

超几何分布及其性质

从定义出发,借助两条组合恒等式(降阶公式与范德蒙德恒等式)完整推导超几何分布的期望与方差,每一步都有据可循。

2026年6月3日 · 高考数学

定义

$X \sim H(n, K, N)$, 称 X 服从超几何分布.

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

预备知识

常用组合恒等式

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

证明1:

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

证明2:

考虑这样一个情景:在 m 个男生和 n 个女生中选出 r 人($r \leq \min m, n$), 不难发现2是算两次的结果.

补充一下, 2正是著名的范德蒙德恒等式

随机变量的性质

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

从方差定义出发:

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

性质

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

$$E(X) = \sum_{i=0}^n [P(X=i) i]$$

$$= \sum_{i=0}^n \left[\frac{C_K^i C_{N-K}^{n-i}}{C_N^n} i \right]$$

运用组合恒等式1消i

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

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

运用组合恒等式2

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

运用组合恒等式1

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

$$\begin{aligned}
E(X^2) &= \sum_{i=0}^n [P(X=i)i^2] \\
&= \sum_{i=0}^n \left[\frac{C_K^i C_{N-K}^{n-i}}{C_N^n} i^2 \right] \\
&\quad \text{运用组合恒等式1消}i \\
&= K \sum_{i=1}^n \left[\frac{C_{K-1}^{i-1} C_{N-K}^{n-i}}{C_N^n} i \right] \\
&= K \sum_{i=1}^n \left[(i-1+1) \frac{C_{K-1}^{i-1} C_{N-K}^{n-i}}{C_N^n} \right] \\
&= K \sum_{i=1}^n \left[(i-1) \frac{C_{K-1}^{i-1} C_{N-K}^{n-i}}{C_N^n} + \frac{C_{K-1}^{i-1} C_{N-K}^{n-i}}{C_N^n} \right] \\
&\quad \text{运用组合恒等式2} \\
&= K \sum_{i=1}^n \left[(i-1) \frac{C_{K-1}^{i-1} C_{N-K}^{n-i}}{C_N^n} \right] + K \frac{C_{N-1}^{n-1}}{C_N^n} \\
&= K \sum_{i=1}^n \left[(i-1) \frac{C_{K-1}^{i-1} C_{N-K}^{n-i}}{C_N^n} \right] + \frac{Kn}{N} \\
&\quad \text{运用组合恒等式1消}i-1 \\
&= K \sum_{i=2}^n \left[(K-1) \frac{C_{K-2}^{i-2} C_{N-K}^{n-i}}{C_N^n} \right] + \frac{Kn}{N} \\
&\quad \text{运用组合恒等式2} \\
&= K \left[(K-1) \frac{C_{N-2}^{n-2}}{C_N^n} \right] + \frac{Kn}{N} \\
&= K(K-1) \frac{n(n-1)}{N(N-1)} + \frac{Kn}{N} \\
&= \frac{Kn}{N} \left[\frac{(K-1)(n-1)}{N-1} + 1 \right] \\
D(X) &= E(X^2) - [E(X)]^2 \\
&= \frac{Kn}{N} \left[\frac{(K-1)(n-1)}{N-1} + 1 \right] - \left(\frac{Kn}{N} \right)^2 \\
&= \frac{Kn}{N} \left[\frac{N^2 - (K+n)N + Kn}{N(N-1)} \right] \\
&= n \frac{K}{N} \frac{N-K}{N} \frac{N-n}{N-1}
\end{aligned}$$

小结

至此,我们仅凭两条组合恒等式,就从分布律出发完整推出了

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

整个过程的关键有两点:一是反复使用恒等式 1 给组合数**降阶**,把带 i 的求和项里的 i 消掉;二是用范德蒙德恒等式把求和**合并**成单个组合数。计算方差时之所以把 i^2 拆成 $(i-1) + 1$,正是为了凑出 $i(i-1)$ 这个能再次降两阶的结构,本质上是在求阶乘矩 $E[X(X-1)]$ 。

记 $p = K/N$,结果可写成 $E(X) = np$ 、 $D(X) = np(1-p) \cdot \frac{N-n}{N-1}$ 。对比二项分布的 $np(1-p)$,多出的因子 $\frac{N-n}{N-1}$ 称为**有限总体修正因子**——它小于 1,刻画了不放回抽样让方差变小的事实;当 $N \gg n$ 时该因子趋于 1,超几何分布便近似为二项分布。