1.9 Functions of Random Variables

If $X$ is a random variable with cdf $F_X(x)$, then any function of $X$, say $g(X) = Y$ is also a random variable. The question then is “what is the distribution of $Y$?”

The function $y = g(x)$ is a mapping from the induced sample space $\mathcal{X}$ of the random variable $X$ to a new sample space, $\mathcal{Y}$, of the random variable $Y$, that is $g(x): \mathcal{X} \to \mathcal{Y}$.

The inverse mapping $g^{-1}$ acts from $\mathcal{Y}$ to $\mathcal{X}$ and we can write

$$g^{-1}(A) = \{x \in \mathcal{X} : g(x) \in A\}$$ where $A \subset \mathcal{Y}$.

Then, we have

$$P(Y \in A) = P(g(X) \in A) = P(\{x \in \mathcal{X} : g(x) \in A\}) = P(X \in g^{-1}(A)).$$

The following theorem relates the cumulative distribution functions of $X$ and $Y = g(X)$.

**Theorem 1.10.** Let $X$ have cdf $F_X(x)$, $Y = g(X)$ and let domain and codomain of $g(X)$, respectively, be

$$\mathcal{X} = \{x : f_X(x) > 0\}, \quad \text{and} \quad \mathcal{Y} = \{y : y = g(x) \text{ for some } x \in \mathcal{X}\}.$$

(a) If $g$ is an increasing function on $\mathcal{X}$ then $F_Y(y) = F_X(g^{-1}(y))$ for $y \in \mathcal{Y}$.

(b) If $g$ is a decreasing function on $\mathcal{X}$, then $F_Y(y) = 1 - F_X(g^{-1}(y))$ for $y \in \mathcal{Y}$.

**Proof.** The cdf of $Y = g(X)$ can be written as

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P(\{x \in \mathcal{X} : g(x) \leq y\}) = \int_{\{x \in \mathcal{X} : g(x) \leq y\}} f_X(x) dx.$$

(a) If $g$ is increasing, then

$$\{x \in \mathcal{X} : g(x) \leq y\} = \{x \in \mathcal{X} : x \leq g^{-1}(y)\}.$$
So, we can write
\[
F_Y(y) = \int_{\{x \in \mathcal{X} : x \leq g^{-1}(y)\}} f_X(x) \, dx \\
= \int_{-\infty}^{g^{-1}(y)} f_X(x) \, dx \\
= F_X(g^{-1}(y)).
\]

(b) Now, if \( g \) is decreasing, then
\[
\{x \in \mathcal{X} : g(x) \leq y\} = \{x \in \mathcal{X} : x \geq g^{-1}(y)\}.
\]
So, we can write
\[
F_Y(y) = \int_{\{x \in \mathcal{X} : x \geq g^{-1}(y)\}} f_X(x) \, dx \\
= \int_{g^{-1}(y)}^{\infty} f_X(x) \, dx \\
= 1 - F_X(g^{-1}(y)).
\]

Example 1.16. Find the distribution of \( Y = g(X) = -\log X \), where \( X \sim \mathcal{U}(0, 1) \). The cdf of \( X \) is
\[
F_X(x) = \begin{cases} 
0, & \text{for } x < 0; \\
x, & \text{for } 0 \leq x \leq 1; \\
1, & \text{for } x > 1.
\end{cases}
\]

For \( x \in (0, 1) \) the function \( g(x) = -\log x \) is defined on \( \mathcal{Y} = (0, \infty) \) and it is decreasing.

For \( y > 0 \), \( y = -\log x \) implies that \( x = e^{-y} \), i.e., \( g^{-1}(y) = e^{-y} \) and
\[
F_Y(y) = 1 - F_X(g^{-1}(y)) = 1 - F_X(e^{-y}) = 1 - e^{-y}.
\]
Hence we may write
\[
F_Y(y) = (1 - e^{-y}) I_{(0, \infty)}.
\]
This is exponential distribution function for \( \lambda = 1 \).

For continuous rvs we have the following result.
CHAPTER 1. ELEMENTS OF PROBABILITY DISTRIBUTION THEORY

Theorem 1.11. Let $X$ have pdf $f_X(x)$ and let $Y = g(X)$, where $g$ is a monotone function. Suppose that $f_X(x)$ is continuous on its support $X = \{x : f_X(x) > 0\}$ and that $g^{-1}(y)$ has a continuous derivative on support $Y = \{y : y = g(x) \text{ for some } x \in X\}$. Then the pdf of $Y$ is given by

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| I_Y.$$

Proof.

$$f_Y(y) = \frac{d}{dy} F_Y(y)$$

$$= \begin{cases} \frac{d}{dy} \{ F_X(g^{-1}(y)) \}, & \text{if } g \text{ is increasing;} \\ \frac{d}{dy} \{ 1 - F_X(g^{-1}(y)) \}, & \text{if } g \text{ is decreasing.} \end{cases}$$

$$= \begin{cases} f_X(g^{-1}(y)) \frac{d}{dy} g^{-1}(y), & \text{if } g \text{ is increasing;} \\ -f_X(g^{-1}(y)) \frac{d}{dy} g^{-1}(y), & \text{if } g \text{ is decreasing.} \end{cases}$$

which gives the thesis of the theorem. \qed

Example 1.17. Suppose that $Z \sim \mathcal{N}(0, 1)$. What is the distribution of $Y = Z^2$?

For $Y > 0$, the cdf of $Y = Z^2$ is

$$F_Y(y) = P(Y \leq y)$$

$$= P(Z^2 \leq y)$$

$$= P(-\sqrt{y} \leq Z \leq \sqrt{y})$$

$$= F_Z(\sqrt{y}) - F_Z(-\sqrt{y}).$$

The pdf can now be obtained by differentiation:

$$f_Y(y) = \frac{d}{dy} F_Y(y)$$

$$= \frac{d}{dy} \left( F_Z(\sqrt{y}) - F_Z(-\sqrt{y}) \right)$$

$$= \frac{1}{2\sqrt{y}} f_Z(\sqrt{y}) + \frac{1}{2\sqrt{y}} f_Z(-\sqrt{y})$$

Now, for the standard normal distribution we have

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad -\infty < z < \infty.$$
1.9. FUNCTIONS OF RANDOM VARIABLES

This gives,

\[
f_Y(y) = \frac{1}{2\sqrt{y}} \left[ \frac{1}{\sqrt{2\pi}} e^{-\sqrt{y}^2/2} + \frac{1}{\sqrt{2\pi}} e^{\sqrt{y}^2/2} \right] \\
= \frac{1}{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-y/2}, \quad 0 < y < \infty.
\]

This is a chi squared random variable with one degree of freedom, \( \chi_1^2 \).

Note that \( g(Z) = Z^2 \) is not a monotone function, but the range of \( Z, (-\infty, \infty) \), can be partitioned so that it is monotone on its sub-sets.

\[ \square \]

Exercise 1.11. Suppose that \( Z \sim \mathcal{N}(0, 1) \). Find the distribution of \( Y = \mu + \sigma Z \) for constant \( \mu \in \mathbb{R} \) and \( \sigma \in \mathbb{R}_+ \).

Exercise 1.12. Let \( X \) be a random variable with moment generating function \( M_X \).

(i) Show that the moment generating function of \( Y = a + bX \), where \( a \) and \( b \) are constants, is given by

\[ M_Y(t) = e^{at} M_X(tb). \]

(ii) Derive the moment generating function of \( Y \sim \mathcal{N}(\mu, \sigma^2) \). Hint: First find \( M_Z(t) \) for a standard normal rv \( Z \).
1.10 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 $X_1 : \Omega \rightarrow X_1 \subset \mathbb{R}$ and $X_2 : \Omega \rightarrow X_2 \subset \mathbb{R}$. Then the pair $X = (X_1, X_2)$ is called a two-dimensional random variable. The induced sample space (range) of the two-dimensional random variable is

$$X = \{(x_1, x_2) : x_1 \in X_1, \ x_2 \in X_2\} \subseteq \mathbb{R}^2.$$ 

We will denote two-dimensional (bi-variate) random variables by bold capital letters.

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

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

1.10.1 Discrete Two-Dimensional Random Variables

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

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

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

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

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

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

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

1. $p_X(x_{1i}, x_{2j}) \geq 0$, for all $i, j$
1.10. TWO-DIMENSIONAL RANDOM VARIABLES

2. \[ \sum_{x_2} \sum_{x_1} p_X(x_{1i}, x_{2j}) = 1 \]

*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.,

\[ \Omega = \{ \omega_{ij} = (i, j) : i, j = 1, \ldots, 6 \} \].

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,

\[ X(\omega_{ij}) = (X_1(\omega_{ij}), X_2(\omega_{ij})) = (i + j, |i - j|) \quad i, j = 1, 2, \ldots, 6. \]

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>( x_2 )</td>
<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>
</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></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></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></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></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></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

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

$$E[g(X)] = 2 \times 0 \times \frac{1}{36} + \ldots + 7 \times 5 \times \frac{1}{18} = \frac{245}{18}.$$ 

□

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

$$p_{X_1}(x_1) = P(X_1 = x_1) = \sum_{x_2} p_{\mathbf{X}}(x_1, x_2) \quad \text{and}$$

$$p_{X_2}(x_2) = P(X_2 = x_2) = \sum_{x_1} p_{\mathbf{X}}(x_1, x_2).$$

**Proof.** For $X_1$:

Let us denote by $A_{x_1} = \{(x_1, x_2) : x_2 \in X_2\}$. Then, for any $x_1 \in X_1$ we may write

$$P(X_1 = x_1) = P(X_1 = x_1, x_2 \in X_2)$$

$$= P((X_1, X_2) \subseteq A_{x_1})$$

$$= \sum_{(x_1, x_2) \in A_{x_1}} P(X_1 = x_1, X_2 = x_2)$$

$$= \sum_{x_2} p_{\mathbf{X}}(x_1, x_2).$$

For $X_2$ the proof is similar. □

Example 1.20. The marginal distributions of the variables $X_1$ and $X_2$ defined in Example 1.18 are following.
Exercise 1.13. Students in a class of 100 were classified according to gender (G) and smoking (S) as follows:

<table>
<thead>
<tr>
<th></th>
<th>s</th>
<th>q</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>male</td>
<td>20</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>female</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>37</td>
<td>33</td>
</tr>
</tbody>
</table>

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.
CHAPTER 1. ELEMENTS OF PROBABILITY DISTRIBUTION THEORY

1.10.2 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 jointly 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

\[
F_X(x_1, x_2) = \int_{-\infty}^{x_2} \int_{-\infty}^{x_1} f_X(t_1, t_2) dt_1 dt_2, \tag{1.11}
\]

where \( f_X(x_1, x_2) \) is the joint probability density function such that

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

The equation (1.11) implies that

\[
\frac{\partial^2 f_X(x_1, x_2)}{\partial x_1 \partial x_2} = f_X(x_1, x_2). \tag{1.12}
\]

Also

\[
P(a \leq X_1 \leq b, c \leq X_2 \leq d) = \int_{c}^{d} \int_{a}^{b} f_X(x_1, x_2) dx_1 dx_2.
\]

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

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

\[
f_{X_2}(x_2) = \int_{-\infty}^{\infty} f_X(x_1, x_2) dx_1, \quad \text{for} \quad -\infty < x_2 < \infty.
\]
1.10. TWO-DIMENSIONAL RANDOM VARIABLES

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

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

Also, we can calculate marginal pdfs.

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

$$f_{X_2}(x_2) = \int_0^1 6x_1x_2^2 dx_1 = 3x_1^2x_2^2 \bigg|_0^1 = 3x_2^2.$$ 

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

$$P \left( \frac{1}{4} < X_1 < \frac{1}{2} \right) = \int_{\frac{1}{4}}^{\frac{1}{2}} 2x_1 dx_1 = \frac{3}{16}. \quad \square$$

Analogously to the discrete case, the expectation of a function $g(X)$ is given by

$$E[g(X)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(X) f_X(x_1, x_2) dx_1 dx_2.$$

Similarly as in the case of univariate rvs the following linear property for the expectation holds for bi-variate rvs.

$$E[ag(X) + bh(X) + c] = aE[g(X)] + bE[h(X)] + c,$$  \hspace{1cm} (1.13)

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