Jointly distributed random variables

Consider collections of 2 or more random variables $X_1, X_2, \ldots, X_n$.
Interested in modeling relationships between them as well.

**Examples:**
- $X_1$ = price of stock 1, $X_2$ = price of stock 2, etc.
- $X_1$ = price today, $X_2$ = price yesterday, etc.
- $X_1$ = expenditures on food, $X_2$ = expenditures on housing, etc.
- $X_1$ = cholesterol level, $X_2$ = blood pressure, etc.
- $X_1$ = rainfall in IL, $X_2$ = rainfall in IN, etc.

Most more advanced
- statistical topics (time series, multivariate analysis, multiple linear regression, factor models, etc) and
- probability topics (Markov chains, stochastic processes, etc)
involve collections of random variables.
Jointly distributed random variables

**Focus on:** Two random variables $X, Y$. All probability questions about $X$ and $Y$ can be answered in terms of their joint c.d.f.

**Joint cumulative distribution function (c.d.f.):**

$$F(a, b) = P(X \leq a, Y \leq b), \quad -\infty < a, b < \infty.$$ 

**For example:** $F$ carries info about $X, Y$ individually: e.g.

$$F_X(a) =$$

But also: e.g.

$$P(X > a, Y > b) =$$

$$P(a_1 < X \leq a_2, b_1 < Y \leq b_2) =$$

Two broad classes of random variables

1. Both $X$ and $Y$ are discrete: characterized through *joint probability mass function* (p.m.f.)

   $$p(x, y) = P(X = x, Y = y).$$

   For example, $p_X(x) = P(X = x) = \sum_y p(x, y)$, etc.

2. $X$ and $Y$ are jointly continuous: there is a non-negative function $f(x, y)$, called *joint probability density function* (p.d.f.), such that, for any set $C$ in the two-dimensional plane,

   $$P((X, Y) \in C) = \int\int_{(x,y) \in C} f(x, y)\,dx\,dy.$$

Next: A number of notes for the jointly continuous case.
Jointly distributed random variables

Note 1: $\int \int_{(x,y) \in C} f(x,y) \, dx \, dy$ is the volume under the surface $f(x,y)$ above the region $C$. In particular, when $f \equiv 1$,

$$\int \int_{(x,y) \in C} \, dx \, dy = \text{Area}(C).$$

Note 2: With $C = A \times B = \{(x,y) : x \in A, y \in B\}$,

$$\mathbb{P}(X \in A, Y \in B) = \int_A \int_B \, dy \, f(x,y)$$

Note 3:

$$F(a,b) = \int_{-\infty}^{a} \, dx \int_{-\infty}^{b} \, dy \, f(x,y), \quad \frac{\partial^2}{\partial a \partial b} F(a,b) = f(a,b)$$

Jointly distributed random variables

Note 4:

$$\mathbb{P}(a < X \leq a + da, b < Y \leq b + db)$$

$$= \int_{a}^{a+da} \, dx \int_{b}^{b+db} \, dy \, f(x,y) \approx f(a,b) \, dadb$$

for small $da, db$, if $f$ is continuous at $(a,b)$. Thus, $f(a,b)$ is a measure of how likely $X, Y$ is near $a, b$.

Note 5: Each individual random variable is continuous. E.g.

$$\mathbb{P}(X \in A) = \mathbb{P}(X \in A, Y \in (-\infty, \infty)) = \int_A \int_{-\infty}^{\infty} \, dy \, f(x,y)$$

and hence the (marginal) density of $X$ is

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

Similarly, $f_Y(y) = \int_{-\infty}^{\infty} f(x,y) \, dx$. 
Problem 1a

Two fair dice are rolled. Find the joint probability mass function of $X$ and $Y$ when $X$ is the largest value obtained on any die and $Y$ is the sum of the values.

Example 1c(b)

The joint density function of $X$ and $Y$ is given by

$$f(x, y) = \begin{cases} 2e^{-x}e^{-2y}, & 0 < x < \infty, 0 < y < \infty \\ 0, & \text{otherwise} \end{cases}$$

Compute $P(X < Y)$. 
Problem 8

The joint density function of $X$ and $Y$ is given by

$$f(x, y) = \begin{cases} 
    c(y^2 - x^2)e^{-y}, & -y \leq x \leq y, 0 < y < \infty \\
    0, & \text{otherwise}
\end{cases}$$

(a) Find $c$.
(b) Find the marginal densities of $X$ and $Y$.
(c) Find $\mathbb{E}X$,
(d) $P(0 < X < 1, Y < 1)$. 

Problem 8 cont’d:
More than two random variables

The notions above can be extended to more than two random variables $X_1, X_2, \ldots, X_n$. For example, the joint c.d.f. is defined as

$$F(a_1, a_2, \ldots, a_n) = \mathbb{P}(X_1 \leq a_1, X_2 \leq a_2, \ldots, X_n \leq a_n).$$

The random variables $X_1, X_2, \ldots, X_n$ are jointly continuous if there is a non-negative function $f(x_1, x_2, \ldots, x_n)$, called joint probability density function (p.d.f.), such that, for any set $C$ in the $n$-dimensional space,

$$\mathbb{P}((X_1, X_2, \ldots, X_n) \in C) = \int \int \ldots \int f(x_1, x_2, \ldots, x_n)\,dx_1\,dx_2\ldots\,x_n.$$
That's all Folks!