JOINT AND CONDITIONAL DISTRIBUTIONS, STOCHASTIC INDEPENDENCE, MORE EXPECTATION

JOINT DISTRIBUTION FUNCTIONS

2.1 Cumulative Distribution Function

Definition 1 Joint cumulative distribution function Let $X_1, X_2, \ldots, X_k$ be k random variables all defined on the same probability space $(\Omega, \mathscr{A}, P[\cdot])$. The *joint cumulative distribution function* of $X_1, \ldots, X_k$, denoted by $F_{X_1, \ldots, X_k}(\cdot, \ldots, \cdot)$, is defined as $P[X_1 \leq x_1; \ldots; X_k \leq x_k]$ for all $(x_1, x_2, \ldots, x_k)$. ////

Thus a joint cumulative distribution function is a function with domain euclidean k space and counterdomain the interval [0, 1]. If $k = 2$, the joint cumulative distribution function is a function of two variables, and so its

of the downturned numbers. The goal is to find $F_{X,Y}(\cdot, \cdot)$, the joint cumulative distribution function of X and Y. Observe first that the random variables X and Y jointly take on only the values

$$\begin{array}{l} (1, 1), (1, 2), (1, 3), (1, 4), \\ (2, 2), (2, 3), (2, 4), \\ (3, 3), (3, 4), \\ (4, 4). \end{array}$$

(The first component is the value of X, and the second the value of Y.)

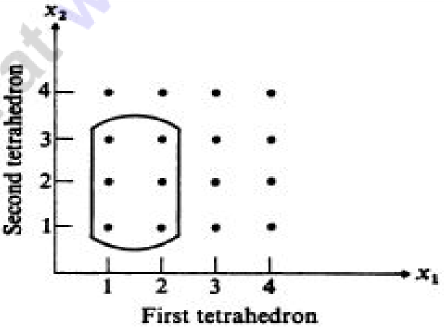

FIGURE 1
Sample space for experiment of tossing two tetrahedra.

The sample space for this experiment is displayed in Fig. 1. The 16 sample points are assumed to be equally likely. Our objective is to find $F_{X,Y}(x, y)$ for each point (x, y). As an example let $(x, y) = (2, 3)$, and find $F_{X,Y}(2, 3) = P[X \le 2;\ Y \le 3]$. Now the event $\{X \le 2 \text{ and } Y \le 3\}$ corresponds to the encircled sample points in Fig. 1; hence $F_{X,Y}(2, 3) = \frac{6}{16}$. Similarly, $F_{X,Y}(x, y)$ can be found for other values of x and y. $F_{X,Y}(x, y)$ is tabled in Fig. 2. ////

We saw that the cumulative distribution function of a unidimensional random variable had certain properties; the same is true of a joint cumulative. We shall list these properties for the joint cumulative distribution function of two random variables; the generalization to k dimensions is straightforward.

TABLE OF VALUES OF $F_{X,Y}(x,y)$

$4 \le y$	0	$\frac{4}{16}$	$\frac{8}{16}$	$\frac{12}{16}$	1
$3 \le y < 4$	0	$\frac{3}{16}$	$\frac{6}{16}$	$\frac{9}{16}$	$\frac{9}{16}$
$2 \le y < 3$	0	$\frac{2}{16}$	$\frac{4}{16}$	$\frac{4}{16}$	$\frac{4}{16}$
$1 \le y < 2$	0	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$
$y < 1$	0	0	0	0	0
	$x < 1$	$1 \le x < 2$	$2 \le x < 3$	$3 \le x < 4$	$4 \le x$

Properties of bivariate cumulative distribution function $F(\cdot, \cdot)$

(i) $F(-\infty, y) = \lim_{x \to -\infty} F(x, y) = 0$ for all y, $F(x, -\infty) = \lim_{y \to -\infty} F(x, y) = 0$ for all x, and $\lim_{\substack{x \to \infty \\ y \to \infty}} F(x, y) = F(\infty, \infty) = 1$.

(ii) If $x_1 < x_2$ and $y_1 < y_2$, then $P[x_1 < X \le x_2;\ y_1 < Y \le y_2]$ $= F(x_2, y_2) - F(x_2, y_1) - F(x_1, y_2) + F(x_1, y_1) \ge 0$.

(iii) $F(x, y)$ is right continuous in each argument; that is,
$$\lim_{0 < h \to 0} F(x + h, y) = \lim_{0 < h \to 0} F(x, y + h) = F(x, y).$$

We will not prove these properties. Property (ii) is a *monotonicity* property of sorts; it is not equivalent to $F(x_1, y_1) \le F(x_2, y_2)$ for $x_1 \le x_2$ and $y_1 \le y_2$.

Definition 2 Bivariate cumulative distribution function Any function satisfying properties (i) to (iii) is defined to be a *bivariate cumulative distribution function* without reference to any random variables. ////

Definition 3 Marginal cumulative distribution function If $F_{X,Y}(\cdot, \cdot)$ is the joint cumulative distribution function of X and Y, then the cumulative distribution functions $F_X(\cdot)$ and $F_Y(\cdot)$ are called *marginal cumulative distribution functions.* ////

Remark $F_X(x) = F_{X,Y}(x, \infty)$, and $F_Y(y) = F_{X,Y}(\infty, y)$; that is, knowledge of the joint cumulative distribution function of X and Y implies knowledge of the two marginal cumulative distribution functions. ////

The converse of the above remark is not generally true; in fact, an example (Example 8) will be given in Subsec. 2.3 below that gives an entire family of joint cumulative distribution functions, and each member of the family has the same marginal distributions.

We will conclude this section with a remark that gives an inequality involving the joint cumulative distribution and marginal distributions. The proof is left as an exercise.

Remark $F_X(x) + F_Y(y) - 1 \leq F_{X,Y}(x, y) \leq \sqrt{F_X(x)F_Y(y)}$ for all x, y. ////

2.2 Joint Density Functions for Discrete Random Variables

If $X_1, X_2, \ldots, X_k$ are random variables defined on the same probability space, then $(X_1, X_2, \ldots, X_k)$ is called a *k-dimensional random variable.*

2.2 Joint Density Functions for Discrete Random Variables

If $X_1, X_2, \ldots, X_k$ are random variables defined on the same probability space, then $(X_1, X_2, \ldots, X_k)$ is called a *k-dimensional random variable*.

Definition 4 Joint discrete random variables The k-dimensional random variable $(X_1, X_2, \ldots, X_k)$ is defined to be a *k-dimensional discrete random variable* if it can assume values only at a countable number of points $(x_1, x_2, \ldots, x_k)$ in k-dimensional real space. We also say that the random variables $X_1, X_2, \ldots, X_k$ are *joint discrete random variables.*

Definition 5 Joint discrete density function If $(X_1, X_2, \ldots, X_k)$ is a k-dimensional discrete random variable, then the *joint discrete density function* of $(X_1, X_2, \ldots, X_k)$, denoted by $f_{X_1, X_2, \ldots, X_k}(\cdot, \cdot, \ldots, \cdot)$, is defined to be

$$f_{X_1, X_2, \ldots, X_k}(x_1, x_2, \ldots, x_k) = P[X_1 = x_1; X_2 = x_2; \ldots; X_k = x_k]$$

for $(x_1, x_2, \ldots, x_k)$, a value of $(X_1, X_2, \ldots, X_k)$ and is defined to be 0 otherwise. ////

Remark $\sum f_{X_1, \ldots, X_k}(x_1, \ldots, x_k) = 1$, where the summation is over all possible values of $(X_1, \ldots, X_k)$. ////

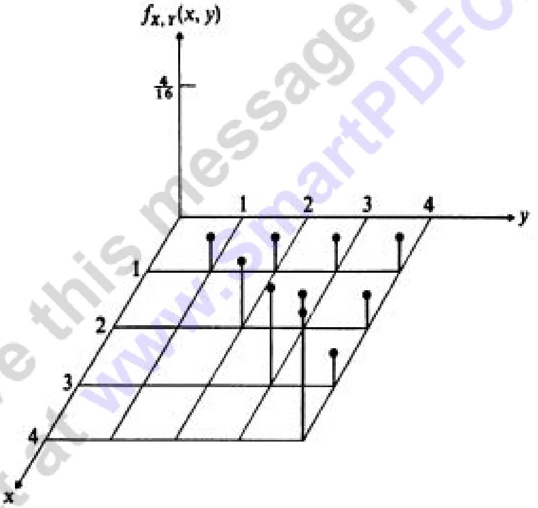

FIGURE 4

EXAMPLE 2 Let X denote the number on the downturned face of the first tetrahedron and Y the larger of the downturned numbers in the experiment of tossing two tetrahedra. The values that (X, Y) can take on are (1, 1), (1, 2), (1, 3), (1, 4), (2, 2), (2, 3), (2, 4), (3, 3), (3, 4), and (4, 4); hence X and Y are jointly discrete. The joint discrete density function of X and Y is given in Fig. 4.

In tabular form it is given as

(x, y)	$(1, 1)$	$(1, 2)$	$(1, 3)$	$(1, 4)$	$(2, 2)$	$(2, 3)$	$(2, 4)$	$(3, 3)$	$(3, 4)$	$(4, 4)$
$f_{X,Y}(x,y)$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{3}{16}$	$\frac{1}{16}$	$\frac{4}{16}$

or in another tabular form as

4	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{4}{16}$
3	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{3}{16}$	
2	$\frac{1}{16}$	$\frac{2}{16}$		
1	$\frac{1}{16}$			
y / x	1	2	3	4

////

Theorem 1 If X and Y are jointly discrete random variables, then knowledge of $F_{X,Y}(\cdot, \cdot)$ is equivalent to knowledge of $f_{X,Y}(\cdot, \cdot)$. Also, the statement extends to k-dimensional discrete random variables.

Definition 6 Marginal discrete density If X and Y are jointly discrete random variables, then $f_X(\cdot)$ and $f_Y(\cdot)$ are called *marginal* discrete density functions. More generally, let $X_{i_1}, \ldots, X_{i_m}$ be any subset of the jointly discrete random variables $X_1, \ldots, X_k$; then $f_{X_{i_1}, \ldots, X_{i_m}}(x_{i_1}, \ldots, x_{i_m})$ is also called a *marginal density*. ////

Remark If $X_1, \ldots, X_k$ are jointly discrete random variables, then any marginal discrete density can be found from the joint density, but not conversely. For example, if X and Y are jointly discrete with values $(x_1, y_1), (x_2, y_2), \ldots$, then

$$f_X(x_k) = \sum_{\{i:\, x_i = x_k\}} f_{X,Y}(x_i, y_i) \quad \text{and} \quad f_Y(y_k) = \sum_{\{i:\, y_i = y_k\}} f_{X,Y}(x_i, y_i). \qquad ////$$

Heretofore we have indexed the values of (X, Y) with a single index, namely i. That is, we listed values as $(x_1, y_1), (x_2, y_2), \ldots, (x_i, y_i), \ldots$. The values of (X, Y) could also be indexed by using separate indices for the X and Y values. For instance, we could let i index the possible X values, say $x_1, \ldots, x_i, \ldots$, and j index the possible Y values, say $y_1, \ldots, y_j, \ldots$. Then the values of (X, Y) would be a subset of the points (x_i, y_j) for $i = 1, 2, \ldots$ and $j = 1, 2, \ldots$. If this latter method of indexing is used, then the marginal density of X is obtained as follows:

$$f_X(x_k) = \sum_j f_{X,Y}(x_k, y_j),$$

where the summation is over all y_j for the fixed x_k. The marginal density of Y is analogously obtained. The following example may help to clarify these two different methods of indexing the values of (X, Y).

EXAMPLE 3 Return to the experiment of tossing two tetrahedra, and define X as the number on the downturned face of the first tetrahedron and Y as the larger of the numbers on the two downturned faces. The joint density of X and Y is given in Fig. 4. The values of (X, Y) can be listed as (1, 1), (1, 2), (1, 3), (1, 4), (2, 2), (2, 3), (2, 4), (3, 3), (3, 4), and (4, 4), 10 points in all. Or, if we note that X has values 1, 2, 3, and 4; Y has values 1, 2, 3, and 4; and Y is greater than or equal X, the values of (X, Y) are $\{(i, j): i = 1, \ldots, 4; j = 1, \ldots, 4; \text{and } i \le j\}$. Let us use each of these methods of indexing to evaluate $F_{X,Y}(2, 3)$ from the joint density. Under the first method of indexing,

$$\begin{aligned} F_{X,Y}(2, 3) &= \sum_{\{i: x_i \le 2, y_i \le 3\}} f_{X,Y}(x_i, y_i) \\ &= f_{X,Y}(1, 1) + f_{X,Y}(1, 2) \\ &\quad + f_{X,Y}(1, 3) + f_{X,Y}(2, 2) + f_{X,Y}(2, 3) = \tfrac{6}{16} \end{aligned}$$

Under the second method of indexing,

$$F_{X,Y}(2, 3) = \sum_{i=1}^{2} \sum_{j=i}^{3} f_{X,Y}(i, j) = \tfrac{6}{16}.$$

Similarly, all other values of $F_{X,Y}(\cdot, \cdot)$ could be obtained. Also

$$\begin{aligned} f_Y(3) &= \sum_{\{i: y_i = 3\}} f_{X,Y}(x_i, y_i) = f_{X,Y}(1, 3) + f_{X,Y}(2, 3) + f_{X,Y}(3, 3) \\ &= \tfrac{1}{16} + \tfrac{1}{16} + \tfrac{3}{16} = \tfrac{5}{16}. \end{aligned}$$

Similarly $f_Y(1) = \frac{1}{16}$, $f_Y(2) = \frac{3}{16}$, and $f_Y(4) = \frac{7}{16}$, which together with $f_Y(3) = \frac{5}{16}$ give the marginal discrete density function of Y. ////

EXAMPLE 4 We mentioned that marginal densities can be obtained from the joint density, but not conversely. The following is an example of a family of joint densities that all have the same marginals, and hence we see that in general the joint density is not uniquely determined from knowledge of the marginals. Consider altering the joint density given in the previous examples as follows:

4	$\frac{1}{16} + \varepsilon$	$\frac{1}{16} - \varepsilon$	$\frac{1}{16}$	$\frac{4}{16}$
3	$\frac{1}{16} - \varepsilon$	$\frac{1}{16} + \varepsilon$	$\frac{3}{16}$	
2	$\frac{1}{16}$	$\frac{2}{16}$		
1	$\frac{1}{16}$			
$y \diagup x$	1	2	3	4