Example: Multivariate Hypergeometric Distribution

Consider the following experiment. We have

\( n_1 \) objects of type 1

\( n_r \) objects of type \( r \)

Let \( N = n_1 + \ldots + n_r \) be the total number of objects. Suppose we draw without replacement \( n \) objects randomly. The underlying probability space is \((S, P)\), where \( S \) is the set of all possible samples of size \( n \) that we could obtain in this way, and \( P \) specifies that every sample in \( S \) is equally likely. Let \( X_i = \# \) of objects of type \( i \) in the sample we draw, \( i = 1, \ldots, r \).

Let \( X = (X_1, \ldots, X_r)^T \). Then \( X \) is a random vector and its distribution is called the Multivariate Hypergeometric distribution with parameter \( n \) and \( n_1, \ldots, n_r \). To consider the joint pmf of \( X \) we should first consider the possible values of \( X \), i.e., the support \( S_X \) of \( X \).

\[ \begin{align*}
\text{e.g.} & \quad r = 2, \quad n_1 = 3, \quad n_2 = 4, \quad N = 7 \\
S_X \text{ when } & \quad n = 1 \\
& \quad \begin{array}{c}
\text{when } n = 2 \\
\text{when } n = 3 \\
\text{when } n = 4 \\
\text{when } n = 5 \\
\text{when } n = 6 \\
\text{when } n = 7
\end{array}
\end{align*} \]

Note that as the sample size \( n \) increases the support of \( X \) starts hitting constraints imposed by the numbers \( n_i \) of each type of object that there are in the population.
In general, we may write the constraints on $S_x$ as follows:

$$S_x = \{ x = (x_1, \ldots, x_r) \in \mathbb{R}^r : x_i \in \{0, \ldots, n_i\} \text{ for } i = 1, \ldots, r \}
\text{ and } x_1 + \cdots + x_r = n \}$$

More explicitly, we can write $S_x$ as

$$S_x = \{ x = (x_1, \ldots, x_r) \in \mathbb{R}^r : x_i + \cdots + x_r = n, 
\max(0, n - (N-n_i)) \leq x_i \leq \min(n, n_i), 
\text{ and } x_i \text{ is an integer, } i = 1, \ldots, r \}$$

Joint pmf

For $x \in S_x$, $(x = (x_1, \ldots, x_r)^T)$

$$p(x = x) = p(x_1 = x_1, \ldots, x_r = x_r)$$

This is a counting problem. Since all samples in $S$ are equally likely, the above probability is

$$\frac{\text{# of samples in } S \text{ that have } x_1 \text{ type 1 objects, } \ldots, x_r \text{ type } r \text{ obj.}}{\text{total # of samples in } S}$$

Note that samples in $S$ are distinct if the objects in the sample are distinct, i.e., 2 samples are distinct if the objects in 1 sample are not all the same as the objects in the other sample (even if they are of the same type). We have

$$p(x_1 = x_1, \ldots, x_n = x_n) = \frac{(n_1)^{x_1} (n_2)^{x_2} \cdots (n_r)^{x_r}}{N^n}$$

So the joint pmf of $X$ is

$$p_X(x) = \begin{cases} \frac{(n_1)^{x_1} \cdots (n_r)^{x_r}}{N^n} & \text{if } x = (x_1, \ldots, x_n)^T \in S_x \\ 0 & \text{otherwise} \end{cases}$$