3.4 Bivariate Distribution

Definition 3.4.1 Suppose that $X$ and $Y$ are random variables. The joint distribution, or bivariate distribution of $X$ and $Y$ is the collection of all probabilities of the form $\Pr((X, Y) \in C)$ for all sets $C \subseteq \mathbb{R}^2$ such that $\{(X, Y) \in C\}$ is an event.

Recall the utility example from section 3.3, where $X =$ water demand and $Y =$ electricity demand. The support of $X$ and $Y$ was defined to be $[4, 200]$ and $[1, 150]$ respectively. Based on the previous discussion of the utility example, $(X, Y)$ has a joint distribution function, and the distribution is uniform on $[4, 200] \times [1, 150]$. Since the integral of $f$ over $[4, 200] \times [1, 150]$ is equal to 1,

$$1 = \int_4^{200} \int_1^{150} c \, dy \, dx \Rightarrow c = \frac{1}{29204}.$$  

Let $A = \{X \geq 115\}$, $B = \{Y \geq 110\}$, and $C = A \cap B$, so that $C$ is the event that water demand is greater than 110 and electric demand is greater than 115. Then,

$$\Pr(X \in C) = \int_C f(x, y) \, dxdy = \int_{110}^{150} \int_{115}^{200} \frac{1}{29204} \, dxdy = .1198.$$  

For convenience, let $c = 1/29204$. The probability of the event $\{(X, Y) \in A \cup B\} = D$ is

$$\Pr(X \in D) = \int_{110}^{150} \int_{115}^{200} c \, dxdy + \int_{110}^{200} \int_{115}^{110} c \, dxdy + \int_{150}^{200} \int_{115}^{110} c \, dxdy = .1469.$$  

The second and third integrals are computing $\Pr(X \in A^c \cap B)$ and $\Pr(X \in A \cap B^c)$, respectively.

Discrete joint distributions

Suppose that $X$ and $Y$ are random variables. If there exist countably many possible values that $(X, Y)$ may take on, then $X$ and $Y$ have a discrete joint distribution.

Theorem 3.4.1 Suppose that $X$ and $Y$ have discrete distributions. Then $(X, Y)$ has a discrete joint distribution. This result follows from the fact that the distributions of both $X$ and $Y$ have countably many points, say $S_x$ and $S_y$ respectively. Consequently, $S_x \times S_y$ is countable, and the possible number of points that $(X, Y)$ can take on is countable.

Definition 3.4.3 The joint probability function (p.f.) of the discrete r.v.s $X$ and $Y$ is the function $f$ such that for every $(x, y) \in \mathbb{R}^2$,

$$f(x, y) = \Pr(X = x \text{ and } Y = y).$$
The notation \( \Pr(X = x, Y = y) \) is often used instead of \( \Pr(X = x \text{ and } Y = y) \). Consequences of this definition are

1. If \((x, y)\) is not in the support of \((X, Y)\), then \(f(x, y) = 0\).
2. \(\sum_{x \in \mathbb{R}} \sum_{y \in \mathbb{R}} f(x, y) = 1\).
3. For any set of ordered pairs \(C\),
   \[
   \Pr[(X, Y) \in C] = \sum_{(x, y) \in C} f(x, y).
   \]

**Example** Suppose that an urn contains 4 green, 6 red and 10 black balls. Two balls are drawn randomly and without replacement. Let \(X\) count the number of green balls and \(Y\) count the number of reds. The (joint) probability distribution for \((X, Y)\) is determined by counting the number of ways to draw \(0 \leq x \leq 2\) from 4 (reds) and \(0 \leq y \leq 2\) from 6 (greens) and the number of ways to draw \(2 - x - y\) from 10, given that \(x + y \leq 2\). The probability distribution function for \((X, Y)\) is

\[
\Pr(X = x, Y = y) = \begin{cases} \frac{\binom{4}{x}\binom{6}{y}\binom{10}{2-x-y}}{\binom{20}{2}}, & x, y \in \{0, 1, 2\} \text{ and } x + y \leq 2, \\ 0, & \text{otherwise.} \end{cases}
\]  

(1)

The joint probability function is

\[
f(x, y) = \begin{cases} \frac{\binom{4}{x}\binom{6}{y}\binom{10}{2-x-y}}{\binom{20}{2}}, & x, y \in \{0, 1, 2\} \text{ and } x + y \leq 2, \\ 0, & \text{otherwise.} \end{cases}
\]  

(2)

Sometimes, it’s sensible to enumerate the probabilities in a table, as illustrated below.

**Table 1:** The joint probability function for the number of red \((X)\) and green \((Y)\) balls from the urn example.

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>.237</td>
<td>.316</td>
<td>.079</td>
</tr>
<tr>
<td>1</td>
<td>.210</td>
<td>.126</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>.032</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

In R, a function was defined to compute the probabilities above:

```r
prob <- function(x,y) { choose(4,x)*choose(6,y)*choose(10,2-x-y)/choose(20,2)}.
```

The function is called by typing `prob(2,0)` in the console (or in a script).
Suppose that \( C = \{ X + Y = 2 \} \). Then \( \Pr[(X, Y) \in C] = f(0, 2) + f(1, 1) + f(2, 0) = .237 \).

**Continuous Joint Distributions**

**Definition 3.4.4.** The random variables \( X \) and \( Y \) have a *continuous joint distribution* if there exists \( f \geq 0 \) defined on \( \mathbb{R}^2 \) such that for every set \( C \subset \mathbb{R}^2 \),

\[
\Pr[(X, Y) \in C] = \int_C f(x, y) \, dx \, dy.
\]

The closure of the set \( \{(x, y) \mid f(x, y) \geq 0\} \) is called the support of \( (X, Y) \), and \( f \) is called the *joint probability distribution function* (p.d.f.).

A joint p.d.f. must satisfy:

1. \( f(x, y) \geq 0 \) for all \( (x, y) \in \mathbb{R}^2 \).
2. \( \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \, dx \, dy = 1 \).

The second condition states that the volume between the surface defined by \( f \) and the Cartesian plane is 1.

**Example** Suppose that

\[
f(x, y) = \begin{cases} 
  cx^2 y, & 0 \leq x^2 \leq y < 1, \\
  0 & \text{otherwise.}
\end{cases}
\]

The support of \( f \) is graphed below and to the left. The lower boundary of the support is the graph of the equation \( y = x^2 \) and the upper boundary is the graph of the line \( y = 1 \). The boundaries are determined by from the constraints \( 0 \leq x^2 \leq y < 1 \).

To determine the value of \( c \), \( f(x, y) \) must be integrated and the equation \( \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} cf(x, y) \, dx \, dy = 1 \) solved for \( c \). We integrate over the support (shown below and left in red).
The integral is

\[ 1 = \int_0^1 \int_{-\sqrt{y}}^{\sqrt{y}} cx^2 y \, dx \, dy \]

\[ = \frac{c}{3} \int_0^1 x^3 y\sqrt{y} \, dy \]

\[ = \frac{2c}{3} \int_0^1 y^{5/2} \, dy = \frac{4c}{27} \Rightarrow c = \frac{27}{4}. \]

DeGroot and Schervish also compute \( c \) by integrating over the \( y \)-variable first. Consequently, their limits of integration are different than those above.

**Example** A common problem\(^1\) is to compute \( \Pr(X \geq Y) \), or some variant such as \( \Pr(X < Y) \). The best approach to solving these problems begins with a sketch of the region of integration. For the current example, computing \( \Pr(X \geq Y) \) requires integration over the region shown in red in the Figure above and right. The region of integration is determined by finding the upper and lower boundaries for \( y \) induced by the constraint that \( y \leq x \) (upper boundary from \( X \geq Y \)) and \( y \leq x^2 \) (lower boundary from the support). We integrate over \( y \) (with \( x \) in the limits of integration), and then integrate over \( x \) over its range \([0, 1]\).

R code for drawing the figures is

\[ x = \text{seq}(\text{from} = -1, \text{to} = 1, \text{by} = .001) \]

\(^1\)particularly on exams
n = length(x)
y = x^2
plot( x, y, type = 'l', xlab = 'x', ylab = 'y')
for (i in 1:n/2) {
    lines(x = c(x[i], -x[i]), y = c(y[i], y[i]), col = 'red')
}
abline(h = 0)
abline(v = 0)

Returning to the computation of Pr(\(X \geq Y\)), the lower limit of integration for \(y\) is \(x^2\) (according to the definition of the joint p.d.f.). The upper limit is \(x\) since the maximum value of \(y\) that is in the region of interest, according to \(Y < X\), is \(x\). Hence,

\[
\Pr(X \geq Y) = \frac{21}{4} \int_0^1 \int_{x^2}^x x^2 y \, dy \, dx
\]

\[
= \frac{21}{4} \int_0^1 \frac{x^2 y^2}{2} \bigg|_x^{x^2} \, dy \, dx
\]

\[
= \frac{21}{8} \int_0^1 (x^4 - x^6) \, dx = \frac{21}{8} \times \frac{1}{5} = \frac{3}{20}.
\]

**Mixed bivariate distributions**

Suppose that \(X\) is discrete and \(Y\) is continuous, and there exists a function \(f(x, y)\) defined on \(\mathbb{R}^2\) such that for every pair \((A, B)\) subsets of real numbers,

\[
\Pr(X \in A \text{ and } Y \in B) = \int_B \sum_A f(x, y) \, dy.
\]

If the integral exists, \(f\) is called the **joint probability function** or **joint probability density function** of \(X\) and \(Y\). Integration and summation operators may be interchanged.

**Example 3.4.12** Suppose that \(X\) is the indicator variable of whether a patient has relapsed and \(P\) represents the probability that a patient will suffer a relapse. The joint p.d.f. is

\[
f(x, p) = p^x (1 - p)^{1-x}, \text{ for } x = 0, 1 \text{ and } 0 \leq p \leq 1.
\]

To compute \(\Pr(X = 0, p \leq .5)\), evaluate \(f(x, y)\) at \(x = 0\), and then integrate with respect to \(p\):

\[
\Pr(X = 0, p \leq .5) = \int_0^{.5} p^0 (1 - p)^1 \, dp = \int_0^{.5} (1 - p) \, dp = \frac{3}{8}.
\]

**Example 3.4.11** Suppose that the joint p.d.f. of \((X, Y)\) is

\[
f(x, y) = \frac{xy^{x-1}}{3}, \text{ for } x \in \{0, 1, 2\} \text{ and } 0 < y < 1.
\]
To check that $f$ satisfies the first condition, compute

\[
\sum_{x=1}^{3} \int_{0}^{1} \frac{xy^{x-1}}{3} \, dy = \sum_{x=1}^{3} \left. \frac{y^x}{3} \right|_{0}^{1} = \sum_{x=1}^{3} \frac{1}{3} = 1.
\]

To compute $\Pr(X > 1, Y \geq .5)$,

\[
\Pr(X > 1, Y \geq .5) = \sum_{x=2}^{3} \int_{.5}^{1} \frac{xy^{x-1}}{3} \, dy
= \sum_{x=2}^{3} \left. \frac{y^x}{3} \right|_{.5}^{1}
= \sum_{x=2}^{3} \frac{1}{3} \left[ 1 - \frac{1}{2^x} \right]
= .5411.
\]

Changing the order of operations confirms the answer:

\[
\Pr(X > 1, Y \geq .5) = \int_{.5}^{1} \sum_{x=2}^{3} \frac{xy^{x-1}}{3} \, dy
= \int_{.5}^{1} \left( \frac{2y}{3} + y^2 \right) \, dy
= \frac{1}{3} \left( y^2 + y^3 \right) \bigg|_{.5}
= .5411.
\]

**Bivariate cumulative distribution functions**

**Definition 3.4.6.** The joint distribution function or joint cumulative distribution function of random variables $X$ and $Y$ is

\[
F(x, y) = \Pr(X \leq x, Y \leq y), x, y \in \mathbb{R}.
\]

To compute the probability that $X$ and $Y$ lie in a rectangle corresponding to $a < X \leq b$ and $c < Y \leq d$, compute

\[
\Pr(a < X \leq b, c < Y \leq d) = F(b, d) - F(a, d) - F(b, d) + F(a, c).
\]

Four terms are needed since $F(b, d) = \Pr(X \leq b, Y \leq d)$, so we need to subtract off the unwanted probability mass.
Theorem 3.4.5. Suppose that \( X \) and \( Y \) have a joint distribution function \( F(x, y) \). The distribution function of \( X \) alone can be obtained from the relationship

\[
\lim_{y \to \infty} \Pr(X \leq x, Y < y) = \Pr(X \leq x).
\]

We utilize this relationship by evaluating \( F(x, y) \) at the upper boundary of the support of \((X,Y)\) with respect to \( y \). The following example demonstrates.

Example 3.4.14 Suppose that \((X,Y)\) have the following joint c.d.f.:

\[
F(x, y) = \begin{cases} 
\frac{xy(x + y)}{16}, & 0 \leq x \leq 2, 0 \leq y \leq 2, \\
1 & \text{if } 2 < x \text{ and } 2 < y, \\
0 & \text{otherwise}.
\end{cases}
\] (3)

Then,

\[
F_1(x) = \lim_{y \to \infty} F(x, y) = F(x, 2)
\]

\[
= \begin{cases} 
0 & x < 0, \\
\frac{2x(x + 2)}{16}, & 0 \leq x \leq 2, \\
1 & \text{if } 2 < x.
\end{cases}
\]

Suppose that \( X \) and \( Y \) have a continuous joint distribution with joint p.d.f. \( f \). Then, the joint c.d.f. at \((x, y)\) is

\[
F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(r, s) \, dr \, ds.
\]

Also,

\[
f(x, y) = \left. \frac{\partial^2 F}{\partial y \partial x} \right|_{(x,y)} = \left. \frac{\partial^2 F}{\partial x \partial y} \right|_{(x,y)}
\]

at every point where the second-order derivatives exist.

Returning to example 3.4.14 (formula 3),

\[
f(x, y) = \frac{1}{16} \left[ \frac{\partial^2 rs(r + s)}{\partial r \partial s} \right]_{(x,y)}
\]

\[
= \frac{1}{8} (r + s) \bigg|_{(r=x,s=y)}
\]

\[
= \begin{cases} 
\frac{x + y}{8}, & 0 \leq x \leq 2, 0 \leq y \leq 2, \\
0 & \text{otherwise}.
\end{cases}
\]