5.4 Joint distributions and independence (discrete)

Most applications of probability to engineering statistics involve not one but several random variables. In some cases, the application is intrinsically multivariate.

Example 5.32. Consider the assembly of a ring bearing with nominal inside diameter 1.00 in. on a rod with nominal diameter .99 in. If

$$
\begin{aligned}
X &= \text{the ring bearing inside diameter} \\
Y &= \text{the rod diameter}
\end{aligned}
$$

One might be interested in

$$P[\text{there is an interference in assembly}] =$$

Even when a situation is univariate, samples larger than size 1 are essentially always used in engineering applications. The n data values in a sample are usually thought of as subject to chance and their simultaneous behavior must then be modeled.

This is actually a very broad and difficult subject, we will only cover a brief introduction to the topic: **jointly discrete random variables**.

5.4.1 Joint distributions

For several discrete random variable, the device typically used to specify probabilities is a *joint probability function.* The two-variable version of this is defined.

Definition 5.21. A *joint probability function (joint pmf)* for discrete random variables X and Y is a nonnegative function $f(x, y)$, giving the probability that (simultaneously) X takes the values x and Y takes the values y. That is,

$$f(x, y) = P[X = x \text{ and } Y = y]$$

Properties:

1.

2.

For the discrete case, it is useful to give $f(x, y)$ in a **table**.

Example 5.33 (Two bolt torques, cont'd). Recall the example of measure the bolt torques on the face plates of a heavy equipment component to the nearest integer. With

$$X = \text{the next torque recorded for bolt 3}$$

$$Y = \text{the next torque recorded for bolt 4}$$

the joint probability function, $f(x, y)$, is

y\x	11	12	13	14	15	16	17	18	19	20
20	0	0	0	0	0	0	0	2/34	2/34	1/34
19	0	0	0	0	0	0	2/34	0	0	0
18	0	0	1/34	1/34	0	0	1/34	1/34	1/34	0
17	0	0	0	0	2/34	1/34	1/34	2/34	0	0
16	0	0	0	1/34	2/34	2/34	0	0	2/34	0
15	1/34	1/34	0	0	3/34	0	0	0	0	0
14	0	0	0	0	1/34	0	0	2/34	0	0
13	0	0	0	0	1/34	0	0	0	0	0

$P[X = 18 \text{ and } Y = 17]$

$P[X = 14 \text{ and } Y = 19]$

By summing up certain values of $f(x, y)$, probabilities associated with X and Y with patterns of interest can be obtained.

Consider:

$$P[X \geq Y]$$

y\x	11	12	13	14	15	16	17	18	19	20
20										
19										
18										
17										
16										
15										
14										
13										

$$P[|X - Y| \leq 1]$$

y\x	11	12	13	14	15	16	17	18	19	20
20										
19										
18										
17										
16										
15										
14										
13										

$$P[X = 17]$$

y\x	11	12	13	14	15	16	17	18	19	20
20										
19										
18										
17										
16										
15										
14										
13										

5.4.2 Maginal distributions

In a bivariate problem, once can add down columns in the (two-way) table of $f(x, y)$ to get values for the probability function of X, $f_X(x)$ and across rows in the same table to get values for the probability distribution of Y, $f_Y(y)$.

Definition 5.22. The individual probability functions for discrete random variables X and Y with joint probability function $f(x, y)$ are called *marginal probability functions.* They are obtained by summing $f(x, y)$ values over all possible values of the other variable.

$$f_X(x) = \sum_y f(x, y)$$
$$f_Y(y) = \sum_x f(x, y)$$

Example 5.34 (Torques, cont'd)**.** Find the marginal probability functions for X and Y from the following joint pmf.

y\x	11	12	13	14	15	16	17	18	19	20
20	0	0	0	0	0	0	0	2/34	2/34	1/34
19	0	0	0	0	0	0	2/34	0	0	0
18	0	0	1/34	1/34	0	0	1/34	1/34	1/34	0
17	0	0	0	0	2/34	1/34	1/34	2/34	0	0
16	0	0	0	1/34	2/34	2/34	0	0	2/34	0
15	1/34	1/34	0	0	3/34	0	0	0	0	0
14	0	0	0	0	1/34	0	0	2/34	0	0
13	0	0	0	0	1/34	0	0	0	0	0

Getting marginal probability functions from joint probability functions begs the question whether the process can be reversed. **Can we find joint probability functions from marginal probability functions?**

5.4.3 Conditional distributions

When working with several random variables, it is often useful to think about what is expected of one of the variables, given the values assumed by all others.

Definition 5.23. For discrete random variables X and Y with joint probability function $f(x, y)$, the *conditional probability function of* X *given* $Y = y$ is the function of x

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{f(x,y)}{\sum_x f(x,y)}$$

and the *conditional probability function of* Y *given* $X = x$ is the function of y

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{f(x,y)}{\sum_y f(x,y)}.$$

Example 5.35 (Torque, cont'd)**.** For the torque example with the following joint distribution, find the following:

1. $f_{Y|X}(20|18)$

2. $f_{Y|X}(y|15)$

3. $f_{Y|X}(y|20)$

4. $f_{X|Y}(x|18)$

$y\backslash x$	11	12	13	14	15	16	17	18	19	20	$f_Y(y)$
20	0/34	0/34	0/34	0/34	0/34	0/34	0/34	2/34	2/34	1/34	5/34
19	0/34	0/34	0/34	0/34	0/34	0/34	2/34	0/34	0/34	0/34	2/34
18	0/34	0/34	1/34	1/34	0/34	0/34	1/34	1/34	1/34	0/34	5/34
17	0/34	0/34	0/34	0/34	2/34	1/34	1/34	2/34	0/34	0/34	6/34
16	0/34	0/34	0/34	1/34	2/34	2/34	0/34	0/34	2/34	0/34	7/34
15	1/34	1/34	0/34	0/34	3/34	0/34	0/34	0/34	0/34	0/34	5/34
14	0/34	0/34	0/34	0/34	1/34	0/34	0/34	2/34	0/34	0/34	3/34
13	0/34	0/34	0/34	0/34	1/34	0/34	0/34	0/34	0/34	0/34	1/34
$f_X(x)$	1/34	1/34	1/34	2/34	9/34	3/34	4/34	7/34	5/34	1/34	34/34

5.4.4 Independence

Recall the following joint distribution:

$y \backslash x$	1	2	3	$f_Y(y)$
3	0.08	0.08	0.04	0.20
2	0.16	0.16	0.08	0.40
1	0.16	0.16	0.08	0.40
$f_X(x)$	0.40	0.40	0.20	1.00

What do you notice?

Definition 5.24. Discrete random variables X and Y are *independent* if their joint distribution function $f(x, y)$ is the product of their respective marginal probability functions. This is, independence means that

$$f(x, y) = f_X(x) f_Y(y) \qquad \text{for all } x, y.$$

If this does not hold, then X and Y are *dependent.*

Alternatively, discrete random variables X and Y are independent if for all x and y,

If X and$Y$ are not only independent but also have the same marginal distribution, then they are **independent and identically distributed (iid)**.

5.5 Functions of several random variables

5.5.1 Linear combinations

5.5.2 Central limit theorem