1 Discrete RVs

1.1 Joint distribution

Let $X, Y$ be random variables. To have all the information about $X, Y$ we need to know $P(X = i, Y = j)$ for all possible $i, j$ in the range of $X, Y$. However, it may NOT be that $X, Y$ are independent. In other words, knowing $P(X = i)$ and $P(Y = j)$ is not enough to determine $P(X = i, Y = j)$. In this case, we need to list all possible pairs of $P(X = i, Y = j)$ in a 2 dimensional table. This is the so-called joint distribution of $X$ and $Y$.

Example 1.1. Suppose 3 balls are randomly selected from an urn containing 3 red, 4 white and 5 blue. Let $X$ and $Y$ denote, respectively, the number of red and white balls in the sample. Find $P(X = i, Y = j)$ for all possible $(i, j)$.

Ans: It’s best to present these informations in a table. We’ll just list the answers here.

\[
p(0, 0) = \frac{\binom{3}{0}}{\binom{12}{3}}; p(0, 1) = \frac{\binom{3}{1}\binom{5}{2}}{\binom{12}{3}}; p(0, 2) = \frac{\binom{3}{2}\binom{4}{1}}{\binom{12}{3}}; p(0, 3) = \frac{\binom{3}{3}}{\binom{12}{3}} \\
p(1, 0) = \frac{\binom{3}{1}\binom{5}{2}}{\binom{12}{3}}; p(1, 1) = \frac{\binom{3}{1}\binom{4}{1}\binom{5}{1}}{\binom{12}{3}}; p(1, 2) = \frac{\binom{3}{1}\binom{4}{2}}{\binom{12}{3}} \\
p(2, 0) = \frac{\binom{3}{2}\binom{4}{1}}{\binom{12}{3}}; p(2, 1) = \frac{\binom{3}{2}\binom{4}{2}}{\binom{12}{3}} \\
p(3, 0) = \frac{\binom{3}{3}}{\binom{12}{3}}.
\]
For a joint distribution, one should check that

$$\sum_{i,j} P(X = i, Y = j) = 1.$$  

Note that the above is a double sum over $i, j$.

### 1.2 Marginal distribution

Intuitively, one can recover the information about $X$ from knowing both about $X, Y$. This can be done, via the following:

$$P(X = i) = \sum_j P(X = i, Y = j)$$  

since

$$\{X = i\} = \bigcup_j \{X = i, Y = j\}.$$  

Similarly, we have

$$P(Y = j) = \sum_i P(X = i, Y = j).$$

The distribution $P(X = i)$ is referred to as the marginal distribution of $X$. The distribution $P(Y = j)$ is referred to as the marginal distribution of $Y$.

### 2 Continuous random variables

#### 2.1 Joint density

Taking the motivation from the definition of a continuous RV, we say two RVs $X$ and $Y$ are *jointly continuous* if there exists a function $f(x, y)$ such that

$$f(x, y) \geq 0$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1.$$  

We refer to $f(x, y)$ as the joint density of $X, Y$. Then we define

$$P(x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2) = \int_{x_1}^{x_2} \int_{y_1}^{y_2} f(x, y) dx dy.$$
From which it follows that

\[ P(X = x, Y = y) = 0, \]

and thus

\[ P(x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2) = P(x_1 < X \leq x_2, y_1 \leq Y \leq y_2) \]
\[ = P(x_1 < X < x_2, y_1 \leq Y \leq y_2) \]
\[ = P(x_1 \leq X \leq x_2, y_1 < Y \leq y_2) \]
\[ = P(x_1 \leq X \leq x_2, y_1 \leq Y < y_2) \cdots \]

as well as

\[ P(X \leq x, Y \leq y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u,v) \, du \, dv. \]

The function \( F(x, y) = P(X \leq x, Y \leq y) \) is referred to as the joint cumulative distribution function of \( X \) and \( Y \). If \( X, Y \) are continuous, then \( F \) is jointly differentiable in \( x, y \) and

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

**Example 2.1.** (Uniform on a unit square)

Let \( X, Y \) have a joint density

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

Then for \( 0 \leq x \leq 1, 0 \leq y \leq 1 \),

\[ F(x, y) = \int_{0}^{x} \int_{0}^{y} dxdy = xy. \]

Note that \( F(x, y) \) is a constant in \( x \) for \( x \geq 1 \) since for \( x \geq 1 \) and \( 0 \leq y \leq 1 \)

\[ F(x, y) = \int_{0}^{1} \int_{0}^{y} dxdy = y, \]

agreeing with the fact that

\[ f(x, y) = \frac{\partial^2}{\partial x \partial y} F(x, y) = 0 \text{ for } x \geq 1. \]

Similarly \( F(x, y) \) is a constant in \( y \) for \( x \geq 1 \). And finally for \( 0 \leq x, y \leq 1 \)

\[ \frac{\partial^2}{\partial x \partial y} F(x, y) = \frac{\partial^2}{\partial x \partial y} xy = 1 = f(x, y). \]

3
2.2 Marginal distribution, marginal density

Again, we have information about $X, Y$. How can we find information about $X$ given this information? First, note that

$$P(X \leq x) = P(X \leq x, Y < \infty) = \lim_{y \to \infty} F(x, y),$$

since the set $\{Y < \infty\}$ has probability 1.

That is

$$P(X \leq x) = \int_{-\infty}^{x} \int_{-\infty}^{\infty} f(u, v) dudv$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(u, v) dvdu.$$

For a fix $u$, note that

$$\int_{-\infty}^{\infty} f(u, v) dv$$

is a function of $u$. Therefore, differentiating with respect to $x$ on both sides gives

$$f_X(x) = \int_{-\infty}^{\infty} f(x, v) dv$$

We call $f_X(x)$ the marginal density of $X$ and $F_X(x) = P(X \leq x)$ the marginal cdf of $X$. Similarly we have

$$f_Y(y) = \int_{-\infty}^{\infty} f(u, y) du,$$

and the cumulative distribution function of $Y$ is

$$P(Y \leq y) = \int_{-\infty}^{y} f_Y(u) du = \int_{-\infty}^{\infty} f(u, v) dvdu.$$

3 Independence

We say, generally that 2 RVs $X, Y$ are independent if for any subsets $A, B$ of the real line:

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B).$$

In the case that $X, Y$ are either jointly discrete or jointly continuous, we can say more.
3.1 Discrete RV

Let \( X, Y \) be discrete RVs with joint distribution \( P(X = i, Y = j) \). Then \( X, Y \) are independent if and only if
\[
P(X = i, Y = j) = P(X = i)P(Y = j).
\]

Proof.
\[
P(X \in A, Y \in B) = \sum_{x \in A} \sum_{y \in B} P(X = x, Y = y) = \sum_{x \in A} P(X = x) \sum_{y \in B} P(Y = y) = P(X \in A)P(Y \in B).
\]

Example 3.1. Suppose that \( n + m \) independent trials having a common probability success \( p \) are performed. If \( X \) is the number of successes in the first \( n \) trials, \( Y \) the number of successes in \( m \) trials, then \( X \) and \( Y \) are independent. However, let \( Z \) be the number of total successes. Then \( X \) and \( Z \) are NOT independent.

3.2 Continuous RV

Let \( X, Y \) be discrete RVs with joint density \( f_{XY}(x, y) \). Then \( X, Y \) are independent if and only if
\[
f_{XY}(x, y) = f_X(x)f_Y(y).
\]

Proof.
\[
P(X \in A, Y \in B) = \int_A \int_B f_{XY}(x, y) \, dx \, dy = \int_A \int_B f_X(x)f_Y(y) \, dx \, dy = \int_A f_X(x) \int_B f_Y(y) = P(X \in A)P(Y \in B).
\]

Example 3.2. Two persons decide to meet at a certain location. If each of them independently arrives at a time uniformly distributed between 12 - 1 pm, find the probability that the first to arrive has to wait longer than 10 minutes.

Ans: Let \( X, Y \) denote the time the first and the second person arrives. Then \( X, Y \) are independent Uniform(0,60). We want to compute
\[
P(X + 10 < Y) + P(Y + 10 < X) = 2P(X + 10 < Y),
\]
by symmetry. We have

\[ 2P(X + 10 < Y) = 2 \int \int_{x+10<y} f(x,y) \, dx \, dy \]
\[ = 2 \int_{10}^{60} \int_{0}^{y-10} (1/60)^2 \, dx \, dy \]
\[ = \frac{2}{60^2} \int_{10}^{60} (y - 10) \, dy \]
\[ = \frac{25}{36}. \]

**Example 3.3.** If the joint density function of \(X\) and \(Y\) is

\[ f(x,y) = 6e^{-2x}e^{-3y}, \quad 0 < x < \infty, \quad 0 < y < \infty, \]
\[ = 0 \quad \text{otherwise} \]

are they independent? What if

\[ f(x,y) = 24xy, \quad 0 < x, y < 1, \quad 0 < x + y < 1 \]
\[ = 0 \quad \text{otherwise} \]

Ans: The RVs are independent in the first case and not in the second. The reason is if we denote

\[ 1(x,y) = 1 \quad \text{if} \quad 0 < x, y < 1, \quad 0 < x + y < 1 \]
\[ = 0 \quad \text{otherwise} \]

then we see that for the second case

\[ f(x,y) = 24x1(x,y), \]

and clearly the function \(1(x,y)\) does not factor.