Lecture 2
Discrete Random Variables

Chao Song

College of Ecology
Lanzhou University

September 8, 2025

Motivating examples

Example 1: Let the random experiment be throwing a die. The sample space associated with this experiment is $S = \{1, 2, 3, 4, 5, 6\}$, with elements of S indicating the number of spots on the side facing up. Let X be a function such that $X(s) = s$. Now, X is a real-valued function that has the outcome space S as its domain and $\{0, 1, 2, 3, 4, 5, 6\}$ as its space.

Example 2: A rat is selected at random from a cage and its sex is determined. The sample space is thus $S = \{female, male\} = \{F, M\}$. Let X be a function that has the outcome space S as its domain and the set of real numbers $\{x : x = 0, 1\}$ as its range.

Definition of random variables

Definition: Given a random experiment with an sample space S, a function X that assigns one and only one real number $X(s) = x$ to each element s in S is called a random variable. The space of X is the set of real number $\{x : X(s) = x, s \in S\}$

Example 1: Let the random experiment be throwing a die. The sample space associated with this experiment is $S = \{1, 2, 3, 4, 5, 6\}$, with elements of S indicating the number of spots on the side facing up. Let $X(s) = s$, the space of the random variable X is $\{1, 2, 3, 4, 5, 6\}$.

Example 2: A rat is selected at random from a cage and its sex is determined. The sample space is $S = \{\textit{female}, \textit{male}\}$. Let X be a function such that $X(F) = 0$ and $X(M) = 1$. X is a random variable with space $\{0, 1\}$

Definition of random variables

A few remarks on the definition of random variable:

- Intuitively, we may view random variable as a quantity whose value is determined by the outcome of an random experiment. For practical purpose, this intuitive interpretation of random variable is sufficient;
- Rigorously, a random variable is a function that maps the outcome of a random experiment to real numbers. This is mainly for mathematical rigor.
- Roughly speaking, because probability is a measure mapping events to unit interval. The argument of probability is events. Thus, if we are going to define probability for random variable, we must be able to interpret $\{X \leqslant x\}$ as an event.
- How to map outcome of random experiment to a real number is not a trivial mathematical question. In practice, the choice is often made based on intuition or convenience.

Discrete random variables

Discrete random variable: a random variable is discrete if it only takes values that are in some countable subsets $\{x_1, x_2, \ldots\}$ of real number.

- Number of heads in 10 coin flips;
- Number of coin flips until we have two heads;
- Species richness in a country;
- Number of students late to this class each week.

Probability mass function

Definition: The probability mass function (PMF) of a discrete random variable X is the function $f(x) : \mathbb{R} \to [0, 1]$ given by $f(x) = P(X = x)$.

Properties of probability mass function:

- $0 \leqslant f(x) \leqslant 1$ for all x;
- $f(x) = 0$ if $x \notin \{x_1, x_2, \ldots\}$;
- $\sum_x f(x) = 1$.

Probability mass function

Example: Let X be the number of heads when tossing two fair coins. What is the probability mass function for random variable X?

Answer: possible number of heads are 0, 1, 2. The PMF of X is

- $P(X = 0) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$;
- $P(X = 1) = \frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{2} = \frac{1}{2}$;
- $P(X = 2) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$;

Cumulative distribution function

Definition: The cumulative distribution function (CDF) of a random variable X is the function $F(x) : \mathbb{R} \to [0, 1]$ given by $F(x) = P(X \leqslant x)$

Properties of cumulative distribution function:

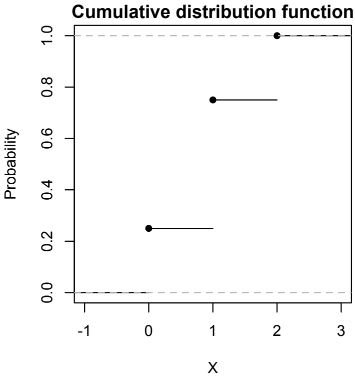
- $F(x)$ is a non-decreasing function: if $x < y$, then $F(x) \leqslant F(y)$;
- $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to+\infty} F(x) = 1$;
- $F(x)$ is right-continuous.

Proposition: Consider real numbers x and y with $x < y$, then

- $P(X > x) = 1 - F(x)$;
- $P(x < X \leqslant y) = F(y) - F(x)$;

Visualize PMF and CDF

Example: Let X be the number of heads when tossing two fair coins. Possible values for X are 0, 1, and 2. The PMF and CDF of X are:



Mathematical expectation

In addition to PMF and CDF, which fully characterize the distribution of a random variable, **mathematical expectation** is an important concept in summarizing characteristics of distribution of probability.

Definition: if $f(x)$ is the probability mass function of the discrete random variable X with space S, and if the summation $\sum_{x \in X} u(x)f(x)$ exists, then the sum is called the mathematical expectation or the expected value of $u(x)$, and it is denoted $E[u(x)]$

Mathematical expectation

Example: Let X be the number of heads when tossing two coins. What is the expected value of X? If one gets two points for each head, what is the expected value of points?

Answer: The expected value of number of heads is

$$E(X) = 0 \times \frac{1}{4} + 1 \times \frac{1}{2} + 2 \times \frac{1}{4} = 1.$$

Let $u(x)$ be the points one get after tossing two coins, $u(x) = 2x$, then

$$E[u(x)] = (2 \times 0) \times \frac{1}{4} + (2 \times 1) \times \frac{1}{2} + (2 \times 2) \times \frac{1}{4} = 2.$$

Properties of mathematical expectation

When exists, the mathematical expectation satisfies the following properties:

- If c is a constant, then $E(c) = c$;
- If c is a constant, $E[cu(x)] = cE[u(x)]$;
- if c_1 and c_2 are constant, $E[c_1 u_1(x) + c_2 u_2(x)] = c_1 E[u_1(x)] + c_2 E[u_2(x)]$.

The above properties arise from the fact that **mathematical expectation is a linear operation**. Thus nonlinear operations cannot be applied the same way. For example, $E(x^2) \neq [E(x)]^2$ in general.

Mean and variance

Mean and **variance** are special cases of the mathematical expectation. Let X be a discrete random variable with probability mass function $f(x)$

- Mean: $\mu = E(X) = \sum_{x \in S} x f(x)$;
- Variance: $\sigma^2 = Var(X) = E[(X - \mu)^2] = \sum_{x \in S} (x - \mu)^2 f(x)$

Mean and variance

Let X be a random variance with mean μ and variance σ. Its variance can be calculated as $\sigma^2 = E(X^2) - \mu^2$

Proof:

$$\begin{aligned} \sigma^2 &= E[(x - \mu)^2] = E[X^2 - 2\mu X + \mu^2] \\ &= E(X^2) - 2\mu E(X) + \mu^2 \\ &= E(X^2) - \mu^2 \end{aligned}$$

Mean and variance

Properties of mean and variance: Let X be a random variable with mean μ and variance σ^2. Let a and b be constants. What is the mean and variance of $aX + b$?

Based on the property of mathematical expectation, we have

- $E(aX + b) = aE(X) + b = a\mu + b$;
- $Var(aX + b) = E[(aX + b - a\mu - b)^2] = E[a^2(X - \mu)^2] = a^2\sigma^2$

Moment

We can view x_i as the distance of that point from the origin. In mechanics, the product of a distance and its weight is called a moment, so $x_i f(x_i)$ is a moment having a moment arm of length x_i. The sum of these products would be the moment of the system of distance and weights.

Definition: For a random variable with probability mass function $f(x)$, we define $\Sigma_{x \in S}(x - a)f(x)$ as the first moment about a. More generally, we call $\Sigma_{x \in S}(x - a)^n f(x)$ the nth moment of X about a.

- The moment about the mean of a random variable μ is called the central moment. The first central moment is always zero.
- The second central moment of a random variable is its variance.

Moment generating function

Definition: Let X be a random variable. We define the moment generating function of X to be

$$m_X(t) = E(e^{tX})$$

Moment generating function, as its name suggests, can be used to find moments of a random variable. For example:

$$\frac{d}{dt} m_X(t) = E(Xe^{tX})$$
$$\frac{d^2}{dt^2} m_X(t) = E(X^2 e^{tX})$$

which when we evaluate at $t = 0$ becomes $E(X)$ and $E(X^2)$. More generally, the nth derivative of $m_X(t)$ evaluated at $t = 0$ is the expected value of X^n, i.e., $m^{(n)}(0) = E(X^n)$

Moment generating function

Example: Suppose random variable has a probability mass function

$$f(x) = q^{x-1}p, \; x = 1, 2, 3, \ldots$$

What is the moment generating function of X? What is the mean of X?

Answer: The moment generating function of X is

$$\begin{aligned} M(t) = E\left(e^{tX}\right) &= \sum_{x=1}^{\infty} e^{tx} q^{x-1} p \\ &= \frac{p}{q} \sum_{x=1}^{\infty} (qe^t)^x \\ &= \frac{p}{q} \sum_{x=1}^{\infty} (qe^t) + (qe^t)^2 + (qe^t)^3 + \cdots \\ &= \frac{p}{q} \frac{qe^t}{1 - qe^t} = \frac{pe^t}{1 - qe^t} \end{aligned}$$

Moment generating function

We use the derivatives of the moment generating function to calculate the mean:

$$
\begin{aligned}
M'(t) &= \frac{(1 - qe^t)pe^t - pe^t(-qe^t)}{(1 - qe^t)^2} \\
&= \frac{pe^t}{(1 - qe^t)^2}
\end{aligned}
$$

Evaluating $M'(t)$ at 0, we have:

$$E(X) = M'(0) = \frac{p}{(1 - q)^2}$$

Bernoulli distribution

A **Bernoulli trial** is a random experiment, the outcome of which can be classified in one of the two mutually exclusive and exhaustive ways–say, success of failure. Let X be a random variable associated with a Bernoulli trial such that $X = 1$ for success and $X = 0$ for failure, X follows a **Bernoulli distribution**.

Example: Suppose that the probability of germination of a beet seed is 0.8 and the germination of a seed is called a success. If we plant 10 seeds and can assume that the germination of one seed is independent of the germination of another seed. This would correspond to 10 Bernoulli trials with $p = 0.8$.

Bernoulli distribution

The probability mass function of X following a Bernoulli distribution is

$$f(x) = \begin{cases} p, & X = 1 \\ 1 - p, & X = 0 \end{cases},$$

Or more concisely, $f(x) = p^x (1 - p)^{1-x}$.

The mean and variance of a Bernoulli distribution is

- $E(X) = 1 \times p + 0 \times (1 - p) = p$
- $Var(X) = (1 - p)^2 p + (0 - p)^2 (1 - p) = p(1 - p)$

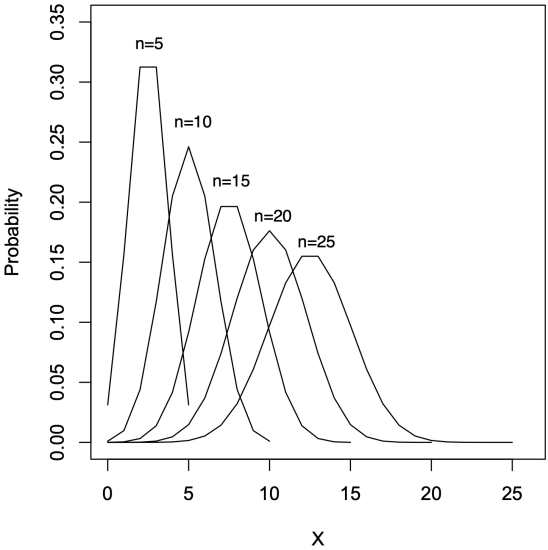
Binomial distribution

In a sequence of Bernoulli trials, we are often interested in the total number of successes, but not the actual order of their occurrences. Let random variable X equal the number of observed successes in n Bernoulli trials.

Binomial distribution: If a random variable X denotes the number of successes in n independent Bernoulli trials, X follows a binomial distribution and its PMF is

$$P(X = k) = \mathbf{C}_n^k p^k (1-p)^{n-k},\ k = 0, 1, \ldots, n$$

Binomial distribution



Binomial distribution

What is the mean and variance of a binomial distribution?

$$
\begin{aligned}
E(X) &= \sum_{x=0}^{n} x \cdot \mathbf{C}_n^x p^x (1-p)^{n-x} \\
&= \sum_{x=1}^{n} x \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \\
&= np \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1} (1-p)^{n-x} \\
&= np
\end{aligned}
$$

because $\sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1}(1-p)^{n-x}$ is the binomial expansion of $(p+1-p)^{n-1}$ and is thus equal to 1.

Binomial distribution

$$
\begin{aligned}
E(X^2) &= \sum_{x=0}^{n} x^2 \cdot \mathbf{C}_n^x p^x (1-p)^{n-x} \\
&= \sum_{x=0}^{n} x(x-1) \cdot \mathbf{C}_n^x p^x (1-p)^{n-x} + \sum_{x=0}^{n} x \cdot \mathbf{C}_n^x p^x (1-p)^{n-x} \\
&= \sum_{x=2}^{n} x(x-1) \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} + np \\
&= \sum_{x=2}^{n} n(n-1)p^2 \frac{(n-2)!}{(x-2)!(n-x)!} p^{x-2} (1-p)^{n-x} + np \\
&= n(n-1)p^2 + np \\
&= n^2p^2 - np^2 + np
\end{aligned}
$$

$$
Var(X) = E(X^2) - [E(X)]^2 = n^2p^2 - np^2 + np - (np)^2 = np(1-p)
$$

Binomial distribution

We can also derive the mean and variance using MGF:

$$
\begin{aligned}
M_X(t) &= E(e^{tX}) = \sum_{x=0}^{n} e^{tx} \mathbf{C}_n^x p^x (1-p)^{n-x} \\
&= \sum_{x=0}^{n} \mathbf{C}_n^x (pe^t)^x (1-p)^{n-x} \\
&= (pe^t + 1 - p)^n
\end{aligned}
$$

$$
\begin{aligned}
E(X) &= M_X'(0) = n(pe^t + 1 - p)^{n-1} pe^t \Big|_{t=0} = np \\
E(X^2) &= M_X''(0) \\
&= n(pe^t + 1 - p)^{n-1} pe^t + n(n-1)(pe^t + 1 - p)^{n-2}(pe^t)^2 \Big|_{t=0} \\
&= np + n^2p^2 - np^2
\end{aligned}
$$

Hypergeometric distribution

A urn contains N balls and K of them are marked. If you randomly select n balls, what is the probability that you get k marked balls?

Let X be the number of marked balls in the n balls one selected,

$$P(X = k) = \frac{\mathbf{C}_K^k \mathbf{C}_{N-K}^{n-k}}{\mathbf{C}_N^n}$$

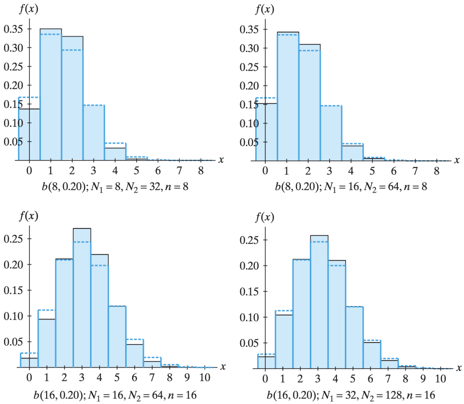
Binomial vs hypergeometric distribution

A urn contains $N1$ white balls and $N2$ of them are marked. Let $p = N_1/(N_1 + N_2)$ and X equal the number of marked balls in a random sample of size n. What is the distribution of X (1) if the sampling is done one at a time with replacement? and (2) if the sampling is one without replacement?

Answer: If sampling is done with replacement, all successive draws are independent. X thus follows a binomial distribution. In contrast, if sampling is done without replacement, one draw will influence the probability of drawing in the next round, we thus have a hypergeometric distribution for X.

Binomial vs hypergeometric distribution

If there are very large number of balls in total compared to the number of balls we draw, i.e., $(N_1 + N_2) \gg n$, hypergeometric distribution and binomial distribution becomes similar.

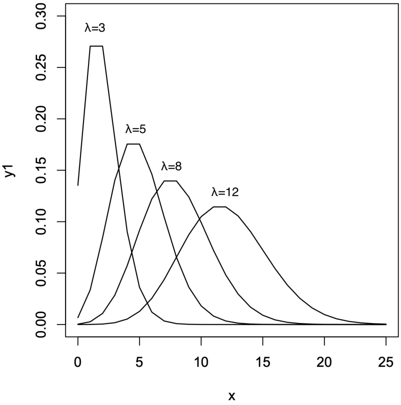


Comparison of binomial and hypergeometric distribution (shaded)

Poisson distribution

Poisson distribution: Let λ be a positive number. A random variable is said to have a Poisson distribution if its probability mass function is

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda},\ k = 0, 1, 2, \ldots$$



Poisson distribution

Let X follows a Poisson distribution with parameter λ. Show that its mean and variance are both λ.

$$
\begin{aligned}
E(X) &= \sum_{x=0}^{\infty} x \cdot \frac{\lambda^x}{x!} e^{-\lambda} \\
&= \sum_{x=1}^{\infty} x \cdot \frac{\lambda^x}{x!} e^{-\lambda} \\
&= \lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} \\
&= \lambda e^{-\lambda} e^{\lambda} \\
&= \lambda
\end{aligned}
$$

given the power series expansion of exponential function $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$

Poisson distribution

To get the variance of X, we first get $E(X^2)$:

$$\begin{aligned}
E(X^2) &= \sum_{x=0}^{\infty} x^2 \frac{\lambda^x}{x!} e^{-\lambda} \\
&= \sum_{x=0}^{\infty} x(x-1) \frac{\lambda^x}{x!} e^{-\lambda} + \sum_{x=0}^{\infty} x \frac{\lambda^x}{x!} e^{-\lambda} \\
&= \sum_{x=2}^{\infty} x(x-1) \frac{\lambda^x}{x!} e^{-\lambda} + \lambda \\
&= \lambda^2 e^{-\lambda} \sum_{x=2}^{\infty} \frac{\lambda^{x-2}}{(x-2)!} + \lambda \\
&= \lambda^2 e^{-\lambda} e^{\lambda} + \lambda \\
&= \lambda^2 + \lambda
\end{aligned}$$

Thus, $Var(X) = E(X^2) - [E(X)]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$

Poisson distribution

We can also derive the mean and variance from the MGF:

$$\begin{aligned}
M_X(t) &= E(e^{tX}) = \sum_{x=0}^{\infty} e^{tx} \frac{\lambda^x}{x!} e^{-\lambda} \\
&= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!} \\
&= e^{-\lambda} e^{\lambda e^t} \\
&= e^{\lambda e^t - \lambda}
\end{aligned}$$

$$\begin{aligned}
E(X) &= M_X'(0) = (e^{\lambda e^t - \lambda} \lambda e^t)\Big|_{t=0} = \lambda \\
E(X^2) &= M_X''(0) = \lambda e^{\lambda e^t - \lambda + t}(\lambda e^t + 1)\Big|_{t=0} = \lambda^2 + \lambda \\
Var(X) &= E(X^2) - [E(X)]^2 = \lambda
\end{aligned}$$

Poisson distribution

What does a Poisson distributed variable model?

Poisson distribution models the number of events in a time interval t.

- Divide t into n segments such that at most one event occur within a segment;
- Probability of occurrence is $\alpha t/n$;
- Number of occurrence is modeled with a binomial distribution.

$$
\begin{aligned}
P(X=k) &= \lim_{n\to\infty} \mathbf{C}_n^k p^k (1-p)^{n-k} \\
&= \lim_{n\to\infty} \frac{n!}{k!(n-k)!} (\frac{\alpha t}{n})^k (1-\frac{\alpha t}{n})^{n-k} \\
&= \lim_{n\to\infty} \frac{(\alpha t)^k}{k!} \frac{n(n-1)\dots(n-k+1)}{n^k} (1-\frac{\alpha t}{n})^{-k} (1-\frac{\alpha t}{n})^n \\
&= \frac{(\alpha t)^k}{k!} e^{-\alpha t}
\end{aligned}
$$

Poisson distribution

Poisson distribution is a limiting case of a binomial distribution. Here, $\lambda = \alpha t$ is often referred to as the rate parameter of the Poisson distribution.

This derivation gives us a mechanistic insights into when we can use Poisson distribution. When some events occur at a constant rate, we can model the count of event with a Poisson distribution.

Poisson distribution

Example: In a large city, telephone calls to 110 come on the average of two every 3 minutes. If one assumes a Poisson process, what is the probability of five or more calls arriving in a 9-minute period?

Let X denote the number of calls in a 9-minute period. We see that $E(X) = 2 \times 9/3 = 6$. Thus, the PMF of X is

$$P(X = k) = \frac{6^x}{x!} e^{-6}$$

Thus, we have

$$\begin{aligned} P(X \geqslant 5) &= 1 - P(X \leqslant 4) \\ &= 1 - \sum_{x=0}^{4} \frac{6^x}{x!} e^{-6} \\ &= 0.715 \end{aligned}$$

Negative binomial distribution

Negative binomial distribution: In a sequence of independent Bernoulli trials with success probability p, let X be the number of failure until r successes. Then X has a negative binomial distribution with probability mass function

$$P(X = k) = \mathbf{C}_{k+r-1}^{k}(1-p)^{k}p^{r}$$

Negative binomial distribution can be defined in alternative ways. For example, in a sequence of independent Bernoulli trials with success probability p, the number of trials X needed to observe r success also has a negative binomial distribution. The PMF here is

$$P(X = m) = \mathbf{C}_{r-1}^{m-1}(1-p)^{m-r}p^{r}$$

Negative binomial distribution

What is the mean and variance of a negative binomial distribution?

To calculate the mean and variance, we first get the MGF:

$$
\begin{aligned}
M(t) &= \sum_{k=0}^{\infty} e^{tk} \mathbf{C}_{k+r-1}^{k} (1-p)^k p^r \\
&= p^r \sum_{k=0}^{\infty} \mathbf{C}_{k+r-1}^{k} [(1-p)e^t]^k \\
&= \frac{p^r}{[1-(1-p)e^t]^r}
\end{aligned}
$$

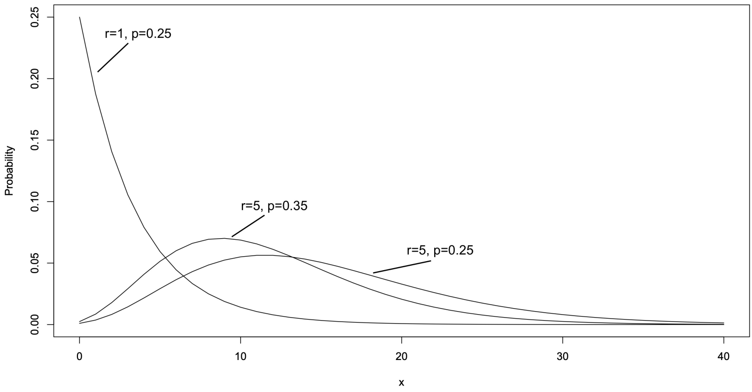
Note: here, we used the Taylor expansion of $(1-w)^{-r}$ at $w=0$

$$
(1-w)^{-r} = \sum_{k=0}^{\infty} \frac{h^{(k)}(0)}{k!} w^k = \sum_{k=0}^{\infty} \mathbf{C}_{k+r-1}^{r-1} w^k = \sum_{k=0}^{\infty} \mathbf{C}_{k+r-1}^{k} w^k
$$

Using the derivatives of $M(t)$ evaluated at $t=0$, we get that the mean of X is $\frac{r(1-p)}{p}$ and the variance of X is $\frac{r(1-p)}{p^2}$.

Negative binomial distribution

The negative binomial distribution can take on a variety of shapes, depending on the parameters *r* and *p*. An important feature of negative binomial distribution is that its variance is larger than the mean.



Geometric distribution

Geometric distribution: In a sequence of independent Bernoulli trials with success probability p, let X be the total number of failures until we have1 successes, X has a geometric distribution with probability mass function:

$$P(X = x) = (1 - p)^x p$$

Geometric distribution is a special case of negative binomial distribution.

The mean and variance of the geometric distribution is $\frac{1-p}{p}$ and $\frac{1-p}{p^2}$, respectively.

Summary of common discrete distributions

Distribution	Probability mass function	Mean	Variance
Bernoulli	$p^x(1-p)^{1-x}$	p	$p(1-p)$
Binomial	$\mathbf{C}_n^k p^k (1-p)^{n-k}$	np	$np(1-p)$
Poisson	$\frac{\lambda^k}{k!} e^{-\lambda}$	λ	λ
Negative binomial	$\mathbf{C}_{k+r-1}^k (1-p)^k p^r$	$\frac{r(1-p)}{p}$	$\frac{r(1-p)}{p^2}$
Geometric	$(1-p)^{k-1} p$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Hypergeometric	$\frac{\mathbf{C}_K^k \mathbf{C}_{N-K}^{n-k}}{\mathbf{C}_N^n}$	$\frac{nK}{N}$	$\frac{nK(N-K)(N-n)}{N^2(N-1)}$