§11. Hypergeometric Random Variables

Context: Consider a set of N objects; r of which are of special type. Suppose that we choose n objects, without replacements and without regards to order. What is the probability that we get exactly x of the special objects?

Definition 1

A **hypergeometric random variable** $X \sim \text{Hypergeom}(N, r, n)$ has pmf

$$P(x) = \frac{{}_rC_x \cdot {}_{N-r}C_{n-x}}{{}_NC_n} \quad ; \quad x = 0, \dots, \min\{n, r\}$$

Proposition

The expected value and the variance are

$$E(X) = \frac{n \cdot r}{N} \qquad \text{Var}(X) = \frac{nr}{N} \cdot \left(1 - \frac{r}{N}\right) \cdot \left(\frac{N-n}{N-1}\right)$$

Example 1

You draw five cards at random without replacement. Let X be the number of face cards. Compute the pmf, $E(X)$ and σ_X.

X	0	1	2	3	4	5
$P(X = x)$	$\frac{{}_{12}C_0 \cdot {}_{40}C_5}{{}_{52}C_5}$ ↓ 0.2532	$\frac{{}_{12}C_1 \cdot {}_{40}C_4}{{}_{52}C_5}$ ↓ 0.4220	$\frac{{}_{12}C_2 \cdot {}_{40}C_3}{{}_{52}C_5}$ ↓ 0.2509	$\frac{{}_{12}C_3 \cdot {}_{40}C_2}{{}_{52}C_5}$ ↓ 0.0660	$\frac{{}_{12}C_4 \cdot {}_{40}C_1}{{}_{52}C_5}$ ↓ 0.0076	$\frac{{}_{12}C_5 \cdot {}_{40}C_0}{{}_{52}C_5}$ ↓ 0.0003

Not necessary to use $E(X) = \sum xp(x)$ and $\text{Var(X)} = \sum (X - \mu)^2 p(x)$

$$E(X) = \frac{5 \cdot 12}{52} = 1.1538$$

$$\text{Var}(X) = \frac{5.12}{52} \cdot \left(1 - \frac{12}{52}\right) \cdot \left(\frac{52-5}{52-1}\right) = 0.81796 \quad \Rightarrow \quad \sigma_X = 0.9044$$

Remark
Taking $N \to \infty; \quad r \to \infty$ with $\frac{r}{N} \to p$ the hypergeometric distribution approaches the binomial distribution: $Hyp(N, r, n) \xrightarrow{N \to \infty} Bin(n, p)$

Example 2

a) In a population of 1000 blood donors, 70 are type O^-. In a random sample of 50 from this population what is the probability that precisely four donnors are type O^-?
b) In a certain population 7% of all blood donors are of type O^-. In a random sample of 50 what is the probability precisely four donnors are of type O^-?

Solution

a) Hypergeometric: $N = 1000, r = 70, n = 50$.

$$p(4) = \frac{{}_{70}C_{4}\ {}_{9930}C_{46}}{{}_{1000}C_{50}} = 0.201672$$

b) Binomial: $n = 50, p = 0.07$.

$$p(4) = {}_{50}C_4\ (0.07)^4(0.93)^{46} = 0.196292$$

Example 3

A lot of 100 items contains 4 defectives. A sample of 5 items is drawn and any defective item in the sample is replaced by a good item. On average, what proportion of defective items remains?

Solution

Let X be the number of defectives in the sample. It has pmf

$$P(x) = \frac{{}_{4}C_{x}\ {}_{96}C_{5-x}}{{}_{100}C_{5}}$$

The number of defective items remaining are $4 - x$. The expected number of defective items remaining is

$$\begin{aligned}\text{Remaining defective } &= \sum_{x=0}^{4}(4-x)p(x) = 4\sum_{x=0}^{4}p(x) - \sum_{x=0}^{4}xp(x) \\ &= 4 - E(X) \\ &= 4 - \frac{5\cdot 4}{100} \\ &= 3.8\end{aligned}$$

The expected proportion of remaining defective items is $\frac{38}{100} = 0.038$

Example 4

In Powerball a player chooses 5 different white balls numbered from 1 to 59 and then also chooses a red ball numbered $1, \ldots, 35$. The Lotto Corp. also chooses (randomly) 5 white and 1 red ball as the winning combination.

a. What is the probability a player wins the jackpot on one selection? (Jackpot today \$90M USD).

b. What is the probability the player matches 4 out of the 5 winning balls and does not match the power ball? (Payout \$100 USD)

Solution

a. $P = \dfrac{1}{{}_{59}C_5} \cdot \dfrac{1}{35} = \dfrac{1}{175223510} \approx 5.7 \times 10^{-9}$

b. $P = \dfrac{{}_5C_4 \cdot {}_{54}C_1}{{}_{59}C_5} \cdot \dfrac{34}{35} = 5.24 \times 10^{-5}$