CHAPTER 1. ELEMENTS OF PROBABILITY DISTRIBUTION THEORY

1.9 Two-Dimensional Random Variables

Definition 1.14. Let $\Omega$ be a sample space and $X_1, X_2$ be functions, each assigning a real number $X_1(\omega), X_2(\omega)$ to every outcome $\omega \in \Omega$, that is $X_1: \Omega \rightarrow \mathcal{X}_1 \subset \mathbb{R}$ and $X_2: \Omega \rightarrow \mathcal{X}_2 \subset \mathbb{R}$. Then the pair $X = (X_1, X_2)$ is called a two-dimensional random variable. The induces sample space (range) of the two-dimensional random variable is

$$\mathcal{X} = \{(x_1, x_2) : x_1 \in \mathcal{X}_1, x_2 \in \mathcal{X}_2\} \subset \mathbb{R}^2.$$ 

We will denote two-dimensional random variables by bold capital letters.

Definition 1.15. The cumulative distribution function of a two-dimensional rv $X = (X_1, X_2)$ is

$$F_X(x_1, x_2) = P(X_1 \leq x_1, X_2 \leq x_2) \quad (1.10)$$

1.9.1 Discrete Two-Dimensional Random Variables

If all values of $X = (X_1, X_2)$ are countable, i.e., the values are in the range

$$\mathcal{X} = \{(x_{1i}, x_{2j}) : i = 1, 2, \ldots, j = 1, 2, \ldots\}$$

then the variable is discrete. The cdf of a discrete rv $X = (X_1, X_2)$ is

$$F_X(x_1, x_2) = \sum_{x_{2j} \leq x_2} \sum_{x_{1i} \leq x_1} p(x_{1i}, x_{2j})$$

where $p(x_{1i}, x_{2j})$ denotes the joint probability mass function and

$$p(x_{1i}, x_{2j}) = P(X_1 = x_{1i}, X_2 = x_{2j})$$

As in the univariate case, the joint pmf satisfies the following conditions.

1. $p(x_{1i}, x_{2j}) \geq 0$, for all $i, j$

2. $\sum_{x_{2j}} \sum_{x_{1i}} p(x_{1i}, x_{2j}) = 1$
Example 1.18. Consider an experiment of tossing two fair dice and noting the outcome on each die. The whole sample space consists of 36 elements, i.e.,

$$\Omega = \{(i, j) : i, j = 1, \ldots, 6\}.$$

Now, with each of these 36 elements associate values of two random variables, $X_1$ and $X_2$, such that

$$X_1 \equiv \text{sum of the outcomes on the two dice},$$

$$X_2 \equiv | \text{difference of the outcomes on the two dice} |.$$

That is,

$$X(i, j) = (i + i, |i - j|) \quad i, j = 1, 2, \ldots, 6.$$

Then, the bivariate rv $X = (X_1, X_2)$ has the following joint probability mass function (empty cells mean that the pmf is equal to zero at the relevant values of the rvs).

<table>
<thead>
<tr>
<th>$x_1$</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1/36</td>
<td>1/36</td>
<td>1/36</td>
<td>1/36</td>
<td>1/36</td>
<td>1/36</td>
<td>1/36</td>
<td>1/36</td>
<td>1/36</td>
<td>1/36</td>
<td>1/36</td>
</tr>
<tr>
<td>1</td>
<td>1/18</td>
<td>1/18</td>
<td>1/18</td>
<td>1/18</td>
<td>1/18</td>
<td>1/18</td>
<td>1/18</td>
<td>1/18</td>
<td>1/18</td>
<td>1/18</td>
<td>1/18</td>
</tr>
<tr>
<td>2</td>
<td>1/18</td>
<td>1/18</td>
<td>1/18</td>
<td>1/18</td>
<td>1/18</td>
<td>1/18</td>
<td>1/18</td>
<td>1/18</td>
<td>1/18</td>
<td>1/18</td>
<td>1/18</td>
</tr>
<tr>
<td>3</td>
<td>1/18</td>
<td>1/18</td>
<td>1/18</td>
<td>1/18</td>
<td>1/18</td>
<td>1/18</td>
<td>1/18</td>
<td>1/18</td>
<td>1/18</td>
<td>1/18</td>
<td>1/18</td>
</tr>
<tr>
<td>4</td>
<td>1/18</td>
<td>1/18</td>
<td>1/18</td>
<td>1/18</td>
<td>1/18</td>
<td>1/18</td>
<td>1/18</td>
<td>1/18</td>
<td>1/18</td>
<td>1/18</td>
<td>1/18</td>
</tr>
<tr>
<td>5</td>
<td>1/18</td>
<td>1/18</td>
<td>1/18</td>
<td>1/18</td>
<td>1/18</td>
<td>1/18</td>
<td>1/18</td>
<td>1/18</td>
<td>1/18</td>
<td>1/18</td>
<td>1/18</td>
</tr>
</tbody>
</table>

Expectations of functions of bivariate random variables are calculated the same way as of the univariate rvs. Let $g(x_1, x_2)$ be a real valued function defined on $\mathcal{X}$. Then $g(X) = g(X_1, X_2)$ is a rv and its expectation is

$$E[g(X)] = \sum_{(x_1, x_2)} g(x_1, x_2)p(x_1, x_2).$$

Example 1.19. Let $X_1$ and $X_2$ be random variables as defined in Example 1.18. Then, for $g(X_1, X_2) = X_1X_2$ we obtain

$$E[g(X)] = 2 \times 0 \times 1/36 + \ldots + 7 \times 5 \times 1/18 = 245/18.$$
Marginal pmfs

Each of the components of the two-dimensional rv is a random variable and so we may be interested in calculating its probabilities, for example \( P(X_1 = x_1) \). Such a uni-variate pmf is then derived in a context of the distribution of the other random variable. We call it the marginal pmf.

**Theorem 1.12.** Let \( X = (X_1, X_2) \) be a discrete bivariate random variable with joint pmf \( p(x_1, x_2) \). Then the marginal pmfs of \( X_1 \) and \( X_2 \), \( p_{X_1} \) and \( p_{X_2} \), are given respectively by

\[
p_{X_1}(x_1) = P(X_1 = x_1) = \sum_{x_2} p(x_1, x_2) \quad \text{and} \quad p_{X_2}(x_2) = P(X_2 = x_2) = \sum_{x_1} p(x_1, x_2).
\]

**Proof.** For \( X_1 \):

Let us denote by \( A_{x_1} = \{(x_1, x_2) : -\infty < x_2 < \infty\} \). Then, for any \( x_1 \in X_1 \) we may write

\[
P(X_1 = x_1) = P(X_1 = x_1, -\infty < x_2 < \infty) \\
= P((X_1, X_2) \subseteq A_{x_1}) \\
= \sum_{(x_1, x_2) \in A_{x_1}} P(X_1 = x_1, X_2 = x_2) \\
= \sum_{A_{x_1}} p(x_1, x_2).
\]

For \( X_2 \) the proof is similar. \( \square \)

**Example 1.20.** The marginal distributions of the variables \( X_1 \) and \( X_2 \) defined in Example 1.18 are following.

<table>
<thead>
<tr>
<th>( x_1 )</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P(X_1 = x_1) )</td>
<td>( \frac{1}{36} )</td>
<td>( \frac{1}{18} )</td>
<td>( \frac{1}{12} )</td>
<td>( \frac{1}{9} )</td>
<td>( \frac{5}{36} )</td>
<td>( \frac{1}{6} )</td>
<td>( \frac{5}{36} )</td>
<td>( \frac{1}{9} )</td>
<td>( \frac{1}{12} )</td>
<td>( \frac{1}{18} )</td>
<td>( \frac{1}{36} )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( x_2 )</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P(X_2 = x_2) )</td>
<td>( \frac{1}{6} )</td>
<td>( \frac{5}{18} )</td>
<td>( \frac{2}{9} )</td>
<td>( \frac{1}{6} )</td>
<td>( \frac{1}{9} )</td>
<td>( \frac{1}{18} )</td>
</tr>
</tbody>
</table>

**Exercise 1.14.** Students in a class of 100 were classified according to gender (\( G \)) and smoking (\( S \)) as follows:
where $s$, $q$ and $n$ denote the smoking status: “now smokes”, “did smoke but quit” and “never smoked”, respectively. Find the probability that a randomly selected student

1. is a male;
2. is a male smoker;
3. is either a smoker or did smoke but quit;
4. is a female who is a smoker or did smoke but quit.

### 1.9.2 Continuous Two-Dimensional Random Variables

If the values of $X = (X_1, X_2)$ are elements of an uncountable set in the Euclidean plane, then the variable is continuous. For example the values might be in the range

$$\mathcal{X} = \{(x_1, x_2) : a \leq x_1 \leq b, c \leq x_2 \leq d\}$$

for some real $a, b, c, d$.

The cdf of a continuous rv $X = (X_1, X_2)$ is defined as

$$F_{X}(x_1, x_2) = \int_{-\infty}^{x_2} \int_{-\infty}^{x_1} f(t_1, t_2) dt_1 dt_2, \quad (1.11)$$

where $f(x_1, x_2)$ is the probability density function such that

1. $f(x_1, x_2) \geq 0$ for all $(x_1, x_2) \in \mathbb{R}^2$
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1, x_2) dx_1 dx_2 = 1.$

The equation (1.11) implies that

$$\frac{\partial^2 F(x_1, x_2)}{\partial x_1 \partial x_2} = f(x_1, x_2). \quad (1.12)$$
Also

\[ P(a \leq X_1 \leq b, c \leq X_2 \leq d) = \int_a^b \int_c^d f(x_1, x_2) \, dx_1 \, dx_2. \]

The marginal pdfs of \(X_1\) and \(X_2\) are defined similarly as in the discrete case, here using integrals.

\[ f_{X_1}(x_1) = \int_{-\infty}^{\infty} f(x_1, x_2) \, dx_2, \quad \text{for} \quad -\infty < x_1 < \infty, \]

\[ f_{X_2}(x_2) = \int_{-\infty}^{\infty} f(x_1, x_2) \, dx_1, \quad \text{for} \quad -\infty < x_2 < \infty. \]

**Example 1.21.** Calculate \(P(X \subseteq A)\), where \(A = \{(x_1, x_2) : x_1 + x_2 \geq 1\}\) and the joint pdf of \(X = (X_1, X_2)\) is defined by

\[ f_X(x_1, x_2) = \begin{cases} 
6x_1x_2^2 & \text{for } 0 < x_1 < 1, \ 0 < x_2 < 1, \\
0 & \text{otherwise.}
\end{cases} \]

The probability is a double integral of the pdf over the region \(A\). The region is however limited by the domain in which the pdf is positive.

We can write

\[ A = \{(x_1, x_2) : x_1 + x_2 \geq 1, \ 0 < x_1 < 1, \ 0 < x_2 < 1\} \]

\[ = \{(x_1, x_2) : x_1 \geq 1 - x_2, \ 0 < x_1 < 1, \ 0 < x_2 < 1\} \]

\[ = \{(x_1, x_2) : 1 - x_2 < x_1 < 1, \ 0 < x_2 < 1\}. \]

Hence, the probability is

\[ P(X \subseteq A) = \int \int_A f(x_1, x_2) \, dx_1 \, dx_2 = \int_0^1 \int_{1-x_2}^1 6x_1x_2^2 \, dx_1 \, dx_2 = 0.9 \]

Also, we can calculate marginal pdfs.

\[ f_{X_1}(x_1) = \int_0^1 6x_1x_2^2 \, dx_2 = 2x_1x_2^3 \bigg|_0^1 = 2x_1, \]

\[ f_{X_2}(x_2) = \int_0^1 6x_1x_2^2 \, dx_1 = 3x_1x_2^2 \bigg|_0^1 = 3x_2^2. \]

These functions allow us to calculate probabilities involving only one variable. For example

\[ P \left( \frac{1}{4} < X_1 < \frac{1}{2} \right) = \int_{\frac{1}{4}}^{\frac{1}{2}} 2x_1 \, dx_1 = \frac{3}{16}. \]
Similarly to the case of a univariate rv the following linear property for the expectation holds.

\[ E[a g(X) + b h(X) + c] = a E[g(X)] + b E[h(X)] + c, \]  

where \( a, b \) and \( c \) are constants and \( g \) and \( h \) are some functions of the bivariate rv \( X = (X_1, X_2) \).

### 1.9.3 Conditional Distributions and Independence

**Definition 1.16.** Let \( X = (X_1, X_2) \) denote a discrete bivariate rv with joint pmf \( p_X(x_1, x_2) \) and marginal pmfs \( p_{X_1}(x_1) \) and \( p_{X_2}(x_2) \). For any \( x_1 \) such that \( p_{X_1}(x_1) > 0 \), the conditional pmf of \( X_2 \) given that \( X_1 = x_1 \) is the function of \( x_2 \) defined by

\[ p_{X_2}(x_2|x_1) = \frac{p_X(x_1, x_2)}{p_{X_1}(x_1)}. \]

Analogously, we define the conditional pmf of \( X_1 \) given \( X_2 = x_2 \)

\[ p_{X_1}(x_1|x_2) = \frac{p_X(x_1, x_2)}{p_{X_2}(x_2)}. \]

It is easy to check that these functions are indeed pdfs. For example,

\[ \sum_{X_2} p_{X_2}(x_2|x_1) = \sum_{X_2} \frac{p_X(x_1, x_2)}{p_{X_1}(x_1)} = \sum_{X_2} p_X(x_1, x_2) / p_{X_1}(x_1) = p_{X_1}(x_1) = 1. \]

**Example 1.22.** Let \( X_1 \) and \( X_2 \) be defined as in Example 1.18. The conditional pmf of \( X_2 \) given \( X_1 = 5 \), is

<table>
<thead>
<tr>
<th>( x_2 )</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p_{X_2}(x_2</td>
<td>5) )</td>
<td>0</td>
<td>( \frac{1}{2} )</td>
<td>0</td>
<td>( \frac{1}{2} )</td>
<td>0</td>
</tr>
</tbody>
</table>

**Exercise 1.15.** Let \( S \) and \( G \) denote the smoking status and gender as defined in Exercise 1.14. Calculate the probability that a randomly selected student is

1. a smoker given that he is a male;
2. female, given that the student smokes.
Analogously to the conditional distribution for discrete rvs, we define the conditional distribution for continuous rvs.

**Definition 1.17.** Let \( \mathbf{X} = (X_1, X_2) \) denote a continuous bivariate rv with joint pdf \( f_X(x_1, x_2) \) and marginal pdfs \( f_{X_1}(x_1) \) and \( f_{X_2}(x_2) \). For any \( x_1 \) such that \( f_{X_1}(x_1) > 0 \), the conditional pdf of \( X_2 \) given that \( X_1 = x_1 \) is the function of \( x_2 \) defined by

\[
f_{X_2}(x_2|x_1) = \frac{f_X(x_1, x_2)}{f_{X_1}(x_1)}.
\]

Analogously, we define the conditional p.d.f. of \( X_1 \) given \( X_2 = x_2 \)

\[
f_{X_1}(x_1|x_2) = \frac{f_X(x_1, x_2)}{f_{X_2}(x_2)}.
\]

Here too, it is easy to verify that these functions are pdfs. For example,

\[
\int_{X_2} f_{X_2}(x_2|x_1) dx_2 = \int_{X_2} \frac{f_X(x_1, x_2)}{f_{X_1}(x_1)} dx_2
\]

\[
= \int_{X_2} f_X(x_1, x_2) dx_2
\]

\[
= f_{X_1}(x_1) = 1.
\]

**Example 1.23.** For the random variables defined in Example 1.21 the conditional pdfs are

\[
f_{X_1}(x_1|x_2) = \frac{f_X(x_1, x_2)}{f_{X_2}(x_2)} = \frac{6x_1x_2^2}{3x_2^2} = 2x_1
\]

and

\[
f_{X_2}(x_2|x_1) = \frac{f_X(x_1, x_2)}{f_{X_1}(x_1)} = \frac{6x_1x_2^2}{2x_1} = 3x_2.
\]

The conditional pdfs allow us to calculate conditional expectations. The conditional expected value of a function \( g(X_2) \) given that \( X_1 = x_1 \) is defined by

\[
E[g(X_2)|x_1] = \left\{ \begin{array}{ll}
\sum_{x_2} g(x_2)p_{X_2}(x_2|x_1) & \text{for a discrete r.v.,} \\
\int_{X_2} g(x_2)f_{X_2}(x_2|x_1)dx_2 & \text{for a continuous r.v.}
\end{array} \right.
\]

(1.14)
Example 1.24. The conditional mean and variance of the $X_2$ given a value of $X_1$, for the variables defined in Example 1.21 are

$$\mu_{X_2|x_1} = \mathbb{E}(X_2|x_1) = \int_0^1 x_2^3 x_2^2 dx_2 = \frac{3}{4},$$

and

$$\sigma^2_{X_2|x_1} = \text{var}(X_2|x_1) = \mathbb{E}(X_2^2|x_1) - [\mathbb{E}(X_2|x_1)]^2 = \int_0^1 x_2^4 x_2^2 dx_2 - \left(\frac{3}{4}\right)^2 = \frac{3}{80}.$$ 

Definition 1.18. Let $X = (X_1, X_2)$ denote a continuous bivariate rv with joint pdf $f_X(x_1, x_2)$ and marginal pdfs $f_{X_1}(x_1)$ and $f_{X_2}(x_2)$. Then $X_1$ and $X_2$ are called independent random variables if, for every $x_1 \in X_1$ and $x_2 \in X_2$

$$f_X(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2). \quad (1.15)$$

We define independent discrete random variables analogously.

If $X_1$ and $X_2$ are independent, then the conditional pdf of $X_2$ given $X_1 = x_1$ is

$$f_{X_2}(x_2|x_1) = \frac{f_X(x_1, x_2)}{f_{X_1}(x_1)} = \frac{f_{X_1}(x_1) f_{X_2}(x_2)}{f_{X_1}(x_1)} = f_{X_2}(x_2)$$

regardless of the value of $x_1$. Analogous property holds for the conditional pdf of $X_1$ given $X_2 = x_2$.

Example 1.25. It is easy to notice that for the variables defined in Example 1.21 we have

$$f_X(x_1, x_2) = 6x_1 x_2^2 = 2x_1 3x_2^2 = f_{X_1}(x_1) f_{X_2}(x_2).$$

So, the variables $X_1$ and $X_2$ are independent. 

In fact, two rvs are independent if and only if there exist functions $g(x_1)$ and $h(x_2)$ such that for every $x_1 \in X_1$ and $x_2 \in X_2$,

$$f_X(x_1, x_2) = g(x_1) h(x_2).$$
CHAPTER 1. ELEMENTS OF PROBABILITY DISTRIBUTION THEORY

Theorem 1.13. Let $X_1$ and $X_2$ be independent random variables. Then

1. For any $A \subset \mathbb{R}$ and $B \subset \mathbb{R}$

   $$P(X_1 \subseteq A, X_2 \subseteq B) = P(X_1 \subseteq A)P(X_2 \subseteq B),$$

   that is, $\{X_1 \subseteq A\}$ and $\{X_2 \subseteq B\}$ are independent events.

2. For $g(X_1)$, a function of $X_1$ only, and for $h(X_2)$, a function of $X_2$ only, we have

   $$E[g(X_1)h(X_2)] = E[g(X_1)]E[h(X_2)].$$

Proof. Assume that $X_1$ and $X_2$ are continuous random variables. To prove the theorem for discrete rvs we follow the same steps with sums instead of integrals.

1. We have

   $$P(X_1 \subseteq A, X_2 \subseteq B) = \int_B \int_A f_{X_1}(x_1, x_2)dx_1dx_2 = \int_B \int_A f_{X_1}(x_1)f_{X_2}(x_2)dx_1dx_2 = \int_B \left( \int_A f_{X_1}(x_1)dx_1 \right) f_{X_2}(x_2)dx_2 = \int_A f_{X_1}(x_1)dx_1 \int_B f_{X_2}(x_2)dx_2 = P(X_1 \subseteq A)P(X_2 \subseteq B).$$

2. Similar arguments as in Part 1 give

   $$E[g(X_1)h(X_2)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x_1)h(x_2)f_{X_1}(x_1, x_2)dx_1dx_2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x_1)h(x_2)f_{X_1}(x_1)f_{X_2}(x_2)dx_1dx_2 = \int_{-\infty}^{\infty} \left( \int_{-\infty}^{\infty} g(x_1)f_{X_1}(x_1)dx_1 \right) h(x_2)f_{X_2}(x_2)dx_2 = \left( \int_{-\infty}^{\infty} g(x_1)f_{X_1}(x_1)dx_1 \right) \left( \int_{-\infty}^{\infty} h(x_2)f_{X_2}(x_2)dx_2 \right) = E[g(X_1)]E[h(X_2)].$$

In the following theorem we will apply this result for the moment generating function of a sum of independent random variables.
Theorem 1.14. Let $X_1$ and $X_2$ be independent random variables with moment generating functions $M_{X_1}(t)$ and $M_{X_2}(t)$, respectively. Then the moment generating function of the sum $Y = X_1 + X_2$ is given by

$$M_Y(t) = M_{X_1}(t)M_{X_2}(t).$$

Proof. By the definition of the mgf and by Theorem 1.13, part 2, we have

$$M_Y(t) = E[e^{tY}] = E[e^{t(X_1+X_2)}] = E(e^{tX_1})E(e^{tX_2}) = M_{X_1}(t)M_{X_2}(t).$$

Note that this result can be easily extended to a sum of any number of mutually independent random variables.

Example 1.26. Let $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$. What is the distribution of $Y = X_1 + X_2$?

Using Theorem 1.14 we can write

$$M_Y(t) = M_{X_1}(t)M_{X_2}(t) \exp\{\mu_1 t + \sigma_1^2 t^2/2\} \exp\{\mu_2 t + \sigma_2^2 t^2/2\}.$$

This is the mgf of a normal rv with $E(Y) = \mu_1 + \mu_2$ and $\text{var}(Y) = \sigma_1^2 + \sigma_2^2$.

Exercise 1.16. A part of an electronic system has two types of components in joint operation. Denote by $X_1$ and $X_2$ the random length of life (measured in hundreds of hours) of component of type I and of type II, respectively. Assume that the joint density function of two rvs is given by

$$f_X(x_1, x_2) = \frac{1}{8} x_1 \exp\left\{-\frac{x_1 + x_2}{2}\right\} I_X,$$

where $X = \{(x_1, x_2) : x_1 > 0, x_2 > 0\}$.

1. Calculate the probability that both components will have a life length longer than 100 hours, that is, find $P(X_1 > 1, X_2 > 1)$.

2. Calculate the probability that a component of type II will have a life length longer than 200 hours, that is, find $P(X_2 > 2)$.

3. Are $X_1$ and $X_2$ independent? Justify your answer.

4. Calculate the expected value of so called relative efficiency of the two components, which is expressed by

$$E\left(\frac{X_2}{X_1}\right).$$