, depending on the particular definition of the geometric distribution.

[7] The special case for $n = 1$ is the famous $1/(1-x) = 1 + x + x^2 + x^3 + \ldots$ (if $|x| < 1$).

[8] Of course, we can also prove the following theorem from the interpretation of $Pas(n,p)$ or using elementary combinatorics; we do not need the probability generating function for the proof.

Theorem 4. Let $X_1, \dots, X_n$ be iid random variables, each distributed according to the geometric distribution $Pas(1, p)$. Then $X = X_1 + \cdots + X_n \sim Pas(n, p)$.

The previous theorem can be further generalized to

Theorem 5. Let $X_1, \dots, X_n$ be iid random variables, $X_i \sim Pas(n_i, p)$, $i = 1, \dots, n$. Then $X = X_1 + \cdots + X_n \sim Pas(\sum_{i=1}^{n} n_i, p)$.

Theorem 5 can be easily proved using the probability generating function.

4 Moments

Theorem 4 is an essential characterization of the Pascal distribution. It can be used, for instance, to derive the mean and the variance of $Pas(n, p)$ from the moments of the geometric distribution.[9]

Theorem 6. Let $X \sim Pas(n, p)$. Then $\alpha_1 = EX = np/q$ and $\beta_2 = DX = np/q^2$.

In particular $EX < DX$ for any n and p; recall that for the Poisson distribution the mean and the variance coincide.

The higher-order moments α_k (absolute) and β_k (central), $k \in \mathbb{N}$, are complex rational functions in p. However, as usual with the fundamental discrete distributions, the general formula for the *factorial* moments is simple:

Theorem 7. Let $X \sim Pas(n, p)$. Then

$$\mu_k = E(X(X-1)\cdots(X-k+1)) = \frac{(n-1+k)!}{(n-1)!}\left(\frac{p}{q}\right)^k.$$

This theorem is straightforward to prove by means of differentiating P_X and setting $s = 1$. It immediately provides Theorem 6, but it can also be used to obtain higher-order moments via the general formulas

$$\alpha_k = E(X^k) = \sum_{j=0}^{k} \left\{ {k \atop j} \right\} \mu_j, \text{ where } \left\{ {k \atop j} \right\} = \frac{1}{j!}\sum_{i=0}^{j}(-1)^i \binom{j}{i}(j-i)^k$$

are the Stirling numbers of the second kind, and

$$\beta_k = E((X - EX)^k) = \sum_{j=0}^{k}(-1)^j \binom{k}{j}(EX)^j \alpha_{k-j}.$$

After some tedious but straightforward calculations based on the formulas above, we can derive

[9] Which should be well known at this point of your studies. Note that you also need the fact that the mean value of the sum of random variables is the sum of the mean values of the random variables. For variance we can use an analogous theorem valid for *independent* random variables.

Theorem 8. The skewness and kurtosis of $X \sim Pas(n, p)$ are

$$\gamma_1 = \frac{1+p}{\sqrt{np}}, \text{ and } \gamma_2 = 3 + \frac{1+4p+p^2}{np}.$$

We see that the Pascal distribution is always positively skewed; however, the skewness tends to 0 with $n \to \infty$. The kurtosis is always larger than 3 (that is, the distribution is "leptokurtic") but converges to 3 with $n \to \infty$. These convergence properties are clear from the relation of the Pascal distribution with the normal distribution, detailed in the next section.

5 Limit cases

Since the distribution of $X \sim Pas(n, p)$ is equal to the one of the sum of n independent geometric random variables, the central limit theorem implies that for $n \to \infty$ the standardized $Pas(n, p)$ converges to $N(0, 1)$:

Theorem 9. Let $X_n \sim Pas(n, p)$ and let $x \in \mathbb{R}$. Then

$$\lim_{n\to\infty} P\left[\frac{X_n - EX_n}{\sqrt{DX_n}} < x\right] = \Phi(x), \tag{4}$$

where Φ is the distribution function of $N(0, 1)$.

Of course, in (4) we could replace EX_n with np/q and DX_n with np/q^2. A less formal statement of Theorem 9 is that for a large n the distribution $Pas(n, p)$ is approximately $N(np/q, np/q^2)$. Note also the similarity of Theorem 9 with the De Moivre-Laplace theorem on the binomial distribution.

A different limit theorem can be obtained by letting $n \to \infty$ but simultaneously decreasing the value of p, such that the mean value np/q converges to some real $\lambda > 0$. One possible version of the theorem is

Theorem 10. Let $X_n \sim Pas(n, \lambda/n)$, $\lambda > 0$, $n = n_0, n_0 + 1, \ldots$, where n_0 is chosen such that $\lambda/n_0 < 1$. Then for any $k \in \mathbb{N}_0$

$$\lim_{n\to\infty} P[X_n = k] = e^{-\lambda}\frac{\lambda^k}{k!}.$$

This can be proved directly or by using the probability generating function (or the characteristic function), because

$$\lim_{n\to\infty} P_{X_n}(s) = \lim_{n\to\infty}\left(\frac{1-\lambda/n}{1-\lambda s/n}\right)^n = \frac{e^{-\lambda}}{e^{-\lambda s}}, \tag{5}$$

which is the probability generating function of $Poi(\lambda)$.

Note that Theorem 10 is analogous to the Poisson theorem for the binomial distribution, and it could be easily "guessed" from the Poisson theorem. Indeed, consider the interpretation of $X \sim Pas(n,p)$ from the introduction. Assume that p is very small and n is very large. Then, the total number of trials will be about n (almost all of the trials will be of the type $\mathcal{Q}$, plus a few of the type $\mathcal{P}$). Therefore, the experiment is similar to the one in which we fix n in advance, and we count the number of $\mathcal{P}$'s. That is, the distribution of X is similar to $Bin(n,p)$ with our large n and small p, i.e., X is approximately $Poi(np)$.

It is also worth mentioning that Theorem 10 is closely related to the central theorem of the Poisson process: If we sequentially add independent exponential random variables with mean value μ until we exceed some $t > 0$, the number of these variables is $Poi(t/\mu)$-distributed. The geometric distribution is a discrete analogue of the exponential distribution, therefore the Pascal distribution is the discrete analogue of the Erlang distribution (a special case of the Gamma distribution).