Multiple Random Variables

Let $X_1, \ldots, X_n$ be $n$ random variables. The vector $X = (X_1, \ldots, X_n)^T$ is called a random vector. So $X$ is a function from $S$ to $\mathbb{R}^n$, where $(S, P)$ is the underlying probability space: $X(s) = (X_1(s), \ldots, X_n(s))^T$.

The distribution of $X$ is the probability measure on $\mathbb{R}^n$ induced by $P$. If $P_X$ denotes the distribution of $X$, then for $A \subset \mathbb{R}^n$, $P_X(A) = P(\{s \in S : X(s) \in A\})$. Just as with random variables, we usually describe the distribution of $X$ via the following non-negative functions on $\mathbb{R}^n$:

1. **Joint distribution function (Jdf)**, in the case when $X_1, \ldots, X_n$ are all discrete.

2. **Joint probability mass function (Jpmf)**, in the case when $X_1, \ldots, X_n$ are all continuous random variables.

For our discussion, we will assume, unless otherwise specified, that all the $X_i$'s are discrete or all the $X_i$'s are continuous.

Joint Distribution Functions

Let $X = (X_1, \ldots, X_n)^T$ be a random vector. The joint distribution function (Jdf) of $X$ is defined by

$$F_X(x) = P(X_1 \leq x_1, \ldots, X_n \leq x_n), \quad \text{where } x \in (x_1, \ldots, x_n)^T \in \mathbb{R}^n$$

$$= P(\{X_1 \leq x_1\} \cap \{X_2 \leq x_2\} \cap \ldots \cap \{X_n \leq x_n\})$$

$$= P(\{s \in S : X_1(s) \leq x_1\} \cap \ldots \cap \{s \in S : X_n(s) \leq x_n\})$$

$$= P((-\infty, x_1] \times (-\infty, x_2] \times \ldots \times (-\infty, x_n])$$

(note: for any 2 sets $A$ and $B$, $A \times B = \{(a,b) : a \in A, b \in B\}$).

$$= P(A_x), \quad \text{where } A_x = (-\infty, x_1] \times \ldots \times (-\infty, x_n] \subset \mathbb{R}^n.$$
Properties of Joint df's

1. \( 0 \leq F_x(x) \leq 1 \) for all \( x \in \mathbb{R}^n \).
2. \( F_x(x) \) is continuous from the right in each component; if \( y_n \downarrow x_i \) as \( n \to \infty \), then
   \[ \lim_{n \to \infty} F_x(x_1, \ldots, x_{i-1}, y_n, x_{i+1}, \ldots, x_n) = F_x(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n) \]

Joint Probability Mass Functions

If \( X = (X_1, \ldots, X_n) \) is a random vector with \( X_i \) discrete for all \( i = 1, \ldots, n \), then the distribution of \( X \) can be specified by its joint pmf. To say that \( X \) is discrete means that there is a finite or countable subset of \( \mathbb{R}^n \), called the support of the distribution of \( X \), and which we will denote by \( S_X \), that satisfies \( P(X \in S_X) = 1 \). The joint pmf of \( X \) is defined by \( p_X(x) = P(X = x) \) for all \( x \in \mathbb{R}^n \). The jpmf \( p_X(x) \) is defined for all \( x \in \mathbb{R}^n \) and \( p_X(x) \geq 0 \) only if \( x \in S_X \). Also, \( 0 \leq p_X(x) \leq 1 \) and \( \sum_{x \in S_X} p_X(x) = 1 \). For any \( A \subseteq \mathbb{R}^n \),
   \[ P(X \in A) = \sum_{x \in A \cap S_X} p_X(x) \]

Joint Probability Density Functions

If \( X = (X_1, \ldots, X_n) \) is a random vector with \( X_i \) continuous for all \( i = 1, \ldots, n \), then if there exists a function \( f_x : \mathbb{R}^n \to [0, \infty) \) satisfying \( P(X \in A) = \int_A f_X(x_1, \ldots, x_n) \, dx_1 \ldots dx_n \) for any \( A \subseteq \mathbb{R}^n \),
then \( f_X \) is called a joint probability density function of \( X \). Just as continuous random variables may not have a pdf, a continuous random vector may not have a jpdf. In the vector case this is not hard to see.
Example. Let $X_1$ be any continuous random variable.

Let $X_2 = X_1^2$ and $X = (X_1, X_2)^T$. Then if

$A = \{(x_1, x_2) \in \mathbb{R}^2 : x_2 = x_1^2\}$, we have $P(X \in A) = 1$.

If a jpdf for $X$ did exist, say $f_X$, then it must satisfy

$1 = P(X \in A) = \int_A f_X(x_1, x_2) \, dx_1 \, dx_2$. But there does not exist any such function because $A$ has 0 area in $\mathbb{R}^2$. More generally, if $X = (X_1, \ldots, X_n)^T$ is an $n$-dimensional continuous random vector whose support is contained in a subset of $\mathbb{R}^n$ of dimension $m$, with $m < n$, then $X$ cannot have a jpdf.
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 $X$.

E.g., $r = 2$, $n_1 = 3$, $n_2 = 4$, $N = 7$

![Support of X](image)

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:

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

More explicitly, we can write $S_x$ as

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

**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

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

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

$$P(x_1 = x_1, \ldots, x_n = x_n) = \frac{(n_1)(n_2) \cdots (n_r)}{(N)_n}$$

So the joint pmf of $X$ is

$$P_x(x) = \begin{cases} \frac{(n_1)(n_2) \cdots (n_r)}{(N)_n} & \text{if } x = (x_1, \ldots, x_n)^T \in S_x \\ 0 & \text{otherwise} \end{cases}$$
Alternative Description of the Multivariate Hypergeometric Dist'n

Observe that if \( X = (X_1, \ldots, X_r)^T \) has a multivariate hypergeometric distribution with parameters \( n, n_1, \ldots, n_r \) then the components of \( X \) have the constraint that
\[
X_1 + \ldots + X_r = n
\]
where \( n \) is the size of the random sample. Then we get
\[
P(X_1 = x_1, \ldots, X_{r-1} = x_{r-1}) = P(X_1 = x_1, \ldots, X_{r-1} = x_{r-1}, X_r = n - x_1 - \ldots - x_{r-1}) = \binom{n}{x_1} \ldots \binom{n_{r-1}}{x_{r-1}} \binom{n}{n - x_1 - \ldots - x_{r-1}} \binom{N}{n}
\]
That is we can equivalently describe this distribution as a distribution on \( (X_1, \ldots, X_{r-1})^T \). The joint pmf of \( X = (X_1, \ldots, X_{r-1})^T \) is
\[
p_X(x_1, \ldots, x_{r-1}) = \binom{n}{x_1} \ldots \binom{n_{r-1}}{x_{r-1}} \binom{n}{n - x_1 - \ldots - x_{r-1}} \binom{N}{n}
\]
When we describe the distribution in this way the support (possible values of \( X_1, \ldots, X_{r-1} \)) becomes
\[
S_X = \{ (x_1, \ldots, x_{r-1}) \in \mathbb{R}^{r-1} : 0 \leq x_1 + \ldots + x_{r-1} \leq n, \max(0, n - (N - n_1)) \leq x_i \leq \min(n, n_i), \text{ and } x_i \text{ is an integer, } i = 1, \ldots, r-1 \}
\]
Alternatively, we can write the support as
\[
S_X = \{ (x_1, \ldots, x_{r-1}) \in \mathbb{R}^{r-1} : 0 \leq x_1 + \ldots + x_{r-1} \leq n, x_i \in \{0, \ldots, n_i\}, i = 1, \ldots, r-1 \}
\]
\[
\max(0, n - (N - n_1)) \leq x_i, i = 1, \ldots, r-1 \}
Marginal Distributions (Discrete case)

Marginal joint pmf's

Let \( X = (X_1, \ldots, X_n)^T \) be a discrete random vector with joint pmf \( p_X(x_1, \ldots, x_n) \). Let \( \{i_1, \ldots, i_d\} \subseteq \{1, \ldots, n\} \) and \( \{j_1, \ldots, j_{n-d}\} = \{1, \ldots, n\} \setminus \{i_1, \ldots, i_d\} \). We are interested in the joint pmf of \( (X_{i_1}, \ldots, X_{i_d})^T \), and in particular how to obtain it from the full joint pmf \( p_X(x_1, \ldots, x_n) \). The joint pmf of \( (X_{i_1}, \ldots, X_{i_d})^T \) is called the marginal joint pmf and the distribution of \( (X_{i_1}, \ldots, X_{i_d})^T \) is called the marginal joint distribution of \( (X_{i_1}, \ldots, X_{i_d})^T \). The term marginal refers to the situation where we are considering the joint distribution of a set of random variables that is a subset of a larger collection of random variables.

The marginal joint pmf of \( (X_{i_1}, \ldots, X_{i_d})^T \) gives probabilities of the form

\[
P(X_{i_1} = x_{i_1}, \ldots, X_{i_d} = x_{i_d}) = P(X_{i_1} = x_{i_1}, \ldots, X_{i_d} = x_{i_d}, X_{j_1} \in \mathbb{R}, \ldots, X_{j_{n-d}} \in \mathbb{R})
\]

\[
= \sum_{(y_1, \ldots, y_n) \in S_X} p_X(y_1, \ldots, y_n)
\]

where \( S_X = \{(x_1, \ldots, x_n) \in \mathbb{R}^n : p_X(x_1, \ldots, x_n) > 0\} \) is the support of \( X \).
Example Marginals of the Multivariate Hypergeometric

Suppose \( \{i_1, \ldots, i_d\} \subseteq \{1, \ldots, n\} \) and consider the joint distribution of \((X_{i_1}, \ldots, X_{i_d})^T\). The simplest way to get this marginal joint distribution is to go back to the experiment giving rise to the Multivariate Hypergeometric distribution and relabel all objects that are not of type \(i_1, \ldots, i_d\) as type \(0\). Then we have \(n_{i_1}\) objects of type \(i_1\), \(n_{i_d}\) objects of type \(i_d\), \(N - n_{i_1} - \ldots - n_{i_d}\) objects of type \(0\), and then, letting \(X_0\) denote the number of objects of type \(0\) in our sample,

\[
P(X_{i_1} = x_{i_1}, \ldots, X_{i_d} = x_{i_d})
\]

\[
= P(X_{i_1} = x_{i_1}, \ldots, X_{i_d} = x_{i_d}, X_0 = n - x_{i_1} - \ldots - x_{i_d})
\]

\[
= \frac{n_{i_1} \ldots n_{i_d} (N - n_{i_1} - \ldots - n_{i_d})}{n^D}
\]

for \(x = (x_{i_1}, \ldots, x_{i_d}) \in S_X \subseteq \mathbb{R}_d^d\), where

\[
S_X = \{x = (x_{i_1}, \ldots, x_{i_d}) \in \mathbb{R}_d^d : 0 \leq x_{i_1} + \ldots + x_{i_d} \leq n,
\]

\[
\max(0, n - (N - n_{i_k})) \leq x_{i_k} \leq \min(n, n_{i_k})
\]

and \(x_{i_k}\) is an integer, for \(k = 1, \ldots, d\)