5.1.1 Cartesian Products of Sets

**Definition 5.1.1.1: Cartesian Product of Sets**

Let $A, B$ be sets. The Cartesian product of $A$ and $B$ is denoted:

$$A \times B = \{(a, b) : a \in A, b \in B\}$$

Further if $A, B$ are finite sets, then $|A \times B| = |A| \cdot |B|$ by the product rule of counting.

**Examples**

1. For example:

$$\{1, 2, 3\} \times \{4, 5\} = \{(1, 4), (1, 5), (2, 4), (2, 5), (3, 4), (3, 5)\}$$

   We have each of the elements of the first set paired with each of the elements of the second set.
   Note that $|\{1, 2, 3\}| = 3$, $|\{4, 5\}| = 2$, and $|\{1, 2, 3\} \times \{4, 5\}| = 6$.

2. Another example is the $xy$-plane (2D space), which is denoted:

$$\mathbb{R}^2 = \mathbb{R} \times \mathbb{R} = \{(x, y) : x \in \mathbb{R}, y \in \mathbb{R}\}$$

Suppose we roll two fair 4-sided die independently, one blue and one red. Let $X$ be the value of the blue die and $Y$ be the value of the red die. Note:

$$\Omega_X = \{1, 2, 3, 4\}$$
$$\Omega_Y = \{1, 2, 3, 4\}$$

Then we can also consider $\Omega_{X,Y}$, the joint range of $X$ and $Y$. This is:

$$\Omega_{X,Y} = \Omega_X \times \Omega_Y$$

This will just be all the ordered pairs of the values that appear on the two die. Further each of these will be equally likely (as shown in the table below):
5.1-2

Lecture 5.1: Alex Tsun

So for the joint PMF $p_{X,Y}(x, y) = \mathbb{P}(X = x, Y = y)$ for $x, y \in \Omega_{X,Y}$ we have:

$$p_{X,Y}(x, y) = \begin{cases} \frac{1}{16}, & x, y \in \Omega_{X,Y} \\ 0, & \text{otherwise} \end{cases}$$

Note that either this piecewise function or the table above are valid ways to express the joint PMF.

### 5.1.2 Joint PMFs and Expectation

**Definition 5.1.2.1: Joint PMFs**

Let $X, Y$ be discrete random variables. The joint PMF of $X$ and $Y$ is:

$$p_{X,Y}(a, b) = \mathbb{P}(X = a, Y = b)$$

The joint range is the set of pairs $(c, d)$ that have nonzero probability:

$$\Omega_{X,Y} = \{(c, d) : p_{X,Y}(c, d) > 0\} \subseteq \Omega_X \times \Omega_Y$$

Note that the probabilities in the table must sum to 1:

$$\sum_{(s,t) \in \Omega_{X,Y}} p_{X,Y}(s, t) = 1$$

Further, note that if $g : \mathbb{R}^2 \to \mathbb{R}$ is a function, then LOTUS extends to the multidimensional case:

$$\mathbb{E}[g(X, Y)] = \sum_{x \in \Omega_X} \sum_{y \in \Omega_Y} g(x, y)p_{X,Y}(x, y)$$
Back to our example of the blue and red die rolls. Again, let $X$ be the value of the blue die and $Y$ be the value of the red die. Now, let $U = \min\{X, Y\}$ and $V = \max\{X, Y\}$. Then:

$$\Omega_U = \{1, 2, 3, 4\}$$
$$\Omega_V = \{1, 2, 3, 4\}$$

Because both random variables can take on any of the four values that appear on the dice. However:

$$\Omega_{U,V} = \{(u, v) \in \Omega_U \times \Omega_V : u \leq v\} \neq \Omega_U \times \Omega_V$$

The reason that these are not equivalent, are there are some pairs in the cartesian product that won’t appear in the joint distribution. The minimum will always be less than the maximum, so for example the case $(u, v) = (4, 1)$ will not occur.

This will just be all the ordered pairs of the values that can appear as $U$ and $V$. Now, however these are not equally likely, as shown in the table below:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1/16</td>
<td>2/16</td>
<td>2/16</td>
<td>2/16</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1/16</td>
<td>2/16</td>
<td>2/16</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>1/16</td>
<td>2/16</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1/16</td>
</tr>
</tbody>
</table>

As discussed earlier, we can’t have the case where $U > V$, so these are all 0. The cases where $U = V$ occurs when the blue and red die have the same value, each which occurs with probability of $\frac{1}{16}$ as shown earlier. The others in which $U < V$ each occur with probability $\frac{2}{16}$ because it could be the red die with the max and the blue die with the min, or the reverse.
So for the joint PMF \( p_{U,V}(u,v) = P(U = u, V = v) \) for \( u, v \in \Omega_{U,V} \) we have:

\[
p_{U,V}(u,v) = \begin{cases} 
\frac{2}{16}, & u, v \in \Omega_U \times \Omega_V, \quad v > u \\
\frac{1}{16}, & u, v \in \Omega_U \times \Omega_V, \quad v = u \\
0, & \text{otherwise}
\end{cases}
\]

Again, the piecewise function and the table are both valid ways to express the joint PMF, and you may choose whichever is easier for you.

### 5.1.3 Marginal PMFs

Suppose that we wanted to solve for the pmf \( p_U(u) \) for \( u \in \Omega_U \). Intuitively, how would you do it? For example, \( P(U = 1) \) would be the sum of the first row, since that is all the cases where \( U = 1 \). Mathematically, we have \( P(U = u) = \sum_v P(U = u, V = v) \). Does this look like anything we learned before? It’s just the law of total probability (intersection version) that we derived in 2.2, as the events \( \{V = v\}_{v \in \Omega_V} \) partition the sample space (\( V \) takes on exactly one value)! We can refer to the table above sum each row (which corresponds to a value of \( u \) to find the probability of that value of \( u \) occurring). That gives us the following:

\[
p_U(u) = \begin{cases} 
\frac{7}{16}, & u = 1 \\
\frac{5}{16}, & u = 2 \\
\frac{3}{16}, & u = 3 \\
\frac{1}{16}, & u = 4
\end{cases}
\]

This brings us to the definition of marginal PMFs.

**Definition 5.1.3.1: Marginal PMFs**

Let \( X, Y \) be discrete random variables. The marginal PMF of \( X \) is:

\[
p_X(a) = \sum_{b \in \Omega_Y} p_{X,Y}(a,b)
\]

Similarly, the marginal PMF of \( Y \) is:

\[
p_Y(d) = \sum_{c \in \Omega_X} p_{X,Y}(c,d)
\]

(Extension) If \( Z \) is also a discrete random variable, then the marginal PMF of \( z \) is:

\[
p_Z(z) = \sum_{x \in \Omega_X} \sum_{y \in \Omega_Y} p_{X,Y,Z}(x,y,z)
\]

This follows from the law of total probability, and is just like taking the sum of a row in the example above.

Now if asked for \( \mathbb{E}[U] \) for example, we actually don’t need the joint PMF anymore. We’ve extracted the pertinent information in the form of \( p_U(u) \), and compute \( \mathbb{E}[U] = \sum_u up_U(u) \) normally.
5.1.4 Independence

**Definition 5.1.4.1: Independence (DRVs)**

Discrete random variables \( X, Y \) are independent, written \( X \perp Y \), if for all \( x \in \Omega_X \) and \( y \in \Omega_Y \):

\[
p_{X,Y}(x, y) = p_X(x)p_Y(y)
\]

**Fact 5.1.4.1: Check for Independence (DRVs)**

Recall \( \Omega_{X,Y} = \{(x, y) : p_{X,Y}(x, y) > 0\} \subseteq \Omega_X \times \Omega_Y \). A necessary but not sufficient condition for independence is that \( \Omega_{X,Y} = \Omega_X \times \Omega_Y \). That is, if \( \Omega_{X,Y} \neq \Omega_X \times \Omega_Y \), then \( X \) and \( Y \) cannot be independent, but if \( \Omega_{X,Y} = \Omega_X \times \Omega_Y \), then we have to check the condition above.

This is because if there is some \( (a, b) \in \Omega_X \times \Omega_Y \) but not in \( \Omega_{X,Y} \), then \( p_{X,Y}(a, b) = 0 \) but \( p_X(a) > 0 \) and \( p_Y(b) > 0 \), violating independence. For example, suppose the joint PMF looks like:

<table>
<thead>
<tr>
<th>( X \setminus Y )</th>
<th>8</th>
<th>9</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1/3</td>
<td>1/2</td>
<td>5/6</td>
</tr>
<tr>
<td>7</td>
<td>1/6</td>
<td>0</td>
<td>1/6</td>
</tr>
<tr>
<td>Col Total</td>
<td>1/2</td>
<td>1/2</td>
<td>1</td>
</tr>
</tbody>
</table>

Also side note that the marginal distributions are named what they are, since we often write the row and column totals in the margins. The joint range \( \Omega_{X,Y} \neq \Omega_X \times \Omega_Y \) since one of the entries is 0, and so \( (7, 9) \notin \Omega_{X,Y} \) but \( (7, 9) \in \Omega_X \times \Omega_Y \). This immediately tells us they cannot be independent - \( p_X(7) > 0 \) and \( p_Y(9) > 0 \), yet \( p_{X,Y}(7, 9) = 0 \).

5.1.5 Variance Adds for Independent Random Variables

You may recall that we started earlier. We will finally prove this.

**Lemma 5.1.5.1: Variance Adds for Independent RVs**

If \( X, Y \) are independent random variables, denoted \( X \perp Y \), then:

\[
\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)
\]

If \( a, b, c \in R \) are scalars, then:

\[
\text{Var}(aX + bY + c) = a^2\text{Var}(X) + b^2\text{Var}(Y)
\]

Note this property relies on the fact that they are independent, whereas linearity of expectation always holds, regardless.

**Proof.** To prove this, we must first prove the following lemma:
Lemma 5.1.5.2: Expected Value of the Product of Independent Random Variables

If \( X \perp Y \), then \( E[XY] = E[X]E[Y] \).

Proof.

\[
E[XY] = \sum_{x \in \Omega_X} \sum_{y \in \Omega_Y} xyp_{X,Y}(x,y) \quad \text{[LOTUS]}
\]
\[
= \sum_{x \in \Omega_X} \sum_{y \in \Omega_Y} xyp_X(x)p_Y(y) \quad \text{[\( X \perp Y \)]}
\]
\[
= \sum_{x \in \Omega_X} xp_X(x) \sum_{y \in \Omega_Y} yp_Y(y)
\]
\[
= E[X]E[Y]
\]

Now we have the following:

\[
Var(X + Y) = E[(X + Y)^2] - (E[X + Y])^2 \quad \text{[def of variance]}
\]
\[
= E[X^2 + 2XY + Y^2] - (E[X] + E[Y])^2 \quad \text{[linearity of expectation]}
\]
\[
= E[X^2] + 2E[XY] + E[Y^2] - (E[X])^2 - 2E[X]E[Y] - (E[Y])^2 \quad \text{[linearity of expectation]}
\]
\[
\]
\[
= E[X^2] - (E[X])^2 + E[Y^2] - (E[Y])^2 + 2E[X]E[Y] - 2E[X]E[Y] \quad \text{[since \( X \perp Y \)]}
\]
\[
= Var(X) + Var(Y) \quad \text{[def of variance]}
\]

5.1.6 Reproducing Linearity of Expectation

Proof of Linearity of Expectation. Let \( X, Y \) be (possibly dependent) random variables. We'll prove that \( E[X + Y] = E[X] + E[Y] \).

\[
E[X + Y] = \sum_{x} \sum_{y} (x + y)p_{X,Y}(x,y) \quad \text{[LOTUS]}
\]
\[
= \sum_{x} \sum_{y} xp_{X,Y}(x,y) + \sum_{y} \sum_{x} yp_{X,Y}(x,y) \quad \text{[split sum]}
\]
\[
= \sum_{x} x \sum_{y} p_{X,Y}(x,y) + \sum_{y} y \sum_{x} p_{X,Y}(x,y) \quad \text{[algebra]}
\]
\[
= \sum_{x} xp_X(x) + \sum_{y} yp_Y(y) \quad \text{[def of marginal PMF]}
\]
\[
= E[X] + E[Y] \quad \text{[def of expectation]}
\]
### 5.1.7 Exercises

1. Suppose $X, Y$ are jointly distributed with joint PMF:

<table>
<thead>
<tr>
<th>$X \setminus Y$</th>
<th>6</th>
<th>9</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>3/12</td>
<td>5/12</td>
<td>?</td>
</tr>
<tr>
<td>2</td>
<td>1/12</td>
<td>2/12</td>
<td>?</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1/12</td>
<td>?</td>
</tr>
<tr>
<td>Col Total</td>
<td>?</td>
<td>?</td>
<td>1</td>
</tr>
</tbody>
</table>

(a) Find $p_X(x)$ and $p_Y(y)$.
(b) Find $E[Y]$.
(c) Are $X$ and $Y$ independent?
(d) Find $E[X^Y]$.

**Solution:**

(a) Actually these can be found by filling in the row and column totals, since

$$p_X(x) = \sum_y p_{X,Y}(x,y), \quad p_Y(y) = \sum_x p_{X,Y}(x,y)$$

For example, $p_X(0) = \sum_y p_{X,Y}(0,y) = p_{X,Y}(0,6) + p_{X,Y}(0,9) = 3/12 + 5/12 = 8/12$.

<table>
<thead>
<tr>
<th>$X \setminus Y$</th>
<th>6</th>
<th>9</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>3/12</td>
<td>5/12</td>
<td>8/12</td>
</tr>
<tr>
<td>2</td>
<td>1/12</td>
<td>2/12</td>
<td>3/12</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1/12</td>
<td>1/12</td>
</tr>
<tr>
<td>Col Total</td>
<td>4/12</td>
<td>8/12</td>
<td>1</td>
</tr>
</tbody>
</table>

Hence,

$$p_X(x) = \begin{cases} 
8/12 & x = 0 \\
3/12 & x = 2 \\
1/12 & x = 3 
\end{cases}$$

$$p_Y(y) = \begin{cases} 
4/12 & y = 6 \\
8/12 & y = 9 
\end{cases}$$

(b) We can actually compute $E[Y]$ just using $p_Y$ now that we’ve eliminated/marginalized out $X$ - we don’t need the joint PMF anymore. We go back to the definition:

$$E[Y] = \sum_y y p_Y(y) = 6 \cdot \frac{4}{12} + 9 \cdot \frac{8}{12} = 8$$

(c) $X, Y$ are independent, if for every table entry $(x, y)$, we have $p_{X,Y}(x,y) = p_X(x)p_Y(y)$. However, notice $p_{X,Y}(3,6) = 0$ but $p_X(3) > 0$ and $p_Y(6) > 0$. Hence we found an entry where this condition isn’t true, so they cannot be independent. This is like the comment mentioned earlier: if $\Omega_{X,Y} \neq \Omega_X \times \Omega_Y$, they have no chance of being independent.

(d) We use the LOTUS formula:

$$E[X^Y] = \sum_x \sum_y x^y p_{X,Y}(x,y) = 0^6 \cdot \frac{3}{12} + 0^9 \cdot \frac{5}{12} + 2^6 \cdot \frac{1}{12} + \ldots$$

This just sums over all the probabilities and takes a weighted average.
2. Suppose there are $N$ marbles in a bag, composed of $r$ different colors. Suppose there are $K_1$ of color 1, $K_2$ of color 2, ..., $K_r$ of color $r$, where $\sum_{i=1}^{r} K_i = N$. We reach in and draw $n$ without replacement. Let $(X_1, \ldots, X_r)$ be a random vector where $X_i$ is the count of how many marbles of color $i$ we drew. What is $p_{X_1, \ldots, X_r}(k_1, \ldots, k_r)$ for valid values of $k_1, \ldots, k_r$? We say the random vector $(X_1, \ldots, X_r) \sim MVHG(N, K_1, \ldots, K_r, n)$ has a multivariate hypergeometric distribution!

**Solution:**

$$p_{X_1, \ldots, X_r}(k_1, \ldots, k_r) = \frac{(K_1)_{k_1} \cdots (K_r)_{k_r}}{(N)_n} = \frac{\prod_{i=1}^{r} \binom{K_i}{k_i}}{\binom{N}{n}}$$