

The Hypergeometric Distribution and Its Properties

Starting from the definition, the expectation and variance of the hypergeometric distribution are completely deduced with the help of two combination identities (reduced order formula and Vandermonde identity), and every step is well documented.

June 3, 2026 • High School Mathematics

Definition

$X \sim H(n, K, N)$, X is said to obey the hypergeometric distribution.

$$P(X = k) = \frac{C_K^k C_{N-K}^{n-k}}{C_N^n} (k = 0, 1, 2, \dots, n)$$

Preliminary knowledge

Commonly used combined identities

1. $C_n^k = \frac{n}{k} C_{n-1}^{k-1} = \frac{n(n-1)}{k(k-1)} C_{n-2}^{k-2} = \dots$
2. $\sum_{k=0}^r C_m^k C_n^{r-k} = C_{m+n}^r$

Proof 1:

$$\begin{aligned} C_n^k &= \frac{n!}{k!(n-k)!} \\ &= \frac{n}{k} \frac{(n-1)!}{(k-1)![(n-1)-(k-1)]!} \\ &= \frac{n}{k} C_{n-1}^{k-1} = \frac{n(n-1)}{k(k-1)} C_{n-2}^{k-2} \end{aligned}$$

Proof 2:

Consider this scenario: select r people ($r \leq \min m, n$) among m boys and n girls. It is not difficult to find that 2 is the result of counting twice.

To add, 2 is exactly the famous **Vandermonde Identity**

Properties of random variables

$$D(X) = E(X^2) - [E(X)]^2$$

Starting from the definition of variance:

$$\begin{aligned} D(X) &= \sum_{i=0}^n p_i [X_i - E(X)]^2 \\ &= \sum_{i=0}^n (p_i X_i^2) - 2E(X) \sum_{i=0}^n (p_i X_i) + [E(X)]^2 \\ &= E(X^2) - [E(X)]^2 \end{aligned}$$

Properties

1. $E(X) = n \frac{K}{N}$
2. $D(X) = n \frac{K}{N} \frac{N-K}{N} \frac{N-n}{N-1}$

$$\begin{aligned} E(X) &= \sum_{i=0}^n [P(X = i)i] \\ &= \sum_{i=0}^n \left[\frac{C_K^i C_{N-K}^{n-i}}{C_N^n} i \right] \end{aligned}$$

apply combinatorial identity 1 to eliminate i

$$\begin{aligned} &= \sum_{i=1}^n \left[\frac{K}{i} \frac{C_{K-1}^{i-1} C_{N-K}^{n-i}}{C_N^n} i \right] \\ &= K \sum_{i=1}^n \left[\frac{C_{K-1}^{i-1} C_{N-K}^{n-i}}{C_N^n} \right] \end{aligned}$$

apply combinatorial identity 2

$$= \frac{K}{C_N^n} C_{N-1}^{n-1}$$

apply combinatorial identity 1

$$= \frac{nK}{N}$$

$$\begin{aligned}
 E(X^2) &= \sum_{i=0}^n [P(X=i)i^2] \\
 &= \sum_{i=0}^n \left[\frac{C_K^i C_{N-K}^{n-i}}{C_N^n} i^2 \right]
 \end{aligned}$$

apply combinatorial identity 1 to eliminate i

$$\begin{aligned}
 &= K \sum_{i=1}^n \left[\frac{C_{K-1}^{i-1} C_{N-K}^{n-i}}{C_N^n} i \right] \\
 &= K \sum_{i=1}^n \left[(i-1+1) \frac{C_{K-1}^{i-1} C_{N-K}^{n-i}}{C_N^n} \right] \\
 &= K \sum_{i=1}^n \left[(i-1) \frac{C_{K-1}^{i-1} C_{N-K}^{n-i}}{C_N^n} + \frac{C_{K-1}^{i-1} C_{N-K}^{n-i}}{C_N^n} \right]
 \end{aligned}$$

apply combinatorial identity 2

$$\begin{aligned}
 &= K \sum_{i=1}^n \left[(i-1) \frac{C_{K-1}^{i-1} C_{N-K}^{n-i}}{C_N^n} \right] + K \frac{C_{N-1}^{n-1}}{C_N^n} \\
 &= K \sum_{i=1}^n \left[(i-1) \frac{C_{K-1}^{i-1} C_{N-K}^{n-i}}{C_N^n} \right] + \frac{Kn}{N}
 \end{aligned}$$

apply combinatorial identity 1 to eliminate $i-1$

$$= K \sum_{i=2}^n \left[(K-1) \frac{C_{K-2}^{i-2} C_{N-K}^{n-i}}{C_N^n} \right] + \frac{Kn}{N}$$

apply combinatorial identity 2

$$\begin{aligned}
 &= K \left[(K-1) \frac{C_{N-2}^{n-2}}{C_N^n} \right] + \frac{Kn}{N} \\
 &= K(K-1) \frac{n(n-1)}{N(N-1)} + \frac{Kn}{N} \\
 &= \frac{Kn}{N} \left[\frac{(K-1)(n-1)}{N-1} + 1 \right] \\
 D(X) &= E(X^2) - [E(X)]^2 \\
 &= \frac{Kn}{N} \left[\frac{(K-1)(n-1)}{N-1} + 1 \right] - \left(\frac{Kn}{N} \right)^2 \\
 &= \frac{Kn}{N} \left[\frac{N^2 - (K+n)N + Kn}{N(N-1)} \right] \\
 &= n \frac{K}{N} \frac{N-K}{N} \frac{N-n}{N-1}
 \end{aligned}$$

Summary

At this point, we have completely derived the distribution law based on only two combination identities.

$$E(X) = n \frac{K}{N}, \quad D(X) = n \frac{K}{N} \cdot \frac{N-K}{N} \cdot \frac{N-n}{N-1}.$$

There are two key points in the whole process: one is to repeatedly use identity 1 to **reduce** the combination number and eliminate i in the summation term with i ; the other is to use the Vandermonde identity to **merge** the summation into a single combination number. The reason why i^2 is split into $(i-1) + 1$ when calculating the variance is precisely to create $i(i-1)$, a structure that can be reduced to two orders again. In essence, it is to find the factorial moment $E[X(X-1)]$.

Note $p = K/N$, and the result can be written as $E(X) = np$, $D(X) = np(1-p) \cdot \frac{N-n}{N-1}$.

Compared with $np(1-p)$ of the binomial distribution, the extra factor $\frac{N-n}{N-1}$ is called the **finite population correction factor** - it is less than 1, which depicts the fact that sampling without replacement makes the variance smaller; when $N \gg n$, the factor tends to 1, and the hypergeometric distribution is approximated as a binomial distribution.